

SPRINGER TEXTS IN STATISTICS

Third Edition

PROBABILITY THEORY

Independence,
Interchangeability,
Martingales

Yuan Shih Chow • Henry Teicher

Copyrighted material

Springer Texts in Statistics

Advisors:

George Casella Stephen Fienberg Ingram Olkin

Springer

New York

Berlin

Heidelberg

Hong Kong

London

Milan

Paris

Tokyo

Springer Texts in Statistics

- Alfred*: Elements of Statistics for the Life and Social Sciences
Berger: An Introduction to Probability and Stochastic Processes
Bilodeau and Brenner: Theory of Multivariate Statistics
Blom: Probability and Statistics: Theory and Applications
Brockwell and Davis: Introduction to Times Series and Forecasting,
Second Edition
Chow and Teicher: Probability Theory: Independence, Interchangeability,
Martingales, Third Edition
Christensen: Advanced Linear Modeling: Multivariate, Time Series, and
Spatial Data; Nonparametric Regression and Response Surface
Maximization, Second Edition
Christensen: Log-Linear Models and Logistic Regression, Second Edition
Christensen: Plane Answers to Complex Questions: The Theory of Linear
Models, Third Edition
Creighton: A First Course in Probability Models and Statistical Inference
Davis: Statistical Methods for the Analysis of Repeated Measurements
Dean and Voss: Design and Analysis of Experiments
du Toit, Steyn, and Stumpf: Graphical Exploratory Data Analysis
Durrett: Essentials of Stochastic Processes
Edwards: Introduction to Graphical Modelling, Second Edition
Finkelstein and Levin: Statistics for Lawyers
Flury: A First Course in Multivariate Statistics
Jobson: Applied Multivariate Data Analysis, Volume I: Regression and
Experimental Design
Jobson: Applied Multivariate Data Analysis, Volume II: Categorical and
Multivariate Methods
Kalbfleisch: Probability and Statistical Inference, Volume I: Probability,
Second Edition
Kalbfleisch: Probability and Statistical Inference, Volume II: Statistical Inference,
Second Edition
Karr: Probability
Keyfitz: Applied Mathematical Demography, Second Edition
Kiefer: Introduction to Statistical Inference
Kokoska and Nevison: Statistical Tables and Formulae
Kulkarni: Modeling, Analysis, Design, and Control of Stochastic Systems
Lange: Applied Probability
Lehmann: Elements of Large-Sample Theory
Lehmann: Testing Statistical Hypotheses, Second Edition
Lehmann and Casella: Theory of Point Estimation, Second Edition
Lindman: Analysis of Variance in Experimental Design
Lindsey: Applying Generalized Linear Models

(continued after index)

Yuan Shih Chow Henry Teicher

Probability Theory

Independence, Interchangeability,
Martingales

Third Edition



Springer

This One



JTF-FTE-GF3B

Yuan Shih Chow
Department of Statistics
Columbia University
New York, NY 10027
USA

Henry Teicher
Department of Statistics
Rutgers University
New Brunswick, NJ 08903
USA

Editorial Board

Stephen Fienberg
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213-3890
USA

Ingram Olkin
Department of Statistics
Stanford University
Stanford, CA 94305
USA

George Casella
Biometrics Unit
Cornell University
Ithaca, NY 14853-7081

Library of Congress Cataloging-in-Publication Data
Chow, Yuan Shih, 1924-

Probability theory : independence, interchangeability, martingales /
Yuan Shih Chow, Henry Teicher. - 3rd ed.
p. cm. - (Springer texts in statistics)
Includes bibliographical references and index.
1. Probabilities. 2. Martingales (Mathematics) I. Teicher, Henry. II. Title. III. Series.
QA273.C573 1997
519.2 - dc21

97-9299

ISBN 0-387-40607-7

Printed on acid-free paper.

First softcover printing, 2003.

© 1997, 1988, 1978 Springer-Verlag New York, Inc.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden. The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

SPIN 10947210

www.springer-ny.com

Springer-Verlag New York Berlin Heidelberg
A member of BertelsmannSpringer Science+Business Media GmbH

Copyrighted material

To our teachers
J. L. Doob and J. Wolfowitz

Copyrighted material

Preface to the Third Edition

Apart from some additional theorems and examples, simplification of proofs, and correction of typographical errors, the main change is the addition of section 7.5, dealing with *U*-statistics. In the first two editions (1978, 1988) Lemma 5.4.5 was contained in the proof of Theorem 10.4.1 but seems worth highlighting here.

Copyrighted material

Preface to the Second Edition

Apart from new examples and exercises, some simplifications of proofs, minor improvements, and correction of typographical errors, the principal change from the first edition is the addition of section 9.5, dealing with the central limit theorem for martingales and more general stochastic arrays.

Copyrighted material

Preface to the First Edition

Probability theory is a branch of mathematics dealing with chance phenomena and has clearly discernible links with the real world. The origins of the subject, generally attributed to investigations by the renowned French mathematician Fermat of problems posed by a gambling contemporary to Pascal, have been pushed back a century earlier to the Italian mathematicians Cardano and Tartaglia about 1570 (Ore, 1953). Results as significant as the Bernoulli weak law of large numbers appeared as early as 1713, although its counterpart, the Borel strong law of large numbers, did not emerge until 1909. Central limit theorems and conditional probabilities were already being investigated in the eighteenth century, but the first serious attempts to grapple with the logical foundations of probability seem to be Keynes (1921), von Mises (1928; 1931), and Kolmogorov (1933).

An axiomatic mold and measure-theoretic framework for probability theory was furnished by Kolmogorov. In this so-called objective or measure-theoretic approach, definitions and axioms are so chosen that the empirical realization of an event is the outcome of a not completely determined physical experiment—an experiment which is at least conceptually capable of indefinite repetition (this notion is due to von Mises). The concrete or intuitive counterpart of the probability of an event is a long run or limiting frequency of the corresponding outcome.

In contradistinction to the objective approach—where typical realizations of events might be: a coin falls heads, more than 50 cars reach a busy intersection during a specified period, a continuously burning light bulb fails within 1000 hours—the subjective approach to probability advocated by Keynes is designed to encompass realizations such as: it will rain tomorrow, life exists on the planet Saturn, the Iliad and the Odyssey were written by the same author—despite the fact that the experiments in question are clearly

unrepeatable. Here the empirical counterpart of probability is degree or intensity of belief.

It is tempting to try to define probability as a limit of frequencies (as advocated by von Mises) rather than as a real number between zero and one satisfying certain postulates (as in the objective approach). Unfortunately, incorporation of repeatability as a postulate (von Mises' "randomness axiom") complicates matters while simultaneously circumscribing the notion of an event. Thus, the probability of the occurrence infinitely often of some particular event in an infinite sequence of repetitions of an experiment—which is of considerable interest in the Kolmogorov schema—is proscribed in (the 1964 rendition of) the von Mises approach (1931). Possibly for these reasons, the frequency approach appears to have lost out to the measure-theoretic. It should be pointed out, however, that justification of the measure-theoretic approach via the Borel strong law of large numbers is circular in that the convergence of the observed frequency of an event to its theoretically defined probability (as the number of repetitions increases) is not pointwise but can only be defined in terms of the concept being justified, viz., probability. If, however, one is willing to ascribe an intuitive meaning to the notion of probability one (hence also, probability zero) then the probability p of any intermediate value can be interpreted in this fashion.

A number of axiomatizations for subjective probability have appeared since Keynes with no single approach dominating. Perhaps the greatest influence of subjective probability is outside the realm of probability theory proper and rather in the recent emergence of the Bayesian school of statistics.

The concern of this book is with the measure-theoretic foundations of probability theory and (a portion of) the body of laws and theorems that emerge therefrom. In the 45 years since the appearance of von Mises' and Kolmogorov's works on the foundations of probability, the theory itself has expanded at an explosive pace. Despite this burgeoning, or perhaps because of the very extent thereof, only the topics of independence, interchangeability, and martingales will be treated here. Thus, such important concepts as Markov and stationary processes will not even be defined, although the special cases of sums of independent random variables and interchangeable random variables will be dealt with extensively. Likewise, continuous parameter stochastic processes, although alluded to, will not be discussed. Indeed, the time seems propitious for the appearance of a book devoted solely to such processes and presupposing familiarity with a significant portion of the material contained here.

Particular emphasis is placed in this book on stopping times—on the one hand, as tools in proving theorems, and on the other, as objects of interest in themselves. Apropos of the latter, randomly stopped sums, optimal stopping problems, and limit distributions of sequences of stopping rules (i.e., finite stopping times) are of special interest. Wald's equation and its second-moment analogue, in turn, show the usefulness of such stopped sums in renewal theory and elsewhere in probability. Martingales provide a natural vehicle for stopping times, but a formal treatment of the latter cannot

await development of the former. Thus, stopping times and, in particular, a sequence of copies of a fixed stopping rule appear as early as Chapter 5, thereby facilitating discussion of the limiting behavior of random walks.

Many of the proofs given and a few of the results are new. Occasionally, a classical notion is looked at through new lenses (e.g., reformulation of the Lindeberg condition). Examples, sprinkled throughout, are used in various guises; to extend theory, to illustrate a theorem that has just appeared, to obtain a classical result from one recently proven.

A novel feature is the attempt to intertwine measure and probability rather than, as is customary, set up between them a sharp demarcation. It is surprising how much probability can be developed (Chapters 2, 3) without even a mention of integration. A number of topics treated later in generality are foreshadowed in the very tractable binomial case of Chapter 2.

This book is intended to serve as a graduate text in probability theory. No knowledge of measure or probability is presupposed, although it is recognized that most students will have been exposed to at least an elementary treatment of the latter. The former is confined for the most part to Chapters 1, 4, 6, with convergence appearing in Section 3.3 (i.e., Section 3 of Chapter 3).¹ Readers familiar with measure theory can plunge into Chapter 5 after reading Section 3.2 and portions of Sections 3.1, 3.3, 4.2, 4.3. In any case, Chapter 2 and also Section 3.4 can be omitted without affecting subsequent developments.

Martingales are introduced in Section 7.4, where the upward case is treated and then developed more generally in Chapter 11. Interchangeable random variables are discussed primarily in Sections 7.3 and 9.2. Apropos of terminology, "interchangeable" is far more indicative of the underlying property than the current "exchangeable," which seems to be a too literal rendition of the french word "échangeable."

A one-year course presupposing measure theory can be built around Chapters 5, 7, 8, 9, 10, 11, 12.

Our warm thanks and appreciation go to Mary Daughaday, Beatrice Williams, and Pat Wolf for their expert typing of the manuscript.

References

- J. M. Keynes, *A Treatise on Probability*, 1921; MacMillan, London, 1943.
- A. Kolmogorov, *Foundations of the Theory of Probability*, 1933; Chelsea, New York, 1950.
- R. von Mises, *Probability, Statistics and Truth*, 1928; Wm. Hodge, London, 1939.
- R. von Mises, *Mathematical Theory of Probability and Statistics*, 1931 (H. Geiringer, editor), Academic Press, N.Y., 1964.
- O. Ore, "Appendix," *Cardano, The Gambling Scholar*, Princeton University Press, 1953; Holt, New York, 1961.
- I. Todhunter, *A History of the Mathematical Theory of Probability*, 1865; Chelsea, New York, 1949.

¹ In the same notational vein, Theorem 3.4.2 signifies Theorem 2 of Section 4 of Chapter 3.

Contents

Preface to the Third Edition	vii
Preface to the Second Edition	ix
Preface to the First Edition	xi
List of Abbreviations	xix
List of Symbols and Conventions	xxi
1 Classes of Sets, Measures, and Probability Spaces	1
1.1 Sets and set operations	1
1.2 Spaces and indicators	4
1.3 Sigma-algebras, measurable spaces, and product spaces	6
1.4 Measurable transformations	12
1.5 Additive set functions, measures, and probability spaces	18
1.6 Induced measures and distribution functions	25
2 Binomial Random Variables	30
2.1 Poisson theorem, interchangeable events, and their limiting probabilities	30
2.2 Bernoulli, Borel theorems	39
2.3 Central limit theorem for binomial random variables, large deviations	45
3 Independence	54
3.1 Independence, random allocation of balls into cells	54
3.2 Borel-Cantelli theorem, characterization of independence, Kolmogorov zero-one law	61

3.3 Convergence in probability, almost certain convergence, and their equivalence for sums of independent random variables	66
3.4 Bernoulli trials	75
4 Integration in a Probability Space	84
4.1 Definition, properties of the integral, monotone convergence theorem	84
4.2 Indefinite integrals, uniform integrability, mean convergence	92
4.3 Jensen, Hölder, Schwarz inequalities	103
5 Sums of Independent Random Variables	113
5.1 Three series theorem	113
5.2 Laws of large numbers	124
5.3 Stopping times, copies of stopping times, Wald's equation	138
5.4 Chung–Fuchs theorem, elementary renewal theorem, optimal stopping	150
6 Measure Extensions, Lebesgue–Stieltjes Measure, Kolmogorov Consistency Theorem	165
6.1 Measure extensions, Lebesgue–Stieltjes measure	165
6.2 Integration in a measure space	171
6.3 Product measure, Fubini's theorem, n -dimensional Lebesgue–Stieltjes measure	184
6.4 Infinite-dimensional product measure space, Kolmogorov consistency theorem	191
6.5 Absolute continuity of measures, distribution functions; Radon–Nikodym theorem	202
7 Conditional Expectation, Conditional Independence, Introduction to Martingales	210
7.1 Conditional expectations	210
7.2 Conditional probabilities, conditional probability measures	222
7.3 Conditional independence, interchangeable random variables	229
7.4 Introduction to martingales	239
7.5 U -statistics	259
8 Distribution Functions and Characteristic Functions	270
8.1 Convergence of distribution functions, uniform integrability, Helly–Bray theorem	270

8.2 Weak compactness, Fréchet–Shohat, Glivenko–Cantelli theorems	281
8.3 Characteristic functions, inversion formula, Lévy continuity theorem	286
8.4 The nature of characteristic functions, analytic characteristic functions, Cramér–Lévy theorem	294
8.5 Remarks on k -dimensional distribution functions and characteristic functions	308
9 Central Limit Theorems	313
9.1 Independent components	313
9.2 Interchangeable components	328
9.3 The martingale case	336
9.4 Miscellaneous central limit theorems	340
9.5 Central limit theorems for double arrays	345
10 Limit Theorems for Independent Random Variables	354
10.1 Laws of large numbers	354
10.2 Law of the iterated logarithm	368
10.3 Marcinkiewicz–Zygmund inequality, dominated ergodic theorems	384
10.4 Maxima of random walks	392
11 Martingales	404
11.1 Upcrossing inequality and convergence	404
11.2 Martingale extension of Marcinkiewicz–Zygmund inequalities	412
11.3 Convex function inequalities for martingales	421
11.4 Stochastic inequalities	432
12 Infinitely Divisible Laws	444
12.1 Infinitely divisible characteristic functions	445
12.2 Infinitely divisible laws as limits	454
12.3 Stable laws	468
Index	479

Copyrighted material

List of Abbreviations

r.v.	random variable
r.v.s	random variables
d.f.	distribution function
c.f.	characteristic function
p.d.f.	probability density function
u.i.	uniform integrability or uniformly integrable
i.o.	infinitely often
a.c.	almost certainly
a.s.	almost surely
a.e.	almost everywhere
i.d.	infinitely divisible
i.i.d.	independent, identically distributed
iff	if and only if
CLT	Central Limit Theorem
WLLN	Weak Law of Large Numbers
SLLN	Strong Law of Large Numbers
LIL	Law of the Iterated Logarithm
m.g.f.	moment generating function
Cov	Covariance

Copyrighted material

List of Symbols and Conventions

$\sigma(\mathcal{G})$	σ -algebra generated by the class \mathcal{G}
$\sigma(X)$	σ -algebra generated by the random variable X
$E X$	expectation of the random variable X
$\int X$	abbreviated form of the integral $\int X dP$
$E^p X$	abbreviated form of $(E X)^p$
$\ X\ _p$	p -norm of X , that is, $(E X ^p)^{1/p}$
$C(F)$	continuity set of the function F
<u>a.c. or a.s. or a.e.</u>	convergence almost certainly or almost surely or almost everywhere
<u>P or d or μ</u>	convergence in probability or in distribution or in μ -measure
<u>L_p</u>	convergence in mean of order p
<u>w or c</u>	weak or complete convergence
\mathcal{B}^n or \mathcal{B}^∞	class of n -dimensional or infinite-dimensional Borel sets
$\Re\{ \cdot \}$	real part of
$\Im\{ \cdot \}$	imaginary part of
\wedge	minimum of
\vee	maximum of
$a \leq \overline{\lim} Y_n \leq b$	simultaneous statement that $a \leq \liminf_{n \rightarrow \infty} Y_n \leq \limsup_{n \rightarrow \infty} Y_n \leq b$
$Z \leq z_2$ $\geq z_1$	simultaneous statement that $Z \leq z_2$ and $Z \geq z_1$

X_F	fictitious r.v. with d.f. F
$m(X)$	median of X
N_{μ, σ^2}	normal r.v. with mean μ , variance σ^2

1 Classes of Sets, Measures, and Probability Spaces

1.1 Sets and Set Operations

A **set**, in the words of Georg Cantor, the founder of modern set theory, is a collection into a whole of definite, well-distinguished objects of our perception or thought. The objects are called elements and the set is the aggregate of these elements. It is very convenient to extend this notion and also envisage a set devoid of elements, a so-called empty set, and this will be denoted by \emptyset . Each element of a set appears only once therein and its order of appearance within the set is irrelevant. A set whose elements are themselves sets will be called a class.

Examples of sets are (i) the set of positive integers denoted by either $\{1, 2, \dots\}$ or $\{\omega : \omega \text{ is a positive integer}\}$ and (ii) the closed interval with end points a and b denoted by either $\{\omega : a \leq \omega \leq b\}$ or $[a, b]$. Analogously, the open interval with end points a and b is denoted by $\{\omega : a < \omega < b\}$ or (a, b) , while $(a, b]$ and $[a, b)$ are designations for $\{\omega : a < \omega \leq b\}$ and $\{\omega : a \leq \omega < b\}$ respectively.

The statement that $\omega \in A$ means that ω is an element of the set A and analogously the assertion $\omega \notin A$ means that ω is not an element of the set A or alternatively that ω does not belong to A . If A and B are sets and every element of A is likewise an element of B , this situation is depicted by writing $A \subset B$ or $B \supset A$, and in such a case the set A is said to be a subset of B or contained in B . If both $A \subset B$ and $B \subset A$, then A and B contain exactly the same elements and are said to be equal, denoted by $A = B$. Note that for every set A , $\emptyset \subset A \subset A$.

A set A is termed **countable** if there exists a one-to-one correspondence between (the elements of) A and (those of) some subset B of the set of all positive integers. If, in this correspondence, $B = \{1, 2, \dots, n\}$, then A is called

a **finite** set (with n elements). It is natural to consider \emptyset as a finite set (with zero elements). A set A which is not countable is called **uncountable** or **nondenumerable**.

If A and B are two sets, the difference $A - B$ is the set of all elements of A which do not belong to B ; the intersection $A \cap B$ or $A \cdot B$ or simply AB is the set of all elements belonging to both A and B ; the union $A \cup B$ is the set of all elements belonging to either A or B (or both); and the symmetric difference $A \Delta B$ is the set of all elements which belong to A or B but not both. Note that

$$A \cup A = A, \quad A \cap A = A, \quad A - A = \emptyset, \quad A - B = A - (AB) \subset A,$$

$$A \cup B = B \cup A \supset A \supset AB = BA, \quad A \Delta B = (A - B) \cup (B - A).$$

Union, intersection, difference, and symmetric difference are termed **set operations**.

If A, B, C are sets and several set operations are indicated, it is, strictly speaking, necessary to indicate via parentheses which operations are to be performed first. However, such specification is frequently unnecessary. For instance, $(A \cup B) \cup C = A \cup (B \cup C)$ and so this double union is independent of order and may be designated simply by $A \cup B \cup C$. Analogously,

$$(AB)C = A(BC) = ABC, \quad (A \Delta B) \Delta C = A \Delta (B \Delta C) = A \Delta B \Delta C,$$

$$A(B \cup C) = AB \cup AC, \quad A(B \Delta C) = AB \Delta AC.$$

If Λ is a nonempty set whose elements λ may be envisaged as tags or labels, $\{A_\lambda : \lambda \in \Lambda\}$ is a nonempty class of sets. The intersection $\bigcap_{\lambda \in \Lambda} A_\lambda$ (resp. union $\bigcup_{\lambda \in \Lambda} A_\lambda$) is defined to be the set of all elements which belong to A_λ for all $\lambda \in \Lambda$ (resp. for some $\lambda \in \Lambda$). Apropos of order of carrying out set operations, if $*$ denotes any one of $\cup, \cap, -, \Delta$, for any set A it follows from the definitions that

$$\bigcup_{\lambda \in \Lambda} A_\lambda * A = \left(\bigcup_{\lambda \in \Lambda} A_\lambda \right) * A, \quad A * \bigcup_{\lambda \in \Lambda} A_\lambda = A * \left(\bigcup_{\lambda \in \Lambda} A_\lambda \right),$$

$$\bigcap_{\lambda \in \Lambda} A_\lambda * A = \left(\bigcap_{\lambda \in \Lambda} A_\lambda \right) * A, \quad A * \bigcap_{\lambda \in \Lambda} A_\lambda = A * \left(\bigcap_{\lambda \in \Lambda} A_\lambda \right).$$

Then

$$A - \bigcup_{\lambda \in \Lambda} A_\lambda = \bigcap_{\lambda \in \Lambda} (A - A_\lambda), \quad A - \bigcap_{\lambda \in \Lambda} A_\lambda = \bigcup_{\lambda \in \Lambda} (A - A_\lambda).$$

For any sequence $\{A_n, n \geq 1\}$ of sets, define

$$\liminf_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k, \quad \limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

and note that, employing the abbreviation i.o. to designate "infinitely often,"

$$\begin{aligned} \liminf_{n \rightarrow \infty} A_n &= \{\omega : \omega \in A_n \text{ for infinitely many } n\} = \{\omega : \omega \in A_n, \text{ i.o.}\} \\ \limsup_{n \rightarrow \infty} A_n &= \{\omega : \omega \in A_n \text{ for all but a finite number of indices } n\}. \end{aligned} \tag{1}$$

To prove, for example, the first relation, let $A = \{\omega : \omega \in A_n, \text{ i.o.}\}$. Then $\omega \in A$ iff for every positive integer m , there exists $n \geq m$ such that $\omega \in A_n$, that is, iff for every positive integer m , $\omega \in \bigcup_{n=m}^{\infty} A_n$ i.e., iff $\omega \in \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n$.

In view of (1), $\liminf A_n \subset \overline{\liminf} A_n$, but these two sets need not be equal (Exercise 3). If $\liminf A_n = \overline{\liminf} A_n = A$ (say), A is called the limit of the sequence A_n ; this situation is depicted by writing $\lim A_n = A$ or $A_n \rightarrow A$. If $A_1 \subset A_2 \subset \dots$ (resp. $A_1 \supset A_2 \supset \dots$) the sequence A_n is said to be **increasing** (resp. **decreasing**). In either case, $\{A_n, n \geq 1\}$ is called **monotone**.

Palpably, for every monotone sequence A_n , $\lim_{n \rightarrow \infty} A_n$ exists; in fact, if $\{A_n\}$ is increasing, $\lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n$, while if $\{A_n\}$ is decreasing, $\lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n$. Consequently, for any sequence of sets A_n ,

$$\overline{\liminf}_{n \rightarrow \infty} A_n = \lim_{n \rightarrow \infty} \bigcup_{k=n}^{\infty} A_k,$$

$$\underline{\liminf}_{n \rightarrow \infty} A_n = \lim_{n \rightarrow \infty} \bigcap_{k=n}^{\infty} A_k.$$

EXERCISES 1.1

- Prove (i) if A_n is countable, $n \geq 1$, so is $\bigcup_{n=1}^{\infty} A_n$; (ii) if A is uncountable and $B \supset A$, then B is uncountable.
- Show that $\bigcup_{n=1}^{\infty} [0, n/(n+1)] = [0, 1]$, $\bigcap_{n=1}^{\infty} (0, 1/n) = \emptyset$.
- Prove that $\overline{\liminf}_{n \rightarrow \infty} A_n \subset \overline{\liminf}_{n \rightarrow \infty} A_n$. Specify $\overline{\liminf} A_n$ and $\underline{\liminf} A_n$ when $A_{2j} = B$, $A_{2j-1} = C$, $j = 1, 2, \dots$
- Verify that $\bigcup_{n=1}^{\infty} A_n = \lim_{n \rightarrow \infty} \bigcup_{j=1}^n A_j$ and $\bigcap_{n=1}^{\infty} A_n = \lim_{n \rightarrow \infty} \bigcap_{j=1}^n A_j$. Moreover, if $\{A_n, n \geq 1\}$ is a sequence of disjoint sets, i.e., $A_i \cap A_j = \emptyset$, $i \neq j$, then

$$\lim_{n \rightarrow \infty} \bigcup_{j=n}^{\infty} A_j = \emptyset.$$

- Prove that $\overline{\liminf}_n (A_n \cup B_n) = \overline{\liminf}_n A_n \cup \overline{\liminf}_n B_n$ and $\underline{\liminf}_n A_n \cdot B_n = \underline{\liminf}_n A_n \cdot \underline{\liminf}_n B_n$. Moreover, $\lim_n A_n = A$ and $\lim_n B_n = B$ imply $\lim_n (A_n \cup B_n) = A \cup B$ and

$$\lim_n A_n B_n = AB.$$

- Demonstrate that if B is a countable set and $B_n = \{(b_1, \dots, b_n) : b_i \in B \text{ for } 1 \leq i \leq n\}$, then B_n is countable, $n \geq 1$.
- Prove that the set S consisting of all infinite sequences with entries 0 or 1 is non-denumerable and conclude that the set of real numbers in $[0, 1]$ or any nondegenerate interval is nondenumerable. Hint: If S were countable, i.e., $S = \{s_n, n \geq 1\}$ where $s_n = (x_{n1}, x_{n2}, \dots)$, then $(1 - x_{11}, 1 - x_{22}, \dots, 1 - x_{nn}, \dots)$ would be an infinite sequence of zeros and ones not in S .
- If a_n is a sequence of real numbers, $0 \leq a_n \leq \infty$, prove that

$$\overline{\bigcup}_{n=1}^{\infty} [0, a_n] = \left[0, \sup_{n \geq 1} a_n\right), \quad \bigcup_{n=1}^{\infty} \left[0, \left(\frac{n+1}{n}\right)^n\right] \neq \left[0, \sup_{n \geq 1} \left(\frac{n+1}{n}\right)^n\right].$$

- For any sequence of sets $\{A_n, n \geq 1\}$, define $B_1 = A_1$, $B_{n+1} = B_n \Delta A_{n+1}$, $n \geq 1$. Prove that $\lim_n B_n$ exists iff $\lim_n A_n$ exists and is empty.

1.2 Spaces and Indicators

A **space** Ω is an arbitrary, nonempty set and is usually postulated as a reference or point of departure for further discussion and investigation. Its elements are referred to as points (of the space) and will be denoted generically by ω . Thus $\Omega = \{\omega : \omega \in \Omega\}$.

For any reference space Ω , the **complement** A^c of a subset A of Ω is defined by $A^c = \Omega - A$ and the **indicator** I_A of $A \subset \Omega$ is a function defined on Ω by

$$I_A(\omega) = 1 \quad \text{for } \omega \in A, \quad I_A(\omega) = 0 \quad \text{for } \omega \in A^c.$$

Similarly, for any real function f on Ω and real constants a, b , $I_{[a \leq f \leq b]}$ signifies the indicator of the set $\{\omega : a \leq f(\omega) \leq b\}$.

For any subsets A, B of Ω

$$\begin{aligned} A \subset B &\text{ iff } A^c \supset B^c, \\ (A^c)^c &= A, \quad A \cup A^c = \Omega, \quad A \cdot A^c = \emptyset, \quad I_{A^c} = 1 - I_A, \\ A - B &= AB^c, \quad I_A \leq I_B \quad \text{iff } A \subset B, \quad I_{A \cup B} \leq I_A + I_B, \end{aligned}$$

with the last inequality becoming an equality for all ω iff $AB = \emptyset$. Let Λ be an arbitrary set and $\{A_\lambda, \lambda \in \Lambda\}$ a class of subsets of Ω . It is convenient to adopt the conventions

$$\bigcup_{\lambda \in \emptyset} A_\lambda = \emptyset, \quad \bigcap_{\lambda \in \emptyset} A_\lambda = \Omega.$$

Moreover,

$$\begin{aligned} \left(\bigcup_{\lambda \in \Lambda} A_\lambda \right)^c &= \bigcap_{\lambda \in \Lambda} A_\lambda^c, & \left(\bigcap_{\lambda \in \Lambda} A_\lambda \right)^c &= \bigcup_{\lambda \in \Lambda} A_\lambda^c, \\ I_{\bigcap_{\lambda \in \Lambda} A_\lambda} &= \inf_{\lambda \in \Lambda} I_{A_\lambda}, & I_{\bigcup_{\lambda \in \Lambda} A_\lambda} &= \sup_{\lambda \in \Lambda} I_{A_\lambda}. \end{aligned}$$

If $A_\lambda \cdot A_{\lambda'} = \emptyset$ for $\lambda, \lambda' \in \Lambda$ and $\lambda \neq \lambda'$, the sets A_λ are called **disjoint**. A class of disjoint sets will be referred to as a **disjoint class**.

If $\{A_n, n \geq 1\}$ is a sequence of subsets of Ω , then $\{I_{A_n}, n \geq 1\}$ is a sequence of functions on Ω with values 0 or 1 and

$$I_{\underline{\lim} A_n} = \varprojlim_{n \rightarrow \infty} I_{A_n}, \quad I_{\overline{\lim} A_n} = \varinjlim_{n \rightarrow \infty} I_{A_n}.$$

Moreover,

$$I_{\bigcup_{\Gamma} A_n} \leq \sum_1^{\infty} I_{A_n}. \tag{1}$$

Equality holds in (1) iff $\{A_n, n \geq 1\}$ is a disjoint class. The following identity (2) is a refinement of the finite counterpart of (1): For $A_i \subset \Omega, 1 \leq i \leq n$, set

$$s_1 = \sum_1^n I_{A_j}, \quad s_2 = \sum_{1 \leq j_1 < j_2 \leq n} I_{A_{j_1} A_{j_2}}, \dots,$$

$$s_n = \sum_{1 \leq j_1 < \dots < j_n \leq n} I_{A_{j_1} A_{j_2} \dots A_{j_n}} = I_{A_1 A_2 \dots A_n}.$$

Then

$$I_{\cup A_j} = s_1 - s_2 + s_3 - \dots + (-1)^{n-1} s_n. \quad (2)$$

In proof of (2), if for some $\omega \in \Omega, I_{\cup A_j}(\omega) = 0$, clearly $s_k(\omega) = 0, 1 \leq k \leq n$, whence (2) obtains. On the other hand, if $I_{\cup A_j}(\omega) = 1$, then $\omega \in A_j$ for at least one $j, 1 \leq j \leq n$. Suppose that ω belongs to exactly m of the sets A_1, \dots, A_n . Then $s_1(\omega) = m, s_2(\omega) = \binom{m}{2}, \dots, s_m(\omega) = 1, s_{m+1}(\omega) = \dots = s_n(\omega) = 0$, whence

$$s_1 - s_2 + \dots + (-1)^{n-1} s_n = m - \binom{m}{2} + \dots + (-1)^{m-1} \binom{m}{m} = 1 = I_{\cup A_j}.$$

EXERCISES 1.2

1. Verify that

$$A \Delta B = A^c \Delta B^c, \quad C = A \Delta B \text{ iff } A = B \Delta C,$$

$$\bigcup_1^\infty A_n \Delta \bigcup_1^\infty B_n \subset \bigcup_1^\infty (A_n \Delta B_n),$$

$$\bigcap_1^\infty A_n \Delta \bigcap_1^\infty B_n \subset \bigcup_1^\infty (A_n \Delta B_n).$$

2. Prove that $(\lim_{n \rightarrow \infty} A_n)^c = \overline{\lim}_{n \rightarrow \infty} A_n^c$ and $(\overline{\lim}_{n \rightarrow \infty} A_n)^c = \lim_{n \rightarrow \infty} A_n^c$; also that $\lim B_n = B$ implies $\lim B_n^c = B^c$ and $\lim A \Delta B_n = A \Delta B$.

3. Prove that $I_{\lim A_n} = \overline{\lim} I_{A_n}$ and that $I_{\lim A_n} = \lim I_{A_n}$ whenever either side exists.

4. If $A_n \subset \Omega, n \geq 1$, show that

$$I_{\bigcup_{n=1}^\infty A_n} = \max_{n \geq 1} I_{A_n}, \quad I_{\bigcap_{n=1}^\infty A_n} = \min_{n \geq 1} I_{A_n}.$$

5. If f is a real function on Ω , then $f^2 = f$ iff f is an indicator of some subset of Ω .

6. Apropos of (2), prove that if B_m is the set of points belonging to exactly $m (1 \leq m \leq n)$ of A_1, \dots, A_n , then

$$I_{B_m} = s_m - \binom{m+1}{m} s_{m+1} + \binom{m+2}{m} s_{m+2} - \dots + (-1)^{n-m} \binom{n}{m} s_n. \quad (3)$$

7. If $\{f_n, n \geq 0\}$ is a sequence of real functions on Ω with $f_n \uparrow f_0$ and $A_n = \{\omega : f_n(\omega) > c\}$, then $A_n \subset A_{n+1}$ and $\lim A_n = A_0$.

8. If $\{f_n, n \geq 0\}$ is a sequence of real functions with $f_n \uparrow f_0$ and $g_n = f_n I_{[a \leq f_n \leq b]}$ for some constants $-\infty \leq a < b \leq \infty$, then $\{g_n, n \geq 1\}$ is not necessarily increasing. However, if for $n \geq 0$

$$f'_n = f_n I_{[a \leq f_n \leq b]} + a I_{[f_n < a]} + b I_{[f_n > b]},$$

then $f'_n \uparrow f'_0$.

9. If f_1 and f_2 are real functions on Ω , prove that for all real x and rational r

$$\{\omega : f_1(\omega) + f_2(\omega) < x\} = \bigcup_{\text{all } r} \{\omega : f_1(\omega) < r\} \cdot \{\omega : f_2(\omega) < x - r\}.$$

1.3 σ -Algebras, Measurable Spaces, and Product Spaces

Let Ω be a space.

Definition. A nonempty class \mathcal{A} of subsets of Ω is an **algebra** if

- i. $A^c \in \mathcal{A}$ whenever $A \in \mathcal{A}$,
- ii. $A_1 \cup A_2 \in \mathcal{A}$ whenever $A_j \in \mathcal{A}, j = 1, 2$.

Moreover, \mathcal{A} is a **σ -algebra** if, in addition,

- iii. $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ whenever $A_n \in \mathcal{A}, n \geq 1$.

Evidently, (ii) implies that for every positive integer n , $\bigcup_1^n A_j \in \mathcal{A}$ whenever $A_j \in \mathcal{A}, 1 \leq j \leq n$, while both (i) and (ii) entail $A_1 A_2 \in \mathcal{A}$, if $A_j \in \mathcal{A}, j = 1, 2$; also, since \mathcal{A} is nonempty, $\Omega \in \mathcal{A}, \emptyset \in \mathcal{A}$. Clearly, (iii) implies (ii) by taking $A_n = A_2, n \geq 2$. Note that a σ -algebra is closed under countable intersections.

Definition. A nonempty class \mathcal{A} of subsets of Ω is a **monotone class** if $\lim A_n \in \mathcal{A}$ for every monotone sequence $A_n \in \mathcal{A}, n \geq 1$.

Obviously, a σ -algebra is a monotone class. Conversely, a monotone algebra \mathcal{A} (i.e., a monotone class which is simultaneously an algebra) is a σ -algebra. For if $A_n \in \mathcal{A}, n \geq 1$, then $B_n = \bigcup_{j=1}^n A_j \in \mathcal{A}, n \geq 1$, whence $\bigcup_{j=1}^{\infty} A_j = \lim_n B_n \in \mathcal{A}$.

Let S_{Ω} be the class of all subsets of Ω and $T_{\Omega} = \{\emptyset, \Omega\}$. Then S_{Ω} and T_{Ω} are σ -algebras and for any σ -algebra U_{Ω} of subsets of Ω , $T_{\Omega} \subset U_{\Omega} \subset S_{\Omega}$.

Definition. The **minimal algebra** \mathcal{E}' (resp. σ -algebra, monotone class) containing a nonempty class \mathcal{E} of subsets of Ω , is an algebra (resp. σ -algebra, monotone class) such that

- i. $\mathcal{E}' \supset \mathcal{E}$,

- ii. $\mathcal{E}'' \supset \mathcal{E}$ whenever $\mathcal{E}'' \supset \mathcal{E}$ and \mathcal{E}'' is an algebra (resp. σ -algebra, monotone class).

Such a minimal algebra \mathcal{E}' (resp. σ -algebra, monotone class) containing \mathcal{E} is also called the algebra (resp. σ -algebra, monotone class) generated by \mathcal{E} and is denoted by $\mathcal{A}(\mathcal{E})$ (resp. $\sigma(\mathcal{E})$, $m(\mathcal{E})$).

