MEASURE THEORY

Volume 2

D.H.Fremlin

By the same author:
Topological Riesz Spaces and Measure Theory, Cambridge University Press, 1974.
Consequences of Martin's Axiom, Cambridge University Press, 1982.

Companions to the present volume:
Measure Theory, vol. 1, Torres Fremlin, 2000;
Measure Theory, vol. 3, Torres Fremlin, 2002;
Measure Theory, vol. 4, Torres Fremlin, 2003;
Measure Theore, vol. 5, Torres Fremlin, 2008.

First edition May 2001

Second edition January 2010

MEASURE THEORY

Volume 2

Broad Foundations

D.H.Fremlin

Research Professor in Mathematics, University of Essex

Dedicated by the Author
to the Publisher

This book may be ordered from the printers, `http://www.lulu.com/buy`.

First published in 2001
by Torres Fremlin, 25 Ireton Road, Colchester CO3 3AT, England

Library of Congress classification QA312.F72

AMS 2010 classification 28-01

ISBN 978-0-9538129-7-4

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

Printed by Lulu.com

Contents

General introduction In this treatise I aim to give a comprehensive description of modern abstract measure theory, with some indication of its principal applications. The first two volumes are set at an introductory level; they are intended for students with a solid grounding in the concepts of real analysis, but possibly with rather limited detailed knowledge. As the book proceeds, the level of sophistication and expertise demanded will increase; thus for the volume on topological measure spaces, familiarity with general topology will be assumed. The emphasis throughout is on the mathematical ideas involved, which in this subject are mostly to be found in the details of the proofs.

My intention is that the book should be usable both as a first introduction to the subject and as a reference work. For the sake of the first aim, I try to limit the ideas of the early volumes to those which are really essential to the development of the basic theorems. For the sake of the second aim, I try to express these ideas in their full natural generality, and in particular I take care to avoid suggesting any unnecessary restrictions in their applicability. Of course these principles are to to some extent contradictory. Nevertheless, I find that most of the time they are very nearly reconcilable, *provided* that I indulge in a certain degree of repetition. For instance, right at the beginning, the puzzle arises: should one develop Lebesgue measure first on the real line, and then in spaces of higher dimension, or should one go straight to the multidimensional case? I believe that there is no single correct answer to this question. Most students will find the one-dimensional case easier, and it therefore seems more appropriate for a first introduction, since even in that case the technical problems can be daunting. But certainly every student of measure theory must at a fairly early stage come to terms with Lebesgue area and volume as well as length; and with the correct formulations, the multidimensional case differs from the one-dimensional case only in a definition and a (substantial) lemma. So what I have done is to write them both out (§§114-115). In the same spirit, I have been uninhibited, when setting out exercises, by the fact that many of the results I invite students to look for will appear in later chapters; I believe that throughout mathematics one has a better chance of understanding a theorem if one has previously attempted something similar alone.

The plan of the work is as follows:

Volume 1: The Irreducible Minimum
Volume 2: Broad Foundations
Volume 3: Measure Algebras
Volume 4: Topological Measure Spaces
Volume 5: Set-theoretic Measure Theory.

Volume 1 is intended for those with no prior knowledge of measure theory, but competent in the elementary techniques of real analysis. I hope that it will be found useful by undergraduates meeting Lebesgue measure for the first time. Volume 2 aims to lay out some of the fundamental results of pure measure theory (the Radon-Nikodým theorem, Fubini's theorem), but also gives short introductions to some of the most important applications of measure theory (probability theory, Fourier analysis). While I should like to believe that most of it is written at a level accessible to anyone who has mastered the contents of Volume 1, I should not myself have the courage to try to cover it in an undergraduate course, though I would certainly attempt to include some parts of it. Volumes 3 and 4 are set at a rather higher level, suitable to postgraduate courses; while Volume 5 assumes a wide-ranging competence over large parts of analysis and set theory.

There is a disclaimer which I ought to make in a place where you might see it in time to avoid paying for this book. I make no attempt to describe the history of the subject. This is not because I think the history uninteresting or unimportant; rather, it is because I have no confidence of saying anything which would not be seriously misleading. Indeed I have very little confidence in anything I have ever read concerning the history of ideas. So while I am happy to honour the names of Lebesgue and Kolmogorov and Maharam in more or less appropriate places, and I try to include in the bibliographies the works which I have myself consulted, I leave any consideration of the details to those bolder and better qualified than myself.

For the time being, at least, printing will be in short runs. I hope that readers will be energetic in commenting on errors and omissions, since it should be possible to correct these relatively promptly. An inevitable consequence of this is that paragraph references may go out of date rather quickly. I shall be most flattered if anyone chooses to rely on this book as a source for basic material; and I am willing to attempt to maintain a concordance to such references, indicating where migratory results have come to rest for the moment, if authors will supply me with copies of papers which use them. In the concordance to the present volume you will find notes on the items which have been referred to in other published volumes of this work.

I mention some minor points concerning the layout of the material. Most sections conclude with lists of 'basic exercises' and 'further exercises', which I hope will be generally instructive and occasionally entertaining. How many of these you should attempt must be for you and your teacher, if any, to decide, as no two students will have quite the same needs. I mark with a **>** those which seem to me to be particularly important. But while you may not need to write out solutions to all the 'basic exercises', if you are in any doubt as to your capacity to do so you should take this as a warning to slow down a bit. The 'further exercises' are unbounded in difficulty, and are unified only by a presumption that each has at least one solution based on ideas already introduced. Occasionally I add a final 'problem', a question to which I do not know the answer and which seems to arise naturally in the course of the work.

The impulse to write this book is in large part a desire to present a unified account of the subject. Cross-references are correspondingly abundant and wide-ranging. In order to be able to refer freely across the whole text, I have chosen a reference system which gives the same code name to a paragraph wherever it is being called from. Thus 132E is the fifth paragraph in the second section of the third chapter of Volume 1, and is referred to by that name throughout. Let me emphasize that cross-references are supposed to help the reader, not distract her. Do not take the interpolation '(121A)' as an instruction, or even a recommendation, to lift Volume 1 off the shelf and hunt for §121. If you are happy with an argument as it stands, independently of the reference, then carry on. If, however, I seem to have made rather a large jump, or the notation has suddenly become opaque, local cross-references may help you to fill in the gaps.

Each volume will have an appendix of 'useful facts', in which I set out material which is called on somewhere in that volume, and which I do not feel I can take for granted. Typically the arrangement of material in these appendices is directed very narrowly at the particular applications I have in mind, and is unlikely to be a satisfactory substitute for conventional treatments of the topics touched on. Moreover, the ideas may well be needed only on rare and isolated occasions. So as a rule I recommend you to ignore the appendices until you have some direct reason to suppose that a fragment may be useful to you.

During the extended gestation of this project I have been helped by many people, and I hope that my friends and colleagues will be pleased when they recognise their ideas scattered through the pages below. But I am especially grateful to those who have taken the trouble to read through earlier drafts and comment on obscurities and errors.

Introduction to Volume 2

For this second volume I have chosen seven topics through which to explore the insights and challenges offered by measure theory. Some, like the Radon-Nikodým theorem (Chapter 23) are necessary for any understanding of the structure of the subject; others, like Fourier analysis (Chapter 28) and the discussion of function spaces (Chapter 24) demonstrate the power of measure theory to attack problems in general real and functional analysis. But all have applications outside measure theory, and all have influenced its development. These are the parts of measure theory which any analyst may find himself using.

Every topic is one which ideally one would wish undergraduates to have seen, but the length of this volume makes it plain that no ordinary undergraduate course could include very much of it. It is directed rather at graduate level, where I hope it will be found adequate to support all but the most ambitious courses in measure theory, though it is perhaps a bit too solid to be suitable for direct use as a course text, except with careful selection of the parts to be covered. If you are using it to teach yourself measure theory, I strongly recommend an eclectic approach, looking for particular subjects and theorems that seem startling or useful, and working backwards from them. My other objective, of course, is to provide an account of the central ideas at this level in measure theory, rather fuller than can easily be found in one volume elsewhere. I cannot claim that it is 'definitive', but I do think I cover a good deal of ground in ways that provide a firm foundation for further study. As in Volume 1, I usually do not shrink from giving 'best' results, like Lindeberg's conditions for the Central Limit Theorem (§274), or the theory of products of arbitrary measure spaces (§251). If I were teaching this material to students in a PhD programme I would rather accept a limitation in the breadth of the course than leave them unaware of what could be done in the areas discussed.

The topics interact in complex ways – one of the purposes of this book is to exhibit their relationships. There is no canonical linear ordering in which they should be taken. Nor do I think organization charts are very helpful, not least because it may be only two or three paragraphs in a section which are needed for a given chapter later on. I do at least try to lay the material of each section out in an order which makes initial segments useful by themselves. But the order in which to take the chapters is to a considerable extent for you to choose, perhaps after a glance at their individual introductions. I have done my best to pitch the exposition at much the same level throughout the volume, sometimes allowing gradients to steepen in the course of a chapter or a section, but always trying to return to something which anyone who has mastered Volume 1 ought to be able to cope with. (Though perhaps the main theorems of Chapter 26 are harder work than the principal results elsewhere, and §286 is only for the most determined.)

I said there were seven topics, and you will see eight chapters ahead of you. This is because Chapter 21 is rather different from the rest. It is the purest of pure measure theory, and is here only because there are places later in the volume where (in my view) the theorems make sense only in the light of some abstract concepts which are not particularly difficult, but are also not obvious. However it is fair to say that the most important ideas of this volume do not really depend on the work of Chapter 21.

As always, it is a puzzle to know how much prior knowledge to assume in this volume. I do of course call on the results of Volume 1 of this treatise whenever they seem to be relevant. I do not doubt, however, that there will be readers who have learnt the elementary theory from other sources. Provided you can, from first principles, construct Lebesgue measure and prove the basic convergence theorems for integrals on arbitrary measure spaces, you ought to be able to embark on the present volume. Perhaps it would be helpful to have in hand the results-only version of Volume 1, since that includes the most important definitions as well as statements of the theorems.

There is also the question of how much material from outside measure theory is needed. Chapter 21 calls for some non-trivial set theory (given in §2A1), but the more advanced ideas are needed only for the counter-examples in §216, and can be passed over to begin with. The problems become acute in Chapter 24. Here we need a variety of results from functional analysis, some of them depending on non-trivial ideas in general topology. For a full understanding of this material there is no substitute for a course in normed spaces up to and including a study of weak compactness. But I do not like to insist on such a preparation, because it is likely to be simultaneously too much and too little. Too much, because I hardly mention linear operators at this stage; too little, because I do ask for some of the theory of non-locally-convex spaces, which is often omitted in first courses on functional analysis. At the risk, therefore, of wasting paper, I have written out condensed accounts of the essential facts (§§2A3-2A5).

Note on second printing, April 2003

For the second printing of this volume, I have made two substantial corrections to inadequate proofs and a large number of minor amendments; I am most grateful to T.D.Austin for his careful reading of the first printing. In addition, I have added a dozen exercises and a handful of straightforward results which turn out to be relevant to the work of later volumes and fit naturally here.

The regular process of revision of this work has led me to make a couple of notational innovations not described explicitly in the early printings of Volume 1. I trust that most readers will find these immediately comprehensible. If, however, you find that there is a puzzling cross-reference which you are unable to match with anything in the version of Volume 1 which you are using, it may be worth while checking the errata pages in `http://www.essex.ac.uk/maths/staff/fremlin/mterr.htm`.

Note on second edition, January 2010

For the new ('Lulu') edition of this volume, I have eliminated a number of further errors; no doubt many remain. There are many new exercises, several new theorems and some corresponding rearrangements of material. The new results are mostly additions with little effect on the structure of the work, but there is a short new section (§266) on the Brunn-Minkowski inequality.

*Chapter 21

Taxonomy of measure spaces

I begin this volume with a 'starred chapter'. The point is that I do not really recommend this chapter for beginners. It deals with a variety of technical questions which are of great importance for the later development of the subject, but are likely to be both abstract and obscure for anyone who has not encountered the problems these techniques are designed to solve. On the other hand, if (as is customary) this work is omitted, and the ideas are introduced only when urgently needed, the student is likely to finish with very vague ideas on which theorems can be expected to apply to which types of measure space, and with no vocabulary in which to express those ideas. I therefore take a few pages to introduce the terminology and concepts which can be used to distinguish 'good' measure spaces from others, with a few of the basic relationships. The only paragraphs which are immediately relevant to the theory set out in Volume 1 are those on 'complete', 'σ-finite' and 'semi-finite' measure spaces (211A, 211D, 211F, 211Lc, §212, 213A-213B, 215B), and on Lebesgue measure (211M). For the rest, I think that a newcomer to the subject can very well pass over this chapter for the time being, and return to it for particular items when the text of later chapters refers to it. On the other hand, it can also be used as an introduction to the flavour of the 'purest' kind of measure theory, the study of measure spaces for their own sake, with a systematic discussion of a few of the elementary constructions.

211 Definitions

I start with a list of definitions, corresponding to the concepts which I have found to be of value in distinguishing different types of measure space. Necessarily, the significance of many of these ideas is likely to be obscure until you have encountered some of the obstacles which arise later on. Nevertheless, you will I hope be able to deal with these definitions on a formal, abstract basis, and to follow the elementary arguments involved in establishing the relationships between them (211L).

In 216C-216E below I will give three substantial examples to demonstrate the rich variety of objects which the definition of 'measure space' encompasses. In the present section, therefore, I content myself with very brief descriptions of sufficient cases to show at least that each of the definitions here discriminates between different spaces (211M-211R).

211A Definition Let (X,Σ,μ) be a measure space. Then μ, or (X,Σ,μ), is (**Carathéodory**) **complete** if whenever $A\subseteq E\in\Sigma$ and $\mu E=0$ then $A\in\Sigma$; that is, if every negligible subset of X is measurable.

211B Definition Let (X,Σ,μ) be a measure space. Then (X,Σ,μ) is a **probability space** if $\mu X=1$. In this case μ is called a **probability** or **probability measure**.

211C Definition Let (X,Σ,μ) be a measure space. Then μ, or (X,Σ,μ), is **totally finite** if $\mu X<\infty$.

211D Definition Let (X,Σ,μ) be a measure space. Then μ, or (X,Σ,μ), is **σ-finite** if there is a sequence $\langle E_n\rangle_{n\in\mathbb{N}}$ of measurable sets of finite measure such that $X=\bigcup_{n\in\mathbb{N}}E_n$.

Remark Note that in this case we can set

$$F_n=E_n\setminus\bigcup_{i<n}E_i,\quad G_n=\bigcup_{i\le n}E_i$$

for each n, to obtain a partition $\langle F_n\rangle_{n\in\mathbb{N}}$ of X (that is, a disjoint cover of X) into measurable sets of finite measure, and a non-decreasing sequence $\langle G_n\rangle_{n\in\mathbb{N}}$ of sets of finite measure covering X.

211E Definition Let (X,Σ,μ) be a measure space. Then μ, or (X,Σ,μ), is **strictly localizable** or **decomposable** if there is a partition $\langle X_i\rangle_{i\in I}$ of X into measurable sets of finite measure such that

$$\Sigma=\{E:E\subseteq X,\ E\cap X_i\in\Sigma\ \forall\ i\in I\},$$

$$\mu E=\sum_{i\in I}\mu(E\cap X_i)\text{ for every }E\in\Sigma.$$

I will call such a family $\langle X_i\rangle_{i\in I}$ a **decomposition** of X.

Remark In this context, we can interpret the sum $\sum_{i\in I}\mu(E\cap X_i)$ simply as

$$\sup\{\textstyle\sum_{i\in J}\mu(E\cap X_i):J\text{ is a finite subset of }I\},$$

taking $\sum_{i\in\emptyset}\mu(E\cap X_i)=0$, because we are concerned only with sums of non-negative terms (cf. 112Bd).

211F Definition Let (X,Σ,μ) be a measure space. Then μ, or (X,Σ,μ), is **semi-finite** if whenever $E\in\Sigma$ and $\mu E=\infty$ there is an $F\subseteq E$ such that $F\in\Sigma$ and $0<\mu F<\infty$.

211G Definition Let (X,Σ,μ) be a measure space. Then μ, or (X,Σ,μ), is **localizable** if it is semi-finite and, for every $\mathcal{E}\subseteq\Sigma$, there is an $H\in\Sigma$ such that (i) $E\setminus H$ is negligible for every $E\in\mathcal{E}$ (ii) if $G\in\Sigma$ and $E\setminus G$ is negligible for every $E\in\mathcal{E}$, then $H\setminus G$ is negligible. It will be convenient to call such a set H **an essential supremum** of $\mathcal{E}$ in Σ.

Remark The definition here is clumsy, because really the concept applies to measure *algebras* rather than to measure *spaces* (see 211Ya-211Yb). However, the present definition can be made to work (see 213N, 241G, 243G below) and enables us to proceed without a formal introduction to the concept of 'measure algebra' before the time comes to do the job properly in Volume 3.

211H Definition Let (X,Σ,μ) be a measure space. Then μ, or (X,Σ,μ), is **locally determined** if it is semi-finite and

$$\Sigma=\{E:E\subseteq X,\ E\cap F\in\Sigma\text{ whenever }F\in\Sigma\text{ and }\mu F<\infty\};$$

that is to say, for any $E\in\mathcal{P}X\setminus\Sigma$ there is an $F\in\Sigma$ such that $\mu F<\infty$ and $E\cap F\notin\Sigma$.

211I Definition Let (X,Σ,μ) be a measure space. A set $E\in\Sigma$ is an **atom** for μ if $\mu E>0$ and whenever $F\in\Sigma$, $F\subseteq E$ one of F, $E\setminus F$ is negligible.

211J Definition Let (X,Σ,μ) be a measure space. Then μ, or (X,Σ,μ), is **atomless** or **diffused** if there are no atoms for μ. (Note that this is *not* the same thing as saying that all finite sets are negligible; see 211R below. Some authors use the word **continuous** in this context.)

211K Definition Let (X,Σ,μ) be a measure space. Then μ, or (X,Σ,μ), is **purely atomic** if whenever $E\in\Sigma$ and E is not negligible there is an atom for μ included in E.

Remark Recall that a measure μ on a set X is **point-supported** if μ measures every subset of X and $\mu E=\sum_{x\in E}\mu\{x\}$ for every $E\subseteq X$ (112Bd). Every point-supported measure is purely atomic, because $\{x\}$ must be an atom whenever $\mu\{x\}>0$, but not every purely atomic measure is point-supported (211R).

211L The relationships between the concepts above are in a sense very straightforward; all the direct implications in which one property implies another are given in the next theorem.

Theorem (a) A probability space is totally finite.

(b) A totally finite measure space is σ-finite.

(c) A σ-finite measure space is strictly localizable.

(d) A strictly localizable measure space is localizable and locally determined.

(e) A localizable measure space is semi-finite.

(f) A locally determined measure space is semi-finite.

proof (a), (b), (e) and (f) are trivial.

(c) Let (X,Σ,μ) be a σ-finite measure space; let $\langle F_n\rangle_{n\in\mathbb{N}}$ be a disjoint sequence of measurable sets of finite measure covering X (see the remark in 211D). If $E\in\Sigma$, then of course $E\cap F_n\in\Sigma$ for every $n\in\mathbb{N}$, and

$$\mu E=\sum_{n=0}^{\infty}\mu(E\cap F_n)=\sum_{n\in\mathbb{N}}\mu(E\cap F_n).$$

If $E\subseteq X$ and $E\cap F_n\in\Sigma$ for every $n\in\mathbb{N}$, then

$$E=\bigcup_{n\in\mathbb{N}}E\cap F_n\in\Sigma.$$

So $\langle F_n\rangle_{n\in\mathbb{N}}$ is a decomposition of X and (X,Σ,μ) is strictly localizable.

(d) Let (X,Σ,μ) be a strictly localizable measure space; let $\langle X_i\rangle_{i\in I}$ be a decomposition of X.

(i) Let $\mathcal{E}$ be a family of measurable subsets of X. Let $\mathcal{F}$ be the family of measurable sets $F\subseteq X$ such that $\mu(F\cap E)=0$ for every $E\in\mathcal{E}$. Note that $\emptyset\in\mathcal{F}$ and, if $\langle F_n\rangle_{n\in\mathbb{N}}$ is any sequence in $\mathcal{F}$, then $\bigcup_{n\in\mathbb{N}}F_n\in\mathcal{F}$. For each $i\in I$, set $\gamma_i=\sup\{\mu(F\cap X_i):F\in\mathcal{F}\}$ and choose a sequence $\langle F_{in}\rangle_{n\in\mathbb{N}}$ in $\mathcal{F}$ such that $\lim_{n\to\infty}\mu(F_{in}\cap X_i)=\gamma_i$; set

$$F_i=\bigcup_{n\in\mathbb{N}}F_{in}\in\mathcal{F}.$$

Set

$$F = \bigcup_{i\in I} F_i \cap X_i \subseteq X$$

and $H = X \setminus F$.

We see that $F \cap X_i = F_i \cap X_i$ for each $i \in I$ (because $\langle X_i\rangle_{i\in I}$ is disjoint), so $F \in \Sigma$ and $H \in \Sigma$. For any $E \in \mathcal{E}$,

$$\mu(E \setminus H) = \mu(E \cap F) = \sum_{i\in I} \mu(E \cap F \cap X_i) = \sum_{i\in I} \mu(E \cap F_i \cap X_i) = 0$$

because every F_i belongs to $\mathcal{F}$. Thus $F \in \mathcal{F}$. If $G \in \Sigma$ and $\mu(E \setminus G) = 0$ for every $E \in \mathcal{E}$, then $X \setminus G$ and $F' = F \cup (X \setminus G)$ belong to $\mathcal{F}$. So $\mu(F' \cap X_i) \le \gamma_i$ for each $i \in I$. But also $\mu(F \cap X_i) \ge \sup_{n\in\mathbb{N}} \mu(F_{in} \cap X_i) = \gamma_i$, so $\mu(F \cap X_i) = \mu(F' \cap X_i)$ for each i. *Because μX_i is finite,* it follows that $\mu((F' \setminus F) \cap X_i) = 0$, for each i. Summing over i, $\mu(F' \setminus F) = 0$, that is, $\mu(H \setminus G) = 0$.

Thus H is an essential supremum for $\mathcal{E}$ in Σ. As $\mathcal{E}$ is arbitrary, (X, Σ, μ) is localizable.

(ii) If $E \in \Sigma$ and $\mu E = \infty$, then there is some $i \in I$ such that

$$0 < \mu(E \cap X_i) \le \mu X_i < \infty;$$

so (X, Σ, μ) is semi-finite. If $E \subseteq X$ and $E \cap F \in \Sigma$ whenever $\mu F < \infty$, then, in particular, $E \cap X_i \in \Sigma$ for every $i \in I$, so $E \in \Sigma$; thus (X, Σ, μ) is locally determined.

211M Example: Lebesgue measure Let us consider Lebesgue measure in the light of the concepts above. Write μ for Lebesgue measure on $\mathbb{R}^r$ and Σ for its domain.

(a) μ is complete, because it is constructed by Carathéodory's method; if $A \subseteq E$ and $\mu E = 0$, then $\mu^* A = \mu^* E = 0$ (writing μ^* for Lebesgue outer measure), so, for any $B \subseteq \mathbb{R}$,

$$\mu^*(B \cap A) + \mu^*(B \setminus A) \le 0 + \mu^* B = \mu^* B,$$

and A must be measurable.

(b) μ is σ-finite, because $\mathbb{R}^r = \bigcup_{n\in\mathbb{N}} [-\mathbf{n}, \mathbf{n}]$, writing $\mathbf{n}$ for the vector $(n, \dots, n)$, and $\mu[-\mathbf{n}, \mathbf{n}] = (2n)^r < \infty$ for every n. Of course μ is neither totally finite nor a probability measure.

(c) Because μ is σ-finite, it is strictly localizable (211Lc), localizable (211Ld), locally determined (211Ld) and semi-finite (211Le or 211Lf).

(d) μ is atomless. **P** Suppose that $E \in \Sigma$. Consider the function

$$a \mapsto f(a) = \mu(E \cap [-\mathbf{a}, \mathbf{a}]) : [0, \infty[\to \mathbb{R}$$

We have

$$f(a) \le f(b) \le f(a) + \mu[-\mathbf{b}, \mathbf{b}] - \mu[-\mathbf{a}, \mathbf{a}] = f(a) + (2b)^r - (2a)^r$$

whenever $a \le b$ in $[0, \infty[$, so f is continuous. Now $f(0) = 0$ and $\lim_{n\to\infty} f(n) = \mu E > 0$. By the Intermediate Value Theorem there is an $a \in [0, \infty[$ such that $0 < f(a) < \mu E$. So we have

$$0 < \mu(E \cap [-\mathbf{a}, \mathbf{a}]) < \mu E.$$

As E is arbitrary, μ is atomless. **Q**

(e) It is now a trivial observation that μ cannot be purely atomic, because $\mathbb{R}^r$ itself is a set of positive measure not including any atom.

211N Counting measure Take X to be any uncountable set (e.g., $X = \mathbb{R}$), and μ to be counting measure on X (112Bd).

(a) μ is complete, because if $A \subseteq E$ and $\mu E = 0$ then

$$A = E = \emptyset \in \Sigma.$$

(b) μ is not σ-finite, because if $\langle E_n\rangle_{n\in\mathbb{N}}$ is any sequence of sets of finite measure then every E_n is finite, therefore countable, and $\bigcup_{n\in\mathbb{N}} E_n$ is countable (1A1F), so cannot be X. *A fortiori,* μ is not a probability measure nor totally finite.

(c) μ is strictly localizable. **P** Set $X_x = \{x\}$ for every $x \in X$. Then $\langle X_x \rangle_{x \in X}$ is a partition of X, and for any $E \subseteq X$

$$\mu(E \cap X_x) = 1 \text{ if } x \in E, \quad 0 \text{ otherwise.}$$

By the definition of μ,

$$\mu E = \sum_{x \in X} \mu(E \cap X_x)$$

for every $E \subseteq X$, and $\langle X_x \rangle_{x \in X}$ is a decomposition of X. **Q**

Consequently μ is localizable, locally determined and semi-finite.

(d) μ is purely atomic. **P** $\{x\}$ is an atom for every $x \in X$, and if $\mu E > 0$ then surely E includes $\{x\}$ for some x. **Q** Obviously, μ is not atomless.

211O A non-semi-finite space Set $X = \{0\}$, $\Sigma = \{\emptyset, X\}$, $\mu\emptyset = 0$ and $\mu X = \infty$. Then μ is not semi-finite, as $\mu X = \infty$ but X has no subset of non-zero finite measure. It follows that μ cannot be localizable, locally determined, σ-finite, totally finite nor a probability measure. Because $\Sigma = \mathcal{P}X$, μ is complete. X is an atom for μ, so μ is purely atomic (indeed, it is point-supported).

211P A non-complete space Write $\mathcal{B}$ for the σ-algebra of Borel subsets of $\mathbb{R}$ (111G), and μ for the restriction of Lebesgue measure to $\mathcal{B}$ (recall that by 114G every Borel subset of $\mathbb{R}$ is Lebesgue measurable). Then $(\mathbb{R}, \mathcal{B}, \mu)$ is atomless, σ-finite and not complete.

proof (a) To see that μ is not complete, recall that there is a continuous, strictly increasing bijection $g : [0, 1] \to [0, 1]$ such that $\mu g[C] > 0$, where C is the Cantor set, so that there is a set $A \subseteq g[C]$ which is not Lebesgue measurable (134Ib). Now $g^{-1}[A] \subseteq C$ cannot be a Borel set, since $\chi A = \chi(g^{-1}[A]) \circ g^{-1}$ is not Lebesgue measurable, therefore not Borel measurable, and the composition of two Borel measurable functions is Borel measurable (121Eg); so $g^{-1}[A]$ is a non-measurable subset of the negligible set C.

(b) The rest of the arguments of 211M apply to μ just as well as to true Lebesgue measure, so μ is σ-finite and atomless.

***Remark** The argument offered in (a) could give rise to a seriously false impression. The set A referred to there can be constructed only with the use of a strong form of the axiom of choice. No such device is necessary for the result here. There are many methods of constructing non-Borel subsets of the Cantor set, all illuminating in different ways. In the absence of any form of the axiom of choice, there are difficulties with the concept of 'Borel set', and others with the concept of 'Lebesgue measure', which I will come to in Chapter 56; but countable choice is quite sufficient for the existence of a non-Borel subset of $\mathbb{R}$. For details of a possible approach see 423L in Volume 4.

211Q Some probability spaces Two obvious constructions of probability spaces, restricting myself to the methods described in Volume 1, are

(a) the subspace measure induced by Lebesgue measure on $[0, 1]$ (131B);

(b) the point-supported measure induced on a set X by a function $h : X \to [0, 1]$ such that $\sum_{x \in X} h(x) = 1$, writing $\mu E = \sum_{x \in E} h(x)$ for every $E \subseteq X$; for instance, if X is a singleton $\{x\}$ and $h(x) = 1$, or if $X = \mathbb{N}$ and $h(n) = 2^{-n-1}$.

Of these two, (a) gives an atomless probability measure and (b) gives a purely atomic probability measure.

211R Countable-cocountable measure The following is one of the basic constructions to keep in mind when considering abstract measure spaces.

(a) Let X be any set. Let Σ be the family of those sets $E \subseteq X$ such that either E or $X \setminus E$ is countable. Then Σ is a σ-algebra of subsets of X. **P** (i) $\emptyset$ is countable, so belongs to Σ. (ii) The condition for E to belong to Σ is symmetric between E and $X \setminus E$, so $X \setminus E \in \Sigma$ for every $E \in \Sigma$. (iii) Let $\langle E_n \rangle_{n \in \mathbb{N}}$ be any sequence in Σ, and set $E = \bigcup_{n \in \mathbb{N}} E_n$. If every E_n is countable, then E is countable, so belongs to Σ. Otherwise, there is some n such that $X \setminus E_n$ is countable, in which case $X \setminus E \subseteq X \setminus E_n$ is countable, so again $E \in \Sigma$. **Q** Σ is called the **countable-cocountable σ-algebra** of X.

(b) Now consider the function $\mu : \Sigma \to \{0, 1\}$ defined by writing $\mu E = 0$ if E is countable, $\mu E = 1$ if $E \in \Sigma$ and E is not countable. Then μ is a measure. **P** (i) $\emptyset$ is countable so $\mu\emptyset = 0$. (ii) Let $\langle E_n \rangle_{n \in \mathbb{N}}$ be a disjoint sequence in Σ, and E its union. (α) If every E_m is countable, then so is E, so

$$\mu E = 0 = \sum_{n=0}^{\infty} \mu E_n.$$

(β) If some E_m is uncountable, then $E \supseteq E_m$ also is uncountable, and $\mu E = \mu E_m = 1$. But in this case, because $E_m \in \Sigma$, $X \setminus E_m$ is countable, so E_n, being a subset of $X \setminus E_m$, is countable for every $n \neq m$; thus $\mu E_n = 0$ for every $n \neq m$, and

$$\mu E = 1 = \sum_{n=0}^{\infty} \mu E_n.$$

As $\langle E_n \rangle_{n \in \mathbb{N}}$ is arbitrary, μ is a measure. **Q** This is the **countable-cocountable measure** on X.

(c) If X is any uncountable set and μ is the countable-cocountable measure on X, then μ is a complete, purely atomic probability measure, but is not point-supported. **P** (i) If $A \subseteq E$ and $\mu E = 0$, then E is countable, so A also is countable and belongs to Σ. Thus μ is complete. (ii) Because X is uncountable, $\mu X = 1$ and μ is a probability measure. (iii) If $\mu E > 0$, then $\mu F = \mu E = 1$ whenever F is a non-negligible measurable subset of E, so E is itself an atom; thus μ is purely atomic. (iv) $\mu X = 1 > 0 = \sum_{x \in X} \mu\{x\}$, so μ is not point-supported. **Q**

211X Basic exercises >(a) Let (X, Σ, μ) be a measure space. Show that μ is σ-finite iff there is a totally finite measure ν on X with the same measurable sets and the same negligible sets as μ.

>(b) Let $g : \mathbb{R} \to \mathbb{R}$ be a non-decreasing function and μ_g the associated Lebesgue-Stieltjes measure (114Xa). Show that μ_g is complete and σ-finite. Show that

(i) μ_g is totally finite iff g is bounded;

(ii) μ_g is a probability measure iff $\lim_{x \to \infty} g(x) - \lim_{x \to -\infty} g(x) = 1$;

(iii) μ_g is atomless iff g is continuous;

(iv) if E is any atom for μ_g, there is a point $x \in E$ such that $\mu_g E = \mu_g\{x\}$;

(v) μ_g is purely atomic iff it is point-supported.

>(c) Let μ be counting measure on a set X. Show that μ is always complete, strictly localizable and purely atomic, and that it is σ-finite iff X is countable, totally finite iff X is finite, a probability measure iff X is a singleton, and atomless iff X is empty.

(d) Show that a point-supported measure is always complete, and is strictly localizable iff it is semi-finite.

(e) Let X be a set. Show that for any σ-ideal $\mathcal{I}$ of subsets of X (definition: 112Db), the set

$$\Sigma = \mathcal{I} \cup \{X \setminus A : A \in \mathcal{I}\}$$

is a σ-algebra of subsets of X, and that there is a measure $\mu : \Sigma \to \{0, 1\}$ given by setting

$$\mu E = 0 \text{ if } E \in \mathcal{I}, \quad \mu E = 1 \text{ if } E \in \Sigma \setminus \mathcal{I}.$$

Show that $\mathcal{I}$ is precisely the null ideal of μ, that μ is complete, totally finite and purely atomic, and is a probability measure iff $X \notin \mathcal{I}$.

(f) Show that a point-supported measure is strictly localizable iff it is semi-finite.

(g) Let (X, Σ, μ) be a measure space such that $\mu X > 0$. Show that the set of conegligible subsets of X is a filter on X.

211Y Further exercises (a) Let (X, Σ, μ) be a measure space, and for E, $F \in \Sigma$ write $E \sim F$ if $\mu(E \triangle F) = 0$. Show that $\sim$ is an equivalence relation on Σ. Let $\mathfrak{A}$ be the set of equivalence classes in Σ for $\sim$; for $E \in \Sigma$, write $E^\bullet \in \mathfrak{A}$ for its equivalence class. Show that there is a partial ordering $\subseteq$ on $\mathfrak{A}$ defined by saying that, for E, $F \in \Sigma$,

$$E^\bullet \subseteq F^\bullet \iff \mu(E \setminus F) = 0.$$

Show that μ is localizable iff for every $A \subseteq \mathfrak{A}$ there is an $h \in \mathfrak{A}$ such that (i) $a \subseteq h$ for every $a \in A$ (ii) whenever $g \in \mathfrak{A}$ is such that $a \subseteq g$ for every $a \in \mathfrak{A}$, then $h \subseteq g$.

(b) Let (X, Σ, μ) be a measure space, and construct $\mathfrak{A}$ as in 211Ya. Show that there are operations $\cup$, $\cap$, $\setminus$ on $\mathfrak{A}$ defined by saying that

$$E^\bullet \cap F^\bullet = (E \cap F)^\bullet, \quad E^\bullet \cup F^\bullet = (E \cup F)^\bullet, \quad E^\bullet \setminus F^\bullet = (E \setminus F)^\bullet$$

for all E, $F \in \Sigma$. Show that if $A \subseteq \mathfrak{A}$ is any *countable* set, then there is an $h \in \mathfrak{A}$ such that (i) $a \subseteq h$ for every $a \in A$ (ii) whenever $g \in \mathfrak{A}$ is such that $a \subseteq g$ for every $a \in \mathfrak{A}$, then $h \subseteq g$. Show that there is a functional $\bar{\mu} : \mathfrak{A} \to [0, \infty]$ defined by saying that $\bar{\mu}(E^\bullet) = \mu E$ for every $E \in \Sigma$. ($(\mathfrak{A}, \bar{\mu})$ is called the **measure algebra** of (X, Σ, μ).)

(c) Let (X,Σ,μ) be a semi-finite measure space. Show that it is atomless iff whenever $\epsilon>0$, $E\in\Sigma$ and $\mu E<\infty$, then there is a finite partition of E into measurable sets of measure at most ϵ.

(d) Let (X,Σ,μ) be a strictly localizable measure space. Show that it is atomless iff for every $\epsilon>0$ there is a decomposition of X consisting of sets of measure at most ϵ.

(e) Let Σ be the countable-cocountable σ-algebra of $\mathbb{R}$. Show that $[0,\infty[\,\notin\Sigma$. Let μ be the restriction of counting measure to Σ. Show that $(\mathbb{R},\Sigma,\mu)$ is complete, semi-finite and purely atomic, but not localizable nor locally determined.

211 Notes and comments The list of definitions in 211A-211K probably strikes you as quite long enough, even though I have omitted many occasionally useful ideas. The concepts here vary widely in importance, and the importance of each varies widely with context. My own view is that it is absolutely necessary, when studying any measure space, to know its classification under the eleven discriminating features listed here, and to be able to describe any atoms which are present. Fortunately, for most 'ordinary' measure spaces, the classification is fairly quick, because if (for instance) the space is σ-finite, and you know the measure of the whole space, the only remaining questions concern completeness and atoms. The distinctions between spaces which are, or are not, strictly localizable, semi-finite, localizable and locally determined are relevant only for spaces which are not σ-finite, and do not arise in elementary applications.

I think it is also fair to say that the notions of 'complete' and 'locally determined' measure space are technical; I mean, that they do not correspond to significant features of the essential structure of a space, though there are some interesting problems concerning incomplete measures. One manifestation of this is the existence of canonical constructions for rendering spaces complete or complete and locally determined (212C, 213D-213E). In addition, measure spaces which are not semi-finite do not really belong to measure theory, but rather to the more general study of σ-algebras and σ-ideals. The most important classifications, in terms of the behaviour of a measure space, seem to me to be 'σ-finite', 'localizable' and 'strictly localizable'; these are the critical features which cannot be forced by elementary constructions.

If you know anything about Borel subsets of the real line, the argument of part (a) of the proof of 211P must look very clumsy. But 'better' proofs rely on ideas which we shall not need until Volume 4, and the proof here is based on a construction which we have to understand for other reasons.

212 Complete spaces

In the next two sections of this chapter I give brief accounts of the theory of measure spaces possessing certain of the properties described in §211. I begin with 'completeness'. I give the elementary facts about complete measure spaces in 212A-212B; then I turn to the notion of 'completion' of a measure (212C) and its relationships with the other concepts of measure theory introduced so far (212D-212G).

212A Proposition Any measure space constructed by Carathéodory's method is complete.

proof Recall that 'Carathéodory's method' starts from an arbitrary outer measure $\theta:\mathcal{P}X\to[0,\infty]$ and sets

$$\Sigma=\{E:E\subseteq X,\ \theta A=\theta(A\cap E)+\theta(A\setminus E)\text{ for every }A\subseteq X\},\qquad \mu=\theta\restriction\Sigma$$

(113C). In this case, if $B\subseteq E\in\Sigma$ and $\mu E=0$, then $\theta B=\theta E=0$ (113A(ii)), so for any $A\subseteq X$ we have

$$\theta(A\cap B)+\theta(A\setminus B)=\theta(A\setminus B)\le\theta A\le\theta(A\cap B)+\theta(A\setminus B),$$

and $B\in\Sigma$.

212B Proposition (a) If (X,Σ,μ) is a complete measure space, then any conegligible subset of X is measurable.

(b) Let (X,Σ,μ) be a complete measure space, and f a $[-\infty,\infty]$-valued function defined on a subset of X. If f is virtually measurable (that is, there is a conegligible set $E\subseteq X$ such that $f\restriction E$ is measurable), then f is measurable.

(c) Let (X,Σ,μ) be a complete measure space, and f a real-valued function defined on a conegligible subset of X. Then the following are equiveridical, that is, if one is true so are the others:

(i) f is integrable;

(ii) f is measurable and $|f|$ is integrable;

(iii) f is measurable and there is an integrable function g such that $|f|\le_{\text{a.e.}} g$.

proof (a) If E is conegligible, then $X\setminus E$ is negligible, therefore measurable, and E is measurable.

(b) Let $a\in\mathbb{R}$. Then there is an $H\in\Sigma$ such that

$$\{x:(f\restriction E)(x)\le a\}=H\cap\operatorname{dom}(f\restriction E)=H\cap E\cap\operatorname{dom}f.$$

Now $F = \{x : x \in \operatorname{dom} f \setminus E,\ f(x) \le a\}$ is a subset of the negligible set $X \setminus E$, so is measurable, and

$$\{x : f(x) \le a\} = (F \cup H) \cap \operatorname{dom} f \in \Sigma_{\operatorname{dom} f},$$

writing $\Sigma_D = \{D \cap E : E \in \Sigma\}$, as in 121A. As a is arbitrary, f is measurable (135E).

(c)(i)⇒(ii) If f is integrable, then by 122P f is virtually measurable and by 122Re $|f|$ is integrable. By (b) here, f is measurable, so (ii) is true.

(ii)⇒(iii) is trivial.

(iii)⇒(i) If f is measurable and g is integrable and $|f| \le_{\text{a.e.}} g$, then f is virtually measurable, $|g|$ is integrable and $|f| \le_{\text{a.e.}} |g|$, so 122P tells us that f is integrable.

212C The completion of a measure Let (X, Σ, μ) be any measure space.

(a) Let $\hat\Sigma$ be the family of those sets $E \subseteq X$ such that there are E', $E'' \in \Sigma$ with $E' \subseteq E \subseteq E''$ and $\mu(E'' \setminus E') = 0$. Then $\hat\Sigma$ is a σ-algebra of subsets of X. **P** (i) Of course $\emptyset$ belongs to $\hat\Sigma$, because we can take $E' = E'' = \emptyset$. (ii) If $E \in \hat\Sigma$, take E', $E'' \in \Sigma$ such that $E' \subseteq E \subseteq E''$ and $\mu(E'' \setminus E') = 0$. Then

$$X \setminus E'' \subseteq X \setminus E \subseteq X \setminus E', \quad \mu((X \setminus E') \setminus (X \setminus E'')) = \mu(E'' \setminus E') = 0,$$

so $X \setminus E \in \hat\Sigma$. (iii) If $\langle E_n \rangle_{n\in\mathbb{N}}$ is a sequence in $\hat\Sigma$, then for each n choose E'_n, $E''_n \in \Sigma$ such that $E'_n \subseteq E_n \subseteq E''_n$ and $\mu(E''_n \setminus E'_n) = 0$. Set

$$E = \bigcup_{n\in\mathbb{N}} E_n, \quad E' = \bigcup_{n\in\mathbb{N}} E'_n, \quad E'' = \bigcup_{n\in\mathbb{N}} E''_n;$$

then $E' \subseteq E \subseteq E''$ and $E'' \setminus E' \subseteq \bigcup_{n\in\mathbb{N}} (E''_n \setminus E'_n)$ is negligible, so $E \in \hat\Sigma$. **Q**

(b) For $E \in \hat\Sigma$, set

$$\hat\mu E = \mu^* E = \min\{\mu F : E \subseteq F \in \Sigma\}$$

(132A). It is worth remarking at once that if $E \in \hat\Sigma$, E', $E'' \in \Sigma$, $E' \subseteq E \subseteq E''$ and $\mu(E'' \setminus E') = 0$, then $\mu E' = \hat\mu E = \mu E''$; this is because

$$\mu E' = \mu^* E' \le \mu^* E \le \mu^* E'' = \mu E'' = \mu E' + \mu(E'' \setminus E) = \mu E'$$

(recalling from 132A, or noting now, that $\mu^* A \le \mu^* B$ whenever $A \subseteq B \subseteq X$, and that μ^* agrees with μ on Σ).

(c) We now find that $(X, \hat\Sigma, \hat\mu)$ is a measure space. **P** (i) Of course $\hat\mu$, like μ, takes values in $[0, \infty]$. (ii) $\hat\mu\emptyset = \mu\emptyset = 0$. (iii) Let $\langle E_n \rangle_{n\in\mathbb{N}}$ be a disjoint sequence in $\hat\Sigma$, with union E. For each $n \in \mathbb{N}$ choose E'_n, $E''_n \in \Sigma$ such that $E'_n \subseteq E_n \subseteq E''_n$ and $\mu(E''_n \setminus E'_n) = 0$. Set $E' = \bigcup_{n\in\mathbb{N}} E'_n$, $E'' = \bigcup_{n\in\mathbb{N}} E''_n$. Then (as in (a-iii) above) $E' \subseteq E \subseteq E''$ and $\mu(E'' \setminus E') = 0$, so

$$\hat\mu E = \mu E' = \sum_{n=0}^{\infty} \mu E'_n = \sum_{n=0}^{\infty} \hat\mu E_n$$

because $\langle E'_n \rangle_{n\in\mathbb{N}}$, like $\langle E_n \rangle_{n\in\mathbb{N}}$, is disjoint. **Q**

(d) The measure space $(X, \hat\Sigma, \hat\mu)$ is called the **completion** of the measure space (X, Σ, μ); equally, I will call $\hat\mu$ the **completion** of μ, and occasionally (if it seems plain which null ideal is under consideration) I will call $\hat\Sigma$ the completion of Σ. Members of $\hat\Sigma$ are sometimes called **μ-measurable**.

212D There is something I had better check at once.

Proposition Let (X, Σ, μ) be any measure space. Then $(X, \hat\Sigma, \hat\mu)$, as defined in 212C, is a complete measure space and $\hat\mu$ is an extension of μ; and $(X, \hat\Sigma, \hat\mu) = (X, \Sigma, \mu)$ iff (X, Σ, μ) is complete.

proof (a) Suppose that $A \subseteq E \in \hat\Sigma$ and $\hat\mu E = 0$. Then (by 212Cb) there is an $E'' \in \Sigma$ such that $E \subseteq E''$ and $\mu E'' = 0$. Accordingly we have

$$\emptyset \subseteq A \subseteq E'', \quad \mu(E'' \setminus \emptyset) = 0,$$

so $A \in \hat\Sigma$. As A is arbitrary, $\hat\mu$ is complete.

(b) If $E \in \Sigma$, then of course $E \in \hat\Sigma$, because $E \subseteq E \subseteq E$ and $\mu(E \setminus E) = 0$; and $\hat\mu E = \mu^* E = \mu E$. Thus $\Sigma \subseteq \hat\Sigma$ and $\hat\mu$ extends μ.

(c) If $\mu = \hat{\mu}$ then of course μ must be complete. If μ is complete, and $E \in \hat{\Sigma}$, then there are E', $E'' \in \Sigma$ such that $E' \subseteq E \subseteq E''$ and $\mu(E'' \setminus E') = 0$. But now $E \setminus E' \subseteq E'' \setminus E'$, so (because (X, Σ, μ) is complete) $E \setminus E' \in \Sigma$ and $E = E' \cup (E \setminus E') \in \Sigma$. As E is arbitrary, $\hat{\Sigma} \subseteq \Sigma$ and $\hat{\Sigma} = \Sigma$ and $\mu = \hat{\mu}$.

212E The importance of this construction is such that it is worth spelling out some further elementary properties.

Proposition Let (X, Σ, μ) be a measure space, and $(X, \hat{\Sigma}, \hat{\mu})$ its completion.

(a) The outer measures $\hat{\mu}^*$, μ^* defined from $\hat{\mu}$ and μ coincide.

(b) μ, $\hat{\mu}$ give rise to the same negligible and conegligible sets and the same sets of full outer measure.

(c) $\hat{\mu}$ is the only measure with domain $\hat{\Sigma}$ which agrees with μ on Σ.

(d) A subset of X belongs to $\hat{\Sigma}$ iff it is expressible as $F \triangle A$ where $F \in \Sigma$ and A is μ-negligible.

proof (a) Take any $A \subseteq X$. (i) If $A \subseteq F \in \Sigma$, then $F \in \hat{\Sigma}$ and $\mu F = \hat{\mu} F$, so

$$\hat{\mu}^* A \le \hat{\mu} F = \mu F;$$

as F is arbitrary, $\hat{\mu}^* A \le \mu^* A$. (ii) If $A \subseteq E \in \hat{\Sigma}$, there is an $E'' \in \Sigma$ such that $E \subseteq E''$ and $\mu E'' = \hat{\mu} E$, so

$$\mu^* A \le \mu E'' = \hat{\mu} E;$$

as E is arbitrary, $\mu^* A \le \hat{\mu}^* A$.

(b) Now, for $A \subseteq X$,

$$A \text{ is } \mu\text{-negligible} \iff \mu^* A = 0 \iff \hat{\mu}^* A = 0 \iff A \text{ is } \hat{\mu}\text{-negligible},$$

$$\begin{aligned} A \text{ is } \mu\text{-conegligible} &\iff \mu^*(X \setminus A) = 0 \\ &\iff \hat{\mu}^*(X \setminus A) = 0 \iff A \text{ is } \hat{\mu}\text{-conegligible.} \end{aligned}$$

If A has full outer measure for μ, $F \in \hat{\Sigma}$ and $F \cap A = \emptyset$, then there is an $F' \in \Sigma$ such that $F' \subseteq F$ and $\mu F' = \hat{\mu} F$; as $F' \cap A = \emptyset$, F' is μ-negligible and F is $\hat{\mu}$-negligible; as F is arbitrary, A has full outer measure for $\hat{\mu}$. In the other direction, of course, if A has full outer measure for $\hat{\mu}$ then

$$\mu^*(F \cap A) = \hat{\mu}^*(F \cap A) = \hat{\mu} F = \mu F$$

for every $F \in \Sigma$, so A has full outer measure for μ.

(c) If $\tilde{\mu}$ is any measure with domain $\hat{\Sigma}$ extending μ, we must have

$$\mu E' \le \tilde{\mu} E \le \mu E'', \quad \mu E' = \hat{\mu} E = \mu E'',$$

so $\tilde{\mu} E = \hat{\mu} E$, whenever E', $E'' \in \Sigma$, $E' \subseteq E \subseteq E''$ and $\mu(E'' \setminus E') = 0$.

(d)(i) If $E \in \hat{\Sigma}$, take E', $E'' \in \Sigma$ such that $E' \subseteq E \subseteq E''$ and $\mu(E'' \setminus E') = 0$. Then $E \setminus E' \subseteq E'' \setminus E'$, so $E \setminus E'$ is μ-negligible, and $E = E' \triangle (E \setminus E')$ is the symmetric difference of a member of Σ and a negligible set.

(ii) If $E = F \triangle A$, where $F \in \Sigma$ and A is μ-negligible, take $G \in \Sigma$ such that $\mu G = 0$ and $A \subseteq G$; then $F \setminus G \subseteq E \subseteq F \cup G$ and $\mu((F \cup G) \setminus (F \setminus G)) = \mu G = 0$, so $E \in \hat{\Sigma}$.

212F Now let us consider integration with respect to the completion of a measure.

Proposition Let (X, Σ, μ) be a measure space and $(X, \hat{\Sigma}, \hat{\mu})$ its completion.

(a) A $[-\infty, \infty]$-valued function f defined on a subset of X is $\hat{\Sigma}$-measurable iff it is μ-virtually measurable.

(b) Let f be a $[-\infty, \infty]$-valued function defined on a subset of X. Then $\int f d\mu = \int f d\hat{\mu}$ if either is defined in $[-\infty, \infty]$; in particular, f is μ-integrable iff it is $\hat{\mu}$-integrable.

proof (a)(i) Suppose that f is a $[-\infty, \infty]$-valued $\hat{\Sigma}$-measurable function. For $q \in \mathbb{Q}$ let $E_q \in \hat{\Sigma}$ be such that $\{x : f(x) \le q\} = \operatorname{dom} f \cap E_q$, and choose E'_q, $E''_q \in \Sigma$ such that $E'_q \subseteq E_q \subseteq E''_q$ and $\mu(E''_q \setminus E'_q) = 0$. Set $H = X \setminus \bigcup_{q \in \mathbb{Q}} (E''_q \setminus E'_q)$; then H is μ-conegligible. For $a \in \mathbb{R}$ set

$$G_a = \bigcup_{q \in \mathbb{Q}, q < a} E'_q \in \Sigma;$$

then

$$\{x : x \in \operatorname{dom}(f \restriction H), \ (f \restriction H)(x) < a\} = G_a \cap \operatorname{dom}(f \restriction H).$$

This shows that $f \restriction H$ is Σ-measurable, so that f is μ-virtually measurable.

(ii) If f is μ-virtually measurable, then there is a μ-conegligible set $H \subseteq X$ such that $f \restriction H$ is Σ-measurable. Since $\Sigma \subseteq \hat{\Sigma}$, $f \restriction H$ is also $\hat{\Sigma}$-measurable. And H is $\hat{\mu}$-conegligible, by 212Eb. But this means that f is $\hat{\mu}$-virtually measurable, therefore $\hat{\Sigma}$-measurable, by 212Bb.

(b)(i) Let $f : D \to [-\infty, \infty]$ be a function, where $D \subseteq X$. If either of $\int f d\mu$, $\int f d\hat{\mu}$ is defined in $[-\infty, \infty]$, then f is virtually measurable, and defined almost everywhere, for one of the appropriate measures, and therefore for both (putting (a) above together with 212Bb).

(ii) Now suppose that f is non-negative and integrable either with respect to μ or with respect to $\hat{\mu}$. Let $E \in \Sigma$ be a conegligible set included in $\operatorname{dom} f$ such that $f \restriction E$ is Σ-measurable. For $n \in \mathbb{N}$, $k \geq 1$ set

$$E_{nk} = \{x : x \in E,\ f(x) \geq 2^{-n}k\};$$

then each E_{nk} belongs to Σ and is of finite measure for both μ and $\hat{\mu}$. (If f is μ-integrable,

$$\hat{\mu} E_{nk} = \mu E_{nk} \leq 2^n \int f d\mu;$$

if f is $\hat{\mu}$-integrable,

$$\mu E_{nk} = \hat{\mu} E_{nk} \leq 2^n \int f d\hat{\mu}.)$$

So

$$f_n = \sum_{k=1}^{4^n} 2^{-n} \chi E_{nk}$$

is both μ-simple and $\hat{\mu}$-simple, and $\int f_n d\mu = \int f_n d\hat{\mu}$. Observe that, for $x \in E$,

$$\begin{aligned} f_n(x) &= 2^{-n}k \text{ if } k < 4^n \text{ and } 2^{-n}k \leq f(x) < 2^{-n}(k+1), \\ &= 2^n \text{ if } f(x) \geq 2^n. \end{aligned}$$

Thus $\langle f_n \rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence of functions converging to f at every point of E, that is, both μ-almost everywhere and $\hat{\mu}$-almost everywhere. So we have, for any $c \in \mathbb{R}$,

$$\begin{aligned} \int f d\mu = c &\iff \lim_{n\to\infty} \int f_n d\mu = c \\ &\iff \lim_{n\to\infty} \int f_n d\hat{\mu} = c \iff \int f d\hat{\mu} = c. \end{aligned}$$

(iii) As for infinite integrals, recall that for a non-negative function I write '$\int f = \infty$' just when f is defined almost everywhere, is virtually measurable, and is not integrable. So (i) and (ii) together show that $\int f d\mu = \int f d\hat{\mu}$ whenever f is non-negative and either integral is defined in $[0, \infty]$.

(iv) Since both μ, $\hat{\mu}$ agree that $\int f$ is to be interpreted as $\int f^+ - \int f^-$ just when this can be defined in $[-\infty, \infty]$, writing $f^+(x) = \max(f(x), 0)$, $f^-(x) = \max(-f(x), 0)$ for $x \in \operatorname{dom} f$, the result for general real-valued f follows at once.

212G I turn now to the question of the effect of the construction on the properties listed in 211B-211K.

Proposition Let (X, Σ, μ) be a measure space, and $(X, \hat{\Sigma}, \hat{\mu})$ its completion.

(a) $(X, \hat{\Sigma}, \hat{\mu})$ is a probability space, or totally finite, or σ-finite, or semi-finite, or localizable, iff (X, Σ, μ) is.

(b) $(X, \hat{\Sigma}, \hat{\mu})$ is strictly localizable if (X, Σ, μ) is, and any decomposition of X for μ is a decomposition for $\hat{\mu}$.

(c) A set $H \in \hat{\Sigma}$ is an atom for $\hat{\mu}$ iff there is an $E \in \Sigma$ such that E is an atom for μ and $\hat{\mu}(H \triangle E) = 0$.

(d) $(X, \hat{\Sigma}, \hat{\mu})$ is atomless or purely atomic iff (X, Σ, μ) is.

proof (a)(i) Because $\hat{\mu} X = \mu X$, $(X, \hat{\Sigma}, \hat{\mu})$ is a probability space, or totally finite, iff (X, Σ, μ) is.

(ii)(α) If (X, Σ, μ) is σ-finite, there is a sequence $\langle E_n \rangle_{n\in\mathbb{N}}$, covering X, with $\mu E_n < \infty$ for each n. Now $\hat{\mu} E_n < \infty$ for each n, so $(X, \hat{\Sigma}, \hat{\mu})$ is σ-finite.

(β) If $(X, \hat{\Sigma}, \hat{\mu})$ is σ-finite, there is a sequence $\langle E_n \rangle_{n\in\mathbb{N}}$, covering X, with $\hat{\mu} E_n < \infty$ for each n. Now we can find, for each n, an $E''_n \in \Sigma$ such that $\mu E''_n < \infty$ and $E_n \subseteq E''_n$; so that $\langle E''_n \rangle_{n\in\mathbb{N}}$ witnesses that (X, Σ, μ) is σ-finite.

(iii)(α) If (X, Σ, μ) is semi-finite and $\hat{\mu} E = \infty$, then there is an $E' \in \Sigma$ such that $E' \subseteq E$ and $\mu E' = \infty$. Next, there is an $F \in \Sigma$ such that $F \subseteq E'$ and $0 < \mu F < \infty$. Of course we now have $F \in \hat{\Sigma}$, $F \subseteq E$ and $0 < \hat{\mu} F < \infty$. As E is arbitrary, $(X, \hat{\Sigma}, \mu)$ is semi-finite.

(β) If $(X,\hat\Sigma,\hat\mu)$ is semi-finite and $\mu E=\infty$, then $\hat\mu E=\infty$, so there is an $F\subseteq E$ such that $0<\hat\mu F<\infty$. Next, there is an $F'\in\Sigma$ such that $F'\subseteq F$ and $\mu F'=\hat\mu F$. Of course we now have $F'\subseteq E$ and $0<\mu F'<\infty$. As E is arbitrary, (X,Σ,μ) is semi-finite.

(iv)(α) If (X,Σ,μ) is localizable and $\mathcal{E}\subseteq\hat\Sigma$, then set

$$\mathcal{F}=\{F:F\in\Sigma,\ \exists\ E\in\mathcal{E},\ F\subseteq E\}.$$

Let H be an essential supremum of $\mathcal{F}$ in Σ, as in 211G.

If $E\in\mathcal{E}$, there is an $E'\in\Sigma$ such that $E'\subseteq E$ and $E\setminus E'$ is negligible; now $E'\in\mathcal{F}$, so

$$\hat\mu(E\setminus H)\le\hat\mu(E\setminus E')+\mu(E'\setminus H)=0.$$

If $G\in\hat\Sigma$ and $\hat\mu(E\setminus G)=0$ for every $E\in\mathcal{E}$, let $G''\in\Sigma$ be such that $G\subseteq G''$ and $\hat\mu(G''\setminus G)=0$; then, for any $F\in\mathcal{F}$, there is an $E\in\mathcal{E}$ including F, so that

$$\mu(F\setminus G'')\le\hat\mu(E\setminus G)=0.$$

As F is arbitrary, $\mu(H\setminus G'')=0$ and $\hat\mu(H\setminus G)=0$. This shows that H is an essential supremum of $\mathcal{E}$ in $\hat\Sigma$. As $\mathcal{E}$ is arbitrary, $(X,\hat\Sigma,\hat\mu)$ is localizable.

(β) Suppose that $(X,\hat\Sigma,\hat\mu)$ is localizable and that $\mathcal{E}\subseteq\Sigma$. Working in $(X,\hat\Sigma,\hat\mu)$, let H be an essential supremum for $\mathcal{E}$ in $\hat\Sigma$. Let $H'\in\Sigma$ be such that $H'\subseteq H$ and $\hat\mu(H\setminus H')=0$. Then

$$\mu(E\setminus H')\le\hat\mu(E\setminus H)+\hat\mu(H\setminus H')=0$$

for every $E\in\mathcal{E}$; while if $G\in\Sigma$ and $\mu(E\setminus G)=0$ for every $E\in\mathcal{E}$, we must have

$$\mu(H'\setminus G)\le\hat\mu(H\setminus G)=0.$$

Thus H' is an essential supremum of $\mathcal{E}$ in Σ. As $\mathcal{E}$ is arbitrary, (X,Σ,μ) is localizable.

(b) Let $\langle X_i\rangle_{i\in I}$ be a decomposition of X for μ, as in 211E. Of course it is a partition of X into sets of finite $\hat\mu$-measure. If $H\subseteq X$ and $H\cap X_i\in\hat\Sigma$ for every i, choose for each $i\in I$ sets E_i', $E_i''\in\Sigma$ such that

$$E_i'\subseteq H\cap X_i\subseteq E_i'',\quad \mu(E_i''\setminus E_i')=0.$$

Set $E'=\bigcup_{i\in I}E_i'$, $E''=\bigcup_{i\in I}(E_i''\cap X_i)$. Then $E'\cap X_i=E_i'$, $E''\cap X_i=E_i''\cap X_i$ for each i, so E' and E'' belong to Σ. Also

$$\mu(E''\setminus E')=\sum_{i\in I}\mu(E_i''\cap X_i\setminus E_i')=0.$$

As $E'\subseteq H\subseteq E''$, $H\in\hat\Sigma$ and

$$\hat\mu H=\mu E'=\sum_{i\in I}\mu E_i'=\sum_{i\in I}\hat\mu(H\cap X_i).$$

As H is arbitrary, $\langle X_i\rangle_{i\in I}$ is a decomposition of X for $\hat\mu$.

Accordingly, $(X,\hat\Sigma,\hat\mu)$ is strictly localizable if such a decomposition exists, which is so if (X,Σ,μ) is strictly localizable.

(c)-(d)(i) Suppose that $E\in\hat\Sigma$ is an atom for $\hat\mu$. Let $E'\in\Sigma$ be such that $E'\subseteq E$ and $\hat\mu(E\setminus E')=0$. Then $\mu E'=\hat\mu E>0$. If $F\in\Sigma$ and $F\subseteq E'$, then $F\subseteq E$, so either $\mu F=\hat\mu F=0$ or $\mu(E'\setminus F)=\hat\mu(E\setminus F)=0$. As F is arbitrary, E' is an atom for μ, and $\hat\mu(E\triangle E')=\hat\mu(E\setminus E')=0$.

(ii) Suppose that $E\in\Sigma$ is an atom for μ, and that $H\in\hat\Sigma$ is such that $\hat\mu(H\triangle E)=0$. Then $\hat\mu H=\mu E>0$. If $F\in\hat\Sigma$ and $F\subseteq H$, let $F'\subseteq F$ be such that $F'\in\Sigma$ and $\hat\mu(F\setminus F')=0$. Then $E\cap F'\subseteq E$ and $\hat\mu(F\triangle(E\cap F'))=0$, so either $\hat\mu F=\mu(E\cap F')=0$ or $\hat\mu(H\setminus F)=\mu(E\setminus F')=0$. As F is arbitrary, H is an atom for $\hat\mu$.

(iii) It follows at once that $(X,\hat\Sigma,\hat\mu)$ is atomless iff (X,Σ,μ) is.

(iv)(α) On the other hand, if (X,Σ,μ) is purely atomic and $\hat\mu H>0$, there is an $E\in\Sigma$ such that $E\subseteq H$ and $\mu E>0$, and an atom F for μ such that $F\subseteq E$; but F is also an atom for $\hat\mu$. As H is arbitrary, $(X,\hat\Sigma,\hat\mu)$ is purely atomic.

(β) And if $(X,\hat\Sigma,\hat\mu)$ is purely atomic and $\mu E>0$, then there is an $H\subseteq E$ which is an atom for $\hat\mu$; now let $F\in\Sigma$ be such that $F\subseteq H$ and $\hat\mu(H\setminus F)=0$, so that F is an atom for μ and $F\subseteq E$. As E is arbitary, (X,Σ,μ) is purely atomic.

212X Basic exercises >(a) Let (X,Σ,μ) be a complete measure space. Suppose that $A\subseteq E\in\Sigma$ and that $\mu^*A+\mu^*(E\setminus A)=\mu E<\infty$. Show that $A\in\Sigma$.

>(b) Let μ and ν be two measures on a set X, with completions $\hat{\mu}$ and $\hat{\nu}$. Show that the following are equiveridical: (i) the outer measures μ^*, ν^* defined from μ and ν coincide; (ii) $\hat{\mu}E = \hat{\nu}E$ whenever either is defined and finite; (iii) $\int f d\mu = \int f d\nu$ whenever f is a real-valued function such that either integral is defined and finite. (*Hint*: for (i)$\Rightarrow$(ii), if $\hat{\mu}E < \infty$, take a measurable envelope F of E for ν and calculate $\nu^*E + \nu^*(F \setminus E)$.)

(c) Let μ be the restriction of Lebesgue measure to the Borel σ-algebra of $\mathbb{R}$, as in 211P. Show that its completion is Lebesgue measure itself. (*Hint*: 134F.)

(d) Repeat 212Xc for (i) Lebesgue measure on $\mathbb{R}^r$ (ii) Lebesgue-Stieltjes measures on $\mathbb{R}$ (114Xa).

(e) Let X be a set and Σ a σ-algebra of subsets of X. Let $\mathcal{I}$ be a σ-ideal of subsets of X (112Db). (i) Show that $\Sigma_1 = \{E \triangle A : E \in \Sigma,\ A \in \mathcal{I}\}$ is a σ-algebra of subsets of X. (ii) Let Σ_2 be the family of sets $E \subseteq X$ such that there are E', $E'' \in \Sigma$ with $E' \subseteq E \subseteq E''$ and $E'' \setminus E' \in \mathcal{I}$. Show that Σ_2 is a σ-algebra of subsets of X and that $\Sigma_2 \subseteq \Sigma_1$. (iii) Show that $\Sigma_2 = \Sigma_1$ iff every member of $\mathcal{I}$ is included in a member of $\Sigma \cap \mathcal{I}$.

(f) Let (X, Σ, μ) be a measure space, Y any set and $\phi : X \to Y$ a function. Set $\theta B = \mu^*\phi^{-1}[B]$ for every $B \subseteq Y$. (i) Show that θ is an outer measure on Y. (ii) Let ν be the measure defined from θ by Carathéodory's method, and T its domain. Show that if $C \subseteq Y$ and $\phi^{-1}[C] \in \Sigma$ then $C \in \mathrm{T}$. (iii) Suppose that (X, Σ, μ) is complete and totally finite. Show that ν is the image measure $\mu\phi^{-1}$.

(g) Let g, h be two non-decreasing functions from $\mathbb{R}$ to itself, and μ_g, μ_h the associated Lebesgue-Stieltjes measures. Show that a real-valued function f defined on a subset of $\mathbb{R}$ is μ_{g+h}-integrable iff it is both μ_g-integrable and μ_h-integrable, and that then $\int f d\mu_{g+h} = \int f d\mu_g + \int f d\mu_h$. (*Hint*: 114Yb).

(h) Let (X, Σ, μ) be a measure space, and $\mathcal{I}$ a σ-ideal of subsets of X; set $\Sigma_1 = \{E \triangle A : E \in \Sigma,\ A \in \mathcal{I}\}$, as in 212Xe. Show that if every member of $\Sigma \cap \mathcal{I}$ is μ-negligible, then there is a unique extension of μ to a measure μ_1 with domain Σ_1 such that $\mu_1 A = 0$ for every $A \in \mathcal{I}$.

(i) Let (X, Σ, μ) be a complete measure space such that $\mu X > 0$, Y a set, $f : X \to Y$ a function and μf^{-1} the image measure on Y. Show that if $\mathcal{F}$ is the filter of μ-conegligible subsets of X, then the image filter $f[[\mathcal{F}]]$ (2A1Ib) is the filter of μf^{-1}-conegligible subsets of Y.

212Y Further exercises (a) Let X be a set and ϕ an inner measure on X, that is, a functional from $\mathcal{P}X$ to $[0, \infty]$ such that

$\phi\emptyset = 0$,

$\phi(A \cup B) \geq \phi A + \phi B$ if $A \cap B = \emptyset$,

$\phi(\bigcap_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} \phi A_n$ whenever $\langle A_n\rangle_{n\in\mathbb{N}}$ is a non-increasing sequence of subsets of X and $\phi A_0 < \infty$,

if $\phi A = \infty$, $a \in \mathbb{R}$ there is a $B \subseteq A$ such that $a \leq \phi B < \infty$.

Let μ be the measure defined from ϕ, that is, $\mu = \phi\restriction\Sigma$, where

$$\Sigma = \{E : \phi(A) = \phi(A \cap E) + \phi(A \setminus E)\ \forall\ A \subseteq X\}$$

(113Yg). Show that μ must be complete.

212 Notes and comments The process of completion is so natural, and so universally applicable, and so convenient, that over large parts of measure theory it is reasonable to use only complete measure spaces. Indeed many authors so phrase their definitions that, explicitly or implicitly, only complete measure spaces are considered. In this treatise I avoid taking quite such a large step, even though it would simplify the statements of many of the theorems in this volume (for instance). I did take the trouble, in Volume 1, to give a definition of 'integrable function' which, in effect, looks at integrability with respect to the completion of a measure (212Fb). There are non-complete measure spaces which are worthy of study (for example, the restriction of Lebesgue measure to the Borel σ-algebra of $\mathbb{R}$ – see 211P), and some interesting questions to be dealt with in Volumes 3 and 5 apply to them. At the cost of rather a lot of verbal manoeuvres, therefore, I prefer to write theorems out in a form in which they can be applied to arbitrary measure spaces, without assuming completeness. But it would be reasonable, and indeed would sharpen your technique, if you regularly sought the alternative formulations which become natural if you are interested only in complete spaces.

213 Semi-finite, locally determined and localizable spaces

In this section I collect a variety of useful facts concerning these types of measure space. I start with the characteristic properties of semi-finite spaces (213A-213B), and continue with the complete locally determined spaces (213C) and the concept of 'c.l.d. version' (213D-213H), the most powerful of the universally available methods of modifying a measure space into a better-behaved one. I briefly discuss 'locally determined negligible sets' (213I-213L), and measurable envelopes (213L-213M), and end with results on localizable spaces (213N) and strictly localizable spaces (213O).

213A Lemma Let (X, Σ, μ) be a semi-finite measure space. Then

$$\mu E = \sup\{\mu F : F \in \Sigma,\ F \subseteq E,\ \mu F < \infty\}$$

for every $E \in \Sigma$.

proof Set $c = \sup\{\mu F : F \in \Sigma,\ F \subseteq E,\ \mu F < \infty\}$. Then surely $c \le \mu E$, so if $c = \infty$ we can stop. If $c < \infty$, let $\langle F_n \rangle_{n\in\mathbb{N}}$ be a sequence of measurable subsets of E, of finite measure, such that $\lim_{n\to\infty} \mu F_n = c$; set $F = \bigcup_{n\in\mathbb{N}} F_n$. For each $n \in \mathbb{N}$, $\bigcup_{k\le n} F_k$ is a measurable set of finite measure included in E, so $\mu(\bigcup_{k\le n} F_k) \le c$, and

$$\mu F = \lim_{n\to\infty} \mu(\textstyle\bigcup_{k\le n} F_k) \le c.$$

Also

$$\mu F \ge \sup_{n\in\mathbb{N}} \mu F_n \ge c,$$

so $\mu F = c$.

If F' is a measurable subset of $E \setminus F$ and $\mu F' < \infty$, then $F \cup F'$ has finite measure and is included in E, so has measure at most $c = \mu F$; it follows that $\mu F' = 0$. But this means that $\mu(E \setminus F)$ cannot be infinite, since then, because (X, Σ, μ) is semi-finite, it would have to include a measurable set of non-zero finite measure. So $E \setminus F$ has finite measure, and is therefore in fact negligible; and $\mu E = c$, as claimed.

213B Proposition Let (X, Σ, μ) be a semi-finite measure space. Let f be a μ-virtually measurable $[0, \infty]$-valued function defined almost everywhere in X. Then

$$\begin{aligned}\int f &= \sup\{\int g : g \text{ is a simple function, } g \le_{\text{a.e.}} f\} \\ &= \sup_{F\in\Sigma, \mu F<\infty} \int_F f\end{aligned}$$

in $[0, \infty]$.

proof (a) For any measure space (X, Σ, μ), a $[0, \infty]$-valued function defined on a subset of X is integrable iff there is a conegligible set E such that

(α) $E \subseteq \operatorname{dom} f$ and $f \restriction E$ is measurable,

(β) $\sup\{\int g : g \text{ is a simple function, } g \le_{\text{a.e.}} f\}$ is finite,

(γ) for every $\epsilon > 0$, $\{x : x \in E,\ f(x) \ge \epsilon\}$ has finite measure,

(δ) f is finite almost everywhere

(see 122Ja, 133B). But if μ is semi-finite, (γ) and (δ) are consequences of the rest. **P** Let $\epsilon > 0$. Set

$$E_\epsilon = \{x : x \in E,\ f(x) \ge \epsilon\},$$

$$c = \sup\{\textstyle\int g : g \text{ is a simple function, } g \le_{\text{a.e.}} f\};$$

we are supposing that c is finite. If $F \subseteq E_\epsilon$ is measurable and $\mu F < \infty$, then $\epsilon\chi F$ is a simple function and $\epsilon\chi F \le_{\text{a.e.}} f$, so

$$\epsilon\mu F = \textstyle\int \epsilon\chi F \le c, \quad \mu F \le c/\epsilon.$$

As F is arbitrary, 213A tells us that $\mu E_\epsilon \le c/\epsilon$ is finite. As ϵ is arbitrary, (γ) is satisfied.

As for (δ), if $F = \{x : x \in E,\ f(x) = \infty\}$ then μF is finite (by (γ)) and $n\chi F \le_{\text{a.e.}} f$, so $n\mu F \le c$, for every $n \in \mathbb{N}$, so $\mu F = 0$. **Q**

(b) Now suppose that $f : D \to [0, \infty]$ is a μ-virtually measurable function, where $D \subseteq X$ is conegligible, so that $\int f$ is defined in $[0, \infty]$ (135F). Then (a) tells us that

$$\int f = \sup_{g \text{ is simple}, g \le f \text{ a.e.}} \int g$$

(if either is finite, and therefore also if either is infinite)

$$= \sup_{g \text{ is simple}, g \le f \text{ a.e.}, \mu F < \infty} \int_F g \le \sup_{\mu F < \infty} \int_F f \le \int f,$$

so we have the equalities we seek.

***213C Proposition** Let (X, Σ, μ) be a complete locally determined measure space, and μ^* the outer measure derived from μ (132A-132B). Then the measure defined from μ^* by Carathéodory's method is μ itself.

proof Write $\check{\mu}$ for the measure defined by Carathéodory's method from μ^*, and $\check{\Sigma}$ for its domain.

(a) If $E \in \Sigma$ and $A \subseteq X$ then $\mu^*(A \cap E) + \mu^*(A \setminus E) = \mu^* A$ (132Af), so $E \in \check{\Sigma}$. Now $\check{\mu} E = \mu^* E = \mu E$ (132Ac). Thus $\Sigma \subseteq \check{\Sigma}$ and $\mu = \check{\mu} \restriction \Sigma$.

(b) Now suppose that $H \in \check{\Sigma}$. Let $E \in \Sigma$ be such that $\mu E < \infty$. Then $H \cap E \in \Sigma$. **P** Let $E_1, E_2 \in \Sigma$ be measurable envelopes of $E \cap H$, $E \setminus H$ respectively, both included in E (132Ee). Because $H \in \check{\Sigma}$,

$$\mu E_1 + \mu E_2 = \mu^*(E \cap H) + \mu^*(E \setminus H) = \mu^* E = \mu E.$$

As $E_1 \cup E_2 = E$,

$$\mu(E_1 \cap E_2) = \mu E_1 + \mu E_2 - \mu E = 0.$$

Now $E_1 \setminus (E \cap H) \subseteq E_1 \cap E_2$; because μ is complete, $E_1 \setminus (E \cap H)$ and $E \cap H$ belong to Σ. **Q**

As E is arbitrary, and μ is locally determined, $H \in \Sigma$. As H is arbitrary, $\check{\Sigma} = \Sigma$ and $\check{\mu} = \mu$.

213D C.l.d. versions: Proposition Let (X, Σ, μ) be a measure space. Write $(X, \hat{\Sigma}, \hat{\mu})$ for its completion (212C) and Σ^f for $\{E : E \in \Sigma, \mu E < \infty\}$. Set

$$\tilde{\Sigma} = \{H : H \subseteq X, H \cap E \in \hat{\Sigma} \text{ for every } E \in \Sigma^f\},$$

and for $H \in \tilde{\Sigma}$ set

$$\tilde{\mu} H = \sup\{\hat{\mu}(H \cap E) : E \in \Sigma^f\}.$$

Then $(X, \tilde{\Sigma}, \tilde{\mu})$ is a complete locally determined measure space.

proof (a) I check first that $\tilde{\Sigma}$ is a σ-algebra. **P** (i) $\emptyset \cap E = \emptyset \in \hat{\Sigma}$ for every $E \in \Sigma^f$, so $\emptyset \in \tilde{\Sigma}$. (ii) if $H \in \tilde{\Sigma}$ then

$$(X \setminus H) \cap E = E \setminus (E \cap H) \in \hat{\Sigma}$$

for every $E \in \Sigma^f$, so $X \setminus H \in \tilde{\Sigma}$. (iii) If $\langle H_n \rangle_{n \in \mathbb{N}}$ is a sequence in $\tilde{\Sigma}$ with union H, then

$$H \cap E = \bigcup_{n \in \mathbb{N}} H \cap H_n \in \hat{\Sigma}$$

for every $E \in \Sigma^f$, so $H \in \tilde{\Sigma}$. **Q**

(b) It is obvious that $\tilde{\mu} \emptyset = 0$. If $\langle H_n \rangle_{n \in \mathbb{N}}$ is a disjoint sequence in $\tilde{\Sigma}$ with union H, then

$$\tilde{\mu} H = \sup\{\hat{\mu}(H \cap E) : E \in \Sigma^f\}$$
$$= \sup\{\sum_{n=0}^{\infty} \hat{\mu}(H_n \cap E) : E \in \Sigma^f\} \le \sum_{n=0}^{\infty} \tilde{\mu} H_n.$$

On the other hand, given $a < \sum_{n=0}^{\infty} \tilde{\mu} H_n$, there is an $m \in \mathbb{N}$ such that $a < \sum_{n=0}^{m} \tilde{\mu} H_n$; now we can find $E_0, \dots, E_m \in \Sigma^f$ such that $a \le \sum_{n=0}^{m} \hat{\mu}(H_n \cap E_n)$. Set $E = \bigcup_{n \le m} E_n \in \Sigma^f$; then

$$\tilde{\mu} H \ge \hat{\mu}(H \cap E) = \sum_{n=0}^{\infty} \hat{\mu}(H_n \cap E) \ge \sum_{n=0}^{m} \hat{\mu}(H_n \cap E_n) \ge a.$$

As a is arbitrary, $\tilde{\mu} H \ge \sum_{n=0}^{\infty} \tilde{\mu} H_n$ and $\tilde{\mu} H = \sum_{n=0}^{\infty} \tilde{\mu} H_n$.

(c) Thus $(X, \tilde{\Sigma}, \tilde{\mu})$ is a measure space. To see that it is semi-finite, note first that $\hat{\Sigma} \subseteq \tilde{\Sigma}$ (because if $H \in \hat{\Sigma}$ then surely $H \cap E \in \hat{\Sigma}$ for every $E \in \Sigma^f$), and that $\tilde{\mu} H = \hat{\mu} H$ whenever $\hat{\mu} H < \infty$ (because then, by the definition in 212Ca, there is an $E \in \Sigma^f$ such that $H \subseteq E$, so that $\tilde{\mu}H = \hat{\mu}(H \cap E) = \hat{\mu}H)$. Now suppose that $H \in \tilde{\Sigma}$ and that $\tilde{\mu}H = \infty$. There is surely an $E \in \Sigma^f$ such that $\hat{\mu}(H \cap E) > 0$; but now $0 < \hat{\mu}(H \cap E) < \infty$, so $0 < \tilde{\mu}(H \cap E) < \infty$.

(d) Thus (X, Σ, μ) is a semi-finite measure space. To see that it is locally determined, let $H \subseteq X$ be such that $H \cap G \in \tilde{\Sigma}$ whenever $G \in \tilde{\Sigma}$ and $\tilde{\mu}G < \infty$. Then, in particular, we must have $H \cap E \in \tilde{\Sigma}$ for every $E \in \Sigma^f$. But this means in fact that $H \cap E \in \hat{\Sigma}$ for every $E \in \Sigma^f$, so that $H \in \tilde{\Sigma}$. As H is arbitrary, (X, Σ, μ) is locally determined.

(e) To see that $(X, \tilde{\Sigma}, \tilde{\mu})$ is complete, suppose that $A \subseteq H \in \tilde{\Sigma}$ and that $\tilde{\mu}H = 0$. Then for every $E \in \Sigma^f$ we must have $\hat{\mu}(H \cap E) = 0$. Because $(X, \hat{\Sigma}, \hat{\mu})$ is complete, and $A \cap E \subseteq H \cap E$, $A \cap E \in \hat{\Sigma}$. As E is arbitrary, $A \in \tilde{\Sigma}$.

213E Definition For any measure space (X, Σ, μ), I will call $(X, \tilde{\Sigma}, \tilde{\mu})$, as constructed in 213D, the **c.l.d. version** ('complete locally determined version') of (X, Σ, μ); and $\tilde{\mu}$ will be the **c.l.d. version** of μ.

213F Following the same pattern as in 212E-212G, I start with some elementary remarks to facilitate manipulation of this construction.

Proposition Let (X, Σ, μ) be any measure space and $(X, \tilde{\Sigma}, \tilde{\mu})$ its c.l.d. version.

(a) $\Sigma \subseteq \tilde{\Sigma}$ and $\tilde{\mu}E = \mu E$ whenever $E \in \Sigma$ and $\mu E < \infty$ – in fact, if $(X, \hat{\Sigma}, \hat{\mu})$ is the completion of (X, Σ, μ), $\hat{\Sigma} \subseteq \tilde{\Sigma}$ and $\tilde{\mu}E = \hat{\mu}E$ whenever $\hat{\mu}E < \infty$.

(b) Writing $\tilde{\mu}^*$ and μ^* for the outer measures defined from $\tilde{\mu}$ and μ respectively, $\tilde{\mu}^*A \le \mu^*A$ for every $A \subseteq X$, with equality if μ^*A is finite. In particular, μ-negligible sets are $\tilde{\mu}$-negligible; consequently, μ-conegligible sets are $\tilde{\mu}$-conegligible.

(c) For every $H \in \tilde{\Sigma}$ there is an $E \in \Sigma$ such that $E \subseteq H$ and $\mu E = \tilde{\mu}H$; if $\tilde{\mu}H < \infty$ then $\tilde{\mu}(H \setminus E) = 0$.

proof (a) This is already covered by remarks in the proof of 213D.

(b) If $\mu^*A = \infty$ then surely $\tilde{\mu}^*A \le \mu^*A$. If $\mu^*A < \infty$, take $E \in \Sigma$ such that $A \subseteq E$ and $\mu E = \mu^*A$ (132Aa). Then

$$\tilde{\mu}^*A \le \tilde{\mu}E = \mu E = \mu^*A.$$

On the other hand, if $A \subseteq H \in \tilde{\Sigma}$, then

$$\tilde{\mu}H \ge \hat{\mu}(H \cap E) \ge \hat{\mu}^*A = \mu^*A,$$

using 212Ea. So $\mu^*A \le \tilde{\mu}^*A$ and $\mu^*A = \tilde{\mu}^*A$.

(c) Write Σ^f for $\{E : E \in \Sigma,\ \mu E < \infty\}$; then, by the definition in 213D, $\tilde{\mu}H = \sup\{\hat{\mu}(H \cap E) : E \in \Sigma^f\}$. Let $\langle E_n \rangle_{n \in \mathbb{N}}$ be a sequence in Σ^f such that $\tilde{\mu}H = \sup_{n \in \mathbb{N}} \hat{\mu}(H \cap E_n)$. For each $n \in \mathbb{N}$ there is an $F_n \in \Sigma$ such that $F_n \subseteq H \cap E_n$ and $\mu F_n = \hat{\mu}(H \cap E_n)$ (212C). Set $E = \bigcup_{n \in \mathbb{N}} F_n$. Then $E \in \Sigma$, $E \subseteq H$ and

$$\begin{aligned}\tilde{\mu}H = \sup_{n \in \mathbb{N}} \mu F_n &\le \lim_{n \to \infty} \mu(\bigcup_{i \le n} F_i) = \mu E \\ &= \lim_{n \to \infty} \tilde{\mu}(\bigcup_{i \le n} F_i) \le \tilde{\mu}H,\end{aligned}$$

so $\mu E = \tilde{\mu}H$, and if $\tilde{\mu}H < \infty$ then $\tilde{\mu}(H \setminus E) = 0$.

213G The next step is to look at functions which are measurable or integrable with respect to $\tilde{\mu}$.

Proposition Let (X, Σ, μ) be a measure space, and $(X, \tilde{\Sigma}, \tilde{\mu})$ its c.l.d. version.

(a) If a real-valued function f defined on a subset of X is μ-virtually measurable, it is $\tilde{\Sigma}$-measurable.

(b) If a real-valued function is μ-integrable, it is $\tilde{\mu}$-integrable with the same integral.

(c) If f is a $\tilde{\mu}$-integrable real-valued function, there is a μ-integrable real-valued function which is equal to f $\tilde{\mu}$-almost everywhere.

proof Write Σ^f for $\{E : E \in \Sigma,\ \mu E < \infty\}$. By 213Fa, $\tilde{\mu}$ and μ agree on Σ^f.

(a) By 212Fa, f is $\hat{\Sigma}$-measurable, where $\hat{\Sigma}$ is the domain of the completion of μ; but since $\hat{\Sigma} \subseteq \tilde{\Sigma}$, f is $\tilde{\Sigma}$-measurable.

(b)(i) If f is a μ-simple function it is $\tilde{\mu}$-simple, and $\int f d\mu = \int f d\tilde{\mu}$, because $\tilde{\mu}E = \mu E$ for every $E \in \Sigma^f$.

(ii) If f is a non-negative μ-integrable function, there is a non-decreasing sequence $\langle f_n \rangle_{n \in \mathbb{N}}$ of μ-simple functions converging to f μ-almost everywhere; now (by 213Fb) μ-negligible sets are $\tilde{\mu}$-negligible, so $\langle f_n \rangle_{n \in \mathbb{N}}$ converges to f $\tilde{\mu}$-a.e. and (by B.Levi's theorem, 123A) f is $\tilde{\mu}$-integrable, with

$$\int f d\tilde{\mu} = \lim_{n\to\infty} \int f_n d\tilde{\mu} = \lim_{n\to\infty} \int f_n d\mu = \int f d\mu.$$

(iii) In general, if $\int f d\mu$ is defined in $\mathbb{R}$, we have

$$\int f d\tilde{\mu} = \int f^+ d\tilde{\mu} - \int f^- d\tilde{\mu} = \int f^+ d\mu - \int f^- d\mu = \int f d\mu,$$

writing $f^+ = f \vee 0$, $f^- = (-f) \vee 0$.

(c)(i) Let f be a $\tilde{\mu}$-simple function. Express it as $\sum_{i=0}^n a_i \chi H_i$ where $\tilde{\mu} H_i < \infty$ for each i. Choose $E_0, \dots, E_n \in \Sigma$ such that $E_i \subseteq H_i$ and $\tilde{\mu}(H_i \setminus E_i) = 0$ for each i (using 213Fc above). Then $g = \sum_{i=0}^n a_i \chi E_i$ is μ-simple, $g = f$ $\tilde{\mu}$-a.e., and $\int g d\mu = \int f d\tilde{\mu}$.

(ii) Let f be a non-negative $\tilde{\mu}$-integrable function. Let $\langle f_n \rangle_{n\in\mathbb{N}}$ be a non-decreasing sequence of $\tilde{\mu}$-simple functions converging $\tilde{\mu}$-almost everywhere to f. For each n, choose a μ-simple function g_n equal $\tilde{\mu}$-almost everywhere to f_n. Then $\{x : g_{n+1}(x) < g_n(x)\}$ belongs to Σ^f and is $\tilde{\mu}$-negligible, therefore μ-negligible. So $\langle g_n \rangle_{n\in\mathbb{N}}$ is non-decreasing μ-almost everywhere. Because

$$\lim_{n\to\infty} \int g_n d\mu = \lim_{n\to\infty} \int f_n d\tilde{\mu} = \int f d\tilde{\mu},$$

B.Levi's theorem tells us that $\langle g_n \rangle_{n\in\mathbb{N}}$ converges μ-almost everywhere to a μ-integrable function g; because μ-negligible sets are $\tilde{\mu}$-negligible,

$$\begin{aligned}(X \setminus \operatorname{dom} f) \cup (X \setminus \operatorname{dom} g)& \\ \cup \bigcup_{n\in\mathbb{N}} &\{x : f_n(x) \neq g_n(x)\} \\ \cup &\{x : x \in \operatorname{dom} f,\ f(x) \neq \sup_{n\in\mathbb{N}} f_n(x)\} \\ \cup &\{x : x \in \operatorname{dom} g,\ g(x) \neq \sup_{n\in\mathbb{N}} g_n(x)\}\end{aligned}$$

is $\tilde{\mu}$-negligible, and $f = g$ $\tilde{\mu}$-a.e.

(iii) If f is $\tilde{\mu}$-integrable, express it as $f_1 - f_2$ where f_1 and f_2 are $\tilde{\mu}$-integrable and non-negative; then there are μ-integrable functions g_1, g_2 such that $f_1 = g_1$, $f_2 = g_2$ $\tilde{\mu}$-a.e., so that $g = g_1 - g_2$ is μ-integrable and equal to f $\tilde{\mu}$-a.e.

213H Thirdly, I turn to the effect of the construction here on the other properties being considered in this chapter.

Proposition Let (X, Σ, μ) be a measure space and $(X, \hat{\Sigma}, \hat{\mu})$ its completion, $(X, \tilde{\Sigma}, \tilde{\mu})$ its c.l.d. version.

(a) If (X, Σ, μ) is a probability space, or totally finite, or σ-finite, or strictly localizable, so is $(X, \tilde{\Sigma}, \tilde{\mu})$, and in all these cases $\tilde{\mu} = \hat{\mu}$;

(b) if (X, Σ, μ) is localizable, so is $(X, \tilde{\Sigma}, \tilde{\mu})$, and for every $H \in \tilde{\Sigma}$ there is an $E \in \Sigma$ such that $\tilde{\mu}(E \triangle H) = 0$;

(c) (X, Σ, μ) is semi-finite iff $\tilde{\mu} F = \mu F$ for every $F \in \Sigma$, and in this case $\int f d\tilde{\mu} = \int f d\mu$ whenever the latter is defined in $[-\infty, \infty]$;

(d) a set $H \in \tilde{\Sigma}$ is an atom for $\tilde{\mu}$ iff there is an atom E for μ such that $\mu E < \infty$ and $\tilde{\mu}(H \triangle E) = 0$;

(e) if (X, Σ, μ) is atomless or purely atomic, so is $(X, \tilde{\Sigma}, \tilde{\mu})$;

(f) (X, Σ, μ) is complete and locally determined iff $\tilde{\mu} = \mu$.

proof (a)(i) I start by showing that if (X, Σ, μ) is strictly localizable, then $\tilde{\mu} = \hat{\mu}$. **P** Let $\langle X_i \rangle_{i\in I}$ be a decomposition of X for μ; then it is also a decomposition for $\hat{\mu}$ (212Gb). If $H \in \tilde{\Sigma}$, we shall have $H \cap X_i \in \hat{\Sigma}$ for every i, and therefore $H \in \hat{\Sigma}$; moreover,

$$\begin{aligned}\hat{\mu} H = \sum_{i\in I} \hat{\mu}(H \cap X_i) &= \sup\{\sum_{i\in J} \hat{\mu}(H \cap X_i) : J \subseteq I \text{ is finite}\} \\ &\le \sup\{\hat{\mu}(H \cap E) : E \in \Sigma,\ \mu E < \infty\} = \tilde{\mu} H \le \hat{\mu} H.\end{aligned}$$

So $\hat{\mu} H = \tilde{\mu} H$ for every $H \in \tilde{\Sigma}$ and $\hat{\mu} = \tilde{\mu}$. **Q**

(ii) Consequently, if (X, Σ, μ) is a probability space, or totally finite, or σ-finite, or strictly localizable, so is $(X, \tilde{\Sigma}, \tilde{\mu})$, using 212Ga-212Gb to see that $(X, \hat{\Sigma}, \hat{\mu})$ has the property involved.

(b) If (X, Σ, μ) is localizable, let $\mathcal{H}$ be any subset of $\tilde{\Sigma}$. Set

$$\mathcal{E} = \{E : E \in \Sigma,\ \exists\ H \in \mathcal{H},\ E \subseteq H\}.$$

Working in (X,Σ,μ), let $F\in\Sigma$ be an essential supremum for $\mathcal{E}$.

(i) ? Suppose, if possible, that there is an $H\in\mathcal{H}$ such that $\tilde\mu(H\setminus F)>0$. Then there is an $E\in\Sigma$ such that $E\subseteq H\setminus F$ and $\mu E=\hat\mu(H\setminus F)>0$ (213Fc). This E belongs to $\mathcal{E}$ and $\mu(E\setminus F)=\mu E>0$; which is impossible if F is an essential supremum of $\mathcal{E}$. **X**

(ii) Thus $\tilde\mu(H\setminus F)=0$ for every $H\in\mathcal{H}$. Now take any $G\in\tilde\Sigma$ such that $\tilde\mu(H\setminus G)=0$ for every $H\in\mathcal{H}$. Let $E_0\in\Sigma$ be such that $E_0\subseteq F\setminus G$ and $\mu E_0=\tilde\mu(F\setminus G)$; note that $F\setminus E_0\supseteq F\cap G$. If $E\in\mathcal{E}$, there is an $H\in\mathcal{H}$ such that $E\subseteq H$, so that

$$\mu(E\setminus(F\setminus E_0))\le\tilde\mu(H\setminus(F\cap G))\le\tilde\mu(H\setminus F)+\tilde\mu(H\setminus G)=0.$$

Because F is an essential supremum for $\mathcal{E}$ in Σ,

$$0=\mu(F\setminus(F\setminus E_0))=\mu E_0=\tilde\mu(F\setminus G).$$

This shows that F is an essential supremum for $\mathcal{H}$ in $\tilde\Sigma$. As $\mathcal{H}$ is arbitrary, $(X,\tilde\Sigma,\tilde\mu)$ is localizable.

(iii) The argument of (i)-(ii) shows in fact that if $\mathcal{H}\subseteq\tilde\Sigma$ then $\mathcal{H}$ has an essential supremum F in $\tilde\Sigma$ such that F actually belongs to Σ. Taking $\mathcal{H}=\{H\}$, we see that if $H\in\tilde\Sigma$ there is an $F\in\Sigma$ such that $\mu(H\triangle F)=0$.

(c) We already know that $\tilde\mu E\le\mu E$ for every $E\in\Sigma$, with equality if $\mu E<\infty$, by 213Fa.

(i) If (X,Σ,μ) is semi-finite, then for any $F\in\Sigma$ we have

$$\begin{aligned}\mu F&=\sup\{\mu E: E\in\Sigma,\ E\subseteq F,\ \mu E<\infty\}\\&=\sup\{\tilde\mu E: E\in\Sigma,\ E\subseteq F,\ \mu E<\infty\}\le\tilde\mu F\le\mu F,\end{aligned}$$

so that $\tilde\mu F=\mu F$.

(ii) Suppose that $\tilde\mu F=\mu F$ for every $F\in\Sigma$. If $\mu F=\infty$, then there must be an $E\in\Sigma$ such that $\mu E<\infty$, $\hat\mu(F\cap E)>0$; in which case $F\cap E\in\Sigma$ and $0<\mu(F\cap E)<\infty$. As F is arbitrary, (X,Σ,μ) is semi-finite.

(iii) If f is non-negative and $\int fd\mu=\infty$, then f is μ-virtually measurable, therefore $\tilde\Sigma$-measurable (213Ga), and defined μ-almost everywhere, therefore $\tilde\mu$-almost everywhere. Now

$$\begin{aligned}\int fd\tilde\mu&=\sup\{\int g\,d\tilde\mu: g\text{ is }\tilde\mu\text{-simple, }0\le g\le f\ \tilde\mu\text{-a.e.}\}\\&\ge\sup\{\int g\,d\mu: g\text{ is }\mu\text{-simple, }0\le g\le f\ \mu\text{-a.e.}\}=\infty\end{aligned}$$

by 213B. With 213Gb, this shows that $\int fd\tilde\mu=\int fd\mu$ whenever f is non-negative and $\int fd\mu$ is defined in $[0,\infty]$. Applying this to the positive and negative parts of f, we see that $\int fd\tilde\mu=\int fd\mu$ whenever the latter is defined in $[-\infty,\infty]$.

(d)(i) If $H\in\tilde\Sigma$ is an atom for $\tilde\mu$, then (because $\tilde\mu$ is semi-finite) there is surely an $H'\in\tilde\Sigma$ such that $H'\subseteq H$ and $0<\tilde\mu H'<\infty$, and we must have $\tilde\mu(H\setminus H')=0$, so that $\tilde\mu H<\infty$. Accordingly there is an $E\in\Sigma$ such that $E\subseteq H$ and $\tilde\mu(H\setminus E)=0$ (213Fc above). We have $\mu E=\tilde\mu H>0$. If $F\in\Sigma$ and $F\subseteq E$, then either $\mu F=\tilde\mu F=0$ or $\mu(E\setminus F)=\tilde\mu(H\setminus F)=0$. Thus $E\in\Sigma$ is an atom for μ with $\tilde\mu(H\triangle E)=0$ and $\mu E=\tilde\mu H<\infty$.

(ii) If $H\in\tilde\Sigma$ and there is an atom E for μ such that $\mu E<\infty$ and $\tilde\mu(H\triangle E)=0$, let $G\in\tilde\Sigma$ be a subset of H. We have

$$\tilde\mu G\le\tilde\mu H=\mu E<\infty,$$

so there is an $F\in\Sigma$ such that $F\subseteq G$ and $\tilde\mu(G\setminus F)=0$. Now either $\tilde\mu G=\mu(E\cap F)=0$ or $\tilde\mu(H\setminus G)=\mu(E\setminus F)=0$. This is true whenever $G\in\tilde\Sigma$ and $G\subseteq H$; also $\tilde\mu H=\mu E>0$. So H is an atom for $\tilde\mu$.

(e) If (X,Σ,μ) is atomless, then $(X,\tilde\Sigma,\tilde\mu)$ must be atomless, by (d).

If (X,Σ,μ) is purely atomic and $H\in\tilde\Sigma$, $\tilde\mu H>0$, then there is an $E\in\Sigma$ such that $0<\hat\mu(H\cap E)<\infty$. Let $E_1\in\Sigma$ be such that $E_1\subseteq H\cap E$ and $\mu E_1>0$. There is an atom F for μ such that $F\subseteq E_1$; now $\mu F<\infty$ so F is an atom for $\tilde\mu$, by (d). Also $F\subseteq H$. As H is arbitrary, $(X,\tilde\Sigma,\tilde\mu)$ is purely atomic.

(f) If $\mu=\tilde\mu$, then of course (X,Σ,μ) must be complete and locally determined, because $(X,\tilde\Sigma,\tilde\mu)$ is. If (X,Σ,μ) is complete and locally determined, then $\hat\mu=\mu$ so (using the definition in 213D) $\tilde\Sigma\subseteq\Sigma$ and $\tilde\mu=\mu$, by (c) above.

213I Locally determined negligible sets The following simple idea is occasionally useful.

Definition A measure space (X,Σ,μ) has **locally determined negligible sets** if for every non-negligible $A\subseteq X$ there is an $E\in\Sigma$ such that $\mu E<\infty$ and $A\cap E$ is not negligible.

213J Proposition If a measure space (X,Σ,μ) is *either* strictly localizable *or* complete and locally determined, it has locally determined negligible sets.

proof Let $A\subseteq X$ be a set such that $A\cap E$ is negligible whenever $\mu E<\infty$; I need to show that A is negligible.

(i) If μ is strictly localizable, let $\langle X_i\rangle_{i\in I}$ be a decomposition of X. For each $i\in I$, $A\cap X_i$ is negligible, so there we can choose a negligible $E_i\in\Sigma$ such that $A\cap X_i\subseteq E_i$. Set $E=\bigcup_{i\in I}E_i\cap X_i$. Then $\mu E=\sum_{i\in I}\mu(E_i\cap X_i)=0$ and $A\subseteq E$, so A is negligible.

(ii) If μ is complete and locally determined, take any measurable set E of finite measure. Then $A\cap E$ is negligible, therefore measurable; as E is arbitrary, A is measurable; as μ is semi-finite, A is negligible.

213K Lemma If a measure space (X,Σ,μ) has locally determined negligible sets, and $\mathcal{E}\subseteq\Sigma$ has an essential supremum $H\in\Sigma$ in the sense of 211G, then $H\setminus\bigcup\mathcal{E}$ is negligible.

proof Set $A=H\setminus\bigcup\mathcal{E}$. Take any $F\in\Sigma$ such that $\mu F<\infty$. Then $F\cap A$ has a measurable envelope V say (132Ee again). If $E\in\mathcal{E}$, then

$$\mu(E\setminus(X\setminus V))=\mu(E\cap V)=\mu^*(E\cap F\cap A)=0,$$

so $H\cap V=H\setminus(X\setminus V)$ is negligible and $F\cap A$ is negligible. As F is arbitrary and μ has locally determined negligible sets, A is negligible, as claimed.

213L Proposition Let (X,Σ,μ) be a localizable measure space with locally determined negligible sets. Then every subset A of X has a measurable envelope.

proof Set

$$\mathcal{E}=\{E:E\in\Sigma,\,\mu^*(A\cap E)=\mu E<\infty\}.$$

Let G be an essential supremum for $\mathcal{E}$ in Σ.

(i) $A\setminus G$ is negligible. **P** Let F be any set of finite measure for μ. Let E be a measurable envelope of $A\cap F$. Then $E\in\mathcal{E}$ so $E\setminus G$ is negligible. But $F\cap A\setminus G\subseteq E\setminus G$, so $F\cap A\setminus G$ is negligible. Because μ has locally determined negligible sets, this is enough to show that $A\setminus G$ is negligible. **Q**

(ii) Let E_0 be a negligible measurable set including $A\setminus G$, and set $\tilde G=E_0\cup G$, so that $\tilde G\in\Sigma$, $A\subseteq\tilde G$ and $\mu(\tilde G\setminus G)=0$. **?** Suppose, if possible, that there is an $F\in\Sigma$ such that $\mu^*(A\cap F)<\mu(\tilde G\cap F)$. Let $F_1\subseteq F$ be a measurable envelope of $A\cap F$. Set $H=X\setminus(F\setminus F_1)$; then $A\subseteq H$. If $E\in\mathcal{E}$ then

$$\mu E=\mu^*(A\cap E)\le\mu(H\cap E),$$

so $E\setminus H$ is negligible; as E is arbitrary, $G\setminus H$ is negligible and $\tilde G\setminus H$ is negligible. But $\tilde G\cap F\setminus F_1\subseteq\tilde G\setminus H$ and

$$\mu(\tilde G\cap F\setminus F_1)=\mu(\tilde G\cap F)-\mu^*(A\cap F)>0.\ \mathbf{X}$$

This shows that $\tilde G$ is a measurable envelope of A, as required.

213M Corollary (a) If (X,Σ,μ) is σ-finite, then every subset of X has a measurable envelope for μ.

(b) If (X,Σ,μ) is localizable, then every subset of X has a measurable envelope for the c.l.d. version of μ.

proof (a) Use 132Ee, or 213L, 211Lc and 213L.

(b) Use 213L and the fact that the c.l.d. version of μ is localizable as well as being complete and locally determined (213Hb).

213N When we come to use the concept of 'localizability', it will frequently be through the following characterization.

Theorem Let (X,Σ,μ) be a localizable measure space. Suppose that Φ is a family of measurable real-valued functions, all defined on measurable subsets of X, such that whenever f, $g\in\Phi$ then $f=g$ almost everywhere in $\operatorname{dom}f\cap\operatorname{dom}g$. Then there is a measurable function $h:X\to\mathbb{R}$ such that every $f\in\Phi$ agrees with h almost everywhere in $\operatorname{dom}f$.

proof For $q \in \mathbb{Q}$, $f \in \Phi$ set

$$E_{fq} = \{x : x \in \operatorname{dom} f,\ f(x) \ge q\} \in \Sigma.$$

For each $q \in \mathbb{Q}$, let E_q be an essential supremum of $\{E_{fq} : f \in \Phi\}$ in Σ. Set

$$h^*(x) = \sup\{q : q \in \mathbb{Q},\ x \in E_q\} \in [-\infty, \infty]$$

for $x \in X$, taking $\sup \emptyset = -\infty$ if necessary.

If f, $g \in \Phi$ and $q \in \mathbb{Q}$, then

$$\begin{aligned} E_{fq} \setminus (X \setminus (\operatorname{dom} g \setminus E_{gq})) &= E_{fq} \cap \operatorname{dom} g \setminus E_{gq} \\ &\subseteq \{x : x \in \operatorname{dom} f \cap \operatorname{dom} g,\ f(x) \ne g(x)\} \end{aligned}$$

is negligible; as f is arbitrary,

$$E_q \cap \operatorname{dom} g \setminus E_{gq} = E_q \setminus (X \setminus (\operatorname{dom} g \setminus E_{gq}))$$

is negligible. Also $E_{gq} \setminus E_q$ is negligible, so $E_{gq} \triangle (E_q \cap \operatorname{dom} g)$ is negligible. Set $H_g = \bigcup_{q \in \mathbb{Q}} E_{gq} \triangle (E_q \cap \operatorname{dom} g)$; then H_g is negligible. But if $x \in \operatorname{dom} g \setminus H_g$, then, for every $q \in \mathbb{Q}$, $x \in E_q \iff x \in E_{gq}$; it follows that for such x, $h^*(x) = g(x)$. Thus $h^* = g$ almost everywhere in $\operatorname{dom} g$; and this is true for every $g \in \Phi$.

The function h^* is not necessarily real-valued. But it is measurable, because

$$\{x : h^*(x) > a\} = \bigcup\{E_q : q \in \mathbb{Q},\ q > a\} \in \Sigma$$

for every real a. So if we modify it by setting

$$\begin{aligned} h(x) &= h^*(x) \text{ if } h(x) \in \mathbb{R}, \\ &= 0 \text{ if } h^*(x) \in \{-\infty, \infty\}, \end{aligned}$$

we shall get a measurable real-valued function $h : X \to \mathbb{R}$; and for any $g \in \Phi$, $h(x)$ will be equal to $g(x)$ at least whenever $h^*(x) = g(x)$, which is true for almost every $x \in \operatorname{dom} g$. Thus h is a suitable function.

213O There is an interesting and useful criterion for a space to be strictly localizable which I introduce at this point, though it will be used rarely in this volume.

Proposition Let (X, Σ, μ) be a complete locally determined space.

(a) Suppose that there is a disjoint family $\mathcal{E} \subseteq \Sigma$ such that (α) $\mu E < \infty$ for every $E \in \mathcal{E}$ (β) whenever $F \in \Sigma$ and $\mu F > 0$ then there is an $E \in \mathcal{E}$ such that $\mu(E \cap F) > 0$. Then (X, Σ, μ) is strictly localizable, $\bigcup \mathcal{E}$ is conegligible, and $\mathcal{E} \cup \{X \setminus \bigcup \mathcal{E}\}$ is a decomposition of X.

(b) Suppose that $\langle X_i \rangle_{i \in I}$ is a partition of X into measurable sets of finite measure such that whenever $E \in \Sigma$ and $\mu E > 0$ there is an $i \in I$ such that $\mu(E \cap X_i) > 0$. Then (X, Σ, μ) is strictly localizable, and $\langle X_i \rangle_{i \in I}$ is a decomposition of X.

proof (a)(i) The first thing to note is that if $F \in \Sigma$ and $\mu F < \infty$, there is a countable $\mathcal{E}' \subseteq \mathcal{E}$ such that $\mu(F \setminus \bigcup \mathcal{E}') = 0$. **P** Set

$$\mathcal{E}'_n = \{E : E \in \mathcal{E},\ \mu(F \cap E) \ge 2^{-n}\} \text{ for each } n \in \mathbb{N},$$

$$\mathcal{E}' = \bigcup_{n \in \mathbb{N}} \mathcal{E}'_n = \{E : E \in \mathcal{E},\ \mu(F \cap E) > 0\}.$$

Because $\mathcal{E}$ is disjoint, we must have

$$\#(\mathcal{E}'_n) \le 2^n \mu F$$

for every $n \in \mathbb{N}$, so that every $\mathcal{E}'_n$ is finite and $\mathcal{E}'$, being the union of a sequence of countable sets, is countable. Set $E' = \bigcup \mathcal{E}'$ and $F' = F \setminus E'$, so that both E' and F' belong to Σ. If $E \in \mathcal{E}'$, then $E \subseteq E'$ so $\mu(E \cap F') = \mu\emptyset = 0$; if $E \in \mathcal{E} \setminus \mathcal{E}'$, then $\mu(E \cap F') = \mu(E \cap F) = 0$. Thus $\mu(E \cap F') = 0$ for every $E \in \mathcal{E}$. By the hypothesis (β) on $\mathcal{E}$, $\mu F' = 0$, so $\mu(F \setminus \bigcup \mathcal{E}') = 0$, as required. **Q**

(ii) Now suppose that $H \subseteq X$ is such that $H \cap E \in \Sigma$ for every $E \in \mathcal{E}$. In this case $H \in \Sigma$. **P** Let $F \in \Sigma$ be such that $\mu F < \infty$. Let $\mathcal{E}' \subseteq \mathcal{E}$ be a countable set such that $\mu(F \setminus E') = 0$, where $E' = \bigcup \mathcal{E}'$. Then $H \cap (F \setminus E') \in \Sigma$ because (X, Σ, μ) is complete. But also $H \cap E' = \bigcup_{E \in \mathcal{E}'} H \cap E \in \Sigma$. So

$$H \cap F = (H \cap (F \setminus E')) \cup (F \cap (H \cap E')) \in \Sigma.$$

As F is arbitrary and (X, Σ, μ) is locally determined, $H \in \Sigma$. **Q**

(iii) We find also that $\mu H = \sum_{E\in\mathcal{E}} \mu(H \cap E)$ for every $H \in \Sigma$. $\mathbf{P}$ (α) Because $\mathcal{E}$ is disjoint, we must have $\sum_{E\in\mathcal{E}'} \mu(H \cap E) \le \mu H$ for every finite $\mathcal{E}' \subseteq \mathcal{E}$, so

$$\textstyle\sum_{E\in\mathcal{E}} \mu(H \cap E) = \sup\{\sum_{E\in\mathcal{E}'} \mu(H \cap E) : \mathcal{E}' \subseteq \mathcal{E} \text{ is finite}\} \le \mu H.$$

(β) For the reverse inequality, consider first the case $\mu H < \infty$. By (i), there is a countable $\mathcal{E}' \subseteq \mathcal{E}$ such that $\mu(H \setminus \bigcup \mathcal{E}') = 0$, so that

$$\textstyle\mu H = \mu(H \cap \bigcup \mathcal{E}') = \sum_{E\in\mathcal{E}'} \mu(H \cap E) \le \sum_{E\in\mathcal{E}} \mu(H \cap E).$$

(γ) In general, because (X, Σ, μ) is semi-finite,

$$\begin{aligned}\mu H &= \sup\{\mu F : F \subseteq H,\ \mu F < \infty\}\\ &\le \sup\{\sum_{E\in\mathcal{E}} \mu(F \cap E) : F \subseteq H,\ \mu F < \infty\} \le \sum_{E\in\mathcal{E}} \mu(H \cap E).\end{aligned}$$

So in all cases we have $\mu H \le \sum_{E\in\mathcal{E}} \mu(H \cap E)$, and the two are equal. $\mathbf{Q}$

(iv) In particular, setting $E_0 = X \setminus \bigcup \mathcal{E}$, $E_0 \in \Sigma$ and $\mu E_0 = 0$; that is, $\bigcup \mathcal{E}$ is conegligible. Consider $\mathcal{E}^* = \mathcal{E} \cup \{E_0\}$. This is a partition of X into sets of finite measure (now using the hypothesis (α) on $\mathcal{E}$). If $H \subseteq X$ is such that $H \cap E \in \Sigma$ for every $E \in \mathcal{E}^*$, then $H \in \Sigma$ and

$$\textstyle\mu H = \sum_{E\in\mathcal{E}} \mu(H \cap E) = \sum_{E\in\mathcal{E}^*} \mu(H \cap E).$$

Thus $\mathcal{E}^*$ (or, if you prefer, the indexed family $\langle E\rangle_{E\in\mathcal{E}^*}$) is a decomposition witnessing that (X, Σ, μ) is strictly localizable.

(b) Apply (a) with $\mathcal{E} = \{X_i : i \in I\}$, noting that E_0 in (iv) is empty, so can be dropped.

213X Basic exercises (a) Let (X, Σ, μ) be any measure space, μ^* the outer measure defined from μ, and $\check\mu$ the measure defined by Carathéodory's method from μ^*; write $\check\Sigma$ for the domain of $\check\mu$. Show that (i) $\check\mu$ extends the completion $\hat\mu$ of μ; (ii) if $H \subseteq X$ is such that $H \cap F \in \check\Sigma$ whenever $F \in \Sigma$ and $\mu F < \infty$, then $H \in \check\Sigma$; (iii) $(\check\mu)^* = \mu^*$, so that the integrable functions for $\check\mu$ and μ are the same (212Xb); (iv) if μ is strictly localizable then $\check\mu = \hat\mu$; (v) if μ is defined by Carathéodory's method from another outer measure, then $\mu = \check\mu$.

>(b) Let μ be counting measure restricted to the countable-cocountable σ-algebra of a set X (211R, 211Ye). (i) Show that the c.l.d. version $\tilde\mu$ of μ is just counting measure on X. (ii) Show that $\check\mu$, as defined in 213Xa, is equal to $\tilde\mu$, and in particular strictly extends the completion of μ.

(c) Let (X, Σ, μ) be any measure space. For $E \in \Sigma$ set

$$\mu_{sf} E = \sup\{\mu(E \cap F) : F \in \Sigma,\ \mu F < \infty\}.$$

(i) Show that (X, Σ, μ_{sf}) is a semi-finite measure space, and is equal to (X, Σ, μ) iff (X, Σ, μ) is semi-finite.

(ii) Show that a μ-integrable real-valued function f is μ_{sf}-integrable, with the same integral.

(iii) Show that if $E \in \Sigma$ and $\mu_{sf} E < \infty$, then E can be expressed as $E_1 \cup E_2$ where E_1, $E_2 \in \Sigma$, $\mu E_1 = \mu_{sf} E_1$ and $\mu_{sf} E_2 = 0$.

(iv) Show that if f is a μ_{sf}-integrable real-valued function on X, it is equal μ_{sf}-almost everywhere to a μ-integrable function.

(v) Show that if (X, Σ, μ_{sf}) is complete, so is (X, Σ, μ).

(vi) Show that μ and μ_{sf} have identical c.l.d. versions.

(d) Let (X, Σ, μ) be any measure space. Define $\check\mu$ as in 213Xa. Show that $(\check\mu)_{sf}$, as constructed in 213Xc, is precisely the c.l.d. version $\tilde\mu$ of μ, so that $\check\mu = \tilde\mu$ iff $\check\mu$ is semi-finite.

(e) Let (X, Σ, μ) be a measure space. For $A \subseteq X$ set $\mu_* A = \sup\{\mu E : E \in \Sigma,\ \mu E < \infty,\ E \subseteq A\}$, as in 113Yh. (i) Show that the measure constructed from μ_* by the method of 113Yg is just the c.l.d. version $\tilde\mu$ of μ. (ii) Show that $\tilde\mu_* = \mu_*$. (iii) Show that if ν is another measure on X, with domain T, then $\tilde\mu = \tilde\nu$ iff $\mu_* = \nu_*$.

(f) Let X be a set and θ an outer measure on X. Show that θ_{sf}, defined by writing

$$\theta_{sf} A = \sup\{\theta B : B \subseteq A,\ \theta B < \infty\}$$

is also an outer measure on X. Show that the measures defined by Carathéodory's method from θ, θ_{sf} have the same domains.

(g) Let (X,Σ,μ) be any measure space. Set

$$\mu^*_{sf}A=\sup\{\mu^*(A\cap E):E\in\Sigma,\,\mu E<\infty\}$$

for every $A\subseteq X$.

(i) Show that

$$\mu^*_{sf}A=\sup\{\mu^*B:B\subseteq A,\,\mu^*B<\infty\}$$

for every A.

(ii) Show that μ^*_{sf} is an outer measure.

(iii) Show that if $A\subseteq X$ and $\mu^*_{sf}A<\infty$, there is an $E\in\Sigma$ such that $\mu^*_{sf}A=\mu^*(A\cap E)=\mu E$, $\mu^*_{sf}(A\setminus E)=0$. (*Hint*: take a non-decreasing sequence $\langle E_n\rangle_{n\in\mathbb{N}}$ of measurable sets of finite measure such that $\mu^*_{sf}A=\lim_{n\to\infty}\mu^*(A\cap E_n)$, and let $E\subseteq\bigcup_{n\in\mathbb{N}}E_n$ be a measurable envelope of $A\cap\bigcup_{n\in\mathbb{N}}E_n$.)

(iv) Show that the measure defined from μ^*_{sf} by Carathéodory's method is precisely the c.l.d. version $\tilde\mu$ of μ.

(v) Show that $\mu^*_{sf}=\tilde\mu^*$, so that if μ is complete and locally determined then $\mu^*_{sf}=\mu^*$.

>(h) Let (X,Σ,μ) be a strictly localizable measure space with a decomposition $\langle X_i\rangle_{i\in I}$. Show that $\mu^*A=\sum_{i\in I}\mu^*(A\cap X_i)$ for every $A\subseteq X$.

>(i) Let (X,Σ,μ) be a complete locally determined measure space, and let $A\subseteq X$ be such that $\max(\mu^*(E\cap A),\mu^*(E\setminus A))<\mu E$ whenever $E\in\Sigma$ and $0<\mu E<\infty$. Show that $A\in\Sigma$. (*Hint*: given $\mu F<\infty$, consider the intersection E of measurable envelopes of $F\cap A$, $F\setminus A$ to see that $\mu^*(F\cap A)+\mu^*(F\setminus A)=\mu F$.)

>(j) Let (X,Σ,μ) be a measure space, $\tilde\mu$ its c.l.d. version, and $\check\mu$ the measure defined by Carathéodory's method from μ^*. (i) Show that the following are equiveridical: (α) μ has locally determined negligible sets; (β) μ and $\tilde\mu$ have the same negligible sets; (γ) $\check\mu=\tilde\mu$. (ii) Show that in this case μ is semi-finite.

(k) Let (X,Σ,μ) be a measure space. Show that the following are equiveridical: (i) (X,Σ,μ) has locally determined negligible sets; (ii) the completion $\hat\mu$ and c.l.d. version $\tilde\mu$ of μ have the same sets of finite measure; (iii) μ and $\tilde\mu$ have the same integrable functions; (iv) $\tilde\mu^*=\mu^*$; (v) the outer measure μ^*_{sf} of 213Xg is equal to μ^*.

(l) Let us say that a measure space (X,Σ,μ) has the **measurable envelope property** if every subset of X has a measurable envelope. (i) Show that a semi-finite space with the measurable envelope property has locally determined negligible sets. (ii) Show that a complete semi-finite space with the measurable envelope property is locally determined.

(m) Let (X,Σ,μ) be a semi-finite measure space, and suppose that it satisfies the conclusion of Theorem 213N. Show that it is localizable. (*Hint*: given $\mathcal{E}\subseteq\Sigma$, set $\mathcal{F}=\{F:F\in\Sigma,\,E\cap F$ is negligible for every $E\in\mathcal{E}\}$. Let Φ be the set of functions f from subsets of X to $\{0,1\}$ such that $f^{-1}[\{1\}]\in\mathcal{E}$ and $f^{-1}[\{0\}]\in\mathcal{F}$.)

(n) Let (X,Σ,μ) be a measure space. Show that its c.l.d. version is strictly localizable iff there is a disjoint family $\mathcal{E}\subseteq\Sigma$ such that $\mu E<\infty$ for every $E\in\mathcal{E}$ and whenever $F\in\Sigma$, $0<\mu F<\infty$ there is an $E\in\mathcal{E}$ such that $\mu(E\cap F)>0$.

(o) Show that the c.l.d. version of any point-supported measure is point-supported.

213Y Further exercises (a) Set $X=\mathbb{N}$, and for $A\subseteq X$ set

$$\theta A=\sqrt{\#(A)}\text{ if }A\text{ is finite, }\infty\text{ if }A\text{ is infinite.}$$

Show that θ is an outer measure on X, that $\theta A=\sup\{\theta B:B\subseteq A,\,\theta B<\infty\}$ for every $A\subseteq X$, but that the measure μ defined from θ by Carathéodory's method is not semi-finite. Show that if $\check\mu$ is the measure defined by Carathéodory's method from μ^* (213Xa), then $\check\mu\ne\mu$.

(b) Set $X=[0,1]\times\{0,1\}$, and let Σ be the family of those subsets E of X such that

$$\{x:x\in[0,1],\,E[\{x\}]\ne\emptyset,\,E[\{x\}]\ne\{0,1\}\}$$

is countable, writing $E[\{x\}]=\{y:(x,y)\in E\}$ for each $x\in[0,1]$. Show that Σ is a σ-algebra of subsets of X. For $E\in\Sigma$, set $\mu E=\#(\{x:(x,1)\in E\})$ if this is finite, ∞ otherwise. Show that μ is a complete semi-finite measure. Show that the measure $\check\mu$ defined from μ^* by Carathéodory's method (213Xa) is not semi-finite. Show that the domain of the c.l.d. version of μ is the whole of $\mathcal{P}X$.

(c) Set $X = \mathbb{N}$, and for $A \subseteq X$ set

$$\phi A = \#(A)^2 \text{ if } A \text{ is finite, } \infty \text{ if } A \text{ is infinite.}$$

Show that ϕ satisfies the conditions of 113Yg/212Ya, but that the measure defined from ϕ by the method of 113Yg is not semi-finite.

(d) Let (X, Σ, μ) be a complete locally determined measure space. Suppose that $D \subseteq X$ and that $f : D \to \mathbb{R}$ is a function. Show that the following are equiveridical: (i) f is measurable; (ii)

$$\mu^*\{x : x \in D \cap E,\ f(x) \le a\} + \mu^*\{x : x \in D \cap E,\ f(x) \ge b\} \le \mu E$$

whenever $a < b$ in $\mathbb{R}$, $E \in \Sigma$ and $\mu E < \infty$ (iii)

$$\max(\mu^*\{x : x \in D \cap E,\ f(x) \le a\}, \mu^*\{x : x \in D \cap E,\ f(x) \ge b\}) < \mu E$$

whenever $a < b$ in $\mathbb{R}$ and $0 < \mu E < \infty$. (*Hint*: for (iii)$\Rightarrow$(i), show that if $E \subseteq X$ then

$$\mu^*\{x : x \in D \cap E,\ f(x) > a\} = \sup_{b>a} \mu^*\{x : x \in D \cap E,\ f(x) \ge b\},$$

and use 213Xi above.)

(e) Let (X, Σ, μ) be a complete locally determined measure space and suppose that $\mathcal{E} \subseteq \Sigma$ is such that $\mu E < \infty$ for every $E \in \mathcal{E}$ and whenever $F \in \Sigma$ and $\mu F < \infty$ there is a countable $\mathcal{E}_0 \subseteq \mathcal{E}$ such that $F \setminus \bigcup \mathcal{E}_0$, $F \cap \bigcup(\mathcal{E} \setminus \mathcal{E}_0)$ are negligible. Show that (X, Σ, μ) is strictly localizable.

(f) Let (X, Σ, μ) be a measure space. Show that μ is semi-finite iff there is a family $\mathcal{E} \subseteq \Sigma$ such that $\mu E < \infty$ for every $E \in \mathcal{E}$ and $\mu F = \sum_{E \in \mathcal{E}} \mu(F \cap E)$ for every $F \in \Sigma$. (*Hint*: take $\mathcal{E}$ maximal subject to the intersection of any two elements being negligible.)

213 Notes and comments I think it is fair to say that if the definition of 'measure space' were re-written to exclude all spaces which are not semi-finite, nothing significant would be lost from the theory. There are solid reasons for not taking such a drastic step, starting with the fact that it would confuse everyone (if you say to an unprepared audience 'let (X, Σ, μ) be a measure space', there is a danger that some will imagine that you mean 'σ-finite measure space', but very few will suppose that you mean 'semi-finite measure space'). But the whole point of measure theory is that we distinguish between sets by their measures, and if every subset of E is either non-measurable, or negligible, or of infinite measure, the classification is too crude to support most of the usual ideas, starting, of course, with ordinary integration.

Let us say that a measurable set E is **purely infinite** if E itself and all its non-negligible measurable subsets have infinite measure. On the definition of the integral which I chose in Volume 1, every simple function, and therefore every integrable function, must be zero almost everywhere in E. This means that the whole theory of integration will ignore E entirely. Looking at the definition of 'c.l.d. version' (213D-213E), you will see that the c.l.d. version of the measure will render E negligible, as does the 'semi-finite version' described in 213Xc. These amendments do not, however, affect sets of finite measure, and consequently leave integrable functions integrable, with the same integrals.

The strongest reason we have yet seen for admitting non-semi-finite spaces into consideration is that Carathéodory's method does not always produce semi-finite spaces. (I give examples in 213Ya-213Yb; more important ones are the Hausdorff measures of §§264-265 below.) In practice the right thing to do is often to take the c.l.d. version of the measure produced by Carathéodory's construction.

It is a reasonable general philosophy, in measure theory, to say that we wish to measure as many sets, and integrate as many functions, as we can manage in a canonical way – I mean, without making blatantly arbitrary choices about the values we assign to our measure or integral. The revision of a measure μ to its c.l.d. version $\tilde{\mu}$ is about as far as we can go with an arbitrary measure space in which we have no other structure to guide our choices.

You will observe that $\tilde{\mu}$ is not as close to μ as the completion $\hat{\mu}$ of μ is; naturally so, because if $E \in \Sigma$ is purely infinite for μ then we have to choose between setting $\tilde{\mu}E = 0 \ne \mu E$ and finding some way of fitting many sets of finite measure into E; which if E is a singleton will be actually impossible, and in any case would be an arbitrary process. However the integrable functions for $\tilde{\mu}$, while not always the same as those for μ (since $\tilde{\mu}$ turns purely infinite sets into negligible ones, so that their characteristic functions become integrable), are 'nearly' the same, in the sense that any $\tilde{\mu}$-integrable function can be changed into a μ-integrable function by adjusting it on a $\tilde{\mu}$-negligible set. This corresponds, of course, to the fact that any set of finite measure for $\tilde{\mu}$ is the symmetric difference of a set of finite measure for μ and a $\tilde{\mu}$-negligible set. For sets of infinite measure this can fail, unless μ is localizable (213Hb, 213Xb).

If (X, Σ, μ) is semi-finite, or localizable, or strictly localizable, then of course it is correspondingly closer to $(X, \tilde{\Sigma}, \tilde{\mu})$, as detailed in 213Ha-c.

It is worth noting that while the measure $\check{\mu}$ obtained by Carathéodory's method directly from the outer measure μ^* defined from μ may fail to be semi-finite, even when μ is (213Yb), a simple modification of μ^* (213Xg) yields the c.l.d. version $\tilde{\mu}$ of μ, which can also be obtained from an appropriate inner measure (213Xe). The measure $\check{\mu}$ is of course related in other ways to $\tilde{\mu}$; see 213Xd.

214 Subspaces

In §131 I described a construction for subspace measures on measurable subsets. It is now time to give the generalization to subspace measures on arbitrary subsets of a measure space. The relationship between this construction and the properties listed in §211 is not quite as straightforward as one might imagine, and in this section I try to give a full account of what can be expected of subspaces in general. I think that for the present volume only (i) general subspaces of σ-finite spaces and (ii) measurable subspaces of general measure spaces will be needed in any essential way, and these do not give any difficulty; but in later volumes we shall need the full theory.

I begin with a general construction for 'subspace measures' (214A-214C), with an account of integration with respect to a subspace measure (214E-214G); these (with 131E-131H) give a solid foundation for the concept of 'integration over a subset' (214D). I present this work in its full natural generality, which will eventually be essential, but even for Lebesgue measure alone it is important to be aware of the ideas here. I continue with answers to some obvious questions concerning subspace measures and the properties of measure spaces so far considered, both for general subspaces (214I) and for measurable subspaces (214K), and I mention a basic construction for assembling measure spaces side-by-side, the 'direct sums' of 214L-214M. At the end of the section I discuss a measure extension problem (214O-214P).

214A Proposition Let (X, Σ, μ) be a measure space, and Y any subset of X. Let μ^* be the outer measure defined from μ (132A-132B), and set $\Sigma_Y = \{E \cap Y : E \in \Sigma\}$; let μ_Y be the restriction of μ^* to Σ_Y. Then (Y, Σ_Y, μ_Y) is a measure space.

proof (a) I have noted in 121A that Σ_Y is a σ-algebra of subsets of Y.

(b) Of course $\mu_Y F \in [0, \infty]$ for every $F \in \Sigma_Y$.

(c) $\mu_Y \emptyset = \mu^* \emptyset = 0$.

(d) If $\langle F_n \rangle_{n \in \mathbb{N}}$ is a disjoint sequence in Σ_Y with union F, then choose E_n, E'_n, $E \in \Sigma$ such that $F_n = Y \cap E_n$, $F_n \subseteq E'_n$ and $\mu_Y F_n = \mu E'_n$ for each n, $F \subseteq E$ and $\mu_Y F = \mu E$ (using 132Aa repeatedly). Set $G_n = E_n \cap E'_n \cap E \setminus \bigcup_{m<n} E_m$ for each $n \in \mathbb{N}$; then $\langle G_n \rangle_{n \in \mathbb{N}}$ is disjoint and $F_n \subseteq G_n \subseteq E'_n$ for each n, so $\mu_Y F_n = \mu G_n$. Also $F \subseteq \bigcup_{n \in \mathbb{N}} G_n \subseteq E$, so

$$\mu_Y F = \mu(\bigcup_{n \in \mathbb{N}} G_n) = \sum_{n=0}^{\infty} \mu G_n = \sum_{n=0}^{\infty} \mu_Y F_n.$$

As $\langle F_n \rangle_{n \in \mathbb{N}}$ is arbitrary, μ_Y is a measure.

214B Definition If (X, Σ, μ) is any measure space and Y is any subset of X, then μ_Y, defined as in 214A, is the **subspace measure** on Y.

It is worth noting the following.

214C Lemma Let (X, Σ, μ) be a measure space, Y a subset of X, μ_Y the subspace measure on Y and Σ_Y its domain. Then

(a) for any $F \in \Sigma_Y$, there is an $E \in \Sigma$ such that $F = E \cap Y$ and $\mu E = \mu_Y F$;

(b) for any $A \subseteq Y$, A is μ_Y-negligible iff it is μ-negligible;

(c)(i) if $A \subseteq X$ is μ-conegligible, then $A \cap Y$ is μ_Y-conegligible;

(ii) if $A \subseteq Y$ is μ_Y-conegligible, then $A \cup (X \setminus Y)$ is μ-conegligible;

(d) $(\mu_Y)^*$, the outer measure on Y defined from μ_Y, agrees with μ^* on $\mathcal{P}Y$;

(e) if $Z \subseteq Y \subseteq X$, then $\Sigma_Z = (\Sigma_Y)_Z$, the subspace σ-algebra of subsets of Z regarded as a subspace of (Y, Σ_Y), and $\mu_Z = (\mu_Y)_Z$ is the subspace measure on Z regarded as a subspace of (Y, μ_Y);

(f) if $Y \in \Sigma$, then μ_Y, as defined here, is exactly the subspace measure on Y defined in 131A-131B; that is, $\Sigma_Y = \Sigma \cap \mathcal{P}Y$ and $\mu_Y = \mu \restriction \Sigma_Y$.

proof (a) By the definition of Σ_Y, there is an $E_0 \in \Sigma$ such that $F = E_0 \cap Y$. By 132Aa, there is an $E_1 \in \Sigma$ such that $F \subseteq E_1$ and $\mu^* F = \mu E_1$. Set $E = E_0 \cap E_1$; this serves.

(b)(i) If A is μ_Y-negligible, there is a set $F \in \Sigma_Y$ such that $A \subseteq F$ and $\mu_Y F = 0$; now $\mu^* A \le \mu^* F = 0$ so A is μ-negligible, by 132Ad. (ii) If A is μ-negligible, there is an $E \in \Sigma$ such that $A \subseteq E$ and $\mu E = 0$; now $A \subseteq E \cap Y \in \Sigma_Y$ and $\mu_Y(E \cap Y) = 0$, so A is μ_Y-negligible.

(c) If $A \subseteq X$ is μ-conegligible, then $A \cap Y$ is μ_Y-conegligible, because $Y \setminus A = Y \cap (X \setminus A)$ is μ-negligible, therefore μ_Y-negligible. If $A \subseteq Y$ is μ_Y-conegligible, then $A \cup (X \setminus Y)$ is μ-conegligible because $X \setminus (A \cup (X \setminus Y)) = Y \setminus A$ is μ_Y-negligible, therefore μ-negligible.

(d) Let $A \subseteq Y$. (i) If $A \subseteq E \in \Sigma$, then $A \subseteq E \cap Y \in \Sigma_Y$, so $\mu_Y^* A \le \mu_Y(E \cap Y) \le \mu E$; as E is arbitrary, $\mu_Y^* A \le \mu^* A$. (ii) If $A \subseteq F \in \Sigma_Y$, there is an $E \in \Sigma$ such that $F \subseteq E$ and $\mu_Y F = \mu^* F = \mu E$; now $A \subseteq E$ so $\mu^* A \le \mu E = \mu_Y F$. As F is arbitrary, $\mu^* A \le \mu_Y^* A$.

(e) That $\Sigma_Z = (\Sigma_Y)_Z$ is immediate from the definition of Σ_Y, etc.; now

$$(\mu_Y)_Z = \mu_Y^* \restriction \Sigma_Z = \mu^* \restriction \Sigma_Z = \mu_Z$$

by (d).

(f) This is elementary, because $E \cap Y \in \Sigma$ and $\mu^*(E \cap Y) = \mu(E \cap Y)$ for every $E \in \Sigma$.

214D Integration over subsets: Definition Let (X, Σ, μ) be a measure space, Y a subset of X and f a $[-\infty, \infty]$-valued function defined on a subset of X. By $\int_Y f$ (or $\int_Y f(x)\mu(dx)$, etc.) I mean $\int (f \restriction Y) d\mu_Y$, if this exists in $[-\infty, \infty]$, following the definitions of 214A-214B, 133A and 135F, and taking $\mathrm{dom}(f \restriction Y) = Y \cap \mathrm{dom}\, f$, $(f \restriction Y)(x) = f(x)$ for $x \in Y \cap \mathrm{dom}\, f$. (Compare 131D.)

214E Proposition Let (X, Σ, μ) be a measure space, $Y \subseteq X$, and f a $[-\infty, \infty]$-valued function defined on a subset $\mathrm{dom}\, f$ of X.

(a) If f is μ-integrable then $f \restriction Y$ is μ_Y-integrable, and $\int_Y f \le \int f$ if f is non-negative.

(b) If $\mathrm{dom}\, f \subseteq Y$ and f is μ_Y-integrable, then there is a μ-integrable function $\tilde{f}$ on X, extending f, such that $\int_F \tilde{f} = \int_{F \cap Y} f$ for every $F \in \Sigma$.

proof (a)(i) If f is μ-simple, it is expressible as $\sum_{i=0}^n a_i \chi E_i$, where $E_0, \dots, E_n \in \Sigma$, $a_0, \dots, a_n \in \mathbb{R}$ and $\mu E_i < \infty$ for each i. Now $f \restriction Y = \sum_{i=0}^n a_i \chi_Y(E_i \cap Y)$, where $\chi_Y(E_i \cap Y) = (\chi E_i) \restriction Y$ is the characteristic function of $E_i \cap Y$ regarded as a subset of Y; and each $E_i \cap Y$ belongs to Σ_Y, with $\mu_Y(E_i \cap Y) \le \mu E_i < \infty$, so $f \restriction Y : Y \to \mathbb{R}$ is μ_Y-simple.

If $f : X \to \mathbb{R}$ is a non-negative simple function, it is expressible as $\sum_{i=0}^n a_i \chi E_i$ where $E_0, \dots, E_n$ are disjoint sets of finite measure (122Cb). Now $f \restriction Y = \sum_{i=0}^n a_i \chi_Y(E_i \cap Y)$ and

$$\int (f \restriction Y) d\mu_Y = \sum_{i=0}^n a_i \mu_Y(E_i \cap Y) \le \sum_{i=0}^n a_i \mu E_i = \int f d\mu$$

because $a_i \ge 0$ whenever $E_i \ne \emptyset$, so that $a_i \mu_Y(E_i \cap Y) \le a_i \mu E_i$ for every i.

(ii) If f is a non-negative μ-integrable function, there is a non-decreasing sequence $\langle f_n \rangle_{n \in \mathbb{N}}$ of non-negative μ-simple functions converging to f μ-almost everywhere; now $\langle f_n \restriction Y \rangle_{n \in \mathbb{N}}$ is a non-decreasing sequence of μ_Y-simple functions increasing to $f \restriction Y$ μ_Y-a.e. (by 214Cb), and

$$\sup_{n \in \mathbb{N}} \int (f_n \restriction Y) d\mu_Y \le \sup_{n \in \mathbb{N}} \int f_n d\mu = \int f d\mu < \infty,$$

so $\int (f \restriction Y) d\mu_Y$ exists and is at most $\int f d\mu$.

(iii) Finally, if f is any μ-integrable real-valued function, it is expressible as $f_1 - f_2$ where f_1 and f_2 are non-negative μ-integrable functions, so that $f \restriction Y = (f_1 \restriction Y) - (f_2 \restriction Y)$ is μ_Y-integrable.

(b) Let us say that if f is a μ_Y-integrable function, then an 'enveloping extension' of f is a μ-integrable function $\tilde{f}$, extending f, real-valued on $X \setminus Y$, such that $\int_F \tilde{f} = \int_{F \cap Y} f$ for every $F \in \Sigma$.

(i) If f is of the form χH, where $H \in \Sigma_Y$ and $\mu_Y H < \infty$, let $E_0 \in \Sigma$ be such that $H = Y \cap E_0$ and $E_1 \in \Sigma$ a measurable envelope for H (132Ee); then $E = E_0 \cap E_1$ is a measurable envelope for H and $H = E \cap Y$. Set $\tilde{f} = \chi E$, regarded as a function from X to $\{0, 1\}$. Then $\tilde{f} \restriction Y = f$, and for any $F \in \Sigma$ we have

$$\int_F \tilde{f} = \mu_F(E \cap F) = \mu(E \cap F) = \mu^*(H \cap F) = \mu_{Y \cap F}(H \cap F) = \int_{Y \cap F} f.$$

So $\tilde{f}$ is an enveloping extension of f.

(ii) If f, g are μ_Y-integrable functions with enveloping extensions $\tilde{f}$, $\tilde{g}$, and a, $b \in \mathbb{R}$, then $a\tilde{f} + b\tilde{g}$ extends $af + bg$ and

$$\int_F a\tilde{f}+b\tilde{g}=a\int_F\tilde{f}+b\int_F\tilde{g}$$
$$=a\int_{F\cap Y}f+b\int_{F\cap Y}g=\int_{F\cap Y}af+bg$$

for every $F\in\Sigma$, so $a\tilde{f}+b\tilde{g}$ is an enveloping extension of $af+bg$.

(iii) Putting (i) and (ii) together, we see that every μ_Y-simple function f has an enveloping extension.

(iv) Now suppose that $\langle f_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence of non-negative μ_Y-simple functions converging μ_Y-almost everywhere to a μ_Y-integrable function f. For each $n\in\mathbb{N}$ let $\tilde{f}_n$ be an enveloping extension of f_n. Then $\tilde{f}_n\le_{\text{a.e.}}\tilde{f}_{n+1}$. **P** If $F\in\Sigma$ then

$$\int_F\tilde{f}_n=\int_{F\cap Y}f_n\le\int_{F\cap Y}f_{n+1}=\int_F\tilde{f}_{n+1}.$$

So $\tilde{f}_n\le_{\text{a.e.}}\tilde{f}_{n+1}$, by 131Ha. **Q** Also

$$\lim_{n\to\infty}\int_F\tilde{f}_n=\lim_{n\to\infty}\int_{F\cap Y}f_n=\int_{F\cap Y}f$$

for every $F\in\Sigma$. Taking $F=X$ to begin with, B.Levi's theorem tells us that $h=\lim_{n\to\infty}\tilde{f}_n$ is defined (as a real-valued function) μ-almost everywhere; now letting F vary, we have $\int_F h=\int_{F\cap Y}f$ for every $F\in\Sigma$, because $h\restriction F=\lim_{n\to\infty}\tilde{f}_n\restriction F$ μ_F-a.e. (I seem to be using 214Cb here.) Now $h\restriction Y=f$ μ_Y-a.e., by 214Cb again. If we define $\tilde{f}$ by setting

$$\tilde{f}(x)=f(x)\text{ for }x\in\operatorname{dom}f,\ h(x)\text{ for }x\in\operatorname{dom}h\setminus\operatorname{dom}f,\ 0\text{ for other }x\in X,$$

then $\tilde{f}$ is defined everywhere in X and is equal to h μ-almost everywhere; so that if $F\in\Sigma$, $\tilde{f}\restriction F$ will be equal to $h\restriction F$ μ_F-almost everywhere, and

$$\int_F\tilde{f}=\int_F h=\int_{F\cap Y}f.$$

As F is arbitrary, $\tilde{f}$ is an enveloping extension of f.

(v) Thus every non-negative μ_Y-integrable function has an enveloping extension. Using (ii) again, every μ_Y-integrable function has an enveloping extension, as claimed.

214F Proposition Let (X,Σ,μ) be a measure space, Y a subset of X, and f a $[-\infty,\infty]$-valued function such that $\int_X f$ is defined in $[-\infty,\infty]$. If *either* Y is of full outer measure in X *or* f is zero almost everywhere in $X\setminus Y$, then $\int_Y f$ is defined and equal to $\int_X f$.

proof (a) Suppose first that f is non-negative, Σ-measurable and defined everywhere in X. In this case $f\restriction Y$ is Σ_Y-measurable. Set $F_{nk}=\{x:x\in X,\ f(x)\ge 2^{-n}k\}$ for $k,\ n\in\mathbb{N}$, $f_n=\sum_{k=1}^{4^n}2^{-n}\chi F_{nk}$ for $n\in\mathbb{N}$, so that $\langle f_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence of real-valued measurable functions converging everywhere to f, and $\int_X f=\lim_{n\to\infty}\int_X f_n$. For each $n\in\mathbb{N}$ and $k\ge 1$,

$$\mu_Y(F_{nk}\cap Y)=\mu^*(F_{nk}\cap Y)=\mu F_{nk}$$

either because $F_{nk}\setminus Y$ is negligible or because X is a measurable envelope of Y. So

$$\int_Y f=\lim_{n\to\infty}\int_Y f_n=\lim_{n\to\infty}\sum_{k=1}^{4^n}2^{-n}\mu_Y(F_{nk}\cap Y)$$
$$=\lim_{n\to\infty}\sum_{k=1}^{4^n}2^{-n}\mu F_{nk}=\lim_{n\to\infty}\int_X f_n=\int_X f.$$

(b) Now suppose that f is non-negative, defined almost everywhere in X and μ-virtually measurable. In this case there is a conegligible measurable set $E\subseteq\operatorname{dom}f$ such that $f\restriction E$ is measurable. Set $\tilde{f}(x)=f(x)$ for $x\in E$, 0 for $x\in X\setminus E$; then $\tilde{f}$ satisfies the conditions of (a) and $f=\tilde{f}$ μ-a.e. Accordingly $f\restriction Y=\tilde{f}\restriction Y$ μ_Y-a.e. (214Cc), and

$$\int_Y f=\int_Y\tilde{f}=\int_X\tilde{f}=\int_X f.$$

(c) Finally, for the general case, we can apply (b) to the positive and negative parts f^+, f^- of f to get

$$\int_Y f=\int_Y f^+-\int_Y f^-=\int_X f^+-\int_X f^-=\int_X f.$$

214G Corollary Let (X, Σ, μ) be a measure space, Y a subset of X, and $E \in \Sigma$ a measurable envelope of Y. If f is a $[-\infty, \infty]$-valued function such that $\int_E f$ is defined in $[-\infty, \infty]$, then $\int_Y f$ is defined and equal to $\int_E f$.

proof By 214Ce, we can identify the subspace measure μ_Y with the subspace measure $(\mu_E)_Y$ induced by the subspace measure on E. Now, regarded as a subspace of E, Y is of full outer measure, so 214F gives the result.

214H Subspaces and Carathéodory's method The following easy technical results will occasionally be useful.

Lemma Let X be a set, $Y \subseteq X$ a subset, and θ an outer measure on X.

(a) $\theta_Y = \theta \restriction \mathcal{P}Y$ is an outer measure on Y.

(b) Let μ, ν be the measures on X, Y defined by Carathéodory's method from the outer measures θ, θ_Y, and Σ, T their domains; let μ_Y be the subspace measure on Y induced by μ, and Σ_Y its domain. Then

(i) $\Sigma_Y \subseteq \mathrm{T}$ and $\nu F \le \mu_Y F$ for every $F \in \Sigma_Y$;

(ii) if $Y \in \Sigma$ then $\nu = \mu_Y$;

(iii) if $\theta = \mu^*$ (that is, θ is 'regular') then ν extends μ_Y;

(iv) if $\theta = \mu^*$ and $\theta Y < \infty$ then $\nu = \mu_Y$.

proof (a) You have only to read the definition of 'outer measure' (113A).

(b)(i) Suppose that $F \in \Sigma_Y$. Then it is of the form $E \cap Y$ where $E \in \Sigma$. If $A \subseteq Y$, then

$$\theta_Y(A \cap F) + \theta_Y(A \setminus F) = \theta(A \cap F) + \theta(A \setminus F) = \theta(A \cap E) + \theta(A \setminus E) = \theta A = \theta_Y A,$$

so $F \in \mathrm{T}$. Now

$$\nu F = \theta_Y F = \theta F \le \mu^* F = \mu_Y F.$$

(ii) Suppose that $F \in \mathrm{T}$. If $A \subseteq X$, then

$$\begin{aligned}\theta A &= \theta(A \cap Y) + \theta(A \setminus Y) = \theta_Y(A \cap Y) + \theta(A \setminus Y)\\ &= \theta_Y(A \cap Y \cap F) + \theta_Y(A \cap Y \setminus F) + \theta(A \setminus Y)\\ &= \theta(A \cap F) + \theta(A \cap Y \setminus F) + \theta(A \setminus Y)\\ &= \theta(A \cap F) + \theta((A \setminus F) \cap Y) + \theta((A \setminus F) \setminus Y) = \theta(A \cap F) + \theta(A \setminus F);\end{aligned}$$

as A is arbitrary, $F \in \Sigma$ and therefore $F \in \Sigma_Y$. Also

$$\mu_Y F = \mu F = \theta F = \theta_Y F = \nu F.$$

Putting this together with (i), we see that μ_Y and ν are identical.

(iii) Let $F \in \Sigma_Y$. Then $F \in \mathrm{T}$, by (i). Now $\nu F = \theta F = \mu^* F = \mu_Y F$. As F is arbitrary, ν extends μ_Y.

(iv) Now suppose that $F \in \mathrm{T}$. Because $\mu^* Y = \theta Y < \infty$, we have measurable envelopes E_1, E_2 of F and $Y \setminus F$ for μ (132Ee). Then

$$\begin{aligned}\theta Y &= \theta_Y Y = \theta_Y F + \theta_Y(Y \setminus F) = \theta F + \theta(Y \setminus F)\\ &= \mu^* F + \mu^*(Y \setminus F) = \mu E_1 + \mu E_2 \ge \mu(E_1 \cup E_2) = \theta(E_1 \cup E_2) \ge \theta Y,\end{aligned}$$

so $\mu E_1 + \mu E_2 = \mu(E_1 \cup E_2)$ and

$$\mu(E_1 \cap E_2) = \mu E_1 + \mu E_2 - \mu(E_1 \cup E_2) = 0.$$

As μ is complete (212A) and $E_1 \cap Y \setminus F \subseteq E_1 \cap E_2$ is μ-negligible, therefore belongs to Σ, $F = Y \cap (E_1 \setminus (E_1 \cap Y \setminus F))$ belongs to Σ_Y. Thus $\mathrm{T} \subseteq \Sigma_Y$; putting this together with (iii), we see that $\nu = \mu_Y$.

214I I now turn to the relationships between subspace measures and the classification of measure spaces developed in this chapter.

Theorem Let (X, Σ, μ) be a measure space and Y a subset of X. Let μ_Y be the subspace measure on Y and Σ_Y its domain.

(a) If (X, Σ, μ) is complete, or totally finite, or σ-finite, or strictly localizable, so is (Y, Σ_Y, μ_Y). If $\langle X_i \rangle_{i \in I}$ is a decomposition of X for μ, then $\langle X_i \cap Y \rangle_{i \in I}$ is a decomposition of Y for μ_Y.

(b) Writing $\hat{\mu}$ for the completion of μ, the subspace measure $\hat{\mu}_Y = (\hat{\mu})_Y$ is the completion of μ_Y.

(c) If (X, Σ, μ) has locally determined negligible sets, then μ_Y is semi-finite.

(d) If (X, Σ, μ) is complete and locally determined, then (Y, Σ_Y, μ_Y) is complete and semi-finite.

(e) If (X, Σ, μ) is complete, locally determined and localizable then so is (Y, Σ_Y, μ_Y).

proof (a)(i) Suppose that (X, Σ, μ) is complete. If $A \subseteq U \in \Sigma_Y$ and $\mu_Y U = 0$, there is an $E \in \Sigma$ such that $U = E \cap Y$ and $\mu E = \mu_Y U = 0$; now $A \subseteq E$ so $A \in \Sigma$ and $A = A \cap Y \in \Sigma_Y$.

(ii) $\mu_Y Y = \mu^* Y \leq \mu X$, so μ_Y is totally finite if μ is.

(iii) If $\langle X_n \rangle_{n\in\mathbb{N}}$ is a sequence of sets of finite measure for μ which covers X, then $\langle X_n \cap Y \rangle_{n\in\mathbb{N}}$ is a sequence of sets of finite measure for μ_Y which covers Y. So (Y, Σ_Y, μ_Y) is σ-finite if (X, Σ, μ) is.

(iv) Suppose that $\langle X_i \rangle_{i\in I}$ is a decomposition of X for μ. Then $\langle X_i \cap Y \rangle_{i\in I}$ is a decomposition of Y for μ_Y. **P** Because $\mu_Y(X_i \cap Y) \leq \mu X_i < \infty$ for each i, $\langle X_i \cap Y \rangle_{i\in I}$ is a partition of Y into sets of finite measure. Suppose that $U \subseteq Y$ is such that $U_i = U \cap X_i \cap Y \in \Sigma_Y$ for every i. For each $i \in I$, choose $E_i \in \Sigma$ such that $U_i = E_i \cap Y$ and $\mu E_i = \mu_Y U_i$; we may of course suppose that $E_i \subseteq X_i$. Set $E = \bigcup_{i\in I} E_i$. Then $E \cap X_i = E_i \in \Sigma$ for every i, so $E \in \Sigma$ and $\mu E = \sum_{i\in I} \mu E_i$. Now $U = E \cap Y$ so $U \in \Sigma_Y$ and

$$\mu_Y U \leq \mu E = \sum_{i\in I} \mu E_i = \sum_{i\in I} \mu_Y U_i.$$

On the other hand, $\mu_Y U$ is surely greater than or equal to $\sum_{i\in I} \mu_Y U_i = \sup_{J\subseteq I \text{ is finite}} \sum_{i\in J} \mu_Y U_i$, so they are equal. As U is arbitrary, $\langle X_i \cap Y \rangle_{i\in I}$ is a decomposition of Y for μ_Y. **Q**

Consequently (Y, Σ_Y, μ_Y) is strictly localizable if (X, Σ, μ) is.

(b) The domain of the completion $(\mu_Y)\hat{}$ is

$$\begin{aligned}\hat{\Sigma}_Y &= \{F \triangle A : F \in \Sigma_Y,\ A \subseteq Y \text{ is } \mu_Y\text{-negligible}\}\\ &= \{(E \cap Y) \triangle (A \cap Y) : E \in \Sigma,\ A \subseteq X \text{ is } \mu\text{-negligible}\}\\ &= \{(E \triangle A) \cap Y : E \in \Sigma,\ A \text{ is } \mu\text{-negligible}\} = \operatorname{dom} \hat{\mu}_Y.\end{aligned}$$

(214Cb)

If $H \in \hat{\Sigma}_Y$ then

$$(\mu_Y)\hat{}(H) = \mu_Y^* H = \mu^* H = (\hat{\mu})^* H = \hat{\mu}_Y H,$$

using 214Cd for the second step, and 212Ea for the third.

(c) Take $U \in \Sigma_Y$ such that $\mu_Y U > 0$. Then there is an $E \in \Sigma$ such that $\mu E < \infty$ and $\mu^*(E \cap U) > 0$. **P?** Otherwise, $E \cap U$ is μ-negligible whenever $\mu E < \infty$; because μ has locally determined negligible sets, U is μ-negligible and $\mu_Y U = \mu^* U = 0$. **XQ** Now $E \cap U \in \Sigma_Y$ and

$$0 < \mu^*(E \cap U) = \mu_Y(E \cap U) \leq \mu E < \infty.$$

(d) By (a), μ_Y is complete; by (c) and 213J, it is semi-finite.

(e) By (d), μ_Y is complete and semi-finite. To see that it is locally determined, take any $U \subseteq Y$ such that $U \cap V \in \Sigma_Y$ whenever $V \in \Sigma_Y$ and $\mu_Y V < \infty$. By 213L and 213J, there is a measurable envelope E of U for μ; of course $E \cap Y \in \Sigma_Y$.

I claim that $\mu(E \cap Y \setminus U) = 0$. **P** Take any $F \in \Sigma$ with $\mu F < \infty$. Then $F \cap U \in \Sigma_Y$, so

$$\mu_Y(F \cap E \cap Y) \leq \mu(F \cap E) = \mu^*(F \cap U) = \mu_Y(F \cap U) \leq \mu_Y(F \cap E \cap Y);$$

thus $\mu_Y(F \cap E \cap Y) = \mu_Y(F \cap U)$ and

$$\mu^*(F \cap E \cap Y \setminus U) = \mu_Y(F \cap E \cap Y \setminus U) = 0.$$

Because μ is complete, $\mu(F \cap E \cap Y \setminus U) = 0$; because μ is locally determined and F is arbitrary, $\mu(E \cap Y \setminus U) = 0$. **Q** But this means that $E \cap Y \setminus U \in \Sigma_Y$ and $U \in \Sigma_Y$. As U is arbitrary, μ_Y is locally determined.

To see that μ_Y is localizable, let $\mathcal{U}$ be any family in Σ_Y. Set

$$\mathcal{E} = \{E : E \in \Sigma,\ \mu E < \infty,\ \mu E = \mu^*(E \cap U) \text{ for some } U \in \mathcal{U}\},$$

and let $G \in \Sigma$ be an essential supremum for $\mathcal{E}$ in Σ. I claim that $G \cap Y$ is an essential supremum for $\mathcal{U}$ in Σ_Y. **P** (i) **?** If $U \in \mathcal{U}$ and $U \setminus (G \cap Y)$ is not negligible, then (because μ_Y is semi-finite) there is a $V \in \Sigma_Y$ such that $V \subseteq U \setminus G$ and $0 < \mu_Y V < \infty$. Now there is an $E \in \Sigma$ such that $V \subseteq E$ and $\mu E = \mu^* V$. We have $\mu^*(E \cap U) \geq \mu^* V = \mu E$, so $E \in \mathcal{E}$ and $E \setminus G$ must be negligible; but $V \subseteq E \setminus G$ is not negligible. **X** Thus $U \setminus (G \cap Y)$ is negligible for every $U \in \mathcal{U}$. (ii) If $W \in \Sigma_Y$ is such that $U \setminus W$ is negligible for every $U \in \mathcal{U}$, express W as $H \cap Y$ where $H \in \Sigma$. If $E \in \mathcal{E}$, there is a $U \in \mathcal{U}$ such that $\mu E = \mu^*(E \cap U)$; now $\mu^*(E \cap U \setminus W) = 0$, so $\mu E = \mu^*(E \cap U \cap W) \leq \mu(E \cap H)$ and $E \setminus H$ is negligible. As E is arbitrary, H is an essential upper bound for $\mathcal{E}$ and $G \setminus H$ is negligible; but this means that $G \cap Y \setminus W$ is negligible. As W is arbitrary, $G \cap Y$ is an essential supremum for $\mathcal{U}$. **Q**

As $\mathcal{U}$ is arbitrary, μ_Y is localizable.

214J Upper and lower integrals The following elementary facts are sometimes useful.

Proposition Let (X, Σ, μ) be a measure space, A a subset of X and f a real-valued function defined almost everywhere in X. Then

(a) if *either* f is non-negative *or* A has full outer measure in X, $\overline{\int}(f \restriction A) d\mu_A \leq \overline{\int} f d\mu$;

(b) if A has full outer measure in X, $\underline{\int} f d\mu \leq \underline{\int}(f \restriction A) d\mu_A$.

proof (a)(i) Suppose that f is non-negative. If $\overline{\int} f d\mu = \infty$, the result is trivial. Otherwise, there is a μ-integrable function g such that $f \leq g$ μ-a.e. and $\overline{\int} f d\mu = \int g\,d\mu$, by 133Ja. Now $f \restriction A \leq g \restriction A$ μ_A-a.e., by 214Cb, and $\int (g \restriction A)\,d\mu_A$ is defined and less than or equal to $\int g\,d\mu$, by 214Ea; so

$$\overline{\int}(f \restriction A) d\mu_A \leq \int (g \restriction A) d\mu_A \leq \int g\,d\mu = \overline{\int} f d\mu.$$

(ii) Now suppose that A has full outer measure in X. If g is such that $f \leq g$ μ-a.e. and $\int g\,d\mu$ is defined in $[-\infty, \infty]$, then $f \restriction A \leq g \restriction A$ μ_A-a.e. and $\int (g \restriction A) d\mu_A = \int g\,d\mu$, by 214F. So $\overline{\int}(f \restriction A) d\mu_A \leq \int g\,d\mu$. As g is arbitrary, $\overline{\int}(f \restriction A) d\mu_A \leq \overline{\int} f\,d\mu$.

(b) Apply (a) to $-f$, and use 133J(b-iv).

214K Measurable subspaces: Proposition Let (X, Σ, μ) be a measure space.

(a) Let $E \in \Sigma$ and let μ_E be the subspace measure, with Σ_E its domain. If (X, Σ, μ) is complete, or totally finite, or σ-finite, or strictly localizable, or semi-finite, or localizable, or locally determined, or atomless, or purely atomic, so is (E, Σ_E, μ_E).

(b) Suppose that $\langle X_i \rangle_{i \in I}$ is a partition of X into measurable sets (not necessarily of finite measure) such that

$$\Sigma = \{E : E \subseteq X,\ E \cap X_i \in \Sigma \text{ for every } i \in I\},$$

$$\mu E = \textstyle\sum_{i \in I} \mu(E \cap X_i) \text{ for every } E \in \Sigma.$$

Then (X, Σ, μ) is complete, or strictly localizable, or semi-finite, or localizable, or locally determined, or atomless, or purely atomic, iff $(X_i, \Sigma_{X_i}, \mu_{X_i})$ has that property for every $i \in I$.

proof I really think that if you have read attentively up to this point, you ought to find this easy. If you are in any doubt, this makes a very suitable set of sixteen exercises to do.

214L Direct sums Let $\langle (X_i, \Sigma_i, \mu_i) \rangle_{i \in I}$ be any indexed family of measure spaces. Set $X = \bigcup_{i \in I}(X_i \times \{i\})$; for $E \subseteq X$, $i \in I$ set $E_i = \{x : (x, i) \in E\}$. Write

$$\Sigma = \{E : E \subseteq X,\ E_i \in \Sigma_i \text{ for every } i \in I\},$$

$$\mu E = \textstyle\sum_{i \in I} \mu_i E_i \text{ for every } E \in \Sigma.$$

Then it is easy to check that (X, Σ, μ) is a measure space; I will call it the **direct sum** of the family $\langle (X_i, \Sigma_i, \mu_i) \rangle_{i \in I}$. Note that if (X, Σ, μ) is any strictly localizable measure space, with decomposition $\langle X_i \rangle_{i \in I}$, then we have a natural isomorphism between (X, Σ, μ) and the direct sum $(X', \Sigma', \mu') = \bigoplus_{i \in I}(X_i, \Sigma_{X_i}, \mu_{X_i})$ of the subspace measures, if we match $(x, i) \in X'$ with $x \in X$ for every $i \in I$ and $x \in X_i$.

For some of the elementary properties (to put it plainly, I know of no properties which are not elementary) of direct sums, see 214M and 214Xh-214Xk.

214M Proposition Let $\langle (X_i, \Sigma_i, \mu_i) \rangle_{i \in I}$ be a family of measure spaces, with direct sum (X, Σ, μ). Let f be a real-valued function defined on a subset of X. For each $i \in I$, set $f_i(x) = f(x, i)$ whenever $(x, i) \in \operatorname{dom} f$.

(a) f is measurable iff f_i is measurable for every $i \in I$.

(b) If f is non-negative, then $\int f d\mu = \sum_{i \in I} \int f_i d\mu_i$ if either is defined in $[0, \infty]$.

proof (a) For $a \in \mathbb{R}$, set $F_a = \{(x, i) : (x, i) \in \operatorname{dom} f,\ f(x, i) \geq a\}$. (i) If f is measurable, $i \in I$ and $a \in \mathbb{R}$, then there is an $E \in \Sigma$ such that $F_a = E \cap \operatorname{dom} f$; now

$$\{x : f_i(x) \geq a\} = \operatorname{dom} f_i \cap \{x : (x, i) \in E\}$$

belongs to the subspace σ-algebra on $\operatorname{dom} f_i$ induced by Σ_i. As a is arbitrary, f_i is measurable. (ii) If every f_i is measurable and $a \in \mathbb{R}$, then for each $i \in I$ there is an $E_i \in \Sigma_i$ such that $\{x : (x, i) \in F_a\} = E_i \cap \operatorname{dom} f$; setting $E = \{(x, i) : i \in I, x \in E_i\}$, $F_a = \operatorname{dom} f \cap E$ belongs to the subspace σ-algebra on $\operatorname{dom} f$. As a is arbitrary, f is measurable.

(b)(i) Suppose first that f is measurable and defined everywhere. Set $F_{nk} = \{(x, i) : (x, i) \in X,\ f(x, i) \geq 2^{-n}k\}$ for k, $n \in \mathbb{N}$, $g_n = \sum_{k=1}^{4^n} 2^{-n}\chi F_{nk}$ for $n \in \mathbb{N}$, $F_{nki} = \{x : (x, i) \in F_{nk}\}$ for k, $n \in \mathbb{N}$ and $i \in I$, $g_{ni}(x) = g_n(x, i)$ for $i \in I$, $x \in X_i$. Then

$$\begin{aligned}\int f d\mu &= \lim_{n\to\infty} \int g_n d\mu = \sup_{n\in\mathbb{N}} \sum_{k=1}^{4^n} 2^{-n}\mu F_{nk} \\ &= \sup_{n\in\mathbb{N}} \sum_{k=1}^{4^n} \sum_{i\in I} 2^{-n}\mu F_{nki} = \sum_{i\in I} \sup_{n\in\mathbb{N}} \sum_{k=1}^{4^n} 2^{-n}\mu F_{nki} \\ &= \sum_{i\in I} \sup_{n\in\mathbb{N}} \int g_{ni} d\mu_i = \sum_{i\in I} \int f_i d\mu_i.\end{aligned}$$

(ii) Generally, if $\int f d\mu$ is defined, there are a measurable $g : X \to [0, \infty[$ and a conegligible measurable set $E \subseteq \operatorname{dom} f$ such that $g = f$ on E. Now $E_i = \{x : (x, i) \in X_i\}$ belongs to Σ_i for each i, and $\sum_{i\in I} \mu_i(X_i \setminus E_i) = \mu(X \setminus E) = 0$, so E_i is μ_i-conegligible for every i. Setting $g_i(x) = g(x, i)$ for $x \in X_i$, (i) tells us that

$$\textstyle\sum_{i\in I} \int f_i d\mu_i = \sum_{i\in I} \int g_i d\mu_i = \int g\, d\mu = \int f d\mu.$$

(iii) On the other hand, if $\int f_i d\mu_i$ is defined for every $i \in I$, then for each $i \in I$ we can find a measurable function $g_i : X_i \to [0, \infty[$ and a μ_i-conegligible measurable set $E_i \subseteq \operatorname{dom} f_i$ such that $g_i = f_i$ on E_i. Setting $g(x, i) = g_i(x)$ for $i \in I$, $x \in X_i$, (a) tells us that g is measurable, while $g = f$ on $\{(x, i) : i \in I, x \in E_i\}$, which is conegligible (by the calculation in (ii) just above); so

$$\textstyle\int f d\mu = \int g\, d\mu = \sum_{i\in I} \int g_i d\mu_i = \sum_{i\in I} \int f_i d\mu_i,$$

again using (i) for the middle step.

214N Corollary Let (X, Σ, μ) be a measure space with a decomposition $\langle X_i\rangle_{i\in I}$. If f is a real-valued function defined on a subset of X, then

(a) f is measurable iff $f{\restriction}X_i$ is measurable for every $i \in I$,

(b) if $f \geq 0$, then $\int f = \sum_{i\in I} \int_{X_i} f$ if either is defined in $[0, \infty]$.

proof Apply 214M to the direct sum of $\langle(X_i, \Sigma_{X_i}, \mu_{X_i})\rangle_{i\in I}$, identified with (X, Σ, μ) as in 214L.

***214O** I make space here for a general theorem which puts rather heavy demands on the reader. So I ought to say that I advise skipping it on first reading. It will not be quoted in this volume, in the full form here I do not expect to use it anywhere in this treatise, only the special case of 214Xm is at all often applied, and the proof depends on a concept ('ideal of sets') and a technique ('transfinite induction', part (d) of the proof of 214P) which are used nowhere else in this volume. However, 'extension of measures' is one of the central themes of Volume 4, and this result may help to make sense of some of the patterns which will appear there.

Lemma Let (X, Σ, μ) be a measure space, and $\mathcal{I}$ an ideal of subsets of X, that is, a family of subsets of X such that $\emptyset \in \mathcal{I}$, $I \cup J \in \mathcal{I}$ for all I, $J \in \mathcal{I}$, and $I \in \mathcal{I}$ whenever $I \subseteq J \in \mathcal{I}$. Then there is a measure λ on X such that $\Sigma \cup \mathcal{I} \subseteq \operatorname{dom} \lambda$, $\mu E = \lambda E + \sup_{I\in\mathcal{I}} \mu^*(E \cap I)$ for every $E \in \Sigma$, and $\lambda I = 0$ for every $I \in \mathcal{I}$.

proof (a) Let Λ be the set of those $F \subseteq X$ such that there are $E \in \Sigma$ and a countable $\mathcal{J} \subseteq \mathcal{I}$ such that $E \triangle F \subseteq \bigcup \mathcal{J}$. Then Λ is a σ-algebra of subsets of X including $\Sigma \cup \mathcal{I}$. **P** $\Sigma \subseteq \Lambda$ because $E \triangle E \subseteq \bigcup \emptyset$ for every $E \in \Sigma$. $\mathcal{I} \subseteq \Lambda$ because $\emptyset \triangle I \subseteq \bigcup\{I\}$ for every $I \in \mathcal{I}$. In particular, $\emptyset \in \Lambda$. If $F \in \Lambda$, let $E \in \Sigma$ and $\mathcal{J} \subseteq \mathcal{I}$ be such that $\mathcal{J}$ is countable and $F \triangle E \subseteq \bigcup \mathcal{J}$; then $(X \setminus F) \triangle (X \setminus E) \subseteq \bigcup \mathcal{J}$ so $X \setminus F \in \Lambda$. If $\langle F_n\rangle_{n\in\mathbb{N}}$ is a sequence in Λ with union F, then for each $n \in \mathbb{N}$ choose $E_n \in \Sigma$, $\mathcal{J}_n \subseteq \mathcal{I}$ such that $\mathcal{J}_n$ is countable and $E_n \triangle F_n \subseteq \bigcup \mathcal{J}_n$; then $E = \bigcup_{n\in\mathbb{N}} E_n$ belongs to Σ, $\mathcal{J} = \bigcup_{n\in\mathbb{N}} \mathcal{J}_n$ is a countable subset of $\mathcal{I}$ and $E \triangle F \subseteq \bigcup \mathcal{J}$, so $F \in \Sigma$. Thus Λ is a σ-algebra. **Q**

(b) For $F \in \Lambda$ set

$$\lambda F = \sup\{\mu E : E \in \Sigma,\ E \subseteq F,\ \mu^*(E \cap I) = 0 \text{ for every } I \in \mathcal{I}\}.$$

Then λ is a measure. **P** The only subset of $\emptyset$ is $\emptyset$, so $\lambda\emptyset = 0$. Let $\langle F_n \rangle_{n\in\mathbb{N}}$ be a disjoint sequence in Λ with union F; set $u = \sum_{n=0}^{\infty} \lambda F_n$. (i) If $E \in \Sigma$, $E \subseteq F$ and $\mu^*(E \cap I) = 0$ for every $I \in \mathcal{I}$, then for each n set $E_n = E \cap F_n$. As $\mu^*(E_n \cap I) = 0$ for every $I \in \mathcal{I}$, $\mu E_n \le \lambda F_n$ for each n. Now $\langle E_n \rangle_{n\in\mathbb{N}}$ is disjoint and has union E, so

$$\mu E = \sum_{n=0}^{\infty} \mu E_n \le \sum_{n=0}^{\infty} \lambda F_n = u.$$

As E is arbitrary, $\lambda F \le u$. (ii) Take any $\gamma < u$. For $n \in \mathbb{N}$, set $\gamma_n = \lambda F_n - 2^{-n-1}\min(1, u-\gamma)$ if λF_n is finite, γ otherwise. For each n, we can find an $E_n \in \Sigma$ such that $E_n \subseteq F_n$, $\mu^*(E_n \cap I) = 0$ for every $I \in \mathcal{I}$, and $\mu E_n \ge \gamma_n$. Set $E = \bigcup_{n\in\mathbb{N}} E_n$; then $E \subseteq F$ and $E \cap I = \bigcup_{n\in\mathbb{N}} E_n \cap I$ is μ-negligible for every $I \in \mathcal{I}$, so $\lambda F \ge \mu E = \sum_{n=0}^{\infty} \mu E_n \ge \gamma$. As γ is arbitrary, $\lambda F \ge u$. (iii) As $\langle F_n \rangle_{n\in\mathbb{N}}$ is arbitrary, λ is a measure. **Q**

(c) Now take any $E \in \Sigma$ and set $u = \sup_{I\in\mathcal{I}} \mu^*(E \cap I)$. If $u = \infty$ then we certainly have $\mu E = \infty = \lambda E + u$. Otherwise, let $\langle I_n \rangle_{n\in\mathbb{N}}$ be a sequence in $\mathcal{I}$ such that $\lim_{n\to\infty} \mu^*(E \cap I_n) = u$; replacing I_n by $\bigcup_{m\le n} I_m$ for each n if necessary, we may suppose that $\langle I_n \rangle_{n\in\mathbb{N}}$ is non-decreasing. Set $A = E \cap \bigcup_{n\in\mathbb{N}} I_n$; because $E \cap I_n$ has finite outer measure for each n, A can be covered by a sequence of sets of finite measure, and has a measurable envelope H for μ included in E (132Ee). Observe that

$$\mu H = \mu^* A = \sup_{n\in\mathbb{N}} \mu^*(E \cap I_n) = u$$

by 132Ae.

Set $G = E \setminus H$. Then $\mu^*(G \cap I) = 0$ for every $I \in \mathcal{I}$. **P** For any $n \in \mathbb{N}$ there is an $F \in \Sigma$ such that $F \supseteq E \cap (I_n \cup I)$ and $\mu F \le u$; in which case

$$\mu^*(G \cap I) + \mu^*(E \cap I_n) \le \mu(F \setminus H) + \mu(F \cap H) \le u.$$

As n is arbitrary, $\mu^*(G \cap I) = 0$. **Q** Accordingly

$$u + \lambda E \ge \mu H + \mu G = \mu E.$$

On the other hand, if $F \in \Sigma$ is such that $F \subseteq E$ and $\mu^*(F \cap I) = 0$ for every $I \in \mathcal{I}$, then

$$\mu^*(E \cap I_n) \le \mu(E \setminus F) + \mu^*(F \cap I_n) = \mu(E \setminus F)$$

for every n, so

$$u + \mu F \le \mu(E \setminus F) + \mu F = \mu E;$$

as F is arbitrary, $u + \lambda E \le \mu E$.

(d) If $J \in \mathcal{I}$, $F \in \Sigma$, $F \subseteq J$ and $\mu^*(F \cap I) = 0$ for every $I \in \mathcal{I}$, then $F \cap J = F$ is μ-negligible; as F is arbitrary, $\lambda J = 0$. Thus λ has all the required properties.

***214P Theorem** Let (X, Σ, μ) be a measure space, and $\mathcal{A}$ a family of subsets of X which is well-ordered by the relation $\subseteq$. Then there is an extension of μ to a measure λ on X such that $\lambda(E \cap A)$ is defined and equal to $\mu^*(E \cap A)$ whenever $E \in \Sigma$ and $A \in \mathcal{A}$.

proof (a) Adding $\emptyset$ and X to $\mathcal{A}$ if necessary, we may suppose that $\mathcal{A}$ has $\emptyset$ as its least member and X as its greatest member. By 2A1Dg, $\mathcal{A}$ is isomorphic, as ordered set, to some ordinal; since $\mathcal{A}$ has a greatest member, this ordinal is a successor, expressible as $\zeta + 1$; let $\xi \mapsto A_\xi : \zeta + 1 \to \mathcal{A}$ be the order-isomorphism, so that $\langle A_\xi \rangle_{\xi\le\zeta}$ is a non-decreasing family of subsets of X, $A_0 = \emptyset$ and $A_\zeta = X$.

(b) For each ordinal $\xi \le \zeta$, write μ_ξ for the subspace measure on A_ξ, Σ_ξ for its domain and $\mathcal{I}_\xi$ for $\bigcup_{\eta<\xi} \mathcal{P}A_\eta$. Because $A_\eta \cup A_{\eta'} = A_{\max(\eta,\eta')}$ for η, $\eta' < \xi$, $\mathcal{I}_\xi$ is an ideal of subsets of A_ξ. By 214O, we have a measure λ_ξ on A_ξ, with domain $\Lambda_\xi \supseteq \Sigma_\xi \cup \mathcal{I}_\xi$, such that $\mu_\xi E = \lambda_\xi E + \sup_{I\in\mathcal{I}_\xi} \mu_\xi^*(E \cap I)$ for every $E \in \Sigma_\xi$ and $\lambda_\xi I = 0$ for every $I \in \mathcal{I}_\xi$. Because every member of $\mathcal{I}_\xi$ is included in A_η for some $\eta < \xi$, we have

$$\mu^*(E \cap A_\xi) = \lambda_\xi(E \cap A_\xi) + \sup_{\eta<\xi} \mu_\xi^*(E \cap A_\eta) = \lambda_\xi(E \cap A_\xi) + \sup_{\eta<\xi} \mu^*(E \cap A_\eta)$$

(214Cd) for every $E \in \Sigma$. Also, of course, $\lambda_\xi A_\eta = 0$ for every $\eta < \xi$.

(c) Now set

$$\Lambda = \{F : F \subseteq X,\ F \cap A_\xi \in \Lambda_\xi \text{ for every } \xi \le \zeta\},$$

$$\lambda F = \sum_{\xi\le\zeta} \lambda_\xi(F \cap A_\xi)$$

for every $F \in \Lambda$. Because Λ_ξ is a σ-algebra of subsets of A_ξ for each ξ, Λ is a σ-algebra of subsets of X; because every λ_ξ is a measure, so is λ. If $E \in \Sigma$, then

$$E \cap A_\xi \in \Sigma_\xi \subseteq \Lambda_\xi$$

for each ξ, so $E \in \Lambda$. If $\eta \le \zeta$, then for each $\xi \le \zeta$ either $\eta < \xi$ and

$$A_\eta \cap A_\xi = A_\eta \in \mathcal{I}_\xi \subseteq \Lambda_\xi$$

or $\eta \ge \xi$ and $A_\eta \cap A_\xi = A_\xi$ belongs to Λ_ξ. So $A_\eta \in \Lambda$ for every $\eta \le \zeta$.

(d) Finally, $\lambda(E \cap A_\xi) = \mu^*(E \cap A_\xi)$ whenever $E \in \Sigma$ and $\xi \le \zeta$. **P?** Otherwise, because the ordinal $\zeta + 1$ is well-ordered, there is a least ξ such that $\lambda(E \cap A_\xi) \ne \mu^*(E \cap A_\xi)$. As $A_0 = \emptyset$ we surely have $\lambda(E \cap A_0) = \mu^*(E \cap A_0)$ and $\xi > 0$. Note that if $\eta > \xi$, then $\lambda_\eta(E \cap A_\xi) = 0$; so

$$\lambda(E \cap A_\xi) = \textstyle\sum_{\eta \le \xi} \lambda_\eta(E \cap A_\xi \cap A_\eta) = \sum_{\eta \le \xi} \lambda_\eta(E \cap A_\eta).$$

Now

$$\mu^*(E \cap A_\xi) = \lambda_\xi(E \cap A_\xi) + \sup_{\xi' < \xi} \mu^*(E \cap A_{\xi'})$$

((b) above)

$$= \lambda_\xi(E \cap A_\xi) + \sup_{\xi' < \xi} \sum_{\eta \le \xi'} \lambda_\eta(E \cap A_\eta)$$

(because ξ was the first problematic ordinal)

$$= \lambda_\xi(E \cap A_\xi) + \sup_{\xi' < \xi} \sup_{K \subseteq \xi' + 1 \text{ is finite}} \sum_{\eta \in K} \lambda_\eta(E \cap A_\eta)$$

(see the definition of 'sum' in 112Bd, or 226A below)

$$= \lambda_\xi(E \cap A_\xi) + \sup_{K \subseteq \xi \text{ is finite}} \sum_{\eta \in K} \lambda_\eta(E \cap A_\eta)$$

$$= \sup_{K \subseteq \xi + 1 \text{ is finite}} \sum_{\eta \in K} \lambda_\eta(E \cap A_\eta) = \sum_{\eta \le \xi} \lambda_\eta(E \cap A_\eta) \ne \mu^*(E \cap A_\xi)$$

by the choice of ξ; but this is absurd. **XQ**

In particular,

$$\lambda E = \lambda(E \cap A_\zeta) = \mu^*(E \cap A_\zeta) = \mu E$$

for every $E \in \Sigma$. This completes the proof of the theorem.

214X Basic exercises (a) Let (X, Σ, μ) be a localizable measure space. Show that there is an $E \in \Sigma$ such that the subspace measure μ_E is purely atomic and $\mu_{X \setminus E}$ is atomless.

(b) Let X be a set, θ a regular outer measure on X, and Y a subset of X. Let μ be the measure on X defined by Carathéodory's method from θ, μ_Y the subspace measure on Y, and ν the measure on Y defined by Carathéodory's method from $\theta \restriction \mathcal{P}Y$. Show that if μ_Y is locally determined (in particular, if μ is locally determined and localizable) then $\nu = \mu_Y$.

(c) Let (X, Σ, μ) be a localizable measure space, and Y a subset of X such that the subspace measure μ_Y is semi-finite. Show that μ_Y is localizable.

>(d) Let (X, Σ, μ) be a measure space, and Y a subset of X such that the subspace measure μ_Y is semi-finite. (i) Show that a set $F \subseteq Y$ is an atom for μ_Y iff it is of the form $E \cap Y$ where E an atom for μ. (ii) Show that if μ is atomless or purely atomic, so is μ_Y.

(e) Let (X, Σ, μ) be a localizable measure space, and Y any subset of X. Show that the c.l.d. version of the subspace measure on Y is localizable.

(f) Let (X, Σ, μ) be a measure space with locally determined negligible sets, and Y a subset of X, with its subspace measure μ_Y. Show that μ_Y has locally determined negligible sets.

>(g) Let (X, Σ, μ) be a measure space. Show that (X, Σ, μ) has locally determined negligible sets iff the subspace measure μ_Y is semi-finite for every $Y \subseteq X$.

>(h) Let $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be a family of measure spaces, with direct sum (X,Σ,μ) (214L). Set $X_i'=X_i\times\{i\}\subseteq X$ for each $i\in I$. Show that X_i', with the subspace measure, is isomorphic to (X_i,Σ_i,μ_i). Under what circumstances is $\langle X_i'\rangle_{i\in I}$ a decomposition of X? Show that μ is complete, or strictly localizable, or localizable, or locally determined, or semi-finite, or atomless, or purely atomic iff every μ_i is. Show that a measure space is strictly localizable iff it is isomorphic to a direct sum of totally finite spaces.

>(i) Let $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be a family of measure spaces, and (X,Σ,μ) their direct sum. Show that the completion of (X,Σ,μ) can be identified with the direct sum of the completions of the (X_i,Σ_i,μ_i), and that the c.l.d. version of (X,Σ,μ) can be identified with the direct sum of the c.l.d. versions of the (X_i,Σ_i,μ_i).

(j) Let $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be a family of measure spaces. Show that their direct sum has locally determined negligible sets iff every μ_i has.

(k) Let $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be a family of measure spaces, and (X,Σ,μ) their direct sum. Show that (X,Σ,μ) has the measurable envelope property (213Xl) iff every (X_i,Σ_i,μ_i) has.

(l) Let (X,Σ,μ) be a measure space, Y a subset of X, and $f:X\to[0,\infty]$ a function such that $\int_Y f$ is defined in $[0,\infty]$. Show that $\int_Y f=\overline{\int}f\times\chi Y\,d\mu$.

>(m) Write out a direct proof of 214P in the special case in which $\mathcal{A}=\{A\}$. (*Hint*: for E, $F\in\Sigma$,

$$\lambda((E\cap A)\cup(F\setminus A))=\mu^*(E\cap A)+\sup\{\mu G:G\in\Sigma,\,G\subseteq F\setminus A\}.)$$

>(n) Let (X,Σ,μ) be a measure space and $\mathcal{A}$ a finite family of subsets of X. Show that there is a measure on X, extending μ, which measures every member of $\mathcal{A}$.

214Y Further exercises (a) Let (X,Σ,μ) be a measure space and A a subset of X such that the subspace measure on A is semi-finite. Set $\alpha=\sup\{\mu E:E\in\Sigma,\,E\subseteq A\}$. Show that if $\alpha\le\gamma\le\mu^*A$ then there is a measure λ on X, extending μ, such that $\lambda A=\gamma$.

(b) Let (X,Σ,μ) be a measure space and $\langle A_n\rangle_{n\in\mathbb{Z}}$ a double-ended sequence of subsets of X such that $A_m\subseteq A_n$ whenever $m\le n$ in $\mathbb{Z}$. Show that there is a measure on X, extending μ, which measures every A_n. (*Hint*: use 214P twice.)

(c) Let X be a set and $\mathcal{A}$ a family of subsets of X. Show that the following are equiveridical: (i) for every measure μ on X there is a measure on X extending μ and measuring every member of $\mathcal{A}$; (ii) for every totally finite measure μ on X there is a measure on X extending μ and measuring every member of $\mathcal{A}$. (*Hint*: 213Xa.)

214 Notes and comments I take the first part of the section, down to 214H, slowly and carefully, because while none of the arguments are deep (214Eb is the longest) the patterns formed by the results are not always easy to predict. There is a counter-example to a tempting extension of 214H/214Xb in 216Xb.

The message of the second part of the section (214I-214L) is that subspaces inherit many, but not all, of the properties of a measure space; and in particular there is a difficulty with semi-finiteness, unless we have locally determined negligible sets (214Xg). (I give an example in 216Xa.) Of course 213Hb shows that if we start with a localizable space, we can convert it into a complete locally determined localizable space without doing great violence to the structure of the space, so the difficulty is ordinarily superable.

By far the most important case of 214P is when $\mathcal{A}=\{A\}$ is a singleton, so that the argument simplifies dramatically (214Xm). In §439 of Volume 4 I will return to the problem of extending a measure to a given larger σ-algebra in the absence of any helpful auxiliary structure. That section will mostly offer counter-examples, in particular showing that there is no general theorem extending 214Xn from finite families to countable families, and that the special conditions in 214P and 214Yb are there for good reasons. But in §552 of Volume 5 I will present some positive results dependent on special axioms beyond those of ZFC.

215 σ-finite spaces and the principle of exhaustion

I interpolate a short section to deal with some useful facts which might get lost if buried in one of the longer sections of this chapter. The great majority of the applications of measure theory involve σ-finite spaces, to the point that many authors skim over any others. I myself prefer to signal the importance of such concepts by explicitly stating just which theorems apply only to the restricted class of spaces. But undoubtedly some facts about σ-finite spaces need to be grasped early on. In 215B I give a list of properties characterizing σ-finite spaces. Some of these make better sense in the light of the principle of exhaustion (215A). I take the opportunity to include a fundamental fact about atomless measure spaces (215D).

215A The principle of exhaustion The following is an example of the use of one of the most important methods in measure theory.

Lemma Let (X, Σ, μ) be any measure space and $\mathcal{E} \subseteq \Sigma$ a non-empty set such that $\sup_{n\in\mathbb{N}} \mu F_n$ is finite for every non-decreasing sequence $\langle F_n \rangle_{n\in\mathbb{N}}$ in $\mathcal{E}$.

(a) There is a non-decreasing sequence $\langle F_n \rangle_{n\in\mathbb{N}}$ in $\mathcal{E}$ such that, for every $E \in \Sigma$, *either* there is an $n \in \mathbb{N}$ such that $E \cup F_n$ is not included in any member of $\mathcal{E}$ *or*, setting $F = \bigcup_{n\in\mathbb{N}} F_n$,

$$\lim_{n\to\infty} \mu(E \setminus F_n) = \mu(E \setminus F) = 0.$$

In particular, if $E \in \mathcal{E}$ and $E \supseteq F$, then $E \setminus F$ is negligible.

(b) If $\mathcal{E}$ is upwards-directed, then there is a non-decreasing sequence $\langle F_n \rangle_{n\in\mathbb{N}}$ in $\mathcal{E}$ such that, setting $F = \bigcup_{n\in\mathbb{N}} F_n$, $\mu F = \sup_{E\in\mathcal{E}} \mu E$ and $E \setminus F$ is negligible for every $E \in \mathcal{E}$, so that F is an essential supremum of $\mathcal{E}$ in Σ in the sense of 211G.

(c) If the union of any non-decreasing sequence in $\mathcal{E}$ belongs to $\mathcal{E}$, then there is an $F \in \mathcal{E}$ such that $E \setminus F$ is negligible whenever $E \in \mathcal{E}$ and $F \subseteq E$.

proof (a) Choose $\langle F_n \rangle_{n\in\mathbb{N}}$, $\langle \mathcal{E}_n \rangle_{n\in\mathbb{N}}$ and $\langle u_n \rangle_{n\in\mathbb{N}}$ inductively, as follows. Take F_0 to be any member of $\mathcal{E}$. Given $F_n \in \mathcal{E}$, set $\mathcal{E}_n = \{E : F_n \subseteq E \in \mathcal{E}\}$ and $u_n = \sup\{\mu E : E \in \mathcal{E}_n\}$ in $[0, \infty]$, and choose $F_{n+1} \in \mathcal{E}_n$ such that $\mu F_{n+1} \geq \min(n, u_n - 2^{-n})$; continue.

Observe that this construction yields a non-decreasing sequence $\langle F_n \rangle_{n\in\mathbb{N}}$ in $\mathcal{E}$. Since $\mathcal{E}_{n+1} \subseteq \mathcal{E}_n$ for every n, $\langle u_n \rangle_{n\in\mathbb{N}}$ is non-increasing, and has a limit u in $[0, \infty]$. Since $\min(n, u - 2^{-n}) \leq \mu F_{n+1} \leq u_n$ for every n, $\lim_{n\to\infty} \mu F_n = u$. Our hypothesis on $\mathcal{E}$ now tells us that u is finite.

If $E \in \Sigma$ is such that for every $n \in \mathbb{N}$ there is an $E_n \in \mathcal{E}$ such that $E \cup F_n \subseteq E_n$, then $E_n \in \mathcal{E}_n$, so

$$\mu F_n \leq \mu(E \cup F_n) \leq \mu E_n \leq u_n$$

for every n, and $\lim_{n\to\infty} \mu(E \cup F_n) = u$. But this means that

$$\mu(E \setminus F) \leq \lim_{n\to\infty} \mu(E \setminus F_n) = \lim_{n\to\infty} \mu(E \cup F_n) - \mu F_n = 0,$$

as stated. In particular, this is so if $E \in \mathcal{E}$ and $E \supseteq F$.

(b) Take $\langle F_n \rangle_{n\in\mathbb{N}}$ from (a). If $E \in \mathcal{E}$, then (because $\mathcal{E}$ is upwards-directed) $E \cup F_n$ is included in some member of $\mathcal{E}$ for every $n \in \mathbb{N}$; so we must have the second alternative of (a), and $E \setminus F$ is negligible. It follows that

$$\sup_{E\in\mathcal{E}} \mu E \leq \mu F = \lim_{n\to\infty} \mu F_n \leq \sup_{E\in\mathcal{E}} \mu E,$$

so $\mu F = \sup_{E\in\mathcal{E}} \mu E$.

If G is any measurable set such that $E \setminus F$ is negligible for every $E \in \mathcal{E}$, then $F_n \setminus G$ is negligible for every n, so that $F \setminus G$ is negligible; thus F is an essential supremum for $\mathcal{E}$.

(c) Again take $\langle F_n \rangle_{n\in\mathbb{N}}$ from (a), and set $F = \bigcup_{n\in\mathbb{N}} E_n$. Our hypothesis now is that $F \in \mathcal{E}$, so has both the properties declared.

215B σ-finite spaces are so important that I think it is worth spelling out the following facts.

Proposition Let (X, Σ, μ) be a semi-finite measure space. Write $\mathcal{N}$ for the family of μ-negligible sets and Σ^f for the family of measurable sets of finite measure. Then the following are equiveridical:

(i) (X, Σ, μ) is σ-finite;

(ii) every disjoint family in $\Sigma^f \setminus \mathcal{N}$ is countable;

(iii) every disjoint family in $\Sigma \setminus \mathcal{N}$ is countable;

(iv) for every $\mathcal{E} \subseteq \Sigma$ there is a countable set $\mathcal{E}_0 \subseteq \mathcal{E}$ such that $E \setminus \bigcup \mathcal{E}_0$ is negligible for every $E \in \mathcal{E}$;

(v) for every non-empty upwards-directed $\mathcal{E} \subseteq \Sigma$ there is a non-decreasing sequence $\langle F_n \rangle_{n\in\mathbb{N}}$ in $\mathcal{E}$ such that $E \setminus \bigcup_{n\in\mathbb{N}} F_n$ is negligible for every $E \in \mathcal{E}$;

(vi) for every non-empty $\mathcal{E} \subseteq \Sigma$, there is a non-decreasing sequence $\langle F_n \rangle_{n\in\mathbb{N}}$ in $\mathcal{E}$ such that $E \setminus \bigcup_{n\in\mathbb{N}} F_n$ is negligible whenever $E \in \mathcal{E}$ and $E \supseteq F_n$ for every $n \in \mathbb{N}$;

(vii) *either* $\mu X = 0$ *or* there is a probability measure ν on X with the same domain and the same negligible sets as μ;

(viii) there is a measurable integrable function $f : X \to]0, 1]$;

(ix) either $\mu X = 0$ or there is a measurable function $f : X \to]0, \infty[$ such that $\int f d\mu = 1$.

proof (i)⇒(vii) and (viii) If $\mu X = 0$, (vii) is trivial and we can take $f = \chi X$ in (viii). Otherwise, let $\langle E_n \rangle_{n\in\mathbb{N}}$ be a disjoint sequence in Σ^f covering X. Then it is easy to see that there is a sequence $\langle \alpha_n \rangle_{n\in\mathbb{N}}$ of strictly positive real numbers such that $\sum_{n=0}^\infty \alpha_n \mu E_n = 1$. Set $\nu E = \sum_{n=0}^\infty \alpha_n \mu(E \cap E_n)$ for $E \in \Sigma$; then ν is a probability measure with domain Σ and the same negligible sets as μ. Also $f = \sum_{n=0}^\infty \min(1, \alpha_n)\chi E_n$ is a strictly positive measurable integrable function.

(vii)⇒(vi) and (v) Assume (vii), and let $\mathcal{E}$ be a non-empty family of measurable sets. If $\mu X = 0$ then (vi) and (v) are certainly true. Otherwise, let ν be a probability measure with domain Σ and the same negligible sets as μ. Since $\sup_{E\in\mathcal{E}} \nu E \le 1$ is finite, we can apply 215Aa to find a non-decreasing sequence $\langle F_n \rangle_{n\in\mathbb{N}}$ in $\mathcal{E}$ such that $E \setminus \bigcup_{n\in\mathbb{N}} F_n$ is negligible whenever $E \in \mathcal{E}$ includes $\bigcup_{n\in\mathbb{N}} F_n$; and if $\mathcal{E}$ is upwards-directed, $E \setminus \bigcup_{n\in\mathbb{N}} F_n$ will be negligible for every $E \in \mathcal{E}$, as in 215Ab.

(vi)⇒(iv) Assume (vi), and let $\mathcal{E}$ be any subset of Σ. Set

$$\mathcal{H} = \{\bigcup \mathcal{E}_0 : \mathcal{E}_0 \subseteq \mathcal{E} \text{ is countable}\}.$$

By (vi), there is a sequence $\langle H_n \rangle_{n\in\mathbb{N}}$ in $\mathcal{H}$ such that $H \setminus \bigcup_{n\in\mathbb{N}} H_n$ is negligible whenever $H \in \mathcal{H}$ and $H \supseteq H_n$ for every $n \in \mathbb{N}$. Now we can express each H_n as $\bigcup \mathcal{E}'_n$, where $\mathcal{E}'_n \subseteq \mathcal{E}$ is countable; setting $\mathcal{E}_0 = \bigcup_{n\in\mathbb{N}} \mathcal{E}'_n$, $\mathcal{E}_0$ is countable. If $E \in \mathcal{E}$, then $E \cup \bigcup_{n\in\mathbb{N}} H_n = \bigcup(\{E\} \cup \mathcal{E}_0)$ belongs to $\mathcal{H}$ and includes every H_n, so that $E \setminus \bigcup \mathcal{E}_0 = E \setminus \bigcup_{n\in\mathbb{N}} H_n$ is negligible. So $\mathcal{E}_0$ has the property we need, and (iv) is true.

(iv)⇒(iii) Assume (iv). If $\mathcal{E}$ is a disjoint family in $\Sigma \setminus \mathcal{N}$, take a countable $\mathcal{E}_0 \subseteq \mathcal{E}$ such that $E \setminus \bigcup \mathcal{E}_0$ is negligible for every $E \in \mathcal{E}$. Then $E = E \setminus \bigcup \mathcal{E}_0$ is negligible for every $E \in \mathcal{E} \setminus \mathcal{E}_0$; but this just means that $\mathcal{E} \setminus \mathcal{E}_0$ is empty, so that $\mathcal{E} = \mathcal{E}_0$ is countable.

(iii)⇒(ii) is trivial.

(ii)⇒(i) Assume (ii). Let $\mathfrak{P}$ be the set of all disjoint subsets of $\Sigma^f \setminus \mathcal{N}$, ordered by $\subseteq$. Then $\mathfrak{P}$ is a partially ordered set, not empty (as $\emptyset \in \mathfrak{P}$), and if $\mathfrak{Q} \subseteq \mathfrak{P}$ is non-empty and totally ordered then it has an upper bound in $\mathfrak{P}$. **P** Set $\mathcal{E} = \bigcup \mathfrak{Q}$, the union of all the disjoint families belonging to $\mathfrak{Q}$. If $E \in \mathcal{E}$ then $E \in \mathcal{C}$ for some $\mathcal{C} \in \mathfrak{Q}$, so $E \in \Sigma^f \setminus \mathcal{N}$. If E, $F \in \mathcal{E}$ and $E \ne F$, then there are $\mathcal{C}$, $\mathcal{D} \in \mathfrak{Q}$ such that $E \in \mathcal{C}$, $F \in \mathcal{D}$; now $\mathfrak{Q}$ is totally ordered, so one of $\mathcal{C}$, $\mathcal{D}$ is larger than the other, and in either case $\mathcal{C} \cup \mathcal{D}$ is a member of $\mathfrak{Q}$ containing both E and F. But since any member of $\mathfrak{Q}$ is a disjoint collection of sets, $E \cap F = \emptyset$. As E and F are arbitrary, $\mathcal{E}$ is a disjoint family of sets and belongs to $\mathfrak{P}$. And of course $\mathcal{C} \subseteq \mathcal{E}$ for every $\mathcal{C} \in \mathfrak{Q}$, so $\mathcal{E}$ is an upper bound for $\mathfrak{Q}$ in $\mathfrak{P}$. **Q**

By Zorn's Lemma (2A1M), $\mathfrak{P}$ has a maximal element $\mathcal{E}$ say. By (ii), $\mathcal{E}$ must be countable, so $\bigcup \mathcal{E} \in \Sigma$. Now $H = X \setminus \bigcup \mathcal{E}$ is negligible. **P?** Suppose, if possible, otherwise. Because (X, Σ, μ) is semi-finite, there is a set G of finite measure such that $G \subseteq H$ and $\mu G > 0$, that is, $G \in \Sigma^f \setminus \mathcal{N}$ and $G \cap E = \emptyset$ for every $E \in \mathcal{E}$. But this means that $\{G\} \cup \mathcal{E}$ is a member of $\mathfrak{P}$ strictly larger than $\mathcal{E}$, which is supposed to be impossible. **XQ**

Let $\langle X_n \rangle_{n\in\mathbb{N}}$ be a sequence running over $\mathcal{E} \cup \{H\}$. Then $\langle X_n \rangle_{n\in\mathbb{N}}$ is a cover of X by a sequence of measurable sets of finite measure, so (X, Σ, μ) is σ-finite.

(v)⇒(i) If (v) is true, then we have a sequence $\langle E_n \rangle_{n\in\mathbb{N}}$ in Σ^f such that $E \setminus \bigcup_{n\in\mathbb{N}} E_n$ is negligible for every $E \in \Sigma^f$. Because μ is semi-finite, $X \setminus \bigcup_{n\in\mathbb{N}} E_n$ must be negligible, so X is covered by a countable family of sets of finite measure and μ is σ-finite.

(viii)⇒(ix) If $\mu X = 0$ this is trivial. Otherwise, if f is a strictly positive measurable integrable function, then $c = \int f > 0$ (122Rc), so $\frac{1}{c} f$ is a strictly positive measurable function with integral 1.

(ix)⇒(i) If $f : X \to]0, \infty[$ is measurable and integrable, $\langle \{x : f(x) \ge 2^{-n}\} \rangle_{n\in\mathbb{N}}$ is a sequence of sets of finite measure covering X.

215C Corollary Let (X, Σ, μ) be a σ-finite measure space, and suppose that $\mathcal{E} \subseteq \Sigma$ is any non-empty set.

(a) There is a non-decreasing sequence $\langle F_n \rangle_{n\in\mathbb{N}}$ in $\mathcal{E}$ such that, for every $E \in \Sigma$, *either* there is an $n \in \mathbb{N}$ such that $E \cup F_n$ is not included in any member of $\mathcal{E}$ *or* $E \setminus \bigcup_{n\in\mathbb{N}} F_n$ is negligible.

(b) If $\mathcal{E}$ is upwards-directed, then there is a non-decreasing sequence $\langle F_n \rangle_{n\in\mathbb{N}}$ in $\mathcal{E}$ such that $\bigcup_{n\in\mathbb{N}} F_n$ is an essential supremum of $\mathcal{E}$ in Σ.

(c) If the union of any non-decreasing sequence in $\mathcal{E}$ belongs to $\mathcal{E}$, then there is an $F \in \mathcal{E}$ such that $E \setminus F$ is negligible whenever $E \in \mathcal{E}$ and $F \subseteq E$.

proof By 215B, there is a totally finite measure ν on X with the same measurable sets and the same negligible sets as μ. Since $\sup_{E\in\mathcal{E}} \nu E$ is finite, we can apply 215A to ν to obtain the results.

215D As a further example of the use of the principle of exhaustion, I give a fundamental fact about atomless measure spaces.

Proposition Let (X, Σ, μ) be an atomless measure space. If $E \in \Sigma$ and $0 \le \alpha \le \mu E < \infty$, there is an $F \in \Sigma$ such that $F \subseteq E$ and $\mu F = \alpha$.

proof (a) We need to know that if $G \in \Sigma$ is non-negligible and $n \in \mathbb{N}$, then there is an $H \subseteq G$ such that $0 < \mu H \le 2^{-n}\mu G$. **P** Induce on n. For $n = 0$ this is trivial. For the inductive step to $n + 1$, use the inductive hypothesis to find $H \subseteq G$ such that $0 < \mu H \le 2^{-n}\mu G$. Because μ is atomless, there is an $H' \subseteq H$ such that $\mu H'$, $\mu(H \setminus H')$ are both defined and non-zero. Now at least one of them has measure less than or equal to $\frac{1}{2}\mu H$, so gives us a subset of G of non-zero measure less than or equal to $2^{-n-1}\mu G$. **Q**

It follows that if $G \in \Sigma$ has non-zero finite measure and $\epsilon > 0$, there is a measurable set $H \subseteq G$ such that $0 < \mu H \le \epsilon$.

(b) Let $\mathcal{H}$ be the family of all those $H \in \Sigma$ such that $H \subseteq E$ and $\mu H \le \alpha$. If $\langle H_n \rangle_{n\in\mathbb{N}}$ is any non-decreasing sequence in $\mathcal{H}$, then $\mu(\bigcup_{n\in\mathbb{N}} H_n) = \lim_{n\to\infty} \mu H_n \le \alpha$, so $\bigcup_{n\in\mathbb{N}} H_n \in \mathcal{H}$. So 215Ac tells us that there is an $F \in \mathcal{H}$ such that $H \setminus F$ is negligible whenever $H \in \mathcal{H}$ and $F \subseteq H$. **?** Suppose, if possible, that $\mu F < \alpha$. By (a), there is an $H \subseteq E \setminus F$ such that $0 < \mu H \le \alpha - \mu F$. But in this case $H \cup F \in \mathcal{H}$ and $\mu((H \cup F) \setminus F) > 0$, which is impossible. **X**

So we have found an appropriate set F.

215X Basic exercises (a) Let (X, Σ, μ) be a measure space and Φ a non-empty set of μ-integrable real-valued functions from X to $\mathbb{R}$. Suppose that $\sup_{n\in\mathbb{N}} \int f_n$ is finite for every sequence $\langle f_n \rangle_{n\in\mathbb{N}}$ in Φ such that $f_n \le_{\text{a.e.}} f_{n+1}$ for every n. Show that there is a sequence $\langle f_n \rangle_{n\in\mathbb{N}}$ in Φ such that $f_n \le_{\text{a.e.}} f_{n+1}$ for every n and, for every integrable real-valued function f on X, *either* $f \le_{\text{a.e.}} \sup_{n\in\mathbb{N}} f_n$ *or* there is an $n \in \mathbb{N}$ such that no member of Φ is greater than or equal to $\max(f, f_n)$ almost everywhere.

>(b) Let (X, Σ, μ) be a measure space. (i) Suppose that $\mathcal{E}$ is a non-empty upwards-directed subset of Σ such that $c = \sup_{E\in\mathcal{E}} \mu E$ is finite. Show that $E \setminus \bigcup_{n\in\mathbb{N}} F_n$ is negligible whenever $E \in \mathcal{E}$ and $\langle F_n \rangle_{n\in\mathbb{N}}$ is a sequence in $\mathcal{E}$ such that $\lim_{n\to\infty} \mu F_n = c$. (ii) Let Φ be a non-empty set of integrable functions on X which is upwards-directed in the sense that for all f, $g \in \Phi$ there is an $h \in \Phi$ such that $\max(f, g) \le_{\text{a.e.}} h$, and suppose that $c = \sup_{f\in\Phi} \int f$ is finite. Show that $f \le_{\text{a.e.}} \sup_{n\in\mathbb{N}} f_n$ whenever $f \in \Phi$ and $\langle f_n \rangle_{n\in\mathbb{N}}$ is a sequence in Φ such that $\lim_{n\to\infty} \int f_n = c$.

(c) Use 215A to shorten the proof of 211Ld.

(d) Give an example of a (non-semi-finite) measure space (X, Σ, μ) satisfying conditions (ii)-(iv) of 215B, but not (i).

>(e) Let (X, Σ, μ) be an atomless σ-finite measure space. Show that for any $\epsilon > 0$ there is a disjoint sequence $\langle E_n \rangle_{n\in\mathbb{N}}$ of measurable sets with measure at most ϵ such that $X = \bigcup_{n\in\mathbb{N}} E_n$.

(f) Let (X, Σ, μ) be an atomless strictly localizable measure space. Show that for any $\epsilon > 0$ there is a decomposition $\langle X_i \rangle_{i\in I}$ of X such that $\mu X_i \le \epsilon$ for every $i \in I$.

215Y Further exercises (a) Let (X, Σ, μ) be a σ-finite measure space and $\langle f_{mn} \rangle_{m,n\in\mathbb{N}}$, $\langle f_m \rangle_{m\in\mathbb{N}}$, f measurable real-valued functions defined almost everywhere in X and such that $\langle f_{mn} \rangle_{n\in\mathbb{N}} \to f_m$ a.e. for each m and $\langle f_m \rangle_{m\in\mathbb{N}} \to f$ a.e. Show that there is a strictly increasing sequence $\langle n_m \rangle_{m\in\mathbb{N}}$ in $\mathbb{N}$ such that $\langle f_{m,n_m} \rangle_{m\in\mathbb{N}} \to f$ a.e. (Compare 134Yb.)

(b) Let (X, Σ, μ) be a σ-finite measure space. Let $\langle f_n \rangle_{n\in\mathbb{N}}$ be a sequence of measurable real-valued functions such that $f = \lim_{n\to\infty} f_n$ is defined almost everywhere in X. Show that there is a non-decreasing sequence $\langle X_k \rangle_{k\in\mathbb{N}}$ of measurable subsets of X such that $\bigcup_{k\in\mathbb{N}} X_k$ is conegligible in X and $\langle f_n \rangle_{n\in\mathbb{N}} \to f$ uniformly on every X_k, in the sense that for any $\epsilon > 0$ there is an $m \in \mathbb{N}$ such that $|f_j(x) - f(x)|$ is defined and less than or equal to ϵ whenever $j \ge m$, $x \in X_k$.

(This is a version of Egorov's theorem.)

(c) Let (X,Σ,μ) be a totally finite measure space and $\langle f_n\rangle_{n\in\mathbb{N}}$, f measurable real-valued functions defined almost everywhere in X. Show that $\langle f_n\rangle_{n\in\mathbb{N}}\to f$ a.e. iff there is a sequence $\langle\epsilon_n\rangle_{n\in\mathbb{N}}$ of strictly positive real numbers, converging to 0, such that

$$\lim_{n\to\infty}\mu^*(\bigcup_{k\ge n}\{x:x\in\operatorname{dom}f_k\cap\operatorname{dom}f,\ |f_k(x)-f(x)|\ge\epsilon_n\})=0.$$

(d) Find a direct proof of (v)$\Rightarrow$(vi) in 215B. (*Hint*: given $\mathcal{E}\subseteq\Sigma$, use Zorn's Lemma to find a maximal totally ordered $\mathcal{E}'\subseteq\mathcal{E}$ such that $E\triangle F\notin\mathcal{N}$ for any distinct E, $F\in\mathcal{E}'$, and apply (v) to $\mathcal{E}'$.)

215 Notes and comments The common ground of 215A, 215B(vi), 215C and 215Xa is actually one of the most fundamental ideas in measure theory. It appears in such various forms that it is often easier to prove an application from first principles than to explain how it can be reduced to the versions here. But I will try henceforth to signal such applications as they arise, calling the method (the proof of 215Aa or 215Xa) the 'principle of exhaustion'. One point which is perhaps worth noting here is the inductive construction of the sequence $\langle F_n\rangle_{n\in\mathbb{N}}$ in the proof of 215Aa. Each F_{n+1} is chosen *after* the preceding one. It is this which makes it possible, in the proof of 215B(vii)$\Rightarrow$(vi), to extract a suitable sequence $\langle F_n\rangle_{n\in\mathbb{N}}$ directly. In many applications (starting with what is surely the most important one in the elementary theory, the Radon-Nikodým theorem of §232, or with part (i) of the proof of 211Ld), this refinement is not needed; we are dealing with an upwards-directed set, as in 215B(v), and can choose the whole sequence $\langle F_n\rangle_{n\in\mathbb{N}}$ at once, no term interacting with any other, as in 215Xb. The axiom of 'dependent choice', which asserts that we can construct sequences term-by-term, is known to be stronger than the axiom of 'countable choice', which asserts only that we can choose countably many objects simultaneously.

In 215B I try to indicate the most characteristic properties of σ-finiteness; in particular, the properties which distinguish σ-finite measures from other strictly localizable measures. This result is in a way more abstract than the manipulations in the rest of the section. Note that it makes an essential use of the axiom of choice in the form of Zorn's Lemma. I spent a paragraph in 134C commenting on the distinction between 'countable choice', which is needed for anything which looks like the standard theory of Lebesgue measure, and the full axiom of choice, which is relatively little used in the elementary theory. The implication (ii)$\Rightarrow$(i) of 215B is one of the points where we do need something beyond countable choice. (I should perhaps remark that the whole theory of non-σ-finite measure spaces looks very odd without the general axiom of choice.) Note also that in 215B the proofs of (i)$\Rightarrow$(vii) and (vii)$\Rightarrow$(vi) are the only points where anything so vulgar as a number appears. The conditions (iii), (iv), (v) and (vi) are linked in ways that have nothing to do with measure theory, and involve only with the structure $(X,\Sigma,\mathcal{N})$. (See 215Yd here, and 316D-316E in Volume 3.) There are similar conditions relating to measurable functions rather than measurable sets; for a fairly abstract example, see 241Ye.

In 215Ya-215Yc are three more standard theorems on almost-everywhere-convergent sequences which depend on σ- or total finiteness.

216 Examples

It is common practice – and, in my view, good practice – in books on pure mathematics, to provide discriminating examples; I mean that whenever we are given a list of new concepts, we expect to be provided with examples to show that we have a fair picture of the relationships between them, and in particular that we are not being kept ignorant of some startling implication. Concerning the concepts listed in 211A-211K, we have ten different properties which some, but not all, measure spaces possess, giving a conceivable total of 2^{10} different types of measure space, classified according to which of these ten properties they have. The list of basic relationships in 211L reduces these 1024 possibilities to 72. Observing that a space can be simultaneously atomless and purely atomic only when the measure of the whole space is 0, we find ourselves with 56 possibilities, being two trivial cases with $\mu X=0$ (because such a measure may or may not be complete) together with $9\times2\times3$ cases, corresponding to the nine classes

probability spaces,
spaces which are totally finite, but not probability spaces,
spaces which are σ-finite, but not totally finite,
spaces which are strictly localizable, but not σ-finite,
spaces which are localizable and locally determined, but not strictly localizable,
spaces which are localizable, but not locally determined,
spaces which are locally determined, but not localizable,
spaces which are semi-finite, but neither locally determined nor localizable,
spaces which are not semi-finite;

the two classes

spaces which are complete,

spaces which are not complete;

and the three classes

spaces which are atomless, not of measure 0,

spaces which are purely atomic, not of measure 0,

spaces which are neither atomless nor purely atomic.

I do not propose to give a complete set of fifty-six examples, particularly as rather fewer than fifty-six different ideas are required. However, I do think that for a proper understanding of abstract measure spaces it is necessary to have seen realizations of some of the critical combinations of properties. I therefore take a few paragraphs to describe three special examples to add to those of 211M-211R.

216A Lebesgue measure Before turning to the new ideas, let me mention Lebesgue measure again. As remarked in 211M, 211P and 211Qa,

(a) Lebesgue measure μ on $\mathbb{R}$ is complete, atomless and σ-finite, therefore strictly localizable, localizable and locally determined.

(b) The subspace measure $\mu_{[0,1]}$ on $[0,1]$ is a complete, atomless probability measure.

(c) The restriction $\mu\restriction\mathcal{B}$ of μ to the Borel σ-algebra $\mathcal{B}$ of $\mathbb{R}$ is atomless, σ-finite and not complete.

216B I now embark on the description of three 'counter-examples'; meaning spaces built specifically for the purpose of showing that there are no unexpected implications among the ten properties under consideration here. Even by the standards of this chapter these must be regarded as dispensable by the student who wants to get on with the real business of understanding the big theorems of the subject. Neither the existence of these examples, nor the techniques needed in constructing them, are vital for anything else we shall look at before Volume 5. But if you are going to take abstract measure theory seriously at all, sooner or later you will need to form some kind of mental picture of the nature of the spaces possessing the different properties here, and a minimal requirement of such a picture is that it should include the discriminations witnessed by these examples.

***216C A complete, localizable, non-locally-determined space** The first example hardly needs an idea beyond what we already have, but it does call for more manipulations than it seems fair to set as an exercise, and may therefore be useful as a demonstration of technique.

(a) Let I be any uncountable set, and set $X=\{0,1\}\times I$. For $E\subseteq X$, $y\in\{0,1\}$ set $E[\{y\}]=\{i:(y,i)\in E\}\subseteq I$. Set

$$\Sigma=\{E:E\subseteq X,\ E[\{0\}]\triangle E[\{1\}]\text{ is countable}\}.$$

Then Σ is a σ-algebra of subsets of X. **P** (i) $\emptyset[\{0\}]\triangle\emptyset[\{1\}]=\emptyset$ is countable, so $\emptyset\in\Sigma$. (ii) If $E\in\Sigma$ then

$$(X\setminus E)[\{0\}]\triangle(X\setminus E)[\{1\}]=E[\{0\}]\triangle E[\{1\}]$$

is countable. (iii) If $\langle E_n\rangle_{n\in\mathbb{N}}$ is a sequence in Σ and $E=\bigcup_{n\in\mathbb{N}}E_n$, then

$$E[\{0\}]\triangle E[\{1\}]\subseteq\bigcup_{n\in\mathbb{N}}E_n[\{0\}]\triangle E_n[\{1\}]$$

is countable. **Q**

For $E\in\Sigma$, set $\mu E=\#(E[\{0\}])$ if this is finite, ∞ otherwise; then (X,Σ,μ) is a measure space.

(b) (X,Σ,μ) is complete. **P** If $A\subseteq E\in\Sigma$ and $\mu E=0$, then $(0,i)\notin E$ for every i. So

$$A[\{0\}]\triangle A[\{1\}]=A[\{1\}]\subseteq E[\{1\}]=E[\{1\}]\triangle E[\{0\}]$$

must be countable, and $A\in\Sigma$. **Q**

(c) (X,Σ,μ) is semi-finite. **P** If $E\in\Sigma$ and $\mu E>0$, there is an $i\in I$ such that $(0,i)\in E$; now $F=\{(0,i)\}\subseteq E$ and $\mu F=1$. **Q**

(d) (X,Σ,μ) is localizable. **P** Let $\mathcal{E}$ be any subset of Σ. Set

$$J=\bigcup_{E\in\mathcal{E}}E[\{0\}],\quad G=\{0,1\}\times J.$$

Then $G\in\Sigma$. If $H\in\Sigma$, then

$$\begin{aligned}\mu(E \setminus H) = 0 \text{ for every } E \in \mathcal{E} \\ &\iff E[\{0\}] \subseteq H[\{0\}] \text{ for every } E \in \mathcal{E} \\ &\iff (0,i) \in H \text{ for every } i \in J \\ &\iff \mu(G \setminus H) = 0.\end{aligned}$$

Thus G is an essential supremum for $\mathcal{E}$ in Σ; as $\mathcal{E}$ is arbitrary, μ is localizable. **Q**

(e) (X, Σ, μ) is not locally determined. **P** Consider $H = \{0\} \times I$. Then $H \notin \Sigma$ because $H[\{0\}] \triangle H[\{1\}] = I$ is uncountable. But let $E \in \Sigma$ be any set such that $\mu E < \infty$. Then

$$(E \cap H)[\{0\}] \triangle (E \cap H)[\{1\}] = (E \cap H)[\{0\}] \subseteq E[\{0\}]$$

is finite, so $E \cap H \in \Sigma$. As E is arbitrary, H witnesses that μ is not locally determined. **Q**

(f) (X, Σ, μ) is purely atomic. **P** Let $E \in \Sigma$ be any set of non-zero measure. Let $i \in I$ be such that $(0,i) \in E$. Then $(0,i) \in E$ and $F = \{(0,i)\}$ is a set of measure 1, included in E; because F is a singleton set, it must be an atom for μ; as E is arbitrary, μ is purely atomic. **Q**

(g) Thus the construction here yields a complete, localizable, purely atomic, non-locally-determined space.

***216D A complete, locally determined space which is not localizable** The next construction requires a little set theory. We need two sets I, J such that I is uncountable (more strictly, I cannot be expressed as the union of countably many countable sets), $I \subseteq J$ and J cannot be expressed as $\bigcup_{i \in I} K_i$ where every K_i is countable. The most natural way of doing this, subject to the axiom of choice, is to take $I = \omega_1$, the first uncountable ordinal, and J to be ω_2, the first ordinal from which there is no injection into ω_1 (see 2A1Fc); but in case you prefer other formulations (e.g., $I = \{\{x\} : x \in \mathbb{R}\}$ and $J = \mathcal{P}\mathbb{R}$), I will write the following argument in terms of I and J, and you can pick your own pair.

(a) Let T be the countable-cocountable σ-algebra of J and ν the countable-cocountable measure on J (211R). Set $X = J \times J$ and for $E \subseteq X$ set

$$E[\{\xi\}] = \{\eta : (\xi, \eta) \in E\}, \quad E^{-1}[\{\xi\}] = \{\eta : (\eta, \xi) \in E\}$$

for every $\xi \in J$. Set

$$\Sigma = \{E : E[\{\xi\}] \text{ and } E^{-1}[\{\xi\}] \text{ belong to } \mathrm{T} \text{ for every } \xi \in J\},$$

$$\mu E = \sum_{\xi \in J} \nu E[\{\xi\}] + \sum_{\xi \in J} \nu E^{-1}[\{\xi\}]$$

for every $E \in \Sigma$. It is easy to check that Σ is a σ-algebra and that μ is a measure.

(b) (X, Σ, μ) is complete. **P** If $A \subseteq E \in \Sigma$ and $\mu E = 0$, then all the sets $E[\{\xi\}]$ and $E^{-1}[\{\xi\}]$ are countable, so the same is true of all the sets $A[\{\xi\}]$ and $A^{-1}[\{\xi\}]$, and $A \in \Sigma$. **Q**

(d) (X, Σ, μ) is semi-finite. **P** For each $\zeta \in J$, set

$$G_\zeta = \{\zeta\} \times J, \quad \tilde{G}_\zeta = J \times \{\zeta\}.$$

Then all the sections $G_\zeta[\{\xi\}]$, $G_\zeta^{-1}[\{\xi\}]$, $\tilde{G}_\zeta[\{\xi\}]$ and $\tilde{G}_\zeta^{-1}[\{\xi\}]$ are either J or $\emptyset$ or $\{\zeta\}$, so belong to T, and all the G_ζ, $\tilde{G}_\zeta$ belong to Σ, with μ-measure 1.

Suppose that $E \in \Sigma$ is a set of strictly positive measure. Then there must be some $\xi \in J$ such that

$$0 < \nu E[\{\xi\}] + \nu E^{-1}[\{\xi\}] = \mu(E \cap G_\xi) + \mu(E \cap \tilde{G}_\xi) < \infty,$$

and one of the sets $E \cap G_\xi$, $E \cap \tilde{G}_\xi$ is a set of non-zero finite measure included in E. **Q**

(e) (X, Σ, μ) is locally determined. **P** Suppose that $H \subseteq X$ is such that $H \cap E \in \Sigma$ whenever $E \in \Sigma$ and $\mu E < \infty$. Then, in particular, $H \cap G_\zeta$ and $H \cap \tilde{G}_\zeta$ belong to Σ, so

$$H[\{\zeta\}] = (H \cap \tilde{G}_\zeta)[\{\zeta\}] \in \mathrm{T},$$

$$H^{-1}[\{\zeta\}] = (H \cap G_\zeta)^{-1}[\{\zeta\}] \in \mathrm{T},$$

for every $\zeta \in J$. This shows that $H \in \Sigma$. As H is arbitrary, μ is locally determined. **Q**

(f) (X, Σ, μ) is not localizable. **P** Set $\mathcal{E} = \{G_\zeta : \zeta \in J\}$. **?** Suppose, if possible, that $G \in \Sigma$ is an essential supremum for $\mathcal{E}$. Then

$$\nu(J\setminus G[\{\xi\}])=\mu(G_\xi\setminus G)=0$$

and $J\setminus G[\{\xi\}]$ is countable, for every $\xi\in J$. Consequently $J\ne\bigcup_{\xi\in I}(J\setminus G[\{\xi\}])$, and there is an η belonging to $J\setminus\bigcup_{\xi\in I}(J\setminus G[\{\xi\}])=\bigcap_{\xi\in I}G[\{\xi\}]$. This means just that $(\xi,\eta)\in G$ for every $\xi\in I$, that is, that $I\subseteq G^{-1}[\{\eta\}]$. Accordingly $G^{-1}[\{\eta\}]$ is uncountable, so that $\nu G^{-1}[\{\eta\}]=\mu(G\cap\tilde G_\eta)=1$. But observe that $\mu(G_\xi\cap\tilde G_\eta)=\mu\{(\xi,\eta)\}=0$ for every $\xi\in J$. This means that, setting $H=X\setminus\tilde G_\eta$, $E\setminus H$ is negligible, for every $E\in\mathcal{E}$; so that we must have $0=\mu(G\setminus H)=\mu(G\cap\tilde G_\eta)=1$, which is absurd. **X**

Thus $\mathcal{E}$ has no essential supremum in Σ, and μ cannot be localizable. **Q**

(g) (X,Σ,μ) is purely atomic. **P** If $E\in\Sigma$ has non-zero measure, there must be some $\xi\in J$ such that one of $E[\{\xi\}]$, $E^{-1}[\{\xi\}]$ is not countable; that is, such that one of $E\cap G_\xi$, $E\cap\tilde G_\xi$ is not negligible. But if now $H\in\Sigma$ and $H\subseteq E\cap G_\xi$, either $H[\{\xi\}]$ is countable, and $\mu H=0$, or $J\setminus H[\{\xi\}]$ is countable, and $\mu(G_\xi\setminus H)=0$; similarly, if $H\subseteq E\cap\tilde G_\xi$, one of μH, $\mu(\tilde G_\xi\setminus H)$ must be 0, according to whether $H^{-1}[\{\xi\}]$ is countable or not. Thus $E\cap G_\xi$ and $E\cap\tilde G_\xi$, if not negligible, must be atoms, and E must include an atom. As E is arbitrary, μ is purely atomic. **Q**

(h) Thus (X,Σ,μ) is complete, locally determined and purely atomic, but is not localizable.

***216E A complete, locally determined, localizable space which is not strictly localizable** For the last, and most interesting, construction, we need a non-trivial result in infinitary combinatorics, which I have written out in 2A1P: if I is any set, and $\langle f_\alpha\rangle_{\alpha\in A}$ is a family in $\{0,1\}^I$, the set of functions from I to $\{0,1\}$, with $\#(A)$ strictly greater than $\mathfrak{c}$, the cardinal of the continuum, and if $\langle K_\alpha\rangle_{\alpha\in A}$ is any family of countable subsets of I, then there must be distinct α, $\beta\in A$ such that f_α and f_β agree on $K_\alpha\cap K_\beta$.

Armed with this fact, I proceed as follows.

(a) Let C be any set of cardinal greater than $\mathfrak{c}$. Set $I=\mathcal{P}C$ and $X=\{0,1\}^I$. For $\gamma\in C$, define $x_\gamma\in X$ by saying that $x_\gamma(\Gamma)=1$ if $\gamma\in\Gamma\subseteq C$ and $x_\gamma(\Gamma)=0$ if $\gamma\notin\Gamma\subseteq C$. Let $\mathcal{K}$ be the family of countable subsets of I, and for $K\in\mathcal{K}$, $\gamma\in C$ set

$$F_{\gamma K}=\{x:x\in X,\,x{\restriction}K=x_\gamma{\restriction}K\}\subseteq X.$$

Let

$$\begin{aligned}\Sigma_\gamma=\{E:E\subseteq X,\ &\text{either there is a }K\in\mathcal{K}\text{ such that }F_{\gamma K}\subseteq E\\ &\text{or there is a }K\in\mathcal{K}\text{ such that }F_{\gamma K}\subseteq X\setminus E\}.\end{aligned}$$

Then Σ_γ is a σ-algebra of subsets of X. **P** (i) $F_{\gamma\emptyset}\subseteq X\setminus\emptyset$ so $\emptyset\in\Sigma_\gamma$. (ii) The definition of Σ_γ is symmetric between E and $X\setminus E$, so $X\setminus E\in\Sigma_\gamma$ whenever $E\in\Sigma_\gamma$. (iii) Let $\langle E_n\rangle_{n\in\mathbb{N}}$ be a sequence in Σ_γ, with union E. (α) If there are $n\in\mathbb{N}$, $K\in\mathcal{K}$ such that $F_{\gamma K}\subseteq E_n$, then $F_{\gamma K}\subseteq E$, so $E\in\Sigma_\gamma$. (β) Otherwise, there is for each $n\in\mathbb{N}$ a $K_n\in\mathcal{K}$ such that $F_{\gamma,K_n}\subseteq X\setminus E_n$. Set $K=\bigcup_{n\in\mathbb{N}}K_n\in\mathcal{K}$. Then

$$\begin{aligned}F_{\gamma K}&=\{x:x{\restriction}K=x_\gamma{\restriction}K\}=\{x:x{\restriction}K_n=x_\gamma{\restriction}K_n\text{ for every }n\in\mathbb{N}\}\\&=\bigcap_{n\in\mathbb{N}}F_{\gamma,K_n}\subseteq\bigcap_{n\in\mathbb{N}}X\setminus E_n=X\setminus E,\end{aligned}$$

so again $E\in\Sigma_\gamma$. As $\langle E_n\rangle_{n\in\mathbb{N}}$ is arbitrary, Σ_γ is a σ-algebra. **Q**

(b) Set

$$\Sigma=\bigcap_{\gamma\in C}\Sigma_\gamma;$$

then Σ, being an intersection of σ-algebras, is a σ-algebra of subsets of X (see 111Ga). Define $\mu:\Sigma\to[0,\infty]$ by setting

$$\begin{aligned}\mu E&=\#(\{\gamma:x_\gamma\in E\})\text{ if this is finite,}\\&=\infty\text{ otherwise;}\end{aligned}$$

then μ is a measure.

(c) It will be convenient later to know something about the sets

$$G_D=\{x:x\in X,\,x(D)=1\}$$

for $D\subseteq C$. In particular, every G_D belongs to Σ. **P** If $\gamma\in D$, then $x_\gamma(D)=1$ so $G_D=F_{\gamma,\{D\}}\in\Sigma_\gamma$. If $\gamma\in C\setminus D$, then $x_\gamma(D)=0$ so $G_D=X\setminus F_{\gamma,\{D\}}\in\Sigma_\gamma$. **Q** Also, of course, $\{\gamma:x_\gamma\in G_D\}=D$.

(d) (X, Σ, μ) is complete. **P** Suppose that $A \subseteq E \subseteq \Sigma$ and that $\mu E = 0$. For every $\gamma \in C$, $E \in \Sigma_\gamma$ and $x_\gamma \notin E$, so $F_{\gamma K} \not\subseteq E$ for any $K \in \mathcal{K}$ and there is a $K \in \mathcal{K}$ such that

$$F_{\gamma K} \subseteq X \setminus E \subseteq X \setminus A.$$

Thus $A \in \Sigma_\gamma$; as γ is arbitrary, $A \in \Sigma$. As A is arbitrary, μ is complete. **Q**

(e) (X, Σ, μ) is semi-finite. **P** Let $E \in \Sigma$ be a set of positive measure. Then there must be some $\gamma \in C$ such that $x_\gamma \in E$. Consider $E' = E \cap G_{\{\gamma\}}$. As $x_\gamma \in E'$, $\mu E' \ge 1 > 0$. On the other hand, $\mu G_{\{\gamma\}} = \#(\{\delta : \delta \in \{\gamma\}\}) = 1$, so $\mu E' = 1$. As E is arbitrary, μ is semi-finite. **Q**

(f) (X, Σ, μ) is localizable. **P** Let $\mathcal{E}$ be any subset of Σ. Set $D = \{\delta : \delta \in C,\ x_\delta \in \bigcup \mathcal{E}\}$. Consider G_D. For $H \in \Sigma$,

$$\begin{aligned}
\mu(E \setminus H) = 0 \text{ for every } E \in \mathcal{E}& \\
&\iff x_\gamma \notin E \setminus H \text{ for every } E \in \mathcal{E}, \gamma \in C \\
&\iff x_\gamma \in H \text{ for every } \gamma \in D \\
&\iff x_\gamma \notin G_D \setminus H \text{ for every } \gamma \in C \\
&\iff \mu(G_D \setminus H) = 0.
\end{aligned}$$

Thus G_D is an essential supremum for $\mathcal{E}$ in Σ. As $\mathcal{E}$ is arbitrary, μ is localizable. **Q**

(g) (X, Σ, μ) is not strictly localizable. **P?** Suppose, if possible, that $\langle X_j \rangle_{j \in J}$ is a decomposition of (X, Σ, μ). Set $J' = \{j : j \in J,\ \mu X_j > 0\}$. For each $j \in J'$, the set $C_j = \{\gamma : x_\gamma \in X_j\}$ must be finite and non-empty. Moreover, for each $\gamma \in C$, there must be some $j \in J$ such that $\mu(G_{\{\gamma\}} \cap X_j) > 0$, and in this case $j \in J'$ and $\gamma \in C_j$. Thus $C = \bigcup_{j \in J'} C_j$. Because $\#(C) > \mathfrak{c}$, $\#(J') > \mathfrak{c}$ (2A1Ld).

For each $j \in J'$, choose $\gamma_j \in C_j$. Then

$$x_{\gamma_j} \in X_j \in \Sigma \subseteq \Sigma_{\gamma_j},$$

so there must be a $K_j \in \mathcal{K}$ such that $F_{\gamma_j, K_j} \subseteq X_j$.

At this point I finally turn to the result cited at the start of this example. Because $\#(J') > \mathfrak{c}$, there must be distinct j, $k \in J'$ such that x_{γ_j} and x_{γ_k} agree on $K_j \cap K_k$. We may therefore define $x \in X$ by saying that

$$\begin{aligned}
x(\delta) &= x_{\gamma_j}(\delta) \text{ if } \delta \in K_j, \\
&= x_{\gamma_k}(\delta) \text{ if } \delta \in K_k, \\
&= 0 \text{ if } \delta \in C \setminus (K_j \cup K_j).
\end{aligned}$$

Now

$$x \in F_{\gamma_j, K_j} \cap F_{\gamma_k, K_k} \subseteq X_j \cap X_k,$$

and $X_j \cap X_k \neq \emptyset$; contradicting the assumption that the X_j formed a decomposition of X. **XQ**

(h) (X, Σ, μ) is purely atomic. **P** If $E \in \Sigma$ and $\mu E > 0$, then (as remarked in (e) above) there is a $\gamma \in C$ such that $\mu(E \cap G_{\{\gamma\}}) = 1$; now $E \cap G_{\{\gamma\}}$ must be an atom. **Q**

(i) Accordingly (X, Σ, μ) is a complete, locally determined, localizable, purely atomic measure space which is not strictly localizable.

216X Basic exercises (a) In the construction of 216C, show that the subspace measure on $\{1\} \times I$ is not semi-finite.

(b) Suppose, in 216D, that $I = \omega_1$. (i) Show that the set $\{(\xi, \eta) : \xi \le \eta < \omega_1\}$ is measured by the measure constructed by Carathéodory's method from $\mu^* \restriction \mathcal{P}(I \times I)$, but not by the subspace measure on $I \times I$. (ii) Hence, or otherwise, show that the subspace measure on $I \times I$ is not locally determined.

(c) In 216Ya, 252Yq and 252Ys below, I indicate how to construct atomless versions of 216C, 216D and 216E, that is, atomless complete measure spaces of which the first is localizable but not locally determined, the second is locally determined spaces but not localizable, and the third is locally determined and localizable but not strictly localizable. Show how direct sums of these, together with counting measure and the examples described in this chapter, can be assembled to provide all 56 examples called for by the discussion in the introduction to this section.

216Y Further exercises (a) Let λ be Lebesgue measure on $[0, 1]$, and Λ its domain. Set $Y = [0, 1] \times \{0, 1\}$ and write

$$\mathrm{T} = \{F : F \subseteq Y,\, F^{-1}[\{0\}] \in \Lambda\},$$

$$\nu F = \lambda F^{-1}[\{0\}] \text{ for every } F \in \mathrm{T}.$$

Set

$$\mathrm{T}_0 = \{F : F \in \mathrm{T},\, F^{-1}[\{0\}] \triangle F^{-1}[\{1\}] \text{ is } \lambda\text{-negligible}\}.$$

Let I be an uncountable set. Set $X = Y \times I$,

$$\Sigma = \{E : E \subseteq X,\, E^{-1}[\{i\}] \in \mathrm{T} \text{ for every } i \in I,\, \{i : E^{-1}[\{i\}] \notin \mathrm{T}_0\} \text{ is countable}\},$$

$$\mu E = \sum_{i \in I} \nu E^{-1}[\{i\}] \text{ for } E \in \Sigma.$$

(i) Show that (Y, T, ν) and $(Y, \mathrm{T}_0, \nu{\restriction}\mathrm{T}_0)$ are complete probability spaces, and that for every $F \in \mathrm{T}$ there is an $F' \in \mathrm{T}_0$ such that $\nu(F \triangle F') = 0$. (ii) Show that (X, Σ, μ) is an atomless complete localizable measure space which is not locally determined.

(b) Define a measure μ on $X = \omega_2 \times \omega_2$ as follows. Take Σ to be the σ-algebra of subsets of X generated by

$$\{A \times \omega_2 : A \subseteq \omega_2\} \cup \{\omega_2 \times \alpha : \alpha < \omega_2\}.$$

For $E \in \Sigma$ set

$$W(E) = \{\xi : \xi < \omega_2,\, \sup E[\{\xi\}] = \omega_2\},$$

and set $\mu E = \#(W(E))$ if this is finite, ∞ otherwise. Show that μ is a measure on X, is localizable and locally determined, but does not have locally determined negligible sets. Find a subspace Y of X such that the subspace measure on Y is not semi-finite.

(c) Show that in the space described in 216E every set has a measurable envelope, but that this is not true in the spaces of 216C and 216D.

(d) Set $X = \omega_1 \times \omega_2$. For $E \subseteq X$ set

$$A(E) = \{\zeta : \text{for some } \xi, \text{ just one of } (\xi, \zeta),\, (\xi, \zeta + 1) \text{ belongs to } E\},$$

$$B(E) = \{\zeta : \text{there are } \xi,\, \zeta' \text{ such that } \zeta < \zeta' < \omega_2 \text{ and just one of } (\xi, \zeta),\, (\xi, \zeta') \text{ belongs to } E\},$$

$$W(E) = \{\xi : \#(E[\{\xi\}]) = \omega_2\}.$$

Let Σ be the set of subsets E of X such that $A(E)$ is countable and $\#(B(E)) \le \omega_1$. For $E \in \Sigma$, set $\mu E = \#(W(E))$ if this is finite, ∞ otherwise. (i) Show that (X, Σ, μ) is a measure space. (ii) Show that if $\hat{\mu}$ is the completion of μ, then its domain is the set of subsets E of X such that $A(E)$ is countable, and $\hat{\mu}$ is strictly localizable. (iii) Show that μ is not strictly localizable.

216 Notes and comments The examples 216C-216E are designed to form, with Lebesgue measure, a basis for constructing a complete set of examples for the concepts listed in 211A-211K. One does not really expect to encounter these phenomena in applications, but a clear understanding of the possibilities demonstrated by these examples is part of a proper appreciation of their rarity. Of course, if we add further properties to our list – for instance, the property of having locally determined negligible sets (213I), or the property that every subset should have a measurable envelope (213Xl) – then there are further positive results to complement 211L, and more examples to hunt for, like 216Yb. But it is time, perhaps past time, that we returned to the classical theorems which apply to the measure spaces at the centre of the subject.

Chapter 22

The Fundamental Theorem of Calculus

In this chapter I address one of the most important properties of the Lebesgue integral. Given an integrable function $f:[a,b]\to\mathbb{R}$, we can form its indefinite integral $F(x)=\int_a^x f(t)dt$ for $x\in[a,b]$. Two questions immediately present themselves. (i) Can we expect to have the derivative F' of F equal to f? (ii) Can we identify which functions F will appear as indefinite integrals? Reasonably satisfactory answers may be found for both of these questions: $F'=f$ almost everywhere (222E) and indefinite integrals are the absolutely continuous functions (225E). In the course of dealing with them, we need to develop a variety of techniques which lead to many striking results both in the theory of Lebesgue measure and in other, apparently unrelated, topics in real analysis.

The first step is 'Vitali's theorem' (§221), a remarkable argument – it is more a method than a theorem – which uses the geometric nature of the real line to extract disjoint subfamilies from collections of intervals. It is the foundation stone not only of the results in §222 but of all geometric measure theory, that is, measure theory on spaces with a geometric structure. I use it here to show that monotonic functions are differentiable almost everywhere (222A). Following this, Fatou's Lemma and Lebesgue's Dominated Convergence Theorem are enough to show that the derivative of an indefinite integral is almost everywhere equal to the integrand. We find that some innocent-looking manipulations of this fact take us surprisingly far; I present these in §223.

I begin the second half of the chapter with a discussion of functions 'of bounded variation', that is, expressible as the difference of bounded monotonic functions (§224). This is one of the least measure-theoretic sections in the volume; only in 224I and 224J are measure and integration even mentioned. But this material is needed for Chapter 28 as well as for the next section, and is also one of the basic topics of twentieth-century real analysis. §225 deals with the characterization of indefinite integrals as the 'absolutely continuous' functions. In fact this is now quite easy; it helps to call on Vitali's theorem again, but everything else is a straightforward application of methods previously used. The second half of the section introduces some new ideas in an attempt to give a deeper intuition into the essential nature of absolutely continuous functions. §226 returns to functions of bounded variation and their decomposition into 'saltus' and 'absolutely continuous' and 'singular' parts, the first two being relatively manageable and the last looking something like the Cantor function.

221 Vitali's theorem in $\mathbb{R}$

I give the first theorem of this chapter a section to itself. It occupies a position between measure theory and geometry (it is, indeed, one of the fundamental results of 'geometric measure theory'), and its proof involves both the measure and the geometry of the real line.

221A Vitali's theorem Let A be a bounded subset of $\mathbb{R}$ and $\mathcal{I}$ a family of non-singleton closed intervals in $\mathbb{R}$ such that every point of A belongs to arbitrarily short members of $\mathcal{I}$. Then there is a countable set $\mathcal{I}_0\subseteq\mathcal{I}$ such that (i) $\mathcal{I}_0$ is disjoint, that is, $I\cap I'=\emptyset$ for all distinct I, $I'\in\mathcal{I}_0$ (ii) $\mu(A\setminus\bigcup\mathcal{I}_0)=0$, where μ is Lebesgue measure on $\mathbb{R}$.

proof (a) If there is a finite disjoint set $\mathcal{I}_0\subseteq\mathcal{I}$ such that $A\subseteq\bigcup\mathcal{I}_0$ (including the possibility that $A=\mathcal{I}_0=\emptyset$), we can stop. So let us suppose henceforth that there is no such $\mathcal{I}_0$.

Let μ^* be Lebesgue outer measure on $\mathbb{R}$. Suppose that $|x|<M$ for every $x\in A$, and set

$$\mathcal{I}'=\{I:I\in\mathcal{I},\,I\subseteq[-M,M]\}.$$

(b) In this case, if $\mathcal{I}_0$ is any finite disjoint subset of $\mathcal{I}'$, there is a $J\in\mathcal{I}'$ which is disjoint from any member of $\mathcal{I}_0$. **P** Take $x\in A\setminus\bigcup\mathcal{I}_0$. Now there is a $\delta>0$ such that $[x-\delta,x+\delta]$ does not meet any member of $\mathcal{I}_0$, and as $|x|<M$ we can suppose that $[x-\delta,x+\delta]\subseteq[-M,M]$. Let J be a member of $\mathcal{I}$, containing x, and of length at most δ; then $J\in\mathcal{I}'$ and $J\cap\bigcup\mathcal{I}_0=\emptyset$. **Q**

(c) We can now choose a sequence $\langle\gamma_n\rangle_{n\in\mathbb{N}}$ of real numbers and a disjoint sequence $\langle I_n\rangle_{n\in\mathbb{N}}$ in $\mathcal{I}'$ inductively, as follows. Given $\langle I_j\rangle_{j<n}$ (if $n=0$, this is the empty sequence, with no members), with $I_j\in\mathcal{I}'$ for each $j<n$, and $I_j\cap I_k=\emptyset$ for $j<k<n$, set

$$\mathcal{J}_n=\{I:I\in\mathcal{I}',\,I\cap I_j=\emptyset\text{ for every }j<n\}.$$

We know from (b) that $\mathcal{J}_n\neq\emptyset$. Set

$$\gamma_n=\sup\{\mu I:I\in\mathcal{J}_n\};$$

then $0<\gamma_n\leq 2M$. We may therefore choose a set $I_n\in\mathcal{J}_n$ such that $\mu I_n\geq\frac{1}{2}\gamma_n$, and this continues the induction.

(e) Because the I_n are disjoint Lebesgue measurable subsets of $[-M, M]$, we have

$$\sum_{n=0}^{\infty} \gamma_n \le 2\sum_{n=0}^{\infty} \mu I_n \le 4M < \infty,$$

and $\lim_{n\to\infty} \gamma_n = 0$. Now define I'_n to be the closed interval with the same midpoint as I_n but five times the length, so that it projects past each end of I_n by at least γ_n. I claim that, for any n,

$$A \subseteq \bigcup_{j<n} I_j \cup \bigcup_{j\ge n} I'_j.$$

P? Suppose, if possible, otherwise. Take any x belonging to $A \setminus (\bigcup_{j<n} I_j \cup \bigcup_{j\ge n} I'_j)$. Let $\delta > 0$ be such that

$$[x-\delta, x+\delta] \subseteq [-M, M] \setminus \bigcup_{j<n} I_j,$$

and let $J \in \mathcal{I}$ be such that

$$x \in J \subseteq [x-\delta, x+\delta].$$

Then

$$\mu J > 0 = \lim_{m\to\infty} \gamma_m;$$

let m be the least integer greater than or equal to n such that $\gamma_m < \mu J$. In this case J cannot belong to $\mathcal{J}_m$, so there must be some $k < m$ such that $J \cap I_k \ne \emptyset$, because certainly $J \in \mathcal{I}'$. By the choice of δ, k cannot be less than n, so $n \le k < m$, and $\gamma_k \ge \mu J$. In this case, the distance from x to the nearest endpoint of I_k is at most $\mu J \le \gamma_k$. But the ends of I'_k project beyond the ends of I_k by at least γ_k, so $x \in I'_k$; which contradicts the choice of x. **XQ**

(f) It follows that

$$\mu^*(A \setminus \bigcup_{j<n} I_j) \le \mu(\bigcup_{j\ge n} I'_j) \le \sum_{j=n}^{\infty} \mu I'_j \le 5\sum_{j=n}^{\infty} \mu I_j.$$

As

$$\sum_{j=0}^{\infty} \mu I_j \le 2M < \infty,$$

we must have

$$\lim_{n\to\infty} \mu^*(A \setminus \bigcup_{j<n} I_j) = 0,$$

and

$$\mu(A \setminus \bigcup_{j\in\mathbb{N}} I_j) = \mu^*(A \setminus \bigcup_{j\in\mathbb{N}} I_j) \le \inf_{n\in\mathbb{N}} \mu^*(A \setminus \bigcup_{j<n} I_j) = 0.$$

Thus in this case we may set $\mathcal{I}_0 = \{I_n : n \in \mathbb{N}\}$ to obtain a countable disjoint family in $\mathcal{I}$ with $\mu(A \setminus \bigcup \mathcal{I}_0) = 0$.

221B Remarks (a) I have expressed this theorem in the form 'there is a countable set $\mathcal{I}_0 \subseteq \mathcal{I}$ such that ...' in an attempt to find a concise way of expressing the three possibilities

(i) $A = \mathcal{I} = \emptyset$, so that we must take $\mathcal{I}_0 = \emptyset$;

(ii) there are disjoint $I_0, \dots, I_n \in \mathcal{I}$ such that $A \subseteq I_0 \cup \ldots \cup I_n$, so that we can take $\mathcal{I}_0 = \{I_0, \dots, I_n\}$;

(iii) there is a disjoint sequence $\langle I_n \rangle_{n\in\mathbb{N}}$ in $\mathcal{I}$ such that $\mu(A \setminus \bigcup_{n\in\mathbb{N}} I_n) = 0$, so that we can take $\mathcal{I}_0 = \{I_n : n \in \mathbb{N}\}$.

Of course many applications, like the proof of 221A itself, will use forms of these three alternatives.

(b) The actual theorem here, as stated, will be used in the next section. But quite as important as the statement of the theorem is the principle of its proof. The I_n are chosen 'greedily', that is, when we come to choose I_n we look at the family $\mathcal{J}_n$ of possible intervals, given the choices $I_0, \dots, I_{n-1}$ already made, and choose an $I_n \in \mathcal{J}_n$ which is 'about' as big as it could be. The supremum of the possibilities for μI_n is γ_n; but since we do not know that there is any $I \in \mathcal{J}_n$ such that $\mu I = \gamma_n$, we must settle for a little less. I follow the standard formula in taking $\mu I_n \ge \frac{1}{2}\gamma_n$, but of course I could have taken $\mu I_n \ge \frac{99}{100}\gamma_n$, or $\mu I_n \ge (1 - 2^{-n})\gamma_n$, if that had helped later on. The remarkable thing is that this works; we can choose the I_n without foresight and without considering their interrelationships (for that matter, without examining the set A) beyond the minimal requirement that $I_n \cap I_j = \emptyset$ for $j < n$, and even this arbitrary and casual procedure yields a suitable sequence.

(c) I have stated the theorem in terms of bounded sets A and closed intervals, which is adequate for our needs, but very small changes in the proof suffice to deal with arbitrary (non-singleton) intervals, and another refinement handles unbounded sets A. (See 221Ya.)

221X Basic exercises (a) Let $\alpha \in]0,1[$. Suppose, in part (c) of the proof of 221A, we take $\mu I_n \ge \alpha\gamma_n$ for each $n \in \mathbb{N}$, rather than $\mu I_n \ge \frac{1}{2}\gamma_n$. What will be the appropriate constant to take in place of 5 in defining the sets I'_j of part (e)?

221Y Further exercises (a) Let A be a subset of $\mathbb{R}$ and $\mathcal{I}$ a family of non-singleton intervals in $\mathbb{R}$ such that every point of A belongs to arbitrarily short members of $\mathcal{I}$. Show that there is a countable disjoint set $\mathcal{I}_0 \subseteq \mathcal{I}$ such that $A \setminus \bigcup \mathcal{I}_0$ is Lebesgue negligible. (*Hint*: apply 221A to the sets $A \cap]n, n+1[$, $\{\overline{I} : I \in \mathcal{I}, \overline{I} \subseteq]n, n+1[\}$, writing $\overline{I}$ for the closed interval with the same endpoints as I.)

(b) Let $\mathcal{J}$ be any family of non-singleton intervals in $\mathbb{R}$. Show that $\bigcup \mathcal{J}$ is Lebesgue measurable. (*Hint*: apply (a) to $A = \bigcup \mathcal{J}$ and the family $\mathcal{I}$ of non-singleton subintervals of members of $\mathcal{J}$.)

(c) Let (X, ρ) be a metric space, A a subset of X, and $\mathcal{I}$ a family of closed balls of non-zero radius in X such that every point of A belongs to arbitrarily small members of $\mathcal{I}$. (I say here that a set is a 'closed ball of non-zero radius' if it is expressible in the form $B(x,\delta) = \{y : \rho(y,x) \le \delta\}$ where $x \in X$ and $\delta > 0$. Of course it is possible for such a ball to be a singleton $\{x\}$.) Show that either A can be covered by a finite disjoint family in $\mathcal{I}$ or there is a disjoint sequence $\langle B(x_n,\delta_n)\rangle_{n\in\mathbb{N}}$ in $\mathcal{I}$ such that

$$A \subseteq \bigcup_{m\le n} B(x_m,\delta_m) \cup \bigcup_{m>n} B(x_m, 5\delta_m) \text{ for every } n \in \mathbb{N}$$

or there is a disjoint sequence $\langle B(x_n,\delta_n)\rangle_{n\in\mathbb{N}}$ in $\mathcal{I}$ such that $\inf_{n\in\mathbb{N}} \delta_n > 0$.

(d) Give an example of a family $\mathcal{I}$ of open intervals such that every point of $\mathbb{R}$ belongs to arbitrarily small members of $\mathcal{I}$, but if $\langle I_n\rangle_{n\in\mathbb{N}}$ is any disjoint sequence in $\mathcal{I}$, and for each $n \in \mathbb{N}$ we write I'_n for the closed interval with the same centre as I_n and ten times the length, then there is an n such that $]0,1[\not\subseteq \bigcup_{m<n} I_m \cup \bigcup_{m\ge n} I'_m$.

(e)(i) Show that if $\mathcal{I}$ is a *finite* family of intervals in $\mathbb{R}$ there are $\mathcal{I}_0$, $\mathcal{I}_1 \subseteq \mathcal{I}$ such that $\bigcup(\mathcal{I}_0 \cup \mathcal{I}_1) = \bigcup \mathcal{I}$ and both $\mathcal{I}_0$ and $\mathcal{I}_1$ are disjoint families. (*Hint*: induce on $\#(\mathcal{I})$.) (ii) Suppose that $\mathcal{I}$ is a family of non-singleton intervals, of length at most 1, covering a bounded set $A \subseteq \mathbb{R}$, and that $\epsilon > 0$. Show that there is a disjoint subfamily $\mathcal{I}_0$ of $\mathcal{I}$ such that $\mu^*(A \setminus \bigcup \mathcal{I}_0) \le \frac{1}{2}\mu^* A + \epsilon$. (*Hint*: replacing each member of $\mathcal{I}$ by a slightly longer one with rational endpoints, reduce to the case in which $\mathcal{I}$ is countable and thence to the case in which $\mathcal{I}$ is finite; now use (i).) (iii) Use (ii) to prove Vitali's theorem. (I learnt this argument from J.Aldaz.)

221 Notes and comments I have headed this section 'Vitali's theorem in $\mathbb{R}$' because there is an r-dimensional version, which will appear in Chapter 26 below. There is an anomaly in the position of this theorem. It is an indispensable element of the proofs of some of the most important theorems in measure theory; on the other hand, the ideas involved in its own proof are not used elsewhere in the elementary theory. I have therefore myself sometimes omitted the proof when teaching this material, and would not reproach any student who left it to one side for the moment. At some stage, of course, any measure theorist must master the method, not just for the sake of completeness, but in order to gain an intuition for possible variations. I must emphasize that it is the *principle* of the proof, rather than its details, which is important, because there are innumerable forms of 'Vitali's theorem'. (I offer some variations in the exercises above and in §261 below, and there are many others which are important in more advanced work; one will appear in §472 in Volume 4.) This principle is, I suppose, that

(i) we choose the I_n greedily, according to some more or less natural criterion applicable to each I_n as we come to choose it, without attempting to look ahead;

(ii) we prove that their sizes tend to zero, even though we seemed to do nothing to ensure that they would (but note the shift from $\mathcal{I}$ to $\mathcal{I}'$ in part (a) of the proof of 221A, which is exactly what is needed to make this step work);

(iii) we check that for a suitable definition of I'_n, enlarging I_n, we shall have $A \subseteq \bigcup_{m<n} I_m \cup \bigcup_{m\ge n} I'_m$ for every n, while $\sum_{n=0}^{\infty} \mu I'_n < \infty$.

In a way, we have to count ourselves lucky every time this works. The reason for studying as many variations as possible of a technique of this kind is to learn to guess when we might be lucky.

222 Differentiating an indefinite integral

I come now to the first of the two questions mentioned in the introduction to this chapter: if f is an integrable function on $[a,b]$, what is $\frac{d}{dx}\int_a^x f$? It turns out that this derivative exists and is equal to f almost everywhere (222E). The argument is based on a striking property of monotonic functions: they are differentiable almost everywhere (222A), and we can bound the integrals of their derivatives (222C).

222A Theorem Let $I \subseteq \mathbb{R}$ be an interval and $f : I \to \mathbb{R}$ a monotonic function. Then f is differentiable almost everywhere in I.

Remark If I seem to be speaking of a measure on $\mathbb{R}$ without naming it, as here, I mean Lebesgue measure.

proof As usual, write μ^* for Lebesgue outer measure on $\mathbb{R}$, μ for Lebesgue measure.

(a) To begin with (down to the end of (c) below), let us suppose that f is non-decreasing and I is a bounded open interval on which f is bounded; say $|f(x)| \le M$ for $x \in I$. For any closed subinterval $J = [a,b]$ of I, write $f^*(J)$ for the open interval $]f(a), f(b)[$. For $x \in I$, write

$$D^*f(x) = \limsup_{h\to 0} \tfrac{1}{h}(f(x+h) - f(x)), \quad D_*f(x) = \liminf_{h\to 0} \tfrac{1}{h}(f(x+h) - f(x)),$$

allowing the value ∞ in both cases. Then f is differentiable at x iff $D^*f(x) = D_*f(x) \in \mathbb{R}$. Because surely $D^*f(x) \ge D_*f(x) \ge 0$, f will be differentiable at x iff $D^*f(x)$ is finite and $D^*f(x) \le D_*f(x)$.

I therefore have to show that the sets

$$\{x : x \in I,\ D^*f(x) = \infty\}, \quad \{x : x \in I,\ D^*f(x) > D_*f(x)\}$$

are negligible.

(b) Let us take $A = \{x : x \in I,\ D^*f(x) = \infty\}$ first. Fix an integer $m \ge 1$ for the moment, and set

$$A_m = \{x : x \in I,\ D^*f(x) > m\} \supseteq A.$$

Let $\mathcal{I}$ be the family of non-trivial closed intervals $[a,b] \subseteq I$ such that $f(b) - f(a) \ge m(b-a)$; then $\mu f^*(J) \ge m\mu J$ for every $J \in \mathcal{I}$. If $x \in A_m$, then for any $\delta > 0$ we have an h with $0 < |h| \le \delta$ and $\frac{1}{h}(f(x+h) - f(x)) > m$, so that

$$[x, x+h] \in \mathcal{I} \text{ if } h > 0, \quad [x+h, x] \in \mathcal{I} \text{ if } h < 0;$$

thus every member of A_m belongs to arbitrarily small intervals in $\mathcal{I}$. By Vitali's theorem (221A), there is a countable disjoint set $\mathcal{I}_0 \subseteq \mathcal{I}$ such that $\mu(A \setminus \bigcup \mathcal{I}_0) = 0$. Now, because f is non-decreasing, $\langle f^*(J)\rangle_{J\in\mathcal{I}_0}$ is disjoint, and all the $f^*(J)$ are included in $[-M, M]$, so $\sum_{J\in\mathcal{I}_0} \mu f^*(J) \le 2M$ and $\sum_{J\in\mathcal{I}_0} \mu J \le 2M/m$. Because $A_m \setminus \bigcup \mathcal{I}_0$ is negligible,

$$\mu^*A \le \mu^*A_m \le \frac{2M}{m}.$$

As m is arbitrary, $\mu^*A = 0$ and A is negligible.

(c) Now consider $B = \{x : x \in I,\ D^*f(x) > D_*f(x)\}$. For q, $q' \in \mathbb{Q}$ with $0 \le q < q'$, set

$$B_{qq'} = \{x : x \in I,\ D_*f(x) < q,\ D^*f(x) > q'\}.$$

Fix such q, q' for the moment, and write $\gamma = \mu^*B_{qq'}$. Take any $\epsilon > 0$, and let G be an open set including $B_{qq'}$ such that $\mu G \le \gamma + \epsilon$ (134Fa). Let $\mathcal{J}$ be the set of non-trivial closed intervals $[a,b] \subseteq I \cap G$ such that $f(b) - f(a) \le q(b-a)$; this time $\mu f^*(J) \le q\mu J$ for $J \in \mathcal{J}$. Then every member of $B_{qq'}$ is included in arbitrarily small members of $\mathcal{J}$, so there is a countable disjoint $\mathcal{J}_0 \subseteq \mathcal{J}$ such that $B_{qq'} \setminus \bigcup \mathcal{J}_0$ is negligible. Let L be the set of endpoints of members of $\mathcal{J}_0$; then L is a countable union of doubleton sets, so is countable, therefore negligible. Set

$$C = B_{qq'} \cap \bigcup \mathcal{J}_0 \setminus L;$$

then $\mu^*C = \gamma$. Let $\mathcal{I}$ be the set of non-trivial closed intervals $J = [a,b]$ such that (i) J is included in one of the members of $\mathcal{J}_0$ (ii) $f(b) - f(a) \ge q'(b-a)$; now $\mu f^*(J) \ge q'\mu J$ for every $J \in \mathcal{I}$. Once again, because every member of C is an interior point of some member of $\mathcal{J}_0$, every point of C belongs to arbitrarily small members of $\mathcal{I}$; so there is a countable disjoint $\mathcal{I}_0 \subseteq \mathcal{I}$ such that $\mu(C \setminus \bigcup \mathcal{I}_0) = 0$.

As in (b) above,

$$\gamma q' \le q'\mu(\bigcup \mathcal{I}_0) = \sum_{I\in\mathcal{I}_0} q'\mu I \le \sum_{I\in\mathcal{I}_0} \mu f^*(I) = \mu(\bigcup_{I\in\mathcal{I}_0} f^*(I)).$$

On the other hand,

$$\mu(\bigcup_{J\in\mathcal{J}_0} f^*(J)) = \sum_{J\in\mathcal{J}_0} \mu f^*(J) \le q \sum_{J\in\mathcal{J}_0} \mu J = q\mu(\bigcup \mathcal{J}_0)$$
$$\le q\mu(\bigcup \mathcal{J}) \le q\mu G \le q(\gamma+\epsilon).$$

But $\bigcup_{I\in\mathcal{I}_0} f^*(I) \subseteq \bigcup_{J\in\mathcal{J}_0} f^*(J)$, because every member of $\mathcal{I}_0$ is included in a member of $\mathcal{J}_0$, so $\gamma q' \le q(\gamma+\epsilon)$ and $\gamma \le \epsilon q/(q'-q)$. As ϵ is arbitrary, $\gamma = 0$.

Thus every $B_{qq'}$ is negligible. Consequently $B = \bigcup_{q,q'\in\mathbb{Q},0\le q<q'} B_{qq'}$ is negligible.

(d) This deals with the case of a bounded open interval on which f is bounded and non-decreasing. Still for non-decreasing f, but for an arbitrary interval I, observe that $K = \{(q,q') : q,\, q' \in I\cap\mathbb{Q},\, q < q'\}$ is countable and that $I\setminus\bigcup_{(q,q')\in K}]q,q'[$ has at most two points (the endpoints of I, if any), so is negligible. If we write S for the set of points of I at which f is not differentiable, then from (a)-(c) we see that $S\cap\,]q,q'[$ is negligible for every $(q,q')\in K$, so that $S\cap\bigcup_{(q,q')\in K}]q,q'[$ is negligible and S is negligible.

(e) Thus we are done if f is non-decreasing. For non-increasing f, apply the above to $-f$, which is differentiable at exactly the same points as f.

222B Remarks(a) I note that in the above argument I am using such formulae as $\sum_{J\in\mathcal{I}_0} \mu f^*(J)$. This is because Vitali's theorem leaves it open whether the families $\mathcal{I}_0$ will be finite or infinite. The sum must be interpreted along the lines laid down in 112Bd in Volume 1; generally, $\sum_{k\in K} a_k$, where K is an arbitrary set and every $a_k \ge 0$, is to be $\sup_{L\subseteq K \text{ is finite}} \sum_{k\in L} a_k$, with the convention that $\sum_{k\in\emptyset} a_k = 0$. Now, in this context, if (X,Σ,μ) is a measure space, K is a countable set, and $\langle E_k\rangle_{k\in K}$ is a family in Σ,

$$\mu(\bigcup_{k\in K} E_k) \le \sum_{k\in K} \mu E_k,$$

with equality if $\langle E_k\rangle_{k\in K}$ is disjoint. **P** If $K=\emptyset$, this is trivial. Otherwise, let $n\mapsto k_n : \mathbb{N}\to K$ be a surjection, and set

$$K_n = \{k_i : i\le n\}, \quad G_n = \bigcup_{i\le n} E_{k_i} = \bigcup_{k\in K_n} E_k$$

for each $n\in\mathbb{N}$. Then $\langle G_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence with union $E = \bigcup_{k\in K} E_k$, so

$$\mu E = \lim_{n\to\infty} \mu G_n = \sup_{n\in\mathbb{N}} \mu G_n;$$

and

$$\mu G_n \le \sum_{k\in K_n} \mu E_k \le \sum_{k\in K} \mu E_k$$

for every n, so $\mu E \le \sum_{k\in K} \mu E_k$. If the E_k are disjoint, then μG_n is precisely $\sum_{k\in K_n} \mu E_k$ for each n; but as $\langle K_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence of sets with union K, every finite subset of K is included in some K_n, and

$$\sum_{k\in K} \mu E_k = \sup_{n\in\mathbb{N}} \sum_{k\in K_n} \mu E_k = \sup_{n\in\mathbb{N}} \mu G_n = \mu E,$$

as required. **Q**

(b) Some readers will prefer to re-index sets regularly, so that all the sums they need to look at will be of the form $\sum_{i=0}^n$ or $\sum_{i=0}^\infty$. In effect, that is what I did in Volume 1, in the proof of 114Da/115Da, when showing that Lebesgue outer measure is indeed an outer measure. The disadvantage of this procedure in the context of 222A is that we must continually check that it doesn't matter whether we have a finite or infinite sum at any particular moment. I believe that it is worth taking the trouble to learn the technique sketched here, because it very frequently happens that we wish to consider unions of sets indexed by sets other than $\mathbb{N}$ and $\{0,\dots,n\}$.

(c) Of course the argument above can be shortened if you know a tiny bit more about countable sets than I have explicitly stated so far. But note that the value assigned to $\sum_{k\in K} a_k$ must not depend on which enumeration $\langle k_n\rangle_{n\in\mathbb{N}}$ we pick on.

222C Lemma Suppose that $a\le b$ in $\mathbb{R}$, and that $F:[a,b]\to\mathbb{R}$ is a non-decreasing function. Then $\int_a^b F'$ exists and is at most $F(b)-F(a)$.

Remark I discussed integration over subsets at length in §131 and §214. For measurable subsets, which are sufficient for our needs in this chapter, we have a simple description: if (X,Σ,μ) is a measure space, $E\in\Sigma$ and f is a real-valued function, then $\int_E f = \int \tilde{f}$ if the latter integral exists, where $\operatorname{dom}\tilde{f} = (E\cap\operatorname{dom} f)\cup(X\setminus E)$ and $\tilde{f}(x) = f(x)$ if $x\in E\cap\operatorname{dom} f$, 0 if $x\in X\setminus E$ (apply 131Fa to $\tilde{f}$). It follows at once that if now $F\in\Sigma$ and $F\subseteq E$, $\int_F f = \int_E f\times\chi F$.

I write $\int_a^x f$ to mean $\int_{[a,x[} f$, which (because $[a,x[$ is measurable) can be dealt with as described above. Note that, as long as we are dealing with Lebesgue measure, so that $[a,x]\setminus]a,x[=\{a,x\}$ is negligible, there is no need to distinguish between $\int_{[a,x]}$, $\int_{]a,x[}$, $\int_{[a,x[}$, $\int_{]a,x]}$; for other measures on $\mathbb{R}$ we may need to take more care. I use half-open intervals to make it obvious that $\int_a^x f+\int_x^y f=\int_a^y f$ if $a\le x\le y$, because

$$f\times\chi[a,y[=f\times\chi[a,x[+f\times\chi[x,y[.$$

proof (a) The result is trivial if $a=b$; let us suppose that $a<b$. By 222A, F' is defined almost everywhere in $[a,b]$.

(b) For each $n\in\mathbb{N}$, define a simple function $g_n:[a,b[\to\mathbb{R}$ as follows. For $0\le k<2^n$, set $a_{nk}=a+2^{-n}k(b-a)$, $b_{nk}=a+2^{-n}(k+1)(b-a)$, $I_{nk}=[a_{nk},b_{nk}[$. For each $x\in[a,b[$, take that $k<2^n$ such that $x\in I_{nk}$, and set

$$g_n(x)=\frac{2^n}{b-a}(F(b_{nk})-F(a_{nk}))$$

for $x\in I_{nk}$, so that g_n gives the slope of the chord of the graph of F defined by the endpoints of I_{nk}. Then

$$\int_a^b g_n=\sum_{k=0}^{2^n-1}F(b_{nk})-F(a_{nk})=F(b)-F(a).$$

(c) On the other hand, if we set

$$C=\{x:x\in\,]a,b[\,,\ F'(x)\text{ exists}\},$$

then $[a,b]\setminus C$ is negligible, by 222A, and $F'(x)=\lim_{n\to\infty}g_n(x)$ for every $x\in C$. **P** Let $\epsilon>0$. Then there is a $\delta>0$ such that $x+h\in[a,b]$ and $|F(x+h)-F(x))-hF'(x)|\le\epsilon|h|$ whenever $|h|\le\delta$. Let $n\in\mathbb{N}$ be such that $2^{-n}(b-a)\le\delta$. Let $k<2^n$ be such that $x\in I_{nk}$. Then

$$x-\delta\le a_{nk}\le x<b_{nk}\le x+\delta,\quad g_n(x)=\frac{2^n}{b-a}(F(b_{nk})-F(a_{nk})).$$

Now we have

$$\begin{aligned}|g_n(x)-F'(x)|&=|\frac{2^n}{b-a}(F(b_{nk})-F(a_{nk}))-F'(x)|\\&=\frac{2^n}{b-a}|F(b_{nk})-F(a_{nk})-(b_{nk}-a_{nk})F'(x)|\\&\le\frac{2^n}{b-a}\big(|F(b_{nk})-F(x)-(b_{nk}-x)F'(x)|\\&\qquad\qquad+|F(x)-F(a_{nk})-(x-a_{nk})F'(x)|\big)\\&\le\frac{2^n}{b-a}(\epsilon|b_{nk}-x|+\epsilon|x-a_{nk}|)=\epsilon.\end{aligned}$$

And this is true whenever $2^{-n}\le\delta$, that is, for all n large enough. As ϵ is arbitrary, $F'(x)=\lim_{n\to\infty}g_n(x)$. **Q**

(d) Thus $g_n\to F'$ almost everywhere in $[a,b]$. By Fatou's Lemma,

$$\int_a^b F'=\int_a^b\liminf_{n\to\infty}g_n\le\liminf_{n\to\infty}\int_a^b g_n=\lim_{n\to\infty}\int_a^b g_n=F(b)-F(a),$$

as required.

Remark There is a generalization of this result in 224I.

222D Lemma Suppose that $a<b$ in $\mathbb{R}$, and that f, g are real-valued functions, both integrable over $[a,b]$, such that $\int_a^x f=\int_a^x g$ for every $x\in[a,b]$. Then $f=g$ almost everywhere in $[a,b]$.

proof The point is that

$$\int_E f=\int_a^b f\times\chi E=\int_a^b g\times\chi E=\int_E g$$

for any measurable set $E\subseteq[a,b[$.

P (i) If $E=[c,d[$ where $a\le c\le d\le b$, then

$$\int_E f=\int_a^d f-\int_a^c f=\int_a^d g-\int_a^c g=\int_E g.$$

(ii) If $E=[a,b[\cap G$ for some open set $G\subseteq\mathbb{R}$, then for each $n\in\mathbb{N}$ set

$$K_n=\{k:k\in\mathbb{Z},\ |k|\le 4^n,\ [2^{-n}k,2^{-n}(k+1)[\subseteq G\},$$

$$H_n = \bigcup_{k\in K_n} [2^{-n}k, 2^{-n}(k+1)[\cap [a,b[;$$

then $\langle H_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence of measurable sets with union E, so $f\times\chi E = \lim_{n\to\infty} f\times\chi H_n$, and (by Lebesgue's Dominated Convergence Theorem, because $|f\times\chi H_n|\le|f|$ almost everywhere for every n, and $|f|$ is integrable)

$$\int_E f = \lim_{n\to\infty}\int_{H_n} f.$$

At the same time, each H_n is a finite disjoint union of half-open intervals in $[a,b[$, so

$$\int_{H_n} f = \sum_{k\in K_n}\int_{[2^{-n}k,2^{-n}(k+1)[\cap[a,b[} f = \sum_{k\in K_n}\int_{[2^{-n}k,2^{-n}(k+1)[\cap[a,b[} g = \int_{H_n} g,$$

and

$$\int_E g = \lim_{n\to\infty}\int_{H_n} g = \lim_{n\to\infty}\int_{H_n} f = \int_E f.$$

(iii) For general measurable $E\subseteq[a,b[$, we can choose for each $n\in\mathbb{N}$ an open set $G_n\supseteq E$ such that $\mu G_n\le\mu E+2^{-n}$ (134Fa). Set $G'_n=\bigcap_{m\le n}G_m$, $E_n=[a,b[\cap G'_n$ for each n,

$$F = [a,b[\cap\bigcap_{n\in\mathbb{N}}G_n = \bigcap_{n\in\mathbb{N}}[a,b[\cap G'_n = \bigcap_{n\in\mathbb{N}}E_n.$$

Then $E\subseteq F$ and

$$\mu F\le\inf_{n\in\mathbb{N}}\mu G_n = \mu E,$$

so $F\setminus E$ is negligible and $f\times\chi(F\setminus E)$ is zero almost everywhere; consequently $\int_{F\setminus E} f = 0$ and $\int_F f = \int_E f$. On the other hand,

$$f\times\chi F = \lim_{n\to\infty} f\times\chi E_n,$$

so by Lebesgue's Dominated Convergence Theorem again

$$\int_E f = \int_F f = \lim_{n\to\infty}\int_{E_n} f.$$

Similarly

$$\int_E g = \lim_{n\to\infty}\int_{E_n} g.$$

But by part (ii) we have $\int_{E_n} g = \int_{E_n} f$ for every n, so $\int_E g = \int_E f$, as required. **Q**

By 131Hb, $f=g$ almost everywhere in $[a,b[$, and therefore almost everywhere in $[a,b]$.

222E Theorem Suppose that $a\le b$ in $\mathbb{R}$ and that f is a real-valued function which is integrable over $[a,b]$. Then $F(x)=\int_a^x f$ exists in $\mathbb{R}$ for every $x\in[a,b]$, and the derivative $F'(x)$ exists and is equal to $f(x)$ for almost every $x\in[a,b]$.

proof (a) For most of this proof (down to the end of (c) below) I suppose that f is non-negative. In this case,

$$F(y) = F(x)+\int_x^y f\ge F(x)$$

whenever $a\le x\le y\le b$; thus F is non-decreasing and therefore differentiable almost everywhere in $[a,b]$, by 222A.

By 222C we know also that $\int_a^x F'$ exists and is less than or equal to $F(x)-F(a)=F(x)$ for every $x\in[a,b]$.

(b) Now suppose, in addition, that f is bounded; say $0\le f(t)\le M$ for every $t\in\operatorname{dom} f$. Then $M-f$ is integrable over $[a,b]$; let G be its indefinite integral, so that $G(x)=M(x-a)-F(x)$ for every $x\in[a,b]$. Applying (a) to $M-f$ and G, we have $\int_a^x G'\le G(x)$ for every $x\in[a,b]$; but of course $G'=M-F'$, so $M(x-a)-\int_a^x F'\le M(x-a)-F(x)$, that is, $\int_a^x F'\ge F(x)$ for every $x\in[a,b]$. Thus $\int_a^x F'=\int_a^x f$ for every $x\in[a,b]$. Now 222D tells us that $F'=f$ almost everywhere in $[a,b]$.

(c) Thus for bounded, non-negative f we are done. For unbounded f, let $\langle f_n\rangle_{n\in\mathbb{N}}$ be a non-decreasing sequence of non-negative simple functions converging to f almost everywhere in $[a,b]$, and let $\langle F_n\rangle_{n\in\mathbb{N}}$ be the corresponding indefinite integrals. Then for any n and any x, y with $a\le x\le y\le b$, we have

$$F(y)-F(x) = \int_x^y f\ge\int_x^y f_n = F_n(y)-F_n(x),$$

so that $F'(x)\ge F'_n(x)$ for any $x\in\,]a,b[$ where both are defined, and $F'(x)\ge f_n(x)$ for almost every $x\in[a,b]$. This is true for every n, so $F'\ge f$ almost everywhere, and $F'-f\ge 0$ almost everywhere. On the other hand, as noted in (a),

$$\int_a^b F'\le F(b)-F(a) = \int_a^b f,$$

so $\int_a^b F' - f \le 0$. It follows that $F' =_{\text{a.e.}} f$ (that is, that $F' = f$ almost everywhere in $[a,b]$)(122Rd).

(d) This completes the proof for non-negative f. For general f, we can express f as $f_1 - f_2$ where f_1, f_2 are non-negative integrable functions; now $F = F_1 - F_2$ where F_1, F_2 are the corresponding indefinite integrals, so $F' =_{\text{a.e.}} F_1' - F_2' =_{\text{a.e.}} f_1 - f_2$, and $F' =_{\text{a.e.}} f$.

222F Corollary Suppose that f is any real-valued function which is integrable over $\mathbb{R}$, and set $F(x) = \int_{-\infty}^x f$ for every $x \in \mathbb{R}$. Then $F'(x)$ exists and is equal to $f(x)$ for almost every $x \in \mathbb{R}$.

proof For each $n \in \mathbb{N}$, set

$$F_n(x) = \int_{-n}^x f$$

for $x \in [-n,n]$. Then $F_n'(x) = f(x)$ for almost every $x \in [-n,n]$. But $F(x) = F(-n) + F_n(x)$ for every $x \in [-n,n]$, so $F'(x)$ exists and is equal to $F_n'(x)$ for every $x \in \,]-n,n[$ for which $F_n'(x)$ is defined; and $F'(x) = f(x)$ for almost every $x \in [-n,n]$. As n is arbitrary, $F' =_{\text{a.e.}} f$.

222G Corollary Suppose that $E \subseteq \mathbb{R}$ is a measurable set and that f is a real-valued function which is integrable over E. Set $F(x) = \int_{E\cap]-\infty,x[} f$ for $x \in \mathbb{R}$. Then $F'(x) = f(x)$ for almost every $x \in E$, and $F'(x) = 0$ for almost every $x \in \mathbb{R} \setminus E$.

proof Apply 222F to $\tilde{f}$, where $\tilde{f}(x) = f(x)$ for $x \in E \cap \operatorname{dom} f$ and $\tilde{f}(x) = 0$ for $x \in \mathbb{R} \setminus E$, so that $F(x) = \int_{-\infty}^x \tilde{f}$ for every $x \in \mathbb{R}$.

222H The result that $\frac{d}{dx}\int_a^x f = f(x)$ for almost every x is satisfying, but is no substitute for the more elementary result that this equality is valid at any point at which f is continuous.

Proposition Suppose that $a \le b$ in $\mathbb{R}$ and that f is a real-valued function which is integrable over $[a,b]$. Set $F(x) = \int_a^x f$ for $x \in [a,b]$. Then $F'(x)$ exists and is equal to $f(x)$ at any point $x \in \operatorname{dom}(f) \cap \,]a,b[$ at which f is continuous.

proof Set $c = f(x)$. Let $\epsilon > 0$. Let $\delta > 0$ be such that $\delta \le \min(b-x, x-a)$ and $|f(t) - c| \le \epsilon$ whenever $t \in \operatorname{dom} f$ and $|t - x| \le \delta$. If $x < y \le x + \delta$, then

$$\left|\frac{F(y)-F(x)}{y-x} - c\right| = \frac{1}{y-x}\left|\int_x^y f - c\right| \le \frac{1}{y-x}\int_x^y |f - c| \le \epsilon.$$

Similarly, if $x - \delta \le y < x$,

$$\left|\frac{F(y)-F(x)}{y-x} - f(x)\right| = \frac{1}{x-y}\left|\int_y^x f - c\right| \le \frac{1}{x-y}\int_y^x |f - c| \le \epsilon.$$

As ϵ is arbitrary,

$$F'(x) = \lim_{y\to x} \frac{F(y)-F(x)}{y-x} = c,$$

as required.

222I Complex-valued functions So far in this section, I have taken every f to be real-valued. The extension to complex-valued f is just a matter of applying the above results to the real and imaginary parts of f. Specifically, we have the following.

(a) If $a \le b$ in $\mathbb{R}$ and f is a complex-valued function which is integrable over $[a,b]$, then $F(x) = \int_a^x f$ is defined in $\mathbb{C}$ for every $x \in [a,b]$, and its derivative $F'(x)$ exists and is equal to $f(x)$ for almost every $x \in [a,b]$; moreover, $F'(x) = f(x)$ whenever $x \in \operatorname{dom}(f) \cap \,]a,b[$ and f is continuous at x.

(b) If f is a complex-valued function which is integrable over $\mathbb{R}$, and $F(x) = \int_{-\infty}^x f$ for each $x \in \mathbb{R}$, then F' exists and is equal to f almost everywhere in $\mathbb{R}$.

(c) If $E \subseteq \mathbb{R}$ is a measurable set and f is a complex-valued function which is integrable over E, and $F(x) = \int_{E\cap]-\infty,x[} f$ for each $x \in \mathbb{R}$, then $F'(x) = f(x)$ for almost every $x \in E$ and $F'(x) = 0$ for almost every $x \in \mathbb{R} \setminus E$.

***222J The Denjoy-Young-Saks theorem** The next result will not be used, on present plans, anywhere in this treatise. It is however central to parts of real analysis for which this volume is supposed to be a foundation, and while the argument requires a certain sophistication it is not really a large step from Lebesgue's theorem 222A. I must begin with some notation.

Definition Let f be any real function, and $A \subseteq \mathbb{R}$ its domain. Write

$$\tilde{A}^+ = \{x : x \in A,\ ]x, x+\delta] \cap A \neq \emptyset \text{ for every } \delta > 0\},$$

$$\tilde{A}^- = \{x : x \in A,\ [x-\delta, x[\cap A \neq \emptyset \text{ for every } \delta > 0\}.$$

Set

$$\bar{D}^+(x) = \limsup_{y \in A, y \downarrow x} \tfrac{f(y)-f(x)}{y-x} = \inf_{\delta>0} \sup_{y \in A, x<y\le x+\delta} \tfrac{f(y)-f(x)}{y-x},$$

$$\underline{D}^+(x) = \liminf_{y \in A, y \downarrow x} \tfrac{f(y)-f(x)}{y-x} = \sup_{\delta>0} \inf_{y \in A, x<y\le x+\delta} \tfrac{f(y)-f(x)}{y-x}$$

for $x \in \tilde{A}^+$, and

$$\bar{D}^-(x) = \limsup_{y \in A, y \uparrow x} \tfrac{f(y)-f(x)}{y-x} = \inf_{\delta>0} \sup_{y \in A, x-\delta\le y< x} \tfrac{f(y)-f(x)}{y-x},$$

$$\underline{D}^-(x) = \liminf_{y \in A, y \uparrow x} \tfrac{f(y)-f(x)}{y-x} = \sup_{\delta>0} \inf_{y \in A, x-\delta\le y< x} \tfrac{f(y)-f(x)}{y-x}$$

for $x \in \tilde{A}^-$, all defined in $[-\infty, \infty]$. (These are the four **Dini derivates** of f. You will also see D^+, d^+, D^-, d^- used in place of my $\bar{D}^+$, $\underline{D}^+$, $\bar{D}^-$ and $\underline{D}^-$.)

Note that we surely have $(\underline{D}^+f)(x) \le (\bar{D}^+f)(x)$ for every $x \in \tilde{A}^+$, while $(\underline{D}^-f)(x) \le (\bar{D}^-f)(x)$ for every $x \in \tilde{A}^-$. The ordinary derivative $f'(x)$ is defined and equal to $c \in \mathbb{R}$ iff (α) x belongs to some open interval included in A (β) $(\bar{D}^+f)(x) = (\underline{D}^+f)(x) = (\bar{D}^-f)(x) = (\underline{D}^+f)(x) = c$.

***222K Lemma** Let A be any subset of $\mathbb{R}$, and define $\tilde{A}^+$ and $\tilde{A}^-$ as in 222J. Then $A \setminus \tilde{A}^*$ and $A \setminus \tilde{A}^-$ are countable, therefore negligible.

proof We have

$$A \setminus \tilde{A}^+ = \bigcup_{q \in \mathbb{Q}} \{x : x \in A,\ x < q,\ A \cap\,]x, q] = \emptyset\}.$$

But for any $q \in \mathbb{Q}$ there can be at most one $x \in A$ such that $x < q$ and $]x, q]$ does not meet A, so $A \setminus \tilde{A}^+$ is a countable union of finite sets and is countable. Similarly,

$$A \setminus \tilde{A}^- = \bigcup_{q \in \mathbb{Q}} \{x : x \in A,\ q < x,\ A \cap [q, x[\, = \emptyset\}$$

is countable.

***222L Theorem** Let f be any real function, and A its domain. Then for almost every $x \in A$
either all four Dini derivates of f at x are defined, finite and equal
or $(\bar{D}^+f)(x) = (\underline{D}^-f)(x)$ is finite, $(\underline{D}^+f)(x) = -\infty$ and $(\bar{D}^+f)(x) = \infty$
or $(\underline{D}^+f)(x) = (\bar{D}^-f)(x)$ is finite, $(\bar{D}^+f)(x) = \infty$ and $(\underline{D}^-f)(x) = -\infty$
or $(\bar{D}^+f)(x) = (\bar{D}^-f)(x) = \infty$ and $(\underline{D}^+f)(x) = (\underline{D}^-f)(x) = -\infty$.

proof (a) Set $\tilde{A} = \tilde{A}^+ \cap \tilde{A}^-$, defining $\tilde{A}^+$ and $\tilde{A}^-$ as in 222J, so that $\tilde{A}$ is a cocountable subset of A and all four Dini derivates are defined on $\tilde{A}$. For $n \in \mathbb{N}$, $q \in \mathbb{Q}$ set

$$E_{qn} = \{x : x \in \tilde{A},\ x < q,\ f(y) \ge f(x) - n(y-x)\} \text{ for every } y \in A \cap [x, q]\}.$$

Observe that

$$\bigcup_{n \in \mathbb{N}, q \in \mathbb{Q}} E_{qn} = \{x : x \in \tilde{A},\ (\underline{D}^+f)(x) > -\infty\}.$$

For those $q \in \mathbb{Q}$, $n \in \mathbb{N}$ such that E_{qn} is not empty, set $\beta_{qn} = \sup E_{qn} \in\,]-\infty, q]$, $\alpha_{qn} = \inf E_{qn} \in [-\infty, \beta_{qn}]$, and for $x \in\,]\alpha_{qn}, \beta_{qn}[$ set $g_{qn}(x) = \inf\{f(y) + ny : y \in A \cap [x, q]\}$. Note that if $x \in E_{qn} \setminus \{\alpha_{qn}, \beta_{qn}\}$ then $g_{qn}(x) = f(x) + nx$ is finite; also g is monotonic, therefore finite everywhere in $]\alpha_{qn}, \beta_{qn}[$, and of course $g_{qn}(x) \le f(x) + nx$ for every $x \in A \cap\,]\alpha_{qn}, \beta_{qn}[$.

By 222A, almost every point of $]\alpha_{qn}, \beta_{qn}[$ belongs to $F_{qn} = \operatorname{dom} g'_{qn}$; in particular, $E_{qn} \setminus F_{qn}$ is negligible. Set $h_{qn}(x) = g_{qn}(x) - nx$ for $x \in\,]\alpha_{qn}, \beta_{qn}[$; then h is differentiable at every point of F_{qn}. Now if $x \in E_{qn} \cap F_{qn}$, we have $h_{qn}(x) = f(x)$, while $h_{qn}(x) \le f(x)$ for $x \in A \cap\,]\alpha_{qn}, \beta_{qn}[$; it follows that

$$\begin{aligned}(\underline{D}^+f)(x) &= \sup_{\delta>0}\inf_{y\in A\cap]x,x+\delta]}\tfrac{f(y)-f(x)}{y-x}\\ &\ge \sup_{\delta>0}\inf_{y\in A\cap]x,x+\delta]}\tfrac{h_{qn}(y)-h_{qn}(x)}{y-x}\\ &\ge \sup_{0<\delta<\beta_{qn}-x}\inf_{y\in]x,x+\delta]}\tfrac{h_{qn}(y)-h_{qn}(x)}{y-x}=h'_{qn}(x),\end{aligned}$$

and similarly $(\bar{D}^-f)(x)\le h'_{qn}(x)$. On the other hand, if $x\in\tilde{E}^+_{qn}$, then

$$\begin{aligned}(\underline{D}^+f)(x) &= \sup_{\delta>0}\inf_{y\in A\cap]x,x+\delta]}\tfrac{f(y)-f(x)}{y-x}\\ &\le \sup_{\delta>0}\inf_{y\in E_{qn}\cap]x,x+\delta]}\tfrac{f(y)-f(x)}{y-x}\\ &= \sup_{\delta>0}\inf_{y\in E_{qn}\cap]x,x+\delta]}\tfrac{h_{qn}(y)-h_{qn}(x)}{y-x}\\ &\le \inf_{\delta>0}\sup_{y\in E_{qn}\cap]x,x+\delta]}\tfrac{h_{qn}(y)-h_{qn}(x)}{y-x}\\ &\le \inf_{0<\delta<\beta_{qn}-x}\sup_{y\in]x,x+\delta]}\tfrac{h_{qn}(y)-h_{qn}(x)}{y-x}=h'_{qn}(x).\end{aligned}$$

Putting these together, we see that if $x\in F_{qn}\cap\tilde{E}^+_{qn}$ then $(\underline{D}^+f)(x)=h'_{qn}(x)\ge(\bar{D}^-f)(x)$.

Conventionally setting $F_{qn}=\emptyset$ if E_{qn} is empty, the last sentence is true for all $q\in\mathbb{Q}$ and $n\in\mathbb{N}$, while $A\setminus\tilde{A}$ and $\bigcup_{q\in\mathbb{Q},n\in\mathbb{N}}E_{qn}\setminus(F_{qn}\cap\tilde{E}^+_{qn})$ are negligible, and $(\underline{D}^+f)(x)=-\infty$ for every $x\in\tilde{A}\setminus\bigcup_{q\in\mathbb{Q},n\in\mathbb{N}}E_{qn}$. So we see that, for almost every $x\in A$, either $(\underline{D}^+f)(x)=-\infty$ or $\infty>(\underline{D}^+f)(x)\ge(\bar{D}^-f)(x)$.

(b) Reflecting the above argument left-to-right or up-to-down, we see that, for almost every $x\in A$,

$$\text{either } (\underline{D}^-f)(x)=-\infty \text{ or } \infty>(\underline{D}^-f)(x)\ge(\bar{D}^+f)(x),$$

$$\text{either } (\bar{D}^+f)(x)=\infty \text{ or } -\infty<(\bar{D}^+f)(x)\le(\underline{D}^-f)(x),$$

$$\text{either } (\bar{D}^-f)(x)=\infty \text{ or } -\infty<(\bar{D}^-f)(x)\le(\underline{D}^+f)(x),$$

and also

$$(\underline{D}^+f)(x)\le(\bar{D}^+f)(x),\quad (\underline{D}^-f)(x)\le(\bar{D}^-f)(x).$$

For such x, therefore,

$$(\underline{D}^+f)(x)>-\infty\implies(\bar{D}^-f)(x)\le(\underline{D}^+f)(x)<\infty\implies(\underline{D}^+f)(x)=(\bar{D}^-f)(x)\in\mathbb{R},$$

and similarly

$$(\underline{D}^-f)(x)>-\infty\implies(\underline{D}^-f)(x)=(\bar{D}^+f)(x)\in\mathbb{R},$$

$$(\bar{D}^+f)(x)<\infty\implies(\bar{D}^+f)(x)=(\underline{D}^-f)(x)\in\mathbb{R},$$

$$(\bar{D}^-f)(x)<\infty\implies(\bar{D}^-f)(x)=(\underline{D}^+f)(x)\in\mathbb{R}.$$

So we have

$$\text{either } (\underline{D}^+f)(x)=(\bar{D}^-f)(x) \text{ is finite or } (\underline{D}^+f)(x)=-\infty \text{ and } (\bar{D}^-f)(x)=\infty,$$

$$\text{either } (\underline{D}^-f)(x)=(\bar{D}^+f)(x) \text{ is finite or } (\underline{D}^-f)(x)=-\infty \text{ and } (\bar{D}^+f)(x)=\infty.$$

These two dichotomies lead to four possibilities; and since

$$(\underline{D}^+f)(x)=(\bar{D}^-f)(x) \text{ is finite},\quad (\underline{D}^-f)(x)=(\bar{D}^+f)(x) \text{ is finite}$$

can be true together only when all four derivates are equal and finite, we have the four cases listed in the statement of the theorem.

222X Basic exercises >(a) Let $F:[0,1]\to[0,1]$ be the Cantor function (134H). Show that $\int_0^1 F'=0<F(1)-F(0)$.

>(b) Suppose that $a<b$ in $\mathbb{R}$ and that h is a real-valued function such that $\int_x^y h$ exists and is non-negative whenever $a\le x\le y\le b$. Show that $h\ge 0$ almost everywhere in $[a,b]$.

>(c) Suppose that $a<b$ in $\mathbb{R}$ and that f, g are integrable complex-valued functions on $[a,b]$ such that $\int_a^x f=\int_a^x g$ for every $x\in[a,b]$. Show that $f=g$ almost everywhere in $[a,b]$.

>(d) Suppose that $a<b$ in $\mathbb{R}$ and that f is a real-valued function which is integrable over $[a,b]$. Show that the indefinite integral $x\mapsto\int_a^x f$ is continuous.

222Y Further exercises (a) Let $\langle F_n\rangle_{n\in\mathbb{N}}$ be a sequence of non-negative, non-decreasing functions on $[0,1]$ such that $F(x)=\sum_{n=0}^\infty F_n(x)$ is finite for every $x\in[0,1]$. Show that $\sum_{n=0}^\infty F_n'(x)=F'(x)$ for almost every $x\in[0,1]$. (*Hint*: take $\langle n_k\rangle_{k\in\mathbb{N}}$ such that $\sum_{k=0}^\infty F(1)-G_k(1)<\infty$, where $G_k=\sum_{j=0}^{n_k}F_j$, and set $H(x)=\sum_{k=0}^\infty F(x)-G_k(x)$. Observe that $\sum_{k=0}^\infty F'(x)-G_k'(x)\le H'(x)$ whenever all the derivatives are defined, so that $F'=\lim_{k\to\infty}G_k'$ almost everywhere.)

(b) Let $F:[0,1]\to\mathbb{R}$ be a continuous non-decreasing function. (i) Show that if $c\in\mathbb{R}$ then $C=\{(x,y):x,\,y\in[0,1],\,F(y)-F(x)=c\}$ is connected. (*Hint*: A set $A\subseteq\mathbb{R}^r$ is **connected** if there is no continuous surjection $h:A\to\{0,1\}$. Show that if $h:C\to\{0,1\}$ is continuous then it is of the form $(x,y)\mapsto h_1(x)$ for some continuous function h_1.) (ii) Now suppose that $F(0)=0$, $F(1)=1$ and that $G:[0,1]\to[0,1]$ is a second continuous non-decreasing function with $G(0)=0$, $G(1)=1$. Show that for any $n\ge 1$ there are x, $y\in[0,1]$ such that $F(y)-F(x)=G(y)-G(x)=\frac{1}{n}$.

(c) Let f, g be non-negative integrable functions on $\mathbb{R}$, and $n\ge 1$. Show that there are $u<v$ in $[-\infty,\infty]$ such that $\int_u^v f=\frac{1}{n}\int f$ and $\int_u^v g=\frac{1}{n}\int g$.

(d) Let $f:\mathbb{R}\to\mathbb{R}$ be measurable. Show that $H=\operatorname{dom}f'$ is a measurable set and that f' is a measurable function.

(e) Construct a Borel measurable function $f:[0,1]\to\{-1,0,1\}$ such that each of the four possibilities described in Theorem 222L occurs on a set of measure $\frac{1}{4}$.

222 Notes and comments I have relegated to an exercise (222Xd) the fundamental fact that an indefinite integral $x\mapsto\int_a^x f$ is always continuous; this is not strictly speaking needed in this section, and a much stronger result is given in 225E. There is also much more to be said about monotonic functions, to which I will return in §224. What we need here is the fact that they are differentiable almost everywhere (222A), which I prove by applying Vitali's theorem three times, once in part (b) of the proof and twice in part (c). Following this, the arguments of 222C-222E form a fine series of exercises in the central ideas of Volume 1, using the concept of integration over a (measurable) subset, Fatou's Lemma (part (d) of the proof of 222C), Lebesgue's Dominated Convergence Theorem (parts (ii) and (iii) of the proof of 222D) and the approximation of Lebesgue measurable sets by open sets (part (iii) of the proof of 222D). Of course knowing that $\frac{d}{dx}\int_a^x f=f(x)$ almost everywhere is not at all the same thing as knowing that this holds for any particular x, and when we come to differentiate any particular indefinite integral we generally turn to 222H first; the point of 222E is that it applies to wildly discontinuous functions, for which more primitive methods give no information at all.

The Denjoy-Young-Saks theorem (222L) is one of the starting points of a flourishing theory of 'typical' phenomena in real analysis. It is easy to build a function f with any prescribed set of values for $(\overline{D}^+f)(0)$, $(\underline{D}^+f)(0)$, $(\overline{D}^-f)(0)$ and $(\underline{D}^-f)(0)$ (subject, of course, to the requirements $(\underline{D}^+f)(0)\le(\overline{D}^+f)(0)$ and $(\underline{D}^-f)(0)\le(\overline{D}^-f)(0)$). But 222L tells us that such combinations as $(\underline{D}^-f)(x)=(\underline{D}^+f)(x)=\infty$ (what we might call '$f'(x)=\infty$') can occur only on negligible sets. The four easily realized possibilities in 222L (see 222Ye) are the only ones which can appear at points which are 'typical' for the given function, from the point of view of Lebesgue measure. For a monotonic function, 222A tells us more: at 'typical' points for a monotonic function, the function is actually differentiable. In the next section we shall see some more ways of generating negligible and conegligible sets from a given set or function, leading to further refinements of the idea.

223 Lebesgue's density theorems

I now turn to a group of results which may be thought of as corollaries of Theorem 222E, but which also have a vigorous life of their own, including the possibility of significant generalizations which will be treated in Chapter 26. The idea is that any measurable function f on $\mathbb{R}$ is almost everywhere 'continuous' in a variety of very weak senses; for almost every x, the value $f(x)$ is determined by the behaviour of f near x, in the sense that $f(y) \doteq f(x)$ for 'most' y near x. I should perhaps say that while I recommend this work as a preparation for Chapter 26, and I also rely on it in Chapter 28, I shall not refer to it again in the present chapter, so that readers in a hurry to characterize indefinite integrals may proceed directly to §224.

223A Lebesgue's Density Theorem: integral form Let I be an interval in $\mathbb{R}$, and let f be a real-valued function which is integrable over I. Then

$$f(x) = \lim_{h \downarrow 0} \frac{1}{h} \int_x^{x+h} f = \lim_{h \downarrow 0} \frac{1}{h} \int_{x-h}^{x} f = \lim_{h \downarrow 0} \frac{1}{2h} \int_{x-h}^{x+h} f$$

for almost every $x \in I$.

proof Setting $F(x) = \int_{I \cap]-\infty, x[} f$, we know from 222G that

$$\begin{aligned} f(x) = F'(x) &= \lim_{h \downarrow 0} \frac{1}{h}(F(x+h) - F(x)) = \lim_{h \downarrow 0} \frac{1}{h} \int_x^{x+h} f \\ &= \lim_{h \downarrow 0} \frac{1}{h}(F(x) - F(x-h)) = \lim_{h \downarrow 0} \frac{1}{h} \int_{x-h}^{x} f \\ &= \lim_{h \downarrow 0} \frac{1}{2h}(F(x+h) - F(x-h)) = \lim_{h \downarrow 0} \frac{1}{2h} \int_{x-h}^{x+h} f \end{aligned}$$

for almost every $x \in I$.

223B Corollary Let $E \subseteq \mathbb{R}$ be a measurable set. Then

$$\lim_{h \downarrow 0} \frac{1}{2h} \mu(E \cap [x-h, x+h]) = 1 \text{ for almost every } x \in E,$$

$$\lim_{h \downarrow 0} \frac{1}{2h} \mu(E \cap [x-h, x+h]) = 0 \text{ for almost every } x \in \mathbb{R} \setminus E.$$

proof Take $n \in \mathbb{N}$. Applying 223A to $f = \chi(E \cap [-n, n])$, we see that

$$\lim_{h \downarrow 0} \frac{1}{2h} \int_{x-h}^{x+h} f = \lim_{h \downarrow 0} \frac{1}{2h} \mu(E \cap [x-h, x+h])$$

whenever $x \in]-n, n[$ and either limit exists, so that

$$\lim_{h \downarrow 0} \frac{1}{2h} \mu(E \cap [x-h, x+h]) = 1 \text{ for almost every } x \in E \cap [-n, n],$$

$$\lim_{h \downarrow 0} \frac{1}{2h} \mu(E \cap [x-h, x+h]) = 0 \text{ for almost every } x \in [-n, n] \setminus E.$$

As n is arbitrary, we have the result.

Remark For a measurable set $E \subseteq \mathbb{R}$, a point x such that $\lim_{h \downarrow 0} \frac{1}{2h} \mu(E \cap [x-h, x+h]) = 1$ is sometimes called a **density point** of E.

223C Corollary Let f be a measurable real-valued function defined almost everywhere in $\mathbb{R}$. Then for almost every $x \in \mathbb{R}$,

$$\lim_{h \downarrow 0} \frac{1}{2h} \mu\{y : y \in \operatorname{dom} f,\ |y - x| \le h,\ |f(y) - f(x)| \le \epsilon\} = 1,$$

$$\lim_{h \downarrow 0} \frac{1}{2h} \mu\{y : y \in \operatorname{dom} f,\ |y - x| \le h,\ |f(y) - f(x)| \ge \epsilon\} = 0$$

for every $\epsilon > 0$.

proof For q, $q' \in \mathbb{Q}$, set

$$D_{qq'} = \{x : x \in \operatorname{dom} f,\, q \le f(x) < q'\},$$

so that $D_{qq'}$ is measurable,

$$C_{qq'} = \{x : x \in D_{qq'},\, \lim_{h\downarrow 0} \frac{1}{2h}\mu(D_{qq'} \cap [x-h, x+h]) = 1\},$$

so that $D_{qq'} \setminus C_{qq'}$ is negligible, by 223B; now set

$$C = \operatorname{dom} f \setminus \bigcup_{q,q'\in\mathbb{Q}}(D_{qq'} \setminus C_{qq'}),$$

so that C is conegligible. If $x \in C$ and $\epsilon > 0$, then there are q, $q' \in \mathbb{Q}$ such that $f(x) - \epsilon \le q \le f(x) < q' \le f(x) + \epsilon$, so that x belongs to $D_{qq'}$ and therefore to $C_{qq'}$, and now

$$\begin{aligned}\liminf_{h\downarrow 0} \frac{1}{2h}\mu\{y : y \in \operatorname{dom} f \cap [x-h, x+h],\, |f(y) - f(x)| \le \epsilon\}\\ \ge \liminf_{h\downarrow 0} \frac{1}{2h}\mu(D_{qq'} \cap [x-h, x+h]) = 1,\end{aligned}$$

so

$$\lim_{h\downarrow 0} \frac{1}{2h}\mu\{y : y \in \operatorname{dom} f \cap [x-h, x+h],\, |f(y) - f(x)| \le \epsilon\} = 1.$$

It follows at once that

$$\lim_{h\downarrow 0} \frac{1}{2h}\mu\{y : y \in \operatorname{dom} f \cap [x-h, x+h],\, |f(y) - f(x)| > \epsilon\} = 0$$

for almost every x; but since ϵ is arbitrary, this is also true of $\frac{1}{2}\epsilon$, so in fact

$$\lim_{h\downarrow 0} \frac{1}{2h}\mu\{y : y \in \operatorname{dom} f \cap [x-h, x+h],\, |f(y) - f(x)| \ge \epsilon\} = 0$$

for almost every x.

223D Theorem Let I be an interval in $\mathbb{R}$, and let f be a real-valued function which is integrable over I. Then

$$\lim_{h\downarrow 0} \frac{1}{2h}\int_{x-h}^{x+h} |f(y) - f(x)|dy = 0$$

for almost every $x \in I$.

proof (a) Suppose first that I is a bounded open interval $]a, b[$. For each $q \in \mathbb{Q}$, set $g_q(x) = |f(x) - q|$ for $x \in I \cap \operatorname{dom} f$; then g is integrable over I, and

$$\lim_{h\downarrow 0} \frac{1}{2h}\int_{x-h}^{x+h} g_q = g_q(x)$$

for almost every $x \in I$, by 223A. Setting

$$E_q = \{x : x \in I \cap \operatorname{dom} f,\, \lim_{h\downarrow 0} \frac{1}{2h}\int_{x-h}^{x+h} g_q = g_q(x)\},$$

we have $I \setminus E_q$ negligible, so $I \setminus E$ is negligible, where $E = \bigcap_{q\in\mathbb{Q}} E_q$. Now

$$\lim_{h\downarrow 0} \frac{1}{2h}\int_{x-h}^{x+h} |f(y) - f(x)|dy = 0$$

for every $x \in E$. **P** Take $x \in E$ and $\epsilon > 0$. Then there is a $q \in \mathbb{Q}$ such that $|f(x) - q| \le \epsilon$, so that

$$|f(y) - f(x)| \le |f(y) - q| + \epsilon = g_q(y) + \epsilon$$

for every $y \in I \cap \operatorname{dom} f$, and

$$\begin{aligned}\limsup_{h\downarrow 0} \frac{1}{2h}\int_{x-h}^{x+h} |f(y) - f(x)|dy &\le \limsup_{h\downarrow 0} \frac{1}{2h}\int_{x-h}^{x+h} g_q(y) + \epsilon\, dy\\ &= \epsilon + g_q(x) \le 2\epsilon.\end{aligned}$$

As ϵ is arbitrary,

$$\lim_{h\downarrow 0}\frac{1}{2h}\int_{x-h}^{x+h}|f(y)-f(x)|dy=0,$$

as required. **Q**

(b) If I is an unbounded open interval, apply (a) to the intervals $I_n=I\cap\,]-n,n[$ to see that the limit is zero almost everywhere in every I_n, and therefore on I. If I is an arbitrary interval, note that it differs by at most two points from an open interval, and that since we are looking only for something to happen almost everywhere we can ignore these points.

Remark The set

$$\{x:x\in\operatorname{dom}f,\ \lim_{h\downarrow 0}\frac{1}{2h}\int_{x-h}^{x+h}|f(y)-f(x)|dy=0\}$$

is sometimes called the **Lebesgue set** of f.

223E Complex-valued functions I have expressed the results above in terms of real-valued functions, this being the most natural vehicle for the ideas. However there are applications of great importance in which the functions involved are complex-valued, so I spell out the relevant statements here. In all cases the proof is elementary, being nothing more than applying the corresponding result (223A, 223C or 223D) to the real and imaginary parts of the function f.

(a) Let I be an interval in $\mathbb{R}$, and let f be a complex-valued function which is integrable over I. Then

$$f(x)=\lim_{h\downarrow 0}\frac{1}{h}\int_x^{x+h}f=\lim_{h\downarrow 0}\frac{1}{h}\int_{x-h}^{x}f=\lim_{h\downarrow 0}\frac{1}{2h}\int_{x-h}^{x+h}f$$

for almost every $x\in I$.

(b) Let f be a measurable complex-valued function defined almost everywhere in $\mathbb{R}$. Then for almost every $x\in\mathbb{R}$,

$$\lim_{h\downarrow 0}\frac{1}{2h}\mu\{y:y\in\operatorname{dom}f,\ |y-x|\le h,\ |f(y)-f(x)|\ge\epsilon\}=0$$

for every $\epsilon>0$.

(c) Let I be an interval in $\mathbb{R}$, and let f be a complex-valued function which is integrable over I. Then

$$\lim_{h\downarrow 0}\frac{1}{2h}\int_{x-h}^{x+h}|f(y)-f(x)|dy=0$$

for almost every $x\in I$.

223X Basic exercises >(a) Let $E\subseteq[0,1]$ be a measurable set for which there is an $\alpha>0$ such that $\mu(E\cap[a,b])\ge\alpha(b-a)$ whenever $0\le a\le b\le 1$. Show that $\mu E=1$.

(b) Let $A\subseteq\mathbb{R}$ be any set. Show that $\lim_{h\downarrow 0}\frac{1}{2h}\mu^*(A\cap[x-h,x+h])=1$ for almost every $x\in A$. (*Hint*: apply 223B to a measurable envelope E of A.)

(c) Let E, $F\subseteq\mathbb{R}$ be measurable sets, and $x\in\mathbb{R}$ a point which is a density point of both. Show that x is a density point of $E\cap F$.

(d) Let $E\subseteq\mathbb{R}$ be a non-negligible measurable set. Show that for any $n\in\mathbb{N}$ there is a $\delta>0$ such that $\bigcap_{i\le n}E+x_i$ is non-empty whenever $x_0,\dots,x_n\in\mathbb{R}$ are such that $|x_i-x_j|\le\delta$ for all i, $j\le n$. (*Hint*: find a non-trivial interval I such that $\mu(E\cap I)>\frac{n}{n+1}\mu I$.)

(e) Let f be any real-valued function defined almost everywhere in $\mathbb{R}$. Show that $\lim_{h\downarrow 0}\frac{1}{2h}\mu^*\{y:y\in\operatorname{dom}f,\ |y-x|\le h,\ |f(y)-f(x)|\le\epsilon\}=1$ for almost every $x\in\mathbb{R}$. (*Hint*: use the argument of 223C, but with 223Xb in place of 223B.)

>(f) Let I be an interval in $\mathbb{R}$, and let f be a real-valued function which is integrable over I. Show that $\lim_{h\downarrow 0}\frac{1}{h}\int_x^{x+h}|f(y)-f(x)|dy=0$ for almost every $x\in I$.

(g) Let E, $F \subseteq \mathbb{R}$ be measurable sets, and suppose that F is bounded and of non-zero measure. Let $x \in \mathbb{R}$ be such that $\lim_{h\downarrow 0} \frac{1}{2h}\mu(E \cap [x-h, x+h]) = 1$. Show that $\lim_{h\downarrow 0} \frac{\mu(E\cap(x+hF))}{h\,\mu F} = 1$. (*Hint*: it helps to know that $\mu(hF) = h\mu F$ (134Ya, 263A). Show that if $F \subseteq [-M, M]$, then

$$\frac{1}{2hM}\mu(E \cap [x-hM, x+hM]) \le 1 - \frac{\mu F}{2M}\bigl(1 - \frac{\mu(E\cap(x+hF))}{h\,\mu F}\bigr).)$$

(Compare 223Ya.)

(h) Let f be a real-valued function which is integrable over $\mathbb{R}$, and E be the Lebesgue set of f. Show that $\lim_{h\downarrow 0} \frac{1}{2h}\int_{x-h}^{x+h} |f(t) - c|dt = |f(x) - c|$ for every $x \in E$ and $c \in \mathbb{R}$.

(i) Let f be an integrable real-valued function defined almost everywhere in $\mathbb{R}$. Let $x \in \operatorname{dom} f$ be such that $\lim_{n\to\infty} \frac{n}{2}\int_{x-1/n}^{x+1/n} |f(y) - f(x)| = 0$. Show that x belongs to the Lebesgue set of f.

(j) Let f be an integrable real-valued function defined almost everywhere in $\mathbb{R}$, and x any point of the Lebesgue set of f. Show that for every $\epsilon > 0$ there is a $\delta > 0$ such that whenever I is a non-trivial interval and $x \in I \subseteq [x-\delta, x+\delta]$, then $|f(x) - \frac{1}{\mu I}\int_I f| \le \epsilon$.

223Y Further exercises (a) Let E, $F \subseteq \mathbb{R}$ be measurable sets, and suppose that $0 < \mu F < \infty$. Let $x \in \mathbb{R}$ be such that $\lim_{h\downarrow 0} \frac{1}{2h}\mu(E \cap [x-h, x+h]) = 1$. Show that

$$\lim_{h\downarrow 0} \frac{\mu(E\cap(x+hF))}{h\,\mu F} = 1.$$

(*Hint*: apply 223Xg to sets of the form $F \cap [-M, M]$.)

(b) Let $\mathfrak{T}$ be the family of measurable sets $G \subseteq \mathbb{R}$ such that every point of G is a density point of G. (i) Show that $\mathfrak{T}$ is a topology on $\mathbb{R}$. (*Hint*: take $\mathcal{G} \subseteq \mathfrak{T}$. By 215B(iv) there is a countable $\mathcal{G}_0 \subseteq \mathcal{G}$ such that $\mu(G \setminus \bigcup \mathcal{G}_0) = 0$ for every $G \in \mathcal{G}$. Show that

$$\bigcup \mathcal{G} \subseteq \{x : \limsup_{h\downarrow 0} \frac{1}{2h}\mu(\bigcup \mathcal{G}_0 \cap [x-h, x+h]) > 0\},$$

so that $\mu(\bigcup \mathcal{G} \setminus \bigcup \mathcal{G}_0) = 0$.) (ii) Show that a function $f : \mathbb{R} \to \mathbb{R}$ is measurable iff it is $\mathfrak{T}$-continuous at almost every $x \in \mathbb{R}$. ($\mathfrak{T}$ is the **density topology** on $\mathbb{R}$. See 414P in Volume 4.)

(c) Show that if $f : [a, b] \to \mathbb{R}$ is bounded and continuous for the density topology on $\mathbb{R}$, then $f(x) = \frac{d}{dx}\int_a^x f$ for every $x \in \,]a, b[$.

(d) Show that a function $f : \mathbb{R} \to \mathbb{R}$ is continuous for the density topology at $x \in \mathbb{R}$ iff $\lim_{h\downarrow 0} \frac{1}{2h}\mu^*\{y : y \in [x-h, x+h], |f(y) - f(x)| \ge \epsilon\} = 0$ for every $\epsilon > 0$.

(e) A set $A \subseteq \mathbb{R}$ is **porous** at a point $x \in \mathbb{R}$ if $\limsup_{y\to x} \frac{\rho(y,A)}{|y-x|} > 0$, where $\rho(y, A) = \inf_{a\in A} |y - a|$. (Take $\rho(y, \emptyset) = \infty$.) Show that if A is porous at every $x \in A$ then A is negligible.

(f) For a measurable set $E \subseteq \mathbb{R}$ write int^*E for the set of its density points. Show that if E, $F \subseteq \mathbb{R}$ are measurable then (i) $\operatorname{int}^*(E\cap F) = \operatorname{int}^*E \cap \operatorname{int}^*F$ (ii) $\operatorname{int}^*E \subseteq \operatorname{int}^*F$ iff $\mu(E \setminus F) = 0$ (iii) $\mu(E \triangle \operatorname{int}^*E) = 0$ (iv) $\operatorname{int}^*(\operatorname{int}^*E) = \operatorname{int}^*E$ (v) for every compact set $K \subseteq \operatorname{int}^*E$ there is a compact set $L \subseteq K \cup E$ such that $K \subseteq \operatorname{int}^*L$.

(g) Let f be an integrable real-valued function defined almost everywhere in $\mathbb{R}$, and x any point of the Lebesgue set of f. Show that for every $\epsilon > 0$ there is a $\delta > 0$ such that $|f(x)\int g - \int f \times g| \le \epsilon \int g$ whenever $g : \mathbb{R} \to [0, \infty[$ is such that g is non-decreasing on $]-\infty, x]$, non-increasing on $[x, \infty[$ and zero outside $[x-\delta, x+\delta]$. (*Hint*: express g as a limit almost everywhere of functions of the form $\frac{g(x)}{n+1}\sum_{i=0}^n \chi\,]a_i, b_i[$, where $x - \delta \le a_0 \le \ldots \le a_n \le x \le b_n \le \ldots \le b_0 \le x + \delta$.)

(h) For each integrable real-valued function f defined almost everywhere in $\mathbb{R}$, let E_f be the Lebesgue set of f. Show that $E_f \cap E_g \subseteq E_{f+g}$, $E_f \subseteq E_{|f|}$ for all integrable f, g.

223 Notes and comments The results of this section can be thought of as saying that a measurable function is in some sense 'almost continuous'; indeed, 223Yb is an attempt to make this notion precise. For an integrable function we have stronger results, of which the furthest-reaching seems to be 223D/223Ec.

There are r-dimensional versions of all these theorems, using balls centered on x in place of intervals $[x-h,x+h]$; I give these in 261C-261E. A new idea is needed for the r-dimensional version of Lebesgue's density theorem (261C), but the rest of the generalization is straightforward. A less natural, and less important, extension, also in §261, involves functions defined on non-measurable sets (compare 223Xb-223Xe).

224 Functions of bounded variation

I turn now to the second of the two problems to which this chapter is devoted: the identification of those real functions which are indefinite integrals. I take the opportunity to offer a brief introduction to the theory of functions of bounded variation, which are interesting in themselves and will be important in Chapter 28. I give the basic characterization of these functions as differences of monotonic functions (224D), with a representative sample of their elementary properties.

224A Definition Let f be a real-valued function and D a subset of $\mathbb{R}$. I define $\operatorname{Var}_D(f)$, the **(total) variation of** f **on** D, as follows. If $D\cap\operatorname{dom} f=\emptyset$, $\operatorname{Var}_D(f)=0$. Otherwise, $\operatorname{Var}_D(f)$ is

$$\sup\{\textstyle\sum_{i=1}^n|f(a_i)-f(a_{i-1})|:a_0,a_1,\dots,a_n\in D\cap\operatorname{dom} f,\ a_0\le a_1\le\dots\le a_n\},$$

allowing $\operatorname{Var}_D(f)=\infty$. If $\operatorname{Var}_D(f)$ is finite, we say that f is **of bounded variation** on D. If the context seems clear, I may write $\operatorname{Var} f$ for $\operatorname{Var}_{\operatorname{dom} f}(f)$, and say that f is simply '**of bounded variation**' if this is finite.

224B Remarks (a) In the present chapter, we shall virtually exclusively be concerned with the case in which D is a bounded closed interval included in $\operatorname{dom} f$. The general formulation will be useful for some technical questions arising in Chapter 28; but if it makes you more comfortable, you will lose nothing by supposing for the moment that D is an interval.

(b) Clearly

$$\operatorname{Var}_D(f)=\operatorname{Var}_{D\cap\operatorname{dom} f}(f)=\operatorname{Var}(f\restriction D)$$

for all D, f.

224C Proposition (a) If f, g are two real-valued functions and $D\subseteq\mathbb{R}$, then

$$\operatorname{Var}_D(f+g)\le\operatorname{Var}_D(f)+\operatorname{Var}_D(g).$$

(b) If f is a real-valued function, $D\subseteq\mathbb{R}$ and $c\in\mathbb{R}$ then $\operatorname{Var}_D(cf)=|c|\operatorname{Var}_D(f)$.

(c) If f is a real-valued function, $D\subseteq\mathbb{R}$ and $x\in\mathbb{R}$ then

$$\operatorname{Var}_D(f)\ge\operatorname{Var}_{D\cap]-\infty,x]}(f)+\operatorname{Var}_{D\cap[x,\infty[}(f),$$

with equality if $x\in D\cap\operatorname{dom} f$.

(d) If f is a real-valued function and $D\subseteq D'\subseteq\mathbb{R}$ then $\operatorname{Var}_D(f)\le\operatorname{Var}_{D'}(f)$.

(e) If f is a real-valued function and $D\subseteq\mathbb{R}$, then $|f(x)-f(y)|\le\operatorname{Var}_D(f)$ for all x, $y\in D\cap\operatorname{dom} f$; so if f is of bounded variation on D then f is bounded on $D\cap\operatorname{dom} f$ and (if $D\cap\operatorname{dom} f\ne\emptyset$)

$$\sup_{y\in D\cap\operatorname{dom} f}|f(y)|\le|f(x)|+\operatorname{Var}_D(f)$$

for every $x\in D\cap\operatorname{dom} f$.

(f) If f is a monotonic real-valued function and $D\subseteq\mathbb{R}$ meets $\operatorname{dom} f$, then $\operatorname{Var}_D(f)=\sup_{x\in D\cap\operatorname{dom} f}f(x)-\inf_{x\in D\cap\operatorname{dom} f}f(x)$.

proof (a) If $D\cap\operatorname{dom}(f+g)=\emptyset$ this is trivial, because $\operatorname{Var}_D(f)$ and $\operatorname{Var}_D(g)$ are surely non-negative. Otherwise, if $a_0\le\dots\le a_n$ in $D\cap\operatorname{dom}(f+g)$, then

$$\begin{aligned}\sum_{i=1}^n|(f+g)(a_i)-(f+g)(a_{i-1})|&\le\sum_{i=1}^n|f(a_i)-f(a_{i-1})|+\sum_{i=1}^n|g(a_i)-g(a_{i-1})|\\&\le\operatorname{Var}_D(f)+\operatorname{Var}_D(g);\end{aligned}$$

as $a_0,\dots,a_n$ are arbitrary, $\operatorname{Var}_D(f+g)\le\operatorname{Var}_D(f)+\operatorname{Var}_D(g)$.

(b)

$$\sum_{i=1}^{n}|(cf)(a_i)-(cf)(a_{i-1})|=|c|\sum_{i=1}^{n}|f(a_i)-f(a_{i-1})|$$

whenever $a_0\le\ldots\le a_n$ in $D\cap\operatorname{dom}f$.

(c)(i) If either $D\cap\,]-\infty,x]\cap\operatorname{dom}f$ or $D\cap[x,\infty[\,\cap\operatorname{dom}f$ is empty, this is trivial. If $a_0\le\ldots\le a_m$ in $D\cap\,]-\infty,x]\cap\operatorname{dom}f$, $b_0\le\ldots\le b_n$ in $D\cap[x,\infty[\,\cap\operatorname{dom}f$, then

$$\begin{aligned}\sum_{i=1}^{m}|f(a_i)-f(a_{i-1})|+\sum_{j=1}^{n}|f(b_i)-f(b_{i-1})|&\le\sum_{i=1}^{m+n+1}|f(a_i)-f(a_{i-1})|\\&\le\operatorname{Var}_{[a,b]}(f),\end{aligned}$$

if we write $a_i=b_{i-m-1}$ for $m+1\le i\le m+n+1$. So

$$\operatorname{Var}_{D\cap]-\infty,x]}(f)+\operatorname{Var}_{D\cap[x,\infty[}(f)\le\operatorname{Var}_D(f).$$

(ii) Now suppose that $x\in D\cap\operatorname{dom}f$. If $a_0\le\ldots\le a_n$ in $D\cap\operatorname{dom}f$, and $a_0\le x\le a_n$, let k be such that $x\in[a_{k-1},a_k]$; then

$$\begin{aligned}\sum_{i=1}^{n}|f(a_i)-f(a_{i-1})|\le\sum_{i=1}^{k-1}|f(a_i)-f(a_{i-1})|&+|f(x)-f(a_{k-1})|\\&+|f(a_k)-f(x)|+\sum_{i=k+1}^{n}|f(a_i)-f(a_{i-1})|\\\le\operatorname{Var}_{D\cap]-\infty,x]}(f)+\operatorname{Var}_{D\cap[x,\infty[}(f)&\end{aligned}$$

(counting empty sums $\sum_{i=1}^{0}$, $\sum_{i=n+1}^{n}$ as 0). If $x\le a_0$ then $\sum_{i=1}^{n}|f(a_i)-f(a_{i-1})|\le\operatorname{Var}_{D\cap[x,\infty[}(f)$; if $x\ge a_n$ then $\sum_{i=1}^{n}|f(a_i)-f(a_{i-1})|\le\operatorname{Var}_{D\cap]-\infty,x]}(f)$. Thus

$$\sum_{i=1}^{n}|f(a_i)-f(a_{i-1})|\le\operatorname{Var}_{D\cap]-\infty,x]}(f)+\operatorname{Var}_{D\cap[x,\infty[}(f)$$

in all cases; as $a_0,\ldots,a_n$ are arbitrary,

$$\operatorname{Var}_D(f)\le\operatorname{Var}_{D\cap]-\infty,x]}(f)+\operatorname{Var}_{D\cap[x,\infty[}(f).$$

So the two sides are equal.

(d) is trivial.

(e) If x, $y\in D\cap\operatorname{dom}f$ and $x\le y$ then

$$|f(x)-f(y)|=|f(y)-f(x)|\le\operatorname{Var}_D(f)$$

by the definition of Var_D; and the same is true if $y\le x$. So of course $|f(y)|\le|f(x)|+\operatorname{Var}_D(f)$.

(f) If f is non-decreasing, then

$$\begin{aligned}\operatorname{Var}_D(f)&=\sup\{\sum_{i=1}^{n}|f(a_i)-f(a_{i-1})|:a_0,a_1,\ldots,a_n\in D\cap\operatorname{dom}f,\ a_0\le a_1\le\ldots\le a_n\}\\&=\sup\{\sum_{i=1}^{n}f(a_i)-f(a_{i-1}):a_0,a_1,\ldots,a_n\in D\cap\operatorname{dom}f,\ a_0\le a_1\le\ldots\le a_n\}\\&=\sup\{f(b)-f(a):a,\ b\in D\cap\operatorname{dom}f,\ a\le b\}\\&=\sup_{b\in D\cap\operatorname{dom}f}f(b)-\inf_{a\in D\cap\operatorname{dom}f}f(a).\end{aligned}$$

If f is non-increasing then

$$\begin{aligned}
\mathrm{Var}_D(f) &= \sup\{\sum_{i=1}^{n} |f(a_i) - f(a_{i-1})| : a_0, a_1, \dots , a_n \in D \cap \operatorname{dom} f,\ a_0 \le a_1 \le \dots \le a_n\} \\
&= \sup\{\sum_{i=1}^{n} f(a_{i-1}) - f(a_i) : a_0, a_1, \dots , a_n \in D \cap \operatorname{dom} f,\ a_0 \le a_1 \le \dots \le a_n\} \\
&= \sup\{f(a) - f(b) : a,\, b \in D \cap \operatorname{dom} f,\ a \le b\} \\
&= \sup_{a\in D\cap\operatorname{dom} f} f(a) - \inf_{b\in D\cap\operatorname{dom} f} f(b).
\end{aligned}$$

224D Theorem For any real-valued function f and any set $D \subseteq \mathbb{R}$, the following are equiveridical:

(i) there are two bounded non-decreasing functions $f_1,\ f_2 : \mathbb{R} \to \mathbb{R}$ such that $f = f_1 - f_2$ on $D \cap \operatorname{dom} f$;

(ii) f is of bounded variation on D;

(iii) there are bounded non-decreasing functions $f_1,\ f_2 : \mathbb{R} \to \mathbb{R}$ such that $f = f_1 - f_2$ on $D \cap \operatorname{dom} f$ and $\mathrm{Var}_D(f) = \mathrm{Var}\, f_1 + \mathrm{Var}\, f_2$.

proof (i)⇒(ii) If $f : \mathbb{R} \to \mathbb{R}$ is bounded and non-decreasing, then $\mathrm{Var}\, f = \sup_{x\in\mathbb{R}} f(x) - \inf_{x\in\mathbb{R}} f(x)$ is finite. So if f agrees on $D \cap \operatorname{dom} f$ with $f_1 - f_2$ where f_1 and f_2 are bounded and non-decreasing, then

$$\begin{aligned}
\mathrm{Var}_D(f) = \mathrm{Var}_{D\cap\operatorname{dom} f}(f) &\le \mathrm{Var}_{D\cap\operatorname{dom} f}(f_1) + \mathrm{Var}_{D\cap\operatorname{dom} f}(f_2) \\
&\le \mathrm{Var}\, f_1 + \mathrm{Var}\, f_2 < \infty,
\end{aligned}$$

using (a), (b) and (d) of 224C.

(ii)⇒(iii) Suppose that f is of bounded variation on D. Set $D' = D \cap \operatorname{dom} f$. If $D' = \emptyset$ we can take both f_j to be the zero function, so henceforth suppose that $D' \ne \emptyset$. Write

$$g(x) = \mathrm{Var}_{D\cap]-\infty,x]}(f)$$

for $x \in D'$. Then $g_1 = g + f$ and $g_2 = g - f$ are both non-decreasing. **P** If $a,\ b \in D'$ and $a \le b$, then

$$g(b) = g(a) + \mathrm{Var}_{D\cap[a,b]}(f) \ge g(a) + |f(b) - f(a)|.$$

So

$$g_1(b) - g_1(a) = g(b) - g(a) + f(b) - f(a), \quad g_2(b) - g_2(a) = g(b) - g(a) - f(b) + f(a)$$

are both non-negative. **Q**

Now there are non-decreasing functions $h_1,\ h_2 : \mathbb{R} \to \mathbb{R}$, extending $g_1,\ g_2$ respectively, such that $\mathrm{Var}\, h_j = \mathrm{Var}\, g_j$ for both j. **P** f is bounded on D, by 224Ce, and g is bounded just because $\mathrm{Var}_D(f) < \infty$, so that g_j is bounded. Set $c_j = \inf_{x\in D'} g_j(x)$ and

$$h_j(x) = \sup(\{c_j\} \cup \{g_j(y) : y \in D',\ y \le x\})$$

for every $x \in \mathbb{R}$; this works. **Q** Observe that for $x \in D'$,

$$h_1(x) + h_2(x) = g_1(x) + g_2(x) = g(x) + f(x) + g(x) - f(x) = 2g(x),$$

$$h_1(x) - h_2(x) = 2f(x).$$

Now, because g_1 and g_2 are non-decreasing,

$$\sup_{x\in D'} g_1(x) + \sup_{x\in D'} g_2(x) = \sup_{x\in D'} g_1(x) + g_2(x) = 2 \sup_{x\in D'} g(x),$$

$$\inf_{x\in D'} g_1(x) + \inf_{x\in D'} g_2(x) = \inf_{x\in D'} g_1(x) + g_2(x) = 2 \inf_{x\in D'} g(x) \ge 0.$$

But this means that

$$\mathrm{Var}\, h_1 + \mathrm{Var}\, h_2 = \mathrm{Var}\, g_1 + \mathrm{Var}\, g_2 = 2\, \mathrm{Var}\, g \le 2\, \mathrm{Var}_D(f),$$

using 224Cf three times. So if we set $f_j(x) = \frac{1}{2}h_j(x)$ for $j \in \{1, 2\}$ and $x \in \mathbb{R}$, we shall have non-decreasing functions such that

$$f_1(x) - f_2(x) = f(x) \text{ for } x \in D', \quad \mathrm{Var}\, f_1 + \mathrm{Var}\, f_2 = \frac{1}{2}\mathrm{Var}\, h_1 + \frac{1}{2}\mathrm{Var}\, h_2 \le \mathrm{Var}_D(f).$$

Since we surely also have

$$\mathrm{Var}_D(f) \le \mathrm{Var}_D(f_1) + \mathrm{Var}_D(f_2) \le \mathrm{Var}\, f_1 + \mathrm{Var}\, f_2,$$

we see that $\mathrm{Var}_D(f) = \mathrm{Var}\, f_1 + \mathrm{Var}\, f_2$, and (iii) is true.

(iii)⇒(i) is trivial.

224E Corollary Let f be a real-valued function and D any subset of $\mathbb{R}$. If f is of bounded variation on D, then

$$\lim_{x\downarrow a} \mathrm{Var}_{D\cap]a,x]}(f) = \lim_{x\uparrow a} \mathrm{Var}_{D\cap[x,a[}(f) = 0$$

for every $a \in \mathbb{R}$, and

$$\lim_{a\to-\infty} \mathrm{Var}_{D\cap]-\infty,a]}(f) = \lim_{a\to\infty} \mathrm{Var}_{D\cap[a,\infty[}(f) = 0.$$

proof (a) Consider first the case in which $D = \mathrm{dom}\, f = \mathbb{R}$ and f is a bounded non-decreasing function. Then

$$\mathrm{Var}_{D\cap]a,x]}(f) = \sup_{y\in]a,x]} f(x) - f(y) = f(x) - \inf_{y>a} f(y) = f(x) - \lim_{y\downarrow a} f(y),$$

so of course

$$\lim_{x\downarrow a} \mathrm{Var}_{D\cap]a,x]}(f) = \lim_{x\downarrow a} f(x) - \lim_{y\downarrow a} f(y) = 0.$$

In the same way

$$\lim_{x\uparrow a} \mathrm{Var}_{D\cap[x,a[}(f) = \lim_{y\uparrow a} f(y) - \lim_{x\uparrow a} f(x) = 0,$$

$$\lim_{a\to-\infty} \mathrm{Var}_{D\cap]-\infty,a]}(f) = \lim_{a\to-\infty} f(a) - \lim_{y\to-\infty} f(y) = 0,$$

$$\lim_{a\to\infty} \mathrm{Var}_{D\cap[a,\infty[}(f) = \lim_{y\to\infty} f(y) - \lim_{a\to\infty} f(a) = 0.$$

(b) For the general case, define f_1, f_2 from f and D as in 224D. Then for every interval I we have

$$\mathrm{Var}_{D\cap I}(f) \le \mathrm{Var}_I(f_1) + \mathrm{Var}_I(f_2),$$

so the results for f follow from those for f_1 and f_2 as established in part (a) of the proof.

224F Corollary Let f be a real-valued function of bounded variation on $[a, b]$, where $a < b$. If $\mathrm{dom}\, f$ meets every interval $]a, a + \delta]$ with $\delta > 0$, then

$$\lim_{t\in\mathrm{dom}\, f, t\downarrow a} f(t)$$

is defined in $\mathbb{R}$. If $\mathrm{dom}\, f$ meets $[b - \delta, b[$ for every $\delta > 0$, then

$$\lim_{t\in\mathrm{dom}\, f, t\uparrow b} f(t)$$

is defined in $\mathbb{R}$.

proof Let f_1, $f_2 : \mathbb{R} \to \mathbb{R}$ be non-decreasing functions such that $f = f_1 - f_2$ on $[a, b] \cap \mathrm{dom}\, f$. Then

$$\lim_{t\in\mathrm{dom}\, f, t\downarrow a} f(t) = \lim_{t\downarrow a} f_1(t) - \lim_{t\downarrow a} f_2(t) = \inf_{t>a} f_1(t) - \inf_{t>a} f_2(t),$$

$$\lim_{t\in\mathrm{dom}\, f, t\uparrow b} f(t) = \lim_{t\uparrow b} f_1(t) - \lim_{t\uparrow b} f_2(t) = \sup_{t<b} f_1(t) - \sup_{t<b} f_2(t).$$

224G Corollary Let f, g be real functions and D a subset of $\mathbb{R}$. If f and g are of bounded variation on D, so is $f \times g$.

proof (a) The point is that there are *non-negative* bounded non-decreasing functions f_1, $f_2 : \mathbb{R} \to \mathbb{R}$ such that $f = f_1 - f_2$ on $D \cap \mathrm{dom}\, f$. **P** We know that there are bounded non-decreasing h_1, h_2 such that $f = h_1 - h_2$ on $D \cap \mathrm{dom}\, f$. Set $\gamma_i = \inf_{x\in\mathbb{R}} h_i(x)$ for $i = 1, 2$,

$$\beta_1 = \max(\gamma_1 - \gamma_2, 0), \quad \beta_2 = \max(\gamma_2 - \gamma_1, 0),$$

$$f_1 = h_1 - \gamma_1 + \beta_1, \quad f_2 = h_1 - \gamma_2 + \beta_2;$$

this works. **Q**

(b) Now taking similar functions g_1, g_2 such that $g = g_1 - g_2$ on $D \cap \mathrm{dom}\, g$, we have

$$f \times g = f_1 \times g_1 - f_2 \times g_1 - f_1 \times g_2 + f_2 \times g_2$$

everywhere in $D \cap \mathrm{dom}(f \times g) = D \cap \mathrm{dom}\, f \cap \mathrm{dom}\, g$; but all the $f_i \times g_j$ are bounded non-decreasing functions, so of bounded variation, and $f \times g$ must be of bounded variation on D.

224H Proposition Let $f : D \to \mathbb{R}$ be a function of bounded variation, where $D \subseteq \mathbb{R}$. Then f is continuous at all except countably many points of D.

proof For $n \ge 1$ set

$$A_n = \{x : x \in D, \text{ for every } \delta > 0 \text{ there is a } y \in D \cap [x-\delta, x+\delta] \text{ such that } |f(y) - f(x)| \ge \frac{1}{n}\}.$$

Then $\#(A_n) \le n \operatorname{Var} f$. **P?** Otherwise, we can find distinct $x_0, \dots, x_k \in A_n$ with $k+1 > n \operatorname{Var} f$. Order these so that $x_0 < x_1 < \ldots < x_k$. Set $\delta = \frac{1}{2}\min_{1 \le i \le k} x_i - x_{i-1} > 0$. For each i, there is a $y_i \in D \cap [x_i - \delta, x_i + \delta]$ such that $|f(y_i) - f(x_i)| \ge \frac{1}{n}$. Take x_i', y_i' to be x_i, y_i in order, so that $x_i' < y_i'$. Now

$$x_0' \le y_0' \le x_1' \le y_1' \le \ldots \le x_k' \le y_k',$$

and

$$\operatorname{Var} f \ge \sum_{i=0}^{k} |f(y_i') - f(x_i')| = \sum_{i=0}^{k} |f(y_i) - f(x_i)| \ge \frac{1}{n}(k+1) > \operatorname{Var} f,$$

which is impossible. **XQ**

It follows that $A = \bigcup_{n\in\mathbb{N}} A_n$ is countable, being a countable union of finite sets. But A is exactly the set of points of D at which f is not continuous.

224I Theorem Let $I \subseteq \mathbb{R}$ be an interval, and $f : I \to \mathbb{R}$ a function of bounded variation. Then f is differentiable almost everywhere in I, and f' is integrable over I, with

$$\int_I |f'| \le \operatorname{Var}_I(f).$$

proof (a) Let f_1 and f_2 be non-decreasing functions such that $f = f_1 - f_2$ everywhere in I (224D). Then f_1 and f_2 are differentiable almost everywhere (222A). At any point of I except possibly its endpoints, if any, f will be differentiable if f_1 and f_2 are, so $f'(x)$ is defined for almost every $x \in I$.

(b) Set $F(x) = \operatorname{Var}_{I\cap]-\infty,x]} f$ for $x \in \mathbb{R}$. If x, $y \in I$ and $x \le y$, then

$$F(y) - F(x) = \operatorname{Var}_{[x,y]} f \ge |f(y) - f(x)|,$$

by 224Cc; so $F'(x) \ge |f'(x)|$ whenever x is an interior point of I and both derivatives exist, which is almost everywhere. So $\int_I |f'| \le \int_I F'$. But if a, $b \in I$ and $a \le b$,

$$\int_a^b F' \le F(b) - F(a) \le F(b) \le \operatorname{Var} f.$$

Now I is expressible as $\bigcup_{n\in\mathbb{N}}[a_n, b_n]$ where $a_{n+1} \le a_n \le b_n \le b_{n+1}$ for every n. So

$$\begin{aligned}\int_I |f'| &\le \int_I F' = \int F' \times \chi I \\ &= \int \sup_{n\in\mathbb{N}} F' \times \chi[a_n, b_n] = \sup_{n\in\mathbb{N}} \int F' \times \chi[a_n, b_n]\end{aligned}$$

(by B.Levi's theorem)

$$= \sup_{n\in\mathbb{N}} \int_{a_n}^{b_n} F' \le \operatorname{Var}_I(f).$$

224J The next result is not needed in this chapter, but is one of the most useful properties of functions of bounded variation, and will be used repeatedly in Chapter 28.

Proposition Let f, g be real-valued functions defined on subsets of $\mathbb{R}$, and suppose that g is integrable over an interval $[a, b]$, where $a < b$, and f is of bounded variation on $]a, b[$ and defined almost everywhere in $]a, b[$. Then $f \times g$ is integrable over $[a, b]$, and

$$\Bigl|\int_a^b f \times g\Bigr| \le \bigl(\lim_{x\in\operatorname{dom} f, x\uparrow b} |f(x)| + \operatorname{Var}_{]a,b[}(f)\bigr) \sup_{c\in[a,b]} \Bigl|\int_a^c g\Bigr|.$$

proof (a) By 224F, $l = \lim_{x\in\operatorname{dom} f, x\uparrow b} f(x)$ is defined. Write $M = |l| + \operatorname{Var}_{]a,b[}(f)$. Note that if y is any point of $\operatorname{dom} f \cap\,]a,b[$,

$$|f(y)| \le |f(x)| + |f(x) - f(y)| \le |f(x)| + \operatorname{Var}_{]a,b[}(f) \to M$$

as $x \uparrow b$ in $\operatorname{dom} f$, so $|f(y)| \le M$. Moreover, f is measurable on $]a,b[$, because there are bounded monotonic functions f_1, $f_2 : \mathbb{R} \to \mathbb{R}$ such that $f = f_1 - f_2$ everywhere in $]a,b[\, \cap \operatorname{dom} f$. So $f \times g$ is measurable and dominated by $M|g|$, and is integrable over $]a,b[$ or $[a,b]$.

(b) For $n \in \mathbb{N}$, $k \le 2^n$ set $a_{nk} = a + 2^{-n}k(b-a)$, and for $1 \le k \le 2^n$ choose $x_{nk} \in \operatorname{dom} f \cap\,]a_{n,k-1}, a_{nk}]$. Define $f_n :]a,b] \to \mathbb{R}$ by setting $f_n(x) = f(x_{nk})$ if $1 \le k \le 2^n$ and $x \in\,]a_{n,k-1}, a_{nk}]$. Then $f(x) = \lim_{n\to\infty} f_n(x)$ whenever $x \in\,]a,b[\, \cap \operatorname{dom} f$ and f is continuous at x, which must be almost everywhere (224H). Note next that all the f_n are measurable, and that they are uniformly bounded, in modulus, by M. So $\{f_n \times g : n \in \mathbb{N}\}$ is dominated by the integrable function $M|g|$, and Lebesgue's Dominated Convergence Theorem tells us that

$$\int_a^b f \times g = \lim_{n\to\infty} \int_a^b f_n \times g.$$

(c) Fix $n \in \mathbb{N}$ for the moment. Set $K = \sup_{c\in[a,b]} |\int_a^c g|$. (Note that K is finite because $c \mapsto \int_a^c g$ is continuous.) Then

$$\begin{aligned}
|\int_a^b f_n \times g| &= |\sum_{k=1}^{2^n} \int_{a_{n,k-1}}^{a_{nk}} f_n \times g| \\
&= |\sum_{k=1}^{2^n} f(x_{nk})(\int_a^{a_{nk}} g - \int_a^{a_{n,k-1}} g)| \\
&= |\sum_{k=1}^{2^n-1} (f(x_{nk}) - f(x_{n,k+1})) \int_a^{a_{nk}} g + f(x_{n,2^n}) \int_a^b g| \\
&\le |f(x_{n,2^n})||\int_a^b g| + \sum_{k=1}^{2^n-1} |f(x_{n,k+1}) - f(x_{nk})||\int_a^{a_{nk}} g| \\
&\le (|f(x_{n,2^n})| + \operatorname{Var}_{]a,b[}(f))K \to MK
\end{aligned}$$

as $n \to \infty$.

(d) Now

$$|\int_a^b f \times g| = \lim_{n\to\infty} |\int_a^b f_n \times g| \le MK,$$

as required.

224K Complex-valued functions So far I have taken all functions to be real-valued. This is adequate for the needs of the present chapter, but in Chapter 28 we shall need to look at complex-valued functions of bounded variation, and I should perhaps spell out the (elementary) adaptations involved in the extension to the complex case.