To demonstrate the existence of $\mathcal{A}(\mathcal{E})$ (resp. $\sigma(\mathcal{E})$, $m(\mathcal{E})$), let

$$Q_{\mathcal{E}} = \{\mathcal{B}: \mathcal{B} \supset \mathcal{E}, \mathcal{B} \text{ is an algebra (resp. } \sigma\text{-algebra, monotone class)}\}$$

Then $Q_{\mathcal{E}}$ is nonempty since $S_{\Omega} \in Q_{\mathcal{E}}$. Since an arbitrary intersection of algebras (resp. σ -algebras, monotone classes) is itself an algebra (resp. σ -algebra, monotone class), if $\mathcal{D} = \bigcap_{\mathcal{B} \in Q_{\mathcal{E}}} \mathcal{B}$, then \mathcal{D} is an algebra (resp. σ -algebra, monotone class) and $\mathcal{D} \supset \mathcal{E}$. Obviously, \mathcal{D} is minimal.

For an arbitrary class \mathcal{E} of sets, it is impossible to give a constructive procedure for obtaining $\sigma(\mathcal{E})$ (see Hausdorff (1957)). However, some indication of its structure is given by

Theorem 1. *If \mathcal{A} is an algebra, $m(\mathcal{A}) = \sigma(\mathcal{A})$.*

PROOF. By a prior observation, $\mathcal{A} \subset m(\mathcal{A}) \subset \sigma(\mathcal{A})$. To prove that $m(\mathcal{A}) \supset \sigma(\mathcal{A})$ it suffices to show that $m(\mathcal{A})$ is a σ -algebra, and indeed merely an algebra, again via a prior comment. To this end, for any $B \subset \Omega$, set $\mathcal{D}_B = \{A: A \cdot B^c \in m(\mathcal{A}), B \cdot A^c \in m(\mathcal{A}), A \cup B \in m(\mathcal{A})\}$. If $A_n \in \mathcal{D}_B$, $A_{n+1} \subset A_n$, $n \geq 1$, and $A = \bigcap_{n=1}^{\infty} A_n$, then $\{A_n \cdot B^c\}$, $\{B \cdot A_n^c\}$, and $\{A_n \cup B\}$ are monotone sequences in $m(\mathcal{A})$. Thus, $A \cdot B^c$, $B \cdot A^c$, and $A \cup B$ belong to $m(\mathcal{A})$, whence $A \in \mathcal{D}_B$. Similarly, for increasing A_n in \mathcal{D}_B , $n \geq 1$, $\lim A_n \in \mathcal{D}_B$. Hence \mathcal{D}_B is a monotone class for all $B \subset \Omega$. If $B \in \mathcal{A}$, $\mathcal{D}_B \supset \mathcal{A}$ since \mathcal{A} is an algebra. Therefore, $\mathcal{D}_B \supset m(\mathcal{A})$ for $B \in \mathcal{A}$. Since $A \in \mathcal{D}_B$ iff $B \in \mathcal{D}_A$, necessarily $\mathcal{D}_B \supset m(\mathcal{A})$ for $B \in m(\mathcal{A})$. This latter means that if $B \in m(\mathcal{A})$ and $E \in m(\mathcal{A})$, then $B \cdot E^c$, $E \cdot B^c$, and $B \cup E$ all are in $m(\mathcal{A})$. In other words, $m(\mathcal{A})$ is an algebra. (Recall that $\Omega \in \mathcal{A} \subset m(\mathcal{A})$). \square

Definitions. A nonempty class \mathcal{A} of subsets of Ω is a π -class if $AB \in \mathcal{A}$ whenever both $A, B \in \mathcal{A}$, whereas it is a λ -class if (i) $\Omega \in \mathcal{A}$, (ii) $A - B \in \mathcal{A}$ for $A, B \in \mathcal{A}$ and $A \supset B$, (iii) $\lim A_n \in \mathcal{A}$ for every increasing sequence $A_n \in \mathcal{A}$, $n \geq 1$.

Note that a λ -class is closed under complementation and hence (iv) $A \cup B \in \mathcal{A}$ if $A, B \in \mathcal{A}$ and $AB = \emptyset$ (since $(A \cup B)^c = B^c - A \in \mathcal{A}$ via (ii)). In view of $A \cup B = A \cup BA^c$, a λ -class which is simultaneously a π -class is a σ -algebra via (iv) and (iii). Furthermore (same proof as for $\mathcal{A}(\mathcal{E})$), there exists a unique minimal λ -class $\lambda(\mathcal{E})$ (resp. π -class) containing a given class \mathcal{E} of sets.

Theorem 2. *If a λ -class \mathcal{A} contains a π -class \mathcal{D} , then $\mathcal{A} \supset \sigma(\mathcal{D})$, the σ -algebra generated by \mathcal{D} . In particular, if \mathcal{D} is a π -class, $\lambda(\mathcal{D}) = \sigma(\mathcal{D})$.*

PROOF. If \mathcal{G} is the minimal λ -class containing \mathcal{D} , it suffices to show that \mathcal{G} is a π -class since for any other λ -class \mathcal{G}' containing \mathcal{D} necessarily $\mathcal{G}' \supset \mathcal{G} \supset \sigma(\mathcal{D})$ because \mathcal{G} is a σ -algebra. To this end, set $\mathcal{G}_1 = \{A : A \subset \Omega, AD \in \mathcal{G} \text{ for all } D \in \mathcal{D}\}$. Clearly, \mathcal{G}_1 is a λ -class containing \mathcal{D} , whence $\mathcal{G}_1 \supset \mathcal{G}$. Hence, $AD \in \mathcal{G}$ whenever $D \in \mathcal{D}, A \in \mathcal{G}$. Thus, if $\mathcal{G}_2 = \{B : B \subset \Omega, AB \in \mathcal{G} \text{ for all } A \in \mathcal{G}\}$, then \mathcal{G}_2 is a λ -class containing \mathcal{D} . Consequently, $\mathcal{G}_2 \supset \mathcal{G}$ and so $AB \in \mathcal{G}$ for both $A, B \in \mathcal{G}$. \square

If A is a fixed subset of Ω and \mathcal{G} is any class of subsets of Ω , the class of all sets of the form BA with $B \in \mathcal{G}$ will be denoted by $A \cdot \mathcal{G}$ or $\mathcal{G} \cdot A$ or $\mathcal{G} \cap A$. If A (rather than Ω) is considered as the reference space (relative to which complementation occurs) then $\mathcal{G} \cap A$ is a σ -algebra (resp. π -class, algebra) relative to A if \mathcal{G} is a σ -algebra (resp. π -class, algebra) relative to Ω . When more than one reference space crops up, this will be appended to the class of sets under consideration. Thus, $\sigma_A(\mathcal{D})$ denotes the σ -algebra generated by the class \mathcal{D} but relative to the reference space A .

Theorem 3. For every non-empty class \mathcal{D} of subsets of Ω and every non-empty set $A \subset \Omega$, $\sigma_\Omega(\mathcal{D}) \cap A = \sigma_A(\mathcal{D} \cap A)$.

PROOF. Let $\mathcal{G} = \{B : B \subset \Omega, B \cdot A \in \sigma_A(\mathcal{D} \cap A)\}$. Then \mathcal{G} is a σ -algebra relative to Ω and $\mathcal{G} \supset \mathcal{D}$, whence $\mathcal{G} \supset \sigma_\Omega(\mathcal{D})$. In other words, $\sigma_\Omega(\mathcal{D}) \cap A \subset \sigma_A(\mathcal{D} \cap A)$. On the other hand, $\sigma_\Omega(\mathcal{D}) \cap A$ is a σ -algebra containing $\mathcal{D} \cap A$, whence $\sigma_\Omega(\mathcal{D}) \cap A \supset \sigma_A(\mathcal{D} \cap A)$. \square

Definition. If \mathcal{A} is a σ -algebra relative to Ω , then the pair (Ω, \mathcal{A}) is called a **measurable space**. The sets of \mathcal{A} are called **measurable sets**.

It is interesting to contrast the definition of a measurable space with that of a topological space. Both measurable and topological spaces engender natural product spaces.

For any measurable spaces $(\Omega_i, \mathcal{A}_i), i = 1, 2, \dots$ define for $n \geq 2$

$$\begin{aligned} \bigtimes_{i=1}^n A_i &= \{(\omega_1, \omega_2, \dots, \omega_n) : \omega_i \in A_i \subset \Omega_i, 1 \leq i \leq n\}, \\ A_1 \times A_2 \times \cdots \times A_n &= \bigtimes_{i=1}^n A_i, \\ \bigtimes_{i=1}^n \mathcal{A}_i &= \sigma \left(\left\{ \bigtimes_{i=1}^n A_i : A_i \in \mathcal{A}_i, 1 \leq i \leq n \right\} \right), \\ \bigtimes_{i=1}^n (\Omega_i, \mathcal{A}_i) &= \left(\bigtimes_{i=1}^n \Omega_i, \bigtimes_{i=1}^n \mathcal{A}_i \right). \end{aligned}$$

Then $\bigtimes_{i=1}^n \Omega_i$ is called the **product space** (with components $\Omega_i, 1 \leq i \leq n$). Moreover, the measurable space $\bigtimes_{i=1}^n (\Omega_i, \mathcal{A}_i)$ is termed the n -dimensional **product-measurable space** and $\bigtimes_{i=1}^n \mathcal{A}_i$ is the **product σ -algebra**.

Sets of the form $\bigtimes_{i=1}^n A_i$ with $A_i \subset \Omega_i$, $1 \leq i \leq n$, are called (n -dimensional) **rectangles** of the product space $\bigtimes_{i=1}^n \Omega_i$, and, moreover, if $A_i \in \mathcal{A}_i$, $1 \leq i \leq n$, they are dubbed **measurable rectangles** or **rectangles with measurable sides**. Clearly, the intersection (but not the union) of any two measurable rectangles of a given product space is a measurable rectangle in that space. In other words, the class of measurable rectangles of $\bigtimes_{i=1}^n \Omega_i$ is a π -class.

Theorem 4. *If $(\Omega_i, \mathcal{A}_i)$, $1 \leq i \leq n$ are measurable spaces, the class \mathcal{A} of all finite unions of disjoint rectangles $\bigtimes_{i=1}^n A_i$ with $A_i \in \mathcal{A}_i$, $1 \leq i \leq n$, is the algebra generated by the class of all measurable rectangles of the product space $\bigtimes_{i=1}^n \Omega_i$.*

PROOF. Let \mathcal{G} denote the π -class of measurable rectangles of $\bigtimes_{i=1}^n \Omega_i$. Now \mathcal{A} is also a π -class since if $A_i = \bigcup_{j=1}^{n_i} E_{ij} \in \mathcal{A}_i$, $i = 1, 2$, with $E_{ij} \in \mathcal{G}$, then $A_1 \cdot A_2 = \bigcup_{k, l} E_{1k} E_{2l} \in \mathcal{A}$. Moreover, if $E = E_1 \times \cdots \times E_n \in \mathcal{G}$,

$$\begin{aligned} E^c &= E_1 \times \cdots \times E_{n-1} \times E_n^c \cup E_1 \times \cdots \times E_{n-2} \times E_{n-1}^c \times \Omega_n \\ &\quad \cup \cdots \cup E_1^c \times \Omega_2 \times \cdots \times \Omega_n = \bigcup_1^n D_i \text{ (say)} \in \mathcal{A} \end{aligned}$$

and so, if $A = \bigcup_{j=1}^r E_j \in \mathcal{A}$ with $E_j \in \mathcal{G}$,

A^c = \bigcap_{j=1}^r E_j^c = \bigcap_{j=1}^r \bigcup_{i=1}^n D_i^{(j)} = \bigcup D_{i_1}^{(1)} D_{i_2}^{(2)} \cdots D_{i_r}^{(r)} \in \mathcal{A}.

Consequently, if $A, B \in \mathcal{A}$, the preceding ensures that

$$A \cup B = A \cup B \cdot A^c \in \mathcal{A}.$$

Hence \mathcal{A} is an algebra, and since $\mathcal{A} \supset \mathcal{G}$, necessarily $\mathcal{A} \supset \mathcal{A}(\mathcal{G})$. On the other hand, every finite union of disjoint rectangles with measurable sides $\in \mathcal{A}(\mathcal{G})$ and so $\mathcal{A} \subset \mathcal{A}(\mathcal{G})$, whence $\mathcal{A} = \mathcal{A}(\mathcal{G})$. \square

Clearly, the σ -algebra $\bigtimes_{i=1}^n \mathcal{A}_i$, generated by the rectangles with measurable sides, is also the σ -algebra generated by the algebra \mathcal{A} of Theorem 4.

The points of $\bigtimes_{i=1}^3 \Omega_i$, $(\Omega_1 \times \Omega_2) \times \Omega_3$, and $\Omega_1 \times (\Omega_2 \times \Omega_3)$ are respectively of the form $(\omega_1, \omega_2, \omega_3)$, $((\omega_1, \omega_2), \omega_3)$, and $(\omega_1, (\omega_2, \omega_3))$, where $\omega_i \in \Omega_i$ for $i = 1, 2, 3$ and formally they are different. However, between any two of the three points there is a natural one-to-one correspondence and so each one of the points will be identified with $(\omega_1, \omega_2, \omega_3)$. Similarly, $((\omega_1, \omega_2), (\omega_3, \omega_4))$ will be identified with $(\omega_1, \omega_2, \omega_3, \omega_4)$, etc. Under this convention it is easily seen that for $1 \leq m < n$

$$\bigtimes_{i=1}^n \Omega_i = \left(\bigtimes_{i=1}^m \Omega_i \right) \times \left(\bigtimes_{i=m+1}^n \Omega_i \right)$$

and not so easily that

$$\bigtimes_{i=1}^n \mathcal{A}_i = \left(\bigtimes_{i=1}^m \mathcal{A}_i \right) \times \left(\bigtimes_{i=m+1}^n \mathcal{A}_i \right).$$

To verify the latter, let

$$\begin{aligned} X &= \bigtimes_{i=1}^n \Omega_i, & Y &= \bigtimes_{i=1}^m \Omega_i, & Z &= \bigtimes_{i=m+1}^n \Omega_i, \\ \mathcal{F} &= \bigtimes_{i=1}^n \mathcal{A}_i, & \mathcal{G} &= \bigtimes_{i=1}^m \mathcal{A}_i, & \mathcal{H} &= \bigtimes_{i=m+1}^n \mathcal{A}_i, \\ \mathcal{F}' &= \left\{ \bigtimes_{i=1}^n A_i : A_i \in \mathcal{A}_i, 1 \leq i \leq n \right\}, & \mathcal{G}' &= \left\{ \bigtimes_{i=1}^m A_i : A_i \in \mathcal{A}_i, 1 \leq i \leq m \right\}, \\ \mathcal{H}' &= \left\{ \bigtimes_{i=m+1}^n A_i : A_i \in \mathcal{A}_i, m+1 \leq i \leq n \right\}. \end{aligned}$$

By definition, $\mathcal{F} = \sigma_X(\mathcal{F}')$, $\mathcal{G} = \sigma_Y(\mathcal{G}')$, $\mathcal{H} = \sigma_Z(\mathcal{H}')$, and $\mathcal{G} \times \mathcal{H} = \sigma(\{A \times B : A \in \mathcal{G}, B \in \mathcal{H}\})$, where σ_Q is the σ -algebra relative to the space Q . Now if $\bigtimes_{i=1}^n A_i \in \mathcal{F}'$, then $\bigtimes_{i=1}^m A_i \in \mathcal{G}'$ and $\bigtimes_{i=m+1}^n A_i \in \mathcal{H}'$, whence $\bigtimes_{i=1}^n A_i = (\bigtimes_{i=1}^m A_i) \times (\bigtimes_{i=m+1}^n A_i) \in \mathcal{G} \times \mathcal{H}$, implying $\mathcal{F} \subset \mathcal{G} \times \mathcal{H}$. On the other hand, if $A \in \mathcal{G}$ and $B \in \mathcal{H}$, then $A \times B = (A \times \Omega_{m+1} \times \dots \times \Omega_n) \cap (\Omega_1 \times \dots \times \Omega_m \times B) \in \mathcal{F}$ and so $\mathcal{G} \times \mathcal{H} \subset \mathcal{F}$. \square

In the product measurable space $\bigtimes_{i=1}^n (\Omega_i, \mathcal{A}_i)$, $n = 1, 2, \dots$, sets of the form $A \times (\bigtimes_{i=m+1}^n \Omega_i)$ with $A \in \bigtimes_{i=1}^m \mathcal{A}_i$ for $1 \leq m < n$ are called **cylinders** with m -dimensional base A . The quintessence of a cylinder set is that all but a finite number of coordinates are unrestricted. The notion of a cylinder is important in the case of an infinite-dimensional product space.

Let $(\Omega_i, \mathcal{A}_i)$, $i = 1, 2, \dots$, be measurable spaces and define

$$\begin{aligned} \bigtimes_{i=1}^{\infty} A_i &= \{(\omega_1, \omega_2, \dots) : \omega_i \in A_i \subset \Omega_i, i = 1, 2, \dots\}, \\ \mathcal{G} &= \bigcup_{m=1}^{\infty} \left\{ \bigtimes_{i=1}^{\infty} A_i : A_i \in \mathcal{A}_i, 1 \leq i \leq m, \text{ and } A_i = \Omega_i, i > m \right\}, \\ \bigtimes_{i=1}^{\infty} \mathcal{A}_i &= \sigma(\mathcal{G}), \\ \bigtimes_{i=1}^{\infty} (\Omega_i, \mathcal{A}_i) &= \left(\bigtimes_{i=1}^{\infty} \Omega_i, \bigtimes_{i=1}^{\infty} \mathcal{A}_i \right). \end{aligned}$$

Then $\bigtimes_{i=1}^{\infty} (\Omega_i, \mathcal{A}_i)$ is the **infinite-dimensional product-measurable space** with components $(\Omega_i, \mathcal{A}_i)$. The σ -algebra in question, $\bigtimes_{i=1}^{\infty} \mathcal{A}_i$, is that generated by cylinders with finite-dimensional (measurable) bases (see Exercise 1.3.6).

In the special case where $(\Omega_i, \mathcal{A}_i) = (\Omega, \mathcal{A})$ for all i , a convenient notation is

$$\Omega^n = \bigtimes_{i=1}^n \Omega_i, \quad \mathcal{A}^n = \bigtimes_{i=1}^n \mathcal{A}_i, \quad (\Omega^n, \mathcal{A}^n) = \bigtimes_{i=1}^n (\Omega_i, \mathcal{A}_i), \quad 1 \leq n \leq \infty.$$

A prominent example of the preceding considerations arises when $\Omega = R = [-\infty, \infty] = \{\omega: -\infty \leq \omega \leq \infty\}$. The set R will be called the **real line** and $\pm \infty$ will be considered as real numbers. In contradistinction, $(-\infty, \infty) = \{\omega: -\infty < \omega < \infty\}$ will be termed the **finite real line**, whose elements are finite real numbers. The relationships between $\pm \infty$ and the finite real numbers are such that for every **finite** real number x

$$-\infty < x < \infty, \quad x + \infty = \infty + x = \infty,$$

$$x + (-\infty) = -\infty + x = -\infty,$$

$$\frac{x}{\infty} = \frac{x}{-\infty} = 0,$$

$$\infty + \infty = (\infty) \cdot (\infty) = (-\infty) \cdot (-\infty) = -(-\infty) = \infty,$$

$$-\infty + (-\infty) = \infty \cdot (-\infty) = (-\infty) \cdot \infty = -(\infty) = -\infty,$$

$$x \cdot \infty = \infty \cdot x = (-x) \cdot (-\infty) = (-\infty) \cdot (-x) = \infty \quad \text{or} \quad 0 \quad \text{or} \quad -\infty,$$

according as $x > 0, x = 0, x < 0$. Note that ∞/∞ and $\infty - \infty$ are not defined. For $x \in R$, the sets $[-\infty, x], (-\infty, x), (-\infty, x], [-\infty, x), [x, \infty], (x, \infty), [x, \infty), (x, \infty]$ are termed infinite intervals.

The elements of the σ -algebra \mathcal{B} generated by the class of infinite intervals of the form $(-\infty, x)$, $-\infty < x < \infty$, are known as the **Borel sets** (of the line) or **linear Borel sets** or Borel sets in R . The measurable space (R, \mathcal{B}) is called the **Borel line** or 1-dimensional Borel space.

The product measurable spaces (R^n, \mathcal{B}^n) and $(R^\infty, \mathcal{B}^\infty)$, emanating from the Borel line (R, \mathcal{B}) , are called the **n -dimensional Borel space** and **infinite-dimensional Borel space** respectively. Moreover, the sets of \mathcal{B}^n are termed **n -dimensional Borel sets**, $1 \leq n \leq \infty$. Since every point $\omega \in R^\infty$ is an infinite sequence of real numbers, the infinite dimensional Borel space $(R^\infty, \mathcal{B}^\infty)$ is also alluded to as a sequence space.

For any interval $J \subset R$ the σ -algebra generated by the class of all subintervals of J coincides with $\mathcal{B} \cap J$ according to Theorem 3. Thus, it is natural to describe the σ -algebra in question as the class of Borel subsets of J . Similar comments apply to any n -dimensional interval (rectangle) $J^n \subset R^n$.

EXERCISES 1.3

1. Prove that \mathcal{A} is a σ -algebra if and only if it is both a λ -class and a π -class. If a class $\mathcal{A} \subset \mathcal{D}$, then $\sigma(\mathcal{A}) \subset \sigma(\mathcal{D})$.
2. If \mathcal{F} is an algebra such that for every disjoint sequence $\{F_n, n \geq 1\}$ of sets in \mathcal{F} $\bigcup_{n=1}^{\infty} F_n \in \mathcal{F}$, then \mathcal{F} is a σ -algebra.
3. Let $(\Omega_i, \mathcal{A}_i)$ be measurable spaces and $\mathcal{A}_i = \sigma(\mathcal{G}_i)$ with $\Omega_i \in \mathcal{G}_i$, $1 \leq i \leq n$. If $\mathcal{G} = \{A_1 \times A_2 \times \dots \times A_n : A_i \in \mathcal{G}_i, 1 \leq i \leq n\}$, then $\sigma(\mathcal{G}) = \mathcal{A}_1 \times \mathcal{A}_2 \times \dots \times \mathcal{A}_n$ on $\Omega_1 \times \Omega_2 \times \dots \times \Omega_n$.

4. If $\mathcal{A}_n, n \geq 1$ is an increasing sequence of σ -algebras, then $\mathcal{A} = \bigcup_{n=1}^{\infty} \mathcal{A}_n$ is merely an algebra.
5. The σ -algebra generated by a countable class of disjoint, nonempty sets whose union $= \Omega$ is the class of all unions of these sets.
6. If $(\Omega_i, \mathcal{A}_i), i \geq 1$, are measurable spaces, the class \mathcal{E} of all cylinder sets of $\bigtimes_{i=1}^{\infty} \Omega_i$ with bases in $\bigtimes_{i=1}^m \mathcal{A}_i$ for some $m \geq 1$ is an algebra, but not a σ -algebra. Moreover, setting $\mathcal{D} = \{\bigtimes_{i=1}^m A_i : A_i \in \mathcal{A}_i\}$ verify that $\bigtimes_{i=1}^m \mathcal{A}_i = \sigma(\mathcal{D}) = \sigma(\mathcal{E})$.
7. Let \mathcal{D} be a π -class of subsets of Ω and \mathcal{G} the class of all finite unions of disjoint sets of \mathcal{D} with $\emptyset \in \mathcal{D}$. If $D^c \in \mathcal{G}$ for every $D \in \mathcal{D}$, prove that \mathcal{G} is the algebra generated by \mathcal{D} .
8. Show that the class of Borel sets \mathcal{B} may be generated by $\{(x, \infty], -\infty < x < \infty\}$ or by $\mathcal{S} = \{[\infty, \infty], [-\infty, \infty], [a, b], -\infty \leq a \leq b \leq \infty\}$.
9. Prove that $A = \{(x, y) : x^2 + y^2 < r^2\}$ is a Borel set of R^2 . Hint: A is a countable union of open (classical) rectangles. Utilize this to prove that $\{(x, y) : x^2 + y^2 \leq r^2\}$ is likewise a Borel set and hence also the circumference of a circle.
10. If \mathcal{A} is the class of open sets of $Q^n = (-\infty, \infty)^n$, then $\sigma(\mathcal{A}) = \mathcal{B}^n \cap Q^n$.

1.4 Measurable Transformations

Let Ω_1 and Ω_2 be reference sets and X a function defined on Ω_1 , with values in Ω_2 , the latter being denoted by $X : \Omega_1 \rightarrow \Omega_2$. For every subset A of Ω_2 and class \mathcal{G} of subsets of Ω_2 , define

$$X^{-1}(A) = \{\omega : \omega \in \Omega_1, X(\omega) \in A\}, \quad X^{-1}(\mathcal{G}) = \{X^{-1}(A) : A \in \mathcal{G}\}.$$

The set $X^{-1}(A)$ and the class $X^{-1}(\mathcal{G})$ are called respectively the **inverse images** of the set A and of the class \mathcal{G} . Clearly, if for each $\lambda \in \Lambda$ the set $A_\lambda \subset \Omega_2$, then

$$X^{-1}\left(\bigcup_{\lambda \in \Lambda} A_\lambda\right) = \bigcup_{\lambda \in \Lambda} X^{-1}(A_\lambda), \quad X^{-1}(A_\lambda - A_{\lambda'}) = X^{-1}(A_\lambda) - X^{-1}(A_{\lambda'}),$$

$$X^{-1}\left(\bigcap_{\lambda \in \Lambda} A_\lambda\right) = \bigcap_{\lambda \in \Lambda} X^{-1}(A_\lambda),$$

and hence the inverse image $X^{-1}(\mathcal{G})$ of a σ -algebra \mathcal{G} on Ω_2 is a σ -algebra on Ω_1 , and the class $\{B : B \subset \Omega_2, X^{-1}(B) \in \mathcal{H}\}$ is a σ -algebra on Ω_2 if \mathcal{H} is a σ -algebra on Ω_1 .

Lemma 1. For any mapping $X : \Omega_1 \rightarrow \Omega_2$ and class \mathcal{A} of subsets of Ω_2 ,

$$X^{-1}(\sigma_{\Omega_2}(\mathcal{A})) = \sigma_{\Omega_1}(X^{-1}(\mathcal{A})).$$

PROOF. Since $X^{-1}(\sigma_{\Omega_2}(\mathcal{A}))$ is a σ -algebra containing $X^{-1}(\mathcal{A})$, it must be shown for any σ -algebra \mathcal{D} of subsets of Ω_1 with $\mathcal{D} \supset X^{-1}(\mathcal{A})$ that $\mathcal{D} \supset X^{-1}(\sigma_{\Omega_2}(\mathcal{A}))$. Since \mathcal{D} is a σ -algebra, the class $\mathcal{B} = \{B : B \subset \Omega_2, X^{-1}(B) \in \mathcal{D}\}$ is likewise a

σ -algebra and this together with the relation $\mathcal{B} \supset \mathcal{A}$ implies $\mathcal{B} \supset \sigma_{\Omega_2}(\mathcal{A})$. Thus

$$\mathcal{D} \supset X^{-1}(\mathcal{B}) \supset X^{-1}(\sigma_{\Omega_2}(\mathcal{A})). \quad \square$$

Suppose that $X: \Omega_1 \rightarrow \Omega_2$ and $Y: \Omega_2 \rightarrow \Omega_3$. If $Y(X)$ is defined by the usual composition, that is,

$$Y(X)(\omega) = Y(X(\omega)), \quad \omega \in \Omega,$$

then $Y(X): \Omega_1 \rightarrow \Omega_3$ and $(Y(X))^{-1}(A) = X^{-1}(Y^{-1}(A))$ for every $A \subset \Omega_3$.

If $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ are measurable spaces and $X: \Omega_1 \rightarrow \Omega_2$ then X is said to be a **measurable transformation** from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_2, \mathcal{A}_2)$ or an \mathcal{A}_1 -**measurable function** from Ω_1 to $(\Omega_2, \mathcal{A}_2)$ provided $X^{-1}(\mathcal{A}_2) \subset \mathcal{A}_1$.

Suppose that X is a measurable transformation from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_2, \mathcal{A}_2)$ and Y is a measurable transformation from $(\Omega_2, \mathcal{A}_2)$ to $(\Omega_3, \mathcal{A}_3)$. It follows immediately from the definition that $Y(X)$ is a measurable transformation from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_3, \mathcal{A}_3)$.

Theorem 1. *If X_i is a measurable transformation from (Ω, \mathcal{A}) to $(\Omega_i, \mathcal{A}_i)$, $1 \leq i \leq n$, where n may be infinite, then $X(\omega) = (X_1(\omega), \dots, X_n(\omega))$ is a measurable transformation from (Ω, \mathcal{A}) to the product measurable space $\bigtimes_{i=1}^n (\Omega_i, \mathcal{A}_i)$.*

PROOF. If $n < \infty$, let

$$\mathcal{G} = \left\{ \bigcap_{i=1}^n A_i : A_i \in \mathcal{A}_i, 1 \leq i \leq n \right\},$$

while if $n = \infty$, take

$$\mathcal{G} = \bigcup_{m=1}^{\infty} \left\{ \bigcap_{i=1}^m A_i : A_i \in \mathcal{A}_i, 1 \leq i \leq m, \text{ and } A_i = \Omega_i, i > m \right\}.$$

Then $\sigma(\mathcal{G}) = \bigtimes_{i=1}^{\infty} \mathcal{A}_i$ and by the prior lemma

$$\sigma(X^{-1}(\mathcal{G})) = X^{-1}(\sigma(\mathcal{G})) = X^{-1}\left(\bigtimes_{i=1}^{\infty} \mathcal{A}_i\right).$$

Since $X_i^{-1}(\mathcal{A}_i) \subset \mathcal{A}$ for each i , $X^{-1}(\mathcal{G}) \subset \mathcal{A}$, whence

$$\mathcal{A} \supset \sigma(X^{-1}(\mathcal{G})) = X^{-1}\left(\bigtimes_{i=1}^{\infty} \mathcal{A}_i\right). \quad \square$$

Next, if $\bigtimes_{i=1}^2 (\Omega_i, \mathcal{A}_i)$ is a product-measurable space, define for $A \subset \Omega_1 \times \Omega_2$

$$\begin{aligned} A^{(1)}(\omega_1) &= \{\omega_2 : (\omega_1, \omega_2) \in A\} \quad \text{for } \omega_1 \in \Omega_1, \\ A^{(2)}(\omega_2) &= \{\omega_1 : (\omega_1, \omega_2) \in A\} \quad \text{for } \omega_2 \in \Omega_2. \end{aligned}$$

The sets $A^{(1)}(\omega_1)$ and $A^{(2)}(\omega_2)$ are called **sections** of A (at ω_1 and ω_2 respectively).

Theorem 2. Let $\bigtimes_{i=1}^2(\Omega_i, \mathcal{A}_i)$ be a product measurable space.

- i. For $A \in \mathcal{A}_1 \times \mathcal{A}_2$ and $\omega_j \in \Omega_j$, the sections $A^{(j)}(\omega_j) \in \mathcal{A}_{3-j}$, $j = 1, 2$.
- ii. If T is a measurable transformation from $\bigtimes_{i=1}^2(\Omega_i, \mathcal{A}_i)$ to a measurable space (Ω, \mathcal{A}) , then for every $\omega_2 \in \Omega_2$, $T(\omega_1, \omega_2)$ defines a measurable transformation from $(\Omega_1, \mathcal{A}_1)$ to (Ω, \mathcal{A}) .

PROOF. Let

$$\begin{aligned}\mathcal{G} &= \{A : A \in \mathcal{A}_1 \times \mathcal{A}_2, A^{(1)}(\omega_1) \in \mathcal{A}_2 \text{ for each } \omega_1 \in \Omega_1\}, \\ \mathcal{D} &= \{D : D = D_1 \times D_2 \text{ where } D_i \in \mathcal{A}_i, i = 1, 2\}.\end{aligned}$$

Then $\mathcal{A}_1 \times \mathcal{A}_2 = \sigma(\mathcal{D})$ by definition. For $D = D_1 \times D_2 \in \mathcal{D}$, $D^{(1)}(\omega_1) = D_2$ or \emptyset according as $\omega_1 \in D_1$ or not. Hence $\mathcal{G} \supset \mathcal{D}$. It is easy to verify that the section of a union (resp. difference) is the union (resp. difference) of the sections, i.e., for $\bigcup_1^\infty A_n \subset \Omega_1 \times \Omega_2$ and all $\omega_1 \in \Omega_1$

$$(A_1 - A_2)^{(1)}(\omega_1) = A_1^{(1)}(\omega_1) - A_2^{(1)}(\omega_1), \left(\bigcup_1^\infty A_n\right)^{(1)}(\omega_1) = \bigcup_1^\infty A_n^{(1)}(\omega_1).$$

Therefore, \mathcal{G} is a σ -algebra and so $\mathcal{G} \supset \sigma(\mathcal{D}) = \mathcal{A}_1 \times \mathcal{A}_2$, proving (i) for $j = 1$. Similarly, for $A^{(2)}(\omega_2)$.

To prove (ii), let $B \in \mathcal{A}$. Then $T^{-1}(B) \in \mathcal{A}_1 \times \mathcal{A}_2$ and for every $\omega_2 \in \Omega_2$, by (i) $\{\omega_1 : T(\omega_1, \omega_2) \in B\} = (T^{-1}(B))^{(2)}(\omega_2) \in \mathcal{A}_1$. Therefore, $T(\omega_1, \omega_2)$ is measurable from $(\Omega_1, \mathcal{A}_1)$ to (Ω, \mathcal{A}) for every $\omega_2 \in \Omega_2$. \square

Since outcomes of interest in a chance experiment can usually be quantized, numerical-valued functions play an important role in probability theory. A measurable transformation X from a measurable space (Ω, \mathcal{A}) to the Borel line (R, \mathcal{B}) is called a **(real) measurable function on (Ω, \mathcal{A})** . Since \mathcal{B} is the σ -algebra generated by the class of intervals $[-\infty, x]$, $-\infty < x < \infty$, or by $(x, \infty]$, $-\infty < x < \infty$, it follows from Lemma 1 that a real-valued function X on Ω is measurable iff for every finite x , $\{X(\omega) < x\} \in \mathcal{A}$ or

$$\{X(\omega) > x\} \in \mathcal{A}, x \in (-\infty, \infty).$$

In the special case of a measurable transformation f from the Borel line (R, \mathcal{B}) to itself, f is termed a **Borel** or **Borel-measurable function**. Since every open set in $(-\infty, \infty)$ is a countable union of open intervals and every open interval is a Borel set, every continuous function on $(-\infty, \infty)$ is a Borel function. Hence a real function on $(-\infty, \infty)$ with range $(-\infty, \infty)$ which is continuous on $(-\infty, \infty)$ is always a Borel function.

Similarly, for any finite or infinite interval $J \subset R$ measurable transformations from $(J, \mathcal{B} \cdot J)$ to (R, \mathcal{B}) are termed Borel functions or **Borel functions on J** . Analogously, for any finite or infinite rectangle $J^n \subset R^n$, measurable functions from $(J^n, \mathcal{B}^n \cdot J^n)$ to (R, \mathcal{B}) are called **Borel functions of n variables** or Borel functions on J^n . Since every open set of $(-\infty, \infty)^n$ is a countable union of open rectangles, continuous functions on $(-\infty, \infty)^n$ are likewise Borel functions.

Let (Ω, \mathcal{A}) be a measurable space and X, Y real functions on Ω . A **complex-valued function** $Z = X + iY$ on Ω is called **\mathcal{A} -measurable** or simply measurable whenever both X and Y are \mathcal{A} -measurable.

If $\{X_n, n \geq 1\}$ are real measurable functions on (Ω, \mathcal{A}) , the set equality $\{\omega : \sup_{n \geq 1} X_n(\omega) > x\} = \bigcup_{n=1}^{\infty} \{\omega : X_n(\omega) > x\}$ reveals that $\sup_{n \geq 1} X_n$ is a measurable function. Analogously, $\inf_{n \geq 1} X_n$, $\overline{\lim}_{n \rightarrow \infty} X_n = \inf_{k \geq 1} \sup_{n \geq k} X_n$, and $\underline{\lim} X_n$ are all measurable. In particular, $\max(X_1, X_2)$ and $\min(X_1, X_2)$ are measurable. Since an identically constant function is trivially measurable, it follows that if X is a real measurable function, so are its **positive** and **negative parts** defined by

$$X^+ = \max(0, X) \quad \text{and} \quad X^- = \max(0, -X).$$

Measurability of the sum (if defined) of two measurable functions is a simple consequence of the relation (Exercise 1.2.9)

$$\{\omega : X_1(\omega) + X_2(\omega) < x\} = \bigcup_{r \text{ rational}} \{\omega : X_1(\omega) < r\} \cdot \{\omega : r < x - X_2(\omega)\}.$$

Thus, $X = X^+ - X^-$ is measurable iff X^+ and X^- are measurable.

Clearly, if n is a positive integer and c is any constant, measurability of X implies that of X^n and cX . Measurability of the product of two measurable functions is then a simple consequence of the identity

$$X_1 X_2 = \frac{1}{4}[(X_1 + X_2)^2 - (X_1 - X_2)^2].$$

Likewise, the ratio of two measurable functions (if defined) is measurable.