(a) Let D be a subset of $\mathbb{R}$ and f a complex-valued function. The **variation** of f on D, $\operatorname{Var}_D(f)$, is zero if $D \cap \operatorname{dom} f = \emptyset$, and otherwise is

$$\sup\{\sum_{j=1}^n |f(a_j) - f(a_{j-1})| : a_0 \le a_1 \le \ldots \le a_n \text{ in } D \cap \operatorname{dom} f\},$$

allowing ∞. If $\operatorname{Var}_D(f)$ is finite, we say that f is **of bounded variation** on D.

(b) Just as in the real case, a complex-valued function of bounded variation must be bounded, and

$$\operatorname{Var}_D(f+g) \le \operatorname{Var}_D(f) + \operatorname{Var}_D(g),$$

$$\operatorname{Var}_D(cf) = |c| \operatorname{Var}_D(f),$$

$$\operatorname{Var}_D(f) \ge \operatorname{Var}_{D\cap]-\infty,x]}(f) + \operatorname{Var}_{D\cap[x,\infty[}(f)$$

for every $x \in \mathbb{R}$, with equality if $x \in D \cap \operatorname{dom} f$,

$$\operatorname{Var}_D(f) \le \operatorname{Var}_{D'}(f) \text{ whenever } D \subseteq D';$$

the arguments of 224C go through unchanged.

(c) A complex-valued function is of bounded variation iff its real and imaginary parts are both of bounded variation (because

$$\max(\mathrm{Var}_D(\mathcal{R}\mathrm{e}\, f), \mathrm{Var}_D(\mathcal{I}\mathrm{m}\, f)) \le \mathrm{Var}_D(f) \le \mathrm{Var}_D(\mathcal{R}\mathrm{e}\, f) + \mathrm{Var}_D(\mathcal{I}\mathrm{m}\, f).)$$

So a complex-valued function f is of bounded variation on D iff there are bounded non-decreasing functions $f_1, \dots, f_4 : \mathbb{R} \to \mathbb{R}$ such that $f = f_1 - f_2 + if_3 - if_4$ on D (224D).

(d) Let f be a complex-valued function and D any subset of $\mathbb{R}$. If f is of bounded variation on D, then

$$\lim_{x\downarrow a} \mathrm{Var}_{D\cap]a,x]}(f) = \lim_{x\uparrow a} \mathrm{Var}_{D\cap[x,a[}(f) = 0$$

for every $a \in \mathbb{R}$, and

$$\lim_{a\to-\infty} \mathrm{Var}_{D\cap]-\infty,a]}(f) = \lim_{a\to\infty} \mathrm{Var}_{D\cap[a,\infty[}(f) = 0.$$

(Apply 224E to the real and imaginary parts of f.)

(e) Let f be a complex-valued function of bounded variation on $[a, b]$, where $a < b$. If $\mathrm{dom}\, f$ meets every interval $]a, a + \delta]$ with $\delta > 0$, then $\lim_{t\in\mathrm{dom}\, f, t\downarrow a} f(t)$ is defined in $\mathbb{C}$. If $\mathrm{dom}\, f$ meets $[b - \delta, b[$ for every $\delta > 0$, then $\lim_{t\in\mathrm{dom}\, f, t\uparrow b} f(t)$ is defined in $\mathbb{C}$. (Apply 224F to the real and imaginary parts of f.)

(f) Let f, g be complex functions and D a subset of $\mathbb{R}$. If f and g are of bounded variation on D, so is $f \times g$. (For $f \times g$ is expressible as a linear combination of the four products $\mathcal{R}\mathrm{e}\, f \times \mathcal{R}\mathrm{e}\, g, \dots, \mathcal{I}\mathrm{m}\, f \times \mathcal{I}\mathrm{m}\, g$, to each of which we can apply 224G.)

(g) Let $I \subseteq \mathbb{R}$ be an interval, and $f : I \to \mathbb{C}$ a function of bounded variation. Then f is differentiable almost everywhere in I, and $\int_I |f'| \le \mathrm{Var}_I(f)$. (As 224I.)

(h) Let f and g be complex-valued functions defined on subsets of $\mathbb{R}$, and suppose that g is integrable over an interval $[a, b]$, where $a < b$, and f is of bounded variation on $]a, b[$ and defined almost everywhere in $]a, b[$. Then $f \times g$ is integrable over $[a, b]$, and

$$\Bigl|\int_a^b f \times g\Bigr| \le \bigl(\lim_{x\in\mathrm{dom}\, f, x\uparrow b} |f(x)| + \mathrm{Var}_{]a,b[}(f)\bigr) \sup_{c\in[a,b]} \Bigl|\int_a^c g\Bigr|.$$

(The argument of 224J applies virtually unchanged.)

224X Basic exercises >(a) Set $f(x) = x^2 \sin \dfrac{1}{x^2}$ for $x \ne 0$, $f(0) = 0$. Show that $f : \mathbb{R} \to \mathbb{R}$ is differentiable everywhere and uniformly continuous, but is not of bounded variation on any non-trivial interval containing 0.

(b) Give an example of a non-negative function $g : [0, 1] \to [0, 1]$, of bounded variation, such that $\sqrt{g}$ is not of bounded variation.

(c) Show that if f is any real-valued function defined on a subset of $\mathbb{R}$, there is a function $\tilde{f} : \mathbb{R} \to \mathbb{R}$, extending f, such that $\mathrm{Var}\, \tilde{f} = \mathrm{Var}\, f$. Under what circumstances is $\tilde{f}$ unique?

(d) Let $f : D \to \mathbb{R}$ be a function of bounded variation, where $D \subseteq \mathbb{R}$ is a non-empty set. Show that if $\inf_{x\in D} |f(x)| > 0$ then $1/f$ is of bounded variation.

(e) Let $f : [a, b] \to \mathbb{R}$ be a continuous function, where $a \le b$ in $\mathbb{R}$. Show that if $c < \mathrm{Var}\, f$ then there is a $\delta > 0$ such that $\sum_{i=1}^n |f(a_i) - f(a_{i-1})| \ge c$ whenever $a = a_0 \le a_1 \le \ldots \le a_n = b$ and $\max_{1\le i\le n} a_i - a_{i-1} \le \delta$.

(f) Let $\langle f_n \rangle_{n\in\mathbb{N}}$ be a sequence of real functions, and set $f(x) = \lim_{n\to\infty} f_n(x)$ whenever the limit is defined. Show that $\mathrm{Var}\, f \le \liminf_{n\to\infty} \mathrm{Var}\, f_n$.

(g) Let f be a real-valued function which is integrable over an interval $[a, b] \subseteq \mathbb{R}$. Set $F(x) = \int_a^x f$ for $x \in [a, b]$. Show that $\mathrm{Var}\, F = \int_a^b |f|$. (*Hint*: start by checking that $\mathrm{Var}\, F \le \int |f|$; for the reverse inequality, consider the case $f \ge 0$ first.)

(h) Show that if f is a real-valued function defined on a non-empty set $D \subseteq \mathbb{R}$, then

$$\mathrm{Var}\, f = \sup\{|\textstyle\sum_{i=1}^n (-1)^i (f(a_i) - f(a_{i-1}))| : a_0 \le a_1 \le \ldots \le a_n \text{ in } D\}.$$

(i) Let f be a real-valued function which is integrable over a bounded interval $[a,b] \subseteq \mathbb{R}$. Show that

$$\int_a^b |f| = \sup\{|\sum_{i=1}^n (-1)^i \int_{a_{i-1}}^{a_i} f| : a = a_0 \le a_1 \le a_2 \le \ldots \le a_n = b\}.$$

(*Hint*: put 224Xg and 224Xh together.)

(j) Let f and g be real-valued functions defined on subsets of $\mathbb{R}$, and suppose that g is integrable over an interval $[a,b]$, where $a < b$, and f is of bounded variation on $]a,b[$ and defined almost everywhere in $]a,b[$. Show that

$$|\int_a^b f \times g| \le (\lim_{x\in\operatorname{dom} f, x\downarrow a} |f(x)| + \operatorname{Var}_{]a,b[}(f)) \sup_{c\in[a,b]} |\int_c^b g|.$$

224Y Further exercises (a) Show that if f is any complex-valued function defined on a subset of $\mathbb{R}$, there is a function $\tilde{f} : \mathbb{R} \to \mathbb{C}$, extending f, such that $\operatorname{Var} \tilde{f} = \operatorname{Var} f$. Under what circumstances is $\tilde{f}$ unique?

(b) Let D be any non-empty subset of $\mathbb{R}$, and let $\mathcal{V}$ be the space of functions $f : D \to \mathbb{R}$ of bounded variation. For $f \in \mathcal{V}$ set

$$\|f\| = \sup\{|f(t_0)| + \sum_{i=1}^n |f(t_i) - f(t_{i-1})| : t_0 \le \ldots \le t_n \in D\}.$$

Show that (i) $\| \;\|$ is a norm on $\mathcal{V}$ (ii) $\mathcal{V}$ is complete under $\| \;\|$ (iii) $\|f \times g\| \le \|f\|\|g\|$ for all f, $g \in \mathcal{V}$, so that $\mathcal{V}$ is a Banach algebra.

(c) Let $f : \mathbb{R} \to \mathbb{R}$ be a function of bounded variation. Show that there is a sequence $\langle f_n \rangle_{n\in\mathbb{N}}$ of differentiable functions such that $\lim_{n\to\infty} f_n(x) = f(x)$ for every $x \in \mathbb{R}$, $\lim_{n\to\infty} \int |f_n - f| = 0$, and $\operatorname{Var}(f_n) \le \operatorname{Var}(f)$ for every $n \in \mathbb{N}$. (*Hint*: start with non-decreasing f.)

(d) For any partially ordered set X and any function $f : X \to \mathbb{R}$, say that $\operatorname{Var}_X(f) = 0$ if $X = \emptyset$ and otherwise

$$\operatorname{Var}_X(f) = \sup\{\sum_{i=1}^n |f(a_i) - f(a_{i-1})| : a_0, a_1, \ldots, a_n \in X,\ a_0 \le a_1 \le \ldots \le a_n\}.$$

State and prove results in this framework generalizing 224D and 224Yb. (*Hints*: f will be 'non-decreasing' if $f(x) \le f(y)$ whenever $x \le y$; interpret $]-\infty, x]$ as $\{y : y \le x\}$.)

(e) Let (X,ρ) be a metric space and $f : [a,b] \to X$ a function, where $a \le b$ in $\mathbb{R}$. Set $\operatorname{Var}_{[a,b]}(f) = \sup\{\sum_{i=1}^n \rho(f(a_i), f(a_{i-1})) : a \le a_0 \le \ldots \le a_n \le b\}$. (i) Show that $\operatorname{Var}_{[a,b]}(f) = \operatorname{Var}_{[a,c]}(f) + \operatorname{Var}_{[c,b]}(f)$ for every $c \in [a,b]$. (ii) Show that if $\operatorname{Var}_{[a,b]}(f)$ is finite then f is continuous at all but countably many points of $[a,b]$. (iii) Show that if X is complete and $\operatorname{Var}_{[a,b]}(f) < \infty$ then $\lim_{t\uparrow x} f(t)$ is defined for every $x \in]a,b]$. (iv) Show that if X is complete then $\operatorname{Var}_{[a,b]}(f)$ is finite iff f is expressible as a composition gh, where $h : [a,b] \to \mathbb{R}$ is non-decreasing and $g : \mathbb{R} \to X$ is 1-Lipschitz, that is, $\rho(g(c), g(d)) \le |c - d|$ for all c, $d \in \mathbb{R}$.

(f) Let U be a normed space and $a \le b$ in $\mathbb{R}$. For functions $f : [a,b] \to U$ define $\operatorname{Var}_{[a,b]}(f)$ as in 224Ye, using the standard metric $\rho(x,y) = \|x - y\|$ for x, $y \in U$. (i) Show that $\operatorname{Var}_{[a,b]}(f+g) \le \operatorname{Var}_{[a,b]}(f) + \operatorname{Var}_{[a,b]}(g)$, $\operatorname{Var}_{[a,b]}(cf) = |c| \operatorname{Var}_{[a,b]}(f)$ for all f, $g : [a,b] \to U$ and all $c \in \mathbb{R}$. (ii) Show that if V is another normed space and $T : U \to V$ is a bounded linear operator then $\operatorname{Var}_{[a,b]}(Tf) \le \|T\| \operatorname{Var}_{[a,b]}(f)$ for every $f : [a,b] \to U$.

(g) Let $f : [0,1] \to \mathbb{R}$ be a continuous function. For $y \in \mathbb{R}$ set $h(y) = \#(f^{-1}[\{y\}])$ if this is finite, ∞ otherwise. Show that (if we allow ∞ as a value of the integral) $\operatorname{Var}_{[0,1]}(f) = \int h$. (*Hint*: for $n \in \mathbb{N}$, $i < 2^n$ set $c_{ni} = \sup\{f(x) - f(y) : x, y \in [2^{-n}i, 2^{-n}(i+1)]\}$, $h_{ni}(y) = 1$ if $y \in f[[2^{-n}i, 2^{-n}(i+1)[]$, 0 otherwise. Show that $c_{ni} = \int h_{ni}$, $\lim_{n\to\infty} \sum_{i=0}^{2^n-1} c_{ni} = \operatorname{Var} f$, $\lim_{n\to\infty} \sum_{i=0}^{2^n-1} h_{ni} = h$.) (See also 226Yb.)

(h) Let ν be any Lebesgue-Stieltjes measure on $\mathbb{R}$, $I \subseteq \mathbb{R}$ an interval (which may be either open or closed, bounded or unbounded), and $D \subseteq I$ a non-empty set. Let $\mathcal{V}$ be the space of functions of bounded variation from D to $\mathbb{R}$, and $\| \;\|$ the norm of 224Yb on $\mathcal{V}$. Let $g : D \to \mathbb{R}$ be a function such that $\int_{[a,b]\cap D} g\,d\nu$ exists whenever $a \le b$ in I, and $K = \sup_{a,b\in I, a\le b} |\int_{[a,b]\cap D} g\,d\nu|$. Show that $|\int_D f \times g\,d\nu| \le K\|f\|$ for every $f \in \mathcal{V}$.

(i) Explain how to apply 224Yh with $D = \mathbb{N}$ to obtain Abel's theorem that the product of a monotonic sequence converging to 0 with a series which has bounded partial sums is summable.

(j) Suppose that $I \subseteq \mathbb{R}$ is an interval, and that $\langle A_n \rangle_{n\in\mathbb{N}}$ is a sequence of sets covering I. Let $f : I \to \mathbb{R}$ be continuous. Show that $\operatorname{Var} f \le \sum_{n=0}^\infty \operatorname{Var}_{A_n} f$. (*Hint*: reduce to the case of closed sets A_n; use Baire's theorem (4A2Ma).)

224 Notes and comments I have taken the ideas above rather farther than we need immediately; for the present chapter, it is enough to consider the case in which $D = \operatorname{dom} f = [a, b]$ for some interval $[a, b] \subseteq \mathbb{R}$. The extension to functions with irregular domains will be useful in Chapter 28, and the extension to irregular sets D, while not important to us here, is of some interest – for instance, taking $D = \mathbb{N}$, we obtain the notion of 'sequence of bounded variation', which is surely relevant to problems of convergence and summability.

The central result of the section is of course the fact that a function of bounded variation can be expressed as the difference of monotonic functions (224D); indeed, one of the objects of the concept is to characterize the linear span of the monotonic functions. Nearly everything else here can be derived as easy consequences of this, as in 224E-224G. In 224I and 224Xg we go a little deeper, and indeed some measure theory appears; this is where the ideas here begin to connect with the real business of this chapter, to be continued in the next section. Another result which is easy enough in itself, but contains the germs of important ideas, is 224Yg.

In 224Yb I mention a natural development in functional analysis, and in 224Yd-224Yf I suggest further wide-ranging generalizations.

225 Absolutely continuous functions

We are now ready for a full characterization of the functions that can appear as indefinite integrals (225E, 225Xh). The essential idea is that of 'absolute continuity' (225B). In the second half of the section (225G-225N) I describe some of the relationships between this concept and those we have already seen.

225A Absolute continuity of the indefinite integral I begin with an easy fundamental result from general measure theory.

Theorem Let (X, Σ, μ) be any measure space and f an integrable real-valued function defined on a conegligible subset of X. Then for any $\epsilon > 0$ there are a measurable set E of finite measure and a real number $\delta > 0$ such that $\int_F |f| \le \epsilon$ whenever $F \in \Sigma$ and $\mu(F \cap E) \le \delta$.

proof There is a non-decreasing sequence $\langle g_n \rangle_{n\in\mathbb{N}}$ of non-negative simple functions such that $|f| =_{\text{a.e.}} \lim_{n\to\infty} g_n$ and $\int |f| = \lim_{n\to\infty} \int g_n$. Take $n \in \mathbb{N}$ such that $\int g_n \ge \int |f| - \frac{1}{2}\epsilon$. Let $M > 0$, $E \in \Sigma$ be such that $\mu E < \infty$ and $g_n \le M\chi E$; set $\delta = \epsilon/2M$. If $F \in \Sigma$ and $\mu(F \cap E) \le \delta$, then

$$\int_F g_n = \int g_n \times \chi F \le M\mu(F \cap E) \le \frac{1}{2}\epsilon;$$

consequently

$$\int_F |f| = \int_F g_n + \int_F |f| - g_n \le \frac{1}{2}\epsilon + \int |f| - g_n \le \epsilon.$$

225B Absolutely continuous functions on $\mathbb{R}$: Definition If $[a, b]$ is a non-empty closed interval in $\mathbb{R}$ and $f : [a, b] \to \mathbb{R}$ is a function, we say that f is **absolutely continuous** if for every $\epsilon > 0$ there is a $\delta > 0$ such that $\sum_{i=1}^n |f(b_i) - f(a_i)| \le \epsilon$ whenever $a \le a_1 \le b_1 \le a_2 \le b_2 \le \ldots \le a_n \le b_n \le b$ and $\sum_{i=1}^n b_i - a_i \le \delta$.

Remark The phrase 'absolutely continuous' is used in various senses in measure theory, closely related (if you look at them in the right way) but not identical; you will need to keep the context of each definition in clear focus.

225C Proposition Let $[a, b]$ be a non-empty closed interval in $\mathbb{R}$.

(a) If $f : [a, b] \to \mathbb{R}$ is absolutely continuous, it is uniformly continuous.

(b) If $f : [a, b] \to \mathbb{R}$ is absolutely continuous it is of bounded variation on $[a, b]$, so is differentiable almost everywhere in $[a, b]$, and its derivative is integrable over $[a, b]$.

(c) If f, $g : [a, b] \to \mathbb{R}$ are absolutely continuous, so are $f + g$ and cf, for every $c \in \mathbb{R}$.

(d) If f, $g : [a, b] \to \mathbb{R}$ are absolutely continuous so is $f \times g$.

(e) If $g : [a, b] \to [c, d]$ and $f : [c, d] \to \mathbb{R}$ are absolutely continuous, and g is non-decreasing, then the composition $fg : [a, b] \to \mathbb{R}$ is absolutely continuous.

proof (a) Let $\epsilon > 0$. Then there is a $\delta > 0$ such that $\sum_{i=1}^n |f(b_i) - f(a_i)| \le \epsilon$ whenever $a \le a_1 \le b_1 \le a_2 \le b_2 \le \ldots \le a_n \le b_n \le b$ and $\sum_{i=1}^n b_i - a_i \le \delta$; but of course now $|f(y) - f(x)| \le \epsilon$ whenever x, $y \in [a, b]$ and $|x - y| \le \delta$. As ϵ is arbitrary, f is uniformly continuous.

(b) Let $\delta > 0$ be such that $\sum_{i=1}^n |f(b_i) - f(a_i)| \le 1$ whenever $a \le a_1 \le b_1 \le a_2 \le b_2 \le \ldots \le a_n \le b_n \le b$ and $\sum_{i=1}^n b_i - a_i \le \delta$. If $a \le c = c_0 \le c_1 \le \ldots \le c_n \le d \le \min(b, c+\delta)$, then $\sum_{i=1}^n |f(c_i) - f(c_{i-1})| \le 1$, so $\mathrm{Var}_{[c,d]}(f) \le 1$; accordingly (inducing on k, using 224Cc for the inductive step) $\mathrm{Var}_{[a,\min(a+k\delta,b)]}(f) \le k$ for every k, and

$$\mathrm{Var}_{[a,b]}(f) \le \lceil (b-a)/\delta \rceil < \infty.$$

It follows that f' is integrable, by 224I.

(c)(i) Let $\epsilon > 0$. Then there are δ_1, $\delta_2 > 0$ such that

$$\sum_{i=1}^n |f(b_i) - f(a_i)| \le \frac{1}{2}\epsilon$$

whenever $a \le a_1 \le b_1 \le a_2 \le b_2 \le \ldots \le a_n \le b_n \le b$ and $\sum_{i=1}^n b_i - a_i \le \delta_1$,

$$\sum_{i=1}^n |g(b_i) - g(a_i)| \le \frac{1}{2}\epsilon$$

whenever $a \le a_1 \le b_1 \le a_2 \le b_2 \le \ldots \le a_n \le b_n \le b$ and $\sum_{i=1}^n b_i - a_i \le \delta_2$. Set $\delta = \min(\delta_1, \delta_2) > 0$, and suppose that $a \le a_1 \le b_1 \le a_2 \le b_2 \le \ldots \le a_n \le b_n \le b$ and $\sum_{i=1}^n b_i - a_i \le \delta$. Then

$$\sum_{i=1}^n |(f+g)(b_i) - (f+g)(a_i)| \le \sum_{i=1}^n |f(b_i) - f(a_i)| + \sum_{i=1}^n |g(b_i) - g(a_i)| \le \epsilon.$$

As ϵ is arbitrary, $f + g$ is absolutely continuous.

(ii) Let $\epsilon > 0$. Then there is a $\delta > 0$ such that

$$\sum_{i=1}^n |f(b_i) - f(a_i)| \le \frac{\epsilon}{1+|c|}$$

whenever $a \le a_1 \le b_1 \le a_2 \le b_2 \le \ldots \le a_n \le b_n \le b$ and $\sum_{i=1}^n b_i - a_i \le \delta$. Now

$$\sum_{i=1}^n |(cf)(b_i) - (cf)(a_i)| \le \epsilon$$

whenever $a \le a_1 \le b_1 \le a_2 \le b_2 \le \ldots \le a_n \le b_n \le b$ and $\sum_{i=1}^n b_i - a_i \le \delta$. As ϵ is arbitrary, cf is absolutely continuous.

(d) By either (a) or (b), f and g are bounded; set $M = \sup_{x\in[a,b]} |f(x)|$, $M' = \sup_{x\in[a,b]} |g(x)|$. Let $\epsilon > 0$. Then there are δ_1, $\delta_2 > 0$ such that

$\sum_{i=1}^n |f(b_i) - f(a_i)| \le \epsilon$ whenever $a \le a_1 \le b_1 \le a_2 \le b_2 \le \ldots \le a_n \le b_n \le b$ and $\sum_{i=1}^n b_i - a_i \le \delta_1$,

$\sum_{i=1}^n |g(b_i) - g(a_i)| \le \epsilon$ whenever $a \le a_1 \le b_1 \le a_2 \le b_2 \le \ldots \le a_n \le b_n \le b$ and $\sum_{i=1}^n b_i - a_i \le \delta_2$.

Set $\delta = \min(\delta_1, \delta_2) > 0$ and suppose that $a \le a_1 \le b_1 \le \ldots \le b_n \le b$ and $\sum_{i=1}^n b_i - a_i \le \delta$. Then

$$\begin{aligned}\sum_{i=1}^n |f(b_i)g(b_i) - f(a_i)g(a_i)| &= \sum_{i=1}^n |(f(b_i) - f(a_i))g(b_i) + f(a_i)(g(b_i) - g(a_i))| \\ &\le \sum_{i=1}^n |f(b_i) - f(a_i)||g(b_i)| + |f(a_i)||g(b_i) - g(a_i)| \\ &\le \sum_{i=1}^n |f(b_i) - f(a_i)|M' + M|g(b_i) - g(a_i)| \\ &\le \epsilon M' + M\epsilon = \epsilon(M + M').\end{aligned}$$

As ϵ is arbitrary, $f \times g$ is absolutely continuous.

(e) Let $\epsilon > 0$. Then there is a $\delta > 0$ such that $\sum_{i=1}^n |f(d_i) - f(c_i)| \le \epsilon$ whenever $c \le c_1 \le d_1 \le \ldots \le c_n \le d_n \le d$ and $\sum_{i=1}^n d_i - c_i \le \delta$; and there is an $\eta > 0$ such that $\sum_{i=1}^n |g(b_i) - g(a_i)| \le \delta$ whenever $a \le a_1 \le b_1 \le \ldots \le a_n \le b_n \le b$ and $\sum_{i=1}^n b_i - a_i \le \eta$. Now suppose that $a \le a_1 \le b_1 \le \ldots \le a_n \le b_n \le b$ and $\sum_{i=1}^n b_i - a_i \le \eta$. Because g is non-decreasing, we have $c \le g(a_1) \le \ldots \le g(b_n) \le d$ and $\sum_{i=1}^n g(b_i) - g(a_i) \le \delta$, so $\sum_{i=1}^n |f(g(b_i)) - f(g(a_i))| \le \epsilon$; as ϵ is arbitrary, fg is absolutely continuous.

225D Lemma Let $[a, b]$ be a non-empty closed interval in $\mathbb{R}$ and $f : [a, b] \to \mathbb{R}$ an absolutely continuous function which has zero derivative almost everywhere in $[a, b]$. Then f is constant on $[a, b]$.

proof Let $x \in [a, b]$, $\epsilon > 0$. Let $\delta > 0$ be such that $\sum_{i=1}^n |f(b_i) - f(a_i)| \le \epsilon$ whenever $a \le a_1 \le b_1 \le a_2 \le b_2 \le \ldots \le a_n \le b_n \le b$ and $\sum_{i=1}^n b_i - a_i \le \delta$. Set $A = \{t : a < t < x,\ f'(t) \text{ exists } = 0\}$; then $\mu A = x - a$, writing μ for Lebesgue measure as usual. Let $\mathcal{I}$ be the set of non-empty non-singleton closed intervals $[c, d] \subseteq [a, x]$ such that

$|f(d)-f(c)| \le \epsilon(d-c)$; then every member of A belongs to arbitrarily short members of $\mathcal{I}$. By Vitali's theorem (221A), there is a countable disjoint family $\mathcal{I}_0 \subseteq \mathcal{I}$ such that $\mu(A \setminus \bigcup \mathcal{I}_0) = 0$, that is,

$$x - a = \mu(\bigcup \mathcal{I}_0) = \sum_{I \in \mathcal{I}_0} \mu I.$$

Now there is a finite $\mathcal{I}_1 \subseteq \mathcal{I}_0$ such that

$$\mu(\bigcup \mathcal{I}_1) = \sum_{I \in \mathcal{I}_1} \mu I \ge x - a - \delta.$$

If $\mathcal{I}_1 = \emptyset$, then $x \le a + \delta$ and $|f(x) - f(a)| \le \epsilon$. Otherwise, express $\mathcal{I}_1$ as $\{[c_0, d_0], \dots , [c_n, d_n]\}$, where $a \le c_0 < d_0 < c_1 < d_1 < \dots < c_n < d_n \le x$. Then

$$(c_0 - a) + \sum_{i=1}^n (c_i - d_{i-1}) + (x - d_n) = \mu([a, x] \setminus \bigcup \mathcal{I}_1) \le \delta,$$

so

$$|f(c_0) - f(a)| + \sum_{i=1}^n |f(c_i) - f(d_{i-1})| + |f(x) - f(d_n)| \le \epsilon.$$

On the other hand, $|f(d_i) - f(c_i)| \le \epsilon(d_i - c_i)$ for each i, so

$$\sum_{i=0}^n |f(d_i) - f(c_i)| \le \epsilon \sum_{i=0}^n d_i - c_i \le \epsilon(x - a).$$

Putting these together,

$$\begin{aligned} |f(x) - f(a)| &\le |f(c_0) - f(a)| + |f(d_0) - f(c_0)| + |f(c_1) - f(d_0)| + \dots \\ &\qquad + |f(d_n) - f(c_n)| + |f(x) - f(d_n)| \\ &= |f(c_0) - f(a)| + \sum_{i=1}^n |f(c_i) - f(d_{i-1})| \\ &\qquad + |f(x) - f(d_n)| + \sum_{i=0}^n |f(d_i) - f(c_i)| \\ &\le \epsilon + \epsilon(x - a) = \epsilon(1 + x - a). \end{aligned}$$

As ϵ is arbitrary, $f(x) = f(a)$. As x is arbitrary, f is constant.

225E Theorem Let $[a, b]$ be a non-empty closed interval in $\mathbb{R}$ and $F : [a, b] \to \mathbb{R}$ a function. Then the following are equiveridical:

(i) there is an integrable real-valued function f such that $F(x) = F(a) + \int_a^x f$ for every $x \in [a, b]$;

(ii) $\int_a^x F'$ exists and is equal to $F(x) - F(a)$ for every $x \in [a, b]$;

(iii) F is absolutely continuous.

proof (i)⇒(iii) Assume (i). Let $\epsilon > 0$. By 225A, there is a $\delta > 0$ such that $\int_H |f| \le \epsilon$ whenever $H \subseteq [a, b]$ and $\mu H \le \delta$, writing μ for Lebesgue measure as usual. Now suppose that $a \le a_1 \le b_1 \le a_2 \le b_2 \le \dots \le a_n \le b_n \le b$ and $\sum_{i=1}^n b_i - a_i \le \delta$. Consider $H = \bigcup_{1 \le i \le n} [a_i, b_i[$. Then $\mu H \le \delta$ and

$$\sum_{i=1}^n |F(b_i) - F(a_i)| = \sum_{i=1}^n |\int_{[a_i,b_i[} f| \le \sum_{i=1}^n \int_{[a_i,b_i[} |f| = \int_F |f| \le \epsilon.$$

As ϵ is arbitrary, F is absolutely continuous.

(iii)⇒(ii) If F is absolutely continuous, then it is of bounded variation (by 225Ba), so $\int_a^b F'$ exists (224I). Set $G(x) = \int_a^x F'$ for $x \in [a, b]$; then $G' =_{\text{a.e.}} F'$ (222E) and G is absolutely continuous (by (i)⇒(iii) just proved). Accordingly $G - F$ is absolutely continuous (225Bb) and is differentiable, with zero derivative, almost everywhere. It follows that $G - F$ must be constant (225D). But as $G(a) = 0$, $G = F + F(a)$; just as required by (ii).

(ii)⇒(i) is trivial.

225F Integration by parts As an application of this result, I give a justification of a familiar formula.

Theorem Let f be a real-valued function which is integrable over an interval $[a, b] \subseteq \mathbb{R}$, and $g : [a, b] \to \mathbb{R}$ an absolutely continuous function. Suppose that F is an indefinite integral of f, so that $F(x) - F(a) = \int_a^x f$ for $x \in [a, b]$. Then

$$\int_a^b f \times g = F(b)g(b) - F(a)g(a) - \int_a^b F \times g'.$$

proof Set $h = F \times g$. Because F is absolutely continuous (225E), so is h (225Cd). Consequently $h(b) - h(a) = \int_a^b h'$, by (iii)⇒(ii) of 225E. But $h' = F' \times g + F \times g'$ wherever F' and g' are defined, which is almost everywhere, and

$F' =_{\text{a.e.}} f$, by 222E; so $h' =_{\text{a.e.}} f \times g + F \times g'$. Finally, g and F are continuous, therefore measurable, and bounded, while f and g' are integrable (using 225E yet again), so $f \times g$ and $F \times g'$ are integrable, and

$$F(b)g(b) - F(a)g(a) = h(b) - h(a) = \int_a^b h' = \int_a^b f \times g + \int_a^b F \times g',$$

as required.

225G I come now to a group of results at a rather deeper level than most of the work of this chapter, being closer to the ideas of Chapter 26.

Proposition Let $[a, b]$ be a non-empty closed interval in $\mathbb{R}$ and $f : [a, b] \to \mathbb{R}$ an absolutely continuous function. Then $f[A]$ is negligible for every negligible set $A \subseteq \mathbb{R}$.

proof Let $\epsilon > 0$. Then there is a $\delta > 0$ such that $\sum_{i=1}^n |f(b_i) - f(a_i)| \le \epsilon$ whenever $a \le a_1 \le b_1 \ldots \le a_n \le b_n \le b$ and $\sum_{i=1}^n b_i - a_i \le \delta$. Now there is a sequence $\langle I_k \rangle_{k\in\mathbb{N}}$ of closed intervals, covering A, with $\sum_{k=0}^\infty \mu I_k \le \delta$. For each $m \in \mathbb{N}$, let F_m be $[a, b] \cap \bigcup_{k\le m} I_k$. Then $\mu f[F_m] \le \epsilon$. $\mathbf{P}$ F_m must be expressible as $\bigcup_{i\le n}[c_i, d_i]$ where $n \le m$ and $a \le c_0 \le d_0 \le \ldots \le c_n \le d_n \le b$. For each $i \le n$ choose x_i, y_i such that $c_i \le x_i$, $y_i \le d_i$ and

$$f(x_i) = \min_{x\in[c_i,d_i]} f(x), \quad f(y_i) = \max_{x\in[c_i,d_i]} f(x);$$

such exist because f is continuous, so is bounded and attains its bounds on $[c_i, d_i]$. Set $a_i = \min(x_i, y_i)$, $b_i = \max(x_i, y_i)$, so that $c_i \le a_i \le b_i \le d_i$. Then

$$\sum_{i=0}^n b_i - a_i \le \sum_{i=0}^n d_i - c_i = \mu F_m \le \mu(\bigcup_{k\in\mathbb{N}} I_k) \le \delta,$$

so

$$\begin{aligned}\mu f[F_m] &= \mu(\bigcup_{i\le m} f[\,[c_i, d_i]\,]) \le \sum_{i=0}^n \mu(f[\,[c_i, d_i]\,]) \\ &= \sum_{i=0}^n \mu[f(x_i), f(y_i)] = \sum_{i=0}^n |f(b_i) - f(a_i)| \le \epsilon. \ \mathbf{Q}\end{aligned}$$

But $\langle f[F_m]\rangle_{m\in\mathbb{N}}$ is a non-decreasing sequence covering $f[A]$, so

$$\mu^* f[A] \le \mu(\bigcup_{m\in\mathbb{N}} f[F_m]) = \sup_{m\in\mathbb{N}} \mu f[F_m] \le \epsilon.$$

As ϵ is arbitrary, $f[A]$ is negligible, as claimed.

225H Semi-continuous functions In preparation for the last main result of this section, I give a general result concerning measurable real-valued functions on subsets of $\mathbb{R}$. It will be convenient here, for once, to consider functions taking values in $[-\infty, \infty]$. If $D \subseteq \mathbb{R}^r$, a function $g : D \to [-\infty, \infty]$ is **lower semi-continuous** if $\{x : g(x) > u\}$ is an open subset of D (for the subspace topology, see 2A3C) for every $u \in [-\infty, \infty]$. Any lower semi-continuous function is Borel measurable, therefore Lebesgue measurable (121B-121D). Now we have the following result.

225I Proposition Suppose that $r \ge 1$ and that f is a real-valued function, defined on a subset D of $\mathbb{R}^r$, which is integrable over D. Then for any $\epsilon > 0$ there is a lower semi-continuous function $g : \mathbb{R}^r \to [-\infty, \infty]$ such that $g(x) \ge f(x)$ for every $x \in D$ and $\int_D g$ is defined and not greater than $\epsilon + \int_D f$.

Remarks This is a result of great general importance, so I give it in a fairly general form; but for the present chapter all we need is the case $r = 1$, $D = [a, b]$ where $a \le b$.

proof (a) We can enumerate $\mathbb{Q}$ as $\langle q_n \rangle_{n\in\mathbb{N}}$. By 225A, there is a $\delta > 0$ such that $\int_F |f| \le \frac{1}{2}\epsilon$ whenever $\mu_D F \le \delta$, where μ_D is the subspace measure on D, so that $\mu_D F = \mu^* F$, the outer Lebesgue measure of F, for every $F \in \Sigma_D$, the domain of μ_D (214A-214B). For each $n \in \mathbb{N}$, set

$$\delta_n = 2^{-n-1} \min(\tfrac{\epsilon}{1+2|q_n|}, \delta),$$

so that $\sum_{n=0}^\infty \delta_n|q_n| \le \frac{1}{2}\epsilon$ and $\sum_{n=0}^\infty \delta_n \le \delta$. For each $n \in \mathbb{N}$, let $E_n \subseteq \mathbb{R}^r$ be a Lebesgue measurable set such that $\{x : f(x) \ge q_n\} = D \cap E_n$, and choose an open set $G_n \supseteq E_n \cap B(\mathbf{0}, n)$ such that $\mu G_n \le \mu(E_n \cap B(\mathbf{0}, n)) + \delta_n$ (134Fa), writing $B(\mathbf{0}, n)$ for the ball $\{x : \|x\| \le n\}$. For $x \in \mathbb{R}^r$, set

$$g(x) = \sup\{q_n : x \in G_n\},$$

allowing $-\infty$ as $\sup \emptyset$ and ∞ as the supremum of a set with no upper bound in $\mathbb{R}$.

(b) Now check the properties of g.

(i) g is lower semi-continuous. **P** If $u \in [-\infty, \infty]$, then

$$\{x : g(x) > u\} = \bigcup\{G_n : q_n > u\}$$

is a union of open sets, therefore open. **Q**

(ii) $g(x) \geq f(x)$ for every $x \in D$. **P** If $x \in D$ and $\eta > 0$, there is an $n \in \mathbb{N}$ such that $\|x\| \leq n$ and $f(x) - \eta \leq q_n \leq f(x)$; now $x \in E_n \subseteq G_n$ so $g(x) \geq q_n \geq f(x) - \eta$. As η is arbitrary, $g(x) \geq f(x)$. **Q**

(iii) Consider the functions h_1, $h_2 : D \to]-\infty, \infty]$ defined by setting

$$\begin{aligned} h_1(x) &= |f(x)| \text{ if } x \in D \cap \bigcup_{n\in\mathbb{N}} (G_n \setminus E_n), \\ &= 0 \text{ for other } x \in D, \\ h_2(x) &= \sum_{n=0}^{\infty} |q_n| \chi(G_n \setminus E_n)(x) \text{ for every } x \in D. \end{aligned}$$

Setting $F = \bigcup_{n\in\mathbb{N}} G_n \setminus E_n$,

$$\mu F \leq \textstyle\sum_{n=0}^{\infty} \mu(G_n \setminus E_n) \leq \delta,$$

so

$$\textstyle\int_D h_1 \leq \int_{D\cap F} |f| \leq \frac{1}{2}\epsilon$$

by the choice of δ. As for h_2, we have (by B.Levi's theorem)

$$\textstyle\int_D h_2 = \sum_{n=0}^{\infty} |q_n| \mu_D(D \cap G_n \setminus F_n) \leq \sum_{n=0}^{\infty} |q_n| \mu(G_n \setminus F_n) \leq \frac{1}{2}\epsilon$$

– because this is finite, $h_2(x) < \infty$ for almost every $x \in D$. Thus $\int_D h_1 + h_2 \leq \epsilon$.

(iv) The point is that $g \leq f + h_1 + h_2$ everywhere in D. **P** Take any $x \in D$. If $n \in \mathbb{N}$ and $x \in G_n$, then either $x \in E_n$, in which case

$$f(x) + h_1(x) + h_2(x) \geq f(x) \geq q_n,$$

or $x \in G_n \setminus E_n$, in which case

$$f(x) + h_1(x) + h_2(x) \geq f(x) + |f(x)| + |q_n| \geq q_n.$$

Thus

$$f(x) + h_1(x) + h_2(x) \geq \sup\{q_n : x \in G_n\} \geq g(x). \text{ } \mathbf{Q}$$

So $g \leq f + h_1 + h_2$ everywhere in D.

(v) Putting (iii) and (iv) together,

$$\textstyle\int_D g \leq \int_D f + h_1 + h_2 \leq \epsilon + \int_D f,$$

as required.

225J We need some results on Borel measurable sets and functions which are of independent interest.

Theorem Let D be a subset of $\mathbb{R}$ and $f : D \to \mathbb{R}$ any function. Then

$$E = \{x : x \in D,\ f \text{ is continuous at } x\}$$

is relatively Borel measurable in D, and

$$F = \{x : x \in D,\ f \text{ is differentiable at } x\}$$

is actually Borel measurable; moreover, $f' : F \to \mathbb{R}$ is Borel measurable.

proof (a) For $k \in \mathbb{N}$ set

$$\mathcal{G}_k = \{]a, b[: a,\ b \in \mathbb{R},\ |f(x) - f(y)| \leq 2^{-k} \text{ for all } x,\ y \in D \cap]a, b[\}.$$

Then $G_k = \bigcup \mathcal{G}_k$ is an open set, so $E_0 = \bigcap_{k\in\mathbb{N}} G_k$ is a Borel set. But $E = D \cap E_0$, so E is a relatively Borel subset of D.

(b)(i) I should perhaps say at once that when interpreting the formula $f'(x) = \lim_{h\to 0}(f(x+h) - f(x))/h$, I insist on the restrictive definition

$$a = \lim_{h\to 0} \tfrac{f(x+h)-f(x)}{h}$$

if

for every $\epsilon > 0$ there is a $\delta > 0$ such that $\frac{f(x+h)-f(x)}{h}$ is defined and

$$|\tfrac{f(x+h)-f(x)}{h} - a| \le \epsilon \text{ whenever } 0 < |h| \le \delta.$$

So $f'(x)$ can be defined only if there is some $\delta > 0$ such that the whole interval $[x-\delta, x+\delta]$ lies within the domain D of f.

(ii) For p, q, $q' \in \mathbb{Q}$ and $k \in \mathbb{N}$ set

$$\begin{aligned} H(k,p,q,q') &= \emptyset \text{ if }]q,q'[\not\subseteq D, \\ &= \{x : x \in E \cap]q,q'[\,, |f(y) - f(x) - p(y-x)| \le 2^{-k} \text{ for every } y \in]q,q'[\} \\ &\qquad \text{if }]q,q'[\subseteq D. \end{aligned}$$

Then $H(k,p,q,q') = E \cap]q,q'[\cap \overline{H(k,p,q,q')}$. **P** If $x \in E \cap]q,q'[\cap \overline{H(k,p,q,q')}$ there is a sequence $\langle x_n\rangle_{n\in\mathbb{N}}$ in $H(k,p,q,q')$ converging to x. Because f is continuous at x,

$$|f(y) - f(x) - p(y-x)| = \lim_{n\to\infty} |f(y) - f(x_n) - p(y-x_n)| \le 2^{-k}$$

for every $y \in]q,q'[$, so that $x \in H(k,p,q,q')$. **Q** Since E is a Borel set, by (a), so is $H(k,p,q,q')$.

(iii) Now

$$F = \bigcap_{k\in\mathbb{N}} \bigcup_{p,q,q'\in\mathbb{Q}} H(k,p,q,q').$$

P (α) Suppose $x \in F$, that is, $f'(x)$ is defined; say $f'(x) = a$. Take any $k \in \mathbb{N}$. Then there are $p \in \mathbb{Q}$, $\delta \in]0,1]$ such that $|p-a| \le 2^{-k-1}$ and $[x-\delta, x+\delta] \subseteq D$ and $|\frac{f(x+h)-f(x)}{h} - a| \le 2^{-k-1}$ whenever $0 < |h| \le \delta$; now take $q \in \mathbb{Q} \cap [x-\delta, x[$, $q' \in \mathbb{Q} \cap]x, x+\delta]$ and see that $x \in H(k,p,q,q')$. As x is arbitrary, $F \subseteq \bigcap_{k\in\mathbb{N}} \bigcup_{p,q,q'\in\mathbb{Q}} H(k,p,q,q')$.
(β) If $x \in \bigcap_{k\in\mathbb{N}} \bigcup_{p,q,q'\in\mathbb{Q}} H(k,p,q,q')$, then for each $k \in \mathbb{N}$ choose p_k, q_k, $q'_k \in \mathbb{Q}$ such that $x \in H(k,p_k,q_k,q'_k)$. If $h \ne 0$, $x+h \in]q_k, q'_k[$ then $|\frac{f(x+h)-f(x)}{h} - p_k| \le 2^{-k}$. But this means, first, that $|p_k - p_l| \le 2^{-k} + 2^{-l}$ for every k, l (since surely there is some $h \ne 0$ such that $x+h \in]q_k, q'_k[\cap]q_l, q'_l[$), so that $\langle p_k\rangle_{k\in\mathbb{N}}$ is a Cauchy sequence, with limit a say; and, second, that $|\frac{f(x+h)-f(x)}{h} - a| \le 2^{-k} + |a - p_k|$ whenever $h \ne 0$ and $x+h \in]q_k, q'_k[$, so that $f'(x)$ is defined and equal to a. **Q**

(iv) Because $\mathbb{Q}$ is countable, all the unions $\bigcup_{p,q,q'\in\mathbb{Q}} H(k,p,q,q')$ are Borel sets, so F also is.

(v) Now enumerate $\mathbb{Q}^3$ as $\langle(p_i,q_i,q'_i)\rangle_{i\in\mathbb{N}}$, and set $H'_{ki} = H(k,p_i,q_i,q'_i) \setminus \bigcup_{j<i} H(k,p_j,q_j,q'_j)$ for each k, $i \in \mathbb{N}$. Every H'_{ki} is Borel measurable, $\langle H'_{ki}\rangle_{i\in\mathbb{N}}$ is disjoint, and

$$\bigcup_{i\in\mathbb{N}} H'_{ki} = \bigcup_{i\in\mathbb{N}} H(k,p_i,q_i,q'_i) \supseteq F$$

for each k. Note that $|f'(x) - p| \le 2^{-k}$ whenever $x \in F \cap H(k,p,q,q')$, so if we set $f_k(x) = p_i$ for every $x \in H'_{ki}$ we shall have a Borel function f_k such that $|f(x) - f_k(x)| \le 2^{-k}$ for every $x \in F$. Accordingly $f' = \lim_{k\to\infty} f_k\restriction F$ is Borel measurable.

225K Proposition Let $[a,b]$ be a non-empty closed interval in $\mathbb{R}$, and $f : [a,b] \to \mathbb{R}$ a function. Set $F = \{x : x \in]a,b[\,, f'(x) \text{ is defined}\}$. Then f is absolutely continuous iff (i) f is continuous (ii) f' is integrable over F (iii) $f[[a,b] \setminus F]$ is negligible.

proof (a) Suppose first that f is absolutely continuous. Then f is surely continuous (225Ca) and f' is integrable over $[a,b]$, therefore over F (225E); also $[a,b] \setminus F$ is negligible, so $f[[a,b] \setminus F]$ is negligible, by 225G.

(b) So now suppose that f satisfies the conditions. Set $f^*(x) = |f'(x)|$ for $x \in F$, 0 for $x \in [a,b] \setminus F$. Then $f(b) \le f(a) + \int_a^b f^*$.

P (i) Because F is a Borel set and f' is a Borel measurable function (225J), f^* is measurable. Let $\epsilon > 0$. Let G be an open subset of $\mathbb{R}$ such that $f[[a,b] \setminus F] \subseteq G$ and $\mu G \le \epsilon$ (134Fa). Let $g : \mathbb{R} \to [0,\infty]$ be a lower semi-continuous function such that $f^*(x) \le g(x)$ for every $x \in [a,b]$ and $\int_a^b g \le \int_a^b f^* + \epsilon$ (225I). Consider

$$A = \{x : a \le x \le b,\ \mu([f(a), f(x)] \setminus G) \le 2\epsilon(x-a) + \int_a^x g\},$$

interpreting $[f(a), f(x)]$ as $\emptyset$ if $f(x) < f(a)$. Then $a \in A \subseteq [a,b]$, so $c = \sup A$ is defined and belongs to $[a,b]$.

Because f is continuous, the function $x \mapsto \mu([f(a), f(x)] \setminus G)$ is continuous; also $x \mapsto 2\epsilon(x-a) + \int_a^x g$ is certainly continuous, so $c \in A$.

(ii) ? If $c \in F$, so that $f^*(c) = |f'(c)|$, then there is a $\delta > 0$ such that

$$a \le c - \delta \le c + \delta \le b,$$

$$g(x) \ge g(c) - \epsilon \ge |f'(c)| - \epsilon \text{ whenever } |x-c| \le \delta,$$

$$|\tfrac{f(x)-f(c)}{x-c} - f'(c)| \le \epsilon \text{ whenever } |x-c| \le \delta.$$

Consider $x = c + \delta$. Then $c < x \le b$ and

$$\begin{aligned}\mu([f(a), f(x)] \setminus G) &\le \mu([f(a), f(c)] \setminus G) + |f(x) - f(c)| \\ &\le 2\epsilon(c-a) + \int_a^c g + \epsilon(x-c) + |f'(c)|(x-c) \\ &\le 2\epsilon(c-a) + \int_a^c g + \epsilon(x-c) + \int_c^x (g+\epsilon)\end{aligned}$$

(because $g(t) \ge |f'(c)| - \epsilon$ whenever $c \le t \le x$)

$$= 2\epsilon(x-a) + \int_a^x g,$$

so $x \in A$; but c is supposed to be an upper bound of A. **X**

Thus $c \in [a,b] \setminus F$.

(iii) ? Now suppose, if possible, that $c < b$. We know that $f(c) \in G$, so there is an $\eta > 0$ such that $[f(c)-\eta, f(c)+\eta] \subseteq G$; now there is a $\delta > 0$ such that $|f(x) - f(c)| \le \eta$ whenever $x \in [a,b]$ and $|x-c| \le \delta$. Set $x = \min(c+\delta, b)$; then $c < x \le b$ and $[f(c), f(x)] \subseteq G$, so

$$\mu([f(a), f(x)] \setminus G) = \mu([f(a), f(c)] \setminus G) \le 2\epsilon(c-a) + \int_a^c g \le 2\epsilon(x-a) + \int_a^x g$$

and once again $x \in A$, even though $x > \sup A$. **X**

(iv) We conclude that $c = b$, so that $b \in A$. But this means that

$$\begin{aligned}f(b) - f(a) &\le \mu([f(a), f(b)]) \le \mu([f(a), f(b)] \setminus G) + \mu G \\ &\le 2\epsilon(b-a) + \int_a^b g + \epsilon \le 2\epsilon(b-a) + \int_a^b f^* + \epsilon + \epsilon \\ &= 2\epsilon(1+b-a) + \int_a^b f^*.\end{aligned}$$

As ϵ is arbitrary, $f(b) - f(a) \le \int_a^b f^*$, as claimed. **Q**

(c) Similarly, or applying (b) to $-f$, $f(a) - f(b) \le \int_a^b f^*$, so that $|f(b) - f(a)| \le \int_a^b f^*$.

Of course the argument applies equally to any subinterval of $[a,b]$, so $|f(d) - f(c)| \le \int_c^d f^*$ whenever $a \le c \le d \le b$. Now let $\epsilon > 0$. By 225A once more, there is a $\delta > 0$ such that $\int_E f^* \le \epsilon$ whenever $E \subseteq [a,b]$ and $\mu E \le \delta$. Suppose that $a \le a_1 \le b_1 \le \ldots \le a_n \le b_n \le b$ and $\sum_{i=1}^n b_i - a_i \le \delta$. Then

$$\sum_{i=1}^n |f(b_i) - f(a_i)| \le \sum_{i=1}^n \int_{a_i}^{b_i} f^* = \int_{\bigcup_{i \le n}[a_i,b_i]} f^* \le \epsilon.$$

So f is absolutely continuous, as claimed.

225L Corollary Let $[a,b]$ be a non-empty closed interval in $\mathbb{R}$. Let $f : [a,b] \to \mathbb{R}$ be a continuous function which is differentiable on the open interval $]a,b[$. If its derivative f' is integrable over $[a,b]$, then f is absolutely continuous, and $f(b) - f(a) = \int_a^b f'$.

proof $f[\,[a,b] \setminus F] = \{f(a), f(b)\}$ is surely negligible, so f is absolutely continuous, by 225K; consequently $f(b) - f(a) = \int_a^b f'$, by 225E.

225M Corollary Let $[a,b]$ be a non-empty closed interval in $\mathbb{R}$, and $f:[a,b]\to\mathbb{R}$ a continuous function. Then f is absolutely continuous iff it is continuous and of bounded variation and $f[A]$ is negligible for every negligible $A\subseteq[a,b]$.

proof (a) Suppose that f is absolutely continuous. By 225C(a-b) it is continuous and of bounded variation, and by 225G we have $f[A]$ negligible for every negligible $A\subseteq[a,b]$.

(b) So now suppose that f satisfies the conditions. Set $F=\{x:x\in\,]a,b[\,,\ f'(x)\text{ is defined}\}$. By 224I, $[a,b]\setminus F$ is negligible, so $f[\,[a,b]\setminus F]$ is negligible. Moreover, also by 224I, f' is integrable over $[a,b]$ or F. So the conditions of 225K are satisfied and f is absolutely continuous.

225N The Cantor function I should mention the standard example of a continuous function of bounded variation which is not absolutely continuous. Let $C\subseteq[0,1]$ be the Cantor set (134G). Recall that the 'Cantor function' is a non-decreasing continuous function $f:[0,1]\to[0,1]$ such that $f'(x)$ is defined and equal to zero for every $x\in[0,1]\setminus C$, but $f(0)=0<1=f(1)$ (134H). Of course f is of bounded variation and not absolutely continuous. C is negligible and $f[C]=[0,1]$ is not. If $x\in C$, then for every $n\in\mathbb{N}$ there is an interval of length 3^{-n}, containing x, on which f increases by 2^{-n}; so f cannot be differentiable at x, and the set $F=\operatorname{dom}f'$ of 225K is precisely $[0,1]\setminus C$, so that $f[\,[0,1]\setminus F]=[0,1]$.

225O Complex-valued functions As usual, I spell out the results above in the forms applicable to complex-valued functions.

(a) Let (X,Σ,μ) be any measure space and f an integrable complex-valued function defined on a conegligible subset of X. Then for any $\epsilon>0$ there are a measurable set E of finite measure and a real number $\delta>0$ such that $\int_F|f|\le\epsilon$ whenever $F\in\Sigma$ and $\mu(F\cap E)\le\delta$. (Apply 225A to $|f|$.)

(b) If $[a,b]$ is a non-empty closed interval in $\mathbb{R}$ and $f:[a,b]\to\mathbb{C}$ is a function, we say that f is **absolutely continuous** if for every $\epsilon>0$ there is a $\delta>0$ such that $\sum_{i=1}^n|f(b_i)-f(a_i)|\le\epsilon$ whenever $a\le a_1\le b_1\le a_2\le b_2\le\ldots\le a_n\le b_n\le b$ and $\sum_{i=1}^n b_i-a_i\le\delta$. Observe that f is absolutely continuous iff its real and imaginary parts are both absolutely continuous.

(c) Let $[a,b]$ be a non-empty closed interval in $\mathbb{R}$.

(i) If $f:[a,b]\to\mathbb{C}$ is absolutely continuous it is of bounded variation on $[a,b]$, so is differentiable almost everywhere in $[a,b]$, and its derivative is integrable over $[a,b]$.

(ii) If f, $g:[a,b]\to\mathbb{C}$ are absolutely continuous, so are $f+g$ and ζf, for any $\zeta\in\mathbb{C}$, and $f\times g$.

(iii) If $g:[a,b]\to[c,d]$ is monotonic and absolutely continuous, and $f:[c,d]\to\mathbb{C}$ is absolutely continuous, then $fg:[a,b]\to\mathbb{C}$ is absolutely continuous.

(d) Let $[a,b]$ be a non-empty closed interval in $\mathbb{R}$ and $F:[a,b]\to\mathbb{C}$ a function. Then the following are equiveridical:

(i) there is an integrable complex-valued function f such that $F(x)=F(a)+\int_a^x f$ for every $x\in[a,b]$;

(ii) $\int_a^x F'$ exists and is equal to $F(x)-F(a)$ for every $x\in[a,b]$;

(iii) F is absolutely continuous.

(Apply 225E to the real and imaginary parts of F.)

(e) Let f be an integrable complex-valued function on an interval $[a,b]\subseteq\mathbb{R}$, and $g:[a,b]\to\mathbb{C}$ an absolutely continuous function. Set $F(x)=\int_a^x f$ for $x\in[a,b]$. Then

$$\int_a^b f\times g=F(b)g(b)-F(a)g(a)-\int_a^b F\times g'.$$

(Apply 225F to the real and imaginary parts of f and g.)

(f) Let f be a continuous complex-valued function on a closed interval $[a,b]\subseteq\mathbb{R}$, and suppose that f is differentiable at every point of the open interval $]a,b[$, with f' integrable over $[a,b]$. Then f is absolutely continuous. (Apply 225L to the real and imaginary parts of f.)

(g) For a result corresponding to 225M, see 264Yp.

225X Basic exercises (a) Show directly from the definition in 225B (without appealing to 225E) that any absolutely continuous real-valued function on a closed interval $[a,b]$ is expressible as the difference of non-decreasing absolutely continuous functions.

(b) Let $f:[a,b]\to\mathbb{R}$ be an absolutely continuous function, where $a\le b$. (i) Show that $|f|:[a,b]\to\mathbb{R}$ is absolutely continuous. (ii) Show that gf is absolutely continuous whenever $g:\mathbb{R}\to\mathbb{R}$ is a differentiable function with bounded derivative.

(c) Show directly from the definition in 225B and the Mean Value Theorem (without appealing to 225K) that if a function f is continuous on a closed interval $[a,b]$, differentiable on the open interval $]a,b[$, and has bounded derivative in $]a,b[$, then f is absolutely continuous, so that $f(x)=f(a)+\int_a^x f'$ for every $x\in[a,b]$.

(d) Show that if $f:[a,b]\to\mathbb{R}$ is absolutely continuous, then $\operatorname{Var} f=\int_a^b|f'|$. (*Hint*: put 224I and 225E together.)

(e) Let $f:[0,\infty[\,\to\mathbb{C}$ be a function which is absolutely continuous on $[0,a]$ for every $a\in[0,\infty[$ and has Laplace transform $F(s)=\int_0^\infty e^{-sx}f(x)dx$ defined on $\{s:\mathcal{R}\mathrm{e}\,s>S\}$. Suppose also that $\lim_{x\to\infty}e^{-Sx}f(x)=0$. Show that f' has Laplace transform $sF(s)-f(0)$ defined whenever $\mathcal{R}\mathrm{e}\,s>S$. (*Hint*: show that

$$f(x)e^{-sx}-f(0)=\int_0^x\frac{d}{dt}(f(t)e^{-st})dt$$

for every $x\ge0$.)

(f) Let $g:\mathbb{R}\to\mathbb{R}$ be a non-decreasing function which is absolutely continuous on every bounded interval; let μ_g be the associated Lebesgue-Stieltjes measure (114Xa), and Σ_g its domain. Show that $\int_E g'=\mu_g E$ for any $E\in\Sigma_g$, if we allow ∞ as a value of the integral. (*Hint*: start with intervals E.)

(g) Let $g:[a,b]\to\mathbb{R}$ be a non-decreasing absolutely continuous function, and $f:[g(a),g(b)]\to\mathbb{R}$ a continuous function. Show that $\int_{g(a)}^{g(b)}f(t)dt=\int_a^b f(g(t))g'(t)dt$. (*Hint*: set $F(x)=\int_{g(a)}^x f$, $G=Fg$ and consider $\int_a^b G'(t)dt$. See also 263I.)

(h) Suppose that $I\subseteq\mathbb{R}$ is any non-trivial interval (bounded or unbounded, open, closed or half-open, but not empty or a singleton), and $f:I\to\mathbb{R}$ a function. Show that f is absolutely continuous on every closed bounded subinterval of I iff there is a function g such that $\int_a^b g=f(b)-f(a)$ whenever $a\le b$ in I, and in this case g is integrable iff f is of bounded variation on I.

(i) Show that $\int_0^1\frac{\ln x}{x-1}dx=\sum_{n=1}^\infty\frac{1}{n^2}$. (*Hint*: use 225F to find $\int_0^1 x^n\ln x\,dx$, and recall that $\frac{1}{1-x}=\sum_{n=0}^\infty x^n$ for $0\le x<1$.)

(j)(i) Show that $\int_0^1 t^a dt$ is finite for every $a>-1$. (ii) Show that $\int_1^\infty t^a e^{-t}dt$ is finite for every $a\in\mathbb{R}$. (*Hint*: show that there is an M such that $t^a\le Me^{t/2}$ for $t\ge M$.) (iii) Show that $\Gamma(a)=\int_0^\infty t^{a-1}e^{-t}dt$ is defined for every $a>0$. (iv) Show that $\Gamma(a+1)=a\Gamma(a)$ for every $a>0$. (v) Show that $\Gamma(n+1)=n!$ for every $n\in\mathbb{N}$.

(Γ is of course the **gamma function**.)

(k) Show that if $b>0$ then $\int_0^\infty u^{b-1}e^{-u^2/2}du=2^{(b-2)/2}\Gamma(\frac{b}{2})$. (*Hint*: consider $f(t)=t^{(b-2)/2}e^{-t}$, $g(u)=u^2/2$ in 225Xg.)

(l) Suppose that f, g are lower semi-continuous functions, defined on subsets of $\mathbb{R}^r$, and taking values in $]-\infty,\infty]$. (i) Show that $f+g$, $f\wedge g$ and $f\vee g$ are lower semi-continuous, and that αf is lower semi-continuous for every $\alpha\ge0$. (ii) Show that if f and g are non-negative, then $f\times g$ is lower semi-continuous. (iii) Show that if f is non-negative and g is continuous, then $f\times g$ is lower semi-continuous. (iv) Show that if f is non-decreasing then the composition fg is lower semi-continuous.

(m) Let A be a non-empty family of lower semi-continuous functions defined on subsets of $\mathbb{R}^r$ and taking values in $[-\infty,\infty]$. Set $g(x)=\sup\{f(x):f\in A,\,x\in\operatorname{dom}f\}$ for $x\in D=\bigcup_{f\in A}\operatorname{dom}f$. Show that g is lower semi-continuous.

(n) Suppose that $f:[a,b]\to\mathbb{R}$ is continuous, and differentiable at all but countably many points of $[a,b]$. Show that f is absolutely continuous iff it is of bounded variation.

(o) Show that if $f:[a,b]\to\mathbb{R}$ is absolutely continuous, then $f[E]$ is Lebesgue measurable for every Lebesgue measurable set $E\subseteq[a,b]$.

225Y Further exercises (a) Show that the composition of two absolutely continuous functions need not be absolutely continuous. (*Hint*: 224Xb.)

(b) Let $f : [a,b] \to \mathbb{R}$ be a continuous function, where $a < b$. Set $G = \{x : x \in \,]a,b[\,, \exists\, y \in \,]x,b]$ such that $f(x) < f(y)\}$. Show that G is open and is expressible as a disjoint union of intervals $]c,d[$ where $f(c) \le f(d)$. Use this to prove 225D without calling on Vitali's theorem.

(c) Let $f : [a,b] \to \mathbb{R}$ be a function of bounded variation and $\gamma > 0$. Show that there is an absolutely continuous function $g : [a,b] \to \mathbb{R}$ such that $|g'(x)| \le \gamma$ wherever the derivative is defined and $\{x : x \in [a,b],\, f(x) \ne g(x)\}$ has measure at most $\gamma^{-1} \operatorname{Var} f$. (*Hint*: reduce to the case of non-decreasing f. Apply 225Yb to the function $x \mapsto f(x) - \gamma x$ and show that $\gamma\mu G \le \operatorname{Var}_{[a,b]}(f)$. Set $g(x) = f(x)$ for $x \in \,]a,b[\, \setminus G$.)

(d) Let f be a non-negative measurable real-valued function defined on a subset D of $\mathbb{R}^r$, where $r \ge 1$. Show that for any $\epsilon > 0$ there is a lower semi-continuous function $g : \mathbb{R}^r \to [-\infty, \infty]$ such that $g(x) \ge f(x)$ for every $x \in D$ and $\int_D g - f \le \epsilon$.

(e) Let f be a measurable real-valued function defined on a subset D of $\mathbb{R}^r$, where $r \ge 1$. Show that for any $\epsilon > 0$ there is a lower semi-continuous function $g : \mathbb{R}^r \to [-\infty, \infty]$ such that $g(x) \ge f(x)$ for every $x \in D$ and $\mu^*\{x : x \in D,\, g(x) > f(x)\} \le \epsilon$. (*Hint*: 134Yd, 134Fb.)

(f)(i) Show that if f is a Lebesgue measurable real function then all its Dini derivates are Lebesgue measurable. (ii) Show that if f is a Borel measurable real function then all its Dini derivates are Borel measurable.

225 Notes and comments There is a good deal more to say about absolutely continuous functions; I will return to the topic in the next section and in Chapter 26. I shall not make direct use of any of the results from 225H on, but it seems to me that this kind of investigation is necessary for any clear picture of the relationships between such concepts as absolute continuity and bounded variation. Of course, in order to apply these results, we do need a store of simple kinds of absolutely continuous function, differentiable functions with bounded derivative forming the most important class (225Xc). A larger family of the same kind is the class of 'Lipschitz' functions (262Bc).

The definition of 'absolutely continuous function' is ordinarily set out for closed bounded intervals, as in 225B. The point is that for other intervals the simplest generalizations of this formulation do not seem quite appropriate. In 225Xh I try to suggest the kind of demands one might make on functions defined on other types of interval.

I should remark that the real prize is still not quite within our grasp. I have been able to give a reasonably satisfactory formulation of simple integration by parts (225F), at least for bounded intervals – a further limiting process is necessary to deal with unbounded intervals. But a companion method from advanced calculus, integration by substitution, remains elusive. The best I think we can do at this point is 225Xg, which insists on a continuous integrand f. It is the case that the result is valid for general integrable f, but there are some further subtleties to be mastered on the way; the necessary ideas are given in the much more general results 235A and 263D below, and applied to the one-dimensional case in 263I.

On the way to the characterization of absolutely continuous functions in 225K, I find myself calling on one of the fundamental relationships between Lebesgue measure and the topology of $\mathbb{R}^r$ (225I). The technique here can be adapted to give many variations of the result; see 225Yd-225Ye. If you have not seen semi-continuous functions before, 225Xl-225Xm give a partial idea of their properties. In 225J I give a fragment of 'descriptive set theory', the study of the kinds of set which can arise from the formulae of analysis. These ideas too will re-surface elsewhere; compare 225Yf and also the proof of 262M below.

226 The Lebesgue decomposition of a function of bounded variation

I end this chapter with some notes on a method of analysing a general function of bounded variation which may help to give a picture of what such functions can be, though it is not directly necessary for anything of great importance dealt with in this volume.

226A Sums over arbitrary index sets To get a full picture of this fragment of real analysis, a bit of preparation will be helpful. This concerns the notion of a sum over an arbitrary index set, which I have rather been skirting around so far.

(a) If I is any set and $\langle a_i \rangle_{i \in I}$ any family in $[0, \infty]$, we set

$$\sum_{i\in I} a_i = \sup\{\sum_{i\in K} a_i : K \text{ is a finite subset of } I\},$$

with the convention that $\sum_{i\in\emptyset} a_i = 0$. (See 112Bd, 222Ba.) For general $a_i \in [-\infty, \infty]$, we can set

$$\sum_{i\in I} a_i = \sum_{i\in I} a_i^+ - \sum_{i\in I} a_i^-$$

if this is defined in $[-\infty, \infty]$, that is, at least one of $\sum_{i\in I} a_i^+$, $\sum_{i\in I} a_i^-$ is finite, where $a^+ = \max(a, 0)$ and $a^- = \max(-a, 0)$ for each a. If $\sum_{i\in I} a_i$ is defined and finite, we say that $\langle a_i\rangle_{i\in I}$ is **summable**.

(b) Since this is a book on measure theory, I will immediately describe the relationship between this kind of summability and an appropriate notion of integration. For any set I, we have the corresponding 'counting measure' μ on I (112Bd). Every subset of I is measurable, so every family $\langle a_i\rangle_{i\in I}$ of real numbers is a measurable real-valued function on I. A subset of I has finite measure iff it is finite; so a real-valued function f on I is 'simple' if $K = \{i : f(i) \neq 0\}$ is finite. In this case,

$$\int f d\mu = \sum_{i\in K} f(i) = \sum_{i\in I} f(i)$$

as defined in part (a). The measure μ is semi-finite (211Nc) so a non-negative function f is integrable iff $\int f = \sup_{\mu K<\infty} \int_K f$ is finite (213B); but of course this supremum is precisely

$$\sup\{\sum_{i\in K} f(i) : K \subseteq I \text{ is finite}\} = \sum_{i\in I} f(i).$$

Now a general function $f : I \to \mathbb{R}$ is integrable iff it is measurable and $\int |f| d\mu < \infty$, that is, iff $\sum_{i\in I} |f(i)| < \infty$, and in this case

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu = \sum_{i\in I} f(i)^+ - \sum_{i\in I} f(i)^- = \sum_{i\in I} f(i),$$

writing $f^\pm(i) = f(i)^\pm$ for each i. Thus we have

$$\sum_{i\in I} a_i = \int_I a_i \mu(di),$$

and the standard rules under which we allow ∞ as the value of an integral (133A, 135F) match well with the interpretations in (a) above.

(c) Accordingly, and unsurprisingly, the operation of summation is a linear operation on the linear space of summable families of real numbers.

I observe here that this notion of summability is 'absolute'; a family $\langle a_i\rangle_{i\in I}$ is summable iff it is absolutely summable. This is necessary because it must also be 'unconditional'; we have no structure on an arbitrary set I to guide us to take the sum in any particular order. See 226Xf. In particular, I distinguish between '$\sum_{n\in\mathbb{N}} a_n$', which in this book will always be interpreted as in 226A above, and '$\sum_{n=0}^\infty a_n$ which (if it makes a difference) should be interpreted as $\lim_{m\to\infty} \sum_{n=0}^m a_n$. So, for instance, $\sum_{n=0}^\infty \frac{(-1)^n}{n+1} = \ln 2$, while $\sum_{n\in\mathbb{N}} \frac{(-1)^n}{n+1}$ is undefined. Of course $\sum_{n=0}^\infty a_n = \sum_{n\in\mathbb{N}} a_n$ whenever the latter is defined in $[-\infty, \infty]$.

(d) There is another, and very important, approach to the sum described here. If $\langle a_i\rangle_{i\in I}$ is an (absolutely) summable family of real numbers, then for every $\epsilon > 0$ there is a finite $K \subseteq I$ such that $\sum_{i\in I\setminus K} |a_i| \le \epsilon$. **P** This is nothing but a special case of 225A; there is a set K with $\mu K < \infty$ such that $\int_{I\setminus K} |a_i|\mu(di) \le \epsilon$, but

$$\int_{I\setminus K} |a_i|\mu(di) = \sum_{i\in I\setminus K} |a_i|. \quad \mathbf{Q}$$

(Of course there are 'direct' proofs of this result from the definition in (a), not mentioning measures or integrals. But I think you will see that they rely on the same idea as that in the proof of 225A.) Consequently, for any family $\langle a_i\rangle_{i\in I}$ of real numbers and any $s \in \mathbb{R}$, the following are equiveridical:

(i) $\sum_{i\in I} a_i = s$;

(ii) for every $\epsilon > 0$ there is a finite $K \subseteq I$ such that $|s - \sum_{i\in J} a_i| \le \epsilon$ whenever J is finite and $K \subseteq J \subseteq I$.

P (i)⇒(ii) Take K such that $\sum_{i\in I\setminus K} |a_i| \le \epsilon$. If $K \subseteq J \subseteq I$, then

$$|s - \sum_{i\in J} a_i| = |\sum_{i\in I\setminus J} a_i| \le \sum_{i\in I\setminus K} |a_i| \le \epsilon.$$

(ii)⇒(i) Let $\epsilon > 0$, and let $K \subseteq I$ be as described in (ii). If $J \subseteq I\setminus K$ is any finite set, then set $J_1 = \{i : i \in J, a_i \ge 0\}$, $J_2 = J \setminus J_1$. We have

$$\sum_{i\in J}|a_i| = |\sum_{i\in J_1\cup K} a_i - \sum_{i\in J_2\cup K} a_i|$$
$$\leq |s - \sum_{i\in J_1\cup K} a_i| + |s - \sum_{i\in J_2\cup K} a_i| \leq 2\epsilon.$$

As J is arbitrary, $\sum_{i\in I\setminus K}|a_i| \leq 2\epsilon$ and

$$\sum_{i\in I}|a_i| \leq \sum_{i\in K}|a_i| + 2\epsilon < \infty.$$

Accordingly $\sum_{i\in I} a_i$ is well-defined in $\mathbb{R}$. Also

$$|s - \sum_{i\in I} a_i| \leq |s - \sum_{i\in K} a_i| + |\sum_{i\in I\setminus K} a_i| \leq \epsilon + \sum_{i\in I\setminus K}|a_i| \leq 3\epsilon.$$

As ϵ is arbitrary, $\sum_{i\in I} a_i = s$, as required. **Q**

In this way, we express $\sum_{i\in I} a_i$ directly as a limit; we could write it as

$$\sum_{i\in I} a_i = \lim_{K\uparrow I}\sum_{i\in K} a_i,$$

on the understanding that we look at finite sets K in the right-hand formula.

(e) Yet another approach is through the following fact. If $\sum_{i\in I}|a_i| < \infty$, then for any $\delta > 0$ the set $\{i : |a_i| \geq \delta\}$ is finite, indeed can have at most $\frac{1}{\delta}\sum_{i\in I}|a_i|$ members; consequently

$$J = \{i : a_i \neq 0\} = \bigcup_{n\in\mathbb{N}}\{i : |a_i| \geq 2^{-n}\}$$

is countable (1A1F). If J is finite, then of course $\sum_{i\in I} a_i = \sum_{i\in J} a_i$ reduces to a finite sum. Otherwise, we can enumerate J as $\langle j_n\rangle_{n\in\mathbb{N}}$, and we shall have

$$\sum_{i\in I} a_i = \sum_{i\in J} a_i = \lim_{n\to\infty}\sum_{k=0}^{n} a_{j_k} = \sum_{n=0}^{\infty} a_{j_n}$$

(using (d) to reduce the sum $\sum_{i\in J} a_i$ to a limit of finite sums). Conversely, if $\langle a_i\rangle_{i\in I}$ is such that there is a countably infinite $J \subseteq \{i : a_i \neq 0\}$ enumerated as $\langle j_n\rangle_{n\in\mathbb{N}}$, and if $\sum_{n=0}^{\infty}|a_{j_n}| < \infty$, then $\sum_{i\in I} a_i$ will be $\sum_{n=0}^{\infty} a_{j_n}$.

(f) It will be useful later to have a fragment of general theory. Let I and J be sets and $\langle a_{ij}\rangle_{i\in I,j\in J}$ a family in $[0,\infty]$. Then

$$\sum_{(i,j)\in I\times J} a_{ij} = \sum_{i\in I}(\sum_{j\in J} a_{ij}) = \sum_{j\in J}(\sum_{i\in I} a_{ij}).$$

P (i) If $\sum_{(i,j)\in I\times J} a_{ij} > u$, then there is a finite set $M \subseteq I\times J$ such that $\sum_{(i,j)\in M} a_{ij} > u$. Now $K = \{i : (i,j) \in M\}$ and $L = \{j : (i,j) \in M\}$ are finite, so

$$\sum_{i\in I}\sum_{j\in J} a_{ij} \geq \sum_{i\in K}\sum_{j\in J} a_{ij} \geq \sum_{i\in K}\sum_{j\in L} a_{ij}$$

(because $\sum_{j\in J} a_{ij} \geq \sum_{j\in L} a_{ij}$ for every i)

$$= \sum_{(i,j)\in K\times L} a_{ij} \geq \sum_{(i,j)\in M} a_{ij} > u.$$

As u is arbitrary, $\sum_{i\in I}\sum_{j\in J} a_{ij} \geq \sum_{(i,j)\in I\times J} a_{ij}$. (ii) If $\sum_{i\in I}\sum_{j\in J} a_{ij} > u$, there is a finite set $K \subseteq I$ such that $\sum_{i\in K}\sum_{j\in J} a_{ij} > u$. Let $\epsilon \in \,]0,1[$ be such that $\sum_{i\in K}\sum_{j\in J} a_{ij} > u + \epsilon$, and set $\delta = \frac{\epsilon}{\#(K)}$. For each $i \in K$ set $\gamma_i = \min(u+1, \sum_{j\in J} a_{ij}) - \delta$; then

$$\epsilon + \sum_{i\in K}\gamma_i = \sum_{i\in K}\min(u+1, \sum_{j\in J} a_{ij}) \geq \min(u+1, \sum_{i\in K}\sum_{j\in J} a_{ij}) > u + \epsilon,$$

so $\sum_{i\in K}\gamma_i > u$. For each $i \in K$, $\gamma_i < \sum_{j\in J} a_{ij}$, so there is a finite $L_i \subseteq J$ such that $\sum_{j\in L_i} a_{ij} \geq \gamma_i$. Set $M = \{(i,j) : i \in K,\, j \in L_i\}$, so that M is a finite subset of $I\times J$; then

$$\sum_{(i,j)\in I\times J} a_{ij} \geq \sum_{(i,j)\in M} a_{ij} = \sum_{i\in K}\sum_{j\in L_i} a_{ij} \geq \sum_{i\in K}\gamma_i > u.$$

As u is arbitrary, $\sum_{(i,j)\in I\times J} a_{ij} \geq \sum_{i\in I}\sum_{j\in J} a_{ij}$ and these two sums are equal. (iii) Similarly, $\sum_{(i,j)\in I\times J} a_{ij} = \sum_{j\in J}\sum_{i\in I} a_{ij}$. **Q**

226B Saltus functions Now we are ready for a special type of function of bounded variation on $\mathbb{R}$. Suppose that $a < b$ in $\mathbb{R}$.

(a) A (real) **saltus function** on $[a,b]$ is a function $F:[a,b]\to\mathbb{R}$ expressible in the form

$$F(x)=\sum_{t\in[a,x[}u_t+\sum_{t\in[a,x]}v_t$$

for $x\in[a,b]$, where $\langle u_t\rangle_{t\in[a,b[}$, $\langle v_t\rangle_{t\in[a,b]}$ are real-valued families such that $\sum_{t\in[a,b[}|u_t|$ and $\sum_{t\in[a,b]}|v_t|$ are finite.

(b) For any function $F:[a,b]\to\mathbb{R}$ we can write

$$F(x^+)=\lim_{y\downarrow x}F(y) \text{ if } x\in[a,b[\text{ and the limit exists,}$$

$$F(x^-)=\lim_{y\uparrow x}F(y) \text{ if } x\in\,]a,b] \text{ and the limit exists.}$$

(I hope that this will not lead to confusion with the alternative interpretation of x^+ as $\max(x,0)$.) Observe that if F is a saltus function, as defined in (b), with associated families $\langle u_t\rangle_{t\in[a,b[}$ and $\langle v_t\rangle_{t\in[a,b]}$, then $v_a=F(a)$, $v_x=F(x)-F(x^-)$ for $x\in\,]a,b]$, $u_x=F(x^+)-F(x)$ for $x\in[a,b[$. **P** Let $\epsilon>0$. As remarked in 226Ad, there is a finite $K\subseteq[a,b]$ such that

$$\sum_{t\in[a,b[\setminus K}|u_t|+\sum_{t\in[a,b]\setminus K}|v_t|\le\epsilon.$$

Given $x\in[a,b]$, let $\delta>0$ be such that $[x-\delta,x+\delta]$ contains no point of K except perhaps x. In this case, if $\max(a,x-\delta)\le y<x$, we must have

$$\begin{aligned}|F(y)-(F(x)-v_x)|&=|\sum_{t\in[y,x[}u_t+\sum_{t\in]y,x[}v_t|\\&\le\sum_{t\in[a,b[\setminus K}|u_t|+\sum_{t\in[a,b]\setminus K}|v_t|\le\epsilon,\end{aligned}$$

while if $x<y\le\min(b,x+\delta)$ we shall have

$$\begin{aligned}|F(y)-(F(x)+u_x)|&=|\sum_{t\in]x,y[}u_t+\sum_{t\in]x,y]}v_t|\\&\le\sum_{t\in[a,b[\setminus K}|u_t|+\sum_{t\in[a,b]\setminus K}|v_t|\le\epsilon.\end{aligned}$$

As ϵ is arbitrary, we get $F(x^-)=F(x)-v_x$ (if $x>a$) and $F(x^+)=F(x)+u_x$ (if $x<b$). **Q**

It follows that F is continuous at $x\in\,]a,b[$ iff $u_x=v_x=0$, while F is continuous at a iff $u_a=0$ and F is continuous at b iff $v_b=0$. In particular, $\{x:x\in[a,b],\ F \text{ is not continuous at } x\}$ is countable (see 226Ae).

(c) If F is a saltus function defined on $[a,b]$, with associated families $\langle u_t\rangle_{t\in[a,b[}$, $\langle v_t\rangle_{t\in[a,b]}$, then F is of bounded variation on $[a,b]$, and

$$\mathrm{Var}_{[a,b]}(F)\le\sum_{t\in[a,b[}|u_t|+\sum_{t\in]a,b]}|v_t|.$$

P If $a\le x<y\le b$, then

$$F(y)-F(x)=u_x+\sum_{t\in]x,y[}(u_t+v_t)+v_y,$$

so

$$|F(y)-F(x)|\le\sum_{t\in[x,y[}|u_t|+\sum_{t\in]x,y]}|v_t|.$$

If $a\le a_0\le a_1\le\ldots\le a_n\le b$, then

$$\begin{aligned}\sum_{i=1}^n|F(a_i)-F(a_{i-1})|&\le\sum_{i=1}^n\Bigl(\sum_{t\in[a_{i-1},a_i[}|u_t|+\sum_{t\in]a_{i-1},a_i]}|v_t|\Bigr)\\&\le\sum_{t\in[a,b[}|u_t|+\sum_{t\in]a,b]}|v_t|.\end{aligned}$$

Consequently

$$\mathrm{Var}_{[a,b]}(F)\le\sum_{t\in[a,b[}|u_t|+\sum_{t\in]a,b]}|v_t|<\infty. \textbf{ Q}$$

(d) The inequality in (c) is actually an equality. To see this, note first that if $a\le x<y\le b$, then $\mathrm{Var}_{[x,y]}(F)\ge|u_x|+|v_y|$. **P** I noted in (b) that $u_x=\lim_{t\downarrow x}F(t)-F(x)$ and $v_y=F(y)-\lim_{t\uparrow y}F(t)$. So, given $\epsilon>0$, we can find t_1, t_2 such that $x<t_1\le t_2\le y$ and

$$|F(t_1)-F(x)|\ge|u_x|-\epsilon,\quad |F(y)-F(t_2)|\ge|v_y|-\epsilon.$$

Now

$$\mathrm{Var}_{[x,y]}(F)\ge|F(t_1)-F(x)|+|F(t_2)-F(t_1)|+|F(y)-F(t_2)|\ge|u_x|+|v_y|-2\epsilon.$$

As ϵ is arbitrary, we have the result. **Q**

Now, given $a\le t_0<t_1<\ldots<t_n\le b$, we must have

$$\mathrm{Var}_{[a,b]}(F)\ge\sum_{i=1}^{n}\mathrm{Var}_{[t_{i-1},t_i]}(F)$$

(using 224Cc)

$$\ge\sum_{i=1}^{n}|u_{t_{i-1}}|+|v_{t_i}|.$$

As $t_0,\ldots,t_n$ are arbitrary,

$$\mathrm{Var}_{[a,b]}(F)\ge\sum_{t\in[a,b[}|u_t|+\sum_{t\in]a,b]}|v_t|,$$

as required.

(e) Because a saltus function is of bounded variation ((c) above), it is differentiable almost everywhere (224I). In fact its derivative is zero almost everywhere. **P** Let $F:[a,b]\to\mathbb{R}$ be a saltus function, with associated families $\langle u_t\rangle_{t\in[a,b[}$, $\langle v_t\rangle_{t\in[a,b]}$. Let $\epsilon>0$. Let $K\subseteq[a,b]$ be a finite set such that

$$\sum_{t\in[a,b[\setminus K}|u_t|+\sum_{t\in[a,b]\setminus K}|v_t|\le\epsilon.$$

Set

$$\begin{aligned}u'_t&=u_t\text{ if }t\in[a,b[\cap K,\\&=0\text{ if }t\in[a,b[\setminus K,\\v'_t&=v_t\text{ if }t\in K,\\&=0\text{ if }t\in[a,b]\setminus K,\\u''_t&=u_t-u'_t\text{ for }t\in[a,b[\,,\\v''_t&=v_t-v'_t\text{ for }t\in[a,b].\end{aligned}$$

Let G, H be the saltus functions corresponding to $\langle u'_t\rangle_{t\in[a,b[}$, $\langle v'_t\rangle_{t\in[a,b]}$ and $\langle u''_t\rangle_{t\in[a,b[}$ $\langle v''_t\rangle_{t\in[a,b]}$, so that $F=G+H$. Then $G'(t)=0$ for every $t\in]a,b[\setminus K$, since $]a,b[\setminus K$ comprises a finite number of open intervals on each of which G is constant. So $G'=0$ a.e. and $F'=_{\text{a.e.}}H'$. On the other hand,

$$\int_a^b|H'|\le\mathrm{Var}_{[a,b]}(H)=\sum_{t\in[a,b[\setminus K}|u_t|+\sum_{t\in]a,b]\setminus K}|v_t|\le\epsilon,$$

using 224I and (d) above. So

$$\int_a^b|F'|=\int_a^b|H'|\le\epsilon.$$

As ϵ is arbitrary, $\int_a^b|F'|=0$ and $F'=0$ a.e., as claimed. **Q**

226C The Lebesgue decomposition of a function of bounded variation Take a, $b\in\mathbb{R}$ with $a<b$.

(a) If $F:[a,b]\to\mathbb{R}$ is non-decreasing, set $v_a=0$, $v_t=F(t)-F(t^-)$ for $t\in]a,b]$, $u_t=F(t^+)-F(t)$ for $t\in[a,b[$, defining $F(t^+)$, $F(t^-)$ as in 226Bb. Then all the v_t, u_t are non-negative, and if $a<t_0<t_1<\ldots<t_n<b$, then

$$\sum_{i=0}^{n}(u_{t_i}+v_{t_i})=\sum_{i=0}^{n}(F(t_i^+)-F(t_i^-))\le F(b)-F(a).$$

Accordingly $\sum_{t\in[a,b[}u_t$ and $\sum_{t\in[a,b]}v_t$ are both finite. Let F_p be the corresponding saltus function, as defined in 226Ba, so that

$$F_p(x)=F(a^+)-F(a)+\sum_{t\in]a,x[}(F(t^+)-F(t^-))+F(x)-F(x^-)$$

if $a<x\le b$. If $a\le x<y\le b$ then

$$
\begin{aligned}
F_p(y) - F_p(x) &= F(x^+) - F(x) + \sum_{t \in]x,y[} (F(t^+) - F(t^-)) + F(y) - F(y^-) \\
&\le F(y) - F(x)
\end{aligned}
$$

because if $x = t_0 < t_1 < \ldots < t_n < t_{n+1} = y$ then

$$
\begin{aligned}
F(x^+) - F(x) &+ \sum_{i=1}^{n} (F(t_i^+) - F(t_i^-)) + F(y) - F(y^-) \\
&= F(y) - F(x) - \sum_{i=1}^{n+1} (F(t_i^-) - F(t_{i-1}^+)) \le F(y) - F(x).
\end{aligned}
$$

Accordingly both F_p and $F_c = F - F_p$ are non-decreasing. Also, because

$$F_p(a) = 0 = v_a,$$

$$F_p(t) - F_p(t^-) = v_t = F(t) - F(t^-) \text{ for } t \in]a, b],$$

$$F_p(t^+) - F_p(t) = u_t = F(t^+) - F(t) \text{ for } t \in [a, b[,$$

we shall have

$$F_c(a) = F(a),$$

$$F_c(t) = F_c(t^-) \text{ for } t \in]a, b],$$

$$F_c(t) = F_c(t^+) \text{ for } t \in [a, b[,$$

and F_c is continuous.

Clearly this expression of $F = F_p + F_c$ as the sum of a saltus function and a continuous function is unique, except that we can freely add a constant to one if we subtract it from the other.

(b) Still taking $F : [a, b] \to \mathbb{R}$ to be non-decreasing, we know that F' is integrable (222C); moreover, $F' =_{\text{a.e.}} F_c'$, by 226Be. Set $F_{ac}(x) = F(a) + \int_a^x F'$ for each $x \in [a, b]$. We have

$$F_{ac}(y) - F_{ac}(x) = \int_x^y F_c' \le F_c(y) - F_c(x)$$

for $a \le x \le y \le b$ (222C again), so $F_{cs} = F_c - F_{ac}$ is still non-decreasing; F_{ac} is continuous (225A), so F_{cs} is continuous; $F_{ac}' =_{\text{a.e.}} F' =_{\text{a.e.}} F_c'$ (222E), so $F_{cs}' = 0$ a.e.

Again, the expression of $F_c = F_{ac} + F_{cs}$ as the sum of an absolutely continuous function and a function with zero derivative almost everywhere is unique, except for the possibility of moving a constant from one to the other, because two absolutely continuous functions whose derivatives are equal almost everywhere must differ by a constant (225D).

(c) Putting all these together: if $F : [a, b] \to \mathbb{R}$ is any non-decreasing function, it is expressible as $F_p + F_{ac} + F_{cs}$, where F_p is a saltus function, F_{ac} is absolutely continuous, and F_{cs} is continuous and differentiable, with zero derivative, almost everywhere; all three components are non-decreasing; and the expression is unique if we say that $F_{ac}(a) = F(a)$, $F_p(a) = F_{cs}(a) = 0$.

The Cantor function $f : [0, 1] \to [0, 1]$ (134H) is continuous and $f' = 0$ a.e. (134Hb), so $f_p = f_{ac} = 0$ and $f = f_{cs}$. Setting $g(x) = \frac{1}{2}(x + f(x))$ for $x \in [0, 1]$, as in 134I, we get $g_p(x) = 0$, $g_{ac}(x) = \frac{x}{2}$ and $g_{cs}(x) = \frac{1}{2} f(x)$.

(d) Now suppose that $F : [a, b] \to \mathbb{R}$ is of bounded variation. Then it is expressible as a difference $G - H$ of non-decreasing functions (224D). So writing $F_p = G_p - H_p$, etc., we can express F as a sum $F_p + F_{cs} + F_{ac}$, where F_p is a saltus function, F_{ac} is absolutely continuous, F_{cs} is continuous, $F_{cs}' = 0$ a.e., $F_{ac}(a) = F(a)$, $F_{cs}(a) = F_p(a) = 0$. Under these conditions the expression is unique, because (for instance) $F_p(t^+) - F_p(t) = F(t^+) - F(t)$ for $t \in [a, b[$, while $F_{ac}' =_{\text{a.e.}} (F - F_p)' =_{\text{a.e.}} F'$.

This is a **Lebesgue decomposition** of the function F. (I have to say 'a' Lebesgue decomposition because of course the assignments $F_{ac}(a) = F(a)$, $F_p(a) = F_{cs}(a) = 0$ are arbitrary.) I will call F_p the **saltus part** of F.

226D Complex functions The modifications needed to deal with complex functions are elementary.

(a) If I is any set and $\langle a_j \rangle_{j \in I}$ is a family of complex numbers, then the following are equiveridical:

(i) $\sum_{j\in I}|a_j|<\infty$;

(ii) there is an $s\in\mathbb{C}$ such that for every $\epsilon>0$ there is a finite $K\subseteq I$ such that $|s-\sum_{j\in J}a_j|\le\epsilon$ whenever J is finite and $K\subseteq J\subseteq I$.

In this case

$$s=\sum_{j\in I}\mathcal{R}\mathrm{e}(a_j)+i\sum_{j\in I}\mathcal{I}\mathrm{m}(a_j)=\int_I a_j\mu(dj),$$

where μ is counting measure on I, and we write $s=\sum_{j\in I}a_j$.

(b) If $a<b$ in $\mathbb{R}$, a complex **saltus function** on $[a,b]$ is a function $F:[a,b]\to\mathbb{C}$ expressible in the form

$$F(x)=\sum_{t\in[a,x[}u_t+\sum_{t\in[a,x]}v_t$$

for $x\in[a,b]$, where $\langle u_t\rangle_{t\in[a,b[}$, $\langle v_t\rangle_{t\in[a,b]}$ are complex-valued families such that $\sum_{t\in[a,b[}|u_t|$ and $\sum_{t\in[a,b]}|v_t|$ are finite; that is, if the real and imaginary parts of F are saltus functions. In this case F is continuous except at countably many points and differentiable, with zero derivative, almost everywhere in $[a,b]$, and

$$u_x=\lim_{t\downarrow x}F(t)-F(x)\text{ for every }x\in[a,b[,$$

$$v_x=\lim_{t\uparrow x}F(x)-F(t)\text{ for every }x\in\,]a,b]$$

(apply the results of 226B to the real and imaginary parts of F). F is of bounded variation, and its variation is

$$\mathrm{Var}_{[a,b]}(F)=\sum_{t\in[a,b[}|u_t|+\sum_{t\in]a,b]}|v_t|$$

(repeat the arguments of 226Bc-d).

(c) If $F:[a,b]\to\mathbb{C}$ is a function of bounded variation, where $a<b$ in $\mathbb{R}$, it is uniquely expressible as $F=F_p+F_{cs}+F_{ac}$, where F_p is a saltus function, F_{ac} is absolutely continuous, F_{cs} is continuous and has zero derivative almost everywhere, and $F_{ac}(a)=F(a)$, $F_p(a)=F_{cs}(a)=0$. (Apply 226C to the real and imaginary parts of F.)

226E As an elementary exercise in the language of 226A, I interpolate a version of a theorem of B.Levi which is sometimes useful.

Proposition Let (X,Σ,μ) be a measure space, I a *countable* set, and $\langle f_i\rangle_{i\in I}$ a family of μ-integrable real- or complex-valued functions such that $\sum_{i\in I}\int|f_i|d\mu$ is finite. Then $f(x)=\sum_{i\in I}f_i(x)$ is defined almost everywhere and $\int fd\mu=\sum_{i\in I}\int f_id\mu$.

proof If I is finite this is elementary. Otherwise, since there must be a bijection between I and $\mathbb{N}$, we may take it that $I=\mathbb{N}$. Setting $g_n=\sum_{i=0}^n|f_i|$ for each n, we have a non-decreasing sequence $\langle g_n\rangle_{n\in\mathbb{N}}$ of integrable functions such that $\int g_n\le\sum_{i\in\mathbb{N}}\int|f_i|$ for every n, so that $g=\sup_{n\in\mathbb{N}}g_n$ is integrable, by B.Levi's theorem as stated in 123A. In particular, g is finite almost everywhere. Now if $x\in X$ is such that $g(x)$ is defined and finite, $\sum_{i\in J}|f_i(x)|\le g(x)$ for every finite $J\subseteq\mathbb{N}$, so $\sum_{i\in\mathbb{N}}|f_i(x)|$ and $\sum_{i\in\mathbb{N}}f_i(x)$ are defined. In this case, of course, $\sum_{i\in\mathbb{N}}f_i(x)=\lim_{n\to\infty}\sum_{i=0}^nf_i(x)$. But $|\sum_{i=0}^nf_i|\le_{\text{a.e.}}g$ for each n, so Lebesgue's Dominated Convergence Theorem tells us that

$$\int\sum_{i\in\mathbb{N}}f_i=\lim_{n\to\infty}\int\sum_{i=0}^nf_i=\lim_{n\to\infty}\sum_{i=0}^n\int f_i=\sum_{i\in\mathbb{N}}f_i.$$

226X Basic exercises >(a) A **step-function** on an interval $[a,b]$ is a function F such that, for suitable $t_0,\dots,t_n$ with $a=t_0\le\ldots\le t_n=b$, F is constant on each interval $]t_{i-1},t_i[$. Show that $F:[a,b]\to\mathbb{R}$ is a saltus function iff for every $\epsilon>0$ there is a step-function $G:[a,b]\to\mathbb{R}$ such that $\mathrm{Var}_{[a,b]}(F-G)\le\epsilon$.

(b) Let F, G be real-valued functions of bounded variation defined on an interval $[a,b]\subseteq\mathbb{R}$. Show that, in the language of 226C,

$$(F+G)_p=F_p+G_p,\quad(F+G)_c=F_c+G_c,$$

$$(F+G)_{cs}=F_{cs}+G_{cs},\quad(F+G)_{ac}=F_{ac}+G_{ac}.$$

>(c) Let F be a real-valued function of bounded variation on an interval $[a,b]\subseteq\mathbb{R}$. Show that, in the language of 226C,

$$\mathrm{Var}_{[a,b]}(F)=\mathrm{Var}_{[a,b]}(F_p)+\mathrm{Var}_{[a,b]}(F_c)=\mathrm{Var}_{[a,b]}(F_p)+\mathrm{Var}_{[a,b]}(F_{cs})+\mathrm{Var}_{[a,b]}(F_{ac}).$$

(d) Let F be a real-valued function of bounded variation on an interval $[a,b]\subseteq\mathbb{R}$. Show that F is absolutely continuous iff $\mathrm{Var}_{[a,b]}(F)=\int_a^b|F'|$.

(e) Consider the function g of 134I/226Cc. Show that $g^{-1}:0,1]\to[0,1]$ is differentiable almost everywhere in $[0,1]$, and find $\mu\{x:(g^{-1})'(x)\le a\}$ for each $a\in\mathbb{R}$.

>(f) Suppose that I and J are sets and that $\langle a_i\rangle_{i\in I}$ is a summable family of real numbers. (i) Show that if $f:J\to I$ is injective then $\langle a_{f(j)}\rangle_{j\in J}$ is summable. (ii) Show that if $g:I\to J$ is any function, then $\sum_{j\in J}\sum_{i\in g^{-1}[\{j\}]}a_i$ is defined and equal to $\sum_{i\in I}a_i$.

226Y Further exercises (a) Explain what modifications are appropriate in the description of the Lebesgue decomposition of a function of bounded variation if we wish to consider functions on open or half-open intervals, including unbounded intervals.

(b) Suppose that $F:[a,b]\to\mathbb{R}$ is a function of bounded variation, and set $h(y)=\#(F^{-1}[\{y\}])$ for $y\in\mathbb{R}$. Show that $\int h=\mathrm{Var}_{[a,b]}(F_c)$, where F_c is the 'continuous part' of F as defined in 226Ca/226Cd.

(c) Show that a set I is countable iff there is a summable family $\langle a_i\rangle_{i\in I}$ of non-zero real numbers.

(d) Suppose that $a<b$ in $\mathbb{R}$, and that $F:[a,b]\to\mathbb{R}$ is a function of bounded variation; let F_p be its saltus part. Show that $|F(b)-F(a)|\le\mu F[\,[a,b]\,]+\mathrm{Var}_{[a,b]}F_p$, where μ is Lebesgue measure on $\mathbb{R}$.

226 Notes and comments In 232I and 232Yb below I will revisit these ideas, linking them to a decomposition of the Lebesgue-Stieltjes measure corresponding to a non-decreasing real function, and thence to more general measures. All this work is peripheral to the main concerns of this volume, but I think it is illuminating, and certainly it is part of the basic knowledge assumed of anyone working in real analysis.

Chapter 23

The Radon-Nikodým Theorem

In Chapter 22, I discussed the indefinite integrals of integrable functions on $\mathbb{R}$, and gave what I hope you feel are satisfying descriptions both of the functions which are indefinite integrals (the absolutely continuous functions) and of how to find which functions they are indefinite integrals of (you differentiate them). For general measure spaces, we have no structure present which can give such simple formulations; but nevertheless the same questions can be asked and, up to a point, answered.

The first section of this chapter introduces the basic machinery needed, the concept of 'countably additive' functional and its decomposition into positive and negative parts. The main theorem takes up the second section: indefinite integrals are the 'truly continuous' additive functionals; on σ-finite spaces, these are the 'absolutely continuous' countably additive functionals. In §233 I discuss the most important single application of the theorem, its use in providing a concept of 'conditional expectation'. This is one of the central concepts of probability theory – as you very likely know; but the form here is a dramatic generalization of the elementary concept of the conditional probability of one event given another, and needs the whole strength of the general theory of measure and integration as developed in Volume 1 and this chapter. I include some notes on convex functions, up to and including versions of Jensen's inequality (233I-233J).

While we are in the area of 'pure' measure theory, I take the opportunity to discuss some further topics. I begin with some essentially elementary constructions, image measures, sums of measures and indefinite-integral measures; I think the details need a little attention, and I work through them in §234. Rather deeper ideas are needed to deal with 'measurable transformations'. In §235 I set out the techniques necessary to provide an abstract basis for a general method of integration-by-substitution, with a detailed account of sufficient conditions for a formula of the type

$$\int g(y)dy = \int g(\phi(x))J(x)dx$$

to be valid.

231 Countably additive functionals

I begin with an abstract description of the objects which will, in appropriate circumstances, correspond to the indefinite integrals of general integrable functions. In this section I give those parts of the theory which do not involve a measure, but only a set with a distinguished σ-algebra of subsets. The basic concepts are those of 'finitely additive' and 'countably additive' functional, and there is one substantial theorem, the 'Hahn decomposition' (231E).

231A Definition Let X be a set and Σ an algebra of subsets of X (136E). A functional $\nu : \Sigma \to \mathbb{R}$ is **finitely additive**, or just **additive**, if $\nu(E \cup F) = \nu E + \nu F$ whenever $E, F \in \Sigma$ and $E \cap F = \emptyset$.

231B Elementary facts Let X be a set, Σ an algebra of subsets of X, and $\nu : \Sigma \to \mathbb{R}$ a finitely additive functional.

(a) $\nu\emptyset = 0$. (For $\nu\emptyset = \nu(\emptyset \cup \emptyset) = \nu\emptyset + \nu\emptyset$.)

(b) If $E_0, \dots , E_n$ are disjoint members of Σ then $\nu(\bigcup_{i\le n} E_i) = \sum_{i=0}^{n} \nu E_i$.

(c) If $E, F \in \Sigma$ and $E \subseteq F$ then $\nu F = \nu E + \nu(F \setminus E)$. More generally, for any $E, F \in \Sigma$,

$$\nu F = \nu(F \cap E) + \nu(F \setminus E).$$

(d) If $E, F \in \Sigma$ then

$$\nu E - \nu F = \nu(E \setminus F) + \nu(E \cap F) - \nu(E \cap F) - \nu(F \setminus E) = \nu(E \setminus F) - \nu(F \setminus E).$$

231C Definition Let X be a set and Σ an algebra of subsets of X. A function $\nu : \Sigma \to \mathbb{R}$ is **countably additive** or **σ-additive** if $\sum_{n=0}^{\infty} \nu E_n$ exists in $\mathbb{R}$ and is equal to $\nu(\bigcup_{n\in\mathbb{N}} E_n)$ for every disjoint sequence $\langle E_n\rangle_{n\in\mathbb{N}}$ in Σ such that $\bigcup_{n\in\mathbb{N}} E_n \in \Sigma$.

Remark Note that when I use the phrase 'countably additive functional' I mean to exclude the possibility of ∞ as a value of the functional. Thus a measure is a countably additive functional iff it is totally finite (211C).

You will sometimes see the phrase '**signed measure**' used to mean what I call a countably additive functional.

231D Elementary facts Let X be a set, Σ a σ-algebra of subsets of X and $\nu : \Sigma \to \mathbb{R}$ a countably additive functional.

(a) ν is finitely additive. **P** (i) Setting $E_n = \emptyset$ for every $n \in \mathbb{N}$, $\sum_{n=0}^{\infty} \nu\emptyset$ must be defined in $\mathbb{R}$ so $\nu\emptyset$ must be 0. (ii) Now if E, $F \in \Sigma$ and $E \cap F = \emptyset$ we can set $E_0 = E$, $E_1 = F$, $E_n = \emptyset$ for $n \geq 2$ and get

$$\nu(E \cup F) = \nu(\bigcup_{n\in\mathbb{N}} E_n) = \sum_{n=0}^{\infty} \nu E_n = \nu E + \nu F. \text{ **Q**}$$

(b) If $\langle E_n \rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence in Σ, with union $E \in \Sigma$, then

$$\nu E = \nu E_0 + \sum_{n=0}^{\infty} \nu(E_{n+1} \setminus E_n) = \lim_{n\to\infty} \nu E_n.$$

(c) If $\langle E_n \rangle_{n\in\mathbb{N}}$ is a non-increasing sequence in Σ with intersection $E \in \Sigma$, then

$$\nu E = \nu E_0 - \lim_{n\to\infty} \nu(E_0 \setminus E_n) = \lim_{n\to\infty} \nu E_n.$$

(d) If $\nu' : \Sigma \to \mathbb{R}$ is another countably additive functional, and $c \in \mathbb{R}$, then $\nu + \nu' : \Sigma \to \mathbb{R}$ and $c\nu : \Sigma \to \mathbb{R}$ are countably additive.

(e) If $H \in \Sigma$, then $\nu_H : \Sigma \to \mathbb{R}$ is countably additive, where $\nu_H E = \nu(E \cap H)$ for every $E \in \Sigma$. **P** If $\langle E_n \rangle_{n\in\mathbb{N}}$ is a disjoint sequence in Σ with union $E \in \Sigma$ then $\langle E_n \cap H \rangle_{n\in\mathbb{N}}$ is disjoint, with union $E \cap H$, so

$$\nu_H E = \nu(H \cap E) = \nu(\bigcup_{n\in\mathbb{N}} (H \cap E_n)) = \sum_{n=0}^{\infty} \nu(H \cap E_n) = \sum_{n=0}^{\infty} \nu_H E_n. \text{ **Q**}$$

Remark For the time being, we shall be using the notion of 'countably additive functional' only on σ-algebras Σ, in which case we can take it for granted that the unions and intersections above belong to Σ.

231E All the ideas above amount to minor modifications of ideas already needed at the very beginning of the theory of measure spaces. We come now to something more substantial.

Theorem Let X be a set, Σ a σ-algebra of subsets of X, and $\nu : \Sigma \to \mathbb{R}$ a countably additive functional. Then
(a) ν is bounded;
(b) there is a set $H \in \Sigma$ such that

$$\nu F \geq 0 \text{ whenever } F \in \Sigma \text{ and } F \subseteq H,$$

$$\nu F \leq 0 \text{ whenever } F \in \Sigma \text{ and } F \cap H = \emptyset.$$

proof (a) ? Suppose, if possible, otherwise. For $E \in \Sigma$, set $M(E) = \sup\{|\nu F| : F \in \Sigma, F \subseteq E\}$; then $M(X) = \infty$. Moreover, whenever E_1, E_2, $F \in \Sigma$ and $F \subseteq E_1 \cup E_2$, then

$$|\nu F| = |\nu(F \cap E_1) + \nu(F \setminus E_1)| \leq |\nu(F \cap E_1)| + |\nu(F \setminus E_1)| \leq M(E_1) + M(E_2),$$

so $M(E_1 \cup E_2) \leq M(E_1) + M(E_2)$. Choose a sequence $\langle E_n \rangle_{n\in\mathbb{N}}$ in Σ as follows. $E_0 = X$. Given that $M(E_n) = \infty$, where $n \in \mathbb{N}$, then surely there is an $F_n \subseteq E_n$ such that $|\nu F_n| \geq 1 + |\nu E_n|$, in which case $|\nu(E_n \setminus F_n)| \geq 1$. Now at least one of $M(F_n)$, $M(E_n \setminus F_n)$ is infinite; if $M(F_n) = \infty$, set $E_{n+1} = F_n$; otherwise, set $E_{n+1} = E_n \setminus F_n$; in either case, note that $|\nu(E_n \setminus E_{n+1})| \geq 1$ and $M(E_{n+1}) = \infty$, so that the induction will continue.

On completing this induction, set $G_n = E_n \setminus E_{n+1}$ for $n \in \mathbb{N}$. Then $\langle G_n \rangle_{n\in\mathbb{N}}$ is a disjoint sequence in Σ, so $\sum_{n=0}^{\infty} \nu G_n$ is defined in $\mathbb{R}$ and $\lim_{n\to\infty} \nu G_n = 0$; but $|\nu G_n| \geq 1$ for every n. **X**

(b)(i) By (a), $\gamma = \sup\{\nu E : E \in \Sigma\} < \infty$. Choose a sequence $\langle E_n \rangle_{n\in\mathbb{N}}$ in Σ such that $\nu E_n \geq \gamma - 2^{-n}$ for every $n \in \mathbb{N}$. For $m \leq n \in \mathbb{N}$, set $F_{mn} = \bigcap_{m\leq i\leq n} E_i$. Then $\nu F_{mn} \geq \gamma - 2 \cdot 2^{-m} + 2^{-n}$ for every $n \geq m$. **P** Induce on n. For $n = m$, this is due to the choice of $E_m = F_{mm}$. For the inductive step, we have $F_{m,n+1} = F_{mn} \cap E_{n+1}$, while surely $\gamma \geq \nu(E_{n+1} \cup F_{mn})$, so

$$\begin{aligned}
\gamma + \nu F_{m,n+1} &\geq \nu(E_{n+1} \cup F_{mn}) + \nu F_{m,n+1} \\
&= \nu E_{n+1} + \nu(F_{mn} \setminus E_{n+1}) + \nu F_{m,n+1} \\
&= \nu E_{n+1} + \nu F_{mn} \\
&\geq \gamma - 2^{-n-1} + \gamma - 2 \cdot 2^{-m} + 2^{-n}
\end{aligned}$$

(by the choice of E_{n+1} and the inductive hypothesis)

$$= 2\gamma - 2\cdot 2^{-m} + 2^{-n-1}.$$

Subtracting γ from both sides, $\nu F_{m,n+1} \geq \gamma - 2\cdot 2^{-m} + 2^{-n-1}$ and the induction proceeds. **Q**

(ii) For $m \in \mathbb{N}$, set

$$F_m = \bigcap_{n\geq m} F_{mn} = \bigcap_{n\geq m} E_n.$$

Then

$$\nu F_m = \lim_{n\to\infty} \nu F_{mn} \geq \gamma - 2\cdot 2^{-m},$$

by 231Dc. Next, $\langle F_m\rangle_{m\in\mathbb{N}}$ is non-decreasing, so setting $H = \bigcup_{m\in\mathbb{N}} F_m$ we have

$$\nu H = \lim_{m\to\infty} \nu F_m \geq \gamma;$$

since νH is surely less than or equal to γ, $\nu H = \gamma$.

If $F \in \Sigma$ and $F \subseteq H$, then

$$\nu H - \nu F = \nu(H \setminus F) \leq \gamma = \nu H,$$

so $\nu F \geq 0$. If $F \in \Sigma$ and $F \cap H = \emptyset$ then

$$\nu H + \nu F = \nu(H \cup F) \leq \gamma = \nu H$$

so $\nu F \leq 0$. This completes the proof.

231F Corollary Let X be a set, Σ a σ-algebra of subsets of X, and $\nu : \Sigma \to \mathbb{R}$ a countably additive functional. Then ν can be expressed as the difference of two totally finite measures with domain Σ.

proof Take $H \in \Sigma$ as described in 231Eb. Set $\nu_1 E = \nu(E \cap H)$, $\nu_2 E = -\nu(E \setminus H)$ for $E \in \Sigma$. Then, as in 231Dd-e, both ν_1 and ν_2 are countably additive functionals on Σ, and of course $\nu = \nu_1 - \nu_2$. But also, by the choice of H, both ν_1 and ν_2 are non-negative, so are totally finite measures.

Remark This is called the '**Jordan decomposition**' of ν. The expression of 231Eb is a '**Hahn decomposition**'.

231X Basic exercises (a) Let Σ be the family of subsets A of $\mathbb{N}$ such that one of A, $\mathbb{N} \setminus A$ is finite. Show that Σ is an algebra of subsets of $\mathbb{N}$. (This is the **finite-cofinite algebra** of subsets of $\mathbb{N}$; compare 211Ra.)

(b) Let X be a set, Σ an algebra of subsets of X and $\nu : \Sigma \to \mathbb{R}$ a finitely additive functional. Show that $\nu(E\cup F\cup G) + \nu(E\cap F) + \nu(E\cap G) + \nu(F\cap G) = \nu E + \nu F + \nu G + \nu(E\cap F\cap G)$ for all E, F, $G \in \Sigma$. Generalize this result to longer sequences of sets.

>(c) Let Σ be the finite-cofinite algebra of subsets of $\mathbb{N}$, as in 231Xa. Define $\nu : \Sigma \to \mathbb{Z}$ by setting

$$\nu E = \lim_{n\to\infty}\bigl(\#(\{i : i \leq n,\, 2i \in E\}) - \#(\{i : i \leq n,\, 2i+1 \in E\})\bigr)$$

for every $E \in \Sigma$. Show that ν is well-defined and finitely additive and unbounded.

(d) Let X be a set and Σ an algebra of subsets of X. (i) Show that if $\nu : \Sigma \to \mathbb{R}$ and $\nu' : \Sigma \to \mathbb{R}$ are finitely additive, so are $\nu + \nu'$ and $c\nu$ for any $c \in \mathbb{R}$. (ii) Show that if $\nu : \Sigma \to \mathbb{R}$ is finitely additive and $H \in \Sigma$, then ν_H is finitely additive, where $\nu_H(E) = \nu(H \cap E)$ for every $E \in \Sigma$.

(e) Let X be a set, Σ an algebra of subsets of X and $\nu : \Sigma \to \mathbb{R}$ a finitely additive functional. Let S be the linear space of those real-valued functions on X expressible in the form $\sum_{i=0}^n a_i\chi E_i$ where $E_i \in \Sigma$ for each i. (i) Show that we have a linear functional $\int : S \to \mathbb{R}$ given by writing

$$\int \sum_{i=0}^n a_i\chi E_i = \sum_{i=0}^n a_i\nu E_i$$

whenever $a_0,\dots,a_n \in \mathbb{R}$ and $E_0,\dots,E_n \in \Sigma$. (ii) Show that if $\nu E \geq 0$ for every $E \in \Sigma$ then $\int f \geq 0$ whenever $f \in S$ and $f(x) \geq 0$ for every $x \in X$. (iii) Show that if ν is bounded and $X \neq \emptyset$ then

$$\sup\{|\int f| : f \in S,\, \|f\|_\infty \leq 1\} = \sup_{E,F\in\Sigma} |\nu E - \nu F|,$$

writing $\|f\|_\infty = \sup_{x\in X} |f(x)|$.

>(f) Let X be a set, Σ a σ-algebra of subsets of X and $\nu : \Sigma \to \mathbb{R}$ a finitely additive functional. Show that the following are equiveridical:

(i) ν is countably additive;

(ii) $\lim_{n\to\infty} \nu E_n = 0$ whenever $\langle E_n\rangle_{n\in\mathbb{N}}$ is a non-increasing sequence in Σ and $\bigcap_{n\in\mathbb{N}} E_n = \emptyset$;

(iii) $\lim_{n\to\infty} \nu E_n = 0$ whenever $\langle E_n\rangle_{n\in\mathbb{N}}$ is a sequence in Σ and $\bigcap_{n\in\mathbb{N}} \bigcup_{m\ge n} E_m = \emptyset$;

(iv) $\lim_{n\to\infty} \nu E_n = \nu E$ whenever $\langle E_n\rangle_{n\in\mathbb{N}}$ is a sequence in Σ and

$$E = \bigcap_{n\in\mathbb{N}} \bigcup_{m\ge n} E_m = \bigcup_{n\in\mathbb{N}} \bigcap_{m\ge n} E_m.$$

(*Hint*: for (i)$\Rightarrow$(iv), consider non-negative ν first.)

(g) Let X be a set and Σ a σ-algebra of subsets of X, and let $\nu : \Sigma \to [-\infty, \infty[$ be a function which is countably additive in the sense that $\nu\emptyset = 0$ and whenever $\langle E_n\rangle_{n\in\mathbb{N}}$ is a disjoint sequence in Σ, $\sum_{n=0}^{\infty} \nu E_n = \lim_{n\to\infty} \sum_{i=0}^{n} \nu E_i$ is defined in $[-\infty, \infty[$ and is equal to $\nu(\bigcup_{n\in\mathbb{N}} E_n)$. Show that ν is bounded above and attains its upper bound (that is, there is an $H \in \Sigma$ such that $\nu H = \sup_{F\in\Sigma} \nu F$). Hence, or otherwise, show that ν is expressible as the difference of a totally finite measure and a measure, both with domain Σ.

231Y Further exercises (a) Let X be a set, Σ an algebra of subsets of X, and $\nu : \Sigma \to \mathbb{R}$ a bounded finitely additive functional. Set

$$\nu^+ E = \sup\{\nu F : F \in \Sigma,\ F \subseteq E\},$$

$$\nu^- E = -\inf\{\nu F : F \in \Sigma,\ F \subseteq E\},$$

$$|\nu| E = \sup\{\nu F_1 - \nu F_2 : F_1,\ F_2 \in \Sigma,\ F_1,\ F_2 \subseteq E\}.$$

Show that ν^+, ν^- and $|\nu|$ are all bounded finitely additive functionals on Σ and that $\nu = \nu^+ - \nu^-$, $|\nu| = \nu^+ + \nu^-$. Show that if ν is countably additive so are ν^+, ν^- and $|\nu|$. ($|\nu|$ is sometimes called the **variation** of ν.)

(b) Let X be a set and Σ an algebra of subsets of X. Let ν_1, ν_2 be two bounded finitely additive functionals defined on Σ. Set

$$(\nu_1 \vee \nu_2)(E) = \sup\{\nu_1 F + \nu_2(E \setminus F) : F \in \Sigma,\ F \subseteq E\},$$

$$(\nu_1 \wedge \nu_2)(E) = \inf\{\nu_1 F + \nu_2(E \setminus F) : F \in \Sigma,\ F \subseteq E\}.$$

Show that $\nu_1 \vee \nu_2$ and $\nu_1 \wedge \nu_2$ are finitely additive functionals, and that $\nu_1 + \nu_2 = \nu_1 \vee \nu_2 + \nu_1 \wedge \nu_2$. Show that, in the language of 231Ya,

$$\nu^+ = \nu \vee 0, \quad \nu^- = (-\nu) \vee 0 = -(\nu \wedge 0), \quad |\nu| = \nu \vee (-\nu) = \nu^+ \vee \nu^- = \nu^+ + \nu^-,$$

$$\nu_1 \vee \nu_2 = \nu_1 + (\nu_2 - \nu_1)^+, \quad \nu_1 \wedge \nu_2 = \nu_1 - (\nu_1 - \nu_2)^+,$$

so that $\nu_1 \vee \nu_2$ and $\nu_1 \wedge \nu_2$ are countably additive if ν_1 and ν_2 are.

(c) Let X be a set and Σ an algebra of subsets of X. Let M be the set of all bounded finitely additive functionals from Σ to $\mathbb{R}$. Show that M is a linear space under the natural definitions of addition and scalar multiplication. Show that M has a partial order $\le$ defined by saying that

$$\nu \le \nu' \text{ iff } \nu E \le \nu' E \text{ for every } E \in \Sigma,$$

and that for this partial order $\nu_1 \vee \nu_2$, $\nu_1 \wedge \nu_2$, as defined in 231Yb, are $\sup\{\nu_1, \nu_2\}$, $\inf\{\nu_1, \nu_2\}$.

(d) Let X be a set and Σ an algebra of subsets of X. Let $\nu_0, \dots, \nu_n$ be bounded finitely additive functionals on Σ and set

$$\check{\nu} E = \sup\{\textstyle\sum_{i=0}^{n} \nu_i F_i : F_0, \dots, F_n \in \Sigma,\ \bigcup_{i\le n} F_i = E,\ F_i \cap F_j = \emptyset \text{ for } i \ne j\},$$

$$\hat{\nu} E = \inf\{\textstyle\sum_{i=0}^{n} \nu_i F_i : F_0, \dots, F_n \in \Sigma,\ \bigcup_{i\le n} F_i = E,\ F_i \cap F_j = \emptyset \text{ for } i \ne j\}$$

for $E \in \Sigma$. Show that $\check{\nu}$ and $\hat{\nu}$ are finitely additive and are, respectively, $\sup\{\nu_0, \dots, \nu_n\}$ and $\inf\{\nu_0, \dots, \nu_n\}$ in the partially ordered set of finitely additive functionals on Σ.

(e) Let X be a set and Σ a σ-algebra of subsets of X; let M be the partially ordered set of all bounded finitely additive functionals from Σ to $\mathbb{R}$. (i) Show that if $A \subseteq M$ is non-empty and bounded above in M, then A has a supremum $\check{\nu}$ in M, given by the formula

$$\check{\nu}E = \sup\{\sum_{i=0}^{n} \nu_i F_i : \nu_0, \dots, \nu_n \in A,\ F_0, \dots, F_n \in \Sigma,\ \bigcup_{i \le n} F_i = E,\ F_i \cap F_j = \emptyset \text{ for } i \ne j\}.$$

(ii) Show that if $A \subseteq M$ is non-empty and bounded below in M then it has an infimum $\hat{\nu} \in M$, given by the formula

$$\hat{\nu}E = \inf\{\sum_{i=0}^{n} \nu_i F_i : \nu_0, \dots, \nu_n \in A,\ F_0, \dots, F_n \in \Sigma,\ \bigcup_{i \le n} F_i = E,\ F_i \cap F_j = \emptyset \text{ for } i \ne j\}.$$

(f) Let X be a set, Σ an algebra of subsets of X, and $\nu : \Sigma \to \mathbb{R}$ a non-negative finitely additive functional. For $E \in \Sigma$ set

$$\nu_{ca}(E) = \inf\{\sup_{n\in\mathbb{N}} \nu F_n : \langle F_n \rangle_{n\in\mathbb{N}} \text{ is a non-decreasing sequence in } \Sigma \text{ with union } E\}.$$

Show that ν_{ca} is a countably additive functional on Σ and that if ν' is any countably additive functional with $\nu' \le \nu$ then $\nu' \le \nu_{ca}$. Show that $\nu_{ca} \wedge (\nu - \nu_{ca}) = 0$.

(g) Let X be a set, Σ an algebra of subsets of X, and $\nu : \Sigma \to \mathbb{R}$ a bounded finitely additive functional. Show that ν is uniquely expressible as $\nu_{ca} + \nu_{pfa}$, where ν_{ca} is countably additive, ν_{pfa} is finitely additive and if $0 \le \nu' \le |\nu_{pfa}|$ and ν' is countably additive then $\nu' = 0$.

(h) Let X be a set and Σ an algebra of subsets of X. Let M be the linear space of bounded finitely additive functionals on Σ, and for $\nu \in M$ set $\|\nu\| = |\nu|(X)$, defining $|\nu|$ as in 231Ya. ($\|\nu\|$ is the **total variation** of ν.) Show that $\| \, \|$ is a norm on M under which M is a Banach space. Show that the space of bounded countably additive functionals on Σ is a closed linear subspace of M.

(i) Repeat as many as possible of the results of this section for complex-valued functionals.

231 Notes and comments The real purpose of this section has been to describe the Hahn decomposition of a countably additive functional (231E). The very leisurely exposition in 231A-231D is intended as a review of the most elementary properties of measures, in the slightly more general context of 'signed measures', with those properties corresponding to 'additivity' alone separated from those which depend on 'countable additivity'. In 231Xf I set out necessary and sufficient conditions for a finitely additive functional on a σ-algebra to be countably additive, designed to suggest that a finitely additive functional is countably additive iff it is 'sequentially order-continuous' in some sense. The fact that a countably additive functional can be expressed as the difference of non-negative countably additive functionals (231F) has an important counterpart in the theory of finitely additive functionals: a finitely additive functional can be expressed as the difference of non-negative finitely additive functionals if (and only if) it is bounded (231Ya). But I do not think that this, or the further properties of bounded finitely additive functionals described in 231Xe and 231Y, will be important to us before Volume 3.

232 The Radon-Nikodým theorem

I come now to the chief theorem of this chapter, one of the central results of measure theory, relating countably additive functionals to indefinite integrals. The objective is to give a complete description of the functionals which can arise as indefinite integrals of integrable functions (232E). These can be characterized as the 'truly continuous' additive functionals (232Ab). A more commonly used concept, and one adequate in many cases, is that of 'absolutely continuous' additive functional (232Aa); I spend the first few paragraphs (232B-232D) on elementary facts about truly continuous and absolutely continuous functionals. I end the section with a discussion of the decomposition of general countably additive functionals (232I).

232A Absolutely continuous functionals Let (X, Σ, μ) be a measure space and $\nu : \Sigma \to \mathbb{R}$ a finitely additive functional.

(a) ν is **absolutely continuous** with respect to μ (sometimes written '$\nu \ll \mu$') if for every $\epsilon > 0$ there is a $\delta > 0$ such that $|\nu E| \le \epsilon$ whenever $E \in \Sigma$ and $\mu E \le \delta$.

(b) ν is **truly continuous** with respect to μ if for every $\epsilon > 0$ there are $E \in \Sigma$, $\delta > 0$ such that μE is finite and $|\nu F| \le \epsilon$ whenever $F \in \Sigma$ and $\mu(E \cap F) \le \delta$.

(c) For reference, I add another definition here. If ν is countably additive, it is **singular** with respect to μ if there is a set $F \in \Sigma$ such that $\mu F = 0$ and $\nu E = 0$ whenever $E \in \Sigma$ and $E \subseteq X \setminus F$.

232B Proposition Let (X, Σ, μ) be a measure space and $\nu : \Sigma \to \mathbb{R}$ a finitely additive functional.

(a) If ν is countably additive, it is absolutely continuous with respect to μ iff $\nu E = 0$ whenever $\mu E = 0$.

(b) ν is truly continuous with respect to μ iff (α) it is countably additive (β) it is absolutely continuous with respect to μ (γ) whenever $E \in \Sigma$ and $\nu E \neq 0$ there is an $F \in \Sigma$ such that $\mu F < \infty$ and $\nu(E \cap F) \neq 0$.

(c) If (X, Σ, μ) is σ-finite, then ν is truly continuous with respect to μ iff it is countably additive and absolutely continuous with respect to μ.

(d) If (X, Σ, μ) is totally finite, then ν is truly continuous with respect to μ iff it is absolutely continuous with respect to μ.

proof (a)(i) If ν is absolutely continuous with respect to μ and $\mu E = 0$, then $\mu E \le \delta$ for every $\delta > 0$, so $|\nu E| \le \epsilon$ for every $\epsilon > 0$ and $\nu E = 0$.

(ii) ? Suppose, if possible, that $\nu E = 0$ whenever $\mu E = 0$, but ν is not absolutely continuous. Then there is an $\epsilon > 0$ such that for every $\delta > 0$ there is an $E \in \Sigma$ such that $\mu E \le \delta$ but $|\nu E| \ge \epsilon$. For each $n \in \mathbb{N}$ we may choose an $F_n \in \Sigma$ such that $\mu F_n \le 2^{-n}$ and $|\nu F_n| \ge \epsilon$. Consider $F = \bigcap_{n\in\mathbb{N}} \bigcup_{k\ge n} F_k$. Then we have

$$\mu F \le \inf_{n\in\mathbb{N}} \mu(\bigcup_{k\ge n} F_k) \le \inf_{n\in\mathbb{N}} \sum_{k=n}^{\infty} 2^{-k} = 0,$$

so $\mu F = 0$.

Now recall that by 231Eb there is an $H \in \Sigma$ such that $\nu G \ge 0$ when $G \in \Sigma$ and $G \subseteq H$, and $\nu G \le 0$ when $G \in \Sigma$ and $G \cap H = \emptyset$. As in 231F, set $\nu_1 G = \nu(G \cap H)$, $\nu_2 G = -\nu(G \setminus H)$ for $G \in \Sigma$, so that ν_1 and ν_2 are totally finite measures, and $\nu_1 F = \nu_2 F = 0$ because $\mu(F \cap H) = \mu(F \setminus H) = 0$. Consequently

$$0 = \nu_i F = \lim_{n\to\infty} \nu_i(\bigcup_{m\ge n} F_m) \ge \limsup_{n\to\infty} \nu_i F_n$$

for both i, and

$$0 = \lim_{n\to\infty}(\nu_1 F_n + \nu_2 F_n) \ge \liminf_{n\to\infty} |\nu F_n| \ge \epsilon > 0,$$

which is absurd. **X**

(b)(i) Suppose that ν is truly continuous with respect to μ. It is obvious from the definitions that ν is absolutely continuous with respect to μ. If $\nu E \neq 0$, there must be an F of finite measure such that $|\nu G| < |\nu E|$ whenever $G \cap F = \emptyset$, so that $|\nu(E \setminus F)| < |\nu E|$ and $\nu(E \cap F) \neq 0$. This deals with the conditions (β) and (γ).

To check that ν is countably additive, let $\langle E_n \rangle_{n\in\mathbb{N}}$ be a disjoint sequence in Σ, with union E, and $\epsilon > 0$. Let $\delta > 0$, $F \in \Sigma$ be such $\mu F < \infty$ and $|\nu G| \le \epsilon$ whenever $G \in \Sigma$ and $\mu(F \cap G) \le \delta$. Then

$$\sum_{n=0}^{\infty} \mu(E_n \cap F) \le \mu F < \infty,$$

so there is an $n \in \mathbb{N}$ such that $\sum_{i=n}^{\infty} \mu(E_i \cap F) \le \delta$. Take any $m \ge n$ and consider $E_m^* = \bigcup_{i\le m} E_i$. We have

$$|\nu E - \sum_{i=0}^{m} \nu E_i| = |\nu E - \nu E_m^*| = |\nu(E \setminus E_m^*)| \le \epsilon,$$

because

$$\mu(F \cap E \setminus E_m^*) = \sum_{i=m+1}^{\infty} \mu(F \cap E_i) \le \delta.$$

As ϵ is arbitrary,

$$\nu E = \sum_{i=0}^{\infty} \nu E_i;$$

as $\langle E_n \rangle_{n\in\mathbb{N}}$ is arbitrary, ν is countably additive.

(ii) Now suppose that ν satisfies the three conditions. By 231F, ν can be expressed as the difference of two non-negative countably additive functionals ν_1, ν_2; set $\nu' = \nu_1 + \nu_2$, so that ν' is a non-negative countably additive functional and $|\nu F| \le \nu' F$ for every $F \in \Sigma$. Set

$$\gamma = \sup\{\nu' F : F \in \Sigma,\ \mu F < \infty\} \le \nu' X < \infty,$$

and choose a sequence $\langle F_n \rangle_{n\in\mathbb{N}}$ of sets of finite measure such that $\lim_{n\to\infty} \nu' F_n = \gamma$; set $F^* = \bigcup_{n\in\mathbb{N}} F_n$. If $G \in \Sigma$ and $G \cap F^* = \emptyset$ then $\nu G = 0$. **P?** Otherwise, by condition (γ), there is an $F \in \Sigma$ such that $\mu F < \infty$ and $\nu(G \cap F) \neq 0$. It follows that

$$\nu'(F \setminus F^*) \geq \nu'(F \cap G) \geq |\nu(F \cap G)| > 0,$$

and there must be an $n \in \mathbb{N}$ such that

$$\gamma < \nu' F_n + \nu'(F \setminus F^*) = \nu'(F_n \cup (F \setminus F^*)) \leq \nu'(F \cup F_n) \leq \gamma$$

because $\mu(F \cup F_n) < \infty$; but this is impossible. **XQ**

Setting $F_n^* = \bigcup_{k \leq n} F_k$ for each n, we have $\lim_{n\to\infty} \nu'(F^* \setminus F_n^*) = 0$. Take any $\epsilon > 0$, and (using condition (β)) let $\delta > 0$ be such that $|\nu E| \leq \frac{1}{2}\epsilon$ whenever $\mu E \leq \delta$. Let n be such that $\nu'(F^* \setminus F_n^*) \leq \frac{1}{2}\epsilon$. Now if $F \in \Sigma$ and $\mu(F \cap F_n^*) \leq \delta$ then

$$\begin{aligned}|\nu F| &\leq |\nu(F \cap F_n^*)| + |\nu(F \cap F^* \setminus F_n^*)| + |\nu(F \setminus F^*)| \\ &\leq \frac{1}{2}\epsilon + \nu'(F \cap F^* \setminus F_n^*) + 0 \\ &\leq \frac{1}{2}\epsilon + \nu'(F^* \setminus F_n^*) \leq \frac{1}{2}\epsilon + \frac{1}{2}\epsilon = \epsilon.\end{aligned}$$

And $\mu F_n^* < \infty$. As ϵ is arbitrary, ν is truly continuous.

(c) Now suppose that (X, Σ, μ) is σ-finite and that ν is countably additive and absolutely continuous with respect to μ. Let $\langle X_n \rangle_{n\in\mathbb{N}}$ be a non-decreasing sequence of sets of finite measure covering X (211D). If $\nu E \neq 0$, then $\lim_{n\to\infty} \nu(E \cap X_n) \neq 0$, so $\nu(E \cap X_n) \neq 0$ for some n. This shows that ν satisfies condition (γ) of (b), so is truly continuous.

Of course the converse of this fact is already covered by (b).

(d) Finally, suppose that $\mu X < \infty$ and that ν is absolutely continuous with respect to μ. Then it must be truly continuous, because we can take $F = X$ in the definition 232Ab.

232C Lemma Let (X, Σ, μ) be a measure space and ν, ν' two countably additive functionals on Σ which are truly continuous with respect to μ. Take $c \in \mathbb{R}$ and $H \in \Sigma$, and set $\nu_H E = \nu(E \cap H)$, as in 231De. Then $\nu + \nu'$, $c\nu$ and ν_H are all truly continuous with respect to μ, and ν is expressible as the difference of non-negative countably additive functionals which are truly continuous with respect to μ.

proof Let $\epsilon > 0$. Set $\eta = \epsilon/(2 + |c|) > 0$. Then there are δ, $\delta' > 0$ and E, $E' \in \Sigma$ such that $\mu E < \infty$, $\mu E' < \infty$ and $|\nu F| \leq \eta$ whenever $\mu(F \cap E) \leq \delta$, $|\nu' F| \leq \eta$ whenever $\mu(F \cap E) \leq \delta'$. Set $\delta^* = \min(\delta, \delta') > 0$, $E^* = E \cup E' \in \Sigma$; then

$$\mu E^* \leq \mu E + \mu E' < \infty.$$

Suppose that $F \in \Sigma$ and $\mu(F \cap E^*) \leq \delta^*$; then

$$\mu(F \cap H \cap E) \leq \mu(F \cap E) \leq \delta^* \leq \delta, \quad \mu(F \cap E') \leq \delta^*$$

so

$$|(\nu + \nu')F| \leq |\nu F| + |\nu' F| \leq \eta + \eta \leq \epsilon,$$

$$|(c\nu)F| = |c||\nu F| \leq |c|\eta \leq \epsilon,$$

$$|\nu_H F| = |\nu(F \cap H)| \leq \eta \leq \epsilon.$$

As ϵ is arbitrary, $\nu + \nu'$, $c\nu$ and ν_H are all truly continuous.

Now, taking H from 231Eb, we see that $\nu_1 = \nu_H$ and $\nu_2 = -\nu_{X\setminus H}$ are truly continuous and non-negative, and $\nu = \nu_1 - \nu_2$ is the difference of truly continuous measures.

232D Proposition Let (X, Σ, μ) be a measure space, and f a μ-integrable real-valued function. For $E \in \Sigma$ set $\nu E = \int_E f$. Then $\nu : \Sigma \to \mathbb{R}$ is a countably additive functional and is truly continuous with respect to μ, therefore absolutely continuous with respect to μ.

proof Recall that $\int_E f = \int f \times \chi E$ is defined for every $E \in \Sigma$ (131Fa). So $\nu : \Sigma \to \mathbb{R}$ is well-defined. If E, $F \in \Sigma$ are disjoint then

$$\begin{aligned}\nu(E \cup F) &= \int f \times \chi(E \cup F) = \int (f \times \chi E) + (f \times \chi F) \\ &= \int f \times \chi E + \int f \times \chi F = \nu E + \nu F,\end{aligned}$$

so ν is finitely additive.

Now 225A, without using the phrase 'truly continuous', proved exactly that ν is truly continuous with respect to μ. It follows from 232Bb that ν is countably additive and absolutely continuous.

Remark The functional $E \mapsto \int_E f$ is called the **indefinite integral** of f.

232E We are now at last ready for the theorem.

The Radon-Nikodým theorem Let (X, Σ, μ) be a measure space and $\nu : \Sigma \to \mathbb{R}$ a function. Then the following are equiveridical:

(i) there is a μ-integrable function f such that $\nu E = \int_E f$ for every $E \in \Sigma$;

(ii) ν is finitely additive and truly continuous with respect to μ.

proof (a) If f is a μ-integrable real-valued function and $\nu E = \int_E f$ for every $E \in \Sigma$, then 232D tells us that ν is finitely additive and truly continuous.

(b) In the other direction, suppose that ν is finitely additive and truly continuous; note that (by 232B(a-b)) $\nu E = 0$ whenever $\mu E = 0$. To begin with, suppose that ν is non-negative and not zero.

In this case, there is a non-negative simple function f such that $\int f > 0$ and $\int_E f \le \nu E$ for every $E \in \Sigma$. **P** Let $H \in \Sigma$ be such that $\nu H > 0$; set $\epsilon = \frac{1}{3}\nu H > 0$. Let $E \in \Sigma$, $\delta > 0$ be such that $\mu E < \infty$ and $\nu F \le \epsilon$ whenever $F \in \Sigma$ and $\mu(F \cap E) \le \delta$; then $\nu(H \setminus E) \le \epsilon$ so $\nu E \ge \nu(H \cap E) \ge 2\epsilon$ and $\mu E \ge \mu(H \cap E) > 0$. Set $\mu_E F = \mu(F \cap E)$ for every $F \in \Sigma$; then μ_E is a countably additive functional on Σ. Set $\nu' = \nu - \alpha\mu_E$, where $\alpha = \epsilon/\mu E$; then ν' is a countably additive functional and $\nu' E > 0$. By 231Eb, as usual, there is a set $G \in \Sigma$ such that $\nu' F \ge 0$ if $F \in \Sigma$, $F \subseteq G$, but $\nu' F \le 0$ if $F \in \Sigma$ and $F \cap G = \emptyset$. As $\nu'(E \setminus G) \le 0$,

$$0 < \nu' E \le \nu'(E \cap G) \le \nu(E \cap G)$$

and $\mu(E \cap G) > 0$. Set $f = \alpha\chi(E \cap G)$; then f is a non-negative simple function and $\int f = \alpha\mu(E \cap G) > 0$.

If $F \in \Sigma$ then $\nu'(F \cap G) \ge 0$, that is,

$$\nu(F \cap G) \ge \alpha\mu_E(F \cap G) = \alpha\mu(F \cap E \cap G) = \int_F f.$$

So

$$\nu F \ge \nu(F \cap G) \ge \int_F f,$$

as required. **Q**

(c) Still supposing that ν is a non-negative, truly continuous additive functional, let Φ be the set of non-negative simple functions $f : X \to \mathbb{R}$ such that $\int_E f \le \nu E$ for every $E \in \Sigma$; then the constant function $\mathbf{0}$ belongs to Φ, so Φ is not empty.

If f, $g \in \Phi$ then $f \vee g \in \Phi$, where $(f \vee g)(x) = \max(f(x), g(x))$ for $x \in X$. **P** Set $H = \{x : (f - g)(x) \ge 0\} \in \Sigma$; then $f \vee g = (f \times \chi H) + (g \times \chi(X \setminus H))$ is a non-negative simple function, and for any $E \in \Sigma$,

$$\int_E f \vee g = \int_{E \cap H} f + \int_{E \setminus H} g \le \nu(E \cap H) + \nu(E \setminus H) = \nu E. \textbf{ Q}$$

Set

$$\gamma = \sup\{\int f : f \in \Phi\} \le \nu X < \infty.$$

Choose a sequence $\langle f_n \rangle_{n \in \mathbb{N}}$ in Φ such that $\lim_{n\to\infty} \int f_n = \gamma$. For each n, set $g_n = f_0 \vee f_1 \vee \ldots \vee f_n$; then $g_n \in \Phi$ and $\int f_n \le \int g_n \le \gamma$ for each n, so $\lim_{n\to\infty} \int g_n = \gamma$. By B.Levi's theorem, $f = \lim_{n\to\infty} g_n$ is integrable and $\int f = \gamma$. Note that if $E \in \Sigma$ then

$$\int_E f = \lim_{n\to\infty} \int_E f_n \le \nu E.$$

? Suppose, if possible, that there is an $H \in \Sigma$ such that $\int_H f \ne \nu H$. Set

$$\nu_1 F = \nu F - \int_F f \ge 0$$

for every $F \in \Sigma$; then by (a) of this proof and 232C, ν_1 is a truly continuous finitely additive functional, and we are supposing that $\nu_1 \ne 0$. By (b) of this proof, there is a non-negative simple function g such that $\int_F g \le \nu_1 F$ for every $F \in \Sigma$ and $\int g > 0$. Take $n \in \mathbb{N}$ such that $\int f_n + \int g > \gamma$. Then $f_n + g$ is a non-negative simple function and

$$\int_F (f_n + g) = \int_F f_n + \int_F g \le \int_F f + \int_F g = \nu F - \nu_1 F + \int_F g \le \nu F$$

for any $F \in \Sigma$, so $f_n + g \in \Phi$, and

$$\gamma < \int f_n + \int g = \int f_n + g \le \gamma,$$

which is absurd. **X** Thus we have $\int_H f = \nu H$ for every $H \in \Sigma$.

(d) This proves the theorem for non-negative ν. For general ν, we need only observe that ν is expressible as $\nu_1 - \nu_2$, where ν_1 and ν_2 are non-negative truly continuous countably additive functionals, by 232C; so that there are integrable functions f_1, f_2 such that $\nu_i F = \int_F f_i$ for both i and every $F \in \Sigma$. Of course $f = f_1 - f_2$ is integrable and $\nu F = \int_F f$ for every $F \in \Sigma$. This completes the proof.

232F Corollary Let (X, Σ, μ) be a σ-finite measure space and $\nu : \Sigma \to \mathbb{R}$ a function. Then there is a μ-integrable function f such that $\nu E = \int_E f$ for every $E \in \Sigma$ iff ν is countably additive and absolutely continuous with respect to μ.

proof Put 232Bc and 232E together.

232G Corollary Let (X, Σ, μ) be a totally finite measure space and $\nu : \Sigma \to \mathbb{R}$ a function. Then there is a μ-integrable function f on X such that $\nu E = \int_E f$ for every $E \in \Sigma$ iff ν is finitely additive and absolutely continuous with respect to μ.

proof Put 232Bd and 232E together.

232H Remarks (a) Most authors are satisfied with 232F as the 'Radon-Nikodým theorem'. In my view the problem of identifying indefinite integrals is of sufficient importance to justify an analysis which applies to all measure spaces, even if it requires a new concept (the notion of 'truly continuous' functional).

(b) I ought to offer an example of an absolutely continuous functional which is not truly continuous. A simple one is the following. Let X be any uncountable set. Let Σ be the countable-cocountable σ-algebra of subsets of X and ν the countable-cocountable measure on X (211R). Let μ be the restriction to Σ of counting measure on X. If $\mu E = 0$ then $E = \emptyset$ and $\nu E = 0$, so ν is absolutely continuous. But for any E of finite measure we have $\nu(X \setminus E) = 1$, so ν is not truly continuous. See also 232Xf(i).

***(c)** The space (X, Σ, μ) of this example is, in terms of the classification developed in Chapter 21, somewhat irregular; for instance, it is neither locally determined nor localizable, and therefore not strictly localizable, though it is complete and semi-finite. Can this phenomenon occur in a strictly localizable measure space? We are led here into a fascinating question. Suppose, in (b), I used the same idea, but with $\Sigma = \mathcal{P}X$. No difficulty arises in constructing μ; but can there now be a ν with the required properties, that is, a non-zero countably additive functional from $\mathcal{P}X$ to $\mathbb{R}$ which is zero on all finite sets? This is the 'Banach-Ulam problem', on which I have written extensively elsewhere (Fremlin 93), and to which I will return in Volume 5. The present question is touched on again in 363S in Volume 3.

(d) Following the Radon-Nikodým theorem, the question immediately arises: for a given ν, how much possible variation is there in the corresponding f? The answer is straightforward enough: two integrable functions f and g give rise to the same indefinite integral iff they are equal almost everywhere (131Hb).

(e) I have stated the Radon-Nikodým theorem in terms of arbitrary integrable functions, meaning to interpret 'integrability' in a wide sense, according to the conventions of Volume 1. However, given a truly continuous countably additive functional ν, we can ask whether there is in any sense a canonical integrable function representing it. The answer is no. But we certainly do not need to take arbitrary integrable functions of the type considered in Chapter 12. If f is any integrable function, there is a conegligible set E such that $f \restriction E$ is measurable, and now we can find a conegligible measurable set $G \subseteq E \cap \operatorname{dom} f$; if we set $g(x) = f(x)$ for $x \in G$, 0 for $x \in X \setminus G$, then $f =_{\text{a.e.}} g$, so g has the same indefinite integral as f (as noted in (d) just above), while g is measurable and has domain X. Thus we can make a trivial, but sometimes convenient, refinement to the theorem: if (X, Σ, μ) is a measure space, and $\nu : \Sigma \to \mathbb{R}$ is finitely additive and truly continuous with respect to μ, then there is a Σ-measurable μ-integrable function $g : X \to \mathbb{R}$ such that $\int_E g = \nu E$ for every $E \in \Sigma$.

(f) It is convenient to introduce now a general definition. If (X, Σ, μ) is a measure space and ν is a $[-\infty, \infty]$-valued functional defined on a family of subsets of X, I will say that a $[-\infty, \infty]$-valued function f defined on a subset of X is a **Radon-Nikodým derivative** of ν with respect to μ if $\int_E f d\mu$ is defined (in the sense of 214D) and equal to νE for every $E \in \operatorname{dom} \nu$. Thus the integrable functions called f in 232E-232G are all 'Radon-Nikodým derivatives'; later on we shall have less well-regulated examples.

When ν is a measure and f is non-negative, f may be called a **density function**.

(g) Throughout the work above I have taken it that ν is defined on the whole domain Σ of μ. In some of the most important applications, however, ν is defined only on some smaller σ-algebra T. In this case we commonly seek to apply the same results with $\mu\restriction\mathrm{T}$ in place of μ.

232I The Lebesgue decomposition of a countably additive functional: Proposition (a) Let (X,Σ,μ) be a measure space and $\nu:\Sigma\to\mathbb{R}$ a countably additive functional. Then ν has unique expressions as

$$\nu=\nu_s+\nu_{ac}=\nu_s+\nu_{tc}+\nu_e,$$

where ν_{tc} is truly continuous with respect to μ, ν_s is singular with respect to ν, and ν_e is absolutely continuous with respect to μ and zero on every set of finite measure.

(b) If $X=\mathbb{R}^r$, Σ is the algebra of Borel sets in $\mathbb{R}^r$ and μ is the restriction of Lebesgue measure to Σ, then ν is uniquely expressible as $\nu_p+\nu_{cs}+\nu_{ac}$ where ν_{ac} is absolutely continuous with respect to μ, ν_{cs} is singular with respect to μ and zero on singletons, and $\nu_pE=\sum_{x\in E}\nu_p\{x\}$ for every $E\in\Sigma$.

proof (a)(i) Suppose first that ν is non-negative. In this case, set

$$\nu_sE=\sup\{\nu(E\cap F):F\in\Sigma,\ \mu F=0\},$$

$$\nu_tE=\sup\{\nu(E\cap F):F\in\Sigma,\ \mu F<\infty\}.$$

Then both ν_s and ν_t are countably additive. **P** Surely $\nu_s\emptyset=\nu_t\emptyset=0$. Let $\langle E_n\rangle_{n\in\mathbb{N}}$ be a disjoint sequence in Σ with union E. **(α)** If $F\in\Sigma$ and $\mu F=0$, then

$$\nu(E\cap F)=\sum_{n=0}^\infty\nu(E_n\cap F)\le\sum_{n=0}^\infty\nu_s(E_n);$$

as F is arbitrary,

$$\nu_sE\le\sum_{n=0}^\infty\nu_sE_n.$$

(β) If $F\in\Sigma$ and $\mu F<\infty$, then

$$\nu(E\cap F)=\sum_{n=0}^\infty\nu(E_n\cap F)\le\sum_{n=0}^\infty\nu_t(E_n);$$

as F is arbitrary,

$$\nu_tE\le\sum_{n=0}^\infty\nu_tE_n.$$

(γ) If $\epsilon>0$, then (because $\sum_{n=0}^\infty\nu E_n=\nu E<\infty$) there is an $n\in\mathbb{N}$ such that $\sum_{k=n+1}^\infty\nu E_k\le\epsilon$. Now, for each $k\le n$, there is an $F_k\in\Sigma$ such that $\mu F_k=0$ and $\nu(E_k\cap F_k)\ge\nu_sE_k-\frac{\epsilon}{n+1}$. In this case, $F=\bigcup_{k\le n}F_k\in\Sigma$, $\mu F=0$ and

$$\nu_sE\ge\nu(E\cap F)\ge\sum_{k=0}^n\nu(E_k\cap F_k)\ge\sum_{k=0}^n\nu_sE_k-\epsilon\ge\sum_{k=0}^\infty\nu_sE_k-2\epsilon,$$

because

$$\sum_{k=n+1}^\infty\nu_sE_k\le\sum_{k=n+1}^\infty\nu E_k\le\epsilon.$$

As ϵ is arbitrary,

$$\nu_sE\ge\sum_{k=0}^\infty\nu_sE_k.$$

(δ) Similarly, for each $k\le n$, there is an $F'_k\in\Sigma$ such that $\mu F'_k<\infty$ and $\nu(E_k\cap F'_k)\ge\nu_tE_k-\frac{\epsilon}{n+1}$. In this case, $F'=\bigcup_{k\le n}F'_k\in\Sigma$, $\mu F'<\infty$ and

$$\nu_tE\ge\nu(E\cap F')\ge\sum_{k=0}^n\nu(E_k\cap F'_k)\ge\sum_{k=0}^n\nu_tE_k-\epsilon\ge\sum_{k=0}^\infty\nu_tE_k-2\epsilon,$$

because

$$\sum_{k=n+1}^\infty\nu_tE_k\le\sum_{k=n+1}^\infty\nu E_k\le\epsilon.$$

As ϵ is arbitrary,

$$\nu_tE\ge\sum_{k=0}^\infty\nu_tE_k.$$

(ϵ) Putting these together, $\nu_sE=\sum_{n=0}^\infty\nu_sE_n$ and $\nu_tE=\sum_{n=0}^\infty\nu_tE_n$. As $\langle E_n\rangle_{n\in\mathbb{N}}$ is arbitrary, ν_s and ν_t are countably additive. **Q**

(ii) Still supposing that ν is non-negative, if we choose a sequence $\langle F_n\rangle_{n\in\mathbb{N}}$ in Σ such that $\mu F_n=0$ for each n and $\lim_{n\to\infty}\nu F_n=\nu_sX$, then $F^*=\bigcup_{n\in\mathbb{N}}F_n$ has $\mu F^*=0$, $\nu F^*=\nu_sX$; so that $\nu_s(X\setminus F^*)=0$, and ν_s is singular with respect to μ in the sense of 232Ac.

Note that $\nu_sF=\nu F$ whenever $\mu F=0$. So if we write $\nu_{ac}=\nu-\nu_s$, then ν_{ac} is a countably additive functional and $\nu_{ac}F=0$ whenever $\mu F=0$; that is, ν_{ac} is absolutely continuous with respect to μ.

If we write $\nu_{tc} = \nu_t - \nu_s$, then ν_{tc} is a non-negative countably additive functional; $\nu_{tc}F = 0$ whenever $\mu F = 0$, and if $\nu_{tc}E > 0$ there is a set F with $\mu F < \infty$ and $\nu_{tc}(E \cap F) > 0$. So ν_{tc} is truly continuous with respect to μ, by 232Bb. Set $\nu_e = \nu - \nu_t = \nu_{ac} - \nu_{tc}$.

Thus for any non-negative countably additive functional ν, we have expressions

$$\nu = \nu_s + \nu_{ac}, \quad \nu_{ac} = \nu_{tc} + \nu_e$$

where ν_s, ν_{ac}, ν_{tc} and ν_e are all non-negative countably additive functionals, ν_s is singular with respect to μ, ν_{ac} and ν_e are absolutely continuous with respect to μ, ν_{tc} is truly continuous with respect to μ, and $\nu_e F = 0$ whenever $\mu F < \infty$.

(iii) For general countably additive functionals $\nu : \Sigma \to \mathbb{R}$, we can express ν as $\nu' - \nu''$, where ν' and ν'' are non-negative countably additive functionals. If we define ν'_s, $\nu''_s, \dots , \nu''_e$ as in (i)-(ii), we get countably additive functionals

$$\nu_s = \nu'_s - \nu''_s, \quad \nu_{ac} = \nu'_{ac} - \nu''_{ac}, \quad \nu_{tc} = \nu'_{tc} - \nu''_{tc}, \quad \nu_e = \nu'_e - \nu''_e$$

such that ν_s is singular with respect to μ (if F', F'' are such that

$$\mu F = \mu F' = \nu'_s(X \setminus F) = \nu''_s(X \setminus F) = 0,$$

then $\mu(F' \cup F'') = 0$ and $\nu_s E = 0$ whenever $E \subseteq X \setminus (F' \cup F'')$), ν_{ac} is absolutely continuous with respect to μ, ν_{tc} is truly continuous with respect to μ, and $\nu_e F = 0$ whenever $\mu F < \infty$, while

$$\nu = \nu_s + \nu_{ac} = \nu_s + \nu_{tc} + \nu_e.$$

(iv) Moreover, these decompositions are unique. $\mathbf{P}(\boldsymbol{\alpha})$ If, for instance, $\nu = \tilde{\nu}_s + \tilde{\nu}_{ac}$, where $\tilde{\nu}_s$ is singular and $\tilde{\nu}_{ac}$ is absolutely continuous with respect to μ, let F, $\tilde{F}$ be such that $\mu F = \mu\tilde{F} = 0$ and $\tilde{\nu}_s E = 0$ whenever $E \cap \tilde{F} = \emptyset$, $\nu_s E = 0$ whenever $E \cap F = \emptyset$; then we must have

$$\nu_{ac}(E \cap (F \cup \tilde{F})) = \tilde{\nu}_{ac}(E \cap (F \cup \tilde{F})) = 0$$

for every $E \in \Sigma$, so

$$\nu_s E = \nu(E \cap (F \cup \tilde{F})) = \tilde{\nu}_s E$$

for every $E \in \Sigma$. Thus $\tilde{\nu}_s = \nu_s$ and $\tilde{\nu}_{ac} = \nu_{ac}$.

$(\boldsymbol{\beta})$ Similarly, if $\nu_{ac} = \tilde{\nu}_{tc} + \tilde{\nu}_e$ where $\tilde{\nu}_{tc}$ is truly continuous with respect to μ and $\tilde{\nu}_e F = 0$ whenever $\mu F < \infty$, then there are sequences $\langle F_n\rangle_{n\in\mathbb{N}}$, $\langle \tilde{F}_n\rangle_{n\in\mathbb{N}}$ of sets of finite measure such that $\nu_{tc}F = 0$ whenever $F \cap \bigcup_{n\in\mathbb{N}} F_n = \emptyset$ and $\tilde{\nu}_{tc}F = 0$ whenever $F \cap \bigcup_{n\in\mathbb{N}} \tilde{F}_n = \emptyset$. Write $F^* = \bigcup_{n\in\mathbb{N}}(F_n \cup \tilde{F}_n)$; then $\tilde{\nu}_e E = \nu_e E = 0$ whenever $E \subseteq F^*$ and $\tilde{\nu}_{tc}E = \nu_{tc}E = 0$ whenever $E \cap F^* = \emptyset$, so $\nu_e E = \nu_{ac}(E \setminus F^*) = \tilde{\nu}_e E$ for every $E \in \Sigma$, and $\nu_e = \tilde{\nu}_e$, $\nu_{tc} = \tilde{\nu}_{tc}$. **Q**

(b) In this case, μ is σ-finite (cf. 211P), so every absolutely continuous countably additive functional is truly continuous (232Bc), and we shall always have $\nu_e = 0$, $\nu_{ac} = \nu_{tc}$. But in the other direction we know that singleton sets, and therefore countable sets, are all measurable. We therefore have a further decomposition $\nu_s = \nu_p + \nu_{cs}$, where there is a countable set $K \subseteq \mathbb{R}^r$ with $\nu_p E = 0$ whenever $E \in \Sigma$, $E \cap K = \emptyset$, and ν_{cs} is singular with respect to μ and zero on countable sets. **P** (i) If $\nu \ge 0$, set

$$\nu_p E = \sup\{\nu(E \cap K) : K \subseteq \mathbb{R}^r \text{ is countable}\};$$

just as with ν_s, dealt with in (a) above, ν_p is countably additive and there is a countable $K \subseteq \mathbb{R}^r$ such that $\nu_p E = \nu(E \cap K)$ for every $E \in \Sigma$. (ii) For general ν, we can express ν as $\nu' - \nu''$ where ν' and ν'' are non-negative, and write $\nu_p = \nu'_p - \nu''_p$. (iii) ν_p is characterized by saying that there is a countable set K such that $\nu_p E = \nu(E \cap K)$ for every $E \in \Sigma$ and $\nu\{x\} = 0$ for every $x \in \mathbb{R}^r \setminus K$. (iv) So if we set $\nu_{cs} = \nu_s - \nu_p$, ν_{cs} will be singular with respect to μ and zero on countable sets. **Q**

Now, for any $E \in \Sigma$,

$$\nu_p E = \nu(E \cap K) = \sum_{x\in K\cap E} \nu\{x\} = \sum_{x\in E} \nu\{x\}.$$

Remark The expression $\nu = \nu_p + \nu_{cs} + \nu_{ac}$ of (b) is the **Lebesgue decomposition** of ν.

232X Basic exercises (a) Let (X, Σ, μ) be a measure space and $\nu : \Sigma \to \mathbb{R}$ a countably additive functional which is absolutely continuous with respect to μ. Show that the following are equiveridical: (i) ν is truly continuous with respect to μ; (ii) there is a sequence $\langle E_n\rangle_{n\in\mathbb{N}}$ in Σ such that $\mu E_n < \infty$ for every $n \in \mathbb{N}$ and $\nu F = 0$ whenever $F \in \Sigma$ and $F \cap \bigcup_{n\in\mathbb{N}} E_n = \emptyset$.

>(b) Let $g : \mathbb{R} \to \mathbb{R}$ be a bounded non-decreasing function and μ_g the associated Lebesgue-Stieltjes measure (114Xa). Show that μ_g is absolutely continuous (equivalently, truly continuous) with respect to Lebesgue measure iff the restriction of g to any closed bounded interval is absolutely continuous in the sense of 225B.

(c) Let X be a set and Σ a σ-algebra of subsets of X; let $\nu : \Sigma \to \mathbb{R}$ be a countably additive functional. Let $\mathcal{I}$ be an **ideal** of Σ, that is, a subset of Σ such that (α) $\emptyset \in \mathcal{I}$ (β) $E \cup F \in \mathcal{I}$ for all E, $F \in \mathcal{I}$ (γ) if $E \in \Sigma$, $F \in \mathcal{I}$ and $E \subseteq F$ then $E \in \mathcal{I}$. Show that ν has a unique decomposition as $\nu = \nu_{\mathcal{I}} + \nu'_{\mathcal{I}}$, where $\nu_{\mathcal{I}}$ and $\nu'_{\mathcal{I}}$ are countably additive functionals, $\nu'_{\mathcal{I}}E = 0$ for every $E \in \mathcal{I}$, and whenever $E \in \Sigma$, $\nu_{\mathcal{I}}E \neq 0$ there is an $F \in \mathcal{I}$ such that $\nu_{\mathcal{I}}(E \cap F) \neq 0$.

>(d) Let X be a non-empty set and Σ a σ-algebra of subsets of X. Show that for any sequence $\langle \nu_n \rangle_{n\in\mathbb{N}}$ of countably additive functionals on Σ there is a probability measure μ on X, with domain Σ, such that every ν_n is absolutely continuous with respect to μ. (*Hint*: start with the case $\nu_n \geq 0$.)

(e) Let (X, Σ, μ) be a measure space and $(X, \hat\Sigma, \hat\mu)$ its completion (212C). Let $\nu : \Sigma \to \mathbb{R}$ be an additive functional such that $\nu E = 0$ whenever $\mu E = 0$. Show that ν has a unique extension to an additive functional $\hat\nu : \hat\Sigma \to \mathbb{R}$ such that $\hat\nu E = 0$ whenever $\hat\mu E = 0$.

(f) Let $\mathcal{F}$ be an ultrafilter on $\mathbb{N}$ including the filter $\{\mathbb{N}\setminus I : I \subseteq \mathbb{N} \text{ is finite}\}$ (2A1O). Define $\nu : \mathcal{P}\mathbb{N} \to \{0, 1\}$ by setting $\nu E = 1$ if $E \in \mathcal{F}$, 0 for $E \in \mathcal{P}\mathbb{N} \setminus \mathcal{F}$. (i) Let μ_1 be counting measure on $\mathcal{P}\mathbb{N}$. Show that ν is additive and absolutely continuous with respect to μ_2, but is not truly continuous. (ii) Define $\mu_2 : \mathcal{P}\mathbb{N} \to [0, 1]$ by setting $\mu_2 E = \sum_{n\in E} 2^{-n-1}$. Show that ν is zero on μ_2-negligible sets, but is not absolutely continuous with respect to μ_2.

(g) Rewrite this section in terms of complex-valued additive functionals.

(h) Let (X, Σ, μ) be a measure space, and ν and λ additive functionals on Σ of which ν is positive and countably additive, so that (X, Σ, ν) also is a measure space. (i) Show that if ν is absolutely continuous with respect to μ and λ is absolutely continuous with respect to ν, then λ is absolutely continuous with respect to μ. (ii) Show that if ν is truly continuous with respect to μ and λ is absolutely continuous with respect to ν then λ is truly continuous with respect to μ.

232Y Further exercises (a) Let (X, Σ, μ) be a measure space and $\nu : \Sigma \to \mathbb{R}$ a finitely additive functional. If E, F, $H \in \Sigma$ and $\mu H < \infty$ set $\rho_H(E, F) = \mu(H \cap (E \triangle F))$. (i) Show that ρ_H is a pseudometric on Σ (2A3Fa). (ii) Let $\mathfrak{T}$ be the topology on Σ generated by $\{\rho_H : H \in \Sigma, \mu H < \infty\}$ (2A3Fc). Show that ν is continuous for $\mathfrak{T}$ iff it is truly continuous in the sense of 232Ab. ($\mathfrak{T}$ is the topology of **convergence in measure** on Σ.)

(b) For a non-decreasing function $F : [a, b] \to \mathbb{R}$, where $a < b$, let ν_F be the corresponding Lebesgue-Stieltjes measure. Show that if we define $(\nu_F)_{ac}$, etc., with regard to Lebesgue measure on $[a, b]$, as in 232I, then

$$(\nu_F)_p = \nu_{F_p}, \quad (\nu_F)_{ac} = \nu_{F_{ac}}, \quad (\nu_F)_{cs} = \nu_{F_{cs}},$$

where F_p, F_{cs} and F_{ac} are defined as in 226C.

(c) Extend the idea of (b) to general functions F of bounded variation.

(d) Extend the ideas of (b) and (c) to open, half-open and unbounded intervals (cf. 226Ya).

(e) Let (X, Σ, μ) be a measure space and $(X, \tilde\Sigma, \tilde\mu)$ its c.l.d. version (213E). Let $\nu : \Sigma \to \mathbb{R}$ be an additive functional which is truly continuous with respect to μ. Show that ν has a unique extension to a functional $\tilde\nu : \tilde\Sigma \to \mathbb{R}$ which is truly continuous with respect to $\tilde\mu$.

(f) Let (X, Σ, μ) be a measure space and f a μ-integrable real-valued function. Show that the indefinite integral of f is the unique countably additive functional $\nu : \Sigma \to \mathbb{R}$ such that whenever $E \in \Sigma$ and $f(x) \in [a, b]$ for almost every $x \in E$, then $a\mu E \leq \nu E \leq b\mu E$.

(g) Say that two bounded additive functionals ν_1, ν_2 on an algebra Σ of sets are **mutually singular** if for any $\epsilon > 0$ there is an $H \in \Sigma$ such that

$$\sup\{|\nu_1 F| : F \in \Sigma, F \subseteq H\} \leq \epsilon,$$

$$\sup\{|\nu_2 F| : F \in \Sigma, F \cap H = \emptyset\} \leq \epsilon.$$

(i) Show that ν_1 and ν_2 are mutually singular iff, in the language of 231Ya-231Yb, $|\nu_1| \wedge |\nu_2| = 0$.

(ii) Show that if Σ is a σ-algebra and ν_1 and ν_2 are countably additive, then they are mutually singular iff there is an $H \in \Sigma$ such that $\nu_1 F = 0$ whenever $F \in \Sigma$ and $F \subseteq H$, while $\nu_2 F = 0$ whenever $F \in \Sigma$ and $F \cap H = \emptyset$.

(iii) Show that if ν_s, ν_{tc} and ν_e are defined from ν and μ as in 232I, then each pair of the three are mutually singular.

(h) Let (X, Σ, μ) be a measure space and f a non-negative real-valued function which is integrable over X; let ν be its indefinite integral. Show that for any function $g : X \to \mathbb{R}$, $\int g\,d\nu = \int f \times g\,d\mu$ in the sense that if one of these is defined in $[-\infty, \infty]$ so is the other, and they are then equal. (*Hint*: start with simple functions g.)

(i) Let (X, Σ, μ) be a measure space, f an integrable function, and $\nu : \Sigma \to \mathbb{R}$ the indefinite integral of f. Show that $|\nu|$, as defined in 231Ya, is the indefinite integral of $|f|$.

(j) Let X be a set, Σ a σ-algebra of subsets of X, and $\nu : \Sigma \to \mathbb{R}$ a countably additive functional. Show that ν has a Radon-Nikodým derivative with respect to $|\nu|$ as defined in 231Ya, and that any such derivative has modulus equal to 1 $|\nu|$-a.e.

(k) (H.König) Let X be a set and μ, ν two measures on X with the same domain Σ. For $\alpha \geq 0$, $E \in \Sigma$ set $(\alpha\mu \wedge \nu)(E) = \inf\{\alpha\mu(E \cap F) + \nu(E \setminus F) : F \in \Sigma\}$ (cf. 112Ya[1]). Show that the following are equiveridical: (i) $\nu E = 0$ whenever $\mu E = 0$; (ii) $\sup_{\alpha \geq 0}(\alpha\mu \wedge \nu)(E) = \nu E$ for every $E \in \Sigma$.

232 Notes and comments The Radon-Nikodým theorem must be on any list of the half-dozen most important theorems of measure theory, and not only the theorem itself, but the techniques necessary to prove it, are at the heart of the subject. In my book FREMLIN 74 I discussed a variety of more or less abstract versions of the theorem and of the method, to some of which I will return in §§327 and 365 of the next volume.

As I have presented it here, the essence of the proof is split between 231E and 232E. I think we can distinguish the following elements. Let ν be a countably additive functional.

(i) ν is bounded (231Ea).

(ii) ν is expressible as the difference of non-negative functionals (231F).

(I gave this as a corollary of 231Eb, but it can also be proved by simpler methods, as in 231Ya.)

(iii) If $\nu > 0$, there is an integrable f such that $0 < \nu_f \leq \nu$,

writing ν_f for the indefinite integral of f. (This is the point at which we really do need the Hahn decomposition 231Eb.)

(iv) The set $\Psi = \{f : \nu_f \leq \nu\}$ is closed under countable suprema, so there is an $f \in \Psi$ maximising $\int f$.

(In part (b) of the proof of 232E, I spoke of simple functions; but this was solely to simplify the technical details, and the same argument works if we apply it to Ψ instead of Φ. Note the use here of B.Levi's theorem.)

(v) Take f from (iv) and use (iii) to show that $\nu - \nu_f = 0$.

Each of the steps (i)-(iv) requires a non-trivial idea, and the importance of the theorem lies not only in its remarkable direct consequences in the rest of this chapter and elsewhere, but in the versatility and power of these ideas.

I introduce the idea of 'truly continuous' functional in order to give a reasonably straightforward account of the status of the Radon-Nikodým theorem in non-σ-finite measure spaces. Of course the whole point is that a truly continuous functional, like an indefinite integral, must be concentrated on a σ-finite part of the space (232Xa), so that 232E, as stated, can be deduced easily from the standard form 232F. I dare to use the word 'truly' in this context because this kind of continuity does indeed correspond to a topological notion (232Ya).

There is a possible trap in the definition I give of 'absolutely continuous' functional. Many authors use the condition of 232Ba as a definition, saying that ν is absolutely continuous with respect to μ if $\nu E = 0$ whenever $\mu E = 0$. For countably additive functionals this coincides with the ϵ-δ formulation in 232Aa; but for other additive functionals this need not be so (232Xf(ii)). Mostly the distinction is insignificant, but I note that in 232Bd it is critical, since ν there is not assumed to be countably additive.

In 232I I describe one of the many ways of decomposing a countably additive functional into mutually singular parts with special properties. In 231Yf-231Yg I have already suggested a method of decomposing an additive functional into the sum of a countably additive part and a 'purely finitely additive' part. All these results have natural expressions in terms of the ordered linear space of bounded additive functionals on an algebra (231Yc).

[1] Formerly 112Yb.

233 Conditional expectations

I devote a section to a first look at one of the principal applications of the Radon-Nikodým theorem. It is one of the most vital ideas of measure theory, and will appear repeatedly in one form or another. Here I give the definition and most basic properties of conditional expectations as they arise in abstract probability theory, with notes on convex functions and a version of Jensen's inequality (233I-233J).

233A σ-subalgebras Let X be a set and Σ a σ-algebra of subsets of X. A **σ-subalgebra** of Σ is a σ-algebra T of subsets of X such that $\mathrm{T} \subseteq \Sigma$. If (X, Σ, μ) is a measure space and T is a σ-subalgebra of Σ, then $(X, \mathrm{T}, \mu\restriction\mathrm{T})$ is again a measure space; this is immediate from the definition (112A). Now we have the following straightforward lemma. It is a special case of 235G below, but I give a separate proof in case you do not wish as yet to embark on the general investigation pursued in §235.

233B Lemma Let (X, Σ, μ) be a measure space and T a σ-subalgebra of Σ. A real-valued function f defined on a subset of X is $\mu\restriction\mathrm{T}$-integrable iff (i) it is μ-integrable (ii) $\operatorname{dom} f$ is $\mu\restriction\mathrm{T}$-conegligible (iii) f is $\mu\restriction\mathrm{T}$-virtually measurable; and in this case $\int f d(\mu\restriction\mathrm{T}) = \int f d\mu$.

proof (a) Note first that if f is a $\mu\restriction\mathrm{T}$-simple function, that is, is expressible as $\sum_{i=0}^n a_i\chi E_i$ where $a_i \in \mathbb{R}$, $E_i \in \mathrm{T}$ and $(\mu\restriction\mathrm{T})E_i < \infty$ for each i, then f is μ-simple and

$$\int f d\mu = \sum_{i=0}^n a_i \mu E_i = \int f d(\mu\restriction\mathrm{T}).$$

(b) Let U_μ be the set of non-negative μ-integrable functions and $U_{\mu\restriction\mathrm{T}}$ the set of non-negative $\mu\restriction\mathrm{T}$-integrable functions.

Suppose $f \in U_{\mu\restriction\mathrm{T}}$. Then there is a non-decreasing sequence $\langle f_n\rangle_{n\in\mathbb{N}}$ of $\mu\restriction\mathrm{T}$-simple functions such that $f(x) = \lim_{n\to\infty} f_n$ $\mu\restriction\mathrm{T}$-a.e. and

$$\int f d(\mu\restriction\mathrm{T}) = \lim_{n\to\infty} \int f_n d(\mu\restriction\mathrm{T}).$$

But now every f_n is also μ-simple, and $\int f_n d\mu = \int f_n d(\mu\restriction\mathrm{T})$ for every n, and $f = \lim_{n\to\infty} f_n$ μ-a.e. So $f \in U_\mu$ and $\int f d\mu = \int f d(\mu\restriction\mathrm{T})$.

(c) Now suppose that f is $\mu\restriction\mathrm{T}$-integrable. Then it is the difference of two members of $U_{\mu\restriction\mathrm{T}}$, so is μ-integrable, and $\int f d\mu = \int f d(\mu\restriction\mathrm{T})$. Also conditions (ii) and (iii) are satisfied, according to the conventions established in Volume 1 (122Nc, 122P-122Q).

(d) Suppose that f satisfies conditions (i)-(iii). Then $|f| \in U_\mu$, and there is a conegligible set $E \subseteq \operatorname{dom} f$ such that $E \in \mathrm{T}$ and $f\restriction E$ is T-measurable. Accordingly $|f|\restriction E$ is T-measurable. Now, if $\epsilon > 0$, then

$$(\mu\restriction\mathrm{T})\{x : x \in E, |f|(x) \ge \epsilon\} = \mu\{x : x \in E, |f|(x) \ge \epsilon\} \le \tfrac{1}{\epsilon}\int |f| d\mu < \infty;$$

moreover,

$$\begin{aligned}
\sup\{\int g\, d(\mu\restriction\mathrm{T}) : g \text{ is a } \mu\restriction\mathrm{T}\text{-simple function}, g \le |f|\ \mu\restriction\mathrm{T}\text{-a.e.}\}&\\
= \sup\{\int g\, d\mu : g \text{ is a } \mu\restriction\mathrm{T}\text{-simple function}, g \le |f|\ \mu\restriction\mathrm{T}\text{-a.e.}\}&\\
\le \sup\{\int g\, d\mu : g \text{ is a } \mu\text{-simple function}, g \le |f|\ \mu\text{-a.e.}\}&\\
\le \int |f| d\mu < \infty.&
\end{aligned}$$

By the criterion of 122Ja, $|f| \in U_{\mu\restriction\mathrm{T}}$. Consequently f, being $\mu\restriction\mathrm{T}$-virtually T-measurable, is $\mu\restriction\mathrm{T}$-integrable, by 122P. This completes the proof.

233C Remarks (a) My argument just above is detailed to the point of pedantry. I think, however, that while I can be accused of wasting paper by writing everything down, every element of the argument is necessary to the result. To be sure, some of the details are needed only because I use such a wide notion of 'integrable function'; if you restrict the notion of 'integrability' to measurable functions defined on the whole measure space, there are simplifications at this stage, to be paid for later when you discover that many of the principal applications are to functions defined by formulae which do not apply on the whole underlying space.

The essential point which does have to be grasped is that while a $\mu\restriction\mathrm{T}$-negligible set is always μ-negligible, a μ-negligible set need not be $\mu\restriction\mathrm{T}$-negligible.

(b) As the simplest possible example of the problems which can arise, I offer the following. Let (X, Σ, μ) be $[0,1]^2$ with Lebesgue measure. Let T be the set of those members of Σ expressible as $F \times [0,1]$ for some $F \subseteq [0,1]$; it is easy to see that T is a σ-subalgebra of Σ. Consider f, $g : X \to [0,1]$ defined by saying that

$$f(t,u) = 1 \text{ if } u > 0, \ 0 \text{ otherwise,}$$

$$g(t,u) = 1 \text{ if } t > 0, \ 0 \text{ otherwise.}$$

Then both f and g are μ-integrable, being constant μ-a.e. But only g is $\mu{\restriction}\mathrm{T}$-integrable, because any non-negligible $E \in \mathrm{T}$ includes a complete vertical section $\{t\} \times [0,1]$, so that f takes both values 0 and 1 on E. If we set

$$h(t,u) = 1 \text{ if } u > 0, \text{ undefined otherwise,}$$

then again (on the conventions I use) h is μ-integrable but not $\mu{\restriction}\mathrm{T}$-integrable, as there is no conegligible member of T included in the domain of h.

(c) If f is defined everywhere in X, and $\mu{\restriction}\mathrm{T}$ is complete, then of course f is $\mu{\restriction}\mathrm{T}$-integrable iff it is μ-integrable and T-measurable. But note that in the example just above, which is one of the archetypes for this topic, $\mu{\restriction}\mathrm{T}$ is not complete, as singleton sets are negligible but not measurable.

233D Conditional expectations Let (X, Σ, μ) be a probability space, that is, a measure space with $\mu X = 1$. (Nearly all the ideas here work perfectly well for any totally finite measure space, but there seems nothing to be gained from the extension, and the traditional phrase 'conditional expectation' demands a probability space.) Let $\mathrm{T} \subseteq \Sigma$ be a σ-subalgebra.

(a) For any μ-integrable real-valued function f defined on a conegligible subset of X, we have a corresponding indefinite integral $\nu_f : \Sigma \to \mathbb{R}$ given by the formula $\nu_f E = \int_E f$ for every $E \in \Sigma$. We know that ν_f is countably additive and truly continuous with respect to μ, which in the present context is the same as saying that it is absolutely continuous (232Bc-232Bd). Now consider the restrictions $\mu{\restriction}\mathrm{T}$, $\nu_f{\restriction}\mathrm{T}$ of μ and ν_f to the σ-algebra T. It follows directly from the definitions of 'countably additive' and 'absolutely continuous' that $\nu_f{\restriction}\mathrm{T}$ is countably additive and absolutely continuous with respect to $\mu{\restriction}\mathrm{T}$, therefore truly continuous with respect to $\mu{\restriction}\mathrm{T}$. Consequently, the Radon-Nikodým theorem (232E) tells us that there is a $\mu{\restriction}\mathrm{T}$-integrable function g such that $(\nu_f{\restriction}\mathrm{T})F = \int_F g\, d(\mu{\restriction}\mathrm{T})$ for every $F \in \mathrm{T}$.

(b) Let us define **a conditional expectation of** f **on** T to be such a function; that is, a $\mu{\restriction}\mathrm{T}$-integrable function g such that $\int_F g\, d(\mu{\restriction}\mathrm{T}) = \int_F f d\mu$ for every $F \in \mathrm{T}$. Looking back at 233B, we see that for such a g we have

$$\int_F g\, d(\mu{\restriction}\mathrm{T}) = \int g \times \chi F\, d(\mu{\restriction}\mathrm{T}) = \int g \times \chi F\, d\mu = \int_F g\, d\mu$$

for every $F \in \mathrm{T}$; also, that g is almost everywhere equal to a T-measurable function defined everywhere in X which is also a conditional expectation of f on T (232He).

(c) I set the word 'a' of the phrase 'a conditional expectation' in bold type to emphasize that there is nothing unique about the function g. In 242J I will return to this point, and describe an object which could properly be called 'the' conditional expectation of f on T. g is 'essentially unique' only in the sense that if g_1, g_2 are both conditional expectations of f on T then $g_1 = g_2$ $\mu{\restriction}\mathrm{T}$-a.e. (131Hb). This does of course mean that a very large number of its properties – for instance, the distribution function $G(a) = \hat{\mu}\{x : g(x) \le a\}$, where $\hat{\mu}$ is the completion of μ (212C) – are independent of which g we take.

(d) A word of explanation of the phrase 'conditional expectation' is in order. This derives from the standard identification of probability with measure, due to Kolmogorov, which I will discuss more fully in Chapter 27. A real-valued random variable may be regarded as a measurable, or virtually measurable, function f on a probability space (X, Σ, μ); its 'expectation' becomes identified with $\int f d\mu$, supposing that this exists. If $F \in \Sigma$ and $\mu F > 0$ then the 'conditional expectation of f given F' is $\frac{1}{\mu F}\int_F f$. If $F_0, \dots, F_n$ is a partition of X into measurable sets of non-zero measure, then the function g given by

$$g(x) = \frac{1}{\mu F_i}\int_{F_i} f \text{ if } x \in F_i$$

is a kind of anticipated conditional expectation; if we are one day told that $x \in F_i$, then $g(x)$ will be our subsequent estimate of the expectation of f. In the terms of the definition above, g is a conditional expectation of f on the finite algebra T generated by $\{F_0, \dots, F_n\}$. An appropriate intuition for general σ-algebras T is that they consist of the events which we shall be able to observe at some stated future time t_0, while the whole algebra Σ consists of all events, including those not observable until times later than t_0, if ever.

233E I list some of the elementary facts concerning conditional expectations.

Proposition Let (X, Σ, μ) be a probability space and T a σ-subalgebra of Σ. Let $\langle f_n\rangle_{n\in\mathbb{N}}$ be a sequence of μ-integrable real-valued functions, and for each n let g_n be a conditional expectation of f_n on T. Then

(a) $g_1 + g_2$ is a conditional expectation of $f_1 + f_2$ on T;

(b) for any $c \in \mathbb{R}$, cg_0 is a conditional expectation of cf_0 on T;

(c) if $f_1 \le_{\text{a.e.}} f_2$ then $g_1 \le_{\text{a.e.}} g_2$;

(d) if $\langle f_n\rangle_{n\in\mathbb{N}}$ is non-decreasing a.e. and $f = \lim_{n\to\infty} f_n$ is μ-integrable, then $\lim_{n\to\infty} g_n$ is a conditional expectation of f on T;

(e) if $f = \lim_{n\to\infty} f_n$ is defined a.e. and there is a μ-integrable function h such that $|f_n| \le_{\text{a.e.}} h$ for every n, then $\lim_{n\to\infty} g_n$ is a conditional expectation of f on T;

(f) if $F \in$ T then $g_0 \times \chi F$ is a conditional expectation of $f_0 \times \chi F$ on T;

(g) if h is a bounded, $\mu{\restriction}$T-virtually measurable real-valued function defined $\mu{\restriction}$T-almost everywhere in X, then $g_0 \times h$ is a conditional expectation of $f_0 \times h$ on T;

(h) if Υ is a σ-subalgebra of T, then a function h_0 is a conditional expectation of f_0 on Υ iff it is a conditional expectation of g_0 on Υ.

proof (a)-(b) We have only to observe that

$$\int_F g_1 + g_2 d(\mu{\restriction}\mathrm{T}) = \int_F g_1 d(\mu{\restriction}\mathrm{T}) + \int_F g_2 d(\mu{\restriction}\mathrm{T}) = \int_F f_1 d\mu + \int_F f_2 d\mu = \int_F f_1 + f_2 d\mu,$$

$$\int_F cg_0 d(\mu{\restriction}\mathrm{T}) = c\int_F g_0 d(\mu{\restriction}\mathrm{T}) = c\int_F f_0 d\mu = \int_F cf_0 d\mu$$

for every $F \in$ T.

(c) If $F \in$ T then

$$\int_F g_1 d(\mu{\restriction}\mathrm{T}) = \int_F f_1 d\mu \le \int_F f_2 d\mu = \int_F g_2 d(\mu{\restriction}\mathrm{T})$$

for every $F \in$ T; consequently $g_1 \le g_2$ $\mu{\restriction}$T-a.e. (131Ha).

(d) By (c), $\langle g_n\rangle_{n\in\mathbb{N}}$ is non-decreasing $\mu{\restriction}$T-a.e.; moreover,

$$\sup_{n\in\mathbb{N}} \int g_n d(\mu{\restriction}\mathrm{T}) = \sup_{n\in\mathbb{N}} \int f_n d\mu = \int f d\mu < \infty.$$

By B.Levi's theorem, $g = \lim_{n\to\infty} g_n$ is defined $\mu{\restriction}$T-almost everywhere, and

$$\int_F g\, d(\mu{\restriction}\mathrm{T}) = \lim_{n\to\infty} \int_F g_n d(\mu{\restriction}\mathrm{T}) = \lim_{n\to\infty} \int_F f_n d\mu = \int_F f d\mu$$

for every $F \in$ T, so g is a conditional expectation of f on T.

(e) Set $f'_n = \inf_{m\ge n} f_m$, $f''_n = \sup_{m\ge n} f_m$ for each $n \in \mathbb{N}$. Then we have

$$-h \le_{\text{a.e.}} f'_n \le f_n \le f''_n \le_{\text{a.e.}} h,$$

and $\langle f'_n\rangle_{n\in\mathbb{N}}$, $\langle f''_n\rangle_{n\in\mathbb{N}}$ are almost-everywhere-monotonic sequences of functions both converging almost everywhere to f. For each n, let g'_n, g''_n be conditional expectations of f'_n, f''_n on T. By (iii) and (iv), $\langle g'_n\rangle_{n\in\mathbb{N}}$ and $\langle g''_n\rangle_{n\in\mathbb{N}}$ are almost-everywhere-monotonic sequences converging almost everywhere to conditional expectations g', g'' of f. Of course $g' = g''$ $\mu{\restriction}$T-a.e. (233Dc). Also, for each n, $g'_n \le_{\text{a.e.}} g_n \le_{\text{a.e.}} g''_n$, so $\langle g_n\rangle_{n\in\mathbb{N}}$ converges to g' $\mu{\restriction}$T-a.e., and $g = \lim_{n\to\infty} g_n$ is defined almost everywhere and is a conditional expectation of f on T.

(f) For any $H \in$ T,

$$\int_H g_0 \times \chi F\, d(\mu{\restriction}\mathrm{T}) = \int_{H\cap F} g_0 d(\mu{\restriction}\mathrm{T}) = \int_{H\cap F} f_0 d\mu = \int_H f_0 \times \chi F\, d\mu.$$

(g)(i) If h is actually $(\mu{\restriction}\mathrm{T})$-simple, say $h = \sum_{i=0}^n a_i \chi F_i$ where $F_i \in$ T for each i, then

$$\int_F g_0 \times h\, d(\mu{\restriction}\mathrm{T}) = \sum_{i=0}^n a_i \int_F g_0 \times \chi F_i d(\mu{\restriction}\mathrm{T}) = \sum_{i=0}^n a_i \int_F f \times \chi F_i\, d\mu = \int_F f \times h\, d\mu$$

for every $F \in$ T. (ii) For the general case, if h is $\mu{\restriction}$T-virtually measurable and $|h(x)| \le M$ $\mu{\restriction}$T-almost everywhere, then there is a sequence $\langle h_n\rangle_{n\in\mathbb{N}}$ of $\mu{\restriction}$T-simple functions converging to h almost everywhere, and with $|h_n(x)| \le M$ for every x, n. Now $f_0 \times h_n \to f_0 \times h$ a.e. and $|f_0 \times h_n| \le_{\text{a.e.}} M|f_0|$ for each n, while $g_0 \times h_n$ is a conditional expectation of $f_0 \times h_n$ for every n, so by (e) we see that $\lim_{n\to\infty} g_0 \times h_n$ will be a conditional expectation of $f_0 \times h$; but this is equal almost everywhere to $g_0 \times h$.

(h) We need note only that $\int_H g_0 d(\mu{\restriction}\mathrm{T}) = \int_H f_0 d\mu$ for every $H \in \Upsilon$, so

$$\begin{aligned}\int_H h_0 d(\mu\restriction\Upsilon) = \int_H g_0 d(\mu\restriction\mathrm{T}) \text{ for every } H\in\Upsilon\\ \iff \int_H h_0 d(\mu\restriction\Upsilon) = \int_H f_0 d\mu \text{ for every } H\in\Upsilon.\end{aligned}$$

233F Remarks Of course the results above are individually nearly trivial (though I think (e) and (g) might give you pause for thought if they were offered without previous preparation of the ground). Cumulatively they amount to some quite strong properties. In §242 I will restate them in language which is syntactically more direct, but relies on a deeper level of abstraction.

As an illustration of the power of conditional expectations to surprise us, I offer the next proposition, which depends on the concept of 'convex' function.

233G Convex functions Recall that a real-valued function ϕ defined on an interval $I\subseteq\mathbb{R}$ is **convex** if

$$\phi(tb+(1-t)c)\le t\phi(b)+(1-t)\phi(c)$$

whenever b, $c\in I$ and $t\in[0,1]$.

Examples The formulae $|x|$, x^2, $e^{\pm x}\pm x$ define convex functions on $\mathbb{R}$; on $]-1,1[$ we have $1/(1-x^2)$; on $]0,\infty[$ we have $1/x$ and $x\ln x$; on $[0,1]$ we have the function which is zero on $]0,1[$ and 1 on $\{0,1\}$.

233H The general theory of convex functions is both extensive and important; I list a few of their more salient properties in 233Xe. For the moment the following lemma covers what we need.

Lemma Let $I\subseteq\mathbb{R}$ be a non-empty open interval (bounded or unbounded) and $\phi:I\to\mathbb{R}$ a convex function.

(a) For every $a\in I$ there is a $b\in\mathbb{R}$ such that $\phi(x)\ge\phi(a)+b(x-a)$ for every $x\in I$.

(b) If we take, for each $q\in I\cap\mathbb{Q}$, a $b_q\in\mathbb{R}$ such that $\phi(x)\ge\phi(q)+b_q(x-q)$ for every $x\in I$, then

$$\phi(x)=\sup_{q\in I\cap\mathbb{Q}}\phi(q)+b_q(x-q)$$

for every $x\in I$.

(c) ϕ is Borel measurable.

proof (a) If c, $c'\in I$ and $c<a<c'$, then a is expressible as $dc+(1-d)c'$ for some $d\in\,]0,1[$, so that $\phi(a)\le d\phi(c)+(1-d)\phi(c')$ and

$$\begin{aligned}\frac{\phi(a)-\phi(c)}{a-c}&\le\frac{d\phi(c)+(1-d)\phi(c')-\phi(c)}{dc+(1-d)c'-c}=\frac{(1-d)(\phi(c')-\phi(c))}{(1-d)(c'-c)}\\&=\frac{d(\phi(c')-\phi(c))}{d(c'-c)}=\frac{\phi(c')-d\phi(c)-(1-d)\phi(c')}{c'-dc-(1-d)c'}\le\frac{\phi(c')-\phi(a)}{c'-a}.\end{aligned}$$

This means that

$$b=\sup_{c<a,c\in I}\frac{\phi(a)-\phi(c)}{a-c}$$

is finite, and $b\le\frac{\phi(c')-\phi(a)}{c'-a}$ whenever $a<c'\in I$; accordingly $\phi(x)\ge\phi(a)+b(x-a)$ for every $x\in I$.

(b) By the choice of the b_q, $\phi(x)\ge\sup_{q\in\mathbb{Q}}\phi_q(x)$. On the other hand, given $x\in I$, fix $y\in I$ such that $x<y$ and let $b\in\mathbb{R}$ be such that $\phi(z)\ge\phi(x)+b(z-x)$ for every $z\in I$. If $q\in\mathbb{Q}$ and $x<q<y$, we have $\phi(y)\ge\phi(q)+b_q(y-q)$, so that $b_q\le\frac{\phi(y)-\phi(q)}{y-q}$ and

$$\begin{aligned}\phi(q)+b_q(x-q)&=\phi(q)-b_q(q-x)\ge\phi(q)-\frac{\phi(y)-\phi(q)}{y-q}(q-x)\\&=\frac{y-x}{y-q}\phi(q)-\frac{q-x}{y-q}\phi(y)\ge\frac{y-x}{y-q}(\phi(x)+b(q-x))-\frac{q-x}{y-q}\phi(y).\end{aligned}$$

Now

$$\begin{aligned}\phi(x) &= \lim_{q\downarrow x} \tfrac{y-x}{y-q}(\phi(x)+b(q-x)) - \tfrac{q-x}{y-q}\phi(y)\\ &\le \sup_{q\in\mathbb{Q}\cap]x,y[} \tfrac{y-x}{y-q}(\phi(x)+b(q-x)) - \tfrac{q-x}{y-q}\phi(y)\\ &\le \sup_{q\in\mathbb{Q}\cap]x,y[} \phi(q)+b_q(x-q) \le \sup_{q\in\mathbb{Q}\cap I} \phi(q)+b_q(x-q).\end{aligned}$$

(c) Writing $\phi_q(x)=\phi(q)+b_q(x-q)$ for every $q\in\mathbb{Q}\cap I$, every ϕ_q is a Borel measurable function, and $\phi=\sup_{q\in I\cap\mathbb{Q}}\phi_q$ is the supremum of a countable family of Borel measurable functions, so is Borel measurable.

233I Jensen's inequality Let (X,Σ,μ) be a measure space and $\phi:\mathbb{R}\to\mathbb{R}$ a convex function.

(a) Suppose that f and g are real-valued μ-virtually measurable functions defined almost everywhere in X and that $g\ge 0$ almost everywhere, $\int g=1$ and $g\times f$ is integrable. Then $\phi(\int g\times f)\le\int g\times\phi f$, where we may need to interpret the right-hand integral as ∞.

(b) In particular, if $\mu X=1$ and f is a real-valued function which is integrable over X, then $\phi(\int f)\le\int\phi f$.

proof (a) For each $q\in\mathbb{Q}$ take b_q such that $\phi(t)\ge\phi_q(t)=\phi(q)+b_q(t-q)$ for every $t\in\mathbb{R}$ (233Ha). Because ϕ is Borel measurable (233Hc), ϕf is μ-virtually measurable (121H), so $g\times\phi f$ also is; since $g\times\phi f$ is defined almost everywhere and almost everywhere greater than or equal to the integrable function $g\times\phi_0 f$, $\int g\times\phi f$ is defined in $]-\infty,\infty]$. Now

$$\begin{aligned}\phi_q\Bigl(\int g\times f\Bigr) &= \phi(q)+b_q\int g\times f-b_q q\\ &= \int g\times(b_q f+(\phi(q)-b_q q)\chi X)=\int g\times\phi_q f\le\int g\times\phi f,\end{aligned}$$

because $\int g=1$ and $g\ge 0$ a.e. By 233Hb,

$$\phi(\textstyle\int g\times f)=\sup_{q\in\mathbb{Q}}\phi_q(\int g\times f)\le\int g\times\phi f.$$

(b) Take g to be the constant function with value 1.

233J Even the special case 233Ib of Jensen's inequality is already very useful. It can be extended as follows.

Theorem Let (X,Σ,μ) be a probability space and T a σ-subalgebra of Σ. Let $\phi:\mathbb{R}\to\mathbb{R}$ be a convex function and f a μ-integrable real-valued function defined almost everywhere in X such that the composition ϕf is also integrable. If g and h are conditional expectations on T of f, ϕf respectively, then $\phi g\le_{\text{a.e.}} h$. Consequently $\int\phi g\le\int\phi f$.

proof We use the same ideas as in 233I. For each $q\in\mathbb{Q}$ take a $b_q\in\mathbb{R}$ such that $\phi(t)\ge\phi_q(t)=\phi(q)+b_q(t-q)$ for every $t\in\mathbb{R}$, so that $\phi(t)=\sup_{q\in\mathbb{Q}}\phi_q(t)$ for every $t\in\mathbb{R}$. Now setting

$$\psi_q(x)=\phi(q)+b_q(g(x)-q)$$

for $x\in\operatorname{dom} g$, we see that $\psi_q=\phi_q g$ is a conditional expectation of $\phi_q f$, and as $\phi_q f\le_{\text{a.e.}}\phi f$ we must have $\psi_q\le_{\text{a.e.}} h$. But also $\phi g=\sup_{q\in\mathbb{Q}}\psi_q$ wherever g is defined, so $\phi g\le_{\text{a.e.}} h$, as claimed.

It follows at once that $\int\phi g\le\int h=\int\phi f$.

233K I give the following proposition, an elaboration of 233Eg, in a very general form, as its applications can turn up anywhere.

Proposition Let (X,Σ,μ) be a probability space, and T a σ-subalgebra of Σ. Suppose that f is a μ-integrable function and h is a $(\mu\restriction\mathrm{T})$-virtually measurable real-valued function defined $(\mu\restriction\mathrm{T})$-almost everywhere in X. Let g, g_0 be conditional expectations of f and $|f|$ on T. Then $f\times h$ is integrable iff $g_0\times h$ is integrable, and in this case $g\times h$ is a conditional expectation of $f\times h$ on T.

proof (a) Suppose that h is a $\mu\restriction\mathrm{T}$-simple function. Then surely $f\times h$ and $g_0\times h$ are integrable, and $g\times h$ is a conditional expectation of $f\times h$ as in 233Eg.

(b) Now suppose that f, $h\ge 0$. Then $g=g_0\ge 0$ a.e. (233Ec). Let $\tilde h$ be a non-negative T-measurable function defined everywhere in X such that $h=_{\text{a.e.}}\tilde h$. For each $n\in\mathbb{N}$ set

$$\begin{aligned} h_n(x) &= 2^{-n}k \text{ if } 0 \le k < 4^n \text{ and } 2^{-n}k \le \tilde h(x) < 2^{-n}(k+1),\\ &= 2^n \text{ if } \tilde h(x) \ge 2^{-n}. \end{aligned}$$

Then h_n is a $(\mu\restriction\mathrm{T})$-simple function, so $g\times h_n$ is a conditional expectation of $f\times h_n$. Both $\langle f\times h_n\rangle_{n\in\mathbb{N}}$ and $\langle g\times h_n\rangle_{n\in\mathbb{N}}$ are almost everywhere non-decreasing sequences of integrable functions, with limits $f\times h$ and $g\times h$ respectively. By B.Levi's theorem,

$$\begin{aligned} f\times h \text{ is integrable} &\iff f\times\tilde h \text{ is integrable}\\ &\iff \sup_{n\in\mathbb{N}}\int f\times h_n<\infty \iff \sup_{n\in\mathbb{N}}\int g\times h_n<\infty \end{aligned}$$

(because $\int g\times h_n=\int f\times h_n$ for each n)

$$\iff g\times h \text{ is integrable} \iff g_0\times h \text{ is integrable.}$$

Moreover, in this case

$$\begin{aligned} \int_E f\times h &= \int_E f\times\tilde h = \lim_{n\to\infty}\int_E f\times h_n\\ &= \lim_{n\to\infty}\int_E g\times h_n = \int_E g\times\tilde h = \int_E g\times h \end{aligned}$$

for every $E\in\mathrm{T}$, while $g\times h$ is $(\mu\restriction\mathrm{T})$-virtually measurable, so $g\times h$ is a conditional expectation of $f\times h$.

(c) Finally, consider the general case of integrable f and virtually measurable h. Set $f^+=f\vee 0$, $f^-=(-f)\vee 0$, so that $f=f^+-f^-$ and $0\le f^+,\,f^-\le|f|$; similarly, set $h^+=h\vee 0$, $h^-=(-h)\vee 0$. Let g_1, g_2 be conditional expectations of f^+, f^- on T. Because $0\le f^+,\,f^-\le|f|$, $0\le g_1,\,g_2\le_{\text{a.e.}} g_0$, while $g=_{\text{a.e.}} g_1-g_2$.

We see that

$$\begin{aligned} f\times h \text{ is integrable} &\iff |f|\times|h|=|f\times h| \text{ is integrable}\\ &\iff g_0\times|h| \text{ is integrable}\\ &\iff g_0\times h \text{ is integrable.} \end{aligned}$$

And in this case all four of $f^+\times h^+,\dots,f^-\times h^-$ are integrable, so

$$(g_1-g_2)\times h = g_1\times h^+ - g_2\times h^+ - g_1\times h^- + g_2\times h^-$$

is a conditional expectation of

$$f^+\times h^+ - f^-\times h^+ - f^+\times h^- + f^-\times h^- = f\times h.$$

Since $g\times h=_{\text{a.e.}}(g_1-g_2)\times h$, this also is a conditional expectation of $f\times h$, and we're done.

233X Basic exercises (a) Let (X,Σ,μ) be a probability space and T a σ-subalgebra of Σ. Let $\langle f_n\rangle_{n\in\mathbb{N}}$ be a sequence of non-negative μ-integrable functions and suppose that g_n is a conditional expectation of f_n on T for each n. Suppose that $f=\liminf_{n\to\infty}f_n$ is integrable and has a conditional expectation g. Show that $g\le_{\text{a.e.}}\liminf_{n\to\infty}g_n$.

(b) Let $I\subseteq\mathbb{R}$ be an interval, and $\phi:I\to\mathbb{R}$ a function. Show that ϕ is convex iff $\{x:x\in I,\,\phi(x)+bx\le c\}$ is an interval for every b, $c\in\mathbb{R}$.

>(c) Let $I\subseteq\mathbb{R}$ be an open interval and $\phi:I\to\mathbb{R}$ a function. (i) Show that if ϕ is differentiable then it is convex iff ϕ' is non-decreasing. (ii) Show that if ϕ is absolutely continuous on every bounded closed subinterval of I then ϕ is convex iff ϕ' is non-decreasing on its domain.

(d) For any $r\ge1$, a subset C of $\mathbb{R}^r$ is **convex** if $tx+(1-t)y\in C$ for all x, $y\in C$ and $t\in[0,1]$. If $C\subseteq\mathbb{R}^r$ is convex, then a function $\phi:C\to\mathbb{R}$ is **convex** if $\phi(tx+(1-t)y)\le t\phi(x)+(1-t)\phi(y)$ for all x, $y\in C$ and $t\in[0,1]$.

Let $C\subseteq\mathbb{R}^r$ be a convex set and $\phi:C\to\mathbb{R}$ a function. Show that the following are equiveridical: (i) the function ϕ is convex; (ii) the set $\{(x,t):x\in C,\,t\in\mathbb{R},\,t\ge\phi(x)\}$ is convex in $\mathbb{R}^{r+1}$; (iii) the set $\{x:x\in C,\,\phi(x)+b\centerdot x\le c\}$ is convex in $\mathbb{R}^r$ for every $b\in\mathbb{R}^r$ and $c\in\mathbb{R}$, writing $b\centerdot x=\sum_{i=1}^r\beta_i\xi_i$ if $b=(\beta_1,\dots,\beta_r)$ and $x=(\xi_1,\dots,\xi_r)$.

(e) Let $I \subseteq \mathbb{R}$ be an interval and $\phi : I \to \mathbb{R}$ a convex function.

(i) Show that if a, $d \in I$ and $a < b \le c < d$ then

$$\frac{\phi(b)-\phi(a)}{b-a} \le \frac{\phi(d)-\phi(c)}{d-c}.$$

(ii) Show that ϕ is continuous at every interior point of I.

(iii) Show that *either* ϕ is monotonic on I *or* there is a $c \in I$ such that $\phi(c) = \min_{x\in I}\phi(x)$ and ϕ is non-increasing on $I \cap \,]-\infty, c]$, monotonic non-decreasing on $I \cap [c, \infty[$.

(iv) Show that ϕ is differentiable at all but countably many points of I, and that its derivative is non-decreasing in the sense that $\phi'(x) \le \phi'(y)$ whenever x, $y \in \operatorname{dom}\phi'$ and $x \le y$.

(v) Show that if I is closed and bounded and ϕ is continuous then ϕ is absolutely continuous.

(vi) Show that if I is closed and bounded and $\psi : I \to \mathbb{R}$ is absolutely continuous with a non-decreasing derivative then ψ is convex.

(f) Show that if $I \subseteq \mathbb{R}$ is an interval and ϕ, $\psi : I \to \mathbb{R}$ are convex functions so is $a\phi + b\psi$ for any a, $b \ge 0$.

(g) In the context of 233K, give an example in which $g \times h$ is integrable but $f \times h$ is not. (*Hint*: take X, μ, T as in 233Cb, and arrange for g to be 0.)

(h) Let $I \subseteq \mathbb{R}$ be an interval and Φ a non-empty family of convex real-valued functions on I such that $\psi(x) = \sup_{\phi\in\Phi}\phi(x)$ is finite for every $x \in I$. Show that ψ is convex.

233Y Further exercises (a) If $I \subseteq \mathbb{R}$ is an interval, a function $\phi : I \to \mathbb{R}$ is **mid-convex** if $\phi(\frac{x+y}{2}) \le \frac{1}{2}(\phi(x)+\phi(y))$ for all x, $y \in I$. Show that a mid-convex function which is bounded on any non-trivial subinterval of I is convex.

(b) Generalize 233Xd to arbitrary normed spaces in place of $\mathbb{R}^r$.

(c) Let (X, Σ, μ) be a probability space and T a σ-subalgebra of Σ. Let ϕ be a convex real-valued function with domain an interval $I \subseteq \mathbb{R}$, and f an integrable real-valued function on X such that $f(x) \in I$ for almost every $x \in X$ and ϕf is integrable. Let g, h be conditional expectations on T of f, ϕf respectively. Show that $g(x) \in I$ for almost every x and that $\phi g \le_{\text{a.e.}} h$.

(d)(i) Show that if $I \subseteq \mathbb{R}$ is a bounded interval, $E \subseteq I$ is Lebesgue measurable, and $\mu E > \frac{2}{3}\mu I$ where μ is Lebesgue measure, then for every $x \in I$ there are y, $z \in E$ such that $z = \frac{x+y}{2}$. (*Hint*: by 134Ya/263A, $\mu(x+E)+\mu(2E) > \mu(2I)$.) (ii) Show that if $f : [0,1] \to \mathbb{R}$ is a mid-convex Lebesgue measurable function (definition: 233Ya), $a > 0$, and $E = \{x : x \in [0,1],\ a \le f(x) < 2a\}$ is not negligible, then there is a non-trivial interval $I \subseteq [0,1]$ such that $f(x) > 0$ for every $x \in I$. (*Hint*: 223B.) (iii) Suppose that $f : [0,1] \to \mathbb{R}$ is a mid-convex function such that $f \le 0$ almost everywhere in $[0,1]$. Show that $f \le 0$ everywhere in $]0,1[$. (*Hint*: for every $x \in \,]0,1[$, $\max(f(x-t), f(x+t)) \le 0$ for almost every $t \in [0, \min(x, 1-x)]$.) (iv) Suppose that $f : [0,1] \to \mathbb{R}$ is a mid-convex Lebesgue measurable function such that $f(0) = f(1) = 0$. Show that $f(x) \le 0$ for every $x \in [0,1]$. (*Hint*: show that $\{x : f(x) \le 0\}$ is dense in $[0,1]$, use (ii) to show that it is conegligible in $[0,1]$ and apply (iii).) (v) Show that if $I \subseteq \mathbb{R}$ is an interval and $f : I \to \mathbb{R}$ is a mid-convex Lebesgue measurable function then it is convex.

(e) Let (X, Σ, μ) be a probability space, T a σ-subalgebra of subsets of X, and $f : X \to [0, \infty]$ a Σ-measurable function. Show that (i) there is a T-measurable $g : X \to [0, \infty]$ such that $\int_F g = \int_F f$ for every $F \in$ T (ii) any two such functions are equal a.e.

(f) Suppose that $r \ge 1$ and $C \subseteq \mathbb{R}^r \setminus \{0\}$ is a convex set. Show that there is a non-zero $b \in \mathbb{R}^r$ such that $b \centerdot z \ge 0$ for every $z \in C$. (*Hint*: if $r = 2$, identify $\mathbb{R}^2$ with $\mathbb{C}$; reduce to the case in which C contains no points which are real and negative; set $\theta = \sup\{\arg z : z \in C\}$ and $b = -ie^{i\theta}$. Now induce on r.)

(g) Suppose that $r \ge 1$, $C \subseteq \mathbb{R}^r$ is a convex set and $\phi : C \to \mathbb{R}$ is a convex function. Show that there is a function $h : \mathbb{R}^r \to [-\infty, \infty[$ such that $\phi(z) = \sup\{h(y)+z\centerdot y : y \in \mathbb{R}^r\}$ for every $z \in C$. (*Hint*: try $h(y) = \inf\{\phi(z)-z\centerdot y : z \in C\}$, and apply 233Yf to a translate of $\{(z,t) : \phi(z) \le t\}$.)

(h) Let (X, Σ, μ) be a probability space, $r \ge 1$ an integer and $C \subseteq \mathbb{R}^r$ a convex set. Let $f_1, \dots, f_r$ be μ-integrable real-valued functions and suppose that $\{x : x \in \bigcap_{j\le r}\operatorname{dom} f_j,\ (f_1(x), \dots, f_r(x)) \in C\}$ is a conegligible subset of X. Show that $(\int f_1, \dots, \int f_r) \in C$. (*Hint*: induce on r.)

(i) Let (X,Σ,μ) be a probability space, $r\ge 1$ an integer, $C\subseteq\mathbb{R}^r$ a convex set and $\phi:C\to\mathbb{R}^r$ a convex function. Let $f_1,\ldots,f_r$ be μ-integrable real-valued functions and suppose that $\{x:x\in\bigcap_{j\le r}\operatorname{dom}f_j,\ (f_1(x),\ldots,f_r(x))\in C\}$ is a conegligible subset of X. Show that $\phi(\int f_1,\ldots,\int f_r)\le\int\phi(f_1,\ldots,f_r)$.

(j) Let (X,Σ,μ) be a measure space with $\mu X>0$, $r\ge 1$ an integer, $C\subseteq\mathbb{R}^r$ a convex set such that $tz\in C$ whenever $z\in C$ and $t>0$, and $\phi:C\to\mathbb{R}$ a convex function. Let $f_1,\ldots,f_r$ be μ-integrable real-valued functions and suppose that $\{x:x\in\bigcap_{j\le r}\operatorname{dom}f_j,\ (f_1(x),\ldots,f_r(x))\in C\}$ is a conegligible subset of X. Show that $(\int f_1,\ldots,\int f_r)\in C$ and that $\phi(\int f_1,\ldots,\int f_r)\le\int\phi(f_1,\ldots,f_r)$. (*Hint*: putting 215B(viii) and 235K below together, show that there are a probability measure ν on X and a function $h:X\to[0,\infty[$ such that $\int f_j d\mu=\int f_j\times h\,d\nu$ for every j.)

233 Notes and comments The concept of 'conditional expectation' is fundamental in probability theory, and will reappear in Chapter 27 in its natural context. I hope that even as an exercise in technique, however, it will strike you as worth taking note of.

I introduced 233E as a 'list of elementary facts', and they are indeed straightforward. But below the surface there are some remarkable ideas waiting for expression. If you take T to be the trivial algebra $\{\emptyset,X\}$, so that the (unique) conditional expectation of an integrable function f is the constant function $(\int f)\chi X$, then 233Ed and 233Ee become versions of B.Levi's theorem and Lebesgue's Dominated Convergence Theorem. (Fatou's Lemma is in 233Xa.) Even 233Eg can be thought of as a generalization of the result that $\int cf=c\int f$, where the constant c has been replaced by a bounded T-measurable function. A recurrent theme in the later parts of this treatise will be the search for 'conditioned' versions of theorems. The proof of 233Ee is a typical example of an argument which has been translated from a proof of the original 'unconditioned' result.

I suggested that 233I-233J are surprising, and I think that most of us find them so, even applied to the list of convex functions given in 233G. But I should remark that in a way 233J has very little to do with conditional expectations. The only properties of conditional expectations used in the proof are (i) that if g is a conditional expectation of f, then $a\chi X+bg$ is a conditional expectation of $a\chi X+bf$ for all real a, b (ii) if g_1, g_2 are conditional expectations of f_1, f_2 and $f_1\le_{\text{a.e.}}f_2$, then $g_1\le_{\text{a.e.}}g_2$. See 244Xm below. Jensen's inequality has an interesting extension to the multidimensional case, explored in 233Yf-233Yj. If you have encountered 'geometric' forms of the Hahn-Banach theorem (see 3A5C in Volume 3) you will find 233Yf and 233Yg very natural, and you may notice that the finite-dimensional case is slightly different from the infinite-dimensional case you have probably been taught. I think that in fact the most delicate step is in 233Yh.

Note that 233Ib can be regarded as the special case of 233J in which $\mathrm{T}=\{\emptyset,X\}$. In fact 233Ia can be derived from 233Ib applied to the measure ν where $\nu E=\int_E g$ for every $E\in\Sigma$.

Like 233B, 233K seems to have rather a lot of technical detail in the argument. The point of this result is that we can deduce the integrability of $f\times h$ from that of $g_0\times h$ (but not from the integrability of $g\times h$; see 233Xg). Otherwise it should be routine.

234 Operations on measures

I take a few pages to describe some standard constructions. The ideas are straightforward, but a number of details need to be worked out if they are to be securely integrated into the general framework I employ. The first step is to formally introduce inverse-measure-preserving functions (234A-234B), the most important class of transformations between measure spaces. For construction of new measures, we have the notions of image measure (234C-234E), sum of measures (234G-234H) and indefinite-integral measure (234I-234O). Finally I mention a way of ordering the measures on a given set (234P-234Q).

234A Inverse-measure-preserving functions It is high time that I introduced the nearest thing in measure theory to a 'morphism'. If (X,Σ,μ) and (Y,T,ν) are measure spaces, a function $\phi:X\to Y$ is **inverse-measure-preserving** if $\phi^{-1}[F]\in\Sigma$ and $\mu(\phi^{-1}[F])=\nu F$ for every $F\in\mathrm{T}$.

234B Proposition Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and $\phi:X\to Y$ an inverse-measure-preserving function.

(a) If $\hat\mu$, $\hat\nu$ are the completions of μ, ν respectively, ϕ is also inverse-measure-preserving for $\hat\mu$ and $\hat\nu$.

(b) μ is a probability measure iff ν is a probability measure.

(c) μ is totally finite iff ν is totally finite.

(d)(i) If ν is σ-finite, then μ is σ-finite.

(ii) If ν is semi-finite and μ is σ-finite, then ν is σ-finite.

(e)(i) If ν is σ-finite and atomless, then μ is atomless.

(ii) If ν is semi-finite and μ is purely atomic, then ν is purely atomic.

(f)(i) $\mu^*\phi^{-1}[B] \le \nu^*B$ for every $B \subseteq Y$.

(ii) $\mu^*A \le \nu^*\phi[A]$ for every $A \subseteq X$.

(g) If (Z, Λ, λ) is another measure space, and $\psi : Y \to Z$ is inverse-measure-preserving, then $\psi\phi : X \to Z$ is inverse-measure-preserving.

proof (a) If $\hat\nu$ measures F, there are F', $F'' \in \mathrm{T}$ such that $F' \subseteq F \subseteq F''$ and $\nu(F'' \setminus F') = 0$. Now

$$\phi^{-1}[F'] \subseteq \phi^{-1}[F] \subseteq \phi^{-1}[F''], \quad \mu(\phi^{-1}[F''] \setminus \phi^{-1}[F']) = \nu(F'' \setminus F') = 0,$$

so $\hat\mu$ measures $\phi^{-1}[F]$ and

$$\hat\mu(\phi^{-1}[F]) = \mu\phi^{-1}[F'] = \nu F' = \hat\nu F.$$

As F is arbitrary, ϕ is inverse-measure-preserving for $\hat\mu$ and $\hat\nu$.

(b)-(c) are surely obvious.

(d)(i) If $\langle F_n\rangle_{n\in\mathbb{N}}$ is a cover of Y by sets of finite measure for ν, then $\langle\phi^{-1}[F_n]\rangle_{n\in\mathbb{N}}$ is a cover of X by sets of finite measure for μ.

(ii) Let $\mathcal{F} \subseteq \mathrm{T}$ be a disjoint family of non-ν-negligible sets. Then $\langle\phi^{-1}[F]\rangle_{F\in\mathcal{F}}$ is a disjoint family of non-μ-negligible sets. By 215B(iii), $\mathcal{F}$ is countable. By 215B(iii) in the opposite direction, ν is σ-finite.

(e)(i) Suppose that $E \in \Sigma$ and $\mu E > 0$. Let $\langle F_n\rangle_{n\in\mathbb{N}}$ be a cover of Y by sets of finite measure for ν. Because ν is atomless, we can find, for each n, a finite partition $\langle F_{ni}\rangle_{i\in I_n}$ of F_n such that $\nu F_{ni} < \mu E$ for every $i \in I_n$ (use 215D repeatedly). Now $X = \bigcup_{n\in\mathbb{N}, i\in I_n} \phi^{-1}[F_{ni}]$, so there are $n \in \mathbb{N}$ and $i \in I_n$ with

$$0 < \mu(E \cap \phi^{-1}[F_{ni}) \le \mu\phi^{-1}[F_{ni}] = \nu F_{ni} < \mu E,$$

and E is not a μ-atom. As E is arbitrary, μ is atomless.

(ii) Suppose that $F \in \mathrm{T}$ and $\nu F > 0$. Because ν is semi-finite, there is an $F_1 \subseteq F$ such that $0 < \nu F_1 < \infty$. Now $\mu\phi^{-1}[F_1] > 0$; because μ is purely atomic, there is a μ-atom $E \subseteq \phi^{-1}[F_1]$.

Let $\mathcal{G}$ be the set of those $G \in \mathrm{T}$ such that $G \subseteq F_1$ and $\mu(E \cap \phi^{-1}[G]) = 0$. Then the union of any sequence in $\mathcal{G}$ belongs to $\mathcal{G}$, so by 215Ac there is an $H \in \mathcal{G}$ such that $\nu(G \setminus H) = 0$ whenever $G \in \mathcal{G}$. Consider $F_1 \setminus H$. We have

$$\nu(F_1 \setminus H) = \mu(\phi^{-1}[F_1] \setminus \phi^{-1}[H]) \ge \mu(E \setminus \phi^{-1}[H]) = \mu E > 0.$$

If $G \in \mathrm{T}$ and $G \subseteq F_1 \setminus H$, then one of $E \cap \phi^{-1}[G]$, $E \setminus \phi^{-1}[G]$ is μ-negligible. In the former case, $G \in \mathcal{G}$ and $G = G \setminus H$ is ν-negligible. In the latter case, $F_1 \setminus G \in \mathcal{G}$ and $(F_1 \setminus H) \setminus G$ is ν-negligible. As G is arbitrary, $F_1 \setminus H$ is a ν-atom included in F; as F is arbitrary, ν is purely atomic.

(f)(i) Let $F \in \mathrm{T}$ be such that $B \subseteq F$ and $\nu^*B = \nu F$ (132Aa); then $\phi^{-1}[B] \subseteq \phi^{-1}[F]$ so

$$\mu^*\phi^{-1}[B] \le \mu\phi^{-1}[F] = \nu F = \nu^*B.$$

(ii) $\mu^*A \le \mu^*(\phi^{-1}[\phi[A]]) \le \nu^*\phi[A]$ by (i).

(g) For any $W \in \Lambda$,

$$\mu(\psi\phi)^{-1}[W] = \mu\phi^{-1}[\psi^{-1}[W]] = \nu\psi^{-1}[W] = \lambda W.$$

234C Image measures The following construction is one of the commonest ways in which new measure spaces appear.

Proposition Let (X, Σ, μ) be a measure space, Y any set, and $\phi : X \to Y$ a function. Set

$$\mathrm{T} = \{F : F \subseteq Y, \phi^{-1}[F] \in \Sigma\}, \quad \nu F = \mu(\phi^{-1}[F]) \text{ for every } F \in \mathrm{T}.$$

Then (Y, T, ν) is a measure space.

proof (a) $\emptyset = \phi^{-1}[\emptyset] \in \Sigma$ so $\emptyset \in \mathrm{T}$.

(b) If $F \in \mathrm{T}$, then $\phi^{-1}[F] \in \Sigma$, so $X \setminus \phi^{-1}[F] \in \Sigma$; but $X \setminus \phi^{-1}[F] = \phi^{-1}[Y \setminus F]$, so $Y \setminus F \in \mathrm{T}$.

(c) If $\langle F_n\rangle_{n\in\mathbb{N}}$ is a sequence in T, then $\phi^{-1}[F_n] \in \Sigma$ for every n, so $\bigcup_{n\in\mathbb{N}} \phi^{-1}[F_n] \in \Sigma$; but $\phi^{-1}[\bigcup_{n\in\mathbb{N}} F_n] = \bigcup_{n\in\mathbb{N}} \phi^{-1}[F_n]$, so $\bigcup_{n\in\mathbb{N}} F_n \in \mathrm{T}$.

Thus T is a σ-algebra.

(d) $\nu\emptyset = \mu\phi^{-1}[\emptyset] = \mu\emptyset = 0$.

(e) If $\langle F_n\rangle_{n\in\mathbb{N}}$ is a disjoint sequence in T, then $\langle\phi^{-1}[F_n]\rangle_{n\in\mathbb{N}}$ is a disjoint sequence in Σ, so

$$\nu(\bigcup_{n\in\mathbb{N}} F_n) = \mu\phi^{-1}[\bigcup_{n\in\mathbb{N}} F_n] = \mu(\bigcup_{n\in\mathbb{N}} \phi^{-1}[F_n]) = \sum_{n=0}^{\infty} \mu\phi^{-1}[F_n] = \sum_{n=0}^{\infty} \nu F_n.$$

So ν is a measure.

234D Definition In the context of 234C, ν is called the **image measure** or **push-forward measure**; I will denote it $\mu\phi^{-1}$.

Remark I ought perhaps to say that this construction does not always produce exactly the 'right' measure on Y; there are circumstances in which some modification of the measure $\mu\phi^{-1}$ described here is more useful. But I will note these explicitly when they occur; when I use the unadorned phrase 'image measure' I shall mean the measure constructed above.

234E Proposition Let (X,Σ,μ) be a measure space, Y a set and $\phi : X \to Y$ a function; let $\mu\phi^{-1}$ be the image measure on Y.

(a) ϕ is inverse-measure-preserving for μ and $\mu\phi^{-1}$.

(b) If μ is complete, so is $\mu\phi^{-1}$.

(c) If Z is another set, and $\psi : Y \to Z$ a function, then the image measures $\mu(\psi\phi)^{-1}$ and $(\mu\phi^{-1})\psi^{-1}$ on Z are the same.

proof (a) Immediate from the definitions.

(b) Write ν for $\mu\phi^{-1}$ and T for its domain. If $\nu^*B = 0$, then $\mu^*\phi^{-1}[B] = 0$, by 234B(f-i); as μ is complete, $\phi^{-1}[B] \in \Sigma$, so $B \in \mathrm{T}$. As B is arbitrary, ν is complete.

(c) For $G \subseteq Z$ and $u \in [0,\infty]$,

$$\begin{aligned}
&(\mu(\psi\phi)^{-1})(G) \text{ is defined and equal to } u\\
&\qquad\iff \mu((\psi\phi)^{-1}[G]) \text{ is defined and equal to } u\\
&\qquad\iff \mu(\phi^{-1}[\psi^{-1}[G]]) \text{ is defined and equal to } u\\
&\qquad\iff (\mu\phi^{-1})(\psi^{-1}[G])) \text{ is defined and equal to } u\\
&\qquad\iff ((\mu\phi^{-1})\psi^{-1})(G) \text{ is defined and equal to } u.
\end{aligned}$$

***234F** In the opposite direction, the following construction of a pull-back measure is sometimes useful.

Proposition Let X be a set, (Y,T,ν) a measure space, and $\phi : X \to Y$ a function such that $\phi[X]$ has full outer measure in Y. Then there is a measure μ on X, with domain $\Sigma = \{\phi^{-1}[F] : F \in \mathrm{T}\}$, such that ϕ is inverse-measure-preserving for μ and ν.

proof The check that Σ is a σ-algebra of subsets of X is straightforward; all we need to know is that $\phi^{-1}[\emptyset] = \emptyset$, $X \setminus \phi^{-1}[F] = \phi^{-1}[Y \setminus F]$ for every $F \subseteq Y$, and that $\phi^{-1}[\bigcup_{n\in\mathbb{N}} F_n] = \bigcup_{n\in\mathbb{N}} \phi^{-1}[F_n]$ for every sequence $\langle F_n\rangle_{n\in\mathbb{N}}$ of subsets of Y. The key fact is that if F_1, $F_2 \in \mathrm{T}$ and $\phi^{-1}[F_1] = \phi^{-1}[F_2]$, then $\phi[X]$ does not meet $F_1\triangle F_2$; because $\phi[X]$ has full outer measure, $F_1\triangle F_2$ is ν-negligible and $\nu F_1 = \nu F_2$. Accordingly the formula $\mu\phi^{-1}[F] = \nu F$ does define a function $\mu : \Sigma \to [0,\infty]$. Now

$$\mu\emptyset = \mu\phi^{-1}[\emptyset] = \nu\emptyset = 0.$$

Next, if $\langle E_n\rangle_{n\in\mathbb{N}}$ is a disjoint sequence in Σ, choose $F_n \in \mathrm{T}$ such that $E_n = \phi^{-1}[F_n]$ for each $n \in \mathbb{N}$. The sequence $\langle F_n\rangle_{n\in\mathbb{N}}$ need not be disjoint, but if we set $F'_n = F_n \setminus \bigcup_{i<n} F_i$ for each $n \in \mathbb{N}$, then $\langle F'_n\rangle_{n\in\mathbb{N}}$ is disjoint and

$$E_n = E_n \setminus \bigcup_{i<n} E_i = \phi^{-1}[F'_n]$$

for each n; so

$$\mu(\bigcup_{n\in\mathbb{N}} E_n) = \nu(\bigcup_{n\in\mathbb{N}} F'_n) = \sum_{n=0}^{\infty} \nu F'_n = \sum_{n=0}^{\infty} \mu E_n.$$

As $\langle E_n\rangle_{n\in\mathbb{N}}$ is arbitrary, μ is a measure on X, as required.

234G Sums of measures I come now to a quite different way of building measures. The idea is an obvious one, but the technical details, in the general case I wish to examine, need watching.

Proposition Let X be a set, and $\langle\mu_i\rangle_{i\in I}$ a family of measures on X. For each $i\in I$, let Σ_i be the domain of μ_i. Set $\Sigma=\mathcal{P}X\cap\bigcap_{i\in I}\Sigma_i$ and define $\mu:\Sigma\to[0,\infty]$ by setting $\mu E=\sum_{i\in I}\mu_i E$ for every $E\in\Sigma$. Then μ is a measure on X.

proof Σ is a σ-algebra of subsets of X because every Σ_i is. (Apply 111Ga with $\mathfrak{S}=\{\Sigma_i:i\in I\}\cup\{\mathcal{P}X\}$.) Of course μ takes values in $[0,\infty]$ (226A). $\mu\emptyset=0$ because $\mu_i\emptyset=0$ for every i. If $\langle E_n\rangle_{n\in\mathbb{N}}$ is a disjoint sequence in Σ with union E, then

$$\mu E=\sum_{i\in I}\mu_i E=\sum_{i\in I}\sum_{n=0}^{\infty}\mu_i E_n=\sum_{n=0}^{\infty}\sum_{i\in I}\mu_i E_n$$

(226Af)

$$=\sum_{n=0}^{\infty}\mu E_n.$$

So μ is a measure.

Remark In this context, I will call μ the **sum** of the family $\langle\mu_i\rangle_{i\in I}$.

234H Proposition Let X be a set and $\langle\mu_i\rangle_{i\in I}$ a family of complete measures on X with sum μ.

(a) μ is complete.

(b)(i) A subset of X is μ-negligible iff it is μ_i-negligible for every $i\in I$.

(ii) A subset of X is μ-conegligible iff it is μ_i-conegligible for every $i\in I$.

(c) Let f be a function from a subset of X to $[-\infty,\infty]$. Then $\int f d\mu$ is defined in $[-\infty,\infty]$ iff $\int f d\mu_i$ is defined in $[-\infty,\infty]$ for every i and one of $\sum_{i\in I}f^+d\mu_i$, $\sum_{i\in I}f^-d\mu_i$ is finite, and in this case $\int f d\mu=\sum_{i\in I}\int f d\mu_i$.

proof Write $\Sigma_i=\operatorname{dom}\mu_i$ for $i\in I$, $\Sigma=\mathcal{P}X\cap\bigcap_{i\in I}\Sigma_i=\operatorname{dom}\mu$.

(a) If $E\subseteq F\in\Sigma$ and $\mu F=0$, then $\mu_i F=0$ for every $i\in I$; because μ_i is complete, $E_i\in\Sigma_i$ for every $i\in I$, and $E\in\Sigma$.

(b) This now follows at once, since a set $A\subseteq X$ is μ-negligible iff $\mu A=0$.

(c)(i) Note first that (b-ii) tells us that, under either hypothesis, $\operatorname{dom}f$ is conegligible both for μ and for every μ_i, so that if we extend f to X by giving it the value 0 on $X\setminus\operatorname{dom}f$ then neither $\int f d\mu$ nor $\sum_{i\in I}\int f d\mu_i$ is affected. So let us assume from now on that f is defined everywhere on X. Now it is plain that either hypothesis ensures that f is Σ-measurable, that is, is Σ_i-measurable for every $i\in I$.

(ii) Suppose that f is non-negative. For $n\in\mathbb{N}$ set $f_n(x)=\sum_{k=1}^{4^n}2^{-n}\chi\{x:f(x)\ge 2^{-n}k\}$, so that $\langle f_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence with supremum f. We have

$$\int f_n d\mu=\sum_{k=1}^{4^n}2^{-n}\mu\{x:f(x)\ge 2^{-n}k\}=\sum_{k=1}^{4^n}\sum_{i\in I}2^{-n}\mu_i\{x:f(x)\ge 2^{-n}k\}$$
$$=\sum_{i\in I}\sum_{k=1}^{4^n}2^{-n}\mu_i\{x:f(x)\ge 2^{-n}k\}=\sum_{i\in I}\int f_n d\mu_i$$

for every n, so

$$\int f d\mu=\sup_{n\in\mathbb{N}}\int f_n d\mu=\sup_{n\in\mathbb{N}}\sup_{J\subseteq I\text{ is finite}}\sum_{i\in J}\int f_n d\mu_i=\sup_{J\subseteq I\text{ is finite}}\sup_{n\in\mathbb{N}}\sum_{i\in J}\int f_n d\mu_i$$
$$=\sup_{J\subseteq I\text{ is finite}}\lim_{n\to\infty}\sum_{i\in J}\int f_n d\mu_i=\sup_{J\subseteq I\text{ is finite}}\sum_{i\in J}\lim_{n\to\infty}\int f_n d\mu_i=\sum_{i\in I}\int f d\mu_i.$$

(iii) Generally,

$$\begin{aligned}
&\int f d\mu \text{ is defined in } [\infty,\infty] \\
&\qquad \iff \int f^+ d\mu \text{ and } \int f^- d\mu \text{ are defined and at most one is infinite} \\
&\qquad \iff \sum_{i\in I}\int f^+ d\mu_i \text{ and } \sum_{i\in I}\int f^- d\mu_i \text{ are defined and at most one is infinite} \\
&\qquad \iff \int f d\mu_i \text{ is defined for every } i \text{ and at most one of } \sum_{i\in I}\int f^+ d\mu_i, \\
&\qquad\qquad\qquad \sum_{i\in I}\int f^- d\mu_i \text{ is infinite,}
\end{aligned}$$

and in this case

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu = \sum_{i\in I}\int f^+ d\mu_i - \sum_{i\in I}\int f^- d\mu_i = \sum_{i\in I}\int f d\mu_i.$$

234I Indefinite-integral measures Extending an idea already used in 232D, we are led to the following construction; once again, we need to take care over the formal details if we want to get full value from it.

Theorem Let (X,Σ,μ) be a measure space, and f a non-negative μ-virtually measurable real-valued function defined on a conegligible subset of X. Write $\nu F = \int f \times \chi F\,d\mu$ whenever $F \subseteq X$ is such that the integral is defined in $[0,\infty]$ according to the conventions of 133A. Then ν is a complete measure on X, and its domain includes Σ.

proof (a) Write T for the domain of ν, that is, the family of sets $F \subseteq X$ such that $\int f \times \chi F\,d\mu$ is defined in $[0,\infty]$, that is, $f \times \chi F$ is μ-virtually measurable (133A). Then T is a σ-algebra of subsets of X. **P** For each $F \in \mathrm{T}$ let $H_F \subseteq X$ be a μ-conegligible set such that $f \times \chi F\restriction H_F$ is Σ-measurable. Because f itself is μ-virtually measurable, $X \in \mathrm{T}$. If $F \in \mathrm{T}$, then

$$f \times \chi(X\setminus F)\restriction(H_X \cap H_F) = f\restriction(H_X\cap H_F) - (f\times\chi F)\restriction(H_X\cap H_F)$$

is Σ-measurable, while $H_X \cap H_F$ is μ-conegligible, so $X \setminus F \in \mathrm{T}$. If $\langle F_n\rangle_{n\in\mathbb{N}}$ is a sequence in T with union F, set $H = \bigcap_{n\in\mathbb{N}} H_{F_n}$; then H is conegligible, $f \times \chi F_n\restriction H$ is Σ-measurable for every $n \in \mathbb{N}$, and $f \times \chi F = \sup_{n\in\mathbb{N}} f \times \chi F_n$, so $f \times \chi F\restriction H$ is Σ-measurable, and $F \in \mathrm{T}$. Thus T is a σ-algebra. If $F \in \Sigma$, then $f \times \chi F\restriction H_X$ is Σ-measurable, so $F \in \mathrm{T}$. **Q**

(b) Next, ν is a measure. **P** Of course $\nu F \in [0,\infty]$ for every $F \in \mathrm{T}$. $f \times \chi\emptyset = 0$ wherever it is defined, so $\nu\emptyset = 0$. If $\langle F_n\rangle_{n\in\mathbb{N}}$ is a disjoint sequence in T with union F, then $f \times \chi F = \sum_{n=0}^\infty f \times \chi F_n$. If $\nu F_m = \infty$ for some m, then we surely have $\nu F = \infty = \sum_{n=0}^\infty \nu F_n$. If $\nu F_m < \infty$ for each m but $\sum_{n=0}^\infty \nu F_n = \infty$, then

$$\int f \times \chi(\bigcup_{n\le m} F_n) = \sum_{n=0}^m \int f \times \chi F_n \to \infty$$

as $m \to \infty$, so again $\nu F = \infty = \sum_{n=0}^\infty \nu F_n$. If $\sum_{n=0}^\infty \nu F_n < \infty$ then by B.Levi's theorem

$$\nu F = \int \sum_{n=0}^\infty f \times \chi F_n = \sum_{n=0}^\infty \int f \times \chi F_n = \sum_{n=0}^\infty \nu F_n. \textbf{ Q}$$

(c) Finally, ν is complete. **P** If $A \subseteq F \in \mathrm{T}$ and $\nu F = 0$, then $f \times \chi F = 0$ a.e., so $f \times \chi A = 0$ a.e. and νA is defined and equal to zero. **Q**

234J Definition Let (X,Σ,μ) be a measure space, and ν another measure on X with domain T. I will call ν an **indefinite-integral measure** over μ, or sometimes a **completed indefinite-integral measure**, if it can be obtained by the method of 234I from some non-negative virtually measurable function f defined almost everywhere on X. In this case, f is a Radon-Nikodým derivative of ν with respect to μ in the sense of 232Hf. As in 232Hf, the phrase **density function** is also used in this context.

234K Remarks Let (X,Σ,μ) be a measure space, and f a μ-virtually measurable non-negative real-valued function defined almost everywhere on X; let ν be the associated indefinite-integral measure.

(a) There is a Σ-measurable function $g : X \to [0,\infty[$ such that $f = g$ μ-a.e. **P** Let $H \subseteq \operatorname{dom} f$ be a measurable conegligible set such that $f\restriction H$ is measurable, and set $g(x) = f(x)$ for $x \in H$, $g(x) = 0$ for $x \in X \setminus H$. **Q** In this case, $\int f \times \chi E\,d\mu = \int g \times \chi E\,d\mu$ if either is defined. So g is a Radon-Nikodým derivative of ν, and ν has a Radon-Nikodým derivative which is Σ-measurable and defined everywhere.

(b) If E is μ-negligible, then $f \times \chi E = 0$ μ-a.e., so $\nu E = 0$. Many authors are prepared to say 'ν is absolutely continuous with respect to μ' in this context. But if ν is not totally finite, it need not be absolutely continuous in the ϵ-δ sense of 232Aa (234Xh), and further difficulties can arise if μ or ν is not σ-finite (see 234Yk, 234Ym).

(c) I have defined 'indefinite-integral measure' in such a way as to produce a complete measure. In my view this is what makes best sense in most applications. There are occasions on which it seems more appropriate to use the measure $\nu_0 : \Sigma \to [0, \infty]$ defined by setting $\nu_0 E = \int_E f d\mu = \int f \times \chi E \, d\mu$ for $E \in \Sigma$. I suppose I would call this the **uncompleted indefinite-integral measure** over μ defined by f. (ν is always the completion of ν_0; see 234Lb.)

(d) Note the way in which I formulated the definition of ν: '$\nu E = \int f \times \chi E \, d\mu$ if the integral is defined', rather than '$\nu E = \int_E f d\mu$'. The point is that the longer formula gives a rule for deciding what the domain of ν must be. Of course it is the case that $\nu E = \int_E f d\mu$ for every $E \in \operatorname{dom} \nu$ (apply 214F to $f \times \chi E$).

(e) Because μ and its completion define the same virtually measurable functions, the same null ideals and the same integrals (212Eb, 212F), they define the same indefinite-integral measures.

234L The domain of an indefinite-integral measure It is sometimes useful to have an explicit description of the domain of a measure constructed in this way.

Proposition Let (X, Σ, μ) be a measure space, f a non-negative μ-virtually measurable function defined almost everywhere in X, and ν the associated indefinite-integral measure. Set $G = \{x : x \in \operatorname{dom} f, \ f(x) > 0\}$, and let $\hat{\Sigma}$ be the domain of the completion $\hat{\mu}$ of μ.

(a) The domain T of ν is $\{E : E \subseteq X, \ E \cap G \in \hat{\Sigma}\}$; in particular, $\mathrm{T} \supseteq \hat{\Sigma} \supseteq \Sigma$.

(b) ν is the completion of its restriction to Σ.

(c) A set $A \subseteq X$ is ν-negligible iff $A \cap G$ is μ-negligible.

(d) In particular, if μ itself is complete, then $\mathrm{T} = \{E : E \subseteq X, \ E \cap G \in \Sigma\}$ and $\nu A = 0$ iff $\mu(A \cap G) = 0$.

proof (a)(i) If $E \in \mathrm{T}$, then $f \times \chi E$ is virtually measurable, so there is a conegligible measurable set $H \subseteq \operatorname{dom} f$ such that $f \times \chi E \restriction H$ is measurable. Now $E \cap G \cap H = \{x : x \in H, \ (f \times \chi E)(x) > 0\}$ must belong to Σ, while $E \cap G \setminus H$ is negligible, so belongs to $\hat{\Sigma}$, and $E \cap G \in \hat{\Sigma}$.

(ii) If $E \cap G \in \hat{\Sigma}$, let $F_1, F_2 \in \Sigma$ be such that $F_1 \subseteq E \cap G \subseteq F_2$ and $F_2 \setminus F_1$ is negligible. Let $H \subseteq \operatorname{dom} f$ be a conegligible set such that $f \restriction H$ is measurable. Then $H' = H \setminus (F_2 \setminus F_1)$ is conegligible and $f \times \chi E \restriction H' = f \times \chi F_1 \restriction H'$ is measurable, so $f \times \chi E$ is virtually measurable and $E \in \mathrm{T}$.

(b) Thus the given formula does indeed describe T. If $E \in \mathrm{T}$, let $F_1, F_2 \in \Sigma$ be such that $F_1 \subseteq E \cap G \subseteq F_2$ and $\mu(F_2 \setminus F_1) = 0$. Because G itself also belongs to $\hat{\Sigma}$, there are $G_1, G_2 \in \Sigma$ such that $G_1 \subseteq G \subseteq G_2$ and $\mu(G_2 \setminus G_1) = 0$. Set $F_2' = F_2 \cup (X \setminus G_1)$. Then $F_2' \in \Sigma$ and $F_1 \subseteq E \subseteq F_2'$. But $(F_2' \setminus F_1) \cap G \subseteq (G_2 \setminus G_1) \cup (F_2 \setminus F_1)$ is μ-negligible, so $\nu(F_2' \setminus F_1) = 0$.

This shows that if ν' is the completion of $\nu \restriction \Sigma$ and T' is its domain, then $\mathrm{T} \subseteq \mathrm{T}'$. But as ν is complete, it surely extends ν', so $\nu = \nu'$, as claimed.

(c) Now take any $A \subseteq X$. Because ν is complete,

$$\begin{aligned} A \text{ is } \nu\text{-negligible} &\iff \nu A = 0 \\ &\iff \int f \times \chi A \, d\mu = 0 \\ &\iff f \times \chi A = 0 \ \mu\text{-a.e.} \\ &\iff A \cap G \text{ is } \mu\text{-negligible.} \end{aligned}$$

(d) This is just a restatement of (a) and (c) when $\mu = \hat{\mu}$.

234M Corollary If (X, Σ, μ) is a complete measure space and $G \in \Sigma$, then the indefinite-integral measure over μ defined by χG is just the measure $\mu \llcorner G$ defined by setting

$$(\mu \llcorner G)(F) = \mu(F \cap G) \text{ whenever } F \subseteq X \text{ and } F \cap G \in \Sigma.$$

proof 234Ld.

***234N** The next two results will not be relied on in this volume, but I include them for future reference, and to give an idea of the scope of these ideas.

Proposition Let (X, Σ, μ) be a measure space, and ν an indefinite-integral measure over μ.

(a) If μ is semi-finite, so is ν.

(b) If μ is complete and locally determined, so is ν.

(c) If μ is localizable, so is ν.

(d) If μ is strictly localizable, so is ν.

(e) If μ is σ-finite, so is ν.

(f) If μ is atomless, so is ν.

proof By 234Ka, we may express ν as the indefinite integral of a Σ-measurable function $f : X \to [0, \infty[$. Let T be the domain of ν, and $\hat{\Sigma}$ the domain of the completion $\hat{\mu}$ of μ; set $G = \{x : x \in X,\, f(x) > 0\} \in \Sigma$.

(a) Suppose that $E \in \mathrm{T}$ and that $\nu E = \infty$. Then $E \cap G$ cannot be μ-negligible. Because μ is semi-finite, there is a non-negligible $F \in \Sigma$ such that $F \subseteq E \cap G$ and $\mu F < \infty$. Now $F = \bigcup_{n \in \mathbb{N}} \{x : x \in F,\, 2^{-n} \le f(x) \le n\}$, so there is an $n \in \mathbb{N}$ such that $F' = \{x : x \in F,\, 2^{-n} \le f(x) \le n\}$ is non-negligible. Because f is measurable, $F' \in \Sigma \subseteq \mathrm{T}$ and $2^{-n}\mu F' \le \nu F' \le n\mu F'$. Thus we have found an $F' \subseteq E$ such that $0 < \nu F' < \infty$. As E is arbitrary, ν is semi-finite.

(b) We already know that ν is complete (234Lb) and semi-finite. Now suppose that $E \subseteq X$ is such that $E \cap F \in \mathrm{T}$, that is, $E \cap F \cap G \in \Sigma$ (234Ld), whenever $F \in \mathrm{T}$ and $\nu F < \infty$. Then $E \cap G \cap F \in \Sigma$ whenever $F \in \Sigma$ and $\mu F < \infty$. **P** Set $F_n = \{x : x \in F \cap G,\, f(x) \le n\}$. Then $\nu F_n \le n\mu F < \infty$, so $E \cap G \cap F_n \in \Sigma$ for every n. But this means that $E \cap G \cap F = \bigcup_{n \in \mathbb{N}} E \cap G \cap F_n \in \Sigma$. **Q** Because μ is locally determined, $E \cap G \in \Sigma$ and $E \in \mathrm{T}$. As E is arbitrary, ν is locally determined.

(c) Let $\mathcal{F} \subseteq \mathrm{T}$ be any set. Set $\mathcal{E} = \{F \cap G : F \in \mathcal{F}\}$, so that $\mathcal{E} \subseteq \hat{\Sigma}$. By 212Ga, $\hat{\mu}$ is localizable, so $\mathcal{E}$ has an essential supremum $H \in \hat{\Sigma}$. But now, for any $H' \in \mathrm{T}$, $H' \cup (X \setminus G) = (H' \cap G) \cup (X \setminus G)$ belongs to $\hat{\Sigma}$, so

$$\begin{aligned}
\nu(F \setminus H') = 0 &\text{ for every } F \in \mathcal{F} \\
&\iff \hat{\mu}(F \cap G \setminus H') = 0 \text{ for every } F \in \mathcal{F} \\
&\iff \hat{\mu}(E \setminus H') = 0 \text{ for every } E \in \mathcal{E} \\
&\iff \hat{\mu}(E \setminus (H' \cup (X \setminus G))) = 0 \text{ for every } E \in \mathcal{E} \\
&\iff \hat{\mu}(H \setminus ((H' \cup (X \setminus G))) = 0 \\
&\iff \hat{\mu}(H \cap G \setminus H') = 0 \\
&\iff \nu(H \setminus H') = 0.
\end{aligned}$$

Thus H is also an essential supremum of $\mathcal{F}$ in T. As $\mathcal{F}$ is arbitrary, ν is localizable.

(d) Let $\langle X_i \rangle_{i \in I}$ be a decomposition of X for μ; then it is also a decomposition for $\hat{\mu}$ (212Gb). Set $F_0 = X \setminus G$, $F_n = \{x : x \in G,\, n-1 < f(x) \le n\}$ for $n \ge 1$. Then $\langle X_i \cap F_n \rangle_{i \in I, n \in \mathbb{N}}$ is a decomposition for ν. **P** (i) $\langle X_i \rangle_{i \in I}$ and $\langle F_n \rangle_{n \in \mathbb{N}}$ are partitions of X into members of $\Sigma \subseteq \mathrm{T}$, so $\langle X_i \cap F_n \rangle_{i \in I, n \in \mathbb{N}}$ also is. (ii) $\nu(X_i \cap F_0) = 0$, $\nu(X_i \cap F_n) \le n\mu X_i < \infty$ for $i \in I$, $n \ge 1$. (iii) If $E \subseteq X$ and $E \cap X_i \cap F_n \in \mathrm{T}$ for every $i \in I$ and $n \in \mathbb{N}$ then $E \cap X_i \cap G = \bigcup_{n \in \mathbb{N}} E \cap X_i \cap F_n \cap G$ belongs to $\hat{\Sigma}$ for every i, so $E \cap G \in \hat{\Sigma}$ and $E \in \mathrm{T}$. (iv) If $E \in \mathrm{T}$, then of course

$$\sum_{i \in I, n \in \mathbb{N}} \nu(E \cap X_i \cap F_n) = \sup_{J \subseteq I \times \mathbb{N} \text{ is finite}} \sum_{(i,n) \in J} \nu(E \cap X_i \cap F_n) \le \nu E.$$

So if $\sum_{i \in I, n \in \mathbb{N}} \nu(E \cap X_i \cap F_n) = \infty$ it is surely equal to νE. If the sum is finite, then $K = \{i : i \in I,\, \nu(E \cap X_i) > 0\}$ must be countable. But for $i \in I \setminus K$, $\int_{E \cap X_i} f d\mu = 0$, so $f = 0$ μ-a.e. on $E \cap X_i$, that is, $\hat{\mu}(E \cap G \cap X_i) = 0$. Because $\langle X_i \rangle_{i \in I}$ is a decomposition for $\hat{\mu}$, $\hat{\mu}(E \cap G \cap \bigcup_{i \in I \setminus K} X_i) = 0$ and $\nu(E \cap \bigcup_{i \in I \setminus K} X_i) = 0$. But this means that

$$\nu E = \sum_{i \in K} \nu(E \cap X_i) = \sum_{i \in K, n \in \mathbb{N}} \nu(E \cap X_i \cap F_n) = \sum_{i \in I, n \in \mathbb{N}} \nu(E \cap X_i \cap F_n).$$

As E is arbitrary, $\langle X_i \cap F_n \rangle_{i \in I, n \in \mathbb{N}}$ is a decomposition for ν. **Q** So ν is strictly localizable.

(e) If μ is σ-finite, then in (d) we can take I to be countable, so that $I \times \mathbb{N}$ also is countable, and ν will be σ-finite.

(f) If μ is atomless, so is $\hat{\mu}$ (212Gd). If $E \in \mathrm{T}$ and $\nu E > 0$, then $\hat{\mu}(E \cap G) > 0$, so there is an $F \in \hat{\Sigma}$ such that $F \subseteq E \cap G$ and neither F nor $E \cap G \setminus F$ is $\hat{\mu}$-negligible. But in this case both $\nu F = \int_F f d\mu$ and $\nu(E \setminus F) = \int_{E \setminus F} f d\mu$ must be greater than 0 (122Rc). As E is arbitrary, ν is atomless.

***234O** For localizable measures, there is a straightforward description of the associated indefinite-integral measures.

Theorem Let (X, Σ, μ) be a localizable measure space. Then a measure ν, with domain $\mathrm{T} \supseteq \Sigma$, is an indefinite-integral measure over μ iff (α) ν is semi-finite and zero on μ-negligible sets (β) ν is the completion of its restriction to Σ (γ) whenever $\nu E > 0$ there is an $F \subseteq E$ such that $F \in \Sigma$, $\mu F < \infty$ and $\nu F > 0$.

proof (a) If ν is an indefinite-integral measure over ν, then by 234Na, 234Kb and 234Lb it is semi-finite, zero on μ-negligible sets and the completion of its restriction to Σ. Now suppose that $E \in \mathrm{T}$ and $\nu E > 0$. Then there is an $E_0 \in \Sigma$ such that $E_0 \subseteq E$ and $\nu E_0 = \nu E$, by 234Lb. If $f : X \to \mathbb{R}$ is a Σ-measurable Radon-Nikodým derivative of ν (234Ka), and $G = \{x : f(x) > 0\}$, then $\mu(G \cap E_0) > 0$; because μ is semi-finite, there is an $F \in \Sigma$ such that $F \subseteq G \cap E_0$ and $0 < \mu F < \infty$, in which case $\nu F > 0$.

(b) So now suppose that ν satisfies the conditions.

(i) Set $\mathcal{E} = \{E : E \in \Sigma, \nu E < \infty\}$. For each $E \in \mathcal{E}$, consider $\nu_E : \Sigma \to \mathbb{R}$, setting $\nu_E G = \nu(G \cap E)$ for every $G \in \Sigma$. Then ν_E is countably additive and truly continuous with respect to μ. **P** ν_E is countably additive, just as in 231De. Because ν is zero on μ-negligible sets, ν_E must be absolutely continuous with respect to μ, by 232Ba. Since ν_E clearly satisfies condition (γ) of 232Bb, it must be truly continuous. **Q**

By 232E, there is a μ-integrable function f_E such that $\nu_E G = \int_G f_E d\mu$ for every $G \in \Sigma$, and we may suppose that f_E is Σ-measurable (232He). Because ν_E is non-negative, $f_E \geq 0$ μ-almost everywhere.

(ii) If E, $F \in \mathcal{E}$ then $f_E = f_F$ μ-a.e. on $E \cap F$, because

$$\int_G f_E d\mu = \nu G = \int_G f_F d\mu$$

whenever $G \in \Sigma$ and $G \subseteq E \cap F$. Because (X, Σ, μ) is localizable, there is a measurable $f : X \to \mathbb{R}$ such that $f_E = f$ μ-a.e. on E for every $E \in \mathcal{E}$ (213N). Because every f_E is non-negative almost everywhere, we may suppose that f is non-negative, since surely $f_E = f \vee \mathbf{0}$ μ-a.e. on E for every $E \in \mathcal{E}$.

(iii) Let ν' be the indefinite-integral measure defined by f. If $E \in \mathcal{E}$ then

$$\nu E = \int_E f_E d\mu = \int_E f d\mu = \nu' E.$$

For $E \in \Sigma \setminus \mathcal{E}$, we have

$$\nu' E \geq \sup\{\nu' F : F \in \mathcal{E}, F \subseteq E\} = \sup\{\nu F : F \in \mathcal{E}, F \subseteq E\} = \nu E = \infty$$

because ν is semi-finite. Thus ν' and ν agree on Σ. But since each is the completion of its restriction to Σ, they must be equal.

234P Ordering measures There are many ways in which one measure can dominate another. Here I will describe one of the simplest.

Definition Let μ, ν be two measures on a set X. I will say that $\mu \leq \nu$ if μE is defined, and $\mu E \leq \nu E$, whenever ν measures E.

234Q Proposition Let X be a set, and write M for the set of all measures on X.

(a) Defining $\leq$ as in 234P, $(\mathrm{M}, \leq)$ is a partially ordered set.

(b) If μ, $\nu \in \mathrm{M}$, then $\mu \leq \nu$ iff there is a $\lambda \in \mathrm{M}$ such that $\mu + \lambda = \nu$.

(c) If $\mu \leq \nu$ in M and f is a $[-\infty, \infty]$-valued function, defined on a subset of X, such that $\int f d\nu$ is defined in $[-\infty, \infty]$, then $\int f d\mu$ is defined; if f is non-negative, $\int f d\mu \leq \int f d\nu$.

proof (a) Of course $\mu \leq \mu$ for every $\mu \in \mathrm{M}$. If $\mu \leq \nu$ and $\nu \leq \lambda$ in M, then $\operatorname{dom} \lambda \subseteq \operatorname{dom} \nu \subseteq \operatorname{dom} \mu$, and $\mu E \leq \nu E \leq \lambda E$ whenever λ measures E. If $\mu \leq \nu$ and $\nu \leq \mu$ then $\operatorname{dom} \mu \subseteq \operatorname{dom} \nu \subseteq \operatorname{dom} \mu$ and $\mu E \leq \nu E \leq \mu E$ for every E in their common domain, so $\mu = \nu$.

(b)(i) If $\mu + \lambda = \nu$, then the definitions in 234G and 234P make it plain that $\mu \leq \nu$.

(ii)$(\boldsymbol{\alpha})$ In the reverse direction, if $\mu \leq \nu$, write T for the domain of ν. Define $\lambda : \mathrm{T} \to [0, \infty]$ by setting

$$\lambda G = \sup\{\nu F - \mu F : F \in \mathrm{T}, F \subseteq G, \mu F < \infty\}$$

for $G \in \mathrm{T}$. Then $\lambda \in \mathrm{M}$. **P** Of course $\operatorname{dom} \lambda = \mathrm{T}$ is a σ-algebra, and $\lambda\emptyset = 0$. Suppose that $\langle G_n \rangle_{n \in \mathbb{N}}$ is a disjoint sequence in T with union G. If $F \in \mathrm{T}$, $F \subseteq G$ and $\mu F < \infty$, then

$$\begin{aligned}\nu F - \mu F &= \sum_{n=0}^{\infty} \nu(F \cap G_n) - \sum_{n=0}^{\infty} \mu(F \cap G_n) \\ &= \sum_{n=0}^{\infty} \nu(F \cap G_n) - \mu(F \cap G_n) \le \sum_{n=0}^{\infty} \lambda G_n;\end{aligned}$$

as F is arbitrary, $\lambda G \le \sum_{n=0}^{\infty} \lambda G_n$. If $\gamma < \sum_{n=0}^{\infty} \lambda G_n$, there are an $m \in \mathbb{N}$ such that $\gamma < \sum_{n=0}^{m} \lambda G_n$, and $F_0, \dots, F_m$ such that $F_n \in \mathrm{T}$, $F_n \subseteq G_n$ and $\mu F_n < \infty$ for every $n \le m$, while $\sum_{n=0}^{m} \nu F_n - \mu F_n \ge \gamma$. Set $F = \bigcup_{n \le m} F_n$; then $F \in \mathrm{T}$, $F \subseteq G$ and $\mu F < \infty$, so

$$\lambda G \ge \nu F - \mu F = \textstyle\sum_{n=0}^{m} \nu F_n - \mu F_n \ge \gamma.$$

As γ is arbitrary, $\lambda G \ge \sum_{n=0}^{\infty} \lambda G_n$; as $\langle G_n \rangle_{n \in \mathbb{N}}$ is arbitrary, λ is countably additive. **Q**

(β) Now $\mu + \lambda = \nu$. **P** The domain of $\mu + \lambda$ is $\operatorname{dom} \mu \cap \operatorname{dom} \lambda = \mathrm{T} = \operatorname{dom} \nu$. Take $G \in \mathrm{T}$. If $\mu G = \infty$, then $\nu G = \infty = (\mu + \lambda)G$. Otherwise,

$$(\mu + \lambda)G \ge \mu G + \nu G - \mu G = \nu G.$$

So if $\nu G = \infty$ we shall certainly have $\nu G = (\mu + \lambda)G$. Finally, if $\nu G < \infty$ then

$$\begin{aligned}(\mu + \lambda)G &= \mu G + \sup\{\nu F - \mu F : F \in \mathrm{T},\ F \subseteq G\} \\ &= \sup\{\nu F + \mu(G \setminus F) : F \in \mathrm{T},\ F \subseteq G\} \\ &\le \sup\{\nu F + \nu(G \setminus F) : F \in \mathrm{T},\ F \subseteq G\} = \nu G,\end{aligned}$$

so again we have equality. **Q**

Thus we have an appropriate expression of ν as a sum of measures.

(c)(i) If f is non-negative, put (b) and 234Hc together.

(ii) In general, if $\int f d\nu$ is defined, so are both $\int f^+ d\nu$ and $\int f^- d\nu$, and at most one is infinite; so $\int f^+ d\mu$ and $\int f^- d\mu$ are defined and at most one is infinite.

234X Basic exercises (a) Let (X, Σ, μ) and (Y, T, ν) be measure spaces, and $\phi : X \to Y$ an inverse-measure-preserving function. Let $A \subseteq X$ be a set of full outer measure in X. Show that $\phi[A]$ has full outer measure in Y, and that $\phi{\restriction}A$ is inverse-measure-preserving for the subspace measures on A and $\phi[A]$.

(b) Let (X, Σ, μ) be a measure space, Y a set and $\phi : X \to Y$ a function. Show that if μ is point-supported, so is the image measure $\mu\phi^{-1}$.

(c) Give an example of a probability space (X, Σ, μ), a set Y, and a function $\phi : X \to Y$ such that the completion of the image measure $\mu\phi^{-1}$ is not the image of the completion of μ. (*Hint*: $\#(X) = 3$.)

(d) Let X, Y be sets, $\phi : X \to Y$ a function and $\langle \mu_i \rangle_{i \in I}$ a family of measures on X with sum μ. Writing $\mu_i\phi^{-1}$, $\mu\phi^{-1}$ for the image measures on Y, show that $\mu\phi^{-1} = \sum_{i \in I} \mu_i\phi^{-1}$.

(e) Let X be a set. (i) Show that if $\langle \mu_i \rangle_{i \in I}$ is a countable family of σ-finite measures on X, and $\mu = \sum_{i \in I} \mu_i$ is semi-finite, then μ is σ-finite. (ii) Show that if $\langle \mu_i \rangle_{i \in I}$ is a family of purely atomic measures on X, and $\mu = \sum_{i \in I} \mu_i$ is semi-finite, then μ is purely atomic. (iii) Show that if $\langle \mu_i \rangle_{i \in I}$ is any family of point-supported measures on X, then $\sum_{i \in I} \mu_i$ is point-supported.

>(f) Let X be a set, and write M for the set of all measures on X. For $\mu \in \mathrm{M}$ and $\alpha \in [0, \infty[$, define $\alpha\mu$ by saying that if $\alpha > 0$ then $(\alpha\mu)(E) = \alpha\mu E$ for $E \in \operatorname{dom} \mu$, while if $\alpha = 0$ then $(\alpha\mu)(E) = 0$ for every $E \subseteq X$. (i) Show that $\alpha\mu \in \mathrm{M}$ for all $\alpha \in [0, \infty[$ and $\mu \in \mathrm{M}$. (ii) Show that $(\alpha + \beta)\mu = \alpha\mu + \beta\mu$, $\alpha(\beta\mu) = (\alpha\beta)\mu$, $\alpha(\mu + \nu) = \alpha\mu + \alpha\nu$ for all α, $\beta \in [0, \infty[$ and μ, $\nu \in \mathrm{M}$.

(g) Let X be a set, and $\langle \mu_i \rangle_{i \in I}$ a family of complete measures on X with sum μ. Show that a $[-\infty, \infty]$-valued function f defined on a subset of X is μ-integrable iff it is μ_i-integrable for every $i \in I$ and $\sum_{i \in I} \int |f| d\mu_i$ is finite.

(h) Let μ be Lebesgue measure on $[0, 1]$, and set $f(x) = \frac{1}{x}$ for $x > 0$. Let ν be the associated indefinite-integral measure. Show that the domain of ν is equal to the domain of μ. Show that for every $\delta \in \left]0, \frac{1}{2}\right]$ there is a measurable set E such that $\mu E = \delta$ but $\nu E = \frac{1}{\delta}$.

(i) Let (X,Σ,μ) be a measure space. (i) Show that if ν_1 and ν_2 are indefinite-integral measures over μ, so is $\nu_1+\nu_2$. (ii) Show that if $\langle\nu_i\rangle_{i\in I}$ is a countable family of indefinite-integral measures over μ, and $\nu=\sum_{i\in I}\nu_i$ is semi-finite, then ν is an indefinite-integral measure over μ.

(j) Let (X,Σ,μ) be a measure space, and ν an indefinite-integral measure over μ. Show that if μ is purely atomic, so is ν.

(k) Let μ be a point-supported measure. Show that any indefinite-integral measure over μ is point-supported.

(l) Let X be a set, and M the set of measures on X, with the partial ordering defined in 234P. Show that (i) M has greatest and least members (to be described); (ii) if $\langle\mu_i\rangle_{i\in I}$ and $\langle\nu_i\rangle_{i\in I}$ are families in M such that $\mu_i\le\nu_i$ for every i, then $\sum_{i\in I}\mu_i\le\sum_{i\in I}\nu_i$; (iii) if we define scalar multiplication as in 234Xf, then $\alpha\mu\le\mu$ whenever $\mu\in\mathrm{M}$ and $\alpha\in[0,1]$; (iv) writing $\hat\mu$ for the completion of μ, $\hat\mu\le\mu$ and $\hat\mu\le\hat\nu$ whenever μ, $\nu\in\mathrm{M}$ and $\mu\le\nu$; (v) writing $\tilde\mu$ for the c.l.d. version of μ, $\tilde\mu\le\mu$ for every $\mu\in\mathrm{M}$; (vi) whenever $A\subseteq\mathrm{M}$ is upwards-directed, it has a least upper bound in M.

(m) Write out an elementary direct proof of 234Qc not depending on 234Qb.

234Y Further exercises (a) Write ν for Lebesgue measure on $Y=[0,1]$, and T for its domain. Let $A\subseteq[0,1]$ be a set such that $\nu^*A=\nu^*([0,1]\setminus A)=1$, and set $X=[0,1]\cup\{x+1:x\in A\}\cup\{x+2:x\in[0,1]\setminus A\}$. Let μ_{LX} be the subspace measure induced on X by Lebesgue measure, and set $\mu E=\frac{1}{3}\mu_{LX}E$ for $E\in\Sigma=\operatorname{dom}\mu_{LX}$. Define $\phi:X\to Y$ by writing $\phi(x)=x$ if $x\in[0,1]$, $\phi(x)=x-1$ if $x\in X\cap\,]1,2]$ and $\phi(x)=x-2$ if $x\in X\cap\,]2,3]$. Show that ν is the image measure $\mu\phi^{-1}$, but that $\nu^*A>\mu^*\phi^{-1}[A]$.

(b) Look for interesting examples of probability spaces (X,Σ,μ) and (Y,T,ν) for which there are functions $\phi:X\to Y$ such that $\phi[E]\in\mathrm{T}$ and $\nu\phi[E]=\mu E$ for every $E\in\Sigma$. (*Hint*: 254K, 343J.)

(c) Let μ be two-dimensional Lebesgue measure on the unit square $[0,1]^2$, and let $\phi:[0,1]^2\to[0,1]$ be the projection onto the first coordinate, so that $\phi(\xi_1,\xi_2)=\xi_1$ for ξ_1, $\xi_2\in[0,1]$. Show that the image measure $\mu\phi^{-1}$ is Lebesgue one-dimensional measure on $[0,1]$.

(d) In 234F, show that the image measure $\mu\phi^{-1}$ extends ν, and is equal to ν if and only if $F\in\mathrm{T}$ for every $F\subseteq Y\setminus\phi[X]$.

(e) Let (Y,T,ν) be a complete measure space, X a set and $\phi:X\to Y$ a surjection. Set

$$\Sigma=\{E:E\subseteq X,\,\phi[E]\in\mathrm{T},\,\nu(\phi[E]\cap\phi[X\setminus E])=0\},\quad \mu E=\nu\phi[E]\text{ for }E\in\mathrm{T}.$$

Show that μ is the completion of the measure constructed by the process of 234F.

(f) Let X be a set, and M the set of measures on X. Show that M, with addition as defined for two measures by the formulae of 234G, is a commutative semigroup with identity; describe the identity.

(g) Give an example of a set X, probability measures μ_1, μ_2 on X and a set $A\subseteq X$ such that A is both μ_1-negligible and μ_2-negligible, but is not μ-negligible, where $\mu=\mu_1+\mu_2$.

(h) In 214O, show that if we set $\nu E=\sup_{I\in\mathcal{I}}\mu^*(E\cap I)$ for every $E\in\Sigma$, then ν is a measure, while $\mu=\nu+\lambda$.

(i) Let (X,Σ,μ) be an atomless semi-finite measure space and ν an indefinite-integral measure over μ. Show that the following are equiveridical: (i) for every $\epsilon>0$ there is a $\delta>0$ such that $\nu E\le\epsilon$ whenever $\mu E\le\delta$ (ii) ν has a Radon-Nikodým derivative expressible as the sum of a bounded function and an integrable function.

(j) Let (X,Σ,μ) be a measure space and ν an indefinite-integral measure over μ, with Radon-Nikodým derivative f. Show that the c.l.d. version of ν is the indefinite-integral measure defined by f over the c.l.d. version of μ.

(k) Let (X,Σ,μ) be a semi-finite measure space which is not localizable. Show that there is a measure $\nu:\Sigma\to[0,\infty]$ such that $\nu E\le\mu E$ for every $E\in\Sigma$ but there is no measurable function f such that $\nu E=\int_E fd\mu$ for every $E\in\Sigma$.

(l) Let (X,Σ,μ) be a localizable measure space with locally determined negligible sets. Show that a measure ν, with domain $\mathrm{T}\supseteq\Sigma$, is an indefinite-integral measure over μ iff (α) ν is complete and semi-finite and zero on μ-negligible sets (β) whenever $\nu E>0$ there is an $F\subseteq E$ such that $F\in\Sigma$ and $\mu F<\infty$ and $\nu F>0$.

(m) Give an example of a localizable measure space (X,Σ,μ) and a complete semi-finite measure ν on X, defined on a σ-algebra $\mathrm{T}\supseteq\Sigma$, zero on μ-negligible sets, and such that whenever $\nu E>0$ there is an $F\subseteq E$ such that $F\in\Sigma$ and $\mu F<\infty$ and $\nu F>0$, but ν is not an indefinite-integral measure over μ. (*Hint*: 216Yb.)

(n) Let (X,Σ,μ) be a localizable measure space, and ν a complete localizable measure on X, with domain $\mathrm{T}\supseteq\Sigma$, which is the completion of its restriction to Σ. Show that if we set $\nu_1 F=\sup\{\nu(F\cap E):E\in\Sigma,\,\mu E<\infty\}$ for every $F\in\mathrm{T}$, then ν_1 is an indefinite-integral measure over μ, and there is an $H\in\Sigma$ such that $\nu_1 F=\nu(F\cap H)$ for every $F\in\mathrm{T}$.

(o) Let X be a set, and M_{sf} the set of semi-finite measures on X. For μ, $\nu\in\mathrm{M}_{\mathrm{sf}}$ say that $\mu\preccurlyeq\nu$ if $\operatorname{dom}\nu\subseteq\operatorname{dom}\mu$, $\mu F\le\nu F$ for every $F\in\operatorname{dom}\nu$, and whenever $E\in\operatorname{dom}\mu$ and $\mu E>0$ there is an $F\in\operatorname{dom}\nu$ such that $F\subseteq E$ and $0<\mu F<\infty$. (i) Show that $(\mathrm{M}_{\mathrm{sf}},\preccurlyeq)$ is a partially ordered set. (ii) Show that if $A\subseteq\mathrm{M}_{\mathrm{sf}}$ is a non-empty set with an upper bound in M_{sf}, then it has a least upper bound λ defined by saying that $\operatorname{dom}\lambda=\bigcap_{\mu\in A}\operatorname{dom}\mu$ and, for $E\in\operatorname{dom}\lambda$,

$$\begin{aligned}\lambda E=\sup\{\sum_{i=0}^n\mu_iF_i:\mu_0,\dots,\mu_n\in A,\ \langle F_i\rangle_{i\le n}\text{ is a partition of }E,\\ F_i\in\operatorname{dom}\lambda\text{ for every }i\le n\}\\ =\sup\{\sum_{i=0}^n\mu_iF_i:\mu_0,\dots,\mu_n\in A,\ F_0,\dots,F_n\text{ are disjoint},\\ F_i\in\operatorname{dom}\mu_i\text{ and }F_i\subseteq E\text{ for every }i\le n\}.\end{aligned}$$

(iii) Suppose that μ, $\nu\in\mathrm{M}_{\mathrm{sf}}$ have completions $\hat\mu$, $\hat\nu$ and c.l.d. versions $\tilde\mu$, $\tilde\nu$. Show that $\tilde\mu\preccurlyeq\hat\mu\preccurlyeq\mu$. Show that if $\mu\preccurlyeq\nu$ then $\hat\mu\preccurlyeq\hat\nu$ and $\tilde\mu\preccurlyeq\tilde\nu$.

234 Notes and comments One of the striking features of measure theory, compared with other comparably abstract branches of pure mathematics, is the relative unimportance of any notion of 'morphism'. The theory of groups, for instance, is dominated by the concept of 'homomorphism', and general topology gives a similar place to 'continuous function'. In my view, the nearest equivalent in measure theory is the idea of 'inverse-measure-preserving function' (234A). I mean in Volumes 3 and 4 to explore this concept more thoroughly. In this volume I will content myself with signalling such functions when they arise, and with the basic facts listed in 234B.

Naturally linked with the idea of inverse-measure-preserving function is the construction of 'image measures' (234C). These appear everywhere in the subject, starting with the not-quite-elementary 234Yc. They are of such importance that it is natural to explore variations, as in 234F and 234Yb, but in my view none are of comparable significance.

Nearly half the section is taken up with 'indefinite-integral measures'. I have taken this part very carefully because the ideas I wish to express here, in so far as they extend the work of §232, rely critically on the details of the formulation in 234I, and it is easy to make a false step once we have left the relatively sheltered context of complete σ-finite measures. I believe that if we take a little trouble at this point we can develop a theory (234K-234N) which will offer a smooth path to later applications; to see what I have in mind, you can refer to the entries under 'indefinite-integral measure' in the index. For the moment I mention only a kind of Radon-Nikodým theorem for localizable measures (234O).

The partial ordering described in 234P-234Q is only one of many which can be considered, and for some purposes it seems unsatisfactory. The most important examples will appear in Chapter 41 of Volume 4, and have a variety of special features for which it might be worth setting out further abstractions. However the version here has the merit of simplicity and supports at least some of the relevant ideas (234Xl). For an alternative notion, see 234Yo.

235 Measurable transformations

I turn now to a topic which is separate from the Radon-Nikodým theorem, but which seems to fit better here than in either of the next two chapters. I seek to give results which will generalize the basic formula of calculus

$$\int g(y)dy=\int g(\phi(x))\phi'(x)dx$$

in the context of a general transformation ϕ between measure spaces. The principal results are I suppose 235A/235E, which are very similar expressions of the basic idea, and 235J, which gives a general criterion for a stronger result. A formulation from a different direction is in 235R.

235A I start with the basic result, which is already sufficient for a large proportion of the applications I have in mind.

Theorem Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and $\phi : D_\phi \to Y$, $J : D_J \to [0,\infty[$ functions defined on conegligible subsets D_ϕ, D_J of X such that

$$\int J \times \chi(\phi^{-1}[F])d\mu \text{ exists} = \nu F$$

whenever $F \in \mathrm{T}$ and $\nu F < \infty$. Then

$$\int_{\phi^{-1}[H]} J \times g\phi\, d\mu \text{ exists} = \int_H g\, d\nu$$

for every ν-integrable function g taking values in $[-\infty,\infty]$ and every $H \in \mathrm{T}$, provided that we interpret $(J \times g\phi)(x)$ as 0 when $J(x) = 0$ and $g(\phi(x))$ is undefined. Consequently, interpreting $J \times f\phi$ in the same way,

$$\underline{\int} f d\nu \le \underline{\int} J \times f\phi\, d\mu \le \overline{\int} J \times f\phi\, d\mu \le \overline{\int} f d\nu$$

for every $[-\infty,\infty]$-valued function f defined almost everywhere in Y.

proof (a) If g is a simple function, say $g = \sum_{i=0}^n a_i \chi F_i$ where $\nu F_i < \infty$ for each i, then

$$\int J \times g\phi\, d\mu = \sum_{i=0}^n a_i \int J \times \chi(\phi^{-1}[F_i])\, d\mu = \sum_{i=0}^n a_i \nu F_i = \int g\, d\nu.$$

(b) If $\nu F = 0$ then $\int J \times \chi(\phi^{-1}[F]) = 0$ so $J = 0$ a.e. on $\phi^{-1}[F]$. So if g is defined ν-a.e., $J = 0$ μ-a.e. on $X \setminus \mathrm{dom}(g\phi) = (X \setminus D_\phi) \cup \phi^{-1}[Y \setminus \mathrm{dom}\, g]$, and, on the convention proposed, $J \times g\phi$ is defined μ-a.e. Moreover, if $\lim_{n\to\infty} g_n = g$ ν-a.e., then $\lim_{n\to\infty} J \times g_n\phi = J \times g\phi$ μ-a.e. So if $\langle g_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence of simple functions converging almost everywhere to g, $\langle J \times g_n\phi\rangle_{n\in\mathbb{N}}$ will be an non-decreasing sequence of integrable functions converging almost everywhere to $J \times g\phi$; by B.Levi's theorem,

$$\int J \times g\phi\, d\mu \text{ exists} = \lim_{n\to\infty} \int J \times g_n\phi\, d\mu = \lim_{n\to\infty} \int g_n d\nu = \int g\, d\nu.$$

(c) If $g = g^+ - g^-$, where g^+ and g^- are ν-integrable functions, then

$$\int J \times g\phi\, d\mu = \int J \times g^+\phi\, d\mu - \int J \times g^-\phi\, d\mu = \int g^+ d\nu - \int g^- d\nu = \int g\, d\nu.$$

(d) This deals with the case $H = Y$. For the general case, we have

$$\begin{aligned} \int_H g\, d\nu &= \int (g \times \chi H) d\nu \\ &= \int J \times (g \times \chi H)\phi\, d\mu = \int J \times g\phi \times \chi(\phi^{-1}[H]) d\mu = \int_{\phi^{-1}[H]} J \times g\phi\, d\mu \end{aligned}$$

(131Fa)

by 214F.

(e) For the upper and lower integrals, I note first that if F is ν-negligible then $\int J \times \chi(\phi^{-1}[F])d\mu = 0$, so that $J = 0$ μ-a.e. on $\phi^{-1}[F]$. It follows that if f and g are $[-\infty,\infty]$-valued functions on subsets of Y and $f \le_{\text{a.e.}} g$, then $J \times f\phi \le_{\text{a.e.}} J \times g\phi$. Now if $\overline{\int} f d\nu = \infty$, we surely have $\overline{\int} J \times f\phi\, d\mu \le \overline{\int} f d\nu$. Otherwise,

$$\begin{aligned} \overline{\int} f d\nu &= \inf\{\int g\, d\nu : g \text{ is } \nu-\text{integrable and } f \le_{\text{a.e.}} g\} \\ &= \inf\{\int J \times g\phi\, d\mu : g \text{ is } \nu\text{-integrable and } f \le_{\text{a.e.}} g\} \\ &\le \inf\{\int h\, d\mu : h \text{ is } \mu\text{-integrable and } J \times f\phi \le_{\text{a.e.}} h\} = \overline{\int} J \times f\phi\, d\mu. \end{aligned}$$

Similarly, or applying this argument to $-f$, we have $\underline{\int} J \times f\phi\, d\mu \le \underline{\int} f\, d\nu$.

235B Remarks (a) Note the particular convention

$$0 \times \text{undefined} = 0$$

which I am applying to the interpretation of $J \times g\phi$. This is the first of a number of technical points which will concern us in this section. The point is that if g is defined ν-almost everywhere, then for any extension of g to a function $g_1 : Y \to \mathbb{R}$ we shall, on this convention, have $J \times g\phi = J \times g_1\phi$ except on $\{x : J(x) > 0,\ \phi(x) \in Y \setminus \operatorname{dom} g\}$, which is negligible; so that

$$\int J \times g\phi \, d\mu = \int J \times g_1\phi \, d\mu = \int g_1 d\nu = \int g \, d\nu$$

if g and g_1 are integrable. Thus the convention is appropriate here, and while it adds a phrase to the statements of many of the results of this section, it makes their application smoother. (But I ought to insist that I am using this as a local convention only, and the ordinary rule $0 \times \text{undefined} = \text{undefined}$ will stand elsewhere in this treatise unless explicitly overruled.)

(b) I have had to take care in the formulation of this theorem to distinguish between the hypothesis

$$\int J(x)\chi(\phi^{-1}[F])(x)\mu(dx) \text{ exists} = \nu F \text{ whenever } \nu F < \infty$$

and the perhaps more elegant alternative

$$\int_{\phi^{-1}[F]} J(x)\mu(dx) \text{ exists} = \nu F \text{ whenever } \nu F < \infty,$$

which is not quite adequate for the theorem. (See 235Q below.) Recall that by $\int_A f$ I mean $\int (f \restriction A) d\mu_A$, where μ_A is the subspace measure on A (214D). It is possible for $\int_A (f \restriction A) d\mu_A$ to be defined even when $\int f \times \chi A \, d\mu$ is not; for instance, take μ to be Lebesgue measure on $[0,1]$, A any non-measurable subset of $[0,1]$, and f the constant function with value 1; then $\int_A f = \mu^* A$, but $f \times \chi A = \chi A$ is not μ-integrable. It is however the case that if $\int f \times \chi A \, d\mu$ is defined, then so is $\int_A f$, and the two are equal; this is a consequence of 214F. While 235P shows that in most of the cases relevant to the present volume the distinction can be passed over, it is important to avoid assuming that $\phi^{-1}[F]$ is measurable for every $F \in \mathrm{T}$. A simple example is the following. Set $X = Y = [0,1]$. Let μ be Lebesgue measure on $[0,1]$, and define ν by setting

$$\mathrm{T} = \{F : F \subseteq [0,1],\ F \cap [0, \tfrac{1}{2}] \text{ is Lebesgue measurable}\},$$

$$\nu F = 2\mu(F \cap [0, \tfrac{1}{2}]) \text{ for every } F \in \mathrm{T}.$$

Set $\phi(x) = x$ for every $x \in [0,1]$. Then we have

$$\nu F = \int_F J \, d\mu = \int J \times \chi(\phi^{-1}[F]) d\mu$$

for every $F \in \mathrm{T}$, where $J(x) = 2$ for $x \in [0, \frac{1}{2}]$ and $J(x) = 0$ for $x \in \left]\frac{1}{2}, 1\right]$. But of course there are subsets F of $[\frac{1}{2}, 1]$ which are not Lebesgue measurable (see 134D), and such an F necessarily belongs to T, even though $\phi^{-1}[F]$ does not belong to the domain Σ of μ.

The point here is that if $\nu F_0 = 0$ then we expect to have $J = 0$ on $\phi^{-1}[F_0]$, and it is of no importance whether $\phi^{-1}[F]$ is measurable for $F \subseteq F_0$.

235C Theorem 235A is concerned with integration, and accordingly the hypothesis $\int J \times \chi(\phi^{-1}[F]) d\mu = \nu F$ looks only at sets F of finite measure. If we wish to consider measurability of non-integrable functions, we need a slightly stronger hypothesis. I approach this version more gently, with a couple of lemmas.

Lemma Let Σ, T be σ-algebras of subsets of X and Y respectively. Suppose that $D \subseteq X$ and that $\phi : D \to Y$ is a function such that $\phi^{-1}[F] \in \Sigma_D$, the subspace σ-algebra, for every $F \in \mathrm{T}$. Then $g\phi$ is Σ-measurable for every $[-\infty, \infty]$-valued T-measurable function g defined on a subset of Y.

proof Set $C = \operatorname{dom} g$ and $B = \operatorname{dom} g\phi = \phi^{-1}[C]$. If $a \in \mathbb{R}$, then there is an $F \in \mathrm{T}$ such that $\{y : g(y) \le a\} = F \cap C$. Now there is an $E \in \Sigma$ such that $\phi^{-1}[F] = E \cap D$. So

$$\{x : g\phi(x) \le a\} = B \cap E \in \Sigma_B.$$

As a is arbitrary, $g\phi$ is Σ-measurable.

235D Some of the results below are easier when we can move freely between measure spaces and their completions (212C). The next lemma is what we need.

Lemma Let (X, Σ, μ) and (Y, T, ν) be measure spaces, with completions $(X, \hat{\Sigma}, \hat{\mu})$ and $(Y, \hat{\mathrm{T}}, \hat{\nu})$. Let $\phi : D_\phi \to Y$, $J : D_J \to [0, \infty[$ be functions defined on conegligible subsets of X.

(a) If $\int J \times \chi(\phi^{-1}[F]) d\mu = \nu F$ whenever $F \in \mathrm{T}$ and $\nu F < \infty$, then $\int J \times \chi(\phi^{-1}[F]) d\hat{\mu} = \hat{\nu} F$ whenever $F \in \hat{\mathrm{T}}$ and $\hat{\nu} F < \infty$.

(b) If $\int J\times\chi(\phi^{-1}[F])d\mu=\nu F$ whenever $F\in\mathrm{T}$, then $\int J\times\chi(\phi^{-1}[F])d\hat{\mu}=\hat{\nu}F$ whenever $F\in\hat{\mathrm{T}}$.

proof Both rely on the fact that either hypothesis is enough to ensure that $\int J\times\chi(\phi^{-1}[F])d\mu=0$ whenever $\nu F=0$. Accordingly, if F is ν-negligible, so that there is an $F'\in\mathrm{T}$ such that $F\subseteq F'$ and $\nu F'=0$, we shall have

$$\int J\times\chi(\phi^{-1}[F])d\mu=\int J\times\chi(\phi^{-1}[F'])d\mu=0.$$

But now, given any $F\in\hat{\mathrm{T}}$, there is an $F_0\in\mathrm{T}$ such that $F_0\subseteq F$ and $\hat{\nu}(F\setminus F_0)=0$, so that

$$\begin{aligned}\int J\times\chi(\phi^{-1}[F])d\hat{\mu}&=\int J\times\chi(\phi^{-1}[F])d\mu\\&=\int J\times\chi(\phi^{-1}[F_0])d\mu+\int J\times\chi(\phi^{-1}[F\setminus F_0])d\mu\\&=\nu F_0=\hat{\nu}F,\end{aligned}$$

provided (for part (a)) that $\hat{\nu}F<\infty$.

Remark Thus if we have the hypotheses of any of the principal results of this section valid for a pair of non-complete measure spaces, we can expect to be able to draw some conclusion by applying the result to the completions of the given spaces.

235E Now I come to the alternative version of 235A.

Theorem Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and $\phi:D_\phi\to Y$, $J:D_J\to[0,\infty[$ two functions defined on conegligible subsets of X such that

$$\int J\times\chi(\phi^{-1}[F])d\mu=\nu F$$

for every $F\in\mathrm{T}$, allowing ∞ as a value of the integral.

(a) $J\times g\phi$ is μ-virtually measurable for every ν-virtually measurable function g defined on a subset of Y.

(b) Let g be a ν-virtually measurable $[-\infty,\infty]$-valued function defined on a conegligible subset of Y. Then $\int J\times g\phi\,d\mu=\int g\,d\nu$ whenever either integral is defined in $[-\infty,\infty]$, if we interpret $(J\times g\phi)(x)$ as 0 when $J(x)=0$ and $g(\phi(x))$ is undefined.

proof Let $(X,\hat{\Sigma},\hat{\mu})$ and $(Y,\hat{\mathrm{T}},\hat{\nu})$ be the completions of (X,Σ,μ) and (Y,T,ν). By 235D,

$$\int J\times\chi(\phi^{-1}[F])d\hat{\mu}=\hat{\nu}F$$

for every $F\in\hat{\mathrm{T}}$. Recalling that a real-valued function is μ-virtually measurable iff it is $\hat{\Sigma}$-measurable (212Fa), and that $\int f d\mu=\int f d\hat{\mu}$ if either is defined in $[-\infty,\infty]$ (212Fb), the conclusions we are seeking are

(a)′ $J\times g\phi$ is $\hat{\Sigma}$-measurable for every $\hat{\mathrm{T}}$-measurable function g defined on a subset of Y;

(b)′ $\int J\times g\phi\,d\hat{\mu}=\int g\,d\hat{\nu}$ whenever g is a $\hat{\mathrm{T}}$-measurable function defined almost everywhere in Y and either integral is defined in $[-\infty,\infty]$.

(a) When I write

$$\int J\times\chi D_\phi d\mu=\int J\times\chi(\phi^{-1}[Y])d\mu=\nu Y,$$

which is part of the hypothesis of this theorem, I mean to imply that $J\times\chi D_\phi$ is μ-virtually measurable, that is, is $\hat{\Sigma}$-measurable. Because D_ϕ is conegligible, it follows that J is $\hat{\Sigma}$-measurable, and its domain D_J, being conegligible, also belongs to $\hat{\Sigma}$. Set $G=\{x:x\in D_J,\ J(x)>0\}\in\hat{\Sigma}$. Then for any set $A\subseteq X$, $J\times\chi A$ is $\hat{\Sigma}$-measurable iff $A\cap G\in\hat{\Sigma}$. So the hypothesis is just that $G\cap\phi^{-1}[F]\in\hat{\Sigma}$ for every $F\in\hat{\mathrm{T}}$.

Now let g be a $[-\infty,\infty]$-valued function, defined on a subset C of Y, which is $\hat{\mathrm{T}}$-measurable. Applying 235C to $\phi{\restriction}G$, we see that $g\phi{\restriction}G$ is $\hat{\Sigma}$-measurable, so $(J\times g\phi){\restriction}G$ is $\hat{\Sigma}$-measurable. On the other hand, $J\times g\phi$ is zero almost everywhere in $X\setminus G$, so (because $G\in\hat{\Sigma}$) $J\times g\phi$ is $\hat{\Sigma}$-measurable, as required.

(b)(i) Suppose first that $g\ge 0$. Then $J\times g\phi\ge 0$, so (a) tells us that $\int J\times g\phi$ is defined in $[0,\infty]$.

(α) If $\int g\,d\hat{\nu}<\infty$ then $\int J\times g\phi\,d\hat{\mu}=\int g\,d\hat{\nu}$ by 235A.

(β) If there is some $\epsilon>0$ such that $\hat{\nu}H=\infty$, where $H=\{y:g(y)\ge\epsilon\}$, then

$$\int J\times g\phi\,d\hat{\mu}\ge\epsilon\int J\times\chi(\phi^{-1}[H])d\hat{\mu}=\epsilon\hat{\nu}H=\infty,$$

so

$$\int J\times g\phi\,d\hat\mu=\infty=\int g\,d\hat\nu.$$

(γ) Otherwise,

$$\begin{aligned}\int J\times g\phi\,d\hat\mu&\ge\sup\{\int J\times h\phi\,d\hat\mu : h\text{ is }\hat\nu\text{-integrable, }0\le h\le g\}\\&=\sup\{\int h\,d\hat\nu : h\text{ is }\hat\nu\text{-integrable, }0\le h\le g\}=\int g\,d\hat\nu=\infty,\end{aligned}$$

so once again $\int J\times\phi\,d\hat\mu=\int g\,d\hat\nu$.

(ii) For general real-valued g, apply (i) to g^+ and g^- where $g^+=\frac{1}{2}(|g|+g)$, $g^-=\frac{1}{2}(|g|-g)$; the point is that $(J\times g\phi)^+=J\times g^+\phi$ and $(J\times g\phi)^-=J\times g^-\phi$, so that

$$\int J\times g\phi=\int J\times g^+\phi-\int J\times g^-\phi=\int g^+-\int g^-=\int g$$

if either side is defined in $[-\infty,\infty]$.

235F Remarks (a) Of course there are two special cases of this theorem which between them carry all its content: the case $J=1$ a.e. and the case in which $X=Y$ and ϕ is the identity function. If $J=\chi X$ we are very close to 235G below, and if ϕ is the identity function we are close to the indefinite-integral measures of §234.

(b) As in 235A, we can strengthen the conclusion of (b) in 235E to

$$\int_{\phi^{-1}[F]}J\times g\phi\,d\mu=\int_F g\,d\nu$$

whenever $F\in\mathrm{T}$ and $\int_F g\,d\nu$ is defined in $[-\infty,\infty]$.

235G Theorem Let (X,Σ,μ) and (Y,T,ν) be measure spaces and $\phi:X\to Y$ an inverse-measure-preserving function. Then

(a) if g is a ν-virtually measurable $[-\infty,\infty]$-valued function defined on a subset of Y, $g\phi$ is μ-virtually measurable;

(b) if g is a ν-virtually measurable $[-\infty,\infty]$-valued function defined on a conegligible subset of Y, $\int g\phi\,d\mu=\int g\,d\nu$ if either integral is defined in $[-\infty,\infty]$;

(c) if g is a ν-virtually measurable $[-\infty,\infty]$-valued function defined on a conegligible subset of Y, and $F\in\mathrm{T}$, then $\int_{\phi^{-1}[F]}g\phi\,d\mu=\int_F g\,d\nu$ if either integral is defined in $[-\infty,\infty]$.

proof (a) This follows immediately from 234Ba and 235C; taking $\hat\Sigma$, $\hat{\mathrm{T}}$ to be the domains of the completions of μ, ν respectively, $\phi^{-1}[F]\in\hat\Sigma$ for every $F\in\hat{\mathrm{T}}$, so if g is $\hat{\mathrm{T}}$-measurable then $g\phi$ will be $\hat\Sigma$-measurable.

(b) Apply 235E with $J=\chi X$; we have

$$\int J\times\chi(\phi^{-1}[F])d\mu=\mu\phi^{-1}[F]=\nu F$$

for every $F\in\mathrm{T}$, so

$$\int g\phi=\int J\times g\phi=\int g$$

if either integral is defined in $[-\infty,\infty]$.

(c) Apply (b) to $g\times\chi F$.

235H The image measure catastrophe Applications of 235A would run much more smoothly if we could say

'$\int g\,d\nu$ exists and is equal to $\int J\times g\phi\,d\mu$ for every $g:Y\to\mathbb{R}$ such that $J\times g\phi$ is μ-integrable'.

Unhappily there is no hope of a universally applicable result in this direction. Suppose, for instance, that ν is Lebesgue measure on $Y=[0,1]$, that $X\subseteq[0,1]$ is a non-Lebesgue-measurable set of outer measure 1 (134D), that μ is the subspace measure ν_X on X, and that $\phi(x)=x$ for $x\in X$. Then

$$\mu\phi^{-1}F=\nu^*(X\cap F)=\nu F$$

for every Lebesgue measurable set $F\subseteq Y$, so we can take $J=\chi X$ and the hypotheses of 235A and 235E will be satisfied. But if we write $g=\chi X:[0,1]\to\{0,1\}$, then $\int g\phi\,d\mu$ is defined even though $\int g\,d\nu$ is not.

The point here is that there is a set $A\subseteq Y$ such that (in the language of 235A/235E) $\phi^{-1}[A]\in\Sigma$ but $A\notin\hat{\mathrm{T}}$. This is the **image measure catastrophe**. The search for contexts in which we can be sure that it does not occur will be one of the motive themes of Volume 4. For the moment, I will offer some general remarks (235I-235J), and describe one of the important cases in which the problem does not arise (235K).

235I Lemma Let Σ, T be σ-algebras of subsets of X, Y respectively, and ϕ a function from a subset D of X to Y. Suppose that $G \subseteq X$ and that

$$\mathrm{T} = \{F : F \subseteq Y,\, G \cap \phi^{-1}[F] \in \Sigma\}.$$

Then a real-valued function g, defined on a member of T, is T-measurable iff $\chi G \times g\phi$ is Σ-measurable.

proof Because surely $Y \in \mathrm{T}$, the hypothesis implies that $G \cap D = G \cap \phi^{-1}[Y]$ belongs to Σ.

Let $g : C \to \mathbb{R}$ be a function, where $C \in \mathrm{T}$. Set $B = \mathrm{dom}(g\phi) = \phi^{-1}[C]$, and for $a \in \mathbb{R}$ set $F_a = \{y : g(y) \geq a\}$,

$$E_a = G \cap \phi^{-1}[F_a] = \{x : x \in G \cap B,\, g\phi(x) \geq a\}.$$

Note that $G \cap B \in \Sigma$ because $C \in \mathrm{T}$.

(i) If g is T-measurable, then $F_a \in \mathrm{T}$ and $E_a \in \Sigma$ for every a. Now

$$G \cap \{x : x \in B,\, g\phi(x) \geq a\} = G \cap \phi^{-1}[F_a] = E_a,$$

so $\{x : x \in B,\, (\chi G \times g\phi)(x) \geq a\}$ is either E_a or $E_a \cup (B \setminus G)$, and in either case is relatively Σ-measurable in B. As a is arbitrary, $\chi G \times g\phi$ is Σ-measurable.

(ii) If $\chi G \times g\phi$ is Σ-measurable, then, for any $a \in \mathbb{R}$,

$$E_a = \{x : x \in G \cap B,\, (\chi G \times g\phi)(x) \geq a\} \in \Sigma$$

because $G \cap B \in \Sigma$ and $\chi G \times g\phi$ is Σ-measurable. So $F_a \in \mathrm{T}$. As a is arbitrary, g is T-measurable.

235J Theorem Let (X, Σ, μ) and (Y, T, ν) be complete measure spaces. Let $\phi : D_\phi \to Y$, $J : D_J \to [0, \infty[$ be functions defined on conegligible subsets of X, and set $G = \{x : x \in D_J,\, J(x) > 0\}$. Suppose that

$$\mathrm{T} = \{F : F \subseteq Y,\, G \cap \phi^{-1}[F] \in \Sigma\},$$

$$\nu F = \int J \times \chi(\phi^{-1}[F]) d\mu \text{ for every } F \in \mathrm{T}.$$

Then, for any real-valued function g defined on a subset of Y, $\int J \times g\phi\, d\mu = \int g\, d\nu$ whenever either integral is defined in $[-\infty, \infty]$, provided that we interpret $(J \times g\phi)(x)$ as 0 when $J(x) = 0$ and $g(\phi(x))$ is undefined.

proof If g is T-measurable and defined almost everywhere, this is a consequence of 235E. So I have to show that if $J \times g\phi$ is measurable and defined almost everywhere, so is g. Set $W = Y \setminus \mathrm{dom}\, g$. Then $J \times g\phi$ is undefined on $G \cap \phi^{-1}[W]$, because $g\phi$ is undefined there and we cannot take advantage of the escape clause available when $J = 0$; so $G \cap \phi^{-1}[W]$ must be negligible, therefore measurable, and $W \in \mathrm{T}$. Next,

$$\nu W = \int J \times \chi(\phi^{-1}[W]) = 0$$

because $J \times \chi(\phi^{-1}[W])$ can be non-zero only on the negligible set $G \cap \phi^{-1}[W]$. So g is defined almost everywhere.

Note that the hypothesis surely implies that $J \times \chi D_\phi = J \times \chi(\phi^{-1}[Y])$ is measurable, so that J is measurable (because D_ϕ is conegligible) and $G \in \Sigma$. Writing $K(x) = 1/J(x)$ for $x \in G$, 0 for $x \in X \setminus G$, the function $K : X \to \mathbb{R}$ is measurable, and

$$\chi G \times g\phi = K \times J \times g\phi$$

is measurable. So 235I tells us that g must be measurable, and we're done.

Remark When $J = \chi X$, the hypothesis of this theorem becomes

$$\mathrm{T} = \{F : F \subseteq Y,\, \phi^{-1}[F] \in \Sigma\}, \quad \nu F = \mu\phi^{-1}[F] \text{ for every } F \in \mathrm{T};$$

that is, ν is the image measure $\mu\phi^{-1}$ as defined in 234D.

235K Corollary Let (X, Σ, μ) be a complete measure space, and J a non-negative measurable function defined on a conegligible subset of X. Let ν be the associated indefinite-integral measure, and T its domain. Then for any real-valued function g defined on a subset of X, g is T-measurable iff $J \times g$ is Σ-measurable, and $\int g\, d\nu = \int J \times g\, d\mu$ if either integral is defined in $[-\infty, \infty]$, provided that we interpret $(J \times g)(x)$ as 0 when $J(x) = 0$ and $g(x)$ is undefined.

proof Put 235J, taking $Y = X$ and ϕ the identity function, together with 234Ld.

235L Applying the Radon-Nikodým theorem In order to use 235A-235J effectively, we need to be able to find suitable functions J. This can be difficult – some very special examples will take up most of Chapter 26 below. But there are many circumstances in which we can be sure that such J exist, even if we do not know what they are. A minimal requirement is that if $\nu F < \infty$ and $\mu^*\phi^{-1}[F] = 0$ then $\nu F = 0$, because $\int J \times \chi(\phi^{-1}[F]) d\mu$ will be zero for any J. A sufficient condition, in the special case of indefinite-integral measures, is in 234O. Another is the following.

235M Theorem Let (X,Σ,μ) be a σ-finite measure space, (Y,T,ν) a semi-finite measure space, and $\phi:D\to Y$ a function such that

(i) D is a conegligible subset of X,

(ii) $\phi^{-1}[F]\in\Sigma$ for every $F\in\mathrm{T}$;

(iii) $\mu\phi^{-1}[F]>0$ whenever $F\in\mathrm{T}$ and $\nu F>0$.

Then there is a Σ-measurable function $J:X\to[0,\infty[$ such that $\int J\times\chi\phi^{-1}[F]\,d\mu=\nu F$ for every $F\in\mathrm{T}$.

proof (a) To begin with (down to the end of (c) below) let us suppose that $D=X$ and that ν is totally finite.

Set $\tilde{\mathrm{T}}=\{\phi^{-1}[F]:F\in\mathrm{T}\}\subseteq\Sigma$. Then $\tilde{\mathrm{T}}$ is a σ-algebra of subsets of X. **P** (i)

$$\emptyset=\phi^{-1}[\emptyset]\in\tilde{\mathrm{T}}.$$

(ii) If $E\in\tilde{\mathrm{T}}$, take $F\in\mathrm{T}$ such that $E=\phi^{-1}[F]$, so that

$$X\setminus E=\phi^{-1}[Y\setminus F]\in\tilde{\mathrm{T}}.$$

(iii) If $\langle E_n\rangle_{n\in\mathbb{N}}$ is any sequence in $\tilde{\mathrm{T}}$, then for each $n\in\mathbb{N}$ choose $F_n\in\mathrm{T}$ such that $E_n=\phi^{-1}[F_n]$; then

$$\bigcup_{n\in\mathbb{N}}E_n=\phi^{-1}[\bigcup_{n\in\mathbb{N}}F_n]\in\tilde{\mathrm{T}}.\text{ } \mathbf{Q}$$

Next, we have a totally finite measure $\tilde\nu:\tilde{\mathrm{T}}\to[0,\nu Y]$ given by setting

$$\tilde\nu(\phi^{-1}[F])=\nu F\text{ for every }F\in\mathrm{T}.$$

P (i) If F, $F'\in\mathrm{T}$ and $\phi^{-1}[F]=\phi^{-1}[F']$, then $\phi^{-1}[F\triangle F']=\emptyset$, so $\mu(\phi^{-1}[F\triangle F'])=0$ and $\nu(F\triangle F')=0$; consequently $\nu F=\nu F'$. This shows that $\tilde\nu$ is well-defined. (ii) Now

$$\tilde\nu\emptyset=\tilde\nu(\phi^{-1}[\emptyset])=\nu\emptyset=0.$$

(iii) If $\langle E_n\rangle_{n\in\mathbb{N}}$ is a disjoint sequence in $\tilde{\mathrm{T}}$, let $\langle F_n\rangle_{n\in\mathbb{N}}$ be a sequence in T such that $E_n=\phi^{-1}[F_n]$ for each n; set $F'_n=F_n\setminus\bigcup_{m<n}F_m$ for each n; then $E_n=\phi^{-1}[F'_n]$ for each n, so

$$\tilde\nu(\bigcup_{n\in\mathbb{N}}E_n)=\tilde\nu(\phi^{-1}[\bigcup_{n\in\mathbb{N}}F'_n])=\nu(\bigcup_{n\in\mathbb{N}}F'_n)=\sum_{n=0}^\infty\nu F'_n=\sum_{n=0}^\infty\tilde\nu E_n.\text{ } \mathbf{Q}$$

Finally, observe that if $\tilde\nu E>0$ then $\mu E>0$, because $E=\phi^{-1}[F]$ where $\nu F>0$.

(b) By 215B(ix) there is a Σ-measurable function $h:X\to]0,\infty[$ such that $\int h\,d\mu$ is finite. Define $\tilde\mu:\tilde{\mathrm{T}}\to[0,\infty[$ by setting $\tilde\mu E=\int_E h\,d\mu$ for every $E\in\tilde{\mathrm{T}}$; then $\tilde\mu$ is a totally finite measure. If $E\in\tilde{\mathrm{T}}$ and $\tilde\mu E=0$, then (because h is strictly positive) $\mu E=0$ and $\tilde\nu E=0$. Accordingly we may apply the Radon-Nikodým theorem to $\tilde\mu$ and $\tilde\nu$ to see that there is a $\tilde{\mathrm{T}}$-measurable function $g:X\to\mathbb{R}$ such that $\int_E g\,d\tilde\mu=\tilde\nu E$ for every $E\in\tilde{\mathrm{T}}$. Because $\tilde\nu$ is non-negative, we may suppose that $g\ge0$.

(c) Applying 235A to μ, $\tilde\mu$, h and the identity function from X to itself, we see that

$$\int_E g\times h\,d\mu=\int_E g\,d\tilde\mu=\tilde\nu E$$

for every $E\in\tilde{\mathrm{T}}$, that is, that

$$\int J\times\chi(\phi^{-1}[F])d\mu=\nu F$$

for every $F\in\mathrm{T}$, writing $J=g\times h$.

(d) This completes the proof when ν is totally finite and $D=X$. For the general case, if $Y=\emptyset$ then $\mu X=0$ and the result is trivial. Otherwise, let $\hat\phi$ be any extension of ϕ to a function from X to Y which is constant on $X\setminus D$; then $\hat\phi^{-1}[F]\in\Sigma$ for every $F\in\mathrm{T}$, because $D=\phi^{-1}[Y]\in\Sigma$ and $\hat\phi^{-1}[F]$ is always either $\phi^{-1}[F]$ or $(X\setminus D)\cup\phi^{-1}[F]$. Now ν must be σ-finite. **P** Use the criterion of 215B(ii). If $\mathcal{F}$ is a disjoint family in $\{F:F\in\mathrm{T},\,0<\nu F<\infty\}$, then $\mathcal{E}=\{\hat\phi^{-1}[F]:F\in\mathcal{F}\}$ is a disjoint family in $\{E:\mu E>0\}$, so $\mathcal{E}$ and $\mathcal{F}$ are countable. **Q**

Let $\langle Y_n\rangle_{n\in\mathbb{N}}$ be a partition of Y into sets of finite ν-measure, and for each $n\in\mathbb{N}$ set $\nu_nF=\nu(F\cap Y_n)$ for every $F\in\mathrm{T}$. Then ν_n is a totally finite measure on Y, and if $\nu_nF>0$ then $\nu F>0$ so

$$\mu\hat\phi^{-1}[F]=\mu\phi^{-1}[F]>0.$$

Accordingly μ, $\hat\phi$ and ν_n satisfy the assumptions of the theorem together with those of (a) above, and there is a Σ-measurable function $J_n:X\to[0,\infty[$ such that

$$\nu_nF=\int J_n\times\chi(\phi^{-1}[F])d\mu$$

for every $F\in\mathrm{T}$. Now set $J=\sum_{n=0}^\infty J_n\times\chi(\phi^{-1}[Y_n])$, so that $J:X\to[0,\infty[$ is Σ-measurable. If $F\in\mathrm{T}$, then

$$\int J \times \chi(\phi^{-1}[F])d\mu = \sum_{n=0}^{\infty} \int J_n \times \chi(\phi^{-1}[Y_n]) \times \chi(\phi^{-1}[F])d\mu$$
$$= \sum_{n=0}^{\infty} \int J_n \times \chi(\phi^{-1}[F \cap Y_n])d\mu = \sum_{n=0}^{\infty} \nu(F \cap Y_n) = \nu F,$$

as required.

235N Remark Theorem 235M can fail if μ is only strictly localizable rather than σ-finite. **P** Let $X = Y$ be an uncountable set, $\Sigma = \mathcal{P}X$, μ counting measure on X (112Bd), T the countable-cocountable σ-algebra of Y, ν the countable-cocountable measure on Y (211R), $\phi : X \to Y$ the identity map. Then $\phi^{-1}[F] \in \Sigma$ and $\mu\phi^{-1}[F] > 0$ whenever $\nu F > 0$. But if J is any μ-integrable function on X, then $F = \{x : J(x) \neq 0\}$ is countable and

$$\nu(Y \setminus F) = 1 \neq 0 = \int_{\phi^{-1}[Y \setminus F]} J\,d\mu. \text{ **Q**}$$

***235O** There are some simplifications in the case of σ-finite spaces; in particular, 235A and 235E become conflated. I will give an adaptation of the hypotheses of 235A which may be used in the σ-finite case. First a lemma.

Lemma Let (X, Σ, μ) be a measure space and f a non-negative integrable function on X. If $A \subseteq X$ is such that $\int_A f + \int_{X \setminus A} f = \int f$, then $f \times \chi A$ is integrable.

proof By 214Eb, there are μ-integrable functions f_1, f_2 such that f_1 extends $f \restriction A$, f_2 extends $f \restriction X \setminus A$, and

$$\int_E f_1 = \int_{E \cap A} f, \quad \int_E f_2 = \int_{E \setminus A} f$$

for every $E \in \Sigma$. Because f is non-negative, $\int_E f_1$ and $\int_E f_2$ are non-negative for every $E \in \Sigma$, and f_1, f_2 are non-negative a.e. Accordingly we have $f \times \chi A \leq_{\text{a.e.}} f_1$ and $f \times \chi(X \setminus A) \leq_{\text{a.e.}} f_2$, so that $f \leq_{\text{a.e.}} f_1 + f_2$. But also

$$\int f_1 + f_2 = \int_X f_1 + \int_X f_2 = \int_A f + \int_{X \setminus A} f = \int f,$$

so $f =_{\text{a.e.}} f_1 + f_2$. Accordingly

$$f_1 =_{\text{a.e.}} f - f_2 \leq_{\text{a.e.}} f - f \times \chi(X \setminus A) = f \times \chi A \leq_{\text{a.e.}} f_1$$

and $f \times \chi A =_{\text{a.e.}} f_1$ is integrable.

***235P Proposition** Let (X, Σ, μ) be a complete measure space and (Y, T, ν) a complete σ-finite measure space. Suppose that $\phi : D_\phi \to Y$, $J : D_J \to [0, \infty[$ are functions defined on conegligible subsets D_ϕ, D_J of X such that $\int_{\phi^{-1}[F]} J\,d\mu$ exists and is equal to νF whenever $F \in \mathrm{T}$ and $\nu F < \infty$.

(a) $J \times g\phi$ is Σ-measurable for every T-measurable real-valued function g defined on a subset of Y.

(b) If g is a T-measurable real-valued function defined almost everywhere in Y, then $\int J \times g\phi\,d\mu = \int g\,d\nu$ whenever either integral is defined in $[-\infty, \infty]$, interpreting $(J \times g\phi)(x)$ as 0 when $J(x) = 0$, $g(\phi(x))$ is undefined.

proof The point is that the hypotheses of 235E are satisfied. To see this, let us write $\Sigma_C = \{E \cap C : E \in \Sigma\}$ and $\mu_C = \mu^* \restriction \Sigma_C$ for the subspace measure on C, for each $C \subseteq X$. Let $\langle Y_n \rangle_{n \in \mathbb{N}}$ be a non-decreasing sequence of sets with union Y and with $\nu Y_n < \infty$ for every $n \in \mathbb{N}$, starting from $Y_0 = \emptyset$.

(i) Take any $F \in \mathrm{T}$ with $\nu F < \infty$, and set $F_n = F \cup Y_n$ for each $n \in \mathbb{N}$; write $C_n = \phi^{-1}[F_n]$.

Fix n for the moment. Then our hypothesis implies that

$$\int_{C_0} J\,d\mu + \int_{C_n \setminus C_0} J\,d\mu = \nu F + \nu(F_n \setminus F) = \nu F_n = \int_{C_n} J\,d\mu.$$

If we regard the subspace measures on C_0 and $C_n \setminus C_0$ as derived from the measure μ_{C_n} of C_n (214Ce), then 235O tells us that $J \times \chi C_0$ is μ_{C_n}-integrable, and there is a μ-integrable function h_n such that h_n extends $(J \times \chi C_0) \restriction C_n$.

Let E be a μ-conegligible set, included in the domain D_ϕ of ϕ, such that $h_n \restriction E$ is Σ-measurable for every n. Because $\langle C_n \rangle_{n \in \mathbb{N}}$ is a non-decreasing sequence with union $\phi^{-1}[\bigcup_{n \in \mathbb{N}} F_n] = D_\phi$,

$$(J \times \chi C_0)(x) = \lim_{n \to \infty} h_n(x)$$

for every $x \in E$, and $(J \times \chi C_0) \restriction E$ is measurable. At the same time, we know that there is a μ-integrable h extending $J \restriction C_0$, and $0 \leq_{\text{a.e.}} J \times \chi C_0 \leq_{\text{a.e.}} |h|$. Accordingly $J \times \chi C_0$ is integrable, and (using 214F)

$$\int J \times \chi\phi^{-1}[F]\,d\mu = \int J \times \chi C_0\,d\mu = \int_{C_0} J \restriction C_0\,d\mu_{C_0} = \nu F.$$

(ii) This deals with F of finite measure. For general $F \in \mathrm{T}$,

$$\int J \times \chi(\phi^{-1}[F])\,d\mu = \lim_{n\to\infty} \int J \times \chi(\phi^{-1}[F \cap Y_n])\,d\mu = \lim_{n\to\infty} \nu(F \cap Y_n) = \nu F.$$

So the hypotheses of 235E are satisfied, and the result follows at once.

***235Q** I remarked in 235Bb that a difficulty can arise in 235A, for general measure spaces, if we speak of $\int_{\phi^{-1}[F]} J\,d\mu$ in the hypothesis, in place of $\int J \times \chi(\phi^{-1}[F])d\mu$. Here is an example.

Example Set $X = Y = [0,2]$. Write Σ_L for the algebra of Lebesgue measurable subsets of $\mathbb{R}$, and μ_L for Lebesgue measure; write μ_c for counting measure on $\mathbb{R}$. Set

$$\Sigma = \mathrm{T} = \{E : E \subseteq [0,2],\ E \cap [0,1[\ \in \Sigma_L\};$$

of course this is a σ-algebra of subsets of $[0,2]$. For $E \in \Sigma = \mathrm{T}$, set

$$\mu E = \nu E = \mu_L(E \cap [0,1[) + \mu_c(E \cap [1,2]);$$

then μ is a complete measure – in effect, it is the direct sum of Lebesgue measure on $[0,1[$ and counting measure on $[1,2]$ (see 214L). It is easy to see that

$$\mu^* B = \mu_L^*(B \cap [0,1[) + \mu_c(B \cap [1,2])$$

for every $B \subseteq [0,2]$. Let $A \subseteq [0,1[$ be a non-Lebesgue-measurable set such that $\mu_L^*(E \setminus A) = \mu_L E$ for every Lebesgue measurable $E \subseteq [0,1[$ (see 134D). Define $\phi : [0,2] \to [0,2]$ by setting $\phi(x) = x+1$ if $x \in A$, $\phi(x) = x$ if $x \in [0,2] \setminus A$.

If $F \in \Sigma$, then $\mu^*(\phi^{-1}[F]) = \mu F$. **P** **(i)** If $F \cap [1,2]$ is finite, then $\mu F = \mu_L(F \cap [0,1]) + \#(F \cap [1,2])$. Now

$$\phi^{-1}[F] = (F \cap [0,1[\ \setminus A) \cup (F \cap [1,2]) \cup \{x : x \in A,\ x+1 \in F\};$$

as the last set is finite, therefore μ-negligible,

$$\mu^*(\phi^{-1}[F]) = \mu_L^*(F \cap [0,1[\ \setminus A) + \#(F \cap [1,2]) = \mu_L(F \cap [0,1[) + \#(F \cap [1,2]) = \mu F.$$

(ii) If $F \cap [1,2]$ is infinite, so is $\phi^{-1}[F] \cap [1,2]$, so

$$\mu^*(\phi^{-1}[F]) = \infty = \mu F. \textbf{ Q}$$

This means that if we set $J(x) = 1$ for every $x \in [0,2]$,

$$\int_{\phi^{-1}[F]} J\,d\mu = \mu_{\phi^{-1}[F]}(\phi^{-1}[F]) = \mu^*(\phi^{-1}[F]) = \mu F$$

for every $F \in \Sigma$, and ϕ, J satisfy the amended hypotheses for 235A. But if we set $g = \chi[0,1[$, then g is μ-integrable, with $\int g\,d\mu = 1$, while

$$J(x)g(\phi(x)) = 1 \text{ if } x \in [0,1] \setminus A,\ 0 \text{ otherwise},$$

so, because $A \notin \Sigma$, $J \times g\phi$ is not measurable, and therefore (since μ is complete) not μ-integrable.

235R Reversing the burden Throughout the work above, I have been using the formula

$$\int J \times g\phi = \int g,$$

as being the natural extension of the formula

$$\int g = \int g\phi \times \phi'$$

of ordinary advanced calculus. But we can also move the 'derivative' J to the other side of the equation, as follows.

Theorem Let (X,Σ,μ), (Y,T,ν) be measure spaces and $\phi : X \to Y$, $J : Y \to [0,\infty[$ functions such that $\int_F J\,d\nu$ and $\mu\phi^{-1}[F]$ are defined in $[0,\infty]$ and equal for every $F \in \mathrm{T}$. Then $\int g\phi\,d\mu = \int J \times g\,d\nu$ whenever g is ν-virtually measurable and defined ν-almost everywhere and either integral is defined in $[-\infty,\infty]$.

proof Let ν_1 be the indefinite-integral measure over ν defined by J, and $\hat\mu$ the completion of μ. Then ϕ is inverse-measure-preserving for $\hat\mu$ and ν_1. **P** If $F \in \mathrm{T}$, then $\nu_1 F = \int_F J\,d\nu = \mu\phi^{-1}[F]$; that is, ϕ is inverse-measure-preserving for μ and $\nu_1{\restriction}\mathrm{T}$. Since ν_1 is the completion of $\nu_1{\restriction}\mathrm{T}$ (234Lb), ϕ is inverse-measure-preserving for μ and ν_1 (234Ba). **Q**

Of course we can also regard ν_1 as being an indefinite-integral measure over the completion $\hat\nu$ of ν (212Fb). So if g is ν-virtually measurable and defined ν-almost everywhere,

$$\int J \times g\,d\nu = \int J \times g\,d\hat\nu = \int g\,d\nu_1 = \int g\phi\,d\hat\mu = \int g\phi\,d\mu$$

if any of the five integrals is defined in $[-\infty,\infty]$, by 235K, 235Gb and 212Fb again.

235X Basic exercises (a) Explain what 235A tells us when $X = Y$, $\mathrm{T} = \Sigma$, ϕ is the identity function and $\nu E = \alpha\mu E$ for every $E \in \Sigma$.

(b) Let (X, Σ, μ) be a measure space, J an integrable non-negative real-valued function on X, and $\phi : D_\phi \to \mathbb{R}$ a measurable function, where D_ϕ is a conegligible subset of X. Set

$$g(a) = \int_{\{x:\phi(x)\le a\}} J$$

for $a \in \mathbb{R}$, and let μ_g be the Lebesgue-Stieltjes measure associated with g. Show that $\int J \times f\phi\, d\mu = \int f d\mu_g$ for every μ_g-integrable real function f.

(c) Let Σ, T and Λ be σ-algebras of subsets of X, Y and Z respectively. Let us say that a function $\phi : A \to Y$, where $A \subseteq X$, is (Σ, T)-measurable if $\phi^{-1}[F] \in \Sigma_A$, the subspace σ-algebra of A, for every $F \in \mathrm{T}$. Suppose that $A \subseteq X$, $B \subseteq Y$, $\phi : A \to Y$ is (Σ, T)-measurable and $\psi : B \to Z$ is (T, Λ)-measurable. Show that $\psi\phi$ is (Σ, Λ)-measurable. Deduce 235C.

(d) Let (X, Σ, μ) be a measure space and (Y, T, ν) a semi-finite measure space. Let $\phi : D_\phi \to Y$ and $J : D_J \to [0, \infty[$ be functions defined on conegligible subsets D_ϕ, D_J of X such that $\int J \times \chi(\phi^{-1}[F])d\mu$ exists $= \nu F$ whenever $F \in \mathrm{T}$ and $\nu F < \infty$. Let g be a T-measurable real-valued function, defined on a conegligible subset of Y. Show that $J \times g\phi$ is μ-integrable iff g is ν-integrable, and the integrals are then equal, provided we interpret $(J \times g\phi)(x)$ as 0 when $J(x) = 0$ and $g(\phi(x))$ is undefined.

(e) Let (X, Σ, μ) be a measure space and $E \in \Sigma$. Define a measure $\mu \llcorner E$ on X by setting $(\mu \llcorner E)(F) = \mu(E \cap F)$ whenever $F \subseteq X$ is such that $F \cap E \in \Sigma$. Show that, for any function f from a subset of X to $[-\infty, \infty]$, $\int f d(\mu \llcorner E) = \int_E f d\mu$ if either is defined in $[-\infty, \infty]$.

>(f) Let $g : \mathbb{R} \to \mathbb{R}$ be a non-decreasing function which is absolutely continuous on every closed bounded interval, and μ_g the associated Lebesgue-Stieltjes measure (114Xa, 225Xf). Write μ for Lebesgue measure on $\mathbb{R}$, and let $f : \mathbb{R} \to \mathbb{R}$ be a function. Show that $\int f \times g'\, d\mu = \int f\, d\mu_g$ in the sense that if one of the integrals exists, finite or infinite, so does the other, and they are then equal.

(g) Let $g : \mathbb{R} \to \mathbb{R}$ be a non-decreasing function and J a non-negative real-valued μ_g-integrable function, where μ_g is the Lebesgue-Stieltjes measure defined from g. Set $h(x) = \int_{]-\infty,x]} J\, d\mu_g$ for each $x \in \mathbb{R}$, and let μ_h be the Lebesgue-Stieltjes measure associated with h. Show that, for any $f : \mathbb{R} \to \mathbb{R}$, $\int f \times J\, d\mu_g = \int f d\mu_h$, in the sense that if one of the integrals is defined in $[-\infty, \infty]$ so is the other, and they are then equal.

>(h) Let X be a set and λ, μ, ν three measures on X such that μ is an indefinite-integral measure over λ, with Radon-Nikodým derivative f, and ν is an indefinite-integral measure over μ, with Radon-Nikodým derivative g. Show that ν is an indefinite-integral measure over λ, and that $f \times g$ is a Radon-Nikodým derivative of ν with respect to λ, provided we interpret $(f \times g)(x)$ as 0 when $f(x) = 0$ and $g(x)$ is undefined.

(i) In 235M, if ν is not semi-finite, show that we can still find a J such that $\int_{\phi^{-1}[F]} J\, d\mu = \nu F$ for every set F of finite measure. (*Hint*: use the 'semi-finite version' of ν, as described in 213Xc.)

(j) Let (X, Σ, μ) be a σ-finite measure space, and T a σ-subalgebra of Σ. Let $\nu : \mathrm{T} \to \mathbb{R}$ be a countably additive functional such that $\nu F = 0$ whenever $F \in \mathrm{T}$ and $\mu F = 0$. Show that there is a μ-integrable function f such that $\int_F f d\mu = \nu F$ for every $F \in \mathrm{T}$. (*Hint*: use the method of 235M, applied to the positive and negative parts of ν.)

(k) Let (X, Σ, μ) and (Y, T, ν) be measure spaces, with completions $(X, \hat{\Sigma}, \hat{\mu})$ and $(Y, \hat{\mathrm{T}}, \hat{\nu})$. Let $\phi : D_\phi \to Y$, $J : D_J \to [0, \infty[$ be functions defined on conegligible subsets of X. Show that if $\int_{\phi^{-1}[F]} J\, d\mu = \nu F$ whenever $F \in \mathrm{T}$ and $\nu F < \infty$, then $\int_{\phi^{-1}[F]} J\, d\mu = \nu F$ whenever $F \in \hat{\mathrm{T}}$ and $\hat{\nu} F < \infty$. Hence, or otherwise, show that 235Pb is valid for non-complete spaces (X, Σ, μ) and (Y, T, ν).

(l) Let (X, Σ, μ) be a complete measure space, Y a set, $\phi : X \to Y$ a function and $\nu = \mu\phi^{-1}$ the corresponding image measure on Y. Let ν_1 be an indefinite-integral measure over ν. Show that there is an indefinite-integral measure μ_1 over μ such that ν_1 is the image measure $\mu_1\phi^{-1}$.

(m) Let (X, Σ, μ) and (Y, T, ν) be measure spaces, and $\phi : X \to Y$ an inverse-measure-preserving function. Let ν_1 be an indefinite-integral measure over ν. Show that there is an indefinite-integral measure μ_1 over μ such that ϕ is inverse-measure-preserving for μ_1 and ν_1.

235Y Further exercises (a) Write T for the algebra of Borel subsets of $Y=[0,1]$, and ν for the restriction of Lebesgue measure to T. Let $A\subseteq[0,1]$ be a set such that both A and $[0,1]\setminus A$ have Lebesgue outer measure 1, and set $X=A\cup[1,2]$. Let Σ be the algebra of relatively Borel subsets of X, and set $\mu E=\mu_A(A\cap E)$ for $E\in\Sigma$, where μ_A is the subspace measure induced on A by Lebesgue measure. Define $\phi:X\to Y$ by setting $\phi(x)=x$ if $x\in A$, $x-1$ if $x\in X\setminus A$. Show that ν is the image measure $\mu\phi^{-1}$, but that, setting $g=\chi([0,1]\setminus A)$, $g\phi$ is μ-integrable while g is not ν-integrable.

(b) Let (X,Σ,μ) be a probability space and T a σ-subalgebra of Σ. Let f be a non-negative μ-integrable function with $\int f d\mu=1$, so that its indefinite-integral measure ν is a probability measure. Let g be a ν-integrable real-valued function and set $h=f\times g$, intepreting $h(x)$ as 0 if $f(x)=0$ and $g(x)$ is undefined. Let f_1, h_1 be conditional expectations of f, h on T with respect to the measure μ, and set $g_1=h_1/f_1$, interpreting $g_1(x)$ as 0 if $h_1(x)=0$ and $f_1(x)$ is either 0 or undefined. Show that g_1 is a conditional expectation of g on T with respect to the measure ν.

235 Notes and comments I see that I have taken up a great deal of space in this section with technicalities; the hypotheses of the theorems vary erratically, with completeness, in particular, being invoked at apparently arbitrary intervals, and ideas repeat themselves in a haphazard pattern. There is nothing deep, and most of the work consists in laboriously verifying details. The trouble with this topic is that it is useful. The results here are abstract expressions of integration-by-substitution; they have applications all over measure theory. I cannot therefore content myself with theorems which will elegantly express the underlying ideas, but must seek formulations which I can quote in later arguments.

I hope that the examples in 235Bb, 235H, 235N, 235Q, 234Ya and 235Ya will go some way to persuade you that there are real traps for the unwary, and that the careful verifications written out at such length are necessary. On the other hand, it is happily the case that in simple contexts, in which the measures μ, ν are σ-finite and the transformations ϕ are Borel isomorphisms, no insuperable difficulties arise, and in particular the image measure catastrophe does not trouble us. But for further work in this direction I refer you to the applications in §263, §265 and §271, and to Volume 4.

Chapter 24

Function spaces

The extraordinary power of Lebesgue's theory of integration is perhaps best demonstrated by its ability to provide structures relevant to questions quite different from those to which it was at first addressed. In this chapter I give the constructions, and elementary properties, of some of the fundamental spaces of functional analysis.

I do not feel called on here to justify the study of normed spaces; if you have not met them before, I hope that the introduction here will show at least that they offer a basis for a remarkable fusion of algebra and analysis. The fragments of the theory of metric spaces, normed spaces and general topology which we shall need are sketched in §§2A2-2A5. The principal 'function spaces' described in this chapter in fact combine three structural elements: they are (infinite-dimensional) linear spaces, they are metric spaces, with associated concepts of continuity and convergence, and they are ordered spaces, with corresponding notions of supremum and infimum. The interactions between these three types of structure provide an inexhaustible wealth of ideas. Furthermore, many of these ideas are directly applicable to a wide variety of problems in more or less applied mathematics, particularly in differential and integral equations, but more generally in any system with infinitely many degrees of freedom.

I have laid out the chapter with sections on L^0 (the space of equivalence classes of all real-valued measurable functions, in which all the other spaces of the chapter are embedded), L^1 (equivalence classes of integrable functions), L^∞ (equivalence classes of bounded measurable functions) and L^p (equivalence classes of pth-power-integrable functions). While ordinary functional analysis gives much more attention to the Banach spaces L^p for $1 \le p \le \infty$ than to L^0, from the special point of view of this book the space L^0 is at least as important and interesting as any of the others. Following these four sections, I return to a study of the standard topology on L^0, the topology of 'convergence in measure' (§245), and then to two linked sections on uniform integrability and weak compactness in L^1 (§§246-247).

There is a technical point here which must never be lost sight of. While it is customary and natural to call L^1, L^2 and the others 'function spaces', their elements are not in fact functions, but equivalence classes of functions. As you see from the language of the preceding paragraph, my practice is to scrupulously maintain the distinction; I give my reasons in the notes to §241.

241 $\mathcal{L}^0$ and L^0

The chief aim of this chapter is to discuss the spaces L^1, L^∞ and L^p of the following three sections. However it will be convenient to regard all these as subspaces of a larger space L^0 of equivalence classes of (virtually) measurable functions, and I have collected in this section the basic facts concerning the ordered linear space L^0.

It is almost the first principle of measure theory that sets of measure zero can often be ignored; the phrase 'negligible set' itself asserts this principle. Accordingly, two functions which agree almost everywhere may often (not always!) be treated as identical. A suitable expression of this idea is to form the space of equivalence classes of functions, saying that two functions are equivalent if they agree on a conegligible set. This is the basis of all the constructions of this chapter. It is a remarkable fact that the spaces of equivalence classes so constructed are actually better adapted to certain problems than the spaces of functions from which they are derived, so that once the technique has been mastered it is easier to do one's thinking in the more abstract spaces.

241A The space $\mathcal{L}^0$: Definition It is time to give a name to a set of functions which has already been used more than once. Let (X,Σ,μ) be a measure space. I write $\mathcal{L}^0$, or $\mathcal{L}^0(\mu)$, for the space of real-valued functions f defined on conegligible subsets of X which are virtually measurable, that is, such that $f{\restriction}E$ is measurable for some conegligible set $E \subseteq X$. Recall that f is μ-virtually measurable iff it is $\hat\Sigma$-measurable, where $\hat\Sigma$ is the completion of Σ (212Fa).

241B Basic properties If (X,Σ,μ) is any measure space, then we have the following facts, corresponding to the fundamental properties of measurable functions listed in §121 of Volume 1. I work through them in order, so that if you have Volume 1 to hand you can see what has to be missed out.

(a) A constant real-valued function defined almost everywhere in X belongs to $\mathcal{L}^0$ (121Ea).

(b) $f+g \in \mathcal{L}^0$ for all f, $g \in \mathcal{L}^0$ (for if $f{\restriction}F$ and $g{\restriction}G$ are measurable, then $(f+g){\restriction}(F\cap G) = (f{\restriction}F)+(g{\restriction}G)$ is measurable)(121Eb).

(c) $cf \in \mathcal{L}^0$ for all $f \in \mathcal{L}^0$, $c \in \mathbb{R}$ (121Ec).

(d) $f \times g \in \mathcal{L}^0$ for all f, $g \in \mathcal{L}^0$ (121Ed).

(e) If $f \in \mathcal{L}^0$ and $h : \mathbb{R} \to \mathbb{R}$ is Borel measurable, then $hf \in \mathcal{L}^0$ (121Eg).

(f) If $\langle f_n \rangle_{n\in\mathbb{N}}$ is a sequence in $\mathcal{L}^0$ and $f = \lim_{n\to\infty} f_n$ is defined (as a real-valued function) almost everywhere in X, then $f \in \mathcal{L}^0$ (121Fa).

(g) If $\langle f_n \rangle_{n\in\mathbb{N}}$ is a sequence in $\mathcal{L}^0$ and $f = \sup_{n\in\mathbb{N}} f_n$ is defined (as a real-valued function) almost everywhere in X, then $f \in \mathcal{L}^0$ (121Fb).

(h) If $\langle f_n \rangle_{n\in\mathbb{N}}$ is a sequence in $\mathcal{L}^0$ and $f = \inf_{n\in\mathbb{N}} f_n$ is defined (as a real-valued function) almost everywhere in X, then $f \in \mathcal{L}^0$ (121Fc).

(i) If $\langle f_n \rangle_{n\in\mathbb{N}}$ is a sequence in $\mathcal{L}^0$ and $f = \limsup_{n\to\infty} f_n$ is defined (as a real-valued function) almost everywhere in X, then $f \in \mathcal{L}^0$ (121Fd).

(j) If $\langle f_n \rangle_{n\in\mathbb{N}}$ is a sequence in $\mathcal{L}^0$ and $f = \liminf_{n\to\infty} f_n$ is defined (as a real-valued function) almost everywhere in X, then $f \in \mathcal{L}^0$ (121Fe).

(k) $\mathcal{L}^0$ is just the set of real-valued functions, defined on subsets of X, which are equal almost everywhere to some Σ-measurable function from X to $\mathbb{R}$. **P** (i) If $g : X \to \mathbb{R}$ is Σ-measurable and $f =_{\text{a.e.}} g$, then $F = \{x : x \in \operatorname{dom} f,\ f(x) = g(x)\}$ is conegligible and $f\restriction F = g\restriction F$ is measurable (121Eh), so $f \in \mathcal{L}^0$. (ii) If $f \in \mathcal{L}^0$, let $E \subseteq X$ be a conegligible set such that $f\restriction E$ is measurable. Then $D = E \cap \operatorname{dom} f$ is conegligible and $f\restriction D$ is measurable, so there is a measurable $h : X \to \mathbb{R}$ agreeing with f on D (121I); and $h =_{\text{a.e.}} f$. **Q**

241C The space L^0: Definition Let (X, Σ, μ) be any measure space. Then $=_{\text{a.e.}}$ is an equivalence relation on $\mathcal{L}^0$. Write L^0, or $L^0(\mu)$, for the set of equivalence classes in $\mathcal{L}^0$ under $=_{\text{a.e.}}$. For $f \in \mathcal{L}^0$, write $f^\bullet$ for its equivalence class in L^0.

241D The linear structure of L^0 Let (X, Σ, μ) be any measure space, and set $\mathcal{L}^0 = \mathcal{L}^0(\mu)$, $L^0 = L^0(\mu)$.

(a) If $f_1, f_2, g_1, g_2 \in \mathcal{L}^0$, $f_1 =_{\text{a.e.}} f_2$ and $g_1 =_{\text{a.e.}} g_2$ then $f_1 + g_1 =_{\text{a.e.}} f_2 + g_2$. Accordingly we may define addition on L^0 by setting $f^\bullet + g^\bullet = (f + g)^\bullet$ for all $f, g \in \mathcal{L}^0$.

(b) If $f_1, f_2 \in \mathcal{L}^0$ and $f_1 =_{\text{a.e.}} f_2$, then $cf_1 =_{\text{a.e.}} cf_2$ for every $c \in \mathbb{R}$. Accordingly we may define scalar multiplication on L^0 by setting $c \cdot f^\bullet = (cf)^\bullet$ for all $f \in \mathcal{L}^0$ and $c \in \mathbb{R}$.

(c) Now L^0 is a linear space over $\mathbb{R}$, with zero $\mathbf{0}^\bullet$, where $\mathbf{0}$ is the function with domain X and constant value 0, and negatives $-(f^\bullet) = (-f)^\bullet$. **P** (i)

$$f + (g + h) = (f + g) + h \text{ for all } f, g, h \in \mathcal{L}^0,$$

so

$$u + (v + w) = (u + v) + w \text{ for all } u, v, w \in L^0.$$

(ii)

$$f + \mathbf{0} = \mathbf{0} + f = f \text{ for every } f \in \mathcal{L}^0,$$

so

$$u + \mathbf{0}^\bullet = \mathbf{0}^\bullet + u = u \text{ for every } u \in L^0.$$

(iii)

$$f + (-f) =_{\text{a.e.}} \mathbf{0} \text{ for every } f \in \mathcal{L}^0,$$

so

$$f^\bullet + (-f)^\bullet = \mathbf{0}^\bullet \text{ for every } f \in \mathcal{L}^0.$$

(iv)

$$f + g = g + f \text{ for all } f, g \in \mathcal{L}^0,$$

so

$$u + v = v + u \text{ for all } u, v \in L^0.$$

(v)

$$c(f+g) = cf + cg \text{ for all } f,\, g \in \mathcal{L}^0 \text{ and } c \in \mathbb{R},$$

so

$$c(u+v) = cu + cv \text{ for all } u,\, v \in L^0 \text{ and } c \in \mathbb{R}.$$

(vi)

$$(a+b)f = af + bf \text{ for all } f \in \mathcal{L}^0 \text{ and } a,\, b \in \mathbb{R},$$

so

$$(a+b)u = au + bu \text{ for all } u \in L^0 \text{ and } a,\, b \in \mathbb{R}.$$

(vii)

$$(ab)f = a(bf) \text{ for all } f \in \mathcal{L}^0 \text{ and } a,\, b \in \mathbb{R},$$

so

$$(ab)u = a(bu) \text{ for all } u \in L^0 \text{ and } a,\, b \in \mathbb{R}.$$

(viii)

$$1f = f \text{ for all } f \in \mathcal{L}^0,$$

so

$$1u = u \text{ for all } u \in L^0. \ \mathbf{Q}$$

241E The order structure of L^0 Let (X, Σ, μ) be any measure space and set $\mathcal{L}^0 = \mathcal{L}^0(\mu)$, $L^0 = L^0(\mu)$.

(a) If f_1, f_2, g_1, $g_2 \in \mathcal{L}^0$, $f_1 =_{\text{a.e.}} f_2$, $g_1 =_{\text{a.e.}} g_2$ and $f_1 \le_{\text{a.e.}} g_1$, then $f_2 \le_{\text{a.e.}} g_2$. Accordingly we may define a relation $\le$ on L^0 by saying that $f^\bullet \le g^\bullet$ iff $f \le_{\text{a.e.}} g$.

(b) Now $\le$ is a partial order on L^0. **P** (i) If f, g, $h \in \mathcal{L}^0$ and $f \le_{\text{a.e.}} g$ and $g \le_{\text{a.e.}} h$, then $f \le_{\text{a.e.}} h$. Accordingly $u \le w$ whenever u, v, $w \in L^0$, $u \le v$ and $v \le w$. (ii) If $f \in \mathcal{L}^0$ then $f \le_{\text{a.e.}} f$; so $u \le u$ for every $u \in L^0$. (iii) If f, $g \in \mathcal{L}^0$ and $f \le_{\text{a.e.}} g$ and $g \le_{\text{a.e.}} f$, then $f =_{\text{a.e.}} g$, so if $u \le v$ and $v \le u$ then $u = v$. **Q**

(c) In fact L^0, with $\le$, is a **partially ordered linear space**, that is, a (real) linear space with a partial order $\le$ such that

if $u \le v$ then $u + w \le v + w$ for every w,

if $0 \le u$ then $0 \le cu$ for every $c \ge 0$.

P (i) If f, g, $h \in \mathcal{L}^0$ and $f \le_{\text{a.e.}} g$, then $f + h \le_{\text{a.e.}} g + h$. (ii) If $f \in \mathcal{L}^0$ and $f \ge 0$ a.e., then $cf \ge 0$ a.e. for every $c \ge 0$. **Q**

(d) More: L^0 is a **Riesz space** or **vector lattice**, that is, a partially ordered linear space such that $u \vee v = \sup\{u, v\}$ and $u \wedge v = \inf\{u, v\}$ are defined for all u, $v \in L^0$. **P** Take f, $g \in \mathcal{L}^0$ such that $f^\bullet = u$ and $g^\bullet = v$. Then $f \vee g$, $f \wedge g \in \mathcal{L}^0$, writing

$$(f \vee g)(x) = \max(f(x), g(x)), \quad (f \wedge g)(x) = \min(f(x), g(x))$$

for $x \in \operatorname{dom} f \cap \operatorname{dom} g$. (Compare 241Bg-h.) Now, for any $h \in \mathcal{L}^0$, we have

$$f \vee g \le_{\text{a.e.}} h \iff f \le_{\text{a.e.}} h \text{ and } g \le_{\text{a.e.}} h,$$

$$h \le_{\text{a.e.}} f \wedge g \iff h \le_{\text{a.e.}} f \text{ and } h \le_{\text{a.e.}} g,$$

so for any $w \in L^0$ we have

$$(f \vee g)^\bullet \le w \iff u \le w \text{ and } v \le w,$$

$$w \le (f \wedge g)^\bullet \iff w \le u \text{ and } w \le v.$$

Thus we have

$$(f \vee g)^\bullet = \sup\{u, v\} = u \vee v, \quad (f \wedge g)^\bullet = \inf\{u, v\} = u \wedge v$$

in L^0. **Q**

(e) In particular, for any $u \in L^0$ we can speak of $|u| = u \vee (-u)$; if $f \in \mathcal{L}^0$ then $|f^\bullet| = |f|^\bullet$.
If f, $g \in \mathcal{L}^0$, $c \in \mathbb{R}$ then

$$|cf| = |c||f|, \quad f \vee g = \frac{1}{2}(f + g + |f - g|),$$

$$f \wedge g = \frac{1}{2}(f + g - |f - g|), \quad |f + g| \leq_{\text{a.e.}} |f| + |g|,$$

so

$$|cu| = |c||u|, \quad u \vee v = \frac{1}{2}(u + v + |u - v|),$$

$$u \wedge v = \frac{1}{2}(u + v - |u - v|), \quad |u + v| \leq |u| + |v|$$

for all u, $v \in L^0$.

(f) A special notation is often useful. If f is a real-valued function, set $f^+(x) = \max(f(x), 0)$, $f^-(x) = \max(-f(x), 0)$ for $x \in \operatorname{dom} f$, so that

$$f = f^+ - f^-, \quad |f| = f^+ + f^- = f^+ \vee f^-,$$

all these functions being defined on $\operatorname{dom} f$. In L^0, the corresponding operations are $u^+ = u \vee 0$, $u^- = (-u) \vee 0$, and we have

$$u = u^+ - u^-, \quad |u| = u^+ + u^- = u^+ \vee u^-, \quad u^+ \wedge u^- = 0.$$

(g) It is perhaps obvious, but I say it anyway: if $u \geq 0$ in L^0, then there is an $f \geq 0$ in $\mathcal{L}^0$ such that $f^\bullet = u$. **P** Take any $g \in \mathcal{L}^0$ such that $u = g^\bullet$, and set $f = g \vee \mathbf{0}$. **Q**

241F Riesz spaces There is an extensive abstract theory of Riesz spaces, which I think it best to leave aside for the moment; a general account may be found in Luxemburg & Zaanen 71 and Zaanen 83; my own book Fremlin 74 covers the elementary material, and Chapter 35 in the next volume repeats the most essential ideas. For our purposes here we need only a few definitions and some simple results which are most easily proved for the special cases in which we need them, without reference to the general theory.

(a) A Riesz space U is **Archimedean** if whenever $u \in U$, $u > 0$ (that is, $u \geq 0$ and $u \neq 0$), and $v \in U$, there is an $n \in \mathbb{N}$ such that $nu \not\leq v$.

(b) A Riesz space U is **Dedekind σ-complete** (or **σ-order-complete**, or **σ-complete**) if every non-empty countable set $A \subseteq U$ which is bounded above has a least upper bound in U.

(c) A Riesz space U is **Dedekind complete** (or **order complete**, or **complete**) if every non-empty set $A \subseteq U$ which is bounded above in U has a least upper bound in U.

241G Now we have the following important properties of L^0.

Theorem Let (X, Σ, μ) be a measure space. Set $L^0 = L^0(\mu)$.
(a) L^0 is Archimedean and Dedekind σ-complete.
(b) If (X, Σ, μ) is semi-finite, then L^0 is Dedekind complete iff (X, Σ, μ) is localizable.

proof Set $\mathcal{L}^0 = \mathcal{L}^0(\mu)$.

(a)(i) If u, $v \in L^0$ and $u > 0$, express u as $f^\bullet$ and v as $g^\bullet$ where f, $g \in \mathcal{L}^0$. Then $E = \{x : x \in \operatorname{dom} f, f(x) > 0\}$ is not negligible. So there is an $n \in \mathbb{N}$ such that

$$E_n = \{x : x \in \operatorname{dom} f \cap \operatorname{dom} g, nf(x) > g(x)\}$$

is not negligible, since $E \cap \operatorname{dom} g \subseteq \bigcup_{n \in \mathbb{N}} E_n$. But now $nu \not\leq v$. As u and v are arbitrary, L^0 is Archimedean.

(ii) Now let $A \subseteq L^0$ be a non-empty countable set with an upper bound w in L^0. Express A as $\{f_n^\bullet : n \in \mathbb{N}\}$ where $\langle f_n \rangle_{n \in \mathbb{N}}$ is a sequence in $\mathcal{L}^0$, and w as $h^\bullet$ where $h \in \mathcal{L}^0$. Set $f = \sup_{n \in \mathbb{N}} f_n$. Then we have $f(x)$ defined in $\mathbb{R}$ at any point $x \in \operatorname{dom} h \cap \bigcap_{n \in \mathbb{N}} \operatorname{dom} f_n$ such that $f_n(x) \leq h(x)$ for every $n \in \mathbb{N}$, that is, for almost every $x \in X$; so $f \in \mathcal{L}^0$ (241Bg). Set $u = f^\bullet \in L^0$. If $v \in L^0$, say $v = g^\bullet$ where $g \in \mathcal{L}^0$, then

$$\begin{aligned} u_n \le v \text{ for every } n \in \mathbb{N} \\ &\iff \text{ for every } n \in \mathbb{N},\ f_n \le_{\text{a.e.}} g \\ &\iff \text{ for almost every } x \in X,\ f_n(x) \le g(x) \text{ for every } n \in \mathbb{N} \\ &\iff f \le_{\text{a.e.}} g \iff u \le v. \end{aligned}$$

Thus $u = \sup_{n\in\mathbb{N}} u_n$ in L^0. As A is arbitrary, L^0 is Dedekind σ-complete.

(b)(i) Suppose that (X, Σ, μ) is localizable. Let $A \subseteq L^0$ be any non-empty set with an upper bound $w_0 \in L^0$. Set

$$\mathcal{A} = \{f : f \text{ is a measurable function from } X \text{ to } \mathbb{R},\ f^\bullet \in A\};$$

then every member of A is of the form $f^\bullet$ for some $f \in \mathcal{A}$ (241Bk). For each $q \in \mathbb{Q}$, let $\mathcal{E}_q$ be the family of subsets of X expressible in the form $\{x : f(x) \ge q\}$ for some $f \in \mathcal{A}$; then $\mathcal{E}_q \subseteq \Sigma$. Because (X, Σ, μ) is localizable, there is a set $F_q \in \Sigma$ which is an essential supremum for $\mathcal{E}_q$. For $x \in X$, set

$$g^*(x) = \sup\{q : q \in \mathbb{Q},\ x \in F_q\},$$

allowing ∞ as the supremum of a set which is not bounded above, and $-\infty$ as $\sup\emptyset$. Then

$$\{x : g^*(x) > a\} = \bigcup_{q\in\mathbb{Q},q>a} F_q \in \Sigma$$

for every $a \in \mathbb{R}$.

If $f \in \mathcal{A}$, then $f \le_{\text{a.e.}} g^*$. **P** For each $q \in \mathbb{Q}$, set

$$E_q = \{x : f(x) \ge q\} \in \mathcal{E}_q;$$

then $E_q \setminus F_q$ is negligible. Set $H = \bigcup_{q\in\mathbb{Q}}(E_q \setminus F_q)$. If $x \in X \setminus H$, then

$$f(x) \ge q \Longrightarrow g^*(x) \ge q,$$

so $f(x) \le g^*(x)$; thus $f \le_{\text{a.e.}} g^*$. **Q**

If $h : X \to \mathbb{R}$ is measurable and $u \le h^\bullet$ for every $u \in A$, then $g^* \le_{\text{a.e.}} h$. **P** Set $G_q = \{x : h(x) \ge q\}$ for each $q \in \mathbb{Q}$. If $E \in \mathcal{E}_q$, there is an $f \in \mathcal{A}$ such that $E = \{x : f(x) \ge q\}$; now $f \le_{\text{a.e.}} h$, so $E \setminus G_q \subseteq \{x : f(x) > h(x)\}$ is negligible. Because F_q is an essential supremum for $\mathcal{E}_q$, $F_q \setminus G_q$ is negligible; and this is true for every $q \in \mathbb{Q}$. Consequently

$$\{x : h(x) < g^*(x)\} \subseteq \bigcup_{q\in\mathbb{Q}} F_q \setminus G_q$$

is negligible, and $g^* \le_{\text{a.e.}} h$. **Q**

Now recall that we are assuming that $A \ne \emptyset$ and that A has an upper bound $w_0 \in L^0$. Take any $f_0 \in \mathcal{A}$ and a measurable $h_0 : X \to \mathbb{R}$ such that $h_0^\bullet = w_0$; then $f \le_{\text{a.e.}} h_0$ for every $f \in \mathcal{A}$, so $f_0 \le_{\text{a.e.}} g^* \le_{\text{a.e.}} h_0$, and g^* must be finite a.e. Setting $g(x) = g^*(x)$ when $g^*(x) \in \mathbb{R}$, we have $g \in \mathcal{L}^0$ and $g =_{\text{a.e.}} g^*$, so that

$$f \le_{\text{a.e.}} g \le_{\text{a.e.}} h$$

whenever f, h are measurable functions from X to $\mathbb{R}$, $f^\bullet \in A$ and $h^\bullet$ is an upper bound for A; that is,

$$u \le g^\bullet \le w$$

whenever $u \in A$ and w is an upper bound for A. But this means that $g^\bullet$ is a least upper bound for A in L^0. As A is arbitrary, L^0 is Dedekind complete.

(ii) Suppose that L^0 is Dedekind complete. We are assuming that (X, Σ, μ) is semi-finite. Let $\mathcal{E}$ be any subset of Σ. Set

$$A = \{0\} \cup \{(\chi E)^\bullet : E \in \mathcal{E}\} \subseteq L^0.$$

Then A is bounded above by $(\chi X)^\bullet$ so has a least upper bound $w \in L^0$. Express w as $h^\bullet$ where $h : X \to \mathbb{R}$ is measurable, and set $F = \{x : h(x) > 0\}$. Then F is an essential supremum for $\mathcal{E}$ in Σ. **P** (α) If $E \in \mathcal{E}$, then $(\chi E)^\bullet \le w$ so $\chi E \le_{\text{a.e.}} h$, that is, $h(x) \ge 1$ for almost every $x \in E$, and $E \setminus F \subseteq \{x : x \in E,\ h(x) < 1\}$ is negligible. (β) If $G \in \Sigma$ and $E \setminus G$ is negligible for every $E \in \mathcal{E}$, then $\chi E \le_{\text{a.e.}} \chi G$ for every $E \in \mathcal{E}$, that is, $(\chi E)^\bullet \le (\chi G)^\bullet$ for every $E \in \mathcal{E}$; so $w \le (\chi G)^\bullet$, that is, $h \le_{\text{a.e.}} \chi G$. Accordingly $F \setminus G \subseteq \{x : h(x) > (\chi G)(x)\}$ is negligible. **Q**

As $\mathcal{E}$ is arbitrary, (X, Σ, μ) is localizable.

241H The multiplicative structure of L^0 Let (X, Σ, μ) be any measure space; write $L^0 = L^0(\mu)$, $\mathcal{L}^0 = \mathcal{L}^0(\mu)$.

(a) If f_1, f_2, g_1, $g_2 \in \mathcal{L}^0$, $f_1 =_{\text{a.e.}} f_2$ and $g_1 =_{\text{a.e.}} g_2$ then $f_1 \times g_1 =_{\text{a.e.}} f_2 \times g_2$. Accordingly we may define multiplication on L^0 by setting $f^\bullet \times g^\bullet = (f \times g)^\bullet$ for all f, $g \in \mathcal{L}^0$.

(b) It is now easy to check that, for all u, v, $w \in L^0$ and $c \in \mathbb{R}$,

$u \times (v \times w) = (u \times v) \times w$,

$u \times e = e \times u = u$,

where $e = \chi X^\bullet$ is the equivalence class of the function with constant value 1,

$c(u \times v) = cu \times v = u \times cv$,

$u \times (v + w) = (u \times v) + (u \times w)$,

$(u + v) \times w = (u \times w) + (v \times w)$,

$u \times v = v \times u$,

$|u \times v| = |u| \times |v|$,

$u \times v = 0$ iff $|u| \wedge |v| = 0$,

$|u| \le |v|$ iff there is a w such that $|w| \le e$ and $u = v \times w$.

241I The action of Borel functions on L^0 Let (X, Σ, μ) be a measure space and $h : \mathbb{R} \to \mathbb{R}$ a Borel measurable function. Then $hf \in \mathcal{L}^0 = \mathcal{L}^0(\mu)$ for every $f \in \mathcal{L}^0$ (241Be) and $hf =_{\text{a.e.}} hg$ whenever $f =_{\text{a.e.}} g$. So we have a function $\bar{h} : L^0 \to L^0$ defined by setting $\bar{h}(f^\bullet) = (hf)^\bullet$ for every $f \in \mathcal{L}^0$. For instance, if $u \in L^0$ and $p \ge 1$, we can consider $|u|^p = \bar{h}(u)$ where $h(x) = |x|^p$ for $x \in \mathbb{R}$.

241J Complex L^0 The ideas of this chapter, like those of Chapters 22-23, are often applied to spaces based on complex-valued functions instead of real-valued functions. Let (X, Σ, μ) be a measure space.

(a) We may write $\mathcal{L}^0_{\mathbb{C}} = \mathcal{L}^0_{\mathbb{C}}(\mu)$ for the space of complex-valued functions f such that $\operatorname{dom} f$ is a conegligible subset of X and there is a conegligible subset $E \subseteq X$ such that $f \restriction E$ is measurable; that is, such that the real and imaginary parts of f both belong to $\mathcal{L}^0(\mu)$. Next, $L^0_{\mathbb{C}} = L^0_{\mathbb{C}}(\mu)$ will be the space of equivalence classes in $\mathcal{L}^0_{\mathbb{C}}$ under the equivalence relation $=_{\text{a.e.}}$.

(b) Using just the same formulae as in 241D, it is easy to describe addition and scalar multiplication rendering $L^0_{\mathbb{C}}$ a linear space over $\mathbb{C}$. We no longer have quite the same kind of order structure, but we can identify a 'real part', being

$$\{f^\bullet : f \in \mathcal{L}^0_{\mathbb{C}} \text{ is real a.e.}\},$$

obviously identifiable with the real linear space L^0, and corresponding maps $u \mapsto \mathcal{R}\mathrm{e}(u)$, $u \mapsto \mathcal{I}\mathrm{m}(u) : L^0_{\mathbb{C}} \to L^0$ such that $u = \mathcal{R}\mathrm{e}(u) + i\,\mathcal{I}\mathrm{m}(u)$ for every u. Moreover, we have a notion of 'modulus', writing

$$|f^\bullet| = |f|^\bullet \in L^0 \text{ for every } f \in \mathcal{L}^0_{\mathbb{C}},$$

satisfying the basic relations $|cu| = |c||u|$, $|u + v| \le |u| + |v|$ for u, $v \in L^0_{\mathbb{C}}$ and $c \in \mathbb{C}$, as in 241Ef. We do of course still have a multiplication on $L^0_{\mathbb{C}}$, for which all the formulae in 241H are still valid.

(c) The following fact is useful. For any $u \in L^0_{\mathbb{C}}$, $|u|$ is the supremum in L^0 of $\{\mathcal{R}\mathrm{e}(\zeta u) : \zeta \in \mathbb{C}, |\zeta| = 1\}$. **P** (i) If $|\zeta| = 1$, then $\mathcal{R}\mathrm{e}(\zeta u) \le |\zeta u| = |u|$. So $|u|$ is an upper bound of $\{\mathcal{R}\mathrm{e}(\zeta u) : |\zeta| = 1\}$. (ii) If $v \in L^0$ and $\mathcal{R}\mathrm{e}(\zeta u) \le v$ whenever $|\zeta| = 1$, then express u, v as $f^\bullet$, $g^\bullet$ where $f : X \to \mathbb{C}$ and $g : X \to \mathbb{R}$ are measurable. For any $q \in \mathbb{Q}$, $x \in X$ set $f_q(x) = \mathcal{R}\mathrm{e}(e^{iqx} f(x))$. Then $f_q \le_{\text{a.e.}} g$. Accordingly $H = \{x : f_q(x) \le g(x) \text{ for every } q \in \mathbb{Q}\}$ is conegligible. But of course $H = \{x : |f(x)| \le g(x)\}$, so $|f| \le_{\text{a.e.}} g$ and $|u| \le v$. As v is arbitrary, $|u|$ is the least upper bound of $\{\mathcal{R}\mathrm{e}(\zeta u) : |\zeta| = 1\}$. **Q**

241X Basic exercises >(a) Let X be a set, and let μ be counting measure on X (112Bd). Show that $L^0(\mu)$ can be identified with $\mathcal{L}^0(\mu) = \mathbb{R}^X$.

>(b) Let (X, Σ, μ) be a measure space and $\hat{\mu}$ the completion of μ. Show that $\mathcal{L}^0(\mu) = \mathcal{L}^0(\hat{\mu})$ and $L^0(\mu) = L^0(\hat{\mu})$.

(c) Let (X, Σ, μ) be a measure space. (i) Show that for every $u \in L^0(\mu)$ we may define an outer measure $\theta_u : \mathcal{P}\mathbb{R} \to [0, \infty]$ by writing $\theta_u(A) = \mu^* f^{-1}[A]$ whenever $A \subseteq \mathbb{R}$ and $f \in \mathcal{L}^0(\mu)$ is such that $f^\bullet = u$. (ii) Show that the measure defined from θ_u by Carathéodory's method measures every Borel subset of $\mathbb{R}$.

(d) Let $\langle (X_i, \Sigma_i, \mu_i) \rangle_{i \in I}$ be a family of measure spaces, with direct sum (X, Σ, μ) (214L). (i) Writing $\phi_i : X_i \to X$ for the canonical maps (in the construction of 214L, $\phi_i(x) = (x, i)$ for $x \in X_i$), show that $f \mapsto \langle f\phi_i \rangle_{i \in I}$ is a bijection between $\mathcal{L}^0(\mu)$ and $\prod_{i \in I} \mathcal{L}^0(\mu_i)$. (ii) Show that this corresponds to a bijection between $L^0(\mu)$ and $\prod_{i \in I} L^0(\mu_i)$.

(e) Let U be a Dedekind σ-complete Riesz space and $A \subseteq U$ a non-empty countable set which is bounded below in U. Show that $\inf A$ is defined in U.

(f) Let U be a Dedekind complete Riesz space and $A \subseteq U$ a non-empty set which is bounded below in U. Show that $\inf A$ is defined in U.

(g) Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and $\phi : X \to Y$ an inverse-measure-preserving function. (i) Show that we have a map $T : L^0(\nu) \to L^0(\mu)$ defined by setting $Tg^\bullet = (g\phi)^\bullet$ for every $g \in \mathcal{L}^0(\nu)$. (ii) Show that T is linear, that $T(v \times w) = Tv \times Tw$ for all v, $w \in L^0(\nu)$, and that $T(\sup_{n\in\mathbb{N}} v_n) = \sup_{n\in\mathbb{N}} Tv_n$ whenever $\langle v_n\rangle_{n\in\mathbb{N}}$ is a sequence in $L^0(\nu)$ with an upper bound in $L^0(\nu)$.

>(h) Let (X,Σ,μ) be a measure space. Suppose that $r \geq 1$ and that $h : \mathbb{R}^r \to \mathbb{R}$ is a Borel measurable function. Show that there is a function $\bar{h} : L^0(\mu)^r \to L^0(\mu)$ defined by writing

$$\bar{h}(f_1^\bullet, \dots, f_r^\bullet) = (h(f_1, \dots, f_r))^\bullet$$

for $f_1, \dots, f_r \in \mathcal{L}^0(\mu)$.

(i) Let (X,Σ,μ) be a measure space and g, h, $\langle g_n\rangle_{n\in\mathbb{N}}$ Borel measurable functions from $\mathbb{R}$ to itself; write $\bar{g}$, $\bar{h}$, $\bar{g}_n$ for the corresponding functions from $L^0 = L^0(\mu)$ to itself (241I). (i) Show that

$$\bar{g}(u) + \bar{h}(u) = \overline{g+h}(u), \quad \bar{g}(u) \times \bar{h}(u) = \overline{g \times h}(u), \quad \bar{g}(\bar{h}(u)) = \overline{gh}(u)$$

for every $u \in L^0$. (ii) Show that if $g(t) \leq h(t)$ for every $t \in \mathbb{R}$, then $\bar{g}(u) \leq \bar{h}(u)$ for every $u \in L^0$. (iii) Show that if g is non-decreasing, then $\bar{g}(u) \leq \bar{g}(v)$ whenever $u \leq v$ in L^0. (iv) Show that if $h(t) = \sup_{n\in\mathbb{N}} g_n(t)$ for every $t \in \mathbb{R}$, then $\bar{h}(u) = \sup_{n\in\mathbb{N}} \bar{g}_n(u)$ in L^0 for every $u \in L^0$.

241Y Further exercises (a) Let U be any Riesz space. For $u \in U$ write $|u| = u \vee (-u)$, $u^+ = u \vee 0$, $u^- = (-u) \vee 0$. Show that, for any u, $v \in U$,

$$u = u^+ - u^-, \quad |u| = u^+ + u^- = u^+ \vee u^-, \quad u^+ \wedge u^- = 0,$$

$$u \vee v = \frac{1}{2}(u + v + |u - v|) = u + (v - u)^+,$$

$$u \wedge v = \frac{1}{2}(u + v - |u - v|) = u - (u - v)^+,$$

$$|u + v| \leq |u| + |v|.$$

(b) Let U be a partially ordered linear space and N a linear subspace of U such that whenever u, $u' \in N$ and $u' \leq v \leq u$ then $v \in N$. (i) Show that the linear space quotient U/N is a partially ordered linear space if we say that $u^\bullet \leq v^\bullet$ in U/N iff there is a $w \in N$ such that $u \leq v + w$ in U. (ii) Show that in this case U/N is a Riesz space if U is a Riesz space and $|u| \in N$ for every $u \in N$.

(c) Let (X,Σ,μ) be a measure space. Write $\mathcal{L}^0_{\text{strict}}$ for the space of all measurable functions from X to $\mathbb{R}$, and $\mathcal{N}$ for the subspace of $\mathcal{L}^0_{\text{strict}}$ consisting of measurable functions which are zero almost everywhere. (i) Show that $\mathcal{L}^0_{\text{strict}}$ is a Dedekind σ-complete Riesz space. (ii) Show that $L^0(\mu)$ can be identified, as ordered linear space, with the quotient $\mathcal{L}^0_{\text{strict}}/\mathcal{N}$ as defined in 241Yb above.

(d) Show that any Dedekind σ-complete Riesz space is Archimedean.

(e) A Riesz space U is said to have the **countable sup property** if for every $A \subseteq U$ with a least upper bound in U, there is a countable $B \subseteq A$ such that $\sup B = \sup A$. Show that if (X,Σ,μ) is a semi-finite measure space, then it is σ-finite iff $L^0(\mu)$ has the countable sup property.

(f) Let (X,Σ,μ) be a measure space and $\tilde{\mu}$ the c.l.d. version of μ (213E). (i) Show that $\mathcal{L}^0(\mu) \subseteq \mathcal{L}^0(\tilde{\mu})$. (ii) Show that this inclusion defines a linear operator $T : L^0(\mu) \to L^0(\tilde{\mu})$ such that $T(u \times v) = Tu \times Tv$ for all u, $v \in L^0(\mu)$. (iii) Show that whenever $v > 0$ in $L^0(\tilde{\mu})$ there is a $u \geq 0$ in $L^0(\mu)$ such that $0 < Tu \leq v$. (iv) Show that $T(\sup A) = \sup T[A]$ whenever $A \subseteq L^0(\mu)$ is a non-empty set with a least upper bound in $L^0(\mu)$. (v) Show that T is injective iff μ is semi-finite. (vi) Show that if μ is localizable, then T is an isomorphism for the linear and order structures of $L^0(\mu)$ and $L^0(\tilde{\mu})$. (*Hint*: 213Hb.)

(g) Let (X,Σ,μ) be a measure space and Y any subset of X; let μ_Y be the subspace measure on Y. (i) Show that $\mathcal{L}^0(\mu_Y) = \{f{\restriction}Y : f \in \mathcal{L}^0(\mu)\}$. (ii) Show that there is a canonical surjection $T : L^0(\mu) \to L^0(\mu_Y)$ defined by setting $T(f^\bullet) = (f{\restriction}Y)^\bullet$ for every $f \in \mathcal{L}^0(\mu)$, which is linear and multiplicative and preserves finite suprema and infima, so that (in particular) $T(|u|) = |Tu|$ for every $u \in L^0(\mu)$. (iii) Show that T is injective iff Y has full outer measure.

(h) Suppose, in 241Yg, that $Y \in \Sigma$. Explain how $L^0(\mu_Y)$ may be identified (as ordered linear space) with the subspace $\{u : u \times \chi(X \setminus Y)^\bullet = 0\}$ of $L^0(\mu)$.

(i) Let (X, Σ, μ) be a measure space, and $h : \mathbb{R} \to \mathbb{R}$ a non-decreasing function which is continuous on the left. Show that if $A \subseteq L^0 = L^0(\mu)$ is a non-empty set with a supremum $v \in L^0$, then $\bar{h}(v) = \sup_{u \in A} \bar{h}(u)$, where $\bar{h} : L^0 \to L^0$ is the function described in 241I.

241 Notes and comments As hinted in 241Ya and 241Yd, the elementary properties of the space L^0 which take up most of this section are strongly interdependent; it is not difficult to develop a theory of 'Riesz algebras' to incorporate the ideas of 241H into the rest. (Indeed, I sketch such a theory in §352 in the next volume.)

If we write $\mathcal{L}^0_{\text{strict}}$ for the space of measurable functions from X to $\mathbb{R}$, then $\mathcal{L}^0_{\text{strict}}$ is also a Dedekind σ-complete Riesz space, and L^0 can be identified with the quotient $\mathcal{L}^0_{\text{strict}}/\mathcal{N}$, writing $\mathcal{N}$ for the set of functions in $\mathcal{L}^0_{\text{strict}}$ which are zero almost everywhere. (To do this properly, we need a theory of quotients of ordered linear spaces; see 241Yb-241Yc above.) Of course $\mathcal{L}^0$, as I define it, is not quite a linear space. I choose the slightly more awkward description of L^0 as a space of equivalence classes in $\mathcal{L}^0$ rather than in $\mathcal{L}^0_{\text{strict}}$ because it frequently happens in practice that a member of L^0 arises from a member of $\mathcal{L}^0$ which is either not defined at every point of the underlying space, or not quite measurable; and to adjust such a function so that it becomes a member of $\mathcal{L}^0_{\text{strict}}$, while trivial, is an arbitrary process which to my mind is liable to distort the true nature of such a construction. Of course the same argument could be used in favour of a slightly larger space, the space $\mathcal{L}^0_\infty$ of μ-virtually measurable $[-\infty, \infty]$-valued functions defined and finite almost everywhere, relying on 135E rather than on 121E-121F. But I maintain that the operation of restricting a function in $\mathcal{L}^0_\infty$ to the set on which it is finite is *not* arbitrary, but canonical and entirely natural.

Reading the exposition above – or, for that matter, scanning the rest of this chapter – you are sure to notice a plethora of $^\bullet$s, adding a distinctive character to the pages which, I expect you will feel, is disagreeable to the eye and daunting, or at any rate wearisome, to the spirit. Many, perhaps most, authors prefer to simplify the typography by using the same symbol for a function in $\mathcal{L}^0$ or $\mathcal{L}^0_{\text{strict}}$ and for its equivalence class in L^0; and indeed it is common to use syntax which does not distinguish between them either, so that an object which has been defined as a member of L^0 will suddenly become a function with actual values at points of the underlying measure space. I prefer to maintain a rigid distinction; you must choose for yourself whether to follow me. Since I have chosen the more cumbersome form, I suppose the burden of proof is on me, to justify my decision. (i) Anyone would agree that there is at least a formal difference between a function and a set of functions. This by itself does not justify insisting on the difference in every sentence; mathematical exposition would be impossible if we always insisted on consistency in such questions as whether (for instance) the number 3 belonging to the set $\mathbb{N}$ of natural numbers is exactly the same object as the number 3 belonging to the set $\mathbb{C}$ of complex numbers, or the ordinal 3. But the difference between an object and a set to which it belongs is a sufficient difference in kind to make any confusion extremely dangerous, and while I agree that you can study this topic without using different symbols for f and $f^\bullet$, I do not think you can ever safely escape a mental distinction for more than a few lines of argument. (ii) As a teacher, I have to say that quite a few students, encountering this material for the first time, are misled by any failure to make the distinction between f and $f^\bullet$ into believing that no distinction need be made; and – as a teacher – I always insist on a student convincing me, by correctly writing out the more pedantic forms of the arguments for a few weeks, that he understands the manipulations necessary, before I allow him to go his own way. (iii) The reason why it *is* possible to evade the distinction in certain types of argument is just that the Dedekind σ-complete Riesz space $\mathcal{L}^0_{\text{strict}}$ parallels the Dedekind σ-complete Riesz space L^0 so closely that any proposition involving only countably many members of these spaces is likely to be valid in one if and only if it is valid in the other. In my view, the implications of this correspondence are at the very heart of measure theory. I prefer therefore to keep it constantly conspicuous, reminding myself through symbolism that every theorem has a Siamese twin, and rising to each challenge to express the twin theorem in an appropriate language. (iv) There are ways in which $\mathcal{L}^0_{\text{strict}}$ and L^0 are actually very different, and many interesting ideas can be expressed only in a language which keeps them clearly separated.

For more than half my life now I have felt that these points between them are sufficient reason for being consistent in maintaining the formal distinction between f and $f^\bullet$. You may feel that in (iii) and (iv) of the last paragraph I am trying to have things both ways; I am arguing that both the similarities and the differences between L^0 and $\mathcal{L}^0$ support my case. Indeed that is exactly my position. If they were totally different, using the same language for both would not give rise to confusion; if they were essentially the same, it would not matter if we were sometimes unclear which we were talking about.

242 L^1

While the space L^0 treated in the previous section is of very great intrinsic interest, its chief use in the elementary theory is as a space in which some of the most important spaces of functional analysis are embedded. In the next few sections I introduce these one at a time.

The first is the space L^1 of equivalence classes of integrable functions. The importance of this space is not only that it offers a language in which to express those many theorems about integrable functions which do not depend on the differences between two functions which are equal almost everywhere. It can also appear as the natural space in which to seek solutions to a wide variety of integral equations, and as the completion of a space of continuous functions.

242A The space L^1 Let (X,Σ,μ) be any measure space.

(a) Let $\mathcal{L}^1=\mathcal{L}^1(\mu)$ be the set of real-valued functions, defined on subsets of X, which are integrable over X. Then $\mathcal{L}^1\subseteq\mathcal{L}^0=\mathcal{L}^0(\mu)$, as defined in §241, and, for $f\in\mathcal{L}^0$, we have $f\in\mathcal{L}^1$ iff there is a $g\in\mathcal{L}^1$ such that $|f|\le_{\text{a.e.}}g$; if $f\in\mathcal{L}^1$, $g\in\mathcal{L}^0$ and $f=_{\text{a.e.}}g$, then $g\in\mathcal{L}^1$. (See 122P-122R.)

(b) Let $L^1=L^1(\mu)\subseteq L^0=L^0(\mu)$ be the set of equivalence classes of members of $\mathcal{L}^1$. If f, $g\in\mathcal{L}^1$ and $f=_{\text{a.e.}}g$ then $\int f=\int g$ (122Rb). Accordingly we may define a functional $\int$ on L^1 by writing $\int f^\bullet=\int f$ for every $f\in\mathcal{L}^1$.

(c) It will be convenient to be able to write $\int_A u$ for $u\in L^1$, $A\subseteq X$; this may be defined by saying that $\int_A f^\bullet=\int_A f$ for every $f\in\mathcal{L}^1$, where the integral is defined in 214D. **P** I have only to check that if $f=_{\text{a.e.}}g$ then $\int_A f=\int_A g$; and this is because $f\restriction A=g\restriction A$ almost everywhere in A. **Q**

If $E\in\Sigma$ and $u\in L^1$ then $\int_E u=\int u\times(\chi E)^\bullet$; this is because $\int_E f=\int f\times\chi E$ for every integrable function f (131Fa).

(d) If $u\in L^1$, there is a Σ-measurable, μ-integrable function $f:X\to\mathbb{R}$ such that $f^\bullet=u$. **P** As noted in 241Bk, there is a measurable $f:X\to\mathbb{R}$ such that $f^\bullet=u$; but of course f is integrable because it is equal almost everywhere to some integrable function. **Q**

242B Theorem Let (X,Σ,μ) be any measure space. Then $L^1(\mu)$ is a linear subspace of $L^0(\mu)$ and $\int:L^1\to\mathbb{R}$ is a linear functional.

proof If u, $v\in L^1=L^1(\mu)$ and $c\in\mathbb{R}$ let f, g be integrable functions such that $u=f^\bullet$ and $v=g^\bullet$; then $f+g$ and cf are integrable, so $u+v=(f+g)^\bullet$ and $cu=(cf)^\bullet$ belong to L^1. Also

$$\int u+v=\int f+g=\int f+\int g=\int u+\int v$$

and

$$\int cu=\int cf=c\int f=c\int u.$$

242C The order structure of L^1 Let (X,Σ,μ) be any measure space.

(a) $L^1=L^1(\mu)$ has an order structure derived from that of $L^0=L^0(\mu)$ (241E); that is, $f^\bullet\le g^\bullet$ iff $f\le g$ a.e. Being a linear subspace of L^0, L^1 must be a partially ordered linear space; the two conditions of 241Ec are obviously inherited by linear subspaces.

Note also that if u, $v\in L^1$ and $u\le v$ then $\int u\le\int v$, because if f, g are integrable functions and $f\le_{\text{a.e.}}g$ then $\int f\le\int g$ (122Od).

(b) If $u\in L^0$, $v\in L^1$ and $|u|\le|v|$ then $u\in L^1$. **P** Let $f\in\mathcal{L}^0=\mathcal{L}^0(\mu)$, $g\in\mathcal{L}^1=\mathcal{L}^1(\mu)$ be such that $u=f^\bullet$ and $v=g^\bullet$; then g is integrable and $|f|\le_{\text{a.e.}}|g|$, so f is integrable and $u\in L^1$. **Q**

(c) In particular, $|u|\in L^1$ whenever $u\in L^1$, and

$$|\int u|=\max(\int u,\int(-u))\le\int|u|,$$

because u, $-u\le|u|$.

(d) Because $|u|\in L^1$ for every $u\in L^1$,

$$u\vee v=\frac{1}{2}(u+v+|u-v|),\quad u\wedge v=\frac{1}{2}(u+v-|u-v|)$$

belong to L^1 for all u, $v \in L^1$. But if $w \in L^1$ we surely have

$$w \le u \ \&\ w \le v \iff w \le u \wedge v,$$

$$w \ge u \ \&\ w \ge v \iff w \ge u \vee v$$

because these are true for all $w \in L^0$, so $u \vee v = \sup\{u, v\}$ and $u \wedge v = \inf\{u, v\}$ in L^1. Thus L^1 is, in itself, a Riesz space.

(e) Note that if $u \in L^1$, then $u \ge 0$ iff $\int_E u \ge 0$ for every $E \in \Sigma$; this is because if f is an integrable function on X and $\int_E f \ge 0$ for every $E \in \Sigma$, then $f \ge 0$ a.e. (131Fb). More generally, if u, $v \in L^1$ and $\int_E u \le \int_E v$ for every $E \in \Sigma$, then $u \le v$. It follows at once that if u, $v \in L^1$ and $\int_E u = \int_E v$ for every $E \in \Sigma$, then $u = v$ (cf. 131Fc).

(f) If $u \ge 0$ in L^1, there is a non-negative $f \in \mathcal{L}^1$ such that $f^\bullet = u$ (compare 241Eg).

242D The norm of L^1 Let (X, Σ, μ) be any measure space.

(a) For $f \in \mathcal{L}^1 = \mathcal{L}^1(\mu)$ I write $\|f\|_1 = \int |f| \in [0, \infty[$. For $u \in L^1 = L^1(\mu)$ set $\|u\|_1 = \int |u|$, so that $\|f^\bullet\|_1 = \|f\|_1$ for every $f \in \mathcal{L}^1$. Then $\| \ \|_1$ is a norm on L^1. **P** (i) If u, $v \in L^1$ then $|u + v| \le |u| + |v|$, by 241Ee, so

$$\|u + v\|_1 = \int |u + v| \le \int |u| + |v| = \int |u| + \int |v| = \|u\|_1 + \|v\|_1.$$

(ii) If $u \in L^1$ and $c \in \mathbb{R}$ then

$$\|cu\|_1 = \int |cu| = \int |c||u| = |c| \int |u| = |c|\|u\|_1.$$

(iii) If $u \in L^1$ and $\|u\|_1 = 0$, express u as $f^\bullet$, where $f \in \mathcal{L}^1$; then $\int |f| = \int |u| = 0$. Because $|f|$ is non-negative, it must be zero almost everywhere (122Rc), so $f = 0$ a.e. and $u = 0$ in L^1. **Q**

(b) Thus L^1, with $\| \ \|_1$, is a normed space and $\int : L^1 \to \mathbb{R}$ is a linear operator; observe that $\| \int \| \le 1$, because

$$|\int u| \le \int |u| = \|u\|_1$$

for every $u \in L^1$.

(c) If u, $v \in L^1$ and $|u| \le |v|$, then

$$\|u\|_1 = \int |u| \le \int |v| = \|v\|_1.$$

In particular, $\|u\|_1 = \||u|\|_1$ for every $u \in L^1$.

(d) Note the following property of the normed Riesz space L^1: if u, $v \in L^1$ and u, $v \ge 0$, then

$$\|u + v\|_1 = \int u + v = \int u + \int v = \|u\|_1 + \|v\|_1.$$

(e) The set $(L^1)^+ = \{u : u \ge 0\}$ is closed in L^1. **P** If $v \in L^1$, $u \in (L^1)^+$ then $\|u - v\|_1 \ge \|v \wedge 0\|_1$; this is because if f, $g \in \mathcal{L}^1$ and $f \ge 0$ a.e., $|f(x) - g(x)| \ge |\min(g(x), 0)|$ whenever $f(x)$ and $g(x)$ are both defined and $f(x) \ge 0$, which is almost everywhere, so

$$\|u - v\|_1 = \int |f - g| \ge \int |g \wedge \mathbf{0}| = \|v \wedge 0\|_1.$$

Now this means that if $v \in L^1$ and $v \not\ge 0$, the ball $\{w : \|w - v\|_1 < \delta\}$ does not meet $(L^1)^+$, where $\delta = \|v \wedge 0\|_1 > 0$ because $v \wedge 0 \ne 0$. Thus $L^1 \setminus (L^1)^+$ is open and $(L^1)^+$ is closed. **Q**

242E For the next result we need a variant of B.Levi's theorem.

Lemma Let (X, Σ, μ) be a measure space and $\langle f_n \rangle_{n \in \mathbb{N}}$ a sequence of μ-integrable real-valued functions such that $\sum_{n=0}^\infty \int |f_n| < \infty$. Then $f = \sum_{n=0}^\infty f_n$ is integrable and

$$\int f = \sum_{n=0}^\infty \int f_n, \quad \int |f| \le \sum_{n=0}^\infty \int |f_n|.$$

proof (a) Suppose first that every f_n is non-negative. Set $g_n = \sum_{k=0}^n f_k$ for each n; then $\langle g_n \rangle_{n \in \mathbb{N}}$ is increasing a.e. and

$$\lim_{n \to \infty} \int g_n = \sum_{k=0}^\infty \int f_k$$

is finite, so by B.Levi's theorem (123A) $f = \lim_{n\to\infty} g_n$ is integrable and

$$\int f = \lim_{n\to\infty} \int g_n = \sum_{k=0}^{\infty} \int f_k.$$

In this case, of course,

$$\int |f| = \int f = \sum_{n=0}^{\infty} \int f_n = \sum_{n=0}^{\infty} \int |f_n|.$$

(b) For the general case, set $f_n^+ = \frac{1}{2}(|f_n| + f_n)$, $f_n^- = \frac{1}{2}(|f_n| - f_n)$, as in 241Ef; then f_n^+ and f_n^- are non-negative integrable functions, and

$$\sum_{n=0}^{\infty} \int f_n^+ + \sum_{n=0}^{\infty} \int f_n^- = \sum_{n=0}^{\infty} \int |f_n| < \infty.$$

So $h_1 = \sum_{n=0}^{\infty} f_n^+$ and $h_2 = \sum_{n=0}^{\infty} f_n^-$ are both integrable. Now $f =_{\text{a.e.}} h_1 - h_2$, so

$$\int f = \int h_1 - \int h_2 = \sum_{n=0}^{\infty} \int f_n^+ - \sum_{n=0}^{\infty} \int f_n^- = \sum_{n=0}^{\infty} \int f_n.$$

Finally

$$\int |f| \le \int |h_1| + \int |h_2| = \sum_{n=0}^{\infty} \int f_n^+ + \sum_{n=0}^{\infty} \int f_n^- = \sum_{n=0}^{\infty} \int |f_n|.$$

242F Theorem For any measure space (X, Σ, μ), $L^1(\mu)$ is complete under its norm $\| \,\|_1$.

proof Let $\langle u_n \rangle_{n\in\mathbb{N}}$ be a sequence in L^1 such that $\|u_{n+1} - u_n\|_1 \le 4^{-n}$ for every $n \in \mathbb{N}$. Choose integrable functions f_n such that $f_0^\bullet = u_0$, $f_{n+1}^\bullet = u_{n+1} - u_n$ for each $n \in \mathbb{N}$. Then

$$\sum_{n=0}^{\infty} \int |f_n| = \|u_0\|_1 + \sum_{n=0}^{\infty} \|u_{n+1} - u_n\|_1 < \infty.$$

So $f = \sum_{n=0}^{\infty} f_n$ is integrable, by 242E, and $u = f^\bullet \in L^1$. Set $g_n = \sum_{j=0}^{n} f_j$ for each n; then $g_n^\bullet = u_n$, so

$$\|u - u_n\|_1 = \int |f - g_n| \le \int \sum_{j=n+1}^{\infty} |f_j| \le \sum_{j=n+1}^{\infty} 4^{-j} = 4^{-n}/3$$

for each n. Thus $u = \lim_{n\to\infty} u_n$ in L^1. As $\langle u_n \rangle_{n\in\mathbb{N}}$ is arbitrary, L^1 is complete (2A4E).

242G Definition It will be convenient, for later reference, to introduce the following phrase. A **Banach lattice** is a Riesz space U together with a norm $\| \,\|$ on U such that (i) $\|u\| \le \|v\|$ whenever u, $v \in U$ and $|u| \le |v|$, writing $|u|$ for $u \vee (-u)$, as in 241Ee (ii) U is complete under $\| \,\|$. Thus 242Dc and 242F amount to saying that the normed Riesz space $(L^1, \| \,\|_1)$ is a Banach lattice.

242H L^1 as a Riesz space We can discuss the ordered linear space L^1 in the language already used in 241E-241G for L^0.

Theorem Let (X, Σ, μ) be any measure space. Then $L^1 = L^1(\mu)$ is Dedekind complete.

proof (a) Let $A \subseteq L^1$ be any non-empty set which is bounded above in L^1. Set

$$A' = \{u_0 \vee \ldots \vee u_n : u_0, \ldots , u_n \in A\}.$$

Then $A \subseteq A'$, A' has the same upper bounds as A and $u \vee v \in A'$ for all u, $v \in A'$. Taking w_0 to be any upper bound of A and A', we have $\int u \le \int w_0$ for every $u \in A'$, so $\gamma = \sup_{u\in A'} \int u$ is defined in $\mathbb{R}$. For each $n \in \mathbb{N}$, choose $u_n \in A'$ such that $\int u_n \ge \gamma - 2^{-n}$. Because $L^0 = L^0(\mu)$ is Dedekind σ-complete (241Ga), $u^* = \sup_{n\in\mathbb{N}} u_n$ is defined in L^0, and $u_0 \le u^* \le w_0$ in L^0. Consequently

$$0 \le u^* - u_0 \le w_0 - u_0$$

in L^0. But $w_0 - u_0 \in L^1$, so $u^* - u_0 \in L^1$ (242Cb) and $u^* \in L^1$.

(b) The point is that u^* is an upper bound for A. **P** If $u \in A$, then $u \vee u_n \in A'$ for every n, so

$$\|u - u \wedge u^*\|_1 = \int u - u \wedge u^* \le \int u - u \wedge u_n$$

(because $u \wedge u_n \le u_n \le u^*$, so $u \wedge u_n \le u \wedge u^*$)

$$= \int u \vee u_n - u_n$$

(because $u \vee u_n + u \wedge u_n = u + u_n$ – see the formulae in 242Cd)

$$= \int u \vee u_n - \int u_n \le \gamma - (\gamma - 2^{-n}) = 2^{-n}$$

for every n; so $\|u - u \wedge u^*\|_1 = 0$. But this means that $u = u \wedge u^*$, that is, that $u \le u^*$. As u is arbitrary, u^* is an upper bound for A. **Q**

(c) On the other hand, any upper bound for A is surely an upper bound for $\{u_n : n \in \mathbb{N}\}$, so is greater than or equal to u^*. Thus $u^* = \sup A$ in L^1. As A is arbitrary, L^1 is Dedekind complete.

Remark Note that the order-completeness of L^1, unlike that of L^0, does not depend on any particular property of the measure space (X, Σ, μ).

242I The Radon-Nikodým theorem I think it is worth re-writing the Radon-Nikodým theorem (232E) in the language of this chapter.

Theorem Let (X, Σ, μ) be a measure space. Then there is a canonical bijection between $L^1 = L^1(\mu)$ and the set of truly continuous additive functionals $\nu : \Sigma \to \mathbb{R}$, given by the formula

$$\nu F = \int_F u \text{ for } F \in \Sigma,\ u \in L^1.$$

Remark Recall that if μ is σ-finite, then the truly continuous additive functionals are just the absolutely continuous countably additive functionals; and that if μ is totally finite, then all absolutely continuous (finitely) additive functionals are truly continuous (232Bd).

proof For $u \in L^1$, $F \in \Sigma$ set $\nu_u F = \int_F u$. If $u \in L^1$, there is an integrable function f such that $f^\bullet = u$, in which case

$$F \mapsto \nu_u F = \int_F f : \Sigma \to \mathbb{R}$$

is additive and truly continuous, by 232D. If $\nu : \Sigma \to \mathbb{R}$ is additive and truly continuous, then by 232E there is an integrable function f such that $\nu F = \int_F f$ for every $F \in \Sigma$; setting $u = f^\bullet$ in L^1, $\nu = \nu_u$. Finally, if u, v are distinct members of L^1, there is an $F \in \Sigma$ such that $\int_F u \ne \int_F v$ (242Ce), so that $\nu_u \ne \nu_v$; thus $u \mapsto \nu_u$ is injective as well as surjective.

242J Conditional expectations revisited We now have the machinery necessary for a new interpretation of some of the ideas of §233.

(a) Let (X, Σ, μ) be a measure space, and T a σ-subalgebra of Σ, as in 233A. Then $(X, \mathrm{T}, \mu\restriction\mathrm{T})$ is a measure space, and $\mathcal{L}^0(\mu\restriction\mathrm{T}) \subseteq \mathcal{L}^0(\mu)$; moreover, if f, $g \in \mathcal{L}^0(\mu\restriction\mathrm{T})$, then $f = g$ $(\mu\restriction\mathrm{T})$-a.e. iff $f = g$ μ-a.e. **P** There are $\mu\restriction\mathrm{T}$-conegligible sets F, $G \in \mathrm{T}$ such that $f\restriction F$ and $g\restriction G$ are T-measurable; set

$$E = \{x : x \in F \cap G,\ f(x) \ne g(x)\} \in \mathrm{T};$$

then

$$f = g \ (\mu\restriction\mathrm{T})\text{-a.e.} \iff (\mu\restriction\mathrm{T})(E) = 0 \iff \mu E = 0 \iff f = g \ \mu\text{-a.e.} \ \mathbf{Q}$$

Accordingly we have a canonical map $S : L^0(\mu\restriction\mathrm{T}) \to L^0(\mu)$ defined by saying that if $u \in L^0(\mu\restriction\mathrm{T})$ is the equivalence class of $f \in \mathcal{L}^0(\mu\restriction\mathrm{T})$, then Su is the equivalence class of f in $L^0(\mu)$. It is easy to check, working through the operations described in 241D, 241E and 241H, that S is linear, injective and order-preserving, and that $|Su| = S|u|$, $S(u \vee v) = Su \vee Sv$ and $S(u \times v) = Su \times Sv$ for u, $v \in L^0(\mu\restriction\mathrm{T})$.

(b) Next, if $f \in \mathcal{L}^1(\mu\restriction\mathrm{T})$, then $f \in \mathcal{L}^1(\mu)$ and $\int f d\mu = \int f d(\mu\restriction\mathrm{T})$ (233B); so $Su \in L^1(\mu)$ and $\|Su\|_1 = \|u\|_1$ for every $u \in L^1(\mu\restriction\mathrm{T})$.

Observe also that every member of $L^1(\mu) \cap S[L^0(\mu\restriction\mathrm{T})]$ is actually in $S[L^1(\mu\restriction\mathrm{T})]$. **P** Take $u \in L^1(\mu) \cap S[L^0(\mu\restriction\mathrm{T})]$. Then u is expressible both as $f^\bullet$ where $f \in \mathcal{L}^1(\mu)$, and as $g^\bullet$ where $g \in \mathcal{L}^0(\mu\restriction\mathrm{T})$. So $g =_{\text{a.e.}} f$, and g is μ-integrable, therefore $(\mu\restriction\mathrm{T})$-integrable (233B again). **Q**

This means that $S : L^1(\mu\restriction\mathrm{T}) \to L^1(\mu) \cap S[L^0(\mu\restriction\mathrm{T})]$ is a bijection.

(c) Now suppose that $\mu X = 1$, so that (X, Σ, μ) is a probability space. Recall that g is a conditional expectation of f on T if g is $\mu\restriction\mathrm{T}$-integrable, f is μ-integrable and $\int_F g = \int_F f$ for every $F \in \mathrm{T}$; and that every μ-integrable function has such a conditional expectation (233D). If g is a conditional expectation of f and $f_1 = f$ μ-a.e. then g is a conditional expectation of f_1, because $\int_F f_1 = \int_F f$ for every F; and I have already remarked in 233Dc that if g, g_1 are conditional expectations of f on T then $g = g_1$ $\mu\restriction\mathrm{T}$-a.e.

(d) This means that we have an operator $P : L^1(\mu) \to L^1(\mu\restriction\mathrm{T})$ defined by saying that $P(f^\bullet) = g^\bullet$ whenever $g \in \mathcal{L}^1(\mu\restriction\mathrm{T})$ is a conditional expectation of $f \in \mathcal{L}^1(\mu)$ on T; that is, that $\int_F Pu = \int_F u$ whenever $u \in L^1(\mu)$ and $F \in \mathrm{T}$. If we identify $L^1(\mu)$, $L^1(\mu\restriction\mathrm{T})$ with the sets of absolutely continuous additive functionals defined on Σ and T, as in 242I, then P corresponds to the operation $\nu \mapsto \nu\restriction\mathrm{T}$.

(e) Because Pu is uniquely defined in $L^1(\mu\restriction\mathrm{T})$ by the requirement $\int_F Pu = \int_F u$ for every $F \in \mathrm{T}$ (242Ce), we see that P must be linear. **P** If u, $v \in L^1(\mu)$ and $c \in \mathbb{R}$, then

$$\int_F Pu + Pv = \int_F Pu + \int_F Pv = \int_F u + \int_F v = \int_F u + v = \int_F P(u+v),$$

$$\int_F P(cu) = \int_F cu = c\int_F u = c\int_F Pu = \int_F cPu$$

for every $F \in \mathrm{T}$. **Q** Also, if $u \ge 0$, then $\int_F Pu = \int_F u \ge 0$ for every $F \in \mathrm{T}$, so $Pu \ge 0$ (242Ce again).

It follows at once that P is order-preserving, that is, that $Pu \le Pv$ whenever $u \le v$. Consequently

$$|Pu| = Pu \vee (-Pu) = Pu \vee P(-u) \le P|u|$$

for every $u \in L^1(\mu)$, because $u \le |u|$ and $-u \le |u|$. Finally, P is a bounded linear operator, with norm 1. **P** The last formula tells us that

$$\|Pu\|_1 \le \|P|u|\|_1 = \int P|u| = \int |u| = \|u\|_1$$

for every $u \in L^1(\mu)$, so $\|P\| \le 1$. On the other hand, $P(\chi X^\bullet) = \chi X^\bullet \ne 0$, so $\|P\| = 1$. **Q**

(f) We may legitimately regard $Pu \in L^1(\mu\restriction\mathrm{T})$ as 'the' conditional expectation of $u \in L^1(\mu)$ on T; P is the **conditional expectation operator**.

(g) If $u \in L^1(\mu\restriction\mathrm{T})$, then we have a corresponding $Su \in L^1(\mu)$, as in (b); now $PSu = u$. **P** $\int_F PSu = \int_F Su = \int_F u$ for every $F \in \mathrm{T}$. **Q** Consequently $SPSP = SP : L^1(\mu) \to L^1(\mu)$.

(h) The distinction drawn above between $u = f^\bullet \in L^0(\mu\restriction\mathrm{T})$ and $Su = f^\bullet \in L^0(\mu)$ is of course pedantic. I believe it is necessary to be aware of such distinctions, even though for nearly all purposes it is safe as well as convenient to regard $L^0(\mu\restriction\mathrm{T})$ as actually a subset of $L^0(\mu)$. If we do so, then (b) tells us that we can identify $L^1(\mu\restriction\mathrm{T})$ with $L^1(\mu) \cap L^0(\mu\restriction\mathrm{T})$, while (g) becomes '$P^2 = P$'.

242K The language just introduced allows the following re-formulations of 233J-233K.

Theorem Let (X, Σ, μ) be a probability space and T a σ-subalgebra of Σ. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a convex function and $\bar\phi : L^0(\mu) \to L^0(\mu)$ the corresponding operator defined by setting $\bar\phi(f^\bullet) = (\phi f)^\bullet$ (241I). If $P : L^1(\mu) \to L^1(\mu\restriction\mathrm{T})$ is the conditional expectation operator, then $\bar\phi(Pu) \le P(\bar\phi u)$ whenever $u \in L^1(\mu)$ is such that $\bar\phi(u) \in L^1(\mu)$.

proof This is just a restatement of 233J.

242L Proposition Let (X, Σ, μ) be a probability space, and T a σ-subalgebra of Σ. Let $P : L^1(\mu) \to L^1(\mu\restriction\mathrm{T})$ be the corresponding conditional expectation operator. If $u \in L^1 = L^1(\mu)$ and $v \in L^0(\mu\restriction\mathrm{T})$, then $u \times v \in L^1$ iff $P|u| \times v \in L^1$, and in this case $P(u \times v) = Pu \times v$; in particular, $\int u \times v = \int Pu \times v$.

proof (I am here using the identification of $L^0(\mu\restriction\mathrm{T})$ as a subspace of $L^0(\mu)$, as suggested in 242Jh.) Express u as $f^\bullet$ and v as $h^\bullet$, where $f \in \mathcal{L}^1 = \mathcal{L}^1(\mu)$ and $h \in \mathcal{L}^0(\mu\restriction\mathrm{T})$. Let g, $g_0 \in \mathcal{L}^1(\mu\restriction\mathrm{T})$ be conditional expectations of f, $|f|$ respectively, so that $Pu = g^\bullet$ and $P|u| = g_0^\bullet$. Then, using 233K,

$$u \times v \in L^1 \iff f \times h \in \mathcal{L}^1 \iff g_0 \times h \in \mathcal{L}^1 \iff P|u| \times v \in L^1,$$

and in this case $g \times h$ is a conditional expectation of $f \times h$, that is, $Pu \times v = P(u \times v)$.

242M L^1 as a completion I mentioned in the introduction to this section that L^1 appears in functional analysis as a completion of some important spaces; put another way, some dense subspaces of L^1 are significant. The first is elementary.

Proposition Let (X, Σ, μ) be any measure space, and write $\mathcal{S}$ for the space of μ-simple functions on X. Then

(a) whenever f is a μ-integrable real-valued function and $\epsilon > 0$, there is an $h \in \mathcal{S}$ such that $\int |f - h| \le \epsilon$;

(b) $S = \{f^\bullet : f \in \mathcal{S}\}$ is a dense linear subspace of $L^1 = L^1(\mu)$.

proof (a)(i) If f is non-negative, then there is a simple function h such that $h \leq_{\text{a.e.}} f$ and $\int h \geq \int f - \frac{1}{2}\epsilon$ (122K), in which case

$$\int |f-h| = \int f-h = \int f - \int h \leq \frac{1}{2}\epsilon.$$

(ii) In the general case, f is expressible as a difference $f_1 - f_2$ of non-negative integrable functions. Now there are h_1, $h_2 \in \mathcal{S}$ such that $\int |f_j - h_j| \leq \frac{1}{2}\epsilon$ for both j and

$$\int |f-h| \leq \int |f_1 - h_1| + \int |f_2 - h_2| \leq \epsilon.$$

(b) Because $\mathcal{S}$ is a linear subspace of $\mathbb{R}^X$ included in $\mathcal{L}^1 = \mathcal{L}^1(\mu)$, S is a linear subspace of L^1. If $u \in L^1$ and $\epsilon > 0$, there are an $f \in \mathcal{L}^1$ such that $f^\bullet = u$ and an $h \in \mathcal{S}$ such that $\int |f-h| \leq \epsilon$; now $v = h^\bullet \in S$ and

$$\|u-v\|_1 = \int |f-h| \leq \epsilon.$$

As u and ϵ are arbitrary, S is dense in L^1.

242N As always, Lebesgue measure on $\mathbb{R}^r$ and its subsets is by far the most important example; and in this case we have further classes of dense subspace of L^1. If you have reached this point without yet troubling to master multi-dimensional Lebesgue measure, just take $r = 1$. If you feel uncomfortable with general subspace measures, take X to be $\mathbb{R}^r$ or $[0,1] \subseteq \mathbb{R}$ or some other particular subset which you find interesting. The following term will be useful.

Definition If f is a real- or complex-valued function defined on a subset of $\mathbb{R}^r$, say that the **support** of f is $\{x : x \in \operatorname{dom} f,\ f(x) \neq 0\}$.

242O Theorem Let X be any subset of $\mathbb{R}^r$, where $r \geq 1$, and let μ be Lebesgue measure on X, that is, the subspace measure on X induced by Lebesgue measure on $\mathbb{R}^r$. Write C_k for the space of bounded continuous functions $f : \mathbb{R}^r \to \mathbb{R}$ which have bounded support, and S_0 for the space of linear combinations of functions of the form χI where $I \subseteq \mathbb{R}^r$ is a bounded half-open interval. Then

(a) whenever $f \in \mathcal{L}^1 = \mathcal{L}^1(\mu)$ and $\epsilon > 0$, there are $g \in C_k$, $h \in S_0$ such that $\int_X |f-g| \leq \epsilon$ and $\int_X |f-h| \leq \epsilon$;

(b) $\{(g{\restriction}X)^\bullet : g \in C_k\}$ and $\{(h{\restriction}X)^\bullet : h \in S_0\}$ are dense linear subspaces of $L^1 = L^1(\mu)$.

Remark Of course there is a redundant 'bounded' in the description of C_k; see 242Xh.

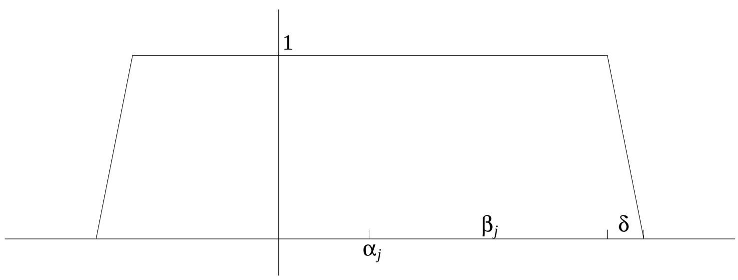
proof (a) I argue in turn that the result is valid for each of an increasing number of members f of $\mathcal{L}^1 = \mathcal{L}^1(\mu)$. Write μ_r for Lebesgue measure on $\mathbb{R}^r$, so that μ is the subspace measure $(\mu_r)_X$.

(i) Suppose first that $f = \chi I{\restriction}X$ where $I \subseteq \mathbb{R}^r$ is a bounded half-open interval. Of course χI is already in S_0, so I have only to show that it is approximated by members of C_k. If $I = \emptyset$ the result is trivial; we can take $g = 0$. Otherwise, express I as $[a-b, a+b[$ where $a = (\alpha_1, \dots, \alpha_r)$, $b = (\beta_1, \dots, \beta_r)$ and $\beta_j > 0$ for each j. Let $\delta > 0$ be such that

$$2^r \prod_{j=1}^r (\beta_j + \delta) \leq \epsilon + 2^r \prod_{j=1}^r \beta_j.$$

For $\xi \in \mathbb{R}$ set

$$\begin{aligned} g_j(\xi) &= 1 \text{ if } |\xi - \alpha_j| \leq \beta_j, \\ &= (\beta_j + \delta - |\xi - \alpha_j|)/\delta \text{ if } \beta_j \leq |\xi - \alpha_j| \leq \beta_j + \delta, \\ &= 0 \text{ if } |\xi - \alpha_j| \geq \beta_j + \delta. \end{aligned}$$



The function g_j

For $x = (\xi_1, \dots, \xi_r) \in \mathbb{R}^r$ set

$$g(x) = \prod_{j=1}^r g_j(\xi_j).$$

Then $g \in C_k$ and $\chi I \le g \le \chi J$, where $J = [a - b - \delta\mathbf{1}, a + b + \delta\mathbf{1}]$ (writing $\mathbf{1} = (1, \dots, 1)$), so that (by the choice of δ) $\mu_r J \le \mu_r I + \epsilon$, and

$$\int_X |g - f| \le \int (\chi(J \cap X) - \chi(I \cap X)) d\mu = \mu((J \setminus I) \cap X)$$
$$\le \mu_r(J \setminus I) = \mu_r J - \mu_r I \le \epsilon,$$

as required.

(ii) Now suppose that $f = \chi(X \cap E)$ where $E \subseteq \mathbb{R}^r$ is a set of finite measure. Then there is a disjoint family $I_0, \dots, I_n$ of half-open intervals such that $\mu_r(E \triangle \bigcup_{j \le n} I_j) \le \frac{1}{2}\epsilon$. **P** There is an open set $G \supseteq E$ such that $\mu_r(G \setminus E) \le \frac{1}{4}\epsilon$ (134Fa). For each $m \in \mathbb{N}$, let $\mathcal{I}_m$ be the family of half-open intervals in $\mathbb{R}^r$ of the form $[a, b[$ where $a = (2^{-m}k_1, \dots, 2^{-m}k_r)$, $k_1, \dots, k_r$ being integers, and $b = a + 2^{-m}\mathbf{1}$; then $\mathcal{I}_m$ is a disjoint family. Set $H_m = \bigcup\{I : I \in \mathcal{I}_m, I \subseteq G\}$; then $\langle H_m \rangle_{m \in \mathbb{N}}$ is a non-decreasing family with union G, so that there is an m such that $\mu_r(G \setminus H_m) \le \frac{1}{4}\epsilon$ and $\mu_r(E \triangle H_m) \le \frac{1}{2}\epsilon$. But now H_m is expressible as a disjoint union $\bigcup_{j \le n} I_j$ where $I_0, \dots, I_n$ enumerate the members of $\mathcal{I}_m$ included in H_m. (The last sentence derails if H_m is empty. But if $H_m = \emptyset$ then we can take $n = 0$, $I_0 = \emptyset$.) **Q**

Accordingly $h = \sum_{j=0}^n \chi I_j \in S_0$ and

$$\int_X |f - h| = \mu(X \cap (E \triangle \textstyle\bigcup_{j \le n} I_j)) \le \frac{1}{2}\epsilon.$$

As for C_k, (i) tells us that there is for each $j \le n$ a $g_j \in C_k$ such that $\int_X |g_j - \chi I_j| \le \epsilon/2(n+1)$, so that $g = \sum_{j=0}^n g_j \in C_k$ and

$$\int_X |f - g| \le \int_X |f - h| + \int_X |h - g| \le \frac{\epsilon}{2} + \sum_{j=0}^n \int_X |g_j - \chi I_j| \le \epsilon.$$

(iii) If f is a simple function, express f as $\sum_{k=0}^n a_k \chi E_k$ where each E_k is of finite measure in X. Each E_k is expressible as $X \cap F_k$ where $\mu_r F_k = \mu E_k$ (214Ca). By (ii), we can find $g_k \in C_k$, $h_k \in S_0$ such that

$$|a_k| \int_X |g_k - \chi F_k| \le \frac{\epsilon}{n+1}, \quad |a_k| \int_X |h_k - \chi F_k| \le \frac{\epsilon}{n+1}$$

for each k. Set $g = \sum_{k=0}^n a_k g_k$ and $h = \sum_{k=0}^n a_k h_k$; then $g \in C_k$, $h \in S_0$ and

$$\int_X |f - g| \le \int_X \sum_{k=0}^n |a_k||\chi F_k - g_k| = \sum_{k=0}^n |a_k| \int_X |\chi F_k - g_k| \le \epsilon,$$

$$\int_X |f - h| \le \sum_{k=0}^n |a_k| \int_X |\chi F_k - h_k| \le \epsilon,$$

as required.

(iv) If f is any integrable function on X, then by 242Ma we can find a simple function f_0 such that $\int |f - f_0| \le \frac{1}{2}\epsilon$, and now by (iii) there are $g \in C_k$, $h \in S_0$ such that $\int_X |f_0 - g| \le \frac{1}{2}\epsilon$, $\int_X |f_0 - h| \le \frac{1}{2}\epsilon$; so that

$$\int_X |f - g| \le \int_X |f - f_0| + \int_X |f_0 - g| \le \epsilon,$$

$$\int_X |f - h| \le \int_X |f - f_0| + \int_X |f_0 - h| \le \epsilon.$$

(b)(i) We must check first that if $g \in C_k$ then $g \restriction X$ is actually μ-integrable. The point here is that if $g \in C_k$ and $a \in \mathbb{R}$ then

$$\{x : x \in X, g(x) > a\}$$

is the intersection of X with an open subset of $\mathbb{R}^r$, and is therefore measured by μ, because all open sets are measured by μ_r (115G). Next, g is bounded and the set $E = \{x : x \in X, g(x) \ne 0\}$ is bounded in $\mathbb{R}^r$, therefore of finite outer measure for μ_r and of finite measure for μ. Thus there is an $M \ge 0$ such that $|g| \le M\chi E$, which is μ-integrable. Accordingly g is μ-integrable.

Of course $h \restriction X$ is μ-integrable for every $h \in S_0$ because (by the definition of subspace measure) $\mu(I \cap X)$ is defined and finite for every bounded half-open interval I.

(ii) Now the rest follows by just the same arguments as in 242Mb. Because $\{g\restriction X : g \in C_k\}$ and $\{h\restriction X : h \in S_0\}$ are linear subspaces of $\mathbb{R}^X$ included in $\mathcal{L}^1(\mu)$, their images $C_k^\#$ and $S_0^\#$ are linear subspaces of L^1. If $u \in L^1$ and $\epsilon > 0$, there are an $f \in \mathcal{L}^1$ such that $f^\bullet = u$, and $g \in C_k$, $h \in S_0$ such that $\int_X |f - g|$, $\int_X |f - h| \le \epsilon$; now $v = (g\restriction X)^\bullet \in C_k^\#$ and $w = (h\restriction X)^\bullet \in S_0^\#$ and

$$\|u - v\|_1 = \int_X |f - g| \le \epsilon, \quad \|u - w\|_1 = \int_X |f - h| \le \epsilon.$$

As u and ϵ are arbitrary, $C_k^\#$ and $S_0^\#$ are dense in L^1.

242P Complex L^1 As you would, I hope, expect, we can repeat the work above with $\mathcal{L}^1_{\mathbb{C}}$, the space of complex-valued integrable functions, in place of $\mathcal{L}^1$, to construct a complex Banach space $L^1_{\mathbb{C}}$. The required changes, based on the ideas of 241J, are minor.

(a) In 242Aa, it is perhaps helpful to remark that, for $f \in \mathcal{L}^0_{\mathbb{C}}$,

$$f \in \mathcal{L}^1_{\mathbb{C}} \iff |f| \in \mathcal{L}^1 \iff \mathcal{R}\mathrm{e}(f), \mathcal{I}\mathrm{m}(f) \in \mathcal{L}^1.$$

Consequently, for $u \in L^0_{\mathbb{C}}$,

$$u \in L^1_{\mathbb{C}} \iff |u| \in L^1 \iff \mathcal{R}\mathrm{e}(u), \mathcal{I}\mathrm{m}(u) \in L^1.$$

(b) To prove a complex version of 242E, observe that if $\langle f_n\rangle_{n\in\mathbb{N}}$ is a sequence in $\mathcal{L}^1_{\mathbb{C}}$ such that $\sum_{n=0}^\infty \int |f_n| < \infty$, then $\sum_{n=0}^\infty \int |\mathcal{R}\mathrm{e}(f_n)|$ and $\sum_{n=0}^\infty \int |\mathcal{I}\mathrm{m}(f_n)|$ are both finite, so we may apply 242E twice and see that

$$\int(\sum_{n=0}^\infty f_n) = \int(\sum_{n=0}^\infty \mathcal{R}\mathrm{e}(f_n)) + \int(\sum_{n=0}^\infty \mathcal{I}\mathrm{m}(f_n)) = \sum_{n=0}^\infty \int f_n.$$

Accordingly we can prove that $L^1_{\mathbb{C}}$ is complete under $\| \|_1$ by the argument of 242F.

(c) Similarly, little change is needed to adapt 242J to give a description of a conditional expectation operator $P : L^1_{\mathbb{C}}(\mu) \to L^1_{\mathbb{C}}(\mu\restriction\mathrm{T})$ when (X, Σ, μ) is a probability space and T is a σ-subalgebra of Σ. In the formula

$$|Pu| \le P|u|$$

of 242Je, we need to know that

$$|Pu| = \sup_{|\zeta|=1} \mathcal{R}\mathrm{e}(\zeta Pu)$$

in $L^0(\mu\restriction\mathrm{T})$ (241Jc), while

$$\mathcal{R}\mathrm{e}(\zeta Pu) = \mathcal{R}\mathrm{e}(P(\zeta u)) = P(\mathcal{R}\mathrm{e}(\zeta u)) \le P|u|$$

whenever $|\zeta| = 1$.

(d) In 242M, we need to replace $\mathcal{S}$ by $\mathcal{S}_{\mathbb{C}}$, the space of 'complex-valued simple functions' of the form $\sum_{k=0}^n a_k\chi E_k$ where each a_k is a complex number and each E_k is a measurable set of finite measure; then we get a dense linear subspace $S_{\mathbb{C}} = \{f^\bullet : f \in \mathcal{S}_{\mathbb{C}}\}$ of $L^1_{\mathbb{C}}$. In 242O, we must replace C_k by $C_k(\mathbb{R}^r;\mathbb{C})$, the space of bounded continuous complex-valued functions of bounded support, and S_0 by the linear span over $\mathbb{C}$ of $\{\chi I : I$ is a bounded half-open interval$\}$.

242X Basic exercises >(a) Let X be a set, and let μ be counting measure on X. Show that $L^1(\mu)$ can be identified with the space $\ell^1(X)$ of absolutely summable real-valued functions on X (see 226A). In particular, the space $\ell^1 = \ell^1(\mathbb{N})$ of absolutely summable real-valued sequences is an L^1 space. Write out proofs of 242F adapted to these special cases.

>(b) Let (X, Σ, μ) be any measure space, and $\hat\mu$ the completion of μ. Show that $\mathcal{L}^1(\hat\mu) = \mathcal{L}^1(\mu)$ and $L^1(\hat\mu) = L^1(\mu)$ (cf. 241Xb).

(c) Let $\langle(X_i, \Sigma_i, \mu_i)\rangle_{i\in I}$ be a family of measure spaces, and (X, Σ, μ) their direct sum. Show that the isomorphism between $L^0(\mu)$ and $\prod_{i\in I} L^0(\mu_i)$ (241Xd) induces an identification between $L^1(\mu)$ and

$$\{u : u \in \prod_{i\in I} L^1(\mu_i), \|u\| = \sum_{i\in I} \|u(i)\|_1 < \infty\} \subseteq \prod_{i\in I} L^1(\mu_i).$$

(d) Let (X, Σ, μ) and (Y, T, ν) be measure spaces, and $\phi : X \to Y$ an inverse-measure-preserving function. Show that $g \mapsto g\phi : \mathcal{L}^1(\nu) \to \mathcal{L}^1(\mu)$ (235G) induces a linear operator $T : L^1(\nu) \to L^1(\mu)$ such that $\|Tv\|_1 = \|v\|_1$ for every $v \in L^1(\nu)$.

(e) Let U be a Riesz space (definition: 241Ed). A **Riesz norm** on U is a norm $\| \ \|$ such that $\|u\| \le \|v\|$ whenever $|u| \le |v|$. Show that if U is given its norm topology (2A4Bb) for such a norm, then (i) $u \mapsto |u| : U \to U$, $(u,v) \mapsto u \vee v : U \times U \to U$ are continuous (ii) $\{u : u \ge 0\}$ is closed.

(f) Show that any Banach lattice must be an Archimedean Riesz space (241Fa).

(g) Let (X,Σ,μ) be a probability space, and T a σ-subalgebra of Σ, Υ a σ-subalgebra of T. Let $P_1 : L^1(\mu) \to L^1(\mu\restriction\mathrm{T})$, $P_2 : L_1(\mu\restriction\mathrm{T}) \to L^1(\mu\restriction\Upsilon)$ and $P : L^1(\mu) \to L^1(\mu\restriction\Upsilon)$ be the corresponding conditional expectation operators. Show that $P = P_2P_1$.

(h) Show that if $g : \mathbb{R}^r \to \mathbb{R}$ is continuous and has bounded support it is bounded and attains its bounds. (*Hint*: 2A2F-2A2G.)

(i) Let μ be Lebesgue measure on $\mathbb{R}$. (i) Take $\delta > 0$. Show that if $\phi_\delta(x) = \exp(-\frac{1}{\delta^2 - x^2})$ for $|x| < \delta$, 0 for $|x| \ge \delta$ then ϕ is **smooth**, that is, differentiable arbitrarily often. (ii) Show that if $F_\delta(x) = \int_{-\infty}^{x} \phi_\delta d\mu$ for $x \in \mathbb{R}$ then F_δ is smooth. (iii) Show that if $a < b < c < d$ in $\mathbb{R}$ there is a smooth function h such that $\chi[b,c] \le h \le \chi[a,d]$. (iv) Write $\mathcal{D}$ for the space of smooth functions $h : \mathbb{R} \to \mathbb{R}$ such that $\{x : h(x) \ne 0\}$ is bounded. Show that $\{h^\bullet : h \in \mathcal{D}\}$ is dense in $L^1(\mu)$. (v) Let f be a real-valued function which is integrable over every bounded subset of $\mathbb{R}$. Show that $f \times h$ is integrable for every $h \in \mathcal{D}$, and that if $\int f \times h = 0$ for every $h \in \mathcal{D}$ then $f = 0$ a.e. (*Hint*: 222D.)

(j) Let (X,Σ,μ) be a probability space, T a σ-subalgebra of Σ and $P : L^1(\mu) \to L^1(\mu\restriction\mathrm{T}) \subseteq L^1(\mu)$ the corresponding conditional expectation operator. Show that if u, $v \in L^1(\mu)$ are such that $P|u| \times P|v| \in L^1(\mu)$, then $\int Pu \times v = \int Pu \times Pv = \int u \times Pv$.

242Y Further exercises (a) Let (X,Σ,μ) be a measure space. Let $A \subseteq L^1 = L^1(\mu)$ be a non-empty downwards-directed set, and suppose that $\inf A = 0$ in L^1. (i) Show that $\inf_{u\in A} \|u\|_1 = 0$. (*Hint*: set $\gamma = \inf_{u\in A} \|u\|_1$; find a non-increasing sequence $\langle u_n\rangle_{n\in\mathbb{N}}$ in A such that $\lim_{n\to\infty} \|u_n\|_1 = \gamma$; set $v = \inf_{n\in\mathbb{N}} u_n$ and show that $u \wedge v = v$ for every $u \in A$, so that $v = 0$.) (ii) Show that if U is any open set containing 0, there is a $u \in A$ such that $v \in U$ whenever $0 \le v \le u$.

(b) Let (X,Σ,μ) be a measure space and Y any subset of X; let μ_Y be the subspace measure on Y and $T : L^0(\mu) \to L^0(\mu_Y)$ the canonical map described in 241Yg. (i) Show that $Tu \in L^1(\mu_Y)$ and $\|Tu\|_1 \le \|u\|_1$ for every $u \in L^1(\mu)$. (ii) Show that if $u \in L^1(\mu)$ then $\|Tu\|_1 = \|u\|_1$ iff $\int_E u = \int_{Y\cap E} Tu$ for every $E \in \Sigma$. (iii) Show that T is surjective and that $\|v\|_1 = \min\{\|u\|_1 : u \in L^1(\mu),\, Tu = v\}$ for every $v \in L^1(\mu_Y)$. (*Hint*: 214Eb.) (See also 244Yd below.)

(c) Let (X,Σ,μ) be a measure space. Write $\mathcal{L}^1_{\text{strict}}$ for the space of all integrable Σ-measurable functions from X to $\mathbb{R}$, and $\mathcal{N}$ for the subspace of $\mathcal{L}^1_{\text{strict}}$ consisting of measurable functions which are zero almost everywhere. (i) Show that $\mathcal{L}^1_{\text{strict}}$ is a Dedekind σ-complete Riesz space. (ii) Show that $L^1(\mu)$ can be identified, as ordered linear space, with the quotient $\mathcal{L}^1_{\text{strict}}/\mathcal{N}$ as defined in 241Yb. (iii) Show that $\| \ \|_1$ is a seminorm on $\mathcal{L}^1_{\text{strict}}$. (iv) Show that $f \mapsto |f| : \mathcal{L}^1_{\text{strict}} \to \mathcal{L}^1_{\text{strict}}$ is continuous if $\mathcal{L}^1_{\text{strict}}$ is given the topology defined from $\| \ \|_1$. (v) Show that $\{f : f = 0 \text{ a.e.}\}$ is closed in $\mathcal{L}^1_{\text{strict}}$, but that $\{f : f \ge 0\}$ need not be.

(d) Let (X,Σ,μ) be a measure space, and $\tilde\mu$ the c.l.d. version of μ (213E). Show that the inclusion $\mathcal{L}^1(\mu) \subseteq \mathcal{L}^1(\tilde\mu)$ induces an isomorphism, as ordered normed linear spaces, between $L^1(\tilde\mu)$ and $L^1(\mu)$.

(e) Let (X,Σ,μ) be a measure space and $u_0,\dots,u_n \in L^1(\mu)$. (i) Suppose $k_0,\dots,k_n \in \mathbb{Z}$ are such that $\sum_{i=0}^n k_i = 1$. Show that $\sum_{i=0}^n \sum_{j=0}^n k_ik_j\|u_i - u_j\|_1 \le 0$. (*Hint*: $\sum_{i=0}^n \sum_{j=0}^n k_ik_j|\alpha_i - \alpha_j| \le 0$ for all $\alpha_0,\dots,\alpha_n \in \mathbb{R}$.) (ii) Suppose $\gamma_0,\dots,\gamma_n \in \mathbb{R}$ are such that $\sum_{i=0}^n \gamma_i = 0$. Show that $\sum_{i=0}^n \sum_{j=0}^n \gamma_i\gamma_j\|u_i - u_j\|_1 \le 0$.

(f) Let (X,Σ,μ) be a measure space, and $A \subseteq L^1 = L^1(\mu)$ a non-empty upwards-directed set. Suppose that A is bounded for the norm $\| \ \|_1$. (i) Show that there is a non-decreasing sequence $\langle u_n\rangle_{n\in\mathbb{N}}$ in A such that $\lim_{n\to\infty} \int u_n = \sup_{u\in A} \int u$, and that $\langle u_n\rangle_{n\in\mathbb{N}}$ is Cauchy. (ii) Show that $w = \sup A$ is defined in L^1 and belongs to the norm-closure of A in L^1, so that, in particular, $\|w\|_1 \le \sup_{u\in A} \|u\|_1$.

(g) A Riesz norm (definition: 242Xe) on a Riesz space U is **order-continuous** if $\inf_{u\in A} \|u\| = 0$ whenever $A \subseteq U$ is a non-empty downwards-directed set with infimum 0. (Thus 242Ya tells us that the norms $\| \ \|_1$ are all order-continuous.) Show that in this case (i) any non-decreasing sequence in U which has an upper bound in U must be Cauchy (ii) if U is a Banach lattice, it is U is Dedekind complete. (*Hint for (i)*: if $\langle u_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence with an upper bound in U, let B be the set of upper bounds of $\{u_n : n \in \mathbb{N}\}$ and show that $A = \{v - u_n : v \in B,\, n \in \mathbb{N}\}$ has infimum 0 because U is Archimedean.)

(h) Let (X,Σ,μ) be any measure space. Show that $L^1(\mu)$ has the countable sup property (241Ye).

(i) More generally, show that any Riesz space with an order-continuous Riesz norm has the countable sup property.

(j) Let (X,Σ,μ) and (Y,T,ν) be measure spaces and $U\subseteq L^0(\mu)$ a linear subspace. Let $T:U\to L^0(\nu)$ be a linear operator such that $Tu\ge 0$ in $L^0(\nu)$ whenever $u\in U$ and $u\ge 0$ in $L^0(\mu)$. Suppose that $w\in U$ is such that $w\ge 0$ and $Tw=(\chi Y)^\bullet$. Show that whenever $\phi:\mathbb{R}\to\mathbb{R}$ is a convex function and $u\in L^0(\mu)$ is such that $w\times u$ and $w\times\bar\phi(u)\in U$, defining $\bar\phi:L^0(\mu)\to L^0(\mu)$ as in 241I, then $\bar\phi T(w\times u)\le T(w\times\bar\phi u)$. Explain how this result may be regarded as a common generalization of Jensen's inequality, as stated in 233I, and 242K above. See also 244M below.

(k)(i) A function $\phi:\mathbb{C}\to\mathbb{R}$ is **convex** if $\phi(ab+(1-a)c)\le a\phi(b)+(1-a)\phi(c)$ for all b, $c\in\mathbb{C}$ and $a\in[0,1]$. (ii) Show that such a function must be bounded on any bounded subset of $\mathbb{C}$. (iii) If $\phi:\mathbb{C}\to\mathbb{R}$ is convex and $c\in\mathbb{C}$, show that there is a $b\in\mathbb{C}$ such that $\phi(x)\ge\phi(c)+\mathcal{R}\mathrm{e}(b(x-c))$ for every $x\in\mathbb{C}$. (iv) If $\langle b_c\rangle_{c\in\mathbb{C}}$ is such that $\phi(x)\ge\phi_c(x)=\phi(c)+\mathcal{R}\mathrm{e}(b_c(x-c))$ for all x, $c\in\mathbb{C}$, show that $\{b_c:c\in I\}$ is bounded for any bounded $I\subseteq\mathbb{C}$. (v) Show that if $D\subseteq\mathbb{C}$ is any dense set, $\phi(x)=\sup_{c\in D}\phi_c(x)$ for every $x\in\mathbb{C}$.

(l) Let (X,Σ,μ) be a probability space and T a σ-subalgebra of Σ. Let $P:L^1_{\mathbb{C}}(\mu)\to L^1_{\mathbb{C}}(\mu\restriction\mathrm{T})$ be the conditional expectation operator. Show that if $\phi:\mathbb{C}\to\mathbb{R}$ is any convex function, and we define $\bar\phi(f^\bullet)=(\phi f)^\bullet$ for every $f\in\mathcal{L}^0_{\mathbb{C}}(\mu)$, then $\bar\phi(Pu)\le P(\bar\phi(u))$ whenever $u\in L^1_{\mathbb{C}}(\mu)$ is such that $\bar\phi(u)\in L^1(\mu)$.

242 Notes and comments Of course L^1-spaces compose one of the most important classes of Riesz space, and accordingly their properties have great prominence in the general theory; 242Xe, 242Xf, 242Ya and 242Yf-242Yi outline some of the interrelations between these properties. I will return to these questions in Chapter 35 in the next volume. I have mentioned in passing (242Dd) the additivity of the norm of L^1 on the positive elements. This elementary fact actually characterizes L^1 spaces among Banach lattices; see 369E in the next volume.

Just as $L^0(\mu)$ can be regarded as a quotient of a linear space $\mathcal{L}^0_{\mathrm{strict}}$, so can $L^1(\mu)$ be regarded as a quotient of a linear space $\mathcal{L}^1_{\mathrm{strict}}$ (242Yc). I have discussed this question in the notes to §241; all I try to do here is to be consistent.

We now have a language in which we can speak of 'the' conditional expectation of a function f, the equivalence class in $L^1(\mu\restriction\mathrm{T})$ consisting precisely of all the conditional expections of f on T. If we think of $L^1(\mu\restriction\mathrm{T})$ as identified with its image in $L^1(\mu)$, then the conditional expectation operator $P:L^1(\mu)\to L^1(\mu\restriction\mathrm{T})$ becomes a projection (242Jh). We therefore have re-statements of 233J-233K, as in 242K, 242L and 242Yj.

I give 242O in a fairly general form; but its importance already appears if we take X to be $[0,1]$ with one-dimensional Lebesgue measure. In this case, we have a natural norm on $C([0,1])$, the space of all continuous real-valued functions on $[0,1]$, given by setting

$$\|f\|_1=\int_0^1|f(x)|dx$$

for every $f\in C([0,1])$. The integral here can, of course, be taken to be the Riemann integral; we do not need the Lebesgue theory to show that $\|\,\|_1$ is a norm on $C([0,1])$. It is easy to check that $C([0,1])$ is not complete for this norm (if we set $f_n(x)=\min(1,2^nx^n)$ for $x\in[0,1]$, then $\langle f_n\rangle_{n\in\mathbb{N}}$ is a $\|\,\|_1$-Cauchy sequence with no $\|\,\|_1$-limit in $C([0,1])$). We can use the abstract theory of normed spaces to construct a completion of $C([0,1])$; but it is much more satisfactory if this completion can be given a relatively concrete form, and this is what the identification of L^1 with the completion of $C([0,1])$ can do. (Note that the remark that $\|\,\|_1$ is a norm on $C([0,1])$, that is, that $\|f\|_1\ne 0$ for every non-zero $f\in C([0,1])$, means just that the map $f\mapsto f^\bullet:C([0,1])\to L^1$ is injective, so that $C([0,1])$ can be identified, as ordered normed space, with its image in L^1.) It would be even better if we could find a realization of the completion of $C([0,1])$ as a space of functions on some set Z, rather than as a space of equivalence classes of functions on $[0,1]$. Unfortunately this is not practical; such realizations do exist, but necessarily involve either a thoroughly unfamiliar base set Z, or an intolerably arbitrary embedding map from $C([0,1])$ into $\mathbb{R}^Z$.

You can get an idea of the obstacle to realizing the completion of $C([0,1])$ as a space of functions on $[0,1]$ itself by considering $f_n(x)=\frac{1}{n}x^n$ for $n\ge 1$. An easy calculation shows that $\sum_{n=1}^\infty\|f_n\|_1<\infty$, so that $\sum_{n=1}^\infty f_n$ must exist in the completion of $C([0,1])$; but there is no natural value to assign to it at the point 1. Adaptations of this idea can give rise to indefinitely complicated phenomena – indeed, 242O shows that every integrable function is associated with some appropriate sequence from $C([0,1])$. In §245 I shall have more to say about what $\|\,\|_1$-convergent sequences look like.

From the point of view of measure theory, narrowly conceived, most of the interesting ideas appear most clearly with real functions and real linearspaces. But some of the most important applications of measure theory – important not only as mathematics in general, but also for the measure-theoretic questions they inspire – deal with complex functions and complex linear spaces. I therefore continue to offer sketches of the complex theory, as in 242P. I note that at irregular intervals we need ideas not already spelt out in the real theory, as in 242Pb and 242Yl.

243 L^∞

The second of the classical Banach spaces of measure theory which I treat is the space L^∞. As will appear below, L^∞ is the polar companion of L^1, the linked opposite; for 'ordinary' measure spaces it is actually the dual of L^1 (243F-243G).

243A Definitions Let (X,Σ,μ) be any measure space. Let $\mathcal{L}^\infty=\mathcal{L}^\infty(\mu)$ be the set of functions $f\in\mathcal{L}^0=\mathcal{L}^0(\mu)$ which are **essentially bounded**, that is, such that there is some $M\ge 0$ such that $\{x:x\in\operatorname{dom}f,\,|f(x)|\le M\}$ is conegligible, and write

$$L^\infty=L^\infty(\mu)=\{f^\bullet:f\in\mathcal{L}^\infty(\mu)\}\subseteq L^0(\mu).$$

Note that if $f\in\mathcal{L}^\infty$, $g\in\mathcal{L}^0$ and $g=_{\text{a.e.}}f$, then $g\in\mathcal{L}^\infty$; thus $\mathcal{L}^\infty=\{f:f\in\mathcal{L}^0,\ f^\bullet\in L^\infty\}$.

243B Theorem Let (X,Σ,μ) be any measure space. Then

(a) $L^\infty=L^\infty(\mu)$ is a linear subspace of $L^0=L^0(\mu)$.

(b) If $u\in L^\infty$, $v\in L^0$ and $|v|\le|u|$ then $v\in L^\infty$. Consequently $|u|$, $u\vee v$, $u\wedge v$, $u^+=u\vee 0$ and $u^-=(-u)\vee 0$ belong to L^∞ for all u, $v\in L^\infty$.

(c) Writing $e=\chi X^\bullet$, the equivalence class in L^0 of the constant function with value 1, then an element u of L^0 belongs to L^∞ iff there is an $M\ge 0$ such that $|u|\le Me$.

(d) If u, $v\in L^\infty$ then $u\times v\in L^\infty$.

(e) If $u\in L^\infty$ and $v\in L^1=L^1(\mu)$ then $u\times v\in L^1$.

proof (a) If f, $g\in\mathcal{L}^\infty=\mathcal{L}^\infty(\mu)$ and $c\in\mathbb{R}$, then $f+g$, $cf\in\mathcal{L}^\infty$. **P** We have M_1, $M_2\ge 0$ such that $|f|\le M_1$ a.e. and $|g|\le M_2$ a.e. Now

$$|f+g|\le|f|+|g|\le M_1+M_2\text{ a.e.},\qquad |cf|\le|c||M_1|\text{ a.e.},$$

so $f+g$, $cf\in\mathcal{L}^\infty$. **Q** It follows at once that $u+v$, $cu\in L^\infty$ whenever u, $v\in L^\infty$ and $c\in\mathbb{R}$.

(b)(i) Take $f\in\mathcal{L}^\infty$, $g\in\mathcal{L}^0=\mathcal{L}^0(\mu)$ such that $u=f^\bullet$ and $v=g^\bullet$. Then $|g|\le_{\text{a.e.}}|f|$. Let $M\ge 0$ be such that $|f|\le M$ a.e.; then $|g|\le M$ a.e., so $g\in\mathcal{L}^\infty$ and $v\in L^\infty$.

(ii) Now $|\,|u|\,|=|u|$ so $|u|\in L^\infty$ whenever $u\in L^\infty$. Also $u\vee v=\frac{1}{2}(u+v+|u-v|)$, $u\wedge v=\frac{1}{2}(u+v-|u-v|)$ belong to L^∞ for all u, $v\in L^\infty$.

(c)(i) If $u\in L^\infty$, take $f\in\mathcal{L}^\infty$ such that $f^\bullet=u$. Then there is an $M\ge 0$ such that $|f|\le M$ a.e., so that $|f|\le_{\text{a.e.}}M\chi X$ and $|u|\le Me$. (ii) Of course $\chi X\in\mathcal{L}^\infty$, so $e\in L^\infty$, and if $u\in L^0$ and $|u|\le Me$ then $u\in L^\infty$ by (b).

(d) $f\times g\in\mathcal{L}^\infty$ whenever f, $g\in\mathcal{L}^\infty$. **P** If $|f|\le M_1$ a.e. and $|g|\le M_2$ a.e., then

$$|f\times g|=|f|\times|g|\le M_1M_2\text{ a.e. } \mathbf{Q}$$

So $u\times v\in L^\infty$ for all u, $v\in L^\infty$.

(e) If $f\in\mathcal{L}^\infty$ and $g\in\mathcal{L}^1=\mathcal{L}^1(\mu)$, then there is an $M\ge 0$ such that $|f|\le M$ a.e., so $|f\times g|\le_{\text{a.e.}}M|g|$; because $M|g|$ is integrable and $f\times g$ is virtually measurable, $f\times g$ is integrable and $u\times v\in L^1$.

243C The order structure of L^∞ Let (X,Σ,μ) be any measure space. Then $L^\infty=L^\infty(\mu)$, being a linear subspace of $L^0=L^0(\mu)$, inherits a partial order which renders it a partially ordered linear space (compare 242Ca). Because $|u|\in L^\infty$ whenever $u\in L^\infty$ (243Bb), $u\wedge v$ and $u\vee v$ belong to L^∞ whenever u, $v\in L^\infty$, and L^∞ is a Riesz space (compare 242Cd).

The behaviour of L^∞ as a Riesz space is dominated by the fact that it has an **order unit** e with the property that

$$\text{for every } u\in L^\infty \text{ there is an } M\ge 0 \text{ such that } |u|\le Me$$

(243Bc).

243D The norm of L^∞ Let (X,Σ,μ) be any measure space.

(a) For $f\in\mathcal{L}^\infty=\mathcal{L}^\infty(\mu)$, say that the **essential supremum** of $|f|$ is

$$\operatorname{ess\,sup}|f|=\inf\{M:M\ge 0,\ \{x:x\in\operatorname{dom}f,\ |f(x)|\le M\}\text{ is conegligible}\}.$$

Then $|f|\le\operatorname{ess\,sup}|f|$ a.e. **P** Set $M=\operatorname{ess\,sup}|f|$. For each $n\in\mathbb{N}$, there is an $M_n\le M+2^{-n}$ such that $|f|\le M_n$ a.e. Now

$$\{x:|f(x)|\le M\}=\bigcap_{n\in\mathbb{N}}\{x:|f(x)|\le M_n\}$$

is conegligible, so $|f|\le M$ a.e. **Q**

(b) If f, $g \in \mathcal{L}^\infty$ and $f =_{\text{a.e.}} g$, then $\operatorname{ess\,sup}|f| = \operatorname{ess\,sup}|g|$. Accordingly we may define a functional $\| \|_\infty$ on $L^\infty = L^\infty(\mu)$ by setting $\|u\|_\infty = \operatorname{ess\,sup}|f|$ whenever $u = f^\bullet$.

(c) From (a), we see that, for any $u \in L^\infty$, $\|u\|_\infty = \min\{\gamma : |u| \le \gamma e\}$, where, as before, $e = \chi X^\bullet \in L^\infty$. Consequently $\| \|_\infty$ is a norm on L^∞. **P**(i) If u, $v \in L^\infty$ then

$$|u+v| \le |u| + |v| \le (\|u\|_\infty + \|v\|_\infty)e$$

so $\|u+v\|_\infty \le \|u\|_\infty + \|v\|_\infty$. (ii) If $u \in L^\infty$ and $c \in \mathbb{R}$ then

$$|cu| = |c||u| \le |c|\|u\|_\infty e,$$

so $\|cu\|_\infty \le |c|\|u\|_\infty$. (iii) If $\|u\|_\infty = 0$, there is an $f \in \mathcal{L}^\infty$ such that $f^\bullet = u$ and $|f| \le \|u\|_\infty$ a.e.; now $f = 0$ a.e. so $u = 0$. **Q**

(d) Note also that if $u \in L^0$, $v \in L^\infty$ and $|u| \le |v|$ then $|u| \le \|v\|_\infty e$ so $u \in L^\infty$ and $\|u\|_\infty \le \|v\|_\infty$; similarly,

$$\|u \times v\|_\infty \le \|u\|_\infty \|v\|_\infty, \quad \|u \vee v\|_\infty \le \max(\|u\|_\infty, \|v\|_\infty)$$

for all u, $v \in L^\infty$. Thus L^∞ is a commutative Banach algebra (2A4J).

(e) Moreover,

$$|\int u \times v| \le \int |u \times v| = \|u \times v\|_1 \le \|u\|_1 \|v\|_\infty$$

whenever $u \in L^1$ and $v \in L^\infty$, because

$$|u \times v| = |u| \times |v| \le |u| \times \|v\|_\infty e = \|v\|_\infty |u|.$$

(f) Observe that if u, v are non-negative members of L^∞ then

$$\|u \vee v\|_\infty = \max(\|u\|_\infty, \|v\|_\infty);$$

this is because, for any $\gamma \ge 0$,

$$u \vee v \le \gamma e \iff u \le \gamma e \text{ and } v \le \gamma e.$$

243E Theorem For any measure space (X, Σ, μ), $L^\infty = L^\infty(\mu)$ is a Banach lattice under $\| \|_\infty$.

proof (a) We already know that $\|u\|_\infty \le \|v\|_\infty$ whenever $|u| \le |v|$ (243Dd); so we have just to check that L^∞ is complete under $\| \|_\infty$. Let $\langle u_n \rangle_{n\in\mathbb{N}}$ be a Cauchy sequence in L^∞. For each $n \in \mathbb{N}$ choose $f_n \in \mathcal{L}^\infty = \mathcal{L}^\infty(\mu)$ such that $f_n^\bullet = u_n$ in L^∞. For all m, $n \in \mathbb{N}$, $(f_m - f_n)^\bullet = u_m - u_n$. Consequently

$$E_{mn} = \{x : |f_m(x) - f_n(x)| > \|u_m - u_n\|_\infty\}$$

is negligible, by 243Da. This means that

$$E = \bigcap_{n\in\mathbb{N}} \{x : x \in \operatorname{dom} f_n,\ |f_n(x)| \le \|u_n\|_\infty\} \setminus \bigcup_{m,n\in\mathbb{N}} E_{mn}$$

is conegligible. But for every $x \in E$, $|f_m(x) - f_n(x)| \le \|u_m - u_n\|_\infty$ for all m, $n \in \mathbb{N}$, so that $\langle f_n(x) \rangle_{n\in\mathbb{N}}$ is a Cauchy sequence, with a limit in $\mathbb{R}$. Thus $f = \lim_{n\to\infty} f_n$ is defined almost everywhere. Also, at least for $x \in E$,

$$|f(x)| \le \sup_{n\in\mathbb{N}} \|u_n\|_\infty < \infty,$$

so $f \in \mathcal{L}^\infty$ and $u = f^\bullet \in L^\infty$. If $m \in \mathbb{N}$, then, for every $x \in E$,

$$|f(x) - f_m(x)| \le \sup_{n \ge m} |f_n(x) - f_m(x)| \le \sup_{n\ge m} \|u_n - u_m\|_\infty,$$

so

$$\|u - u_m\|_\infty \le \sup_{n\ge m} \|u_n - u_m\|_\infty \to 0$$

as $m \to \infty$, and $u = \lim_{m\to\infty} u_m$ in L^∞. As $\langle u_n \rangle_{n\in\mathbb{N}}$ is arbitrary, L^∞ is complete.

243F The duality between L^∞ and L^1 Let (X, Σ, μ) be any measure space.

(a) I have already remarked that if $u \in L^1 = L^1(\mu)$ and $v \in L^\infty = L^\infty(\mu)$, then $u \times v \in L^1$ and $|\int u \times v| \le \|u\|_1 \|v\|_\infty$ (243Bd, 243De).

(b) Consequently we have a bounded linear operator T from L^∞ to the normed space dual $(L^1)^*$ of L^1, given by writing

$$(Tv)(u) = \int u \times v \text{ for all } u \in L^1,\, v \in L^\infty.$$

P (i) By (a), $(Tv)(u)$ is well-defined for $u \in L^1$, $v \in L^\infty$. (ii) If $v \in L^\infty$, u, u_1, $u_2 \in L^1$ and $c \in \mathbb{R}$, then

$$\begin{aligned}(Tv)(u_1 + u_2) &= \int (u_1 + u_2) \times v = \int (u_1 \times v) + (u_2 \times v)\\ &= \int u_1 \times v + \int u_2 \times v = (Tv)(u_1) + (Tv)(u_2),\end{aligned}$$

$$(Tv)(cu) = \int cu \times v = \int c(u \times v) = c\int u \times v = c(Tv)(u).$$

This shows that $Tv : L^1 \to \mathbb{R}$ is a linear functional for each $v \in L^\infty$. (iii) Next, for any $u \in L^1$ and $v \in L^\infty$,

$$|(Tv)(u)| = |\int u \times v| \le \|u \times v\|_1 \le \|u\|_1 \|v\|_\infty,$$

as remarked in (a). This means that $Tv \in (L^1)^*$ and $\|Tv\| \le \|v\|_\infty$ for every $v \in L^\infty$. (iv) If v, v_1, $v_2 \in L^\infty$, $u \in L^1$ and $c \in \mathbb{R}$, then

$$\begin{aligned}T(v_1 + v_2)(u) &= \int (v_1 + v_2) \times u = \int (v_1 \times u) + (v_2 \times u)\\ &= \int v_1 \times u + \int v_2 \times u = (Tv_1)(u) + (Tv_2)(u)\\ &= (Tv_1 + Tv_2)(u),\end{aligned}$$

$$T(cv)(u) = \int cv \times u = c\int v \times u = c(Tv)(u) = (cTv)(u).$$

As u is arbitrary, $T(v_1 + v_2) = Tv_1 + Tv_2$ and $T(cv) = c(Tv)$; thus $T : L^\infty \to (L^1)^*$ is linear. (v) Recalling from (iii) that $\|Tv\| \le \|v\|_\infty$ for every $v \in L^\infty$, we see that $\|T\| \le 1$. **Q**

(c) Exactly the same arguments show that we have a linear operator $T' : L^1 \to (L^\infty)^*$, given by writing

$$(T'u)(v) = \int u \times v \text{ for all } u \in L^1,\, v \in L^\infty,$$

and that $\|T'\|$ also is at most 1.

243G Theorem Let (X, Σ, μ) be a measure space, and $T : L^\infty(\mu) \to (L^1(\mu))^*$ the canonical map described in 243F. Then

(a) T is injective iff (X, Σ, μ) is semi-finite, and in this case is norm-preserving;

(b) T is bijective iff (X, Σ, μ) is localizable, and in this case is a normed space isomorphism.

proof (a)(i) Suppose that T is injective, and that $E \in \Sigma$ has $\mu E = \infty$. Then χE is not equal almost everywhere to $\mathbf{0}$, so $(\chi E)^\bullet \ne 0$ in L^∞, and $T(\chi E)^\bullet \ne 0$; let $u \in L^1$ be such that $T(\chi E)^\bullet(u) \ne 0$, that is, $\int u \times (\chi E)^\bullet \ne 0$. Express u as $f^\bullet$ where f is integrable; then $\int_E f \ne 0$ so $\int_E |f| \ne 0$. Let g be a simple function such that $0 \le g \le_{\text{a.e.}} |f|$ and $\int g > \int |f| - \int_E |f|$; then $\int_E g \ne 0$. Express g as $\sum_{i=0}^n a_i \chi E_i$ where $\mu E_i < \infty$ for each i; then $0 \ne \sum_{i=0}^n a_i \mu(E_i \cap E)$, so there is an $i \le n$ such that $\mu(E \cap E_i) \ne 0$, and now $E \cap E_i$ is a measurable subset of E of non-zero finite measure.

As E is arbitrary, this shows that (X, Σ, μ) must be semi-finite if T is injective.

(ii) Now suppose that (X, Σ, μ) is semi-finite, and that $v \in L^\infty$ is non-zero. Express v as $g^\bullet$ where $g : X \to \mathbb{R}$ is measurable; then $g \in \mathcal{L}^\infty$. Take any $a \in \,]0, \|v\|_\infty[$; then $E = \{x : |g(x)| \ge a\}$ has non-zero measure. Let $F \subseteq E$ be a measurable set of non-zero finite measure, and set $f(x) = |g(x)|/g(x)$ if $x \in F$, 0 otherwise; then $f \in \mathcal{L}^1$ and $(f \times g)(x) \ge a$ for $x \in F$, so, setting $u = f^\bullet \in L^1$, we have

$$(Tv)(u) = \int u \times v = \int f \times g \ge a\mu F = a\int |f| = a\|u\|_1 > 0.$$

This shows that $\|Tv\| \ge a$; as a is arbitrary, $\|Tv\| \ge \|v\|_\infty$. We know already from 243F that $\|Tv\| \le \|v\|_\infty$, so $\|Tv\| = \|v\|_\infty$ for every non-zero $v \in L^\infty$; the same is surely true for $v = 0$, so T is norm-preserving and injective.

(b)(i) Using (a) and the definition of 'localizable', we see that under either of the conditions proposed (X, Σ, μ) is semi-finite and T is injective and norm-preserving. I therefore have to show just that it is surjective iff (X, Σ, μ) is localizable.

(ii) Suppose that T is surjective and that $\mathcal{E} \subseteq \Sigma$. Let $\mathcal{F}$ be the family of finite unions of members of $\mathcal{E}$, counting $\emptyset$ as the union of no members of $\mathcal{E}$, so that $\mathcal{F}$ is closed under finite unions and, for any $G \in \Sigma$, $E \setminus G$ is negligible for every $E \in \mathcal{E}$ iff $E \setminus G$ is negligible for every $E \in \mathcal{F}$.

If $u \in L^1$, then $h(u) = \lim_{E\in\mathcal{F},E\uparrow} \int_E u$ exists in $\mathbb{R}$. **P** If u is non-negative, then

$$h(u) = \sup\{\int_E u : E \in \mathcal{F}\} \le \int u < \infty.$$

For other u, we can express u as $u_1 - u_2$, where u_1 and u_2 are non-negative, and now $h(u) = h(u_1) - h(u_2)$. **Q**

Evidently $h : L^1 \to \mathbb{R}$ is linear, being a limit of the linear functionals $u \mapsto \int_E u$, and also

$$|h(u)| \le \sup_{E\in\mathcal{F}} |\int_E u| \le \int |u|$$

for every u, so $h \in (L^1)^*$. Since we are supposing that T is surjective, there is a $v \in L^\infty$ such that $Tv = h$. Express v as $g^\bullet$ where $g : X \to \mathbb{R}$ is measurable and essentially bounded. Set $G = \{x : g(x) > 0\} \in \Sigma$.

If $F \in \Sigma$ and $\mu F < \infty$, then

$$\int_F g = \int (\chi F)^\bullet \times g^\bullet = (Tv)(\chi F)^\bullet = h(\chi F)^\bullet = \sup_{E\in\mathcal{F}} \mu(E \cap F).$$

? If $E \in \mathcal{E}$ and $E \setminus G$ is not negligible, then there is a set $F \subseteq E \setminus G$ such that $0 < \mu F < \infty$; now

$$\mu F = \mu(E \cap F) \le \int_F g \le 0,$$

as $g(x) \le 0$ for $x \in F$. **X** Thus $E \setminus G$ is negligible for every $E \in \mathcal{E}$.

Let $H \in \Sigma$ be such that $E \setminus H$ is negligible for every $E \in \mathcal{E}$. **?** If $G \setminus H$ is not negligible, there is a set $F \subseteq G \setminus H$ of non-zero finite measure. Now

$$\mu(E \cap F) \le \mu(H \cap F) = 0$$

for every $E \in \mathcal{E}$, so $\mu(E \cap F) = 0$ for every $E \in \mathcal{F}$, and $\int_F g = 0$; but $g(x) > 0$ for every $x \in F$, so $\mu F = 0$, which is impossible. **X** Thus $G \setminus H$ is negligible.

Accordingly G is an essential supremum of $\mathcal{E}$ in Σ. As $\mathcal{E}$ is arbitrary, (X, Σ, μ) is localizable.

(iii) For the rest of this proof, I will suppose that (X, Σ, μ) is localizable and seek to show that T is surjective.

Take $h \in (L^1)^*$ such that $\|h\| = 1$. Write $\Sigma^f = \{F : F \in \Sigma,\ \mu F < \infty\}$, and for $F \in \Sigma^f$ define $\nu_F : \Sigma \to \mathbb{R}$ by setting

$$\nu_F E = h(\chi(E \cap F)^\bullet)$$

for every $E \in \Sigma$. Then $\nu_F \emptyset = h(0) = 0$, and if E, $E' \in \Sigma$ are disjoint

$$\begin{aligned}\nu_F E + \nu_F E' &= h(\chi(E \cap F)^\bullet) + h(\chi(E' \cap F)^\bullet) = h((\chi(E \cap F) + \chi(E' \cap F))^\bullet)\\ &= h(\chi((E \cup E') \cap F)^\bullet) = \nu_F(E \cup E').\end{aligned}$$

Thus ν_F is additive. Also

$$|\nu_F E| \le \|\chi(E \cap F)^\bullet\|_1 = \mu(E \cap F)$$

for every $E \in \Sigma$, so ν_F is truly continuous in the sense of 232Ab. By the Radon-Nikodým theorem (232E), there is an integrable function g_F such that $\int_E g_F = \nu_F E$ for every $E \in \Sigma$; we may take it that g_F is measurable and has domain X (232He).

(iv) It is worth noting that $|g_F| \le 1$ a.e. **P** If $G = \{x : g_F(x) > 1\}$, then

$$\int_G g_F = \nu_F G \le \mu(F \cap G) \le \mu G;$$

but this is possible only if $\mu G = 0$. Similarly, if $G' = \{x : g_F(x) < -1\}$, then

$$\int_{G'} g_F = \nu_F G' \ge -\mu G',$$

so again $\mu G' = 0$. **Q**

(v) If F, $F' \in \Sigma^f$, then $g_F = g_{F'}$ almost everywhere in $F \cap F'$. **P** If $E \in \Sigma$ and $E \subseteq F \cap F'$, then

$$\int_E g_F = h(\chi(E \cap F)^\bullet) = h(\chi(E \cap F')^\bullet) = \int_E g_{F'}.$$

So 131Hb gives the result. **Q** 213N (applied to $\{g_F \restriction F : F \in \Sigma^f\}$) now tells us that, because μ is localizable, there is a measurable function $g : X \to \mathbb{R}$ such that $g = g_F$ almost everywhere in F, for every $F \in \Sigma^f$.

(vi) For any $F \in \Sigma^f$, the set

$$\{x : x \in F,\ |g(x)| > 1\} \subseteq \{x : |g_F(x)| > 1\} \cup \{x : x \in F,\ g(x) \ne g_F(x)\}$$

is negligible; because μ is semi-finite, $\{x : |g(x)| > 1\}$ is negligible, and $g \in \mathcal{L}^\infty$, with $\operatorname{ess\,sup}|g| \le 1$. Accordingly $v = g^\bullet \in L^\infty$, and we may speak of $Tv \in (L^1)^*$.

(vii) If $F \in \Sigma^f$, then

$$\int_F g = \int_F g_F = \nu_F X = h(\chi F^\bullet).$$

It follows at once that

$$(Tv)(f^\bullet) = \int f \times g = h(f^\bullet)$$

for every simple function $f : X \to \mathbb{R}$. Consequently $Tv = h$, because both Tv and h are continuous and the equivalence classes of simple functions form a dense subset of L^1 (242Mb, 2A3Uc). Thus $h = Tv$ is a value of T.

(viii) The argument as written above has assumed that $\|h\| = 1$. But of course any non-zero member of $(L^1)^*$ is a scalar multiple of an element of norm 1, so is a value of T. So $T : L^\infty \to (L^1)^*$ is indeed surjective, and is therefore an isometric isomorphism, as claimed.

243H Recall that L^0 is always Dedekind σ-complete and sometimes Dedekind complete (241G), while L^1 is always Dedekind complete (242H). In this respect L^∞ follows L^0.

Theorem Let (X, Σ, μ) be a measure space.
(a) $L^\infty(\mu)$ is Dedekind σ-complete.
(b) If μ is localizable, $L^\infty(\mu)$ is Dedekind complete.

proof These are both consequences of 241G. If $A \subseteq L^\infty = L^\infty(\mu)$ is bounded above in L^∞, fix $u_0 \in A$ and an upper bound w_0 of A in L^∞. If B is the set of upper bounds for A in $L^0 = L^0(\mu)$, then $B \cap L^\infty$ is the set of upper bounds for A in L^∞. Moreover, if B has a least member v_0, then we must have $u_0 \le v_0 \le w_0$, so that

$$0 \le v_0 - u_0 \le w_0 - u_0 \in L^\infty$$

and $v_0 - u_0$, v_0 belong to L^∞. (Compare part (a) of the proof of 242H.) Thus $v_0 = \sup A$ in L^∞.

Now we know that L^0 is Dedekind σ-complete; if $A \subseteq L^\infty$ is a non-empty countable set which is bounded above in L^∞, it is surely bounded above in L^0, so has a supremum in L^0 which is also its supremum in L^∞. As A is arbitrary, L^∞ is Dedekind σ-complete. While if μ is localizable, we can argue in the same way with arbitrary non-empty subsets of L^∞ to see that L^∞ is Dedekind complete because L^0 is.

243I A dense subspace of L^∞ In 242M and 242O I described a couple of important dense linear subspaces of L^1 spaces. The position concerning L^∞ is a little different. However I can describe one important dense subspace.

Proposition Let (X, Σ, μ) be a measure space.

(a) Write $\mathcal{S}$ for the space of 'Σ-simple' functions on X, that is, the space of functions from X to $\mathbb{R}$ expressible as $\sum_{k=0}^n a_k \chi E_k$ where $a_k \in \mathbb{R}$ and $E_k \in \Sigma$ for every $k \le n$. Then for every $f \in \mathcal{L}^\infty = \mathcal{L}^\infty(\mu)$ and every $\epsilon > 0$, there is a $g \in \mathcal{S}$ such that $\operatorname{ess\,sup}|f - g| \le \epsilon$.

(b) $S = \{f^\bullet : f \in \mathcal{S}\}$ is a $\| \|_\infty$-dense linear subspace of $L^\infty = L^\infty(\mu)$.

(c) If (X, Σ, μ) is totally finite, then $\mathcal{S}$ is the space of μ-simple functions, so S becomes just the space of equivalence classes of simple functions, as in 242Mb.

proof (a) Let $\tilde{f} : X \to \mathbb{R}$ be a bounded measurable function such that $f =_{\text{a.e.}} \tilde{f}$. Let $n \in \mathbb{N}$ be such that $|f(x)| \le n\epsilon$ for every $x \in X$. For $-n \le k \le n$ set

$$E_k = \{x : k\epsilon \le \tilde{f}(x) < k + 1)\epsilon.$$

Set

$$g = \sum_{k=-n}^n k\epsilon\chi E_k \in \mathcal{S};$$

then $0 \le \tilde{f}(x) - g(x) \le \epsilon$ for every $x \in X$, so

$$\operatorname{ess\,sup}|f - g| = \operatorname{ess\,sup}|\tilde{f} - g| \le \epsilon.$$

(b) This follows immediately, as in 242Mb.

(c) also is elementary.

243J Conditional expectations Conditional expectations are so important that it is worth considering their interaction with every new concept.

(a) If (X, Σ, μ) is any measure space, and T is a σ-subalgebra of Σ, then the canonical embedding $S : L^0(\mu\restriction\mathrm{T}) \to L^0(\mu)$ (242Ja) embeds $L^\infty(\mu\restriction\mathrm{T})$ as a subspace of $L^\infty(\mu)$, and $\|Su\|_\infty = \|u\|_\infty$ for every $u \in L^\infty(\mu\restriction\mathrm{T})$. As in 242Jb, we can identify $L^\infty(\mu\restriction\mathrm{T})$ with its image in $L^\infty(\mu)$.

(b) Now suppose that $\mu X = 1$, and let $P : L^1(\mu) \to L^1(\mu\restriction\mathrm{T})$ be the conditional expectation operator (242Jd). Then $L^\infty(\mu)$ is actually a linear subspace of $L^1(\mu)$. Setting $e = \chi X^\bullet \in L^\infty(\mu)$, we see that $\int_F e = (\mu\restriction\mathrm{T})(F)$ for every $F \in \mathrm{T}$, so

$$Pe = \chi X^\bullet \in L^\infty(\mu\restriction\mathrm{T}).$$

If $u \in L^\infty(\mu)$, then setting $M = \|u\|_\infty$ we have $-Me \le u \le Me$, so $-MPe \le Pu \le MPe$, because P is order-preserving (242Je); accordingly $\|Pu\|_\infty \le M = \|u\|_\infty$. Thus $P\restriction L^\infty(\mu) : L^\infty(\mu) \to L^\infty(\mu\restriction\mathrm{T})$ is an operator of norm 1.

If $u \in L^\infty(\mu\restriction\mathrm{T})$, then $Pu = u$; so $P[L^\infty]$ is the whole of $L^\infty(\mu\restriction\mathrm{T})$.

243K Complex L^∞ All the ideas needed to adapt the work above to complex L^∞ spaces have already appeared in 241J and 242P. Let $\mathcal{L}^\infty_{\mathbb{C}}$ be

$$\{f : f \in \mathcal{L}^0_{\mathbb{C}}, \text{ ess sup}\,|f| < \infty\} = \{f : \mathcal{R}\mathrm{e}(f) \in \mathcal{L}^\infty, \mathcal{I}\mathrm{m}(f) \in \mathcal{L}^\infty\}.$$

Then

$$L^\infty_{\mathbb{C}} = \{f^\bullet : f \in \mathcal{L}^\infty_{\mathbb{C}}\} = \{u : u \in L^0_{\mathbb{C}}, \mathcal{R}\mathrm{e}(u) \in L^\infty, \mathcal{I}\mathrm{m}(u) \in L^\infty\}.$$

Setting

$$\|u\|_\infty = \||u|\|_\infty = \text{ess sup}\,|f| \text{ whenever } f^\bullet = u,$$

we have a norm on $L^\infty_{\mathbb{C}}$ rendering it a Banach space. We still have $u \times v \in L^\infty_{\mathbb{C}}$ and $\|u \times v\|_\infty \le \|u\|_\infty\|v\|_\infty$ for all u, $v \in L^\infty_{\mathbb{C}}$.

We now have a duality between $L^1_{\mathbb{C}}$ and $L^\infty_{\mathbb{C}}$ giving rise to a linear operator $T : L^\infty_{\mathbb{C}} \to (L^1_{\mathbb{C}})^*$ of norm at most 1, defined by the formula

$$(Tv)(u) = \int u \times v \text{ for every } u \in L^1,\, v \in L^\infty.$$

T is injective iff the underlying measure space is semi-finite, and is a bijection iff the underlying measure space is localizable. (This can of course be proved by re-working the arguments of 243G; but it is perhaps easier to note that $T(\mathcal{R}\mathrm{e}(v)) = \mathcal{R}\mathrm{e}(Tv)$, $T(\mathcal{I}\mathrm{m}(v)) = \mathcal{I}\mathrm{m}(Tv)$ for every v, so that the result for complex spaces can be deduced from the result for real spaces.) To check that T is norm-preserving when it is injective, the quickest route seems to be to imitate the argument of (a-ii) of the proof of 243G.

243X Basic exercises (a) Let (X, Σ, μ) be any measure space, and $\hat\mu$ the completion of μ (212C, 241Xb). Show that $\mathcal{L}^\infty(\hat\mu) = \mathcal{L}^\infty(\mu)$ and $L^\infty(\hat\mu) = L^\infty(\mu)$.

>(b) Let (X, Σ, μ) be a non-empty measure space. Write $\mathcal{L}^\infty_{\text{strict}}$ for the space of bounded Σ-measurable real-valued functions with domain X. (i) Show that $L^\infty(\mu) = \{f^\bullet : f \in \mathcal{L}^\infty_{\text{strict}}\} \subseteq L^0 = L^0(\mu)$. (ii) Show that $\mathcal{L}^\infty_{\text{strict}}$ is a Dedekind σ-complete Banach lattice if we give it the norm

$$\|f\|_\infty = \sup_{x\in X} |f(x)| \text{ for every } f \in \mathcal{L}^\infty_{\text{strict}}.$$

(iii) Show that for every $u \in L^\infty = L^\infty(\mu)$, $\|u\|_\infty = \min\{\|f\|_\infty : f \in \mathcal{L}^\infty_{\text{strict}}, f^\bullet = u\}$.

>(c) Let (X, Σ, μ) be any measure space, and A a subset of $L^\infty(\mu)$. Show that A is bounded for the norm $\| \|_\infty$ iff it is bounded above and below for the ordering of L^∞.

(d) Let (X, Σ, μ) be any measure space, and $A \subseteq L^\infty(\mu)$ a non-empty set with a least upper bound w in $L^\infty(\mu)$. Show that $\|w\|_\infty \le \sup_{u\in A} \|u\|_\infty$.

(e) Let $\langle(X_i, \Sigma_i, \mu_i)\rangle_{i\in I}$ be a family of measure spaces, and (X, Σ, μ) their direct sum (214L). Show that the canonical isomorphism between $L^0(\mu)$ and $\prod_{i\in I} L^0(\mu_i)$ (241Xd) induces an isomorphism between $L^\infty(\mu)$ and the subspace

$$\{u : u \in \prod_{i\in I} L^\infty(\mu_i),\, \|u\| = \sup_{i\in I} \|u(i)\|_\infty < \infty\}$$

of $\prod_{i\in I} L^\infty(\mu_i)$.

(f) Let (X,Σ,μ) be any measure space, and $u\in L^1(\mu)$. Show that there is a $v\in L^\infty(\mu)$ such that $\|v\|_\infty\le 1$ and $\int u\times v=\|u\|_1$.

(g) Let (X,Σ,μ) be a semi-finite measure space and $v\in L^\infty(\mu)$. Show that

$$\|v\|_\infty=\sup\{\int u\times v: u\in L^1,\ \|u\|_1\le 1\}=\sup\{\|u\times v\|_1 : u\in L^1,\ \|u\|_1\le 1\}.$$

(h) Give an example of a probability space (X,Σ,μ) and a $v\in L^\infty(\mu)$ such that $\|u\times v\|_1<\|v\|_\infty$ whenever $u\in L^1(\mu)$ and $\|u\|_1\le 1$.

(i) Write out proofs of 243G adapted to the special cases (i) $\mu X=1$ (ii) (X,Σ,μ) is σ-finite.

(j) Let (X,Σ,μ) be any measure space. Show that $L^0(\mu)$ is Dedekind complete iff $L^\infty(\mu)$ is Dedekind complete.

(k) Let (X,Σ,μ) be a totally finite measure space and $\nu:\Sigma\to\mathbb{R}$ a functional. Show that the following are equiveridical: (i) there is a continuous linear functional $h:L^1(\mu)\to\mathbb{R}$ such that $h((\chi E)^\bullet)=\nu E$ for every $E\in\Sigma$ (ii) ν is additive and there is an $M\ge 0$ such that $|\nu E|\le M\mu E$ for every $E\in\Sigma$.

>(l) Let X be any set, and let μ be counting measure on X. In this case it is customary to write $\ell^\infty(X)$ for $\mathcal{L}^\infty(\mu)$, and to identify it with $L^\infty(\mu)$. Write out statements and proofs of the results of this chapter adapted to this special case – if you like, with $X=\mathbb{N}$. In particular, write out a direct proof that $(\ell^1)^*$ can be identified with ℓ^∞. What happens when X has just two members? or three?

(m) Show that if (X,Σ,μ) is any measure space and $u\in L^\infty_{\mathbb{C}}(\mu)$, then

$$\|u\|_\infty=\sup\{\|\mathcal{R}\mathrm{e}(\zeta u)\|_\infty:\zeta\in\mathbb{C},\ |\zeta|=1\}.$$

(n) Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and $\phi:X\to Y$ an inverse-measure-preserving function. Show that $g\phi\in\mathcal{L}^\infty(\mu)$ for every $g\in\mathcal{L}^\infty(\nu)$, and that the map $g\mapsto g\phi$ induces a linear operator $T:L^\infty(\nu)\to L^\infty(\mu)$ defined by setting $T(g^\bullet)=(g\phi)^\bullet$ for every $g\in\mathcal{L}^\infty(\nu)$. (Compare 241Xg.) Show that $\|Tv\|_\infty=\|v\|_\infty$ for every $v\in L^\infty(\nu)$.

(o) For f, $g\in C=C([0,1])$, the space of continuous real-valued functions on the unit interval $[0,1]$, say

$$f\le g \text{ iff } f(x)\le g(x) \text{ for every } x\in[0,1],$$

$$\|f\|_\infty=\sup_{x\in[0,1]}|f(x)|.$$

Show that C is a Banach lattice, and that moreover

$$\|f\vee g\|_\infty=\max(\|f\|_\infty,\|g\|_\infty) \text{ whenever } f,\ g\ge 0,$$

$$\|f\times g\|_\infty\le\|f\|_\infty\|g\|_\infty \text{ for all } f,\ g\in C,$$

$$\|f\|_\infty=\min\{\gamma:|f|\le\gamma\chi X\} \text{ for every } f\in C.$$

243Y Further exercises (a) Let (X,Σ,μ) be a measure space, and Y a subset of X; write μ_Y for the subspace measure on Y. Show that the canonical map from $L^0(\mu)$ onto $L^0(\mu_Y)$ (241Yg) induces a canonical map from $L^\infty(\mu)$ onto $L^\infty(\mu_Y)$, which is norm-preserving iff it is injective iff Y has full outer measure.

243 Notes and comments I mention the formula

$$\|u\vee v\|_\infty=\max(\|u\|_\infty,\|v\|_\infty) \text{ for } u,\ v\ge 0$$

(243Df) because while it does not characterize L^∞ spaces among Banach lattices (see 243Xo), it is in a sense dual to the characteristic property

$$\|u+v\|_1=\|u\|_1+\|v\|_1 \text{ for } u,\ v\ge 0$$

of the norm of L^1. (I will return to this in Chapter 35 in the next volume.)

The particular set $\mathcal{L}^\infty$ I have chosen (243A) is somewhat arbitrary. The space L^∞ can very well be described entirely as a subspace of L^0, without going back to functions at all; see 243Bc, 243Dc. Just as with $\mathcal{L}^0$ and $\mathcal{L}^1$, there are occasions when it would be simpler to work with the linear space of essentially bounded measurable functions from

X to $\mathbb{R}$; and we now have a third obvious candidate, the linear space $\mathcal{L}^\infty_{\mathrm{strict}}$ of measurable functions from X to $\mathbb{R}$ which are literally, rather than essentially, bounded, which is itself a Banach lattice (243Xb).

I suppose the most important theorem of this section is 243G, identifying L^∞ with $(L^1)^*$. This identification is the chief reason for setting 'localizable' measure spaces apart. The proof of 243Gb is long because it depends on two separate ideas. The Radon-Nikodým theorem deals, in effect, with the totally finite case, and then in parts (b-v) and (b-vi) of the proof localizability is used to link the partial solutions g_F together. Exercise 243Xi is supposed to help you to distinguish the two operations. The map $T' : L^1 \to (L^\infty)^*$ (243Fc) is also very interesting in its way, but I shall leave it for Chapter 36.

243G gives another way of looking at conditional expectation operators. If (X, Σ, μ) is a probability space and T is a σ-subalgebra of Σ, of course both μ and $\mu\restriction\mathrm{T}$ are localizable, so $L^\infty(\mu)$ can be identified with $(L^1(\mu))^*$ and $L^\infty(\mu\restriction\mathrm{T})$ can be identified with $(L^1(\mu\restriction\mathrm{T}))^*$. Now we have the canonical embedding $S : L^1(\mu\restriction\mathrm{T}) \to L^1(\mu)$ (242Jb) which is a norm-preserving linear operator, so gives rise to an adjoint operator $S' : L^1(\mu)^* \to L^1(\mu\restriction\mathrm{T})^*$ defined by the formula

$$(S'h)(v) = h(Sv) \text{ for all } v \in L^1(\mu\restriction\mathrm{T}),\ h \in L^1(\mu)^*.$$

Writing $T_\mu : L^\infty(\mu) \to L^1(\mu)^*$ and $T_{\mu\restriction\mathrm{T}} : L^\infty(\mu\restriction\mathrm{T}) \to L^1(\mu\restriction\mathrm{T})^*$ for the canonical maps, we get a map $Q = T^{-1}_{\mu\restriction\mathrm{T}} S' T_\mu : L^\infty(\mu) \to L^\infty(\mu\restriction\mathrm{T})$, defined by saying that

$$\int Qu \times v = (T_{\mu\restriction\mathrm{T}}Qu)(v) = (S'T_\mu u)(v) = (T_\mu v)(Su) = \int Su \times v = \int u \times v$$

whenever $v \in L^1(\mu\restriction\mathrm{T})$ and $u \in L^\infty(\mu)$. But this agrees with the formula of 242L: we have

$$\int Qu \times v = \int u \times v = \int P(u \times v) = \int Pu \times v.$$

Because v is arbitrary, we must have $Qu = Pu$ for every $u \in L^\infty(\mu)$. Thus a conditional expectation operator is, in a sense, the adjoint of the appropriate embedding operator.

The discussion in the last paragraph applies, of course, only to the restriction $P\restriction L^\infty(\mu)$ of the conditional expectation operator to the L^∞ space. Because μ is totally finite, $L^\infty(\mu)$ is a subspace of $L^1(\mu)$, and the real qualities of the operator P are related to its behaviour on the whole space L^1. $P : L^1(\mu) \to L^1(\mu\restriction\mathrm{T})$ can also be expressed as an adjoint operator, but the expression needs more of the theory of Riesz spaces than I have space for here. I will return to this topic in Chapter 36.

244 L^p

Continuing with our tour of the classical Banach spaces, we come to the L^p spaces for $1 < p < \infty$. The case $p = 2$ is more important than all the others put together, and it would be reasonable, perhaps even advisable, to read this section first with this case alone in mind. But the other spaces provide instructive examples and remain a basic part of the education of any functional analyst.

244A Definitions Let (X, Σ, μ) be any measure space, and $p \in \,]1, \infty[$. Write $\mathcal{L}^p = \mathcal{L}^p(\mu)$ for the set of functions $f \in \mathcal{L}^0 = \mathcal{L}^0(\mu)$ such that $|f|^p$ is integrable, and $L^p = L^p(\mu)$ for $\{f^\bullet : f \in \mathcal{L}^p\} \subseteq L^0 = L^0(\mu)$.

Note that if $f \in \mathcal{L}^p$, $g \in \mathcal{L}^0$ and $f =_{\text{a.e.}} g$, then $|f|^p =_{\text{a.e.}} |g|^p$ so $|g|^p$ is integrable and $g \in \mathcal{L}^p$; thus $\mathcal{L}^p = \{f : f \in \mathcal{L}^0,\ f^\bullet \in L^p\}$.

Alternatively, we can define u^p whenever $u \in L^0$, $u \ge 0$ by writing $(f^\bullet)^p = (f^p)^\bullet$ for every $f \in \mathcal{L}^0$ such that $f(x) \ge 0$ for every $x \in \operatorname{dom} f$ (compare 241I), and say that $L^p = \{u : u \in L^0,\ |u|^p \in L^1(\mu)\}$.

244B Theorem Let (X, Σ, μ) be any measure space, and $p \in [1, \infty]$.

(a) $L^p = L^p(\mu)$ is a linear subspace of $L^0 = L^0(\mu)$.

(b) If $u \in L^p$, $v \in L^0$ and $|v| \le |u|$, then $v \in L^p$. Consequently $|u|$, $u \vee v$ and $u \wedge v$ belong to L^p for all u, $v \in L^p$.

proof The cases $p = 1$, $p = \infty$ are covered by 242B, 242C and 243B; so I suppose that $1 < p < \infty$.

(a)(i) Suppose that f, $g \in \mathcal{L}^p = \mathcal{L}^p(\mu)$. If a, $b \in \mathbb{R}$ then $|a + b|^p \le 2^p \max(|a|^p, |b|^p)$, so $|f + g|^p \le_{\text{a.e.}} 2^p(|f|^p \vee |g|^p)$; now $|f + g|^p \in \mathcal{L}^0$ and $2^p(|f|^p \vee |g|^p) \in \mathcal{L}^1$ so $|f + g|^p \in \mathcal{L}^1$. Thus $f + g \in \mathcal{L}^p$ for all f, $g \in \mathcal{L}^p$; it follows at once that $u + v \in L^p$ for all u, $v \in L^p$.

(ii) If $f \in \mathcal{L}^p$ and $c \in \mathbb{R}$ then $|cf|^p = |c|^p|f|^p \in \mathcal{L}^1$, so $cf \in \mathcal{L}^p$. Accordingly $cu \in L^p$ whenever $u \in L^p$ and $c \in \mathbb{R}$.

(b)(i) Express u as $f^\bullet$ and v as $g^\bullet$, where $f \in \mathcal{L}^p$ and $g \in \mathcal{L}^0$. Then $|g| \le_{\text{a.e.}} |f|$, so $|g|^p \le_{\text{a.e.}} |f|^p$ and $|g|^p$ is integrable; accordingly $g \in \mathcal{L}^p$ and $v \in L^p$.

(ii) Now $|\,|u|\,| = |u|$ so $|u| \in L^p$ whenever $u \in L^p$. Finally $u \vee v = \frac{1}{2}(u + v + |u - v|)$ and $u \wedge v = \frac{1}{2}(u + v - |u - v|)$ belong to L^p for all u, $v \in L^p$.

244C The order structure of L^p Let (X,Σ,μ) be any measure space, and $p\in[1,\infty]$. Then 244B is enough to ensure that the partial order inherited from $L^0(\mu)$ makes $L^p(\mu)$ a Riesz space (compare 242C, 243C).

244D The norm of L^p Let (X,Σ,μ) be a measure space, and $p\in\,]1,\infty[$.

(a) For $f\in\mathcal{L}^p=\mathcal{L}^p(\mu)$, set $\|f\|_p=(\int|f|^p)^{1/p}$. If f, $g\in\mathcal{L}^p$ and $f=_{\text{a.e.}}g$ then $|f|^p=_{\text{a.e.}}|g|^p$ so $\|f\|_p=\|g\|_p$. Accordingly we may define $\|\;\|_p:L^p=L^p(\mu)\to[0,\infty[$ by writing $\|f^\bullet\|_p=\|f\|_p$ for every $f\in\mathcal{L}^p$.

Alternatively, we can say just that $\|u\|_p=(\int|u|^p)^{1/p}$ for every $u\in L^p=L^p(\mu)$.

(b) The notation $\|\;\|_p$ carries a promise that it is a norm on L^p; this is indeed so, but I hold the proof over to 244F below. For the moment, however, let us note just that $\|cu\|_p=|c|\|u\|_p$ for all $u\in L^p$ and $c\in\mathbb{R}$, and that if $\|u\|_p=0$ then $\int|u|^p=0$ so $|u|^p=0$ and $u=0$.

(c) If $|u|\le|v|$ in L^p then $\|u\|_p\le\|v\|_p$; this is because $|u|^p\le|v|^p$.

244E I now work through the lemmas required to show that $\|\;\|_p$ is a norm on L^p and, eventually, that the normed space dual of L^p may be identified with a suitable L^q.

Lemma Suppose (X,Σ,μ) is a measure space, and that p, $q\in\,]1,\infty[$ are such that $\frac{1}{p}+\frac{1}{q}=1$.

(a) $ab\le\frac{1}{p}a^p+\frac{1}{q}b^q$ for all real a, $b\ge 0$.

(b)(i) $f\times g$ is integrable and

$$|\int f\times g|\le\int|f\times g|\le\|f\|_p\|g\|_q$$

for all $f\in\mathcal{L}^p=\mathcal{L}^p(\mu)$, $g\in\mathcal{L}^q=\mathcal{L}^q(\mu)$;

(ii) $u\times v\in L^1=L^1(\mu)$ and

$$|\int u\times v|\le\|u\times v\|_1\le\|u\|_p\|v\|_q$$

for all $u\in L^p=L^p(\mu)$, $v\in L^q=L^q(\mu)$.

proof (a) If either a or b is 0, this is trivial. If both are non-zero, we may argue as follows. The function $x\mapsto x^{1/p}:[0,\infty[\,\to\mathbb{R}$ is concave, with second derivative strictly less than 0, so lies entirely below any of its tangents; in particular, below its tangent at the point $(1,1)$, which has equation $y=1+\frac{1}{p}(x-1)$. Thus we have

$$x^{1/p}\le\frac{1}{p}x+1-\frac{1}{p}=\frac{1}{p}x+\frac{1}{q}$$

for every $x\in[0,\infty[$. So if c, $d>0$, then

$$(\frac{c}{d})^{1/p}\le\frac{1}{p}\frac{c}{d}+\frac{1}{q};$$

multiplying both sides by d,

$$c^{1/p}d^{1/q}\le\frac{1}{p}c+\frac{1}{q}d;$$

setting $c=a^p$ and $d=b^q$, we get

$$ab\le\frac{1}{p}a^p+\frac{1}{q}b^q,$$

as claimed.

(b)(i)(α) Suppose first that $\|f\|_p=\|g\|_q=1$. For every $x\in\operatorname{dom}f\cap\operatorname{dom}g$ we have

$$|f(x)g(x)|\le\frac{1}{p}|f(x)|^p+\frac{1}{q}|g(x)|^q$$

by (a). So

$$|f\times g|\le_{\text{a.e.}}\frac{1}{p}|f|^p+\frac{1}{q}|g|^q\in\mathcal{L}^1(\mu)$$

and $f\times g$ is integrable; also

$$\int|f\times g|\le\frac{1}{p}\int|f|^p+\frac{1}{q}\int|g|^q=\frac{1}{p}\|f\|_p^p+\frac{1}{q}\|g\|_q^q=\frac{1}{p}+\frac{1}{q}=1.$$

(β) If $\|f\|_p = 0$, then $\int |f|^p = 0$ so $|f|^p =_{\text{a.e.}} \mathbf{0}$, $f =_{\text{a.e.}} \mathbf{0}$, $f \times g =_{\text{a.e.}} \mathbf{0}$ and

$$\int |f \times g| = 0 = \|f\|_p \|g\|_q.$$

Similarly, if $\|g\|_q = 0$, then $g =_{\text{a.e.}} \mathbf{0}$ and again

$$\int |f \times g| = 0 = \|f\|_p \|g\|_q.$$

(γ) Finally, for general $f \in \mathcal{L}^p$, $g \in \mathcal{L}^q$ such that $c = \|f\|_p$ and $d = \|g\|_q$ are both non-zero, we have $\|\frac{1}{c}f\|_p = \|\frac{1}{d}g\|_q = 1$ so

$$f \times g = cd(\frac{1}{c}f \times \frac{1}{d}g)$$

is integrable, and

$$\int |f \times g| = cd \int |\frac{1}{c}f \times \frac{1}{d}g| \le cd,$$

as required.

(ii) Now if $u \in L^p$, $v \in L^q$ take $f \in \mathcal{L}^p$, $g \in \mathcal{L}^q$ such that $u = f^\bullet$ and $v = g^\bullet$; $f \times g$ is integrable, so $u \times v \in L^1$, and

$$|\int u \times v| \le \|u \times v\|_1 = \int |f \times g| \le \|f\|_p \|g\|_q = \|u\|_p \|v\|_q.$$

Remark Part (b) is '**Hölder's inequality**'. In the case $p = q = 2$ it is '**Cauchy's inequality**'.

244F Proposition Let (X, Σ, μ) be a measure space and $p \in]1, \infty[$. Set $q = p/(p-1)$, so that $\frac{1}{p} + \frac{1}{q} = 1$.

(a) For every $u \in L^p = L^p(\mu)$, $\|u\|_p = \max\{\int u \times v : v \in L^q(\mu),\ \|v\|_q \le 1\}$.

(b) $\| \|_p$ is a norm on L^p.

proof (a) For $u \in L^p$, set

$$\tau(u) = \sup\{\int u \times v : v \in L^q(\mu),\ \|v\|_q \le 1\}.$$

By 244E(b-ii), $\|u\|_p \ge \tau(u)$. If $\|u\|_p = 0$ then surely

$$0 = \|u\|_p = \tau(u) = \max\{\int u \times v : v \in L^q(\mu),\ \|v\|_q \le 1\}.$$

If $\|u\|_p = c > 0$, consider

$$v = c^{-p/q} \operatorname{sgn} u \times |u|^{p/q},$$

where for $a \in \mathbb{R}$ I write $\operatorname{sgn} a = |a|/a$ if $a \ne 0$, 0 if $a = 0$, so that $\operatorname{sgn} : \mathbb{R} \to \mathbb{R}$ is a Borel measurable function; for $f \in \mathcal{L}^0$ I write $(\operatorname{sgn} f)(x) = \operatorname{sgn}(f(x))$ for $x \in \operatorname{dom} f$, so that $\operatorname{sgn} f \in \mathcal{L}^0$; and for $f \in \mathcal{L}^0$ I write $\operatorname{sgn}(f^\bullet) = (\operatorname{sgn} f)^\bullet$ to define a function $\operatorname{sgn} : L^0 \to L^0$ (cf. 241I). Then $v \in L^q = L^q(\mu)$ and

$$\|v\|_q = (\int |v|^q)^{1/q} = c^{-p/q}(\int |u|^p)^{1/q} = c^{-p/q} c^{p/q} = 1.$$

So

$$\begin{aligned}\tau(u) \ge \int u \times v &= c^{-p/q} \int \operatorname{sgn} u \times |u| \times \operatorname{sgn} u \times |u|^{p/q} \\ &= c^{-p/q} \int |u|^{1+\frac{p}{q}} = c^{-p/q} \int |u|^p = c^{p-\frac{p}{q}} = c,\end{aligned}$$

recalling that $1 + \frac{p}{q} = p$, $p - \frac{p}{q} = 1$. Thus $\tau(u) \ge \|u\|_p$ and

$$\tau(u) = \|u\|_p = \int u \times v.$$

(b) In view of the remarks in 244Db, I have only to check that $\|u + v\|_p \le \|u\|_p + \|v\|_p$ for all u, $v \in L^p$. But given u and v, let $w \in L^q$ be such that $\|w\|_q = 1$ and $\int (u + v) \times w = \|u + v\|_p$. Then

$$\|u + v\|_p = \int (u + v) \times w = \int u \times w + \int v \times w \le \|u\|_p + \|v\|_p,$$

as required.

Remark The triangle inequality '$\|u + v\|_p \le \|u\|_p + \|v\|_p$' is **Minkowski's inequality**.

244G Theorem Let (X,Σ,μ) be any measure space, and $p\in[1,\infty]$. Then $L^p=L^p(\mu)$ is a Banach lattice under its norm $\| \|_p$.

proof The cases $p=1$, $p=\infty$ are covered by 242F and 243E, so let us suppose that $1<p<\infty$. We know already that $\|u\|_p\le\|v\|_p$ whenever $|u|\le|v|$, so it remains only to show that L^p is complete.

Let $\langle u_n\rangle_{n\in\mathbb{N}}$ be a sequence in L^p such that $\|u_{n+1}-u_n\|_p\le 4^{-n}$ for every $n\in\mathbb{N}$. Note that

$$\|u_n\|_p\le\|u_0\|_p+\sum_{k=0}^{n-1}\|u_{k+1}-u_k\|_p\le\|u_0\|_p+\sum_{k=0}^{\infty}4^{-k}\le\|u_0\|_p+2$$

for every n. For each $n\in\mathbb{N}$, choose $f_n\in\mathcal{L}^p$ such that $f_0^\bullet=u_0$, $f_n^\bullet=u_n-u_{n-1}$ for $n\ge1$; do this in such a way that $\operatorname{dom}f_n=X$ and f_n is Σ-measurable (241Bk). Then $\|f_n\|_p\le4^{-n+1}$ for $n\ge1$.

For m, $n\in\mathbb{N}$, set

$$E_{mn}=\{x:|f_m(x)|\ge2^{-n}\}\in\Sigma.$$

Then $|f_m(x)|^p\ge2^{-np}$ for $x\in E_{mn}$, so

$$2^{-np}\mu E_{mn}\le\int|f_m|^p<\infty$$

and $\mu E_{mn}<\infty$. So $\chi E_{mn}\in\mathcal{L}^q=\mathcal{L}^q(\mu)$ and

$$\int_{E_{mn}}|f_k|=\int|f_k|\times\chi E_{mn}\le\|f_k\|_p\|\chi E_{mn}\|_q$$

for each k, by 244E(b-i). Accordingly

$$\sum_{k=0}^{\infty}\int_{E_{mn}}|f_k|\le\|\chi E_{mn}\|_q\sum_{k=0}^{\infty}\|f_k\|_p<\infty,$$

and $\sum_{k=0}^{\infty}f_k(x)$ exists for almost every $x\in E_{mn}$, by 242E. This is true for all m, $n\in\mathbb{N}$. But if $x\in X\setminus\bigcup_{m,n\in\mathbb{N}}E_{mn}$, $f_n(x)=0$ for every n, so $\sum_{k=0}^{\infty}f_k(x)$ certainly exists. Thus $g(x)=\sum_{k=0}^{\infty}f_k(x)$ is defined in $\mathbb{R}$ for almost every $x\in X$.

Set $g_n=\sum_{k=0}^{n}f_k$; then $g_n^\bullet=u_n\in L^p$ for each n, and $g(x)=\lim_{n\to\infty}g_n(x)$ is defined for almost every x. Now consider $|g|^p=_{\text{a.e.}}\lim_{n\to\infty}|g_n|^p$. We know that

$$\liminf_{n\to\infty}\int|g_n|^p=\liminf_{n\to\infty}\|u_n\|_p^p\le(2+\|u_0\|_p)^p<\infty,$$

so by Fatou's Lemma

$$\int|g|^p\le\liminf_{k\to\infty}\int|g_k|^p<\infty.$$

Thus $u=g^\bullet\in L^p$. Moreover, for any $m\in\mathbb{N}$,

$$\begin{aligned}\int|g-g_m|^p&\le\liminf_{n\to\infty}\int|g_n-g_m|^p=\liminf_{n\to\infty}\|u_n-u_m\|_p^p\\&\le\liminf_{n\to\infty}\sum_{k=m}^{n-1}4^{-kp}=\sum_{k=m}^{\infty}4^{-kp}=4^{-mp}/(1-4^{-p}).\end{aligned}$$

So

$$\|u-u_m\|_p=(\int|g-g_m|^p)^{1/p}\le4^{-m}/(1-4^{-p})^{1/p}\to0$$

as $m\to\infty$. Thus $u=\lim_{m\to\infty}u_m$ in L^p. As $\langle u_n\rangle_{n\in\mathbb{N}}$ is arbitrary, L^p is complete.

244H Following 242M and 242O, I note that L^p behaves like L^1 in respect of certain dense subspaces.

Proposition (a) Let (X,Σ,μ) be any measure space, and $p\in[1,\infty[$. Then the space S of equivalence classes of μ-simple functions is a dense linear subspace of $L^p=L^p(\mu)$.

(b) Let X be any subset of $\mathbb{R}^r$, where $r\ge1$, and let μ be the subspace measure on X induced by Lebesgue measure on $\mathbb{R}^r$. Write C_k for the set of (bounded) continuous functions $g:\mathbb{R}^r\to\mathbb{R}$ such that $\{x:g(x)\ne0\}$ is bounded, and S_0 for the space of linear combinations of functions of the form χI, where $I\subseteq\mathbb{R}^r$ is a bounded half-open interval. Then $\{(g\restriction X)^\bullet:g\in C_k\}$ and $\{(h\restriction X)^\bullet:h\in S_0\}$ are dense in $L^p(\mu)$.

proof (a) I repeat the argument of 242M with a tiny modification.

(i) Suppose that $u\in L^p(\mu)$, $u\ge0$ and $\epsilon>0$. Express u as $f^\bullet$ where $f:X\to[0,\infty[$ is a measurable function. Let $g:X\to\mathbb{R}$ be a simple function such that $0\le g\le f^p$ and $\int g\ge\int f^p-\epsilon^p$. Set $h=g^{1/p}$. Then h is a simple function and $h\le f$. Because $p>1$, $(f-h)^p+h^p\le f^p$ and

$$\int(f-h)^p\le\int f^p-g\le\epsilon^p,$$

so

$$\|u - h^\bullet\|_p = (\int |f - h|^p)^{1/p} \le \epsilon,$$

while $h^\bullet \in S$.

(ii) For general $u \in L^p$, $\epsilon > 0$, u can be expressed as $u^+ - u^-$ where $u^+ = u \vee 0$, $u^- = (-u) \vee 0$ belong to L^p and are non-negative. By (i), we can find $v_1, v_2 \in S$ such that $\|u^+ - v_1\|_p \le \frac{1}{2}\epsilon$ and $\|u^- - v_2\|_p \le \frac{1}{2}\epsilon$, so that $v = v_1 - v_2 \in S$ and $\|u - v\|_p \le \epsilon$. As u and ϵ are arbitrary, S is dense.

(b) Again, all the ideas are to be found in 242O; the changes needed are in the formulae, not in the method. They will go more easily if I note at the outset that whenever $f_1, f_2 \in \mathcal{L}^p(\mu)$ and $\int |f_1|^p \le \epsilon^p$, $\int |f_2|^p \le \delta^p$ (where $\epsilon, \delta \ge 0$), then $\int |f_1 + f_2|^p \le (\epsilon + \delta)^p$; this is just the triangle inequality for $\| \|_p$ (244Fb). Also I will regularly express the target relationships in the form '$\int_X |f - g|^p \le \epsilon^p$', '$\int_X |f - g|^p \le \epsilon^p$'. Now let me run through the argument of 242Oa, rather more briskly than before.

(i) Suppose first that $f = \chi I \restriction X$ where $I \subseteq \mathbb{R}^r$ is a bounded half-open interval. As before, we can set $h = \chi I$ and get $\int_X |f - h|^p = 0$. This time, use the same construction to find an interval J and a function $g \in C_k$ such that $\chi I \le g \le \chi J$ and $\mu_r(J \setminus I) \le \epsilon^p$; this will ensure that $\int_X |f - g|^p \le \epsilon^p$.

(ii) Now suppose that $f = \chi(X \cap E)$ where $E \subseteq \mathbb{R}^r$ is a set of finite measure. Then, for the same reasons as before, there is a disjoint family $I_0, \dots, I_n$ of half-open intervals such that $\mu_r(E \triangle \bigcup_{j \le n} I_j) \le (\frac{1}{2}\epsilon)^p$. Accordingly $h = \sum_{j=0}^n \chi I_j \in S_0$ and $\int_X |f - h|^p \le (\frac{1}{2}\epsilon)^p$. And (i) tells us that there is for each $j \le n$ a $g_j \in C_k$ such that $\int_X |g_j - \chi I_j|^p \le (\epsilon/2(n+1))^p$, so that $g = \sum_{j=0}^n g_j \in C_k$ and $\int_X |f - g|^p \le \epsilon^p$.

(iii) The move to simple functions, and thence to arbitrary members of $\mathcal{L}^p(\mu)$, is just as before, but using $\|f\|_p$ in place of $\int_X |f|$. Finally, the translation from $\mathcal{L}^p$ to L^p is again direct – remembering, as before, to check that $g \restriction X$, $h \restriction X$ belong to $\mathcal{L}^p(\mu)$ whenever $g \in C_k$ and $h \in S_0$.

***244I Corollary** In the context of 244Hb, $L^p(\mu)$ is separable.

proof Let A be the set

$$\{(\sum_{j=0}^n q_j \chi([a_j, b_j[\cap X))^\bullet : n \in \mathbb{N},\ q_0, \dots, q_n \in \mathbb{Q},\ a_0, \dots, a_n, b_0, \dots, b_n \in \mathbb{Q}^r\}.$$

Then A is a countable subset of $L^p(\mu)$, and its closure must contain $(\sum_{j=0}^n c_j \chi([a_j, b_j[\cap X))^\bullet$ whenever $c_0, \dots, c_n \in \mathbb{R}$ and $a_0, \dots, a_n, b_0, \dots, b_n \in \mathbb{R}^r$; that is, $\overline{A}$ is a closed set including $\{(h \restriction X)^\bullet : h \in S_0\}$, and is the whole of $L^p(\mu)$, by 244Hb.

244J Duality in L^p spaces Let (X, Σ, μ) be any measure space, and $p \in]1, \infty[$. Set $q = p/(p-1)$; note that $\frac{1}{p} + \frac{1}{q} = 1$ and that $p = q/(q-1)$; the relation between p and q is symmetric. Now $u \times v \in L^1(\mu)$ and $\|u \times v\|_1 \le \|u\|_p \|v\|_q$ whenever $u \in L^p = L^p(\mu)$ and $v \in L^q = L^q(\mu)$ (244E). Consequently we have a bounded linear operator T from L^q to the normed space dual $(L^p)^*$ of L^p, given by writing

$$(Tv)(u) = \int u \times v$$

for all $u \in L^p$ and $v \in L^q$, exactly as in 243F.

244K Theorem Let (X, Σ, μ) be a measure space, and $p \in]1, \infty[$; set $q = p/(p-1)$. Then the canonical map $T : L^q(\mu) \to L^p(\mu)^*$, described in 244J, is a normed space isomorphism.

Remark I should perhaps remind anyone who is reading this section to learn about L^2 that the basic theory of Hilbert spaces yields this theorem in the case $p = q = 2$ without any need for the more generally applicable argument given below (see 244N, 244Yk).

proof We know that T is a bounded linear operator of norm at most 1; I need to show (i) that T is actually an isometry (that is, that $\|Tv\| = \|v\|_q$ for every $v \in L^q$), which will show incidentally that T is injective (ii) that T is surjective, which is the really substantial part of the theorem.

(a) If $v \in L^q$, then by 244Fa (recalling that $p = q/(q-1)$) there is a $u \in L^p$ such that $\|u\|_p \le 1$ and $\int u \times v = \|v\|_q$; thus $\|Tv\| \ge (Tv)(u) = \|v\|_q$. As we know already that $\|Tv\| \le \|v\|_q$, we have $\|Tv\| = \|v\|_q$ for every v, and T is an isometry.

(b) The rest of the proof, therefore, will be devoted to showing that $T : L^q \to (L^p)^*$ is surjective. Fix $h \in (L^p)^*$ with $\|h\| = 1$.

I need to show that h 'lives on' a countable union of sets of finite measure in X, in the following sense: there is a non-decreasing sequence $\langle E_n\rangle_{n\in\mathbb{N}}$ of sets of finite measure such that $h(f^\bullet)=0$ whenever $f\in\mathcal{L}^p$ and $f(x)=0$ for $x\in\bigcup_{n\in\mathbb{N}}E_n$. **P** Choose a sequence $\langle u_n\rangle_{n\in\mathbb{N}}$ in L^p such that $\|u_n\|_p\le 1$ for every n and $\lim_{n\to\infty}h(u_n)=\|h\|=1$. For each n, express u_n as $f_n^\bullet$, where $f_n:X\to\mathbb{R}$ is a measurable function. Set

$$E_n=\{x:\sum_{k=0}^n|f_k(x)|^p\ge 2^{-n}\}$$

for $n\in\mathbb{N}$; because $|f_k|^p$ is measurable and integrable and has domain X for every k, $E_n\in\Sigma$ and $\mu E_n<\infty$ for each n.

Now suppose that $f\in\mathcal{L}^p(X)$ and that $f(x)=0$ for $x\in\bigcup_{n\in\mathbb{N}}E_n$; set $u=f^\bullet\in L^p$. **?** Suppose, if possible, that $h(u)\ne 0$, and consider $h(cu)$, where

$$\operatorname{sgn}c=\operatorname{sgn}h(u),\quad 0<|c|<(p\,|h(u)|\,\|u\|_p^{-p})^{1/(p-1)}.$$

(Of course $\|u\|_p\ne 0$ if $h(u)\ne 0$.) For each n, we have

$$\{x:f_n(x)\ne 0\}\subseteq\bigcup_{m\in\mathbb{N}}E_m\subseteq\{x:f(x)=0\},$$

so $|f_n+cf|^p=|f_n|^p+|cf|^p$ and

$$h(u_n+cu)\le\|u_n+cu\|_p=(\|u_n\|_p^p+\|cu\|_p^p)^{1/p}\le(1+|c|^p\|u\|_p^p)^{1/p}.$$

Letting $n\to\infty$,

$$1+ch(u)\le(1+|c|^p\|u\|_p^p)^{1/p}.$$

Because $\operatorname{sgn}c=\operatorname{sgn}h(u)$, $ch(u)=|c||h(u)|$ and we have

$$1+p|c||h(u)|\le(1+ch(u))^p\le 1+|c|^p\|u\|_p^p,$$

so that

$$p|h(u)|\le|c|^{p-1}\|u\|_p^p<p|h(u)|$$

by the choice of c; which is impossible. **X**

This means that $h(f^\bullet)=0$ whenever $f:X\to\mathbb{R}$ is measurable, belongs to $\mathcal{L}^q$, and is zero on $\bigcup_{n\in\mathbb{N}}E_n$. **Q**

(c) Set $H_n=E_n\setminus\bigcup_{k<n}E_k$ for each $n\in\mathbb{N}$; then $\langle H_n\rangle_{n\in\mathbb{N}}$ is a disjoint sequence of sets of finite measure. Now $h(u)=\sum_{n=0}^\infty h(u\times(\chi H_n)^\bullet)$ for every $u\in L^p$. **P** Express u as $f^\bullet$, where $f:X\to\mathbb{R}$ is measurable. Set $f_n=f\times\chi H_n$ for each n, $g=f\times\chi(X\setminus\bigcup_{n\in\mathbb{N}}H_n)$; then $h(g^\bullet)=0$, by (a), because $\bigcup_{n\in\mathbb{N}}H_n=\bigcup_{n\in\mathbb{N}}E_n$. Consider

$$g_n=g+\sum_{k=0}^n f_k\in\mathcal{L}^p$$

for each n. Then $\lim_{n\to\infty}f-g_n=\mathbf{0}$, and

$$|f-g_n|^p\le|f|^p\in\mathcal{L}^1$$

for every n, so by either Fatou's Lemma or Lebesgue's Dominated Convergence Theorem

$$\lim_{n\to\infty}\int|f-g_n|^p=0,$$

and

$$\begin{aligned}\lim_{n\to\infty}\|u-g^\bullet-\sum_{k=0}^n u\times(\chi H_k)^\bullet\|_p&=\lim_{n\to\infty}\|u-g_n^\bullet\|_p\\&=\lim_{n\to\infty}\Bigl(\int|f-g_n|^p\Bigr)^{1/p}=0,\end{aligned}$$

that is,

$$u=g^\bullet+\sum_{k=0}^\infty u\times\chi H_k^\bullet$$

in L^p. Because $h:L^p\to\mathbb{R}$ is linear and continuous, it follows that

$$h(u)=h(g^\bullet)+\sum_{k=0}^\infty h(u\times\chi H_k^\bullet)=\sum_{k=0}^\infty h(u\times\chi H_k^\bullet),$$

as claimed. **Q**

(d) For each $n\in\mathbb{N}$, define $\nu_n:\Sigma\to\mathbb{R}$ by setting

$$\nu_nF=h(\chi(F\cap H_n)^\bullet)$$

for every $F\in\Sigma$. (Note that ν_nF is always defined because $\mu(F\cap H_n)<\infty$, so that

$$\|\chi(F \cap H_n)\|_p = \mu(F \cap H_n)^{1/p} < \infty.)$$

Then $\nu_n\emptyset = h(0) = 0$, and if $\langle F_k \rangle_{k\in\mathbb{N}}$ is a disjoint sequence in Σ,

$$\|\chi(\bigcup_{k\in\mathbb{N}} H_n \cap F_k) - \sum_{k=0}^{m} \chi(H_n \cap F_k)\|_p = \mu(H_n \cap \bigcup_{k=m+1}^{\infty} F_k)^{1/p} \to 0$$

as $m \to \infty$, so

$$\nu_n(\bigcup_{k\in\mathbb{N}} F_k) = \sum_{k=0}^{\infty} \nu_n F_k.$$

So ν_n is countably additive. Further, $|\nu_n F| \le \mu(H_n \cap F)^{1/p}$ for every $F \in \Sigma$, so ν_n is truly continuous in the sense of 232Ab.

There is therefore an integrable function g_n such that $\nu_n F = \int_F g_n$ for every $F \in \Sigma$; let us suppose that g_n is measurable and defined on the whole of X. Set $g(x) = g_n(x)$ whenever $n \in \mathbb{N}$ and $x \in H_n$, $g(x) = 0$ for $x \in X \setminus \bigcup_{n\in\mathbb{N}} H_n$.

(e) $g = \sum_{n=0}^{\infty} g_n \times \chi H_n$ is measurable and has the property that $\int_F g = h(\chi F^\bullet)$ whenever $n \in \mathbb{N}$ and F is a measurable subset of H_n; consequently $\int_F g = h(\chi F^\bullet)$ whenever $n \in \mathbb{N}$ and F is a measurable subset of $E_n = \bigcup_{k\le n} H_k$. Set $G = \{x : g(x) > 0\} \subseteq \bigcup_{n\in\mathbb{N}} E_n$. If $F \subseteq G$ and $\mu F < \infty$, then

$$\lim_{n\to\infty} \int g \times \chi(F \cap E_n) \le \sup_{n\in\mathbb{N}} h(\chi(F \cap E_n)^\bullet) \le \sup_{n\in\mathbb{N}} \|\chi(F \cap E_n)\|_p = (\mu F)^{1/p},$$

so by B.Levi's theorem

$$\int_F g = \int g \times \chi F = \lim_{n\to\infty} \int g \times \chi(F \cap E_n)$$

exists. Similarly, $\int_F g$ exists if $F \subseteq \{x : g(x) < 0\}$ has finite measure; while obviously $\int_F g$ exists if $F \subseteq \{x : g(x) = 0\}$. Accordingly $\int_F g$ exists for every set F of finite measure. Moreover, by Lebesgue's Dominated Convergence Theorem,

$$\int_F g = \lim_{n\to\infty} \int_{F\cap E_n} g = \lim_{n\to\infty} h(\chi(F \cap E_n)^\bullet) = \sum_{n=0}^{\infty} h(\chi(F \cap H_n)^\bullet) = h(\chi F^\bullet)$$

for such F, by (c) above. It follows at once that

$$\int g \times f = h(f^\bullet)$$

for every simple function $f : X \to \mathbb{R}$.

(f) Now $g \in L^q$. **P** (i) We already know that $|g|^q : X \to \mathbb{R}$ is measurable, because g is measurable and $a \mapsto |a|^q$ is continuous. (ii) Suppose that f is a non-negative simple function and $f \le_{\text{a.e.}} |g|^q$. Then $f^{1/p}$ is a simple function, and $\operatorname{sgn} g$ is measurable and takes only the values 0, 1 and -1, so $f_1 = f^{1/p} \times \operatorname{sgn} g$ is simple. We see that $\int |f_1|^p = \int f$, so $\|f_1\|_p = (\int f)^{1/p}$. Accordingly

$$(\int f)^{1/p} \ge h(f_1^\bullet) = \int g \times f_1 = \int |g \times f^{1/p}|$$

$$\ge \int f^{1/q} \times f^{1/p}$$

(because $0 \le f^{1/q} \le_{\text{a.e.}} |g|$)

$$= \int f,$$

and we must have $\int f \le 1$. (iii) Thus

$$\sup\{\int f : f \text{ is a non-negative simple function, } f \le_{\text{a.e.}} |g|^q\} \le 1 < \infty.$$

But now observe that if $\epsilon > 0$ then

$$\{x : |g(x)|^q \ge \epsilon\} = \bigcup_{n\in\mathbb{N}} \{x : x \in E_n,\ |g(x)|^q \ge \epsilon\},$$

and for each $n \in \mathbb{N}$

$$\mu\{x : x \in E_n,\ |g(x)|^q \ge \epsilon\} \le \tfrac{1}{\epsilon},$$

because $f = \epsilon\chi\{x : x \in E_n, |g(x)|^q \ge \epsilon\}$ is a simple function less than or equal to $|g|^q$, so has integral at most 1. Accordingly

$$\mu\{x : |g(x)|^q \ge \epsilon\} = \sup_{n\in\mathbb{N}} \mu\{x : x \in E_n,\ |g(x)|^q \ge \epsilon\} \le \tfrac{1}{\epsilon} < \infty.$$

Thus $|g|^q$ is integrable, by the criterion in 122Ja. **Q**

(g) We may therefore speak of $h_1 = T(g^\bullet) \in (L^p)^*$, and we know that it agrees with h on members of L^p of the form $f^\bullet$ where f is a simple function. But these form a dense subset of L^p, by 244Ha, and both h and h_1 are continuous, so $h = h_1$, by 2A3Uc, and h is a value of T. The argument as written so far has assumed that $\|h\| = 1$. But of course any non-zero member of $(L^p)^*$ is a scalar multiple of an element of norm 1, so is a value of T. So $T : L^q \to (L^p)^*$ is indeed surjective, and is therefore an isometric isomorphism, as claimed.

244L Continuing with the same topics as in §§242 and 243, I turn to the order-completeness of L^p.

Theorem Let (X, Σ, μ) be any measure space, and $p \in [1, \infty[$. Then $L^p = L^p(\mu)$ is Dedekind complete.

proof I use 242H. Let $A \subseteq L^p$ be a non-empty set which is bounded above in L^p. Fix $u_0 \in A$ and set

$$A' = \{u_0 \vee u : u \in A\},$$

so that A' has the same upper bounds as A and is bounded below by u_0. Fixing an upper bound w_0 of A in L^p, then $u_0 \le u \le w_0$ for every $u \in A'$. Set

$$B = \{(u - u_0)^p : u \in A'\}.$$

Then

$$0 \le v \le (w_0 - u_0)^p \in L^1 = L^1(\mu)$$

for every $v \in B$, so B is a non-empty subset of L^1 which is bounded above in L^1, and therefore has a least upper bound v_1 in L^1. Now $v_1^{1/p} \in L^p$; consider $w_1 = u_0 + v_1^{1/p}$. If $u \in A'$ then $(u - u_0)^p \le v_1$ so $u - u_0 \le v_1^{1/p}$ and $u \le w_1$; thus w_1 is an upper bound for A'. If $w \in L^p$ is an upper bound for A', then $u - u_0 \le w - u_0$ and $(u - u_0)^p \le (w - u_0)^p$ for every $u \in A'$, so $(w - u_0)^p$ is an upper bound for B and $v_1 \le (w - u_0)^p$, $v_1^{1/p} \le w - u_0$ and $w_1 \le w$. Thus $w = \sup A' = \sup A$ in L^p. As A is arbitrary, L^p is Dedekind complete.

244M As in the last two sections, the theory of conditional expectations is worth revisiting.

Theorem Let (X, Σ, μ) be a probability space, and T a σ-subalgebra of Σ. Take $p \in [1, \infty]$. Regard $L^0(\mu \restriction \mathrm{T})$ as a subspace of $L^0 = L^0(\mu)$, as in 242Jh, so that $L^p(\mu \restriction \mathrm{T})$ becomes $L^p(\mu) \cap L^0(\mu \restriction \mathrm{T})$. Let $P : L^1(\mu) \to L^1(\mu \restriction \mathrm{T})$ be the conditional expectation operator, as described in 242Jd. Then whenever $u \in L^p = L^p(\mu)$, $|Pu|^p \le P(|u|^p)$, so $Pu \in L^p(\mu \restriction \mathrm{T})$ and $\|Pu\|_p \le \|u\|_p$. Moreover, $P[L^p] = L^p(\mu \restriction \mathrm{T})$.

proof For $p = \infty$, this is 243Jb, so I assume henceforth that $p < \infty$. Concerning the identification of $L^p(\mu \restriction \mathrm{T})$ with $L^p \cap L^0(\mu \restriction \mathrm{T})$, if $S : L^0(\mu \restriction \mathrm{T}) \to L^0$ is the canonical embedding described in 242J, we have $|Su|^p = S(|u|^p)$ for every $u \in L^0(\mu \restriction \mathrm{T})$, so that $Su \in L^p$ iff $|u|^p \in L^1(\mu \restriction \mathrm{T})$ iff $u \in L^p(\mu \restriction \mathrm{T})$.

Set $\phi(t) = |t|^p$ for $t \in \mathbb{R}$; then ϕ is a convex function (because it is absolutely continuous on any bounded interval, and its derivative is non-decreasing), and $|u|^p = \bar{\phi}(u)$ for every $u \in L^0 = L^0(\mu)$, where $\bar{\phi}$ is defined as in 241I. Now if $u \in L^p = L^p(\mu)$, we surely have $u \in L^1$ (because $|u| \le |u|^p \vee (\chi X)^\bullet$, or otherwise); so 242K tells us that $|Pu|^p \le P|u|^p$. But this means that $Pu \in L^p \cap L^1(\mu \restriction \mathrm{T}) = L^p(\mu \restriction \mathrm{T})$, and

$$\|Pu\|_p = (\int |Pu|^p)^{1/p} \le (\int P|u|^p)^{1/p} = (\int |u|^p)^{1/p} = \|u\|_p,$$

as claimed. If $u \in L^p(\mu \restriction \mathrm{T})$, then $Pu = u$, so $P[L^p]$ is the whole of $L^p(\mu \restriction \mathrm{T})$.

244N The space L^2 (a) As I have already remarked, the really important function spaces are L^0, L^1, L^2 and L^∞. L^2 has the special property of being an inner product space; if (X, Σ, μ) is any measure space and u, $v \in L^2 = L^2(\mu)$ then $u \times v \in L^1(\mu)$, by 244Eb, and we may write $(u|v) = \int u \times v$. This makes L^2 a real inner product space (because

$$(u_1 + u_2|v) = (u_1|v) + (u_2|v), \quad (cu|v) = c(u|v), \quad (u|v) = (v|u),$$

$$(u|u) \ge 0, \quad u = 0 \text{ whenever } (u|u) = 0$$

for all u, u_1, u_2, $v \in L^2$ and $c \in \mathbb{R}$) and its norm $\| \ \|_2$ is the associated norm (because $\|u\|_2 = \sqrt{(u|u)}$ whenever $u \in L^2$). Because L^2 is complete (244G), it is a real Hilbert space. The fact that it may be identified with its own dual (244K) can of course be deduced from this.

I will use the phrase '**square-integrable**' to describe functions in $\mathcal{L}^2(\mu)$.

(b) Conditional expectations take a special form in the case of L^2. Let (X, Σ, μ) be a probability space, T a σ-subalgebra of Σ, and $P : L^1 = L^1(\mu) \to L^1(\mu \restriction \mathrm{T}) \subseteq L^1$ the corresponding conditional expectation operator. Then $P[L^2] \subseteq L^2$, where $L^2 = L^2(\mu)$ (244M), so we have an operator $P_2 = P \restriction L^2$ from L^2 to itself. Now P_2 is an orthogonal projection and its kernel is $\{u : u \in L^2, \int_F u = 0$ for every $F \in \mathrm{T}\}$. **P** (i) If $u \in L^1$ then $Pu = 0$ iff $\int_F u = 0$ for every $F \in \mathrm{T}$ (cf. 242Je); so surely the kernel of P_2 is the set described. (ii) Since $P^2 = P$, P_2 also is a projection; because P_2 has norm at most 1 (244M), and is therefore continuous,

$$U = P_2[L^2] = L^2(\mu\restriction\mathrm{T}) = \{u : u \in L^2,\, P_2 u = u\}, \quad V = \{u : P_2 u = 0\}$$

are closed linear subspaces of L^2 such that $U \oplus V = L^2$. (iii) Now suppose that $u \in U$ and $v \in V$. Then $P|v| \in L^2$, so $u \times P|v| \in L^1$ and $P(u \times v) = u \times Pv$, by 242L. Accordingly

$$(u|v) = \int u \times v = \int P(u \times v) = \int u \times Pv = 0.$$

Thus U and V are orthogonal subspaces of L^2, which is what we mean by saying that P_2 is an orthogonal projection. (Some readers will know that every projection of norm at most 1 on an inner product space is orthogonal.) **Q**

***244O** This is not the place for a detailed discussion of the geometry of L^p spaces. However there is a particularly important fact about the shape of the unit ball which is accessible by the methods available to us here.

Theorem (Clarkson 36) Suppose that $p \in]1, \infty[$ and (X, Σ, μ) is a measure space. Then $L^p = L^p(\mu)$ is uniformly convex (definition: 2A4K).

proof (Hanner 56, Naor 04)

(a)(i) For $0 < t \le 1$ and $a, b \in \mathbb{R}$, set

$$\phi_0(t) = (1+t)^{p-1} + (1-t)^{p-1},$$

$$\phi_1(t) = \tfrac{(1+t)^{p-1} - (1-t)^{p-1}}{t^{p-1}} = (\frac{1}{t} + 1)^{p-1} - (\frac{1}{t} - 1)^{p-1},$$

$$\psi_{ab}(t) = |a|^p \phi_0(t) + |b|^p \phi_1(t),$$

$$\phi_2(b) = (1+b)^p + |1-b|^p.$$

(ii) We have

$$\phi_0'(t) = (p-1)((1+t)^{p-2} - (1-t)^{p-2}), \text{ which has the same sign as } p-2,$$

(of course it is zero if $p = 2$),

$$\begin{aligned}\phi_1'(t) &= -\tfrac{p-1}{t^2}((\frac{1}{t} - 1)^{p-2} - (\frac{1}{t} - 1)^{p-2}) \\ &= -\tfrac{p-1}{t^p}((1+t)^{p-2} - (1-t)^{p-2}) = -\tfrac{1}{t^p}\phi_0'(t)\end{aligned}$$

for every $t \in]0, 1[$. Accordingly $\phi_0' - \phi_1'$ has the same sign as $p - 2$ everywhere on $]0, 1[$. Also

$$\phi_0(1) = 2^{p-1} = \phi_1(1),$$

so $\phi_0 - \phi_1$ has the same sign as $2 - p$ everywhere on $]0, 1]$.

(iii) ϕ_2 is strictly increasing on $[0, \infty[$. **P** For $b > 0$,

$$\begin{aligned}\phi_2'(b) &= p((1+b)^{p-1} - (1-b)^{p-1}) > 0 \text{ if } b \le 1, \\ &= p((1+b)^{p-1} + (b-1)^{p-1}) > 0 \text{ if } b \ge 1. \text{ **Q**}\end{aligned}$$

(iv) If $0 < b \le a$, then

$$\begin{aligned}\psi_{ab}(\frac{b}{a}) &= a^p \phi_0(\frac{b}{a}) + b^p \phi_1(\frac{b}{a}) \\ &= a^p(1 + \frac{b}{a})^{p-1} + a^p(1 - \frac{b}{a})^{p-1} + b^p(\frac{a}{b} + 1)^{p-1} - b^p(\frac{a}{b} - 1)^{p-1} \\ &= a(a+b)^{p-1} + a(a-b)^{p-1} + b(a+b)^{p-1} - b(a-b)^{p-1} \\ &= (a+b)^p + (a-b)^p = (a+b)^p + |a-b|^p. \qquad (\dagger)\end{aligned}$$

Also $\psi'_{ab}(t) = (a^p - \frac{b^p}{t^p})\phi'_0(t)$ has the sign of $2-p$ if $0 < t < \frac{b}{a}$ and the sign of $p-2$ if $\frac{b}{a} < t < 1$. Accordingly

—— if $1 < p \le 2$, $\psi_{ab}(t) \le \psi_{ab}(\frac{b}{a}) = (a+b)^p + |a-b|^p$ for every $t \in \,]0,1]$,

—— if $p \ge 2$, $\psi_{ab}(t) \ge \psi_{ab}(\frac{b}{a}) = (a+b)^p + |a-b|^p$ for every $t \in \,]0,1]$.

(v) Now consider the case $0 < a \le b$. If $1 < p \le 2$,

$$\psi_{ab}(t) = a^p\phi_0(t) + b^p\phi_1(t) \le a^p\phi_0(t) + b^p\phi_1(t) + (b^p - a^p)(\phi_0(t) - \phi_1(t))$$

(by (ii))

$$= b^p\phi_0(t) + a^p\phi_1(t) \le (b+a)^p + (b-a)^p = (a+b)^p + |a-b|^p$$

for every $t \in \,]0,1]$. If $p \ge 2$, on the other hand,

$$\begin{aligned}\psi_{ab}(t) &= a^p\phi_0(t) + b^p\phi_1(t) \ge a^p\phi_0(t) + b^p\phi_1(t) + (b^p - a^p)(\phi_0(t) - \phi_1(t))\\ &= b^p\phi_0(t) + a^p\phi_1(t) \ge (a+b)^p + |a-b|^p\end{aligned}$$

for every t.

(vi) Thus we have the inequalities

$$\begin{aligned}\psi_{ab}(t) &\le |a+b|^p + |a-b|^p \text{ if } p \in \,]1,2]\,,\\ &\ge |a+b|^p + |a-b|^p \text{ if } p \in [2,\infty[\end{aligned} \qquad (*)$$

whenever $t \in \,]0,1]$ and $a, b \in \,]0,\infty[$. Since $(a,b) \mapsto \psi_{ab}(t)$ is continuous for every t, the same inequalities are valid for all $a, b \in [0,\infty[$. And since

$$\psi_{ab}(t) = \psi_{|a|,|b|}(t), \quad |a+b|^p + |a-b|^p = \big||a|+|b|\big|^p + \big||a|-|b|\big|^p$$

for all $a, b \in \mathbb{R}$ and $t \in \,]0,1]$, the inequalities (*) are valid for all $a, b \in \mathbb{R}$ and $t \in \,]0,1]$.

(b) Suppose that $p \ge 2$.

(i)

$$\|u+v\|_p^p + \|u-v\|_p^p \le (\|u\|_p + \|v\|_p)^p + |\|u\|_p - \|v\|_p|^p$$

for all $u, v \in L^p$. **P** First consider the case $0 < \|v\|_p \le \|u\|_p$. Let $f, g : X \to \mathbb{R}$ be Σ-measurable functions such that $f^\bullet = u$ and $g^\bullet = v$. Then for any $t \in \,]0,1]$,

$$\begin{aligned}\|u+v\|_p^p + \|u-v\|_p^p &= \int |f(x)+g(x)|^p + |f(x)-g(x)|^p \mu(dx)\\ &\le \int \psi_{f(x),g(x)}(t)\mu(dx)\end{aligned}$$

(by the second inequality in (*))

$$= \int |f(x)|^p\phi_0(t) + |g(x)|^p\phi_1(t)\mu(dx) = \|u\|_p^p\phi_0(t) + \|v\|_p^p\phi_1(t).$$

In particular, taking $t = \|v\|_p/\|u\|_p$, and applying (†) from (a-iv),

$$\|u+v\|_p^p + \|u-v\|_p^p \le (\|u\|_p + \|v\|_p)^p + |\|u\|_p - \|v\|_p|^p.$$

Of course the result will also be true if $0 < \|u\|_p \le \|v\|_p$, and the case in which either u or v is zero is trivial. **Q**

(ii) Let $\epsilon \in \,]0,2]$. Set $\delta = 2 - (2^p - \epsilon^p)^{1/p} > 0$. If $u, v \in L^p$, $\|u\|_p = \|v\|_p = 1$ and $\|u-v\|_p \ge \epsilon$, then

$$\|u+v\|_p^p + \epsilon^p \le \|u+v\|_p^p + \|u-v\|_p^p \le (\|u\|_p + \|v\|_p)^p + |\|u\|_p - \|v\|_p|^p = 2^p,$$

so $\|u+v\|_p \le (2^p - \epsilon^p)^{1/p} = 2 - \delta$. As u, v and ϵ are arbitrary, L^p is uniformly convex.

(c) Next suppose that $p \in \,]1,2]$.

(i)

$$(\|u\|_p + \|v\|_p)^p + |\|u\|_p - \|v\|_p|^p \le \|u+v\|_p^p + \|u-v\|_p^p$$

for all u, $v \in L^p$. **P** We can repeat all the ideas, and most of the formulae, of (b-i). As before, start with the case $0 < \|v\|_p \le \|u\|_p$. Let f, $g : X \to \mathbb{R}$ be Σ-measurable functions such that $f^\bullet = u$ and $g^\bullet = v$. Taking $t = \|v\|_p/\|u\|_p$,

$$\begin{aligned}\|u+v\|_p^p + \|u-v\|_p^p &= \int |f(x)+g(x)|^p + |f(x)-g(x)|^p \mu(dx)\\ &\ge \int \psi_{f(x),g(x)}(t)\mu(dx)\end{aligned}$$

(by the first inequality in (*))

$$= \|u\|_p^p \phi_0(t) + \|v\|_p^p \phi_1(t) = (\|u\|_p + \|v\|_p)^p + |\|u\|_p - \|v\|_p|^p.$$

Similarly if $0 < \|u\|_p \le \|v\|_p$, and the case in which either u or v is zero is trivial. **Q**

(ii) Let $\epsilon > 0$. Set $\gamma = \phi_2(\frac{\epsilon}{2}) > 2$ (see (a-iii) above) and $\delta = 2\bigl(1-(\frac{2}{\gamma})^{1/p}\bigr) > 0$. Now suppose that $\|u\|_p = \|v\|_p = 1$ and $\|u-v\|_p \ge \epsilon$. Then $\|u+v\|_p \le 2-\delta$. **P** If $u+v=0$ this is trivial. Otherwise, set $a = \|u+v\|_p$ and $b = \|u-v\|_p$. Then $a \le 2$ and $b \ge \epsilon$, so

$$a^p\gamma = a^p\phi_2(\tfrac{\epsilon}{2}) \le a^p\phi_2(\tfrac{b}{a})$$

(by (a-iii) again)

$$\begin{aligned}&= (a+b)^p + |a-b|^p = (\|u+v\|_p + \|u-v\|_p)^p + |\|u+v\|_p - \|u-v\|_p|^p\\ &\le \|2u\|_p^p + \|2v\|_p^p\end{aligned}$$

(by (i) here)

$$= 2^{p+1}$$

and $a \le 2\bigl(\frac{2}{\gamma}\bigr)^{1/p} = 2-\delta$. **Q** As u, v and ϵ are arbitrary, L^p is uniformly convex.

Remark The inequalities in (b-i) and (c-i) of the proof are called **Hanner's inequalities**.

244P Complex L^p Let (X,Σ,μ) be any measure space.

(a) For any $p \in \,]1,\infty[$, set

$$\mathcal{L}^p_{\mathbb{C}} = \mathcal{L}^p_{\mathbb{C}}(\mu) = \{f : f \in \mathcal{L}^0_{\mathbb{C}}(\mu),\ |f|^p \text{ is integrable}\},$$

$$\begin{aligned}L^p_{\mathbb{C}} = L^p_{\mathbb{C}}(\mu) &= \{f^\bullet : f \in \mathcal{L}^p_{\mathbb{C}}\}\\ &= \{u : u \in L^0_{\mathbb{C}}(\mu),\ \mathcal{R}\mathrm{e}(u) \in L^p(\mu) \text{ and } \mathcal{I}\mathrm{m}(u) \in L^p(\mu)\}\\ &= \{u : u \in L^0_{\mathbb{C}}(\mu),\ |u| \in L^p(\mu)\}.\end{aligned}$$

Then $L^p_{\mathbb{C}}$ is a linear subspace of $L^0_{\mathbb{C}}(\mu)$. Set $\|u\|_p = \||u|\|_p = (\int |u|^p)^{1/p}$ for $u \in L^p_{\mathbb{C}}$.

(b) The proof of 244E(b-i) applies unchanged to complex-valued functions, so taking $q = p/(p-1)$ we get

$$\|u \times v\|_1 \le \|u\|_p\|v\|_q$$

for all $u \in L^p_{\mathbb{C}}$, $v \in L^q_{\mathbb{C}}$. 244Fa becomes

for every $u \in L^p_{\mathbb{C}}$ there is a $v \in L^q_{\mathbb{C}}$ such that $\|v\|_q \le 1$ and

$$\int u \times v = |\int u \times v| = \|u\|_p;$$

the same proof works, if you omit all mention of the functional τ, and allow me to write $\operatorname{sgn} a = |a|/a$ for all non-zero complex numbers – it would perhaps be more natural to write $\overline{\operatorname{sgn}} a$ in place of $\operatorname{sgn} a$. So, just as before, we find that $\|\ \|_p$ is a norm. We can use the argument of 244G to show that $L^p_{\mathbb{C}}$ is complete. (Alternatively, note that a sequence $\langle u_n\rangle_{n\in\mathbb{N}}$ in $L^0_{\mathbb{C}}$ is Cauchy, or convergent, iff its real and imaginary parts are.) The space $S_{\mathbb{C}}$ of equivalence classes of 'complex-valued simple functions' is dense in $L^p_{\mathbb{C}}$. If X is a subset of $\mathbb{R}^r$ and μ is Lebesgue measure on X, then the space of equivalence classes of continuous complex-valued functions on X with bounded support is dense in $L^p_{\mathbb{C}}$.

(c) The canonical map $T : L^q_{\mathbb{C}} \to (L^p_{\mathbb{C}})^*$, defined by writing $(Tv)(u) = \int u \times v$, is surjective because $T \restriction L^q : L^q \to (L^p)^*$ is surjective; and it is an isometry by the remarks in (b) just above. Thus we can still identify $L^q_{\mathbb{C}}$ with $(L^p_{\mathbb{C}})^*$.

(d) When we come to the complex form of Jensen's inequality, it seems that a new idea is needed. I have relegated this to 242Yk-242Yl. But for the complex form of 244M a simpler argument will suffice. If (X, Σ, μ) is a probability space, T is a σ-subalgebra of Σ and $P : L^1_{\mathbb{C}}(\mu) \to L^1_{\mathbb{C}}(\mu \restriction \mathrm{T})$ is the corresponding conditional expectation operator, then for any $u \in L^p_{\mathbb{C}}$ we shall have

$$|Pu|^p \le (P|u|)^p \le P(|u|^p),$$

applying 242Pc and 244M. So $\|Pu\|_p \le \|u\|_p$, as before.

(e) There is a special point arising with $L^2_{\mathbb{C}}$. We now have to define

$$(u|v) = \int u \times \bar{v}$$

for u, $v \in L^2_{\mathbb{C}}$, so that $(u|u) = \int |u|^2 = \|u\|_2^2$ for every u; this means that $(v|u)$ is the complex conjugate of $(u|v)$.

244X Basic exercises >(a) Let (X, Σ, μ) be a measure space, and $(X, \hat{\Sigma}, \hat{\mu})$ its completion. Show that $\mathcal{L}^p(\hat{\mu}) = \mathcal{L}^p(\mu)$ and $L^p(\hat{\mu}) = L^p(\mu)$ for every $p \in [1, \infty]$.

(b) Let (X, Σ, μ) be a measure space, and $1 \le p \le q \le r \le \infty$. Show that $L^p(\mu) \cap L^r(\mu) \subseteq L^q(\mu) \subseteq L^p(\mu) + L^r(\mu) \subseteq L^0(\mu)$. (See also 244Yh.)

(c) Let (X, Σ, μ) be a measure space. Suppose that p, q, $r \in [1, \infty]$ and that $\frac{1}{p} + \frac{1}{q} = \frac{1}{r}$, setting $\frac{1}{\infty} = 0$ as usual. Show that $u \times v \in L^r(\mu)$ and $\|u \times v\|_r \le \|u\|_p \|v\|_q$ whenever $u \in L^p(\mu)$ and $v \in L^q(\mu)$. (*Hint*: if $r < \infty$ apply Hölder's inequality to $|u|^r \in L^{p/r}$, $|v|^r \in L^{q/r}$.)

>(d)(i) Let (X, Σ, μ) be a probability space. Show that if $1 \le p \le r \le \infty$ then $\|f\|_p \le \|f\|_r$ for every $f \in \mathcal{L}^r(\mu)$. (*Hint*: use Hölder's inequality to show that $\int |f|^p \le \| |f|^p \|_{r/p}$.) In particular, $\mathcal{L}^p(\mu) \supseteq \mathcal{L}^r(\mu)$. (ii) Let (X, Σ, μ) be a measure space such that $\mu E \ge 1$ whenever $E \in \Sigma$ and $\mu E > 0$. (This happens, for instance, when μ is 'counting measure' on X.) Show that if $1 \le p \le r \le \infty$ then $L^p(\mu) \subseteq L^r(\mu)$ and $\|u\|_p \ge \|u\|_r$ for every $u \in L^p(\mu)$. (*Hint*: look first at the case $\|u\|_p = 1$.)

(e) Let (X, Σ, μ) be a semi-finite measure space, and p, $q \in [1, \infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Show that if $u \in L^0(\mu) \setminus L^p(\mu)$ then there is a $v \in L^q(\mu)$ such that $u \times v \notin L^1(\mu)$. (*Hint*: reduce to the case $u \ge 0$. Show that in this case there is for each $n \in \mathbb{N}$ a $u_n \le u$ such that $4^n \le \|u_n\|_p < \infty$; take $v_n \in L^q$ such that $\|v_n\|_q \le 2^{-n}$ and $\int u_n \times v_n \ge 2^n$, and set $v = \sum_{n=0}^{\infty} v_n$.)

(f) Let $\langle (X_i, \Sigma_i, \mu_i) \rangle_{i \in I}$ be a family of measure spaces, and (X, Σ, μ) their direct sum (214L). Take any $p \in [1, \infty[$. Show that the canonical isomorphism between $L^0(\mu)$ and $\prod_{i \in I} L^0(\mu_i)$ (241Xd) induces an isomorphism between $L^p(\mu)$ and the subspace

$$\{u : u \in \textstyle\prod_{i \in I} L^p(\mu_i),\ \|u\| = \bigl(\sum_{i \in I} \|u(i)\|_p^p\bigr)^{1/p} < \infty\}$$

of $\prod_{i \in I} L^p(\mu_i)$.

(g) Let (X, Σ, μ) be a measure space. Set $M^{\infty,1} = L^1(\mu) \cap L^\infty(\mu)$. Show that for $u \in M^{\infty,1}$ the function $p \mapsto \|u\|_p : [1, \infty[\to [0, \infty[$ is continuous, and that $\|u\|_\infty = \lim_{p \to \infty} \|u\|_p$. (*Hint*: consider first the case in which u is the equivalence class of a simple function.)

(h) Let μ be counting measure on $X = \{1, 2\}$, so that $\mathcal{L}^0(\mu) = \mathbb{R}^2$ and $L^p(\mu) = L^0(\mu)$ can be identified with $\mathbb{R}^2$ for every $p \in [1, \infty]$. Sketch the unit balls $\{u : \|u\|_p \le 1\}$ in $\mathbb{R}^2$ for $p = 1$, $\frac{3}{2}$, 2, 3 and ∞.

(i) Let μ be counting measure on $X = \{1, 2, 3\}$, so that $\mathcal{L}^0(\mu) = \mathbb{R}^3$ and $L^p(\mu) = L^0(\mu)$ can be identified with $\mathbb{R}^3$ for every $p \in [1, \infty]$. Describe the unit balls $\{u : \|u\|_p \le 1\}$ in $\mathbb{R}^3$ for $p = 1$, 2 and ∞.

(j) At which points does the argument of 244Hb break down if we try to apply it to L^∞ with $\| \|_\infty$?

(k) Let $p \in [1, \infty[$. (i) Show that $|a^p - b^p| \ge |a - b|^p$ for all a, $b \ge 0$. (*Hint*: for $a > b$, differentiate both sides with respect to a.) (ii) Let (X, Σ, μ) be a measure space and U a linear subspace of $L^0(\mu)$ such that (α) $|u| \in U$ for every $u \in U$ (β) $u^{1/p} \in U$ for every $u \in U$ (γ) $U \cap L^1$ is dense in $L^1 = L^1(\mu)$. Show that $U \cap L^p$ is dense in $L^p = L^p(\mu)$. (*Hint*: check first that $\{u : u \in U \cap L^1,\ u \ge 0\}$ is dense in $\{u : u \in L^1,\ u \ge 0\}$.) (iii) Use this to prove 244H from 242M and 242O.

(l) For any measure space (X,Σ,μ) write $M^{1,\infty}=M^{1,\infty}(\mu)$ for $\{v+w: v\in L^1(\mu),\, w\in L^\infty(\mu)\}\subseteq L^0(\mu)$. Show that $M^{1,\infty}$ is a linear subspace of L^0 including L^p for every $p\in[1,\infty]$, and that if $u\in L^0$, $v\in M^{1,\infty}$ and $|u|\le|v|$ then $u\in M^{1,\infty}$. (*Hint*: $u=v\times w$ where $|w|\le\chi X^\bullet$.)

(m) Let (X,Σ,μ) and (Y,T,ν) be two measure spaces, and let $\mathcal{T}^+$ be the set of linear operators $T: M^{1,\infty}(\mu)\to M^{1,\infty}(\nu)$ such that (α) $Tu\ge0$ whenever $u\ge0$ in $M^{1,\infty}(\mu)$ (β) $Tu\in L^1(\nu)$ and $\|Tu\|_1\le\|u\|_1$ whenever $u\in L^1(\mu)$ (γ) $Tu\in L^\infty(\nu)$ and $\|Tu\|_\infty\le\|u\|_\infty$ whenever $u\in L^\infty(\mu)$. (i) Show that if $\phi:\mathbb{R}\to\mathbb{R}$ is a convex function such that $\phi(0)=0$, and $u\in M^{1,\infty}(\mu)$ is such that $\bar\phi(u)\in M^{1,\infty}(\mu)$ (interpreting $\bar\phi: L^0(\mu)\to L^0(\mu)$ as in 241I), then $\bar\phi(Tu)\in M^{1,\infty}(\nu)$ and $\bar\phi(Tu)\le T(\bar\phi(u))$ for every $T\in\mathcal{T}^+$. (ii) Hence show that if $p\in[1,\infty]$ and $u\in L^p(\mu)$, $Tu\in L^p(\nu)$ and $\|Tu\|_p\le\|u\|_p$ for every $T\in\mathcal{T}^+$.

>(n) Let X be any set, and let μ be counting measure on X. In this case it is customary to write $\ell^p(X)$ for $\mathcal{L}^p(\mu)$, and to identify it with $L^p(\mu)$. In particular, $L^2(\mu)$ becomes identified with $\ell^2(X)$, the space of square-summable functions on X. Write out statements and proofs of the results of this section adapted to this special case.

(o) Let (X,Σ,μ) and (Y,T,ν) be measure spaces and $\phi:X\to Y$ an inverse-measure-preserving function. Show that the map $g\mapsto g\phi:\mathcal{L}^0(\nu)\to\mathcal{L}^0(\mu)$ (241Xg) induces a norm-preserving map from $L^p(\nu)$ to $L^p(\mu)$ for every $p\in[1,\infty]$, and also a map from $M^{1,\infty}(\nu)$ to $M^{1,\infty}(\mu)$ which belongs to the class $\mathcal{T}^+$ of 244Xm.

244Y Further exercises (a) Let (X,Σ,μ) be a measure space, and $(X,\tilde\Sigma,\tilde\mu)$ its c.l.d. version. Show that $\mathcal{L}^p(\mu)\subseteq\mathcal{L}^p(\tilde\mu)$ and that this embedding induces a Banach lattice isomorphism between $L^p(\mu)$ and $L^p(\tilde\mu)$, for every $p\in[1,\infty[$.

(b) Let (X,Σ,μ) be any measure space, and $p\in[1,\infty[$. Show that $L^p(\mu)$ has the countable sup property in the sense of 241Ye. (*Hint*: 242Yh.)

(c) Suppose that (X,Σ,μ) is a measure space, and that $p\in\,]0,1[$, $q<0$ are such that $\frac1p+\frac1q=1$. (i) Show that $ab\ge\frac1p a^p+\frac1q b^q$ for all real $a\ge0$, $b>0$. (*Hint*: set $p'=\frac1p$, $q'=\frac{p'}{p'-1}$, $c=(ab)^p$, $d=b^{-p}$ and apply 244Ea.) (ii) Show that if f, $g\in\mathcal{L}^0(\mu)$ are non-negative and $E=\{x:x\in\operatorname{dom}g,\ g(x)>0\}$, then

$$(\int_E f^p)^{1/p}(\int_E g^q)^{1/q}\le\int f\times g.$$

(iii) Show that if f, $g\in\mathcal{L}^0(\mu)$ are non-negative, then

$$(\int f^p)^{1/p}+(\int g^p)^{1/p}\le(\int(f+g)^p)^{1/p}.$$

(d) Let (X,Σ,μ) be a measure space, and Y a subset of X; write μ_Y for the subspace measure on Y. Show that the canonical map T from $L^0(\mu)$ onto $L^0(\mu_Y)$ (241Yg) includes a surjection from $L^p(\mu)$ onto $L^p(\mu_Y)$ for every $p\in[1,\infty]$, and also a map from $M^{1,\infty}(\mu)$ to $M^{1,\infty}(\mu_Y)$ which belongs to the class $\mathcal{T}^+$ of 244Xm. Show that the following are equiveridical: (i) there is some $p\in[1,\infty[$ such that $T\restriction L^p(\mu)$ is injective; (ii) $T:L^p(\mu)\to L^p(\mu_Y)$ is norm-preserving for every $p\in[1,\infty[$; (iii) $F\cap Y\ne\emptyset$ whenever $F\in\Sigma$ and $0<\mu F<\infty$.

(e) Let (X,Σ,μ) be any measure space, and $p\in[1,\infty[$. Show that the norm $\|\,\|_p$ on $L^p(\mu)$ is order-continuous in the sense of 242Yg.

(f) Let (X,Σ,μ) be any measure space, and $p\in[1,\infty]$. Show that if $A\subseteq L^p(\mu)$ is upwards-directed and norm-bounded, then it is bounded above. (*Hint*: 242Yf.)

(g) Let (X,Σ,μ) be any measure space, and $p\in[1,\infty]$. Show that if a non-empty set $A\subseteq L^p(\mu)$ is upwards-directed and has a supremum in $L^p(\mu)$, then $\|\sup A\|_p\le\sup_{u\in A}\|u\|_p$. (*Hint*: consider first the case $0\in A$.)

(h) Let (X,Σ,μ) be a measure space and $u\in L^0(\mu)$. (i) Show that $I=\{p:p\in[1,\infty[,\ u\in L^p(\mu)\}$ is an interval. Give examples to show that it may be open, closed or half-open. (ii) Show that $p\mapsto p\ln\|u\|_p: I\to\mathbb{R}$ is convex. (*Hint*: if $p<q$ and $t\in\,]0,1[$, observe that $\int|u|^{tp+(1-t)q}\le\||u|^{pt}\|_{1/t}\||u|^{q(1-t)}\|_{1/(1-t)}$.) (iii) Show that if $p\le q\le r$ in I, then $\|u\|_q\le\max(\|u\|_p,\|u\|_r)$.

(i) Let $[a,b]$ be a non-trivial closed interval in $\mathbb{R}$ and $F:[a,b]\to\mathbb{R}$ a function; take $p\in\,]1,\infty[$. Show that the following are equiveridical: (i) F is absolutely continuous and its derivative F' belongs to $\mathcal{L}^p(\mu)$, where μ is Lebesgue measure on $[a,b]$ (ii)

$$c=\sup\{\sum_{i=1}^n\frac{|F(a_i)-F(a_{i-1})|^p}{(a_i-a_{i-1})^{p-1}}: a\le a_0<a_1<\ldots<a_n\le b\}$$

is finite, and that in this case $c = \|F'\|_p$. (*Hint*: (i) if F is absolutely continuous and $F' \in \mathcal{L}^p$, use Hölder's inequality to show that $|F(b') - F(a')|^p \le (b'-a')^{p-1} \int_{a'}^{b'} |F'|^p$ whenever $a \le a' \le b' \le b$. (ii) If F satisfies the condition, show that $(\sum_{i=0}^n |F(b_i) - F(a_i)|)^p \le c(\sum_{i=0}^n (b_i - a_i))^{p-1}$ whenever $a \le a_0 \le b_0 \le a_1 \le \ldots \le b_n \le b$, so that F is absolutely continuous. Take a sequence $\langle F_n \rangle_{n\in\mathbb{N}}$ of polygonal functions approximating F; use 223Xj to show that $F_n' \to F'$ a.e., so that $\int |F'|^p \le \sup_{n\in\mathbb{N}} \int |F_n'|^p \le c^p$.)

(j) Let G be an open set in $\mathbb{R}^r$ and write μ for Lebesgue measure on G. Let $C_k(G)$ be the set of continuous functions $f : G \to \mathbb{R}$ such that $\inf\{\|x-y\| : x \in G,\ f(x) \ne 0,\ y \in \mathbb{R}^r \setminus G\} > 0$ (counting $\inf\emptyset$ as ∞). Show that for any $p \in [1,\infty[$ the set $\{f^\bullet : f \in C_k(G)\}$ is a dense linear subspace of $L^p(\mu)$.

(k) Let U be any Hilbert space. (i) Show that if $C \subseteq U$ is convex (that is, $tu + (1-t)v \in C$ whenever $u, v \in C$ and $t \in [0,1]$; see 233Xd), closed and not empty, and $u \in U$, then there is a unique $v \in C$ such that $\|u-v\| = \inf_{w\in C} \|u-w\|$, and $(u-v|v-w) \ge 0$ for every $w \in C$. (ii) Show that if $h \in U^*$ there is a unique $v \in U$ such that $h(w) = (w|v)$ for every $w \in U$. (*Hint*: apply (i) with $C = \{w : h(w) = 1\}$, $u = 0$.) (iii) Show that if $V \subseteq U$ is a closed linear subspace then there is a unique linear projection P on U such that $P[U] = V$ and $(u - Pu|v) = 0$ for all $u \in U$, $v \in V$ (P is 'orthogonal'). (*Hint*: take Pu to be the point of V nearest to u.)

(l) Let (X, Σ, μ) be a probability space, and T a σ-subalgebra of Σ. Use part (iii) of 244Yk to show that there is an orthogonal projection $P : L^2(\mu) \to L^2(\mu\restriction\mathrm{T})$ such that $\int_F Pu = \int_F u$ for every $u \in L^2(\mu)$ and $F \in \mathrm{T}$. Show that $Pu \ge 0$ whenever $u \ge 0$ and that $\int Pu = \int u$ for every u, so that P has a unique extension to a continuous operator from $L^1(\mu)$ onto $L^1(\mu\restriction\mathrm{T})$. Use this to develop the theory of conditional expectations without using the Radon-Nikodým theorem.

(m) (Roselli & Willem 02) (i) Set $C = [0,\infty[^2 \subseteq \mathbb{R}^2$. Let $\phi : C \to \mathbb{R}$ be a continuous function such that $\phi(tz) = t\phi(z)$ for all $z \in C$. Show that ϕ is convex (definition: 233Xd) iff $t \mapsto \phi(1,t) : [0,\infty[\to \mathbb{R}$ is convex. (ii) Show that if $p \in]1,\infty[$ and $q = \frac{p}{p-1}$ then $(s,t) \mapsto -s^{1/p}t^{1/q}$, $(s,t) \mapsto -(s^{1/p} + t^{1/p})^p : C \to \mathbb{R}$ are convex. (iii) Show that if $p \in [1,2]$ then $(s,t) \mapsto |s^{1/p} + t^{1/p}|^p + |s^{1/p} - t^{1/p}|^p$ is convex. (iv) Show that if $p \in [2,\infty[$ then $(s,t) \mapsto -|s^{1/p} + t^{1/p}|^p - |s^{1/p} - t^{1/p}|^p$ is convex. (v) Use (ii) and 233Yj to prove Hölder's and Minkowski's inequalities. (vi) Use (iii) and (iv) to prove Hanner's inequalities. (vii) Adapt the method to answer (ii) and (iii) of 244Yc.

(n)(i) Show that any inner product space is uniformly convex. (ii) Let U be a uniformly convex Banach space, $C \subseteq U$ a non-empty closed convex set, and $u \in U$. Show that there is a unique $v_0 \in C$ such that $\|u - v_0\| = \inf_{v\in C} \|u - v\|$. (iii) Let U be a uniformly convex Banach space, and $A \subseteq U$ a non-empty bounded set. Set $\delta_0 = \inf\{\delta :$ there is some $u \in U$ such that $A \subseteq B(u,\delta) = \{v : \|v - u\| \le \delta\}\}$. Show that there is a unique $u_0 \in U$ such that $A \subseteq B(u_0, \delta_0)$.

(o) Let (X, Σ, μ) be a measure space, and $u \in L^0(\mu)$. Suppose that $\langle p_n \rangle_{n\in\mathbb{N}}$ is a sequence in $[1,\infty]$ with limit $p \in [1,\infty]$. Show that if $\limsup_{n\to\infty} \|u\|_{p_n}$ is finite then $\lim_{n\to\infty} \|u\|_{p_n}$ is defined and is equal to $\|u\|_p$.

244 Notes and comments At this point I feel we must leave the investigation of further function spaces. The next stage would have to be a systematic abstract analysis of general Banach lattices. The L^p spaces give a solid foundation for such an analysis, since they introduce the basic themes of norm-completeness, order-completeness and identification of dual spaces. I have tried in the exercises to suggest the importance of the next layer of concepts: order-continuity of norms and the relationship between norm-boundedness and order-boundedness. What I have not had space to discuss is the subject of order-preserving linear operators between Riesz spaces, which is the key to understanding the order structure of the dual spaces here. (But you can make a start by re-reading the theory of conditional expectation operators in 242J-242L, 243J and 244M.) All these topics are treated in Fremlin 74 and in Chapters 35 and 36 of the next volume.

I remember that one of my early teachers of analysis said that the L^p spaces (for $p \ne 1, 2, \infty$) had somehow got into the syllabus and had never been got out again. I would myself call them classics, in the sense that they have been part of the common experience of all functional analysts since functional analysis began; and while you are at liberty to dislike them, you can no more ignore them than you can ignore Milton if you are studying English poetry. Hölder's inequality, in particular, has a wealth of applications; not only 244F and 244K, but also 244Xc-244Xd and 244Yh-244Yi, for instance.

The L^p spaces, for $1 \le p \le \infty$, form a kind of continuum. In terms of the concepts dealt with here, there is no distinction to be drawn between different L^p spaces for $1 < p < \infty$ except the observation that the norm of L^2 is an inner product norm, corresponding to a Euclidean geometry on its finite-dimensional subspaces. To discriminate between the other L^p spaces we need much more refined concepts in the geometry of normed spaces.

In terms of the theorems given here, L^1 seems closer to the middle range of L^p for $1 < p < \infty$ than L^∞ does; thus, for all $1 \le p < \infty$, we have L^p Dedekind complete (independent of the measure space involved), the space S of equivalence classes of simple functions is dense in L^p (again, for every measure space), and the dual $(L^p)^*$ is (almost) identifiable as another function space. All of these should be regarded as consequences in one way or another of the order-continuity of the norm of L^p for $p < \infty$. The chief obstacle to the universal identification of $(L^1)^*$ with L^∞ is that for non-σ-finite measure spaces the space L^∞ can be inadequate, rather than any pathology in the L^1 space itself. (This point, at least, I mean to return to in Volume 3.) There is also the point that for a non-semi-finite measure space the purely infinite sets can contribute to L^∞ without any corresponding contribution to L^1. For $1 < p < \infty$, neither of these problems can arise. Any member of any such L^p is supported entirely by a σ-finite part of the measure space, and the same applies to the dual – see part (c) of the proof of 244K.

Of course L^1 does have a markedly different geometry from the other L^p spaces. The first sign of this is that it is not reflexive as a Banach space (except when it is finite-dimensional), whereas for $1 < p < \infty$ the identifications of $(L^p)^*$ with L^q and of $(L^q)^*$ with L^p, where $q = p/(p-1)$, show that the canonical embedding of L^p in $(L^p)^{**}$ is surjective, that is, that L^p is reflexive. But even when L^1 is finite-dimensional the unit balls of L^1 and L^∞ are clearly different in kind from the unit balls of L^p for $1 < p < \infty$; they have corners instead of being smoothly rounded (244Xh-244Xi). A direct expression of the difference is in 244O. As usual, the case $p = 2$ is both much more important than the general case and enormously easier (244Yn(i)); and note how Hanner's inequalities reverse at $p = 2$. (See 244Yc for the reversal of Hölder's and Minkowski's inequalities at $p = 1$.) There are occasions on which it is useful to know that $\| \|_1$ and $\| \|_\infty$ can be approximated, in an exactly describable way, by uniformly convex norms (244Yo). I have written out a proof of 244O based on ingenuity and advanced calculus, like that of 244E. With a bit more about convex sets and functions, sketched in 233Yf-233Yj, there is a striking alternative proof (244Ym). Of course the proof of 244Ea also uses convexity, upside down.

The proof of 244K, identifying $(L^p)^*$, is a fairly long haul, and it is natural to ask whether we really have to work so hard, especially since in the case of L^2 we have a much easier argument (244Yk). Of course we can go faster if we know a bit more about Banach lattices (§369 in Volume 3 has the relevant facts), though this route uses some theorems quite as hard as 244K as given. There are alternative routes using the geometry of the L^p spaces, following the ideas of 244Yk, but I do not think they are any easier, and the argument I have presented here at least has the virtue of using some of the same ideas as the identification of $(L^1)^*$ in 243G. The difference is that whereas in 243G we may have to piece together a large family of functions g_F (part (b-v) of the proof), in 244K there are only countably many g_n; consequently we can make the argument work for arbitrary measure spaces, not just localizable ones.

The geometry of Hilbert space gives us an approach to conditional expectations which does not depend on the Radon-Nikodým theorem (244Yl). To turn these ideas into a proof of the Radon-Nikodým theorem itself, however, requires qualities of determination and ingenuity which can be better employed elsewhere.

The convexity arguments of 233J/242K can be used on many operators besides conditional expectations (see 244Xm). The class '$\mathcal{T}^+$' described there is not in fact the largest for which these arguments work; I take the ideas farther in Chapter 37. There is also a great deal more to be said if you put an arbitrary pair of L^p spaces in place of L^1 and L^∞ in 244Xl. 244Yh is a start, but for the real thing (the 'Riesz convexity theorem') I refer you to Zygmund 59, XII.1.11 or Dunford & Schwartz 57, VI.10.11.

245 Convergence in measure

I come now to an important and interesting topology on the spaces $\mathcal{L}^0$ and L^0. I start with the definition (245A) and with properties which echo those of the L^p spaces for $p \ge 1$ (245D-245E). In 245G-245J I describe the most useful relationships between this topology and the norm topologies of the L^p spaces. For σ-finite spaces, it is metrizable (245Eb) and sequential convergence can be described in terms of pointwise convergence of sequences of functions (245K-245L).

245A Definitions Let (X, Σ, μ) be a measure space.

(a) For any measurable set $F \subseteq X$ of finite measure, we have a functional τ_F on $\mathcal{L}^0 = \mathcal{L}^0(\mu)$ defined by setting

$$\tau_F(f) = \int |f| \wedge \chi F$$

for every $f \in \mathcal{L}^0$. (The integral exists in $\mathbb{R}$ because $|f| \wedge \chi F$ belongs to $\mathcal{L}^0$ and is dominated by the integrable function χF). Now $\tau_F(f + g) \le \tau_F(f) + \tau_F(g)$ whenever $f, g \in \mathcal{L}^0$. **P** We need only observe that

$$\min(|(f + g)(x)|, \chi F(x)) \le \min(|f(x)|, \chi F(x)) + \min(|g(x)|, \chi F(x))$$

for every $x \in \operatorname{dom} f \cap \operatorname{dom} g$, which is almost every $x \in X$. **Q** Consequently, setting $\rho_F(f, g) = \tau_F(f - g)$, we have

$$\rho_F(f,h) = \tau_F((f-g)+(g-h)) \le \tau_F(f-g) + \tau_F(g-h) = \rho_F(f,g) + \rho_F(g,h),$$

$$\rho_F(f,g) = \tau_F(f-g) \ge 0,$$

$$\rho_F(f,g) = \tau_F(f-g) = \tau_F(g-f) = \rho_F(g,f)$$

for all f, g, $h \in \mathcal{L}^0$; that is, ρ_F is a pseudometric on $\mathcal{L}^0$.

(b) The family

$$\{\rho_F : F \in \Sigma,\ \mu F < \infty\}$$

now defines a topology on $\mathcal{L}^0$ (2A3F); I will call it the topology of **convergence in measure** on $\mathcal{L}^0$.

(c) If f, $g \in \mathcal{L}^0$ and $f =_{\text{a.e.}} g$, then $|f| \wedge \chi F =_{\text{a.e.}} |g| \wedge \chi F$ and $\tau_F(f) = \tau_F(g)$, for every set F of finite measure. Consequently we have functionals $\bar\tau_F$ on $L^0 = L^0(\mu)$ defined by writing

$$\bar\tau_F(f^\bullet) = \tau_F(f)$$

whenever $f \in \mathcal{L}^0$, $F \in \Sigma$ and $\mu F < \infty$. Corresponding to these we have pseudometrics $\bar\rho_F$ defined by either of the formulae

$$\bar\rho_F(u,v) = \bar\tau_F(u-v), \quad \bar\rho_F(f^\bullet, g^\bullet) = \rho_F(f,g)$$

for u, $v \in L^0$, f, $g \in \mathcal{L}^0$ and F of finite measure. The family of these pseudometrics defines the **topology of convergence in measure** on L^0.

(d) I shall allow myself to say that a sequence (in $\mathcal{L}^0$ or L^0) **converges in measure** if it converges for the topology of convergence in measure (in the sense of 2A3M).

245B Remarks (a) Of course the topologies of $\mathcal{L}^0$, L^0 are about as closely related as it is possible for them to be. Not only is the topology of L^0 the quotient of the topology on $\mathcal{L}^0$ (that is, a set $G \subseteq L^0$ is open iff $\{f : f^\bullet \in G\}$ is open in $\mathcal{L}^0$), but every open set in $\mathcal{L}^0$ is the inverse image under the quotient map of an open set in L^0.

(b) It is convenient to note that if $F_0, \dots, F_n$ are measurable sets of finite measure with union F, then, in the notation of 245A, $\tau_{F_i} \le \tau_F$ for every i; this means that a set $G \subseteq \mathcal{L}^0$ is open for the topology of convergence in measure iff for every $f \in G$ we can find a single set F of finite measure and a $\delta > 0$ such that

$$\rho_F(g,f) \le \delta \Longrightarrow g \in G.$$

Similarly, a set $G \subseteq L^0$ is open for the topology of convergence in measure iff for every $u \in G$ we can find a set F of finite measure and a $\delta > 0$ such that

$$\bar\rho_F(v,u) \le \delta \Longrightarrow v \in G.$$

(c) The phrase 'topology of convergence in measure' agrees well enough with standard usage when (X,Σ,μ) is totally finite. But a **warning!** the phrase 'topology of convergence in measure' is also used for the topology defined by the metric of 245Ye below, even when $\mu X = \infty$. I have seen the phrase **local convergence in measure** used for the topology of 245A. Most authors ignore non-σ-finite spaces in this context. However I hold that 245D-245E below are of sufficient interest to make the extension worth while.

245C Pointwise convergence The topology of convergence in measure is almost definable in terms of 'pointwise convergence', which is one of the roots of measure theory. The correspondence is closest in σ-finite measure spaces (see 245K), but there is still a very important relationship in the general case, as follows. Let (X,Σ,μ) be a measure space, and write $\mathcal{L}^0 = \mathcal{L}^0(\mu)$, $L^0 = L^0(\mu)$.

(a) If $\langle f_n\rangle_{n\in\mathbb{N}}$ is a sequence in $\mathcal{L}^0$ converging almost everywhere to $f \in \mathcal{L}^0$, then $\langle f_n\rangle_{n\in\mathbb{N}} \to f$ in measure. **P** By 2A3Mc, I have only to show that $\lim_{n\to\infty} \rho_F(f_n,f) = 0$ whenever $\mu F < \infty$. But $\langle |f_n - f| \wedge \chi F\rangle_{n\in\mathbb{N}}$ converges to 0 a.e. and is dominated by the integrable function χF, so by Lebesgue's Dominated Convergence Theorem

$$\lim_{n\to\infty} \rho_F(f_n,f) = \lim_{n\to\infty} \int |f_n - f| \wedge \chi F = 0. \ \mathbf{Q}$$

(b) To formulate a corresponding result applicable to L^0, we need the following concept. If $\langle f_n\rangle_{n\in\mathbb{N}}$, $\langle g_n\rangle_{n\in\mathbb{N}}$ are sequences in $\mathcal{L}^0$ such that $f_n^\bullet = g_n^\bullet$ for every n, and f, $g \in \mathcal{L}^0$ are such that $f^\bullet = g^\bullet$, and $\langle f_n\rangle_{n\in\mathbb{N}} \to f$ a.e., then $\langle g_n\rangle_{n\in\mathbb{N}} \to g$ a.e., because

$$\{x : x \in \operatorname{dom} f \cap \operatorname{dom} g \cap \bigcap_{n\in\mathbb{N}} \operatorname{dom} f_n \cap \bigcap_{n\in\mathbb{N}} g_n,$$
$$g(x) = f(x) = \lim_{n\to\infty} f_n(x),\ f_n(x) = g_n(x)\ \forall\ n \in \mathbb{N}\}$$

is conegligible. Consequently we have a definition applicable to sequences in L^0; we can say that, for f, $f_n \in \mathcal{L}^0$, $\langle f_n^\bullet\rangle_{n\in\mathbb{N}}$ is **order*-convergent**, or **order*-converges**, to $f^\bullet$ iff $f =_{\text{a.e.}} \lim_{n\to\infty} f_n$. In this case, of course, $\langle f_n\rangle_{n\in\mathbb{N}} \to f$ in measure. Thus, in L^0, a sequence $\langle u_n\rangle_{n\in\mathbb{N}}$ which order*-converges to $u \in L^0$ also converges to u in measure.

Remark I suggest alternative descriptions of order-convergence in 245Xc; the conditions (iii)-(vi) there are in forms adapted to more general structures.

(c) For a typical example of a sequence which is convergent in measure without being order-convergent, consider the following. Take μ to be Lebesgue measure on $[0,1]$, and set $f_n(x) = 2^m$ if $x \in [2^{-m}k, 2^{-m}(k+1)]$, 0 otherwise, where $k = k(n) \in \mathbb{N}$, $m = m(n) \in \mathbb{N}$ are defined by saying that $n+1 = 2^m + k$ and $0 \le k < 2^m$. Then $\langle f_n\rangle_{n\in\mathbb{N}} \to 0$ for the topology of convergence in measure (since $\rho_F(f_n, 0) \le 2^{-m}$ if $F \subseteq [0,1]$ is measurable and $2^m - 1 \le n$), though $\langle f_n\rangle_{n\in\mathbb{N}}$ is not convergent to 0 almost everywhere (indeed, $\limsup_{n\to\infty} f_n = \infty$ everywhere).

245D Proposition Let (X, Σ, μ) be any measure space.
(a) The topology of convergence in measure is a linear space topology on $L^0 = L^0(\mu)$.
(b) The maps $\vee$, $\wedge : L^0 \times L^0 \to L^0$, and $u \mapsto |u|$, $u \mapsto u^+$, $u \mapsto u^- : L^0 \to L^0$ are all continuous.
(c) The map $\times : L^0 \times L^0 \to L^0$ is continuous.
(d) For any continuous function $h : \mathbb{R} \to \mathbb{R}$, the corresponding function $\bar{h} : L^0 \to L^0$ (241I) is continuous.

proof (a) The point is that the functionals $\bar\tau_F$, as defined in 245Ac, satisfy the conditions of 2A5B below. **P** Fix a set $F \in \Sigma$ of finite measure. I noted in 245Aa that

$$\tau_F(f+g) \le \tau_F(f) + \tau_F(g) \text{ for all } f,\ g \in \mathcal{L}^0,$$

so

$$\bar\tau_F(u+v) \le \bar\tau_F(u) + \bar\tau_F(v) \text{ for all } u,\ v \in L^0.$$

Next,

$$\bar\tau_F(cu) \le \bar\tau_F(u) \text{ whenever } u \in L^0 \text{ and } |c| \le 1 \qquad (*)$$

because $|cf| \wedge \chi F \le_{\text{a.e.}} |f| \wedge \chi F$ whenever $f \in \mathcal{L}^0$ and $|c| \le 1$. Finally, given $u \in L^0$ and $\epsilon > 0$, let $f \in \mathcal{L}^0$ be such that $f^\bullet = u$. Then

$$\lim_{n\to\infty} |2^{-n} f| \wedge \chi F = 0 \text{ a.e.},$$

so by Lebesgue's Dominated Convergence Theorem

$$\lim_{n\to\infty} \bar\tau_F(2^{-n}u) = \lim_{n\to\infty} \int |2^{-n}f| \wedge \chi F = 0,$$

and there is an n such that $\bar\tau_F(2^{-n}u) \le \epsilon$. It follows (by (*) just above) that $\bar\tau_F(cu) \le \epsilon$ whenever $|c| \le 2^{-n}$. As ϵ is arbitrary, $\lim_{c\to 0} \bar\tau_F(u) = 0$ for every $u \in L^0$; which is the third condition in 2A5B. **Q**

Now 2A5B tells us that the topology defined by the $\bar\tau_F$ is a linear space topology.

(b) For any u, $v \in L^0$, $||u| - |v|| \le |u - v|$, so $\bar\rho_F(|u|, |v|) \le \bar\rho_F(u, v)$ for every set F of finite measure. By 2A3H, $|\ | : L^0 \to L^0$ is continuous. Now

$$u \vee v = \frac{1}{2}(u + v + |u - v|), \quad u \wedge v = \frac{1}{2}(u + v - |u - v|),$$

$$u^+ = u \wedge 0, \quad u^- = (-u) \wedge 0.$$

As addition and subtraction are continuous, so are $\vee$, $\wedge$, $^+$ and $^-$.

(c) Take u_0, $v_0 \in L^0$, $F \in \Sigma$ a set of finite measure and $\epsilon > 0$. Represent u_0 and v_0 as $f_0^\bullet$, $g_0^\bullet$ respectively, where f_0, $g_0 : X \to \mathbb{R}$ are Σ-measurable (241Bk). If we set

$$F_m = \{x : x \in F,\ |f_0(x)| + |g_0(x)| \le m\},$$

then $\langle F_m\rangle_{m\in\mathbb{N}}$ is a non-decreasing sequence of sets with union F, so there is an $m\in\mathbb{N}$ such that $\mu(F\setminus F_m)\le\frac{1}{2}\epsilon$. Let $\delta>0$ be such that $(2m+\mu F)\delta^2+2\delta\le\frac{1}{2}\epsilon$.

Now suppose that u, $v\in L^0$ are such that $\bar\rho_F(u,u_0)\le\delta^2$ and $\bar\rho_F(v,v_0)\le\delta^2$. Let f, $g:X\to\mathbb{R}$ be measurable functions such that $f^\bullet=u$ and $v^\bullet=v$. Then

$$\mu\{x:x\in F,\,|f(x)-f_0(x)|\ge\delta\}\le\delta,\quad \mu\{x:x\in F,\,|g(x)-g_0(x)|\ge\delta\}\le\delta,$$

so that

$$\mu\{x:x\in F,\,|f(x)-f_0(x)||g(x)-g_0(x)|\ge\delta^2\}\le 2\delta$$

and

$$\int_F\min(1,|f-f_0|\times|g-g_0|)\le\delta^2\mu F+2\delta.$$

Also

$$|f\times g-f_0\times g_0|\le|f-f_0|\times|g-g_0|+|f_0|\times|g-g_0|+|f-f_0|\times|g_0|,$$

so that

$$\begin{aligned}\bar\rho_F(u\times v,u_0\times v_0)&=\int_F\min(1,|f\times g-f_0\times g_0|)\\&\le\frac{1}{2}\epsilon+\int_{F_m}\min(1,|f\times g-f_0\times g_0|)\\&\le\frac{1}{2}\epsilon+\int_{F_m}\min(1,|f-f_0|\times|g-g_0|+m|g-g_0|+m|f-f_0|)\\&\le\frac{1}{2}\epsilon+\int_F\min(1,|f-f_0|\times|g-g_0|)\\&\qquad\qquad+m\int_F\min(1,|g-g_0|)+m\int_F\min(1,|f-f_0|)\\&\le\frac{1}{2}\epsilon+\delta^2\mu F+2\delta+2m\delta^2\le\epsilon.\end{aligned}$$

As F and ϵ are arbitrary, $\times$ is continuous at (u_0,v_0); as u_0 and v_0 are arbitrary, $\times$ is continuous.

(d) Take $u\in L^0$, $F\in\Sigma$ of finite measure and $\epsilon>0$. Then there is a $\delta>0$ such that $\rho_F(\bar h(v),\bar h(u))\le\epsilon$ whenever $\rho_F(v,u)\le\delta$. **P?** Otherwise, we can find, for each $n\in\mathbb{N}$, a v_n such that $\bar\rho_F(v_n,u)\le 4^{-n}$ but $\bar\rho_F(\bar h(v_n),\bar h(u))>\epsilon$. Express u as $f^\bullet$ and v_n as $g_n^\bullet$ where f, $g_n:X\to\mathbb{R}$ are measurable. Set

$$E_n=\{x:x\in F,\,|g_n(x)-f(x)|\ge 2^{-n}\}$$

for each n. Then $\bar\rho_F(v_n,u)\ge 2^{-n}\mu E_n$, so $\mu E_n\le 2^{-n}$ for each n, and $E=\bigcap_{n\in\mathbb{N}}\bigcup_{m\ge n}E_m$ is negligible. But $\lim_{n\to\infty}g_n(x)=f(x)$ for every $x\in F\setminus E$, so (because h is continuous) $\lim_{n\to\infty}h(g_n(x))=h(f(x))$ for every $x\in F\setminus E$. Consequently (by Lebesgue's Dominated Convergence Theorem, as always)

$$\lim_{n\to\infty}\bar\rho_F(\bar h(v_n),\bar h(u))=\lim_{n\to\infty}\int_F\min(1,|h(g_n(x))-h(f(x))|\mu(dx)=0,$$

which is impossible. **XQ**

By 2A3H, $\bar h$ is continuous.

Remark I cannot say that the topology of convergence in measure on $\mathcal{L}^0$ is a linear space topology solely because (on the definitions I have chosen) $\mathcal{L}^0$ is not in general a linear space.

245E I turn now to the principal theorem relating the properties of the topological linear space $L^0(\mu)$ to the classification of measure spaces in Chapter 21.

Theorem Let (X,Σ,μ) be a measure space. Let $\mathfrak{T}$ be the topology of convergence in measure on $L^0=L^0(\mu)$, as described in 245A.

(a) (X,Σ,μ) is semi-finite iff $\mathfrak{T}$ is Hausdorff.

(b) (X,Σ,μ) is σ-finite iff $\mathfrak{T}$ is metrizable.

(c) (X,Σ,μ) is localizable iff $\mathfrak{T}$ is Hausdorff and L^0 is complete under $\mathfrak{T}$.

proof I use the pseudometrics ρ_F on $\mathcal{L}^0 = \mathcal{L}^0(\mu)$, $\bar\rho_F$ on L^0 described in 245A.

(a)(i) Suppose that (X,Σ,μ) is semi-finite and that u, v are distinct members of L^0. Express them as $f^\bullet$ and $g^\bullet$ where f and g are measurable functions from X to $\mathbb{R}$. Then $\mu\{x : f(x) \ne g(x)\} > 0$ so, because (X,Σ,μ) is semi-finite, there is a set $F \in \Sigma$ of finite measure such that $\mu\{x : x \in F,\ f(x) \ne g(x)\} > 0$. Now

$$\bar\rho_F(u,v) = \int_F \min(|f(x) - g(x)|, 1)dx > 0$$

(see 122Rc). As u and v are arbitrary, $\mathfrak{T}$ is Hausdorff (2A3L).

(ii) Suppose that $\mathfrak{T}$ is Hausdorff and that $E \in \Sigma$, $\mu E > 0$. Then $u = \chi E^\bullet \ne 0$ so there is an $F \in \Sigma$ such that $\mu F < \infty$ and $\bar\rho_F(u,0) \ne 0$, that is, $\mu(E \cap F) > 0$. Now $E \cap F$ is a non-negligible set of finite measure included in E. As E is arbitrary, (X,Σ,μ) is semi-finite.

(b)(i) Suppose that (X,Σ,μ) is σ-finite. Let $\langle E_n\rangle_{n\in\mathbb{N}}$ be a non-decreasing sequence of sets of finite measure covering X. Set

$$\bar\rho(u,v) = \sum_{n=0}^{\infty} \frac{\bar\rho_{E_n}(u,v)}{1 + 2^n \mu E_n}$$

for u, $v \in L^0$. Then $\bar\rho$ is a metric on L^0. **P** Because every $\bar\rho_{E_n}$ is a pseudometric, so is $\bar\rho$. If $\bar\rho(u,v) = 0$, express u as $f^\bullet$, v as $g^\bullet$ where f, $g \in \mathcal{L}^0(\mu)$; then

$$\int |f - g| \wedge \chi E_n = \bar\rho_{E_n}(u,v) = 0,$$

so $f = g$ almost everywhere in E_n, for every n. Because $X = \bigcup_{n\in\mathbb{N}} E_n$, $f =_{\text{a.e.}} g$ and $u = v$. **Q**

If $F \in \Sigma$ and $\mu F < \infty$ and $\epsilon > 0$, take n such that $\mu(F \setminus E_n) \le \frac{1}{2}\epsilon$. If u, $v \in L^0$ and $\bar\rho(u,v) \le \epsilon/2(1 + 2^n\mu E_n)$, then $\bar\rho_F(u,v) \le \epsilon$. **P** Express u as $f^\bullet$, $v = g^\bullet$ where f, $g \in \mathcal{L}^0$. Then

$$\int |u - v| \wedge \chi E_n = \bar\rho_{E_n}(u,v) \le (1 + 2^n \mu E_n)\bar\rho(u,v) \le \frac{\epsilon}{2},$$

while

$$\int |f - g| \wedge \chi(F \setminus E_n) \le \mu(F \setminus E_n) \le \frac{\epsilon}{2},$$

so

$$\bar\rho_F(u,v) = \int |f - g| \wedge \chi F \le \int |f - g| \wedge \chi E_n + \int |f - g| \wedge \chi(F \setminus E_n) \le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon. \textbf{ Q}$$

In the other direction, given $\epsilon > 0$, take $n \in \mathbb{N}$ such that $2^{-n} \le \frac{1}{2}\epsilon$; then $\bar\rho(u,v) \le \epsilon$ whenever $\bar\rho_{E_n}(u,v) \le \epsilon/2(n+1)$.

These show that $\bar\rho$ defines the same topology as the $\bar\rho_F$ (2A3Ib), so that $\mathfrak{T}$, the topology defined by the $\bar\rho_F$, is metrizable.

(ii) Suppose that $\mathfrak{T}$ is metrizable. Let $\bar\rho$ be a metric defining $\mathfrak{T}$. For each $n \in \mathbb{N}$ there must be a measurable set F_n of finite measure and a $\delta_n > 0$ such that

$$\bar\rho_{F_n}(u,0) \le \delta_n \Longrightarrow \bar\rho(u,0) \le 2^{-n}.$$

Set $E = X \setminus \bigcup_{n\in\mathbb{N}} F_n$. **?** If E is not negligible, then $u = \chi E^\bullet \ne 0$; because $\bar\rho$ is a metric, there is an $n \in \mathbb{N}$ such that $\bar\rho(u,0) > 2^{-n}$; now

$$\mu(E \cap F_n) = \bar\rho_{F_n}(u,0) > \delta_n.$$

But $E \cap F_n = \emptyset$. **X**

So $\mu E = 0 < \infty$. Now $X = E \cup \bigcup_{n\in\mathbb{N}} F_n$ is a countable union of sets of finite measure, so μ is σ-finite.

(c) By (a), either hypothesis ensures that μ is semi-finite and that $\mathfrak{T}$ is Hausdorff.

(i) Suppose that (X,Σ,μ) is localizable. Let $\mathcal{F}$ be a Cauchy filter on L^0 (2A5F). For each measurable set F of finite measure, choose a sequence $\langle A_n(F)\rangle_{n\in\mathbb{N}}$ of members of $\mathcal{F}$ such that $\bar\rho_F(u,v) \le 4^{-n}$ for every u, $v \in A_n(F)$ and every n (2A5G). Choose $u_{Fn} \in \bigcap_{k\le n} A_n(F)$ for each n; then $\bar\rho_F(u_{F,n+1}, u_{Fn}) \le 2^{-n}$ for each n. Express each u_{Fn} as $f_{Fn}^\bullet$ where f_{Fn} is a measurable function from X to $\mathbb{R}$. Then

$$\mu\{x : x \in F,\ |f_{F,n+1}(x) - f_{Fn}(x)| \ge 2^{-n}\} \le 2^n \bar\rho_F(u_{F,n+1}, u_{Fn}) \le 2^{-n}$$

for each n. It follows that $\langle f_{Fn}\rangle_{n\in\mathbb{N}}$ must converge almost everywhere in F. **P** Set

$$H_n = \{x : x \in F,\ |f_{F,n+1}(x) - f_{Fn}(x)| \ge 2^{-n}\}.$$

Then $\mu H_n \le 2^{-n}$ for each n, so

$$\mu(\bigcap_{n\in\mathbb{N}}\bigcup_{m\ge n} H_m) \le \inf_{n\in\mathbb{N}}\sum_{m=n}^{\infty} 2^{-m} = 0.$$

If $x \in F \setminus \bigcap_{n\in\mathbb{N}}\bigcup_{m\ge n} H_m$, then there is some k such that $x \in F \setminus \bigcup_{m\ge k} H_m$, so that $|f_{F,m+1}(x) - f_{Fm}(x)| \le 2^{-m}$ for every $m \ge k$, and $\langle f_{Fn}(x)\rangle_{n\in\mathbb{N}}$ is Cauchy, therefore convergent. **Q**

Set $f_F(x) = \lim_{n\to\infty} f_{Fn}(x)$ for every $x \in F$ such that the limit is defined in $\mathbb{R}$, so that f_F is measurable and defined almost everywhere in F.

If E, F are two sets of finite measure and $E \subseteq F$, then $\bar\rho_E(u_{En}, u_{Fn}) \le 2\cdot 4^{-n}$ for each n. **P** $A_n(E)$ and $A_n(F)$ both belong to $\mathcal{F}$, so must have a point w in common; now

$$\begin{aligned}\bar\rho_E(u_{En}, u_{Fn}) &\le \bar\rho_E(u_{En}, w) + \bar\rho_E(w, u_{Fn})\\ &\le \bar\rho_E(u_{En}, w) + \bar\rho_F(w, u_{Fn}) \le 4^{-n} + 4^{-n}. \textbf{ Q}\end{aligned}$$

Consequently

$$\mu\{x : x \in E,\ |f_{Fn}(x) - f_{En}(x)| \ge 2^{-n}\} \le 2^n\bar\rho_E(u_{Fn}, u_{En}) \le 2^{-n+1}$$

for each n, and $\lim_{n\to\infty} f_{Fn} - f_{En} = 0$ almost everywhere in E; so that $f_E = f_F$ a.e. on E.

Consequently, if E and F are any two sets of finite measure, $f_E = f_F$ a.e. on $E \cap F$, because both are equal almost everywhere on $E \cap F$ to $f_{E\cup F}$.

Because μ is localizable, it follows that there is an $f \in \mathcal{L}^0$ such that $f = f_E$ a.e. on E for every measurable set E of finite measure (213N). Consider $u = f^\bullet \in L^0$. For any set E of finite measure,

$$\bar\rho_E(u, u_{En}) = \int_E \min(1, |f(x) - f_{En}(x)|)dx = \int_E \min(1, |f_E(x) - f_{En}(x)|)dx \to 0$$

as $n \to \infty$, using Lebesgue's Dominated Convergence Theorem. Now

$$\begin{aligned}\inf_{A\in\mathcal{F}}\sup_{v\in A}\bar\rho_E(v, u) &\le \inf_{n\in\mathbb{N}}\sup_{v\in A_{En}}\bar\rho_E(v, u)\\ &\le \inf_{n\in\mathbb{N}}\sup_{v\in A_{En}}(\bar\rho_E(v, u_{En}) + \bar\rho_E(u, u_{En}))\\ &\le \inf_{n\in\mathbb{N}}(4^{-n} + \bar\rho_E(u, u_{En})) = 0.\end{aligned}$$

As E is arbitrary, $\mathcal{F} \to u$. As $\mathcal{F}$ is arbitrary, L^0 is complete under $\mathfrak{T}$.

(ii) Now suppose that L^0 is complete under $\mathfrak{T}$ and let $\mathcal{E}$ be any family of sets in Σ. Let $\mathcal{E}'$ be

$$\{\bigcup\mathcal{E}_0 : \mathcal{E}_0 \text{ is a finite subset of } \mathcal{E}\}.$$

Then the union of any two members of $\mathcal{E}'$ belongs to $\mathcal{E}'$. Let $\mathcal{F}$ be the set

$$\{A : A \subseteq L^0,\ A \supseteq A_E \text{ for some } E \in \mathcal{E}'\},$$

where for $E \in \mathcal{E}'$ I write

$$A_E = \{\chi F^\bullet : F \in \mathcal{E}',\ F \supseteq E\}.$$

Then $\mathcal{F}$ is a filter on L^0, because $A_E \cap A_F = A_{E\cup F}$ for all E, $F \in \mathcal{E}'$.

In fact $\mathcal{F}$ is a Cauchy filter. **P** Let H be any set of finite measure and $\epsilon > 0$. Set $\gamma = \sup_{E\in\mathcal{E}'}\mu(H \cap E)$ and take $E \in \mathcal{E}'$ such that $\mu(H \cap E) \ge \gamma - \epsilon$. Consider $A_E \in \mathcal{F}$. If F, $G \in \mathcal{E}'$ and $F \supseteq E$, $G \supseteq E$ then

$$\begin{aligned}\bar\rho_H(\chi F^\bullet, \chi G^\bullet) &= \mu(H \cap (F \triangle G)) = \mu(H \cap (F \cup G)) - \mu(H \cap F \cap G)\\ &\le \gamma - \mu(H \cap E) \le \epsilon.\end{aligned}$$

Thus $\bar\rho_H(u, v) \le \epsilon$ for all u, $v \in A_E$. As H and ϵ are arbitrary, $\mathcal{F}$ is Cauchy. **Q**

Because L^0 is complete under $\mathfrak{T}$, $\mathcal{F}$ has a limit w say. Express w as $h^\bullet$, where $h : X \to \mathbb{R}$ is measurable, and consider $G = \{x : h(x) > \frac{1}{2}\}$.

? If $E \in \mathcal{E}$ and $\mu(E \setminus G) > 0$, let $F \subseteq E \setminus G$ be a set of non-zero finite measure. Then $\{u : \bar\rho_F(u, w) < \frac{1}{2}\mu F\}$ belongs to $\mathcal{F}$, so meets A_E; let $H \in \mathcal{E}'$ be such that $E \subseteq H$ and $\bar\rho_F(\chi H^\bullet, w) < \frac{1}{2}\mu F$. Then

$$\int_F \min(1, |1 - h(x)|) = \bar\rho_F(\chi H^\bullet, w) < \frac{1}{2}\mu F;$$

but because $F \cap G = \emptyset$, $1 - h(x) \ge \frac{1}{2}$ for every $x \in F$, so this is impossible. **X**

Thus $E \setminus G$ is negligible for every $E \in \mathcal{E}$.

Now suppose that $H \in \Sigma$ and $\mu(G \setminus H) > 0$. Then there is an $E \in \mathcal{E}$ such that $\mu(E \setminus H) > 0$. **P** Let $F \subseteq G \setminus H$ be a set of non-zero finite measure. Let $u \in A_\emptyset$ be such that $\bar\rho_F(u, w) < \frac{1}{2}\mu F$. Then u is of the form $\chi C^\bullet$ for some $C \in \mathcal{E}'$, and

$$\int_F \min(1, |\chi C(x) - h(x)|) < \frac{1}{2}\mu F.$$

As $h(x) \geq \frac{1}{2}$ for every $x \in F$, $\mu(C \cap F) > 0$. But C is a finite union of members of $\mathcal{E}$, so there is an $E \in \mathcal{E}$ such that $\mu(E \cap F) > 0$, and now $\mu(E \setminus H) > 0$. **Q**

As H is arbitrary, G is an essential supremum of $\mathcal{E}$ in Σ. As $\mathcal{E}$ is arbitrary, (X, Σ, μ) is localizable.

245F Alternative description of the topology of convergence in measure Let us return to arbitrary measure spaces (X, Σ, μ).

(a) For any $F \in \Sigma$ of finite measure and $\epsilon > 0$ define $\tau_{F\epsilon} : \mathcal{L}^0 \to [0, \infty[$ by

$$\tau_{F\epsilon}(f) = \mu^*\{x : x \in F \cap \operatorname{dom} f,\, |f(x)| > \epsilon\}$$

for $f \in \mathcal{L}^0$, taking μ^* to be the outer measure defined from μ (132B). If f, $g \in \mathcal{L}^0$ and $f =_{\text{a.e.}} g$, then

$$\{x : x \in F \cap \operatorname{dom} f,\, |f(x)| > \epsilon\} \triangle \{x : x \in F \cap \operatorname{dom} g,\, |g(x)| > \epsilon\}$$

is negligible, so $\tau_{F\epsilon}(f) = \tau_{F\epsilon}(g)$; accordingly we have a functional from L^0 to $[0, \infty[$, given by

$$\bar\tau_{F\epsilon}(u) = \tau_{F\epsilon}(f)$$

whenever $f \in \mathcal{L}^0$ and $u = f^\bullet \in L^0$.

(b) Now $\tau_{F\epsilon}$ is not (except in trivial cases) subadditive, so does not define a pseudometric on $\mathcal{L}^0$ or L^0. But we can say that, for $f \in \mathcal{L}^0$,

$$\tau_F(f) \leq \epsilon \min(1, \epsilon) \Longrightarrow \tau_{F\epsilon}(f) \leq \epsilon \Longrightarrow \tau_F(f) \leq \epsilon(1 + \mu F).$$

(The point is that if $E \subseteq \operatorname{dom} f$ is a measurable conegligible set such that $f \restriction E$ is measurable, then

$$\tau_F(f) = \int_{E \cap F} \min(f(x), 1) dx, \quad \tau_{F\epsilon}(f) = \mu\{x : x \in E \cap F,\, f(x) > \epsilon\}.)$$

This means that a set $G \subseteq \mathcal{L}^0$ is open for the topology of convergence in measure iff for every $f \in G$ we can find a set F of finite measure and ϵ, $\delta > 0$ such that

$$\tau_{F\epsilon}(g - f) \leq \delta \Longrightarrow g \in G.$$

Of course $\tau_{F\delta}(f) \geq \tau_{F\epsilon}(f)$ whenever $\delta \leq \epsilon$, so we can equally say: $G \subseteq \mathcal{L}^0$ is open for the topology of convergence in measure iff for every $f \in G$ we can find a set F of finite measure and $\epsilon > 0$ such that

$$\tau_{F\epsilon}(g - f) \leq \epsilon \Longrightarrow g \in G.$$

Similarly, $G \subseteq L^0$ is open for the topology of convergence in measure on L^0 iff for every $u \in G$ we can find a set F of finite measure and $\epsilon > 0$ such that

$$\bar\tau_{F\epsilon}(v - u) \leq \epsilon \Longrightarrow v \in G.$$

(c) It follows at once that a sequence $\langle f_n \rangle_{n \in \mathbb{N}}$ in $\mathcal{L}^0 = \mathcal{L}^0(\mu)$ converges in measure to $f \in \mathcal{L}^0$ iff

$$\lim_{n \to \infty} \mu^*\{x : x \in F \cap \operatorname{dom} f \cap \operatorname{dom} f_n,\, |f_n(x) - f(x)| > \epsilon\} = 0$$

whenever $F \in \Sigma$, $\mu F < \infty$ and $\epsilon > 0$. Similarly, a sequence $\langle u_n \rangle_{n \in \mathbb{N}}$ in L^0 converges in measure to u iff $\lim_{n \to \infty} \bar\tau_{F\epsilon}(u - u_n) = 0$ whenever $\mu F < \infty$ and $\epsilon > 0$.

(d) In particular, if (X, Σ, μ) is totally finite, $\langle f_n \rangle_{n \in \mathbb{N}} \to f$ in $\mathcal{L}^0$ iff

$$\lim_{n \to \infty} \mu^*\{x : x \in \operatorname{dom} f \cap \operatorname{dom} f_n,\, |f(x) - f_n(x)| > \epsilon\} = 0$$

for every $\epsilon > 0$, and $\langle u_n \rangle_{n \in \mathbb{N}} \to u$ in L^0 iff

$$\lim_{n \to \infty} \bar\tau_{X\epsilon}(u - u_n) = 0$$

for every $\epsilon > 0$.

245G Embedding L^p in L^0: Proposition Let (X, Σ, μ) be any measure space. Then for any $p \in [1, \infty]$, the embedding of $L^p = L^p(\mu)$ in $L^0 = L^0(\mu)$ is continuous for the norm topology of L^p and the topology of convergence in measure on L^0.

proof Suppose that u, $v \in L^p$ and that $\mu F < \infty$. Then $(\chi F)^\bullet$ belongs to L^q, where $q = p/(p-1)$ (taking $q = 1$ if $p = \infty$, $q = \infty$ if $p = 1$ as usual), and

$$\bar\rho_F(u, v) \le \int |u - v| \times (\chi F)^\bullet \le \|u - v\|_p \|\chi F^\bullet\|_q$$

(244Eb). By 2A3H, this is enough to ensure that the embedding $u \mapsto u : L^p \to L^0$ is continuous.

245H The case of L^1 is so important that I go farther with it.

Proposition Let (X, Σ, μ) be a measure space.

(a)(i) If $f \in \mathcal{L}^1 = \mathcal{L}^1(\mu)$ and $\epsilon > 0$, there are a $\delta > 0$ and a set $F \in \Sigma$ of finite measure such that $\int |f - g| \le \epsilon$ whenever $g \in \mathcal{L}^1$, $\int |g| \le \int |f| + \delta$ and $\rho_F(f, g) \le \delta$.

(ii) For any sequence $\langle f_n \rangle_{n\in\mathbb{N}}$ in $\mathcal{L}^1$ and any $f \in \mathcal{L}^1$, $\lim_{n\to\infty} \int |f - f_n| = 0$ iff $\langle f_n \rangle_{n\in\mathbb{N}} \to f$ in measure and $\limsup_{n\to\infty} \int |f_n| \le \int |f|$.

(b)(i) If $u \in L^1 = L^1(\mu)$ and $\epsilon > 0$, there are a $\delta > 0$ and a set $F \in \Sigma$ of finite measure such that $\|u - v\|_1 \le \epsilon$ whenever $v \in L^1$, $\|v\|_1 \le \|u\|_1 + \delta$ and $\bar\rho_F(u, v) \le \delta$.

(ii) For any sequence $\langle u_n \rangle_{n\in\mathbb{N}}$ in L^1 and any $u \in L^1$, $\langle u_n \rangle_{n\in\mathbb{N}} \to u$ for $\| \ \|_1$ iff $\langle u_n \rangle_{n\in\mathbb{N}} \to u$ in measure and $\limsup_{n\to\infty} \|u_n\|_1 \le \|u\|_1$.

proof (a)(i) We know that there are a set F of finite measure and an $\eta > 0$ such that $\int_E |f| \le \frac{1}{5}\epsilon$ whenever $\mu(E \cap F) \le \eta$ (225A). Take $\delta > 0$ such that $\delta(\epsilon + 5\mu F) \le \epsilon\eta$ and $\delta \le \frac{1}{5}\epsilon$. Then if $\int |g| \le \int |f| + \delta$ and $\rho_F(f, g) \le \delta$, let $G \subseteq \operatorname{dom} f \cap \operatorname{dom} g$ be a conegligible measurable set such that $f{\restriction}G$ and $g{\restriction}G$ are both measurable. Set

$$E = \{x : x \in F \cap G, \ |f(x) - g(x)| \ge \tfrac{\epsilon}{\epsilon + 5\mu F}\};$$

then

$$\delta \ge \rho_F(f, g) \ge \tfrac{\epsilon}{\epsilon + 5\mu F} \mu E,$$

so $\mu E \le \eta$. Set $H = F \setminus E$, so that $\mu(F \setminus H) \le \eta$ and $\int_{X\setminus H} |f| \le \frac{1}{5}\epsilon$. On the other hand, for almost every $x \in H$, $|f(x) - g(x)| \le \frac{\epsilon}{\epsilon + 5\mu F}$, so $\int_H |f - g| \le \frac{1}{5}\epsilon$ and

$$\int_H |g| \ge \int_H |f| - \frac{1}{5}\epsilon \ge \int |f| - \int_{X\setminus H} |f| - \frac{1}{5}\epsilon \ge \int |f| - \frac{2}{5}\epsilon.$$

Since $\int |g| \le \int |f| + \frac{1}{5}\epsilon$, $\int_{X\setminus H} |g| \le \frac{3}{5}\epsilon$. Now this means that

$$\int |g - f| \le \int_{X\setminus H} |g| + \int_{X\setminus H} |f| + \int_H |g - f| \le \frac{3}{5}\epsilon + \frac{1}{5}\epsilon + \frac{1}{5}\epsilon = \epsilon,$$

as required.

(ii) If $\lim_{n\to\infty} \int |f - f_n| = 0$, that is, $\lim_{n\to\infty} f_n^\bullet = f^\bullet$ in L^1, then by 245G we must have $\langle f_n^\bullet \rangle_{n\in\mathbb{N}} \to f^\bullet$ in L^0, that is, $\langle f_n \rangle_{n\in\mathbb{N}} \to f$ for the topology of convergence in measure. Also, of course, $\lim_{n\to\infty} \int |f_n| = \int |f|$.

In the other direction, if $\limsup_{n\to\infty} \int |f_n| \le \int |f|$ and $\langle f_n \rangle_{n\in\mathbb{N}} \to f$ for the topology of convergence in measure, then whenever $\delta > 0$ and $\mu F < \infty$ there must be an $m \in \mathbb{N}$ such that $\int |f_n| \le \int |f| + \delta$, $\rho_F(f, f_n) \le \delta$ for every $n \ge m$; so (i) tells us that $\lim_{n\to\infty} \int |f_n - f| = 0$.

(b) This now follows immediately if we express u as $f^\bullet$, v as $g^\bullet$ and u_n as $f_n^\bullet$.

245I Remarks (a) I think the phenomenon here is so important that it is worth looking at some elementary examples.

(i) If μ is counting measure on $\mathbb{N}$, and we set $f_n(n) = 1$, $f_n(i) = 0$ for $i \ne n$, then $\langle f_n \rangle_{n\in\mathbb{N}} \to 0$ in measure, while $\int |f_n| = 1$ for every n.

(ii) If μ is Lebesgue measure on $[0, 1]$, and we set $f_n(x) = 2^n$ for $0 < x \le 2^{-n}$, 0 for other x, then again $\langle f_n \rangle_{n\in\mathbb{N}} \to 0$ in measure, while $\int |f_n| = 1$ for every n.

(iii) In 245Cc we have another sequence $\langle f_n \rangle_{n\in\mathbb{N}}$ converging to 0 in measure, while $\int |f_n| = 1$ for every n. In all these cases (as required by Fatou's Lemma, at least in (i) and (ii)) we have $\int |f| \le \liminf_{n\to\infty} \int |f_n|$. (The next proposition shows that this applies to any sequence which is convergent in measure.)

The common feature of these examples is the way in which somehow the f_n escape to infinity, either laterally (in (i)) or vertically (in (iii)) or both (in (ii)). Note that in all three examples we can set $f_n' = 2^n f_n$ to obtain a sequence still converging to 0 in measure, but with $\lim_{n\to\infty} \int |f_n'| = \infty$.

(b) In 245H, I have used the explicit formulations '$\lim_{n\to\infty}\int|f_n-f|=0$' (for sequences of functions), '$\langle u_n\rangle_{n\in\mathbb{N}}\to u$ for $\| \|_1$' (for sequences in L^1). These are often expressed by saying that $\langle f_n\rangle_{n\in\mathbb{N}}$, $\langle u_n\rangle_{n\in\mathbb{N}}$ are **convergent in mean** to f, u respectively.

245J For semi-finite spaces we have a further relationship.

Proposition Let (X,Σ,μ) be a semi-finite measure space. Write $\mathcal{L}^0=\mathcal{L}^0(\mu)$, etc.

(a)(i) For any $a\ge 0$, the set $\{f:f\in\mathcal{L}^1,\int|f|\le a\}$ is closed in $\mathcal{L}^0$ for the topology of convergence in measure.

(ii) If $\langle f_n\rangle_{n\in\mathbb{N}}$ is a sequence in $\mathcal{L}^1$ which is convergent in measure to $f\in\mathcal{L}^0$, and $\liminf_{n\to\infty}\int|f_n|<\infty$, then f is integrable and $\int|f|\le\liminf_{n\to\infty}\int|f_n|$.

(b)(i) For any $a\ge 0$, the set $\{u:u\in L^1,\|u\|_1\le a\}$ is closed in L^0 for the topology of convergence in measure.

(ii) If $\langle u_n\rangle_{n\in\mathbb{N}}$ is a sequence in L^1 which is convergent in measure to $u\in L^0$, and $\liminf_{n\to\infty}\|u_n\|_1<\infty$, then $u\in L^1$ and $\|u\|_1\le\liminf_{n\to\infty}\|u_n\|_1$.

proof (a)(i) Set $A=\{f:f\in\mathcal{L}^1,\int|f|\le a\}$, and let g be any member of the closure of A in $\mathcal{L}^0$. Let h be any simple function such that $0\le h\le_{\text{a.e.}}|g|$, and $\epsilon>0$. If $h=0$ then of course $\int h\le a$. Otherwise, setting $F=\{x:h(x)>0\}$ and $M=\sup_{x\in X}h(x)$, there is an $f\in A$ such that $\mu^*\{x:x\in F\cap\operatorname{dom}f\cap\operatorname{dom}g,|f(x)-g(x)|\ge\epsilon\}\le\epsilon$ (245F); let $E\supseteq\{x:x\in F\cap\operatorname{dom}f\cap\operatorname{dom}g,|f(x)-g(x)|\ge\epsilon\}$ be a measurable set of measure at most ϵ. Then $h\le_{\text{a.e.}}|f|+\epsilon\chi F+M\chi E$, so $\int h\le a+\epsilon(M+\mu F)$. As ϵ is arbitrary, $\int h\le a$. But we are supposing that μ is semi-finite, so this is enough to ensure that g is integrable and that $\int|g|\le a$ (213B), that is, that $g\in A$. As g is arbitrary, A is closed.

(ii) Now if $\langle f_n\rangle_{n\in\mathbb{N}}$ is convergent in measure to f, and $\liminf_{n\to\infty}\int|f_n|=a$, then for any $\epsilon>0$ there is a subsequence $\langle f_{n(k)}\rangle_{k\in\mathbb{N}}$ such that $\int|f_{n(k)}|\le a+\epsilon$ for every k; since $\langle f_{n(k)}\rangle_{k\in\mathbb{N}}$ still converges in measure to f, $\int|f|\le a+\epsilon$. As ϵ is arbitrary, $\int|f|\le a$.

(b) As in 245H, this is just a translation of part (a) into the language of L^1 and L^0.

245K For σ-finite measure spaces, the topology of convergence in measure on L^0 is metrizable, so can be described effectively in terms of convergent sequences; it is therefore important that we have, in this case, a sharp characterisation of sequential convergence in measure.

Proposition Let (X,Σ,μ) be a σ-finite measure space. Then

(a) a sequence $\langle f_n\rangle_{n\in\mathbb{N}}$ in $\mathcal{L}^0$ converges in measure to $f\in\mathcal{L}^0$ iff every subsequence of $\langle f_n\rangle_{n\in\mathbb{N}}$ has a sub-subsequence converging to f almost everywhere;

(b) a sequence $\langle u_n\rangle_{n\in\mathbb{N}}$ in L^0 converges in measure to $u\in L^0$ iff every subsequence of $\langle u_n\rangle_{n\in\mathbb{N}}$ has a sub-subsequence which order*-converges to u.

proof (a)(i) Suppose that $\langle f_n\rangle_{n\in\mathbb{N}}\to f$, that is, that $\lim_{n\to\infty}\int|f-f_n|\wedge\chi F=0$ for every set F of finite measure. Let $\langle E_k\rangle_{n\in\mathbb{N}}$ be a non-decreasing sequence of sets of finite measure covering X, and let $\langle n(k)\rangle_{k\in\mathbb{N}}$ be a strictly increasing sequence in $\mathbb{N}$ such that $\int|f-f_{n(k)}|\wedge\chi E_k\le 4^{-k}$ for every $k\in\mathbb{N}$. Then $\sum_{k=0}^\infty|f-f_{n(k)}|\wedge\chi E_k$ is finite almost everywhere (242E); but $\lim_{k\to\infty}f_{n(k)}(x)=f(x)$ whenever $\sum_{k=0}^\infty\min(1,|f(x)-f_{n(k)}(x)|)<\infty$, so $\langle f_{n(k)}\rangle_{k\in\mathbb{N}}\to f$ a.e.

(ii) The same applies to every subsequence of $\langle f_n\rangle_{n\in\mathbb{N}}$, so that every subsequence of $\langle f_n\rangle_{n\in\mathbb{N}}$ has a sub-subsequence converging to f almost everywhere.

(iii) Now suppose that $\langle f_n\rangle_{n\in\mathbb{N}}$ does not converge to f. Then there is an open set U containing f such that $\{n:f_n\notin U\}$ is infinite, that is, $\langle f_n\rangle_{n\in\mathbb{N}}$ has a subsequence $\langle f'_n\rangle_{n\in\mathbb{N}}$ with $f'_n\notin U$ for every n. But now no sub-subsequence of $\langle f'_n\rangle_{n\in\mathbb{N}}$ converges to f in measure, so no such sub-subsequence can converge almost everywhere, by 245Ca.

(b) This follows immediately from (a) if we express u as $f^\bullet$, u_n as $f_n^\bullet$.

245L Corollary Let (X,Σ,μ) be a σ-finite measure space.

(a) A subset A of $\mathcal{L}^0=\mathcal{L}^0(\mu)$ is closed for the topology of convergence in measure iff $f\in A$ whenever $f\in\mathcal{L}^0$ and there is a sequence $\langle f_n\rangle_{n\in\mathbb{N}}$ in A such that $f=_{\text{a.e.}}\lim_{n\to\infty}f_n$.

(b) A subset A of $L^0=L^0(\mu)$ is closed for the topology of convergence in measure iff $u\in A$ whenever $u\in L^0$ and there is a sequence $\langle u_n\rangle_{n\in\mathbb{N}}$ in A order*-converging to u.

proof (a)(i) If A is closed for the topology of convergence in measure, and $\langle f_n\rangle_{n\in\mathbb{N}}$ is a sequence in A converging to f almost everywhere, then $\langle f_n\rangle_{n\in\mathbb{N}}$ converges to f in measure, so surely $f\in A$ (since otherwise all but finitely many of the f_n would have to belong to the open set $\mathcal{L}^0\setminus A$).

(ii) If A is not closed, there is an $f \in \overline{A} \setminus A$. The topology can be defined by a metric ρ (245Eb), and we can choose a sequence $\langle f_n \rangle_{n\in\mathbb{N}}$ in A such that $\rho(f_n, f) \le 2^{-n}$ for every n, so that $\langle f_n \rangle_{n\in\mathbb{N}} \to f$ in measure. By 245K, $\langle f_n \rangle_{n\in\mathbb{N}}$ has a subsequence $\langle f'_n \rangle_{n\in\mathbb{N}}$ converging a.e. to f, and this witnesses that A fails to satisfy the condition.

(b) This follows immediately, because $A \subseteq L^0$ is closed iff $\{f : f^\bullet \in A\}$ is closed in $\mathcal{L}^0$.

245M Complex L^0 In 241J I briefly discussed the adaptations needed to construct the complex linear space $L^0_{\mathbb{C}}$. The formulae of 245A may be used unchanged to define topologies of convergence in measure on $\mathcal{L}^0_{\mathbb{C}}$ and $L^0_{\mathbb{C}}$. I think that every word of 245B-245L still applies if we replace each L^0 or $\mathcal{L}^0$ with $L^0_{\mathbb{C}}$ or $\mathcal{L}^0_{\mathbb{C}}$. Alternatively, to relate the 'real' and 'complex' forms of 245E, for instance, we can observe that because

$$\begin{aligned}\max(\rho_F(\mathcal{R}\mathrm{e}(u), \mathcal{R}\mathrm{e}(v)), \rho_F(\mathcal{I}\mathrm{m}(u), \mathcal{I}\mathrm{m}(v))) &\le \rho_F(u, v) \\ &\le \rho_F(\mathcal{R}\mathrm{e}(u), \mathcal{R}\mathrm{e}(v)) + \rho_F(\mathcal{I}\mathrm{m}(u), \mathcal{I}\mathrm{m}(v))\end{aligned}$$

for all u, $v \in L^0$ and all sets F of finite measure, $L^0_{\mathbb{C}}$ can be identified, as uniform space, with $L^0 \times L^0$, so is Hausdorff, or metrizable, or complete iff L^0 is.

245X Basic exercises >(a) Let X be any set, and μ counting measure on X. Show that the topology of convergence in measure on $\mathcal{L}^0(\mu) = \mathbb{R}^X$ is just the product topology on $\mathbb{R}^X$ regarded as a product of copies of $\mathbb{R}$.

>(b) Let (X, Σ, μ) be any measure space, and $(X, \hat{\Sigma}, \hat{\mu})$ its completion. Show that the topologies of convergence in measure on $\mathcal{L}^0(\mu) = \mathcal{L}^0(\hat{\mu})$ (241Xb), corresponding to the families $\{\rho_F : F \in \Sigma, \mu F < \infty\}$, $\{\rho_F : F \in \hat{\Sigma}, \hat{\mu} F < \infty\}$ are the same.

>(c) Let (X, Σ, μ) be any measure space; set $L^0 = L^0(\mu)$. Let u, $u_n \in L^0$ for $n \in \mathbb{N}$. Show that the following are equiveridical:

(i) $\langle u_n \rangle_{n\in\mathbb{N}}$ order*-converges to u in the sense of 245C;

(ii) there are measurable functions f, $f_n : X \to \mathbb{R}$ such that $f^\bullet = u$, $f_n^\bullet = u_n$ for every $n \in \mathbb{N}$, and $f(x) = \lim_{n\to\infty} f_n(x)$ for every $x \in X$;

(iii) $u = \inf_{n\in\mathbb{N}} \sup_{m\ge n} u_m = \sup_{n\in\mathbb{N}} \inf_{m\ge n} u_m$, the infima and suprema being taken in L^0;

(iv) $\inf_{n\in\mathbb{N}} \sup_{m\ge n} |u - u_m| = 0$ in L^0;

(v) there is a non-increasing sequence $\langle v_n \rangle_{n\in\mathbb{N}}$ in L^0 such that $\inf_{n\in\mathbb{N}} v_n = 0$ in L^0 and $u - v_n \le u_n \le u + v_n$ for every $n \in \mathbb{N}$;

(vi) there are sequences $\langle v_n \rangle_{n\in\mathbb{N}}$, $\langle w_n \rangle_{n\in\mathbb{N}}$ in L^0 such that $\langle v_n \rangle_{n\in\mathbb{N}}$ is non-decreasing, $\langle w_n \rangle_{n\in\mathbb{N}}$ is non-increasing, $\sup_{n\in\mathbb{N}} v_n = u = \inf_{n\in\mathbb{N}} w_n$ and $v_n \le u_n \le w_n$ for every $n \in \mathbb{N}$.

(d) Let (X, Σ, μ) be a semi-finite measure space. Show that a sequence $\langle u_n \rangle_{n\in\mathbb{N}}$ in $L^0 = L^0(\mu)$ is order*-convergent to $u \in L^0$ iff $\{|u_n| : n \in \mathbb{N}\}$ is bounded above in L^0 and $\langle \sup_{m\ge n} |u_m - u| \rangle_{n\in\mathbb{N}} \to 0$ for the topology of convergence in measure.

(e) Write out proofs that $L^0(\mu)$ is complete (as linear topological space) adapted to the special cases (i) $\mu X = 1$ (ii) μ is σ-finite, taking advantage of any simplifications you can find.

(f) Let (X, Σ, μ) be a measure space and $r \ge 1$; let $h : \mathbb{R}^r \to \mathbb{R}$ be a continuous function. (i) Suppose that for $1 \le k \le r$ we are given a sequence $\langle f_{kn} \rangle_{n\in\mathbb{N}}$ in $\mathcal{L}^0 = \mathcal{L}^0(\mu)$ converging in measure to $f_k \in \mathcal{L}^0$. Show that $\langle h(f_{1n}, \dots, f_{kn}) \rangle_{n\in\mathbb{N}}$ converges in measure to $h(f_1, \dots, f_k)$. (ii) Generally, show that $(f_1, \dots, f_r) \mapsto h(f_1, \dots, f_r) : (\mathcal{L}^0)^r \to \mathcal{L}^0$ is continuous for the topology of convergence in measure. (iii) Show that the corresponding function $\bar{h} : (L^0)^r \to L^0$ (241Xh) is continuous for the topology of convergence in measure.

(g) Let (X, Σ, μ) be a measure space and $u \in L^1(\mu)$. Show that $v \mapsto \int u \times v : L^\infty \to \mathbb{R}$ is continuous for the topology of convergence in measure on the unit ball of L^∞, but not, as a rule, on the whole of L^∞.

(h) Let (X, Σ, μ) be a measure space and v a non-negative member of $L^1 = L^1(\mu)$. Show that on the set $A = \{u : u \in L^1, |u| \le v\}$ the subspace topologies (2A3C) induced by the norm topology of L^1 and the topology of convergence in measure are the same. (*Hint*: given $\epsilon > 0$, take $F \in \Sigma$ of finite measure and $M \ge 0$ such that $\int (|v| - M\chi F^\bullet)^+ \le \epsilon$. Show that $\|u - u'\|_1 \le \epsilon + M\bar{\rho}_F(u, u')$ for all u, $u' \in A$.)

(i) Let (X, Σ, μ) be a measure space and $\mathcal{F}$ a filter on $L^1 = L^1(\mu)$ which is convergent, for the topology of convergence in measure, to $u \in L^1$. Show that $\mathcal{F} \to u$ for the norm topology of L^1 iff $\inf_{A\in\mathcal{F}} \sup_{v\in A} \|v\|_1 \le \|u\|_1$.

(j) Let (X,Σ,μ) be a measure space and $p\in[1,\infty[$. Suppose that $\langle u_n\rangle_{n\in\mathbb{N}}$ is a sequence in $L^p(\mu)$ which converges for $\|\;\|_p$ to $u\in L^p(\mu)$. Show that $\langle |u_n|^p\rangle_{n\in\mathbb{N}}\to|u|^p$ for $\|\;\|_1$. (*Hint*: 245G, 245Dd, 245H.)

>(k) Let (X,Σ,μ) be a semi-finite measure space and $p\in[1,\infty]$, $a\ge 0$. Show that $\{u:u\in L^p(\mu),\,\|u\|_p\le a\}$ is closed in $L^0(\mu)$ for the topology of convergence in measure.

(l) Let (X,Σ,μ) be a measure space, and $\langle u_n\rangle_{n\in\mathbb{N}}$ a sequence in $L^p=L^p(\mu)$, where $1\le p<\infty$. Let $u\in L^p$. Show that the following are equiveridical: (i) $u=\lim_{n\to\infty}u_n$ for the norm topology of L^p (ii) $\langle u_n\rangle_{n\in\mathbb{N}}\to u$ for the topology of convergence in measure and $\lim_{n\to\infty}\|u_n\|_p=\|u\|_p$ (iii) $\langle u_n\rangle_{n\in\mathbb{N}}\to u$ for the topology of convergence in measure and $\limsup_{n\to\infty}\|u_n\|_p\le\|u\|_p$.

(m) Let X be a set and μ, ν two measures on X with the same measurable sets and the same negligible sets. (i) Show that $\mathcal{L}^0(\mu)=\mathcal{L}^0(\nu)$ and $L^0(\mu)=L^0(\nu)$. (ii) Show that if both μ and ν are semi-finite, then they define the same topology of convergence in measure on $\mathcal{L}^0$ and L^0. (*Hint*: use 215A to show that if $\mu E<\infty$ then $\mu E=\sup\{\mu F:F\subseteq E,\,\nu F<\infty\}$.)

245Y Further exercises (a) Let (X,Σ,μ) be a measure space and give Σ the topology described in 232Ya. Show that $\chi:\Sigma\to\mathcal{L}^0(\mu)$ is a homeomorphism between Σ and its image $\chi[\Sigma]$ in $\mathcal{L}^0$, if $\mathcal{L}^0$ is given the topology of convergence in measure and $\chi[\Sigma]$ the subspace topology.

(b) Let (X,Σ,μ) be a measure space and Y any subset of X; let μ_Y be the subspace measure on Y. Let $T:L^0(\mu)\to L^0(\mu_Y)$ be the canonical map defined by setting $T(f^\bullet)=(f\restriction Y)^\bullet$ for every $f\in\mathcal{L}^0(\mu)$ (241Yg). Show that T is continuous for the topologies of convergence in measure on $L^0(\mu)$ and $L^0(\mu_Y)$.

(c) Let (X,Σ,μ) be a measure space, and $\tilde\mu$ the c.l.d. version of μ. Show that the map $T:L^0(\mu)\to L^0(\tilde\mu)$ induced by the inclusion $\mathcal{L}^0(\mu)\subseteq\mathcal{L}^0(\tilde\mu)$ (241Yf) is continuous for the topologies of convergence in measure.

(d) Let (X,Σ,μ) be a measure space, and give $L^0=L^0(\mu)$ the topology of convergence in measure. Let $A\subseteq L^0$ be a non-empty downwards-directed set, and suppose that $\inf A=0$ in L^0. (i) Let $F\in\Sigma$ be any set of finite measure, and define $\bar\tau_F$ as in 245A; show that $\inf_{u\in A}\bar\tau_F(u)=0$. (*Hint*: set $\gamma=\inf_{u\in A}\bar\tau_F(u)$; find a non-increasing sequence $\langle u_n\rangle_{n\in\mathbb{N}}$ in A such that $\lim_{n\to\infty}\bar\tau_F(u_n)=\gamma$; set $v=(\chi F)^\bullet\wedge\inf_{n\in\mathbb{N}}u_n$ and show that $u\wedge v=v$ for every $u\in A$, so that $v=0$.) (ii) Show that if U is any open set containing 0, there is a $u\in A$ such that $v\in U$ whenever $0\le v\le u$.

(e) Let (X,Σ,μ) be a measure space. (i) Show that for $u\in L^0=L^0(\mu)$ we may define $\psi_a(u)$, for $a\ge 0$, by setting $\psi_a(u)=\mu\{x:|f(x)|\ge a\}$ whenever $f:X\to\mathbb{R}$ is a measurable function and $f^\bullet=u$. (ii) Define $\rho:L^0\times L^0\to[0,1]$ by setting $\rho(u,v)=\min(\{1\}\cup\{a:a\ge 0,\ \psi_a(u-v)\le a\}$. Show that ρ is a metric on L^0, that L^0 is complete under ρ, and that $+,\ -,\ \wedge,\ \vee:L^0\times L^0\to L^0$ are continuous for ρ. (iii) Show that $c\mapsto cu:\mathbb{R}\to L^0$ is continuous for every $u\in L^0$ iff (X,Σ,μ) is totally finite, and that in this case ρ defines the topology of convergence in measure on L^0.

(f) Let (X,Σ,μ) be a localizable measure space and $A\subseteq L^0=L^0(\mu)$ a non-empty upwards-directed set which is bounded in the linear topological space sense (i.e., such that for every neighbourhood U of 0 in L^0 there is a $k\in\mathbb{N}$ such that $A\subseteq kU$). Show that A is bounded above in L^0, and that its supremum belongs to its closure.

(g) Let (X,Σ,μ) be a measure space, $p\in[1,\infty[$ and v a non-negative member of $L^p=L^p(\mu)$. Show that on the set $A=\{u:u\in L^p,\ |u|\le v\}$ the subspace topologies induced by the norm topology of L^p and the topology of convergence in measure are the same.

(h) Let S be the set of all sequences $s:\mathbb{N}\to\mathbb{N}$ such that $\lim_{n\to\infty}s(n)=\infty$. For every $s\in S$, let (X_s,Σ_s,μ_s) be $[0,1]$ with Lebesgue measure, and let (X,Σ,μ) be the direct sum of $\langle(X_s,\Sigma_s,\mu_s)\rangle_{s\in S}$ (214L). For $s\in S$, $t\in[0,1]$, $n\in\mathbb{N}$ set $h_n(s,t)=f_{s(n)}(t)$, where $\langle f_n\rangle_{n\in\mathbb{N}}$ is the sequence of 245Cc. Show that $\langle h_n\rangle_{n\in\mathbb{N}}\to 0$ for the topology of convergence in measure on $\mathcal{L}^0(\mu)$, but that $\langle h_n\rangle_{n\in\mathbb{N}}$ has no subsequence which is convergent to 0 almost everywhere.

(i) Let X be a set, and suppose we are given a relation $\rightharpoonup$ between sequences in X and members of X such that (α) if $x_n=x$ for every n then $\langle x_n\rangle_{n\in\mathbb{N}}\rightharpoonup x$ (β) $\langle x'_n\rangle_{n\in\mathbb{N}}\rightharpoonup x$ whenever $\langle x_n\rangle_{n\in\mathbb{N}}\rightharpoonup x$ and $\langle x'_n\rangle_{n\in\mathbb{N}}$ is a subsequence of $\langle x_n\rangle_{n\in\mathbb{N}}$. Show that we have a topology $\mathfrak{T}$ on X defined by saying that a subset G of X belongs to $\mathfrak{T}$ iff whenever $\langle x_n\rangle_{n\in\mathbb{N}}$ is a sequence in X and $\langle x_n\rangle_{n\in\mathbb{N}}\rightharpoonup x\in G$ then some x_n belongs to G. Show that a sequence $\langle x_n\rangle_{n\in\mathbb{N}}$ in X is $\mathfrak{T}$-convergent to x iff every subsequence of $\langle x_n\rangle_{n\in\mathbb{N}}$ has a sub-subsequence $\langle x''_n\rangle_{n\in\mathbb{N}}$ such that $\langle x''_n\rangle_{n\in\mathbb{N}}\rightharpoonup x$.

(j) Let μ be Lebesgue measure on $\mathbb{R}^r$. Show that $L^0(\mu)$ is separable for the topology of convergence in measure. (*Hint*: 244I.)

245 Notes and comments In this section I am inviting you to regard the topology of (local) convergence in measure as the standard topology on L^0, just as the norms define the standard topologies on L^p spaces for $p \geq 1$. The definition I have chosen is designed to make addition and scalar multiplication and the operations $\vee$, $\wedge$ and $\times$ continuous (245D); see also 245Xf. From the point of view of functional analysis these properties are more important than metrizability or even completeness.

Just as the algebraic and order structure of L^0 can be described in terms of the general theory of Riesz spaces, the more advanced results 241G and 245E also have interpretations in the general theory. It is not an accident that (for semi-finite measure spaces) L^0 is Dedekind complete iff it is complete as uniform space; you may find the relevant generalizations in 23K and 24E of Fremlin 74. Of course it is exactly because the two kinds of completeness are interrelated that I feel it necessary to use the phrase 'Dedekind completeness' to distinguish this particular kind of order-completeness from the more familiar uniformity-completeness described in 2A5F.

The usefulness of the topology of convergence in measure derives in large part from 245G-245J and the L^p versions 245Xk and 245Xl. Some of the ideas here can be related to a question arising out of the basic convergence theorems. If $\langle f_n\rangle_{n\in\mathbb{N}}$ is a sequence of integrable functions converging (pointwise) to a function f, in what ways can $\int f$ fail to be $\lim_{n\to\infty}\int f_n$? In the language of this section, this translates into: if we have a sequence (or filter) in L^1 converging for the topology of convergence in measure, in what ways can it fail to converge for the norm topology of L^1? The first answer is Lebesgue's Dominated Convergence Theorem: this cannot happen if the sequence is dominated, that is, lies within some set of the form $\{u : |u| \leq v\}$ where $v \in L^1$. (See 245Xh and 245Yg.) I will return to this in the next section. For the moment, though, 245H tells us that if $\langle u_n\rangle_{n\in\mathbb{N}}$ converges in measure to $u \in L^1$, but not for the topology of L^1, it is because $\limsup_{n\to\infty}\|u_n\|_1$ is too big; some of its weight is being lost at infinity, as in the examples of 245I. If $\langle u_n\rangle_{n\in\mathbb{N}}$ actually order*-converges to u, then Fatou's Lemma tells us that $\liminf_{n\to\infty}\|u_n\|_1 \geq \|u\|_1$, that is, that the limit cannot have greater weight (as measured by $\| \|_1$) than the sequence provides. 245J and 245Xk are generalizations of this to convergence in measure. If you want a generalization of B.Levi's theorem, then 242Yf remains the best expression in the language of this chapter; but 245Yf is a version in terms of the concepts of the present section.

In the case of σ-finite spaces, we have an alternative description of the topology of convergence in measure (245L) which makes no use of any of the functionals or pseudo-metrics in 245A. This can be expressed, at least in the context of L^0, in terms of a standard result from general topology (245Yi). You will see that that result gives a recipe for a topology on L^0 which could be applied in any measure space. What is remarkable is that for σ-finite spaces we get a linear space topology.

246 Uniform integrability

The next topic is a fairly specialized one, but it is of great importance, for different reasons, in both probability theory and functional analysis, and it therefore seems worth while giving a proper treatment straight away.

246A Definition Let (X, Σ, μ) be a measure space.

(a) A set $A \subseteq \mathcal{L}^1(\mu)$ is **uniformly integrable** if for every $\epsilon > 0$ we can find a set $E \in \Sigma$, of finite measure, and an $M \geq 0$ such that

$$\int(|f| - M\chi E)^+ \leq \epsilon \text{ for every } f \in A.$$

(b) A set $A \subseteq L^1(\mu)$ is **uniformly integrable** if for every $\epsilon > 0$ we can find a set $E \in \Sigma$, of finite measure, and an $M \geq 0$ such that

$$\int(|u| - M\chi E^\bullet)^+ \leq \epsilon \text{ for every } u \in A.$$

246B Remarks (a) Recall the formulae from 241Ef: $u^+ = u \vee 0$, so $(u - v)^+ = u - u \wedge v$.

(b) The phrase 'uniformly integrable' is not particularly helpful. But of course we can observe that for any particular integrable function f, there are simple functions approximating f for $\| \|_1$ (242M), and such functions will be bounded (in modulus) by functions of the form $M\chi E$, with $\mu E < \infty$; thus singleton subsets of $\mathcal{L}^1$ and L^1 are uniformly integrable. A general uniformly integrable set of functions is one in which M and E can be chosen uniformly over the set.

(c) It will I hope be clear from the definitions that $A \subseteq \mathcal{L}^1$ is uniformly integrable iff $\{f^\bullet : f \in A\} \subseteq L^1$ is uniformly integrable.

(d) There is a useful simplification in the definition if $\mu X < \infty$ (in particular, if (X, Σ, μ) is a probability space). In this case a set $A \subseteq L^1(\mu)$ is uniformly integrable iff

$$\inf_{M \geq 0} \sup_{u \in A} \int (|u| - Me)^+ = 0$$

iff

$$\lim_{M \to \infty} \sup_{u \in A} \int (|u| - Me)^+ = 0,$$

writing $e = \chi X^\bullet \in L^1(\mu)$. (For if $\sup_{u \in A} \int (|u| - M\chi E^\bullet)^+ \leq \epsilon$, then $\int (|u| - M'e)^+ \leq \epsilon$ for every $M' \geq M$.) Similarly, $A \subseteq \mathcal{L}^1(\mu)$ is uniformly integrable iff

$$\lim_{M \to \infty} \sup_{f \in A} \int (|f| - M\chi X)^+ = 0$$

iff

$$\inf_{M \geq 0} \sup_{f \in A} \int (|f| - M\chi X)^+ = 0.$$

Warning! Some authors use the phrase 'uniformly integrable' for sets satisfying the conditions in (d) even when μ is not totally finite.

246C We have the following wide-ranging stability properties of the class of uniformly integrable sets in L^1 or $\mathcal{L}^1$.

Proposition Let (X, Σ, μ) be a measure space and A a uniformly integrable subset of $L^1(\mu)$.

(a) A is bounded for the norm $\| \|_1$.

(b) Any subset of A is uniformly integrable.

(c) For any $a \in \mathbb{R}$, $aA = \{au : u \in A\}$ is uniformly integrable.

(d) There is a uniformly integrable $C \supseteq A$ such that C is convex and $\| \|_1$-closed and $v \in C$ whenever $u \in C$ and $|v| \leq |u|$.

(e) If B is another uniformly integrable subset of L^1, then $A \cup B$ and $A + B = \{u + v : u \in A, v \in B\}$ are uniformly integrable.

proof Write Σ^f for $\{E : E \in \Sigma, \mu E < \infty\}$.

(a) There must be $E \in \Sigma^f$, $M \geq 0$ such that $\int (|u| - M\chi E^\bullet)^+ \leq 1$ for every $u \in A$; now

$$\|u\|_1 \leq \int (|u| - M\chi E^\bullet)^+ + \int M\chi E^\bullet \leq 1 + M\mu E$$

for every $u \in A$, so A is bounded.

(b) This is immediate from the definition 246Ab.

(c) Given $\epsilon > 0$, we can find $E \in \Sigma^f$, $M \geq 0$ such that $|a| \int_E (|u| - M\chi E^\bullet)^+ \leq \epsilon$ for every $u \in A$; now $\int_E (|v| - |a|M\chi E^\bullet)^+ \leq \epsilon$ for every $v \in aA$.

(d) If A is empty, take $C = A$. Otherwise, try

$$C = \{v : v \in L^1, \int (|v| - w)^+ \leq \sup_{u \in A} \int (|u| - w)^+ \text{ for every } w \in L^1(\mu)\}.$$

Evidently $A \subseteq C$, and C satisfies the definition 246Ab because A does, considering w of the form $M\chi E^\bullet$ where $E \in \Sigma^f$ and $M \geq 0$. The functionals

$$v \mapsto \int (|v| - w)^+ : L^1(\mu) \to \mathbb{R}$$

are all continuous for $\| \|_1$ (because the operators $v \mapsto |v|$, $v \mapsto v - w$, $v \mapsto v^+$, $v \mapsto \int v$ are continuous), so C is closed. If $|v'| \leq |v|$ and $v \in C$, then

$$\int (|v'| - w)^+ \leq \int (|v| - w)^+ \leq \sup_{u \in A} \int (|u| - w)^+$$

for every w, and $v' \in C$. If $v = av_1 + bv_2$ where v_1, $v_2 \in C$, $a \in [0, 1]$ and $b = 1 - a$, then $|v| \leq a|v_1| + b|v_2|$, so

$$|v| - w \leq (a|v_1| - aw) + (b|v_2| - bw) \leq (a|v_1| - aw)^+ + (b|v_2| - bw)^+$$

and

$$(|v|-w)^+ \le a(|v_1|-w)^+ + b(|v_2|-w)^+$$

for every w; accordingly

$$\begin{aligned}\int(|v|-w)^+ &\le a\int(|v_1|-w)^+ + b\int(|v_2|-w)^+\\ &\le (a+b)\sup_{u\in A}\int(|u|-w)^+ = \sup_{u\in A}\int(|u|-w)^+\end{aligned}$$

for every w, and $v \in C$.

Thus C has all the required properties.

(e) I show first that $A \cup B$ is uniformly integrable. **P** Given $\epsilon > 0$, let M_1, $M_2 \ge 0$ and E_1, $E_2 \in \Sigma^f$ be such that

$$\int(|u|-M_1\chi E_1^\bullet)^+ \le \epsilon \text{ for every } u \in A,$$

$$\int(|u|-M_2\chi E_2^\bullet)^+ \le \epsilon \text{ for every } u \in B.$$

Set $M = \max(M_1, M_2)$, $E = E_1 \cup E_2$; then $\mu E < \infty$ and

$$\int(|u|-M\chi E^\bullet)^+ \le \epsilon \text{ for every } u \in A \cup B.$$

As ϵ is arbitrary, $A \cup B$ is uniformly integrable. **Q**

Now (d) tells us that there is a convex uniformly integrable set C including $A \cup B$, and in this case $A + B \subseteq 2C$, so $A + B$ is also uniformly integrable, using (b) and (c).

246D Proposition Let (X, Σ, μ) be a probability space and $A \subseteq L^1(\mu)$ a uniformly integrable set. Then there is a convex, $\| \|_1$-closed uniformly integrable set $C \subseteq L^1$ such that $A \subseteq C$, $w \in C$ whenever $v \in C$ and $|w| \le |v|$, and $Pv \in C$ whenever $v \in C$ and P is the conditional expectation operator associated with a σ-subalgebra of Σ.

proof Set

$$C = \{v : v \in L^1(\mu),\ \int(|v|-Me)^+ \le \sup_{u\in A}\int(|u|-Me)^+ \text{ for every } M \ge 0\},$$

writing $e = \chi X^\bullet$ as usual. The arguments in the proof of 246Cd make it plain that $C \supseteq A$ is uniformly integrable, convex and closed, and that $w \in C$ whenever $v \in C$ and $|w| \le |v|$. As for the conditional expectation operators, if $v \in C$, T is a σ-subalgebra of Σ, P is the associated conditional expectation operator, and $M \ge 0$, then

$$|Pv| \le P|v| = P((|v| \wedge Me) + (|v|-Me)^+) \le Me + P((|v|-Me)^+),$$

so

$$(|Pv|-Me)^+ \le P((|v|-Me)^+)$$

and

$$\int(|Pv|-Me)^+ \le \int P(|v|-Me)^+ = \int(|v|-Me)^+ \le \sup_{u\in A}\int(|u|-Me)^+;$$

as M is arbitrary, $Pv \in C$.

246E Remarks (a) Of course 246D has an expression in terms of $\mathcal{L}^1$ rather than L^1: if (X, Σ, μ) is a probability space and $A \subseteq \mathcal{L}^1(\mu)$ is uniformly integrable, then there is a uniformly integrable set $C \supseteq A$ such that (i) $af + (1-a)g \in C$ whenever f, $g \in C$ and $a \in [0,1]$ (ii) $g \in C$ whenever $f \in C$, $g \in \mathcal{L}^0(\mu)$ and $|g| \le_{\text{a.e.}} |f|$ (iii) $f \in C$ whenever there is a sequence $\langle f_n\rangle_{n\in\mathbb{N}}$ in C such that $\lim_{n\to\infty}\int|f-f_n| = 0$ (iv) $g \in C$ whenever there is an $f \in C$ such that g is a conditional expectation of f with respect to some σ-subalgebra of Σ.

(b) In fact, there are obvious extensions of 246D; the proof there already shows that $T[C] \subseteq C$ whenever $T : L^1(\mu) \to L^1(\mu)$ is an order-preserving linear operator such that $\|Tu\|_1 \le \|u\|_1$ for every $u \in L^1(\mu)$ and $\|Tu\|_\infty \le \|u\|_\infty$ for every $u \in L^1(\mu) \cap L^\infty(\mu)$ (246Yc). If we had done a bit more of the theory of operators on Riesz spaces I should be able to take you a good deal farther along this road; for instance, it is not in fact necessary to assume that the operators T of the last sentence are order-preserving. I will return to this in Chapter 37 in the next volume.

(c) Moreover, the main theorem of the next section will show that for any measure spaces (X, Σ, μ), (Y, T, ν), $T[A]$ will be uniformly integrable in $L^1(\nu)$ whenever $A \subseteq L^1(\mu)$ is uniformly integrable and $T : L^1(\mu) \to L^1(\nu)$ is a continuous linear operator (247D).

246F We shall need an elementary lemma which I have not so far spelt out.

Lemma Let (X,Σ,μ) be a measure space. Then for any $u\in L^1(\mu)$,

$$\|u\|_1\le 2\sup_{E\in\Sigma}|\int_E u|.$$

proof Express u as $f^\bullet$ where $f:X\to\mathbb{R}$ is measurable. Set $F=\{x:f(x)\ge 0\}$. Then

$$\|u\|_1=\int|f|=|\int_F f|+|\int_{X\setminus F}f|\le 2\sup_{E\in\Sigma}|\int_E f|=2\sup_{E\in\Sigma}|\int_E u|.$$

246G Now we come to some of the remarkable alternative descriptions of uniform integrability.

Theorem Let (X,Σ,μ) be any measure space and A a non-empty subset of $L^1(\mu)$. Then the following are equiveridical:

(i) A is uniformly integrable;

(ii) $\sup_{u\in A}|\int_F u|<\infty$ for every μ-atom $F\in\Sigma$, and for every $\epsilon>0$ there are $E\in\Sigma$, $\delta>0$ such that $\mu E<\infty$ and $|\int_F u|\le\epsilon$ whenever $u\in A$, $F\in\Sigma$ and $\mu(F\cap E)\le\delta$;

(iii) $\sup_{u\in A}|\int_F u|<\infty$ for every μ-atom $F\in\Sigma$, and $\lim_{n\to\infty}\sup_{u\in A}|\int_{F_n}u|=0$ whenever $\langle F_n\rangle_{n\in\mathbb{N}}$ is a disjoint sequence in Σ;

(iv) $\sup_{u\in A}|\int_F u|<\infty$ for every μ-atom $F\in\Sigma$, and $\lim_{n\to\infty}\sup_{u\in A}|\int_{F_n}u|=0$ whenever $\langle F_n\rangle_{n\in\mathbb{N}}$ is a non-increasing sequence in Σ with empty intersection.

Remark I use the phrase 'μ-atom' to emphasize that I mean an atom in the measure space sense (211I).

proof (a)(i)$\Rightarrow$(iv) Suppose that A is uniformly integrable. Then surely if $F\in\Sigma$ is a μ-atom,

$$\sup_{u\in A}|\int_F u|\le\sup_{u\in A}\|u\|_1<\infty,$$

by 246Ca. Now suppose that $\langle F_n\rangle_{n\in\mathbb{N}}$ is a non-increasing sequence in Σ with empty intersection, and that $\epsilon>0$. Take $E\in\Sigma$, $M\ge 0$ such that $\mu E<\infty$ and $\int(|u|-M\chi E^\bullet)^+\le\frac{1}{2}\epsilon$ whenever $u\in A$. Then for all n large enough, $M\mu(F_n\cap E)\le\frac{1}{2}\epsilon$, so that

$$|\int_{F_n}u|\le\int_{F_n}|u|\le\int(|u|-M\chi E^\bullet)^+ +\int_{F_n}M\chi E^\bullet\le\frac{\epsilon}{2}+M\mu(F_n\cap E)\le\epsilon$$

for every $u\in A$. As ϵ is arbitrary, $\lim_{n\to\infty}\sup_{u\in A}|\int_{F_n}u|=0$, and (iv) is true.

(b)(iv)$\Rightarrow$(iii) Suppose that (iv) is true. Then of course $\sup_{u\in A}|\int_F u|<\infty$ for every μ-atom $F\in\Sigma$. **?** Suppose, if possible, that $\langle F_n\rangle_{n\in\mathbb{N}}$ is a disjoint sequence in Σ such that $\epsilon=\limsup_{n\to\infty}\sup_{u\in A}\min(1,\frac{1}{3}|\int_{F_n}u|)>0$. Set $H_n=\bigcup_{i\ge n}F_i$ for each n, so that $\langle H_n\rangle_{n\in\mathbb{N}}$ is non-increasing and has empty intersection, and $\int_{H_n}u\to 0$ as $n\to\infty$ for every $u\in L^1(\mu)$. Choose $\langle n_i\rangle_{i\in\mathbb{N}}$, $\langle m_i\rangle_{i\in\mathbb{N}}$, $\langle u_i\rangle_{i\in\mathbb{N}}$ inductively, as follows. $n_0=0$. Given $n_i\in\mathbb{N}$, take $m_i\ge n_i$, $u_i\in A$ such that $|\int_{F_{m_i}}u_i|\ge 2\epsilon$. Take $n_{i+1}>m_i$ such that $\int_{H_{n_{i+1}}}|u_i|\le\epsilon$. Continue.

Set $G_k=\bigcup_{i\ge k}F_{m_i}$ for each k. Then $\langle G_k\rangle_{k\in\mathbb{N}}$ is a non-increasing sequence in Σ with empty intersection. But $F_{m_i}\subseteq G_i\subseteq F_{m_i}\cup H_{n_{i+1}}$, so

$$|\int_{G_i}u_i|\ge|\int_{F_{m_i}}u_i|-|\int_{G_i\setminus F_{m_i}}u_i|\ge 2\epsilon-\int_{H_{n_{i+1}}}|u_i|\ge\epsilon$$

for every i, contradicting the hypothesis (iv). **X**

This means that $\lim_{n\to\infty}\sup_{u\in A}|\int_{F_n}u|$ must be zero, and (iii) is true.

(c)(iii)$\Rightarrow$(ii) We still have $\sup_{u\in A}|\int_F u|<\infty$ for every μ-atom F. **?** Suppose, if possible, that there is an $\epsilon>0$ such that for every measurable set E of finite measure and every $\delta>0$ there are $u\in A$, $F\in\Sigma$ such that $\mu(F\cap E)\le\delta$ and $|\int_F u|\ge\epsilon$. Choose a sequence $\langle E_n\rangle_{n\in\mathbb{N}}$ of sets of finite measure, a sequence $\langle G_n\rangle_{n\in\mathbb{N}}$ in Σ, a sequence $\langle\delta_n\rangle_{n\in\mathbb{N}}$ of strictly positive real numbers and a sequence $\langle u_n\rangle_{n\in\mathbb{N}}$ in A as follows. Given u_k, E_k, δ_k for $k<n$, choose $u_n\in A$ and $G_n\in\Sigma$ such that $\mu(G_n\cap\bigcup_{k<n}E_k)\le 2^{-n}\min(\{1\}\cup\{\delta_k:k<n\})$ and $|\int_{G_n}u_n|\ge\epsilon$; then choose a set E_n of finite measure and a $\delta_n>0$ such that $\int_F|u_n|\le\frac{1}{2}\epsilon$ whenever $F\in\Sigma$ and $\mu(F\cap E_n)\le\delta_n$ (see 225A). Continue.

On completing the induction, set $F_n=E_n\cap G_n\setminus\bigcup_{k>n}G_k$ for each n; then $\langle F_n\rangle_{n\in\mathbb{N}}$ is a disjoint sequence in Σ. By the choice of G_k,

$$\mu(E_n\cap\bigcup_{k>n}G_k)\le\sum_{k=n+1}^\infty 2^{-k}\delta_n\le\delta_n,$$

so $\mu(E_n\cap(G_n\setminus F_n))\le\delta_n$ and $\int_{G_n\setminus F_n}|u_n|\le\frac{1}{2}\epsilon$. This means that $|\int_{F_n}u_n|\ge|\int_{G_n}u_n|-\frac{1}{2}\epsilon\ge\frac{1}{2}\epsilon$. But this is contrary to the hypothesis (iii). **X**

(d)(ii)⇒(i)($\boldsymbol{\alpha}$) Assume (ii). Let $\epsilon > 0$. Then there are $E \in \Sigma$, $\delta > 0$ such that $\mu E < \infty$ and $|\int_F u| \le \epsilon$ whenever $u \in A$, $F \in \Sigma$ and $\mu(F \cap E) \le \delta$. Now $\sup_{u\in A}\int_E |u| < \infty$. **P** Write $\mathcal{I}$ for the family of those $F \in \Sigma$ such that $F \subseteq E$ and $\sup_{u\in A}\int_F |u|$ is finite. If $F \subseteq E$ is an atom for μ, then $\sup_{u\in A}\int_F |u| = \sup_{u\in A}|\int_F u| < \infty$, so $F \in \mathcal{I}$. (The point is that if $f : X \to \mathbb{R}$ is a measurable function such that $f^\bullet = u$, then one of $F' = \{x : x \in F,\, f(x) \ge 0\}$, $F'' = \{x : x \in F,\, f(x) < 0\}$ must be negligible, so that $\int_F |u|$ is either $\int_{F'} u = \int_F u$ or $-\int_{F''} u = -\int_F u$.) If $F \in \Sigma$, $F \subseteq E$ and $\mu F \le \delta$ then

$$\sup_{u\in A}\int_F |u| \le 2\sup_{u\in A, G\in\Sigma, G\subseteq F}|\int_G u| \le 2\epsilon$$

(by 246F), so $F \in \mathcal{I}$. Next, if F, $G \in \mathcal{I}$ then $\sup_{u\in A}\int_{F\cup G}|u| \le \sup_{u\in A}\int_F |u| + \sup_{u\in A}\int_G |u|$ is finite, so $F \cup G \in \mathcal{I}$. Finally, if $\langle F_n\rangle_{n\in\mathbb{N}}$ is any sequence in $\mathcal{I}$, and $F = \bigcup_{n\in\mathbb{N}} F_n$, there is some $n \in \mathbb{N}$ such that $\mu(F \setminus \bigcup_{i\le n} F_i) \le \delta$; now $\bigcup_{i\le n} F_i$ and $F \setminus \bigcup_{i\le n} F_i$ both belong to $\mathcal{I}$, so $F \in \mathcal{I}$.

By 215Ab, there is an $F \in \mathcal{I}$ such that $H \setminus F$ is negligible for every $H \in \mathcal{I}$. Now observe that $E \setminus F$ cannot include any non-negligible member of $\mathcal{I}$; in particular, cannot include either an atom or a non-negligible set of measure less than δ. But this means that the subspace measure on $E \setminus F$ is atomless, totally finite and has no non-negligible measurable sets of measure less than δ; by 215D, $\mu(E \setminus F) = 0$ and $E \setminus F$ and E belong to $\mathcal{I}$, as required. **Q**

Since $\int_{X\setminus E}|u| \le \delta$ for every $u \in A$, $\gamma = \sup_{u\in A}\int |u|$ is finite.

($\boldsymbol{\beta}$) Set $M = \gamma/\delta$. If $u \in A$, express u as $f^\bullet$, where $f : X \to \mathbb{R}$ is measurable, and consider

$$F = \{x : f(x) \ge M\chi E(x)\}.$$

Then

$$M\mu(F \cap E) \le \int_F f = \int_F u \le \gamma,$$

so $\mu(F \cap E) \le \gamma/M = \delta$. Accordingly $\int_F u \le \epsilon$. Similarly, $\int_{F'}(-u) \le \epsilon$, writing $F' = \{x : -f(x) \ge M\chi E(x)\}$. But this means that

$$\int(|u| - M\chi E^\bullet)^+ = \int(|f| - M\chi E)^+ \le \int_{F\cup F'}|f| = \int_{F\cup F'}|u| \le 2\epsilon$$

for every $u \in A$. As ϵ is arbitrary, A is uniformly integrable.

246H Remarks (a) Of course conditions (ii)-(iv) of this theorem, like (i), have direct translations in terms of members of $\mathcal{L}^1$. Thus a non-empty set $A \subseteq \mathcal{L}^1$ is uniformly integrable iff $\sup_{f\in A}|\int_F f|$ is finite for every atom $F \in \Sigma$ and

either for every $\epsilon > 0$ we can find $E \in \Sigma$, $\delta > 0$ such that $\mu E < \infty$ and $|\int_F f| \le \epsilon$ whenever $f \in A$, $F \in \Sigma$ and $\mu(F \cap E) \le \delta$

or $\lim_{n\to\infty}\sup_{f\in A}|\int_{F_n} f| = 0$ for every disjoint sequence $\langle F_n\rangle_{n\in\mathbb{N}}$ in Σ

or $\lim_{n\to\infty}\sup_{f\in A}|\int_{F_n} f| = 0$ for every non-increasing sequence $\langle F_n\rangle_{n\in\mathbb{N}}$ in Σ with empty intersection.

(b) There are innumerable further equivalent expressions characterizing uniform integrability; every author has his own favourite. Many of them are variants on (i)-(iv) of this theorem, as in 246I and 246Yd-246Yf. For a condition of a quite different kind, see Theorem 247C.

246I Corollary Let (X, Σ, μ) be a probability space. For $f \in \mathcal{L}^0(\mu)$, $M \ge 0$ set $F(f, M) = \{x : x \in \operatorname{dom} f,\, |f(x)| \ge M\}$. Then a non-empty set $A \subseteq \mathcal{L}^1(\mu)$ is uniformly integrable iff $\lim_{M\to\infty}\sup_{f\in A}\int_{F(f,M)}|f| = 0$.

proof (a) If A satisfies the condition, then

$$\inf_{M\ge 0}\sup_{f\in A}\int(|f| - M\chi X)^+ \le \inf_{M\ge 0}\sup_{f\in A}\int_{F(f,M)}|f| = 0,$$

so A is uniformly integrable.

(b) If A is uniformly integrable, and $\epsilon > 0$, there is an $M_0 \ge 0$ such that $\int(|f| - M_0\chi X)^+ \le \epsilon$ for every $f \in A$; also, $\gamma = \sup_{f\in A}\int |f|$ is finite (246Ca). Take any $M \ge M_0\max(1, (1+\gamma)/\epsilon)$. If $f \in A$, then

$$|f| \times \chi F(f, M) \le (|f| - M_0\chi X)^+ + M_0\chi F(f, M) \le (|f| - M_0\chi X)^+ + \tfrac{\epsilon}{\gamma+1}|f|$$

everywhere on $\operatorname{dom} f$, so

$$\int_{F(f,M)}|f| \le \int(|f| - M_0\chi X)^+ + \tfrac{\epsilon}{\gamma+1}\int |f| \le 2\epsilon.$$

As ϵ is arbitrary, $\lim_{M\to\infty}\sup_{f\in A}\int_{F(f,M)}|f| = 0$.

246J The next step is to set out some remarkable connexions between uniform integrability and the topology of convergence in measure discussed in the last section.

Theorem Let (X,Σ,μ) be a measure space.

(a) If $\langle f_n\rangle_{n\in\mathbb{N}}$ is a uniformly integrable sequence of real-valued functions on X, and $f(x)=\lim_{n\to\infty} f_n(x)$ for almost every $x\in X$, then f is integrable and $\lim_{n\to\infty}\int|f_n-f|=0$; consequently $\int f=\lim_{n\to\infty}\int f_n$.

(b) If $A\subseteq L^1=L^1(\mu)$ is uniformly integrable, then the norm topology of L^1 and the topology of convergence in measure of $L^0=L^0(\mu)$ agree on A.

(c) For any $u\in L^1$ and any sequence $\langle u_n\rangle_{n\in\mathbb{N}}$ in L^1, the following are equiveridical:

(i) $u=\lim_{n\to\infty}u_n$ for $\| \|_1$;

(ii) $\{u_n:n\in\mathbb{N}\}$ is uniformly integrable and $\langle u_n\rangle_{n\in\mathbb{N}}$ converges to u in measure.

(d) If (X,Σ,μ) is semi-finite, and $A\subseteq L^1$ is uniformly integrable, then the closure $\overline{A}$ of A in L^0 for the topology of convergence in measure is still a uniformly integrable subset of L^1.

proof (a) Note first that because $\sup_{n\in\mathbb{N}}\int|f_n|<\infty$ (246Ca) and $|f|=\liminf_{n\to\infty}|f_n|$, Fatou's Lemma assures us that $|f|$ is integrable, with $\int|f|\le\limsup_{n\to\infty}\int|f_n|$. It follows immediately that $\{f_n-f:n\in\mathbb{N}\}$ is uniformly integrable, being the sum of two uniformly integrable sets (246Cc, 246Ce).

Given $\epsilon>0$, there are $M\ge0$, $E\in\Sigma$ such that $\mu E<\infty$ and $\int(|f_n-f|-M\chi E)^+\le\epsilon$ for every $n\in\mathbb{N}$. Also $|f_n-f|\wedge M\chi E\to0$ a.e., so

$$\begin{aligned}\limsup_{n\to\infty}\int|f_n-f|&\le\limsup_{n\to\infty}\int(|f_n-f|-M\chi E)^+\\&\qquad+\limsup_{n\to\infty}\int|f_n-f|\wedge M\chi E\\&\le\epsilon,\end{aligned}$$

by Lebesgue's Dominated Convergence Theorem. As ϵ is arbitrary, $\lim_{n\to\infty}\int|f_n-f|=0$ and $\lim_{n\to\infty}\int f_n-f=0$.

(b) Let $\mathfrak{T}_A$, $\mathfrak{S}_A$ be the topologies on A induced by the norm topology of L^1 and the topology of convergence in measure on L^0 respectively.

(i) Given $\epsilon>0$, let $F\in\Sigma$, $M\ge0$ be such that $\mu F<\infty$ and $\int(|v|-M\chi F^\bullet)^+\le\epsilon$ for every $v\in A$, and consider $\bar\rho_F$, defined as in 245A. Then for any f, $g\in\mathcal{L}^0$,

$$|f-g|\le(|f|-M\chi F)^++(|g|-M\chi F)^++M(|f-g|\wedge\chi F)$$

everywhere on $\operatorname{dom}f\cap\operatorname{dom}g$, so

$$|u-v|\le(|u|-M\chi F^\bullet)^++(|v|-M\chi F^\bullet)^++M(|u-v|\wedge\chi F^\bullet)$$

for all u, $v\in L^0$. Consequently

$$\|u-v\|_1\le2\epsilon+M\bar\rho_F(u,v)$$

for all u, $v\in A$.

This means that, given $\epsilon>0$, we can find F, M such that, for u, $v\in A$,

$$\bar\rho_F(u,v)\le\tfrac{\epsilon}{1+M}\implies\|u-v\|_1\le3\epsilon.$$

It follows that every subset of A which is open for $\mathfrak{T}_A$ is open for $\mathfrak{S}_A$ (2A3Ib).

(ii) In the other direction, we have $\bar\rho_F(u,v)\le\|u-v\|_1$ for every $u\in L^1$ and every set F of finite measure, so every subset of A which is open for $\mathfrak{S}_A$ is open for $\mathfrak{T}_A$.

(c) If $\langle u_n\rangle_{n\in\mathbb{N}}\to u$ for $\| \|_1$, $A=\{u_n:n\in\mathbb{N}\}$ is uniformly integrable. **P** Given $\epsilon>0$, let m be such that $\|u_n-u\|_1\le\epsilon$ whenever $n\ge m$. Set $v=|u|+\sum_{i\le m}|u_i|\in L^1$, and let $M\ge0$, $E\in\Sigma$ be such that μE is finite and $\int_E(v-M\chi E^\bullet)^+\le\epsilon$. Then, for $w\in A$,

$$(|w|-M\chi E^\bullet)^+\le(|w|-v)^++(v-M\chi E^\bullet)^+,$$

so

$$\int_E(|w|-M\chi E^\bullet)^+\le\|(|w|-v)^+\|_1+\int_E(v-M\chi E^\bullet)^+\le2\epsilon.\ \mathbf{Q}$$

Thus on either hypothesis we can be sure that $\{u_n:n\in\mathbb{N}\}$ and $A=\{u\}\cup\{u_n:n\in\mathbb{N}\}$ are uniformly integrable, so that the two topologies agree on A (by (b)) and $\langle u_n\rangle_{n\in\mathbb{N}}$ converges to u in one topology iff it converges to u in the other.

(d) Because A is $\| \|_1$-bounded (246Ca) and μ is semi-finite, $\overline{A} \subseteq L^1$ (245J(b-i)). Given $\epsilon > 0$, let $M \geq 0$, $E \in \Sigma$ be such that $\mu E < \infty$ and $\int (|u| - M\chi E^\bullet)^+ \leq \epsilon$ for every $u \in A$. Now the maps $u \mapsto |u|$, $u \mapsto u - M\chi E^\bullet$, $u \mapsto u^+ : L^0 \to L^0$ are all continuous for the topology of convergence in measure (245D), while $\{u : \|u\|_1 \leq \epsilon\}$ is closed for the same topology (245J again), so $\{u : u \in L^0, \int (|u| - M\chi E^\bullet)^+ \leq \epsilon\}$ is closed and must include $\overline{A}$. Thus $\int (|u| - M\chi E^\bullet)^+ \leq \epsilon$ for every $u \in \overline{A}$. As ϵ is arbitrary, $\overline{A}$ is uniformly integrable.

246K Complex $\mathcal{L}^1$ and L^1 The definitions and theorems above can be repeated without difficulty for spaces of (equivalence classes of) complex-valued functions, with just one variation: in the complex equivalent of 246F, the constant must be changed. It is easy to see that, for $u \in L^1_{\mathbb{C}}(\mu)$,

$$\begin{aligned}\|u\|_1 &\leq \|\mathcal{R}\mathrm{e}(u)\|_1 + \|\mathcal{I}\mathrm{m}(u)\|_1 \\ &\leq 2 \sup_{F \in \Sigma} \Big|\int_F \mathcal{R}\mathrm{e}(u)\Big| + 2 \sup_{F \in \Sigma} \Big|\int_F \mathcal{I}\mathrm{m}(u)\Big| \leq 4 \sup_{F \in \Sigma} \Big|\int_F u\Big|.\end{aligned}$$

(In fact, $\|u\|_1 \leq \pi \sup_{F \in \Sigma} |\int_F u|$; see 246Yl and 252Yt.) Consequently some of the arguments of 246G need to be written out with different constants, but the results, as stated, are unaffected.

246X Basic exercises (a) Let (X, Σ, μ) be a measure space and A a subset of $L^1 = L^1(\mu)$. Show that the following are equiveridical: (i) A is uniformly integrable; (ii) for every $\epsilon > 0$ there is a $w \geq 0$ in L^1 such that $\int (|u| - w)^+ \leq \epsilon$ for every $u \in A$; (iii) $\langle (|u_{n+1}| - \sup_{i \leq n} |u_i|)^+ \rangle_{n \in \mathbb{N}} \to 0$ in L^1 for every sequence $\langle u_n \rangle_{n \in \mathbb{N}}$ in A. (*Hint*: for (ii)⇒(iii), set $v_n = \sup_{i \leq n} |u_i|$ and note that $\langle v_n \wedge w \rangle_{n \in \mathbb{N}}$ is convergent in L^1 for every $w \geq 0$.)

>(b) Let (X, Σ, μ) be a totally finite measure space. Show that for any $p > 1$ and $M \geq 0$ the set $\{f : f \in \mathcal{L}^p(\mu), \|f\|_p \leq M\}$ is uniformly integrable. (*Hint*: $\int (|f| - M\chi X)^+ \leq M^{1-p} \int |f|^p$.)

>(c) Let μ be counting measure on $\mathbb{N}$. Show that a set $A \subseteq \mathcal{L}^1(\mu) = \ell^1$ is uniformly integrable iff (i) $\sup_{f \in A} |f(n)| < \infty$ for every $n \in \mathbb{N}$ (ii) for every $\epsilon > 0$ there is an $m \in \mathbb{N}$ such that $\sum_{n=m}^{\infty} |f(n)| \leq \epsilon$ for every $f \in A$.

(d) Let X be a set, and let μ be counting measure on X. Show that a set $A \subseteq \mathcal{L}^1(\mu) = \ell^1(X)$ is uniformly integrable iff (i) $\sup_{f \in A} |f(x)| < \infty$ for every $x \in X$ (ii) for every $\epsilon > 0$ there is a finite set $I \subseteq X$ such that $\sum_{x \in X \setminus I} |f(x)| \leq \epsilon$ for every $f \in A$. Show that in this case A is relatively compact for the norm topology of $\ell^1(X)$.

(e) Let (X, Σ, μ) be a measure space, $\delta > 0$, and $\mathcal{I} \subseteq \Sigma$ a family such that (i) every atom belongs to $\mathcal{I}$ (ii) $E \in \mathcal{I}$ whenever $E \in \Sigma$ and $\mu E \leq \delta$ (iii) $E \cup F \in \mathcal{I}$ whenever E, $F \in \mathcal{I}$ and $E \cap F = \emptyset$. Show that every set of finite measure belongs to $\mathcal{I}$.

(f) Let (X, Σ, μ) and (Y, T, ν) be measure spaces and $\phi : X \to Y$ an inverse-measure-preserving function. Show that a set $A \subseteq \mathcal{L}^1(\nu)$ is uniformly integrable iff $\{g\phi : g \in A\}$ is uniformly integrable in $\mathcal{L}^1(\mu)$. (*Hint*: use 246G for 'if', 246A for 'only if'.)

>(g) Let (X, Σ, μ) be a measure space and $p \in [1, \infty[$. Let $\langle f_n \rangle_{n \in \mathbb{N}}$ be a sequence in $\mathcal{L}^p = \mathcal{L}^p(\mu)$ such that $\{|f_n|^p : n \in \mathbb{N}\}$ is uniformly integrable and $f_n \to f$ a.e. Show that $f \in \mathcal{L}^p$ and $\lim_{n \to \infty} \int |f_n - f|^p = 0$.

(h) Let (X, Σ, μ) be a semi-finite measure space and $p \in [1, \infty[$. Let $\langle u_n \rangle_{n \in \mathbb{N}}$ be a sequence in $L^p = L^p(\mu)$ and $u \in L^0(\mu)$. Show that the following are equiveridical: (i) $u \in L^p$ and $\langle u_n \rangle_{n \in \mathbb{N}}$ converges to u for $\| \|_p$ (ii) $\langle u_n \rangle_{n \in \mathbb{N}}$ converges in measure to u and $\{|u_n|^p : n \in \mathbb{N}\}$ is uniformly integrable. (*Hint*: 245Xl.)

(i) Let (X, Σ, μ) be a totally finite measure space, and $1 \leq p < r \leq \infty$. Let $\langle u_n \rangle_{n \in \mathbb{N}}$ be a $\| \|_r$-bounded sequence in $L^r(\mu)$ which converges in measure to $u \in L^0(\mu)$. Show that $\langle u_n \rangle_{n \in \mathbb{N}}$ converges to u for $\| \|_p$. (*Hint*: show that $\{|u_n|^p : n \in \mathbb{N}\}$ is uniformly integrable.)

246Y Further exercises (a) Let (X, Σ, μ) be a totally finite measure space. Show that $A \subseteq \mathcal{L}^1(\mu)$ is uniformly integrable iff there is a convex function $\phi : [0, \infty[\to \mathbb{R}$ such that $\lim_{a \to \infty} \phi(a)/a = \infty$ and $\sup_{f \in A} \int \phi(|f|) < \infty$.

(b) For any metric space (Z, ρ), let $\mathcal{C}_Z$ be the family of closed subsets of Z, and for F, $F' \in \mathcal{C}_Z \setminus \{\emptyset\}$ set $\tilde{\rho}(F, F') = \min(1, \max(\sup_{z \in F} \inf_{z' \in F'} \rho(z, z'), \sup_{z' \in F'} \inf_{z \in F} \rho(z, z')))$. Show that $\tilde{\rho}$ is a metric on $\mathcal{C}_Z \setminus \{\emptyset\}$ (it is the **Hausdorff metric**). Show that if (Z, ρ) is complete then the family $\mathcal{K}_Z \setminus \{\emptyset\}$ of non-empty compact subsets of Z is closed for $\tilde{\rho}$. Now let (X, Σ, μ) be any measure space and take $Z = L^1 = L^1(\mu)$, $\rho(z, z') = \|z - z'\|_1$ for z, $z' \in Z$. Show that the family of non-empty closed uniformly integrable subsets of L^1 is a closed subset of $\mathcal{C}_Z \setminus \{\emptyset\}$ including $\mathcal{K}_Z \setminus \{\emptyset\}$.

(c) Let (X,Σ,μ) be a totally finite measure space and $A \subseteq L^1(\mu)$ a uniformly integrable set. Show that there is a uniformly integrable set $C \supseteq A$ such that (i) C is convex and closed in $L^0(\mu)$ for the topology of convergence in measure (ii) if $u \in C$ and $|v| \le |u|$ then $v \in C$ (iii) if T belongs to the set $\mathcal{T}^+$ of operators from $L^1(\mu) = M^{1,\infty}(\mu)$ to itself, as described in 244Xm, then $T[C] \subseteq C$.

(d) Let μ be Lebesgue measure on $\mathbb{R}$. Show that a set $A \subseteq \mathcal{L}^1(\mu)$ is uniformly integrable iff $\lim_{n\to\infty} \int_{F_n} f_n = 0$ for every disjoint sequence $\langle F_n\rangle_{n\in\mathbb{N}}$ of compact sets in $\mathbb{R}$ and every sequence $\langle f_n\rangle_{n\in\mathbb{N}}$ in A.

(e) Let μ be Lebesgue measure on $\mathbb{R}$. Show that a set $A \subseteq \mathcal{L}^1(\mu)$ is uniformly integrable iff $\lim_{n\to\infty} \int_{G_n} f_n = 0$ for every disjoint sequence $\langle G_n\rangle_{n\in\mathbb{N}}$ of open sets in $\mathbb{R}$ and every sequence $\langle f_n\rangle_{n\in\mathbb{N}}$ in A.

(f) Repeat 246Yd and 246Ye for Lebesgue measure on arbitrary subsets of $\mathbb{R}^r$.

(g) Let X be a set and Σ a σ-algebra of subsets of X. Let $\langle \nu_n\rangle_{n\in\mathbb{N}}$ be a sequence of countably additive functionals on Σ such that $\nu E = \lim_{n\to\infty} \nu_n E$ is defined for every $E \in \Sigma$. Show that $\lim_{n\to\infty} \nu_n F_n = 0$ whenever $\langle F_n\rangle_{n\in\mathbb{N}}$ is a disjoint sequence in Σ. (*Hint*: suppose otherwise. By taking suitable subsequences reduce to the case in which $|\nu_n F_i - \nu F_i| \le 2^{-n}\epsilon$ for $i < n$, $|\nu_n F_n| \ge 3\epsilon$, $|\nu_n F_i| \le 2^{-i}\epsilon$ for $i > n$. Set $F = \bigcup_{i\in\mathbb{N}} F_{2i+1}$ and show that $|\nu_{2n+1}F - \nu_{2n}F| \ge \epsilon$ for every n.)

(h) Let (X,Σ,μ) be a measure space and $\langle u_n\rangle_{n\in\mathbb{N}}$ a sequence in $L^1 = L^1(\mu)$ such that $\lim_{n\to\infty} \int_F u_n$ is defined for every $F \in \Sigma$. Show that $\{u_n : n \in \mathbb{N}\}$ is uniformly integrable. (*Hint*: suppose not. Then there are a disjoint sequence $\langle F_n\rangle_{n\in\mathbb{N}}$ in Σ and a subsequence $\langle u'_n\rangle_{n\in\mathbb{N}}$ of $\langle u_n\rangle_{n\in\mathbb{N}}$ such that $\inf_{n\in\mathbb{N}} |\int_{F_n} u'_n| = \epsilon > 0$. But this contradicts 246Yg.)

(i) In 246Yg, show that ν is countably additive. (*Hint*: Set $\mu = \sum_{n=0}^{\infty} a_n\nu_n$ for a suitable sequence $\langle a_n\rangle_{n\in\mathbb{N}}$ of strictly positive numbers. For each n choose a Radon-Nikodým derivative f_n of ν_n with respect to μ. Show that $\{f_n : n \in \mathbb{N}\}$ is uniformly integrable, so that ν is truly continuous.)(This is the **Vitali-Hahn-Saks theorem**.)

(j) Let (X,Σ,μ) be any measure space, and $A \subseteq L^1(\mu)$. Show that the following are equiveridical: (i) A is $\| \|_1$-bounded; (ii) $\sup_{u\in A} |\int_F u| < \infty$ for every μ-atom $F \in \Sigma$ and $\limsup_{n\to\infty} \sup_{u\in A} |\int_{F_n} u| < \infty$ for every disjoint sequence $\langle F_n\rangle_{n\in\mathbb{N}}$ of measurable sets of finite measure; (iii) $\sup_{u\in A} |\int_E u| < \infty$ for every $E \in \Sigma$. (*Hint*: show that $\langle a_n u_n\rangle_{n\in\mathbb{N}}$ is uniformly integrable whenever $\lim_{n\to\infty} a_n = 0$ in $\mathbb{R}$ and $\langle u_n\rangle_{n\in\mathbb{N}}$ is a sequence in A.)

(k) Let (X,Σ,μ) be a measure space and $A \subseteq L^1(\mu)$ a non-empty set. Show that the following are equiveridical: (i) A is uniformly integrable; (ii) whenever $B \subseteq L^\infty(\mu)$ is non-empty and downwards-directed and has infimum 0 in $L^\infty(\mu)$ then $\inf_{v\in B} \sup_{u\in A} |\int u \times v| = 0$. (*Hint*: for (i)$\Rightarrow$(ii), note that $\inf_{v\in B} w \times v = 0$ for every $w \ge 0$ in L^0. For (ii)$\Rightarrow$(i), use 246G(iv).)

(l) Set $f(x) = e^{ix}$ for $x \in [-\pi,\pi]$. Show that $|\int_E f| \le 2$ for every $E \subseteq [-\pi,\pi]$.

246 Notes and comments I am holding over to the next section the most striking property of uniformly integrable sets (they are the relatively weakly compact sets in L^1) because this demands some non-trivial ideas from functional analysis and general topology. In this section I give the results which can be regarded as essentially measure-theoretic in inspiration. The most important new concept, or technique, is that of 'disjoint-sequence theorem'. A typical example is in condition (iii) of 246G, relating uniform integrability to the behaviour of functionals on disjoint sequences of sets. I give variants of this in 246Yd-246Yf, and 246Yg-246Yj are further results in which similar methods can be used. The central result of the next section (247C) will also use disjoint sequences in the proof, and they will appear more than once in Chapter 35 in the next volume.

The phrase 'uniformly integrable' ought to mean something like 'uniformly approximable by simple functions', and the definition 246A can be forced into such a form, but I do not think it very useful to do so. However condition (ii) of 246G amounts to something like 'uniformly truly continuous', if we think of members of L^1 as truly continuous functionals on Σ, as in 242I. (See 246Yi.) Note that in each of the statements (ii)-(iv) of 246G we need to take special note of any atoms for the measure, since they are not controlled by the main condition imposed. In an atomless measure space, of course, we have a simplification here, as in 246Yd-246Yf.

Another way of justifying the 'uniformly' in 'uniformly integrable' is by considering functionals θ_w where $w \ge 0$ in L^1, setting $\theta_w(u) = \int(|u| - w)^+$ for $u \in L^1$; then $A \subseteq L^1$ is uniformly integrable iff $\theta_w \to 0$ uniformly on A as w rises in L^1 (246Xa). It is sometimes useful to know that if this is true at all then it is necessarily witnessed by elements w which can be built directly from materials at hand (see (iii) of 246Xa). Furthermore, the sets $A_{w\epsilon} = \{u : \theta_w(u) \le \epsilon\}$ are always convex, $\| \|_1$-closed and 'solid' (if $u \in A_{w\epsilon}$ and $|v| \le |u|$ then $v \in A_{w\epsilon}$)(246Cd); they are closed under pointwise convergence of sequences (246Ja) and in semi-finite measure spaces they are closed for the topology of convergence in measure (246Jd); in probability spaces, for level w, they are closed under conditional expectations (246D) and similar operators (246Yc). Consequently we can expect that any uniformly integrable set will be included in a uniformly integrable set which is closed under operations of several different types.

Yet another 'uniform' property of uniformly integrable sets is in 246Yk. The norm $\| \|_\infty$ is never (in interesting cases) order-continuous in the way that other $\| \|_p$ are (244Ye); but the uniformly integrable subsets of L^1 provide interesting order-continuous seminorms on L^∞.

246J supplements results from §245. In the notes to that section I mentioned the question: if $\langle f_n\rangle_{n\in\mathbb{N}} \to f$ a.e., in what ways can $\langle\int f_n\rangle_{n\in\mathbb{N}}$ fail to converge to $\int f$? Here we find that $\langle\int |f_n - f|\rangle_{n\in\mathbb{N}} \to 0$ iff $\{f_n : n \in \mathbb{N}\}$ is uniformly integrable; this is a way of making precise the expression 'none of the weight of the sequence is lost at infinity'. Generally, for sequences, convergence in $\| \|_p$, for $p \in [1,\infty[$, is convergence in measure for pth-power-uniformly-integrable sequences (246Xh).

247 Weak compactness in L^1

I now come to the most striking feature of uniform integrability: it provides a description of the relatively weakly compact subsets of L^1 (247C). I have put this into a separate section because it demands some knowledge of functional analysis – in particular, of course, of weak topologies on Banach spaces. I will try to give an account in terms which are accessible to novices in the theory of normed spaces because the result is essentially measure-theoretic, as well as being of vital importance to applications in probability theory. I have written out the essential definitions in §§2A3-2A5.

247A Part of the argument of the main theorem below will run more smoothly if I separate out an idea which is, in effect, a simple special case of a theme which has been running through the exercises of this chapter (241Yg, 242Yb, 243Ya, 244Yd).

Lemma Let (X,Σ,μ) be a measure space, and G any member of Σ. Let μ_G be the subspace measure on G, so that $\mu_G E = \mu E$ for $E \subseteq G$, $E \in \Sigma$. Set

$$U = \{u : u \in L^1(\mu),\ u \times \chi G^\bullet = u\} \subseteq L^1(\mu).$$

Then we have an isomorphism S between the ordered normed spaces U and $L^1(\mu_G)$, given by writing

$$S(f^\bullet) = (f{\restriction}G)^\bullet$$

for every $f \in \mathcal{L}^1(\mu)$ such that $f^\bullet \in U$.

proof Of course I should remark explicitly that U is a linear subspace of $L^1(\mu)$. I have discussed integration over subspaces in §§131 and 214; in particular, I noted that $f{\restriction}G$ is integrable, and that

$$\int |f{\restriction}G| d\mu_G = \int |f| \times \chi G\, d\mu \le \int |f| d\mu$$

for every $f \in \mathcal{L}^1(\mu)$ (131Fa). If f, $g \in \mathcal{L}^1(\mu)$ and $f = g$ μ-a.e., then $f{\restriction}G = g{\restriction}G$ μ_G-a.e.; so the proposed formula for S does indeed define a map from U to $L^1(\mu_G)$.

Because

$$(f+g){\restriction}G = (f{\restriction}G) + (g{\restriction}G), \quad (cf){\restriction}G = c(f{\restriction}G)$$

for all f, $g \in \mathcal{L}^1(\mu)$ and all $c \in \mathbb{R}$, S is linear. Because

$$f \le g \ \mu\text{-a.e.} \Longrightarrow f{\restriction}G \le g{\restriction}G \ \mu_G\text{-a.e.},$$

S is order-preserving. Because $\int |f{\restriction}G| d\mu_G \le \int |f| d\mu$ for every $f \in \mathcal{L}^1(\mu)$, $\|Su\|_1 \le \|u\|_1$ for every $u \in U$.

To see that S is surjective, take any $v \in L^1(\mu_G)$. Express v as $g^\bullet$ where $g \in \mathcal{L}^1(\mu_G)$. By 131E, $f \in \mathcal{L}^1(\mu)$, where $f(x) = g(x)$ for $x \in \operatorname{dom} g$, 0 for $x \in X \setminus G$; so that $f^\bullet \in U$ and $f{\restriction}G = g$ and $v = S(f^\bullet) \in S[U]$.

To see that S is norm-preserving, note that, for any $f \in \mathcal{L}^1(\mu)$,

$$\int |f{\restriction}G| d\mu_G = \int |f| \times \chi G\, d\mu,$$

so that if $u = f^\bullet \in U$ we shall have

$$\|Su\|_1 = \int |f{\restriction}G| d\mu_G = \int |f| \times \chi G\, d\mu = \|u \times \chi G^\bullet\|_1 = \|u\|_1.$$

247B Corollary Let (X, Σ, μ) be any measure space, and let $G \in \Sigma$ be a measurable set expressible as a countable union of sets of finite measure. Define U as in 247A, and let $h : L^1(\mu) \to \mathbb{R}$ be any continuous linear functional. Then there is a $v \in L^\infty(\mu)$ such that $h(u) = \int u \times v \, d\mu$ for every $u \in U$.

proof Let $S : U \to L^1(\mu_G)$ be the isomorphism described in 247A. Then $S^{-1} : L^1(\mu_G) \to U$ is linear and continuous, so $h_1 = hS^{-1}$ belongs to the normed space dual $(L^1(\mu_G))^*$ of $L^1(\mu_G)$. Now of course μ_G is σ-finite, therefore localizable (211L), so 243Gb tells us that there is a $v_1 \in L^\infty(\mu_G)$ such that

$$h_1(u) = \int u \times v_1 d\mu_G$$

for every $u \in L^1(\mu_G)$.

Express v_1 as $g_1^\bullet$ where $g_1 : G \to \mathbb{R}$ is a bounded measurable function. Set $g(x) = g_1(x)$ for $x \in G$, 0 for $x \in X \setminus G$; then $g : X \to \mathbb{R}$ is a bounded measurable function, and $v = g^\bullet \in L^\infty(\mu)$. If $u \in U$, express u as $f^\bullet$ where $f \in \mathcal{L}^1(\mu)$; then

$$\begin{aligned} h(u) = h(S^{-1}Su) = h_1((f\restriction G)^\bullet) &= \int (f\restriction G) \times g_1 d\mu_G \\ = \int (f \times g)\restriction G \, d\mu_G &= \int f \times g \times \chi G \, d\mu = \int f \times g \, d\mu = \int u \times v. \end{aligned}$$

As u is arbitrary, this proves the result.

247C Theorem Let (X, Σ, μ) be any measure space and A a subset of $L^1 = L^1(\mu)$. Then A is uniformly integrable iff it is relatively compact in L^1 for the weak topology of L^1.

proof (a) Suppose that A is relatively compact for the weak topology. I seek to show that it satisfies the condition (iii) of 246G.

(i) If $F \in \Sigma$, then surely $\sup_{u \in A} |\int_F u| < \infty$, because $u \mapsto \int_F u$ belongs to $(L^1)^*$, and if $h \in (L^1)^*$ then the image of any relatively weakly compact set under h must be bounded (2A5Ie).

(ii) Now suppose that $\langle F_n \rangle_{n \in \mathbb{N}}$ is a disjoint sequence in Σ. **?** Suppose, if possible, that $\langle \sup_{u \in A} |\int_{F_n} u| \rangle_{n \in \mathbb{N}}$ does not converge to 0. Then there is a strictly increasing sequence $\langle n(k) \rangle_{k \in \mathbb{N}}$ in $\mathbb{N}$ such that

$$\gamma = \frac{1}{2} \inf_{k \in \mathbb{N}} \sup_{u \in A} |\int_{F_{n(k)}} u| > 0.$$

For each k, choose $u_k \in A$ such that $|\int_{F_{n(k)}} u_k| \geq \gamma$. Because A is relatively compact for the weak topology, there is a cluster point u of $\langle u_k \rangle_{k \in \mathbb{N}}$ in L^1 for the weak topology (2A3Ob). Set $\eta_j = 2^{-j}\gamma/6 > 0$ for each $j \in \mathbb{N}$.

We can now choose a strictly increasing sequence $\langle k(j) \rangle_{j \in \mathbb{N}}$ inductively so that, for each j,

$$\int_{F_{n(k(j))}} (|u| + \sum_{i=0}^{j-1} |u_{k(i)}|) \leq \eta_j$$

$$\sum_{i=0}^{j-1} |\int_{F_{n(k(i))}} u - \int_{F_{n(k(i))}} u_{k(j)}| \leq \eta_j$$

for every j, interpreting $\sum_{i=0}^{-1}$ as 0. **P** Given $\langle k(i) \rangle_{i<j}$, set $v^* = |u| + \sum_{i=0}^{j-1} |u_{k(i)}|$; then $\lim_{k \to \infty} \int_{F_{n(k)}} v^* = 0$, by Lebesgue's Dominated Convergence Theorem or otherwise, so there is a k^* such that $k^* > k(i)$ for every $i < j$ and $\int_{F_{n(k)}} v^* \leq \eta_j$ for every $k \geq k^*$. Next,

$$w \mapsto \sum_{i=0}^{j-1} |\int_{F_{n(k(i))}} u - \int_{F_{n(k(i))}} w| : L^1 \to \mathbb{R}$$

is continuous for the weak topology of L^1 and zero at u, and u belongs to every weakly open set containing $\{u_k : k \geq k^*\}$, so there is a $k(j) \geq k^*$ such that $\sum_{i=0}^{j-1} |\int_{F_{n(k(i))}} u - \int_{F_{n(k(i))}} u_{k(j)}| < \eta_j$, which continues the construction. **Q**

Let v be any cluster point in L^1, for the weak topology, of $\langle u_{k(j)} \rangle_{j \in \mathbb{N}}$. Setting $G_i = F_{n(k(i))}$, we have $|\int_{G_i} u - \int_{G_i} u_{k(j)}| \leq \eta_j$ whenever $i < j$, so $\lim_{j \to \infty} \int_{G_i} u_{k(j)}$ exists $= \int_{G_i} u$ for each i, and $\int_{G_i} v = \int_{G_i} u$ for every i; setting $G = \bigcup_{i \in \mathbb{N}} G_i$,

$$\int_G v = \sum_{i=0}^{\infty} \int_{G_i} v = \sum_{i=0}^{\infty} \int_{G_i} u = \int_G u,$$

by 232D, because $\langle G_i \rangle_{i \in \mathbb{N}}$ is disjoint.

For each $j \in \mathbb{N}$,

$$\begin{aligned}\sum_{i=0}^{j-1}|\int_{G_i}u_{k(j)}|+\sum_{i=j+1}^{\infty}|\int_{G_i}u_{k(j)}|\\&\le\sum_{i=0}^{j-1}\int_{G_i}|u|+\sum_{i=0}^{j-1}|\int_{G_i}u-\int_{G_i}u_{k(j)}|+\sum_{i=j+1}^{\infty}\int_{G_i}|u_{k(j)}|\\&\le\sum_{i=0}^{j-1}\eta_i+\eta_j+\sum_{i=j+1}^{\infty}\eta_i=\sum_{i=0}^{\infty}\eta_i=\frac{\gamma}{3}.\end{aligned}$$

On the other hand, $|\int_{G_j}u_{k(j)}|\ge\gamma$. So

$$|\int_G u_{k(j)}|=|\sum_{i=0}^{\infty}\int_{G_i}u_{k(j)}|\ge\frac{2}{3}\gamma.$$

This is true for every j; because every weakly open set containing v meets $\{u_{k(j)}:j\in\mathbb{N}\}$, $|\int_G v|\ge\frac{2}{3}\gamma$ and $|\int_G u|\ge\frac{2}{3}\gamma$. On the other hand,

$$|\int_G u|=|\sum_{i=0}^{\infty}\int_{G_i}u|\le\sum_{i=0}^{\infty}\int_{G_i}|u|\le\sum_{i=0}^{\infty}\eta_i=\frac{\gamma}{3},$$

which is absurd. **X**

This contradiction shows that $\lim_{n\to\infty}\sup_{u\in A}|\int_{F_n}u|=0$. As $\langle F_n\rangle_{n\in\mathbb{N}}$ is arbitrary, A satisfies the condition 246G(iii) and is uniformly integrable.

(b) Now assume that A is uniformly integrable. I seek a weakly compact set $C\supseteq A$.

(i) For each $n\in\mathbb{N}$, choose $E_n\in\Sigma$, $M_n\ge 0$ such that $\mu E_n<\infty$ and $\int(|u|-M_n\chi E_n^\bullet)^+\le 2^{-n}$ for every $u\in A$. Set

$$C=\{v:v\in L^1,\ |\int_F v|\le M_n\mu(F\cap E_n)+2^{-n}\ \forall\ n\in\mathbb{N},\ F\in\Sigma\},$$

and note that $A\subseteq C$, because if $u\in A$ and $F\in\Sigma$,

$$|\int_F u|\le\int_F(|u|-M_n\chi E_n^\bullet)^+ +\int_F M_n\chi E_n^\bullet\le 2^{-n}+M_n\mu(F\cap E_n)$$

for every n. Observe also that C is $\| \|_1$-bounded, because

$$\|u\|_1\le 2\sup_{F\in\Sigma}|\int_F u|\le 2(1+M_0\mu(F\cap E_0))\le 2(1+M_0\mu E_0)$$

for every $u\in C$ (using 246F).

(ii) Because I am seeking to prove this theorem for arbitrary measure spaces (X,Σ,μ), I cannot use 243G to identify the dual of L^1. Nevertheless, 247B above shows that 243Gb it is 'nearly' valid, in the following sense: if $h\in(L^1)^*$, there is a $v\in L^\infty$ such that $h(u)=\int u\times v$ for every $u\in C$. **P** Set $G=\bigcup_{n\in\mathbb{N}}E_n\in\Sigma$, and define $U\subseteq L^1$ as in 247A-247B. By 247B, there is a $v\in L^\infty$ such that $h(u)=\int u\times v$ for every $u\in U$. But if $u\in C$, we can express u as $f^\bullet$ where $f:X\to\mathbb{R}$ is measurable. If $F\in\Sigma$ and $F\cap G=\emptyset$, then

$$|\int_F f|=|\int_F u|\le 2^{-n}+M_n\mu(F\cap E_n)=2^{-n}$$

for every $n\in\mathbb{N}$, so $\int_F f=0$; it follows that $f=0$ a.e. on $X\setminus G$ (131Fc), so that $f\times\chi G=_{\text{a.e.}}f$ and $u=u\times\chi G^\bullet$, that is, $u\in U$, and $h(u)=\int u\times v$, as required. **Q**

(iii) So we may proceed, having an adequate description, not of $(L^1(\mu))^*$ itself, but of its action on C.

Let $\mathcal{F}$ be any ultrafilter on L^1 containing C (see 2A3R). For each $F\in\Sigma$, set

$$\nu F=\lim_{u\to\mathcal{F}}\int_F u;$$

because

$$\sup_{u\in C}|\int_F u|\le\sup_{u\in C}\|u\|_1<\infty,$$

this is well-defined in $\mathbb{R}$ (2A3Se). If E, F are disjoint members of Σ, then $\int_{E\cup F}u=\int_E u+\int_F u$ for every $u\in C$, so

$$\nu(E\cup F)=\lim_{u\to\mathcal{F}}\int_{E\cup F}u=\lim_{u\to\mathcal{F}}\int_E u+\lim_{u\to\mathcal{F}}\int_F u=\nu E+\nu F$$

(2A3Sf). Thus $\nu : \Sigma \to \mathbb{R}$ is additive. Next, it is truly continuous with respect to μ. **P** Given $\epsilon > 0$, take $n \in \mathbb{N}$ such that $2^{-n} \le \frac{1}{2}\epsilon$, set $\delta = \epsilon/2(M_n + 1) > 0$ and observe that

$$|\nu F| \le \sup_{u\in C} |\int_F u| \le 2^{-n} + M_n\mu(F \cap E_n) \le \epsilon$$

whenever $\mu(F \cap E_n) \le \delta$. **Q** By the Radon-Nikodým theorem (232E), there is an $f_0 \in \mathcal{L}^1$ such that $\int_F f_0 = \nu F$ for every $F \in \Sigma$. Set $u_0 = f_0^\bullet \in L^1$. If $n \in \mathbb{N}$, $F \in \Sigma$ then

$$|\int_F u_0| = |\nu F| \le \sup_{u\in C} |\int_F u| \le 2^{-n} + M_n\mu(F \cap E_n),$$

so $u_0 \in C$.

(iv) Of course the point is that $\mathcal{F}$ converges to u_0. **P** Let $h \in (L^1)^*$. Then there is a $v \in L^\infty$ such that $h(u) = \int u \times v$ for every $u \in C$. Express v as $g^\bullet$, where $g : X \to \mathbb{R}$ is bounded and Σ-measurable. Let $\epsilon > 0$. Take $a_0 \le a_1 \le \ldots \le a_n$ such that $a_{i+1} - a_i \le \epsilon$ for each i while $a_0 \le g(x) < a_n$ for each $x \in X$. Set $F_i = \{x : a_{i-1} \le g(x) < a_i\}$ for $1 \le i \le n$, and set $\tilde{g} = \sum_{i=1}^n a_i\chi F_i$, $\tilde{v} = \tilde{g}^\bullet$; then $\|\tilde{v} - v\|_\infty \le \epsilon$. We have

$$\begin{aligned}\int u_0 \times \tilde{v} &= \sum_{i=1}^n a_i \int_{F_i} u = \sum_{i=1}^n a_i\nu F_i \\ &= \sum_{i=1}^n a_i \lim_{u\to\mathcal{F}} \int_{F_i} u = \lim_{u\to\mathcal{F}} \sum_{i=1}^n a_i \int_{F_i} u = \lim_{u\to\mathcal{F}} \int u \times \tilde{v}.\end{aligned}$$

Consequently

$$\begin{aligned}\limsup_{u\to\mathcal{F}} |\int u \times v - \int u_0 \times v| &\le |\int u_0 \times v - \int u_0 \times \tilde{v}| + \sup_{u\in C} |\int u \times v - \int u \times \tilde{v}| \\ &\le \|u_0\|_1\|v - \tilde{v}\|_\infty + \sup_{u\in C} \|u\|_1\|v - \tilde{v}\|_\infty \\ &\le 2\epsilon \sup_{u\in C} \|u\|_1.\end{aligned}$$

As ϵ is arbitrary,

$$\limsup_{u\to\mathcal{F}} |h(u) - h(u_0)| = \limsup_{u\to\mathcal{F}} |\int u \times v - \int u_0 \times v| = 0.$$

As h is arbitrary, u_0 is a limit of $\mathcal{F}$ in C for the weak topology of L^1. **Q**

As $\mathcal{F}$ is arbitrary, C is weakly compact in L^1, and the proof is complete.

247D Corollary Let (X, Σ, μ) and (Y, T, ν) be any two measure spaces, and $T : L^1(\mu) \to L^1(\nu)$ a continuous linear operator. Then $T[A]$ is a uniformly integrable subset of $L^1(\nu)$ whenever A is a uniformly integrable subset of $L^1(\mu)$.

proof The point is that T is continuous for the respective weak topologies (2A5If). If $A \subseteq L^1(\mu)$ is uniformly integrable, then there is a weakly compact $C \supseteq A$, by 247C; $T[C]$, being the image of a compact set under a continuous map, must be weakly compact (2A3N(b-ii)); so $T[C]$ and $T[A]$ are uniformly integrable by the other half of 247C.

247E Complex L^1 There are no difficulties, and no surprises, in proving 247C for $L^1_{\mathbb{C}}$. If we follow the same proof, everything works, but of course we must remember to change the constant when applying 246F, or rather 246K, in part (b-i) of the proof.

247X Basic exercises >(a) Let (X, Σ, μ) be any measure space. Show that if $A \subseteq L^1 = L^1(\mu)$ is relatively weakly compact, then $\{v : v \in L^1, |v| \le |u| \text{ for some } u \in A\}$ is relatively weakly compact.

(b) Let (X, Σ, μ) be a measure space. On $L^1 = L^1(\mu)$ define pseudometrics ρ_F, ρ'_w for $F \in \Sigma$, $w \in L^\infty(\mu)$ by setting $\rho_F(u, v) = |\int_F u - \int_F v|$, $\rho'_w(u, v) = |\int u \times w - \int v \times w|$ for u, $v \in L^1$. Show that on any $\| \|_1$-bounded subset of L^1, the topology defined by $\{\rho_F : F \in \Sigma\}$ agrees with the topology generated by $\{\rho'_w : w \in L^\infty\}$.

>(c) Show that for any set X a subset of $\ell^1 = \ell^1(X)$ is compact for the weak topology of ℓ^1 iff it is compact for the norm topology of ℓ^1. (*Hint*: 246Xd.)

(d) Use the argument of (a-ii) in the proof of 247C to show directly that if $A \subseteq \ell^1(\mathbb{N})$ is weakly compact then $\inf_{n\in\mathbb{N}} |u_n(n)| = 0$ for any sequence $\langle u_n\rangle_{n\in\mathbb{N}}$ in A.

(e) Let (X, Σ, μ) and (Y, T, ν) be measure spaces, and $T : L^2(\nu) \to L^1(\mu)$ any bounded linear operator. Show that $\{Tu : u \in L^2(\nu),\ \|u\|_2 \le 1\}$ is uniformly integrable in $L^1(\mu)$. (*Hint*: use 244K to see that $\{u : \|u\|_2 \le 1\}$ is weakly compact in $L^2(\nu)$.)

247Y Further exercises (a) Let (X, Σ, μ) be a measure space. Take $1 < p < \infty$ and $M \ge 0$ and set $A = \{u : u \in L^p = L^p(\mu),\ \|u\|_p \le M\}$. Write $\mathfrak{S}_A$ for the topology of convergence in measure on A, that is, the subspace topology induced by the topology of convergence in measure on $L^0(\mu)$. Show that if $h \in (L^p)^*$ then $h\restriction A$ is continuous for $\mathfrak{S}_A$; so that if $\mathfrak{T}$ is the weak topology on L^p, then the subspace topology $\mathfrak{T}_A$ is included in $\mathfrak{S}_A$.

(b) Let (X, Σ, μ) be a measure space and $\langle u_n\rangle_{n\in\mathbb{N}}$ a sequence in $L^1 = L^1(\mu)$ such that $\lim_{n\to\infty} \int_F u_n$ is defined for every $F \in \Sigma$. Show that $\{u_n : n \in \mathbb{N}\}$ is weakly convergent. (*Hint*: 246Yh.) Find an alternative argument relying on 2A5J and the result of 246Yj.

247 Notes and comments In 247D and 247Xa I try to suggest the power of the identification between weak compactness and uniform integrability. That a continuous image of a weakly compact set should be weakly compact is a commonplace of functional analysis; that the solid hull of a uniformly integrable set should be uniformly integrable is immediate from the definition. But I see no simple arguments to show that a continuous image of a uniformly integrable set should be uniformly integrable, or that the solid hull of a weakly compact set should be relatively weakly compact. (Concerning the former, an alternative route does exist; see 371Xf in the next volume.)

I can distinguish two important ideas in the proof of 247C. The first, in (a-ii) of the proof, is a careful manipulation of sequences; it is the argument needed to show that a weakly compact subset of ℓ^1 is norm-compact. (You may find it helpful to write out a solution to 247Xd.) The $F_{n(k)}$ and u_k are chosen to mimic the situation in which we have a sequence in ℓ^1 such that $u_k(k) = 1$ for each k. The $k(i)$ are chosen so that the 'hump' moves sufficiently rapidly along for $u_{k(j)}(k(i))$ to be very small whenever $i \ne j$. But this means that $\sum_{i=0}^\infty u_{k(j)}(k(i))$ (corresponding to $\int_G u_{k(j)}$ in the proof) is always substantial, while $\sum_{i=0}^\infty v(k(i))$ will be small for any putative cluster point v of $\langle u_{k(j)}\rangle_{j\in\mathbb{N}}$. I used similar techniques in §246; compare 246Yg.