If $\{X_n, n \geq 1\}$ are real measurable functions, measurability of $\underline{\lim}_{n \rightarrow \infty} X_n$ and $\overline{\lim}_{n \rightarrow \infty} X_n$ and Exercise 1.4.5 ensure that $\lim_{n \rightarrow \infty} X_n$ (if it exists) is measurable.

As an aid in the discussion of integration in later chapters, it is helpful to define classes of functions with properties mirroring those of a λ -class of sets.

Definition. A family \mathcal{H} of nonnegative functions is called a **λ -system** if

- i. $1 \in \mathcal{H}$,
- ii. $X_n \in \mathcal{H}, n \geq 1, X_n \uparrow X \Rightarrow X \in \mathcal{H}$,
- iii. $X_i \in \mathcal{H}, c_i$ real, finite, $i = 1, 2$, and $c_1 X_1 + c_2 X_2 \geq 0 \Rightarrow c_1 X_1 + c_2 X_2 \in \mathcal{H}$.

Definition. A nonempty family \mathcal{M} of nonnegative functions is called a **monotone system** if the following conditions are satisfied:

- i. If $X_i \in \mathcal{M}$ and c_i is a finite nonnegative number, $i = 1, 2$, then $c_1 X_1 + c_2 X_2 \in \mathcal{M}$,
- ii. if $X_n \in \mathcal{M}$ and $X_n \uparrow X$, then $X \in \mathcal{M}$.

Note that a λ -system is always a monotone system but not conversely. A connection between classes of functions and classes of sets beyond that furnished by indicator functions is given next. In the course of the proof, an

arbitrary nonnegative measurable function is shown to be an increasing limit of simple (see Exercise 1.4.3) nonnegative measurable functions, a phenomenon that will reappear frequently in Chapters 4 and 6.

Theorem 3. *Let \mathcal{H} be a family of nonnegative functions on Ω which contains all indicators of sets of some class \mathcal{D} of subsets of Ω . If either (i) \mathcal{D} is a π -class and \mathcal{H} is a λ -system, or (ii) \mathcal{D} is a σ -algebra and \mathcal{H} is a monotone system, then \mathcal{H} contains all nonnegative $\sigma(\mathcal{D})$ -measurable functions.*

PROOF. Set $\mathcal{G} = \{A : I_A \in \mathcal{H}\}$. (i) By hypothesis $\mathcal{G} \supset \mathcal{D}$ and, according to (i), (ii), (iii) of the definition of a λ -system, \mathcal{G} is a λ -class, whence $\mathcal{G} \supset \sigma(\mathcal{D})$ by Theorem 1.3.2. If X is a nonnegative $\sigma(\mathcal{D})$ -measurable function, then $J_{k,n} = \{k/2^n \leq X < (k+1)/2^n\} \in \mathcal{G}$, $0 \leq k \leq 4^n - 1$, as does $J_{4^n,n} = \{X \geq 2^n\}$. Hence, if

$$X_n = \sum_{k=1}^{4^n} \frac{k}{2^n} I_{J_{k,n}},$$

$X_n \in \mathcal{H}$ by (iii). Clearly $X_n \uparrow X$ and so $X \in \mathcal{H}$ by (ii). In case (ii), where \mathcal{H} is a monotone system and \mathcal{D} is a σ -algebra, $\mathcal{G} \supset \mathcal{D} = \sigma(\mathcal{D})$ and the rest follows as before. \square

If X is a real \mathcal{A} -measurable function so that $X^{-1}(\mathcal{B}) \subset \mathcal{A}$, it is natural to speak of $X^{-1}(\mathcal{B})$ as the σ -algebra generated by X and hence to denote it by $\sigma(X)$. Then X is also a measurable function from $(\Omega, \sigma(X))$ to (R, \mathcal{B}) and indeed $\sigma(X)$ is the smallest σ -algebra of subsets of Ω for which that statement is true. Moreover, if g is a Borel function on R , then $g(X)$ is a real \mathcal{A} -measurable function and $\sigma(g(X)) \subset \sigma(X)$.

Definition. The σ -algebra $\sigma(X_\lambda, \lambda \in \Lambda)$ generated by a nonempty family $\{X_\lambda, \lambda \in \Lambda\}$ of real measurable functions on (Ω, \mathcal{A}) is defined to be the σ -algebra generated by the class of sets $\{\omega : X_\lambda(\omega) \in B\}$, $B \in \mathcal{B}$, $\lambda \in \Lambda$, or equivalently by $\{X_\lambda < x\}$, $x \in (-\infty, \infty)$, $\lambda \in \Lambda$.

Clearly, $\sigma(X_\lambda, \lambda \in \Lambda) \subset \mathcal{A}$ and $\sigma(X_\lambda, \lambda \in \Lambda)$ is the smallest σ -algebra relative to which all X_λ , $\lambda \in \Lambda$, are measurable.

In particular, for $\Lambda = \{1, 2, \dots, n\}$ the σ -algebra generated by X_1, \dots, X_n , namely $\sigma(X_1, \dots, X_n)$, is the σ -algebra generated by the class of sets $\{X_i \in B_i\}$, $B_i \in \mathcal{B}$, $1 \leq i \leq n$, or equivalently by the sets of $\sigma(X_i)$, $1 \leq i \leq n$. Thus, $\sigma(X_1, \dots, X_n) = \sigma(\bigcup_1^n \sigma(X_i))$.

When $\Lambda = \{1, 2, \dots\}$, it follows analogously that $\sigma(X_1, X_2, \dots) = \sigma(\bigcup_1^\infty \sigma(X_i))$. Consequently, for $n \geq 2$

$$\sigma(X_1) \subset \sigma(X_1, X_2) \subset \dots \subset \sigma(X_1, \dots, X_n),$$

$$\sigma(X_1, X_2, \dots) \supset \sigma(X_2, X_3, \dots) \supset \dots \supset \sigma(X_n, X_{n+1}, \dots).$$

The σ -algebra $\bigcap_{n=1}^\infty \sigma(X_n, X_{n+1}, \dots)$, which is called the tail σ -algebra of $\{X_n, n \geq 1\}$, will be encountered in Chapter 3.

If X_i is a measurable function from (Ω, \mathcal{A}) to (R, \mathcal{B}) , $1 \leq i \leq n$, then X_i is likewise a (real) measurable function from $(\Omega, \sigma(X_i))$ to (R, \mathcal{B}) , $1 \leq i \leq n$, and, moreover, each X_i is a measurable function from $(\Omega, \sigma(X_1 \dots X_n))$ to (R, \mathcal{B}) , $1 \leq i \leq n$. The next theorem characterizes functions which are $\sigma(X_1, \dots, X_n)$ -measurable.

Theorem 4. Let X_1, \dots, X_n be real measurable functions on (Ω, \mathcal{A}) . A (real) function on Ω is $\sigma(X_1, \dots, X_n)$ -measurable iff it has the form $f(X_1, \dots, X_n)$, where f is a Borel function on R^n .

PROOF. Let \mathcal{G} denote the class of Borel functions on R^n and define

$$\begin{aligned}\mathcal{H} &= \{f(X_1, \dots, X_n); f \in \mathcal{G}, f(X_1, \dots, X_n) \geq 0\}, \\ \mathcal{D} &= \{\{(\omega; X_i(\omega) \leq x_i, 1 \leq i \leq n); -\infty < x_i < \infty, 1 \leq i \leq n\}\}\end{aligned}$$

Then \mathcal{D} is a π -class of subsets of Ω , $\sigma(\mathcal{D}) = \sigma(X_1, \dots, X_n)$, and $I_D \in \mathcal{H}$ for $D \in \mathcal{D}$. Moreover, \mathcal{H} is a λ -system. In fact, $1 \in \mathcal{H}$, and if $f_i(X_1, \dots, X_n) \in \mathcal{H}$, c_i is a finite real number for $i = 1, 2$, and

$$c_1 f_1(X_1, \dots, X_n) + c_2 f_2(X_1, \dots, X_n) \geq 0,$$

then $c_1 f_1(X_1, \dots, X_n) + c_2 f_2(X_1, \dots, X_n) = f(X_1, \dots, X_n) \in \mathcal{H}$ where

$$f = (c_1 f_1 + c_2 f_2) I_{\{c_1 f_1 + c_2 f_2 \neq \pm \infty\}}.$$

Now let $f_m(X_1, \dots, X_n) \in \mathcal{H}$, $m \geq 1$, and $f_m(X_1, \dots, X_n) \uparrow f(X_1, \dots, X_n)$. Set $g_m = \max(f_1, \dots, f_m)$. Then $g_m \uparrow g \in \mathcal{G}$ and $0 \leq f_m(X_1, \dots, X_n) = g_m(X_1, \dots, X_n) \leq f_{m+1}(X_1, \dots, X_n)$. Therefore,

$$f(X_1, \dots, X_n) = g(X_1, \dots, X_n) \in \mathcal{H}$$

and \mathcal{H} is a λ -system. By Theorem 3, \mathcal{H} contains all nonnegative $\sigma(\mathcal{D})$ -measurable functions. Hence, if Y is $\sigma(X_1, \dots, X_n)$ -measurable, both Y^+ and Y^- are in \mathcal{H} so that $Y = Y^+ - Y^- = f(X_1, \dots, X_n)$ for some Borel function f on R^n .

Conversely, let f be a Borel function on R^n and

$$X(\omega) = (X_1(\omega), \dots, X_n(\omega)), \quad \omega \in \Omega.$$

Since X_i is measurable from $(\Omega, \sigma(X_1, \dots, X_n))$ to (R, \mathcal{B}) for $1 \leq i \leq n$, X is measurable from $(\Omega, \sigma(X_1, X_2, \dots, X_n))$ to (R^n, \mathcal{B}^n) by Theorem 1. By hypothesis, f is measurable from (R^n, \mathcal{B}^n) to (R, \mathcal{B}) , whence $f(X_1, \dots, X_n) = f(X)$ is measurable from $(\Omega, \sigma(X_1, \dots, X_n))$ to (R, \mathcal{B}) , that is, $f(X_1, \dots, X_n)$ is $\sigma(X_1, \dots, X_n)$ -measurable. \square

In similar fashion one can prove

Theorem 5. If $\{X_\lambda, \lambda \in \Lambda\}$ is an infinite family of real measurable functions on (Ω, \mathcal{A}) , then a real measurable function on Ω is $\sigma(X_\lambda, \lambda \in \Lambda)$ -measurable iff it is of the form $f(X_{\lambda_1}, X_{\lambda_2}, \dots, X_{\lambda_n}, \dots)$, where $\lambda_i \in \Lambda$, $i = 1, 2, \dots$, and f is an infinite-dimensional Borel function.

EXERCISES 1.4

1. Let $T: \Omega_1 \rightarrow \Omega_2$ and $A_\lambda \subset \Omega_1$ for $\lambda \in \Lambda$. Show that $T(\bigcup_\lambda A_\lambda) = \bigcup_\lambda T(A_\lambda)$ but that $T(A_\lambda - A_{\lambda'})$ need not equal $T(A_\lambda) - T(A_{\lambda'})$, where $T(A) = \{T(\omega_1): \omega_1 \in A\}$, $T(\mathcal{A}) = \{T(A): A \in \mathcal{A}\}$. If T is one-to-one and onto, then \mathcal{A} a σ -algebra on Ω_1 entails $T(\mathcal{A})$ is a σ -algebra on Ω_2 .
2. In the notation of Exercise 1, prove that $A \subset T^{-1}(T(A))$ with set equality holding if T is one-to-one. Also, $T(T^{-1}(B)) \subset B \subset \Omega_2$ with equality if T is onto.
3. A function X on (Ω, \mathcal{F}, P) is called **simple** if for some finite integer n and real numbers x_1, \dots, x_n it is representable as $X = \sum_{i=1}^n x_i I_{A_i}$ for $\{A_i, 1 \leq i \leq n\}$ a disjoint subclass of \mathcal{F} . Prove that any non-negative measurable function is an increasing limit of non-negative simple functions.
4. Prove that $(X + Y)^+ \leq X^+ + Y^+$, $(X + Y)^- \leq X^- + Y^-$, $X^+ \leq (X + Y)^+ + Y^-$. Also if $B_n \uparrow B$, then $X^{-1}(B_n) \uparrow X^{-1}(B)$.
5. Let X and Y be real measurable functions on (Ω, \mathcal{A}) and c any real constant. Show that $\{\omega: X(\omega) < Y(\omega) + c\}$, $\{\omega: X(\omega) \leq Y(\omega) + c\}$, $\{\omega: X(\omega) = Y(\omega)\} \in \mathcal{A}$.
6. Prove that if X is a real measurable function on (Ω, \mathcal{A}) , so is $|X|$. Is the converse true?
7. If X and Y are real functions on (Ω, \mathcal{A}) then $X + iY$ is measurable iff (X, Y) is measurable from (Ω, \mathcal{A}) to (R^2, \mathcal{B}^2) .
8. If $(\Omega_i, \mathcal{A}_i)$, $i = 1, 2$ are measurable spaces and $X: \Omega_1 \rightarrow \Omega_2$ then X is a measurable transformation from $(\Omega_1, \mathcal{A}_1)$ to $(\Omega_2, \mathcal{A}_2)$ provided $X^{-1}(\mathcal{D}) \subset \mathcal{A}_1$ for some class \mathcal{D} of sets for which $\sigma(\mathcal{D}) = \mathcal{A}_2$.
9. Prove that any real monotone function on R is Borel measurable and has at most a countable number of discontinuities.
10. For any linear Borel set B and r.v. X prove that $\sigma(XI_{[X \in B]}) = X^{-1}(B) \cap \sigma(X)$.

1.5 Additive Set Functions, Measures, and Probability Spaces

Let Ω be a space and \mathcal{A} be a nonempty class of subsets of Ω . A **set function** μ on \mathcal{A} is a real-valued function defined on \mathcal{A} . If $\mu(A)$ is finite for each $A \in \mathcal{A}$, μ is said to be **finite**, and if there exists $\{A_n, n \geq 1\} \subset \mathcal{A}$ such that $\mu(A_n)$ is finite for each n and $\bigcup_{n=1}^\infty A_n = \Omega$, μ is said to be **σ -finite** on \mathcal{A} .

If $A \subset \Omega$ and $A = \bigcup_{n=1}^m A_n$, where $\{A_n, n = 1, \dots, m\}$ is a disjoint subclass of \mathcal{A} , the latter subclass is called a **finite partition of A in \mathcal{A}** . If $\{A_n, n = 1, 2, \dots\}$ is a disjoint subclass of \mathcal{A} and $\bigcup_{n=1}^\infty A_n = A$, it is called a **σ -partition of A in \mathcal{A}** .

Definition. A set function μ on \mathcal{A} , denoted $\mu(A)$ or $\mu\{A\}$ for $A \in \mathcal{A}$ is **additive** (or more precisely, **finitely additive**), if for every $A \in \mathcal{A}$ and every finite partition $\{A_n, n = 1, \dots, m\}$ of A in \mathcal{A} , $\sum_1^m \mu(A_n)$ is defined and

$$\mu(A) = \sum_1^m \mu(A_n);$$

moreover, μ is **σ -additive** or **countably additive** or **completely additive** if for every $A \in \mathcal{A}$ and every σ -partition $\{A_n, n = 1, 2, \dots\}$ of A in \mathcal{A} , $\sum_1^\infty \mu(A_n)$ is defined and

$$\mu(A) = \sum_1^\infty \mu(A_n).$$

Note in the σ -additive case that the definition precludes conditional convergence of the series. Clearly, if $\emptyset \in \mathcal{A}$, countable additivity implies finite additivity.

If an additive set function is finite on some set A of an algebra \mathcal{A} , it is necessarily finite on all $B \in \mathcal{A}$ with $B \subset A$. Examples of set functions that are additive but not σ -additive appear in Exercises 1.5.6 and 1.5.7.

Definition. A nonnegative σ -additive set function μ on a class \mathcal{A} containing \emptyset with $\mu(\emptyset) = 0$ is called a **measure**. If μ is a measure on a σ -algebra \mathcal{F} of subsets of Ω , the triplet $(\Omega, \mathcal{F}, \mu)$ is called a **measure space**. A measure space (Ω, \mathcal{F}, P) is a **probability space** if $P(\Omega) = 1$.

From a purely formal vantage point the prior definition relegates probability theory to a special case of measure theory. However, such important notions as independence (Chapter 3) and other links between probability and the real world have nurtured probability theory and given it a life and direction of its own.

In addition to the basic property of σ -additivity, a measure μ on an algebra \mathcal{A} is **monotone**, that is, $\mu(A_1) \leq \mu(A_2)$ whenever $A_1 \subset A_2, A_i \in \mathcal{A}, i = 1, 2$ (Exercise 1), and, moreover, **subadditive**, that is, if $\bigcup_1^\infty A_j \in \mathcal{A}, A_j \in \mathcal{A}, j \geq 1$, then

$$\mu\left\{\bigcup_1^\infty A_j\right\} \leq \sum_{j=1}^\infty \mu(A_j), \quad (1)$$

as noted in (iii) of the forthcoming Theorem 2. A prominent example of a measure is ordinary "length" on the class \mathcal{A} of all finite half-open intervals $[a, b)$, i.e., $\mu([a, b)) = b - a$. The extension of the definition of length to $\sigma(\mathcal{A})$ (= class of all Borel sets) is known as Lebesgue measure and will be encountered in Chapter 6.

Probability spaces will underlie most if not all of this book. In a probability space (Ω, \mathcal{F}, P) , the sets $A \in \mathcal{F}$ are called **events** and the nonnegative, real number $P(A)$ is referred to as the **probability** of the event A . The monotone property of a measure ensures that for every event A

$$0 = P(\emptyset) \leq P(A) \leq P(\Omega) = 1. \quad (2)$$

Thus, in probability theory "event" is simply a name attached to an element of a σ -algebra of subsets of a basic reference set Ω , frequently called the **sample space**. From an intuitive or applicational standpoint an event is the

abstract counterpart of an observable outcome of a (not completely determined) physical experiment. The numerical-valued probability of an event is in some sense an idealization of the intuitive notion of a long-run frequency as attested to by (2) and the additivity property.

Events of probability zero are called **null events**. A real-valued measurable function X on a probability space (Ω, \mathcal{F}, P) is called a **random variable** (abbreviated r.v.) if $\{\omega : |X(\omega)| = \infty\}$ is a null event. If some property obtains except on a null event, this property is said to obtain **almost surely** (abbreviated a.s.) or **almost certainly** (abbreviated a.c.) or with probability one. Hence, a r.v. on a probability space (Ω, \mathcal{F}, P) is just an a.c. finite \mathcal{F} -measurable function on Ω .

It is an extremely useful state of affairs that a probability or more generally a measure defined on an algebra \mathcal{A} or even a semi-algebra (definition forthcoming) may be uniquely extended to $\sigma(\mathcal{A})$. The proof will be deferred to Chapter 6 but a first step in this direction is Theorem 1.

Definition. A **semi-algebra** \mathcal{S} is a π -class of subsets of Ω such that $\Omega \in \mathcal{S}$, $\emptyset \in \mathcal{S}$ and for each $A \in \mathcal{S}$ there is a finite partition of A^c in \mathcal{S} .

Let \mathcal{S} be a semi-algebra. It follows easily that

- i. the class \mathcal{G} of all finite unions of disjoint sets of \mathcal{S} is the algebra generated by \mathcal{S} (clearly, $\mathcal{G} \subset \mathcal{A}(\mathcal{S})$ and \mathcal{G} is an algebra containing \mathcal{S} and hence $\mathcal{A}(\mathcal{S})$);
- ii. for $\{A_n, n = 1, \dots, m\} \subset \mathcal{S}$, there is a finite partition of $A_1 A_2^c \dots A_m^c$ in \mathcal{S} ;
- iii. for each $A \in \mathcal{S}$ and each countable class $\{A_n, n = 1, 2, \dots\} \subset \mathcal{S}$ such that $\bigcup_1^\infty A_n \supset A$, there is a σ -partition $\{B_n, n = 1, 2, \dots\}$ of A in \mathcal{S} for which each B_n is a subset of some A_m (write $A = \bigcup_n A_n$ as a disjoint union and utilize (ii)).

Let \mathcal{G} and \mathcal{H} be two classes of subsets of Ω with $\mathcal{G} \subset \mathcal{H}$. If μ and v are set functions defined on \mathcal{G} and \mathcal{H} respectively such that $\mu(A) = v(A)$ for $A \in \mathcal{G}$, v is said to be an **extension of μ to \mathcal{H}** , and μ the **restriction of v to \mathcal{G}** , denoted by $\mu = v|_{\mathcal{G}}$.

Theorem 1. If μ is a nonnegative additive set function on a semi-algebra \mathcal{S} , then there exists a unique extension v of μ to the algebra \mathcal{A} generated by \mathcal{S} such that v is additive. Moreover, if μ is σ -additive on \mathcal{S} , so is v on \mathcal{A} .

PROOF. Since \mathcal{A} is the class of all finite unions of disjoint sets in \mathcal{S} , every $A \in \mathcal{A}$ has a finite partition $\{A_n, 1 \leq n \leq m\}$ in \mathcal{S} . For such an A , define

$$v(A) = \sum_1^m \mu(A_n).$$

Then v is consistently defined on \mathcal{A} , since if $A \in \mathcal{A}$ has two distinct finite partitions $\{A_n, 1 \leq n \leq m\}$ and $\{B_j, 1 \leq j \leq k\}$ in \mathcal{S} ,

$$\begin{aligned} v(A) &= \sum_{n=1}^m \mu(A_n) = \sum_{n=1}^m \mu(AB_n) = \sum_{n=1}^m \sum_{j=1}^k \mu(A_n B_j) \\ &= \sum_{j=1}^k \sum_{n=1}^m \mu(A_n B_j) = \sum_{j=1}^k \mu(AB_j) = \sum_{j=1}^k \mu(B_j). \end{aligned}$$

It is easy to see that v is additive on \mathcal{A} . The uniqueness of v follows from the fact that if v^* is additive on \mathcal{A} and $v^*|_{\mathcal{S}} = \mu$, then for any finite partition $\{A_n, 1 \leq n \leq m\}$ in \mathcal{S} of $A \in \mathcal{A}$,

$$v^*(A) = \sum_{n=1}^m v^*(A_n) = \sum_{n=1}^m \mu(A_n) = v(A).$$

Suppose next that μ is σ -additive on \mathcal{S} , $\{A_n, n \geq 1\}$ is a σ -partition in \mathcal{A} of $A \in \mathcal{A}$ and $\{C_n, 1 \leq n \leq m\}$ is a partition of A in \mathcal{S} . For each n , let $\{B_j, j_{n-1} < j \leq j_n\}$ be a finite partition of A_n in \mathcal{S} , where $j_0 = 0$. Then $\{B_j, j \geq 1\}$ is a σ -partition of A in \mathcal{S} and

$$\begin{aligned} \sum_1^\infty v(A_n) &= \sum_{n=1}^\infty \sum_{j=j_{n-1}+1}^{j_n} \mu(B_j) = \sum_1^\infty \mu(B_j) = \sum_1^\infty \mu(AB_j) \\ &= \sum_{j=1}^\infty \sum_{n=1}^m \mu(C_n B_j) = \sum_{n=1}^m \sum_{j=1}^\infty \mu(C_n B_j) \\ &= \sum_1^m \mu(C_n) = v(A). \end{aligned} \quad \square$$

The subadditive property of a measure and alternative characterizations of countable additivity appear in

Theorem 2. *Let μ be a nonnegative additive set function on an algebra \mathcal{A} and $\{A_n, n \geq 0\} \subset \mathcal{A}$.*

- i. *if $\{A_n, n \geq 1\}$ is a disjoint class and $\bigcup_1^\infty A_n \subset A_0$, then*

$$\mu(A_0) \geq \sum_1^\infty \mu(A_n).$$

- ii. *if $A_0 \subset \bigcup_1^m A_n$ for some $m = 1, 2, \dots$, then*

$$\mu(A_0) \leq \sum_1^m \mu(A_n).$$

- iii. *μ is σ -additive iff for every increasing sequence $\{A_n, n \geq 1\} \subset \mathcal{A}$ with $\lim_n A_n = A \in \mathcal{A}$,*

$$\lim_n \mu(A_n) = \mu(A). \quad (3)$$

In this case μ is subadditive on \mathcal{A} , i.e., if $\bigcup_1^\infty A_j \in \mathcal{A}$, $A_j \in \mathcal{A}$, $j \geq 1$, then

$$\mu\left(\bigcup_1^\infty A_j\right) \leq \sum_1^\infty \mu(A_j).$$

iv. if μ is σ -additive, then for every decreasing sequence $A_n \in \mathcal{A}$ with $\mu(A_1) < \infty$ and $\lim_n A_n = A \in \mathcal{A}$,

$$\lim_n \mu(A_n) = \mu(A). \quad (4)$$

Conversely, if for every decreasing sequence $A_n \in \mathcal{A}$ with $\lim_n A_n = \emptyset$,

$$\lim_n \mu(A_n) = 0, \quad (5)$$

then μ is σ -additive.

PROOF. (i) Since (i) is obvious if $\mu(A_0) = \infty$, let $\mu(A_0) < \infty$. Now $A_0 - \bigcup_1^m A_n \in \mathcal{A}$ ($m = 1, 2, \dots$), whence by subtractivity (see Exercise 1.5.1)

$$\begin{aligned} 0 \leq \mu\left(A_0 - \bigcup_1^m A_n\right) &= \mu(A_0) - \mu\left(\bigcup_1^m A_n\right) = \mu(A_0) - \sum_1^m \mu(A_n), \\ \mu(A_0) &\geq \sum_1^m \mu(A_n) \xrightarrow{m \rightarrow \infty} \sum_1^\infty \mu(A_n). \end{aligned}$$

(ii) Since $A_0 = \bigcup_1^m A_0 A_n = A_0 A_1 \cup A_0 A_2 A_1^c \cup \dots \cup A_0 A_m A_{m-1}^c \dots A_1^c$, by additivity and monotonicity of μ

$$\begin{aligned} \mu(A_0) &= \mu\left(\bigcup_1^m A_0 A_n\right) = \mu(A_0 A_1) + \mu(A_0 A_2 A_1^c) + \dots \\ &\quad + \mu(A_0 A_m A_{m-1}^c \dots A_1^c) \\ &\leq \mu(A_1) + \mu(A_2) + \dots + \mu(A_m). \end{aligned}$$

(iii) If (3) holds for every increasing sequence $\{A_n, n \geq 1\} \subset \mathcal{A}$ with $\lim_n A_n = A \in \mathcal{A}$, then for any σ -partition $\{B_n, n \geq 1\}$ in \mathcal{A} of $B \in \mathcal{A}$,

$$\begin{aligned} \mu(B) &= \mu\left(\bigcup_1^\infty B_n\right) = \mu\left(\lim_m \bigcup_1^m B_n\right) = \lim_m \mu\left(\bigcup_1^m B_n\right) \\ &= \lim_m \sum_1^m \mu(B_n) = \sum_1^\infty \mu(B_n). \end{aligned}$$

Hence, μ is σ -additive. Conversely, let μ be σ -additive and $A_n, n \geq 1$, an increasing sequence of \mathcal{A} sets with $\lim_n A_n = A \in \mathcal{A}$. Then

$$\begin{aligned} \mu(A) &= \mu\left(\bigcup_1^\infty A_n\right) = \mu(A_1 \cup (A_2 - A_1) \cup (A_3 - A_2) \cup \dots) \\ &= \mu(A_1) + \mu(A_2 - A_1) + \mu(A_3 - A_2) + \dots \\ &= \lim_n (\mu(A_1) + \mu(A_2 - A_1) + \dots + \mu(A_n - A_{n-1})) \\ &= \lim_n \mu(A_n). \end{aligned}$$

The second part of (iii) follows from the first part and (ii).

(iv) If μ is σ -additive, and $A_n, n \geq 1$, is a decreasing sequence of \mathcal{A} sets with $\mu(A_1) < \infty$ and $\lim_n A_n = A \in \mathcal{A}$, then $A_1 - A_n$ is an increasing sequence in \mathcal{A} , whence by (iii)

$$\begin{aligned}\mu(A_1) - \mu(A) &= \mu(A_1 - A) = \mu\left(\lim_n (A_1 - A_n)\right) \\ &= \lim_n \mu(A_1 - A_n) = \mu(A_1) - \lim_n \mu(A_n).\end{aligned}$$

Since $\mu(A_1) < \infty$, (4) holds. Conversely, if $\{A_n\}$ is any σ -partition in \mathcal{A} of some set $A \in \mathcal{A}$, then $\bigcup_{j=n}^{\infty} A_j \in \mathcal{A}$, $n \geq 1$, and $\bigcup_{j=n}^{\infty} A_j \downarrow \overline{\lim} A_n = \emptyset$, whence by hypothesis $\lim \mu\{\bigcup_{j=n}^{\infty} A_j\} = 0$. By finite additivity,

$$\mu\{A\} = \sum_1^{n-1} \mu\{A_j\} + \mu\left\{\bigcup_{j=n}^{\infty} A_j\right\},$$

and so countable additivity follows upon letting $n \rightarrow \infty$. \square

Corollary 1. *A nonnegative additive set function μ on a semi-algebra \mathcal{S} satisfies (i) and (ii) on \mathcal{S} . Moreover, a measure μ on \mathcal{S} is subadditive on \mathcal{S} .*

PROOF. Extend μ to the algebra $\mathcal{A}(\mathcal{S})$ generated by \mathcal{S} via Theorem 1. By Theorem 2, the set function (resp. measure) μ satisfies (i) and (ii) (resp. σ -additivity, hence subadditivity) on $\mathcal{A}(\mathcal{S})$ and *a fortiori* on \mathcal{S} . \square

A finite measure space is a measure space $(\Omega, \mathcal{F}, \mu)$ with μ finite. In such a case, the finiteness proviso in Theorem 2 (iv) is superfluous.

Corollary 2. *If (Ω, \mathcal{F}, P) is a probability space, $P\{\lim_n A_n\} = \lim_n P\{A_n\}$ for every sequence of sets $A_n \in \mathcal{F}$ whose limit exists.*

Theorem 3. *Let (Ω, \mathcal{F}, P) be a probability space with $\mathcal{F} = \sigma(\mathcal{G})$, where \mathcal{G} is an algebra of subsets of Ω . Then for $A \in \mathcal{F}$ and all $\varepsilon > 0$, there exists a set $B_\varepsilon \in \mathcal{G}$ such that $P\{A \Delta B_\varepsilon\} < \varepsilon$ and so $|P\{A\} - P\{B_\varepsilon\}| < \varepsilon$.*

PROOF. Let $\mathcal{A} = \{A: A \in \mathcal{F}$ and for every $\varepsilon > 0$ there exists $B_\varepsilon \in \mathcal{G}$ with $P\{A \Delta B_\varepsilon\} < \varepsilon\}$. Then $\mathcal{A} \supset \mathcal{G}$, and moreover \mathcal{A} is a σ -algebra since, setting $B = B_\varepsilon$,

$$A \Delta B = A^c \Delta B^c \Rightarrow A^c \in \mathcal{A} \quad \text{if } A \in \mathcal{A},$$

$$\bigcup_1^{\infty} A_n \Delta \bigcup_1^{\infty} B_n \subset \bigcup_1^{\infty} (A_n \Delta B_n) \Rightarrow \bigcup_1^{\infty} A_n \in \mathcal{A} \quad \text{if } A_j \in \mathcal{A}, \quad j \geq 1,$$

(recall Exercise 1.2.1), noting that

$$\lim_{n \rightarrow \infty} P\left\{\bigcup_1^{\infty} A_j \Delta \bigcup_{j=1}^n B_j\right\} = P\left\{\bigcup_1^{\infty} A_j \Delta \bigcup_{j=1}^{\infty} B_j\right\}.$$

Hence $\mathcal{A} \supset \sigma(\mathcal{G}) = \mathcal{F}$. \square

EXERCISES 1.5

- Show that a finite measure μ on a class \mathcal{A} is (i) **subtractive**, i.e., $A_i \in \mathcal{A}$, $i = 1, 2$, $A_1 \subset A_2$ and $A_2 - A_1 \in \mathcal{A}$ implies $\mu(A_2 - A_1) = \mu(A_2) - \mu(A_1)$, and (ii) **monotone**, i.e., $\mu(A_1) \leq \mu(A_2)$ with A_i as in (i). In view of (i), if there is one set in \mathcal{A} with finite measure, the proviso $\mu(\emptyset) = 0$ is automatically satisfied.
- If $(\Omega, \mathcal{A}, \mu)$ is a measure space, prove that $\mu(\liminf A_n) \leq \liminf \mu(A_n)$ for $A_n \in \mathcal{A}$. Analogously, if $\mu(\bigcup_{i=n}^{\infty} A_i) < \infty$, some $n \geq 1$, then $\mu(\overline{\limsup} A_n) \geq \overline{\limsup} \mu(A_n)$.
- If $\Omega = R = [-\infty, \infty]$, then $\mathcal{S} = \{\{+\infty\}, R, [a, b], -\infty \leq a \leq b \leq \infty\}$ is a semi-algebra, not an algebra. Also, if $(\Omega_i, \mathcal{A}_i)$, $i = 1, 2$, are measurable spaces,

$$\{A_1 \times A_2 : A_i \in \mathcal{A}_i, i = 1, 2\}$$

is a semi-algebra. Also, if $n \geq 2$ and \mathcal{S}_i is a semi-algebra of subsets of Ω_i , $1 \leq i \leq n$, then $\mathcal{S} = \{\bigtimes_{i=1}^n S_i : S_i \in \mathcal{S}_i, 1 \leq i \leq n\}$ is a semi-algebra of subsets of $\bigtimes_{i=1}^n \Omega_i$.

- Prove that the class of all finite unions of disjoint sets of a semi-algebra \mathcal{S} is an algebra.
- If (Ω, \mathcal{A}, P) is a probability space and $\{A_n, n \geq 1\}$ is a sequence of events,

(i) verify that

$$P\left\{\varliminf_{n \rightarrow \infty} A_n\right\} = \lim_{n \rightarrow \infty} P\left\{\bigcap_{j=n}^{\infty} A_j\right\}, \quad P\left\{\varlimsup_{n \rightarrow \infty} A_n\right\} = \lim_{n \rightarrow \infty} P\left\{\bigcup_{j=n}^{\infty} A_j\right\}$$

(ii) show that $P(\bigcap_{n=1}^{\infty} A_n) = 1$ if $P(A_n) = 1$, $n \geq 1$.

- Let $\Omega = \{\text{positive integers}\}$, $\mathcal{A} = \{A : A \subset \Omega\}$. Define $\mu_1\{A\} = \text{number of elements in } A$ and $\mu_2\{A\} = 0$ or ∞ according as A is a finite or infinite set. Note that if $A_n = \{n\}$, then $\bigcup_{j=n}^{\infty} A_j \downarrow \overline{\lim} A_n = \emptyset$ and $\mu_1(\bigcup_{j=n}^{\infty} A_j) = \infty$. Since μ_1 is a bona fide measure on \mathcal{A} , called **counting measure**, the finite hypothesis in the first part of (iv) of Theorem 2 is indispensable. The set function μ_2 is additive but not σ -additive. Thus, if in the second part of (iv), condition (5), were stipulated only when $A_n \downarrow \emptyset$ and $\mu\{A_n\} < \infty$, some $n \geq 1$, then a finiteness requirement for μ would also be necessary.

- For $\Omega, \mathcal{A}, \mu_1$ as in Exercise 6, define $N_n(A) = \mu_1\{A[1, n]\}$ and

$$\mathcal{D} = \left\{A \in \mathcal{A} : \mu\{A\} = \lim_{n \rightarrow \infty} \frac{N_n(A)}{n} \text{ exists}\right\}.$$

Prove that \mathcal{D} is closed under complementation and finite disjoint unions but that nonetheless \mathcal{D} is not an algebra. Also $\mu\{A\}$, called the asymptotic density of A , is additive but not σ -additive. Hint: Let $B_k = \{\text{odd integers in } [2^{2k}, 2^{2k+1}]\}$ and $B'_k = \{\text{even integers in } [2^{2k-1}, 2^{2k}]\}$. If $B = \bigcup_{k=1}^{\infty} (B_k \cup B'_k)$ and $A = \{\text{odd integers of } \Omega\}$, then $A \in \mathcal{D}$, $B \in \mathcal{D}$, but $AB \notin \mathcal{D}$.

- Let f be a monotone increasing function on $\Omega = [0, 1]$ such that $0 \leq f \leq 1$, and \mathcal{G} the class of finite unions of disjoint intervals $[a, b] \subset \Omega$.

i. \mathcal{G} is an algebra.