In the other half of the proof of 247C, the strategy is clearer. Members of L^1 correspond to truly continuous functionals on Σ; the uniform integrability of C makes the corresponding set of functionals 'uniformly truly continuous', so that any limit functional will also be truly continuous and will give us a member of L^1 via the Radon-Nikodým theorem. A straightforward approximation argument ((a-iv) in the proof, and 247Xb) shows that $\lim_{u\in\mathcal{F}} \int u \times w = \int v \times w$ for every $w \in L^\infty$. For localizable measures μ, this would complete the proof. For the general case, we need another step, here done in 247A-247B; a uniformly integrable subset of L^1 effectively lives on a σ-finite part of the measure space, so that we can ignore the rest of the measure and suppose that we have a localizable measure space.

The conditions (ii)-(iv) of 246G make it plain that weak compactness in L^1 can be effectively discussed in terms of sequences; see also 246Yh. I should remark that this is a general feature of weak compactness in Banach spaces (2A5J). Of course the disjoint-sequence formulations in 246G are characteristic of L^1 – I mean that while there are similar results applicable elsewhere (see Fremlin 74, chap. 8), the ideas are clearest and most dramatically expressed in their application to L^1.

Chapter 25

Product Measures

I come now to another chapter on 'pure' measure theory, discussing a fundamental construction – or, as you may prefer to consider it, two constructions, since the problems involved in forming the product of two arbitrary measure spaces (§251) are rather different from those arising in the product of arbitrarily many probability spaces (§254). This work is going to stretch our technique to the utmost, for while the fundamental theorems to which we are moving are natural aims, the proofs are lengthy and there are many pitfalls beside the true paths.

The central idea is that of 'repeated integration'. You have probably already seen formulae of the type '$\iint f(x,y)dxdy$' used to calculate the integral of a function of two real variables over a region in the plane. One of the basic techniques of advanced calculus is reversing the order of integration; for instance, we expect $\int_0^1(\int_y^1 f(x,y)dx)dy$ to be equal to $\int_0^1(\int_0^x f(x,y)dy)dx$. As I have developed the subject, we already have a third calculation to compare with these two: $\int_D f$, where $D=\{(x,y):0\le y\le x\le 1\}$ and the integral is taken with respect to Lebesgue measure on the plane. The first two sections of this chapter are devoted to an analysis of the relationship between one- and two-dimensional Lebesgue measure which makes these operations valid – some of the time; part of the work has to be devoted to a careful description of the exact conditions which must be imposed on f and D if we are to be safe.

Repeated integration, in one form or another, appears everywhere in measure theory, and it is therefore necessary sooner or later to develop the most general possible expression of the idea. The standard method is through the theory of products of general measure spaces. Given measure spaces (X,Σ,μ) and (Y,T,ν), the aim is to find a measure λ on $X\times Y$ which will, at least, give the right measure $\mu E\cdot\nu F$ to a 'rectangle' $E\times F$ where $E\in\Sigma$ and $F\in\mathrm{T}$. It turns out that there are already difficulties in deciding what 'the' product measure is, and to do the job properly I find I need, even at this stage, to describe two related but distinguishable constructions. These constructions and their elementary properties take up the whole of §251. In §252 I turn to integration over the product, with Fubini's and Tonelli's theorems relating $\int f d\lambda$ with $\iint f(x,y)\mu(dx)\nu(dy)$. Because the construction of λ is symmetric between the two factors, this automatically provides theorems relating $\iint f(x,y)\mu(dx)\nu(dy)$ with $\iint f(x,y)\nu(dy)\mu(dx)$. §253 looks at the space $L^1(\lambda)$ and its relationship with $L^1(\mu)$ and $L^1(\nu)$.

For general measure spaces, there are obstacles in the way of forming an infinite product; to start with, if $\langle(X_n,\mu_n)\rangle_{n\in\mathbb{N}}$ is a sequence of measure spaces, then a product measure λ on $X=\prod_{n\in\mathbb{N}}X_n$ ought to set $\lambda X=\prod_{n=0}^{\infty}\mu_nX_n$, and there is no guarantee that the product will converge, or behave well when it does. But for probability spaces, when $\mu_nX_n=1$ for every n, this problem at least evaporates. It is possible to define the product of any family of probability spaces; this is the burden of §254.

I end the chapter with three sections which are a preparation for Chapters 27 and 28, but are also important in their own right as an investigation of the way in which the group structure of $\mathbb{R}^r$ interacts with Lebesgue and other measures. §255 deals with the 'convolution' $f*g$ of two functions, where $(f*g)(x)=\int f(y)g(x-y)dy$ (the integration being with respect to Lebesgue measure). In §257 I show that some of the same ideas, suitably transformed, can be used to describe a convolution $\nu_1*\nu_2$ of two measures on $\mathbb{R}^r$; in preparation for this I include a section on Radon measures on $\mathbb{R}^r$ (§256).

251 Finite products

The first construction to set up is the product of a pair of measure spaces. It turns out that there are already substantial technical difficulties in the way of finding a canonical universally applicable method. I find myself therefore describing two related, but distinct, constructions, the 'primitive' and 'c.l.d.' product measures (251C, 251F). After listing the fundamental properties of the c.l.d product measure (251I-251J), I work through the identification of the product of Lebesgue measure with itself (251N) and a fairly thorough discussion of subspaces (251O-251S).

251A Definition Let (X,Σ,μ) and (Y,T,ν) be two measure spaces. For $A\subseteq X\times Y$ set

$$\theta A=\inf\{\textstyle\sum_{n=0}^{\infty}\mu E_n\cdot\nu F_n : E_n\in\Sigma,\ F_n\in\mathrm{T}\ \forall\ n\in\mathbb{N},\ A\subseteq\bigcup_{n\in\mathbb{N}}E_n\times F_n\}.$$

Remark In the products $\mu E_n\cdot\nu F_n$, $0\cdot\infty$ is to be taken as 0, as in §135.

251B Lemma In the context of 251A, θ is an outer measure on $X\times Y$.

proof (a) Setting $E_n=F_n=\emptyset$ for every $n\in\mathbb{N}$, we see that $\theta\emptyset=0$.

(b) If $A\subseteq B\subseteq X\times Y$, then whenever $B\subseteq\bigcup_{n\in\mathbb{N}}E_n\times F_n$ we shall have $A\subseteq\bigcup_{n\in\mathbb{N}}E_n\times F_n$; so $\theta A\le\theta B$.

(c) Let $\langle A_n\rangle_{n\in\mathbb{N}}$ be a sequence of subsets of $X\times Y$, with union A. For any $\epsilon>0$, we may choose, for each $n\in\mathbb{N}$, sequences $\langle E_{nm}\rangle_{m\in\mathbb{N}}$ in Σ and $\langle F_{nm}\rangle_{m\in\mathbb{N}}$ in T such that $A_n\subseteq\bigcup_{m\in\mathbb{N}}E_{nm}\times F_{nm}$ and $\sum_{m=0}^{\infty}\mu E_{nm}\cdot\nu F_{nm}\le\theta A_n+2^{-n}\epsilon$. Because $\mathbb{N}\times\mathbb{N}$ is countable, we have a bijection $k\mapsto(n_k,m_k):\mathbb{N}\to\mathbb{N}\times\mathbb{N}$, and now

$$A\subseteq\bigcup_{n,m\in\mathbb{N}}E_{nm}\times F_{nm}=\bigcup_{k\in\mathbb{N}}E_{n_km_k}\times F_{n_km_k},$$

so that

$$\begin{aligned}\theta A&\le\sum_{k=0}^{\infty}\mu E_{n_km_k}\cdot\nu F_{n_km_k}=\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\mu E_{nm}\cdot\nu F_{nm}\\&\le\sum_{n=0}^{\infty}\theta A_n+2^{-n}\epsilon=2\epsilon+\sum_{n=0}^{\infty}\theta A_n.\end{aligned}$$

As ϵ is arbitrary, $\theta A\le\sum_{n=0}^{\infty}\theta A_n$.

As $\langle A_n\rangle_{n\in\mathbb{N}}$ is arbitrary, θ is an outer measure.

251C Definition Let (X,Σ,μ) and (Y,T,ν) be measure spaces. By the **primitive product measure** on $X\times Y$ I shall mean the measure λ_0 derived by Carathéodory's method (113C) from the outer measure θ defined in 251A.

Remark I ought to point out that there is no general agreement on what 'the' product measure on $X\times Y$ should be. Indeed in 251F below I will introduce an alternative one, and in the notes to this section I will mention a third.

251D Definition It is convenient to have a name for a natural construction for σ-algebras. If X and Y are sets with σ-algebras $\Sigma\subseteq\mathcal{P}X$ and $\mathrm{T}\subseteq\mathcal{P}Y$, I will write $\Sigma\widehat{\otimes}\mathrm{T}$ for the σ-algebra of subsets of $X\times Y$ generated by $\{E\times F:E\in\Sigma,\ F\in\mathrm{T}\}$.

251E Proposition Let (X,Σ,μ) and (Y,T,ν) be measure spaces; let λ_0 be the primitive product measure on $X\times Y$, and Λ its domain. Then $\Sigma\widehat{\otimes}\mathrm{T}\subseteq\Lambda$ and $\lambda_0(E\times F)=\mu E\cdot\nu F$ for all $E\in\Sigma$ and $F\in\mathrm{T}$.

proof Throughout this proof, write $\Sigma^f=\{E:E\in\Sigma,\ \mu E<\infty\}$, $\mathrm{T}^f=\{F:F\in\mathrm{T},\ \nu F<\infty\}$.

(a) Suppose that $E\in\Sigma$ and $A\subseteq X\times Y$. For any $\epsilon>0$, there are sequences $\langle E_n\rangle_{n\in\mathbb{N}}$ in Σ and $\langle F_n\rangle_{n\in\mathbb{N}}$ in T such that $A\subseteq\bigcup_{n\in\mathbb{N}}E_n\times F_n$ and $\sum_{n=0}^{\infty}\mu E_n\cdot\nu E_n\le\theta A+\epsilon$. Now

$$A\cap(E\times Y)\subseteq\bigcup_{n\in\mathbb{N}}(E_n\cap E)\times F_n,\quad A\setminus(E\times Y)\subseteq\bigcup_{n\in\mathbb{N}}(E_n\setminus E)\times F_n,$$

so

$$\begin{aligned}\theta(A\cap(E\times Y))+\theta(A\setminus(E\times Y))&\le\sum_{n=0}^{\infty}\mu(E_n\cap E)\cdot\nu F_n+\sum_{n=0}^{\infty}\mu(E_n\setminus E)\cdot\nu F_n\\&=\sum_{n=0}^{\infty}\mu E_n\cdot\nu F_n\le\theta A+\epsilon.\end{aligned}$$

As ϵ is arbitrary, $\theta(A\cap(E\times Y))+\theta(A\setminus(E\times Y))\le\theta A$. And this is enough to ensure that $E\times Y\in\Lambda$ (see 113D).

(b) Similarly, $X\times F\in\Lambda$ for every $F\in\mathrm{T}$, so $E\times F=(E\times Y)\cap(X\times F)\in\Lambda$ for every $E\in\Sigma$, $F\in\mathrm{T}$.

Because Λ is a σ-algebra, it must include the smallest σ-algebra containing all the products $E\times F$, that is, $\Lambda\supseteq\Sigma\widehat{\otimes}\mathrm{T}$.

(c) Take $E\in\Sigma$, $F\in\mathrm{T}$. We know that $E\times F\in\Lambda$; setting $E_0=E$, $F_0=F$, $E_n=F_n=\emptyset$ for $n\ge1$ in the definition of θ, we have

$$\lambda_0(E\times F)=\theta(E\times F)\le\mu E\cdot\nu F.$$

We have come to the central idea of the construction. In fact $\theta(E\times F)=\mu E\cdot\nu F$. **P** Suppose that $E\times F\subseteq\bigcup_{n\in\mathbb{N}}E_n\times F_n$ where $E_n\in\Sigma$ and $F_n\in\mathrm{T}$ for every n. Set $u=\sum_{n=0}^{\infty}\mu E_n\cdot\nu F_n$. If $u=\infty$ or $\mu E=0$ or $\nu F=0$ then of course $\mu E\cdot\nu F\le u$. Otherwise, set

$$I=\{n:n\in\mathbb{N},\ \mu E_n=0\},\quad J=\{n:n\in\mathbb{N},\ \nu F_n=0\},\quad K=\mathbb{N}\setminus(I\cup J),$$

$$E'=E\setminus\bigcup_{n\in I}E_n,\quad F'=F\setminus\bigcup_{n\in J}F_n.$$

Then $\mu E'=\mu E$ and $\nu F'=\nu F$; $E'\times F'\subseteq\bigcup_{n\in K}E_n\times F_n$; and for $n\in K$, $\mu E_n<\infty$ and $\nu F_n<\infty$, since $\mu E_n\cdot\nu F_n\le u<\infty$ and neither μE_n nor νF_n is zero. Set

$$f_n = \nu F_n \chi E_n : X \to \mathbb{R}$$

if $n \in K$, and $f_n = \mathbf{0} : X \to \mathbb{R}$ if $n \in I \cup J$. Then f_n is a simple function and $\int f_n = \nu F_n \mu E_n$ for $n \in K$, 0 otherwise, so

$$\sum_{n=0}^{\infty} \int f_n(x)\mu(dx) = \sum_{n=0}^{\infty} \mu E_n \cdot \nu F_n \le u.$$

By B.Levi's theorem (123A), applied to $\langle \sum_{k=0}^{n} f_k \rangle_{n\in\mathbb{N}}$, $g = \sum_{n=0}^{\infty} f_n$ is integrable and $\int g\,d\mu \le u$. Write E'' for $\{x : x \in E',\, g(x) < \infty\}$, so that $\mu E'' = \mu E' = \mu E$. Now take any $x \in E''$ and set $K_x = \{n : n \in K,\, x \in E_n\}$. Because $E' \times F' \subseteq \bigcup_{n\in K} E_n \times F_n$, $F' \subseteq \bigcup_{n\in K_x} F_n$ and

$$\nu F = \nu F' \le \sum_{n\in K_x} \nu F_n = \sum_{n=0}^{\infty} f_n(x) = g(x).$$

Thus $g(x) \ge \nu F$ for every $x \in E''$. We are supposing that $0 < \mu E = \mu E''$ and $0 < \nu F$, so we must have $\nu F < \infty$, $\mu E'' < \infty$. Now $g \ge \nu F \chi E''$, so

$$\mu E \cdot \nu F = \mu E'' \cdot \nu F = \int \nu F \chi E'' \le \int g \le u = \sum_{n=0}^{\infty} \mu E_n \cdot \nu F_n.$$

As $\langle E_n \rangle_{n\in\mathbb{N}}$, $\langle F_n \rangle_{n\in\mathbb{N}}$ are arbitrary, $\theta(E \times F) \ge \mu E \cdot \nu F$ and $\theta(E \times F) = \mu E \cdot \nu F$. **Q**

Thus

$$\lambda_0(E \times F) = \theta(E \times F) = \mu E \cdot \nu F$$

for all $E \in \Sigma$, $F \in \mathrm{T}$.

251F Definition Let (X, Σ, μ) and (Y, T, ν) be measure spaces, and λ_0 the primitive product measure defined in 251C. By the **c.l.d. product measure** on $X \times Y$ I shall mean the function $\lambda : \operatorname{dom}\lambda_0 \to [0, \infty]$ defined by setting

$$\lambda W = \sup\{\lambda_0(W \cap (E \times F)) : E \in \Sigma,\, F \in \mathrm{T},\, \mu E < \infty,\, \nu F < \infty\}$$

for $W \in \operatorname{dom}\lambda_0$.

251G Remark I had better show at once that λ is a measure. **P** Of course its domain $\Lambda = \operatorname{dom}\lambda_0$ is a σ-algebra, and $\lambda\emptyset = \lambda_0\emptyset = 0$. If $\langle W_n \rangle_{n\in\mathbb{N}}$ is a disjoint sequence in Λ, then for any $E \in \Sigma$, $F \in \mathrm{T}$ of finite measure

$$\lambda_0(\bigcup_{n\in\mathbb{N}} W_n \cap (E \times F)) = \sum_{n=0}^{\infty} \lambda_0(W_n \cap (E \times F)) \le \sum_{n=0}^{\infty} \lambda W_n,$$

so $\lambda(\bigcup_{n\in\mathbb{N}} W_n) \le \sum_{n=0}^{\infty} \lambda W_n$. On the other hand, if $a < \sum_{n=0}^{\infty} \lambda W_n$, then we can find $m \in \mathbb{N}$ and $a_0, \dots, a_m$ such that $a \le \sum_{n=0}^{m} a_n$ and $a_n < \lambda W_n$ for each $n \le m$; now there are $E_0, \dots, E_m \in \Sigma$ and $F_0, \dots, F_m \in \mathrm{T}$, all of finite measure, such that $a_n \le \lambda_0(W_n \cap (E_n \times F_n))$ for each n. Setting $E = \bigcup_{n\le m} E_n$ and $F = \bigcup_{n\le m} F_n$, we have $\mu E < \infty$ and $\nu F < \infty$, so

$$\begin{aligned}\lambda(\bigcup_{n\in\mathbb{N}} W_n) &\ge \lambda_0(\bigcup_{n\in\mathbb{N}} W_n \cap (E \times F)) = \sum_{n=0}^{\infty} \lambda_0(W_n \cap (E \times F)) \\ &\ge \sum_{n=0}^{m} \lambda_0(W_n \cap (E_n \times F_n)) \ge \sum_{n=0}^{m} a_n \ge a.\end{aligned}$$

As a is arbitrary, $\lambda(\bigcup_{n\in\mathbb{N}} W_n) \ge \sum_{n=0}^{\infty} \lambda W_n$ and $\lambda(\bigcup_{n\in\mathbb{N}} W_n) = \sum_{n=0}^{\infty} \lambda W_n$. As $\langle W_n \rangle_{n\in\mathbb{N}}$ is arbitrary, λ is a measure. **Q**

251H We need a simple property of the measure λ_0.

Lemma Let (X, Σ, μ) and (Y, T, ν) be two measure spaces; let λ_0 be the primitive product measure on $X \times Y$, and Λ its domain. If $H \subseteq X \times Y$ and $H \cap (E \times F) \in \Lambda$ whenever $\mu E < \infty$ and $\nu F < \infty$, then $H \in \Lambda$.

proof Let θ be the outer measure described in 251A. Suppose that $A \subseteq X \times Y$ and $\theta A < \infty$. Let $\epsilon > 0$. Let $\langle E_n \rangle_{n\in\mathbb{N}}$, $\langle F_n \rangle_{n\in\mathbb{N}}$ be sequences in Σ, T respectively such that $A \subseteq \bigcup_{n\in\mathbb{N}} E_n \times F_n$ and $\sum_{n=0}^{\infty} \mu E_n \cdot \nu F_n \le \theta A + \epsilon$. Now, for each n, the product of the measures μE_n, νE_n is finite, so either one is zero or both are finite. If $\mu E_n = 0$ or $\nu F_n = 0$ then of course

$$\mu E_n \cdot \nu F_n = 0 = \theta((E_n \times F_n) \cap H) + \theta((E_n \times F_n) \setminus H).$$

If $\mu E_n < \infty$ and $\nu F_n < \infty$ then

$$\begin{aligned}\mu E_n\cdot\nu F_n &= \lambda_0(E_n\times F_n)\\ &= \lambda_0((E_n\times F_n)\cap H)+\lambda_0((E_n\times F_n)\setminus H)\\ &= \theta((E_n\times F_n)\cap H)+\theta((E_n\times F_n)\setminus H).\end{aligned}$$

Accordingly, because θ is an outer measure,

$$\begin{aligned}\theta(A\cap H)+\theta(A\setminus H) &\le \sum_{n=0}^{\infty}\theta((E_n\times F_n)\cap H)+\sum_{n=0}^{\infty}\theta((E_n\times F_n)\setminus H)\\ &= \sum_{n=0}^{\infty}\mu E_n\cdot\nu F_n\le\theta A+\epsilon.\end{aligned}$$

As ϵ is arbitrary, $\theta(A\cap H)+\theta(A\setminus H)\le\theta A$. As A is arbitrary, $H\in\Lambda$.

251I Now for the fundamental properties of the c.l.d. product measure.

Theorem Let (X,Σ,μ) and (Y,T,ν) be measure spaces; let λ be the c.l.d. product measure on $X\times Y$, and Λ its domain. Then

(a) $\Sigma\widehat{\otimes}\mathrm{T}\subseteq\Lambda$ and $\lambda(E\times F)=\mu E\cdot\nu F$ whenever $E\in\Sigma$, $F\in\mathrm{T}$ and $\mu E\cdot\nu F<\infty$;

(b) for every $W\in\Lambda$ there is a $V\in\Sigma\widehat{\otimes}\mathrm{T}$ such that $V\subseteq W$ and $\lambda V=\lambda W$;

(c) $(X\times Y,\Lambda,\lambda)$ is complete and locally determined, and in fact is the c.l.d. version of $(X\times Y,\Lambda,\lambda_0)$ as described in 213D-213E; in particular, $\lambda W=\lambda_0 W$ whenever $\lambda_0 W<\infty$;

(d) if $W\in\Lambda$ and $\lambda W>0$ then there are $E\in\Sigma$, $F\in\mathrm{T}$ such that $\mu E<\infty$, $\nu F<\infty$ and $\lambda(W\cap(E\times F))>0$;

(e) if $W\in\Lambda$ and $\lambda W<\infty$, then for every $\epsilon>0$ there are $E_0,\dots,E_n\in\Sigma$, $F_0,\dots,F_n\in\mathrm{T}$, all of finite measure, such that $\lambda(W\triangle\bigcup_{i\le n}(E_i\times F_i))\le\epsilon$.

proof Take θ to be the outer measure of 251A and λ_0 the primitive product measure of 251C. Set $\Sigma^f=\{E:E\in\Sigma,\ \mu E<\infty\}$ and $\mathrm{T}^f=\{F:F\in\mathrm{T},\ \nu F<\infty\}$.

(a) By 251E, $\Sigma\widehat{\otimes}\mathrm{T}\subseteq\Lambda$. If $E\in\Sigma$ and $F\in\mathrm{T}$ and $\mu E\cdot\nu F<\infty$, either $\mu E\cdot\nu F=0$ and $\lambda(E\times F)=\lambda_0(E\times F)=0$ or both μE and νF are finite and again $\lambda(E\times F)=\lambda_0(E\times F)=\mu E\cdot\nu F$.

(b)(i) Take any $a<\lambda W$. Then there are $E\in\Sigma^f$, $F\in\mathrm{T}^f$ such that $\lambda_0(W\cap(E\times F))>a$ (251F); now

$$\begin{aligned}\theta((E\times F)\setminus W) &= \lambda_0((E\times F)\setminus W)\\ &= \lambda_0(E\times F)-\lambda_0(W\cap(E\times F))<\lambda_0(E\times F)-a.\end{aligned}$$

Let $\langle E_n\rangle_{n\in\mathbb{N}}$, $\langle F_n\rangle_{n\in\mathbb{N}}$ be sequences in Σ, T respectively such that $(E\times F)\setminus W\subseteq\bigcup_{n\in\mathbb{N}}E_n\times F_n$ and $\sum_{n=0}^{\infty}\mu E_n\cdot\nu F_n\le\lambda_0(E\times F)-a$. Consider

$$V=(E\times F)\setminus\bigcup_{n\in\mathbb{N}}E_n\times F_n\in\Sigma\widehat{\otimes}\mathrm{T};$$

then $V\subseteq W$, and

$$\begin{aligned}\lambda V=\lambda_0 V &= \lambda_0(E\times F)-\lambda_0((E\times F)\setminus V)\\ &\ge \lambda_0(E\times F)-\lambda_0(\bigcup_{n\in\mathbb{N}}E_n\times F_n)\end{aligned}$$

(because $(E\times F)\setminus V\subseteq\bigcup_{n\in\mathbb{N}}E_n\times F_n$)

$$\ge\lambda_0(E\times F)-\sum_{n=0}^{\infty}\mu E_n\cdot\nu F_n\ge a$$

(by the choice of the E_n, F_n).

(ii) Thus for every $a<\lambda W$ there is a $V\in\Sigma\widehat{\otimes}\mathrm{T}$ such that $V\subseteq W$ and $\lambda V\ge a$. Now choose a sequence $\langle a_n\rangle_{n\in\mathbb{N}}$ strictly increasing to λW, and for each a_n a corresponding V_n; then $V=\bigcup_{n\in\mathbb{N}}V_n$ belongs to the σ-algebra $\Sigma\widehat{\otimes}\mathrm{T}$, is included in W, and has measure at least $\sup_{n\in\mathbb{N}}\lambda V_n$ and at most λW; so $\lambda V=\lambda W$, as required.

(c)(i) If $H \subseteq X \times Y$ is λ-negligible, there is a $W \in \Lambda$ such that $H \subseteq W$ and $\lambda W = 0$. If $E \in \Sigma$, $F \in \mathrm{T}$ are of finite measure, $\lambda_0(W \cap (E \times F)) = 0$; but λ_0, being derived from the outer measure θ by Carathéodory's method, is complete (212A), so $H \cap (E \times F) \in \Lambda$ and $\lambda_0(H \cap (E \times F)) = 0$. Because E and F are arbitrary, $H \in \Lambda$, by 251H. As H is arbitrary, λ is complete.

(ii) If $W \in \Lambda$ and $\lambda W = \infty$, then there must be $E \in \Sigma$, $F \in \mathrm{T}$ such that $\mu E < \infty$, $\nu F < \infty$ and $\lambda_0(W \cap (E \times F)) > 0$; now

$$0 < \lambda(W \cap (E \times F)) \le \mu E \cdot \nu F < \infty.$$

Thus λ is semi-finite.

(iii) If $H \subseteq X \times Y$ and $H \cap W \in \Lambda$ whenever $\lambda W < \infty$, then, in particular, $H \cap (E \times F) \in \Lambda$ whenever $\mu E < \infty$ and $\nu F < \infty$; by 251H again, $H \in \Lambda$. Thus λ is locally determined.

(iv) If $W \in \Lambda$ and $\lambda_0 W < \infty$, then we have sequences $\langle E_n \rangle_{n\in\mathbb{N}}$ in Σ, $\langle F_n \rangle_{n\in\mathbb{N}}$ in T such that $W \subseteq \bigcup_{n\in\mathbb{N}}(E_n \times F_n)$ and $\sum_{n=0}^{\infty} \mu E_n \cdot \nu F_n < \infty$. Set

$$I = \{n : \mu E_n = \infty\}, \quad J = \{n : \nu F_n = \infty\}, \quad K = \mathbb{N} \setminus (I \cup J);$$

then $\nu(\bigcup_{n\in I} F_n) = \mu(\bigcup_{n\in J} E_n) = 0$, so $\lambda_0(W \setminus W') = 0$, where

$$W' = W \cap \bigcup_{n\in K}(E_n \times F_n) \supseteq W \setminus ((\bigcup_{n\in J} E_n \times Y) \cup (X \times \bigcup_{n\in I} F_n)).$$

Now set $E'_n = \bigcup_{i\in K, i\le n} E_i$, $F'_n = \bigcup_{i\in K, i\le n} F_i$ for each n. We have $W' = \bigcup_{n\in\mathbb{N}} W' \cap (E'_n \times F'_n)$, so

$$\lambda W \le \lambda_0 W = \lambda_0 W' = \lim_{n\to\infty} \lambda_0(W' \cap (E'_n \times F'_n)) \le \lambda W' \le \lambda W,$$

and $\lambda W = \lambda_0 W$.

(v) Following the terminology of 213D, let us write

$$\tilde{\Lambda} = \{W : W \subseteq X \times Y,\ W \cap V \in \Lambda \text{ whenever } V \in \Lambda \text{ and } \lambda_0 V < \infty\},$$

$$\tilde{\lambda} W = \sup\{\lambda_0(W \cap V) : V \in \Lambda,\ \lambda_0 V < \infty\}.$$

Because $\lambda_0(E \times F) < \infty$ whenever $\mu E < \infty$ and $\nu F < \infty$, $\tilde{\Lambda} \subseteq \Lambda$ and $\tilde{\Lambda} = \Lambda$.

Now for any $W \in \Lambda$ we have

$$\begin{aligned}
\tilde{\lambda} W &= \sup\{\lambda_0(W \cap V) : V \in \Lambda,\ \lambda_0 V < \infty\} \\
&\ge \sup\{\lambda_0(W \cap (E \times F)) : E \in \Sigma^f,\ F \in \mathrm{T}^f\} \\
&= \lambda W \\
&\ge \sup\{\lambda(W \cap V) : V \in \Lambda,\ \lambda_0 V < \infty\} \\
&= \sup\{\lambda_0(W \cap V) : V \in \Lambda,\ \lambda_0 V < \infty\},
\end{aligned}$$

using (iv) just above, so that $\lambda = \tilde{\lambda}$ is the c.l.d. version of λ_0.

(d) If $W \in \Lambda$ and $\lambda W > 0$, there are $E \in \Sigma^f$ and $F \in \mathrm{T}^f$ such that $\lambda(W \cap (E \times F)) = \lambda_0(W \cap (E \times F)) > 0$.

(e) There are $E \in \Sigma^f$, $F \in \mathrm{T}^f$ such that $\lambda_0(W \cap (E \times F)) \ge \lambda W - \frac{1}{3}\epsilon$; set $V_1 = W \cap (E \times F)$; then

$$\lambda(W \setminus V_1) = \lambda W - \lambda V_1 = \lambda W - \lambda_0 V_1 \le \frac{1}{3}\epsilon.$$

There are sequences $\langle E'_n \rangle_{n\in\mathbb{N}}$ in Σ, $\langle F'_n \rangle_{n\in\mathbb{N}}$ in T such that $V_1 \subseteq \bigcup_{n\in\mathbb{N}} E'_n \times F'_n$ and $\sum_{n=0}^{\infty} \mu E'_n \cdot \nu F'_n \le \lambda_0 V_1 + \frac{1}{3}\epsilon$. Replacing E'_n, F'_n by $E'_n \cap E$, $F'_n \cap F$ if necessary, we may suppose that $E'_n \in \Sigma^f$ and $F'_n \in \mathrm{T}^f$ for every n. Set $V_2 = \bigcup_{n\in\mathbb{N}} E'_n \times F'_n$; then

$$\lambda(V_2 \setminus V_1) \le \lambda_0(V_2 \setminus V_1) \le \sum_{n=0}^{\infty} \mu E'_n \cdot \nu F'_n - \lambda_0 V_1 \le \frac{1}{3}\epsilon.$$

Let $m \in \mathbb{N}$ be such that $\sum_{n=m+1}^{\infty} \mu E'_n \cdot \nu F'_n \le \frac{1}{3}\epsilon$, and set

$$V = \bigcup_{n=0}^{m} E'_n \times F'_n.$$

Then

$$\lambda(V_2 \setminus V) \le \sum_{n=m+1}^{\infty} \mu E'_n \cdot \nu F'_n \le \frac{1}{3}\epsilon.$$

Putting these together, we have $W \triangle V \subseteq (W \setminus V_1) \cup (V_2 \setminus V_1) \cup (V_2 \setminus V)$, so

$$\lambda(W \triangle V) \le \lambda(W \setminus V_1) + \lambda(V_2 \setminus V_1) + \lambda(V_2 \setminus V) \le \frac{1}{3}\epsilon + \frac{1}{3}\epsilon + \frac{1}{3}\epsilon = \epsilon.$$

And V is of the required form.

251J Proposition If (X, Σ, μ) and (Y, T, ν) are semi-finite measure spaces and λ is the c.l.d. product measure on $X \times Y$, then $\lambda(E \times F) = \mu E \cdot \nu F$ for all $E \in \Sigma$, $F \in \mathrm{T}$.

proof Setting $\Sigma^f = \{E : E \in \Sigma,\, \mu E < \infty\}$, $\mathrm{T}^f = \{F : F \in \mathrm{T},\, \nu F < \infty\}$, we have

$$\begin{aligned}\lambda(E \times F) &= \sup\{\lambda_0((E \cap E_0) \times (F \cap F_0)) : E_0 \in \Sigma^f,\, F_0 \in \mathrm{T}^f\}\\ &= \sup\{\mu(E \cap E_0) \cdot \nu(F \cap F_0)) : E_0 \in \Sigma^f,\, F_0 \in \mathrm{T}^f\}\\ &= \sup\{\mu(E \cap E_0) : E_0 \in \Sigma^f\} \cdot \sup\{\nu(F \cap F_0) : F_0 \in \mathrm{T}^f\} = \mu E \cdot \nu F\end{aligned}$$

(using 213A).

251K σ-finite spaces Of course most of the measure spaces we shall apply these results to are σ-finite, and in this case there are some useful simplifications.

Proposition Let (X, Σ, μ) and (Y, T, ν) be σ-finite measure spaces. Then the c.l.d. product measure on $X \times Y$ is equal to the primitive product measure, and is the completion of its restriction to $\Sigma \widehat{\otimes} \mathrm{T}$; moreover, this common product measure is σ-finite.

proof Write λ_0, λ for the primitive and c.l.d. product measures, as usual, and Λ for their domain. Let $\langle E_n \rangle_{n \in \mathbb{N}}$, $\langle F_n \rangle_{n \in \mathbb{N}}$ be non-decreasing sequences of sets of finite measure covering X, Y respectively (see 211D).

(a) For each $n \in \mathbb{N}$, $\lambda(E_n \times F_n) = \mu E_n \cdot \nu F_n$ is finite, by 251Ia. Since $X \times Y = \bigcup_{n \in \mathbb{N}} E_n \times F_n$, λ is σ-finite.

(b) For any $W \in \Lambda$,

$$\lambda_0 W = \lim_{n \to \infty} \lambda_0(W \cap (E_n \times F_n)) = \lim_{n \to \infty} \lambda(W \cap (E_n \times F_n)) = \lambda W.$$

So $\lambda = \lambda_0$.

(c) Write $\lambda_{\mathcal{B}}$ for the restriction of $\lambda = \lambda_0$ to $\Sigma \widehat{\otimes} \mathrm{T}$, and $\hat{\lambda}_{\mathcal{B}}$ for its completion.

(i) Suppose that $W \in \operatorname{dom} \hat{\lambda}_{\mathcal{B}}$. Then there are W', $W'' \in \Sigma \widehat{\otimes} \mathrm{T}$ such that $W' \subseteq W \subseteq W''$ and $\lambda_{\mathcal{B}}(W'' \setminus W') = 0$ (212C). In this case, $\lambda(W'' \setminus W') = 0$; as λ is complete, $W \in \Lambda$ and

$$\lambda W = \lambda W' = \lambda_{\mathcal{B}} W' = \hat{\lambda}_{\mathcal{B}} W.$$

Thus λ extends $\hat{\lambda}_{\mathcal{B}}$.

(ii) If $W \in \Lambda$, then there is a $V \in \Sigma \widehat{\otimes} \mathrm{T}$ such that $V \subseteq W$ and $\lambda(W \setminus V) = 0$. **P** For each $n \in \mathbb{N}$ there is a $V_n \in \Sigma \widehat{\otimes} \mathrm{T}$ such that $V_n \subseteq W \cap (E_n \times F_n)$ and $\lambda V_n = \lambda(W \cap (E_n \times F_n))$ (251Ib). But as $\lambda(E_n \times F_n) = \mu E_n \cdot \nu F_n$ is finite, this means that $\lambda(W \cap (E_n \times F_n) \setminus V_n) = 0$. So if we set $V = \bigcup_{n \in \mathbb{N}} V_n$, we shall have $V \in \Sigma \widehat{\otimes} \mathrm{T}$, $V \subseteq W$ and

$$W \setminus V = \bigcup_{n \in \mathbb{N}} W \cap (E_n \times F_n) \setminus V \subseteq \bigcup_{n \in \mathbb{N}} W \cap (E_n \times F_n) \setminus V_n$$

is λ-negligible. **Q**

Similarly, there is a $V' \in \Sigma \widehat{\otimes} \mathrm{T}$ such that $V' \subseteq (X \times Y) \setminus W$ and $\lambda(((X \times Y) \setminus W) \setminus V') = 0$. Setting $V'' = (X \times Y) \setminus V'$, $V'' \in \Sigma \widehat{\otimes} \mathrm{T}$, $W \subseteq V''$ and $\lambda(V'' \setminus W) = 0$. So

$$\lambda_{\mathcal{B}}(V'' \setminus V) = \lambda(V'' \setminus V) = \lambda(V'' \setminus W) + \lambda(W \setminus V) = 0,$$

and W is measured by $\hat{\lambda}_{\mathcal{B}}$, with $\hat{\lambda}_{\mathcal{B}} W = \lambda_{\mathcal{B}} V = \lambda W$. As W is arbitrary, $\hat{\lambda}_{\mathcal{B}} = \lambda$.

***251L** The following result fits in naturally here; I star it because it will be used seldom (there is a more important version of the same idea in 254G) and the proof can be skipped until you come to need it.

Proposition Let (X_1, Σ_1, μ_1), (X_2, Σ_2, μ_2), $(Y_1, \mathrm{T}_1, \nu_1)$ and $(Y_2, \mathrm{T}_2, \nu_2)$ be σ-finite measure spaces; let λ_1, λ_2 be the product measures on $X_1 \times Y_1$, $X_2 \times Y_2$ respectively. Suppose that $f : X_1 \to X_2$ and $g : Y_1 \to Y_2$ are inverse-measure-preserving functions, and that $h(x, y) = (f(x), g(y))$ for $x \in X_1$, $y \in Y_1$. Then h is inverse-measure-preserving.

proof Write Λ_1, Λ_2 for the domains of λ_1, λ_2 respectively.

(a) Suppose that $E \in \Sigma_2$ and $F \in \mathrm{T}_2$ have finite measure. Then $\lambda_1 h^{-1}[W \cap (E \times F)]$ is defined and equal to $\lambda_2(W \cap (E \times F))$ for every $W \in \Lambda_2$. **P**

$$\begin{aligned}\lambda_1 h^{-1}[E \times F] &= \lambda_1(f^{-1}[E] \times g^{-1}[F]) = \mu_1 f^{-1}[E] \cdot \nu_1 g^{-1}[F] \\ &= \mu_2 E \cdot \nu_2 F = \lambda_2(E \times F)\end{aligned}$$

by 251E/251J. **Q**

(b) Take $E_0 \in \Sigma_2$ and $F_0 \in \mathrm{T}_2$ of finite measure. Let $\tilde{\lambda}_1$, $\tilde{\lambda}_2$ be the subspace measures on $f^{-1}[E_0] \times g^{-1}[F_0]$ and $E_0 \times F_0$ respectively. Set $\tilde{h} = h \restriction f^{-1}[E_0] \times g^{-1}[F_0]$, and write ι for the identity map from $E_0 \times F_0$ to $X_2 \times Y_2$; let $\lambda = \tilde{\lambda}_1 \tilde{h}^{-1}$ and $\lambda' = \tilde{\lambda}_2 \iota^{-1}$ be the image measures on $X_2 \times Y_2$. Then (a) tells us that

$$\begin{aligned}\lambda(E \times F) &= \lambda_1(h^{-1}[(E \cap E_0) \times (F \cap F_0)]) \\ &= \lambda_2((E \cap E_0) \times (F \cap F_0)) = \lambda'(E \times F)\end{aligned}$$

whenever $E \in \Sigma_2$ and $F \in \mathrm{T}_2$. By the Monotone Class Theorem (136C), λ and λ' agree on $\Sigma_2 \widehat{\otimes} \mathrm{T}_2$, that is, $\lambda_1(h^{-1}[W \cap (E_0 \times F_0)]) = \lambda_2(W \cap (E_0 \times F_0))$ for every $W \in \Sigma_2 \widehat{\otimes} \mathrm{T}_2$.

If W is any member of Λ_2, there are W', $W'' \in \Sigma_2 \widehat{\otimes} \mathrm{T}_2$ such that $W' \subseteq W \subseteq W''$ and $\lambda_2(W'' \setminus W') = 0$ (251K). Now we must have

$$h^{-1}[W' \cap (E_0 \times F_0)] \subseteq h^{-1}[W \cap (E_0 \times F_0)] \subseteq h^{-1}[W'' \cap (E_0 \times F_0)],$$

$$\lambda_1(h^{-1}[W'' \cap (E_0 \times F_0)] \setminus h^{-1}[W' \cap (E_0 \times F_0)]) = \lambda_2((W'' \setminus W') \cap (E_0 \times F_0)) = 0;$$

because λ_1 is complete, $\lambda_1 h^{-1}[W \cap (E_0 \times F_0)]$ is defined and equal to

$$\lambda_1 h^{-1}[W' \cap (E_0 \times F_0)] = \lambda_2(W' \cap (E_0 \times F_0)) = \lambda_2(W \cap (E_0 \times F_0)).$$

(c) Now suppose that $\langle E_n \rangle_{n \in \mathbb{N}}$, $\langle F_n \rangle_{n \in \mathbb{N}}$ are non-decreasing sequences of sets of finite measure with union X_2, Y_2 respectively. If $W \in \Lambda_2$,

$$\lambda_1 h^{-1}[W] = \sup_{n \in \mathbb{N}} \lambda_1 h^{-1}[W \cap (E_n \times F_n)] = \sup_{n \in \mathbb{N}} \lambda_2(W \cap (E_n \times F_n)) = \lambda_2 W.$$

So h is inverse-measure-preserving, as claimed.

251M It is time that I gave some examples. Of course the central example is Lebesgue measure. In this case we have the only reasonable result. I pause to describe the leading example of the product $\Sigma \widehat{\otimes} \mathrm{T}$ introduced in 251D.

Proposition Let r, $s \geq 1$ be integers. Then we have a natural bijection $\phi : \mathbb{R}^r \times \mathbb{R}^s \to \mathbb{R}^{r+s}$, defined by setting

$$\phi((\xi_1, \ldots, \xi_r), (\eta_1, \ldots, \eta_s)) = (\xi_1, \ldots, \xi_r, \eta_1, \ldots, \eta_s)$$

for $\xi_1, \ldots, \xi_r, \eta_1, \ldots, \eta_s \in \mathbb{R}$. If we write $\mathcal{B}_r$, $\mathcal{B}_s$ and $\mathcal{B}_{r+s}$ for the Borel σ-algebras of $\mathbb{R}^r$, $\mathbb{R}^s$ and $\mathbb{R}^{r+s}$ respectively, then ϕ identifies $\mathcal{B}_{r+s}$ with $\mathcal{B}_r \widehat{\otimes} \mathcal{B}_s$.

proof (a) Write $\mathcal{B}$ for the σ-algebra $\{\phi^{-1}[W] : W \in \mathcal{B}_{r+s}\}$ copied onto $\mathbb{R}^r \times \mathbb{R}^s$ by the bijection ϕ; we are seeking to prove that $\mathcal{B} = \mathcal{B}_r \widehat{\otimes} \mathcal{B}_s$. We have maps $\pi_1 : \mathbb{R}^{r+s} \to \mathbb{R}^r$, $\pi_2 : \mathbb{R}^{r+s} \to \mathbb{R}^s$ defined by setting $\pi_1(\phi(x, y)) = x$, $\pi_2(\phi(x, y)) = y$. Each co-ordinate of π_1 is continuous, therefore Borel measurable (121Db), so $\pi_1^{-1}[E] \in \mathcal{B}_{r+s}$ for every $E \in \mathcal{B}_r$, by 121K. Similarly, $\pi_2^{-1}[F] \in \mathcal{B}_{r+s}$ for every $F \in \mathcal{B}_s$. So $\phi[E \times F] = \pi_1^{-1}[E] \cap \pi_1^{-1}[F]$ belongs to $\mathcal{B}_{r+s}$, that is, $E \times F \in \mathcal{B}$, whenever $E \in \mathcal{B}_r$ and $F \in \mathcal{B}_s$. Because $\mathcal{B}$ is a σ-algebra, $\mathcal{B}_r \widehat{\otimes} \mathcal{B}_s \subseteq \mathcal{B}$.

(b) Now examine sets of the form

$$\{(x, y) : \xi_i \leq \alpha\} = \{x : \xi_i \leq \alpha\} \times \mathbb{R}^s,$$

$$\{(x, y) : \eta_j \leq \alpha\} = \mathbb{R}^r \times \{y : \eta_j \leq \alpha\}$$

for $\alpha \in \mathbb{R}$, $i \leq r$ and $j \leq s$, taking $x = (\xi_1, \ldots, \xi_r)$ and $y = (\eta_1, \ldots, \eta_s)$. All of these belong to $\mathcal{B}_r \widehat{\otimes} \mathcal{B}_s$. But the σ-algebra they generate is just $\mathcal{B}$, by 121J. So $\mathcal{B} \subseteq \mathcal{B}_r \widehat{\otimes} \mathcal{B}_s$ and $\mathcal{B} = \mathcal{B}_r \widehat{\otimes} \mathcal{B}_s$.

251N Theorem Let r, $s \geq 1$ be integers. Then the bijection $\phi : \mathbb{R}^r \times \mathbb{R}^s \to \mathbb{R}^{r+s}$ described in 251M identifies Lebesgue measure on $\mathbb{R}^{r+s}$ with the c.l.d. product λ of Lebesgue measure on $\mathbb{R}^r$ and Lebesgue measure on $\mathbb{R}^s$.

proof Write μ_r, μ_s, μ_{r+s} for the three versions of Lebesgue measure, μ_r^*, μ_s^* and μ_{r+s}^* for the corresponding outer measures, and θ for the outer measure on $\mathbb{R}^r \times \mathbb{R}^s$ derived from μ_r and μ_s by the formula of 251A.

(a) If $I \subseteq \mathbb{R}^r$ and $J \subseteq \mathbb{R}^s$ are half-open intervals, then $\phi[I \times J] \subseteq \mathbb{R}^{r+s}$ is also a half-open interval, and

$$\mu_{r+s}(\phi[I \times J]) = \mu_r I \cdot \mu_s J;$$

this is immediate from the definition of the Lebesgue measure of an interval. (I speak of 'half-open' intervals here, that is, intervals of the form $\prod_{1 \le j \le r} [\alpha_j, \beta_j[$, because I used them in the definition of Lebesgue measure in §115. If you prefer to work with open intervals or closed intervals it makes no difference.) Note also that every half-open interval in $\mathbb{R}^{r+s}$ is expressible as $\phi[I \times J]$ for suitable I, J.

(b) For any $A \subseteq \mathbb{R}^{r+s}$, $\theta(\phi^{-1}[A]) \le \mu_{r+s}^*(A)$. **P** For any $\epsilon > 0$, there is a sequence $\langle K_n \rangle_{n\in\mathbb{N}}$ of half-open intervals in $\mathbb{R}^{r+s}$ such that $A \subseteq \bigcup_{n\in\mathbb{N}} K_n$ and $\sum_{n=0}^{\infty} \mu_{r+s}(K_n) \le \mu_{r+s}^*(A) + \epsilon$. Express each K_n as $\phi[I_n \times J_n]$, where I_n and J_n are half-open intervals in $\mathbb{R}^r$ and $\mathbb{R}^s$ respectively; then $\phi^{-1}[A] \subseteq \bigcup_{n\in\mathbb{N}} I_n \times J_n$, so that

$$\theta(\phi^{-1}[A]) \le \sum_{n=0}^{\infty} \mu_r I_n \cdot \mu_s J_n = \sum_{n=0}^{\infty} \mu_{r+s}(K_n) \le \mu_{r+s}^*(A) + \epsilon.$$

As ϵ is arbitrary, we have the result. **Q**

(c) If $E \subseteq \mathbb{R}^r$ and $F \subseteq \mathbb{R}^s$ are measurable, then $\mu_{r+s}^*(\phi[E \times F]) \le \mu_r E \cdot \mu_s F$.

P (i) Consider first the case $\mu_r E < \infty$, $\mu_s F < \infty$. In this case, given $\epsilon > 0$, there are sequences $\langle I_n \rangle_{n\in\mathbb{N}}$, $\langle J_n \rangle_{n\in\mathbb{N}}$ of half-open intervals such that $E \subseteq \bigcup_{n\in\mathbb{N}} I_n$, $F \subseteq \bigcup_{n\in\mathbb{N}} F_n$,

$$\sum_{n=0}^{\infty} \mu_r I_n \le \mu_r^* E + \epsilon = \mu_r E + \epsilon,$$

$$\sum_{n=0}^{\infty} \mu_s J_n \le \mu_s^* F + \epsilon = \mu_s F + \epsilon.$$

Accordingly $E \times F \subseteq \bigcup_{m,n\in\mathbb{N}} I_m \times J_n$ and $\phi[E \times F] \subseteq \bigcup_{m,n\in\mathbb{N}} \phi[I_m \times J_n]$, so that

$$\begin{aligned}\mu_{r+s}^*(\phi[E \times F]) &\le \sum_{m,n=0}^{\infty} \mu_{r+s}(\phi[I_m \times J_n]) = \sum_{m,n=0}^{\infty} \mu_r I_m \cdot \mu_s J_n \\ &= \sum_{m=0}^{\infty} \mu_r I_m \cdot \sum_{n=0}^{\infty} \mu_s J_n \le (\mu_r E + \epsilon)(\mu_s F + \epsilon).\end{aligned}$$

As ϵ is arbitrary, we have the result.

(ii) Next, if $\mu_r E = 0$, there is a sequence $\langle F_n \rangle_{n\in\mathbb{N}}$ of sets of finite measure covering $\mathbb{R}^s \supseteq F$, so that

$$\mu_{r+s}^*(\phi[E \times F]) \le \sum_{n=0}^{\infty} \mu_{r+s}^*(\phi[E \times F_n]) \le \sum_{n=0}^{\infty} \mu_r E \cdot \mu_s F_n = 0 = \mu_r E \cdot \mu_s F.$$

Similarly, $\mu_{r+s}^*(\phi[E \times F]) \le \mu_r E \cdot \mu_s F$ if $\mu_s F = 0$. The only remaining case is that in which both of $\mu_r E$, $\mu_s F$ are strictly positive and one is infinite; but in this case $\mu_r E \cdot \mu_s F = \infty$, so surely $\mu_{r+s}^*(\phi[E \times F]) \le \mu_r E \cdot \mu_s F$. **Q**

(d) If $A \subseteq \mathbb{R}^{r+s}$, then $\mu_{r+s}^*(A) \le \theta(\phi^{-1}[A])$. **P** Given $\epsilon > 0$, there are sequences $\langle E_n \rangle_{n\in\mathbb{N}}$, $\langle F_n \rangle_{n\in\mathbb{N}}$ of measurable sets in $\mathbb{R}^r$, $\mathbb{R}^s$ respectively such that $\phi^{-1}[A] \subseteq \bigcup_{n\in\mathbb{N}} E_n \times F_n$ and $\sum_{n=0}^{\infty} \mu_r E_n \cdot \mu_s F_n \le \theta(\phi^{-1}[A]) + \epsilon$. Now $A \subseteq \bigcup_{n\in\mathbb{N}} \phi[E_n \times F_n]$, so

$$\mu_{r+s}^*(A) \le \sum_{n=0}^{\infty} \mu_{r+s}^*(\phi[E_n \times F_n]) \le \sum_{n=0}^{\infty} \mu_r E_n \cdot \mu_s F_n \le \theta(\phi^{-1}[A]) + \epsilon.$$

As ϵ is arbitrary, we have the result. **Q**

(e) Putting (c) and (d) together, we have $\theta(\phi^{-1}[A]) = \mu_{r+s}^*(A)$ for every $A \subseteq \mathbb{R}^{r+s}$. Thus θ on $\mathbb{R}^r \times \mathbb{R}^s$ corresponds exactly to μ_{r+s}^* on $\mathbb{R}^{r+s}$. So the associated measures λ_0, μ_{r+s} must correspond in the same way, writing λ_0 for the primitive product measure. But 251K tells us that $\lambda_0 = \lambda$, so we have the result.

251O In fact, a large proportion of the applications of the constructions here are to subspaces of Euclidean space, rather than to the whole product $\mathbb{R}^r \times \mathbb{R}^s$. It would not have been especially difficult to write 251N out to deal with arbitrary subspaces, but I prefer to give a more general description of the product of subspace measures, as I feel that it illuminates the method. I start with a straightforward result on strictly localizable spaces.

Proposition Let (X, Σ, μ) and (Y, T, ν) be strictly localizable measure spaces. Then the c.l.d. product measure on $X \times Y$ is strictly localizable; moreover, if $\langle X_i \rangle_{i\in I}$ and $\langle Y_j \rangle_{j\in J}$ are decompositions of X and Y respectively, $\langle X_i \times Y_j \rangle_{(i,j)\in I\times J}$ is a decomposition of $X \times Y$.

proof Let $\langle X_i\rangle_{i\in I}$ and $\langle Y_j\rangle_{j\in J}$ be decompositions of X, Y respectively. Then $\langle X_i\times Y_j\rangle_{(i,j)\in I\times J}$ is a partition of $X\times Y$ into measurable sets of finite measure. If $W\subseteq X\times Y$ and $\lambda W>0$, there are sets $E\in\Sigma$, $F\in\mathrm{T}$ such that $\mu E<\infty$, $\nu F<\infty$ and $\lambda(W\cap(E\times F))>0$. We know that $\mu E=\sum_{i\in I}\mu(E\cap X_i)$ and $\mu F=\sum_{j\in J}\mu(F\cap Y_j)$, so there must be finite sets $I_0\subseteq I$, $J_0\subseteq J$ such that

$$\mu E\cdot\nu F-(\textstyle\sum_{i\in I_0}\mu(E\cap X_i))(\sum_{j\in J_0}\nu(F\cap Y_j))<\lambda(W\cap(E\times F)).$$

Setting $E'=\bigcup_{i\in I_0}X_i$ and $F'=\bigcup_{j\in J_0}Y_j$ we have

$$\lambda((E\times F)\setminus(E'\times F'))=\lambda(E\times F)-\lambda((E\cap E')\times(F\cap F'))<\lambda(W\cap(E\times F)),$$

so that $\lambda(W\cap(E'\times F'))>0$. There must therefore be some $i\in I_0$, $j\in J_0$ such that $\lambda(W\cap(X_i\times Y_j))>0$.

This shows that $\{X_i\times Y_j:i\in I,\,j\in J\}$ satisfies the criterion of 213O, so that λ, being complete and locally determined, must be strictly localizable. Because $\langle X_i\times Y_j\rangle_{(i,j)\in I\times J}$ covers $X\times Y$, it is actually a decomposition of $X\times Y$ (213Ob).

251P Lemma Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and λ the c.l.d. product measure on $X\times Y$. Let λ^* be the corresponding outer measure (132B). Then

$$\lambda^*C=\sup\{\theta(C\cap(E\times F)):E\in\Sigma,\,F\in\mathrm{T},\,\mu E<\infty,\,\nu F<\infty\}$$

for every $C\subseteq X\times Y$, where θ is the outer measure of 251A.

proof Write Λ for the domain of λ, Σ^f for $\{E:E\in\Sigma,\mu E<\infty\}$, T^f for $\{F:F\in\mathrm{T},\nu F<\infty\}$; set $u=\sup\{\theta(C\cap(E\times F)):E\in\Sigma^f,\,F\in\mathrm{T}^f\}$.

(a) If $C\subseteq W\in\Lambda$, $E\in\Sigma^f$ and $F\in\mathrm{T}^f$, then

$$\theta(C\cap(E\times F))\le\theta(W\cap(E\times F))=\lambda_0(W\cap(E\times F))$$

(where λ_0 is the primitive product measure)

$$\le\lambda W.$$

As E and F are arbitrary, $u\le\lambda W$; as W is arbitrary, $u\le\lambda^*C$.

(b) If $u=\infty$, then of course $\lambda^*C=u$. Otherwise, let $\langle E_n\rangle_{n\in\mathbb{N}}$, $\langle F_n\rangle_{n\in\mathbb{N}}$ be sequences in Σ^f, T^f respectively such that

$$u=\sup_{n\in\mathbb{N}}\theta(C\cap(E_n\times F_n)).$$

Consider $C'=C\setminus(\bigcup_{n\in\mathbb{N}}E_n\times\bigcup_{n\in\mathbb{N}}F_n)$. If $E\in\Sigma^f$ and $F\in\mathrm{T}^f$, then for every $n\in\mathbb{N}$ we have

$$\begin{aligned}u&\ge\theta(C\cap((E\cup E_n)\times(F\cup F_n)))\\&=\theta(C\cap((E\cup E_n)\times(F\cup F_n))\cap(E_n\times F_n))\\&\qquad\qquad+\theta(C\cap((E\cup E_n)\times(F\cup F_n))\setminus(E_n\times F_n))\end{aligned}$$

(because $E_n\times F_n\in\Lambda$, by 251E)

$$\ge\theta(C\cap(E_n\times F_n))+\theta(C'\cap(E\times F)).$$

Taking the supremum of the right-hand expression as n varies, we have $u\ge u+\theta(C'\cap(E\times F))$ so

$$\lambda(C'\cap(E\times F))=\theta(C'\cap(E\times F))=0.$$

As E and F are arbitrary, $\lambda C'=0$.

But this means that

$$\begin{aligned}\lambda^*C&\le\lambda^*(C\cap(\bigcup_{n\in\mathbb{N}}E_n\times\bigcup_{n\in\mathbb{N}}F_n))+\lambda^*C'\\&=\lim_{n\to\infty}\lambda^*(C\cap(\bigcup_{i\le n}E_i\times\bigcup_{i\le n}F_i))\end{aligned}$$

(using 132Ae)

$$\le u,$$

as required.

251Q Proposition Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and $A\subseteq X$, $B\subseteq Y$ subsets; write μ_A, ν_B for the subspace measures on A, B respectively. Let λ be the c.l.d. product measure on $X\times Y$, and $\lambda^\#$ the subspace measure it induces on $A\times B$. Let $\tilde\lambda$ be the c.l.d. product measure of μ_A and ν_B on $A\times B$. Then

(i) $\tilde\lambda$ extends $\lambda^\#$.

(ii) If

either (α) $A\in\Sigma$ and $B\in\mathrm{T}$

or (β) A and B can both be covered by sequences of sets of finite measure

or (γ) μ and ν are both strictly localizable,

then $\tilde\lambda=\lambda^\#$.

proof Let θ be the outer measure on $X\times Y$ defined from μ and ν by the formula of 251A, and $\tilde\theta$ the outer measure on $A\times B$ similarly defined from μ_A and ν_B. Write Λ for the domain of λ, $\tilde\Lambda$ for the domain of $\tilde\lambda$, and $\Lambda^\#=\{W\cap(A\times B):W\in\Lambda\}$ for the domain of $\lambda^\#$. Set $\Sigma^f=\{E:\mu E<\infty\}$, $\mathrm{T}^f=\{F:\nu F<\infty\}$.

(a) The first point to observe is that $\tilde\theta C=\theta C$ for every $C\subseteq A\times B$. **P** (i) If $\langle E_n\rangle_{n\in\mathbb{N}}$ and $\langle F_n\rangle_{n\in\mathbb{N}}$ are sequences in Σ, T respectively such that $C\subseteq\bigcup_{n\in\mathbb{N}}E_n\times F_n$, then

$$C=C\cap(A\times B)\subseteq\textstyle\bigcup_{n\in\mathbb{N}}(E_n\cap A)\times(F_n\cap B),$$

so

$$\begin{aligned}\tilde\theta C&\le\sum_{n=0}^{\infty}\mu_A(E_n\cap A)\cdot\nu_B(F_n\cap B)\\&=\sum_{n=0}^{\infty}\mu^*(E_n\cap A)\cdot\nu^*(F_n\cap B)\le\sum_{n=0}^{\infty}\mu E_n\cdot\nu F_n.\end{aligned}$$

As $\langle E_n\rangle_{n\in\mathbb{N}}$ and $\langle F_n\rangle_{n\in\mathbb{N}}$ are arbitrary, $\tilde\theta C\le\theta C$. (ii) If $\langle\tilde E_n\rangle_{n\in\mathbb{N}}$, $\langle\tilde F_n\rangle_{n\in\mathbb{N}}$ are sequences in $\Sigma_A=\operatorname{dom}\mu_A$, $\mathrm{T}_B=\operatorname{dom}\nu_B$ respectively such that $C\subseteq\bigcup_{n\in\mathbb{N}}\tilde E_n\times\tilde F_n$, then for each $n\in\mathbb{N}$ we can choose $E_n\in\Sigma$, $F_n\in\mathrm{T}$ such that

$$\tilde E_n\subseteq E_n,\quad \mu E_n=\mu^*\tilde E_n=\mu_A\tilde E_n,$$

$$\tilde F_n\subseteq F_n,\quad \nu F_n=\nu^*\tilde F_n=\nu_B\tilde F_n,$$

and now

$$\theta C\le\textstyle\sum_{n=0}^{\infty}\mu E_n\cdot\nu F_n=\sum_{n=0}^{\infty}\mu_A\tilde E_n\cdot\nu_B\tilde F_n.$$

As $\langle\tilde E_n\rangle_{n\in\mathbb{N}}$, $\langle\tilde F_n\rangle_{n\in\mathbb{N}}$ are arbitrary, $\theta C\le\tilde\theta C$. **Q**

(b) It follows that $\Lambda^\#\subseteq\tilde\Lambda$. **P** Suppose that $V\in\Lambda^\#$ and that $C\subseteq A\times B$. In this case there is a $W\in\Lambda$ such that $V=W\cap(A\times B)$. So

$$\tilde\theta(C\cap V)+\tilde\theta(C\setminus V)=\theta(C\cap W)+\theta(C\setminus W)=\theta C=\tilde\theta C.$$

As C is arbitrary, $V\in\tilde\Lambda$. **Q**

Accordingly, for $V\in\Lambda^\#$,

$$\begin{aligned}\lambda^\#V=\lambda^*V&=\sup\{\theta(V\cap(E\times F)):E\in\Sigma^f,\,F\in\mathrm{T}^f\}\\&=\sup\{\theta(V\cap(\tilde E\times\tilde F)):\tilde E\in\Sigma_A,\,\tilde F\in\mathrm{T}_B,\,\mu_A\tilde E<\infty,\,\nu_B\tilde F<\infty\}\\&=\sup\{\tilde\theta(V\cap(\tilde E\times\tilde F)):\tilde E\in\Sigma_A,\,\tilde F\in\mathrm{T}_B,\,\mu_A\tilde E<\infty,\,\nu_B\tilde F<\infty\}=\tilde\lambda V,\end{aligned}$$

using 251P twice.

This proves part (i) of the proposition.

(c) The next thing to check is that if $V\in\tilde\Lambda$ and $V\subseteq E\times F$ where $E\in\Sigma^f$ and $F\in\mathrm{T}^f$, then $V\in\Lambda^\#$. **P** Let $W\subseteq E\times F$ be a measurable envelope of V with respect to λ (132Ee). Then

$$\theta(W\cap(A\times B)\setminus V)=\tilde\theta(W\cap(A\times B)\setminus V)=\tilde\lambda(W\cap(A\times B)\setminus V)$$

(because $W\cap(A\times B)\in\Lambda^\#\subseteq\tilde\Lambda$, $V\in\tilde\Lambda$)

$$= \tilde{\lambda}(W \cap (A \times B)) - \tilde{\lambda}V = \tilde{\theta}(W \cap (A \times B)) - \tilde{\theta}V$$
$$= \theta(W \cap (A \times B)) - \theta V = \lambda^*(W \cap (A \times B)) - \lambda^* V$$
$$\leq \lambda W - \lambda^* V = 0.$$

But this means that $W' = W \cap (A \times B) \setminus V \in \Lambda$ and $V = (A \times B) \cap (W \setminus W')$ belongs to $\Lambda^\#$. **Q**

(d) Now fix any $V \in \tilde{\Lambda}$, and look at the conditions (α)-(γ) of part (ii) of the proposition.

(α) If $A \in \Sigma$ and $B \in \mathrm{T}$, and $C \subseteq X \times Y$, then $A \times B \in \Lambda$ (251E), so

$$\begin{aligned}\theta(C \cap V) + \theta(C \setminus V) &= \theta(C \cap V) + \theta((C \setminus V) \cap (A \times B)) + \theta((C \setminus V) \setminus (A \times B)) \\ &= \tilde{\theta}(C \cap V) + \tilde{\theta}(C \cap (A \times B) \setminus V) + \theta(C \setminus (A \times B)) \\ &= \tilde{\theta}(C \cap (A \times B)) + \theta(C \setminus (A \times B)) \\ &= \theta(C \cap (A \times B)) + \theta(C \setminus (A \times B)) = \theta C.\end{aligned}$$

As C is arbitrary, $V \in \Lambda$, so $V = V \cap (A \times B)$ belongs to $\Lambda^\#$.

(β) If $A \subseteq \bigcup_{n\in\mathbb{N}} E_n$ and $B \subseteq \bigcup_{n\in\mathbb{N}} F_n$ where all the E_n, F_n are of finite measure, then $V = \bigcup_{m,n\in\mathbb{N}} V \cap (E_m \times F_n) \in \Lambda^\#$, by (c).

(γ) If $\langle X_i\rangle_{i\in I}$, $\langle Y_j\rangle_{j\in J}$ are decompositions of X, Y respectively, then for each $i \in I$, $j \in J$ we have $V \cap (X_i \times Y_j) \in \Lambda^\#$, that is, there is a $W_{ij} \in \Lambda$ such that $V \cap (X_i \times Y_j) = W_{ij} \cap (A \times B)$. Now $\langle X_i \times Y_j\rangle_{(i,j)\in I\times J}$ is a decomposition of $X \times Y$ for λ (251O), so that

$$W = \bigcup_{i\in I, j\in J} W_{ij} \cap (X_i \times Y_j) \in \Lambda,$$

and $V = W \cap (A \times B) \in \Lambda^\#$.

(e) Thus any of the three conditions is sufficient to ensure that $\tilde{\Lambda} = \Lambda^\#$, in which case (a) tells us that $\tilde{\lambda} = \lambda^\#$.

251R Corollary Let r, $s \geq 1$ be integers, and $\phi : \mathbb{R}^r \times \mathbb{R}^s \to \mathbb{R}^{r+s}$ the natural bijection. If $A \subseteq \mathbb{R}^r$ and $B \subseteq \mathbb{R}^s$, then the restriction of ϕ to $A \times B$ identifies the product of Lebesgue measure on A and Lebesgue measure on B with Lebesgue measure on $\phi[A \times B] \subseteq \mathbb{R}^{r+s}$.

Remark Note that by 'Lebesgue measure on A' I mean the subspace measure μ_{rA} on A induced by r-dimensional Lebesgue measure μ_r on $\mathbb{R}^r$, whether or not A is itself a measurable set.

proof By 251Q, using either of the conditions (ii-β) or (ii-γ), the product measure $\tilde{\lambda}$ on $A \times B$ is just the subspace measure $\lambda^\#$ on $A \times B$ induced by the product measure λ on $\mathbb{R}^r \times \mathbb{R}^s$. But by 251N we know that ϕ is an isomorphism between $(\mathbb{R}^r \times \mathbb{R}^s, \lambda)$ and $(\mathbb{R}^{r+s}, \mu_{r+s})$; so it must also identify $\tilde{\lambda}$ with the subspace measure on $\phi[A \times B]$.

251S Corollary Let (X, Σ, μ) and (Y, T, ν) be measure spaces, and λ the c.l.d. product measure on $X \times Y$. If $A \subseteq X$ and $B \subseteq Y$ can be covered by sequences of sets of finite measure, then $\lambda^*(A \times B) = \mu^* A \cdot \nu^* B$.

proof In the language of 251Q,

$$\lambda^*(A \times B) = \lambda^\#(A \times B) = \tilde{\lambda}(A \times B) = \mu_A A \cdot \nu_B B$$

(by 251K and 251E)

$$= \mu^* A \cdot \nu^* B.$$

251T The next proposition gives an idea of how the technical definitions here fit together.

Proposition Let (X, Σ, μ) and (Y, T, ν) be measure spaces. Write $(X, \hat{\Sigma}, \hat{\mu})$ and $(X, \tilde{\Sigma}, \tilde{\mu})$ for the completion and c.l.d. version of (X, Σ, μ) (212C, 213E). Let λ, $\hat{\lambda}$ and $\tilde{\lambda}$ be the three c.l.d. product measures on $X \times Y$ obtained from the pairs (μ, ν), $(\hat{\mu}, \nu)$ and $(\tilde{\mu}, \nu)$ of factor measures. Then $\lambda = \hat{\lambda} = \tilde{\lambda}$.

proof Write Λ, $\hat{\Lambda}$ and $\tilde{\Lambda}$ for the domains of λ, $\hat{\lambda}$, $\tilde{\lambda}$ respectively; and θ, $\hat{\theta}$, $\tilde{\theta}$ for the outer measures on $X \times Y$ obtained by the formula of 251A from the three pairs of factor measures.

(a) If $E \in \Sigma$ and $\mu E < \infty$, then θ, $\hat{\theta}$ and $\tilde{\theta}$ agree on subsets of $E \times Y$. **P** Take $A \subseteq E \times Y$ and $\epsilon > 0$.

(i) There are sequences $\langle E_n \rangle_{n\in\mathbb{N}}$ in Σ, $\langle F_n \rangle_{n\in\mathbb{N}}$ in T such that $A \subseteq \bigcup_{n\in\mathbb{N}} E_n \times F_n$ and $\sum_{n=0}^{\infty} \mu E_n \cdot \nu F_n \le \theta A + \epsilon$. Now $\tilde{\mu} E_n \le \mu E_n$ for every n (213Fb), so

$$\tilde{\theta} A \le \sum_{n=0}^{\infty} \tilde{\mu} E_n \cdot \nu F_n \le \sum_{n=0}^{\infty} \mu E_n \cdot \nu F_n \le \theta A + \epsilon.$$

(ii) There are sequences $\langle \hat{E}_n \rangle_{n\in\mathbb{N}}$ in $\hat{\Sigma}$, $\langle \hat{F}_n \rangle_{n\in\mathbb{N}}$ in T such that $A \subseteq \bigcup_{n\in\mathbb{N}} \hat{E}_n \times \hat{F}_n$ and $\sum_{n=0}^{\infty} \hat{\mu}\hat{E}_n \cdot \nu\hat{F}_n \le \hat{\theta} A + \epsilon$. Now for each n there is an $E'_n \in \Sigma$ such that $\hat{E}_n \subseteq E'_n$ and $\mu E'_n = \hat{\mu}\hat{E}_n$, so that

$$\theta A \le \sum_{n=0}^{\infty} \mu E'_n \cdot \nu\hat{F}_n = \sum_{n=0}^{\infty} \hat{\mu}\hat{E}_n \cdot \nu\hat{F}_n \le \hat{\theta} A + \epsilon.$$

(iii) There are sequences $\langle \tilde{E}_n \rangle_{n\in\mathbb{N}}$ in $\tilde{\Sigma}$, $\langle \tilde{F}_n \rangle_{n\in\mathbb{N}}$ in T such that $A \subseteq \bigcup_{n\in\mathbb{N}} \tilde{E}_n \times \tilde{F}_n$ and $\sum_{n=0}^{\infty} \tilde{\mu}\tilde{E}_n \cdot \nu\tilde{F}_n \le \tilde{\theta} A + \epsilon$. Now for each n, $\tilde{E}_n \cap E \in \hat{\Sigma}$, so

$$\hat{\theta} A \le \sum_{n=0}^{\infty} \hat{\mu}(\tilde{E}_n \cap E) \cdot \nu\tilde{F}_n \le \sum_{n=0}^{\infty} \tilde{\mu}\tilde{E}_n \cdot \nu\tilde{F}_n \le \tilde{\theta} A + \epsilon.$$

(iv) Since A and ϵ are arbitrary, $\theta = \hat{\theta} = \tilde{\theta}$ on $\mathcal{P}(E \times Y)$. **Q**

(b) Consequently, the outer measures λ^*, $\hat{\lambda}^*$ and $\tilde{\lambda}^*$ are identical. **P** Use 251P. Take $A \subseteq X \times Y$, $E \in \Sigma$, $\hat{E} \in \hat{\Sigma}$, $\tilde{E} \in \tilde{\Sigma}$, $F \in$ T such that μE, $\hat{\mu}\hat{E}$, $\tilde{\mu}\tilde{E}$ and νF are all finite. Then

(i)

$$\theta(A \cap (E \times F)) = \hat{\theta}(A \cap (E \times F)) \le \hat{\lambda}^* A, \quad \theta(A \cap (E \times F)) = \tilde{\theta}(A \cap (E \times F)) \le \tilde{\lambda}^* A$$

because $\hat{\mu} E$ and $\tilde{\mu} E$ are both finite.

(ii) There is an $E' \in \Sigma$ such that $\hat{E} \subseteq E'$ and $\mu E' < \infty$, so that

$$\hat{\theta}(A \cap (\hat{E} \times F)) \le \hat{\theta}(A \cap (E' \times F)) = \theta(A \cap (E' \times F)) \le \lambda^* A.$$

(iii) There is an $E'' \in \Sigma$ such that $E'' \subseteq \tilde{E}$ and $\tilde{\mu}(\tilde{E} \setminus E'') = 0$ (213Fc), so that $\tilde{\theta}((\tilde{E} \setminus E'') \times Y) = 0$ and $\mu E'' < \infty$; accordingly

$$\tilde{\theta}(A \cap (\tilde{E} \times F)) = \tilde{\theta}(A \cap (E'' \times F)) = \theta(A \cap (E'' \times F)) \le \lambda^* A.$$

(iv) Taking the suprema over E, $\hat{E}$, $\tilde{E}$ and F, we get

$$\lambda^* A \le \hat{\lambda}^* A, \quad \lambda^* A \le \tilde{\lambda}^* A, \quad \hat{\lambda}^* A \le \lambda^* A, \quad \tilde{\lambda}^* A \le \lambda^* A.$$

As A is arbitrary, $\lambda^* = \hat{\lambda}^* = \tilde{\lambda}^*$. **Q**

(c) Now λ, $\hat{\lambda}$ and $\tilde{\lambda}$ are all complete and locally determined, so by 213C are the measures defined by Carathéodory's method from their own outer measures, and are therefore identical.

251U It is 'obvious', and an easy consequence of theorems so far proved, that the set $\{(x, x) : x \in \mathbb{R}\}$ is negligible for Lebesgue measure on $\mathbb{R}^2$. The corresponding result is true in the square of any *atomless* measure space.

Proposition Let (X, Σ, μ) be an atomless measure space, and let λ be the c.l.d. measure on $X \times X$. Then $\Delta = \{(x, x) : x \in X\}$ is λ-negligible.

proof Let E, $F \in \Sigma$ be sets of finite measure, and $n \in \mathbb{N}$. Applying 215D repeatedly, we can find a disjoint family $\langle F_i \rangle_{i<n}$ of measurable subsets of F such that $\mu F_i = \frac{\mu F}{n+1}$ for each i; setting $F_n = F \setminus \bigcup_{i<n} F_i$, we also have $\mu F_n = \frac{\mu F}{n+1}$. Now

$$\Delta \cap (E \times F) \subseteq \bigcup_{i\le n} (E \cap F_i) \times F_i,$$

so

$$\lambda^*(\Delta \cap (E \times F)) \le \sum_{i=0}^{n} \mu(E \cap F_i) \cdot \mu F_i = \frac{\mu F}{n+1} \sum_{i=0}^{n} \mu(E \cap F_i) \le \frac{1}{n+1} \mu E \cdot \mu F.$$

As n is arbitrary, $\lambda(\Delta \cap (E \times F)) = 0$; as E and F are arbitrary, $\lambda\Delta = 0$.

***251W Products of more than two spaces** The whole of this section can be repeated for arbitrary finite products. The labour is substantial but no new ideas are required. By the time we need the general construction in any formal way, it should come very naturally, and I do not think it is necessary to work through the next couple of pages before proceeding, especially as products of *probability* spaces are dealt with in §254. However, for completeness, and to help locate results when applications do appear, I list them here. They do of course constitute a very instructive set of exercises. The most important fragments are repeated in 251Xe-251Xf.

Let $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be a finite family of measure spaces, and set $X=\prod_{i\in I}X_i$. Write $\Sigma_i^f=\{E:E\in\Sigma_i,\,\mu_iE<\infty\}$ for each $i\in I$.

(a) For $A\subseteq X$ set

$$\theta A=\inf\{\sum_{n=0}^{\infty}\prod_{i\in I}\mu_iE_{ni}:E_{ni}\in\Sigma_i\ \forall\ i\in I,\,n\in\mathbb{N},\,A\subseteq\bigcup_{n\in\mathbb{N}}\prod_{i\in I}E_{ni}\}.$$

Then θ is an outer measure on X. Let λ_0 be the measure on X derived by Carathéodory's method from θ, and Λ its domain.

(b) If $\langle X_i\rangle_{i\in I}$ is a finite family of sets and Σ_i is a σ-algebra of subsets of X_i for each $i\in I$, then $\widehat{\bigotimes}_{i\in I}\Sigma_i$ is the σ-algebra of subsets of $X=\prod_{i\in I}X_i$ generated by $\{\prod_{i\in I}E_i:E_i\in\Sigma_i$ for every $i\in I\}$. (For the corresponding construction when I is infinite, see 254E.)

(c) $\lambda_0(\prod_{i\in I}E_i)$ is defined and equal to $\prod_{i\in I}\mu_iE_i$ whenever $E_i\in\Sigma_i$ for each $i\in I$.

(d) The **c.l.d. product measure** on X is the measure λ defined by setting

$$\lambda W=\sup\{\lambda_0(W\cap\textstyle\prod_{i\in I}E_i):E_i\in\Sigma_i^f\text{ for each }i\in I\}$$

for $W\in\Lambda$. If I is empty, so that $X=\{\emptyset\}$, then the appropriate convention is to set $\lambda X=1$.

(e) If $H\subseteq X$, then $H\in\Lambda$ iff $H\cap\prod_{i\in I}E_i\in\Lambda$ whenever $E_i\in\Sigma_i^f$ for each $i\in I$.

(f)(i) $\widehat{\bigotimes}_{i\in I}\Sigma_i\subseteq\Lambda$ and $\lambda(\prod_{i\in I}E_i)=\prod_{i\in I}\mu_iE_i$ whenever $E_i\in\Sigma_i^f$ for each i.

(ii) For every $W\in\Lambda$ there is a $V\in\widehat{\bigotimes}_{i\in I}\Sigma_i$ such that $V\subseteq W$ and $\lambda V=\lambda W$.

(iii) λ is complete and locally determined, and is the c.l.d. version of λ_0.

(iv) If $W\in\Lambda$ and $\lambda W>0$ then there are $E_i\in\Sigma_i^f$, for $i\in I$, such that $\lambda(W\cap\prod_{i\in I}E_i)>0$.

(v) If $W\in\Lambda$ and $\lambda W<\infty$, then for every $\epsilon>0$ there are $n\in\mathbb{N}$ and $E_{0i},\dots,E_{ni}\in\Sigma_i^f$, for each $i\in I$, such that $\lambda(W\triangle\bigcup_{k\le n}\prod_{i\in I}E_{ki})\le\epsilon$.

(g) If each μ_i is σ-finite, so is λ, and $\lambda=\lambda_0$ is the completion of its restriction to $\widehat{\bigotimes}_{i\in I}\Sigma_i$.

(h) If $\langle I_j\rangle_{j\in J}$ is any partition of I, then λ can be identified with the c.l.d. product of $\langle\lambda_j\rangle_{j\in J}$, where λ_j is the c.l.d. product of $\langle\mu_i\rangle_{i\in I_j}$. (See the arguments in 251N and also in 254N below.)

(i) If $I=\{1,\dots,n\}$ and each μ_i is Lebesgue measure on $\mathbb{R}$, then λ can be identified with Lebesgue measure on $\mathbb{R}^n$.

(j) If, for each $i\in I$, we have a decomposition $\langle X_{ij}\rangle_{j\in J_i}$ of X_i, then $\langle\prod_{i\in I}X_{i,f(i)}\rangle_{f\in\prod_{i\in I}J_i}$ is a decomposition of X.

(k) For any $C\subseteq X$,

$$\lambda^*C=\sup\{\theta(C\cap\textstyle\prod_{i\in I}E_i):E_i\in\Sigma_i^f\text{ for every }i\in I\}.$$

(l) Suppose that $A_i\subseteq X_i$ for each $i\in I$. Write $\lambda^\#$ for the subspace measure on $A=\prod_{i\in I}A_i$, and $\tilde\lambda$ for the c.l.d. product of the subspace measures on the A_i. Then $\tilde\lambda$ extends $\lambda^\#$, and if

either $A_i\in\Sigma_i$ for every i

or every A_i can be covered by a sequence of sets of finite measure

or every μ_i is strictly localizable,

then $\tilde\lambda=\lambda^\#$.

(m) If $A_i\subseteq X_i$ can be covered by a sequence of sets of finite measure for each $i\in I$, then $\lambda^*(\prod_{i\in I}A_i)=\prod_{i\in I}\mu_i^*A_i$.

(n) Writing $\hat{\mu}_i$, $\tilde{\mu}_i$ for the completion and c.l.d. version of each μ_i, λ is the c.l.d. product of $\langle\hat{\mu}_i\rangle_{i\in I}$ and also of $\langle\tilde{\mu}_i\rangle_{i\in I}$.

(o) If all the (X_i,Σ_i,μ_i) are the same atomless measure space, then $\{x : x\in X,\, i\mapsto x(i)$ is injective$\}$ is λ-conegligible.

(p) Now suppose that we have another family $\langle(Y_i,\mathrm{T}_i,\nu_i)\rangle_{i\in I}$ of measure spaces, with product (Y,Λ',λ'), and inverse-measure-preserving functions $f_i : X_i\to Y_i$ for each i; define $f : X\to Y$ by saying that $f(x)(i)=f_i(x(i))$ for $x\in X$ and $i\in I$. If all the ν_i are σ-finite, then f is inverse-measure-preserving for λ and λ'.

251X Basic exercises (a) Let X and Y be sets, $\mathcal{A}\subseteq\mathcal{P}X$ and $\mathcal{B}\subseteq\mathcal{P}Y$. Let Σ be the σ-algebra of subsets of X generated by $\mathcal{A}$ and T the σ-algebra of subsets of Y generated by $\mathcal{B}$. Show that $\Sigma\widehat{\otimes}\mathrm{T}$ is the σ-algebra of subsets of $X\times Y$ generated by $\{A\times B : A\in\mathcal{A},\, B\in\mathcal{B}\}$.

(b) Let (X,Σ,μ) and (Y,T,ν) be measure spaces; let λ_0 be the primitive product measure on $X\times Y$, and λ the c.l.d. product measure. Show that $\lambda_0W<\infty$ iff $\lambda W<\infty$ and W is included in a set of the form

$$(E\times Y)\cup(X\times F)\cup\bigcup_{n\in\mathbb{N}}E_n\times F_n$$

where $\mu E=\nu F=0$ and $\mu E_n<\infty$, $\nu F_n<\infty$ for every n.

>(c) Show that if X and Y are any sets, with their respective counting measures, then the primitive and c.l.d. product measures on $X\times Y$ are both counting measure on $X\times Y$.

(d) Let (X,Σ,μ) and (Y,T,ν) be measure spaces; let λ_0 be the primitive product measure on $X\times Y$, and λ the c.l.d. product measure. Show that

$$\begin{aligned}&\lambda_0 \text{ is locally determined}\\ &\iff \lambda_0 \text{ is semi-finite}\\ &\iff \lambda_0=\lambda\\ &\iff \lambda_0 \text{ and } \lambda \text{ have the same negligible sets.}\end{aligned}$$

>(e) (See 251W.) Let $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be a family of measure spaces, where I is a non-empty finite set. Set $X=\prod_{i\in I}X_i$. For $A\subseteq X$, set

$$\theta A=\inf\{\sum_{n=0}^{\infty}\prod_{i\in I}\mu_iE_{ni} : E_{ni}\in\Sigma_i\ \forall\, n\in\mathbb{N},\, i\in I,\, A\subseteq\bigcup_{n\in\mathbb{N}}\prod_{i\in I}E_{ni}\}.$$

Show that θ is an outer measure on X. Let λ_0 be the measure defined from θ by Carathéodory's method, and for $W\in\operatorname{dom}\lambda_0$ set

$$\lambda W=\sup\{\lambda_0(W\cap\prod_{i\in I}E_i) : E_i\in\Sigma_i,\, \mu_iE_i<\infty \text{ for every } i\in I\}.$$

Show that λ is a measure on X, and is the c.l.d. version of λ_0.

>(f) (See 251W.) Let I be a non-empty finite set and $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ a family of measure spaces. For non-empty $K\subseteq I$ set $X^{(K)}=\prod_{i\in K}X_i$ and let $\lambda_0^{(K)}$, $\lambda^{(K)}$ be the measures on $X^{(K)}$ constructed as in 251Xe. Show that if K is a non-empty proper subset of I, then the natural bijection between $X^{(I)}$ and $X^{(K)}\times X^{(I\setminus K)}$ identifies $\lambda_0^{(I)}$ with the primitive product measure of $\lambda_0^{(K)}$ and $\lambda_0^{(I\setminus K)}$, and $\lambda^{(I)}$ with the c.l.d. product measure of $\lambda^{(K)}$ and $\lambda^{(I\setminus K)}$.

>(g) Using 251Xe-251Xf above, or otherwise, show that if (X_1,Σ_1,μ_1), (X_2,Σ_2,μ_2), (X_3,Σ_3,μ_3) are measure spaces then the primitive and c.l.d. product measures λ_0, λ of $(X_1\times X_2)\times X_3$, constructed by first taking the appropriate product measure on $X_1\times X_2$ and then taking the product of this with the measure of X_3, are identified with the corresponding product measures on $X_1\times(X_2\times X_3)$ by the canonical bijection between the sets $(X_1\times X_2)\times X_3$ and $X_1\times(X_2\times X_3)$.

(h)(i) What happens in 251Xe when I is a singleton? (ii) Devise an appropriate convention to make 251Xe-251Xf remain valid when one or more of the sets I, K, $I\setminus K$ there is empty.

>(i) Let (X,Σ,μ) be a complete locally determined measure space, and I any non-empty set; let ν be counting measure on I. Show that the c.l.d. product measure on $X\times I$ is equal to (or at any rate identifiable with) the direct sum measure of the family $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$, if we set $(X_i,\Sigma_i,\mu_i)=(X,\Sigma,\mu)$ for every i.

>(j) Let $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be a family of measure spaces, with direct sum (X,Σ,μ) (214L). Let (Y,T,ν) be any measure space, and give $X\times Y$, $X_i\times Y$ their c.l.d. product measures. Show that the natural bijection between $X\times Y$ and $Z=\bigcup_{i\in I}(X_i\times Y)\times\{i\}$ is an isomorphism between the measure of $X\times Y$ and the direct sum measure on Z.

>(k) Let (X,Σ,μ) be any measure space, and Y a singleton set $\{y\}$; let ν be the measure on Y such that $\nu Y=1$. Show that the natural bijection between $X\times\{y\}$ and X identifies the primitive product measure on $X\times\{y\}$ with $\check\mu$ as defined in 213Xa, and the c.l.d. product measure with the c.l.d. version of μ. Explain how to put this together with 251Xg and 251Ic to prove 251T.

>(l) Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and λ the c.l.d. product measure on $X\times Y$. Show that λ is the c.l.d. version of its restriction to $\Sigma\widehat{\otimes}\mathrm{T}$.

(m) Let (X,Σ,μ) and (Y,T,ν) be measure spaces, with primitive and c.l.d. product measures λ_0, λ. Let λ_1 be any measure with domain $\Sigma\widehat{\otimes}\mathrm{T}$ such that $\lambda_1(E\times F)=\mu E\cdot\nu F$ whenever $E\in\Sigma$ and $F\in\mathrm{T}$. Show that $\lambda W\le\lambda_1 W\le\lambda_0 W$ for every $W\in\Sigma\widehat{\otimes}\mathrm{T}$.

(n) Let (X,Σ,μ) and (Y,T,ν) be two measure spaces, and λ_0 the primitive product measure on $X\times Y$. Show that the corresponding outer measure λ_0^* is just the outer measure θ of 251A.

(o) Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and $A\subseteq X$, $B\subseteq Y$ subsets; write μ_A, ν_B for the subspace measures. Let λ_0 be the primitive product measure on $X\times Y$, and $\lambda_0^\#$ the subspace measure it induces on $A\times B$. Let $\tilde\lambda_0$ be the primitive product measure of μ_A and ν_B on $A\times B$. Show that $\tilde\lambda_0$ extends $\lambda_0^\#$. Show that if *either* (α) $A\in\Sigma$ and $B\in\mathrm{T}$ *or* (β) A and B can both be covered by sequences of sets of finite measure *or* (γ) μ and ν are both strictly localizable, then $\tilde\lambda_0=\lambda_0^\#$.

(p) In 251Q, show that $\tilde\lambda$ and $\lambda^\#$ will have the same null ideals, even if none of the conditions of 251Q(ii) are satisfied.

(q) Let (X,Σ,μ) and (Y,T,ν) be any measure spaces, and λ_0 the primitive product measure on $X\times Y$. Show that $\lambda_0^*(A\times B)=\mu^*A\cdot\nu^*B$ for any $A\subseteq X$ and $B\subseteq Y$.

(r) Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and $\hat\mu$ the completion of μ. Show that μ, ν and $\hat\mu$, ν have the same primitive product measures.

(s) Let (X,Σ,μ) be a semi-finite measure space. Show that μ is atomless iff the diagonal $\{(x,x):x\in X\}$ is negligible for the c.l.d. product measure on $X\times X$.

(t) Let (X,Σ,μ) be an atomless measure space, and (Y,T,ν) any measure space. Show that the c.l.d. product measure on $X\times Y$ is atomless.

>(u) Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and λ the c.l.d. product measure on $X\times Y$. (i) Show that if μ and ν are purely atomic, so is λ. (ii) Show that if μ and ν are point-supported, so is λ.

251Y Further exercises (a) Let X, Y be sets with σ-algebras of subsets Σ, T. Suppose that $h:X\times Y\to\mathbb{R}$ is $\Sigma\widehat{\otimes}\mathrm{T}$-measurable and $\phi:X\to Y$ is (Σ,T)-measurable (121Yc). Show that $x\mapsto h(x,\phi(x)):X\to\mathbb{R}$ is Σ-measurable.

(b) Show that there are measure spaces (X_1,Σ_1,μ_1) and (X_2,Σ_2,μ_2), a probability space (Y,T,ν) and an inverse-measure-preserving function $f:X_1\to X_2$ such that $h:X_1\times Y\to X_2\times Y$ is not inverse-measure-preserving for the c.l.d. product measures on $X_1\times Y$ and $X_2\times Y$, where $h(x,y)=(f(x),y)$ for $x\in X_1$ and $y\in Y$.

(c) Let (X,Σ,μ) be a complete locally determined measure space with a subspace A whose measure is not locally determined (see 216Xb). Set $Y=\{0\}$, $\nu Y=1$ and consider the c.l.d. product measures on $X\times Y$ and $A\times Y$; write Λ, $\tilde\Lambda$ for their domains. Show that $\tilde\Lambda$ properly includes $\{W\cap(A\times Y):W\in\Lambda\}$.

(d) Let (X,Σ,μ) be any measure space, (Y,T,ν) an atomless measure space, and $f:X\to Y$ a (Σ,T)-measurable function. Show that $\{(x,f(x)):x\in X\}$ is negligible for the c.l.d. product measure on $X\times Y$.

251 Notes and comments There are real difficulties in deciding which construction to declare as 'the' product of two arbitrary measures. My phrase 'primitive product measure', and notation λ_0, betray a bias; my own preference is for the c.l.d. product λ, for two principal reasons. The first is that λ_0 is likely to be 'bad', in particular, not semi-finite, even if μ and ν are 'good' (251Xd, 252Yk), while λ inherits some of the most important properties of μ and ν (e.g., 251O); the second is that in the case of topological measure spaces X and Y, there is often a canonical topological measure on $X \times Y$, which is likely to be more closely related to λ than to λ_0. But for elucidation of this point I must ask you to wait until §417 in Volume 4.

It would be possible to remove the 'primitive' product measure entirely from the exposition, or at least to relegate it to the exercises. This is indeed what I expect to do in the rest of this treatise, since (in my view) all significant features of product measures on finitely many factors can be expressed in terms of the c.l.d. product measure. For the first introduction to product measures, however, a direct approach to the c.l.d. product measure (through the description of λ^* in 251P, for instance) is an uncomfortably large bite, and I have some sort of duty to present the most natural rival to the c.l.d. product measure prominently enough for you to judge for yourself whether I am right to dismiss it. There certainly are results associated with the primitive product measure (251Xn, 251Xq, 252Yc) which have an agreeable simplicity.

The clash is avoided altogether, of course, if we specialize immediately to σ-finite spaces, in which the two constructions coincide (251K). But even this does not solve all problems. There is a popular alternative measure often called 'the' product measure: the restriction $\lambda_{0\mathcal{B}}$ of λ_0 to the σ-algebra $\Sigma\widehat{\otimes}\mathrm{T}$. (See, for instance, Halmos 50.) The advantage of this is that if a function f on $X \times Y$ is $\Sigma\widehat{\otimes}\mathrm{T}$-measurable, then $x \mapsto f(x,y)$ is Σ-measurable for every $y \in Y$. (This is because

$$\{W : W \subseteq X \times Y,\, \{x : (x,y) \in W\} \in \Sigma \;\forall\; y \in Y\}$$

is a σ-algebra of subsets of $X \times Y$ containing $E \times F$ whenever $E \in \Sigma$ and $F \in \mathrm{T}$, and therefore including $\Sigma\widehat{\otimes}\mathrm{T}$.) The primary objection, to my mind, is that Lebesgue measure on $\mathbb{R}^2$ is no longer 'the' product of Lebesgue measure on $\mathbb{R}$ with itself. Generally, it is right to seek measures which measure as many sets as possible, and I prefer to face up to the technical problems (which I acknowledge are off-putting) by seeking appropriate definitions on the approach to major theorems, rather than rely on ad hoc fixes when the time comes to apply them.

I omit further examples of product measures for the moment, because the investigation of particular examples will be much easier with the aid of results from the next section. Of course the leading example, and the one which should come always to mind in response to the words 'product measure', is Lebesgue measure on $\mathbb{R}^2$, the case $r = s = 1$ of 251N and 251R. For an indication of what can happen when one of the factors is not σ-finite, you could look ahead to 252K.

I hope that you will see that the definition of the outer measure θ in 251A corresponds to the standard definition of Lebesgue outer measure, with 'measurable rectangles' $E \times F$ taking the place of intervals, and the functional $E \times F \mapsto \mu E \cdot \nu F$ taking the place of 'length' or 'volume' of an interval; moreover, thinking of E and F as intervals, there is an obvious relation between Lebesgue measure on $\mathbb{R}^2$ and the product measure on $\mathbb{R} \times \mathbb{R}$. Of course an 'obvious relationship' is not the same thing as a proper theorem with exact hypotheses and conclusions, but Theorem 251N is clearly central. Long before that, however, there is another parallel between the construction of 251A and that of Lebesgue measure. In both cases, the proof that we have an outer measure comes directly from the defining formula (in 113Yd I gave as an exercise a general result covering 251B), and consequently a very general construction can lead us to a measure. But the measure would be of far less interest and value if it did not measure, and measure correctly, the basic sets, in this case the measurable rectangles. Thus 251E corresponds to the theorem that intervals are Lebesgue measurable, with the right measure (114Db, 114G). This is the real key to the construction, and is one of the fundamental ideas of measure theory.

Yet another parallel is in 251Xn; the outer measure defining the primitive product measure λ_0 is exactly equal to the outer measure defined from λ_0. I described the corresponding phenomenon for Lebesgue measure in 132C.

Any construction which claims the title 'canonical' must satisfy a variety of natural requirements; for instance, one expects the canonical bijection between $X \times Y$ and $Y \times X$ to be an isomorphism between the corresponding product measure spaces. 'Commutativity' of the product in this sense is I think obvious from the definitions in 251A-251C. It is obviously desirable – not, I think, obviously true – that the product should be 'associative' in that the canonical bijection between $(X \times Y) \times Z$ and $X \times (Y \times Z)$ should also be an isomorphism between the corresponding products of product measures. This is in fact valid for both the primitive and c.l.d. product measures (251Wh, 251Xe-251Xg).

Working through the classification of measure spaces presented in §211, we find that the primitive product measure λ_0 of arbitrary factor measures μ, ν is complete, while the c.l.d. product measure λ is always complete and locally determined. λ_0 may not be semi-finite, even if μ and ν are strictly localizable (252Yk); but λ will be strictly localizable if μ and ν are (251O). Of course this is associated with the fact that the c.l.d. product measure is distributive over direct sums (251Xj). If either μ or ν is atomless, so is λ (251Xt). Both λ and λ_0 are σ-finite if μ and ν are (251K). It is possible for both μ and ν to be localizable but λ not (254U).

At least if you have worked through Chapter 21, you have now done enough 'pure' measure theory for this kind of investigation, however straightforward, to raise a good many questions. Apart from direct sums, we also have the constructions of 'completion', 'subspace', 'outer measure' and (in particular) 'c.l.d. version' to integrate into the new ideas; I offer some results in 251T and 251Xk. Concerning subspaces, some possibly surprising difficulties arise. The problem is that the product measure on the product of two subspaces can have a larger domain than one might expect. I give a simple example in 251Yc and a more elaborate one in 254Ye. For strictly localizable spaces, there is no problem (251Q); but no other criterion drawn from the list of properties considered in §251 seems adequate to remove the possibility of a disconcerting phenomenon.

252 Fubini's theorem

Perhaps the most important feature of the concept of 'product measure' is the fact that we can use it to discuss repeated integrals. In this section I give versions of Fubini's theorem and Tonelli's theorem (252B, 252G) with a variety of corollaries, the most useful ones being versions for σ-finite spaces (252C, 252H). As applications I describe the relationship between integration and measuring ordinate sets (252N) and calculate the r-dimensional volume of a ball in $\mathbb{R}^r$ (252Q, 252Xi). I mention counter-examples showing the difficulties which can arise with non-σ-finite measures and non-integrable functions (252K-252L, 252Xf-252Xg).

252A Repeated integrals Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and f a real-valued function defined on a set $\operatorname{dom} f \subseteq X\times Y$. We can seek to form the **repeated integral**

$$\iint f(x,y)\nu(dy)\mu(dx) = \int\Bigl(\int f(x,y)\nu(dy)\Bigr)\mu(dx),$$

which should be interpreted as follows: set

$$D = \{x : x\in X,\ \int f(x,y)\nu(dy) \text{ is defined in } [-\infty,\infty]\},$$

$$g(x) = \int f(x,y)\nu(dy) \text{ for } x\in D,$$

and then write $\iint f(x,y)\nu(dy)\mu(dx) = \int g(x)\mu(dx)$ if this is defined. Of course the subset of Y on which $y\mapsto f(x,y)$ is defined may vary with x, but it must always be conegligible, as must D.

Similarly, exchanging the roles of X and Y, we can seek a repeated integral

$$\iint f(x,y)\mu(dx)\nu(dy) = \int\Bigl(\int f(x,y)\mu(dx)\Bigr)\nu(dy).$$

The point is that, under appropriate conditions on μ and ν, we can relate these repeated integrals to each other by connecting them both with the integral of f itself with respect to the product measure on $X\times Y$.

As will become apparent shortly, it is essential here to allow oneself to discuss the integral of a function which is not everywhere defined. It is of less importance whether one allows integrands and integrals to take infinite values, but for definiteness let me say that I shall be following the rules of 135F; that is, $\int f = \int f^+ - \int f^-$ provided that f is defined almost everywhere, takes values in $[-\infty,\infty]$ and is virtually measurable, and at most one of $\int f^+$, $\int f^-$ is infinite.

252B Theorem Let (X,Σ,μ) and (Y,T,ν) be measure spaces, with c.l.d. product $(X\times Y,\Lambda,\lambda)$ (251F). Suppose that ν is σ-finite and that μ is *either* strictly localizable *or* complete and locally determined. Let f be a $[-\infty,\infty]$-valued function such that $\int f d\lambda$ is defined in $[-\infty,\infty]$. Then $\iint f(x,y)\nu(dy)\mu(dx)$ is defined and is equal to $\int f d\lambda$.

proof The proof of this result involves substantial technical difficulties. If you have not seen these ideas before, you should almost certainly not go straight to the full generality of the version announced above. I will therefore start by writing out a proof in the case in which both μ and ν are totally finite; this is already lengthy enough. I will present it in such a way that only the central section (part (b) below) needs to be amended in the general case, and then, after completing the proof of the special case, I will give the alternative version of (b) which is required for the full result.

(a) Write $\mathcal{L}$ for the family of $[0,\infty]$-valued functions f such that $\int f d\lambda$ and $\iint f(x,y)\nu(dy)\mu(dx)$ are defined and equal. My aim is to show first that $f\in\mathcal{L}$ whenever f is non-negative and $\int f d\lambda$ is defined, and then to look at differences of functions in $\mathcal{L}$. To prove that enough functions belong to $\mathcal{L}$, my strategy will be to start with 'elementary' functions and work outwards through progressively larger classes. It is most efficient to begin by describing ways of building new members of $\mathcal{L}$ from old, as follows.

(i) $f_1+f_2\in\mathcal{L}$ for all f_1, $f_2\in\mathcal{L}$, and $cf\in\mathcal{L}$ for all $f\in\mathcal{L}$, $c\in[0,\infty[$; this is because

$$\int (f_1+f_2)(x,y)\nu(dy) = \int f_1(x,y)\nu(dy) + \int f_2(x,y)\nu(dy),$$

$$\int (cf)(x,y)\nu(dy) = c\int f(x,y)\nu(dy)$$

whenever the right-hand sides are defined, which we are supposing to be the case for almost every x, so that

$$\begin{aligned}\iint (f_1+f_2)(x,y)\nu(dy)\mu(dx) &= \iint f_1(x,y)\nu(dy)\mu(dx) + \iint f_2(x,y)\nu(dy)\mu(dx)\\ &= \int f_1 d\lambda + \int f_2 d\lambda = \int (f_1+f_2)d\lambda,\end{aligned}$$

$$\iint (cf)(x,y)\nu(dy)\mu(dx) = c\int f(x,y)\nu(dy)\mu(dx) = c\int f d\lambda = \int (cf)d\lambda.$$

(ii) If $\langle f_n\rangle_{n\in\mathbb{N}}$ is a sequence in $\mathcal{L}$ such that $f_n(x,y) \le f_{n+1}(x,y)$ whenever $n\in\mathbb{N}$ and $(x,y)\in \operatorname{dom} f_n \cap \operatorname{dom} f_{n+1}$, then $\sup_{n\in\mathbb{N}} f_n \in \mathcal{L}$. **P** Set $f = \sup_{n\in\mathbb{N}} f_n$; for $x\in X$, $n\in\mathbb{N}$ set $g_n(x) = \int f_n(x,y)\nu(dy)$ when the integral is defined in $[0,\infty]$. Since here I am allowing ∞ as a value of a function, it is natural to regard f as defined on $\bigcap_{n\in\mathbb{N}} \operatorname{dom} f_n$. By B.Levi's theorem, $\int f d\lambda = \sup_{n\in\mathbb{N}} \int f_n d\lambda$; write u for this common value in $[0,\infty]$. Next, because $f_n \le f_{n+1}$ wherever both are defined, $g_n \le g_{n+1}$ wherever both are defined, for each n; we are supposing that $f_n\in\mathcal{L}$, so g_n is defined μ-almost everywhere for each n, and

$$\sup_{n\in\mathbb{N}} \int g_n d\mu = \sup_{n\in\mathbb{N}} \int f_n d\lambda = u.$$

By B.Levi's theorem again, $\int g\,d\mu = u$, where $g = \sup_{n\in\mathbb{N}} g_n$. Now take any $x\in\bigcap_{n\in\mathbb{N}} \operatorname{dom} g_n$, and consider the functions f_{xn} on Y, setting $f_{xn}(y) = f_n(x,y)$ whenever this is defined. Each f_{xn} has an integral in $[0,\infty]$, and $f_{xn}(y) \le f_{x,n+1}(y)$ whenever both are defined, and

$$\sup_{n\in\mathbb{N}} \int f_{xn} d\nu = g(x);$$

so, using B.Levi's theorem for a third time, $\int(\sup_{n\in\mathbb{N}} f_{xn})d\nu$ is defined and equal to $g(x)$, that is,

$$\int f(x,y)\nu(dy) = g(x).$$

This is true for almost every x, so

$$\iint f(x,y)\nu(dy)\mu(dx) = \int g\,d\mu = u = \int f d\lambda.$$

Thus $f\in\mathcal{L}$, as claimed. **Q**

(iii) The expression of the ideas in the next section of the proof will go more smoothly if I introduce another term. Write $\mathcal{W}$ for $\{W : W\subseteq X\times Y,\ \chi W\in\mathcal{L}\}$. Then

(α) if W, $W'\in\mathcal{W}$ and $W\cap W'=\emptyset$, $W\cup W'\in\mathcal{W}$

by (i), because $\chi(W\cup W') = \chi W + \chi W'$,

(β) $\bigcup_{n\in\mathbb{N}} W_n \in \mathcal{W}$ whenever $\langle W_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence in $\mathcal{W}$

because $\langle \chi W_n\rangle_{n\in\mathbb{N}} \uparrow \chi W$, and we can use (ii).

It is also helpful to note that, for any $W\subseteq X\times Y$ and any $x\in X$, $\int \chi W(x,y)\nu(dy) = \nu W[\{x\}]$, at least whenever $W[\{x\}] = \{y : (x,y)\in W\}$ is measured by ν. Moreover, because λ is complete, a set $W\subseteq X\times Y$ belongs to Λ iff χW is λ-virtually measurable iff $\int \chi W\,d\lambda$ is defined in $[0,\infty]$, and in this case $\lambda W = \int \chi W\,d\lambda$.

(iv) Finally, we need to observe that, in appropriate circumstances, the difference of two members of $\mathcal{W}$ will belong to $\mathcal{W}$: if W, $W'\in\mathcal{W}$ and $W\subseteq W'$ and $\lambda W'<\infty$, then $W'\setminus W\in\mathcal{W}$. **P** We are supposing that $g(x) = \int \chi W(x,y)\nu(dy)$ and $g'(x) = \int \chi W'(x,y)\nu(dy)$ are defined for almost every x, and that $\int g\,d\mu = \lambda W$, $\int g' d\mu = \lambda W'$. Because $\lambda W'$ is finite, g' must be finite almost everywhere, and $D = \{x : x\in \operatorname{dom} g\cap\operatorname{dom} g',\ g'(x)<\infty\}$ is conegligible. Now, for any $x\in D$, both $g(x)$ and $g'(x)$ are finite, so

$$y\mapsto \chi(W'\setminus W)(x,y) = \chi W'(x,y) - \chi W(x,y)$$

is the difference of two integrable functions, and

$$\begin{aligned}\int \chi(W'\setminus W)(x,y)\nu(dy) &= \int \chi W'(x,y) - \chi W(x,y)\nu(dy)\\ &= \int \chi W'(x,y)\nu(dy) - \int \chi W(x,y)\nu(dy) = g'(x) - g(x).\end{aligned}$$

Accordingly

$$\iint \chi(W' \setminus W)(x,y)\nu(dy)\mu(dx) = \int g'(x) - g(x)\mu(dx) = \lambda W' - \lambda W = \lambda(W' \setminus W),$$

and $W' \setminus W$ belongs to $\mathcal{W}$. **Q**

(Of course the argument just above can be shortened by a few words if we allow ourselves to assume that μ and ν are totally finite, since then $g(x)$ and $g'(x)$ will be finite whenever they are defined; but the key idea, that the difference of integrable functions is integrable, is unchanged.)

(b) Now let us examine the class $\mathcal{W}$, assuming that μ and ν are totally finite.

(i) $E \times F \in \mathcal{W}$ for all $E \in \Sigma$, $F \in \mathrm{T}$. **P** $\lambda(E \times F) = \mu E \cdot \nu F$ (251J), and

$$\int \chi(E \times F)(x,y)\nu(dy) = \nu F\,\chi E(x)$$

for each x, so

$$\iint \chi(E \times F)(x,y)\nu(dy)\mu(dx) = \int (\nu F\,\chi E(x))\mu(dx) = \mu E \cdot \nu F$$

$$= \lambda(E \times F) = \int \chi(E \times F)d\lambda. \text{ **Q**}$$

(ii) Let $\mathcal{C}$ be $\{E \times F : E \in \Sigma,\ F \in \mathrm{T}\}$. Then $\mathcal{C}$ is closed under finite intersections (because $(E \times F) \cap (E' \times F') = (E \cap E') \times (F \cap F')$) and is included in $\mathcal{W}$. In particular, $X \times Y \in \mathcal{W}$. But this, together with (a-iv) and (a-iii-β) above, means that $\mathcal{W}$ is a Dynkin class (definition: 136A), so includes the σ-algebra of subsets of $X \times Y$ generated by $\mathcal{C}$, by the Monotone Class Theorem (136B); that is, $\mathcal{W} \supseteq \Sigma\widehat{\otimes}\mathrm{T}$ (definition: 251D).

(iii) Next, $W \in \mathcal{W}$ whenever $W \subseteq X \times Y$ is λ-negligible. **P** By 251Ib, there is a $V \in \Sigma\widehat{\otimes}\mathrm{T}$ such that $V \subseteq (X \times Y) \setminus W$ and $\lambda V = \lambda((X \times Y) \setminus W)$. Because $\lambda(X \times Y) = \mu X \cdot \nu Y$ is finite, $V' = (X \times Y) \setminus V$ is λ-negligible, and we have $W \subseteq V' \in \Sigma\widehat{\otimes}\mathrm{T}$. Consequently

$$0 = \lambda V' = \iint \chi V'(x,y)\nu(dy)\mu(dx).$$

But this means that

$$D = \{x : \int \chi V'(x,y)\nu(dy) \text{ is defined and equal to } 0\}$$

is conegligible. If $x \in D$, then we must have $\chi V'(x,y) = 0$ for ν-almost every y, that is, $V'[\{x\}]$ is negligible; in which case $W[\{x\}] \subseteq V'[\{x\}]$ also is negligible, and $\int \chi W(x,y)\nu(dy) = 0$. And this is true for every $x \in D$, so $\int \chi W(x,y)\nu(dy)$ is defined and equal to 0 for almost every x, and

$$\iint \chi W(x,y)\nu(dy)\mu(dx) = 0 = \lambda W,$$

as required. **Q**

(iv) It follows that $\Lambda \subseteq \mathcal{W}$. **P** If $W \in \Lambda$, then, by 251Ib again, there is a $V \in \Sigma\widehat{\otimes}\mathrm{T}$ such that $V \subseteq W$ and $\lambda V = \lambda W$, so that $\lambda(W \setminus V) = 0$. Now $V \in \mathcal{W}$ by (ii) and $W \setminus V \in \mathcal{W}$ by (iii), so $W \in \mathcal{W}$ by (a-iii-α). **Q**

(c) I return to the class $\mathcal{L}$.

(i) If $f \in \mathcal{L}$ and g is a $[0,\infty]$-valued function defined and equal to f λ-a.e., then $g \in \mathcal{L}$. **P** Set

$$W = (X \times Y) \setminus \{(x,y) : (x,y) \in \operatorname{dom} f \cap \operatorname{dom} g,\ f(x,y) = g(x,y)\},$$

so that $\lambda W = 0$. (Remember that λ is complete.) By (b), $\iint \chi W(x,y)\nu(dy)\mu(dx) = 0$, that is, $W[\{x\}]$ is ν-negligible for μ-almost every x. Let D be $\{x : x \in X,\ W[\{x\}] \text{ is } \nu\text{-negligible}\}$. Then D is μ-conegligible. If $x \in D$, then

$$W[\{x\}] = Y \setminus \{y : (x,y) \in \operatorname{dom} f \cap \operatorname{dom} g,\ f(x,y) = g(x,y)\}$$

is negligible, so that $\int f(x,y)\nu(dy) = \int g(x,y)\nu(dy)$ if either is defined. Thus the functions $x \mapsto \int f(x,y)\nu(dy)$, $x \mapsto \int g(x,y)\nu(dy)$ are equal almost everywhere, and

$$\iint g(x,y)\nu(dy)\mu(dx) = \iint f(x,y)\nu(dy)\mu(dx) = \int f d\lambda = \int g\,d\lambda,$$

so that $g \in \mathcal{L}$. **Q**

(ii) Now let f be any non-negative function such that $\int f d\lambda$ is defined in $[0,\infty]$. Then $f \in \mathcal{L}$. **P** For k, $n \in \mathbb{N}$ set

$$W_{nk} = \{(x,y) : (x,y) \in \operatorname{dom} f,\ f(x,y) \geq 2^{-n}k\}.$$

Because λ is complete and f is λ-virtually measurable and $\operatorname{dom} f$ is conegligible, every W_{nk} belongs to Λ, so $\chi W_{nk} \in \mathcal{L}$, by (b). Set $f_n = \sum_{k=1}^{4^n} 2^{-n}\chi W_{nk}$, so that

$$\begin{aligned} f_n(x,y) &= 2^{-n}k \text{ if } k \le 4^n \text{ and } 2^{-n}k \le f(x,y) < 2^{-n}(k+1), \\ &= 2^n \text{ if } f(x,y) \ge 2^n, \\ &= 0 \text{ if } (x,y) \in (X \times Y) \setminus \operatorname{dom} f. \end{aligned}$$

By (a-i), $f_n \in \mathcal{L}$ for every $n \in \mathbb{N}$, while $\langle f_n \rangle_{n\in\mathbb{N}}$ is non-decreasing, so $f' = \sup_{n\in\mathbb{N}} f_n \in \mathcal{L}$, by (a-ii). But $f =_{\text{a.e.}} f'$, so $f \in \mathcal{L}$, by (i) just above. **Q**

(iii) Finally, let f be any $[-\infty,\infty]$-valued function such that $\int f d\lambda$ is defined in $[-\infty,\infty]$. Then $\int f^+ d\lambda$, $\int f^- d\lambda$ are both defined and at most one is infinite. By (ii), both f^+ and f^- belong to $\mathcal{L}$. Set $g(x) = \int f^+(x,y)\nu(dy)$, $h(x) = \int f^-(x,y)\nu(dy)$ whenever these are defined; then $\int g\,d\mu = \int f^+ d\lambda$ and $\int h\,d\mu = \int f^- d\lambda$ are both defined in $[0,\infty]$.

Suppose first that $\int f^- d\lambda$ is finite. Then $\int h\,d\mu$ is finite, so h must be finite μ-almost everywhere; set

$$D = \{x : x \in \operatorname{dom} g \cap \operatorname{dom} h,\ h(x) < \infty\}.$$

For any $x \in D$, $\int f^+(x,y)\nu(dy)$ and $\int f^-(x,y)\nu(dy)$ are defined in $[0,\infty]$, and the latter is finite; so

$$\int f(x,y)\nu(dy) = \int f^+(x,y)\nu(dy) - \int f^-(x,y)\nu(dy) = g(x) - h(x)$$

is defined in $]-\infty,\infty]$. Because D is conegligible,

$$\begin{aligned} \iint f(x,y)\nu(dy)\mu(dx) &= \int g(x) - h(x)\mu(dx) = \int g\,d\mu - \int h\,d\mu \\ &= \int f^+ d\lambda - \int f^- d\lambda = \int f d\lambda, \end{aligned}$$

as required.

Thus we have the result when $\int f^- d\lambda$ is finite. Similarly, or by applying the argument above to $-f$, $\iint f(x,y)\nu(dy)\mu(dx) = \int f d\lambda$ if $\int f^+ d\lambda$ is finite.

Thus the theorem is proved, at least when μ and ν are totally finite.

(b*) The only point in the argument above where we needed to know anything special about the measures μ and ν was in part (b), when showing that $\Lambda \subseteq \mathcal{W}$. I now return to this point under the hypotheses of the theorem as stated, that ν is σ-finite and μ is either strictly localizable or complete and locally determined.

(i) It will be helpful to note that the completion $\hat\mu$ of μ (212C) is identical with its c.l.d. version $\tilde\mu$ (213E). **P** If μ is strictly localizable, then $\hat\mu = \tilde\mu$ by 213Ha. If μ is complete and locally determined, then $\hat\mu = \mu = \tilde\mu$ (212D, 213Hf). **Q**

(ii) Write $\Sigma^f = \{G : G \in \Sigma,\ \mu G < \infty\}$, $\mathrm{T}^f = \{H : H \in \mathrm{T},\ \nu H < \infty\}$. For $G \in \Sigma^f$, $H \in \mathrm{T}^f$ let μ_G, ν_H and $\lambda_{G\times H}$ be the subspace measures on G, H and $G \times H$ respectively; then $\lambda_{G\times H}$ is the c.l.d. product measure of μ_G and ν_H (251Q(ii-α)). Now $W \cap (G \times H) \in \mathcal{W}$ for every $W \in \Lambda$. **P** $W \cap (G \times H)$ belongs to the domain of $\lambda_{G\times H}$, so by (b) of this proof, applied to the totally finite measures μ_G and ν_H,

$$\begin{aligned} \lambda(W \cap (G \times H)) &= \lambda_{G\times H}(W \cap (G \times H)) \\ &= \int_G \int_H \chi(W \cap (G \times H))(x,y)\nu_H(dy)\mu_G(dx) \\ &= \int_G \int_Y \chi(W \cap (G \times H))(x,y)\nu(dy)\mu_G(dx) \end{aligned}$$

(because $\chi(W \cap (G \times H))(x,y) = 0$ if $y \in Y \setminus H$, so we can use 131E)

$$= \int_X \int_Y \chi(W \cap (G \times H))(x,y)\nu(dy)\mu(dx)$$

by 131E again, because $\int_Y \chi(W \cap (G \times H))(x,y)\nu(dy) = 0$ if $x \in X \setminus G$. So $W \cap (G \times H) \in \mathcal{W}$. **Q**

(iii) In fact, $W \in \mathcal{W}$ for every $W \in \Lambda$. **P** Remember that we are supposing that ν is σ-finite. Let $\langle Y_n \rangle_{n\in\mathbb{N}}$ be a non-decreasing sequence in T^f covering Y, and for each $n \in \mathbb{N}$ set $W_n = W \cap (X \times Y_n)$, $g_n(x) = \int \chi W_n(x,y)\nu(dy)$ whenever this is defined. For any $G \in \Sigma^f$,

$$\int_G g_n d\mu = \iint \chi(W \cap (G \times Y_n))(x,y)\nu(dy)\mu(dx)$$

is defined and equal to $\lambda(W \cap (G \times Y_n))$, by (ii). But this means, first, that $G \setminus \operatorname{dom} g_n$ is negligible, that is, that $\hat{\mu}(G \setminus \operatorname{dom} g_n) = 0$. Since this is so whenever μG is finite, $\tilde{\mu}(X \setminus \operatorname{dom} g_n) = 0$, and g_n is defined $\tilde{\mu}$-a.e.; but $\tilde{\mu} = \hat{\mu}$, so g_n is defined $\hat{\mu}$-a.e., that is, μ-a.e. (212Eb). Next, if we set $E_{na} = \{x : x \in \operatorname{dom} g_n,\ g_n(x) \geq a\}$ for $a \in \mathbb{R}$, then $E_{na} \cap G \in \hat{\Sigma}$ whenever $G \in \Sigma^f$, where $\hat{\Sigma}$ is the domain of $\hat{\mu}$; by the definition in 213D, E_{na} is measured by $\tilde{\mu} = \hat{\mu}$. As a is arbitrary, g_n is μ-virtually measurable (212Fa).

We can therefore speak of $\int g_n d\mu$. Now

$$\iint \chi W_n(x,y)\nu(dy)\mu(dx) = \int g_n d\mu = \sup_{G \in \Sigma^f} \int_G g_n$$

(213B, because μ is semi-finite)

$$= \sup_{G \in \Sigma^f} \lambda(W \cap (G \times Y_n)) = \lambda(W \cap (X \times Y_n))$$

by the definition in 251F. Thus $W \cap (X \times Y_n) \in \mathcal{W}$.

This is true for every $n \in \mathbb{N}$. Because $\langle Y_n \rangle_{n \in \mathbb{N}} \uparrow Y$, $W \in \mathcal{W}$, by (a-iii-β). **Q**

(iv) We can therefore return to part (c) of the argument above and conclude as before.

252C The theorem above is of course asymmetric, in that different hypotheses are imposed on the two factor measures μ and ν. If we want a 'symmetric' theorem we have to suppose that they are both σ-finite, as follows.

Corollary Let (X, Σ, μ) and (Y, T, ν) be two σ-finite measure spaces, and λ the c.l.d. product measure on $X \times Y$. If f is λ-integrable, then $\iint f(x,y)\nu(dy)\mu(dx)$ and $\iint f(x,y)\mu(dx)\nu(dy)$ are defined, finite and equal.

proof Since μ and ν are surely strictly localizable (211Lc), we can apply 252B from either side to conclude that

$$\iint f(x,y)\nu(dy)\mu(dx) = \int f\,d\lambda = \iint f(x,y)\mu(dx)\nu(dy).$$

252D So many applications of Fubini's theorem are to characteristic functions that I take a few lines to spell out the form which 252B takes in this case, as in parts (b)-(b*) of the proof there.

Corollary Let (X, Σ, μ) and (Y, T, ν) be measure spaces and λ the c.l.d. product measure on $X \times Y$. Suppose that ν is σ-finite and that μ is either strictly localizable or complete and locally determined.

(i) If $W \in \operatorname{dom} \lambda$, then $\int \nu^* W[\{x\}]\mu(dx)$ is defined in $[0, \infty]$ and equal to λW.

(ii) If ν is complete, we can write $\int \nu W[\{x\}]\mu(dx)$ in place of $\int \nu^* W[\{x\}]\mu(dx)$.

proof The point is just that $\int \chi W(x,y)\nu(dy) = \hat{\nu} W[\{x\}]$ whenever either is defined, where $\hat{\nu}$ is the completion of ν (212F). Now 252B tells us that

$$\lambda W = \iint \chi W(x,y)\nu(dy)\mu(dx) = \int \hat{\nu} W[\{x\}]\mu(dx).$$

We always have $\hat{\nu} W[\{x\}] = \nu^* W[\{x\}]$, by the definition of $\hat{\nu}$ (212C); and if ν is complete, then $\hat{\nu} = \nu$ so $\lambda W = \int \nu W[\{x\}]\mu(dx)$.

252E Corollary Let (X, Σ, μ) and (Y, T, ν) be measure spaces, with c.l.d. product $(X \times Y, \Lambda, \lambda)$. Suppose that ν is σ-finite and that μ is has locally determined negligible sets (213I). Then if f is a Λ-measurable real-valued function defined on a subset of $X \times Y$, $y \mapsto f(x,y)$ is ν-virtually measurable for μ-almost every $x \in X$.

proof Let $\tilde{f}$ be a Λ-measurable extension of f to a real-valued function defined everywhere in $X \times Y$ (121I), and set $\tilde{f}_x(y) = \tilde{f}(x,y)$ for all $x \in X$, $y \in Y$,

$$D = \{x : x \in X,\ \tilde{f}_x \text{ is } \nu\text{-virtually measurable}\}.$$

If $G \in \Sigma$ and $\mu G < \infty$, then $G \setminus D$ is negligible. **P** Let $\langle Y_n \rangle_{n \in \mathbb{N}}$ be a non-decreasing sequence of sets of finite measure covering Y respectively, and set

$$\begin{aligned} \tilde{f}_n(x,y) &= \tilde{f}(x,y) \text{ if } x \in G,\ y \in Y_n \text{ and } |\tilde{f}(x,y)| \leq n, \\ &= 0 \text{ for other } x \in X \times Y. \end{aligned}$$

Then each $\tilde{f}_n$ is λ-integrable, being bounded and Λ-measurable and zero off $G \times Y_n$. Consequently, setting $\tilde{f}_{nx}(y) = \tilde{f}_n(x,y)$,

$$\int(\int \tilde f_{nx}d\nu)\mu(dx) \text{ exists} = \int \tilde f_n d\lambda.$$

But this surely means that $\tilde f_{nx}$ is ν-integrable, therefore ν-virtually measurable, for almost every $x \in X$. Set

$$D_n = \{x : x \in X,\ \tilde f_{nx} \text{ is } \nu\text{-virtually measurable}\};$$

then every D_n is μ-conegligible, so $\bigcap_{n\in\mathbb{N}} D_n$ is conegligible. But for any $x \in G \cap \bigcap_{n\in\mathbb{N}} D_n$, $\tilde f_x = \lim_{n\to\infty} \tilde f_{nx}$ is ν-virtually measurable. Thus $G \setminus D \subseteq X \setminus \bigcap_{n\in\mathbb{N}} D_n$ is negligible. **Q**

This is true whenever $\mu G < \infty$. As G is arbitrary and μ has locally determined negligible sets, D is conegligible. But, for any $x \in D$, $y \mapsto f(x, y)$ is a restriction of $\tilde f_x$ and must be ν-virtually measurable.

252F As a further corollary we can get some useful information about the c.l.d. product measure for arbitrary measure spaces.

Corollary Let (X, Σ, μ) and (Y, T, ν) be two measure spaces, λ the c.l.d. product measure on $X \times Y$, and Λ its domain. Let $W \in \Lambda$ be such that the vertical section $W[\{x\}]$ is ν-negligible for μ-almost every $x \in X$. Then $\lambda W = 0$.

proof Take $E \in \Sigma$, $F \in \mathrm{T}$ of finite measure. Let $\lambda_{E\times F}$ be the subspace measure on $E \times F$. By 251Q(ii-α) again, this is just the product of the subspace measures μ_E and ν_F. We know that $W \cap (E \times F)$ is measured by $\lambda_{E\times F}$. At the same time, the vertical section $(W \cap (E \times F))[\{x\}] = W[\{x\}] \cap F$ is ν_F-negligible for μ_E-almost every $x \in E$. Applying 252B to μ_E and ν_F and $\chi(W \cap (E \times F))$,

$$\lambda(W \cap (E \times F)) = \lambda_{E\times F}(W \cap (E \times F)) = \int_E \nu_F(W[\{x\}] \cap F)\mu_E(dx) = 0.$$

But looking at the definition in 251F, we see that this means that $\lambda W = 0$, as claimed.

252G Theorem 252B and its corollaries depend on the factor measures μ and ν belonging to restricted classes. There is a partial result which applies to all c.l.d. product measures, as follows.

Tonelli's theorem Let (X, Σ, μ) and (Y, T, ν) be measure spaces, and $(X \times Y, \Lambda, \lambda)$ their c.l.d. product. Let f be a Λ-measurable $[-\infty, \infty]$-valued function defined on a member of Λ, and suppose that either $\iint |f(x, y)|\mu(dx)\nu(dy)$ or $\iint |f(x, y)|\nu(dy)\mu(dx)$ exists in $\mathbb{R}$. Then f is λ-integrable.

proof Because the construction of the product measure is symmetric in the two factors, it is enough to consider the case in which $\iint |f(x, y)|\nu(dy)\mu(dx)$ is defined and finite, as the same ideas will surely deal with the other case also.

(a) The first step is to check that f is defined and finite λ-a.e. **P** Set $W = \{(x, y) : (x, y) \in \operatorname{dom} f,\ f(x, y) \text{ is finite}\}$. Then $W \in \Lambda$. The hypothesis

$$\iint |f(x, y)|\nu(dy)\mu(dx) \text{ is defined and finite}$$

includes the assertion

$$\int |f(x, y)|\nu(dy) \text{ is defined and finite for } \mu\text{-almost every } x,$$

which implies that

$$\text{for } \mu\text{-almost every } x,\ f(x, y) \text{ is defined and finite for } \nu\text{-almost every } y;$$

that is, that

$$\text{for } \mu\text{-almost every } x,\ W[\{x\}] \text{ is } \nu\text{-conegligible.}$$

But by 252F this implies that $(X \times Y) \setminus W$ is λ-negligible, as required. **Q**

(b) Let h be any non-negative λ-simple function such that $h \le |f|$ λ-a.e. Then $\int h$ cannot be greater than $\iint |f(x, y)|\nu(dy)\mu(dx)$. **P** Set

$$W = \{(x, y) : (x, y) \in \operatorname{dom} f,\ h(x, y) \le |f(x, y)|\}, \quad h' = h \times \chi W;$$

then h' is a simple function and $h' =_{\text{a.e.}} h$. Express h' as $\sum_{i=0}^n a_i \chi W_i$ where $a_i \ge 0$ and $\lambda W_i < \infty$ for each i. Let $\epsilon > 0$. For each $i \le n$ there are $E_i \in \Sigma$, $F_i \in \mathrm{T}$ such that $\mu E_i < \infty$, $\nu F_i < \infty$ and $\lambda(W_i \cap (E_i \times F_i)) \ge \lambda W_i - \epsilon$. Set $E = \bigcup_{i\le n} E_i$ and $F = \bigcup_{i\le n} F_i$. Consider the subspace measures μ_E and ν_F and their product $\lambda_{E\times F}$ on $E \times F$; then $\lambda_{E\times F}$ is the subspace measure on $E \times F$ defined from λ (251Q(ii-α) once more). Accordingly, applying 252B to the product $\mu_E \times \nu_F$,

$$\int_{E\times F} h'\, d\lambda = \int_{E\times F} h'\, d\lambda_{E\times F} = \int_E \int_F h'(x, y)\nu_F(dy)\mu_E(dx).$$

For any x, we know that $h'(x, y) \le |f(x, y)|$ whenever $f(x, y)$ is defined. So we can be sure that

$$\int_F h'(x,y)\nu_F(dy) = \int h'(x,y)\chi F(y)\nu(dy) \le \int |f(x,y)|\nu(dy)$$

at least whenever $\int_F h'(x,y)\nu_F(dy)$ and $\int |f(x,y)|\nu(dy)$ are both defined, which is the case for almost every $x \in E$. Consequently

$$\begin{aligned}\int_{E\times F} h'\,d\lambda &= \int_E \int_F h'(x,y)\nu_F(dy)\mu_E(dx)\\ &\le \int_E \int |f(x,y)|\nu(dy)\mu(dx) \le \iint |f(x,y)|\nu(dy)\mu(dx).\end{aligned}$$

On the other hand,

$$\begin{aligned}\int h'\,d\lambda - \int_{E\times F} h'\,d\lambda &= \sum_{i=0}^{n} a_i\lambda(W_i \setminus (E\times F))\\ &\le \sum_{i=0}^{n} a_i\lambda(W_i \setminus (E_i\times F_i)) \le \epsilon\sum_{i=0}^{n} a_i.\end{aligned}$$

So

$$\int h\,d\lambda = \int h'd\lambda \le \iint |f(x,y)|\nu(dy)\mu(dx) + \epsilon\textstyle\sum_{i=0}^n a_i.$$

As ϵ is arbitrary, $\int h\,d\lambda \le \iint |f(x,y)|\nu(dy)\mu(dx)$, as claimed. **Q**

(c) This is true whenever h is a λ-simple function less than or equal to $|f|$ λ-a.e. But $|f|$ is Λ-measurable and λ is semi-finite (251Ic), so this is enough to ensure that $|f|$ is λ-integrable (213B), which (because f is supposed to be Λ-measurable) in turn implies that f is λ-integrable.

252H Corollary Let (X,Σ,μ) and (Y,T,ν) be σ-finite measure spaces, λ the c.l.d. product measure on $X\times Y$, and Λ its domain. Let f be a Λ-measurable $[-\infty,\infty]$-valued function defined on a member of Λ. Then if one of

$$\int_{X\times Y} |f(x,y)|\lambda(d(x,y)),\quad \int_Y\int_X |f(x,y)|\mu(dx)\nu(dy),\quad \int_X\int_Y |f(x,y)|\nu(dy)\mu(dx)$$

exists in $\mathbb{R}$, so do the other two, and in this case

$$\int_{X\times Y} f(x,y)\lambda(d(x,y)) = \int_Y\int_X f(x,y)\mu(dx)\nu(dy) = \int_X\int_Y f(x,y)\nu(dy)\mu(dx).$$

proof (a) Suppose that $\int |f|d\lambda$ is finite. Because both μ and ν are σ-finite, 252B tells us that

$$\iint |f(x,y)|\mu(dx)\nu(dy),\quad \iint |f(x,y)|\nu(dy)\mu(dx)$$

both exist and are equal to $\int |f|d\lambda$, while

$$\iint f(x,y)\mu(dx)\nu(dy),\quad \iint f(x,y)\nu(dy)\mu(dx)$$

both exist and are equal to $\int f d\lambda$.

(b) Now suppose that $\iint |f(x,y)|\nu(dy)\mu(dx)$ exists in $\mathbb{R}$. Then 252G tells us that $|f|$ is λ-integrable, so we can use (a) to complete the argument. Exchanging the coordinates, the same argument applies if $\iint |f(x,y)|\mu(dx)\nu(dy)$ exists in $\mathbb{R}$.

252I Corollary Let (X,Σ,μ) and (Y,T,ν) be measure spaces, λ the c.l.d. product measure on $X\times Y$, and Λ its domain. Take $W\in\Lambda$. If either of the integrals

$$\int \mu^*W^{-1}[\{y\}]\nu(dy),\quad \int \nu^*W[\{x\}]\mu(dx)$$

exists and is finite, then $\lambda W < \infty$.

proof Apply 252G with $f = \chi W$, remembering that

$$\mu^*W^{-1}[\{y\}] = \int \chi W(x,y)\mu(dx),\quad \nu^*W[\{x\}] = \int \chi W(x,y)\nu(dy)$$

whenever the integrals are defined, as in the proof of 252D.

252J Remarks 252H is the basic form of Fubini's theorem; it is not a coincidence that most authors avoid non-σ-finite spaces in this context. The next two examples exhibit some of the difficulties which can arise if we leave the familiar territory of more-or-less Borel measurable functions on σ-finite spaces. The first is a classic.

252K Example Let (X,Σ,μ) be $[0,1]$ with Lebesgue measure, and let (Y,T,ν) be $[0,1]$ with counting measure.

(a) Consider the set

$$W=\{(t,t):t\in[0,1]\}\subseteq X\times Y.$$

We observe that W is expressible as

$$\bigcap_{n\in\mathbb{N}}\bigcup_{k=0}^{n}[\tfrac{k}{n+1},\tfrac{k+1}{n+1}]\times[\tfrac{k}{n+1},\tfrac{k+1}{n+1}]\in\Sigma\widehat{\otimes}\mathrm{T}.$$

If we look at the sections

$$W^{-1}[\{t\}]=W[\{t\}]=\{t\}$$

for $t\in[0,1]$, we have

$$\iint\chi W(x,y)\mu(dx)\nu(dy)=\int\mu W^{-1}[\{y\}]\nu(dy)=\int 0\,\nu(dy)=0,$$

$$\iint\chi W(x,y)\nu(dy)\mu(dx)=\int\nu W[\{x\}]\mu(dx)=\int 1\,\mu(dx)=1,$$

so the two repeated integrals differ. It is therefore not generally possible to reverse the order of repeated integration, even for a non-negative measurable function in which both repeated integrals exist and are finite.

(b) Because the set W of part (a) actually belongs to $\Sigma\widehat{\otimes}\mathrm{T}$, we know that it is measured by the c.l.d. product measure λ, and 252F (applied with the coordinates reversed) tells us that $\lambda W=0$.

(c) It is in fact easy to give a full description of λ.

(i) The point is that a set $W\subseteq[0,1]\times[0,1]$ belongs to the domain Λ of λ iff every horizontal section $W^{-1}[\{y\}]$ is Lebesgue measurable. **P** (α) If $W\in\Lambda$, then, for every $b\in[0,1]$, $\lambda([0,1]\times\{b\})$ is finite, so $W\cap([0,1]\times\{b\})$ is a set of finite measure, and

$$\lambda(W\cap([0,1]\times\{b\}))=\int\mu(W\cap([0,1]\times\{b\}))^{-1}[\{y\}]\nu(dy)=\mu W^{-1}[\{b\}]$$

by 252D, because μ is σ-finite, ν is both strictly localizable and complete and locally determined, and

$$\begin{aligned}(W\cap([0,1]\times\{b\}))^{-1}[\{y\}]&=W^{-1}[\{b\}]\text{ if }y=b,\\&=\emptyset\text{ otherwise.}\end{aligned}$$

As b is arbitrary, every horizontal section of W is measurable. (β) If every horizontal section of W is measurable, let $F\subseteq[0,1]$ be any set of finite measure for ν; then F is finite, so

$$W\cap([0,1]\times F)=\bigcup_{y\in F}W^{-1}[\{y\}]\times\{y\}\in\Sigma\widehat{\otimes}\mathrm{T}\subseteq\Lambda.$$

But it follows that W itself belongs to Λ, by 251H. **Q**

(ii) Now some of the same calculations show that for every $W\in\Lambda$,

$$\lambda W=\sum_{y\in[0,1]}\mu W^{-1}[\{y\}].$$

P For any finite $F\subseteq[0,1]$,

$$\begin{aligned}\lambda(W\cap([0,1]\times F))&=\int\mu(W\cap([0,1]\times F))^{-1}[\{y\}]\nu(dy)\\&=\int_F\mu W^{-1}[\{y\}]\nu(dy)=\sum_{y\in F}\mu W^{-1}[\{y\}].\end{aligned}$$

So

$$\lambda W=\sup_{F\subseteq[0,1]\text{ is finite}}\sum_{y\in F}\mu W^{-1}[\{y\}]=\sum_{y\in[0,1]}\mu W^{-1}[\{y\}].\ \mathbf{Q}$$

252L Example For the second example, I turn to a problem that can arise if we neglect to check that a function is measurable as a function of two variables.

Let $(X,\Sigma,\mu)=(Y,\mathrm{T},\nu)$ be ω_1, the first uncountable ordinal (2A1Fc), with the countable-cocountable measure (211R). Set

$$W=\{(\xi,\eta):\xi\le\eta<\omega_1\}\subseteq X\times Y.$$

Then all the horizontal sections $W^{-1}[\{\eta\}] = \{\xi : \xi \le \eta\}$ are countable, so

$$\int \mu W^{-1}[\{\eta\}]\nu(d\eta) = \int 0\,\nu(d\eta) = 0,$$

while all the vertical sections $W[\{\xi\}] = \{\eta : \xi \le \eta < \omega_1\}$ are cocountable, so

$$\int \nu W[\{\xi\}]\mu(d\xi) = \int 1\,\mu(d\xi) = 1.$$

Because the two repeated integrals are different, they cannot both be equal to the measure of W, and the sole resolution is to say that W is not measured by the product measure.

252M Remark A third kind of difficulty in the formula

$$\iint f(x,y)dxdy = \iint f(x,y)dydx$$

can arise even on probability spaces with $\Sigma\widehat{\otimes}\mathrm{T}$-measurable real-valued functions defined everywhere if we neglect to check that f is integrable with respect to the product measure. In 252H, we do need the hypothesis that one of

$$\int_{X\times Y} |f(x,y)|\lambda(d(x,y)), \quad \int_Y \int_X |f(x,y)|\mu(dx)\nu(dy), \quad \int_X \int_Y |f(x,y)|\nu(dy)\mu(dx)$$

is finite. For examples to show this, see 252Xf and 252Xg.

252N Integration through ordinate sets I: Proposition Let (X,Σ,μ) be a complete locally determined measure space, and λ the c.l.d. product measure on $X\times\mathbb{R}$, where $\mathbb{R}$ is given Lebesgue measure; write Λ for the domain of λ. For any $[0,\infty]$-valued function f defined on a conegligible subset of X, write Ω_f, Ω'_f for the **ordinate sets**

$$\Omega_f = \{(x,a) : x \in \operatorname{dom} f,\ 0 \le a \le f(x)\} \subseteq X\times\mathbb{R},$$

$$\Omega'_f = \{(x,a) : x \in \operatorname{dom} f,\ 0 \le a < f(x)\} \subseteq X\times\mathbb{R}.$$

Then

$$\lambda\Omega_f = \lambda\Omega'_f = \int f d\mu$$

in the sense that if one of these is defined in $[0,\infty]$, so are the other two, and they are equal.

proof (a) If $\Omega_f \in \Lambda$, then

$$\int f(x)\mu(dx) = \int \nu\{y : (x,y) \in \Omega_f\}\mu(dx) = \lambda\Omega_f$$

by 252D, writing μ for Lebesgue measure, because f is defined almost everywhere. Similarly, if $\Omega'_f \in \Lambda$,

$$\int f(x)\mu(dx) = \int \nu\{y : (x,y) \in \Omega'_f\}\mu(dx) = \lambda\Omega'_f.$$

(b) If $\int f d\mu$ is defined, then f is μ-virtually measurable, therefore measurable (because μ is complete); again because μ is complete, $\operatorname{dom} f \in \Sigma$. So

$$\Omega'_f = \bigcup_{q\in\mathbb{Q},q>0}\{x : x \in \operatorname{dom} f,\ f(x) > q\}\times[0,q],$$

$$\Omega_f = \bigcap_{n\ge1}\bigcup_{q\in\mathbb{Q},q>0}\{x : x \in \operatorname{dom} f,\ f(x) \ge q - \frac{1}{n}\}\times[0,q]$$

belong to Λ, so that $\lambda\Omega_f$ and $\lambda\Omega'_f$ are defined. Now both are equal to $\int f d\mu$, by (a).

252O Integration through ordinate sets II: Proposition Let (X,Σ,μ) be a measure space, and f a non-negative μ-virtually measurable function defined on a conegligible subset of X. Then

$$\int f d\mu = \int_0^\infty \mu^*\{x : x \in \operatorname{dom} f,\ f(x) \ge t\}dt = \int_0^\infty \mu^*\{x : x \in \operatorname{dom} f,\ f(x) > t\}dt$$

in $[0,\infty]$, where the integrals $\int \dots dt$ are taken with respect to Lebesgue measure.

proof Completing μ does not change the integral of f or the outer measure μ^* (212Fb, 212Ea), so we may suppose that μ is complete, in which case $\operatorname{dom} f$ and f will be measurable. For n, $k \in \mathbb{N}$ set $E_{nk} = \{x : x \in \operatorname{dom} f,\ f(x) > 2^{-n}k\}$, $g_n(x) = 2^{-n}\sum_{k=1}^{4^n}\chi E_{nk}$. Then $\langle g_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence of measurable functions converging to f at every point of $\operatorname{dom} f$, so $\int f d\mu = \lim_{n\to\infty}\int g_n d\mu$ and $\mu\{x : f(x) > t\} = \lim_{n\to\infty}\mu\{x : g_n(x) > t\}$ for every $t \ge 0$; consequently

$$\int_0^\infty \mu\{x : f(x) > t\}dt = \lim_{n\to\infty}\int_0^\infty \mu\{x : g_n(x) > t\}dt.$$

On the other hand, $\mu\{x : g_n(x) > t\} = \mu E_{nk}$ if $1 \le k \le 4^n$ and $2^{-n}(k-1) \le t < 2^{-n}k$, 0 if $t \ge 2^n$, so that

$$\int_0^\infty \mu\{x : g_n(x) > t\}dt = \sum_{k=1}^{4^n} 2^{-n}\mu E_{nk} = \int g_n \, d\mu,$$

for every $n \in \mathbb{N}$. So $\int_0^\infty \mu\{x : f(x) > t\}dt = \int f d\mu$.

Now $\mu\{x : f(x) \ge t\} = \mu\{x : f(x) > t\}$ for almost all t. **P** Set $C = \{t : \mu\{x : f(x) > t\} < \infty\}$, $h(t) = \mu\{x : f(x) > t\}$ for $t \in C$. If C is not empty, $h : C \to [0, \infty[$ is monotonic, so is continuous almost everywhere in C (222A). But at any point of $C \setminus \{\inf C\}$ at which h is continuous,

$$\mu\{x : f(x) \ge t\} = \lim_{s \uparrow t} \mu\{x : f(x) > s\} = \mu\{x : f(x) > t\}.$$

So we have the result, since $\mu\{x : f(x) \ge t\} = \mu\{x : f(x) > t\} = \infty$ for any $t \in [0, \infty[\setminus C$. **Q**

Accordingly $\int_0^\infty \mu\{x : f(x) \ge t\}dt$ is also equal to $\int f d\mu$.

***252P** If we work through the ideas of 252B for $\Sigma\widehat{\otimes}\mathrm{T}$-measurable functions, we get the following, which is sometimes useful.

Proposition Let (X, Σ, μ) be a measure space, and (Y, T, ν) a σ-finite measure space. Then for any $\Sigma\widehat{\otimes}\mathrm{T}$-measurable function $f : X \times Y \to [0, \infty]$, $x \mapsto \int f(x, y)\nu(dy) : X \to [0, \infty]$ is Σ-measurable; and if μ is semi-finite, $\iint f(x, y)\nu(dy)\mu(dx) = \int f d\lambda$, where λ is the c.l.d. product measure on $X \times Y$.

proof (a) Let $\langle Y_n \rangle_{n\in\mathbb{N}}$ be a non-decreasing sequence of subsets of Y of finite measure with union Y. Set

$$\begin{aligned}\mathcal{A} = \{W : W \subseteq X \times Y,\ & W[\{x\}] \in \mathrm{T} \text{ for every } x \in X, \\ & x \mapsto \nu(Y_n \cap W[\{x\}]) \text{ is } \Sigma\text{-measurable for every } n \in \mathbb{N}\}.\end{aligned}$$

Then $\mathcal{A}$ is a Dynkin class of subsets of $X \times Y$ including $\{E \times F : E \in \Sigma,\ F \in \mathrm{T}\}$, so includes $\Sigma\widehat{\otimes}\mathrm{T}$, by the Monotone Class Theorem again (136B).

This means that if $W \in \Sigma\widehat{\otimes}\mathrm{T}$, then

$$\mu W[\{x\}] = \sup_{n\in\mathbb{N}} \nu(Y_n \cap W[\{x\}])$$

is defined for every $x \in X$ and is a Σ-measurable function of x.

(b) Now, for n, $k \in \mathbb{N}$, set

$$W_{nk} = \{(x, y) : f(x, y) \ge 2^{-n}k\}, \quad g_n = \sum_{k=1}^{4^n} 2^{-n}\chi W_{nk}.$$

Then if we set

$$h_n(x) = \int g_n(x, y)\nu(dy) = \sum_{k=1}^{4^n} 2^{-n}\nu W_{nk}[\{x\}]$$

for $n \in \mathbb{N}$ and $x \in X$, $h_n : X \to [0, \infty]$ is Σ-measurable, and

$$\lim_{n\to\infty} h_n(x) = \int (\lim_{n\to\infty} g_n(x, y))\nu(dy) = \int f(x, y)\nu(dy)$$

for every x, because $\langle g_n(x, y) \rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence with limit $f(x, y)$ for all $x \in X$, $y \in Y$. So $x \mapsto \int f(x, y)\nu(dy)$ is defined everywhere in X and is Σ-measurable.

(c) If $E \subseteq X$ is measurable and has finite measure, then $\int_E \int f(x, y)\nu(dy)\mu(dx) = \int_{E\times Y} f d\lambda$, applying 252B to the product of the subspace measure μ_E and ν (and using 251Q to check that the product of μ_E and ν is the subspace measure on $E \times Y$). Now if λW is defined and finite, there must be a non-decreasing sequence $\langle E_n \rangle_{n\in\mathbb{N}}$ of subsets of X of finite measure such that $\lambda W = \sup_{n\in\mathbb{N}} \lambda(W \cap (E_n \times Y))$, so that $W \setminus \bigcup_{n\in\mathbb{N}}(E_n \times Y)$ is negligible, and

$$\int_W f d\lambda = \lim_{n\to\infty} \int_{W\cap(E_n\times Y)} f d\lambda$$

(by B.Levi's theorem applied to $\langle f \times \chi(W \cap (E_n \times Y)) \rangle_{n\in\mathbb{N}}$)

$$\le \lim_{n\to\infty} \int_{E_n\times Y} f d\lambda = \lim_{n\to\infty} \int_{E_n} \int f(x, y)\nu(dy)\mu(dx)$$

$$\le \iint f(x, y)\nu(dy)\mu(dx).$$

By 213B once more,

$$\int f d\lambda = \sup_{\lambda W<\infty} \int_W f d\lambda \le \iint f(x, y)\nu(dy)\mu(dx).$$

But also, if μ is semi-finite,

$$\iint f(x,y)\nu(dy)\mu(dx) = \sup_{\mu E<\infty} \int_E \int f(x,y)\nu(dy)\mu(dx) \le \int f d\lambda,$$

so $\int f d\lambda = \iint f(x,y)\nu(dy)\mu(dx)$, as claimed.

252Q The volume of a ball We now have all the essential machinery to perform a little calculation which is, I suppose, desirable simply as general knowledge: the volume of the unit ball $\{x : \|x\| \le 1\} = \{(\xi_1,\dots,\xi_r) : \sum_{i=1}^r \xi_i^2 \le 1\}$ in $\mathbb{R}^r$. In fact, from a theoretical point of view, I think we could very nearly just call it β_r and leave it at that; but since there is a general formula in terms of $\beta_2 = \pi$ and factorials, it seems shameful not to present it. The calculation has nothing to do with Lebesgue integration, and I could dismiss it as mere advanced calculus; but since only a minority of mathematicians are now taught calculus to this level with reasonable rigour before being introduced to the Lebesgue integral, I do not doubt that many readers, like myself, missed some of the subtleties involved. I therefore take the space to spell the details out in the style used elsewhere in this volume, recognising that the machinery employed is a good deal more elaborate than is really necessary for this result.

(a) The first basic fact we need is that, for any $n \ge 1$,

$$\begin{aligned} I_n = \int_{-\pi/2}^{\pi/2} \cos^n t\,dt &= \frac{(2k)!}{(2^k k!)^2}\pi \text{ if } n = 2k \text{ is even,} \\ &= 2\frac{(2^k k!)^2}{(2k+1)!} \text{ if } n = 2k+1 \text{ is odd.} \end{aligned}$$

P For $n = 0$, of course,

$$I_0 = \int_{-\pi/2}^{\pi/2} 1\,dt = \pi = \frac{0!}{(2^0 0!)^2}\pi,$$

while for $n = 1$ we have

$$I_1 = \sin\tfrac{\pi}{2} - \sin(-\tfrac{\pi}{2}) = 2 = 2\frac{(2^0 0!)^2}{1!},$$

using the Fundamental Theorem of Calculus (225L) and the fact that $\sin' = \cos$ is bounded. For the inductive step to $n+1 \ge 2$, we can use integration by parts (225F):

$$\begin{aligned} I_{n+1} &= \int_{-\pi/2}^{\pi/2} \cos t \cos^n t\,dt \\ &= \sin\frac{\pi}{2}\cos^n\frac{\pi}{2} - \sin(-\frac{\pi}{2})\cos^n(-\frac{\pi}{2}) + \int_{-\pi/2}^{\pi/2} \sin t \cdot n\cos^{n-1} t \cdot \sin t\,dt \\ &= n\int_{-\pi/2}^{\pi/2} (1-\cos^2 t)\cos^{n-1} t\,dt = n(I_{n-1} - I_{n+1}), \end{aligned}$$

so that $I_{n+1} = \frac{n}{n+1} I_{n-1}$. Now the given formulae follow by an easy induction. **Q**

(b) The next result is that, for any $n \in \mathbb{N}$ and any $a \ge 0$,

$$\int_{-a}^{a} (a^2 - s^2)^{n/2} ds = I_{n+1} a^{n+1}.$$

P Of course this is an integration by substitution; but the singularity of the integrand at $s = \pm a$ complicates the issue slightly. I offer the following argument. If $a = 0$ the result is trivial; take $a > 0$. For $-a \le b \le a$, set $F(b) = \int_{-a}^{b} (a^2 - s^2)^{n/2} ds$. Because the integrand is continuous, $F'(b)$ exists and is equal to $(a^2 - b^2)^{n/2}$ for $-a < b < a$ (222H). Set $G(t) = F(a\sin t)$; then G is continuous and

$$G'(t) = aF'(a\sin t)\cos t = a^{n+1}\cos^{n+1} t$$

for $-\frac{\pi}{2} < t < \frac{\pi}{2}$. Consequently

$$\int_{-a}^{a} (a^2 - s^2)^{n/2} ds = F(a) - F(-a) = G(\frac{\pi}{2}) - G(-\frac{\pi}{2}) = \int_{-\pi/2}^{\pi/2} G'(t)dt$$

(by 225L, as before)

$$= a^{n+1} I_{n+1},$$

as required. **Q**

(c) Now at last we are ready for the balls $B_r = \{x : x \in \mathbb{R}^r, \|x\| \le 1\}$. Let μ_r be Lebesgue measure on $\mathbb{R}^r$, and set $\beta_r = I_1 I_2 \ldots I_r$ for $r \ge 1$. I claim that, writing

$$B_r(a) = \{x : x \in \mathbb{R}^r, \|x\| \le a\},$$

we have $\mu_r(B_r(a)) = \beta_r a^r$ for every $a \ge 0$. **P** Induce on r. For $r = 1$ we have $\beta_1 = 2$, $B_1(a) = [-a, a]$, so the result is trivial. For the inductive step to $r+1$, we have

$$\mu_{r+1} B_{r+1}(a) = \int \mu_r \{x : (x,t) \in B_{r+1}(a)\} dt$$

(putting 251N and 252D together, and using the fact that $B_{r+1}(a)$ is closed, therefore measurable)

$$= \int_{-a}^{a} \mu_r B_r(\sqrt{a^2 - t^2}) dt$$

(because $(x,t) \in B_{r+1}(a)$ iff $|t| \le a$ and $\|x\| \le \sqrt{a^2 - t^2}$)

$$= \int_{-a}^{a} \beta_r (a^2 - t^2)^{r/2} dt$$

(by the inductive hypothesis)

$$= \beta_r a^{r+1} I_{r+1}$$

(by (b) above)

$$= \beta_{r+1} a^{r+1}$$

(by the definition of β_{r+1}). Thus the induction continues. **Q**

(d) In particular, the r-dimensional Lebesgue measure of the r-dimensional ball $B_r = B_r(1)$ is just $\beta_r = I_1 \ldots I_r$. Now an easy induction on k shows that

$$\beta_r = \frac{1}{k!}\pi^k \text{ if } r = 2k \text{ is even,}$$

$$= \frac{2^{2k+1}k!}{(2k+1)!}\pi^k \text{ if } r = 2k+1 \text{ is odd.}$$

(e) Note that in part (c) of the proof we saw that $\{x : x \in \mathbb{R}^r, \|x\| \le a\}$ has measure $\beta_r a^r$ for every $a \ge 0$.

The formulae here are consistent with the assignation $\beta_0 = 1$; which corresponds to saying that $\mathbb{R}^0 = \{\emptyset\}$, that $\mu_0 \mathbb{R}^0 = 1$ and that $B_0 = \{\emptyset\}$. Taking $\mu_0 \mathbb{R}^0$ to be 1 is itself consistent with the idea that, following 251N, the product measure $\mu_0 \times \mu_r$ ought to match μ_{0+r} on $\mathbb{R}^{0+r}$.

252R Complex-valued functions It is easy to apply the results of 252B-252I above to complex-valued functions, by considering their real and imaginary parts. Specifically:

(a) Let (X, Σ, μ) and (Y, T, ν) be measure spaces, and λ the c.l.d. product measure on $X \times Y$. Suppose that ν is σ-finite and that μ is either strictly localizable or complete and locally determined. Let f be a λ-integrable complex-valued function. Then $\iint f(x,y)\nu(dy)\mu(dx)$ is defined and equal to $\int f d\lambda$.

(b) Let (X, Σ, μ) and (Y, T, ν) be measure spaces, λ the c.l.d. product measure on $X \times Y$, and Λ its domain. Let f be a Λ-measurable complex-valued function defined on a member of Λ, and suppose that either $\iint |f(x,y)|\mu(dx)\nu(dy)$ or $\iint |f(x,y)|\nu(dy)\mu(dx)$ is defined and finite. Then f is λ-integrable.

(c) Let (X, Σ, μ) and (Y, T, ν) be σ-finite measure spaces, λ the c.l.d. product measure on $X \times Y$, and Λ its domain. Let f be a Λ-measurable complex-valued function defined on a member of Λ. Then if one of

$$\int_{X \times Y} |f(x,y)|\lambda(d(x,y)), \quad \int_Y \int_X |f(x,y)|\mu(dx)\nu(dy), \quad \int_X \int_Y |f(x,y)|\nu(dy)\mu(dx)$$

exists in $\mathbb{R}$, so do the other two, and in this case

$$\int_{X \times Y} f(x,y)\lambda(d(x,y)) = \int_Y \int_X f(x,y)\mu(dx)\nu(dy) = \int_X \int_Y f(x,y)\nu(dy)\mu(dx).$$

252X Basic exercises (a) Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and λ the c.l.d. product measure on $X\times Y$. Let f be a λ-integrable real-valued function such that $\int_{E\times F} f=0$ whenever $E\in\Sigma$, $F\in\mathrm{T}$, $\mu E<\infty$ and $\nu F<\infty$. Show that $f=0$ λ-a.e. (*Hint*: use 251Ie to show that $\int_W f=0$ whenever $\lambda W<\infty$.)

(b) Let f, $g:\mathbb{R}\to\mathbb{R}$ be two non-decreasing functions, and μ_f, μ_g the associated Lebesgue-Stieltjes measures (see 114Xa). Set

$$f(x^+)=\lim_{t\downarrow x} f(t),\quad f(x^-)=\lim_{t\uparrow x} f(t)$$

for each $x\in\mathbb{R}$, and define $g(x^+)$, $g(x^-)$ similarly. Show that whenever $a\le b$ in $\mathbb{R}$,

$$\begin{aligned}\int_{[a,b]} f(x^-)\mu_g(dx)+\int_{[a,b]} g(x^+)\mu_f(dx)&=g(b^+)f(b^+)-g(a^-)f(a^-)\\ &=\int_{[a,b]}\frac{1}{2}(f(x^-)+f(x^+))\mu_g(dx)+\int_{[a,b]}\frac{1}{2}((g(x^-)+g(x^+))\mu_f(dx).\end{aligned}$$

(*Hint*: find two expressions for $(\mu_f\times\mu_g)\{(x,y):a\le x<y\le b\}$.)

>(c) Let (X,Σ,μ) and (Y,T,ν) be complete locally determined measure spaces, λ the c.l.d. product measure on $X\times Y$, and Λ its domain. Suppose that $A\subseteq X$ and $B\subseteq Y$. Show that $A\times B\in\Lambda$ iff either $\mu A=0$ or $\nu B=0$ or $A\in\Sigma$ and $B\in\mathrm{T}$. (*Hint*: if B is not negligible and $A\times B\in\Lambda$, take H such that $\nu H<\infty$ and $B\cap H$ is not negligible. Then $W=A\times(B\cap H)$ is measured by $\mu\times\nu_H$, where ν_H is the subspace measure on H. Now apply 252D to μ, ν_H and W to see that $A\in\Sigma$.)

>(d) Let (X_1,Σ_1,μ_1), (X_2,Σ_2,μ_2), (X_2,Σ_3,μ_3) be three σ-finite measure spaces, and f a real-valued function defined almost everywhere on $X_1\times X_2\times X_3$ and Λ-measurable, where Λ is the domain of the product measure described in 251W or 251Xg. Show that if $\iiint|f(x_1,x_2,x_3)|dx_1dx_2dx_3$ is defined in $\mathbb{R}$, then $\iiint f(x_1,x_2,x_3)dx_2dx_3dx_1$ and $\iiint f(x_1,x_2,x_3)dx_3dx_1dx_2$ exist and are equal.

(e) Give an example of strictly localizable measure spaces (X,Σ,μ), (Y,T,ν) and a $W\in\Sigma\widehat{\otimes}\mathrm{T}$ such that $x\mapsto\nu W[\{x\}]$ is not Σ-measurable. (*Hint*: in 252Kb, try Y a proper subset of $[0,1]$.)

>(f) Set $f(x,y)=\sin(x-y)$ if $0\le y\le x\le y+2\pi$, 0 for other x, $y\in\mathbb{R}^2$. Show that $\iint f(x,y)dx\,dy=0$ and $\iint f(x,y)dy\,dx=2\pi$, taking all integrals with respect to Lebesgue measure.

>(g) Set $f(x,y)=\frac{x^2-y^2}{(x^2+y^2)^2}$ for x, $y\in\,]0,1]$. Show that $\int_0^1\int_0^1 f(x,y)dydx=\frac{\pi}{4}$, $\int_0^1\int_0^1 f(x,y)dxdy=-\frac{\pi}{4}$.

>(h) Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and f a $\Sigma\widehat{\otimes}\mathrm{T}$-measurable function defined on a subset of $X\times Y$. Show that $y\mapsto f(x,y)$ is T-measurable for every $x\in X$.

(i) Let $r\ge1$ be an integer, and write β_r for the Lebesgue measure of the unit ball in $\mathbb{R}^r$. Set $g_r(t)=r\beta_r t^{r-1}$ for $t\ge0$, $\phi(x)=\|x\|$ for $x\in\mathbb{R}^r$. (i) Writing μ_r for Lebesgue measure on $\mathbb{R}^r$, show that $\mu_r\phi^{-1}[E]=\int_E r\beta_r t^{r-1}\mu_1(dt)$ for every Lebesgue measurable set $E\subseteq[0,\infty[$. (*Hint*: start with intervals E, noting from 115Xe that $\mu_r\{x:\|x\|\le a\}=\beta_r a^r$ for $a\ge0$, and progress to open sets, negligible sets and general measurable sets.) (ii) Using 235R, show that

$$\begin{aligned}\int e^{-\|x\|^2/2}\mu_r(dx)=r\beta_r\int_0^\infty t^{r-1}e^{-t^2/2}\mu_1(dt)&=2^{(r-2)/2}r\beta_r\Gamma(\tfrac{r}{2})\\ &=2^{r/2}\beta_r\Gamma(1+\tfrac{r}{2})=(\sqrt{2}\Gamma(\tfrac{1}{2}))^r\end{aligned}$$

where Γ is the Γ-function (225Xj). (iii) Show that

$$2\Gamma(\tfrac{1}{2})^2=2\beta_2\Gamma(2)=2\beta_2\int_0^\infty te^{-t^2/2}dt=2\pi,$$

and hence that $\beta_r=\frac{\pi^{r/2}}{\Gamma(1+\frac{r}{2})}$ and $\int_{-\infty}^\infty e^{-t^2/2}dt=\sqrt{2\pi}$.

252Y Further exercises (a) Let (X,Σ,μ) be a measure space. Show that the following are equiveridical: (i) the completion of μ is locally determined; (ii) the completion of μ coincides with the c.l.d. version of μ; (iii) whenever (Y,T,ν) is a σ-finite measure space and λ the c.l.d. product measure on $X\times Y$ and f is a function such that $\int fd\lambda$ is defined in $[-\infty,\infty]$, then $\iint f(x,y)\nu(dy)\mu(dx)$ is defined and equal to $\int fd\lambda$.

(b) Let (X,Σ,μ) be a measure space. Show that the following are equiveridical: (i) μ has locally determined negligible sets; (ii) whenever (Y,T,ν) is a σ-finite measure space and λ the c.l.d. product measure on $X\times Y$, then $\iint f(x,y)\nu(dy)\mu(dx)$ is defined and equal to $\int f d\lambda$ for any λ-integrable function f.

(c) Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and λ_0 the primitive product measure on $X\times Y$ (251C). Let f be any λ_0-integrable real-valued function. Show that $\iint f(x,y)\nu(dy)\mu(dx)=\int f d\lambda_0$. (*Hint*: show that there are sequences $\langle G_n\rangle_{n\in\mathbb{N}}$, $\langle H_n\rangle_{n\in\mathbb{N}}$ of sets of finite measure such that $f(x,y)$ is defined and equal to 0 for every $(x,y)\in(X\times Y)\setminus\bigcup_{n\in\mathbb{N}}G_n\times H_n$.)

(d) Let (X,Σ,μ) and (Y,T,ν) be measure spaces; let λ_0 be the primitive product measure on $X\times Y$, and λ the c.l.d. product measure. Show that if f is a λ_0-integrable real-valued function, it is λ-integrable, and $\int f d\lambda=\int f d\lambda_0$.

(e) Let (X,Σ,μ) be a complete locally determined measure space and $a<b$ in $\mathbb{R}$, endowed with Lebesgue measure; let Λ be the domain of the c.l.d. product measure λ on $X\times[a,b]$. Let $f:X\times\,]a,b[\,\to\mathbb{R}$ be a Λ-measurable function such that $t\mapsto f(x,t):[a,b]\to\mathbb{R}$ is continuous on $[a,b]$ and differentiable on $]a,b[$ for every $x\in X$. (i) Show that the partial derivative $\dfrac{\partial f}{\partial t}$ with respect to the second variable is Λ-measurable. (ii) Now suppose that $\dfrac{\partial f}{\partial t}$ is λ-integrable and that $\int f(x,t_0)\mu(dx)$ is defined and finite for some $t_0\in\,]a,b[$. Show that $F(t)=\int f(x,t)\mu(dx)$ is defined in $\mathbb{R}$ for every $t\in[a,b]$, that F is absolutely continuous, and that $F'(t)=\int\dfrac{\partial f}{\partial t}(x,t)\mu(dx)$ for almost every $t\in\,]a,b[$. (*Hint*: $F(c)=F(a)+\int_{X\times[a,c]}\dfrac{\partial f}{\partial t}d\lambda$ for every $c\in[a,b]$.)

(f) Show that $\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}=\int_0^1 t^{a-1}(1-t)^{b-1}dt$ for all a, $b>0$. (*Hint*: show that

$$\int_0^\infty t^{a-1}\int_t^\infty e^{-x}(x-t)^{b-1}dxdt=\int_0^\infty e^{-x}\int_0^x t^{a-1}(x-t)^{b-1}dtdx.)$$

(g) Let (X,Σ,μ) and (Y,T,ν) be σ-finite measure spaces and λ the c.l.d. product measure on $X\times Y$. Suppose that $f\in\mathcal{L}^0(\lambda)$ and that $1<p<\infty$. Show that $(\int|\int f(x,y)dx|^p dy)^{1/p}\le\int(\int|f(x,y)|^p dy)^{1/p}dx$. (*Hint*: set $q=\frac{p}{p-1}$ and consider the integral $\int|f(x,y)g(y)|\lambda(d(x,y))$ for $g\in\mathcal{L}^q(\nu)$, using 244K.)

(h) Let ν be Lebesgue measure on $[0,\infty[$; suppose that $f\in\mathcal{L}^p(\nu)$ where $1<p<\infty$. Set $F(y)=\dfrac{1}{y}\int_0^y f$ for $y>0$. Show that $\|F\|_p\le\frac{p}{p-1}\|f\|_p$. (*Hint*: $F(y)=\int_0^1 f(xy)dx$; use 252Yg with $X=[0,1]$, $Y=[0,\infty[$.)

(i) Show that if p is any non-zero (real) polynomial in r variables, then $\{x:x\in\mathbb{R}^r,\ p(x)=0\}$ is Lebesgue negligible.

(j) Let (X,Σ,μ) and (Y,T,ν) be measure spaces, with c.l.d. product $(X\times Y,\Lambda,\lambda)$. Let f be a non-negative Λ-measurable real-valued function defined on a λ-conegligible set, and suppose that

$$\overline{\int}\big(\overline{\int}f(x,y)\mu(dx)\big)\nu(dy)$$

is finite. Show that f is λ-integrable.

(k) Let (X,Σ,μ) be the unit interval $[0,1]$ with Lebesgue measure, and (Y,T,ν) the interval with counting measure, as in 252K; let λ_0 be the primitive product measure on $[0,1]^2$. (i) Setting $\Delta=\{(t,t):t\in[0,1]\}$, show that $\lambda_0\Delta=\infty$. (ii) Show that λ_0 is not semi-finite. (iii) Show that if $W\in\operatorname{dom}\lambda_0$, then $\lambda_0 W=\sum_{y\in[0,1]}\mu W^{-1}[\{y\}]$ if there are a countable set $A\subseteq[0,1]$ and a Lebesgue negligible set $E\subseteq[0,1]$ such that $W\subseteq([0,1]\times A)\cup(E\times[0,1])$, ∞ otherwise.

(l) Let (X,Σ,μ) be a measure space, and λ_0 the primitive product measure on $X\times\mathbb{R}$, where $\mathbb{R}$ is given Lebesgue measure; write Λ for its domain. For any $[0,\infty]$-valued function f defined on a conegligible subset of X, write Ω_f, Ω'_f for the corresponding ordinate sets, as in 252N. Show that if any of $\lambda_0\Omega_f$, $\lambda_0\Omega'_f$, $\int f d\mu$ is defined and finite, so are the others, and all three are equal.

(m) Let (X,Σ,μ) be a complete locally determined measure space, and f a non-negative function defined on a conegligible subset of X. Write Ω_f, Ω'_f for the corresponding ordinate sets, as in 252N. Let λ be the c.l.d. product measure on $X\times\mathbb{R}$, where $\mathbb{R}$ is given Lebesgue measure. Show that $\overline{\int}f\,d\mu=\lambda^*\Omega_f=\lambda^*\Omega'_f$.

(n) Let (X,Σ,μ) be a measure space and $f : X \to [0,\infty[$ a function. Show that $\overline{\int} f d\mu = \int_0^\infty \mu^*\{x : f(x) \ge t\}dt$.

(o) Let (X,Σ,μ) be a complete measure space and write $\mathcal{M}^{0,\infty}$ for the set $\{f : f \in \mathcal{L}^0(\mu),\, \mu\{x : |f(x)| \ge a\}$ is finite for some $a \in [0,\infty[\}$. (i) Show that for each $f \in \mathcal{M}^{0,\infty}$ there is a non-increasing $f^* :]0,\infty[\to \mathbb{R}$ such that $\mu_L\{t : f^*(t) \ge \alpha\} = \mu\{x : |f(x)| \ge \alpha\}$ for every $\alpha > 0$, writing μ_L for Lebesgue measure. (ii) Show that $\int_E |f| d\mu \le \int_0^{\mu E} f^* d\mu_L$ for every $E \in \Sigma$ (allowing ∞). (*Hint*: $(f \times \chi E)^* \le f^*$.) (iii) Show that $\|f^*\|_p = \|f\|_p$ for every $p \in [1,\infty]$, $f \in \mathcal{M}^{0,\infty}$. (*Hint*: $(|f|^p)^* = (f^*)^p$.) (iv) Show that if f, $g \in \mathcal{M}^{0,\infty}$ then $\int |f \times g| d\mu \le \int f^* \times g^* d\mu_L$. (*Hint*: look at simple functions first.) (v) Show that if μ is atomless then $\int_0^a f^* d\mu_L = \sup_{E\in\Sigma,\mu E\le a} \int_E |f|$ for every $a \ge 0$. (*Hint*: 215D.) (vi) Show that $A \subseteq \mathcal{L}^1(\mu)$ is uniformly integrable iff $\{f^* : f \in A\}$ is uniformly integrable in $\mathcal{L}^1(\mu_L)$. (f^* is called the **decreasing rearrangement** of f.)

(p) Let (X,Σ,μ) be a complete locally determined measure space, and write ν for Lebesgue measure on $[0,1]$. Show that the c.l.d. product measure λ on $X \times [0,1]$ is localizable iff μ is localizable. (*Hints*: (i) if $\mathcal{E} \subseteq \Sigma$, show that $F \in \Sigma$ is an essential supremum for $\mathcal{E}$ in Σ iff $F \times [0,1]$ is an essential supremum for $\{E \times [0,1] : E \in \mathcal{E}\}$ in $\Lambda = \operatorname{dom}\lambda$. (ii) For $W \in \Lambda$, $n \in \mathbb{N}$, $k < 2^n$ set

$$W_{nk} = \{x : x \in X,\, \nu^*\{t : (x,t) \in W,\, 2^{-n}k \le t \le 2^{-n}(k+1)\} \ge 2^{-n-1}\}.$$

Show that if $\mathcal{W} \subseteq \Lambda$ and F_{nk} is an essential supremum for $\{W_{nk} : W \in \mathcal{W}\}$ in Σ for all n, k, then

$$\bigcup_{n\in\mathbb{N}} \bigcap_{m\ge n} \bigcup_{k<2^m} F_{mk} \times [2^{-m}k, 2^{-m}(k+1)]$$

is an essential supremum for $\mathcal{W}$ in Λ.)

(q) Let (X,Σ,μ) be the space of Example 216D, and give Lebesgue measure to $[0,1]$. Show that the c.l.d. product measure on $X \times [0,1]$ is complete, locally determined, atomless and not localizable.

(r) Let (X,Σ,μ) be a complete locally determined measure space and (Y,T,ν) a semi-finite measure space with $\nu Y > 0$. Show that if the c.l.d. product measure on $X \times Y$ is strictly localizable, then μ is strictly localizable. (*Hint*: take $F \in \mathrm{T}$, $0 < \nu F < \infty$. Let $\langle W_i\rangle_{i\in I}$ be a decomposition of $X \times Y$. For $i \in I$, $n \in \mathbb{N}$ set $E_{in} = \{x : \nu^*\{y : y \in F,\, (x,y) \in W_i\} \ge 2^{-n}\}$. Apply 213Ye to $\{E_{in} : i \in I,\, n \in \mathbb{N}\}$.)

(s) Let (X,Σ,μ) be the space of Example 216E, and give Lebesgue measure to $[0,1]$. Show that the c.l.d. product measure on $X \times [0,1]$ is complete, locally determined, atomless and localizable, but not strictly localizable.

(t) Let (X,Σ,μ) be a measure space and f a μ-integrable complex-valued function. For $\alpha \in]-\pi,\pi]$ set $H_\alpha = \{x : x \in \operatorname{dom} f,\, \mathcal{R}\mathrm{e}(e^{-i\alpha} f(x)) > 0\}$. Show that $\int_{-\pi}^{\pi} \mathcal{R}\mathrm{e}(e^{-i\alpha} \int_{H_\alpha} f) d\alpha = 2\int |f|$, and hence that there is some α such that $|\int_{H_\alpha} f| \ge \frac{1}{\pi}\int |f|$. (Compare 246F.)

(u) Set $f(t) = t - \ln(t+1)$ for $t > -1$. (i) Show that $\Gamma(a+1) = a^{a+1}e^{-a}\int_{-1}^\infty e^{-af(u)}du$ for every $a > 0$. (*Hint*: substitute $u = \frac{t}{a} - 1$ in 225Xj(iii).) (ii) Show that there is a $\delta > 0$ such that $f(t) \ge \frac{1}{3}t^2$ for $-1 \le t \le \delta$. (iii) Setting $\alpha = \frac{1}{2}f(\delta)$, show that (for $a \ge 1$)

$$\sqrt{a}\int_\delta^\infty e^{-af(t)}dt \le \sqrt{a}e^{-a\alpha}\int_0^\infty e^{-f(t)/2}dt \to 0$$

as $a \to \infty$. (iv) Set $g_a(t) = e^{-af(t/\sqrt{a})}$ if $-\sqrt{a} < t \le \delta\sqrt{a}$, 0 otherwise. Show that $g_a(t) \le e^{-t^2/3}$ for all a, t and that $\lim_{a\to\infty} g_a(t) = e^{-t^2/2}$ for all t, so that

$$\begin{aligned}\lim_{a\to\infty} \frac{e^a\Gamma(a+1)}{a^{a+\frac{1}{2}}} &= \lim_{a\to\infty} \sqrt{a}\int_{-1}^\infty e^{-af(t)}dt = \lim_{a\to\infty} \sqrt{a}\int_{-1}^\delta e^{-af(t)}dt \\ &= \lim_{a\to\infty} \int_{-\infty}^\infty g_a(t)dt = \int_{-\infty}^\infty e^{-t^2/2}dt = \sqrt{2\pi}.\end{aligned}$$

(v) Show that $\lim_{n\to\infty} \frac{n!}{e^{-n}n^n\sqrt{n}} = \sqrt{2\pi}$. (This is **Stirling's formula**.)

(v) Let (X,Σ,μ) be a complete locally determined measure space and f, g two real-valued, μ-virtually measurable functions defined almost everywhere in X. (i) Let λ be the c.l.d. product of μ and Lebesgue measure on $\mathbb{R}$. Setting $\Omega_f^* = \{(x,a) : x \in \operatorname{dom} f,\, a \in \mathbb{R},\, a \le f(x)\}$ and $\Omega_g^* = \{(x,a) : x \in \operatorname{dom} g,\, a \in \mathbb{R},\, a \le g(x)\}$, show that $\lambda(\Omega_f^* \setminus \Omega_g^*) = \int (f-g)^+ d\mu$ and $\lambda(\Omega^* \triangle \Omega_g^*) = \int |f-g| d\mu$. (ii) Suppose that μ is σ-finite. Show that

$$\int |f-g|d\mu = \int_{-\infty}^{\infty} \mu(\{x : x \in \operatorname{dom} f \cap \operatorname{dom} g,\ (f(x)-a)(g(x)-a) < 0\}da.$$

(iii) Suppose that μ is σ-finite, that T is a σ-subalgebra of Σ, that $E \in \Sigma$ and that $g : X \to [0,1]$ is T-measurable. Show that there is an $F \in \mathrm{T}$ such that $\mu(E \triangle F) \le \int |\chi E - g|d\mu$.

252 Notes and comments For a volume and a half now I have asked you to accept the idea of integrating partially-defined functions, insisting that sooner or later they would appear at the core of the subject. The moment has now come. If we wish to apply Fubini's and Tonelli's theorems in the most fundamental of all cases, with both factors equal to Lebesgue measure on the unit interval, it is surely natural to look at all functions which are integrable on the square for two-dimensional Lebesgue measure. Now two-dimensional Lebesgue measure is a complete measure, so, in particular, assigns zero measure to any set of the form $\{(x,b) : x \in A\}$ or $\{(a,y) : y \in A\}$, whether or not the set A is measured by one-dimensional measure. Accordingly, if f is a function of two variables which is integrable for two-dimensional Lebesgue measure, there is no reason why any particular section $x \mapsto f(x,b)$ or $y \mapsto f(a,y)$ should be measurable, let alone integrable. Consequently, even if f itself is defined everywhere, the outer integral of $\iint f(x,y)dxdy$ is likely to be applied to a function which is not defined for every y. Let me remark that the problem does not concern '∞'; the awkward functions are those with sections so irregular that they cannot be assigned an integral at all.

I have seen many approaches to this particular nettle, generally less whole-hearted than the one I have determined on for this treatise. Part of the difficulty is that Fubini's theorem really is at the centre of measure theory. Over large parts of the subject, it is possible to assert that a result is non-trivial if and only if it depends on Fubini's theorem. I am therefore unwilling to insert any local fix, saying that 'in this chapter, we shall integrate functions which are not defined everywhere'; before long, such a provision would have to be interpolated into the preambles to half the best theorems, or an explanation offered of why it wasn't necessary in their particular contexts. I suppose that one of the commonest responses is (like Halmos 50) to restrict attention to $\Sigma\widehat{\otimes}\mathrm{T}$-measurable functions, which eliminates measurability problems for the moment (252Xh, 252P); but unhappily (or rather, to my mind, happily) there are crucial applications in which the functions are not actually $\Sigma\widehat{\otimes}\mathrm{T}$-measurable, but belong to some wider class, and this restriction sooner or later leads to undignified contortions as we are forced to adapt limited results to unforeseen contexts. Besides, it leaves unsaid the really rather important information that if f is a measurable function of two variables then (under appropriate conditions) almost all its sections are measurable (252E).

In 252B and its corollaries there is a clumsy restriction: we assume that one of the measures is σ-finite and the other is either strictly localizable or complete and locally determined. The obvious question is, whether we need these hypotheses. From 252K we see that the hypothesis 'σ-finite' on the second factor can certainly not be abandoned, even when the first factor is a complete probability measure. The requirement 'μ is either strictly localizable or complete and locally determined' is in fact fractionally stronger than what is needed, as well as disagreeably elaborate. The 'right' hypothesis is that the completion of μ should be locally determined (see 252Ya). The point is that because the product of two measures is the same as the product of their c.l.d. versions (251T), no theorem which leads from the product measure to the factor measures can distinguish between a measure and its c.l.d. version; so that, in 252B, we must expect to need μ and its c.l.d. version to give rise to the same integrals. The proof of 252B would be better focused if the hypothesis was simplified to 'ν is σ-finite and μ is complete and locally determined'. But this would just transfer part of the argument into the proof of 252C.

We also have to work a little harder in 252B in order to cover functions and integrals taking the values $\pm\infty$. Fubini's theorem is so central to measure theory that I believe it is worth taking a bit of extra trouble to state the results in maximal generality. This is especially important because we frequently apply it in multiply repeated integrals, as in 252Xd, in which we have even less control than usual over the intermediate functions to be integrated.

I have expressed all the main results of this section in terms of the 'c.l.d.' product measure. In the case of σ-finite spaces, of course, which is where the theory works best, we could just as well use the 'primitive' product measure. Indeed, Fubini's theorem itself has a version in terms of the primitive product measure which is rather more elegant than 252B as stated (252Yc), and covers the great majority of applications. (Integrals with respect to the primitive and c.l.d. product measures are of course very closely related; see 252Yd.) But we do sometimes need to look at non-σ-finite spaces, and in these cases the asymmetric form in 252B is close to the best we can do. Using the primitive product measure does not help at all with the most substantial obstacle, the phenomenon in 252K (see 252Yk).

The pre-calculus concept of an integral as 'the area under a curve' is given expression in 252N: the integral of a non-negative function is the measure of its ordinate set. This is unsatisfactory as a definition of the integral, not just because of the requirement that the base space should be complete and locally determined (which can be dealt with by using the primitive product measure, as in 252Yl), but because the construction of the product measure involves integration (part (c) of the proof of 251E). The idea of 252N is to relate the measure of an ordinate set to the integral of the measures of its vertical sections. Curiously, if instead we integrate the measures of its *horizontal* sections, as in

252O, we get a more versatile result. (Indeed this one does not involve the concept of 'product measure', and could have appeared at any point after §123.) Note that the integral $\int_0^\infty \ldots dt$ here is applied to a monotonic function, so may be interpreted as an improper Riemann integral. If you think you know enough about the Riemann integral to make this a tempting alternative to the construction in §122, the tricky bit now becomes the proof that the integral is additive.

A different line of argument is to use integration over sections to define a product measure. The difficulty with this approach is that unless we take great care we may find ourselves with an asymmetric construction. My own view is that such an asymmetry is acceptable only when there is no alternative. But in Chapter 43 of Volume 4 I will describe a couple of examples.

Of the two examples I give here, 252K is supposed to show that when I call for σ-finite spaces they are really necessary, while 252L is supposed to show that joint measurability is essential in Tonelli's theorem and its corollaries. The factor spaces in 252K, Lebesgue measure and counting measure, are chosen to show that it is only the lack of σ-finiteness that can be the problem; they are otherwise as regular as one can reasonably ask. In 252L I have used the countable-cocountable measure on ω_1, which you may feel is fit only for counter-examples; and the question does arise, whether the same phenomenon occurs with Lebesgue measure. This leads into deep water, and I will return to it in Chapter 53 of Volume 5.

I ought perhaps to note explicitly that in Fubini's theorem, we really do need to have a function which is integrable for the product measure. I include 252Xf and 252Xg to remind you that even in the best-regulated circumstances, the repeated integrals $\iint f\,dxdy$, $\iint f\,dydx$ may fail to be equal if f is not integrable as a function of two variables.

There are many ways to calculate the volume β_r of an r-dimensional ball; the one I have used in 252Q follows a line that would have been natural to me before I ever heard of measure theory. In 252Xi I suggest another method. The idea of integration-by-substitution, used in part (b) of the argument for 252Q, is there supported by an ad hoc argument; I will present a different, more generally applicable, approach in Chapter 26. Elsewhere (252Xi, 252Yf, 252Yh, 252Yu) I find myself taking for granted substitutions of the form $t \mapsto at$, $t \mapsto a+t$, $t \mapsto t^2$; for a systematic justification, see §263. Of course an enormous number of other formulae of advanced calculus are also based on repeated integration of one kind or another, and I give a sample handful of such results (252Xb, 252Ye-252Yh, 252Yu).

253 Tensor products

The theorems of the last section show that the integrable functions on a product of two measure spaces can be effectively studied in terms of integration on each factor space separately. In this section I present a very striking relationship between the L^1 space of a product measure and the L^1 spaces of its factors, which actually determines the product L^1 up to isomorphism as Banach lattice. I start with a brief note on bilinear operators (253A) and a description of the canonical bilinear operator from $L^1(\mu) \times L^1(\nu)$ to $L^1(\mu \times \nu)$ (253B-253E). The main theorem of the section is 253F, showing that this canonical map is universal for continuous bilinear operators from $L^1(\mu) \times L^1(\nu)$ to Banach spaces; it also determines the ordering of $L^1(\mu \times \nu)$ (253G). I end with a description of a fundamental type of conditional expectation operator (253H) and notes on products of indefinite-integral measures (253I) and upper integrals of special kinds of function (253J, 253K).

253A Bilinear operators Before looking at any of the measure theory in this section, I introduce a concept from the theory of linear spaces.

(a) Let U, V and W be linear spaces over $\mathbb{R}$ (or, indeed, any field). A map $\phi : U \times V \to W$ is **bilinear** if it is linear in each variable separately, that is,

$$\phi(u_1 + u_2, v) = \phi(u_1, v) + \phi(u_2, v),$$

$$\phi(u, v_1 + v_2) = \phi(u, v_1) + \phi(u, v_2),$$

$$\phi(\alpha u, v) = \alpha\phi(u, v) = \phi(u, \alpha v)$$

for all u, u_1, $u_2 \in U$, v, v_1, $v_2 \in V$ and scalars α. Observe that such a ϕ gives rise to, and in turn can be defined by, a linear operator $T : U \to \mathsf{L}(V;W)$, writing $\mathsf{L}(V;W)$ for the space of linear operators from V to W, where

$$(Tu)(v) = \phi(u, v)$$

for all $u \in U$, $v \in V$. Hence, or otherwise, we can see, for instance, that $\phi(0, v) = \phi(u, 0) = 0$ whenever $u \in U$ and $v \in V$.

If W' is another linear space over the same field, and $S : W \to W'$ is a linear operator, then $S\phi : U \times V \to W'$ is bilinear.

(b) Now suppose that U, V and W are normed spaces, and $\phi : U \times V \to W$ a bilinear operator. Then we say that ϕ is **bounded** if $\sup\{\|\phi(u,v)\| : \|u\| \le 1, \|v\| \le 1\}$ is finite, and in this case we call this supremum the norm $\|\phi\|$ of ϕ. Note that $\|\phi(u,v)\| \le \|\phi\|\|u\|\|v\|$ for all $u \in U$, $v \in V$ (because

$$\|\phi(u,v)\| = \alpha\beta\|\phi(\alpha^{-1}u, \beta^{-1}v)\| \le \alpha\beta\|\phi\|$$

whenever $\alpha > \|u\|$, $\beta > \|v\|$).

If W' is another normed space and $S : W \to W'$ is a bounded linear operator, then $S\phi : U \times V \to W'$ is a bounded bilinear operator, and $\|S\phi\| \le \|S\|\|\phi\|$.

253B Definition The most important bilinear operators of this section are based on the following idea. Let f and g be real-valued functions. I will write $f \otimes g$ for the function $(x,y) \mapsto f(x)g(y) : \operatorname{dom} f \times \operatorname{dom} g \to \mathbb{R}$.

253C Proposition (a) Let X and Y be sets, and Σ, T σ-algebras of subsets of X, Y respectively. If f is a Σ-measurable real-valued function defined on a subset of X, and g is a T-measurable real-valued function defined on a subset of Y, then $f \otimes g$, as defined in 253B, is $\Sigma\widehat{\otimes}\mathrm{T}$-measurable.

(b) Let (X, Σ, μ) and (Y, T, ν) be measure spaces, and λ the c.l.d. product measure on $X \times Y$. If $f \in \mathcal{L}^0(\mu)$ and $g \in \mathcal{L}^0(\nu)$, then $f \otimes g \in \mathcal{L}^0(\lambda)$.

Remark Recall from 241A that $\mathcal{L}^0(\mu)$ is the space of μ-virtually measurable real-valued functions defined on μ-conegligible subsets of X.

proof (a) The point is that $f \otimes \chi Y$ is $\Sigma\widehat{\otimes}\mathrm{T}$-measurable, because for any $\alpha \in \mathbb{R}$ there is an $E \in \Sigma$ such that

$$\{x : f(x) \ge \alpha\} = E \cap \operatorname{dom} f,$$

so that

$$\{(x,y) : (f \otimes \chi Y)(x,y) \ge \alpha\} = (E \cap \operatorname{dom} f) \times Y = (E \times Y) \cap \operatorname{dom}(f \otimes \chi Y),$$

and of course $E \times Y \in \Sigma\widehat{\otimes}\mathrm{T}$. Similarly, $\chi X \otimes g$ is $\Sigma\widehat{\otimes}\mathrm{T}$-measurable and $f \otimes g = (f \otimes \chi Y) \times (\chi X \otimes g)$ is $\Sigma\widehat{\otimes}\mathrm{T}$-measurable.

(b) Let $E \in \Sigma$, $F \in \mathrm{T}$ be conegligible subsets of X, Y respectively such that $E \subseteq \operatorname{dom} f$, $F \subseteq \operatorname{dom} g$, $f\restriction E$ is Σ-measurable and $g\restriction F$ is T-measurable. Write Λ for the domain of λ. Then $\Sigma\widehat{\otimes}\mathrm{T} \subseteq \Lambda$ (251Ia). Also $E \times F$ is λ-conegligible, because

$$\begin{aligned}\lambda((X \times Y) \setminus (E \times F)) &\le \lambda((X \setminus E) \times Y) + \lambda(X \times (Y \setminus F)) \\ &= \mu(X \setminus E) \cdot \nu Y + \mu X \cdot \nu(Y \setminus F) = 0\end{aligned}$$

(also from 251Ia). So $\operatorname{dom}(f \otimes g) \supseteq E \times F$ is conegligible. Also, by (a), $(f \otimes g)\restriction(E \times F) = (f\restriction E) \otimes (g\restriction F)$ is $\Sigma\widehat{\otimes}\mathrm{T}$-measurable, therefore Λ-measurable, and $f \otimes g$ is virtually measurable. Thus $f \otimes g \in \mathcal{L}^0(\lambda)$, as claimed.

253D Now we can apply the ideas of 253B-253C to integrable functions.

Proposition Let (X, Σ, μ) and (Y, T, ν) be measure spaces, and write λ for the c.l.d. product measure on $X \times Y$. If $f \in \mathcal{L}^1(\mu)$ and $g \in \mathcal{L}^1(\nu)$, then $f \otimes g \in \mathcal{L}^1(\lambda)$ and $\int f \otimes g \, d\lambda = \int f \, d\mu \int g \, d\nu$.

Remark I follow §242 in writing $\mathcal{L}^1(\mu)$ for the space of μ-integrable real-valued functions.

proof (a) Consider first the case $f = \chi E$, $g = \chi F$ where $E \in \Sigma$, $F \in \mathrm{T}$ have finite measure; then $f \otimes g = \chi(E \times F)$ is λ-integrable with integral

$$\lambda(E \times F) = \mu E \cdot \nu F = \int f \, d\mu \cdot \int g \, d\nu,$$

by 251Ia.

(b) It follows at once that $f \otimes g$ is λ-simple, with $\int f \otimes g \, d\lambda = \int f \, d\mu \int g \, d\nu$, whenever f is a μ-simple function and g is a ν-simple function.

(c) If f and g are non-negative integrable functions, there are non-decreasing sequences $\langle f_n \rangle_{n \in \mathbb{N}}$, $\langle g_n \rangle_{n \in \mathbb{N}}$ of non-negative simple functions converging almost everywhere to f, g respectively; now note that if $E \subseteq X$, $F \subseteq Y$ are conegligible, $E \times F$ is conegligible in $X \times Y$, as remarked in the proof of 253C, so the non-decreasing sequence $\langle f_n \times g_n \rangle_{n \in \mathbb{N}}$ of λ-simple functions converges almost everywhere to $f \otimes g$, and

$$\int f \otimes g \, d\lambda = \lim_{n\to\infty} \int f_n \otimes g_n d\lambda = \lim_{n\to\infty} \int f_n d\mu \int g_n d\nu = \int f \, d\mu \int g \, d\nu$$

by B.Levi's theorem.

(d) Finally, for general f and g, we can express them as the differences $f^+ - f^-$, $g^+ - g^-$ of non-negative integrable functions, and see that

$$\int f \otimes g\,d\lambda = \int f^+ \otimes g^+ - f^+ \otimes g^- - f^- \otimes g^+ + f^- \otimes g^- d\lambda = \int f\,d\mu \int g\,d\nu.$$

253E The canonical map $L^1 \times L^1 \to L^1$ I continue the argument from 253D. Because $E \times F$ is conegligible in $X \times Y$ whenever E and F are conegligible subsets of X and Y, $f_1 \otimes g_1 = f \otimes g$ λ-a.e. whenever $f = f_1$ μ-a.e. and $g = g_1$ ν-a.e. We may therefore define $u \otimes v \in L^1(\lambda)$, for $u \in L^1(\mu)$ and $v \in L^1(\nu)$, by saying that $u \otimes v = (f \otimes g)^\bullet$ whenever $u = f^\bullet$ and $v = g^\bullet$.

Now if f, f_1, $f_2 \in \mathcal{L}^1(\mu)$, g, g_1, $g_2 \in \mathcal{L}^1(\nu)$ and $a \in \mathbb{R}$,

$$(f_1 + f_2) \otimes g = (f_1 \otimes g) + (f_2 \otimes g),$$

$$f \otimes (g_1 + g_2) = (f \otimes g_1) + (f \otimes g_2),$$

$$(af) \otimes g = a(f \otimes g) = f \otimes (ag).$$

It follows at once that the map $(u, v) \mapsto u \otimes v$ is bilinear.

Moreover, if $f \in \mathcal{L}^1(\mu)$ and $g \in \mathcal{L}^1(\nu)$, $|f| \otimes |g| = |f \otimes g|$, so $\int |f \otimes g| d\lambda = \int |f| d\mu \int |g| d\nu$. Accordingly

$$\|u \otimes v\|_1 = \|u\|_1 \|v\|_1$$

for all $u \in L^1(\mu)$, $v \in L^1(\nu)$. In particular, the bilinear operator $\otimes$ is bounded, with norm 1 (except in the trivial case in which one of $L^1(\mu)$, $L^1(\nu)$ is 0-dimensional).

253F We are now ready for the main theorem of this section.

Theorem Let (X, Σ, μ) and (Y, T, ν) be measure spaces, and let λ be the c.l.d. product measure on $X \times Y$. Let W be any Banach space and $\phi : L^1(\mu) \times L^1(\nu) \to W$ a bounded bilinear operator. Then there is a unique bounded linear operator $T : L^1(\lambda) \to W$ such that $T(u \otimes v) = \phi(u, v)$ for all $u \in L^1(\mu)$ and $v \in L^1(\nu)$, and $\|T\| = \|\phi\|$.

proof (a) The centre of the argument is the following fact: if $E_0, \dots, E_n$ are measurable sets of finite measure in X, $F_0, \dots, F_n$ are measurable sets of finite measure in Y, $a_0, \dots, a_n \in \mathbb{R}$ and $\sum_{i=0}^n a_i \chi(E_i \times F_i) = 0$ λ-a.e., then $\sum_{i=0}^n a_i \phi(\chi E_i^\bullet, \chi F_i^\bullet) = 0$ in W. **P** We can find a disjoint family $\langle G_j \rangle_{j \le m}$ of measurable sets of finite measure in X such that each E_i is expressible as a union of some subfamily of the G_j; so that χE_i is expressible in the form $\sum_{j=0}^m b_{ij} \chi G_j$ (see 122Ca). Similarly, we can find a disjoint family $\langle H_k \rangle_{k \le l}$ of measurable sets of finite measure in Y such that each χF_i is expressible as $\sum_{k=0}^l c_{ik} \chi H_k$. Now

$$\sum_{j=0}^m \sum_{k=0}^l \bigl(\sum_{i=0}^n a_i b_{ij} c_{ik}\bigr) \chi(G_j \times H_k) = \sum_{i=0}^n a_i \chi(E_i \times F_i) = 0 \text{ } \lambda\text{-a.e.}$$

Because the $G_j \times H_k$ are disjoint, and $\lambda(G_j \times H_k) = \mu G_j \cdot \nu H_k$ for all j, k, it follows that for every $j \le m$, $k \le l$ we have either $\mu G_j = 0$ or $\nu H_k = 0$ or $\sum_{i=0}^n a_i b_{ij} c_{ik} = 0$. In any of these three cases, $\sum_{i=0}^n a_i b_{ij} c_{ik} \phi(\chi G_j^\bullet, \chi H_k^\bullet) = 0$ in W. But this means that

$$0 = \sum_{j=0}^m \sum_{k=0}^l \bigl(\sum_{i=0}^n a_i b_{ij} c_{ik}\bigr) \phi(\chi G_j^\bullet, \chi H_k^\bullet) = \sum_{i=0}^n a_i \phi(\chi E_i^\bullet, \chi F_i^\bullet),$$

as claimed. **Q**

(b) It follows that if $E_0, \dots, E_n, E_0', \dots, E_m'$ are measurable sets of finite measure in X, $F_0, \dots, F_n, F_0', \dots, F_m'$ are measurable sets of finite measure in Y, $a_0, \dots, a_n, a_0', \dots, a_m' \in \mathbb{R}$ and $\sum_{i=0}^n a_i \chi(E_i \times F_i) = \sum_{i=0}^m a_i' \chi(E_i' \times F_i')$ λ-a.e., then

$$\sum_{i=0}^n a_i \phi(\chi E_i^\bullet, \chi F_i^\bullet) = \sum_{i=0}^m a_i' \phi(\chi E_i'^\bullet, \chi F_i'^\bullet)$$

in W. Let M be the linear subspace of $L^1(\lambda)$ generated by

$$\{\chi(E \times F)^\bullet : E \in \Sigma,\ \mu E < \infty,\ F \in \mathrm{T},\ \nu F < \infty\};$$

then we have a unique map $T_0 : M \to W$ such that

$$T_0\bigl(\sum_{i=0}^n a_i \chi(E_i \times F_i)^\bullet\bigr) = \sum_{i=0}^n a_i \phi(\chi E_i^\bullet, \chi F_i^\bullet)$$

whenever $E_0, \dots, E_n$ are measurable sets of finite measure in X, $F_0, \dots, F_n$ are measurable sets of finite measure in Y and $a_0, \dots, a_n \in \mathbb{R}$. Of course T_0 is linear.

(c) Some of the same calculations show that $\|T_0u\| \le \|\phi\|\|u\|_1$ for every $u \in M$. **P** If $u \in M$, then, by the arguments of (a), we can express u as $\sum_{j=0}^{m}\sum_{k=0}^{l} a_{jk}\chi(G_j \times H_k)^\bullet$, where $\langle G_j\rangle_{j\le m}$ and $\langle H_k\rangle_{k\le l}$ are disjoint families of sets of finite measure. Now

$$\begin{aligned}\|T_0u\| &= \|\sum_{j=0}^{m}\sum_{k=0}^{l} a_{jk}\phi(\chi G_j^\bullet, \chi H_k^\bullet)\| \le \sum_{j=0}^{m}\sum_{k=0}^{l}|a_{jk}|\|\phi(\chi G_j^\bullet, \chi H_k^\bullet)\| \\ &\le \sum_{j=0}^{m}\sum_{k=0}^{l}|a_{jk}|\|\phi\|\|\chi G_j^\bullet\|_1\|\chi H_k^\bullet\|_1 = \|\phi\|\sum_{j=0}^{m}\sum_{k=0}^{l}|a_{jk}|\mu G_j\cdot\nu H_k \\ &= \|\phi\|\sum_{j=0}^{m}\sum_{k=0}^{l}|a_{jk}|\lambda(G_j\times H_k) = \|\phi\|\|u\|_1,\end{aligned}$$

as claimed. **Q**

(d) The next point is to observe that M is dense in $L^1(\lambda)$ for $\|\,\|_1$. **P** Repeating the ideas above once again, we observe that if $E_0,\dots,E_n$ are sets of finite measure in X and $F_0,\dots,F_n$ are sets of finite measure in Y, then $\chi(\bigcup_{i\le n} E_i\times F_i)^\bullet \in M$; this is because, expressing each E_i as a union of G_j, where the G_j are disjoint, we have

$$\bigcup_{i\le n} E_i\times F_i = \bigcup_{j\le m} G_j\times F_j',$$

where $F_j' = \bigcup\{F_i : G_j \subseteq E_i\}$ for each j; now $\langle G_j\times F_j'\rangle_{j\le m}$ is disjoint, so

$$\chi(\bigcup_{j\le m} G_j\times F_j)^\bullet = \sum_{j=0}^{m}\chi(G_j\times F_j')^\bullet \in M.$$

So 251Ie tells us that whenever $\lambda H < \infty$ and $\epsilon > 0$ there is a G such that $\lambda(H\triangle G) \le \epsilon$ and $\chi G^\bullet \in M$; now

$$\|\chi H^\bullet - \chi G^\bullet\|_1 = \lambda(G\triangle H) \le \epsilon,$$

so $\chi H^\bullet$ is approximated arbitrarily closely by members of M, and belongs to the closure $\overline{M}$ of M in $L^1(\lambda)$. Because M is a linear subspace of $L^1(\lambda)$, so is $\overline{M}$ (2A4Cb); accordingly $\overline{M}$ contains the equivalence classes of all λ-simple functions; but these are dense in $L^1(\lambda)$ (242Mb), so $\overline{M} = L^1(\lambda)$, as claimed. **Q**

(e) Because W is a Banach space, it follows that there is a bounded linear operator $T : L^1(\lambda) \to W$ extending T_0, with $\|T\| = \|T_0\| \le \|\phi\|$ (2A4I). Now $T(u\otimes v) = \phi(u,v)$ for all $u \in L^1(\mu)$, $v \in L^1(\nu)$. **P** If $u = \chi E^\bullet$ and $v = \chi F^\bullet$, where E, F are measurable sets of finite measure, then

$$T(u\otimes v) = T(\chi(E\times F)^\bullet) = T_0(\chi(E\times F)^\bullet) = \phi(\chi E^\bullet, \chi F^\bullet) = \phi(u,v).$$

Because ϕ and $\otimes$ are bilinear and T is linear,

$$T(f^\bullet\otimes g^\bullet) = \phi(f^\bullet, g^\bullet)$$

whenever f and g are simple functions. Now whenever $u \in L^1(\mu)$, $v \in L^1(\nu)$ and $\epsilon > 0$, there are simple functions f, g such that $\|u - f^\bullet\|_1 \le \epsilon$, $\|v - g^\bullet\|_1 \le \epsilon$ (242Mb again); so that

$$\begin{aligned}\|\phi(u,v) - \phi(f^\bullet, g^\bullet)\| &\le \|\phi(u - f^\bullet, v - g^\bullet)\| + \|\phi(u, g^\bullet - v)\| + \|\phi(f^\bullet - u, v)\| \\ &\le \|\phi\|(\epsilon^2 + \epsilon\|u\|_1 + \epsilon\|v\|_1).\end{aligned}$$

Similarly

$$\|u\otimes v - f^\bullet\otimes g^\bullet\|_1 \le \epsilon(\epsilon + \|u\|_1 + \|v\|_1),$$

so

$$\|T(u\otimes v) - T(f^\bullet\otimes g^\bullet)\| \le \epsilon\|T\|(\epsilon + \|u\|_1 + \|v\|_1);$$

because $T(f^\bullet\otimes g^\bullet) = \phi(f^\bullet, g^\bullet)$,

$$\|T(u\otimes v) - \phi(u,v)\| \le \epsilon(\|T\| + \|\phi\|)(\epsilon + \|u\|_1 + \|v\|_1).$$

As ϵ is arbitrary, $T(u\otimes v) = \phi(u,v)$, as required. **Q**

(f) The argument of (e) ensured that $\|T\| \le \|\phi\|$. Because $\|u\otimes v\|_1 \le \|u\|_1\|v\|_1$ for all $u \in L^1(\mu)$ and $v \in L^1(\nu)$, $\|\phi(u,v)\| \le \|T\|\|u\|_1\|v\|_1$ for all u, v, and $\|\phi\| \le \|T\|$; so $\|T\| = \|\phi\|$.

(g) Thus T has the required properties. To see that it is unique, we have only to observe that any bounded linear operator $S : L^1(\lambda) \to W$ such that $S(u\otimes v) = \phi(u,v)$ for all $u \in L^1(\mu)$, $v \in L^1(\nu)$ must agree with T on objects of the form $\chi(E\times F)^\bullet$ where E and F are of finite measure, and therefore on every member of M; because M is dense and both S and T are continuous, they agree everywhere in $L^1(\lambda)$.

253G The order structure of L^1 In 253F I have treated the L^1 spaces exclusively as normed linear spaces. In general, however, the order structure of an L^1 space (see 242C) is as important as its norm. The map $\otimes$: $L^1(\mu)\times L^1(\nu)\to L^1(\lambda)$ respects the order structures of the three spaces in the following strong sense.

Proposition Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and λ the c.l.d. product measure on $X\times Y$. Then

(a) $u\otimes v\ge 0$ in $L^1(\lambda)$ whenever $u\ge 0$ in $L^1(\mu)$ and $v\ge 0$ in $L^1(\nu)$.

(b) The positive cone $\{w : w\ge 0\}$ of $L^1(\lambda)$ is precisely the closed convex hull C of $\{u\otimes v : u\ge 0,\ v\ge 0\}$ in $L^1(\lambda)$.

*(c) Let W be any Banach lattice, and $T : L^1(\lambda)\to W$ a bounded linear operator. Then the following are equiveridical:

(i) $Tw\ge 0$ in W whenever $w\ge 0$ in $L^1(\lambda)$;

(ii) $T(u\otimes v)\ge 0$ in W whenever $u\ge 0$ in $L^1(\mu)$ and $v\ge 0$ in $L^1(\nu)$.

proof (a) If u, $v\ge 0$ then they are expressible as $f^\bullet$, $g^\bullet$ where $f\in\mathcal{L}^1(\mu)$, $g\in\mathcal{L}^1(\nu)$, $f\ge 0$ and $g\ge 0$. Now $f\otimes g\ge 0$ so $u\otimes v=(f\otimes g)^\bullet\ge 0$.

(b)(i) Write $L^1(\lambda)^+$ for $\{w : w\in L^1(\lambda),\ w\ge 0\}$. Then $L^1(\lambda)^+$ is a closed convex set in $L^1(\lambda)$ (242De); by (a), it contains $u\otimes v$ whenever $u\in L^1(\mu)^+$ and $v\in L^1(\nu)^+$, so it must include C.

(ii)(α) Of course $0=0\otimes 0\in C$. (β) If $u\in M$, as defined in the proof of 253F, and $u>0$, then u is expressible as $\sum_{j\le m,k\le l}a_{jk}\chi(G_j\times H_k)^\bullet$, where $G_0,\dots,G_m$ and $H_0,\dots,H_l$ are disjoint sequences of sets of finite measure, as in (a) of the proof of 253F. Now a_{jk} can be negative only if $\chi(G_j\times H_k)^\bullet=0$, so replacing every a_{jk} by $\max(0,a_{jk})$ if necessary, we can suppose that $a_{jk}\ge 0$ for all j, k. Not all the a_{jk} can be zero, so $a=\sum_{j\le m,k\le l}a_{jk}>0$, and

$$u=\sum_{j\le m,k\le l}\frac{a_{jk}}{a}\cdot a\chi(G_j\times H_k)^\bullet=\sum_{j\le m,k\le l}\frac{a_{jk}}{a}\cdot(a\chi G_j^\bullet)\otimes\chi H_k^\bullet\in C.$$

(γ) If $w\in L^1(\lambda)^+$ and $\epsilon>0$, express w as $h^\bullet$ where $h\ge 0$ in $\mathcal{L}^1(\lambda)$. There is a simple function $h_1\ge 0$ such that $h_1\le_{\text{a.e.}}h$ and $\int h\le\int h_1+\epsilon$. Express h_1 as $\sum_{i=0}^n a_i\chi H_i$ where $\lambda H_i<\infty$ and $a_i\ge 0$ for each i, and for each $i\le n$ choose sets $G_{i0},\dots,G_{im_i}\in\Sigma$, $F_{i0},\dots,F_{im_i}\in\mathrm{T}$, all of finite measure, such that $G_{i0},\dots,G_{im_i}$ are disjoint and $\lambda(H_i\triangle\bigcup_{j\le m_i}G_{ij}\times F_{ij})\le\epsilon/(n+1)(a_i+1)$, as in (d) of the proof of 253F. Set

$$w_0=\sum_{i=0}^n a_i\sum_{j=0}^{m_i}\chi(G_{ij}\times F_{ij})^\bullet.$$

Then $w_0\in C$ because $w_0\in M$ and $w_0\ge 0$. Also

$$\begin{aligned}\|w-w_0\|_1&\le\|w-h_1^\bullet\|_1+\|h_1^\bullet-w_0\|_1\\&\le\int(h-h_1)d\lambda+\sum_{i=0}^n a_i\int|\chi H_i-\sum_{j=0}^{m_i}\chi(G_{ij}\times F_{ij})|d\lambda\\&\le\epsilon+\sum_{i=0}^n a_i\lambda(H\triangle\bigcup_{j\le m_i}G_{ij}\times F_{ij})\le 2\epsilon.\end{aligned}$$

As ϵ is arbitrary and C is closed, $w\in C$. As w is arbitrary, $L^1(\lambda)^+\subseteq C$ and $C=L^1(\lambda)^+$.

(c) Part (a) tells us that (i)$\Rightarrow$(ii). For the reverse implication, we need a fragment from the theory of Banach lattices: $W^+=\{w : w\in W,\ w\ge 0\}$ is a closed set in W. **P** If w, $w'\in W$, then

$$w=(w-w')+w'\le|w-w'|+w'\le|w-w'|+|w'|,$$

$$-w=(w'-w)-w'\le|w-w'|-w'\le|w-w'|+|w'|,$$

$$|w|\le|w-w'|+|w'|,\quad |w|-|w'|\le|w-w'|,$$

because $|w|=w\vee(-w)$ and the order of W is translation-invariant (241Ec). Similarly, $|w'|-|w|\le|w-w'|$ and $||w|-|w'||\le|w-w'|$, so $\||w|-|w'|\|\le\|w-w'\|$, by the definition of Banach lattice (242G). Setting $\phi(w)=|w|-w$, we see that $\|\phi(w)-\phi(w')\|\le 2\|w-w'\|$ for all w, $w'\in W$, so that ϕ is continuous.

Now, because the order is invariant under multiplication by positive scalars,

$$w\ge 0\iff 2w\ge 0\iff w\ge -w\iff w=|w|\iff\phi(w)=0,$$

so $W^+=\{w : \phi(w)=0\}$ is closed. **Q**

Now suppose that (ii) is true, and set $C_1 = \{w : w \in L^1(\lambda),\, Tw \ge 0\}$. Then C_1 contains $u \otimes v$ whenever u, $v \ge 0$; but also it is convex, because T is linear, and closed, because T is continuous and $C_1 = T^{-1}[W^+]$. By (b), C_1 includes $\{w : w \in L^1(\lambda),\, w \ge 0\}$, as required by (i).

253H Conditional expectations The ideas of this section and the preceding one provide us with some of the most important examples of conditional expectations.

Theorem Let (X, Σ, μ) and (Y, T, ν) be complete probability spaces, with c.l.d. product $(X \times Y, \Lambda, \lambda)$. Set $\Lambda_1 = \{E \times Y : E \in \Sigma\}$. Then Λ_1 is a σ-subalgebra of Λ. Given a λ-integrable real-valued function f, set

$$g(x, y) = \int f(x, z)\nu(dz)$$

whenever $x \in X$, $y \in Y$ and the integral is defined in $\mathbb{R}$. Then g is a conditional expectation of f on Λ_1.

proof We know that $\Lambda_1 \subseteq \Lambda$, by 251Ia, and Λ_1 is a σ-algebra of sets because Σ is. Fubini's theorem (252B, 252C) tells us that $f_1(x) = \int f(x, z)\nu(dz)$ is defined for almost every x, and therefore that $g = f_1 \otimes \chi Y$ is defined almost everywhere in $X \times Y$. f_1 is μ-virtually measurable; because μ is complete, f_1 is Σ-measurable, so g is Λ_1-measurable (since $\{(x, y) : g(x, y) \le \alpha\} = \{x : f_1(x) \le \alpha\} \times Y$ for every $\alpha \in \mathbb{R}$). Finally, if $W \in \Lambda_1$, then $W = E \times Y$ for some $E \in \Sigma$, so

$$\int_W g\,d\lambda = \int (f_1 \otimes \chi Y) \times (\chi E \otimes \chi Y) d\lambda = \int f_1 \times \chi E\,d\mu \int \chi Y\,d\nu$$

(by 253D)

$$= \iint \chi E(x) f(x, y)\nu(dy)\mu(dx) = \int f \times \chi(E \times Y) d\lambda$$

(by Fubini's theorem)

$$= \int_W f\,d\lambda.$$

So g is a conditional expectation of f.

253I This is a convenient moment to set out a useful result on products of indefinite-integral measures.

Proposition Let (X, Σ, μ) and (Y, T, ν) be measure spaces, and $f \in \mathcal{L}^0(\mu)$, $g \in \mathcal{L}^0(\nu)$ non-negative functions. Let μ', ν' be the corresponding indefinite-integral measures (see §234). Let λ be the c.l.d. product of μ and ν, and λ' the indefinite-integral measure defined from λ and $f \otimes g \in \mathcal{L}^0(\lambda)$ (253Cb). Then λ' is the c.l.d. product of μ' and ν'.

proof Write θ for the c.l.d. product of μ' and ν'.

(a) If we replace μ by its completion, we do not change μ' (234Ke); at the same time, we do not change λ, by 251T. The same applies to ν. So it will be enough to prove the result on the assumption that μ and ν are complete; in which case f and g are measurable and have measurable domains.

Set $F = \{x : x \in \operatorname{dom} f,\, f(x) > 0\}$ and $G = \{y : y \in \operatorname{dom} g,\, g(y) > 0\}$, so that $F \times G = \{w : w \in \operatorname{dom}(f \otimes g),\, (f \otimes g)(w) > 0\}$. Then F is μ'-conegligible and G is ν'-conegligible, so $F \times G$ is θ-conegligible as well as λ'-conegligible. Because both θ and λ' are complete (251Ic, 234I), it will be enough to show that the subspace measures $\theta_{F\times G}$, $\lambda'_{F\times G}$ on $F \times G$ are equal. But note that $\theta_{F\times G}$ can be identified with the product of μ'_F and ν'_G, where μ'_F and ν'_G are the subspace measures on F, G respectively (251Q(ii-α)). At the same time, μ'_F is the indefinite-integral measure defined from the subspace measure μ_F on F and the function $f{\restriction}F$, ν'_G is the indefinite-integral measure defined from the subspace measure ν_G on G and $g{\restriction}G$, and $\lambda'_{F\times G}$ is defined from the subspace measure $\lambda_{F\times G}$ and $(f{\restriction}F) \otimes (g{\restriction}G)$. Finally, by 251Q again, $\lambda_{F\times G}$ is the product of μ_F and ν_G.

What all this means is that it will be enough to deal with the case in which $F = X$ and $G = Y$, that is, f and g are everywhere defined and strictly positive; which is what I will suppose from now on.

(b) In this case $\operatorname{dom}\mu' = \Sigma$ and $\operatorname{dom}\nu' = \mathrm{T}$ (234La). Similarly, $\operatorname{dom}\lambda' = \Lambda$ is just the domain of λ. Set

$$F_n = \{x : x \in X,\, 2^{-n} \le f(x) \le 2^n\}, \quad G_n = \{y : y \in Y,\, 2^{-n} \le g(y) \le 2^n\}$$

for $n \in \mathbb{N}$.

(c) Set

$$\mathcal{A} = \{W : W \in \operatorname{dom}\theta \cap \operatorname{dom}\lambda',\, \theta(W) = \lambda'(W)\}.$$

If $\mu' E$ and $\nu' H$ are defined and finite, then $f \times \chi E$ and $g \times \chi H$ are integrable, so

$$\lambda'(E \times H) = \int (f \otimes g) \times \chi(E \times H) d\lambda = \int (f \times \chi E) \otimes (g \times \chi H) d\lambda$$
$$= \int f \times \chi E \, d\mu \cdot \int g \times \chi H \, d\nu = \theta(E \times H)$$

by 253D and 251Ia, that is, $E \times H \in \mathcal{A}$. If we now look at $\mathcal{A}_{EH} = \{W : W \subseteq X \times Y, \, W \cap (E \times H) \in \mathcal{A}\}$, then we see that

$\mathcal{A}_{EH}$ contains $E' \times H'$ for every $E' \in \Sigma$, $H' \in \mathrm{T}$,

if $\langle W_n \rangle_{n \in \mathbb{N}}$ is a non-decreasing sequence in $\mathcal{A}_{EH}$ then $\bigcup_{n \in \mathbb{N}} W_n \in \mathcal{A}_{EH}$,

if W, $W' \in \mathcal{A}_{EH}$ and $W \subseteq W'$ then $W' \setminus W \in \mathcal{A}_{EH}$.

Thus $\mathcal{A}_{EH}$ is a Dynkin class of subsets of $X \times Y$, and by the Monotone Class Theorem (136B) includes the σ-algebra generated by $\{E' \times H' : E' \in \Sigma, \, H' \in \mathrm{T}\}$, which is $\Sigma \widehat{\otimes} \mathrm{T}$.

(d) Now suppose that $W \in \Lambda$. In this case $W \in \operatorname{dom} \theta$ and $\theta W \le \lambda' W$. **P** Take $n \in \mathbb{N}$, and $E \in \Sigma$, $H \in \mathrm{T}$ such that $\mu' E$ and $\nu' H$ are both finite. Set $E' = E \cap F_n$, $H' = H \cap G_n$ and $W' = W \cap (E' \times H')$. Then $W' \in \Lambda$, while $\mu E' \le 2^n \mu' E$ and $\nu H' \le 2^n \nu' H$ are finite. By 251Ib there is a $V \in \Sigma \widehat{\otimes} \mathrm{T}$ such that $V \subseteq W'$ and $\lambda V = \lambda W'$. Similarly, there is a $V' \in \Sigma \widehat{\otimes} \mathrm{T}$ such that $V' \subseteq (E' \times H') \setminus W'$ and $\lambda V' = \lambda((E' \times H') \setminus W')$. This means that $\lambda((E' \times H') \setminus (V \cup V')) = 0$, so $\lambda'((E' \times H') \setminus (V \cup V')) = 0$. But $(E' \times H') \setminus (V \cup V') \in \mathcal{A}$, by (c), so $\theta((E' \times H') \setminus (V \cup V')) = 0$ and $W' \in \operatorname{dom} \theta$, while

$$\theta W' = \theta V = \lambda' V \le \lambda' W.$$

Since E and H are arbitrary, $W \cap (F_n \times G_n) \in \operatorname{dom} \theta$ (251H) and $\theta(W \cap (F_n \times G_n)) \le \lambda' W$. Since $\langle F_n \rangle_{n \in \mathbb{N}}$, $\langle G_n \rangle_{n \in \mathbb{N}}$ are non-decreasing sequences with unions X, Y respectively,

$$\theta W = \sup_{n \in \mathbb{N}} \theta(W \cap (F_n \times G_n)) \le \lambda' W. \ \mathbf{Q}$$

(e) In the same way, $\lambda' W$ is defined and less than or equal to θW for every $W \in \operatorname{dom} \theta$. **P** The arguments are very similar, but a refinement seems to be necessary at the last stage. Take $n \in \mathbb{N}$, and $E \in \Sigma$, $H \in \mathrm{T}$ such that μE and νH are both finite. Set $E' = E \cap F_n$, $H' = H \cap G_n$ and $W' = W \cap (E' \times H')$. Then $W' \in \operatorname{dom} \theta$, while $\mu' E' \le 2^n \mu E$ and $\nu' H' \le 2^n \nu H$ are finite. This time, there are V, $V' \in \Sigma \widehat{\otimes} \mathrm{T}$ such that $V \subseteq W'$, $V' \subseteq (E' \times H') \setminus W'$, $\theta V = \theta W'$ and $\theta V' = \theta((E' \times H') \setminus W')$. Accordingly

$$\lambda' V + \lambda' V' = \theta V + \theta V' = \theta(E' \times H') = \lambda'(E' \times H'),$$

so that $\lambda' W'$ is defined and equal to $\theta W'$.

What this means is that $W \cap (F_n \times G_n) \cap (E \times H) \in \mathcal{A}$ whenever μE and νH are finite. So $W \cap (F_n \times G_n) \in \Lambda$, by 251H; as n is arbitrary, $W \in \Lambda$ and $\lambda' W$ is defined.

? Suppose, if possible, that $\lambda' W > \theta W$. Then there is some $n \in \mathbb{N}$ such that $\lambda'(W \cap (F_n \times G_n)) > \theta W$. Because λ is semi-finite, 213B tells us that there is some λ-simple function h such that $h \le (f \otimes g) \times \chi(W \cap (F_n \times G_n))$ and $\int h \, d\lambda > \theta W$; setting $V = \{(x, y) : h(x, y) > 0\}$, we see that $V \subseteq W \cap (F_n \times G_n)$, λV is defined and finite and $\lambda' V > \theta W$. Now there must be sets $E \in \Sigma$, $H \in \mathrm{T}$ such that μE and νF are both finite and $\lambda(V \setminus (E \times H)) < 4^{-n}(\lambda' V - \theta W)$. But in this case $V \in \Lambda \subseteq \operatorname{dom} \theta$ (by (d)), so we can apply the argument just above to V and conclude that $V \cap (E \times H) = V \cap (F_n \times G_n) \cap (E \times H)$ belongs to $\mathcal{A}$. And now

$$\lambda' V = \lambda'(V \cap (E \times H)) + \lambda'(V \setminus (E \times H))$$
$$\le \theta(V \cap (E \times H)) + 4^n \lambda(V \setminus (E \times H)) < \theta V + \lambda' V - \theta W \le \lambda' V,$$

which is absurd. **X**

So $\lambda' W$ is defined and not greater than θW. **Q**

(f) Putting this together with (d), we see that $\lambda' = \theta$, as claimed.

Remark If μ' and ν' are totally finite, so that they are 'truly continuous' with respect to μ and ν in the sense of 232Ab, then f and g are integrable, so $f \otimes g$ is λ-integrable, and $\theta = \lambda'$ is truly continuous with respect to λ.

The proof above can be simplified using fragments of the general theory of complete locally determined spaces, which will be given in §412 in Volume 4.

***253J Upper integrals** The idea of 253D can be repeated in terms of upper integrals, as follows.

Proposition Let (X, Σ, μ) and (Y, T, ν) be σ-finite measure spaces, with c.l.d. product measure λ. Then for any functions f and g, defined on conegligible subsets of X and Y respectively, and taking values in $[0, \infty]$,

$$\overline{\int} f \otimes g \, d\lambda = \overline{\int} f d\mu \cdot \overline{\int} g \, d\nu.$$

Remark Here $(f \otimes g)(x, y) = f(x)g(y)$ for all $x \in \operatorname{dom} f$ and $y \in \operatorname{dom} g$, taking $0 \cdot \infty = 0$, as in §135.

proof (a) I show first that $\overline{\int} f \otimes g \le \overline{\int} f \overline{\int} g$. **P** If $\overline{\int} f = 0$, then $f = 0$ a.e., so $f \otimes g = 0$ a.e. and the result is immediate. The same argument applies if $\overline{\int} g = 0$. If both $\overline{\int} f$ and $\overline{\int} g$ are non-zero, and either is infinite, the result is trivial. So let us suppose that both are finite. In this case there are integrable f_0, g_0 such that $f \le_{\text{a.e.}} f_0$, $g \le_{\text{a.e.}} g_0$, $\overline{\int} f = \int f_0$ and $\overline{\int} g = \int g_0$ (133Ja/135Ha). So $f \otimes g \le_{\text{a.e.}} f_0 \otimes g_0$, and

$$\overline{\int} f \otimes g \le \int f_0 \otimes g_0 = \int f_0 \int g_0 = \overline{\int} f \overline{\int} g,$$

by 253D. **Q**

(b) For the reverse inequality, we need consider only the case in which $\overline{\int} f \otimes g$ is finite, so that there is a λ-integrable function h such that $f \otimes g \le_{\text{a.e.}} h$ and $\overline{\int} f \otimes g = \int h$. Set

$$f_0(x) = \int h(x, y) \nu(dy)$$

whenever this is defined in $\mathbb{R}$, which is almost everywhere, by Fubini's theorem (252B-252C). Then $f_0(x) \ge f(x) \overline{\int} g \, d\nu$ for every $x \in \operatorname{dom} f_0 \cap \operatorname{dom} f$, which is a conegligible set in X; so

$$\overline{\int} f \otimes g = \int h \, d\lambda = \int f_0 d\mu \ge \overline{\int} f \overline{\int} g,$$

as required.

***253K** A similar argument applies to upper integrals of sums, as follows.

Proposition Let (X, Σ, μ) and (Y, T, ν) be probability spaces, with c.l.d. product measure λ. Then for any real-valued functions f, g defined on conegligible subsets of X, Y respectively,

$$\overline{\int} f(x) + g(y) \, \lambda(d(x, y)) = \overline{\int} f(x) \mu(dx) + \overline{\int} g(y) \nu(dy),$$

at least when the right-hand side is defined in $[-\infty, \infty]$.

proof Set $h(x, y) = f(x) + g(y)$ for $x \in \operatorname{dom} f$ and $y \in \operatorname{dom} g$, so that $\operatorname{dom} h$ is λ-conegligible.

(a) As in 253J, I start by showing that $\overline{\int} h \le \overline{\int} f + \overline{\int} g$. **P** If either $\overline{\int} f$ or $\overline{\int} g$ is ∞, this is trivial. Otherwise, take integrable functions f_0, g_0 such that $f \le_{\text{a.e.}} f_0$ and $g \le_{\text{a.e.}} g_0$. Set $h_0 = (f_0 \otimes \chi Y) + (\chi X \otimes g_0)$; then $h \le h_0$ λ-a.e., so

$$\overline{\int} h \, d\lambda \le \int h_0 d\lambda = \int f_0 d\mu + \int g_0 d\nu.$$

As f_0, g_0 are arbitrary, $\overline{\int} h \le \overline{\int} f + \overline{\int} g$. **Q**

(b) For the reverse inequality, suppose that $h \le h_0$ for λ-almost every (x, y), where h_0 is λ-integrable. Set $f_0(x) = \int h_0(x, y) \nu(dy)$ whenever this is defined in $\mathbb{R}$. Then $f_0(x) \ge f(x) + \overline{\int} g \, d\nu$ whenever $x \in \operatorname{dom} f \cap \operatorname{dom} f_0$, so

$$\int h_0 \, d\lambda = \int f_0 d\mu \ge \overline{\int} f d\mu + \overline{\int} g \, d\nu.$$

As h_0 is arbitrary, $\overline{\int} h \ge \overline{\int} f + \overline{\int} g$, as required.

253L Complex spaces As usual, the ideas of 253F and 253H apply essentially unchanged to complex L^1 spaces. Writing $L^1_{\mathbb{C}}(\mu)$, etc., for the complex L^1 spaces involved, we have the following results. Throughout, let (X, Σ, μ) and (Y, T, ν) be measure spaces, and λ the c.l.d. product measure on $X \times Y$.

(a) If $f \in \mathcal{L}^0_{\mathbb{C}}(\mu)$ and $g \in \mathcal{L}^0_{\mathbb{C}}(\nu)$ then $f \otimes g$, defined by the formula $(f \otimes g)(x, y) = f(x)g(y)$ for $x \in \operatorname{dom} f$ and $y \in \operatorname{dom} g$, belongs to $\mathcal{L}^0_{\mathbb{C}}(\lambda)$.

(b) If $f \in \mathcal{L}^1_{\mathbb{C}}(\mu)$ and $g \in \mathcal{L}^1_{\mathbb{C}}(\nu)$ then $f \otimes g \in \mathcal{L}^1_{\mathbb{C}}(\lambda)$ and $\int f \otimes g \, d\lambda = \int f d\mu \int g \, d\nu$.

(c) We have a bilinear operator $(u, v) \mapsto u \otimes v : L^1_{\mathbb{C}}(\mu) \times L^1_{\mathbb{C}}(\nu) \to L^1_{\mathbb{C}}(\lambda)$ defined by writing $f^\bullet \otimes g^\bullet = (f \otimes g)^\bullet$ for all $f \in \mathcal{L}^1_{\mathbb{C}}(\mu)$, $g \in \mathcal{L}^1_{\mathbb{C}}(\nu)$.

(d) If W is any complex Banach space and $\phi : L^1_{\mathbb{C}}(\mu) \times L^1_{\mathbb{C}}(\nu) \to W$ is any bounded bilinear operator, then there is a unique bounded linear operator $T : L^1_{\mathbb{C}}(\lambda) \to W$ such that $T(u \otimes v) = \phi(u, v)$ for every $u \in L^1_{\mathbb{C}}(\mu)$ and $v \in L^1_{\mathbb{C}}(\nu)$, and $\|T\| = \|\phi\|$.

(e) If μ and ν are complete probability measures, and $\Lambda_1 = \{E \times Y : E \in \Sigma\}$, then for any $f \in \mathcal{L}^1_{\mathbb{C}}(\lambda)$ we have a conditional expectation g of f on Λ_1 given by setting $g(x, y) = \int f(x, z)\nu(dz)$ whenever this is defined.

253X Basic exercises >(a) Let U, V and W be linear spaces. Show that the set of bilinear operators from $U \times V$ to W has a natural linear structure agreeing with those of $\mathsf{L}(U; \mathsf{L}(V; W))$ and $\mathsf{L}(V; \mathsf{L}(U; W))$, writing $\mathsf{L}(U; W)$ for the linear space of linear operators from U to W.

>(b) Let U, V and W be normed spaces. (i) Show that for a bilinear operator $\phi : U \times V \to W$ the following are equiveridical: (α) ϕ is bounded in the sense of 253Ab; (β) ϕ is continuous; (γ) ϕ is continuous at some point of $U \times V$. (ii) Show that the space of bounded bilinear operators from $U \times V$ to W is a linear subspace of the space of all bilinear operators from $U \times V$ to W, and that the functional $\| \|$ defined in 253Ab is a norm, agreeing with the norms of $\mathsf{B}(U; \mathsf{B}(V; W))$ and $\mathsf{B}(V; \mathsf{B}(U; W))$, writing $\mathsf{B}(U; W)$ for the normed space of bounded linear operators from U to W.

(c) Let $(X_1, \Sigma_1, \mu_1), \dots, (X_n, \Sigma_n, \mu_n)$ be measure spaces, and λ the c.l.d. product measure on $X_1 \times \dots \times X_n$, as described in 251W. Let W be a Banach space, and suppose that $\phi : L^1(\mu_1) \times \dots \times L^1(\mu_n) \to W$ is **multilinear** (that is, linear in each variable separately) and **bounded** (that is, $\|\phi\| = \sup\{\phi(u_1, \dots, u_n) : \|u_i\|_1 \le 1\ \forall\ i \le n\} < \infty$). Show that there is a unique bounded linear operator $T : L^1(\lambda) \to W$ such that $T\otimes = \phi$, where $\otimes : L^1(\mu_1) \times \dots \times L^1(\mu_n) \to L^1(\lambda)$ is a canonical multilinear operator (to be defined).

(d) Let (X, Σ, μ) and (Y, T, ν) be measure spaces, and λ the c.l.d. product measure on $X \times Y$. Show that if $A \subseteq L^1(\mu)$ and $B \subseteq L^1(\nu)$ are both uniformly integrable, then $\{u \otimes v : u \in A, v \in B\}$ is uniformly integrable in $L^1(\lambda)$.

>(e) Let (X, Σ, μ) and (Y, T, ν) be measure spaces and λ the c.l.d. product measure on $X \times Y$. Show that

(i) we have a bilinear operator $(u, v) \mapsto u \otimes v : L^0(\mu) \times L^0(\nu) \to L^0(\lambda)$ given by setting $f^\bullet \otimes g^\bullet = (f \otimes g)^\bullet$ for all $f \in \mathcal{L}^0(\mu)$ and $g \in \mathcal{L}^0(\nu)$;

(ii) if $1 \le p \le \infty$ then $u \otimes v \in L^p(\lambda)$ and $\|u \otimes v\|_p = \|u\|_p \|v\|_p$ for all $u \in L^p(\mu)$ and $v \in L^p(\nu)$;

(iii) if u, $u' \in L^2(\mu)$ and v, $v' \in L^2(\nu)$ then the inner product $(u \otimes v | u' \otimes v')$, taken in $L^2(\lambda)$, is just $(u|u')(v|v')$;

(iv) the map $(u, v) \mapsto u \otimes v : L^0(\mu) \times L^0(\nu) \to L^0(\lambda)$ is continuous if $L^0(\mu)$, $L^0(\nu)$ and $L^0(\lambda)$ are all given their topologies of convergence in measure.

(f) In 253Xe, assume that μ and ν are semi-finite. Show that if $u_0, \dots, u_n$ are linearly independent members of $L^0(\mu)$ and $v_0, \dots, v_n \in L^0(\nu)$ are not all 0, then $\sum_{i=0}^n u_i \otimes v_i \ne 0$ in $L^0(\lambda)$. (*Hint*: start by finding sets $E \in \Sigma$, $F \in \mathrm{T}$ of finite measure such that $u_0 \times \chi E^\bullet, \dots, u_n \times \chi E^\bullet$ are linearly independent and $v_0 \times \chi F^\bullet, \dots, v_n \times \chi F^\bullet$ are not all 0.)

(g) In 253Xe, assume that μ and ν are semi-finite. If U, V are linear subspaces of $L^0(\mu)$ and $L^0(\nu)$ respectively, write $U \otimes V$ for the linear subspace of $L^0(\lambda)$ generated by $\{u \otimes v : u \in U, v \in V\}$. Show that if W is any linear space and $\phi : U \times V \to W$ is a bilinear operator, there is a unique linear operator $T : U \otimes V \to W$ such that $T(u \otimes v) = \phi(u, v)$ for all $u \in U$, $v \in V$. (*Hint*: start by showing that if $u_0, \dots, u_n \in U$ and $v_0, \dots, v_n \in V$ are such that $\sum_{i=0}^n u_i \otimes v_i = 0$, then $\sum_{i=0}^n \phi(u_i, v_i) = 0$ – do this by expressing the u_i as linear combinations of some linearly independent family and applying 253Xf.)

>(h) Let (X, Σ, μ) and (Y, T, ν) be complete probability spaces, with c.l.d. product measure λ. Suppose that $p \in [1, \infty]$ and that $f \in \mathcal{L}^p(\lambda)$. Set $g(x) = \int f(x, y)\nu(dy)$ whenever this is defined. Show that $g \in \mathcal{L}^p(\mu)$ and that $\|g\|_p \le \|f\|_p$. (*Hint*: 253H, 244M.)

(i) Let (X, Σ, μ) and (Y, T, ν) be measure spaces, with c.l.d. product measure λ, and $p \in [1, \infty[$. Show that $\{w : w \in L^p(\lambda), w \ge 0\}$ is the closed convex hull in $L^p(\lambda)$ of $\{u \otimes v : u \in L^p(\mu), v \in L^p(\nu), u \ge 0, v \ge 0\}$ (see 253Xe(ii) above).

253Y Further exercises (a) Let (X, Σ, μ) and (Y, T, ν) be measure spaces, and λ_0 the primitive product measure on $X \times Y$. Show that if $f \in \mathcal{L}^0(\mu)$ and $g \in \mathcal{L}^0(\nu)$, then $f \otimes g \in \mathcal{L}^0(\lambda_0)$.

(b) Let (X, Σ, μ) and (Y, T, ν) be measure spaces, and λ_0 the primitive product measure on $X \times Y$. Show that if $f \in \mathcal{L}^1(\mu)$ and $g \in \mathcal{L}^1(\nu)$, then $f \otimes g \in \mathcal{L}^1(\lambda_0)$ and $\int f \otimes g \, d\lambda_0 = \int f \, d\mu \int g \, d\nu$.

(c) Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and λ_0, λ the primitive and c.l.d. product measures on $X\times Y$. Show that the embedding $\mathcal{L}^1(\lambda_0)\subsetneq\mathcal{L}^1(\lambda)$ induces a Banach lattice isomorphism between $L^1(\lambda_0)$ and $L^1(\lambda)$.

(d) Let (X,Σ,μ), (Y,T,ν) be strictly localizable measure spaces, with c.l.d. product measure λ. Show that $L^\infty(\lambda)$ can be identified with $L^1(\lambda)^*$. Show that under this identification $\{w:w\in L^\infty(\lambda),\,w\ge0\}$ is the weak*-closed convex hull of $\{u\otimes v:u\in L^\infty(\mu),\,v\in L^\infty(\nu),\,u\ge0,\,v\ge0\}$.

(e) Find a version of 253J valid when one of μ, ν is not σ-finite.

(f) Let (X,Σ,μ) be any measure space and V any Banach space. Write $\mathcal{L}^1_V=\mathcal{L}^1_V(\mu)$ for the set of functions f such that (α) $\operatorname{dom}f$ is a conegligible subset of X (β) f takes values in V (γ) there is a conegligible set $D\subseteq\operatorname{dom}f$ such that $f[D]$ is separable and $D\cap f^{-1}[G]\in\Sigma$ for every open set $G\subseteq V$ (δ) the integral $\int\|f(x)\|\mu(dx)$ is finite. (These are the **Bochner integrable** functions from X to V.) For f, $g\in\mathcal{L}^1_V$ write $f\sim g$ if $f=g$ μ-a.e.; let L^1_V be the set of equivalence classes in $\mathcal{L}^1_V$ under $\sim$. Show that

(i) $f+g$, $cf\in\mathcal{L}^1_V$ for all f, $g\in\mathcal{L}^1_V$, $c\in\mathbb{R}$;

(ii) L^1_V has a natural linear space structure, defined by writing $f^\bullet+g^\bullet=(f+g)^\bullet$, $cf^\bullet=(cf)^\bullet$ for f, $g\in\mathcal{L}^1_V$ and $c\in\mathbb{R}$;

(iii) L^1_V has a norm $\|\;\|$, defined by writing $\|f^\bullet\|=\int\|f(x)\|\mu(dx)$ for $f\in\mathcal{L}^1_V$;

(iv) L^1_V is a Banach space under this norm;

(v) there is a natural map $\otimes:\mathcal{L}^1\times V\to\mathcal{L}^1_V$ defined by writing $(f\otimes v)(x)=f(x)v$ when $f\in\mathcal{L}^1=\mathcal{L}^1_{\mathbb{R}}(\mu)$, $v\in V$ and $x\in\operatorname{dom}f$;

(vi) there is a canonical bilinear operator $\otimes:L^1\times V\to L^1_V$ defined by writing $f^\bullet\otimes v=(f\otimes v)^\bullet$ for $f\in\mathcal{L}^1$ and $v\in V$;

(vii) whenever W is a Banach space and $\phi:L^1\times V\to W$ is a bounded bilinear operator, there is a unique bounded linear operator $T:L^1_V\to W$ such that $T(u\otimes v)=\phi(u,v)$ for all $u\in L^1$ and $v\in V$, and $\|T\|=\|\phi\|$. (When $W=V$ and $\phi(u,v)=(\int u)v$ for $u\in L^1$ and $v\in V$, $Tf^\bullet$ is called the **Bochner integral** of f.)

(g) Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and λ_0 the primitive product measure on $X\times Y$. If f is a λ_0-integrable function, write $f_x(y)=f(x,y)$ whenever this is defined. Show that we have a map $x\mapsto f_x^\bullet$ from a conegligible subset D_0 of X to $L^1(\nu)$. Show that this map is a Bochner integrable function, as defined in 253Yf, and that its Bochner integral is $\int f d\lambda_0$.

(h) Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and suppose that ϕ is a function from X to a separable subset of $L^1(\nu)$ which is measurable in the sense that $\phi^{-1}[G]\in\Sigma$ for every open $G\subseteq L^1(\nu)$. Show that there is a Λ-measurable function f from $X\times Y$ to $\mathbb{R}$, where Λ is the domain of the c.l.d. product measure on $X\times Y$, such that $\phi(x)=f_x^\bullet$ for every $x\in X$, writing $f_x(y)=f(x,y)$ for $x\in X$, $y\in Y$.

(i) Let (X,Σ,μ) and (Y,T,ν) be measure spaces, and λ the c.l.d. product measure on $X\times Y$. Show that 253Yg provides a canonical identification between $L^1(\lambda)$ and $L^1_{L^1(\nu)}(\mu)$.

(j) Let (X,Σ,μ) and (Y,T,ν) be complete locally determined measure spaces, with c.l.d. product measure λ. (i) Suppose that $K\in\mathcal{L}^2(\lambda)$, $f\in\mathcal{L}^2(\mu)$. Show that $h(y)=\int K(x,y)f(x)dx$ is defined for almost all $y\in Y$ and that $h\in\mathcal{L}^2(\nu)$. (*Hint*: to see that h is defined a.e., consider $\int_{E\times F}K(x,y)f(x)d(x,y)$ for μE, $\nu F<\infty$; to see that $h\in\mathcal{L}^2$ consider $\int h\times g$ where $g\in\mathcal{L}^2(\nu)$.) (ii) Show that the map $f\mapsto h$ corresponds to a bounded linear operator $T_K:L^2(\mu)\to L^2(\nu)$. (iii) Show that the map $K\mapsto T_K$ corresponds to a bounded linear operator, of norm at most 1, from $L^2(\lambda)$ to $\mathsf{B}(L^2(\mu);L^2(\nu))$.

(k) Suppose that p, $q\in[1,\infty]$ and that $\frac{1}{p}+\frac{1}{q}=1$, interpreting $\frac{1}{\infty}$ as 0 as usual. Let (X,Σ,μ), (Y,T,ν) be complete locally determined measure spaces with c.l.d. product measure λ. Show that the ideas of 253Yj can be used to define a bounded linear operator, of norm at most 1, from $L^p(\lambda)$ to $\mathsf{B}(L^q(\mu);L^p(\nu))$.

(l) In 253Xc, suppose that W is a Banach lattice. Show that the following are equiveridical: (i) $Tu\ge0$ whenever $u\in L^1(\lambda)$; (ii) $\phi(u_1,\dots,u_n)\ge0$ whenever $u_i\ge0$ in $L^1(\mu_i)$ for each $i\le n$.

253 Notes and comments Throughout the main arguments of this section, I have written the results in terms of the c.l.d. product measure; of course the isomorphism noted in 253Yc means that they could just as well have been expressed in terms of the primitive product measure. The more restricted notion of integrability with respect to the primitive product measure is indeed the one appropriate for the ideas of 253Yg.

Theorem 253F is a 'universal mapping theorem'; it asserts that every bounded bilinear operator on $L^1(\mu) \times L^1(\nu)$ factors through $\otimes : L^1(\mu) \times L^1(\nu) \to L^1(\lambda)$, at least if the range space is a Banach space. It is easy to see that this property defines the pair $(L^1(\lambda), \otimes)$ up to Banach space isomorphism, in the following sense: if V is a Banach space, and $\psi : L^1(\mu) \times L^1(\nu) \to V$ is a bounded bilinear operator such that for every bounded bilinear operator ϕ from $L^1(\mu) \times L^1(\nu)$ to any Banach space W there is a unique bounded linear operator $T : V \to W$ such that $T\psi = \phi$ and $\|T\| = \|\phi\|$, then there is an isometric Banach space isomorphism $S : L^1(\lambda) \to V$ such that $S\otimes = \psi$. There is of course a general theory of bilinear operators between Banach spaces; in the language of this theory, $L^1(\lambda)$ is, or is isomorphic to, the 'projective tensor product' of $L^1(\mu)$ and $L^1(\nu)$. For an introduction to this subject, see Defant & Floret 93, §I.3, or Semadeni 71, §20. I should perhaps emphasise, for the sake of those who have not encountered tensor products before, that this theorem is special to L^1 spaces. While some of the same ideas can be applied to other function spaces (see 253Xe-253Xg), there is no other class to which 253F applies.

There is also a theory of tensor products of Banach lattices, for which I do not think we are quite ready (it needs general ideas about ordered linear spaces for which I mean to wait until Chapter 35 in the next volume). However 253G shows that the ordering, and therefore the Banach lattice structure, of $L^1(\lambda)$ is determined by the ordering of $L^1(\mu)$ and $L^1(\nu)$ and the map $\otimes : L^1(\mu) \times L^1(\nu) \to L^1(\lambda)$.

The conditional expectation operators described in 253H are of very great importance, largely because in this special context we have a realization of the conditional expectation operator as a function P_0 from $\mathcal{L}^1(\lambda)$ to $\mathcal{L}^1(\lambda\restriction\Lambda_1)$, not just as a function from $L^1(\lambda)$ to $L^1(\lambda\restriction\Lambda_1)$, as in 242J. As described here, $P_0(f + f')$ need not be equal, in the strict sense, to $P_0 f + P_0 f'$; it can have a larger domain. In applications, however, one might be willing to restrict attention to the linear space $\mathfrak{U}$ of bounded $\Sigma\widehat{\otimes}\mathrm{T}$-measurable functions defined everywhere on $X \times Y$, so that P_0 becomes an operator from $\mathfrak{U}$ to itself (see 252P).

254 Infinite products

I come now to the second basic idea of this chapter: the description of a product measure on the product of a (possibly large) family of probability spaces. The section begins with a construction on similar lines to that of §251 (254A-254F) and its defining property in terms of inverse-measure-preserving functions (254G). I discuss the usual measure on $\{0,1\}^I$ (254J-254K), subspace measures (254L) and various properties of subproducts (254M-254T), including a study of the associated conditional expectation operators (254R-254T).

254A Definitions (a) Let $\langle(X_i, \Sigma_i, \mu_i)\rangle_{i\in I}$ be a family of probability spaces. Set $X = \prod_{i\in I} X_i$, the family of functions x with domain I such that $x(i) \in X_i$ for every $i \in I$. In this context, I will say that a **measurable cylinder** is a subset of X expressible in the form

$$C = \prod_{i\in I} C_i,$$

where $C_i \in \Sigma_i$ for every $i \in I$ and $\{i : C_i \neq X_i\}$ is finite. Note that for a non-empty $C \subseteq X$ this expression is unique. **P** Suppose that $C = \prod_{i\in I} C_i = \prod_{i\in I} C'_i$. For each $i \in I$ set

$$D_i = \{x(i) : x \in C\}.$$

Of course $D_i \subseteq C_i$. Because $C \neq \emptyset$, we can fix on some $z \in C$. If $i \in I$ and $\xi \in C_i$, consider $x \in X$ defined by setting

$$x(i) = \xi, \quad x(j) = z(j) \text{ for } j \neq i;$$

then $x \in C$ so $\xi = x(i) \in D_i$. Thus $D_i = C_i$ for $i \in I$. Similarly, $D_i = C'_i$. **Q**

(b) We can therefore define a functional $\theta_0 : \mathcal{C} \to [0,1]$, where $\mathcal{C}$ is the set of measurable cylinders, by setting

$$\theta_0 C = \prod_{i\in I} \mu_i C_i$$

whenever $C_i \in \Sigma_i$ for every $i \in I$ and $\{i : C_i \neq X_i\}$ is finite, noting that only finitely many terms in the product can differ from 1, so that it can safely be treated as a finite product. If $C = \emptyset$, one of the C_i must be empty, so $\theta_0 C$ is surely 0, even though the expression of C as $\prod_{i\in I} C_i$ is no longer unique.

(c) Now define $\theta : \mathcal{P}X \to [0,1]$ by setting

$$\theta A = \inf\{\sum_{n=0}^\infty \theta_0 C_n : C_n \in \mathcal{C} \text{ for every } n \in \mathbb{N},\ A \subseteq \bigcup_{n\in\mathbb{N}} C_n\}.$$

254B Lemma The functional θ defined in 254Ac is always an outer measure on X.

proof Use exactly the same arguments as those in 251B above.

254C Definition Let $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be any indexed family of probability spaces, and X the Cartesian product $\prod_{i\in I}X_i$. The **product measure** on X is the measure defined by Carathéodory's method (113C) from the outer measure θ defined in 254A.

254D Remarks (a) In 254Ab, I asserted that if $C\in\mathcal{C}$ and no C_i is empty, then nor is $C=\prod_{i\in I}C_i$. This is the 'Axiom of Choice': the product of any family $\langle C_i\rangle_{i\in I}$ of non-empty sets is non-empty, that is, there is a 'choice function' x with domain I picking out a distinguished member $x(i)$ of each C_i. In this volume I have not attempted to be scrupulous in indicating uses of the axiom of choice. In fact the use here is not an absolutely vital one; I mean, the theory of infinite products, even uncountable products, of probability spaces does not change character completely in the absence of the full axiom of choice (provided, that is, that we allow ourselves to use the countable axiom of choice). The point is that all we really need, in the present context, is that $X=\prod_{i\in I}X_i$ should be non-empty; and in many contexts we can prove this, for the particular cases of interest, without using the axiom of choice, by actually exhibiting a member of X. The simplest case in which this is difficult is when the X_i are uncontrolled Borel subsets of $[0,1]$; and even then, if they are presented with coherent descriptions, we may, with appropriate labour, be able to construct a member of X. But clearly such a process is liable to slow us down a good deal, and for the moment I think there is no great virtue in taking so much trouble.

(b) I have given this section the title 'infinite products', but it is useful to be able to apply the ideas to finite I; I should mention in particular the cases $\#(I)\le 2$.

(i) If $I=\emptyset$, X consists of the unique function with domain I, the empty function. If we identify a function with its graph, then X is actually $\{\emptyset\}$; in any case, X is to be a singleton set, with $\lambda X=1$.

(ii) If I is a singleton $\{i\}$, then we can identify X with X_i; $\mathcal{C}$ becomes identified with Σ_i and θ_0 with μ_i, so that θ can be identified with μ_i^* and the 'product measure' becomes the measure on X_i defined from μ_i^*, that is, the completion of μ_i (213Xa(iv)).

(iii) If I is a doubleton $\{i,j\}$, then we can identify X with $X_i\times X_j$; in this case the definitions of 254A and 254C match exactly with those of 251A and 251C, so that λ here can be identified with the primitive product measure as defined in 251C. Because μ_i and μ_j are both totally finite, this agrees with the c.l.d. product measure of 251F.

254E Definition Let $\langle X_i\rangle_{i\in I}$ be any family of sets, and $X=\prod_{i\in I}X_i$. If Σ_i is a σ-subalgebra of subsets of X_i for each $i\in I$, I write $\widehat{\bigotimes}_{i\in I}\Sigma_i$ for the σ-algebra of subsets of X generated by

$$\{\{x:x\in X,\,x(i)\in E\}:i\in I,\,E\in\Sigma_i\}.$$

(Compare 251D.)

254F Theorem Let $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be a family of probability spaces, and let λ be the product measure on $X=\prod_{i\in I}X_i$ defined as in 254C; let Λ be its domain.

(a) $\lambda X=1$.

(b) If $E_i\in\Sigma_i$ for every $i\in I$, and $\{i:E_i\ne X_i\}$ is countable, then $\prod_{i\in I}E_i\in\Lambda$, and $\lambda(\prod_{i\in I}E_i)=\prod_{i\in I}\mu_iE_i$. In particular, $\lambda C=\theta_0C$ for every measurable cylinder C, as defined in 254A, and if $j\in I$ then $x\mapsto x(j):X\to X_j$ is inverse-measure-preserving.

(c) $\widehat{\bigotimes}_{i\in I}\Sigma_i\subseteq\Lambda$.

(d) λ is complete.

(e) For every $W\in\Lambda$ and $\epsilon>0$ there is a finite family $C_0,\dots,C_n$ of measurable cylinders such that $\lambda(W\triangle\bigcup_{k\le n}C_k)\le\epsilon$.

(f) For every $W\in\Lambda$ there are $W_1,\,W_2\in\widehat{\bigotimes}_{i\in I}\Sigma_i$ such that $W_1\subseteq W\subseteq W_2$ and $\lambda(W_2\setminus W_1)=0$.

Remark Perhaps I should pause to interpret the product $\prod_{i\in I}\mu_iE_i$. Because all the μ_iE_i belong to $[0,1]$, this is simply $\inf_{J\subseteq I,J\text{ is finite}}\prod_{i\in J}\mu_iE_i$, taking the empty product to be 1.

proof Throughout this proof, define $\mathcal{C}$, θ_0 and θ as in 254A. I will write out an argument which applies to finite I as well as infinite I, but you may reasonably prefer to assume that I is infinite on first reading.

(a) Of course $\lambda X=\theta X$, so I have to show that $\theta X=1$. Because $X,\,\emptyset\in\mathcal{C}$ and $\theta_0X=\prod_{i\in I}\mu_iX_i=1$ and $\theta_0\emptyset=0$,

$$\theta X\le\theta_0X+\theta_0\emptyset+\ldots=1.$$

I therefore have to show that $\theta X\ge 1$. **?** Suppose, if possible, otherwise.

(i) There is a sequence $\langle C_n\rangle_{n\in\mathbb{N}}$ in $\mathcal{C}$, covering X, such that $\sum_{n=0}^{\infty}\theta_0C_n<1$. For each $n\in\mathbb{N}$, express C_n as $\{x:x(i)\in E_{ni}\ \forall\ i\in I\}$, where every $E_{ni}\in\Sigma_i$ and $J_n=\{i:E_{ni}\ne X_i\}$ is finite. No J_n can be empty, because $\theta_0C_n<1=\theta_0X$; set $J=\bigcup_{n\in\mathbb{N}}J_n$. Then J is a countable non-empty subset of I. Set $K=\mathbb{N}$ if J is infinite, $\{k:0\le k<\#(J)\}$ if J is finite; let $k\mapsto i_k:K\to J$ be a bijection.

For each $k\in K$, set $L_k=\{i_j:j<k\}\subseteq J$, and set $\alpha_{nk}=\prod_{i\in I\setminus L_k}\mu_iE_{ni}$ for $n\in\mathbb{N}$, $k\in K$. If J is finite, then we can identify $L_{\#(J)}$ with J, and set $\alpha_{n,\#(J)}=1$ for every n. We have $\alpha_{n0}=\theta_0C_n$ for each n, so $\sum_{n=0}^{\infty}\alpha_{n0}<1$. For $n\in\mathbb{N}$, $k\in K$ and $t\in X_{i_k}$ set

$$\begin{aligned}f_{nk}(t)&=\alpha_{n,k+1}\text{ if }t\in E_{n,i_k},\\&=0\text{ otherwise.}\end{aligned}$$

Then

$$\int f_{nk}d\mu_{i_k}=\alpha_{n,k+1}\mu_{i_k}E_{n,i_k}=\alpha_{nk}.$$

(ii) Choose $t_k\in X_{i_k}$ inductively, for $k\in K$, as follows. The inductive hypothesis will be that $\sum_{n\in M_k}\alpha_{nk}<1$, where $M_k=\{n:n\in\mathbb{N},\ t_j\in E_{n,i_j}\ \forall\ j<k\}$; of course $M_0=\mathbb{N}$, so the induction starts. Given that

$$1>\sum_{n\in M_k}\alpha_{nk}=\sum_{n\in M_k}\int f_{nk}d\mu_{i_k}=\int(\sum_{n\in M_k}f_{nk})d\mu_{i_k}$$

(by B.Levi's theorem), there must be a $t_k\in X_{i_k}$ such that $\sum_{n\in M_k}f_{nk}(t_k)<1$. Now for such a choice of t_k, $\alpha_{n,k+1}=f_{nk}(t_k)$ for every $n\in M_{k+1}$, so that $\sum_{n\in M_{k+1}}\alpha_{n,k+1}<1$, and the induction continues, unless J is finite and $k+1=\#(J)$. In this last case we must just have $M_{\#(J)}=\emptyset$, because $\alpha_{n,\#(J)}=1$ for every n.

(iii) If J is infinite, we obtain a full sequence $\langle t_k\rangle_{k\in\mathbb{N}}$; if J is finite, we obtain just a finite sequence $\langle t_k\rangle_{k<\#(J)}$. In either case, there is an $x\in X$ such that $x(i_k)=t_k$ for each $k\in K$. Now there must be some $m\in\mathbb{N}$ such that $x\in C_m$. Because $J_m=\{i:E_{mi}\ne X_i\}$ is finite, there is a $k\in\mathbb{N}$ such that $J_m\subseteq L_k$ (allowing $k=\#(J)$ if J is finite). Now $m\in M_k$, so in fact we cannot have $k=\#(J)$, and $\alpha_{mk}=1$, so $\sum_{n\in M_k}\alpha_{nk}\ge 1$, contrary to the inductive hypothesis. **X**

This contradiction shows that $\theta X=1$.

(b)(i) I take the particular case first. Let $j\in I$ and $E\in\Sigma_j$, and let $C\in\mathcal{C}$; set $W=\{x:x\in X,\ x(j)\in E\}$; then $C\cap W$ and $C\setminus W$ both belong to $\mathcal{C}$, and $\theta_0C=\theta_0(C\cap W)+\theta_0(C\setminus W)$. **P** If $C=\prod_{i\in I}C_i$, where $C_i\in\Sigma_i$ for each i, then $C\cap W=\prod_{i\in I}C_i'$, where $C_i'=C_i$ if $i\ne j$, and $C_j'=C_j\cap E$; similarly, $C\setminus W=\prod_{i\in I}C_i''$, where $C_i''=C_i$ if $i\ne j$, and $C_j''=C_j\setminus E$. So both belong to $\mathcal{C}$, and

$$\theta_0(C\cap W)+\theta_0(C\setminus W)=(\mu_j(C_j\cap E)+\mu_j(C_j\setminus E))\prod_{i\ne j}\mu C_i=\prod_{i\in I}\mu C_i=\theta_0C.\ \mathbf{Q}$$

(ii) Now suppose that $A\subseteq X$ is any set, and $\epsilon>0$. Then there is a sequence $\langle C_n\rangle_{n\in\mathbb{N}}$ in $\mathcal{C}$ such that $A\subseteq\bigcup_{n\in\mathbb{N}}C_n$ and $\sum_{n=0}^{\infty}\theta_0C_n\le\theta A+\epsilon$. In this case

$$A\cap W\subseteq\bigcup_{n\in\mathbb{N}}C_n\cap W,\quad A\setminus W\subseteq\bigcup_{n\in\mathbb{N}}C_n\setminus W,$$

so

$$\theta(A\cap W)\le\sum_{n=0}^{\infty}\theta_0(C_n\cap W),\quad\theta(A\setminus W)\le\sum_{n=0}^{\infty}\theta_0(C_n\setminus W),$$

and

$$\theta(A\cap W)+\theta(A\setminus W)\le\sum_{n=0}^{\infty}\theta_0(C_n\cap W)+\theta_0(C_n\setminus W)=\sum_{n=0}^{\infty}\theta_0C_n\le\theta A+\epsilon.$$

As ϵ is arbitrary, $\theta(A\cap W)+\theta(A\setminus W)\le\theta A$; as A is arbitrary, $W\in\Lambda$.

(iii) I show next that if $J\subseteq I$ is finite and $C_i\in\Sigma_i$ for each $i\in J$, and $C=\{x:x\in X,\ x(i)\in C_i\ \forall\ i\in J\}$, then $C\in\Lambda$ and $\lambda C=\prod_{i\in J}\mu_iC_i$. **P** Induce on $\#(J)$. If $\#(J)=0$, that is, $J=\emptyset$, then $C=X$ and this is part (a). For the inductive step to $\#(J)=n+1$, take any $j\in J$ and set $J'=J\setminus\{j\}$,

$$C'=\{x:x\in X,\ x(i)\in C_i\ \forall\ i\in J'\},$$

$$C''=C'\setminus C=\{x:x\in C',\ x(j)\in X_j\setminus C_j\}.$$

Then C, C', C'' all belong to $\mathcal{C}$, and $\theta_0C'=\prod_{i\in J'}\mu_iC_i=\alpha$ say, $\theta_0C=\alpha\mu_jC_j$, $\theta_0C''=\alpha(1-\mu_jC_j)$. Moreover, by the inductive hypothesis, $C'\in\Lambda$ and $\alpha=\lambda C'=\theta C'$. So $C=C'\cap\{x:x(j)\in C_j\}\in\Lambda$ by (ii), and $C''=C'\setminus C\in\Lambda$.

We surely have $\lambda C=\theta C\le\theta_0C$, $\lambda C''\le\theta_0C''$; but also

$$\alpha=\lambda C'=\lambda C+\lambda C''\le\theta_0C+\theta_0C''=\alpha,$$

so in fact

$$\lambda C=\theta_0 C=\alpha\mu_j C_j=\prod_{i\in J}\mu C_i,$$

and the induction proceeds. **Q**

(iv) Now let us return to the general case of a set W of the form $\prod_{i\in I}E_i$ where $E_i\in\Sigma_i$ for each i, and $K=\{i:E_i\ne X_i\}$ is countable. If K is finite then $W=\{x:x(i)\in E_i\ \forall\ i\in K\}$ so $W\in\Lambda$ and

$$\lambda W=\prod_{i\in K}\mu_i E_i=\prod_{i\in I}\mu_i E_i.$$

Otherwise, let $\langle i_n\rangle_{n\in\mathbb{N}}$ be an enumeration of K. For each $n\in\mathbb{N}$ set $W_n=\{x:x\in X,\ x(i_k)\in E_{i_k}\ \forall\ k\le n\}$; then we know that $W_n\in\Lambda$ and that $\lambda W_n=\prod_{k=0}^n\mu_{i_k}E_{i_k}$. But $\langle W_n\rangle_{n\in\mathbb{N}}$ is a non-increasing sequence with intersection W, so $W\in\Lambda$ and

$$\lambda W=\lim_{n\to\infty}\lambda W_n=\prod_{i\in K}\mu_i E_i=\prod_{i\in I}\mu_i E_i.$$

(c) is an immediate consequence of (b) and the definition of $\widehat{\bigotimes}_{i\in I}\Sigma_i$.

(d) Because λ is constructed by Carathéodory's method it must be complete.

(e) Let $\langle C_n\rangle_{n\in\mathbb{N}}$ be a sequence in $\mathcal{C}$ such that $W\subseteq\bigcup_{n\in\mathbb{N}}C_n$ and $\sum_{n=0}^\infty\theta_0 C_n\le\theta W+\frac{1}{2}\epsilon$. Set $V=\bigcup_{n\in\mathbb{N}}C_n$; by (b), $V\in\Lambda$. Let $n\in\mathbb{N}$ be such that $\sum_{i=n+1}^\infty\theta_0 C_i\le\frac{1}{2}\epsilon$, and consider $W'=\bigcup_{k\le n}C_k$. Since $V\setminus W'\subseteq\bigcup_{i>n}C_i$,

$$\begin{aligned}\lambda(W\triangle W')&\le\lambda(V\setminus W')+\lambda(V\setminus W)=\lambda V-\lambda W+\lambda(V\setminus W)=\theta V-\theta W+\theta(V\setminus W)\\&\le\sum_{i=0}^\infty\theta_0 C_i-\theta W+\sum_{i=n+1}\theta_0 C_i\le\frac{1}{2}\epsilon+\frac{1}{2}\epsilon=\epsilon.\end{aligned}$$

(f)(i) If $W\in\Lambda$ and $\epsilon>0$ there is a $V\in\widehat{\bigotimes}_{i\in I}\Sigma_i$ such that $W\subseteq V$ and $\lambda V\le\lambda W+\epsilon$. **P** Let $\langle C_n\rangle_{n\in\mathbb{N}}$ be a sequence in $\mathcal{C}$ such that $W\subseteq\bigcup_{n\in\mathbb{N}}C_n$ and $\sum_{n=0}^\infty\theta_0 C_n\le\theta W+\epsilon$. Then $C_n\in\widehat{\bigotimes}_{i\in I}\Sigma_i$ for each n, so $V=\bigcup_{n\in\mathbb{N}}C_n\in\widehat{\bigotimes}_{i\in I}\Sigma_i$. Now $W\subseteq V$, and

$$\lambda V=\theta V\le\sum_{n=0}^\infty\theta_0 C_n\le\theta W+\epsilon=\lambda W+\epsilon.\ \mathbf{Q}$$

(ii) Now, given $W\in\Lambda$, let $\langle V_n\rangle_{n\in\mathbb{N}}$ be a sequence of sets in $\widehat{\bigotimes}_{i\in I}\Sigma_i$ such that $W\subseteq V_n$ and $\lambda V_n\le\lambda W+2^{-n}$ for each n; then $W_2=\bigcap_{n\in\mathbb{N}}V_n$ belongs to $\widehat{\bigotimes}_{i\in I}\Sigma_i$ and $W\subseteq W_2$ and $\lambda W_2=\lambda W$. Similarly, there is a $W_2'\in\widehat{\bigotimes}_{i\in I}\Sigma_i$ such that $X\setminus W\subseteq W_2'$ and $\lambda W_2'=\lambda(X\setminus W)$, so we may take $W_1=X\setminus W_2'$ to complete the proof.

254G The following is a fundamental, indeed defining, property of product measures. (Compare 251L.)

Lemma Let $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be a family of probability spaces with product (X,Λ,λ). Let (Y,T,ν) be a complete probability space and $\phi:Y\to X$ a function. Suppose that $\nu^*\phi^{-1}[C]\le\lambda C$ for every measurable cylinder $C\subseteq X$. Then ϕ is inverse-measure-preserving. In particular, ϕ is inverse-measure-preserving iff $\phi^{-1}[C]\in\mathrm{T}$ and $\nu\phi^{-1}[C]=\lambda C$ for every measurable cylinder $C\subseteq X$.

Remark By ν^* I mean the usual outer measure defined from ν as in §132.

proof (a) First note that, writing θ for the outer measure of 254A, $\nu^*\phi^{-1}[A]\le\theta A$ for every $A\subseteq X$. **P** Given $\epsilon>0$, there is a sequence $\langle C_n\rangle_{n\in\mathbb{N}}$ of measurable cylinders such that $A\subseteq\bigcup_{n\in\mathbb{N}}C_n$ and $\sum_{n=0}^\infty\theta_0 C_n\le\theta A+\epsilon$, where θ_0 is the functional of 254A. But we know that $\theta_0 C=\lambda C$ for every measurable cylinder C (254Fb), so

$$\nu^*\phi^{-1}[A]\le\nu^*(\bigcup_{n\in\mathbb{N}}\phi^{-1}[C_n])\le\sum_{n=0}^\infty\nu^*\phi^{-1}[C_n]\le\sum_{n=0}^\infty\lambda C_n\le\theta A+\epsilon.$$

As ϵ is arbitrary, $\nu^*\phi^{-1}[A]\le\theta A$. **Q**

(b) Now take any $W\in\Lambda$. Then there are F, $F'\in\mathrm{T}$ such that

$$\phi^{-1}[W]\subseteq F,\quad\phi^{-1}[X\setminus W]\subseteq F',$$

$$\nu F=\nu^*\phi^{-1}[W]\le\theta W=\lambda W,\quad\nu F'\le\lambda[X\setminus W].$$

We have

$$F\cup F'\supseteq\phi^{-1}[W]\cup\phi^{-1}[X\setminus W]=Y,$$

so

$$\nu(F \cap F') = \nu F + \nu F' - \nu(F \cup F') \le \lambda W + \lambda(X \setminus W) - 1 = 0.$$

Now

$$F \setminus \phi^{-1}[W] \subseteq F \cap \phi^{-1}[X \setminus W] \subseteq F \cap F'$$

is ν-negligible. Because ν is complete, $F \setminus \phi^{-1}[W] \in \mathrm{T}$ and $\phi^{-1}[W] = F \setminus (F \setminus \phi^{-1}[W])$ belongs to T. Moreover,

$$1 = \nu F + \nu F' \le \lambda W + \lambda(X \setminus W) = 1,$$

so we must have $\nu F = \lambda W$; but this means that $\nu\phi^{-1}[W] = \nu W$. As W is arbitrary, ϕ is inverse-measure-preserving.

254H Corollary Let $\langle(X_i, \Sigma_i, \mu_i)\rangle_{i\in I}$ and $\langle(Y_i, \mathrm{T}_i, \nu_i)\rangle_{i\in I}$ be two families of probability spaces, with products (X, Λ, λ) and (Y, Λ', λ'). Suppose that for each $i \in I$ we are given an inverse-measure-preserving function $\phi_i : X_i \to Y_i$. Set $\phi(x) = \langle\phi_i(x(i))\rangle_{i\in I}$ for $x \in X$. Then $\phi : X \to Y$ is inverse-measure-preserving.

proof If $C = \prod_{i\in I} C_i$ is a measurable cylinder in Y, then $\phi^{-1}[C] = \prod_{i\in I} \phi_i^{-1}[C_i]$ is a measurable cylinder in X, and

$$\lambda\phi^{-1}[C] = \prod_{i\in I} \mu_i\phi_i^{-1}[C_i] = \prod_{i\in I} \nu_i C_i = \lambda' C.$$

Since λ is a complete probability measure, 254G tells us that ϕ is inverse-measure-preserving.

254I Corresponding to 251T we have the following.

Proposition Let $\langle(X_i, \Sigma_i, \mu_i)\rangle_{i\in I}$ be a family of probability spaces, λ the product measure on $X = \prod_{i\in I} X_i$, and Λ its domain. Then λ is also the product of the completions $\hat\mu_i$ of the μ_i (212C).

proof Write $\hat\lambda$ for the product of the $\hat\mu_i$, and $\hat\Lambda$ for its domain. (i) The identity map from X_i to itself is inverse-measure-preserving if regarded as a map from $(X_i, \hat\mu_i)$ to (X_i, μ_i), so the identity map on X is inverse-measure-preserving if regarded as a map from $(X, \hat\lambda)$ to (X, λ), by 254H; that is, $\Lambda \subseteq \hat\Lambda$ and $\lambda = \hat\lambda\restriction\Lambda$. (ii) If C is a measurable cylinder for $\langle\hat\mu_i\rangle_{i\in I}$, that is, $C = \prod_{i\in I} C_i$ where $C_i \in \hat\Sigma_i$ for every i and $\{i : C_i \ne X_i\}$ is finite, then for each $i \in I$ we can find a $C_i' \in \Sigma_i$ such that $C_i \subseteq C_i'$ and $\mu_i C_i' = \hat\mu_i C_i$; setting $C' = \prod_{i\in I} C_i'$, we get

$$\lambda^* C \le \lambda C' = \prod_{i\in I} \mu_i C_i' = \prod_{i\in I} \hat\mu_i C_i = \hat\lambda C.$$

By 254G, λW must be defined and equal to $\hat\lambda W$ whenever $W \in \hat\Lambda$. Putting this together with (i), we see that $\lambda = \hat\lambda$.

254J The product measure on $\{0,1\}^I$ **(a)** Perhaps the most important of all examples of infinite product measures is the case in which each factor X_i is just $\{0,1\}$ and each μ_i is the 'fair-coin' probability measure, setting

$$\mu_i\{0\} = \mu_i\{1\} = \frac{1}{2}.$$

In this case, the product $X = \{0,1\}^I$ has a family $\langle E_i\rangle_{i\in I}$ of measurable sets such that, writing λ for the product measure on X,

$$\lambda(\bigcap_{i\in J} E_i) = 2^{-\#(J)} \text{ if } J \subseteq I \text{ is finite.}$$

(Just take $E_i = \{x : x(i) = 1\}$ for each i.) I will call this λ the **usual measure** on $\{0,1\}^I$. Observe that if I is finite then $\lambda\{x\} = 2^{-\#(I)}$ for each $x \in X$ (using 254Fb). On the other hand, if I is infinite, then $\lambda\{x\} = 0$ for every $x \in X$ (because, again using 254Fb, $\lambda^*\{x\} \le 2^{-n}$ for every n).

(b) There is a natural bijection between $\{0,1\}^I$ and $\mathcal{P}I$, matching $x \in \{0,1\}^I$ with $\{i : i \in I, x(i) = 1\}$. So we get a standard measure $\tilde\lambda$ on $\mathcal{P}I$, which I will call the **usual measure on** $\mathcal{P}I$. Note that for any finite $b \subseteq I$ and any $c \subseteq b$ we have

$$\tilde\lambda\{a : a \cap b = c\} = \lambda\{x : x(i) = 1 \text{ for } i \in c,\ x(i) = 0 \text{ for } i \in b \setminus c\} = 2^{-\#(b)}.$$

(c) Of course we can apply 254G to these measures; if (Y, T, ν) is a complete probability space, a function $\phi : Y \to \{0,1\}^I$ is inverse-measure-preserving iff

$$\nu\{y : y \in Y,\ \phi(y)\restriction J = z\} = 2^{-\#(J)}$$

whenever $J \subseteq I$ is finite and $z \in \{0,1\}^J$; this is because the measurable cylinders in $\{0,1\}^I$ are precisely the sets of the form $\{x : x\restriction J = z\}$ where $J \subseteq I$ is finite.

254K In the case of countably infinite I, we have a very important relationship between the usual product measure of $\{0,1\}^I$ and Lebesgue measure on $[0,1]$.

Proposition Let λ be the usual measure on $X=\{0,1\}^{\mathbb{N}}$, and let μ be Lebesgue measure on $[0,1]$; write Λ for the domain of λ and Σ for the domain of μ.

(i) For $x\in X$ set $\phi(x)=\sum_{i=0}^\infty 2^{-i-1}x(i)$. Then

$\phi^{-1}[E]\in\Lambda$ and $\lambda\phi^{-1}[E]=\mu E$ for every $E\in\Sigma$;

$\phi[F]\in\Sigma$ and $\mu\phi[F]=\lambda F$ for every $F\in\Lambda$.

(ii) There is a bijection $\tilde\phi:X\to[0,1]$ which is equal to ϕ at all but countably many points, and any such bijection is an isomorphism between (X,Λ,λ) and $([0,1],\Sigma,\mu)$.

proof (a) The first point to observe is that ϕ is nearly a bijection. Setting

$$H=\{x:x\in X,\,\exists\,m\in\mathbb{N},\,x(i)=x(m)\;\forall\,i\ge m\},$$

$$H'=\{2^{-n}k:n\in\mathbb{N},\,k\le 2^n\},$$

then H and H' are countable and $\phi\restriction X\setminus H$ is a bijection between $X\setminus H$ and $[0,1]\setminus H'$. (For $t\in[0,1]\setminus H'$, $\phi^{-1}(t)$ is the binary expansion of t.) Because H and H' are countably infinite, there is a bijection between them; combining this with $\phi\restriction X\setminus H$, we have a bijection between X and $[0,1]$ equal to ϕ except at countably many points. For the rest of this proof, let $\tilde\phi$ be any such bijection. Let M be the countable set $\{x:x\in X,\,\phi(x)\ne\tilde\phi(x)\}$, and N the countable set $\phi[M]\cup\tilde\phi[M]$; then $\phi[A]\triangle\tilde\phi[A]\subseteq N$ for every $A\subseteq X$.

(b) To see that $\lambda\tilde\phi^{-1}[E]$ exists and is equal to μE for every $E\in\Sigma$, I consider successively more complex sets E.

($\boldsymbol\alpha$) If $E=\{t\}$ then $\lambda\tilde\phi^{-1}[E]=\lambda\{\tilde\phi^{-1}(t)\}$ exists and is zero.

($\boldsymbol\beta$) If E is of the form $[2^{-n}k,2^{-n}(k+1)[$, where $n\in\mathbb{N}$ and $0\le k<2^n$, then $\phi^{-1}[E]$ differs by at most two points from a set of the form $\{x:x(i)=z(i)\;\forall\,i<n\}$, so $\tilde\phi^{-1}[E]$ differs from this by a countable set, and

$$\lambda\tilde\phi^{-1}[E]=2^{-n}=\mu E.$$

($\boldsymbol\gamma$) If E is of the form $[2^{-n}k,2^{-n}l[$, where $n\in\mathbb{N}$ and $0\le k<l\le 2^n$, then $E=\bigcup_{k\le i<l}[2^{-n}i,2^{-n}(i+1)[$, so

$$\lambda\tilde\phi^{-1}[E]=2^{-n}(l-k)=\mu E.$$

($\boldsymbol\delta$) If E is of the form $[t,u[$, where $0\le t<u\le 1$, then for each $n\in\mathbb{N}$ let k_n be the integral part of 2^nt and l_n the integral part of 2^nu; set $E_n=[2^{-n}(k_n+1),2^{-n}l_n[$; then $\langle E_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence and $\bigcup_{n\in\mathbb{N}}E_n$ is $]t,u[$. So (using (α))

$$\lambda\tilde\phi^{-1}[E]=\lambda\tilde\phi^{-1}[\bigcup_{n\in\mathbb{N}}E_n]=\lim_{n\to\infty}\lambda\tilde\phi^{-1}[E_n]$$
$$=\lim_{n\to\infty}\mu E_n=\mu E.$$

($\boldsymbol\epsilon$) If $E\in\Sigma$, then for any $\epsilon>0$ there is a sequence $\langle I_n\rangle_{n\in\mathbb{N}}$ of half-open subintervals of $[0,1[$ such that $E\setminus\{1\}\subseteq\bigcup_{n\in\mathbb{N}}I_n$ and $\sum_{n=0}^\infty\mu I_n\le\mu E+\epsilon$; now $\tilde\phi^{-1}[E]\subseteq\{\tilde\phi^{-1}(1)\}\cup\bigcup_{n\in\mathbb{N}}\phi^{-1}[I_n]$, so

$$\lambda^*\tilde\phi^{-1}[E]\le\lambda(\textstyle\bigcup_{n\in\mathbb{N}}\tilde\phi^{-1}[I_n])\le\sum_{n=0}^\infty\lambda\tilde\phi^{-1}[I_n]=\sum_{n=0}^\infty\mu I_n\le\mu E+\epsilon.$$

As ϵ is arbitrary, $\lambda^*\tilde\phi^{-1}[E]\le\mu E$, and there is a $V\in\Lambda$ such that $\phi^{-1}[E]\subseteq V$ and $\lambda V\le\mu E$.

($\boldsymbol\zeta$) Similarly, there is a $V'\in\Lambda$ such that $V'\supseteq\tilde\phi^{-1}[[0,1]\setminus E]$ and $\lambda V'\le\mu([0,1]\setminus E)$. Now $V\cup V'=X$, while

$$\lambda V+\lambda V'\le\mu E+(1-\mu E)=1=\lambda(V\cup V'),$$

so $\lambda(V\cap V')=0$ and

$$\tilde\phi^{-1}[E]=(X\setminus V')\cup(V\cap V'\cap\tilde\phi^{-1}[E])$$

belongs to Λ, with

$$\lambda\tilde\phi^{-1}[E]\le\lambda V\le\mu E;$$

at the same time,

$$1-\lambda\tilde\phi^{-1}[E]\le\lambda V'\le 1-\mu E$$

so $\lambda\tilde{\phi}^{-1}[E] = \mu E$.

(c) Now suppose that $C \subseteq X$ is a measurable cylinder of the special form $\{x : x(0) = \epsilon_0, \dots , x(n) = \epsilon_n\}$ for some $\epsilon_0, \dots , \epsilon_n \in \{0, 1\}$. Then $\phi[C] = [t, t + 2^{-n-1}]$ where $t = \sum_{i=0}^{n} 2^{-i-1}\epsilon_i$, so that $\mu\phi[C] = \lambda C$. Since $\tilde{\phi}[C]\triangle\phi[C] \subseteq N$ is countable, $\mu\tilde{\phi}[C] = \lambda C$.

If $C \subseteq X$ is any measurable cylinder, then it is of the form $\{x : x\restriction J = z\}$ for some finite $J \subseteq \mathbb{N}$; taking n so large that $J \subseteq \{0, \dots , n\}$, C is expressible as a disjoint union of $2^{n+1-\#(J)}$ sets of the form just considered, being just those in which $\epsilon_i = z(i)$ for $i \in J$. Summing their measures, we again get $\mu\tilde{\phi}[C] = \lambda C$. Now 254G tells us that $\tilde{\phi}^{-1} : [0, 1] \to X$ is inverse-measure-preserving, that is, $\tilde{\phi}[W]$ is Lebesgue measurable, with measure λW, for every $W \in \Lambda$.

Putting this together with (b), $\tilde{\phi}$ must be an isomorphism between (X, Λ, λ) and $([0, 1], \Sigma, \mu)$, as claimed in (ii) of the proposition.

(d) As for (i), if $E \in \Sigma$ then $\phi^{-1}[E]\triangle\tilde{\phi}^{-1}[E] \subseteq M$ is countable, so $\lambda\phi^{-1}[E] = \lambda\tilde{\phi}^{-1}[E] = \mu E$. While if $W \in \Lambda$, $\phi[F]\triangle\tilde{\phi}[W] \subseteq N$ is countable, so $\mu\phi[W] = \mu\tilde{\phi}[W] = \lambda W$.

(e) Finally, if $\psi : X \to [0, 1]$ is any other bijection which agrees with ϕ at all but countably many points, set $M' = \{x : \psi(x) \neq \phi(x)\}$, $N' = \psi[M'] \cup \phi[M']$. Then

$$\psi^{-1}[E]\triangle\phi^{-1}[E] \subseteq M', \quad \lambda\psi^{-1}[E] = \lambda\phi^{-1}[E] = \mu E$$

for every $E \in \Sigma$, and

$$\psi[F]\triangle\phi[F] \subseteq N', \quad \mu\psi[F] = \mu\phi[F] = \lambda F$$

for every $F \in \Lambda$.

254L Subspaces Just as in 251Q, we can consider the product of subspace measures. There is a simplification in the form of the result because in the present context we are restricted to probability measures.

Theorem Let $\langle(X_i, \Sigma_i, \mu_i)\rangle_{i\in I}$ be a family of probability spaces, and (X, Λ, λ) their product.

(a) For each $i \in I$, let $A_i \subseteq X_i$ be a set of full outer measure, and write $\tilde{\mu}_i$ for the subspace measure on A_i (214B). Let $\tilde{\lambda}$ be the product measure on $A = \prod_{i\in I} A_i$. Then $\tilde{\lambda}$ is the subspace measure on A induced by λ.

(b) $\lambda^*(\prod_{i\in I} A_i) = \prod_{i\in I} \mu_i^* A_i$ whenever $A_i \subseteq X_i$ for every i.

proof (a) Write λ_A for the subspace measure on A defined from λ, and Λ_A for its domain; write $\tilde{\Lambda}$ for the domain of $\tilde{\lambda}$.

(i) Let $\phi : A \to X$ be the identity map. If $C \subseteq X$ is a measurable cylinder, say $C = \prod_{i\in I} C_i$ where $C_i \in \Sigma_i$ for each i, then $\phi^{-1}[C] = \prod_{i\in I}(C_i \cap A_i)$ is a measurable cylinder in A, and

$$\tilde{\lambda}\phi^{-1}[C] = \prod_{i\in I} \tilde{\mu}_i(C_i \cap A_i) \le \prod_{i\in I} \mu_i C_i = \mu C.$$

By 254G, ϕ is inverse-measure-preserving, that is, $\tilde{\lambda}(A \cap W) = \lambda W$ for every $W \in \Lambda$. But this means that $\tilde{\lambda}V$ is defined and equal to $\lambda_A V = \lambda^* V$ for every $V \in \Lambda_A$, since for any such V there is a $W \in \Lambda$ such that $V = A \cap W$ and $\lambda W = \lambda_A V$. In particular, $\lambda_A A = 1$.

(ii) Now regard ϕ as a function from the measure space $(A, \Lambda_A, \lambda_A)$ to $(A, \tilde{\Lambda}, \tilde{\lambda})$. If D is a measurable cylinder in A, we can express it as $\prod_{i\in I} D_i$ where every D_i belongs to the domain of $\tilde{\mu}_i$ and $D_i = A_i$ for all but finitely many i. Now for each i we can find $C_i \in \Sigma_i$ such that $D_i = C_i \cap A_i$ and $\mu C_i = \tilde{\mu}_i D_i$, and we can suppose that $C_i = X_i$ whenever $D_i = A_i$. In this case $C = \prod_{i\in I} C_i \in \Lambda$ and

$$\lambda C = \prod_{i\in I} \mu_i C_i = \prod_{i\in I} \tilde{\mu}_i D_i = \tilde{\lambda} D.$$

Accordingly

$$\lambda_A\phi^{-1}[D] = \lambda_A(A \cap C) \le \lambda C = \tilde{\lambda} D.$$

By 254G again, ϕ is inverse-measure-preserving in this manifestation, that is, $\lambda_A V$ is defined and equal to $\tilde{\lambda}V$ for every $V \in \tilde{\Lambda}$. Putting this together with (i), we have $\lambda_A = \tilde{\lambda}$, as claimed.

(b) For each $i \in I$, choose a set $E_i \in \Sigma_i$ such that $A_i \subseteq E_i$ and $\mu_i E_i = \mu_i^* A_i$; do this in such a way that $E_i = X_i$ whenever $\mu_i^* A_i = 1$. Set $B_i = A_i \cup (X_i \setminus E_i)$, so that $\mu_i^* B_i = 1$ for each i (if $F \in \Sigma_i$ and $F \supseteq B_i$ then $F \cap E_i \supseteq A_i$, so

$$\mu_i F = \mu_i(F \cap E_i) + \mu_i(F \setminus E_i) = \mu_i E_i + \mu_i(X_i \setminus E_i) = 1.)$$

By (a), we can identify the subspace measure λ_B on $B = \prod_{i\in I} B_i$ with the product of the subspace measures $\tilde{\mu}_i$ on B_i. In particular, $\lambda^* B = \lambda_B B = 1$. Now $A_i = B_i \cap E_i$ so (writing $A = \prod_{i\in I} A_i$), $A = B \cap \prod_{i\in I} E_i$.

If $\prod_{i\in I}\mu_i^*A_i=0$, then for every $\epsilon>0$ there is a finite $J\subseteq I$ such that $\prod_{i\in J}\mu_i^*A_i\le\epsilon$; consequently (using 254Fb)

$$\lambda^*A\le\lambda\{x:x(i)\in E_i\text{ for every }i\in J\}=\textstyle\prod_{i\in J}\mu_iE_i\le\epsilon.$$

As ϵ is arbitrary, $\lambda^*A=0$. If $\prod_{i\in I}\mu_i^*A_i>0$, then for every $n\in\mathbb{N}$ the set $\{i:\mu^*A_i\le 1-2^{-n}\}$ must be finite, so

$$J=\{i:\mu^*A_i<1\}=\{i:E_i\ne X_i\}$$

is countable. By 254Fb again, applied to $\langle E_i\cap B_i\rangle_{i\in I}$ in the product $\prod_{i\in I}B_i$,

$$\begin{aligned}\lambda^*(\prod_{i\in I}A_i)&=\lambda_B(\prod_{i\in I}A_i)=\lambda_B\{x:x\in B,\,x(i)\in E_i\cap B_i\text{ for every }i\in J\}\\&=\prod_{i\in J}\tilde\mu_i(E_i\cap B_i)=\prod_{i\in I}\mu_i^*A_i,\end{aligned}$$

as required.

254M I now turn to the basic results which make it possible to use these product measures effectively. First, I offer a vocabulary for dealing with subproducts. Let $\langle X_i\rangle_{i\in I}$ be a family of sets, with product X.

(a) For $J\subseteq I$, write X_J for $\prod_{i\in J}X_i$. We have a canonical bijection $x\mapsto(x\restriction J,x\restriction I\setminus J):X\to X_I\times X_{I\setminus J}$. Associated with this we have the map $x\mapsto\pi_J(x)=x\restriction J:X\to X_J$. Now I will say that a set $W\subseteq X$ is **determined by coordinates in** J if there is a $V\subseteq X_J$ such that $W=\pi_J^{-1}[V]$; that is, W corresponds to $V\times X_{I\setminus J}\subseteq X_J\times X_{I\setminus J}$.

It is easy to see that

$$\begin{aligned}&W\text{ is determined by coordinates in }J\\&\quad\iff x'\in W\text{ whenever }x\in W,\,x'\in X\text{ and }x'\restriction J=x\restriction J\\&\quad\iff W=\pi_J^{-1}[\pi_J[W]].\end{aligned}$$

It follows that if W is determined by coordinates in J, and $J\subseteq K\subseteq I$, W is also determined by coordinates in K. The family $\mathcal{W}_J$ of subsets of X determined by coordinates in J is closed under complementation and arbitrary unions and intersections. **P** If $W\in\mathcal{W}_J$, then

$$X\setminus W=X\setminus\pi_J^{-1}[\pi_J[W]]=\pi_J^{-1}[X_J\setminus\pi_J[W]]\in\mathcal{W}_J.$$

If $\mathcal{V}\subseteq\mathcal{W}_J$, then

$$\textstyle\bigcup\mathcal{V}=\bigcup_{V\in\mathcal{V}}\pi_J^{-1}[\pi_J[V]]=\pi_J^{-1}[\bigcup_{V\in\mathcal{V}}\pi_J[V]]\in\mathcal{W}_J.\ \mathbf{Q}$$

(b) It follows that

$$\mathcal{W}=\bigcup\{\mathcal{W}_J:J\subseteq I\text{ is countable}\},$$

the family of subsets of X determined by coordinates in some countable set, is a σ-algebra of subsets of X. **P** (i) X and $\emptyset$ are determined by coordinates in $\emptyset$ (recall that $X_\emptyset$ is a singleton, and that $X=\pi_\emptyset^{-1}[X_\emptyset]$, $\emptyset=\pi_\emptyset^{-1}[\emptyset]$). (ii) If $W\in\mathcal{W}$, there is a countable $J\subseteq I$ such that $W\in\mathcal{W}_J$; now

$$X\setminus W=\pi_J^{-1}[X_J\setminus\pi_J[W]]\in\mathcal{W}_J\subseteq\mathcal{W}.$$

(iii) If $\langle W_n\rangle_{n\in\mathbb{N}}$ is a sequence in $\mathcal{W}$, then for each $n\in\mathbb{N}$ there is a countable $J_n\subseteq I$ such that $W\in\mathcal{W}_{J_n}$. Now $J=\bigcup_{n\in\mathbb{N}}J_n$ is a countable subset of I, and every W_n belongs to $\mathcal{W}_J$, so

$$\textstyle\bigcup_{n\in\mathbb{N}}W_n\in\mathcal{W}_J\subseteq\mathcal{W}.\ \mathbf{Q}$$

(c) If $i\in I$ and $E\subseteq X_i$ then $\{x:x\in X,\,x(i)\in E\}$ is determined by the single coordinate i, so surely belongs to $\mathcal{W}$; accordingly $\mathcal{W}$ must include $\widehat{\bigotimes}_{i\in I}\mathcal{P}X_i$. *A fortiori*, if Σ_i is a σ-algebra of subsets of X_i for each i, $\mathcal{W}\supseteq\widehat{\bigotimes}_{i\in I}\Sigma_i$; that is, every member of $\widehat{\bigotimes}_{i\in I}\Sigma_i$ is determined by coordinates in some countable set.

254N Theorem Let $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be a family of probability spaces and $\langle K_j\rangle_{j\in J}$ a partition of I. For each $j\in J$ let λ_j be the product measure on $Z_j=\prod_{i\in K_j}X_i$, and write λ for the product measure on $X=\prod_{i\in I}X_i$. Then the natural bijection

$$x\mapsto\phi(x)=\langle x\restriction K_j\rangle_{j\in J}:X\to\textstyle\prod_{j\in J}Z_j$$

identifies λ with the product of the family $\langle\lambda_j\rangle_{j\in J}$.

In particular, if $K\subseteq I$ is any set, then λ can be identified with the c.l.d. product of the product measures on $\prod_{i\in K}X_i$ and $\prod_{i\in I\setminus K}X_i$.

proof (Compare 251N.) Write $Z=\prod_{j\in J}Z_j$ and $\tilde\lambda$ for the product measure on Z; let Λ, $\tilde\Lambda$ be the domains of λ and $\tilde\lambda$.

(a) Let $C\subseteq Z$ be a measurable cylinder. Then $\lambda^*\phi^{-1}[C]\le\tilde\lambda C$. **P** Express C as $\prod_{j\in J}C_j$ where $C_j\subseteq Z_j$ belongs to the domain Λ_j of λ_j for each j. Set $L=\{j:C_j\ne Z_j\}$, so that L is finite. Let $\epsilon>0$. For each $j\in L$ let $\langle C_{jn}\rangle_{n\in\mathbb{N}}$ be a sequence of measurable cylinders in $Z_j=\prod_{i\in K_j}X_i$ such that $C_j\subseteq\bigcup_{n\in\mathbb{N}}C_{jn}$ and $\sum_{n=0}^\infty\lambda_jC_{jn}\le\lambda C_j+\epsilon$. Express each C_{jn} as $\prod_{i\in K_j}C_{jni}$ where $C_{jni}\in\Sigma_i$ for $i\in K_j$ (and $\{i:C_{jni}\ne X_i\}$ is finite).

For $f\in\mathbb{N}^L$, set

$$D_f=\{x:x\in X,\,x(i)\in C_{j,f(j),i}\text{ whenever }j\in L,\,i\in K_j\}.$$

Because $\bigcup_{j\in L}\{i:C_{j,f(j),i}\ne X_i\}$ is finite, D_f is a measurable cylinder in X, and

$$\lambda D_f=\prod_{j\in L}\prod_{i\in K_j}\mu_iC_{j,f(j),i}=\prod_{j\in L}\lambda_jC_{j,f(j)}.$$

Also

$$\bigcup\{D_f:f\in\mathbb{N}^L\}\supseteq\phi^{-1}[C]$$

because if $\phi(x)\in C$ then $\phi(x)(j)\in C_j$ for each $j\in L$, so there must be an $f\in\mathbb{N}^L$ such that $\phi(x)(j)\in C_{j,f(j)}$ for every $j\in L$. But (because $\mathbb{N}^L$ is countable) this means that

$$\begin{aligned}\lambda^*\phi^{-1}[C]&\le\sum_{f\in\mathbb{N}^L}\lambda D_f=\sum_{f\in\mathbb{N}^L}\prod_{j\in L}\lambda_jC_{j,f(j)}\\&=\prod_{j\in L}\sum_{n=0}^\infty\lambda_jC_{jn}\le\prod_{j\in L}(\lambda_jC_j+\epsilon).\end{aligned}$$

As ϵ is arbitrary,

$$\lambda^*\phi^{-1}[C]\le\prod_{j\in L}\lambda_jC_j=\tilde\lambda C.\ \mathbf{Q}$$

By 254G, it follows that $\lambda\phi^{-1}[W]$ is defined, and equal to $\tilde\lambda W$, whenever $W\in\tilde\Lambda$.

(b) Next, $\tilde\lambda\phi[D]=\lambda D$ for every measurable cylinder $D\subseteq X$. **P** This is easy. Express D as $\prod_{i\in I}D_i$ where $D_i\in\Sigma_i$ for every $i\in I$ and $\{i:D_i\ne\Sigma_i\}$ is finite. Then $\phi[D]=\prod_{j\in J}\tilde D_j$, where $\tilde D_j=\prod_{i\in K_j}D_i$ is a measurable cylinder for each $j\in J$. Because $\{j:\tilde D_j\ne Z_j\}$ must also be finite (in fact, it cannot have more members than the finite set $\{i:D_i\ne X_i\}$), $\prod_{j\in J}\tilde D_j$ is itself a measurable cylinder in Z, and

$$\tilde\lambda\phi[D]=\prod_{j\in J}\lambda_j\tilde D_j=\prod_{j\in J}\prod_{i\in K_j}\mu D_i=\lambda D.\ \mathbf{Q}$$

Applying 254G to $\phi^{-1}:Z\to X$, it follows that $\tilde\lambda\phi[W]$ is defined, and equal to λW, for every $W\in\Lambda$. But together with (a) this means that for any $W\subseteq X$,

if $W\in\Lambda$ then $\phi[W]\in\tilde\Lambda$ and $\tilde\lambda\phi[W]=\lambda W$,

if $\phi[W]\in\tilde\Lambda$ then $W\in\Lambda$ and $\lambda W=\tilde\lambda\phi[W]$.

And of course this is just what is meant by saying that ϕ is an isomorphism between (X,Λ,λ) and $(Z,\tilde\Lambda,\tilde\lambda)$.

254O Proposition Let $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be a family of probability spaces. For each $J\subseteq I$ let λ_J be the product probability measure on $X_J=\prod_{i\in J}X_i$, and Λ_J its domain; write $X=X_I$, $\lambda=\lambda_I$ and $\Lambda=\Lambda_I$. For $x\in X$ and $J\subseteq I$ set $\pi_J(x)=x\restriction J\in X_J$.

(a) For every $J\subseteq I$, λ_J is the image measure $\lambda\pi_J^{-1}$ (234D); in particular, $\pi_J:X\to X_J$ is inverse-measure-preserving for λ and λ_J.

(b) If $J\subseteq I$ and $W\in\Lambda$ is determined by coordinates in J (254M), then $\lambda_J\pi_J[W]$ is defined and equal to λW. Consequently there are W_1, W_2 belonging to the σ-algebra of subsets of X generated by

$$\{\{x:x(i)\in E\}:i\in J,\,E\in\Sigma_i\}$$

such that $W_1\subseteq W\subseteq W_2$ and $\lambda(W_2\setminus W_1)=0$.

(c) For every $W\in\Lambda$, we can find a countable set J and W_1, $W_2\in\Lambda$, both determined by coordinates in J, such that $W_1\subseteq W\subseteq W_2$ and $\lambda(W_2\setminus W_1)=0$.

(d) For every $W \in \Lambda$, there is a countable set $J \subseteq I$ such that $\pi_J[W] \in \Lambda_J$ and $\lambda_J \pi_J[W] = \lambda W$; so that $W' = \pi_J^{-1}[\pi_J[W]]$ belongs to Λ, and $\lambda(W' \setminus W) = 0$.

proof (a)(i) By 254N, we can identify λ with the product of λ_J and $\lambda_{I\setminus J}$ on $X_J \times X_{I\setminus J}$. Now $\pi_J^{-1}[E] \subseteq X$ corresponds to $E \times X_{I\setminus J} \subseteq X_J \times X_{I\setminus J}$, so

$$\lambda(\pi^{-1}[E]) = \lambda_J E \cdot \lambda_{I\setminus J} X_{I\setminus J} = \lambda_J E,$$

by 251E or 251Ia, whenever $E \in \Lambda_J$. This shows that π_J is inverse-measure-preserving.

(ii) To see that λ_J is actually the image measure, suppose that $E \subseteq X_J$ is such that $\pi_J^{-1}[E] \in \Lambda$. Identifying $\pi_J^{-1}[E]$ with $E \times X_{I\setminus J}$, as before, we are supposing that $E \times X_{I\setminus J}$ is measured by the product measure on $X_J \times X_{I\setminus J}$. But this means that for $\lambda_{I\setminus J}$-almost every $z \in X_{I\setminus J}$, $E_z = \{y : (y,z) \in E \times X_{I\setminus J}\}$ belongs to Λ_J (252D(ii), because λ_J is complete). Since $E_z = E$ for every z, E itself belongs to Λ_J, as claimed.

(b) If $W \in \Lambda$ is determined by coordinates in J, set $H = \pi_J[W]$; then $\pi_J^{-1}[H] = W$, so $H \in \Lambda_J$ by (a) just above. By 254Ff, there are H_1, $H_2 \in \widehat{\bigotimes}_{i\in J}\Sigma_i$ such that $H_1 \subseteq H \subseteq H_2$ and $\lambda_J(H_2 \setminus H_1) = 0$.

Let T_J be the σ-algebra of subsets of X generated by sets of the form $\{x : x(i) \in E\}$ where $i \in J$ and $E \in \Sigma_J$. Consider $\mathrm{T}'_J = \{G : G \subseteq X_J,\ \pi_J^{-1}[G] \in \mathrm{T}_J\}$. This is a σ-algebra of subsets of X_J, and it contains $\{y : y \in X_J,\ y(i) \in E\}$ whenever $i \in J$, $E \in \Sigma_J$ (because

$$\pi_J^{-1}[\{y : y \in X_J,\ y(i) \in E\}] = \{x : x \in X,\ x(i) \in E\}$$

whenever $i \in J$, $E \subseteq X_i$). So T'_J must include $\widehat{\bigotimes}_{i\in J}\Sigma_i$. In particular, H_1 and H_2 both belong to T'_J, that is, $W_k = \pi_J^{-1}[H_k]$ belongs to T_J for both k. Of course $W_1 \subseteq W \subseteq W_2$, because $H_1 \subseteq H \subseteq H_2$, and

$$\lambda(W_2 \setminus W_1) = \lambda_J(H_2 \setminus H_1) = 0,$$

as required.

(c) Now take any $W \in \Lambda$. By 254Ff, there are W_1 and $W_2 \in \widehat{\bigotimes}_{i\in I}\Sigma_i$ such that $W_1 \subseteq W \subseteq W_2$ and $\lambda(W_2 \setminus W_1) = 0$. By 254Mc, there are countable sets J_1, $J_2 \subseteq I$ such that, for each k, W_k is determined by coordinates in J_k. Setting $J = J_1 \cup J_2$, J is a countable subset of I and both W_1 and W_2 are determined by coordinates in J.

(d) Continuing the argument from (c), $\pi_J[W_1], \pi_J[W_2] \in \Lambda_J$, by (b), and $\lambda_J(\pi_J[W_2] \setminus \pi_J[W_1]) = 0$. Since $\pi_J[W_1] \subseteq \pi_J[W] \subseteq \pi_J[W_2]$, it follows that $\pi_J[W] \in \Lambda_J$, with $\lambda_J\pi_J[W] = \lambda_J\pi_J[W_2]$; so that, setting $W' = \pi_J^{-1}[\pi_J[W]]$, $W' \in \Lambda$, and

$$\lambda W' = \lambda_J\pi_J[W] = \lambda_J\pi_J[W_2] = \lambda\pi_J^{-1}[\pi_J[W_2]] = \lambda W_2 = \lambda W.$$

254P Proposition Let $\langle(X_i, \Sigma_i, \mu_i)\rangle_{i\in I}$ be a family of probability spaces, and for each $J \subseteq I$ let λ_J be the product probability measure on $X_J = \prod_{i\in J} X_i$, and Λ_J its domain; write $X = X_I$, $\Lambda = \Lambda_I$ and $\lambda = \lambda_I$. For $x \in X$ and $J \subseteq I$ set $\pi_J(x) = x{\restriction}J \in X_J$.

(a) If $J \subseteq I$ and g is a real-valued function defined on a subset of X_J, then g is Λ_J-measurable iff $g\pi_J$ is Λ-measurable.

(b) Whenever f is a Λ-measurable real-valued function defined on a λ-conegligible subset of X, we can find a countable set $J \subseteq I$ and a Λ_J-measurable function g defined on a λ_J-conegligible subset of X_J such that f extends $g\pi_J$.

proof (a)(i) If g is Λ_J-measurable and $a \in \mathbb{R}$, there is an $H \in \Lambda_J$ such that $\{y : y \in \operatorname{dom} g,\ g(y) \ge a\} = H \cap \operatorname{dom} g$. Now $\pi_J^{-1}[H] \in \Lambda$, by 254Oa, and $\{x : x \in \operatorname{dom} g\pi_J,\ g\pi_J(x) \ge a\} = \pi_J^{-1}[H] \cap \operatorname{dom} g\pi_J$. So $g\pi_J$ is Λ-measurable.

(ii) If $g\pi_J$ is Λ-measurable and $a \in \mathbb{R}$, then there is a $W \in \Lambda$ such that $\{x : g\pi_J(x) \ge a\} = W \cap \operatorname{dom} g\pi_J$. As in the proof of 254Oa, we may identify λ with the product of λ_J and $\lambda_{I\setminus J}$, and 252D(ii) tells us that, if we identify W with the corresponding subset of $X_J \times X_{I\setminus J}$, there is at least one $z \in X_{I\setminus J}$ such that $W_z = \{y : y \in X_I,\ (y,z) \in W\}$ belongs to Λ_J. But since (on this convention) $g\pi_J(y,z) = g(y)$ for every $y \in X_J$, we see that $\{y : y \in \operatorname{dom} g,\ g(y) \ge a\} = W_z \cap \operatorname{dom} g$. As a is arbitrary, g is Λ_J-measurable.

(b) For rational numbers q, set $W_q = \{x : x \in \operatorname{dom} f,\ f(x) \ge q\}$. By 254Oc we can find for each q a countable set $J_q \subseteq I$ and sets W'_q, W''_q, both determined by coordinates in J_q, such that $W'_q \subseteq W_q \subseteq W''_q$ and $\lambda(W''_q \setminus W'_q) = 0$. Set $J = \bigcup_{q\in\mathbb{Q}} J_q$, $V = X \setminus \bigcup_{q\in\mathbb{Q}}(W''_q \setminus W'_q)$; then J is a countable subset of I and V is a conegligible subset of X; moreover, V is determined by coordinates in J because all the W'_q, W''_q are.

For every $q \in \mathbb{Q}$, $W_q \cap V = W'_q \cap V$, because $V \cap (W_q \setminus W'_q) \subseteq V \cap (W''_q \setminus W'_q) = \emptyset$; so $W_q \cap V$ is determined by coordinates in J. Consequently $V \cap \operatorname{dom} f = \bigcup_{q\in\mathbb{Q}} V \cap W_q$ also is determined by coordinates in J. Also

$$\{x : x \in V \cap \operatorname{dom} f,\ f(x) \geq a\} = \bigcap_{q \leq a} V \cap W_q$$

is determined by coordinates in J. What this means is that if x, $x' \in V$ and $\pi_J x = \pi_J x'$, then $x \in \operatorname{dom} f$ iff $x' \in \operatorname{dom} f$ and in this case $f(x) = f(x')$. Setting $H = \pi_J[V \cap \operatorname{dom} f]$, we have $\pi_J^{-1}[H] = V \cap \operatorname{dom} f$ a conegligible subset of X, so (because $\lambda_J = \lambda\pi_J^{-1}$) H is conegligible in X_J. Also, for $y \in H$, $f(x) = f(x')$ whenever $\pi_J x = \pi_J x' = y$, so there is a function $g : H \to \mathbb{R}$ defined by saying that $g\pi_J(x) = f(x)$ whenever $x \in V \cap \operatorname{dom} f$. Thus g is defined almost everywhere in X_J and f extends $g\pi_J$. Finally, for any $a \in \mathbb{R}$,

$$\pi_J^{-1}[\{y : g(y) \geq a\}] = \{x : x \in V \cap \operatorname{dom} f,\ f(x) \geq a\} \in \Lambda;$$

by 254Oa, $\{y : g(y) \geq a\} \in \Lambda_J$; as a is arbitrary, g is measurable.

254Q Proposition Let $\langle(X_i, \Sigma_i, \mu_i)\rangle_{i\in I}$ be a family of probability spaces, and for each $J \subseteq I$ let λ_J be the product probability measure on $X_J = \prod_{i\in J} X_i$; write $X = X_I$, $\lambda = \lambda_I$. For $x \in X$, $J \subseteq I$ set $\pi_J(x) = x{\restriction}J \in X_J$.

(a) Let $\mathcal{S}$ be the linear subspace of $\mathbb{R}^X$ spanned by $\{\chi C : C \subseteq X$ is a measurable cylinder$\}$. Then for every λ-integrable real-valued function f and every $\epsilon > 0$ there is a $g \in \mathcal{S}$ such that $\int |f - g| d\lambda \leq \epsilon$.

(b) Whenever $J \subseteq I$ and g is a real-valued function defined on a subset of X_J, then $\int g\, d\lambda_J = \int g\pi_J d\lambda$ if either integral is defined in $[-\infty, \infty]$.

(c) Whenever f is a λ-integrable real-valued function, we can find a countable set $J \subseteq X$ and a λ_J-integrable function g such that f extends $g\pi_J$.

proof (a)(i) Write $\overline{\mathcal{S}}$ for the set of functions f satisfying the assertion, that is, such that for every $\epsilon > 0$ there is a $g \in \mathcal{S}$ such that $\int |f - g| \leq \epsilon$. Then $f_1 + f_2$ and $cf_1 \in \overline{\mathcal{S}}$ whenever f_1, $f_2 \in \overline{\mathcal{S}}$. **P** Given $\epsilon > 0$ there are g_1, $g_2 \in \mathcal{S}$ such that $\int |f_1 - g_1| \leq \frac{\epsilon}{2+|c|}$, $\int |f_2 - g_2| \leq \frac{\epsilon}{2}$; now $g_1 + g_2$, $cg_1 \in \mathcal{S}$ and $\int |(f_1 + f_2) - (g_1 + g_2)| \leq \epsilon$, $\int |cf_1 - cg_1| \leq \epsilon$. **Q** Also, of course, $f \in \overline{\mathcal{S}}$ whenever $f_0 \in \overline{\mathcal{S}}$ and $f =_{\text{a.e.}} f_0$.

(ii) Write $\mathcal{W}$ for $\{W : W \subseteq X,\ \chi W \in \overline{\mathcal{S}}\}$, and $\mathcal{C}$ for the family of measurable cylinders in X. Then it is plain from the definition in 254A that $C \cap C' \in \mathcal{C}$ for all C, $C' \in \mathcal{C}$, and of course $C \in \mathcal{W}$ for every $C \in \mathcal{C}$, because $\chi C \in \mathcal{S}$. Next, $W \setminus V \in \mathcal{W}$ whenever W, $V \in \mathcal{W}$ and $V \subseteq W$, because then $\chi(W \setminus V) = \chi W - \chi V$. Thirdly, $\bigcup_{n\in\mathbb{N}} W_n \in \mathcal{W}$ for every non-decreasing sequence $\langle W_n\rangle_{n\in\mathbb{N}}$ in $\mathcal{W}$. **P** Set $W = \bigcup_{n\in\mathbb{N}} W_n$. Given $\epsilon > 0$, there is an $n \in \mathbb{N}$ such that $\lambda(W \setminus W_n) \leq \frac{\epsilon}{2}$. Now there is a $g \in \mathcal{S}$ such that $\int |\chi W_n - g| \leq \frac{\epsilon}{2}$, so that $\int |\chi W - g| \leq \epsilon$. **Q** Thus $\mathcal{W}$ is a Dynkin class of subsets of X.

By the Monotone Class Theorem (136B), $\mathcal{W}$ must include the σ-algebra of subsets of X generated by $\mathcal{C}$, which is $\widehat{\bigotimes}_{i\in I}\Sigma_i$. But this means that $\mathcal{W}$ contains every measurable subset of X, since by 254Ff any measurable set differs by a negligible set from some member of $\widehat{\bigotimes}_{i\in I}\Sigma_i$.

(iii) Thus $\overline{\mathcal{S}}$ contains the characteristic function of any measurable subset of X. Because it is closed under addition and scalar multiplication, it contains all simple functions. But this means that it must contain all integrable functions. **P** If f is a real-valued function which is integrable over X, and $\epsilon > 0$, there is a simple function $h : X \to \mathbb{R}$ such that $\int |f - h| \leq \frac{\epsilon}{2}$ (242M), and now there is a $g \in \mathcal{S}$ such that $\int |h - g| \leq \frac{\epsilon}{2}$, so that $\int |f - g| \leq \epsilon$. **Q**

This proves part (a) of the proposition.

(b) Put 254Oa and 235J together.

(c) By 254Pb, there are a countable $J \subseteq I$ and a real-valued function g defined on a conegligible subset of X_J such that f extends $g\pi_J$. Now $\operatorname{dom}(g\pi_J) = \pi_J^{-1}[\operatorname{dom} g]$ is conegligible, so $f =_{\text{a.e.}} g\pi_J$ and $g\pi_J$ is λ-integrable. By (b), g is λ_J-integrable.

254R Conditional expectations again Putting the ideas of 253H together with the work above, we obtain some results which are important not only for their direct applications but for the light they throw on the structures here.

Theorem Let $\langle(X_i, \Sigma_i, \mu_i)\rangle_{i\in I}$ be a family of probability spaces with product (X, Λ, λ). For $J \subseteq I$ let $\Lambda_J \subseteq \Lambda$ be the σ-subalgebra of sets determined by coordinates in J (254Mb). Then we may regard $L^0(\lambda{\restriction}\Lambda_J)$ as a subspace of $L^0(\lambda)$ (242Jh). Let $P_J : L^1(\lambda) \to L^1(\lambda{\restriction}\Lambda_J) \subseteq L^1(\lambda)$ be the corresponding conditional expectation operator (242Jd). Then

(a) for any J, $K \subseteq I$, $P_{K\cap J} = P_K P_J$;

(b) for any $u \in L^1(\lambda)$, there is a countable set $J^* \subseteq I$ such that $P_J u = u$ iff $J \supseteq J^*$;

(c) for any $u \in L^0(\lambda)$, there is a unique smallest set $J^* \subseteq I$ such that $u \in L^0(\lambda{\restriction}\Lambda_{J^*})$, and this J^* is countable;

(d) for any $W \in \Lambda$ there is a unique smallest set $J^* \subseteq I$ such that $W \triangle W'$ is negligible for some $W' \in \Lambda_{J^*}$, and this J^* is countable;

(e) for any Λ-measurable real-valued function $f : X \to \mathbb{R}$ there is a unique smallest set $J^* \subseteq I$ such that f is equal almost everywhere to a Λ^*_J-measurable function, and this J^* is countable.

proof For $J \subseteq I$, write $X_J = \prod_{i\in J} X_i$, let λ_J be the product measure on X_J, and set $\phi_J(x) = x\restriction J$ for $x \in X$. Write L^0_J for $L^0(\lambda\restriction\Lambda_J)$, regarded as a subset of $L^0 = L^0_I$, and L^1_J for $L^1(\lambda\restriction\Lambda_J) = L^1(\lambda) \cap L^0_J$, as in 242Jb; thus L^1_J is the set of values of the projection P_J.

(a)(i) Let $C \subseteq X$ be a measurable cylinder, expressed as $\prod_{i\in I} C_i$ where $C_i \in \Sigma_i$ for every i and $L = \{i : C_i \neq X_i\}$ is finite. Set

$$C'_i = C_i \text{ for } i \in J,\ X_i \text{ for } i \in I \setminus J, \quad C' = \prod_{i\in I} C'_i, \quad \alpha = \prod_{i\in I\setminus J} \mu_i C_i.$$

Then $\alpha\chi C'$ is a conditional expectation of χC on Λ_J. **P** By 254N, we can identify λ with the product of λ_J and $\lambda_{I\setminus J}$. This identifies Λ_J with $\{E \times X_{I\setminus J} : E \in \operatorname{dom}\lambda_J\}$. By 253H we have a conditional expectation g of χC defined by setting

$$g(y, z) = \int \chi C(y, t)\lambda_{I\setminus J}(dt)$$

for $y \in X_J$, $z \in X_{I\setminus J}$. But C is identified with $C_J \times C_{I\setminus J}$, where $C_J = \prod_{i\in J} C_i$, so that $g(y, z) = 0$ if $y \notin C_J$ and otherwise is $\lambda_{I\setminus J} C_{I\setminus J} = \alpha$. Thus $g = \alpha\chi(C_J \times X_{I\setminus J})$. But the identification between $X_I \times X_{I\setminus J}$ and X matches $C_J \times X_{I\setminus J}$ with C', as described above. So g becomes identified with $\alpha\chi C'$ and $\alpha\chi C'$ is a conditional expectation of χC. **Q**

(ii) Next, setting

$$C''_i = C'_i \text{ for } i \in K,\ X_i \text{ for } i \in I \setminus K, \quad C'' = \prod_{i\in I} C''_i,$$

$$\beta = \prod_{i\in I\setminus K} \mu_i C'_i = \prod_{i\in I\setminus(J\cup K)} \mu_i C_i,$$

the same arguments show that $\beta\chi C''$ is a conditional expectation of $\chi C'$ on Λ_K. So we have

$$P_K P_J(\chi C)^\bullet = \beta\alpha(\chi C'')^\bullet.$$

But if we look at $\beta\alpha$, this is just $\prod_{i\in I\setminus(K\cap J)} \mu_i C_i$, while $C''_i = C_i$ if $i \in K \cap J$, X_i for other i. So $\beta\alpha\chi C''$ is a conditional expectation of χC on $\Lambda_{K\cap J}$, and

$$P_K P_J(\chi C)^\bullet = P_{K\cap J}(\chi C)^\bullet.$$

(iii) Thus we see that the operators $P_K P_J$, $P_{K\cap J}$ agree on elements of the form $\chi C^\bullet$ where C is a measurable cylinder. Because they are both linear, they agree on linear combinations of these, that is, $P_K P_J v = P_{K\cap J} v$ whenever $v = g^\bullet$ for some g in the space $\mathcal{S}$ of 254Q. But if $u \in L^1(\lambda)$ and $\epsilon > 0$, there is a λ-integrable function f such that $f^\bullet = u$ an there is a $g \in \mathcal{S}$ such that $\int |f - g| \le \epsilon$ (254Qa), so that $\|u - v\|_1 \le \epsilon$, where $v = g^\bullet$. Since P_J, P_K and $P_{K\cap J}$ are all linear operators of norm 1,

$$\|P_K P_J u - P_{K\cap J} u\|_1 \le 2\|u - v\|_1 + \|P_K P_J v - P_{K\cap J} v\|_1 \le 2\epsilon.$$

As ϵ is arbitrary, $P_K P_J u = P_{K\cap J} u$; as u is arbitrary, $P_K P_J = P_{K\cap J}$.

(b) Take $u \in L^1(\lambda)$. Let $\mathcal{J}$ be the family of all subsets J of I such that $P_J u = u$. By (a), $J \cap K \in \mathcal{J}$ for all J, $K \in \mathcal{J}$. Next, $\mathcal{J}$ contains a countable set J_0. **P** Let f be a λ-integrable function such that $f^\bullet = u$. By 254Qc, we can find a countable set $J_0 \subseteq I$ and a λ_{J_0}-integrable function g such that $f =_{\text{a.e.}} g\pi_{J_0}$. Now $g\pi_{J_0}$ is Λ_{J_0}-measurable and $u = (g\pi_{J_0})^\bullet$ belongs to $L^1_{J_0}$, so $J_0 \in \mathcal{J}$. **Q**

Write $J^* = \bigcap \mathcal{J}$, so that $J^* \subseteq J_0$ is countable. Then $J^* \in \mathcal{J}$. **P** Let $\epsilon > 0$. As in the proof of (a) above, there is a $g \in \mathcal{S}$ such that $\|u - v\|_1 \le \epsilon$, where $v = g^\bullet$. But because g is a finite linear combination of characteristic functions of measurable cylinders, each determined by coordinates in some finite set, there is a finite $K \subseteq I$ such that g is Λ_K-measurable, so that $P_K v = v$. Because K is finite, there must be $J_1, \dots, J_n \in \mathcal{J}$ such that $J^* \cap K = \bigcap_{1\le i\le n} J_i \cap K$; but as $\mathcal{J}$ is closed under finite intersections, $J = J_1 \cap \ldots \cap J_n \in \mathcal{J}$, and $J^* \cap K = J \cap K$.

Now we have

$$P_{J^*} v = P_{J^*} P_K v = P_{J^*\cap K} v = P_{J\cap K} v = P_J P_K v = P_J v,$$

using (a) twice. Because both P_J and P_{J^*} have norm 1,

$$\begin{aligned}\|P_{J^*} u - u\|_1 &\le \|P_{J^*} u - P_{J^*} v\|_1 + \|P_{J^*} v - P_J v\|_1 + \|P_J v - P_J u\|_1 + \|P_J u - u\|_1 \\ &\le \|u - v\|_1 + 0 + \|u - v\|_1 + 0 \le 2\epsilon.\end{aligned}$$

As ϵ is arbitrary, $P_{J^*} u = u$ and $J^* \in \mathcal{J}$. **Q**

Now, for any $J \subseteq I$,

$$P_J u = u \Longrightarrow J \in \mathcal{J} \Longrightarrow J \supseteq J^*$$
$$\Longrightarrow P_J u = P_J P_{J^*} u = P_{J\cap J^*} u = P_{J^*} u = u.$$

Thus J^* has the required properties.

(c) Set $e = (\chi X)^\bullet$, $u_n = (-ne)\vee(u\wedge ne)$ for each $n \in \mathbb{N}$. Then, for any $J \subseteq I$, $u \in L^0_J$ iff $u_n \in L^0_J$ for every n. **P** (α) If $u \in L^0_J$, then u is expressible as $f^\bullet$ for some Λ_J-measurable f; now $f_n = (-n\chi X) \vee (f \wedge n\chi X)$ is Λ_J-measurable, so $u_n = f_n^\bullet \in L^0_J$ for every n. (β) If $u_n \in L^0_J$ for each n, then for each n we can find a Λ_J-measurable function f_n such that $f_n^\bullet = u_n$. But there is also a Λ-measurable function f such that $u = f^\bullet$, and we must have $f_n =_{\text{a.e.}} (-n\chi X)\vee(f\wedge n\chi X)$ for each n, so that $f =_{\text{a.e.}} \lim_{n\to\infty} f_n$ and $u = (\lim_{n\to\infty} f_n)^\bullet$. Since $\lim_{n\to\infty} f_n$ is Λ_J-measurable, $u \in L^0_J$. **Q**

As every u_n belongs to L^1, we know that

$$u_n \in L^0_J \iff u_n \in L^1_J \iff P_J u_n = u_n.$$

By (b), there is for each n a countable J_n^* such that $P_J u_n = u_n$ iff $J \supseteq J_n^*$. So we see that $u \in L^0_J$ iff $J \supseteq J_n^*$ for every n, that is, $J \supseteq \bigcup_{n\in\mathbb{N}} J_n^*$. Thus $J^* = \bigcup_{n\in\mathbb{N}} J_n^*$ has the property claimed.

(d) Applying (c) to $u = (\chi W)^\bullet$, we have a (countable) unique smallest J^* such that $u \in L^0_{J^*}$. But if $J \subseteq I$, then there is a $W' \in \Lambda_J$ such that $W'\triangle W$ is negligible iff $u \in L^0_J$. So this is the J^* we are looking for.

(e) Again apply (c), this time to $f^\bullet$.

254S Proposition Let $\langle(X_i, \Sigma_i, \mu_i)\rangle_{i\in I}$ be a family of probability spaces, with product (X, Λ, λ).

(a) If $A \subseteq X$ is determined by coordinates in $I \setminus \{j\}$ for every $j \in I$, then its outer measure λ^*A must be either 0 or 1.

(b) If $W \in \Lambda$ and $\lambda W > 0$, then for every $\epsilon > 0$ there are a $W' \in \Lambda$ and a finite set $J \subseteq I$ such that $\lambda W' \geq 1 - \epsilon$ and for every $x \in W'$ there is a $y \in W$ such that $x\restriction I \setminus J = y\restriction I \setminus J$.

proof For $J \subseteq I$ write X_J for $\prod_{i\in J} X_i$ and λ_J for the product measure on X_J.

(a) Let W be a measurable envelope of A. By 254Rd, there is a smallest $J \subseteq I$ for which there is a $W' \in \Lambda$, determined by coordinates in J, with $\lambda(W\triangle W') = 0$. Now $J = \emptyset$. **P** Take any $j \in I$. Then A is determined by coordinates in $I\setminus\{j\}$, that is, can be regarded as $X_j \times A'$ for some $A' \subseteq X_{I\setminus\{j\}}$. We can also think of λ as the product of $\lambda_{\{j\}}$ and $\lambda_{I\setminus\{j\}}$ (254N). Let $\Lambda_{I\setminus\{j\}}$ be the domain of $\lambda_{I\setminus\{j\}}$. By 251S,

$$\lambda^*A = \lambda^*_{\{j\}} X_j \cdot \lambda^*_{I\setminus\{j\}} A' = \lambda^*_{I\setminus\{j\}} A'.$$

Let $V \in \Lambda_{I\setminus\{j\}}$ be measurable envelope of A'. Then $W' = X_j \times V$ belongs to Λ, includes A and has measure λ^*A, so $\lambda(W \cap W') = \lambda W = \lambda W'$ and $W\triangle W'$ is negligible. At the same time, W' is determined by coordinates in $I \setminus \{j\}$. This means that J must be included in $I \setminus \{j\}$. As j is arbitrary, $J = \emptyset$. **Q**

But the only subsets of X which are determined by coordinates in $\emptyset$ are X and $\emptyset$. Since W differs from one of these by a negligible set, $\lambda^*A = \lambda W \in \{0, 1\}$, as claimed.

(b) Set $\eta = \frac{1}{2}\min(\epsilon, 1)\lambda W$. By 254Fe, there is a measurable set V, determined by coordinates in a finite subset J of I, such that $\lambda(W\triangle V) \leq \eta$. Note that

$$\lambda V \geq \lambda W - \eta \geq \frac{1}{2}\lambda W > 0,$$

so

$$\lambda(W\triangle V) \leq \frac{1}{2}\epsilon\lambda W \leq \epsilon\lambda V.$$

We may identify λ with the c.l.d. product of λ_J and $\lambda_{I\setminus J}$ (254N). Let $\tilde{W}$, $\tilde{V} \subseteq X_I \times X_{I\setminus J}$ be the sets corresponding to W, $V \subseteq X$. Then $\tilde{V}$ can be expressed as $U \times X_{I\setminus J}$ where $\lambda_J U = \lambda V > 0$. Set $U' = \{z : z \in X_{I\setminus J},\ \lambda_J \tilde{W}^{-1}[\{z\}] = 0\}$. Then U' is measured by $\lambda_{I\setminus J}$ (252D(ii) again, because both λ_J and $\lambda_{I\setminus J}$ are complete), and

$$\lambda_J U \cdot \lambda_{I\setminus J} U' \leq \int \lambda_J(\tilde{W}^{-1}[\{z\}]\triangle U)\lambda_{I\setminus J}(dz)$$

(because if $z \in U'$ then $\lambda_J(\tilde{W}^{-1}[\{z\}]\triangle U) = \lambda_J U$)

$$= \int \lambda_J(\tilde{W}\triangle\tilde{V})^{-1}[\{z\}]\lambda_{I\setminus J}(dz)$$
$$= (\lambda_J \times \lambda_{I\setminus J})(\tilde{W}\triangle\tilde{V})$$

(252D once more)

$$= \lambda(W\triangle V) \le \epsilon\lambda V = \epsilon\lambda_J U.$$

This means that $\lambda_{I\setminus J}U' \le \epsilon$. Set $W' = \{x : x \in X,\, x\restriction I\setminus J \notin U'\}$; then $\lambda W' \ge 1-\epsilon$. If $x \in W'$, then $z = x\restriction I\setminus J \notin U'$, so $\tilde{W}^{-1}[\{z\}]$ is not empty, that is, there is a $y \in W$ such that $y\restriction I \setminus J = z$. So this W' has the required properties.

254T Remarks It is important to understand that the results above apply to L^0 and L^1 and measurable-sets-up-to-a-negligible set, not to sets and functions themselves. One idea does apply to sets and functions, whether measurable or not.

(a) Let $\langle X_i\rangle_{i\in I}$ be a family of sets with Cartesian product X. For each $J \subseteq I$ let $\mathcal{W}_J$ be the set of subsets of X determined by coordinates in J. Then $\mathcal{W}_J \cap \mathcal{W}_K = \mathcal{W}_{J\cap K}$ for all J, $K \subseteq I$. **P** Of course $\mathcal{W}_J \cap \mathcal{W}_K \supseteq \mathcal{W}_{J\cap K}$, because $\mathcal{W}_J \supseteq \mathcal{W}_{J'}$ whenever $J' \subseteq J$. On the other hand, suppose $W \in \mathcal{W}_J \cap \mathcal{W}_K$, $x \in W$, $y \in X$ and $x\restriction J\cap K = y\restriction J\cap K$. Set $z(i) = x(i)$ for $i \in J$, $y(i)$ for $i \in I\setminus J$. Then $z\restriction J = x\restriction J$ so $z \in W$. Also $y\restriction K = z\restriction K$ so $y \in W$. As x, y are arbitrary, $W \in \mathcal{W}_{J\cap K}$; as W is arbitrary, $\mathcal{W}_J \cap \mathcal{W}_K \subseteq \mathcal{W}_{J\cap K}$. **Q** Accordingly, for any $W \subseteq X$, $\mathcal{F} = \{J : W \in \mathcal{W}_J\}$ is a filter on I (unless $W = X$ or $W = \emptyset$, in which case $\mathcal{F} = \mathcal{P}X$). But $\mathcal{F}$ does not necessarily have a least element, as the following example shows.

(b) Set $X = \{0,1\}^{\mathbb{N}}$,

$$W = \{x : x \in X,\, \lim_{i\to\infty} x(i) = 0\}.$$

Then for every $n \in \mathbb{N}$ W is determined by coordinates in $J_n = \{i : i \ge n\}$. But W is not determined by coordinates in $\bigcap_{n\in\mathbb{N}} J_n = \emptyset$. Note that

$$W = \bigcup_{n\in\mathbb{N}} \bigcap_{i\ge n} \{x : x(i) = 0\}$$

is measured by the usual measure on X. But it is also negligible (since it is countable); in 254Rd we have $J = \emptyset$, $W' = \emptyset$.

***254U** I am now in a position to describe a counter-example answering a natural question arising out of §251.

Example There are a localizable measure space (X, Σ, μ) and a probability space (Y, T, ν) such that the c.l.d. product measure λ on $X \times Y$ is not localizable.

proof (a) Take (X, Σ, μ) to be the space of 216E, so that $X = \{0,1\}^I$, where $I = \mathcal{P}C$ for some set C of cardinal greater than $\mathfrak{c}$. For each $\gamma \in C$ write E_γ for $\{x : x \in X,\, x(\{\gamma\}) = 1\}$ (that is, $G_{\{\gamma\}}$ in the notation of 216Ec); then $E_\gamma \in \Sigma$ and $\mu E_\gamma = 1$; also every measurable set of non-zero measure meets some E_γ in a set of non-zero measure, while $E_\gamma \cap E_\delta$ is negligible for all distinct γ, δ (see 216Ee).

Let (Y, T, ν) be $\{0,1\}^C$ with the usual measure (254J). For $\gamma \in C$, let F_γ be $\{y : y \in Y,\, y(\gamma) = 1\}$, so that $\nu F_\gamma = \frac{1}{2}$. Let λ be the c.l.d. product measure on $X \times Y$, and Λ its domain.

(b) Consider the family $\mathcal{W} = \{E_\gamma \times F_\gamma : \gamma \in C\} \subseteq \Lambda$. **?** Suppose, if possible, that V were an essential supremum of $\mathcal{W}$ in Λ in the sense of 211G. For $\gamma \in C$ write $H_\gamma = \{x : V[\{x\}]\triangle F_\gamma$ is negligible$\}$. Because $F_\gamma \triangle F_\delta$ is non-negligible, $H_\gamma \cap H_\delta = \emptyset$ for all $\gamma \ne \delta$.

Now $E_\gamma \setminus H_\gamma$ is μ-negligible for every $\gamma \in C$. **P** $\lambda((E_\gamma \times F_\gamma) \setminus V) = 0$, so $F_\gamma \setminus V[\{x\}]$ is negligible for almost every $x \in E_\gamma$, by 252D. On the other hand, if we set $F'_\gamma = Y \setminus F_\gamma$, $W_\gamma = (X \times Y) \setminus (E_\gamma \times F'_\gamma)$, then we see that

$$(E_\gamma \times F'_\gamma) \cap (E_\gamma \times F_\gamma) = \emptyset, \quad E_\gamma \times F_\gamma \subseteq W_\gamma,$$

$$\lambda((E_\delta \times F_\delta) \setminus W_\gamma) = \lambda((E_\gamma \times F'_\gamma) \cap (E_\delta \times F_\delta)) \le \mu(E_\gamma \cap E_\delta) = 0$$

for every $\delta \ne \gamma$, so W_γ is an essential upper bound for $\mathcal{W}$ and $V \cap (E_\gamma \times F'_\gamma) = V \setminus W_\gamma$ must be λ-negligible. Accordingly $V[\{x\}] \setminus F_\gamma = V[\{x\}] \cap F'_\gamma$ is ν-negligible for μ-almost every $x \in E_\gamma$. But this means that $V[\{x\}]\triangle F_\gamma$ is ν-negligible for μ-almost every $x \in E_\gamma$, that is, $\nu(E_\gamma \setminus H_\gamma) = 0$. **Q**

Now consider the family $\langle E_\gamma \cap H_\gamma\rangle_{\gamma\in C}$. This is a disjoint family of sets of finite measure in X. If $E \in \Sigma$ has non-zero measure, there is a $\gamma \in C$ such that $\mu(E_\gamma \cap H_\gamma \cap E) = \nu(E_\gamma \cap E) > 0$. But this means that $\mathcal{E} = \{E_\gamma \cap H_\gamma : \gamma \in C\}$ satisfies the conditions of 213O, and μ must be strictly localizable; which it isn't. **X**

(c) Thus we have found a family $\mathcal{W} \subseteq \Lambda$ with no essential supremum in Λ, and λ is not localizable.

Remark If (X, Σ, μ) and (Y, T, ν) are any localizable measure spaces with a non-localizable c.l.d. product measure, then their c.l.d. versions are still localizable (213Hb) and still have a non-localizable product (251T), which cannot be strictly localizable; so that at least one of the factors is not strictly localizable (251O). Thus any example of the type here must involve a complete locally determined localizable space which is not strictly localizable, as in 216E.

254V Corresponding to 251U and 251Wo, we have the following result on countable powers of atomless probability spaces.

Proposition Let (X, Σ, μ) be an atomless probability space and I a countable set. Let λ be the product probability measure on X^I. Then $\{x : x \in X^I,\ x \text{ is injective}\}$ is λ-conegligible.

proof For any pair $\{i, j\}$ of distinct elements of X, the set $\{z : z \in X^{\{i,j\}},\ z(i) = z(j)\}$ is negligible for the product measure on $X^{\{i,j\}}$, by 251U. By 254Oa, $\{x : x \in X,\ x(i) = x(j)\}$ is λ-negligible. Because I is countable, there are only countably many such pairs $\{i, j\}$, so $\{x : x \in X,\ x(i) = x(j)$ for some distinct $i,\ j \in I\}$ is negligible, and its complement is conegligible; but this complement is just the set of injective functions from I to X.

254X Basic exercises (a) Let $\langle (X_i, \Sigma_i, \mu_i)\rangle_{i\in I}$ be any family of probability spaces, with product (X, Λ, μ). Write $\mathcal{E}$ for the family of subsets of X expressible as the union of a finite disjoint family of measurable cylinders. (i) Show that if $C \subseteq X$ is a measurable cylinder then $X \setminus C \in \mathcal{E}$. (ii) Show that $W \cap V \in \mathcal{E}$ for all $W,\ V \in \mathcal{E}$. (iii) Show that $X \setminus W \in \mathcal{E}$ for every $W \in \mathcal{E}$. (iv) Show that $\mathcal{E}$ is an algebra of subsets of X. (v) Show that for any $W \in \Lambda$, $\epsilon > 0$ there is a $V \in \mathcal{E}$ such that $\lambda(W \triangle V) \le \epsilon^2$. (vi) Show that for any $W \in \Lambda$, $\epsilon > 0$ there are disjoint measurable cylinders $C_0, \dots, C_n$ such that $\lambda(W \cap C_j) \ge (1-\epsilon)\lambda C_j$ for every j and $\lambda(W \setminus \bigcup_{j\le n} C_j) \le \epsilon$. (*Hint*: select the C_j from the measurable cylinders composing a set V as in (v).) (vii) Show that if f, g are λ-integrable functions and $\int_C f \le \int_C g$ for every measurable cylinder $C \subseteq X$, then $f \le_{\text{a.e.}} g$. (*Hint*: show that $\int_W f \le \int_W f$ for every $W \in \Lambda$.)

>(b) Let $\langle (X_i, \Sigma_i, \mu_i)$ be a family of probability spaces, with product (X, Λ, λ). Show that the outer measure λ^* defined by λ is exactly the outer measure θ described in 254A, that is, that θ is a regular outer measure.

(c) Let $\langle (X_i, \Sigma_i, \mu_i)$ be a family of probability spaces, with product (X, Λ, λ). Write λ_0 for the restriction of λ to $\widehat{\bigotimes}_{i\in I}\Sigma_i$, and $\mathcal{C}$ for the family of measurable cylinders in X. Suppose that (Y, T, ν) is a probability space and $\phi : Y \to X$ a function. (i) Show that ϕ is inverse-measure-preserving when regarded as a function from (Y, T, ν) to $(X, \widehat{\bigotimes}_{i\in I}\Sigma_i, \lambda_0)$ iff $\phi^{-1}[C]$ belongs to T and $\nu\phi^{-1}[C] = \lambda_0 C$ for every $C \in \mathcal{C}$. (ii) Show that λ_0 is the only measure on X with this property. (*Hint*: 136C.)

>(d) Let I be a set and (Y, T, ν) a complete probability space. Show that a function $\phi : Y \to \{0,1\}^I$ is inverse-measure-preserving for ν and the usual measure on $\{0,1\}^I$ iff $\nu\{y : \phi(y)(i) = 1 \text{ for every } i \in J\} = 2^{-\#(J)}$ for every finite $J \subseteq I$.

>(e) Let I be any set and λ the usual measure on $X = \{0,1\}^I$. Define addition on X by setting $(x+y)(i) = x(i) +_2 y(i)$ for every $i \in I$, x, $y \in X$, where $0 +_2 0 = 1 +_2 1 = 0$, $0 +_2 1 = 1 +_2 0 = 1$. (i) Show that for any $y \in X$, the map $x \mapsto x+y : X \to X$ is inverse-measure-preserving. (*Hint*: Use 254G.) (ii) Show that the map $(x, y) \mapsto x+y : X \times X \to X$ is inverse-measure-preserving, if $X \times X$ is given its product measure.

>(f) Let I be any set and λ the usual measure on $\mathcal{P}I$. (i) Show that the map $a \mapsto a \triangle b : \mathcal{P}I \to \mathcal{P}I$ is inverse-measure-preserving for any $b \subseteq I$; in particular, $a \mapsto I \setminus a$ is inverse-measure-preserving. (ii) Show that the map $(a, b) \mapsto a \triangle b : \mathcal{P}I \times \mathcal{P}I \to \mathcal{P}I$ is inverse-measure-preserving.

>(g) Show that for any $q \in [0,1]$ and any set I there is a measure λ on $\mathcal{P}I$ such that $\lambda\{a : J \subseteq a\} = q^{\#(J)}$ for every finite $J \subseteq I$.

>(h) Let (Y, T, ν) be a complete probability space, and write μ for Lebesgue measure on $[0,1]$. Suppose that $\phi : Y \to [0,1]$ is a function such that $\nu\phi^{-1}[I]$ exists and is equal to μI for every interval I of the form $[2^{-n}k, 2^{-n}(k+1)]$, where $n \in \mathbb{N}$ and $0 \le k < 2^n$. Show that ϕ is inverse-measure-preserving for ν and μ.

(i) Let $\langle X_i\rangle_{i\in I}$ be a family of sets, and for each $i \in I$ let Σ_i be a σ-algebra of subsets of X_i. Show that for every $E \in \widehat{\bigotimes}_{i\in I}\Sigma_i$ there is a countable set $J \subseteq I$ such that E is expressible as $\pi_J^{-1}[F]$ for some $F \in \widehat{\bigotimes}_{i\in J} X_j$, writing $\pi_J(x) = x{\restriction}J \in \prod_{i\in J} X_i$ for $x \in \prod_{i\in I} X_i$.

(j)(i) Let ν be the usual measure on $X=\{0,1\}^{\mathbb{N}}$. Show that for any $k\ge 1$, (X,ν) is isomorphic to (X^k,ν_k), where ν_k is the measure on X^k which is the product measure obtained by giving each factor X the measure ν. (ii) Writing $\mu_{[0,1]}$ for Lebesgue measure on $[0,1]$, etc., show that for any $k\ge 1$, $([0,1]^k,\mu_{[0,1]^k})$ is isomorphic to $([0,1],\mu_{[0,1]})$.

(k)(i) Writing $\mu_{[0,1]}$ for Lebesgue measure on $[0,1]$, etc., show that $([0,1],\mu_{[0,1]})$ is isomorphic to $([0,1[\,,\mu_{[0,1[})$. (ii) Show that for any $k\ge 1$, $([0,1[^k,\mu_{[0,1[^k})$ is isomorphic to $([0,1[\,,\mu_{[0,1[})$. (iii) Show that for any $k\ge 1$, $(\mathbb{R},\mu_{\mathbb{R}})$ is isomorphic to $(\mathbb{R}^k,\mu_{\mathbb{R}^k})$.

(l) Let μ be Lebesgue measure on $[0,1]$ and λ the product measure on $[0,1]^{\mathbb{N}}$. Show that $([0,1],\mu)$ and $([0,1]^{\mathbb{N}},\lambda)$ are isomorphic.

(m) Let $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be a family of complete probability spaces and λ the product measure on $\prod_{i\in I}X_i$, with domain Λ. Suppose that $A_i\subseteq X_i$ for each $i\in I$. Show that $\prod_{i\in I}A_i\in\Lambda$ iff either (i) $\prod_{i\in I}\mu_i^*A_i=0$ or (ii) $A_i\in\Sigma_i$ for every i and $\{i:A_i\ne X_i\}$ is countable. (*Hint*: assemble ideas from 252Xc, 254F, 254L and 254N.)

(n) Let $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be a family of probability spaces with product (X,Λ,λ). (i) Show that, for any $A\subseteq X$,

$$\lambda^*A=\min\{\lambda_J^*\pi_J[A]:J\subseteq I\text{ is countable}\},$$

where for $J\subseteq I$ I write λ_J for the product probability measure on $X_J=\prod_{i\in J}X_i$ and $\pi_J:X\to X_J$ for the canonical map. (ii) Show that if J, $K\subseteq I$ are disjoint and A, $B\subseteq X$ are determined by coordinates in J, K respectively, then $\lambda^*(A\cap B)=\lambda^*A\cdot\lambda^*B$.

(o) Let $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be a family of probability spaces with product (X,Λ,λ). Let $\mathcal{S}$ be the linear span of the set of characteristic functions of measurable cylinders in X, as in 254Q. Show that $\{f^\bullet:f\in\mathcal{S}\}$ is dense in $L^p(\mu)$ for every $p\in[1,\infty[$.

(p) Let $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be a family of probability spaces, and (X,Λ,λ) their product; for $J\subseteq I$ let Λ_J be the σ-algebra of members of Λ determined by coordinates in J and $P_J:L^1=L^1(\lambda)\to L^1_J=L^1(\lambda\restriction\Lambda_J)$ the corresponding conditional expectation. (i) Show that if $u\in L^1_J$ and $v\in L^1_{I\setminus J}$ then $u\times v\in L^1$ and $\int u\times v=\int u\cdot\int v$. (*Hint*: 253D.) (ii) Show that if $u\in L^1$ then $u\in L^1_J$ iff $\int_C u=\lambda C\cdot\int u$ for every measurable cylinder $C\subseteq X$ which is determined by coordinates in $I\setminus J$. (*Hint*: 254Xa(vii).) (iii) Show that if $\mathcal{J}\subseteq\mathcal{P}I$ is non-empty, with $J^*=\bigcap\mathcal{J}$, then $L^1_{J^*}=\bigcap_{J\in\mathcal{J}}L^1_J$.

(q)(i) Let I be any set and λ the usual measure on $\mathcal{P}I$. Let $A\subseteq\mathcal{P}I$ be such that $a\triangle b\in A$ whenever $a\in A$ and $b\subseteq I$ is finite. Show that λ^*A must be either 0 or 1. (ii) Let λ be the usual measure on $\{0,1\}^{\mathbb{N}}$, and Λ its domain. Let $f:\{0,1\}^{\mathbb{N}}\to\mathbb{R}$ be a function such that, for x, $y\in\{0,1\}^{\mathbb{N}}$, $f(x)=f(y)\iff\{n:n\in\mathbb{N},\,x(n)\ne y(n)\}$ is finite. Show that f is not Λ-measurable. (*Hint*: for any $q\in\mathbb{Q}$, $\lambda^*\{x:f(x)\le q\}$ is either 0 or 1.)

(r) Let $\langle X_i\rangle_{i\in I}$ be any family of sets and $A\subseteq B\subseteq\prod_{i\in I}X_i$. Suppose that A is determined by coordinates in $J\subseteq I$ and that B is determined by coordinates in K. Show that there is a set C such that $A\subseteq C\subseteq B$ and C is determined by coordinates in $J\cap K$.

(s) Show that if $\tilde\phi:\{0,1\}^{\mathbb{N}}\to[0,1]$ is any bijection constructed by the method of 254K, then $\{\tilde\phi^{-1}[E]:E\subseteq[0,1]$ is a Borel set$\}$ is just the σ-algebra of subsets of $\{0,1\}^{\mathbb{N}}$ generated by the sets $\{x:x(i)=1\}$ for $i\in\mathbb{N}$.

254Y Further exercises (a) Let $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be a family of probability spaces, and for $J\subseteq I$ let λ_J be the product measure on $X_J=\prod_{i\in J}X_i$; write $X=X_I$, $\lambda=\lambda_I$ and $\pi_J(x)=x\restriction J$ for $x\in X$ and $J\subseteq I$.

(i) Show that for $K\subseteq J\subseteq I$ we have a natural linear, order-preserving and norm-preserving map $T_{JK}:L^1(\lambda_K)\to L^1(\lambda_J)$ defined by writing $T_{JK}(f^\bullet)=(f\pi_{KJ})^\bullet$ for every λ_K-integrable function f, where $\pi_{KJ}(y)=y\restriction K$ for $y\in X_J$.

(ii) Write $\mathcal{K}$ for the set of finite subsets of I. Show that if W is any Banach space and $\langle T_K\rangle_{K\in\mathcal{K}}$ is a family such that (α) T_K is a bounded linear operator from $L^1(\lambda_K)$ to W for every $K\in\mathcal{K}$ (β) $T_K=T_JT_{JK}$ whenever $K\subseteq J\in\mathcal{K}$ (γ) $\sup_{K\in\mathcal{K}}\|T_K\|<\infty$, then there is a unique bounded linear operator $T:L^1(\lambda)\to W$ such that $T_K=TT_{IK}$ for every $K\in\mathcal{K}$.

(iii) Write $\mathcal{J}$ for the set of countable subsets of I. Show that $L^1(\lambda)=\bigcup_{J\in\mathcal{J}}T_{IJ}[L^1(\lambda_J)]$.

(b) Let $\langle(X_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be a family of probability spaces, and λ a complete measure on $X=\prod_{i\in I}X_i$. Suppose that for every complete probability space (Y,T,ν) and function $\phi:Y\to X$, ϕ is inverse-measure-preserving for ν and λ iff $\nu\phi^{-1}[C]$ is defined and equal to θ_0C for every measurable cylinder $C\subseteq X$, writing θ_0 for the functional of 254A. Show that λ is the product measure on X.

(c) Let I be a set, and λ the usual measure on $\{0,1\}^I$. Show that $L^1(\lambda)$ is separable, in its norm topology, iff I is countable.

(d) Let I be a set, and λ the usual measure on $\mathcal{P}I$. Show that if $\mathcal{F}$ is a non-principal ultrafilter on I then $\lambda^*\mathcal{F} = 1$. (*Hint*: 254Xq, 254Xf.)

(e) Let (X,Σ,μ), (Y,T,ν) and λ be as in 254U. Set $A = \{x_\gamma : \gamma \in C\}$ as defined in 216E. Let μ_A be the subspace measure on A, and $\tilde{\lambda}$ the c.l.d. product measure of μ_A and ν on $A \times Y$. Show that $\tilde{\lambda}$ is a proper extension of the subspace measure $\lambda_{A\times Y}$. (*Hint*: consider $\tilde{W} = \{(f_\gamma, y) : \gamma \in C,\, y \in F_\gamma\}$, in the notation of 254U.)

(f) Let (X,Σ,μ) be an atomless probability space, I a set with cardinal at most $\#(X)$, and A the set of injective functions from I to X. Show that A has full outer measure for the product measure on X^I.

254 Notes and comments While there are many reasons for studying infinite products of probability spaces, one stands pre-eminent, from the point of view of abstract measure theory: they provide constructions of essentially new kinds of measure space. I cannot describe the nature of this 'newness' effectively without venturing into the territory of Volume 3. But the function spaces of Chapter 24 do give at least a form of words we can use: these are the first *probability* spaces (X,Λ,λ) we have seen for which $L^1(\lambda)$ need not be separable for its norm topology (254Yc).

The formulae of 254A, like those of 251A, lead very naturally to measures; the point at which they become more than a curiosity is when we find that the product measure λ is a probability measure (254Fa), which must be regarded as the crucial argument of this section, just as 251E is the essential basis of §251. It is I think remarkable that it makes no difference to the result here whether I is finite, countably infinite or uncountable. If you write out the proof for the case $I = \mathbb{N}$, it will seem natural to expand the sets J_n until they are initial segments of I itself, thereby avoiding altogether the auxiliary set K; but this is a misleading simplification, because it hides an essential feature of the argument, which is that any sequence in $\mathcal{C}$ involves only countably many coordinates, so that as long as we are dealing with only one such sequence the uncountability of the whole set I is irrelevant. This general principle naturally permeates the whole of the section; in 254O I have tried to spell out the way in which many of the questions we are interested in can be expressed in terms of countable subproducts of the factor spaces X_i. See also the exercises 254Xi, 254Xm and 254Ya(iii).

There is a slightly paradoxical side to this principle: even the best-behaved subsets E_i of X_i may fail to have measurable products $\prod_{i\in I} E_i$ if $E_i \ne X_i$ for uncountably many i. For instance, $]0,1[^I$ is not a measurable subset of $[0,1]^I$ if I is uncountable (254Xm). It has full outer measure and its own product measure is just the subspace measure (254L), but any measurable subset must have measure zero. The point is that the empty set is the only member of $\widehat{\bigotimes}_{i\in I}\Sigma_i$, where Σ_i is the algebra of Lebesgue measurable subsets of $[0,1]$ for each i, which is included in $]0,1[^I$ (see 254Xi).

As in §251, I use a construction which automatically produces a complete measure on the product space. I am sure that this is the best choice for 'the' product measure. But there are occasions when its restriction to the σ-algebra generated by the measurable cylinders is worth looking at; see 254Xc.

Lemma 254G is a result of a type which will be commoner in Volume 3 than in the present volume. It describes the product measure in terms not of what it *is* but of what it *does*; specifically, in terms of a property of the associated family of inverse-measure-preserving functions. It is therefore a 'universal mapping theorem'. (Compare 253F.) Because this description is sufficient to determine the product measure completely (254Yb), it is not surprising that I use it repeatedly.

The 'usual measure' on $\{0,1\}^I$ (254J) is sometimes called 'coin-tossing measure' because it can be used to model the concept of tossing a coin arbitrarily many times indexed by the set I, taking an $x \in \{0,1\}^I$ to represent the outcome in which the coin is 'heads' for just those $i \in I$ for which $x(i) = 1$. The sets, or 'events', in the class $\mathcal{C}$ are just those which can be specified by declaring the outcomes of finitely many tosses, and the probability of any particular sequence of n results is $1/2^n$, regardless of which tosses we look at or in which order. In Chapter 27 I will return to the use of product measures to represent probabilities involving independent events.

In 254K I come to the first case in this treatise of a non-trivial isomorphism between two measure spaces. If you have been brought up on a conventional diet of modern abstract pure mathematics based on algebra and topology, you may already have been struck by the absence of emphasis on any concept of 'homomorphism' or 'isomorphism'. Here indeed I start to speak of 'isomorphisms' between measure spaces without even troubling to define them; I hope it really is obvious that an isomorphism between measure spaces (X,Σ,μ) and (Y,T,ν) is a bijection $\phi : X \to Y$ such that $\mathrm{T} = \{F : F \subseteq Y,\, \phi^{-1}[F] \in \Sigma\}$ and $\nu F = \mu\phi^{-1}[F]$ for every $F \in \mathrm{T}$, so that Σ is necessarily $\{E : E \subseteq X,\, \phi[E] \in \mathrm{T}\}$ and $\mu E = \nu\phi[E]$ for every $E \in \Sigma$. Put like this, you may, if you worked through the exercises of Volume 1, be reminded of some constructions of σ-algebras in 111Xc-111Xd and of the 'image measures' in 234C-234D. The result in 254K (see also 134Yo) naturally leads to two distinct notions of 'homomorphism' between two measure spaces (X,Σ,μ) and (Y,T,ν):

(i) a function $\phi : X \to Y$ such that $\phi^{-1}[F] \in \Sigma$ and $\mu\phi^{-1}[F] = \nu F$ for every $F \in \mathrm{T}$,

(ii) a function $\phi : X \to Y$ such that $\phi[E] \in \mathrm{T}$ and $\nu\phi[E] = \mu E$ for every $E \in \Sigma$.

On either definition, we find that a bijection $\phi : X \to Y$ is an isomorphism iff ϕ and ϕ^{-1} are both homomorphisms. (Also, of course, the composition of homomorphisms will be a homomorphism.) My own view is that (i) is the more important, and in this treatise I study such functions at length, calling them 'inverse-measure-preserving'. But both have their uses. The function ϕ of 254K not only satisfies both definitions, but is also 'nearly' an isomorphism in several different ways, of which possibly the most important is that there are conegligible sets $X' \subseteq \{0,1\}^{\mathbb{N}}$, $Y' \subseteq [0,1]$ such that $\phi\restriction X'$ is an isomorphism between X' and Y' when both are given their subspace measures.

Having once established the isomorphism between $[0,1]$ and $\{0,1\}^{\mathbb{N}}$, we are led immediately to many more; see 254Xj-254Xl. In fact Lebesgue measure on $[0,1]$ is isomorphic to a large proportion of the probability spaces arising in applications. In Volumes 3 and 4 I will discuss these isomorphisms at length.

The general notion of 'subproduct' is associated with some of the deepest and most characteristic results in the theory of product measures. Because we are looking at products of arbitrary families of probability spaces, the definition must ignore any possible structure in the index set I of 254A-254C. But many applications, naturally enough, deal with index sets with favoured subsets or partitions, and the first essential step is the 'associative law' (254N; compare 251Xe-251Xf and 251Wh). This is, for instance, the tool by which we can apply Fubini's theorem within infinite products. The natural projection maps from $\prod_{i\in I} X_i$ to $\prod_{i\in J} X_i$, where $J \subseteq I$, are related in a way which has already been used as the basis of theorems in §235; the product measure on $\prod_{i\in J} X_i$ is precisely the image of the product measure on $\prod_{i\in I} X_i$ (254Oa). In 254O-254Q I explore the consequences of this fact and the fact already noted that all measurable sets in the product are 'essentially' determined by coordinates in some countable set.

In 254R I go more deeply into this notion of a set $W \subseteq \prod_{i\in I} X_i$ 'determined by coordinates in' a set $J \subseteq I$. In its primitive form this is a purely set-theoretic notion (254M, 254Ta). I think that even a three-element set I can give us surprises; I invite you to try to visualize subsets of $[0,1]^3$ which are determined by pairs of coordinates. But the interactions of this with measure-theoretic ideas, and in particular with a willingness to add or discard negligible sets, lead to much more, and in particular to the unique minimal sets of coordinates associated with measurable sets and functions (254R). Of course these results can be elegantly and effectively described in terms of L^1 and L^0 spaces, in which negligible sets are swept out of sight as the spaces are constructed. The basis of all this is the fact that the conditional expectation operators associated with subproducts multiply together in the simplest possible way (254Ra); but some further idea is needed to show that if $\mathcal{J}$ is a non-empty family of subsets of I, then $L^0_{\bigcap\mathcal{J}} = \bigcap_{J\in\mathcal{J}} L^0_J$ (see part (b) of the proof of 254R, and 254Xp(iii)).

254Sa is a version of the 'zero-one law' (272O below). 254Sb is a strong version of the principle that measurable sets in a product must be approximable by sets determined by a *finite* set of coordinates (254Fe, 254Qa, 254Xa). Evidently it is not a coincidence that the set W of 254Tb is negligible. In §272 I will revisit many of the ideas of 254R-254S and 254Xp, in particular, in the more general context of 'independent σ-algebras'.

Finally, 254U and 254Ye hardly belong to this section at all; they are unfinished business from §251. They are here because the construction of 254A-254C is the simplest way to produce an adequately complex probability space (Y,T,ν).

255 Convolutions of functions

I devote a section to a construction which is of great importance – and will in particular be very useful in Chapters 27 and 28 – and may also be regarded as a series of exercises on the work so far.

I find it difficult to know how much repetition to indulge in in this section, because the natural unified expression of the ideas is in the theory of topological groups, and I do not think we are yet ready for the general theory (I will come to it in Chapter 44 in Volume 4). The groups we need for this volume are

$\mathbb{R}$;

$\mathbb{R}^r$, for $r \ge 2$;

$S^1 = \{z : z \in \mathbb{C}, |z| = 1\}$, the 'circle group';

$\mathbb{Z}$, the group of integers.

All the ideas already appear in the theory of convolutions on $\mathbb{R}$, and I will therefore present this material in relatively detailed form, before sketching the forms appropriate to the groups $\mathbb{R}^r$ and S^1 (or $]-\pi,\pi]$); $\mathbb{Z}$ can I think be safely left to the exercises.

255A This being a book on measure theory, it is perhaps appropriate for me to emphasize, as the basis of the theory of convolutions, certain measure space isomorphisms.

Theorem Let μ be Lebesgue measure on $\mathbb{R}$ and μ_2 Lebesgue measure on $\mathbb{R}^2$; write Σ, Σ_2 for their domains.

(a) For any $a \in \mathbb{R}$, the map $x \mapsto a + x : \mathbb{R} \to \mathbb{R}$ is a measure space automorphism of $(\mathbb{R}, \Sigma, \mu)$.

(b) The map $x \mapsto -x : \mathbb{R} \to \mathbb{R}$ is a measure space automorphism of $(\mathbb{R}, \Sigma, \mu)$.

(c) For any $a \in \mathbb{R}$, the map $x \mapsto a - x : \mathbb{R} \to \mathbb{R}$ is a measure space automorphism of $(\mathbb{R}, \Sigma, \mu)$.

(d) The map $(x, y) \mapsto (x + y, y) : \mathbb{R}^2 \to \mathbb{R}^2$ is a measure space automorphism of $(\mathbb{R}^2, \Sigma_2, \mu_2)$.

(e) The map $(x, y) \mapsto (x - y, y) : \mathbb{R}^2 \to \mathbb{R}^2$ is a measure space automorphism of $(\mathbb{R}^2, \Sigma_2, \mu_2)$.

Remark I ought to remark that (b), (d) and (e) may be regarded as simple special cases of Theorem 263A in the next chapter. I nevertheless feel that it is worth writing out separate proofs here, partly because the general case of linear operators dealt with in 263A requires some extra machinery not needed here, but more because the result here has nothing to do with the *linear* structure of $\mathbb{R}$ and $\mathbb{R}^2$; it is exclusively dependent on the *group* structure of $\mathbb{R}$, together with the links between its topology and measure, and the arguments I give now are adaptable to the proper generalizations to abelian topological groups.

proof (a) This is just the translation-invariance of Lebesgue measure, dealt with in §134. There I showed that if $E \in \Sigma$ then $E + a \in \Sigma$ and $\mu(E + a) = \mu E$ (134Ab); that is, writing $\phi(x) = x + a$, $\mu(\phi[E])$ exists and is equal to μE for every $E \in \Sigma$. But of course we also have

$$\mu(\phi^{-1}[E]) = \mu(E + (-a)) = \mu E$$

for every $E \in \Sigma$, so ϕ is an automorphism.

(b) The point is that $\mu^*(A) = \mu^*(-A)$ for every $A \subseteq \mathbb{R}$. **P** (I follow the definitions of Volume 1.) If $\epsilon > 0$, there is a sequence $\langle I_n \rangle_{n\in\mathbb{N}}$ of half-open intervals covering A with $\sum_{n=0}^\infty \mu I_n \le \mu^* A + \epsilon$. Now $-A \subseteq \bigcup_{n\in\mathbb{N}}(-I_n)$. But if $I_n = [a_n, b_n[$ then $-I_n =]-b_n, a_n]$, so

$$\mu^*(-A) \le \sum_{n=0}^\infty \mu(-I_n) = \sum_{n=0}^\infty \max(0, -a_n - (-b_n)) = \sum_{n=0}^\infty \mu I_n \le \mu^* A + \epsilon.$$

As ϵ is arbitrary, $\mu^*(-A) \le \mu^* A$. Also of course $\mu^* A \le \mu^*(-(-A)) = \mu^* A$, so $\mu^*(-A) = \mu^* A$. **Q**

This means that, setting $\phi(x) = -x$ this time, ϕ is an automorphism of the structure $(\mathbb{R}, \mu^*)$. But since μ is defined from μ^* by the abstract procedure of Carathéodory's method, ϕ must also be an automorphism of the structure $(\mathbb{R}, \Sigma, \mu)$.

(c) Put (a) and (b) together; $x \mapsto a - x$ is the composition of the automorphisms $x \mapsto -x$ and $x \mapsto a + x$, and the composition of automorphisms is surely an automorphism.

(d)(i) Write T for the set $\{E : E \in \Sigma_2,\ \phi[E] \in \Sigma_2\}$, where this time $\phi(x, y) = (x + y, y)$ for x, $y \in \mathbb{R}$, so that $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ is a bijection. Then T is a σ-algebra, being the intersection of the σ-algebras Σ_2 and $\{E : \phi[E] \in \Sigma_2\} = \{\phi^{-1}[F] : F \in \Sigma_2\}$. Moreover, $\mu_2 E = \mu_2(\phi[E])$ for every $E \in$ T. **P** By 252D, we have

$$\mu_2 E = \int \mu\{x : (x, y) \in E\}\mu(dy).$$

But applying the same result to $\phi[E]$ we have

$$\mu_2\phi[E] = \int \mu\{x : (x, y) \in \phi[E]\}\mu(dy) = \int \mu\{x : (x - y, y) \in E\}\mu(dy)$$
$$= \int \mu(E^{-1}[\{y\}] + y)\mu(dy) = \int \mu E^{-1}[\{y\}]\mu(dy)$$

(because Lebesgue measure is translation-invariant)

$$= \mu_2 E. \ \textbf{Q}$$

(ii) Now ϕ and ϕ^{-1} are clearly continuous, so that $\phi[G]$ is open, and therefore measurable, for every open G; consequently all open sets must belong to T. Because T is a σ-algebra, it contains all Borel sets. Now let E be any measurable set. Then there are Borel sets H_1, H_2 such that $H_1 \subseteq E \subseteq H_2$ and $\mu_2(H_2 \setminus H_1) = 0$ (134Fb). We have $\phi[H_1] \subseteq \phi[E] \subseteq \phi[H_2]$ and

$$\mu(\phi[H_2] \setminus \phi[H_1]) = \mu\phi[H_2 \setminus H_1] = \mu(H_2 \setminus H_1) = 0.$$

Thus $\phi[E] \setminus \phi[H_1]$ must be negligible, therefore measurable, and $\phi[E] = \phi[H_1] \cup (\phi[E] \setminus \phi[H_1])$ is measurable. This shows that $\phi[E]$ is measurable whenever E is.

(iii) Repeating the same arguments with $-y$ in the place of y, we see that $\phi^{-1}[E]$ is measurable, and $\mu_2\phi^{-1}[E] = \mu_2 E$, for every $E \in \Sigma_2$. So ϕ is an automorphism of the structure $(\mathbb{R}^2, \Sigma_2, \mu_2)$.

(e) Of course this is an immediate corollary either of the proof of (d) or of (d) itself as stated, since $(x,y) \mapsto (x-y,y)$ is just the inverse of $(x,y) \mapsto (x+y,y)$.

255B Corollary (a) If $a \in \mathbb{R}$, then for any complex-valued function f defined on a subset of $\mathbb{R}$

$$\int f(x)dx = \int f(a+x)dx = \int f(-x)dx = \int f(a-x)dx$$

in the sense that if one of the integrals exists so do the others, and they are then all equal.

(b) If f is a complex-valued function defined on a subset of $\mathbb{R}^2$, then

$$\int f(x+y,y)d(x,y) = \int f(x-y,y)d(x,y) = \int f(x,y)d(x,y)$$

in the sense that if one of the integrals exists and is finite so does the other, and they are then equal.

255C Remarks (a) I am not sure whether it ought to be 'obvious' that if (X, Σ, μ), (Y, T, ν) are measure spaces and $\phi : X \to Y$ is an isomorphism, then for any function f defined on a subset of Y

$$\int f(\phi(x))\mu(dx) = \int f(y)\nu(dy)$$

in the sense that if one is defined so is the other, and they are then equal. If it is obvious then the obviousness must be contingent on the nature of the definition of integration: integrability with respect to the measure μ is something which depends on the structure (X, Σ, μ) and on no other properties of X. If it is not obvious then it is an easy deduction from Theorem 235A above, applied in turn to ϕ and ϕ^{-1} and to the real and imaginary parts of f. In any case the isomorphisms of 255A are just those needed to prove 255B.

(b) Note that in 255Bb I write $\int f(x,y)d(x,y)$ to emphasize that I am considering the integral of f with respect to two-dimensional Lebesgue measure. The fact that

$$\int\Bigl(\int f(x,y)dx\Bigr)dy = \int\Bigl(\int f(x+y,y)dx\Bigr)dy = \int\Bigl(\int f(x-y,y)dx\Bigr)dy$$

is actually easier, being an immediate consequence of the equality $\int f(a+x)dx = \int f(x)dx$. But applications of this result often depend essentially on the fact that the functions $(x,y) \mapsto f(x+y,y)$, $(x,y) \mapsto f(x-y,y)$ are measurable as functions of two variables.

(c) I have moved directly to complex-valued functions because these are necessary for the applications in Chapter 28. If however they give you any discomfort, either technically or aesthetically, all the measure-theoretic ideas of this section are already to be found in the real case, and you may wish at first to read it as if only real numbers were involved.

255D A further corollary of 255A will be useful.

Corollary Let f be a complex-valued function defined on a subset of $\mathbb{R}$.

(a) If f is measurable, then the functions $(x,y) \mapsto f(x+y)$, $(x,y) \mapsto f(x-y)$ are measurable.

(b) If f is defined almost everywhere in $\mathbb{R}$, then the functions $(x,y) \mapsto f(x+y)$, $(x,y) \mapsto f(x-y)$ are defined almost everywhere in $\mathbb{R}^2$.

proof Writing $g_1(x,y) = f(x+y)$, $g_2(x,y) = f(x-y)$ whenever these are defined, we have

$$g_1(x,y) = (f \otimes \chi\mathbb{R})(\phi(x,y)), \quad g_2(x,y) = (f \otimes \chi\mathbb{R})(\phi^{-1}(x,y)),$$

writing $\phi(x,y) = (x+y,y)$ as in 255B(d-e), and $(f \otimes \chi\mathbb{R})(x,y) = f(x)$, following the notation of 253B. By 253C, $f \otimes \chi\mathbb{R}$ is measurable if f is, and defined almost everywhere if f is. Because ϕ is a measure space automorphism, $(f \otimes \chi\mathbb{R})\phi = g_1$ and $(f \otimes \chi\mathbb{R})\phi^{-1} = g_2$ are measurable, or defined almost everywhere, if f is.

255E The basic formula Let f and g be measurable complex-valued functions defined almost everywhere in $\mathbb{R}$. Write $f * g$ for the function defined by the formula

$$(f * g)(x) = \int f(x-y)g(y)dy$$

whenever the integral exists (with respect to Lebesgue measure, naturally) as a complex number. Then $f * g$ is the **convolution** of the functions f and g.

Observe that $\mathrm{dom}(|f| * |g|) = \mathrm{dom}(f * g)$, and that $|f * g| \le |f| * |g|$ everywhere on their common domain, for all f and g.

Remark Note that I am here prepared to contemplate the convolution of f and g for arbitrary members of $\mathcal{L}^0_{\mathbb{C}}$, the space of almost-everywhere-defined measurable complex-valued functions, even though the domain of $f * g$ may be empty.

255F Elementary properties (a) Because integration is linear, we surely have

$$((f_1 + f_2) * g)(x) = (f_1 * g)(x) + (f_2 * g)(x),$$

$$(f * (g_1 + g_2))(x) = (f * g_1)(x) + (f * g_2)(x),$$

$$(cf * g)(x) = (f * cg)(x) = c(f * g)(x)$$

whenever the right-hand sides of the formulae are defined.

(b) If f and g are measurable complex-valued functions defined almost everywhere in $\mathbb{R}$, then $f * g = g * f$, in the strict sense that they have the same domain and the same value at each point of that common domain.

P Take $x \in \mathbb{R}$ and apply 255Ba to see that

$$\begin{aligned}(f * g)(x) &= \int f(x-y)g(y)dy = \int f(x-(x-y))g(x-y)dy \\ &= \int f(y)g(x-y)dy = (g * f)(x)\end{aligned}$$

if either is defined. **Q**

(c) If f_1, f_2, g_1, g_2 are measurable complex-valued functions defined almost everywhere in $\mathbb{R}$, $f_1 =_{\text{a.e.}} f_2$ and $g_1 =_{\text{a.e.}} g_2$, then $f_1 * g_1 = f_2 * g_2$. **P** For every $x \in \mathbb{R}$ we shall have $f_1(x-y) = f_2(x-y)$ for almost every $y \in \mathbb{R}$, by 255Ac. Consequently $f_1(x-y)g_1(y) = f_2(x-y)g_2(y)$ for almost every y, and $(f_1 * g_1)(x) = (f_2 * g_2)(x)$ in the sense that if one of these is defined so is the other, and they are then equal. **Q**

It follows that if u, $v \in L^0_{\mathbb{C}}$, then we have a function $\theta(u,v)$ which is equal to $f * g$ whenever f, $g \in \mathcal{L}^0_{\mathbb{C}}$ are such that $f^\bullet = u$ and $g^\bullet = u$. Observe that $\theta(u,v) = \theta(v,u)$, and that $\theta(u_1+u_2,v)$ extends $\theta(u_1,v)+\theta(u_2,v)$, $\theta(cu,v)$ extends $c\theta(u,v)$ for all u, u_1, u_2, $v \in L^0_{\mathbb{C}}$ and $c \in \mathbb{C}$.

255G I have grouped 255Fa-255Fc together because they depend only on ideas up to and including 255Ac and 255Ba. Using the second halves of 255A and 255B we get much deeper. I begin with what seems to be the fundamental result.

Theorem Let f, g and h be measurable complex-valued functions defined almost everywhere in $\mathbb{R}$.

(a) Suppose that $\int h(x+y)f(x)g(y)d(x,y)$ exists in $\mathbb{C}$. Then

$$\begin{aligned}\int h(x)(f * g)(x)dx &= \int h(x+y)f(x)g(y)d(x,y) \\ &= \iint h(x+y)f(x)g(y)dxdy = \iint h(x+y)f(x)g(y)dydx\end{aligned}$$

provided that in the expression $h(x)(f * g)(x)$ we interpret the product as 0 if $h(x) = 0$ and $(f * g)(x)$ is undefined.

(b) If, on a similar interpretation of $|h(x)|(|f| * |g|)(x)$, the integral $\int |h(x)|(|f| * |g|)(x)dx$ is finite, then $\int h(x+y)f(x)g(y)d(x,y)$ exists in $\mathbb{C}$.

proof Consider the functions

$$k_1(x,y) = h(x)f(x-y)g(y), \quad k_2(x,y) = h(x+y)f(x)g(y)$$

wherever these are defined. 255D tells us that k_1 and k_2 are measurable and defined almost everywhere. Now setting $\phi(x,y) = (x+y,y)$, we have $k_2 = k_1\phi$, so that

$$\int k_1(x,y)d(x,y) = \int k_2(x,y)d(x,y)$$

if either exists, by 255Bb.

If

$$\int h(x+y)f(x)g(y)d(x,y) = \int k_2$$

exists, then by Fubini's theorem we have

$$\int k_2 = \int k_1(x,y)d(x,y) = \int(\int h(x)f(x-y)g(y)dy)dx$$

so $\int h(x)f(x-y)g(y)dy$ exists almost everywhere, that is, $(f*g)(x)$ exists for almost every x such that $h(x) \neq 0$; on the interpretation I am using here, $h(x)(f*g)(x)$ exists almost everywhere, and

$$\begin{aligned}\int h(x)(f*g)(x)dx &= \int\Big(\int h(x)f(x-y)g(y)dy\Big)dx = \int k_1\\ &= \int k_2 = \int h(x+y)f(x)g(y)d(x,y)\\ &= \iint h(x+y)f(x)g(y)dxdy = \iint h(x+y)f(x)g(y)dydx\end{aligned}$$

by Fubini's theorem again.

If (on the same interpretation) $|h| \times (|f| * |g|)$ is integrable,

$$|k_1(x,y)| = |h(x)||f(x-y)||g(y)|$$

is measurable, and

$$\iint |h(x)||f(x-y)||g(y)|dydx = \int |h(x)|(|f|*|g|)(x)dx$$

is finite, so by Tonelli's theorem (252G, 252H) k_1 and k_2 are integrable.

255H Certain standard results are now easy.

Corollary If f, g are complex-valued functions which are integrable over $\mathbb{R}$, then $f*g$ is integrable, with

$$\int f*g = \int f\int g, \quad \int |f*g| \le \int |f|\int |g|.$$

proof In 255G, set $h(x) = 1$ for every $x \in \mathbb{R}$; then

$$\int h(x+y)f(x)g(y)d(x,y) = \int f(x)g(y)d(x,y) = \int f\int g$$

by 253D, so

$$\int f*g = \int h(x)(f*g)(x)dx = \int h(x+y)f(x)g(y)d(x,y) = \int f\int g,$$

as claimed. Now

$$\int |f*g| \le \int |f|*|g| = \int |f|\int |g|.$$

255I Corollary For any measurable complex-valued functions f, g defined almost everywhere in $\mathbb{R}$, $f*g$ is measurable and has measurable domain.

proof Set $f_n(x) = f(x)$ if $x \in \operatorname{dom} f$, $|x| \le n$ and $|f(x)| \le n$, and 0 elsewhere in $\mathbb{R}$; define g_n similarly from g. Then f_n and g_n are integrable, $|f_n| \le |f|$ and $|g_n| \le |g|$ almost everywhere, $f =_{\text{a.e.}} \lim_{n\to\infty} f_n$ and $g =_{\text{a.e.}} \lim_{n\to\infty} g_n$. Consequently, by Lebesgue's Dominated Convergence Theorem,

$$\begin{aligned}(f*g)(x) &= \int f(x-y)g(y)dy = \int \lim_{n\to\infty} f_n(x-y)g_n(y)dy\\ &= \lim_{n\to\infty}\int f_n(x-y)g_n(y)dy = \lim_{n\to\infty}(f_n*g_n)(x)\end{aligned}$$

for every $x \in \operatorname{dom} f*g$. But f_n*g_n is integrable, therefore measurable, for every n, so that $f*g$ must be measurable.

As for the domain of $f*g$,

$$\begin{aligned} x \in \operatorname{dom}(f*g) &\iff \int f(x-y)g(y)dy \text{ is defined in } \mathbb{C} \\ &\iff \int |f(x-y)||g(y)|dy \text{ is defined in } \mathbb{R} \\ &\iff \int |f_n(x-y)||g_n(y)|dy \text{ is defined in } \mathbb{R} \text{ for every } n \\ &\qquad \text{and } \sup_{n\in\mathbb{N}} \int |f_n(x-y)||g_n(y)|dy < \infty. \end{aligned}$$

Because every $|f_n|*|g_n|$ is integrable, therefore measurable and with measurable domain,

$$\operatorname{dom}(f*g) = \{x : x \in \bigcap_{n\in\mathbb{N}} \operatorname{dom}(|f_n|*|g_n|),\ \sup_{n\in\mathbb{N}}(|f_n|*|g_n|)(x) < \infty\}$$

is measurable.

255J Theorem Let f, g and h be complex-valued measurable functions, defined almost everywhere in $\mathbb{R}$, such that $f*g$ and $g*h$ are defined a.e. Suppose that $x \in \mathbb{R}$ is such that one of $(|f|*(|g|*|h|))(x)$, $((|f|*|g|)*|h|)(x)$ is defined in $\mathbb{R}$. Then $f*(g*h)$ and $(f*g)*h$ are defined and equal at x.

proof Set $k(y) = f(x-y)$ when this is defined, so that k is measurable and defined almost everywhere (255D).

(a) If $(|f|*(|g|*|h|))(x)$ is defined, this is $\int |k(y)|(|g|*|h|)(y)dy$, so by 255G we have

$$\int k(y)(g*h)(y)dy = \int k(y+z)g(y)h(z)d(y,z),$$

that is,

$$\begin{aligned} (f*(g*h))(x) &= \int f(x-y)(g*h)(y)dy = \int k(y)(g*h)(y)dy \\ &= \int k(y+z)g(y)h(z)d(y,z) = \iint k(y+z)g(y)h(z)dydz \\ &= \iint f(x-y-z)g(y)h(z)dydz = \int (f*g)(x-z)h(z)dz \\ &= ((f*g)*h)(x). \end{aligned}$$

(b) If $((|f|*|g|)*|h|)(x)$ is defined, this is

$$\begin{aligned} \int (|f|*|g|)(x-z)|h(z)|dz &= \iint |f(x-z-y)||g(y)||h(z)|dydz \\ &= \iint |k(y+z)||g(y)||h(z)|dydz. \end{aligned}$$

By 255D again, $(y,z) \mapsto k(y+z)$ is measurable, so we can apply Tonelli's theorem to see that $\int k(y+z)g(y)h(z)d(y,z)$ is defined, and is equal to $\int k(y)(g*h)(y)dy = (f*(g*h))(x)$ by 255Ga. On the other side, by the last two lines of the proof of (a), $\int k(y+z)g(y)h(z)d(y,z)$ is also equal to $((f*g)*h)(x)$.

255K I do not think we shall need an exhaustive discussion of the question of just when $(f*g)(x)$ is defined; this seems to be complicated. However there is a fundamental case in which we can be sure that $(f*g)(x)$ is defined everywhere.

Proposition Suppose that f, g are measurable complex-valued functions defined almost everywhere in $\mathbb{R}$, and that $f \in \mathcal{L}^p_{\mathbb{C}}$, $g \in \mathcal{L}^q_{\mathbb{C}}$ where p, $q \in [1,\infty]$ and $\frac{1}{p}+\frac{1}{q} = 1$ (writing $\frac{1}{\infty} = 0$ as usual). Then $f*g$ is defined everywhere in $\mathbb{R}$, is uniformly continuous, and

$$\begin{aligned} \sup_{x\in\mathbb{R}} |(f*g)(x)| &\le \|f\|_p\|g\|_q \text{ if } 1 < p < \infty,\ 1 < q < \infty, \\ &\le \|f\|_1 \operatorname{ess\,sup}|g| \text{ if } p = 1,\ q = \infty, \\ &\le \operatorname{ess\,sup}|f| \cdot \|g\|_1 \text{ if } p = \infty,\ q = 1. \end{aligned}$$

proof (a) (For an introduction to $\mathcal{L}^p$ spaces, see §244.) For any $x \in \mathbb{R}$, the function f_x, defined by setting $f_x(y) = f(x-y)$ whenever $x-y \in \operatorname{dom} f$, must also belong to $\mathcal{L}^p$, because $f_x = f\phi$ for an automorphism ϕ of the measure space. Consequently $(f*g)(x) = \int f_x \times g$ is defined, and of modulus at most $\|f\|_p\|g\|_q$ or $\|f\|_1 \operatorname{ess\,sup}|g|$ or $\operatorname{ess\,sup}|f| \cdot \|g\|_1$, by 244Eb/244Pb and 243Fa/243K.

(b) To see that $f*g$ is uniformly continuous, argue as follows. Suppose first that $p<\infty$. Let $\epsilon>0$. Let $\eta>0$ be such that $(2+2^{1/p})\|g\|_q\eta \le \epsilon$. Then there is a bounded continuous function $h:\mathbb{R}\to\mathbb{C}$ such that $\{x : h(x)\ne 0\}$ is bounded and $\|f-h\|_p \le \eta$ (244Hb/244Pb); let $M\ge 1$ be such that $h(x)=0$ whenever $|x|\ge M-1$. Next, h is uniformly continuous, so there is a $\delta \in \,]0,1]$ such that $|h(x)-h(x')| \le M^{-1/p}\eta$ whenever $|x-x'|\le\delta$.

Suppose that $|x-x'|\le\delta$. Defining $h_x(y)=h(x-y)$, as before, we have

$$\int |h_x - h_{x'}|^p = \int |h(x-y)-h(x'-y)|^p dy = \int |h(t)-h(x'-x+t)|^p dt$$

(substituting $t=x-y$)

$$= \int_{-M}^{M} |h(t)-h(x'-x+t)|^p dt$$

(because $h(t)=h(x'-x+t)=0$ if $|t|\ge M$)

$$\le 2M(M^{-1/p}\eta)^p$$

(because $|h(t)-h(x'-x+t)| \le M^{-1/p}\eta$ for every t)

$$= 2\eta^p.$$

So $\|h_x-h_{x'}\|_p \le 2^{1/p}\eta$. On the other hand,

$$\int |h_x - f_x|^p = \int |h(x-y)-f(x-y)|^p dy = \int |h(y)-f(y)|^p dy,$$

so $\|h_x-f_x\|_p = \|h-f\|_p \le \eta$, and similarly $\|h_{x'}-f_{x'}\|_p \le \eta$. So

$$\|f_x - f_{x'}\|_p \le \|f_x - h_x\|_p + |h_x - h_{x'}\|_p + \|h_{x'} - f_{x'}\|_p \le (2+2^{1/p})\eta.$$

This means that

$$|(f*g)(x)-(f*g)(x')| = |\int f_x \times g - \int f_{x'} \times g| = |\int (f_x - f_{x'}) \times g|$$
$$\le \|f_x - f_{x'}|_p\|g\|_q \le (2+2^{1/p})\|g\|_q\eta \le \epsilon.$$

As ϵ is arbitrary, $f*g$ is uniformly continuous.

The argument here supposes that p is finite. But if $p=\infty$ then $q=1$ is finite, so we can apply the method with g in place of f to show that $g*f$ is uniformly continuous, and $f*g=g*f$ by 255Fb.

255L The r-dimensional case I have written 255A-255K out as theorems about Lebesgue measure on $\mathbb{R}$. However they all apply equally well to Lebesgue measure on $\mathbb{R}^r$ for any $r\ge 1$, and the modifications required are so small that I think I need do no more than ask you to read through the arguments again, turning every $\mathbb{R}$ into an $\mathbb{R}^r$, and every $\mathbb{R}^2$ into an $(\mathbb{R}^r)^2$. In 255A and elsewhere, the measure μ_2 should be read either as Lebesgue measure on $\mathbb{R}^{2r}$ or as the product measure on $(\mathbb{R}^r)^2$; by 251N the two may be identified. There is a trivial modification required in part (b) of the proof; if $I_n = [a_n, b_n[$ then

$$\mu I_n = \mu(-I_n) = \prod_{i=1}^r \max(0, \beta_{ni} - \alpha_{ni}),$$

writing $a_n = (\alpha_{n1}, \dots, \alpha_{nr})$. In the proof of 255I, the functions f_n should be defined by saying that $f_n(x)=f(x)$ if $|f(x)|\le n$ and $\|x\|\le n$, 0 otherwise.

In quoting these results, therefore, I shall be uninhibited in referring to the paragraphs 255A-255K as if they were actually written out for general $r\ge 1$.

255M The case of $]-\pi,\pi]$ The same ideas also apply to the circle group S^1 and to the interval $]-\pi,\pi]$, but here perhaps rather more explanation is in order.

(a) The first thing to establish is the appropriate group operation. If we think of S^1 as the set $\{z : z\in\mathbb{C}, |z|=1\}$, then the group operation is complex multiplication, and in the formulae above $x+y$ must be rendered as xy, while $x-y$ must be rendered as xy^{-1}. On the interval $]-\pi,\pi]$, the group operation is $+_{2\pi}$, where for x, $y \in \,]-\pi,\pi]$ I write

$x +_{2\pi} y$ for whichever of $x+y$, $x+y+2\pi$, $x+y-2\pi$ belongs to $]-\pi,\pi]$. To see that this is indeed a group operation, one method is to note that it corresponds to multiplication on S^1 if we use the canonical bijection $x \mapsto e^{ix} :]-\pi,\pi] \to S^1$; another, to note that it corresponds to the operation on the quotient group $\mathbb{R}/2\pi\mathbb{Z}$. Thus in this interpretation of the ideas of 255A-255K, we shall wish to replace $x+y$ by $x +_{2\pi} y$, $-x$ by $-_{2\pi}x$, and $x-y$ by $x -_{2\pi} y$, where

$$-_{2\pi}x = -x \text{ if } x \in \,]-\pi,\pi[, \quad -_{2\pi}\pi = \pi,$$

and $x -_{2\pi} y$ is whichever of $x-y$, $x-y+2\pi$, $x-y-2\pi$ belongs to $]-\pi,\pi]$.

(b) As for the measure, the measure to use on $]-\pi,\pi]$ is just Lebesgue measure. Note that because $]-\pi,\pi]$ is Lebesgue measurable, there will be no confusion concerning the meaning of 'measurable subset', as the relatively measurable subsets of $]-\pi,\pi]$ are actually measured by Lebesgue measure on $\mathbb{R}$. Also we can identify the product measure on $]-\pi,\pi] \times]-\pi,\pi]$ with the subspace measure induced by Lebesgue measure on $\mathbb{R}^2$ (251R).

On S^1, we need the corresponding measure induced by the canonical bijection between S^1 and $]-\pi,\pi]$, which indeed is often called 'Lebesgue measure on S^1'. (We shall see in 265E that it is also equal to Hausdorff one-dimensional measure on S^1.) We are very close to the level at which it would become reasonable to move to S^1 and this measure (or its normalized version, in which it is reduced by a factor of 2π, so as to make S^1 a probability space). However, the elementary theory of Fourier series, which will be the principal application of this work in the present volume, is generally done on intervals in $\mathbb{R}$, so that formulae based on $]-\pi,\pi]$ are closer to the standard expressions. Henceforth, therefore, I will express the work in terms of $]-\pi,\pi]$.

(c) The result corresponding to 255A now takes a slightly different form, so I spell it out.

255N Theorem Let μ be Lebesgue measure on $]-\pi,\pi]$ and μ_2 Lebesgue measure on $]-\pi,\pi] \times]-\pi,\pi]$; write Σ, Σ_2 for their domains.

(a) For any $a \in \,]-\pi,\pi]$, the map $x \mapsto a +_{2\pi} x :]-\pi,\pi] \to]-\pi,\pi]$ is a measure space automorphism of $(]-\pi,\pi],\Sigma,\mu)$.

(b) The map $x \mapsto -_{2\pi}x :]-\pi,\pi] \to]-\pi,\pi]$ is a measure space automorphism of $(]-\pi,\pi],\Sigma,\mu)$.

(c) For any $a \in \,]-\pi,\pi]$, the map $x \mapsto a -_{2\pi} x :]-\pi,\pi] \to]-\pi,\pi]$ is a measure space automorphism of $(]-\pi,\pi],\Sigma,\mu)$.

(d) The map $(x,y) \mapsto (x +_{2\pi} y, y) :]-\pi,\pi]^2 \to]-\pi,\pi]^2$ is a measure space automorphism of $(]-\pi,\pi]^2,\Sigma_2,\mu_2)$.

(e) The map $(x,y) \mapsto (x -_{2\pi} y, y) :]-\pi,\pi]^2 \to]-\pi,\pi]^2$ is a measure space automorphism of $(]-\pi,\pi]^2,\Sigma_2,\mu_2)$.

proof (a) Set $\phi(x) = a +_{2\pi} x$. Then for any $E \subseteq \,]-\pi,\pi]$,

$$\phi[E] = ((E+a) \cap]-\pi,\pi]) \cup (((E+a) \cap]\pi,3\pi]) - 2\pi) \cup (((E+a) \cap]-3\pi,-\pi]) + 2\pi),$$

and these three sets are disjoint, so that

$$\begin{aligned}\mu\phi[E] &= \mu((E+a) \cap]-\pi,\pi]) + \mu(((E+a) \cap]\pi,3\pi]) - 2\pi) \\ &\qquad + \mu(((E+a) \cap]-3\pi,-\pi]) + 2\pi) \\ &= \mu_L((E+a) \cap]-\pi,\pi]) + \mu_L(((E+a) \cap]\pi,3\pi]) - 2\pi) \\ &\qquad + \mu_L(((E+a) \cap]-3\pi,-\pi]) + 2\pi)\end{aligned}$$

(writing μ_L for Lebesgue measure on $\mathbb{R}$)

$$\begin{aligned}&= \mu_L((E+a) \cap]-\pi,\pi]) + \mu_L((E+a) \cap]\pi,3\pi]) + \mu_L((E+a) \cap]-3\pi,-\pi]) \\ &= \mu_L(E+a) = \mu_L E = \mu E.\end{aligned}$$

Similarly, $\mu\phi^{-1}[E]$ is defined and equal to μE for every $E \in \Sigma$, so that ϕ is an automorphism of $(]-\pi,\pi],\Sigma,\mu)$.

(b) Of course this is quicker. Setting $\phi(x) = -_{2\pi}x$ for $x \in \,]-\pi,\pi]$, we have

$$\begin{aligned}\mu(\phi[E]) &= \mu(\phi[E] \cap]-\pi,\pi[) = \mu(-(E \cap]-\pi,\pi[) \\ &= \mu_L(-(E \cap]-\pi,\pi[)) = \mu_L(E \cap]-\pi,\pi[) \\ &= \mu(E \cap]-\pi,\pi[) = \mu E\end{aligned}$$

for every $E \in \Sigma$.

(c) This is just a matter of putting (a) and (b) together, as in 255A.

(d) We can argue as in (a), but with a little more elaboration. If $E \in \Sigma_2$, and $\phi(x,y) = (x +_{2\pi} y, y)$ for x, $y \in \,]-\pi,\pi]$, set $\psi(x,y) = (x+y,y)$ for x, $y \in \mathbb{R}$, and write $c = (2\pi,0) \in \mathbb{R}^2$, $H = \,]-\pi,\pi]^2$, $H' = H + c$, $H'' = H - c$. Then for any $E \in \Sigma_2$,

$$\phi[E] = (\psi[E] \cap H) \cup ((\psi[E] \cap H') - c) \cup ((\psi[E] \cap H'') + c),$$

so

$$\begin{aligned}\mu_2\phi[E] &= \mu_2(\psi[E] \cap H) + \mu_2((\psi[E] \cap H') - c) + \mu_2((\psi[E] \cap H'') + c)\\ &= \mu_L(\psi[E] \cap H) + \mu_L((\psi[E] \cap H') - c) + \mu_L((\psi[E] \cap H'') + c)\end{aligned}$$

(this time writing μ_L for Lebesgue measure on $\mathbb{R}^2$)

$$\begin{aligned}&= \mu_L(\psi[E] \cap H) + \mu_L(\psi[E] \cap H') + \mu_L(\psi[E] \cap H'')\\ &= \mu_L\psi[E] = \mu_L E = \mu_2 E.\end{aligned}$$

In the same way, $\mu_2(\phi^{-1}[E]) = \mu_2 E$ for every $E \in \Sigma_2$, so ϕ is an automorphism of $(\left]-\pi,\pi\right]^2, \Sigma_2, \mu_2)$, as required.

(e) Finally, (e) is just a restatement of (d), as before.

255O Convolutions on $\left]-\pi,\pi\right]$ With the fundamental result established, the same arguments as in 255B-255K now yield the following. Write μ for Lebesgue measure on $\left]-\pi,\pi\right]$.

(a) Let f and g be measurable complex-valued functions defined almost everywhere in $\left]-\pi,\pi\right]$. Write $f * g$ for the function defined by the formula

$$(f * g)(x) = \int_{-\pi}^{\pi} f(x -_{2\pi} y)g(y)dy$$

whenever $x \in \left]-\pi,\pi\right]$ and the integral exists as a complex number. Then $f * g$ is the **convolution** of the functions f and g.

(b) If f and g are measurable complex-valued functions defined almost everywhere in $\left]-\pi,\pi\right]$, then $f * g = g * f$.

(c) Let f, g and h be measurable complex-valued functions defined almost everywhere in $\left]-\pi,\pi\right]$. Then

(i)

$$\int_{-\pi}^{\pi} h(x)(f * g)(x)dx = \int_{\left]-\pi,\pi\right]^2} h(x +_{2\pi} y)f(x)g(y)d(x,y)$$

whenever the right-hand side exists and is finite, provided that in the expression $h(x)(f*g)(x)$ we interpret the product as 0 if $h(x) = 0$ and $(f * g)(x)$ is undefined.

(ii) If, on the same interpretation of $|h(x)|(|f| * |g|)(x)$, the integral $\int_{-\pi}^{\pi} |h(x)|(|f| * |g|)(x)dx$ is finite, then $\int_{\left]-\pi,\pi\right]^2} h(x +_{2\pi} y)f(x)g(y)d(x,y)$ exists in $\mathbb{C}$, so again we shall have

$$\int_{-\pi}^{\pi} h(x)(f * g)(x)dx = \int_{\left]-\pi,\pi\right]^2} h(x +_{2\pi} y)f(x)g(y)d(x,y).$$

(d) If f, g are complex-valued functions which are integrable over $\left]-\pi,\pi\right]$, then $f * g$ is integrable, with

$$\int_{-\pi}^{\pi} f * g = \int_{-\pi}^{\pi} f \int_{-\pi}^{\pi} g, \quad \int_{-\pi}^{\pi} |f * g| \le \int_{-\pi}^{\pi} |f| \int_{-\pi}^{\pi} |g|.$$

(e) Let f, g, h be complex-valued measurable functions defined almost everywhere in $\left]-\pi,\pi\right]$, such that $f*g$ and $g*h$ are also defined almost everywhere. Suppose that $x \in \left]-\pi,\pi\right]$ is such that one of $(|f| * (|g| * |h|))(x)$, $((|f| * |g|) * |h|)(x)$ is defined in $\mathbb{R}$. Then $f * (g * h)$ and $(f * g) * h$ are defined and equal at x.

(f) Suppose that $f \in \mathcal{L}^p_{\mathbb{C}}(\mu)$, $g \in \mathcal{L}^q_{\mathbb{C}}(\mu)$ where p, $q \in [1,\infty]$ and $\frac{1}{p} + \frac{1}{q} = 1$. Then $f * g$ is defined everywhere in $\left]-\pi,\pi\right]$, and $\sup_{x\in\left]-\pi,\pi\right]} |(f * g)(x)| \le \|f\|_p\|g\|_q$, interpreting $\|\ \|_\infty$ as ess sup$|\ |$, as in 255K.

255X Basic exercises >(a) Let f, g be complex-valued functions defined almost everywhere in $\mathbb{R}$. Show that for any $x \in \mathbb{R}$, $(f * g)(x) = \int f(x + y)g(-y)dy$ if either is defined.

>(b) Let f and g be complex-valued functions defined almost everywhere in $\mathbb{R}$. (i) Show that if f and g are even functions, so is $f * g$. (ii) Show that if f is even and g is odd then $f * g$ is odd. (iii) Show that if f and g are odd then $f * g$ is even.

(c) Suppose that f, g are real-valued measurable functions defined almost everywhere in $\mathbb{R}^r$ and such that $f > 0$ a.e., $g \ge 0$ a.e. and $\{x : g(x) > 0\}$ is not negligible. Show that $f * g > 0$ everywhere in $\mathrm{dom}(f * g)$.

>(d) Suppose that $f : \mathbb{R} \to \mathbb{C}$ is a bounded differentiable function and that f' is bounded. Show that for any integrable complex-valued function g on $\mathbb{R}$, $f * g$ is differentiable and $(f * g)' = f' * g$ everywhere. (*Hint*: 123D.)

(e) A complex-valued function g defined almost everywhere in $\mathbb{R}$ is **locally integrable** if $\int_a^b g$ is defined in $\mathbb{C}$ whenever $a < b$ in $\mathbb{R}$. Suppose that g is such a function and that $f : \mathbb{R} \to \mathbb{C}$ is a differentiable function, with continuous derivative, such that $\{x : f(x) \ne 0\}$ is bounded. Show that $(f * g)' = f' * g$ everywhere.

>(f) Set $\phi_\delta(x) = \exp(-\frac{1}{\delta^2 - x^2})$ if $|x| < \delta$, 0 if $|x| \ge \delta$, as in 242Xi. Set $\alpha_\delta = \int \phi_\delta$, $\psi_\delta = \alpha_\delta^{-1}\phi_\delta$. Let f be a locally integrable complex-valued function on $\mathbb{R}$. (i) Show that $f * \psi_\delta$ is a smooth function defined everywhere on $\mathbb{R}$ for every $\delta > 0$. (ii) Show that $\lim_{\delta\downarrow 0}(f * \psi_\delta)(x) = f(x)$ for almost every $x \in \mathbb{R}$. (*Hint*: 223Yg.) (iii) Show that if f is integrable then $\lim_{\delta\downarrow 0}\int |f - f * \psi_\delta| = 0$. (*Hint*: use (ii) and 245H(a-ii) *or* look first at the case $f = \chi[a,b]$ and use 242O, noting that $\int |f * \psi_\delta| \le \int |f|$.) (iv) Show that if f is uniformly continuous and defined everywhere in $\mathbb{R}$ then $\lim_{\delta\downarrow 0}\sup_{x\in\mathbb{R}}|f(x) - (f * \psi_\delta)(x)| = 0$.

>(g) For $\alpha > 0$, set $g_\alpha(t) = \dfrac{1}{\Gamma(\alpha)}t^{\alpha-1}$ for $t > 0$, 0 for $t \le 0$. Show that $g_\alpha * g_\beta = g_{\alpha+\beta}$ for all α, $\beta > 0$. (*Hint*: 252Yf.)

>(h) Let μ be Lebesgue measure on $\mathbb{R}$. For u, v, $w \in L^0_{\mathbb{C}} = L^0_{\mathbb{C}}(\mu)$, say that $u * v = w$ if $f * g$ is defined almost everywhere and $(f * g)^\bullet = w$ whenever f, $g \in \mathcal{L}^0_{\mathbb{C}}(\mu)$, $f^\bullet = u$ and $g^\bullet = w$. (i) Show that $(u_1 + u_2) * v = u_1 * v + u_2 * v$ whenever u_1, u_2, $v \in L^0_{\mathbb{C}}$ and $u_1 * v$ and $u_2 * v$ are defined in this sense. (ii) Show that $u * v = v * u$ whenever u, $v \in L^0(\mathbb{C})$ and either $u * v$ or $v * u$ is defined. (iii) Show that if u, v, $w \in L^0_{\mathbb{C}}$, $u * v$ and $v * w$ are defined, and either $|u| * (|v| * |w|)$ or $(|u| * |v|) * |w|$ is defined, then then $u * (v * w) = (u * v) * w$ are defined and equal.

>(i) Let μ be Lebesgue measure on $\mathbb{R}$. (i) Show that $u * v$, as defined in 255Xh, belongs to $L^1_{\mathbb{C}}(\mu)$ whenever u, $v \in L^1_{\mathbb{C}}(\mu)$. (ii) Show that $L^1_{\mathbb{C}}$ is a commutative Banach algebra under $*$ (definition: 2A4Jb).

(j)(i) Show that if h is an integrable function on $\mathbb{R}^2$, then $(Th)(x) = \int h(x - y, y)dy$ exists for almost every $x \in \mathbb{R}$, and that $\int (Th)(x)dx = \int h(x,y)d(x,y)$. (ii) Write μ_2 for Lebesgue measure on $\mathbb{R}^2$, μ for Lebesgue measure on $\mathbb{R}$. Show that there is a linear operator $\tilde{T} : L^1(\mu_2) \to L^1(\mu)$ defined by setting $\tilde{T}(h^\bullet) = (Th)^\bullet$ for every integrable function h on $\mathbb{R}^2$. (iii) Show that in the language of 253E and 255Xh, $\tilde{T}(u \otimes v) = u * v$ for all u, $v \in L^1(\mu)$.

>(k) For $\boldsymbol{a}$, $\boldsymbol{b} \in \mathbb{C}^{\mathbb{Z}}$ set $(\boldsymbol{a} * \boldsymbol{b})(n) = \sum_{i\in\mathbb{Z}}\boldsymbol{a}(n-i)\boldsymbol{b}(i)$ whenever $\sum_{i\in\mathbb{Z}}|\boldsymbol{a}(n-i)\boldsymbol{b}(i)| < \infty$. Show that
(i) $\boldsymbol{a} * \boldsymbol{b} = \boldsymbol{b} * \boldsymbol{a}$;
(ii) $\sum_{i\in\mathbb{Z}}\boldsymbol{c}(i)(\boldsymbol{a} * \boldsymbol{b})(i) = \sum_{i,j\in\mathbb{Z}}\boldsymbol{c}(i+j)\boldsymbol{a}(i)\boldsymbol{b}(j)$ if $\sum_{i,j\in\mathbb{Z}}|\boldsymbol{c}(i+j)\boldsymbol{a}(i)\boldsymbol{b}(j)| < \infty$;
(iii) if $\boldsymbol{a}$, $\boldsymbol{b} \in \ell^1(\mathbb{Z})$ then $\boldsymbol{a} * \boldsymbol{b} \in \ell^1(\mathbb{Z})$ and $\|\boldsymbol{a} * \boldsymbol{b}\|_1 \le \|\boldsymbol{a}\|_1\|\boldsymbol{b}\|_1$;
(iv) If $\boldsymbol{a}$, $\boldsymbol{b} \in \ell^2(\mathbb{Z})$ then $\boldsymbol{a} * \boldsymbol{b} \in \ell^\infty(\mathbb{Z})$ and $\|\boldsymbol{a} * \boldsymbol{b}\|_\infty \le \|\boldsymbol{a}\|_2\|\boldsymbol{b}\|_2$;
(v) if $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c} \in \mathbb{C}^{\mathbb{Z}}$ and $(|\boldsymbol{a}| * (|\boldsymbol{b}| * |\boldsymbol{c}|))(n)$ is well-defined, then $(\boldsymbol{a} * (\boldsymbol{b} * \boldsymbol{c}))(n) = ((\boldsymbol{a} * \boldsymbol{b}) * \boldsymbol{c})(n)$.

255Y Further exercises (a) Let f be a complex-valued function which is integrable over $\mathbb{R}$. (i) Let x be any point of the Lebesgue set of f. Show that for any $\epsilon > 0$ there is a $\delta > 0$ such that $|f(x) - (f * g)(x)| \le \epsilon$ whenever $g : \mathbb{R} \to [0, \infty[$ is a function which is non-decreasing on $]-\infty, 0]$, non-decreasing on $[0, \infty[$, and has $\int g = 1$ and $\int_{-\delta}^{\delta} g \ge 1 - \delta$. (ii) Show that for any $\epsilon > 0$ there is a $\delta > 0$ such that $\|f - f * g\|_1 \le \epsilon$ whenever $g : \mathbb{R} \to [0, \infty[$ is a function which is non-decreasing on $]-\infty, 0]$, non-decreasing on $[0, \infty[$, and has $\int g = 1$ and $\int_{-\delta}^{\delta} g \ge 1 - \delta$.

(b) Let f be a complex-valued function which is integrable over $\mathbb{R}$. Show that, for almost every $x \in \mathbb{R}$,

$$\lim_{a\to\infty}\frac{a}{\pi}\int_{-\infty}^{\infty}\frac{f(y)}{1+a^2(x-y)^2}dy, \quad \lim_{a\to\infty}\frac{1}{a}\int_x^{\infty}f(y)e^{-a(y-x)}dy,$$

$$\lim_{\sigma\downarrow 0}\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty}f(y)e^{-(y-x)^2/2\sigma^2}dy$$

all exist and are equal to $f(x)$. (*Hint*: 263G.)

(c) Set $f(x) = 1$ for all $x \in \mathbb{R}$, $g(x) = \dfrac{x}{|x|}$ for $0 < |x| \le 1$ and 0 otherwise, $h(x) = \tanh x$ for all $x \in \mathbb{R}$. Show that $f * (g * h)$ and $(f * g) * h$ are both defined (and constant) everywhere, and are different.

(d) Discuss what can happen if, in the context of 255J, we know that $(|f| * (|g| * |h|))(x)$ is defined, but have no information on the domain of $f * g$.

(e) Suppose that $p \in [1, \infty[$ and that $f \in \mathcal{L}^p_{\mathbb{C}}(\mu)$, where μ is Lebesgue measure on $\mathbb{R}^r$. For $a \in \mathbb{R}^r$ set $(S_a f)(x) = f(a + x)$ whenever $a + x \in \operatorname{dom} f$. Show that $S_a f \in \mathcal{L}^p_{\mathbb{C}}(\mu)$, and that for every $\epsilon > 0$ there is a $\delta > 0$ such that $\|S_a f - f\|_p \le \epsilon$ whenever $|a| \le \delta$.

(f) Suppose that $p, q \in]1, \infty[$ and $\frac{1}{p} + \frac{1}{q} = 1$. Take $f \in \mathcal{L}^p_{\mathbb{C}}(\mu)$ and $g \in \mathcal{L}^q_{\mathbb{C}}(\mu)$, where μ is Lebesgue measure on $\mathbb{R}^r$. Show that $\lim_{\|x\| \to \infty} (f * g)(x) = 0$. (*Hint*: use 244Hb.)

(g) Repeat 255Ye and 255K, this time taking μ to be Lebesgue measure on $]-\pi, \pi]$, and setting $(S_a f)(x) = f(a +_{2\pi} x)$ for $a \in]-\pi, \pi]$; show that in the new version of 255K, $(f * g)(\pi) = \lim_{x \downarrow -\pi} (f * g)(x)$.

(h) Let μ be Lebesgue measure on $\mathbb{R}$. For $a \in \mathbb{R}$, $f \in \mathcal{L}^0 = \mathcal{L}^0(\mu)$ set $(S_a f)(x) = f(a + x)$ whenever $a + x \in \operatorname{dom} f$.

(i) Show that $S_a f \in \mathcal{L}^0$ for every $f \in \mathcal{L}^0$.

(ii) Show that we have a map $\tilde{S}_a : L^0 \to L^0$ defined by setting $\tilde{S}_a(f^\bullet) = (S_a f)^\bullet$ for every $f \in \mathcal{L}^0$.

(iii) Show that $\tilde{S}_a$ is a Riesz space isomorphism and is a homeomorphism for the topology of convergence in measure; moreover, that $\tilde{S}_a(u \times v) = \tilde{S}_a u \times \tilde{S}_a v$ for all $u, v \in L^0$.

(iv) Show that $\tilde{S}_{a+b} = \tilde{S}_a \tilde{S}_b$ for all $a, b \in \mathbb{R}$.

(v) Show that $\lim_{a \to 0} \tilde{S}_a u = u$ for the topology of convergence in measure, for every $u \in L^0$.

(vi) Show that if $1 \le p \le \infty$ then $\tilde{S}_a \restriction L^p$ is an isometric isomorphism of the Banach lattice L^p.

(vii) Show that if $p \in [1, \infty[$ then $\lim_{a \to 0} \|\tilde{S}_a u - u\|_p = 0$ for every $u \in L^p$.

(viii) Show that if $A \subseteq L^1$ is uniformly integrable and $M \ge 0$, then $\{\tilde{S}_a u : u \in A, |a| \le M\}$ is uniformly integrable.

(ix) Suppose that $u, v \in L^0$ are such that $u * v$ is defined in L^0 in the sense of 255Xh. Show that $\tilde{S}_a(u * v) = (\tilde{S}_a u) * v = u * (\tilde{S}_a v)$ for every $a \in \mathbb{R}$.

(i) Prove 255Nd from 255Na by the method used to prove 255Ad from 255Aa, rather than by quoting 255Ad.

(j) Let μ be Lebesgue measure on $\mathbb{R}$, and $\phi : \mathbb{R} \to \mathbb{R}$ a convex function; let $\bar{\phi} : L^0 \to L^0 = L^0(\mu)$ be the associated operator (see 241I). Show that if $u \in L^1 = L^1(\mu)$, $v \in L^0$ are such that $u \ge 0$, $\int u = 1$ and $u * v$, $u * \bar{\phi}(v)$ are both defined in the sense of 255Xh, then $\bar{\phi}(u * v) \le u * \bar{\phi}(v)$. (*Hint*: 233I.)

(k) Let μ be Lebesgue measure on $\mathbb{R}$, and $p \in [1, \infty]$. Let $f \in \mathcal{L}^1_{\mathbb{C}}(\mu)$, $g \in \mathcal{L}^p_{\mathbb{C}}(\mu)$. Show that $f * g \in \mathcal{L}^p_{\mathbb{C}}(\mu)$ and that $\|f * g\|_p \le \|f\|_1 \|g\|_p$. (*Hint*: argue from 255Yj, as in 244M.)

(l) Suppose that $p, q, r \in]1, \infty[$ and that $\frac{1}{p} + \frac{1}{q} = 1 + \frac{1}{r}$. Let μ be Lebesgue measure on $\mathbb{R}$. (i) Show that

$$\int f \times g \le \|f\|_p^{1-p/r} \|g\|_q^{1-q/r} (\int f^p \times g^q)^{1/r}$$

whenever $f, g \ge 0$ and $f \in \mathcal{L}^p(\mu)$, $g \in \mathcal{L}^q(\mu)$. (*Hint*: set $p' = p/(p-1)$, etc.; $f_1 = f^{p/q'}$, $g_1 = g^{q/p'}$, $h = (f^p \times g^q)^{1/r}$. Use 244Xc to see that $\|f_1 \times g_1\|_{r'} \le \|f_1\|_{q'} \|g_1\|_{p'}$, so that $\int f_1 \times g_1 \times h \le \|f_1\|_{q'} \|g_1\|_{p'} \|h\|_r$.) (ii) Show that $f * g$ is defined a.e. and that $\|f * g\|_r \le \|f\|_p \|g\|_q$ for all $f \in \mathcal{L}^p(\mu)$, $g \in \mathcal{L}^q(\mu)$. (*Hint*: take $f, g \ge 0$. Use (i) to see that $(f * g)(x)^r \le \|f\|_p^{r-p} \|g\|_q^{r-q} \int f(y)^p g(x - y)^q dy$, so that $\|f * g\|_r^r \le \|f\|_p^{r-p} \|g\|_q^{r-q} \int f(y)^p \|g\|_q^q dy$.) (This is **Young's inequality**.)

(m) Repeat the results of this section for the group $(S^1)^r$, where $r \ge 2$, given its product measure.

(n) Let G be a group and μ a σ-finite measure on G such that (α) for every $a \in G$, the map $x \mapsto ax$ is an automorphism of (G, μ) (β) the map $(x, y) \mapsto (x, xy)$ is an automorphism of (G^2, μ_2), where μ_2 is the c.l.d. product measure on $G \times G$. For $f, g \in \mathcal{L}^0_{\mathbb{C}}(\mu)$ write $(f * g)(x) = \int f(y) g(y^{-1}x) dy$ whenever this is defined. Show that

(i) if $f, g, h \in \mathcal{L}^0_{\mathbb{C}}(\mu)$ and $\int h(xy) f(x) g(y) d(x, y)$ is defined in $\mathbb{C}$, then $\int h(x)(f * g)(x) dx$ exists and is equal to $\int h(xy) f(x) g(y) d(x, y)$, provided that in the expression $h(x)(f * g)(x)$ we interpret the product as 0 if $h(x) = 0$ and $(f * g)(x)$ is undefined;

(ii) if $f, g \in \mathcal{L}^1_{\mathbb{C}}(\mu)$ then $f * g \in \mathcal{L}^1_{\mathbb{C}}(\mu)$ and $\int f * g = \int f \int g$, $\|f * g\|_1 \le \|f\|_1 \|g\|_1$;

(iii) if $f, g, h \in \mathcal{L}^1_{\mathbb{C}}(\mu)$ then $f * (g * h) = (f * g) * h$.

(See Halmos 50, §59.)

(o) Repeat 255Yn for counting measure on any group G.

255 Notes and comments I have tried to set this section out in such a way that it will be clear that the basis of all the work here is 255A, and the crucial application is 255G. I hope that if and when you come to look at general topological groups (for instance, in Chapter 44), you will find it easy to trace through the ideas in any abelian topological group for which you can prove a version of 255A. For non-abelian groups, of course, rather more care is necessary, especially as in some important examples we no longer have $\mu\{x^{-1} : x \in E\} = \mu E$ for every E; see 255Yn-255Yo for a little of what can be done without using topological ideas.

The critical point in 255A is the move from the one-dimensional results in 255Aa-255Ac, which are just the translation- and reflection-invariance of Lebesgue measure, to the two-dimensional results in 255Ac-255Ad. And the living centre of the argument, as I present it, is the fact that the shear transformation ϕ is an automorphism of the structure $(\mathbb{R}^2, \Sigma_2)$. The actual calculation of $\mu_2\phi[E]$, assuming that it is measurable, is an easy application of Fubini's and Tonelli's theorems and the translation-invariance of μ. It is for this step that we absolutely need the topological properties of Lebesgue measure. I should perhaps remind you that the fact that ϕ is a homeomorphism is not sufficient; in 134I I described a homeomorphism of the unit interval which does not preserve measurability, and it is easy to adapt this to produce a homeomorphism $\psi : \mathbb{R}^2 \to \mathbb{R}^2$ such that $\psi[E]$ is not always measurable for measurable E. The argument of 255A is dependent on the special relationships between all three of the measure, topology and group structure of $\mathbb{R}$.

I have already indulged in a few remarks on what ought, or ought not, to be 'obvious' (255C). But perhaps I can add that such results as 255B and the later claim, in the proof of 255K, that a reflected version of a function in $\mathcal{L}^p$ is also in $\mathcal{L}^p$, can only be trivial consequences of results like 255A if every step in the construction of the integral is done in the abstract context of general measure spaces. Even though we are here working exclusively with the Lebesgue integral, the argument will become untrustworthy if we have at any stage in the definition of the integral even mentioned that we are thinking of Lebesgue measure. I advance this as a solid reason for defining 'integration' on abstract measure spaces from the beginning, as I did in Volume 1. Indeed, I suggest that generally in pure mathematics there are good reasons for casting arguments into the forms appropriate to the arguments themselves.

I am writing this book for readers who are interested in proofs, and as elsewhere I have written the proofs of this section out in detail. But most of us find it useful to go through some material in 'advanced calculus' mode, by which I mean starting with a formula such as

$$(f * g)(x) = \int f(x-y)g(y)dy,$$

and then working out consequences by formal manipulations, for instance

$$\int h(x)(f * g)(x)dx = \iint h(x)f(x-y)g(y)dydx = \iint h(x+y)f(x)g(y)dydx,$$

without troubling about the precise applicability of the formulae to begin with. In some ways this formula-driven approach can be more truthful to the structure of the subject than the careful analysis I habitually present. The exact hypotheses necessary to make the theorems strictly true are surely secondary, in such contexts as this section, to the pattern formed by the ensemble of the theorems, which can be adequately and elegantly expressed in straightforward formulae. Of course I do still insist that we cannot properly appreciate the structure, nor safely use it, without mastering the ideas of the proofs – and as I have said elsewhere, I believe that mastery of ideas necessarily includes mastery of the formal details, at least in the sense of being able to reconstruct them fairly fluently on demand.

Throughout the main exposition of this section, I have worked with functions rather than equivalence classes of functions. But all the results here have interpretations of great importance for the theory of the 'function spaces' of Chapter 24. In 255Xh and the succeeding exercises, I have pointed to a definition of convolution as an operator from a subset of $L^0 \times L^0$ to L^0. It is an interesting point that if u, $v \in L^0$ then $u * v$ can be interpreted as a *function*, not as a member of L^0 (255Fc). Thus 255H can be regarded as saying that $u * v \in \mathcal{L}^1$ for u, $v \in L^1$. We cannot quite say that convolution is a bilinear operator from $L^1 \times L^1$ to $\mathcal{L}^1$, because $\mathcal{L}^1$, as I define it, is not strictly speaking a linear space. If we want a bilinear operator, then we have to regard convolution as a function from $L^1 \times L^1$ to L^1. But when we look at convolution as a function on $L^2 \times L^2$, for instance, then our functions $u * v$ are defined everywhere (255K), and indeed are continuous functions vanishing at ∞ (255Ye-255Yf). So in this case it seems more appropriate to regard convolution as a bilinear operator from $L^2 \times L^2$ to some space of continuous functions, and not as an operator from $L^2 \times L^2$ to L^∞. For an example of an interesting convolution which is not naturally representable in terms of an operator on L^p spaces, see 255Xg.

Because convolution acts as a continuous bilinear operator from $L^1(\mu) \times L^1(\mu)$ to $L^1(\mu)$, where μ is Lebesgue measure on $\mathbb{R}$, Theorem 253F tells us that it must correspond to a linear operator from $L^1(\mu_2)$ to $L^1(\mu)$, where μ_2 is Lebesgue measure on $\mathbb{R}^2$. This is the operator $\tilde{T}$ of 255Xj.

So far in these notes I have written as though we were concerned only with Lebesgue measure on $\mathbb{R}$. However many applications of the ideas involve $\mathbb{R}^r$ or $]-\pi, \pi]$ or S^1. The move to $\mathbb{R}^r$ should be elementary. The move to S^1 does require a re-formulation of the basic result 255A/255N. It should also be clear that there will be no new difficulties in moving to $]-\pi,\pi]^r$ or $(S^1)^r$. Moreover, we can also go through the whole theory for the groups $\mathbb{Z}$ and $\mathbb{Z}^r$, where the appropriate measure is now counting measure, so that $L^0_{\mathbb{C}}$ becomes identified with $\mathbb{C}^{\mathbb{Z}}$ or $\mathbb{C}^{\mathbb{Z}^r}$ (255Xk, 255Yo).

256 Radon measures on $\mathbb{R}^r$

In the next section, and again in Chapters 27 and 28, we need to consider the principal class of measures on Euclidean spaces. For a proper discussion of this class, and the interrelationships between the measures and the topologies involved, we must wait until Volume 4. For the moment, therefore, I present definitions adapted to the case in hand, warning you that the correct generalizations are not quite obvious. I give the definition (256A) and a characterization (256C) of Radon measures on Euclidean spaces, and theorems on the construction of Radon measures as indefinite integrals (256E, 256J), as image measures (256G) and as product measures (256K). In passing I give a version of Lusin's theorem concerning measurable functions on Radon measure spaces (256F).

256A Definitions Let ν be a measure on $\mathbb{R}^r$, where $r \geq 1$, and Σ its domain.

(a) ν is a **topological measure** if every open set belongs to Σ. Note that in this case every Borel set, and in particular every closed set, belongs to Σ.

(b) ν is **locally finite** if every bounded set has finite outer measure.

(c) If ν is a topological measure, it is **inner regular with respect to the compact sets** if

$$\nu E = \sup\{\nu K : K \subseteq E \text{ is compact}\}$$

for every $E \in \Sigma$. (Because ν is a topological measure, and compact sets are closed (2A2Ec), νK is defined for every compact set K.)

(d) ν is a **Radon measure** if it is a complete locally finite topological measure which is inner regular with respect to the compact sets.

256B It will be convenient to be able to call on the following elementary facts.

Lemma Let ν be a Radon measure on $\mathbb{R}^r$, and Σ its domain.

(a) ν is σ-finite.

(b) For any $E \in \Sigma$ and any $\epsilon > 0$ there are a closed set $F \subseteq E$ and an open set $G \supseteq E$ such that $\nu(G \setminus F) \leq \epsilon$.

(c) For every $E \in \Sigma$ there is a set $H \subseteq E$, expressible as the union of a sequence of compact sets, such that $\nu(E \setminus H) = 0$.

(d) Every continuous real-valued function on $\mathbb{R}^r$ is Σ-measurable.

(e) If $h : \mathbb{R}^r \to \mathbb{R}$ is continuous and has bounded support, then h is ν-integrable.

proof (a) For each $n \in \mathbb{N}$, $B(\mathbf{0},n) = \{x : \|x\| \leq n\}$ is a closed bounded set, therefore Borel. So if ν is a Radon measure on $\mathbb{R}^r$, $\langle B(\mathbf{0},n)\rangle_{n\in\mathbb{N}}$ is a cover of $\mathbb{R}^r$ by a sequence of sets of finite measure.

(b) Set $E_n = \{x : x \in E,\, n \leq \|x\| < n+1\}$ for each n. Then $\nu E_n < \infty$, so there is a compact set $K_n \subseteq E_n$ such that $\nu K_n \geq \nu E_n - 2^{-n-2}\epsilon$. Set $F = \bigcup_{n\in\mathbb{N}} K_n$; then

$$\nu(E \setminus F) = \sum_{n=0}^{\infty} \nu(E_n \setminus K_n) \leq \frac{1}{2}\epsilon.$$

Also $F \subseteq E$ and F is closed because

$$F \cap B(\mathbf{0},n) = \bigcup_{i \leq n} K_i \cap B(\mathbf{0},n)$$

is closed for each n.

In the same way, there is a closed set $F' \subseteq \mathbb{R}^r \setminus E$ such that $\nu((\mathbb{R}^r \setminus E) \setminus F') \leq \frac{1}{2}\epsilon$. Setting $G = \mathbb{R}^r \setminus F'$, we see that G is open, that $G \supseteq E$ and that $\nu(G \setminus E) \leq \frac{1}{2}\epsilon$, so that $\nu(G \setminus F) \leq \epsilon$, as required.

(c) By (b), we can choose for each $n \in \mathbb{N}$ a closed set $F_n \subseteq E$ such that $\nu(E \setminus F_n) \leq 2^{-n}$. Set $H = \bigcup_{n\in\mathbb{N}} F_n$; then $H \subseteq E$ and $\nu(E \setminus H) = 0$, and also $H = \bigcup_{m,n\in\mathbb{N}} B(\mathbf{0},m) \cap F_n$ is a countable union of compact sets.

(d) If $h : \mathbb{R}^r \to \mathbb{R}$ is continuous, all the sets $\{x : h(x) > a\}$ are open, so belong to Σ.

(e) By (d), h is measurable. Now we are supposing that there is some $n \in \mathbb{N}$ such that $h(x) = 0$ whenever $x \notin B(\mathbf{0},n)$. Since $B(\mathbf{0},n)$ is compact (2A2F), h is bounded on $B(\mathbf{0},n)$ (2A2G), and we have $|h| \leq \gamma\chi B(\mathbf{0},n)$ for some γ; since $\nu B(\mathbf{0},n)$ is finite, h is ν-integrable.

256C Theorem A measure ν on $\mathbb{R}^r$ is a Radon measure iff it is the completion of a locally finite measure defined on the σ-algebra $\mathcal{B}$ of Borel subsets of $\mathbb{R}^r$.

proof (a) Suppose first that ν is a Radon measure. Write Σ for its domain.

(i) Set $\nu_0 = \nu\restriction\mathcal{B}$. Then ν_0 is a measure with domain $\mathcal{B}$, and it is locally finite because $\nu_0 B(\mathbf{0}, n) = \nu B(\mathbf{0}, n)$ is finite for every n. Let $\hat{\nu}_0$ be the completion of ν_0 (212C).

(ii) If $\hat{\nu}_0$ measures E, there are E_1, $E_2 \in \mathcal{B}$ such that $E_1 \subseteq E \subseteq E_2$ and $\nu_0(E_2 \setminus E_1) = 0$. Now $E \setminus E_1 \subseteq E_2 \setminus E_1$ must be ν-negligible; as ν is complete, $E \in \Sigma$ and

$$\nu E = \nu E_1 = \nu_0 E_1 = \hat{\nu}_0 E.$$

(iii) If $E \in \Sigma$, then by 256Bc there is a Borel set $H \subseteq E$ such that $\nu(E \setminus H) = 0$. Equally, there is a Borel set $H' \subseteq \mathbb{R}^r \setminus E$ such that $\nu((\mathbb{R}^r \setminus E) \setminus H') = 0$, so that we have $H \subseteq E \subseteq \mathbb{R}^r \setminus H'$ and

$$\nu_0((\mathbb{R}^r \setminus H') \setminus H) = \nu((\mathbb{R}^r \setminus H') \setminus H) = 0.$$

So $\hat{\nu}_0 E$ is defined and equal to $\nu_0 E_1 = \nu E$.

This shows that $\nu = \hat{\nu}_0$ is the completion of the locally finite Borel measure $\nu\restriction\mathcal{B}$. And this is true for any Radon measure ν on $\mathbb{R}^r$.

(b) For the rest of the proof, I suppose that ν_0 is a locally finite measure on $\mathbb{R}^r$ and ν is its completion. Write Σ for the domain of ν. We say that a subset of $\mathbb{R}^r$ is a **K_σ set** if it is expressible as the union of a sequence of compact sets. Note that every K_σ set is a Borel set, so belongs to Σ. Set

$$\mathcal{A} = \{E : E \in \Sigma, \text{ there is a } \mathrm{K}_\sigma \text{ set } H \subseteq E \text{ such that } \nu(E \setminus H) = 0\},$$

$$\Sigma = \{E : E \in \mathcal{A}, \mathbb{R}^r \setminus E \in \mathcal{A}\}.$$

(c)(i) Every open set is itself a K_σ set, so belongs to $\mathcal{A}$. **P** Let $G \subseteq \mathbb{R}^r$ be open. If $G = \emptyset$ then G is compact and the result is trivial. Otherwise, let $\mathcal{I}$ be the set of closed intervals of the form $[q, q']$, where q, $q' \in \mathbb{Q}^r$, which are included in G. Then all the members of $\mathcal{I}$ are closed and bounded, therefore compact. If $x \in G$, there is a $\delta > 0$ such that $B(x, \delta) = \{y : \|y - x\| \le \delta\} \subseteq G$; now there is an $I \in \mathcal{I}$ such that $x \in I \subseteq B(x, \delta)$. Thus $G = \bigcup \mathcal{I}$. But $\mathcal{I}$ is countable, so G is K_σ. **Q**

(ii) Every closed subset of $\mathbb{R}$ is K_σ, so belongs to $\mathcal{A}$. **P** If $F \subseteq \mathbb{R}$ is closed, then $F = \bigcup_{n\in\mathbb{N}} F \cap B(\mathbf{0}, n)$; but every $F \cap B(\mathbf{0}, n)$ is closed and bounded, therefore compact. **Q**

(iii) If $\langle E_n \rangle_{n\in\mathbb{N}}$ is any sequence in $\mathcal{A}$, then $E = \bigcup_{n\in\mathbb{N}} E_n$ belongs to $\mathcal{A}$. **P** For each $n \in \mathbb{N}$ we have a countable family $\mathcal{K}_n$ of compact subsets of E_n such that $\nu(E_n \setminus \bigcup \mathcal{K}_n) = 0$; now $\mathcal{K} = \bigcup_{n\in\mathbb{N}} \mathcal{K}_n$ is a countable family of compact subsets of E, and $E \setminus \bigcup \mathcal{K} \subseteq \bigcup_{n\in\mathbb{N}} (E_n \setminus \bigcup \mathcal{K}_n)$ is ν-negligible. **Q**

(iv) If $\langle E_n \rangle_{n\in\mathbb{N}}$ is any sequence in $\mathcal{A}$, then $F = \bigcap_{n\in\mathbb{N}} E_n \in \mathcal{A}$. **P** For each $n \in \mathbb{N}$, let $\langle K_{ni} \rangle_{i\in\mathbb{N}}$ be a sequence of compact subsets of E_n such that $\nu(E_n \setminus \bigcup_{i\in\mathbb{N}} K_{ni}) = 0$. Set $K'_{nj} = \bigcup_{i\le j} K_{ni}$ for each j, so that

$$\nu(E_n \cap H) = \lim_{j\to\infty} \nu(K'_{nj} \cap H)$$

for every $H \in \Sigma$. Now, for each m, $n \in \mathbb{N}$, choose $j(m, n)$ such that

$$\nu(E_n \cap B(\mathbf{0}, m) \cap K'_{n,j(m,n)}) \ge \nu(E_n \cap B(\mathbf{0}, m)) - 2^{-(m+n)}.$$

Set $K_m = \bigcap_{n\in\mathbb{N}} K'_{n,j(m,n)}$; then K_m is closed (being an intersection of closed sets) and bounded (being a subset of $K'_{0,j(m,0)}$), therefore compact. Also $K_m \subseteq F$, because $K'_{n,j(m,n)} \subseteq E_n$ for each n, and

$$\nu(F \cap B(\mathbf{0}, m) \setminus K_m) \le \sum_{n=0}^{\infty} \nu(E_n \cap B(\mathbf{0}, m) \setminus K'_{n,j(m,n)}) \le \sum_{n=0}^{\infty} 2^{-(m+n)} = 2^{-m+1}.$$

Consequently $H = \bigcup_{m\in\mathbb{N}} K_m$ is a K_σ subset of F and

$$\nu(F \cap B(\mathbf{0}, m) \setminus H) \le \inf_{k\ge m} \nu(F \cap B(\mathbf{0}, k) \setminus H_k) = 0$$

for every m, so $\nu(F \setminus H) = 0$ and $F \in \mathcal{A}$. **Q**

(d) Σ is a σ-algebra of subsets of $\mathbb{R}$. **P** (i) $\emptyset$ and its complement are open, so belong to $\mathcal{A}$ and therefore to Σ. (ii) If $E \in \Sigma$ then both $\mathbb{R}^r \setminus E$ and $\mathbb{R}^r \setminus (\mathbb{R}^r \setminus E) = E$ belong to $\mathcal{A}$, so $\mathbb{R}^r \setminus E \in \Sigma$. (iii) Let $\langle E_n \rangle_{n\in\mathbb{N}}$ be a sequence in Σ with union E. By (a-iii) and (a-iv),

$$E \in \mathcal{A}, \quad \mathbb{R}^r \setminus E = \bigcap_{n\in\mathbb{N}} (\mathbb{R}^r \setminus E_n) \in \mathcal{A},$$

so $E \in \Sigma$. **Q**

(e) By (c-i) and (c-ii), every open set belongs to Σ; consequently every Borel set belongs to Σ and therefore to $\mathcal{A}$. Now if E is any member of Σ, there is a Borel set $E_1 \subseteq E$ such that $\nu(E \setminus E_1) = 0$ and a K_σ set $H \subseteq E_1$ such that $\nu(E_1 \setminus H) = 0$. Express H as $\bigcup_{n\in\mathbb{N}} K_n$ where every K_n is compact; then

$$\nu E = \nu H = \lim_{n\to\infty} \nu(\bigcup_{i\le n} K_i) \le \sup_{K\subseteq E \text{ is compact}} \nu K \le \nu E$$

because $\bigcup_{i\in n} K_i$ is a compact subset of E for every n.

(f) Thus ν is inner regular with respect to the compact sets. But of course it is complete (being the completion of ν_0) and a locally finite topological measure (because ν_0 is); so it is a Radon measure. This completes the proof.

256D Proposition If ν and ν' are two Radon measures on $\mathbb{R}^r$, the following are equiveridical:
(i) $\nu = \nu'$;
(ii) $\nu K = \nu' K$ for every compact set $K \subseteq \mathbb{R}^r$;
(iii) $\nu G = \nu' G$ for every open set $G \subseteq \mathbb{R}^r$;
(iv) $\int h\,d\nu = \int h\,d\nu'$ for every continuous function $h : \mathbb{R}^r \to \mathbb{R}$ with bounded support.

proof (a)(i)$\Rightarrow$(iv) is trivial.

(b)(iv)$\Rightarrow$(iii) If (iv) is true, and $G \subseteq \mathbb{R}^r$ is an open set, then for each $n \in \mathbb{N}$ set

$$h_n(x) = \min(1, 2^n \inf_{y\in\mathbb{R}^r\setminus(G\cap B(\mathbf{0},n))} \|y - x\|)$$

for $x \in \mathbb{R}^r$. Then h_n is continuous (in fact $|h_n(x) - h_n(x')| \le 2^n\|x - x'\|$ for all x, $x' \in \mathbb{R}^r$) and zero outside $B(\mathbf{0}, n)$, so $\int h_n d\nu = \int h_n d\nu'$. Next, $\langle h_n(x)\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence converging to $\chi G(x)$ for every $x \in \mathbb{R}^r$. So

$$\nu G = \lim_{n\to\infty} \int h_n d\nu = \lim_{n\to\infty} \int h_n d\nu' = \nu' G,$$

by 135Ga. As G is arbitrary, (iii) is true.

(c)(iii)$\Rightarrow$(ii) If (iii) is true, and $K \subseteq \mathbb{R}^r$ is compact, let n be so large that $\|x\| < n$ for every $x \in K$. Set $G = \{x : \|x\| < n\}$, $H = G \setminus K$. Then G and H are open and G is bounded, so $\nu G = \nu' G$ is finite, and

$$\nu K = \nu G - \nu H = \nu' G - \nu' H = \nu' K.$$

As K is arbitrary, (ii) is true.

(d)(ii)$\Rightarrow$(i) If ν, ν' agree on the compact sets, then

$$\nu E = \sup_{K\subseteq E \text{ is compact}} \nu K = \sup_{K\subseteq E \text{ is compact}} \nu' K = \nu' E$$

for every Borel set E. So $\nu\restriction\mathcal{B} = \nu'\restriction\mathcal{B}$, where $\mathcal{B}$ is the algebra of Borel sets. But since ν and ν' are both the completions of their restrictions to $\mathcal{B}$, they are identical.

256E It is I suppose time I gave some examples of Radon measures. However it will save a few lines if I first establish some basic constructions. You may wish to glance ahead to 256H at this point.

Theorem Let ν be a Radon measure on $\mathbb{R}^r$, with domain Σ, and f a non-negative Σ-measurable function defined on a ν-conegligible subset of $\mathbb{R}^r$. Suppose that f is **locally integrable** in the sense that $\int_E f d\nu < \infty$ for every bounded set E. Then the indefinite-integral measure ν' on $\mathbb{R}^r$ defined by saying that

$$\nu' E = \int_E f d\nu \text{ whenever } E \cap \{x : x \in \operatorname{dom} f,\ f(x) > 0\} \in \Sigma$$

is a Radon measure on $\mathbb{R}^r$.

proof For the construction of ν', see 234I-234L. It is a topological measure because every open set belongs to Σ and therefore to the domain Σ' of ν'. ν' is locally finite because f is locally integrable. To see that ν' is inner regular with respect to the compact sets, take any set $E \in \Sigma'$, and set $E' = \{x : x \in E \cap \operatorname{dom} f,\ f(x) > 0\}$. Then $E' \in \Sigma$, so there is a set $H \subseteq E'$, expressible as the union of a sequence of compact sets, such that $\nu(E' \setminus H) = 0$. In this case

$$\nu'(E \setminus H) = \int_{E\setminus H} f d\nu = 0.$$

Let $\langle K_n\rangle_{n\in\mathbb{N}}$ be a sequence of compact sets with union H; then

$$\nu' E = \nu' H = \lim_{n\to\infty} \nu'(\bigcup_{i\le n} K_i) \le \sup_{K\subseteq E \text{ is compact}} \nu' K \le \nu' E.$$

As E is arbitrary, ν' is inner regular with respect to the compact sets.

256F Theorem Let ν be a Radon measure on $\mathbb{R}^r$, and Σ its domain. Let $f : D \to \mathbb{R}$ be a Σ-measurable function, where $D \subseteq \mathbb{R}^r$. Then for every $\epsilon > 0$ there is a closed set $F \subseteq \mathbb{R}^r$ such that $\nu(\mathbb{R}^r \setminus F) \le \epsilon$ and $f{\restriction}F$ is continuous.

proof By 121I, there is a Σ-measurable function $h : \mathbb{R}^r \to \mathbb{R}$ extending f. Enumerate $\mathbb{Q}$ as $\langle q_n\rangle_{n\in\mathbb{N}}$. For each $n \in \mathbb{N}$ set $E_n = \{x : h(x) \le q_n\}$, $E'_n = \{x : h(x) > q_n\}$ and use 256Bb to choose closed sets $F_n \subseteq E_n$, $F'_n \subseteq E'_n$ such that $\nu(E_n \setminus F_n) \le 2^{-n-2}\epsilon$, $\nu(E'_n \setminus F'_n) \le 2^{-n-2}\epsilon$. Set $F = \bigcap_{n\in\mathbb{N}}(F_n \cup F'_n)$; then F is closed and

$$\nu(\mathbb{R}^r \setminus F) \le \sum_{n=0}^{\infty} \nu(\mathbb{R}^r \setminus (F_n \cup F'_n)) \le \sum_{n=0}^{\infty} \nu(E_n \setminus F_n) + \nu(E'_n \setminus F'_n) \le \epsilon.$$

I claim that $h{\restriction}F$ is continuous. **P** Suppose that $x \in F$ and $\delta > 0$. Then there are m, $n \in \mathbb{N}$ such that

$$h(x) - \delta \le q_m < h(x) \le q_n \le h(x) + \delta.$$

This means that $x \in E'_m \cap E_n$; consequently $x \notin F_m \cup F'_n$. Because $F_m \cup F'_n$ is closed, there is an $\eta > 0$ such that $y \notin F_m \cup F'_n$ whenever $\|y - x\| \le \eta$. Now suppose that $y \in F$ and $\|y - x\| \le \eta$. Then $y \in (F_m \cup F'_m) \cap (F_n \cup F'_n)$ and $y \notin F_m \cup F'_n$, so $y \in F'_m \cap F_n \subseteq E'_m \cap E_n$ and $q_m < h(y) \le q_n$. Consequently $|h(y) - h(x)| \le \delta$. As x and δ are arbitrary, $h{\restriction}F$ is continuous. **Q** Consequently $f{\restriction}F = (h{\restriction}F){\restriction}D$ is continuous, as required.

256G Theorem Let ν be a Radon measure on $\mathbb{R}^r$, with domain Σ, and suppose that $\phi : \mathbb{R}^r \to \mathbb{R}^s$ is measurable in the sense that all its coordinates are Σ-measurable. If the image measure $\nu' = \nu\phi^{-1}$ (234D) is locally finite, it is a Radon measure.

proof Write Σ for the domain of ν and Σ' for the domain of ν'. If $\phi = (\phi_1, \dots, \phi_s)$, then

$$\phi^{-1}[\{y : \eta_j \le \alpha\}] = \{x : \phi_j(x) \le \alpha\} \in \Sigma,$$

so $\{y : \eta_j \le \alpha\} \in \Sigma'$ for every $j \le s$, $\alpha \in \mathbb{R}$, where I write $y = (\eta_1, \dots, \eta_s)$ for $y \in \mathbb{R}^s$. Consequently every Borel subset of $\mathbb{R}^s$ belongs to Σ' (121J), and ν' is a topological measure. It is complete by 234Eb.

The point is of course that ν' is inner regular with respect to the compact sets. **P** Suppose that $F \in \Sigma'$ and that $\gamma < \nu' F$. For each $j \le s$, there is a closed set $H_j \subseteq \mathbb{R}^r$ such that $\phi_j{\restriction}H_j$ is continuous and $\nu(\mathbb{R}^r \setminus H_j) < \frac{1}{s}(\nu' F - \gamma)$, by 256F. Set $H = \bigcap_{j\le s} H_j$; then H is closed and $\phi{\restriction}H$ is continuous and

$$\nu(\mathbb{R}^r \setminus H) < \nu' F - \gamma = \nu\phi^{-1}[F] - \gamma,$$

so that $\nu(\phi^{-1}[F] \cap H) > \gamma$. Let $K \subseteq \phi^{-1}[F] \cap H$ be a compact set such that $\nu K \ge \gamma$, and set $L = \phi[K]$. Because $K \subseteq H$ and $\phi{\restriction}H$ is continuous, L is compact (2A2Eb). Of course $L \subseteq F$, and

$$\nu' L = \nu\phi^{-1}[L] \ge \nu K \ge \gamma.$$

As F and γ are arbitrary, ν' is inner regular with respect to the compact sets. **Q**

Since ν' is locally finite by the hypothesis of the theorem, it is a Radon measure.

256H Examples I come at last to the promised examples.

(a) Lebesgue measure on $\mathbb{R}^r$ is a Radon measure. (It is a topological measure by 115G, and inner regular with respect to the compact sets by 134Fb.)

(b) Let $\langle t_n\rangle_{n\in\mathbb{N}}$ be any sequence in $\mathbb{R}^r$, and $\langle a_n\rangle_{n\in\mathbb{N}}$ any summable sequence in $[0, \infty[$. For every $E \subseteq \mathbb{R}^r$ set

$$\nu E = \sum\{a_n : t_n \in E\}.$$

so that ν is a totally finite point-supported measure. Then ν is a (totally finite) Radon measure on $\mathbb{R}^r$. **P** Clearly ν is complete and defined on every Borel set and gives finite measure to bounded sets. To see that it is inner regular with respect to the compact sets, observe that for any $E \subseteq \mathbb{R}^r$ the sets

$$K_n = E \cap \{t_i : i \le n\}$$

are compact and $\nu E = \lim_{n\to\infty} \nu K_n$. **Q**

(c) Now we come to a new idea. Recall that the Cantor set C (134G) is a closed negligible subset of $[0, 1]$, and that the Cantor function (134H) is a non-decreasing continuous function $f : [0, 1] \to [0, 1]$ such that $f(0) = 0$, $f(1) = 1$ and f is constant on each of the intervals composing $[0, 1] \setminus C$. It follows that if we set $g(x) = \frac{1}{2}(x + f(x))$ for $x \in [0, 1]$, then $g : [0, 1] \to [0, 1]$ is a continuous bijection such that the Lebesgue measure of $g[C]$ is $\frac{1}{2}$ (134I); consequently $g^{-1} : [0, 1] \to [0, 1]$ is continuous. Now extend g to a bijection $h : \mathbb{R} \to \mathbb{R}$ by setting $h(x) = x$ for $x \in \mathbb{R} \setminus [0, 1]$. Then h and h^{-1} are continuous. Note that $h[C] = g[C]$ has Lebesgue measure $\frac{1}{2}$.

Let ν_1 be the indefinite-integral measure defined from Lebesgue measure μ on $\mathbb{R}$ and the function $2\chi(h[C])$; that is, $\nu_1 E = 2\mu(E\cap h[C])$ whenever this is defined. By 256E, ν_1 is a Radon measure, and $\nu_1 h[C] = \nu_1\mathbb{R} = 1$. Let ν be the measure $\nu_1 h$, that is, $\nu E = \nu_1 h[E]$ for just those $E \subseteq \mathbb{R}$ such that $h[E] \in \operatorname{dom}\nu_1$. Then ν is a Radon probability measure on $\mathbb{R}$, by 256G, and $\nu C = 1$, $\nu(\mathbb{R}\setminus C) = \mu C = 0$.

256I Remarks (a) The measure ν of 256Hc, sometimes called **Cantor measure**, is a classic example, and as such has many constructions, some rather more natural than the one I use here (see 256Xk, and also 264Ym below). But I choose the method above because it yields directly, without further investigation or any appeal to more advanced general theory, the fact that ν is a Radon measure.

(b) The examples above are chosen to represent the extremes under the 'Lebesgue decomposition' described in 232I. If ν is a (totally finite) Radon measure on $\mathbb{R}^r$, we can use 232Ib to express its restriction $\nu\restriction\mathcal{B}$ to the Borel σ-algebra as $\nu_p+\nu_{ac}+\nu_{cs}$, where ν_p is the 'point-mass' or 'atomic' part of $\nu\restriction\mathcal{B}$, ν_{ac} is the 'absolutely continuous' part (with respect to Lebesgue measure), and ν_{cs} is the 'atomless singular part'. In the example of 256Hb, we have $\nu\restriction\mathcal{B} = \nu_p$; in 256E, if we start from Lebesgue measure, we have $\nu\restriction\mathcal{B} = \nu_{ac}$; and in 256Hc we have $\nu\restriction\mathcal{B} = \nu_{cs}$.

256J Absolutely continuous Radon measures It is worth pausing a moment over the indefinite-integral measures described in 256E.

Proposition Let ν be a Radon measure on $\mathbb{R}^r$, where $r \geq 1$, and write μ for Lebesgue measure on $\mathbb{R}^r$. Then the following are equiveridical:

(i) ν is an indefinite-integral measure over μ;

(ii) $\nu E = 0$ whenever E is a Borel subset of $\mathbb{R}^r$ and $\mu E = 0$.

In this case, if $g \in \mathcal{L}^0(\mu)$ and $\int_E g\,d\mu = \nu E$ for every Borel set $E \subseteq \mathbb{R}^r$, then g is a Radon-Nikodým derivative of ν with respect to μ in the sense of 232Hf.

proof (a)(i)⇒(ii) If f is a Radon-Nikodým derivative of ν with respect to μ, then of course

$$\nu E = \int_E f d\mu = 0$$

whenever $\mu E = 0$.

(ii)⇒(i) If $\nu E = 0$ for every μ-negligible Borel set E, then νE is defined and equal to 0 for every μ-negligible set E, because ν is complete and any μ-negligible set is included in a μ-negligible Borel set. Consequently $\operatorname{dom}\nu$ includes the domain Σ of μ, since every Lebesgue measurable set is expressible as the union of a Borel set and a negligible set.

For each $n \in \mathbb{N}$ set $E_n = \{x : n \leq \|x\| < n+1\}$, so that $\langle E_n\rangle_{n\in\mathbb{N}}$ is a partition of $\mathbb{R}^r$ into bounded Borel sets. Set $\nu_n E = \nu(E\cap E_n)$ for every Lebesgue measurable set E and every $n \in \mathbb{N}$. Now ν_n is absolutely continuous with respect to μ (232Ba), so by the Radon-Nikodým theorem (232F) there is a μ-integrable function f_n such that $\int_E f_n d\mu = \nu_n E$ for every Lebesgue measurable set E. Because $\nu_n E \geq 0$ for every $E \in \Sigma$, $f_n \geq 0$ a.e.; because $\nu_n(\mathbb{R}^r\setminus E_n) = 0$, $f_n = 0$ a.e. on $\mathbb{R}^r \setminus E_n$. Now if we set

$$f = \max(0, \sum_{n=0}^{\infty} f_n),$$

f will be defined μ-a.e. and we shall have

$$\int_E f d\mu = \sum_{n=0}^{\infty}\int_E f_n d\mu = \sum_{n=0}^{\infty}\nu(E\cap E_n) = \nu E$$

for every Borel set E, so that the indefinite-integral measure ν' defined by f and μ agrees with ν on the Borel sets. Since this ensures that ν' is locally finite, ν' is a Radon measure, by 256E, and is equal to ν, by 256D. Accordingly ν is an indefinite-integral measure over μ.

(b) As in (a-ii) above, h must be locally integrable and the indefinite-integral measure defined by h agrees with ν on the Borel sets, so is identical with ν.

256K Products The class of Radon measures on Euclidean spaces is stable under a wide variety of operations, as we have already seen; in particular, we have the following.

Theorem Let ν_1, ν_2 be Radon measures on $\mathbb{R}^r$ and $\mathbb{R}^s$ respectively, where r, $s \geq 1$. Let λ be their c.l.d. product measure on $\mathbb{R}^r\times\mathbb{R}^s$. Then λ is a Radon measure.

Remark When I say that λ is 'Radon' according to the definition in 256A, I am of course identifying $\mathbb{R}^r\times\mathbb{R}^s$ with $\mathbb{R}^{r+s}$, as in 251M-251N.

proof (a) I hope the following rather voluminous notation will seem natural. Write Σ_1, Σ_2 for the domains of ν_1, ν_2; $\mathcal{B}_r$, $\mathcal{B}_s$ for the Borel σ-algebras of $\mathbb{R}^r$, $\mathbb{R}^s$; Λ for the domain of λ; and $\mathcal{B}$ for the Borel σ-algebra of $\mathbb{R}^{r+s}$.

Because each ν_i is the completion of its restriction to the Borel sets (256C), λ is the product of $\nu_1{\restriction}\mathcal{B}_r$ and $\nu_2{\restriction}\mathcal{B}_s$ (251T). Because $\nu_1{\restriction}\mathcal{B}_r$ and $\nu_2{\restriction}\mathcal{B}_s$ are σ-finite (256Ba, 212Ga), λ must be the completion of its restriction to $\mathcal{B}_r\widehat{\otimes}\mathcal{B}_s$, which by 251M is identified with $\mathcal{B}$. Setting $Q_n = \{(x,y) : \|x\| \le n,\ \|y\| \le n\}$ we have

$$\lambda Q_n = \nu_1\{x : \|x\| \le n\} \cdot \nu_2\{y : \|y\| \le n\} < \infty$$

for every n, while every bounded subset of $\mathbb{R}^{r+s}$ is included in some Q_n. So $\lambda{\restriction}\mathcal{B}$ is locally finite, and its completion λ is a Radon measure, by 256C.

256L Remark We see from 253I that if ν_1 and ν_2 are Radon measures on $\mathbb{R}^r$ and $\mathbb{R}^s$ respectively, and both are indefinite-integral measures over Lebesgue measure, then their product measure on $\mathbb{R}^{r+s}$ is also an indefinite-integral measure over Lebesgue measure.

***256M** For the sake of applications in §286 below, I include another result, which is in fact one of the fundamental properties of Radon measures, as will appear in §414.

Proposition Let ν be a Radon measure on $\mathbb{R}^r$, and D any subset of $\mathbb{R}^r$. Let Φ be a non-empty upwards-directed family of non-negative continuous functions from D to $\mathbb{R}$. For $x \in D$ set $g(x) = \sup_{f\in\Phi} f(x)$ in $[0,\infty]$. Then

(a) $g : D \to [0,\infty]$ is lower semi-continuous, therefore Borel measurable;

(b) $\int_D g\,d\nu = \sup_{f\in\Phi} \int_D f d\nu$.

proof (a) For any $u \in [-\infty,\infty]$,

$$\{x : x \in D,\ g(x) > u\} = \bigcup_{f\in\Phi}\{x : x \in D,\ f(x) > u\}$$

is an open set for the subspace topology on D (2A3C), so is the intersection of D with a Borel subset of $\mathbb{R}^r$. This is enough to show that g is Borel measurable (121B-121C).

(b) Accordingly $\int_D g\,d\nu$ will be defined in $[0,\infty]$, and of course $\int_D g\,d\nu \ge \sup_{f\in\Phi}\int_D f d\nu$.

For the reverse inequality, observe that there is a countable set $\Psi \subseteq \Phi$ such that $g(x) = \sup_{f\in\Psi} f(x)$ for every $x \in D$. **P** For $a \in \mathbb{Q}$, q, $q' \in \mathbb{Q}^r$ set

$$\Phi_{aqq'} = \{f : f \in \Phi,\ f(y) > a \text{ whenever } y \in D \cap [q,q']\},$$

interpreting $[q,q']$ as in 115G. Choose $f_{aqq'} \in \Phi_{aqq'}$ if $\Phi_{aqq'}$ is not empty, and arbitrarily in Φ otherwise; and set $\Psi = \{f_{aqq'} : a \in \mathbb{Q},\ q,\ q' \in \mathbb{Q}^r\}$, so that Ψ is a countable subset of Φ. If $x \in D$ and $b < g(x)$, there is an $a \in \mathbb{Q}$ such that $b \le a < g(x)$; there is an $\hat{f} \in \Phi$ such that $\hat{f}(x) > a$; because $\hat{f}$ is continuous, there are q, $q' \in \mathbb{Q}^r$ such that $q \le x \le q'$ and $\hat{f}(y) \ge a$ whenever $y \in D \cap [q,q']$; so that $\hat{f} \in \Phi_{aqq'}$, $\Phi_{aqq'} \ne \emptyset$, $f_{aqq'} \in \Phi_{aqq'}$ and $\sup_{f\in\Psi} f(x) \ge f_{aqq'}(x) \ge b$. As b is arbitrary, $g(x) = \sup_{f\in\Psi} f(x)$. **Q**

Let $\langle f_n\rangle_{n\in\mathbb{N}}$ be a sequence running over Ψ. Because Φ is upwards-directed, we can choose $\langle f'_n\rangle_{n\in\mathbb{N}}$ in Φ inductively in such a way that $f'_{n+1} \ge \max(f'_n, f_n)$ for every $n \in \mathbb{N}$. So $\langle f'_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence in Φ and $\sup_{n\in\mathbb{N}} f'_n(x) \ge \sup_{f\in\Psi} f(x) = g(x)$ for every $x \in D$. By B.Levi's theorem,

$$\int_D g\,d\nu \le \sup_{n\in\mathbb{N}} \int_D f'_n d\nu \le \sup_{f\in\Phi}\int_D f d\nu,$$

and we have the required inequality.

256X Basic exercises >(a) Let ν be a measure on $\mathbb{R}^r$. (i) Show that it is locally finite, in the sense of 256Ab, iff for every $x \in \mathbb{R}^r$ there is a $\delta > 0$ such that $\nu^*B(x,\delta) < \infty$. (*Hint*: the sets $B(\mathbf{0},n)$ are compact.) (ii) Show that in this case ν is σ-finite.

>(b) Let ν be a Radon measure on $\mathbb{R}^r$ and $\mathcal{G}$ a non-empty upwards-directed family of open sets in $\mathbb{R}^r$. (i) Show that $\nu(\bigcup\mathcal{G}) = \sup_{G\in\mathcal{G}} \nu G$. (*Hint*: observe that if $K \subseteq \bigcup\mathcal{G}$ is compact, then $K \subseteq G$ for some $G \in \mathcal{G}$.) (ii) Show that $\nu(E \cap \bigcup\mathcal{G}) = \sup_{G\in\mathcal{G}} \nu(E \cap G)$ for every set E which is measured by ν.

>(c) Let ν be a Radon measure on $\mathbb{R}^r$ and $\mathcal{F}$ a non-empty downwards-directed family of closed sets in $\mathbb{R}^r$ such that $\inf_{F\in\mathcal{F}} \nu F < \infty$. (i) Show that $\nu(\bigcap\mathcal{F}) = \inf_{F\in\mathcal{F}} \nu F$. (*Hint*: apply 256Xb(ii) to $\mathcal{G} = \{\mathbb{R}^r \setminus F : F \in \mathcal{F}\}$.) (ii) Show that $\nu(E \cap \bigcap\mathcal{F}) = \inf_{F\in\mathcal{F}} \nu(E \cap F)$ for every E in the domain of ν.

>(d) Show that a Radon measure ν on $\mathbb{R}^r$ is atomless iff $\nu\{x\} = 0$ for every $x \in \mathbb{R}^r$. (*Hint*: apply 256Xc with $\mathcal{F} = \{F : F \subseteq E$ is closed, not negligible$\}$.)

(e) Let ν_1, ν_2 be Radon measures on $\mathbb{R}^r$, and α_1, $\alpha_2 \in \,]0,\infty[$. Set $\Sigma = \operatorname{dom}\nu_1 \cap \operatorname{dom}\nu_2$, and for $E \in \Sigma$ set $\nu E = \alpha_1\nu_1 E + \alpha_2\nu_2 E$. Show that ν is a Radon measure on $\mathbb{R}^r$. Show that ν is an indefinite-integral measure over Lebesgue measure iff ν_1, ν_2 are, and that in this case a linear combination of of Radon-Nikodým derivatives of ν_1 and ν_2 is a Radon-Nikodým derivative of ν.

>(f) Let ν be a Radon measure on $\mathbb{R}^r$. (i) Show that there is a unique closed set $F \subseteq \mathbb{R}^r$ such that, for open sets $G \subseteq \mathbb{R}^r$, $\nu G > 0$ iff $G \cap F \neq \emptyset$. (F is called the **support** of ν.) (ii) Generally, a set $A \subseteq \mathbb{R}^r$ is called **self-supporting** if $\nu^*(A \cap G) > 0$ whenever $G \subseteq \mathbb{R}^r$ is an open set meeting A. Show that for every closed set $F \subseteq \mathbb{R}^r$ there is a unique self-supporting closed set $F' \subseteq F$ such that $\nu(F \setminus F') = 0$.

>(g) Show that a measure ν on $\mathbb{R}$ is a Radon measure iff it is a Lebesgue-Stieltjes measure as described in 114Xa. Show that in this case ν is an indefinite-integral measure over Lebesgue measure iff the function $x \mapsto \nu\,]-\infty, x]$ is absolutely continuous on every bounded interval.

(h) Let ν be a Radon measure on $\mathbb{R}^r$. Let C_k be the space of continuous real-valued functions on $\mathbb{R}^r$ with bounded supports. Show that for every ν-integrable function f and every $\epsilon > 0$ there is a $g \in C_k$ such that $\int |f - g| d\nu \leq \epsilon$. (*Hint*: use arguments from 242O, but in (a-i) of the proof there start with *closed* intervals I.)

(i) Let ν be a Radon measure on $\mathbb{R}^r$. Show that $\nu E = \inf\{\nu G : G \supseteq E \text{ is open}\}$ for every set E in the domain of ν.

(j) Let ν, ν' be two Radon measures on $\mathbb{R}^r$, and suppose that $\nu I = \nu' I$ for every half-open interval $I \subseteq \mathbb{R}^r$ (definition: 115Ab). Show that $\nu = \nu'$.

(k) Let ν be Cantor measure (256Hc). (i) Show that if C_n is the nth set used in the construction of the Cantor set, so that C_n consists of 2^n intervals of length 3^{-n}, then $\nu I = 2^{-n}$ for each of the intervals I composing C_n. (ii) Let λ be the usual measure on $\{0,1\}^{\mathbb{N}}$ (254J). Define $\phi : \{0,1\}^{\mathbb{N}} \to \mathbb{R}$ by setting $\phi(x) = \frac{2}{3}\sum_{n=0}^{\infty} 3^{-n} x(n)$ for each $x \in \{0,1\}^{\mathbb{N}}$. Show that ϕ is a bijection between $\{0,1\}^{\mathbb{N}}$ and C. (iii) Show that if $\mathcal{B}$ is the Borel σ-algebra of $\mathbb{R}$, then $\{\phi^{-1}[E] : E \in \mathcal{B}\}$ is precisely the σ-algebra of subsets of $\{0,1\}^{\mathbb{N}}$ generated by the sets $\{x : x(n) = i\}$ for $n \in \mathbb{N}$, $i \in \{0,1\}$. (iv) Show that ϕ is an isomorphism between $(\{0,1\}^{\mathbb{N}}, \lambda)$ and (C, ν_C), where ν_C is the subspace measure on C induced by ν.

(l) Let ν and ν' be two Radon measures on $\mathbb{R}^r$. Show that ν' is an indefinite-integral measure over ν iff $\nu' E = 0$ whenever $\nu E = 0$, and in this case a function f is a Radon-Nikodým derivative of ν' with respect to ν iff $\int_E f d\nu = \nu' E$ for every Borel set E.

256Y Further exercises (a) Let ν be a Radon measure on $\mathbb{R}^r$, and X any subset of $\mathbb{R}^r$; let ν_X be the subspace measure on X and Σ_X its domain, and give X its subspace topology (2A3C). Show that ν_X has the following properties: (i) ν_X is complete and locally determined; (ii) every open subset of X belongs to Σ_X; (iii) $\nu_X E = \sup\{\nu_X F : F \subseteq E$ is closed in $X\}$ for every $E \in \Sigma_X$; (iv) whenever $\mathcal{G}$ is a non-empty upwards-directed family of open subsets of X, $\nu_X(\bigcup \mathcal{G}) = \sup_{G \in \mathcal{G}} \nu_X G$; (v) every point of X belongs to an open set of finite measure.

(b) Let ν be a Radon measure on $\mathbb{R}^r$, with domain Σ, and $f : \mathbb{R}^r \to \mathbb{R}$ a function. Show that the following are equiveridical: (i) f is Σ-measurable; (ii) for every non-negligible set $E \in \Sigma$ there is a non-negligible $F \in \Sigma$ such that $F \subseteq E$ and $f \restriction F$ is continuous; (iii) for every set $E \in \Sigma$, $\nu E = \sup_{K \in \mathcal{K}_f, K \subseteq E} \nu K$, where $\mathcal{K}_f = \{K : K \subseteq \mathbb{R}^r$ is compact, $f \restriction K$ is continuous$\}$. (*Hint*: for (ii)$\Rightarrow$(i), apply 215B(iv) to $CalK_f$.)

(c) Take ν, X, ν_X and Σ_X as in 256Ya. Suppose that $f : X \to \mathbb{R}$ is a function. Show that f is Σ_X-measurable iff for every non-negligible measurable set $E \subseteq X$ there is a non-negligible measurable $F \subseteq E$ such that $f \restriction F$ is continuous.

(d) Let $\langle \nu_n \rangle_{n \in \mathbb{N}}$ be a sequence of Radon measures on $\mathbb{R}^r$. Show that there is a Radon measure ν on $\mathbb{R}^r$ such that every ν_n is an indefinite-integral measure over ν. (*Hint*: find a sequence $\langle \alpha_n \rangle_{n \in \mathbb{N}}$ of strictly positive numbers such that $\sum_{n=0}^{\infty} \alpha_n \nu_n B(\mathbf{0}, k) < \infty$ for every k, and set $\nu = \sum_{n=0}^{\infty} \alpha_n \nu_n$, using the idea of 256Xe.)

(e) A set $G \subseteq \mathbb{R}^{\mathbb{N}}$ is **open** if for every $x \in G$ there are $n \in \mathbb{N}$, $\delta > 0$ such that

$$\{y : y \in \mathbb{R}^{\mathbb{N}},\ |y(i) - x(i)| < \delta \text{ for every } i \leq n\} \subseteq G.$$

The **Borel σ-algebra** of $\mathbb{R}^{\mathbb{N}}$ is the σ-algebra $\mathcal{B}$ of subsets of $\mathbb{R}^{\mathbb{N}}$ generated, in the sense of 111Gb, by the family $\mathfrak{T}$ of open sets. (i) Show that $\mathfrak{T}$ is a topology (2A3A). (ii) Show that a filter $\mathcal{F}$ on $\mathbb{R}^{\mathbb{N}}$ converges to $x \in \mathbb{R}^{\mathbb{N}}$ iff $\pi_i[[\mathcal{F}]] \to x(i)$ for every $i \in \mathbb{N}$, where $\pi_i(y) = y(i)$ for $i \in \mathbb{N}$, $y \in \mathbb{R}^{\mathbb{N}}$. (iii) Show that $\mathcal{B}$ is the σ-algebra generated by sets of the form $\{x : x \in \mathbb{R}^{\mathbb{N}}, x(i) \le a\}$, where i runs over $\mathbb{N}$ and a runs over $\mathbb{R}$. (iv) Show that if $\alpha_i \ge 0$ for every $i \in \mathbb{N}$, then $\{x : |x(i)| \le \alpha_i \ \forall\ i \in \mathbb{N}\}$ is compact. (*Hint*: 2A3R.) (v) Show that any open set in $\mathbb{R}^{\mathbb{N}}$ is the union of a sequence of closed sets. (*Hint*: look at sets of the form $\{x : q_i \le x(i) \le q'_i \ \forall\ i \le n\}$, where q_i, $q'_i \in \mathbb{Q}$ for $i \le n$.) (vi) Show that if ν_0 is any probability measure with domain $\mathcal{B}$, then its completion ν is inner regular with respect to the compact sets, and therefore may be called a 'Radon measure on $\mathbb{R}^{\mathbb{N}}$'. (*Hint*: show that there are compact sets of measure arbitrarily close to 1, and therefore that every open set, and every closed set, includes a K_σ set of the same measure.)

256 Notes and comments Radon measures on Euclidean spaces are very special, and the results of this section do not give clear pointers to the direction the theory takes when applied to other kinds of topological space. With the material here you could make a stab at developing a theory of Radon measures on complete separable metric spaces, provided you use 256Xa as the basis for your definition of 'locally finite'. These are the spaces for which a version of 256C is true. (See 256Ye.) But for generalizations to other types of topological space, and for the more interesting parts of the theory on $\mathbb{R}^r$, I must ask you to wait for Volume 4. My purpose in introducing Radon measures here is strictly limited; I wish only to give a basis for §257 and §271 sufficiently solid not to need later revision. In fact I think that all we really need are the Radon probability measures.

The chief technical difficulty in the definition of 'Radon measure' here lies in the insistence on completeness. It may well be that for everything studied in this volume, it would be simpler to look at locally finite measures with domain the algebra of Borel sets. This would involve us in a number of circumlocutions when dealing with Lebesgue measure itself and its derivates, since Lebesgue measure is defined on a larger σ-algebra; but the serious objection arises in the more advanced theory, when non-Borel sets of various kinds become central. Since my aim in this book is to provide secure foundations for the study of all aspects of measure theory, I ask you to take a little extra trouble now in order to avoid the possibility of having to re-work all your ideas later. The extra trouble arises, for instance, in 256D, 256Xe and 256Xj; since different Radon measures are defined on different σ-algebras, we have to check that two Radon measures which agree on the compact sets, or on the open sets, have the same domains. On the credit side, some of the power of 256G arises from the fact that the Radon image measure $\nu\phi^{-1}$ is defined on the whole σ-algebra $\{F : \phi^{-1}[F] \in \mathrm{dom}(\nu)\}$, not just on the Borel sets.

The further technical point that Radon measures are expected to be locally finite gives less difficulty; its effect is that from most points of view there is little difference between a general Radon measure and a totally finite Radon measure. The extra condition which obviously has to be put into the hypotheses of such results as 256E and 256G is no burden on either intuition or memory.

In effect, we have two definitions of Radon measures on Euclidean spaces: they are the inner regular locally finite topological measures, and they are also the completions of the locally finite Borel measures. The equivalence of these definitions is Theorem 256C. The latter definition is the better adapted to 256K, and the former to 256G. The 'inner regularity' of the basic definition refers to compact sets; we also have forms of inner regularity with respect to closed sets (256Bb) and K_σ sets (256Bc), and a complementary notion of 'outer regularity' with respect to open sets (256Xi).

257 Convolutions of measures

The ideas of this chapter can be brought together in a satisfying way in the theory of convolutions of Radon measures, which will be useful in §272 and again in §285. I give just the definition (257A) and the central property (257B) of the convolution of totally finite Radon measures, with a few corollaries and a note on the relation between convolution of functions and convolution of measures (257F).

257A Definition Let $r \ge 1$ be an integer and ν_1, ν_2 two totally finite Radon measures on $\mathbb{R}^r$. Let λ be the product measure on $\mathbb{R}^r \times \mathbb{R}^r$; then λ also is a (totally finite) Radon measure, by 256K. Define $\phi : \mathbb{R}^r \times \mathbb{R}^r \to \mathbb{R}^r$ by setting $\phi(x, y) = x + y$; then ϕ is continuous, therefore measurable in the sense of 256G. The **convolution** of ν_1 and ν_2, $\nu_1 * \nu_2$, is the image measure $\lambda\phi^{-1}$; by 256G, this is a Radon measure.

Note that if ν_1 and ν_2 are Radon probability measures, then λ and $\nu_1 * \nu_2$ are also probability measures.

257B Theorem Let $r \ge 1$ be an integer, and ν_1 and ν_2 two totally finite Radon measures on $\mathbb{R}^r$; let $\nu = \nu_1 * \nu_2$ be their convolution, and λ their product on $\mathbb{R}^r \times \mathbb{R}^r$. Then for any real-valued function h defined on a subset of $\mathbb{R}^r$,

$$\int h(x + y)\lambda(d(x, y)) \text{ exists } = \int h(x)\nu(dx)$$

if either integral is defined in $[-\infty, \infty]$.

proof Apply 235J with $J(x, y) = 1$, $\phi(x, y) = x + y$ for all x, $y \in \mathbb{R}^r$.

257C Corollary Let $r \geq 1$ be an integer, and ν_1, ν_2 two totally finite Radon measures on $\mathbb{R}^r$; let $\nu = \nu_1 * \nu_2$ be their convolution, and λ their product on $\mathbb{R}^r \times \mathbb{R}^r$; write Λ for the domain of λ. Let h be a Λ-measurable function defined λ-almost everywhere in $\mathbb{R}^r$. Suppose that any one of the integrals

$$\iint |h(x+y)|\nu_1(dx)\nu_2(dy), \quad \iint |h(x+y)|\nu_2(dy)\nu_1(dx), \quad \int h(x+y)\lambda(d(x,y))$$

exists and is finite. Then h is ν-integrable and

$$\int h(x)\nu(dx) = \iint h(x+y)\nu_1(dx)\nu_2(dy) = \iint h(x+y)\nu_2(dy)\nu_1(dx).$$

proof Put 257B together with Fubini's and Tonelli's theorems (252H).

257D Corollary If ν_1 and ν_2 are totally finite Radon measures on $\mathbb{R}^r$, then $\nu_1 * \nu_2 = \nu_2 * \nu_1$.

proof For any Borel set $E \subseteq \mathbb{R}^r$, apply 257C to $h = \chi E$ to see that

$$\begin{aligned}(\nu_1 * \nu_2)(E) &= \iint \chi E(x+y)\nu_1(dx)\nu_2(dy) = \iint \chi E(x+y)\nu_2(dy)\nu_1(dx) \\ &= \iint \chi E(y+x)\nu_2(dy)\nu_1(dx) = (\nu_2 * \nu_1)(E).\end{aligned}$$

Thus $\nu_1 * \nu_2$ and $\nu_2 * \nu_1$ agree on the Borel sets of $\mathbb{R}^r$; because they are both Radon measures, they must be identical (256D).

257E Corollary If ν_1, ν_2 and ν_3 are totally finite Radon measures on $\mathbb{R}^r$, then $(\nu_1 * \nu_2) * \nu_3 = \nu_1 * (\nu_2 * \nu_3)$.

proof For any Borel set $E \subseteq \mathbb{R}^r$, apply 257B to $h = \chi E$ to see that

$$\begin{aligned}((\nu_1 * \nu_2) * \nu_3)(E) &= \iint \chi E(x+z)(\nu_1 * \nu_2)(dx)\nu_3(dz) \\ &= \iiint \chi E(x+y+z)\nu_1(dx)\nu_2(dy)\nu_3(dz)\end{aligned}$$

(because $x \mapsto \chi E(x+z)$ is Borel measurable for every z)

$$= \iint \chi E(x+y)\nu_1(dx)(\nu_2 * \nu_3)(dy)$$

(because $(x,y) \mapsto \chi E(x+y)$ is Borel measurable, so $y \mapsto \int \chi E(x+y)\nu_1(dx)$ is $(\nu_2 * \nu_3)$-integrable)

$$= (\nu_1 * (\nu_2 * \nu_3))(E).$$

Thus $(\nu_1 * \nu_2) * \nu_3$ and $\nu_1 * (\nu_2 * \nu_3)$ agree on the Borel sets of $\mathbb{R}^r$; because they are both Radon measures, they must be identical.

257F Theorem Suppose that ν_1 and ν_2 are totally finite Radon measures on $\mathbb{R}^r$ which are indefinite-integral measures over Lebesgue measure μ. Then $\nu_1 * \nu_2$ also is an indefinite-integral measure over μ; if f_1 and f_2 are Radon-Nikodým derivatives of ν_1, ν_2 respectively, then $f_1 * f_2$ is a Radon-Nikodým derivative of $\nu_1 * \nu_2$.

proof By 255H/255L, $f_1 * f_2$ is integrable with respect to μ, with $\int f_1 * f_2 d\mu = 1$, and of course $f_1 * f_2$ is non-negative. If $E \subseteq \mathbb{R}^r$ is a Borel set,

$$\int_E f_1 * f_2 d\mu = \iint \chi E(x+y)f_1(x)f_2(y)\mu(dx)\mu(dy)$$

(255G)

$$= \iint \chi E(x+y)f_2(y)\nu_1(dx)\mu(dy)$$

(because $x \mapsto \chi E(x+y)$ is Borel measurable)

$$= \iint \chi E(x+y)\nu_1(dx)\nu_2(dy)$$

(because $(x,y) \mapsto \chi E(x+y)$ is Borel measurable, so $y \mapsto \int \chi E(x+y)\nu_1(dx)$ is ν_2-integrable)

$$= (\nu_1 * \nu_2)(E).$$

So $f_1 * f_2$ is a Radon-Nikodým derivative of ν with respect to μ, by 256J.

257X Basic exercises >(a) Let $r \geq 1$ be an integer. Let δ_0 be the Dirac measure on $\mathbb{R}^r$ concentrated at 0. Show that δ_0 is a Radon probability measure on $\mathbb{R}^r$ and that $\delta_0 * \nu = \nu$ for every totally finite Radon measure on $\mathbb{R}^r$.

(b) Let μ and ν be totally finite Radon measures on $\mathbb{R}^r$, and E any set measured by their convolution $\mu * \nu$. Show that $\int \mu(E - y)\nu(dy)$ is defined in $[0, \infty]$ and equal to $(\mu * \nu)(E)$.

(c) Let $\nu_1, \dots , \nu_n$ be totally finite Radon measures on $\mathbb{R}^r$, and let ν be the convolution $\nu_1 * \ldots * \nu_n$ (using 257E to see that such a bracketless expression is legitimate). Show that

$$\int h(x)\nu(dx) = \int \ldots \int h(x_1 + \ldots + x_n)\nu_1(dx_1) \ldots \nu_n(dx_n)$$

for every ν-integrable function h.

(d) Let ν_1 and ν_2 be totally finite Radon measures on $\mathbb{R}^r$, with supports F_1, F_2 (256Xf). Show that the support of $\nu_1 * \nu_2$ is $\overline{\{x + y : x \in F_1,\, y \in F_2\}}$.

>(e) Let ν_1 and ν_2 be totally finite Radon measures on $\mathbb{R}^r$, and suppose that ν_1 has a Radon-Nikodým derivative f with respect to Lebesgue measure μ. Show that $\nu_1 * \nu_2$ has a Radon-Nikodým derivative g, where $g(x) = \int f(x-y)\nu_2(dy)$ for μ-almost every $x \in \mathbb{R}^r$.

(f) Suppose that ν_1, ν_2, ν'_1 and ν'_2 are totally finite Radon measures on $\mathbb{R}^r$, and that ν'_1, ν'_2 are absolutely continuous with respect to ν_1, ν_2 respectively. Show that $\nu'_1 * \nu'_2$ is absolutely continuous with respect to $\nu_1 * \nu_2$.

257Y Further exercises (a) Let M be the space of countably additive functionals defined on the algebra $\mathcal{B}$ of Borel subsets of $\mathbb{R}$, with its norm $\|\nu\| = |\nu|(\mathbb{R})$ (see 231Yh). (i) Show that we have a unique bilinear operator $* : M \times M \to M$ such that $(\mu_1 \restriction \mathcal{B}) * (\mu_2 \restriction \mathcal{B}) = (\mu_1 * \mu_2) \restriction \mathcal{B}$ for all totally finite Radon measures μ_1, μ_2 on $\mathbb{R}$. (ii) Show that $*$ is commutative and associative. (iii) Show that $\|\nu_1 * \nu_2\| \leq \|\nu_1\|\|\nu_2\|$ for all ν_1, $\nu_2 \in M$, so that M is a Banach algebra under this multiplication. (iv) Show that M has a multiplicative identity. (v) Show that $L^1(\mu)$ can be regarded as a closed subalgebra of M, where μ is Lebesgue measure on $\mathbb{R}^r$ (cf. 255Xi).

(b) Let us say that a **Radon measure on** $]-\pi, \pi]$ is a measure ν, with domain Σ, on $]-\pi, \pi]$ such that (i) every Borel subset of $]-\pi, \pi]$ belongs to Σ (ii) for every $E \in \Sigma$ there are Borel sets E_1, E_2 such that $E_1 \subseteq E \subseteq E_2$ and $\nu(E_2 \setminus E_1) = 0$ (iii) every compact subset of $]-\pi, \pi]$ has finite measure. Show that for any two totally finite Radon measures ν_1, ν_2 on $]-\pi, \pi]$ there is a unique totally finite Radon measure ν on $]-\pi, \pi]$ such that

$$\int h(x)\nu(dx) = \int h(x +_{2\pi} y)\nu_1(dx)\nu_2(dy)$$

for every ν-integrable function h, where $+_{2\pi}$ is defined as in 255Ma.

257 Notes and comments Of course convolution of functions and convolution of measures are very closely connected; the obvious link being 257F, but the correspondence between 255G and 257B is also very marked. In effect, they give us the same notion of convolution $u * v$ when u, v are positive members of L^1 and $u * v$ is interpreted in L^1 rather than as a function (257Ya). But we should have to go rather deeper than the arguments here to find ideas in the theory of convolution of measures to correspond to such results as 255K. I will return to questions of this type in §444 in Volume 4.

All the theorems of this section can be extended to general abelian locally compact Hausdorff topological groups; but for such generality we need much more advanced ideas (see §444), and for the moment I leave only the suggestion in 257Yb that you should try to adapt the ideas here to $]-\pi, \pi]$ or S^1.

Chapter 26

Change of Variable in the Integral

I suppose most courses on basic calculus still devote a substantial amount of time to practice in the techniques of integrating standard functions. Surely the most powerful single technique is that of substitution: replacing $\int g(y)dy$ by $\int g(\phi(x))\phi'(x)dx$ for an appropriate function ϕ. At this level one usually concentrates on the skills of guessing at appropriate ϕ and getting the formulae right. I will not address such questions here, except for rare special cases; in this book I am concerned rather with validating the process. For functions of one variable, it can usually be justified by an appeal to the Fundamental Theorem of Calculus, and for any particular case I would normally go first to §225 in the hope that the results there would cover it. But for functions of two or more variables some much deeper ideas are necessary.

I have already treated the general problem of integration-by-substitution in abstract measure spaces in §235. There I described conditions under which $\int g(y)dy = \int g(\phi(x))J(x)dx$ for an appropriate function J. The context there gave very little scope for suggestions as to how to compute J; at best, it could be presented as a Radon-Nikodým derivative (235M). In this chapter I give a form of the fundamental theorem for the case of Lebesgue measure, in which ϕ is a more or less differentiable function between Euclidean spaces, and J is a 'Jacobian', the modulus of the determinant of the derivative of ϕ (263D). This necessarily depends on a serious investigation of the relationship between Lebesgue measure and geometry. The first step is to establish a form of Vitali's theorem for r-dimensional space, together with r-dimensional density theorems; I do this in §261, following closely the scheme of §§221 and 223 above. We need to know quite a lot about differentiable functions between Euclidean spaces, and it turns out that the theory is intertwined with that of 'Lipschitz' functions; I treat these in §262.

In the next two sections of the chapter, I turn to a separate problem for which some of the same techniques turn out to be appropriate: the description of surface measure on (smooth) surfaces in Euclidean space, like the surface of a cone or sphere. I suppose there is no difficulty in forming a robust intuition as to what is meant by the 'area' of such a surface and of suitably simple regions within it, and there is a very strong presumption that there ought to be an expression for this intuition in terms of measure theory as presented in this book; but the details are not I think straightforward. The first point to note is that for any calculation of the area of a region G in a surface S, one would always turn at once to a parametrization of the region, that is, a bijection $\phi : D \to G$ from some subset D of Euclidean space. But obviously one needs to be sure that the result of the calculation is independent of the parametrization chosen, and while it would be possible to base the theory on results showing such independence directly, that does not seem to me to be a true reflection of the underlying intuition, which is that the area of simple surfaces, at least, is something intrinsic to their geometry. I therefore see no acceptable alternative to a theory of 'r-dimensional measure' which can be described in purely geometric terms. This is the burden of §264, in which I give the definition and most fundamental properties of Hausdorff r-dimensional measure in Euclidean spaces. With this established, we find that the techniques of §§261-263 are sufficient to relate it to calculations through parametrizations, which is what I do in §265.

The chapter ends with a brief account of the Brunn-Minkowski inequality (266C), which is an essential tool for the geometric measure theory of convex sets.

261 Vitali's theorem in $\mathbb{R}^r$

The main aim of this section is to give r-dimensional versions of Vitali's theorem and Lebesgue's Density Theorem, following ideas already presented in §§221 and 223.

261A Notation For most of this chapter, we shall be dealing with the geometry and measure of Euclidean space; it will save space to fix some notation.

Throughout this section and the two following, $r \geq 1$ will be an integer. I will use Roman letters for members of $\mathbb{R}^r$ and Greek letters for their coordinates, so that $a = (\alpha_1, \dots , \alpha_r)$, etc.; if you see any Greek letter with a subscript you should look first for a nearby vector of which it might be a coordinate. The measure under consideration will nearly always be Lebesgue measure on $\mathbb{R}^r$; so unless otherwise indicated μ should be interpreted as Lebesgue measure, and μ^* as Lebesgue outer measure. Similarly, $\int \dots dx$ will always be integration with respect to Lebesgue measure (in a dimension determined by the context).

For $x = (\xi_1, \dots , \xi_r) \in \mathbb{R}^r$, write $\|x\| = \sqrt{\xi_1^2 + \dots + \xi_r^2}$. Recall that $\|x + y\| \leq \|x\| + \|y\|$ (1A2C) and that $\|\alpha x\| = |\alpha|\|x\|$ for any vectors x, y and scalar α.

I will use the same notation as in §115 for 'intervals', so that, in particular,

$$[a, b[= \{x : \alpha_i \leq \xi_i < \beta_i \ \forall\, i \leq r\},$$

$$]a, b[= \{x : \alpha_i < \xi_i < \beta_i \ \forall\ i \le r\},$$

$$[a, b] = \{x : \alpha_i \le \xi_i \le \beta_i \ \forall\ i \le r\}$$

whenever a, $b \in \mathbb{R}^r$.

$\mathbf{0} = (0, \dots, 0)$ will be the zero vector in $\mathbb{R}^r$, and $\mathbf{1}$ will be $(1, \dots, 1)$. If $x \in \mathbb{R}^r$ and $\delta > 0$, $B(x, \delta)$ will be the closed ball with centre x and radius δ, that is, $\{y : y \in \mathbb{R}^r, \|y - x\| \le \delta\}$. Note that $B(x, \delta) = x + B(\mathbf{0}, \delta)$; so that by the translation-invariance of Lebesgue measure we have

$$\mu B(x, \delta) = \mu B(\mathbf{0}, \delta) = \beta_r \delta^r,$$

where

$$\beta_r = \frac{1}{k!}\pi^k \text{ if } r = 2k \text{ is even,}$$

$$= \frac{2^{2k+1}k!}{(2k+1)!}\pi^k \text{ if } r = 2k+1 \text{ is odd}$$

(252Q).

261B Vitali's theorem in $\mathbb{R}^r$ Let $A \subseteq \mathbb{R}^r$ be any set, and $\mathcal{I}$ a family of closed non-trivial (that is, non-singleton, or, equivalently, non-negligible) balls in $\mathbb{R}^r$ such that every point of A is contained in arbitrarily small members of $\mathcal{I}$. Then there is a countable disjoint set $\mathcal{I}_0 \subseteq \mathcal{I}$ such that $\mu(A \setminus \bigcup \mathcal{I}_0) = 0$.

proof (a) To begin with (down to the end of (f) below), suppose that $\|x\| < M$ for every $x \in A$, and set

$$\mathcal{I}' = \{I : I \in \mathcal{I}, I \subseteq B(\mathbf{0}, M)\}.$$

If there is a finite disjoint set $\mathcal{I}_0 \subseteq \mathcal{I}'$ such that $A \subseteq \bigcup \mathcal{I}_0$ (including the possibility that $A = \mathcal{I}_0 = \emptyset$), we can stop. So let us suppose henceforth that there is no such $\mathcal{I}_0$.

(b) In this case, if $\mathcal{I}_0$ is any finite disjoint subset of $\mathcal{I}'$, there is a $J \in \mathcal{I}'$ which is disjoint from any member of $\mathcal{I}_0$. **P** Take $x \in A \setminus \bigcup \mathcal{I}_0$. Because every member of $\mathcal{I}_0$ is closed, there is a $\delta > 0$ such that $B(x, \delta)$ does not meet any member of $\mathcal{I}_0$, and as $\|x\| < M$ we can suppose that $B(x, \delta) \subseteq B(\mathbf{0}, M)$. Let J be a member of $\mathcal{I}$, containing x, and of diameter at most δ; then $J \in \mathcal{I}'$ and $J \cap \bigcup \mathcal{I}_0 = \emptyset$. **Q**

(c) We can therefore choose a sequence $\langle \gamma_n \rangle_{n \in \mathbb{N}}$ of real numbers and a disjoint sequence $\langle I_n \rangle_{n \in \mathbb{N}}$ in $\mathcal{I}'$ inductively, as follows. Given $\langle I_j \rangle_{j<n}$ (if $n = 0$, this is the empty sequence, with no members), with $I_j \in \mathcal{I}'$ for each $j < n$, and $I_j \cap I_k = \emptyset$ for $j < k < n$, set $\mathcal{J}_n = \{I : I \in \mathcal{I}', I \cap I_j = \emptyset \text{ for every } j < n\}$. We know from (b) that $\mathcal{J}_n \ne \emptyset$. Set

$$\gamma_n = \sup\{\operatorname{diam} I : I \in \mathcal{J}_n\};$$

then $\gamma_n \le 2M$, because every member of $\mathcal{J}_n$ is included in $B(\mathbf{0}, M)$. We can therefore find a set $I_n \in \mathcal{J}_n$ such that $\operatorname{diam} I_n \ge \frac{1}{2}\gamma_n$, and this continues the induction.

(e) Because the I_n are disjoint measurable subsets of the bounded set $B(\mathbf{0}, M)$, we have

$$\textstyle\sum_{n=0}^{\infty} \mu I_n \le \mu B(\mathbf{0}, M) < \infty,$$

and $\lim_{n\to\infty} \mu I_n = 0$. Also $\mu I_n \ge \beta_r(\frac{1}{4}\gamma_n)^r$ for each n, so $\lim_{n\to\infty} \gamma_n = 0$.

Now define I_n' to be the closed ball with the same centre as I_n but five times the diameter, so that it contains every point within a distance γ_n of I_n. I claim that, for any n, $A \subseteq \bigcup_{j<n} I_j \cup \bigcup_{j \ge n} I_j'$. **P?** Suppose, if possible, otherwise. Take any $x \in A \setminus (\bigcup_{j<n} I_j \cup \bigcup_{j\ge n} I_j')$. Let $\delta > 0$ be such that

$$B(x, \delta) \subseteq B(\mathbf{0}, M) \setminus \textstyle\bigcup_{j<n} I_j,$$

and let $J \in \mathcal{I}$ be such that $x \in J \subseteq B(x, \delta)$. Then

$$\lim_{m\to\infty} \gamma_m = 0 < \operatorname{diam} J$$

(this is where we use the hypothesis that all the balls in $\mathcal{I}$ are non-trivial); let m be the least integer greater than or equal to n such that $\gamma_m < \operatorname{diam} J$. In this case J cannot belong to $\mathcal{J}_m$, so there must be some $k < m$ such that $J \cap I_k \ne \emptyset$, because certainly $J \in \mathcal{I}'$. By the choice of δ, k cannot be less than n, so $n \le k < m$, and $\gamma_k \ge \operatorname{diam} J$. So the distance from x to the nearest point of I_k is at most $\operatorname{diam} J \le \gamma_k$. But this means that $x \in I_k'$; which contradicts the choice of x. **XQ**

(f) It follows that

$$\mu^*(A \setminus \bigcup_{j<n} I_j) \le \mu(\bigcup_{j\ge n} I'_j) \le \sum_{j=n}^{\infty} \mu I'_j \le 5^r \sum_{j=n}^{\infty} \mu I_j.$$

As

$$\sum_{j=0}^{\infty} \mu I_j \le \mu B(\mathbf{0}, M) < \infty,$$

$\lim_{n\to\infty} \mu^*(A \setminus \bigcup_{j<n} I_j) = 0$ and

$$\mu(A \setminus \bigcup_{j\in\mathbb{N}} I_j) = \mu^*(A \setminus \bigcup_{j\in\mathbb{N}} I_j) = 0.$$

Thus in this case we may set $\mathcal{I}_0 = \{I_n : n \in \mathbb{N}\}$ to obtain a countable disjoint family in $\mathcal{I}$ with $\mu(A \setminus \bigcup \mathcal{I}_0) = 0$.

(g) This completes the proof if A is bounded. In general, set

$$U_n = \{x : x \in \mathbb{R}^r,\ n < \|x\| < n+1\}, \quad A_n = A \cap U_n, \quad \mathcal{J}_n = \{I : I \in \mathcal{I},\ I \subseteq U_n\},$$

for each $n \in \mathbb{N}$. Then for each n we see that every point of A_n belongs to arbitrarily small members of $\mathcal{J}_n$, so there is a countable disjoint $\mathcal{J}'_n \subseteq \mathcal{J}_n$ such that $A_n \setminus \bigcup \mathcal{J}'_n$ is negligible. Now (because the U_n are disjoint) $\mathcal{I}_0 = \bigcup_{n\in\mathbb{N}} \mathcal{J}'_n$ is disjoint, and of course $\mathcal{I}_0$ is a countable subset of $\mathcal{I}$; moreover,

$$A \setminus \bigcup \mathcal{I}_0 \subseteq (\mathbb{R}^r \setminus \bigcup_{n\in\mathbb{N}} U_n) \cup \bigcup_{n\in\mathbb{N}} (A_n \setminus \bigcup \mathcal{J}'_n)$$

is negligible. (To see that $\mathbb{R}^r \setminus \bigcup_{n\in\mathbb{N}} U_n = \{x : \|x\| \in \mathbb{N}\}$ is negligible, note that for any $n \in \mathbb{N}$ the set

$$\{x : \|x\| = n\} \subseteq B(\mathbf{0}, n) \setminus B(\mathbf{0}, \delta n)$$

has measure at most $\beta_r n^r - \beta_r(\delta n)^r$ for every $\delta \in [0, 1[$, so must be negligible.)

261C Just as in §223, we can use the r-dimensional Vitali theorem to prove theorems on the approximation of functions by their local mean values.

Density Theorem in $\mathbb{R}^r$: integral form Let D be a subset of $\mathbb{R}^r$, and f a real-valued function which is integrable over D. Then

$$f(x) = \lim_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{D\cap B(x,\delta)} f d\mu$$

for almost every $x \in D$.

proof (a) To begin with (down to the end of (b)), let us suppose that $D = \operatorname{dom} f = \mathbb{R}^r$.

Take $n \in \mathbb{N}$ and q, $q' \in \mathbb{Q}$ with $q < q'$, and set

$$A = A_{nqq'} = \{x : \|x\| \le n,\ f(x) \le q,\ \limsup_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{B(x,\delta)} f d\mu > q'\}.$$

? Suppose, if possible, that $\mu^* A > 0$. Let $\epsilon > 0$ be such that $\epsilon(1 + |q|) < (q' - q)\mu^* A$, and let $\eta \in]0, \epsilon]$ be such that $\int_E |f| \le \epsilon$ whenever $\mu E \le \eta$ (225A). Let $G \supseteq A$ be an open set of measure at most $\mu^* A + \eta$ (134Fa). Let $\mathcal{I}$ be the set of non-trivial closed balls $B \subseteq G$ such that $\frac{1}{\mu B} \int_B f d\mu \ge q'$. Then every point of A is contained in (indeed, is the centre of) arbitrarily small members of $\mathcal{I}$. So there is a countable disjoint set $\mathcal{I}_0 \subseteq \mathcal{I}$ such that $\mu(A \setminus \bigcup \mathcal{I}_0) = 0$, by 261B; set $H = \bigcup \mathcal{I}_0$.

Because $\int_I f d\mu \ge q' \mu I$ for each $I \in \mathcal{I}_0$, we have

$$\int_H f d\mu = \sum_{I\in\mathcal{I}_0} \int_I f d\mu \ge q' \sum_{I\in\mathcal{I}_0} \mu I = q' \mu H \ge q' \mu^* A.$$

Set

$$E = \{x : x \in G,\ f(x) \le q\}.$$

Then E is measurable, and $A \subseteq E \subseteq G$; so

$$\mu^* A \le \mu E \le \mu G \le \mu^* A + \eta \le \mu^* A + \epsilon.$$

Also

$$\mu(H \setminus E) \le \mu G - \mu E \le \eta,$$

so by the choice of η, $\int_{H\setminus E} f \le \epsilon$ and

$$\begin{aligned}\int_H f &\le \epsilon + \int_{H\cap E} f \le \epsilon + q\mu(H \cap E)\\ &\le \epsilon + q\mu^* A + |q|(\mu(H \cap E) - \mu^* A) \le q\mu^* A + \epsilon(1 + |q|)\end{aligned}$$

(because $\mu^* A = \mu^*(A \cap H) \le \mu(H \cap E)$)

$$< q'\mu^* A \le \int_H f,$$

which is impossible. **X**

Thus $A_{nqq'}$ is negligible. This is true for all $q < q'$ and all n, so

$$A^* = \bigcup_{q,q' \in \mathbb{Q}, q<q'} \bigcup_{n\in\mathbb{N}} A_{nqq'}$$

is negligible. But

$$f(x) \ge \limsup_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{B(x,\delta)} f$$

for every $x \in \mathbb{R}^r \setminus A^*$, that is, for almost all $x \in \mathbb{R}^r$.

(b) Similarly, or applying this result to $-f$.

$$f(x) \le \liminf_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{B(x,\delta)} f$$

for almost every x, so

$$f(x) = \lim_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{B(x,\delta)} f$$

for almost every x.

(c) For the (superficially) more general case enunciated in the theorem, let $\tilde{f}$ be a μ-integrable function extending $f\restriction D$, defined everywhere on $\mathbb{R}^r$, and such that $\int_F \tilde{f} = \int_{D\cap F} f$ for every measurable $F \subseteq \mathbb{R}^r$ (applying 214Eb to $f\restriction D$). Then

$$f(x) = \tilde{f}(x) = \lim_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{B(x,\delta)} \tilde{f} = \lim_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{D\cap B(x,\delta)} f$$

for almost every $x \in D$.

261D Corollary (a) If $D \subseteq \mathbb{R}^r$ is any set, then

$$\lim_{\delta\downarrow 0} \frac{\mu^*(D\cap B(x,\delta))}{\mu B(x,\delta)} = 1$$

for almost every $x \in D$.

(b) If $E \subseteq \mathbb{R}^r$ is a measurable set, then

$$\lim_{\delta\downarrow 0} \frac{\mu(E\cap B(x,\delta))}{\mu B(x,\delta)} = \chi E(x)$$

for almost every $x \in \mathbb{R}^r$.

(c) If $D \subseteq \mathbb{R}^r$ and $f : D \to \mathbb{R}$ is any function, then for almost every $x \in D$,

$$\lim_{\delta\downarrow 0} \frac{\mu^*(\{y : y\in D, |f(y)-f(x)| \le \epsilon\}\cap B(x,\delta))}{\mu B(x,\delta)} = 1$$

for every $\epsilon > 0$.

(d) If $D \subseteq \mathbb{R}^r$ and $f : D \to \mathbb{R}$ is measurable, then for almost every $x \in D$,

$$\lim_{\delta\downarrow 0} \frac{\mu^*(\{y : y\in D, |f(y)-f(x)| \ge \epsilon\}\cap B(x,\delta))}{\mu B(x,\delta)} = 0$$

for every $\epsilon > 0$.

proof (a) Apply 261C with $f = \chi B(\mathbf{0}, n)$ to see that, for any $n \in \mathbb{N}$,

$$\lim_{\delta\downarrow 0} \frac{\mu^*(D\cap B(x,\delta))}{\mu B(x,\delta)} = 1$$

for almost every $x \in D$ with $\|x\| < n$.

(b) Apply (a) to E to see that

$$\liminf_{\delta\downarrow 0} \frac{\mu(E\cap B(x,\delta))}{\mu B(x,\delta)} \ge \chi E(x)$$

for almost every $x \in \mathbb{R}^r$, and to $E' = \mathbb{R}^r \setminus E$ to see that

$$\limsup_{\delta\downarrow 0}\frac{\mu(E\cap B(x,\delta))}{\mu B(x,\delta)}=1-\liminf_{\delta\downarrow 0}\frac{\mu(E'\cap B(x,\delta))}{\mu B(x,\delta)}\le 1-\chi E'(x)=\chi E(x)$$

for almost every x.

(c) For q, $q'\in\mathbb{Q}$, set

$$D_{qq'}=\{x:x\in D,\,q\le f(x)\le q'\},$$

$$C_{qq'}=\{x:x\in D_{qq'},\,\lim_{\delta\downarrow 0}\frac{\mu^*(D_{qq'}\cap B(x,\delta))}{\mu B(x,\delta)}=1\};$$

now set

$$C=D\setminus\bigcup_{q,q'\in\mathbb{Q}}(D_{qq'}\setminus C_{qq'}),$$

so that $D\setminus C$ is negligible. If $x\in C$ and $\epsilon>0$, then there are q, $q'\in\mathbb{Q}$ such that $f(x)-\epsilon\le q\le f(x)\le q'\le f(x)+\epsilon$, and now

$$\liminf_{\delta\downarrow 0}\frac{\mu^*\{y:y\in D\cap B(x,\delta),\,|f(y)-f(x)|\le\epsilon\}}{\mu B(x,\delta)}\ge\liminf_{\delta\downarrow 0}\frac{\mu^*(D_{qq'}\cap B(x,\delta))}{\mu B(x,\delta)}=1,$$

so

$$\lim_{\delta\downarrow 0}\frac{\mu^*\{y:y\in D\cap B(x,\delta),\,|f(y)-f(x)|\le\epsilon\}}{\mu B(x,\delta)}=1.$$

(d) Define C as in (c). We know from (a) that $\mu(D\setminus C')=0$, where

$$C'=\{x:x\in D,\lim_{\delta\downarrow 0}\frac{\mu^*(D\cap B(x,\delta))}{\mu B(x,\delta)}=1\}.$$

If $x\in C\cap C'$ and $\epsilon>0$, we know from (c) that

$$\lim_{\delta\downarrow 0}\frac{\mu^*\{y:y\in D\cap B(x,\delta),\,|f(y)-f(x)|\le\epsilon/2\}}{\mu B(x,\delta)}=1.$$

But because f is measurable, we have

$$\mu^*\{y:y\in D\cap B(x,\delta),\,|f(y)-f(x)|\ge\epsilon\}$$
$$+\mu^*\{y:y\in D\cap B(x,\delta),\,|f(y)-f(x)|\le\frac{1}{2}\epsilon\}\le\mu^*(D\cap B(x,\delta))$$

for every $\delta>0$. Accordingly

$$\limsup_{\delta\downarrow 0}\frac{\mu^*\{y:y\in D\cap B(x,\delta),\,|f(y)-f(x)|\ge\epsilon\}}{\mu B(x,\delta)}$$
$$\le\lim_{\delta\downarrow 0}\frac{\mu^*(D\cap B(x,\delta))}{\mu B(x,\delta)}-\lim_{\delta\downarrow 0}\frac{\mu^*\{y:y\in D\cap B(x,\delta),\,|f(y)-f(x)|\le\epsilon/2\}}{\mu B(x,\delta)}=0,$$

and

$$\lim_{\delta\downarrow 0}\frac{\mu^*\{y:y\in D\cap B(x,\delta),\,|f(y)-f(x)|\ge\epsilon\}}{\mu B(x,\delta)}=0$$

for every $x\in C\cap C'$, that is, for almost every $x\in D$.

261E Theorem Let f be a locally integrable function defined on a conegligible subset of $\mathbb{R}^r$. Then

$$\lim_{\delta\downarrow 0}\frac{1}{\mu B(x,\delta)}\int_{B(x,\delta)}|f(y)-f(x)|dy=0$$

for almost every $x\in\mathbb{R}^r$.

proof (Compare 223D.)

(a) Fix $n\in\mathbb{N}$ for the moment, and set $G=\{x:\|x\|<n\}$. For each $q\in\mathbb{Q}$, set $g_q(x)=|f(x)-q|$ for $x\in G\cap\operatorname{dom}f$; then g_q is integrable over G, and

$$\lim_{\delta\downarrow 0}\frac{1}{\mu B(x,\delta)}\int_{G\cap B(x,\delta)}g_q=g_q(x)$$

for almost every $x\in G$, by 261C. Setting

$$E_q = \{x : x \in G \cap \operatorname{dom} f,\ \lim_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{G\cap B(x,\delta)} g_q = g_q(x)\},$$

we have $G \setminus E_q$ negligible for every q, so $G \setminus E$ is negligible, where $E = \bigcap_{q\in\mathbb{Q}} E_q$. Now

$$\lim_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{G\cap B(x,\delta)} |f(y) - f(x)| dy = 0$$

for every $x \in E$. **P** Take $x \in E$ and $\epsilon > 0$. Then there is a $q \in \mathbb{Q}$ such that $|f(x) - q| \le \epsilon$, so that

$$|f(y) - f(x)| \le |f(y) - q| + \epsilon = g_q(y) + \epsilon$$

for every $y \in G \cap \operatorname{dom} f$, and

$$\begin{aligned}\limsup_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{G\cap B(x,\delta)} |f(y) - f(x)| dy &\le \limsup_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{G\cap B(x,\delta)} g_q(y) + \epsilon\, dy \\ &= \epsilon + g_q(x) \le 2\epsilon.\end{aligned}$$

As ϵ is arbitrary,

$$\lim_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{G\cap B(x,\delta)} |f(y) - f(x)| dy = 0,$$

as required. **Q**

(b) Because G is open,

$$\lim_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{B(x,\delta)} |f(y) - f(x)| dy = \lim_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{G\cap B(x,\delta)} |f(y) - f(x)| dy = 0$$

for almost every $x \in G$. As n is arbitrary,

$$\lim_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{B(x,\delta)} |f(y) - f(x)| dy = 0$$

for almost every $x \in \mathbb{R}^r$.

Remark The set

$$\{x : x \in \operatorname{dom} f,\ \lim_{\delta\downarrow 0} \frac{1}{\mu B(x,\delta)} \int_{B(x,\delta)} |f(y) - f(x)| dy = 0\}$$

is sometimes called the **Lebesgue set** of f.

261F Another very useful consequence of 261B is the following.

Proposition Let $A \subseteq \mathbb{R}^r$ be any set, and $\epsilon > 0$. Then there is a sequence $\langle B_n\rangle_{n\in\mathbb{N}}$ of closed balls in $\mathbb{R}^r$, all of radius at most ϵ, such that $A \subseteq \bigcup_{n\in\mathbb{N}} B_n$ and $\sum_{n=0}^{\infty} \mu B_n \le \mu^* A + \epsilon$. Moreover, we may suppose that the balls in the sequence whose centres do not lie in A have measures summing to at most ϵ.

proof (a) Set $\beta_r = \mu B(\mathbf{0}, 1)$. The first step is the obvious remark that if $x \in \mathbb{R}^r$, $\delta > 0$ then the half-open cube $I = [x, x + \delta\mathbf{1}[$ is a subset of the ball $B(x, \delta\sqrt{r})$, which has measure $\gamma_r \delta^r = \gamma_r \mu I$, where $\gamma_r = \beta_r r^{r/2}$. It follows that if $G \subseteq \mathbb{R}^r$ is any open set, then G can be covered by a sequence of balls of total measure at most $\gamma_r \mu G$. **P** If G is empty, we can take all the balls to be singletons. Otherwise, for each $k \in \mathbb{N}$, set

$$Q_k = \{z : z \in \mathbb{Z}^r,\ \bigl[2^{-k} z, 2^{-k}(z + \mathbf{1})\bigr[\subseteq G\},$$

$$E_k = \bigcup_{z\in Q_k} \bigl[2^{-k} z, 2^{-k}(z + \mathbf{1}\bigr[).$$

Then $\langle E_k\rangle_{k\in\mathbb{N}}$ is a non-decreasing sequence of sets with union G, and E_0 and each of the differences $E_{k+1} \setminus E_k$ is expressible as a disjoint union of half-open cubes. Thus G also is expressible as a disjoint union of a sequence $\langle I_n\rangle_{n\in\mathbb{N}}$ of half-open cubes. Each I_n is covered by a ball B_n of measure $\gamma_r \mu I_n$; so that $G \subseteq \bigcup_{n\in\mathbb{N}} B_n$ and

$$\sum_{n=0}^{\infty} \mu B_n \le \gamma_r \sum_{n=0}^{\infty} \mu I_n = \gamma_r \mu G. \textbf{ Q}$$

(b) It follows at once that if $\mu A = 0$ then for any $\epsilon > 0$ there is a sequence $\langle B_n\rangle_{n\in\mathbb{N}}$ of balls covering A of measures summing to at most ϵ, because there is certainly an open set including A with measure at most ϵ/γ_r.

(c) Now take any set A, and $\epsilon > 0$. Let $G \supseteq A$ be an open set with $\mu G \le \mu^* A + \frac{1}{2}\epsilon$. Let $\mathcal{I}$ be the family of non-trivial closed balls included in G, of radius at most ϵ and with centres in A. Then every point of A belongs to arbitrarily small members of $\mathcal{I}$, so there is a countable disjoint $\mathcal{I}_0 \subseteq \mathcal{I}$ such that $\mu(A \setminus \bigcup \mathcal{I}_0) = 0$. Let $\langle B'_n\rangle_{n\in\mathbb{N}}$ be a sequence of balls covering $A\setminus\bigcup\mathcal{I}_0$ with $\sum_{n=0}^{\infty}\mu B'_n\le\min(\frac{1}{2}\epsilon,\beta_r\epsilon^r)$; these surely all have radius at most ϵ. Let $\langle B_n\rangle_{n\in\mathbb{N}}$ be a sequence amalgamating $\mathcal{I}_0$ with $\langle B'_n\rangle_{n\in\mathbb{N}}$; then $A\subseteq\bigcup_{n\in\mathbb{N}}B_n$, every B_n has radius at most ϵ and

$$\sum_{n=0}^{\infty}\mu B_n=\sum_{B\in\mathcal{I}_0}\mu B+\sum_{n=0}^{\infty}\mu B'_n\le\mu G+\frac{1}{2}\epsilon\le\mu A+\epsilon,$$

while the B_n whose centres do not lie in A must come from the sequence $\langle B'_n\rangle_{n\in\mathbb{N}}$, so their measures sum to at most $\frac{1}{2}\epsilon\le\epsilon$.

Remark In fact we can (if A is not empty) arrange that the centre of *every* B_n belongs to A. This is an easy consequence of Besicovitch's Covering Lemma (see §472 in Volume 4).

261X Basic exercises (a) Show that 261C is valid for any locally integrable real-valued function f; in particular, for any $f\in\mathcal{L}^p(\mu_D)$ for any $p\ge 1$, writing μ_D for the subspace measure on D.

(b) Show that 261C, 261Dc, 261Dd and 261E are valid for complex-valued functions f.

>(c) Take three disks in the plane, each touching the other two, so that they enclose an open region R with three cusps. In R let D be a disk tangent to each of the three original disks, and R_0, R_1, R_2 the three components of $R\setminus D$. In each R_j let D_j be a disk tangent to each of the disks bounding R_j, and R_{j0}, R_{j1}, R_{j2} the three components of $R_j\setminus D_j$. Continue, obtaining 27 regions at the next step, 81 regions at the next, and so on.

Show that the total area of the residual regions converges to zero as the process continues indefinitely. (*Hint*: compare with the process in the proof of 261B.)

261Y Further exercises (a) Formulate an abstract definition of 'Vitali cover', meaning a family of sets satisfying the conclusion of 261B in some sense, and corresponding generalizations of 261C-261E, covering (at least) (b)-(d) below.

(b) For $x\in\mathbb{R}^r$, $k\in\mathbb{N}$ let $C(x,k)$ be the half-open cube of the form $\left[2^{-k}z,2^{-k}(z+\mathbf{1})\right[$, with $z\in\mathbb{Z}^r$, containing x. Show that if f is an integrable function on $\mathbb{R}^r$ then

$$\lim_{k\to\infty}2^{kr}\int_{C(x,k)}f=f(x)$$

for almost every $x\in\mathbb{R}^r$.

(c) Let f be a real-valued function which is integrable over $\mathbb{R}^r$. Show that

$$\lim_{\delta\downarrow 0}\frac{1}{\delta^r}\int_{[x,x+\delta\mathbf{1}[}f=f(x)$$

for almost every $x\in\mathbb{R}^r$.

(d) Give $X=\{0,1\}^{\mathbb{N}}$ its usual measure ν (254J). For $x\in X$, $k\in\mathbb{N}$ set $C(x,k)=\{y:y\in X,\ y(i)=x(i)$ for $i<k\}$. Show that if f is any real-valued function which is integrable over X then $\lim_{k\to\infty}2^k\int_{C(x,k)}f d\nu=f(x)$, $\lim_{k\to\infty}2^k\int_{C(x,k)}|f(y)-f(x)|\nu(dy)=0$ for almost every $x\in X$.

(e) Let f be a real-valued function which is integrable over $\mathbb{R}^r$, and x a point in the Lebesgue set of f. Show that for every $\epsilon>0$ there is a $\delta>0$ such that $|f(x)-\int f(x-y)g(\|y\|)dy|\le\epsilon$ whenever $g:[0,\infty[\,\to[0,\infty[$ is a non-increasing function such that $\int_{\mathbb{R}^r}g(\|y\|)dy=1$ and $\int_{B(\mathbf{0},\delta)}g(\|y\|)dy\ge 1-\delta$. (*Hint*: 223Yg.)

(f) Let $\mathfrak{T}$ be the family of those measurable sets $G\subseteq\mathbb{R}^r$ such that $\lim_{\delta\downarrow 0}\frac{\mu(G\cap B(x,\delta))}{\mu B(x,\delta)}=1$ for every $x\in G$. Show that $\mathfrak{T}$ is a topology on $\mathbb{R}^r$, the **density topology** of $\mathbb{R}^r$. Show that a function $f:\mathbb{R}^r\to\mathbb{R}$ is measurable iff it is $\mathfrak{T}$-continuous at almost every point of $\mathbb{R}^r$.

(g) A set $A\subseteq\mathbb{R}^r$ is said to be **porous** at $x\in\mathbb{R}^r$ if $\limsup_{y\to x}\frac{\rho(y,A)}{\|y-x\|}>0$, writing $\rho(y,A)=\inf_{z\in A}\|y-z\|$ (or ∞ if A is empty). (i) Show that if A is porous at all its points then it is negligible. (ii) Show that in the construction of 261A the residual set $A\setminus\bigcup\mathcal{I}_0$ is always porous.

(h) Let $A\subseteq\mathbb{R}^r$ be a bounded set and $\mathcal{I}$ a non-empty family of non-trivial closed balls covering A. Show that for any $\epsilon>0$ there are disjoint $B_0,\dots,B_n\in\mathcal{I}$ such that $\mu^*A\le(3+\epsilon)^r\sum_{k=0}^{n}\mu B_k$.

(i) Let (X,ρ) be a metric space and $A \subseteq X$ any set, $x \mapsto \delta_x : A \to [0,\infty[$ any bounded function. Show that if $\gamma > 3$ then there is an $A' \subseteq A$ such that (i) $\rho(x,y) > \delta_x + \delta_y$ for all distinct x, $y \in A'$ (ii) $\bigcup_{x\in A} B(x,\delta_x) \subseteq \bigcup_{x\in A'} B(x,\gamma\delta_x)$, writing $B(x,\alpha)$ for the closed ball $\{y : \rho(y,x) \le \alpha\}$.

(j) Show that any union of non-trivial closed balls in $\mathbb{R}^r$ is Lebesgue measurable. (*Hint*: induce on r. Compare 415Ye in Volume 4.)

(k) Suppose that $A \subseteq \mathbb{R}^r$ and that $\mathcal{I}$ is a family of closed subsets of $\mathbb{R}^r$ such that

for every $x \in A$ there is an $\eta > 0$ such that for every $\epsilon > 0$ there is an $I \in \mathcal{I}$ such that $x \in I$ and $0 < \eta(\operatorname{diam} I)^r \le \mu I \le \epsilon$.

Show that there is a countable disjoint set $\mathcal{I}_0 \subseteq \mathcal{I}$ such that $A \setminus \bigcup \mathcal{I}_0$ is negligible.

(l) Let $\mathfrak{T}'$ be the family of measurable sets $G \subseteq \mathbb{R}^r$ such that whenever $x \in G$ and $\epsilon > 0$ there is a $\delta > 0$ such that $\mu(G \cap I) \ge (1-\epsilon)\mu I$ whenever I is an interval containing x and included in $B(x,\delta)$. Show that $\mathfrak{T}'$ is a topology on $\mathbb{R}^r$ intermediate between the density topology (261Yf) and the Euclidean topology.

261 Notes and comments In the proofs of 261B-261E above, I have done my best to follow the lines of the one-dimensional case; this section amounts to a series of generalizations of the work of §§221 and 223.

It will be clear that the idea of 261A/261B can be used on other shapes than balls. To make it work in the form above, we need a family $\mathcal{I}$ such that there is a constant K for which

$$\mu I' \le K\mu I$$

for every $I \in \mathcal{I}$, where we write

$$I' = \{x : \inf_{y\in I} \|x-y\| \le \operatorname{diam}(I)\}.$$

Evidently this will be true for many classes $\mathcal{I}$ determined by the shapes of the sets involved; for instance, if $E \subseteq \mathbb{R}^r$ is any bounded set of strictly positive measure, the family $\mathcal{I} = \{x + \delta E : x \in \mathbb{R}^r,\ \delta > 0\}$ will satisfy the condition.

In 261Ya I challenge you to find an appropriate generalization of the arguments depending on the conclusion of 261B.

Another way of using 261B is to say that because sets can be essentially covered by *disjoint* sequences of balls, it ought to be possible to use balls, rather than half-open intervals, in the definition of Lebesgue measure on $\mathbb{R}^r$. This is indeed so (261F). The difficulty in using balls in the basic definition comes right at the start, in proving that if a ball is covered by finitely many balls then the sum of the volumes of the covering balls is at least the volume of the covered ball. (There is a trick, using the compactness of closed balls and the openness of open balls, to extend such a proof to infinite covers.) Of course you could regard this fact as 'elementary', on the ground that Archimedes would have noticed if it weren't true, but nevertheless it would be something of a challenge to prove it, unless you were willing to wait for a version of Fubini's theorem, as some authors do.

I have given the results in 261C-261D for arbitrary subsets D of $\mathbb{R}^r$ not because I have any applications in mind in which non-measurable subsets are significant, but because I wish to make it possible to notice when measurability matters. Of course it is necessary to interpret the integrals $\int_D f d\mu$ in the way laid down in §214. The game is given away in part (c) of the proof of 261C, where I rely on the fact that if f is integrable over D then there is an integrable $\tilde f : \mathbb{R}^r \to \mathbb{R}$ such that $\int_F \tilde f = \int_{D\cap F} f$ for every measurable $F \subseteq \mathbb{R}^r$. In effect, for all the questions dealt with here, we can replace f, D by $\tilde f$, $\mathbb{R}^r$.

The idea of 261C is that, for almost every x, $f(x)$ is approximated by its mean value on small balls $B(x,\delta)$, ignoring the missing values on $B(x,\delta) \setminus (D \cap \operatorname{dom} f)$; 261E is a sharper version of the same idea. The formulae of 261C-261E mostly involve the expression $\mu B(x,\delta)$. Of course this is just $\beta_r\delta^r$. But I think that leaving it unexpanded is actually more illuminating, as well as avoiding sub- and superscripts, since it makes it clearer what these density theorems are really about. In §472 of Volume 4 I will revisit this material, showing that a surprisingly large proportion of the ideas can be applied to arbitrary Radon measures on $\mathbb{R}^r$, even though Vitali's theorem (in the form stated here) is no longer valid.

262 Lipschitz and differentiable functions

In preparation for the main work of this chapter in §263, I devote a section to two important classes of functions between Euclidean spaces. What we really need is the essentially elementary material down to 262I, together with the technical lemma 262M and its corollaries. Theorem 262Q is not relied on in this volume, though I believe that it makes the patterns which will develop more natural and comprehensible.

262A Lipschitz functions Suppose that r, $s \geq 1$ and $\phi : D \to \mathbb{R}^s$ is a function, where $D \subseteq \mathbb{R}^r$. We say that ϕ is **γ-Lipschitz**, where $\gamma \in [0, \infty[$, if

$$\|\phi(x) - \phi(y)\| \leq \gamma\|x - y\|$$

for all x, $y \in D$, writing $\|x\| = \sqrt{\xi_1^2 + \ldots + \xi_r^2}$ if $x = (\xi_1, \ldots, \xi_r) \in \mathbb{R}^r$, $\|z\| = \sqrt{\zeta_1^2 + \ldots + \zeta_s^2}$ if $z = (\zeta_1, \ldots, \zeta_s) \in \mathbb{R}^s$. In this case, γ is **a Lipschitz constant for** ϕ.

A **Lipschitz function** is a function ϕ which is γ-Lipschitz for some $\gamma \geq 0$. Note that in this case ϕ has a least Lipschitz constant (since if A is the set of Lipschitz constants for ϕ, and $\gamma_0 = \inf A$, then γ_0 is a Lipschitz constant for ϕ).

262B We need the following easy facts.

Lemma Let $D \subseteq \mathbb{R}^r$ be a set and $\phi : D \to \mathbb{R}^s$ a function.

(a) ϕ is Lipschitz iff $\phi_i : D \to \mathbb{R}$ is Lipschitz for every i, writing $\phi(x) = (\phi_1(x), \ldots, \phi_s(x))$ for every $x \in D = \operatorname{dom}\phi \subseteq \mathbb{R}^r$.

(b) In this case, there is a Lipschitz function $\tilde{\phi} : \mathbb{R}^r \to \mathbb{R}^s$ extending ϕ.

(c) If $r = s = 1$ and $D = [a, b]$ is an interval, then ϕ is Lipschitz iff it is absolutely continuous and has a bounded derivative.

proof (a) For any x, $y \in D$ and $i \leq s$,

$$|\phi_i(x) - \phi_i(y)| \leq \|\phi(x) - \phi(y)\| \leq \sqrt{s}\sup_{j \leq s}|\phi_j(x) - \phi_j(y)|,$$

so any Lipschitz constant for ϕ will be a Lipschitz constant for every ϕ_i, and if γ_j is a Lipschitz constant for ϕ_j for each j, then $\sqrt{s}\sup_{j \leq s}\gamma_j$ will be a Lipschitz constant for ϕ.

(b) By (a), it is enough to consider the case $s = 1$, for if every ϕ_i has a Lipschitz extension $\tilde{\phi}_i$, we can set $\tilde{\phi}(x) = (\tilde{\phi}_1(x), \ldots, \tilde{\phi}_s(x))$ for every x to obtain a Lipschitz extension of ϕ. Taking $s = 1$, then, note that the case $D = \emptyset$ is trivial; so suppose that $D \neq \emptyset$. Let γ be a Lipschitz constant for ϕ, and write

$$\tilde{\phi}(z) = \sup_{y \in D}\phi(y) - \gamma\|y - z\|$$

for every $z \in \mathbb{R}^r$. If $x \in D$, then, for any $z \in \mathbb{R}^r$ and $y \in D$,

$$\phi(y) - \gamma\|y - z\| \leq \phi(x) + \gamma\|y - x\| - \gamma\|y - z\| \leq \phi(x) + \gamma\|z - x\|,$$

so that $\tilde{\phi}(z) \leq \phi(x) + \gamma\|z - x\|$; this shows, in particular, that $\tilde{\phi}(z) < \infty$. Also, if $z \in D$, we must have

$$\phi(z) - \gamma\|z - z\| \leq \tilde{\phi}(z) \leq \phi(z) + \gamma\|z - z\|,$$

so that $\tilde{\phi}$ extends ϕ. Finally, if w, $z \in \mathbb{R}^r$ and $y \in D$,

$$\phi(y) - \gamma\|y - w\| \leq \phi(y) - \gamma\|y - z\| + \gamma\|w - z\| \leq \tilde{\phi}(z) + \gamma\|w - z\|;$$

and taking the supremum over $y \in D$,

$$\tilde{\phi}(w) \leq \tilde{\phi}(z) + \gamma\|w - z\|.$$

As w and z are arbitrary, $\tilde{\phi}$ is Lipschitz.

(c)(i) Suppose that ϕ is γ-Lipschitz. If $\epsilon > 0$ and $a \leq a_1 \leq b_1 \leq \ldots \leq a_n \leq b_n \leq b$ and $\sum_{i=1}^n b_i - a_i \leq \epsilon/(1 + \gamma)$, then

$$\sum_{i=1}^n |\phi(b_i) - \phi(a_i)| \leq \sum_{i=1}^n \gamma|b_i - a_i| \leq \epsilon.$$

As ϵ is arbitrary, ϕ is absolutely continuous. If $x \in [a, b]$ and $\phi'(x)$ is defined, then

$$|\phi'(x)| = \lim_{y \to x}\frac{|\phi(y) - \phi(x)|}{|y - x|} \leq \gamma,$$

so ϕ' is bounded.

(ii) Now suppose that ϕ is absolutely continuous and that $|\phi'(x)| \le \gamma$ for every $x \in \operatorname{dom}\phi'$, where $\gamma \ge 0$. Then whenever $a \le x \le y \le b$,

$$|\phi(y) - \phi(x)| = |\int_x^y \phi'| \le \int_x^y |\phi'| \le \gamma(y-x)$$

(using 225E for the first equality). As x and y are arbitrary, ϕ is γ-Lipschitz.

262C Remark The argument for (b) above shows that if $\phi : D \to \mathbb{R}$ is a Lipschitz function, where $D \subseteq \mathbb{R}^r$, then ϕ has an extension to $\mathbb{R}^r$ with the same Lipschitz constants. In fact it is the case that if $\phi : D \to \mathbb{R}^s$ is a Lipschitz function, then ϕ has an extension to $\tilde{\phi} : \mathbb{R}^r \to \mathbb{R}^s$ with the same Lipschitz constants; this is 'Kirzbraun's theorem' (Kirzbraun 34, or Federer 69, 2.10.43).

262D Proposition If $\phi : D \to \mathbb{R}^r$ is a γ-Lipschitz function, where $D \subseteq \mathbb{R}^r$, then $\mu^*\phi[A] \le \gamma^r\mu^*A$ for every $A \subseteq D$, where μ is Lebesgue measure on $\mathbb{R}^r$. In particular, $\phi[D \cap A]$ is negligible for every negligible set $A \subseteq \mathbb{R}^r$.

proof Let $\epsilon > 0$. By 261F, there is a sequence $\langle B_n\rangle_{n\in\mathbb{N}} = \langle B(x_n, \delta_n)\rangle_{n\in\mathbb{N}}$ of closed balls in $\mathbb{R}^r$, covering A, such that $\sum_{n=0}^\infty \mu B_n \le \mu^*A + \epsilon$ and $\sum_{n\in\mathbb{N}\setminus K} \mu B_n \le \epsilon$, where $K = \{n : n \in \mathbb{N},\, x_n \in A\}$. Set

$$L = \{n : n \in \mathbb{N} \setminus K,\, B_n \cap D \ne \emptyset\},$$

and for $n \in L$ choose $y_n \in D \cap B_n$. Now set

$$\begin{aligned} B'_n &= B(\phi(x_n), \gamma\delta_n) \text{ if } n \in K,\\ &= B(\phi(y_n), 2\gamma\delta_n) \text{ if } n \in L,\\ &= \emptyset \text{ if } n \in \mathbb{N} \setminus (K \cup L). \end{aligned}$$

Then $\phi[B_n \cap D] \subseteq B'_n$ for every n, so $\phi[D \cap A] \subseteq \bigcup_{n\in\mathbb{N}} B'_n$, and

$$\begin{aligned} \mu^*\phi[A \cap D] &\le \sum_{n=0}^\infty \mu B'_n = \gamma^r \sum_{n\in K} \mu B_n + 2^r\gamma^r \sum_{n\in L} \mu B_n\\ &\le \gamma^r(\mu^*A + \epsilon) + 2^r\gamma^r\epsilon. \end{aligned}$$

As ϵ is arbitrary, $\mu^*\phi[A \cap D] \le \gamma^r\mu^*A$, as claimed.

262E Corollary Let $\phi : D \to \mathbb{R}^r$ be an injective Lipschitz function, where $D \subseteq \mathbb{R}^r$, and f a measurable function from a subset of $\mathbb{R}^r$ to $\mathbb{R}$.

(a) If ϕ^{-1} is defined almost everywhere in a subset H of $\mathbb{R}^r$ and f is defined almost everywhere in $\mathbb{R}^r$, then $f\phi^{-1}$ is defined almost everywhere in H.

(b) If $E \subseteq D$ is Lebesgue measurable then $\phi[E]$ is measurable.

(c) If D is measurable then $f\phi^{-1}$ is measurable.

proof Set

$$C = \operatorname{dom}(f\phi^{-1}) = \{y : y \in \phi[D],\, \phi^{-1}(y) \in \operatorname{dom} f\} = \phi[D \cap \operatorname{dom} f].$$

(a) Because f is defined almost everywhere, $\phi[D \setminus \operatorname{dom} f]$ is negligible. But now

$$C = \phi[D] \setminus \phi[D \setminus \operatorname{dom} f] = \operatorname{dom}\phi^{-1} \setminus \phi[D \setminus \operatorname{dom} f],$$

so

$$H \setminus C \subseteq (H \setminus \operatorname{dom}\phi^{-1}) \cup \phi[D \setminus \operatorname{dom} f]$$

is negligible.

(b) Now suppose that $E \subseteq D$ and that E is measurable. Let $\langle F_n\rangle_{n\in\mathbb{N}}$ be a sequence of closed bounded subsets of E such that $\mu(E \setminus \bigcup_{n\in\mathbb{N}} F_n) = 0$ (134Fb). Because ϕ is Lipschitz, it is continuous, so $\phi[F_n]$ is compact, therefore closed, therefore measurable for every n (2A2F, 2A2E, 115G); also $\phi[E \setminus \bigcup_{n\in\mathbb{N}} F_n]$ is negligible, by 262D, therefore measurable. So

$$\phi[E] = \phi[E \setminus \textstyle\bigcup_{n\in\mathbb{N}} F_n] \cup \bigcup_{n\in\mathbb{N}} \phi[F_n]$$

is measurable.

(c) For any $a \in \mathbb{R}$, take a measurable set $E \subseteq \mathbb{R}^r$ such that $\{x : f(x) \ge a\} = E \cap \operatorname{dom} f$. Then

$$\{y : y \in C,\ f\phi^{-1}(y) \ge a\} = C \cap \phi[D \cap E].$$

But $\phi[D \cap E]$ is measurable, by (b), so $\{y : f\phi^{-1}(y) \ge a\}$ is relatively measurable in C. As a is arbitrary, $f\phi^{-1}$ is measurable.

262F Differentiability I come now to the class of functions whose properties will take up most of the rest of the chapter.

Definitions Suppose r, $s \ge 1$ and that ϕ is a function from a subset $D = \operatorname{dom}\phi$ of $\mathbb{R}^r$ to $\mathbb{R}^s$.

(a) ϕ is **differentiable** at $x \in D$ if there is a real $s \times r$ matrix T such that

$$\lim_{y\to x} \frac{\|\phi(y)-\phi(x)-T(y-x)\|}{\|y-x\|} = 0;$$

in this case we may write $T = \phi'(x)$.

(b) I will say that ϕ is **differentiable relative to its domain** at x, and that T is **a** derivative of ϕ at x, if $x \in D$ and for every $\epsilon > 0$ there is a $\delta > 0$ such that $\|\phi(y) - \phi(x) - T(y - x)\| \le \epsilon\|y - x\|$ for every $y \in B(x, \delta) \cap D$.

262G Remarks (a) The standard definition in 262Fa, involving an all-sided limit '$\lim_{y\to x}$', implicitly requires ϕ to be defined on some non-trivial ball centered on x, so that we can calculate $\phi(y) - \phi(x) - T(y - x)$ for all y sufficiently near x. It has the advantage that the derivative $T = \phi'(x)$ is uniquely defined (because if $\lim_{z\to\mathbf{0}} \frac{\|T_1z-T_2z\|}{\|z\|} = 0$ then

$$\frac{\|(T_1-T_2)z\|}{\|z\|} = \lim_{\alpha\to 0} \frac{\|T_1(\alpha z)-T_2(\alpha z)\|}{\|\alpha z\|} = 0$$

for every non-zero z, so $T_1 - T_2$ must be the zero matrix). For our purposes here, there is some advantage in relaxing this slightly to the form in 262Fb, so that we do not need to pay special attention to the boundary of $\operatorname{dom}\phi$.

(b) If you have not seen this concept of 'differentiability' before, but have some familiarity with partial differentiation, it is necessary to emphasize that the concept of 'differentiable' function (at least in the strict sense demanded by 262Fa) is strictly stronger than the concept of 'partially differentiable' function. For purposes of computation, the most useful method of finding true derivatives is through 262Id below. For a simple example of a function with a full set of partial derivatives, which is not everywhere differentiable, consider $\phi : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\begin{aligned}\phi(\xi_1, \xi_2) &= \frac{\xi_1\xi_2}{\xi_1^2 + \xi_2^2} \text{ if } \xi_1^2 + \xi_2^2 \ne 0,\\ &= 0 \text{ if } \xi_1 = \xi_2 = 0.\end{aligned}$$

Then ϕ is not even continuous at $\mathbf{0}$, although both partial derivatives $\frac{\partial\phi}{\partial\xi_j}$ are defined everywhere.

(c) In the definition above, I speak of a derivative as being a matrix. Properly speaking, the derivative of a function defined on a subset of $\mathbb{R}^r$ and taking values in $\mathbb{R}^s$ should be thought of as a bounded linear operator from $\mathbb{R}^r$ to $\mathbb{R}^s$; the formulation in terms of matrices is acceptable just because there is a natural one-to-one correspondence between $s \times r$ real matrices and linear operators from $\mathbb{R}^r$ to $\mathbb{R}^s$, and all these linear operators are bounded. I use the 'matrix' description because it makes certain calculations more direct; in particular, the relationship between ϕ' and the partial derivatives of ϕ (262Ic), and the notion of the determinant $\det\phi'(x)$, used throughout §§263 and 265.

262H The norm of a matrix Some of the calculations below will rely on the notion of 'norm' of a matrix. The one I will use (in fact, for our purposes here, any norm would do) is the 'operator norm', defined by saying

$$\|T\| = \sup\{\|Tx\| : x \in \mathbb{R}^r,\ \|x\| \le 1\}$$

for any $s \times r$ matrix T. For the basic facts concerning these norms, see 2A4F-2A4G. The following will also be useful.

(a) If all the coefficients of T are small, so is $\|T\|$; in fact, if $T = \langle\tau_{ij}\rangle_{i\le s,j\le r}$, and $\|x\| \le 1$, then $|\xi_j| \le 1$ for each j, so

$$\|Tx\| = \big(\sum_{i=1}^s(\sum_{j=1}^r \tau_{ij}\xi_j)^2\big)^{1/2} \le \big(\sum_{i=1}^s(\sum_{j=1}^r |\tau_{ij}|)^2\big)^{1/2} \le r\sqrt{s}\max_{i\le s,j\le r} |\tau_{ij}|,$$

and $\|T\| \le r\sqrt{s}\max_{i\le s,j\le r} |\tau_{ij}|$. (This is a singularly crude inequality. A better one is in 262Ya. But it tells us, in particular, that $\|T\|$ is always finite.)

(b) If $\|T\|$ is small, so are all the coefficients of T; in fact, writing e_j for the jth unit vector of $\mathbb{R}^r$, then the ith coordinate of Te_j is τ_{ij}, so $|\tau_{ij}| \le \|Te_j\| \le \|T\|$.

262I Lemma Let $\phi : D \to \mathbb{R}^s$ be a function, where $D \subseteq \mathbb{R}^r$. For $i \le s$ let $\phi_i : D \to \mathbb{R}$ be its ith coordinate, so that $\phi(x) = (\phi_1(x), \dots, \phi_s(x))$ for $x \in D$.

(a) If ϕ is differentiable relative to its domain at $x \in D$, then ϕ is continuous at x.

(b) If $x \in D$, then ϕ is differentiable relative to its domain at x iff each ϕ_i is differentiable relative to its domain at x.

(c) If ϕ is differentiable at $x \in D$, then all the partial derivatives $\frac{\partial\phi_i}{\partial\xi_j}$ of ϕ are defined at x, and the derivative of ϕ at x is the matrix $\langle\frac{\partial\phi_i}{\partial\xi_j}(x)\rangle_{i\le s, j\le r}$.

(d) If all the partial derivatives $\frac{\partial\phi_i}{\partial\xi_j}$, for $i \le s$ and $j \le r$, are defined in a neighbourhood of $x \in D$ and are continuous at x, then ϕ is differentiable at x.

proof (a) Let T be a derivative of ϕ at x. Applying the definition 262Fb with $\epsilon = 1$, we see that there is a $\delta > 0$ such that

$$\|\phi(y) - \phi(x) - T(y - x)\| \le \|y - x\|$$

whenever $y \in D$ and $\|y - x\| \le \delta$. Now

$$\|\phi(y) - \phi(x)\| \le \|T(y - x)\| + \|y - x\| \le (1 + \|T\|)\|y - x\|$$

whenever $y \in D$ and $\|y - x\| \le \delta$, so ϕ is continuous at x.

(b)(i) If ϕ is differentiable relative to its domain at $x \in D$, let T be a derivative of ϕ at x. For $i \le s$ let T_i be the $1 \times r$ matrix consisting of the ith row of T. Let $\epsilon > 0$. Then we have a $\delta > 0$ such that

$$\begin{aligned}|\phi_i(y) - \phi_i(x) - T_i(y - x)| &\le \|\phi(y) - \phi(x) - T(y - x)\|\\ &\le \epsilon\|y - x\|\end{aligned}$$

whenever $y \in D$ and $\|y - x\| \le \delta$, so that T_i is a derivative of ϕ_i at x.

(ii) If each ϕ_i is differentiable relative to its domain at x, with corresponding derivatives T_i, let T be the $s \times r$ matrix with rows $T_1, \dots, T_s$. Given $\epsilon > 0$, there is for each $i \le s$ a $\delta_i > 0$ such that

$$|\phi_i(y) - \phi_i(x) - T_i y| \le \epsilon\|y - x\| \text{ whenever } y \in D,\ \|y - x\| \le \delta_i;$$

set $\delta = \min_{i\le s}\delta_i > 0$; then if $y \in D$ and $\|y - x\| \le \delta$, we shall have

$$\|\phi(y) - \phi(x) - T(y - x)\|^2 = \sum_{i=1}^s |\phi_i(y) - \phi_i(x) - T_i(y - x)|^2 \le s\epsilon^2\|y - x\|^2,$$

so that

$$\|\phi(y) - \phi(x) - T(y - x)\| \le \epsilon\sqrt{s}\|y - x\|.$$

As ϵ is arbitrary, T is a derivative of ϕ at x.

(c) Set $T = \phi'(x)$. We have

$$\lim_{y\to x}\frac{\|\phi(y)-\phi(x)-T(y-x)\|}{\|y-x\|} = 0;$$

fix $j \le r$, and consider $y = x + \eta e_j$, where $e_j = (0, \dots, 0, 1, 0, \dots, 0)$ is the jth unit vector in $\mathbb{R}^r$. Then we must have

$$\lim_{\eta\to 0}\frac{\|\phi(x+\eta e_j)-\phi(x)-\eta T(e_j)\|}{|\eta|} = 0.$$

Looking at the ith coordinate of $\phi(x + \eta e_j) - \phi(x) - \eta T(e_j)$, we have

$$|\phi_i(x + \eta e_j) - \phi_i(x) - \tau_{ij}\eta| \le \|\phi(x + \eta e_j) - \phi(x) - \eta T(e_j)\|,$$

where τ_{ij} is the (i, j)th coefficient of T; so that

$$\lim_{\eta\to 0}\frac{|\phi_i(x+\eta e_j)-\phi_i(x)-\tau_{ij}\eta|}{|\eta|} = 0.$$

But this just says that the partial derivative $\frac{\partial\phi_i}{\partial\xi_j}(x)$ exists and is equal to τ_{ij}, as claimed.

(d) Now suppose that the partial derivatives $\frac{\partial\phi_i}{\partial\xi_j}$ are defined near x and continuous at x. Let $\epsilon > 0$. Let $\delta > 0$ be such that

$$|\frac{\partial\phi_i}{\partial\xi_j}(y)-\tau_{ij}|\le\epsilon$$

whenever $\|y-x\|\le\delta$, writing $\tau_{ij}=\frac{\partial\phi_i}{\partial\xi_j}(x)$. Now suppose that $\|y-x\|\le\delta$. Set

$$y=(\eta_1,\dots,\eta_r),\quad x=(\xi_1,\dots,\xi_r),$$

$$y_j=(\eta_1,\dots,\eta_j,\xi_{j+1},\dots,\xi_r)\text{ for }0\le j\le r,$$

so that $y_0=x$, $y_r=y$ and the line segment between y_{j-1} and y_j lies wholly within δ of x whenever $1\le j\le r$, since if z lies on this line segment then ζ_i lies between ξ_i and η_i for every i. By the ordinary mean value theorem for differentiable real functions, applied to the function

$$t\mapsto\phi_i(\eta_1,\dots,\eta_{j-1},t,\xi_{j+1},\dots,\xi_r),$$

there is for each $i\le s$, $j\le r$ a point z_{ij} on the line segment between y_{j-1} and y_j such that

$$\phi_i(y_j)-\phi_i(y_{j-1})=(\eta_j-\xi_j)\frac{\partial\phi_i}{\partial\xi_j}(z_{ij}).$$

But

$$|\frac{\partial\phi_i}{\partial\xi_j}(z_{ij})-\tau_{ij}|\le\epsilon,$$

so

$$|\phi_i(y_j)-\phi_i(y_{j-1})-\tau_{ij}(\eta_j-\xi_j)|\le\epsilon|\eta_j-\xi_j|\le\epsilon\|y-x\|.$$

Summing over j,

$$|\phi_i(y)-\phi_i(x)-\textstyle\sum_{j=1}^r\tau_{ij}(\eta_j-\xi_j)|\le r\epsilon\|y-x\|$$

for each i. Summing the squares and taking the square root,

$$\|\phi(y)-\phi(x)-T(y-x)\|\le\epsilon r\sqrt{s}\|y-x\|,$$

where $T=\langle\tau_{ij}\rangle_{i\le s,j\le r}$. And this is true whenever $\|y-x\|\le\delta$. As ϵ is arbitrary, $\phi'(x)=T$ is defined.

262J Remark I am not sure if I ought to apologize for the notation $\frac{\partial}{\partial\xi_j}$. In such formulae as $(\eta_j-\xi_j)\frac{\partial\phi_i}{\partial\xi_j}(z_{ij})$ above, the two appearances of ξ_j clash most violently. But I do not think that any person of good will is likely to be misled, provided that the labels ξ_j (or whatever symbols are used to represent the variables involved) are adequately described when the domain of ϕ is first introduced (and always remembering that in partial differentiation, we are not only moving one variable – a ξ_j in the present context – but holding fixed some further list of variables, not listed in the notation). I believe that the traditional notation $\frac{\partial}{\partial\xi_j}$ has survived for solid reasons, and I should like to offer a welcome to those who are more comfortable with it than with any of the many alternatives which have been proposed, but have never taken root.

262K The Cantor function revisited It is salutary to re-examine the examples of 134H-134I in the light of the present considerations. Let $f:[0,1]\to[0,1]$ be the Cantor function (134H) and set $g(x)=\frac{1}{2}(x+f(x))$ for $x\in[0,1]$. Then $g:[0,1]\to[0,1]$ is a homeomorphism (134I); set $\phi=g^{-1}:[0,1]\to[0,1]$. We see that if $0\le x\le y\le 1$ then $g(y)-g(x)\ge\frac{1}{2}(y-x)$; equivalently, $\phi(y)-\phi(x)\le 2(y-x)$ whenever $0\le x\le y\le 1$, so that ϕ is a Lipschitz function, therefore absolutely continuous (262Bc). If $D=\{x:\phi'(x)\text{ is defined}\}$, then $[0,1]\setminus D$ is negligible (225Cb), so $[0,1]\setminus\phi[D]=\phi[[0,1]\setminus D]$ is negligible (262Da). I noted in 134I that there is a measurable function $h:[0,1]\to\mathbb{R}$ such that the composition $h\phi$ is not measurable; now $h(\phi\restriction D)=(h\phi)\restriction D$ cannot be measurable, even though $\phi\restriction D$ is differentiable.

262L It will be convenient to be able to call on the following straightforward result.

Lemma Suppose that $D\subseteq\mathbb{R}^r$ and $x\in\mathbb{R}^r$ are such that $\lim_{\delta\downarrow 0}\frac{\mu^*(D\cap B(x,\delta))}{\mu B(x,\delta)}=1$. Then $\lim_{z\to\mathbf{0}}\frac{\rho(x+z,D)}{\|z\|}=0$, where $\rho(x+z,D)=\inf_{y\in D}\|x+z-y\|$.

proof Let $\epsilon>0$. Let $\delta_0>0$ be such that

$$\mu^*(D\cap B(x,\delta))>(1-(\tfrac{\epsilon}{1+\epsilon})^r)\mu B(x,\delta)$$

whenever $0<\delta\le\delta_0$. Take any z such that $0<\|z\|\le\delta_0/(1+\epsilon)$. **?** Suppose, if possible, that $\rho(x+z,D)>\epsilon\|z\|$. Then $B(x+z,\epsilon\|z\|)\subseteq B(x,(1+\epsilon)\|z\|)\setminus D$, so

$$\mu^*(D\cap B(x,(1+\epsilon)\|z\|)) \le \mu B(x,(1+\epsilon)\|z\|) - \mu B(x+z,\epsilon\|z\|)$$
$$= (1-(\tfrac{\epsilon}{1+\epsilon})^r)\mu B(x,(1+\epsilon)\|z\|),$$

which is impossible, as $(1+\epsilon)\|z\| \le \delta_0$. **X** Thus $\rho(x+z,D) \le \epsilon\|z\|$. As ϵ is arbitrary, this proves the result.

Remark There is a word for this; see 261Yg.

262M I come now to the first result connecting Lipschitz functions with differentiable functions. I approach it through a substantial lemma which will be the foundation of §263.

Lemma Let r, $s \ge 1$ be integers and ϕ a function from a subset D of $\mathbb{R}^r$ to $\mathbb{R}^s$ which is differentiable at each point of its domain. For each $x \in D$ let $T(x)$ be a derivative of ϕ. Let M_{sr} be the set of $s\times r$ matrices and $\zeta : A \to \,]0,\infty[$ a strictly positive function, where $A \subseteq M_{sr}$ is a non-empty set containing $T(x)$ for every $x \in D$. Then we can find sequences $\langle D_n\rangle_{n\in\mathbb{N}}$, $\langle T_n\rangle_{n\in\mathbb{N}}$ such that

(i) $\langle D_n\rangle_{n\in\mathbb{N}}$ is a partition of D into sets which are relatively measurable in D, that is, are intersections of D with measurable subsets of $\mathbb{R}^r$;

(ii) $T_n \in A$ for every n;

(iii) $\|\phi(x)-\phi(y)-T_n(x-y)\| \le \zeta(T_n)\|x-y\|$ for every $n\in\mathbb{N}$ and x, $y \in D_n$;

(iv) $\|T(x)-T_n\| \le \zeta(T_n)$ for every $x \in D_n$.

proof (a) The first step is to note that there is a sequence $\langle S_n\rangle_{n\in\mathbb{N}}$ in A such that

$$A \subseteq \bigcup_{n\in\mathbb{N}}\{T : T\in M_{sr},\ \|T-S_n\| < \zeta(S_n)\}.$$

P (Of course this is a standard result about separable metric spaces.) Write Q for the set of matrices in M_{sr} with rational coefficients; then there is a natural bijection between Q and $\mathbb{Q}^{sr}$, so Q and $Q\times\mathbb{N}$ are countable. Enumerate $Q\times\mathbb{N}$ as $\langle(R_n,k_n)\rangle_{n\in\mathbb{N}}$. For each $n\in\mathbb{N}$, choose $S_n \in A$ by the rule

— if there is an $S\in A$ such that $\{T : \|T-R_n\| \le 2^{-k_n}\} \subseteq \{T : \|T-S\| < \zeta(S)\}$, take such an S for S_n;

— otherwise, take S_n to be any member of A.

I claim that this works. For let $S\in A$. Then $\zeta(S) > 0$; take $k\in\mathbb{N}$ such that $2^{-k} < \zeta(S)$. Take $R^*\in Q$ such that $\|R^*-S\| < \min(\zeta(S)-2^{-k},2^{-k})$; this is possible because $\|R-S\|$ will be small whenever all the coefficients of R are close enough to the corresponding coefficients of S (262Ha), and we can find rational numbers to achieve this. Let $n\in\mathbb{N}$ be such that $R^*=R_n$ and $k=k_n$. Then

$$\{T : \|T-R_n\| \le 2^{-k_n}\} \subseteq \{T : \|T-S\| < \zeta(S)\}$$

(because $\|T-S\| \le \|T-R_n\| + \|R_n-S\|$), so we must have chosen S_n by the first part of the rule above, and

$$S \in \{T : \|T-R_n\| \le 2^{-k_n}\} \subseteq \{T : \|T-S_n\| < \zeta(S_n)\}.$$

As S is arbitrary, this proves the result. **Q**

(b) Enumerate $\mathbb{Q}^r\times\mathbb{Q}^r\times\mathbb{N}$ as $\langle(q_n,q'_n,m_n)\rangle_{n\in\mathbb{N}}$. For each $n\in\mathbb{N}$, set

$$\begin{aligned}H_n &= \{x : x\in[q_n,q'_n]\cap D,\ \|\phi(y)-\phi(x)-S_{m_n}(y-x)\| \le \zeta(S_{m_n})\|y-x\| \\ &\qquad\qquad \text{for every } y\in[q_n,q'_n]\cap D\} \\ &= [q_n,q'_n]\cap D\cap\bigcap_{y\in[q_n,q'_n]\cap D}\{x : x\in D, \\ &\qquad\qquad \|\phi(y)-\phi(x)-S_{m_n}(y-x)\| \le \zeta(S_{m_n})\|y-x\|\}.\end{aligned}$$

Because ϕ is continuous, $H_n = D\cap\overline{H}_n$, writing $\overline{H}_n$ for the closure of H_n, so H_n is relatively measurable in D. Note that if x, $y\in H_n$, then $y\in D\cap[q_n,q'_n]$, so that

$$\|\phi(y)-\phi(x)-S_{m_n}(y-x)\| \le \zeta(S_{m_n})\|y-x\|.$$

Set

$$H'_n = \{x : x\in H_n,\ \|T(x)-S_{m_n}\| \le \zeta(S_{m_n})\}.$$

(c) $D = \bigcup_{n\in\mathbb{N}}H'_n$. **P** Let $x\in D$. Then $T(x)\in A$, so there is a $k\in\mathbb{N}$ such that $\|T(x)-S_k\| < \zeta(S_k)$. Let $\delta > 0$ be such that

$$\|\phi(y)-\phi(x)-T(x)(x-y)\| \le (\zeta(S_k)-\|T(x)-S_k\|)\|x-y\|$$

whenever $y \in D$ and $\|y-x\| \le \delta$. Then

$$\|\phi(y)-\phi(x)-S_k(x-y)\| \le (\zeta(S_k)-\|T(x)-S_k\|)\|x-y\|+\|T(x)-S_k\|\|x-y\|$$
$$\le \zeta(S_k)\|x-y\|$$

whenever $y \in D\cap B(x,\delta)$. Let q, $q' \in \mathbb{Q}^r$ be such that $x \in [q,q'] \subseteq B(x,\delta)$. Let n be such that $q=q_n$, $q'=q'_n$ and $k=m_n$. Then $x \in H'_n$. **Q**

(d) Write

$$C_n=\{x : x \in H_n,\, \lim_{\delta\downarrow 0}\frac{\mu^*(H_n\cap B(x,\delta))}{\mu B(x,\delta)}=1\}.$$

Then $C_n \subseteq H'_n$.

P (i) Take $x \in C_n$, and set $\tilde{T}=T(x)-S_{m_n}$. I have to show that $\|\tilde{T}\| \le \zeta(S_{m_n})$. Take $\epsilon>0$. Let $\delta_0>0$ be such that

$$\|\phi(y)-\phi(x)-T(x)(y-x)\| \le \epsilon\|y-x\|$$

whenever $y \in D$ and $\|y-x\| \le \delta_0$. Since

$$\|\phi(y)-\phi(x)-S_{m_n}(y-x)\| \le \zeta(S_{m_n})\|y-x\|$$

whenever $y \in H_n$, we have

$$\|\tilde{T}(y-x)\| \le (\epsilon+\zeta(S_{m_n}))\|y-x\|$$

whenever $y \in H_n$ and $\|y-x\| \le \delta_0$.

(ii) By 262L, there is a $\delta_1>0$ such that $(1+2\epsilon)\delta_1 \le \delta_0$ and $\rho(x+z,H_n) \le \epsilon\|z\|$ whenever $0<\|z\| \le \delta_1$. So if $\|z\| \le \delta_1$ there is a $y \in H_n$ such that $\|x+z-y\| \le 2\epsilon\|z\|$. (If $z=0$ we can take $y=x$.) Now $\|x-y\| \le (1+2\epsilon)\|z\| \le \delta_0$, so

$$\begin{aligned}\|\tilde{T}z\| &\le \|\tilde{T}(y-x)\|+\|\tilde{T}(x+z-y)\|\\ &\le (\epsilon+\zeta(S_{m_n}))\|y-x\|+\|\tilde{T}\|\|x+z-y\|\\ &\le (\epsilon+\zeta(S_{m_n}))\|z\|+(\epsilon+\zeta(S_{m_n})+\|\tilde{T}\|)\|x+z-y\|\\ &\le (\epsilon+\zeta(S_{m_n})+2\epsilon^2+2\epsilon\zeta(S_{m_n})+2\epsilon\|\tilde{T}\|)\|z\|.\end{aligned}$$

And this is true whenever $0<\|z\| \le \delta_1$. But multiplying this inequality by suitable positive scalars we see that

$$\|\tilde{T}z\| \le \bigl(\epsilon+\zeta(S_{m_n})+2\epsilon^2+2\epsilon\zeta(S_{m_n})+2\epsilon\|\tilde{T}\|\bigr)\|z\|$$

for all $z \in \mathbb{R}^r$, and

$$\|\tilde{T}\| \le \epsilon+\zeta(S_{m_n})+2\epsilon^2+2\epsilon\zeta(S_{m_n})+2\epsilon\|\tilde{T}\|.$$

As ϵ is arbitrary, $\|\tilde{T}\| \le \zeta(S_{m_n})$, as claimed. **Q**

(e) By 261Da, $H_n\setminus C_n$ is negligible for every n, so $H_n\setminus H'_n$ is negligible, and

$$H'_n=D\cap(\overline{H}_n\setminus(H_n\setminus H'_n))$$

is relatively measurable in D. Set

$$D_n=H'_n\setminus\bigcup_{k<n}H'_k,\quad T_n=S_{m_n}$$

for each n; these serve.

262N Corollary Let ϕ be a function from a subset D of $\mathbb{R}^r$ to $\mathbb{R}^s$, and suppose that ϕ is differentiable relative to its domain at each point of D. Then D can be expressed as the union of a sequence $\langle D_n\rangle_{n\in\mathbb{N}}$ of sets such that $\phi\restriction D_n$ is Lipschitz for each $n \in \mathbb{N}$.

proof In 262M, take $\zeta(T)=1$ for every $T \in A=M_{sr}$. If x, $y \in D_n$ then

$$\|\phi(x)-\phi(y)\| \le \|\phi(x)-\phi(y)-T_n(x-y)\|+\|T_n(x-y)\|$$
$$\le \|x-y\|+\|T_n\|\|x-y\|,$$

so $\phi\restriction D_n$ is $(1+\|T_n\|)$-Lipschitz.

262O Corollary Suppose that ϕ is an injective function from a measurable subset D of $\mathbb{R}^r$ to $\mathbb{R}^r$, and that ϕ is differentiable relative to its domain at every point of D.

(a) If $A \subseteq D$ is negligible, $\phi[A]$ is negligible.

(b) If $E \subseteq D$ is measurable, then $\phi[E]$ is measurable.

(c) If D is measurable and f is a measurable function defined on a subset of $\mathbb{R}^r$, then $f\phi^{-1}$ is measurable.

(d) If $H \subseteq \mathbb{R}^r$ and ϕ^{-1} is defined almost everywhere on H, and if f is a function defined almost everywhere in $\mathbb{R}^r$, then $f\phi^{-1}$ is defined almost everywhere in H.

proof Let $\langle D_n\rangle_{n\in\mathbb{N}}$ be a sequence of measurable sets with union D such that $\phi\restriction D_n$ is Lipschitz for each n. Then $\phi[A\cap D] = \bigcup_{n\in\mathbb{N}}(\phi\restriction D_n)[A\cap D_n]$ is negligible for every negligible $A \subseteq \mathbb{R}^r$, by 262D.

Now parts (b)-(d) follow from (a) (because ϕ is continuous), just as in 262E.

262P Corollary Let ϕ be a function from a a subset D of $\mathbb{R}^r$ to $\mathbb{R}^s$, and suppose that ϕ is differentiable relative to its domain, with a derivative $T(x)$, at each point $x \in D$. Then the function $x \mapsto T(x)$ is measurable in the sense that $\tau_{ij} : D \to \mathbb{R}$ is measurable for all $i \le s$ and $j \le r$, where $\tau_{ij}(x)$ is the (i,j)th coefficient of the matrix $T(x)$ for all i, j and x.

proof For each $k \in \mathbb{N}$, apply 262M with $\zeta(T) = 2^{-k}$ for each $T \in A = M_{sr}$, obtaining sequences $\langle D_{kn}\rangle_{n\in\mathbb{N}}$ of relatively measurable subsets of D and $\langle T_{kn}\rangle_{n\in\mathbb{N}}$ in M_{sr}. Let $\tau_{ij}^{(kn)}$ be the (i,j)th coefficient of T_{kn}. Then we have functions $f_{ijk} : D \to \mathbb{R}$ defined by setting

$$f_{ijk}(x) = \tau_{ij}^{(kn)} \text{ if } x \in D_{kn}.$$

Because the D_{kn} are relatively measurable, the f_{ijk} are measurable functions. For $x \in D_{kn}$,

$$|\tau_{ij}(x) - f_{ijk}(x)| \le \|T(x) - T_n\| \le 2^{-k},$$

so $|\tau_{ij}(x) - f_{ijk}(x)| \le 2^{-k}$ for every $x \in D$, and

$$\tau_{ij} = \lim_{k\to\infty} f_{ijk}$$

is measurable, as claimed.

***262Q** This concludes the part of the section which is essential for the rest of the chapter. However the main results of §263 will I think be better understood if you are aware of the fact that any Lipschitz function is differentiable (relative to its domain) almost everywhere in its domain. I devote the next couple of pages to a proof of this fact, which apart from its intrinsic interest is a useful exercise.

Rademacher's theorem Let ϕ be a Lipschitz function from a subset of $\mathbb{R}^r$ to $\mathbb{R}^s$, where $s \ge 1$. Then ϕ is differentiable relative to its domain almost everywhere in its domain.

proof (a) By 262Ba and 262Ib, it will be enough to deal with the case $s = 1$. By 262Bb, there is a Lipschitz function $\tilde\phi : \mathbb{R}^r \to \mathbb{R}$ extending ϕ; now ϕ is differentiable with respect to its domain at any point of $\operatorname{dom}\phi$ at which $\tilde\phi$ is differentiable, so it will be enough if I can show that $\tilde\phi$ is differentiable almost everywhere. To make the notation more agreeable to the eye, I will suppose that ϕ itself was defined everywhere in $\mathbb{R}^r$. Let γ be a Lipschitz constant for ϕ.

The proof proceeds by induction on r. If $r = 1$, we have a Lipschitz function $\phi : \mathbb{R} \to \mathbb{R}$; now ϕ is absolutely continuous in any bounded interval (262Bc), therefore differentiable almost everywhere. Thus the induction starts. The rest of the proof is devoted to the inductive step to $r > 1$.

(b) The first step is to show that all the partial derivatives $\frac{\partial\phi}{\partial\xi_j}$ are defined almost everywhere and are Borel measurable. **P** Take $j \le r$. For $q \in \mathbb{Q}\setminus\{0\}$ set

$$\Delta_q(x) = \frac{1}{q}(\phi(x + qe_j) - \phi(x)),$$

writing e_j for the jth unit vector of $\mathbb{R}^r$. Because ϕ is continuous, so is Δ_q, so that Δ_q is a Borel measurable function for each q. Next, for any $x \in \mathbb{R}^r$,

$$D^+(x) = \limsup_{\delta\to0}\frac{1}{\delta}(\phi(x+\delta e_j) - \phi(x)) = \lim_{n\to\infty}\sup_{q\in\mathbb{Q},0<|q|\le2^{-n}}\Delta_q(x),$$

so that the set on which $D^+(x)$ is defined in $\mathbb{R}$ is Borel and D^+ is a Borel measurable function. Similarly,

$$D^-(x) = \liminf_{\delta\to0}\frac{1}{\delta}(\phi(x+\delta e_j) - \phi(x))$$

is a Borel measurable function with Borel domain. So

$$E = \{x : \frac{\partial\phi}{\partial\xi_j}(x) \text{ exists in } \mathbb{R}\} = \{x : D^+(x) = D^-(x) \in \mathbb{R}\}$$

is a Borel set, and $\frac{\partial\phi}{\partial\xi_j}$ is a Borel measurable function.

On the other hand, if we identify $\mathbb{R}^r$ with $\mathbb{R}^J \times \mathbb{R}$, taking J to be $\{1,\dots,j-1,j+1,\dots,r\}$, then we can think of the measure μ on $\mathbb{R}^r$ as being the product of Lebesgue measure μ_J on $\mathbb{R}^J$ with Lebesgue measure μ_1 on $\mathbb{R}$ (251N). Now for every $y \in \mathbb{R}^J$ we have a function $\phi_y : \mathbb{R} \to \mathbb{R}$ defined by writing

$$\phi_y(\sigma) = \phi(y,\sigma),$$

and E becomes

$$\{(y,\sigma) : \phi'_y(\sigma) \text{ is defined}\},$$

so that all the sections

$$\{\sigma : (y,\sigma) \in E\}$$

are conegligible subsets of $\mathbb{R}$, because every ϕ_y is Lipschitz, therefore differentiable almost everywhere, as remarked in part (a) of the proof. Since we know that E is measurable, it must be conegligible, by Fubini's theorem (apply 252D or 252F to the complement of E). Thus $\frac{\partial\phi}{\partial\xi_j}$ is defined almost everywhere, as claimed. **Q**

Write

$$H = \{x : x \in \mathbb{R}^r,\ \frac{\partial\phi}{\partial\xi_j}(x) \text{ exists for every } j \le r\},$$

so that H is a conegligible Borel set in $\mathbb{R}^r$.

(c) For the rest of this proof, I fix on the natural identification of $\mathbb{R}^r$ with $\mathbb{R}^{r-1} \times \mathbb{R}$, identifying $(\xi_1,\dots,\xi_r)$ with $((\xi_1,\dots,\xi_{r-1}),\xi_r)$. For $x \in H$, let $T(x)$ be the $1 \times r$ matrix $(\frac{\partial\phi}{\partial\xi_1}(x),\dots,\frac{\partial\phi}{\partial\xi_r}(x))$.

(d) Set

$$H_1 = \{x : x \in H,\ \lim_{u\to\mathbf{0} \text{ in } \mathbb{R}^{r-1}} \frac{|\phi(x+(u,0))-\phi(x)-T(x)(u,0)|}{\|u\|} = 0\}.$$

I claim that H_1 is conegligible in $\mathbb{R}^r$. **P** This is really the same idea as in (b). For $x \in H$, $x \in H_1$ iff

for every $\epsilon > 0$ there is a $\delta > 0$ such that

$$|\phi(x+(u,0)) - \phi(x) - T(x)(u,0)| \le \epsilon\|u\|$$

whenever $\|u\| \le \delta$,

that is, iff

for every $m \in \mathbb{N}$ there is an $n \in \mathbb{N}$ such that

$$|\phi(x+(u,0)) - \phi(x) - T(x)(u,0)| \le 2^{-m}\|u\|$$

whenever $u \in \mathbb{Q}^{r-1}$ and $\|u\| \le 2^{-n}$.

But for any particular $m \in \mathbb{N}$ and $u \in \mathbb{Q}^{r-1}$ the set

$$\{x : |\phi(x+(u,0)) - \phi(x) - T(x)(u,0)| \le 2^{-m}\|u\|\}$$

is measurable, indeed Borel, because all the functions $x \mapsto \phi(x+(u,0))$, $x \mapsto \phi(x)$, $x \mapsto T(x)(u,0)$ are Borel measurable. So H_1 is of the form

$$\bigcap_{m\in\mathbb{N}} \bigcup_{n\in\mathbb{N}} \bigcap_{u\in\mathbb{Q}^{r-1},\|u\|\le 2^{-n}} E_{mnu}$$

where every E_{mnu} is a measurable set, and H_1 is therefore measurable.

Now however observe that for any $\sigma \in \mathbb{R}$, the function

$$v \mapsto \phi_\sigma(v) = \phi(v,\sigma) : \mathbb{R}^{r-1} \to \mathbb{R}$$

is Lipschitz, therefore (by the inductive hypothesis) differentiable almost everywhere in $\mathbb{R}^{r-1}$; and that $(v,\sigma) \in H_1$ iff $(v,\sigma) \in H$ and $\phi'_\sigma(v)$ is defined. Consequently $\{v : (v,\sigma) \in H_1\}$ is conegligible whenever $\{v : (v,\sigma) \in H\}$ is, that is, for almost every $\sigma \in \mathbb{R}$; so that H_1, being measurable, must be conegligible. **Q**

(e) Now, for q, $q' \in \mathbb{Q}$ and $n \in \mathbb{N}$, set

$$F(q,q',n) = \{x : x \in \mathbb{R}^r,\ q \le \frac{\phi(x+(\mathbf{0},\eta))-\phi(x)}{\eta} \le q' \text{ whenever } 0 < |\eta| \le 2^{-n}\}.$$

Set

$$F_*(q,q',n)=\{x:x\in F(q,q',n),\ \lim_{\delta\downarrow 0}\frac{\mu^*(F(q,q',n)\cap B(x,\delta))}{\mu B(x,\delta)}=1\}.$$

By 261Da, $F(q,q',n)\setminus F_*(q,q',n)$ is negligible for all q, q', n, so that

$$H_2=H_1\setminus\bigcup_{q,q'\in\mathbb{Q},n\in\mathbb{N}}(F(q,q',n)\setminus F_*(q,q',n))$$

is conegligible.

(f) I claim that ϕ is differentiable at every point of H_2. **P** Take $x=(u,\sigma)\in H_2$. Then $\alpha=\dfrac{\partial\phi}{\partial\xi_r}(x)$ and $T=T(x)$ are defined. Let γ be a Lipschitz constant for ϕ.

Take $\epsilon>0$; take q, $q'\in\mathbb{Q}$ such that $\alpha-\epsilon\le q<\alpha<q'\le\alpha+\epsilon$. There must be an $n\in\mathbb{N}$ such that $x\in F(q,q',n)$; consequently $x\in F_*(q,q',n)$, by the definition of H_2. By 262L, there is a $\delta_0>0$ such that $\rho(x+z,F(q,q',n))\le\epsilon\|z\|$ whenever $\|z\|\le\delta_0$. Next, there is a $\delta_1>0$ such that $|\phi(x+(v,0))-\phi(x)-T(v,0)|\le\epsilon\|v\|$ whenever $v\in\mathbb{R}^{r-1}$ and $\|v\|\le\delta_1$. Set

$$\delta=\min(\delta_0,\delta_1,2^{-n})/(1+2\epsilon)>0.$$

Suppose that $z=(v,\tau)\in\mathbb{R}^r$ and that $\|z\|\le\delta$. Because $\|z\|\le\delta_0$ there is an $x'=(u',\sigma')\in F(q,q',n)$ such that $\|x+z-x'\|\le 2\epsilon\|z\|$; set $x^*=(u',\sigma)$. Now

$$\max(\|u-u'\|,|\sigma-\sigma'|)\le\|x-x'\|\le(1+2\epsilon)\|z\|\le\min(\delta_1,2^{-n}).$$

so

$$|\phi(x^*)-\phi(x)-T(x^*-x)|\le\epsilon\|u'-u\|\le\epsilon(1+2\epsilon)\|z\|.$$

But also

$$|\phi(x')-\phi(x^*)-T(x'-x^*)|=|\phi(x')-\phi(x^*)-\alpha(\sigma'-\sigma)|\le\epsilon|\sigma'-\sigma|\le\epsilon(1+2\epsilon)\|z\|,$$

because $x'\in F(q,q',n)$ and $|\sigma-\sigma'|\le 2^{-n}$, so that (if $x'\ne x^*$)

$$\alpha-\epsilon\le q\le\frac{\phi(x^*)-\phi(x')}{\sigma-\sigma'}\le q'\le\alpha+\epsilon$$

and

$$\Bigl|\frac{\phi(x')-\phi(x^*)}{\sigma'-\sigma}-\alpha\Bigr|\le\epsilon.$$

Finally,

$$|\phi(x+z)-\phi(x')|\le\gamma\|x+z-x'\|\le 2\gamma\epsilon\|z\|,$$

$$|Tz-T(x'-x)|\le\|T\|\|x+z-x'\|\le 2\epsilon\|T\|\|z\|.$$

Putting all these together,

$$\begin{aligned}|\phi(x+z)-\phi x-Tz|&\le|\phi(x+z)-\phi(x')|+|T(x'-x)-Tz|\\&\qquad+|\phi(x')-\phi(x^*)-T(x'-x^*)|+|\phi(x^*)-\phi(x)-T(x^*-x)|\\&\le 2\gamma\epsilon\|z\|+2\epsilon\|T\|\|z\|+\epsilon(1+2\epsilon)\|z\|+\epsilon(1+2\epsilon)\|z\|\\&=\epsilon(2\gamma+2\|T\|+2+4\epsilon)\|z\|.\end{aligned}$$

And this is true whenever $\|z\|\le\delta$. As ϵ is arbitrary, ϕ is differentiable at x. **Q**

Thus $\{x:\phi$ is differentiable at $x\}$ includes H_2 and is conegligible; and the induction continues.

262X Basic exercises (a) Let ϕ and ψ be Lipschitz functions from subsets of $\mathbb{R}^r$ to $\mathbb{R}^s$. Show that $\phi+\psi$ is a Lipschitz function from $\operatorname{dom}\phi\cap\operatorname{dom}\psi$ to $\mathbb{R}^s$.

(b) Let ϕ be a Lipschitz function from a subset of $\mathbb{R}^r$ to $\mathbb{R}^s$, and $c\in\mathbb{R}$. Show that $c\phi$ is a Lipschitz function.

(c) Suppose $\phi:D\to\mathbb{R}^s$ and $\psi:E\to\mathbb{R}^q$ are Lipschitz functions, where $D\subseteq\mathbb{R}^r$ and $E\subseteq\mathbb{R}^s$. Show that the composition $\psi\phi:D\cap\phi^{-1}[E]\to\mathbb{R}^q$ is Lipschitz.

(d) Suppose ϕ, ψ are functions from subsets of $\mathbb{R}^r$ to $\mathbb{R}^s$, and suppose that $x\in\operatorname{dom}\phi\cap\operatorname{dom}\psi$ is such that each function is differentiable relative to its domain at x, with derivatives S, T there. Show that $\phi+\psi$ is differentiable relative to its domain at x, and that $S+T$ is a derivative of $\phi+\psi$ at x.

(e) Suppose that ϕ is a function from a subset of $\mathbb{R}^r$ to $\mathbb{R}^s$, and is differentiable relative to its domain at $x \in \operatorname{dom}\phi$. Show that $c\phi$ is differentiable relative to its domain at x for every $c \in \mathbb{R}$.

>(f) Suppose $\phi : D \to \mathbb{R}^s$ and $\psi : E \to \mathbb{R}^q$ are functions, where $D \subseteq \mathbb{R}^r$ and $E \subseteq \mathbb{R}^s$; suppose that ϕ is differentiable relative to its domain at $x \in D \cap \phi^{-1}[E]$, with an $s \times r$ matrix T a derivative there, and that ψ is differentiable relative to its domain at $\phi(x)$, with a $q \times s$ matrix S a derivative there. Show that the composition $\psi\phi$ is differentiable relative to its domain at x, and that the $q \times r$ matrix ST is a derivative of $\psi\phi$ at x.

(g) Let $\phi : \mathbb{R}^r \to \mathbb{R}^s$ be a linear operator, with associated matrix T. Show that ϕ is differentiable everywhere, with $\phi'(x) = T$ for every x.

>(h) Let $G \subseteq \mathbb{R}^r$ be a convex open set, and $\phi : G \to \mathbb{R}^s$ a function such that all the partial derivatives $\frac{\partial \phi_i}{\partial \xi_j}$ are defined everywhere in G. Show that ϕ is Lipschitz iff all the partial derivatives are bounded on G.

(i) Let $\phi : \mathbb{R}^r \to \mathbb{R}^s$ be a function. Show that ϕ is differentiable at $x \in \mathbb{R}^r$ iff for every $m \in \mathbb{N}$ there are an $n \in \mathbb{N}$ and an $r \times s$ matrix T with rational coefficients such that $\|\phi(y) - \phi(x) - T(y - x)\| \le 2^{-m}\|y - x\|$ whenever $\|y - x\| \le 2^{-n}$.

>(j) Suppose that f is a real-valued function which is integrable over $\mathbb{R}^r$, and that $g : \mathbb{R}^r \to \mathbb{R}$ is a bounded differentiable function such that the partial derivative $\dfrac{\partial g}{\partial \xi_j}$ is bounded, where $j \le r$. Let $f * g$ be the convolution of f and g (255L). Show that $\dfrac{\partial}{\partial \xi_j}(f * g)$ is defined everywhere and equal to $f * \dfrac{\partial g}{\partial \xi_j}$. (*Hint*: 255Xd.)

>(k) Let (X, Σ, μ) be a measure space, $G \subseteq \mathbb{R}^r$ an open set, and $f : X \times G \to \mathbb{R}$ a function. Suppose that

(i) for every $x \in X$, $t \mapsto f(x, t) : G \to \mathbb{R}$ is differentiable;

(ii) there is an integrable function g on X such that $|\frac{\partial f}{\partial \tau_j}(x, t)| \le g(x)$ whenever $x \in X$, $t \in G$ and $j \le r$;

(iii) $\int |f(x, t)|\mu(dx)$ exists in $\mathbb{R}$ for every $t \in G$.

Show that $t \mapsto \int f(x, t)\mu(dx) : G \to \mathbb{R}$ is differentiable. (*Hint*: show first that, for a suitable M, $|f(x, t) - f(x, t')| \le M|g(x)|\|t - t'\|$ for every t, $t' \in G$ and $x \in X$.)

262Y Further exercises (a) Show that if $T = \langle \tau_{ij} \rangle_{i \le s, j \le r}$ is an $s \times r$ matrix then the operator norm $\|T\|$, as defined in 262H, is at most $\sqrt{\sum_{i=1}^{s} \sum_{j=1}^{r} \tau_{ij}^2}$.

(b) Give an example of a measurable function $\phi : \mathbb{R}^2 \to \mathbb{R}$ such that $\operatorname{dom} \dfrac{\partial \phi}{\partial \xi_1}$ is not measurable.

(c) Let $\phi : D \to \mathbb{R}$ be any function, where $D \subseteq \mathbb{R}^r$. Show that $H = \{x : x \in D,\ \phi$ is differentiable relative to its domain at $x\}$ is relatively measurable in D, and that $\dfrac{\partial \phi}{\partial \xi_j} \restriction H$ is measurable for every $j \le r$.

(d) A function $\phi : \mathbb{R}^r \to \mathbb{R}$ is **smooth** if all its partial derivatives $\dfrac{\partial ... \partial \phi}{\partial \xi_i \partial \xi_j ... \partial \xi_l}$ are defined everywhere in $\mathbb{R}^r$ and are continuous. Show that if f is integrable over $\mathbb{R}^r$ and $\phi : \mathbb{R}^r \to \mathbb{R}$ is smooth and has bounded support then the convolution $f * \phi$ is smooth. (*Hint*: 262Xj, 262Xk.)

(e) For $\delta > 0$ set $\tilde{\phi}_\delta(x) = e^{1/(\delta^2 - \|x\|^2)}$ if $\|x\| < \delta$, 0 if $\|x\| \ge \delta$; set $\alpha_\delta = \int \tilde{\phi}_\delta(x)dx$, $\phi_\delta(x) = \alpha_\delta^{-1}\tilde{\phi}_\delta(x)$ for every x. (i) Show that $\phi_\delta : \mathbb{R}^r \to \mathbb{R}$ is smooth and has bounded support. (ii) Show that if f is integrable over $\mathbb{R}^r$ then $\lim_{\delta \downarrow 0} \int |f(x) - (f * \phi_\delta)(x)|dx = 0$. (*Hint*: start with continuous functions f with bounded support, and use 242O.)

(f) Show that if f is integrable over $\mathbb{R}^r$ and $\epsilon > 0$ there is a smooth function h with bounded support such that $\int |f - h| \le \epsilon$. (*Hint*: *either* reduce to the case in which f has bounded support and use 262Ye *or* adapt the method of 242Xi.)

(g) Suppose that f is a real function which is integrable over every bounded subset of $\mathbb{R}^r$. (i) Show that $f \times \phi$ is integrable whenever $\phi : \mathbb{R}^r \to \mathbb{R}$ is a smooth function with bounded support. (ii) Show that if $\int f \times \phi = 0$ for every smooth function with bounded support then $f = 0$ a.e. (*Hint*: show that $\int_{B(x,\delta)} f = 0$ for every $x \in \mathbb{R}^r$ and $\delta > 0$, and use 261C. Alternatively show that $\int_E f = 0$ first for $E = [b, c]$, then for open sets E, then for arbitrary measurable sets E.)

(h) Let f be integrable over $\mathbb{R}^r$, and for $\delta > 0$ let $\phi_\delta : \mathbb{R}^r \to \mathbb{R}$ be the function of 262Ye. Show that $\lim_{\delta\downarrow 0}(f*\phi_\delta)(x) = f(x)$ for every x in the Lebesgue set of f. (*Hint*: 261Ye.)

(i) Let L be the space of all Lipschitz functions from $\mathbb{R}^r$ to $\mathbb{R}^s$ and for $\phi \in L$ set

$$\|\phi\| = \|\phi(0)\| + \inf\{\gamma : \gamma \in [0,\infty[, \ \|\phi(y) - \phi(x)\| \le \gamma\|y - x\| \text{ for every } x,\, y \in \mathbb{R}^r\}.$$

Show that $(L, \|\ \|)$ is a Banach space.

262 Notes and comments The emphasis of this section has turned out to be on the connexions between the concepts of 'Lipschitz function' and 'differentiable function'. It is the delight of classical real analysis that such intimate relationships arise between concepts which belong to different categories. 'Lipschitz functions' clearly belong to the theory of metric spaces (I will return to this in §264), while 'differentiable functions' belong to the theory of differentiable manifolds, which is outside the scope of this volume. I have written this section out carefully just in case there are readers who have so far missed the theory of differentiable mappings between multi-dimensional Euclidean spaces; but it also gives me a chance to work through the notion of 'function differentiable relative to its domain', which will make it possible in the next section to ride smoothly past a variety of problems arising at boundaries. The difficulties I am concerned with arise in the first place with such functions as the polar-coordinate transformation

$$(\rho,\theta) \mapsto (\rho\cos\theta, \rho\sin\theta) : \{(0,0)\} \cup (]0,\infty[\times]{-\pi},\pi]) \to \mathbb{R}^2.$$

In order to make this a bijection we have to do something rather arbitrary, and the domain of the transformation cannot be an open set. On the definitions I am using, this function is differentiable relative to its domain at every point of its domain, and we can apply such results as 262O uninhibitedly. You will observe that in this case the non-interior points of the domain form a negligible set $\{(0,0)\} \cup (]0,\infty[\times \{\pi\})$, so we can expect to be able to ignore them; and for most of the geometrically straightforward transformations that the theory is applied to, judicious excision of negligible sets will reduce problems to the case of honestly differentiable functions with open domains. But while open-domain theory will deal with a large proportion of the most important examples, there is a danger that you would be left with real misapprehensions concerning the scope of these methods.

The essence of differentiability is that a differentiable function ϕ is approximable, near any given point of its domain, by an affine function. The idea of 262M is to describe a widely effective method of dissecting $D = \operatorname{dom}\phi$ into countably many pieces on each of which ϕ is well-behaved. This will be applied in §§263 and 265 to investigate the measure of $\phi[D]$; but we already have several straightforward consequences (262N-262P).

263 Differentiable transformations in $\mathbb{R}^r$

This section is devoted to the proof of a single major theorem (263D) concerning differentiable transformations between subsets of $\mathbb{R}^r$. There will be a generalization of this result in §265, and those with some familiarity with the topic, or sufficient hardihood, may wish to read §264 before taking this section and §265 together. I end with a few simple corollaries and an extension of the main result which can be made in the one-dimensional case (263I).

Throughout this section, as in the rest of the chapter, μ will denote Lebesgue measure on $\mathbb{R}^r$.

263A Linear transformations I begin with the special case of linear operators, which is not only the basis of the proof of 263D, but is also one of its most important applications, and is indeed sufficient for many very striking results.

Theorem Let T be a real $r \times r$ matrix; regard T as a linear operator from $\mathbb{R}^r$ to itself. Let $J = |\det T|$ be the modulus of its determinant. Then

$$\mu T[E] = J\mu E$$

for every measurable set $E \subseteq \mathbb{R}^r$. If T is a bijection (that is, if $J \ne 0$), then

$$\mu F = J\mu T^{-1}[F]$$

for every measurable $F \subseteq \mathbb{R}^r$, and

$$\int_F g\,d\mu = J\int_{T^{-1}[F]} gT\,d\mu$$

for every integrable function g and measurable set F.

proof (a) The first step is to show that $T[I]$ is measurable for every half-open interval $I \subseteq \mathbb{R}^r$. **P** Any non-empty half-open interval $I = [a,b[$ is a countable union of closed intervals $I_n = [a, b - 2^{-n}\mathbf{1}]$, and each I_n is compact (2A2F), so that $T[I_n]$ is compact (2A2Eb), therefore closed (2A2Ec), therefore measurable (115G), and $T[I] = \bigcup_{n\in\mathbb{N}} T[I_n]$ is measurable. **Q**

(b) Set $J^* = \mu T[\,[\mathbf{0},\mathbf{1}[\,]$, where $\mathbf{0} = (0,\dots,0)$ and $\mathbf{1} = (1,\dots,1)$; because $T[\,[\mathbf{0},\mathbf{1}[\,]$ is bounded, $J^* < \infty$. (I will eventually show that $J^* = J$.) It is convenient to deal with the case of singular T first. Recall that T, regarded as a linear transformation from $\mathbb{R}^r$ to itself, is either bijective or onto a proper linear subspace. In the latter case, take any $e \in \mathbb{R}^r \setminus T[\mathbb{R}^r]$; then the sets

$$T[\,[\mathbf{0},\mathbf{1}[\,] + \gamma e,$$

as γ runs over $[0,1]$, are disjoint and all of the same measure J^*, because μ is translation-invariant (134A); moreover, their union is bounded, so has finite outer measure. As there are infinitely many such γ, the common measure J^* must be zero. Now observe that

$$T[\mathbb{R}^r] = \bigcup_{z\in\mathbb{Z}^r} T[\,[\mathbf{0},\mathbf{1}[\,] + Tz,$$

and

$$\mu(T[\,[\mathbf{0},\mathbf{1}[\,] + Tz) = J^* = 0$$

for every $z \in \mathbb{Z}^r$, while $\mathbb{Z}^r$ is countable, so $\mu T[\mathbb{R}^r] = 0$. At the same time, because T is singular, it has zero determinant, and $J = 0$. Accordingly

$$\mu T[E] = 0 = J\mu E$$

for every measurable $E \subseteq \mathbb{R}^r$, and we're done.

(c) Henceforth, therefore, let us assume that T is non-singular. Note that it and its inverse are continuous, so that T is a homeomorphism, and $T[G]$ is open iff G is open.

If $a \in \mathbb{R}^r$ and $k \in \mathbb{N}$, then

$$\mu T[\,\big[a, a + 2^{-k}\mathbf{1}\big[\,] = 2^{-kr}J^*.$$

P Set $J^*_k = \mu T[\,\big[\mathbf{0}, 2^{-k}\mathbf{1}\big[\,]$. Now $T[\,\big[a, a+2^{-k}\mathbf{1}\big[\,] = T[\,\big[\mathbf{0}, 2^{-k}\mathbf{1}\big[\,] + Ta$; because μ is translation-invariant, its measure is also J^*_k. Next, $[\mathbf{0},\mathbf{1}[$ is expressible as a disjoint uion of 2^{kr} sets of the form $\big[a, a+2^{-k}\mathbf{1}\big[$; consequently, $T[\,[\mathbf{0},\mathbf{1}[\,]$ is expressible as a disjoint uion of 2^{kr} sets of the form $T[\,\big[a, a+2^{-k}\mathbf{1}\big[\,]$, and

$$J^* = \mu T[\,[\mathbf{0},\mathbf{1}[\,] = 2^{kr}J^*_k,$$

that is, $J^*_k = 2^{-kr}J^*$, as claimed. **Q**

(d) Consequently $\mu T[G] = J^*\mu G$ for every open set $G \subseteq \mathbb{R}^r$. **P** For each $k \in \mathbb{N}$, set

$$Q_k = \{z : z \in \mathbb{Z}^r,\ \big[2^{-k}z, 2^{-k}z + 2^{-k}\mathbf{1}\big[\subseteq G,$$

$$G_k = \bigcup_{z\in Q_k} \big[2^{-k}z, 2^{-k}z + 2^{-k}\mathbf{1}\big[.$$

Then G_k is a disjoint union of $\#(Q_k)$ sets of the form $\big[2^{-k}z, 2^{-k}z + 2^{-k}\mathbf{1}\big[$, so $\mu G_k = 2^{kr}\#(Q_k)$; also, $T[G_k]$ is a disjoint union of $\#(Q_k)$ sets of the form $T[\,\big[2^{-k}z, 2^{-k}z + 2^{-k}\mathbf{1}\big[\,]$, so has measure $2^{-kr}J^*\#(Q_k) = J^*\mu G_k$, using (c).

Observe next that $\langle G_k\rangle_{k\in\mathbb{N}}$ is a non-decreasing sequence with union G, so that

$$\mu T[G] = \lim_{k\to\infty} \mu T[G_k] = \lim_{k\to\infty} J^*\mu G_k = J^*\mu G. \text{ } \mathbf{Q}$$

(e) It follows that $\mu^*T[A] = J^*\mu^*A$ for every $A \subseteq \mathbb{R}^r$. **P** Given $A \subseteq \mathbb{R}^r$ and $\epsilon > 0$, there are open sets G, H such that $G \supseteq A$, $H \supseteq T[A]$, $\mu G \le \mu^*A + \epsilon$ and $\mu H \le \mu^*T[A] + \epsilon$ (134Fa). Set $G_1 = G \cap T^{-1}[H]$; then G_1 is open because $T^{-1}[H]$ is. Now $\mu T[G_1] = J^*\mu G_1$, so

$$\begin{aligned}\mu^*T[A] &\le \mu T[G_1] = J^*\mu G_1 \le J^*\mu^*A + J^*\epsilon \\ &\le J^*\mu G_1 + J^*\epsilon = \mu T[G_1] + J^*\epsilon \le \mu H + J^*\epsilon \\ &\le \mu^*T[A] + \epsilon + J^*\epsilon.\end{aligned}$$

As ϵ is arbitrary, $\mu^*T[A] = J^*\mu^*A$. **Q**

(f) Consequently $\mu T[E]$ exists and is equal to $J^*\mu E$ for every measurable $E \subseteq \mathbb{R}^r$. **P** Let $E \subseteq \mathbb{R}^r$ be measurable, and take any $A \subseteq \mathbb{R}^r$. Set $A' = T^{-1}[A]$. Then

$$\begin{aligned}\mu^*(A \cap T[E]) + \mu^*(A \setminus T[E]) &= \mu^*(T[A' \cap E]) + \mu^*(T[A' \setminus E]) \\ &= J^*(\mu^*(A' \cap E) + \mu^*(A' \setminus E)) \\ &= J^*\mu^*A' = \mu^*T[A'] = \mu^*A.\end{aligned}$$

As A is arbitrary, $T[E]$ is measurable, and now

$$\mu T[E] = \mu^*T[E] = J^*\mu^*E = J^*\mu E. \ \mathbf{Q}$$

(g) We are at last ready for the calculation of J^*. Recall that the matrix T must be expressible as PDQ, where P and Q are orthogonal matrices and D is diagonal, with non-negative diagonal entries (2A6C). Now we must have

$$T[\,[\mathbf{0},\mathbf{1}[\,] = P[D[Q[\,[\mathbf{0},\mathbf{1}[\,]]],$$

so, using (f),

$$J^* = J^*_P J^*_D J^*_Q,$$

where $J^*_P = \mu P[\,[\mathbf{0},\mathbf{1}[\,]$, etc. Now we find that $J^*_P = J^*_Q = 1$. **P** Let $B = B(\mathbf{0}, 1)$ be the unit ball of $\mathbb{R}^r$. Because B is closed, it is measurable; because it is bounded, $\mu B < \infty$; and because B includes the non-empty half-open interval $\big[\mathbf{0}, r^{-1/2}\mathbf{1}\big[$, $\mu B > 0$. Now $P[B] = Q[B] = B$, because P and Q are orthogonal matrices; so we have

$$\mu B = \mu P[B] = J^*_P \mu B,$$

and J^*_P must be 1; similarly, $J^*_Q = 1$. **Q**

(h) So we have only to calculate J^*_D. Suppose the coefficients of D are $\delta_1, \dots, \delta_r \ge 0$, so that $Dx = (\delta_1\xi_1, \dots, \delta_r\xi_r) = d \times x$. We have been assuming since the beginning of (c) that T is non-singular, so no δ_i can be 0. Accordingly

$$D[\,[\mathbf{0},\mathbf{1}[\,] = [\mathbf{0}, d[,$$

and

$$J^*_D = \mu\,[\mathbf{0}, d[= \textstyle\prod_{i=1}^r \delta_i = \det D.$$

Now because P and Q are orthogonal, both have determinant ± 1, so $\det T = \pm \det D$ and $J^* = \pm \det T$; because J^* is surely non-negative, $J^* = |\det T| = J$.

(i) Thus $\mu T[E] = J\mu E$ for every Lebesgue measurable $E \subseteq \mathbb{R}^r$. If T is non-singular, then we may use the above argument to show that $T^{-1}[F]$ is measurable for every measurable F, and

$$\mu F = \mu T[T^{-1}[F]] = J\mu T^{-1}[F] = \int J \times \chi(T^{-1}[F])\,d\mu,$$

identifying J with the constant function with value J. By 235A,

$$\int_F g\,d\mu = \int_{T^{-1}[F]} JgT\,d\mu = J\int_{T^{-1}[F]} gT\,d\mu$$

for every integrable function g and measurable set F.

263B Remark Perhaps I should have warned you that I should be calling on the results of §235. But if they were fresh in your mind the formulae of the statement of the theorem will have recalled them, and if not then it is perhaps better to turn back to them now rather than before reading the theorem, since they are used only in the last sentence of the proof.

I have taken the argument above at a leisurely, not to say pedestrian, pace. The point is that while the translation-invariance of Lebesgue measure, and its behaviour under simple magnification of a single coordinate, are more or less built into the definition, its behaviour under general rotations is not, since a rotation takes half-open intervals into skew cuboids. Of course the calculation of the measure of such an object is not really anything to do with the Lebesgue theory, and it will be clear that much of the argument would apply equally to any geometrically reasonable notion of r-dimensional volume.

We come now to the central result of the chapter. We have already done some of the detail work in 262M. The next basic element is the following lemma.

263C Lemma Let T be any $r \times r$ matrix; set $J = |\det T|$. Then for any $\epsilon > 0$ there is a $\zeta = \zeta(T, \epsilon) > 0$ such that

(i) $|\det S - \det T| \le \epsilon$ whenever S is an $r \times r$ matrix and $\|S - T\| \le \zeta$;

(ii) whenever $D \subseteq \mathbb{R}^r$ is a bounded set and $\phi : D \to \mathbb{R}^r$ is a function such that $\|\phi(x) - \phi(y) - T(x - y)\| \le \zeta\|x - y\|$ for all x, $y \in D$, then $|\mu^*\phi[D] - J\mu^*D| \le \epsilon\mu^*D$.

proof (a) Of course (i) is the easy part. Because $\det S$ is a continuous function of the coefficients of S, and the coefficients of S must be close to those of T if $\|S-T\|$ is small (262Hb), there is surely a $\zeta_0 > 0$ such that $|\det S-\det T| \le \epsilon$ whenever $\|S - T\| \le \zeta_0$.

(b)(i) Write $B = B(\mathbf{0}, 1)$ for the unit ball of $\mathbb{R}^r$, and consider $T[B]$. We know that $\mu T[B] = J\mu B$ (263A). Let $G \supseteq T[B]$ be an open set such that $\mu G \le (J + \epsilon)\mu B$ (134Fa). Because B is compact (2A2F) so is $T[B]$, so there is a $\zeta_1 > 0$ such that $T[B] + \zeta_1 B \subseteq G$ (2A2Ed). This means that $\mu^*(T[B] + \zeta_1 B) \le (J + \epsilon)\mu B$.

(ii) Now suppose that $D \subseteq \mathbb{R}^r$ is a bounded set, and that $\phi : D \to \mathbb{R}^r$ is a function such that $\|\phi(x) - \phi(y) - T(x - y)\| \le \zeta_1\|x - y\|$ for all x, $y \in D$. Then if $x \in D$ and $\delta > 0$,

$$\phi[D \cap B(x, \delta)] \subseteq \phi(x) + \delta T[B] + \delta\zeta_1 B,$$

because if $y \in D \cap B(x, \delta)$ then $T(y - x) \in \delta T[B]$ and

$$\begin{aligned}\phi(y) &= \phi(x) + T(y - x) + (\phi(y) - \phi(x) - T(y - x))\\ &\in \phi(x) + \delta T[B] + \zeta_1\|y - x\|B \subseteq \phi(x) + \delta T[B] + \zeta_1\delta B.\end{aligned}$$

Accordingly

$$\begin{aligned}\mu^*\phi[D \cap B(x, \delta)] &\le \mu^*(\delta T[B] + \delta\zeta_1 B) = \delta^r\mu^*(T[B] + \zeta_1 B)\\ &\le \delta^r(J + \epsilon)\mu B = (J + \epsilon)\mu B(x, \delta).\end{aligned}$$

Let $\eta > 0$. Then there is a sequence $\langle B_n\rangle_{n\in\mathbb{N}}$ of balls in $\mathbb{R}^r$ such that $D \subseteq \bigcup_{n\in\mathbb{N}} B_n$, $\sum_{n=0}^{\infty} \mu B_n \le \mu^* D + \eta$ and the sum of the measures of those B_n whose centres do not lie in D is at most η (261F). Let K be the set of those n such that the centre of B_n lies in D. Then $\mu^*\phi[D\cap B_n] \le (J+\epsilon)\mu B_n$ for every $n \in K$. Also, of course, ϕ is $(\|T\| + \zeta_1)$-Lipschitz, so $\mu^*\phi[D \cap B_n] \le (\|T\| + \zeta_1)^r\mu B_n$ for $n \in \mathbb{N} \setminus K$ (262D). Now

$$\begin{aligned}\mu^*\phi[D] &\le \sum_{n=0}^{\infty}\mu^*\phi[D \cap B_n]\\ &\le \sum_{n\in K}(J + \epsilon)\mu B_n + \sum_{n\in\mathbb{N}\setminus K}(\|T\| + \zeta_1)^r\mu B_n\\ &\le (J + \epsilon)(\mu^*D + \eta) + \eta(\|T\| + \zeta_1)^r.\end{aligned}$$

As η is arbitrary,

$$\mu^*\phi[D] \le (J + \epsilon)\mu^*D.$$

(c) If $J = 0$, we can stop here, setting $\zeta = \min(\zeta_0, \zeta_1)$; for then we surely have $|\det S - \det T| \le \epsilon$ whenever $\|S - T\| \le \zeta$, while if $\phi : D \to \mathbb{R}^r$ is such that $\|\phi(x) - \phi(y) - T(x - y)\| \le \zeta\|x - y\|$ for all x, $y \in D$, then

$$|\mu^*\phi[D] - J\mu^*D| = \mu^*\phi[D] \le \epsilon\mu^*D.$$

If $J \ne 0$, we have more to do. Because T has non-zero determinant, it has an inverse T^{-1}, and $|\det T^{-1}| = J^{-1}$. As in (b-i) above, there is a $\zeta_2 > 0$ such that $\mu^*(T^{-1}[B] + \zeta_2 B) \le (J^{-1} + \epsilon')\mu B$, where $\epsilon' = \epsilon/J(J + \epsilon)$. Repeating (b), we see that if $C \subseteq \mathbb{R}^r$ is bounded and $\psi : C \to \mathbb{R}^r$ is such that $\|\psi(u) - \psi(v) - T^{-1}(u - v)\| \le \zeta_2\|u - v\|$ for all u, $v \in C$, then $\mu^*\psi[C] \le (J^{-1} + \epsilon')\mu^*C$.

Now suppose that $D \subseteq \mathbb{R}^r$ is bounded and $\phi : D \to \mathbb{R}^r$ is such that $\|\phi(x) - \phi(y) - T(x - y)\| \le \zeta_2'\|x - y\|$ for all x, $y \in D$, where $\zeta_2' = \min(\zeta_2, \|T^{-1}\|)/2\|T^{-1}\|^2 > 0$. Then

$$\|T^{-1}(\phi(x) - \phi(y)) - (x - y)\| \le \|T^{-1}\|\zeta_2'\|x - y\| \le \frac{1}{2}\|x - y\|$$

for all x, $y \in D$, so ϕ must be injective; set $C = \phi[D]$ and $\psi = \phi^{-1} : C \to D$. Note that C is bounded, because

$$\|\phi(x) - \phi(y)\| \le (\|T\| + \zeta_2')\|x - y\|$$

whenever x, $y \in D$. Also

$$\|T^{-1}(u - v) - (\psi(u) - \psi(v))\| \le \|T^{-1}\|\zeta_2'\|\psi(u) - \psi(v)\| \le \frac{1}{2}\|\psi(u) - \psi(v)\|$$

for all u, $v \in C$. But this means that

$$\|\psi(u) - \psi(v)\| - \|T^{-1}\|\|u - v\| \le \frac{1}{2}\|\psi(u) - \psi(v)\|$$

and $\|\psi(u)-\psi(v)\| \le 2\|T^{-1}\|\|u-v\|$ for all u, $v \in C$, so that

$$\|\psi(u)-\psi(v)-T^{-1}(u-v)\| \le 2\zeta_2'\|T^{-1}\|^2\|u-v\| \le \zeta_2\|u-v\|$$

for all u, $v \in C$.

By (b) just above, it follows that

$$\mu^*D = \mu^*\psi[C] \le (J^{-1}+\epsilon')\mu^*C = (J^{-1}+\epsilon')\mu^*\phi[D],$$

and

$$J\mu^*D \le (1+J\epsilon')\mu^*\phi[D].$$

(d) So if we set $\zeta = \min(\zeta_0,\zeta_1,\zeta_2') > 0$, and if $D \subseteq \mathbb{R}^r$, $\phi : D \to \mathbb{R}^r$ are such that D is bounded and $\|\phi(x)-\phi(y)-T(x-y)\| \le \zeta\|x-y\|$ for all x, $y \in D$, we shall have

$$\mu^*\phi[D] \le (J+\epsilon)\mu^*D,$$

$$\mu^*\phi[D] \ge J\mu^*D - J\epsilon'\mu^*\phi[D] \ge J\mu^*D - J\epsilon'(J+\epsilon)\mu^*D = J\mu^*D - \epsilon\mu^*D,$$

so we get the required formula

$$|\mu^*\phi[D] - J\mu^*D| \le \epsilon\mu^*D.$$

263D We are ready for the theorem.

Theorem Let $D \subseteq \mathbb{R}^r$ be any set, and $\phi : D \to \mathbb{R}^r$ a function differentiable relative to its domain at each point of D. For each $x \in D$ let $T(x)$ be a derivative of ϕ relative to D at x, and set $J(x) = |\det T(x)|$. Then

(i) $J : D \to [0,\infty[$ is a measurable function,

(ii) $\mu^*\phi[D] \le \int_D J\,d\mu$,

allowing ∞ as the value of the integral. If D is measurable, then

(iii) $\phi[D]$ is measurable.

If D is measurable and ϕ is injective, then

(iv) $\mu\phi[D] = \int_D J\,d\mu$,

(v) for every real-valued function g defined on a subset of $\phi[D]$,

$$\int_{\phi[D]} g\,d\mu = \int_D J \times g\phi\,d\mu$$

if either integral is defined in $[-\infty,\infty]$, provided we interpret $J(x)g(\phi(x))$ as zero when $J(x) = 0$ and $g(\phi(x))$ is undefined.

proof (a) To see that J is measurable, use 262P; the function $T \mapsto |\det T|$ is a continuous function of the coefficients of T, and the coefficients of $T(x)$ are measurable functions of x, by 262P, so $x \mapsto |\det T(x)|$ is measurable (121K). We also know that if D is measurable, $\phi[D]$ will be measurable, by 262Ob. Thus (i) and (iii) are done.

(b) For the moment, assume that D is bounded, and fix $\epsilon > 0$. For $r \times r$ matrices T, take $\zeta(T,\epsilon) > 0$ as in 263C. Take $\langle D_n\rangle_{n\in\mathbb{N}}$, $\langle T_n\rangle_{n\in\mathbb{N}}$ as in 262M, so that $\langle D_n\rangle_{n\in\mathbb{N}}$ is a disjoint cover of D by sets which are relatively measurable in D, and each T_n is an $r \times r$ matrix such that

$$\|T(x)-T_n\| \le \zeta(T_n,\epsilon) \text{ whenever } x \in D_n,$$

$$\|\phi(x)-\phi(y)-T_n(x-y)\| \le \zeta(T_n,\epsilon)\|x-y\| \text{ for all } x,\ y \in D_n.$$

Then, setting $J_n = |\det T_n|$, we have

$$|J(x)-J_n| \le \epsilon \text{ for every } x \in D_n,$$

$$|\mu^*\phi[D_n] - J_n\mu^*D_n| \le \epsilon\mu^*D_n,$$

by the choice of $\zeta(T_n,\epsilon)$. So we have

$$\int_D J\,d\mu \le \sum_{n=0}^{\infty} J_n\mu^*D_n + \epsilon\mu^*D \le \int_D J\,d\mu + 2\epsilon\mu^*D;$$

I am using here the fact that all the D_n are relatively measurable in D, so that, in particular, $\mu^*D = \sum_{n=0}^{\infty}\mu^*D_n$. Next,

$$\mu^*\phi[D] \le \sum_{n=0}^{\infty}\mu^*\phi[D_n] \le \sum_{n=0}^{\infty} J_n\mu^*D_n + \epsilon\mu^*D.$$

Putting these together,

$$\mu^*\phi[D] \le \int_D J\,d\mu + 2\epsilon\mu^* D.$$

If D is measurable and ϕ is injective, then all the D_n are measurable subsets of $\mathbb{R}^r$, so all the $\phi[D_n]$ are measurable, and they are also disjoint. Accordingly

$$\int_D J\,d\mu \le \sum_{n=0}^{\infty} J_n\mu D_n + \epsilon\mu D \le \sum_{n=0}^{\infty}(\mu\phi[D_n] + \epsilon\mu D_n) + \epsilon\mu D = \mu\phi[D] + 2\epsilon\mu D.$$

Since ϵ is arbitrary, we get

$$\mu^*\phi[D] \le \int_D J\,d\mu,$$

and if D is measurable and ϕ is injective,

$$\int_D J\,d\mu \le \mu\phi[D];$$

thus we have (ii) and (iv), on the assumption that D is bounded.

(c) For a general set D, set $B_k = B(\mathbf{0}, k)$; then

$$\mu^*\phi[D] = \lim_{k\to\infty} \mu^*\phi[D \cap B_k] \le \lim_{k\to\infty} \int_{D\cap B_k} J\,d\mu = \int_D J\,d\mu,$$

with equality if ϕ is injective and D is measurable.

(d) For part (v), I seek to show that the hypotheses of 235J are satisfied, taking $X = D$ and $Y = \phi[D]$. **P** Set $G = \{x : x \in D,\ J(x) > 0\}$.

(α) If $F \subseteq \phi[D]$ is measurable, then there are Borel sets F_1, F_2 such that $F_1 \subseteq F \subseteq F_2$ and $\mu(F_2 \setminus F_1) = 0$. Set $E_j = \phi^{-1}[F_j]$ for each j, so that $E_1 \subseteq \phi^{-1}[F] \subseteq E_2$, and both the sets E_j are measurable, because ϕ and $\operatorname{dom}\phi$ are measurable. Now, applying (iv) to $\phi\restriction E_j$,

$$\int_{E_j} J\,d\mu = \mu\phi[E_j] = \mu(F_j \cap \phi[D]) = \mu F$$

for both j, so $\int_{E_2\setminus E_1} J\,d\mu = 0$ and $J = 0$ a.e. on $E_2\setminus E_1$. Accordingly $J\times\chi(\phi^{-1}[F]) =_{\text{a.e.}} J\times\chi E_1$, and $\int J\times\chi(\phi^{-1}[F])d\mu$ exists and is equal to $\int_{E_1} J\,d\mu = \mu F$. At the same time, $(\phi^{-1}[F]\cap G)\triangle(E_1\cap G)$ is negligible, so $\phi^{-1}[F]\cap G$ is measurable.

(β) If $F \subseteq \phi[D]$ and $G \cap \phi^{-1}[F]$ is measurable, then we know that $\mu\phi[D \setminus G] = \int_{D\setminus G} J = 0$ (by (iv)), so $F \setminus \phi[G]$ must be negligible; while $F \cap \phi[G] = \phi[G \cap \phi^{-1}[F]]$ also is measurable, by (iii). Accordingly F is measurable whenever $G \cap \phi^{-1}[F]$ is measurable.

Thus all the hypotheses of 235J are satisfied. **Q** Now (v) can be read off from the conclusion of 235J.

263E Remarks (a) This is a version of the classical result on change of variable in a many-dimensional integral. What I here call $J(x)$ is the **Jacobian** of ϕ at x; it describes the change in volumes of objects near x, following the rule already established in 263A for functions with constant derivative. The idea of the proof is also the classical one: to break the set D up into small enough pieces D_m for us to be able to approximate ϕ by affine operators $y \mapsto \phi(x) + T_m(y - x)$ on each. The potential irregularity of the set D, which in this theorem may be any set, is compensated for by a corresponding freedom in choosing the sets D_m. In fact there is a further decomposition of the sets D_m hidden in part (b-ii) of the proof of 263C; each D_m is essentially covered by a disjoint family of balls, the measures of whose images we can estimate with an adequate accuracy. There is always a danger of a negligible exceptional set, and we need the crude inequalities of the proof of 262D to deal with it.

(b) Throughout the work of this chapter, from 261B to 263D, I have chosen balls $B(x, \delta)$ as the basic shapes to work with. I think it should be clear that in fact any reasonable shapes would do just as well. In particular, the 'balls'

$$B_1(x, \delta) = \{y : \sum_{i=1}^r |\eta_i - \xi_i| \le \delta\}, \quad B_\infty(x, \delta) = \{y : |\eta_i - \xi_i| \le \delta\ \forall\ i\}$$

would serve perfectly. There are many alternatives. We could use sets of the form $C(x, k)$, for $x \in \mathbb{R}^r$ and $k \in \mathbb{N}$, defined to be the half-open cube of the form $\left[2^{-k}z, 2^{-k}(z + \mathbf{1})\right[$ with $z \in \mathbb{Z}^r$ containing x, instead; or even $C'(x, \delta) = [x, x + \delta\mathbf{1}[$. In all such cases we have versions of the density theorems (261Yb-261Yc) which support the remaining theory.

(c) I have presented 263D as a theorem about differentiable functions, because that is the normal form in which one uses it in elementary applications. However, the proof depends essentially on the fact that a differentiable function is a countable union of Lipschitz functions, and 263D would follow at once from the same theorem proved for Lipschitz functions only. Now the fact is that the theorem applies to *any* countable union of Lipschitz functions, because a Lipschitz function is differentiable almost everywhere. For more advanced work (see Federer 69 or Evans & Gariepy 92, or Chapter 47 in Volume 4) it seems clear that Lipschitz functions are the vital ones, so I spell out the result.

***263F Corollary** Let $D \subseteq \mathbb{R}^r$ be any set and $\phi : D \to \mathbb{R}^r$ a Lipschitz function. Let D_1 be the set of points at which ϕ has a derivative relative to D, and for each $x \in D_1$ let $T(x)$ be such a derivative, with $J(x) = |\det T(x)|$. Then

(i) $D \setminus D_1$ is negligible;

(ii) $J : D_1 \to [0, \infty[$ is measurable;

(iii) $\mu^*\phi[D] \le \int_D J(x)dx$.

If D is measurable, then

(iv) $\phi[D]$ is measurable.

If D is measurable and ϕ is injective, then

(v) $\mu\phi[D] = \int_D J\,d\mu$,

(vi) for every real-valued function g defined on a subset of $\phi[D]$,

$$\int_{\phi[D]} g\,d\mu = \int_D J \times g\phi\,d\mu$$

if either integral is defined in $[-\infty, \infty]$, provided we interpret $J(x)g(\phi(x))$ as zero when $J(x) = 0$ and $g(\phi(x))$ is undefined.

proof This is now just a matter of putting 262Q and 263D together, with a little help from 262D. Use 262Q to show that $D \setminus D_1$ is negligible, 262D to show that $\phi[D \setminus D_1]$ is negligible, and apply 263D to $\phi\restriction D_1$.

263G Polar coordinates in the plane I offer an elementary example with a useful consequence. Define $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ by setting $\phi(\rho, \theta) = (\rho\cos\theta, \rho\sin\theta)$ for ρ, $\theta \in \mathbb{R}^2$. Then $\phi'(\rho, \theta) = \begin{pmatrix} \cos\theta & -\rho\sin\theta \\ \sin\theta & \rho\cos\theta \end{pmatrix}$, so $J(\rho, \theta) = |\rho|$ for all ρ, θ. Of course ϕ is not injective, but if we restrict it to the domain $D = \{(0, 0)\} \cup \{(\rho, \theta) : \rho > 0,\ -\pi < \theta \le \pi\}$ then $\phi\restriction D$ is a bijection between D and $\mathbb{R}^2$, and

$$\int g\,d\xi_1 d\xi_2 = \int_D g(\phi(\rho, \theta))\rho\,d\rho d\theta$$

for every real-valued function g which is integrable over $\mathbb{R}^2$.

Suppose, in particular, that we set

$$g(x) = e^{-\|x\|^2/2} = e^{-\xi_1^2/2}e^{-\xi_2^2/2}$$

for $x = (\xi_1, \xi_2) \in \mathbb{R}$. Then

$$\int g(x)dx = \int e^{-\xi_1^2/2}d\xi_1 \int e^{-\xi_2^2/2}d\xi_2,$$

as in 253D. Setting $I = \int e^{-t^2/2}dt$, we have $\int g = I^2$. (To see that I is well-defined in $\mathbb{R}$, note that the integrand is continuous, therefore measurable, and that

$$\int_{-1}^{1} e^{-t^2/2}dt \le 2,$$

$$\int_{-\infty}^{-1} e^{-t^2/2}dt = \int_1^\infty e^{-t^2/2}dt \le \int_1^\infty e^{-t/2}dt = \lim_{a\to\infty} \int_1^a e^{-t/2}dt = \frac{1}{2}e^{-1/2}$$

are both finite.) Now looking at the alternative expression we have

$$\begin{aligned} I^2 &= \int g(x)dx = \int_D g(\rho\cos\theta, \rho\sin\theta)\rho\,d(\rho, \theta) \\ &= \int_D e^{-\rho^2/2}\rho\,d(\rho, \theta) = \int_0^\infty \int_{-\pi}^{\pi} \rho e^{-\rho^2/2}d\theta d\rho \end{aligned}$$

(ignoring the point $(0, 0)$, which has zero measure)

$$\begin{aligned} &= \int_0^\infty 2\pi\rho e^{-\rho^2/2}d\rho = 2\pi \lim_{a\to\infty} \int_0^a \rho e^{-\rho^2/2}d\rho \\ &= 2\pi \lim_{a\to\infty} (-e^{a^2/2} + 1) = 2\pi. \end{aligned}$$

Consequently

$$\int_{-\infty}^{\infty} e^{-t^2/2}dt = I = \sqrt{2\pi},$$

which is one of the many facts every mathematician should know, and in particular is vital for Chapter 27 below.

263H Corollary If $k \in \mathbb{N}$ is odd,

$$\int_{-\infty}^{\infty} x^k e^{-x^2/2}dx = 0;$$

if $k = 2l \in \mathbb{N}$ is even, then

$$\int_{-\infty}^{\infty} x^k e^{-x^2/2}dx = \frac{(2l)!}{2^l l!}\sqrt{2\pi}.$$

proof (a) To see that all the integrals are well-defined and finite, observe that $\lim_{x\to\pm\infty} x^k e^{-x^2/4} = 0$, so that $M_k = \sup_{x\in\mathbb{R}} |x^k e^{-x^2/4}|$ is finite, and

$$\int_{-\infty}^{\infty} |x^k e^{-x^2/2}|dx \le M_k \int_{-\infty}^{\infty} e^{-x^2/4}dx < \infty.$$

(b) If k is odd, then substituting $y = -x$ we get

$$\int_{-\infty}^{\infty} x^k e^{-x^2/2}dx = -\int_{-\infty}^{\infty} y^k e^{-y^2/2}dy,$$

so that both integrals must be zero.

(c) For even k, proceed by induction. Set $I_l = \int_{-\infty}^{\infty} x^{2l}e^{-x^2/2}dx$. $I_0 = \sqrt{2\pi} = \frac{0!}{2^0 0!}\sqrt{2\pi}$ by 263G. For the inductive step to $l + 1 \ge 1$, integrate by parts to see that

$$\int_{-a}^{a} x^{2l+1} \cdot xe^{-x^2/2}dx = -a^{2l+1}e^{-a^2/2} + (-a)^{2l+1}e^{-a^2/2} + \int_{-a}^{a}(2l+1)x^{2l}e^{-x^2/2}dx$$

for every $a \ge 0$. Letting $a \to \infty$,

$$I_{l+1} = (2l + 1)I_l.$$

Because

$$\frac{(2(l+1))!}{2^{l+1}(l+1)!}\sqrt{2\pi} = (2l + 1)\frac{(2l)!}{2^l l!}\sqrt{2\pi},$$

the induction proceeds.

263I The one-dimensional case The restriction to injective functions ϕ in 263D(v) is unavoidable in the context of the result there. But in the substitutions of elementary calculus it is not always essential. In the hope of clarifying the position I give a result here which covers many of the standard tricks.

Theorem Let $I \subseteq \mathbb{R}$ be an interval with more than one point, and $\phi : I \to \mathbb{R}$ a function which is absolutely continuous on any closed bounded subinterval of I. Write $u = \inf I$, $u' = \sup I$ in $[-\infty, \infty]$, and suppose that $v = \lim_{x\downarrow u} \phi(x)$ and $v' = \lim_{x\uparrow u'} \phi(x)$ are defined in $[-\infty, \infty]$. Let g be a Lebesgue measurable real-valued function defined almost everywhere in $\phi[I]$. Then

$$\int_v^{v'} g = \int_I g(\phi(x))\phi'(x)dx$$

whenever the right-hand side is defined in $\mathbb{R}$, on the understanding that we interpret $\int_v^{v'} g$ as $-\int_{v'}^{v} g$ when $v' < v$, and $g(\phi(x))\phi'(x)$ as 0 when $\phi'(x) = 0$ and $g(\phi(x))$ is undefined.

proof (a) Recall that ϕ is differentiable almost everywhere on I (225Cb) and that $\phi[A]$ is negligible for every negligible $A \subseteq I$ (225G). (These results are stated for closed bounded intervals; but since any interval is expressible as the union of a sequence of closed bounded intervals, they remain valid in the present context.) Set $D = \operatorname{dom}\phi'$, so that $I \setminus D$ and $\phi[I \setminus D]$ are negligible. Next, setting $D_0 = \{x : x \in D, \phi'(x) = 0\}$, D and D_0 are Borel sets (225J) and $\phi[D_0]$ is negligible, by 263D(ii), while $\int_I g(\phi(x))\phi'(x)dx = \int_{D\setminus D_0} g(\phi(x))\phi'(x)dx$.

Applying 262M with $A = \mathbb{R}\setminus\{0\}$ and $\zeta(\alpha) = \frac{1}{2}|\alpha|$ for $\alpha \in A$, we have sequences $\langle E_n\rangle_{n\in\mathbb{N}}$, $\langle\alpha_n\rangle_{n\in\mathbb{N}}$ such that $\langle E_n\rangle_{n\in\mathbb{N}}$ is a partition of $D \setminus D_0$ into measurable sets, every α_n is non-zero, and $|\phi(x) - \phi(y) - \alpha_n(x - y)| \le \frac{1}{2}|\alpha_n||x - y|$ for all x, $y \in E_n$; so that, in particular, $\phi\restriction E_n$ is injective, while $\operatorname{sgn}\phi'(x) = \operatorname{sgn}\alpha_n$ for every $x \in E_n$, writing $\operatorname{sgn}\alpha = \alpha/|\alpha|$ as usual. Set $\epsilon_n = \operatorname{sgn}\alpha_n$ for each n. Now 263D(v) tells us that

$$\sum_{n=0}^{\infty} \int |g| \times \chi(\phi[E_n]) = \sum_{n=0}^{\infty} \int_{E_n} |g(\phi(x))\phi'(x)|dx$$

is finite.

Note that 263D(v) also shows that if $B \subseteq \mathbb{R}$ is negligible, then $E_n \cap \phi^{-1}[B]$ must be negligible for every n, so that $\int_{\phi^{-1}[B]} g(\phi(x))\phi'(x)dx = 0$.

Consequently, setting

$$C_0 = \{y : y \in (\phi[I] \cap \operatorname{dom} g) \setminus (\{v, v'\} \cup \phi[I \setminus D] \cup \phi[D_0]),\ \textstyle\sum_{n=0}^\infty |g(y)\chi(\phi[E_n])(y)| < \infty\},$$

$\phi[I] \setminus C_0$ is negligible, and if we set $C = \{y : y \in C_0,\ g(y) \neq 0\}$,

$$\textstyle\int_{J\cap C} g = \int_J g$$

for every $J \subseteq \phi[I]$.

(b) The point of the argument is the following fact: if $y \in C$ then

$$\begin{aligned}\sum_{n=0}^\infty \epsilon_n \chi(\phi[E_n])(y) &= 1 \text{ if } v < y < v',\\ &= -1 \text{ if } v' < y < v,\\ &= 0 \text{ if } y < v \leq v' \text{ or } v' \leq v < y.\end{aligned}$$

P Because $g(y) \neq 0$ and $\sum_{n=0}^\infty |g(y)\chi(\phi[E_n])(y)|$ is finite, $\{n : y \in \phi[E_n]\}$ is finite; because $y \notin \phi[I \setminus D] \cup \phi[D_0]$, and $\phi{\restriction}E_n$ is injective for every n, and $\bigcup_{n\in\mathbb{N}} E_n = D \setminus D_0$, $K = \phi^{-1}[\{y\}]$ is finite. For each $x \in K$, let n_x be such that $x \in E_{n_x}$; then $\epsilon_{n_x} = \operatorname{sgn}\phi'(x)$. So $\sum_{n=0}^\infty \epsilon_n\chi(\phi[E_n])(y) = \sum_{x\in K} \operatorname{sgn}\phi'(x)$.

If $J \subseteq \mathbb{R} \setminus K$ is an interval, $\phi(z) \neq y$ for $z \in J$; since ϕ is continuous, the Intermediate Value Theorem tells us that $\operatorname{sgn}(\phi(z) - y)$ is constant on J. A simple induction on $\#(K \cap {]-\infty, z[})$ shows that $\operatorname{sgn}(\phi(z) - y) = \operatorname{sgn}(v - y) + 2\sum_{x\in K, x<z} \operatorname{sgn}\phi'(x)$ for every $z \in \mathbb{R} \setminus K$; taking the limit as $z \uparrow u'$, $\sum_{x\in K} \operatorname{sgn}\phi'(x) = \frac{1}{2}(\operatorname{sgn}(v' - y) - \operatorname{sgn}(v - y))$. (Here we may have to interpret $\operatorname{sgn}(\pm\infty)$ as ±1 in the obvious way.) This turns out to be just what we need to know. **Q**

(c) What this means is that

$$\int_v^{v'} g = \int_v^{v'} g \times \chi C = \int_C g \times \sum_{n=0}^\infty \epsilon_n \chi(\phi[E_n])$$

(allowing for the conventional reversal if $v' < v$)

$$= \sum_{n=0}^\infty \epsilon_n \int_C g \times \chi(\phi[E_n]) = \sum_{n=0}^\infty \epsilon_n \int_{E_n \cap \phi^{-1}[C]} g(\phi(x))|\phi'(x)|dx$$

(263D(v) again, applied to $\phi{\restriction}(E_n \cap \phi^{-1}[C])$ for each n)

$$= \sum_{n=0}^\infty \epsilon_n \int_{E_n} g(\phi(x))|\phi'(x)|dx$$

(because $g(\phi(x))\phi'(x) = 0$ almost everywhere in $E_n \setminus \phi^{-1}[C]$)

$$= \sum_{n=0}^\infty \int_{E_n} g(\phi(x))\phi'(x)dx = \int_{D\setminus D_0} g(\phi(x))\phi'(x)dx = \int_I g(\phi(x))\phi'(x)dx,$$

as claimed.

263X Basic exercises (a) Let (X, Σ, μ) be any measure space, $f \in \mathcal{L}^0(\mu)$ and $p \in [1, \infty[$. Show that $f \in \mathcal{L}^p(\mu)$ iff

$$\gamma = p\textstyle\int_0^\infty \alpha^{p-1}\mu^*\{x : x \in \operatorname{dom} f,\ |f(x)| > \alpha\}d\alpha$$

is finite, and in this case $\|f\|_p = \gamma^{1/p}$. (*Hint*: $\int |f|^p = \int_0^\infty \mu^*\{x : |f(x)|^p > \beta\}d\beta$, by 252O; now substitute $\beta = \alpha^p$.)

(b) Let f be an integrable function defined almost everywhere in $\mathbb{R}^r$. Show that if $\alpha < r - 1$ then $\sum_{n=1}^\infty n^\alpha|f(nx)|$ is finite for almost every $x \in \mathbb{R}^r$. (*Hint*: estimate $\sum_{n=0}^\infty n^\alpha \int_B |f(nx)|dx$ for any ball B centered at the origin.)

(c) Let $A \subseteq {]0, 1[}$ be a set such that $\mu^*A = \mu^*([0, 1] \setminus A) = 1$, where μ is Lebesgue measure on $\mathbb{R}$. Set $D = A \cup \{-x : x \in {]0, 1[} \setminus A\} \subseteq [-1, 1]$, and set $\phi(x) = |x|$ for $x \in D$. Show that ϕ is injective, that ϕ is differentiable relative to its domain everywhere in D, and that $\mu^*\phi[D] < \int_D |\phi'(x)|dx$.

(d) Let $\phi : D \to \mathbb{R}^r$ be a function differentiable relative to D at each point of $D \subseteq \mathbb{R}^r$, and suppose that for each $x \in D$ there is a non-singular derivative $T(x)$ of ϕ at x; set $J(x) = |\det T(x)|$. Show that D is expressible as $\bigcup_{k\in\mathbb{N}} D_k$ where $D_k = D \cap \overline{D}_k$ and $\phi{\restriction}D_k$ is injective for each k.

>(e)(i) Show that for any Lebesgue measurable $E \subseteq \mathbb{R}$, $t \in \mathbb{R}\setminus\{0\}$, $\int_{tE} \frac{1}{|u|}du = \int_E \frac{1}{|u|}du$. (ii) For $t \in \mathbb{R}$, $u \in \mathbb{R}\setminus\{0\}$ set $\phi(t,u) = (\frac{t}{u}, u)$. Show that $\int_{\phi[E]} \frac{1}{|tu|}d(t,u) = \int_E \frac{1}{|tu|}d(t,u)$ for any Lebesgue measurable $E \subseteq \mathbb{R}^2$.

>(f) Define $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ by setting

$$\phi(\rho,\theta,\alpha) = (\rho\sin\theta\sin\alpha, \rho\sin\theta\cos\alpha, \rho\cos\theta).$$

Show that $\det\phi'(\rho,\theta,\alpha) = \rho^2\sin\theta$.

(g) Show that if $k = 2l+1$ is odd, then $\int_0^\infty x^k e^{-x^2/2}dx = 2^l l!$. (Compare 252Xi.)

263Y Further exercises (a) Define a measure ν on $\mathbb{R}$ by setting $\nu E = \int_E \frac{1}{|x|}dx$ for Lebesgue measurable sets $E \subseteq \mathbb{R}$. For f, $g \in \mathcal{L}^1(\nu)$ set $(f*g)(x) = \int f(\frac{x}{t})g(t)\nu(dt)$ whenever this is defined in $\mathbb{R}$. (i) Show that $f*g = g*f \in \mathcal{L}^1(\nu)$. (ii) Show that $\int h(x)(f*g)(x)\nu(dx) = \int h(xy)f(x)f(y)\nu(dx)\nu(dy)$ for every $h \in \mathcal{L}^\infty(\nu)$. (iii) Show that $f*(g*h) = (f*g)*h$ for every $h \in \mathcal{L}^1(\nu)$. (*Hint*: 263Xe.)

(b) Let $E \subseteq \mathbb{R}^2$ be a measurable set such that $\limsup_{\alpha\to\infty} \frac{1}{\alpha^2}\mu_2(E \cap B(\mathbf{0},\alpha)) > 0$, writing μ_2 for Lebesgue measure on $\mathbb{R}^2$. Show that there is some $\theta \in \,]-\pi,\pi]$ such that $\mu_1 E_\theta = \infty$, where $E_\theta = \{\rho : \rho \ge 0,\ (\rho\cos\theta, \rho\sin\theta) \in E\}$. (*Hint*: show that $\frac{1}{\alpha^2}\mu_2(E \cap B(\mathbf{0},\alpha)) \le \int_{-\pi}^{\pi} \min(\frac{1}{2}, \frac{1}{\alpha}\mu_1 E_\theta)d\theta$.) Generalize to higher dimensions and to functions other than χE.

(c) Let $E \subseteq \mathbb{R}^r$ be a measurable set, and $\phi : E \to \mathbb{R}^r$ a function differentiable relative to its domain, with a derivative $T(x)$, at each point x of E; set $J(x) = |\det T(x)|$. Show that for any integrable function g defined on $\phi[E]$,

$$\int g(y)\#(\phi^{-1}[\{y\}])dy = \int_E J(x)g(\phi(x))dx$$

(*Hint*: 263I.)

(d) Find a proof of 263I based on the ideas of §225. (*Hint*: 225Xg.)

(e) Let $f : [a,b] \to \mathbb{R}$ be a function of bounded variation, where $a < b$ in $\mathbb{R}$, with Lebesgue decomposition $f = f_p + f_{cs} + f_{ac}$ as in 226Cd; let μ be Lebesgue measure on $\mathbb{R}$. Show that the following are equiveridical: (i) f_{cs} is constant; (ii) $\mu f[\,[a,b]\,] \le \int_a^b |f'|d\mu$; (iii) $\mu^* f[A] \le \int_A |f'|d\mu$ for every $A \subseteq [a,b]$; (iv) $f[A]$ is negligible for every negligible set $A \subseteq [a,b]$. (*Hint*: for (iv)$\Rightarrow$(i) put 226Yd and 263D(ii) together to show that $|f(d) - f(c)| \le \int_c^d |f'|d\mu + \mathrm{Var}_{[c,d]} f_p$ whenever $a \le c \le d \le b$, and therefore that $\mathrm{Var}_{[a,b]} f \le \mathrm{Var}_{[a,b]} f_p + \mathrm{Var}_{[a,b]} f_{ac}$.)

263 Notes and comments Yet again, approaching 263D, I find myself having to choose between giving an accessible, relatively weak result and making the extra effort to set out a theorem which is somewhere near the natural boundary of what is achievable within the concepts being developed in this volume; and, as usual, I go for the more powerful form. There are three basic sources of difficulty: (i) the fact that we are dealing with more than one dimension; (ii) the fact that we are dealing with irregular domains; (iii) the fact that we are dealing with arbitrary integrable functions. I do not think I need to apologise for (iii) in a book on measure theory. Concerning (ii), it is quite true that the principal applications of these results are to cases in which the transformation ϕ is differentiable everywhere, with continuous derivative, and the set D has negligible boundary; and in these cases there are substantial simplifications available – mostly because the sets D_m of the proof of 263D can be taken to be cubes. Nevertheless, I think any form of the result which makes such assumptions is deeply unsatisfactory at this level, being an awkward compromise between ideas natural to the Riemann integral and those natural to the Lebesgue integral. Concerning (i), it might even have been right to lay out the whole argument for the case $r = 1$ before proceeding to the general case, as I did in §§114-115, because the one-dimensional case is already important and interesting; and if you find the work above difficult – which it is – and your immediate interests are in one-dimensional integration by substitution, then I think you might find it worth your time to reproduce the $r = 1$ argument yourself, up to a proof of 263I. In fact the biggest difference is in 263A, which becomes nearly trivial; the work of 262M and 263C becomes more readable, because all the matrices turn into scalars and we can drop the word ‘determinant’, but I do not think we can dispense with any of the ideas, at least if we wish to obtain 263D as stated. (But see 263Yd.)

I found myself insisting, in the last paragraph, that a distinction can be made between 'ideas natural to the Riemann integral and those natural to the Lebesgue integral'. We are approaching deep questions here, like 'what are books on measure theory for?', which I do not think can be answered without some – possibly unconscious – reference to the question 'what is mathematics for?'. I do of course want to present here some of the wonderful general theorems which arise in the Lebesgue theory. But more important than any specific theorem is a general idea of what can be proved by these methods. It is the essence of modern measure theory that continuity does not matter, or, if you prefer, that measurable functions are in some sense so nearly continuous that we do not have to add hypotheses of continuity in our theorems. Now this is in a sense a great liberation, and the Lebesgue integral is now the standard one. But you must not regard the Riemann integral as outdated. The intuitions on which it is founded – for instance, that the surface of a solid body has zero volume – remain of great value in their proper context, which certainly includes the study of differentiable functions with continuous derivatives. What I am saying here is that I believe we can use these intuitions best if we maintain a division, a flexible and permeable one, of course, between the ideas of the two theories; and that when transferring a theorem from one side of the boundary to the other we should do so whole-heartedly, seeking to express the full power of the methods we are using.

I have already said that the essential difference between the one-dimensional and multi-dimensional cases lies in 263A, where the Jacobian $J = |\det T|$ enters the argument. Shorn of the technical devices necessary to deal with arbitrary Lebesgue measurable sets, this amounts to a calculation of the volume of the parallelepiped $T[I]$ where I is the interval $[\mathbf{0},\mathbf{1}[$. I have dealt with this by a little bit of algebra, saying that the result is essentially obvious if T is diagonal, whereas if T is an isometry it follows from the fact that the unit ball is left invariant; and the algebra comes in to express an arbitrary matrix as a product of diagonal and orthogonal matrices (2A6C). It is also plain from 261F that Lebesgue measure must be rotation-invariant as well as translation-invariant; that is to say, it is invariant under all isometries. Another way of looking at this will appear in the next section.

I feel myself that the centre of the argument for 263D is in the lemma 263C. This is where we turn the exact result for linear operators into an approximate result for almost-linear functions; and the whole point of differentiability is that a differentiable function is well approximated, in a neighbourhood of any point of its domain, by a linear operator. The lemma involves two rather different ideas. To show that $\mu^*\phi[D] \le (J+\epsilon)\mu^*D$, we look first at balls and then use Vitali's theorem to see that D is economically covered by balls, so that an upper bound for $\mu^*\phi[D]$ in terms of a sum $\sum_{B\in\mathcal{I}_0}\mu^*\phi[D\cap B]$ is adequate. To obtain a lower bound, we need to reverse the argument by looking at $\psi = \phi^{-1}$, which involves checking first that ϕ is invertible, and then that ψ is appropriately linked to T^{-1}. I have written out exact formulae for ϵ', ζ_2' and so on, but this is only in case you do not trust your intuition; the fact that $\|\phi^{-1}(u)-\phi^{-1}(v)-T^{-1}(u-v)\|$ is small compared with $\|u-v\|$ is pretty clearly a consequence of the hypothesis that $\|\phi(x)-\phi(y)-T(x-y)\|$ is small compared with $\|x-y\|$.

The argument of 263D itself is now a matter of breaking the set D up into appropriate pieces on each of which ϕ is sufficiently nearly linear for 263C to apply, so that

$$\mu^*\phi[D] \le \sum_{m=0}^{\infty}\mu^*\phi[D_m] \le \sum_{m=0}^{\infty}(J_m+\epsilon)\mu^*D_m.$$

With a little care (taken in 263C, with its condition (i)), we can also ensure that the Jacobian J is well approximated by J_m almost everywhere in D_m, so that $\sum_{m=0}^{\infty}J_m\mu^*D_m \simeq \int_D J(x)dx$.

These ideas, joined with the results of §262, bring us to the point

$$\int_E J\,d\mu = \mu\phi[E]$$

when ϕ is injective and $E \subseteq D$ is measurable. We need a final trick, involving Borel sets, to translate this into

$$\int_{\phi^{-1}[F]} J\,d\mu = \mu F$$

whenever $F \subseteq \phi[D]$ is measurable, which is what is needed for the application of 235J.

I hope that you long ago saw, and were delighted by, the device in 263G. Once again, this is not really Lebesgue integration; but I include it just to show that the machinery of this chapter can be turned to deal with the classical results, and that indeed we have a tiny profit from our labour, in that no apology need be made for the boundary of the set D into which the polar coordinate system maps the plane. I have already given the actual result as an exercise in 252Xi. That involved (if you chase through the references) a one-dimensional substitution (performed in 225Xj), Fubini's theorem and an application of the formulae of §235; that is to say, very much the same elements as those used above, though in a different order. I could present this with no mention of differentiation in higher dimensions because the first change of variable was in one dimension, and the second (involving the function $x \mapsto \|x\|$, in 252Xi(i)) was of a particularly simple type, so that a different method could be used to find the function J.

The abstract ideas to which this treatise is devoted do not, indeed, lead us to many particular examples on which to practise the ideas of this section. The ones which do arise tend to be very straightforward, as in 263G, 263Xa-263Xb and 263Xe. I mention the last because it provides a formula needed to discuss a new type of convolution (263Ya).

In effect, this depends on the multiplicative group $\mathbb{R} \setminus \{0\}$ in place of the additive group $\mathbb{R}$ treated in §255. The formula $\frac{1}{x}$ in the definition of ν is of course the derivative of $\ln x$, and ln is an isomorphism between $(]0, \infty[, \cdot, \nu)$ and $(\mathbb{R}, +, \text{Lebesgue measure})$.

264 Hausdorff measures

The next topic I wish to approach is the question of 'surface measure'; a useful example to bear in mind throughout this section and the next is the notion of area for regions on the sphere, but any other smoothly curved two-dimensional surface in three-dimensional space will serve equally well. It is I think more than plausible that our intuitive concepts of 'area' for such surfaces should correspond to appropriate measures. But formalizing this intuition is non-trivial, especially if we seek the generality that simple geometric ideas lead us to; I mean, not contenting ourselves with arguments that depend on the special nature of the sphere, for instance, to describe spherical surface area. I divide the problem into two parts. In this section I will describe a construction which enables us to define the r-dimensional measure of an r-dimensional surface – among other things – in s-dimensional space. In the next section I will set out the basic theorems making it possible to calculate these measures effectively in the leading cases.

264A Definitions Let $s \ge 1$ be an integer, and $r > 0$. (I am primarily concerned with integral r, but will not insist on this until it becomes necessary, since there are some very interesting ideas which involve non-integral 'dimension' r.) For any $A \subseteq \mathbb{R}^s$, $\delta > 0$ set

$$\theta_{r\delta}A = \inf\{\sum_{n=0}^{\infty}(\operatorname{diam} A_n)^r : \langle A_n\rangle_{n\in\mathbb{N}} \text{ is a sequence of subsets of } \mathbb{R}^s \text{ covering } A,$$

$$\operatorname{diam} A_n \le \delta \text{ for every } n \in \mathbb{N}\}.$$

It is convenient in this context to say that $\operatorname{diam}\emptyset = 0$. Now set

$$\theta_r A = \sup_{\delta>0} \theta_{r\delta}A;$$

θ_r is **r-dimensional Hausdorff outer measure** on $\mathbb{R}^s$.

264B Of course we must immediately check the following:

Lemma θ_r, as defined in 264A, is always an outer measure.

proof You should be used to these arguments by now, but there is an extra step in this one, so I spell out the details.

(a) Interpreting the diameter of the empty set as 0, we have $\theta_{r\delta}\emptyset = 0$ for every $\delta > 0$, so $\theta_r\emptyset = 0$.

(b) If $A \subseteq B \subseteq \mathbb{R}^s$, then every sequence covering B also covers A, so $\theta_{r\delta}A \le \theta_{r\delta}B$ for every δ and $\theta_r A \le \theta_r B$.

(c) Let $\langle A_n\rangle_{n\in\mathbb{N}}$ be a sequence of subsets of $\mathbb{R}^s$ with union A, and take any $a < \theta_r A$. Then there is a $\delta > 0$ such that $a \le \theta_{r\delta}A$. Now $\theta_{r\delta}A \le \sum_{n=0}^{\infty}\theta_{r\delta}(A_n)$. **P** Let $\epsilon > 0$, and for each $n \in \mathbb{N}$ choose a sequence $\langle A_{nm}\rangle_{m\in\mathbb{N}}$ of sets, covering A, with $\operatorname{diam} A_{nm} \le \delta$ for every m and $\sum_{m=0}^{\infty}(\operatorname{diam} A_{nm})^r \le \theta_{r\delta} + 2^{-n}\epsilon$. Then $\langle A_{nm}\rangle_{m,n\in\mathbb{N}}$ is a cover of A by countably many sets of diameter at most δ, so

$$\theta_{r\delta}A \le \sum_{n=0}^{\infty}\sum_{m=0}^{\infty}(\operatorname{diam} A_{nm})^r \le \sum_{n=0}^{\infty}\theta_{r\delta}A_n + 2^{-n}\epsilon = 2\epsilon + \sum_{n=0}^{\infty}\theta_{r\delta}A_n.$$

As ϵ is arbitrary, we have the result. **Q**

Accordingly

$$a \le \theta_{r\delta}A \le \sum_{n=0}^{\infty}\theta_{r\delta}A_n \le \sum_{n=0}^{\infty}\theta_r A_n.$$

As a is arbitrary,

$$\theta_r A \le \sum_{n=0}^{\infty}\theta_r A_n;$$

as $\langle A_n\rangle_{n\in\mathbb{N}}$ is arbitrary, θ_r is an outer measure.

264C Definition If $s \ge 1$ is an integer, and $r > 0$, then **Hausdorff r-dimensional measure** on $\mathbb{R}^s$ is the measure μ_{Hr} on $\mathbb{R}^s$ defined by Carathéodory's method from the outer measure θ_r of 264A-264B.

264D Remarks (a) It is important to note that the sets used in the definition of the $\theta_{r\delta}$ need not be balls; even in $\mathbb{R}^2$ not every set A can be covered by a ball of the same diameter as A.

(b) In the definitions above I require $r > 0$. It is sometimes appropriate to take μ_{H0} to be counting measure. This is nearly the result of applying the formulae above with $r = 0$, but there can be difficulties if we interpret them over-literally.

(c) All Hausdorff measures must be complete, because they are defined by Carathéodory's method (212A). For $r > 0$, they are atomless (264Yg). In terms of the other criteria of §211, however, they are very ill-behaved; for instance, if r, s are integers and $1 \le r < s$, then μ_{Hr} on $\mathbb{R}^s$ is not semi-finite. (I will give a proof of this in 439H in Volume 4.) Nevertheless, they do have some striking properties which make them reasonably tractable.

(d) In 264A, note that $\theta_{r\delta}A \le \theta_{r\delta'}A$ when $0 < \delta' \le \delta$; consequently, for instance, $\theta_r A = \lim_{n\to\infty} \theta_{r,2^{-n}}A$. I have allowed arbitrary sets A_n in the covers, but it makes no difference if we restrict our attention to covers consisting of open sets or of closed sets (264Xc).

264E Theorem Let $s \ge 1$ be an integer, and $r \ge 0$; let μ_{Hr} be Hausdorff r-dimensional measure on $\mathbb{R}^s$, and Σ_{Hr} its domain. Then every Borel subset of $\mathbb{R}^s$ belongs to Σ_{Hr}.

proof This is trivial if $r = 0$; so suppose henceforth that $r > 0$.

(a) The first step is to note that if A, B are subsets of $\mathbb{R}^s$ and $\eta > 0$ is such that $\|x-y\| \ge \eta$ for all $x \in A$, $y \in B$, then $\theta_r(A\cup B) = \theta_r A+\theta_r B$, where θ_r is r-dimensional Hausdorff outer measure on $\mathbb{R}^s$. **P** Of course $\theta_r(A\cup B) \le \theta_r A+\theta_r B$, because θ_r is an outer measure. For the reverse inequality, we may suppose that $\theta_r(A \cup B) < \infty$, so that $\theta_r A$ and $\theta_r B$ are both finite. Let $\epsilon > 0$ and let δ_1, $\delta_2 > 0$ be such that

$$\theta_r A + \theta_r B \le \theta_{r\delta_1} A + \theta_{r\delta_2} B + \epsilon.$$

Set $\delta = \min(\delta_1, \delta_2, \frac{1}{2}\eta) > 0$ and let $\langle A_n\rangle_{n\in\mathbb{N}}$ be a sequence of sets of diameter at most δ, covering $A \cup B$, and such that $\sum_{n=0}^{\infty}(\operatorname{diam} A_n)^r \le \theta_{r\delta}(A \cup B) + \epsilon$. Set

$$K = \{n : A_n \cap A \ne \emptyset\}, \quad L = \{n : A_n \cap B \ne \emptyset\}.$$

Because

$$\|x - y\| \ge \eta > \operatorname{diam} A_n$$

whenever $x \in A$, $y \in B$ and $n \in \mathbb{N}$, $K \cap L = \emptyset$; and of course $A \subseteq \bigcup_{n\in K} A_k$, $B \subseteq \bigcup_{n\in L} A_n$. Consequently

$$\begin{aligned}\theta_r A + \theta_r B &\le \epsilon + \theta_{r\delta_1} A + \theta_{r\delta_2} B \\ &\le \epsilon + \sum_{n\in K}(\operatorname{diam} A_n)^r + \sum_{n\in L}(\operatorname{diam} A_n)^r \\ &\le \epsilon + \sum_{n=0}^{\infty}(\operatorname{diam} A_n)^r \le 2\epsilon + \theta_{r\delta}(A \cup B) \le 2\epsilon + \theta_r(A \cup B).\end{aligned}$$

As ϵ is arbitrary, $\theta_r(A \cup B) \ge \theta_r A + \theta_r B$, as required. **Q**

(b) It follows that $\theta_r A = \theta_r(A \cap G) + \theta_r(A \setminus G)$ whenever $A \subseteq \mathbb{R}^s$ and G is open. **P** As usual, it is enough to consider the case $\theta_r A < \infty$ and to show that in this case $\theta_r(A \cap G) + \theta_r(A \setminus G) \le \theta_r A$. Set

$$A_n = \{x : x \in A, \|x - y\| \ge 2^{-n} \text{ for every } y \in A \setminus G\},$$

$$B_0 = A_0, \quad B_n = A_n \setminus A_{n-1} \text{ for } n > 1.$$

Observe that $A_n \subseteq A_{n+1}$ for every n and $\bigcup_{n\in\mathbb{N}} A_n = \bigcup_{n\in\mathbb{N}} B_n = A\cap G$. The point is that if m, $n \in \mathbb{N}$ and $n \ge m+2$, and if $x \in B_m$ and $y \in B_n$, then there is a $z \in A \setminus G$ such that $\|y - z\| < 2^{-n+1} \le 2^{-m-1}$, while $\|x - z\|$ must be at least 2^{-m}, so $\|x - y\| \ge \|x - z\| - \|y - z\| \ge 2^{-m-1}$. It follows that for any $k \ge 0$

$$\textstyle\sum_{m=0}^{k} \theta_r B_{2m} = \theta_r(\bigcup_{m\le k} B_{2m}) \le \theta_r(A \cap G) < \infty,$$

$$\textstyle\sum_{m=0}^{k} \theta_r B_{2m+1} = \theta_r(\bigcup_{m\le k} B_{2m+1}) \le \theta_r(A \cap G) < \infty,$$

(inducing on k, using (a) above for the inductive step). Consequently $\sum_{n=0}^{\infty} \theta_r B_n < \infty$.

But now, given $\epsilon > 0$, there is an m such that $\sum_{n=m}^{\infty} \theta_r B_m \le \epsilon$, so that

$$\theta_r(A\cap G)+\theta_r(A\setminus G)\le\theta_rA_m+\sum_{n=m}^{\infty}\theta_rB_n+\theta_r(A\setminus G)$$
$$\le\epsilon+\theta_rA_m+\theta_r(A\setminus G)=\epsilon+\theta_r(A_m\cup(A\setminus G))$$

(by (a) again, since $\|x-y\|\ge 2^{-m}$ for $x\in A_m$, $y\in A\setminus G$)

$$\le\epsilon+\theta_rA.$$

As ϵ is arbitrary, $\theta_r(A\cap G)+\theta_r(A\setminus G)\le\theta_rA$, as required. **Q**

(c) Part (b) shows exactly that open sets belong to Σ_{Hr}. It follows at once that the Borel σ-algebra of $\mathbb{R}^s$ is included in Σ_{Hr}, as claimed.

264F Proposition Let $s\ge 1$ be an integer, and $r>0$; let θ_r be r-dimensional Hausdorff outer measure on $\mathbb{R}^s$, and write μ_{Hr} for r-dimensional Hausdorff measure on $\mathbb{R}^s$, Σ_{Hr} for its domain. Then

(a) for every $A\subseteq\mathbb{R}^s$ there is a Borel set $E\supseteq A$ such that $\mu_{Hr}E=\theta_rA$;

(b) $\theta_r=\mu_{Hr}^*$, the outer measure defined from μ_{Hr};

(c) if $E\in\Sigma_{Hr}$ is expressible as a countable union of sets of finite measure, there are Borel sets E', E'' such that $E'\subseteq E\subseteq E''$ and $\mu_{Hr}(E''\setminus E')=0$.

proof (a) If $\theta_rA=\infty$ this is trivial – take $E=\mathbb{R}^s$. Otherwise, for each $n\in\mathbb{N}$ choose a sequence $\langle A_{nm}\rangle_{m\in\mathbb{N}}$ of sets of diameter at most 2^{-n}, covering A, and such that $\sum_{m=0}^{\infty}(\operatorname{diam}A_{nm})^r\le\theta_{r,2^{-n}}A+2^{-n}$. Set $F_{nm}=\overline{A}_{nm}$, $E=\bigcap_{n\in\mathbb{N}}\bigcup_{m\in\mathbb{N}}F_{nm}$; then E is a Borel set in $\mathbb{R}^s$. Of course

$$A\subseteq\bigcap_{n\in\mathbb{N}}\bigcup_{m\in\mathbb{N}}A_{mn}\subseteq\bigcap_{n\in\mathbb{N}}\bigcup_{m\in\mathbb{N}}F_{nm}=E.$$

For any $n\in\mathbb{N}$,

$$\operatorname{diam}F_{nm}=\operatorname{diam}A_{nm}\le 2^{-n}\text{ for every }m\in\mathbb{N},$$

$$\sum_{m=0}^{\infty}(\operatorname{diam}F_{nm})^r=\sum_{m=0}^{\infty}(\operatorname{diam}A_{nm})^r\le\theta_{r,2^{-n}}A+2^{-n},$$

so

$$\theta_{r,2^{-n}}E\le\theta_{r,2^{-n}}A+2^{-n}.$$

Letting $n\to\infty$,

$$\theta_rE=\lim_{n\to\infty}\theta_{r,2^{-n}}E\le\lim_{n\to\infty}\theta_{r,2^{-n}}A+2^{-n}=\theta_rA;$$

of course it follows that $\theta_rA=\theta_rE$, because $A\subseteq E$. Now by 264E we know that $E\in\Sigma_{Hr}$, so we can write $\mu_{Hr}E$ in place of θ_rE.

(b) This follows at once, because we have

$$\mu_{Hr}^*A=\inf\{\mu_{Hr}E:E\in\Sigma_{Hr},\ A\subseteq E\}=\inf\{\theta_rE:E\in\Sigma_{Hr},\ A\subseteq E\}\ge\theta_rA$$

for every $A\subseteq\mathbb{R}^s$. On the other hand, if $A\subseteq\mathbb{R}^s$, we have a Borel set $E\supseteq A$ such that $\theta_rA=\mu_{Hr}E$, so that $\mu_{Hr}^*A\le\mu_{Hr}E=\theta_rA$.

(c)(i) Suppose first that $\mu_{Hr}E<\infty$. By (a), there are Borel sets $E''\supseteq E$, $H\supseteq E''\setminus E$ such that $\mu_{Hr}E''=\theta_rE$,

$$\mu_{Hr}H=\theta_r(E''\setminus E)=\mu_{Hr}(E''\setminus E)=\mu_{Hr}E''-\mu_{Hr}E=\mu_{Hr}E''-\theta_rE=0.$$

So setting $E'=E''\setminus H$, we obtain a Borel set included in E, and

$$\mu_{Hr}(E''\setminus E')\le\mu_{Hr}H=0.$$

(ii) For the general case, express E as $\bigcup_{n\in\mathbb{N}}E_n$ where $\mu_{Hr}E_n<\infty$ for each n; take Borel sets E'_n, E''_n such that $E'_n\subseteq E_n\subseteq E''_n$ and $\mu_{Hr}(E''_n\setminus E'_n)=0$ for each n; and set $E'=\bigcup_{n\in\mathbb{N}}E'_n$, $E''=\bigcup_{n\in\mathbb{N}}E''_n$.

264G Lipschitz functions The definition of Hausdorff measure is exactly adapted to the following result, corresponding to 262D.

Proposition Let m, $s \ge 1$ be integers, and $\phi : D \to \mathbb{R}^s$ a γ-Lipschitz function, where D is a subset of $\mathbb{R}^m$. Then for any $A \subseteq D$ and $r \ge 0$,

$$\mu_{Hr}^*(\phi[A]) \le \gamma^r \mu_{Hr}^* A$$

for every $A \subseteq D$, writing μ_{Hr} for r-dimensional Hausdorff outer measure on either $\mathbb{R}^m$ or $\mathbb{R}^s$.

proof (a) The case $r = 0$ is trivial, since then $\gamma^r = 1$ and $\mu_{Hr}^* A = \mu_{H0} A = \#(A)$ if A is finite, ∞ otherwise, while $\#(\phi[A]) \le \#(A)$.

(b) If $r > 0$, then take any $\delta > 0$. Set $\eta = \delta/(1+\gamma)$ and consider $\theta_{r\eta} : \mathcal{P}\mathbb{R}^m \to [0, \infty]$, defined as in 264A. We know from 264Fb that

$$\mu_{Hr}^* A = \theta_r A \ge \theta_{r\eta} A,$$

so there is a sequence $\langle A_n \rangle_{n\in\mathbb{N}}$ of sets, all of diameter at most η, covering A, with $\sum_{n=0}^{\infty} (\operatorname{diam} A_n)^r \le \mu_{Hr}^* A + \delta$. Now $\phi[A] \subseteq \bigcup_{n\in\mathbb{N}} \phi[A_n \cap D]$ and

$$\operatorname{diam} \phi[A_n \cap D] \le \gamma \operatorname{diam} A_n \le \gamma\eta \le \delta$$

for every n. Consequently

$$\theta_{r\delta}(\phi[A]) \le \sum_{n=0}^{\infty} (\operatorname{diam} \phi[A_n])^r \le \sum_{n=0}^{\infty} \gamma^r (\operatorname{diam} A_n)^r \le \gamma^r (\mu_{Hr}^* A + \delta),$$

and

$$\mu_{Hr}^*(\phi[A]) = \lim_{\delta\downarrow 0} \theta_{r\delta}(\phi[A]) \le \gamma^r \mu_{Hr}^* A,$$

as claimed.

264H The next step is to relate r-dimensional Hausdorff measure on $\mathbb{R}^r$ to Lebesgue measure on $\mathbb{R}^r$. The basic fact we need is the following, which is even more important for the idea in its proof than for the result.

Theorem Let $r \ge 1$ be an integer, and A a bounded subset of $\mathbb{R}^r$; write μ_r for Lebesgue measure on $\mathbb{R}^r$ and $d = \operatorname{diam} A$. Then

$$\mu_r^*(A) \le \mu_r B(\mathbf{0}, \tfrac{d}{2}) = 2^{-r} \beta_r d^r,$$

where $B(\mathbf{0}, \frac{d}{2})$ is the ball with centre $\mathbf{0}$ and diameter d, so that $B(\mathbf{0}, 1)$ is the unit ball in $\mathbb{R}^r$, and has measure

$$\begin{aligned} \beta_r &= \frac{1}{k!}\pi^k \text{ if } r = 2k \text{ is even,} \\ &= \frac{2^{2k+1}k!}{(2k+1)!}\pi^k \text{ if } r = 2k+1 \text{ is odd.} \end{aligned}$$

proof (a) For the calculation of β_r, see 252Q or 252Xi.

(b) The case $r = 1$ is elementary, for in this case A is included in an interval of length $\operatorname{diam} A$, so that $\mu_1^* A \le \operatorname{diam} A$. So henceforth let us suppose that $r \ge 2$.

(c) For $1 \le i \le r$ let $S_i : \mathbb{R}^r \to \mathbb{R}^r$ be reflection in the ith coordinate, so that $S_i x = (\xi_1, \dots, \xi_{i-1}, -\xi_i, \xi_{i+1}, \dots, \xi_r)$ for every $x = (\xi_1, \dots, \xi_r) \in \mathbb{R}^r$. Let us say that a set $C \subseteq \mathbb{R}^r$ is **symmetric in coordinates in** J, where $J \subseteq \{1, \dots, r\}$, if $S_i[C] = C$ for $i \in J$. Now the centre of the argument is the following fact: if $C \subseteq \mathbb{R}$ is a bounded set which is symmetric in coordinates in J, where J is a proper subset of $\{1, \dots, r\}$, and $j \in \{1, \dots, r\} \setminus J$, then there is a set D, symmetric in coordinates in $J \cup \{j\}$, such that $\operatorname{diam} D \le \operatorname{diam} C$ and $\mu_r^* C \le \mu_r^* D$.

P (i) Because Lebesgue measure is invariant under permutation of coordinates, it is enough to deal with the case $j = r$. Start by writing $F = \overline{C}$, so that $\operatorname{diam} F = \operatorname{diam} C$ and $\mu_r F \ge \mu_r^* C$. Note that because S_i is a homeomorphism for every i,

$$S_i[F] = S_i[\overline{C}] = \overline{S_i[C]} = \overline{C} = F$$

for $i \in J$, and F is symmetric in coordinates in J.

For $y = (\eta_1, \dots, \eta_{r-1}) \in \mathbb{R}^{r-1}$, set

$$F_y = \{\xi : (\eta_1, \dots, \eta_{r-1}, \xi) \in F\}, \quad f(y) = \mu_1 F_y,$$

where μ_1 is Lebesgue measure on $\mathbb{R}$. Set

$$D = \{(y, \xi) : y \in \mathbb{R}^{r-1},\ |\xi| < \frac{1}{2}f(y)\} \subseteq \mathbb{R}^r.$$

(ii) If $H \subseteq \mathbb{R}^r$ is measurable and $H \supseteq D$, then, writing μ_{r-1} for Lebesgue measure on $\mathbb{R}^{r-1}$, we have

$$\mu_r H = \int \mu_1\{\xi : (y, \xi) \in H\}\mu_{r-1}(dy)$$

(using 251N and 252D)

$$\geq \int \mu_1\{\xi : (y, \xi) \in D\}\mu_{r-1}(dy) = \int f(y)\mu_{r-1}(dy)$$
$$= \int \mu_1\{\xi : (y, \xi) \in F\}\mu_{r-1}(dy) = \mu_r F \geq \mu_r^* C.$$

As H is arbitrary, $\mu_r^* D \geq \mu_r^* C$.

(iii) The next step is to check that $\operatorname{diam} D \leq \operatorname{diam} C$. If x, $x' \in D$, express them as (y, ξ_r) and (y', ξ_r'). Because F is a bounded closed set in $\mathbb{R}^r$, F_y and $F_{y'}$ are bounded closed subsets of $\mathbb{R}$. Also both $f(y)$ and $f(y')$ must be greater than 0, so that F_y, $F_{y'}$ are both non-empty. Consequently

$$\alpha = \inf F_y, \quad \beta = \sup F_y, \quad \alpha' = \inf F_{y'}, \quad \beta' = \sup F_{y'}$$

are all defined in $\mathbb{R}$, and α, $\beta \in F_y$, while α' and β' belong to $F_{y'}$. We have

$$|\xi_r - \xi_r'| \leq |\xi_r| + |\xi_r'| < \frac{1}{2}f(y) + \frac{1}{2}f(y')$$
$$= \frac{1}{2}(\mu_1 F_y + \mu_1 F_{y'}) \leq \frac{1}{2}(\beta - \alpha + \beta' - \alpha') \leq \max(\beta' - \alpha, \beta - \alpha').$$

So taking (ξ, ξ') to be one of (α, β') or (β, α'), we can find $\xi \in F_y$, $\xi' \in F_{y'}$ such that $|\xi - \xi'| \geq |\xi_r - \xi_r'|$. Now $z = (y, \xi)$, $z' = (y', \xi')$ both belong to F, so

$$\|x - x'\|^2 = \|y - y'\|^2 + |\xi_r - \xi_r'|^2 \leq \|y - y'\|^2 + |\xi - \xi'|^2 = \|z - z'\|^2 \leq (\operatorname{diam} F)^2,$$

and $\|x - x'\| \leq \operatorname{diam} F$. As x and x' are arbitrary, $\operatorname{diam} D \leq \operatorname{diam} F = \operatorname{diam} C$, as claimed.

(iv) Evidently $S_r[D] = D$. Moreover, if $i \in J$, then (interpreting S_i as an operator on $\mathbb{R}^{r-1}$)

$$F_{S_i(y)} = F_y \text{ for every } y \in \mathbb{R}^{r-1},$$

so $f(S_i(y)) = f(y)$ and, for $\xi \in \mathbb{R}$, $y \in \mathbb{R}^{r-1}$,

$$(y, \xi) \in D \iff |\xi| < \tfrac{1}{2}f(y) \iff |\xi| < \tfrac{1}{2}f(S_i(y)) \iff (S_i(y), \xi) \in D,$$

so that $S_i[D] = D$. Thus D is symmetric in coordinates in $J \cup \{r\}$. **Q**

(d) The rest is easy. Starting from any bounded $A \subseteq \mathbb{R}^r$, set $A_0 = A$ and construct inductively $A_1, \ldots, A_r$ such that

$$d = \operatorname{diam} A = \operatorname{diam} A_0 \geq \operatorname{diam} A_1 \geq \ldots \geq \operatorname{diam} A_r,$$

$$\mu_r^* A = \mu_r^* A_0 \leq \ldots \leq \mu_r^* A_r,$$

$$A_j \text{ is symmetric in coordinates in } \{1, \ldots, j\} \text{ for every } j \leq r.$$

At the end, we have A_r symmetric in coordinates in $\{1, \ldots, r\}$. But this means that if $x \in A_r$ then

$$-x = S_1 S_2 \ldots S_r x \in A_r,$$

so that

$$\|x\| = \frac{1}{2}\|x - (-x)\| \leq \frac{1}{2}\operatorname{diam} A_r \leq \frac{d}{2}.$$

Thus $A_r \subseteq B(\mathbf{0}, \frac{d}{2})$, and

$$\mu_r^* A \leq \mu_r^* A_r \leq \mu_r B(\mathbf{0}, \frac{d}{2}),$$

as claimed.

264I Theorem Let $r \geq 1$ be an integer; let μ be Lebesgue measure on $\mathbb{R}^r$, and let μ_{Hr} be r-dimensional Hausdorff measure on $\mathbb{R}^r$. Then μ and μ_{Hr} have the same measurable sets and

$$\mu E = 2^{-r}\beta_r\mu_{Hr}E$$

for every measurable set $E \subseteq \mathbb{R}^r$, where $\beta_r = \mu B(\mathbf{0}, 1)$, so that the normalizing factor is

$$2^{-r}\beta_r = \frac{1}{2^{2k}k!}\pi^k \text{ if } r = 2k \text{ is even,}$$
$$= \frac{k!}{(2k+1)!}\pi^k \text{ if } r = 2k+1 \text{ is odd.}$$

proof (a) Of course if $B = B(x, \alpha)$ is any ball of radius α,

$$2^{-r}\beta_r(\operatorname{diam} B)^r = \beta_r\alpha^r = \mu B.$$

(b) The point is that $\mu^* = 2^{-r}\beta_r\mu^*_{Hr}$. **P** Let $A \subseteq \mathbb{R}^r$.

(i) Let δ, $\epsilon > 0$. By 261F, there is a sequence $\langle B_n\rangle_{n\in\mathbb{N}}$ of balls, all of diameter at most δ, such that $A \subseteq \bigcup_{n\in\mathbb{N}} B_n$ and $\sum_{n=0}^\infty \mu B_n \leq \mu^*A + \epsilon$. Now, defining $\theta_{r\delta}$ as in 264A,

$$2^{-r}\beta_r\theta_{r\delta}(A) \leq 2^{-r}\beta_r\sum_{n=0}^\infty(\operatorname{diam} B_n)^r = \sum_{n=0}^\infty \mu B_n \leq \mu^*A + \epsilon.$$

Letting $\delta \downarrow 0$,

$$2^{-r}\beta_r\mu^*_{Hr}A \leq \mu^*A + \epsilon.$$

As ϵ is arbitrary, $2^{-r}\beta_r\mu^*_{Hr}A \leq \mu^*A$.

(ii) Let $\epsilon > 0$. Then there is a sequence $\langle A_n\rangle_{n\in\mathbb{N}}$ of sets of diameter at most 1 such that $A \subseteq \bigcup_{n\in\mathbb{N}} A_n$ and $\sum_{n=0}^\infty(\operatorname{diam} A_n)^r \leq \theta_{r1}A + \epsilon$, so that

$$\mu^*A \leq \sum_{n=0}^\infty \mu^*A_n \leq \sum_{n=0}^\infty 2^{-r}\beta_r(\operatorname{diam} A_n)^r \leq 2^{-r}\beta_r(\theta_{r1}A + \epsilon) \leq 2^{-r}\beta_r(\mu^*_{Hr}A + \epsilon)$$

by 264H. As ϵ is arbitrary, $\mu^*A \leq 2^{-r}\beta_r\mu^*_{Hr}A$. **Q**

(c) Because μ, μ_{Hr} are the measures defined from their respective outer measures by Carathéodory's method, it follows at once that $\mu = 2^{-r}\beta_r\mu_{Hr}$ in the strict sense required.

***264J The Cantor set** I remarked in 264A that fractional 'dimensions' r were of interest. I have no space for these here, and they are off the main lines of this volume, but I will give one result for its intrinsic interest.

Proposition Let C be the Cantor set in $[0, 1]$. Set $r = \ln 2/\ln 3$. Then the r-dimensional Hausdorff measure of C is 1.

proof (a) Recall that $C = \bigcap_{n\in\mathbb{N}} C_n$, where each C_n consists of 2^n closed intervals of length 3^{-n}, and C_{n+1} is obtained from C_n by deleting the middle (open) third of each interval of C_n. (See 134G.) Because C is closed, $\mu_{Hr}C$ is defined (264E). Note that $3^r = 2$.

(b) If $\delta > 0$, take n such that $3^{-n} \leq \delta$; then C can be covered by 2^n intervals of diameter 3^{-n}, so

$$\theta_{r\delta}C \leq 2^n(3^{-n})^r = 1.$$

Consequently

$$\mu_{Hr}C = \mu^*_{Hr}C = \lim_{\delta\downarrow 0}\theta_{r\delta}C \leq 1.$$

(c) We need the following elementary fact: if α, β, $\gamma \geq 0$ and $\max(\alpha, \gamma) \leq \beta$, then $\alpha^r + \gamma^r \leq (\alpha + \beta + \gamma)^r$. **P** Because $0 < r \leq 1$,

$$\xi \mapsto (\xi + \eta)^r - \xi^r = r\int_0^\eta (\xi + \zeta)^{r-1}d\zeta$$

is non-increasing for every $\eta \geq 0$. Consequently

$$\begin{aligned}(\alpha + \beta + \gamma)^r - \alpha^r - \gamma^r &\geq (\beta + \beta + \gamma)^r - \beta^r - \gamma^r \\ &\geq (\beta + \beta + \beta)^r - \beta^r - \beta^r = \beta^r(3^r - 2) = 0,\end{aligned}$$

as required. **Q**

(d) Now suppose that $I \subseteq \mathbb{R}$ is any interval, and $m \in \mathbb{N}$; write $j_m(I)$ for the number of the intervals composing C_m which are included in I. Then $2^{-m}j_m(I) \leq (\operatorname{diam} I)^r$. **P** If I does not meet C_m, this is trivial. Otherwise, induce on

$$l = \min\{i : I \text{ meets only one of the intervals composing } C_{m-i}\}.$$

If $l = 0$, so that I meets only one of the intervals composing C_m, then $j_m(I) \le 1$, and if $j_m(I) = 1$ then $\operatorname{diam} I \ge 3^{-m}$ so $(\operatorname{diam} I)^r \ge 2^{-m}$; thus the induction starts. For the inductive step to $l > 1$, let J be the interval of C_{m-l} which meets I, and J', J'' the two intervals of C_{m-l+1} included in J, so that I meets both J' and J'', and

$$j_m(I) = j_m(I \cap J) = j_m(I \cap J') + j_m(I \cap J'').$$

By the inductive hypothesis,

$$(\operatorname{diam}(I \cap J'))^r + (\operatorname{diam}(I \cap J''))^r \ge 2^{-m} j_m(I \cap J') + 2^{-m} j_m(I \cap J'') = 2^{-m} j_m(I).$$

On the other hand, by (c),

$$\begin{aligned}(\operatorname{diam}(I \cap J'))^r + (\operatorname{diam}(I \cap J''))^r &\le (\operatorname{diam}(I \cap J') + 3^{-m+l-1} + \operatorname{diam}(I \cap J''))^r \\ &= (\operatorname{diam}(I \cap J))^r \le (\operatorname{diam} I)^r\end{aligned}$$

because J', J'' both have diameter at most $3^{-(m-l+1)}$, the length of the interval between them. Thus the induction continues. **Q**

(e) Now suppose that $\epsilon > 0$. Then there is a sequence $\langle A_n \rangle_{n \in \mathbb{N}}$ of sets, covering C, such that $\sum_{n=0}^{\infty} (\operatorname{diam} A_n)^r < \mu_{Hr} C + \epsilon$. Take $\eta_n > 0$ such that $\sum_{n=0}^{\infty} (\operatorname{diam} A_n + \eta_n)^r \le \mu_{Hr} C + \epsilon$, and for each n take an open interval $I_n \supseteq A_n$ of length at most $\operatorname{diam} A_n + \eta_n$ and with neither endpoint belonging to C; this is possible because C does not include any non-trivial interval. Now $C \subseteq \bigcup_{n \in \mathbb{N}} I_n$; because C is compact, there is a $k \in \mathbb{N}$ such that $C \subseteq \bigcup_{n \le k} I_n$. Next, there is an $m \in \mathbb{N}$ such that no endpoint of any I_n, for $n \le k$, belongs to C_m. Consequently each of the intervals composing C_m must be included in some I_n, and (in the terminology of (d) above) $\sum_{n=0}^{k} j_m(I_n) \ge 2^m$. Accordingly

$$1 \le \sum_{n=0}^{k} 2^{-m} j_m(I_n) \le \sum_{n=0}^{k} (\operatorname{diam} I_n)^r \le \sum_{n=0}^{\infty} (\operatorname{diam} A_n + \eta_n)^r \le \mu_{Hr} C + \epsilon.$$

As ϵ is arbitrary, $\mu_{Hr} C \ge 1$, as required.

***264K General metric spaces** While this chapter deals exclusively with Euclidean spaces, readers familiar with the general theory of metric spaces may find the nature of the theory clearer if they use the language of metric spaces in the basic definitions and results. I therefore repeat the definition here, and spell out the corresponding results in the exercises 264Yb-264Yl.

Let (X, ρ) be a metric space, and $r > 0$. For any $A \subseteq X$, $\delta > 0$ set

$$\begin{aligned}\theta_{r\delta} A = \inf\{\sum_{n=0}^{\infty} (\operatorname{diam} A_n)^r : \langle A_n \rangle_{n \in \mathbb{N}} &\text{ is a sequence of subsets of } X \text{ covering } A, \\ &\operatorname{diam} A_n \le \delta \text{ for every } n \in \mathbb{N}\},\end{aligned}$$

interpreting the diameter of the empty set as 0, and $\inf \emptyset$ as ∞, so that $\theta_{r\delta} A = \infty$ if A cannot be covered by a sequence of sets of diameter at most δ. Say that $\theta_r A = \sup_{\delta > 0} \theta_{r\delta} A$ is the **r-dimensional Hausdorff outer measure** of A, and take the measure μ_{Hr} defined by Carathéodory's method from this outer measure to be **r-dimensional Hausdorff measure** on X.

264X Basic exercises >(a) Show that all the functions $\theta_{r\delta}$ of 264A are outer measures. Show that in that context, $\theta_{r\delta}(A) = 0$ iff $\theta_r(A) = 0$, for any $\delta > 0$ and any $A \subseteq \mathbb{R}^s$.

(b) Let $s \ge 1$ be an integer, and θ an outer measure on $\mathbb{R}^s$ such that $\theta(A \cup B) = \theta A + \theta B$ whenever A, B are non-empty subsets of $\mathbb{R}^s$ and $\inf_{x \in A, y \in B} \|x - y\| > 0$. Show that every Borel subset of $\mathbb{R}^s$ is measured by the measure defined from θ by Carathéodory's method.

>(c) Let $s \ge 1$ be an integer and $r > 0$; define $\theta_{r\delta}$ as in 264A. Show that for any $A \subseteq \mathbb{R}^s$, $\delta > 0$,

$$\begin{aligned}\theta_{r\delta} A = \inf\{\sum_{n=0}^{\infty} (\operatorname{diam} F_n)^r : \langle F_n \rangle_{n \in \mathbb{N}} &\text{ is a sequence of closed subsets of } X \\ &\text{covering } A, \operatorname{diam} F_n \le \delta \text{ for every } n \in \mathbb{N}\} \\ = \inf\{\sum_{n=0}^{\infty} (\operatorname{diam} G_n)^r : \langle G_n \rangle_{n \in \mathbb{N}} &\text{ is a sequence of open subsets of } X \\ &\text{covering } A, \operatorname{diam} G_n \le \delta \text{ for every } n \in \mathbb{N}\}.\end{aligned}$$

>(d) Let $s \geq 1$ be an integer and $r \geq 0$; let μ_{Hr} be r-dimensional Hausdorff measure on $\mathbb{R}^s$. Show that for every $A \subseteq \mathbb{R}^s$ there is a G_δ set (that is, a set expressible as the intersection of a sequence of open sets) $H \supseteq A$ such that $\mu_{Hr}H = \mu^*_{Hr}A$. (*Hint*: use 264Xc.)

>(e) Let $s \geq 1$ be an integer, and $0 \leq r < r'$. Show that if $A \subseteq \mathbb{R}^s$ and the r-dimensional Hausdorff outer measure $\mu^*_{Hr}A$ of A is finite, then $\mu^*_{Hr'}A$ must be zero.

(f)(i) Suppose that $f : [a,b] \to \mathbb{R}$ has graph $\Gamma_f \subseteq \mathbb{R}^2$, where $a \leq b$ in $\mathbb{R}$. Show that the outer measure $\mu^*_{H1}(\Gamma_f)$ of Γ for one-dimensional Hausdorff measure on $\mathbb{R}^2$ is at most $b - a + \mathrm{Var}_{[a,b]}(f)$. (*Hint*: if f has finite variation, show that $\mathrm{diam}(\Gamma_{f\restriction]t,u[}) \leq u - t + \mathrm{Var}_{]t,u[}(f)$; then use 224E.) (ii) Let $f : [0,1] \to [0,1]$ be the Cantor function (134H). Show that $\mu_{H1}(Gamma_f) = 2$. (*Hint*: 264G.)

(g) In 264A, show that

$$\theta_{r\delta}A = \inf\{\sum_{n=0}^{\infty}(\mathrm{diam}\,A_n)^r : \langle A_n\rangle_{n\in\mathbb{N}} \text{ is a sequence of convex sets covering } A,$$
$$\mathrm{diam}\,A_n \leq \delta \text{ for every } n \in \mathbb{N}\}$$

for any $A \subseteq \mathbb{R}^s$.

264Y Further exercises (a) Let θ_{11} be the outer measure on $\mathbb{R}^2$ defined in 264A, with $r = \delta = 1$, and μ_{11} the measure derived from θ_{11} by Carathéodory's method, Σ_{11} its domain. Show that any set in Σ_{11} is either negligible or conegligible.

(b) Let (X,ρ) be a metric space and $r \geq 0$. Show that if $A \subseteq X$ and $\mu^*_{Hr}A < \infty$, then A is separable.

(c) Let (X,ρ) be a metric space, and θ an outer measure on X such that $\theta(A \cup B) = \theta A + \theta B$ whenever A, B are non-empty subsets of X and $\inf_{x\in A,y\in B}\rho(x,y) > 0$. (Such an outer measure is called a **metric outer measure**.) Show that every open subset of X is measured by the measure defined from θ by Carathéodory's method.

(d) Let (X,ρ) be a metric space and $r > 0$; define $\theta_{r\delta}$ as in 264K. Show that for any $A \subseteq X$,

$$\begin{aligned}\mu^*_{Hr}A &= \sup_{\delta>0}\inf\{\sum_{n=0}^{\infty}(\mathrm{diam}\,F_n)^r : \langle F_n\rangle_{n\in\mathbb{N}} \text{ is a sequence of closed subsets of } X \\ &\qquad\qquad \text{covering } A,\ \mathrm{diam}\,F_n \leq \delta \text{ for every } n \in \mathbb{N}\} \\ &= \sup_{\delta>0}\inf\{\sum_{n=0}^{\infty}(\mathrm{diam}\,G_n)^r : \langle G_n\rangle_{n\in\mathbb{N}} \text{ is a sequence of open subsets of } X \\ &\qquad\qquad \text{covering } A,\ \mathrm{diam}\,G_n \leq \delta \text{ for every } n \in \mathbb{N}\}.\end{aligned}$$

(e) Let (X,ρ) be a metric space and $r \geq 0$; let μ_{Hr} be r-dimensional Hausdorff measure on X. Show that for every $A \subseteq X$ there is a G_δ set $H \supseteq A$ such that $\mu_{Hr}H = \mu^*_{Hr}A$ is the r-dimensional Hausdorff outer measure of A.

(f) Let (X,ρ) be a metric space and $r \geq 0$; let Y be any subset of X, and give Y its induced metric ρ_Y. (i) Show that the r-dimensional Hausdorff outer measure $\mu^{(Y)*}_{Hr}$ on Y is just the restriction to $\mathcal{P}Y$ of the outer measure μ^*_{Hr} on X. (ii) Show that if *either* $\mu^*_{Hr}Y < \infty$ *or* μ_{Hr} measures Y then r-dimensional Hausdorff measure $\mu^{(Y)}_{Hr}$ on Y is just the subspace measure on Y induced by the measure μ_{Hr} on X.

(g) Let (X,ρ) be a metric space and $r > 0$. Show that r-dimensional Hausdorff measure on X is atomless. (*Hint*: Let $E \in \mathrm{dom}\,\mu_{Hr}$. (i) If E is not separable, there is an open set G such that $E \cap G$ and $E \setminus G$ are both non-separable, therefore both non-negligible. (ii) If there is an $x \in E$ such that $\mu_{Hr}(E \cap B(x,\delta)) > 0$ for every $\delta > 0$, then one of these sets has non-negligible complement in E. (iii) Otherwise, $\mu_{Hr}E = 0$.)

(h) Let (X,ρ) be a metric space and $r \geq 0$; let μ_{Hr} be r-dimensional Hausdorff measure on X. Show that if $\mu_{Hr}E < \infty$ then $\mu_{Hr}E = \sup\{\mu_{Hr}F : F \subseteq E \text{ is closed and totally bounded}\}$. (*Hint*: given $\epsilon > 0$, use 264Yd to find a closed totally bounded set F such that $\mu_{Hr}(F \setminus E) = 0$ and $\mu_{Hr}(E \setminus F) \leq \epsilon$, and now apply 264Ye to $F \setminus E$.)

(i) Let (X,ρ) be a complete metric space and $r \ge 0$; let μ_{Hr} be r-dimensional Hausdorff measure on X. Show that if $\mu_{Hr}E < \infty$ then $\mu_{Hr}E = \sup\{\mu_{Hr}F : F \subseteq E \text{ is compact}\}$.

(j) Let (X,ρ) and (Y,σ) be metric spaces. If $D \subseteq X$ and $\phi : D \to Y$ is a function, then ϕ is **γ-Lipschitz** if $\sigma(\phi(x),\phi(x')) \le \gamma\rho(x,x')$ for every x, $x' \in D$. (i) Show that in this case, if $r \ge 0$, $\mu^*_{Hr}(\phi[A]) \le \gamma^r\mu^*_{Hr}A$ for every $A \subseteq D$, writing μ^*_{Hr} for r-dimensional Hausdorff outer measure on either X or Y. (ii) Show that if X is complete and $\mu_{Hr}E$ is defined and finite, then $\mu_{Hr}(\phi[E])$ is defined. (*Hint*: 264Yi.)

(k) Let (X,ρ) be a metric space, and for $r \ge 0$ let μ_{Hr} be Hausdorff r-dimensional measure on X. Show that there is a unique $\Delta = \Delta(X) \in [0,\infty]$ such that $\mu_{Hr}X = \infty$ if $r \in [0,\Delta[$, 0 if $r \in]\Delta,\infty[$.

(l) Let (X,ρ) be a metric space and $\phi : I \to X$ a continuous function, where $I \subseteq \mathbb{R}$ is an interval. Write μ_{H1} for one-dimensional Hausdorff measure on X. Show that

$$\mu_{H1}(\phi[I]) \le \sup\{\textstyle\sum_{i=1}^n \rho(\phi(t_i),\phi(t_{i-1})) : t_0,\dots,t_n \in I,\ t_0 \le \dots \le t_n\},$$

the length of the curve ϕ, with equality if ϕ is injective.

(m) Set $r = \ln 2/\ln 3$, as in 264J, and write μ_{Hr} for r-dimensional Hausdorff measure on the Cantor set C. Let λ be the usual measure on $\{0,1\}^{\mathbb{N}}$ (254J). Define $\phi : \{0,1\}^{\mathbb{N}} \to C$ by setting $\phi(x) = \frac{2}{3}\sum_{n=0}^{\infty} 3^{-n}x(n)$ for $x \in \{0,1\}^{\mathbb{N}}$. Show that ϕ is an isomorphism between $(\{0,1\}^{\mathbb{N}},\lambda)$ and (C,μ_{Hr}).

(n) Set $r = \ln 2/\ln 3$ and write μ_{Hr} for r-dimensional Hausdorff measure on the Cantor set C. Let $f : [0,1] \to [0,1]$ be the Cantor function and let μ be Lebesgue measure on $\mathbb{R}$. Show that $\mu f[E] = \mu_{Hr}E$ for every $E \in \operatorname{dom}\mu_{Hr}$ and $\mu_{Hr}(C \cap f^{-1}[F]) = \mu F$ for every Lebesgue measurable set $F \subseteq [0,1]$.

(o) Let (X,ρ) be a metric space and $h : [0,\infty[\to [0,\infty[$ a non-decreasing function. For $A \subseteq X$ set

$$\theta_h A = \sup_{\delta>0}\inf\{\sum_{n=0}^{\infty} h(\operatorname{diam} A_n) : \langle A_n\rangle_{n\in\mathbb{N}} \text{ is a sequence of subsets of } X$$
$$\text{covering } A,\ \operatorname{diam} A_n \le \delta \text{ for every } n \in \mathbb{N}\},$$

interpreting $\operatorname{diam}\emptyset$ as 0, $\inf\emptyset$ as ∞ as usual. Show that θ_h is an outer measure on X. State and prove theorems corresponding to 264E and 264F. Look through 264X and 264Y for further results which might be generalizable, perhaps on the assumption that h is continuous on the right.

(p) Let (X,ρ) be a metric space. Let us say that if $a < b$ in $\mathbb{R}$ and $f : [a,b] \to X$ is a function, then f is **absolutely continuous** if for every $\epsilon > 0$ there is a $\delta > 0$ such that $\sum_{i=1}^n \rho(f(a_i),f(b_i)) \le \epsilon$ whenever $a \le a_0 \le b_0 \le \dots \le a_n \le b_n \le b$ and $\sum_{i=0}^n b_i - a_i \le \delta$. Show that $f : [a,b] \to X$ is absolutely continuous iff it is continuous and of bounded variation (in the sense of 224Ye) and $\mu_{H1}f[A] = 0$ whenever $A \subseteq [a,b]$ is Lebesgue negligible, where μ_{H1} is 1-dimensional Hausdorff measure on X. (Compare 225M.) Show that in this case $\mu_{H1}f[\,[a,b]\,] < \infty$.

(q) Let $s \ge 1$ be an integer, and $r \in [1, infty[$. For x, $y \in \mathbb{R}^s$ set $\rho(x,y) = \|x-y\|^{s/r}$. (i) Show that ρ is a metric on $\mathbb{R}^s$ inducing the Euclidean topology. (ii) Let μ_{Hr} be the associated r-dimensional Hausdorff measure. Show that $\mu_{Hr}B(\mathbf{0},1) = 2^s$.

264 Notes and comments In this section we have come to the next step in 'geometric measure theory'. I am taking this very slowly, because there are real difficulties in the subject, and for the purposes of this volume we do not need to master very much of it. The idea here is to find a definition of r-dimensional Lebesgue measure which will be 'geometric' in the strict sense, that is, dependent only on the metric structure of $\mathbb{R}^r$, and therefore applicable to sets which have a metric structure but no linear structure. As has happened before, the definition of Hausdorff measure from an outer measure gives no problems – the only new idea in 264A-264C is that of using a supremum $\theta_r = \sup_{\delta>0}\theta_{r\delta}$ of outer measures – and the difficult part is proving that our new measure has any useful properties. Concerning the properties of Hausdorff measure, there are two essential objectives; first, to check that these measures, in general, share a reasonable proportion of the properties of Lebesgue measure; and second, to justify the term 'r-dimensional measure' by relating Hausdorff r-dimensional measure on $\mathbb{R}^r$ to Lebesgue measure on $\mathbb{R}^r$.

As for the properties of general Hausdorff measures, we have to go rather carefully. I do not give counter-examples here because they involve concepts which belong to Volumes 4 and 5 rather than this volume, but I must warn you to expect the worst. However, we do at least have open sets measurable, so that all Borel sets are measurable (264E).

The outer measure of a set A can be defined in terms of the Borel sets including A (264Fa), though not in general in terms of the open sets including A; but the measure of a measurable set E is not necessarily the supremum of the measures of the Borel sets included in E, unless E is of finite measure (264Fc). We do find that the outer measure θ_r defined in 264A is the outer measure defined from μ_{Hr} (264Fb), so that the phrase 'r-dimensional Hausdorff outer measure' is unambiguous. A crucial property of Lebesgue measure is the fact that the measure of a measurable set E is the supremum of the measures of the compact subsets of E; this is not generally shared by Hausdorff measures, but is valid for sets E of finite measure in complete spaces (264Yi). Concerning subspaces, there are no problems with the outer measures, and for sets of finite measure the subspace measures are also consistent (264Yf). Because Hausdorff measure is defined in metric terms, it behaves regularly for Lipschitz maps (264G); one of the most natural classes of functions to consider when studying metric spaces is that of 1-Lipschitz functions, so that (in the language of 264G) $\mu^*_{Hr}\phi[A] \le \mu^*_{Hr}A$ for every A.

The second essential feature of Hausdorff measure, its relation with Lebesgue measure in the appropriate dimension, is Theorem 264I. Because both Hausdorff measure and Lebesgue measure are translation-invariant, this can be proved by relatively elementary means, except for the evaluation of the normalizing constant; all we need to know is that $\mu\,[0,1[^r = 1$ and $\mu_{Hr}\,[0,1[^r$ are both finite and non-zero, and this is straightforward. (The arguments of part (a) of the proof of 261F are relevant.) For the purposes of this chapter, we do not I think have to know the value of the constant; but I cannot leave it unsettled, and therefore give Theorem 264H, the **isodiametric inequality**, to show that it is just the Lebesgue measure of an r-dimensional ball of diameter 1, as one would hope and expect. The critical step in the argument of 264H is in part (c) of the proof. This is called 'Steiner symmetrization'; the idea is that given a set A, we transform A through a series of steps, at each stage lowering, or at least not increasing, its diameter, and raising, or at least not decreasing, its outer measure, progressively making A more symmetric, until at the end we have a set which is sufficiently constrained to be amenable. The particular symmetrization operation used in this proof is important enough; but the idea of progressive regularization of an object is one of the most powerful methods in measure theory, and you should give all your attention to mastering any example you encounter. In my experience, the idea is principally useful when seeking an inequality involving disparate quantities – in the present example, the diameter and volume of a set.

Of course it is awkward having two measures on $\mathbb{R}^r$, differing by a constant multiple, and for the purposes of the next section it would actually have been a little more convenient to follow Federer 69 in using 'normalized Hausdorff measure' $2^{-r}\beta_r\mu_{Hr}$. (For non-integral r, we could take $\beta_r = \pi^{r/2}/\Gamma(1+\frac{r}{2})$, as suggested in 252Xi.) However, I believe this to be a minority position, and the striking example of Hausdorff measure on the Cantor set (264J, 264Ym-264Yn) looks much better in the non-normalized version.

Hausdorff $(\ln 2/\ln 3)$-dimensional measure on the Cantor set is of course but one, perhaps the easiest, of a large class of examples. Because the Hausdorff r-dimensional outer measure of a set A, regarded as a function of r, behaves dramatically (falling from ∞ to 0) at a certain critical value $\Delta(A)$ (see 264Xe, 264Yk), it gives us a metric space invariant of A; $\Delta(A)$ is the **Hausdorff dimension** of A. Evidently the Hausdorff dimension of C is $\ln 2/\ln 3$, while that of r-dimensional Euclidean space is r.

265 Surface measures

In this section I offer a new version of the arguments of §263, this time not with the intention of justifying integration-by-substitution, but instead to give a practically effective method of computing the Hausdorff r-dimensional measure of a smooth r-dimensional surface in an s-dimensional space. The basic case to bear in mind is $r = 2$, $s = 3$, though any other combination which you can easily visualize will also be a valuable aid to intuition. I give a fundamental theorem (265E) providing a formula from which we can hope to calculate the r-dimensional measure of a surface in s-dimensional space which is parametrized by a differentiable function, and work through some of the calculations in the case of the r-sphere (265F-265H).

265A Normalized Hausdorff measure As I remarked at the end of the last section, Hausdorff measure, as defined in 264A-264C, is not quite the most appropriate measure for our work here; so in this section I will use **normalized Hausdorff measure**, meaning $\nu_r = 2^{-r}\beta_r\mu_{Hr}$, where μ_{Hr} is r-dimensional Hausdorff measure (interpreted in whichever space is under consideration) and $\beta_r = \mu_r B(\mathbf{0},1)$ is the Lebesgue measure of any ball of radius 1 in $\mathbb{R}^r$. It will be convenient to take $\beta_0 = 1$. As shown in 264H-264I, this normalization makes ν_r on $\mathbb{R}^r$ agree with Lebesgue measure μ_r. Observe that of course $\nu^*_r = 2^{-r}\beta_r\mu^*_{Hr}$ (264Fb).

265B Linear subspaces Just as in §263, the first step is to deal with linear operators.