- ii. Put $\mu(A) = \sum_{j=1}^n (f(b_j) - f(a_j))$ for $A = \bigcup_{j=1}^n [a_j, b_j] \in \mathcal{G}$, where $0 \leq a_1 < b_1 \leq a_2 < b_2 \leq \dots \leq a_n < b_n \leq 1$. Then μ is additive on \mathcal{G} .
- iii. If μ is σ -additive on \mathcal{G} , then f is **left continuous**, i.e., $f(t) = \lim_{0 < \delta \rightarrow 0} f(t - \delta)$ for every $t \in \Omega$.
- iv. If f is left continuous, then μ is σ -additive on \mathcal{G} . Hint: For part (iv), see the proof of Lemma 6.1.1.
9. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and \mathcal{G} a sub- σ -algebra of \mathcal{F} . Prove that

$$\mathcal{G}^* = \{G \Delta N : G \in \mathcal{G}, N \in \mathcal{F}, \mu\{N\} = 0\}$$

is a sub- σ -algebra of \mathcal{F} . Then \mathcal{G}^* is called the **completion** of \mathcal{G} . When $\mathcal{G} = \mathcal{F}$ then $\mathcal{G}^* = \mathcal{F}$ and so the completion of \mathcal{F} is rather defined by $\mathcal{F}^* = \{A \Delta M : A \in \mathcal{F}, M \subset N \in \mathcal{F}, \mu\{N\} = 0\}$. If μ^* on \mathcal{F}^* is defined by $\mu^*\{A \Delta M\} = \mu\{A\}$, check that $(\Omega, \mathcal{F}^*, \mu^*)$ is a measure space. It is called the completion of $(\Omega, \mathcal{F}, \mu)$.

10. (Cantor ternary set.) Let $I_1 = I_{1,1} = (\frac{1}{3}, \frac{2}{3})$ be the “open middle third” of $I = [0, 1]$. Analogously, for $j \geq 2$, let $I_{j,k}$, $k = 1, \dots, 2^{j-1}$, be the open middle third of each of the 2^{j-1} intervals of $I - \bigcup_{j=1}^{j-1} I_j$, where $I_j = \bigcup_{k=1}^{2^{j-1}} I_{j,k}$, $j \geq 1$. Show that the Cantor set $C = I - \bigcup_{j=1}^{\infty} I_j$ satisfies: (i) $x \in C$ iff the digit 1 is not needed in the triadic expansion $x = \sum_{i=1}^{\infty} x_i 3^{-i}$, where, in general, $x_i = 0, 1, 2$; (ii) C is uncountable; (iii) C is perfect (all its points are limit points); (iv) C is nowhere dense (the closure of C contains no nondegenerate interval); (v) if μ is a measure on the Borel sets of $[0, 1]$ whose restriction to intervals coincides with “length” (such a “Lebesgue measure” appears in Section 6.1), then $\mu\{C\} = 0$.

1.6 Induced Measures and Distribution Functions

Consider a measurable transformation X from a measure space $(\Omega, \mathcal{A}, \mu)$ to a measurable space $(\tilde{\Omega}, \tilde{\mathcal{A}})$, i.e., $X: \Omega \rightarrow \tilde{\Omega}$ and $X^{-1}(\tilde{\mathcal{A}}) \subset \mathcal{A}$. The measure μ on \mathcal{A} induces in a very natural fashion a measure $\tilde{\mu}$ on $\tilde{\mathcal{A}}$, referred to as the **induced measure**, denoted by μ_X and defined for all $\tilde{A} \in \tilde{\mathcal{A}}$ by

$$\mu_X\{\tilde{A}\} = \tilde{\mu}\{\tilde{A}\} = \mu\{X^{-1}(\tilde{A})\}.$$

It is readily verified that μ_X is a measure on $\tilde{\mathcal{A}}$ and consequently that $(\tilde{\Omega}, \tilde{\mathcal{A}}, \mu_X)$ is a measure space.

In particular, a random variable X on a probability space (Ω, \mathcal{F}, P) engenders a new (Borel) probability space (R, \mathcal{B}, P_X) , where \mathcal{B} is the class of Borel sets of $R = [-\infty, \infty]$. To focus attention on the latter, i.e., to consider X of primary interest, is tantamount to restricting the former to $(\Omega, \sigma(X), P)$. Associated with any random variable X on a probability space (Ω, \mathcal{F}, P) is a real function F_X called the **distribution function** of the random variable X and defined for all $x \in [-\infty, \infty]$ by

$$F_X(x) = P\{\omega: X(\omega) < x\} = P\{X^{-1}((-\infty, x])\} \quad (1)$$

Definition. A real-valued function $G(x)$ on $R = [-\infty, \infty]$ is called a **distribution function** (abbreviated d.f.) if

- i. G is nondecreasing,
- ii. G is left continuous, i.e.,

$$\lim_{\substack{y \rightarrow x \\ y < x}} G(y) = G(x), \quad \text{all } x \in R,$$

$$\text{iii. } G(-\infty) = \lim_{x \rightarrow -\infty} G(x) = 0, \quad G(\infty) = \lim_{x \rightarrow \infty} G(x) = 1.$$

It follows from Theorem 1.5.2 and the fact that a r.v. is finite a.s., that the “distribution function of any r.v.” is a distribution function in the sense of the prior definition. Conversely, for every distribution function G there exists a r.v. X on some probability space (Ω, \mathcal{F}, P) such that $F_X = G$. The proof for a special case is given in this section but a general treatment will be deferred to Section 6.3.

Analogously to the preceding, n random variables X_1, \dots, X_n on a probability space (Ω, \mathcal{F}, P) induce via the map $X = (X_1, \dots, X_n)$, the new (Borel) probability space $(R^n, \mathcal{B}^n, P_X)$, where \mathcal{B}^n is the class of n -dimensional Borel sets of R^n and P_X is the induced measure. Rather than shift from the basic probability space, one may restrict attention to $(\Omega, \sigma(X_1, \dots, X_n), P)$. Associated with the **random vector** X i.e., with the n -tuple of random variables (X_1, \dots, X_n) , is a real function on R^n called the (joint) distribution function of (X_1, \dots, X_n) , denoted by F_{X_1, \dots, X_n} and defined by

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P\{\omega: X_1 < x_1, \dots, X_n < x_n\}, \quad x = (x_1, \dots, x_n) \in R^n. \quad (2)$$

The joint d.f. F_{X_1, \dots, X_n} satisfies properties akin to (i), (ii), (iii) (see Section 6.3), but monotonicity in each argument is insufficient and the analogue of (i) involves a mixed difference. Nonetheless, if a d.f. on R^n is properly defined, it is again true that the “d.f. of (X_1, \dots, X_n) ” is a d.f. on R^n and that, conversely, given a d.f. $G(x_1, \dots, x_n)$ on R^n , there always exist random variables X_1, \dots, X_n on some probability space whose joint d.f. $F_{X_1, \dots, X_n} = G$.

In the same vein, a sequence $\{X_n, n \geq 1\}$ of r.v.s on a probability space (Ω, \mathcal{F}, P) induces a Borel probability space $(R^\infty, \mathcal{B}^\infty, P_X)$ via the map $X = (X_1, X_2, \dots)$ or alternatively places in relief $(\Omega, \sigma(X_n, n \geq 1), P)$. Note that the d.f.s of all finite n -tuples of r.v.s $(X_{i_1}, \dots, X_{i_n})$ are determined via

$$F_{X_{i_1}, \dots, X_{i_n}}(x_1, \dots, x_n) = P\{\omega: X_{i_1} < x_1, \dots, X_{i_n} < x_n\}.$$

It will be proved in Section 6.4 that if the d.f.s G_{i_1, \dots, i_n} are prescribed in a consistent manner for all choices of indices $1 \leq i_1 < \dots < i_n$ and all $n \geq 1$, there exists a sequence $\{X_n, n \geq 1\}$ of r.v.s on a probability space such that

$$F_{X_{i_1}, \dots, X_{i_n}} = G_{i_1, \dots, i_n}.$$

A d.f. G on R is called **discrete** if

$$G(x) = \sum_{j: x_j < x} p_j, \quad x \in R, \quad (3)$$

where $p_j > 0$ for all j , $\sum_{\text{all } j} p_j = 1$, and $S = \{x_j: 1 \leq j < n + 1 \leq \infty\}$ is a subset of $(-\infty, \infty)$. The associated function

$$f(x) \begin{cases} = p_j & \text{for } x = x_j, \\ = 0 & \text{for } x \neq x_j \end{cases} \quad 1 \leq j < n + 1 \leq \infty \quad (4)$$

is termed a **probability density function** (abbreviated p.d.f.) on

$$S = \{x_j: 1 \leq j < n + 1 \leq \infty\}.$$

Clearly, a probability density function is completely determined by $\{x_j, p_j, 1 \leq j < n + 1 \leq \infty\}$. Typically, S is the set of positive or nonnegative integers or some finite subset thereof. This will be the case with the binomial and Poisson d.f.s occurring in Chapter 2.

To construct a probability space (Ω, \mathcal{F}, P) and a r.v. X on it whose d.f. F_X is equal to a preassigned discrete d.f. G say (as in (3)), it suffices to choose $\Omega = S$, \mathcal{F} = class of all subsets of Ω , $P\{A\} = \sum_{j: x_j \in A} p_j$, and $X(\omega) = \omega$. Note that then $P\{\omega: X(\omega) = x_j\} = p_j$, $1 \leq j < n + 1 \leq \infty$, where $\sum_1^n p_j = 1$.

A d.f. G is called **absolutely continuous** if there exists a Borel function g on $R = [-\infty, \infty]$ such that

$$G(x) = \int_{-\infty}^x g(t)dt, \quad x \in R. \quad (5)$$

The associated function g is termed a **density function** and necessarily satisfies

$$g \geq 0, \quad \text{a.e.}, \quad \int_{-\infty}^{\infty} g(t)dt = 1. \quad (6)$$

Here, a.e. abbreviates "almost everywhere" (the analogue of "almost certainly" when probability is replaced by Lebesgue measure) and both integrals will be interpreted in the Lebesgue sense. The Lebesgue integral, Lebesgue measure, and related questions will be discussed in Chapter 6.

A third type of d.f., called singular, is defined in Chapter 8. The most general d.f. on R occurring in probability theory is a (convex) linear combination of the three types mentioned (see Section 8.1).

Distribution functions occupy a preeminent position in probability theory, which is primarily concerned with properties of r.v.s verifiable via their d.f.s.

Random variables with discrete d.f.s as in (3) will be called **discrete r.v.s** (with values x_j) while r.v.s with absolutely continuous d.f.s as in (5) will be called **absolutely continuous r.v.s** (with density g). The next chapter deals with some important special cases thereof.

Theorem 1. If (X_1, \dots, X_n) and (Y_1, \dots, Y_n) are random vectors with identical distribution functions, that is,

$$F_{X_1, \dots, X_n} = F_{Y_1, \dots, Y_n},$$

then $g(X_1, \dots, X_n)$ and $g(Y_1, \dots, Y_n)$ have identical d.f.s for any finite Borel function g on R^n .

PROOF. It follows from Theorem 1.4.4 that $g(X_1, \dots, X_n)$ and $g(Y_1, \dots, Y_n)$ are r.v.s. Set

$$\mathcal{G} = \{B: B \in \mathcal{B}^n, P\{(X_1, \dots, X_n) \in B\} = P\{(Y_1, \dots, Y_n) \in B\}\},$$

$$\mathcal{D} = \left\{ D: D = \bigcup_{j=1}^n (-\infty, c_j), c_j \text{ real} \right\}.$$

Then $\mathcal{G} \supset \mathcal{D}$ by hypothesis, and, moreover, it is easy to verify that \mathcal{G} is a λ -class and \mathcal{D} is a π -class. By Theorem 1.3.2, $\mathcal{G} \supset \sigma(\mathcal{D}) = \mathcal{B}^n$. Hence, since for $\lambda \in (-\infty, \infty)$, $A \equiv \{(x_1, \dots, x_n) \in R^n: g(x_1, \dots, x_n) < \lambda\} \in \mathcal{B}^n$,

$$\begin{aligned} P\{g(X_1, \dots, X_n) < \lambda\} &= P\{(X_1, \dots, X_n) \in A\} = P\{(Y_1, \dots, Y_n) \in A\} \\ &= P\{g(Y_1, \dots, Y_n) < \lambda\}. \end{aligned} \quad \square$$

Corollary 1. If (X_1, \dots, X_n) and (Y_1, \dots, Y_n) are random vectors with identical distribution functions, then for any Borel function g on R^n and linear Borel set B

$$P\{g(X_1, \dots, X_n) \in B\} = P\{g(Y_1, \dots, Y_n) \in B\}.$$

EXERCISES 1.6

1. Prove that G as defined in (3) is a d.f.; verify that (Ω, \mathcal{F}, P) as defined thereafter is a probability space and that $F_X = G$.
2. If $p_{\rho, t} \geq 0$, $\rho \in S$, $t \in T$, and $\sum_{t \in T} \sum_{\rho \in S} p_{\rho, t} = 1$, where S and T are countable subsets of $(-\infty, \infty)$, define a probability space and random variables X, Y on it with

$$F_X(x) = \sum_{\rho < x} \sum_{t \in T} p_{\rho, t} \quad F_Y(y) = \sum_{t < y} \sum_{\rho \in S} p_{\rho, t}.$$

Hint: Take $\Omega = S \times T$.

3. Prove that if $H(x) = P\{X \leq x\}$, where X is a r.v. on (Ω, \mathcal{F}, P) , then H is a nondecreasing right-continuous function with $H(-\infty) = 0$, $H(+\infty) = 1$. Some authors define such functions to be d.f.s.
4. For a d.f. F , the set

$$S(F) = \{x: F(x + \varepsilon) - F(x - \varepsilon) > 0 \text{ for all } \varepsilon > 0\}$$

is called the **support** of F . Show that each jump point of F belongs to the support and that each isolated point of the support is a jump point. Prove that $S(F)$ is a closed set and give an example of a discrete d.f. whose support is $(-\infty, \infty)$. Any point $x \in S(F)$ is called a **point of increase** of F .

Comments

The notion of a π -class and λ -class (Section 3) seems to have originated with Dynkin (1961).

References

- J. L. Doob, "Supplement," *Stochastic Processes*, Wiley, New York, 1953.
E. B. Dynkin, *Theory of Markov Processes* (D. E. Brown, translator), Prentice-Hall, Englewood Cliffs, New Jersey, 1961.
Paul R. Halmos, *Measure Theory*, Van Nostrand, Princeton, 1950; Springer-Verlag, Berlin and New York, 1974.
Felix Hausdorff, *Set Theory* (J. Aumann *et al.*, translators), Chelsea, New York, 1957.
Stanislaw Saks, *Theory of the Integral*, (L. C. Young, translator), Stechert-Hafner, New York, 1937.

2 Binomial Random Variables

The major theorems of probability theory fall into a natural dichotomy—those which are analytic in character and those which are measure-theoretic. In the latter category are zero-one laws, the Borel–Cantelli lemma, strong laws of large numbers, and indeed any result which requires the apparatus of a probability space.

On the other hand, findings such as central limit theorems, weak laws of large numbers, etc., may be stated entirely in terms of distribution functions and hence are intrinsically analytic. The fact that these distribution functions (d.f.s) are frequently attached to random variables (r.v.s) does not alter the fact that the underlying probability space (on which the r.v.s are defined) is of no consequence in the statement of the analytic result. Indeed, it would be possible, although in many cases unnatural, to express the essential finding without any mention of r.v.s (and *a fortiori* of a probability space).

In presenting theorems, distributions will generally be attached to r.v.s. For analytic results, the r.v.s are inessential but provide a more colorful and intuitive background. In the case of measure-theoretic results, it suffices to recognize that a probability space and r.v.s on it with the given distributions can always be constructed.

2.1 Poisson Theorem, Interchangeable Events, and Their Limiting Probabilities

The term **distribution** will be used in lieu of either d.f. or p.d.f. The binomial distribution is not only one of the most frequently encountered in practice but plays a considerable theoretical role due in part to its elemental simplicity

and in part to its leading (in the limit) to even more important distributions, such as the Poisson and the normal.

As is customary, the combinatorial symbol $\binom{n}{j}$ abbreviates $n!/j!(n-j)!$ for integers $n \geq j \geq 0$ and is defined as zero when $j > n$ or $j < 0$.

The **binomial d.f.** is a discrete d.f. with jumps of sizes $\binom{n}{j}p^j(1-p)^{n-j}$ at the points $j = 0, 1, \dots, n$. In other words, its p.d.f. is completely characterized (Section 1.6) by $S = \{0, 1, 2, \dots, n\}$ and $\{\binom{n}{j}p^j(1-p)^{n-j}, 0 \leq j \leq n\}$. Here, $p \in (0, 1)$ and n is a positive integer.

The construction of Section 1.6 shows that it is a simple matter to define a probability space and a r.v. X on this space whose associated d.f. is binomial. Thus, a r.v. X will be alluded to as a **binomial r.v.** with p.d.f. $b(k; n, p)$ if for some positive integer n and $p \in (0, 1)$

$$P\{X = k\} = \binom{n}{k} p^k q^{n-k} \equiv b(k; n, p), \quad k = 0, 1, \dots, n, \quad q = 1 - p. \quad (1)$$

The **Poisson d.f.** is a discrete d.f. with jumps of sizes $\lambda^j e^{-\lambda}/j!$ at the points $j = 0, 1, 2, \dots$, and so X will be called a **Poisson r.v.** with p.d.f. $p(k; \lambda)$ if for some $\lambda \in (0, \infty)$

$$P\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!} \equiv p(k; \lambda), \quad k = 0, 1, 2, \dots \quad (2)$$

The quantities n and p of the binomial distribution and λ of the Poisson distribution are referred to as **parameters**. If $\lambda = 0$ in (2) or $p = 0$ in (1), the r.v.s (or their distributions) are called **degenerate** since then $P\{X = 0\} = 1$ (i.e., its d.f. has a single jump of size one). Degeneracy likewise occurs if $p = 1$ in (1) since then $P\{X = n\} = 1$.

The normal d.f. is an absolutely continuous d.f. (Section 1.6) G with

$$G(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-(u-\theta)^2/2\sigma^2} du, \quad x \in R = [-\infty, \infty]. \quad (3)$$

The parameters of the distribution are $\theta \in (-\infty, \infty)$ and $\sigma^2 \in (0, \infty)$. Here, $\sigma^2 = 0$ may be regarded as a degenerate distribution. It is customary to denote the **standard normal distribution** corresponding to the case $\theta = 0$, $\sigma^2 = 1$ by

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du, \quad \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \quad (4)$$

Here $\varphi(x)$ is called the (standard) normal density function. A r.v. X is normally distributed with parameters θ and σ^2 , written $X \sim N(\theta, \sigma^2)$ if

$$P\{X < x\} = G(x) = \Phi\left(\frac{x-\theta}{\sigma}\right), \quad x \in R. \quad (5)$$

Commencing here and sporadically throughout the remaining chapters, a simple but incredibly expedient device for asymptotic calculations, introduced by the English mathematician G. H. Hardy, will be utilized. Thus, if $r_n > 0$, $n \geq 1$, a sequence $\{b_n, n \geq 1\}$ of real numbers is said to be **little o of r_n** (resp. **capital O of r_n**), denoted $b_n = o(r_n)$ (resp. $b_n = O(r_n)$) if $\lim_{n \rightarrow \infty} (b_n/r_n) = 0$ (resp. $|b_n|/r_n < C < \infty, n \geq 1$). It is a simple matter to check that the sum of two quantities that are $o(r_n)$ (resp. $O(r_n)$) is likewise $o(r_n)$ (resp. $O(r_n)$). Thus, a veritable calculus can be established involving one or the other or both of o and O . In a similar vein, $b_n \sim r_n$ if $\lim_{n \rightarrow \infty} (b_n/r_n) = 1$. The same notations and calculus apply to real functions $f(x)$ where either $x \rightarrow \infty$ or $x \rightarrow 0$. Thus, for example, $\log(1 + x) = x + O(x^2) = O(|x|)$ as $x \rightarrow 0$.

The first theorem, purely analytic in character, states that the limit of binomial p.d.f.s may be a Poisson p.d.f. and hence implies the analogous statement for d.f.s.

Theorem 1 (Poisson, 1832). *If S_n is a binomial r.v. with p.d.f. $b(k; n, p_n)$, $n \geq 1$, and as $n \rightarrow \infty$, $np_n = \lambda + o(1)$ for some $\lambda \in [0, \infty)$, then for $k = 0, 1, 2, \dots$*

$$\lim_{n \rightarrow \infty} P\{S_n = k\} = \frac{\lambda^k e^{-\lambda}}{k!}.$$

PROOF. Set $q_n = 1 - p_n$. Since $n \log(1 - (\lambda/n) + o(1/n)) = n((-\lambda/n) + o(1/n)) \rightarrow -\lambda$,

$$\begin{aligned} P\{S_n = k\} &= \binom{n}{k} p_n^k q_n^{n-k} \\ &= \frac{n(n-1) \cdots (n-k+1)}{k!} \left[\frac{\lambda}{n} + o\left(\frac{1}{n}\right) \right]^k \left[1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right) \right]^{n-k} \\ &= \frac{n}{n} \left(\frac{n-1}{n} \right) \cdots \left(\frac{n-k+1}{n} \right) \frac{1}{k!} [\lambda + o(1)]^k \\ &\quad \times \left[1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right) \right]^n \left[1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right) \right]^{-k} \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}. \quad \square \end{aligned}$$

By refining the arguments of the preceding proof, a more general result is obtainable. To pave the way for the introduction of a new p.d.f. subsuming the binomial p.d.f., two lemmas which are probabilistic or measure analogues of (2) and (3) of Section 1.2 will be proved.

Lemma 1. *Let φ be an additive set function on an algebra \mathcal{A} and $A_j \in \mathcal{A}$, $1 \leq j \leq n$. If $|\varphi(\bigcup_1^n A_j)| < \infty$, then*

$$\varphi\left(\bigcup_1^n A_j\right) = \sum_{j=1}^n \varphi(A_j) - \sum_{1 \leq j_1 < j_2 \leq n} \varphi(A_{j_1} A_{j_2}) + \cdots + (-1)^{n-1} \varphi(A_1 A_2 \cdots A_n). \quad (6)$$

PROOF. Set $\varphi_j = \varphi(A_j)$, $\varphi_{j_1, j_2, \dots, j_r} = \varphi(A_{j_1} A_{j_2} \cdots A_{j_r})$, $r \geq 2$. Clearly, (6) holds for $n = 1$. Suppose, inductively, that it holds for a fixed but arbitrary positive integer m in $[1, n]$, whence $|\varphi(\bigcup_1^{m+1} A_j)| < \infty$. Now,

$$\begin{aligned}\varphi\left(\bigcup_1^{m+1} A_j\right) &= \varphi\left(\bigcup_1^m A_j\right) + \varphi\left(A_{m+1} - \bigcup_1^m A_j\right), \\ \varphi(A_{m+1}) &= \varphi\left(\bigcup_1^m A_j A_{m+1}\right) + \varphi\left(A_{m+1} - \bigcup_1^m A_j\right),\end{aligned}\tag{7}$$

where $|\varphi(\bigcup_1^m A_j)| < \infty$, $|\varphi(A_{m+1})| < \infty$, $|\varphi(\bigcup_1^m A_j A_{m+1})| < \infty$. By the induction hypothesis

$$\begin{aligned}\varphi\left(\bigcup_1^m A_j\right) &= \sum_{j=1}^m \varphi_j - \sum_{1 \leq j_1 < j_2 \leq m} \varphi_{j_1, j_2} + \cdots + (-1)^{m-1} \varphi_{1, 2, \dots, m}, \\ \varphi\left(\bigcup_1^m A_j A_{m+1}\right) &= \sum_{j=1}^m \varphi_{j, m+1} - \sum_{1 \leq j_1 < j_2 \leq m} \varphi_{j_1, j_2, m+1} + \cdots \\ &\quad + (-1)^{m-1} \varphi_{1, 2, \dots, m, m+1},\end{aligned}\tag{8}$$

and so employing (8) in (7) the conclusion (6) with $n = m + 1$ follows. \square

Lemma 1 yields immediately

Corollary 1 (Poincaré Formula). *If A_1, \dots, A_n are events in a probability space (Ω, \mathcal{F}, P) and $T_k = \sum_{1 \leq j_1 < j_2 < \dots < j_k \leq n} P\{A_{j_1} A_{j_2} \cdots A_{j_k}\}$, then*

$$P\left\{\bigcup_1^n A_j\right\} = \sum_1^n (-1)^{k-1} T_k.\tag{9}$$

Definition. Events A_1, \dots, A_n of a probability space (Ω, \mathcal{F}, P) are called **interchangeable** (exchangeable) if for all choices of $1 \leq i_1 < \cdots < i_j \leq n$ and all $1 \leq j \leq n$,

$$P\{A_{i_1} \cdot A_{i_2} \cdots A_{i_j}\} = p_j.\tag{10}$$

Interchangeable events seem to have been introduced by Haag (1924–1928) but the basic results involving these are due to de Finetti (1937).

Corollary 2. *If $\{A_1, \dots, A_n\}$ are interchangeable events of a probability space (Ω, \mathcal{F}, P) with p_j defined by (10), then*

$$P\left\{\bigcup_{j=1}^n A_j\right\} = np_1 - \binom{n}{2} p_2 + \binom{n}{3} p_3 - \cdots + (-1)^{n-1} p_n.\tag{11}$$

The novel part of the next lemma is (13), since for $A = \Omega$, (12) is merely the complementary version of the Poincaré formula (9).

Lemma 2. Let φ be an additive set function on an algebra \mathcal{A} and $A_j \in \mathcal{A}$, $1 \leq j \leq n$. If $|\varphi(\bigcup_1^n A_j)| < \infty$ and

$$B_m = \{\omega : \omega \text{ exactly } m \text{ of the events } A_1, \dots, A_n\}, \quad 0 \leq m \leq n,$$

then for any $A \in \mathcal{A}$

$$\begin{aligned} \varphi(B_0 A) &= \varphi(A) - \sum_{j=1}^n \varphi(A_j A) + \sum_{1 \leq j_1 < j_2 \leq n} \varphi(A_{j_1} A_{j_2} A) - \dots \\ &\quad + (-1)^n \varphi(A_1 \dots A_n A), \end{aligned} \quad (12)$$

$$\begin{aligned} \varphi(B_m) &= \sum_{1 \leq j_1 < \dots < j_m \leq n} \varphi(A_{j_1} \dots A_{j_m}) - \binom{m+1}{m} \sum_{1 \leq j_1 < \dots < j_{m+1} \leq n} \\ &\quad \times \varphi(A_{j_1} \dots A_{j_{m+1}}) + \dots + (-1)^{n-m} \binom{n}{m} \varphi(A_1 \dots A_n). \end{aligned} \quad (13)$$

PROOF. By Lemma 1

$$\begin{aligned} \varphi(B_0 A) &= \varphi(A) - \varphi\left(\bigcup_1^n A_j A\right) \\ &= \varphi(A) - \sum_{j=1}^n \varphi(A_j A) + \sum_{1 \leq j_1 < j_2 \leq n} \varphi(A_{j_1} A_{j_2} A) - \dots \\ &\quad + (-1)^n \varphi(A_1 \dots A_n A), \end{aligned}$$

proving (12). For any choice of j_i , $1 \leq i \leq m$ with $1 \leq j_1 < \dots < j_m \leq n$, let $1 \leq i_1 < \dots < i_{n-m} \leq n$ signify the remaining integers in $[1, n]$. Then taking $A = A_{j_1} \dots A_{j_m}$ in (12)

$$\begin{aligned} \varphi(A_{j_1} \dots A_{j_m} A_{i_1}^c \dots A_{i_{n-m}}^c) &= \varphi(A \cdot A_{i_1}^c \dots A_{i_{n-m}}^c) \\ &= \varphi(A) - \sum_{h=1}^{n-m} \varphi(A_{i_h} A) + \sum_{1 \leq h_1 < h_2 \leq n-m} \varphi(A_{i_{h_1}} A_{i_{h_2}} A) - \dots \\ &\quad + (-1)^{n-m} \varphi(A_{i_1} \dots A_{i_{n-m}} A). \end{aligned} \quad (14)$$

Since for any $1 \leq m \leq n$

$$B_m = \bigcup_{1 \leq j_1 < \dots < j_m \leq n} A_{j_1} \dots A_{j_m} \cdot A_{i_1}^c \dots A_{i_{n-m}}^c$$

represents B_m as a disjoint union, (13) follows by summing over j_1, \dots, j_m in (14). \square

The following corollaries are direct consequences of Lemma 2.

Corollary 3. If A_1, \dots, A_n are events of a probability space (Ω, \mathcal{F}, P) and $T_k = \sum_{1 \leq j_1 < \dots < j_k \leq n} P\{A_{j_1} A_{j_2} \dots A_{j_k}\}$ for $k = 1, 2, \dots, n$, then

$$\begin{aligned} P\{\text{exactly } m \text{ of } A_1, \dots, A_n \text{ occur}\} \\ = T_m - \binom{m+1}{m} T_{m+1} + \dots + (-1)^{n-m} \binom{n}{m} T_n. \end{aligned} \quad (15)$$

In the case of interchangeable events, (15) simplifies considerably and may be expressed in terms of the classical difference operator:

$$\Delta p_n = \Delta^1 p_n = p_{n+1} - p_n, \quad \Delta^k p_n = \Delta(\Delta^{k-1} p_n), \quad k \geq 2.$$

Corollary 4. If A_1, \dots, A_n are interchangeable events of a probability space (Ω, \mathcal{F}, P) with $p_j = P\{A_1 A_2 \cdots A_j\}$, $1 \leq j \leq n$, then, setting $p_0 = 1$, for any integer m in $[0, n]$

$$\begin{aligned} P\{\text{exactly } m \text{ of } A_1, \dots, A_n \text{ occur}\} &= \sum_{i=m}^n (-1)^{i-m} \binom{n}{i} \binom{i}{m} p_i \\ &= \binom{n}{m} \sum_{j=0}^{n-m} (-1)^j \binom{n-m}{j} p_{m+j} \\ &= \binom{n}{m} (-1)^{n-m} \Delta^{n-m} p_m. \end{aligned} \quad (16)$$

Since the events B_m , $0 \leq m \leq n$, of (16) constitute a finite partition of Ω , their probabilities are nonnegative and add up to one. This shows that the probabilities p_j in (10) cannot be arbitrary positive numbers (see Exercise 9).

In view of the correspondence between events A_j and their indicators I_{A_j} , it should not be surprising, setting $S_n = \sum_{j=1}^n I_{A_j}$, that the event B_m of (16) is equivalent to $\{\omega : S_n = m\}$. Indeed, discussion of the limiting behavior of the probabilities in (16) will be couched in terms of the r.v.s S_n .

Theorem 2. If for each $n \geq 1$, S_n is a random variable with

$$P\{S_n = m\} = \sum_{j=0}^{n-m} (-1)^j \binom{n}{m+j} \binom{m+j}{m} p_{m+j}^{(n)}, \quad 0 \leq m \leq n, \quad (17)$$

such that for some $\lambda_0 \in (0, \infty)$ and $\lambda_n \in [0, \infty)$

$$0 \leq n(n-1) \cdots (n-k+1) p_k^{(n)} \leq \lambda_0^k, \quad 1 \leq k \leq n, \quad (18)$$

$$n^k p_k^{(n)} = \lambda_n^k + o(1), \quad k \geq 1, \quad \text{as } n \rightarrow \infty, \quad (19)$$

then

$$P\{S_n = m\} - \frac{\lambda_n^m e^{-\lambda_n}}{m!} \rightarrow 0, \quad m \geq 0.$$

PROOF. For $m \geq 0$, fixed n_0 and all large n

$$\begin{aligned} P\{S_n = m\} - \frac{\lambda_n^m e^{-\lambda_n}}{m!} &= \sum_{j=0}^{n_0-1} \frac{(-1)^j}{m! j!} \left[\frac{n!}{(n-m-j)!} p_{m+j}^{(n)} - \lambda_n^{m+j} \right] \\ &\quad + \sum_{j=n_0}^{n-m} (-1)^j \frac{n!}{m! j! (n-m-j)!} p_{m+j}^{(n)} \\ &\quad - \sum_{j=n_0}^{\infty} (-1)^j \frac{\lambda_n^{m+j}}{m! j!} = I_1 + I_2 + I_3 \text{ (say).} \end{aligned}$$

Since $\lambda_n + o(1) = np_1^{(n)} \leq \lambda_0$, there exists $\lambda > \lambda_0$ such that $\lambda_n < \lambda$, all n . For any $\varepsilon > 0$ and fixed m , choose n_0 to satisfy

$$\sum_{j=n_0}^{\infty} \frac{\lambda^{m+j}}{m!j!} < \varepsilon,$$

whence $|I_3| < \varepsilon$ and via (18)

$$|I_2| \leq \sum_{j=n_0}^{n-m} \frac{n(n-1)\cdots(n-m-j+1)}{m!j!} p_{m+j}^{(n)} \leq \sum_{j=n_0}^{n-m} \frac{\lambda_0^{m+j}}{m!j!} < \varepsilon.$$

Now, for $1 \leq m+j \leq m+n_0$, by (19)

$$\begin{aligned} \frac{n!}{(n-m-j)!} p_{m+j}^{(n)} - \lambda_n^{m+j} &\leq n^{m+j} p_{m+j}^{(n)} - \lambda_n^{m+j} \\ &= \lambda_n^{m+j} + o(1) - \lambda_n^{m+j} = o(1), \\ \frac{n!}{(n-m-j)!} p_{m+j}^{(n)} - \lambda_n^{m+j} &\geq (n-m-j)^{m+j} p_{m+j}^{(n)} - \lambda_n^{m+j} \\ &= \left(1 - \frac{m+j}{n}\right)^{m+j} n^{m+j} p_{m+j}^{(n)} - \lambda_n^{m+j} \\ &= \left(1 - \frac{m+j}{n}\right)^{m+j} (\lambda_n^{m+j} + o(1)) - \lambda_n^{m+j} = o(1), \end{aligned}$$

implying $I_1 = o(1)$, and the theorem follows. \square

This result will be utilized in Theorem 3.1.2 to obtain (under suitable conditions) a limiting Poisson distribution for the number of empty cells in the casting of balls into cells at random. Note that this implies Theorem 1.

It follows immediately from Theorem 1 that the sequence $F_n(x) = \sum_{k < x} b(k; n, p)$ of binomial d.f.s tends to the Poisson d.f. $F(x) = \sum_{k < x} p(k; \lambda)$ as $n \rightarrow \infty$, $np \rightarrow \lambda \in (0, \infty)$. In this case, not only does

$$b(k; n, p) \rightarrow p(k; \lambda), \quad (20)$$

for every positive integer k , but also

$$\sum_{k=0}^{\infty} |b(k; n, p) - p(k; \lambda)| \rightarrow 0 \quad (21)$$

as $n \rightarrow \infty$, $np \rightarrow \lambda \in (0, \infty)$, where $b(k; n, p)$ is defined as zero for $k > n$. This is an instance of the more general phenomenon of the following

EXAMPLE 1. Let $\Omega = \{\omega_1, \omega_2, \dots\}$, \mathcal{F} = class of all subsets of Ω , and let $\{P_n, P_n, n \geq 1\}$ be a sequence of probability measures on (Ω, \mathcal{F}) such that $P_n \rightarrow P$ as $n \rightarrow \infty$, that is,

$$p_{n,j} \equiv P_n\{\omega_j\} \rightarrow P\{\omega_j\} \equiv p_j, \quad j = 1, 2, \dots \quad (22)$$

Then, as $n \rightarrow \infty$

$$\sum_{j=1}^{\infty} |p_{n,j} - p_j| \rightarrow 0. \quad (23)$$

PROOF. For any $\varepsilon > 0$, choose $N = N_\varepsilon$ such that

$$\sum_{j>N} p_j < \varepsilon. \quad (24)$$

Now (22) ensures that as $n \rightarrow \infty$

$$\sum_{j=1}^N |p_{n,j} - p_j| \rightarrow 0, \quad (25)$$

whence as $n \rightarrow \infty$

$$\left| \sum_{j>N} p_{n,j} - \sum_{j>N} p_j \right| = \left| \sum_{j=1}^N p_{n,j} - \sum_{j=1}^N p_j \right| \rightarrow 0. \quad (26)$$

Then (24) and (26) entail

$$\overline{\lim}_{n \rightarrow \infty} \sum_{j>N} p_{n,j} \leq \varepsilon, \quad (27)$$

so that via (24), (25), (26), and (27)

$$\overline{\lim}_{n \rightarrow \infty} \sum_{j=1}^{\infty} |p_{n,j} - p_j| = \overline{\lim}_{n \rightarrow \infty} \sum_{j>N} |p_{n,j} - p_j| \leq \overline{\lim}_{n \rightarrow \infty} \sum_{j>N} (p_{n,j} + p_j) \leq 2\varepsilon,$$

which, in view of the arbitrariness of ε , is tantamount to (23). \square

EXERCISES 2.1

1. Prove that for $n = 1, 2, \dots$

$$\int_{-\infty}^{\infty} \varphi(x) dx = 1, \quad \int_{-\infty}^{\infty} x^{2n-1} \varphi(x) dx = 0, \quad \int_{-\infty}^{\infty} x^{2n} \varphi(x) dx = 1 \cdot 3 \cdots (2n-1).$$

2. Verify that

$$\sum_{k=0}^{\infty} kp(k; \lambda) = \lambda = \sum_{k=0}^{\infty} k^2 p(k; \lambda) - \lambda^2,$$

$$\sum_{k=0}^n p(k; \lambda) = \frac{1}{n!} \int_{\lambda}^{\infty} e^{-x} x^n dx,$$

$$\sum_{k=0}^{\infty} k^3 p(k; \lambda) = \lambda^3 + 3\lambda^2 + \lambda,$$

$$\sum_{k=0}^n k^j b(k; n, p) = np \quad \text{or} \quad npq + (np)^2 \quad \text{as} \quad j = 1 \text{ or } 2.$$

3. Verify for all positive integers n_1, n_2 and nonnegative integers k that

$$\sum_{j=0}^{n_1} \binom{n_1}{j} \binom{n_2}{k-j} = \binom{n_1 + n_2}{k}.$$

4. Find the d.f. of X^2 when

- i. X is a Poisson r.v. with parameter λ ,
- ii. X is $N(0, 1)$.

Prove under (i) that $\lambda^n e^{-\lambda}/n! \leq P\{X \geq n\} \leq \lambda^n/n!$ for $n \geq 0$.

5. A deck of N cards numbered $1, 2, \dots, N$ is shuffled. A "match" occurs at position j if the card numbered j occupies the j th place in the deck. If $p_m^{(N)}$ denotes the probability of exactly m matches, $0 \leq m \leq N$, prove that

$$p_m^{(N)} = \frac{1}{m!} \left(1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^{N-m}}{(N-m)!} \right), \quad (\text{i})$$

$$\left| p_m^{(N)} - \frac{e^{-1}}{m!} \right| < \frac{1}{m!(N-m+1)!}, \quad (\text{ii})$$

$$q_2^{(N)} \rightarrow e^{-1} \left(\frac{1}{2!} + \frac{1}{3!} + \cdots \right), \quad (\text{iii})$$

where $q_m^{(N)}$ is the probability of at least m matches.

6. For $k = 0, 1, 2, \dots, r = 1, 2, \dots, 0 < p < 1$, and $q = 1 - p$, put $f(k; r, p) = \binom{r+k-1}{k} p^k q^{r-k}$. Prove that $\sum_{k=0}^{\infty} f(k; r, p) = 1$ for every r and p and that if $q = q(r)$ and $rq \rightarrow \lambda \in (0, 1)$ as $r \rightarrow \infty$, then

$$f(k; r, p) \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}.$$

($f(k; r, p)$ is called a **negative binomial** p.d.f. with parameters r and p .)

7. **Bonferroni Inequalities.** If A_1, \dots, A_n are events and

$$T_k = \sum_{1 \leq j_1 < \cdots < j_k \leq n} P\{A_{j_1} A_{j_2} \cdots A_{j_k}\},$$

then

$$\sum_1^{2m} (-1)^{k-1} T_k \leq P\left\{\bigcup_1^n A_j\right\} \leq \sum_1^{2m-1} (-1)^{j-1} T_j, \quad 2m \leq n.$$

8. Find the formula analogous to (15) for the probability that at least m of the events A_1, \dots, A_n occur.

9. Let $p_0 = 1$ and p_m , $1 \leq m \leq n$, be real numbers such that

$$q_m = \binom{n}{m} \sum_{j=0}^{n-m} (-1)^j \binom{n-m}{j} p_{m+j} \geq 0$$

for $0 \leq m \leq n$. Then $\sum_{m=0}^n q_m = p_0 = 1$ and there exists a probability space (Ω, \mathcal{F}, P) and interchangeable events A_1, \dots, A_n with $p_j = P\{A_1 A_2 \cdots A_j\}$, $1 \leq j \leq n$. Hint: Take $\Omega = \{\omega = (\omega_1, \dots, \omega_n) : \omega_j = 0 \text{ or } 1, 1 \leq j \leq n\}$, \mathcal{F} = class of all subsets of Ω , and $P(\{\omega\}) = q_m/\binom{n}{m}$ whenever $\sum_{j=1}^n \omega_j = m$. Set

$$A_j = \{\omega = (\omega_1, \dots, \omega_n) : \omega_j = 1\}, 1 \leq j \leq n.$$

10. Verify that if $p_j = p^j$ in (16) i.e., in Exercise 9, the probabilities q_m coalesce to the binomial and $P\{A_{i_1} A_{i_2} \cdots A_{i_k}\} = \prod_{j=1}^k P\{A_{i_j}\}$, $1 \leq i_1 < \cdots < i_k \leq n$.
11. Prove that if $p_j = \binom{N-j}{r-j}/\binom{N}{r}$, $j = 0, \dots, r$, $r \leq N$, the p.d.f. in (16) is **hypergeometric**, i.e., $q_m = \binom{m}{r} \binom{N-m}{r-m} / \binom{N}{r}$; also if $p_j = 1/(j+1)$, $\{q_m\}$ is the discrete **uniform** distribution, i.e., $q_m = 1/(n+1)$, $0 \leq m \leq n$.
12. Prove that if $r = r_n = \lambda_n N/n \rightarrow \infty$ and $\lambda_n \rightarrow \lambda \in [0, \infty)$ in the hypergeometric case of Exercise 11, then $p_j = p_j^{(n)}$ satisfies (18) and (19) of Theorem 2, so that the hypergeometric p.d.f. tends to the Poisson p.d.f. under the stated condition.
13. Let $\{A_n, n \geq 1\}$ be an infinite sequence of interchangeable events on a probability space (Ω, \mathcal{F}, P) , that is, for all $n \geq 1$ and indices

$$1 \leq i_1 < \cdots < i_n \quad P\{A_{i_1} A_{i_2} \cdots A_{i_n}\} = p_n.$$

Prove that

$$\begin{aligned} P\{A_n, \text{i.o.}\} &= P\left\{\bigcup_{j=1}^{\infty} A_j\right\} = 1 - \lim_{k \rightarrow \infty} (-1)^k \Delta^k p_0, \\ P\{\liminf A_n\} &= P\left\{\bigcap_{n=1}^{\infty} A_n\right\} = \lim_{k \rightarrow \infty} p_k. \end{aligned}$$

Hint: Recall Exercise 1.5.5.

2.2 Bernoulli, Borel Theorems

Due to the ease of computations involving the binomial distribution, many important notions such as weak and strong laws of large numbers and the central limit theorem applicable to wide classes of random variables may be adumbrated here.

Theorem 1 (Bernoulli Weak Law of Large Numbers, 1713). *Let $\{S_n\}$ be a sequence of binomial random variables with p.d.f.s $b(k; n, p)$, $n \geq 1$. Then for all $\varepsilon > 0$,*

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right\} = 0.$$

PROOF.

$$\begin{aligned} P\left\{\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right\} &= P\{|S_n - np| \geq n\varepsilon\} = \sum_{|k-np| \geq n\varepsilon} P\{S_n = k\} \\ &\leq \sum_{|k-np| \geq n\varepsilon} \frac{(k-np)^2}{n^2 \varepsilon^2} P\{S_n = k\} \\ &\leq \frac{1}{n^2 \varepsilon^2} \sum_{k=0}^n (k-np)^2 P\{S_n = k\}. \end{aligned}$$

The prior inequality is a special case of the simple but very useful Tchebychev inequality discussed in Chapter 4. Moreover, the last sum which will be identified in Chapter 4 as the variance of S_n (or its d.f.) equals

$$\begin{aligned}
 & \sum_{k=0}^n (k - np)^2 \binom{n}{k} p^k q^{n-k} \\
 &= \sum_{k=0}^n [k(k-1) - (2np-1)k + n^2 p^2] \binom{n}{k} p^k q^{n-k} \\
 &= n(n-1)p^2 \sum_{k=2}^n \frac{(n-2)!}{(k-2)!(n-k)!} p^{k-2} q^{n-k} \\
 &\quad - (2np-1)np \sum_{k=1}^n \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k} + n^2 p^2 \\
 &= n(n-1)p^2 - np(2np-1) + n^2 p^2 = npq. \tag{1}
 \end{aligned}$$

Therefore,

$$P \left\{ \left| \frac{S_n}{n} - p \right| \geq \varepsilon \right\} \leq \frac{pq}{n\varepsilon^2} = o(1). \quad \square \tag{2}$$

The prior theorem may be strengthened by the simple device of replacing $(k - np)^2$ by $(k - np)^4$.

Theorem 2. Let S_n , $n \geq 1$, be binomial r.v.s with p.d.f.s $b(k; n, p)$, $0 < p < 1$, $n \geq 1$. Then for every $\varepsilon > 0$

$$\sum_{n=1}^{\infty} P \left\{ \left| \frac{S_n}{n} - p \right| \geq \varepsilon \right\} < \infty.$$

PROOF. Set $k_j = k(k-1)\cdots(k-j+1)$, $1 \leq j \leq k$, whence

$$\begin{aligned}
 k &= k_1, & k^2 &= k_2 + k_1, & k^3 &= k_3 + 3k_2 + k_1, \\
 k^4 &= k_4 + 6k_3 + 7k_2 + k_1.
 \end{aligned}$$

Since

$$\sum_{k=0}^n k_j \binom{n}{k} p^k q^{n-k} = n_j p^j, \quad 0 \leq j \leq n,$$

it follows that

$$\begin{aligned}
 & \sum_{k=0}^n (k - np)^4 \binom{n}{k} p^k q^{n-k} \\
 &= \sum_{k=0}^n (k^4 - 4k^3 np + 6k^2 n^2 p^2 - 4kn^3 p^3 + n^4 p^4) \binom{n}{k} p^k q^{n-k} \\
 &= (n_4 p^4 + 6n_3 p^3 + 7n_2 p^2 + n_1 p) - 4np(n_3 p^3 + 3n_2 p^2 + n_1 p) \\
 &\quad + 6n^2 p^2(n_2 p^2 + n_1 p) - 4n^3 p^3(n_1 p) + n^4 p^4 \\
 &= p^4(n_4 - 4nn_3 + 6n^2 n_2 - 4n^3 n_1 + n^4) + p^3(6n_3 - 12nn_2 + 6n^2 n_1) \\
 &\quad + p^2(7n_2 - 4nn_1) + n_1 p \\
 &= (3n^2 - 6n)p^4 - (6n^2 - 12n)p^3 + (3n^2 - 7n)p^2 + np \\
 &= 3n(n-2)p^2(p-1)^2 + npq = npq(3npq - 6pq + 1).
 \end{aligned}$$

Consequently, proceeding as at the outset of Theorem 1,

$$\begin{aligned}
 P\left\{\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right\} &= \sum_{|k-np| \geq n\varepsilon} P\{S_n = k\} \leq \frac{1}{n^4 \varepsilon^4} \sum_{k=0}^n (k - np)^4 \binom{n}{k} p^k q^{n-k}, \\
 &= \frac{pq(3npq - 6pq + 1)}{n^3 \varepsilon^4} = O\left(\frac{1}{n^2}\right),
 \end{aligned}$$

and therefore the series in question converges. \square

The strong law of large numbers involves the pointwise convergence of a sequence of random variables on a probability space. A discussion of this will be facilitated by

Lemma 1. *If $\{Y_n, n \geq 1\}$ is any sequence of random variables on a probability space (Ω, \mathcal{F}, P) , then $P\{\lim_{n \rightarrow \infty} Y_n = 0\} = 1$ iff*

$$P\left\{|Y_n| > \frac{1}{k}, \text{ i.o.}\right\} = 0, \quad k = 1, 2, \dots$$

PROOF. Let $A = \bigcup_{k=1}^{\infty} A_k$, where $A_k = \{|Y_n| > 1/k, \text{ i.o.}\}$. If $\omega \notin A$, then $|Y_n(\omega)| > 1/k$ for only finitely many n for every positive integer k , implying $\lim Y_n(\omega) = 0$. Conversely, if $\lim Y_n(\omega) = 0$, then $\omega \notin A_k$ for $k = 1, 2, \dots$, and so $\omega \notin A$. Thus $A^c = \{\lim Y_n = 0\}$, whence $P\{\lim Y_n = 0\} = 1$ iff $P\{A\} = 0$ or equivalently

$$P\{A_k\} = 0, \quad k \geq 1.$$

\square

The lemma which follows plays a pivotal role in probability theory in establishing the existence of limits and constitutes one-half the Borel-Cantelli theorem of Section 3.2.

Lemma 2 (Borel–Cantelli Lemma). *If $\{A_n, n \geq 1\}$ is a sequence of events for which $\sum_1^\infty P\{A_n\} < \infty$, then $P\{A_n, i.o.\} = 0$.*

PROOF. Since $\{A_n, i.o.\} = \bigcap_{k=1}^{\infty} \bigcup_{n=k}^{\infty} A_n \subset \bigcup_{n=k}^{\infty} A_n$, all $k \geq 1$, by Theorem 1.5.2

$$P\{A_n, i.o.\} \leq P\left\{\bigcup_{n=k}^{\infty} A_n\right\} \leq \sum_{n=k}^{\infty} P\{A_n\},$$

and so

$$0 \leq P\{A_n, i.o.\} \leq \lim_{k \rightarrow \infty} \sum_{n=k}^{\infty} P\{A_n\} = 0. \quad \square$$

The last two lemmas in conjunction with Theorem 2 yield

Theorem 3 (Borel Strong Law of Large Numbers, 1909). *Let S_n constitute a sequence of binomial r.v.s on some probability space (Ω, \mathcal{F}, P) with p.d.f.s $b(k; n, p)$, $n \geq 1$. Then*

$$P\left\{\lim_{n \rightarrow \infty} \frac{S_n}{n} = p\right\} = 1.$$

PROOF. According to Theorem 2, for every $\varepsilon > 0$

$$\sum_{n=1}^{\infty} P\left\{\left|\frac{S_n}{n} - p\right| > \varepsilon\right\} < \infty,$$

whence the Borel–Cantelli lemma guarantees that

$$P\left\{\left|\frac{S_n}{n} - p\right| > \varepsilon, i.o.\right\} = 0, \quad \varepsilon > 0.$$

Thus, by Lemma 1

$$P\left\{\lim_{n \rightarrow \infty} \left(\frac{S_n}{n} - p\right) = 0\right\} = 1,$$

which is tantamount to that which was to be proved. \square

The existence of such a probability space will follow from Theorems 3.1.1 and 6.4.3.

S. Bernstein ingeniously exploited the binomial distribution and Theorem 1 to prove Weierstrass' approximation theorem, which asserts that every continuous function on $[0, 1]$ can be uniformly approximated by polynomials.

EXAMPLE 1. If f is a continuous function on $[0, 1]$ and the Bernstein polynomials are defined by

$$B_n(p) = \sum_{j=0}^n f\left(\frac{j}{n}\right) \binom{n}{j} p^j (1-p)^{n-j}, \quad p \in [0, 1], \quad (3)$$

then

$$\lim_{n \rightarrow \infty} B_n(p) = f(p) \quad \text{uniformly for } p \in [0, 1]. \quad (4)$$

PROOF. Let S_n be a binomial r.v. with p.d.f. $b(k; n, p)$. Since every continuous function on $[0, 1]$ is bounded and uniformly continuous thereon, $|f(p)| \leq M < \infty$ for $p \in [0, 1]$, and for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(p) - f(p')| < \varepsilon$ if $|p - p'| < \delta$ and $0 \leq p, p' \leq 1$. Then, setting $q = 1 - p$ and $A_n = \{j: |j/n - p| < \delta\}$,

$$\begin{aligned} |B_n(p) - f(p)| &= \left| \sum_{j=0}^n \binom{n}{j} p^j q^{n-j} \left[f\left(\frac{j}{n}\right) - f(p) \right] \right| \\ &\leq \sum_{j=0}^n \binom{n}{j} p^j q^{n-j} \left| f\left(\frac{j}{n}\right) - f(p) \right| \\ &\leq \varepsilon \sum_{A_n} b(j; n, p) + 2M \sum_{A_n^c} b(j; n, p) \\ &\leq \varepsilon + 2M P \left\{ \left| \frac{S_n}{n} - p \right| \geq \delta \right\}. \end{aligned}$$

By (2),

$$P \left\{ \left| \frac{S_n}{n} - p \right| \geq \delta \right\} \leq \frac{p(1-p)}{n\delta^2} \leq \frac{1}{4n\delta^2},$$

and so if $n \geq M(\varepsilon\delta^2)^{-1}$,

$$|B_n(p) - f(p)| \leq \varepsilon + \varepsilon = 2\varepsilon, \quad 0 \leq p \leq 1,$$

yielding (4). \square

If $\{Y_n, n \geq 0\}$ are r.v.s on a probability space (Ω, \mathcal{F}, P) , then $\{Y_n, n \geq 1\}$ is said to **converge in probability** to Y_0 , denoted $Y_n \xrightarrow{P} Y_0$, if

$$\lim_{n \rightarrow \infty} P\{|Y_n - Y_0| \leq \varepsilon\} = 1,$$

all $\varepsilon > 0$. Alternatively, $\{Y_n, n \geq 1\}$ **converges almost certainly** (a.c.) to Y_0 , denoted $Y_n \xrightarrow{a.c.} Y_0$, if $P\{\lim_{n \rightarrow \infty} Y_n = Y_0\} = 1$. Theorems 1 and 3 of this section assert that if $\{S_n, n \geq 1\}$ are binomial r.v.s on (Ω, \mathcal{F}, P) , then both types of convergence hold with $Y_n = S_n/n$ and $Y_0 = p$. A detailed discussion of these concepts is given in Section 3.3, where it is shown that $Y_n \xrightarrow{a.c.} Y_0$ implies $Y_n \xrightarrow{P} Y_0$, the converse being untrue in general. However, the case of a countable sample space Ω is exceptional according to

EXAMPLE 2. If $\{Y_n, n \geq 0\}$ is a sequence of r.v.s on a countable probability space (Ω, \mathcal{F}, P) where $\mathcal{F} = \{\text{all subsets of } \Omega\}$ and $Y_n \xrightarrow{P} Y_0$, then $Y_n \xrightarrow{a.c.} Y_0$.

PROOF. Set $A = \{\omega : \lim_{n \rightarrow \infty} Y_n = Y_0\}$ and suppose that $P\{A\} < 1$ or equivalently $P\{A^c\} > 0$. Since A^c is countable and, moreover, a countable union of null events is a null event, there exists $\omega_0 \in A^c$ with $P\{\omega_0\} = \delta > 0$. Moreover, $\omega_0 \in A^c$ implies that for some $\varepsilon > 0$ and subsequence n_j of the positive integers, $|Y_{n_j}(\omega_0) - Y_0(\omega_0)| > \varepsilon, j \geq 1$, whence

$$P\{|Y_{n_j} - Y_0| > \varepsilon\} \geq P\{\omega_0\} \geq \delta,$$

contradicting $Y_n \xrightarrow{P} Y_0$. \square

EXERCISES 2.2

- Verify that if $np \rightarrow \lambda \in (0, \infty)$, then $\sum_0^\infty k^j b(k; n, p) \rightarrow \sum_0^\infty k^j p(k; \lambda)$, $j = 1, 2$. Hint: Recall Exercise 2.1.2.
 - If $g_m(\lambda) = \sum_{k=0}^\infty (k - \lambda)^m p(k; \lambda)$, show that $g_1(\lambda) = 0$, $g_2(\lambda) = g_3(\lambda) = \lambda$, $g_4(\lambda) = 3\lambda^2 + \lambda$, $g_6(\lambda) = 15\lambda^3 + 25\lambda^2 + \lambda$.
 - (i) Prove a weak law of large numbers where S_n has p.d.f. $p(k; n\lambda)$, $n \geq 1$, that is, $P\{|(S_n/n) - \lambda| > \varepsilon\} = o(1)$, $\varepsilon > 0$. (ii) If $\{X_i, X_n, n \geq 1\}$ are r.v.s with a common d.f. and $n P\{|X_i| > n\} = o(1)$, then $(1/n) \max_{1 \leq i \leq n} |X_i| \xrightarrow{P} 0$.
 - Prove a strong law of large numbers where S_n has p.d.f. $p(k; n\lambda)$, $n \geq 1$, that is, $\lim_{n \rightarrow \infty} (S_n/n) = \lambda$, a.c. Hint: Consider $P\{|S_n - n\lambda|^4 > n^4 \varepsilon^4\}$.
 - Show that the Borel–Cantelli lemma is valid on any measure space $(\Omega, \mathcal{A}, \mu)$, that is, $\sum_1^\infty \mu\{A_n\} < \infty$ implies $\mu(\overline{\lim}_n A_n) = 0$ for $A_n \in \mathcal{A}, n \geq 1$.
 - Let $\{X_{n,j}, j \geq 1, n \geq 1\}$ be a sequence of r.v.s such that $\sum_{j=1}^\infty P\{|X_{n,j}| > \varepsilon\} \xrightarrow{n \rightarrow \infty} 0$, all $\varepsilon > 0$. Prove that $\sup_{j \geq 1} |X_{n,j}| \xrightarrow{P} 0$.
 - Prove that for $0 < p < 1$
- $$\lim_{n \rightarrow \infty} \sum_{k=0}^n \left(\frac{k - np}{\sqrt{npq}} \right)^4 b(k; n, p) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^4 e^{-t^2/2} dt = \lim_{n \rightarrow \infty} \sum_{k=0}^n \left(\frac{k - \lambda}{\sqrt{\lambda}} \right)^4 p(k; \lambda).$$
- A sequence $\{S_n, n \geq 1\}$ of r.v.s is said to **converge completely** (Hsu–Robbins) if for every $\varepsilon > 0$, $\sum P\{|S_n| > \varepsilon\} < \infty$. Prove that if $\{S_n, n \geq 1\}$ converges completely, then $\lim S_n = 0$, a.c.
 - (i) For any r.v. X and constants $a_n \rightarrow 0$, verify that $a_n X \xrightarrow{a.e.} 0$. Hint: It suffices to prove $c_n |X| \xrightarrow{a.e.} 0$, where $c_n = \sup_{j \geq n} |a_j|$. (ii) If $\{X_n, n \geq 1\}$ are r.v.s with identical d.f.s and $\{a_n\}$ are as in (i), then $a_n X_n \xrightarrow{P} 0$.

2.3 Central Limit Theorem for Binomial Random Variables, Large Deviations

To show that the limit of binomial p.d.f.s may be normal an asymptotic estimate of the combinatorial occurring therein is essential and this follows readily from

Lemma 1 (Stirling's Formula). *For every positive integer $n \geq 2$*

$$n! = n^{n+(1/2)} e^{-n+\varepsilon_n} \sqrt{2\pi} \quad \text{where } \frac{1}{12n+1} < \varepsilon_n < \frac{1}{12n}. \quad (1)$$

PROOF. Define

$$a_n = \left(1 + \frac{1}{n}\right)^{n+(1/2)}, \quad b_n = \frac{1}{2n+1}, \quad n \geq 1.$$

Then

$$\begin{aligned} \log a_n &= (n + \frac{1}{2}) \log \frac{n+1}{n} = \frac{1}{2b_n} \log \frac{1+b_n}{1-b_n} = 1 + \frac{b_n^2}{3} + \frac{b_n^4}{5} + \dots \\ &= 1 + \delta_n \text{ (say), where} \end{aligned}$$

$$\begin{aligned} \frac{1}{12n+1} - \frac{1}{12(n+1)+1} &< \frac{b_n^2}{3} < \delta_n < \frac{b_n^2}{3}(1 + b_n^2 + b_n^4 + \dots) \\ &= \frac{b_n^2}{3(1-b_n^2)} = \frac{1}{12n(n+1)}, \end{aligned} \quad (2)$$

so that

$$0 < \sum_{n=1}^{\infty} \delta_n = C < \infty.$$

Therefore,

$$\begin{aligned} \log \frac{(n+1)^{n+(1/2)}}{n!} &= \sum_1^n \log a_j = n + \sum_{j=1}^n \delta_j = n + C - \sum_{j=n+1}^{\infty} \delta_j \\ &= n + C - \varepsilon_{n+1} \text{ (say), where via (2)} \end{aligned}$$

$$\frac{1}{12n+1} < \varepsilon_n < \frac{1}{12n}.$$

Hence, for $n \geq 1$

$$n! = (n+1)^{n+(1/2)} \exp(-n - C + \varepsilon_{n+1}),$$

implying

$$(n+1)! = (n+1)^{n+(3/2)} \exp(-n - C + \varepsilon_{n+1}). \quad (3)$$

Set $K = e^{1-c} > 0$. Then, replacing n by $n - 1$ in (3),

$$n! = n^{n+(1/2)}e^{-n+\varepsilon_n}K, \quad n \geq 2. \quad (4)$$

The identification of K as $\sqrt{2\pi}$ will be made at the end of this section via probabilistic reasoning, thereby completing the proof. \square

Lemma 2 (DeMoivre–Laplace, 1730). *For $n = 1, 2, \dots$ let $k = k_n$ be a non-negative integer and set $x = x_k = (k - np)(npq)^{-1/2}$, where $q = 1 - p$, $0 < p < 1$. If $x = o(n^{1/6})$ and $\varphi(x)$ is the standard normal density, there exist positive constants A, B, C such that*

$$\left| \frac{b(k; n, p)}{(npq)^{-1/2}\varphi(x)} - 1 \right| < \frac{A}{n} + \frac{B|x|^3}{\sqrt{n}} + \frac{C|x|}{\sqrt{n}}. \quad (5)$$

PROOF. Since $x = o(n^{1/6})$, necessarily $(k/n) \rightarrow p$. By Stirling's formula

$$\begin{aligned} b(k; n, p) &= \binom{n}{k} p^k q^{n-k} = \frac{n^{n+(1/2)} \exp(-n + \varepsilon_n) (2\pi)^{-1/2} p^k q^{n-k}}{k^{k+(1/2)} (n - k)^{n-k+(1/2)} \exp(-n + \varepsilon_k + \varepsilon_{n-k})} \\ &= \frac{e^\varepsilon}{\sqrt{2\pi}} \left(\frac{k}{np} \right)^{-k-(1/2)} \left(\frac{n-k}{nq} \right)^{-n+k-(1/2)} (npq)^{-1/2} \end{aligned}$$

where $\varepsilon = \varepsilon_n - \varepsilon_k - \varepsilon_{n-k}$. Since $k/n \rightarrow p$, $\varepsilon = O(n^{-1})$. Now

$$\begin{aligned} \log\{(2\pi npq)^{1/2}b(k; n, p)\} &= \varepsilon - (k + \frac{1}{2})\log \frac{k}{np} - (n - k + \frac{1}{2})\log \frac{n - k}{nq} \\ &= \varepsilon - (np + x\sqrt{npq} + \frac{1}{2})\log \left(1 + x\sqrt{\frac{q}{np}} \right) \\ &\quad - (nq - x\sqrt{npq} + \frac{1}{2})\log \left(1 - x\sqrt{\frac{p}{nq}} \right) \\ &= \varepsilon - (np + x\sqrt{npq} + \frac{1}{2}) \left[x\sqrt{\frac{q}{np}} - \frac{x^2q}{2np} + O\left(\frac{|x|^3}{n^{3/2}}\right) \right] \\ &\quad - (nq - x\sqrt{npq} + \frac{1}{2}) \left[-x\sqrt{\frac{p}{nq}} - \frac{x^2p}{2nq} + O\left(\frac{|x|^3}{n^{3/2}}\right) \right] \\ &= \varepsilon - \left[x\sqrt{npq} + x^2q - \frac{x^2q}{2} + \frac{x}{2}\sqrt{\frac{q}{np}} + O\left(\frac{|x|^3}{n^{1/2}}\right) + O\left(\frac{x^2}{n}\right) \right] \\ &\quad - \left[-x\sqrt{npq} + x^2p - \frac{x^2p}{2} - \frac{x}{2}\sqrt{\frac{p}{nq}} + O\left(\frac{|x|^3}{n^{1/2}}\right) + O\left(\frac{x^2}{n}\right) \right] \\ &= -\frac{x^2}{2} + O\left(\frac{|x|^3}{\sqrt{n}}\right) + O\left(\frac{|x|}{\sqrt{n}}\right) + O\left(\frac{1}{n}\right), \end{aligned}$$

whence

$$\begin{aligned}(npq)^{1/2}b(k; n, p) &= \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2} + O\left(\frac{|x|^3}{\sqrt{n}}\right) + O\left(\frac{|x|}{\sqrt{n}}\right) + O\left(\frac{1}{n}\right)\right) \\ &= \varphi(x)\left[1 + O\left(\frac{|x|^3}{\sqrt{n}}\right) + O\left(\frac{|x|}{\sqrt{n}}\right) + O\left(\frac{1}{n}\right)\right]\end{aligned}$$

yielding (5). \square

The preceding lemma is an example of a **local limit theorem** since it concerns densities, while the forthcoming corollary provides a **global limit theorem** involving d.f.s.

Theorem 1 (DeMoivre–Laplace (Central Limit) Theorem). *If S_n is a sequence of binomial r.v.s with p.d.f.s $b(k; n, p)$, $n \geq 1$, and $\{m_n\}$, $\{M_n\}$ are sequences of nonnegative integers satisfying $m_n \leq M_n$, $n \geq 1$, for which*

$$\frac{(m_n - np)^3}{n^2} \rightarrow 0, \quad \frac{(M_n - np)^3}{n^2} \rightarrow 0 \quad (6)$$

as $n \rightarrow \infty$, then

$$P\{m_n \leq S_n \leq M_n\} \sim \Phi\left(\frac{M_n - np + \frac{1}{2}}{(npq)^{1/2}}\right) - \Phi\left(\frac{m_n - np - \frac{1}{2}}{(npq)^{1/2}}\right). \quad (7)$$

PROOF. For $k = m_n, m_n + 1, \dots, M_n$ let $x_k = (k - np)h_n$, where $h_n = (npq)^{-1/2}$. By Lemma 2

$$P\{m_n \leq S_n \leq M_n\} = \sum_{k=m_n}^{M_n} b(k; n, p) = \sum_{k=m_n}^{M_n} \varphi(x_k) \cdot h_n(1 + o_k(1)), \quad (8)$$

where $o_k(1) \rightarrow 0$ uniformly for $m_n \leq k \leq M_n$ (via (6)). Now

$$\int_{(k-(1/2)-np)h_n}^{(k+(1/2)-np)h_n} \varphi(u)du = h_n \varphi(\zeta_k) = h_n \exp(\frac{1}{2}(x_k^2 - \zeta_k^2))\varphi(x_k),$$

where $x_k - (h_n/2) < \zeta_k < x_k + (h_n/2)$. Hence,

$$h_n \varphi(x_k) = \exp(\frac{1}{2}(\zeta_k^2 - x_k^2))[\Phi(x_k + \frac{1}{2}h_n) - \Phi(x_k - \frac{1}{2}h_n)]. \quad (9)$$

Since $|\zeta_k^2 - x_k^2| = |\zeta_k - x_k| \cdot |\zeta_k + x_k| \leq h_n(|x_k| + (h_n/4))$, hypothesis (6) ensures that $\exp(\frac{1}{2}(\zeta_k^2 - x_k^2)) \rightarrow 1$ uniformly for $m_n \leq k \leq M_n$.

Consequently, from (8) and (9)

$$\begin{aligned}P\{m_n \leq S_n \leq M_n\} &= (1 + o(1)) \sum_{k=m_n}^{M_n} [\Phi(x_k + \frac{1}{2}h_n) - \Phi(x_k - \frac{1}{2}h_n)] \\ &= (1 + o(1))[\Phi(x_{M_n} + \frac{1}{2}h_n) - \Phi(x_{m_n} - \frac{1}{2}h_n)] \\ &\sim \Phi\left(\frac{M_n - np + \frac{1}{2}}{\sqrt{npq}}\right) - \Phi\left(\frac{m_n - np - \frac{1}{2}}{\sqrt{npq}}\right).\end{aligned} \quad \square$$

Note that the second statement in (11) of Corollary 1 asserts that the d.f. of the normalized binomial r.v. S_n^* as defined in (10) tends to the normal d.f. Such a tendency toward normality is a widespread phenomenon and exemplifies a class of theorems known as central limit theorems (Chapter 9).

Corollary 1. If $\{S_n, n \geq 1\}$ is a sequence of binomial r.v.s with p.d.f.s $b(k; n, p)$, $0 < p < 1$, and

$$S_n^* = \frac{S_n - np}{\sqrt{npq}}, \quad (10)$$

then for every pair of finite constants $a < b$ and all x

$$\lim_{n \rightarrow \infty} P\{a \leq S_n^* \leq b\} = \Phi(b) - \Phi(a), \quad \lim_{n \rightarrow \infty} P\{S_n^* < x\} = \Phi(x). \quad (11)$$

PROOF. Let M_n be the largest integer $\leq np + b\sqrt{npq}$ and m_n the smallest integer $\geq np + a\sqrt{npq}$. Then $M_n \geq m_n \geq 0$ for all large n and (6) obtains. Thus, by the DeMoivre–Laplace theorem

$$\begin{aligned} P\{a \leq S_n^* \leq b\} &= P\{m_n \leq S_n \leq M_n\} \\ &\sim \Phi\left(\frac{M_n - np + \frac{1}{2}}{\sqrt{npq}}\right) - \Phi\left(\frac{m_n - np - \frac{1}{2}}{\sqrt{npq}}\right). \end{aligned} \quad (12)$$

Since the right side of (12) approaches $\Phi(b) - \Phi(a)$ as $n \rightarrow \infty$ via the continuity of $\Phi(\cdot)$, the first part of (11) follows. As for the second, for

$$-\infty < a < b < \infty$$

$$\varliminf_{n \rightarrow \infty} P\{S_n^* \leq b\} \geq \lim_{n \rightarrow \infty} P\{a \leq S_n^* \leq b\} = \Phi(b) - \Phi(a),$$

whence, letting $a \rightarrow -\infty$,

$$\varliminf_{n \rightarrow \infty} P\{S_n^* \leq b\} \geq \Phi(b).$$

Thus, for $a < b$,

$$\varliminf_{n \rightarrow \infty} P\{S_n^* < b\} \geq \varliminf_{n \rightarrow \infty} P\{S_n^* \leq a\} \geq \Phi(a),$$

yielding as $a \rightarrow b$

$$\varliminf_{n \rightarrow \infty} P\{S_n^* < b\} \geq \Phi(b).$$

Similarly,

$$1 - \varlimsup_{n \rightarrow \infty} P\{S_n^* < a\} = \varlimsup_{n \rightarrow \infty} P\{S_n^* \geq a\} \geq \varlimsup_{n \rightarrow \infty} P\{a \leq S_n^* \leq b\} = \Phi(b) - \Phi(a),$$

yielding as $b \rightarrow \infty$

$$\varlimsup_{n \rightarrow \infty} P\{S_n^* < a\} \leq \Phi(a),$$

and the second part of (11) follows. \square

As an application of Corollary 1, the constant K occurring in the proof of Stirling's formula will be shown to equal $(2\pi)^{1/2}$: if (4) rather than (1) is utilized in the proof of the prior theorem, then setting $a = -b$, (11) becomes

$$\lim_{n \rightarrow \infty} P\{|S_n^*| \leq b\} = \frac{1}{K} \int_{-b}^b e^{-x^2/2} dx, \quad (13)$$

which readily implies $K \geq (2\pi)^{1/2}$. On the other hand,

$$\begin{aligned} P\{|S_n^*| > b\} &\leq \frac{1}{b^2} \sum_{|k-np|>b\sqrt{npq}} \binom{k-np}{\sqrt{npq}}^2 \binom{n}{k} p^k q^{n-k} \\ &\leq \frac{1}{b^2} \sum_{k=0}^n \frac{(k-np)^2}{npq} \binom{n}{k} p^k q^{n-k} = \frac{1}{b^2} \end{aligned}$$

via (1) of Section 2. Hence,

$$\begin{aligned} 1 &= P\{|S_n^*| \leq b\} + P\{|S_n^*| > b\} \leq P\{|S_n^*| \leq b\} + \frac{1}{b^2} \\ &\xrightarrow{n \rightarrow \infty} \frac{1}{K} \int_{-b}^b e^{-x^2/2} dx + \frac{1}{b^2} \xrightarrow{b \rightarrow \infty} \frac{(2\pi)^{1/2}}{K} \end{aligned}$$

yielding $K \leq (2\pi)^{1/2}$. Consequently, $K = (2\pi)^{1/2}$. \square

As will become increasingly apparent, an important characteristic of any d.f. $F(x)$ is the manner in which it increases to one and decreases to zero—in other words, the order of magnitude of its **tails** $1 - F(x)$ and $F(-x)$ as $x \rightarrow +\infty$. In the case of a normal r.v., symmetry of the distribution entails identical behavior of the **upper and lower tails**.

Lemma 3. For all $x > 0$,

$$\frac{x}{1+x^2} e^{-x^2/2} < \int_x^\infty e^{-u^2/2} du < \frac{1}{x} e^{-x^2/2}.$$

PROOF. For $x > 0$

$$\frac{1}{x^2} \int_x^\infty e^{-u^2/2} du > \int_x^\infty \frac{1}{u^2} e^{-u^2/2} du = \frac{1}{x} e^{-x^2/2} - \int_x^\infty e^{-u^2/2} du$$

whence

$$\frac{1}{x} e^{-x^2/2} = \int_x^\infty \left(1 + \frac{1}{u^2}\right) e^{-u^2/2} du > \int_x^\infty e^{-u^2/2} du > \frac{x}{1+x^2} e^{-x^2/2}. \quad \square$$

Corollary 2. For the standard normal distribution $\Phi(x)$,

$$1 - \Phi(x) \sim \frac{1}{x(2\pi)^{1/2}} e^{-x^2/2}, \quad \text{as } x \rightarrow \infty,$$

Corollary 1 ensures that the tails of the distribution of the normalized binomial r.v. $S_n^* = (S_n - np)/\sqrt{npq}$, say $1 - F_n^*(x)$ and $F_n^*(-x)$, tend to $1 - \Phi(x)$ and $\Phi(-x)$ respectively as $n \rightarrow \infty$. The next assertion is that this remains true for $x = x_n \rightarrow \infty$ provided the approach of x_n to ∞ is sufficiently slow. This is the prototype of what is called a large deviation theorem.