Theorem Suppose that r, s are integers with $1 \le r \le s$, and that T is a real $s \times r$ matrix; regard T as a linear operator from $\mathbb{R}^r$ to $\mathbb{R}^s$. Set $J = \sqrt{\det T'T}$, where T' is the transpose of T. Write ν_r for normalized r-dimensional Hausdorff measure on $\mathbb{R}^s$, T_r for its domain, and μ_r for Lebesgue measure on $\mathbb{R}^r$. Then

$$\nu_r T[E] = J\mu_r E$$

for every measurable set $E \subseteq \mathbb{R}^r$. If T is injective (that is, if $J \neq 0$), then

$$\nu_r F = J\mu_r T^{-1}[F]$$

whenever $F \in \mathrm{T}_r$ and $F \subseteq T[\mathbb{R}^r]$.

proof The formula for J assumes that $\det T'T$ is non-negative, which is a fact not in evidence; but the argument below will establish it adequately soon.

(a) Let V be the linear subspace of $\mathbb{R}^s$ consisting of vectors $y = (\eta_1, \dots, \eta_s)$ such that $\eta_i = 0$ whenever $r < i \le s$. Let R be the $r \times s$ matrix $\langle \rho_{ij} \rangle_{i \le r, j \le s}$, where $\rho_{ij} = 1$ if $i = j \le r$, 0 otherwise; then the $s \times r$ matrix R' may be regarded as a bijection from $\mathbb{R}^r$ to V. Let W be an r-dimensional linear subspace of $\mathbb{R}^s$ including $T[\mathbb{R}^r]$, and let P be an orthogonal $s \times s$ matrix such that $P[W] = V$. Then $S = RPT$ is an $r \times r$ matrix. We have $R'Ry = y$ for $y \in V$, so $R'RPT = PT$ and

$$S'S = T'P'R'RPT = T'P'PT = T'T;$$

accordingly

$$\det T'T = \det S'S = (\det S)^2 \ge 0$$

and $J = |\det S|$. At the same time,

$$P'R'S = P'R'RPT = P'PT = T.$$

Observe that $J = 0$ iff S is not injective, that is, T is not injective.

(b) If we consider the $s \times r$ matrix $P'R'$ as a map from $\mathbb{R}^r$ to $\mathbb{R}^s$, we see that $\phi = P'R'$ is an isometry between $\mathbb{R}^r$ and W, with inverse $\phi^{-1} = RP{\restriction}W$. It follows that ϕ is an isomorphism between the measure spaces $(\mathbb{R}^r, \mu_{Hr}^{(r)})$ and $(W, \mu_{HrW}^{(s)})$, where $\mu_{Hr}^{(r)}$ is r-dimensional Hausdorff measure on $\mathbb{R}^r$ and $\mu_{HrW}^{(s)}$ is the subspace measure on W induced by r-dimensional Hausdorff measure $\mu_{Hr}^{(s)}$ on $\mathbb{R}^s$.

P **(i)** If $A \subseteq \mathbb{R}^r$, $A' \subseteq W$,

$$\mu_{Hr}^{(s)*}(\phi[A]) \le \mu_{Hr}^{(r)*}(A), \quad \mu_{Hr}^{(r)*}(\phi^{-1}[A']) \le \mu_{Hr}^{(s)*}(A'),$$

using 264G twice. Thus $\mu_{Hr}^{(s)*}(\phi[A]) = \mu_{Hr}^{(r)*}(A)$ for every $A \subseteq \mathbb{R}^r$.

(ii) Now because W is closed, therefore in the domain of $\mu_{Hr}^{(s)}$ (264E), the subspace measure $\mu_{HrW}^{(s)}$ is just the measure induced by $\mu_{Hr}^{(s)*}{\restriction}W$ by Carathéodory's method (214H(b-ii)). Because ϕ is an isomorphism between $(\mathbb{R}^r, \mu_{Hr}^{(r)*})$ and $(W, \mu_{Hr}^{(s)*}{\restriction}W)$, it is an isomorphism between $(\mathbb{R}^r, \mu_{Hr}^{(r)})$ and $(W, \mu_{HrW}^{(s)})$. **Q**

(c) It follows that ϕ is also an isomorphism between the normalized versions $(\mathbb{R}^r, \mu_r)$ and (W, ν_{rW}), writing ν_{rW} for the subspace measure on W induced by ν_r.

Now if $E \subseteq \mathbb{R}^r$ is Lebesgue measurable, we have $\mu_r S[E] = J\mu_r E$, by 263A; so that

$$\nu_r T[E] = \nu_r(P'R'[S[E]]) = \nu_r(\phi[S[E]]) = \mu_r S[E] = J\mu_r E.$$

If T is injective, then $S = \phi^{-1}T$ must also be injective, so that $J \neq 0$ and

$$\nu_r F = \mu_r(\phi^{-1}[F]) = J\mu_r(S^{-1}[\phi^{-1}[F]]) = J\mu_r T^{-1}[F]$$

whenever $F \in \mathrm{T}_r$ and $F \subseteq W = T[\mathbb{R}^r]$.

265C Corollary Under the conditions of 265B,

$$\nu_r^* T[A] = J\mu_r^* A$$

for every $A \subseteq \mathbb{R}^r$.

proof (a) If E is Lebesgue measurable and $A \subseteq E$, then $T[A] \subseteq T[E]$, so

$$\nu_r^* T[A] \le \nu_r T[E] = J\mu_r E;$$

as E is arbitrary, $\nu_r^* T[A] \le J\mu_r^* A$.

(b) If $J=0$ we can stop. If $J\ne0$ then T is injective, so if $F\in\mathrm{T}_r$ and $T[A]\subseteq F$ we shall have

$$J\mu_r^*A\le J\mu_rT^{-1}[F\cap W]=\nu_r(F\cap W)\le\nu_rF;$$

as F is arbitrary, $J\mu_r^*A\le\nu_r^*T[A]$.

265D I now proceed to the lemma corresponding to 263C.

Lemma Suppose that $1\le r\le s$ and that T is an $s\times r$ matrix; set $J=\sqrt{\det T'T}$, and suppose that $J\ne0$. Then for any $\epsilon>0$ there is a $\zeta=\zeta(T,\epsilon)>0$ such that

(i) $|\sqrt{\det S'S}-J|\le\epsilon$ whenever S is an $s\times r$ matrix and $\|S-T\|\le\zeta$;

(ii) whenever $D\subseteq\mathbb{R}^r$ is a bounded set and $\phi:D\to\mathbb{R}^s$ is a function such that $\|\phi(x)-\phi(y)-T(x-y)\|\le\zeta\|x-y\|$ for all x, $y\in D$, then $|\nu_r^*\phi[D]-J\mu_r^*D|\le\epsilon\mu_r^*D$.

proof (a) Because $\det S'S$ is a continuous function of the coefficients of S, 262Hb tells us that there must be a $\zeta_0>0$ such that $|J-\sqrt{\det S'S}|\le\epsilon$ whenever $\|S-T\|\le\zeta_0$.

(b) Because $J\ne0$, T is injective, and there is an $r\times s$ matrix T^* such that T^*T is the identity $r\times r$ matrix. Take $\zeta>0$ such that $\zeta\le\zeta_0$, $\zeta\|T^*\|<1$, $J(1+\zeta\|T^*\|)^r\le J+\epsilon$ and $1-J^{-1}\epsilon\le(1-\zeta\|T^*\|)^r$.

Let $\phi:D\to\mathbb{R}^s$ be such that $\|\phi(x)-\phi(y)-T(x-y)\|\le\zeta\|x-y\|$ whenever x, $y\in D$. Set $\psi=\phi T^*$, so that $\phi=\psi T$. Then for u, $v\in T[D]$

$$\|\psi(u)-\psi(v)\|\le(1+\zeta\|T^*\|)\|u-v\|,\quad \|u-v\|\le(1-\zeta\|T^*\|)^{-1}\|\psi(u)-\psi(v)\|.$$

P Take x, $y\in D$ such that $u=Tx$, $v=Ty$; of course $x=T^*u$, $y=T^*v$. Then

$$\begin{aligned}\|\psi(u)-\psi(v)\|&=\|\phi(T^*u)-\phi(T^*v)\|=\|\phi(x)-\phi(y)\|\\&\le\|T(x-y)\|+\zeta\|x-y\|\\&=\|u-v\|+\zeta\|T^*u-T^*v\|\le\|u-v\|(1+\zeta\|T^*\|).\end{aligned}$$

Next,

$$\begin{aligned}\|u-v\|&=\|Tx-Ty\|\le\|\phi(x)-\phi(y)\|+\zeta\|x-y\|\\&=\|\psi(u)-\psi(v)\|+\zeta\|T^*u-T^*v\|\\&\le\|\psi(u)-\psi(v)\|+\zeta\|T^*\|\|u-v\|,\end{aligned}$$

so that $(1-\zeta\|T^*\|)\|u-v\|\le\|\psi(u)-\psi(v)\|$ and $\|u-v\|\le(1-\zeta\|T^*\|)^{-1}\|\psi(u)-\psi(v)\|$. **Q**

(c) Now from 264G and 265C we see that

$$\nu_r^*\phi[D]=\nu_r^*\psi[T[D]]\le(1+\zeta\|T^*\|)^r\nu_r^*T[D]=(1+\zeta\|T^*\|)^rJ\mu_r^*D\le(J+\epsilon)\mu_r^*D,$$

and (provided $\epsilon\le J$)

$$\begin{aligned}(J-\epsilon)\mu_r^*D&=(1-J^{-1}\epsilon)\nu_r^*T[D]\le(1-J^{-1}\epsilon)(1-\zeta\|T^*\|)^{-r}\nu_r^*\psi[T[D]]\\&\le\nu_r^*\psi[T[D]]=\nu_r^*\phi[D].\end{aligned}$$

(Of course, if $\epsilon\ge J$, then surely $(J-\epsilon)\mu_r^*D\le\nu_r^*\phi[D]$.) Thus

$$(J-\epsilon)\mu_r^*D\le\nu_r^*\phi[D]\le(J+\epsilon)\mu_r^*D$$

as required, and we have an appropriate ζ.

265E Theorem Suppose that $1\le r\le s$; write μ_r for Lebesgue measure on $\mathbb{R}^r$, ν_r for normalized Hausdorff measure on $\mathbb{R}^s$, and T_r for the domain of ν_r. Let $D\subseteq\mathbb{R}^r$ be any set, and $\phi:D\to\mathbb{R}^s$ a function differentiable relative to its domain at each point of D. For each $x\in D$ let $T(x)$ be a derivative of ϕ at x relative to D, and set $J(x)=\sqrt{\det T(x)'T(x)}$. Set $D'=\{x:x\in D,\ J(x)>0\}$. Then

(i) $J:D\to[0,\infty[$ is a measurable function;

(ii) $\nu_r^*\phi[D]\le\int_DJ(x)\mu_r(dx)$,

allowing ∞ as the value of the integral;

(iii) $\nu_r^*\phi[D\setminus D']=0$.

If D is Lebesgue measurable, then

(iv) $\phi[D]\in\mathrm{T}_r$.

If D is measurable and ϕ is injective, then

(v) $\nu_r\phi[D] = \int_D J\,d\mu_r$;

(vi) for any set $E \subseteq \phi[D]$, $E \in \mathrm{T}_r$ iff $\phi^{-1}[E] \cap D'$ is Lebesgue measurable, and in this case

$$\nu_r E = \int_{\phi^{-1}[E]} J(x)\mu_r(dx) = \int_D J \times \chi(\phi^{-1}[E])d\mu_r;$$

(vii) for every real-valued function g defined on a subset of $\phi[D]$,

$$\int_{\phi[D]} g\,d\nu_r = \int_D J \times g\phi\,d\mu_r,$$

if either integral is defined in $[-\infty, \infty]$, provided we interpret $J(x)g(\phi(x))$ as zero when $J(x) = 0$ and $g(\phi(x))$ is undefined.

proof I seek to follow the line laid out in the proof of 263D.

(a) Just as in 263D, we know that $J : D \to \mathbb{R}$ is measurable, since $J(x)$ is a continuous function of the coefficients of $T(x)$, all of which are measurable, by 262P. If D is Lebesgue measurable, then there is a sequence $\langle F_n\rangle_{n\in\mathbb{N}}$ of compact subsets of D such that $D\setminus\bigcup_{n\in\mathbb{N}} F_n$ is μ_r-negligible. Now $\phi[F_n]$ is compact, therefore belongs to T_r, for each $n \in \mathbb{N}$. As for $\phi[D \setminus \bigcup_{n\in\mathbb{N}} F_n]$, this must be ν_r-negligible by 264G, because ϕ is a countable union of Lipschitz functions (262N). So

$$\phi[D] = \bigcup_{n\in\mathbb{N}} \phi[F_n] \cup \phi[D \setminus \bigcup_{n\in\mathbb{N}} F_n] \in \mathrm{T}_r.$$

This deals with (i) and (iv).

(b) For the moment, assume that D is bounded and that $J(x) > 0$ for every $x \in D$, and fix $\epsilon > 0$. Let M^*_{sr} be the set of $s \times r$ matrices T such that $\det T'T \neq 0$, that is, the corresponding map $T : \mathbb{R}^r \to \mathbb{R}^s$ is injective. For $T \in M^*_{sr}$ take $\zeta(T, \epsilon) > 0$ as in 265D.

Take $\langle D_n\rangle_{n\in\mathbb{N}}$, $\langle T_n\rangle_{n\in\mathbb{N}}$ as in 262M, with $A = M^*_{sr}$, so that $\langle D_n\rangle_{n\in\mathbb{N}}$ is a partition of D into sets which are relatively measurable in D, and each T_n is an $s \times r$ matrix such that

$$\|T(x) - T_n\| \le \zeta(T_n, \epsilon) \text{ whenever } x \in D_n,$$

$$\|\phi(x) - \phi(y) - T_n(x - y)\| \le \zeta(T_n, \epsilon)\|x - y\| \text{ for all } x,\ y \in D_n.$$

Then, setting $J_n = \sqrt{\det T_n'T_n}$, we have

$$|J(x) - J_n| \le \epsilon \text{ for every } x \in D_n,$$

$$|\nu^*_r\phi[D_n] - J_n\mu^*_r D_n| \le \epsilon\mu^*_r D_n,$$

by the choice of $\zeta(T_n, \epsilon)$. So

$$\nu^*_r\phi[D] \le \sum_{n=0}^{\infty} \nu^*_r\phi[D_n]$$

(because $\phi[D] = \bigcup_{n\in\mathbb{N}} \phi[D_n]$)

$$\le \sum_{n=0}^{\infty} J\mu^*_r D_n + \epsilon\mu^*_r D_n \le \epsilon\mu^*_r D + \sum_{n=0}^{\infty} J_n\mu^*_r D_n$$

(because the D_n are disjoint and relatively measurable in D)

$$= \epsilon\mu^*_r D + \int_D \sum_{n=0}^{\infty} J_n\chi D_n d\mu$$

$$\le \epsilon\mu^*_r D + \int_D J(x) + \epsilon\mu_r(dx) = 2\epsilon\mu^*_r D + \int_D J\,d\mu_r.$$

If D is measurable and ϕ is injective, then all the D_n are Lebesgue measurable subsets of $\mathbb{R}^r$, so all the $\phi[D_n]$ are measured by ν_r, and they are also disjoint. Accordingly

$$\int_D J\,d\mu \le \sum_{n=0}^{\infty} J_n\mu_r D_n + \epsilon\mu_r D$$

$$\le \sum_{n=0}^{\infty}(\nu_r\phi[D_n] + \epsilon\mu_r D_n) + \epsilon\mu_r D = \nu_r\phi[D] + 2\epsilon\mu_r D.$$

Since ϵ is arbitrary, we get

$$\nu_r^*\phi[D] \le \int_D J\,d\mu_r,$$

and if D is measurable and ϕ is injective,

$$\int_D J\,d\mu_r \le \nu_r\phi[D];$$

thus we have (ii) and (v), on the assumption that D is bounded and $J > 0$ everywhere on D.

(c) Just as in 263D, we can now relax the assumption that D is bounded by considering $B_k = B(\mathbf{0}, k) \subseteq \mathbb{R}^r$; provided $J > 0$ everywhere on D, we get

$$\nu_r^*\phi[D] = \lim_{k\to\infty} \mu_r^*\phi[D \cap B_k] \le \lim_{k\to\infty} \int_{D\cap B_k} J\,d\mu_r = \int_D J\,d\mu_r,$$

with equality if D is measurable and ϕ is injective.

(d) Now we find that $\nu_r^*\phi[D \setminus D'] = 0$.

P (i) Let $\eta \in \,]0, 1]$. Define $\psi_\eta : D \to \mathbb{R}^{s+r}$ by setting $\psi(x) = (\phi(x), \eta x)$, identifying $\mathbb{R}^{s+r}$ with $\mathbb{R}^s \times \mathbb{R}^r$. ψ_η is differentiable relative to its domain at each point of D, with derivative $\tilde{T}_\eta(x)$, being the $(s+r)\times r$ matrix in which the top s rows consist of the $s \times r$ matrix $T(x)$, and the bottom r rows are ηI_r, writing I_r for the $r \times r$ identity matrix. (Use 262Ib.) Now of course $\tilde{T}_\eta(x)$, regarded as a map from $\mathbb{R}^r$ to $\mathbb{R}^{s+r}$, is injective, so

$$\tilde{J}_\eta(x) = \sqrt{\det \tilde{T}_\eta(x)'\tilde{T}_\eta(x)} = \sqrt{\det(T(x)'T(x) + \eta^2 I)} > 0.$$

We have $\lim_{\eta\downarrow 0} \tilde{J}_\eta(x) = J(x) = 0$ for $x \in D \setminus D'$.

(ii) Express $T(x)$ as $\langle\tau_{ij}(x)\rangle_{i\le s, j\le r}$ for each $x \in D$. Set

$$C_m = \{x : x \in D,\ \|x\| \le m,\ |\tau_{ij}(x)| \le m \text{ for all } i \le s,\ j \le r\}$$

for each $m \ge 1$. For $x \in C_m$, all the coefficients of $\tilde{T}_\eta(x)$ have moduli at most m; consequently (giving the crudest and most immediately available inequalities) all the coefficients of $\tilde{T}_\eta(x)'\tilde{T}_\eta(x)$ have moduli at most $(r+s)m^2$ and $\tilde{J}_\eta(x) \le \sqrt{r!(s+r)^r}m^r$. Consequently we can use Lebesgue's Dominated Convergence Theorem to see that

$$\lim_{\eta\downarrow 0} \int_{C_m\setminus D'} \tilde{J}_\eta d\mu_r = 0.$$

(iii) Let $\tilde{\nu}_r$ be normalized Hausdorff r-dimensional measure on $\mathbb{R}^{s+r}$. Applying (b) of this proof to $\psi_\eta{\restriction}C_m \setminus D'$, we see that

$$\tilde{\nu}_r^*\psi_\eta[C_m \setminus D'] \le \int_{C_m\setminus D'} \tilde{J}_\eta d\mu_r.$$

Now we have a natural map $P : \mathbb{R}^{s+r} \to \mathbb{R}^s$ given by setting $P(\xi_1, \dots, \xi_{s+r}) = (\xi_1, \dots, \xi_s)$, and P is 1-Lipschitz, so by 264G we have (allowing for the normalizing constants $2^{-r}\beta_r$)

$$\nu_r^*P[A] \le \tilde{\nu}_r^*A$$

for every $A \subseteq \mathbb{R}^{s+r}$. In particular,

$$\nu_r^*\phi[C_m \setminus D'] = \nu_r^*P[\psi_\eta[C_m \setminus D']] \le \tilde{\nu}_r^*\psi_\eta[C_m \setminus D'] \le \int_{C_m\setminus D'} \tilde{J}_\eta d\mu_r \to 0$$

as $\eta \downarrow 0$. But this means that $\nu_r^*\phi[C_m \setminus D'] = 0$. As $D = \bigcup_{m\ge 1} C_m$, $\nu_r^*\phi[D \setminus D'] = 0$, as claimed. **Q**

(d) This proves (iii) of the theorem. But of course this is enough to give (ii) and (v), because we must have

$$\nu_r^*\phi[D] = \nu_r^*\phi[D'] \le \int_{D'} J\,d\mu_r = \int_D J\,d\mu_r,$$

with equality if D is measurable and ϕ is injective.

(e) So let us turn to part (vi). Assume that D is measurable and that ϕ is injective.

(i) Suppose that $E \subseteq \phi[D]$ belongs to T_r. Let

$$H_k = \{x : x \in D,\ \|x\| \le k,\ J(x) \le k\}$$

for each k; then each H_k is Lebesgue measurable, so (applying (iii) to $\phi{\restriction}H_k$) $\phi[H_k] \in \mathrm{T}_r$, and

$$\nu_r\phi[H_k] \le k\mu_r H_k < \infty.$$

Thus $\phi[D]$ can be covered by a sequence of sets of finite measure for ν_r, which of course are of finite measure for r-dimensional Hausdorff measure on $\mathbb{R}^s$. By 264Fc, there are Borel sets E_1, $E_2 \subseteq \mathbb{R}^s$ such that $E_1 \subseteq E \subseteq E_2$ and $\nu_r(E_2 \setminus E_1) = 0$. Now $F_1 = \phi^{-1}[E_1]$, $F_2 = \phi^{-1}[E_2]$ are Lebesgue measurable subsets of D, and

$$\int_{F_2\setminus F_1} J\,d\mu_r = \nu_r\phi[F_2 \setminus F_1] = \nu_r(\phi[D] \cap E_2 \setminus E_1) = 0.$$

Accordingly $\mu_r(D' \cap (F_2 \setminus F_1)) = 0$. But as

$$D' \cap F_1 \subseteq D' \cap \phi^{-1}[E] \subseteq D' \cap F_2,$$

it follows that $D' \cap \phi^{-1}[E]$ is measurable, and that

$$\int_{\phi^{-1}[E]} J\,d\mu_r = \int_{D'\cap\phi^{-1}[E]} J\,d\mu_r = \int_{D'\cap F_1} J\,d\mu_r$$
$$= \int_{D\cap F_1} J\,d\mu_r = \nu_r\phi[D \cap F_1] = \nu_r E_1 = \nu_r E.$$

Moreover, $J \times \chi(\phi^{-1}[E]) = J \times \chi(D' \cap \phi^{-1}[E])$ is measurable, so we can write $\int J \times \chi(\phi^{-1}[E])$ in place of $\int_{\phi^{-1}[E]} J$.

(ii) If $E \subseteq \phi[D]$ and $D' \cap \phi^{-1}[E]$ is measurable, then of course

$$E = \phi[D' \cap \phi^{-1}[E]] \cup \phi[(D \setminus D') \cap \phi^{-1}[E]] \in \mathrm{T}_r,$$

because $\phi[G] \in \mathrm{T}_r$ for every measurable $G \subseteq D$ and $\phi[D \setminus D']$ is ν_r-negligible.

(f) Finally, (vii) follows at once from (vi), applying 235J to μ_r and the subspace measure induced by ν_r on $\phi[D]$.

265F The surface of a sphere To show how these ideas can be applied to one of the basic cases, I give the details of a method of describing spherical surface measure in s-dimensional space. Take $r \geq 1$ and $s = r + 1$. Write S_r for $\{z : z \in \mathbb{R}^{r+1}, \|z\| = 1\}$, the r-sphere. Then we have a parametrization ϕ_r of S_r given by setting

$$\phi_r\begin{pmatrix}\xi_1\\ \xi_2\\ \ldots\\ \ldots\\ \ldots\\ \xi_r\end{pmatrix} = \begin{pmatrix}\sin\xi_1\sin\xi_2\sin\xi_3\ldots\sin\xi_r\\ \cos\xi_1\sin\xi_2\sin\xi_3\ldots\sin\xi_r\\ \cos\xi_2\sin\xi_3\ldots\sin\xi_r\\ \ldots\\ \cos\xi_{r-2}\sin\xi_{r-1}\sin\xi_r\\ \cos\xi_{r-1}\sin\xi_r\\ \cos\xi_r\end{pmatrix}.$$

I choose this formulation because I wish to use an inductive argument based on the fact that

$$\phi_{r+1}\begin{pmatrix}x\\ \xi\end{pmatrix} = \begin{pmatrix}\sin\xi\,\phi_r(x)\\ \cos\xi\end{pmatrix}$$

for $x \in \mathbb{R}^r$, $\xi \in \mathbb{R}$. Every ϕ_r is differentiable, by 262Id. If we set

$$D_r = \{x : \xi_1 \in \left]-\pi,\pi\right],\ \xi_2,\ldots,\xi_r \in [0,\pi],$$
$$\text{if } \xi_j \in \{0,\pi\} \text{ then } \xi_i = 0 \text{ for } i < j\},$$

then it is easy to check that D_r is a Borel subset of $\mathbb{R}^r$ and that $\phi_r{\restriction}D_r$ is a bijection between D_r and S_r. Now let $T_r(x)$ be the $(r+1)\times r$ matrix $\phi_r'(x)$. Then

$$T_{r+1}\begin{pmatrix}x\\ \xi\end{pmatrix} = \begin{pmatrix}\sin\xi\,T_r(x) & \cos\xi\,\phi_r(x)\\ \mathbf{0} & -\sin\xi\end{pmatrix}.$$

So

$$(T_{r+1}\begin{pmatrix}x\\ \xi\end{pmatrix})'T_{r+1}\begin{pmatrix}x\\ \xi\end{pmatrix} = \begin{pmatrix}\sin^2\xi\,T_r(x)'T_r(x) & \sin\xi\cos\xi\,T_r(x)'\phi_r(x)\\ \cos\xi\sin\xi\,\phi_r(x)'T_r(x) & \cos^2\xi\phi_r(x)'\phi_r(x) + \sin^2\xi\end{pmatrix}.$$

But of course $\phi_r(x)'\phi_r(x) = \|\phi_r(x)\|^2 = 1$ for every x, and (differentiating with respect to each coordinate of x, if you wish) $T_r(x)'\phi_r(x) = \mathbf{0}$, $\phi_r(x)'T_r(x) = \mathbf{0}$. So we get

$$(T_{r+1}\begin{pmatrix}x\\ \xi\end{pmatrix})'T_{r+1}\begin{pmatrix}x\\ \xi\end{pmatrix} = \begin{pmatrix}\sin^2\xi\,T_r(x)'T_r(x) & \mathbf{0}\\ \mathbf{0} & 1\end{pmatrix},$$

and writing $J_r(x)=\sqrt{\det T_r(x)'T_r(x)}$,

$$J_{r+1}\begin{pmatrix}x\\ \xi\end{pmatrix}=|\sin^r\xi|J_r(x).$$

At this point we induce on r to see that

$$J_r(x)=|\sin^{r-1}\xi_r\sin^{r-2}\xi_{r-1}\ldots\sin\xi_2|$$

(since of course the induction starts with the case $r=1$, $\phi_1(x)=\begin{pmatrix}\sin x\\ \cos x\end{pmatrix}$, $T_1(x)=\begin{pmatrix}\cos x\\ -\sin x\end{pmatrix}$, $T_1(x)'T_1(x)=1$, $J_1(x)=1$).

To find the surface measure of S_r, we need to calculate

$$\begin{aligned}\int_{D_r}J_r d\mu_r&=\int_0^\pi\ldots\int_0^\pi\int_{-\pi}^\pi\sin^{r-1}\xi_r\ldots\sin\xi_2 d\xi_1 d\xi_2\ldots d\xi_r\\ &=2\pi\prod_{k=2}^r\int_0^\pi\sin^{k-1}t\,dt=2\pi\prod_{k=1}^{r-1}\int_{-\pi/2}^{\pi/2}\cos^k t\,dt\end{aligned}$$

(substituting $\frac{\pi}{2}-t$ for t). But in the language of 252Q, this is just

$$2\pi\textstyle\prod_{k=1}^{r-1}I_k=2\pi\beta_{r-1},$$

where β_{r-1} is the volume of the unit ball of $\mathbb{R}^{r-1}$ (interpreting β_0 as 1, if you like).

265G The surface area of a sphere can also be calculated through the following result.

Theorem Let μ_{r+1} be Lebesgue measure on $\mathbb{R}^{r+1}$, and ν_r normalized r-dimensional Hausdorff measure on $\mathbb{R}^{r+1}$. If f is a locally μ_{r+1}-integrable real-valued function, $y\in\mathbb{R}^{r+1}$ and $\delta>0$,

$$\textstyle\int_{B(y,\delta)}f d\mu_{r+1}=\int_0^\delta\int_{\partial B(y,t)}f d\nu_r dt,$$

where I write $\partial B(y,s)$ for the sphere $\{x:\|x-y\|=s\}$ and the integral $\int\ldots dt$ is to be taken with respect to Lebesgue measure on $\mathbb{R}$.

proof Take any differentiable function $\phi:\mathbb{R}^r\to S_r$ with a Borel set $F\subseteq\mathbb{R}^r$ such that $\phi\restriction F$ is a bijection between F and S_r; such a pair (ϕ,F) is described in 265F. Define $\psi:\mathbb{R}^r\times\mathbb{R}\to\mathbb{R}^{r+1}$ by setting $\psi(z,t)=y+t\phi(z)$; then ψ is differentiable and $\psi\restriction F\times]0,\delta]$ is a bijection between $F\times]0,\delta]$ and $B(y,\delta)\setminus\{y\}$. For $t\in]0,\delta]$, $z\in\mathbb{R}^r$ set $\psi_t(z)=\psi(z,t)$; then $\psi_t\restriction F$ is a bijection between F and the sphere $\{x:\|x-y\|=t\}=\partial B(y,t)$.

The derivative of ϕ at z is an $(r+1)\times r$ matrix $T_1(z)$ say, and the derivative $T_t(z)$ of ψ_t at z is just $tT_1(z)$; also the derivative of ψ at (z,t) is the the $(r+1)\times(r+1)$ matrix $T(z,t)=\begin{pmatrix}tT_1(z) & \phi(z)\end{pmatrix}$, where $\phi(z)$ is interpreted as a column vector. If we set

$$J_t(z)=\sqrt{\det T_t(z)'T_t(z)},\quad J(z,t)=|\det T(z,t)|,$$

then

$$\begin{aligned}J(z,t)^2=\det T(z,t)'T(z,t)&=\det\begin{pmatrix}tT_1(z)'\\ \phi(z)'\end{pmatrix}\begin{pmatrix}tT_1z & \phi(z)\end{pmatrix}\\ &=\det\begin{pmatrix}t^2T_1(z)'T_1(z) & \mathbf{0}\\ \mathbf{0} & 1\end{pmatrix}=J_t(z)^2,\end{aligned}$$

because when we come to calculate the $(i,r+1)$-coefficient of $T(z,t)'T(z,t)$, for $1\le i\le r$, it is

$$\textstyle\sum_{j=1}^{r+1}t\frac{\partial\phi_j}{\partial\zeta_i}(z)\phi_j(z)=\frac{t}{2}\frac{\partial}{\partial\zeta_i}(\sum_{j=1}^{r+1}\phi_j(z)^2)=0,$$

where ϕ_j is the jth coordinate of ϕ; while the $(r+1,r+1)$-coefficient of $T(z,t)'T(z,t)$ is just $\sum_{j=1}^{r+1}\phi_j(z)^2=1$. So in fact $J(z,t)=J_t(z)$ for all $z\in\mathbb{R}^r$, $t>0$.

Now, given $f\in\mathcal{L}^1(\mu_{r+1})$, we can calculate

$$\int_{B(y,\delta)} f d\mu_{r+1} = \int_{B(y,\delta)\setminus\{y\}} f d\mu_{r+1}$$
$$= \int_{F\times]0,\delta]} f(\psi(z,t))J(z,t)\mu_{r+1}(d(z,t))$$

(by 263D)

$$= \int_0^\delta \int_F f(\psi_t(z))J_t(z)\mu_r(dz)dt$$

(where μ_r is Lebesgue measure on $\mathbb{R}^r$, by Fubini's theorem, 252B)

$$= \int_0^\delta \int_{\partial B(y,t)} f d\nu_r dt$$

by 265E(vii).

265H Corollary If ν_r is normalized r-dimensional Hausdorff measure on $\mathbb{R}^{r+1}$, then $\nu_r S_r = (r+1)\beta_{r+1}$.

proof In 265G, take $y = \mathbf{0}$, $\delta = 1$, and $f = \chi B(\mathbf{0},1)$; then

$$\beta_{r+1} = \int f d\mu_{r+1} = \int_0^1 \nu_r(\partial B(\mathbf{0},t))dt = \int_0^1 t^r \nu_r S_r dt = \frac{1}{r+1}\nu_r S_r$$

applying 264G to the maps $x \mapsto tx$, $x \mapsto \frac{1}{t}x$ from $\mathbb{R}^{r+1}$ to itself to see that $\nu_r(\partial B(\mathbf{0},t)) = t^r \nu_r S_r$ for $t > 0$.

265X Basic exercises (a) Let $r \ge 1$, and let $S_r(\alpha) = \{z : z \in \mathbb{R}^{r+1},\ \|z\| = \alpha\}$ be the r-sphere of radius α. Show that $\nu_r S_r(\alpha) = 2\pi\beta_{r-1}\alpha^r = (r+1)\beta_{r+1}\alpha^r$ for every $\alpha \ge 0$.

>(b) Let $r \ge 1$, and for $a \in [-1,1]$ set $C_a = \{z : z \in \mathbb{R}^{r+1},\ \|z\| = 1,\ \zeta_1 \ge a\}$, writing $z = (\zeta_1, \dots, \zeta_{r+1})$ as usual. Show that

$$\nu_r C_a = r\beta_r \int_0^{\arccos a} \sin^{r-1} t\, dt.$$

>(c) Again write $C_a = \{z : z \in S_r,\ \zeta_1 \ge a\}$, where $S_r \subseteq \mathbb{R}^{r+1}$ is the unit sphere. Show that, for any $a \in]0,1]$, $\nu_r C_a \le \frac{\nu_r S_r}{2(r+1)a^2}$. (*Hint*: calculate $\sum_{i=1}^{r+1} \int_{S_r} \|\xi_i\|^2 \nu_r(dx)$.)

>(d) Let $\phi :]0,1[\to \mathbb{R}^r$ be an injective differentiable function. Show that the 'length' or one-dimensional Hausdorff measure of $\phi[]0,1[]$ is just $\int_0^1 \|\phi'(t)\| dt$.

(e)(i) Show that if I is the identity $r \times r$ matrix and $z \in \mathbb{R}^r$, then $\det(I + zz') = 1 + \|z\|^2$. (*Hint*: induce on r.) (ii) Write U_{r-1} for the open unit ball in $\mathbb{R}^{r-1}$, where $r \ge 2$. Define $\phi : U_{r-1} \times \mathbb{R} \to S_r$ by setting

$$\phi\begin{pmatrix} x \\ \xi \end{pmatrix} = \begin{pmatrix} x \\ \theta(x)\cos\xi \\ \theta(x)\sin\xi \end{pmatrix},$$

where $\theta(x) = \sqrt{1 - \|x\|^2}$. Show that

$$\phi'\begin{pmatrix} x \\ \xi \end{pmatrix}' \phi'\begin{pmatrix} x \\ \xi \end{pmatrix} = \begin{pmatrix} I + \frac{1}{\theta(x)^2}xx' & \mathbf{0} \\ \mathbf{0} & \theta(x)^2 \end{pmatrix},$$

so that $J\begin{pmatrix} x \\ \xi \end{pmatrix} = 1$ for all $x \in U_{r-1}$, $\xi \in \mathbb{R}$. (iii) Hence show that the normalized r-dimensional Hausdorff measure of $\{y : y \in S_r,\ \sum_{i=1}^{r-1} \eta_i^2 < 1\}$ is just $2\pi\beta_{r-1}$, where β_{r-1} is the Lebesgue measure of U_{r-1}. (iv) By considering $\psi z = \begin{pmatrix} z \\ 0 \\ 0 \end{pmatrix}$ for $z \in S_{r-2}$, or otherwise, show that the normalized r-dimensional Hausdorff measure of S_r is $2\pi\beta_{r-1}$. (v) Setting $C_a = \{z : z \in \mathbb{R}^{r+1},\ \|z\| = 1,\ \zeta_r \ge a\}$, as in 265Xb and 265Xc, show that $\nu_r C_a = 2\pi\mu_{r-1}\{x : x \in \mathbb{R}^{r-1},\ \|x\| \le 1,\ \xi_1 \ge a\}$ for every $a \in [-1,1]$.

265Y Further exercises (a) Take $a < b$ in $\mathbb{R}$. (i) Show that $\phi : [a,b] \to \mathbb{R}^r$ is absolutely continuous in the sense of 264Yp iff all its coordinates $\phi_i : [a,b] \to \mathbb{R}$, for $i \le r$, are absolutely continuous in the sense of §225. (ii) Let $\phi : [a,b] \to \mathbb{R}^r$ be a continuous function, and set $F = \{x : x \in \,]a,b[\,, \phi$ is differentiable at $x\}$. Show that ϕ is absolutely continuous iff $\int_F \|\phi'(x)\| dx$ is finite and $\nu_1(\phi[[a,b] \setminus F]) = 0$, where ν_1 is normalized Hausdorff one-dimensional measure on $\mathbb{R}^r$. (*Hint*: 225K.) (iii) Show that if $\phi : [a,b] \to \mathbb{R}^r$ is absolutely continuous then $\nu_1^*(\phi[D]) \le \int_D \|\phi'(x)\| dx$ for every $D \subseteq [a,b]$, with equality if D is measurable and $\phi{\restriction}D$ is injective.

(b) Suppose that $a \le b$ in $\mathbb{R}$, and that $f : [a,b] \to \mathbb{R}$ is a continuous function of bounded variation with graph Γ_f. Show that the one-dimensional Hausdorff measure of Γ_f is $\operatorname{Var}_{[a,b]}(f) + \int_a^b (\sqrt{1+(f')^2} - |f'|)$. (*Hint*: set $E = \operatorname{dom} f'$ and examine $\Gamma_{f{\restriction}E}$, $\Gamma_f \setminus \Gamma_{f{\restriction}E}$ separately; use 264Xf and/or 264Yl.)

265 Notes and comments The proof of 265B seems to call on most of the second half of the alphabet. The idea is supposed to be straightforward enough. Because $T[\mathbb{R}^r]$ has dimension at most r, it can be rotated by an orthogonal transformation P into a subspace of the canonical r-dimensional subspace V, which is a natural copy of $\mathbb{R}^r$; the matrix R represents the copying process from V to $\mathbb{R}^r$, and ϕ or $P'R'$ is a copy of $\mathbb{R}^r$ onto a subspace including $T[\mathbb{R}^r]$. All this copying back and forth is designed to turn T into a linear operator $S : \mathbb{R}^r \to \mathbb{R}^r$ to which we can apply 263A, and part (b) of the proof is the check that we are copying the measures as well as the linear structures.

In 265D-265E I have tried to follow 263C-263D as closely as possible. In fact only one new idea is needed. When $s = r$, we have a special argument available to show that $\mu_r^*\phi[D] \le J\mu_r^*D + \epsilon\mu_r^*D$ (in the language of 263C) which applies whether or not $J = 0$. When $s > r$, this approach fails, because we can no longer approximate $\nu_r T[B]$ by $\nu_r G$ where $G \supseteq T[B]$ is open. (See part (b-i) of the proof of 263C.) I therefore turn to a different argument, valid only when $J > 0$, and accordingly have to find a separate method to show that $\{\phi(x) : x \in D, J(x) = 0\}$ is ν_r-negligible. Since we are working without restrictions on the dimensions r, s except that $r \le s$, we can use the trick of approximating $\phi : D \to \mathbb{R}^s$ by $\psi_\eta : D \to \mathbb{R}^{s+r}$, as in part (d) of the proof of 265E.

I give three methods by which the area of the r-sphere can be calculated; a bare-hands approach (265F), the surrounding-cylinder method (265Xe) and an important repeated-integral theorem (265G). The first two provide formulae for the area of a cap (265Xb, 265Xe(v)). The surrounding-cylinder method is attractive because the Jacobian comes out to be 1, that is, we have an inverse-measure-preserving function. I note that despite having developed a technique which allows irregular domains, I am still forced by the singularity in the function θ of 265Xe to take the sphere in two bites. Theorem 265G is a special case of the Coarea Theorem (Evans & Gariepy 92, §3.4; Federer 69, 3.2.12).

For the next step in the geometric theory of measures on Euclidean space, see Chapter 47 in Volume 4.

*266 The Brunn-Minkowski inequality

We now have most of the essential ingredients for a proof of the Brunn-Minkowski inequality (266C) in a strong form. I do not at present expect to use it in this treatise, but it is one of the basic results of geometric measure theory and from where we now stand is not difficult, so I include it here. The preliminary results on arithmetic and geometric means (266A) and essential closures (266B) are of great importance for other reasons.

266A Arithmetic and geometric means We shall need the following standard result.

Proposition If $u_0, \dots, u_n, p_0, \dots, p_n \in [0,\infty[$ and $\sum_{i=0}^n p_i = 1$, then $\prod_{i=0}^n u_i^{p_i} \le \sum_{i=0}^n p_i u_i$.

proof Induce on n. For $n = 0$, $p_0 = 1$ the result is trivial. If $n = 1$, then if $u_1 = 0$ the result is trivial (even if, as is standard in this book, we interpret 0^0 as 1). Otherwise, set $t = \frac{u_0}{u_1}$; then

$$t^{p_0} \le p_0 t + 1 - p_0 = p_0 t + p_1$$

(as in part (a) of the proof of 244E), so

$$u_0^{p_0} u_1^{p_1} = t^{p_0} u_1 \le p_0 t u_1 + p_1 u_1 = p_0 u_0 + p_1 u_1.$$

For the inductive step to $n \ge 2$, if $p_0 = \ldots = p_{n-1} = 0$ the result is trivial. Otherwise, set $q = p_0 + \ldots + p_{n-1} = 1 - p_n$; then

$$\prod_{i=0}^n u_i^{p_i} = (\prod_{i=0}^{n-1} u_i^{p_i/q})^q u_n^{p_n} \le (\sum_{i=0}^{n-1} \frac{p_i}{q} u_i)^q u_n^{p_n}$$

(by the inductive hypothesis)

$$\leq q(\sum_{i=0}^{n-1}\tfrac{p_i}{q}u_i) + p_n u_n$$

(by the two-term case just examined)

$$= \sum_{i=0}^{n} p_i u_i,$$

and the induction continues.

266B Proposition For any set $D \subseteq \mathbb{R}^r$ set

$$\mathrm{cl}^*D = \{x : \limsup_{\delta\downarrow 0}\tfrac{\mu^*(D\cap B(x,\delta))}{\mu B(x,\delta)} > 0\},$$

where μ is Lebesgue measure on $\mathbb{R}^r$.

(a) $D \setminus \mathrm{cl}^*D$ is negligible.

(b) $\mathrm{cl}^*D \subseteq \overline{D}$.

(c) cl^*D is a Borel set.

(d) $\mu(\mathrm{cl}^*D) = \mu^*D$.

(e) If $C \subseteq \mathbb{R}$ then $\overline{C} + \mathrm{cl}^*D \subseteq \mathrm{cl}^*(C+D)$, writing $C + D$ for $\{x + y : x \in C,\, y \in D\}$.

proof (a) 261Da.

(b) If $x \in \mathbb{R}^r \setminus \overline{D}$ then $D \cap B(x,\delta) = \emptyset$ for all small δ.

(c) The point is just that $(x,\delta) \mapsto \mu^*(D \cap B(x,\delta))$ is continuous. **P** For any x, $y \in \mathbb{R}^r$ and δ, $\eta \geq 0$ we have

$$\begin{aligned}|\mu^*(D \cap B(y,\eta)) - \mu^*(D \cap B(x,\delta))| &\leq \mu(B(y,\eta)\triangle B(x,\delta))\\ &= 2\mu(B(x,\delta) \cup B(y,\eta)) - \mu B(x,\delta) - \mu B(y,\eta)\\ &\leq \beta_r\bigl(2(\max(\delta,\eta) + \|x-y\|)^r - \delta^r - \eta^r\bigr)\end{aligned}$$

(where $\beta_r = \mu B(\mathbf{0},1)$)

$$\to 0$$

as $(y,\eta) \to (x,\delta)$. **Q** So

$$x \mapsto \limsup_{\delta\downarrow 0}\tfrac{\mu^*(D\cap B(x,\delta))}{\mu B(x,\delta)} = \inf_{\alpha\in\mathbb{Q},\alpha>0}\sup_{\beta\in\mathbb{Q},0<\beta\leq\alpha}\tfrac{1}{\beta_r\beta^r}\mu^*(D \cap B(x,\beta))$$

is Borel measurable, and

$$\mathrm{cl}^*D = \{x : \limsup_{\delta\downarrow 0}\tfrac{\mu^*(D\cap B(x,\delta))}{\mu B(x,\delta)} > 0\}$$

is a Borel set.

(d) By (c), $\mu(\mathrm{cl}^*D)$ is defined; by (a), $\mu(\mathrm{cl}^*D) \geq \mu^*D$. On the other hand, let E be a measurable envelope of D (132Ee); then 261Db tells us that

$$\limsup_{\delta\downarrow 0}\tfrac{\mu^*(D\cap B(x,\delta))}{\mu B(x,\delta)} \leq \limsup_{\delta\downarrow 0}\tfrac{\mu(E\cap B(x,\delta))}{\mu B(x,\delta)} = 0$$

for almost every $x \in \mathbb{R}^r \setminus E$, so $\mathrm{cl}^*D \setminus E$ is negligible and

$$\mu(\mathrm{cl}^*D) \leq \mu E = \mu^*D.$$

(e) If $x \in \overline{C}$ and $y \in \mathrm{cl}^*D$, set

$$\gamma = \frac{1}{3}\limsup_{\delta\downarrow 0}\tfrac{\mu^*(D\cap B(y,\delta))}{\mu B(y,\delta)} > 0.$$

For any $\eta > 0$, there is a $\delta \in \,]0,\eta]$ such that $\mu^*(D \cap B(y,\delta)) \geq 2\gamma\mu B(x,\delta)$. Let $\delta_1 \in [0,\delta[$ be such that $\delta^r - \delta_1^r \leq \gamma\delta^r$. Then there is an $x' \in C$ such that $\|x - x'\| \leq \delta - \delta_1$. In this case,

$$\begin{aligned}\mu^*((C + D) \cap B(x + y,\delta)) &\geq \mu^*((x' + D) \cap B(x' + y,\delta_1)) = \mu^*(D \cap B(y,\delta_1))\\ &\geq \mu^*(D \cap B(y,\delta)) - \mu B(y,\delta) + \mu B(y,\delta_1)\\ &\geq 2\beta_r\gamma\delta^r - \beta_r\delta^r + \beta_r\delta_1^r \geq \beta_r\gamma\delta^r.\end{aligned}$$

As η is arbitrary,

$$\limsup_{\delta\downarrow 0}\frac{\mu^*((C+D)\cap B(x+y,\delta))}{\mu B(y,\delta)}\ge\gamma$$

and $x+y\in\mathrm{cl}^*(C+D)$; as x and y are arbitrary, $\overline{C}+\mathrm{cl}^*D\subseteq\mathrm{cl}^*(C+D)$.

Remark In this context, cl^*D is called the **essential closure** of D.

266C Theorem Let A, $B\subseteq\mathbb{R}^r$ be non-empty sets, where $r\ge 1$ is an integer. If μ is Lebesgue measure on $\mathbb{R}^r$, and $A+B=\{x+y:x\in A,\ y\in B\}$, then $\mu^*(A+B)^{1/r}\ge(\mu^*A)^{1/r}+(\mu^*B)^{1/r}$.

proof (a) Consider first the case in which $A=[a,a'[$ and $B=[b,b'[$ are half-open intervals. In this case $A+B=[a+b,a'+b'[$; writing $a=(\alpha_1,\dots,\alpha_r)$, etc., as in §115, set

$$u_i=\frac{\alpha_i'-\alpha_i}{\alpha_i'+\beta_i'-\alpha_i-\beta_i},\quad v_i=\frac{\beta_i'-\beta_i}{\alpha_i'+\beta_i'-\alpha_i-\beta_i}$$

for each i. Then we have

$$\begin{aligned}(\mu A)^{1/r}+(\mu B)^{1/r}&=\prod_{i=0}^{r}(\alpha_i'-\alpha_i)^{1/r}+\prod_{i=0}^{r}(\beta_i'-\beta_i)^{1/r}\\&=\mu(A+B)^{1/r}(\prod_{i=1}^{r}u_i^{1/r}+\prod_{i=1}^{r}v_i^{1/r})\\&\le\mu(A+B)^{1/r}(\frac{1}{r}\sum_{i=1}^{r}u_i+\frac{1}{r}\sum_{i=1}^{r}v_i)\\&=\mu(A+B)^{1/r}.\end{aligned}$$

(266A)

(b) Now I show by induction on $m+n$ that if $A=\bigcup_{j=0}^m A_j$ and $B=\bigcup_{j=0}^n B_j$, where $\langle A_j\rangle_{j\le m}$ and $\langle B_j\rangle_{j\le n}$ are both disjoint families of non-empty half-open intervals, then $\mu(A+B)^{1/r}\ge(\mu A)^{1/r}+(\mu B)^{1/r}$. **P** The induction starts with the case $m=n=0$, dealt with in (a). For the inductive step to $m+n=l\ge 1$, one of m, n is non-zero; the argument is the same in both cases; suppose the former. Since $A_0\cap A_1=\emptyset$, there must be some $j\le r$ and $\alpha\in\mathbb{R}$ such that A_0 and A_1 are separated by the hyperplane $\{x:\xi_j=\alpha\}$. Set $A'=\{x:x\in A,\ \xi_j<\alpha\}$ and $A''=\{x:x\in A,\ \xi_j\ge\alpha\}$; then both A' and A'' are non-empty and can be expressed as the union of at most $m-1$ disjoint half-open intervals. Set $\gamma=\frac{\mu A'}{\mu A}\in\,]0,1[$. The function $\beta\mapsto\mu\{x:x\in B,\ \xi_j<\beta\}$ is continuous, so there is a $\beta\in\mathbb{R}$ such that $\mu B'=\gamma\mu B$, where $B'=\{x:x\in B,\ \xi_j<\beta\}$; set $B''=B\setminus B$. Then B' and B'' can be expressed as unions of at most n half-open intervals. By the inductive hypothesis,

$$\mu(A'+B')^{1/r}\ge(\mu A')^{1/r}+(\mu B')^{1/r},\quad\mu(A''+B'')^{1/r}\ge(\mu A'')^{1/r}+(\mu B'')^{1/r}.$$

Now $A'+B'\subseteq\{x:\xi_j<\alpha+\beta\}$, while $A''+B''\subseteq\{x:\xi_j\ge\alpha+\beta\}$. So

$$\begin{aligned}\mu(A+B)&\ge\mu(A'+B')+\mu(A''+B'')\\&\ge\big((\mu A')^{1/r}+(\mu B')^{1/r}\big)^r+\big((\mu A'')^{1/r}+(\mu B'')^{1/r}\big)^r\\&=\big((\gamma\mu A)^{1/r}+(\gamma\mu B)^{1/r}\big)^r+\big(((1-\gamma)\mu A)^{1/r}+((1-\gamma)\mu B)^{1/r}\big)^r\\&=\big((\mu A)^{1/r}+(\mu B)^{1/r}\big)^r.\end{aligned}$$

Taking rth roots, $\mu(A+B)^{1/r}\ge(\mu A)^{1/r}+(\mu B)^{1/r}$ and the induction proceeds. **Q**

(c) Now suppose that A and B are compact non-empty subsets of $\mathbb{R}^r$. Then $\mu(A+B)^{1/r}\ge(\mu A)^{1/r}+(\mu B)^{1/r}$. **P** $A+B$ is compact (because $A\times B\subseteq\mathbb{R}^r\times\mathbb{R}^r$ is compact, being closed and bounded, and addition is continuous, so we can use 2A2Eb). Let $\epsilon>0$. Let $G\supseteq A+B$ be an open set such that $\mu G\le\mu(A+B)+\epsilon$ (134Fa); then there is a $\delta>0$ such that $B(x,2\delta)\subseteq G$ for every $x\in A+B$ (2A2Ed). Let $n\in\mathbb{N}$ be such that $2^{-n}\sqrt{r}\le\delta$, and let A_1 be the union of all the half-open intervals of the form $[2^{-n}z,2^{-n}z+2^{-n}e[$ which meet A, where $z\in\mathbb{Z}^r$ and $e=(1,1,\dots,1)$. Then A_1 is a finite disjoint union of half-open intervals, $A\subseteq A_1$ and every point of A_1 is within a distance δ of some point of A. Similarly, we can find a set B_1, a finite disjoint union of half-open intervals, including B and such that every point of

B_1 is within δ of some point of B. But this means that every point of A_1+B_1 is within a distance 2δ of some point of $A+B$, and belongs to G. Accordingly

$$(\mu(A+B)+\epsilon)^{1/r} \ge (\mu G)^{1/r} \ge \mu(A_1+B_1)^{1/r} \ge (\mu A_1)^{1/r} + (\mu B_1)^{1/r}$$

(by (b))

$$\ge (\mu A)^{1/r} + (\mu B)^{1/r}.$$

As ϵ is arbitrary, $\mu(A+B)^{1/r} \ge (\mu A)^{1/r} + (\mu B)^{1/r}$. **Q**

(d) Next suppose that A, $B \subseteq \mathbb{R}^r$ are Lebesgue measurable. Then

$$(\mu A)^{1/r} + (\mu B)^{1/r} = \sup\{(\mu K)^{1/r} + (\mu L)^{1/r} : K \subseteq A \text{ and } L \subseteq B \text{ are compact}\}$$

(134Fb)

$$\le \sup\{\mu(K+L)^{1/r} : K \subseteq A \text{ and } L \subseteq B \text{ are compact}\}$$

(by (c))

$$\le \mu^*(A+B)^{1/r}.$$

(e) For the penultimate step, suppose that A, $B \subseteq \mathbb{R}^r$ have non-zero outer Lebesgue measure. Consider cl^*A, cl^*B and $\mathrm{cl}^*(A+B)$ as defined in 266B. Then cl^*A and cl^*B are non-empty and their sum is included in $\mathrm{cl}^*(A+B)$, by 266Bb and 266Be. So we have

$$(\mu^*A)^{1/r} + (\mu^*B)^{1/r} = \mu(\mathrm{cl}^*A)^{1/r} + \mu(\mathrm{cl}^*B)^{1/r}$$

(266Bd)

$$\le \mu^*(\mathrm{cl}^*A + \mathrm{cl}^*B)^{1/r}$$

(by (d) here)

$$\le \mu(\mathrm{cl}^*(A+B))^{1/r} = \mu^*(A+B)^{1/r}.$$

(f) Finally, for arbitrary non-empty sets A, $B \subseteq \mathbb{R}^r$, note that if (for instance) A is negligible then we can take any $x \in A$ and see that

$$\mu^*(A+B)^{1/r} \ge \mu^*(x+B)^{1/r} = (\mu B)^{1/r} = (\mu^*A)^{1/r} + (\mu^*B)^{1/r},$$

and the result is similarly trivial if B is negligible. So all cases are covered.

266X Basic exercises (a) Let D, D' be subsets of $\mathbb{R}^r$. Show that (i) $\mathrm{cl}^*(D \cup D') = \mathrm{cl}^*D \cup \mathrm{cl}^*D'$ (ii) $\mathrm{cl}^*D = \mathrm{cl}^*D'$ iff D and D' have a common measurable envelope (iii) $\mathrm{cl}^*D \setminus \mathrm{cl}^*(\mathbb{R}^r \setminus D') \subseteq \mathrm{cl}^*(D \cap D')$ (iv) D is Lebesgue measurable iff $\mathrm{cl}^*D \cap \mathrm{cl}^*(\mathbb{R}^r \setminus D)$ is Lebesgue negligible (v) $D \cup \mathrm{cl}^*D$ is a measurable envelope of D (vi) $\mathrm{cl}^*(\mathrm{cl}^*D) = \mathrm{cl}^*D$.

(b) Show that, for a measurable set $E \subseteq \mathbb{R}$, cl^*E is just the set of real numbers which are not density points of $\mathbb{R} \setminus E$.

(c) In 266C, show that if A and B are similar convex sets in the same orientation then $A+B$ is a convex set similar to both and $\mu(A+B)^{1/r} = (\mu A)^{1/r} + (\mu B)^{1/r}$.

266 Notes and comments The proof of 266C is taken from Federer 69. There is a slightly specious generality in the form given here. If the sets A and B are at all irregular, then $\mu^*(A+B)^{1/r}$ is likely to be much greater than $(\mu^*A)^{1/r} + (\mu^*B)^{1/r}$. The critical case, in which A and B are similar convex sets, is much easier (266Xc). The theorem is therefore most useful when A and B are non-similar convex sets and we get a non-trivial estimate which may be hard to establish by other means. For this case we do not need 266B. Theorem 266C is an instructive example of the way in which the dimension r enters formulae when we seek results applying to general Euclidean spaces. There will be many more when I return to geometric measure theory in Chapter 47 of Volume 4.

Chapter 27

Probability theory

Lebesgue created his theory of integration in response to a number of problems in real analysis, and all his life seems to have thought of it as a tool for use in geometry and calculus (LEBESGUE 72, vols. 1 and 2). Remarkably, it turned out, when suitably adapted, to provide a solid foundation for probability theory. The development of this approach is generally associated with the name of Kolmogorov. It has so come to dominate modern abstract probability theory that many authors ignore all other methods. I do not propose to commit myself to any view on whether σ-additive measures are the only way to give a rigorous foundation to probability theory, or whether they are adequate to deal with all probabilistic ideas; there are some serious philosophical questions here, since probability theory, at least in its applied aspects, seeks to help us to understand the material world outside mathematics. But from my position as a measure theorist, it is incontrovertible that probability theory is among the central applications of the concepts and theorems of measure theory, and is one of the most vital sources of new ideas; and that every measure theorist must be alert to the intuitions which probabilistic methods can provide.

I have written the preceding paragraph in terms suggesting that 'probability theory' is somehow distinguishable from the rest of measure theory; this is another point on which I should prefer not to put forward any opinion as definitive. But undoubtedly there is a distinction, rather deeper than the elementary point that probability deals (almost) exclusively with spaces of measure 1. M.Loève argues persuasively (LOÈVE 77, §10.2) that the essence of probability theory is the artificial nature of the probability spaces themselves. In measure theory, when we wish to integrate a function, we usually feel that we have a proper function with a domain and values. In probability theory, when we take the expectation of a random variable, the variable is an 'observable' or 'the result of an experiment'; we are generally uncertain, or ignorant, or indifferent concerning the factors underlying the variable. Let me give an example from the theorems below. In the proof of the Central Limit Theorem (274F), I find that I need an auxiliary list $Z_0,\dots,Z_n$ of random variables, independent of each other and of the original sequence $X_0,\dots,X_n$. I create such a sequence by taking a product space $\Omega\times\Omega'$, and writing $X_i'(\omega,\omega')=X_i(\omega)$, while the Z_i are functions of ω'. Now the difference between the X_i and the X_i' is of a type which a well-trained analyst would ordinarily take seriously. We do not think that the function $x\mapsto x^2:[0,1]\to[0,1]$ is the same thing as the function $(x_1,x_2)\mapsto x_1^2:[0,1]^2\to[0,1]$. But a probabilist is likely to feel that it is positively pedantic to start writing X_i' instead of X_i. He did not believe in the space Ω in the first place, and if it turns out to be inadequate for his intuition he enlarges it without a qualm. Loève calls probability spaces 'fictions', 'inventions of the imagination' in Larousse's words; they are necessary in the models Kolmogorov has taught us to use, but we have a vast amount of freedom in choosing them, and in their essence they are nothing so definite as a set with points.

A probability space, therefore, is somehow a more shadowy entity in probability theory than it is in measure theory. The important objects in probability theory are random variables and distributions, particularly joint distributions. In this volume I shall deal exclusively with random variables which can be thought of as taking values in some power of $\mathbb{R}$; but this is not the central point. What is vital is that somehow the *codomain*, the potential set of values, of a random variable, is much better defined than its *domain*. Consequently our attention is focused not on any features of the artificial space which it is convenient to use as the underlying probability space – I write 'underlying', though it is the most superficial and easily changed aspect of the model – but on the distribution on the codomain induced by the random variable. Thus the Central Limit Theorem, which speaks only of distributions, is actually more important in applied probability than the Strong Law of Large Numbers, which claims to tell us what a long-term average will almost certainly be.

W.Feller (FELLER 66) goes even farther than Loève, and as far as possible works entirely with distributions, setting up machinery which enables him to go for long stretches without mentioning probability spaces at all. I make no attempt to emulate him. But the approach is instructive and faithful to the essence of the subject.

Probability theory includes more mathematics than can easily be encompassed in a lifetime, and I have selected for this introductory chapter the two limit theorems I have already mentioned, the Strong Law of Large Numbers and the Central Limit Theorem, together with some material on martingales (§§275-276). They illustrate not only the special character of probability theory – so that you will be able to form your own judgement on the remarks above – but also some of its chief contributions to 'pure' measure theory, the concepts of 'independence' and 'conditional expectation'.

271 Distributions

I start this chapter with a discussion of 'probability distributions', the probability measures on $\mathbb{R}^n$ defined by families $(X_1, \ldots, X_n)$ of random variables. I give the basic results describing the circumstances under which two distributions are equal (271G), integration with respect to a distribution (271E), and probability density functions (271H-271K).

271A Notation I have just spent some paragraphs on an attempt to describe the essential difference between probability theory and measure theory. But there is a quicker test by which you may discover whether your author is a measure theorist or a probabilist: open any page, and look for the phrases 'measurable function' and 'random variable', and the formulae '$\int f d\mu$' and '$\mathbb{E}(X)$'. The first member of each pair will enable you to diagnose 'measure' and the second 'probability', with little danger of error. So far in this treatise I have firmly used measure theorists' terminology, with a few individual quirks. But in a chapter on probability theory I find that measure-theoretic notation, while perfectly adequate in a formal sense, does such violence to the familiar formulations as to render them unnatural. Moreover, you must surely at some point – if you have not already done so – become familiar with probabilists' language. So in this chapter I will make a substantial step in that direction. Happily, I think that this can be done without setting up any direct conflicts, so that I shall be able, in later volumes, to call upon this work in whichever notation then seems appropriate, without needing to re-formulate it.

(a) So let (Ω, Σ, μ) be a probability space. I take the opportunity given by a new phrase to make a technical move. A **real-valued random variable** on Ω will be a member of $\mathcal{L}^0(\mu)$, as defined in 241A; that is, a real-valued function X defined on a conegligible subset of Ω such that X is measurable with respect to the completion $\hat{\mu}$ of μ, or, if you prefer, such that $X \restriction E$ is Σ-measurable for some conegligible set $E \subseteq \Omega$.[1]

(b) If X is a real-valued random variable on a probability space (Ω, Σ, μ), write $\mathbb{E}(X) = \int X \, d\mu$ if this is defined in $[-\infty, \infty]$ in the sense of Chapter 12 and §133. In this case I will call $\mathbb{E}(X)$ the **mean** or **expectation** of X. Thus we may say that 'X **has a finite expectation**' in place of 'X is integrable'. 133A says that '$\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ whenever $\mathbb{E}(X)$ and $\mathbb{E}(Y)$ and their sum are defined in $[-\infty, \infty]$', and 122P becomes 'a real-valued random variable X has a finite expectation iff $\mathbb{E}(|X|) < \infty$'.

(c) If X is a real-valued random variable with finite expectation, the **variance** of X is

$$\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}(X))^2 = \mathbb{E}(X^2 - 2\mathbb{E}(X)X + \mathbb{E}(X)^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

(Note that this formula shows that $\mathbb{E}(X)^2 \le \mathbb{E}(X^2)$; compare 244Xd(i).) $\mathrm{Var}(X)$ is finite iff $\mathbb{E}(X^2) < \infty$, that is, iff $X \in \mathcal{L}^2(\mu)$ (244A). In particular, $X + Y$ and cX have finite variance whenever X and Y do and $c \in \mathbb{R}$.

(d) I shall allow myself to use such formulae as

$$\Pr(X > a), \quad \Pr(X - \epsilon \le Y \le X + \delta),$$

where X and Y are random variables on the same probability space (Ω, Σ, μ), to mean respectively

$$\hat{\mu}\{\omega : \omega \in \mathrm{dom}\, X,\ X(\omega) > a\},$$

$$\hat{\mu}\{\omega : \omega \in \mathrm{dom}\, X \cap \mathrm{dom}\, Y,\ X(\omega) - \epsilon \le Y(\omega) \le X(\omega) + \delta\},$$

writing $\hat{\mu}$ for the completion of μ as usual. There are two points to note here. First, Pr depends on $\hat{\mu}$, not on μ; in effect, the notation automatically directs us to complete the probability space (Ω, Σ, μ). I could, of course, equally well write

$$\Pr(X^2 + Y^2 > 1) = \mu^*\{\omega : \omega \in \mathrm{dom}\, X \cap \mathrm{dom}\, Y,\ X(\omega)^2 + Y(\omega)^2 > 1\},$$

taking μ^* to be the outer measure on Ω associated with μ (132B). Secondly, I will use this notation *only for predicates corresponding to Borel measurable sets*; that is to say, I shall write

$$\Pr(\psi(X_1, \ldots, X_n)) = \hat{\mu}\{\omega : \omega \in \bigcap_{i \le n} \mathrm{dom}\, X_i,\ \psi(X_1(\omega), \ldots, X_n(\omega))\}$$

only when the set

$$\{(\alpha_1, \ldots, \alpha_n) : \psi(\alpha_1, \ldots, \alpha_n)\}$$

is a Borel set in $\mathbb{R}^n$. Part of the reason for this restriction will appear in the next few paragraphs; $\Pr(\psi(X_1, \ldots, X_n))$ must be something calculable from knowledge of the joint distribution of $X_1, \ldots, X_n$, as defined in 271C. In fact we

[1] For an account of how this terminology became standard, see `http://www.dartmouth.edu/~chance/Doob/conversation.html`.

can safely extend the idea to 'universally measurable' predicates ψ, to be discussed in Volume 4. But it could happen that μ gave a measure to a set of the form $\{\omega : X(\omega) \in A\}$ for some exceedingly irregular set A, and in such a case it would be prudent to regard this as an accidental pathology of the probability space, and to treat it in a rather different way.

(I see that I have rather glibly assumed that the formula above defines $\Pr(\psi(X_1, \dots , X_n))$ for every Borel predicate ψ. This is a consequence of 271Bb below.)

271B Theorem Let (Ω, Σ, μ) be a probability space, and $X_1, \dots , X_n$ real-valued random variables on Ω. Set $\boldsymbol{X}(\omega) = (X_1(\omega), \dots , X_n(\omega))$ for $\omega \in \bigcap_{i\le n} \operatorname{dom} X_i$.

(a) There is a unique Radon measure ν on $\mathbb{R}^n$ such that

$$\nu\,]-\infty, a] = \Pr(X_i \le \alpha_i \text{ for every } i \le n)$$

whenever $a = (\alpha_1, \dots , \alpha_n) \in \mathbb{R}^n$, writing $]-\infty, a]$ for $\prod_{i\le n}]-\infty, \alpha_i]$;

(b) $\nu\mathbb{R}^n = 1$ and $\nu E = \hat\mu(\boldsymbol{X}^{-1}[E])$ whenever νE is defined, where $\hat\mu$ is the completion of μ; in particular, $\nu E = \Pr((X_1, \dots , X_n) \in E)$ for every Borel set $E \subseteq \mathbb{R}^n$.

proof Let $\hat\Sigma$ be the domain of $\hat\mu$, and set $D = \bigcap_{i\le n} \operatorname{dom} X_i = \operatorname{dom} \boldsymbol{X}$; then D is conegligible, so belongs to $\hat\Sigma$. Let $\hat\mu_D = \hat\mu\restriction\mathcal{P}D$ be the subspace measure on D (131B, 214B), and ν_0 the image measure $\hat\mu_D\boldsymbol{X}^{-1}$ (234D); let T be the domain of ν_0.

Write $\mathcal{B}$ for the algebra of Borel sets in $\mathbb{R}^n$. Then $\mathcal{B} \subseteq \mathrm{T}$. **P** For $i \le n$, $\alpha \in \mathbb{R}$ set $F_{i\alpha} = \{x : x \in \mathbb{R}^n, \xi_i \le \alpha\}$, $H_{i\alpha} = \{\omega : \omega \in \operatorname{dom} X_i,\ X_i(\omega) \le \alpha\}$. X_i is $\hat\Sigma$-measurable and its domain is in $\hat\Sigma$, so $H_{i\alpha} \in \hat\Sigma$, and $\boldsymbol{X}^{-1}[F_{i\alpha}] = D \cap H_{i\alpha}$ is $\hat\mu_D$-measurable. Thus $F_{i\alpha} \in \mathrm{T}$. As T is a σ-algebra of subsets of $\mathbb{R}^n$, $\mathcal{B} \subseteq \mathrm{T}$ (121J). **Q**

Accordingly $\nu_0\restriction\mathcal{B}$ is a measure on $\mathbb{R}^n$ with domain $\mathcal{B}$; of course $\nu_0\mathbb{R}^n = \hat\mu D = 1$. By 256C, the completion ν of $\nu_0\restriction\mathcal{B}$ is a Radon measure on $\mathbb{R}^n$, and $\nu\mathbb{R}^n = \nu_0\mathbb{R}^n = 1$.

For $E \in \mathcal{B}$,

$$\nu E = \nu_0 E = \hat\mu_D\boldsymbol{X}^{-1}[E] = \hat\mu\boldsymbol{X}^{-1}[E] = \Pr((X_1, \dots , X_n) \in E).$$

More generally, if $E \in \operatorname{dom}\nu$, then there are Borel sets E', E'' such that $E' \subseteq E \subseteq E''$ and $\nu(E'' \setminus E') = 0$, so that $\boldsymbol{X}^{-1}[E'] \subseteq \boldsymbol{X}^{-1}[E] \subseteq \boldsymbol{X}^{-1}[E'']$ and $\hat\mu(\boldsymbol{X}^{-1}[E''] \setminus \boldsymbol{X}^{-1}[E']) = 0$. This means that $\boldsymbol{X}^{-1}[E] \in \hat\Sigma$ and

$$\hat\mu\boldsymbol{X}^{-1}[E] = \hat\mu\boldsymbol{X}^{-1}[E'] = \nu E' = \nu E.$$

As for the uniqueness of ν, if ν' is any Radon measure on $\mathbb{R}^n$ such that $\nu'\,]-\infty, a] = \Pr(X_i \le \alpha_i\ \forall\ i \le n)$ for every $a \in \mathbb{R}^n$, then surely

$$\nu'\mathbb{R}^n = \lim_{k\to\infty} \nu'\,]-\infty, k\mathbf{1}] = \lim_{k\to\infty} \nu\,]-\infty, k\mathbf{1}] = 1 = \nu\mathbb{R}^n.$$

Also $\mathcal{I} = \{]-\infty, a] : a \in \mathbb{R}^n\}$ is closed under finite intersections, and ν and ν' agree on $\mathcal{I}$. By the Monotone Class Theorem (or rather, its corollary 136C), ν and ν' agree on the σ-algebra generated by $\mathcal{I}$, which is $\mathcal{B}$ (121J), and are identical (256D).

271C Definition Let (Ω, Σ, μ) be a probability space and $X_1, \dots , X_n$ real-valued random variables on Ω. By the (**joint**) **distribution** or **law** $\nu_{\boldsymbol{X}}$ of the family $\boldsymbol{X} = (X_1, \dots , X_n)$ I shall mean the Radon probability measure ν of 271B. If we think of $\boldsymbol{X}$ as a function from $\bigcap_{i\le n} \operatorname{dom} X_i$ to $\mathbb{R}^n$, then $\nu_{\boldsymbol{X}}E = \Pr(\boldsymbol{X} \in E)$ for every Borel set $E \subseteq \mathbb{R}^n$.

271D Remarks (a) The choice of the Radon probability measure $\nu_{\boldsymbol{X}}$ as 'the' distribution of $\boldsymbol{X}$, with the insistence that 'Radon measures' should be complete, is of course somewhat arbitrary. Apart from the general principle that one should always complete measures, these conventions fit better with some of the work in Volume 4 and with such results as 272G below.

(b) Observe that in order to speak of the distribution of a family $\boldsymbol{X} = (X_1, \dots , X_n)$ of random variables, it is essential that all the X_i should be based on the same probability space.

(c) I see that the language I have chosen allows the X_i to have different domains, so that the family $(X_1, \dots , X_n)$ may not be exactly identifiable with the corresponding function from $\bigcap_{i\le n} \operatorname{dom} X_i$ to $\mathbb{R}^n$. I hope however that using the same symbol $\boldsymbol{X}$ for both will cause no confusion.

(d) It is not useful to think of the whole image measure $\nu_0 = \hat\mu_D\boldsymbol{X}^{-1}$ in the proof of 271B as the distribution of $\boldsymbol{X}$, unless it happens to be equal to $\nu = \nu_{\boldsymbol{X}}$. The 'distribution' of a random variable is exactly that aspect of it which can be divorced from any consideration of the underlying space (Ω, Σ, μ), and the point of such results as 271K and 272G is that distributions can be calculated from each other, without going back to the relatively fluid and uncertain model of a random variable in terms of a function on a probability space.

(e) If $\boldsymbol{X} = (X_1, \dots, X_n)$ and $\boldsymbol{Y} = (Y_1, \dots, Y_n)$ are such that $X_i =_{\text{a.e.}} Y_i$ for each i, then

$$\{\omega : \omega \in \bigcap_{i \le n} \operatorname{dom} X_i,\ X_i(\omega) \le \alpha_i\ \forall\ i \le n\} \triangle \{\omega : \omega \in \bigcap_{i \le n} \operatorname{dom} Y_i,\ Y_i(\omega) \le \alpha_i\ \forall\ i \le n\}$$

is negligible, so

$$\begin{aligned}\Pr(X_i \le \alpha_i\ \forall\ i \le n) &= \hat\mu\{\omega : \omega \in \bigcap_{i \le n} \operatorname{dom} X_i,\ X_i(\omega) \le \alpha_i\ \forall\ i \le n\} \\ &= \Pr(Y_i \le \alpha_i\ \forall\ i \le n)\end{aligned}$$

for all $\alpha_0, \dots, \alpha_n \in \mathbb{R}$, and $\nu_{\boldsymbol{X}} = \nu_{\boldsymbol{Y}}$. This means that we can, if we wish, think of a distribution as a measure $\nu_{\boldsymbol{u}}$ where $\boldsymbol{u} = (u_0, \dots, u_n)$ is a finite sequence in $L^0(\mu)$. In the present chapter I shall not emphasize this approach, but it will always be at the back of my mind.

271E Measurable functions of random variables: Proposition Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be a family of random variables (as always in such a context, I mean them all to be on the same probability space (Ω, Σ, μ)); write $\mathrm{T}_{\boldsymbol{X}}$ for the domain of the distribution $\nu_{\boldsymbol{X}}$, and let h be a $\mathrm{T}_{\boldsymbol{X}}$-measurable real-valued function defined $\nu_{\boldsymbol{X}}$-a.e. on $\mathbb{R}^n$. Then we have a random variable $Y = h(X_1, \dots, X_n)$ defined by setting

$$h(X_1, \dots, X_n)(\omega) = h(X_1(\omega), \dots, X_n(\omega)) \text{ for every } \omega \in \boldsymbol{X}^{-1}[\operatorname{dom} h].$$

The distribution ν_Y of Y is the measure on $\mathbb{R}$ defined by the formula

$$\nu_Y F = \nu_{\boldsymbol{X}} h^{-1}[F]$$

for just those sets $F \subseteq \mathbb{R}$ such that $h^{-1}[F] \in \mathrm{T}_{\boldsymbol{X}}$. Also

$$\mathbb{E}(Y) = \int h\, d\nu_{\boldsymbol{X}}$$

in the sense that if one of these exists in $[-\infty, \infty]$, so does the other, and they are then equal.

proof (a)(i) Once again, write $(\Omega, \hat\Sigma, \hat\mu)$ for the completion of (Ω, Σ, μ). Since

$$\Omega \setminus \operatorname{dom} Y \subseteq \bigcup_{i \le n} (\Omega \setminus \operatorname{dom} X_i) \cup \boldsymbol{X}^{-1}[\mathbb{R}^n \setminus \operatorname{dom} h]$$

is negligible (using 271Bb), $\operatorname{dom} Y$ is conegligible. If $a \in \mathbb{R}$, then

$$E = \{x : x \in \operatorname{dom} h,\ h(x) \le a\} \in \mathrm{T}_{\boldsymbol{X}},$$

so

$$\{\omega : \omega \in \Omega,\ Y(\omega) \le a\} = \boldsymbol{X}^{-1}[E] \in \hat\Sigma.$$

As a is arbitrary, Y is $\hat\Sigma$-measurable, and is a random variable.

(ii) Let $\tilde h : \mathbb{R}^n \to \mathbb{R}$ be any extension of h to the whole of $\mathbb{R}^n$. Then $\tilde h$ is $\mathrm{T}_{\boldsymbol{X}}$-measurable, so the ordinary image measure $\nu_{\boldsymbol{X}} \tilde h^{-1}$, defined on $\{F : \tilde h^{-1}[F] \in \operatorname{dom} \nu_{\boldsymbol{X}}\}$, is a Radon probability measure on $\mathbb{R}$ (256G). But for any $A \subseteq \mathbb{R}$,

$$\tilde h^{-1}[A] \triangle h^{-1}[A] \subseteq \mathbb{R}^n \setminus \operatorname{dom} h$$

is $\nu_{\boldsymbol{X}}$-negligible, so $\nu_{\boldsymbol{X}} h^{-1}[F] = \nu_{\boldsymbol{X}} \tilde h^{-1}[F]$ if either is defined.

If $F \subseteq \mathbb{R}$ is a Borel set, then

$$\nu_Y F = \hat\mu\{\omega : Y(\omega) \in F\} = \hat\mu(\boldsymbol{X}^{-1}[h^{-1}[F]]) = \nu_{\boldsymbol{X}}(h^{-1}[F]).$$

So ν_Y and $\nu_{\boldsymbol{X}} \tilde h^{-1}$ agree on the Borel sets and are equal (256D again).

(b) Now apply Theorem 235E to the measures $\hat\mu$ and $\nu_{\boldsymbol{X}}$ and the function $\phi = \boldsymbol{X}$. We have

$$\int \chi(\boldsymbol{X}^{-1}[F]) d\hat\mu = \hat\mu(\boldsymbol{X}^{-1}[F]) = \nu_{\boldsymbol{X}} F$$

for every $F \in \mathrm{T}_{\boldsymbol{X}}$, by 271Bb. Because h is $\nu_{\boldsymbol{X}}$-virtually measurable and defined $\nu_{\boldsymbol{X}}$-a.e., 235Eb tells us that

$$\int h(\boldsymbol{X}) d\mu = \int h(\boldsymbol{X}) d\hat\mu = \int h\, d\nu_{\boldsymbol{X}}$$

whenever either side is defined in $[-\infty, \infty]$, which is exactly the result we need.

271F Corollary If X is a single random variable with distribution ν_X, then

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \, \nu_X(dx)$$

if either is defined in $[-\infty, \infty]$. Similarly

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 \, \nu_X(dx)$$

(whatever X may be). If X, Y are two random variables (on the same probability space!) then we have

$$\mathbb{E}(X \times Y) = \int xy \, \nu_{(X,Y)} d(x, y)$$

if either side is defined in $[-\infty, \infty]$.

271G Distribution functions (a) If X is a real-valued random variable, its **distribution function** is the function $F_X : \mathbb{R} \to [0, 1]$ defined by setting

$$F_X(a) = \Pr(X \le a) = \nu_X \,]-\infty, a]$$

for every $a \in \mathbb{R}$. (**Warning!** some authors prefer $F_X(a) = \Pr(X < a)$.) Observe that F_X is non-decreasing, that $\lim_{a \to -\infty} F_X(a) = 0$, that $\lim_{a \to \infty} F_X(a) = 1$ and that $\lim_{x \downarrow a} F_X(x) = F_X(a)$ for every $a \in \mathbb{R}$. By 271Ba, X and Y have the same distribution iff $F_X = F_Y$.

(b) If $X_1, \dots, X_n$ are real-valued random variables on the same probability space, their **(joint) distribution function** is the function $F_{\boldsymbol{X}} : \mathbb{R}^n \to [0, 1]$ defined by writing

$$F_{\boldsymbol{X}}(a) = \Pr(X_i \le \alpha_i \; \forall \, i \le n)$$

whenever $a = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ have the same distribution function, they have the same distribution, by the n-dimensional version of 271B.

271H Densities Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be a family of random variables, all defined on the same probability space. A **density function** for $(X_1, \dots, X_n)$ is a Radon-Nikodým derivative, with respect to Lebesgue measure, for the distribution $\nu_{\boldsymbol{X}}$; that is, a non-negative function f, integrable with respect to Lebesgue measure μ_L on $\mathbb{R}^n$, such that

$$\int_E f d\mu_L = \nu_{\boldsymbol{X}} E = \Pr(\boldsymbol{X} \in E)$$

for every Borel set $E \subseteq \mathbb{R}^n$ (256J) – if there is such a function, of course.

271I Proposition Let $\boldsymbol{X} = (X_1, \dots, X_n)$ be a family of random variables, all defined on the same probability space. Write μ_L for Lebesgue measure on $\mathbb{R}^n$.

(a) There is a density function for $\boldsymbol{X}$ iff $\Pr(\boldsymbol{X} \in E) = 0$ for every Borel set E such that $\mu_L E = 0$.

(b) A non-negative Lebesgue integrable function f is a density function for $\boldsymbol{X}$ iff $\int_{]-\infty, a]} f d\mu_L = \Pr(\boldsymbol{X} \in \,]-\infty, a])$ for every $a \in \mathbb{R}^n$.

(c) Suppose that f is a density function for $\boldsymbol{X}$, and $G = \{x : f(x) > 0\}$. Then if h is a Lebesgue measurable real-valued function defined almost everywhere in G,

$$\mathbb{E}(h(\boldsymbol{X})) = \int h \, d\nu_{\boldsymbol{X}} = \int h \times f d\mu_L$$

if any of the three integrals is defined in $[-\infty, \infty]$, interpreting $(h \times f)(x)$ as 0 if $f(x) = 0$ and $x \notin \operatorname{dom} h$.

proof (a) Apply 256J to the Radon probability measure $\nu_{\boldsymbol{X}}$.

(b) Of course the condition is necessary. If it is satisfied, then (by B.Levi's theorem)

$$\int f d\mu_L = \lim_{k \to \infty} \int_{]-\infty, k\mathbf{1}]} f d\mu_L = \lim_{k \to \infty} \nu_{\boldsymbol{X}} \,]-\infty, k\mathbf{1}] = 1.$$

So we have a Radon probability measure ν defined by writing

$$\nu E = \int_E f d\mu_L$$

whenever $E \cap \{x : f(x) > 0\}$ is Lebesgue measurable (256E). We are supposing that $\nu \,]-\infty, a] = \nu_{\boldsymbol{X}} \,]-\infty, a]$ for every $a \in \mathbb{R}^n$; by 271Ba, as usual, $\nu = \nu_{\boldsymbol{X}}$, so

$$\int_E f d\mu_L = \nu E = \nu_{\boldsymbol{X}} E = \Pr(\boldsymbol{X} \in E)$$

for every Borel set $E \subseteq \mathbb{R}^n$, and f is a density function for $\boldsymbol{X}$.

(c) By 256E, $\nu_{\boldsymbol{X}}$ is the indefinite-integral measure over μ associated with f. So, writing $G = \{x : f(x) > 0\}$, we have

$$\int h\,d\nu_{\boldsymbol{X}} = \int h \times f d\mu_L$$

whenever either is defined in $[-\infty, \infty]$ (235K). By 234La, h is $\mathrm{T}_{\boldsymbol{X}}$-measurable and defined $\nu_{\boldsymbol{X}}$-almost everywhere, where $\mathrm{T}_{\boldsymbol{X}} = \operatorname{dom}\nu_{\boldsymbol{X}}$, so $\mathbb{E}(h(\boldsymbol{X})) = \int h\,d\nu_{\boldsymbol{X}}$ by 271E.

271J The machinery developed in §263 is sufficient to give a very general result on the densities of random variables of the form $\phi(\boldsymbol{X})$, as follows.

Theorem Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a family of random variables, and $D \subseteq \mathbb{R}^n$ a Borel set such that $\Pr(\boldsymbol{X} \in D) = 1$. Let $\phi : D \to \mathbb{R}^n$ be a function which is differentiable relative to its domain everywhere in D; for $x \in D$, let $T(x)$ be a derivative of ϕ at x, and set $J(x) = |\det T(x)|$. Suppose that $J(x) \neq 0$ for each $x \in D$, and that $\boldsymbol{X}$ has a density function f; and suppose moreover that $\langle D_k\rangle_{k\in\mathbb{N}}$ is a disjoint sequence of Borel sets, with union D, such that $\phi\restriction D_k$ is injective for every k. Then $\phi(\boldsymbol{X})$ has a density function $g = \sum_{k=0}^{\infty} g_k$ where

$$\begin{aligned} g_k(y) &= \tfrac{f(\phi^{-1}(y))}{J(\phi^{-1}(y))} \text{ for } y \in \phi[D_k \cap \operatorname{dom} f], \\ &= 0 \text{ for } y \in \mathbb{R}^n \setminus \phi[D_k]. \end{aligned}$$

proof By 262Ia, ϕ is continuous, therefore Borel measurable, so $\phi(\boldsymbol{X})$ is a random variable.

For the moment, fix $k \in \mathbb{N}$ and a Borel set $F \subseteq \mathbb{R}^n$. By 263D(iii), $\phi[D_k]$ is measurable, and by 263D(ii) $\phi[D_k \setminus \operatorname{dom} f]$ is negligible. The function g_k is such that $f(x) = J(x)g_k(\phi(x))$ for every $x \in D_k \cap \operatorname{dom} f$, so by 263D(v) we have

$$\begin{aligned} \int_F g_k\,d\mu &= \int_{\phi[D_k]} g_k \times \chi F\,d\mu = \int_{D_k} J(x)g_k(\phi(x))\chi F(\phi(x))\mu(dx) \\ &= \int_{D_k \cap \phi^{-1}[F]} f d\mu = \Pr(\boldsymbol{X} \in D_k \cap \phi^{-1}[F]). \end{aligned}$$

(The integral $\int_{\phi[D_k]} g_k \times \chi F$ is defined because $\int_{D_k} J \times (g_k \times \chi F)\phi$ is defined, and the integral $\int g_k \times \chi F$ is defined because $\phi[D_k]$ is measurable and g is zero off $\phi[D_k]$.)

Now sum over k. Every g_k is non-negative, so by B.Levi's theorem (123A, 123Xa)

$$\begin{aligned} \int_F g\,d\mu &= \sum_{k=0}^{\infty} \int_F g_k\,d\mu = \sum_{k=0}^{\infty} \Pr(\boldsymbol{X} \in D_k \cap \phi^{-1}[F]) \\ &= \Pr(\boldsymbol{X} \in \phi^{-1}[F]) = \Pr(\phi(\boldsymbol{X}) \in F). \end{aligned}$$

As F is arbitrary, g is a density function for $\phi(\boldsymbol{X})$, as claimed.

271K The application of the last theorem to ordinary transformations is sometimes indirect, so I give an example.

Proposition Let X, Y be two random variables with a joint density function f. Then $X \times Y$ has a density function h, where

$$h(u) = \int_{-\infty}^{\infty} \frac{1}{|v|} f(\tfrac{u}{v}, v)dv$$

whenever this is defined in $\mathbb{R}$.

proof Set $\phi(x, y) = (xy, y)$ for x, $y \in \mathbb{R}^2$. Then ϕ is differentiable, with derivative $T(x, y) = \begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}$, so $J(x, y) = |\det T(x, y)| = |y|$. Set $D = \{(x, y) : y \neq 0\}$; then D is a conegligible Borel set in $\mathbb{R}^2$ and $\phi\restriction D$ is injective. Now $\phi[D] = D$ and $\phi^{-1}(u, v) = (\frac{u}{v}, v)$ for $v \neq 0$. So $\phi(X, Y) = (X \times Y, Y)$ has a density function g, where

$$g(u, v) = \tfrac{f(u/v, v)}{|v|} \text{ if } v \neq 0.$$

To find a density function for $X \times Y$, we calculate

$$\Pr(X \times Y \leq a) = \int_{]-\infty, a] \times \mathbb{R}} g = \int_{-\infty}^{a} \int_{-\infty}^{\infty} g(u, v)dv\,du = \int_{-\infty}^{a} h$$

by Fubini's theorem (252B, 252C). In particular, h is defined and finite almost everywhere; and by 271Ib it is a density function for $X \times Y$.

***271L** When a random variable is presented as the limit of a sequence of random variables the following can be very useful.

Proposition Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a sequence of real-valued random variables converging in measure to a random variable X (definition: 245A). Writing F_{X_n}, F_X for the distribution functions of X_n, X respectively,

$$F_X(a) = \inf_{b>a} \liminf_{n\to\infty} F_{X_n}(b) = \inf_{b>a} \limsup_{n\to\infty} F_{X_n}(b)$$

for every $a \in \mathbb{R}$.

proof Set $\gamma = \inf_{b>a} \liminf_{n\to\infty} F_{X_n}(b)$, $\gamma' = \inf_{b>a} \limsup_{n\to\infty} F_{X_n}(b)$.

(a) $F_X(a) \le \gamma$. **P** Take any $b > a$ and $\epsilon > 0$. Then there is an $n_0 \in \mathbb{N}$ such that $\Pr(|X_n - X| \ge b - a) \le \epsilon$ for every $n \ge n_0$ (245F). Now, for $n \ge n_0$,

$$F_X(a) = \Pr(X \le a) \le \Pr(X_n \le b) + \Pr(X_n - X \ge b - a) \le F_{X_n}(b) + \epsilon.$$

So $F_X(a) \le \liminf_{n\to\infty} F_{X_n}(b) + \epsilon$; as ϵ is arbitrary, $F_X(a) \le \liminf_{n\to\infty} F_{X_n}(b)$; as b is arbitrary, $F_X(a) \le \gamma$. **Q**

(b) $\gamma' \le F_X(a)$. **P** Let $\epsilon > 0$. Then there is a $\delta > 0$ such that $F_X(a + 2\delta) \le F_X(a) + \epsilon$ (271Ga). Next, there is an $n_0 \in \mathbb{N}$ such that $\Pr(|X_n - X| \ge \delta) \le \epsilon$ for every $n \ge n_0$. In this case, for $n \ge n_0$,

$$\begin{aligned} F_{X_n}(a+\delta) &= \Pr(X_n \le a+\delta) \le \Pr(X \le a + 2\delta) + \Pr(X - X_n \ge \delta) \\ &\le F_X(a + 2\delta) + \epsilon \le F_X(a) + 2\epsilon. \end{aligned}$$

Accordingly

$$\gamma' \le \limsup_{n\to\infty} F_{X_n}(a+\delta) \le F_X(a) + 2\epsilon.$$

As ϵ is arbitrary, $\gamma' \le F_X(a)$. **Q**

(c) Since of course $\gamma \le \gamma'$, we must have $F_X(a) = \gamma = \gamma'$, as claimed.

271X Basic exercises >(a) Let X be a real-valued random variable with finite expectation, and $\epsilon > 0$. Show that $\Pr(|X - \mathbb{E}(X)| \ge \epsilon) \le \frac{1}{\epsilon^2}\operatorname{Var}(X)$. (This is **Chebyshev's inequality**.)

>(b) Let $F : \mathbb{R} \to [0,1]$ be a non-decreasing function such that (i) $\lim_{a\to-\infty} F(a) = 0$ (ii) $\lim_{a\to\infty} F(a) = 1$ (iii) $\lim_{x\downarrow a} F(x) = F(a)$ for every $a \in \mathbb{R}$. Show that there is a unique Radon probability measure ν in $\mathbb{R}$ such that $F(a) = \nu\,]-\infty, a]$ for every $a \in \mathbb{R}$. (*Hint*: 114Xa.) Hence show that F is the distribution function of some random variable.

>(c) Let X be a real-valued random variable with a density function f. (i) Show that $|X|$ has a density function g_1 where $g_1(x) = f(x) + f(-x)$ whenever $x \ge 0$ and $f(x)$, $f(-x)$ are both defined, 0 otherwise. (ii) Show that X^2 has a density function g_2 where $g_2(x) = (f(\sqrt{x}) + f(-\sqrt{x}))/2\sqrt{x}$ whenever $x > 0$ and this is defined, 0 for other x. (iii) Show that if $\Pr(X = 0) = 0$ then $1/X$ has a density function g_3 where $g_3(x) = \frac{1}{x^2} f(\frac{1}{x})$ whenever this is defined. (iv) Show that if $\Pr(X < 0) = 0$ then $\sqrt{X}$ has a density function g_4 where $g_4(x) = 2xf(x^2)$ if $x \ge 0$ and $f(x^2)$ is defined, 0 otherwise.

>(d) Let X and Y be random variables with a joint density function $f : \mathbb{R}^2 \to \mathbb{R}$. Show that $X + Y$ has a density function h where $h(u) = \int f(u - v, v)dv$ for almost every u.

(e) Let X, Y be random variables with a joint density function $f : \mathbb{R}^2 \to \mathbb{R}$. Show that X/Y has a density function h where $h(u) = \int |v| f(uv, v)dv$ for almost every u.

(f) Devise an alternative proof of 271K by using Fubini's theorem and one-dimensional substitutions to show that

$$\int_a^b \int_{-\infty}^{\infty} \frac{1}{|v|} f(\tfrac{u}{v}, v)dv\,du = \int_{\{(u,v):a\le uv\le b\}} f$$

whenever $a \le b$ in $\mathbb{R}$.

271Y Further exercises (a) Let $\mathfrak{T}$ be the topology of $\mathbb{R}^{\mathbb{N}}$ and $\mathcal{B}$ the σ-algebra of Borel sets (256Ye). (i) Let $\mathcal{I}$ be the family of sets of the form

$$\{x : x \in \mathbb{R}^{\mathbb{N}},\ x(i) \le \alpha_i\ \forall\, i \le n\},$$

where $n \in \mathbb{N}$ and $\alpha_i \in \mathbb{R}$ for each $i \le n$. Show that $\mathcal{B}$ is the smallest family of subsets of $\mathbb{R}^{\mathbb{N}}$ such that (α) $\mathcal{I} \subseteq \mathcal{B}$ (β) $B \setminus A \in \mathcal{B}$ whenever A, $B \in \mathcal{B}$ and $A \subseteq B$ (γ) $\bigcup_{k\in\mathbb{N}} A_k \in \mathcal{B}$ for every non-decreasing sequence $\langle A_k\rangle_{k\in\mathbb{N}}$ in $\mathcal{B}$. (ii) Show that if μ, μ' are two totally finite measures defined on $\mathbb{R}^{\mathbb{N}}$, and μF and $\mu' F$ are defined and equal for every $F \in \mathcal{I}$, then μE and $\mu' E$ are defined and equal for every $E \in \mathcal{B}$. (iii) Show that if Ω is a set and Σ a σ-algebra of subsets of Ω and $X : \Omega \to \mathbb{R}^{\mathbb{N}}$ is a function, then $X^{-1}[E] \in \Sigma$ for every $E \in \mathcal{B}$ iff $\pi_i X$ is Σ-measurable for every $i \in \mathbb{N}$, where $\pi_i(x) = x(i)$ for each $x \in \mathbb{R}^{\mathbb{N}}$, $i \in \mathbb{N}$. (iv) Show that if $\boldsymbol{X} = \langle X_i\rangle_{i\in\mathbb{N}}$ is a sequence of real-valued random variables on a probability space (Ω, Σ, μ), then there is a unique probability measure $\nu^{\mathcal{B}}_{\boldsymbol{X}}$, with domain $\mathcal{B}$, such that $\nu^{\mathcal{B}}_{\boldsymbol{X}}\{x : x(i) \le \alpha_i \ \forall\ i \le n\} = \Pr(X_i \le \alpha_i \ \forall\ i \le n)$ for every $\alpha_0, \dots, \alpha_n \in \mathbb{R}$. (v) Under the conditions of (iv), show that there is a unique Radon measure $\nu_{\boldsymbol{X}}$ on $\mathbb{R}^{\mathbb{N}}$ (in the sense of 256Ye) such that $\nu_{\boldsymbol{X}}\{x : x(i) \le \alpha_i \ \forall\ i \le n\} = \Pr(X_i \le \alpha_i \ \forall\ i \le n)$ for every $\alpha_0, \dots, \alpha_n \in \mathbb{R}$.

(b) Let $F : \mathbb{R}^2 \to [0, 1]$ be a function. Show that the following are equiveridical: (i) F is the distribution function of some pair (X_1, X_2) of random variables (ii) there is a probability measure ν on $\mathbb{R}^2$ such that $\nu\,]-\infty, a] = F(a)$ for every $a \in \mathbb{R}^2$ (iii)(α) $F(\alpha_1, \alpha_2) + F(\beta_1, \beta_2) \ge F(\alpha_1, \beta_2) + F(\alpha_2, \beta_1)$ whenever $\alpha_1 \le \beta_1$ and $\alpha_2 \le \beta_2$ (β) $F(\alpha_1, \alpha_2) = \lim_{\xi_1\downarrow\alpha_1,\xi_2\downarrow\alpha_2} F(\xi_1, \xi_2)$ for every α_1, α_2 (γ) $\lim_{\alpha\to-\infty} F(\alpha, \beta) = \lim_{\alpha\to-\infty} F(\beta, \alpha) = 0$ for all β (δ) $\lim_{\alpha\to\infty} F(\alpha, \alpha) = 1$. (*Hint*: for non-empty half-open intervals $]a, b]$, set $\lambda\,]a, b] = F(\alpha_1, \alpha_2) + F(\beta_1, \beta_2) - F(\alpha_1, \beta_2) - F(\alpha_2, \beta_1)$, and continue as in 115B-115F.)

(c) Generalize (b) to higher dimensions, finding a suitable formula to stand in place of that in (iii-α) of (b).

(d) Let (Ω, Σ, μ) be a probability space and $\mathcal{F}$ a filter on $\mathcal{L}^0(\mu)$ converging to $X_0 \in \mathcal{L}^0(\mu)$ for the topology of convergence in measure. Show that, writing F_X for the distribution function of $X \in \mathcal{L}^0(\mu)$,

$$F_{X_0}(a) = \inf_{b>a} \liminf_{X\to\mathcal{F}} F_X(b) = \inf_{b>a} \limsup_{X\to\mathcal{F}} F_X(b)$$

for every $a \in \mathbb{R}$.

(e) Let X, Y be non-negative random variables with the same distribution, and $h : [0, \infty[\to [0, \infty[$ a non-decreasing function. Show that $\mathbb{E}(X \times hY) \le \mathbb{E}(Y \times hY)$. (*Hint*: in the language of 252Yo, $(Y \times hY)^* = Y^* \times (hY)^*$.)

271 Notes and comments Most of this section seems to have been taken up with technicalities. This is perhaps unsurprising in view of the fact that it is devoted to the relationship between a vector random variable $\boldsymbol{X}$ and the associated distribution $\nu_{\boldsymbol{X}}$, and this necessarily leads us into the minefield which I attempted to chart in §235. Indeed, I call on results from §235 twice; once in 271E, with a $\phi(\omega) = \boldsymbol{X}(\omega)$ and $J(\omega) = 1$, and once in 271I, with $\phi(x) = x$ and $J(x) = f(x)$.

Distribution functions of one-dimensional random variables are easily characterized (271Xb); in higher dimensions we have to work harder (271Yb-271Yc). Distributions, rather than distribution functions, can be described for infinite sequences of random variables (271Ya); indeed, these ideas can be extended to uncountable families, but this requires proper topological measure theory, and belongs in Volume 4.

The statement of 271J is lengthy, not to say cumbersome. The point is that many of the most important transformations ϕ are not themselves injective, but can easily be dissected into injective fragments (see, for instance, 271Xc and 263Xd). The point of 271K is that we frequently wish to apply the ideas here to transformations which are singular, and indeed change the dimension of the random variable. I have not given the theorems which make such applications routine and suggest rather that you seek out tricks such as that used in the proof of 271K, which in any case are necessary if you want amenable formulae. Of course other methods are available (271Xf).

272 Independence

I introduce the concept of 'independence' for families of events, σ-algebras and random variables. The first part of the section, down to 272G, amounts to an analysis of the elementary relationships between the three manifestations of the idea. In 272G I give the fundamental result that the joint distribution of a (finite) independent family of random variables is just the product of the individual distributions. Further expressions of the connexion between independence and product measures are in 272J, 272M and 272N. I give a version of the zero-one law (272O), and I end the section with a group of basic results from probability theory concerning sums and products of independent random variables (272R-272W).

272A Definitions Let (Ω,Σ,μ) be a probability space.

(a) A family $\langle E_i\rangle_{i\in I}$ in Σ is **(stochastically) independent** if

$$\mu(E_{i_1}\cap E_{i_2}\cap\ldots\cap E_{i_n})=\prod_{j=1}^n\mu E_{i_j}$$

whenever $i_1,\ldots,i_n$ are distinct members of I.

(b) A family $\langle\Sigma_i\rangle_{i\in I}$ of σ-subalgebras of Σ is **(stochastically) independent** if

$$\mu(E_1\cap E_2\cap\ldots\cap E_n)=\prod_{j=1}^n\mu E_j$$

whenever $i_1,\ldots,i_n$ are distinct members of I and $E_j\in\Sigma_{i_j}$ for every $j\le n$.

(c) A family $\langle X_i\rangle_{i\in I}$ of real-valued random variables on Ω is **(stochastically) independent** if

$$\Pr(X_{i_j}\le\alpha_j\text{ for every }j\le n)=\prod_{j=1}^n\Pr(X_{i_j}\le\alpha_j)$$

whenever $i_1,\ldots,i_n$ are distinct members of I and $\alpha_1,\ldots,\alpha_n\in\mathbb{R}$.

272B Remarks (a) This is perhaps the central contribution of probability theory to measure theory, and as such deserves the most careful scrutiny. The idea of 'independence' comes from outside mathematics altogether, in the notion of events which have independent causes. I suppose that 272G and 272M are the results below which most clearly show the measure-theoretic aspects of the concept. It is not an accident that both involve product measures; one of the wonders of measure theory is the fact that the same technical devices are used in establishing the probability theory of stochastic independence and the geometry of multi-dimensional volume.

(b) In the following paragraphs I will try to describe some relationships between the three notions of independence just defined. But it is worth noting at once the fact that, in all three cases, a family is independent iff all its finite subfamilies are independent. Consequently any subfamily of an independent family is independent. Another elementary fact which is immediate from the definitions is that if $\langle\Sigma_i\rangle_{i\in I}$ is an independent family of σ-algebras, and Σ'_i is a σ-subalgebra of Σ_i for each i, then $\langle\Sigma'_i\rangle_{i\in I}$ is an independent family.

(c) A useful reformulation of 272Ab is the following: A family $\langle\Sigma_i\rangle_{i\in I}$ of σ-subalgebras of Σ is independent iff

$$\mu(\bigcap_{i\in I}E_i)=\prod_{i\in I}\mu E_i$$

whenever $E_i\in\Sigma_i$ for every i and $\{i:E_i\ne\Omega\}$ is finite. (Here I follow the convention of 254F, saying that for a family $\langle\alpha_i\rangle_{i\in I}$ in $[0,1]$ we take $\prod_{i\in I}\alpha_i=1$ if $I=\emptyset$, and otherwise it is to be $\inf_{J\subseteq I,J\text{ is finite}}\prod_{i\in J}\alpha_j$.)

(d) In 272Aa-b I speak of sets $E_i\in\Sigma$ and algebras $\Sigma_i\subseteq\Sigma$. In fact (272Ac already gives a hint of this) we shall more often than not be concerned with $\hat\Sigma$ rather than with Σ, if there is a difference, where $(\Omega,\hat\Sigma,\hat\mu)$ is the completion of (Ω,Σ,μ).

272C The σ-subalgebra defined by a random variable To relate 272Ab to 272Ac we need the following notion. Let (Ω,Σ,μ) be a probability space and X a real-valued random variable defined on Ω. Write $\mathcal{B}$ for the σ-algebra of Borel subsets of $\mathbb{R}$, and Σ_X for

$$\{X^{-1}[F]:F\in\mathcal{B}\}\cup\{(\Omega\setminus\operatorname{dom}X)\cup X^{-1}[F]:F\in\mathcal{B}\}.$$

Then Σ_X is a σ-algebra of subsets of Ω. **P**

$$\emptyset=X^{-1}[\emptyset]\in\Sigma_X;$$

if $F\in\mathcal{B}$ then

$$\Omega\setminus X^{-1}[F]=(\Omega\setminus\operatorname{dom}X)\cup X^{-1}[\mathbb{R}\setminus F]\in\Sigma_X,$$

$$\Omega\setminus((\Omega\setminus\operatorname{dom}X)\cup X^{-1}[F])=X^{-1}[\mathbb{R}\setminus F]\in\Sigma_X;$$

if $\langle F_k\rangle_{k\in\mathbb{N}}$ is any sequence in $\mathcal{B}$ then

$$\bigcup_{k\in\mathbb{N}}X^{-1}[F_k]=X^{-1}[\bigcup_{k\in\mathbb{N}}F_k],$$

so

$$\bigcup_{k\in\mathbb{N}}X^{-1}[F_k],\quad(\Omega\setminus\operatorname{dom}X)\cup\bigcup_{k\in\mathbb{N}}X^{-1}[F_k]$$

belong to Σ_X. **Q**

Evidently Σ_X is the smallest σ-algebra of subsets of Ω, containing $\operatorname{dom} X$, for which X is measurable. Also Σ_X is a subalgebra of $\hat{\Sigma}$, where $\hat{\Sigma}$ is the domain of the completion of μ (271Aa).

Now we have the following result.

272D Proposition Let (Ω, Σ, μ) be a probability space and $\langle X_i \rangle_{i\in I}$ a family of real-valued random variables on Ω. For each $i \in I$, let Σ_i be the σ-algebra defined by X_i, as in 272C. Then the following are equiveridical:

(i) $\langle X_i \rangle_{i\in I}$ is independent;

(ii) whenever $i_1, \dots , i_n$ are distinct members of I and $F_1, \dots , F_n$ are Borel subsets of $\mathbb{R}$, then

$$\Pr(X_{i_j} \in F_j \text{ for every } j \le n) = \prod_{j=1}^{n} \Pr(X_{i_j} \in F_j);$$

(iii) whenever $\langle F_i \rangle_{i\in I}$ is a family of Borel subsets of $\mathbb{R}$, and $\{i : F_i \ne \mathbb{R}\}$ is finite, then

$$\hat{\mu}\bigl(\bigcap_{i\in I}(X_i^{-1}[F_i] \cup (\Omega \setminus \operatorname{dom} X_i))\bigr) = \prod_{i\in I} \Pr(X_i \in F_i),$$

where $\hat{\mu}$ is the completion of μ;

(iv) $\langle \Sigma_i \rangle_{i\in I}$ is independent with respect to $\hat{\mu}$.

proof (a)(i)$\Rightarrow$(ii) Write $\boldsymbol{X} = (X_{i_1}, \dots , X_{i_n})$. Write $\nu_{\boldsymbol{X}}$ for the joint distribution of $\boldsymbol{X}$, and for each $j \le n$ write ν_j for the distribution of X_{i_j}; let ν be the product of $\nu_1, \dots , \nu_n$ as described in 254A-254C. (I wrote §254 out as for infinite products. If you are interested only in finite products of probability spaces, which are adequate for our needs in this paragraph, I recommend reading §§251-252 with the mental proviso that all measures are probabilities, and then §254 with the proviso that the set I is finite.) By 256K, ν is a Radon measure on $\mathbb{R}^n$. (This is an induction on n, relying on 254N for assurance that we can regard ν as the repeated product $(\dots((\nu_1 \times \nu_2) \times \nu_3) \times \dots \nu_{n-1}) \times \nu_n$.) Then for any $a = (\alpha_1, \dots , \alpha_n) \in \mathbb{R}^n$, we have

$$\nu\,]-\infty, a] = \nu\bigl(\prod_{j=1}^{n}\,]-\infty, \alpha_j]\bigr) = \prod_{j=1}^{n} \nu_j\,]-\infty, \alpha_j]$$

(using 254Fb)

$$= \prod_{j=1}^{n} \Pr(X_{i_j} \le \alpha_j) = \Pr(X_{i_j} \le \alpha_j \text{ for every } j \le n)$$

(using the condition (i))

$$= \nu_{\boldsymbol{X}}\,]-\infty, a]\,.$$

By the uniqueness assertion in 271Ba, $\nu = \nu_{\boldsymbol{X}}$. In particular, if $F_1, \dots , F_n$ are Borel subsets of $\mathbb{R}$,

$$\begin{aligned}\Pr(X_{i_j} \in F_j \text{ for every } j \le n) &= \Pr(\boldsymbol{X} \in \prod_{j\le n} F_j) = \nu_{\boldsymbol{X}}(\prod_{j\le n} F_j)\\ &= \nu(\prod_{j\le n} F_j) = \prod_{j=1}^{n} \nu_j F_j = \prod_{j=1}^{n} \Pr(X_{i_j} \in F_j),\end{aligned}$$

as required.

(b)(ii)$\Rightarrow$(i) is trivial, if we recall that all sets $]-\infty, \alpha]$ are Borel sets, so that the definition of independence given in 272Ac is just a special case of (ii).

(c)(ii)$\Rightarrow$(iv) Assume (ii), and suppose that $i_1, \dots , i_n$ are distinct members of I and $E_j \in \Sigma_{i_j}$ for each $j \le n$. For each j, set $E'_j = E_j \cap \operatorname{dom} X_{i_j}$, so that E'_j may be expressed as $X_{i_j}^{-1}[F_j]$ for some Borel set $F_j \subseteq \mathbb{R}$. Then $\hat{\mu}(E_j \setminus E'_j) = 0$ for each j, so

$$\begin{aligned}\hat{\mu}(\bigcap_{1\le j\le n} E_j) &= \hat{\mu}(\bigcap_{1\le j\le n} E'_j) = \Pr(X_{i_1} \in F_1, \dots , X_{i_n} \in F_n)\\ &= \prod_{j=1}^{n} \Pr(X_{i_j} \in F_j)\end{aligned}$$

(using (ii))

$$= \prod_{i=1}^{n} \hat{\mu} E_j.$$

As $E_1, \dots , E_k$ are arbitrary, $\langle \Sigma_i \rangle_{i \in I}$ is independent.

(d)(iv)⇒(ii) Now suppose that $\langle \Sigma_i \rangle_{i \in I}$ is independent. If $i_1, \dots , i_n$ are distinct members of I and $F_1, \dots , F_n$ are Borel sets in $\mathbb{R}$, then $X_{i_j}^{-1}[F_j] \in \Sigma_{i_j}$ for each j, so

$$\begin{aligned}\Pr(X_{i_1} \in F_1, \dots , X_{i_n} \in F_n) &= \hat{\mu}(\bigcap_{1 \le j \le n} X_{i_j}^{-1}[F_j]) \\ &= \prod_{i=1}^{n} \hat{\mu} X_{i_j}^{-1}[F_j] = \prod_{j=1}^{n} \Pr(X_{i_j} \in F_j)\end{aligned}$$

.

(e) Finally, observe that (iii) is nothing but a re-formulation of (ii), because if $F_i = \mathbb{R}$ then $\Pr(X_i \in F_i) = 1$ and $X_i^{-1}[F_i] \cup (\Omega \setminus \operatorname{dom} X_i) = \Omega$.

272E Corollary Let $\langle X_i \rangle_{i \in I}$ be an independent family of real-valued random variables, and $\langle h_i \rangle_{i \in I}$ any family of Borel measurable functions from $\mathbb{R}$ to $\mathbb{R}$. Then $\langle h_i(X_i) \rangle_{i \in I}$ is independent.

proof Writing Σ_i for the σ-algebra defined by X_i, Σ'_i for the σ-algebra defined by $h(X_i)$, $h(X_i)$ is Σ_i-measurable (121Eg) so $\Sigma'_i \subseteq \Sigma_i$ for every i and $\langle \Sigma'_i \rangle_{i \in I}$ is independent, as in 272Bb.

272F Similarly, we can relate the definition in 272Aa to the others.

Proposition Let (Ω, Σ, μ) be a probability space, and $\langle E_i \rangle_{i \in I}$ a family in Σ. Set $\Sigma_i = \{\emptyset, E_i, \Omega \setminus E_i, \Omega\}$, the ($\sigma$-)algebra of subsets of Ω generated by E_i, and $X_i = \chi E_i$, the characteristic function of E_i. Then the following are equiveridical:

(i) $\langle E_i \rangle_{i \in I}$ is independent;
(ii) $\langle \Sigma_i \rangle_{i \in I}$ is independent;
(iii) $\langle X_i \rangle_{i \in I}$ is independent.

proof (i)⇒(iii) If $i_1, \dots , i_n$ are distinct members of I and $\alpha_1, \dots , \alpha_n \in \mathbb{R}$, then for each $j \le n$ the set $G_j = \{\omega : X_{i_j}(\omega) \le \alpha_j\}$ is either E_{i_j} or $\emptyset$ or Ω. If any G_j is empty, then

$$\Pr(X_{i_j} \le \alpha_j \text{ for every} j \le n\} = 0 = \prod_{j=1}^{n} \Pr(X_{i_j} \le \alpha_j).$$

Otherwise, set $K = \{j : G_j = E_{i_j}\}$; then

$$\begin{aligned}\Pr(X_{i_j} \le \alpha_j \text{ for every} j \le n\} &= \mu(\bigcap_{j \le n} G_j) = \mu(\bigcap_{j \in K} E_{i_j}) \\ &= \prod_{j \in K} \mu E_{i_j} = \prod_{j=1}^{n} \Pr(X_{i_j} \le \alpha_j).\end{aligned}$$

As $i_1, \dots , i_n$ and $\alpha_1, \dots , \alpha_n$ are arbitrary, $\langle X_i \rangle_{i \in I}$ is independent.

(iii)⇒(ii) follows from (i)⇒(iii) of 272D, because Σ_i is the σ-algebra defined by X_i.

(ii)⇒(i) is trivial, because $E_i \in \Sigma_i$ for each i.

Remark You will I hope feel that while the theory of product measures might be appropriate to 272D, it is surely rather heavy machinery to use on what ought to be a simple combinatorial problem like (iii)⇒(ii) of this proposition. I suggest that you construct an 'elementary' proof, and examine which of the ideas of the theory of product measures (and the Monotone Class Theorem, 136B) are actually needed here.

272G Distributions of independent random variables I have not tried to describe the 'joint distribution' of an infinite family of random variables. (Indications of how to deal with a countable family are offered in 271Ya and 272Yb. For uncountable families I will wait until §454 in Volume 4.) As, however, the independence of a family of random variables is determined by the behaviour of finite subfamilies, we can approach it through the following proposition.

Theorem Let $\boldsymbol{X} = (X_1, \ldots , X_n)$ be a finite family of real-valued random variables on a probability space. Let $\nu_{\boldsymbol{X}}$ be the corresponding distribution on $\mathbb{R}^n$. Then the following are equiveridical:

(i) $X_1, \ldots , X_n$ are independent;

(ii) $\nu_{\boldsymbol{X}}$ can be expressed as a product of n probability measures $\nu_1, \ldots , \nu_n$, one for each factor $\mathbb{R}$ of $\mathbb{R}^n$;

(iii) $\nu_{\boldsymbol{X}}$ is the product measure of $\nu_{X_1}, \ldots , \nu_{X_n}$, writing ν_{X_i} for the distribution of the random variable X_i.

proof (a)(i)⇒(iii) In the proof of (i)⇒(ii) of 272D above I showed that $\nu_{\boldsymbol{X}}$ is the product ν of $\nu_{X_1}, \ldots , \nu_{X_n}$.

(b)(iii)⇒(ii) is trivial.

(c)(ii)⇒(i) Suppose that $\nu_{\boldsymbol{X}}$ is expressible as a product $\nu_1 \times \ldots \times \nu_n$. Take $a = (\alpha_1, \ldots , \alpha_n)$ in $\mathbb{R}^n$. Then

$$\Pr(X_i \le \alpha_i \;\forall\; i \le n) = \Pr(\boldsymbol{X} \in \,]-\infty, a]) = \nu_{\boldsymbol{X}}(]-\infty, a]) = \prod_{i=1}^n \nu_i \,]-\infty, \alpha_i].$$

On the other hand, setting $F_i = \{(\xi_1, \ldots , \xi_n) : \xi_i \le \alpha_i\}$, we must have

$$\nu_i \,]-\infty, \alpha_i] = \nu_{\boldsymbol{X}} F_i = \Pr(\boldsymbol{X} \in F_i) = \Pr(X_i \le \alpha_i)$$

for each i. So we get

$$\Pr(X_i \le \alpha_i \text{ for every } i \le n) = \prod_{i=1}^n \Pr(X_i \le \alpha_i),$$

as required.

272H Corollary Suppose that $\langle X_i\rangle_{i\in I}$ is an independent family of real-valued random variables on a probability space (Ω, Σ, μ), and that for each $i \in I$ we are given another real-valued random variable Y_i on Ω such that $Y_i =_{\text{a.e.}} X_i$. Then $\langle Y_i\rangle_{i\in I}$ is independent.

proof For every distinct $i_1, \ldots , i_n \in I$, if we set $\boldsymbol{X} = (X_{i_1}, \ldots , X_{i_n})$ and $\boldsymbol{Y} = (Y_{i_1}, \ldots , Y_{i_n})$, then $\boldsymbol{X} =_{\text{a.e.}} \boldsymbol{Y}$, so $\nu_{\boldsymbol{X}}$, $\nu_{\boldsymbol{Y}}$ are equal (271De). By 272G, $Y_{i_1}, \ldots , Y_{i_n}$ must be independent because $X_{i_1}, \ldots , X_{i_n}$ are. As $i_1, \ldots , i_n$ are arbitrary, the whole family $\langle Y_i\rangle_{i\in I}$ is independent.

Remark It follows that we may speak of independent families in the space $L^0(\mu)$ of equivalence classes of random variables (241C), saying that $\langle X_i^\bullet\rangle_{i\in I}$ is independent iff $\langle X_i\rangle_{i\in I}$ is.

272I Corollary Suppose that $X_1, \ldots , X_n$ are independent real-valued random variables with density functions $f_1, \ldots , f_n$ (271H). Then $\boldsymbol{X} = (X_1, \ldots , X_n)$ has a density function f given by setting $f(x) = \prod_{i=1}^n f_i(\xi_i)$ whenever $x = (\xi_1, \ldots , \xi_n) \in \prod_{i\le n} \operatorname{dom}(f_i) \subseteq \mathbb{R}^n$.

proof For $n = 2$ this is covered by 253I; the general case follows by induction on n.

272J The most important theorems of the subject refer to independent families of random variables, rather than independent families of σ-algebras. The value of the concept of independent σ-algebras lies in such results as the following.

Proposition Let (Ω, Σ, μ) be a complete probability space, and $\langle\Sigma_i\rangle_{i\in I}$ a family of σ-subalgebras of Σ. For each $i \in I$ let μ_i be the restriction of μ to Σ_i, and let $(\Omega^I, \Lambda, \lambda)$ be the product probability space of the family $\langle(\Omega, \Sigma_i, \mu_i)\rangle_{i\in I}$. Define $\phi : \Omega \to \Omega^I$ by setting $\phi(\omega)(i) = \omega$ whenever $\omega \in \Omega$ and $i \in I$. Then ϕ is inverse-measure-preserving iff $\langle\Sigma_i\rangle_{i\in I}$ is independent.

proof This is virtually a restatement of 254Fb and 254G. (i) If ϕ is inverse-measure-preserving, $i_1, \ldots , i_n \in I$ are distinct and $E_j \in \Sigma_{i_j}$ for each j, then $\bigcap_{j\le n} E_{i_j} = \phi^{-1}[\{x : x(i_j) \in E_j \text{ for every } j \le n\}]$, so that

$$\mu(\bigcap_{j\le n} E_{i_j}) = \lambda\{x : x(i_j) \in E_j \text{ for every } j \le n\} = \prod_{j=1}^n \mu_{i_j} E_{i_j} = \prod_{j=1}^n \mu E_{i_j}.$$

(ii) If $\langle\Sigma_i\rangle_{i\in I}$ is independent, $E_i \in \Sigma_i$ for every $i \in I$ and $\{i : E_i \ne \Omega\}$ is finite, then

$$\mu\phi^{-1}[\prod_{i\in I} E_i] = \mu(\bigcap_{i\in I} E_i) = \prod_{i\in I} \mu E_i = \prod_{i\in I} \mu_i E_i.$$

So the conditions of 254G are satisfied and $\mu\phi^{-1}[W] = \lambda W$ for every $W \in \Lambda$.

272K Proposition Let (Ω, Σ, μ) be a probability space and $\langle\Sigma_i\rangle_{i\in I}$ an independent family of σ-subalgebras of Σ. Let $\langle J(s)\rangle_{s\in S}$ be a disjoint family of subsets of I, and for each $s \in S$ let $\tilde\Sigma_s$ be the σ-algebra of subsets of Ω generated by $\bigcup_{i\in J(s)} \Sigma_i$. Then $\langle\tilde\Sigma_s\rangle_{s\in S}$ is independent.

proof Let $(\Omega, \hat\Sigma, \hat\mu)$ be the completion of (Ω, Σ, μ). On Ω^I let λ be the product of the measures $\mu{\restriction}\Sigma_i$, and let $\phi : \Omega \to \Omega^I$ be the diagonal map, as in 272J. ϕ is inverse-measure-preserving for $\hat\mu$ and λ, by 272J.

We can identify λ with the product of $\langle\lambda_s\rangle_{s\in S}$, where for each $s\in S$ λ_s is the product of $\langle\mu\restriction\Sigma_i\rangle_{i\in J(s)}$ (254N). For $s\in S$, let Λ_s be the domain of λ_s, and set $\pi_s(x)=x\restriction J(s)$ for $x\in\Omega^I$, so that π_s is inverse-measure-preserving for λ and λ_s (254Oa), and $\phi_s=\pi_s\phi$ is inverse-measure-preserving for $\hat\mu$ and λ_s; of course ϕ_s is the diagonal map from Ω to $\Omega^{J(s)}$. Set $\Sigma_s^*=\{\phi_s^{-1}[H]:H\in\Lambda_s\}$. Then Σ_s^* is a σ-subalgebra of $\hat\Sigma$, and $\Sigma_s^*\supseteq\tilde\Sigma_s$, because

$$E=\phi_s^{-1}[\{x:x(i)\in E\}]\in\Sigma_s^*$$

whenever $i\in J(s)$a dn $E\in\Sigma_i$.

Now suppose that $s_1,\dots,s_n\in S$ are distinct and that $E_j\in\tilde\Sigma_{s_j}$ for each j. Then $E_j\in\Sigma_{s_j}^*$, so there are $H_j\in\Lambda_{s_j}$ such that $E_j=\phi_{s_j}^{-1}[H_j]$ for each j. Set

$$W=\{x:x\in\Omega^I,\ x\restriction J(s_j)\in H_j\text{ for every }j\le n\}.$$

Because we can identify λ with the product of the λ_s, we have

$$\lambda W=\prod_{j=1}^n\lambda_{s_j}H_j=\prod_{j=1}^n\hat\mu(\phi_{s_j}^{-1}[H_j])=\prod_{j=1}^n\hat\mu E_j=\prod_{j=1}^n\mu E_j.$$

On the other hand, $\phi^{-1}[W]=\bigcap_{j\le n}E_j$, so, because ϕ is inverse-measure-preserving,

$$\mu(\bigcap_{j\le n}E_j)=\hat\mu(\bigcap_{j\le n}E_j)=\lambda W=\prod_{j=1}^n\mu E_j.$$

As $E_1,\dots,E_n$ are arbitrary, $\langle\tilde\Sigma_s\rangle_{s\in S}$ is independent.

272L I give a typical application of this result as a sample.

Corollary Let $X,X_1,\dots,X_n$ be independent real-valued random variables and $h:\mathbb{R}^n\to\mathbb{R}$ a Borel function. Then X and $h(X_1,\dots,X_n)$ are independent.

proof Let Σ_X, Σ_{X_i} be the σ-algebras defined by X, X_i (272C). Then $\Sigma_X,\Sigma_{X_1},\dots,\Sigma_{X_n}$ are independent (272D). Let Σ^* be the σ-algebra generated by $\Sigma_{X_1}\cup\ldots\cup\Sigma_{X_n}$. Then 272K (perhaps working in the completion of the original probability space) tells us that Σ_X and Σ^* are independent. But every X_j is Σ^*-measurable so $Y=h(X_1,\dots,X_n)$ is Σ^*-measurable (121Kb); also $\operatorname{dom}Y\in\Sigma^*$, so $\Sigma_Y\subseteq\Sigma^*$ and Σ_X, Σ_Y are independent. By 272D again, X and Y are independent, as claimed.

Remark Nearly all of us, when teaching elementary probability theory, would invite our students to treat this corollary (with an explicit function h, of course) as 'obvious'. In effect, the proof here is a confirmation that the formal definition of 'independence' offered is a faithful representation of our intuition of independent events having independent causes.

272M Products of probability spaces and independent families of random variables We have already seen that the concept of 'independent random variables' is intimately linked with that of 'product measure'. I now give some further manifestations of the connexion.

Proposition Let $\langle(\Omega_i,\Sigma_i,\mu_i)\rangle_{i\in I}$ be a family of probability spaces, and (Ω,Σ,μ) their product.

(a) For each $i\in I$ write $\tilde\Sigma_i=\{\pi_i^{-1}[E]:E\in\Sigma_i\}$, where $\pi_i:\Omega\to\Omega_i$ is the coordinate map. Then $\langle\tilde\Sigma_i\rangle_{i\in I}$ is an independent family of σ-subalgebras of Σ.

(b) For each $i\in I$ let $\langle X_{ij}\rangle_{j\in J(i)}$ be an independent family of real-valued random variables on Ω_i, and for $i\in I$, $j\in J(i)$ write $\tilde X_{ij}(\omega)=X_{ij}(\omega(i))$ for those $\omega\in\Omega$ such that $\omega(i)\in\operatorname{dom}X_{ij}$. Then $\langle\tilde X_{ij}\rangle_{i\in I,j\in J(i)}$ is an independent family of random variables, and each $\tilde X_{ij}$ has the same distribution as the corresponding X_{ij}.

proof (a) It is easy to check that each $\tilde\Sigma_i$ is a σ-algebra of sets. The rest amounts just to recalling from 254Fb that if $J\subseteq I$ is finite and $E_i\in\Sigma_i$ for $i\in J$, then

$$\mu(\bigcap_{i\in J}\pi_i^{-1}[E_i])=\mu\{\omega:\omega(i)\in E_i\text{ for every }i\in I\}=\prod_{i\in I}\mu_iE_i$$

if we set $E_i=X_i$ for $i\in I\setminus J$.

(b) We know also that (Ω,Σ,μ) is the product of the completions $(\Omega_i,\hat\Sigma_i,\hat\mu_i)$ (254I). From this, we see that each $\tilde X_{ij}$ is defined μ-a.e., and is Σ-measurable, with the same distribution as X_{ij}. Now apply condition (iii) of 272D. Suppose that $\langle F_{ij}\rangle_{i\in I,j\in J(i)}$ is a family of Borel sets in $\mathbb{R}$, and that $\{(i,j):F_{ij}\ne\mathbb{R}\}$ is finite. Consider

$$E_i=\bigcap_{j\in J(i)}(X_{ij}^{-1}[F_{ij}]\cup(\Omega_i\setminus\operatorname{dom}X_{ij})),$$

$$E=\prod_{i\in I}E_i=\bigcap_{i\in I,j\in J(i)}(\tilde X_{ij}^{-1}[F_{ij}]\cup(\Omega\setminus\operatorname{dom}\tilde X_{ij})).$$

Because each family $\langle X_{ij}\rangle_{j\in J(i)}$ is independent, and $\{j:F_{ij}\ne\mathbb{R}\}$ is finite,

$$\hat{\mu}_i E_i = \prod_{j\in J(i)} \Pr(X_{ij} \in E_{ij})$$

for each $i \in I$. Because

$$\{i : E_i \neq \Omega_i\} \subseteq \{i : \exists\, j \in J(i),\, F_{ij} \neq \mathbb{R}\}$$

is finite,

$$\mu E = \prod_{i\in I} \hat{\mu}_i E_i = \prod_{i\in I, j\in J(i)} \Pr(\tilde{X}_{ij} \in F_{ij});$$

as $\langle F_{ij}\rangle_{i\in I, j\in J(i)}$ is arbitrary, $\langle \tilde{X}_{ij}\rangle_{i\in I, j\in J(i)}$ is independent.

Remark The formulation in (b) is more complicated than is necessary to express the idea, but is what is needed for an application below.

272N A special case of 272J is of particular importance in general measure theory, and is most useful in an adapted form.

Proposition Let (Ω, Σ, μ) be a complete probability space, and $\langle E_i\rangle_{i\in I}$ an independent family in Σ such that $\mu E_i = \frac{1}{2}$ for every $i \in I$. Define $\phi : \Omega \to \{0,1\}^I$ by setting $\phi(\omega)(i) = 1$ if $\omega \in E_i$, 0 if $\omega \in \Omega \setminus E_i$. Then ϕ is inverse-measure-preserving for the usual measure λ on $\{0,1\}^I$ (254J).

proof I use 254G again. For each $i \in I$ let Σ_i be the algebra $\{\emptyset, E_i, \Omega \setminus E_i, \Omega\}$; then $\langle \Sigma_i\rangle_{i\in I}$ is independent (272F). For $i \in I$ set $\phi_i(\omega) = \phi(\omega)(i)$. Let ν be the usual measure of $\{0,1\}$. Then it is easy to check that

$$\mu\phi_i^{-1}[H] = \frac{1}{2}\#(H) = \nu H$$

for every $H \subseteq \{0,1\}$. If $\langle H_i\rangle_{i\in I}$ is a family of subsets of $\{0,1\}$, and $\{i : H_i \neq \{0,1\}\}$ is finite, then

$$\mu\phi^{-1}[\bigcap_{i\in I} H_i] = \mu(\bigcap_{i\in I} \phi_i^{-1}[H_i]) = \prod_{i\in J} \mu\phi_i^{-1}[H_i]$$

(because $\phi^{-1}[H_i] \in \Sigma_i$ for each i, and $\langle \Sigma_i\rangle_{i\in I}$ is independent)

$$= \prod_{i\in I} \nu H_i = \lambda(\prod_{i\in I} H_i).$$

As $\langle H_i\rangle_{i\in I}$ is arbitrary, 254G gives the result.

272O Tail σ-algebras and the zero-one law I have never been able to make up my mind whether the following result is 'deep' or not. I think it is one of the many cases in mathematics where a theorem is surprising and exciting if one comes on it unprepared, but is natural and straightforward if one approaches it from the appropriate angle.

Proposition Let (Ω, Σ, μ) be a probability space and $\langle \Sigma_n\rangle_{n\in\mathbb{N}}$ an independent sequence of σ-subalgebras of Σ. Let Σ_n^* be the σ-algebra generated by $\bigcup_{m\geq n} \Sigma_m$ for each n, and set $\Sigma_\infty^* = \bigcap_{n\in\mathbb{N}} \Sigma_n^*$. Then μE is either 0 or 1 for every $E \in \Sigma_\infty^*$.

proof For each n, the family $(\Sigma_0, \dots, \Sigma_n, \Sigma_{n+1}^*)$ is independent, by 272K. So $(\Sigma_0, \dots, \Sigma_n, \Sigma_\infty^*)$ is independent, because $\Sigma_\infty^* \subseteq \Sigma_{n+1}^*$. But this means that every finite subfamily of $(\Sigma_\infty^*, \Sigma_0, \Sigma_1, \dots)$ is independent, and therefore that the whole family is (272Bb). Consequently $(\Sigma_\infty^*, \Sigma_0^*)$ must be independent, by 272K again.

Now if $E \in \Sigma_\infty^*$, then E also belongs to Σ_0^*, so we must have

$$\mu(E \cap E) = \mu E \cdot \mu E,$$

that is, $\mu E = (\mu E)^2$; so that $\mu E \in \{0,1\}$, as claimed.

272P To support the claim that somewhere we have achieved a non-trivial insight, I give a corollary, which will be fundamental to the understanding of the limit theorems in the next section, and does not seem to be obvious.

Corollary Let (Ω, Σ, μ) be a probability space, and $\langle X_n\rangle_{n\in\mathbb{N}}$ an independent sequence of real-valued random variables on Ω. Then

$$\limsup_{n\to\infty} \frac{1}{n+1}(X_0 + \ldots + X_n)$$

is almost everywhere constant – that is, there is some $u \in [-\infty, \infty]$ such that

$$\limsup_{n\to\infty}\frac{1}{n+1}(X_0+\ldots+X_n)=u$$

almost everywhere.

proof We may suppose that each X_n is Σ-measurable and defined everywhere in Ω, because (as remarked in 272H) changing the X_n on a negligible set does not affect their independence, and it affects $\limsup_{n\to\infty}\frac{1}{n+1}(X_0+\ldots+X_n)$ only on a negligible set. For each n, let Σ_n be the σ-algebra generated by X_n (272C), and Σ_n^* the σ-algebra generated by $\bigcup_{m\ge n}\Sigma_m$; set $\Sigma_\infty^*=\bigcap_{n\in\mathbb{N}}\Sigma_n^*$. By 272D, $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ is independent, so $\mu E\in\{0,1\}$ for every $E\in\Sigma_\infty^*$ (272O).

Now take any $a\in\mathbb{R}$ and set

$$E_a=\{\omega:\limsup_{m\to\infty}\frac{1}{m+1}(X_0(\omega)+\ldots+X_m(\omega))\le a\}.$$

Then

$$\limsup_{m\to\infty}\frac{1}{m+1}(X_0+\ldots+X_m)=\limsup_{m\to\infty}\frac{1}{m+1}(X_n+\ldots+X_{m+n}),$$

so

$$E_a=\{\omega:\limsup_{m\to\infty}\frac{1}{m+1}(X_n(\omega)+\ldots+X_{n+m}(\omega))\le a\}$$

belongs to Σ_n^* for every n, because X_i is Σ_n^*-measurable for every $i\ge n$. So $E\in\Sigma_\infty^*$ and

$$\Pr(\limsup_{m\to\infty}\frac{1}{m+1}(X_0+\ldots+X_m)\le a)=\mu E_a$$

must be either 0 or 1. Setting

$$u=\sup\{a:a\in\mathbb{R},\,\mu E_a=0\}$$

(allowing $\sup\emptyset=-\infty$ and $\sup\mathbb{R}=\infty$, as usual in such contexts), we see that

$$\limsup_{n\to\infty}\frac{1}{n+1}(X_0+\ldots+X_n)=u$$

almost everywhere.

***272Q** I add here a result which will be useful in Volume 5 and which gives further insight into the nature of large independent families.

Theorem Let (Ω,Σ,μ) be a probability space, and $\langle\Sigma_i\rangle_{i\in I}$ an independent family of σ-subalgebras of Σ. Let $\mathcal{E}\subseteq\Sigma$ be a family of measurable sets, and T the σ-algebra generated by $\mathcal{E}$. Then there is a set $J\subseteq I$ such that $\#(I\setminus J)\le\max(\omega,\#(\mathcal{E}))$ and T, $\langle\Sigma_j\rangle_{j\in J}$ are independent, in the sense that $\mu(F\cap\bigcap_{r\le n}E_r)=\mu F\cdot\prod_{r=1}^n\mu E_r$ whenever $F\in\mathrm{T}$, $j_1,\ldots,j_r$ are distinct members of J and $E_r\in\Sigma_{j_r}$ for each $r\le n$.

proof (a) As in 272J, give Ω^I the probability measure λ which is the product of the measures $\mu\restriction\Sigma_i$, and let $\phi:\Omega\to\Omega^I$ be the diagonal map, so that ϕ is inverse-measure-preserving for $\hat\mu$ and λ, where $\hat\mu$ is the completion of μ. Write Λ for the domain of λ. Set $\kappa=\max(\omega,\#(\mathcal{E}))$, and let $\mathcal{E}^*$ be the set $\{\bigcap_{r\le n}F_r:n\in\mathbb{N},\,F_r\in\mathcal{E}$ for every $r\le n\}$. Because $\#(\mathcal{E}^n)\le\kappa$ for each n (2A1Lc), $\#(\mathcal{E}^*)\le\kappa$ (2A1Ld). For each $F\in\mathcal{E}^*$, define $\nu_F:\Lambda\to[0,1]$ by setting $\nu_F W=\hat\mu(F\cap\phi^{-1}[W])$; then ν_F is countably additive and dominated by λ. It therefore has a Radon-Nikodým derivative h_F with respect to λ, so that $\hat\mu(F\cap\phi^{-1}[W])=\int_W h_F d\lambda$ for every $W\in\Lambda$ (232F). By 254Qc or 254Rb, we can find a function h'_F equal λ-almost everywhere to h_F and determined by coordinates in a countable set J_F, in the sense that $h'_F(w)=h'_F(w')$ whenever w, $w'\in\Omega^I$ and $w\restriction J_F=w'\restriction J_F$. (I am taking it for granted that we chose h'_F to be defined everywhere on Ω^I.)

(b) Set $J=I\setminus\bigcup_{F\in\mathcal{E}^*}J_F$; by 2A1Ld, $I\setminus J=\bigcup_{F\in\mathcal{E}^*}J_F$ has cardinal at most κ. If $F\in\mathcal{E}^*$, $j_1,\ldots,j_r$ are distinct members of J and $E_r\in\Sigma_{j_r}$ for each $r\le n$, $\mu(F\cap\bigcap_{r\le n}E_r)=\mu F\cdot\prod_{r=1}^n\mu E_r$. **P** Set $W=\{w:w\in\Omega^I,\,w(j_r)\in E_r$ for each $r\le n\}$. Then

$$\mu(F\cap\bigcap_{r\le n}E_r)=\hat\mu(F\cap\phi^{-1}[W])=\int_W h'_F d\lambda=\int h'_F\times\chi W\,d\lambda.$$

But observe that W is determined by coordinates in J, while h'_F is determined by coordinates in $J_F\subseteq I\setminus J$; putting 272Ma, 272K and 272R together (or otherwise), we have

$$\mu(F\cap\bigcap_{r\le n}E_r)=\int h'_F\times\chi W\,d\lambda=\int h'_F d\lambda\cdot\lambda W=\mu F\cdot\prod_{r=1}^n\mu E_r.\ \mathbf{Q}$$

(c) Now consider the family $\mathcal{A}$ of those sets $F\in\Sigma$ such that $\mu(F\cap\bigcap_{r\le n}E_r)=\mu F\cdot\prod_{r=1}^n\mu E_r$ whenever $j_1,\dots,j_n\in J$ are distinct and $E_r\in\Sigma_{j_r}$ for every $r\le n$. It is easy to check that $\mathcal{A}$ is a Dynkin class, and we have just seen that $\mathcal{A}$ includes $\mathcal{E}^*$; as $\mathcal{E}^*$ is closed under $\cap$, $\mathcal{A}$ includes the σ-algebra T of sets generated by $\mathcal{E}^*$ (136B). And this is just what the theorem asserts.

272R I must now catch up on some basic facts from elementary probability theory.

Proposition Let X, Y be independent real-valued random variables with finite expectation (271Ab). Then $\mathbb{E}(X\times Y)$ exists and is equal to $\mathbb{E}(X)\mathbb{E}(Y)$.

proof Let $\nu_{(X,Y)}$ be the joint distribution of the pair (X,Y). Then $\nu_{(X,Y)}$ is the product of the distributions ν_X and ν_Y (272G). Also $\int x\nu_X(dx)=\mathbb{E}(X)$ and $\int y\nu_Y(dy)=\mathbb{E}(Y)$ exist in $\mathbb{R}$ (271F). So

$$\int xy\nu_{(X,Y)}d(x,y)\text{ exists }=\mathbb{E}(X)\mathbb{E}(Y)$$

(253D). But this is just $\mathbb{E}(X\times Y)$, by 271E with $h(x,y)=xy$.

272S Bienaymé's Equality Let $X_1,\dots,X_n$ be independent real-valued random variables. Then $\mathrm{Var}(X_1+\ldots+X_n)=\mathrm{Var}(X_1)+\ldots+\mathrm{Var}(X_n)$.

proof (a) Suppose first that all the X_i have finite variance. Set $a_i=\mathbb{E}(X_i)$, $Y_i=X_i-a_i$, $X=X_1+\ldots+X_n$, $Y=Y_1+\ldots+Y_n$; then $\mathbb{E}(X)=a_1+\ldots+a_n$, so $Y=X-\mathbb{E}(X)$ and

$$\begin{aligned}\mathrm{Var}(X)&=\mathbb{E}(Y^2)=\mathbb{E}(\sum_{i=1}^n Y_i)^2\\&=\mathbb{E}(\sum_{i=1}^n\sum_{j=1}^n Y_i\times Y_j)=\sum_{i=1}^n\sum_{j=1}^n\mathbb{E}(Y_i\times Y_j).\end{aligned}$$

Now observe that if $i\ne j$ then $\mathbb{E}(Y_i\times Y_j)=\mathbb{E}(Y_i)\mathbb{E}(Y_j)=0$, because Y_i and Y_j are independent (by 272E) and we may use 272R, while if $i=j$ then

$$\mathbb{E}(Y_i\times Y_j)=\mathbb{E}(Y_i^2)=\mathbb{E}(X_i-\mathbb{E}(X_i))^2=\mathrm{Var}(X_i).$$

So

$$\mathrm{Var}(X)=\textstyle\sum_{i=1}^n\mathbb{E}(Y_i^2)=\sum_{i=1}^n\mathrm{Var}(X_i).$$

(b)(i) I show next that if $\mathrm{Var}(X_1+X_2)<\infty$ then $\mathrm{Var}(X_1)<\infty$. **P** We have

$$\iint(x+y)^2\nu_{X_1}(dx)\nu_{X_2}(dy)=\int(x+y)^2\nu_{(X_1,X_2)}(d(x,y))$$

(by 272G and Fubini's theorem)

$$=\mathbb{E}((X_1+X_2)^2)$$

(by 271E)

$$<\infty.$$

So there must be some $a\in\mathbb{R}$ such that $\int(x+a)^2\mu_{X_1}(dx)$ is finite, that is, $\mathbb{E}((X_1+a)^2)<\infty$; consequently $\mathbb{E}(X_1^2)$ and $\mathrm{Var}(X_1)$ are finite. **Q**

(ii) Now an easy induction (relying on 272L!) shows that if $\mathrm{Var}(X_1+\ldots+X_n)$ is finite, so is $\mathrm{Var}\,X_j$ for every j. Turning this round, if $\sum_{j=1}^n\mathrm{Var}(X_j)=\infty$, then $\mathrm{Var}(X_1+\ldots+X_n)=\infty$, and again the two are equal.

272T The distribution of a sum of independent random variables: Theorem Let X, Y be independent real-valued random variables on a probability space (Ω,Σ,μ), with distributions ν_X, ν_Y. Then the distribution of $X+Y$ is the convolution $\nu_X*\nu_Y$ (257A).

proof Set $\nu = \nu_X * \nu_Y$. Take $a \in \mathbb{R}$ and set $h = \chi\,]-\infty, a]$. Then h is ν-integrable, so

$$\nu\,]-\infty, a] = \int h\,d\nu = \int h(x+y)(\nu_X \times \nu_Y)(d(x,y))$$

(by 257B, writing $\nu_X \times \nu_Y$ for the product measure on $\mathbb{R}^2$)

$$= \int h(x+y)\nu_{(X,Y)}(d(x,y))$$

(by 272G, writing $\nu_{(X,Y)}$ for the joint distribution of (X,Y); this is where we use the hypothesis that X and Y are independent)

$$= \mathbb{E}(h(X+Y))$$

(applying 271E to the function $(x,y) \mapsto h(x+y)$)

$$= \Pr(X + Y \le a).$$

As a is arbitrary, $\nu_X * \nu_Y$ is the distribution of $X + Y$.

272U Corollary Suppose that X and Y are independent real-valued random variables, and that they have densities f and g. Then the convolution $f * g$ is a density function for $X + Y$.

proof By 257F, $f * g$ is a density function for $\nu_X * \nu_Y = \nu_{X+Y}$.

272V The following simple result will be very useful when we come to stochastic processes in Volume 4, as well as in the next section.

Etemadi's lemma (ETEMADI 96) Let $X_0, \dots, X_n$ be independent real-valued random variables. For $m \le n$, set $S_m = \sum_{i=0}^m X_i$. Then

$$\Pr(\sup_{m\le n} |S_m| \ge 3\gamma) \le 3\max_{m\le n} \Pr(|S_m| \ge \gamma)$$

for every $\gamma > 0$.

proof As in 272P, we may suppose that every X_i is a measurable function defined everywhere on a measure space Ω. Set $\alpha = \max_{m\le n} \Pr(|S_m| \ge \gamma)$. For each $r \le n$, set

$$E_r = \{\omega : |S_m(\omega)| < 3\gamma \text{ for every } m < r,\ |S_r(\omega)| \ge 3\gamma\}.$$

Then $E_0, \dots, E_n$ is a partition of $\{\omega : \max_{m\le n} |S_m(\omega)| \ge 3\gamma\}$. Set $E'_r = \{\omega : \omega \in E_r,\ |S_n(\omega)| < \gamma\}$. Then $E'_r \subseteq \{\omega : \omega \in E_r,\ |(S_n - S_r)(\omega)| > 2\gamma\}$. But E_r depends on $X_0, \dots, X_r$ so is independent of $\{\omega : |(S_n - S_r)(\omega)| > 2\gamma\}$, which can be calculated from $X_{r+1}, \dots, X_n$ (272K). So

$$\begin{aligned}\mu E'_r &\le \mu\{\omega : \omega \in E_r,\ |(S_n - S_r)(\omega)| > 2\gamma\} = \mu E_r \cdot \Pr(|S_n - S_r| > 2\gamma)\\ &\le \mu E_r(\Pr(|S_n| > \gamma) + \Pr(|S_r| > \gamma)) \le 2\alpha\mu E_r,\end{aligned}$$

and $\mu(E_r \setminus E'_r) \ge (1-2\alpha)\mu E_r$. On the other hand, $\langle E_r \setminus E'_r\rangle_{r\le n}$ is a disjoint family of sets all included in $\{\omega : |S_n(\omega)| \ge \gamma\}$. So

$$\alpha \ge \mu\{\omega : |S_n(\omega)| \ge \gamma\} \ge \sum_{r=0}^n \mu(E_r \setminus E'_r) \ge (1 - 2\alpha)\sum_{r=0}^n \mu E_r,$$

and

$$\Pr(\sup_{r\le n} |S_r| \ge 3\gamma) = \sum_{r=0}^n \mu E_r \le \min(1, \tfrac{\alpha}{1-2\alpha}) \le 3\alpha,$$

(considering $\alpha \le \frac{1}{3}$, $\alpha \ge \frac{1}{3}$ separately), as required.

***272W** The next result is a similarly direct application of the ideas of this section. While it will not be used in this volume, it is an accessible and useful representative of a very large number of results on tails of sums of independent random variables.

Theorem (HOEFFDING 63) Let $X_0, \dots, X_n$ be independent real-valued random variables such that $0 \le X_i \le 1$ a.e. for every i. Set $S = \frac{1}{n+1}\sum_{i=0}^n X_i$ and $a = \mathbb{E}(S)$. Then

$$\Pr(S - a \ge c) \le \exp(-2(n+1)c^2)$$

for every $c \geq 0$.

proof (a) Set $a_i = \mathbb{E}(X_i)$ for each i. If $b \geq 0$ and $i \leq n$, then

$$\mathbb{E}(e^{bX_i}) \leq \exp(ba_i + \frac{1}{8}b^2).$$

P Set $\phi(x) = e^{bx}$ for $x \in \mathbb{R}$. Then ϕ is convex, so

$$\phi(x) \leq 1 + x(e^b - 1)$$

whenever $x \in [0, 1]$,

$$\phi(X_i) \leq_{\text{a.e.}} 1 + (e^b - 1)X_i$$

and

$$\mathbb{E}(e^{bX_i}) = \mathbb{E}(\phi(X_i)) \leq 1 + (e^b - 1)a_i = e^{h(b)}$$

where $h(t) = \ln(1 - a_i + a_i e^t)$ for $t \in \mathbb{R}$. Now $h(0) = 0$,

$$h'(t) = \tfrac{a_i e^t}{1-a_i+a_i e^t} = 1 - \tfrac{1-a_i}{1-a_i+a_i e^t}, \quad h'(0) = a_i,$$

$$h''(t) = \tfrac{1-a_i}{1-a_i+a_i e^t} \cdot \tfrac{a_i e^t}{1-a_i+a_i e^t} \leq \frac{1}{4}$$

because $a_i e^t$ and $1 - a_i$ are both greater than or equal to 0. By Taylor's theorem with remainder, there is some $t \in [0, b]$ such that

$$h(b) = h(0) + bh'(0) + \frac{1}{2}b^2 h_i''(t) \leq ba_i + \frac{1}{8}b^2,$$

and

$$\mathbb{E}(e^{bX_i}) \leq \exp(ba_i + \frac{1}{8}b^2). \text{ } \mathbf{Q}$$

(b) Take any $b \geq 0$. Then

$$\Pr(S - a \geq c) = \Pr(\sum_{i=0}^{n}(X_i - a_i - c) \geq 0) \leq \mathbb{E}(\exp(b\sum_{i=0}^{n} X_i - a_i - c))$$

(because $\exp(b\sum_{i=0}^{n} X_i - a_i - c) \geq 1$ whenever $\sum_{i=0}^{n} X_i - a_i - c \geq 0$)

$$= e^{-(n+1)bc} \prod_{i=0}^{n} e^{-ba_i} \mathbb{E}(\prod_{i=0}^{n} \exp(bX_i))$$

$$= e^{-(n+1)bc} \prod_{i=0}^{n} e^{-ba_i} \prod_{i=0}^{n} \mathbb{E}(\exp(bX_i))$$

(because the random variables $\exp(bX_i)$ are independent, by 272E, so the expectation of the product is the product of the expectations, by 272R)

$$\leq e^{-(n+1)bc} \prod_{i=0}^{n} e^{-ba_i} \exp(ba_i + \frac{1}{8}b^2)$$

((a) above)

$$= \exp(-(n+1)bc + \tfrac{n+1}{8}b^2).$$

Now the minimum value of the quadratic $\frac{n+1}{8}b^2 - (n+1)cb$ is $-2(n+1)c^2$ when $b = 4c$, so $\Pr(S - a \geq c) \leq \exp(-2(n+1)c^2)$, as claimed.

272X Basic exercises (a) Let (Ω, Σ, μ) be an atomless probability space, and $\langle \epsilon_n \rangle_{n\in\mathbb{N}}$ any sequence in $[0, 1]$. Show that there is an independent sequence $\langle E_n \rangle_{n\in\mathbb{N}}$ in Σ such that $\mu E_n = \epsilon_n$ for every n. (*Hint*: 215D.)

>(b) Let $\langle X_i\rangle_{i\in I}$ be a family of real-valued random variables. Show that it is independent iff

$$\mathbb{E}(h_1(X_{i_1})\times\ldots\times h_n(X_{i_n}))=\prod_{j=1}^n\mathbb{E}(h_j(X_{i_j}))$$

whenever $i_1,\ldots,i_n$ are distinct members of I and $h_1,\ldots,h_n$ are Borel measurable functions from $\mathbb{R}$ to $\mathbb{R}$ such that $\mathbb{E}(h_j(X_{i_j}))$ are all finite.

(c) Write out a proof of 272F which does not use the theory of product measures.

(d) Let $\boldsymbol{X}=(X_1,\ldots,X_n)$ be a family of real-valued random variables all defined on the same probability space, and suppose that $\boldsymbol{X}$ has a density function f expressible in the form $f(\xi_1,\ldots,\xi_n)=f_1(\xi_1)f_2(\xi_2)\ldots f_n(\xi_n)$ for suitable functions $f_1,\ldots,f_n$ of one real variable. Show that $X_1,\ldots,X_n$ are independent.

(e) Let X_1, X_2 be independent real-valued random variables both with distribution ν and distribution function F. Set $Y=\max(X_1,X_2)$. Show that the distribution of Y is absolutely continuous with respect to ν, with a Radon-Nikodým derivative $F+F^-$, where $F^-(x)=\lim_{t\uparrow x}F(t)$ for every $x\in\mathbb{R}$.

(f) Use 254Sa and the idea of 272J to give another proof of 272O.

(g) Let (Ω,Σ,μ) be a probability space and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of Σ. Let Σ_∞ be the σ-algebra generated by $\bigcup_{n\in\mathbb{N}}\Sigma_n$. Let T be another σ-subalgebra of Σ such that Σ_n and T are independent for each n. Show that Σ_∞ and T are independent. (*Hint*: apply the Monotone Class Theorem to $\{E:\mu(E\cap F)=\mu E\cdot\mu F$ for every $F\in\mathrm{T}\}$.) Use this to prove 272O.

(h) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a sequence of real-valued random variables and Y a real-valued random variable such that Y and X_n are independent for each $n\in\mathbb{N}$. Suppose that $\Pr(Y\in\mathbb{N})=1$ and that $\sum_{n=0}^\infty\Pr(Y\ge n)\mathbb{E}(|X_n|)$ is finite. Set $Z=\sum_{n=0}^Y X_n$ (that is, $Z(\omega)=\sum_{n=0}^{Y(\omega)}X_n(\omega)$ whenever $\omega\in\operatorname{dom}Y$ is such that $Y(\omega)\in\mathbb{N}$ and $\omega\in\operatorname{dom}X_n$ for every $n\le Y(\omega)$). (i) Show that $\mathbb{E}(Z)=\sum_{n=0}^\infty\Pr(Y\ge n)\mathbb{E}(X_n)$. (*Hint*: set $X_n'(\omega)=X_n(\omega)$ if $Y(\omega)\ge n$, 0 otherwise.) (ii) Show that if $\mathbb{E}(X_n)=\gamma$ for every $n\in\mathbb{N}$ then $\mathbb{E}(Z)=\gamma\mathbb{E}(Y)$. (This is **Wald's equation**.)

(i) Let $X_1,\ldots,X_n$ be independent real-valued random variables. Show that if $X_1+\ldots+X_n$ has finite expectation so does every X_j. (*Hint*: part (b) of the proof of 272S.)

>(j) Let X and Y be independent real-valued random variables with densities f and g. Show that $X\times Y$ has a density function h where $h(x)=\int_{-\infty}^\infty\frac{1}{|y|}g(y)f(\frac{x}{y})dy$ for almost every x. (*Hint*: 271K.)

(k) Let $X_0,\ldots,X_n$ be independent real-valued random variables such that $d_i\le X_i\le d_i'$ a.e. for every i. (i) Show that if $b\ge0$ then $\mathbb{E}(e^{bX_i})\le\exp(ba_i+\frac{1}{8}b^2(d_i'-d_i)^2)$ for each i, where $a_i=\mathbb{E}(X_i)$. (ii) Set $S=\frac{1}{n+1}\sum_{i=0}^nX_i$ and $a=\mathbb{E}(S)$. Show that

$$\Pr(S-a\ge c)\le\exp(-\tfrac{2(n+1)^2c^2}{d})$$

for every $c\ge0$, where $d=\sum_{i=0}^n(d_i'-d_i)^2$.

(l) Suppose that $X_0,\ldots,X_n$ are independent real-valued random variables, all with expectation 0, such that $\Pr(|X_i|\le1)=1$ for every i. Set $S=\frac{1}{\sqrt{n+1}}\sum_{i=0}^nX_i$. Show that $\Pr(S\ge c)\le\exp(-c^2/2)$ for every $c\ge0$.

272Y Further exercises (a) Let $X_0,\ldots,X_n$ be independent real-valued random variables with distributions $\nu_0,\ldots,\nu_n$ and distribution functions $F_0,\ldots,F_n$. Show that, for any Borel set $E\subseteq\mathbb{R}$,

$$\Pr(\sup_{i\le n}X_i\in E)=\sum_{i=0}^n\int_E\prod_{j=0}^{i-1}F_j^-(x)\prod_{j=i+1}^nF_j(x)\nu_i(dx),$$

where $F_j^-(x)=\lim_{t\uparrow x}F_j(t)$ for each j, and we interpret the empty products $\prod_{j=0}^{-1}F_j^-(x)$, $\prod_{j=n+1}^nF_j(x)$ as 1.

(b) Let $\boldsymbol{X}=\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables on a complete probability space (Ω,Σ,μ). Let $\mathcal{B}$ be the Borel σ-algebra of $\mathbb{R}^{\mathbb{N}}$ (271Ya). Let $\nu_{\boldsymbol{X}}^{(\mathcal{B})}$ be the probability measure with domain $\mathcal{B}$ defined by setting $\nu_{\boldsymbol{X}}^{(\mathcal{B})}E=\mu\boldsymbol{X}^{-1}[E]$ for every $E\in\mathcal{B}$, and write $\nu_{\boldsymbol{X}}$ for the completion of $\nu_{\boldsymbol{X}}^{(\mathcal{B})}$. Show that $\nu_{\boldsymbol{X}}$ is just the product of the distributions ν_{X_n}.

(c) Let $X_1,\dots,X_n$ be real-valued random variables such that for each $j<n$ the family

$$(X_1,\dots,X_j,-X_{j+1},\dots,-X_n)$$

has the same joint distribution as the original family $(X_1,\dots,X_n)$. Set $S_j=X_1+\ldots+X_j$ for each $j\le n$. (i) Show that for any $a\ge 0$

$$\Pr(\sup_{1\le j\le n}|S_j|\ge a)\le 2\Pr(|S_n|\ge a).$$

(*Hint*: show that if $E_j=\{\omega:\omega\in\bigcap_{i\le n}\operatorname{dom}X_i,\ |S_i(\omega)|<a \text{ for } i<j,\ |S_j(\omega)|\ge a\}$ then $\mu\{\omega:\omega\in E_j,\ |S_n(\omega)|\ge |S_j(\omega)|\}\ge\frac{1}{2}\mu E_j$.) (ii) Show that $\mathbb{E}(\sup_{j\le n}|S_j|)\le 2\mathbb{E}(|S_n|)$. (iii) Show that $\mathbb{E}(\sup_{i\le n}S_i^2)\le 2\mathbb{E}(S_n^2)$.

(d) Let $\langle X_i\rangle_{i\in\mathbb{N}}$ be an independent sequence of real-valued random variables, and set $S_n=\sum_{i=0}^n X_i$ for each n. Show that if $\langle S_n\rangle_{n\in\mathbb{N}}$ converges to S in $\mathcal{L}^0$ for the topology of convergence in measure, then $\langle S_n\rangle_{n\in\mathbb{N}}$ converges to S a.e.

(e) Let (Ω,Σ,μ) be a probability space.

(i) Let $\langle E_n\rangle_{n\in\mathbb{N}}$ be an independent sequence in Σ. Show that for any real-valued random variable X with finite expectation,

$$\lim_{n\to\infty}\int_{E_n}X\,d\mu-\mu E_n\mathbb{E}(X)=0.$$

(*Hint*: let T_0 be the subalgebra of Σ generated by $\{E_n:n\in\mathbb{N}\}$ and T the σ-subalgebra of Σ generated by $\{E_n:n\in\mathbb{N}\}$. Start by considering $X=\chi E$ for $E\in\mathrm{T}_0$ and then $X=\chi E$ for $E\in\mathrm{T}$. Move from $\mathcal{L}^1(\mu\restriction\mathrm{T})$ to $\mathcal{L}^1(\mu)$ by using conditional expectations.)

(ii) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a uniformly integrable independent sequence of real-valued random variables on Ω. Show that for any bounded real-valued random variable Y,

$$\lim_{n\to\infty}\mathbb{E}(X_n\times Y)-\mathbb{E}(X_n)\mathbb{E}(Y)=0.$$

(iii) Suppose that $1<p\le\infty$ and set $q=p/(p-1)$ (taking $q=1$ if $p=\infty$). Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables with $\sup_{n\in\mathbb{N}}\|X_n\|_p<\infty$, and Y a real-valued random variable with $\|Y\|_q<\infty$. Show that

$$\lim_{n\to\infty}\mathbb{E}(X_n\times Y)-\mathbb{E}(X_n)\mathbb{E}(Y)=0.$$

(f) Let (Ω,Σ,μ) be a probability space and $\langle Z_n\rangle_{n\in\mathbb{N}}$ a sequence of random variables on Ω such that $\Pr(Z_n\in\mathbb{N})=1$ for each n, and $\Pr(Z_m=Z_n)=0$ for all $m\ne n$. Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a sequence of real-valued random variables on Ω, all with the same distribution ν, and independent of each other and the Z_n, in the sense that if Σ_n is the σ-algebra defined by X_n, and T_n the σ-algebra defined by Z_n, and T is the σ-algebra generated by $\bigcup_{n\in\mathbb{N}}\mathrm{T}_n$, then $(\mathrm{T},\Sigma_0,\Sigma_1,\dots)$ is independent. Set $Y_n(\omega)=X_{Z_n(\omega)}(\omega)$ whenever this is defined, that is, $\omega\in\operatorname{dom}Z_n$, $Z_n(\omega)\in\mathbb{N}$ and $\omega\in\operatorname{dom}X_{Z_n(\omega)}$. Show that $\langle Y_n\rangle_{n\in\mathbb{N}}$ is an independent sequence of random variables and that every Y_n has the distribution ν.

(g) Show that all the ideas of this section apply equally to complex-valued random variables, subject to suitable adjustments (to be devised).

(h) Develop a theory of independence for random variables taking values in $\mathbb{R}^r$, following through as many as possible of the ideas of this section.

272 Notes and comments This section is lengthy for two reasons: I am trying to pack in the basic results associated with one of the most fertile concepts of mathematics, and it is hard to know where to stop; and I am trying to do this in language appropriate to abstract measure theory, insisting on a variety of distinctions which are peripheral to the essential ideas. For while I am prepared to be flexible on the question of whether the letter X should denote a space or a function, some of the applications of these results which are most important to me are in contexts where we expect to be exactly clear what the domains of our functions are. Consequently it is necessary to form an opinion on such matters as what the σ-algebra defined by a random variable really is (272C).

The point of 272Q is that the family $\mathcal{E}$ does not have to be related in any way to the family $\langle\Sigma_i\rangle_{i\in I}$, except, of course, that we are dealing with measurable sets. All we need to know is that I should be large compared with $\mathcal{E}$; for instance, that $\mathcal{E}$ is countable and I is uncountable. The family $\langle\Sigma_j\rangle_{j\in J}$ is now a kind of 'tail' of $\langle\Sigma_i\rangle_{i\in I}$, safely independent of the 'head' σ-algebra generated by $\mathcal{E}$.

Of course I should emphasize again that such proofs as those in 272R-272S are to be thought of as confirmations that we have a suitable model of probability theory, rather than as reasons for believing the results to be valid in statistical contexts. Similarly, 272T-272U can be approached by a variety of intuitions concerning discrete random variables and random variables with continuous densities, and while the elegant general results are delightful, they are more important to the pure mathematician than to the statistician. But I came to an odd obstacle in the proof of 272S, when showing that if $X_1+\ldots+X_n$ has finite variance then so does every X_j. We have done enough measure theory for this to be readily dealt with, but the connexion with ordinary probabilistic intuition, both here and in 272Xi, remains unclear to me.

There are four ideas in 272W worth storing for future use. The first is the estimate

$$\mathbb{E}(e^{bX_i}) \leq 1 - a_i + e^b a_i$$

in part (a), a crude but effective way of using the hypothesis that X_i is bounded. The second is the use of Taylor's theorem to show that $1 - a_i + e^b a_i \leq \exp(a_i + \frac{1}{8}b^2)$. The third is the estimate

$$\Pr(Y \geq 0) \leq \mathbb{E}(e^{bY}) \text{ if } b \geq 0$$

used in part (b); and the fourth is 272R. After this one need only be sufficiently determined to reach 272Xk. But even the special case 272W is both striking and useful.

273 The strong law of large numbers

I come now to the first of the three main theorems of this chapter. Perhaps I should call it a 'principle', rather than a 'theorem', as I shall not attempt to enunciate any fully general form, but will give three theorems (273D, 273H, 273I), with a variety of corollaries, each setting out conditions under which the averages of a sequence of independent random variables will almost surely converge. At the end of the section (273N) I add a result on norm-convergence of averages.

273A It will be helpful to start with an explicit statement of a very simple but very useful lemma.

Lemma Let $\langle E_n\rangle_{n\in\mathbb{N}}$ be a sequence of measurable sets in a measure space (Ω,Σ,μ), and suppose that $\sum_{n=0}^\infty \mu E_n < \infty$. Then $\{n : \omega \in E_n\}$ is finite for almost every $\omega \in \Omega$.

proof We have

$$\mu\{\omega : \{n : \omega \in E_n\} \text{ is infinite}\} = \mu(\bigcap_{n\in\mathbb{N}}\bigcup_{m\geq n} E_m) = \inf_{n\in\mathbb{N}} \mu(\bigcup_{m\geq n} E_m)$$
$$\leq \inf_{n\in\mathbb{N}} \sum_{m=n}^\infty \mu E_m = 0.$$

273B Lemma Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables, and set $S_n = \sum_{i=0}^n X_i$ for each $n \in \mathbb{N}$.

(a) If $\langle S_n\rangle_{n\in\mathbb{N}}$ is convergent in measure, then it is convergent almost everywhere.

(b) In particular, if $\mathbb{E}(X_n) = 0$ for every n and $\sum_{n=0}^\infty \mathbb{E}(X_n^2) < \infty$, then $\sum_{n=0}^\infty X_n$ is defined, and finite, almost everywhere.

proof (a) Let (Ω,Σ,μ) be the underlying probability space. If we change each X_n on a negligible set, we do not change the independence of $\langle X_n\rangle_{n\in\mathbb{N}}$ (272H), and the S_n are also changed only on a negligible set; so we may suppose from the beginning that every X_n is a measurable function defined on the whole of Ω.

Because the functional $X \mapsto \mathbb{E}(\min(1,|X|))$ is one of the pseudometrics defining the topology of convergence in measure (245A), $\lim_{m,n\to\infty}\mathbb{E}(\min(1,|S_m-S_n|)) = 0$, and we can find for each $k \in \mathbb{N}$ an $n_k \in \mathbb{N}$ such that $\mathbb{E}(\min(1,|S_m-S_{n_k}|)) \leq 4^{-k}$ for every $m \geq n_k$. So $\Pr(|S_m - S_{n_k}| \geq 2^{-k}) \leq 2^{-k}$ for every $m \geq n_k$. By Etemadi's lemma (272V) applied to $\langle X_i\rangle_{i\geq n_k}$,

$$\Pr(\sup_{n_k\leq m\leq n}|S_m - S_{n_k}| \geq 3\cdot 2^{-k}) \leq 3\cdot 2^{-k}$$

for every $n \geq n_k$. Setting

$$H_{kn} = \{\omega : \sup_{n_k\leq m\leq n}|S_m(\omega) - S_{n_k}(\omega)| \geq 3\cdot 2^{-k}\} \text{ for } n \geq n_k,$$

$$H_k = \bigcup_{n\geq n_k} H_{kn},$$

we have

$$\mu H_k = \lim_{n\to\infty} \mu H_{kn} \le 3 \cdot 2^{-k}$$

for each k, so $\sum_{k=0}^{\infty} \mu H_k$ is finite and almost every $\omega \in \Omega$ belongs to only finitely many of the H_k (273A).

Now take any such ω. Then there is some $r \in \mathbb{N}$ such that $\omega \notin H_k$ for any $k \ge r$. In this case, for every $k \ge r$, $\omega \notin \bigcup_{n \ge n_k} H_{kn}$, that is, $|S_n(\omega) - S_{n_k}(\omega)| < 3 \cdot 2^{-k}$ for every $n \ge n_k$. But this means that $\langle S_n(\omega)\rangle_{n\in\mathbb{N}}$ is a Cauchy sequence, therefore convergent. Since this is true for almost every ω, $\langle S_n\rangle_{n\in\mathbb{N}}$ converges almost everywhere, as claimed.

(b) Now suppose that $\mathbb{E}(X_n) = 0$ for every n and that $\sum_{n=0}^{\infty} \mathbb{E}(X_n^2) < \infty$. In this case, for any $m < n$,

$$\|S_n - S_m\|_1^2 \le \|\chi\Omega\|_2^2 \|S_n - S_m\|_2^2$$

(by Cauchy's inequality, 244Eb)

$$= \mathbb{E}(S_n - S_m)^2 = \mathrm{Var}(S_n - S_m)$$

(because $\mathbb{E}(S_n - S_m) = \sum_{i=m+1}^{n} \mathbb{E}(X_i) = 0$)

$$= \sum_{i=m+1}^{n} \mathrm{Var}(X_i)$$

(by Bienaymé's equality, 272S)

$$\to 0$$

as $m \to \infty$. So $\langle S_n^\bullet\rangle_{n\in\mathbb{N}}$ is a Cauchy sequence in $L^1(\mu)$ and converges in $L^1(\mu)$, by 242F; by 245G, it converges in measure in $L^0(\mu)$, that is, $\langle S_n\rangle_{n\in\mathbb{N}}$ converges in measure in $\mathcal{L}^0(\mu)$. By (a), $\langle S_n\rangle_{n\in\mathbb{N}}$ converges almost everywhere, that is, $\sum_{i=0}^{\infty} X_i$ is defined and finite almost everywhere.

Remark The proof above assumes familiarity with the ideas of Chapter 24. However part (b), at least, can be established without any of these; see 273Xa. In 276B there is a generalization of (b) based on a different approach.

273C We now need a lemma (part (b) below) from the theory of summability. I take the opportunity to include an elementary fact which will be useful later in this section and elsewhere.

Lemma (a) If $\lim_{n\to\infty} x_n = x$, then $\lim_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} x_i = x$.

(b) Let $\langle x_n\rangle_{n\in\mathbb{N}}$ be summable, and $\langle b_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence in $[0,\infty[$ diverging to ∞. Then $\lim_{n\to\infty} \frac{1}{b_n} \sum_{k=0}^{n} b_k x_k$
0.

proof (a) Let $\epsilon > 0$. Let m be such that $|x_n - x| \le \epsilon$ whenever $n \ge m$. Let $m' \ge m$ be such that $|\sum_{i=0}^{m-1} x - x_i| \le \epsilon m'$. Then for $n \ge m'$ we have

$$\begin{aligned} |x - \frac{1}{n+1}\sum_{i=0}^{n} x_i| &= \frac{1}{n+1}|\sum_{i=0}^{n} x - x_i| \\ &\le \frac{1}{n+1}|\sum_{i=0}^{m-1} x - x_i| + \frac{1}{n+1}\sum_{i=m}^{n}|x - x_i| \\ &\le \frac{\epsilon m'}{n+1} + \frac{\epsilon(n-m+1)}{n+1} \le 2\epsilon. \end{aligned}$$

As ϵ is arbitrary, $\lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} x_i = x$.

(b) Let $\epsilon > 0$. Write $s_n = \sum_{i=0}^{n} x_i$ for each n, and

$$s = \lim_{n\to\infty} s_n = \sum_{i=0}^{\infty} x_i;$$

set $s^* = \sup_{n\in\mathbb{N}} |s_n| < \infty$. Let $m \in \mathbb{N}$ be such that $|s_n - s| \le \epsilon$ whenever $n \ge m$; then $|s_n - s_j| \le 2\epsilon$ whenever j, $n \ge m$. Let $m' \ge m$ be such that $b_m s^* \le \epsilon b_{m'}$.

Take any $n \ge m'$. Then

$$
\begin{aligned}
|\sum_{k=0}^{n} b_k x_k| &= |b_0 s_0 + b_1(s_1 - s_0) + \ldots + b_n(s_n - s_{n-1})| \\
&= |(b_0 - b_1)s_0 + (b_1 - b_2)s_1 + \ldots + (b_{n-1} - b_n)s_{n-1} + b_n s_n| \\
&= |b_0 s_n + \sum_{i=0}^{n-1}(b_{i+1} - b_i)(s_n - s_i)| \\
&\le b_0|s_n| + \sum_{i=0}^{m-1}(b_{i+1} - b_i)|s_n - s_i| + \sum_{i=m}^{n-1}(b_{i+1} - b_i)|s_n - s_i| \\
&\le b_0 s^* + 2s^* \sum_{i=0}^{m-1}(b_{i+1} - b_i) + 2\epsilon \sum_{i=m}^{n-1}(b_{i+1} - b_i) \\
&= b_0 s^* + 2s^*(b_m - b_0) + 2\epsilon(b_n - b_m) \le 2s^* b_m + 2\epsilon b_n.
\end{aligned}
$$

Consequently, because $b_n \ge b_{m'}$,

$$|\frac{1}{b_n}\sum_{k=0}^{n} b_k x_k| \le 2\frac{s^* b_m}{b_n} + 2\epsilon \le 4\epsilon.$$

As ϵ is arbitrary,

$$\lim_{n\to\infty} \frac{1}{b_n}\sum_{k=0}^{n} b_k x_k = 0,$$

as required.

Remark Part (b) above is sometimes called 'Kronecker's lemma'.

273D The strong law of large numbers: first form Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables, and suppose that $\langle b_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence in $]0,\infty[$, diverging to ∞, such that $\sum_{n=0}^{\infty} \frac{1}{b_n^2} \operatorname{Var}(X_n) < \infty$. Then

$$\lim_{n\to\infty} \frac{1}{b_n}\sum_{i=0}^{n}(X_i - \mathbb{E}(X_i)) = 0$$

almost everywhere.

proof As usual, write (Ω, Σ, μ) for the underlying probability space. Set

$$Y_n = \frac{1}{b_n}(X_n - \mathbb{E}(X_n))$$

for each n; then $\langle Y_n\rangle_{n\in\mathbb{N}}$ is independent (272E), $\mathbb{E}(Y_n) = 0$ for each n, and

$$\sum_{n=0}^{\infty} \mathbb{E}(Y_n^2) = \sum_{n=0}^{\infty} \frac{1}{b_n^2} \operatorname{Var}(X_n) < \infty.$$

By 273B, $\langle Y_n(\omega)\rangle_{n\in\mathbb{N}}$ is summable for almost every $\omega \in \Omega$. But by 273Cb,

$$\lim_{n\to\infty} \frac{1}{b_n}\sum_{i=0}^{n}(X_i(\omega) - \mathbb{E}(X_i)) = \lim_{n\to\infty} \frac{1}{b_n}\sum_{i=0}^{n} b_i Y_i(\omega) = 0$$

for all such ω. So we have the result.

273E Corollary Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of random variables such that $\mathbb{E}(X_n) = 0$ for every n and $\sup_{n\in\mathbb{N}} \mathbb{E}(X_n^2) < \infty$. Then

$$\lim_{n\to\infty} \frac{1}{b_n}(X_0 + \ldots + X_n) = 0$$

almost everywhere whenever $\langle b_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence of strictly positive numbers and $\sum_{n=0}^{\infty} \frac{1}{b_n^2}$ is finite. In particular,

$$\lim_{n\to\infty} \frac{1}{n+1}(X_0 + \ldots + X_n) = 0$$

almost everywhere.

Remark For most of the rest of this section, we shall take $b_n = n+1$. The special virtue of 273D is that it allows other b_n, e.g., $b_n = \sqrt{n}\ln n$. A direct strengthening of this theorem is in 276C below.

273F Corollary Let $\langle E_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of measurable sets in a probability space (Ω,Σ,μ). and suppose that

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}\mu E_i = c.$$

Then

$$\lim_{n\to\infty}\frac{1}{n+1}\#(\{i : i\le n,\,\omega\in E_i\}) = c$$

for almost every $\omega\in\Omega$.

proof In 273D, set $X_n = \chi E_n$, $b_n = n+1$. For almost every ω, we have

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}(\chi E_i(\omega) - a_i) = 0,$$

writing $a_i = \mu E_i = \mathbb{E}(X_i)$ for each i. (I see that I am relying on 272F to support the claim that $\langle X_n\rangle_{n\in\mathbb{N}}$ is independent.) But for any such ω,

$$\begin{aligned}\lim_{n\to\infty}\Bigl(\frac{1}{n+1}\#(\{i : i\le n,\,\omega\in E_i\}) - \frac{1}{n+1}\sum_{i=0}^{n}a_i\Bigr)\\ = \lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}(\chi E_i(\omega) - a_i) = 0;\end{aligned}$$

because we are supposing that $\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}a_i = c$, we must have

$$\lim_{n\to\infty}\frac{1}{n+1}\#(\{i : i\le n,\,\omega\in E_i\}) = c,$$

as required.

273G Corollary Let μ be the usual measure on $\mathcal{P}\mathbb{N}$, as described in 254Jb. Then for μ-almost every set $a\subseteq\mathbb{N}$,

$$\lim_{n\to\infty}\frac{1}{n+1}\#(a\cap\{0,\dots,n\}) = \tfrac{1}{2}.$$

proof The sets $E_n = \{a : n\in a\}$ are independent, with measure $\frac{1}{2}$.

Remark The limit $\lim_{n\to\infty}\frac{1}{n+1}\#(a\cap\{0,\dots,n\})$ is called the **asymptotic density** of a.

273H Strong law of large numbers: second form Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables, and suppose that $\sup_{n\in\mathbb{N}}\mathbb{E}(|X_n|^{1+\delta}) < \infty$ for some $\delta > 0$. Then

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}(X_i - \mathbb{E}(X_i)) = 0$$

almost everywhere.

proof As usual, call the underlying probability space (Ω,Σ,μ); as in 273B we can adjust the X_n on negligible sets so as to make them measurable and defined everywhere on Ω, without changing $\mathbb{E}(X_n)$, $\mathbb{E}(|X_n|)$ or the convergence of the averages except on a negligible set.

(a) For each n, define a random variable Y_n on Ω by setting

$$\begin{aligned}Y_n(\omega) &= X_n(\omega) \text{ if } |X_n(\omega)|\le n,\\ &= 0 \text{ if } |X_n(\omega)| > n.\end{aligned}$$

Then $\langle Y_n\rangle_{n\in\mathbb{N}}$ is independent (272E). For each $n\in\mathbb{N}$,

$$\mathrm{Var}(Y_n) \le \mathbb{E}(Y_n^2) \le \mathbb{E}(n^{1-\delta}|X_n|^{1+\delta}) \le n^{1-\delta}K,$$

where $K = \sup_{n\in\mathbb{N}} \mathbb{E}(|X_n|^{1+\delta})$, so

$$\sum_{n=0}^{\infty} \frac{1}{(n+1)^2} \mathrm{Var}(Y_n) \le \sum_{n=0}^{\infty} \frac{n^{1-\delta}}{(n+1)^2} K < \infty.$$

By 273D,

$$G = \{\omega : \lim_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} (Y_i(\omega) - \mathbb{E}(Y_i)) = 0\}$$

is conegligible.

(b) On the other hand, setting

$$E_n = \{\omega : Y_n(\omega) \ne X_n(\omega)\} = \{\omega : |X_n(\omega)| > n\},$$

we have $K \ge n^{1+\delta}\mu E_n$ for each n, so

$$\sum_{n=0}^{\infty} \mu E_n \le 1 + K \sum_{n=1}^{\infty} \frac{1}{n^{1+\delta}} < \infty,$$

and the set $H = \{\omega : \{n : \omega \in E_n\}$ is finite$\}$ is conegligible (273A). But of course

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} (X_i(\omega) - Y_i(\omega)) = 0$$

for every $\omega \in H$.

(c) Finally,

$$|\mathbb{E}(Y_n) - \mathbb{E}(X_n)| \le \int_{E_n} |X_n| \le \int_{E_n} n^{-\delta}|X_n|^{1+\delta} \le n^{-\delta}K$$

whenever $n \ge 1$, so $\lim_{n\to\infty} \mathbb{E}(Y_n) - \mathbb{E}(X_n) = 0$ and

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} \mathbb{E}(Y_i) - \mathbb{E}(X_i) = 0$$

(273Ca). Putting these three together, we get

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} X_i(\omega) - \mathbb{E}(X_i) = 0$$

whenever ω belongs to the conegligible set $G \cap H$. So

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} X_i - \mathbb{E}(X_i) = 0$$

almost everywhere, as required.

273I Strong law of large numbers: third form Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables of finite expectation, and suppose that they are **identically distributed**, that is, all have the same distribution. Then

$$\lim_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} (X_i - \mathbb{E}(X_i)) = 0$$

almost everywhere.

proof The proof follows the same line as that of 273H, but some of the inequalities require more delicate arguments. As usual, call the underlying probability space (Ω, Σ, μ) and suppose that the X_n are all measurable and defined everywhere on Ω. (We need to remember that changing a random variable on a negligible set does not change its distribution.) Let ν be the common distribution of the X_n.

(a) For each n, define a random variable Y_n on Ω by setting

$$\begin{aligned} Y_n(\omega) &= X_n(\omega) \text{ if } |X_n(\omega)| \le n, \\ &= 0 \text{ if } |X_n(\omega)| > n. \end{aligned}$$

Then $\langle Y_n\rangle_{n\in\mathbb{N}}$ is independent (272E). For each $n \in \mathbb{N}$,

$$\mathrm{Var}(Y_n) \le \mathbb{E}(Y_n^2) = \int_{[-n,n]} x^2 \nu(dx)$$

(271E). To estimate $\sum_{n=0}^{\infty}\frac{1}{(n+1)^2}\mathbb{E}(Y_n^2)$, set

$$f_n(x)=\frac{x^2}{(n+1)^2}\text{ if }|x|\le n,\ 0\text{ if }|x|>n,$$

so that $\frac{1}{(n+1)^2}\operatorname{Var}(Y_n)\le\int f_n d\nu$. If $r\ge 1$ and $r<|x|\le r+1$ then

$$\begin{aligned}\sum_{n=0}^{\infty}f_n(x)&\le\sum_{n=r+1}^{\infty}\frac{1}{(n+1)^2}(r+1)|x|\\&\le(r+1)|x|\sum_{n=r+1}^{\infty}(\frac{1}{n}-\frac{1}{n+1})\le|x|,\end{aligned}$$

while if $|x|\le 1$ then

$$\sum_{n=0}^{\infty}f_n(x)\le\sum_{n=0}^{\infty}\frac{1}{(n+1)^2}=\frac{\pi^2}{6}\le 2<\infty.$$

(You do not need to know that the sum is $\frac{\pi^2}{6}$, only that it is finite; but see 282Xo.) Consequently

$$f(x)=\sum_{n=0}^{\infty}f_n(x)\le 2+|x|$$

for every x, and $\int fd\nu<\infty$, because $\int|x|\nu(dx)$ is the common value of $\mathbb{E}(|X_n|)$, and is finite. By any of the great convergence theorems,

$$\sum_{n=0}^{\infty}\frac{1}{(n+1)^2}\operatorname{Var}(Y_n)\le\sum_{n=0}^{\infty}\int f_n d\nu=\int fd\nu<\infty.$$

By 273D,

$$G=\{\omega:\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}(Y_i(\omega)-\mathbb{E}(Y_i))=0\}$$

is conegligible.

(b) Next, setting

$$E_n=\{\omega:X_n(\omega)\ne Y_n(\omega)\}=\{\omega:|X_n(\omega)|>n\},$$

we have

$$E_n=\bigcup_{i\ge n}F_{ni},$$

where

$$F_{ni}=\{\omega:i<|X_n(\omega)|\le i+1\}.$$

Now

$$\mu F_{ni}=\nu\{x:i<|x|\le i+1\}$$

for every n and i. So

$$\begin{aligned}\sum_{n=0}^{\infty}\mu E_n&=\sum_{n=0}^{\infty}\sum_{i=n}^{\infty}\mu F_{ni}=\sum_{i=0}^{\infty}\sum_{n=0}^{i}\mu F_{ni}\\&=\sum_{i=0}^{\infty}(i+1)\nu\{x:i<|x|\le i+1\}\le\int(1+|x|)\nu(dx)<\infty.\end{aligned}$$

Consequently the set $H=\{\omega:\{n:X_n(\omega)\ne Y_n(\omega)\}\text{ is finite}\}$ is conegligible (273A). But of course

$$\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}X_i(\omega)-Y_i(\omega)=0$$

for every $\omega\in H$.

(c) Finally,

$$|\mathbb{E}(Y_n)-\mathbb{E}(X_n)|\le\int_{E_n}|X_n|=\int_{\mathbb{R}\setminus[-n,n]}|x|\nu(dx)$$

whenever $n\in\mathbb{N}$, so $\lim_{n\to\infty}\mathbb{E}(Y_n)-\mathbb{E}(X_n)=0$ and

$$\lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} \mathbb{E}(Y_i) - \mathbb{E}(X_i) = 0$$

(273Ca). Putting these three together, we get

$$\lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} X_i(\omega) - \mathbb{E}(X_i) = 0$$

whenever ω belongs to the conegligible set $G \cap H$. So

$$\lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} X_i - \mathbb{E}(X_i) = 0$$

almost everywhere, as required.

Remarks In my own experience, this is the most important form of the strong law from the point of view of 'pure' measure theory. I note that 273G above can also be regarded as a consequence of this form.

For a very striking alternative proof, see 275Yn. Yet another proof treats this result as a special case of the Ergodic Theorem (see 372Xg in Volume 3).

273J Corollary Let (Ω, Σ, μ) be a probability space. If f is a real-valued function such that $\int f d\mu$ is defined in $[-\infty, \infty]$, then

$$\lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} f(\omega_i) = \int f d\mu$$

for λ-almost every $\boldsymbol{\omega} = \langle \omega_n \rangle_{n\in\mathbb{N}} \in \Omega^{\mathbb{N}}$, where λ is the product measure on $\Omega^{\mathbb{N}}$ (254A-254C).

proof (a) To begin with, suppose that f is integrable. Define functions X_n on $\Omega^{\mathbb{N}}$ by setting

$$X_n(\boldsymbol{\omega}) = f(\omega_n) \text{ whenever } \omega_n \in \operatorname{dom} f.$$

Then $\langle X_n \rangle_{n\in\mathbb{N}}$ is an independent sequence of random variables, all with the same distribution as f (272M). So

$$\lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} f(\omega_i) - \int f d\mu = \lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} X_i(\boldsymbol{\omega}) - \mathbb{E}(X_i) = 0$$

for almost every $\boldsymbol{\omega}$, by 273I, and

$$\lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} f(\omega_i) = \int f d\mu$$

for almost every $\boldsymbol{\omega}$.

(b) Next, suppose that $f \ge 0$ and $\int f = \infty$. In this case, for every $m \in \mathbb{N}$,

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} f(\omega_i) &\ge \liminf_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} \min(m, f(\omega_i)) \\ &= \int \min(m, f(\omega))\mu(d\omega)\end{aligned}$$

for almost every $\boldsymbol{\omega}$, so

$$\liminf_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} f(\omega_i) \ge \sup_{m\in\mathbb{N}} \int \min(m, f(\omega))\mu(d\omega) = \infty$$

and

$$\lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} f(\omega_i) = \infty = \int f$$

for almost every $\boldsymbol{\omega}$.

(c) In general, if $\int f = \infty$, this is because $\int f^+ = \infty$ and f^- is integrable, so

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} f(\omega_i) &= \lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} f^+(\omega_i) - \lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} f^-(\omega_i) \\ &= \infty - \int f^- = \int f\end{aligned}$$

for almost every $\boldsymbol{\omega}$. Similarly,

$$\lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} f(\omega_i) = -\infty$$

for almost every $\boldsymbol{\omega}$ if $\int f d\mu = -\infty$.

Remark I find myself slipping here into measure-theorists' terminology; this corollary is one of the basic applications of the strong law to measure theory. Obviously, in view of 272J and 272M, this corollary covers 273I. It could also (in theory) be used as a *definition* of integration on a probability space (see 273Ya); it is sometimes called the 'Monte Carlo' method of integration.

273K It is tempting to seek extensions of 273I in which the X_n are not identically distributed, but are otherwise well-behaved. Any such idea should be tested against the following example. I find that I need another standard result, complementing that in 273A.

Borel-Cantelli lemma Let (Ω, Σ, μ) be a probability space and $\langle E_n\rangle_{n\in\mathbb{N}}$ a sequence of measurable subsets of Ω such that $\sum_{n=0}^{\infty} \mu E_n = \infty$ and $\mu(E_m \cap E_n) \le \mu E_m \cdot \mu E_n$ whenever $m \ne n$. Then almost every point of Ω belongs to infinitely many of the E_n.

proof For $n, k \in \mathbb{N}$ set $X_n = \sum_{i=0}^{n} \chi E_i$, $\beta_n = \sum_{i=0}^{n} \mu E_n = \mathbb{E}(X_n)$ and $F_{nk} = \{x : x \in \Omega, \#(\{i : i \le n, x \in E_i\}) \le k\}$. Then

$$\begin{aligned}\mathbb{E}(X_n^2) &= \sum_{i=0}^{n}\sum_{j=0}^{n} \mu(E_i \cap E_j) \le \sum_{i=0}^{n} \mu E_i + \sum_{i=0}^{n}\sum_{j\ne i} \mu E_i \cdot \mu E_j \\ &= \beta_n + \beta_n^2 - \sum_{i=0}^{n} (\mu E_i)^2,\end{aligned}$$

so

$$\operatorname{Var}(X_n) = \beta_n - \sum_{i=0}^{n} (\mu E_i)^2 \le \beta_n.$$

Now if $k < \beta_n$,

$$(\beta_n - k)^2 \mu F_{nk} = (\beta_n - k)^2 \Pr(X_n \le k) \le \mathbb{E}(X_n - \beta_n)^2 = \operatorname{Var}(X_n) \le \beta_n$$

and $\mu F_{nk} \le \frac{\beta_n}{(\beta_n - k)^2}$.

Now recall that we are assuming that $\lim_{n\to\infty} \beta_n = \infty$. So for any $k \in \mathbb{N}$,

$$\mu(\bigcap_{n\in\mathbb{N}} F_{nk}) = \lim_{n\to\infty} \mu F_{nk} \le \lim_{n\to\infty} \frac{\beta_n}{(\beta_n - k)^2} = 0.$$

Accordingly

$$\mu\{x : x \text{ belongs to only finitely many } E_n\} = \mu(\bigcup_{k\in\mathbb{N}} \bigcap_{n\in\mathbb{N}} F_{nk}) = 0,$$

and almost every point of Ω belongs to infinitely many E_n.

Remark Of course this result is usually applied to an independent sequence $\langle E_n\rangle_{n\in\mathbb{N}}$. But very occasionally it is of interest to know that it is enough to assume that weaker hypotheses suffice. See also 273Yb.

273L Now for the promised example.

Example There is an independent sequence $\langle X_n\rangle_{n\in\mathbb{N}}$ of non-negative random variables such that $\lim_{n\to\infty} \mathbb{E}(X_n) = 0$ but

$$\limsup_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{\infty} X_i - \mathbb{E}(X_i) = \infty,$$

$$\liminf_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{\infty} X_i - \mathbb{E}(X_i) = 0$$

almost everywhere.

proof Let (Ω, Σ, μ) be a probability space with an independent sequence $\langle E_n\rangle_{n\in\mathbb{N}}$ of measurable sets such that $\mu E_n = \frac{1}{(n+3)\ln(n+3)}$ for each n. (I have nowhere explained exactly how to build such a sequence. Two obvious methods are available to us, and another a trifle less obvious. (i) Take $\Omega = \{0,1\}^{\mathbb{N}}$ and μ to be the product of the probabilities μ_n on $\{0,1\}$, defined by saying that $\mu_n\{1\} = \frac{1}{(n+3)\ln(n+3)}$ for each n; set $E_n = \{\omega : \omega(n) = 1\}$, and appeal to 272M to check that the E_n are independent. (ii) Build the E_n inductively as subsets of $[0,1]$, arranging that each E_n should be a finite union of intervals, so that when you come to choose E_{n+1} the sets $E_0, \dots, E_n$ define a partition $\mathcal{I}_n$ of $[0,1]$ into intervals, and you can take E_{n+1} to be the union of (say) the left-hand subintervals of length a proportion $\frac{1}{(n+3)\ln(n+3)}$ of the intervals in $\mathcal{I}_n$. (iii) Use 215D to see that the method of (ii) can be used on any atomless probability space, as in 272Xa.)

Set $X_n = (n+3)\ln\ln(n+3)\chi E_n$ for each n; then $\langle X_n\rangle_{n\in\mathbb{N}}$ is an independent sequence of real-valued random variables (272F) and $\mathbb{E}(X_n) = \frac{\ln\ln(n+3)}{\ln(n+3)}$ for each n, so that $\mathbb{E}(X_n) \to 0$ as $n \to \infty$. Thus, for instance, $\{X_n : n \in \mathbb{N}\}$ is uniformly integrable and $\langle X_n\rangle_{n\in\mathbb{N}} \to 0$ in measure (246Jc); while surely $\lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^n \mathbb{E}(X_i) = 0$.

On the other hand,

$$\sum_{n=0}^{\infty} \mu E_n = \sum_{n=0}^{\infty} \frac{1}{(n+3)\ln(n+3)} \ge \int_0^{\infty} \frac{1}{(x+3)\ln(x+3)} dx$$
$$= \lim_{a\to\infty} (\ln\ln(a+3) - \ln\ln 3) = \infty,$$

so almost every ω belongs to infinitely many of the E_n, by the Borel-Cantelli lemma (273K). Now if we write $Y_n = \frac{1}{n+1}\sum_{i=0}^n X_i$, then if $\omega \in E_n$ we have $X_n(\omega) = (n+3)\ln\ln(n+3)$ so

$$Y_n(\omega) \ge \frac{n+3}{n+1}\ln\ln(n+3).$$

This means that

$$\{\omega : \limsup_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n}(X_i(\omega) - \mathbb{E}(X_i)) = \infty\} = \{\omega : \limsup_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n} X_i(\omega) = \infty\}$$
$$= \{\omega : \sup_{n\in\mathbb{N}} Y_n(\omega) = \infty\} \supseteq \{\omega : \{n : \omega \in E_n\} \text{ is infinite}\}$$

is conegligible, and the strong law of large numbers does not apply to $\langle X_n\rangle_{n\in\mathbb{N}}$.

Because

$$\lim_{n\to\infty} \|Y_n\|_1 = \lim_{n\to\infty} \mathbb{E}(Y_n) = \lim_{n\to\infty} \mathbb{E}(X_n) = 0$$

(273Ca), $\langle Y_n\rangle_{n\in\mathbb{N}} \to 0$ for the topology of convergence in measure, and $\langle Y_n\rangle_{n\in\mathbb{N}}$ has a subsequence converging to 0 almost everywhere (245K). So

$$\liminf_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^{n}(X_i(\omega) - \mathbb{E}(X_i)) = \liminf_{n\to\infty} Y_n(\omega) = 0$$

for almost every ω. The fact that both $\limsup_{n\to\infty} Y_n$ and $\liminf_{n\to\infty} Y_n$ are constant almost everywhere is of course a consequence of the zero-one law (272P).

***273M** All the above has been concerned with pointwise convergence of the averages of independent random variables, and that is the important part of the work of this section. But it is perhaps worth complementing it with a brief investigation of norm-convergence. To deal efficiently with convergence in $\mathcal{L}^p$, we need the following. (I should perhaps remark that, compared with the general case treated here, the case $p = 2$ is trivial; see 273Xl.)

Lemma For any $p \in \,]1,\infty[$ and $\epsilon > 0$, there is a $\delta > 0$ such that $\|S + X\|_p \le 1 + \epsilon\|X\|_p$ whenever S and X are independent random variables, $\|S\|_p = 1$, $\|X\|_p \le \delta$ and $\mathbb{E}(X) = 0$.

proof (a) Take $\zeta \in \,]0,1]$ such that $p\zeta \le 2$ and

$$(1+\xi)^p \le 1 + p\xi + \frac{p^2}{2}\xi^2$$

whenever $|\xi| \le \zeta$; such exists because

$$\lim_{\xi\to 0} \frac{(1+\xi)^p - 1 - p\xi}{\xi^2} = \frac{p(p-1)}{2} < \frac{p^2}{2}.$$

Observe that

$$(1+\xi)^p \le (1+\frac{1}{\zeta})^p + \xi^p + 2p\xi^{p-1}$$

for every $\xi \ge 0$. **P** If $\xi \le \frac{1}{\zeta}$, this is trivial. If $\xi \ge \frac{1}{\zeta}$, then

$$(1+\xi)^p = \xi^p(1+\frac{1}{\xi})^p \le \xi^p(1+\frac{p}{\xi}+\frac{p^2}{2\xi^2})$$
$$\le \xi^p(1+\frac{p}{\xi}+\frac{p^2\zeta}{2\xi}) = \xi^p + p\xi^{p-1}(1+\frac{p\zeta}{2}) \le \xi^p + 2p\xi^{p-1}. \text{ **Q**}$$

Define $\eta > 0$ by declaring that $3\eta^{p-1} = \frac{\epsilon}{2}$ (this is one of the places where we need to know that $p > 1$). Let $\delta > 0$ be such that

$$\delta \le \eta\zeta, \quad \frac{p^2}{2\eta^2}\delta + (1+\frac{1}{\zeta})^p\delta^{p-1} \le \frac{p\epsilon}{2}.$$

(b) Now suppose that S and X are independent random variables with $\|S\|_p = 1$, $\|X\|_p \le \delta$ and $\mathbb{E}(X) = 0$. If $\|X\|_p = 0$ then of course $\|S+X\|_p \le 1+\epsilon\|X\|_p$, so suppose that X is non-trivial. Write (Ω,Σ,μ) for the underlying probability space and adjust S and X on negligible sets so that they are measurable and defined everywhere on Ω. Set $\alpha = \|X\|_p$, $\gamma = \alpha/\eta$,

$$E = \{\omega : S(\omega) \ne 0\}, \quad F = \{\omega : |X(\omega)| > \gamma|S(\omega)|\}, \quad \beta = \|S\times\chi F\|_p.$$

Then

$$\int |S+X|^p = \int_F |S+X|^p + \int_{E\setminus F} |S+X|^p$$

(because S and X are both zero on $\Omega\setminus(E\cup F)$)

$$= \|(S\times\chi F)+(X\times\chi F)\|_p^p + \int_{E\setminus F}|S|^p|1+\frac{X}{S}|^p$$
$$\le (\|S\times\chi F\|_p + \|X\times\chi F)\|_p)^p + \int_{E\setminus F}|S|^p(1+p\frac{X}{S}+\frac{p^2}{2}\gamma^2)$$

(because $|\frac{X}{S}| \le \gamma \le \frac{\delta}{\eta} \le \zeta \le 1$ everywhere on $E\setminus F$)

$$\le (\beta+\alpha)^p + (1+\frac{p^2}{2}\gamma^2)\int_{E\setminus F}|S|^p + p\int_{E\setminus F}|S|^{p-1}\times\operatorname{sgn} S\times X$$

(writing $\operatorname{sgn}(\xi) = \xi/|\xi|$ if $\xi \ne 0$, 0 if $\xi = 0$)

$$= (\beta+\alpha)^p + (1+\frac{p^2}{2}\gamma^2)\int_{\Omega\setminus F}|S|^p + p\int_{\Omega\setminus F}|S|^{p-1}\times\operatorname{sgn} S\times X$$

(because $S = 0$ on $\Omega\setminus E$)

$$= \alpha^p(1+\frac{\beta}{\alpha})^p + (1+\frac{p^2}{2}\gamma^2)(1-\beta^p) - p\int_F|S|^{p-1}\times\operatorname{sgn} S\times X$$

(because X and $|S|^{p-1}\times\operatorname{sgn} S$ are independent, by 272L, so $\int|S|^{p-1}\times\operatorname{sgn} S\times X = \mathbb{E}(|S|^{p-1}\times\operatorname{sgn} S)\mathbb{E}(X) = 0$)

$$\le \alpha^p\bigl((1+\frac{1}{\zeta})^p + 2p(\frac{\beta}{\alpha})^{p-1} + (\frac{\beta}{\alpha})^p\bigr) + (1+\frac{p^2}{2}\gamma^2)(1-\beta^p)$$
$$+ p\int_F|S|^{p-1}\times|X|$$

(see (a) above)

$$\le \alpha^p(1+\frac{1}{\zeta})^p + \beta^p + 2p\beta^{p-1}\alpha + (1+\frac{p^2}{2}\gamma^2)(1-\beta^p)$$
$$+ p\int_F\frac{1}{\gamma^{p-1}}|X|^p$$
$$\le \alpha^p(1+\frac{1}{\zeta})^p + 2p\frac{\alpha^p}{\gamma^{p-1}} + 1 + \frac{p^2}{2}\gamma^2 + p\frac{\alpha^p}{\gamma^{p-1}}$$

(because $\beta = \|S\times\chi F\|_p \le \frac{1}{\gamma}\|X\times\chi F\|_p \le \frac{\alpha}{\gamma}$)

$$= \alpha^p(1+\frac{1}{\zeta})^p + 3p\eta^{p-1}\alpha + 1 + \frac{p^2\alpha^2}{2\eta^2}$$
$$= 1 + \bigl(\alpha^{p-1}(1+\frac{1}{\zeta})^p + 3p\eta^{p-1} + \frac{p^2\alpha}{2\eta^2}\bigr)\alpha$$
$$\le 1 + \bigl(\delta^{p-1}(1+\frac{1}{\zeta})^p + 3p\eta^{p-1} + \frac{p^2\delta}{2\eta^2}\bigr)\alpha$$
$$\le 1 + p\alpha\epsilon \le (1+\epsilon\|X\|_p)^p.$$

So $\|S+X\|_p \le 1+\epsilon\|X\|_p$, as required.

***Remark** What is really happening here is that $\phi = \| \|_p^p : L^p \to \mathbb{R}$ is differentiable (as a real-valued function on the normed space L^p) and

$$\phi'(S^\bullet)(X^\bullet) = p\int |S|^{p-1} \times \operatorname{sgn} S \times X,$$

so that in the context here

$$\phi((S+X)^\bullet) = \phi(S^\bullet) + \phi'(S^\bullet)(X^\bullet) + o(\|X\|_p) = 1 + o(\|X\|_p)$$

and $\|S+X\|_p = 1+o(\|X\|_p)$. The calculations above are elaborate partly because they do not appeal to any non-trivial ideas about normed spaces, and partly because we need the estimates to be uniform in S.

273N Theorem Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables with zero expectation, and set $Y_n = \frac{1}{n+1}(X_0+\ldots+X_n)$ for each $n\in\mathbb{N}$.

(a) If $\langle X_n\rangle_{n\in\mathbb{N}}$ is uniformly integrable, then $\lim_{n\to\infty}\|Y_n\|_1 = 0$.

*(b) If $p \in \,]1,\infty[$ and $\sup_{n\in\mathbb{N}}\|X_n\|_p < \infty$, then $\lim_{n\to\infty}\|Y_n\|_p = 0$.

proof (a) Let $\epsilon > 0$. Then there is an $M \ge 0$ such that $\mathbb{E}(|X_n|-M)^+ \le \epsilon$ for every $n\in\mathbb{N}$. Set

$$X_n' = (-M\chi\Omega)\vee(X_n\wedge M\chi\Omega), \quad \alpha_n = \mathbb{E}(X_n'), \quad \tilde X_n = X_n' - \alpha_n, \quad X_n'' = X_n - X_n'$$

for each $n\in\mathbb{N}$. Then $\langle X_n'\rangle_{n\in\mathbb{N}}$ and $\langle\tilde X_n\rangle_{n\in\mathbb{N}}$ are independent and uniformly bounded, and $\|X_n''\|_1 \le \epsilon$ for every n. So if we write

$$\tilde Y_n = \frac{1}{n+1}\sum_{i=0}^n \tilde X_i, \quad Y_n'' = \frac{1}{n+1}\sum_{i=0}^n X_i'',$$

$\langle\tilde Y_n\rangle_{n\in\mathbb{N}} \to 0$ almost everywhere, by 273E (for instance), while $\|Y_n''\|_1 \le \epsilon$ for every n. Moreover,

$$|\alpha_n| = |\mathbb{E}(X_n' - X_n)| \le \mathbb{E}(|X_n''|) \le \epsilon$$

for every n. As $|\tilde Y_n| \le 2M$ almost everywhere for each n, $\lim_{n\to\infty}\|\tilde Y_n\|_1 = 0$, by Lebesgue's Dominated Convergence Theorem. So

$$\begin{aligned}\limsup_{n\to\infty}\|Y_n\|_1 &= \limsup_{n\to\infty}\|\tilde Y_n + Y_n'' + \alpha_n\|_1\\ &\le \lim_{n\to\infty}\|\tilde Y_n\|_1 + \sup_{n\in\mathbb{N}}\|Y_n''\|_1 + \sup_{n\in\mathbb{N}}|\alpha_n|\\ &\le 2\epsilon.\end{aligned}$$

As ϵ is arbitrary, $\lim_{n\to\infty}\|Y_n\|_1 = 0$, as claimed.

***(b)** Set $M = \sup_{n\in\mathbb{N}}\|X_n\|_p$. For $n\in\mathbb{N}$, set $S_n = \sum_{i=0}^n X_i$. Let $\epsilon > 0$. Then there is a $\delta > 0$ such that $\|S+X\|_p \le 1+\epsilon\|X\|_p$ whenever S and X are independent random variables, $\|S\|_p = 1$, $\|X\|_p \le \delta$ and $\mathbb{E}(X) = 0$ (273M). It follows that $\|S+X\|_p \le \|S\|_p + \epsilon\|X\|_p$ whenever S and X are independent random variables, $\|S\|_p$ is finite, $\|X\|_p \le \delta\|S\|_p$ and $\mathbb{E}(X) = 0$. In particular, $\|S_{n+1}\|_p \le \|S_n\|_p + \epsilon M$ whenever $\|S_n\|_p \ge M/\delta$. An easy induction shows that

$$\|S_n\|_p \le \frac{M}{\delta} + M + n\epsilon M$$

for every $n\in\mathbb{N}$. But this means that

$$\limsup_{n\to\infty}\|Y_n\|_p = \limsup_{n\to\infty}\frac{1}{n+1}\|S_n\|_p \le \epsilon M.$$

As ϵ is arbitrary, $\lim_{n\to\infty}\|Y_n\|_p = 0$.

Remark There are strengthenings of (a) in 276Xd, and of (b) in 276Ya.

273X Basic exercises (a) In part (b) of the proof of 273B, use Bienaymé's equality to show that $\lim_{m\to\infty}\sup_{n\ge m}\Pr(|S_n-S_m|\ge\epsilon)=0$ for every $\epsilon>0$, so that we can apply the argument of part (a) of the proof directly, without appealing to 242F or 245G or even 244E.

(b) Show that $\sum_{n=0}^\infty\frac{(-1)^{\omega(n)}}{n+1}$ is defined in $\mathbb{R}$ for almost every $\boldsymbol{\omega}=\langle\omega(n)\rangle_{n\in\mathbb{N}}$ in $\{0,1\}^{\mathbb{N}}$, where $\{0,1\}^{\mathbb{N}}$ is given its usual measure (254J).

(c) Let $\langle E_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of measurable sets in a probability space, all with the same non-zero measure. Let $\langle a_n\rangle_{n\in\mathbb{N}}$ be a sequence of non-negative real numbers such that $\sum_{n=0}^\infty a_n=\infty$. Show that $\sum_{n=0}^\infty a_n\chi E_n=\infty$ a.e. (*Hint*: Take a strictly increasing sequence $\langle k_n\rangle_{n\in\mathbb{N}}$ such that $d_n=\sum_{i=k_n+1}^{k_{n+1}}a_i\ge 1$ for each n. Set $c_i=\frac{a_i}{(n+1)d_n}$ for $k_n<i\le k_{n+1}$; show that $\sum_{n=0}^\infty c_n^2<\infty=\sum_{n=0}^\infty c_n$. Apply 273D with $X_n=c_n\chi E_n$ and $b_n=\sqrt{\sum_{i=0}^n c_i}$.)

>(d) Take any $q\in[0,1]$, and give $\mathcal{P}\mathbb{N}$ a measure μ such that

$$\mu\{a:I\subseteq a\}=q^{\#(I)}$$

for every $I\subseteq\mathbb{N}$, as in 254Xg. Show that for μ-almost every $a\subseteq\mathbb{N}$,

$$\lim_{n\to\infty}\frac{1}{n+1}\#(a\cap\{0,\dots,n\})=q.$$

>(e) Let μ be the usual probability measure on $\mathcal{P}\mathbb{N}$ (254Jb), and for $r\ge 1$ let μ^r be the product probability measure on $(\mathcal{P}\mathbb{N})^r$. Show that

$$\lim_{n\to\infty}\frac{1}{n+1}\#(a_1\cap\ldots\cap a_r\cap\{0,\dots,n\})=2^{-r},$$

$$\lim_{n\to\infty}\frac{1}{n+1}\#((a_1\cup\ldots\cup a_r)\cap\{0,\dots,n\})=1-2^{-r}$$

for μ^r-almost every $(a_1,\dots,a_r)\in(\mathcal{P}\mathbb{N})^r$.

(f) Let μ be the usual probability measure on $\mathcal{P}\mathbb{N}$, and b any infinite subset of $\mathbb{N}$. Show that $\lim_{n\to\infty}\frac{\#(a\cap b\cap\{0,\dots,n\})}{\#(b\cap\{0,\dots,n\})}=\frac{1}{2}$ for almost every $a\subseteq\mathbb{N}$.

>(g) For each $x\in[0,1]$, let $\epsilon_k(x)$ be the kth digit in the decimal expansion of x (choose for yourself what to do with $0{\cdot}100\ldots=0{\cdot}099\ldots$). Show that $\lim_{k\to\infty}\frac{1}{k}\#(\{j:j\le k,\ \epsilon_j(x)=7\})=\frac{1}{10}$ for almost every $x\in[0,1]$.

(h) Let $\langle F_n\rangle_{n\in\mathbb{N}}$ be a sequence of distribution functions for real-valued random variables, in the sense of 271Ga, and F another distribution function; suppose that $\lim_{n\to\infty}F_n(q)=F(q)$ for every $q\in\mathbb{Q}$ and $\lim_{n\to\infty}F_n(a^-)=F(a^-)$ whenever $F(a^-)<F(a)$, where I write $F(a^-)$ for $\lim_{x\uparrow a}F(x)$. Show that $F_n\to F$ uniformly.

>(i) Let (Ω,Σ,μ) be a probability space and $\langle X_n\rangle_{n\in\mathbb{N}}$ an independent identically distributed sequence of real-valued random variables on Ω with common distribution function F. For $a\in\mathbb{R}$, $n\in\mathbb{N}$ and $\omega\in\bigcap_{i\le n}\operatorname{dom}X_i$ set

$$F_n(\omega,a)=\frac{1}{n+1}\#(\{i:i\le n,\ X_i(\omega)\le a\}).$$

Show that

$$\lim_{n\to\infty}\sup_{a\in\mathbb{R}}|F_n(\omega,a)-F(a)|=0$$

for almost every $\omega\in\Omega$.

(j) Let (Ω,Σ,μ) be a probability space, and λ the product measure on $\Omega^{\mathbb{N}}$. Let $f:\Omega\to\mathbb{R}$ be a function, and set $f^*(\boldsymbol{\omega})=\limsup_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^n f(\omega_i)$ for $\boldsymbol{\omega}=\langle\omega_n\rangle_{n\in\mathbb{N}}\in\Omega^{\mathbb{N}}$. Show that $\overline{\int}f^*d\lambda=\overline{\int}fd\mu$ whenever the right-hand-side is finite. (*Hint*: 133J(a-i).)

(k) Find an independent sequence $\langle X_n\rangle_{n\in\mathbb{N}}$ of random variables with zero expectation such that $\|X_n\|_1=1$ and $\|\frac{1}{n+1}\sum_{i=0}^n X_i\|_1\ge\frac{1}{2}$ for every $n\in\mathbb{N}$. (*Hint*: take $\Pr(X_n\ne 0)$ very small.)

(l) Use 272S to prove 273Nb in the case $p = 2$.

(m) Find an independent sequence $\langle X_n\rangle_{n\in\mathbb{N}}$ of random variables with zero expectation such that $\|X_n\|_\infty = \|\frac{1}{n+1}\sum_{i=0}^n X_i\|_\infty = 1$ for every $n \in \mathbb{N}$.

(n) Repeat the work of this section for complex-valued random variables.

273Y Further exercises (a) Let (Ω, Σ, μ) be a probability space, and λ the product measure on $\Omega^{\mathbb{N}}$. Suppose that f is a real-valued function, defined on a subset of Ω, such that

$$h(\boldsymbol{\omega}) = \lim_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^n f(\omega_i)$$

exists in $\mathbb{R}$ for λ-almost every $\boldsymbol{\omega} = \langle\omega_n\rangle_{n\in\mathbb{N}}$ in $\Omega^{\mathbb{N}}$. Show (i) that f has conegligible domain (ii) f is $\hat\Sigma$-measurable, where $\hat\Sigma$ is the domain of the completion of μ (iii) there is an $a \in \mathbb{R}$ such that $h = a$ almost everywhere in $\Omega^{\mathbb{N}}$ (iv) f is integrable, with $\int f d\mu = a$.

(b) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a sequence of random variables with finite variance. Suppose that $\lim_{n\to\infty}\mathbb{E}(X_n) = \infty$ and $\liminf_{n\to\infty}\frac{\mathbb{E}(X_n^2)}{(\mathbb{E}(X_n))^2} \le 1$. Show that $\limsup_{n\to\infty} X_n = \infty$ a.e.

273 Notes and comments I have tried in this section to offer the most useful of the standard criteria for pointwise convergence of averages of independent random variables. In my view the strong law of large numbers, like Fubini's theorem, is one of the crucial steps in measure theory, where the subject changes character. Theorems depending on the strong law have a kind of depth and subtlety to them which is missing in other parts of the subject. I have described only a handful of applications here, but I hope that 273G, 273J, 273Xd, 273Xg and 273Xi will give an idea of what is to be expected. These do have rather different weights. Of the four, only 273J requires the full resources of this chapter; the others can be deduced from the essentially simpler version in 273Xi.

273Xi is the 'fundamental theorem of statistics' or 'Glivenko-Cantelli theorem'. The $F_n(., a)$ are 'statistics', computed from the X_i; they are the 'empirical distributions', and the theorem says that, almost surely, $F_n \to F$ uniformly. (I say 'uniformly' to make the result look more striking, but of course the real content is that $F_n(., a) \to F(a)$ almost surely for each a; the extra step is just 273Xh.)

I have included 273N to show that independence is quite as important in questions of norm-convergence as it is in questions of pointwise convergence. It does not really rely on any form of the strong law; in the proof I quote 273E as a quick way of disposing of the 'uniformly bounded parts' X'_n, but of course Bienaymé's equality (272S) is already enough to show that if $\langle X'_n\rangle_{n\in\mathbb{N}}$ is an independent uniformly bounded sequence of random variables with zero expectation, then $\|\frac{1}{n+1}(X_0 + \ldots + X_n)\|_p \to 0$ for $p = 2$, and therefore for every $p < \infty$.

The proofs of 273H, 273I and 273Na all involve 'truncation'; the expression of a random variable X as the sum of a bounded random variable and a tail. This is one of the most powerful techniques in the subject, and will appear again in §§274 and 276. In 273Na I used a slightly different formulation of the method, solely because it matched the definition of 'uniformly integrable' more closely.

274 The central limit theorem

The second of the great theorems to which this chapter is devoted is of a new type. It is a limit theorem, but the limit involved is a limit of *distributions*, not of functions (as in the strong limit theorem above or the martingale theorem below), nor of equivalence classes of functions (as in Chapter 24). I give three forms of the theorem, in 274I-274K, all drawn as corollaries of Theorem 274G; the proof is spread over 274C-274G. In 274A-274B and 274M I give the most elementary properties of the normal distribution.

274A The normal distribution We need some facts from basic probability theory.

(a) Recall that

$$\int_{-\infty}^{\infty} e^{-x^2/2}dx = \sqrt{2\pi}$$

(263G). Consequently, if we set

$$\mu_G E = \frac{1}{\sqrt{2\pi}}\int_E e^{-x^2/2}dx$$

for every Lebesgue measurable set E, μ_G is a Radon probability measure (256E); we call it the **standard normal distribution**. The corresponding distribution function is

$$\Phi(a) = \mu_G\,]-\infty, a] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-x^2/2}dx$$

for $a \in \mathbb{R}$; for the rest of this section I will reserve the symbol Φ for this function.

Writing Σ for the algebra of Lebesgue measurable subsets of $\mathbb{R}$, $(\mathbb{R}, \Sigma, \mu_G)$ is a probability space. Note that it is complete, and has the same negligible sets as Lebesgue measure, because $e^{-x^2/2} > 0$ for every x (cf. 234Lc).

(b) A random variable X is **standard normal** if its distribution is μ_G; that is, if the function $x \mapsto \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ is a density function for X. The point of the remarks in (a) is that there are such random variables; for instance, take the probability space $(\mathbb{R}, \Sigma, \mu_G)$ there, and set $X(x) = x$ for every $x \in \mathbb{R}$.

(c) If X is a standard normal random variable, then

$$\mathbb{E}(X) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} xe^{-x^2/2}dx = 0,$$

$$\mathrm{Var}(X) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^2e^{-x^2/2}dx = 1$$

by 263H.

(d) More generally, a random variable X is **normal** if there are $a \in \mathbb{R}$, $\sigma > 0$ such that $Z = (X - a)/\sigma$ is standard normal. In this case $X = \sigma Z + a$ so $\mathbb{E}(X) = \sigma\mathbb{E}(Z) + a = a$, $\mathrm{Var}(X) = \sigma^2\,\mathrm{Var}(Z) = \sigma^2$.

We have, for any $c \in \mathbb{R}$,

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{\infty}^{c} e^{-(x-a)^2/2\sigma^2}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{(c-a)/\sigma} e^{-y^2/2}dy$$

(substituting $x = a + \sigma y$ for $-\infty < y \le (c - a)/\sigma$)

$$= \Pr(Z \le \tfrac{c-a}{\sigma}) = \Pr(X \le c).$$

So $x \mapsto \frac{1}{\sigma\sqrt{2\pi}}e^{-(x-a)^2/2\sigma^2}$ is a density function for X (271Ib). Conversely, of course, a random variable with such a density function is normal, with expectation a and variance σ^2. The **normal distributions** are the distributions with these density functions.

(e) If Z is standard normal, so is $-Z$, because

$$\Pr(-Z \le a) = \Pr(Z \ge -a) = \frac{1}{\sqrt{2\pi}}\int_{-a}^{\infty} e^{-x^2/2}dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-x^2/2}dx.$$

The definition in the first sentence of (d) now makes it obvious that if X is normal, so is $a + bX$ for any $a \in \mathbb{R}$, $b \in \mathbb{R} \setminus \{0\}$.

274B Proposition Let $X_1, \dots, X_n$ be independent normal random variables. Then $Y = X_1 + \ldots + X_n$ is normal, with $\mathbb{E}(Y) = \mathbb{E}(X_1) + \ldots + \mathbb{E}(X_n)$ and $\mathrm{Var}(Y) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)$.

proof There are innumerable proofs of this fact; the following one gives me a chance to show off the power of Chapter 26, but of course (at the price of some disagreeable algebra) 272U also gives the result.

(a) Consider first the case $n = 2$. Setting $a_i = \mathbb{E}(X_i)$, $\sigma_i = \sqrt{\mathrm{Var}(X_i)}$, $Z_i = (X_i - a_i)/\sigma_i$ we get independent standard normal variables Z_1, Z_2. Set $\rho = \sqrt{\sigma_1^2 + \sigma_2^2}$, and express σ_1, σ_2 as $\rho\cos\theta$, $\rho\sin\theta$. Consider $U = \cos\theta Z_1 + \sin\theta Z_2$. We know that (Z_1, Z_2) has a density function

$$(\zeta_1, \zeta_2) \mapsto g(\zeta_1, \zeta_2) = \frac{1}{2\pi\sigma_1\sigma_2}e^{-(\zeta_1^2+\zeta_2^2)/2}$$

(272I). Consequently, for any $c \in \mathbb{R}$,

$$\Pr(U \le c) = \int_F g(z)dz,$$

where $F=\{(\zeta_1,\zeta_2):\zeta_1\cos\theta+\zeta_2\sin\theta\le c\}$. But now let T be the matrix

$$\begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix}.$$

Then it is easy to check that

$$T^{-1}[F]=\{(\eta_1,\eta_2):\eta_1\le c\},$$

$$\det T=1,$$

$$g(Ty)=g(y)\text{ for every }y\in\mathbb{R}^2,$$

so by 263A

$$\Pr(U\le c)=\int_F g(z)dz=\int_{T^{-1}[F]}g(Ty)dy=\int_{]-\infty,c]\times\mathbb{R}}g(y)dy=\Pr(Z_1\le c)=\Phi(c).$$

As this is true for every $c\in\mathbb{R}$, U also is standard normal (I am appealing to 271Ga again). But

$$X_1+X_2=\sigma_1Z_1+\sigma_2Z_2+a_1+a_2=\rho U+a_1+a_2,$$

so X_1+X_2 is normal.

(b) Now we can induce on n. If $n=1$ the result is trivial. For the inductive step to $n+1\ge 2$, we know that $X_1+\ldots+X_n$ is normal, by the inductive hypothesis, and that X_{n+1} is independent of $X_1+\ldots+X_n$, by 272L. So $X_1+\ldots+X_n+X_{n+1}$ is normal, by (a).

The computation of the expectation and variance of $X_1+\ldots+X_n$ is immediate from 271Ab and 272S.

274C Lemma Let $U_0,\ldots,U_n,V_0,\ldots,V_n$ be independent real-valued random variables and $h:\mathbb{R}\to\mathbb{R}$ a bounded Borel measurable function. Then

$$|\mathbb{E}\bigl(h(\sum_{i=0}^n U_i)-h(\sum_{i=0}^n V_i)\bigr)|\le\sum_{i=0}^n\sup_{t\in\mathbb{R}}|\mathbb{E}\bigl(h(t+U_i)-h(t+V_i)\bigr)|.$$

proof For $0\le j\le n+1$, set $Z_j=\sum_{i=0}^{j-1}U_i+\sum_{i=j}^n V_i$, taking $Z_0=\sum_{i=0}^n V_i$ and $Z_{n+1}=\sum_{i=0}^n U_i$, and for $j\le n$ set $W_j=\sum_{i=0}^{j-1}U_j+\sum_{i=j+1}^n V_j$, so that $Z_j=W_j+V_j$ and $Z_{j+1}=W_j+U_j$ and W_j, U_j and V_j are independent (I am appealing to 272K, as in 272L). Then

$$\begin{aligned}|\mathbb{E}\bigl(h(\sum_{i=0}^n U_i)-h(\sum_{i=0}^n V_i)\bigr)|&=|\mathbb{E}\bigl(\sum_{i=0}^n h(Z_{i+1})-h(Z_i)\bigr)|\\&\le\sum_{i=0}^n|\mathbb{E}\bigl(h(Z_{i+1})-h(Z_i)\bigr)|\\&=\sum_{i=0}^n|\mathbb{E}\bigl(h(W_i+U_i)-h(W_i+V_i)\bigr)|.\end{aligned}$$

To estimate this sum I turn it into a sum of integrals, as follows. For each i, let ν_{W_i} be the distribution of W_i, and so on. Because $(w,u)\mapsto w+u$ is continuous, therefore Borel measurable, $(w,u)\mapsto h(w,u)$ also is Borel measurable; accordingly $(w,u,v)\mapsto h(w+u)-h(w+v)$ is measurable for each of the product measures $\nu_{W_i}\times\nu_{U_i}\times\nu_{V_i}$ on $\mathbb{R}^3$, and 271E and 272G give us

$$\begin{aligned}\bigl|\mathbb{E}\bigl(h(W_i+U_i)&-h(W_i+V_i)\bigr)\bigr|\\&=\bigl|\int h(w+u)-h(w+v)(\nu_{W_i}\times\nu_{U_i}\times\nu_{V_i})d(w,u,v)\bigr|\\&=\bigl|\int\bigl(\int h(w+u)-h(w+v)(\nu_{U_i}\times\nu_{V_i})d(u,v)\bigr)\nu_{W_i}(dw)\bigr|\\&\le\int\bigl|\int h(w+u)-h(w+v)(\nu_{U_i}\times\nu_{V_i})d(u,v)\bigr|\nu_{W_i}(dw)\\&=\int\bigl|\mathbb{E}\bigl(h(w+U_i)-h(w+V_i)\bigr)\bigr|\nu_{W_i}(dw)\\&\le\sup_{t\in\mathbb{R}}\bigl|\mathbb{E}\bigl(h(t+U_i)-h(t+V_i)\bigr)\bigr|.\end{aligned}$$

So we get

$$\begin{aligned}|\mathbb{E}\big(h(\sum_{i=0}^n U_i) - h(\sum_{i=0}^n V_i)\big)| &\le \sum_{i=0}^n |\mathbb{E}\big(h(W_i + U_i) - h(W_i + V_i)\big)| \\ &\le \sum_{i=0}^n \sup_{t\in\mathbb{R}} |\mathbb{E}\big(h(t + U_i) - h(t + V_i)\big)|,\end{aligned}$$

as required.

274D Lemma Let $h : \mathbb{R} \to \mathbb{R}$ be a bounded three-times-differentiable function such that $M_2 = \sup_{x\in\mathbb{R}} |h''(x)|$, $M_3 = \sup_{x\in\mathbb{R}} |h'''(x)|$ are both finite. Let $\epsilon > 0$.

(a) Let U be a real-valued random variable of zero expectation and finite variance σ^2. Then for any $t \in \mathbb{R}$ we have

$$|\mathbb{E}(h(t + U)) - h(t) - \frac{\sigma^2}{2}h''(t)| \le \frac{1}{6}\epsilon M_3\sigma^2 + M_2\mathbb{E}(\psi_\epsilon(U))$$

where $\psi_\epsilon(x) = 0$ if $|x| \le \epsilon$, x^2 if $|x| > \epsilon$.

(b) Let $U_0, \dots , U_n, V_0, \dots , V_n$ be independent random variables with finite variances, and suppose that $\mathbb{E}(U_i) = \mathbb{E}(V_i) = 0$, $\mathrm{Var}(U_i) = \mathrm{Var}(V_i) = \sigma_i^2$ for every $i \le n$. Then

$$\begin{aligned}&|\mathbb{E}\big(h(\sum_{i=0}^n U_i) - h(\sum_{i=0}^n V_i)\big)| \\ &\qquad\le \frac{1}{3}\epsilon M_3 \sum_{i=0}^n \sigma_i^2 + M_2\sum_{i=0}^n \mathbb{E}\big(\psi_\epsilon(U_i)\big) + M_2\sum_{i=0}^n \mathbb{E}\big(\psi_\epsilon(V_i)\big).\end{aligned}$$

proof (a) The point is that, by Taylor's theorem with remainder,

$$|h(t + x) - h(t) - xh'(t)| \le \frac{1}{2}M_2x^2,$$

$$|h(t + x) - h(t) - xh'(t) - \frac{1}{2}x^2h''(t)| \le \frac{1}{6}M_3|x|^3$$

for every $x \in \mathbb{R}$. So

$$|h(t + x) - h(t) - xh'(t) - \frac{1}{2}x^2h''(t)| \le \min(\frac{1}{6}M_3|x|^3, M_2x^2) \le \frac{1}{6}\epsilon M_3x^2 + M_2\psi_\epsilon(x).$$

Integrating with respect to the distribution of U, we get

$$\begin{aligned}|\mathbb{E}\big(h(t + U)) - h(t) - \frac{1}{2}h''(t)\sigma^2\big)| &= |\mathbb{E}(h(t + U)) - h(t) - h'(t)\mathbb{E}(U) - \frac{1}{2}h''(t)\mathbb{E}(U^2)| \\ &= |\mathbb{E}\big(h(t + U) - h(t) - h'(t)U - \frac{1}{2}h''(t)U^2\big)| \\ &\le \mathbb{E}\big(|h(t + U) - h(t) - h'(t)U - \frac{1}{2}h''(t)U^2|\big) \\ &\le \mathbb{E}\big(\frac{1}{6}\epsilon M_3U^2 + M_2\psi_\epsilon(U)\big) \\ &= \frac{1}{6}\epsilon M_3\sigma^2 + M_2\mathbb{E}(\psi_\epsilon(U)),\end{aligned}$$

as claimed.

(b) By 274C,

$$\begin{aligned}|\mathbb{E}\big(h(\sum_{i=0}^n U_i) - h(\sum_{i=0}^n V_i)\big)| &\le \sum_{i=0}^n \sup_{t\in\mathbb{R}} |\mathbb{E}\big(h(t + U_i) - h(t + V_i)\big)| \\ &\le \sum_{i=0}^n \sup_{t\in\mathbb{R}}\big(|\mathbb{E}(h(t + U_i)) - h(t) - \frac{1}{2}h''(t)\sigma_i^2| \\ &\qquad + |\mathbb{E}(h(t + V_i)) - h(t) - \frac{1}{2}h''(t)\sigma_i^2|\big),\end{aligned}$$

which by (a) above is at most

$$\sum_{i=0}^n \frac{1}{3}\epsilon M_3\sigma_i^2 + M_2\mathbb{E}(\psi_\epsilon(U_i)) + M_2\mathbb{E}(\psi_\epsilon(V_i)),$$

as claimed.

274E Lemma For any $\epsilon > 0$, there is a three-times-differentiable function $h : \mathbb{R} \to [0,1]$, with continuous third derivative, such that $h(x) = 1$ for $x \le -\epsilon$ and $h(x) = 0$ for $x \ge \epsilon$.

proof Let $f :]-\epsilon,\epsilon[\to]0,\infty[$ be any twice-differentiable function such that

$$\lim_{x\downarrow-\epsilon} f^{(n)}(x) = \lim_{x\uparrow\epsilon} f^{(n)}(x) = 0$$

for $n = 0$, 1 and 2, writing $f^{(n)}$ for the nth derivative of f; for instance, you could take $f(x) = (\epsilon^2 - x^2)^3$, or $f(x) = \exp(-\frac{1}{\epsilon^2-x^2})$. Now set

$$h(x) = 1 - \int_{-\epsilon}^x f/\int_{-\epsilon}^\epsilon f$$

for $|x| \le \epsilon$.

274F Lindeberg's theorem Let $\epsilon > 0$. Then there is a $\delta > 0$ such that whenever $X_0,\dots,X_n$ are independent real-valued random variables such that

$$\mathbb{E}(X_i) = 0 \text{ for every } i \le n,$$

$$\sum_{i=0}^n \mathrm{Var}(X_i) = 1,$$

$$\sum_{i=0}^n \mathbb{E}(\psi_\delta(X_i)) \le \delta$$

(writing $\psi_\delta(x) = 0$ if $|x| \le \delta$, x^2 if $|x| > \delta$), then

$$\left|\Pr(\sum_{i=0}^n X_i \le a) - \Phi(a)\right| \le \epsilon$$

for every $a \in \mathbb{R}$.

proof (a) Let $h : \mathbb{R} \to [0,1]$ be a three-times-differentiable function, with continuous third derivative, such that $\chi\,]-\infty,-\epsilon] \le h \le \chi\,]-\infty,\epsilon]$, as in 274E. Set

$$M_2 = \sup_{x\in\mathbb{R}} |h''(x)| = \sup_{|x|\le\epsilon} |h''(x)|,$$

$$M_3 = \sup_{x\in\mathbb{R}} |h'''(x)| = \sup_{|x|\le\epsilon} |h'''(x)|;$$

because h''' is continuous, both are finite. Write $\epsilon' = \epsilon(1 - \frac{2}{\sqrt{2\pi}}) > 0$, and let $\eta > 0$ be such that

$$(\frac{1}{3}M_3 + 2M_2)\eta \le \epsilon'.$$

Note that $\lim_{m\to\infty} \psi_m(x) = 0$ for every x, so if X is a random variable of finite variance we must have $\lim_{m\to\infty} \mathbb{E}(\psi_m(X)) = 0$, by Lebesgue's Dominated Convergence Theorem; let $m \ge 1$ be such that $\mathbb{E}(\psi_m(Z)) \le \eta$, where Z is some (or any) standard normal random variable. Finally, take $\delta > 0$ such that $\delta \le \eta$, $\delta + \delta^2 \le (\eta/m)^2$.

(I hope that you have seen enough ϵ-δ arguments not to be troubled by any expectation of understanding the reasons for each particular formula here before reading the rest of the argument. But the formula $\frac{1}{3}M_3 + 2M_2$, in association with ψ_δ, should recall 274D.)

(b) Let $X_0,\dots,X_n$ be independent random variables with zero expectation such that $\sum_{i=0}^n \mathrm{Var}(X_i) = 1$ and $\sum_{i=0}^n \mathbb{E}(\psi_\delta(X_i)) \le \delta$. We need an auxiliary sequence $Z_0,\dots,Z_n$ of standard normal random variables to match against the X_i. To create this, I use the following device. Suppose that the probability space underlying $X_0,\dots,X_n$ is (Ω,Σ,μ). Set $\Omega' = \Omega \times \mathbb{R}^{n+1}$, and let μ' be the product measure on Ω', where Ω is given the measure μ and each factor $\mathbb{R}$ of $\mathbb{R}^{n+1}$ is given the measure μ_G. Set $X'_i(\omega,z) = X_i(\omega)$ and $Z_i(\omega,z) = \zeta_i$ for $\omega \in \mathrm{dom}\,X_i$, $z = (\zeta_0,\dots,\zeta_n) \in \mathbb{R}^{n+1}$, $i \le n$. Then $X'_0,\dots,X'_n,Z_0,\dots,Z_n$ are independent, and each X'_i has the same distribution as X_i (272Mb). Consequently $S' = X'_0+\dots+X'_n$ has the same distribution as $S = X_0+\dots+X_n$ (using 272T, or otherwise); so that $\mathbb{E}(g(S')) = \mathbb{E}(g(S))$ for any bounded Borel measurable function g (using 271E). Also each Z_i has distribution μ_G, so is standard normal.

(c) Write $\sigma_i = \sqrt{\mathrm{Var}(X_i)}$ for each i, and set $K = \{i : i \le n,\ \sigma_i > 0\}$. Observe that $\eta/\sigma_i \ge m$ for each $i \in K$. **P** We know that

$$\sigma_i^2 = \mathrm{Var}(X_i) = \mathbb{E}(X_i^2) \le \mathbb{E}(\delta^2 + \psi_\delta(X_i)) = \delta^2 + \mathbb{E}(\psi_\delta(X_i)) \le \delta^2 + \delta,$$

so

$$\eta/\sigma_i \ge \eta/\sqrt{\delta+\delta^2} \ge m$$

by the choice of δ. **Q**

(d) Consider the independent normal random variables $\sigma_i Z_i$. We have $\mathbb{E}(\sigma_i Z_i) = \mathbb{E}(X_i') = 0$ and $\mathrm{Var}(\sigma_i Z_i) = \mathrm{Var}(X_i') = \sigma_i^2$ for each i, so that $Z = Z_0 + \ldots + Z_n$ has expectation 0 and variance $\sum_{i=0}^n \sigma_i^2 = 1$; moreover, by 274B, Z is normal, so in fact it is standard normal. Now we have

$$\sum_{i=0}^n \mathbb{E}(\psi_\eta(\sigma_i Z_i)) = \sum_{i\in K} \mathbb{E}(\psi_\eta(\sigma_i Z_i)) = \sum_{i\in K} \sigma_i^2 \mathbb{E}(\psi_{\eta/\sigma_i}(Z_i))$$

(because $\sigma^2\psi_{\eta/\sigma}(x) = \psi_\eta(\sigma x)$ whenever $x \in \mathbb{R}$, $\sigma > 0$)

$$= \sum_{i\in K} \sigma_i^2 \mathbb{E}(\psi_{\eta/\sigma_i}(Z)) \le \sum_{i\in K} \sigma_i^2 \mathbb{E}(\psi_m(Z))$$

(because, by (c), $\eta/\sigma_i \ge m$ for every $i \in K$, so $\psi_{\eta/\sigma_i}(t) \le \psi_m(t)$ for every t)

$$\le \sum_{i\in K} \sigma_i^2 \eta$$

(by the choice of m)

$$= \eta.$$

On the other hand, we surely have

$$\textstyle\sum_{i=0}^n \mathbb{E}(\psi_\eta(X_i')) = \sum_{i=0}^n \mathbb{E}(\psi_\eta(X_i)) \le \sum_{i=0}^n \mathbb{E}(\psi_\delta(X_i)) \le \delta \le \eta.$$

(e) For any real number t, set

$$h_t(x) = h(x-t)$$

for each $x \in \mathbb{R}$. Then h_t is three-times-differentiable, with $\sup_{x\in\mathbb{R}} |h_t''(x)| = M_2$ and $\sup_{x\in\mathbb{R}} |h'''(x)| = M_3$. Consequently

$$|\mathbb{E}(h_t(S)) - \mathbb{E}(h_t(Z))| \le \epsilon'.$$

P By 274Db,

$$\begin{aligned}
|\mathbb{E}(h_t(S)) - \mathbb{E}(h_t(Z))| &= |\mathbb{E}(h_t(S')) - \mathbb{E}(h_t(Z))| \\
&= |\mathbb{E}(h_t(\sum_{i=0}^n X_i')) - \mathbb{E}(h_t(\sum_{i=0}^n \sigma_i Z_i))| \\
&\le \frac{1}{3}\eta M_3 \sum_{i=0}^n \sigma_i^2 + M_2 \sum_{i=0}^n \mathbb{E}(\psi_\eta(X_i)) + M_2 \sum_{i=0}^n \mathbb{E}(\psi_\eta(\sigma_i Z_i)) \\
&\le \frac{1}{3}\eta M_3 + M_2\eta + M_2\eta \le \epsilon',
\end{aligned}$$

by the choice of η. **Q**

(f) Now take any $a \in \mathbb{R}$. We have

$$\chi\,]-\infty, a-2\epsilon] \le h_{a-\epsilon} \le \chi\,]\infty, a] \le h_{a+\epsilon} \le \chi\,]-\infty, a+\epsilon].$$

Note also that, for any b,

$$\Phi(b+2\epsilon) = \Phi(b) + \tfrac{1}{\sqrt{2\pi}}\int_b^{b+2\epsilon} e^{-x^2/2}dx \le \Phi(b) + \tfrac{2\epsilon}{\sqrt{2\pi}} = \Phi(b) + \epsilon - \epsilon'.$$

Consequently

$$\begin{aligned}
\Phi(a) - \epsilon &\le \Phi(a-2\epsilon) - \epsilon' = \Pr(Z \le a-2\epsilon) - \epsilon' \\
&\le \mathbb{E}(h_{a-\epsilon}(Z)) - \epsilon' \le \mathbb{E}(h_{a-\epsilon}(S)) \le \Pr(S \le a) \\
&\le \mathbb{E}(h_{a+\epsilon}(S)) \le \mathbb{E}(h_{a+\epsilon}(Z)) + \epsilon' \le \Pr(Z \le a+2\epsilon) + \epsilon' \\
&= \Phi(a+2\epsilon) + \epsilon' \le \Phi(a) + \epsilon.
\end{aligned}$$

But this means just that

$$\left|\Pr(\sum_{i=0}^{n} X_i \le a) - \Phi(a)\right| \le \epsilon,$$

as claimed.

274G Central Limit Theorem Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of random variables, all with zero expectation and finite variance; write $s_n = \sqrt{\sum_{i=0}^{n}\operatorname{Var}(X_i)}$ for each n. Suppose that

$$\lim_{n\to\infty}\frac{1}{s_n^2}\sum_{i=0}^{n}\mathbb{E}(\psi_{\delta s_n}(X_i)) = 0 \text{ for every } \delta > 0,$$

writing $\psi_\delta(x) = 0$ if $|x| \le \delta$, x^2 if $|x| > \delta$. Set

$$S_n = \frac{1}{s_n}(X_0 + \ldots + X_n)$$

for each $n \in \mathbb{N}$ such that $s_n > 0$. Then

$$\lim_{n\to\infty}\Pr(S_n \le a) = \Phi(a)$$

uniformly for $a \in \mathbb{R}$.

proof Given $\epsilon > 0$, take $\delta > 0$ as in Lindeberg's theorem (274F). Then for all n large enough,

$$\frac{1}{s_n^2}\sum_{i=0}^{n}\mathbb{E}(\psi_{\delta s_n}(X_i)) \le \delta.$$

Fix on any such n. Of course we have $s_n > 0$. Set

$$X'_i = \frac{1}{s_n}X_i \text{ for } i \le n;$$

then $X'_0, \ldots, X'_n$ are independent, with zero expectation,

$$\sum_{i=0}^{n}\operatorname{Var}(X'_i) = \sum_{i=0}^{n}\frac{1}{s_n^2}\operatorname{Var}(X_i) = 1,$$

$$\sum_{i=0}^{n}\mathbb{E}(\psi_\delta(X'_i)) = \sum_{i=0}^{n}\frac{1}{s_n^2}\mathbb{E}(\psi_{\delta s_n}(X_i)) \le \delta.$$

By 274F,

$$\left|\Pr(S_n \le a) - \Phi(a)\right| = \left|\Pr(\sum_{i=0}^{n} X'_i \le a) - \Phi(a)\right| \le \epsilon$$

for every $a \in \mathbb{R}$. Since this is true for all n large enough, we have the result.

274H Remarks (a) The condition

$$\lim_{n\to\infty}\frac{1}{s_n^2}\sum_{i=0}^{n}\mathbb{E}(\psi_{\epsilon s_n}(X_i)) = 0 \text{ for every } \epsilon > 0$$

is called **Lindeberg's condition**, following Lindeberg 22.

(b) Lindeberg's condition is necessary as well as sufficient, in the following sense. Suppose that $\langle X_n\rangle_{n\in\mathbb{N}}$ is an independent sequence of real-valued random variables with zero expectation and finite variance; write $\sigma_n = \sqrt{\operatorname{Var}(X_n)}$, $s_n = \sqrt{\sum_{i=0}^{n}\operatorname{Var}(X_i)}$ for each n. Suppose that $\lim_{n\to\infty} s_n = \infty$, $\lim_{n\to\infty}\frac{\sigma_n}{s_n} = 0$ and that $\lim_{n\to\infty}\Pr(S_n \le a) = \Phi(a)$ for each $a \in \mathbb{R}$, where $S_n = \frac{1}{s_n}(X_0 + \ldots + X_n)$. Then

$$\lim_{n\to\infty}\frac{1}{s_n^2}\sum_{i=0}^{n}\mathbb{E}(\psi_{\epsilon s_n}(X_i)) = 0$$

for every $\epsilon > 0$. (Feller 66, §XV.6, Theorem 3; Loève 77, §21.2.)

(c) The proof of 274F-274G here is adapted from Feller 66, §VIII.4. It has the virtue of being 'elementary', in that it does not involve characteristic functions. Of course this has to be paid for by a number of detailed estimations; and – what is much more serious – it leaves us without one of the most powerful techniques for describing distributions. The proof does offer a method of bounding

$$|\Pr(S_n \le a) - \Phi(a)|;$$

but it should be said that the bounds obtained are not useful ones, being grossly over-pessimistic, at least in the readily analysable cases. (For instance, a better bound, in many cases, is given by the Berry-Esséen theorem: if $\langle X_n\rangle_{n\in\mathbb{N}}$ is independent and identically distributed, with zero expectation, and the common values of $\sqrt{\mathbb{E}(X_n^2)}$, $\mathbb{E}(|X_n|^3)$ are σ, $\rho<\infty$, then

$$|\Pr(S_n\le a)-\Phi(a)|\le\frac{33\rho}{4\sigma^3\sqrt{n+1}};$$

see Feller 66, §XVI.5, Loève 77, §21.3, or Hall 82.) Furthermore, when $|a|$ is large, $\Phi(a)$ is exceedingly close to either 0 or 1, so that any uniform bound for $|\Pr(S\le a)-\Phi(a)|$ gives very little information; a great deal of work has been done on estimating the tails of such distributions more precisely, subject to special conditions. For instance, if $X_0,\dots,X_n$ are independent random variables, of zero expectation, uniformly bounded with $|X_i|\le K$ almost everywhere for each i, $Y=X_0+\ldots+X_n$, $s=\sqrt{\mathrm{Var}(Y)}>0$, $S=\frac{1}{s}Y$, then for any $\alpha\in[0,s/K]$

$$\Pr(|S|\ge\alpha)\le 2\exp\Bigl(\frac{-\alpha^2}{2(1+\frac{\alpha K}{2s})^2}\Bigr)\simeq 2e^{-\alpha^2/2}$$

if $s\gg\alpha K$ (Rényi 70, §VII.4, Theorem 1).

I now list some of the standard cases in which Lindeberg's conditions are satisfied, so that we may apply the theorem.

274I Corollary Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables, all with the same distribution, and suppose that their common expectation is 0 and their common variance is finite and not zero. Write σ for the common value of $\sqrt{\mathrm{Var}(X_n)}$, and set

$$S_n=\frac{1}{\sigma\sqrt{n+1}}(X_0+\ldots+X_n)$$

for each $n\in\mathbb{N}$. Then

$$\lim_{n\to\infty}\Pr(S_n\le a)=\Phi(a)$$

uniformly for $a\in\mathbb{R}$.

proof In the language of 274G-274H, we have $\sigma_n=\sigma$, $s_n=\sigma\sqrt{n+1}$, so the first two conditions are surely satisfied; moreover, if ν is the common distribution of the X_n, then

$$\mathbb{E}(\psi_{\epsilon s_n}(X_n))=\int_{\{x:|x|>\epsilon\sigma\sqrt{n}\}}x^2\nu(dx)\to 0$$

by Lebesgue's Dominated Convergence Theorem; so that

$$\frac{1}{s_n^2}\sum_{i=0}^n\mathbb{E}(\psi_{\epsilon s_n}(X_n))\to 0$$

by 273Ca. Thus Lindeberg's conditions are satisfied and 274G gives the result.

274J Corollary Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables with zero expectation, and suppose that $\{X_n^2:n\in\mathbb{N}\}$ is uniformly integrable and that

$$\liminf_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^n\mathrm{Var}(X_i)>0.$$

Set

$$s_n=\sqrt{\sum_{i=0}^n\mathrm{Var}(X_i)},\qquad S_n=\frac{1}{s_n}(X_0+\ldots+X_n)$$

for large $n\in\mathbb{N}$. Then

$$\lim_{n\to\infty}\Pr(S_n\le a)=\Phi(a)$$

uniformly for $a\in\mathbb{R}$.

proof The condition

$$\liminf_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^n\mathrm{Var}(X_i)>0$$

means that there are $c>0$, $n_0\in\mathbb{N}$ such that $s_n\ge c\sqrt{n+1}$ for every $n\ge n_0$. Let the underlying space be (Ω,Σ,μ), and take ϵ, $\eta>0$. Writing $\psi_\delta(x)=0$ for $|x|\le\delta$, x^2 for $|x|>\delta$, as in 274F-274G, we have

$$\mathbb{E}(\psi_{\epsilon s_n}(X_i)) \le \mathbb{E}(\psi_{c\epsilon\sqrt{n+1}}(X_i)) = \int_{F(i,c\epsilon\sqrt{n+1})} X_i^2 d\mu$$

for $n \ge n_0$, $i \le n$, where $F(i,\gamma) = \{\omega : \omega \in \operatorname{dom} X_i,\ |X_i(\omega)| > \gamma\}$. Because $\{X_i^2 : i \in \mathbb{N}\}$ is uniformly integrable, there is a $\gamma \ge 0$ such that $\int_{F(i,\gamma)} X_i^2 d\mu \le \eta c^2$ for every $i \in \mathbb{N}$ (246I). Let $n_1 \ge n_0$ be such that $c\epsilon\sqrt{n_1+1} \ge \gamma$; then for any $n \ge n_1$

$$\frac{1}{s_n^2}\sum_{i=0}^n \mathbb{E}(\psi_{\epsilon s_n}(X_i)) \le \frac{1}{c^2(n+1)}\sum_{i=0}^n \eta c^2 = \eta.$$

As ϵ, η are arbitrary, the conditions of 274G are satisfied and the result follows.

274K Corollary Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables with zero expectation, and suppose that

(i) there is some $\delta > 0$ such that $\sup_{n\in\mathbb{N}} \mathbb{E}(|X_n|^{2+\delta}) < \infty$,

(ii) $\liminf_{n\to\infty} \frac{1}{n+1}\sum_{i=0}^n \operatorname{Var}(X_i) > 0$.

Set $s_n = \sqrt{\sum_{i=0}^n \operatorname{Var}(X_i)}$ and

$$S_n = \frac{1}{s_n}(X_0 + \ldots + X_n)$$

for large $n \in \mathbb{N}$. Then

$$\lim_{n\to\infty} \Pr(S_n \le a) = \Phi(a)$$

uniformly for $a \in \mathbb{R}$.

proof The point is that $\{X_n^2 : n \in \mathbb{N}\}$ is uniformly integrable. **P** Set $K = 1 + \sup_{n\in\mathbb{N}} \mathbb{E}(|X_n|^{2+\delta})$. Given $\epsilon > 0$, set $M = (K/\epsilon)^{1/\delta}$. Then $(X_n^2 - M)^+ \le M^{-\delta}|X_n|^{2+\delta}$, so

$$\mathbb{E}(X_n^2 - M)^+ \le KM^{-\delta} = \epsilon$$

for every $n \in \mathbb{N}$. As ϵ is arbitrary, $\{X_n^2 : n \in \mathbb{N}\}$ is uniformly integrable. **Q**

Accordingly the conditions of 274J are satisfied and we have the result.

274L Remarks (a) All the theorems of this section are devoted to finding conditions under which a random variable S is 'nearly' standard normal, in the sense that $\Pr(S \le a) \doteq \Pr(Z \le a)$ uniformly for $a \in \mathbb{R}$, where Z is some (or any) standard normal random variable. In all cases the random variable S is normalized to have expectation 0 and variance 1, and is a sum of a large number of independent random variables. (In 274G and 274I-274K it is explicit that there must be many X_i, since they refer to a limit as $n \to \infty$. This is not said in so many words in the formulation I give of Lindeberg's theorem, but the proof makes it evident that $n\sqrt{\delta + \delta^2} \ge 1$, so surely n will have to be large there also.)

(b) I cannot leave this section without remarking that the form of the definition of 'nearly standard normal' may lead your intuition astray if you try to apply it to other distributions. If we take F to be the distribution function of S, so that $F(a) = \Pr(S \le a)$, I am saying that S is 'nearly standard normal' if $\sup_{a\in\mathbb{R}} |F(a) - \Phi(a)|$ is small. It is natural to think of this as approximation in a metric, writing

$$\tilde\rho(\nu, \nu') = \sup_{a\in\mathbb{R}} |F_\nu(a) - F_{\nu'}(a)|$$

for distributions ν, ν' on $\mathbb{R}$, where $F_\nu(a) = \nu\,]-\infty, a]$. In this form, the theorems above can be read as finding conditions under which $\lim_{n\to\infty} \tilde\rho(\nu_{S_n}, \mu_G) = 0$. But the point is that $\tilde\rho$ is not really the right metric to use. It works here because μ_G is atomless. But suppose, for instance, that ν is the distribution which gives mass 1 to the point 0 (I mean, that $\nu E = 1$ if $0 \in E \subseteq \mathbb{R}$, 0 if $0 \notin E \subseteq \mathbb{R}$), and that ν_n is the distribution of a normal random variable with expectation 0 and variance $\frac{1}{n}$, for each $n \ge 1$. Then $F_\nu(0) = 1$ and $F_{\nu_n}(0) = \frac{1}{2}$, so $\tilde\rho(\nu_n, \nu) = \frac{1}{2}$ for each $n \ge 1$. However, for most purposes one would regard the difference between ν_n and ν as small, and surely ν is the only distribution which one could reasonably call a limit of the ν_n.

(c) The difficulties here present themselves in more than one form. A statistician would be unhappy with the idea that the ν_n of the last paragraph were far from ν (and from each other), on the grounds that any measurement involving random variables with these distributions must be subject to error, and small errors of measurement will render them indistinguishable. A pure mathematician, looking forward to the possibility of generalizing these results, will be unhappy with the emphasis given to the values of $\nu\,]-\infty, a]$, for which it may be difficult to find suitable equivalents in more abstract spaces.

(d) These considerations join together to lead us to a rather different definition for a topology on the space P of probability distributions on $\mathbb{R}$. For any bounded continuous function $h:\mathbb{R}\to\mathbb{R}$ we have a pseudometric $\rho_h : P\times P\to [0,\infty[$ defined by writing

$$\rho_h(\nu,\nu') = |\int h\,d\nu - \int h\,d\nu'|$$

for all ν, $\nu'\in P$. The **vague topology** on P is that generated by the pseudometrics ρ_h (2A3F). I will not go into its properties in detail here (some are sketched in 274Ya-274Yd below; see also 285K-285L, 285S and 437J-437P in Volume 4). But I maintain that the right way to look at the results of this chapter is to say that (i) the distributions ν_S are close to μ_G for the vague topology (ii) the sets $\{\nu:\tilde\rho(\nu,\mu_G)<\epsilon\}$ are open for that topology, and that is why $\tilde\rho(\nu_S,\mu_G)$ is small.

***274M** I conclude with a simple pair of inequalities which are frequently useful when studying normal random variables.

Lemma (a) $\int_x^\infty e^{-t^2/2}dt \le \frac{1}{x}e^{-x^2/2}$ for every $x>0$.

(b) $\int_x^\infty e^{-t^2/2}dt \ge \frac{1}{2x}e^{-x^2/2}$ for every $x\ge 1$.

proof (a)

$$\int_x^\infty e^{-t^2/2}dt = \int_0^\infty e^{-(x+s)^2/2}ds \ge e^{-x^2/2}\int_0^\infty e^{-xs}ds = \frac{1}{x}e^{-x^2/2}.$$

(b) Set

$$f(t) = e^{-t^2/2} - (1-x(t-x))e^{-x^2/2}.$$

Then $f(x)=f'(x)=0$ and $f''(t)=(t^2-1)e^{-t^2/2}$ is positive for $t\ge x$ (because $x\ge 1$). Accordingly $f(t)\ge 0$ for every $t\ge x$, and $\int_x^{x+1/x} f(t)dt\ge 0$. But this means just that

$$\int_x^\infty e^{-t^2/2}dt \ge \int_x^{x+\frac{1}{x}} e^{-t^2/2}dt \ge \int_x^{x+\frac{1}{x}} (1-x(t-x))e^{-x^2/2}dt = \frac{1}{2x}e^{-x^2/2},$$

as required.

274X Basic exercises >(a) Use 272U to give an alternative proof of 274B.

(b) Prove 274D when h'' is M_3-Lipschitz but not necessarily differentiable.

(c) Let $\langle m_k\rangle_{k\in\mathbb{N}}$ be a strictly increasing sequence in $\mathbb{N}$ such that $m_0=0$ and $\lim_{k\to\infty} m_k/m_{k+1}=0$. Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of random variables such that $\Pr(X_n=\sqrt{m_k})=\Pr(X_n=-\sqrt{m_k})=1/2m_k$, $\Pr(X_n=0)=1-1/m_k$ whenever $m_{k-1}\le n<m_k$. Show that the Central Limit Theorem is not valid for $\langle X_n\rangle_{n\in\mathbb{N}}$. (*Hint*: setting $W_k=(X_0+\ldots+X_{m_k-1})/\sqrt{m_k}$, show that $\Pr(W_k\in G)\to 1$ for every open set G including $\mathbb{Z}$.)

(d) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be any independent sequence of random variables all with the same distribution; suppose that they all have finite variance $\sigma^2>0$, and that their common expectation is c. Set $S_n=\frac{1}{\sqrt{n+1}}(X_0+\ldots+X_n)$ for each n, and let Y be a normal random variable with expectation c and variance σ^2. Show that $\lim_{n\to\infty}\Pr(S_n\le a)=\Pr(Y\le a)$ uniformly for $a\in\mathbb{R}$.

>(e) Show that for any $a\in\mathbb{R}$,

$$\lim_{n\to\infty}\frac{1}{2^n}\sum_{r=0}^{\lfloor\frac{n}{2}+a\frac{\sqrt{n}}{2}\rfloor}\frac{n!}{r!(n-r)!} = \lim_{n\to\infty}\frac{1}{2^n}\#(\{I:I\subseteq n,\ \#(I)\le\frac{n}{2}+a\frac{\sqrt{n}}{2}\}) = \Phi(a).$$

(f) Show that 274I is a special case of 274J.

(g) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables with zero expectation. Set $s_n=\sqrt{\sum_{i=0}^n \mathrm{Var}(X_i)}$ and

$$S_n = \frac{1}{s_n}(X_0 + \ldots + X_n)$$

for each $n \in \mathbb{N}$. Suppose that there is some $\delta > 0$ such that

$$\lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{i=0}^{n} \mathbb{E}(|X_i|^{2+\delta}) = 0.$$

Show that $\lim_{n\to\infty} \Pr(S_n \le a) = \Phi(a)$ uniformly for $a \in \mathbb{R}$. (This is a form of **Liapounoff's central limit theorem**; see LIAPOUNOFF 1901.)

(h) Let P be the set of Radon probability measures on $\mathbb{R}$. Let $\nu_0 \in P$, $a \in \mathbb{R}$. Show that the map $\nu \mapsto \nu\,]-\infty, a] : P \to [0,1]$ is continuous at ν_0 for the vague topology on P iff $\nu_0\{a\} = 0$.

(i) Suppose that $f : \mathbb{R} \to \mathbb{R}$ is absolutely continuous on every closed bounded interval, and that $\int_{-\infty}^{\infty} |f'(x)| e^{-ax^2} dx < \infty$ for every $a > 0$. Let X be a normal random variable with zero expectation. Show that $\mathbb{E}(Xf(X))$ and $\mathbb{E}(X^2)\mathbb{E}(f'(X))$ are defined and equal.

(j) (STEELE 86) Suppose that $X_0, \ldots, X_n, Y_0, \ldots, Y_n$ are independent random variables such that, for each $i \le n$, X_i and Y_i have the same distribution. Let $h : \mathbb{R}^{n+1} \to \mathbb{R}$ be a Borel measurable function, and set $Z = h(X_0, \ldots, X_n)$, $Z_i = h(X_0, \ldots, X_{i-1}, Y_i, X_{i+1}, \ldots, X_n)$ for each i (with $Z_0 = h(Y_0, X_1, \ldots, X_n)$ and $Z_n = h(X_0, \ldots, X_{n-1}, Y_n)$, of course). Suppose that Z has finite expectation. Show that $\operatorname{Var}(Z) \le \frac{1}{2} \sum_{i=0}^{n} \mathbb{E}(Z_i - Z)^2$.

(k) Let $\langle X_n \rangle_{n\in\mathbb{N}}$ be an independent identically distributed sequence of random variables with non-zero finite variance. Let $\langle t_n \rangle_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ such that $\sum_{n=0}^{\infty} t_n^2 = \infty$. Show that $\sum_{n=0}^{\infty} t_n X_n$ is undefined or infinite a.e. (*Hint*: First deal with the case in which $\langle t_n \rangle_{n\in\mathbb{N}}$ does not converge to 0. Otherwise, use 274G to show that, for any $n \in \mathbb{N}$, $\lim_{m\to\infty} \Pr(|\sum_{i=n}^{m} t_i X_i| \ge 1) \ge \frac{1}{2}$).)

(l) Let $\langle X_n \rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables with zero expectation. Suppose that $M \ge 0$ is such that $|X_n| \le M$ a.e. for every n, and that $\sum_{n=0}^{\infty} \operatorname{Var}(X_n) = \infty$. Set $s_n = \sqrt{\sum_{i=0}^{n} \operatorname{Var}(X_i)}$ for each n, and $S_n = \frac{1}{s_n} \sum_{i=0}^{n} X_i$ when $s_n > 0$. Show that $\lim_{n\to\infty} \Pr(S_n \le a) = \Phi(a)$ for every $a \in \mathbb{R}$.

274Y Further exercises (a) Write P for the set of Radon probability measures on $\mathbb{R}$. For ν, $\nu' \in P$ set

$$\rho(\nu, \nu') = \inf\{\epsilon : \epsilon \ge 0,\ \nu\,]-\infty, a-\epsilon] - \epsilon \le \nu'\,]-\infty, a] \le \nu\,]-\infty, a+\epsilon] + \epsilon \text{ for every } a \in \mathbb{R}\}.$$

Show that ρ is a metric on P and that it defines the vague topology on P. (ρ is called **Lévy's metric**.)

(b) Write P for the set of Radon probability measures on $\mathbb{R}$, and let $\tilde{\rho}$ be the metric on P defined in 274Lb. Show that if $\nu \in P$ is atomless and $\epsilon > 0$, then $\{\nu' : \nu' \in P, \tilde{\rho}(\nu', \nu) < \epsilon\}$ is open for the vague topology on $\mathbb{R}$.

(c) Let $\langle S_n \rangle_{n\in\mathbb{N}}$ be a sequence of real-valued random variables, and Z a standard normal random variable. Show that the following are equiveridical:

(i) $\mu_G = \lim_{n\to\infty} \nu_{S_n}$ for the vague topology, writing ν_{S_n} for the distribution of S_n;

(ii) $\mathbb{E}(h(Z)) = \lim_{n\to\infty} \mathbb{E}(h(S_n))$ for every bounded continuous function $h : \mathbb{R} \to \mathbb{R}$;

(iii) $\mathbb{E}(h(Z)) = \lim_{n\to\infty} \mathbb{E}(h(S_n))$ for every bounded function $h : \mathbb{R} \to \mathbb{R}$ such that (α) h has continuous derivatives of all orders (β) $\{x : h(x) \ne 0\}$ is bounded;

(iv) $\lim_{n\to\infty} \Pr(S_n \le a) = \Phi(a)$ for every $a \in \mathbb{R}$;

(v) $\lim_{n\to\infty} \Pr(S_n \le a) = \Phi(a)$ uniformly for $a \in \mathbb{R}$;

(vi) $\{a : \lim_{n\to\infty} \Pr(S_n \le a) = \Phi(a)\}$ is dense in $\mathbb{R}$.

(See also 285L.)

(d) Let (Ω, Σ, μ) be a probability space, and P the set of Radon probability measures on $\mathbb{R}$. Show that $X \mapsto \nu_X : \mathcal{L}^0(\mu) \to P$ is continuous for the topology of convergence in measure on $\mathcal{L}^0(\mu)$ and the vague topology on P.

(e) Let $\langle X_n \rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables. Suppose that there is an $M \ge 0$ such that $|X_n| \le M$ a.e. for every $n \in \mathbb{N}$, and that $\sum_{n=0}^{\infty} X_n$ is defined, as a real number, almost everywhere. Show that $\sum_{n=0}^{\infty} \operatorname{Var}(X_n) < \infty$.

274 Notes and comments For more than two hundred years the Central Limit Theorem has been one of the glories of mathematics, and no branch of mathematics or science would be the same without it. I suppose it is the most important single theorem of probability theory; and I observe that the proof hardly uses measure theory. To be sure, I have clothed the arguments above in the language of measure and integration. But if you look at their essence, the vital elements of the proof are

(i) a linear combination of independent normal random variables is normal (274Ae, 274B);

(ii) if U, V, W are independent random variables, and h is a bounded continuous function, then $|\mathbb{E}(h(U,V,W))| \le \sup_{t\in\mathbb{R}} |\mathbb{E}(h(U,V,t))|$ (274C);

(iii) if $(X_0,\dots,X_n)$ are independent random variables, then we can find independent random variables $(X_0',\dots,X_n',Z_0,\dots,Z_n)$ such that Z_j is standard normal and X_j' has the same distribution as X_j, for each j (274F).

The rest of the argument consists of elementary calculus, careful estimations and a few of the most fundamental properties of expectations and independence. Now (ii) and (iii) are justified above by appeals to Fubini's theorem, but surely they belong to the list of probabilistic intuitions which take priority over the identification of probabilities with countably additive functionals. If they had given any insuperable difficulty it would have been a telling argument against the model of probability we were using, but would not have affected the Central Limit Theorem. In fact (i) seems to be the place where we really need a mathematical model of the concept of 'distribution', and all the relevant calculations can be done in terms of the Riemann integral on the plane, with no mention of countable additivity. So while I am happy and proud to have written out a version of these beautiful ideas, I have to admit that they are in no essential way dependent on the rest of this treatise.

In §285 I will describe a quite different approach to the theorem, using much more sophisticated machinery; but it will again be the case, perhaps more thoroughly hidden, that the relevance of measure theory will not be to the theorem itself, but to our imagination of what an arbitrary distribution is. For here I do have a claim to make for my subject. The characterization of distribution functions as arbitrary monotonic functions, continuous on the right, and with the right limits at $\pm\infty$ (271Xb), together with the analysis of monotonic functions in §226, gives us a chance of forming a mental picture of the proper class of objects to which such results as the Central Limit Theorem can be applied.

Theorem 274F is a trifling modification of Theorem 3 of Lindeberg 22. Like the original, it emphasizes what I believe to be vital to all the limit theorems of this chapter: they are best founded on a proper understanding of finite sequences of random variables. Lindeberg's condition was the culmination of a long search for the most general conditions under which the Central Limit Theorem would be valid. I offer a version of Laplace's theorem (274Xe) as the starting place, and Liapounoff's condition (274Xg) as an example of one of the intermediate stages. Naturally the corollaries 274I, 274J, 274K and 274Xd are those one seeks to apply by choice. There is an intriguing, but as far as I know purely coincidental, parallel between 273H/274K and 273I/274Xd. As an example of an independent sequence $\langle X_n\rangle_{n\in\mathbb{N}}$ of random variables, all with expectation zero and variance 1, to which the Central Limit Theorem does *not* apply, I offer 274Xc.

275 Martingales

This chapter so far has been dominated by independent sequences of random variables. I now turn to another of the remarkable concepts to which probabilistic intuitions have led us. Here we study evolving systems, in which we gain progressively more information as time progresses. I give the basic theorems on pointwise convergence of martingales (275F-275H, 275K) and a very brief account of 'stopping times' (275L-275P).

275A Definition Let (Ω,Σ,μ) be a probability space with completion $(\Omega,\hat\Sigma,\hat\mu)$, and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of $\hat\Sigma$. (Such sequences $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ are called **filtrations**.) A **martingale adapted to** $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ is a sequence $\langle X_n\rangle_{n\in\mathbb{N}}$ of integrable real-valued random variables on Ω such that (i) $\operatorname{dom} X_n \in \Sigma_n$ and X_n is Σ_n-measurable for each $n\in\mathbb{N}$ (ii) whenever $m\le n\in\mathbb{N}$ and $E\in\Sigma_m$ then $\int_E X_n = \int_E X_m$.

Note that for (ii) it is enough if $\int_E X_{n+1} = \int_E X_n$ whenever $n\in\mathbb{N}$ and $E\in\Sigma_n$.

275B Examples We have seen many contexts in which such sequences appear naturally; here are a few.

(a) Let (Ω,Σ,μ) be a probability space and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of Σ. Let X be any real-valued random variable on Ω with finite expectation, and for each $n\in\mathbb{N}$ let X_n be a conditional expectation of X on Σ_n, as in §233. Subject to the conditions that $\operatorname{dom} X_n\in\Sigma_n$ and X_n is actually Σ_n-measurable for each n (a purely technical point – see 232He), $\langle X_n\rangle_{n\in\mathbb{N}}$ will be a martingale adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$, because $\int_E X_{n+1} = \int_E X = \int_E X_n$ whenever $E\in\Sigma_n$.

(b) Let (Ω,Σ,μ) be a probability space and $\langle X_n\rangle_{n\in\mathbb{N}}$ an independent sequence of random variables all with zero expectation. For each $n\in\mathbb{N}$ let $\tilde\Sigma_n$ be the σ-algebra generated by $\bigcup_{i\le n}\Sigma_{X_i}$, writing Σ_{X_i} for the σ-algebra defined by X_i (272C), and set $S_n=X_0+\ldots+X_n$. Then $\langle S_n\rangle_{n\in\mathbb{N}}$ is a martingale adapted to $\tilde\Sigma_n$. (Use 272K to see that $\Sigma_{X_{n+1}}$ is independent of $\tilde\Sigma_n$, so that $\int_E X_{n+1}=\int X_{n+1}\times\chi E=0$ for every $E\in\tilde\Sigma_n$, by 272R.)

(c) Let (Ω,Σ,μ) be a probability space and $\langle X_n\rangle_{n\in\mathbb{N}}$ an independent sequence of random variables all with expectation 1. For each $n\in\mathbb{N}$ let $\tilde\Sigma_n$ be the σ-algebra generated by $\bigcup_{i\le n}\Sigma_{X_i}$, writing Σ_{X_i} for the σ-algebra defined by X_i, and set $W_n=X_0\times\ldots\times X_n$. Then $\langle W_n\rangle_{n\in\mathbb{N}}$ is a martingale adapted to $\langle\tilde\Sigma_n\rangle_{n\in\mathbb{N}}$.

275C Remarks (a) It seems appropriate to the concept of a random variable X being 'adapted' to a σ-algebra Σ to require that $\operatorname{dom}X\in\Sigma$ and that X should be Σ-measurable, even though this may mean that other random variables, equal almost everywhere to X, may fail to be 'adapted' to Σ.

(b) Technical problems of this kind evaporate, of course, if all μ-negligible subsets of X belong to Σ_0. But examples such as 275Bb make it seem unreasonable to insist on such a simplification as a general rule.

(c) The concept of 'martingale' can readily be extended to other index sets than $\mathbb{N}$; indeed, if I is any partially ordered set, we can say that $\langle X_i\rangle_{i\in I}$ is a martingale on (Ω,Σ,μ) adapted to $\langle\Sigma_i\rangle_{i\in I}$ if (i) each Σ_i is a σ-subalgebra of $\hat\Sigma$ (ii) each X_i is an integrable real-valued Σ_i-measurable random variable such that $\operatorname{dom}X_i\in\Sigma_i$ (iii) whenever $i\le j$ in I, then $\Sigma_i\subseteq\Sigma_j$ and $\int_E X_i=\int_E X_j$ for every $E\in\Sigma_i$. The principal case, after $I=\mathbb{N}$, is $I=[0,\infty[$; $I=\mathbb{Z}$ also is interesting, and I think it is fair to say that the most important ideas can already be expressed in theorems about martingales indexed by finite sets I. But in this volume I will generally take martingales to be indexed by $\mathbb{N}$.

(d) Given just a sequence $\langle X_n\rangle_{n\in\mathbb{N}}$ of integrable real-valued random variables on a probability space (Ω,Σ,μ), we can say simply that $\langle X_n\rangle_{n\in\mathbb{N}}$ is a **martingale** on (Ω,Σ,μ) if there is some non-decreasing sequence $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ of σ-subalgebras of $\hat\Sigma$ (the completion of Σ) such that $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$. If we write $\tilde\Sigma_n$ for the σ-algebra generated by $\bigcup_{i\le n}\Sigma_{X_i}$, where Σ_{X_i} is the σ-algebra defined by X_i, as in 275Bb, then it is easy to see that $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale iff it is a martingale adapted to $\langle\tilde\Sigma_n\rangle_{n\in\mathbb{N}}$.

(e) Continuing from (d), it is also easy to see that if $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale on (Ω,Σ,μ), and $X'_n=_{\text{a.e.}}X_n$ for every n, then $\langle X'_n\rangle_{n\in\mathbb{N}}$ is a martingale on (Ω,Σ,μ). (The point is that if $\langle X_n\rangle_{n\in\mathbb{N}}$ is adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$, then both $\langle X_n\rangle_{n\in\mathbb{N}}$ and $\langle X'_n\rangle_{n\in\mathbb{N}}$ are adapted to $\langle\hat\Sigma_n\rangle_{n\in\mathbb{N}}$, where

$$\hat\Sigma_n=\{E\triangle F:E\in\Sigma_n,\ F\text{ is negligible}\}.)$$

Consequently we have a concept of 'martingale' as a sequence in $L^1(\mu)$, saying that a sequence $\langle X_n^\bullet\rangle_{n\in\mathbb{N}}$ in $L^1(\mu)$ is a martingale iff $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale.

Nevertheless, I think that the concept of 'martingale adapted to a sequence of σ-algebras' is the primary one, since in all the principal applications the σ-algebras reflect some essential aspect of the problem, which may not be fully encompassed by the random variables alone.

(f) The word 'martingale' originally (in English; the history in French is more complex) referred to a strap used to prevent a horse from throwing its head back. Later it was used as the name of a gambling system in which the gambler doubles his stake each time he loses, and (in French) as a general term for gambling systems. These may be regarded as a class of 'stopped-time martingales', as described in 275L-275P below.

275D A large part of the theory of martingales consists of inequalities of various kinds. I give two of the most important, both due to J.L.Doob. (See also 276Xa-276Xb.)

Lemma Let (Ω,Σ,μ) be a probability space, and $\langle X_n\rangle_{n\in\mathbb{N}}$ a martingale on Ω. Fix $n\in\mathbb{N}$ and set $X^*=\max(X_0,\dots,X_n)$. Then for any $\epsilon>0$,

$$\Pr(X^*\ge\epsilon)\le\frac{1}{\epsilon}\mathbb{E}(X_n^+),$$

writing $X_n^+=\max(0,X_n)$.

proof Write $\hat\mu$ for the completion of μ, and $\hat\Sigma$ for its domain. Let $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ be a non-decreasing sequence of σ-subalgebras of $\hat\Sigma$ to which $\langle X_n\rangle_{n\in\mathbb{N}}$ is adapted. For each $i\le n$ set

$$E_i=\{\omega:\omega\in\operatorname{dom}X_i,\ X_i(\omega)\ge\epsilon\},$$

$$F_i = E_i \setminus \bigcup_{j<i} E_j.$$

Then $F_0, \dots, F_n$ are disjoint and $F = \bigcup_{i\le n} F_i = \bigcup_{i\le n} E_i$; moreover, writing H for the conegligible set $\bigcap_{i\le n} \operatorname{dom} X_i$,

$$\{\omega : X^*(\omega) \ge \epsilon\} = F \cap H,$$

so that

$$\Pr(X^* \ge \epsilon) = \hat\mu\{\omega : X^*(\omega) \ge \epsilon\} = \hat\mu F = \sum_{i=0}^n \hat\mu F_i.$$

On the other hand, E_i and F_i belong to Σ_i for each $i \le n$, so

$$\int_{F_i} X_n = \int_{F_i} X_i \ge \epsilon\hat\mu F_i$$

for every i, and

$$\epsilon\hat\mu F = \epsilon\sum_{i=0}^n \hat\mu F_i \le \sum_{i=0}^n \int_{F_i} X_n = \int_F X_n \le \int_F X_n^+ \le \mathbb{E}(X_n^+),$$

as required.

Remark Note that in fact we have $\epsilon\hat\mu F \le \int_F X_n$, where $F = \{\omega : X^*(\omega) \ge \epsilon\}$; this is of great importance in many applications.

275E Up-crossings The next lemma depends on the notion of 'up-crossing'. Let $x_0, \dots, x_n$ be any list of real numbers, and $a < b$ in $\mathbb{R}$. The **number of up-crossings from** a **to** b in the list $x_0, \dots, x_n$ is the number of pairs (j, k) such that $0 \le j < k \le n$, $x_j \le a$, $x_k \ge b$ and $a < x_i < b$ for $j < i < k$. Note that this is also the largest m such that $s_m < \infty$, if we write

$$r_1 = \inf\{i : i \le n,\ x_i \le a\},$$

$$s_1 = \inf\{i : r_1 < i \le n,\ x_i \ge b\},$$

$$r_2 = \inf\{i : s_1 < i \le n,\ x_i \le a\},$$

$$s_2 = \inf\{i : r_2 < i \le n,\ x_i \ge b\}$$

and so on, taking $\inf\emptyset = \infty$.

275F Lemma Let (Ω, Σ, μ) be a probability space and $\langle X_n\rangle_{n\in\mathbb{N}}$ a martingale on Ω. Suppose that $n \in \mathbb{N}$ and that $a < b$ in $\mathbb{R}$. For each $\omega \in \bigcap_{i\le n} \operatorname{dom} X_i$, let $U(\omega)$ be the number of up-crossings from a to b in the list $X_0(\omega), \dots, X_n(\omega)$. Then

$$\mathbb{E}(U) \le \frac{1}{b-a}\mathbb{E}((X_n - X_0)^+),$$

writing $(X_n - X_0)^+(\omega) = \max(0, X_n(\omega) - X_0(\omega))$ for $\omega \in \operatorname{dom} X_n \cap \operatorname{dom} X_0$.

proof Each individual step in the proof is 'elementary', but the structure as a whole is non-trivial.

(a) The following fact will be useful. Suppose that $x_0, \dots, x_n$ are real numbers; let u be the number of up-crossings from a to b in the list $x_0, \dots, x_n$. Set $y_i = \max(x_i, a)$ for each i; then u is also the number of up-crossings from a to b in the list $y_0, \dots, y_n$. For each $k \le n$, set $c_k = 1$ if there is a $j \le k$ such that $x_j \le a$ and $x_i < b$ for $j \le i \le k$, 0 otherwise. Then

$$(b-a)u \le \sum_{k=0}^{n-1} c_k(y_{k+1} - y_k).$$

P I induce on m to show that (defining r_m, s_m as in 275E)

$$(b-a)m \le \sum_{k=0}^{s_m-1} c_k(y_{k+1} - y_k)$$

whenever $m \le u$. For $m = 0$ (taking $s_0 = -1$) we have $0 = 0$. For the inductive step to $m \ge 1$, we have $s_{m-1} < r_m < s_m \le n$ (because I am supposing that $m \le u$), and $c_k = 0$ if $s_{m-1} \le k < r_m$, $c_k = 1$ if $r_m \le k < s_m$. So

$$\begin{aligned}\sum_{k=0}^{s_m-1} c_k(y_{k+1} - y_k) &= \sum_{k=0}^{s_{m-1}-1} c_k(y_{k+1} - y_k) + \sum_{k=r_m}^{s_m-1} (y_{k+1} - y_k)\\ &\ge (b-a)(m-1) + y_{s_m} - y_{r_m}\end{aligned}$$

(by the inductive hypothesis)

$$\geq (b-a)m$$

(because $y_{s_m} \geq b$, $y_{r_m} = a$), and the induction proceeds.

Accordingly

$$\sum_{k=0}^{s_u-1} c_k(y_{k+1} - y_k) \geq (b-a)u.$$

As for the sum $\sum_{k=s_u}^{n-1} c_k(y_{k+1} - y_k)$, we have $c_k = 0$ for $s_u \leq k < r_{u+1}$, $c_k = 1$ for $r_u \leq k < s_{u+1}$, while $s_{u+1} > n$, so if $n \leq r_{u+1}$ we have

$$\sum_{k=0}^{n-1} c_k(y_{k+1} - y_k) = \sum_{k=0}^{s_u-1} c_k(y_{k+1} - y_k) \geq (b-a)u,$$

while if $n > r_{u+1}$ we have

$$\begin{aligned}\sum_{k=0}^{n-1} c_k(y_{k+1} - y_k) &= \sum_{k=0}^{s_u-1} c_k(y_{k+1} - y_k) + \sum_{k=r_{u+1}}^{n-1} y_{k+1} - y_k \\ &\geq (b-a)u + y_n - y_{r_{u+1}} \\ &\geq (b-a)u\end{aligned}$$

because $y_n \geq a = y_{r_{u+1}}$. Thus in both cases we have the required result. **Q**

(b)(i) Now define

$$Y_k(\omega) = \max(a, X_k(\omega)) \text{ for } \omega \in \operatorname{dom} X_k,$$

$$F_k = \{\omega : \omega \in \bigcap_{i \leq k} \operatorname{dom} X_i,\ \exists\, j \leq k,\ X_j(\omega) \leq a,\ X_i(\omega) < b \text{ if } j \leq i \leq k\}$$

for each $k \in \mathbb{N}$. If $\langle \Sigma_n \rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence of σ-algebras to which $\langle X_n \rangle_{n\in\mathbb{N}}$ is adapted, then $F_k \in \Sigma_k$ (because if $j \leq k$ all the sets $\operatorname{dom} X_j$, $\{\omega : X_j(\omega) \leq a\}$, $\{\omega : X_j(\omega) < b\}$ belong to $\Sigma_j \subseteq \Sigma_k$).

(ii) We find that $\int_F Y_k \leq \int_F Y_{k+1}$ if $F \in \Sigma_k$. **P** Set $G = \{\omega : X_k(\omega) > a\} \in \Sigma_k$. Then

$$\begin{aligned}\int_F Y_k &= \int_{F\cap G} X_k + a\hat{\mu}(F \setminus G) \\ &= \int_{F\cap G} X_{k+1} + a\hat{\mu}(F \setminus G) \\ &\leq \int_{F\cap G} Y_{k+1} + \int_{F\setminus G} Y_{k+1} = \int_F Y_{k+1}. \ \mathbf{Q}\end{aligned}$$

(iii) Consequently $\int_F Y_{k+1} - Y_k \leq \int Y_{k+1} - Y_k$ for every $F \in \Sigma_k$.

$$\mathbf{P}\ \int (Y_{k+1} - Y_k) - \int_F (Y_{k+1} - Y_k) = \int_{\Omega\setminus F} Y_{k+1} - \int_{\Omega\setminus F} Y_k \geq 0.\ \mathbf{Q}$$

(c) Let H be the conegligible set $\operatorname{dom} U = \bigcap_{i\leq n} \operatorname{dom} X_i \in \Sigma_n$. We ought to check at some point that U is Σ_n-measurable; but this is clearly true, because all the relevant sets $\{\omega : X_i(\omega) \leq a\}$, $\{\omega : X_i(\omega) \geq b\}$ belong to Σ_n. For each $\omega \in H$, apply (a) to the list $X_0(\omega), \dots, X_n(\omega)$ to see that

$$(b-a)U(\omega) \leq \sum_{k=0}^{n-1} \chi F_k(\omega)(Y_{k+1}(\omega) - Y_k(\omega)).$$

Because H is conegligible, it follows that

$$(b-a)\mathbb{E}(U) \leq \sum_{k=0}^{n-1} \int_{F_k} Y_{k+1} - Y_k \leq \sum_{k=0}^{n-1} \int Y_{k+1} - Y_k$$

(using (b-iii))

$$= \mathbb{E}(Y_n - Y_0) \leq \mathbb{E}((X_n - X_0)^+)$$

because $Y_n - Y_0 \leq (X_n - X_0)^+$ everywhere on $\operatorname{dom} X_n \cap \operatorname{dom} X_0$. This completes the proof.

275G We are now ready for the principal theorems of this section.

Doob's Martingale Convergence Theorem Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale on a probability space (Ω,Σ,μ), and suppose that $\sup_{n\in\mathbb{N}}\mathbb{E}(|X_n|)<\infty$. Then $\lim_{n\to\infty}X_n(\omega)$ is defined in $\mathbb{R}$ for almost every ω in Ω.

proof (a) Set $H=\bigcap_{n\in\mathbb{N}}\operatorname{dom}X_n$, and for $\omega\in H$ set $Y(\omega)=\liminf_{n\to\infty}X_n(\omega)$, $\quad Z(\omega)=\limsup_{n\to\infty}X_n(\omega)$, allowing $\pm\infty$ in both cases. But note that $Y\le\liminf_{n\to\infty}|X_n|$, so by Fatou's Lemma $Y(\omega)<\infty$ for almost every ω; similarly $Z(\omega)>-\infty$ for almost every ω. It will therefore be enough if I can show that $Y=_{\text{a.e.}}Z$, for then $Y(\omega)=Z(\omega)\in\mathbb{R}$ for almost every ω, and $\langle X_n(\omega)\rangle_{n\in\mathbb{N}}$ will be convergent for almost every ω.

(b) ? So suppose, if possible, that Y and Z are not equal almost everywhere. Of course both are $\hat\Sigma$-measurable, where $(\Omega,\hat\Sigma,\hat\mu)$ is the completion of (Ω,Σ,μ), so we must have

$$\hat\mu\{\omega:\omega\in H,\,Y(\omega)<Z(\omega)\}>0.$$

Accordingly there are rational numbers q, q' such that $q<q'$ and $\hat\mu G>0$, where

$$G=\{\omega:\omega\in H,\,Y(\omega)<q<q'<Z(\omega)\}.$$

Now, for each $\omega\in H$, $n\in\mathbb{N}$, let $U_n(\omega)$ be the number of up-crossings from q to q' in the list $X_0(\omega),\dots,X_n(\omega)$. Then 275F tells us that

$$\mathbb{E}(U_n)\le\frac{1}{q'-q}\mathbb{E}((X_n-X_0)^+)\le\frac{1}{q'-q}\mathbb{E}(|X_n|+|X_0|)\le\frac{2M}{q'-q},$$

if we write $M=\sup_{i\in\mathbb{N}}\mathbb{E}(|X_i|)$. By B.Levi's theorem, $U(\omega)=\sup_{n\in\mathbb{N}}U_n(\omega)<\infty$ for almost every ω. On the other hand, if $\omega\in G$, then there are arbitrarily large j, k such that $X_j(\omega)<q$ and $X_k(\omega)>q'$, so $U(\omega)=\infty$. This means that $\hat\mu G$ must be 0, contrary to the choice of q, q'. **X**

(c) Thus we must in fact have $Y=_{\text{a.e.}}Z$, and $\langle X_n(\omega)\rangle_{n\in\mathbb{N}}$ is convergent for almost every ω, as claimed.

275H Theorem Let (Ω,Σ,μ) be a probability space, and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of Σ. Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$. Then the following are equiveridical:

(i) there is a random variable X, of finite expectation, such that X_n is a conditional expectation of X on Σ_n for every n;

(ii) $\{X_n:n\in\mathbb{N}\}$ is uniformly integrable;

(iii) $X_\infty(\omega)=\lim_{n\to\infty}X_n(\omega)$ is defined in $\mathbb{R}$ for almost every ω, and $\mathbb{E}(|X_\infty|)=\lim_{n\to\infty}\mathbb{E}(|X_n|)<\infty$.

proof (i)⇒(ii) By 246D, the set of all conditional expectations of X is uniformly integrable, so $\{X_n:n\in\mathbb{N}\}$ is surely uniformly integrable.

(ii)⇒(iii) If $\{X_n:n\in\mathbb{N}\}$ is uniformly integrable, we surely have $\sup_{n\in\mathbb{N}}\mathbb{E}(|X_n|)<\infty$, so 275G tells us that X_∞ is defined almost everywhere. By 246Ja, X_∞ is integrable and $\lim_{n\to\infty}\mathbb{E}(|X_n-X_\infty|)=0$. Consequently $\mathbb{E}(|X_\infty|)=\lim_{n\to\infty}\mathbb{E}(|X_n|)<\infty$.

(iii)⇒(i) Because $\mathbb{E}(|X_\infty|)=\lim_{n\to\infty}\mathbb{E}(|X_n|)$, $\lim_{n\to\infty}\mathbb{E}(|X_n-X_\infty|)=0$ (245H(a-ii)). Now let $n\in\mathbb{N}$, $E\in\Sigma_n$. Then

$$\int_E X_n=\lim_{m\to\infty}\int_E X_m=\int_E X_\infty.$$

As E is arbitrary, X_n is a conditional expectation of X_∞ on Σ_n.

275I Theorem Let (Ω,Σ,μ) be a probability space, and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of Σ; write Σ_∞ for the σ-algebra generated by $\bigcup_{n\in\mathbb{N}}\Sigma_n$. Let X be any real-valued random variable on Ω with finite expectation, and for each $n\in\mathbb{N}$ let X_n be a conditional expectation of X on Σ_n. Then $X_\infty(\omega)=\lim_{n\to\infty}X_n(\omega)$ is defined almost everywhere; $\lim_{n\to\infty}\mathbb{E}(|X_\infty-X_n|)=0$, and X_∞ is a conditional expectation of X on Σ_∞.

proof (a) By 275G-275H, we know that X_∞ is defined almost everywhere, and, as remarked in 275H, $\lim_{n\to\infty}\mathbb{E}(|X_\infty-X_n|)=0$. To see that X_∞ is a conditional expectation of X on Σ_∞, set

$$\mathcal{A}=\{E:E\in\Sigma_\infty,\,\int_E X_\infty=\int_E X\},\quad\mathcal{I}=\bigcup_{n\in\mathbb{N}}\Sigma_n.$$

Now $\mathcal{I}$ and $\mathcal{A}$ satisfy the conditions of the Monotone Class Theorem (136B). **P** $(\boldsymbol{\alpha})$ Of course $\Omega\in\mathcal{I}$ and $\mathcal{I}$ is closed under finite intersections, because $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence of σ-algebras; in fact $\mathcal{I}$ is a subalgebra of $\mathcal{P}\Omega$, and is closed under finite unions and complements. $(\boldsymbol{\beta})$ If $E\in\mathcal{I}$, say $E\in\Sigma_n$; then

$$\int_E X_\infty=\lim_{m\to\infty}\int_E X_m=\int_E X,$$

as in (iii)⇒(i) of 275H, so $E\in\mathcal{A}$. Thus $\mathcal{I}\subseteq\mathcal{A}$. $(\boldsymbol{\gamma})$ If E, $F\in\mathcal{A}$ and $E\subseteq F$, then

$$\int_{F\setminus E} X_\infty = \int_F X_\infty - \int_E X_\infty = \int_F X - \int_E X = \int_{F\setminus E} X,$$

so $F \setminus E \in \mathcal{A}$. **($\delta$)** If $\langle E_k\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence in $\mathcal{A}$ with union E, then

$$\int_E X_\infty = \lim_{k\to\infty} \int_{E_k} X_\infty = \lim_{k\to\infty} \int_{E_k} X = \int_E X,$$

so $E \in \mathcal{A}$. Thus $\mathcal{A}$ is a Dynkin class. **Q**

Consequently, by 136B, $\mathcal{A}$ includes Σ_∞; that is, X_∞ is a conditional expectation of X on Σ_∞.

Remark I have written '$\lim_{n\to\infty} \mathbb{E}(|X_n - X_\infty|) = 0$'; but you may prefer to say '$X_\infty^\bullet = \lim_{n\to\infty} X_n^\bullet$ in $L^1(\mu)$', as in Chapter 24.

The importance of this theorem is such that you may be interested in a proof based on 275D rather than 275E-275G; see 275Xd.

***275J** As a corollary of this theorem I give an important result, a kind of density theorem for product measures.

Proposition Let $\langle(\Omega_n, \Sigma_n, \mu_n)\rangle_{n\in\mathbb{N}}$ be a sequence of probability spaces with product (Ω, Σ, μ). Let X be a real-valued random variable on Ω with finite expectation. For each $n \in \mathbb{N}$ define X_n by setting

$$X_n(\boldsymbol{\omega}) = \int X(\omega_0, \dots, \omega_n, \xi_{n+1}, \dots) d(\xi_{n+1}, \dots)$$

wherever this is defined, where I write '$\int \dots d(\xi_{n+1}, \dots)$' to mean integration with respect to the product measure λ_n' on $\prod_{i\ge n+1} \Omega_i$. Then $X(\boldsymbol{\omega}) = \lim_{n\to\infty} X_n(\boldsymbol{\omega})$ for almost every $\boldsymbol{\omega} = (\omega_0, \omega_1, \dots)$ in Ω, and $\lim_{n\to\infty} \mathbb{E}(|X - X_n|) = 0$.

proof For each n, we can identify μ with the product of λ_n and λ_n', where λ_n is the product measure on $\Omega_0 \times \dots \times \Omega_n$ (254N). So 253H tells us that X_n is a conditional expectation of X on the σ-algebra $\Lambda_n = \{E \times \prod_{i>n} \Omega_i : E \in \operatorname{dom} \lambda_n\}$. Since (by 254N again) we can think of λ_{n+1} as the product of λ_n and μ_{n+1}, $\Lambda_n \subseteq \Lambda_{n+1}$ for each n. So 275I tells us that $\langle X_n\rangle_{n\in\mathbb{N}}$ converges almost everywhere to a conditional expectation X_∞ of X on the σ-algebra Λ_∞ generated by $\bigcup_{n\in\mathbb{N}} \Lambda_n$. Now $\Lambda_\infty \subseteq \Sigma$ and also $\widehat{\bigotimes}_{n\in\mathbb{N}}\Sigma_n \subseteq \Lambda_\infty$, so every member of Σ is sandwiched between two members of Λ_∞ of the same measure (254Ff), and X_∞ must be equal to X almost everywhere. Moreover, 275I also tells us that

$$\lim_{n\to\infty} \mathbb{E}(|X - X_n|) = \lim_{n\to\infty} \mathbb{E}(|X_\infty - X_n|) = 0,$$

as required.

275K Reverse martingales We have a result corresponding to 275I for *decreasing* sequences of σ-algebras. While this is used less often than 275G-275I, it does have very important applications.

Theorem Let (Ω, Σ, μ) be a probability space, and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-increasing sequence of σ-subalgebras of Σ, with intersection Σ_∞. Let X be any real-valued random variable with finite expectation, and for each $n \in \mathbb{N}$ let X_n be a conditional expectation of X on Σ_n. Then $X_\infty = \lim_{n\to\infty} X_n$ is defined almost everywhere and is a conditional expectation of X on Σ_∞.

proof (a) Set $H = \bigcap_{n\in\mathbb{N}} \operatorname{dom} X_n$, so that H is conegligible. For $n \in \mathbb{N}$, $a < b$ in $\mathbb{R}$, and $\omega \in H$, write $U_{abn}(\omega)$ for the number of up-crossings from a to b in the list $X_n(\omega), X_{n-1}(\omega), \dots, X_0(\omega)$ (275E). Then

$$\mathbb{E}(U_{abn}) \le \frac{1}{b-a}\mathbb{E}((X_0 - X_n)^+)$$

(275F)

$$\le \frac{1}{b-a}\mathbb{E}(|X_0| + |X_n|) \le \frac{2}{b-a}\mathbb{E}(|X_0|) < \infty.$$

So $\lim_{n\to\infty} U_{abn}(\omega)$ is finite for almost every ω. But this means that

$$\{\omega : \liminf_{n\to\infty} X_n(\omega) < a, \limsup_{n\to\infty} X_n(\omega) > b\}$$

is negligible. As a and b are arbitrary, $\langle X_n\rangle_{n\in\mathbb{N}}$ is convergent a.e., just as in 275G. Set $X_\infty(\omega) = \lim_{n\to\infty} X_n(\omega)$ whenever this is defined in $\mathbb{R}$.

(b) By 246D, $\{X_n : n \in \mathbb{N}\}$ is uniformly integrable, so $\mathbb{E}(|X_n - X_\infty|) \to 0$ as $n \to \infty$ (246Ja), and

$$\int_E X_\infty = \lim_{n\to\infty} \int_E X_n = \int_E X_0$$

for every $E \in \Sigma_\infty$.

(c) Now there is a conegligible set $G \in \Sigma_\infty$ such that $G \subseteq \operatorname{dom} X_\infty$ and $X_\infty \restriction G$ is Σ_∞-measurable. **P** For each $n \in \mathbb{N}$, there is a conegligible set $G_n \in \Sigma_n$ such that $G_n \subseteq \operatorname{dom} X_n$ and $X_n \restriction G_n$ is Σ_n-measurable. Set $G' = \bigcup_{n\in\mathbb{N}} \bigcap_{m\ge n} G_m$; then, for any $r \in \mathbb{N}$, $G' = \bigcup_{n\ge r} \bigcap_{m\ge n} G_m$ belongs to Σ_r, so $G' \in \Sigma_\infty$, while of course G' is conegligible. For $n \in \mathbb{N}$, set $X'_n(\omega) = X_n(\omega)$ for $\omega \in G_n$, 0 for $\omega \in \Omega \setminus G_n$; then for $\omega \in G'$, $\lim_{n\to\infty} X'_n(\omega) = \lim_{n\to\infty} X_n(\omega)$ if either is defined in $\mathbb{R}$. Writing $X'_\infty = \lim_{n\to\infty} X'_n$ whenever this is defined in $\mathbb{R}$, 121F and 121H tell us that X'_∞ is Σ_r-measurable and $\operatorname{dom} X'_\infty \in \Sigma_r$ for every $r \in \mathbb{N}$, so that $G'' = \operatorname{dom} X'_\infty$ belongs to Σ_∞ and X'_∞ is Σ_∞-measurable. We also know, from (b), that G'' is conegligible. So setting $G = G' \cap G''$ we have the result. **Q**

Thus X_∞ is a conditional expectation of X on Σ_∞.

275L Stopping times In a sense, the main work of this section is over; I have no room for any more theorems of importance comparable to 275G-275I. However, it would be wrong to leave this chapter without briefly describing one of the most fruitful ideas of the subject.

Definition Let (Ω, Σ, μ) be a probability space, with completion $(\Omega, \hat{\Sigma}, \hat{\mu})$, and $\langle \Sigma_n \rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of $\hat{\Sigma}$. A **stopping time adapted to** $\langle \Sigma_n \rangle_{n\in\mathbb{N}}$ (also called '**optional time**', '**Markov time**') is a function τ from Ω to $\mathbb{N} \cup \{\infty\}$ such that $\{\omega : \tau(\omega) \le n\} \in \Sigma_n$ for every $n \in \mathbb{N}$.

Remark Of course the condition

$$\{\omega : \tau(\omega) \le n\} \in \Sigma_n \text{ for every } n \in \mathbb{N}$$

can be replaced by the equivalent condition

$$\{\omega : \tau(\omega) = n\} \in \Sigma_n \text{ for every } n \in \mathbb{N}.$$

I give priority to the former expression because it is more appropriate to other index sets (see 275Cc).

275M Examples (a) If $\langle X_n \rangle_{n\in\mathbb{N}}$ is a martingale adapted to a sequence $\langle \Sigma_n \rangle_{n\in\mathbb{N}}$ of σ-algebras, and H_n is a Borel subset of $\mathbb{R}^{n+1}$ for each n, then we have a stopping time τ adapted to $\langle \Sigma_n \rangle_{n\in\mathbb{N}}$ defined by the formula

$$\tau(\omega) = \inf\{n : \omega \in \textstyle\bigcap_{i\le n} \operatorname{dom} X_i,\ (X_0(\omega), \dots , X_n(\omega)) \in H_n\},$$

setting $\inf \emptyset = \infty$ as usual. (For by 121Ka the set $E_n = \{\omega : (X_0(\omega), \dots , X_n(\omega)) \in H_n\}$ belongs to Σ_n for each n, and $\{\omega : \tau(\omega) \le n\} = \bigcup_{i\le n} E_i$.) In particular, for instance, the formulae

$$\inf\{n : X_n(\omega) \ge a\}, \quad \inf\{n : |X_n(\omega)| > a\}$$

define stopping times.

(b) Any constant function $\tau : \Omega \to \mathbb{N} \cup \{\infty\}$ is a stopping time. If τ, τ' are two stopping times adapted to the same sequence $\langle \Sigma_n \rangle_{n\in\mathbb{N}}$ of σ-algebras, then $\tau \wedge \tau'$ is a stopping time adapted to $\langle \Sigma_n \rangle_{n\in\mathbb{N}}$, setting $(\tau \wedge \tau')(\omega) = \min(\tau(\omega), \tau'(\omega))$ for $\omega \in \Omega$.

275N Lemma Let (Ω, Σ, μ) be a complete probability space, and $\langle \Sigma_n \rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of Σ. Suppose that τ and τ' are stopping times on Ω, and $\langle X_n \rangle_{n\in\mathbb{N}}$ a martingale, all adapted to $\langle \Sigma_n \rangle_{n\in\mathbb{N}}$.

(a) The family

$$\tilde{\Sigma}_\tau = \{E : E \in \Sigma,\ E \cap \{\omega : \tau(\omega) \le n\} \in \Sigma_n \text{ for every } n \in \mathbb{N}\}$$

is a σ-subalgebra of Σ.

(b) If $\tau(\omega) \le \tau'(\omega)$ for every ω, then $\tilde{\Sigma}_\tau \subseteq \tilde{\Sigma}_{\tau'}$.

(c) Now suppose that τ is finite almost everywhere. Set

$$\tilde{X}_\tau(\omega) = X_{\tau(\omega)}(\omega)$$

whenever $\tau(\omega) < \infty$ and $\omega \in \operatorname{dom} X_{\tau(\omega)}$. Then $\operatorname{dom} \tilde{X}_\tau \in \tilde{\Sigma}_\tau$ and $\tilde{X}_\tau$ is $\tilde{\Sigma}_\tau$-measurable.

(d) If τ is essentially bounded, that is, there is some $m \in \mathbb{N}$ such that $\tau \le m$ almost everywhere, then $\mathbb{E}(\tilde{X}_\tau)$ exists and is equal to $\mathbb{E}(X_0)$.

(e) If $\tau \le \tau'$ almost everywhere, and τ' is essentially bounded, then $\tilde{X}_\tau$ is a conditional expectation of $\tilde{X}_{\tau'}$ on $\tilde{\Sigma}_\tau$.

proof (a) This is elementary. Write $H_n = \{\omega : \tau(\omega) \le n\}$ for each $n \in \mathbb{N}$. The empty set belongs to $\tilde{\Sigma}_\tau$ because it belongs to Σ_n for every n. If $E \in \tilde{\Sigma}_\tau$, then

$$(\Omega \setminus E) \cap H_n = H_n \setminus (E \cap H_n) \in \Sigma_n$$

because $H_n \in \Sigma_n$; this is true for for every n, so $X \setminus E \in \tilde{\Sigma}_\tau$. If $\langle E_k\rangle_{k\in\mathbb{N}}$ is any sequence in $\tilde{\Sigma}_\tau$ then

$$(\bigcup_{k\in\mathbb{N}} E_k) \cap H_n = \bigcup_{k\in\mathbb{N}} E_k \cap H_n \in \Sigma_n$$

for every n, so $\bigcup_{k\in\mathbb{N}} E_k \in \tilde{\Sigma}_\tau$.

(b) If $E \in \tilde{\Sigma}_\tau$ then of course $E \in \Sigma$, and if $n \in \mathbb{N}$ then $\{\omega : \tau'(\omega) \le n\} \subseteq \{\omega : \tau(\omega) \le n\}$, so that

$$E \cap \{\omega : \tau'(\omega) \le n\} = E \cap \{\omega : \tau(\omega) \le n\} \cap \{\omega : \tau'(\omega) \le n\}$$

belongs to Σ_n; as n is arbitrary, $E \in \tilde{\Sigma}_{\tau'}$.

(c) Set $H_n = \{\omega : \tau(\omega) \le n\}$ for each $n \in \mathbb{N}$. For any $a \in \mathbb{R}$,

$$H_n \cap \{\omega : \omega \in \operatorname{dom} \tilde{X}_\tau,\ \tilde{X}_\tau(\omega) \le a\} = \bigcup_{k\le n} \{\omega : \tau(\omega) = k,\ \omega \in \operatorname{dom} X_k,\ X_k(\omega) \le a\} \in \Sigma_n.$$

As n is arbitrary,

$$G_a = \{\omega : \omega \in \operatorname{dom} \tilde{X}_\tau,\ \tilde{X}_\tau(\omega) \le a\} \in \tilde{\Sigma}_\tau.$$

As a is arbitrary, $\operatorname{dom} \tilde{X}_\tau = \bigcup_{m\in\mathbb{N}} G_m \in \tilde{\Sigma}_\tau$ and $\tilde{X}_\tau$ is $\tilde{\Sigma}_\tau$-measurable.

(d) Set $H_k = \{\omega : \tau(\omega) = k\}$ for $k \le m$. Then $\bigcup_{k\le m} H_k$ is conegligible, so

$$\mathbb{E}(X_\tau) = \sum_{k=0}^m \int_{H_k} X_k = \sum_{k=0}^m \int_{H_k} X_m = \int_\Omega X_m = \int_\Omega X_0.$$

(e) Suppose $\tau' \le n$ almost everywhere. Set $H_k = \{\omega : \tau(\omega) = k\}$, $H'_k = \{\omega : \tau'(\omega) = k\}$ for each k; then both $\langle H_k\rangle_{k\le n}$ and $\langle H'_k\rangle_{k\le n}$ are partitions of conegligible subsets of X. Now suppose that $E \in \Sigma_\tau$. Then

$$\int_E \tilde{X}_\tau = \sum_{k=0}^n \int_{E\cap H_k} \tilde{X}_\tau = \sum_{k=0}^n \int_{E\cap H_k} X_k = \sum_{k=0}^n \int_{E\cap H_k} X_n = \int_E X_n$$

because $E \cap H_k \in \Sigma_k$ for every k. By (b), $E \in \tilde{\Sigma}_{\tau'}$, so we also have $\int_E \tilde{X}_{\tau'} = \int_E X_n$. Thus $\int_E \tilde{X}_\tau = \int_E \tilde{X}_{\tau'}$ for every $E \in \tilde{\Sigma}_\tau$, as claimed.

275O Proposition Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale and τ a stopping time, both adapted to the same sequence $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ of σ-algebras. For each n, set $(\tau\wedge n)(\omega) = \min(\tau(\omega), n)$ for $\omega \in \Omega$; then $\tau \wedge n$ is a stopping time, and $\langle\tilde{X}_{\tau\wedge n}\rangle_{n\in\mathbb{N}}$ is a martingale adapted to $\langle\tilde{\Sigma}_{\tau\wedge n}\rangle_{n\in\mathbb{N}}$, defining $\tilde{X}_{\tau\wedge n}$ and $\tilde{\Sigma}_{\tau\wedge n}$ as in 275N.

proof As remarked in 275Mb, each $\tau \wedge n$ is a stopping time. If $m \le n$, then $\tilde{\Sigma}_{\tau\wedge m} \subseteq \tilde{\Sigma}_{\tau\wedge n}$ by 275Nb. Each $\tilde{X}_{\tau\wedge m}$ is $\tilde{\Sigma}_{\tau\wedge m}$-measurable, with domain belonging to $\tilde{\Sigma}_{\tau\wedge m}$, by 275Nc, and has finite expectation, by 275Nd; finally, if $m \le n$, then $\tilde{X}_{\tau\wedge m}$ is a conditional expectation of $\tilde{X}_{\tau\wedge n}$ on $\tilde{\Sigma}_{\tau\wedge m}$, by 275Ne.

275P Corollary Suppose that (Ω, Σ, μ) is a probability space and $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale on Ω such that $W = \sup_{n\in\mathbb{N}} |X_{n+1} - X_n|$ is finite almost everywhere and has finite expectation. Then for almost every $\omega \in \Omega$, either $\lim_{n\to\infty} X_n(\omega)$ exists in $\mathbb{R}$ or $\sup_{n\in\mathbb{N}} X_n(\omega) = \infty$ and $\inf_{n\in\mathbb{N}} X_n(\omega) = -\infty$.

proof Let $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ be a non-decreasing sequence of σ-algebras to which $\langle X_n\rangle_{n\in\mathbb{N}}$ is adapted. Let H be the conegligible set $\bigcap_{n\in\mathbb{N}} \operatorname{dom} X_n \cap \{\omega : W(\omega) < \infty\}$. For each $m \in \mathbb{N}$, set

$$\tau_m(\omega) = \inf\{n : \omega \in \operatorname{dom} X_n,\ X_n(\omega) > m\}.$$

As in 275Ma, τ_m is a stopping time adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$. Set

$$Y_{mn} = \tilde{X}_{\tau_m\wedge n},$$

defined as in 275Nc, so that $\langle Y_{mn}\rangle_{n\in\mathbb{N}}$ is a martingale, by 275O. If $\omega \in H$, then either $\tau_m(\omega) > n$ and

$$Y_{mn}(\omega) = X_n(\omega) \le m,$$

or $0 < \tau_m(\omega) \le n$ and

$$Y_{mn}(\omega) = X_{\tau_m(\omega)}(\omega) \le W(\omega) + X_{\tau_m(\omega)-1}(\omega) \le W(\omega) + m,$$

or $\tau_m(\omega) = 0$ and

$$Y_{mn}(\omega) = X_0(\omega).$$

Thus

$$Y_{mn}(\omega) \le |X_0(\omega)| + W(\omega) + m$$

for every $\omega \in H$, and

$$|Y_{mn}(\omega)| = 2\max(0, Y_{mn}(\omega)) - Y_{mn}(\omega) \le 2(|X_0(\omega)| + W(\omega) + m) - Y_{mn}(\omega),$$

$$\mathbb{E}(|Y_{mn}|) \le 2\mathbb{E}(|X_0|) + 2\mathbb{E}(W) + 2m - \mathbb{E}(Y_{mn}) = 2\mathbb{E}(|X_0|) + 2\mathbb{E}(W) + 2m - \mathbb{E}(X_0)$$

by 275Nd. As this is true for every $n \in \mathbb{N}$, $\sup_{n\in\mathbb{N}} \mathbb{E}(|Y_{mn}|) < \infty$, and $\lim_{n\to\infty} Y_{mn}$ is defined in $\mathbb{R}$ almost everywhere, by Doob's Martingale Convergence Theorem (275G). Let F_m be the conegligible set on which $\langle Y_{mn}\rangle_{n\in\mathbb{N}}$ converges. Set $H^* = H \cap \bigcap_{m\in\mathbb{N}} F_m$, so that H^* is conegligible.

Now consider

$$E = \{\omega : \omega \in H^*, \sup_{n\in\mathbb{N}} X_n(\omega) < \infty\}.$$

For any $\omega \in E$, there must be an $m \in \mathbb{N}$ such that $\sup_{n\in\mathbb{N}} X_n(\omega) \le m$. Now this means that $Y_{mn}(\omega) = X_n(\omega)$ for every n, and as $\omega \in F_m$ we have

$$\lim_{n\to\infty} X_n(\omega) = \lim_{n\to\infty} Y_{mn}(\omega) \in \mathbb{R}.$$

This means that $\langle X_n(\omega)\rangle_{n\in\mathbb{N}}$ is convergent for almost every ω such that $\{X_n(\omega) : n \in \mathbb{N}\}$ is bounded above.

Similarly, $\langle X_n(\omega)\rangle_{n\in\mathbb{N}}$ is convergent for almost every ω such that $\{X_n(\omega) : n \in \mathbb{N}\}$ is bounded below, which completes the proof.

275X Basic exercises >(a) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of random variables with zero expectation and finite variance. Set $s_n = (\sum_{i=0}^n \mathrm{Var}(X_i))^{1/2}$, $Y_n = (X_0 + \ldots + X_n)^2 - s_n^2$ for each n. Show that $\langle Y_n\rangle_{n\in\mathbb{N}}$ is a martingale.

>(b) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale. Show that for any $\epsilon > 0$, $\Pr(\sup_{n\in\mathbb{N}} |X_n| \ge \epsilon) \le \frac{1}{\epsilon}\sup_{n\in\mathbb{N}} \mathbb{E}(|X_n|)$.

(c) Pólya's urn scheme Imagine a box containing red and white balls. At each move, a ball is drawn at random from the box and replaced together with another of the same colour. (i) Writing R_n, W_n for the numbers of red and white balls after the nth move and $X_n = R_n/(R_n + W_n)$, show that $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale. (ii) Starting from $R_0 = W_0 = 1$, find the distribution of (R_n, W_n) for each n. (iii) Show that $X = \lim_{n\to\infty} X_n$ is defined almost everywhere, and find its distribution when $R_0 = W_0 = 1$. (See Feller 66 for a discussion of other starting values.)

>(d) Let (Ω, Σ, μ) be a probability space, and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of Σ; for each $n \in \mathbb{N}$ let $P_n : L^1 \to L^1$ be the conditional expectation operator corresponding to Σ_n, where $L^1 = L^1(\mu)$ (242J). (i) Show that $V = \{u : u \in L^1, \lim_{n\to\infty} \|P_n u - u\|_1 = 0\}$ is a $\|\,\|_1$-closed linear subspace of L^1. (ii) Show that $\{E : E \in \Sigma, \chi E^\bullet \in V\}$ is a Dynkin class including $\bigcup_{n\in\mathbb{N}} \Sigma_n$, so includes the σ-algebra Σ_∞ generated by $\bigcup_{n\in\mathbb{N}} \Sigma_n$. (iii) Show that if $u \in L^1$ then $v = \sup_{n\in\mathbb{N}} P_n|u|$ is defined in L^1 and is of the form $W^\bullet$ where $\Pr(W \ge \epsilon) \le \frac{1}{\epsilon}\|u\|_1$ for every $\epsilon > 0$. (*Hint*: 275D.) (iv) Show that if X is a Σ_∞-measurable random variable with finite expectation, and for each $n \in \mathbb{N}$ X_n is a conditional expectation of X on Σ_n, then $X^\bullet \in V$ and $X =_{\text{a.e.}} \lim_{n\to\infty} X_n$. (*Hint*: apply (iii) to $u = (X - X_m)^\bullet$ for large m.)

(e) Let (Ω, Σ, μ) be a probability space, $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of Σ, and Σ_∞ the σ-algebra generated by $\bigcup_{n\in\mathbb{N}} \Sigma_n$. For each $n \in \mathbb{N} \cup \{\infty\}$ let $P_n : L^1 \to L^1$ be the conditional expectation operator corresponding to Σ_n, where $L^1 = L^1(\mu)$. Show that $\lim_{n\to\infty} \|P_n u - u\|_p = 0$ whenever $p \in [1, \infty[$ and $u \in L^p(\mu)$. (*Hint*: 275Xd, 233J/242K, 246Xg.)

(f) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale, and suppose that $p \in\,]1, \infty[$ is such that $\sup_{n\in\mathbb{N}} \|X_n\|_p < \infty$. Show that $X = \lim_{n\to\infty} X_n$ is defined almost everywhere and that $\lim_{n\to\infty} \|X_n - X\|_p = 0$.

>(g) Let (Ω, Σ, μ) be $[0, 1]$ with Lebesgue measure. For each $n \in \mathbb{N}$ let Σ_n be the finite subalgebra of Σ generated by intervals of the type $[0, 2^{-n}r]$ for $r \le 2^{-n}$. Use 275I to show that for any integrable $X : [0, 1] \to \mathbb{R}$ we must have $X(t) = \lim_{n\to\infty} 2^n \int_{I_n(t)} X$ for almost every $t \in [0, 1[$, where $I_n(t)$ is the interval of the form $[2^{-n}r, 2^{-n}(r + 1)[$ containing t. Compare this result with 223A and 261Yd.

(h) In 275K, show that $\lim_{n\to\infty} \|X_n - X_\infty\|_p = 0$ for any $p \in [1, \infty[$ such that $\|X_0\|_p$ is finite. (Compare 275Xe.)

(i) Let (Ω,Σ,μ) be a probability space, with completion $(\Omega,\hat{\Sigma},\hat{\mu})$, and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of $\hat{\Sigma}$. Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a uniformly integrable martingale adapted to Σ_n, and set $X_\infty=\lim_{n\to\infty}X_n$. Let τ be a stopping time adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$, and set $\tilde{X}_\tau(\omega)=X_{\tau(\omega)}(\omega)$ whenever $\omega\in\operatorname{dom}X_{\tau(\omega)}$, allowing ∞ as a value of $\tau(\omega)$. Show that $\tilde{X}_\tau$ is a conditional expectation of X_∞ on $\tilde{\Sigma}_\tau$, as defined in 275N.

(j) Let (Ω,Σ,μ) be a probability space, with completion $(\Omega,\hat{\Sigma},\hat{\mu})$, and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of $\hat{\Sigma}$. Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale and τ a stopping time, both adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$. Suppose that $\sup_{n\in\mathbb{N}}\mathbb{E}(|X_n|)<\infty$ and that τ is finite almost everywhere. Show that $\tilde{X}_\tau$, as defined in 275Nc, has finite expectation, but that $\mathbb{E}(\tilde{X}_\tau)$ need not be equal to $\mathbb{E}(X_0)$.

(k)(i) Find a martingale $\langle X_n\rangle_{n\in\mathbb{N}}$ such that $\langle X_{2n}\rangle_{n\in\mathbb{N}}\to 0$ a.e. but $|X_{2n+1}|\ge 1$ a.e. for every $n\in\mathbb{N}$. (ii) Find a martingale which converges in measure but is not convergent a.e.

275Y Further exercises (a) Let (Ω,Σ,μ) be a complete probability space, $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of Σ all containing every negligible set, and $\langle X_n\rangle_{n\in\mathbb{N}}$ a martingale adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$. Let ν be another probability measure with domain Σ which is absolutely continuous with respect to μ, with Radon-Nikodým derivative Z. For each $n\in\mathbb{N}$ let Z_n be a conditional expectation of Z on Σ_n (with respect to the measure μ). (i) Show that Z_n is a Radon-Nikodým derivative of $\nu{\restriction}\Sigma_n$ with respect to $\mu{\restriction}\Sigma_n$. (ii) Set $W_n(\omega)=Z_n(\omega)/Z_{n-1}(\omega)$ if this is defined in $\mathbb{R}$, otherwise 0. For $n\ge 1$, let V_n be a conditional expectation of $W_n\times(X_n-X_{n-1})$ on Σ_{n-1} (with respect to the measure μ). Set $Y_0=X_0$, $Y_n=X_n-\sum_{k=1}^n V_k$ for $n\ge 1$. Show that $\langle Y_n\rangle_{n\in\mathbb{N}}$ is a martingale adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ with respect to the measure ν.

(b) Combine the ideas of 275Cc with those of 275Cd-275Ce to describe a notion of 'martingale indexed by I', where I is an arbitrary partially ordered set.

(c) Let $\langle X_k\rangle_{k\in\mathbb{N}}$ be a martingale on a complete probability space (Ω,Σ,μ), and fix $n\in\mathbb{N}$. Set $X^*=\max(|X_0|,\dots,|X_n|)$. Let $p\in\,]1,\infty[$. Show that $\|X^*\|_p\le\frac{p}{p-1}\|X_n\|_p$. (*Hint*: set $F_t=\{\omega:X^*(\omega)\ge t\}$. Show that $t\mu F_t\le\int_{F_t}|X_n|$. Using Fubini's theorem on $\Omega\times[0,\infty[$ and on $\Omega\times[0,\infty[\,\times[0,\infty[$, show that

$$\mathbb{E}((X^*)^p)=p\int_0^\infty t^{p-1}\hat{\mu}F_t dt,$$

$$\int_0^\infty t^{p-2}\int_{F_t}|X_n|dt=\frac{1}{p-1}\mathbb{E}(|X_n|\times(X^*)^{p-1}),$$

$$\mathbb{E}(|X_n|\times(X^*)^{p-1})\le\|X_n\|_p\|X^*\|_p^{p-1}.$$

Compare 286A.)

(d) Let $\langle X_k\rangle_{k\in\mathbb{N}}$ be a martingale on a complete probability space (Ω,Σ,μ), and fix $n\in\mathbb{N}$. Set $X^*=\max(|X_0|,\dots,|X_n|)$, $F_t=\{\omega:X^*(\omega)\ge t\}$, $G_t=\{\omega:|X_n(\omega)|\ge\frac{1}{2}t\}$ for $t\ge 0$. (i) Show that $t\mu F_t\le 2\int_{G_t}|X_n|$ for every $t\ge 0$. (ii) Show that $\mathbb{E}(X^*)\le 1+2\ln 2\mathbb{E}(|X_n|)+2\mathbb{E}(|X_n|\times\ln^+|X_n|)$, where $\ln^+ t=\ln t$ for $t\ge 1$, 0 for $t\in[0,1]$.

(e) Let (Ω,Σ,μ) be a probability space and $\langle\Sigma_i\rangle_{i\in I}$ a countable family of σ-subalgebras of Σ such that for any i, $j\in I$ either $\Sigma_i\subseteq\Sigma_j$ or $\Sigma_j\subseteq\Sigma_i$. Let X be a real-valued random variable on Ω such that $\|X\|_p<\infty$, where $1<p<\infty$, and suppose that X_i is a conditional expectation of X on Σ_i for each $i\in I$. Show that $\|\sup_{i\in I}|X_i|\|_p\le\frac{p}{p-1}\|X\|_p$.

(f) Let (Ω,Σ,μ) be a probability space, with completion $(\Omega,\hat{\Sigma},\hat{\mu})$, and let $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ be a non-decreasing sequence of σ-subalgebras of $\hat{\Sigma}$. Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a sequence of μ-integrable real-valued functions such that $\operatorname{dom}X_n\in\Sigma_n$ and X_n is Σ_n-measurable for each $n\in\mathbb{N}$. We say that $\langle X_n\rangle_{n\in\mathbb{N}}$ is a **submartingale adapted to** $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ (also called '**semi-martingale**') if $\int_E X_{n+1}\ge\int_E X_n$ for every $n\in\mathbb{N}$ and every $E\in\Sigma_n$. Prove versions of 275D, 275F, 275G, 275Xf for submartingales.

(g) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale, and $\phi:\mathbb{R}\to\mathbb{R}$ a convex function. Show that $\langle\phi(X_n)\rangle_{n\in\mathbb{N}}$ is a submartingale. (*Hint*: 233J.) Re-examine part (b-ii) of the proof of 275F in the light of this fact.

(h) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of non-negative random variables all with expectation 1. Set $W_n=X_0\times\ldots\times X_n$ for every n. (i) Show that $W=\lim_{n\to\infty}W_n$ is defined a.e. (ii) Show that $\mathbb{E}(W)$ is either 0 or 1. (*Hint*:

suppose $\mathbb{E}(W) > 0$. Set $Z_n = \lim_{m\to\infty} X_n \times \ldots \times X_m$. Show that $\lim_{n\to\infty} Z_n = 1$ when $0 < W < \infty$, therefore a.e., by the zero-one law, while $\mathbb{E}(Z_n) \le 1$, by Fatou's lemma, so $\lim_{n\to\infty} \mathbb{E}(Z_n) = 1$, while $\mathbb{E}(W) = \mathbb{E}(W_n)\mathbb{E}(Z_{n+1})$ for every n.) (iii) Set $\gamma = \prod_{n=0}^{\infty} \mathbb{E}(\sqrt{X_n})$. Show that $\gamma > 0$ iff $\mathbb{E}(W) = 1$. (*Hint*: $\Pr(W_n \ge \frac{1}{4}\gamma^2) \ge \frac{1}{4}\gamma^2$ for every n, so if $\gamma > 0$ then W cannot be zero a.e.; while $\mathbb{E}(\sqrt{W}) \le \gamma$.)

(i) Let $\langle(\Omega_n, \Sigma_n, \mu_n)\rangle_{n\in\mathbb{N}}$ be a sequence of probability spaces with product (Ω, Σ, μ). Suppose that for each $n \in \mathbb{N}$ we have a probability measure ν_n, with domain Σ_n, which is absolutely continuous with respect to μ_n, with Radon-Nikodým derivative f_n, and suppose that $\prod_{n=0}^{\infty} \int \sqrt{f_n} d\mu_n > 0$. Let ν be the product of $\langle\nu_n\rangle_{n\in\mathbb{N}}$. Show that ν is an indefinite-integral measure over μ, with Radon-Nikodým derivative f, where $f(\boldsymbol{\omega}) = \prod_{n=0}^{\infty} f_n(\omega_n)$ for μ-almost every $\boldsymbol{\omega} = \langle\omega_n\rangle_{n\in\mathbb{N}}$ in Ω. (*Hint*: use 275Yh to show that $\int f d\mu = 1$.)

(j) Let $\langle p_n\rangle_{n\in\mathbb{N}}$ be a sequence in $[0, 1]$. Let μ be the usual measure on $\{0, 1\}^{\mathbb{N}}$ (254J) and ν the product of $\langle\nu_n\rangle_{n\in\mathbb{N}}$, where ν_n is the probability measure on $\{0, 1\}$ defined by setting $\nu_n\{1\} = p_n$. Show that ν is an indefinite-integral measure over μ iff $\sum_{n=0}^{\infty} |p_n - \frac{1}{2}|^2 < \infty$.

(k) Let (Ω, Σ, μ) be a probability space, $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of Σ and $\langle X_n\rangle_{n\in\mathbb{N}}$ a sequence of random variables on Ω such that $\mathbb{E}(\sup_{n\in\mathbb{N}} |X_n|)$ is finite and $X = \lim_{n\to\infty} X_n$ is defined almost everywhere. For each n, let Y_n be a conditional expectation of X_n on Σ_n. Show that $\langle Y_n\rangle_{n\in\mathbb{N}}$ converges almost everywhere to a conditional expectation of X on the σ-algebra generated by $\bigcup_{n\in\mathbb{N}} \Sigma_n$.

(l) Show that 275Yk can fail if $\langle X_n\rangle_{n\in\mathbb{N}}$ is merely uniformly integrable, rather than dominated by an integrable function.

(m) Let (Ω, Σ, μ) be a probability space, $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ an independent sequence of σ-subalgebras of Σ, and X a random variable on Ω with finite variance. Let X_n be a conditional expectation of X on Σ_n for each n. Show that $\lim_{n\to\infty} X_n = \mathbb{E}(X)$ almost everywhere. (*Hint*: consider $\sum_{n=0}^{\infty} \mathrm{Var}(X_n)$.)

(n) Let (Ω, Σ, μ) be a complete probability space, and $\langle X_n\rangle_{n\in\mathbb{N}}$ an independent sequence of random variables on Ω, all with the same distribution, and of finite expectation. For each n, set $S_n = \frac{1}{n+1}(X_0 + \ldots + X_n)$; let Σ_n be the σ-algebra defined by S_n and Σ_n^* the σ-algebra generated by $\bigcup_{m\ge n} \Sigma_m$. Show that S_n is a conditional expectation of X_0 on Σ_n^*. (*Hint*: assume every X_i defined everywhere on Ω. Set $\phi(\omega) = \langle X_i(\omega)\rangle_{i\in\mathbb{N}}$. Show that $\phi : \Omega \to \mathbb{R}^{\mathbb{N}}$ is inverse-measure-preserving for a suitable product measure on $\mathbb{R}^{\mathbb{N}}$, and that every set in Σ_n^* is of the form $\phi^{-1}[H]$ where $H \subseteq \mathbb{R}^{\mathbb{N}}$ is a Borel set invariant under permutations of coordinates in the set $\{0, \ldots, n\}$, so that $\int_E X_i = \int_E X_j$ whenever $i \le j \le n$ and $E \in \Sigma_n^*$.) Hence show that $\langle S_n\rangle_{n\in\mathbb{N}}$ converges almost everywhere. (Compare 273I.)

(o) Formulate and prove versions of the results of this chapter for martingales consisting of functions taking values in $\mathbb{C}$ or $\mathbb{R}^r$ rather than $\mathbb{R}$.

(p)(i) Find a martingale $\langle X_n\rangle_{n\in\mathbb{N}}$ which is convergent in measure, but is not convergent a.e. (Compare 272Yd.) (ii) Find a martingale $\langle X_n\rangle_{n\in\mathbb{N}}$ such that the sequence ν_{X_n} of distributions (271C) is convergent for the vague topology (274Ld), but $\langle X_n\rangle_{n\in\mathbb{N}}$ is not convergent in measure.

(q) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables such that $\sum_{n=0}^{\infty} X_n$ is defined in $\mathbb{R}$ almost everywhere. Suppose that there is an $M \ge 0$ such that $|X_n| \le M$ a.e. for every n. Show that $\sum_{n=0}^{\infty} \mathbb{E}(X_n)$ is defined in $\mathbb{R}$. (*Hint*: 274Ye, 275G.)

(r) Let (Ω, Σ, μ) be a probability space and $\langle X_n\rangle_{n\in\mathbb{N}}$ an independent sequence of real-valued random variables on Ω; set $E_n = \{\omega : \omega \in \mathrm{dom}\, X_n,\ |X_n(\omega)| > 1\}$, $Y_n = X_n \times \chi(\Omega \setminus E_n)$ for each n, and $Z_n(\omega) = \mathrm{med}(-1, X_n(\omega), 1)$ for $n \in \mathbb{N}$ and $\omega \in \mathrm{dom}\, X_n$. Show that the following are equiveridical: (i) $\sum_{n=0}^{\infty} X_n(\omega)$ is defined in $\mathbb{R}$ for almost every ω; (ii) $\sum_{n=0}^{\infty} \hat{\mu}E_n < \infty$, $\sum_{n=0}^{\infty} \mathbb{E}(Y_n)$ is defined in $\mathbb{R}$, and $\sum_{n=0}^{\infty} \mathrm{Var}(Y_n) < \infty$; (iii) $\sum_{n=0}^{\infty} \hat{\mu}E_n < \infty$, $\sum_{n=0}^{\infty} \mathbb{E}(Z_n)$ is defined in $\mathbb{R}$, and $\sum_{n=0}^{\infty} \mathrm{Var}(Z_n) < \infty$. (*Hint*: 273K, 275Yq.) (This is a version of the **Three Series Theorem**.)

275 Notes and comments I hope that the sketch above, though distressingly abbreviated, has suggested some of the richness of the concepts involved, and will provide a foundation for further study. All the theorems of this section have far-reaching implications, but the one which is simply indispensable in advanced measure theory is 275I, 'Lévy's martingale convergence theorem', which I will use in the proof of the Lifting Theorem in Chapter 34 of the next volume.

As for stopping times, I mention them partly in an attempt to cast further light on what martingales are for (see 276Ed below), and partly because the ideas of 275N-275O are so important in modern probability theory that, just as a matter of general knowledge, you should be aware that there is something there. I add 275P as one of the most accessible of the standard results which may be obtained by this method.

276 Martingale difference sequences

Hand in hand with the concept of 'martingale' is that of 'martingale difference sequence' (276A), a direct generalization of the notion of 'independent sequence'. In this section I collect results which can be naturally expressed in terms of difference sequences, including yet another strong law of large numbers (276C). I end the section with a proof of Komlós' theorem (276H).

276A Martingale difference sequences (a) If $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale adapted to a sequence $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ of σ-algebras, then we have

$$\int_E X_{n+1}-X_n=0$$

whenever $E\in\Sigma_n$. Let us say that if (Ω,Σ,μ) is a probability space, with completion $(\Omega,\hat\Sigma,\hat\mu)$, and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence of σ-subalgebras of $\hat\Sigma$, then a **martingale difference sequence adapted to** $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ is a sequence $\langle X_n\rangle_{n\in\mathbb{N}}$ of real-valued random variables on Ω, all with finite expectation, such that (i) $\operatorname{dom}X_n\in\Sigma_n$ and X_n is Σ_n-measurable, for each $n\in\mathbb{N}$ (ii) $\int_E X_{n+1}=0$ whenever $n\in\mathbb{N}$, $E\in\Sigma_n$.

(b) Evidently $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale difference sequence adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ iff $\langle\sum_{i=0}^n X_i\rangle_{n\in\mathbb{N}}$ is a martingale adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$.

(c) Just as in 275Cd, we can say that a sequence $\langle X_n\rangle_{n\in\mathbb{N}}$ is in itself a **martingale difference sequence** if $\langle\sum_{i=0}^n X_i\rangle_{n\in\mathbb{N}}$ is a martingale, that is, if $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale difference sequence adapted to $\langle\tilde\Sigma_n\rangle_{n\in\mathbb{N}}$, where $\tilde\Sigma_n$ is the σ-algebra generated by $\bigcup_{i\le n}\Sigma_{X_i}$.

(d) If $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale difference sequence then $\langle a_nX_n\rangle_{n\in\mathbb{N}}$ is a martingale difference sequence for any real a_n.

(e) If $\langle X_n\rangle_{n\in\mathbb{N}}$ is a martingale difference sequence and $X'_n=_{\text{a.e.}}X_n$ for every n, then $\langle X'_n\rangle_{n\in\mathbb{N}}$ is a martingale difference sequence. (Compare 275Ce.)

(f) Of course the most important example of 'martingale difference sequence' is that of 275Bb: any independent sequence of random variables with zero expectation is a martingale difference sequence. It turns out that some of the theorems of §273 concerning such independent sequences may be generalized to martingale difference sequences.

276B Proposition Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale difference sequence such that $\sum_{n=0}^\infty\mathbb{E}(X_n^2)<\infty$. Then $\sum_{n=0}^\infty X_n$ is defined, and finite, almost everywhere.

proof (a) Let (Ω,Σ,μ) be the underlying probability space, $(\Omega,\hat\Sigma,\hat\mu)$ its completion, and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of $\hat\Sigma$ such that $\langle X_n\rangle_{n\in\mathbb{N}}$ is adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$. Set $Y_n=\sum_{i=0}^n X_i$ for each $n\in\mathbb{N}$. Then $\langle Y_n\rangle_{n\in\mathbb{N}}$ is a martingale adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$.

(b) $\mathbb{E}(Y_n\times X_{n+1})=0$ for each n. **P** Y_n is a sum of random variables with finite variance, so $\mathbb{E}(Y_n^2)<\infty$, by 244Ba; it follows that $Y_n\times X_{n+1}$ has finite expectation, by 244Eb. Because the constant function $\mathbf{0}$ is a conditional expectation of X_{n+1} on Σ_n,

$$\mathbb{E}(Y_n\times X_{n+1})=\mathbb{E}(Y_n\times\mathbf{0})=0,$$

by 242L. **Q**

(c) It follows that $\mathbb{E}(Y_n^2)=\sum_{i=0}^n\mathbb{E}(X_i^2)$ for every n. **P** Induce on n. For the inductive step, we have

$$\mathbb{E}(Y_{n+1}^2)=\mathbb{E}(Y_n^2+2Y_n\times X_{n+1}+X_{n+1}^2)=\mathbb{E}(Y_n^2)+\mathbb{E}(X_{n+1}^2)$$

because, by (b), $\mathbb{E}(Y_n\times X_{n+1})=0$. **Q**

(d) Of course

$$\mathbb{E}(|Y_n|)=\int|Y_n|\times\chi\Omega\le\|Y_n\|_2\|\chi\Omega\|_2=\sqrt{\mathbb{E}(Y_n^2)},$$

so

$$\sup_{n\in\mathbb{N}}\mathbb{E}(|Y_n|)\le\sup_{n\in\mathbb{N}}\sqrt{\mathbb{E}(Y_n^2)}=\sqrt{\textstyle\sum_{i=0}^\infty\mathbb{E}(X_i^2)}<\infty.$$

By 275G, $\lim_{n\to\infty}Y_n$ is defined and finite almost everywhere, that is, $\sum_{i=0}^\infty X_i$ is defined and finite almost everywhere.

276C The strong law of large numbers: fourth form Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale difference sequence, and suppose that $\langle b_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence in $]0,\infty[$, diverging to ∞, such that $\sum_{n=0}^{\infty}\frac{1}{b_n^2}\operatorname{Var}(X_n)<\infty$. Then

$$\lim_{n\to\infty}\frac{1}{b_n}\sum_{i=0}^{n}X_i=0$$

almost everywhere.

proof (Compare 273D.) As usual, write (Ω,Σ,μ) for the underlying probability space. Set

$$\tilde{X}_n=\frac{1}{b_n}X_n$$

for each n; then $\langle\tilde{X}_n\rangle_{n\in\mathbb{N}}$ also is a martingale difference sequence, and

$$\sum_{n=1}^{\infty}\mathbb{E}(\tilde{X}_n^2)=\sum_{n=1}^{\infty}\frac{1}{b_n^2}\operatorname{Var}(X_n)<\infty.$$

By 276B, $\langle\tilde{X}_n(\omega)\rangle_{n\in\mathbb{N}}$ is summable for almost every $\omega\in\Omega$. But by 273Cb,

$$\lim_{n\to\infty}\frac{1}{b_n}\sum_{i=0}^{n}X_i(\omega)=\lim_{n\to\infty}\frac{1}{b_n}\sum_{i=0}^{n}b_i\tilde{X}_i(\omega)=0$$

for all such ω. So we have the result.

276D Corollary Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale such that $b_n=\mathbb{E}(X_n^2)$ is finite for each n.

(a) If $\sup_{n\in\mathbb{N}}b_n$ is infinite, then $\lim_{n\to\infty}\frac{1}{b_n}X_n=0$ a.e.

(b) If $\sup_{n\geq 1}\frac{1}{n}b_n<\infty$, then $\lim_{n\to\infty}\frac{1}{n}X_n=0$ a.e.

proof Consider the martingale difference sequence $\langle Y_n\rangle_{n\in\mathbb{N}}=\langle X_{n+1}-X_n\rangle_{n\in\mathbb{N}}$. Then $\mathbb{E}(Y_n\times X_n)=0$, so $\mathbb{E}(Y_n^2)+\mathbb{E}(X_n^2)=\mathbb{E}(X_{n+1}^2)$ for each n. In particular, $\langle b_n\rangle_{n\in\mathbb{N}}$ must be non-decreasing.

(a) If $\lim_{n\to\infty}b_n=\infty$, take m such that $b_m>0$; then

$$\sum_{n=m}^{\infty}\frac{1}{b_{n+1}^2}\operatorname{Var}(Y_n)=\sum_{n=m}^{\infty}\frac{1}{b_{n+1}^2}(b_{n+1}-b_n)\leq\int_{b_m}^{\infty}\frac{1}{t^2}dt<\infty.$$

By 276C (modifying b_i for $i<m$, if necessary),

$$\lim_{n\to\infty}\frac{1}{b_n}X_n=\lim_{n\to\infty}\frac{1}{b_{n+1}}(X_0+\sum_{i=0}^{n}Y_i)=\lim_{n\to\infty}\frac{1}{b_{n+1}}\sum_{i=0}^{n}Y_i=0$$

almost everywhere.

(b) If $\gamma=\sup_{n\geq 1}\frac{1}{n}b_n<\infty$, then $\frac{1}{(n+1)^2}\leq\min(1,\gamma^2/t^2)$ for $b_n<t\leq b_{n+1}$, so

$$\sum_{n=0}^{\infty}\frac{1}{(n+1)^2}(b_{n+1}-b_n)\leq\gamma+\gamma^2\int_{\gamma}^{\infty}\frac{1}{t^2}dt<\infty,$$

and, by the same argument as before, $\lim_{n\to\infty}\frac{1}{n}X_n=0$ a.e.

276E 'Impossibility of systems' (a) I return to the word 'martingale' and the idea of a gambling system. Consider a gambler who takes a sequence of 'fair' bets, that is, bets which have payoff expectations of zero, but who chooses which bets to take on the basis of past experience. The appropriate model for such a sequence of random events is a martingale in the sense of 275A, taking Σ_n to be the algebra of all events which are observable up to and including the outcome of the nth bet, and X_n to be the gambler's net gain at that time. (In this model it is natural to take $\Sigma_0=\{\emptyset,\Omega\}$ and $X_0=0$.) Certain paradoxes can arise if we try to imagine this model with atomless Σ_n; to begin with it is perhaps easier to work with the discrete case, in which each Σ_n is finite, or is the set of unions of some countable family of atomic events. Now suppose that the bets involved are just two-way bets, with two equally likely outcomes, but that the gambler chooses his stake each time. In this case we can think of the outcomes as corresponding to an independent sequence $\langle W_n\rangle_{n\in\mathbb{N}}$ of random variables, each taking the values ± 1 with equal probability. The gambler's system must be of the form

$$X_{n+1}=X_n+Z_{n+1}\times W_{n+1},$$

where Z_{n+1} is his stake on the $(n+1)$-st bet, and must be constant on each atom of the σ-algebra Σ_n generated by $W_1,\dots,W_n$. The point is that because $\int_E W_{n+1}=0$ for each $E\in\Sigma_n$, $\mathbb{E}(Z_{n+1}\times W_{n+1})=0$, so $\mathbb{E}(X_{n+1})=\mathbb{E}(X_n)$.

(b) The general result, of which this is a special case, is the following. If $\langle W_n\rangle_{n\in\mathbb{N}}$ is a martingale difference sequence adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$, and $\langle Z_n\rangle_{n\ge 1}$ is a sequence of random variables such that (i) Z_n is Σ_{n-1}-measurable (ii) $Z_n\times W_n$ has finite expectation for each $n\ge 1$, then $W_0, Z_1\times W_1, Z_2\times W_2,\dots$ is a martingale difference sequence adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$; the proof that $\int_E Z_{n+1}\times W_{n+1}=0$ for every $E\in\Sigma_n$ is exactly the argument of (b) of the proof of 276B.

(c) I invited you to restrict your ideas to the discrete case for a moment; but if you feel that you understand what it means to say that a 'system' or **predictable sequence** $\langle Z_n\rangle_{n\ge 1}$ must be adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$, in the sense that every Z_n is Σ_{n-1}-measurable, then any further difficulty lies in the measure theory needed to show that the integrals $\int_E Z_{n+1}\times W_{n+1}$ are zero, which is what this book is about.

(d) Consider the gambling system mentioned in 275Cf. Here the idea is that $W_n=\pm 1$, as in (a), and $Z_{n+1}=2^n a$ if $X_n\le 0$, 0 if $X_n>0$; that is, the gambler doubles his stake each time until he wins, and then quits. Of course he is almost sure to win eventually, so we have $\lim_{n\to\infty}X_n=a$ almost everywhere, even though $\mathbb{E}(X_n)=0$ for every n. We can compute the distribution of X_n: for $n\ge 1$ we have $\Pr(X_n=a)=1-2^{-n}$, $\Pr(X_n=-(2^n-1)a)=2^{-n}$. Thus $\mathbb{E}(|X_n|)=(2-2^{-n+1})a$ and the almost-everywhere convergence of the X_n is an example of Doob's Martingale Convergence Theorem.

In the language of stopping times (275N), $X_n=\tilde Y_{\tau\wedge n}$, where $Y_n=\sum_{k=0}^n 2^k aW_k$ and $\tau=\min\{n:Y_n>0\}$.

***276F** I come now to Komlós' theorem. The first step is a trifling refinement of 276C.

Lemma Let (Ω,Σ,μ) be a probability space, and $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ a non-decreasing sequence of σ-subalgebras of Σ. Suppose that $\langle X_n\rangle_{n\in\mathbb{N}}$ is a sequence of random variables on Ω such that (i) X_n is Σ_n-measurable for each n (ii) $\sum_{n=0}^\infty\frac{1}{(n+1)^2}\mathbb{E}(X_n^2)$ is finite (iii) $\lim_{n\to\infty}X'_n=0$ a.e., where X'_n is a conditional expectation of X_n on Σ_{n-1} for each $n\ge 1$. Then $\lim_{n\to\infty}\frac{1}{n+1}\sum_{k=0}^n X_k=0$ a.e.

proof Making suitable adjustments on a negligible set if necessary, we may suppose that X'_n is Σ_{n-1}-measurable for $n\ge 1$ and that every X_n and X'_n is defined on the whole of Ω. Set $X'_0=X_0$ and $Y_n=X_n-X'_n$ for $n\in\mathbb{N}$. Then $\langle Y_n\rangle_{n\in\mathbb{N}}$ is a martingale difference sequence adapted to $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$. Also $\mathbb{E}(Y_n^2)\le\mathbb{E}(X_n^2)$ for every n. **P** If $n\ge 1$, X'_n is square-integrable (244M), and $\mathbb{E}(Y_n\times X'_n)=0$, as in part (b) of the proof of 276B. Now

$$\mathbb{E}(X_n^2)=\mathbb{E}(Y_n+X'_n)^2=\mathbb{E}(Y_n^2)+2\mathbb{E}(Y_n\times X'_n)+\mathbb{E}(X'_n)^2\ge\mathbb{E}(Y_n^2).\ \mathbf{Q}$$

This means that $\sum_{n=0}^\infty\frac{1}{(n+1)^2}\mathbb{E}(Y_n^2)$ must be finite. By 276C, $\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^n Y_i=0$ a.e. But by 273Ca we also have $\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^n X'_i=0$ whenever $\lim_{n\to\infty}X'_n=0$, which is almost everywhere. So $\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^n X_i=0$ a.e.