Theorem 2. If S_n is a sequence of binomial r.v.s with p.d.f.s $b(k; n, p)$, $n \geq 1$, and $\{a_n, n \geq 1\}$ is a sequence of real numbers with $a_n \rightarrow \infty$, $a_n = o(n^{1/6})$ as $n \rightarrow \infty$, then setting $S_n^* = (S_n - np)/(npq)^{1/2}$ as earlier,

$$\Pr\{S_n^* \geq a_n\} \sim 1 - \Phi(a_n) \sim \frac{1}{a_n} \varphi(a_n). \quad (14)$$

PROOF. For simplicity suppose $a_n > 1$, $n \geq 1$, and define M_n to be the largest integer $\leq np + (a_n + \log a_n)(npq)^{1/2}$ and m_n the largest integer $\leq np + a_n(npq)^{1/2}$. Then

$$x_{M_n} = \frac{M_n - np}{(npq)^{1/2}} = a_n + \log a_n + O(n^{-1/2}), \quad x_{m_n} = a_n + O(n^{-1/2})$$

whence the DeMoivre–Laplace theorem is applicable, yielding

$$\begin{aligned} \Pr\{m_n \leq S_n \leq M_n\} &\sim \Phi\left(\frac{M_n + \frac{1}{2} - np}{(npq)^{1/2}}\right) - \Phi\left(\frac{m_n - \frac{1}{2} - np}{(npq)^{1/2}}\right) \\ &= \Phi(a_n + \log a_n + O(n^{-1/2})) - \Phi(a_n + O(n^{-1/2})). \end{aligned}$$

By Lemma 3 and the fact that $a_n^3 = o(n^{1/2})$,

$$\begin{aligned} 1 - \Phi(a_n + O(n^{-1/2})) &\sim \frac{(2\pi)^{-1/2}}{(a_n + O(n^{-1/2}))} \exp\left\{-\frac{1}{2}(a_n + O(n^{-1/2}))^2\right\} \sim \frac{1}{a_n} \varphi(a_n) \\ 1 - \Phi(a_n + \log a_n + O(n^{-1/2})) &\sim \frac{(2\pi)^{-1/2}}{a_n + \log a_n + O(n^{-1/2})} \\ &\quad \times \exp\{-\frac{1}{2}(a_n + \log a_n + O(n^{-1/2}))^2\} = o\left(\frac{\varphi(a_n)}{a_n}\right). \end{aligned}$$

Hence,

$$\Pr\{m_n \leq S_n \leq M_n\} \sim \frac{1}{a_n} \varphi(a_n).$$

Now, for $k = M_n + 1, \dots, n$

$$\frac{\Pr\{S_n = k\}}{\Pr\{S_n = k - 1\}} = \frac{n - k + 1}{k} \frac{p}{q} \leq \frac{n - M_n}{M_n + 1} \frac{p}{q},$$

whence

$$\begin{aligned} \mathbb{P}\{S_n \geq M_n\} &= \sum_{j=0}^{n-M_n} \mathbb{P}\{S_n = M_n + j\} \leq \sum_{j=0}^{n-M_n} \left(\frac{n - M_n}{M_n + 1} \frac{p}{q}\right)^j \mathbb{P}\{S_n = M_n\} \\ &\leq \frac{(M_n + 1)q}{M_n + q - np} \mathbb{P}\{S_n = M_n\} \leq \frac{n}{M_n - np} \mathbb{P}\{S_n = M_n\}. \end{aligned}$$

Since $M_n - np = x_{M_n}(npq)^{1/2}$, by Lemma 2

$$\begin{aligned} \frac{n}{M_n - np} \mathbb{P}\{S_n = M_n\} &\sim \frac{\varphi(x_{M_n})}{p \cdot q x_{M_n}} \\ &= \frac{(2\pi p^2 q^2)^{-1/2}}{[a_n + \log a_n + O(n^{-1/2})]} \exp(-\frac{1}{2}[a_n + \log a_n + O(n^{-1/2})]^2) \\ &= o\left(\frac{\varphi(a_n)}{a_n}\right). \end{aligned}$$

Analogously,

$$\mathbb{P}\{S_n = m_n\} \sim \frac{\varphi(a_n)}{(npq)^{1/2}} = o\left(\frac{\varphi(a_n)}{a_n}\right). \quad (15)$$

Thus,

$$\mathbb{P}\{S_n \geq M_n\} = o\left(\frac{\varphi(a_n)}{a_n}\right)$$

and

$$\begin{aligned} \mathbb{P}\{S_n \geq m_n\} &= \mathbb{P}\{m_n \leq S_n \leq M_n\} + \mathbb{P}\{S_n > M_n\} \\ &\sim \frac{\varphi(a_n)}{a_n} + o\left(\frac{\varphi(a_n)}{a_n}\right) \sim \frac{1}{a_n} \varphi(a_n). \end{aligned} \quad (16)$$

Consequently, (14) follows from (15), (16) and

$$\mathbb{P}\{S_n > m_n\} \leq \mathbb{P}\{S_n^* \geq a_n\} \leq \mathbb{P}\{S_n \geq m_n\}. \quad \square$$

EXERCISES 2.3

1. Show that the Bernoulli weak law of large numbers is consequence of the DeMoivre–Laplace theorem.
2. Prove that as $x \rightarrow \infty$.

$$\begin{aligned} \int_x^\infty u \varphi(u) du &\sim x \int_x^\infty \varphi(u) du, \\ \frac{1 - \Phi(x + (a/x))}{1 - \Phi(x)} &\rightarrow e^{-a}. \end{aligned}$$

3. Use Lemma 2 to show that $b(n; 2n, \frac{1}{2}) \sim (n\pi)^{-1/2}$.

4. (Normal approximation to the Poisson distribution.) If X_λ is a Poisson r.v. with parameter λ , prove that for any $-\infty < a < b < \infty$

$$\lim_{\lambda \rightarrow \infty} P\left[a < \frac{X_\lambda - \lambda}{\sqrt{\lambda}} < b\right] = \Phi(b) - \Phi(a)$$

by first showing for

$$z = \frac{k - \lambda}{\sqrt{\lambda}} = o(\lambda^{1/6}) \quad \text{that} \quad \left| \frac{p(k; \lambda)}{\lambda^{-1/2} \phi(z)} - 1 \right| \leq \frac{A}{\lambda} + \frac{B|z|^3}{\sqrt{\lambda}} + \frac{C|z|}{\sqrt{\lambda}}.$$

Obtain a large deviation result akin to Theorem 2.

5. Let S_n be binomial r.v.s with p.d.f. $b(k; n, p)$, $0 < p < 1$, and $S_n^* = (S_n - np)/\sqrt{npq}$. Prove that for every $\alpha > 1$,

$$P\{|S_n^*| > (2\alpha \log n)^{1/2}, \text{i.o.}\} = 0.$$

6. Utilize Exercise 5 to prove that for S_n as defined there and for every $\beta > \frac{1}{2}$,

$$\lim_{n \rightarrow \infty} \frac{S_n - np}{n^\beta} = 0 \text{ a.c.}$$

(If $\beta = 1$, this yields the Borel strong law of large numbers.)

7. Prove for X_λ as in Exercise 4 that $P\{X_\lambda \leq \lambda\} > c_\lambda$, $\lambda > 0$, where $c_\lambda = \frac{1}{2}$ for integer λ and $c_\lambda = e^{-1}$ otherwise. Hint: $A_{n,\lambda} \equiv \sum_{j=0}^n (\lambda^j/j!) e^{-\lambda} = P\{X_\lambda \leq \lambda\}$ for $n \leq \lambda < n+1$, and $A_{n,n} - A_{n,n+1} = \int_n^{n+1} (\lambda^x e^{-\lambda}/x!) d\lambda > A_{n+1,n+1} - A_{n,n+1}$, implying $A_{n,n} > A_{n+1,n+1}$. Also $A_{n,n} - A_{n,n+1} < A_{n,n} - A_{n-1,n}$ and by Exercise 4 $P\{X_\lambda \leq \lambda\} \rightarrow \frac{1}{2}$ as $\lambda \rightarrow \infty$.

8. For $q > 0$ and $k = 0, 1, \dots, n$, let

$$e_k = b\left(k; n, \frac{1}{q+1}\right) = \binom{n}{k} (q+1)^{-k} q^{n-k}, \quad c = \frac{q+1}{2q},$$

$$N = \left[\frac{n+1}{q+1} \right], \quad h = k - N.$$

Prove that if $\frac{1}{2} < \alpha < \frac{3}{2}$, then

$$\sum_{|h| > n^\alpha} e_h = O(\exp(-n^\eta)) \tag{i}$$

where $\eta < 2\alpha - 1$, and that for $|h| \leq n^\alpha$,

$$e_h = \left(\frac{c}{\pi N} \right)^{1/2} e^{-ch^2/N} \left\{ 1 + O\left(\frac{|h|+1}{n}\right) + O\left(\frac{|h|^3}{n^2}\right) \right\}. \tag{ii}$$

(Hint: For (i) apply Theorem 2, and for (ii) apply Lemma 2.)

9. (Renyi) Let S_n be a binomial r.v. with p.d.f. $b(k; n, p)$, where $0 < p < 1$. If $[np]$ is the largest integer $\leq np$ and j_n, k_n are integers such that $j_n = O(n^\alpha)$, $k_n = O(n^\alpha)$ for some α in $(0, \frac{2}{3})$, and $k_n^2 - j_n^2 = o(n)$, prove that

$$\lim_{n \rightarrow \infty} \frac{P\{S_n = k_n + [np]\}}{P\{S_n = j_n + [np]\}} = 1.$$

Hint: Apply Exercise 8(ii).

10. If $0 < \sigma \leq 1$ and $-\infty < a < \infty$ show that $\sup_x |\Phi(x\sigma^{-1} - a) - \Phi(x)| \leq (2\pi)^{-1/2} [\sigma^{-1} - 1 + |a|]$. Hint: Add and subtract $\Phi(x\sigma^{-1})$, and use the mean value theorem.

References

- J. Bernoulli, *Ars Conjectandi*, Basel, 1713.
- S. Bernstein, "Demonstration du théorème de Weierstrass fondée sur le calcul des probabilités," *Soob. Charkov. Mat. Obs.* **13** (1912), 1–2.
- É. Borel, "Sur les probabilités dénombrables et leurs applications arithmétiques," *Rend. Circ. Mat. Palermo* **27** (1909), 247–271.
- W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd ed., Wiley, New York, 1950.
- B. de Finetti, "La prévision, ses lois logiques, ses sources subjectives," *Annales de l'Institut Henri Poincaré* **7** (1937), 1–68.
- J. Haag, "Sur un problème général de probabilités et ses diverses applications," *Proc. Inst. Congr. Math., Toronto, 1924, 1928*, 629–674.
- G. H. Hardy, *Divergent Series*, Clarendon Press, Oxford, 1949.
- P. L. Hsu and H. Robbins, "Complete convergence and the law of large numbers," *Proc. Nat. Acad. Sci. U.S.A.* **33** (1947), 25–31.
- P. S. Laplace, *Théorie analytique de probabilités*, 1812 [Vol. 7 in *Oeuvres complètes de Laplace*, Gauthier-Villars, Paris, 1886].
- A. de Moivre, *The Doctrine of Chances*, 1718; 3rd ed., London, 1756.
- S. D. Poisson, *Recherches sur la probabilité des juggements*, Paris, 1837.
- A. Renyi, *Foundations of Probability*, Holden-Day, San Francisco, 1970.
- H. Robbins, "A remark on Stirling's formula," *Amer. Math. Monthly* **62** (1955), 26–29.
- J. Stirling, *Methodus Differentialis*, London, 1730.
- H. Teicher, "An inequality on Poisson probabilities," *Ann. Math. Stat.* **26** (1955), 147–149.

3 Independence

Independence may be considered the single most important concept in probability theory, demarcating the latter from measure theory and fostering an independent development. In the course of this evolution, probability theory has been fortified by its links with the real world, and indeed the definition of independence is the abstract counterpart of a highly intuitive and empirical notion. Independence of random variables $\{X_i\}$, the definition of which involves the events of $\sigma(X_i)$, will be shown in Section 2 to concern only the joint distribution functions.

3.1 Independence, Random Allocation of Balls into Cells

Definition. If (Ω, \mathcal{F}, P) is a probability space and T a nonempty index set, classes \mathcal{G}_t of events, $t \in T$, are termed **independent** if for each $m = 2, 3, \dots$, each choice of distinct $t_j \in T$, and events $A_j \in \mathcal{G}_{t_j}$, $1 \leq j \leq m$,

$$P\{A_1 \cdot A_2 \cdots A_m\} = P\{A_1\} \cdot P\{A_2\} \cdots P\{A_m\}. \quad (1)$$

Events A_t , $t \in T$, are called independent if the one-element classes $\mathcal{G}_t = \{A_t\}$, $t \in T$, are independent.

Clearly, nonempty subclasses of independent classes are likewise independent classes. Conversely, if for every nonempty finite subset $T_1 \subset T$ the classes \mathcal{G}_t , $t \in T_1$, are independent, then so are the classes \mathcal{G}_t , $t \in T$. It may be noted that the validity of (1) for some fixed integer $m > 2$ is not sufficient to guarantee independence of the events A_1, A_2, \dots, A_m (Exercise 4).

On the other hand, it is easily verified via (14) of Lemma 2.1.2 that $A_t, t \in T$, are independent events iff the classes $\mathcal{G}_t = \{\emptyset, \Omega, A_t, A_t^c\}$ are independent.

Definition. $\{X_n, n \geq 1\}$ are termed **independent random variables** if the classes $\mathcal{G}_n = \sigma(X_n), n \geq 1$, are independent. More generally, **stochastic processes**, i.e., families of r.v.s $\{X_t^{(n)}, t \in T_n\}, n \geq 1$, are independent (of one another) if the classes $\mathcal{G}_n = \sigma(X_t^{(n)}, t \in T_n), n \geq 1$, are independent.

Random variables which are not independent are generally referred to as **dependent**. Clearly, subsets of independent families or of independent random variables are themselves independent. Of course, independence of the families (or random vectors) (X_1, X_2) and (X_3, X_4) postulates nothing about the independence or dependence of X_1 and X_2 . Note that random variables X and Y are independent if and only if for all $A \in \sigma(X), B \in \sigma(Y)$

$$P\{A \cdot B\} = P\{A\} \cdot P\{B\}.$$

A sequence $\{X_n, n \geq 1\}$ (or the random variables comprising this sequence) is called independent, identically distributed (abbreviated i.i.d.) if $X_n, n \geq 1$, are independent and their distribution functions are identical.

If $\{X_n, n \geq 1\}$ are random variables with $P\{X_n \in A_n\} = 1$, where $A_n \subset A = \{a_1, a_2, \dots\}, n \geq 1$, define

$$p_{1, \dots, n}(a_1, \dots, a_n) = P\{X_1 = a_1, \dots, X_n = a_n\}. \quad (2)$$

Then $p_{1, \dots, n}(a_1, \dots, a_n)$ is called the **joint probability density function** of X_1, \dots, X_n and the latter are termed discrete r.v.s. It is not difficult to ascertain that the discrete random variables X_1, \dots, X_n are independent iff for every choice of (a_1, \dots, a_n) in $A \times A \times \dots \times A$

$$p_{1, \dots, n}(a_1, \dots, a_n) = p_1(a_1) \cdots p_n(a_n), \quad (3)$$

where p_j is the (one-dimensional) probability density function of X_j , $1 \leq j \leq n$. It follows from (3) that if the discrete random variables X_1, \dots, X_n are independent then for all real $\lambda_1, \dots, \lambda_n$

$$P\{X_1 < \lambda_1, \dots, X_n < \lambda_n\} = P\{X_1 < \lambda_1\} \cdots P\{X_n < \lambda_n\}, \quad (4)$$

and the converse is also true. Even without the proviso of discreteness (4) is still equivalent to independence, but the proof is more involved and hence deferred to Section 2. This condition may be rephrased in terms of the joint and one-dimensional d.f.s, namely, for all real $\lambda_i, 1 \leq i \leq n$,

$$F_{X_1, \dots, X_n}(\lambda_1, \dots, \lambda_n) = \prod_{i=1}^n F_{X_i}(\lambda_i). \quad (5)$$

If r.v.s X_1 and X_2 on (Ω, \mathcal{F}, P) are finite everywhere or even if merely $A_1 = \{X_1 = \infty, X_2 = -\infty\} = \emptyset, A_2 = \{X_1 = -\infty, X_2 = \infty\} = \emptyset$, then the definition of their sum is standard, i.e., $(X_1 + X_2)(\omega) = X_1(\omega) + X_2(\omega)$.

However, if $A = A_1 \cup A_2 \neq \emptyset$, then $X_1(\omega) + X_2(\omega)$ is undefined on A . For definiteness, set $(X_1 + X_2)(\omega) = 0$, $\omega \in A$. Since $P\{|X_i| = \infty\} = 0$, $i = 1, 2$, $P\{A\} = 0$ and so $(X_1 + X_2)(\omega) = X_1(\omega) + X_2(\omega)$, a.c. Hence, for any r.v. X_3 on (Ω, \mathcal{F}, P) ,

$$X_1 + X_2 + X_3 = (X_1 + X_2) + X_3 = X_1 + (X_2 + X_3), \quad \text{a.c.}$$

Unless the contrary is stated, any subsequent relationship among r.v.s will be interpreted in the a.c. sense. Since $f(t_1, t_2) = t_1 + t_2$ is a Borel function on (R^2, \mathcal{B}^2) (with the convention that $\infty + (-\infty) = (-\infty) + \infty = 0$), the sum $X_1 + X_2$ is a r.v. by Theorem 1.4.4. By induction, the sum $\sum_1^n X_i$ of n r.v.s is a r.v. for every $n = 1, 2, \dots$.

Definition. Random variables X_j , $1 \leq j \leq n \leq \infty$, are called **Bernoulli trials with success probability p** if X_j , $1 \leq j \leq n \leq \infty$, are i.i.d. with $P\{X_j = 1\} = p \in (0, 1)$ and $P\{X_j = 0\} = q = 1 - p$.

The event $\{X_j = 1\}$ is frequently interpreted as a success on the j th trial (and concomitantly $\{X_j = 0\}$ as a failure on the j th trial) so that it is natural to refer to p as the **success probability**. Then $S_n = \sum_1^n X_i$ is the number of successes in the first n trials.

Theorem 1. If $\{X_n, n \geq 1\}$ are i.i.d. r.v.s on a probability space (Ω, \mathcal{F}, P) with $P\{X_1 = 1\} = p \in (0, 1)$ and $P\{X_1 = 0\} = q = 1 - p$, then $S_n = \sum_1^n X_i$, $n \geq 1$, is a sequence of binomial r.v.s on (Ω, \mathcal{F}, P) with p.d.f.s $b(k; n, p)$, $n \geq 1$, and

$$\lim_{n \rightarrow \infty} P\left\{\frac{S_n - np}{(npq)^{1/2}} < x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du, \quad x \in (-\infty, \infty), \quad (6)$$

$$\frac{S_n}{n} \xrightarrow{\text{a.c.}} p. \quad (7)$$

PROOF. Clearly, $\{S_n, n \geq 1\}$ are r.v.s, and since $P\{X_n = x\} = p^x q^{n-x}$ for $x = 0$ or 1 , by independence for $k = 0, 1, \dots, n$

$$P\{S_n = k\} = \sum P\{X_1 = x_1, \dots, X_n = x_n\} = \sum p^{\sum_i x_i} q^{n - \sum_i x_i},$$

where the summation is over all $\binom{n}{k}$ choices of x_1, \dots, x_n such that $\sum_i x_i = k$ and $x_i = 0$ or 1 for $i = 1, \dots, n$. Hence,

$$P\{S_n = k\} = \sum p^k q^{n-k} = \binom{n}{k} p^k q^{n-k}.$$

Now, (6) and (7) follow from Corollary 2.3.1 and Theorem 2.2.3. \square

To say that r balls have been cast **at random** into n cells will mean that every possible arrangement of these r balls into the n cells has the same

probability, i.e., is "equally likely." A mathematical model or probability space describing such an experiment is concocted by taking

$$\begin{aligned}\Omega &= \{\omega: \omega = (\omega_1, \dots, \omega_r), \omega_j = 1, 2, \dots, n, 1 \leq j \leq r\} \\ &= (1, \dots, n) \times \cdots \times (1, \dots, n),\end{aligned}$$

\mathcal{F} = class of all subsets of Ω ,

$$P\{A\} = \sum_{\omega \in A} n^{-r} = mn^{-r}, \quad A \in \mathcal{F},$$

where m = number of elements in A . Define X_1, \dots, X_r to be the **coordinate random variables** of Ω , i.e., $X_j(\omega) = \omega_j, 1 \leq j \leq r$. In words, X_j is the number of the cell containing the j th ball. Then X_1, \dots, X_r are discrete r.v.s with joint p.d.f.

$$p_{1, \dots, r}(k_1, \dots, k_r) = P\{X_1 = k_1, \dots, X_r = k_r\} = n^{-r},$$

$$k_i = 1, \dots, n, 1 \leq i \leq r.$$

Moreover, if p_i denotes the p.d.f. of $X_i, 1 \leq i \leq r$, for $k_i = 1, \dots, n, 1 \leq i \leq r$,

$$p_i(k_i) = P\{X_i = k_i\} = \sum_{k_j=1, j \neq i}^n p_{1, \dots, r}(k_1, \dots, k_r) = \frac{n^{r-1}}{n^r} = \frac{1}{n}, \quad (8)$$

whence for all choices of k_1, \dots, k_r

$$p_{1, \dots, r}(k_1, \dots, k_r) = \prod_{i=1}^r p_i(k_i).$$

According to (3), X_1, \dots, X_r are independent and, taking cognizance of (8), $X_j, 1 \leq j \leq r$, are i.i.d. random variables. Thus, a random allocation of r balls into n cells is tantamount to considering i.i.d. r.v.s X_1, \dots, X_r with $P\{X_1 = k\} = 1/n, k = 1, \dots, n$. Let

$$A_i = \{X_1 \neq i, \dots, X_r \neq i\} = \{\omega: \text{cell } i \text{ is empty}\}.$$

Then in view of independence and (8), for $1 \leq i \leq n$

$$P\{A_i\} = \prod_{j=1}^r P\{X_j \neq i\} = \left(1 - \frac{1}{n}\right)^r;$$

for $1 \leq i_1 < i_2 \leq n$

$$\begin{aligned}P\{A_{i_1} A_{i_2}\} &= P\{X_j \neq i_1 \text{ or } i_2, 1 \leq j \leq r\} \\ &= (1 - P\{X_j = i_1\} - P\{X_j = i_2\})^r = \left(1 - \frac{2}{n}\right)^r,\end{aligned}$$

and, in general, for $1 \leq m \leq n$ and $1 \leq i_1 < \cdots < i_m \leq n$

$$P\{A_{i_1} \cdots A_{i_m}\} = P\{X_j \neq i_1 \text{ or } \cdots \text{ or } i_m, 1 \leq j \leq r\} = \left(1 - \frac{m}{n}\right)^r.$$

In other words, A_1, \dots, A_n are interchangeable events and consequently if

$$p_m(r, n) = P\{\text{exactly } m \text{ cells are vacant}\}, \quad 0 \leq m \leq n,$$

it follows from Corollary 2.1.4 that

$$p_m(r, n) = \sum_{j=m}^n (-1)^{j-m} \binom{j}{m} \binom{n}{j} \left(1 - \frac{j}{n}\right)^r, \quad 0 \leq m \leq n. \quad (9)$$

Theorem 2 (von Mises). *If r_n balls are cast at random into n cells so that each arrangement has probability n^{-r_n} , where*

$$ne^{-r_n/n} = \lambda_n \leq \lambda_0 < \infty, \quad n \geq 1, \quad (10)$$

then as $n \rightarrow \infty$ the probability $p_m(r_n, n)$ of exactly m vacant cells satisfies

$$p_m(r_n, n) - \frac{\lambda_n^m e^{-\lambda_n}}{m!} = o(1) \quad \text{for } m = 0, 1, \dots \quad (11)$$

In particular, if $\lambda_n = \lambda + o(1)$, then $\{p_m(r_n, n), m \geq 0\}$ tends as $n \rightarrow \infty$ to the Poisson p.d.f. with parameter λ .

PROOF. Set $p_j^{(n)} = (1 - (j/n))^{r_n}$ and rewrite (9) as

$$p_m(r_n, n) = \sum_{j=0}^{n-m} (-1)^j \binom{n}{m+j} \binom{m+j}{m} p_{m+j}^{(n)}.$$

Since $1 - x \leq e^{-x}$ for $0 < x < 1$,

$$n(n-1) \cdots (n-k+1) p_k^{(n)} \leq n^k e^{-kr_n/n} = \lambda_n^k \leq \lambda_0^k$$

by (10). Moreover, for any fixed $k \geq 1$, if $r_n \leq n^{3/2}$, then

$$\begin{aligned} n^k p_k^{(n)} &= n^k \left(1 - \frac{k}{n}\right)^{r_n} = n^k [e^{-k/n} + O(n^{-2})]^{r_n} \\ &= n^k e^{-kr_n/n} [1 + O(n^{-2} e^{k/n})]^{r_n} = \lambda_n^k [1 + o(1)] = \lambda_n^k + o(1) \end{aligned}$$

by (10), while if $r_n > n^{3/2}$, then

$$\begin{aligned} n^k p_k^{(n)} &= n^k \left(1 - \frac{k}{n}\right)^{r_n} \leq n^k \left[\left(1 - \frac{k}{n}\right)^n\right]^{\sqrt{n}} \leq n^k e^{-k\sqrt{n}} = o(1), \\ \lambda_n^k &= n^k e^{-kr_n/n} \leq n^k e^{-k\sqrt{n}} = o(1), \\ n^k p_k^{(n)} &= \lambda_n^k + o(1), \end{aligned}$$

and so the desired conclusion (11) follows from Theorem 2.1.2. \square

In Chapter 9 it will be shown that the d.f. of the number of empty cells tends to the normal distribution when $r_n/n \rightarrow \alpha > 0$ and even in certain cases if $r_n/n \rightarrow 0$ or ∞ .

Definition. The **conditional probability** of an event A given an event B of positive probability, denoted $P\{A|B\}$, is defined by

$$P\{A|B\} = \frac{P\{AB\}}{P\{B\}}. \quad (12)$$

Clearly, when the events A and B are independent, $P\{A|B\} = P\{A\}$.

Theorem 3. If $\{A_n, n \geq 1\}$ is a sequence of events with $\sum_{n=1}^{\infty} P\{A_n\} = \infty$, then

$$\limsup_{n \rightarrow \infty} P\left\{ \bigcup_{j=n+1}^{\infty} A_j | A_n \right\} = 1. \quad (13)$$

PROOF. Evidently, $P\{A_n\} > 0$ for infinitely many values of n , and it may be supposed without loss of generality that $P\{A_n\} > 0, n \geq 1$. Since

$$1 \geq P\left\{ \bigcup_{n=1}^{\infty} A_n \right\} \geq \sum_{n=1}^{\infty} P\left\{ A_n \cap \bigcap_{j=n+1}^{\infty} A_j^c \right\} = \sum_{n=1}^{\infty} P\{A_n\} \left[1 - P\left\{ \bigcup_{j=n+1}^{\infty} A_j | A_n \right\} \right],$$

divergence of the series requires

$$\liminf_{n \rightarrow \infty} \left[1 - P\left\{ \bigcup_{j=n+1}^{\infty} A_j | A_n \right\} \right] = 0,$$

which is tantamount to (13).

EXERCISES 3.1

- Events $A_n, n \geq 1$, are independent iff their indicator functions $I_{A_n}, n \geq 1$, are independent r.v.s iff $\mathcal{G}_n = \{\emptyset, \Omega, A_n, A_n^c\}, n \geq 1$, are independent classes.
- If the classes of events \mathcal{A}_n and \mathcal{D} are independent, $n \geq 1$, so are $\bigcup_{n=1}^{\infty} \mathcal{A}_n$ and \mathcal{D} .
- (i) Any r.v. X is independent of a degenerate r.v. Y . (ii) Two disjoint events are independent iff one of them has probability zero. (iii) If $P\{X = \pm 1, Y = \pm 1\} = \frac{1}{4}$ for all four pairs of signs, then X and Y are independent r.v.s.
- Let $\Omega = \{\omega_0, \omega_1, \omega_2, \omega_3\}$, $\mathcal{F} = \{A : A \subset \Omega\}$, $P\{\omega_i\} = \frac{1}{4}, 0 \leq i \leq 3$, and otherwise P is defined by additivity. If $A_i = \{\omega_0, \omega_i\}, 1 \leq i \leq 3$, then each of $\{A_1, A_2\}, \{A_1, A_3\}, \{A_2, A_3\}$ is a pair of independent events but the events A_1, A_2, A_3 are not independent. On the other hand, if $B_1 = \{\omega_0\}, B_2 = A_2, B_3 = \emptyset$, then (1) obtains for $m = 3$ but the events B_1, B_2 are not independent.
- (i) If X_1, \dots, X_n are independent r.v.s and $g_i, 1 \leq i \leq n$, are finite Borel functions on $(-\infty, \infty)$, then $Y_i = g_i(X_i), 1 \leq i \leq n$ are independent r.v.s. In particular, $-X_1, \dots, -X_n$ are independent r.v.s. (ii) If $B_i, 1 \leq i \leq n$, are linear Borel sets and $Y_i = X_i I_{\{X_i \in B_i\}}$, where I_A is the indicator function of the set A , show that $\{Y_i, 1 \leq i \leq n\}$ are independent r.v.s.
- If $\{X_n, n \geq 1\}$ are i.i.d. r.v.s with $P\{X_1 = 0\} < 1$ and $S_n = \sum_1^n X_i, n \geq 1$, then for every $c > 0$ there exists an integer $n = n_c$ such that $P\{|S_n| > c\} > 0$.

7. If X, Y, Z are r.v.s on (Ω, \mathcal{F}, P) and \circ signifies the relation of independence, prove or give counter examples for:
- $X \circ Y$ iff $X^2 \circ Y^2$.
 - $X \circ Y, X \circ Z$ iff $X \circ (Y + Z)$.
 - $X \circ Y, Y \circ Z$ imply $X \circ Z$.
 - $X \circ (Y, Z), Y \circ Z$ imply X, Y, Z independent r.v.s.
8. If X and Y are independent r.v.s and $X + Y$ is degenerate, then both X and Y are degenerate.
9. Let $\{X_n, n \geq 1\}$ be i.i.d. r.v.s with $P\{X_i = X_j\} = 0$ for $i \neq j$. If $Y_i = \sum_{j=1}^i I_{\{X_j \leq x_i\}}$, $i \geq 1$, prove that $\{Y_i, i \geq 1\}$ are independent r.v.s with $P\{Y_i = j\} = 1/i, j = 1, 2, \dots, i$, and $i \geq 1$.
10. In Bernoulli trials with success probability p , let Y be the waiting time beyond the 1st trial until the first success occurs, i.e., $Y = j$ iff $X_{j+1} = 1, X_i = 0, i \leq j$, where $\{X_1, \dots, X_j\}$ are i.i.d. with $P\{X_i = 1\} = p = 1 - P\{X_i = 0\}$. Find $P\{Y = j\}$, $j = 0, 1, \dots$, which for obvious reasons is called the **geometric distribution**. If Y_1, \dots, Y_r are i.i.d. r.v.s with a geometric distribution, find the p.d.f. of $S_r = \sum_{i=1}^r Y_i$, known as the **negative binomial distribution** (with parameter r). Hint: S_r may be envisaged as the waiting time beyond the r th trial until the r th success.
11. If Y and Z are independent Binomial (resp. Poisson, negative binomial) r.v.s with parameters (n, p) and (m, p) (resp. λ_1 and λ_2, r_1 and r_2), then $Y + Z$ is a binomial (resp. Poisson, negative binomial) r.v. with parameter $(m + n, p)$ (resp. $\lambda_1 + \lambda_2, r_1 + r_2$). Thus, if $\{X_n, n \geq 1\}$ are i.i.d. Poisson r.v.s with parameter λ , the sum $S_n = \sum_{i=1}^n X_i$ is a Poisson r.v. with parameter $n\lambda$.
12. In the random casting of r balls into n cells, let $Y_i = 1$ if the i th cell is empty and $= 0$ otherwise. For any $k < n$, show that the p.d.f. of Y_{i_1}, \dots, Y_{i_k} depends only on k and not on i_1, \dots, i_k . Hint: It suffices to consider $P\{Y_{i_1} = 1, \dots, Y_{i_k} = 1\}$ for all $k \leq n$.
13. If $\{A_n\}$ is a sequence of independent events with $P\{A_n\} < 1, n \geq 1$, and $P\{\bigcup_{n=1}^{\infty} A_n\} = 1$, then $P\{A_n, \text{i.o.}\} = 1$.
14. Let $\Omega = [0, 1]$ and $\mathcal{A} = \text{Borel subsets of } \Omega$, and let P be a probability measure on \mathcal{A} such that $P\{[a, b)\} = b - a$ for $0 \leq a \leq b \leq 1$. Such a "Lebesgue measure" exists and is unique (Section 6.1). For $\omega \in \Omega$, let $\omega = \omega_1\omega_2\dots$ be the decimal expansion of ω (for definiteness, no "finite expansion" is permitted). Prove that $\{\omega_n, n \geq 1\}$ are independent r.v.s with $P\{\omega_n = j\} = \frac{1}{10}, j = 0, 1, \dots, 9$.
15. If N_p is a geometric r.v. with parameter p , that is, $P\{N_p = k\} = pq^k, k \geq 0$, prove that $\lim_{p \rightarrow 0} P\{pN_p < x\} = 1 - e^{-x}, x > 0$, and check that $F(x; \lambda) = (1 - e^{-\lambda x})I_{\{x > 0\}}$ is a d.f. for any $\lambda > 0$. A r.v. with d.f. $F(x; \lambda)$ is said to have an **exponential distribution** with parameter λ .
16. If $\{A_n, n \geq 1\}$ is a sequence of events with $\sum_{n=1}^{\infty} P\{A_n\} = \infty$, show that $\limsup_{n \rightarrow \infty} P\{\bigcup_{j=n}^{n-1} A_j | A_n\} = 1$ for all $m \geq 1$.

3.2 Borel–Cantelli Theorem, Characterization of Independence, Kolmogorov Zero–One Law

The Borel–Cantelli theorem is a *sine qua non* of probability theory and is instrumental in proving strong laws of large numbers, the law of the iterated logarithm (Chapter 10), etc. The portion of Theorem 1 that is complementary to Lemma 2.2.2 postulates independent events and while this proviso can be weakened (Lemma 4.2.4) some such restriction is necessary (see Example 1).

Theorem 1 (Borel–Cantelli). *If $\{A_n, n \geq 1\}$ is a sequence of events with $\sum_1^\infty P\{A_n\} < \infty$, then $P\{A_n, \text{i.o.}\} = 0$. Conversely, if the events $\{A_n, n \geq 1\}$ are independent and $\sum_1^\infty P\{A_n\} = \infty$, then $P\{A_n, \text{i.o.}\} = 1$.*

PROOF. The first part is just the Borel–Cantelli lemma of the prior chapter. If the events are independent, Theorem 3.1.3 ensures that $\limsup_{n \rightarrow \infty} P\{\bigcup_{j=n+1}^\infty A_j\} = 1$; therefore, since $P\{\bigcup_{j=n}^\infty A_j\}$ is a monotone sequence,

$$1 = \lim_{n \rightarrow \infty} P\left\{\bigcup_{j=n}^\infty A_j\right\} = P\{A_n, \text{i.o.}\}. \quad \square$$

Corollary 1 (Borel Zero–One Criterion). *If $\{A_n, n \geq 1\}$ are independent events, $P\{A_n, \text{i.o.}\} = 0$ or 1 according as $\sum_1^\infty P\{A_n\} < \infty$ or $\sum_1^\infty P\{A_n\} = \infty$.*

EXAMPLE 1 (D. J. Newman).¹ In a sequence of Bernoulli trials $\{X_n\}$ with success probability $p \in (0, 1)$, define N_n to be the length of the maximal success run commencing at trial n , that is,

$$\{\omega: N_n = j\} = \{\omega: X_{n+j} = 0, X_i = 1, n \leq i < n+j\}, \quad j \geq 0.$$

If $\log n$ denotes logarithm to the base $1/p$, then

$$P\left\{\overline{\lim}_{n \rightarrow \infty} \frac{N_n}{\log n} = 1\right\} = 1. \quad (1)$$

PROOF. Since (Exercise 3.1.10) N_n has a geometric distribution

$$P\{N_n > a \log n\} = \sum_{j>a \log n} qp^j \leq p^{a \log n} = \frac{1}{n^a},$$

and so by the Borel–Cantelli theorem

$$P\{N_n > a \log n, \text{i.o.}\} = 0, \quad a > 1,$$

implying

$$P\left\{\overline{\lim}_{n \rightarrow \infty} \frac{N_n}{\log n} \leq 1\right\} = 1. \quad (2)$$

¹ See (Feller, 1950, p. 210).

To circumvent the dependence of the events $\{N_n > a \log n\}$, $n \geq 1$, define $k_n = [n \log n] =$ greatest integer equal to or less than $n \log n$. If $\log_2 n$ denotes $\log \log n$ and $0 < a < 1$,

$$\begin{aligned} k_n + [a \log k_n] &\leq (n+1) \log n - \log n + a(\log n + \log_2 n) \\ &\leq k_{n+1} - (1-a)\log n + a \log_2 n + 1, \end{aligned}$$

whence

$$k_{n+1} - (k_n + [a \log k_n]) \geq (1-a)\log n - a \log_2 n - 1 > 1$$

for $n \geq n_0$. Consequently, the events

$$A_n = \{N_{k_n} \geq [a \log k_n] + 1\}, \quad n \geq n_0,$$

are independent and, moreover,

$$\begin{aligned} P\{A_n\} &= P\{X_{k_n} = 1, \dots, X_{k_n + [a \log k_n]} = 1\} = p^{[a \log k_n] + 1} \\ &\geq p^{a \log k_n + 1} \geq \frac{p}{(n \log n)^a}, \end{aligned}$$

implying $\sum_{n=n_0}^{\infty} P\{A_n\} = \infty$. Thus, by the Borel–Cantelli theorem, for $a \in (0, 1)$

$$P\{N_{k_n} \geq a \log k_n, \text{i.o.}\} \geq P\{N_{k_n} \geq [a \log k_n] + 1, \text{i.o.}\} = 1,$$

yielding

$$P\left\{\overline{\lim}_{n \rightarrow \infty} \frac{N_n}{\log n} \geq 1\right\} \geq P\left\{\overline{\lim}_{n \rightarrow \infty} \frac{N_{k_n}}{\log k_n} \geq 1\right\} = 1,$$

which, in conjunction with (2), proves (1). \square

Next, it will be shown that independence of a finite set of r.v.s depends only upon their joint d.f. Some preliminary lemmas permitting enlargement of independent π -classes are needed.

Lemma 1. *If \mathcal{G} and \mathcal{D} are independent classes of events and \mathcal{D} is a π -class, then \mathcal{G} and $\sigma(\mathcal{D})$ are independent.*

PROOF. For any $B \in \mathcal{G}$, define

$$\mathcal{D}^* = \{A : A \in \sigma(\mathcal{D}) \text{ and } P\{A \cdot B\} = P\{A\} \cdot P\{B\}\}.$$

Then: $\Omega \in \mathcal{D}^*$; $A_1 - A_2 \in \mathcal{D}^*$ if $A_1, A_2 \in \mathcal{D}^*$ and $A_1 \supset A_2$; $A \in \mathcal{D}^*$ if $A = \lim A_n$, where $A_n \in \mathcal{D}^*$ and $A_n \subset A_{n+1}$, $n \geq 1$. Thus, \mathcal{D}^* is a λ -class which contains \mathcal{D} by hypothesis. Consequently, $\mathcal{D}^* \supset \sigma(\mathcal{D})$ by Theorem 1.3.2. In other words, $P\{A \cdot B\} = P\{A\} \cdot P\{B\}$ for any $B \in \mathcal{G}$ and every $A \in \sigma(\mathcal{D})$. \square

Lemma 2. *Let $\{X_t, t \in T\}$ be a stochastic process on (Ω, \mathcal{F}, P) and T_1, T_2 non-empty disjoint subsets of T . For $i = 1, 2$ define \mathcal{D}_i to be the class of all sets*

$D_m^i = \{X_{t_1} < \lambda_1, \dots, X_{t_m} < \lambda_m\}$, where m is a positive integer, λ_j a real number, and $t_j^{(i)} \in T_i$, $1 \leq j \leq m$. If \mathcal{D}_1 and \mathcal{D}_2 are independent classes, then so are $\sigma(X_t, t \in T_1)$ and $\sigma(X_t, t \in T_2)$.

PROOF. Since \mathcal{D}_1 and \mathcal{D}_2 are π -classes and \mathcal{D}_1 is independent of \mathcal{D}_2 , Lemma 1 ensures that \mathcal{D}_1 and $\sigma(\mathcal{D}_2)$ are independent. A second application of Lemma 1 yields independence of $\sigma(\mathcal{D}_1)$ and $\sigma(\mathcal{D}_2)$. Since $\sigma(\mathcal{D}_i) = \sigma(X_t, t \in T_i)$, $i = 1, 2$, Lemma 2 is proved.

Corollary 2. If the r.v.s X_t , $t \in T$, are independent and T_1 , T_2 are nonempty disjoint subsets of T , then $\sigma(X_t, t \in T_1)$ and $\sigma(X_t, t \in T_2)$ are independent.

The joint d.f. of any finite set of r.v.s X_1, \dots, X_n has been defined in (2) of Section 1.6. by

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P\{X_1 < x_1, \dots, X_n < x_n\}. \quad (3)$$

Theorem 2. Random variables X_1, \dots, X_n on a probability space (Ω, \mathcal{F}, P) are independent if and only if

$$F_{X_1, \dots, X_n} = \prod_{j=1}^n F_{X_j}. \quad (4)$$

PROOF. Necessity is trivial since $\{X_j < x_j\}$ is an event in $\sigma(X_j)$, $1 \leq j \leq n$. To prove sufficiency, let \mathcal{D}_1 be the class of sets of the type $\{X_n < \lambda_n\}$ while \mathcal{D}_2 is the class of sets of the form $\{X_1 < \lambda_1, \dots, X_{n-1} < \lambda_{n-1}\}$. Then \mathcal{D}_1 and \mathcal{D}_2 are π -classes and independent. It follows from Lemma 2 that $\sigma(X_n)$ and $\sigma(X_1, \dots, X_{n-1})$ are independent. Consequently, if $A_n \in \sigma(X_n)$ and $A_i \in \sigma(X_i)$ $\subset \sigma(X_1, \dots, X_{n-1})$, $1 \leq i < n$,

$$P\{A_1 \cdots A_{n-1} A_n\} = P\{A_1 \cdots A_{n-1}\} P\{A_n\}. \quad (5)$$

If $n = 2$, the proof is complete. Otherwise, since (4) holds with n replaced by $n - 1$ so does (5), and repeating the argument a finite number of times

$$P\{A_1 A_2 \cdots A_n\} = \prod_{i=1}^n P\{A_i\}$$

for all $A_i \in \sigma(X_i)$, $1 \leq i \leq n$. Thus, the classes $\sigma(X_i)$, $1 \leq i \leq n$, and hence also the r.v.s X_i , $1 \leq i \leq n$, are independent. \square

Corollary 3. (i) The random variables comprising a stochastic process $\{X_t, t \in T\}$ are independent iff for all finite subsets of indices $\{t_j \in T, 1 \leq j \leq m\}$ the joint distribution of X_{t_1}, \dots, X_{t_m} coincides with $\prod_{j=1}^m F_{X_{t_j}}$.

(ii) If $\{X_n, n \geq 1\}$ are independent r.v.s, $\{Y_n, n \geq 1\}$ are independent r.v.s and X_n and Y_n are identically distributed for $n \geq 1$, then $F_{X_1, \dots, X_n} = F_{Y_1, \dots, Y_n}$, $n \geq 1$.

It is a remarkable fact that probabilities of sets of a certain class defined in terms of independent r.v.s cannot be other than zero or one. The sets in question are encompassed by the following

Definition. The **tail σ -algebra** of a sequence $\{X_n, n \geq 1\}$ of r.v.s on a probability space (Ω, \mathcal{F}, P) is $\bigcap_{n=1}^{\infty} \sigma(X_j, j \geq n)$. The sets of the tail σ -algebra are called **tail events** and functions measurable relative to the tail σ -algebra are dubbed **tail functions**.

A typical example of a tail event of $\{X_n, n \geq 1\}$ is $\{\omega : \sum_{n=1}^{\infty} X_n(\omega) \text{ converges}\}$ since convergence depends entirely upon the "tail of the series."

If $A_n, n \geq 1$, are independent events, then $X_n = I_{A_n}$ are independent r.v.s. Moreover, $\bigcup_{j=n}^{\infty} A_j \in \sigma(X_j, j \geq m)$ for $n \geq m \geq 1$ whence

$$\{A_n, \text{i.o.}\} = \bigcap_{n=m}^{\infty} \bigcup_{j=n}^{\infty} A_j \in \sigma(X_m, X_{m+1}, \dots), \quad \text{all } m \geq 1.$$

The Kolmogorov zero-one law (below) confines the value of $P\{A_n, \text{i.o.}\}$ to zero or one while the Borel zero-one criterion provides a touchstone. The zero-one law, however, applies to independent r.v.s other than indicator functions.

Theorem 3 (Kolmogorov Zero-One Law). *Tail events of a sequence $\{X_n, n \geq 1\}$ of independent random variables have probabilities zero or one.*

PROOF. By Corollary 2, for each $n \geq 1$ the classes $\sigma(X_i, 1 \leq i \leq n)$ and $\sigma(X_j, j > n)$ are independent and so *a fortiori* are $\sigma(X_i, 1 \leq i \leq n)$ and $\bigcap_{n=0}^{\infty} \sigma(X_j, j > n) = \mathcal{D}$ (say) for every $n \geq 1$. The latter implies that $\mathcal{A} = \bigcup_{n=1}^{\infty} \sigma(X_j, 1 \leq j \leq n)$ is independent of \mathcal{D} . Since \mathcal{A} is an algebra (Exercise 1.3.4) and hence a π -class, Lemma 1 ensures that $\sigma(\mathcal{A})$ and \mathcal{D} are independent. But $\mathcal{D} \subset \sigma(X_n, n \geq 1) = \sigma(\mathcal{A})$, whence the tail σ -algebra \mathcal{D} is independent of itself! In other words, for every $B \in \mathcal{D}$, $P\{B\} = P\{B \cdot B\} = P^2\{B\}$, implying $P\{B\} = 0$ or 1. \square

Corollary 4. *Tail functions of a sequence of independent r.v.s are degenerate, that is, a.c. constant.*

PROOF. For any tail function Y , by the zero-one law $P\{Y < c\} = 0$ or 1 for every c in $(-\infty, \infty)$. If $P\{Y < c\} = 0$ for all c , then $P\{Y = \infty\} = 1$, whereas if $P\{Y < c\} = 1$ for all c , then $P\{Y = -\infty\} = 1$. Otherwise $c_0 = \inf\{c : P\{Y < c\} = 1\}$ is finite, whence $P\{Y = c_0\} = 1$ via the definition of c_0 . \square

Corollary 5. *If $\{X_n, n \geq 1\}$ is a sequence of independent r.v.s then $\overline{\lim}_{n \rightarrow \infty} X_n$ and $\underline{\lim}_{n \rightarrow \infty} X_n$ are degenerate.*

PROOF. Since for each $n \geq k \geq 1$, X_n is $\sigma(X_j, j \geq k)$ -measurable, $Y_k = \sup_{n \geq k} X_n$ is $\sigma(X_j, j \geq k)$ -measurable (Section 1.4), whence Y_n is $\sigma(X_j, j \geq k)$ -measurable for $n \geq k \geq 1$, implying $\overline{\lim}_{n \rightarrow \infty} X_n = \lim_{n \rightarrow \infty} \sup_{j \geq n} X_j = \lim_{n \rightarrow \infty} Y_n$ is $\sigma(X_j, j \geq k)$ -measurable for all $k \geq 1$. Thus, $\overline{\lim}_{n \rightarrow \infty} X_n$ is $\bigcap_{k=1}^{\infty} \sigma(X_j, j \geq k)$ -measurable, i.e., a tail function, and similarly for $\underline{\lim} X_n$, so that the conclusion follows from Corollary 4. \square

EXERCISES 3.2

1. Find a trivial example of dependent events A_n with $\sum P\{A_n\}$ divergent but $P\{A_n, \text{i.o.}\} < 1$.

2. If $X_n, n \geq 1$, are independent r.v.s, prove that $P\{\lim_{n \rightarrow \infty} X_n = 0\} = 1$ iff

$$\sum_{n=1}^{\infty} P\{|X_n| \geq \varepsilon\} < \infty,$$

all $\varepsilon > 0$.

3. Prove that if (i) $\{X_n, n \geq 1\}$ are i.i.d. $N(0, \sigma^2)$ random variables, then

$$P\left\{\overline{\lim}_{n \rightarrow \infty} \frac{X_n}{\sqrt{2 \log n}} = \sigma\right\} = 1.$$

If, rather, $\{X_n, n \geq 1\}$ are i.i.d. exponential r.v.s with parameter λ (Exercise 3.1.15) then $P\{\overline{\lim}_{n \rightarrow \infty} X_n / \log n = \lambda^{-1}\} = 1$. (ii) If $\{X_n, n \geq 1\}$ are i.i.d. Poisson r.v.s with parameter λ , then $P\{\overline{\lim}_{n \rightarrow \infty} X_n (\log \log n) / \log n = 1\} = 1$ irrespective of λ . Hint: Recall Exercise 2.1.4.

4. Show that random vectors $X = (X_1, \dots, X_m)$ and $Y = (Y_1, \dots, Y_n)$ are independent (of one another) iff the joint d.f. of X and Y is the product of the d.f.s of X and of Y .
5. If $\{X_n, n \geq 1\}$ is a sequence of finite-valued r.v.s and $S_n = \sum_{i=1}^n X_i$, determine which of the following are tail events of $\{X_n, n \geq 1\}$: $\{S_n \text{ converges}\}$; $\{\overline{\lim} S_n > \underline{\lim} S_n\}$; $\{\overline{\lim} S_n = \infty\}$; $\{X_n > c, \text{i.o.}\}$; $\{\underline{\lim} X_n = 0\}$; $\{S_n > c_n, \text{i.o.}\}$.
6. If $\{X_n, n \geq 1\}$ is a sequence of independent finite-valued r.v.s and $\{a_n, n \geq 1\}$ is a sequence of finite constants with $0 < a_n \rightarrow \infty$, show that $\overline{\lim} S_n/a_n$ and $\underline{\lim} S_n/a_n$ are degenerate.
7. (Barndorff-Nielsen.) If $\{A_n, n \geq 1\}$ are events satisfying $\sum_{n=1}^{\infty} P\{A_n A_{n+1}^c\} < \infty$ and $P\{A_n\} = o(1)$, then $P\{A_n, \text{i.o.}\} = 0$.
8. Let $\{X_n, n \geq 1\}$ be i.i.d. r.v.s with $P\{X_1 = 1\} = P\{X_1 = -1\} = \frac{1}{2}$ and set $S_n = \sum_{i=1}^n X_i$. Prove that
- $\lim_{n \rightarrow \infty} P\{S_n / \sqrt{n} < x\} = \Phi(x)$, $-\infty < x < \infty$,
 - $P\{\inf_{n \geq 1} S_n = -\infty\} = P\{\sup_{n \geq 1} S_n = \infty\} = 1$.
9. If $X_n, n \geq 1$, are i.i.d. demonstrate equivalence of the following relations:
- $\overline{\lim}_{n \rightarrow \infty} |X_n|/n \leq 1$, a.s.;
 - $\sum_{n=1}^{\infty} P\{|X_n| > n\} < \infty$;
 - $\sum_{n=1}^{\infty} n P\{n-1 \leq |X_1| < n\} < \infty$.
10. Verify for i.i.d. r.v.s $X_n, n \geq 1$, that $P\{\overline{\lim}_{n \rightarrow \infty} X_n = \infty\} = 1$ if and only if X_1 is unbounded above, i.e., $P\{X_1 < C\} < 1$, all $C < \infty$.

11. For each $n \geq 1$, let $\{X_{n,j}, j \geq 1\}$ be a sequence of independent r.v.s. Then $\sup_n |X_{n,j}| \xrightarrow{P} 0$ iff $\sum_{j=1}^{\infty} P\{|X_{n,j}| > \varepsilon\} = o(1)$, all $\varepsilon > 0$.
12. If $\{X_n, n \geq 1\}$ are independent r.v.s with $E X_n = 0$, $E X_n^2 = 1$ which obey a central limit theorem, i.e., $\lim_{n \rightarrow \infty} P\{S_n \geq xn^{1/2}\} = 1 - \Phi(x)$, $x \in (-\infty, \infty)$, where $S_n = \sum_i^n X_i$, $n \geq 1$, prove that $\overline{\lim}_{n \rightarrow \infty} S_n/n^{1/2} = \infty$, a.c. Hint: Utilize the zero-one law.

3.3 Convergence in Probability, Almost Certain Convergence, and Their Equivalence for Sums of Independent Random Variables

If $\{X_n, n \geq 1\}$ is a sequence of r.v.s on (Ω, \mathcal{F}, P) , the set C , where $X_n, n \geq 1$, converges (i.e., $\lim X_n(\omega)$ exists and is finite) is a measurable set since

$$C = \bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{i=1}^{\infty} \left\{ |X_{n+i}(\omega) - X_n(\omega)| < \frac{1}{k} \right\}.$$

Moreover, if $P\{C\} = 1$, there exists a r.v. X on (Ω, \mathcal{F}, P) such that

$$P\left\{\lim_{n \rightarrow \infty} X_n = X\right\} = 1. \quad (1)$$

It suffices to define

$$\begin{aligned} X(\omega) &= \lim_{n \rightarrow \infty} X_n(\omega), & \omega \in C, \\ &= \overline{\lim}_{n \rightarrow \infty} X_n(\omega), & \omega \in C^c, \end{aligned}$$

and clearly X is a r.v. In such a situation, the r.v.s $X_n, n \geq 1$, are said to **converge almost certainly** (or almost surely) to the r.v. X , denoted symbolically by $X_n \xrightarrow{a.s.} X$. Even if $X(\omega)$ is not a r.v. it is still possible that $\underline{\lim}_{n \rightarrow \infty} X_n(\omega) = \overline{\lim}_{n \rightarrow \infty} X_n(\omega) = X(\omega)$ with probability one, likewise denoted by $X_n \xrightarrow{a.s.} X$. The former will be distinguished from the latter on occasion by writing $X_n \xrightarrow{a.s.} \text{a r.v. } X$ or $X_n \xrightarrow{a.s.} X$, finite.

Definition. A sequence of r.v.s $\{X_n, n \geq 1\}$ on a probability space (Ω, \mathcal{F}, P) is said to **converge in probability** (to X) if there exists a r.v. X on (Ω, \mathcal{F}, P) such that

$$P\{|X_n - X| > \varepsilon\} = o(1), \quad \text{all } \varepsilon > 0.$$

This will be denoted symbolically by $X_n \xrightarrow{P} X$.

If r.v.s $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{P} Y$, then $X = Y$, a.c. (Exercise 1). Moreover, the calculus of convergence of real numbers carries over to convergence in probability (Exercise 1(ii), 1(iv)).

The next lemmas analyze a.c. convergence and convergence in probability.

Lemma 1. Random variables $X_n \xrightarrow{\text{a.c.}} a$ r.v. X iff $\sup_{j \geq n} |X_j - X| \xrightarrow{P} 0$ iff for all $\varepsilon, \delta > 0$

$$P\left\{\bigcap_{j=n}^{\infty} [|X_j - X| \leq \varepsilon]\right\} \geq 1 - \delta \quad \text{for } n \geq N(\varepsilon, \delta).$$

PROOF. By Lemma 2.2.1, $X_n \xrightarrow{\text{a.c.}} X$, a r.v., or equivalently $X_n - X \xrightarrow{\text{a.c.}} 0$, iff

$$P\{|X_n - X| > \varepsilon, \text{i.o.}\} = 0, \quad \varepsilon > 0.$$

Hence, $X_n \xrightarrow{\text{a.c.}} X$ iff for all $\varepsilon > 0$

$$\begin{aligned} \lim_{n \rightarrow \infty} P\left\{\sup_{j \geq n} |X_j - X| > \varepsilon\right\} &= \lim_{n \rightarrow \infty} P\left\{\bigcup_{j=n}^{\infty} [|X_j - X| > \varepsilon]\right\} \\ &= P\left\{\bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} [|X_j - X| > \varepsilon]\right\} = 0, \end{aligned} \quad (2)$$

that is, iff $\sup_{j \geq n} |X_j - X| \xrightarrow{P} 0$. The final condition of the lemma is simply a restatement in finite terms of the following alternative form of (2):

$$\lim_{n \rightarrow \infty} P\left\{\bigcap_{j=n}^{\infty} [|X_j - X| \leq \varepsilon]\right\} = P\left\{\bigcup_{n=1}^{\infty} \bigcap_{j=n}^{\infty} [|X_j - X| \leq \varepsilon]\right\} = 1, \quad \varepsilon > 0.$$

□

If $X, X_n, n \geq 1$ are r.v.s with $\sum_{n=1}^{\infty} P\{|X_n - X| > \varepsilon\} < \infty$, all $\varepsilon > 0$ then the Borel-Cantelli lemma ensures that $X_n \xrightarrow{\text{a.c.}} X$.

Corollary 1. If random variables $X_n \xrightarrow{\text{a.c.}} a$ r.v. X , then $X_n \xrightarrow{P} X$.

Although a.c. convergence of r.v.s $X_n, n \geq 1$, is, in general, stronger than convergence in probability, the latter does circumscribe $\overline{\lim} X_n$ and $\underline{\lim} X_n$ (Exercise 5).

Corollary 2. If random variables $X_n \xrightarrow{\text{a.c.}} a$ r.v. X , then $g(X_n) \xrightarrow{\text{a.c.}} g(X)$ for g continuous.

Strictly speaking, the latter is not a corollary of the lemma but rather an immediate consequence of the definition and continuity.

Lemma 2. Random variables $X_n \xrightarrow{P} a$ r.v. X iff (i) $\sup_{m > n} P\{|X_m - X_n| > \varepsilon\} = o(1), \varepsilon > 0$, iff (ii) every subsequence of $\{X_n, n \geq 1\}$ has itself a subsequence converging a.c. to the same r.v. X .

PROOF. If $X_n \xrightarrow{P} X$, then for all $\varepsilon > 0, n \geq n_0(\varepsilon)$ implies $P\{|X_n - X| > \varepsilon\} < \varepsilon$. Hence, for $m > n \geq n_0(\varepsilon/2)$

$$P\{|X_m - X_n| > \varepsilon\} \leq P\left\{|X_m - X| > \frac{\varepsilon}{2}\right\} + P\left\{|X - X_n| > \frac{\varepsilon}{2}\right\} < \varepsilon,$$

which implies (i). Conversely, if (i) obtains, then for any integer $k \geq 1$, $P\{|X_n - X_m| > 2^{-k}\} < 2^{-k}$ provided $n > m \geq m_k$. Set $n_1 = m_1$, $n_{i+1} = \max(n_i + 1, m_{i+1})$, and $X'_k = X_{n_k}$ and $A_k = \{|X'_{k+1} - X'_k| > 2^{-k}\}$. Then $\sum_{k=1}^{\infty} P(A_k) < \infty$, whence the Borel-Cantelli theorem ensures that, apart from an ω -set A of measure zero, $|X'_{k+1}(\omega) - X'_k(\omega)| \leq 2^{-k}$ provided $k \geq$ some integer $k_0(\omega)$. Hence, for $\omega \in A^c$, as $n \rightarrow \infty$

$$\sup_{m>n} |X'_m - X'_n| \leq \sum_{k=n}^{\infty} |X'_{k+1} - X'_k| \leq \sum_{k=n}^{\infty} 2^{-k} = \frac{1}{2^{n-1}} = o(1),$$

and so

$$P\left\{\overline{\lim}_{k \rightarrow \infty} X'_k = \underline{\lim}_{k \rightarrow \infty} X'_k, \text{ finite}\right\} = 1. \quad (3)$$

If $X = \underline{\lim} X'_k$, then X is a r.v. and, according to (3), $X_{n_k} \xrightarrow{a.c.} X$ as $k \rightarrow \infty$. By Corollary 1, $X_{n_k} \xrightarrow{P} X$. Since for any $\varepsilon > 0$

$$P\{|X_k - X| > \varepsilon\} \leq P\left\{|X_k - X_{n_k}| > \frac{\varepsilon}{2}\right\} + P\left\{|X_{n_k} - X| > \frac{\varepsilon}{2}\right\} = o(1)$$

as $k \rightarrow \infty$, $X_k \xrightarrow{P} X$.

Next, if $X_n \xrightarrow{P} X$, any subsequence of $\{X_n\}$, say $X''_n \xrightarrow{P} X$, whence, as already shown, there exists a subsequence of $\{X''_n\}$, say $X'_n \xrightarrow{a.c.} Y$. Then $X'_n \xrightarrow{P} Y$ but also $X'_n \xrightarrow{P} X$, necessitating $X = Y$, a.c. (Exercise 1). Thus, the subsequence X''_n has a further subsequence $X'_n \xrightarrow{a.c.} X$.

Finally, if X_n does not converge in probability to X , there exists an $\varepsilon > 0$ and a subsequence X_{n_k} with $P\{|X_{n_k} - X| > \varepsilon\} > \varepsilon$. But then no subsequence of X_{n_k} converges in probability to X (*a fortiori* almost certainly to X) in violation of (ii).

Corollary 3. *If random variables $X_n \xrightarrow{P} X$, then $g(X_n) \xrightarrow{P} g(X)$ for every continuous function g .*

PROOF. Every subsequence of $Y_n = g(X_n)$ has a further subsequence $Y_{n_k} = g(X_{n_k})$ with $X_{n_k} \xrightarrow{a.c.} X$. By Corollary 2, $Y_{n_k} = g(X_{n_k}) \xrightarrow{a.c.} g(X)$, whence, by Lemma 2, $g(X_n) = Y_n \xrightarrow{P} g(X)$.

Corollary 4. *Random variables $X_n \xrightarrow{a.c.} a$ r.v. X iff (iii) $\sup_{m>n} |X_m - X_n| \xrightarrow{P} 0$.*

PROOF. If $X_n \xrightarrow{a.c.} X$, both $\sup_{m>n} |X_m - X| \xrightarrow{P} 0$ and $|X_n - X| \xrightarrow{P} 0$ via Lemma 1, whence for any $\varepsilon > 0$

$$\begin{aligned} P\left\{\sup_{m \geq n} |X_m - X_n| > \varepsilon\right\} &\leq P\left\{\sup_{m \geq n} |X_m - X| > \frac{\varepsilon}{2}\right\} \\ &\quad + P\left\{|X - X_n| > \frac{\varepsilon}{2}\right\} = o(1). \end{aligned}$$

Conversely, (iii) entails $\sup_{m > n} P\{|X_m - X_n| > \varepsilon\} = o(1)$, all $\varepsilon > 0$, and so by Lemma 2 there exists a r.v. X with $X_n \xrightarrow{P} X$. Thus, for all $\varepsilon > 0$

$$\begin{aligned} P\left\{\sup_{m \geq n} |X_m - X| > \varepsilon\right\} &\leq P\left\{\sup_{m \geq n} |X_m - X_n| > \frac{\varepsilon}{2}\right\} \\ &\quad + P\left\{|X_n - X| > \frac{\varepsilon}{2}\right\} = o(1) \end{aligned}$$

as $n \rightarrow \infty$, implying $X_n \xrightarrow{\text{a.c.}} X$ by Lemma 1. \square

The question of a.c. convergence of a sequence of r.v.s $\{X_n, n \geq 1\}$ depends only upon the corresponding sequence of joint d.f.s $\{F_{X_1, \dots, X_n}, n \geq 1\}$. In other words, if $\{X_n, n \geq 1\}$ and $\{Y_n, n \geq 1\}$ have the same finite-dimensional joint d.f.s, then $X_n \xrightarrow{\text{a.c.}} X$ iff $Y_n \xrightarrow{\text{a.c.}} Y$. In fact, setting $A_m = \{|X_m - X_n| > \varepsilon\}$,

$$\begin{aligned} P\left\{\sup_{m > n} |X_m - X_n| > \varepsilon\right\} &= P\left\{\bigcup_{m=n+1}^{\infty} A_m\right\} \\ &= P\{A_{n+1}\} + P\{A_{n+1}^c A_{n+2}\} \\ &\quad + P\{A_{n+1}^c A_{n+2}^c A_{n+3}\} + \dots \\ &= P\left\{\sup_{m > n} |Y_m - Y_n| > \varepsilon\right\} \end{aligned}$$

by Theorem 1.6.1, and so the equivalence follows from Lemma 1.

Suppose that for some constants $0 < b_n \uparrow \infty$, $b_{n+1}/b_n \rightarrow 1$, i.i.d. r.v.s $\{X_n, n \geq 1\}$ and partial sums $S_n = \sum_1^n X_i, n \geq 1$,

$$\frac{S_n}{b_n} \xrightarrow{\text{a.c.}} S, \quad \text{finite.} \quad (4)$$

Then, clearly, it is necessary that

$$\frac{X_n}{b_n} = \frac{S_n}{b_n} - \frac{b_{n-1}}{b_n} \frac{S_{n-1}}{b_{n-1}} \xrightarrow{\text{a.c.}} 0, \quad (5)$$

and via the Borel–Cantelli theorem

$$\sum_{n=1}^{\infty} P\{|X_1| > \varepsilon b_n\} = \sum_{n=1}^{\infty} P\{|X_n| > \varepsilon b_n\} < \infty, \quad \text{all } \varepsilon > 0. \quad (6)$$

This is a restriction on the d.f. of X_1 (see Corollary 4.1.3) and thereby provides a necessary condition for (4). On the other hand, if (4) is replaced by

$$\frac{1}{b_n} \sum_{j=1}^n \frac{X_{n+1-j}}{a_j} \xrightarrow{\text{a.c.}} 0, \quad (7)$$

the d.f. of X_1 should likewise be constrained, but the simple subtraction of (5) leads nowhere. However, (6) is still necessary for (7) if the r.v.s $\{X_n, n \geq 1\}$ are i.i.d. as follows from the second of the ensuing lemmas.

Lemma 3 (Feller-Chung). *Let $\{A_n, n \geq 1\}$ and $\{B_n, n \geq 1\}$ be sequences of events on (Ω, \mathcal{F}, P) and set $A_0 = \emptyset$. If either (i) B_n and $A_n A_{n-1}^c \cdots A_0^c$ are independent for all $n \geq 1$ or (ii) the classes $\{B_n\}$ and*

$$\{A_n, A_n A_{n+1}^c, A_n A_{n+1}^c A_{n+2}^c, \dots\}$$

are independent for all $n \geq 1$, then

$$P\left\{\bigcup_{n=1}^{\infty} A_n B_n\right\} \geq P\left\{\bigcup_{n=1}^{\infty} A_n\right\} \inf_{n \geq 1} P\{B_n\}.$$

PROOF. In case (i)

$$\begin{aligned} P\left\{\bigcup_{n=1}^{\infty} A_n B_n\right\} &= P\left\{\bigcup_{n=1}^{\infty} B_n A_n \bigcap_{j=0}^{n-1} (B_j A_j)^c\right\} = \sum_{n=1}^{\infty} P\left\{B_n A_n \bigcap_{j=0}^{n-1} (B_j A_j)^c\right\} \\ &\geq \sum_{n=1}^{\infty} P\left\{B_n A_n \bigcap_{j=0}^{n-1} A_j^c\right\} = \sum_{n=1}^{\infty} P\{B_n\} P\left\{A_n \bigcap_{j=0}^{n-1} A_j^c\right\} \\ &\geq \left(\inf_{n \geq 1} P\{B_n\}\right) P\left\{\bigcup_{n=1}^{\infty} A_n\right\}, \end{aligned}$$

and in case (ii)

$$\begin{aligned} P\left\{\bigcup_{j=1}^n A_j B_j\right\} &= \sum_{j=1}^n P\left\{A_j B_j \bigcap_{i=j+1}^n (A_i B_i)^c\right\} \geq \sum_{j=1}^n P\left\{A_j B_j \bigcap_{i=j+1}^n A_i^c\right\} \\ &= \sum_{j=1}^n P\{B_j\} P\left\{A_j \bigcap_{i=j+1}^n A_i^c\right\} \geq \inf_{1 \leq j \leq n} P\{B_j\} P\left\{\bigcup_{j=1}^n A_j\right\}, \end{aligned}$$

whence the conclusion follows as $n \rightarrow \infty$. \square

Lemma 4. *Let $\{Y_n, n \geq 1\}$ and $\{Z_n, n \geq 1\}$ be sequences of r.v.s such that either (i) Y_n and (Z_1, \dots, Z_n) are independent for all $n \geq 1$ or (ii) Y_n and (Z_n, Z_{n+1}, \dots) are independent for all $n \geq 1$. Then for any constants $\varepsilon_n, \delta_n, \varepsilon$, and δ ,*

$$P\left\{\bigcup_{n=1}^{\infty} [Z_n + Y_n > \varepsilon_n]\right\} \geq P\left\{\bigcup_{n=1}^{\infty} [Z_n > \varepsilon_n + \delta_n]\right\} \inf_{n \geq 1} P\{Y_n \geq -\delta_n\}, \quad (8)$$

$$P\left\{\overline{\lim}_{n \rightarrow \infty} (Z_n + Y_n) \geq \varepsilon\right\} \geq P\left\{\overline{\lim}_{n \rightarrow \infty} Z_n > \varepsilon + \delta\right\} \cdot \underline{\lim}_{n \rightarrow \infty} P\{Y_n \geq -\delta\}. \quad (9)$$

Moreover, if $\underline{\lim}_{n \rightarrow \infty} P\{Y_n \geq -\delta\} > 0$, all $\delta > 0$, then $\overline{\lim}_{n \rightarrow \infty} (Z_n + Y_n) < \varepsilon$.

a.c., entails $\overline{\lim}_{n \rightarrow \infty} Z_n \leq \varepsilon$, a.c. Furthermore, if $\overline{\lim}_{n \rightarrow \infty} P\{Y_n > -\delta\} \cdot \overline{\lim}_{n \rightarrow \infty} P\{Y_n < \delta\} > 0$ for all $\delta > 0$ (a fortiori, if $Y_n \xrightarrow{P} 0$), then $Z_n + Y_n \xrightarrow{a.c.} 0$ implies $Z_n \xrightarrow{a.c.} 0$.

PROOF. Set $A_n = \{Z_n > \varepsilon_n + \delta_n\}$, $B_n = \{Y_n \geq -\delta_n\}$. By Lemma 3, for $m \geq 1$

$$P\left\{\bigcup_{n=m}^{\infty} [Y_n + Z_n > \varepsilon_n]\right\} \geq P\left\{\bigcup_{n=m}^{\infty} A_n B_n\right\} \geq P\left\{\bigcup_{n=m}^{\infty} A_n\right\} \inf_{n \geq m} P\{B_n\},$$

yielding (8) for $m = 1$ and (9) via $\varepsilon_n \equiv \varepsilon$, $\delta_n \equiv \delta$, $m \rightarrow \infty$. The penultimate statement follows easily from (9), and since both (8) and (9) also hold for $\{-Y_n\}$, $\{-Z_n\}$ the final assertion likewise obtains. \square

EXAMPLE 1 (Chung). Let $\{X_n, n \geq 1\}$ be independent r.v.s with partial sums $S_n = \sum_1^n X_i$, $n \geq 1$. If $S_n/n \xrightarrow{P} 0$ and $S_{2^n}/2^n \xrightarrow{a.c.} 0$, then $S_n/n \xrightarrow{a.c.} 0$.

PROOF. For $k = 1, 2, \dots$ there is a unique integer $n(k)$ such that $2^{n(k)-1} \leq k < 2^{n(k)}$. Take $\varepsilon > 0$ and set

$$\begin{aligned} A_k &= \{|S_k| > 2^{n(k)+1}\varepsilon\}, & B_k &= \{|S_{2^{n(k)}} - S_k| \leq 2^{n(k)}\varepsilon\}, \\ C_n &= \{|S_{2^n}| > 2^n\varepsilon\}. \end{aligned}$$

Since B_k and $A_k A_{k-1}^c \cdots A_1^c A_0^c$ ($A_0 = \emptyset$) are independent and $C_n \supset \bigcup_{k=2^{n-1}}^{2^n-1} A_k B_k$, by Lemma 3

$$P\left\{\bigcup_{m+1}^{\infty} C_m\right\} \geq P\left\{\bigcup_{2^m}^{\infty} A_k B_k\right\} \geq P\left\{\bigcup_{2^m}^{\infty} A_k\right\} \inf_{k \geq 2^m} P\{B_k\}. \quad (10)$$

By hypothesis and Lemma 1 the left side of (10) is $o(1)$ as $m \rightarrow \infty$, and moreover,

$$P\{B_k^c\} \leq P\{|S_{2^{n(k)}}| > 2^{n(k)-1}\varepsilon\} + P\{|S_k| > 2^{n(k)-1}\varepsilon\} = o(1)$$

as $k \rightarrow \infty$. Consequently, $P\{\bigcup_{2^m}^{\infty} A_k\} = o(1)$ as $m \rightarrow \infty$, and so $S_n/n \xrightarrow{a.c.} 0$. \square

Although, in general, a.c. convergence is much stronger than convergence in probability, in the special case of sums S_n of independent random variables the two are equivalent. A basic tool in demonstrating this equivalence for sums of independent random variables is an inequality due to P. Lévy. This in turn necessitates the

Definition. For any r.v. X a real number $m(X)$ is called a **median** of X if

$$P\{X \leq m(X)\} \geq \frac{1}{2} \leq P\{X \geq m(X)\}.$$

In fact, if $a = \inf\{\lambda: P\{X \leq \lambda\} \geq \frac{1}{2}\}$, then $|a| < \infty$ and, since (Exercise 1.6.3) $P\{X \leq \lambda\}$ is right continuous, $P\{X \leq a\} \geq \frac{1}{2}$. By definition,

$$P\{X \leq a - \varepsilon\} < \frac{1}{2}$$

for all $\varepsilon > 0$, and so letting $\varepsilon \rightarrow 0$, $P\{X < a\} \leq \frac{1}{2}$ or equivalently $P\{X \geq a\} \geq \frac{1}{2}$. Thus, a is a median of X .