***276G Lemma** Let (Ω,Σ,μ) be a probability space, and $\langle X_n\rangle_{n\in\mathbb{N}}$ a sequence of random variables on Ω such that $\sup_{n\in\mathbb{N}}\mathbb{E}(|X_n|)$ is finite. For $k\in\mathbb{N}$ and $x\in\mathbb{R}$ set $F_k(x)=x$ if $|x|\le k$, 0 otherwise. Let $\mathcal{F}$ be an ultrafilter on $\mathbb{N}$.

(a) For each $k\in\mathbb{N}$ there is a measurable function $Y_k:\Omega\to[-k,k]$ such that $\lim_{n\to\mathcal{F}}\int_E F_k(X_n)=\int_E Y_k$ for every $E\in\Sigma$.

(b) $\lim_{n\to\mathcal{F}}\mathbb{E}((F_k(X_n)-Y_k)^2)\le\lim_{n\to\mathcal{F}}\mathbb{E}(F_k(X_n)^2)$ for each k.

(c) $Y=\lim_{k\to\infty}Y_k$ is defined a.e. and $\lim_{k\to\infty}\mathbb{E}(|Y-Y_k|)=0$.

proof (a) For each k, $|F_k(X_n)|\le_{\text{a.e.}}k\chi\Omega$ for every n, so that $\{F_k(X_n):n\in\mathbb{N}\}$ is uniformly integrable, and $\{F_k(X_n)^\bullet:n\in\mathbb{N}\}$ is relatively weakly compact in $L^1=L^1(\mu)$ (247C). Accordingly $v_k=\lim_{n\to\mathcal{F}}F_k(X_n)^\bullet$ is defined in L^1 (2A3Se); take $Y_k:\Omega\to\mathbb{R}$ to be a measurable function such that $Y_k^\bullet=v_k$. For any $E\in\Sigma$,

$$\int_E Y_k=\int v_k\times(\chi E)^\bullet=\lim_{n\to\mathcal{F}}\int_E F_k(X_n).$$

In particular,

$$|\int_E Y_k|\le\sup_{n\in\mathbb{N}}|\int_E F_k(X_n)|\le k\mu E$$

for every E, so that $\{\omega:Y_k(\omega)>k\}$ and $\{\omega:Y_k(\omega)<-k\}$ are both negligible; changing Y_k on a negligible set if necessary, we may suppose that $|Y_k(\omega)|\le k$ for every $\omega\in\Omega$.

(b) Because Y_k is bounded, $Y_k^\bullet\in L^\infty(\mu)$, and

$$\lim_{n\to\mathcal{F}}\int F_k(X_n)\times Y_k=\lim_{n\to\mathcal{F}}\int F_k(X_n)^\bullet\times Y_k^\bullet=\int Y_k^\bullet\times Y_k^\bullet=\int Y_k^2.$$

Accordingly

$$\lim_{n\to\mathcal{F}}\int(F_k(X_n)-Y_k)^2=\lim_{n\to\mathcal{F}}\int F_k(X_n)^2-2\lim_{n\to\mathcal{F}}\int F_k(X_n)\times Y_k+\int Y_k^2$$
$$=\lim_{n\to\mathcal{F}}\int F_k(X_n)^2-\int Y_k^2\le\lim_{n\to\mathcal{F}}\int F_k(X_n)^2.$$

(c) Set $W_0=Y_0=0$, $W_k=Y_k-Y_{k-1}$ for $k\ge 1$. Then $\mathbb{E}(|W_k|)\le\lim_{n\to\mathcal{F}}\mathbb{E}(|F_k(X_n)-F_{k-1}(X_n)|)$ for every $k\ge 1$. **P** Set $E=\{\omega:W_k(\omega)\ge 0\}$. Then

$$\int_E W_k=\int_E Y_k-\int_E Y_{k-1}$$
$$=\lim_{n\to\mathcal{F}}\int_E F_k(X_n)-\lim_{n\to\mathcal{F}}\int_E F_{k-1}(X_n)$$
$$=\lim_{n\to\mathcal{F}}\int_E F_k(X_n)-F_{k-1}(X_n)\le\lim_{n\to\mathcal{F}}\int_E|F_k(X_n)-F_{k-1}(X_n)|.$$

Similarly,

$$|\int_{X\setminus E}W_k|\le\lim_{n\to\mathcal{F}}\int_{X\setminus E}|F_k(X_n)-F_{k-1}(X_n)|.$$

So

$$\mathbb{E}(|W_k|)=\int_E W_k-\int_{X\setminus E}W_k\le\lim_{n\to\mathcal{F}}\int|F_k(X_n)-F_{k-1}(X_n)|.\ \mathbf{Q}$$

It follows that $\sum_{k=0}^{\infty}\mathbb{E}(|W_k|)$ is finite. **P** For any $m\ge 1$,

$$\sum_{k=0}^{m}\mathbb{E}(|W_k|)\le\sum_{k=1}^{m}\lim_{n\to\mathcal{F}}\mathbb{E}(|F_k(X_n)-F_{k-1}(X_n)|)$$
$$=\lim_{n\to\mathcal{F}}\mathbb{E}(\sum_{k=1}^{m}|F_k(X_n)-F_{k-1}(X_n)|)$$
$$=\lim_{n\to\mathcal{F}}\mathbb{E}(|F_m(X_n)|)\le\sup_{n\in\mathbb{N}}\mathbb{E}(|X_n|).$$

So $\sum_{k=0}^{\infty}\mathbb{E}(|W_k|)\le\sup_{n\in\mathbb{N}}\mathbb{E}(|X_n|)$ is finite. **Q**

By B.Levi's theorem (123A), $\lim_{m\to\infty}\sum_{k=0}^{m}|W_k|$ is finite a.e., so that

$$Y=\lim_{m\to\infty}Y_m=\sum_{k=0}^{\infty}W_k$$

is defined a.e.; and moreover

$$\mathbb{E}(|Y-Y_k|)\le\lim_{m\to\infty}\mathbb{E}(\sum_{j=k+1}^{m}|W_j|)\to 0$$

as $k\to\infty$.

***276H Komlós' theorem** (Komlós 67) Let (Ω,Σ,μ) be any measure space, and $\langle X_n\rangle_{n\in\mathbb{N}}$ a sequence of integrable real-valued functions on Ω such that $\sup_{n\in\mathbb{N}}\int|X_n|$ is finite. Then there are a subsequence $\langle X'_n\rangle_{n\in\mathbb{N}}$ of $\langle X_n\rangle_{n\in\mathbb{N}}$ and an integrable function Y such that $Y=_{\text{a.e.}}\lim_{n\to\infty}\frac{1}{n+1}\sum_{i=0}^{n}X''_i$ whenever $\langle X''_n\rangle_{n\in\mathbb{N}}$ is a subsequence of $\langle X'_n\rangle_{n\in\mathbb{N}}$.

proof Since neither the hypothesis nor the conclusion is affected by changing the X_n on a negligible set, we may suppose throughout that every X_n is measurable and defined on the whole of Ω. In addition, to begin with (down to the end of (e) below), let us suppose that $\mu X=1$. As in 276G, set $F_k(x)=x$ for $|x|\le k$, 0 for $|x|>k$.

(a) Let $\mathcal{F}$ be any non-principal ultrafilter on $\mathbb{N}$ (2A1O). For $j\in\mathbb{N}$ set $p_j=\lim_{n\to\mathcal{F}}\Pr(|X_n|>j)$. Then $\sum_{j=0}^{\infty}p_j$ is finite. **P** For any $k\in\mathbb{N}$,

$$\sum_{j=0}^{k}p_j=\sum_{j=0}^{k}\lim_{n\to\mathcal{F}}\Pr(|X_n|>j)=\lim_{n\to\mathcal{F}}\sum_{j=0}^{k}\Pr(|X_n|>j)$$
$$\le\lim_{n\to\mathcal{F}}(1+\int|X_n|)\le 1+\sup_{n\in\mathbb{N}}\int|X_n|.$$

So $\sum_{j=0}^{\infty} p_j \le 1 + \sup_{n\in\mathbb{N}} \int |X_n|$ is finite. **Q**

Setting

$$p'_j = p_j - p_{j+1} = \lim_{n\to\mathcal{F}} \Pr(j < |X_n| \le j+1)$$

for each j, we have

$$\begin{aligned}\sum_{j=0}^{\infty}(j+1)p'_j &= \lim_{m\to\infty}\Bigl(\sum_{j=0}^{m}(j+1)p_j - \sum_{j=0}^{m}(j+1)p_{j+1}\Bigr)\\ &= \lim_{m\to\infty}\sum_{j=0}^{m} p_j - (m+1)p_{m+1} \le \sum_{j=0}^{\infty} p_j < \infty.\end{aligned}$$

Next,

$$\lim_{n\to\mathcal{F}} \int F_k(X_n)^2 \le \sum_{j=0}^{k}(j+1)^2 p'_j$$

for each k. **P** Setting $E_{jn} = \{\omega : j \le |X_n(\omega)| < j+1\}$ for j, $n \in \mathbb{N}$, $F_k(X_n)^2 \le \sum_{j=0}^{k}(j+1)^2\chi E_{jn}$, so

$$\lim_{n\to\mathcal{F}} \int F_k(X_n)^2 \le \lim_{n\to\mathcal{F}} \sum_{j=0}^{k}(j+1)^2\mu E_{jn} = \sum_{j=0}^{k}(j+1)^2 p'_j. \text{ **Q**}$$

(b) Define $\langle Y_k\rangle_{k\in\mathbb{N}}$ and $Y =_{\text{a.e.}} \lim_{k\to\infty} Y_k$ from $\langle X_n\rangle_{n\in\mathbb{N}}$ and $\mathcal{F}$ as in Lemma 276G. Then

$$J_k = \{n : n \in \mathbb{N}, \int (F_k(X_n) - Y_k)^2 \le 1 + \sum_{j=0}^{k}(j+1)^2 p'_j\}$$

belongs to $\mathcal{F}$ for every $k \in \mathbb{N}$. **P** By (a) above and 276Gb,

$$\lim_{n\to\mathcal{F}} \int (F_k(X_n) - Y_k)^2 \le \lim_{n\to\mathcal{F}} \int F_k(X_n)^2 \le \sum_{j=0}^{k}(j+1)^2 p'_j. \text{ **Q**}$$

Also, of course,

$$K_k = \{n : n \in \mathbb{N}, \Pr(F_j(X_n) \ne X_n) \le p_j + 2^{-j} \text{ for every } j \le k\}$$

belongs to $\mathcal{F}$ for every k.

(c) For n, $k \in \mathbb{N}$ let Z_{kn} be a simple function such that $|Z_{kn}| \le |F_k(X_n) - Y_k|$ and $\int |F_k(X_n) - Y_k - Z_{kn}| \le 2^{-k}$. For $m \in \mathbb{N}$ let Σ_m be the algebra of subsets of Ω generated by sets of the form $\{\omega : Z_{kn}(\omega) = \alpha\}$ for k, $n \le m$ and $\alpha \in \mathbb{R}$. Because each Z_{kn} takes only finitely many values, Σ_m is finite (and is therefore a σ-subalgebra of Σ); and of course $\Sigma_m \subseteq \Sigma_{m+1}$ for every m.

We need to look at conditional expectations on the Σ_m, and because Σ_m is always finite these have a particularly straightforward expression. Let $\mathcal{A}_m$ be the set of 'atoms', or minimal non-empty sets, in Σ_m; that is, the set of equivalence classes in Ω under the relation $\omega \sim \omega'$ if $Z_{kn}(\omega) = Z_{kn}(\omega')$ for all k, $n \le m$. For any integrable random variable X on Ω, define $\mathbb{E}_m(X)$ by setting

$$\begin{aligned}\mathbb{E}_m(X)(\omega) &= \frac{1}{\mu A}\int_A X \text{ if } x \in A \in \mathcal{A}_m \text{ and } \mu A > 0,\\ &= 0 \text{ if } x \in A \in \mathcal{A}_m \text{ and } \mu A = 0.\end{aligned}$$

Then $\mathbb{E}_m(X)$ is a conditional expectation of X on Σ_m.

Now

$$\begin{aligned}\lim_{n\to\mathcal{F}} \int |\mathbb{E}_m(F_k(X_n) - Y_k)| &= \lim_{n\to\mathcal{F}} \sum_{A\in\mathcal{A}_m} \int_A |\mathbb{E}_m(F_k(X_n) - Y_k)|\\ &= \lim_{n\to\mathcal{F}} \sum_{A\in\mathcal{A}_m} \Bigl|\int_A \mathbb{E}_m(F_k(X_n) - Y_k)\Bigr|\end{aligned}$$

(because $\mathbb{E}_m(F_k(X_n) - Y_k)$ is constant on each $A \in \mathcal{A}_m$)

$$\begin{aligned}&= \lim_{n\to\mathcal{F}} \sum_{A\in\mathcal{A}_m} \Bigl|\int_A F_k(X_n) - Y_k\Bigr|\\ &= \sum_{A\in\mathcal{A}_m} \lim_{n\to\mathcal{F}} \Bigl|\int_A F_k(X_n) - Y_k\Bigr| = 0\end{aligned}$$

by the choice of Y_k. So if we set

$$I_m = \{n : n \in \mathbb{N},\ \int |\mathbb{E}_m(F_k(X_n) - Y_k)| \le 2^{-k} \text{ for every } k \le m\},$$

then $I_m \in \mathcal{F}$ for every m.

(d) Suppose that $\langle r(n)\rangle_{n\in\mathbb{N}}$ is any strictly increasing sequence in $\mathbb{N}$ such that $r(0) > 0$, $r(n) \in J_n \cap K_n$ for every n and $r(n) \in I_{r(n-1)}$ for $n \ge 1$. Then $\frac{1}{n+1}\sum_{i=0}^n X_{r(i)} \to Y$ a.e. as $n \to \infty$. **P** Express $X_{r(n)}$ as

$$(X_{r(n)} - F_n(X_{r(n)})) + (F_n(X_{r(n)}) - Y_n - Z_{n,r(n)}) + Y_n + Z_{n,r(n)}$$

for each n. Taking these pieces in turn:

(i)

$$\sum_{n=0}^{\infty} \Pr(X_{r(n)} \ne F_n(X_{r(n)})) \le \sum_{n=0}^{\infty} p_n + 2^{-n}$$

(because $r(n) \in K_n$ for every n)

$$< \infty$$

by (a). But this means that $X_{r(n)} - F_n(X_{r(n)}) \to 0$ a.e., since the sequence is eventually zero at almost every point, and $\frac{1}{n+1}\sum_{i=0}^n X_{r(i)} - F_i(X_{r(i)}) \to 0$ a.e. by 273Ca.

(ii) By the choice of the $Z_{n,r(n)}$,

$$\sum_{n=0}^{\infty} \int |F_n(X_{r(n)}) - Y_n - Z_{n,r(n)}| \le \sum_{n=0}^{\infty} 2^{-n}$$

is finite, so $F_n(X_{r(n)}) - Y_n - Z_{n,r(n)} \to 0$ a.e. and $\frac{1}{n+1}\sum_{i=0}^n F_i(X_{r(i)}) - Y_i - Z_{i,r(i)} \to 0$ a.e.

(iii) By 276G, $Y_n \to Y$ a.e. and $\frac{1}{n+1}\sum_{i=0}^n Y_i \to Y$ a.e.

(iv) We know that, for each $n \ge 1$, $r(n) \in I_{r(n-1)}$. So (because $r(n-1) \ge n$) $\int |\mathbb{E}_{r(n-1)}(F_n(X_{r(n)}) - Y_n| \le 2^{-n}$. But as also

$$\int \bigl|\mathbb{E}_{r(n-1)}\bigl(F_n(X_{r(n)}) - Y_n - Z_{n,r(n)}\bigr)\bigr| \le \int |F_n(X_{r(n)}) - Y_n - Z_{n,r(n)}| \le 2^{-n}$$

by 244M and the choice of $Z_{n,r(n)}$,

$$\begin{aligned}\int |\mathbb{E}_{r(n-1)}Z_{n,r(n)}| &= \int \bigl|\mathbb{E}_{r(n-1)}(F_n(X_{r(n)}) - Y_n) - \mathbb{E}_{r(n-1)}(F_n(X_{r(n)}) - Y_n - Z_{n,r(n)})\bigr| \\ &\le 2^{-n+1}\end{aligned}$$

for every n. Accordingly $\mathbb{E}_{r(n-1)}Z_{n,r(n)} \to 0$ a.e.

On the other hand,

$$\begin{aligned}\sum_{n=0}^{\infty} \frac{1}{(n+1)^2} \int Z_{n,r(n)}^2 &\le \sum_{n=0}^{\infty} \frac{1}{(n+1)^2} \int F_n(X_{r(n)} - Y_n)^2 \\ &\le \sum_{n=0}^{\infty} \frac{1}{(n+1)^2}(1 + \sum_{j=0}^{n}(j+1)^2 p'_j)\end{aligned}$$

(because $r(n) \in J_n$)

$$\begin{aligned}&\le \sum_{n=0}^{\infty} \frac{1}{(n+1)^2} + \sum_{j=0}^{\infty}(j+1)^2 p'_j \sum_{n=j}^{\infty} \frac{1}{(n+1)^2} \\ &\le \frac{\pi^2}{6} + 2\sum_{j=0}^{\infty}(j+1)p'_j\end{aligned}$$

is finite. (I am using the estimate

$$\sum_{n=j}^{\infty} \frac{1}{(n+1)^2} \le \sum_{n=j}^{\infty} \frac{2}{n+1} - \frac{2}{n+2} = \frac{2}{j+1}.)$$

By 276F, applied to $\langle \Sigma_{r(n)} \rangle_{n\in\mathbb{N}}$ and $\langle Z_{n,r(n)} \rangle_{n\in\mathbb{N}}$, $\frac{1}{n+1}\sum_{i=0}^{n} Z_{i,r(i)} \to 0$ a.e.

(v) Adding these four components, we see that $\frac{1}{n+1}\sum_{i=0}^{n} X_{r(i)} \to 0$, as claimed. **Q**

(e) Now fix any strictly increasing sequence $\langle s(n) \rangle_{n\in\mathbb{N}}$ in $\mathbb{N}$ such that $s(0) > 0$, $s(n) \in J_n \cap K_n$ for every n and $s(n) \in I_{s(n-1)}$ for $n \ge 1$; such a sequence exists because $J_n \cap K_n \cap I_{s(n-1)}$ belongs to $\mathcal{F}$, so is infinite, for every $n \ge 1$. Set $X'_n = X_{s(n)}$ for every n. If $\langle X''_n \rangle_{n\in\mathbb{N}}$ is a subsequence of $\langle X'_n \rangle_{n\in\mathbb{N}}$, then it is of the form $\langle X_{s(r(n))} \rangle_{n\in\mathbb{N}}$ for some strictly increasing sequence $\langle r(n) \rangle_{n\in\mathbb{N}}$. In this case

$$s(r(0)) \ge s(0) > 0,$$

$$s(r(n)) \in J_{r(n)} \cap K_{r(n)} \subseteq J_n \cap K_n \text{ for every } n,$$

$$s(r(n)) \in I_{s(r(n)-1)} \subseteq I_{s(r(n-1))} \text{ for every } n \ge 1.$$

So (d) tells us that $\frac{1}{n+1}\sum_{i=0}^{n} X''_i \to Y$ a.e.

(f) Thus the theorem is proved in the case in which (Ω, Σ, μ) is a probability space. Now suppose that μ is σ-finite and $\mu\Omega > 0$. In this case there is a strictly positive measurable function $f : \Omega \to \mathbb{R}$ such that $\int f d\mu = 1$ (215B(ix)). Let ν be the corresponding indefinite-integral measure (234J), so that ν is a probability measure on Ω, and $\langle \frac{1}{f} \times X_n \rangle_{n\in\mathbb{N}}$ is a sequence of ν-integrable functions such that $\sup_{n\in\mathbb{N}} \int \frac{1}{f} \times X_n d\nu$ is finite (235K). From (a)-(e) we see that there must be a ν-integrable function Y and a subsequence $\langle X'_n \rangle_{n\in\mathbb{N}}$ of $\langle X_n \rangle_{n\in\mathbb{N}}$ such that $\frac{1}{n+1}\sum_{i=0}^{n} \frac{1}{f} \times X''_i \to Y$ ν-a.e. for every subsequence $\langle X''_n \rangle_{n\in\mathbb{N}}$ of $\langle X'_n \rangle_{n\in\mathbb{N}}$. But μ and ν have the same negligible sets (234Lc), so $\frac{1}{n+1}\sum_{i=0}^{n} X''_i \to f \times Y$ μ-a.e. for every subsequence $\langle X''_n \rangle_{n\in\mathbb{N}}$ of $\langle X'_n \rangle_{n\in\mathbb{N}}$.

(g) Since the result is trivial if $\mu\Omega = 0$, the theorem is true whenever μ is σ-finite. For the general case, set

$$\tilde{\Omega} = \bigcup_{n\in\mathbb{N}} \{\omega : X_n(\omega) \ne 0\} = \bigcup_{m,n\in\mathbb{N}} \{\omega : |X_n(\omega)| \ge 2^{-m}\},$$

so that the subspace measure $\mu_{\tilde{\Omega}}$ is σ-finite. Then there are a $\mu_{\tilde{\Omega}}$-integrable function $\tilde{Y}$ and a subsequence $\langle X'_n \rangle_{n\in\mathbb{N}}$ of $\langle X_n \rangle_{n\in\mathbb{N}}$ such that $\frac{1}{n+1}\sum_{i=0}^{n} X''_i \restriction \tilde{\Omega} \to \tilde{Y}$ $\mu_{\tilde{\Omega}}$-a.e. for every subsequence $\langle X''_n \rangle_{n\in\mathbb{N}}$ of $\langle X'_n \rangle_{n\in\mathbb{N}}$. Setting $Y(\omega) = \tilde{Y}(\omega)$ if $\omega \in \tilde{\Omega}$, 0 for $\omega \in \Omega \setminus \tilde{\Omega}$, we see that Y is μ-integrable and that $\frac{1}{n+1}\sum_{i=0}^{n} X''_i \to Y$ μ-a.e. whenever $\langle X''_n \rangle_{n\in\mathbb{N}}$ is a subsequence of $\langle X'_n \rangle_{n\in\mathbb{N}}$. This completes the proof.

276X Basic exercises >(a) Let $\langle X_n \rangle_{n\in\mathbb{N}}$ be a martingale adapted to a sequence $\langle \Sigma_n \rangle_{n\in\mathbb{N}}$ of σ-algebras. Show that $\int_E X_n^2 \le \int_E X_{n+1}^2$ for every $n \in \mathbb{N}$, $E \in \Sigma_n$ (allowing ∞ as a value of an integral). (*Hint*: see the proof of 276B.)

>(b) Let $\langle X_n \rangle_{n\in\mathbb{N}}$ be a martingale. Show that for any $\epsilon > 0$,

$$\Pr(\sup_{n\in\mathbb{N}} |X_n| \ge \epsilon) \le \frac{1}{\epsilon^2} \sup_{n\in\mathbb{N}} \mathbb{E}(X_n^2).$$

(*Hint*: put 276Xa together with the argument for 275D.)

(c) When does 276Xb give a sharper result than 275Xb?

(d) Let $\langle X_n \rangle_{n\in\mathbb{N}}$ be a martingale difference sequence and set $Y_n = \frac{1}{n+1}(X_0 + \ldots + X_n)$ for each $n \in \mathbb{N}$. Show that if $\langle X_n \rangle_{n\in\mathbb{N}}$ is uniformly integrable then $\lim_{n\to\infty} \|Y_n\|_1 = 0$. (*Hint*: use the argument of 273Na, with 276C in place of 273D, and setting $\tilde{X}_n = X'_n - Z_n$, where Z_n is an appropriate conditional expectation of X'_n.)

>(e) Strong law of large numbers: fifth form A sequence $\langle X_n \rangle_{n\in\mathbb{N}}$ of random variables is **exchangeable** if $(X_{n_0}, \ldots, X_{n_k})$ has the same joint distribution as $(X_0, \ldots, X_k)$ whenever $n_0, \ldots, n_k$ are distinct. Show that if $\langle X_n \rangle_{n\in\mathbb{N}}$ is an exchangeable sequence of random variables with finite expectation, then $\langle \frac{1}{n+1}\sum_{i=0}^{\infty} X_i \rangle_{n\in\mathbb{N}}$ converges a.e. (*Hint*: 276H.)

(f) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be an independent identically distributed sequence of random variables with zero expectation and non-zero finite variance, and $\langle t_n\rangle_{n\in\mathbb{N}}$ a sequence in $\mathbb{R}$. Show that (i) if $\sum_{n=0}^\infty t_n^2<\infty$, then $\sum_{n=0}^\infty t_nX_n$ is defined in $\mathbb{R}$ a.e. (ii) if $\sum_{n=0}^\infty t_n^2=\infty$ then $\sum_{n=0}^\infty t_nX_n$ is undefined a.e. (*Hint*: 276B, 274Xk.)

(g) Suppose that $\langle X_n\rangle_{n\in\mathbb{N}}$ is a uniformly bounded martingale difference sequence and $\langle a_n\rangle_{n\in\mathbb{N}}\in\ell^2$. Show that $\lim_{n\to\infty}\prod_{i=0}^n(1+a_iX_i)$ is defined and finite almost everywhere. (*Hint*: $\langle a_nX_n\rangle_{n\in\mathbb{N}}$ is summable and square-summable a.e.)

276Y Further exercises (a) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale difference sequence such that $\sup_{n\in\mathbb{N}}\|X_n\|_p$ is finite, where $p\in\,]1,\infty[$. Show that $\lim_{n\to\infty}\|\frac{1}{n+1}\sum_{i=0}^n X_i\|_p=0$. (*Hint*: 273Nb.)

(b) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a uniformly integrable martingale difference sequence and Y a bounded random variable. Show that $\lim_{n\to\infty}\mathbb{E}(X_n\times Y)=0$. (Compare 272Ye.)

(c) Use 275Yg to prove 276Xa.

(d) Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a sequence of random variables such that, for some $\delta>0$, $\sup_{n\in\mathbb{N}}n^\delta\mathbb{E}(|X_n|)$ is finite. Set $S_n=\frac{1}{n+1}(X_0+\ldots+X_n)$ for each n. Show that $\lim_{n\to\infty}S_n=0$ a.e. (*Hint*: set $Z_k=2^{-k}(|X_0|+\ldots+|X_{2^k-1}|)$. Show that $\sum_{k=0}^\infty\mathbb{E}(Z_k)<\infty$. Show that $S_n\le 2Z_{k+1}$ if $2^k<n\le 2^{k+1}$.)

(e) Strong law of large numbers: sixth form Let $\langle X_n\rangle_{n\in\mathbb{N}}$ be a martingale difference sequence such that, for some $\delta>0$, $\sup_{n\in\mathbb{N}}\mathbb{E}(|X_n|^{1+\delta})$ is finite. Set $S_n=\frac{1}{n+1}(X_0+\ldots+X_n)$ for each n. Show that $\lim_{n\to\infty}S_n=0$ a.e. (*Hint*: take a non-decreasing sequence $\langle\Sigma_n\rangle_{n\in\mathbb{N}}$ to which $\langle X_n\rangle_{n\in\mathbb{N}}$ is adapted. Set $Y_n=X_n$ when $|X_n|\le n$, 0 otherwise. Let U_n be a conditional expectation of Y_n on Σ_{n-1} and set $Z_n=Y_n-U_n$. Use ideas from 273H, 276C and 276Yd above to show that $\frac{1}{n+1}\sum_{i=0}^n V_i\to 0$ a.e. for $V_i=Z_i$, $V_i=U_i$, $V_i=X_i-Y_i$.)

(f) Show that there is a martingale $\langle X_n\rangle_{n\in\mathbb{N}}$ which converges in measure but is not convergent a.e. (Compare 273Ba.) (*Hint*: arrange that $\{\omega:X_{n+1}(\omega)\ne 0\}=E_n\subseteq\{\omega:|X_{n+1}(\omega)-X_n(\omega)|\ge 1\}$, where $\langle E_n\rangle_{n\in\mathbb{N}}$ is an independent sequence of sets and $\mu E_n=\dfrac{1}{n+1}$ for each n.)

(g) Give an example of an identically distributed martingale difference sequence $\langle X_n\rangle_{n\in\mathbb{N}}$ such that $\langle\frac{1}{n+1}(X_0+\ldots+X_n)\rangle_{n\in\mathbb{N}}$ does not converge to 0 almost everywhere. (*Hint*: start by devising a uniformly bounded sequence $\langle U_n\rangle_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty}\mathbb{E}(|U_n|)=0$ but $\langle\frac{1}{n+1}(U_0+\ldots+U_n)\rangle_{n\in\mathbb{N}}$ does not converge to 0 almost everywhere. Now repeat your construction in such a context that the U_n can be derived from an identically distributed martingale difference sequence by the formulae of 276Ye.)

(h) Construct a proof of Komlós' theorem which does not involve ultrafilters, or any other use of the full axiom of choice, but proceeds throughout by selecting appropriate sub-subsequences. Remember to check that you can prove any fact you use about weakly convergent sequences in L^1 on the same rules.

276 Notes and comments I include two more versions of the strong law of large numbers (276C, 276Ye) not because I have any applications in mind but because I think that if you know the strong law for $\|\,\|_{1+\delta}$-bounded independent sequences, and what a martingale difference sequence is, then there is something missing if you do not know the strong law for $\|\,\|_{1+\delta}$-bounded martingale difference sequences. And then, of course, I have to add 276Yf and 276Yg, lest you be tempted to think that the strong law is 'really' about martingale difference sequences rather than about independent sequences. (Compare 272Yd and 275Yp.)

Komlós' theorem is rather outside the scope of this volume; it is quite hard work and surely much less important, to most probabilists, than many results I have omitted. It does provide a quick proof of 276Xe. However it is relevant to questions arising in some topics treated in Volumes 3 and 4, and the proof fits naturally into this section.

Chapter 28

Fourier analysis

For the last chapter of this volume, I attempt a brief account of one of the most important topics in analysis. This is a bold enterprise, and I cannot hope to satisfy the reasonable demands of anyone who knows and loves the subject as it deserves. But I also cannot pass it by without being false to my own subject, since problems contributed by the study of Fourier series and transforms have led measure theory throughout its history. What I will try to do, therefore, is to give versions of those results which everyone ought to know in language unifying them with the rest of this treatise, aiming to open up a channel for the transfer of intuitions and techniques between the abstract general study of measure spaces, which is the centre of our work, and this particular family of applications of the theory of integration.

I have divided the material of this chapter, conventionally enough, into three parts: Fourier series, Fourier transforms and the characteristic functions of probability theory. While it will be obvious that many ideas are common to all three, I do not think it useful, at this stage, to try to formulate an explicit generalization to unify them; that belongs to a more general theory of harmonic analysis on groups, which must wait until Volume 4. I begin however with a section on the Stone-Weierstrass theorem (§281), which is one of the basic tools of functional analysis, as well as being useful for this chapter. The final section (§286), a proof of Carleson's theorem, is at a rather different level from the rest.

281 The Stone-Weierstrass theorem

Before we begin work on the real subject of this chapter, it will be helpful to have a reasonably general statement of a fundamental theorem on the approximation of continuous functions. In fact I give a variety of forms (281A, 281E, 281F and 281G, together with 281Ya, 281Yd and 281Yg), all of which are sometimes useful. I end the section with a version of Weyl's Equidistribution Theorem (281M-281N).

281A Stone-Weierstrass theorem: first form Let X be a topological space and K a compact subset of X. Write $C_b(X)$ for the space of all bounded continuous real-valued functions on X, so that $C_b(X)$ is a linear space over $\mathbb{R}$. Let $A \subseteq C_b(X)$ be such that

A is a linear subspace of $C_b(X)$;

$|f| \in A$ for every $f \in A$;

$\chi X \in A$;

whenever x, y are distinct points of K there is an $f \in A$ such that $f(x) \neq f(y)$.

Then for every continuous $h : K \to \mathbb{R}$ and $\epsilon > 0$ there is an $f \in A$ such that

$|f(x) - h(x)| \leq \epsilon$ for every $x \in K$,

if $K \neq \emptyset$, $\inf_{x \in X} f(x) \geq \inf_{x \in K} h(x)$ and $\sup_{x \in X} f(x) \leq \sup_{x \in K} h(x)$.

Remark I have stated this theorem in its natural context, that of general topological spaces. But if these are unfamiliar to you, you do not in fact need to know what they are. If you read 'let X be a topological space' as 'let X be a subset of $\mathbb{R}^r$' and 'K is a compact subset of X' as 'K is a subset of X which is closed and bounded in $\mathbb{R}^r$', you will have enough for all the applications in this chapter. In order to follow the proof, of course, you will need to know a little about compactness in $\mathbb{R}^r$; I have written out the necessary facts in §2A2.

proof (a) If K is empty, then we can take $f = \mathbf{0}$ to be the constant function with value 0. So henceforth let us suppose that $K \neq \emptyset$.

(b) The first point to note is that if f, $g \in A$ then $f \wedge g$ and $f \vee g$ belong to A, where

$$(f \wedge g)(x) = \min(f(x), g(x)), \quad (f \vee g)(x) = \max(f(x), g(x))$$

for every $x \in X$; this is because

$$f \wedge g = \frac{1}{2}(f + g - |f - g|), \quad f \vee g = \frac{1}{2}(f + g + |f - g|).$$

It follows by induction on n that $f_0 \wedge \ldots \wedge f_n$ and $f_0 \vee \ldots \vee f_n$ belong to A for all $f_0, \ldots , f_n \in A$.

(c) If x, y are distinct points of K, and a, $b \in \mathbb{R}$, there is an $f \in A$ such that $f(x) = a$ and $f(y) = b$. **P** Start from $g \in A$ such that $g(x) \neq g(y)$; this is the point at which we use the last of the list of four hypotheses on A. Set

$$\alpha = \frac{a-b}{g(x)-g(y)}, \quad \beta = \frac{bg(x)-ag(y)}{g(x)-g(y)}, \quad f = \alpha g + \beta \chi X \in A. \ \mathbf{Q}$$

(d) (The heart of the proof lies in the next two paragraphs.) Let $h : K \to [0, \infty[$ be a continuous function and x any point of K. For any $\epsilon > 0$, there is an $f \in A$ such that $f(x) = h(x)$ and $f(y) \leq h(y) + \epsilon$ for every $y \in K$. **P** Let $\mathcal{G}_x$ be the family of those open sets $G \subseteq X$ for which there is some $f \in A$ such that $f(x) = h(x)$ and $f(w) \leq h(w) + \epsilon$ for every $w \in K \cap G$. I claim that $K \subseteq \bigcup \mathcal{G}_x$. To see this, take any $y \in K$. By (c), there is an $f \in A$ such that $f(x) = h(x)$ and $f(y) = h(y)$. Now $h - f\restriction K : K \to \mathbb{R}$ is a continuous function, taking the value 0 at y, so there is an open subset G of X, containing y, such that $(h - f\restriction K)(w) \geq -\epsilon$ for every $w \in G \cap K$, that is, $f(w) \leq h(w) + \epsilon$ for every $w \in G \cap K$. Thus $G \in \mathcal{G}_x$ and $y \in \bigcup \mathcal{G}_x$, as required.

Because K is compact, $\mathcal{G}_x$ has a finite subcover $G_0, \dots, G_n$ say. For each $i \leq n$, take $f_i \in A$ such that $f_i(x) = h(x)$ and $f_i(w) \leq h(w) + \epsilon$ for every $w \in G_i \cap K$. Then

$$f = f_0 \wedge f_1 \wedge \ldots \wedge f_n \in A,$$

by (b), and evidently $f(x) = h(x)$, while if $y \in K$ there is some $i \leq n$ such that $y \in G_i$, so that

$$f(y) \leq f_i(y) \leq h(y) + \epsilon. \quad \mathbf{Q}$$

(e) If $h : K \to \mathbb{R}$ is any continuous function and $\epsilon > 0$, there is an $f \in A$ such that $|f(y) - h(y)| \leq \epsilon$ for every $y \in K$. **P** This time, let $\mathcal{G}$ be the set of those open subsets G of X for which there is some $f \in A$ such that $f(y) \leq h(y) + \epsilon$ for every $y \in K$ and $f(x) \geq h(x) - \epsilon$ for every $x \in G \cap K$. Once again, $\mathcal{G}$ is an open cover of K. To see this, take any $x \in K$. By (d), there is an $f \in A$ such that $f(x) = h(x)$ and $f(y) \leq h(y) + \epsilon$ for every $y \in K$. Now $h - f\restriction K : K \to \mathbb{R}$ is a continuous function which is zero at x, so there is an open subset G of X, containing x, such that $(h - f\restriction K)(w) \leq \epsilon$ for every $w \in G \cap K$, that is, $f(w) \geq h(w) - \epsilon$ for every $w \in G \cap K$. Thus $G \in \mathcal{G}$ and $x \in \bigcup \mathcal{G}$, as required.

Because K is compact, $\mathcal{G}$ has a finite subcover $G_0, \dots, G_m$ say. For each $j \leq m$, take $f_j \in A$ such that $f_j(y) \leq h(y) + \epsilon$ for every $y \in K$ and $f_j(w) \geq h(w) - \epsilon$ for every $w \in G_j \cap K$. Then

$$f = f_0 \vee f_1 \vee \ldots \vee f_m \in A,$$

by (b), and evidently $f(y) \leq h(y) + \epsilon$ for every $y \in K$, while if $x \in K$ there is some $j \leq m$ such that $x \in G_j$, so that

$$f(x) \geq f_j(x) \geq h(x) - \epsilon.$$

Thus $|f(x) - h(x)| \leq \epsilon$ for every $x \in K$, as required. **Q**

(f) Thus we have an f satisfying the first of the two requirements of the theorem. But for the second, set $M_0 = \inf_{x \in K} h(x)$ and $M_1 = \sup_{x \in K} h(x)$, and

$$f_1 = (M_0 \chi X) \vee (f \wedge M_1 \chi X);$$

f_1 satisfies the second condition as well as the first. (I am tacitly assuming here what is in fact the case, that M_0 and M_1 are finite; this is because K is compact – see 2A2G or 2A3N.)

281B We need some simple tools, belonging to the basic theory of normed spaces; but I hope they will be accessible even if you have not encountered 'normed spaces' before, if you keep a finger at the beginning of §2A4 as you read the next lemma.

Lemma Let X be any set. Write $\ell^\infty(X)$ for the set of bounded functions from X to $\mathbb{R}$. For $f \in \ell^\infty(X)$, set

$$\|f\|_\infty = \sup_{x \in X} |f(x)|,$$

counting the supremum as 0 if X is empty. Then

(a) $\ell^\infty(X)$ is a normed space.

(b) Let $A \subseteq \ell^\infty(X)$ be a subset and $\overline{A}$ its closure (2A3D).

(i) If A is a linear subspace of $\ell^\infty(X)$, so is $\overline{A}$.

(ii) If $f \times g \in A$ whenever $f, g \in A$, then $f \times g \in \overline{A}$ whenever $f, g \in \overline{A}$.

(iii) If $|f| \in A$ whenever $f \in A$, then $|f| \in \overline{A}$ whenever $f \in \overline{A}$.

proof (a) This is a routine verification. To confirm that $\ell^\infty(X)$ is a linear space over $\mathbb{R}$, we have to check that $f + g$, cf belong to $\ell^\infty(X)$ whenever $f, g \in \ell^\infty(X)$ and $c \in \mathbb{R}$; simultaneously we can confirm that $\| \, \|_\infty$ is a norm on $\ell^\infty(X)$ by observing that

$$|(f + g)(x)| \leq |f(x)| + |g(x)| \leq \|f\|_\infty + \|g\|_\infty,$$

$$|cf(x)| = |c||f(x)| \leq |c|\|f\|_\infty$$

whenever $f, g \in \ell^\infty(X)$ and $c \in \mathbb{R}$. It is worth noting at the same time that if $f, g \in \ell^\infty(X)$, then

$$|(f \times g)(x)| = |f(x)||g(x)| \leq \|f\|_\infty \|g\|_\infty$$

for every $x \in X$, so that $\|f \times g\|_\infty \le \|f\|_\infty \|g\|_\infty$.

(Of course all these remarks are very elementary special cases of parts of §243; see 243Xl.)

(b) Recall that

$$\overline{A} = \{f : f \in \ell^\infty(X), \forall\, \epsilon > 0\ \exists\, f_1 \in A,\ \|f - f_1\|_\infty \le \epsilon\}$$

(2A3Kb). Take f, $g \in \overline{A}$ and $c \in \mathbb{R}$, and let $\epsilon > 0$. Set

$$\eta = \min(1, \tfrac{\epsilon}{2+|c|+\|f\|_\infty+\|g\|_\infty}) > 0.$$

Then there are f_1, $g_1 \in \overline{A}$ such that $\|f - f_1\|_\infty \le \eta$, $\|g - g_1\|_\infty \le \eta$.

Now

$$\|(f + g) - (f_1 + g_1)\|_\infty \le \|f - f_1\|_\infty + \|g - g_1\|_\infty \le 2\eta \le \epsilon,$$

$$\|cf - cf_1\|_\infty = |c|\|f - f_1\|_\infty \le |c|\eta \le \epsilon,$$

$$\begin{aligned}\|(f \times g) - (f_1 \times g_1)\|_\infty &= \|(f - f_1) \times g + f \times (g - g_1) - (f - f_1) \times (g - g_1)\|_\infty \\ &\le \|(f - f_1) \times g\|_\infty + \|f \times (g - g_1)\|_\infty + \|(f - f_1) \times (g - g_1)\|_\infty \\ &\le \|f - f_1\|_\infty \|g\|_\infty + \|f\|_\infty \|g - g_1)\|_\infty + \|f - f_1\|_\infty \|g - g_1\|_\infty \\ &\le \eta(\|g\|_\infty + \|f\|_\infty + \eta) \le \eta(\|g\|_\infty + \|f\|_\infty + 1) \le \epsilon,\end{aligned}$$

$$\||f| - |f_1|\|_\infty \le \|f - f_1\|_\infty \le \eta \le \epsilon.$$

(i) If A is a linear subspace, then $f_1 + g_1$ and cf_1 belong to A. As ϵ is arbitrary, $f + g$ and cf belong to $\overline{A}$. As f, g and c are arbitrary, $\overline{A}$ is a linear subspace of $\ell^\infty(X)$.

(ii) If A is closed under multiplication, then $f_1 \times g_1 \in A$. As ϵ is arbitrary, $f \times g \in \overline{A}$.

(iii) If the absolute values of functions in A belong to A, then $|f_1| \in A$. As ϵ is arbitrary, $|f| \in \overline{A}$.

281C Lemma There is a sequence $\langle p_n\rangle_{n\in\mathbb{N}}$ of real polynomials such that $\lim_{n\to\infty} p_n(x) = |x|$ uniformly for $x \in [-1, 1]$.

proof (a) By the Binomial Theorem we have

$$(1 - x)^{1/2} = 1 - \frac{1}{2}x - \frac{1}{4\cdot 2!}x^2 - \frac{1\cdot 3}{2^3\cdot 3!}x^3 - \ldots = -\sum_{n=0}^\infty \frac{(2n)!}{(2n-1)(2^n n!)^2} x^n$$

whenever $|x| < 1$, with the convergence being uniform on any interval $[-a, a]$ with $0 \le a < 1$. (For a proof of this, see almost any book on real or complex analysis. If you have no favourite text to hand, you can try to construct a proof from the following facts: (i) the radius of convergence of the series is 1, so on any interval $[-a, a]$, with $0 \le a < 1$, it is uniformly absolutely summable (ii) writing $f(x)$ for the sum of the series for $|x| < 1$, use Lebesgue's Dominated Convergence Theorem to find expressions for the indefinite integrals $\int_0^x f$, $-\int_{-x}^0 f$ and show that these are $\frac{2}{3}(1 - (1 - x)f(x))$, $\frac{2}{3}(1 - (1 + x)f(-x))$ for $0 \le x < 1$ (iii) use the Fundamental Theorem of Calculus to show that $f(x) + 2(1 - x)f'(x) = 0$ (iv) show that $\frac{d}{dx}\big(\frac{f(x)^2}{1-x}\big) = 0$ and hence (v) that $f(x)^2 = 1 - x$ whenever $|x| < 1$. Finally, show that because f is continuous and non-zero in $]-1, 1[$, $f(x)$ must be the positive square root of $1 - x$ throughout.)

We have a further fragment of information. If we set

$$q_0(x) = 1, \quad q_1(x) = 1 - \frac{1}{2}x, \quad q_n(x) = -\sum_{k=0}^n \frac{(2k)!}{(2k-1)(2^k k!)^2} x^k$$

for $n \ge 2$ and $x \in [0, 1]$, so that q_n is the nth partial sum of the binomial series for $(1 - x)^{1/2}$, then we have $\lim_{n\to\infty} q_n(x) = (1 - x)^{1/2}$ for every $x \in [0, 1[$. But also every q_n is non-increasing on $[0, 1]$, and $\langle q_n(x)\rangle_{n\in\mathbb{N}}$ is a non-increasing sequence for each $x \in [0, 1]$. So we must have

$$\sqrt{1 - x} \le q_n(x)\ \forall\, n \in \mathbb{N},\ x \in [0, 1[,$$

and therefore, because all the q_n are continuous,

$$\sqrt{1 - x} \le q_n(x)\ \forall\, n \in \mathbb{N},\ x \in [0, 1].$$

Moreover, given $\epsilon > 0$, set $a = 1 - \frac{1}{4}\epsilon^2$, so that $\sqrt{1-a} = \frac{\epsilon}{2}$. Then there is an $n_0 \in \mathbb{N}$ such that $q_n(x) - \sqrt{1-x} \le \frac{\epsilon}{2}$ for every $x \in [0, a]$ and $n \ge n_0$. In particular, $q_n(a) \le \epsilon$, so $q_n(x) \le \epsilon$ and $q_n(x) - \sqrt{1-x} \le \epsilon$ for every $x \in [a, 1]$, $n \ge n_0$. This means that

$$0 \le q_n(x) - \sqrt{1-x} \le \epsilon \ \forall\, n \ge n_0,\ x \in [0,1];$$

as ϵ is arbitrary, $\langle q_n(x)\rangle_{n\in\mathbb{N}} \to \sqrt{1-x}$ uniformly on $[0, 1]$.

(b) Now set $p_n(x) = q_n(1 - x^2)$ for $x \in \mathbb{R}$. Because each q_n is a real polynomial of degree n, each p_n is a real polynomial of degree $2n$. Next,

$$\begin{aligned}\sup_{|x|\le 1} |p_n(x) - |x|| &= \sup_{|x|\le 1} |q_n(1-x^2) - \sqrt{1-(1-x^2)}| \\ &= \sup_{y\in[0,1]} |q_n(y) - \sqrt{1-y}| \to 0\end{aligned}$$

as $n \to \infty$, so $\lim_{n\to\infty} p_n(x) = |x|$ uniformly for $|x| \le 1$, as required.

281D Corollary Let X be a set, and A a norm-closed linear subspace of $\ell^\infty(X)$ containing χX and such that $f \times g \in A$ whenever f, $g \in A$. Then $|f| \in A$ for every $f \in A$.

proof Set

$$f_1 = \frac{1}{1+\|f\|_\infty} f,$$

so that $f_1 \in A$ and $\|f_1\|_\infty \le 1$. Because A contains χX and is closed under multiplication, $p\circ f_1 \in A$ for every polynomial p with real coefficients. In particular, $g_n = p_n \circ f_1 \in A$ for every n, where $\langle p_n\rangle_{n\in\mathbb{N}}$ is the sequence of 281C. Now, because $|f_1(x)| \le 1$ for every $x \in X$,

$$\|g_n - |f_1|\|_\infty = \sup_{x\in X} |p_n(f_1(x)) - |f_1(x)|| \le \sup_{|y|\le 1} |p_n(y) - |y|| \to 0$$

as $n \to \infty$. Because A is $\| \|_\infty$-closed, $|f_1| \in A$; consequently $|f| \in A$, as claimed.

281E Stone-Weierstrass theorem: second form Let X be a topological space and K a compact subset of X. Write $C_b(X)$ for the space of all bounded continuous real-valued functions on X. Let $A \subseteq C_b(X)$ be such that

A is a linear subspace of $C_b(X)$;

$f \times g \in A$ for every f, $g \in A$;

$\chi X \in A$;

whenever x, y are distinct points of K there is an $f \in A$ such that $f(x) \ne f(y)$.

Then for every continuous $h : K \to \mathbb{R}$ and $\epsilon > 0$ there is an $f \in A$ such that

$|f(x) - h(x)| \le \epsilon$ for every $x \in K$,

if $K \ne \emptyset$, $\inf_{x\in X} f(x) \ge \inf_{x\in K} h(x)$ and $\sup_{x\in X} f(x) \le \sup_{x\in K} h(x)$.

proof Let $\overline{A}$ be the $\| \|_\infty$-closure of A in $\ell^\infty(X)$. It is helpful to know that $\overline{A} \subseteq C_b(X)$; this is because the uniform limit of continuous functions is continuous. (But if this is new to you, or your memory has faded, don't take time to look it up now; just read '$\overline{A}\cap C_b(X)$' in place of '$\overline{A}$' in the rest of this argument.) By 281B-281D, $\overline{A}$ is a linear subspace of $C_b(X)$ and $|f| \in \overline{A}$ for every $f \in \overline{A}$, so the conditions of 281A apply to $\overline{A}$.

Take a continuous $h : K \to \mathbb{R}$ and an $\epsilon > 0$. The cases in which $K = \emptyset$ or h is constant are trivial, because all constant functions belong to A; so I suppose that $M_0 = \inf_{x\in K} h(x)$ and $M_1 = \sup_{x\in K} h(x)$ are defined and distinct. As observed at the end of the proof of 281A, M_0 and M_1 are finite. Set

$$\eta = \min(\tfrac{1}{3}\epsilon, \tfrac{1}{2}(M_1 - M_2)) > 0, \quad \tilde{h}(x) = \operatorname{med}(M_0 + \eta, h(x), M_1 - \eta) \text{ for } x \in K$$

(definition: 2A1Ac), so that $\tilde{h} : K \to \mathbb{R}$ is continuous and $M_0 + \eta \le \tilde{h}(x) \le M_1 - \eta$ for every $x \in K$. By 281A, there is an $f_0 \in \overline{A}$ such that $|f_0(x) - \tilde{h}(x)| \le \eta$ for every $x \in K$ and $M_0 + \eta \le f_0(x) \le M_1 - \eta$ for every $x \in X$. Now there is an $f \in A$ such that $\|f - f_0\|_\infty \le \eta$, so that

$$|f(x) - h(x)| \le |f(x) - f_0(x)| + |f_0(x) - \tilde{h}(x)| + |\tilde{h}(x) - h(x)| \le 3\eta \le \epsilon$$

for every $x \in K$, while

$$M_0 \le f_0(x) - \eta \le f(x) \le f_0(x) + \eta \le M_1$$

for every $x \in X$.

281F Corollary: Weierstrass' theorem Let K be any closed bounded subset of $\mathbb{R}$. Then every continuous $h : K \to \mathbb{R}$ can be uniformly approximated on K by polynomials.

proof Apply 281E with $X = K$ (noting that K, being closed and bounded, is compact), and A the set of polynomials with real coefficients, regarded as functions from K to $\mathbb{R}$.

281G Stone-Weierstrass theorem: third form Let X be a topological space and K a compact subset of X. Write $C_b(X;\mathbb{C})$ for the space of all bounded continuous complex-valued functions on X, so that $C_b(X;\mathbb{C})$ is a linear space over $\mathbb{C}$. Let $A \subseteq C_b(X;\mathbb{C})$ be such that

A is a linear subspace of $C_b(X;\mathbb{C})$;

$f \times g \in A$ for every f, $g \in A$;

$\chi X \in A$;

the complex conjugate $\bar{f}$ of f belongs to A for every $f \in A$;

whenever x, y are distinct points of K there is an $f \in A$ such that $f(x) \neq f(y)$.

Then for every continuous $h : K \to \mathbb{C}$ and $\epsilon > 0$ there is an $f \in A$ such that

$|f(x) - h(x)| \leq \epsilon$ for every $x \in K$,

if $K \neq \emptyset$, $\sup_{x\in X} |f(x)| \leq \sup_{x\in K} |h(x)|$.

proof If $K = \emptyset$, or h is identically zero, we can take $f = \mathbf{0}$. So let us suppose that $M = \sup_{x\in K} |h(x)| > 0$.

(a) Set

$$A_{\mathbb{R}} = \{f : f \in A,\ f(x) \text{ is real for every } x \in X\}.$$

Then $A_{\mathbb{R}}$ satisfies the conditions of 281E. **P** (i) Evidently $A_{\mathbb{R}}$ is a subset of $C_b(X) = C_b(X;\mathbb{R})$, is closed under addition, multiplication by real scalars and pointwise multiplication of functions, and contains χX. If x, y are distinct points of K, there is an $f \in A$ such that $f(x) \neq f(y)$. Now

$$\mathcal{R}\mathrm{e}\, f = \frac{1}{2}(f + \bar{f}), \quad \mathcal{I}\mathrm{m}\, f = \frac{1}{2i}(f - \bar{f})$$

both belong to A and are real-valued, so belong to $A_{\mathbb{R}}$, and at least one of them takes different values at x and y. **Q**

(b) Consequently, given a continuous function $h : K \to \mathbb{C}$ and $\epsilon > 0$, we may apply 281E twice to find f_1, $f_2 \in A_{\mathbb{R}}$ such that

$$|f_1(x) - \mathcal{R}\mathrm{e}(h(x))| \leq \eta, \quad |f_2(x) - \mathcal{I}\mathrm{m}(h(x))| \leq \eta$$

for every $x \in K$, where $\eta = \min(\frac{1}{2}, \frac{\epsilon M}{6M+4}) > 0$. Setting $g = f_1 + if_2$, we have $g \in A$ and $|g(x) - h(x)| \leq 2\eta$ for every $x \in K$.

(c) Set $L = \|g\|_\infty$. If $L \leq M$ we can take $f = g$ and stop. Otherwise, consider the function

$$\phi(t) = \frac{M-\eta}{\max(M,\sqrt{t})}$$

for $t \in [0, L^2]$. By Weierstrass' theorem (281F), there is a real polynomial p such that $|\phi(t) - p(t)| \leq \frac{\eta}{L}$ whenever $0 \leq t \leq L^2$. Note that $|g|^2 = g \times \bar{g} \in A$, so that

$$f = g \times p(|g|^2) \in A.$$

Now

$$|p(t)| \leq \phi(t) + \frac{\eta}{L} \leq \phi(t) + \frac{\eta}{\max(M,\sqrt{t})} = \frac{M}{\max(M,\sqrt{t})}$$

whenever $0 \leq t \leq L^2$, so

$$|f(x)| \leq |g(x)|\frac{M}{\max(M,|g(x)|)} \leq M$$

for every $x \in X$. Next, if $0 \leq t \leq \min(L, M + 2\eta)^2$,

$$|1 - p(t)| \leq \frac{\eta}{L} + 1 - \phi(t) \leq \frac{\eta}{M} + 1 - \frac{M-\eta}{M+2\eta} \leq \frac{4\eta}{M}.$$

Consequently, if $x \in K$, so that

$$|g(x)| \leq \min(L, |h(x)| + 2\eta) \leq \min(L, M + 2\eta),$$

we shall have

$$|1 - p(|g(x)|^2)| \le \frac{4\eta}{M},$$

and

$$\begin{aligned}|f(x) - h(x)| &\le |g(x) - h(x)| + |g(x)||1 - p(|g(x)|^2)| \\ &\le 2\eta + \frac{4\eta}{M}(M + 2\eta) \le 2\eta + \frac{4\eta}{M}(M + 1) \le \epsilon,\end{aligned}$$

as required.

Remark Of course we could have saved ourselves effort by settling for

$$\sup_{x\in X} |f(x)| \le 2\sup_{x\in K} |h(x)|,$$

which would be quite good enough for the applications below.

281H Corollary Let $[a, b] \subseteq \mathbb{R}$ be a non-empty bounded closed interval and $h : [a, b] \to \mathbb{C}$ a continuous function. Then for any $\epsilon > 0$ there are $y_0, \dots , y_n \in \mathbb{R}$ and $c_0, \dots , c_n \in \mathbb{C}$ such that

$$|h(x) - \sum_{k=0}^{n} c_k e^{iy_k x}| \le \epsilon \text{ for every } x \in [a, b],$$

$$\sup_{x\in\mathbb{R}} |\sum_{k=0}^{n} c_k e^{iy_k x}| \le \sup_{x\in[a,b]} |h(x)|.$$

proof Apply 281G with $X = \mathbb{R}$, $K = [a, b]$ and A the linear span of the functions $x \mapsto e^{iyx}$ as y runs over $\mathbb{R}$.

281I Corollary Let S^1 be the unit circle $\{z : |z| = 1\} \subseteq \mathbb{C}$. Then for any continuous function $h : S^1 \to \mathbb{C}$ and $\epsilon > 0$, there are $n \in \mathbb{N}$ and $c_{-n}, c_{-n+1}, \dots , c_0, \dots , c_n \in \mathbb{C}$ such that $|h(z) - \sum_{k=-n}^{n} c_k z^k| \le \epsilon$ for every $z \in S^1$.

proof Apply 281G with $X = K = S^1$ and A the linear span of the functions $z \mapsto z^k$ for $k \in \mathbb{Z}$.

281J Corollary Let $h : [-\pi, \pi] \to \mathbb{C}$ be a continuous function such that $h(\pi) = h(-\pi)$. Then for any $\epsilon > 0$ there are $n \in \mathbb{N}$, $c_{-n}, \dots , c_n \in \mathbb{C}$ such that $|h(x) - \sum_{k=-n}^{n} c_k e^{ikx}| \le \epsilon$ for every $x \in [-\pi, \pi]$.

proof The point is that $\tilde{h} : S^1 \to \mathbb{C}$ is continuous on S^1, where $\tilde{h}(z) = h(\arg z)$; this is because arg is continuous everywhere except at -1, and

$$\lim_{x\downarrow-\pi} h(x) = h(-\pi) = h(\pi) = \lim_{x\uparrow\pi} h(x),$$

so

$$\lim_{z\in S^1, z\to-1} \tilde{h}(z) = h(\pi) = \tilde{h}(-1).$$

Now by 281H there are $c_{-n}, \dots , c_n \in \mathbb{C}$ such that $|\tilde{h}(z) - \sum_{k=-n}^{n} c_k z^k| \le \epsilon$ for every $z \in S^1$, and these coefficients serve equally for h.

281K Corollary Suppose that $r \ge 1$ and that $K \subseteq \mathbb{R}^r$ is a non-empty closed bounded set. Let $h : K \to \mathbb{C}$ be a continuous function, and $\epsilon > 0$. Then there are $y_0, \dots , y_n \in \mathbb{Q}^r$ and $c_0, \dots , c_n \in \mathbb{C}$ such that

$$|h(x) - \sum_{k=0}^{n} c_k e^{iy_k \centerdot x}| \le \epsilon \text{ for every } x \in K,$$

$$\sup_{x\in\mathbb{R}^r} |\sum_{k=0}^{n} c_k e^{iy_k \centerdot x}| \le \sup_{x\in K} |h(x)|,$$

writing $y\centerdot x = \sum_{j=1}^{r} \eta_j \xi_j$ when $y = (\eta_1, \dots , \eta_r)$ and $x = (\xi_1, \dots , \xi_r)$ belong to $\mathbb{R}^r$.

proof Apply 281G with $X = \mathbb{R}^r$ and A the linear span of the functions $x \mapsto e^{iy\centerdot x}$ as y runs over $\mathbb{Q}^r$.

281L Corollary Suppose that $r \ge 1$ and that $K \subseteq \mathbb{R}^r$ is a non-empty closed bounded set. Let $h : K \to \mathbb{R}$ be a continuous function, and $\epsilon > 0$. Then there are $y_0, \dots , y_n \in \mathbb{R}^r$ and $c_0, \dots , c_n \in \mathbb{C}$ such that, writing $g(x) = \sum_{k=0}^{n} c_k e^{iy_k \centerdot x}$, g is real-valued and

$$|h(x) - g(x)| \le \epsilon \text{ for every } x \in K,$$

$$\inf_{y\in K} h(y) \le g(x) \le \sup_{y\in K} h(y) \text{ for every } x \in \mathbb{R}^r.$$

proof Apply 281E with $X = \mathbb{R}^r$ and A the set of *real*-valued functions on $\mathbb{R}^r$ which are *complex* linear combinations of the functions $x \mapsto e^{iy\centerdot x}$; as remarked in part (a) of the proof of 281G, A satisfies the conditions of 281E.

281M Weyl's Equidistribution Theorem We are now ready for one of the basic results of number theory. I shall actually apply it to provide an example in §285 below, but (at least in the one-variable case) it is surely on the (rather long) list of things which every pure mathematician should know. For the sake of the application I have in mind, I give the full r-dimensional version, but you may wish to take it in the first place with $r = 1$.

It will be helpful to have a notation for 'fractional part'. For any real number x, write $<x>$ for that number in $[0,1[$ such that $x - <x>$ is an integer. Now for the theorem.

281N Theorem Let $\eta_1, \dots, \eta_r$ be real numbers such that $1, \eta_1, \dots, \eta_r$ are linearly independent over $\mathbb{Q}$. Then whenever $0 \le \alpha_j \le \beta_j \le 1$ for each $j \le r$,

$$\lim_{n\to\infty} \frac{1}{n+1} \#(\{m : m \le n,\ <m\eta_j> \in [\alpha_j, \beta_j] \text{ for every } j \le r\}) = \prod_{j=1}^r (\beta_j - \alpha_j).$$

Remark Thus the theorem says that the long-term proportion of the r-tuples $(<m\eta_1>, \dots, <m\eta_r>)$ which belong to the interval $[a, b] \subseteq [\mathbf{0}, \mathbf{1}]$ is just the Lebesgue measure $\mu[a, b]$ of the interval. Of course the condition '$\eta_1, \dots, \eta_r$ are linearly independent over $\mathbb{Q}$' is necessary as well as sufficient (281Xg).

proof (a) Write $y = (\eta_1, \dots, \eta_r) \in \mathbb{R}^r$,

$$<my> = (<m\eta_1>, \dots, <m\eta_r>) \in [\mathbf{0}, \mathbf{1}[= [0,1[^r$$

for each $m \in \mathbb{N}$. Set $I = [\mathbf{0}, \mathbf{1}] = [0,1]^r$, and for any function $f : I \to \mathbb{R}$ write

$$\overline{L}(f) = \limsup_{n\to\infty} \frac{1}{n+1} \sum_{m=0}^n f(<my>),$$

$$\underline{L}(f) = \liminf_{n\to\infty} \frac{1}{n+1} \sum_{m=0}^n f(<my>);$$

and for $f : I \to \mathbb{C}$ write

$$L(f) = \lim_{n\to\infty} \frac{1}{n+1} \sum_{m=0}^n f(<my>)$$

if the limit exists. It will be worth noting that for non-negative functions f, g, $h : I \to \mathbb{R}$ such that $h \le f + g$,

$$\overline{L}(h) \le \overline{L}(f) + \overline{L}(g),$$

and that $L(cf + g) = cL(f) + L(g)$ for any two functions f, $g : I \to \mathbb{C}$ such that $L(f)$ and $L(g)$ exist, and any $c \in \mathbb{C}$.

(b) I mean to show that $L(f)$ exists and is equal to $\int_I f$ for (many) continuous functions f. The key step is to consider functions of the form

$$f(x) = e^{2\pi i k \centerdot x},$$

where $k = (\kappa_1, \dots, \kappa_r) \in \mathbb{Z}^r$. In this case, if $k \ne \mathbf{0}$,

$$k \centerdot y = \sum_{j=1}^r \kappa_j \eta_j \notin \mathbb{Z}$$

because $1, \eta_1, \dots, \eta_r$ are linearly independent over $\mathbb{Q}$. So

$$L(f) = \lim_{n\to\infty} \frac{1}{n+1} \sum_{m=0}^n e^{2\pi i k \centerdot <my>} = \lim_{n\to\infty} \frac{1}{n+1} \sum_{m=0}^n e^{2\pi i m k \centerdot y}$$

(because $mk \centerdot y - k \centerdot <my> = \sum_{j=1}^r \kappa_j (m\eta_j - <m\eta_j>)$ is an integer)

$$= \lim_{n\to\infty} \frac{1 - e^{2\pi i (n+1) k \centerdot y}}{(n+1)(1 - e^{2\pi i k \centerdot y})}$$

(because $e^{2\pi i k \centerdot y} \ne 1$)

$$= 0,$$

because $|1 - e^{2\pi i (n+1) k \centerdot y}| \le 2$ for every n. Of course we can also calculate the integral of f over I, which is

$$\int_I f(x) dx = \int_I e^{2\pi i k \centerdot x} dx = \int_I \prod_{j=1}^r e^{2\pi i \kappa_j \xi_j} dx$$

(writing $x = (\xi_1, \dots, \xi_r)$)

$$= \int_0^1 \dots \int_0^1 \prod_{j=1}^r e^{2\pi i\kappa_j\xi_j} d\xi_r \dots d\xi_1$$
$$= \int_0^1 e^{2\pi i\kappa_r\xi_r} d\xi_r \dots \int_0^1 e^{2\pi i\kappa_1\xi_1} d\xi_1 = 0$$

because at least one κ_j is non-zero, and for this j we must have

$$\int_0^1 e^{2\pi i\kappa_j\xi_j} d\xi_j = \frac{1}{2\pi i\kappa_j}(e^{2\pi i\kappa_j} - 1) = 0.$$

So we have $L(f) = \int_I f = 0$ when $k \neq \mathbf{0}$. On the other hand, if $k = \mathbf{0}$, then f is constant with value 1, so

$$L(f) = \lim_{n\to\infty} \frac{1}{n+1}\sum_{m=0}^n f(<my>) = \lim_{n\to\infty} 1 = 1 = \int_I f(x)dx.$$

(c) Now write $\partial I = [\mathbf{0}, \mathbf{1}] \setminus]\mathbf{0}, \mathbf{1}[$, the boundary of I. If $f : I \to \mathbb{C}$ is continuous and $f(x) = 0$ for $x \in \partial I$, then $L(f) = \int_I f$. **P** As in 281I, let S^1 be the unit circle $\{z : z \in \mathbb{C}, |z| = 1\}$, and set $K = (S^1)^r \subseteq \mathbb{C}^r$. If we think of K as a subset of $\mathbb{R}^{2r}$, it is closed and bounded. Let $\phi : K \to I$ be given by

$$\phi(\zeta_1, \dots, \zeta_r) = (\frac{1}{2} + \frac{\arg\zeta_1}{2\pi}, \dots, \frac{1}{2} + \frac{\arg\zeta_r}{2\pi})$$

for $\zeta_1, \dots, \zeta_r \in S^1$. Then $h = f\phi : K \to \mathbb{C}$ is continuous, because ϕ is continuous on $(S^1 \setminus \{-1\})^r$ and

$$\lim_{w\to z} f\phi(w) = f\phi(z) = 0$$

for any $z \in K \setminus (S^1 \setminus \{-1\})^r$. (Compare 281J.) Now apply 281G with $X = K$ and A the set of polynomials in $\zeta_1, \dots, \zeta_r, \zeta_1^{-1}, \dots, \zeta_r^{-1}$ to see that, given $\epsilon > 0$, there is a function of the form

$$g(z) = \sum_{k\in J} c_k \zeta_1^{\kappa_1} \dots \zeta_r^{\kappa_r},$$

for some finite set $J \subseteq \mathbb{Z}^r$ and constants $c_k \in \mathbb{C}$ for $k \in J$, such that

$$|g(z) - h(z)| \le \epsilon \text{ for every } z \in K.$$

Set

$$\tilde{g}(x) = g(e^{\pi i(2\xi_1-1)}, \dots, e^{\pi i(2\xi_r-1)}) = \sum_{k\in J} c_k e^{\pi i k\,.\,(2x-\mathbf{1})} = \sum_{k\in J} (-1)^{k.\mathbf{1}} c_k e^{2\pi i k\,.\,x},$$

so that $\tilde{g}\phi = g$, and see that

$$\sup_{x\in I} |\tilde{g}(x) - f(x)| = \sup_{z\in K} |g(z) - h(z)| \le \epsilon.$$

Now $\tilde{g}$ is of the form dealt with in (a), so we must have $L(\tilde{g}) = \int_I \tilde{g}$. Let n_0 be such that

$$\left|\int_I \tilde{g} - \frac{1}{n+1}\sum_{m=0}^n \tilde{g}(<my>)\right| \le \epsilon$$

for every $n \ge n_0$. Then

$$|\int_I f - \int_I \tilde{g}| \le \int_I |f - \tilde{g}| \le \epsilon$$

and

$$|\frac{1}{n+1}\sum_{m=0}^n \tilde{g}(<my>) - \frac{1}{n+1}\sum_{m=0}^n f(<my>)| \le \frac{1}{n+1}\sum_{m=0}^n |\tilde{g}(<my>) - f(<my>)|$$
$$\le \frac{1}{n+1}(n+1)\epsilon = \epsilon$$

for every $n \in \mathbb{N}$. So for $n \ge n_0$ we must have

$$|\frac{1}{n+1}\sum_{m=0}^n f(<my>) - \int_I f| \le 3\epsilon.$$

As ϵ is arbitrary, $L(f) = \int_I f$, as required. **Q**

(d) Observe next that if $a, b \in]\mathbf{0}, \mathbf{1}[=]0, 1[^r$, and $\epsilon > 0$, there are continuous functions f_1, f_2 such that

$$f_1 \le \chi[a, b] \le f_2 \le \chi\,]\mathbf{0}, \mathbf{1}[, \quad \int_I f_2 - \int_I f_1 \le \epsilon.$$

P This is elementary. For $n \in \mathbb{N}$, define $h_n : \mathbb{R} \to [0,1]$ by setting $h_n(\xi) = 0$ if $\xi \le 0$, $2^n\xi$ if $0 \le \xi \le 2^{-n}$ and 1 if $\xi \ge 2^{-n}$. Set

$$f_{1n}(x) = \prod_{j=1}^r h_n(\xi_j - \alpha_j)h_n(\beta_j - \xi_j),$$

$$f_{2n}(x) = \prod_{j=1}^r (1 - h_n(\alpha_j - \xi_j))(1 - h_n(\xi_j - \beta_j))$$

for $x = (\xi_1, \dots, \xi_r) \in \mathbb{R}^r$. (Compare the proof of 242Oa.) Then $f_{1n} \le \chi[a,b] \le f_{2n}$ for each n, $f_{2n} \le \chi\,]\mathbf{0},\mathbf{1}[$ for all n so large that

$$2^{-n} \le \min(\min_{j\le r}\alpha_j, \min_{j\le r}(1-\beta_j)),$$

and $\lim_{n\to\infty} f_{2n}(x) - f_{1n}(x) = 0$ for every x, so

$$\lim_{n\to\infty} \int_I f_{2n} - \int_I f_{1n} = 0.$$

Thus we can take $f_1 = f_{1n}$, $f_2 = f_{2n}$ for any n large enough. **Q**

(e) It follows that if a, $b \in]\mathbf{0},\mathbf{1}[$ and $a \le b$, $L(\chi[a,b]) = \mu[a,b]$. **P** Let $\epsilon > 0$. Take f_1, f_2 as in (d). Then, using (c),

$$\overline{L}(\chi[a,b]) \le \overline{L}(f_2) = L(f_2) = \int_I f_2 \le \int_I f_1 + \epsilon \le \mu[a,b] + \epsilon,$$

$$\underline{L}(\chi[a,b]) \ge \underline{L}(f_1) = L(f_1) = \int_I f_1 \ge \int_I f_2 - \epsilon \ge \mu[a,b] - \epsilon,$$

so

$$\mu[a,b] - \epsilon \le \underline{L}(\chi[a,b]) \le \overline{L}(\chi[a,b]) \le \mu[a,b] + \epsilon.$$

As ϵ is arbitrary,

$$\mu[a,b] = \overline{L}(\chi[a,b]) = \underline{L}(\chi[a,b]) = L(\chi[a,b]),$$

as required. **Q**

(f) To complete the proof, take any a, $b \in I$ with $a \le b$. For $0 \le \epsilon \le \frac{1}{2}$, set $I_\epsilon = [\epsilon\mathbf{1}, (1-\epsilon)\mathbf{1}]$, so that I_ϵ is a closed interval included in $]\mathbf{0},\mathbf{1}[$ and $\mu I_\epsilon = (1-2\epsilon)^r$. Of course $L(\chi I) = \mu I = 1$, so

$$L(\chi(I \setminus I_\epsilon)) = L(\chi I) - L(\chi I_\epsilon) = 1 - \mu I_\epsilon,$$

and

$$\begin{aligned} \mu[a,b] - 1 + \mu I_\epsilon &\le \mu[a,b] + \mu I_\epsilon - \mu([a,b] \cup I_\epsilon) = \mu([a,b] \cap I_\epsilon) \\ &= L(\chi([a,b] \cap I_\epsilon)) \le \underline{L}(\chi([a,b])) \\ &\le \overline{L}(\chi([a,b])) \le \overline{L}(\chi([a,b] \cap I_\epsilon)) + \overline{L}(\chi(I \setminus I_\epsilon)) \\ &= L(\chi([a,b] \cap I_\epsilon)) + 1 - \mu I_\epsilon \\ &= \mu([a,b] \cap I_\epsilon) + 1 - \mu I_\epsilon \le \mu[a,b] + 1 - \mu I_\epsilon. \end{aligned}$$

As ϵ is arbitrary,

$$\mu[a,b] = \overline{L}(\chi[a,b]) = \underline{L}(\chi[a,b]) = L(\chi[a,b]),$$

as stated.

281X Basic exercises (a) Let A be the set of those bounded continuous functions $f : \mathbb{R}^r \times \mathbb{R}^r \to \mathbb{R}$ which are expressible in the form $f(x,y) = \sum_{k=0}^n g_k(x)g'_k(y)$, where all the g_k, g'_k are continuous functions from $\mathbb{R}^r$ to $\mathbb{R}$. Show that for any bounded continuous function $h : \mathbb{R}^r \times \mathbb{R}^r \to \mathbb{R}$ and any bounded set $K \subseteq \mathbb{R}^r \times \mathbb{R}^r$ and any $\epsilon > 0$, there is an $f \in A$ such that $|f(x,y) - h(x,y)| \le \epsilon$ for every $(x,y) \in K$ and $\sup_{x,y\in\mathbb{R}^r} |f(x,y)| \le \sup_{x,y\in\mathbb{R}^r} |h(x,y)|$.

(b) Let K be a closed bounded set in $\mathbb{R}^r$, where $r \ge 1$, and $h : K \to \mathbb{R}$ a continuous function. Show that for any $\epsilon > 0$ there is a polynomial p in r variables such that $|h(x) - p(x)| \le \epsilon$ for every $x \in K$.

>(c) Let $[a,b]$ be a non-empty closed interval of $\mathbb{R}$ and $h : [a,b] \to \mathbb{R}$ a continuous function. Show that for any $\epsilon > 0$ there are $y_0, \dots, y_n, a_0, \dots, a_n, b_0, \dots, b_n \in \mathbb{R}$ such that

$$|h(x) - \sum_{k=0}^n (a_k \cos y_k x + b_k \sin y_k x)| \le \epsilon \text{ for every } x \in [a,b],$$

$$\sup_{x\in\mathbb{R}} |\sum_{k=0}^n (a_k \cos y_k x + b_k \sin y_k x)| \le \sup_{x\in[a,b]} |h(x)|.$$

(d) Let h be a complex-valued function on $]-\pi,\pi]$ such that $|h|^p$ is integrable, where $1 \le p < \infty$. Show that for every $\epsilon > 0$ there is a function of the form $x \mapsto f(x) = \sum_{k=-n}^{n} c_k e^{ikx}$, where $c_{-k},\dots,c_k \in \mathbb{C}$, such that $\int_{-\pi}^{\pi} |h-f|^p \le \epsilon$. (Compare 244H.)

>(e) Let $h : [-\pi,\pi] \to \mathbb{R}$ be a continuous function such that $h(\pi) = h(-\pi)$, and $\epsilon > 0$. Show that there are $a_0,\dots,a_n,b_1,\dots,b_n \in \mathbb{R}$ such that

$$|h(x) - \frac{1}{2}a_0 - \sum_{k=1}^{n}(a_k \cos kx + b_k \sin kx)| \le \epsilon$$

for every $x \in [-\pi,\pi]$.

(f) Let K be a non-empty closed bounded set in $\mathbb{R}^r$, where $r \ge 1$, and $h : K \to \mathbb{R}$ a continuous function. Show that for any $\epsilon > 0$ there are $y_0,\dots,y_n \in \mathbb{R}^r$, $a_0,\dots,a_n,b_0,\dots,b_n \in \mathbb{R}$ such that

$$|h(x) - \sum_{k=0}^{n}(a_k \cos y_k . x + b_k \sin y_k . x)| \le \epsilon \text{ for every } x \in K,$$

$$\sup_{x\in\mathbb{R}} |\sum_{k=0}^{n}(a_k \cos y_k . x + b_k \sin y_k . x)| \le \sup_{x\in K} |h(x)|,$$

interpreting $y . x$ as in 281K.

(g) Let $y_1,\dots,y_r$ be real numbers which are not linearly independent over $\mathbb{Q}$. Show that there is a non-trivial interval $[a,b] \subseteq [\mathbf{0},\mathbf{1}] \subseteq \mathbb{R}^r$ such that $(<ky_1>,\dots,<ky_r>) \notin [a,b]$ for every $k \in \mathbb{Z}$.

(h) Let $\eta_1,\dots,\eta_r$ be real numbers such that $1,\eta_1,\dots,\eta_r$ are linearly independent over $\mathbb{Q}$. Suppose that $0 \le \alpha_j \le \beta_j \le 1$ for each $j \le r$. Show that for every $\epsilon > 0$ there is an $n_0 \in \mathbb{N}$ such that

$$|\prod_{j=1}^{r}(\beta_j - \alpha_j) - \frac{1}{n+1}\#(\{m : k \le m \le k+n,\ <m\eta_j> \in [\alpha_j,\beta_j] \text{ for every } j \le r\})| \le \epsilon$$

whenever $n \ge n_0$ and $k \in \mathbb{N}$. (*Hint*: in 281N, set

$$\overline{L}(f) = \limsup_{n\to\infty} \sup_{k\in\mathbb{N}} \frac{1}{n+1}\sum_{m=k}^{k+n} f(<my>).)$$

281Y Further exercises (a) Show that under the hypotheses of 281A, there is an $f \in \overline{A}$, the $\| \|_\infty$-closure of A in $C_b(X)$, such that $f\restriction K = h$. (*Hint*: take $f = \lim_{n\to\infty} f_n$ where

$$\|f_{n+1} - f_n\|_\infty \le \sup_{x\in K} |f_n(x) - h(x)| \le 2^{-n}$$

for every $n \in \mathbb{N}$.)

(b) Let X be a topological space and $K \subseteq X$ a compact subset. Suppose that for any distinct points x, y of K there is a continuous function $f : X \to \mathbb{R}$ such that $f(x) \ne f(y)$. Show that for any $r \in \mathbb{N}$ and any continuous $h : K \to \mathbb{R}^r$ there is a continuous $f : X \to \mathbb{R}^r$ extending h. (*Hint*: consider $r = 1$ first.)

(c) Let $\langle X_i\rangle_{i\in I}$ be any family of compact Hausdorff spaces, and X their product as topological spaces. For each i, write $C(X_i)$ for the set of continuous functions from X_i to $\mathbb{R}$, and $\pi_i : X \to X_i$ for the coordinate map. Show that the subalgebra of $C(X)$ generated by $\{f\pi_i : i \in I,\ f \in C(X_i)\}$ is $\| \|_\infty$-dense in $C(X)$. (*Note*: you will need to know that X is compact, and that if Z is any compact Hausdorff space then for any distinct z, $w \in Z$ there is an $f \in C(Z)$ such that $f(z) \ne f(w)$. For references see 3A3J and 3A3Bf in the next volume.)

(d) Let X be a topological space and K a compact subset of X. Let A be a linear subspace of the space $C_b(X)$ of real-valued continuous functions on X such that $|f| \in A$ for every $f \in A$. Let $h : K \to \mathbb{R}$ be a continuous function such that whenever x, $y \in K$ there is an $f \in A$ such that $f(x) = h(x)$ and $f(y) = h(y)$. Show that for every $\epsilon > 0$ there is an $f \in A$ such that $|f(x) - h(x)| \le \epsilon$ for every $x \in K$.

(e) Let X be a compact topological space and write $C(X)$ for the set of continuous functions from X to $\mathbb{R}$. Suppose that $h \in C(X)$, and let $A \subseteq C(X)$ be such that

A is a linear subspace of $C(X)$;

either $|f| \in A$ for every $f \in A$ *or* $f \times g \in A$ for every f, $g \in A$ *or* $f \times f \in A$ for every $f \in A$;

whenever x, $y \in X$ and $\delta > 0$ there is an $f \in A$ such that $|f(x) - h(x)| \le \delta$, $|f(y) - h(y)| \le \delta$.

Show that for every $\epsilon > 0$ there is an $f \in A$ such that $|h(x) - f(x)| \le \epsilon$ for every $x \in X$.

(f) Let X be a compact topological space and A a $\| \|_\infty$-closed linear subspace of the space $C(X)$ of continuous functions from X to $\mathbb{R}$. Show that the following are equiveridical:

(i) $|f| \in A$ for every $f \in A$;

(ii) $f \times f \in A$ for every $f \in A$;

(iii) $f \times g \in A$ for all $f, g \in A$,

and that in this case A is closed in $C(X)$ for the topology defined by the pseudometrics

$$(f, g) \mapsto |f(x) - g(x)| : C(X) \times C(X) \to [0, \infty[$$

as x runs over X (the 'topology of pointwise convergence' on $C(X)$).

(g) Show that under the hypotheses of 281G there is an $f \in \overline{A}$, the $\| \|_\infty$-closure of A in $C_b(X; \mathbb{C})$, such that $f \restriction K = h$ and (if $K \neq \emptyset$) $\|f\|_\infty = \sup_{x \in K} |h(x)|$.

(h) Let $y \in \mathbb{R}$ be irrational. Show that for any Riemann integrable function $f : [0, 1] \to \mathbb{R}$,

$$\int_0^1 f(x)dx = \lim_{n\to\infty} \frac{1}{n+1} \sum_{m=0}^n f(<my>),$$

writing $<my>$ for the fractional part of my. (*Hint*: recall *Riemann's criterion:* for any $\epsilon > 0$, there are $a_0, \ldots, a_n$ with $0 = a_0 \le a_1 \le \ldots \le a_n = 1$ and

$$\sum\{a_j - a_{j-1} : j \le n, \sup_{x\in[a_{j-1},a_j]} f(x) - \inf_{x\in[a_{j-1},a_j]} f(x) \ge \epsilon\} \le \epsilon.)$$

(i) Let $\langle t_n\rangle_{n\in\mathbb{N}}$ be a sequence in $[0, 1]$. Show that the following are equiveridical: (i) $\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^n f(t_k) = \int_0^1 f$ for every continuous function $f : [0, 1] \to \mathbb{R}$; (ii) $\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^n f(t_k) = \int_0^1 f$ for every Riemann integrable function $f : [0, 1] \to \mathbb{R}$; (iii) $\liminf_{n\to\infty} \frac{1}{n+1} \#(\{k : k \le n, t_k \in G\}) \ge \mu G$ for every open set $G \subseteq [0, 1]$; (iv) $\lim_{n\to\infty} \frac{1}{n+1} \#(\{k : k \le n, t_k \le \alpha\}) = \alpha$ for every $\alpha \in [0, 1]$; (v) $\lim_{n\to\infty} \frac{1}{n+1} \#(\{k : k \le n, t_k \in E\}) = \mu E$ for every $E \subseteq [0, 1]$ such that $\mu(\operatorname{int} E) = \mu\overline{E}$ (vi) $\lim_{n\to\infty} \frac{1}{n+1} \sum_{k=0}^n e^{2\pi imt_k} = 0$ for every $m \ge 1$. (Cf. 273J. Such sequences $\langle t_n\rangle_{n\in\mathbb{N}}$ are called **equidistributed** or **uniformly distributed**.)

(j) Show that the sequence $\langle < \ln(n + 1)>\rangle_{n\in\mathbb{N}}$ is not equidistributed.

(k) Give $[0, 1]^{\mathbb{N}}$ its product measure λ. Show that λ-almost every sequence $\langle t_n\rangle_{n\in\mathbb{N}} \in [0, 1]^{\mathbb{N}}$ is equidistributed in the sense of 281Yi. (*Hint*: 273J.)

(l) Let $f : [0, 1]^2 \to \mathbb{C}$ be a continuous function. Show that if $\gamma \in \mathbb{R}$ is irrational then $\int_{[0,1]^2} f = \lim_{a\to\infty} \frac{1}{a} \int_0^a f(<t>, <\gamma t>)dt$. (*Hint*: consider first functions of the form $x \mapsto e^{2\pi ik\,.\,x}$.)

(m) A sequence $\langle t_n\rangle_{n\in\mathbb{N}}$ in $[0, 1]$ is **well-distributed** if $\liminf_{n\to\infty} \inf_{l\in\mathbb{N}} \frac{1}{n+1} \#(\{k : l \le k \le l + n, t_k \in G\}) \ge \mu G$ for every open set $G \subseteq [0, 1]$. (i) Show that $\langle t_n\rangle_{n\in\mathbb{N}}$ is well-distributed iff $\lim_{n\to\infty} \sup_{l\in\mathbb{N}} |\int_0^1 f - \frac{1}{n+1} \sum_{k=l}^{k+n} f(t_k)| = 0$ for every continuous $f : [0, 1] \to \mathbb{R}$. (ii) Show that $\langle <n\alpha>\rangle_{n\in\mathbb{N}}$ is well-distributed for every irrational α.

281 Notes and comments I have given three statements (281A, 281E and 281G) of the Stone-Weierstrass theorem, with an acknowledgement (281F) of Weierstrass' own version, and three further forms (281Ya, 281Yd, 281Yg) in the exercises. Yet another will appear in §4A6 in Volume 4. Faced with such a multiplicity, you may wish to try your own hand at writing out theorems which will cover some or all of these versions. I myself see no way of doing it without setting up a confusing list of alternative hypotheses and conclusions. At which point, I ask 'what is a theorem, anyway?', and answer, it is a stopping-place on our journey; it is a place where we can rest, and congratulate ourselves on our achievement; it is a place which we can learn to recognise, and use as a starting point for new adventures; it is a place we can describe, and share with others. For some theorems, like Fermat's last theorem, there is a canonical statement, an exactly locatable point. For others, like the Stone-Weierstrass theorem here, we reach a mass of closely related results, all depending on some arrangement of the arguments laid out in 281A-281G and 281Ya (which introduces a new idea), and all useful in different ways. I suppose, indeed, that most authors would prefer the versions 281Ya and 281Yg, which eliminate the variable ϵ which appears in 281A, 281E and 281G, at the expense of taking a closed subspace A. But I find that the corollaries which will be useful later (281H-281L) are more naturally expressed in terms of linear subspaces which are not closed.

The applications of the theorem, or the theorems, or the method – choose your own expression – are legion; only a few of them are here. An apparently innocent one is in 281Xa and, in a different variant, in 281Yc; these are enormously important in their own domains. In this volume the principal application will be to 285L below, depending on 281K, and it is perhaps right to note that there is an alternative approach to this particular result, based on ideas in 282G. But I offer Weyl's equidistribution theorem (281M-281N) as evidence that we can expect to find good use for these ideas in almost any branch of mathematics.

282 Fourier series

Out of the enormous theory of Fourier series, I extract a few results which may at least provide a foundation for further study. I give the definitions of Fourier and Fejér sums (282A), with five of the most important results concerning their convergence (282G, 282H, 282J, 282L, 282O). On the way I include the Riemann-Lebesgue lemma (282E). I end by mentioning convolutions (282Q).

282A Definition Let f be an integrable complex-valued function defined almost everywhere in $]-\pi,\pi]$.

(a) The **Fourier coefficients** of f are the complex numbers

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-ikx}dx$$

for $k \in \mathbb{Z}$.

(b) The **Fourier sums** of f are the functions

$$s_n(x) = \sum_{k=-n}^{n} c_k e^{ikx}$$

for $x \in \,]-\pi,\pi]$, $n \in \mathbb{N}$.

(c) The **Fourier series** of f is the series $\sum_{k=-\infty}^{\infty} c_k e^{ikx}$, or (because we ordinarily consider the symmetric partial sums s_n) the series $c_0 + \sum_{k=1}^{\infty}(c_k e^{ikx} + c_{-k}e^{-ikx})$.

(d) The **Fejér sums** of f are the functions

$$\sigma_m = \frac{1}{m+1}\sum_{n=0}^{m} s_n$$

for $m \in \mathbb{N}$.

(e) It will be convenient to have a further phrase available. If f is any function with $\operatorname{dom} f \subseteq \,]-\pi,\pi]$, its **periodic extension** is the function $\tilde{f}$, with domain $\bigcup_{k\in\mathbb{Z}}(\operatorname{dom} f + 2k\pi)$, such that $\tilde{f}(x) = f(x-2k\pi)$ whenever $k \in \mathbb{Z}$ and $x \in \operatorname{dom} f + 2k\pi$.

282B Remarks I have made two more or less arbitrary choices here.

(a) I have chosen to express Fourier series in their 'complex' form rather than their 'real' form. From the point of view of pure measure theory (and, indeed, from the point of view of the nineteenth-century origins of the subject) there are gains in elegance from directing attention to real functions f and looking at the real coefficients

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos kx\,dx \text{ for } k \in \mathbb{N},$$

$$b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\sin kx\,dx \text{ for } k \geq 1.$$

If we do this we have

$$c_0 = \frac{1}{2}a_0,$$

and for $k \geq 1$ we have

$$c_k = \frac{1}{2}(a_k - ib_k), \quad c_{-k} = \frac{1}{2}(a_k + ib_k), \quad a_k = c_k + c_{-k}, \quad b_k = i(c_k - c_{-k}),$$

so that the Fourier sums become

$$s_n(x) = \frac{1}{2}a_0 + \sum_{k=1}^{n} a_k \cos kx + b_k \sin kx.$$

The advantage of this is that real functions f correspond to real coefficients a_k, b_k, so that it is obvious that if f is real-valued so are its Fourier and Fejér sums. The disadvantages are that we have to use a variety of trigonometric equalities which are rather more complicated than the properties of the complex exponential function which they reflect, and that we are farther away from the natural generalizations to locally compact abelian groups. So both electrical engineers and harmonic analysts tend to prefer the coefficients c_k.

(b) I have taken the functions f to be defined on the interval $]-\pi, \pi]$ rather than on the circle $S^1 = \{z : z \in \mathbb{C}, |z| = 1\}$. There would be advantages in elegance of language in using S^1, though I do not recall often seeing the formula

$$c_k = \int z^k f(z) dz$$

which is the natural translation of $c_k = \frac{1}{2\pi}\int e^{ikx} f(x)dx$ under the substitution $x = \arg z$, $dx = 2\pi\nu(dz)$. However, applications of the theory tend to deal with periodic functions on the real line, so I work with $]-\pi, \pi]$, and accept the fact that its group operation $+_{2\pi}$, writing $x +_{2\pi} y$ for whichever of $x + y$, $x + y + 2\pi$, $x + y - 2\pi$ belongs to $]-\pi, \pi]$, is less familiar than multiplication on S^1.

(c) The remarks in (b) are supposed to remind you of §255.

(d) Observe that if $f =_{\text{a.e.}} g$ then f and g have the same Fourier coefficients, Fourier sums and Fejér sums. This means that we could, if we wished, regard the c_k, s_n and σ_m as associated with a member of $L^1_{\mathbb{C}}$, the space of equivalence classes of integrable functions (§242), rather than as associated with a particular function f. Since however the s_n and σ_m appear as actual functions, and since many of the questions we are interested in refer to their values at particular points, it is more natural to express the theory in terms of integrable functions f rather than in terms of members of $L^1_{\mathbb{C}}$.

282C The problems (a) Under what conditions, and in what senses, do the Fourier and Fejér sums s_n and σ_m of a function f converge to f?

(b) How do the properties of the double-ended sequence $\langle c_k \rangle_{k\in\mathbb{Z}}$ reflect the properties of f, and vice versa?

Remark The theory of Fourier series has been one of the leading topics of analysis for nearly two hundred years, and innumerable further problems have contributed greatly to our understanding. (For instance: can one characterize those sequences $\langle c_k \rangle_{k\in\mathbb{Z}}$ which are the Fourier coefficients of some integrable function?) But in this outline I will concentrate on the question (a) above, with one and a half results (282K, 282Rb) addressing (b), which will give us more than enough material to work on.

While most people would feel that the Fourier sums are somehow closer to what we really want to know, it turns out that the Fejér sums are easier to analyse, and there are advantages in dealing with them first. So while you may wish to look ahead to the statements of 282J, 282L and 282O for an idea of where we are going, the first half of this section will be largely about Fejér sums. Note that in any case in which we know that the Fourier sums converge (which is quite common; see, for instance, the examples in 282Xh and 282Xo), then if we know that the Fejér sums converge to f, we can deduce that the Fourier sums also do, by 273Ca.

The first step is a basic lemma showing that both the Fourier and Fejér sums of a function f can be thought of as convolutions of f with kernels describable in terms of familiar functions.

282D Lemma Let f be a complex-valued function which is integrable over $]-\pi, \pi]$, and

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-ikx}dx, \quad s_n(x) = \sum_{k=-n}^{n} c_k e^{ikx}, \quad \sigma_m(x) = \frac{1}{m+1}\sum_{n=0}^{m} s_n(x)$$

its Fourier coefficients, Fourier sums and Fejér sums. Write $\tilde{f}$ for the periodic extension of f (282Ae). For $m \in \mathbb{N}$, write

$$\psi_m(t) = \frac{1-\cos(m+1)t}{2\pi(m+1)(1-\cos t)}$$

for $0 < |t| \le \pi$. (If you like, you can set $\psi_m(0) = \frac{m+1}{2\pi}$ to make ψ_m continuous on $[-\pi, \pi]$.)

(a) For each $n \in \mathbb{N}$, $x \in \,]-\pi, \pi]$,

$$\begin{aligned} s_n(x) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\frac{\sin(n+\frac{1}{2})(x-t)}{\sin\frac{1}{2}(x-t)}dt \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{f}(x+t)\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}dt \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x -_{2\pi} t)\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}dt, \end{aligned}$$

writing $x -_{2\pi} t$ for whichever of $x - t$, $x - t - 2\pi$, $x - t + 2\pi$ belongs to $]-\pi, \pi]$.

(b) For each $m \in \mathbb{N}$, $x \in \,]-\pi, \pi]$,

$$\begin{aligned} \sigma_m(x) &= \int_{-\pi}^{\pi} \tilde{f}(x+t)\psi_m(t)dt \\ &= \int_0^{\pi} (\tilde{f}(x+t) + \tilde{f}(x-t))\psi_m(t)dt \\ &= \int_{-\pi}^{\pi} f(x -_{2\pi} t)\psi_m(t)dt. \end{aligned}$$

(c) For any $n \in \mathbb{N}$,

$$\frac{1}{2\pi}\int_{-\pi}^{0} \frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}dt = \frac{1}{2\pi}\int_{0}^{\pi} \frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}dt = \frac{1}{2}, \quad \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}dt = 1.$$

(d) For any $m \in \mathbb{N}$,

(i) $0 \le \psi_m(t) \le \dfrac{m+1}{2\pi}$ for every t;

(ii) for any $\delta > 0$, $\lim_{m\to\infty} \psi_m(t) = 0$ uniformly on $\{t : \delta \le |t| \le \pi\}$;

(iii) $\int_{-\pi}^{0} \psi_m = \int_0^{\pi} \psi_m = \dfrac{1}{2}$, $\quad \int_{-\pi}^{\pi} \psi_m = 1$.

proof Really all that these amount to is summing geometric series.

(a) For (a), we have

$$\begin{aligned} \sum_{k=-n}^{n} e^{-ikt} &= \frac{e^{int} - e^{-i(n+1)t}}{1 - e^{-it}} \\ &= \frac{e^{i(n+\frac{1}{2})t} - e^{-i(n+\frac{1}{2})t}}{e^{\frac{1}{2}it} - e^{-\frac{1}{2}it}} = \frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}. \end{aligned}$$

So

$$\begin{aligned} s_n(x) &= \sum_{k=-n}^{n} c_k e^{ikx} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\Bigl(\sum_{k=-n}^{n} e^{ik(x-t)}\Bigr)dt \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)\frac{\sin(n+\frac{1}{2})(x-t)}{\sin\frac{1}{2}(x-t)}dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{f}(t)\frac{\sin(n+\frac{1}{2})(x-t)}{\sin\frac{1}{2}(x-t)}dt \\ &= \frac{1}{2\pi}\int_{-\pi-x}^{\pi-x} \tilde{f}(x+t)\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{f}(x+t)\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}dt \end{aligned}$$

because $\tilde{f}$ and $t \mapsto \frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}$ are periodic with period 2π, so that the integral from $-\pi - x$ to $-\pi$ must be the same as the integral from $\pi - x$ to π.

For the expression in terms of $f(x -_{2\pi} t)$, we have

$$s_n(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{f}(x+t)\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{f}(x-t)\frac{\sin(n+\frac{1}{2})(-t)}{\sin\frac{1}{2}(-t)}dt$$

(substituting $-t$ for t)

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x -_{2\pi} t)\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}dt$$

because (for x, $t \in]-\pi,\pi]$) $f(x -_{2\pi} t) = \tilde{f}(x-t)$ whenever either is defined, and sin is an odd function.

(b) In the same way, we have

$$\begin{aligned}\sum_{n=0}^{m}\sin(n+\frac{1}{2})t &= \mathcal{I}\mathrm{m}\big(\sum_{n=0}^{m} e^{i(n+\frac{1}{2})t}\big) = \mathcal{I}\mathrm{m}\big(e^{\frac{1}{2}it}\sum_{n=0}^{m} e^{int}\big)\\ &= \mathcal{I}\mathrm{m}\big(e^{\frac{1}{2}it}\frac{1-e^{i(m+1)t}}{1-e^{it}}\big) = \mathcal{I}\mathrm{m}\big(\frac{1-e^{i(m+1)t}}{e^{-\frac{1}{2}it}-e^{\frac{1}{2}it}}\big)\\ &= \mathcal{I}\mathrm{m}\big(\frac{1-e^{i(m+1)t}}{-2i\sin\frac{1}{2}t}\big) = \mathcal{I}\mathrm{m}\big(\frac{i(1-e^{i(m+1)t})}{2\sin\frac{1}{2}t}\big)\\ &= \frac{1-\cos(m+1)t}{2\sin\frac{1}{2}t}.\end{aligned}$$

So

$$\textstyle\sum_{n=0}^{m}\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t} = \frac{1-\cos(m+1)t}{2\sin^2\frac{1}{2}t} = \frac{1-\cos(m+1)t}{1-\cos t} = 2\pi(m+1)\psi_m(t).$$

Accordingly,

$$\begin{aligned}\sigma_m(x) &= \frac{1}{m+1}\sum_{n=0}^{m} s_n(x)\\ &= \frac{1}{m+1}\sum_{n=0}^{m}\frac{1}{2\pi}\int_{-\pi}^{\pi}\tilde{f}(x+t)\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}dt\\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\tilde{f}(x+t)\big(\frac{1}{m+1}\sum_{n=0}^{m}\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}\big)dt\\ &= \int_{-\pi}^{\pi}\tilde{f}(x+t)\psi_m(t)dt = \int_{-\pi}^{\pi} f(x -_{2\pi} t)\psi_m(t)dt\end{aligned}$$

as in (a), because cos and ψ_m are even functions. For the same reason,

$$\int_0^{\pi}\tilde{f}(x-t)\psi_m(t)dt = \int_{-\pi}^{0}\tilde{f}(x+t)\psi_m(t)dt,$$

so

$$\sigma_m(x) = \int_0^{\pi}(\tilde{f}(x+t)+\tilde{f}(x-t))\psi_m(t)dt.$$

(c) We need only look at where the formula $\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}$ came from to see that

$$\begin{aligned}\frac{1}{2\pi}\int_I\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}dt &= \frac{1}{2\pi}\int_I\sum_{k=-n}^{n} e^{ikt}dt\\ &= \frac{1}{2\pi}\int_I(1+2\sum_{k=1}^{n}\cos kt)dt = \frac{1}{2}\end{aligned}$$

for both $I=[-\pi,0]$ and $I=[0,\pi]$, because $\int_I\cos kt\,dt = 0$ for every $k\ne 0$.

(d)(i) $\psi_m(t)\ge 0$ for every t because $1-\cos(m+1)t$, $1-\cos t$ are always greater than or equal to 0. For the upper bound, we have, using the constructions in (a) and (b),

$$\big|\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}\big| = \big|\sum_{k=-n}^{n} e^{ikt}\big| \le 2n+1$$

for every n, so

$$\psi_m(t) = \frac{1}{2\pi(m+1)} \sum_{n=0}^{m} \frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}$$
$$\le \frac{1}{2\pi(m+1)} \sum_{n=0}^{m} 2n+1 = \frac{m+1}{2\pi}.$$

(ii) If $\delta \le |t| \le \pi$,

$$\psi_m(t) \le \frac{1}{\pi(m+1)(1-\cos t)} \le \frac{1}{\pi(m+1)(1-\cos\delta)} \to 0$$

as $m \to \infty$.

(iii) also follows from the construction in (b), because

$$\int_I \psi_m = \frac{1}{2\pi(m+1)} \sum_{n=0}^{m} \int_I \frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t} dt = \frac{1}{m+1} \sum_{n=0}^{m} \frac{1}{2} = \frac{1}{2}$$

for both $I = [-\pi, 0]$ and $I = [0, \pi]$, using (c).

Remarks For a discussion of substitution in integrals, if you feel any need to justify the manipulations in part (a) of the proof, see 263I.

The functions

$$t \mapsto \frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}, \quad t \mapsto \frac{1-\cos(m+1)t}{(m+1)(1-\cos t)}$$

are called respectively the **Dirichlet kernel** and the **Fejér kernel**.

I give the formulae in terms of $f(x -_{2\pi} t)$ in (a) and (b) in order to provide a link with the work of 255O.

282E The next step is a vital lemma, with a suitably distinguished name which (you will be glad to know) reflects its importance rather than its difficulty.

The Riemann-Lebesgue lemma Let f be a complex-valued function which is integrable over $\mathbb{R}$. Then

$$\lim_{y\to\infty} \int f(x)e^{-iyx}dx = \lim_{y\to-\infty} \int f(x)e^{-iyx}dx = 0.$$

proof (a) Consider first the case in which $f = \chi\,]a, b[$, where $a < b$. Then

$$|\int f(x)e^{-iyx}dx| = |\int_a^b e^{-iyx}dx| = |\frac{1}{-iy}(e^{-iyb} - e^{-iya})| \le \frac{2}{|y|}$$

if $y \ne 0$. So in this case the result is obvious.

(b) It follows at once that the result is true if f is a step-function with bounded support, that is, if there are $a_0 \le a_1 \ldots \le a_n$ such that f is constant on every interval $]a_{j-1}, a_j[$ and zero outside $[a_0, a_n]$.

(c) Now, for a given integrable f and $\epsilon > 0$, there is a step-function g such that $\int |f - g| \le \epsilon$ (242Oa). So

$$|\int f(x)e^{-iyx}dx - \int g(x)e^{-iyx}dx| \le \int |f(x) - g(x)|dx \le \epsilon$$

for every y, and

$$\limsup_{y\to\infty} |\int f(x)e^{-iyx}dx| \le \epsilon,$$

$$\limsup_{y\to-\infty} |\int f(x)e^{-iyx}dx| \le \epsilon.$$

As ϵ is arbitrary, we have the result.

282F Corollary (a) Let f be a complex-valued function which is integrable over $]-\pi, \pi]$, and $\langle c_k\rangle_{k\in\mathbb{Z}}$ its sequence of Fourier coefficients. Then $\lim_{k\to\infty} c_k = \lim_{k\to-\infty} c_k = 0$.

(b) Let f be a complex-valued function which is integrable over $\mathbb{R}$. Then $\lim_{y\to\infty} \int f(x)\sin yx\,dx = 0$.

proof (a) We need only identify

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-ikx}dx$$

with $\int g(x)e^{-ikx}dx$, where $g(x) = f(x)/2\pi$ for $x \in \operatorname{dom} f$ and 0 for $|x| > \pi$.

(b) This is just because

$$\int f(x)\sin yx\,dx = \tfrac{1}{2i}(\int f(x)e^{iyx}dx - \int f(x)e^{-iyx}dx).$$

282G We are now ready for theorems on the convergence of Fejér sums. I start with an easy one, almost a warming-up exercise.

Theorem Let $f :]-\pi,\pi] \to \mathbb{C}$ be a continuous function such that $\lim_{t\downarrow-\pi} f(t) = f(\pi)$. Then its sequence $\langle\sigma_m\rangle_{m\in\mathbb{N}}$ of Fejér sums converges uniformly to f on $]-\pi,\pi]$.

proof The conditions on f amount just to saying that its periodic extension $\tilde{f}$ is defined and continuous everywhere on $\mathbb{R}$. Consequently it is bounded and uniformly continuous on any bounded interval, in particular, on the interval $[-2\pi, 2\pi]$. Set $K = \sup_{|t|\le 2\pi}|\tilde{f}(t)| = \sup_{t\in]-\pi,\pi]}|f(t)|$. Write

$$\psi_m(t) = \tfrac{1-\cos(m+1)t}{2\pi(m+1)(1-\cos t)}$$

for $m \in \mathbb{N}$, $0 < |t| \le \pi$, as in 282D.

Given $\epsilon > 0$ we can find a $\delta \in]0,\pi]$ such that $|\tilde{f}(x+t) - \tilde{f}(x)| \le \epsilon$ whenever $x \in [-\pi,\pi]$ and $|t| \le \delta$. Next, we can find an $m_0 \in \mathbb{N}$ such that $M_m \le \frac{\epsilon}{4\pi K}$ for every $m \ge m_0$, where $M_m = \sup_{\delta\le|t|\le\pi}\psi_m(t)$ (282D(d-ii)). Now suppose that $m \ge m_0$ and $x \in]-\pi,\pi]$. Set $g(t) = \tilde{f}(x+t) - f(x)$ for $|t| \le \pi$. Then $|g(t)| \le 2K$ for all $t \in [-\pi,\pi]$ and $|g(t)| \le \epsilon$ if $|t| \le \delta$, so

$$\begin{aligned}|\int_{-\pi}^{\pi} g \times \psi_m| &\le \int_{-\pi}^{-\delta}|g| \times \psi_m + \int_{-\delta}^{\delta}|g| \times \psi_m + \int_{\delta}^{\pi}|g| \times \psi_m \\ &\le 2M_mK(\pi-\delta) + \epsilon\int_{-\delta}^{\delta}\psi_m + 2M_mK(\pi-\delta) \\ &\le 4\pi M_mK + \epsilon \le 2\epsilon.\end{aligned}$$

Consequently, using 282Db and 282D(d-iii),

$$|\sigma_m(x) - f(x)| = |\int_{-\pi}^{\pi}(\tilde{f}(x+t) - f(x))\psi_m(t)dt| \le 2\epsilon$$

for every $m \ge m_0$; and this is true for every $x \in]-\pi,\pi]$. As ϵ is arbitrary, $\langle\sigma_m\rangle_{m\in\mathbb{N}}$ converges to f uniformly on $]-\pi,\pi]$.

282H I come now to a theorem describing the behaviour of the Fejér sums of general functions f. The hypothesis of the theorem may take a little bit of digesting; you can get an idea of its intended scope by glancing at Corollary 282I.

Theorem Let f be a complex-valued function which is integrable over $]-\pi,\pi]$, and $\langle\sigma_m\rangle_{m\in\mathbb{N}}$ its sequence of Fejér sums. Suppose that $x \in]-\pi,\pi]$ and $c \in \mathbb{C}$ are such that

$$\lim_{\delta\downarrow 0}\frac{1}{\delta}\int_0^{\delta}|\tilde{f}(x+t) + \tilde{f}(x-t) - 2c|dt = 0,$$

writing $\tilde{f}$ for the periodic extension of f, as usual; then $\lim_{m\to\infty}\sigma_m(x) = c$.

proof Set $\phi(t) = |\tilde{f}(x+t) + \tilde{f}(x-t) - 2c|$ when this is defined, which is almost everywhere, and $\Phi(t) = \int_0^t \phi$, which is defined for every $t \ge 0$, because $\tilde{f}$ is integrable over $]-\pi,\pi]$ and therefore over every bounded interval.

As in 282D, set

$$\psi_m(t) = \tfrac{1-\cos(m+1)t}{2\pi(m+1)(1-\cos t)}$$

for $m \in \mathbb{N}$, $0 < |t| \le \pi$. We have

$$|\sigma_m(x) - c| = |\int_0^\pi (\tilde{f}(x+t) + \tilde{f}(x-t) - 2c)\psi_m(t)dt| \le \int_0^\pi \phi(t)\psi_m$$

by (b) and (d) of 282D.

Let $\epsilon > 0$. By hypothesis, $\lim_{t\downarrow 0} \Phi(t)/t = 0$; let $\delta \in \,]0,\pi]$ be such that $\Phi(t) \le \epsilon t$ for every $t \in [0,\delta]$. Take any $m \ge \pi/\delta$. I break the integral $\int_0^\pi \phi \times \psi_m$ up into three parts.

(i) For the integral from 0 to $1/m$, we have

$$\int_0^{1/m} \phi \times \psi_m \le \int_0^{1/m} \tfrac{m+1}{2\pi}\phi = \tfrac{m+1}{2\pi}\Phi(\tfrac{1}{m}) \le \tfrac{\epsilon(m+1)}{2\pi m} \le \epsilon,$$

because $\psi_m(t) \le \frac{m+1}{2\pi}$ for every t (282D(d-i)).

(ii) For the integral from $1/m$ to δ, we have

$$\int_{1/m}^\delta \phi \times \psi_m \le \frac{1}{2\pi(m+1)}\int_{1/m}^\delta \phi(t)\frac{1}{1-\cos t}dt \le \frac{\pi}{4(m+1)}\int_{1/m}^\delta \frac{\phi(t)}{t^2}dt$$

(because $1 - \cos t \ge \frac{2t^2}{\pi^2}$ for $|t| \le \pi$)

$$= \frac{\pi}{4(m+1)}\Bigl(\frac{\Phi(\delta)}{\delta^2} - \frac{\Phi(\frac{1}{m})}{(\frac{1}{m})^2} + \int_{1/m}^\delta \frac{2\Phi(t)}{t^3}dt\Bigr)$$

(integrating by parts – see 225F)

$$\le \frac{\pi}{4(m+1)}\Bigl(\frac{\epsilon}{\delta} + \int_{1/m}^\delta \frac{2\epsilon}{t^2}dt\Bigr)$$

(because $\Phi(t) \le \epsilon t$ for $0 \le t \le \delta$)

$$\le \frac{\pi}{4(m+1)}\Bigl(\frac{\epsilon}{\delta} + 2\epsilon m\Bigr) \le \frac{\pi\epsilon}{4(m+1)\delta} + \frac{\pi\epsilon}{2} \le \frac{\epsilon}{4} + \frac{\pi\epsilon}{2} \le 2\epsilon.$$

(iii) For the integral from δ to π, we have

$$\int_\delta^\pi \phi \times \psi_m \le \int_\delta^\pi \tfrac{1}{\pi(m+1)(1-\cos\delta)}\phi \to 0 \text{ as } m \to \infty$$

because ϕ is integrable over $[-\pi,\pi]$. There must therefore be an $m_0 \in \mathbb{N}$ such that

$$\int_\delta^\pi \phi \times \psi_m \le \epsilon$$

for every $m \ge m_0$.

Putting these together, we see that

$$\int_0^\pi \phi \times \psi_m \le \epsilon + 2\epsilon + \epsilon = 4\epsilon$$

for every $m \ge \max(m_0, \frac{\pi}{\delta})$. As ϵ is arbitrary, $\lim_{m\to\infty} \sigma_m(x) = c$, as claimed.

282I Corollary Let f be a complex-valued function which is integrable over $]-\pi,\pi]$, and $\langle\sigma_m\rangle_{m\in\mathbb{N}}$ its sequence of Fejér sums.

(a) $f(x) = \lim_{m\to\infty} \sigma_m(x)$ for almost every $x \in \,]-\pi,\pi]$.

(b) $\lim_{m\to\infty} \int_{-\pi}^\pi |f - \sigma_m| = 0$.

(c) If g is another integrable function with the same Fourier coefficients, then $f =_{\text{a.e.}} g$.

(d) If $x \in \,]-\pi,\pi[$ is such that $a = \lim_{t\in\operatorname{dom} f, t\uparrow x} f(t)$ and $b = \lim_{t\in\operatorname{dom} f, t\downarrow x} f(t)$ are both defined in $\mathbb{C}$, then

$$\lim_{m\to\infty} \sigma_m(x) = \tfrac{1}{2}(a+b).$$

(e) If $a = \lim_{t\in\operatorname{dom} f, t\uparrow\pi} f(t)$ and $b = \lim_{t\in\operatorname{dom} f, t\downarrow-\pi} f(t)$ are both defined in $\mathbb{C}$, then

$$\lim_{m\to\infty} \sigma_m(\pi) = \frac{1}{2}(a+b).$$

(f) If f is defined and continuous at $x \in \,]-\pi,\pi[$, then

$$\lim_{m\to\infty} \sigma_m(x) = f(x).$$

(g) If $\tilde{f}$, the periodic extension of f, is defined and continuous at π, then

$$\lim_{m\to\infty} \sigma_m(\pi) = f(\pi).$$

proof (a) We have only to recall that by 223D

$$\begin{aligned}\limsup_{\delta\downarrow 0} \frac{1}{\delta}\int_0^\delta &|f(x+t)+f(x-t)-2f(x)|dt \\ &\le \limsup_{\delta\downarrow 0}\frac{1}{\delta}\Bigl(\int_0^\delta |f(x+t)-f(x)|dt + \int_0^\delta |f(x-t)-f(x)|dt\Bigr) \\ &= \limsup_{\delta\downarrow 0}\frac{1}{\delta}\int_{-\delta}^\delta |f(x+t)-f(x)|dt = 0\end{aligned}$$

for almost every $x \in \,]-\pi,\pi[$.

(b) Next observe that, in the language of 255O,

$$\sigma_m = f * \psi_m,$$

by the last formula in 282Db. Consequently, by 255Od,

$$\|\sigma_m\|_1 \le \|f\|_1 \|\psi_m\|_1,$$

writing $\|\sigma_m\|_1 = \int_{-\pi}^{\pi} |\sigma_m|$. But this means that we have

$$f(x) = \lim_{m\to\infty} \sigma_m(x) \text{ for almost every } x, \quad \limsup_{m\to\infty} \|\sigma_m\|_1 \le \|f\|_1;$$

and it follows from 245H that $\lim_{m\to\infty} \|f-\sigma_m\|_1 = 0$.

(c) If g has the same Fourier coefficients as f, then it has the same Fourier and Fejér sums, so we have

$$g(x) = \lim_{m\to\infty} \sigma_m(x) = f(x)$$

almost everywhere.

(d)-(e) Both of these amount to considering $x \in \,]-\pi,\pi]$ such that

$$\lim_{t\in\operatorname{dom}\tilde{f}, t\uparrow x} \tilde{f}(t) = a, \quad \lim_{t\in\operatorname{dom}\tilde{f}, t\downarrow x} \tilde{f}(t) = b.$$

Setting $c = \frac{1}{2}(a+b)$, $\phi(t) = |\tilde{f}(x+t)+\tilde{f}(x-t)-2c|$ whenever this is defined, we have $\lim_{t\in\operatorname{dom}\phi, t\downarrow 0} \phi(t) = 0$, so surely $\lim_{\delta\downarrow 0} \frac{1}{\delta}\int_0^\delta \phi = 0$, and the theorem applies.

(f)-(g) are special cases of (d) and (e).

282J I now turn to conditions for the convergence of Fourier sums. Probably the easiest result – one which is both striking and satisfying – is the following.

Theorem Let f be a complex-valued function which is square-integrable over $]-\pi,\pi]$. Let $\langle c_k\rangle_{k\in\mathbb{Z}}$ be its Fourier coefficients and $\langle s_n\rangle_{n\in\mathbb{N}}$ its Fourier sums (282A). Then

(i) $\sum_{k=-\infty}^{\infty} |c_k|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |f|^2$,

(ii) $\lim_{n\to\infty} \int_{-\pi}^{\pi} |f-s_n|^2 = 0$.

proof (a) I recall some notation from 244N/244P. Let $\mathcal{L}^2_{\mathbb{C}}$ be the space of square-integrable complex-valued functions on $]-\pi,\pi]$. For g, $h \in \mathcal{L}^2_{\mathbb{C}}$, write

$$(g|h) = \int_{-\pi}^{\pi} g \times \bar{h}, \quad \|g\|_2 = \sqrt{(g|g)}.$$

Recall that $\|g+h\|_2 \le \|g\|_2 + \|h\|_2$ for all g, $h \in \mathcal{L}^2_{\mathbb{C}}$ (244Fb/244Pb). For $k \in \mathbb{Z}$, $x \in \,]-\pi,\pi]$ set $e_k(x) = e^{ikx}$, so that

$$(f|e_k) = \int_{-\pi}^{\pi} f(x)e^{-ikx}dx = 2\pi c_k.$$

Moreover, if $|k| \le n$,

$$(s_n|e_k) = \sum_{j=-n}^{n} c_j \int_{-\pi}^{\pi} e^{ijx}e^{-ikx}dx = 2\pi c_k,$$

because

$$\begin{aligned}\int_{-\pi}^{\pi} e^{ijx}e^{-ikx}dx &= 2\pi \text{ if } j = k,\\ &= 0 \text{ if } j \ne k.\end{aligned}$$

So

$$(f - s_n|e_k) = 0 \text{ whenever } |k| \le n;$$

in particular,

$$(f - s_n|s_n) = \sum_{k=-n}^{n} \bar{c}_k(f - s_n|e_k) = 0$$

for every $n \in \mathbb{N}$.

(b) Fix $\epsilon > 0$. The next element of the proof is the fact that there are $m \in \mathbb{N}$, $a_{-m}, \dots, a_m \in \mathbb{C}$ such that $\|f-h\|_2 \le \epsilon$, where $h = \sum_{k=-m}^{m} a_k e_k$. **P** By 244Hb/244Pb we know that there is a continuous function $g : [-\pi,\pi] \to \mathbb{C}$ such that $\|f-g\|_2 \le \frac{\epsilon}{3}$. Next, modifying g on a suitably short interval $]\pi-\delta,\pi]$, we can find a continuous function $g_1 : [-\pi,\pi] \to \mathbb{C}$ such that $\|g - g_1\|_2 \le \frac{\epsilon}{3}$ and $g_1(-\pi) = g_1(\pi)$. (Set $M = \sup_{x\in[-\pi,\pi]} |g(x)|$, take $\delta \in \,]0,2\pi]$ such that $(2M)^2\delta \le (\epsilon/3)^2$, and set $g_1(\pi - t\delta) = tg(\pi-\delta) + (1-t)g(-\pi)$ for $t \in [0,1]$.) Either by the Stone-Weierstrass theorem (281J), or by 282G above, there are $a_{-m}, \dots, a_m$ such that $|g_1(x) - \sum_{k=-m}^{m} a_k e^{ikx}| \le \frac{\epsilon}{3\sqrt{2\pi}}$ for every $x \in [-\pi,\pi]$; setting $h = \sum_{k=-m}^{m} a_k e_k$, we have $\|g_1 - h\|_2 \le \frac{1}{3}\epsilon$, so that

$$\|f-h\|_2 \le \|f-g\|_2 + \|g-g_1\|_2 + \|g_1-h\|_2 \le \epsilon. \textbf{ Q}$$

(c) Now take any $n \ge m$. Then $s_n - h$ is a linear combination of $e_{-n}, \dots, e_n$, so $(f-s_n|s_n-h) = 0$. Consequently

$$\begin{aligned}\epsilon^2 &\ge (f-h|f-h)\\ &= (f-s_n|f-s_n) + (f-s_n|s_n-h) + (s_n-h|f-s_n) + (s_n-h|s_n-h)\\ &= \|f-s_n\|_2^2 + \|s_n-h\|_2^2 \ge \|f-s_n\|_2^2.\end{aligned}$$

Thus $\|f-s_n\|_2 \le \epsilon$ for every $n \ge m$. As ϵ is arbitrary, $\lim_{n\to\infty} \|f-s_n\|_2^2 = 0$, which proves (ii).

(d) As for (i), we have

$$\sum_{k=-n}^{n} |c_k|^2 = \frac{1}{2\pi}\sum_{k=-n}^{n} \bar{c}_k(s_n|e_k) = \frac{1}{2\pi}(s_n|s_n) = \frac{1}{2\pi}\|s_n\|_2^2.$$

But of course

$$\big|\|s_n\|_2 - \|f\|_2\big| \le \|s_n - f\|_2 \to 0$$

as $n \to \infty$, so

$$\sum_{k=-\infty}^{\infty} |c_k|^2 = \frac{1}{2\pi}\lim_{n\to\infty} \|s_n\|_2^2 = \frac{1}{2\pi}\|f\|_2^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |f|^2,$$

as required.

282K Corollary Let $L^2_{\mathbb{C}}$ be the Hilbert space of equivalence classes of square-integrable complex-valued functions on $]-\pi,\pi]$, with the inner product

$$(f^\bullet|g^\bullet)=\int_{-\pi}^{\pi} f\times\bar g$$

and norm

$$\|f^\bullet\|_2=\bigl(\int_{-\pi}^{\pi}|f|^2\bigr)^{1/2},$$

writing $f^\bullet\in L^2_{\mathbb{C}}$ for the equivalence class of a square-integrable function f. Let $\ell^2_{\mathbb{C}}(\mathbb{Z})$ be the Hilbert space of square-summable double-ended complex sequences, with the inner product

$$(\boldsymbol{c}|\boldsymbol{d})=\sum_{k=-\infty}^{\infty}c_k\bar d_k$$

and norm

$$\|\boldsymbol{c}\|_2=\bigl(\sum_{k=-\infty}^{\infty}|c_k|^2\bigr)^{1/2}$$

for $\boldsymbol{c}=\langle c_k\rangle_{k\in\mathbb{Z}}$, $\boldsymbol{d}=\langle d_k\rangle_{k\in\mathbb{Z}}$ in $\ell^2_{\mathbb{C}}(\mathbb{Z})$. Then we have an inner-product-space isomorphism $S:L^2_{\mathbb{C}}\to\ell^2_{\mathbb{C}}(\mathbb{Z})$ defined by saying that

$$S(f^\bullet)(k)=\tfrac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi}f(x)e^{-ikx}dx$$

for every square-integrable function f and every $k\in\mathbb{Z}$.

proof (a) As in 282J, write $\mathcal{L}^2_{\mathbb{C}}$ for the space of square-integrable functions. If f, $g\in\mathcal{L}^2_{\mathbb{C}}$ and $f^\bullet=g^\bullet$, then $f=_{\text{a.e.}}g$, so

$$\tfrac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi}f(x)e^{-ikx}dx=\tfrac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi}g(x)e^{-ikx}dx$$

for every $k\in\mathbb{N}$. Thus S is well-defined.

(b) S is linear. **P** This is elementary. If f, $g\in\mathcal{L}^2_{\mathbb{C}}$ and $c\in\mathbb{C}$,

$$\begin{aligned}S(f^\bullet+g^\bullet)(k)&=\frac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi}(f(x)+g(x))e^{-ikx}dx\\&=\frac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi}f(x)e^{-ikx}dx+\frac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi}g(x)e^{-ikx}dx\\&=S(f^\bullet)(k)+S(g^\bullet)(k)\end{aligned}$$

for every $k\in\mathbb{Z}$, so that $S(f^\bullet+g^\bullet)=S(f^\bullet)+S(g^\bullet)$. Similarly,

$$S(cf^\bullet)(k)=\tfrac{1}{\sqrt{2\pi}}\int_{-\pi}^{\pi}cf(x)e^{-ikx}dx=\tfrac{c}{\sqrt{2\pi}}\int_{-\pi}^{\pi}f(x)e^{-ikx}dx=cS(f^\bullet)(k)$$

for every $k\in\mathbb{Z}$, so that $S(cf^\bullet)=cS(f^\bullet)$. **Q**

(c) If $f\in\mathcal{L}^2_{\mathbb{C}}$ has Fourier coefficients c_k, then $S(f^\bullet)=\langle c_k\sqrt{2\pi}\rangle_{k\in\mathbb{Z}}$, so by 282J(i)

$$\|S(f^\bullet)\|_2^2=2\pi\sum_{k=-\infty}^{\infty}|c_k|^2=\int_{-\pi}^{\pi}|f|^2=\|f^\bullet\|_2^2.$$

Thus $Su\in\ell^2_{\mathbb{C}}(\mathbb{Z})$ and $\|Su\|_2=\|u\|_2$ for every $u\in L^2_{\mathbb{C}}$. Because S is linear and norm-preserving, it is surely injective.

(d) It now follows that $(Sv|Su)=(v|u)$ for every u, $v\in L^2_{\mathbb{C}}$. **P** (This is of course a standard fact about Hilbert spaces.) We know that for any $t\in\mathbb{R}$

$$\begin{aligned}\|u\|_2^2 + 2\,\mathcal{R}\mathrm{e}(e^{it}(v|u)) + \|v\|_2^2 &= (u|u) + e^{it}(v|u) + e^{-it}(u|v) + (v|v)\\ &= (u+e^{it}v|u+e^{it}v)\\ &= \|u+e^{it}v\|_2^2 = \|S(u+e^{it}v)\|_2^2\\ &= \|Su\|_2^2 + 2\,\mathcal{R}\mathrm{e}(e^{it}(Sv|Su)) + \|Sv\|_2^2\\ &= \|u\|_2^2 + 2\,\mathcal{R}\mathrm{e}(e^{it}(Sv|Su)) + \|v\|_2^2,\end{aligned}$$

so that $\mathcal{R}\mathrm{e}(e^{it}(Sv|Su)) = \mathcal{R}\mathrm{e}(e^{it}(v|u))$. As t is arbitrary, $(Sv|Su) = (v|u)$. **Q**

(e) Finally, S is surjective. **P** Let $\boldsymbol{c} = \langle c_k\rangle_{k\in\mathbb{Z}}$ be any member of $\ell^2_{\mathbb{C}}(\mathbb{Z})$. Set $c_k^{(n)} = c_k$ if $|k| \le n$, 0 otherwise, and $\boldsymbol{c}^{(n)} = \langle c_k^{(n)}\rangle_{k\in\mathbb{N}}$. Consider

$$s_n = \sum_{k=-n}^{n} c_k e_k, \quad u_n = s_n^\bullet$$

where I write $e_k(x) = \frac{1}{\sqrt{2\pi}}e^{ikx}$ for $x \in \,]-\pi,\pi]$. Then $Su_n = \boldsymbol{c}^{(n)}$, by the same calculations as in part (a) of the proof of 282J. Now

$$\|\boldsymbol{c}^{(n)} - \boldsymbol{c}\|_2 = \sqrt{\textstyle\sum_{|k|>n}|c_k|^2} \to 0$$

as $n \to \infty$, so

$$\|u_m - u_n\|_2 = \|\boldsymbol{c}^{(m)} - \boldsymbol{c}^{(n)}\|_2 \to 0$$

as m, $n \to \infty$, and $\langle u_n\rangle_{n\in\mathbb{N}}$ is a Cauchy sequence in $L^2_{\mathbb{C}}$. Because $L^2_{\mathbb{C}}$ is complete (244G/244Pb), $\langle u_n\rangle_{n\in\mathbb{N}}$ has a limit $u \in L^2_{\mathbb{C}}$, and now

$$Su = \lim_{n\to\infty} Su_n = \lim_{n\to\infty} \boldsymbol{c}^{(n)} = \boldsymbol{c}. \ \mathbf{Q}$$

Thus $S : L^2_{\mathbb{C}} \to \ell^2_{\mathbb{C}}(\mathbb{Z})$ is an inner-product-space isomorphism.

Remark In the language of Hilbert spaces, all that is happening here is that $\langle e_k^\bullet\rangle_{k\in\mathbb{Z}}$ is a 'Hilbert space basis' or 'complete orthonormal sequence' in $L^2_{\mathbb{C}}$, which is matched by S with the standard basis of $\ell^2_{\mathbb{C}}(\mathbb{Z})$. The only step which calls on non-trivial real analysis, as opposed to the general theory of Hilbert spaces, is the check that the linear subspace generated by $\{e_k^\bullet : k \in \mathbb{Z}\}$ is dense; this is part (b) of the proof of 282J.

Observe that while $S : L^2 \to \ell^2$ is readily described, its inverse is more of a problem. If $\boldsymbol{c} \in \ell^2$, we should like to say that $S^{-1}\boldsymbol{c}$ is the equivalence class of f, where $f(x) = \frac{1}{\sqrt{2\pi}}\sum_{k=-\infty}^{\infty} c_k e^{ikx}$ for every x. This works very well if $\{k : c_k \ne 0\}$ is finite, but for the general case it is less clear how to interpret the sum. It is in fact the case that if $\boldsymbol{c} \in \ell^2$ then

$$g(x) = \frac{1}{\sqrt{2\pi}} \lim_{n\to\infty} \textstyle\sum_{k=-n}^{n} c_k e^{ikx}$$

is defined for almost every $x \in \,]-\pi,\pi]$, and that $S^{-1}\boldsymbol{c} = g^\bullet$ in L^2; this is, in effect, Carleson's theorem (286V). A proof of Carleson's theorem is out of our reach for the moment. What is covered by the results of this section is that

$$h(x) = \frac{1}{\sqrt{2\pi}} \lim_{m\to\infty} \frac{1}{m+1}\textstyle\sum_{n=0}^{m}\sum_{k=-n}^{n} c_k e^{ikx}$$

is defined for almost every $x \in \,]-\pi,\pi]$, and that $h^\bullet = S^{-1}\boldsymbol{c}$. (The point is that we know from the result just proved that there is *some* square-integrable f such that $\boldsymbol{c}$ is the sequence of Fourier coefficients of f; now 282Ia declares that the Fejér sums of f converge to f almost everywhere, that is, that $h =_{\text{a.e.}} \frac{1}{\sqrt{2\pi}}f$.)

282L The next result is the easiest, and one of the most useful, theorems concerning pointwise convergence of Fourier sums.

Theorem Let f be a complex-valued function which is integrable over $]-\pi,\pi]$, and $\langle s_n\rangle_{n\in\mathbb{N}}$ its sequence of Fourier sums.

(i) If f is differentiable at $x \in \,]-\pi,\pi[$, then $f(x) = \lim_{n\to\infty} s_n(x)$.

(ii) If the periodic extension $\tilde{f}$ of f is differentiable at π, then $f(\pi) = \lim_{n\to\infty} s_n(\pi)$.

proof (a) Take $x \in \,]-\pi,\pi]$ such that $\tilde{f}$ is differentiable at x; of course this covers both parts. We have

$$s_n(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\tilde{f}(x+t)}{\sin\frac{1}{2}t}\sin(n+\frac{1}{2})t\,dt$$

for each n, by 282Da.

(b) Next,

$$\int_{-\pi}^{\pi} \frac{\tilde{f}(x+t)-\tilde{f}(x)}{t}dt$$

exists in $\mathbb{C}$, because there is surely some $\delta \in \,]0,\pi]$ such that $(\tilde{f}(x+t)-\tilde{f}(x))/t$ is bounded on $\{t : 0 < |t| \le \delta\}$, while

$$\int_{-\pi}^{-\delta} \frac{\tilde{f}(x+t)-\tilde{f}(x)}{t}dt, \quad \int_{\delta}^{\pi} \frac{\tilde{f}(x+t)-\tilde{f}(x)}{t}dt$$

exist because $1/t$ is bounded on those intervals. It follows that

$$\int_{-\pi}^{\pi} \frac{\tilde{f}(x+t)-\tilde{f}(x)}{\sin\frac{1}{2}t}dt$$

exists, because $|t| \le \pi|\sin\frac{1}{2}t|$ if $|t| \le \pi$. So by the Riemann-Lebesgue lemma (282Fb),

$$\lim_{n\to\infty}\int_{-\pi}^{\pi} \frac{\tilde{f}(x+t)-\tilde{f}(x)}{\sin\frac{1}{2}t}\sin(n+\frac{1}{2})t\,dt = 0.$$

(c) Because

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \tilde{f}(x)\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}dt = \tilde{f}(x)$$

for every n (282Dc),

$$s_n(x) = \tilde{f}(x) + \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\tilde{f}(x+t)-\tilde{f}(x)}{\sin\frac{1}{2}t}\sin(n+\frac{1}{2})t\,dt \to \tilde{f}(x)$$

as $n \to \infty$, as required.

282M Lemma Suppose that f is a complex-valued function, defined almost everywhere and of bounded variation on $]-\pi,\pi]$. Then $\sup_{k\in\mathbb{Z}}|kc_k| < \infty$, where c_k is the kth Fourier coefficient of f, as in 282A.

proof Set

$$M = \lim_{x\in\operatorname{dom} f, x\uparrow\pi}|f(x)| + \operatorname{Var}_{]-\pi,\pi[}(f).$$

By 224J,

$$|kc_k| = \frac{1}{2\pi}\Bigl|\int_{-\pi}^{\pi} kf(t)e^{-ikt}dt\Bigr| \le \frac{1}{2\pi}M\sup_{c\in[-\pi,\pi]}\Bigl|\int_{-\pi}^{c} ke^{-ikt}dt\Bigr|$$
$$= \frac{M}{2\pi}\sup_{c\in[-\pi,\pi]}|e^{-ikc}-e^{ik\pi}| \le \frac{M}{\pi}$$

for every k.

282N I give another lemma, extracting the technical part of the proof of the next theorem. (Its most natural application is in 282Xn.)

Lemma Let $\langle d_k\rangle_{k\in\mathbb{N}}$ be a complex sequence, and set $t_n = \sum_{k=0}^{n} d_k$, $\tau_m = \frac{1}{m+1}\sum_{n=0}^{m} t_n$ for n, $m \in \mathbb{N}$. Suppose that $\sup_{k\in\mathbb{N}}|kd_k| = M < \infty$. Then for any $j \ge 1$ and any $c \in \mathbb{C}$,

$$|t_n - c| \le \frac{M}{j} + (2j+3)\sup_{m\ge n-n/j}|\tau_m - c|$$

for every $n \ge j^2$.

proof (a) The first point to note is that for any n, $n' \in \mathbb{N}$,

$$|t_n - t_{n'}| \le \frac{M|n-n'|}{1+\min(n,n')}.$$

$\mathbf{P}$ If $n = n'$ this is trivial. Suppose that $n' < n$. Then

$$|t_n - t_{n'}| = |\sum_{k=n'+1}^{n} d_k| \le \sum_{k=n'+1}^{n} \frac{M}{k} \le \frac{M(n-n')}{n'+1} = \frac{M|n-n'|}{1+\min(n',n)}.$$

Of course the case $n < n'$ is identical. $\mathbf{Q}$

(b) Now take any $n \ge j^2$. Set $\eta = \sup_{m \ge n-n/j} |\tau_m - c|$. Let $m \ge j$ be such that $jm \le n < j(m+1)$; then $n < jm + m$; also

$$n(1 - \tfrac{1}{j}) \le m(j+1)(1 - \tfrac{1}{j}) \le mj.$$

Set

$$\tau^* = \frac{1}{m} \sum_{n'=jm+1}^{jm+m} t_{n'} = \frac{jm+m+1}{m} \tau_{jm+m} - \frac{jm+1}{m} \tau_{jm}.$$

Then

$$\begin{aligned} |\tau^* - c| &= |\frac{jm+m+1}{m} \tau_{jm+m} - \frac{jm+1}{m} \tau_{jm} - c| \\ &= |\frac{jm+m+1}{m} (\tau_{jm+m} - c) - \frac{jm+1}{m} (\tau_{jm} - c)| \\ &\le \frac{jm+m+1}{m} \eta + \frac{jm+1}{m} \eta \le (2j+3)\eta. \end{aligned}$$

On the other hand,

$$\begin{aligned} |\tau^* - t_n| &= \Bigl|\frac{1}{m} \sum_{n'=jm+1}^{jm+m} (t_{n'} - t_n)\Bigr| \le \frac{1}{m} \sum_{n'=jm+1}^{jm+m} \frac{M|n-n'|}{1+\min(n,n')} \\ &\le \frac{1}{m} \sum_{n'=jm+1}^{jm+m} \frac{Mm}{1+jm} = \frac{Mm}{1+jm} \le \frac{M}{j}. \end{aligned}$$

Putting these together, we have

$$|t_n - c| \le |t_n - \tau^*| + |\tau^* - c| \le \frac{M}{j} + (2j+3)\eta = \frac{M}{j} + (2j+3) \sup_{m \ge n-n/j} |\tau_m - c|,$$

as required.

282O Theorem Let f be a complex-valued function of bounded variation, defined almost everywhere in $]-\pi, \pi]$, and let $\langle s_n \rangle_{n \in \mathbb{N}}$ be its sequence of Fourier sums.

(i) If $x \in \,]-\pi, \pi[$, then

$$\lim_{n\to\infty} s_n(x) = \frac{1}{2}(\lim_{t \in \operatorname{dom} f, t \uparrow x} f(t) + \lim_{t \in \operatorname{dom} f, t \downarrow x} f(t)).$$

(ii) $\lim_{n\to\infty} s_n(\pi) = \frac{1}{2}(\lim_{t \in \operatorname{dom} f, t \uparrow \pi} f(t) + \lim_{t \in \operatorname{dom} f, t \downarrow -\pi} f(t))$.

(iii) If f is defined throughout $]-\pi, \pi]$, is continuous, and $\lim_{t \downarrow -\pi} f(t) = f(\pi)$, then $s_n(x) \to f(x)$ uniformly on $]-\pi, \pi]$.

proof (a) Note first that 224F shows that the limits $\lim_{t \in \operatorname{dom} f, t \downarrow x} f(t)$, $\lim_{t \in \operatorname{dom} f, t \uparrow x} f(t)$ required in the formulae above always exist. We know also from 282M that $M = \sup_{k \in \mathbb{Z}} |kc_k| < \infty$, where c_k is the kth Fourier coefficient of f.

Take any $x \in \,]-\pi, \pi]$, and set

$$c = \tfrac{1}{2}(\lim_{t \in \operatorname{dom} f, t \uparrow x} \tilde{f}(t) + \lim_{t \in \operatorname{dom} \tilde{f}, t \downarrow x} \tilde{f}(t)),$$

writing $\tilde{f}$ for the periodic extension of f, as usual. We know from 282Id-282Ie that $c = \lim_{m\to\infty} \sigma_m(x)$, writing σ_m for the Fejér sums of f. Let $\epsilon > 0$. Take any $j \ge \max(2, 2M/\epsilon)$, and $m_0 \ge 1$ such that $|\sigma_m(x) - c| \le \epsilon/(2j+3)$ for every $m \ge m_0$.

Now if $n \ge \max(j^2, 2m_0)$, apply Lemma 282N with

$$d_0 = c_0, \quad d_k = c_k e^{ikx} + c_{-k}e^{-ikx} \text{ for } k \ge 1,$$

so that $t_n = s_n(x)$, $\tau_m = \sigma_m(x)$ and $|kd_k| \le 2M$ for every k, n, $m \in \mathbb{N}$. We have $n - n/j \ge \frac{1}{2}n \ge m_0$, so

$$\eta = \sup_{m\ge n-n/j} |\tau_m - c| \le \sup_{m\ge m_0} |\tau_m - c| \le \frac{\epsilon}{2j+3}.$$

So 282N tells us that

$$|s_n(x) - c| = |t_n - c| \le \frac{2M}{j} + (2j+3)\sup_{m\ge n-n/j} |\tau_m - c| \le \epsilon + (2j+3)\eta \le 2\epsilon.$$

As ϵ is arbitrary, $\lim_{n\to\infty} s_n(x) = c$, as required.

(b) This proves (i) and (ii) of this theorem. Finally, for (iii), observe that under these conditions $\sigma_m(x) \to f(x)$ uniformly as $m \to \infty$, by 282G. So given $\epsilon > 0$ we choose $j \ge \max(2, 2M/\epsilon)$ and $m_0 \in \mathbb{N}$ such that $|\sigma_m(x) - f(x)| \le \epsilon/(2j+3)$ whenever $m \ge m_0$ and $x \in \left]-\pi, \pi\right]$. By the same calculation as before,

$$|s_n(x) - f(x)| \le 2\epsilon$$

for every $n \ge \max(j^2, 2m_0)$ and every $x \in \left]-\pi, \pi\right]$. As ϵ is arbitrary, $\lim_{n\to\infty} s_n(x) = f(x)$ uniformly for $x \in \left]-\pi, \pi\right]$.

282P Corollary Let f be a complex-valued function which is integrable over $\left]-\pi, \pi\right]$, and $\langle s_n\rangle_{n\in\mathbb{N}}$ its sequence of Fourier sums.

(i) Suppose that $x \in \left]-\pi, \pi\right[$ is such that f is of bounded variation on some neighbourhood of x. Then

$$\lim_{n\to\infty} s_n(x) = \frac{1}{2}(\lim_{t\in\operatorname{dom} f, t\uparrow x} f(t) + \lim_{t\in\operatorname{dom} f, t\downarrow x} f(t)).$$

(ii) If there is a $\delta > 0$ such that f is of bounded variation on both $\left]-\pi, -\pi+\delta\right]$ and $[\pi-\delta, \pi]$, then

$$\lim_{n\to\infty} s_n(\pi) = \frac{1}{2}(\lim_{t\in\operatorname{dom} f, t\uparrow\pi} f(t) + \lim_{t\in\operatorname{dom} f, t\downarrow -\pi} f(t)).$$

proof In case (i), take $\delta > 0$ such that f is of bounded variation on $[x-\delta, x+\delta]$ and set $f_1(t) = f(t)$ if $x \in \operatorname{dom} f \cap [x-\delta, x+\delta]$, 0 for other $t \in \left]-\pi, \pi\right]$; in case (ii), set $f_1(t) = f(t)$ if $t \in \operatorname{dom} f$ and $|t| \ge \pi - \delta$, 0 for other $t \in \left]-\pi, \pi\right]$, and say that $x = \pi$. In either case, f_1 is of bounded variation, so by 282O the Fourier sums $\langle s'_n\rangle_{n\in\mathbb{N}}$ of f_1 converge at x to the value given by the formulae above. But now observe that, writing $\tilde{f}$ and $\tilde{f}_1$ for the periodic extensions of f and f_1, $\tilde{f} - \tilde{f}_1 = 0$ on a neighbourhood of x, so

$$\int_{-\pi}^{\pi} \frac{\tilde{f}(x+t) - \tilde{f}_1(x+t)}{\sin\frac{1}{2}t} dt$$

exists in $\mathbb{C}$, and by 282Fb

$$\lim_{n\to\infty} \int_{-\pi}^{\pi} \frac{\tilde{f}(x+t) - \tilde{f}_1(x+t)}{\sin\frac{1}{2}t} \sin(n+\frac{1}{2})t\, dt = 0,$$

that is, $\lim_{n\to\infty} s_n(x) - s'_n(x) = 0$. So $\langle s_n\rangle_{n\in\mathbb{N}}$ also converges to the right limit.

282Q I cannot leave this section without mentioning one of the most important facts about Fourier series, even though I have no space here to discuss its consequences.

Theorem Let f and g be complex-valued functions which are integrable over $\left]-\pi, \pi\right]$, and $\langle c_k\rangle_{k\in\mathbb{N}}$, $\langle d_k\rangle_{k\in\mathbb{N}}$ their Fourier coefficients. Let $f * g$ be their convolution, defined by the formula

$$(f*g)(x) = \int_{-\pi}^{\pi} f(x -_{2\pi} t)g(t)dt = \int_{-\pi}^{\pi} \tilde{f}(x-t)g(t)dt,$$

as in 255O, writing $\tilde{f}$ for the periodic extension of f. Then the Fourier coefficients of $f * g$ are $\langle 2\pi c_k d_k\rangle_{k\in\mathbb{Z}}$.

proof By 255O(c-i),

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}(f*g)(x)e^{-ikx}dx=\frac{1}{2\pi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}e^{-ik(t+u)}f(t)g(u)dtdu$$
$$=\frac{1}{2\pi}\int_{-\pi}^{\pi}e^{-ikt}f(t)dt\int_{-\pi}^{\pi}e^{-iku}g(u)du=2\pi c_k d_k.$$

***282R** In my hurry to get to the theorems on convergence of Fejér and Fourier sums, I have rather neglected the elementary manipulations which are essential when applying the theory. One basic result is the following.

Proposition (a) Let $f:[-\pi,\pi]\to\mathbb{C}$ be an absolutely continuous function such that $f(-\pi)=f(\pi)$, and $\langle c_k\rangle_{k\in\mathbb{Z}}$ its sequence of Fourier coefficients. Then the Fourier coefficients of f' are $\langle ikc_k\rangle_{k\in\mathbb{Z}}$.

(b) Let $f:\mathbb{R}\to\mathbb{C}$ be a differentiable function such that f' is absolutely continuous on $[-\pi,\pi]$, and $f(\pi)=f(-\pi)$. If $\langle c_k\rangle_{k\in\mathbb{Z}}$ are the Fourier coefficients of $f\restriction\,]-\pi,\pi]$, then $\sum_{k=-\infty}^{\infty}|c_k|$ is finite.

proof (a) By 225Cb, f' is integrable over $[-\pi,\pi]$; by 225E, f is an indefinite integral of f'. So 225F tells us that

$$\int_{-\pi}^{\pi}f'(x)e^{-ikx}dx=f(\pi)e^{-ik\pi}-f(-\pi)e^{ik\pi}+ik\int_{-\pi}^{\pi}f(x)e^{-ikx}dx=ikc_k$$

for every $k\in\mathbb{Z}$.

(b)(i) Suppose first that $f'(\pi)=f'(-\pi)$. By (a), applied twice, the Fourier coefficients of f'' are $\langle -k^2c_k\rangle_{k\in\mathbb{Z}}$, so $\sup_{k\in\mathbb{Z}}k^2|c_k|$ is finite; because $\sum_{k=1}^{\infty}\frac{1}{k^2}<\infty$, $\sum_{k=-\infty}^{\infty}|c_k|<\infty$.

(ii) Next, suppose that $f(x)=x^2$ for every x. Then, for $k\ne 0$,

$$c_k=\frac{1}{2\pi}\int x^2e^{-ikx}dx=\frac{1}{2\pi}\bigl(-\frac{1}{ik}(\pi^2e^{-ik\pi}-\pi^2e^{ik\pi})+\int_{-\pi}^{\pi}\frac{2x}{ik}e^{-ikx}dx\bigr)$$
$$=\frac{1}{ik\pi}\bigl(-\frac{1}{ik}(\pi e^{-ik\pi}+\pi e^{ik\pi})+\frac{1}{ik}\int_{-\pi}^{\pi}e^{ikx}dx\bigr)=\frac{2}{k^2}(-1)^k,$$

so $\sum_{k\in\mathbb{Z}}|c_k|\le|c_0|+4\sum_{k=1}^{\infty}\frac{1}{k^2}$ is finite.

(iii) In general, we can express f as f_1+cf_2 where $f_2(x)=x^2$ for every x, $c=\frac{1}{4\pi}(f'(\pi)-f'(-\pi))$, and f_1 satisfies the conditions of (i); so that $\langle c_k\rangle_{k\in\mathbb{Z}}$ is the sum of two summable sequences and is itself summable.

282X Basic exercises >(a) Suppose that $\langle c_k\rangle_{k\in\mathbb{N}}$ is an absolutely summable double-ended sequence of complex numbers. Show that $f(x)=\sum_{k=-\infty}^{\infty}c_ke^{ikx}$ exists for every $x\in\mathbb{R}$, that f is continuous and periodic, and that its Fourier coefficients are the c_k.

(c) Set $\phi_n(t)=\frac{2}{t}\sin(n+\frac{1}{2}t)$ for $t\ne 0$. (This is sometimes called the **modified Dirichlet kernel**.) Show that for any integrable function f on $]-\pi,\pi]$, with Fourier sums $\langle s_n\rangle_{n\in\mathbb{N}}$ and periodic extension $\tilde f$,

$$\lim_{n\to\infty}|s_n(x)-\frac{1}{2\pi}\int_{-\pi}^{\pi}\phi_n(t)\tilde f(x+t)dt|=0$$

for every $x\in\,]-\pi,\pi]$. (*Hint*: show that $\frac{2}{t}-\frac{1}{\sin\frac{1}{2}t}$ is bounded, and use 282E.)

(d) Give a proof of 282Ib from 242O, 255O and 282G.

(e) Give another proof of 282Ic, based on 242O and 281J instead of on 282H.

(f) Use the idea of 255Ya to shorten one of the steps in the proof of 282H, taking $g_m(t)=\min(\frac{m+1}{2\pi},\frac{\pi}{4(m+1)t^2})$ for $|t|\le\delta$, so that $g_m\ge\psi_m$ on $[-\delta,\delta]$.

>(g)(i) Let f be a real square-integrable function on $]-\pi,\pi]$, and $\langle a_k\rangle_{k\in\mathbb{N}}$, $\langle b_k\rangle_{k\ge1}$ its real Fourier coefficients (282Ba). Show that $\frac{1}{2}a_0^2+\sum_{k=1}^{\infty}(a_k^2+b_k^2)=\frac{1}{\pi}\int_{-\pi}^{\pi}|f|^2$. (ii) Show that $f\mapsto(\sqrt{\frac{\pi}{2}}a_0,\sqrt{\pi}a_1,\sqrt{\pi}b_1,\dots)$ defines an inner-product-space isomorphism between the real Hilbert space $L^2_{\mathbb{R}}$ of equivalence classes of real square-integrable functions on $]-\pi,\pi]$ and the real Hilbert space $\ell^2_{\mathbb{R}}$ of square-summable sequences.

(h) Show that $\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \ldots$. (*Hint*: find the Fourier series of f where $f(x) = x/|x|$, and compute the sum of the series at $\frac{\pi}{2}$. Of course there are other methods, e.g., examining the Taylor series for arctan $\frac{\pi}{4}$.)

(i) Let f be an integrable complex-valued function on $]-\pi,\pi]$, and $\langle s_n\rangle_{n\in\mathbb{N}}$ its sequence of Fourier sums. Suppose that $x \in]-\pi,\pi[$, $a \in \mathbb{C}$ are such that $\int_{-\pi}^{\pi} \frac{f(t)-a}{t-x} dt$ exists and is finite. Show that $\lim_{n\to\infty} s_n(x) = a$. Explain how this generalizes 282L. What modification is appropriate to obtain a limit $\lim_{n\to\infty} s_n(\pi)$?

(j) Suppose that $\alpha > 0$, $K \ge 0$ and $f :]-\pi,\pi[\to \mathbb{C}$ are such that $|f(x) - f(y)| \le K|x-y|^\alpha$ for all x, $y \in]-\pi,\pi[$. (Such functions are called **Hölder continuous**.) Show that the Fourier sums of f converge to f everywhere in $]-\pi,\pi[$. (*Hint*: use 282Xi.) (Compare 282Yb.)

(k) In 282L, show that it is enough if $\tilde{f}$ is differentiable with respect to its domain at x or π (see 262Fb), rather than differentiable in the strict sense.

(l) Show that $\lim_{a\to\infty} \int_0^a \frac{\sin t}{t} dt$ exists and is finite. (*Hint*: use 224J to estimate $\int_a^b \frac{\sin t}{t} dt$ for $0 < a \le b$.)

(m) Show that $\int_0^\infty \frac{|\sin t|}{t} dt = \infty$. (*Hint*: show that $\sup_{a\ge 0} |\int_1^a \frac{\cos 2t}{t} dt| < \infty$, and therefore that $\sup_{a\ge 0} \int_1^a \frac{\sin^2 t}{t} dt = \infty$.)

>(n) Let $\langle d_k\rangle_{k\in\mathbb{N}}$ be a sequence in $\mathbb{C}$ such that $\sup_{k\in\mathbb{N}} |kd_k| < \infty$ and

$$\lim_{m\to\infty} \frac{1}{m+1} \sum_{n=0}^m \sum_{k=0}^n d_k = c \in \mathbb{C}.$$

Show that $c = \sum_{k=0}^\infty d_k$. (*Hint*: 282N.)

>(o) Show that $\sum_{n=1}^\infty \frac{1}{n^2} = \frac{\pi^2}{6}$. (*Hint*: (b-ii) of the proof of 282R.)

(p) Let f be an integrable complex-valued function on $]-\pi,\pi]$, and $\langle s_n\rangle_{n\in\mathbb{N}}$ its sequence of Fourier sums. Suppose that $x \in]-\pi,\pi[$ is such that
(i) there is an $a \in \mathbb{C}$ such that
either $\int_{-\pi}^{x} \frac{a-f(t)}{x-t} dt$ exists in $\mathbb{C}$
or there is some $\delta > 0$ such that f is of bounded variation on $[x-\delta, x]$, and $a = \lim_{t\in\mathrm{dom}\, f, t\uparrow x} f(t)$
(ii) there is a $b \in \mathbb{C}$ such that
either $\int_x^\pi \frac{f(t)-b}{t-x} dt$ exists in $\mathbb{C}$
or there is some $\delta > 0$ such that f is of bounded variation on $[x, x+\delta]$, and $b = \lim_{t\in\mathrm{dom}\, f, t\downarrow x} f(t)$.
Show that $\lim_{n\to\infty} s_n(x) = \frac{1}{2}(a+b)$. What modification is appropriate to obtain a limit $\lim_{n\to\infty} s_n(\pi)$?

>(q) Let f, g be integrable complex-valued functions on $]-\pi,\pi]$, and $\boldsymbol{c} = \langle c_k\rangle_{k\in\mathbb{Z}}$, $\boldsymbol{d} = \langle d_k\rangle_{k\in\mathbb{Z}}$ their sequences of Fourier coefficients. Suppose that *either* $\sum_{k=-\infty}^\infty |c_k| < \infty$ *or* $\sum_{k=-\infty}^\infty |c_k|^2 + |d_k|^2 < \infty$. Show that the sequence of Fourier coefficients of $f \times g$ is just the convolution $\boldsymbol{c} * \boldsymbol{d}$ of $\boldsymbol{c}$ and $\boldsymbol{d}$ (255Xk).

(r) In 282Ra, what happens if $f(\pi) \ne f(-\pi)$?

(s) Suppose that $\langle c_k\rangle_{k\in\mathbb{N}}$ is a double-ended sequence of complex numbers such that $\sum_{k=-\infty}^\infty |kc_k| < \infty$. Show that $f(x) = \sum_{k=-\infty}^\infty c_k e^{ikx}$ exists for every $x \in \mathbb{R}$ and that f is differentiable everywhere.

(t) Let $\langle c_k\rangle_{k\in\mathbb{Z}}$ be a double-ended sequence of complex numbers such that $\sup_{k\in\mathbb{Z}} |kc_k| < \infty$. Show that there is a square-integrable function f on $]-\pi,\pi]$ such that the c_k are the Fourier coefficients of f, that f is the limit almost everywhere of its Fourier sums, and that $f * f * f$ is differentiable. (*Hint*: use 282K to show that there is an f, and 282Xn to show that its Fourier sums converge wherever its Fejér sums do; use 282Q and 282Xs to show that $f * f * f$ is differentiable.)

282Y Further exercises (a) Let f be a non-negative integrable function on $]-\pi,\pi]$, with Fourier coefficients $\langle c_k\rangle_{k\in\mathbb{Z}}$. Show that

$$\sum_{j=0}^n \sum_{k=0}^n a_j \bar{a}_k c_{j-k} \ge 0$$

for all complex numbers $a_0, \ldots, a_n$. (See also 285Xr below.)

(b) Let $f:]-\pi,\pi]\to\mathbb{C}$, $K\ge 0$, $\alpha>0$ be such that $|f(x)-f(y)|\le K|x-y|^\alpha$ for all x, $y\in]-\pi,\pi]$. Let c_k, s_n be the Fourier coefficients and sums of f. (i) Show that $\sup_{k\in\mathbb{Z}}|k|^\alpha|c_k|<\infty$. (*Hint*: show that $c_k=\frac{1}{4\pi}\int_{-\pi}^{\pi}(f(x)-\tilde f(x+\frac{\pi}{k}))e^{-ikx}dx$.) (ii) Show that if $f(\pi)=\lim_{x\downarrow-\pi}f(x)$ then $s_n\to f$ uniformly. (Compare 282Xj.)

(c) Let f be a measurable complex-valued function on $]-\pi,\pi]$, and suppose that $p\in[1,\infty[$ is such that $\int_{-\pi}^{\pi}|f|^p<\infty$. Let $\langle\sigma_m\rangle_{m\in\mathbb{N}}$ be the sequence of Fejér sums of f. Show that $\lim_{m\to\infty}\int_{-\pi}^{\pi}|f-\sigma_m|^p=0$. (*Hint*: use 245Xl, 255Yk and the ideas in 282Ib.)

(d) Construct a continuous function $h:[-\pi,\pi]\to\mathbb{R}$ such that $h(\pi)=h(-\pi)$ but the Fourier sums of h are unbounded at 0, as follows. Set $\alpha(m,n)=\int_0^\pi\frac{\sin(m+\frac{1}{2})t\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}dt$. Show that $\lim_{n\to\infty}\alpha(m,n)=0$ for every m, but $\lim_{n\to\infty}\alpha(n,n)=\infty$. Set $h_0(x)=\sum_{k=0}^\infty\delta_k\sin(m_k+\frac{1}{2})x$ for $0\le x\le\pi$, 0 for $-\pi\le x\le 0$, where $\delta_k>0$, $m_k\in\mathbb{N}$ are such that (α) $\delta_k\le 2^{-k}$, $\delta_k|\alpha(m_k,m_n)|\le 2^{-k}$ for every $n<k$ (choosing δ_k) (β) $\delta_k\alpha(m_k,m_k)\ge k$, $\delta_n|\alpha(m_k,m_n)|\le 2^{-n}$ for every $n<k$ (choosing m_k). Now modify h_0 on $[-\pi,0[$ by adding a function of bounded variation.

(e)(i) Show that $\lim_{n\to\infty}\int_{-\pi}^{\pi}|\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}|dt=\infty$. (*Hint*: 282Xm.) (ii) Show that for any $\delta>0$ there are $n\in\mathbb{N}$, $f\ge 0$ such that $\int_{-\pi}^{\pi}f\le\delta$, $\int_{-\pi}^{\pi}|s_n|\ge 1$, where s_n is the nth Fourier sum of f. (*Hint*: take n such that $\frac{1}{2\pi}\int_{-\pi}^{\pi}|\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}|dt>\frac{1}{\delta}$ and set $f(x)=\frac{\delta}{\eta}$ for $0\le x\le\eta$, 0 otherwise, where η is small.) (iii) Show that there is an integrable function $f:]-\pi,\pi]\to\mathbb{R}$ such that $\sup_{n\in\mathbb{N}}\|s_n\|_1$ is infinite, where $\langle s_n\rangle_{n\in\mathbb{N}}$ is the sequence of Fourier sums of f. (*Hint*: it helps to know the 'Uniform Boundedness Theorem' of functional analysis, but f can also be constructed bare-handed by the method of 282Yd.)

(f) Let $u:[-\pi,\pi]\to\mathbb{R}$ be an absolutely continuous function such that $u(\pi)=u(-\pi)$ and $\int_{-\pi}^{\pi}u=0$. Show that $\|u\|_2\le\|u'\|_2$. (This is **Wirtinger's inequality**.)

(g) For $0\le r<1$, $t\in\mathbb{R}$ set $A_r(t)=\frac{1-r^2}{1-2r\cos t+r^2}$. ($A_r$ is the **Poisson kernel**; see 478Xl[1] in Volume 4.) (i) Show that $\frac{1}{2\pi}\int_{-\pi}^{\pi}A_r=1$. (ii) For a real function f which is integrable over $]-\pi,\pi]$, with real Fourier coefficients a_k, b_k (282Ba), set $S_r(x)=\frac{1}{2}a_0+\sum_{k=1}^\infty r^k(a_k\cos kx+b_k\sin kx)$ for $x\in]-\pi,\pi]$, $r\in[0,1[$. Show that $S_r(x)=\frac{1}{2\pi}\int_{-\pi}^{\pi}A_r(x-t)f(t)dt$ for every $x\in]-\pi,\pi]$. (*Hint*: $A_r(t)=1+2\sum_{n=1}^\infty r^n\cos nt$.) (iii) Show that $\lim_{r\uparrow1}S_r(x)=f(x)$ for every $x\in]-\pi,\pi[$ which is in the Lebesgue set of f. (*Hint*: 223Yg.) (iv) Show that $\lim_{r\uparrow1}\int_{-\pi}^{\pi}|S_r-f|=0$. (v) Show that if f is defined everywhere on $]-\pi,\pi]$, is continuous, and $f(\pi)=\lim_{x\downarrow-\pi}f(x)$, then $\lim_{r\uparrow1}\sup_{x\in]-\pi,\pi]}|S_r(x)-f(x)|=0$.

282 Notes and comments This has been a long section with a potentially confusing collection of results, so perhaps I should recapitulate. Associated with any integrable function on $]-\pi,\pi]$ we have the corresponding Fourier sums, being the symmetric partial sums $\sum_{k=-n}^n c_ke^{ikx}$ of the complex series $\sum_{k=-\infty}^\infty c_ke^{ikx}$, or, equally, the partial sums $\frac{1}{2}a_0+\sum_{k=1}^n a_k\cos kx+b_k\sin kx$ of the real series $\frac{1}{2}a_0+\sum_{k=1}^\infty a_k\cos kx+b_k\sin kx$. The Fourier coefficients c_k, a_k, b_k are the only natural ones, because if the series is to converge with any regularity at all then

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\Bigl(\sum_{k=-\infty}^\infty c_ke^{ikx}\Bigr)e^{-ilx}dx$$

ought to be simultaneously

$$\sum_{k=-\infty}^\infty\frac{1}{2\pi}\int_{-\pi}^{\pi}c_ke^{ikx}e^{-ilx}dx=c_l$$

and

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}f(x)e^{-ilx}dx.$$

(Compare the calculations in 282J.) The effect of taking Fejér sums $\sigma_m(x)$ rather than the Fourier sums $s_n(x)$ is to smooth the sequence out; recall that if $\lim_{n\to\infty}s_n(x)=c$ then $\lim_{m\to\infty}\sigma_m(x)=c$, by 273Ca in the last chapter.

Most of the work above is concerned with the question of when Fourier or Fejér sums converge, in some sense, to the original function f. As has happened before, in §245 and elsewhere, we have more than one kind of convergence to

[1]Later editions only.

consider. *Norm* convergence, for $\| \|_1$ or $\| \|_2$ or $\| \|_\infty$, is the simplest; the three theorems 282G, 282Ib and 282J at least are relatively straightforward. (I have given 282Ib as a corollary of 282Ia; but there is an easier proof from 282G. See 282Xd.) Respectively, we have

if f is continuous (and matches at $\pm\pi$, that is, $f(\pi) = \lim_{t\downarrow-\pi} f(t)$) then $\sigma_m \to f$ uniformly, that is, for $\| \|_\infty$ (282G);

if f is any integrable function, then $\sigma_m \to f$ for $\| \|_1$ (282Ib);

if f is a square-integrable function, then $s_n \to f$ for $\| \|_2$ (282J);

if f is continuous and of bounded variation (and matches at $\pm\pi$), then $s_n \to f$ uniformly (282O).

There are some similar results for other $\| \|_p$ (282Yc); but note that the Fourier sums need not converge for $\| \|_1$ (282Ye).

Pointwise convergence is harder. The results I give are

if f is any integrable function, then $\sigma_m \to f$ almost everywhere (282Ia);

this relies on some careful calculations in 282H, and also on the deep result 223D. Next we have the results which look at the average of the limits of f from the two sides. Suppose I write

$$f^{\pm}(x) = \frac{1}{2}(\lim_{t\uparrow x} f(t) + \lim_{t\downarrow x} f(t))$$

whenever this is defined, taking $f^{\pm}(\pi) = \frac{1}{2}(\lim_{t\uparrow\pi} f(t) + \lim_{t\downarrow-\pi} f(t))$. Then we have

if f is any integrable function, $\sigma_m \to f^{\pm}$ wherever $f^{\pm}$ is defined (282I);

if f is of bounded variation, $s_n \to f^{\pm}$ everywhere (282O).

Of course these apply at any point at which f is continuous, in which case $f(x) = f^{\pm}(x)$. Yet another result of this type is

if f is any integrable function, $s_n \to f$ at any point at which f is differentiable (282L);

in fact, this can be usefully extended for very little extra labour (282Xi, 282Xp).

I cannot leave this list without mentioning the theorem I have *not* given. This is **Carleson's theorem**:

if f is square-integrable, $s_n \to f$ almost everywhere

(Carleson 66). I will come to this in §286. There is an elementary special case in 282Xt. The result is in fact valid for many other f (see the notes to §286).

The next glaring lacuna in the exposition here is the absence of any examples to show how far these results are best possible. There is no suggestion, indeed, that there are any natural necessary and sufficient conditions for

$s_n \to f$ at every point.

Nevertheless, we have to make an effort to find a continuous function for which this is not so, and the construction of an example by du Bois-Reymond (Bois-Reymond 1876) was an important moment in the history of analysis, not least because it forced mathematicians to realise that some comfortable assumptions about the classification of functions – essentially, that functions are either 'good' or so bad that one needn't trouble with them – were false. The example is instructive but I have had to omit it for lack of space; I give an outline of a possible method in 282Yd. (You can find a detailed construction in Körner 88, chapter 18, and a proof that such a function exists in Dudley 89, 7.4.3.) If you allow general integrable functions, then you can do much better, or perhaps I should say much worse; there is an integrable f such that $\sup_{n\in\mathbb{N}} |s_n(x)| = \infty$ for every $x \in \,]-\pi, \pi]$ (Kolmogorov 26; see Zygmund 59, §§VIII.3-4).

In 282C I mentioned two types of problem. The first – when is a Fourier series summable? – has at least been treated at length, even though I cannot pretend to have given more than a sample of what is known. The second – how do properties of the c_k reflect properties of f? – I have hardly touched on. I do give what seem to me to be the three most important results in this area. The first is

if f and g have the same Fourier coefficients, they are equal almost everywhere (282Ic).

This at least tells us that we ought in principle to be able to learn almost anything about f by looking at its Fourier series. (For instance, 282Ya describes a necessary and sufficient condition for f to be non-negative almost everywhere.) The second is

f is square-integrable iff $\sum_{k=-\infty}^{\infty} |c_k|^2 < \infty$;

in fact,

$\sum_{k=-\infty}^{\infty} |c_k|^2 = \frac{1}{2\pi}\int_{\pi}^{\pi} |f|^2$ (282J).

Of course this is fundamental, since it shows that Fourier coefficients provide a natural Hilbert space isomorphism between L^2 and ℓ^2 (282K). I should perhaps remark that while the real Hilbert spaces $L^2_{\mathbb{R}}$, $\ell^2_{\mathbb{R}}$ are isomorphic as inner product spaces (282Xg), they are certianly not isomorphic as Banach lattices; for instance, $\ell^2_{\mathbb{R}}$ has 'atomic' elements $\boldsymbol{c}$ such that if $0 \le \boldsymbol{d} \le \boldsymbol{c}$ then $\boldsymbol{d}$ is a multiple of $\boldsymbol{c}$, while $L^2_{\mathbb{R}}$ does not. Perhaps even more important is the Fourier coefficients of a convolution $f*g$ are just a scalar multiple of the products of the Fourier coefficients of f and g (282Q);

but to use this effectively we need to study the Banach algebra structure of L^1, and I have no choice but to abandon this path immediately. (It will form a conspicuous part of Chapter 44 in Volume 4.) 282Xt gives an elementary consequence, and 282Xq a very partial description of the relationship between a product $f\times g$ of two functions and the convolution product of their sequences of Fourier coefficients.

The Fejér sums considered in this section are one way of working around the convergence difficulties associated with Fourier sums. When we come to look at Fourier transforms in the next two sections we shall need some further manoeuvres. A different type of smoothing is obtained by using the Poisson kernel in place of the Dirichlet or Fejér kernel (282Yg).

I end these notes with a remark on the number 2π. This enters nearly every formula involving Fourier series, but could I think be removed totally from the present section, at least, by re-normalizing the measure of $]-\pi,\pi]$. If instead of Lebesgue measure μ we took the measure $\nu=\frac{1}{2\pi}\mu$ throughout, then every 2π would disappear. (Compare the remark in 282Bb concerning the possibility of doing integrals over S^1.) But I think most of us would prefer to remember the location of a 2π in every formula than to deal with an unfamiliar measure.

283 Fourier transforms I

I turn now to the theory of Fourier transforms on $\mathbb{R}$. In the first of two sections on the subject, I present those parts of the elementary theory which can be dealt with using the methods of the previous section on Fourier series. I find no way of making sense of the theory, however, without introducing a fragment of L.Schwartz' theory of distributions, which I present in §284. As in §282, of course, this treatment also is nothing but a start in the topic.

The whole theory can also be done in $\mathbb{R}^r$. I leave this extension to the exercises, however, since there are few new ideas, the formulae are significantly more complicated, and I shall not, in this volume at least, have any use for the multidimensional versions of these particular theorems, though some of the same ideas will appear, in multidimensional form, in §285.

283A Definitions Let f be a complex-valued function which is integrable over $\mathbb{R}$.

(a) The **Fourier transform** of f is the function $\hat{f}:\mathbb{R}\to\mathbb{C}$ defined by setting

$$\hat{f}(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-iyx}f(x)dx$$

for every $y\in\mathbb{R}$. (Of course the integral is always defined because $x\mapsto e^{-iyx}$ is bounded and continuous, therefore measurable.)

(b) The **inverse Fourier transform** of f is the function $\check{f}:\mathbb{R}\to\mathbb{C}$ defined by setting

$$\check{f}(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{iyx}f(x)dx$$

for every $y\in\mathbb{R}$.

283B Remarks (a) It is a mildly vexing feature of the theory of Fourier transforms – vexing, that is, for outsiders like myself – that there is in fact no standard definition of 'Fourier transform'. The commonest definitions are, I think,

$$\hat{f}(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{\mp iyx}f(x)dx,$$

$$\hat{f}(y)=\int_{-\infty}^{\infty}e^{\mp iyx}f(x)dx,$$

$$\hat{f}(y)=\int_{-\infty}^{\infty}e^{\mp 2\pi iyx}f(x)dx,$$

corresponding to inverse transforms

$$\check{f}(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{\pm iyx}f(x)dx,$$

$$\check{f}(y)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{\pm iyx}f(x)dx,$$

$$\overset{\vee}{f}(y) = \int_{-\infty}^{\infty} e^{\pm 2\pi i y x} f(x) dx.$$

I leave it to you to check that the whole theory can be carried through with any of these six pairs, and to investigate other possibilities (see 283Xa-283Xb below).

(b) The phrases 'Fourier transform', 'inverse Fourier transform' make it plain that $(\hat{f})^\vee$ is supposed to be f, at least some of the time. This is indeed the case, but the class of f for which this is true in the literal sense is somewhat constrained, and we shall have to wait a little while before investigating it.

(c) No amount of juggling with constants, in the manner of (a) above, can make $\hat{f}$ and $\check{f}$ quite the same. However, on the definitions I have chosen, we do have $\check{f}(y) = \hat{f}(-y)$ for every y, so that $\check{f}$ and $\hat{f}$ will share essentially all the properties of interest to us here; in particular, everything in the next proposition will be valid with $^\vee$ in place of $^\wedge$, if you change signs at the right points in parts (c), (h) and (i).

283C Proposition Let f and g be complex-valued functions which are integrable over $\mathbb{R}$.

(a) $(f+g)^\wedge = \hat{f} + \hat{g}$.

(b) $(cf)^\wedge = c\hat{f}$ for every $c \in \mathbb{C}$.

(c) If $c \in \mathbb{R}$ and $h(x) = f(x+c)$ whenever this is defined, then $\hat{h}(y) = e^{icy}\hat{f}(y)$ for every $y \in \mathbb{R}$.

(d) If $c \in \mathbb{R}$ and $h(x) = e^{icx}f(x)$ for every $x \in \operatorname{dom} f$, then $\hat{h}(y) = \hat{f}(y-c)$ for every $y \in \mathbb{R}$.

(e) If $c > 0$ and $h(x) = f(cx)$ whenever this is defined, then $\hat{h}(y) = \frac{1}{c}\hat{f}(cy)$ for every $y \in \mathbb{R}$.

(f) $\hat{f} : \mathbb{R} \to \mathbb{C}$ is continuous.

(g) $\lim_{y\to\infty} \hat{f}(y) = \lim_{y\to-\infty} \hat{f}(y) = 0$.

(h) If $\int_{-\infty}^{\infty} |xf(x)|dx < \infty$, then $\hat{f}$ is differentiable, and its derivative is

$$\hat{f}'(y) = -\frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-iyx} xf(x) dx$$

for every $y \in \mathbb{R}$.

(i) If f is absolutely continuous on every bounded interval and f' is integrable, then $(f')^\wedge(y) = iy\hat{f}(y)$ for every $y \in \mathbb{R}$.

proof (a) and **(b)** are trivial, and **(c)**, **(d)** and **(e)** are elementary substitutions.

(f) If $\langle y_n \rangle_{n\in\mathbb{N}}$ is any convergent sequence in $\mathbb{R}$ with limit y, then

$$\begin{aligned}\hat{f}(y) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \lim_{n\to\infty} e^{-iy_n x} f(x) dx \\ &= \lim_{n\to\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-iy_n x} f(x) dx = \lim_{n\to\infty} \hat{f}(y_n)\end{aligned}$$

by Lebesgue's Dominated Convergence Theorem, because $|e^{-iy_n x} f(x)| \le |f(x)|$ for every $n \in \mathbb{N}$, $x \in \operatorname{dom} f$. As $\langle y_n \rangle_{n\in\mathbb{N}}$ is arbitrary, $\hat{f}$ is continuous.

(g) This is just the Riemann-Lebesgue lemma (282E).

(h) The point is that $|\frac{\partial}{\partial y} e^{-iyx} f(x)| = |xf(x)|$ for every $x \in \operatorname{dom} f$, $y \in \mathbb{R}$. So by 123D

$$\begin{aligned}\hat{f}'(y) &= \frac{1}{\sqrt{2\pi}} \frac{d}{dy} \int_{-\infty}^{\infty} e^{-iyx} f(x) dx = \frac{1}{\sqrt{2\pi}} \frac{d}{dy} \int_{\operatorname{dom} f} e^{-iyx} f(x) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\operatorname{dom} f} \frac{\partial}{\partial y} e^{-iyx} f(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} -ixe^{-iyx} f(x) dx \\ &= -\frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} xe^{-iyx} f(x) dx.\end{aligned}$$

(i) Because f is absolutely continuous on every bounded interval,

$$f(x) = f(0) + \int_0^x f' \text{ for } x \ge 0, \quad f(x) = f(0) - \int_x^0 f' \text{ for } x \le 0.$$

Because f' is integrable,

$$\lim_{x\to\infty} f(x) = f(0) + \int_0^\infty f', \quad \lim_{x\to-\infty} f(x) = f(0) - \int_{-\infty}^0 f'$$

both exist. Because f also is integrable, both limits must be zero. Now we can integrate by parts (225F) to see that

$$\begin{aligned}(f')^\wedge(y) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-iyx} f'(x)dx = \frac{1}{\sqrt{2\pi}}\lim_{a\to\infty}\int_{-a}^{a} e^{-iyx} f'(x)dx \\ &= \frac{1}{\sqrt{2\pi}}\bigl(\lim_{a\to\infty} e^{-iya} f(a) - \lim_{a\to-\infty} e^{-iya} f(a)\bigr) + \frac{iy}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-iyx} f(x)dx \\ &= iy\hat{f}(y).\end{aligned}$$

283D Lemma (a) $\lim_{a\to\infty}\int_0^a \frac{\sin x}{x}dx = \frac{\pi}{2}$, $\lim_{a\to\infty}\int_{-a}^a \frac{\sin x}{x}dx = \pi$.

(b) There is a $K < \infty$ such that $|\int_a^b \frac{\sin cx}{x}dx| \le K$ whenever $a \le b$ and $c \in \mathbb{R}$.

proof (a)(i) Set

$$F(a) = \int_0^a \frac{\sin x}{x}dx \text{ if } a \ge 0, \quad F(a) = -\int_{-a}^0 \frac{\sin x}{x}dx \text{ if } a \le 0,$$

so that $F(a) = -F(-a)$ and $\int_a^b \frac{\sin x}{x}dx = F(b) - F(a)$ for all $a \le b$.

If $0 < a \le b$, then by 224J

$$|\int_a^b \frac{\sin x}{x}dx| \le (\frac{1}{b} + \frac{1}{a} - \frac{1}{b})\sup_{c\in[a,b]}|\int_a^c \sin x\,dx| \le \frac{1}{a}\sup_{c\in[a,b]}|\cos c - \cos a| \le \frac{2}{a}.$$

In particular, $|F(n) - F(m)| \le \frac{2}{m}$ if $m \le n$ in $\mathbb{N}$, and $\langle F(n)\rangle_{n\in\mathbb{N}}$ is a Cauchy sequence with limit γ say; now

$$|\gamma - F(a)| = \lim_{n\to\infty}|F(n) - F(a)| \le \frac{2}{a}$$

for every $a > 0$, so $\lim_{a\to\infty} F(a) = \gamma$. Of course we also have

$$\lim_{a\to\infty}\int_{-a}^a \frac{\sin x}{x}dx = \lim_{a\to\infty}(F(a) - F(-a)) = \lim_{a\to\infty} 2F(a) = 2\gamma.$$

(ii) So now I have to calculate γ. For this, observe first that

$$2\gamma = \lim_{a\to\infty}\int_{-\pi a}^{\pi a}\frac{\sin x}{x}dx = \lim_{a\to\infty}\int_{-\pi}^{\pi}\frac{\sin at}{t}dt$$

(substituting $x = t/a$). Next,

$$\lim_{t\to0}\frac{1}{t} - \frac{1}{2\sin\frac{1}{2}t} = \lim_{u\to0}\frac{\sin u - u}{2u\sin u} = 0,$$

so

$$\int_{-\pi}^{\pi}\Bigl|\frac{1}{t} - \frac{1}{2\sin\frac{1}{2}t}\Bigr|dt < \infty,$$

and by the Riemann-Lebesgue lemma (282Fb)

$$\lim_{a\to\infty}\int_{-\pi}^{\pi}\Bigl(\frac{1}{t} - \frac{1}{2\sin\frac{1}{2}t}\Bigr)\sin at\,dt = 0.$$

But we know that

$$\int_{-\pi}^{\pi}\frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}dt = \pi$$

for every n (using 282Dc), so we must have

$$\begin{aligned}\lim_{a\to\infty}\int_{-a}^{a}\frac{\sin t}{t}dt &= \lim_{a\to\infty}\int_{-\pi}^{\pi}\frac{\sin at}{t}dt = \lim_{a\to\infty}\int_{-\pi}^{\pi}\frac{\sin at}{2\sin\frac{1}{2}t}dt \\ &= \lim_{n\to\infty}\int_{-\pi}^{\pi}\frac{\sin(n+\frac{1}{2})t}{2\sin\frac{1}{2}t}dt = \pi,\end{aligned}$$

and $\gamma = \pi/2$, as claimed.

(b) Because F is continuous and

$$\lim_{a\to\infty} F(a) = \gamma = \frac{\pi}{2}, \quad \lim_{a\to-\infty} F(a) = -\gamma = -\frac{\pi}{2},$$

F is bounded; say $|F(a)| \le K_1$ for all $a \in \mathbb{R}$. Try $K = 2K_1$. Now suppose that $a < b$ and $c \in \mathbb{R}$. If $c > 0$, then

$$|\int_a^b \frac{\sin cx}{x} dx| = |\int_{ac}^{bc} \frac{\sin t}{t} dt| = |F(bc) - F(ac)| \le 2K_1 = K,$$

substituting $x = t/c$. If $c < 0$, then

$$|\int_a^b \frac{\sin cx}{x} dx| = | - \int_a^b \frac{\sin(-c)x}{x} dx| \le K;$$

while if $c = 0$ then

$$|\int_a^b \frac{\sin cx}{x} dx| = 0 \le K.$$

283E The hardest work of this section will lie in the 'pointwise inversion theorems' 283I and 283K below. I begin however with a relatively easy, and at least equally important, result, showing (among other things) that an integrable function f can (essentially) be recovered from its Fourier transform.

Lemma Whenever $c < d$ in $\mathbb{R}$,

$$\begin{aligned}\lim_{a\to\infty} \int_{-a}^{a} e^{-iyx} \frac{e^{idy}-e^{icy}}{y} dy &= 2\pi i \text{ if } c < x < d,\\ &= \pi i \text{ if } x = c \text{ or } x = d,\\ &= 0 \text{ if } x < c \text{ or } x > d.\end{aligned}$$

proof We know that for any $b > 0$

$$\lim_{a\to\infty} \int_{-a}^{a} \frac{\sin bx}{x} dx = \lim_{a\to\infty} \int_{-ab}^{ab} \frac{\sin t}{t} dt = \pi$$

(subsituting $x = t/b$), and therefore that for any $b < 0$

$$\lim_{a\to\infty} \int_{-a}^{a} \frac{\sin bx}{x} dx = -\lim_{a\to\infty} \int_{-a}^{a} \frac{\sin(-b)x}{x} dx = -\pi.$$

Now consider, for $x \in \mathbb{R}$,

$$\lim_{a\to\infty} \int_{-a}^{a} e^{-iyx} \frac{e^{idy}-e^{icy}}{y} dy.$$

First note that all the integrals $\int_{-a}^{a}$ exist, because

$$\lim_{y\to 0} \frac{e^{idy}-e^{icy}}{y} = i(d-c)$$

is finite, and the integrand is certainly continuous except at 0. Now we have

$$\begin{aligned}\int_{-a}^{a} & e^{-iyx} \frac{e^{idy}-e^{icy}}{y} dy\\ &= \int_{-a}^{a} \frac{e^{i(d-x)y}-e^{i(c-x)y}}{y} dy\\ &= \int_{-a}^{a} \frac{\cos(d-x)y-\cos(c-x)y}{y} dy + i \int_{-a}^{a} \frac{\sin(d-x)y-\sin(c-x)y}{y} dy\\ &= i \int_{-a}^{a} \frac{\sin(d-x)y-\sin(c-x)y}{y} dy\end{aligned}$$

because cos is an even function, so

$$\int_{-a}^{a} \frac{\cos(d-x)y-\cos(c-x)y}{y} dy = 0$$

for every $a \geq 0$. (Once again, this integral exists because

$$\lim_{y\to 0} \frac{\cos(d-x)y-\cos(c-x)y}{y} = 0.)$$

Accordingly

$$\begin{aligned}\lim_{a\to\infty}\int_{-a}^{a} e^{-iyx}\frac{e^{idy}-e^{icy}}{y}dy &= i\lim_{a\to\infty}\int_{-a}^{a}\frac{\sin(d-x)y}{y}dy - i\lim_{a\to\infty}\int_{-a}^{a}\frac{\sin(c-x)y}{y}dy\\ &= i\pi - i\pi = 0 \text{ if } x < c,\\ &= i\pi - 0 = \pi i \text{ if } x = c,\\ &= i\pi + i\pi = 2\pi i \text{ if } c < x < d,\\ &= 0 + i\pi = \pi i \text{ if } x = d,\\ &= -i\pi + i\pi = 0 \text{ if } x > d.\end{aligned}$$

283F Theorem Let f be a complex-valued function which is integrable over $\mathbb{R}$, and $\hat{f}$ its Fourier transform. Then whenever $c \leq d$ in $\mathbb{R}$,

$$\int_c^d f = \frac{i}{\sqrt{2\pi}}\lim_{a\to\infty}\int_{-a}^{a}\frac{e^{icy}-e^{idy}}{y}\hat{f}(y)dy.$$

proof If $c = d$ this is trivial; let us suppose that $c < d$.

(a) Writing

$$\theta_a(x) = \int_{-a}^{a} e^{-iyx}\,\frac{e^{idy}-e^{icy}}{y}dy$$

for $x \in \mathbb{R}$, $a \geq 0$, 283E tells us that

$$\lim_{a\to\infty}\theta_a(x) = 2\pi i\theta(x)$$

where $\theta = \frac{1}{2}(\chi[c,d] + \chi]c,d[)$ takes the value 1 inside the interval $[c,d]$, 0 outside and the value $\frac{1}{2}$ at the endpoints. At the same time,

$$\begin{aligned}|\theta_a(x)| &= |\int_{-a}^{a}\frac{\sin(d-x)y-\sin(c-x)y}{y}dy|\\ &\leq |\int_{-a}^{a}\frac{\sin(d-x)y}{y}dy| + |\int_{-a}^{a}\frac{\sin(c-x)y}{y}dy| \leq 2K\end{aligned}$$

for all $a \geq 0$, $x \in \mathbb{R}$, where K is the constant of 283Db. Consequently $|f\times\theta_a| \leq 2K|f|$ everywhere on $\operatorname{dom} f$, for every $a \geq 0$, and (applying Lebesgue's Dominated Convergence Theorem to sequences $\langle f\times\theta_{a_n}\rangle_{n\in\mathbb{N}}$, where $a_n \to \infty$)

$$\lim_{a\to\infty}\int f\times\theta_a = 2\pi i\int f\times\theta = 2\pi i\int_c^d f.$$

(b) Now consider the limit in the statement of the theorem. We have

$$\begin{aligned}\int_{-a}^{a}\frac{e^{icy}-e^{idy}}{y}\hat{f}(y)dy &= \frac{1}{\sqrt{2\pi}}\int_{-a}^{a}\int_{-\infty}^{\infty}\frac{e^{icy}-e^{idy}}{y}e^{-iyx}f(x)dxdy\\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\int_{-a}^{a}\frac{e^{icy}-e^{idy}}{y}e^{-iyx}f(x)dydx\\ &= -\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f\times\theta_a,\end{aligned}$$

by Fubini's and Tonelli's theorems (252H), using the fact that $(e^{icy}-e^{idy})/y$ is bounded to see that

$$\int_{-\infty}^{\infty}\int_{-a}^{a}\Big|\frac{e^{icy}-e^{idy}}{y}e^{-iyx}f(x)\Big|dydx$$

is finite. Accordingly

$$\frac{i}{\sqrt{2\pi}}\lim_{a\to\infty}\int_{-a}^{a}\frac{e^{icy}-e^{idy}}{y}\hat{f}(y)dy=-\frac{i}{2\pi}\lim_{a\to\infty}\int_{-\infty}^{\infty}f\times\theta_a$$
$$=-\frac{i}{2\pi}2\pi i\int_c^d f=\int_c^d f,$$

as required.

283G Corollary If f and g are complex-valued functions which are integrable over $\mathbb{R}$, then $\hat{f}=\hat{g}$ iff $f=_{\text{a.e.}}g$.

proof If $f=_{\text{a.e.}}g$ then of course

$$\hat{f}(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{iyx}f(x)dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{iyx}g(x)dx=\hat{g}(y)$$

for every $y\in\mathbb{R}$. Conversely, if $\hat{f}=\hat{g}$, then by the last theorem

$$\int_c^d f=\int_c^d g$$

for all $c\le d$, so $f=g$ almost everywhere, by 222D.

283H Lemma Let f be a complex-valued function which is integrable over $\mathbb{R}$, and $\hat{f}$ its Fourier transform. Then for any $a>0$, $x\in\mathbb{R}$,

$$\frac{1}{\sqrt{2\pi}}\int_{-a}^{a}e^{ixy}\hat{f}(y)dy=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin a(x-t)}{x-t}f(t)dt=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin at}{t}f(x-t)dt.$$

proof We have

$$\int_{-a}^{a}\int_{-\infty}^{\infty}|e^{ixy}e^{-iyt}f(t)|dtdy\le 2a\int_{-\infty}^{\infty}|f(t)|dt<\infty,$$

so (because the function $(t,y)\mapsto e^{ixy}e^{-iyt}f(t)$ is surely jointly measurable) we may reverse the order of integration, and get

$$\frac{1}{\sqrt{2\pi}}\int_{-a}^{a}e^{ixy}\hat{f}(y)dy=\frac{1}{2\pi}\int_{-a}^{a}\int_{-\infty}^{\infty}e^{ixy}e^{-iyt}f(t)dt\,dy$$
$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}f(t)\int_{-a}^{a}e^{i(x-t)y}dy\,dt$$
$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{2\sin(x-t)a}{x-t}f(t)dt=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin au}{u}f(x-u)du,$$

substituting $t=x-u$.

283I Theorem Let f be a complex-valued function which is integrable over $\mathbb{R}$, and suppose that f is differentiable at $x\in\mathbb{R}$. Then

$$f(x)=\frac{1}{\sqrt{2\pi}}\lim_{a\to\infty}\int_{-a}^{a}e^{ixy}\hat{f}(y)dy=\frac{1}{\sqrt{2\pi}}\lim_{a\to\infty}\int_{-a}^{a}e^{-ixy}\check{f}(y)dy.$$

proof Set $g(u)=f(x)$ if $|u|\le 1$, 0 otherwise, and observe that $\lim_{u\to 0}\frac{1}{u}(f(x-u)-g(u))=-f'(x)$ is finite, so that there is a $\delta\in\,]0,1]$ such that

$$K=\sup_{0<|u|\le\delta}\left|\frac{f(x-u)-g(u)}{u}\right|<\infty.$$

Consequently

$$\int_{-\infty}^{\infty}\left|\frac{f(x-u)-g(u)}{u}\right|du\le\frac{1}{\delta}\int_{-\infty}^{-\delta}|f(x-u)|du+\frac{1}{\delta}\int_{-1}^{1}|g|$$
$$+\int_{-\delta}^{\delta}K+\frac{1}{\delta}\int_{\delta}^{\infty}|f(x-u)|du$$
$$\le\frac{1}{\delta}\int_{-\infty}^{\infty}|f|+\frac{2}{\delta}|f(x)|+2\delta K<\infty.$$

By the Riemann-Lebesgue lemma (282Fb),

$$\lim_{a\to\infty}\int_{-\infty}^{\infty}\frac{\sin au}{u}(f(x-u)-g(u))du=0.$$

If we now examine $\int\frac{\sin au}{u}g(u)du$, we get

$$\int_{-\infty}^{\infty}\frac{\sin au}{u}g(u)du=\int_{-1}^{1}\frac{\sin au}{u}f(x)du=f(x)\int_{-a}^{a}\frac{\sin v}{v}dv,$$

substituting $u=v/a$. So we get

$$\begin{aligned}\lim_{a\to\infty}\int_{-\infty}^{\infty}\frac{\sin au}{u}f(x-u)du&=\lim_{a\to\infty}\int_{-\infty}^{\infty}\frac{\sin au}{u}g(u)du\\&=\lim_{a\to\infty}f(x)\int_{-a}^{a}\frac{\sin v}{v}dv=\pi f(x),\end{aligned}$$

by 283D. Accordingly

$$\frac{1}{\sqrt{2\pi}}\lim_{a\to\infty}\int_{-a}^{a}e^{ixy}\hat{f}(y)dy=\frac{1}{\pi}\lim_{a\to\infty}\int_{-\infty}^{\infty}\frac{\sin au}{u}f(x-u)du=f(x),$$

using 283H. As for the second equality,

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\lim_{a\to\infty}\int_{-a}^{a}e^{-ixy}\check{f}(y)dy&=\frac{1}{\sqrt{2\pi}}\lim_{a\to\infty}\int_{-a}^{a}e^{-ixy}\hat{f}(-y)dy\\&=\frac{1}{\sqrt{2\pi}}\lim_{a\to\infty}\int_{-a}^{a}e^{ixu}\hat{f}(u)du=f(x)\end{aligned}$$

(substituting $y=-u$).

Remark Compare 282L.

283J Corollary Let $f:\mathbb{R}\to\mathbb{C}$ be an integrable function such that f is differentiable and $\hat{f}$ is integrable. Then $f=(\hat{f})^\vee=(\check{f})^\wedge$.

proof Because $\hat{f}$ is integrable,

$$\hat{f}^\vee(x)=\lim_{a\to\infty}\frac{1}{\sqrt{2\pi}}\int_{-a}^{a}e^{ixy}\hat{f}(y)dy=f(x)$$

for every $x\in\mathbb{R}$. Similarly,

$$\check{f}^\wedge(x)=\lim_{a\to\infty}\frac{1}{\sqrt{2\pi}}\int_{-a}^{a}e^{-ixy}\check{f}(y)dy=f(x).$$

283K The next proposition gives a class of functions to which the last corollary can be applied.

Proposition Suppose that f is a twice-differentiable function from $\mathbb{R}$ to $\mathbb{C}$ such that f, f' and f'' are all integrable. Then $\hat{f}$ is integrable.

proof Because f' and f'' are integrable, f and f' are absolutely continuous on any bounded interval (225L). So by 283Ci we have

$$(f'')^\wedge(y)=iy(f')^\wedge(y)=-y^2\hat{f}(y)$$

for every $y\in\mathbb{R}$. At the same time, by 283Cf-283Cg, $(f'')^\wedge$ and $\hat{f}$ must be bounded; say $|\hat{f}(y)|+|(f'')^\wedge(y)|\le K$ for every $y\in\mathbb{R}$. Now

$$|\hat{f}(y)|\le\frac{K}{1+y^2}$$

for every y, so that

$$\int_{-\infty}^{\infty}|\hat{f}|\le K\int_{-\infty}^{-1}\frac{1}{y^2}dy+2K+K\int_{1}^{\infty}\frac{1}{y^2}dy=4K<\infty.$$

Remark Compare 282Rb.

283L I turn now to the result corresponding to 282O, using a slightly different approach.

Theorem Let f be a complex-valued function which is integrable over $\mathbb{R}$, with Fourier transform $\hat{f}$ and inverse Fourier transform $\check{f}$, and suppose that f is of bounded variation on some neighbourhood of $x \in \mathbb{R}$. Set $a = \lim_{t\in\operatorname{dom} f, t\uparrow x} f(t)$, $b = \lim_{t\in\operatorname{dom} f, t\downarrow x} f(t)$. Then

$$\frac{1}{\sqrt{2\pi}} \lim_{\gamma\to\infty} \int_{-\gamma}^{\gamma} e^{ixy}\hat{f}(y)dy = \frac{1}{\sqrt{2\pi}} \lim_{\gamma\to\infty} \int_{-\gamma}^{\gamma} e^{-ixy}\check{f}(y)dy = \frac{1}{2}(a+b).$$

proof (a) The limits $\lim_{t\in\operatorname{dom} f, t\uparrow x} f(t)$ and $\lim_{t\in\operatorname{dom} f, t\downarrow x} f(t)$ exist because f is of bounded variation near x (224F). Recall from 283Db that there is a constant $K < \infty$ such that

$$|\int_{\gamma}^{\delta} \frac{\sin cx}{x} dx| \le K$$

whenever $\gamma \le \delta$ and $c \in \mathbb{R}$.

(b) Let $\epsilon > 0$. The hypothesis is that there is some $\delta > 0$ such that $\operatorname{Var}_{[x-\delta,x+\delta]}(f) < \infty$. Consequently

$$\lim_{\eta\downarrow 0} \operatorname{Var}_{]x,x+\eta]}(f) = \lim_{\eta\downarrow 0} \operatorname{Var}_{[x-\eta,x[}(f) = 0$$

(224E). There is therefore an $\eta > 0$ such that

$$\max(\operatorname{Var}_{[x-\eta,x[}(f), \operatorname{Var}_{]x,x+\eta]}(f)) \le \epsilon.$$

Of course

$$|f(t) - f(u)| \le \operatorname{Var}_{[x-\eta,x[}(f) \le \epsilon$$

whenever t, $u \in \operatorname{dom} f$ and $x - \eta \le t \le u < x$, so we shall have

$$|f(t) - a| \le \epsilon \text{ for every } t \in \operatorname{dom} f \cap [x-\eta, x[,$$

and similarly

$$|f(t) - b| \le \epsilon \text{ whenever } t \in \operatorname{dom} f \cap]x, x+\eta].$$

(c) Now set

$$g_1(t) = f(t) \text{ when } t \in \operatorname{dom} f \text{ and } |x-t| > \eta, \ 0 \text{ otherwise},$$

$$g_2(t) = a \text{ when } x-\eta \le t < x, \ b \text{ when } x < t \le x+\eta, \ 0 \text{ otherwise},$$

$$g_3 = f - g_1 - g_2.$$

Then $f = g_1 + g_2 + g_3$; each g_j is integrable; g_1 is zero on a neighbourhood of x;

$$\sup_{t\in\operatorname{dom} g_3, t\ne x} |g_3(t)| \le \epsilon,$$

$$\operatorname{Var}_{[x-\eta,x[}(g_3) \le \epsilon, \quad \operatorname{Var}_{]x,x+\eta]}(g_3) \le \epsilon.$$

(d) Consider the three parts g_1, g_2, g_3 separately.

(i) For the first, we have

$$\lim_{\gamma\to\infty} \frac{1}{\sqrt{2\pi}} \int_{-\gamma}^{\gamma} e^{ixy}\hat{g}_1(y)dy = 0$$

by 283I.

(ii) Next,

$$\frac{1}{\sqrt{2\pi}} \int_{-\gamma}^{\gamma} e^{ixy}\hat{g}_2(y)dy = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{\sin(x-t)\gamma}{x-t} g_2(t)dt$$

(by 283H)

$$= \frac{a}{\pi}\int_{x-\eta}^{x}\frac{\sin(x-t)\gamma}{x-t}dt + \frac{b}{\pi}\int_{x}^{x+\eta}\frac{\sin(x-t)\gamma}{x-t}dt$$

$$= \frac{a}{\pi}\int_{0}^{\gamma\eta}\frac{\sin u}{u}du + \frac{b}{\pi}\int_{0}^{\gamma\eta}\frac{\sin u}{u}du$$

(substituting $t = x - \frac{1}{\gamma}u$ in the first integral, $t = -x + \frac{1}{\gamma}u$ in the second)

$$\to \frac{a+b}{2} \text{ as } \gamma \to \infty$$

by 283Da.

(iii) As for the third, we have, for any $\gamma > 0$,

$$\begin{aligned}
\Bigl|\frac{1}{\sqrt{2\pi}}\int_{-\gamma}^{\gamma}e^{ixy}\hat{g}_3(y)dy\Bigr| &= \frac{1}{\pi}\Bigl|\int_{-\infty}^{\infty}\frac{\sin(x-t)\gamma}{x-t}g_3(t)dt\Bigr| = \frac{1}{\pi}\Bigl|\int_{-\infty}^{\infty}\frac{\sin t\gamma}{t}g_3(x-t)dt\Bigr| \\
&\le \frac{1}{\pi}\Bigl|\int_{-\eta}^{0}\frac{\sin t\gamma}{t}g_3(x-t)dt\Bigr| + \frac{1}{\pi}\Bigl|\int_{0}^{\eta}\frac{\sin t\gamma}{t}g_3(x-t)dt\Bigr| \\
&\le \frac{K}{\pi}\Bigl(\sup_{t\in\operatorname{dom} g_3\cap]x-\eta,x[}|g_3(t)| + \operatorname*{Var}_{]x-\eta,x[}(g_3) \\
&\qquad + \sup_{t\in\operatorname{dom} g_3\cap]x,x+\eta[}|g_3(t)| + \operatorname*{Var}_{]x,x+\eta[}(g_3)\Bigr) \\
&\le 4\epsilon\frac{K}{\pi},
\end{aligned}$$

using 224J to bound the integrals in terms of the variation and supremum of g_3 and integrals of $\frac{\sin\gamma t}{t}$ over subintervals.

(e) We therefore have

$$\begin{aligned}
&\limsup_{\gamma\to\infty}\Bigl|\frac{1}{\sqrt{2\pi}}\int_{-\gamma}^{\gamma}e^{ixy}\hat{f}(y)dy - \frac{a+b}{2}\Bigr| \\
&\le \limsup_{\gamma\to\infty}\frac{1}{\sqrt{2\pi}}\Bigl|\int_{-\gamma}^{\gamma}e^{ixy}\hat{g}_1(y)dy\Bigr| \\
&\qquad + \limsup_{\gamma\to\infty}\Bigl|\frac{1}{\sqrt{2\pi}}\int_{-\gamma}^{\gamma}e^{ixy}\hat{g}_2(y)dy - \frac{a+b}{2}\Bigr| \\
&\qquad + \limsup_{\gamma\to\infty}\frac{1}{\sqrt{2\pi}}\Bigl|\int_{-\gamma}^{\gamma}e^{ixy}\hat{g}_3ydy\Bigr| \\
&\le 0 + 0 + \frac{4K}{\pi}\epsilon
\end{aligned}$$

by the calculations in (d). As ϵ is arbitrary,

$$\lim_{\gamma\to\infty}\frac{1}{\sqrt{2\pi}}\int_{-\gamma}^{\gamma}e^{ixy}\hat{f}(y)dy - \frac{a+b}{2} = 0.$$

(f) This is the first half of the theorem. But of course the second half follows at once, because

$$\begin{aligned}
\frac{1}{\sqrt{2\pi}}\lim_{\gamma\to\infty}\int_{-\gamma}^{\gamma}e^{-ixy}\check{f}(y)dy &= \frac{1}{\sqrt{2\pi}}\lim_{\gamma\to\infty}\int_{-\gamma}^{\gamma}e^{-ixy}\hat{f}(-y)dy \\
&= \frac{1}{\sqrt{2\pi}}\lim_{\gamma\to\infty}\int_{-\gamma}^{\gamma}e^{ixy}\hat{f}(y)dy = \frac{a+b}{2}.
\end{aligned}$$

Remark You will see that this argument uses some of the same ideas as those in 282O-282P. It is more direct because (i) I am not using any concept corresponding to Fejér sums (though a very suitable one is available; see 283Xf) (ii) I do not trouble to give the result concerning uniform convergence of the Fourier integrals when f is continuous and of bounded variation (283Xj) (iii) I do not give any pointer to the significance of the fact that if f is of bounded variation then $\sup_{y\in\mathbb{R}}|y\hat{f}(y)| < \infty$ (283Xk).

283M Corresponding to 282Q, we have the following.

Theorem Let f and g be complex-valued functions which are integrable over $\mathbb{R}$, and $f*g$ their convolution product, defined by setting

$$(f*g)(x)=\int_{-\infty}^{\infty}f(t)g(x-t)dt$$

whenever this is defined (255E). Then

$$(f*g)^{\wedge}(y)=\sqrt{2\pi}\hat{f}(y)\hat{g}(y),\quad (f*g)^{\vee}(y)=\sqrt{2\pi}\check{f}(y)\check{g}(y)$$

for every $y\in\mathbb{R}$.

proof For any y,

$$(f*g)^{\wedge}(y)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-iyx}(f*g)(x)dx$$
$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-iy(t+u)}f(t)g(u)dtdu$$

(using 255G)

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{-iyt}f(t)dt\int_{-\infty}^{\infty}e^{-iyu}g(u)du=\sqrt{2\pi}\hat{f}(y)\hat{g}(y).$$

Now, of course,

$$(f*g)^{\vee}(y)=(f*g)^{\wedge}(-y)=\sqrt{2\pi}\hat{f}(-y)\hat{g}(-y)=\sqrt{2\pi}\check{f}(y)\check{g}(y).$$

283N I show how to compute a special Fourier transform, which will be used repeatedly in the next section.

Lemma For $\sigma>0$, set $\psi_\sigma(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}$ for $x\in\mathbb{R}$. Then its Fourier transform and inverse Fourier transform are

$$\hat{\psi}_\sigma=\check{\psi}_\sigma=\frac{1}{\sigma}\psi_{1/\sigma}.$$

In particular, $\hat{\psi}_1=\psi_1$.

proof (a) I begin with the special case $\sigma=1$, using the Maclaurin series

$$e^{-iyx}=\sum_{k=0}^{\infty}\frac{(-iyx)^k}{k!}$$

and the expressions for $\int_{-\infty}^{\infty}x^ke^{-x^2/2}dx$ from §263.

Fix $y\in\mathbb{R}$. Writing

$$g_k(x)=\frac{(-iyx)^k}{k!}e^{-x^2/2},\quad h_n(x)=\sum_{k=0}^{n}g_k(x),\qquad h(x)=e^{|yx|-x^2/2},$$

we see that

$$|g_k(x)|\le\frac{|yx|^k}{k!}e^{-x^2/2},$$

so that

$$|h_n(x)|\le\sum_{k=0}^{\infty}|g_k(x)|\le e^{|yx|}e^{-x^2/2}=h(x)$$

for every n; moreover, h is integrable, because $|h(x)|\le e^{-|x|}$ whenever $|x|\ge 2(1+|y|)$. Consequently, using Lebesgue's Dominated Convergence Theorem,

$$\hat{\psi}_1(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \lim_{n\to\infty} h_n = \frac{1}{2\pi}\lim_{n\to\infty}\int_{-\infty}^{\infty} h_n$$
$$= \frac{1}{2\pi}\sum_{k=0}^{\infty}\int_{-\infty}^{\infty} g_k = \frac{1}{2\pi}\sum_{k=0}^{\infty}\frac{(-iy)^k}{k!}\int_{-\infty}^{\infty} x^k e^{-x^2/2}dx$$
$$= \frac{1}{2\pi}\sum_{j=0}^{\infty}\frac{(-iy)^{2j}}{(2j)!}\frac{(2j)!}{2^j j!}\sqrt{2\pi}$$

(by 263H)

$$= \frac{1}{\sqrt{2\pi}}\sum_{j=0}^{\infty}\frac{(-y^2)^j}{2^j j!} = \frac{1}{\sqrt{2\pi}}e^{-y^2/2} = \psi_1(y),$$

as claimed.

(b) For the general case, $\psi_\sigma(x) = \frac{1}{\sigma}\psi_1(\frac{x}{\sigma})$, so that

$$\hat{\psi}_\sigma(y) = \frac{1}{\sigma}\cdot\sigma\hat{\psi}_1(\sigma y) = \frac{1}{\sigma}\hat{\psi}_{1/\sigma}(y)$$

by 283Ce. Of course we now have

$$\check{\psi}_\sigma(y) = \hat{\psi}_\sigma(-y) = \frac{1}{\sigma}\psi_{1/\sigma}(y)$$

because $\psi_{1/\sigma}$ is an even function.

283O To lead into the ideas of the next section, I give the following very simple fact.

Proposition Let f and g be two complex-valued functions which are integrable over $\mathbb{R}$. Then $\int_{-\infty}^{\infty} f\times\hat{g} = \int_{-\infty}^{\infty}\hat{f}\times g$ and $\int_{-\infty}^{\infty} f\times\check{g} = \int_{-\infty}^{\infty}\check{f}\times g$.

proof Of course

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|e^{-ixy}f(x)g(y)|dxdy = \int_{-\infty}^{\infty}|f|\int_{-\infty}^{\infty}|g| < \infty,$$

so

$$\int_{\infty}^{\infty} f\times\hat{g} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y)e^{-iyx}g(x)dxdy$$
$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y)e^{-ixy}g(x)dydx = \int_{-\infty}^{\infty}\hat{f}\times g.$$

For the other half of the proposition, replace every e^{-ixy} in the argument by e^{ixy}.

283W Higher dimensions I offer a series of exercises designed to provide hints on how the work of this section may be done in the r-dimensional case, where $r \ge 1$.

(a) Let f be an integrable complex-valued function defined almost everywhere in $\mathbb{R}^r$. Its **Fourier transform** is the function $\hat{f}:\mathbb{R}^r\to\mathbb{C}$ defined by the formula

$$\hat{f}(y) = \frac{1}{(\sqrt{2\pi})^r}\int e^{-iy\,.\,x}f(x)dx,$$

writing $y\,.\,x = \eta_1\xi_1 + \ldots + \eta_r\xi_r$ for $x = (\xi_1,\ldots,\xi_r)$ and $y = (\eta_1,\ldots,\eta_r)\in\mathbb{R}^r$, and $\int\ldots dx$ for integration with respect to Lebesgue measure on $\mathbb{R}^r$. Similarly, the **inverse Fourier transform** of f is the function $\check{f}$ given by

$$\check{f}(y) = \frac{1}{(\sqrt{2\pi})^r}\int e^{iy\,.\,x}f(x)dx = \hat{f}(-y).$$

Show that, for any integrable complex-valued function f on $\mathbb{R}^r$,

(i) $\hat{f}:\mathbb{R}^r\to\mathbb{C}$ is continuous;

(ii) $\lim_{\|y\|\to\infty} \hat{f}(y) = 0$, writing $\|y\| = \sqrt{y\,.\,y}$ as usual;

(iii) if $\int \|x\||f(x)|dx < \infty$, then $\hat{f}$ is differentiable, and

$$\frac{\partial}{\partial \eta_j}\hat{f}(y) = -\frac{i}{(\sqrt{2\pi})^r}\int e^{-iy\,.\,x}\xi_j f(x)dx$$

for $j \le r$, $y \in \mathbb{R}^r$, always taking ξ_j to be the jth coordinate of $x \in \mathbb{R}^r$;

(iv) if $j \le r$ and $\frac{\partial f}{\partial \xi_j}$ is defined everywhere and is integrable, and if $\lim_{\|x\|\to\infty} f(x) = 0$, then $(\frac{\partial f}{\partial \xi_j})^\wedge(y) = i\eta_j\hat{f}(y)$ for every $y \in \mathbb{R}^r$.

(b) Let f be an integrable complex-valued function on $\mathbb{R}^r$, and $\hat{f}$ its Fourier transform. If $c \le d$ in $\mathbb{R}^r$, show that

$$\int_{[c,d]} f = (\frac{i}{\sqrt{2\pi}})^r \lim_{\alpha_1,\dots,\alpha_r\to\infty}\int_{[-a,a]}\prod_{j=1}^{r}\frac{e^{i\gamma_j\eta_j} - e^{i\delta_j\eta_j}}{\eta_j}\hat{f}(y)dy,$$

setting $a = (\alpha_1,\dots)$, $c = (\gamma_1,\dots)$, $d = (\delta_1,\dots)$.

(c) Let f be an integrable complex-valued function on $\mathbb{R}^r$, and $\hat{f}$ its Fourier transform. Show that if we write

$$B_\infty(\mathbf{0}, a) = \{y : |\eta_j| \le a \text{ for every } j \le r\},$$

then

$$\frac{1}{(\sqrt{2\pi})^r}\int_{B_\infty(\mathbf{0},a)} e^{ix\,.\,y}\hat{f}(y)dy = \int \phi_a(t)f(x-t)dt$$

for every $a \ge 0$, where

$$\phi_a(t) = \frac{1}{\pi^r}\prod_{j=1}^{r}\frac{\sin a\tau_j}{\tau_j}$$

for $t = (\tau_1,\dots,\tau_r) \in \mathbb{R}^r$.

(d) Let f and g be integrable complex-valued functions on $\mathbb{R}^r$. Show that $f * \check{g} = (\sqrt{2\pi})^r(\hat{f} \times g)^\vee$.

(e) For $\sigma > 0$, define $\psi_\sigma : \mathbb{R}^r \to \mathbb{C}$ by setting

$$\psi_\sigma(x) = \frac{1}{(\sigma\sqrt{2\pi})^r}e^{-x\,.\,x/2\sigma^2}$$

for every $x \in \mathbb{R}^r$. Show that

$$\hat{\psi}_\sigma = \check{\psi}_\sigma = \frac{1}{\sigma^r}\psi_{1/\sigma}.$$

(f) Defining ψ_σ as in (e), show that $\lim_{\sigma\to 0}(f * \psi_\sigma)(x) = f(x)$ for every continuous integrable $f : \mathbb{R}^r \to \mathbb{C}$, $x \in \mathbb{R}^r$.

(g) Show that if $f : \mathbb{R}^r \to \mathbb{C}$ is continuous and integrable, and $\hat{f}$ also is integrable, then $f = \check{\hat{f}}$. (*Hint*: Show that both are equal at every point to

$$\lim_{\sigma\to\infty}(\sigma\sqrt{2\pi})^r(\hat{f} \times \psi_\sigma)^\vee = \lim_{\sigma\to\infty} f * \psi_{1/\sigma}.)$$

(h) Show that

$$\int_{\mathbb{R}^r}\frac{1}{1+\|x\|^{r+1}}dx < \infty.$$

(i) Show that if $f : \mathbb{R}^r \to \mathbb{C}$ can be partially differentiated $r+1$ times, and f and all its partial derivatives $\frac{\partial^k f}{\partial\xi_{j_1}\partial\xi_{j_2}\dots\partial\xi_{j_k}}$ are integrable for $k \le r+1$, then $\hat{f}$ is integrable.

(j) Show that if f and g are integrable complex-valued functions on $\mathbb{R}^r$, then $(f * g)^\wedge = (\sqrt{2\pi})^r\hat{f} \times \hat{g}$.

(k) Show that if f and g are integrable complex-valued functions on $\mathbb{R}^r$, then $\int f \times \hat{g} = \int \hat{f} \times g$.

(l) Show that if $f_1, \dots, f_r$ are integrable complex-valued functions on $\mathbb{R}$ with Fourier transforms $g_1, \dots, g_r$, and we write $f(x) = f_1(\xi_1) \dots f_r(\xi_r)$ for $x = (\xi_1, \dots, \xi_r) \in \mathbb{R}^r$, then the Fourier transform of f is $y \mapsto g_1(\eta_1) \dots g_r(\eta_r)$.

(m)(i) Show that $\int_{2k\pi}^{2(k+1)\pi} \frac{\sin t}{t\sqrt{t}} dt > 0$ for every $k \in \mathbb{N}$, and hence that $\int_0^\infty \frac{\sin t}{t\sqrt{t}} dt > 0$.

(ii) Set $f_1(\xi) = 1/\sqrt{|\xi|}$ for $0 < |\xi| \le 1$, 0 for other ξ. Show that $\lim_{a\to\infty} \frac{1}{\sqrt{a}} \int_{-a}^{a} \hat{f}_1(\eta) d\eta$ exists in $\mathbb{R}$ and is greater than 0.

(iii) Construct an integrable function f_2, zero on some neighbourhood of 0, such that there are infinitely many $m \in \mathbb{N}$ for which $|\int_{-m}^{m} \hat{f}_2(\eta) d\eta| \ge \frac{1}{\sqrt{m}}$. (*Hint*: take $f_2(\xi) = 2^{-k} \sin m_k \xi$ for $k+1 \le \xi < k+2$, for a sufficiently rapidly increasing sequence $\langle m_k \rangle_{k\in\mathbb{N}}$.)

(iv) Set $f(x) = f_1(\xi_1) f_2(\xi_2)$ for $x \in \mathbb{R}^2$. Show that f is integrable, that f is zero in a neighbourhood of $\mathbf{0}$, but that

$$\limsup_{a\to\infty} \frac{1}{2\pi} \Bigl|\int_{B_\infty(\mathbf{0},a)} \hat{f}(y) dy\Bigr| > 0,$$

defining B_∞ as in (c).

283X Basic exercises (a) Confirm that the six alternative definitions of the transforms $\hat{f}$, $\check{f}$ offered in 283B all lead to the same theory; find the constants involved in the new versions of 283Ch, 283Ci, 283L, 283M and 283N.

(b) If we redefined $\hat{f}(y)$ to be $\alpha \int_{-\infty}^{\infty} e^{i\beta xy} f(x) dx$, what would $\check{f}(y)$ be?

(c) Show that nearly every 2π would disappear from the theorems of this section if we defined a measure ν on $\mathbb{R}$ by saying that $\nu E = \frac{1}{\sqrt{2\pi}} \mu E$ for every Lebesgue measurable set E, where μ is Lebesgue measure, and wrote

$$\hat{f}(y) = \int_{-\infty}^{\infty} e^{-iyx} f(x) \nu(dx), \quad \check{f}(y) = \int_{-\infty}^{\infty} e^{iyx} f(x) \nu(dx),$$

$$(f * g)(x) = \int_{-\infty}^{\infty} f(t) g(x-t) \nu(dt).$$

What is $\lim_{a\to\infty} \int_{-a}^{a} \frac{\sin t}{t} \nu(dt)$?

>(d) Let f be an integrable complex-valued function on $\mathbb{R}$, with Fourier transform $\hat{f}$. Show that (i) if $g(x) = f(-x)$ whenever this is defined, then $\hat{g}(y) = \hat{f}(-y)$ for every $y \in \mathbb{R}$; (ii) if $g(x) = \overline{f(x)}$ whenever this is defined, then $\hat{g}(y) = \overline{\hat{f}(-y)}$ for every y.

(e) Let f be an integrable complex-valued function on $\mathbb{R}$, with Fourier transform $\hat{f}$. Show that

$$\int_c^d \hat{f}(y) dy = \frac{i}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{e^{-idx} - e^{-icx}}{x} f(x) dx$$

whenever $c \le d$ in $\mathbb{R}$.

>(f) For an integrable complex-valued function f on $\mathbb{R}$, let its **Fejér integrals** be

$$\sigma_c(x) = \frac{1}{c\sqrt{2\pi}} \int_0^c \Bigl(\int_{-a}^{a} e^{ixy} \hat{f}(y) dy \Bigr) da$$

for $c > 0$. Show that

$$\sigma_c(x) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1 - \cos ct}{ct^2} f(x-t) dt.$$

(g) Show that $\int_{-\infty}^{\infty} \frac{1-\cos at}{at^2} dt = \pi$ for every $a > 0$. (*Hint*: integrate by parts and use 283Da.) Show that

$$\lim_{a\to\infty} \int_\delta^\infty \frac{1-\cos at}{at^2} dt = \lim_{a\to\infty} \sup_{t\ge\delta} \frac{1-\cos at}{at^2} = 0$$

for every $\delta > 0$.

(h) Let f be an integrable complex-valued function on $\mathbb{R}$, and define its Fejér integrals σ_a as in 283Xf above. Show that if $x \in \mathbb{R}$, $c \in \mathbb{C}$ are such that

$$\lim_{\delta\downarrow 0} \frac{1}{\delta}\int_0^\delta |f(x+t)+f(x-t)-2c|dt = 0,$$

then $\lim_{a\to\infty} \sigma_a(x) = c$. (*Hint*: adapt the argument of 282H.)

>(i) Let f be an integrable complex-valued function on $\mathbb{R}$, and define its Fejér integrals σ_a as in 283Xf above. Show that $f(x) = \lim_{a\to\infty} \sigma_a(x)$ for almost every $x \in \mathbb{R}$.

(j) Let $f : \mathbb{R} \to \mathbb{C}$ be a continuous integrable complex-valued function of bounded variation, and define its Fejér integrals σ_a as in 283Xf above. Show that $f(x) = \lim_{a\to\infty} \sigma_a(x)$ uniformly for $x \in \mathbb{R}$.

>(k) Let f be an integrable complex-valued function of bounded variation on $\mathbb{R}$, and $\hat{f}$ its Fourier transform. Show that $\sup_{y\in\mathbb{R}} |y\hat{f}(y)| < \infty$.

(l) Let f and g be integrable complex-valued functions on $\mathbb{R}$. Show that $f * \check{g} = \sqrt{2\pi}(\hat{f} \times g)^\vee$.

(m) Let f be an integrable complex-valued function on $\mathbb{R}$, and fix $x \in \mathbb{R}$. Set

$$\hat{f}_x(y) = \int_{-\infty}^{\infty} f(t)\cos y(x-t)dt$$

for $y \in \mathbb{R}$. Show that

(i) if f is differentiable at x,

$$f(x) = \frac{1}{\pi}\lim_{a\to\infty}\int_0^a \tilde{f}_x(y)dy;$$

(ii) if there is a neighbourhood of x in which f has bounded variation, then

$$\frac{1}{\pi}\lim_{a\to\infty}\int_0^a \hat{f}_x(y)dy = \frac{1}{2}(\lim_{t\in\mathrm{dom}\,f,t\uparrow 0} f(t) + \lim_{t\in\mathrm{dom}\,f,t\downarrow 0} f(t));$$

(iii) if f is twice differentiable and f', f'' are integrable then $\hat{f}_x$ is integrable and $f(x) = \frac{1}{\pi}\int_0^\infty \hat{f}_x$. (The formula

$$f(x) = \frac{1}{\pi}\int_0^\infty \Bigl(\int_{-\infty}^\infty f(t)\cos y(x-t)dt\Bigr)dy,$$

valid for such functions f, is called **Fourier's integral formula**.)

(n) Show that if f is a complex-valued function of bounded variation, defined almost everywhere in $\mathbb{R}$, and converging to 0 at $\pm\infty$, then

$$g(y) = \frac{1}{\sqrt{2\pi}}\lim_{a\to\infty}\int_{-a}^a e^{-iyx}f(x)dx$$

is defined in $\mathbb{C}$ for every $y \neq 0$, and that the limit is uniform in any region bounded away from 0.

(o) Let f be an integrable complex-valued function on $\mathbb{R}$. Set

$$\hat{f}_c(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty \cos yx\, f(x)dx, \quad \hat{f}_s(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty \sin yx\, f(x)dx$$

for $y \in \mathbb{R}$. Show that

$$\frac{1}{\sqrt{2\pi}}\int_{-a}^a e^{ixy}\hat{f}(y)dy = \sqrt{\frac{2}{\pi}}\int_0^a \cos xy\,\hat{f}_c(y)dy + \sqrt{\frac{2}{\pi}}\int_0^a \sin xy\,\hat{f}_s(y)dy$$

for every $x \in \mathbb{R}$, $a \geq 0$.

(p) Use the fact that $\int_0^a\int_0^\infty e^{-xy}\sin y\,dxdy = \int_0^\infty\int_0^a e^{-xy}\sin y\,dydx$ whenever $a \geq 0$ to show that $\lim_{a\to\infty}\int_0^a \frac{\sin y}{y}dy = \int_0^\infty \frac{1}{1+x^2}dx$.

(q) Let $f : \mathbb{R} \to \mathbb{C}$ be an integrable function which is absolutely continuous on every bounded interval, and suppose that its derivative f' is of bounded variation on $\mathbb{R}$. Show that $\hat{f}$ is integrable and that $f = \hat{f}^\vee$. (*Hint*: 283Ci, 283Xk.)

>(r) Show that if $f(x) = e^{-\sigma|x|}$, where $\sigma > 0$, then $\hat{f}(y) = \frac{2\sigma}{\sqrt{2\pi}(\sigma^2+y^2)}$. Hence, or otherwise, find the Fourier transform of $y \mapsto \frac{1}{1+y^2}$.

(s) Find the inverse Fourier transform of the characteristic function of a bounded interval in $\mathbb{R}$. Show that in a formal sense 283F can be regarded as a special case of 283O.

(t) Let f be a non-negative integrable function on $\mathbb{R}$, with Fourier transform $\hat{f}$. Show that $\sum_{j=0}^{n}\sum_{k=0}^{n} a_j\bar{a}_k\hat{f}(y_j - y_k) \ge 0$ whenever $y_0, \dots, y_n$ in $\mathbb{R}$ and $a_0, \dots, a_n \in \mathbb{C}$.

(u) Let f be an integrable complex-valued function on $\mathbb{R}$. Show that $\tilde{f}(x) = \sum_{n=-\infty}^{\infty} f(x + 2\pi n)$ is defined in $\mathbb{C}$ for almost every x. (*Hint*: $\sum_{n=-\infty}^{\infty}\int_{-\pi}^{\pi}|f(x+2\pi n)|dx < \infty$.) Show that $\tilde{f}$ is periodic. Show that the Fourier coefficients of $\tilde{f}\restriction\,]-\pi,\pi]$ are $\langle\frac{1}{\sqrt{2\pi}}\hat{f}(k)\rangle_{k\in\mathbb{Z}}$.

283Y Further exercises (a) Show that if $f : \mathbb{R} \to \mathbb{C}$ is absolutely continuous in every bounded interval, f' is of bounded variation on $\mathbb{R}$, and $\lim_{x\to\infty} f(x) = \lim_{x\to-\infty} f(x) = 0$, then

$$g(y) = \frac{1}{\sqrt{2\pi}}\lim_{a\to\infty}\int_{-a}^{a} e^{-iyx}f(x)dx = -\frac{i}{y\sqrt{2\pi}}\lim_{a\to\infty}\int_{-a}^{a} e^{-iyx}f'(x)dx$$

is defined, with

$$y^2|g(y)| \le \frac{4}{\sqrt{2\pi}}\operatorname{Var}_{\mathbb{R}}(f'),$$

for every $y \ne 0$.

(b) Let $f : \mathbb{R} \to [0,\infty[$ be an even function such that f is convex on $[0,\infty[$ and $\lim_{x\to\infty} f(x) = 0$.

(i) Show that, for any $y > 0$, $k \in \mathbb{N}$, $\int_{-2k\pi/y}^{2k\pi/y} e^{-iyx}f(x)dx \ge 0$.

(ii) Show that $g(y) = \frac{1}{\sqrt{2\pi}}\lim_{a\to\infty}\int_{-a}^{a} e^{-iyx}f(x)dx$ exists in $[0,\infty[$ for every $y \ne 0$.

(iii) For $n \in \mathbb{N}$, set $f_n(x) = e^{-|x|/(n+1)}f(x)$ for every x. Show that f_n is integrable and convex on $[0,\infty[$.

(iv) Show that $g(y) = \lim_{n\to\infty}\hat{f}_n(y)$ for every $y \ne 0$.

(vi) Show that if f is integrable then

$$\int_{-a}^{a}\hat{f} = \frac{4}{\sqrt{2\pi}}\int_0^\infty \frac{\sin at}{t}f(t)dt \le \frac{4a}{\sqrt{2\pi}}\int_0^{\pi/a} f \le 2\sqrt{2\pi}f(1)$$

for every $a \ge 0$. Hence show that whether f is integrable or not, g is integrable and $f_n = (\hat{f}_n)^\vee$ for every n.

(vii) Show that $\lim_{a\downarrow 0}\sup_{n\in\mathbb{N}}\int_{-a}^{a}\hat{f}_n = 0$.

(viii) Show that if f' is bounded (on its domain) then $\{\hat{f}_n : n \in \mathbb{N}\}$ is uniformly integrable (*hint*: use (vii) and 283Ya), so that $\lim_{n\to\infty}\|\hat{f}_n - g\|_1 = 0$ and $f = \check{g}$.

(ix) Show that if f' is unbounded then for every $\epsilon > 0$ we can find h_1, $h_2 : \mathbb{R} \to [0,\infty[$, both even, convex and converging to 0 at ∞, such that $f = h_1 + h_2$, h_1' is bounded, $\int h_2 \le \epsilon$ and $h_2(1) \le \epsilon$. Hence show that in this case also $f = \check{g}$.

(c) Suppose that $f : \mathbb{R} \to \mathbb{R}$ is even, twice differentiable and convergent to 0 at ∞, that f'' is continuous and that $\{x : f''(x) = 0\}$ is bounded in $\mathbb{R}$. Show that f is the Fourier transform of an integrable function. (*Hint*: use 283Yb and 283Xq.)

(d) Let $g : \mathbb{R} \to \mathbb{R}$ be an odd function of bounded variation such that $\int_1^\infty \frac{1}{x}g(x)dx = \infty$. Show that g cannot be the Fourier transform of any integrable function f. (*Hint*: show that if $g = \hat{f}$ then

$$\int_0^1 f = \frac{2i}{\sqrt{2\pi}}\lim_{a\to\infty}\int_0^a \frac{1-\cos x}{x}g(x)dx = \infty.)$$

283 Notes and comments I have tried in this section to give the elementary theory of Fourier transforms of integrable functions on $\mathbb{R}$, with an eye to the extension of the concept which will be attempted in the next section. Following §282, I have given prominence to two theorems (283I and 283L) describing conditions for the inversion of the Fourier transform to return to the original function; we find ourselves looking at improper integrals $\lim_{a\to\infty}\int_{-a}^{a}$, just as earlier we needed to look at symmetric sums $\lim_{n\to\infty}\sum_{k=-n}^{n}$. I do not go quite so far as in §282, and in particular I leave the study of square-integrable functions for the moment, since their Fourier transforms may not be describable by the simple formulae used here.

One of the most fundamental obstacles in the subject is the lack of any effective criteria for determining which functions are the Fourier transforms of integrable functions. (Happily, things are better for square-integrable functions; see 284O-284P.) In 283Yb-283Yc I sketch an argument showing that 'ordinary' non-oscillating *even* functions which converge to 0 at $\pm\infty$ are Fourier transforms of integrable functions. Strikingly, this is not true of *odd* functions; thus $y\mapsto\frac{1}{\ln(e+y^2)}$ is the Fourier transform of an integrable function, but $y\mapsto\frac{\arctan y}{\ln(e+y^2)}$ is not (283Yd).

In 283W I sketch the corresponding theory of Fourier transforms in $\mathbb{R}^r$. There are few surprises. One point to note is that where in the one-dimensional case we ask for a well-behaved second derivative, in the r-dimensional case we may need to differentiate $r+1$ times (283Wi). Another is that we lose the 'localization principle'. In the one-dimensional case, if f is integrable and zero on an interval $]c,d[$, then $\lim_{a\to\infty}\int_{-a}^{a}e^{ixy}\hat{f}(y)dy=0$ for every $x\in\,]c,d[$; this is immediate from either 283I or 283L. But in higher dimensions the most natural formulation of a corresponding result is false (283Wm).

284 Fourier transforms II

The basic paradox of Fourier transforms is the fact that while for certain functions (see 283J-283K) we have $(\hat{f})^{\vee}=f$, 'ordinary' integrable functions f (for instance, the characteristic functions of non-trivial intervals) give rise to non-integrable Fourier transforms $\hat{f}$ for which there is no direct definition available for $\hat{f}^{\vee}$, making it a puzzle to decide in what sense the formula $f=\hat{f}^{\vee}$ might be true. What now seems by far the most natural resolution of the problem lies in declaring the Fourier transform to be an operation on *distributions* rather than on *functions*. I shall not attempt to describe this theory properly (almost any book on 'Distributions' will cover the ground better than I can possibly do here), but will try to convey the fundamental ideas, so far as they are relevant to the questions dealt with here, in language which will make the transition to a fuller treatment straightforward. At the same time, these methods make it easy to prove strong versions of the 'classical' theorems concerning Fourier transforms.

284A Test functions: Definition Throughout this section, a **rapidly decreasing test function** or **Schwartz function** will be a function $h:\mathbb{R}\to\mathbb{C}$ such that h is **smooth**, that is, differentiable everywhere any finite number of times, and moreover

$$\sup_{x\in\mathbb{R}}|x|^k|h^{(m)}(x)|<\infty$$

for all k, $m\in\mathbb{N}$, writing $h^{(m)}$ for the mth derivative of h.

284B The following elementary facts will be useful.

Lemma (a) If g and h are rapidly decreasing test functions, so are $g+h$ and ch, for any $c\in\mathbb{C}$.

(b) If h is a rapidly decreasing test function and $y\in\mathbb{R}$, then $x\mapsto h(y-x)$ is a rapidly decreasing test function.

(c) If h is any rapidly decreasing test function, then h and h^2 are integrable.

(d) If h is a rapidly decreasing test function, so is its derivative h'.

(e) If h is a rapidly decreasing test function, so is the function $x\mapsto xh(x)$.

(f) For any $\epsilon>0$, the function $x\mapsto e^{-\epsilon x^2}$ is a rapidly decreasing test function.

proof (a) is trivial.

(b) Write $g(x)=h(y-x)$ for $x\in\mathbb{R}$. Then $g^{(m)}(x)=(-1)^m h^{(m)}(y-x)$ for every m, so g is smooth. For any $k\in\mathbb{N}$,

$$|x|^k\le 2^k(|y|^k+|y-x|^k)$$

for every x, so

$$\begin{aligned}\sup_{x\in\mathbb{R}}|x|^k|g^{(m)}(x)|&=\sup_{x\in\mathbb{R}}|x|^k|h^{(m)}(y-x)|\\&\le 2^k|y|^k\sup_{x\in\mathbb{R}}|h^{(m)}(y-x)|+2^k\sup_{x\in\mathbb{R}}|y-x|^k|h^{(m)}(y-x)|\\&=2^k|y|^k\sup_{x\in\mathbb{R}}|h^{(m)}(x)|+2^k\sup_{x\in\mathbb{R}}|x|^k|h^{(m)}(x)|<\infty.\end{aligned}$$

(c) Because

$$M = \sup_{x\in\mathbb{R}} |h(x)| + x^2|h(x)|$$

is finite, we have

$$\int |h| \le \int \frac{M}{1+x^2} dx < \infty.$$

Of course we now have $|h^2| \le M|h|$, so h^2 also is integrable.

(d) This is immediate from the definition, as every derivative of h' is a derivative of h.

(e) Setting $g(x) = xh(x)$, $g^{(m)}(x) = xh^{(m)}(x) + mh^{(m-1)}(x)$ for $m \ge 1$, so

$$\sup_{x\in\mathbb{R}} |x^k g^{(m)}(x)| \le \sup_{x\in\mathbb{R}} |x^{k+1} h^{(m)}(x)| + m \sup_{x\in\mathbb{R}} |x^k h^{(m-1)}(x)|$$

is finite, for all $k \in \mathbb{N}$, $m \ge 1$.

(f) If $h(x) = e^{-\epsilon x^2}$, then for each $m \in \mathbb{N}$ we have $h^{(m)}(x) = p_m(x)h(x)$, where $p_0(x) = 1$ and $p_{m+1}(x) = p'_m(x) - 2\epsilon x p_m(x)$, so that p_m is a polynomial. Because $e^{\epsilon x^2} \ge \epsilon^{k+1}x^{2k+2}/(k+1)!$ for all x, $k \ge 0$,

$$\lim_{|x|\to\infty} |x|^k h(x) = \lim_{x\to\infty} x^k/e^{\epsilon x^2} = 0$$

for every k, and $\lim_{|x|\to\infty} p(x)h(x) = 0$ for every polynomial p; consequently

$$\lim_{|x|\to\infty} x^k h^{(m)}(x) = \lim_{|x|\to\infty} x^k p_m(x)h(x) = 0$$

for all k, m, and h is a rapidly decreasing test function.

284C Proposition Let $h : \mathbb{R} \to \mathbb{C}$ be a rapidly decreasing test function. Then $\hat{h} : \mathbb{R} \to \mathbb{C}$ and $\check{h} : \mathbb{R} \to \mathbb{C}$ are rapidly decreasing test functions, and $\check{\hat{h}} = \hat{\check{h}} = h$.

proof (a) Let k, $m \in \mathbb{N}$. Then $\sup_{x\in\mathbb{R}}(|x|^m + |x|^{m+2})|h^{(k)}(x)| < \infty$ and $\int_{-\infty}^{\infty} |x^m h^{(k)}(x)|dx < \infty$. We may therefore use 283Ch-283Ci to see that $y \mapsto i^{k+m}y^k\hat{h}^{(m)}(y)$ is the Fourier transform of $x \mapsto x^m h^{(k)}(x)$, and therefore that $\lim_{|y|\to\infty} y^k\hat{h}^{(m)}(y) = 0$, by 283Cg, so that (because $\hat{h}^{(m)}$ is continuous) $\sup_{y\in\mathbb{R}} |y^k\hat{h}^{(m)}(y)|$ is finite. As k and m are arbitrary, $\hat{h}$ is a rapidly decreasing test function.

(b) Since $\check{h}(y) = \hat{h}(-y)$ for every y, it follows at once that $\check{h}$ is a rapidly decreasing test function.

(c) By 283J, it follows from (a) and (b) that $\check{\hat{h}} = \hat{\check{h}} = h$.

284D Definition I will use the phrase **tempered function** on $\mathbb{R}$ to mean a measurable complex-valued function f, defined almost everywhere in $\mathbb{R}$, such that

$$\int_{-\infty}^{\infty} \frac{1}{1+|x|^k} |f(x)|dx < \infty$$

for some $k \in \mathbb{N}$.

284E As in 284B I spell out some elementary facts.

Lemma (a) If f and g are tempered functions, so are $|f|$, $f + g$ and cf, for any $c \in \mathbb{C}$.
(b) If f is a tempered function then it is integrable over any bounded interval.
(c) If f is a tempered function and $x \in \mathbb{R}$, then $t \mapsto f(x + t)$ and $t \mapsto f(x - t)$ are both tempered functions.

proof (a) is elementary; if

$$\int_{-\infty}^{\infty} \frac{1}{1+|x|^j} f(x)dx < \infty, \quad \int_{-\infty}^{\infty} \frac{1}{1+|x|^k} g(x)dx < \infty,$$

then

$$\int_{-\infty}^{\infty} \frac{1}{1+|x|^{j+k}} |(f + g)(x)| < \infty$$

because

$$1 + |x|^{j+k} \ge \max(1, |x|^{j+k}) \ge \max(1, |x|^j, |x|^k) \ge \frac{1}{2} \max(1 + |x|^j, 1 + |x|^k)$$

for all x.

(b) If

$$\int_{-\infty}^{\infty} \frac{1}{1+|x|^k} |f(x)| dx = M < \infty,$$

then for any $a \le b$

$$\int_a^b |f| \le M(1 + |a|^k + |b|^k)(b - a) < \infty.$$

(c) The idea is the same as in 284Bb. If $k \in \mathbb{N}$ is such that

$$\int_{-\infty}^{\infty} \frac{1}{1+|t|^k} |f(t)| dt = M < \infty,$$

then we have

$$1 + |x + t|^k \le 2^k (1 + |x|^k)(1 + |t|^k)$$

so that

$$\frac{1}{1+|t|^k} \le 2^k (1 + |x|^k) \frac{1}{1+|x+t|^k}$$

for every t, and

$$\int_{-\infty}^{\infty} \frac{|f(x+t)|}{1+|t|^k} dt \le 2^k (1 + |x|^k) \int_{-\infty}^{\infty} \frac{|f(x+t)|}{1+|x+t|^k} dt \le 2^k (1 + |x|^k) M < \infty.$$

Similarly,

$$\int_{-\infty}^{\infty} \frac{|f(x-t)|}{1+|t|^k} dt \le 2^k (1 + |x|^k) M < \infty.$$

284F Linking the two concepts, we have the following.

Lemma Let f be a tempered function on $\mathbb{R}$ and h a rapidly decreasing test function. Then $f \times h$ is integrable.

proof Of course $f \times h$ is measurable. Let $k \in \mathbb{N}$ be such that $\int_{-\infty}^{\infty} \frac{1}{1+|x|^k} |f(x)| dx < \infty$. There is a M such that $(1 + |x|^k)|h(x)| \le M$ for every $x \in \mathbb{R}$, so that

$$\int_{-\infty}^{\infty} |f \times h| \le M \int_{-\infty}^{\infty} \frac{1}{1+|x|^k} |f(x)| dx < \infty.$$

284G Lemma Suppose that f_1 and f_2 are tempered functions and that $\int f_1 \times h = \int f_2 \times h$ for every rapidly decreasing test function h. Then $f_1 =_{\text{a.e.}} f_2$.

proof (a) Set $g = f_1 - f_2$; then $\int g \times h = 0$ for every rapidly decreasing test function h. Of course g is a tempered function, so is integrable over any bounded interval. By 222D, it will be enough if I can show that $\int_a^b g = 0$ whenever $a < b$, since then we shall have $g = 0$ a.e. on every bounded interval and $f_1 =_{\text{a.e.}} f_2$.

(b) Consider the function $\phi_0(x) = e^{-1/x}$ for $x > 0$. Then ϕ_0 is differentiable arbitrarily often everywhere in $]0, \infty[$, $0 < \phi_0(x) < 1$ for every $x > 0$, and $\lim_{x\to\infty} \phi_0(x) = 1$. Moreover, writing $\phi_0^{(m)}$ for the mth derivative of ϕ_0,

$$\lim_{x\downarrow 0} \phi_0^{(m)}(x) = \lim_{x\downarrow 0} \frac{1}{x} \phi_0^{(m)}(x) = 0$$

for every $m \in \mathbb{N}$. **P** (Compare 284Bf.) We have $\phi_0^{(m)}(x) = p_m(\frac{1}{x})\phi_0(x)$, where $p_0(t) = 1$ and $p_{m+1}(t) = t^2(p_m(t) - p_m'(t))$, so that p_m is a polynomial for each $m \in \mathbb{N}$. Now for any $k \in \mathbb{N}$,

$$\lim_{t\to\infty} t^k e^{-t} \le \lim_{t\to\infty} \frac{(k+1)!t^k}{t^{k+1}} = 0,$$

so

$$\lim_{x\downarrow 0} \phi_0^{(m)}(x) = \lim_{t\to\infty} p_m(t)e^{-t} = 0,$$

$$\lim_{x\downarrow 0} \frac{1}{x} \phi_0^{(m)}(x) = \lim_{t\to\infty} t p_m(t)e^{-t} = 0. \text{ **Q**}$$

(c) Consequently, setting $\phi(x) = 0$ for $x \le 0$, $e^{-1/x}$ for $x > 0$, ϕ is smooth, with mth derivative

$$\phi^{(m)}(x) = 0 \text{ for } x \le 0, \quad \phi^{(m)}(x) = \phi_0^{(m)}(x) \text{ for } x > 0.$$

(The proof is an easy induction on m.) Also $0 \le \phi(x) \le 1$ for every $x \in \mathbb{R}$, and $\lim_{x\to\infty} \phi(x) = 1$.

(d) Now take any $a < b$, and for $n \in \mathbb{N}$ set

$$\psi_n(x) = \phi(n(x - a))\phi(n(b - x)).$$

Then ψ_n will be smooth and $\psi_n(x) = 0$ if $x \notin \,]a, b[$, so surely ψ_n is a rapidly decreasing test function, and

$$\int_{-\infty}^{\infty} g \times \psi_n = 0.$$

Next, $0 \le \psi_n(x) \le 1$ for every x, n, and if $a < x < b$ then $\lim_{n\to\infty} \psi_n(x) = 1$. So

$$\int_a^b g = \int g \times \chi(]a, b[) = \int g \times (\lim_{n\to\infty} \psi_n) = \lim_{n\to\infty} \int g \times \psi_n = 0,$$

using Lebesgue's Dominated Convergence Theorem. As a and b are arbitrary, $g = 0$ a.e., as required.

284H Definition Let f and g be tempered functions in the sense of 284D. Then I will say that g **represents the Fourier transform of** f if

$$\int_{-\infty}^{\infty} g \times h = \int_{-\infty}^{\infty} f \times \hat{h}$$

for every rapidly decreasing test function h.

284I Remarks (a) As usual, when shifting definitions in this way, we have some checking to do. If f is an integrable complex-valued function on $\mathbb{R}$, $\hat{f}$ its Fourier transform, then surely $\hat{f}$ is a tempered function, being a bounded continuous function; and if h is any rapidly decreasing test function, then $\int \hat{f} \times h = \int f \times \hat{h}$ by 283O. Thus $\hat{f}$ 'represents the Fourier transform of f' in the sense of 284H above.

(b) Note also that 284G assures us that if g_1, g_2 are two tempered functions both representing the Fourier transform of f, then $g_1 =_{\text{a.e.}} g_2$, since we must have

$$\int g_1 \times h = \int f \times \hat{h} = \int g_2 \times h$$

for every rapidly decreasing test function h.

(c) Of course the value of this indirect approach is that we can assign Fourier transforms, in a sense, to many more functions. But we must note at once that if g 'represents the Fourier transform of f' then so will any function equal almost everywhere to g; we can no longer expect to be able to speak of 'the' Fourier transform of f as a function. We could say that 'the' Fourier transform of f is a functional ϕ on the space of rapidly decreasing test functions, defined by setting $\phi(h) = \int f \times \hat{h}$; alternatively, we could say that 'the' Fourier transform of f is a member of $L^0_{\mathbb{C}}$, the space of equivalence classes of almost-everywhere-defined measurable functions (241J).

(d) It is now natural to say that g **represents the inverse Fourier transform of** f just when f represents the Fourier transform of g; that is, when $\int f \times h = \int g \times \hat{h}$ for every rapidly decreasing test function h. Because $\check{\hat{h}} = \hat{\check{h}} = h$ for every such h, this is the same thing as saying that $\int f \times \check{h} = \int g \times h$ for every rapidly decreasing test function h, which is the other natural expression of what it might mean to say that 'g represents the inverse Fourier transform of f'.

(e) If f, g are tempered functions and we write $\overleftrightarrow{g}(x) = g(-x)$ whenever this is defined, then $\overleftrightarrow{g}$ will also be a tempered function, and we shall always have

$$\int \overleftrightarrow{g} \times \hat{h} = \int g(-x)\hat{h}(x)dx = \int g(x)\hat{h}(-x)dx = \int g \times \check{h},$$

so that

g represents the Fourier transform of f

$\iff \int g \times h = \int f \times \hat{h}$ for every test function h

$\iff \int g \times \check{h} = \int f \times \hat{\check{h}}$ for every h

$\iff \int \overleftrightarrow{g} \times \hat{h} = \int f \times h$ for every h

$\iff \overset{\leftrightarrow}{g}$ represents the inverse Fourier transform of f.

Combining this with (d), we get

g represents the Fourier transform of f

$\iff \overset{\leftrightarrow}{\overset{\leftrightarrow}{f}} = f$ represents the inverse Fourier transform of g

$\iff \overset{\leftrightarrow}{f}$ represents the Fourier transform of g.

(f) Yet again, I ought to spell out the check: if f is integrable and $\check{f}$ is its inverse Fourier transform as defined in 283Ab, then

$$\int \check{f} \times \hat{h} = \int f \times \check{\hat{h}} = \int f \times h$$

for every rapidly decreasing test function h, so $\check{f}$ 'represents the inverse Fourier transform of f' in the sense given here.

284J Lemma Let f be any tempered function and h a rapidly decreasing test function. Then $f * h$, defined by the formula

$$(f * h)(y) = \int_{-\infty}^{\infty} f(t)h(y-t)dt,$$

is defined everywhere.

proof Take any $y \in \mathbb{R}$. By 284Bb, $t \mapsto h(y-t)$ is a rapidly decreasing test function, so the integral is always defined in $\mathbb{C}$, by 284F.

284K Proposition Let f and g be tempered functions such that g represents the Fourier transform of f, and h a rapidly decreasing test function.

(a) The Fourier transform of the integrable function $f \times h$ is $\frac{1}{\sqrt{2\pi}} g * \hat{h}$, where $g * \hat{h}$ is the convolution of g and $\hat{h}$.

(b) The Fourier transform of the continuous function $f * h$ is represented by the product $\sqrt{2\pi} g \times \hat{h}$.

proof (a) Of course $f \times h$ is integrable, by 284F, while $g * \hat{h}$ is defined everywhere, by 284C and 284J.

Fix $y \in \mathbb{R}$. Set $\phi(x) = \hat{h}(y-x)$ for $x \in \mathbb{R}$; then ϕ is a rapidly decreasing test function because $\hat{h}$ is (284Bb). Now

$$\begin{aligned}\hat{\phi}(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-itx} \hat{h}(y-x)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-it(y-x)} \hat{h}(x)dx \\ &= \frac{1}{\sqrt{2\pi}} e^{-ity} \int_{-\infty}^{\infty} e^{itx} \hat{h}(x)dx = e^{-ity} \check{\hat{h}}(t) = e^{-ity} h(t),\end{aligned}$$

using 284C. Accordingly

$$\begin{aligned}(f \times h)^{\wedge}(y) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-ity} f(t)h(t)dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)\hat{\phi}(t)dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(t)\phi(t)dt\end{aligned}$$

(because g represents the Fourier transform of f)

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} g(t)\hat{h}(y-t)dt = \frac{1}{\sqrt{2\pi}} (g * \hat{h})(y).$$

As y is arbitrary, $\frac{1}{\sqrt{2\pi}} g * \hat{h}$ is the Fourier transform of $f \times h$.

(b) Write ψ for the Fourier transform of $g \times \hat{h}$, $\overset{\leftrightarrow}{f}(x) = f(-x)$ when this is defined, and $\overset{\leftrightarrow}{h}(x) = h(-x)$ for every x, so that $\overset{\leftrightarrow}{f}$ represents the Fourier transform of g, by 284Ie, and $\overset{\leftrightarrow}{h}$ is the Fourier transform of $\hat{h}$. By (a), we have $\psi = \frac{1}{\sqrt{2\pi}} \overset{\leftrightarrow}{f} * \overset{\leftrightarrow}{h}$. This means that the inverse Fourier transform of $\sqrt{2\pi} g \times \hat{h}$ must be $\sqrt{2\pi}\overset{\leftrightarrow}{\psi} = (\overset{\leftrightarrow}{f} * \overset{\leftrightarrow}{h})^{\leftrightarrow}$; and as

$$\begin{aligned}(\overset{\leftrightarrow}{f} * \overset{\leftrightarrow}{h})^{\leftrightarrow}(y) &= (\overset{\leftrightarrow}{f} * \overset{\leftrightarrow}{h})(-y)\\ &= \int_{-\infty}^{\infty} \overset{\leftrightarrow}{f}(t)\overset{\leftrightarrow}{h}(-y-t)dt\\ &= \int_{-\infty}^{\infty} f(-t)h(y+t)dt\\ &= \int_{-\infty}^{\infty} f(t)h(y-t)dt = (f*h)(y),\end{aligned}$$

the inverse Fourier transform of $\sqrt{2\pi}g \times \hat{h}$ is $f * h$ (which is therefore continuous), and $\sqrt{2\pi}g \times \hat{h}$ must represent the Fourier transform of $f * h$.

Remark Compare 283M. It is typical of the theory of Fourier transforms that we have formulae valid in a wide variety of contexts, each requiring a different interpretation and a different proof.

284L We are now ready for a result corresponding to 282H. I use a different method, or at least a different arrangement of the ideas, through the following fact, which is important in other ways.

Proposition Let f be any tempered function. Writing $\psi_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}$ for $x \in \mathbb{R}$ and $\sigma > 0$, then

$$\lim_{\sigma\downarrow 0}(f * \psi_\sigma)(x) = c$$

whenever $x \in \mathbb{R}$ and $c \in \mathbb{C}$ are such that

$$\lim_{\delta\downarrow 0}\frac{1}{\delta}\int_0^\delta |f(x+t) + f(x-t) - 2c|dt = 0.$$

proof (a) By 284Bf, every ψ_σ is a rapidly decreasing test function, so that $f * \psi_\sigma$ is defined everywhere, by 284J. We need to know that $\int_{-\infty}^{\infty} \psi_\sigma = 1$; this is because (substituting $u = x/\sigma$)

$$\int_{-\infty}^{\infty} \psi_\sigma = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-u^2/2}du = 1,$$

by 263G. The argument now follows the lines of 282H. Set

$$\phi(t) = |f(x+t) + f(x-t) - 2c|$$

when this is defined, which is almost everywhere, and $\Phi(t) = \int_0^t \phi$, defined for all $t \geq 0$ because f is integrable over every bounded interval (284Eb). We have

$$\begin{aligned}|(f * \psi_\sigma)(x) - c| &= |\int_{-\infty}^{\infty} f(x-t)\psi_\sigma(t)dt - c\int_{-\infty}^{\infty} \psi_\sigma(t)dt|\\ &= |\int_{-\infty}^{0} (f(x-t)-c)\psi_\sigma(t)dt + \int_0^{\infty} (f(x-t)-c)\psi_\sigma(t)dt|\\ &= |\int_0^{\infty} (f(x+t)-c)\psi_\sigma(t)dt + \int_0^{\infty} (f(x-t)-c)\psi_\sigma(t)dt|\end{aligned}$$

(because ψ_σ is an even function)

$$\begin{aligned}&= |\int_0^{\infty} (f(x+t)+f(x-t)-2c)\psi_\sigma(t)dt|\\ &\leq \int_0^{\infty} |f(x+t)+f(x-t)-2c|\psi_\sigma(t)dt = \int_0^{\infty} \phi \times \psi_\sigma.\end{aligned}$$

(b) I should explain why this last integral is finite. Because f is a tempered function, so are the functions $t \mapsto f(x+t)$, $t \mapsto f(x-t)$ (284Ec); of course constant functions are tempered, so $t \mapsto \phi(t) = |f(x+t) + f(x-t) - 2c|$ is tempered, and because ψ_σ is a rapidly decreasing test function we may apply 284F to see that the product is integrable.

(c) Let $\epsilon > 0$. By hypothesis, $\lim_{t\downarrow 0}\Phi(t)/t = 0$; let $\delta > 0$ be such that $\Phi(t) \leq \epsilon t$ for every $t \in [0, \delta]$. Take any $\sigma \in \,]0, \delta]$. I break the integral $\int_0^\infty \phi \times \psi_\sigma$ up into three parts.

(i) For the integral from 0 to σ, we have

$$\int_0^\sigma \phi\times\psi_\sigma \le \int_0^\sigma \frac{1}{\sigma\sqrt{2\pi}}\phi = \frac{1}{\sigma\sqrt{2\pi}}\Phi(\sigma) \le \frac{\epsilon\sigma}{\sigma\sqrt{2\pi}} \le \epsilon,$$

because $\psi_\sigma(t) \le \frac{1}{\sigma\sqrt{2\pi}}$ for every t.

(ii) For the integral from σ to δ, we have

$$\int_\sigma^\delta \phi\times\psi_\sigma \le \frac{1}{\sigma\sqrt{2\pi}}\int_\sigma^\delta \phi(t)\frac{2\sigma^2}{t^2}dt$$

(because $e^{-t^2/2\sigma^2} = 1/e^{t^2/2\sigma^2} \le 1/(t^2/2\sigma^2) = 2\sigma^2/t^2$ for every $t \ne 0$)

$$= \sigma\sqrt{\frac{2}{\pi}}\int_\sigma^\delta \frac{\phi(t)}{t^2}dt = \sigma\sqrt{\frac{2}{\pi}}\Bigl(\frac{\Phi(\delta)}{\delta^2} - \frac{\Phi(\sigma)}{\sigma} + \int_\sigma^\delta \frac{2\Phi(t)}{t^3}dt\Bigr)$$

(integrating by parts – see 225F)

$$\le \sigma\Bigl(\frac{\epsilon}{\delta} + \int_\sigma^\delta \frac{2\epsilon}{t^2}dt\Bigr)$$

(because $\Phi(t) \le \epsilon t$ for $0 \le t \le \delta$ and $\sqrt{2/\pi} \le 1$)

$$\le \sigma\bigl(\frac{\epsilon}{\delta} + \frac{2\epsilon}{\sigma}\bigr) \le 3\epsilon.$$

(iii) For the integral from δ to ∞, we have

$$\int_\delta^\infty \phi\times\psi_\sigma = \frac{1}{\sqrt{2\pi}}\int_\delta^\infty \phi(t)\frac{e^{-t^2/2\sigma^2}}{\sigma}dt.$$

Now for any $t \ge \delta$,

$$\sigma \mapsto \frac{1}{\sigma}e^{-t^2/2\sigma^2} :]0,\delta] \to \mathbb{R}$$

is monotonically increasing, because its derivative

$$\frac{d}{d\sigma}\frac{1}{\sigma}e^{-t^2/2\sigma^2} = \frac{1}{\sigma^2}\bigl(\frac{t^2}{\sigma^2} - 1\bigr)e^{-t^2/2\sigma^2}$$

is positive, and

$$\lim_{\sigma\downarrow 0}\frac{1}{\sigma}e^{-t^2/2\sigma^2} = \lim_{a\to\infty} ae^{-a^2t^2/2} = 0.$$

So we may apply Lebesgue's Dominated Convergence Theorem to see that

$$\lim_{n\to\infty}\int_\delta^\infty \phi(t)\frac{e^{-t^2/2\sigma_n^2}}{\sigma_n}dt = 0$$

whenever $\langle\sigma_n\rangle_{n\in\mathbb{N}}$ is a sequence in $]0,\delta]$ converging to 0, so that

$$\lim_{\sigma\downarrow 0}\int_\delta^\infty \phi(t)\frac{e^{-t^2/2\sigma^2}}{\sigma}dt = 0.$$

There must therefore be a $\sigma_0 \in]0,\delta]$ such that

$$\int_\delta^\infty \phi\times\psi_\sigma \le \epsilon$$

for every $\sigma \le \sigma_0$.

Putting these together, we see that

$$|(f*\psi_\sigma)(x) - c| \le \int_0^\infty \phi\times\psi_\sigma \le \epsilon + 3\epsilon + \epsilon = 5\epsilon$$

whenever $0 < \sigma \le \sigma_0$. As ϵ is arbitrary, $\lim_{\sigma\downarrow 0}(f*\psi_\sigma)(x) = c$, as claimed.

284M Theorem Let f and g be tempered functions such that g represents the Fourier transform of f. Then

(a)(i) $g(y) = \lim_{\epsilon\downarrow 0} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-iyx} e^{-\epsilon x^2} f(x)dx$ for almost every $y \in \mathbb{R}$.

(ii) If $y \in \mathbb{R}$ is such that $a = \lim_{t\in\mathrm{dom}\, g, t\uparrow y} g(t)$ and $b = \lim_{t\in\mathrm{dom}\, g, t\downarrow y} g(t)$ are both defined in $\mathbb{C}$, then

$$\lim_{\epsilon\downarrow 0} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-iyx} e^{-\epsilon x^2} f(x)dx = \frac{1}{2}(a+b).$$

(b)(i) $f(x) = \lim_{\epsilon\downarrow 0} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ixy} e^{-\epsilon y^2} g(y)dy$ for almost every $x \in \mathbb{R}$.

(ii) If $x \in \mathbb{R}$ is such that $a = \lim_{t\in\mathrm{dom}\, f, t\uparrow x} f(t)$ and $b = \lim_{t\in\mathrm{dom}\, f, t\downarrow x} f(t)$ are both defined in $\mathbb{C}$, then

$$\lim_{\epsilon\downarrow 0} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ixy} e^{-\epsilon y^2} g(y)dy = \frac{1}{2}(a+b).$$

proof (a)(i) By 223D,

$$\lim_{\delta\downarrow 0} \frac{1}{2\delta} \int_{-\delta}^{\delta} |g(y+t) - g(y)|dt = 0$$

for almost every $y \in \mathbb{R}$, because g is integrable over any bounded interval. Fix any such y. Set $\phi(t) = |g(y+t) + g(y-t) - 2g(y)|$ whenever this is defined. Then, as in 282Ia,

$$\int_0^{\delta} \phi \le \int_{-\delta}^{\delta} |g(y+t) - g(y)|dt,$$

so $\lim_{\delta\downarrow 0} \frac{1}{\delta} \int_0^{\delta} \phi = 0$. Consequently, by 284L,

$$g(y) = \lim_{\sigma\to\infty} (g * \psi_{1/\sigma})(y).$$

We know from 283N that the Fourier transform of ψ_σ is $\frac{1}{\sigma}\psi_{1/\sigma}$ for any $\sigma > 0$. Accordingly, by 284K, $g * \psi_{1/\sigma}$ is the Fourier transform of $\sigma\sqrt{2\pi} f \times \psi_\sigma$, that is,

$$(g * \psi_{1/\sigma})(y) = \int_{-\infty}^{\infty} e^{-iyx} \sigma\psi_\sigma(x) f(x)dx.$$

So

$$\begin{aligned} g(y) &= \lim_{\sigma\to\infty} \int_{-\infty}^{\infty} e^{-iyx} \sigma\psi_\sigma(x) f(x)dx \\ &= \lim_{\sigma\to\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-iyx} e^{-x^2/2\sigma^2} f(x)dx \\ &= \lim_{\epsilon\downarrow 0} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-iyx} e^{-\epsilon x^2} f(x)dx. \end{aligned}$$

And this is true for almost every y.

(ii) Again, setting $c = \frac{1}{2}(a+b)$, $\phi(t) = |g(y+t) + g(y-t) - 2c|$ whenever this is defined, we have $\lim_{t\in\mathrm{dom}\,\phi, t\downarrow 0} \phi(t) = 0$, so of course $\lim_{\delta\downarrow 0} \frac{1}{\delta} \int_0^{\delta} \phi = 0$, and

$$c = \lim_{\sigma\to\infty} (g * \psi_{1/\sigma})(y) = \lim_{\epsilon\downarrow 0} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-iyx} e^{-\epsilon x^2} f(x)dx$$

as before.

(b) This can be shown by similar arguments; or it may be actually deduced from (a), by observing that $x \mapsto \overleftrightarrow{f}(x) = f(-x)$ represents the Fourier transform of g (see 284Id), and applying (a) to g and $\overleftrightarrow{f}$.

284N L^2 spaces We are now ready for results corresponding to 282J-282K.

Lemma Let $\mathcal{L}^2_{\mathbb{C}}$ be the space of square-integrable complex-valued functions on $\mathbb{R}$, and $\mathcal{S}$ the space of rapidly decreasing test functions. Then for every $f \in \mathcal{L}^2_{\mathbb{C}}$ and $\epsilon > 0$ there is an $h \in \mathcal{S}$ such that $\|f - h\|_2 \le \epsilon$.

proof Set $\phi(x) = e^{-1/x}$ for $x > 0$, zero for $x \le 0$; recall from the proof of 284G that ϕ is smooth. For any $a < b$, the functions

$$x \mapsto \psi_n(x) = \phi(n(x-a))\phi(n(b-x))$$

provide a sequence of test functions converging to $\chi\,]a,b[$ from below, so (as in 284G)

$$\inf_{h\in\mathcal{S}} \|\chi\,]a,b[-h\|_2^2 \le \lim_{n\to\infty} \int_a^b |1-\psi_n|^2 = 0.$$

Because $\mathcal{S}$ is a linear space (284Ba), it follows that for every step-function g with bounded support and every $\epsilon > 0$ there is an $h \in \mathcal{S}$ such that $\|g-h\|_2 \le \frac{1}{2}\epsilon$. But we know from 244H that for every $f \in \mathcal{L}^2_{\mathbb{C}}$ and $\epsilon > 0$ there is a step-function g with bounded support such that $\|f-g\|_2 \le \frac{1}{2}\epsilon$; so there must be an $h \in \mathcal{S}$ such that

$$\|f-h\|_2 \le \|f-g\|_2 + \|g-h\|_2 \le \epsilon.$$

As f and ϵ are arbitrary, we have the result.

284O Theorem (a) Let f be any complex-valued function which is square-integrable over $\mathbb{R}$. Then f is a tempered function and its Fourier transform is represented by another square-integrable function g, and $\|g\|_2 = \|f\|_2$.

(b) If f_1 and f_2 are complex-valued functions, square-integrable over $\mathbb{R}$, with Fourier transforms represented by functions g_1, g_2, then

$$\int_{-\infty}^{\infty} f_1 \times \bar{f}_2 = \int_{-\infty}^{\infty} g_1 \times \bar{g}_2.$$

(c) If f_1 and f_2 are complex-valued functions, square-integrable over $\mathbb{R}$, with Fourier transforms represented by functions g_1, g_2, then the integrable function $f_1 \times f_2$ has Fourier transform $\frac{1}{\sqrt{2\pi}} g_1 * g_2$.

(d) If f_1 and f_2 are complex-valued functions, square-integrable over $\mathbb{R}$, with Fourier transforms represented by functions g_1, g_2, then $\sqrt{2\pi} g_1 \times g_2$ represents the Fourier transform of the continuous function $f_1 * f_2$.

proof (a)(i) Consider first the case in which f is a rapidly decreasing test function and g is its Fourier transform; we know that g is also a rapidly decreasing test function, and that f is the inverse Fourier transform of g (284C). Now the complex conjugate $\overline{g}$ of g is given by the formula

$$\overline{g}(y) = \overline{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-iyx} f(x)dx} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{iyx} \overline{f}(x)dx,$$

so that $\overline{g}$ is the inverse Fourier transform of $\overline{f}$. Accordingly

$$\int f \times \overline{f} = \int \check{g} \times \overline{f} = \int g \times \check{\overline{f}} = \int g \times \overline{g},$$

using 283O for the middle equality.

(ii) Now suppose that $f \in \mathcal{L}^2_{\mathbb{C}}$. I said that f is a tempered function; this is simply because

$$\int_{-\infty}^{\infty} \Bigl(\frac{1}{1+|x|}\Bigr)^2 dx < \infty,$$

so

$$\int_{-\infty}^{\infty} \frac{|f(x)|}{1+|x|} dx < \infty$$

(244Eb). By 284N, there is a sequence $\langle f_n\rangle_{n\in\mathbb{N}}$ of rapidly decreasing test functions such that $\lim_{n\to\infty} \|f-f_n\|_2 = 0$. By (i),

$$\lim_{m,n\to\infty} \|\hat{f}_m - \hat{f}_n\|_2 = \lim_{m,n\to\infty} \|f_m - f_n\|_2 = 0,$$

and the sequence $\langle \hat{f}^\bullet_n\rangle_{n\in\mathbb{N}}$ of equivalence classes is a Cauchy sequence in $L^2_{\mathbb{C}}$. Because $L^2_{\mathbb{C}}$ is complete (244G), $\langle \hat{f}^\bullet_n\rangle_{n\in\mathbb{N}}$ has a limit in $L^2_{\mathbb{C}}$, which is representable as $g^\bullet$ for some $g \in \mathcal{L}^2_{\mathbb{C}}$. Like f, g must be a tempered function. Of course

$$\|g\|_2 = \lim_{n\to\infty} \|\hat{f}_n\|_2 = \lim_{n\to\infty} \|f_n\|_2 = \|f\|_2.$$

Now if h is any rapidly decreasing test function, $h \in \mathcal{L}^2_{\mathbb{C}}$ (284Bc), so we shall have

$$\int g \times h = \lim_{n\to\infty} \int \hat{f}_n \times h = \lim_{n\to\infty} \int f_n \times \hat{h} = \int f \times \hat{h}.$$

So g represents the Fourier transform of f.

(b) Of course any functions representing the Fourier transforms of f_1 and f_2 must be equal almost everywhere to square-integrable functions, and therefore square-integrable, with the right norms. It follows as in 282K (part (d) of the proof) that if g_1, g_2 represent the Fourier transforms of f_1, f_2, so that $ag_1 + bg_2$ represents the Fourier transform of $af_1 + bf_2$ and $\|ag_1 + bg_2\|_2 = \|af_1 + bf_2\|_2$ for all a, $b \in \mathbb{C}$, we must have

$$\int f_1 \times \overline{f}_2 = (f_1|f_2) = (g_1|g_2) = \int g_1 \times \overline{g}_2.$$

(c) Of course $f_1 \times f_2$ is integrable because it is the product of two square-integrable functions (244E).

(i) Let $y \in \mathbb{R}$ and set $f(x) = \overline{f_2(x)e^{iyx}}$ for $x \in \mathbb{R}$. Then $f \in \mathcal{L}^2_{\mathbb{C}}$. We need to know that the Fourier transform of f is represented by g, where $g(u) = \overline{g_2(y-u)}$. **P** Let h be a rapidly decreasing test function. Then

$$\begin{aligned}\int g \times h &= \int \overline{g_2(y-u)}h(u)du = \int \overline{g_2(u)}h(y-u)du \\ &= \overline{\int g_2 \times h_1} = \overline{\int f_2 \times \hat{h}_1},\end{aligned}$$

where $h_1(u) = \overline{h(y-u)}$. To compute $\hat{h}_1$, we have

$$\begin{aligned}\hat{h}_1(v) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-ivu}h_1(u)du = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-ivu}\overline{h(y-u)}du \\ &= \overline{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{ivu}h(y-u)du} = \overline{\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{iv(y-u)}h(u)du} = \overline{e^{ivy}\hat{h}(v)}.\end{aligned}$$

So

$$\begin{aligned}\int g \times h &= \overline{\int f_2 \times \hat{h}_1} = \int \overline{f_2(v)\hat{h}_1(v)}dv \\ &= \int \overline{f_2(v)}e^{ivy}\hat{h}(v)dv = \int f \times \hat{h} :\end{aligned}$$

as h is arbitrary, g represents the Fourier transform of f. **Q**

(ii) We now have

$$\begin{aligned}(f_1 \times f_2)^\wedge(y) &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-iyx}f_1(x)f_2(x)dx \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f_1 \times \bar{f} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g_1 \times \bar{g}\end{aligned}$$

(using part (b))

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g_1(u)g_2(y-u)du = \frac{1}{\sqrt{2\pi}}(g_1 * g_2)(y).$$

As y is arbitrary, $(f_1 \times f_2)^\wedge = \frac{1}{\sqrt{2\pi}}g_1 * g_2$, as claimed.

(d) By (c), the Fourier transform of $\sqrt{2\pi}g_1 \times g_2$ is $\overset{\leftrightarrow}{f}_1 * \overset{\leftrightarrow}{f}_2$, writing $\overset{\leftrightarrow}{f}_1(x) = f_1(-x)$, so that $\overset{\leftrightarrow}{f}_1$ represents the Fourier transform of g_1. So the inverse Fourier transform of $\sqrt{2\pi}g_1 \times g_2$ is $(\overset{\leftrightarrow}{f}_1 * \overset{\leftrightarrow}{f}_2)^\leftrightarrow$. But, just as in the proof of 284Kb, $(\overset{\leftrightarrow}{f}_1 * \overset{\leftrightarrow}{f}_2)^\leftrightarrow = f_1 * f_2$, so $f_1 * f_2$ is the inverse Fourier transform of $\sqrt{2\pi}g_1 \times g_2$, and $\sqrt{2\pi}g_1 \times g_2$ represents the Fourier transform of $f_1 * f_2$, as claimed. Also $f_1 * f_2$, being the Fourier transform of an integrable function, is continuous (283Cf; see also 255K).

284P Corollary Writing $L^2_{\mathbb{C}}$ for the Hilbert space of equivalence classes of square-integrable complex-valued functions on $\mathbb{R}$, we have a linear isometry $T : L^2_{\mathbb{C}} \to L^2_{\mathbb{C}}$ given by saying that $T(f^\bullet) = g^\bullet$ whenever f, $g \in \mathcal{L}^2_{\mathbb{C}}$ and g represents the Fourier transform of f.

284Q Remarks (a) 284P corresponds, of course, to 282K, where the similar isometry between $\ell^2_{\mathbb{C}}(\mathbb{Z})$ and $L^2_{\mathbb{C}}(]-\pi,\pi])$ is described. In that case there was a marked asymmetry which is absent from the present situation; because the relevant measure on $\mathbb{Z}$, counting measure, gives non-zero mass to every point, members of $\ell^2_{\mathbb{C}}$ are true functions, and it is not surprising that we have a straightforward formula for $S(f^\bullet) \in \ell^2_{\mathbb{C}}$ for every $f \in \mathcal{L}^2_{\mathbb{C}}(]-\pi,\pi])$. The difficulty of describing $S^{-1} : \ell^2_{\mathbb{C}}(\mathbb{Z}) \to L^2_{\mathbb{C}}(]-\pi,\pi])$ is very similar to the difficulty of describing $T : L^2_{\mathbb{C}}(\mathbb{R}) \to L^2_{\mathbb{C}}(\mathbb{R})$ and its inverse. 284Yg and 286U-286V show just how close this similarity is.

(b) I have spelt out parts (c) and (d) of 284O in detail, perhaps in unnecessary detail, because they give me an opportunity to insist on the difference between '$\sqrt{2\pi}g_1 \times g_2$ *represents* the Fourier transform of $f_1 * f_2$' and '$\frac{1}{\sqrt{2\pi}}g_1 * g_2$ *is* the Fourier transform of $f_1 \times f_2$'. The actual functions g_1 and g_2 are not well-defined by the hypothesis that they represent the Fourier transforms of f_1 and f_2, though their equivalence classes $g_1^\bullet$, $g_2^\bullet \in L^2_{\mathbb{C}}$ are. So the product $g_1 \times g_2$ is also not uniquely defined as a function, though its equivalence class $(g_1 \times g_2)^\bullet = g_1^\bullet \times g_2^\bullet$ is well-defined as a member of $L^1_{\mathbb{C}}$. However the continuous function $g_1 * g_2$ is unaffected by changes to g_1 and g_2 on negligible sets, so is well defined as a function; and since $f_1 \times f_2$ is integrable, and has a true Fourier transform, it is to be expected that $(f_1 \times f_2)^\wedge$ should be exactly equal to $\frac{1}{\sqrt{2\pi}} g_1 * g_2$.

(c) Of course 284Oc-284Od also exhibit a characteristic feature of arguments involving Fourier transforms, the extension by continuity of relations valid for test functions.

(d) 284Oa is a version of **Plancherel's theorem**. The formula $\|f\|_2 = \|\hat{f}\|_2$ is **Parseval's identity**.

284R Dirac's delta function Consider the tempered function $\chi\mathbb{R}$ with constant value 1. In what sense, if any, can we assign a Fourier transform to $\chi\mathbb{R}$?

If we examine $\int \chi\mathbb{R} \times \hat{h}$, as suggested in 284H, we get

$$\int_{-\infty}^{\infty} \chi\mathbb{R} \times \hat{h} = \int_{-\infty}^{\infty} \hat{h} = \sqrt{2\pi}\check{\hat{h}}(0) = \sqrt{2\pi}h(0)$$

for every rapidly decreasing test function h. Of course there is no *function* g such that $\int g \times h = \sqrt{2\pi}h(0)$ for every rapidly decreasing test function h, since (using the arguments of 284G) we should have to have $\int_a^b g = \sqrt{2\pi}$ whenever $a < 0 < b$, so that the indefinite integral of g could not be continuous at 0. However there is a *measure* on $\mathbb{R}$ with exactly the right property, the Dirac measure δ_0 concentrated at 0; this is a Radon probability measure (257Xa), and $\int h\,d\delta_0 = h(0)$ for every function h defined at 0. So we shall have

$$\int_{-\infty}^{\infty} \chi\mathbb{R} \times \hat{h} = \sqrt{2\pi}\int h\,d\delta_0$$

for every rapidly decreasing test function h, and we can reasonably say that the measure $\nu = \sqrt{2\pi}\delta_0$ 'represents the Fourier transform of $\chi\mathbb{R}$'.

We note with pleasure at this point that

$$\frac{1}{\sqrt{2\pi}}\int e^{ixy}\nu(dy) = 1$$

for every $x \in \mathbb{R}$, so that $\chi\mathbb{R}$ can be called the inverse Fourier transform of ν.

If we look at the formulae of Theorem 284M, we get ideas consistent with this pairing of $\chi\mathbb{R}$ with ν. We have

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-iyx}e^{-\epsilon x^2}\chi\mathbb{R}(x)dx = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-iyx}e^{-\epsilon x^2}dx = \frac{1}{\sqrt{2\epsilon}}e^{-y^2/4\epsilon}$$

for every $y \in \mathbb{R}$, using 283N with $\sigma = 1/\sqrt{2\epsilon}$. So

$$\lim_{\epsilon\downarrow 0} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-iyx}e^{-\epsilon x^2}\chi\mathbb{R}(x)dx = 0$$

for every $y \neq 0$, and the Fourier transform of $\chi\mathbb{R}$ should be zero everywhere except at 0. On the other hand, the functions $y \mapsto \frac{1}{\sqrt{2\epsilon}}e^{-y^2/4\epsilon}$ all have integral $\sqrt{2\pi}$, concentrated more and more closely about 0 as ϵ decreases to 0, so also point us directly to ν, the measure which gives mass $\sqrt{2\pi}$ to 0.

Thus allowing measures, as well as functions, enables us to extend the notion of Fourier transform. Of course we can go very much farther than this. If h is any rapidly decreasing test function, then

$$\int_{-\infty}^{\infty} x\hat{h}(x)dx = -i\sqrt{2\pi}h'(0),$$

so that the identity function $x \mapsto x$ can be assigned, as a Fourier transform, the operator $h \mapsto -i\sqrt{2\pi}h'(0)$.

At this point we are entering the true theory of (Schwartzian) distributions or 'generalized functions', and I had better stop. The 'Dirac delta function' is most naturally regarded as the measure δ_0 above; alternatively, as $\frac{1}{\sqrt{2\pi}}\widehat{\chi\mathbb{R}}$.

284W The multidimensional case As in §283, I give exercises designed to point the way to the r-dimensional generalization.

(a) A **rapidly decreasing test function** on $\mathbb{R}^r$ is a function $h : \mathbb{R}^r \to \mathbb{C}$ such that (i) h is **smooth**, that is, all repeated partial derivatives

$$\frac{\partial^m h}{\partial\xi_{j_1}\ldots\partial\xi_{j_m}}$$

are defined and continuous everywhere in $\mathbb{R}^r$ (ii)

$$\sup_{x\in\mathbb{R}^r}\|x\|^k|h(x)|<\infty,\quad \sup_{x\in\mathbb{R}^r}\|x\|^k|\frac{\partial^m h}{\partial\xi_{j_1}\ldots\partial\xi_{j_m}}(x)|<\infty$$

for every $k\in\mathbb{N}$, $j_1,\ldots,j_m\le r$. A **tempered function** on $\mathbb{R}^r$ is a measurable complex-valued function f, defined almost everywhere in $\mathbb{R}^r$, such that, for some $k\in\mathbb{N}$,

$$\int_{\mathbb{R}^r}\frac{1}{1+\|x\|^k}|f(x)|dx<\infty.$$

Show that if f is a tempered function on $\mathbb{R}^r$ and h is a rapidly decreasing test function on $\mathbb{R}^r$ then $f\times h$ is integrable.

(b) Show that if h is a rapidly decreasing test function on $\mathbb{R}^r$ so is $\hat{h}$, and that in this case $\hat{h}^\vee=h$.

(c) Show that if f is a tempered function on $\mathbb{R}^r$ and $\int f\times h=0$ for every rapidly decreasing test function h on $\mathbb{R}^r$, then $f=0$ a.e.

(d) If f and g are tempered functions on $\mathbb{R}^r$, I say that g **represents the Fourier transform of** f if $\int g\times h=\int f\times\hat{h}$ for every rapidly decreasing test function h on $\mathbb{R}^r$. Show that if f is integrable then $\hat{f}$ represents the Fourier transform of f in this sense.

(e) Let f be any tempered function on $\mathbb{R}^r$. Writing $\psi_\sigma(x)=\frac{1}{(\sigma\sqrt{2\pi})^r}e^{-x\,.\,x/2\sigma^2}$ for $x\in\mathbb{R}^r$, show that $\lim_{\sigma\downarrow0}(f*\psi_\sigma)(x)=c$ whenever $x\in\mathbb{R}^r$, $c\in\mathbb{C}$ are such that $\lim_{\delta\downarrow0}\frac{1}{\delta^r}\int_{B(x,\delta)}|f(t)-c|dt=0$, writing $B(x,\delta)=\{t:\|t-x\|\le\delta\}$.

(f) Let f and g be tempered functions on $\mathbb{R}^r$ such that g represents the Fourier transform of f, and h a rapidly decreasing test function. Show that (i) the Fourier transform of $f\times h$ is $\frac{1}{(\sqrt{2\pi})^r}g*\hat{h}$ (ii) $(\sqrt{2\pi})^r g\times\hat{h}$ represents the Fourier transform of $f*h$.

(g) Let f and g be tempered functions on $\mathbb{R}^r$ such that g represents the Fourier transform of f. Show that

$$g(y)=\lim_{\epsilon\downarrow0}\frac{1}{(\sqrt{2\pi})^r}\int_{\mathbb{R}^r}e^{-iy\,.\,x}e^{-\epsilon x\,.\,x}f(x)dx$$

for almost every $y\in\mathbb{R}^r$.

(h) Show that for any square-integrable complex-valued function f on $\mathbb{R}^r$ and any $\epsilon>0$ there is a rapidly decreasing test function h such that $\|f-h\|_2\le\epsilon$.

(i) Let $\mathcal{L}^2_{\mathbb{C}}$ be the space of square-integrable complex-valued functions on $\mathbb{R}^r$. Show that
(i) for every $f\in\mathcal{L}^2_{\mathbb{C}}$ there is a $g\in\mathcal{L}^2_{\mathbb{C}}$ which represents the Fourier transform of f, and in this case $\|g\|_2=\|f\|_2$;
(ii) if g_1, $g_2\in\mathcal{L}^2_{\mathbb{C}}$ represent the Fourier transforms of f_1, $f_2\in\mathcal{L}^2_{\mathbb{C}}$, then $\frac{1}{(\sqrt{2\pi})^r}g_1*g_2$ is the Fourier transform of $f_1\times f_2$, and $(\sqrt{2\pi})^r g_1\times g_2$ represents the Fourier transform of f_1*f_2.

(j) Let T be an invertible real $r\times r$ matrix, regarded as a linear operator from $\mathbb{R}^r$ to itself. (i) Show that $\hat{f}=|\det T|(f\circ T)^\wedge\circ T'$ for every integrable complex-valued function on $\mathbb{R}^r$. (ii) Show that $h\circ T$ is a rapidly decreasing test function for every rapidly decreasing test function h. (iii) Show that if f, g are a tempered functions and g represents the Fourier transform of f, then $\frac{1}{|\det T|}g\circ(T')^{-1}$ represents the Fourier transform of $f\circ T$; so that if T is orthogonal, then $g\circ T$ represents the Fourier transform of $f\circ T$.

284X Basic exercises (a) Show that if g and h are rapidly decreasing test functions, so is $g\times h$.

(b) Show that there are non-zero continuous integrable functions f, $g:\mathbb{R}\to\mathbb{C}$ such that $f*g=0$ everywhere. (*Hint*: take them to be Fourier transforms of suitable test functions.)

(c) Suppose that $f:\mathbb{R}\to\mathbb{C}$ is a differentiable function such that its derivative f' is a tempered function and, for some $k\in\mathbb{N}$,

$$\lim_{x\to\infty}x^{-k}f(x)=\lim_{x\to-\infty}x^{-k}f(x)=0.$$

(i) Show that $\int f\times h'=-\int f'\times h$ for every rapidly decreasing test function h. (ii) Show that if g is a tempered function representing the Fourier transform of f, then $y\mapsto iyg(y)$ represents the Fourier transform of f'.

(d) Show that if h is a rapidly decreasing test function and f is any measurable complex-valued function, defined almost everywhere in $\mathbb{R}$, such that $\int_{-\infty}^{\infty}|x|^k|f(x)|dx<\infty$ for every $k\in\mathbb{N}$, then the convolution $f*h$ is a rapidly decreasing test function. (*Hint*: show that the Fourier transform of $f*h$ is a test function.)

>(e) Let f be a tempered function such that $\lim_{a\to\infty}\int_{-a}^{a}f$ exists in $\mathbb{C}$. Show that this limit is also equal to $\lim_{\epsilon\downarrow 0}\int_{-\infty}^{\infty}e^{-\epsilon x^2}f(x)dx$. (*Hint*: set $g(x)=f(x)+f(-x)$. Use 224J to show that if $0\le a\le b$ then $|\int_a^b g(x)e^{-\epsilon x^2}dx|\le\sup_{c\in[a,b]}|\int_a^c g|$, so that $\lim_{a\to\infty}\int_0^a g(x)e^{-\epsilon x^2}dx$ exists uniformly in ϵ, while $\lim_{\epsilon\downarrow 0}\int_0^a g(x)e^{-\epsilon x^2}dx=\int_0^a g$ for every $a\ge 0$.)

>(f) Let f and g be tempered functions on $\mathbb{R}$ such that g represents the Fourier transform of f. Show that

$$g(y)=\lim_{a\to\infty}\frac{1}{\sqrt{2\pi}}\int_{-a}^{a}e^{-iyx}f(x)dx$$

at almost all points y for which the limit exists. (*Hint*: 284Xe, 284M.)

>(g) Let f be an integrable complex-valued function on $\mathbb{R}$ such that $\hat{f}$ also is integrable. Show that $\check{\hat{f}}=f$ at any point at which f is continuous.

(h) Show that for every $p\in[1,\infty[$, $f\in\mathcal{L}^p_{\mathbb{C}}$ and $\epsilon>0$ there is a rapidly decreasing test function h such that $\|f-h\|_p\le\epsilon$.

>(i) Let f and g be square-integrable complex-valued functions on $\mathbb{R}$ such that g represents the Fourier transform of f. Show that

$$\int_c^d f=\frac{i}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{e^{icy}-e^{idy}}{y}g(y)dy$$

whenever $c<d$ in $\mathbb{R}$.

(j) Let f be a measurable complex-valued function, defined almost everywhere in $\mathbb{R}$, such that $\int|f|^p<\infty$, where $1<p\le 2$. Show that f is a tempered function and that there is a tempered function g representing the Fourier transform of f. (*Hint*: express f as f_1+f_2, where f_1 is integrable and f_2 is square-integrable.) (**Remark** Defining $\|f\|_p$, $\|g\|_q$ as in 244D, where $q=p/(p-1)$, we have $\|g\|_q\le(2\pi)^{(p-2)/2p}\|f\|_p$; see ZYGMUND 59, XVI.3.2.)

(k) Let f, g be square-integrable complex-valued functions on $\mathbb{R}$ such that g represents the Fourier transform of f.

(i) Show that

$$\frac{1}{\sqrt{2\pi}}\int_{-a}^{a}e^{ixy}g(y)dy=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\sin at}{t}f(x-t)dt$$

for every $x\in\mathbb{R}$, $a>0$. (*Hint*: find the inverse Fourier transform of $y\mapsto e^{-ixy}\chi[-a,a](y)$, and use 284Ob.)

(ii) Show that if $f(x)=0$ for $x\in\,]c,d[$ then

$$\frac{1}{\sqrt{2\pi}}\lim_{a\to\infty}\int_{-a}^{a}e^{ixy}g(y)dy=0$$

for $x\in\,]c,d[$.

(iii) Show that if f is differentiable at $x\in\mathbb{R}$, then

$$\frac{1}{\sqrt{2\pi}}\lim_{a\to\infty}\int_{-a}^{a}e^{ixy}g(y)dy=f(x).$$

(iv) Show that if f has bounded variation over some interval properly containing x, then

$$\frac{1}{\sqrt{2\pi}}\lim_{a\to\infty}\int_{-a}^{a}e^{ixy}g(y)dy=\frac{1}{2}(\lim_{t\in\operatorname{dom}f,t\uparrow x}f(t)+\lim_{t\in\operatorname{dom}f,t\downarrow x}f(t)).$$

(l) Let f be an integrable complex function on $\mathbb{R}$. Show that if $\hat{f}$ is square-integrable, so is f.

(m) Let f_1, f_2 be square-integrable complex-valued functions on $\mathbb{R}$ with Fourier transforms represented by g_1, g_2. Show that $\int_{-\infty}^{\infty}f_1(t)f_2(-t)dt=\int_{-\infty}^{\infty}g_1(t)g_2(t)dt$.

(n) Suppose $x\in\mathbb{R}$. Write δ_x for Dirac measure on $\mathbb{R}$ concentrated at x. Describe a sense in which $\sqrt{2\pi}\delta_x$ can be regarded as the Fourier transform of the function $t\mapsto e^{ixt}$.

(o) For any tempered function f and $x \in \mathbb{R}$, let δ_x be the Dirac measure on $\mathbb{R}$ concentrated at x, and set

$$(\delta_x * f)(u) = \int f(u-t)\delta_x(dt) = f(u-x)$$

for every u for which $u - x \in \operatorname{dom} f$ (cf. 257Xe). If g represents the Fourier transform of f, find a corresponding representation of the Fourier transform of $\delta_x * f$, and relate it to the product of g with the Fourier transform of δ_x.

(p) Show that

$$\lim_{\delta\downarrow 0, a\to\infty}\Bigl(\int_{-a}^{-\delta}\frac{1}{x}e^{-iyx}dx + \int_{\delta}^{a}\frac{1}{x}e^{-iyx}dx\Bigr) = -\pi i \operatorname{sgn} y$$

for every $y \in \mathbb{R}$, writing $\operatorname{sgn} y = y/|y|$ if $y \neq 0$ and $\operatorname{sgn} 0 = 0$. (*Hint*: 283Da.)

(ii) Show that

$$\lim_{c\to\infty}\frac{1}{c}\int_0^c\int_{-a}^{a} e^{ixy}\operatorname{sgn} y\,dy\,da = \frac{2i}{x}$$

for every $x \neq 0$.

(iii) Show that for any rapidly decreasing test function h,

$$\begin{aligned}\int_0^\infty \frac{1}{x}(\hat{h}(x) - \hat{h}(-x))dx &= \lim_{\delta\downarrow 0, a\to\infty}\Bigl(\int_{-a}^{-\delta}\frac{1}{x}\hat{h}(x)dx + \int_{\delta}^{a}\frac{1}{x}\hat{h}(x)dx\Bigr)\\ &= -\frac{i\pi}{\sqrt{2\pi}}\int_{-\infty}^{\infty} h(y)\operatorname{sgn} y\,dy.\end{aligned}$$

(iv) Show that for any rapidly decreasing test function h,

$$\frac{i\pi}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\hat{h}(x)\operatorname{sgn} x\,dx = \int_0^\infty \frac{1}{y}(h(y) - h(-y))dy.$$

(q) Let $\langle h_n\rangle_{n\in\mathbb{N}}$ be a sequence of rapidly decreasing test functions such that $\phi(f) = \lim_{n\to\infty}\int_{-\infty}^{\infty} h_n \times f$ is defined for every rapidly decreasing test function f. Show that $\lim_{n\to\infty}\int_{-\infty}^{\infty} h_n' \times f$, $\lim_{n\to\infty}\int_{-\infty}^{\infty}\hat{h}_n \times f$ and $\lim_{n\to\infty}\int_{-\infty}^{\infty}(h_n * g)\times f$ are defined for all rapidly decreasing test functions f and g, and are zero if ϕ is identically zero. (*Hint*: 255G will help with the last.)

284Y Further exercises (a) Let f be an integrable complex-valued function on $]-\pi,\pi]$, and $\tilde{f}$ its periodic extension, as in 282Ae. Show that $\tilde{f}$ is a tempered function. Show that for any rapidly decreasing test function h, $\int \tilde{f}\times\hat{h} = \sqrt{2\pi}\sum_{k=-\infty}^{\infty} c_k h(k)$, where $\langle c_k\rangle_{k\in\mathbb{N}}$ is the sequence of Fourier coefficients of f. (*Hint*: begin with the case $f(x) = e^{inx}$. Next show that

$$M = \sum_{k=-\infty}^{\infty}|h(k)| + \sum_{k=-\infty}^{\infty}\sup_{x\in[(2k-1)\pi,(2k+1)\pi]}|\hat{h}(x)| < \infty,$$

and that

$$\Bigl|\int \tilde{f}\times\hat{h} - \sqrt{2\pi}\sum_{k=-\infty}^{\infty} c_k h(k)\Bigr| \le M\|f\|_1.$$

Finally apply 282Ib.)

(b) Let f be a complex-valued function, defined almost everywhere in $\mathbb{R}$, such that $f \times h$ is integrable for every rapidly decreasing test function h. Show that f is tempered.

(c) Let f and g be tempered functions on $\mathbb{R}$ such that g represents the Fourier transform of f. Show that

$$\int_c^d f = \frac{i}{\sqrt{2\pi}}\lim_{\sigma\to\infty}\int_{-\infty}^{\infty}\frac{e^{icy}-e^{idy}}{y}e^{-y^2/2\sigma^2}g(y)dy$$

whenever $c \le d$ in $\mathbb{R}$. (*Hint*: set $\theta = \chi[c,d]$. Show that both sides are $\lim_{\sigma\to\infty}\int f\times(\theta * \psi_{1/\sigma})$, defining ψ_σ as in 283N.)

(d) Show that if $g : \mathbb{R} \to \mathbb{R}$ is an odd function of bounded variation such that $\int_1^\infty \frac{1}{x}g(x)dx = \infty$, then g does not represent the Fourier transform of any tempered function. (*Hint*: 283Yd, 284Yc.)

(e) Let $\mathcal{S}$ be the space of rapidly decreasing test functions. For k, $m \in \mathbb{N}$ set $\tau_{km}(h) = \sup_{x\in\mathbb{R}}|x|^k|h^{(m)}(x)|$ for every $h \in \mathcal{S}$, writing $h^{(m)}$ for the mth derivative of h as usual. (i) Show that each τ_{km} is a seminorm and that $\mathcal{S}$ is complete and separable for the metrizable linear space topology $\mathfrak{T}$ they define. (ii) Show that $h \mapsto \hat{h} : \mathcal{S} \to \mathcal{S}$ is continuous for $\mathfrak{T}$. (iii) Show that if f is any tempered function, then $h \mapsto \int f \times h$ is $\mathfrak{T}$-continuous. (iv) Show that if f is an integrable function such that $\int |x^k f(x)|dx < \infty$ for every $k \in \mathbb{N}$, then $h \mapsto f * h : \mathcal{S} \to \mathcal{S}$ is $\mathfrak{T}$-continuous.

(f) Show that if f is a tempered function on $\mathbb{R}$ and

$$\gamma = \lim_{c\to\infty} \frac{1}{c}\int_0^c \int_{-a}^a f(x)dxda$$

is defined in $\mathbb{C}$, then γ is also

$$\lim_{\epsilon\downarrow 0} \int_{-\infty}^{\infty} f(x)e^{-\epsilon|x|}dx.$$

(g) Let f, g be square-integrable complex-valued functions on $\mathbb{R}$ such that g represents the Fourier transform of f. Suppose that $m \in \mathbb{Z}$ and that $(2m-1)\pi < x < (2m+1)\pi$. Set $\tilde{f}(t) = f(t+2m\pi)$ for those $t \in \left]-\pi,\pi\right]$ such that $t+2m\pi \in \operatorname{dom} f$. Let $\langle c_k\rangle_{k\in\mathbb{Z}}$ be the sequence of Fourier coefficients of $\tilde{f}$. Show that

$$\frac{1}{\sqrt{2\pi}} \lim_{a\to\infty} \int_{-a}^a e^{ixy}g(y)dy = \lim_{n\to\infty} \sum_{k=-n}^n c_k e^{ikx}$$

in the sense that if one limit exists in $\mathbb{C}$ so does the other, and they are then equal. (*Hint*: 284Xk(i), 282Da.)

(h) Show that if f is integrable over $\mathbb{R}$ and there is some $M \ge 0$ such that $f(x) = \hat{f}(x) = 0$ for $|x| \ge M$, then $f = 0$ a.e. (*Hint*: reduce to the case $M = \pi$. Looking at the Fourier series of $f\restriction\left]-\pi,\pi\right]$, show that f is expressible in the form $f(x) = \sum_{k=-m}^m c_k e^{ikx}$ for almost every $x \in \left]-\pi,\pi\right]$. Now compute $\hat{f}(2n+\frac{1}{2})$ for large n.)

(i) Let ν be a Radon measure on $\mathbb{R}$ (definition: 256A) which is 'tempered' in the sense that $\int_{-\infty}^{\infty} \frac{1}{1+|x|^k}\nu(dx)$ is finite for some $k \in \mathbb{N}$. (i) Show that every rapidly decreasing test function is ν-integrable. (ii) Show that if ν has bounded support (definition: 256Xf), and h is a rapidly decreasing test function, then $\nu * h$ is a rapidly decreasing test function, where $(\nu * h)(x) = \int_{-\infty}^{\infty} h(x-y)\nu(dy)$ for $x \in \mathbb{R}$. (iii) Show that there is a sequence $\langle h_n\rangle_{n\in\mathbb{N}}$ of rapidly decreasing test functions such that $\lim_{n\to\infty} \int_{-\infty}^{\infty} h_n \times f = \int_{-\infty}^{\infty} f d\nu$ for every rapidly decreasing test function f.

(j) Let $\phi : \mathcal{S} \to \mathbb{R}$ be a functional defined by the formula of 284Xq. Show that ϕ is continuous for the topology of 284Ye. (*Note*: it helps to know a little more about metrizable linear topological spaces than is covered in §2A5.)

284 Notes and comments Yet again I must warn you that the material above gives a very restricted view of the subject. I have tried to indicate how the theory of Fourier transforms of 'good' functions – here taken to be the rapidly decreasing test functions – may be extended, through a kind of duality, to a very much wider class of functions, the 'tempered functions'. Evidently, writing $\mathcal{S}$ for the linear space of rapidly decreasing test functions, we can seek to investigate a Fourier transform of any linear functional $\phi : \mathcal{S} \to \mathbb{C}$, writing $\hat{\phi}(h) = \phi(\hat{h})$ for any $h \in \mathcal{S}$. (It is actually commoner at this point to restrict attention to functionals ϕ which are continuous for the standard topology on $\mathcal{S}$, described in 284Ye; these are called **tempered distributions**.) By 284F-284G, we can identify some of these functionals with equivalence classes of tempered functions, and then set out to investigate those tempered functions whose Fourier transforms can again be represented by tempered functions.

I suppose the structure of the theory of Fourier transforms is best laid out through the formulae involved. Our aim is to set up pairs $(f,g) = (f,\hat{f}) = (\check{g},g)$ in such a way that we have

Inversion: $\check{\hat{h}} = \hat{\check{h}} = h$;

Reversal: $\check{h}(y) = \hat{h}(-y)$;

Linearity: $(h_1+h_2)^\wedge = \hat{h}_1 + \hat{h}_2$, $\quad (ch)^\wedge = c\hat{h}$;

Differentiation: $(h')^\wedge(y) = iy\hat{h}(y)$;

Shift: if $h_1(x) = h(x+c)$ then $\hat{h}_1(y) = e^{iyc}\hat{h}(y)$;

Modulation: if $h_1(x) = e^{icx}h(x)$ then $\hat{h}_1(y) = \hat{h}(y-c)$;

Symmetry: if $h_1(x) = h(-x)$ then $\hat{h}_1(y) = \hat{h}(-y)$;

Complex Conjugate: $(\overline{h})^\wedge(y) = \overline{\hat{h}(-y)}$;

Dilation: if $h_1(x) = h(cx)$, where $c > 0$, then $\hat{h}_1(y) = \frac{1}{c}\hat{h}(\frac{y}{c})$;

Convolution: $(h_1 * h_2)^\wedge = \sqrt{2\pi}\hat{h}_1 \times \hat{h}_2$, $\quad (h_1 \times h_2)^\wedge = \frac{1}{\sqrt{2\pi}}\hat{h}_1 * \hat{h}_2$;

Duality: $\int_{-\infty}^{\infty} h_1 \times \hat{h}_2 = \int_{-\infty}^{\infty} \hat{h}_1 \times h_2$;

Parseval: $\int_{-\infty}^{\infty} h_1 \times \overline{h_2} = \int_{-\infty}^{\infty} \hat{h}_1 \times \overline{\hat{h}_2}$;

and, of course,

$$\hat{h}(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-iyx}h(x)dx,$$

$$\int_c^d \hat{h}(y)dy = \frac{i}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \frac{e^{-icy}-e^{-idy}}{y}h(y)dy.$$

(I have used the letter h in the list above to suggest what is in fact the case, that all the formulae here are valid for rapidly decreasing test functions.) On top of all this, it is often important that the operation $h \mapsto \hat{h}$ should be continuous in some sense.

The challenge of the 'pure' theory of Fourier transforms is to find the widest possible variety of objects h for which the formulae above will be valid, subject to appropriate interpretations of $^\wedge$, $*$ and $\int_{-\infty}^{\infty}$. I must of course remark here that from the very beginnings, the subject has been enriched by its applications in other parts of mathematics, the physical sciences and the social sciences, and that again and again these have suggested further possible pairs $(f, \hat{f})$, making new demands on our power to interpret the rules we seek to follow. Even the theory of distributions does not seem to give a full canonical account of what can be done. First, there are great difficulties in interpreting the 'product' of two arbitrary distributions, making several of the formulae above problematic; and second, it is not obvious that only one kind of distribution need be considered. In this section I have looked at just one space of 'test functions', the space $\mathcal{S}$ of rapidly decreasing test functions; but at least two others are significant, the space $\mathcal{D}$ of smooth functions with bounded support and the space $\mathcal{Z}$ of Fourier transforms of functions in $\mathcal{D}$. The advantage of starting with $\mathcal{S}$ is that it gives a symmetric theory, since $\hat{h} \in \mathcal{S}$ for every $h \in \mathcal{S}$; but it is easy to find objects (e.g., the function $x \mapsto e^{x^2}$, or the function $x \mapsto 1/|x|$) which cannot be interpreted as functionals on $\mathcal{S}$, so that their Fourier transforms must be investigated by other methods, if at all. In 284Xp I sketch some of the arguments which can be used to justify the assertion that the Fourier transform of the function $x \mapsto 1/x$ is, or can be represented by, the function $y \mapsto -i\sqrt{\frac{\pi}{2}}\operatorname{sgn} y$; the general principle in this case being that we approach both 0 and ∞ symmetrically. For a variety of such matching pairs, established by arguments based on the idea in 284Xq, see LIGHTHILL 59, chap. 3.

Accordingly it seems that, after nearly two centuries, we must still proceed by carefully examining particular classes of function, and checking appropriate interpretations of the formulae. In the work above I have repeatedly used the concepts

$$\lim_{a\to\infty}\int_{-a}^{a} f, \quad \lim_{\epsilon\downarrow 0}\int_{-\infty}^{\infty} e^{-\epsilon x^2} f(x)dx$$

as alternative interpretations of $\int_{-\infty}^{\infty} f$. (Of course they are closely related; see 284Xe.) The reasons for using the particular kernel $e^{-\epsilon x^2}$ are that it belongs to $\mathcal{S}$, it is an even function, its Fourier transform is calculable and easy to manipulate, and it is associated with the normal probability density function $\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}$, so that any miscellaneous facts we gather have a chance of being valuable elsewhere. But there are applications in which alternative kernels are more manageable – e.g., $e^{-\epsilon|x|}$ (283Xr, 283Yb, 284Yf).

One of the guiding principles here is that purely formal manipulations, along the lines of those in the list above, and (especially) changes in the order of integration, with other exchanges of limit, again and again give rise to formulae which, suitably interpreted, are valid. First courses in analysis are often inhibitory; students are taught to distrust any manipulation which they cannot justify. To my own eye, the delight of this subject lies chiefly in the variety of the arguments demanded by a rigorous approach, the ground constantly shifting with the context; but there is no doubt that cheerful sanguinity is often the best guide to the manipulations which it will be right to try to justify.

This being a book on measure theory, I am of course particularly interested in the possibility of a measure appearing as a Fourier transform. This is what happens if we seek the Fourier transform of the constant function $\chi\mathbb{R}$ (284R). More generally, any periodic tempered function f with period 2π can be assigned a Fourier transform which is a 'signed measure' (for our present purposes, a complex linear combination of measures) concentrated on $\mathbb{Z}$, the mass at each $k \in \mathbb{Z}$ being determined by the corresponding Fourier coefficient of $f\restriction]-\pi,\pi]$ (284Xn, 284Ya). In the next section I will go farther in this direction, with particular reference to probability distributions on $\mathbb{R}^r$. But the reason why *positive* measures have not forced themselves on our attention so far is that we do not expect to get a positive function as a Fourier transform unless some very special conditions are satisfied, as in 283Yb.

As in §282, I have used the Hilbert space structure of $L^2_{\mathbb{C}}$ as the basis of the discussion of Fourier transforms of functions in $\mathcal{L}^2_{\mathbb{C}}$ (284O-284P). But as with Fourier series, Carleson's theorem (286U) provides a more direct description.

285 Characteristic functions

I come now to one of the most effective applications of Fourier transforms, the use of 'characteristic functions' to analyse probability distributions. It turns out not only that the Fourier transform of a probability distribution determines the distribution (285M) but that many of the things we want to know about a distribution are easily calculated from its transform (285G, 285Xf). Even more strikingly, pointwise convergence of Fourier transforms corresponds (for sequences) to convergence for the vague topology in the space of distributions, so they provide a new and extremely powerful method for proving such results as the Central Limit Theorem and Poisson's theorem (285Q).

As the applications of the ideas here mostly belong to probability theory, I return to probabilists' terminology, as in Chapter 27. There will nevertheless be many points at which it is appropriate to speak of integrals, and there will often be more than one measure in play; so I should say directly that an integral $\int f(x)dx$ will always be with respect to Lebesgue measure (usually, but not always, one-dimensional), as in the rest of the chapter, while integrals with respect to other measures will be expressed in the forms $\int f d\nu$ or $\int f(x)\nu(dx)$.

285A Definition (a) Let ν be a Radon probability measure on $\mathbb{R}^r$ (256A). Then the **characteristic function** of ν is the function $\phi_\nu : \mathbb{R}^r \to \mathbb{C}$ given by the formula

$$\phi_\nu(y) = \int e^{iy \cdot x}\nu(dx)$$

for every $y \in \mathbb{R}^r$, writing $y \centerdot x = \eta_1\xi_1 + \ldots + \eta_r\xi_r$ if $y = (\eta_1, \ldots , \eta_r)$, $x = (\xi_1, \ldots , \xi_r)$.

(b) Let $X_1, \ldots , X_r$ be real-valued random variables on the same probability space. The **characteristic function** of $\boldsymbol{X} = (X_1, \ldots , X_r)$ is the characteristic function $\phi_{\boldsymbol{X}} = \phi_{\nu_{\boldsymbol{X}}}$ of their joint probability distribution $\nu_{\boldsymbol{X}}$ as defined in 271C.

285B Remarks (a) By one of the ordinary accidents of history, the definitions of 'characteristic function' and 'Fourier transform' have evolved independently. In 283Ba I remarked that the definition of the Fourier transform remains unfixed, and that the formulae

$$\hat{f}(y) = \int_{-\infty}^{\infty} e^{iyx}f(x)dx,$$

$$\check{f}(y) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iyx}f(x)dx$$

are sometimes used. On the other hand, I think that nearly all authors agree on the definition of the characteristic function as given above. You may feel therefore that I should have followed their lead, and chosen the definition of Fourier transform which best matches the definition of characteristic function. I did not do so largely because I wished to emphasise the symmetry between the Fourier transform and the inverse Fourier transform, and the correspondence between Fourier transforms and Fourier series. The principal advantage of matching the definitions up would be to make the constants in such theorems as 283F, 285Xh the same, and would be balanced by the need to remember different constants for $\hat{f}$, $\check{f}$ in such results as 283M.

(b) A secondary reason for not trying too hard to make the formulae of this section match directly those of §§283-284 is that the r-dimensional case is at the heart of some of the most important applications of characteristic functions, so that it seems right to introduce it from the beginning; and consequently the formulae of this section will necessarily have new features compared with those in the body of the work so far.

285C Of course there is a direct way to describe the characteristic function of a family $(X_1, \ldots , X_r)$ of random variables, as follows.

Proposition Let $X_1, \ldots , X_r$ be real-valued random variables on the same probability space (Ω, Σ, μ), and $\nu_{\boldsymbol{X}}$ their joint distribution. Then their characteristic function $\phi_{\nu_{\boldsymbol{X}}}$ is given by

$$\phi_{\nu_{\boldsymbol{X}}}(y) = \mathbb{E}(e^{iy \cdot \boldsymbol{X}}) = \mathbb{E}(e^{i\eta_1 X_1} e^{i\eta_2 X_2} \ldots e^{i\eta_r X_r})$$

for every $y = (\eta_1, \ldots , \eta_r) \in \mathbb{R}^r$.

proof Apply 271E to the functions h_1, $h_2 : \mathbb{R}^r \to \mathbb{R}$ defined by

$$h_1(x) = \cos(y \,\boldsymbol{.}\, x), \quad h_2(y) = \sin(y \,\boldsymbol{.}\, x),$$

to see that

$$\begin{aligned}\phi_{\nu_{\boldsymbol{X}}}(y) &= \int h_1(x)\nu_{\boldsymbol{X}}(dx) + i\int h_2(x)\nu_{\boldsymbol{X}}(dx)\\ &= \mathbb{E}(h_1(\boldsymbol{X})) + i\mathbb{E}(h_2(\boldsymbol{X})) = \mathbb{E}(e^{iy\,\boldsymbol{.}\,\boldsymbol{X}}).\end{aligned}$$

285D I ought to spell out the correspondence between Fourier transforms, as defined in 283A, and characteristic functions.

Proposition Let ν be a Radon probability measure on $\mathbb{R}$. Write

$$\hat{\nu}(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-iyx}\nu(dx)$$

for every $y \in \mathbb{R}$, and ϕ_ν for the characteristic function of ν.

(a) $\hat{\nu}(y) = \frac{1}{\sqrt{2\pi}}\phi_\nu(-y)$ for every $y \in \mathbb{R}$.

(b) For any Lebesgue integrable complex-valued function h defined almost everywhere in $\mathbb{R}$,

$$\int_{-\infty}^{\infty} \hat{\nu}(y)h(y)dy = \int_{-\infty}^{\infty} \hat{h}(x)\nu(dx).$$

(c) For any rapidly decreasing test function h on $\mathbb{R}$ (see §284),

$$\int_{-\infty}^{\infty} h(x)\nu(dx) = \int_{-\infty}^{\infty} \check{h}(y)\hat{\nu}(y)dy.$$

(d) If ν is an indefinite-integral measure over Lebesgue measure, with Radon-Nikodým derivative f, then $\hat{\nu}$ is the Fourier transform of f.

proof (a) This is immediate from the definitions of ϕ_ν and $\hat{\nu}$.

(b) Because

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |h(y)|\nu(dx)dy = \int_{-\infty}^{\infty} |h(y)|dy < \infty,$$

we may change the order of integration to see that

$$\begin{aligned}\int_{-\infty}^{\infty} \hat{\nu}(y)h(y)dy &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-iyx}h(y)\nu(dx)dy\\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-iyx}h(y)dy\,\nu(dx) = \int_{-\infty}^{\infty} \hat{h}(x)\nu(dx).\end{aligned}$$

(c) This follows immediately from (b), because $\check{h}$ is integrable and $\hat{\check{h}} = h$ (284C).

(d) The point is just that

$$\int h\,d\nu = \int h(x)f(x)dx$$

for every bounded Borel measurable $h : \mathbb{R} \to \mathbb{R}$ (235K), and therefore for the functions $x \mapsto e^{-iyx} : \mathbb{R} \to \mathbb{C}$. Now

$$\hat{\nu}(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-iyx}\nu(dx) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-iyx}f(x)dx = \hat{f}(y)$$

for every y.

285E Lemma Let X be a normal random variable with expectation a and variance σ^2, where $\sigma > 0$. Then the characteristic function of X is given by

$$\phi(y) = e^{iya}e^{-\sigma^2y^2/2}.$$

proof This is just 283N with the constants changed. We have

$$\phi(y) = \mathbb{E}(e^{iyX}) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{iyx}e^{-(x-a)^2/2\sigma^2}dx$$

(taking the density function for X given in 274Ad, and applying 271Ic)

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{iy(\sigma t+a)} e^{-t^2/2} dt$$

(substituting $x = \sigma t + a$)

$$= e^{iya}\sqrt{2\pi}\hat{\psi}_1(-y\sigma)$$

(setting $\psi_1(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, as in 283N)

$$= e^{iya} e^{-\sigma^2 y^2/2}.$$

285F I now give results corresponding to parts of 283C, with an extra refinement concerning independent random variables (285I).

Proposition Let ν be a Radon probability measure on $\mathbb{R}^r$, and ϕ its characteristic function.

(a) $\phi(0) = 1$.

(b) $\phi : \mathbb{R}^r \to \mathbb{C}$ is uniformly continuous.

(c) $\phi(-y) = \overline{\phi(y)}$, $|\phi(y)| \le 1$ for every $y \in \mathbb{R}^r$.

(d) If $r = 1$ and $\int |x|\nu(dx) < \infty$, then $\phi'(y)$ exists and is equal to $i\int xe^{ixy}\nu(dx)$ for every $y \in \mathbb{R}$.

(e) If $r = 1$ and $\int x^2\nu(dx) < \infty$, then $\phi''(y)$ exists and is equal to $-\int x^2 e^{ixy}\nu(dx)$ for every $y \in \mathbb{R}$.

proof (a) $\phi(0) = \int \chi\mathbb{R}^r \nu(dx) = \nu(\mathbb{R}^r) = 1$.

(b) Let $\epsilon > 0$. Let $M > 0$ be such that

$$\nu\{x : \|x\| \ge M\} \le \epsilon,$$

writing $\|x\| = \sqrt{x\,.\,x}$ as usual. Let $\delta > 0$ be such that $|e^{ia} - 1| \le \epsilon$ whenever $|a| \le \delta$. Now suppose that y, $y' \in \mathbb{R}^r$ are such that $\|y - y'\| \le \delta/M$. Then whenever $\|x\| \le M$,

$$|e^{iy\,.\,x} - e^{iy'\,.\,x}| = |e^{iy'\,.\,x}||e^{i(y-y')\,.\,x} - 1| = |e^{i(y-y')\,.\,x} - 1| \le \epsilon$$

because

$$|(y - y')\,.\,x| \le \|y - y'\|\|x\| \le \delta.$$

Consequently

$$\begin{aligned} |\phi(y) - \phi(y')| &\le \int_{\|x\|\le M} |e^{iy\,.\,x} - e^{iy'\,.\,x}|\nu(dx) + \int_{\|x\|>M} |e^{iy\,.\,x}|\nu(dx) \\ &\qquad + \int_{\|x\|>M} |e^{iy'\,.\,x}|\nu(dx) \\ &\le \epsilon + \epsilon + \epsilon = 3\epsilon. \end{aligned}$$

As ϵ is arbitrary, ϕ is uniformly continuous.

(c) This is elementary;

$$\phi(-y) = \int e^{-iy\,.\,x}\nu(dx) = \overline{\int e^{iy\,.\,x}\nu(dx)} = \overline{\phi(y)},$$

$$|\phi(y)| = |\int e^{iy\,.\,x}\nu(dx)| \le \int |e^{iy\,.\,x}|\nu(dx) = \int \chi\mathbb{R}^r\,\nu(dx) = 1.$$

(d) The point is that $|\frac{\partial}{\partial y} e^{iyx}| = |x|$ for every x, $y \in \mathbb{R}$. So by 123D (applied, strictly speaking, to the real and imaginary parts of the function)

$$\phi'(y) = \frac{d}{dy}\int e^{iyx}\nu(dx) = \int \frac{\partial}{\partial y} e^{iyx}\nu(dx) = \int ixe^{iyx}\nu(dx).$$

(e) Since we now have $|\frac{\partial}{\partial y} xe^{iyx}| = x^2$ for every x, y, we can repeat the argument to get

$$\phi''(y) = i\frac{d}{dy}\int xe^{iyx}\nu(dx) = i\int \frac{\partial}{\partial y} xe^{iyx}\nu(dx) = -\int x^2 e^{iyx}\nu(dx).$$

285G Corollary (a) Let X be a real-valued random variable with finite expectation, and ϕ its characteristic function. Then $\phi'(0)=i\mathbb{E}(X)$.

(b) Let X be a real-valued random variable with finite variance, and ϕ its characteristic function. Then $\phi''(0)=-\mathbb{E}(X^2)$.

proof We have only to match X to its distribution ν, and say that

$$\text{`}X\text{ has finite expectation'}$$

corresponds to

$$\text{`}\int|x|\nu(dx)=\mathbb{E}(|X|)<\infty\text{'},$$

so that

$$\phi'(0)=i\int x\,\nu(dx)=i\mathbb{E}(X),$$

and that

$$\text{`}X\text{ has finite variance'}$$

corresponds to

$$\text{`}\int x^2\nu(dx)=\mathbb{E}(X^2)<\infty\text{'},$$

so that

$$\phi''(0)=-\int x^2\,\nu(dx)=-\mathbb{E}(X^2),$$

as in 271E.

285H Remark Observe that there is no result corresponding to 283Cg ('$\lim_{|y|\to\infty}\hat{f}(y)=0$'). If ν is the Dirac measure on $\mathbb{R}$ concentrated at 0, that is, the distribution of a random variable which is zero almost everywhere, then $\phi(y)=1$ for every y.

285I Proposition Let $X_1,\dots,X_n$ be independent real-valued random variables, with characteristic functions $\phi_1,\dots,\phi_n$. Let ϕ be the characteristic function of their sum $X=X_1+\ldots+X_n$. Then

$$\phi(y)=\prod_{j=1}^n\phi_j(y)$$

for every $y\in\mathbb{R}$.

proof Let $y\in\mathbb{R}$. By 272E, the variables

$$Y_j=e^{iyX_j}$$

are independent, so by 272R

$$\phi(y)=\mathbb{E}(e^{iyX})=\mathbb{E}(e^{iy(X_1+\ldots+X_n)})=\mathbb{E}(\prod_{j=1}^n Y_j)=\prod_{j=1}^n\mathbb{E}(Y_j)=\prod_{j=1}^n\phi_j(y),$$

as required.

Remark See also 285R below.

285J There is an inversion theorem for characteristic functions, corresponding to 283F; I give it in 285Xh, with an r-dimensional version in 285Yb. However, this does not seem to be as useful as the following group of results.

Lemma Let ν be a Radon probability measure on $\mathbb{R}^r$, and ϕ its characteristic function. Then for any $j\le r$ and $a>0$,

$$\nu\{x:|\xi_j|\ge a\}\le 7a\int_0^{1/a}(1-\mathcal{R}\mathrm{e}\,\phi(te_j))dt,$$

where $e_j\in\mathbb{R}^r$ is the jth unit vector.

proof We have

$$7a\int_0^{1/a}(1-\mathcal{R}\mathrm{e}\,\phi(te_j))dt = 7a\int_0^{1/a}\bigl(1-\mathcal{R}\mathrm{e}\int_{\mathbb{R}^r}e^{it\xi_j}\nu(dx)\bigr)dt$$
$$= 7a\int_0^{1/a}\int_{\mathbb{R}^r}1-\cos(t\xi_j)\nu(dx)dt$$
$$= 7a\int_{\mathbb{R}^r}\int_0^{1/a}1-\cos(t\xi_j)dt\,\nu(dx)$$

(because $(x,t)\mapsto 1-\cos(t\xi_j)$ is bounded and $\nu\mathbb{R}^r\cdot\frac{1}{a}$ is finite)

$$= 7a\int_{\mathbb{R}^r}(\frac{1}{a}-\frac{1}{\xi_j}\sin\frac{\xi_j}{a})\nu(dx)$$
$$\ge 7a\int_{|\xi_j|\ge a}(\frac{1}{a}-\frac{1}{\xi_j}\sin\frac{\xi_j}{a})\nu(dx)$$

(because $\frac{1}{\xi}\sin\frac{\xi}{a}\le\frac{1}{a}$ for every $\xi\ne 0$)

$$\ge \nu\{x:|\xi_j|\ge a\},$$

because

$$\frac{\sin\eta}{\eta}\le\frac{\sin 1}{1}\le\frac{6}{7}\text{ if }\eta\ge 1,$$

so

$$a(\frac{1}{a}-\frac{1}{\xi_j}\sin\frac{\xi_j}{a})\ge\frac{1}{7}$$

if $|\xi_j|\ge a$.

285K Characteristic functions and the vague topology The time has come to return to ideas mentioned briefly in 274L. Fix $r\ge 1$ and let P be the set of all Radon probability measures on $\mathbb{R}^r$. For any bounded continuous function $h:\mathbb{R}^r\to\mathbb{R}$, define $\rho_h:P\times P\to\mathbb{R}$ by setting

$$\rho_h(\nu,\nu')=|\int h\,d\nu-\int h\,d\nu'|$$

for ν, $\nu'\in P$. Then the vague topology on P is the topology generated by the pseudometrics ρ_h (274Ld).

285L Theorem Let ν, $\langle\nu_n\rangle_{n\in\mathbb{N}}$ be Radon probability measures on $\mathbb{R}^r$, with characteristic functions ϕ, $\langle\phi_n\rangle_{n\in\mathbb{N}}$. Then the following are equiveridical:

(i) $\nu=\lim_{n\to\infty}\nu_n$ for the vague topology;

(ii) $\int h\,d\nu=\lim_{n\to\infty}\int h\,d\nu_n$ for every bounded continuous $h:\mathbb{R}^r\to\mathbb{R}$;

(iii) $\lim_{n\to\infty}\phi_n(y)=\phi(y)$ for every $y\in\mathbb{R}^r$.

proof (a) The equivalence of (i) and (ii) is virtually the definition of the vague topology; we have

$$\lim_{n\to\infty}\nu_n=\nu\text{ for the vague topology}$$
$$\iff \lim_{n\to\infty}\rho_h(\nu_n,\nu)=0\text{ for every bounded continuous }h$$

(2A3Mc)

$$\iff \lim_{n\to\infty}|\int h\,d\nu_n-\int h\,d\nu|=0\text{ for every bounded continuous }h.$$

(b) Next, (ii) obviously implies (iii), because

$$\mathcal{R}\mathrm{e}\,\phi(y)=\int h_y\,d\nu=\lim_{n\to\infty}h_y\,d\nu_n=\lim_{n\to\infty}\mathcal{R}\mathrm{e}\,\phi_n(y),$$

setting $h_y(x)=\cos x\centerdot y$ for each x, and similarly

$$\mathcal{I}\mathrm{m}\,\phi(y)=\lim_{n\to\infty}\mathcal{I}\mathrm{m}\,\phi_n(y)$$

for every $y \in \mathbb{R}^r$.

(c) So we are left to prove that (iii)⇒(ii). I start by showing that, given $\epsilon > 0$, there is a closed bounded set K such that

$$\nu_n(\mathbb{R}^r \setminus K) \le \epsilon \text{ for every } n \in \mathbb{N}.$$

P We know that $\phi(0) = 1$ and that ϕ is continuous at 0 (285Fb). Let $a > 0$ be so large that for every $j \le r$, $|t| \le 1/a$ we have

$$1 - \mathcal{R}\mathrm{e}\,\phi(te_j) \le \frac{\epsilon}{14r},$$

writing e_j for the jth unit vector, as in 285J. Then

$$7a\int_0^{1/a}(1 - \mathcal{R}\mathrm{e}\,\phi(te_j))dt \le \frac{\epsilon}{2r}$$

for each $j \le r$. By Lebesgue's Dominated Convergence Theorem (since of course the functions $t \mapsto 1 - \mathcal{R}\mathrm{e}\,\phi_n(te_j)$ are uniformly bounded on $[0, \frac{1}{a}]$), there is an $n_0 \in \mathbb{N}$ such that

$$7a\int_0^{1/a}(1 - \mathcal{R}\mathrm{e}\,\phi_n(te_j))dt \le \frac{\epsilon}{r}$$

for every $j \le r$, $n \ge n_0$. But 285J tells us that now

$$\nu_n\{x : |\xi_j| \ge a\} \le \frac{\epsilon}{r}$$

for every $j \le r$, $n \ge n_0$. On the other hand, there is surely a $b \ge a$ such that

$$\nu_n\{x : |\xi_j| \ge b\} \le \frac{\epsilon}{r}$$

for every $j \le r$, $n < n_0$. So, setting $K = \{x : |\xi_j| \le b \text{ for every } j \le r\}$,

$$\nu_n(\mathbb{R}^r \setminus K) \le \epsilon$$

for every $n \in \mathbb{N}$, as required. **Q**

(d) Now take any bounded continuous $h : \mathbb{R}^r \to \mathbb{R}$ and $\epsilon > 0$. Set $M = 1 + \sup_{x\in\mathbb{R}^r} |h(x)|$, and let K be a bounded closed set such that

$$\nu_n(\mathbb{R}^r \setminus K) \le \frac{\epsilon}{M} \text{ for every } n \in \mathbb{N}, \quad \nu(\mathbb{R}^r \setminus K) \le \frac{\epsilon}{M},$$

using (b) just above. By the Stone-Weierstrass theorem (281K) there are $y_0, \dots , y_m \in \mathbb{Q}^r$ and $c_0, \dots , c_m \in \mathbb{C}$ such that

$$|h(x) - g(x)| \le \epsilon \text{ for every } x \in K,$$

$$|g(x)| \le M \text{ for every } x \in \mathbb{R}^r,$$

writing $g(x) = \sum_{k=0}^m c_k e^{iy_k \cdot x}$ for $x \in \mathbb{R}^r$. Now

$$\lim_{n\to\infty} \int g\,d\nu_n = \lim_{n\to\infty}\sum_{k=0}^m c_k\phi_n(y_k) = \sum_{k=0}^m c_k\phi(y_k) = \int g\,d\nu.$$

On the other hand, for every $n \in \mathbb{N}$,

$$|\int g\,d\nu_n - \int h\,d\nu_n| \le \int_K |g - h|d\nu_n + 2M\nu_n(\mathbb{R} \setminus K) \le 3\epsilon,$$

and similarly $|\int g\,d\nu - \int h\,d\nu| \le 3\epsilon$. Consequently

$$\limsup_{n\to\infty} |\int h\,d\nu_n - \int h\,d\nu| \le 6\epsilon.$$

As ϵ is arbitrary,

$$\lim_{n\to\infty} \int h\,d\nu_n = \int h\,d\nu,$$

and (ii) is true.

285M Corollary (a) Let ν, ν' be two Radon probability measures on $\mathbb{R}^r$ with the same characteristic functions. Then they are equal.

(b) Let $(X_1,\dots,X_r)$ and $(Y_1,\dots,Y_r)$ be two families of real-valued random variables. If

$$\mathbb{E}(e^{i\eta_1 X_1+\ldots+i\eta_r X_r})=\mathbb{E}(e^{i\eta_1 Y_1+\ldots+i\eta_r Y_r})$$

for all $\eta_1,\dots,\eta_r\in\mathbb{R}$, then $(X_1,\dots,X_r)$ has the same joint distribution as $(Y_1,\dots,Y_r)$.

proof (a) Applying 285L with $\nu_n=\nu'$ for every n, we see that $\int h\,d\nu'=\int h\,d\nu$ for every bounded continuous $h:\mathbb{R}^r\to\mathbb{R}$. By 256D(iv), $\nu=\nu'$.

(b) Apply (a) with ν, ν' the two joint distributions.

285N Remarks Probably the most important application of this theorem is to the standard proof of the Central Limit Theorem. I sketch the ideas in 285Xn and 285Yj-285Ym; details may be found in most serious probability texts; two on my shelf are Shiryayev 84, §III.4, and Feller 66, §XV.6. However, to get the full strength of Lindeberg's version of the Central Limit Theorem we have to work quite hard, and I therefore propose to illustrate the method with a version of Poisson's theorem (285Q) instead. I begin with two lemmas which are very frequently used in results of this kind.

285O Lemma Let $c_0,\dots,c_n,d_0,\dots,d_n$ be complex numbers of modulus at most 1. Then

$$|\prod_{k=0}^n c_k-\prod_{k=0}^n d_k|\le\sum_{k=0}^n|c_k-d_k|.$$

proof Induce on n. The case $n=0$ is trivial. For the case $n=1$ we have

$$\begin{aligned}|c_0c_1-d_0d_1|&=|c_0(c_1-d_1)+(c_0-d_0)d_1|\\&\le|c_0||c_1-d_1|+|c_0-d_0||d_1|\le|c_1-d_1|+|c_0-d_0|,\end{aligned}$$

which is what we need. For the inductive step to $n+1$, we have

$$|\prod_{k=0}^{n+1}c_k-\prod_{k=0}^{n+1}d_k|\le|\prod_{k=0}^{n}c_k-\prod_{k=0}^{n}d_k|+|c_{n+1}-d_{n+1}|$$

(by the case just done, because c_{n+1}, d_{n+1}, $\prod_{k=0}^n c_k$ and $\prod_{k=0}^n d_k$ all have modulus at most 1)

$$\le\sum_{k=0}^n|c_k-d_k|+|c_{n+1}-d_{n+1}|$$

(by the inductive hypothesis)

$$=\sum_{k=0}^{n+1}|c_k-d_k|,$$

so the induction continues.

285P Lemma Let M, $\epsilon>0$. Then there are $\eta>0$ and $y_0,\dots,y_n\in\mathbb{R}$ such that whenever X, Z are two real-valued random variables with $\mathbb{E}(|X|)\le M$, $\mathbb{E}(|Z|)\le M$ and $|\phi_X(y_j)-\phi_Z(y_j)|\le\eta$ for every $j\le n$, then $F_X(a)\le F_Z(a+\epsilon)+\epsilon$ for every $a\in\mathbb{R}$, where I write ϕ_X for the characteristic function of X and F_X for the distribution function of X.

proof Set $\delta=\frac{\epsilon}{7}>0$, $b=M/\delta$.

(a) Define $h_0:\mathbb{R}\to[0,1]$ by setting

$$h_0(x)=1\text{ if }x\le0,\ h_0(x)=1-x/\delta\text{ if }0\le x\le\delta,\ h_0(x)=0\text{ if }x\ge\delta.$$

Then h_0 is continuous. Let m be the integral part of b/δ, and for $-m\le k\le m+1$ set $h_k(x)=h_0(x-k\delta)$.

By the Stone-Weierstrass theorem (281K), there are $y_0,\dots,y_n\in\mathbb{R}$ and $c_0,\dots,c_n\in\mathbb{C}$ such that, writing $g_0(x)=\sum_{j=0}^n c_je^{iy_jx}$,

$$|h_0(x)-g_0(x)|\le\delta\text{ for every }x\in[-b-(m+1)\delta,b+m\delta],$$

$$|g_0(x)|\le1\text{ for every }x\in\mathbb{R}.$$

For $-m \le k \le m+1$, set

$$g_k(x) = g_0(x - k\delta) = \sum_{j=0}^{n} c_j e^{-iy_j k\delta} e^{iy_j x}.$$

Set $\eta = \delta/(1 + \sum_{j=0}^{n} |c_j|) > 0$.

(b) Now suppose that X, Z are random variables such that $\mathbb{E}(|X|) \le M$, $\mathbb{E}(|Z|) \le M$ and $|\phi_X(y_j) - \phi_Z(y_j)| \le \eta$ for every $j \le n$. Then for any k we have

$$\mathbb{E}(g_k(X)) = \mathbb{E}(\sum_{j=0}^{n} c_j e^{-iy_j k\delta} e^{iy_j X}) = \sum_{j=0}^{n} c_j e^{-iy_j k\delta} \phi_X(y_j),$$

and similarly

$$\mathbb{E}(g_k(Z)) = \sum_{j=0}^{n} c_j e^{-iy_j k\delta} \phi_Z(y_j),$$

so

$$|\mathbb{E}(g_k(X)) - \mathbb{E}(g_k(Z))| \le \sum_{j=0}^{n} |c_j||\phi_X(y_j) - \phi_Z(y_j)| \le \sum_{j=0}^{n} |c_j|\eta \le \delta.$$

Next,

$$|h_k(x) - g_k(x)| \le \delta \text{ for every } x \in [-b - (m+1)\delta + k\delta, b + m\delta + k\delta] \supseteq [-b, b],$$

$$|h_k(x) - g_k(x)| \le 2 \text{ for every } x,$$

$$\Pr(|X| \ge b) \le \frac{M}{b} = \delta,$$

so $\mathbb{E}(|h_k(X) - g_k(X)|) \le 3\delta$; and similarly $\mathbb{E}(|h_k(Z) - g_k(Z)|) \le 3\delta$. Putting these together,

$$|\mathbb{E}(h_k(X)) - \mathbb{E}(h_k(Z))| \le 7\delta = \epsilon$$

whenever $-m \le k \le m+1$.

(c) Now suppose that $-b \le a \le b$. Then there is a k such that $-m \le k \le m+1$ and $a \le k\delta \le a + \delta$. Since

$$\chi\,]-\infty, a] \le \chi\,]-\infty, k\delta] \le h_k \le \chi\,]-\infty, (k+1)\delta] \le \chi\,]-\infty, a + 2\delta],$$

we must have

$$\Pr(X \le a) \le \mathbb{E}(h_k(X)),$$

$$\mathbb{E}(h_k(Z)) \le \Pr(Z \le a + 2\delta) \le \Pr(Z \le a + \epsilon).$$

But this means that

$$\Pr(X \le a) \le \mathbb{E}(h_k(X)) \le \mathbb{E}(h_k(Z)) + \epsilon \le \Pr(Z \le a + \epsilon) + \epsilon$$

whenever $a \in [-b, b]$.

(d) As for the cases $a \ge b$, $a \le -b$, we surely have

$$b(1 - F_Z(b)) = b\Pr(Z \ge b) \le \mathbb{E}(|Z|) \le M,$$

so if $a \ge b$ then

$$F_X(a) \le 1 \le F_Z(a) + 1 - F_Z(b) \le F_Z(a) + \frac{M}{b} = F_Z(a) + \delta \le F_Z(a + \epsilon) + \epsilon.$$

Similarly,

$$bF_X(-b) \le \mathbb{E}(|X|) \le M,$$

so

$$F_X(a) \le \delta \le F_Z(a + \epsilon) + \epsilon$$

for every $a \le -b$. This completes the proof.

285Q Law of Rare Events: Theorem For any $M \ge 0$ and $\epsilon > 0$ there is a $\delta > 0$ such that whenever $X_0, \dots, X_n$ are independent $\{0, 1\}$-valued random variables with $\Pr(X_k = 1) = p_k \le \delta$ for every $k \le n$, and $\sum_{k=0}^{n} p_k = \lambda \le M$, and $X = X_0 + \ldots + X_n$, then

$$|\Pr(X = m) - \frac{\lambda^m}{m!} e^{-\lambda}| \le \epsilon$$

for every $m \in \mathbb{N}$.

proof (a) We should begin by calculating some characteristic functions. First, the characteristic function ϕ_k of X_k will be given by

$$\phi_k(y) = (1-p_k)e^{iy0} + p_k e^{iy1} = 1 + p_k(e^{iy}-1).$$

Next, if Z is a Poisson random variable with parameter λ (that is, if $\Pr(Z=m) = \lambda^m e^{-\lambda}/m!$ for every $m \in \mathbb{N}$; all you need to know at this point about the Poisson distribution is that $\sum_{m=0}^{\infty} \lambda^m e^{-\lambda}/m! = 1$), then its characteristic function ϕ_Z is given by

$$\phi_Z(y) = \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} e^{-\lambda} e^{iym} = e^{-\lambda} \sum_{m=0}^{\infty} \frac{(\lambda e^{iy})^m}{m!} = e^{-\lambda} e^{\lambda e^{iy}} = e^{\lambda(e^{iy}-1)}.$$

(b) Before getting down to δ's and η's, I show how to estimate $\phi_X(y) - \phi_Z(y)$. We know that

$$\phi_X(y) = \prod_{k=0}^{n} \phi_k(y)$$

(using 285I), while

$$\phi_Z(y) = \prod_{k=0}^{n} e^{p_k(e^{iy}-1)}.$$

Because $\phi_k(y)$, $e^{p_k(e^{iy}-1)}$ all have modulus at most 1 (we have

$$|e^{p_k(e^{iy}-1)}| = e^{-p_k(1-\cos y)} \le 1,)$$

285O tells us that

$$|\phi_X(y) - \phi_Z(y)| \le \sum_{k=0}^{n} |\phi_k(y) - e^{p_k(e^{iy}-1)}| = \sum_{k=0}^{n} |e^{p_k(e^{iy}-1)} - 1 - p_k(e^{iy}-1)|.$$

(c) So we have a little bit of analysis to do. To estimate $|e^z - 1 - z|$ where $\mathcal{R}\mathrm{e}\, z \le 0$, consider the function

$$g(t) = \mathcal{R}\mathrm{e}(c(e^{tz} - 1 - tz))$$

where $|c| = 1$. We have $g(0) = g'(0) = 0$ and

$$|g''(t)| = |\,\mathcal{R}\mathrm{e}(c(z^2 e^{tz}))| \le |c||z^2||e^{tz}| \le |z|^2$$

for every $t \ge 0$, so that

$$|g(1)| \le \frac{1}{2}|z|^2$$

by the (real-valued) Taylor theorem with remainder, or otherwise. As c is arbitrary,

$$|e^z - 1 - z| \le \frac{1}{2}|z|^2$$

whenever $\mathcal{R}\mathrm{e}\, z \le 0$. In particular,

$$|e^{p_k(e^{iy}-1)} - 1 - p_k(e^{iy}-1)| \le \frac{1}{2}p_k^2|e^{iy}-1|^2 \le 2p_k^2$$

for each k, and

$$|\phi_X(y) - \phi_Z(y)| \le \sum_{k=0}^{n} |e^{p_k(e^{iy}-1)} - 1 - p_k(e^{iy}-1)| \le 2\sum_{k=0}^{n} p_k^2$$

for each $y \in \mathbb{R}$.

(d) Now for the detailed estimates. Given $M \ge 0$ and $\epsilon > 0$, let $\eta > 0$ and $y_0, \dots, y_l \in \mathbb{R}$ be such that

$$\Pr(X \le a) \le \Pr(Z \le a + \frac{1}{2}) + \frac{\epsilon}{2}$$

whenever X, Z are real-valued random variables, $\mathbb{E}(|X|) \le M$, $\mathbb{E}(|Z|) \le M$ and $|\phi_X(y_j) - \phi_X(y_j)| \le \eta$ for every $j \le l$ (285P). Take $\delta = \eta/(2M+1)$ and suppose that $X_0, \dots, X_n$ are independent $\{0,1\}$-valued random variables with $\Pr(X_k = 1) = p_k \le \delta$ for every $k \le n$, $\lambda = \sum_{k=0}^{n} p_k \le M$. Set $X = X_0 + \ldots + X_n$ and let Z be a Poisson random variable with parameter λ; then by the arguments of (a)-(c),

$$|\phi_X(y) - \phi_Z(y)| \le 2\sum_{k=0}^{n} p_k^2 \le 2\delta\sum_{k=0}^{n} p_k = 2\delta\lambda \le \eta$$

for every $y \in \mathbb{R}$. Also

$$\mathbb{E}(|X|) = \mathbb{E}(X) = \sum_{k=0}^{n} p_k = \lambda \le M,$$

$$\mathbb{E}(|Z|) = \mathbb{E}(Z) = \sum_{m=0}^{\infty} m\frac{\lambda^m}{m!}e^{-\lambda} = e^{-\lambda}\sum_{m=1}^{\infty}\frac{\lambda^m}{(m-1)!} = e^{-\lambda}\sum_{m=0}^{\infty}\frac{\lambda^{m+1}}{m!} = \lambda \le M.$$

So

$$\Pr(X \le a) \le \Pr(Z \le a + \frac{1}{2}) + \frac{\epsilon}{2},$$

$$\Pr(Z \le a) \le \Pr(X \le a + \frac{1}{2}) + \frac{\epsilon}{2}$$

for every a. But as both X and Z take all their values in $\mathbb{N}$,

$$|\Pr(X \le m) - \Pr(Z \le m)| \le \frac{\epsilon}{2}$$

for every $m \in \mathbb{N}$, and

$$|\Pr(X = m) - \frac{\lambda^m}{m!}e^{-\lambda}| = |\Pr(X = m) - \Pr(Z = m)| \le \epsilon$$

for every $m \in \mathbb{N}$, as required.

285R Convolutions Recall from 257A that if ν, $\tilde{\nu}$ are Radon probability measures on $\mathbb{R}^r$ then they have a convolution $\nu * \tilde{\nu}$ defined by writing

$$(\nu * \tilde{\nu})(E) = (\nu \times \tilde{\nu})\{(x, y) : x + y \in E\}$$

for every Borel set $E \subseteq \mathbb{R}^r$, which is also a Radon probability measure. We can readily compute the characteristic function $\phi_{\nu*\tilde{\nu}}$ from 257B: we have

$$\begin{aligned}\phi_{\nu*\tilde{\nu}}(y) &= \int e^{iy\,\cdot\,x}(\nu * \tilde{\nu})(dx) = \int e^{iy\,\cdot\,(x+x')}\nu(dx)\tilde{\nu}(dx') \\ &= \int e^{iy\,\cdot\,x}e^{iy\,\cdot\,x'}\nu(dx)\tilde{\nu}(dx') = \int e^{iy\,\cdot\,x}\nu(dx)\int e^{iy\,\cdot\,x'}\tilde{\nu}(dx') = \phi_\nu(y)\phi_{\tilde{\nu}}(y).\end{aligned}$$

(Thus convolution of measures corresponds to pointwise multiplication of characteristic functions, just as convolution of functions corresponds to pointwise multiplication of Fourier transforms.) Recalling that the sum of independent random variables corresponds to convolution of their distributions (272T), this gives another way of looking at 285I. Remember also that if ν, $\tilde{\nu}$ have Radon-Nikodým derivatives f, $\tilde{f}$ with respect to Lebesgue measure then $f * \tilde{f}$ is a Radon-Nikodým derivative of $\nu * \tilde{\nu}$ (257F).

285S The vague topology and pointwise convergence of characteristic functions In 285L we saw that a sequence $\langle\nu_n\rangle_{n\in\mathbb{N}}$ of Radon probability measures on $\mathbb{R}^r$ converges in the vague topology to a Radon probability measure ν if and only if

$$\lim_{n\to\infty}\int e^{iy\,\cdot\,x}\nu_n(dx) = \int e^{iy\,\cdot\,x}\nu(dx)$$

for every $y \in \mathbb{R}^r$; that is, iff

$$\lim_{n\to\infty}\rho'_y(\nu_n, \nu) = 0 \text{ for every } y \in \mathbb{R}^r,$$

writing

$$\rho'_y(\nu, \nu') = |\int e^{iy\,\cdot\,x}\nu(dx) - \int e^{iy\,\cdot\,x}\nu'(dx)|$$

for Radon probability measures ν, ν' on $\mathbb{R}^r$ and $y \in \mathbb{R}^r$. It is natural to ask whether the pseudometrics ρ'_y actually define the vague topology. Writing $\mathfrak{T}$ for the vague topology and $\mathfrak{S}$ for the topology defined by $\{\rho'_y : y \in \mathbb{R}^r\}$, we surely have $\mathfrak{S} \subseteq \mathfrak{T}$, just because every ρ'_y is one of the pseudometrics used in the definition of $\mathfrak{T}$. Also we know that $\mathfrak{S}$ and $\mathfrak{T}$ give the same convergent sequences, and incidentally that $\mathfrak{T}$ is metrizable (see 285Xq). But all this does not quite amount to saying that the two topologies are the same, and indeed they are not, as the next result shows.

285T Proposition Let $y_0, \dots, y_n \in \mathbb{R}$ and $\eta > 0$. Then there are infinitely many $m \in \mathbb{N}$ such that $|1 - e^{iy_k m}| \le \eta$ for every $k \le n$.

proof Let $\eta_1, \dots, \eta_r \in \mathbb{R}$ be such that $1 = \eta_0, \eta_1, \dots, \eta_r$ are linearly independent over $\mathbb{Q}$ and every $y_k/2\pi$ is a linear combination of the η_j over $\mathbb{Q}$; say $y_k = 2\pi\sum_{j=0}^{r} q_{kj}\eta_j$ where every $q_{kj} \in \mathbb{Q}$. Express the q_{kj} as p_{kj}/p where each $p_{kj} \in \mathbb{Z}$ and $p \in \mathbb{N} \setminus \{0\}$. Set $M = \max_{k\le n}\sum_{j=0}^{r}|p_{kj}|$.

Take any $m_0 \in \mathbb{N}$ and let $\delta > 0$ be such that $|1 - e^{2\pi ix}| \le \eta$ whenever $|x| \le 2\pi M\delta$. By Weyl's Equidistribution Theorem (281N), there are infinitely many m such that $<m\eta_j> \le \delta$ whenever $1 \le j \le r$; in particular, there is such an $m \ge m_0$. Let m_j be the integral part of $m\eta_j$, so that $|m\eta_j - m_j| \le \delta$ for $0 \le j \le r$. Then

$$|mpy_k - 2\pi\sum_{j=0}^{r} p_{kj}m_j| \le 2\pi\sum_{j=0}^{r} |p_{kj}||m\eta_j - m_j| \le 2\pi M\delta,$$

so that

$$|1 - e^{iy_k mp}| = |1 - \exp(i(mpy_k - 2\pi\sum_{j=0}^{r} p_{kj}m_j))| \le \eta$$

for every $k \le n$. As $mp \ge m_0$ and m_0 is arbitrary, this proves the result.

285U Corollary The topologies $\mathfrak{S}$ and $\mathfrak{T}$ on the space of Radon probability measures on $\mathbb{R}$, as described in 285S, are different.

proof Let δ_x be the Dirac measure on $\mathbb{R}$ concentrated at x. By 285T, every member of $\mathfrak{S}$ which contains δ_0 also contains δ_m for infinitely many $m \in \mathbb{N}$. On the other hand, the set

$$G = \{\nu : \int e^{-x^2}\nu(dx) > \frac{1}{2}\}$$

is a member of $\mathfrak{T}$, containing δ_0, which does not contain δ_m for any integer $m \ne 0$. So $G \in \mathfrak{T} \setminus \mathfrak{S}$ and $\mathfrak{T} \ne \mathfrak{S}$.

285X Basic exercises (a) Let ν be a Radon probability measure on $\mathbb{R}^r$, where $r \ge 1$, and suppose that $\int \|x\|\nu(dx) < \infty$. Show that the characteristic function ϕ of ν is differentiable (in the full sense of 262Fa) and that $\frac{\partial\phi}{\partial\eta_j}(y) = i\int \xi_j e^{iy\cdot x}\nu(dx)$ for every $j \le r$, $y \in \mathbb{R}^r$, using ξ_j, η_j to represent the coordinates of x and y as usual.

>(b) Let $\boldsymbol{X} = (X_1, \dots, X_r)$ be a family of real-valued random variables, with characteristic function $\phi_{\boldsymbol{X}}$. Show that the characteristic function ϕ_{X_j} of X_j is given by

$$\phi_{X_j}(y) = \phi_{\boldsymbol{X}}(ye_j) \text{ for every } y \in \mathbb{R},$$

where e_j is the jth unit vector of $\mathbb{R}^r$.

>(c) Let X be a real-valued random variable and ϕ_X its characteristic function. Show that

$$\phi_{aX+b}(y) = e^{iyb}\phi_X(ay)$$

for any a, b, $y \in \mathbb{R}$.

(d) Let X be a real-valued random variable and ϕ its characteristic function.
(i) Show that for any integrable complex-valued function h on $\mathbb{R}$,

$$\mathbb{E}(\hat{h}(X)) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \phi(-y)h(y)dy,$$

writing $\hat{h}$ for the Fourier transform of h.
(ii) Show that for any rapidly decreasing test function h,

$$\mathbb{E}(h(X)) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} \phi(y)\hat{h}(y)dy.$$

(e) Let ν be a Radon probability measure on $\mathbb{R}$, and suppose that its characteristic function ϕ is square-integrable. Show that ν is an indefinite-integral measure over Lebesgue measure and that its Radon-Nikodým derivatives are also square-integrable. (*Hint*: use 284O to find a square-integrable f such that $\int h \times f = \frac{1}{\sqrt{2\pi}}\int \phi \times \hat{h}$ for every rapidly decreasing test function h, and ideas from the proof of 284G to show that $\int_a^b f = \nu\,]a, b[$ whenever $a < b$ in $\mathbb{R}$.)

>(f) Let $\boldsymbol{X} = (X_1, \dots, X_r)$ be a family of real-valued random variables with characteristic function $\phi_{\boldsymbol{X}}$. Suppose that $\phi_{\boldsymbol{X}}$ is expressible in the form

$$\phi_{\boldsymbol{X}}(y) = \prod_{j=1}^{r} \phi_j(\eta_j)$$

for some functions $\phi_1, \dots, \phi_r$, writing $y = (\eta_1, \dots, \eta_r)$ as usual. Show that $X_1, \dots, X_r$ are independent. (*Hint*: show that the ϕ_j must be the characteristic functions of the X_j; now show that the distribution of $\boldsymbol{X}$ has the same characteristic function as the product of the distributions of the X_j.)

(g) Let X_1, X_2 be independent real-valued random variables with the same distribution, and ϕ the characteristic function of $X_1 - X_2$. Show that $\phi(t) = \phi(-t) \ge 0$ for every $t \in \mathbb{R}$.

(h) Let ν be a Radon probability measure on $\mathbb{R}$, with characteristic function ϕ. Show that

$$\frac{1}{2}(\nu[c,d] + \nu\,]c,d[) = \frac{i}{2\pi}\lim_{a\to\infty}\int_{-a}^{a}\frac{e^{-idy}-e^{-icy}}{y}\phi(y)dy$$

whenever $c \le d$ in $\mathbb{R}$. (*Hint*: use part (a) of the proof of 283F.)

(i) Let X be a real-valued random variable and ϕ_X its characteristic function. Show that

$$\Pr(|X| \ge a) \le 7a\int_0^{1/a}(1 - \mathcal{R}\mathrm{e}(\phi_X(y))dy$$

for every $a > 0$.

(j) We say that a set Q of Radon probability measures on $\mathbb{R}$ is **uniformly tight** if for every $\epsilon > 0$ there is an $M \ge 0$ such that $\nu(\mathbb{R} \setminus [-M,M]) \le \epsilon$ for every $\nu \in Q$. Show that if Q is any uniformly tight family of Radon probability measures on $\mathbb{R}$, and $\epsilon > 0$, then there are $\eta > 0$ and $y_0, \dots, y_n \in \mathbb{R}$ such that

$$\nu\,]-\infty, a] \le \nu'\,]-\infty, a+\epsilon] + \epsilon$$

whenever ν, $\nu' \in Q$ and $|\phi_\nu(y_j) - \phi_{\nu'}(y_j)| \le \eta$ for every $j \le n$, writing ϕ_ν for the characteristic function of ν.

(k) Let $\langle\nu_n\rangle_{n\in\mathbb{N}}$ be a sequence of Radon probability measures on $\mathbb{R}$, and suppose that it converges for the vague topology to a Radon probability measure ν. Show that $\{\nu\} \cup \{\nu_n : n \in \mathbb{N}\}$ is uniformly tight in the sense of 285Xj.

>(l) Let ν, ν' be two totally finite Radon measures on $\mathbb{R}^r$ which agree on all closed half-spaces, that is, sets of the form $\{x : x\centerdot y \ge c\}$ for $c \in \mathbb{R}$, $y \in \mathbb{R}^r$. Show that $\nu = \nu'$. (*Hint*: reduce to the case $\nu\mathbb{R}^r = \nu'\mathbb{R}^r = 1$ and use 285M.)

>(m) For $\gamma > 0$, the **Cauchy distribution** with centre 0 and scale parameter γ is the Radon probability measure ν_γ defined by the formula

$$\nu_\gamma(E) = \frac{\gamma}{\pi}\int_E \frac{1}{\gamma^2+t^2}dt.$$

(i) Show that if X is a random variable with distribution ν_γ then $\Pr(X \ge 0) = \Pr(|X| \ge \gamma) = \frac{1}{2}$. (ii) Show that the characteristic function of ν_γ is $y \mapsto e^{-\gamma|y|}$. (*Hint*: 283Xr.) (iii) Show that if X and Y are independent random variables with Cauchy distributions, both centered at 0 and with scale parameters γ, δ respectively, and α, β are not both 0, then $\alpha X + \beta Y$ has a Cauchy distribution centered at 0 and with scale parameter $|\alpha|\gamma + |\beta|\delta$. (iv) Show that if X and Y are independent normally distributed random variables with expectation 0 then X/Y has a Cauchy distribution.

>(n) Let $X_1, X_2, \dots$ be an independent identically distributed sequence of random variables, all of zero expectation and variance 1; let ϕ be their common characteristic function. For each $n \ge 1$, set $S_n = \frac{1}{\sqrt{n}}(X_1 + \ldots + X_n)$.

(i) Show that the characteristic function ϕ_n of S_n is given by the formula $\phi_n(y) = (\phi(\frac{y}{\sqrt{n}}))^n$ for each n.

(ii) Show that $|\phi_n(y) - e^{-y^2/2}| \le n|\phi(\frac{y}{\sqrt{n}}) - e^{-y^2/2n}|$.

(iii) Setting $h(y) = \phi(y) - e^{-y^2/2}$, show that $h(0) = h'(0) = h''(0) = 0$ and therefore that $\lim_{n\to\infty} nh(y/\sqrt{n}) = 0$, so that $\lim_{n\to\infty}\phi_n(y) = e^{-y^2/2}$ for every $y \in \mathbb{R}$.

(iv) Show that $\lim_{n\to\infty}\Pr(S_n \le a) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-x^2/2}dx$ for every $a \in \mathbb{R}$.

>(o) A random variable X has a **Poisson distribution** with parameter $\lambda > 0$ if $\Pr(X = n) = e^{-\lambda}\lambda^n/n!$ for every $n \in \mathbb{N}$. (i) Show that in this case $\mathbb{E}(X) = \mathrm{Var}(X) = \lambda$. (ii) Show that if X and Y are independent random variables with Poisson distributions then $X + Y$ has a Poisson distribution. (iii) Find a proof of (ii) based on 285Q.

>(p) For $x \in \mathbb{R}^r$, let δ_x be the Dirac measure on $\mathbb{R}^r$ concentrated at x. Show that $\delta_x * \delta_y = \delta_{x+y}$ for all x, $y \in \mathbb{R}^r$.

(q) Let P be the set of Radon probability measures on $\mathbb{R}^r$. For $y \in \mathbb{R}^r$, set $\rho'_y(\nu,\nu') = |\phi_\nu(y) - \phi_{\nu'}(y)|$ for all ν, $\nu' \in P$, writing ϕ_ν for the characteristic function of ν. Set $\psi(x) = \frac{1}{(\sqrt{2\pi})^r}e^{-x\centerdot x/2}$ for $x \in \mathbb{R}^r$. Show that the vague topology on P is defined by the family $\{\rho_\psi\} \cup \{\rho'_y : y \in \mathbb{Q}^r\}$, defining ρ_ψ as in 285K, and is therefore metrizable. (*Hint*: 281K; cf. 285Xj.)

>(r) Let $\phi : \mathbb{R}^r \to \mathbb{C}$ be the characteristic function of a Radon probability measure on $\mathbb{R}^r$. Show that $\phi(0) = 1$ and that $\sum_{j=0}^n \sum_{k=0}^n c_j \bar{c}_k \phi(a_j - a_k) \ge 0$ whenever $a_0, \dots, a_n \in \mathbb{R}^r$ and $c_0, \dots, c_n \in \mathbb{C}$. ('Bochner's theorem' states that these conditions are sufficient, as well as necessary, for ϕ to be a characteristic function; see 445N in Volume 4.)

(s) Let $\langle X_n \rangle_{n\in\mathbb{N}}$ be an independent sequence of real-valued random variables and set $S_n = \sum_{j=0}^n X_j$ for each $n \in \mathbb{N}$. Suppose that the sequence $\langle \nu_{S_n} \rangle_{n\in\mathbb{N}}$ of distributions is convergent for the vague topology to a distribution. Show that $\langle S_n \rangle_{n\in\mathbb{N}}$ converges in measure, therefore a.e. (*Hint*: 285J, 273B.)

(t) Let X be a normal random variable with expectation a and variance σ^2. Show that $\mathbb{E}(e^X) = \exp(a + \frac{1}{2}\sigma^2)$.

285Y Further exercises (a) Let ν be a Radon probability measure on $\mathbb{R}^r$. Write

$$\hat{\nu}(y) = \frac{1}{(\sqrt{2\pi})^r} \int e^{-iy\cdot x} \nu(dx)$$

for every $y \in \mathbb{R}^r$.

(i) Writing ϕ_ν for the characteristic function of ν, show that $\hat{\nu}(y) = \frac{1}{(\sqrt{2\pi})^r} \phi_\nu(-y)$ for every $y \in \mathbb{R}^r$.

(ii) Show that $\int \hat{\nu}(y) h(y) dy = \int \hat{h}(x) \nu(dx)$ for any Lebesgue integrable complex-valued function h on $\mathbb{R}^r$, defining the Fourier transform $\hat{h}$ as in 283Wa.

(iii) Show that $\int \check{h}(x) \nu(dx) = \int h(y) \hat{\nu}(y) dy$ for any rapidly decreasing test function h on $\mathbb{R}^r$.

(iv) Show that if ν is an indefinite-integral measure over Lebesgue measure, with Radon-Nikodým derivative f, then $\hat{\nu}$ is the Fourier transform of f.

(b) Let ν be a Radon probability measure on $\mathbb{R}^r$, with characteristic function ϕ. Show that whenever $c \le d$ in $\mathbb{R}^r$ then

$$\Big(\frac{i}{2\pi}\Big)^r \lim_{\alpha_1, \dots, \alpha_r \to \infty} \int_{-\alpha_1}^{\alpha_1} \dots \int_{-\alpha_r}^{\alpha_r} \Big(\prod_{j=1}^r \frac{e^{-i\delta_j \eta_j} - e^{-i\gamma_j \eta_j}}{\eta_j}\Big) \phi(y) dy$$

exists and lies between $\nu\,]c, d[$ and $\nu[c, d]$, writing $]c, d[= \prod_{j\le r}]\gamma_j, \delta_j[$ if $c = (\gamma_1, \dots, \gamma_r)$ and $d = (\delta_1, \dots, \delta_r)$.

(c) Let $\langle X_n \rangle_{n\in\mathbb{N}}$ be an independent identically distributed sequence of (not-essentially-constant) random variables, and set $S_n = \sum_{k=0}^n X_{2k+1} - X_{2k}$ for each $n \in \mathbb{N}$. Show that $\lim_{n\to\infty} \Pr(|S_n| \ge \alpha) = 1$ for every $\alpha \in \mathbb{R}$. (*Hint*: 285Xg, proof of 285J.) Hence, or otherwise, show that $\lim_{n\to\infty} \Pr(|\sum_{k=0}^n X_k| \ge \alpha) = 1$ for every $\alpha \in \mathbb{R}$.

(d) For Radon probability measures ν, ν' on $\mathbb{R}^r$ set

$$\rho(\nu, \nu') = \inf\{\epsilon : \epsilon \ge 0,\ \nu\,]-\infty, a] \le \nu'\,]-\infty, a + \epsilon\mathbf{1}] + \epsilon \le \nu\,]-\infty, a + 2\epsilon\mathbf{1}] + 2\epsilon \quad \text{for every } a \in \mathbb{R}^r\},$$

writing $]-\infty, a] = \{(\xi_1, \dots, \xi_r) : \xi_j \le \alpha_j \text{ for every } j \le r\}$ when $a = (\alpha_1, \dots, \alpha_r)$, and $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^r$. Show that ρ is a metric on the set of Radon probability measures on $\mathbb{R}^r$, and that the topology it defines is the vague topology. (Cf. 274Ya.)

(e) Let $r \ge 1$. We say that a set Q of Radon probability measures on $\mathbb{R}^r$ is **uniformly tight** if for every $\epsilon > 0$ there is a compact set $K \subseteq \mathbb{R}^r$ such that $\nu(\mathbb{R}^r \setminus K) \le \epsilon$ for every $\nu \in Q$. Show that if Q is any uniformly tight family of Radon probability measures on $\mathbb{R}^r$, and $\epsilon > 0$, then there are $\eta > 0$, $y_0, \dots, y_n \in \mathbb{R}^r$ such that $\nu\,]-\infty, a] \le \nu'\,]-\infty, a + \epsilon\mathbf{1}] + \epsilon$ whenever ν, $\nu' \in Q$ and $a \in \mathbb{R}^r$ and $|\phi_\nu(y_j) - \phi_{\nu'}(y_j)| \le \eta$ for every $j \le n$, writing ϕ_ν for the characteristic function of ν.

(f) Show that for any $M \ge 0$ the set of Radon probability measures ν on $\mathbb{R}^r$ such that $\int \|x\| \nu(dx) \le M$ is uniformly tight in the sense of 285Ye.

(g) Let $C_b(\mathbb{R}^r)$ be the Banach space of bounded continuous real-valued functions on $\mathbb{R}^r$.

(i) Show that any Radon probability measure ν on $\mathbb{R}^r$ corresponds to a continuous linear functional $h_\nu : C_b(\mathbb{R}^r) \to \mathbb{R}$, writing $h_\nu(f) = \int f d\nu$ for $f \in C_b(\mathbb{R}^r)$.

(ii) Show that if $h_\nu = h_{\nu'}$ then $\nu = \nu'$.

(iii) Show that the vague topology on the set of Radon probability measures corresponds to the weak* topology on the dual $(C_b(\mathbb{R}^r))^*$ of $C_b(\mathbb{R}^r)$.

(h) Let $r \geq 1$ and let P be the set of Radon probability measures on $\mathbb{R}^r$. For $m \in \mathbb{N}$ let ρ^*_m be the pseudometric on P defined by setting $\rho^*_m(\nu, \nu') = \sup_{\|y\| \leq m} |\phi_\nu(y) - \phi_{\nu'}(y)|$ for ν, $\nu' \in P$, writing ϕ_ν for the characteristic function of ν. Show that $\{\rho^*_m : m \in \mathbb{N}\}$ defines the vague topology on P.

(i) Let $r \geq 1$ and let P be the set of Radon probability measures on $\mathbb{R}^r$. For $m \in \mathbb{N}$ let $\tilde{\rho}^*_m$ be the pseudometric on P defined by setting

$$\tilde{\rho}^*_m(\nu, \nu') = \int_{\{y : \|y\| \leq m\}} |\phi_\nu(y) - \phi_{\nu'}(y)| dy$$

for ν, $\nu' \in P$, writing ϕ_ν for the characteristic function of ν. Show that $\{\tilde{\rho}^*_m : m \in \mathbb{N}\}$ defines the vague topology on P.

(j) Let X be a real-valued random variable with finite variance. Show that for any $\eta \geq 0$,

$$|\phi(y) - 1 - iy\mathbb{E}(X) + \frac{1}{2}y^2\mathbb{E}(X^2)| \leq \frac{1}{6}\eta|y^3|\mathbb{E}(X^2) + y^2\mathbb{E}(\psi_\eta(X)),$$

writing ϕ for the characteristic function of X and $\psi_\eta(x) = 0$ for $|x| \leq \eta$, x^2 for $|x| > \eta$.

(k) Suppose that $\epsilon \geq \delta > 0$ and that $X_0, \dots , X_n$ are independent real-valued random variables such that

$$\mathbb{E}(X_k) = 0 \text{ for every } k \leq n, \quad \textstyle\sum_{k=0}^n \mathrm{Var}(X_k) = 1, \quad \sum_{k=0}^n \mathbb{E}(\psi_\delta(X_k)) \leq \delta$$

(writing $\psi_\delta(x) = 0$ if $|x| \leq \delta$, x^2 if $|x| > \delta$). Set $\gamma = \epsilon/\sqrt{\delta^2 + \delta}$, and let Z be a standard normal random variable. Show that

$$|\phi(y) - e^{-y^2/2}| \leq \frac{1}{3}\epsilon|y|^3 + y^2(\delta + \mathbb{E}(\psi_\gamma(Z)))$$

for every $y \in \mathbb{R}$, writing ϕ for the characteristic function of $X = \sum_{k=0}^n X_k$. (*Hint*: write ϕ_k for the characteristic function of X_k and $\tilde{\phi}_k$ for the characteristic function of $\sigma_k Z$, where $\sigma_k = \sqrt{\mathrm{Var}(X_k)}$. Show that

$$|\phi_k(y) - \tilde{\phi}_k(y)| \leq \frac{1}{3}\epsilon|y^3|\sigma_k^2 + y^2\big(\mathbb{E}(\psi_\epsilon(X_k)) + \sigma_k^2\mathbb{E}(\psi_\gamma(Z))\big).)$$

(l) Show that for every $\epsilon > 0$ there is a $\delta > 0$ such that whenever $X_0, \dots , X_n$ are independent real-valued random variables such that

$$\mathbb{E}(X_k) = 0 \text{ for every } k \leq n, \quad \textstyle\sum_{k=0}^n \mathrm{Var}(X_k) = 1, \quad \sum_{k=0}^n \mathbb{E}(\psi_\delta(X_k)) \leq \delta$$

(writing $\psi_\delta(x) = 0$ if $|x| \leq \delta$, x^2 if $|x| > \delta$), then $|\phi(y) - e^{-y^2/2}| \leq \epsilon(y^2 + |y^3|)$ for every $y \in \mathbb{R}$, writing ϕ for the characteristic function of $X = X_0 + \ldots + X_n$.

(m) Use 285Yl to prove Lindeberg's theorem (274F).

(n) Let $r \geq 1$ and let P be the set of Radon probability measures on $\mathbb{R}^r$. Show that convolution, regarded as a map from $P \times P$ to P, is continuous when P is given the vague topology. (*Hint*: 281Xa and 257B will help.)

(o) Let $\mathfrak{S}$ be the topology on $\mathbb{R}$ defined by $\{\rho'_y : y \in \mathbb{R}\}$, where $\rho'_y(x, x') = |e^{iyx} - e^{iyx'}|$ (compare 285S). Show that addition and subtraction are continuous for $\mathfrak{S}$ in the sense of 2A5A.

(p) Let ν be a Radon probability measure on $\mathbb{R}^r$ with bounded support (definition: 256Xf). Show that its characteristic function is smooth.

(q) Let (Ω, Σ, μ) be a probability space. Suppose that $\langle X_n \rangle_{n\in\mathbb{N}}$ is a sequence of real-valued random variables on Ω, and X another real-valued random variable on Ω; let ϕ_{X_n}, ϕ_X be the corresponding characteristic functions. Show that the following are equiveridical: (i) $\lim_{n\to\infty} \mathbb{E}(f(X_n)) = \mathbb{E}(f(X))$ for every bounded continuous function $f : \mathbb{R} \to \mathbb{R}$; (ii) $\lim_{n\to\infty} \phi_{X_n}(y) = \phi_X(y)$ for every $y \in \mathbb{R}$. (In this case we say that $\langle X_n \rangle_{n\in\mathbb{N}}$ **converges in distribution** to X.)

(r) Let (Ω, Σ, μ) be a probability space, and P the set of Radon probability measures on $\mathbb{R}$. (i) Show that we have a function $\psi : L^0(\mu) \to P$ defined by saying that $\psi(X^\bullet)$ is the distribution of X whenever X is a real-valued random variable on Ω. (ii) Show that ψ is continuous for the topology of convergence in measure on $L^0(\mu)$ and the vague topology on P. (Compare 271Yd.)

285 Notes and comments Just as with Fourier transforms, the power of methods which use the characteristic functions of distributions is based on three points: (i) the characteristic function of a distribution determines the distribution (285M); (ii) the properties of interest in a distribution are reflected in accessible properties of its characteristic function (285G, 285I, 285J) (iii) these properties of the characteristic function are actually *different* from the corresponding properties of the distribution, and are amenable to different kinds of investigation. Above all, the fact that (for sequences!) convergence in the vague topology of distributions corresponds to pointwise convergence for characteristic functions (285L) provides us with a path to the classic limit theorems, as in 285Q and 285Xn. In 285S-285U I show that this result for sequences does not correspond immediately to any alternative characterization of the vague topology, though it can be adapted in more than one way to give such a characterization (see 285Yh-285Yi).

Concerning the Central Limit Theorem there is one conspicuous difference between the method suggested here and that of §274. The previous approach offered at least a theoretical possibility of giving an explicit formula for δ in 274F as a function of ϵ, and hence an estimate of the rate of convergence to be expected in the Central Limit Theorem. The arguments in the present chapter, involving as they do an entirely non-constructive compactness argument in 281A, leave us with no way of achieving such an estimate. But in fact the method of characteristic functions, suitably refined, is the basis of the best estimates known, such as the Berry-Esséen theorem (274Hc).

In 285D I try to show how the characteristic function ϕ_ν of a Radon probability measure can be related to a 'Fourier transform' $\hat{\nu}$ of ν which corresponds directly to the Fourier transforms of functions discussed in §§283-284. If f is a non-negative Lebesgue integrable function and we take ν to be the corresponding indefinite-integral measure, then $\hat{\nu} = \hat{f}$. Thus the concept of 'Fourier transform of a measure' is a natural extension of the Fourier transform of an integrable function. Looking at it from the other side, the formula of 285Dc shows that ν can be thought of as representing the inverse Fourier transform of $\hat{\nu}$ in the sense of 284H-284I. Taking ν to be the measure which assigns a mass 1 to the point 0, we get the Dirac delta function, with Fourier transform the constant function $\chi\mathbb{R}$. These ideas can be extended without difficulty to handle convolutions of measures (285R).

It is a striking fact that while there is no satisfactory characterization of the functions which are Fourier transforms of integrable functions, there is a characterization of the characteristic functions of probability distributions. This is 'Bochner's theorem'. I give the condition in 285Xr, asking you to prove its necessity as an exercise; we already have three-quarters of the machinery to prove its sufficiency, but the last step will have to wait for Volume 4.

286 Carleson's theorem

Carleson's theorem (Carleson 66) was the (unexpected) solution to a long-standing problem. Remarkably, it can be proved by 'elementary' arguments. The hardest part of the work below, in 286I-286L, involves only the laborious verification of inequalities. How the inequalities were chosen is a different matter; for once, some of the ideas of the proof lie in the statements of the lemmas. The argument here is a greatly expanded version of Lacey & Thiele 00.

The Hardy-Littlewood Maximal Theorem (286A) is important, and worth learning even if you leave the rest of the section as an unexamined monument. I bring 286B-286D forward to the beginning of the section, even though they are little more than worked exercises, because they also have potential uses in other contexts.

The complexity of the argument is such that it is useful to introduce a substantial number of special notations. Rather than include these in the general index, I give a list in 286W. Among them are ten constants $C_1, \ldots, C_{10}$. The values of these numbers are of no significance. The method of proof here is quite inappropriate if we want to estimate rates of convergence. I give recipes for the calculation of the C_n only for the sake of the linear logic in which this treatise is written, and because they occasionally offer clues concerning the tactics being used.

In this section all integrals are with respect to Lebesgue measure μ on $\mathbb{R}$ unless otherwise stated.

286A The Maximal Theorem Suppose that $1 < p < \infty$ and that $f \in \mathcal{L}^p_{\mathbb{C}}(\mu)$ (definition: 244P). Set

$$f^*(x) = \sup\{\frac{1}{b-a}\int_a^b |f| : a \le x \le b,\, a < b\}$$

for $x \in \mathbb{R}$. Then $\|f^*\|_p \le \frac{2^{1/p}p}{p-1}\|f\|_p$.

proof (a) It is enough to consider the case $f \ge 0$. Note that if $E \subseteq \mathbb{R}$ has finite measure, then

$$\int_E f = \int (f \times \chi E) \times \chi E \le \|f \times \chi E\|_p (\mu E)^{1/q} \le \|f\|_p (\mu E)^{1/q}$$

is finite, where $q = \frac{p}{p-1}$, by Hölder's inequality (244Eb). Consequently, if $t > 0$ and $\int_E f \ge t\mu E$, we must have $t\mu E \le \|f \times \chi E\|_p (\mu E)^{1/q}$ and

$$\mu E = (\mu E)^{p-p/q} \le \frac{1}{t^p}\|f \times \chi E\|_p^p = \frac{1}{t^p}\int_E f^p.$$

(b) For $t > 0$, set

$$G_t = \{x : \int_x^a f > (a-x)t \text{ for some } a > x\}.$$

(i) G_t is an open set. **P** For any $a \in \mathbb{R}$,

$$G_{ta} = \{x : x < a,\ \int_x^a f > (a-x)t\}$$

is open, because $x \mapsto \int_x^a f$ and $x \mapsto (a-x)t$ are continuous (225A); so $G_t = \bigcup_{a\in\mathbb{R}} G_{ta}$ is open. **Q**

(ii) By 2A2I, there is a partition $\mathcal{C}$ of G_t into open intervals. Now C is bounded and $t\mu C \le \int_C f$ for every $C \in \mathcal{C}$. **P** Express C as $]a, b[$ (for the moment, we have to allow for the possibility that one or both of a, b is infinite).

(α) If $x \in C$, there is some (finite) $c > x$ such that $\int_x^c f > (c-x)t$. Set $d = \min(b, c) > x$. If $d = c$, then of course $\int_x^d f > (d-x)t$. If $d = b < c$, then (because $b \notin G_t$) $\int_b^c f \le (c-b)t$, so again

$$\int_x^d f = \int_x^b f = \int_x^c f - \int_b^c f > (c-x)t - (c-b)t = (b-x)t = (d-x)t.$$

Thus we always have some $d \in]x, b]$ such that $\int_x^d f > (d-x)t$.

(β) Now take any $z \in C$, and consider

$$A_z = \{x : z \le x \le b,\ \int_z^x f \ge (x-z)t\}.$$

Then $z \in A_z$, and A_z is closed, again because the functions $x \mapsto \int_z^x f$ and $x \mapsto (x-z)t$ are continuous. Moreover, A_z is bounded, because $x - z \le \frac{1}{t^p}\|f\|_p^p$ for every $x \in A_z$, by (a). **?** If $\sup A_z = x_0 < b$, then $x_0 \in A_z$, and there is a $d \in]x_0, b]$ such that $\int_{x_0}^d f \ge t(d-x_0)$, by ($\alpha$); but in this case $d \in A_z$, which is impossible. **X** Thus $b = \sup A_z \in A_z$ (in particular, $b < \infty$), and $\int_z^b f \ge (b-z)t$.

(γ) Letting z decrease to a, we see that $b - a \le \frac{1}{t^p}\|f\|_p^p$, so a is finite, and also

$$\int_a^b f = \lim_{z\downarrow a}\int_z^b f \ge \lim_{z\downarrow a}(b-z)t = (b-a)t,$$

as required. **Q**

(iii) Accordingly, because $\mathcal{C}$ is countable and f is non-negative,

$$\mu G_t = \sum_{C\in\mathcal{C}} \mu C \le \sum_{C\in\mathcal{C}} \frac{1}{t^p}\int_C f^p \le \frac{1}{t^p}\int_{-\infty}^{\infty} f^p$$

is finite, and

$$\int_{G_t} f = \sum_{C\in\mathcal{C}} \int_C f \ge \sum_{C\in\mathcal{C}} t\mu C = t\mu G_t.$$

(c) All this is true for every $t > 0$. Now if we set

$$f_1^*(x) = \sup_{a>x} \frac{1}{a-x}\int_x^a f$$

for $x \in \mathbb{R}$, we have $\{x : f_1^*(x) > t\} = G_t$ for every $t > 0$.

For any $t > 0$,

$$\frac{1}{p}t\mu G_t = (1 - \frac{1}{q})t\mu G_t \le \int_{G_t} f - \frac{1}{q}t\chi\mathbb{R} \le \int_{-\infty}^{\infty}(f - \frac{1}{q}t\chi\mathbb{R})^+.$$

So

$$\int_{-\infty}^{\infty}(f_1^*)^p = \int_0^{\infty} \mu\{x : f_1^*(x)^p > t\}dt$$

(see 252O)

$$= p\int_0^{\infty} u^{p-1}\mu\{x : f_1^*(x) > u\}du$$

(substituting $t = u^p$)

$$\le p\int_0^\infty u^{p-1}\mu G_u du = p^2\int_0^\infty u^{p-2}\Bigl(\int_{-\infty}^\infty (f-\frac{1}{q}u\chi\mathbb{R})^+\Bigr)du$$
$$= p^2\int_{-\infty}^\infty\int_0^\infty \max(0,f(x)-\frac{1}{q}u)u^{p-2}dudx$$

(by Fubini's theorem, 252B, because $(x,u)\mapsto u^{p-2}\max(0,f(x)-\frac{1}{q}u)$ is measurable and non-negative)

$$= p^2\int_{-\infty}^\infty\int_0^{qf(x)} u^{p-2}(f(x)-\frac{1}{q}u)dudx$$
$$= \tfrac{p^2q^{p-1}}{p(p-1)}\int_{-\infty}^\infty f^p = (\tfrac{p}{p-1})^p\|f\|_p^p.$$

(d) Similarly, setting $f_2^*(x)=\sup_{a<x}\frac{1}{x-a}\int_a^x f$ for $x\in\mathbb{R}$, $\int_{-\infty}^\infty (f_2^*)^p\le(\frac{p}{p-1})^p\|f\|_p^p$. But $f^*=\max(f_1^*,f_2^*)$. **P** Of course $f_1^*\le f^*$ and $f_2^*\le f^*$. But also, if $f^*(x)>t$, there must be a non-trivial interval I containing x such that $\int_I f>t\mu I$; if $a=\inf I$ and $b=\sup I$, then either $\int_a^x f>(x-a)t$ and $f_2^*(x)>t$, or $\int_x^b f>(b-x)t$ and $f_1^*(x)>t$. As x and t are arbitrary, $f^*=\max(f_1^*,f_2^*)$. **Q**

Accordingly

$$\|f^*\|_p^p=\int_{-\infty}^\infty (f^*)^p=\int_{-\infty}^\infty \max((f_1^*)^p,(f_2^*)^p)$$
$$\le\int_{-\infty}^\infty (f_1^*)^p+(f_2^*)^p\le 2(\tfrac{p}{p-1})^p\|f\|_p^p.$$

Taking pth roots, we have the inequality we seek.

286B Lemma Let $g:\mathbb{R}\to[0,\infty[$ be a function which is non-decreasing on $]-\infty,\alpha]$, non-increasing on $[\beta,\infty[$ and constant on $[\alpha,\beta]$, where $\alpha\le\beta$. Then for any measurable function $f:\mathbb{R}\to[0,\infty]$, $\int_{-\infty}^\infty f\times g\le\int_{-\infty}^\infty g\cdot\sup_{a\le\alpha,b\ge\beta,a<b}\frac{1}{b-a}\int_a^b f$.

proof Set $\gamma=\sup_{a\le\alpha,b\ge\beta,a<b}\frac{1}{b-a}\int_a^b f$. For n, $k\in\mathbb{N}$ set $E_{nk}=\{x:\alpha-2^n\le x\le\beta+2^n,\ g(x)\ge 2^{-n}(k+1)\}$, so that E_{nk} is either empty or a bounded interval including $[\alpha,\beta]$, and $\int_{E_{nk}}f\le\gamma\mu E_{nk}$. For $n\in\mathbb{N}$, set $g_n=2^{-n}\sum_{k=0}^{4^n-1}\chi E_{nk}$; then $\langle g_n\rangle_{n\in\mathbb{N}}$ is a non-decreasing sequence of functions with supremum g, and

$$\int_{-\infty}^\infty f\times g=\sup_{n\in\mathbb{N}}\int_{-\infty}^\infty f\times g_n=\sup_{n\in\mathbb{N}}2^{-n}\sum_{k=0}^{4^n-1}\int_{E_{nk}}f$$
$$\le\sup_{n\in\mathbb{N}}2^{-n}\sum_{k=0}^{4^n-1}\gamma\mu E_{nk}=\sup_{n\in\mathbb{N}}\gamma\int_{-\infty}^\infty g_n=\gamma\int_{-\infty}^\infty g,$$

as claimed.

286C Shift, modulation and dilation Some of the calculations below will be easier if we use the following formalism. For any function f with domain included in $\mathbb{R}$, and $\alpha\in\mathbb{R}$, we can define

$$(S_\alpha f)(x)=f(x+\alpha),\quad (M_\alpha f)(x)=e^{i\alpha x}f(x),\quad (D_\alpha f)(x)=f(\alpha x)$$

whenever the right-hand sides are defined. In the case of $S_\alpha f$ and $D_\alpha f$ it is sometimes convenient to allow $\pm\infty$ as a value of the function. We have the following elementary facts.

(a) $S_{-\alpha}S_\alpha f=f$, $D_{1/\alpha}D_\alpha f=f$ if $\alpha\ne 0$.

(b) $S_\alpha(f\times g)=S_\alpha f\times S_\alpha g$, $D_\alpha(f\times g)=D_\alpha f\times D_\alpha g$.

(c) $D_\alpha|f|=|D_\alpha f|$.

(d) If f is integrable, then

$$(M_\alpha f)^\wedge = S_{-\alpha}\hat{f}, \quad (S_\alpha f)^\wedge = M_\alpha \hat{f}, \quad (S_\alpha f)^\vee = M_{-\alpha}\check{f};$$

if moreover $\alpha > 0$, then

$$\alpha(D_\alpha f)^\wedge = D_{1/\alpha}\hat{f}, \quad \alpha(D_\alpha f)^\vee = D_{1/\alpha}\check{f}$$

(283Cc-283Ce).

(e) If f belongs to $\mathcal{L}^1_{\mathbb{C}} = \mathcal{L}^1_{\mathbb{C}}(\mu)$, so do $S_\alpha f$, $M_\alpha f$ and (if $\alpha \neq 0$) $D_\alpha f$, and in this case

$$\|S_\alpha f\|_1 = \|M_\alpha f\|_1 = \|f\|_1, \quad \|D_\alpha f\|_1 = \frac{1}{|\alpha|}\|f\|_1.$$

(f) If f belongs to $\mathcal{L}^2_{\mathbb{C}}$ so do $S_\alpha f$, $M_\alpha f$ and (if $\alpha \neq 0$) $D_\alpha f$, and in this case

$$\|S_\alpha f\|_2 = \|M_\alpha f\|_2 = \|f\|_2, \quad \|D_\alpha f\|_2 = \frac{1}{\sqrt{|\alpha|}}\|f\|_2.$$

(g) If f is a rapidly decreasing test function (284A), so are $M_\alpha f$ and $S_\alpha f$ and (if $\alpha \neq 0$) $D_\alpha f$.

286D Lemma Suppose that $f : \mathbb{R} \to [0,\infty]$ is a measurable function such that, for some constant $C \geq 0$, $\int_E f \leq C\sqrt{\mu E}$ whenever $\mu E < \infty$. Then f is finite almost everywhere and $\int_{-\infty}^{\infty} \frac{1}{1+|x|} f(x)dx$ is finite.

proof For any $n \geq 1$, set $E_n = \{x : |x| \leq n,\ f(x) \geq n\}$; then

$$n\mu E_n \leq \int_{E_n} f \leq C\sqrt{\mu E_n},$$

so $\mu E_n \leq \frac{C^2}{n^2}$ and

$$\{x : f(x) = \infty\} = \bigcap_{n\geq 1}\bigcup_{m\geq n} E_m$$

has measure at most $\inf_{n\geq 1}\sum_{m=n}^{\infty}\mu E_m = 0$.

As for the integral, set $F(x) = \int_0^x f$ for $x \geq 0$. Then, for any $a \geq 0$,

$$\int_0^a \frac{f(x)}{1+x}dx = \frac{F(a)}{1+a} + \int_0^a \frac{F(x)}{(1+x)^2}dx$$

(225F)

$$\leq C\Bigl(\frac{\sqrt{a}}{1+a} + \int_0^a \frac{\sqrt{x}}{(1+x)^2}dx\Bigr) \leq C\Bigl(1 + \int_0^\infty \frac{\sqrt{x}}{(1+x)^2}dx\Bigr),$$

so

$$\int_0^\infty \frac{f(x)}{1+x}dx \leq C\Bigl(1 + \int_0^\infty \frac{\sqrt{x}}{(1+x)^2}dx\Bigr)$$

is finite. Similarly, $\int_{-\infty}^0 \frac{f(x)}{1-x}dx$ is finite, so we have the result.

286E The Lacey-Thiele construction (a) Let $\mathcal{I}$ be the family of all **dyadic intervals** of the form $\left[2^k n, 2^k(n+1)\right[$ where k, $n \in \mathbb{Z}$. The essential geometric property of $\mathcal{I}$ is that if I, $J \in \mathcal{I}$ then either $I \subseteq J$ or $J \subseteq I$ or $I \cap J = \emptyset$. Let Q be the set of all pairs $\sigma = (I_\sigma, J_\sigma) \in \mathcal{I}^2$ such that $\mu I_\sigma \cdot \mu J_\sigma = 1$. For $\sigma \in Q$, let $k_\sigma \in \mathbb{Z}$ be such that $\mu I_\sigma = 2^{-k_\sigma}$ and $\mu J_\sigma = 2^{k_\sigma}$; let x_σ be the midpoint of I_σ, y_σ the midpoint of J_σ, $J^l_\sigma \in \mathcal{I}$ the left-hand half-interval of J_σ, $J^r_\sigma \in \mathcal{I}$ the right-hand half-interval of J_σ, and y^l_σ the lower quartile of J_σ, that is, the midpoint of J^l_σ.

(b) There is a rapidly decreasing test function ϕ such that $\hat{\phi}$ is real-valued and $\chi[-\frac{1}{6},\frac{1}{6}] \leq \hat{\phi} \leq \chi[-\frac{1}{5},\frac{1}{5}]$. **P** Look at parts (b)-(d) of the proof of 284G. The process there can be used to provide us with a smooth function ψ_1 which is zero outside the interval $[\frac{1}{6},\frac{1}{5}]$ and strictly positive on $\left]\frac{1}{6},\frac{1}{5}\right[$; multiplying by a suitable factor, we can arrange that $\int_{-\infty}^{\infty}\psi_1 = 1$. So if we set $\psi_2(x) = 1 - \int_{-\infty}^x \psi_1$ for $x \in \mathbb{R}$, ψ_2 will be smooth, and $\chi\left]-\infty,\frac{1}{6}\right] \leq \psi_2 \leq \chi\left]-\infty,\frac{1}{5}\right]$. Now set $\psi_0(x) = \psi_2(x)\psi_2(-x)$ for $x \in \mathbb{R}$, and $\phi = \check{\psi}_0$; $\hat{\phi} = \psi_0$ (284C) will have the required property. **Q**

For $\sigma \in Q$, set $\phi_\sigma = 2^{k_\sigma/2} M_{y^l_\sigma} S_{-x_\sigma} D_{2^{k_\sigma}} \phi$, so that

$$\phi_\sigma(x) = 2^{k_\sigma/2} e^{iy^l_\sigma x} \phi(2^{k_\sigma}(x - x_\sigma)).$$

Observe that ϕ_σ is a rapidly decreasing test function. Now $\hat{\phi}_\sigma = 2^{-k_\sigma/2} S_{-y^l_\sigma} M_{-x_\sigma} D_{2^{-k_\sigma}} \hat{\phi}$, that is,

$$\hat{\phi}_\sigma(y) = 2^{-k_\sigma/2} e^{-ix_\sigma(y - y^l_\sigma)} \hat{\phi}(2^{-k_\sigma}(y - y^l_\sigma)),$$

which is zero unless $|y - y^l_\sigma| \le \frac{1}{5} 2^{k_\sigma}$; since the length of J^l_σ is $\frac{1}{2} 2^{k_\sigma}$, this can be true only when $y \in J^l_\sigma$. We have the following simple facts.

(i) $\|\phi_\sigma\|_2 = 2^{k_\sigma/2} \cdot 2^{-k_\sigma/2} \|\phi\|_2 = \|\phi\|_2$ for every $\sigma \in Q$.

(ii) $\|\hat{\phi}_\sigma\|_1 = 2^{-k_\sigma/2} \cdot 2^{k_\sigma} \|\hat{\phi}\|_1 = 2^{k_\sigma/2} \|\hat{\phi}\|_1$ for every $\sigma \in Q$.

(iii) If σ, $\tau \in Q$ and $J_\sigma \ne J_\tau$ and $J^r_\sigma \cap J^r_\tau$ is non-empty, then $J^l_\sigma \cap J^l_\tau = \emptyset$ so

$$(\phi_\sigma | \phi_\tau) = (\hat{\phi}_\sigma | \hat{\phi}_\tau) = 0,$$

by 284Ob. (For f, $g \in \mathcal{L}^2_{\mathbb{C}}$, I write $(f|g)$ for $\int f \times \bar{g}$.)

(c) Set $w(x) = \frac{1}{(1+|x|)^3}$ for $x \in \mathbb{R}$. Then there is a $C_1 > 0$ such that $|\phi(x)| \le C_1 \min(w(3), w(x)^2)$ for every $x \in \mathbb{R}$ (because $\lim_{x\to\infty} x^6 \phi(x) = \lim_{x\to-\infty} x^6 \phi(x) = 0$). For $\sigma \in Q$, set $w_\sigma = 2^{k_\sigma} S_{-x_\sigma} D_{2^{k_\sigma}} w$, so that $w_\sigma(x) = 2^{k_\sigma} w(2^{k_\sigma}(x - x_\sigma))$ for every x. Elementary calculations show that

(i) w_σ depends only on I_σ;

(ii) $\int_{-\infty}^{\infty} w_\sigma = \int_{-\infty}^{\infty} w = 1$ for every σ;

(iii)

$$|\phi_\sigma(x)| \le C_1 \min(2^{-k_\sigma/2} w_\sigma(x), 2^{-3k_\sigma/2} w_\sigma(x)^2)$$

for every x and σ (because $|\phi(x)| \le C_1 w(x)^2 \le C_1 w(x)$ for every $x \in \mathbb{R}$).

286F Two partial orders (a) For σ, $\tau \in Q$ say that $\sigma \le \tau$ if $I_\sigma \subseteq I_\tau$ and $J_\tau \subseteq J_\sigma$. Then $\le$ is a partial order on Q. We have the following elementary facts.

(i) If $\sigma \le \tau$, then $k_\sigma \ge k_\tau$.

(ii) If σ and τ are incomparable (that is, $\sigma \not\le \tau$ and $\tau \not\le \sigma$), then $(I_\sigma \times J_\sigma) \cap (I_\tau \times J_\tau)$ is empty. **P** We may suppose that $k_\sigma \le k_\tau$. If $J_\sigma \cap J_\tau \ne \emptyset$, then $J_\sigma \subseteq J_\tau$, because both are dyadic intervals, and J_σ is the shorter; but as $\tau \not\le \sigma$, this means that $I_\tau \not\subseteq I_\sigma$ and $I_\tau \cap I_\sigma = \emptyset$. **Q**

(iii) If σ, σ' are incomparable and both less than or equal to τ, then $I_\sigma \cap I_{\sigma'} = \emptyset$, because $J_\tau \subseteq J_\sigma \cap J_{\sigma'}$.

(iv) If $\sigma \le \tau$ and $k_\sigma \ge k \ge k_\tau$, then there is a (unique) σ' such that $\sigma \le \sigma' \le \tau$ and $k_{\sigma'} = k$. (The point is that there is a unique $I \in \mathcal{I}$ such that $I_\sigma \subseteq I \subseteq I_\tau$ and $\mu I = 2^{-k}$; and similarly there is just one candidate for $J_{\sigma'}$.)

(b) For σ, $\tau \in Q$ say that $\sigma \le_r \tau$ if $I_\sigma \subseteq I_\tau$ and $J^r_\tau \subseteq J^r_\sigma$ (that is, either $\tau = \sigma$ or $J_\tau \subseteq J^r_\sigma$), so that, in particular, $\sigma \le \tau$. Note that if σ, $\sigma' \le_r \tau$ and $k_\sigma \ne k_{\sigma'}$ then $J^r_\sigma \cap J^r_{\sigma'} \ne \emptyset$, so $(\phi_\sigma | \phi_{\sigma'}) = 0$ (286E(b-iii)).

(c) It will be convenient to have a shorthand for the following: if P, $R \subseteq Q$, say that $P \preccurlyeq R$ if for every $\sigma \in P$ there is a $\tau \in R$ such that $\sigma \le \tau$.

286G We shall need the results of some elementary calculations. The first three are nearly trivial.

Lemma (a) For any $m \in \mathbb{N}$, $\sum_{n=m}^{\infty} w(n + \frac{1}{2}) \le \frac{1}{2(1+m)^2}$.

(b) Suppose that $\sigma \in P$ and that I is an interval not containing x_σ in its interior. Then $\int_I w_\sigma \ge w_\sigma(x) \mu I$, where x is the midpoint of I.

(c) For any $x \in \mathbb{R}$, $\sum_{n=-\infty}^{\infty} w(x - n) \le 2$.

(d) There is a constant $C_2 \ge 0$ such that $\int_{-\infty}^{\infty} w(x) w(\alpha x + \beta) dx \le C_2 w(\beta)$ whenever $0 \le \alpha \le 1$ and $\beta \in \mathbb{R}$.

(e) There is a constant $C_3 \ge 0$ such that $|(\phi_\sigma | \phi_\tau)| \le 2^{-k_\sigma/2} 2^{k_\tau/2} C_3 \int_{I_\tau} w_\sigma$ whenever σ, $\tau \in Q$ and $k_\sigma \le k_\tau$.

(f) There is a constant $C_4 \ge 0$ such that whenever $\tau \in Q$ and $k \in \mathbb{Z}$, then

$$\sum_{\sigma \in Q, \sigma \le \tau, k_\sigma = k} \int_{\mathbb{R} \setminus I_\tau} w_\sigma \le C_4.$$

proof (a) The point is just that w is convex on $]-\infty, 0]$ and $[0, \infty[$. So we can apply 233Ib with $f(x) = x$, or argue directly from the fact that $w(n + \frac{1}{2}) \le \frac{1}{2}(w(n + \frac{1}{2} + x) + w(n + \frac{1}{2} - x))$ for $|x| \le \frac{1}{2}$, to see that $w(n + \frac{1}{2}) \le \int_n^{n+1} w$ for every $n \ge 0$. Accordingly

$$\sum_{n=m}^{\infty} w(n+\tfrac{1}{2}) \le \int_m^\infty w = \tfrac{1}{2(1+m)^2}.$$

(b) Similarly, because I lies all on the same side of x_σ, w_σ is convex on I, so the same inequality yields $w_\sigma(x)\mu I \le \int_I w_\sigma$.

(c) Let m be such that $|x-m| \le \frac{1}{2}$. Then, using the same inequalities as before to estimate $w(x-n)$ for $n \ne m$, we have

$$\begin{aligned}\sum_{n=-\infty}^{\infty} w(x-n) &\le w(x-m) + \int_{-\infty}^{x-m-\frac{1}{2}} w + \int_{x-m+\frac{1}{2}}^{\infty} w \\ &\le 1 + \int_{-\infty}^{\infty} w = 2.\end{aligned}$$

(d)(i) The first step is to note that

$$\frac{w(\frac{1}{2}(1+\beta))}{w(\beta)} = \frac{8(1+\beta)^3}{(3+\beta)^3} \le 8$$

for every $\beta \ge 0$. Now $\alpha w(\alpha+\alpha\beta) \le 4w(\beta)$ whenever $\beta \ge 0$ and $\alpha \ge \frac{1}{2}$. **P** For $t \ge \frac{1}{2}$,

$$\frac{d}{dt} t w(t+t\beta) = \tfrac{1-2t(1+\beta)}{(1+t+t\beta)^4} \le 0,$$

so

$$\alpha w(\alpha+\alpha\beta) \le \tfrac{1}{2} w(\tfrac{1}{2}+\tfrac{1}{2}\beta) \le 4w(\beta). \text{ } \mathbf{Q}$$

Of course this means that

$$\frac{1}{\alpha} w(\tfrac{1+\beta}{2\alpha}) \le 8w(\beta)$$

whenever $\beta \ge 0$ and $0 < \alpha \le 1$.

(ii) Try $C_2 = 16$. If $0 < \alpha \le 1$ and $\beta \ge 0$, set $\gamma = \frac{1+\beta}{2\alpha}$. Then, for any $x \ge -\gamma$,

$$1 + \alpha x + \beta = (1+\beta)(1 + \tfrac{\alpha x}{1+\beta}) \ge \frac{1}{2}(1+\beta),$$

so $w(\alpha x+\beta) \le 8w(\beta)$ and

$$\int_{-\gamma}^{\infty} w(x)w(\alpha x+\beta)dx \le 8w(\beta)\int_{-\gamma}^{\infty} w \le 8w(\beta).$$

On the other hand,

$$\begin{aligned}\int_{-\infty}^{-\gamma} w(x)w(\alpha x+\beta)dx &\le w(\gamma)\int_{-\infty}^{\infty} w(\alpha x+\beta)dx \\ &= \frac{1}{\alpha} w(\tfrac{1+\beta}{2\alpha}) \int_{-\infty}^{\infty} w \le 8w(\beta).\end{aligned}$$

Putting these together, $\int_{-\infty}^{\infty} w(x)w(\alpha x+\beta)dx \le 16w(\beta)$; and this is true whenever $0 < \alpha \le 1$ and $\beta \ge 0$.

(iii) If $\alpha = 0$, then

$$\int_{-\infty}^{\infty} w(x)w(\alpha x+\beta)dx = w(\beta)\int_{-\infty}^{\infty} w = w(\beta) \le C_2 w(\beta)$$

for any β. If $0 < \alpha \le 1$ and $\beta < 0$, then

$$\int_{-\infty}^{\infty} w(x)w(\alpha x+\beta)dx = \int_{-\infty}^{\infty} w(-x)w(-\alpha x-\beta)dx$$

(because w is an even function)

$$= \int_{-\infty}^{\infty} w(x)w(\alpha x-\beta)dx \le C_2 w(-\beta)$$

(by (ii) above)

$$= C_2 w(\beta).$$

So we have the required inequality in all cases.

(e) Set $C_3 = \max(C_1^2 C_2, \|\phi\|_2^2 / \int_{-1/2}^{1/2} w)$.

(i) It is worth disposing immediately of the case $\sigma = \tau$. In this case,

$$|(\phi_\sigma|\phi_\tau)| = \|\phi_\sigma\|_2^2 = \|\phi\|_2^2,$$

while

$$\int_{I_\tau} w_\sigma = 2^{k_\sigma} \int_{x_\sigma - 2^{-k_\sigma - 1}}^{x_\sigma + 2^{-k_\sigma - 1}} w(2^{k_\sigma}(x - x_\sigma))dx = \int_{-1/2}^{1/2} w,$$

so certainly $|(\phi_\sigma|\phi_\tau)| \le C_3 \int_{I_\tau} w_\sigma$.

(ii) Now suppose that $I_\sigma \ne I_\tau$. In this case, because $k_\sigma \le k_\tau$, I_τ must all lie on the same side of x_σ, so $\int_{I_\tau} w_\sigma \ge w_\sigma(x_\tau)\mu I_\tau$, by (b).

We know from 286E(c-iii) that $|\phi_\sigma(x)| \le 2^{-k_\sigma/2} C_1 w_\sigma(x)$ for every x. So

$$\begin{aligned}
|(\phi_\sigma|\phi_\tau)| &\le 2^{-k_\sigma/2} 2^{-k_\tau/2} C_1^2 \int_{-\infty}^{\infty} w_\sigma \times w_\tau \\
&= 2^{k_\sigma/2} 2^{k_\tau/2} C_1^2 \int_{-\infty}^{\infty} w(2^{k_\sigma}(x - x_\sigma)) w(2^{k_\tau}(x - x_\tau)) dx \\
&= 2^{k_\sigma/2} 2^{-k_\tau/2} C_1^2 \int_{-\infty}^{\infty} w(2^{k_\sigma - k_\tau} x + 2^{k_\sigma}(x_\tau - x_\sigma)) w(x) dx \\
&\le 2^{k_\sigma/2} 2^{-k_\tau/2} C_1^2 C_2 w(2^{k_\sigma}(x_\tau - x_\sigma))
\end{aligned}$$

(by (d), since $2^{k_\sigma - k_\tau} \le 1$)

$$\le 2^{-k_\sigma/2} 2^{-k_\tau/2} C_3 w_\sigma(x_\tau) \le 2^{-k_\sigma/2} 2^{k_\tau/2} C_3 \int_{I_\tau} w_\sigma,$$

as required.

(f) Set $C_4 = 2\sum_{j=0}^{\infty} \int_{j+\frac{1}{2}}^{\infty} w$; this is finite because $\int_\alpha^\infty w = \frac{1}{2(1+\alpha)^2}$ for every $\alpha \ge 0$.

If $k < k_\tau$ then $k_\sigma \ne k$ for any $\sigma \le \tau$, so the result is trivial. If $k \ge k_\tau$, then for each dyadic subinterval I of I_τ of length 2^{-k} there is exactly one $\sigma \le \tau$ such that $I_\sigma = I$. List these as $\sigma_0, \dots$ in ascending order of the centres x_{σ_j}, so that if $I_\tau = \left[2^{-k_\tau} m, 2^{-k_\tau}(m+1)\right[$ then $x_{\sigma_j} = 2^{-k_\tau} m + 2^{-k}(j + \frac{1}{2})$, for $j < 2^{k - k_\tau}$. Now

$$\begin{aligned}
\sum_{j=0}^{2^{k-k_\tau}-1} \int_{-\infty}^{2^{-k_\tau} m} w_{\sigma_j} &= 2^k \sum_{j=0}^{2^{k-k_\tau}-1} \int_{-\infty}^{2^{-k_\tau} m} w(2^k(x - 2^{-k_\tau} m) - j - \tfrac{1}{2}) dx \\
&= \sum_{j=0}^{2^{k-k_\tau}-1} \int_{-\infty}^{0} w(x - j - \tfrac{1}{2}) dx \\
&\le \sum_{j=0}^{\infty} \int_{j+\frac{1}{2}}^{\infty} w = \tfrac{1}{2} C_4.
\end{aligned}$$

Similarly (since w is an even function, so the whole picture is symmetric about x_τ)

$$\sum_{j=0}^{2^{k-k_\tau}-1} \int_{2^{-k_\tau}(m+1)}^{\infty} w_{\sigma_j} \le \tfrac{1}{2} C_4,$$

and

$$\sum_{\sigma \le \tau, k_\sigma = k} \int_{\mathbb{R} \setminus I_\tau} w_\sigma \le C_4,$$

as required.

286H 'Mass' and 'energy' (Lacey & Thiele 00) If P is a subset of Q, $E \subseteq \mathbb{R}$ is measurable, $h : \mathbb{R} \to \mathbb{R}$ is measurable, and $f \in \mathcal{L}^2_{\mathbb{C}}$, set

$$\mathrm{mass}_{Eh}(P) = \sup_{\sigma\in P,\tau\in Q,\sigma\le\tau} \int_{E\cap h^{-1}[J_\tau]} w_\tau \le \sup_{\tau\in Q} \int_{-\infty}^{\infty} w_\tau = 1,$$

$$\mathrm{energy}_f(P) = \sup_{\tau\in Q} 2^{k_\tau/2}\sqrt{\sum_{\sigma\in P,\sigma\le_r\tau} |(f|\phi_\sigma)|^2}.$$

If $P' \subseteq P$ then $\mathrm{mass}_{Eh}(P') \le \mathrm{mass}_{Eh}(P)$ and $\mathrm{energy}_f(P') \le \mathrm{energy}_f(P)$. Note that $\mathrm{energy}_f(\{\sigma\}) = 2^{k_\sigma/2}|(f|\phi_\sigma)|$ for any $\sigma \in Q$, since if $\sigma \le_r \tau$ then $k_\tau \le k_\sigma$.

286I Lemma Set $C_5 = 2^{12}$. If $P \subseteq Q$ is finite, $E \subseteq \mathbb{R}$ is measurable, $h : \mathbb{R} \to \mathbb{R}$ is measurable, and $\gamma \ge \mathrm{mass}_{Eh}(P)$, then we can find sets $P_1 \subseteq P$, $P_2 \subseteq Q$ such that $\mathrm{mass}_{Eh}(P_1) \le \frac{1}{4}\gamma$, $\gamma\sum_{\tau\in P_2} \mu I_\tau \le C_5 \mu E$ and $P \setminus P_1 \preccurlyeq P_2$ (in the notation of 286Fc).

proof (a) Set $P_1 = \{\sigma : \sigma \in P,\ \mathrm{mass}_{Eh}(\{\sigma\}) \le \frac{1}{4}\gamma\}$. Then $\mathrm{mass}_{Eh}(P_1) \le \frac{1}{4}\gamma$. If $\gamma = 0$ we can stop here, as $P_1 = P$. Otherwise, for each $\sigma \in P \setminus P_1$ let $\sigma' \in Q$ be such that $\sigma \le \sigma'$ and $\int_{E\cap h^{-1}[J_{\sigma'}]} w_{\sigma'} > \frac{1}{4}\gamma$. Let P_2 be the set of elements of $\{\sigma' : \sigma \in P \setminus P_1\}$ which are maximal for $\le$; then $P \setminus P_1 \preccurlyeq P_2$.

(b) For $k \in \mathbb{N}$ set

$$R_k = \{\tau : \tau \in P_2,\ 2^{k_\tau}\mu(E \cap h^{-1}[J_\tau] \cap I_\tau^{(k)}) \ge 2^{2k-9}\gamma\},$$

where $I_\tau^{(k)}$ is the interval with the same centre as I_τ and 2^k times its length. Now $P_2 = \bigcup_{k\in\mathbb{N}} R_k$. **P** Take $\tau \in P_2$. If $k \in \mathbb{N}$ and $x \in \mathbb{R} \setminus I_\tau^{(k)}$, then $|x - x_\tau| \ge 2^{k-k_\tau-1}$, so

$$w_\tau(x) = 2^{k_\tau} w(2^{k_\tau}(x - x_\tau)) \le 2^{k_\tau} w(2^{k-1}) = 2^{k_\tau}(1 + 2^{k-1})^{-3}.$$

So

$$\begin{aligned}\frac{1}{4}\gamma < \int_{E\cap h^{-1}[J_\tau]} w_\tau &= \int_{E\cap h^{-1}[J_\tau]\cap I_\tau} w_\tau + \sum_{k=0}^{\infty} \int_{E\cap h^{-1}[J_\tau]\cap I_\tau^{(k+1)}\setminus I_\tau^{(k)}} w_\tau \\ &\le 2^{k_\tau}\mu(E \cap h^{-1}[J_\tau] \cap I_\tau) + \sum_{k=0}^{\infty} 2^{k_\tau}\mu(E \cap h^{-1}[J_\tau] \cap I_\tau^{(k+1)})(1 + 2^{k-1})^{-3}.\end{aligned}$$

It follows that either

$$2^{k_\tau}\mu(E \cap h^{-1}[J_\tau] \cap I_\tau) \ge \frac{1}{8}\gamma$$

and $\tau \in R_0$, or there is some $k \in \mathbb{N}$ such that

$$2^{k_\tau}\mu(E \cap h^{-1}[J_\tau] \cap I_\tau^{(k+1)})(1 + 2^{k-1})^{-3} \ge 2^{-k-4}\gamma$$

and

$$2^{k_\tau}\mu(E \cap h^{-1}[J_\tau] \cap I_\tau^{(k+1)}) \ge (1 + 2^{k-1})^3 2^{-k-4}\gamma \ge 2^{2k-7}\gamma,$$

so that $\tau \in R_{k+1}$. **Q**

(c) For every $k \in \mathbb{N}$, $\sum_{\tau\in R_k} \mu I_\tau \le 2^{11-k}\mu E$. **P** If $R_k = \emptyset$, this is trivial. Otherwise, enumerate R_k as $\langle\tau_j\rangle_{j\le n}$ in such a way that $k_{\tau_j} \le k_{\tau_l}$ if $j \le l \le n$. Define $q : \{0, \dots, n\} \to \{0, \dots, n\}$ inductively by the rule

$$q(l) = \min(\{l\} \cup \{q(j) : j < l,\ (I^{(k)}_{\tau_{q(j)}} \times J_{\tau_{q(j)}}) \cap (I^{(k)}_{\tau_l} \times J_{\tau_l}) \ne \emptyset\})$$

for each $l \le n$. A simple induction shows that $q(q(l)) = q(l) \le l$ for every $l \le n$. Note that, for $l \le n$, $I^{(k)}_{\tau_{q(l)}} \cap I^{(k)}_{\tau_l} \ne \emptyset$, so that

$$I_{\tau_l} \subseteq I^{(k)}_{\tau_l} \subseteq I^{(k+2)}_{\tau_{q(l)}},$$

because $\mu I^{(k)}_{\tau_l} \le \mu I^{(k)}_{\tau_{q(l)}}$. Moreover, if $j < l \le n$ and $q(j) = q(l)$, then both J_{τ_j} and J_{τ_l} meet $J_{\tau_{q(j)}}$, therefore include it, and $J_{\tau_j} \subseteq J_{\tau_l}$. But as τ_j and τ_l are distinct members of P_2, $\tau_l \not\le \tau_j$ and $I_{\tau_j} \cap I_{\tau_l}$ must be empty.

Set $M = \{q(j) : j \le n\}$. We have

$$\begin{aligned}\gamma\sum_{\tau\in R_k}\mu I_\tau&=\gamma\sum_{m\in M}\sum_{\substack{j\le n\\ q(j)=m}}\mu I_{\tau_j}\\&\le\gamma\sum_{m\in M}\mu I^{(k+2)}_{\tau_m}=2^{k+2}\gamma\sum_{m\in M}\mu I_{\tau_m}\\&\le 2^{k+2}\sum_{m\in M}2^{9-2k}\mu(E\cap h^{-1}[J_{\tau_m}]\cap I^{(k)}_{\tau_m})\\&\le 2^{k+2}\cdot 2^{9-2k}\mu E=2^{11-k}\mu E\end{aligned}$$

because if l, $m\in M$ and $l<m$ then $I^{(k)}_{\tau_l}\times J_{\tau_l}$ and $I^{(k)}_{\tau_m}\times J_{\tau_m}$ are disjoint (since otherwise $q(m)\le l$ and there can be no j such that $q(j)=m$), so that $h^{-1}[J_{\tau_l}]\cap I^{(k)}_{\tau_l}$ and $h^{-1}[J_{\tau_m}]\cap I^{(k)}_{\tau_m}$ are disjoint. **Q**

(d) Accordingly

$$\gamma\textstyle\sum_{\tau\in P_2}\mu I_\tau\le\gamma\sum_{k=0}^\infty\sum_{\tau\in R_k}\mu I_\tau\le 2^{12}\mu E,$$

as required.

286J Lemma If $P\subseteq Q$ is finite and $f\in\mathcal{L}^2_{\mathbb{C}}$, then

$$\begin{aligned}\sum_{\sigma,\tau\in P,J_\sigma=J_\tau}\bigl|(f|\phi_\sigma)(\phi_\sigma|\phi_\tau)(\phi_\tau|f)\bigr|&\le C_3\sum_{\sigma\in P}|(f|\phi_\sigma)|^2\\&\le C_3\|\sum_{\sigma\in P}(f|\phi_\sigma)\phi_\sigma\|_2\|f\|_2.\end{aligned}$$

proof

$$\sum_{\sigma,\tau\in P,J_\sigma=J_\tau}\bigl|(f|\phi_\sigma)(\phi_\sigma|\phi_\tau)(\phi_\tau|f)\bigr|\le\sum_{\sigma,\tau\in P,J_\sigma=J_\tau}\frac12\bigl(|(f|\phi_\sigma)|^2+|(f|\phi_\tau)|^2\bigr)|(\phi_\sigma|\phi_\tau)|$$

(because $|\xi\zeta|\le\frac12(|\xi|^2+|\zeta|^2)$ for all complex numbers ξ, ζ)

$$\begin{aligned}&=\sum_{\sigma\in P}\sum_{\substack{\tau\in P\\ J_\sigma=J_\tau}}|(f|\phi_\sigma)|^2|(\phi_\sigma|\phi_\tau)|\\&\le\sum_{\sigma\in P}|(f|\phi_\sigma)|^2\sum_{\tau\in P,J_\sigma=J_\tau}C_3\int_{I_\tau}w_\sigma\end{aligned}$$

(by 286Ge, since $k_\sigma=k_\tau$ if $J_\sigma=J_\tau$)

$$\le\sum_{\sigma\in P}|(f|\phi_\sigma)|^2C_3\int_{-\infty}^\infty w_\sigma$$

(because if τ, τ' are distinct members of P and $J_\tau=J_{\tau'}$, then I_τ and $I_{\tau'}$ are disjoint)

$$\begin{aligned}&=C_3\sum_{\sigma\in P}|(f|\phi_\sigma)|^2=C_3\sum_{\sigma\in P}(f|\phi_\sigma)(\phi_\sigma|f)\\&=C_3\bigl(\sum_{\sigma\in P}(f|\phi_\sigma)\phi_\sigma\big|f\bigr)\le C_3\|\sum_{\sigma\in P}(f|\phi_\sigma)\phi_\sigma\|_2\|f\|_2\end{aligned}$$

by Cauchy's inequality (244Eb).

286K Lemma Set

$$C_6=4(C_3+4C_3\sqrt{2C_4}).$$

Let $P\subseteq Q$ be a finite set, $f\in\mathcal{L}^2_{\mathbb{C}}$ and $\|f\|_2=1$. Suppose that $\gamma\ge\text{energy}_f(P)$. Then we can find finite sets $P_1\subseteq P$ and $P_2\subseteq Q$ such that $\text{energy}_f(P_1)\le\frac12\gamma$, $\gamma^2\sum_{\tau\in P_2}\mu I_\tau\le C_6$, and $P\setminus P_1\preccurlyeq P_2$.

proof (a) We may suppose that $\gamma>0$ and that $P\ne\emptyset$, since otherwise we can take $P_1=P$ and $P_2=\emptyset$.

(i) For $\tau\in Q$, $A\subseteq Q$ set

$$T_\tau=\{\sigma:\sigma\in P,\,\sigma\le_r\tau\},\quad \Delta(A)=\sum\nolimits_{\sigma\in A}|(f|\phi_\sigma)|^2.$$

There are only finitely many sets of the form T_τ; let $R\subseteq Q$ be a non-empty finite set such that whenever $\tau\in Q$ and T_τ is not empty, there is a $\tau'\in R$ such that $T_\tau=T_{\tau'}$ and $k_{\tau'}\ge k_\tau$; this is possible because if $A\subseteq P$ is not empty then $k_\tau\le\min_{\sigma\in A}k_\sigma$ whenever $A=T_\tau$.

(ii) Choose $\tau_0,\tau_1,\dots$, $P'_0,P'_1,\dots$ inductively, as follows. $P'_0=P$. Given that $P'_j\subseteq P$ is not empty, consider

$$R_j=\{\tau:\tau\in R,\,2^{k_\tau}\Delta(P'_j\cap T_\tau)\ge\frac14\gamma^2\}.$$

If $R_j=\emptyset$, stop the induction and set $n=j$, $P_2=\{\tau_l:l<j\}$, $P_1=P'_j$. Otherwise, among the members of R_j take one with y_τ as far to the left as possible, and call it τ_j; set $P'_{j+1}=P'_j\setminus\{\sigma:\sigma\in P,\,\sigma\le\tau_j\}$, and continue. Note that as $R_{j+1}\subseteq R_j$ for every j, $y_{\tau_{j+1}}\ge y_{\tau_j}$ for every j.

The induction must stop at a finite stage because if it does not stop with $n=j$ then $\Delta(P'_j\cap T_{\tau_j})>0$, so $P'_j\cap T_{\tau_j}$ is not empty and $P'_{j+1}\subseteq P'_j\setminus T_{\tau_j}$ is a proper subset of P'_j, while $P'_0=P$ is finite. Since $R_n=\emptyset$,

$$\begin{aligned}\operatorname*{energy}_f(P_1)=\operatorname*{energy}_f(P'_n)&=\sup_{\tau\in Q}2^{k_\tau/2}\sqrt{\Delta(P'_n\cap T_\tau)}\\&=\max_{\tau\in R}2^{k_\tau/2}\sqrt{\Delta(P'_n\cap T_\tau)}\le\frac12\gamma.\end{aligned}$$

We also have $P\setminus P_1\preccurlyeq\{\tau_j:j<n\}$.

(iii) Set $P''_j=P'_j\cap T_{\tau_j}\subseteq P'_j\setminus P'_{j+1}$ for $j<n$, so that $\langle P''_j\rangle_{j<n}$ is disjoint, and $P'=\bigcup_{j<n}P''_j\subseteq P$. Then if $\sigma\in P'$, $j<n$ and $J_{\tau_j}\subseteq J^l_\sigma$, $I_\sigma\cap I_{\tau_j}=\emptyset$. **P?** Otherwise, take $l<n$ such that $\sigma\in P''_l$. Then $J_{\tau_j}\subseteq J_\sigma$, so $k_{\tau_j}\le k_\sigma$ and I_σ must be included in I_{τ_j}; thus $\sigma\le\tau_j$ and $\sigma\notin P'_{j+1}$. On the other hand,

$$y_{\tau_j}\in J_{\tau_j}\subseteq J^l_\sigma,\quad y_{\tau_l}\in J^r_{\tau_l}\subseteq J^r_\sigma,$$

so $y_{\tau_j}<y_{\tau_l}$ and $j<l$. But this means that $\sigma\notin P'_l$, while we chose l such that $\sigma\in P''_l$. **XQ**

It follows that if σ, $\tau\in P'$ are distinct and $J^l_\sigma\cap J^l_\tau$ is not empty, then $I_\sigma\cap I_\tau=\emptyset$. **P** If $J_\sigma=J_\tau$ this is true just because $\sigma\ne\tau$. Otherwise, since J_σ and J_τ intersect, one is included in the other; suppose that $J_\sigma\subset J_\tau$. Since J_σ meets J^l_τ, $J_\sigma\subseteq J^l_\tau$. Now let $j<n$ be such that $\sigma\in P''_j$; then $\sigma\le\tau_j$, so $J_{\tau_j}\subseteq J_\sigma\subseteq J^l_\tau$, and $I_\sigma\cap I_\tau\subseteq I_{\tau_j}\cap I_\tau=\emptyset$ by the last remark. **Q**

(b) Now let us estimate

$$\begin{aligned}\gamma^2\sum_{j<n}\mu I_{\tau_j}&=\sum_{j<n}2^{-k_{\tau_j}}\gamma^2\le4\sum_{j<n}\Delta(P''_j)\\&=4\sum_{j<n}\sum_{\sigma\in P''_j}|(f|\phi_\sigma)|^2=4\sum_{\sigma\in P'}|(f|\phi_\sigma)|^2=4\alpha\end{aligned}$$

say.

Because $\|f\|_2=1$, we have $\alpha\le\|\sum_{\sigma\in P'}(f|\phi_\sigma)\phi_\sigma\|_2$ (see 286J). So

$$\begin{aligned}\alpha^2\le\|\sum_{\sigma\in P'}(f|\phi_\sigma)\phi_\sigma\|_2^2&=\sum_{\sigma,\tau\in P'}(f|\phi_\sigma)(\phi_\sigma|\phi_\tau)(\phi_\tau|f)\\&=\sum_{\sigma,\tau\in P',J_\sigma=J_\tau}(f|\phi_\sigma)(\phi_\sigma|\phi_\tau)(\phi_\tau|f)+2\sum_{\sigma,\tau\in P',J_\sigma\subseteq J^l_\tau}(f|\phi_\sigma)(\phi_\sigma|\phi_\tau)(\phi_\tau|f)\end{aligned}$$

because $(\phi_\sigma|\phi_\tau)=0$ unless $J^l_\sigma\cap J^l_\tau\ne\emptyset$, as noted in 286Eb. Take these two terms separately. For the first, we have

$$\sum\nolimits_{\sigma,\tau\in P',J_\sigma=J_\tau}\bigl|(f|\phi_\sigma)(\phi_\sigma|\phi_\tau)(\phi_\tau|f)\bigr|\le C_3\alpha$$

by 286J. For the second term, we have

$$\sum_{\sigma,\tau\in P',J_\sigma\subseteq J^l_\tau}\bigl|(f|\phi_\sigma)(\phi_\sigma|\phi_\tau)(\phi_\tau|f)\bigr| \le \sum_{\sigma\in P'}|(f|\phi_\sigma)|\sum_{\tau\in P',J_\sigma\subseteq J^l_\tau}|(\phi_\sigma|\phi_\tau)(\phi_\tau|f)|$$
$$\le \sqrt{\sum_{\sigma\in P'}|(f|\phi_\sigma)|^2}\sqrt{\sum_{\sigma\in P'}\bigl(\sum_{\tau\in P',J_\sigma\subseteq J^l_\tau}|(\phi_\sigma|\phi_\tau)(\phi_\tau|f)|\bigr)^2}$$
$$= \sqrt{\alpha}\sqrt{\sum_{j<n}H_j},$$

where for $j<n$ I set

$$H_j = \sum_{\sigma\in P''_j}\bigl(\sum_{\tau\in P',J_\sigma\subseteq J^l_\tau}|(\phi_\sigma|\phi_\tau)(\phi_\tau|f)|\bigr)^2.$$

Now we can estimate H_j by observing that, for any $\tau\in P'$,

$$|(\phi_\tau|f)| = 2^{-k_\tau/2}\,\mathrm{energy}_f(\{\tau\}) \le 2^{-k_\tau/2}\gamma,$$

while if σ, $\tau\in P'$ and $J^l_\tau\supseteq J_\sigma$ then

$$|(\phi_\sigma|\phi_\tau)| \le 2^{-k_\sigma/2}2^{k_\tau/2}C_3\textstyle\int_{I_\tau}w_\sigma$$

by 286Ge. We also need to know that if $\sigma\in P''_j$ and τ, τ' are distinct elements of P' such that $J_\sigma\subseteq J^l_\tau\cap J^l_{\tau'}$, then I_τ, $I_{\tau'}$ and I_{τ_j} are all disjoint, by (a-iii) above, because $J_{\tau_j}\subseteq J_\sigma$. So we have

$$H_j \le \sum_{\sigma\in P''_j}\bigl(\sum_{\tau\in P',J_\sigma\subseteq J^l_\tau}2^{-k_\tau/2}\gamma 2^{-k_\sigma/2}2^{k_\tau/2}C_3\int_{I_\tau}w_\sigma\bigr)^2$$
$$= C_3^2\gamma^2\sum_{\sigma\in P''_j}2^{-k_\sigma}\bigl(\sum_{\tau\in P',J_\sigma\subseteq J^l_\tau}\int_{I_\tau}w_\sigma\bigr)^2$$
$$\le C_3^2\gamma^2\sum_{\sigma\in P''_j}2^{-k_\sigma}\bigl(\int_{\mathbb{R}\setminus I_{\tau_j}}w_\sigma\bigr)^2$$
$$\le C_3^2\gamma^2\sum_{k=k_{\tau_j}}^\infty 2^{-k}\sum_{\sigma\in P''_j,k_\sigma=k}\int_{\mathbb{R}\setminus I_{\tau_j}}w_\sigma\cdot\int_{-\infty}^\infty w_\sigma$$
$$\le C_3^2\gamma^2\sum_{k=k_{\tau_j}}^\infty 2^{-k}C_4$$

(by 286Gf, since $\sigma\le\tau_j$ for every $\sigma\in P''_j$)

$$= C_3^2\gamma^2 2^{-k_{\tau_j}+1}C_4.$$

Accordingly

$$\textstyle\sum_{j<n}H_j \le 2C_3^2\gamma^2C_4\sum_{j<n}2^{-k_{\tau_j}} \le 2C_3^2C_4\cdot 4\alpha.$$

Putting these together,

$$\alpha^2 \le \sum_{\sigma,\tau\in P',J_\sigma=J_\tau}\bigl|(f|\phi_\sigma)(\phi_\sigma|\phi_\tau)(\phi_\tau|f)\bigr| + 2\sum_{\sigma,\tau\in P',J_\sigma\subseteq J^l_\tau}\bigl|(f|\phi_\sigma)(\phi_\sigma|\phi_\tau)(\phi_\tau|f)\bigr|$$
$$\le C_3\alpha + 2\sqrt{\alpha}\sqrt{\sum_{j<n}H_j} \le C_3\alpha + 4C_3\sqrt{\alpha}\sqrt{2C_4\alpha}$$
$$= \alpha\bigl(C_3 + 4C_3\sqrt{2C_4}\bigr) = \frac{1}{4}\alpha C_6.$$

But this means that

$$\gamma^2\textstyle\sum_{j<n}\mu I_{\tau_j} \le 4\alpha \le C_6,$$

and $P_2=\{\tau_j : j<n\}$ has the property required.

286L Lemma Set

$$C_7 = C_1\bigl(6 + \frac{28}{w(3/2)} + \frac{4\sqrt{14C_3}}{w(3/2)}\bigr).$$

Suppose that P is a finite subset of Q with an upper bound τ in Q. Suppose that $E \subseteq \mathbb{R}$ is measurable, $h : \mathbb{R} \to \mathbb{R}$ is measurable and $f \in \mathcal{L}^2_{\mathbb{C}}$. Then

$$\sum_{\sigma\in P} |(f|\phi_\sigma)\int_{E\cap h^{-1}[J^r_\sigma]} \phi_\sigma| \le 2^{-k_\tau} C_7 \operatorname{energy}_f(P) \operatorname{mass}_{Eh}(P).$$

proof Set $\gamma = \operatorname{energy}_f(P)$, $\gamma' = \operatorname{mass}_{Eh}(P)$. If $P = \emptyset$ the result is trivial, so suppose that $P \ne \emptyset$.

(a)(i) For a dyadic interval $J = \left[2^k n, 2^k(n+1)\right[$ set $J^* = \left[2^k(n-1), 2^k(n+2)\right[$, the half-open interval with the same centre as J and three times its length. Let $\mathcal{J}$ be the family of those $J \in \mathcal{I}$ such that $I_\sigma \not\subseteq J^*$ for any $\sigma \in P$ such that $\mu I_\sigma \le \mu J$. Because P is finite, all sufficiently small intervals belong to $\mathcal{J}$, and $\bigcup \mathcal{J} = \mathbb{R}$; let $\mathcal{K}$ be the set of maximal members of $\mathcal{J}$, so that $\mathcal{K}$ is disjoint. Then $\bigcup \mathcal{K} = \mathbb{R}$. **P** The point is that $P \ne \emptyset$; fix $\sigma \in P$ for the moment. If $J \in \mathcal{J}$, consider for each $n \in \mathbb{N}$ the interval $\tilde J^{(n)} \in \mathcal{I}$ including J with length $2^n \mu J$. Then there is some $n \in \mathbb{N}$ such that $\mu \tilde J^{(n)} \ge \mu I_\sigma$ and $I_\sigma \subseteq (\tilde J^{(n)})^*$, so that $\tilde J^{(k)} \notin \mathcal{J}$ for any $k \ge n$, and there must be some $k < n$ such that $\tilde J^{(k)} \in \mathcal{K}$. Thus $J \subseteq \tilde J^{(k)} \subseteq \bigcup \mathcal{K}$; as J is arbitrary, $\bigcup \mathcal{K} = \bigcup \mathcal{J} = \mathbb{R}$. **Q**

(ii) For $K \in \mathcal{K}$, let $l_K \in \mathbb{Z}$ be such that $\mu K = 2^{-l_K}$. If $l_K \ge k_\tau$, so that $\mu K \le \mu I_\tau$, then K must lie within the interval $\hat I$ with centre x_τ and length $7\mu I_\tau$, since otherwise we should have $I_\tau \cap \tilde K^* = \emptyset$, where $\tilde K$ is the dyadic interval of length $2\mu K$ including K, and $\tilde K$ would belong to $\mathcal{J}$. But this means that

$$\sum_{K\in\mathcal{K}, l_K \ge k_\tau} \mu K \le 7\mu I_\tau = 7\cdot 2^{-k_\tau},$$

because $\mathcal{K}$ is disjoint.

(iii) For any $l < k_\tau$, there are just three members K of $\mathcal{K}$ such that $l_K = l$. **P** If $J \in \mathcal{I}$ and $\mu J > \mu I_\tau$, then either $I_\tau \subseteq J^*$ or $I_\tau \cap J^* = \emptyset$, and $J \in \mathcal{J}$ iff $I_\tau \cap J^*$ is empty. This means that if $K \in \mathcal{I}$ and $\mu K = 2^{-l}$, $K \in \mathcal{K}$ iff $I_\tau \cap K^*$ is empty and $I_\tau \subseteq \tilde K^*$. So if $I_\tau \subseteq \left[2^{-l}n, 2^{-l}(n+1)\right[$ and $K = \left[2^{-l}m, 2^{-l}(m+1)\right[$, we shall have $K \in \mathcal{K}$ iff

$$\textit{either } m = n-2 \textit{ or } m = n+2 \textit{ or } m = n-3 \text{ is even } \textit{or } m = n+3 \text{ is odd};$$

which for any given n happens for just three values of m. **Q**

(b) For $\sigma \in P$, let ζ_σ be a complex number of modulus 1 such that

$$\zeta_\sigma (f|\phi_\sigma)\int_{E\cap h^{-1}[J^r_\sigma]} \phi_\sigma = |(f|\phi_\sigma)\int_{E\cap h^{-1}[J^r_\sigma]} \phi_\sigma|.$$

Set $W = P \times \mathcal{K}$. For $(\sigma, K) \in W$, set

$$\alpha_{\sigma K} = (f|\phi_\sigma)\int_{E\cap h^{-1}[J^r_\sigma]\cap K} \phi_\sigma.$$

The aim of the proof is to estimate

$$\sum_{\sigma\in P}\left|(f|\phi_\sigma)\int_{E\cap h^{-1}[J^r_\sigma]} \phi_\sigma\right| = \sum_{(\sigma,K)\in W} \zeta_\sigma \alpha_{\sigma K}.$$

It will be helpful to note straight away that

$$\sum_{(\sigma,K)\in W} |\alpha_{\sigma K}| \le \sum_{\sigma\in P} |(f|\phi_\sigma)| \int_{-\infty}^{\infty} |\phi_\sigma|$$

is finite.

Set

$$W_0 = \{(\sigma, K) : \sigma \in P,\ K \in \mathcal{K},\ k_\tau \le l_K \le k_\sigma\},$$

$$W_1 = \{(\sigma, K) : \sigma \in P,\ K \in \mathcal{K},\ l_K < k_\tau\},$$

$$W_2 = \{(\sigma, K) : \sigma \in P,\ K \in \mathcal{K},\ k_\sigma < l_K,\ \sigma \not\le_r \tau\},$$

$$W_3 = \{(\sigma, K) : \sigma \in P,\ K \in \mathcal{K},\ k_\sigma < l_K,\ \sigma \le_r \tau\}.$$

Because $k_\sigma \ge k_\tau$ for every $\sigma \in P$, $W = W_0 \cup W_1 \cup W_2 \cup W_3$. I will give estimates for

$$\alpha_j = \sum_{(\sigma,K)\in W_j} \zeta_\sigma \alpha_{\sigma K}$$

for each j; the three components in the expression for C_7 given above are bounds for $|\alpha_0| + |\alpha_1|$, $|\alpha_2|$ and $|\alpha_3|$ respectively.

(c)(i) Whenever $K\in\mathcal{K}$ and $\sigma\in P$,

$$|(f|\phi_\sigma)|\le 2^{-k_\sigma/2}\gamma,$$

by 286H, and

$$\int_{E\cap h^{-1}[J_\sigma^r]\cap K}|\phi_\sigma|\le 2^{-3k_\sigma/2}C_1\int_{E\cap h^{-1}[J_\sigma^r]\cap K}w_\sigma^2$$

(286E(c-iii))

$$\le 2^{-3k_\sigma/2}C_1\int_{E\cap h^{-1}[J_\sigma]}w_\sigma\cdot\sup_{x\in K}w_\sigma(x)$$

$$\le 2^{-3k_\sigma/2}C_1\gamma'\sup_{x\in K}w_\sigma(x)$$

$$=2^{-k_\sigma/2}C_1\gamma' w(2^{k_\sigma}\rho(x_\sigma,K)),$$

where I write $\rho(x_\sigma,K)$ for $\inf_{x\in K}|x-x_\sigma|$. So, for fixed $K\in\mathcal{K}$ and $k\ge l_K$,

$$\sum_{\sigma\in P,k_\sigma=k}|\alpha_{\sigma K}|\le 2^{-k}C_1\gamma\gamma'\sum_{\sigma\in P,k_\sigma=k}w(2^k\rho(x_\sigma,K))$$

$$\le 2^{-k}C_1\gamma\gamma'\cdot 2\sum_{n=2^{k-l_K}}^{\infty}w(n+\tfrac{1}{2})$$

because the x_σ, for $\sigma\in P$ and $k_\sigma=k$, are all distinct and all a distance at least 2^{-l_K} from K (because $I_\sigma\not\subseteq K^*$); so there are at most two σ with $\rho(x_\sigma,K)=2^{-k}(n+\frac{1}{2})$ for each $n\ge 2^{k-l_K}$. So we have

$$\textstyle\sum_{\sigma\in P,k_\sigma=k}|\alpha_{\sigma K}|\le 2^{-k}C_1\gamma\gamma'(1+2^{k-l_K})^{-2}\le 2^{-k-2}C_1\gamma\gamma'$$

by 286Ga. And this is true whenever $K\in\mathcal{K}$ and $k\ge l_K$.

(ii) Now

$$|\alpha_0|\le\sum_{(\sigma,K)\in W_0}|\alpha_{\sigma K}|=\sum_{\substack{K\in\mathcal{K}\\ l_K\ge k_\tau}}\sum_{\substack{\sigma\in P\\ k_\sigma\ge l_K}}|\alpha_{\sigma K}|$$

$$=\sum_{\substack{K\in\mathcal{K}\\ l_K\ge k_\tau}}\sum_{k=l_K}^{\infty}\sum_{\substack{\sigma\in P\\ k_\sigma=k}}|\alpha_{\sigma K}|$$

$$\le\sum_{\substack{K\in\mathcal{K}\\ l_K\ge k_\tau}}\sum_{k=l_K}^{\infty}2^{-k-2}C_1\gamma\gamma'=C_1\gamma\gamma'\sum_{K\in\mathcal{K},l_K\ge k_\tau}2^{-l_K-1}$$

$$=\frac{1}{2}C_1\gamma\gamma'\sum_{K\in\mathcal{K},l_K\ge k_\tau}\mu K\le 4\cdot 2^{-k_\tau}C_1\gamma\gamma'$$

by the formula in (a-ii). This deals with α_0.

(d) Next consider W_1. We have

$$|\alpha_1|\le\sum_{(\sigma,K)\in W_1}|\alpha_{\sigma K}|=\sum_{\substack{K\in\mathcal{K}\\ l_K<k_\tau}}\sum_{k=k_\tau}^{\infty}\sum_{\substack{\sigma\in P\\ k_\sigma=k}}|\alpha_{\sigma K}|$$

$$\le C_1\gamma\gamma'\sum_{\substack{K\in\mathcal{K}\\ l_K<k_\tau}}\sum_{k=k_\tau}^{\infty}2^{-k}(1+2^{k-l_K})^{-2}$$

(by (c-i) above)

$$\leq C_1\gamma\gamma' \sum_{l=-\infty}^{k_\tau-1} \sum_{\substack{K\in\mathcal{K}\\ l_K=l}} (1+2^{k_\tau-l})^{-2} \sum_{k=k_\tau}^{\infty} 2^{-k}$$

$$= 2^{-k_\tau+1}C_1\gamma\gamma' \sum_{l=-\infty}^{k_\tau-1} \sum_{\substack{K\in\mathcal{K}\\ l_K=l}} (1+2^{k_\tau-l})^{-2}$$

$$= 6\cdot 2^{-k_\tau}C_1\gamma\gamma' \sum_{l=-\infty}^{k_\tau-1} (1+2^{k_\tau-l})^{-2}$$

(by (a-iii))

$$= 6\cdot 2^{-k_\tau}C_1\gamma\gamma' \sum_{l=1}^{\infty}(1+2^l)^{-2} \leq 2\cdot 2^{-k_\tau}C_1\gamma\gamma'$$

because

$$\textstyle\sum_{l=1}^{\infty}(1+2^l)^{-2} \leq \sum_{l=1}^{\infty}\frac{1}{4^l} = \frac{1}{3}.$$

This deals with α_1.

(e) For $K \in \mathcal{K}$, set $G_K = K \cap E \cap \bigcup_{\sigma\in P, k_\sigma<l_K} h^{-1}[J_\sigma]$. Then $\mu G_K \leq 2\gamma'\mu K/w(\frac{3}{2})$. **P** If $l_K \leq k_\tau$, then $G_K = \emptyset$, so we may suppose that $l_K > k_\tau$. Let $\tilde{K} \in \mathcal{I}$ be the dyadic interval containing K and with twice the length. Then $\tilde{K} \notin \mathcal{J}$, so there is a $\sigma \in P$ such that $\tilde{K}^* \supseteq I_\sigma$ and $\mu I_\sigma \leq \mu\tilde{K}$, so that $k_\sigma \geq l_K - 1 \geq k_\tau$. Let $v \in Q$ be such that $\sigma \leq v \leq \tau$ and $k_v = l_K - 1$ (286F(a-iv)). Then I_v meets $\tilde{K}^*$, so $\tilde{K}$ is either equal to I_v or adjacent to it, and $|x - x_v| \leq \frac{3}{2}\cdot 2^{-k_v}$ for every $x \in \tilde{K}$, therefore for every $x \in K$. Accordingly

$$w_v(x) \geq 2^{k_v}w(\tfrac{3}{2}) = w(\tfrac{3}{2})/2\mu K$$

for every $x \in K$. On the other hand, because $\sigma \in P$ and $\sigma \leq v$, $\int_{E\cap h^{-1}[J_v]} w_v \leq \gamma'$. So

$$\mu(E \cap h^{-1}[J_v] \cap K) \leq 2\gamma'\mu K/w(\tfrac{3}{2}).$$

Now suppose that $\sigma' \in P$ and $k_{\sigma'} < l_K$. Then $J_{\sigma'}$ is the dyadic interval of length $2^{k_{\sigma'}}$ including J_τ. But J_v is the dyadic interval of length 2^{k_v} including J_τ, so includes $J_{\sigma'}$, and $h^{-1}[J_{\sigma'}] \subseteq h^{-1}[J_v]$. As σ' is arbitrary, $G_K \subseteq E \cap h^{-1}[J_v] \cap K$ and $\mu G_K \leq 2\gamma'\mu K/w(\frac{3}{2})$, as claimed. **Q**

(f)(i) For $x \in \mathbb{R}$, set

$$v_2(x) = \bigl|\textstyle\sum_{(\sigma,K)\in W_2} \zeta_\sigma(f|\phi_\sigma)\phi_\sigma(x)\chi(E\cap h^{-1}[J_\sigma^r]\cap K)(x)\bigr|.$$

Then, for any $x \in \mathbb{R}$, either $v_2(x) = 0$ or there is a $k \geq k_\tau$ such that

$$v_2(x) = \bigl|\textstyle\sum_{\sigma\in P, k_\sigma=k} \zeta_\sigma(f|\phi_\sigma)\phi_\sigma(x)\bigr|.$$

P If $v_2(x) \neq 0$, then we have a pair $(v, L) \in W_2$ such that $x \in E \cap h^{-1}[J_v^r] \cap L$. Now suppose that $(\sigma, K) \in W_2$ and either $K \neq L$ or $k_\sigma \neq k_v$. If $K \neq L$ then of course $x \notin L$, because $\mathcal{K}$ is a disjoint family. If $k_\sigma \neq k_v$, then examine J_σ and J_v. These are dyadic intervals of different lengths, and both include J_τ. On the other hand, neither of the right-hand halves J_σ^r, J_v^r includes J_τ^r, because $\sigma \not\leq_r \tau$ and $v \not\leq_r \tau$. So either $J_\sigma \cap J_v^r = \emptyset$ (if $k_\sigma < k_v$) or $J_v \cap J_\sigma^r = \emptyset$ (if $k_\sigma > k_v$); in either case, $J_\sigma^r \cap J_v^r$ is empty, and $x \notin h^{-1}[J_\sigma^r]$.

On the other hand, of couse, if $\sigma \in P$ and $k_\sigma = k_v$, then $k_\sigma < l_L$ and $J_\sigma^r = J_v^r$ does not include J_τ^r, so that $(\sigma, L) \in W_2$ and $x \in E \cap h^{-1}[J_\sigma^r] \cap L$.

Thus

$$v_2(x) = |\textstyle\sum_{(\sigma,K)\in W_2, x\in h^{-1}[J_\sigma^r]\cap K} \zeta_\sigma(f|\phi_\sigma)\phi_\sigma(x)| = |\sum_{\sigma\in P, k_\sigma=k_v} \zeta_\sigma(f|\phi_\sigma)\phi_\sigma(x)|,$$

and we can set $k = k_v$. **Q**

(ii) It follows that $v_2(x) \leq 2C_1\gamma$ for every $x \in \mathbb{R}$. **P** If $v_2(x) = 0$ this is trivial. Otherwise, take k from (i). Then

$$v_2(x) \leq \sum_{\sigma\in P, k_\sigma=k} |(f|\phi_\sigma)\phi_\sigma(x)| \leq \sum_{\sigma\in P, k_\sigma=k} 2^{-k/2}\gamma\cdot 2^{-k/2}C_1 w_\sigma(x)$$

(by 286H and 286E(c-iii))

$$= C_1\gamma\sum_{\sigma\in P,k_\sigma=k} w(2^k(x-x_\sigma)) \le C_1\gamma\sum_{n=-\infty}^{\infty} w(2^kx-n-\tfrac12)$$

(because the x_σ, for $\sigma\in P$ and $k_\sigma=k$, are all distinct and of the form $2^{-k}(n+\frac12)$)

$$\le 2C_1\gamma$$

by 286Gc. **Q**

(iii) Note also that, if $v_2(x)>0$, there is a pair $(\sigma,K)\in W_2$ such that $x\in h^{-1}[J_\sigma]\cap K$, so that $k_\tau\le k_\sigma<l_K$ and $x\in G_K$. But now we have

$$|\alpha_2| = \Bigl|\sum_{(\sigma,K)\in W_2}\zeta_\sigma(f|\phi_\sigma)\int_{-\infty}^{\infty}\phi_\sigma\times\chi(E\cap h^{-1}[J^r_\sigma]\cap K)\Bigr|$$

$$\le\int_{-\infty}^{\infty} v_2 = \sum_{K\in\mathcal{K},l_K>k_\tau}\int_{G_K} v_2 \le \sum_{K\in\mathcal{K},l_K>k_\tau} 4C_1\gamma\gamma'\mu K/w(\tfrac32)$$

(putting the estimates in (e) and (ii) just above together)

$$\le 28\cdot 2^{-k_\tau}C_1\gamma\gamma'/w(\tfrac32)$$

by (a-ii). This deals with α_2.

(g) Set $P'=\{\sigma:\sigma\in P,\ \sigma\le_r\tau\}$ and $\tilde g=\sum_{\sigma\in P'}\zeta_\sigma(f|\phi_\sigma)\phi_\sigma$. Then

$$\|\tilde g\|_2^2\le 2^{-k_\tau}C_3\gamma^2.$$

P If σ, $\sigma'\in P'$ and $k_\sigma\ne k_{\sigma'}$, then $(\phi_\sigma|\phi_{\sigma'})=0$ (286Fb). While if $k_\sigma=k_{\sigma'}$, then $J_\sigma=J_{\sigma'}$, because σ and σ' have a common upper bound τ. So

$$\|\tilde g\|_2^2=\sum_{\sigma,\sigma'\in P'}\zeta_\sigma(f|\phi_\sigma)(\phi_\sigma|\phi_{\sigma'})(\phi_{\sigma'}|f)\bar\zeta_{\sigma'}$$

$$\le\sum_{\sigma,\sigma'\in P',J_\sigma=J_{\sigma'}}\bigl|(f|\phi_\sigma)(\phi_\sigma|\phi_{\sigma'})(\phi_{\sigma'}|f)\bigr|\le C_3\sum_{\sigma\in P'}|(f|\phi_\sigma)|^2$$

(by 286J)

$$\le C_3\cdot 2^{-k_\tau}\gamma^2$$

by the definition of ‘energy’. **Q**

(h) For $m\in\mathbb{N}$, set

$$\tilde g_m=\textstyle\sum_{\sigma\in P',k_\sigma\le m}\zeta_\sigma(f|\phi_\sigma)\phi_\sigma.$$

Then whenever x, $x'\in\mathbb{R}$ and $|x-x'|\le 2^{-m}$, $|\tilde g_m(x)|\le\frac12 C_1\tilde g^*(x')$, where

$$\tilde g^*(x')=\sup_{a\le x'\le b,a<b}\frac{1}{b-a}\int_a^b|\tilde g|$$

as in 286A. **P (i)** Since $k_\sigma\ge k_\tau$ for every $\sigma\in P'$, we may take it that $m\ge k_\tau$. Let $\hat J$ be the dyadic interval of length 2^m including J_τ, and $\hat y$ its midpoint. Set $\psi=S_{-\hat y}D_{2^{-m}/3}\overset{\wedge}{\phi}$, that is, $\psi(y)=\overset{\wedge}{\phi}(\frac13 2^{-m}(y-\hat y))$ for $y\in\mathbb{R}$.

(ii) If $\sigma\in P'$ and $k_\sigma\le m$ and $\phi_\sigma(y)\ne0$, then $y\in J_\sigma$. But $J_\sigma\cap\hat J\supseteq J_\tau$ is not empty, so $J_\sigma\subseteq\hat J$, $|y-\hat y|\le\frac12 2^m$, $|\frac13 2^{-m}(y-\hat y)|\le\frac16$ and $\psi(y)=1$.

(iii) If $\sigma\in P'$ and $k_\sigma>m$ and $\phi_\sigma(y)\ne0$, then $J^r_\sigma\cap\hat J\supseteq J^r_\tau$ is non-empty, so $\hat J\subseteq J^r_\sigma$ and $y\le y_\sigma\le\hat y$; now

$$\hat y-y=(\hat y-y_\sigma)+(y_\sigma-y)\ge\frac12\cdot2^m+\frac1{20}\cdot2^{k_\sigma}\ge\frac35\cdot2^m,\quad|\frac13\cdot2^{-m}(y-\hat y)|\ge\frac15$$

and $\psi(y)=0$.

(iv) What this means is that if $\sigma\in P'$ then

$$\overset{\wedge}{\phi}_\sigma \times \psi = \overset{\wedge}{\phi}_\sigma \text{ if } k_\sigma \le m,$$
$$= 0 \text{ if } k_\sigma > m,$$

so that $\overset{\wedge}{\tilde{g}}_m = \psi \times \overset{\wedge}{\tilde{g}}$.

(v) By 283M, $\tilde{g}_m = \frac{1}{\sqrt{2\pi}}\tilde{g} * \overset{\vee}{\psi}$, where $\tilde{g} * \overset{\vee}{\psi}$ is the convolution of $\tilde{g}$ and the inverse Fourier transform $\overset{\vee}{\psi}$ of ψ. (Strictly speaking, 283M, with the help of 284C, tells us that $\tilde{g}_m$ and $\frac{1}{\sqrt{2\pi}}\tilde{g} * \overset{\vee}{\psi}$ have the same Fourier transforms. By 283G, they are equal almost everywhere; by 255K, the convolution is defined everywhere and is continuous; so in fact they are the same function.) Now

$$\overset{\vee}{\psi} = 3 \cdot 2^m M_{\hat{y}} D_{3\cdot 2^m} \overset{\wedge\vee}{\phi} = 3 \cdot 2^m M_{\hat{y}} D_{3\cdot 2^m}\phi,$$

that is,

$$\overset{\vee}{\psi}(x) = 3 \cdot 2^m e^{ix\hat{y}}\phi(3 \cdot 2^m x)$$

for $y \in \mathbb{R}$.

(vi) Set $\theta(x) = \min(w(3), w(x))$ for $x \in \mathbb{R}$, so that θ is non-decreasing on $]-\infty, -3]$, non-increasing on $[3, \infty[$, and constant on $[-3, 3]$, and $|\phi(x)| \le C_1\theta(x)$ for every x, by the choice of C_1 (286Ec). Take any x, $x' \in \mathbb{R}$ such that $|x - x'| \le 2^{-m}$. Then

$$|\tilde{g}_m(x)| \le \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |\tilde{g}(x-t)||\overset{\vee}{\psi}(t)|dt = \frac{3\cdot 2^m}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |\tilde{g}(x-t)||\phi(3 \cdot 2^m t)|dt$$
$$\le \frac{3\cdot 2^m}{\sqrt{2\pi}} C_1 \int_{-\infty}^{\infty} |\tilde{g}(x-t)|\theta(3 \cdot 2^m t)dt = \frac{3\cdot 2^m}{\sqrt{2\pi}} C_1 \int_{-\infty}^{\infty} |\tilde{g}(x+t)|\theta(3 \cdot 2^m t)dt$$

(because θ is an even function)

$$\le \frac{3\cdot 2^m}{\sqrt{2\pi}} C_1 \int_{-\infty}^{\infty} \theta(3 \cdot 2^m t)dt \cdot \sup_{a \le -2^{-m}, b \ge 2^{-m}} \frac{1}{b-a} \int_a^b |\tilde{g}(x+t)|dt$$

(by 286B, because $t \mapsto \theta(3 \cdot 2^m t)$ is non-decreasing on $]-\infty, -2^{-m}]$, non-increasing on $[2^{-m}, \infty[$ and constant on $[-2^{-m}, 2^{-m}]$)

$$= \frac{1}{\sqrt{2\pi}} C_1 \int_{-\infty}^{\infty} \theta \cdot \sup_{a \le x - 2^{-m}, b \ge x + 2^{-m}} \frac{1}{b-a} \int_a^b |\tilde{g}| \le \frac{1}{2} C_1 \int_{-\infty}^{\infty} w \cdot \tilde{g}^*(x')$$

(because if $a \le x - 2^{-m}$ and $b \ge x + 2^{-m}$ then $a \le x' \le b$)

$$= \frac{1}{2} C_1 \tilde{g}^*(x')$$

as required. **Q**

(i) For $x \in \mathbb{R}$, set

$$v_3(x) = \bigl|\textstyle\sum_{(\sigma,K)\in W_3} \zeta_\sigma(f|\phi_\sigma)\phi_\sigma(x)\chi(E \cap h^{-1}[J_\sigma^r] \cap K)(x)\bigr|.$$

Then whenever $L \in \mathcal{K}$ and x, $x' \in L$, $|v_3(x)| \le C_1\tilde{g}^*(x')$. **P** We may suppose that $v_3(x) \ne 0$, so that, in particular, $x \in E$. The only pairs (σ, K) contributing to the sum forming $v_3(x)$ are those in which $x \in K$, so that $K = L$, and $h(x) \in J_\sigma^r$. Moreover, since we are looking only at σ such that $\sigma \le_r \tau$, so that $J_\tau^r \subseteq J_\sigma^r$, J_σ^r will always be the dyadic interval of length $2^{k_\sigma - 1}$ including J_τ^r. So these intervals are nested, and there will be some m such that (for $\sigma \le_r \tau$) $h(x) \in J_\sigma^r$ iff $k_\sigma \ge m$. Accordingly

$$v_3(x) = \bigl|\textstyle\sum_{\sigma \in P', m \le k_\sigma < l_L} \zeta_\sigma(f|\phi_\sigma)\phi_\sigma(x)\bigr| = |\tilde{g}_{l_L - 1}(x) - \tilde{g}_{m-1}(x)|$$

(we must have $m < l_L$ because $v_3(x) \ne 0$). Now $|x - x'| \le 2^{-l_L+1} \le 2^{-m+1}$, so (h) tells us that both $|\tilde{g}_{l_L-1}(x)|$ and $|\tilde{g}_{m-1}(x)|$ are at most $\frac{1}{2}C_1\tilde{g}^*(x')$, and $v_3(x) \le C_1\tilde{g}^*(x')$, as claimed. **Q**

It follows that $v_3(x) \le \frac{C_1}{\mu L}\int_L \tilde{g}^*$ for every $x \in L$.

(j) Now we are in a position to estimate

$$|\alpha_3| = |\sum_{(\sigma,K)\in W_3} \zeta_\sigma \alpha_{\sigma K}| \le \int_{-\infty}^{\infty} v_3 \le \sum_{K\in\mathcal{K},l_K>k_\tau} \int_{G_K} v_3$$

(because if $v_3(x) \ne 0$ there are $(\sigma, K) \in W_3$ such that $x \in K$, and in this case $x \in G_K$ and $l_K > k_\sigma \ge k_\tau$)

$$\le \sum_{K\in\mathcal{K},l_K>k_\tau} \mu G_K \cdot \frac{C_1}{\mu K} \int_K \tilde{g}^*$$

(by (i) above, because $G_K \subseteq K$)

$$\le C_1 \sum_{K\in\mathcal{K},l_K>k_\tau} \frac{2\gamma'}{w(\frac{3}{2})} \int_K \tilde{g}^*$$

(by (e))

$$\le \frac{2C_1\gamma'}{w(\frac{3}{2})} \int_{\hat{I}} \tilde{g}^*$$

(because if $l_K > k_\tau$ then $K \subseteq \hat{I}$, as noted in (a-ii))

$$\le \frac{2C_1\gamma'}{w(\frac{3}{2})} \sqrt{\mu\hat{I}} \cdot \|\tilde{g}^*\|_2$$

(by Cauchy's inequality)

$$\le \frac{2C_1\gamma'}{w(\frac{3}{2})} \sqrt{7} \cdot 2^{-k_\tau/2}\sqrt{8}\|\tilde{g}\|_2$$

(by the Maximal Theorem, 286A)

$$\le \frac{4C_1\gamma'\sqrt{14}}{w(\frac{3}{2})} 2^{-k_\tau/2} \cdot 2^{-k_\tau/2}\sqrt{C_3}\gamma$$

(by (g))

$$= \frac{4C_1\sqrt{14C_3}}{w(\frac{3}{2})} \gamma\gamma' 2^{-k_\tau}.$$

(k) Assembling these,

$$\begin{aligned}\sum_{\sigma\in P} \Bigl|(f|\phi_\sigma) \int_{E\cap h^{-1}[J_\sigma^r]} \phi_\sigma\Bigr| &= \sum_{\sigma\in P,K\in\mathcal{K}} \zeta_\sigma \alpha_{\sigma K} = \sum_{j=0}^{3} \sum_{(\sigma,K)\in W_j} \zeta_\sigma \alpha_{\sigma K} \le \sum_{j=0}^{3} |\alpha_j| \\ &\le 4 \cdot 2^{-k_\tau} C_1\gamma\gamma' + 2 \cdot 2^{-k_\tau} C_1\gamma\gamma' + 28 \cdot 2^{-k_\tau} C_1\gamma\gamma'/w(\tfrac{3}{2}) \\ &\qquad + 4\sqrt{14C_3} \cdot 2^{-k_\tau} C_1\gamma\gamma'/w(\tfrac{3}{2}) \\ &= 2^{-k_\tau} C_7\gamma\gamma',\end{aligned}$$

as claimed.

286M The Lacey-Thiele lemma Set $C_8 = 3C_7(C_5 + C_6)$. Then

$$\textstyle\sum_{\sigma\in Q} |(f|\phi_\sigma)\int_{E\cap h^{-1}[J_\sigma^r]} \phi_\sigma| \le C_8$$

whenever $f \in \mathcal{L}^2_{\mathbb{C}}$, $\|f\|_2 = 1$, $\mu E \le 1$ and $h : \mathbb{R} \to \mathbb{R}$ is measurable.

proof (a) The first step is to combine 286I and 286K, as follows: if $P \subseteq Q$ is finite and $\max(\sqrt{\mathrm{mass}_{Eh}(P)}, \mathrm{energy}_f(P)) \le \gamma$, there are $P' \subseteq P$ and $R \subseteq Q$ such that $\max(\sqrt{\mathrm{mass}_{Eh}(P')}, \mathrm{energy}_f(P')) \le \frac{1}{2}\gamma$, $\gamma^2 \sum_{\tau\in R} 2^{-k_\tau} \le C_5 + C_6$, and $P \setminus P' \preccurlyeq R$. **P** Since $\mathrm{mass}_{Eh}(P) \le \gamma^2$, 286I tells us that there are $P_0 \subseteq P$, $R_0 \subseteq Q$ such that $\mathrm{mass}_{Eh}(P_0) \le \frac{1}{4}\gamma^2$, $\gamma^2 \sum_{\tau\in R_0} 2^{-k_\tau} \le C_5$, and $P \setminus P_0 \preccurlyeq R_0$. Now turn to 286K: since $\mathrm{energy}_f(P_0) \le \mathrm{energy}_f(P) \le \gamma$, we can find $P' \subseteq P_0$ and $R_1 \subseteq Q$ such that $\mathrm{energy}_f(P') \le \frac{1}{2}\gamma$, $\gamma^2 \sum_{\tau\in R_1} 2^{-k_\tau} \le C_6$, and $P_0 \setminus P' \preccurlyeq R_1$. Now $\mathrm{mass}_{Eh}(P') \le \mathrm{mass}_{Eh}(P_0) \le \frac{1}{4}\gamma^2$, so $\max(\sqrt{\mathrm{mass}_{Eh}(P')}, \mathrm{energy}_f(P')) \le \frac{1}{2}\gamma$; and setting $R = R_0 \cup R_1$, $\gamma^2 \sum_{\tau\in R} 2^{-k_\tau} \le C_5 + C_6$, while $P \setminus P' \preccurlyeq R$. **Q**

(b) Now take any finite $P \subseteq Q$. Let $k \in \mathbb{N}$ be such that $\max(\sqrt{\mathrm{mass}_{Eh}(P)}, \mathrm{energy}_f(P)) \le 2^k$. By (a), we can choose $\langle P_n\rangle_{n\in\mathbb{N}}$, $\langle R_n\rangle_{n\in\mathbb{N}}$ inductively such that $P_0 = P$ and, for each $n \in \mathbb{N}$,

$$P_{n+1} \subseteq P_n, \quad P_n \setminus P_{n+1} \preccurlyeq R_n,$$

$$\max(\sqrt{\mathrm{mass}_{Eh}(P_n)}, \mathrm{energy}_f(P_n)) \le 2^{k-n}, \quad 2^{2k-2n}\textstyle\sum_{\tau\in R_n} 2^{-k_\tau} \le C_5 + C_6.$$

Since $\mathrm{energy}_f(\{\sigma\}) = 2^{k_\sigma/2}|(f|\phi_\sigma)| > 0$ whenever $(f|\phi_\sigma) \ne 0$ (286H), $(f|\phi_\sigma) = 0$ whenever $\sigma \in \bigcap_{n\in\mathbb{N}} P_n$, and

$$\begin{aligned}\sum_{\sigma\in P}\bigl|(f|\phi_\sigma)\int_{E\cap h^{-1}[J^r_\sigma]}\phi_\sigma\bigr| &= \sum_{\sigma\in\bigcup_{n\in\mathbb{N}} P_n\setminus P_{n+1}}\bigl|(f|\phi_\sigma)\int_{E\cap h^{-1}[J^r_\sigma]}\phi_\sigma\bigr| \\ &= \sum_{n=0}^{\infty}\sum_{\sigma\in P_n\setminus P_{n+1}}\bigl|(f|\phi_\sigma)\int_{E\cap h^{-1}[J^r_\sigma]}\phi_\sigma\bigr| \\ &\le \sum_{n=0}^{\infty}\sum_{\tau\in R_n}\sum_{\substack{\sigma\in P_n\\ \sigma\le\tau}}\bigl|(f|\phi_\sigma)\int_{E\cap h^{-1}[J^r_\sigma]}\phi_\sigma\bigr| \\ &\le \sum_{n=0}^{\infty}\sum_{\tau\in R_n} 2^{-k_\tau} C_7 \,\underset{f}{\mathrm{energy}}(P_n)\,\underset{Eh}{\mathrm{mass}}(P_n)\end{aligned}$$

(by 286L)

$$\le \sum_{n=0}^{\infty} C_7(C_5+C_6)2^{2n-2k}2^{k-n}\min(1, 2^{2k-2n})$$

(because $\mathrm{mass}_{Eh}(P_n) \le 1$ for every n, as noted in 286H)

$$\begin{aligned}&= C_7(C_5+C_6)\sum_{n=0}^{\infty}\min(2^{n-k}, 2^{k-n}) \\ &\le C_7(C_5+C_6)\sum_{n=-\infty}^{\infty}\min(2^n, 2^{-n}) = 3C_7(C_5+C_6).\end{aligned}$$

(c) Since this true for every finite $P \subseteq Q$,

$$\textstyle\sum_{\sigma\in Q}|(f|\phi_\sigma)\int_{E\cap h^{-1}[J^r_\sigma]}\phi_\sigma| \le C_8,$$

as claimed.

286N Lemma Set $C_9 = C_8\sqrt{2}$. Suppose that $f \in \mathcal{L}^2_{\mathbb{C}}$, $h : \mathbb{R} \to \mathbb{R}$ is measurable and $\mu F < \infty$. Then

$$\textstyle\sum_{\sigma\in Q}|(f|\phi_\sigma)\int_{F\cap h^{-1}[J^r_\sigma]}\phi_\sigma| \le C_9\|f\|_2\sqrt{\mu F}.$$

proof This is trivial if $\|f\|_2 = 0$, that is, $f = 0$ a.e. So we may take it that $\|f\|_2 > 0$. Dividing both sides by $\|f\|_2$, we may suppose that $\|f\|_2 = 1$.

Let $k \in \mathbb{Z}$ be such that $2^{k-1} < \mu F \le 2^k$. We have a bijection $\sigma \mapsto \sigma^* : Q \to Q$ defined by saying that $\sigma^* = (2^{-k}I_\sigma, 2^kJ_\sigma)$; so that $k_{\sigma^*} = k_\sigma + k$, $x_{\sigma^*} = 2^{-k}x_\sigma$, $y^l_{\sigma^*} = 2^ky^l_\sigma$, $J^r_{\sigma^*} = 2^kJ^r_\sigma$, and for every $x \in \mathbb{R}$

$$\begin{aligned}\phi_\sigma(2^kx) &= 2^{k_\sigma/2}e^{2^kiy^l_\sigma x}\phi(2^{k_\sigma}(2^kx - x_\sigma)) \\ &= 2^{k_\sigma/2}e^{iy^l_{\sigma^*}x}\phi(2^{k_\sigma+k}(x - 2^{-k}x_\sigma)) \\ &= 2^{k_\sigma/2}e^{iy^l_{\sigma^*}x}\phi(2^{k_{\sigma^*}}(x - x_{\sigma^*})) = 2^{-k/2}\phi_{\sigma^*}(x).\end{aligned}$$

Write $\tilde F = 2^{-k}F$, so that $\mu\tilde F \le 1$, and $\tilde h(x) = 2^kh(2^kx)$ for every x. Then, for $\sigma \in Q$,

$$\begin{aligned}F\cap h^{-1}[J^r_\sigma] &= \{x : x \in F,\ h(x) \in J^r_\sigma\} = \{x : 2^{-k}x \in \tilde F,\ 2^{-k}\tilde h(2^{-k}x) \in J^r_\sigma\} \\ &= \{x : 2^{-k}x \in \tilde F,\ \tilde h(2^{-k}x) \in J^r_{\sigma^*}\} = 2^k\{x : x \in \tilde F,\ \tilde h(x) \in J^r_{\sigma^*}\}.\end{aligned}$$

Write $\tilde f(x) = 2^{k/2}f(2^kx)$, so that

$$\|\tilde{f}\|_2 = 2^{k/2}\|D_{2^k}f\|_2 = \|f\|_2 = 1,$$

while

$$(f|\phi_\sigma) = \int_{-\infty}^{\infty} f \times \bar{\phi}_\sigma = 2^k \int_{-\infty}^{\infty} f(2^k x)\overline{\phi_\sigma(2^k x)}dx = (\tilde{f}|\phi_{\sigma^*})$$

for every $\sigma \in Q$. Putting all these together,

$$\begin{aligned}
\sum_{\sigma\in Q}\bigl|(f|\phi_\sigma)\int_{F\cap h^{-1}[J_\sigma^r]}\phi_\sigma\bigr| &= 2^k\sum_{\sigma\in Q}\bigl|(f|\phi_\sigma)\int_{2^{-k}(F\cap h^{-1}[J_\sigma^r])}\phi_\sigma(2^k x)dx\bigr| \\
&= 2^{k/2}\sum_{\sigma\in Q}\bigl|(\tilde{f}|\phi_{\sigma^*})\int_{\tilde{F}\cap\tilde{h}^{-1}[J_{\sigma^*}^r]}\phi_{\sigma^*}\bigr| \\
&= 2^{k/2}\sum_{\tau\in Q}\bigl|(\tilde{f}|\phi_\tau)\int_{\tilde{F}\cap\tilde{h}^{-1}[J_\tau^r]}\phi_\tau\bigr| \\
&\le 2^{k/2}C_8
\end{aligned}$$

(by the Lacey-Thiele lemma, applied to $\tilde{h}$, $\tilde{F}$ and $\tilde{f}$)

$$\le C_9\sqrt{\mu F} = C_9\|f\|_2\sqrt{\mu F}.$$

286O Lemma Suppose that $f \in \mathcal{L}^2_{\mathbb{C}}$. For $x \in \mathbb{R}$, set

$$(Af)(x) = \sup_{z\in\mathbb{R}}\sum_{\sigma\in Q, z\in J_\sigma^r}|(f|\phi_\sigma)\phi_\sigma(x)|.$$

Then $Af : \mathbb{R} \to [0,\infty]$ is Borel measurable, and $\int_F Af \le C_9\|f\|_2\sqrt{\mu F}$ whenever $\mu F < \infty$.

proof (a) For $z \in \mathbb{R}$ and finite $P \subseteq Q$, set $g_{Pz} = \sum_{\sigma\in P, z\in J_\sigma^r}|(f|\phi_\sigma)\phi_\sigma|$, so that

$$Af(x) = \sup\{g_{Pz}(x) : P \subseteq Q \text{ is finite, } z \in \mathbb{R}\}.$$

Because every g_{Pz} is continuous, Af is Borel measurable and

$$\int_F Af = \sup\{\int_F g_{P_0z_0} \vee \ldots \vee g_{P_nz_n} : P_0, \ldots, P_n \subseteq Q \text{ are finite, } z_0, \ldots, z_n \in \mathbb{R}\}$$

for every measurable set F (256M).

(b) Given finite sets $P_0, \ldots, P_n \subseteq Q$ and $z_0, \ldots, z_n, x \in \mathbb{R}$, set

$$g(x) = \max_{j\le n} g_{P_jz_j}(x), \quad l(x) = \min\{j : j \le n,\, g(x) = g_{P_jz_j}(x)\}, \quad h(x) = z_{l(x)};$$

because every $g_{P_jz_j}$ is continuous, $l : \mathbb{R} \to \{0, \ldots, n\}$ and $h : \mathbb{R} \to \mathbb{R}$ are measurable. Now

$$\begin{aligned}
\int_F g = \int_F g_{P_{l(x)},h(x)}(x)dx &= \int_F \sum_{\sigma\in P_{l(x)}, h(x)\in J_\sigma^r}|(f|\phi_\sigma)\phi_\sigma(x)|dx \\
&\le \int_F \sum_{\sigma\in Q, h(x)\in J_\sigma^r}|(f|\phi_\sigma)\phi_\sigma(x)|dx \\
&= \sum_{\sigma\in Q}\int_{F\cap h^{-1}[J_\sigma^r]}|(f|\phi_\sigma)\phi_\sigma| \le C_9\|f\|_2\sqrt{\mu F}
\end{aligned}$$

by 286N. Since $P_0, \ldots, P_n$ and $z_0, \ldots, z_n$ are arbitrary, $\int_F Af \le C_9\|f\|_2\sqrt{\mu F}$.

286P Lemma For any $z \in \mathbb{R}$, define $\theta_z : \mathbb{R} \to [0,1]$ by setting

$$\theta_z(y) = \hat{\phi}(2^{-k}(y - \hat{y}))^2$$

whenever there is a dyadic interval $J \in \mathcal{I}$ of length 2^k such that z belongs to the right-hand half of J and y belongs to the left-hand half of J and $\hat{y}$ is the lower quartile of J, and zero if there is no such J. Then $0 \le \theta_z(y) \le 1$ for every $y \in \mathbb{R}$, $\theta_z(y) = 0$ if $y \ge z$, and $2\pi|(\hat{f} \times \theta_z)^\vee| \le Af$ for any rapidly decreasing test function f.

proof (a) I had better start by explaining why the recipe above defines a function θ_z. Let M be the set of those $k\in\mathbb{Z}$ such that z belongs to the right-hand half of the dyadic interval $\hat J_k$ of length 2^k containing z. For $k\in M$, let $\hat y_k$ be the midpoint of the left-hand half $\hat J^l_k$ of $\hat J_k$, and set $\psi_k(y)=\hat\phi(2^{-k}(y-\hat y_k))^2$ for $y\in\mathbb{R}$; then ψ_k is smooth and zero outside $\hat J^l_k$. But now observe that if k, k' are distinct members of M, then $\hat J^l_k$ and $\hat J^l_{k'}$ are disjoint, as observed in 286E(b-iii). So θ_z is just the sum $\sum_{k\in M}\psi_k$. Because $\hat\phi$ takes values in $[0,1]$, so does θ_z. If $y\ge z$, then of course $y\notin\hat J^l_k$ for any $k\in M$, so $\theta_z(y)=0$.

(b) Fix a rapidly decreasing test function f. For $k\in M$, set $R_k=\{(I,\hat J_k):I\in\mathcal{I},\,\mu I=2^{-k}\}$, so that $\{\sigma:\sigma\in Q,\,z\in J^r_\sigma\}=\bigcup_{k\in M}R_k$, and $\sum_{k\in M}\sum_{\sigma\in R_k}|(f|\phi_\sigma)\phi_\sigma(x)|\le(Af)(x)$.

Now, for $k\in M$, $2\pi(\hat f\times\psi_k)^\vee=\sum_{\sigma\in R_k}(f|\phi_\sigma)\phi_\sigma$. **P** If $\sigma\in R_k$, $y^l_\sigma=\hat y_k$ and x_σ is of the form $2^{-k}(n+\frac12)$ for some $n\in\mathbb{Z}$. So

$$(f|\phi_\sigma)=\int_{-\infty}^{\infty}\hat f\times\bar{\hat\phi}_\sigma$$

(284O)

$$=\int_{-\infty}^{\infty}\hat f(t)\cdot 2^{-k/2}e^{2^{-k}i(n+\frac12)(t-\hat y_k)}\hat\phi(2^{-k}(t-\hat y_k))dt$$

(by the formula in 286Eb, because $\hat\phi$ is real-valued)

$$=2^{k/2}\int_{-\infty}^{\infty}\hat f(2^kt+\hat y_k)e^{i(n+\frac12)t}\hat\phi(t)dt$$

$$=2^{k/2}\int_{-\pi}^{\pi}\hat f(2^kt+\hat y_k)e^{i(n+\frac12)t}\hat\phi(t)dt$$

(because $\hat\phi(t)=0$ if $|t|\ge\frac15$)

$$=2^{k/2}\int_{-\pi}^{\pi}g(t)e^{int}dt,$$

where $g(t)=\hat f(2^kt+\hat y_k)e^{it/2}\hat\phi(t)$ for $-\pi<t\le\pi$. So if we set $c_n=\frac{1}{2\pi}\int_{-\pi}^{\pi}g(t)e^{-int}dt$, as in 282A, we have

$$(f|\phi_\sigma)=2^{k/2}\cdot 2\pi c_{-n}$$

when $\sigma\in R_k$ and $x_\sigma=2^{-k}(n+\frac12)$. Note that as g is smooth and zero outside $[-\frac15,\frac15]$, $\sum_{n=-\infty}^{\infty}|c_n|<\infty$ (282Rb).

Now, for any $y\in\hat J^l_k$,

$$\sum_{\sigma\in R_k}(f|\phi_\sigma)\hat\phi_\sigma(y)=\sum_{n=-\infty}^{\infty}2^{k/2}\cdot 2\pi c_{-n}\cdot 2^{-k/2}e^{-2^{-k}i(n+\frac12)(y-\hat y_k)}\hat\phi(2^{-k}(y-\hat y_k))$$

$$=2\pi\hat\phi(2^{-k}(y-\hat y_k))e^{-2^{-k-1}i(y-\hat y_k)}\sum_{n=-\infty}^{\infty}c_{-n}e^{-2^{-k}in(y-\hat y_k)}$$

$$=2\pi\hat\phi(2^{-k}(y-\hat y_k))e^{-2^{-k-1}i(y-\hat y_k)}\sum_{n=-\infty}^{\infty}c_ne^{in2^{-k}(y-\hat y_k)}$$

$$=2\pi\hat\phi(2^{-k}(y-\hat y_k))e^{-2^{-k-1}i(y-\hat y_k)}g(2^{-k}(y-\hat y_k))$$

(by 282L, because $2^{-k}|y-\hat y_k|\le\frac14<\pi$)

$$=2\pi e^{-2^{-k-1}i(y-\hat y_k)}\hat\phi(2^{-k}(y-\hat y_k))\hat f(y)e^{2^{-k-1}i(y-\hat y_k)}\hat\phi(2^{-k}(y-\hat y_k))$$

$$=2\pi\hat f(y)\psi_k(y).$$

On the other hand, if $y\in\mathbb{R}\setminus\hat J^l_k$, $\psi_k(y)=\hat\phi_\sigma(y)=0$ for every $\sigma\in R_k$, so again $\sum_{\sigma\in R_k}(f|\phi_\sigma)\hat\phi_\sigma(y)=2\pi\hat f(y)\psi_k(y)$.

Next,

$$\textstyle\sum_{\sigma\in R_k}|(f|\phi_\sigma)|=2\pi\cdot 2^{k/2}\sum_{n=-\infty}^{\infty}|c_n|$$

and

$$\sup_{\sigma\in R_k}\int_{-\infty}^{\infty}|\hat{\phi}_\sigma| = 2^{k/2}\int_{-\infty}^{\infty}|\hat{\phi}|$$

are finite. So $\sum_{\sigma\in R_k}\int|(f|\phi_\sigma)\hat{\phi}_\sigma|$ is finite. Accordingly, for any $x\in\mathbb{R}$,

$$\begin{aligned}2\pi(\hat{f}\times\psi_k)^\vee(x) &= \bigl(\sum_{\sigma\in R_k}(f|\phi_\sigma)\hat{\phi}_\sigma\bigr)^\vee(x)\\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\sum_{\sigma\in R_k}(f|\phi_\sigma)\hat{\phi}_\sigma(y)e^{ixy}dy\\ &= \frac{1}{\sqrt{2\pi}}\sum_{\sigma\in R_k}\int_{-\infty}^{\infty}(f|\phi_\sigma)\hat{\phi}_\sigma(y)e^{ixy}dy\end{aligned}$$

(226E)

$$= \frac{1}{\sqrt{2\pi}}\sum_{\sigma\in R_k}(f|\phi_\sigma)(\hat{\phi}_\sigma)^\vee(x) = \frac{1}{\sqrt{2\pi}}\sum_{\sigma\in R_k}(f|\phi_\sigma)\phi_\sigma(x)$$

by 284C. **Q**

(c) Because every ψ_k is non-negative, $\theta_z=\sum_{k\in M}\psi_k$ is bounded above by 1, and $\hat{f}$ is integrable,

$$\begin{aligned}2\pi(\hat{f}\times\theta_z)^\vee(x) &= 2\pi\cdot\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{ixy}\hat{f}(y)\theta_z(y)dy\\ &= \sqrt{2\pi}\int_{-\infty}^{\infty}\sum_{k\in M}e^{ixy}\hat{f}(y)\psi_k(y)dy\\ &= \sqrt{2\pi}\sum_{k\in M}\int_{-\infty}^{\infty}e^{ixy}\hat{f}(y)\psi_k(y)dy\end{aligned}$$

(by 226E again)

$$= 2\pi\sum_{k\in M}(\hat{f}\times\psi_k)^\vee(x) = \sum_{k\in M}\sum_{\sigma\in R_k}(f|\phi_\sigma)\phi_\sigma(x),$$

and

$$2\pi|(\hat{f}\times\theta_z)^\vee(x)| \le \sum_{k\in M}\sum_{\sigma\in R_k}|(f|\phi_\sigma)\phi_\sigma(x)| \le (Af)(x).$$

286Q Lemma For $\alpha>0$ and y, z, $\beta\in\mathbb{R}$, set $\theta'_{z\alpha\beta}(y)=\theta_{\alpha z+\beta}(\alpha y+\beta)$. Then
(a) the function $(\alpha,\beta,y,z)\mapsto\theta'_{z\alpha\beta}(y):\,]0,\infty[\,\times\mathbb{R}^3\to[0,1]$ is Borel measurable;
(b) for any rapidly decreasing test function f,

$$2\pi|(\hat{f}\times\theta'_{z\alpha\beta})^\vee| \le D_{1/\alpha}A(M_\beta D_\alpha f)$$

(in the notation of 286C) at every point.

proof (a) We need only observe that $(y,z)\mapsto\theta_z(y):\mathbb{R}^2\to\mathbb{R}$ is Borel measurable, and that $(\alpha,\beta,y,z)\mapsto\theta'_{z\alpha\beta}(y)$ is built up from this, + and ×.

(b) Set $v=\alpha z+\beta$, so that $\theta'_{z\alpha\beta}=D_\alpha S_\beta\theta_v$. Then

$$\begin{aligned}\hat{f}\times\theta'_{z\alpha\beta} &= \hat{f}\times D_\alpha S_\beta\theta_v = D_\alpha S_\beta(S_{-\beta}D_{1/\alpha}\hat{f}\times\theta_v)\\ &= \alpha D_\alpha S_\beta(S_{-\beta}(D_\alpha f)^\wedge\times\theta_v) = \alpha D_\alpha S_\beta((M_\beta D_\alpha f)^\wedge\times\theta_v),\end{aligned}$$

so

$$\begin{aligned}(\hat{f}\times\theta'_{z\alpha\beta})^\vee &= \alpha\bigl(D_\alpha S_\beta((M_\beta D_\alpha f)^\wedge\times\theta_v)\bigr)^\vee\\ &= D_{1/\alpha}\bigl(S_\beta((M_\beta D_\alpha f)^\wedge\times\theta_v)\bigr)^\vee\\ &= D_{1/\alpha}M_{-\beta}\bigl((M_\beta D_\alpha f)^\wedge\times\theta_v\bigr)^\vee\end{aligned}$$

and

$$2\pi|(\hat{f} \times \theta'_{z\alpha\beta})^\vee| = 2\pi D_{1/\alpha}\big|((M_\beta D_\alpha f)^\wedge \times \theta_v)^\vee\big| \le D_{1/\alpha}A(M_\beta D_\alpha f)$$

by 286P.

286R Lemma For any y, $z \in \mathbb{R}$,

$$\tilde{\theta}_z(y) = \int_1^2 \frac{1}{\alpha}\Big(\lim_{n\to\infty}\frac{1}{n}\int_0^n \theta'_{z\alpha\beta}(y)d\beta\Big)d\alpha$$

is defined, and

$$\begin{aligned}\tilde{\theta}_z(y) &= \tilde{\theta}_1(0) > 0 \text{ if } y < z,\\ &= 0 \text{ if } y \ge z.\end{aligned}$$

proof (a) The case $y \ge z$ is trivial; because if $y \ge z$ then $\alpha y + \beta \ge \alpha z + \beta$ for all $\alpha > 0$ and $\beta \in \mathbb{R}$, so that $\theta'_{z\alpha\beta}(y) = 0$ for every $\alpha > 0$, $\beta \in \mathbb{R}$ and $\tilde{\theta}_z(y) = 0$. For the rest of the proof, therefore, I look at the case $y < z$.

(b)(i) Given $y < z \in \mathbb{R}$ and $\alpha > 0$, set $l = \lfloor \log_2(20\alpha(z-y))\rfloor$. Then $\theta'_{z,\alpha,\beta+2^l}(y) = \theta'_{z\alpha\beta}(y)$ for every $\beta \in \mathbb{R}$. **P** If $\theta'_{z\alpha\beta}(y) = \theta_{\alpha z+\beta}(\alpha y + \beta)$ is non-zero, there must be k, $m \in \mathbb{Z}$ such that

$$2^k(m+\frac{1}{2}) \le \alpha z + \beta < 2^k(m+1)$$

and

$$\hat{\phi}(2^{-k}(\alpha y + \beta) - (m+\tfrac{1}{4}))^2 = \theta'_{z\alpha\beta}(y) \ne 0,$$

so

$$2^k m \le \alpha y + \beta \le 2^k(m+\frac{9}{20})$$

because $\hat{\phi}$ is zero outside $[-\frac{1}{5}, \frac{1}{5}]$. In this case, $\frac{1}{20}\cdot 2^k < \alpha(z-y)$, so that $k \le l$. We therefore have

$$2^k(m+2^{l-k}+\frac{1}{2}) \le \alpha z + \beta + 2^l < 2^k(m+2^{l-k}+1),$$

$$2^k(m+2^{l-k}) \le \alpha y + \beta + 2^l < 2^k(m+2^{l-k}+\frac{1}{2}),$$

so

$$\theta'_{z,\alpha,\beta+2^l}(y) = \hat{\phi}(2^{-k}(\alpha y + \beta + 2^l) - (m+2^{l-k}+\frac{1}{4}))^2 = \theta'_{z\alpha\beta}(y).$$

Similarly,

$$2^k(m-2^{l-k}+\frac{1}{2}) \le \alpha z + \beta - 2^l < 2^k(m-2^{l-k}+1),$$

$$2^k(m-2^{l-k}) \le \alpha y + \beta - 2^l < 2^k(m-2^{l-k}+\frac{1}{2}),$$

so

$$\theta'_{z,\alpha,\beta-2^l}(y) = \hat{\phi}(2^{-k}(\alpha y + \beta - 2^l) - (m-2^{l-k}+\frac{1}{4}))^2 = \theta'_{z\alpha\beta}(y).$$

What this shows is that $\theta'_{z,\alpha,\beta+2^l}(y) = \theta'_{z\alpha\beta}(y)$ if either is non-zero, so we have the equality in any case. **Q**

(ii) It follows that $g(\alpha, y, z) = \lim_{b\to\infty}\frac{1}{b}\int_0^b \theta'_{z\alpha\beta}(y)d\beta$ is defined. **P** Set

$$\gamma = 2^{-l}\int_0^{2^l} \theta'_{z\alpha\beta}(y)d\beta.$$

From (i) we see that

$$\gamma = 2^{-l}\int_{2^l m}^{2^l(m+1)} \theta'_{z\alpha\beta}(y)d\beta$$

for every $m \in \mathbb{Z}$, and therefore that

$$\gamma = \frac{1}{2^l m}\int_0^{2^l m} \theta'_{z\alpha\beta}(y)d\beta$$

for every $m \ge 1$. Now $\theta_{z\alpha\beta}(y)$ is always greater than or equal to 0, so if $2^l m \le b \le 2^l(m+1)$ then

$$\begin{aligned}\frac{m}{m+1}\gamma = \frac{1}{2^l(m+1)}\int_0^{2^l m}\theta'_{z\alpha\beta} &\le \frac{1}{b}\int_0^b \theta'_{z\alpha\beta}\\ &\le \frac{1}{2^l m}\int_0^{2^l(m+1)}\theta'_{z\alpha\beta} = \frac{m+1}{m}\gamma,\end{aligned}$$

which approach γ as $b \to \infty$. **Q**

(c) Because $(\alpha,\beta) \mapsto \theta'_{z\alpha\beta}(y)$ is Borel measurable, each of the functions $\alpha \mapsto \frac{1}{n}\int_0^n \theta'_{z\alpha\beta}$, for $n \ge 1$, is Borel measurable (putting 251M and 252P together), and $\alpha \mapsto g(\alpha,y,z) :]0,\infty[\to \mathbb{R}$ is Borel measurable; at the same time, since $0 \le \theta'_{z\alpha\beta}(y) \le 1$ for all α and β, $0 \le g(\alpha,y,z) \le 1$ for every α, and $\tilde\theta_z(y) = \int_1^2 \frac{1}{\alpha}g(\alpha,y,z)d\alpha$ is defined in $[0,1]$.

(d) For any $y < z$, $\gamma \in \mathbb{R}$ and $\alpha > 0$, $g(\alpha,y+\gamma,z+\gamma) = g(\alpha,y,z)$. **P** It is enough to consider the case $\gamma \ge 0$. In this case

$$\begin{aligned}g(\alpha,y+\gamma,z+\gamma) &= \lim_{b\to\infty}\frac{1}{b}\int_0^b \theta'_{z+\gamma,\alpha,\beta}(y+\gamma)d\beta\\ &= \lim_{b\to\infty}\frac{1}{b}\int_0^b \theta_{\alpha z+\alpha\gamma+\beta}(\alpha y+\alpha\gamma+\beta)d\beta\\ &= \lim_{b\to\infty}\frac{1}{b}\int_{\alpha\gamma}^{b+\alpha\gamma} \theta_{\alpha z+\beta}(\alpha y+\beta)d\beta\\ &= \lim_{b\to\infty}\frac{1}{b}\int_{\alpha\gamma}^{b+\alpha\gamma} \theta'_{z\alpha\beta}(y)d\beta,\end{aligned}$$

so

$$\begin{aligned}|g(\alpha,y+\gamma,z+\gamma) - g(\alpha,y,z)| &= \lim_{b\to\infty}\frac{1}{b}\Bigl|\int_b^{b+\alpha\gamma}\theta'_{z\alpha\beta}(y)d\beta - \int_0^{\alpha\gamma}\theta'_{z\alpha\beta}(y)d\beta\Bigr|\\ &\le \lim_{b\to\infty}\frac{2\alpha\gamma}{b} = 0. \textbf{ Q}\end{aligned}$$

It follows that whenever $y < z$ and $\gamma \in \mathbb{R}$,

$$\tilde\theta_{z+\gamma}(y+\gamma) = \int_0^1 \frac{1}{\alpha}g(\alpha,y+\gamma,z+\gamma)d\alpha = \int_0^1 \frac{1}{\alpha}g(\alpha,y,z)d\alpha = \tilde\theta_z(y).$$

(e) The next essential fact to note is that $\theta_{2z}(2y)$ is always equal to $\theta_z(y)$. **P** If $\theta_z(y) \ne 0$, then (as in (b) above) there are k, $m \in \mathbb{Z}$ such that

$$2^k(m+\tfrac{1}{2}) \le z < 2^k(m+1), \quad 2^k m \le y \le 2^k(m+\tfrac{1}{2}), \quad \theta_z(y) = \hat\phi(2^{-k}y-(m+\tfrac{1}{4}))^2.$$

In this case,

$$2^{k+1}(m+\tfrac{1}{2}) \le 2z < 2^{k+1}(m+1), \quad 2^{k+1}m \le 2y \le 2^{k+1}(m+\tfrac{1}{2}),$$

so

$$\theta_{2z}(2y) = \hat\phi(2^{-k-1}\cdot 2y-(m+\tfrac{1}{4}))^2 = \theta_z(y).$$

Similarly,

$$2^{k-1}(m+\tfrac{1}{2}) \le \tfrac{1}{2}z < 2^{k-1}(m+1), \quad 2^{k-1}m \le \tfrac{1}{2}y \le 2^{k-1}(m+\tfrac{1}{2}),$$

so

$$\theta_{\frac{1}{2}z}(\frac{1}{2}y) = \hat{\phi}(2^{-k+1} \cdot \frac{1}{2}y - (m + \frac{1}{4}))^2 = \theta_z(y).$$

This shows that $\theta_{2z}(2y) = \theta_z(y)$ if either is non-zero, and therefore in all cases. **Q**

Accordingly

$$\theta'_{z,2\alpha,2\beta}(y) = \theta_{2\alpha z+2\beta}(2\alpha y + 2\beta) = \theta_{\alpha z+\beta}(\alpha y + \beta) = \theta'_{z\alpha\beta}(y)$$

for all y, z, $\beta \in \mathbb{R}$ and all $\alpha > 0$.

(f) Consequently

$$\begin{aligned} g(2\alpha, y, z) &= \lim_{b\to\infty} \frac{1}{b} \int_0^b \theta'_{z,2\alpha,\beta}(y)d\beta = \lim_{b\to\infty} \frac{2}{b} \int_0^{b/2} \theta'_{z,2\alpha,2\beta}(y)d\beta \\ &= \lim_{b\to\infty} \frac{2}{b} \int_0^{b/2} \theta'_{z\alpha\beta}(y)d\beta = \lim_{b\to\infty} \frac{1}{b} \int_0^b \theta'_{z\alpha\beta}(y)d\beta = g(\alpha, y, z) \end{aligned}$$

whenever $\alpha > 0$ and y, $z \in \mathbb{R}$. It follows that

$$\int_\gamma^\delta \frac{1}{\alpha} g(\alpha, y, z)d\alpha = \int_\gamma^\delta \frac{1}{\alpha} g(2\alpha, y, z)d\alpha = \int_{2\gamma}^{2\delta} \frac{1}{\alpha} g(\alpha, y, z)d\alpha$$

whenever $0 < \gamma \le \delta$, and therefore that

$$\int_\gamma^{2\gamma} \frac{1}{\alpha} g(\alpha, y, z)d\alpha = \int_1^2 \frac{1}{\alpha} g(\alpha, y, z)d\alpha$$

for every $\gamma > 0$. **P** Take $k \in \mathbb{Z}$ such that $2^k \le \gamma < 2^{k+1}$. Then

$$\begin{aligned} \int_\gamma^{2\gamma} \frac{1}{\alpha} g(\alpha, y, z)d\alpha &= \int_{2^k}^{2^{k+1}} \frac{1}{\alpha} g(\alpha, y, z)d\alpha - \int_{2^k}^{\gamma} \frac{1}{\alpha} g(\alpha, y, z)d\alpha + \int_{2^{k+1}}^{2\gamma} \frac{1}{\alpha} g(\alpha, y, z)d\alpha \\ &= \int_{2^k}^{2^{k+1}} \frac{1}{\alpha} g(\alpha, y, z)d\alpha = \int_1^2 \frac{1}{\alpha} g(\alpha, y, z)d\alpha. \text{ } \mathbf{Q} \end{aligned}$$

(g) Now if α, $\gamma > 0$ and $y < z$,

$$g(\alpha, \gamma y, \gamma z) = \lim_{b\to\infty} \frac{1}{b} \int_0^b \theta_{\alpha\gamma z+\beta}(\alpha\gamma y + \beta)d\beta = g(\alpha\gamma, y, z).$$

So if $\gamma > 0$ and $y < z$,

$$\begin{aligned} \tilde{\theta}_{\gamma z}(\gamma y) &= \int_1^2 \frac{1}{\alpha} g(\alpha, \gamma y, \gamma z)d\alpha = \int_1^2 \frac{1}{\alpha} g(\alpha\gamma, y, z)d\alpha \\ &= \int_\gamma^{2\gamma} \frac{1}{\alpha} g(\alpha, y, z)d\alpha = \int_1^2 \frac{1}{\alpha} g(\alpha, y, z)d\alpha = \tilde{\theta}_z(y). \end{aligned}$$

Putting this together with (d), we see that if $y < z$ then

$$\tilde{\theta}_z(y) = \tilde{\theta}_{z-y}(0) = \tilde{\theta}_1(0).$$

(h) I have still to check that $\tilde{\theta}_1(0)$ is not zero. But suppose that $1 \le \alpha < \frac{7}{6}$ and that there is some $m \in \mathbb{Z}$ such that $2(m + \frac{1}{12}) \le \beta \le 2(m + \frac{5}{12})$. Then $2(m + \frac{1}{2}) \le \alpha + \beta < 2(m + 1)$, while $|\frac{1}{2}\beta - (m + \frac{1}{4})| \le \frac{1}{6}$, so

$$\theta_{\alpha+\beta}(\beta) = \hat{\phi}(\frac{1}{2}\beta - (m + \frac{1}{4}))^2 = 1.$$

What this means is that, for $1 \le \alpha < \frac{7}{6}$,

$$\begin{aligned} g(\alpha, 1, 0) &= \lim_{m\to\infty} \frac{1}{2m} \int_0^{2m} \theta_{\alpha+\beta}(\beta)d\beta \\ &\ge \lim_{m\to\infty} \frac{1}{2m} \sum_{j=0}^{m-1} \mu[2(j + \frac{1}{12}), 2(j + \frac{5}{12})] = \frac{1}{3}. \end{aligned}$$

So

$$\tilde{\theta}_1(0)=\int_1^2\frac{1}{\alpha}g(\alpha,1,0)d\alpha\ge\frac{1}{3}\int_1^{7/6}\frac{1}{\alpha}d\alpha>0.$$

This completes the proof.

286S Lemma Suppose that $f\in\mathcal{L}^2_{\mathbb{C}}$.

(a) For every $x\in\mathbb{R}$,

$$(\tilde{A}f)(x)=\liminf_{n\to\infty}\frac{1}{n}\int_1^2\frac{1}{\alpha}\int_0^n(D_{1/\alpha}AM_\beta D_\alpha f)(x)d\beta d\alpha$$

is defined in $[0,\infty]$, and $\tilde{A}f:\mathbb{R}\to[0,\infty]$ is Borel measurable.

(b) $\int_F\tilde{A}f\le C_9\|f\|_2\sqrt{\mu F}$ whenever $\mu F<\infty$.

(c) If f is a rapidly decreasing test function and $z\in\mathbb{R}$, $2\pi|(\hat{f}\times\tilde{\theta}_z)^\vee|\le\tilde{A}f$ at every point.

proof (a) The point here is that the function

$$(\alpha,\beta,x)\mapsto(D_{1/\alpha}AM_\beta D_\alpha f)(x):\,]0,\infty[\,\times\mathbb{R}^2\to[0,\infty]$$

is Borel measurable. **P**

$$\begin{aligned}(D_{1/\alpha}AM_\beta D_\alpha f)(x)&=(AM_\beta D_\alpha f)(\frac{x}{\alpha})\\&=\sup_{z\in\mathbb{R},P\subseteq Q\text{ is finite}}\sum_{\sigma\in P,z\in J^r_\sigma}|(M_\beta D_\alpha f|\phi_\sigma)\phi_\sigma(\frac{x}{\alpha})|.\end{aligned}$$

Look at the central term in this formula. For any $\sigma\in Q$, we have

$$\begin{aligned}(M_\beta D_\alpha f|\phi_\sigma)&=\int_{-\infty}^{\infty}e^{i\beta t}f(\alpha t)\overline{\phi_\sigma(t)}dt\\&=\frac{1}{\alpha}\int_{-\infty}^{\infty}e^{i\beta t/\alpha}f(t)\overline{\phi_\sigma(t/\alpha)}dt.\end{aligned}$$

Now ϕ_σ is a rapidly decreasing test function, so there is some $\gamma\ge0$ such that $|\phi_\sigma(t)|\le\gamma/(1+t^2)$ for every $t\in\mathbb{R}$. This means that if $\alpha>0$ and $\langle\alpha_n\rangle_{n\in\mathbb{N}}$ is a sequence in $\left[\frac{1}{2}\alpha,\infty\right[$ and we set $g(t)=\sup_{n\in\mathbb{N}}|\phi_\sigma(t/\alpha_n)|$ for $t\in\mathbb{R}$, then $g(t)\le4\gamma/(4+t^2)$ for every t and g is integrable. So Lebesgue's Dominated Convergence Theorem tells us that if $\langle\alpha_n\rangle_{n\in\mathbb{N}}\to\alpha$ and $\langle\beta_n\rangle_{n\in\mathbb{N}}\to\beta$,

$$\frac{1}{\alpha_n}\int_{-\infty}^{\infty}e^{i\beta_nt/\alpha_n}f(t)\overline{\phi_\sigma(t/\alpha_n)}dt\to\frac{1}{\alpha}\int_{-\infty}^{\infty}e^{i\beta t/\alpha}f(t)\overline{\phi_\sigma(t/\alpha)}dt.$$

Thus

$$(\alpha,\beta)\mapsto(M_\beta D_\alpha f|\phi_\sigma):\,]0,\infty[\,\times\mathbb{R}\to\mathbb{R}$$

is continuous; and this is true for every $\sigma\in Q$.

Accordingly

$$(\alpha,\beta,x)\mapsto\sum_{\sigma\in P,z\in J^r_\sigma}|(M_\beta D_\alpha f|\phi_\sigma)\phi_\sigma(\frac{x}{\alpha})|$$

is continuous for every $z\in\mathbb{R}$ and every finite $P\subseteq Q$, and $(\alpha,\beta,x)\mapsto(D_{1/\alpha}AM_\beta D_\alpha f)(x)$ is Borel measurable by 256Ma. **Q**

It follows that the repeated integrals

$$\int_1^2\frac{1}{\alpha}\int_0^n(D_{1/\alpha}AM_\beta D_\alpha f)(x)d\beta d\alpha$$

are defined in $[0,\infty]$ and are Borel measurable functions of x (252P again), so that $\tilde{A}f$ is Borel measurable.

(b) For any $n\in\mathbb{N}$,

$$\begin{aligned}\int_F\frac{1}{n}\int_1^2\frac{1}{\alpha}\int_0^n&(D_{1/\alpha}AM_\beta D_\alpha f)(x)d\beta d\alpha dx\\&=\frac{1}{n}\int_1^2\frac{1}{\alpha}\int_0^n\int_F(D_{1/\alpha}AM_\beta D_\alpha f)(x)dxd\beta d\alpha\end{aligned}$$

(by Fubini's theorem, 252H)

$$= \frac{1}{n}\int_1^2\int_0^n\int_F \frac{1}{\alpha}(AM_\beta D_\alpha f)(\frac{x}{\alpha})dxd\beta d\alpha$$

$$= \frac{1}{n}\int_1^2\int_0^n\int_{\alpha^{-1}F}(AM_\beta D_\alpha f)(x)dxd\beta d\alpha$$

$$\le \frac{1}{n}\int_1^2\int_0^n C_9\|M_\beta D_\alpha f\|_2\sqrt{\mu(\alpha^{-1}F)}d\beta d\alpha$$

(286O)

$$= C_9\cdot\frac{1}{n}\int_1^2\int_0^n \frac{1}{\sqrt{\alpha}}\|f\|_2\cdot\frac{1}{\sqrt{\alpha}}\sqrt{\mu F}d\beta d\alpha$$

$$= C_9\|f\|_2\sqrt{\mu F}\cdot\frac{1}{n}\int_1^2\frac{1}{\alpha}\int_0^n d\beta d\alpha$$

$$= C_9\|f\|_2\sqrt{\mu F}\ln 2 \le C_9\|f\|_2\sqrt{\mu F}.$$

So

$$\int_F \tilde{A}f = \int_F \liminf_{n\to\infty}\frac{1}{n}\int_1^2\frac{1}{\alpha}\int_0^n(D_{1/\alpha}AM_\beta D_\alpha f)(x)d\beta d\alpha dx$$

$$\le \liminf_{n\to\infty}\int_F\frac{1}{n}\int_1^2\frac{1}{\alpha}\int_0^n(D_{1/\alpha}AM_\beta D_\alpha f)(x)d\beta d\alpha dx$$

(by Fatou's lemma)

$$\le C_9\|f\|_2\sqrt{\mu F}.$$

(c) For any $x\in\mathbb{R}$,

$$\int_{-\infty}^\infty|\hat{f}(y)|\int_1^2\frac{1}{\alpha}\Bigl(\sup_{n\in\mathbb{N}}\frac{1}{n}\int_0^n\theta'_{z\alpha\beta}(y)d\beta\Bigr)d\alpha dy \le \ln 2\cdot\int_{-\infty}^\infty|\hat{f}|$$

is finite. So

$$(\hat{f}\times\tilde{\theta}_z)^\vee(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{ixy}\hat{f}(y)\tilde{\theta}_z(y)dy$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty e^{ixy}\hat{f}(y)\int_1^2\frac{1}{\alpha}\lim_{n\to\infty}\frac{1}{n}\int_0^n\theta'_{z\alpha\beta}(y)d\beta d\alpha dy$$

$$= \frac{1}{\sqrt{2\pi}}\lim_{n\to\infty}\int_{-\infty}^\infty e^{ixy}\hat{f}(y)\int_1^2\frac{1}{\alpha n}\int_0^n\theta'_{z\alpha\beta}(y)d\beta d\alpha dy$$

(by Lebesgue's Dominated Convergence Theorem)

$$= \frac{1}{\sqrt{2\pi}}\lim_{n\to\infty}\int_1^2\frac{1}{\alpha n}\int_0^n\int_{-\infty}^\infty e^{ixy}\hat{f}(y)\theta'_{z\alpha\beta}(y)dyd\beta d\alpha$$

(by Fubini's theorem)

$$= \lim_{n\to\infty}\int_1^2\frac{1}{\alpha n}\int_0^n(\hat{f}\times\theta'_{z\alpha\beta})^\vee(x)d\beta d\alpha,$$

and

$$2\pi|(\hat{f}\times\tilde{\theta}_z)^\vee(x)| = 2\pi\Bigl|\lim_{n\to\infty}\int_1^2\frac{1}{\alpha n}\int_0^n(\hat{f}\times\theta'_{z\alpha\beta})^\vee(x)d\beta d\alpha\Bigr|$$

$$\le 2\pi\liminf_{n\to\infty}\int_1^2\frac{1}{\alpha n}\int_0^n|(\hat{f}\times\theta'_{z\alpha\beta})^\vee(x)|d\beta d\alpha$$

$$\le \liminf_{n\to\infty}\int_1^2\frac{1}{\alpha n}\int_0^n(D_{1/\alpha}AM_\beta D_\alpha f)(x)d\beta d\alpha$$

(286Qb)

$$= (\tilde{A}f)(x).$$

286T Lemma Set $C_{10} = C_9/\pi\tilde{\theta}_1(0)$. For $f \in \mathcal{L}^2_{\mathbb{C}}$, define $\hat{A}f : \mathbb{R} \to [0,\infty]$ by setting

$$(\hat{A}f)(y) = \sup_{a\le b} \frac{1}{\sqrt{2\pi}} |\int_a^b e^{-ixy} f(x)dx|$$

for each $y \in \mathbb{R}$. Then $\int_F \hat{A}f \le C_{10}\|f\|_2\sqrt{\mu F}$ whenever $\mu F < \infty$.

proof (a) As usual, the first step is to confirm that $\hat{A}f$ is measurable. **P** For $a \le b$, $y \mapsto |\frac{1}{\sqrt{2\pi}}\int_a^b e^{-ixy}f(x)dx|$ is continuous (by 283Cf, since $f \times \chi[a,b]$ is integrable), so 256M gives the result. **Q**

(b) Suppose that f is a rapidly decreasing test function. Then

$$(\hat{A}f)(y) \le \frac{1}{\pi\tilde{\theta}_1(0)}(\tilde{A}\check{f})(-y)$$

for every $y \in \mathbb{R}$. **P** If $a \in \mathbb{R}$ then

$$\frac{1}{\sqrt{2\pi}}|\int_{-\infty}^a e^{-ixy}f(x)dx| = \frac{1}{\tilde{\theta}_1(0)\sqrt{2\pi}}|\int_{-\infty}^{\infty} e^{-ixy}\tilde{\theta}_a(x)f(x)dx|$$

(286R)

$$= \frac{1}{\tilde{\theta}_1(0)}|(f \times \tilde{\theta}_a)^\vee(-y)| = \frac{1}{\tilde{\theta}_1(0)}|(\overset{\vee\wedge}{f} \times \tilde{\theta}_a)^\vee(-y)|$$

(284C)

$$\le \frac{1}{2\pi\tilde{\theta}_1(0)}(\tilde{A}\check{f})(-y)$$

(286Sc). So if $a \le b$ in $\mathbb{R}$,

$$\frac{1}{\sqrt{2\pi}}|\int_a^b e^{-ixy}f(x)dx| \le \frac{1}{\pi\tilde{\theta}_1(0)}(\tilde{A}\check{f})(-y);$$

taking the supremum over a and b, we have the result. **Q**

It follows that

$$\int_F \hat{A}f \le \frac{1}{\pi\tilde{\theta}_1(0)}\int_{-F}\tilde{A}\check{f} \le \frac{1}{\pi\tilde{\theta}_1(0)}C_9\|f\|_2\sqrt{\mu(-F)}$$

(286Sb, 284Oa)

$$= C_{10}\|f\|_2\sqrt{\mu F}.$$

(c) For general square-integrable f, take any $\epsilon > 0$ and any $n \in \mathbb{N}$. Set

$$(\hat{A}_n f)(y) = \sup_{-n\le a\le b\le n} \frac{1}{\sqrt{2\pi}}|\int_a^b e^{-ixy}f(x)dx|$$

for each $y \in \mathbb{R}$. Let g be a rapidly decreasing test function such that $\|f - g\|_2 \le \epsilon$ (284N). Then

$$\hat{A}g \ge \hat{A}_n g \ge \hat{A}_n f - \frac{\sqrt{2n}}{\sqrt{2\pi}}\epsilon$$

(using Cauchy's inequality), so

$$\int_F \hat{A}_n f \le \int_F \hat{A}g + \sqrt{\frac{n}{\pi}}\epsilon\mu F \le C_{10}(\|f\|_2 + \epsilon)\sqrt{\mu F} + \sqrt{\frac{n}{\pi}}\epsilon\mu F.$$

As ϵ is arbitrary, $\int_F \hat{A}_n f \le C_{10}\|f\|_2\sqrt{\mu F}$; letting $n \to \infty$, we get $\int_F \hat{A}f \le C_{10}\|f\|_2\sqrt{\mu F}$.

286U Theorem If $f \in \mathcal{L}^2_{\mathbb{C}}$ then

$$g(y) = \lim_{a\to-\infty, b\to\infty} \frac{1}{\sqrt{2\pi}}\int_a^b e^{-ixy}f(x)dx$$

is defined in $\mathbb{C}$ for almost every $y \in \mathbb{R}$, and g represents the Fourier transform of f.

proof (a) For $n \in \mathbb{N}$, $y \in \mathbb{R}$ set

$$\gamma_n(y) = \sup_{a\le -n, b\ge n} \frac{1}{\sqrt{2\pi}}\Bigl|\int_a^b e^{-ixy}f(x)dx - \int_{-n}^n e^{-ixy}f(x)dx\Bigr|.$$

Then $g(y)$ is defined whenever $\inf_{n\in\mathbb{N}}\gamma_n(y)=0$. **P** If $\inf_{n\in\mathbb{N}}\gamma_n(y)=0$ and $\epsilon>0$, take $m\in\mathbb{N}$ such that $\gamma_m(y)\le\frac{1}{2}\epsilon$; then $\frac{1}{\sqrt{2\pi}}|\int_a^b e^{-ixy}f(x)dx - \int_{-n}^n e^{-ixy}f(x)dx|\le\epsilon$ whenever $n\ge m$ and $a\le -n$, $b\ge n$. But this means, first, that $\langle\int_{-n}^n e^{-ixy}f(x)dx\rangle_{n\in\mathbb{N}}$ is a Cauchy sequence, so has a limit ζ say, and, second, that $\zeta=\lim_{a\to-\infty,b\to\infty}\int_a^b e^{-ixy}f(x)dx$, so that $g(y)=\frac{\zeta}{\sqrt{2\pi}}$ is defined. **Q**

Also each γ_n is a measurable function (cf. part (a) of the proof of 286T).

(b) ? Suppose, if possible, that $\{y:\inf_{n\in\mathbb{N}}\gamma_n(y)>0\}$ is not negligible. Then

$$\lim_{m\to\infty}\mu\{y:|y|\le m,\ \inf_{n\in\mathbb{N}}\gamma_n(y)\ge\frac{1}{m}\}>0,$$

so there is an $\epsilon>0$ such that

$$F=\{y:|y|\le\frac{1}{\epsilon},\ \inf_{n\in\mathbb{N}}\gamma_n(y)\ge\epsilon\}$$

has measure greater than ϵ. Let $n\in\mathbb{N}$ be such that

$$4C_{10}^2\Bigl(\int_{-\infty}^{\infty}|f(x)|^2dx - \int_{-n}^n|f(x)|^2dx\Bigr)<\epsilon^3,$$

and set $f_1=f-f\times\chi[-n,n]$; then $2C_{10}\|f_1\|_2\le\epsilon^{3/2}$.

We have

$$\begin{aligned}\gamma_n(y)&=\sup_{a\le -n,b\ge n}\frac{1}{\sqrt{2\pi}}\Bigl|\int_a^b e^{-ixy}f_1(x)dx-\int_{-n}^n e^{-ixy}f_1(x)dx\Bigr|\\&\le 2\sup_{a\le b}\frac{1}{\sqrt{2\pi}}\Bigl|\int_a^b e^{-ixy}f_1(x)dx\Bigr|\le 2(\hat{A}f_1)(y),\end{aligned}$$

so that

$$\begin{aligned}\epsilon\mu F&\le\int_F\gamma_n\le 2\int_F\hat{A}f_1\le 2C_{10}\|f_1\|_2\sqrt{\mu F}\\&\le\epsilon^{3/2}\sqrt{\mu F}\end{aligned}$$

(286T)

and $\mu F\le\epsilon$; but we chose ϵ so that μF would be greater than ϵ. **X**

(c) Thus $g(y)$ is defined for almost every $y\in\mathbb{R}$. Now g represents the Fourier transform of f. **P** Let h be a rapidly decreasing test function. Then the restriction of $\hat{A}f$ to the set on which it is finite is a tempered function, by 286D, so $\int_{-\infty}^{\infty}(\hat{A}f)(y)|h|$ is finite, by 284F. Now

$$\begin{aligned}\int_{-\infty}^{\infty}g\times h&=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\bigl(\lim_{n\to\infty}\int_{-n}^n e^{-ixy}f(x)dx\bigr)h(y)dy\\&=\frac{1}{\sqrt{2\pi}}\lim_{n\to\infty}\int_{-\infty}^{\infty}\int_{-n}^n e^{-ixy}f(x)h(y)dxdy\end{aligned}$$

(because $\frac{1}{\sqrt{2\pi}}|\int_{-n}^n e^{-ixy}f(x)dx|\le\hat{A}f(y)$ for every n and y, so we can use Lebesgue's Dominated Convergence Theorem)

$$=\frac{1}{\sqrt{2\pi}}\lim_{n\to\infty}\int_{-n}^n\int_{-\infty}^{\infty}e^{-ixy}f(x)h(y)dydx$$

(because $\int_{-\infty}^{\infty}\int_{-n}^n|f(x)h(y)|dxdy$ is finite for each n)

$$=\lim_{n\to\infty}\int_{-n}^n f\times\hat{h}=\int_{-\infty}^{\infty}f\times\hat{h}$$

because $f\times\hat{h}$ is certainly integrable. As h is arbitrary, g represents the Fourier transform of f. **Q**

286V Theorem For any square-integrable complex-valued function on $]-\pi,\pi]$, its sequence of Fourier sums converges to it almost everywhere.

proof Suppose that $f\in\mathcal{L}^2_{\mathbb{C}}(\mu_{]-\pi,\pi]})$. Set $f_1(x)=f(x)$ for $x\in\operatorname{dom}f$, 0 for $x\in\mathbb{R}\setminus\,]-\pi,\pi]$; then $f_1\in\mathcal{L}^2_{\mathbb{C}}(\mu)$. Let $g\in\mathcal{L}^2_{\mathbb{C}}(\mu)$ represent the inverse Fourier transform of f_1 (284O). Then 286U tells us that $f_2(x)=\lim_{a\to\infty}\frac{1}{\sqrt{2\pi}}\int_{-a}^a e^{-ixy}g(y)dy$ is defined for almost every x, and that f_2 represents the Fourier transform of g, so is equal almost everywhere to f_1 (284Ib).

Now, for any $a\ge 0$, $x\in\mathbb{R}$,

$$\int_{-a}^a e^{-ixy}g(y)dy=(g|h_{ax})$$

(where $h_{ax}(y)=e^{ixy}$ if $|y|\le a$, 0 otherwise)

$$=(f_2|\hat{h}_{ax})$$

(284Ob)

$$=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}f_2(t)\overline{\int_{-\infty}^{\infty}e^{-ity}h_{ax}(y)dy}\,dt$$

$$=\frac{2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{\sin(x-t)a}{x-t}f_2(t)dt=\frac{2}{\sqrt{2\pi}}\int_{-\pi}^{\pi}\frac{\sin(x-t)a}{x-t}f(t)dt.$$

So

$$f(x)=f_2(x)=\lim_{a\to\infty}\frac{1}{\pi}\int_{-\pi}^{\pi}\frac{\sin(x-t)a}{x-t}f(t)dt$$

for almost every $x\in\,]-\pi,\pi]$.

On the other hand, writing $\langle s_n\rangle_{n\in\mathbb{N}}$ for the sequence of Fourier sums of f, we have, for any $x\in\,]-\pi,\pi[$,

$$s_n(x)=\frac{1}{2\pi}\int_{-\pi}^{\pi}f(t)\frac{\sin(n+\frac{1}{2})(x-t)}{\sin\frac{1}{2}(x-t)}dt$$

for each n, by 282Da. Now

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}f(t)\frac{\sin(n+\frac{1}{2})(x-t)}{\sin\frac{1}{2}(x-t)}dt-\frac{1}{\pi}\int_{-\pi}^{\pi}f(t)\frac{\sin(n+\frac{1}{2})(x-t)}{x-t}dt$$

$$=\frac{1}{\pi}\int_{-\pi}^{\pi}f(t)\Bigl(\frac{\sin(n+\frac{1}{2})(x-t)}{2\sin\frac{1}{2}(x-t)}dt-\frac{\sin(n+\frac{1}{2})(x-t)}{x-t}\Bigr)dt$$

$$=\frac{1}{\pi}\int_{x-\pi}^{x+\pi}f(x-t)\sin(n+\tfrac{1}{2})t\Bigl(\frac{1}{2\sin\frac{1}{2}t}-\frac{1}{t}\Bigr)dt.$$

But if we look at the function

$$p_x(t)=f(x-t)\Bigl(\frac{1}{2\sin\frac{1}{2}t}-\frac{1}{t}\Bigr)\text{ if }x-\pi<t<x+\pi\text{ and }t\ne 0,$$

$$=0\text{ otherwise,}$$

p_x is integrable, because f is integrable over $]-\pi,\pi]$ and $\lim_{t\to 0}\frac{1}{2\sin\frac{1}{2}t}-\frac{1}{t}=0$, so $\sup_{t\ne 0,x-\pi\le t\le x+\pi}|\frac{1}{2\sin\frac{1}{2}t}-\frac{1}{t}|$ is finite. (This is where we need to know that $|x|<\pi$.) So

$$\lim_{n\to\infty}s_n(x)-\frac{1}{\pi}\int_{-\pi}^{\pi}f(t)\frac{\sin(n+\frac{1}{2})(x-t)}{x-t}dt=\lim_{n\to\infty}\int_{-\infty}^{\infty}p_x(t)\sin(n+\tfrac{1}{2})t\,dt=0$$

by the Riemann-Lebesgue lemma (282Fb). But this means that $\lim_{n\to\infty}s_n(x)=f(x)$ for any $x\in\,]-\pi,\pi[$ such that $f(x)=\lim_{a\to\infty}\frac{1}{\pi}\int_{-\pi}^{\pi}\frac{\sin(x-t)a}{x-t}f(t)dt$, which is almost every $x\in\,]-\pi,\pi]$.

286W Glossary The following special notations are used in more than one paragraph of this section:

μ for Lebesgue measure on $\mathbb{R}$.
286A: f^*.
286C: $S_\alpha f$, $M_\alpha f$, $D_\alpha f$.
286Ea: $\mathcal{I}$, Q, I_σ, J_σ, k_σ, x_σ, y_σ, J^l_σ, J^r_σ, y^l_σ.
286Eb: ϕ, ϕ_σ, $(f|g)$.
286Ec: w, w_σ, C_1.
286F: $\le$, $\le_r$, $\preccurlyeq$.
286G: C_2, C_3, C_4.
286H: mass, energy.
286I: C_5.
286K: C_6.
286L: C_7.
286M: C_8.
286N: C_9.
286O: Af.
286P: $\theta_z(y)$.
286Q: $\theta'_{z\alpha\beta}(y)$.
286R: $\tilde\theta_z(y)$.
286S: $\tilde A f$.
286T: C_{10}, $\hat A f$.

286X Basic exercises (a) Use 284Oa and 284Xf to shorten part (c) of the proof of 286U.

(b) Show that if $\langle c_k\rangle_{k\in\mathbb{N}}$ is a sequence of complex numbers such that $\sum_{k=0}^\infty |c_k|^2$ is finite, then $\sum_{k=0}^\infty c_k e^{ikx}$ is defined in $\mathbb{C}$ for almost all $x\in\mathbb{R}$.

286Y Further exercises (a) Show that if f is a square-integrable function on $\mathbb{R}^r$, where $r\ge 2$, then

$$g(y)=\frac{1}{(\sqrt{2\pi})^r}\lim_{\alpha_1,\dots,\alpha_r\to-\infty,\beta_1,\dots,\beta_r\to\infty}\int_a^b e^{-iy\cdot x}f(x)dx$$

is defined in $\mathbb{C}$ for almost every $y\in\mathbb{R}^r$, and that g represents the Fourier transform of f.

286 Notes and comments This is not the longest single section in this treatise as a whole, but it is by a substantial margin the longest in the present volume, and thirty pages of sub-superscripts must tax the endurance of the most enthusiastic. You will easily understand why Carleson's theorem is not usually presented at this level. But I am trying in this book to present complete proofs of the principal theorems, there is no natural place for Carleson's theorem in later volumes as at present conceived, and it is (just) accessible at this point; so I take the space to do it here.

The proof here divides naturally into two halves: the 'combinatorial' part in 286E-286M, up to the Lacey-Thiele lemma, followed by the 'analytic' part in 286N-286V, in which the averaging process

$$\int_1^2\frac{1}{\alpha}\lim_{b\to\infty}\frac{1}{b}\int_0^b\dots d\beta d\alpha$$

is used to transform the geometrically coherent, but analytically irregular, functions θ_z into the characteristic functions $\frac{1}{\tilde\theta_1(0)}\tilde\theta_z$. From the standpoint of ordinary Fourier analysis, this second part is essentially routine; there are many paths we could follow, and we have only to take the ordinary precautions against illegitimate operations.

Carleson (Carleson 66) stated his theorem in the Fourier-series form of 286V; but it had long been understood that this was equiveridical with the Fourier-transform version in 286U. There are of course many ways of extending the theorem. In particular, there are corresponding results for functions in $\mathcal{L}^p$ for any $p>1$, and even for functions f such that $f\times\ln(1+|f|)\times\ln\ln\ln(16+|f|)$ is integrable (Antonov 96). The methods here do not seem to reach so far. I ought also to remark that if we define $\hat Af$ as in 286T, then there is for every $p>1$ a constant C such that $\|\hat Af\|_p\le C\|f\|_p$ for every $f\in\mathcal{L}^p_{\mathbb{C}}$ (Hunt 67, Mozzochi 71, Jørsboe & Mejlbro 82, Arias de Reyna 02).

Note that the point of Carleson's theorem, in either form, is that we take special limits. In the formulae

$$\hat f(y)=\frac{1}{\sqrt{2\pi}}\lim_{a\to-\infty,b\to\infty}\int_a^b e^{-ixy}f(x)dx,$$

$$f(x)=\lim_{n\to\infty}\sum_{-n}^n c_k e^{ikx},$$

valid almost everywhere for square-integrable functions f, we are not taking the ordinary integral $\int_{-\infty}^\infty e^{-ixy}f(x)dx$ or the unconditional sum $\sum_{k\in\mathbb{Z}}c_k e^{ikx}$. If f is not integrable, or $\sum_{k=-\infty}^\infty|c_k|$ is infinite, these will not be defined at even one point. Carleson's theorem makes sense only because we have a natural preference for particular kinds of improper integral and conditional sum. So when we return, in Chapter 44 of Volume 4, to Fourier analysis on general topological groups, there will simply be no language in which to express the theorem, and while versions have been proved for other groups (e.g., Schipp 78), they necessarily depend on some structure beyond the simple notion of 'locally compact Hausdorff abelian topological group'. Even in $\mathbb{R}^2$, I understand that it is still unknown whether

$$\lim_{a\to\infty}\frac{1}{2\pi}\int_{B(\mathbf{0},a)}e^{-iy\cdot x}f(x)dx$$

will be defined a.e. for any square-integrable function f, if we use ordinary Euclidean balls $B(\mathbf{0},a)$ in place of the rectangles in 286Ya.

Appendix to Volume 2

Useful Facts

In the course of writing this volume, I have found that a considerable number of concepts and facts from various branches of mathematics are necessary to us. Nearly all of them are embedded in important and well-established theories for which many excellent textbooks are available and which I very much hope that you will one day study in depth. Nevertheless, I am reluctant to send you off immediately to courses in general topology, functional analysis and set theory, as if these were essential prerequisites for our work here, along with real analysis and basic linear algebra. For this reason I have written this Appendix, setting out those results which we actually need at some point in this volume. The great majority of them really are elementary – indeed, some are so elementary that they are not always spelt out in detail in orthodox treatments of their subjects.

While I do not put this book forward as the proper place to learn any of these topics, I have tried to set them out in a way that you will find easy to integrate into regular approaches. I do not expect anybody to read systematically through this work, and I hope that the references given in the main chapters of this volume will be adequate to guide you to the particular items you need.

2A1 Set theory

Especially for the examples in Chapter 21, we need some non-trivial set theory, which is best approached through the standard theory of cardinals and ordinals; and elsewhere in this volume I make use of Zorn's Lemma. Here I give a very brief outline of the results involved, largely omitting proofs. Most of this material should be in any sound introduction to set theory. The references I give are to books which happen to have come my way and which I can recommend as reasonably suitable for beginners.

I do not discuss axiom systems or logical foundations. The set theory I employ is 'naive' in the sense that I rely on my understanding of the collective experience of the last ninety years, rather than on any attempt at formal description, to distinguish legitimate from unsafe arguments. There are, however, points in Volume 5 at which such a relaxed philosophy becomes inappropriate, and I therefore use arguments which can, I believe, be translated into standard Zermelo-Fraenkel set theory without new ideas being invoked.

Although in this volume I use the axiom of choice without scruple whenever appropriate, I will divide this section into two parts, starting with ideas and results not dependent on the axiom of choice (2A1A-2A1I) and continuing with the remainder (2A1J-2A1P). I believe that even at this level it helps us to understand the nature of the arguments better if we maintain a degree of separation.

2A1A Ordered sets (a) Recall that a **partially ordered set** is a set P together with a relation $\leq$ on P such that

if $p \leq q$ and $q \leq r$ then $p \leq r$

$p \leq p$ for every $p \in P$

if $p \leq q$ and $q \leq p$ then $p = q$.

In this context, I will write $p \geq q$ to mean $q \leq p$, and $p < q$ or $q > p$ to mean '$p \leq q$ and $p \neq q$'. $\leq$ is a **partial order** on P.

(b) Let $(P, \leq)$ be a partially ordered set, and $A \subseteq P$. A **maximal** element of A is a $p \in A$ such that $p \not< a$ for any $a \in A$. Note that A may have more than one maximal element. An **upper bound** for A is a $p \in P$ such that $a \leq p$ for every $a \in A$; a **supremum** or **least upper bound** is an upper bound p such that $p \leq q$ for every upper bound q of A. There can be at most one such, because if p, p' are both least upper bounds then $p \leq p'$ and $p' \leq p$. Accordingly we may safely write $p = \sup A$ if p is the least upper bound of A.

Similarly, a **minimal** element of A is a $p \in A$ such that $p \not> a$ for every $a \in A$; a **lower bound** of A is a $p \in P$ such that $p \leq a$ for every $a \in A$; and $\inf A = a$ means that

$$\forall\ q \in P,\ a \geq q \iff p \geq q \text{ for every } p \in A.$$

A subset A of P is **order-bounded** if it has both an upper bound and a lower bound.

A subset A of P is **upwards-directed** if for any p, $p' \in A$ there is a $q \in A$ such that $p \leq q$ and $p' \leq q$; that is, if any non-empty finite subset of A has an upper bound in A. Similarly, $A \subseteq P$ is **downwards-directed** if for any p, $p' \in A$ there is a $q \in A$ such that $q \leq p$ and $q \leq p'$; that is, if any non-empty finite subset of A has a lower bound in A.

It is sometimes convenient to adapt the notation for closed intervals to arbitrary partially ordered sets: $[p, q]$ will be $\{r : p \leq r \leq q\}$.

(c) A **totally ordered set** is a partially ordered set $(P, \leq)$ such that
for any p, $q \in P$, either $p \leq q$ or $q \leq p$.
$\leq$ is then a **total** or **linear** order on P. In any totally ordered set we have a **median function**; for p, q, $r \in P$ set

$$\begin{aligned}\operatorname{med}(p, q, r) &= \max(\min(p, q), \min(p, r), \min(q, r)) \\ &= \min(\max(p, q), \max(p, r), \max(q, r)),\end{aligned}$$

so that $\operatorname{med}(p, q, r) = q$ if $p \leq q \leq r$.

(d) A **lattice** is a partially ordered set $(P, \leq)$ such that
for any p, $q \in P$, $p \vee q = \sup\{p, q\}$ and $p \wedge q = \inf\{p, q\}$ are defined in P.

(e) A **well-ordered set** is a totally ordered set $(P, \leq)$ such that $\inf A$ exists and belongs to A for every non-empty set $A \subseteq P$; that is, every non-empty subset of P has a least element. In this case $\leq$ is a **well-ordering** of P.

2A1B Transfinite Recursion: Theorem Let $(P, \leq)$ be a well-ordered set and X any class. For $p \in P$ write L_p for the set $\{q : q \in P, q < p\}$ and X^{L_p} for the class of all functions from L_p to X. Let $F : \bigcup_{p \in P} X^{L_p} \to X$ be any function. Then there is a unique function $f : P \to X$ such that $f(p) = F(f \restriction L_p)$ for every $p \in P$.

proof There are versions of this result in Enderton 77 (p. 175) and Halmos 60 (§18). Nevertheless I write out a proof, since it seems to me that most elementary books on set theory do not give it its proper place at the very beginning of the theory of well-ordered sets.

(a) Let Φ be the class of all functions ϕ such that
(α) $\operatorname{dom} \phi$ is a subset of P, and $L_p \subseteq \operatorname{dom} \phi$ for every $p \in \operatorname{dom} \phi$;
(β) $\phi(p) \in X$ for every $p \in \operatorname{dom} \phi$, and $\phi(p) = F(\phi \restriction L_p)$ for every $p \in \operatorname{dom} \phi$.

(b) If ϕ, $\psi \in \Phi$ then ϕ and ψ agree on $\operatorname{dom} \phi \cap \operatorname{dom} \psi$. **P?** If not, then $A = \{q : q \in \operatorname{dom} \phi \cap \operatorname{dom} \psi, \phi(q) \neq \psi(q)\}$ is non-empty. Because P is well-ordered, A has a least element p say. Now $L_p \subseteq \operatorname{dom} \phi \cap \operatorname{dom} \psi$ and $L_p \cap A = \emptyset$, so

$$\phi(p) = F(\phi \restriction L_p) = F(\psi \restriction L_p) = \psi(p),$$

which is impossible. **XQ**

(c) It follows that Φ is a set, since the function $\phi \mapsto \operatorname{dom} \phi$ is an injective function from Φ to $\mathcal{P}P$, and its inverse is a surjection from a subset of $\mathcal{P}P$ onto Φ. We can therefore, without inhibitions, define a function f by writing

$$\operatorname{dom} f = \bigcup_{\phi \in \Phi} \operatorname{dom} \phi, \quad f(p) = \phi(p) \text{ whenever } \phi \in \Phi,\ p \in \operatorname{dom} \phi.$$

(If you think that a function ϕ is just the set of ordered pairs $\{(p, \phi(p)) : p \in \operatorname{dom} \phi\}$, then f becomes $\bigcup \Phi$.) Then $f \in \Phi$. **P** Of course f is a function from a subset of P to X. If $p \in \operatorname{dom} f$, then there is a $\phi \in \Phi$ such that $p \in \operatorname{dom} \phi$, in which case

$$L_p \subseteq \operatorname{dom} \phi \subseteq \operatorname{dom} f, \quad f(p) = \phi(p) = F(\phi \restriction L_p) = F(f \restriction L_p). \ \mathbf{Q}$$

(d) f is defined everywhere in P. **P?** Otherwise, $P \setminus \operatorname{dom} f$ is non-empty and has a least element r say. Now $L_r \subseteq \operatorname{dom} f$. Define a function ψ by saying that $\operatorname{dom} \psi = \{r\} \cup \operatorname{dom} f$, $\psi(p) = f(p)$ for $p \in \operatorname{dom} f$ and $\psi(r) = F(f \restriction L_r)$. Then $\psi \in \Phi$, because if $p \in \operatorname{dom} \psi$

either $p \in \operatorname{dom} f$ so $L_p \subseteq \operatorname{dom} f \subseteq \operatorname{dom} \psi$ and

$$\psi(p) = f(p) = F(f \restriction L_p) = F(\psi \restriction L_p)$$

or $p = r$ so $L_p = L_r \subseteq \operatorname{dom} f \subseteq \operatorname{dom} \psi$ and

$$\psi(p) = F(f \restriction L_r) = F(\psi \restriction L_r).$$

Accordingly $\psi \in \Phi$ and $r \in \operatorname{dom} \psi \subseteq \operatorname{dom} f$. **XQ**

(e) Thus $f : P \to X$ is a function such that $f(p) = F(f \restriction L_p)$ for every p. To see that f is unique, observe that any function of this type must belong to Φ, so must agree with f on their common domain, which is the whole of P.

Remark If you have been taught to distinguish between the words 'set' and 'class', you will observe that my naive set theory is a relatively tolerant one in that it is willing to allow class variables in its theorems.

2A1C Ordinals An **ordinal** (sometimes called a 'von Neumann ordinal') is a set ξ such that

if $\eta \in \xi$ then η is a set and $\eta \notin \eta$,

if $\eta \in \zeta \in \xi$ then $\eta \in \xi$,

writing '$\eta \le \zeta$' to mean '$\eta \in \zeta$ or $\eta = \zeta$', $(\xi, \le)$ is well-ordered

(Enderton 77, p. 191; Halmos 60, §19; Henle 86, p. 27; Krivine 71, p. 24; Roitman 90, 3.2.8. Of course many set theories do not allow sets to belong to themselves, and/or take it for granted that every object of discussion is a set, but I prefer not to take a view on such points in general.)

2A1D Basic facts about ordinals (a) If ξ is an ordinal, then every member of ξ is an ordinal. (Enderton 77, p. 192; Henle 86, 6.4; Krivine 71, p. 14; Roitman 90, 3.2.10.)

(b) If ξ, η are ordinals then either $\xi \in \eta$ or $\xi = \eta$ or $\eta \in \xi$ (and no two of these can occur together). (Enderton 77, p. 192; Henle 86, 6.4; Krivine 71, p. 14; Lipschutz 64, 11.12; Roitman 90, 3.2.13.) It is customary, in this case, to write $\eta < \xi$ if $\eta \in \xi$ and $\eta \le \xi$ if either $\eta \in \xi$ or $\eta = \xi$. Note that $\eta \le \xi$ iff $\eta \subseteq \xi$.

(c) If A is any non-empty class of ordinals, then there is an $\alpha \in A$ such that $\alpha \le \xi$ for every $\xi \in A$. (Henle 86, 6.7; Krivine 71, p. 15.)

(d) If ξ is an ordinal, so is $\xi \cup \{\xi\}$; call it '$\xi + 1$'. If $\xi < \eta$ then $\xi + 1 \le \eta$; $\xi + 1$ is the least ordinal greater than ξ. (Enderton 77, p. 193; Henle 86, 6.3; Krivine 71, p. 15.) For any ordinal ξ, either there is a greatest ordinal $\eta < \xi$, in which case $\xi = \eta + 1$ and we call ξ a **successor ordinal**, or $\xi = \bigcup \xi$, in which case we call ξ a **limit ordinal**.

(e) The first few ordinals are $0 = \emptyset$, $1 = 0 + 1 = \{0\} = \{\emptyset\}$, $2 = 1 + 1 = \{0, 1\} = \{\emptyset, \{\emptyset\}\}$, $3 = 2 + 1 = \{0, 1, 2\}$, $\dots$. The first infinite ordinal is $\omega = \{0, 1, 2, \dots\}$, which may be identified with $\mathbb{N}$.

(f) The union of any set of ordinals is an ordinal. (Enderton 77, p. 193; Henle 86, 6.8; Krivine 71, p. 15; Roitman 90, 3.2.19.)

(g) If $(P, \le)$ is any well-ordered set, there is a unique ordinal ξ such that P is order-isomorphic to ξ, and the order-isomorphism is unique. (Enderton 77, pp. 187-189; Henle 86, 6.13; Halmos 60, §20.)

2A1E Initial ordinals An **initial ordinal** is an ordinal κ such that there is no bijection between κ and any member of κ. (Enderton 77, p. 197; Halmos 60, §25; Henle 86, p. 34; Krivine 71, p. 24; Roitman 90, 5.1.10, p. 79).

2A1F Basic facts about initial ordinals (a) All finite ordinals, and the first infinite ordinal ω, are initial ordinals.

(b) For every well-ordered set P there is a unique initial ordinal κ such that there is a bijection between P and κ.

(c) For every ordinal ξ there is a least initial ordinal greater than ξ. (Enderton 77, p. 195; Henle 86, 7.2.1.) If κ is an initial ordinal, write κ^+ for the least initial ordinal greater than κ. We write ω_1 for ω^+, ω_2 for ω_1^+, and so on.

(d) For any initial ordinal $\kappa \ge \omega$ there is a bijection between $\kappa \times \kappa$ and κ; consequently there are bijections between κ and κ^r for every $r \ge 1$.

2A1G Schröder-Bernstein theorem I remind you of the following fundamental result: if X and Y are sets and there are injections $f : X \to Y$, $g : Y \to X$ then there is a bijection $h : X \to Y$. (Enderton 77, p. 147; Halmos 60, §22; Henle 86, 7.4; Lipschutz 64, p. 145; Roitman 90, 5.1.2. It is also a special case of 344D in Volume 3.)

2A1H Countable subsets of $\mathcal{P}\mathbb{N}$ The following results will be needed below.

(a) There is a bijection between $\mathcal{P}\mathbb{N}$ and $\mathbb{R}$. (Enderton 77, p. 149; Lipschutz 64, p. 146.)

(b) Suppose that X is any set such that there is an injection from X into $\mathcal{P}\mathbb{N}$. Let $\mathcal{C}$ be the set of countable subsets of X. Then there is a surjection from $\mathcal{P}\mathbb{N}$ onto $\mathcal{C}$. **P** Let $f : X \to \mathcal{P}\mathbb{N}$ be an injection. Set $f_1(x) = \{0\} \cup \{i+1 : i \in f(x)\}$; then $f_1 : X \to \mathcal{P}\mathbb{N}$ is injective and $f_1(x) \ne \emptyset$ for every $x \in X$. Define $g : \mathcal{P}\mathbb{N} \to \mathcal{P}X$ by setting

$$g(A) = \{x : \exists\, n \in \mathbb{N},\ f_1(x) = \{i : 2^n(2i + 1) \in A\}\}$$

for each $A \subseteq \mathbb{N}$. Then $g(A)$ is countable, since we have an injection

$$x \mapsto \min\{n : f_1(x) = \{i : 2^n(2i + 1) \in A\}\}$$

from $g(A)$ to $\mathbb{N}$. Thus g is a function from $\mathcal{P}\mathbb{N}$ to $\mathcal{C}$. To see that g is surjective, observe that $\emptyset = g(\emptyset)$, while if $C \subseteq X$ is countable and not empty there is a surjection $h : \mathbb{N} \to C$; now set

$$A = \{2^n(2i+1) : n \in \mathbb{N},\ i \in f_1(h(n))\},$$

and see that $g(A) = C$. **Q**

(c) Again suppose that X is a set such that there is an injection from X to $\mathcal{P}\mathbb{N}$, and write H for the set of functions h such that $\operatorname{dom} h$ is a countable subset of X and h takes values in $\{0,1\}$. Then there is a surjection from $\mathcal{P}\mathbb{N}$ onto H. **P** Let $\mathcal{C}$ be the set of countable subsets of X and let $g : \mathcal{P}\mathbb{N} \to \mathcal{C}$ be a surjection, as in (a). For $A \subseteq \mathbb{N}$ set

$$g_0(A) = g(\{i : 2i \in A\}), \quad g_1(A) = g(\{i : 2i+1 \in A\}),$$

so that $g_0(A)$, $g_1(A)$ are countable subsets of X, and $A \mapsto (g_0(A), g_1(A))$ is a surjection from $\mathcal{P}\mathbb{N}$ onto $\mathcal{C} \times \mathcal{C}$. Let h_A be the function with domain $g_0(A) \cup g_1(A)$ such that $h_A(x) = 1$ if $x \in g_1(A)$, 0 if $x \in g_0(A) \setminus g_1(A)$. Then $A \mapsto h_A$ is a surjection from $\mathcal{P}\mathbb{N}$ onto H. **Q**

2A1I Filters I pause for a moment to discuss a construction which is of great value in investigating topological spaces, but has other uses, and in its nature belongs to elementary set theory (much more elementary, indeed, than the work above).

(a) Let X be a non-empty set. A **filter** on X is a family $\mathcal{F}$ of subsets of X such that

$X \in \mathcal{F}$, $\emptyset \notin \mathcal{F}$,

$E \cap F \in \mathcal{F}$ whenever $E, F \in \mathcal{F}$,

$E \in \mathcal{F}$ whenever $X \supseteq E \supseteq F \in \mathcal{F}$.

The second condition implies (inducing on n) that $F_0 \cap \ldots \cap F_n \in \mathcal{F}$ whenever $F_0, \ldots, F_n \in \mathcal{F}$.

(b) Let X, Y be non-empty sets, $\mathcal{F}$ a filter on X and $f : D \to Y$ a function, where $D \in \mathcal{F}$. Then

$$\{E : E \subseteq Y,\ f^{-1}[E] \in \mathcal{F}\}$$

is a filter on Y (because $f^{-1}[Y] = D$, $f^{-1}[\emptyset] = \emptyset$, $f^{-1}[E \cap F] = f^{-1}[E] \cap f^{-1}[F]$, $X \supseteq f^{-1}[E] \supseteq f^{-1}[F]$ whenever $Y \supseteq E \supseteq F$); I will call it $f[[\mathcal{F}]]$, the **image filter** of $\mathcal{F}$ under f.

Remark Of course there is a hidden variable in this notation. Ordinarily in this book I regard a function f as being defined by its domain $\operatorname{dom} f$ and its values on its domain; that is, it is determined by its graph $\{(x, f(x)) : x \in \operatorname{dom} f\}$, and indeed I normally do not distinguish between a function and its graph. This means that when I write '$f : D \to Y$ is a function' then the class $D = \operatorname{dom} f$ can be recovered from the function, but the class Y cannot; all I promise is that Y includes the class $f[D]$ of values of f. Now in the notation $f[[\mathcal{F}]]$ above we do actually need to know which set Y it is to be a filter on, even though this cannot be discovered from knowledge of f and $\mathcal{F}$. So you will always have to infer it from the context.

2A1J The Axiom of Choice I come now to the second half of this section, in which I discuss concepts and theorems dependent on the Axiom of Choice. Let me remind you of the statement of this axiom:

(AC) 'whenever I is a set and $\langle X_i \rangle_{i \in I}$ is a family of non-empty sets indexed by I, there is a function f, with domain I, such that $f(i) \in X_i$ for every $i \in I$'.

The function f is a **choice function**; it picks out one member of each of the given family of non-empty sets X_i.

I believe that one's attitude to this principle is a matter for individual choice. It is an indispensable foundation for very large parts of twentieth-century pure mathematics, including a substantial fraction of the present volume; but there are also significant areas in which principles actually contradictory to it can be employed to striking effect, leading – in my view – to equally valid mathematics. (I will describe one of these in §567 of Volume 5.) At present it is the case that more current mathematical activity, by volume, depends on asserting the axiom of choice than on all its rivals put together; but it is a matter of judgement and taste where the most important, or exciting, ideas are to be found. For the present volume I follow standard practice in twentieth-century abstract analysis, using the axiom of choice whenever necessary.

2A1K Zermelo's Well-Ordering Theorem (a) The Axiom of Choice is equiveridical with each of the statements

'for every set X there is a well-ordering of X',

'for every set X there is a bijection between X and some ordinal',

'for every set X there is a unique initial ordinal κ such that there is a bijection between X and κ.'

(Enderton 77, p. 196 et seq.; Halmos 60, §17; Henle 86, 9.1-9.3; Krivine 71, p. 20; Lipschutz 64, 12.1; Roitman 90, 3.6.38.)

(b) When assuming the axiom of choice, as I do nearly everywhere in this treatise, I write $\#(X)$ for that initial ordinal κ such that there is a bijection between κ and X; I call this the **cardinal** of X.

2A1L Fundamental consequences of the Axiom of Choice (a) For any two sets X and Y, there is a bijection between X and Y iff $\#(X) = \#(Y)$. More generally, there is an injection from X to Y iff $\#(X) \le \#(Y)$, and a surjection from X onto Y iff either $\#(X) \ge \#(Y) > 0$ or $\#(X) = \#(Y) = 0$.

(b) In particular, $\#(\mathcal{P}\mathbb{N}) = \#(\mathbb{R})$; write $\mathfrak{c}$ for this common value, the **cardinal of the continuum**. Cantor's theorem that $\mathcal{P}\mathbb{N}$ and $\mathbb{R}$ are uncountable becomes the result $\omega < \mathfrak{c}$, that is, $\omega_1 \le \mathfrak{c}$.

(c) If X is any infinite set, and $r \ge 1$, then there is a bijection between X^r and X. (Enderton 77, p. 162; Halmos 60, §24.) (I note that we need some form of the axiom of choice to prove the result in this generality. But of course for most of the infinite sets arising naturally in mathematics – sets like $\mathbb{N}$ and $\mathcal{P}\mathbb{R}$ – it is easy to prove the result without appeal to the axiom of choice.)

(d) Suppose that κ is an infinite cardinal. If I is a set of cardinal at most κ and $\langle A_i\rangle_{i\in I}$ is a family of sets with $\#(A_i) \le \kappa$ for every $i \in I$, then $\#(\bigcup_{i\in I} A_i) \le \kappa$. Consequently $\#(\bigcup \mathcal{A}) \le \kappa$ whenever $\mathcal{A}$ is a family of sets such that $\#(\mathcal{A}) \le \kappa$ and $\#(A) \le \kappa$ for every $A \in \mathcal{A}$. In particular, ω_1 cannot be expressed as a countable union of countable sets, and ω_2 cannot be expressed as a countable union of sets of cardinal at most ω_1.

(e) Now we can rephrase 2A1Hc as: if $\#(X) \le \mathfrak{c}$, then $\#(H) \le \mathfrak{c}$, where H is the set of functions from a countable subset of X to $\{0,1\}$. **P** For we have an injection from X into $\mathcal{P}\mathbb{N}$, and therefore a surjection from $\mathcal{P}\mathbb{N}$ onto H. **Q**

(f) Any non-empty class of cardinals has a least member (by 2A1Dc).

2A1M Zorn's Lemma In 2A1K I described the well-ordering principle. I come now to another proposition which is equiveridical with the axiom of choice:

> 'Let $(P, \le)$ be a non-empty partially ordered set such that every non-empty totally ordered subset of P has an upper bound in P. Then P has a maximal element.'

This is **Zorn's Lemma**. For the proof that the axiom of choice implies, and is implied by, Zorn's Lemma, see Enderton 77, p. 151; Halmos 60, §16; Henle 86, 9.1-9.3; Roitman 90, 3.6.38.

2A1N Ultrafilters A filter $\mathcal{F}$ on a set X is an **ultrafilter** if for every $A \subseteq X$ either $A \in \mathcal{F}$ or $X \setminus A \in \mathcal{F}$.

If $\mathcal{F}$ is an ultrafilter on X and $f : D \to Y$ is a function, where $D \in \mathcal{F}$, then $f[[\mathcal{F}]]$ is an ultrafilter on Y (because $f^{-1}[Y \setminus A] = D \setminus f^{-1}[A]$ for every $A \subseteq Y$).

One type of ultrafilter can be described easily: if x is any point of a set X, then $\mathcal{F} = \{F : x \in F \subseteq X\}$ is an ultrafilter on X. (You need only read the definitions. Ultrafilters of this type are called **principal ultrafilters**.) But it is not obvious that there are any further ultrafilters, and indeed it is not possible to prove that there are any, without using a strong form of the axiom of choice, as follows.

2A1O The Ultrafilter Theorem As an example of the use of Zorn's lemma which will be of great value in studying compact topological spaces (2A3N *et seq.*, and §247), I give the following result.

Theorem Let X be any non-empty set, and $\mathcal{F}$ a filter on X. Then there is an ultrafilter $\mathcal{H}$ on X such that $\mathcal{F} \subseteq \mathcal{H}$.

proof (Cf. Henle 86, 9.4; Roitman 90, 3.6.37.) Let $\mathfrak{P}$ be the set of all filters on X including $\mathcal{F}$, and order $\mathfrak{P}$ by inclusion, so that, for $\mathcal{G}_1, \mathcal{G}_2 \in \mathfrak{P}$, $\mathcal{G}_1 \le \mathcal{G}_2$ in $\mathfrak{P}$ iff $\mathcal{G}_1 \subseteq \mathcal{G}_2$. It is easy to see that $\mathfrak{P}$ is a partially ordered set, and it is non-empty because $\mathcal{F} \in \mathfrak{P}$. If $\mathfrak{Q}$ is any non-empty totally ordered subset of $\mathfrak{P}$, then $\mathcal{H}_{\mathfrak{Q}} = \bigcup \mathfrak{Q} \in \mathfrak{P}$. **P** Of course $\mathcal{H}_{\mathfrak{Q}}$ is a family of subsets of X. (i) Take any $\mathcal{G}_0 \in \mathfrak{Q}$; then $X \in \mathcal{G}_0 \subseteq \mathcal{H}_{\mathfrak{Q}}$. If $\mathcal{G} \in \mathfrak{Q}$, then $\mathcal{G}$ is a filter, so $\emptyset \notin \mathcal{G}$; accordingly $\emptyset \notin \mathcal{H}_{\mathfrak{Q}}$. (ii) If $E, F \in \mathcal{H}_{\mathfrak{Q}}$, then there are $\mathcal{G}_1, \mathcal{G}_2 \in \mathfrak{Q}$ such that $E \in \mathcal{G}_1$ and $F \in \mathcal{G}_2$. Because $\mathfrak{Q}$ is totally ordered, either $\mathcal{G}_1 \subseteq \mathcal{G}_2$ or $\mathcal{G}_2 \subseteq \mathcal{G}_1$. In either case, $\mathcal{G} = \mathcal{G}_1 \cup \mathcal{G}_2 \in \mathfrak{Q}$. Now $\mathcal{G}$ is a filter containing both E and F, so it contains $E \cap F$, and $E \cap F \in \mathcal{H}_{\mathfrak{Q}}$. (iii) If $X \supseteq E \supseteq F \in \mathcal{H}_{\mathfrak{Q}}$, there is a $\mathcal{G} \in \mathfrak{Q}$ such that $F \in \mathcal{G}$; and $E \in \mathcal{G} \subseteq \mathcal{H}_{\mathfrak{Q}}$. This shows that $\mathcal{H}_{\mathfrak{Q}}$ is a filter on X. (iv) Finally, $\mathcal{H}_{\mathfrak{Q}} \supseteq \mathcal{G}_0 \supseteq \mathcal{F}$, so $\mathcal{H}_{\mathfrak{Q}} \in \mathfrak{P}$. **Q** Now $\mathcal{H}_{\mathfrak{Q}}$ is evidently an upper bound for $\mathfrak{Q}$ in $\mathfrak{P}$.

We may therefore apply Zorn's Lemma to find a maximal element $\mathcal{H}$ of $\mathfrak{P}$. This $\mathcal{H}$ is surely a filter on X including $\mathcal{F}$.

Now let $A \subseteq X$ be such that $A \notin \mathcal{H}$. Consider

$$\mathcal{H}_1 = \{E : E \subseteq X,\ E \cup A \in \mathcal{H}\}.$$

This is a filter on X. **P** Of course it is a family of subsets of X. (i) $X \cup A = X \in \mathcal{H}$, so $X \in \mathcal{H}_1$. $\emptyset \cup A = A \notin \mathcal{H}$ so $\emptyset \notin \mathcal{H}_1$. (ii) If E, $F \in \mathcal{H}_1$ then

$$(E \cap F) \cup A = (E \cup A) \cap (F \cup A) \in \mathcal{H},$$

so $E \cap F \in \mathcal{H}_1$. (iii) If $X \supseteq E \supseteq F \in \mathcal{H}_1$ then $E \cup A \supseteq F \cup A \in \mathcal{H}$, so $E \cup A \in \mathcal{H}$ and $E \in \mathcal{H}_1$. **Q** Also $\mathcal{H}_1 \supseteq \mathcal{H}$, so $\mathcal{H}_1 \in \mathfrak{P}$. But $\mathcal{H}$ is a maximal element of $\mathfrak{P}$, so $\mathcal{H}_1 = \mathcal{H}$. Since $(X \setminus A) \cup A = X \in \mathcal{H}$, $X \setminus A \in \mathcal{H}_1$ and $X \setminus A \in \mathcal{H}$.

As A is arbitrary, $\mathcal{H}$ is an ultrafilter, as required.

2A1P I come now to a result from infinitary combinatorics for which I give a detailed proof, not because it cannot be found in many textbooks, but because it is usually given in enormously greater generality, to the point indeed that it may be harder to understand why the stated theorem covers the present result than to prove the latter from first principles.

Theorem (a) Let $\langle K_\alpha \rangle_{\alpha \in A}$ be a family of countable sets, with $\#(A)$ strictly greater than $\mathfrak{c}$, the cardinal of the continuum. Then there are a set M, of cardinal at most $\mathfrak{c}$, and a set $B \subseteq A$, of cardinal strictly greater than $\mathfrak{c}$, such that $K_\alpha \cap K_\beta \subseteq M$ whenever α, β are distinct members of B.

(b) Let I be a set, and $\langle f_\alpha \rangle_{\alpha \in A}$ a family in $\{0,1\}^I$, the set of functions from I to $\{0,1\}$, with $\#(A) > \mathfrak{c}$. If $\langle K_\alpha \rangle_{\alpha \in A}$ is any family of countable subsets of I, then there is a set $B \subseteq A$, of cardinal greater than $\mathfrak{c}$, such that f_α and f_β agree on $K_\alpha \cap K_\beta$ for all α, $\beta \in B$.

(c) In particular, under the conditions of (b), there are distinct α, $\beta \in A$ such that f_α and f_β agree on $K_\alpha \cap K_\beta$.

proof (a) Choose inductively a family $\langle M_\xi \rangle_{\xi < \omega_1}$ of sets by the rule

if there is any set N such that

$(*)$ N is disjoint from $\bigcup_{\eta < \xi} M_\eta$, $\#(N) \le \mathfrak{c}$ and

$\#(\{\alpha : \alpha \in A,\, K_\alpha \cap N = \emptyset\}) \le \mathfrak{c}$,

choose such a set for M_ξ;

otherwise set $M_\xi = \emptyset$.

When M_ξ has been chosen for every $\xi < \omega_1$, set $M = \bigcup_{\xi < \omega_1} M_\xi$. The rule ensures that $\langle M_\xi \rangle_{\xi < \omega_1}$ is disjoint and that $\#(M_\xi) \le \mathfrak{c}$ for every $\xi < \omega_1$, while $\omega_1 \le \mathfrak{c}$, so $\#(M) \le \mathfrak{c}$.

Let $\mathfrak{P}$ be the family of sets $P \subseteq A$ such that $K_\alpha \cap K_\beta \subseteq M$ for all distinct α, $\beta \in P$. Order $\mathfrak{P}$ by inclusion, so that it is a partially ordered set. If $\mathfrak{Q} \subseteq \mathfrak{P}$ is totally ordered, then $\bigcup \mathfrak{Q} \in \mathfrak{P}$. **P** If α, β are distinct members of $\bigcup \mathfrak{Q}$, there are Q_1, $Q_2 \in \mathfrak{Q}$ such that $\alpha \in Q_1$, $\beta \in Q_2$; now $P = Q_1 \cup Q_2$ is equal to one of Q_1, Q_2, and in either case belongs to $\mathfrak{P}$ and contains both α and β, so $K_\alpha \cap K_\beta \subseteq M$. **Q** By Zorn's Lemma, $\mathfrak{P}$ has a maximal element B, and we surely have $K_\alpha \cap K_\beta \subseteq M$ for all distinct α, $\beta \in B$.

? Suppose, if possible, that $\#(B) \le \mathfrak{c}$. Set $N = \bigcup_{\alpha \in B} K_\alpha \setminus M$. Then N has cardinal at most $\mathfrak{c}$, being included in a union of at most $\mathfrak{c}$ countable sets. For every $\gamma \in A \setminus B$, $B \cup \{\gamma\} \notin \mathfrak{P}$, so there must be some $\alpha \in B$ such that $K_\alpha \cap K_\gamma \not\subseteq M$; that is, $K_\gamma \cap N \ne \emptyset$. Thus $\{\gamma : K_\gamma \cap N = \emptyset\} \subseteq B$ has cardinal at most $\mathfrak{c}$. But this means that in the rule for choosing M_ξ, there was always an N satisfying the condition $(*)$, and therefore M_ξ also did. Thus $C_\xi = \{\alpha : K_\alpha \cap M_\xi = \emptyset\}$ has cardinal at most $\mathfrak{c}$ for every $\xi < \omega_1$. So $C = \bigcup_{\xi < \omega_1} C_\xi$ also has. But the original hypothesis was that $\#(A) > \mathfrak{c}$, so there is an $\alpha \in A \setminus C$. In this case, $K_\alpha \cap M_\xi \ne \emptyset$ for every $\xi < \omega_1$. But this means that we have a surjection $\phi : K_\alpha \cap M \to \omega_1$ given by setting

$$\phi(i) = \xi \text{ if } i \in K_\alpha \cap M_\xi.$$

Since $\#(K_\alpha) \le \omega < \omega_1$, this is impossible. **X**

Accordingly $\#(B) > \mathfrak{c}$ and we have found a suitable pair M, B.

(b) By (a), we can find a set M, of cardinal at most $\mathfrak{c}$, and a set $B_0 \subseteq A$, of cardinal greater than $\mathfrak{c}$, such that $K_\alpha \cap K_\beta \subseteq M$ for all distinct α, $\beta \in B_0$. Let H be the set of functions from countable subsets of M to $\{0,1\}$; then $f'_\alpha = f_\alpha \restriction (K_\alpha \cap M) \in H$ for each $\alpha \in B_0$. Now $B_0 = \bigcup_{h \in H} \{\alpha : \alpha \in B_0,\, f'_\alpha = h\}$ has cardinal greater than $\mathfrak{c}$, while $\#(H) \le \mathfrak{c}$ (2A1Le), so there must be some $h \in H$ such that $B = \{\alpha : \alpha \in B_0,\, f'_\alpha = h\}$ has cardinal greater than $\mathfrak{c}$.

If α, β are distinct members of B, then $K_\alpha \cap K_\beta \subseteq M$, because α, $\beta \in B_0$; but this means that

$$f_\alpha \restriction K_\alpha \cap K_\beta = h \restriction K_\alpha \cap K_\beta = f_\beta \restriction K_\alpha \cap K_\beta.$$

Thus B has the required property. (Of course f_α and f_β agree on $K_\alpha \cap K_\beta$ if $\alpha = \beta$.)

(c) follows at once.

Remark The result we need in this volume (in 216E) is part (c) above. There are other proofs of this, perhaps a little simpler; but the stronger result in part (b) will be useful in Volume 3.

2A2 The topology of Euclidean space

In the appendix to Volume 1 (§1A2) I discussed open and closed sets in $\mathbb{R}^r$; the chief aim there was to support the idea of 'Borel set', which is vital in the theory of Lebesgue measure, but of course they are also fundamental to the study of continuous functions, and indeed to all aspects of real analysis. I give here a very brief introduction to the further elementary facts about closed and compact sets and continuous functions which we need for this volume. Much of this material can be derived from the generalizations in §2A3, but nevertheless I sketch the proofs, since for the greater part of the volume (most of the exceptions are in Chapter 24) Euclidean space is sufficient for our needs.

2A2A Closures: Definition For any $r \geq 1$ and any $A \subseteq \mathbb{R}^r$, the **closure** of A, $\overline{A}$, is the intersection of all the closed subsets of $\mathbb{R}^r$ including A. This is itself closed (being the intersection of a non-empty family of closed sets, see 1A2Fd), so is the smallest closed set including A. In particular, A is closed iff $\overline{A} = A$.

2A2B Lemma Let $A \subseteq \mathbb{R}^r$ be any set. Then for $x \in \mathbb{R}^r$ the following are equiveridical:
(i) $x \in \overline{A}$, the closure of A;
(ii) $B(x,\delta) \cap A \neq \emptyset$ for every $\delta > 0$, where $B(x,\delta) = \{y : \|y - x\| \leq \delta\}$;
(iii) there is a sequence $\langle x_n \rangle_{n\in\mathbb{N}}$ in A such that $\lim_{n\to\infty} \|x_n - x\| = 0$.

proof (a)(i)⇒(ii) Suppose that $x \in \overline{A}$ and $\delta > 0$. Then $U(x,\delta) = \{y : \|y - x\| < \delta\}$ is an open set (1A2D), so $F = \mathbb{R}^r \setminus U(x,\delta)$ is closed, while $x \notin F$. Now

$$x \in \overline{A} \setminus F \Longrightarrow \overline{A} \not\subseteq F \Longrightarrow A \not\subseteq F \Longrightarrow A \cap U(x,\delta) \neq \emptyset \Longrightarrow A \cap B(x,\delta) \neq \emptyset.$$

As δ is arbitrary, (ii) is true.

(b)(ii)⇒(iii) If (ii) is true, then for each $n \in \mathbb{N}$ we can find an $x_n \in A$ such that $\|x_n - x\| \leq 2^{-n}$, and now $\lim_{n\to\infty} \|x_n - x\| = 0$.

(c)(iii)⇒(i) Assume (iii). **?** Suppose, if possible, that $x \notin \overline{A}$. Then x belongs to the open set $\mathbb{R}^r \setminus \overline{A}$ and there is a $\delta > 0$ such that $U(x,\delta) \subseteq \mathbb{R}^r \setminus \overline{A}$. But now there is an n such that $\|x_n - x\| < \delta$, in which case $x_n \in U(x,\delta) \cap A \subseteq U(x,\delta) \cap \overline{A}$. **X**

2A2C Continuous functions (a) I begin with a characterization of continuous functions in terms of open sets. If r, $s \geq 1$, $D \subseteq \mathbb{R}^r$ and $\phi : D \to \mathbb{R}^s$ is a function, we say that ϕ is **continuous** if for every $x \in D$ and $\epsilon > 0$ there is a $\delta > 0$ such that $\|\phi(y) - \phi(x)\| \leq \epsilon$ whenever $y \in D$ and $\|y - x\| \leq \delta$. Now ϕ is continuous iff for every open set $G \subseteq \mathbb{R}^s$ there is an open set $H \subseteq \mathbb{R}^r$ such that $\phi^{-1}[G] = D \cap H$.

P (i) Suppose that ϕ is continuous and that $G \subseteq \mathbb{R}^s$ is open. Set

$$H = \bigcup\{U : U \subseteq \mathbb{R}^r \text{ is open, } \phi[U \cap D] \subseteq G\}.$$

Then H is a union of open sets, therefore open (1A2Bd), and $H \cap D \subseteq \phi^{-1}[G]$. If $x \in \phi^{-1}[G]$, then $\phi(x) \in G$, so there is an $\epsilon > 0$ such that $U(\phi(x),\epsilon) \subseteq G$; now there is a $\delta > 0$ such that $\|\phi(y) - \phi(x)\| \leq \frac{1}{2}\epsilon$ whenever $y \in D$ and $\|y - x\| \leq \delta$, so that

$$\phi[U(x,\delta) \cap D] \subseteq U(\phi(x),\epsilon) \subseteq G$$

and

$$x \in U(x,\delta) \subseteq H.$$

As x is arbitrary, $\phi^{-1}[G] = H \cap D$. As G is arbitrary, ϕ satisfies the condition.

(ii) Now suppose that ϕ satisfies the condition. Take $x \in D$ and $\epsilon > 0$. Then $U(\phi(x),\epsilon)$ is open, so there is an open $H \subseteq \mathbb{R}^r$ such that $H \cap D = \phi^{-1}[U(\phi(x),\epsilon)]$; we see that $x \in H$, so there is a $\delta > 0$ such that $U(x,\delta) \subseteq H$; now if $y \in D$ and $\|y - x\| \leq \frac{1}{2}\delta$ then $y \in D \cap H$, $\phi(y) \in U(\phi(x),\epsilon)$ and $\|\phi(y) - \phi(x)\| \leq \epsilon$. As x and ϵ are arbitrary, ϕ is continuous. **Q**

(b) Using the ϵ-δ definition of continuity, it is easy to see that a function ϕ from a subset D of $\mathbb{R}^r$ to $\mathbb{R}^s$ is continuous iff all its components ϕ_i are continuous, writing $\phi(x) = (\phi_1(x), \dots, \phi_s(x))$ for $x \in D$. **P (i)** If ϕ is continuous, $i \leq s$, $x \in D$ and $\epsilon > 0$, then there is a $\delta > 0$ such that

$$|\phi_i(y) - \phi_i(x)| \leq \|\phi(y) - \phi(x)\| \leq \epsilon$$

whenever $y \in D$ and $\|y - x\| \leq \delta$. **(ii)** If every ϕ_i is continuous, $x \in D$ and $\epsilon > 0$, then there are $\delta_i > 0$ such that $|\phi_i(y) - \phi_i(x)| \leq \epsilon/\sqrt{s}$ whenever $y \in D$ and $\|y - x\| \leq \delta_i$; setting $\delta = \min_{1\leq i\leq r} \delta_i > 0$, we have $\|\phi(y) - \phi(x)\| \leq \epsilon$ whenever $y \in D$ and $\|y - x\| \leq \delta$. **Q**

2A2D Compactness in $\mathbb{R}^r$: Definition A subset F of $\mathbb{R}^r$ is called **compact** if whenever $\mathcal{G}$ is a family of open sets covering F then there is a finite subset $\mathcal{G}_0$ of $\mathcal{G}$ still covering F.

2A2E Elementary properties of compact sets Take any $r \geq 1$, and subsets D, F, G and K of $\mathbb{R}^r$.

(a) If K is compact and F is closed, then $K \cap F$ is compact. **P** Let $\mathcal{G}$ be an open cover of $F \cap K$. Then $\mathcal{G} \cup \{\mathbb{R}^r \setminus F\}$ is an open cover of K, so has a finite subcover $\mathcal{G}_0$ say. Now $\mathcal{G}_0 \setminus \{\mathbb{R}^r \setminus F\}$ is a finite subset of $\mathcal{G}$ covering $K \cap F$. As $\mathcal{G}$ is arbitrary, $K \cap F$ is compact. **Q**

(b) If $s \geq 1$, $\phi : D \to \mathbb{R}^s$ is a continuous function, K is compact and $K \subseteq D$, then $\phi[K]$ is compact. **P** Let $\mathcal{V}$ be an open cover of $\phi[K]$. Let $\mathcal{H}$ be

$$\{H : H \subseteq \mathbb{R}^r \text{ is open, } \exists\ V \in \mathcal{V},\ \phi^{-1}[V] = D \cap H\}.$$

If $x \in K$, then $\phi(x) \in \phi[K]$ so there is a $V \in \mathcal{V}$ such that $\phi(x) \in V$; now there is an $H \in \mathcal{H}$ such that $D \cap H\phi^{-1}[V]$ contains x (2A2Ca); as x is arbitrary, $K \subseteq \bigcup \mathcal{H}$. Let $\mathcal{H}_0$ be a finite subset of H covering K. For each $H \in \mathcal{H}_0$, let $V_H \in \mathcal{V}$ be such that $\phi^{-1}[V_H] = D \cap H$; then $\{V_H : H \in \mathcal{H}_0\}$ is a finite subset of $\mathcal{V}$ covering $\phi[K]$. As $\mathcal{V}$ is arbitrary, $\phi[K]$ is compact. **Q**

(c) If K is compact, it is closed. **P** Write $H = \mathbb{R}^r \setminus K$. Take any $x \in H$. Then $G_n = \mathbb{R}^r \setminus B(x, 2^{-n})$ is open for every $n \in \mathbb{N}$ (1A2G). Also

$$\bigcup_{n\in\mathbb{N}} G_n = \{y : y \in \mathbb{R}^r,\ \|y - x\| > 0\} = \mathbb{R}^r \setminus \{x\} \supseteq K.$$

So there is some finite set $\mathcal{G}_0 \subseteq \{G_n : n \in \mathbb{N}\}$ which covers K. There must be an n such that $\mathcal{G}_0 \subseteq \{G_i : i \leq n\}$, so that

$$K \subseteq \bigcup \mathcal{G}_0 \subseteq \bigcup_{i\leq n} G_i = G_n,$$

and $B(x, 2^{-n}) \subseteq H$. As x is arbitrary, H is open and K is closed. **Q**

(d) If K is compact and G is open and $K \subseteq G$, then there is a $\delta > 0$ such that $K + B(\mathbf{0}, \delta) \subseteq G$. **P** If $K = \emptyset$, this is trivial, as then

$$K + B(\mathbf{0}, 1) = \{x + y : x \in K,\ y \in B(\mathbf{0}, 1)\} = \emptyset.$$

Otherwise, set

$$\mathcal{G} = \{U(x, \delta) : x \in \mathbb{R}^r,\ \delta > 0,\ U(x, 2\delta) \subseteq G\}.$$

Then $\mathcal{G}$ is a family of open sets and $\bigcup \mathcal{G} = G$ (because G is open), so $\mathcal{G}$ is an open cover of K and has a finite subcover $\mathcal{G}_0$. Express $\mathcal{G}_0$ as $\{U(x_0, \delta_0), \dots , U(x_n, \delta_n)\}$ where $U(x_i, 2\delta_i) \subseteq G$ for each i. Set $\delta = \min_{i\leq n} \delta_i > 0$. If $x \in K$ and $y \in B(\mathbf{0}, \delta)$, then there is an $i \leq n$ such that $x \in U(x_i, \delta_i)$; now

$$\|(x + y) - x_i\| \leq \|x - x_i\| + \|y\| < \delta_i + \delta \leq 2\delta_i,$$

so $x + y \in U(x_i, 2\delta_i) \subseteq G$. As x and y are arbitrary, $K + B(\mathbf{0}, \delta) \subseteq G$. **Q**

Remark This result is a simple form of the **Lebesgue covering lemma**.

2A2F The value of the concept of 'compactness' is greatly increased by the fact that there is an effective characterization of the compact subsets of $\mathbb{R}^r$.

Theorem For any $r \geq 1$, a subset K of $\mathbb{R}^r$ is compact iff it is closed and bounded.

proof (a) Suppose that K is compact. By 2A2Ec, it is closed. To see that it is bounded, consider $\mathcal{G} = \{U(\mathbf{0}, n) : n \in \mathbb{N}\}$. $\mathcal{G}$ consists entirely of open sets, and $\bigcup \mathcal{G} = \mathbb{R}^r \supseteq K$, so there is a finite $\mathcal{G}_0 \subseteq \mathcal{G}$ covering K. There must be an n such that $\mathcal{G}_0 \subseteq \{G_i : i \leq n\}$, so that

$$K \subseteq \bigcup \mathcal{G}_0 \subseteq \bigcup_{i\leq n} U(\mathbf{0}, i) = U(\mathbf{0}, n),$$

and K is bounded.

(b) Thus we are left with the converse; I have to show that a closed bounded set is compact. The main part of the argument is a proof by induction on r that the closed interval $[-\mathbf{n}, \mathbf{n}]$ is compact for all $n \in \mathbb{N}$, writing $\mathbf{n} = (n, \dots , n) \in \mathbb{R}^r$.

(i) If $r = 1$ and $n \in \mathbb{N}$ and $\mathcal{G}$ is a family of open sets in $\mathbb{R}$ covering $[-n, n]$, set

$$A = \{x : x \in [-n, n],\ \text{there is a finite } \mathcal{G}_0 \subseteq \mathcal{G} \text{ such that } [-n, x] \subseteq \bigcup \mathcal{G}_0\}.$$

Then $-n \in A$, because if $-n \in G \in \mathcal{G}$ then $[-n,-n] \subseteq \bigcup\{G\}$, and A is bounded above by n, so $c = \sup A$ exists and belongs to $[-n,n]$.

Next, $c \in [-n,n] \subseteq \bigcup \mathcal{G}$, so there is a $G \in \mathcal{G}$ containing c. Let $\delta > 0$ be such that $U(c,\delta) \subseteq G$. There is an $x \in A$ such that $x \geq c - \delta$. Let $\mathcal{G}_0$ be a finite subset of $\mathcal{G}$ covering $[-n,x]$. Then $\mathcal{G}_1 = \mathcal{G}_0 \cup \{G\}$ is a finite subset of $\mathcal{G}$ covering $[-n, c+\frac{1}{2}\delta]$. But $c + \frac{1}{2}\delta \notin A$ so $c + \frac{1}{2}\delta > n$ and $\mathcal{G}_1$ is a finite subset of $\mathcal{G}$ covering $[-n,n]$. As $\mathcal{G}$ is arbitrary, $[-n,n]$ is compact and the induction starts.

(ii) For the inductive step to $r+1$, regard the closed interval $F = [-\mathbf{n}, \mathbf{n}]$, taken in $\mathbb{R}^{r+1}$, as the product of the closed interval $E = [-\mathbf{n}, \mathbf{n}]$, taken in $\mathbb{R}^r$, with the closed interval $[-n,n] \subseteq \mathbb{R}$; by the inductive hypothesis, both E and $[-n,n]$ are compact. Let $\mathcal{G}$ be a family of open subsets of $\mathbb{R}^{r+1}$ covering F. Write $\mathcal{H}$ for the family of open subsets H of $\mathbb{R}^r$ such that $H \times [-n,n]$ is covered by a finite subfamily of $\mathcal{G}$. Then $E \subseteq \bigcup \mathcal{H}$. **P** Take $x \in E$. Set

$$\mathcal{U}_x = \{U : U \subseteq \mathbb{R} \text{ is open, } \exists\ G \in \mathcal{G}, \text{ open } H \subseteq \mathbb{R}^r,\ x \in H \text{ and } H \times U \subseteq G\}.$$

Then $\mathcal{U}_x$ is a family of open subsets of $\mathbb{R}$. If $\xi \in [-n,n]$, there is a $G \in \mathcal{G}$ containing (x,ξ); there is a $\delta > 0$ such that $U((x,\xi),\delta) \subseteq G$; now $U(x,\frac{1}{2}\delta)$ and $U(\xi,\frac{1}{2}\delta)$ are open sets in $\mathbb{R}^r$, $\mathbb{R}$ respectively and

$$U(x,\tfrac{1}{2}\delta) \times U(\xi,\tfrac{1}{2}\delta) \subseteq U((x,\xi),\delta) \subseteq G,$$

so $U(\xi,\frac{1}{2}\delta) \in \mathcal{U}_x$. As ξ is arbitrary, $\mathcal{U}_x$ is an open cover of $[-n,n]$ in $\mathbb{R}$. By (i), it has a finite subcover $U_0, \dots, U_k$ say. For each $j \leq k$ we can find H_j, G_j such that H_j is an open subset of $\mathbb{R}^r$ containing x and $H_j \times U_j \subseteq G_j \in \mathcal{G}$. Now set $H = \bigcap_{j \leq k} H_j$. This is an open subset of $\mathbb{R}^r$ containing x, and $H \times [-n,n] \subseteq \bigcup_{j \leq n} G_j$ is covered by a finite subfamily of $\mathcal{G}$. So $x \in H \in \mathcal{H}$. As x is arbitrary, $\mathcal{H}$ covers E. **Q**

(iii) Now the inductive hypothesis tells us that E is compact, so there is a finite subfamily $\mathcal{H}_0$ of $\mathcal{H}$ covering E. For each $H \in \mathcal{H}_0$ let $\mathcal{G}_H$ be a finite subfamily of $\mathcal{G}$ covering $H \times [-n,n]$. Then $\bigcup_{H \in \mathcal{H}_0} \mathcal{G}_H$ is a finite subfamily of $\mathcal{G}$ covering $E \times [-n,n] = F$. As $\mathcal{G}$ is arbitrary, F is compact and the induction proceeds.

(iv) Thus the interval $[-\mathbf{n}, \mathbf{n}]$ is compact in $\mathbb{R}^r$ for every r, n. Now suppose that K is a closed bounded set in $\mathbb{R}^r$. Then there is an $n \in \mathbb{N}$ such that $K \subseteq [-\mathbf{n}, \mathbf{n}]$, that is, $K = K \cap [-\mathbf{n}, \mathbf{n}]$. As K is closed and $[-\mathbf{n}, \mathbf{n}]$ is compact, K is compact, by 2A2Ea.

This completes the proof.

2A2G Corollary If $\phi : D \to \mathbb{R}$ is continuous, where $D \subseteq \mathbb{R}^r$, and $K \subseteq D$ is a non-empty compact set, then ϕ is bounded and attains its bounds on K.

proof By 2A2Eb, $\phi[K]$ is compact; by 2A2F it is closed and bounded. To say that $\phi[K]$ is bounded is just to say that ϕ is bounded on K. Because $\phi[K]$ is a non-empty bounded set, it has an infimum a and a supremum b; now both belong to $\overline{\phi[K]}$ (by the criterion 2A2B(ii), or otherwise); because $\phi[K]$ is closed, both belong to $\phi[K]$, that is, ϕ attains its bounds.

2A2H Lim sup and lim inf revisited In §1A3 I briefly discussed $\limsup_{n\to\infty} a_n$, $\liminf_{n\to\infty} a_n$ for real sequences $\langle a_n \rangle_{n\in\mathbb{N}}$. In this volume we need the notion of $\limsup_{\delta\downarrow 0} f(\delta)$, $\liminf_{\delta\downarrow 0} f(\delta)$ for real functions f. I say that $\limsup_{\delta\downarrow 0} f(\delta) = u \in [-\infty,\infty]$ if (i) for every $v > u$ there is an $\eta > 0$ such that $f(\delta)$ is defined and less than or equal to v for every $\delta \in]0,\eta]$ (ii) for every $v < u$ and $\eta > 0$ there is a $\delta \in]0,\eta]$ such that $f(\delta)$ is defined and greater than or equal to v. Similarly, $\liminf_{\delta\downarrow 0} f(\delta) = u \in [-\infty,\infty]$ if (i) for every $v < u$ there is an $\eta > 0$ such that $f(\delta)$ is defined and greater than or equal to v for every $\delta \in]0,\eta]$ (ii) for every $v > u$ and $\eta > 0$ there is an $\delta \in]0,\eta]$ such that $f(\delta)$ is defined and less than or equal to v.

2A2I In the one-dimensional case, we have a particularly simple description of the open sets.

Proposition If $G \subseteq \mathbb{R}$ is any open set, it is expressible as the union of a countable disjoint family of open intervals.

proof For x, $y \in G$ write $x \sim y$ if either $x \leq y$ and $[x,y] \subseteq G$ or $y \leq x$ and $[y,x] \subseteq G$. It is easy to check that $\sim$ is an equivalence relation on G. Let $\mathcal{C}$ be the set of equivalence classes under $\sim$. Then $\mathcal{C}$ is a partition of G. Now every $C \in \mathcal{C}$ is an open interval. **P** Set $a = \inf C$, $b = \sup C$ (allowing $a = -\infty$ and/or $b = \infty$ if C is unbounded). If $a < x < b$, there are y, $z \in C$ such that $y \leq x \leq z$, so that $[y,x] \subseteq [y,z] \subseteq G$ and $y \sim x$ and $x \in C$; thus $]a,b[\subseteq C$. If $x \in C$, there is an open interval I containing x and included in G; since $x \sim y$ for every $y \in I$, $I \subseteq C$; so

$$a \leq \inf I < x < \sup I \leq b$$

and $x \in]a,b[$. Thus $C =]a,b[$ is an open interval. **Q**

To see that $\mathcal{C}$ is countable, observe that every member of $\mathcal{C}$ contains a member of $\mathbb{Q}$, so that we have a surjective function from a subset of $\mathbb{Q}$ onto $\mathcal{C}$, and $\mathcal{C}$ is countable (1A1E).

2A3 General topology

At various points – principally §§245-247, but also for certain ideas in Chapter 27 – we need to know something about non-metrizable topologies. I must say that you should probably take the time to look at some book on elementary functional analysis which has the phrases 'weak compactness' or 'weakly compact' in the index. But I can list here the concepts actually used in this volume, in a good deal less space than any orthodox, complete treatment would employ.

2A3A Topologies First we need to know what a 'topology' is. If X is any set, a **topology** on X is a family $\mathfrak{T}$ of subsets of X such that (i) $\emptyset$, $X \in \mathfrak{T}$ (ii) if G, $H \in \mathfrak{T}$ then $G \cap H \in \mathfrak{T}$ (iii) if $\mathcal{G} \subseteq \mathfrak{T}$ then $\bigcup \mathcal{G} \in \mathfrak{T}$ (cf. 1A2B). The pair $(X, \mathfrak{T})$ is now a **topological space**. In this context, members of $\mathfrak{T}$ are called **open** and their complements (in X) are called **closed** (cf. 1A2E-1A2F).

2A3B Continuous functions (a) If $(X, \mathfrak{T})$ and $(Y, \mathfrak{S})$ are topological spaces, a function $\phi : X \to Y$ is **continuous** if $\phi^{-1}[G] \in \mathfrak{T}$ for every $G \in \mathfrak{S}$. (By 2A2Ca above, this is consistent with the ϵ-δ definition of continuity for functions from one Euclidean space to another. See also 2A3H below.)

(b) If $(X, \mathfrak{T})$, $(Y, \mathfrak{S})$ and $(Z, \mathfrak{U})$ are topological spaces and $\phi : X \to Y$ and $\psi : Y \to Z$ are continuous, then $\psi\phi : X \to Z$ is continuous. **P** If $G \in \mathfrak{U}$ then $\psi^{-1}[G] \in \mathfrak{S}$ so $(\psi\phi)^{-1}[G] = \phi^{-1}[\psi^{-1}[G]] \in \mathfrak{T}$. **Q**

(c) If $(X, \mathfrak{T})$ is a topological space, a function $f : X \to \mathbb{R}$ is continuous iff $\{x : a < f(x) < b\}$ is open whenever $a < b$ in $\mathbb{R}$. **P (i)** Every interval $]a, b[$ is open in $\mathbb{R}$, so if f is continuous its inverse image $\{x : a < f(x) < b\}$ must be open. **(ii)** Suppose that $f^{-1}[]a, b[]$ is open whenever $a < b$, and let $H \subseteq \mathbb{R}$ be any open set. By the definition of 'open' set in $\mathbb{R}$ (1A2A),

$$H = \bigcup\{]y - \delta, y + \delta[: y \in \mathbb{R},\, \delta > 0,\,]y - \delta, y + \delta[\subseteq H\},$$

so

$$f^{-1}[H] = \bigcup\{f^{-1}[]y - \delta, y + \delta[] : y \in \mathbb{R},\, \delta > 0,\,]y - \delta, y + \delta[\subseteq H\}$$

is a union of open sets in X, therefore open. **Q**

(d) If $r \ge 1$, $(X, \mathfrak{T})$ is a topological space, and $\phi : X \to \mathbb{R}^r$ is a function, then ϕ is continuous iff $\phi_i : X \to \mathbb{R}$ is continuous for each $i \le r$, where $\phi(x) = (\phi_1(x), \dots, \phi_r(x))$ for each $x \in X$. **P (i)** Suppose that ϕ is continuous. For $i \le r$, $y = (\eta_1, \dots, \eta_r) \in \mathbb{R}^r$, set $\pi_i(y) = \eta_i$. Then $|\pi_i(y) - \pi_i(z)| \le \|y - z\|$ for all y, $z \in \mathbb{R}^r$ so $\pi_i : \mathbb{R}^r \to \mathbb{R}$ is continuous. Consequently $\phi_i = \pi_i\phi$ is continuous, by (b) above. **(ii)** Suppose that every ϕ_i is continuous, and that $H \subseteq \mathbb{R}^r$ is open. Set

$$\mathcal{G} = \{G : G \subseteq X \text{ is open}, G \subseteq \phi^{-1}[H]\}.$$

Then $G_0 = \bigcup \mathcal{G}$ is open, and $G_0 \subseteq \phi^{-1}[H]$. But suppose that x_0 is any point of $\phi^{-1}[H]$. Then there is a $\delta > 0$ such that $U(\phi(x_0), \delta) \subseteq H$, because H is open and contains $\phi(x_0)$. For $1 \le i \le r$ set $V_i = \{x : \phi_i(x_0) - \frac{\delta}{\sqrt{r}} < \phi_i(x) < \phi_i(x_0) + \frac{\delta}{\sqrt{r}}\}$; then V_i is the inverse image of an open set under the continuous map ϕ_i, so is open. Set $G = \bigcap_{i \le r} V_i$. Then G is open (using (ii) of the definition 2A3A), $x_0 \in G$, and $\|\phi(x) - \phi(x_0\| < \delta$ for every $x \in G$, so $G \subseteq \phi^{-1}[H]$, $G \in \mathcal{G}$ and $x_0 \in G_0$. This shows that $\phi^{-1}[H] = G_0$ is open. As H is arbitrary, ϕ is continuous. **Q**

(e) If $(X, \mathfrak{T})$ is a topological space, $f_1, \dots, f_r$ are continuous functions from X to $\mathbb{R}$, and $h : \mathbb{R}^r \to \mathbb{R}$ is continuous, then $h(f_1, \dots, f_r) : X \to \mathbb{R}$ is continuous. **P** Set $\phi(x) = (f_1(x), \dots, f_r(x)) \in \mathbb{R}^r$ for $x \in X$. By (d), ϕ is continuous, so by 2A3Bb $h(f_1, \dots, f_r) = h\phi$ is continuous. **Q** In particular, $f + g$, $f \times g$ and $f - g$ are continuous for all continuous functions f, $g : X \to \mathbb{R}$.

(f) If $(X, \mathfrak{T})$ and $(Y, \mathfrak{S})$ are topological spaces and $\phi : X \to Y$ is a continuous function, then $\phi^{-1}[F]$ is closed in X for every closed set $F \subseteq Y$. (For $X \setminus \phi^{-1}[F] = \phi^{-1}[Y \setminus F]$ is open.)

2A3C Subspace topologies If $(X, \mathfrak{T})$ is a topological space and $D \subseteq X$, then $\mathfrak{T}_D = \{G \cap D : G \in \mathfrak{T}\}$ is a topology on D. **P** (i) $\emptyset = \emptyset \cap D$ and $D = X \cap D$ belong to $\mathfrak{T}_D$. (ii) If G, $H \in \mathfrak{T}_D$ there are G', $H' \in \mathfrak{T}$ such that $G = G' \cap D$, $H = H' \cap D$; now $G \cap H = G' \cap H' \cap D \in \mathfrak{T}_D$. (iii) If $\mathcal{G} \subseteq \mathfrak{T}_D$ set $\mathcal{H} = \{H : H \in \mathfrak{T}, H \cap D \in \mathcal{G}\}$; then $\bigcup \mathcal{G} = (\bigcup \mathcal{H}) \cap D \in \mathfrak{T}_D$. **Q**

$\mathfrak{T}_D$ is called the **subspace topology** on D, or the topology on D **induced** by $\mathfrak{T}$. If $(Y, \mathfrak{S})$ is another topological space, and $\phi : X \to Y$ is $(\mathfrak{T}, \mathfrak{S})$-continuous, then $\phi \restriction D : D \to Y$ is $(\mathfrak{T}_D, \mathfrak{S})$-continuous. (For if $H \in \mathfrak{S}$ then

$$(\phi \restriction D)^{-1}[H] = D \cap \phi^{-1}[H] \in \mathfrak{T}_D.)$$

2A3D Closures and interiors (a) In the proof of 2A3Bd I have already used the following idea. Let $(X,\mathfrak{T})$ be any topological space and A any subset of X. Write

$$\operatorname{int} A = \bigcup\{G : G \in \mathfrak{T},\, G \subseteq A\}.$$

Then $\operatorname{int} A$ is an open set, being a union of open sets, and is of course included in A; it must be the largest open set included in A, and is called the **interior** of A.

(b) Because a set is closed iff its complement is open, we have a complementary notion:

$$\begin{aligned}\overline{A} &= \bigcap\{F : F \text{ is closed, } A \subseteq F\}\\ &= X \setminus \bigcup\{X \setminus F : F \text{ is closed, } A \subseteq F\}\\ &= X \setminus \bigcup\{G : G \text{ is open, } A \cap G = \emptyset\}\\ &= X \setminus \bigcup\{G : G \text{ is open, } G \subseteq X \setminus A\} = X \setminus \operatorname{int}(X \setminus A).\end{aligned}$$

$\overline{A}$ is closed (being the complement of an open set) and is the smallest closed set including A; it is called the **closure** of A. (Compare 2A2A.) Because the union of two closed sets is closed (cf. 1A2Fc), $\overline{A \cup B} = \overline{A} \cup \overline{B}$ for all A, $B \subseteq X$.

(c) There are innumerable ways of looking at these concepts; a useful description of the closure of a set is

$$\begin{aligned} x \in \overline{A} &\iff x \notin \operatorname{int}(X \setminus A)\\ &\iff \text{there is no open set containing } x \text{ and included in } X \setminus A\\ &\iff \text{every open set containing } x \text{ meets } A.\end{aligned}$$

2A3E Hausdorff topologies (a) The concept of 'topological space' is so widely drawn, and so widely applicable, that a vast number of different types of topological space have been studied. For this volume we shall not need much of the (very extensive) vocabulary which has been developed to describe this variety. But one useful word (and one of the most important concepts) is that of 'Hausdorff space'; a topological space X is **Hausdorff** if for all distinct x, $y \in X$ there are disjoint open sets G, $H \subseteq X$ such that $x \in G$ and $y \in H$.

(b) In a Hausdorff space X, finite sets are closed. **P** If $z \in X$, then for any $x \in X \setminus \{z\}$ there is an open set containing x but not z, so $X \setminus \{z\}$ is open and $\{z\}$ is closed. So a finite set is a finite union of closed sets and is therefore closed. **Q**

2A3F Pseudometrics Many important topologies (not all!) can be defined by families of pseudometrics; it will be useful to have a certain amount of technical skill with these.

(a) Let X be a set. A **pseudometric** on X is a function $\rho : X \times X \to [0,\infty[$ such that

$\rho(x,z) \le \rho(x,y) + \rho(y,z)$ for all x, y, $z \in X$

(the 'triangle inequality';)

$\rho(x,y) = \rho(y,x)$ for all x, $y \in X$;

$\rho(x,x) = 0$ for all $x \in X$.

A **metric** is a pseudometric ρ satisfying the further condition

if $\rho(x,y) = 0$ then $x = y$.

(b) Examples (i) For x, $y \in \mathbb{R}$, set $\rho(x,y) = |x-y|$; then ρ is a metric on $\mathbb{R}$ (the 'usual metric' on $\mathbb{R}$).

(ii) For x, $y \in \mathbb{R}^r$, where $r \ge 1$, set $\rho(x,y) = \|x-y\|$, defining $\|z\| = \sqrt{\sum_{i=1}^r \zeta_i^2}$, as usual. Then ρ is a metric, the **Euclidean metric** on $\mathbb{R}^r$. (The triangle inequality for ρ comes from Cauchy's inequality in 1A2C: if x, y, $z \in \mathbb{R}^r$, then

$$\rho(x,z) = \|x-z\| = \|(x-y)+(y-z)\| \le \|x-y\| + \|y-z\| = \rho(x,y) + \rho(y,z).$$

The other required properties of ρ are elementary. Compare 2A4Bb below.)

(iii) For an example of a pseudometric which is not a metric, take $r \ge 2$ and define $\rho : \mathbb{R}^r \times \mathbb{R}^r \to [0,\infty[$ by setting $\rho(x,y) = |\xi_1 - \eta_1|$ whenever $x = (\xi_1,\dots,\xi_r)$, $y = (\eta_1,\dots,\eta_r) \in \mathbb{R}^r$.

(c) Now let X be a set and P a non-empty family of pseudometrics on X. Let $\mathfrak{T}$ be the family of those subsets G of X such that for every $x \in G$ there are $\rho_0, \dots, \rho_n \in \mathrm{P}$ and $\delta > 0$ such that

$$U(x; \rho_0, \dots, \rho_n; \delta) = \{y : y \in X, \max_{i \le n} \rho_i(y, x) < \delta\} \subseteq G.$$

Then $\mathfrak{T}$ is a topology on X.

P (Compare 1A2B.) (i) $\emptyset \in \mathfrak{T}$ because the condition is vacuously satisfied. $X \in \mathfrak{T}$ because $U(x; \rho; 1) \subseteq X$ for any $x \in X$, $\rho \in \mathrm{P}$. (ii) If G, $H \in \mathfrak{T}$ and $x \in G \cap H$, take $\rho_0, \dots, \rho_m, \rho'_0, \dots, \rho'_n \in \mathrm{P}$, δ, $\delta' > 0$ such that $U(x; \rho_0, \dots, \rho_m; \delta) \subseteq G$, $U(x; \rho'_0, \dots, \rho'_n; \delta') \subseteq G$; then

$$U(x; \rho_0, \dots, \rho_m, \rho'_0, \dots, \rho'_n; \min(\delta, \delta')) \subseteq G \cap H.$$

As x is arbitrary, $G \cap H \in \mathfrak{T}$. (iii) If $\mathcal{G} \subseteq \mathfrak{T}$ and $x \in \bigcup \mathcal{G}$, there is a $G \in \mathcal{G}$ such that $x \in G$; now there are $\rho_0, \dots, \rho_n \in \mathrm{P}$ and $\delta > 0$ such that

$$U(x; \rho_0, \dots, \rho_n; \delta) \subseteq G \subseteq \bigcup \mathcal{G}.$$

As x is arbitrary, $\bigcup \mathcal{G} \in \mathfrak{T}$. **Q**

$\mathfrak{T}$ is the **topology defined by** P.

(d) You may wish to have a convention to deal with the case in which P is the empty set; in this case the topology on X defined by P is $\{\emptyset, X\}$.

(e) In many important cases, P is upwards-directed in the sense that for any ρ_1, $\rho_2 \in \mathrm{P}$ there is a $\rho \in \mathrm{P}$ such that $\rho_i(x, y) \le \rho(x, y)$ for all x, $y \in X$ and both i. In this case, of course, any set $U(x; \rho_0, \dots, \rho_n; \delta)$, where $\rho_0, \dots, \rho_n \in \mathrm{P}$, includes some set of the form $U(x; \rho; \delta)$, where $\rho \in \mathrm{P}$. Consequently, for instance, a set $G \subseteq X$ is open iff for every $x \in G$ there are $\rho \in \mathrm{P}$, $\delta > 0$ such that $U(x; \rho; \delta) \subseteq G$.

(f) A topology $\mathfrak{T}$ is **metrizable** if it is the topology defined by a family P consisting of a single metric. Thus the **Euclidean topology** on $\mathbb{R}^r$ is the metrizable topology defined by $\{\rho\}$, where ρ is the metric of (b-ii) above.

2A3G Proposition Let X be a set with a topology defined by a non-empty set P of pseudometrics on X. Then $U(x; \rho_0, \dots, \rho_n; \epsilon)$ is open for all $x \in X$, $\rho_0, \dots, \rho_n \in \mathrm{P}$ and $\epsilon > 0$.

proof (Compare 1A2D.) Take $y \in U(x; \rho_0, \dots, \rho_n; \epsilon)$. Set

$$\eta = \max_{i \le n} \rho_i(y, x), \quad \delta = \epsilon - \eta > 0.$$

If $z \in U(y; \rho_0, \dots, \rho_n; \delta)$ then

$$\rho_i(z, x) \le \rho_i(z, y) + \rho_i(y, x) < \delta + \eta = \epsilon$$

for each $i \le n$, so $U(y; \rho_0, \dots, \rho_n; \delta) \subseteq U(x; \rho_0, \dots, \rho_n; \epsilon)$. As y is arbitrary, $U(x; \rho_0, \dots, \rho_n; \epsilon)$ is open.

2A3H Now we have a result corresponding to 2A2Ca, describing continuous functions between topological spaces defined by families of pseudometrics.

Proposition Let X and Y be sets; let P be a non-empty family of pseudometrics on X, and Θ a non-empty family of pseudometrics on Y; let $\mathfrak{T}$ and $\mathfrak{S}$ be the corresponding topologies. Then a function $\phi : X \to Y$ is continuous iff whenever $x \in X$, $\theta \in \Theta$ and $\epsilon > 0$, there are $\rho_0, \dots, \rho_n \in \mathrm{P}$ and $\delta > 0$ such that $\theta(\phi(y), \phi(x)) \le \epsilon$ whenever $y \in X$ and $\max_{i \le n} \rho_i(y, x) \le \delta$.

proof (a) Suppose that ϕ is continuous; take $x \in X$, $\theta \in \Theta$ and $\epsilon > 0$. By 2A3G, $U(\phi(x); \theta; \epsilon) \in \mathfrak{S}$. So $G = \phi^{-1}[U(\phi(x); \theta; \epsilon)] \in \mathfrak{T}$. Now $x \in G$, so there are $\rho_0, \dots, \rho_n \in \mathrm{P}$ and $\delta > 0$ such that $U(x; \rho_0, \dots, \rho_n; \delta) \subseteq G$. In this case $\theta(\phi(y), \phi(x)) \le \epsilon$ whenever $y \in X$ and $\max_{i \le n} \rho_i(y, x) \le \frac{1}{2}\delta$. As x, θ and ϵ are arbitrary, ϕ satisfies the condition.

(b) Suppose ϕ satisfies the condition. Take $H \in \mathfrak{S}$ and consider $G = \phi^{-1}[H]$. If $x \in G$, then $\phi(x) \in H$, so there are $\theta_0, \dots, \theta_n \in \Theta$ and $\epsilon > 0$ such that $U(\phi(x); \theta_0, \dots, \theta_n; \epsilon) \subseteq H$. For each $i \le n$ there are $\rho_{i0}, \dots, \rho_{i,m_i} \in \mathrm{P}$ and $\delta_i > 0$ such that $\theta(\phi(y), \phi(x)) \le \frac{1}{2}\epsilon$ whenever $y \in X$ and $\max_{j \le m_i} \rho_{ij}(y, x) \le \delta_i$. Set $\delta = \min_{i \le n} \delta_i > 0$; then

$$U(x; \rho_{00}, \dots, \rho_{0,m_0}, \dots, \rho_{n0}, \dots, \rho_{n,m_n}; \delta) \subseteq G.$$

As x is arbitrary, $G \in \mathfrak{T}$. As H is arbitrary, ϕ is continuous.

2A3I Remarks (a) If P is upwards-directed, the condition simplifies to: for every $x \in X$, $\theta \in \Theta$ and $\epsilon > 0$, there are $\rho \in \mathrm{P}$ and $\delta > 0$ such that $\theta(\phi(y), \phi(x)) \le \epsilon$ whenever $y \in X$ and $\rho(y, x) \le \delta$.

(b) Suppose we have a set X and two non-empty families P, Θ of pseudometrics on X, generating topologies $\mathfrak{T}$ and $\mathfrak{S}$ on X. Then $\mathfrak{S} \subseteq \mathfrak{T}$ iff the identity map ϕ from X to itself is a continuous function when regarded as a map from $(X, \mathfrak{T})$ to $(X, \mathfrak{S})$, because this will mean that $G = \phi^{-1}[G]$ belongs to $\mathfrak{T}$ whenever $G \in \mathfrak{S}$. Applying the proposition above to ϕ, we see that this happens iff for every $\theta \in \Theta$, $x \in X$ and $\epsilon > 0$ there are $\rho_0, \dots, \rho_n \in \mathrm{P}$ and $\delta > 0$ such that $\theta(y, x) \le \epsilon$ whenever $y \in X$ and $\max_{i \le n} \rho_i(y, x) \le \delta$. Similarly, reversing the roles of P and Θ, we get a criterion for when $\mathfrak{T} \subseteq \mathfrak{S}$, and putting the two together we obtain a criterion to determine when $\mathfrak{T} = \mathfrak{S}$.

2A3J Subspaces: Proposition If X is a set, P a non-empty family of pseudometrics on X defining a topology $\mathfrak{T}$ on X, and $D \subseteq X$, then

(a) for every $\rho \in \mathrm{P}$, the restriction $\rho^{(D)}$ of ρ to $D \times D$ is a pseudometric on D;

(b) the topology defined by $\mathrm{P}_D = \{\rho^{(D)} : \rho \in \mathrm{P}\}$ on D is precisely the subspace topology $\mathfrak{T}_D$ described in 2A3C.

proof (a) is just a matter of reading through the definition in 2A3Fa. For (b), we have to think for a moment.

(i) Suppose that G belongs to the topology defined by P_D. Set

$$\mathcal{H} = \{H : H \in \mathfrak{T},\ H \cap D \subseteq G\},$$

$$H^* = \bigcup \mathcal{H} \in \mathfrak{T}, \quad G^* = H^* \cap D \in \mathfrak{T}_D;$$

then $G^* \subseteq G$. On the other hand, if $x \in G$, then there are $\rho_0, \dots, \rho_n \in \mathrm{P}$ and $\delta > 0$ such that

$$U(x; \rho_0^{(D)}, \dots, \rho_n^{(D)}; \delta) = \{y : y \in D,\ \max_{i \le n} \rho_i^{(D)}(y, x) < \delta\} \subseteq G.$$

Consider

$$H = U(x; \rho_0, \dots, \rho_n; \delta) = \{y : y \in X,\ \max_{i \le n} \rho_i(y, x) < \delta\} \subseteq X.$$

Evidently

$$H \cap D = U(x; \rho_0^{(D)}, \dots, \rho_n^{(D)}; \delta) \subseteq G.$$

Also $H \in \mathfrak{T}$. So $H \in \mathcal{H}$ and

$$x \in H \cap D \subseteq H^* \cap D = G^*.$$

Thus $G = G^* \in \mathfrak{T}_D$.

(ii) Now suppose that $G \in \mathfrak{T}_D$. Then there is an $H \in \mathfrak{T}$ such that $G = H \cap D$. Consider the identity map $\phi : D \to X$, defined by saying that $\phi(x) = x$ for every $x \in D$. ϕ obviously satisfies the criterion of 2A3H, if we endow D with P_D and X with P, because $\rho(\phi(x), \phi(y)) = \rho^{(D)}(x, y)$ whenever x, $y \in D$ and $\rho \in \mathrm{P}$; so ϕ must be continuous for the associated topologies, and $\phi^{-1}[H]$ must belong to the topology defined by P_D. But $\phi^{-1}[H] = G$. Thus every set in $\mathfrak{T}_D$ belongs to the topology defined by P_D, and the two topologies are the same, as claimed.

2A3K Closures and interiors Let X be a set, P a non-empty family of pseudometrics on X and $\mathfrak{T}$ the topology defined by P.

(a) For any $A \subseteq X$ and $x \in X$,

$$\begin{aligned} x \in \operatorname{int} A &\iff \text{there is an open set included in } A \text{ containing } x \\ &\iff \text{there are } \rho_0, \dots, \rho_n \in \mathrm{P},\ \delta > 0 \text{ such that } U(x; \rho_0, \dots, \rho_n; \delta) \subseteq A. \end{aligned}$$

(b) For any $A \subseteq X$ and $x \in X$, $x \in \overline{A}$ iff $U(x; \rho_0, \dots, \rho_n; \delta) \cap A \ne \emptyset$ for every $\rho_0, \dots, \rho_n \in \mathrm{P}$ and $\delta > 0$. (Compare 2A2B(ii), 2A3Dc.)

2A3L Hausdorff topologies Recall that a topology $\mathfrak{T}$ is Hausdorff if any two points can be separated by open sets (2A3E). Now a topology defined on a set X by a non-empty family P of pseudometrics is Hausdorff iff for any two different points x, y of X there is a $\rho \in \mathrm{P}$ such that $\rho(x, y) > 0$. **P** **(i)** Suppose that the topology is Hausdorff and that x, y are distinct points in X. Then there is an open set G containing x but not containing y. Now there are $\rho_0, \dots, \rho_n \in \mathrm{P}$ and $\delta > 0$ such that $U(x; \rho_0), \dots, \rho_n; \delta) \subseteq G$, in which case $\rho_i(y, x) \ge \delta > 0$ for some $i \le n$. **(ii)** If P satisfies the condition, and x, y are distinct points of X, take $\rho \in \mathrm{P}$ such that $\rho(x, y) > 0$, and set $\delta = \frac{1}{2}\rho(x, y)$. Then $U(x; \rho; \delta)$ and $U(y; \rho; \delta)$ are disjoint (because if $z \in X$, then

$$\rho(z, x) + \rho(z, y) \ge \rho(x, y) = 2\delta,$$

so at least one of $\rho(z,x)$, $\rho(z,y)$ is greater than or equal to δ), and they are open sets containing x, y respectively. As x and y are arbitrary, the topology is Hausdorff. **Q**

In particular, metrizable topologies are Hausdorff.

2A3M Convergence of sequences (a) If $(X,\mathfrak{T})$ is any topological space, and $\langle x_n\rangle_{n\in\mathbb{N}}$ is a sequence in X, we say that $\langle x_n\rangle_{n\in\mathbb{N}}$ **converges** to $x\in X$, or that x is a **limit** of $\langle x_n\rangle_{n\in\mathbb{N}}$, or $\langle x_n\rangle_{n\in\mathbb{N}}\to x$, if for every open set G containing x there is an $n_0\in\mathbb{N}$ such that $x_n\in G$ for every $n\ge n_0$.

(b) Warning In general topological spaces, it is possible for a sequence to have more than one limit, and we cannot safely write $x=\lim_{n\to\infty}x_n$. But in Hausdorff spaces, this does not occur. **P** If $\mathfrak{T}$ is Hausdorff, and x, y are distinct points of X, there are disjoint open sets G, H such that $x\in G$ and $y\in H$. If now $\langle x_n\rangle_{n\in\mathbb{N}}$ converges to x, there is an n_0 such that $x_n\in G$ for every $n\ge n_0$, so $x_n\notin H$ for every $n\ge n_0$, and $\langle x_n\rangle_{n\in\mathbb{N}}$ cannot converge to y. **Q** In particular, a sequence in a metrizable space can have at most one limit.

(c) Let X be a set, and P a non-empty family of pseudometrics on X, generating a topology $\mathfrak{T}$; let $\langle x_n\rangle_{n\in\mathbb{N}}$ be a sequence in X and $x\in X$. Then $\langle x_n\rangle_{n\in\mathbb{N}}$ converges to x iff $\lim_{n\to\infty}\rho(x_n,x)=0$ for every $\rho\in\mathrm{P}$. **P (i)** Suppose that $\langle x_n\rangle_{n\in\mathbb{N}}\to x$ and that $\rho\in\mathrm{P}$. Then for any $\epsilon>0$ the set $G=U(x;\rho;\epsilon)$ is an open set containing x, so there is an n_0 such that $x_n\in G$ for every $n\ge n_0$, that is, $\rho(x_n,x)<\epsilon$ for every $n\ge n_0$. As ϵ is arbitrary, $\lim_{n\to\infty}\rho(x_n,x)=0$. **(ii)** If the condition is satisfied, take any open set G containing X. Then there are $\rho_0,\dots,\rho_k\in\mathrm{P}$ and $\delta>0$ such that $U(x;\rho_0,\dots,\rho_k;\delta)\subseteq G$. For each $i\le k$ there is an $n_i\in\mathbb{N}$ such that $\rho_i(x_n,x)<\delta$ for every $n\ge n_i$. Set $n^*=\max(n_0,\dots,n_k)$; then $x_n\in U(x;\rho_0,\dots,\rho_k;\delta)\subseteq G$ for every $n\ge n^*$. As G is arbitrary, $\langle x_n\rangle_{n\in\mathbb{N}}\to x$. **Q**

(d) Let (X,ρ) be a metric space, A a subset of X and $x\in X$. Then $x\in\overline{A}$ iff there is a sequence in A converging to x. **P(i)** If $x\in\overline{A}$, then for every $n\in\mathbb{N}$ we can choose a point $x_n\in A\cap U(x;\rho;2^{-n})$ (2A3Kb); now $\langle x_n\rangle_{n\in\mathbb{N}}\to x$. **(ii)** If $\langle x_n\rangle_{n\in\mathbb{N}}$ is a sequence in A converging to x, then for every open set G containing x there is an n such that $x_n\in G$, so that $A\cap G\ne\emptyset$; by 2A3Dc, $x\in\overline{A}$. **Q**

2A3N Compactness The next concept we need is the idea of 'compactness' in general topological spaces.

(a) If $(X,\mathfrak{T})$ is any topological space, a subset K of X is **compact** if whenever $\mathcal{G}$ is a family in $\mathfrak{T}$ covering K, then there is a finite $\mathcal{G}_0\subseteq\mathcal{G}$ covering K. (Cf. 2A2D. A **warning**: many authors reserve the term 'compact' for Hausdorff spaces.) A set $A\subseteq X$ is **relatively compact** in X if there is a compact subset of X including A. (**Warning!** in non-Hausdorff spaces, this is not the same thing as saying that $\overline{A}$ is compact.)

(b) Just as in 2A2E-2A2G (and the proofs are the same in the general case), we have the following results.

(i) If K is compact and E is closed, then $K\cap E$ is compact.

(ii) If $K\subseteq X$ is compact and $\phi:K\to Y$ is continuous, where $(Y,\mathfrak{S})$ is another topological space, then $\phi[K]$ is a compact subset of Y.

(iii) If $K\subseteq X$ is compact and $\phi:K\to\mathbb{R}$ is continuous, then ϕ is bounded and attains its bounds.

2A3O Cluster points (a) If $(X,\mathfrak{T})$ is a topological space, and $\langle x_n\rangle_{n\in\mathbb{N}}$ is a sequence in X, then a **cluster point** of $\langle x_n\rangle_{n\in\mathbb{N}}$ is an $x\in X$ such that whenever G is an open set containing x and $n\in\mathbb{N}$ then there is a $k\ge n$ such that $x_k\in G$.

(b) Now if $(X,\mathfrak{T})$ is a topological space and $A\subseteq X$ is relatively compact, every sequence $\langle x_n\rangle_{n\in\mathbb{N}}$ in A has a cluster point in X. **P** Let K be a compact subset of X including A. Set

$$\mathcal{G}=\{G:G\in\mathfrak{T},\ \{n:x_n\in G\}\text{ is finite}\}.$$

? If $\mathcal{G}$ covers K, then there is a finite $\mathcal{G}_0\subseteq\mathcal{G}$ covering K. Now

$$\mathbb{N}=\{n:x_n\in A\}=\{n:x_n\in\bigcup\mathcal{G}_0\}=\bigcup_{G\in\mathcal{G}_0}\{n:x_n\in G\}$$

is a finite union of finite sets, which is absurd. **X** Thus $\mathcal{G}$ does not cover K. Take any $x\in K\setminus\bigcup\mathcal{G}$. If $G\in\mathfrak{T}$ and $x\in G$ and $n\in\mathbb{N}$, then $G\notin\mathcal{G}$ so $\{k:x_k\in G\}$ is infinite and there is a $k\ge n$ such that $x_k\in G$. Thus x is a cluster point of $\langle x_n\rangle_{n\in\mathbb{N}}$, as required. **Q**

2A3P Filters In $\mathbb{R}^r$, and more generally in all metrizable spaces, topological ideas can be effectively discussed in terms of convergent sequences. (To be sure, this occasionally necessitates the use of a weak form of the axiom of choice, in order to choose a sequence; but as measure theory without such choices is changed utterly – see Chapter 56 in Volume 5 – there is no point in fussing about them here.) For topological spaces in general, however, sequences are quite inadequate, for very interesting reasons which I shall not enlarge upon. Instead we need to use 'nets' or 'filters'. The latter take a moment's more effort at the beginning, but are then (in my view) much easier to work with, so I describe this method now.

2A3Q Convergent filters (a) Let $(X,\mathfrak{T})$ be a topological space, $\mathcal{F}$ a filter on X (see 2A1I) and x a point of X. We say that $\mathcal{F}$ is **convergent** to x, or that x is a **limit** of $\mathcal{F}$, and write $\mathcal{F}\to x$, if every open set containing x belongs to $\mathcal{F}$.

(b) Let $(X,\mathfrak{T})$ and $(Y,\mathfrak{S})$ be topological spaces, $\phi:X\to Y$ a continuous function, $x\in X$ and $\mathcal{F}$ a filter on X converging to x. Then $\phi[[\mathcal{F}]]$ (as defined in 2A1Ib) converges to $\phi(x)$ (because $\phi^{-1}[G]$ is an open set containing x whenever G is an open set containing $\phi(x)$).

2A3R Now we have the following characterization of compactness.

Theorem Let X be a topological space, and K a subset of X. Then K is compact iff every ultrafilter on X containing K has a limit in K.

proof (a) Suppose that K is compact and that $\mathcal{F}$ is an ultrafilter on X containing K. Set

$$\mathcal{G}=\{G:G\subseteq X\text{ is open, }X\setminus G\in\mathcal{F}\}.$$

Then the union of any two members of $\mathcal{G}$ belongs to $\mathcal{G}$, so the union of any finite number of members of $\mathcal{G}$ belongs to $\mathcal{G}$; also no member of $\mathcal{G}$ can include K, because $X\setminus K\notin\mathcal{F}$. Because K is compact, it follows that $\mathcal{G}$ cannot cover K. Let x be any point of $K\setminus\bigcup\mathcal{G}$. If G is any open set containing x, then $G\notin\mathcal{G}$ so $X\setminus G\notin\mathcal{F}$; but this means that G must belong to $\mathcal{F}$, because $\mathcal{F}$ is an ultrafilter. As G is arbitrary, $\mathcal{F}\to x$. Thus every ultrafilter on X containing K has a limit in K.

(b) Now suppose that every ultrafilter on X containing K has a limit in K. Let $\mathcal{G}$ be a cover of K by open sets in X. **?** Suppose, if possible, that $\mathcal{G}$ has no finite subcover. Set

$$\mathcal{F}=\{F:\text{there is a finite }\mathcal{G}_0\subseteq\mathcal{G},\ F\cup\bigcup\mathcal{G}_0\supseteq K\}.$$

Then $\mathcal{F}$ is a filter on X. **P** (i) $X\cup\bigcup\emptyset\supseteq K$ so $X\in\mathcal{F}$.

$$\emptyset\cup\bigcup\mathcal{G}_0=\bigcup\mathcal{G}_0\not\supseteq K$$

for any finite $\mathcal{G}_0\subseteq\mathcal{G}$, by hypothesis, so $\emptyset\notin\mathcal{F}$. (ii) If E, $F\in\mathcal{F}$ there are finite sets $\mathcal{G}_1$, $\mathcal{G}_2\subseteq\mathcal{G}$ such that $E\cup\bigcup\mathcal{G}_1$ and $F\cup\bigcup\mathcal{G}_2$ both include K; now $(E\cap F)\cup\bigcup(\mathcal{G}_1\cup\mathcal{G}_2)\supseteq K$ so $E\cap F\in\mathcal{F}$. (iii) If $X\supseteq E\supseteq F\in\mathcal{F}$ then there is a finite $\mathcal{G}_0\subseteq\mathcal{G}$ such that $F\cup\mathcal{G}_0\supseteq K$; now $E\cup\bigcup\mathcal{G}_0\supseteq K$ and $E\in\mathcal{F}$. **Q**

By the Ultrafilter Theorem (2A1O), there is an ultrafilter $\mathcal{F}^*$ on X including $\mathcal{F}$. Of course K itself belongs to $\mathcal{F}$, so $K\in\mathcal{F}^*$. By hypothesis, $\mathcal{F}^*$ has a limit $x\in K$. But now there is a set $G\in\mathcal{G}$ containing x, and $(X\setminus G)\cup G\supseteq K$, so $X\setminus G\in\mathcal{F}\subseteq\mathcal{F}^*$; which means that G cannot belong to $\mathcal{F}^*$, and x cannot be a limit of $\mathcal{F}^*$. **X**

So $\mathcal{G}$ has a finite subcover. As $\mathcal{G}$ is arbitrary, K must be compact.

Remark Note that this theorem depends vitally on the Ultrafilter Theorem and therefore on the axiom of choice.

2A3S Further calculations with filters (a) In general, it is possible for a filter to have more than one limit; but in Hausdorff spaces this does not occur. **P** (Compare 2A3Mb.) If $(X,\mathfrak{T})$ is Hausdorff, and x, y are distinct points of X, there are disjoint open sets G, H such that $x\in G$ and $y\in H$. If now a filter $\mathcal{F}$ on X converges to x, $G\in\mathcal{F}$ so $H\notin\mathcal{F}$ and $\mathcal{F}$ does not converge to y. **Q**

Accordingly we can safely write $x=\lim\mathcal{F}$ when $\mathcal{F}\to x$ in a Hausdorff space.

(b) Now suppose that X is a set, $\mathcal{F}$ is a filter on X, $(Y,\mathfrak{S})$ is a Hausdorff space, $D\in\mathcal{F}$ and $\phi:D\to Y$ is a function. Then we write $\lim_{x\to\mathcal{F}}\phi(x)$ for $\lim\phi[[\mathcal{F}]]$ if this is defined in Y; that is, $\lim_{x\to\mathcal{F}}\phi(x)=y$ iff $\phi^{-1}[H]\in\mathcal{F}$ for every open set H containing y.

In the special case $Y=\mathbb{R}$, $\lim_{x\to\mathcal{F}}\phi(x)=a$ iff $\{x:|\phi(x)-a|\le\epsilon\}\in\mathcal{F}$ for every $\epsilon>0$ (because every open set containing a includes a set of the form $[a-\epsilon,a+\epsilon]$, which in turn includes the open set $]a-\epsilon,a+\epsilon[$).

(c) Suppose that X and Y are sets, $\mathcal{F}$ is a filter on X, Θ is a non-empty family of pseudometrics on Y defining a topology $\mathfrak{S}$ on Y, and $\phi:X\to Y$ is a function. Then the image filter $\phi[[\mathcal{F}]]$ converges to $y\in Y$ iff $\lim_{x\to\mathcal{F}}\theta(\phi(x),y)=0$ in $\mathbb{R}$ for every $\theta\in\Theta$. **P (i)** Suppose that $\phi[[\mathcal{F}]]\to y$. For every $\theta\in\Theta$ and $\epsilon>0$, $U(y;\theta;\epsilon)=\{z:\theta(z,y)<\epsilon\}$ is an open set containing y (2A3G), so belongs to $\phi[[\mathcal{F}]]$, and its inverse image $\{x : 0 \le \theta(\phi(x), y) < \epsilon\}$ belongs to $\mathcal{F}$. As ϵ is arbitrary, $\lim_{x\to\mathcal{F}} \theta(\phi(x), y) = 0$. As θ is arbitrary, ϕ satisfies the condition. **(ii)** Now suppose that $\lim_{x\to\mathcal{F}} \theta(\phi(x), y) = 0$ for every $\theta \in \Theta$. Let G be any open set in Y containing y. Then there are $\theta_0, \dots, \theta_n \in \Theta$ and $\epsilon > 0$ such that

$$U(y; \theta_0, \dots, \theta_n; \epsilon) = \bigcap_{i\le n} U(y; \theta_i; \epsilon) \subseteq G.$$

For each $i \le n$,

$$\phi^{-1}[U(y; \theta_i; \epsilon)] = \{x : \theta(\phi(x), y) < \epsilon\}$$

belongs to $\mathcal{F}$; because $\mathcal{F}$ is closed under finite intersections, so do $\phi^{-1}[U(y; \theta_0, \dots, \theta_n; \epsilon)]$ and its superset $\phi^{-1}[G]$. Thus $G \in \phi[[\mathcal{F}]]$. As G is arbitrary, $\phi[[\mathcal{F}]] \to y$. **Q**

(d) In particular, taking $X = Y$ and ϕ the identity map, if X has a topology $\mathfrak{T}$ defined by a non-empty family P of pseudometrics, then a filter $\mathcal{F}$ on X converges to $x \in X$ iff $\lim_{y\to\mathcal{F}} \rho(y, x) = 0$ for every $\rho \in$ P.

(e)(i) If X is any set, $\mathcal{F}$ is an ultrafilter on X, $(Y, \mathfrak{S})$ is a Hausdorff space, and $h : X \to Y$ is a function such that $h[F]$ is relatively compact in Y for some $F \in \mathcal{F}$, then $\lim_{x\to\mathcal{F}} h(x)$ is defined in Y. **P** Let $K \subseteq Y$ be a compact set including $h[F]$. Then $K \in h[[\mathcal{F}]]$, which is an ultrafilter (2A1N), so $h[[\mathcal{F}]]$ has a limit in Y (2A3R), which is $\lim_{x\to\mathcal{F}} h(x)$. **Q**

(ii) If X is any set, $\mathcal{F}$ is an ultrafilter on X, and $h : X \to \mathbb{R}$ is a function such that $h[F]$ is bounded in $\mathbb{R}$ for some set $F \in \mathcal{F}$, then $\lim_{x\to\mathcal{F}} h(x)$ exists in $\mathbb{R}$. **P** $\overline{h[F]}$ is closed and bounded, therefore compact (2A2F), so $h[F]$ is relatively compact and we can use (i). **Q**

(f) The concepts of lim sup, lim inf can be applied to filters. Suppose that $\mathcal{F}$ is a filter on a set X, and that $f : X \to [-\infty, \infty]$ is any function. Then

$$\begin{aligned}\limsup_{x\to\mathcal{F}} f(x) &= \inf\{u : u \in [-\infty, \infty],\ \{x : f(x) \le u\} \in \mathcal{F}\}\\ &= \inf_{F\in\mathcal{F}} \sup_{x\in F} f(x) \in [-\infty, \infty],\end{aligned}$$

$$\begin{aligned}\liminf_{x\to\mathcal{F}} f(x) &= \sup\{u : u \in [-\infty, \infty],\ \{x : f(x) \ge u\} \in \mathcal{F}\}\\ &= \sup_{F\in\mathcal{F}} \inf_{x\in F} f(x).\end{aligned}$$

It is easy to see that, for any two functions f, $g : X \to \mathbb{R}$,

$$\lim_{x\to\mathcal{F}} f(x) = a \quad \text{iff} \quad a = \limsup_{x\to\mathcal{F}} f(x) = \liminf_{x\to\mathcal{F}} f(x),$$

and

$$\limsup_{x\to\mathcal{F}} f(x) + g(x) \le \limsup_{x\to\mathcal{F}} f(x) + \limsup_{x\to\mathcal{F}} g(x),$$

$$\liminf_{x\to\mathcal{F}} f(x) + g(x) \ge \liminf_{x\to\mathcal{F}} f(x) + \liminf_{x\to\mathcal{F}} g(x),$$

$$\liminf_{x\to\mathcal{F}}(-f(x)) = -\limsup_{x\to\mathcal{F}} f(x), \quad \limsup_{x\to\mathcal{F}}(-f(x)) = -\liminf_{x\to\mathcal{F}} f(x),$$

$$\liminf_{x\to\mathcal{F}} cf(x) = c\liminf_{x\to\mathcal{F}} f(x), \quad \limsup_{x\to\mathcal{F}} cf(x) = c\limsup_{x\to\mathcal{F}} f(x)$$

whenever the right-hand-sides are defined in $[-\infty, \infty]$ and $c \ge 0$. So if $a = \lim_{x\to\mathcal{F}} f(x)$ and $b = \lim_{x\to\mathcal{F}}(x)$ exist in $\mathbb{R}$, $\lim_{x\to\mathcal{F}} f(x) + g(x)$ exists and is equal to $a + b$ and $\lim_{x\to\mathcal{F}} cf(x)$ exists and is equal to $c\lim_{x\to\mathcal{F}} f(x)$ for every $c \in \mathbb{R}$.

We also see that if $f : X \to \mathbb{R}$ is such that

for every $\epsilon > 0$ there is an $F \in \mathcal{F}$ such that $\sup_{x\in F} f(x) \le \epsilon + \inf_{x\in F} f(x)$,

then $\limsup_{x\to\mathcal{F}} f(x) \le \epsilon + \liminf_{x\to\mathcal{F}} f(x)$ for every $\epsilon > 0$, so that $\lim_{x\to\mathcal{F}} f(x)$ is defined in $[-\infty, \infty]$.

(g) Note that the standard limits of real analysis can be represented in the form described here. For instance, $\lim_{n\to\infty}$, $\limsup_{n\to\infty}$, $\liminf_{n\to\infty}$ correspond to $\lim_{n\to\mathcal{F}_0}$, $\limsup_{n\to\mathcal{F}_0}$, $\liminf_{n\to\mathcal{F}_0}$ where $\mathcal{F}_0$ is the **Fréchet filter** on $\mathbb{N}$, the filter $\{\mathbb{N} \setminus A : A \subseteq \mathbb{N}$ is finite$\}$ of cofinite subsets of $\mathbb{N}$. Similarly, $\lim_{\delta\downarrow a}$, $\limsup_{\delta\downarrow a}$, $\liminf_{\delta\downarrow a}$ correspond to $\lim_{\delta\to\mathcal{F}}$, $\limsup_{\delta\to\mathcal{F}}$, $\liminf_{\delta\to\mathcal{F}}$ where

$$\mathcal{F} = \{A : A \subseteq \mathbb{R},\ \exists\, h > 0 \text{ such that }]a, a + h] \subseteq A\}.$$

2A3T Product topologies We need some brief remarks concerning topologies on product spaces.

(a) Let $(X,\mathfrak{T})$ and $(Y,\mathfrak{S})$ be topological spaces. Let $\mathfrak{U}$ be the set of subsets U of $X\times Y$ such that for every $(x,y)\in U$ there are $G\in\mathfrak{T}$, $H\in\mathfrak{S}$ such that $(x,y)\in G\times H\subseteq U$. Then $\mathfrak{U}$ is a topology on $X\times Y$. **P** (i) $\emptyset\in\mathfrak{U}$ because the condition for membership of $\mathfrak{U}$ is vacuously satisfied. $X\times Y\in\mathfrak{U}$ because $X\in\mathfrak{T}$, $Y\in\mathfrak{S}$ and $(x,y)\in X\times Y\subseteq X\times Y$ for every $(x,y)\in X\times Y$. (ii) If U, $V\in\mathfrak{U}$ and $(x,y)\in U\cap V$, then there are G, $G'\in\mathfrak{T}$, H, $H'\in\mathfrak{S}$ such that

$$(x,y)\in G\times H\subseteq U,\quad (x,y)\in G'\times H'\subseteq V;$$

now $G\cap G'\in\mathfrak{T}$, $H\cap H'\in\mathfrak{S}$ and

$$(x,y)\in(G\cap G')\times(H\cap H')\subseteq U\cap V.$$

As (x,y) is arbitrary, $U\cap V\in\mathfrak{U}$. (iii) If $\mathcal{U}\subseteq\mathfrak{U}$ and $(x,y)\in\bigcup\mathcal{U}$, then there is a $U\in\mathcal{U}$ such that $(x,y)\in U$; now there are $G\in\mathfrak{T}$, $H\in\mathfrak{S}$ such that $(x,y)\in G\times H\subseteq U\subseteq\bigcup\mathcal{U}$. As (x,y) is arbitrary, $\bigcup\mathcal{U}\in\mathfrak{U}$. **Q**

$\mathfrak{U}$ is called the **product topology** on $X\times Y$.

(b) Suppose, in (a), that $\mathfrak{T}$ and $\mathfrak{S}$ are defined by non-empty families P, Θ of pseudometrics in the manner of 2A3F. Then $\mathfrak{U}$ is defined by the family $\Upsilon=\{\tilde\rho:\rho\in\mathrm{P}\}\cup\{\bar\theta:\theta\in\Theta\}$ of pseudometrics on $X\times Y$, where

$$\tilde\rho((x,y),(x',y'))=\rho(x,x'),\quad \bar\theta((x,y),(x',y'))=\theta(y,y')$$

whenever x, $x'\in X$, y, $y'\in Y$, $\rho\in\mathrm{P}$ and $\theta\in\Theta$.

P (i) Of course you should check that every $\tilde\rho$, $\bar\theta$ is a pseudometric on $X\times Y$.

(ii) If $U\in\mathfrak{U}$ and $(x,y)\in U$, then there are $G\in\mathfrak{T}$, $H\in\mathfrak{S}$ such that $(x,y)\in G\times H\subseteq U$. There are $\rho_0,\dots,\rho_m\in\mathrm{P}$, $\theta_0,\dots,\theta_n\in\Theta$, δ, $\delta'>0$ such that (in the language of 2A3Fc) $U(x;\rho_0,\dots,\rho_m;\delta)\subseteq G$, $U(x;\theta_0,\dots,\theta_n;\delta)\subseteq H$. Now

$$U((x,y);\tilde\rho_0,\dots,\tilde\rho_m,\bar\theta_0,\dots,\bar\theta_n;\min(\delta,\delta'))\subseteq U.$$

As (x,y) is arbitrary, U is open for the topology generated by Υ.

(iii) If $U\subseteq X\times Y$ is open for the topology defined by Υ, take any $(x,y)\in U$. Then there are $\upsilon_0,\dots,\upsilon_k\in\Upsilon$ and $\delta>0$ such that $U((x,y);\upsilon_0,\dots,\upsilon_k;\delta)\subseteq U$. Take $\rho_0,\dots,\rho_m\in\mathrm{P}$ and $\theta_0,\dots,\theta_n\in\Theta$ such that $\{\upsilon_0,\dots,\upsilon_k\}\subseteq\{\tilde\rho_0,\dots,\tilde\rho_m,\bar\theta_0,\dots,\bar\theta_n\}$; then $G=U(x;\rho_0,\dots,\rho_m;\delta)\in\mathfrak{T}$ (2A3G), $H=U(y;\theta_0,\dots,\theta_n;\delta)\in\mathfrak{S}$, and

$$(x,y)\in G\times H=U((x,y);\tilde\rho_0,\dots,\rho_m,\bar\theta_0,\dots,\bar\theta_n;\delta)\subseteq U((x,y);\upsilon_0,\dots,\upsilon_k;\delta)\subseteq U.$$

As (x,y) is arbitrary, $U\in\mathfrak{U}$. This completes the proof that $\mathfrak{U}$ is the topology defined by Υ. **Q**

(c) In particular, the product topology on $\mathbb{R}^r\times\mathbb{R}^s$ is the Euclidean topology if we identify $\mathbb{R}^r\times\mathbb{R}^s$ with $\mathbb{R}^{r+s}$. **P** The product topology is defined by the two pseudometrics υ_1, υ_2, where for x, $x'\in\mathbb{R}^r$ and y, $y'\in\mathbb{R}^s$ I write

$$\upsilon_1((x,y),(x',y'))=\|x-x'\|,\quad \upsilon_2((x,y),(x',y'))=\|y-y'\|$$

(2A3F(b-ii)). Similarly, the Euclidean topology on $\mathbb{R}^r\times\mathbb{R}^s\cong\mathbb{R}^{r+s}$ is defined by the metric ρ, where

$$\rho((x,y),(x',y'))=\|(x-y)-(x',y')\|=\sqrt{\|x-x'\|^2+\|y-y'\|^2}.$$

Now if $(x,y)\in\mathbb{R}^r\times\mathbb{R}^s$ and $\epsilon>0$, then

$$U((x,y);\rho;\epsilon)\subseteq U((x,y);\upsilon_j;\epsilon)$$

for both j, while

$$U((x,y);\upsilon_1,\upsilon_2;\frac{\epsilon}{\sqrt{2}})\subseteq U((x,y);\rho;\epsilon).$$

Thus, as remarked in 2A3Ib, each topology is included in the other, and they are the same. **Q**

2A3U Dense sets (a) If X is a topological space, a set $D\subseteq X$ is **dense** in X if $\overline{D}=X$, that is, if every non-empty open set meets D. More generally, if $D\subseteq A\subseteq X$, then D is dense in A if it is dense for the subspace topology of A (2A3C), that is, if $A\subseteq\overline{D}$.

(b) If $\mathfrak{T}$ is defined by a non-empty family P of pseudometrics on X, then $D\subseteq X$ is dense iff $U(x;\rho_0,\dots,\rho_n;\delta)\cap D\ne\emptyset$ whenever $x\in X$, $\rho_0,\dots,\rho_n\in\mathrm{P}$ and $\delta>0$.

(c) If $(X,\mathfrak{T})$, $(Y,\mathfrak{S})$ are topological spaces, of which Y is Hausdorff (in particular, if (X,ρ) and (Y,θ) are metric spaces), and f, $g:X\to Y$ are continuous functions which agree on some dense subset D of X, then $f=g$. **P?** Suppose, if possible, that there is an $x\in X$ such that $f(x)\ne g(x)$. Then there are open sets G, $H\subseteq Y$ such that $f(x)\in G$, $g(x)\in H$ and $G\cap H=\emptyset$. Now $f^{-1}[G]\cap g^{-1}[H]$ is an open set, containing x and therefore not empty, but it cannot meet D, so $x\notin\overline{D}$ and D is not dense. **XQ**

(d) A topological space is called **separable** if it has a countable dense subset. For instance, $\mathbb{R}^r$ is separable for every $r \ge 1$, since $\mathbb{Q}^r$ is dense.

2A4 Normed spaces

In Chapter 24 I discuss the spaces L^p, for $1 \le p \le \infty$, and describe their most basic properties. These spaces form a group of fundamental examples for the general theory of 'normed spaces', the basis of functional analysis. This is not the book from which you should learn that theory, but once again it may save you trouble if I briefly outline those parts of the general theory which are essential if you are to make sense of the ideas here.

2A4A The real and complex fields While the most important parts of the theory, from the point of view of measure theory, are most effectively dealt with in terms of *real* linear spaces, there are many applications in which *complex* linear spaces are essential. I will therefore use the phrase

$$\text{`}U \text{ is a linear space over } {}^{\mathbb{R}}_{\mathbb{C}}\text{'}$$

to mean that U is either a linear space over the field $\mathbb{R}$ or a linear space over the field $\mathbb{C}$; it being understood that in any particular context all linear spaces considered will be over the same field. In the same way, I will write '$\alpha \in {}^{\mathbb{R}}_{\mathbb{C}}$' to mean that α belongs to whichever is the current underlying field.

2A4B Definitions (a) A **normed space** is a linear space U over ${}^{\mathbb{R}}_{\mathbb{C}}$ together with a **norm**, that is, a functional $\| \; \| : U \to [0, \infty[$ such that

$\|u + v\| \le \|u\| + \|v\|$ for all u, $v \in U$,

$\|\alpha u\| = |\alpha|\|u\|$ for $u \in U$, $\alpha \in {}^{\mathbb{R}}_{\mathbb{C}}$,

$\|u\| = 0$ only when $u = 0$, the zero vector of U.

(Observe that if $u = 0$ (the zero vector) then $0u = u$ (where this 0 is the zero scalar) so that $\|u\| = |0|\|u\| = 0$.)

(b) If U is a normed space, then we have a metric ρ on U defined by saying that $\rho(u, v) = \|u - v\|$ for u, $v \in U$. **P** $\rho(u, v) \in [0, \infty[$ for all u, v because $\|u\| \in [0, \infty[$ for every u. $\rho(u, v) = \rho(v, u)$ for all u, v because $\|v - u\| = |-1|\|u - v\| = \|u - v\|$ for all u, v. If u, v, $w \in U$ then

$$\rho(u, w) = \|u - w\| = \|(u - v) + (v - w)\| \le \|u - v\| + \|v - w\| = \rho(u, v) + \rho(v, w).$$

If $\rho(u, v) = 0$ then $\|u - v\| = 0$ so $u - v = 0$ and $u = v$. **Q**

We therefore have a corresponding topology, with open and closed sets, closures, convergent sequences and so on.

(c) If U is a normed space, a set $A \subseteq U$ is **bounded** (for the norm) if $\{\|u\| : u \in A\}$ is bounded in $\mathbb{R}$; that is, there is some $M \ge 0$ such that $\|u\| \le M$ for every $u \in A$.

2A4C Linear subspaces (a) If U is any normed space and V is a linear subspace of U, then V is also a normed space, if we take the norm of V to be just the restriction to V of the norm of U; the verification is trivial.

(b) If V is a linear subspace of U, so is its closure $\overline{V}$. **P** Take u, $u' \in \overline{V}$ and $\alpha \in {}^{\mathbb{R}}_{\mathbb{C}}$. If $\epsilon > 0$, set $\delta = \epsilon/(2 + |\alpha|) > 0$; then there are v, $v' \in V$ such that $\|u - v\| \le \delta$, $\|u' - v'\| \le \delta$. Now $v + v'$, $\alpha v \in V$ and

$$\|(u + u') - (v + v')\| \le \|u - v\| + \|u' - v'\| \le \epsilon, \quad \|\alpha u - \alpha v\| \le |\alpha|\|u - v\| \le \epsilon.$$

As ϵ is arbitrary, $u + u'$ and αu belong to $\overline{V}$; as u, u' and α are arbitrary, and 0 surely belongs to $V \subseteq \overline{V}$, $\overline{V}$ is a linear subspace of U. **Q**

2A4D Banach spaces (a) If U is a normed space, a sequence $\langle u_n \rangle_{n \in \mathbb{N}}$ in U is **Cauchy** if $\|u_m - u_n\| \to 0$ as m, $n \to \infty$, that is, for every $\epsilon > 0$ there is an $n_0 \in \mathbb{N}$ such that $\|u_m - u_n\| \le \epsilon$ for all m, $n \ge n_0$.

(b) A normed space U is **complete** if every Cauchy sequence has a limit; a complete normed space is called a **Banach space**.

2A4E It is helpful to know the following result.

Lemma Let U be a normed space such that $\langle u_n \rangle_{n \in \mathbb{N}}$ is convergent (that is, has a limit) in U whenever $\langle u_n \rangle_{n \in \mathbb{N}}$ is a sequence in U such that $\|u_{n+1} - u_n\| \le 4^{-n}$ for every $n \in \mathbb{N}$. Then U is complete.

proof Let $\langle u_n\rangle_{n\in\mathbb{N}}$ be any Cauchy sequence in U. For each $k\in\mathbb{N}$, let $n_k\in\mathbb{N}$ be such that $\|u_m-u_n\|\le 4^{-k}$ whenever $m,\,n\ge n_k$. Set $v_k=u_{n_k}$ for each k. Then $\|v_{k+1}-v_k\|\le 4^{-k}$ (whether $n_k\le n_{k+1}$ or $n_{k+1}\le n_k$). So $\langle v_k\rangle_{k\in\mathbb{N}}$ has a limit $v\in U$. I seek to show that v is the required limit of $\langle u_n\rangle_{n\in\mathbb{N}}$. Given $\epsilon>0$, let $l\in\mathbb{N}$ be such that $\|v_k-v\|\le\epsilon$ for every $k\ge l$; let $k\ge l$ be such that $4^{-k}\le\epsilon$; then if $n\ge n_k$,

$$\|u_n-v\|=\|(u_n-v_k)+(v_k-v)\|\le\|u_n-v_k\|+\|v_k-v\|\le\|u_n-u_{n_k}\|+\epsilon\le 2\epsilon.$$

As ϵ is arbitrary, v is a limit of $\langle u_n\rangle_{n\in\mathbb{N}}$. As $\langle u_n\rangle_{n\in\mathbb{N}}$ is arbitrary, U is complete.

2A4F Bounded linear operators (a) Let U, V be two normed spaces. A linear operator $T:U\to V$ is **bounded** if $\{\|Tu\|:u\in U,\,\|u\|\le 1\}$ is bounded. (**Warning!** in this context, we do not ask for the whole set of values $T[U]$ to be bounded; a 'bounded linear operator' need not be what we ordinarily call a 'bounded function'.) Write $\mathsf{B}(U;V)$ for the space of all bounded linear operators from U to V, and for $T\in\mathsf{B}(U;V)$ write $\|T\|=\sup\{\|Tu\|:u\in U,\,\|u\|\le 1\}$.

(b) A useful fact: $\|Tu\|\le\|T\|\|u\|$ for every $T\in\mathsf{B}(U;V)$, $u\in U$. **P** If $|\alpha|>\|u\|$ then

$$\|\frac{1}{\alpha}u\|=\frac{1}{|\alpha|}\|u\|\le 1,$$

so

$$\|Tu\|=\|\alpha T(\frac{1}{\alpha}u)\|=|\alpha|\|T(\frac{1}{\alpha}u)\|\le|\alpha|\|T\|;$$

as α is arbitrary, $\|Tu\|\le\|T\|\|u\|$. **Q**

(c) A linear operator $T:U\to V$ is bounded iff it is continuous for the norm topologies on U and V. **P** (i) If T is bounded, $u_0\in U$ and $\epsilon>0$, then

$$\|Tu-Tu_0\|=\|T(u-u_0)\|\le\|T\|\|u-u_0\|\le\epsilon$$

whenever $\|u-u_0\|\le\frac{\epsilon}{1+\|T\|}$; by 2A3H, T is continuous. (ii) If T is continuous, then there is some $\delta>0$ such that $\|Tu\|=\|Tu-T0\|\le 1$ whenever $\|u\|=\|u-0\|\le\delta$. If now $\|u\|\le 1$,

$$\|Tu\|=\frac{1}{\delta}\|T(\delta u)\|\le\frac{1}{\delta},$$

so T is a bounded operator. **Q**

2A4G Theorem $\mathsf{B}(U;V)$ is a linear space over $\genfrac{}{}{0pt}{}{\mathbb{R}}{\mathbb{C}}$, and $\|\;\|$ is a norm on $\mathsf{B}(U;V)$.

proof I am rather supposing that you are aware, but in any case you will find it easy to check, that if $S:U\to V$ and $T:U\to V$ are linear operators, and $\alpha\in\genfrac{}{}{0pt}{}{\mathbb{R}}{\mathbb{C}}$, then we have linear operators $S+T$ and αT from U to V defined by the formulae

$$(S+T)(u)=Su+Tu,\quad(\alpha T)(u)=\alpha(Tu)$$

for every $u\in U$; moreover, that under these definitions of addition and scalar multiplication the space of all linear operators from U to V is a linear space. Now we see that whenever $S,\,T\in\mathsf{B}(U;V)$, $\alpha\in\genfrac{}{}{0pt}{}{\mathbb{R}}{\mathbb{C}}$, $u\in U$ and $\|u\|\le 1$,

$$\|(S+T)(u)\|=\|Su+Tu\|\le\|Su\|+\|Tu\|\le\|S\|+\|T\|,$$

$$\|(\alpha T)u\|=\|\alpha(Tu)\|=|\alpha|\|Tu\|\le|\alpha|\|T\|;$$

so that $S+T$ and αT belong to $\mathsf{B}(U;V)$, with $\|S+T\|\le\|S\|+\|T\|$ and $\|\alpha T\|\le|\alpha|\|T\|$. This shows that $\mathsf{B}(U;V)$ is a linear subspace of the space of all linear operators and is therefore a linear space over $\genfrac{}{}{0pt}{}{\mathbb{R}}{\mathbb{C}}$ in its own right. To check that the given formula for $\|T\|$ defines a norm, most of the work has just been done; I suppose I should remark, for the sake of form, that $\|T\|\in[0,\infty[$ for every T; if $\alpha=0$, then of course $\|\alpha T\|=0=|\alpha|\|T\|$; for other α,

$$|\alpha|\|T\|=|\alpha|\|\alpha^{-1}\alpha T\|\le|\alpha||\alpha^{-1}|\|\alpha T\|=\|\alpha T\|,$$

so $\|\alpha T\|=|\alpha|\|T\|$. Finally, if $\|T\|=0$ then $\|Tu\|\le\|T\|\|u\|=0$ for every $u\in U$, so $Tu=0$ for every u and T is the zero operator (in the space of all linear operators, and therefore in its subspace $\mathsf{B}(U;V)$).

2A4H Dual spaces The most important case of $\mathsf{B}(U;V)$ is when V is the scalar field $\genfrac{}{}{0pt}{}{\mathbb{R}}{\mathbb{C}}$ itself (of course we can think of $\genfrac{}{}{0pt}{}{\mathbb{R}}{\mathbb{C}}$ as a normed space over itself, writing $\|\alpha\|=|\alpha|$ for each scalar α). In this case we call $\mathsf{B}(U;\genfrac{}{}{0pt}{}{\mathbb{R}}{\mathbb{C}})$ the **dual** of U; it is commonly denoted U' or U^*; I use the latter.

2A4I Extensions of bounded operators: Theorem Let U be a normed space and $V \subseteq U$ a dense linear subspace. Let W be a Banach space and $T_0 : V \to W$ a bounded linear operator; then there is a unique bounded linear operator $T : U \to W$ extending T_0, and $\|T\| = \|T_0\|$.

proof (a) For any $u \in U$, there is a sequence $\langle v_n \rangle_{n\in\mathbb{N}}$ in V converging to u. Now

$$\|T_0 v_m - T_0 v_n\| = \|T_0(v_m - v_n)\| \le \|T_0\|\|v_m - v_n\| \le \|T_0\|(\|v_m - u\| + \|u - v_n\|) \to 0$$

as m, $n \to \infty$, so $\langle T_0 v_n \rangle_{n\in\mathbb{N}}$ is Cauchy and $w = \lim_{n\to\infty} T_0 v_n$ is defined in W. If $\langle v'_n \rangle_{n\in\mathbb{N}}$ is another sequence in V converging to u, then

$$\begin{aligned}\|w - T_0 v'_n\| &\le \|w - T_0 v_n\| + \|T_0(v_n - v'_n)\| \\ &\le \|w - T_0 v_n\| + \|T_0\|(\|v_n - u\| + \|u - v'_n\|) \to 0\end{aligned}$$

as $n \to \infty$, so w is also the limit of $\langle T_0 v'_n \rangle_{n\in\mathbb{N}}$.

(b) We may therefore define $T : U \to W$ by setting $Tu = \lim_{n\to\infty} T_0 v_n$ whenever $\langle v_n \rangle_{n\in\mathbb{N}}$ is a sequence in V converging to u. If $v \in V$, then we can set $v_n = v$ for every n to see that $Tv = T_0 v$; thus T extends T_0. If u, $u' \in U$ and $\alpha \in {}^{\mathbb{R}}_{\mathbb{C}}$, take sequences $\langle v_n \rangle_{n\in\mathbb{N}}$, $\langle v'_n \rangle_{n\in\mathbb{N}}$ in V converging to u, u' respectively; in this case

$$\|(u + u') - (v_n + v'_n)\| \le \|u - v_n\| + \|u' - v'_n\| \to 0, \quad \|\alpha u - \alpha u_n\| = |\alpha|\|u - u_n\| \to 0$$

as $n \to \infty$, so that $T(u + u') = \lim_{n\to\infty} T_0(v_n + v'_n)$, $T(\alpha u) = \lim_{n\to\infty} T_0(\alpha v_n)$, and

$$\begin{aligned}\|T(u + u') - Tu - Tu'\| &\le \|T(u + u') - T_0(v_n + v'_n)\| + \|T_0 v_n - Tu\| + \|T_0 v'_n - Tu'\| \\ &\to 0,\end{aligned}$$

$$\|T(\alpha u) - \alpha Tu\| \le \|T(\alpha u) - T_0(\alpha v_n)\| + |\alpha|\|T_0 v_n - Tu\| \to 0$$

as $n \to \infty$. This means that $\|T(u+u') - Tu - Tu'\| = 0$, $\|T(\alpha u) - \alpha Tu\| = 0$ so $T(u + u') = Tu + Tu'$, $T(\alpha u) = \alpha Tu$; as u, u' and α are arbitrary, T is linear.

(c) For any $u \in U$, let $\langle v_n \rangle_{n\in\mathbb{N}}$ be a sequence in V converging to u. Then

$$\begin{aligned}\|Tu\| &\le \|T_0 v_n\| + \|Tu - T_0 v_n\| \le \|T_0\|\|v_n\| + \|Tu - T_0 v_n\| \\ &\le \|T_0\|(\|u\| + \|v_n - u\|) + \|Tu - T_0 v_n\| \to \|T_0\|\|u\|\end{aligned}$$

as $n \to \infty$, so $\|Tu\| \le \|T_0\|\|u\|$. As u is arbitrary, T is bounded and $\|T\| \le \|T_0\|$. Of course $\|T\| \ge \|T_0\|$ just because T extends T_0.

(d) Finally, let $\tilde{T}$ be any other bounded linear operator from U to W extending T. If $u \in U$, there is a sequence $\langle v_n \rangle_{n\in\mathbb{N}}$ in V converging to u; now

$$\|\tilde{T}u - Tu\| \le \|\tilde{T}(u - v_n)\| + \|T(v_n - u)\| \le (\|\tilde{T}\| + \|T\|)\|u - v_n\| \to 0$$

as $n \to \infty$, so $\|\tilde{T}u - Tu\| = 0$ and $\tilde{T}u = Tu$. As u is arbitrary, $\tilde{T} = T$. Thus T is unique.

2A4J Normed algebras (a) A **normed algebra** is a normed space $(U, \| \|)$ together with a multiplication, a binary operator $\times$ on U, such that

$$u \times (v \times w) = (u \times v) \times w,$$

$$u \times (v + w) = (u \times v) + (u \times w), \quad (u + v) \times w = (u \times w) + (v \times w),$$

$$(\alpha u) \times v = u \times (\alpha v) = \alpha(u \times v),$$

$$\|u \times v\| \le \|u\|\|v\|$$

for all u, v, $w \in U$ and $\alpha \in {}^{\mathbb{R}}_{\mathbb{C}}$.

(b) A **Banach algebra** is a normed algebra which is a Banach space. A normed algebra U is **commutative** if its multiplication is commutative, that is, $u \times v = v \times u$ for all u, $v \in U$.

***2A4K Definition** A normed space U is **uniformly convex** if for every $\epsilon > 0$ there is a $\delta > 0$ such that $\|u + v\| \le 2 - \delta$ whenever u, $v \in U$, $\|u\| = \|v\| = 1$ and $\|u - v\| \ge \epsilon$.

2A5 Linear topological spaces

The principal objective of §2A3 is in fact the study of certain topologies on the linear spaces of Chapter 24. I give some fragments of the general theory.

2A5A Linear space topologies Something which is not covered in detail by every introduction to functional analysis is the general concept of 'linear topological space'. The ideas needed for the work of §245 are reasonably briefly expressed.

Definition A **linear topological space** or **topological vector space** over $\substack{\mathbb{R}\\\mathbb{C}}$ is a linear space U over $\substack{\mathbb{R}\\\mathbb{C}}$ together with a topology $\mathfrak{T}$ such that the maps

$$(u,v)\mapsto u+v:U\times U\to U,$$

$$(\alpha,u)\mapsto \alpha u:\substack{\mathbb{R}\\\mathbb{C}}\times U\to U$$

are both continuous, where the product spaces $U\times U$ and $\substack{\mathbb{R}\\\mathbb{C}}\times U$ are given their product topologies (2A3T). Given a linear space U, a topology on U satisfying the conditions above is a **linear space topology**. Note that

$$(u,v)\mapsto u-v=u+(-1)v:U\times U\to U$$

will also be continuous.

2A5B All the linear topological spaces we need turn out to be readily presentable in the following terms.

Proposition Suppose that U is a linear space over $\substack{\mathbb{R}\\\mathbb{C}}$, and T is a family of functionals $\tau:U\to[0,\infty[$ such that
(i) $\tau(u+v)\le\tau(u)+\tau(v)$ for all u, $v\in U$, $\tau\in\mathrm{T}$;
(ii) $\tau(\alpha u)\le\tau(u)$ if $u\in U$, $|\alpha|\le 1$, $\tau\in\mathrm{T}$;
(iii) $\lim_{\alpha\to 0}\tau(\alpha u)=0$ for every $u\in U$, $\tau\in\mathrm{T}$.
For $\tau\in\mathrm{T}$, define $\rho_\tau:U\times U\to[0,\infty[$ by setting $\rho_\tau(u,v)=\tau(u-v)$ for all u, $v\in U$. Then each ρ_τ is a pseudometric on U, and the topology defined by $\mathrm{P}=\{\rho_\tau:\tau\in\mathrm{T}\}$ renders U a linear topological space.

proof (a) It is worth noting immediately that

$$\tau(0)=\lim_{\alpha\to 0}\tau(\alpha 0)=0$$

for every $\tau\in\mathrm{T}$.

(b) To see that every ρ_τ is a pseudometric, argue as follows.

(i) ρ_τ takes values in $[0,\infty[$ because τ does.

(ii) If u, v, $w\in U$ then

$$\begin{aligned}\rho_\tau(u,w)&=\tau(u-w)=\tau((u-v)+(v-w))\\&\le\tau(u-v)+\tau(v-w)=\rho_\tau(u,v)+\rho_\tau(v,w).\end{aligned}$$

(iii) If u, $v\in U$, then

$$\rho(v,u)=\tau(v-u)=\tau(-1(u-v))\le\tau(u,v)=\rho_\tau(u,v),$$

and similarly $\rho_\tau(u,v)\le\rho_\tau(v,u)$, so the two are equal.

(iv) If $u\in U$ then $\rho_\tau(u,u)=\tau(0)=0$.

(c) Let $\mathfrak{T}$ be the topology on U defined by $\{\rho_\tau:\tau\in\mathrm{T}\}$ (2A3F).

(i) Addition is continuous because, given $\tau\in\mathrm{T}$, we have

$$\begin{aligned}\rho_\tau(u'+v',u+v)&=\tau((u'+v')-(u+v))\\&\le\tau(u'-u)+\tau(v'-v)\le\rho_\tau(u',u)+\rho_\tau(v',v)\end{aligned}$$

for all u, v, u', $v'\in U$. This means that, given $\epsilon>0$ and $(u,v)\in U\times U$, we shall have

$$\rho_\tau(u'+v',u+v)\le\epsilon \text{ whenever } (u',v')\in U((u,v);\tilde\rho_\tau,\bar\rho_\tau;\tfrac{\epsilon}{2}),$$

using the language of 2A3Tb. Because $\tilde{\rho}_\tau$, $\bar{\rho}_\tau$ are two of the pseudometrics defining the product topology of $U \times U$ (2A3Tb), $(u, v) \mapsto u + v$ is continuous, by the criterion of 2A3H.

(ii) Scalar multiplication is continuous because if $u \in U$ and $n \in \mathbb{N}$ then $\tau(nu) \le n\tau(u)$ for every $\tau \in \mathrm{T}$ (induce on n). Consequently, if $\tau \in \mathrm{T}$,

$$\tau(\alpha u) \le n\tau(\frac{\alpha}{n}u) \le n\tau(u)$$

whenever $|\alpha| < n \in \mathbb{N}$ and $\tau \in \mathrm{T}$. Now, given $(\alpha, u) \in {}^{\mathbb{R}}_{\mathbb{C}} \times U$ and $\epsilon > 0$, take $n > |\alpha|$ and $\delta > 0$ such that $\delta \le \min(n - |\alpha|, \frac{\epsilon}{2n})$ and $\tau(\gamma u) \le \frac{\epsilon}{2}$ whenever $|\gamma| \le \delta$; then

$$\begin{aligned}\rho_\tau(\alpha' u', \alpha u) = \tau(\alpha' u' - \alpha u) &\le \tau(\alpha'(u' - u)) + \tau((\alpha' - \alpha)u) \\ &\le n\tau(u' - u) + \tau((\alpha' - \alpha)u)\end{aligned}$$

whenever $u' \in U$ and $\alpha' \in {}^{\mathbb{R}}_{\mathbb{C}}$ and $|\alpha'| < n \in \mathbb{N}$. Accordingly, setting $\theta(\alpha', \alpha) = |\alpha' - \alpha|$ for α', $\alpha \in {}^{\mathbb{R}}_{\mathbb{C}}$,

$$\rho_\tau(\alpha' u', \alpha u) \le n\delta + \frac{\epsilon}{2} \le \epsilon$$

whenever

$$(\alpha', u') \in U((\alpha, u); \tilde{\theta}, \bar{\rho}_\tau; \delta).$$

Because $\tilde{\theta}$ and $\bar{\rho}_\tau$ are among the pseudometrics defining the topology of ${}^{\mathbb{R}}_{\mathbb{C}} \times U$, the map $(\alpha, u) \mapsto \alpha u$ satisfies the criterion of 2A3H and is continuous.

Thus $\mathfrak{T}$ is a linear space topology on U.

***2A5C** We do not need it for Chapter 24, but the following is worth knowing.

Theorem Let U be a linear space and $\mathfrak{T}$ a linear space topology on U.

(a) There is a family T of functionals satisfying the conditions (i)-(iii) of 2A5B and defining $\mathfrak{T}$.

(b) If $\mathfrak{T}$ is metrizable, we can take T to consist of a single functional.

proof (a) Kelley & Namioka 76, p. 50.

(b) Köthe 69, §15.11.

2A5D Definition Let U be a linear space over ${}^{\mathbb{R}}_{\mathbb{C}}$. Then a **seminorm** on U is a functional $\tau : U \to [0, \infty[$ such that

(i) $\tau(u + v) \le \tau(u) + \tau(v)$ for all u, $v \in U$;

(ii) $\tau(\alpha u) = |\alpha|\tau(u)$ if $u \in U$, $\alpha \in {}^{\mathbb{R}}_{\mathbb{C}}$.

Observe that a norm is always a seminorm, and that a seminorm is always a functional of the type described in 2A5B. In particular, the association of a metric with a norm (2A4Bb) is a special case of 2A5B.

2A5E Convex sets (a) Let U be a linear space over ${}^{\mathbb{R}}_{\mathbb{C}}$. A subset C of U is **convex** if $\alpha u + (1 - \alpha)v \in C$ whenever u, $v \in C$ and $\alpha \in [0, 1]$. The intersection of any family of convex sets is convex, so for every set $A \subseteq U$ there is a smallest convex set including A; this is just the set of vectors expressible as $\sum_{i=0}^n \alpha_i u_i$ where $u_0, \ldots, u_n \in A$, $\alpha_0, \ldots, \alpha_n \in [0, 1]$ and $\sum_{i=0}^n \alpha_i = 1$ (Bourbaki 87, II.2.3); it is the **convex hull** of A.

(b) If U is a linear topological space, the closure of any convex set is convex (Bourbaki 87, II.2.6). It follows that, for any $A \subseteq U$, the closure of the convex hull of A is the smallest closed convex set including A; this is the **closed convex hull** of A.

(c) I note for future reference that in a linear topological space, the closure of any linear subspace is a linear subspace. (Bourbaki 87, I.1.3; Köthe 69, §15.2. Compare 2A4Cb.)

2A5F Completeness in linear topological spaces In normed spaces, completeness can be described in terms of Cauchy sequences (2A4D). In general linear topological spaces this is inadequate. The true theory of 'completeness' demands the concept of 'uniform space' (see §3A4 in the next volume, or Kelley 55, chap. 6; Engelking 89, §8.1: Bourbaki 66, chap. II); I shall not describe this here, but will give a version adapted to linear spaces. I mention this only because you will I hope some day come to the general theory (in Volume 3 of this treatise, if not before), and you should be aware that the special case described here gives a misleading emphasis at some points.

Definitions Let U be a linear space over ${}^{\mathbb{R}}_{\mathbb{C}}$, and $\mathfrak{T}$ a linear space topology on U. A filter $\mathcal{F}$ on U is **Cauchy** if for every open set G in U containing 0 there is an $F \in \mathcal{F}$ such that $F - F = \{u - v : u, v \in F\}$ is included in G. U is **complete** if every Cauchy filter on U is convergent.

2A5G Cauchy filters have a simple description when a linear space topology is defined by the method of 2A5B.

Lemma Let U be a linear space over ${}^{\mathbb{R}}_{\mathbb{C}}$, and let T be a family of functionals defining a linear space topology on U, as in 2A5B. Then a filter $\mathcal{F}$ on U is Cauchy iff for every $\tau \in \mathrm{T}$ and $\epsilon > 0$ there is an $F \in \mathcal{F}$ such that $\tau(u - v) \le \epsilon$ for all u, $v \in F$.

proof (a) Suppose that $\mathcal{F}$ is Cauchy, $\tau \in \mathrm{T}$ and $\epsilon > 0$. Then $G = U(0; \rho_\tau; \epsilon)$ is open (using the language of 2A3F-2A3G), so there is an $F \in \mathcal{F}$ such that $F - F \subseteq G$; but this just means that $\tau(u - v) < \epsilon$ for all u, $v \in F$.

(b) Suppose that $\mathcal{F}$ satisfies the criterion, and that G is an open set containing 0. Then there are $\tau_0, \dots, \tau_n \in \mathrm{T}$ and $\epsilon > 0$ such that $U(0; \rho_{\tau_0}, \dots, \rho_{\tau_n}; \epsilon) \subseteq G$. For each $i \le n$ there is an $F_i \in \mathcal{F}$ such that $\tau_i(u, v) < \frac{\epsilon}{2}$ for all u, $v \in F_i$; now $F = \bigcap_{i \le n} F_i \in \mathcal{F}$ and $u - v \in G$ for all u, $v \in F$.

2A5H Normed spaces and sequential completeness I had better point out that for normed spaces the definition of 2A5F agrees with that of 2A4D.

Proposition Let $(U, \| \, \|)$ be a normed space over ${}^{\mathbb{R}}_{\mathbb{C}}$, and let $\mathfrak{T}$ be the linear space topology on U defined by the method of 2A5B from the set $\mathrm{T} = \{\| \, \|\}$. Then U is complete in the sense of 2A5F iff it is complete in the sense of 2A4D.

proof (a) Suppose first that U is complete in the sense of 2A5F. Let $\langle u_n \rangle_{n \in \mathbb{N}}$ be a sequence in U which is Cauchy in the sense of 2A4Da. Set

$$\mathcal{F} = \{F : F \subseteq U, \ \{n : u_n \notin F\} \text{ is finite}\}.$$

Then it is easy to check that $\mathcal{F}$ is a filter on U, the image of the Fréchet filter under the map $n \mapsto u_n : \mathbb{N} \to U$. If $\epsilon > 0$, take $m \in \mathbb{N}$ such that $\|u_j - u_k\| \le \epsilon$ whenever j, $k \ge m$; then $F = \{u_j : j \ge m\}$ belongs to $\mathcal{F}$, and $\|u - v\| \le \epsilon$ for all u, $v \in F$. So $\mathcal{F}$ is Cauchy in the sense of 2A5F, and has a limit u say. Now, for any $\epsilon > 0$, the set $\{v : \|v - u\| < \epsilon\} = U(u; \rho_{\| \, \|}; \epsilon)$ is an open set containing u, so belongs to $\mathcal{F}$, and $\{n : \|u_n - u\| \ge \epsilon\}$ is finite, that is, there is an $m \in \mathbb{N}$ such that $\|u_m - u\| < \epsilon$ whenever $n \ge m$. As ϵ is arbitrary, $u = \lim_{n \to \infty} u_n$ in the sense of 2A3M. As $\langle u_n \rangle_{n \in \mathbb{N}}$ is arbitrary, U is complete in the sense of 2A4D.

(b) Now suppose that U is complete in the sense of 2A4D. Let $\mathcal{F}$ be a Cauchy filter on U. For each $n \in \mathbb{N}$, choose a set $F_n \in \mathcal{F}$ such that $\|u - v\| \le 2^{-n}$ for all u, $v \in F_n$. For each $n \in \mathbb{N}$, $F'_n = \bigcap_{i \le n} F_i$ belongs to $\mathcal{F}$, so is not empty; choose $u_n \in F'_n$. If $m \in \mathbb{N}$ and j, $k \ge m$, then both u_j and u_k belong to F_m, so $\|u_j - u_k\| \le 2^{-m}$; thus $\langle u_n \rangle_{n \in \mathbb{N}}$ is a Cauchy sequence in the sense of 2A4Da, and has a limit u say. Now take any $\epsilon > 0$ and $m \in \mathbb{N}$ such that $2^{-m+1} \le \epsilon$. There is surely a $k \ge m$ such that $\|u_k - u\| \le 2^{-m}$; now $u_k \in F_m$, so

$$F_m \subseteq \{v : \|v - u_k\| \le 2^{-m}\} \subseteq \{v : \|v - u\| \le 2^{-m+1}\} \subseteq \{v : \rho_{\| \, \|}(v, u) \le \epsilon\},$$

and $\{v : \rho_{\| \, \|}(v, u) \le \epsilon\} \in \mathcal{F}$. As ϵ is arbitrary, $\mathcal{F}$ converges to u, by 2A3Sd. As $\mathcal{F}$ is arbitrary, U is complete.

(c) Thus the two definitions coincide, provided at least that we allow the countably many simultaneous choices of the u_n in part (b) of the proof.

2A5I Weak topologies I come now to brief notes on 'weak topologies' on normed spaces; from the point of view of this volume, these are in fact the primary examples of linear space topologies. Let U be a normed linear space over ${}^{\mathbb{R}}_{\mathbb{C}}$.

(a) Write U^* for its dual $\mathrm{B}(U; {}^{\mathbb{R}}_{\mathbb{C}})$ (2A4H). Let T be the set $\{|h| : h \in U^*\}$; then T satisfies the conditions of 2A5B, so defines a linear space topology on U; this is called the **weak topology** of U.

(b) A filter $\mathcal{F}$ on U converges to $u \in U$ for the weak topology of U iff $\lim_{v \to \mathcal{F}} \rho_{|h|}(v, u) = 0$ for every $h \in U^*$ (2A3Sd), that is, iff $\lim_{v \to \mathcal{F}} |h(v - u)| = 0$ for every $h \in U^*$, that is, iff $\lim_{v \to \mathcal{F}} h(v) = h(u)$ for every $h \in U^*$.

(c) A set $C \subseteq U$ is called **weakly compact** if it is compact for the weak topology of U. So (subject to the axiom of choice) a set $C \subseteq U$ is weakly compact iff for every ultrafilter $\mathcal{F}$ on U containing C there is a $u \in C$ such that $\lim_{v \to \mathcal{F}} h(v) = h(u)$ for every $h \in U^*$ (put 2A3R together with (b) above).

(d) A subset A of U is called **relatively weakly compact** if it is a subset of some weakly compact subset of U.

(e) If $h \in U^*$, then $h : U \to {}^{\mathbb{R}}_{\mathbb{C}}$ is continuous for the weak topology on U and the usual topology of ${}^{\mathbb{R}}_{\mathbb{C}}$; this is obvious if we apply the criterion of 2A3H. So if $A \subseteq U$ is relatively weakly compact, $h[A]$ must be bounded in ${}^{\mathbb{R}}_{\mathbb{C}}$. **P** Let $C \supseteq A$ be a weakly compact set. Then $h[C]$ is compact in ${}^{\mathbb{R}}_{\mathbb{C}}$, by 2A3Nb, so is bounded, by 2A2F (noting that if the underlying field is $\mathbb{C}$, then it can be identified, as metric space, with $\mathbb{R}^2$). Accordingly $h[A]$ also is bounded. **Q**

(f) If V is another normed space and $T : U \to V$ is a bounded linear operator, then T is continuous for the respective weak topologies. **P** If $h \in V^*$ then the composition hT belongs to U^*. Now, for any $u, v \in U$,

$$\rho_{|h|}(Tu, Tv) = |h(Tu - Tv)| = |hT(u - v)| = \rho_{|hT|}(u, v),$$

taking $\rho_{|h|}$, $\rho_{|hT|}$ to be the pseudometrics on V, U respectively defined by the formula of 2A5B. By 2A3H, T is continuous. **Q**

(g) Corresponding to the weak topology on a normed space U, we have the **weak*** or **w*-**topology on its dual U^*, defined by the set $\mathrm{T} = \{|\hat{u}| : u \in U\}$, where I write $\hat{u}(f) = f(u)$ for every $f \in U^*$, $u \in U$. As in (a), this is a linear space topology on U^*. (It is essential to distinguish between the 'weak*' topology and the 'weak' topology on U^*. The former depends only on the action of U on U^*, the latter on the action of $U^{**} = (U^*)^*$. You will have no difficulty in checking that $\hat{u} \in U^{**}$ for every $u \in U$, but the point is that there may be members of U^{**} not representable in this way, leading to open sets for the weak topology which are not open for the weak* topology.)

***2A5J Angelic spaces** I do not rely on the following ideas, but they may throw light on some results in §§246-247. First, a topological space X is **regular** if whenever $G \subseteq X$ is open and $x \in G$ then there is an open set H such that $x \in H \subseteq \overline{H} \subseteq G$. Next, a regular Hausdorff space X is **angelic** if whenever $A \subseteq X$ is such that every sequence in A has a cluster point in X, then $\overline{A}$ is compact and every point of $\overline{A}$ is the limit of a sequence in A. What this means is that compactness in X, and the topologies of compact subsets of X, can be effectively described in terms of sequences. Now the theorem (due to Eberlein and Šmulian) is that any normed space is angelic in its weak topology. (462D in Volume 4; Köthe 69, §24; Dunford & Schwartz 57, V.6.1.) In particular, this is true of L^1 spaces, which makes it less surprising that there should be criteria for weak compactness in L^1 spaces which deal only with sequences.

2A6 Factorization of matrices

I spend a couple of pages on the linear algebra of $\mathbb{R}^r$ required for Chapter 26. I give only one proof, because this is material which can be found in any textbook of elementary linear algebra; but I think it may be helpful to run through the basic ideas in the language which I use for this treatise.

2A6A Determinants We need to know the following things about determinants.

(i) Every $r \times r$ real matrix T has a real determinant $\det T$.

(ii) For any $r \times r$ matrices S and T, $\det ST = \det S \det T$.

(iii) If T is a diagonal matrix, its determinant is just the product of its diagonal entries.

(iv) For any $r \times r$ matrix T, $\det T' = \det T$, where T' is the transpose of T.

(v) $\det T$ is a continuous function of the coefficients of T.

There are so many routes through this topic that I avoid even a definition of 'determinant'; I invite you to check your memory, or your favourite text, to confirm that you are indeed happy with the facts above.

2A6B Orthonormal families For $x = (\xi_1, \dots, \xi_r)$, $y = (\eta_1, \dots, \eta_r) \in \mathbb{R}^r$, write $x\,\boldsymbol{.}\,y = \sum_{i=1}^r \xi_i \eta_i$; of course $\|x\|$, as defined in 1A2A, is $\sqrt{x\,\boldsymbol{.}\,x}$. Recall that $x_1, \dots, x_k$ are **orthonormal** if $x_i\,\boldsymbol{.}\,x_j = 0$ for $i \neq j$, 1 for $i = j$. The results we need here are:

(i) If $x_1, \dots, x_k$ are orthonormal vectors in $\mathbb{R}^r$, where $k < r$, then there are vectors $x_{k+1}, \dots, x_r$ in $\mathbb{R}^r$ such that $x_1, \dots, x_r$ are orthonormal.

(ii) An $r \times r$ matrix P is **orthogonal** if $P'P$ is the identity matrix; equivalently, if the columns of P are orthonormal.

(iii) For an orthogonal matrix P, $\det P$ must be ± 1 (put (ii)-(iv) of 2A6A together).

(iv) If P is orthogonal, then $Px\,\boldsymbol{.}\,Py = P'Px\,\boldsymbol{.}\,y = x\,\boldsymbol{.}\,y$ for all $x, y \in \mathbb{R}^r$.

(v) If P is orthogonal, so is $P' = P^{-1}$.

(vi) If P and Q are orthogonal, so is PQ.

2A6C I now give a proposition which is not always included in elementary presentations. Of course there are many approaches to this; I offer a direct one.

Proposition Let T be any real $r \times r$ matrix. Then T is expressible as PDQ where P and Q are orthogonal matrices and D is a diagonal matrix with non-negative coefficients.

proof I induce on r.

(a) If $r = 1$, then $T = (\tau_{11})$. Set $D = (|\tau_{11}|)$, $P = (1)$ and $Q = (1)$ if $\tau_{11} \ge 0$, (-1) otherwise.

(b)(i) For the inductive step to $r + 1 \ge 2$, consider the unit ball $B = \{x : x \in \mathbb{R}^{r+1}, \|x\| \le 1\}$. This is a closed bounded set in $\mathbb{R}^{r+1}$, so is compact (2A2F). The maps $x \mapsto Tx : \mathbb{R}^{r+1} \to \mathbb{R}^{r+1}$ and $x \mapsto \|x\| : \mathbb{R}^{r+1} \to \mathbb{R}$ are continuous, so the function $x \mapsto \|Tx\| : B \to \mathbb{R}$ is bounded and attains its bounds (2A2G), and there is a $u \in B$ such that $\|Tu\| \ge \|Tx\|$ for every $x \in B$. Observe that $\|Tu\|$ must be the norm $\|T\|$ of T as defined in 262H. Set $\delta = \|T\| = \|Tu\|$. If $\delta = 0$, then T must be the zero matrix, and the result is trivial; so let us suppose that $\delta > 0$. In this case $\|u\|$ must be exactly 1, since otherwise we should have $u = \|u\|u'$ where $\|u'\| = 1$ and $\|Tu'\| > \|Tu\|$.

(ii) If $x \in \mathbb{R}^{r+1}$ and $x \boldsymbol{.} u = 0$, then $Tx \boldsymbol{.} Tu = 0$. **P?** If not, set $\gamma = Tx \boldsymbol{.} Tu \ne 0$. Consider $y = u + \eta\gamma x$ for small $\eta > 0$. We have

$$\|y\|^2 = y \boldsymbol{.} y = u \boldsymbol{.} u + 2\eta\gamma u \boldsymbol{.} x + \eta^2\gamma^2 x \boldsymbol{.} x = \|u\|^2 + \eta^2\gamma^2\|x\|^2 = 1 + \eta^2\gamma^2\|x\|^2,$$

while

$$\|Ty\|^2 = Ty \boldsymbol{.} Ty = Tu \boldsymbol{.} Tu + 2\eta\gamma Tu \boldsymbol{.} Tx + \eta^2\gamma^2 Tx \boldsymbol{.} Tx = \delta^2 + 2\eta\gamma^2 + \eta^2\gamma^2\|Tx\|^2.$$

But also $\|Ty\|^2 \le \delta^2\|y\|^2$ (2A4Fb), so

$$\delta^2 + 2\eta\gamma^2 + \eta^2\gamma^2\|Tx\|^2 \le \delta^2(1 + \eta^2\gamma^2\|x\|^2)$$

and

$$2\eta\gamma^2 \le \delta^2\eta^2\gamma^2\|x\|^2 - \eta^2\gamma^2\|Tx\|^2,$$

that is,

$$2 \le \eta(\delta^2\|x\|^2 - \|Tx\|^2).$$

But this surely cannot be true for all $\eta > 0$, so we have a contradiction. **XQ**

(iii) Set $v = \delta^{-1}Tu$, so that $\|v\| = 1$. Let $u_1, \dots, u_{r+1}$ be orthonormal vectors such that $u_{r+1} = u$, and let Q_0 be the orthogonal $(r+1) \times (r+1)$ matrix with columns $u_1, \dots, u_{r+1}$; then, writing $e_1, \dots, e_{r+1}$ for the standard orthonormal basis of $\mathbb{R}^{r+1}$, we have $Q_0 e_i = u_i$ for each i, and $Q_0 e_{r+1} = u$. Similarly, there is an orthogonal matrix P_0 such that $P_0 e_{r+1} = v$.

Set $T_1 = P_0^{-1}TQ_0$. Then

$$T_1 e_{r+1} = P_0^{-1}Tu = \delta P_0^{-1}v = \delta e_{r+1},$$

while if $x \boldsymbol{.} e_{r+1} = 0$ then $Q_0 x \boldsymbol{.} u = 0$ (2A6B(iv)), so that

$$T_1 x \boldsymbol{.} e_{r+1} = P_0 T_1 x \boldsymbol{.} P_0 e_{r+1} = TQ_0 x \boldsymbol{.} v = 0,$$

by (ii). This means that T_1 must be of the form

$$\begin{pmatrix} S & 0 \\ 0 & \delta \end{pmatrix},$$

where S is an $r \times r$ matrix.

(iv) By the inductive hypothesis, S is expressible as $\tilde{P}\tilde{D}\tilde{Q}$, where $\tilde{P}$ and $\tilde{Q}$ are orthogonal $r \times r$ matrices and $\tilde{D}$ is a diagonal $r \times r$ matrix with non-negative coefficients. Set

$$P_1 = \begin{pmatrix} \tilde{P} & 0 \\ 0 & 1 \end{pmatrix}, \quad Q_1 = \begin{pmatrix} \tilde{Q} & 0 \\ 0 & 1 \end{pmatrix}, \quad D = \begin{pmatrix} \tilde{D} & 0 \\ 0 & \delta \end{pmatrix}.$$

Then P_1 and Q_1 are orthogonal and D is diagonal, with non-negative coefficients, and $P_1 D Q_1 = T_1$. Now set

$$P = P_0 P_1, \quad Q = Q_1 Q_0^{-1},$$

so that P and Q are orthogonal and

$$PDQ = P_0 P_1 D Q_1 Q_0^{-1} = P_0 T_1 Q_0^{-1} = T.$$

Thus the induction proceeds.

Concordance

I list here the section and paragraph numbers which have (to my knowledge) appeared in print in references to this chapter, and which have since been changed.

211Ya Countable-cocountable algebra of $\mathbb{R}$ This exercise, referred to in the 2002 edition of Volume 3, has been moved to 211Ye.

214J Subspace measures on measurable subspaces, direct sums 214J-214M, referred to in the 2002 and 2004 editions of Volume 3, the 2003 and 2006 editions of Volume 4, and the 2008 edition of Volume 5, have been moved to 214K-214N.

214N Upper and lower integrals This result, referred to in the 2008 edition of Volume 5, has been moved to 214J.

215Yc Measurable envelopes This exercise, referred to in the 2000 edition of Volume 1, has been moved to 216Yc.

§**234** Section §234 has been rewritten, with a good deal of new material. The former paragraphs 234A-234G, referred to in the 2002 and 2004 editions of Volume 3 and the 2003 and 2006 editions of Volume 4, are now 234I-234O.

§**235** Section §235 has been re-organized, with some material moved to §234. Specifically, 235H, 235I, 235J, 235L, 235M, 235T and 235Xe, referred to in the 2002 and 2004 editions of Volume 3 and the 2003 and 2006 editions of Volume 4, are now dealt with in 234B, 235G, 235H, 235J, 235K, 235R and 234A.

241Yd Countable sup property This exercise, referred to in the 2002 edition of Volume 3, has been moved to 241Ye.

241Yh Quotient Riesz spaces This exercise, referred to in the 2002 edition of Volume 3, has been moved to 241Yc.

242Xf Inverse-measure-preserving functions This exercise, referred to in the 2002 edition of Volume 3, has been moved to 242Xd.

242Yc Order-continuous norms This exercise, referred to in the 2002 edition of Volume 3, has been moved to 242Yg.

244O Complex L^p This paragraph, referred to in the 2002 and 2004 editions of Volume 3, and the 2003 and 2006 editions of Volume 4, is now 244P.

244Xf L^p and L^q This exercise, referred to in the 2003 edition of Volume 4, has been moved to 244Xe.

244Yd-244Yf L^p as Banach lattice These exercises, referred to in the 2002 and 2004 editions of Volume 3, are now 244Ye-244Yg.

251N Paragraph numbers in the second half of §251, referred to in editions of Volumes 3 and 4 up to and including 2006, have been changed, so that 251M-251S are now 251N-251T.

252Yf Exercise This exercise, referred to in the first edition of Volume 1, has been moved to 252Ym.

272S Distribution of a sum of independent random variables This result, referred to in the 2002 and 2004 editions of Volume 3, and the 2003 and 2006 editions of Volume 4, is now 272T.

272U Etemadi's lemma This result, referred to in the 2003 and 2006 editions of Volume 4, is now 272V.

272Yd This exercise, referred to in the 2002 and 2004 editions of Volume 3, is now 272Ye.

273Xh This exercise, referred to in the 2006 edition of Volume 4, is now 273Xi.

References for Volume 2

Alexits G. [78] (ed.) *Fourier Analysis and Approximation Theory.* North-Holland, 1978 (Colloq. Math. Soc. Janos Bolyai 19).

Antonov N.Yu. [96] 'Convergence of Fourier series', East J. Approx. 7 (1996) 187-196. [§286 *notes*.]

Arias de Reyna J. [02] *Pointwise Convergence of Fourier Series.* Springer, 2002 (Lecture Notes in Mathematics 1785). [§286 *notes*.]

Bergelson V., March P. & Rosenblatt J. [96] (eds.) *Convergence in Ergodic Theory and Probability.* de Gruyter, 1996.

du Bois-Reymond P. [1876] 'Untersuchungen über die Convergenz und Divergenz der Fouriersche Darstellungformeln', Abh. Akad. München 12 (1876) 1-103. [§282 *notes*.]

Bourbaki N. [66] *General Topology.* Hermann/Addison-Wesley, 1968. [2A5F.]

Bourbaki N. [87] *Topological Vector Spaces.* Springer, 1987. [2A5E.]

Carleson L. [66] 'On convergence and growth of partial sums of Fourier series', Acta Math. 116 (1966) 135-157. [§282 *notes*, §286 *intro.*, §286 *notes*.]

Clarkson J.A. [36] 'Uniformly convex spaces', Trans. Amer. Math. Soc. 40 (1936) 396-414. [244O.]

Defant A. & Floret K. [93] *Tensor Norms and Operator Ideals*, North-Holland, 1993. [§253 *notes*.]

Dudley R.M. [89] *Real Analysis and Probability.* Wadsworth & Brooks/Cole, 1989. [§282 *notes*.]

Dunford N. & Schwartz J.T. [57] *Linear Operators I.* Wiley, 1957 (reprinted 1988). [§244 *notes*, 2A5J.]

Enderton H.B. [77] *Elements of Set Theory.* Academic, 1977. [§2A1.]

Engelking R. [89] *General Topology.* Heldermann, 1989 (Sigma Series in Pure Mathematics 6). [2A5F.]

Etemadi N. [96] 'On convergence of partial sums of independent random variables', pp. 137-144 in Bergelson March & Rosenblatt 96. [272V.]

Evans L.C. & Gariepy R.F. [92] *Measure Theory and Fine Properties of Functions.* CRC Press, 1992. [263Ec, §265 *notes*.]

Federer H. [69] *Geometric Measure Theory.* Springer, 1969 (reprinted 1996). [262C, 263Ec, §264 *notes*, §265 *notes*, §266 *notes*.]

Feller W. [66] *An Introduction to Probability Theory and its Applications,* vol. II. Wiley, 1966. [Chap. 27 *intro.*, 274H, 275Xc, 285N.]

Fremlin D.H. [74] *Topological Riesz Spaces and Measure Theory.* Cambridge U.P., 1974. [§232 *notes*, 241F, §244 *notes*, §245 *notes*, §247 *notes*.]

Fremlin D.H. [93] 'Real-valued-measurable cardinals', pp. 151-304 in Judah 93. [232Hc.]

Haimo D.T. [67] (ed.) *Orthogonal Expansions and their Continuous Analogues.* Southern Illinois University Press, 1967.

Hall P. [82] *Rates of Convergence in the Central Limit Theorem.* Pitman, 1982. [274Hc.]

Halmos P.R. [50] *Measure Theory.* Van Nostrand, 1950. [§251 *notes*, §252 *notes*, 255Yn.]

Halmos P.R. [60] *Naive Set Theory.* Van Nostrand, 1960. [§2A1.]

Hanner O. [56] 'On the uniform convexity of L^p and l^p', Arkiv för Matematik 3 (1956) 239-244. [244O.]

Henle J.M. [86] *An Outline of Set Theory.* Springer, 1986. [§2A1.]

Hoeffding W. [63] 'Probability inequalities for sums of bounded random variables', J. Amer. Statistical Association 58 (1963) 13-30. [272W.]

Hunt R.A. [67] 'On the convergence of Fourier series', pp. 235-255 in Haimo 67. [§286 *notes*.]

Jorsbøe O.G. & Mejlbro L. [82] *The Carleson-Hunt Theorem on Fourier Series.* Springer, 1982 (Lecture Notes in Mathematics 911). [§286 *notes*.]

Judah H. [93] (ed.) Proceedings of the Bar-Ilan Conference on Set Theory and the Reals, 1991. Amer. Math. Soc. (Israel Mathematical Conference Proceedings 6), 1993.

Kelley J.L. [55] *General Topology.* Van Nostrand, 1955. [2A5F.]

Kelley J.L. & Namioka I. [76] *Linear Topological Spaces.* Springer, 1976. [2A5C.]

Kirzbraun M.D. [34] 'Über die zusammenziehenden und Lipschitzian Transformationen', Fund. Math. 22 (1934) 77-108. [262C.]

Kolmogorov A.N. [26] 'Une série de Fourier-Lebesgue divergente partout', C. R. Acad. Sci. Paris 183 (1926) 1327-1328. [§282 *notes*.]

Komlós J. [67] 'A generalization of a problem of Steinhaus', Acta Math. Acad. Sci. Hung. 18 (1967) 217-229. [276H.]

Körner T.W. [88] *Fourier Analysis.* Cambridge U.P., 1988. [§282 *notes*.]

Köthe G. [69] *Topological Vector Spaces I.* Springer, 1969. [2A5C, 2A5J.]

Krivine J.-L. [71] *Introduction to Axiomatic Set Theory.* D. Reidel, 1971. [§2A1.]

Lacey M. & Thiele C. [00] 'A proof of boundedness of the Carleson operator', Math. Research Letters 7 (2000) 1-10. [§286 *intro.*, 286H.]

Lebesgue H. [72] *Oeuvres Scientifiques.* L'Enseignement Mathématique, Institut de Mathématiques, Univ. de Genève, 1972. [Chap. 27 *intro.*]

Liapounoff A. [1901] 'Nouvelle forme du théorème sur la limite de probabilité', Mém. Acad. Imp. Sci. St-Pétersbourg 12(5) (1901) 1-24. [274Xg.]

Lighthill M.J. [59] *Introduction to Fourier Analysis and Generalised Functions.* Cambridge U.P., 1959. [§284 *notes*.]

Lindeberg J.W. [22] 'Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung', Math. Zeitschrift 15 (1922) 211-225. [274Ha, §274 *notes*.]

Lipschutz S. [64] *Set Theory and Related Topics.* McGraw-Hill, 1964 (Schaum's Outline Series). [§2A1.]

Loève M. [77] *Probability Theory I.* Springer, 1977. [Chap. 27 *intro.*, 274H.]

Luxemburg W.A.J. & Zaanen A.C. [71] *Riesz Spaces I.* North-Holland, 1971. [241F.]

Mozzochi C.J. [71] *On the Pointwise Convergence of Fourier Series.* Springer, 1971 (Lecture Notes in Mathematics 199). [§286 *notes*.]

Naor A. [04] 'Proof of the uniform convexity lemma', `http://www.cims.nyu.edu/~naor/homepage files/inequality.pdf`, 26.2.04. [244O.]

Rényi A. [70] *Probability Theory.* North-Holland, 1970. [274Hc.]

Roitman J. [90] *An Introduction to Set Theory.* Wiley, 1990. [§2A1.]

Roselli P. & Willem M. [02] 'A convexity inequality', Amer. Math. Monthly 109 (2002) 64-70. [244Ym.]

Schipp F. [78] 'On Carleson's method', pp. 679-695 in ALEXITS 78. [§286 *notes*.]

Semadeni Z. [71] *Banach spaces of continuous functions I.* Polish Scientific Publishers, 1971. [§253 *notes*.]

Shiryayev A. [84] *Probability.* Springer, 1984. [285N.]

Sjölin P. [71] 'Convergence almost everywhere of certain singular integrals and multiple Fourier series', Arkiv för Math. 9 (1971) 65-90. [§286 *notes*.]

Steele J.M. [86] 'An Efron-Stein inequality of nonsymmetric statistics', Annals of Statistics 14 (1986) 753-758. [274Xj.]

Zaanen A.C. [83] *Riesz Spaces II.* North-Holland, 1983. [241F.]

Zygmund A. [59] *Trigonometric Series.* Cambridge U.P, 1959. [§244 *notes*, §282 *notes*, 284Xj.]

Index to volumes 1 and 2

Principal topics and results

The general index below is intended to be comprehensive. Inevitably the entries are voluminous to the point that they are often unhelpful. I have therefore prepared a shorter, better-annotated, index which will, I hope, help readers to focus on particular areas. It does not mention definitions, as the bold-type entries in the main index are supposed to lead efficiently to these; and if you draw blank here you should always, of course, try again in the main index. Entries in the form of mathematical assertions frequently omit essential hypotheses and should be checked against the formal statements in the body of the work.

General index

References in **bold** type refer to definitions; references in *italics* are passing references. Definitions marked with **>** are those in which my usage is dangerously at variance with that of some other author or authors.