A pertinent observation concerning medians is that if for some constant c $P\{|X| \geq c\} < \varepsilon \leq \frac{1}{2}$, then $|m(X)| \leq c$. Moreover, if c is any finite constant, $cm(X)$ and $m(X) + c$ are medians of cX and $X + c$ respectively.

Lemma 5 (Lévy Inequalities). *If $\{X_j, 1 \leq j \leq n\}$ are independent r.v.s, $S_n = \sum_{j=1}^n X_j$, and $m(Y)$ denotes a median of Y , then for every $\varepsilon > 0$*

$$P\left\{\max_{1 \leq j \leq n} [S_j - m(S_j - S_n)] \geq \varepsilon\right\} \leq 2 P\{S_n \geq \varepsilon\}, \quad (11)$$

$$P\left\{\max_{1 \leq j \leq n} |S_j - m(S_j - S_n)| \geq \varepsilon\right\} \leq 2 P\{|S_n| \geq \varepsilon\}. \quad (12)$$

PROOF. Set $S_0 = 0$ and define T to be the smallest integer j in $[1, n]$ for which $S_j - m(S_j - S_n) \geq \varepsilon$ (if such an integer exists) and $T = n + 1$ otherwise. If

$$B_j = \{m(S_j - S_n) \geq S_j - S_n\}, \quad 1 \leq j \leq n,$$

then $P\{B_j\} \geq \frac{1}{2}$. Since $\{\omega: T = j\} \in \sigma(X_1, \dots, X_j)$, $B_j \in \sigma(X_{j+1}, \dots, X_n)$, and $\{S_n \geq \varepsilon\} \supset \bigcup_{j=1}^n B_j \{T = j\}$,

$$P\{S_n \geq \varepsilon\} \geq \sum_{j=1}^n P\{B_j \{T = j\}\} = \sum_{j=1}^n P\{B_j\} \cdot P\{T = j\} \geq \frac{1}{2} P\{1 \leq T \leq n\},$$

which is tantamount to (11). Rewrite (11) with X_j replaced by $-X_j$, $1 \leq j \leq n$, recalling that $m(-Y) = -m(Y)$, and add this to (11) to obtain (12). \square

Definition. A r.v. X is called **symmetric** or said to have a symmetric d.f. if X and $-X$ have the same distribution function.

It is easy to verify that X is symmetric iff $P\{X < x\} = P\{X > -x\}$ for every real x and also that zero is a median of a symmetric r.v. It follows from Corollary 3.2.3 that sums of independent symmetric r.v.s are themselves symmetric r.v.s. This leads directly to

Corollary 5. *If $\{X_j, 1 \leq j \leq n\}$ are independent, symmetric r.v.s with partial sums $S_n = \sum_1^n X_j$, then for every $\varepsilon > 0$*

$$P\left\{\max_{1 \leq j \leq n} S_j \geq \varepsilon\right\} \leq 2 P\{S_n \geq \varepsilon\}, \quad (13)$$

$$P\left\{\max_{1 \leq j \leq n} |S_j| \geq \varepsilon\right\} \leq 2 P\{|S_n| \geq \varepsilon\}. \quad (14)$$

Theorem 1 (Lévy). *If $\{X_n, n \geq 1\}$ is a sequence of independent r.v.s, then $S_n = \sum_1^n X_i$ converges a.c. iff it converges in probability.*

PROOF. It suffices to verify sufficiency. By Lemma 2, for any ε in $(0, \frac{1}{4})$, there exists an integer h_0 such that $n > h \geq h_0$ implies, setting $S_{h,n} = S_n - S_h$, that $P\{|S_{h,n}| > \varepsilon\} < \varepsilon$. In view of an earlier comment, this entails $|m(S_{h,n})| \leq \varepsilon$ for $n > h \geq h_0$. By Lévy's inequality (12), for $k > h \geq h_0$

$$\begin{aligned} P\left\{\max_{h < n \leq k}|S_{h,n}| > 2\varepsilon\right\} &= P\left\{\max_{h < n \leq k}|S_{h,n}| > 2\varepsilon, \max_{h < n \leq k}|m(S_{n,k})| \leq \varepsilon\right\} \\ &\leq P\left\{\max_{h < n \leq k}|S_{h,n} - m(S_{h,n} - S_{h,k})| > \varepsilon\right\} \\ &\leq 2P\{|S_{h,k}| > \varepsilon\} < 2\varepsilon. \end{aligned}$$

Hence, letting $k \rightarrow \infty$, if $h \geq h_0$,

$$P\left\{\sup_{n > h}|S_{h,n}| > 2\varepsilon\right\} \leq 2\varepsilon,$$

and so $S_n \xrightarrow{\text{a.e.}} \text{some r.v. } S$ by Corollary 4. \square

Lemmas 4 and 2 may be exploited to give an alternative proof of Lévy's Theorem. Since $S_n \xrightarrow{P} S$, the latter ensures the existence of a subsequence k_n with $S_{k_n} \xrightarrow{\text{a.e.}} S$. Now for every integer $m > 0$, there is an integer $n = n(m)$ such that $k_n \leq m < k_{n+1}$. Clearly, $m \rightarrow \infty$ entails $k_n \rightarrow \infty$.

Set $Y_m = S - S_m$ and $Z_m = S_m - S_{k_n}$. By hypothesis, $Y_m \xrightarrow{P} 0$ and moreover $Y_m + Z_m \xrightarrow{\text{a.e.}} 0$ by the choice of k_n . Clearly, (Z_1, \dots, Z_m) is $\sigma(X_1, \dots, X_m)$ -measurable and as noted in the proof of Corollary 3.2.5, Y_m is $\sigma(X_j, j > m)$ -measurable. Since the two σ -algebras are independent via Corollary 3.2.2, so are (Z_1, \dots, Z_m) and Y_m for every $m \geq 1$. Hence, by Lemma 4, $Z_m \xrightarrow{\text{a.e.}} 0$ implying $Y_m \xrightarrow{\text{a.e.}} 0$. \square

EXERCISES 3.3

1. i. If $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{P} Y$, then $P\{X = Y\} = 1$.
 ii. $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$ imply $X_n + Y_n \xrightarrow{P} X + Y$.
 iii. $X_n \xrightarrow{L} 0$ implies $m(X_n) \rightarrow 0$.
 iv. If $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$, and g is a continuous function on R^2 , then $g(X_n, Y_n) \xrightarrow{P} g(X, Y)$.
2. Let $\{X_n, n \geq 1\}$ and $\{Y_n, n \geq 1\}$ be two sequences of r.v.s with $F_{X_1, \dots, X_n} = F_{Y_1, \dots, Y_n}$ for $n \geq 1$. If $X_n \xrightarrow{P} X$, prove that $Y_n \xrightarrow{P} Y$ and that X and Y are identically distributed.
Hint: Apply Lemma 3.3.2.
3. i. What is wrong with the following "proof" of Corollary 3?

$$P\{|g(X_n) - g(X)| > \varepsilon\} \leq P\{|X_n - X| > \delta\} = o(1).$$

- ii. A r.v. X is symmetric iff X^+ and X^- have identical d.f.s.
 - iii. If 0 is a median of a r.v. X , it is also a median of $X\mathbf{1}_{\{|X|<c\}}$ for any $c > 0$.
4. If independent r.v.s $X_n \xrightarrow{P} X$, then X is degenerate. Prove for nondegenerate i.i.d. r.v.s $\{X_n\}$ that $P\{X_n \text{ converges}\} = 0$.
5. For any sequence of r.v.s $\{X_n, n \geq 0\}$ with $X_n \xrightarrow{P} X_0$,
- $$P\left\{\liminf_{n \rightarrow \infty} X_n \leq X_0 \leq \limsup_{n \rightarrow \infty} X_n\right\} = 1.$$
- Conversely, if $\liminf_{n \rightarrow \infty} X_n = X_0$ (resp. $\limsup_{n \rightarrow \infty} X_n = X_0$), a.c., then for any $\varepsilon > 0$ $P\{X_n > X_0 + \varepsilon\} = o(1)$ (resp. $P\{X_n < X_0 - \varepsilon\} = o(1)$).
6. If $\{X_n, n \geq 1\}$ are independent, symmetric r.v.s such that $(1/b_n) \sum_1^n X_i \xrightarrow{P} 0$ for some positive constants b_n , then $(1/b_n) \max_{1 \leq i \leq n} X_i \xrightarrow{P} 0$. Hint:
- $$P\left\{\max_{1 \leq j \leq n} |X_j| > \varepsilon b_n\right\} \leq P\left\{\max_{1 \leq j \leq n} \left|\sum_1^j X_i\right| > \frac{\varepsilon}{2} b_n\right\}.$$
7. If the r.v.s $X_n/b_n \xrightarrow{P} 0$, where the constants b_n satisfy $0 < b_n \uparrow \infty$, then
- $$\max_{1 \leq j \leq n} |m(X_j - X_n)| = o(b_n).$$
8. Prove for any r.v.s $\{X_n\}$ and constants $\{b_n\}$ with $0 < b_n \uparrow \infty$ that (i) $(1/b_n) \sum_1^n X_i \xrightarrow{P} 0$ implies $X_n/b_n \xrightarrow{a.e.} 0$, (ii) if for identically distributed $\{X_n\}$, some nonzero constant c and $0 < b_n \rightarrow \infty$, $(1/b_n) \sum_1^n X_i \xrightarrow{P} c$, then $b_n \sim b_{n-1}$.
9. Let the stochastic processes $\{X_n, 1 \leq n \leq k\}$ and $\{X'_n, 1 \leq n \leq k\}$ be independent of one another and have the same joint distributions. If m_n is a median of X_n , $1 \leq n \leq k$, then for $\varepsilon > 0$ (see Lemma 10.1.1)
- $$P\left\{\max_{1 \leq n \leq k} |X_n - m_n| \geq \varepsilon\right\} \leq 2 P\left\{\max_{1 \leq n \leq k} |X_n - X'_n| \geq \varepsilon\right\}.$$
10. If r.v.s $X_n \xrightarrow{a.e.} X$ as $n \rightarrow \infty$ and $\{N_n, n \geq 1\}$ are positive-integer valued r.v.s with (i) $N_n \xrightarrow{a.e.} \infty$, then $X_{N_n} \xrightarrow{a.e.} X$. If, rather, (ii) $N_n \xrightarrow{P} \infty$, that is, $P\{N_n < C\} = o(1)$ all $C > 0$ and X is a r.v., then $X_{N_n} \xrightarrow{P} X$.
11. If the r.v. W_n on (Ω, \mathcal{F}, P) is \mathcal{F}_n -measurable, $n \geq 1$, where \mathcal{F}_n is a decreasing sequence of sub- σ -algebras of \mathcal{F} and $W_n \xrightarrow{a.e.} W$, then W is $\bigcap_{n=1}^{\infty} \mathcal{F}_n$ -measurable.
12. If X_1, X_2, X_3 are independent, symmetric r.v.s with $P\{|X_1 + X_2 + X_3| \leq M\} = 1$, then $P\{\sum_{i=1}^3 |X_i| \leq M\} = 1$.
13. If $\{X, X_n, n \geq 1\}$ are finite measurable functions on a measure space (S, Σ, μ) with $\mu\{|X_n - X| > \varepsilon\} = o(1)$, $\varepsilon > 0$, then $\sup_{m > n} \mu\{|X_m - X_n| > \varepsilon\} = o(1)$, $\varepsilon > 0$, and there exists a subsequence X_{n_j} with $\mu\{\lim_{j \rightarrow \infty} X_{n_j} \neq X\} = 0$.
14. Let $\{X_n, n \geq 1\}$ be r.v.s such that $P\{|X_n| \geq c > 0\} \geq \delta > 0$, $n \geq 1$. If $\{a_n\}$ are finite constants for which $a_n X_n \xrightarrow{P} 0$, then $a_n \rightarrow 0$.

15. If r.v.s $X_n \xrightarrow{a.s.} X$, finite, prove that for every $\varepsilon > 0$ there is a set A_ε with $P\{A_\varepsilon\} < \varepsilon$ such that $\lim X_n(\omega) = X(\omega)$ uniformly on A_ε^c . This is known as **Egorov's theorem**. (*Hint:* Verify that if $A_{n,k} = \bigcap_{j=n}^{\infty} \{|X_j - X| < 2^{-k}\}$ and $A = \{\lim X_n = X\}$, then $\lim_{n \rightarrow \infty} P\{A_{n,k}\} = P\{\lim_{n \rightarrow \infty} A_{n,k}\} \geq P\{A\} = 1$, whence for some integers n_k $P\{A_{n_k,k}^c\} < \varepsilon/2^k$. Take $A_\varepsilon = \bigcup_{k=1}^{\infty} A_{n_k,k}^c$). State and prove the converse.
16. If $\{X_n, n \geq 1\}$ are independent r.v.s, $S_{m,n} = \sum_{j=m+1}^n X_j$, $S_n = S_{0,n}$, then for any $\varepsilon > 0$

$$P\left\{\max_{1 \leq j \leq n} |S_j| > 2\varepsilon\right\} \leq \frac{P\{|S_n| > \varepsilon\}}{\min_{1 \leq j \leq n} P\{|S_{j,n}| \leq \varepsilon\}}.$$

This is **Ottaviani's inequality**. *Hint:* If $T = \inf\{j \geq 1 : |S_j| > 2\varepsilon\}$, then

$$\bigcup_{j=1}^T \{T = j, |S_{j,n}| \leq \varepsilon\} \subset \{|S_n| > \varepsilon\}.$$

17. If $\{X, X_n, n \geq 1\}$ are i.i.d. symmetric r.v.s and $S_n = \sum_1^n X_i$, then (i)

$$P\{S_n > x\} \geq \frac{n}{2} P\{X > 2x\} P^{n-1}\{X \leq 2x\}$$

for $x > 0$, and (ii) $P\{S_n > x\} \geq (n/2)P\{X > x\}[1 - (n-1)P\{X > x\}]$. Part (i) is untrue if all "twos" are deleted (Take $n = 2$, $x = \frac{1}{2}$, and $P\{X = \pm 1\} = \frac{1}{2}$). *Hint:* Apropos of (ii), define $T = \inf\{j \geq 1 : X_j > x\}$.

18. Let $S_n = \sum_1^n X_i$ where $\{X_n, n \geq 1\}$ are independent r.v.s and suppose that $\lim_{n \rightarrow \infty} P\{S_{n-1} \geq -\delta n\} > 0$, all $\delta > 0$. Then $\overline{\lim} S_n/n \leq C < \infty$, a.e. implies $\sum_{n=1}^{\infty} P\{X_n > \epsilon n\} < \infty$, all $\epsilon > C$.

3.4 Bernoulli Trials

A sequence $\{X_n, n \geq 1\}$ of i.i.d. r.v.s with $P\{X_n = 1\} = p \in (0, 1)$ and $P\{X_n = -1\} = q = 1 - p$ constitutes a **sequence of Bernoulli trials with parameter p** . Define $S_n = \sum_1^n X_i, n \geq 1$. If $Y_n = (X_n + 1)/2$, clearly $\{Y_n, n \geq 1\}$ is a sequence of Bernoulli trials with success probability p (Section 1) and so $\{(S_n + n)/2, n \geq 1\}$ is a sequence of binomial r.v.s. Thus, the DeMoivre-Laplace, Bernoulli, and Borel theorems all pertain to Bernoulli trials with parameter p .

According to the intuitive notion of fairness, a sequence of tosses of a fair coin should at any stage n assign equal probabilities to each of the 2^n n -tuples of outcomes. If a gambler bets one dollar on correctly guessing each individual outcome and X_i denotes his gain (or loss) at the i th toss, this is tantamount to requiring that $P\{X_1 = \pm 1, \dots, X_n = \pm 1\} = 2^{-n}$ for each of the 2^n choices of sign, $n \geq 1$. Thus, his cumulative gain (or loss) after n tosses is $S_n = \sum_1^n X_j$, where $\{X_n, n \geq 1\}$ are Bernoulli trials with parameter $\frac{1}{2}$. The graph of $S_n, n \geq 1$, shows the random cumulative fortunes (gains) of the gambler as a function of n (which may be envisaged as time), and the fortunes $S_n, n \geq 1$, are said to undergo or constitute a **random walk**.

The distribution of the length of time T_k to achieve a gain of k dollars is implicit in Theorem 1 (Exercise 5) while the limit distribution of the "first passage time" T_k as well as that of $\max_{1 \leq j \leq n} S_j$ appear in Theorem 2.

The same limit distributions are shown in Chapter 9 to hold for a large class of **random walks**, i.e., sequence of partial sums S_n , $n \geq 1$, of i.i.d. random variables.

Clearly the r.v.s $\{X_n, n \geq 1\}$ constituting Bernoulli trials with parameter $p = \frac{1}{2}$ are independent symmetric r.v.s, and so by Corollary 3.2.3 the joint d.f.s of (X_1, \dots, X_n) and $(-X_1, \dots, -X_n)$ are identical and Theorem 1.6.1 guarantees

$$P\{(X_1, \dots, X_n) \in B\} = P\{(-X_1, \dots, -X_n) \in B\}$$

for any Borel set B of R^n .

Theorem 1. If $\{X_n, n \geq 1\}$ are i.i.d. with $P\{X_1 = 1\} = P\{X_1 = -1\} = \frac{1}{2}$ and $S_n = \sum_{i=1}^n X_i$, then for every positive integer N :

$$P\left\{\max_{1 \leq j \leq n} S_j \geq N, S_n < N\right\} = P\{S_n > N\}; \quad (1)$$

$$P\left\{\max_{1 \leq j \leq n} S_j \geq N\right\} = 2 P\{S_n \geq N\} - P\{S_n = N\}; \quad (2)$$

$$\begin{aligned} P\left\{\max_{1 \leq j \leq n} S_j = N\right\} &= P\{S_n = N\} + P\{S_n = N + 1\} \\ &= 2^{-n} \binom{n}{[(n+N+1)/2]}, \end{aligned} \quad (3)$$

where $[\lambda]$ is the integral part of λ if $\lambda \geq 0$, $[-\lambda] = -[\lambda]$ if $\lambda \geq 0$;

$$\begin{aligned} P\{S_j \neq 0, 1 \leq j \leq n+1\} &= P\left\{\max_{1 \leq j \leq n} S_j \leq 0\right\} \\ &= P\{S_n = 0\} + P\{S_n = 1\} = 2^{-n} \binom{n}{[n/2]}; \end{aligned} \quad (4)$$

$$P\{S_1 \neq 0, \dots, S_n \neq 0, S_{n+1} = 0\} = P\left\{\max_{1 \leq j \leq n+1} S_j \leq 0, S_n > 0\right\}. \quad (5)$$

PROOF. Define T to be the smallest integer j in $[1, n]$ for which $S_j = N$ if such j exists and $T = n + 1$ otherwise. Then $[T = k] \in \sigma(X_1, \dots, X_k)$ and $S_T = N$ on $[T \leq n]$ since $N \geq 1$. Hence, in view of independence,

$$\begin{aligned} P\left\{\max_{1 \leq j \leq n} S_j \geq N, S_n < N\right\} &= P\{T \leq n, S_n < N\} = P\{T < n, S_n < N\} \\ &= \sum_{k=1}^{n-1} P\left\{T = k, \sum_{i=k+1}^n X_i < 0\right\} \end{aligned}$$

$$\begin{aligned}
&= \sum_{k=1}^{n-1} P\{T = k\} P\left\{\sum_{i=k+1}^n X_i < 0\right\} \\
&= \sum_{k=1}^{n-1} P\{T = k\} P\left\{\sum_{i=k+1}^n X_i > 0\right\} \\
&= \sum_{k=1}^{n-1} P\{T = k, S_n > N\} \\
&= P\{T < n, S_n > N\} = P\{S_n > N\},
\end{aligned}$$

yielding (1). To obtain (2), note that via (1)

$$\begin{aligned}
P\left\{\max_{1 \leq j \leq n} S_j \geq N\right\} &= P\left\{\max_{1 \leq j \leq n} S_j \geq N, S_n < N\right\} + P\left\{\max_{1 \leq j \leq n} S_j \geq N, S_n \geq N\right\} \\
&= P\{S_n > N\} + P\{S_n \geq N\} = 2 P\{S_n \geq N\} - P\{S_n = N\}.
\end{aligned}$$

The first equality of (3) follows via (2) since

$$\begin{aligned}
P\left\{\max_{1 \leq j \leq n} S_j = N\right\} &= P\left\{\max_{1 \leq j \leq n} S_j \geq N\right\} - P\left\{\max_{1 \leq j \leq n} S_j \geq N + 1\right\} \\
&= 2 P\{S_n \geq N\} - 2 P\{S_n \geq N + 1\} - P\{S_n = N\} \\
&\quad + P\{S_n = N + 1\} \\
&= P\{S_n = N\} + P\{S_n = N + 1\}.
\end{aligned}$$

In proving the second equality of (3), if $n + N = 2m$ for some $m = 0, 1, \dots$,

$$P\{S_n = N\} = P\left\{\frac{S_n + n}{2} = \frac{n + N}{2}\right\} = \binom{n}{m} 2^{-n} = \binom{n}{[(n + N)/2]} 2^{-n}, \quad (6)$$

and clearly $P\{S_n = N + 1\} = 0$. Similarly, if $n + N + 1 = 2m$ for some $m = 0, 1, \dots$ then (6) holds with N replaced by $N + 1$.

Apropos of (4),

$$\begin{aligned}
P\{S_1 \neq 0, \dots, S_n \neq 0\} &= P\{S_1 = -1, S_2 \leq -1, \dots, S_n \leq -1\} \\
&\quad + P\{S_1 = 1, S_2 \geq 1, \dots, S_n \geq 1\} \\
&= 2 P\{S_1 = -1, S_2 \leq -1, \dots, S_n \leq -1\} \\
&= 2 P\{X_1 = -1, X_2 \leq 0, \dots, X_2 + \dots + X_n \leq 0\} \\
&= P\{X_2 \leq 0, \dots, X_2 + \dots + X_n \leq 0\} \\
&= P\{X_1 \leq 0, \dots, X_1 + \dots + X_{n-1} \leq 0\} \\
&= P\left\{\max_{1 \leq j \leq n-1} S_j \leq 0\right\},
\end{aligned}$$

which is tantamount to the first equality of (4). To obtain the second, note that via (2)

$$\begin{aligned} \mathbb{P}\left\{\max_{1 \leq j \leq n} S_j \leq 0\right\} &= 1 - \mathbb{P}\left\{\max_{1 \leq j \leq n} S_j \geq 1\right\} = 1 - 2 \mathbb{P}\{S_n \geq 1\} + \mathbb{P}\{S_n = 1\} \\ &= \mathbb{P}\{S_n = 0\} + \mathbb{P}\{S_n = 1\}. \end{aligned}$$

The last equality of (4) derives from

$$\mathbb{P}\{S_{2m} = 0\} = \binom{2m}{m} 2^{-2m}, \quad \mathbb{P}\{S_{2m+1} = 1\} = \binom{2m+1}{m} 2^{-2m-1}.$$

Finally, to obtain (5), note that via (4)

$$\begin{aligned} \mathbb{P}\{S_1 \neq 0, \dots, S_n \neq 0, S_{n+1} = 0\} &= \mathbb{P}\{S_1 \neq 0, \dots, S_n \neq 0\} \\ &\quad - \mathbb{P}\{S_1 \neq 0, \dots, S_{n+1} \neq 0\} \\ &= \mathbb{P}\left\{\max_{1 \leq j \leq n-1} S_j \leq 0\right\} \\ &\quad - \mathbb{P}\left\{\max_{1 \leq j \leq n} S_j \leq 0\right\} \\ &= \mathbb{P}\left\{\max_{1 \leq j \leq n-1} S_j \leq 0, S_n > 0\right\}. \quad \square \end{aligned}$$

Next, a local limit theorem (7) and global limit (8) will be obtained for $\bar{S}_n = \max_{1 \leq j \leq n} S_j$. According to (8), the d.f. of $n^{-1/2} \bar{S}_n$ tends as $n \rightarrow \infty$ to the **positive normal** d.f. The name arises from the fact that if X is $N(0, 1)$ then $|X|$ has the d.f. $2\Phi(x) - 1$, $x > 0$.

Theorem 2. Let $\{X_n, n \geq 1\}$ be i.i.d. r.v.s with $\mathbb{P}\{X_1 = 1\} = \mathbb{P}\{X_1 = -1\} = \frac{1}{2}$ and partial sums $S_n = \sum_1^n X_i$, $n \geq 1$, $S_0 = 0$. If $\bar{S}_n = \max_{1 \leq j \leq n} S_j$ and $T_k = \inf\{n \geq 1 : S_n = k\}$, $k \geq 1$ then for any sequence of positive integers N_n with $N_n = o(n^{2/3})$

$$\mathbb{P}\{\bar{S}_n = N_n\} \sim \left(\frac{2}{\pi n}\right)^{1/2} e^{-N_n^2/2n}, \quad (7)$$

$$\lim_{n \rightarrow \infty} \mathbb{P}\{\bar{S}_n < xn^{1/2}\} = 2\Phi(x) - 1 = \left(\frac{2}{\pi}\right)^{1/2} \int_0^x e^{-u^2/2} du, \quad x > 0, \quad (8)$$

$$\lim_{k \rightarrow \infty} \mathbb{P}\{T_k < xk^2\} = 2[1 - \Phi(x^{-1/2})], \quad x > 0. \quad (9)$$

PROOF. By (3) of Theorem 1

$$\mathbb{P}\{\bar{S}_n = N\} = \mathbb{P}\{S_n = N\} + \mathbb{P}\{S_n = N + 1\},$$

and so, by DeMoivre–Laplace local central limit theorem (Lemma 2.3.2), if $N_n + n$ is even,

$$P\{\bar{S}_n = N_n\} = P\{S_n = N_n\} = b\left(\frac{N_n + n}{2}; n, \frac{1}{2}\right) \sim \frac{2}{(2\pi n)^{1/2}} e^{-N_n^2/2n},$$

and similarly when $N_n + n$ is odd.

Next, if $x > 0$, setting $N = [xn^{1/2}]$ and employing (2) and Lemma 2.3.2

$$\begin{aligned} P\{\bar{S}_n \leq xn^{1/2}\} &= P\{\bar{S}_n \leq N\} = 1 - P\{\bar{S}_n \geq N + 1\} \\ &= 1 - 2P\{S_n \geq N + 1\} + P\{S_n = N + 1\} \\ &= 2P\{S_n \leq N\} - 1 + o(1). \end{aligned} \quad (10)$$

By the DeMoivre–Laplace global central limit theorem (Corollary 2.3.1)

$$P\{S_n \leq N\} = P\left\{\frac{S_n}{n^{1/2}} \leq x(1 + o(1))\right\} \rightarrow \Phi(x),$$

whence (8) follows from (10) and (7).

Finally, if $x > 0$, setting $n = [xk^2]$, via (8) and continuity of the normal d.f.

$$\begin{aligned} P\{T_k \leq xk^2\} &= P\{T_k \leq n\} = P\{\bar{S}_n \geq k\} \\ &\rightarrow 1 - [2\Phi(x^{-1/2}) - 1] = 2[1 - \Phi(x^{-1/2})], \end{aligned}$$

and since

$$P\{T_k = xk^2\} \leq P\{T_k = n\} \leq P\{S_n = k\} = o(1)$$

by the local central limit theorem, (9) follows. \square

Theorem 3. Let $\{X_n, n \geq 1\}$ be i.i.d. with $P\{X_1 = 1\} = P\{X_1 = -1\} = \frac{1}{2}$ and set $S_n = \sum_{i=1}^n X_i, n \geq 1, S_0 = 0$. Then

$$P\left\{\lim_{n \rightarrow \infty} S_n = \infty\right\} = 1 = P\left\{\lim_{n \rightarrow \infty} S_n = -\infty\right\}, \quad (11)$$

$$P\{S_n \text{ assumes every integer value, i.o.}\} = 1, \quad (12)$$

$$P\{S_n, n \geq 0 \text{ reaches } k \text{ before } -j\} = j/(j+k) \text{ for every pair of positive integers } j, k. \quad (13)$$

PROOF. Since, probabilistically speaking, the sequence $\{X_n, n \geq 2\}$ does not differ from $\{X_n, n \geq 1\}$, defining $q_j = P\{\sup_{n \geq 0} S_n \geq j\}, j \geq 0$, it follows that for $j \geq 1$

$$q_j = P\left\{\sup_{n \geq 1} S_n \geq j\right\}$$

$$\begin{aligned}
&= \frac{1}{2} P\left\{\sup_{n \geq 1} \sum_{i=2}^n X_i \geq j - 1\right\} + \frac{1}{2} P\left\{\sup_{n \geq 1} \sum_{i=2}^n X_i \geq j + 1\right\} \\
&= \frac{1}{2} P\left\{\sup_{n \geq 0} S_n \geq j - 1\right\} + \frac{1}{2} P\left\{\sup_{n \geq 1} S_n \geq j + 1\right\} = \frac{1}{2}(q_{j-1} + q_{j+1}).
\end{aligned}$$

Hence, for $j \geq 1$, $q_j - q_{j-1} = q_{j+1} - q_j = \text{constant} = c$ (say). Therefore, for $j \geq 1$, $q_j = cj + q_0$. Since $0 \leq q_j \leq 1$, necessarily $c = 0$, whence $q_j = q_0 = 1$ for $j \geq 1$. That is, $P\{\sup_{n \geq 1} S_n \geq j\} = 1$ for every $j \geq 1$, whence

$$P\left\{\sup_{n \geq 1} S_n = \infty\right\} = P\left\{\overline{\lim}_{n \rightarrow \infty} S_n = \infty\right\} = 1.$$

By symmetry, $P\{\underline{\lim}_{n \rightarrow \infty} S_n = -\infty\} = 1$ and (11) is proved.

Next, if $A = \{\sup_{n \geq 1} S_n = \infty, \inf_{n \geq 1} S_n = -\infty\}$, then by (11), $P\{A\} = 1$, necessitating (12).

To prove (13), set $r = j + k$ and

$$y_i = P\{\{S_n, n \geq 0\} \text{ reaches } i \text{ before } i - r\}, \quad i \geq 0.$$

Then $y_0 = 1$, $y_r = 0$. For $0 < i < r$

$$\begin{aligned}
y_i &= P\{X_1 = 1, \{S_n - X_1, n \geq 1\} \text{ reaches } i - 1 \text{ before } i - r - 1\} \\
&\quad + P\{X_1 = -1, \{S_n - X_1, n \geq 1\} \text{ reaches } i + 1 \text{ before } i - r + 1\} \\
&= py_{i-1} + qy_{i+1},
\end{aligned} \tag{14}$$

where $p = q = \frac{1}{2}$.

As earlier, $y_i = c + y_{i-1} = ci + y_0$, $1 \leq i \leq r$. Since $y_0 = 1$, $y_r = 0$, necessarily $y_i = 1 - (i/r) = (r - i)/r$, $0 \leq i \leq r$, and $y_k = (r - k)/r = j/(j + k)$. \square

Theorem 4. Let $\{X_n, n \geq 1\}$ be a sequence of Bernoulli trials with parameter $p \neq \frac{1}{2}$. If the partial sums are defined by $S_n = \sum_1^n X_i$, $n \geq 1$, and $S_0 = 0$, then for any positive integers j, k

$$P\{\text{partial sums } \{S_n, n \geq 0\} \text{ reach } k \text{ before } -j\} = \frac{(p/q)^k - (p/q)^{k+j}}{1 - (p/q)^{k+j}}. \tag{15}$$

PROOF. Set $r = j + k$ and $s = p/q$, where, as usual, $q = 1 - p$, and define for any integer $i \geq 0$

$$y_i = P\{\{S_n, n \geq 0\} \text{ reaches } i \text{ before } i - r\}.$$

They $y_0 = 1$, $y_r = 0$, and (14) obtains but with $p \neq \frac{1}{2}$. Hence, for $0 < i < r$

$$p(y_i - y_{i-1}) = q(y_{i+1} - y_i) \quad \text{or} \quad y_{i+1} - y_i = s(y_i - y_{i-1}).$$

Thus, for $0 < i < r$

$$y_{i+1} - y_i = s^2(y_{i-1} - y_{i-2}) = \cdots = s^i(y_1 - y_0), \tag{16}$$

and clearly (16) holds for $i = 0$. Since $s \neq 1$, for $0 < i \leq r$

$$y_i - y_0 = \sum_{m=1}^i (y_m - y_{m-1}) = \sum_{m=1}^i s^{m-1}(y_1 - y_0) = \frac{1-s^i}{1-s}(y_1 - y_0). \quad (17)$$

Taking $i = r$ in (17) reveals that $-(1-s) = (1-s^r)(y_1 - y_0)$, and hence

$$y_i - y_0 = \frac{-(1-s^i)}{(1-s^r)} \quad \text{or} \quad y_i = \frac{s^i - s^r}{1-s^r} = \frac{s^i - s^{j+k}}{1-s^{j+k}},$$

yielding (15) for $i = k$. \square

When $p = q$, the right side of (15) becomes 0/0 and by l'Hospital's rule

$$\frac{s^k - s^{j+k}}{1 - s^{j+k}} \rightarrow \frac{j}{j+k} \quad \text{as } s \rightarrow 1.$$

If it were known that the left side of (15) was a continuous function of p , then (15) would imply (13).

EXAMPLE 1. Suppose that a gambler A with a capital of j dollars and his adversary B whose capital is k dollars play the following game. A coin is tossed repeatedly (sequentially) and at each toss A wins one dollar if the outcome is a head while B wins a dollar if the outcome is a tail. The game terminates when one of the players is ruined, i.e., when either A loses j dollars or B loses k dollars. If the probability of heads is $p \in (0, 1)$, then by Theorems 3 and 4

$$\begin{aligned} P\{A \text{ ultimately wins}\} &= \frac{j}{j+k} && \text{if } p = q \\ &= \frac{s^k - s^{j+k}}{1 - s^{j+k}} && \text{if } s = \frac{p}{q} \neq 1. \end{aligned} \quad (18)$$

Interchanging p with q and j with k ,

$$\begin{aligned} P\{B \text{ ultimately wins}\} &= \frac{k}{j+k} && \text{if } p = q \\ &= \frac{s^{-j} - s^{-(j+k)}}{1 - s^{-(j+k)}} = \frac{1 - s^k}{1 - s^{j+k}} && \text{if } s = \frac{p}{q} \neq 1, \end{aligned} \quad (19)$$

and so for all $p \in (0, 1)$

$$P\{A \text{ ultimately wins}\} + P\{B \text{ ultimately wins}\} = 1,$$

that is, the game terminates (in a finite number of tosses) with probability one.

If $p \leq q$, that is, $s \leq 1$ and B has infinite capital ($k = \infty$), then letting $k \rightarrow \infty$ in (18) and (19)

$$P\{A \text{ ultimately wins}\} = 0,$$

$$P\{B \text{ ultimately wins}\} = 1,$$

whereas if $p > q$, that is, $s > 1$ and B has infinite capital,

$$\mathbb{P}\{\text{games terminates}\} = \mathbb{P}\{B \text{ ultimately wins}\} = s^{-j}.$$

The next result demonstrates that when the gamblers A and B have the same initial capital r , the duration of the game and the final capital of A are independent random variables. The duration of the game foreshadows the notion of a finite stopping time (Chapter 5.3).

EXAMPLE 2 (S. Samuels). Let $S_n = \sum_{j=1}^n X_j$ where $\{X_j, j \geq 1\}$ are Bernoulli trials with parameter $p \in (0, 1)$ and define $T = \inf\{n \geq 1 : |S_n| = r\}$, $r > 0$ and $T = \infty$ otherwise. If $U = r + S_T$ where $S_T = \sum_{n=1}^{\infty} S_n \cdot I_{\{T=n\}}$, then U and T are independent random variables.

PROOF. Note that according to Example 1, $\mathbb{P}\{T < \infty\} = 1$. Let $C(n, r) =$ number of n -tuples (X_1, \dots, X_n) with $S_n = r$, $|X_j| < r$, $1 \leq j < n$. Then

$$\mathbb{P}\{T = n, U = 2r\} = \mathbb{P}\{T = n, S_n = r\} = C(n, r)p^{(n+r)/2}q^{(n-r)/2}, \quad n \geq r > 0.$$

and by symmetry

$$\mathbb{P}\{T = n, U = 0\} = \mathbb{P}\{T = n, S_n = -r\} = C(n, r)p^{(n-r)/2}q^{(n+r)/2}.$$

Hence,

$$\mathbb{P}\{T = n\} = [1 + (p/q)^r] \cdot \mathbb{P}\{T = n, U = 0\}, \quad n \geq r > 0 \quad (20)$$

and so

$$\mathbb{P}\{U = 0\} = \sum_{n=r}^{\infty} \mathbb{P}\{T = n, U = 0\} = [1 + (p/q)^r]^{-1}.$$

Consequently, according to (20)

$$\mathbb{P}\{T = n, U = 0\} = \mathbb{P}\{T = n\} \cdot \mathbb{P}\{U = 0\}, \quad n \geq r > 0$$

which, in turn, implies

$$\mathbb{P}\{T = n, U = 2r\} = \mathbb{P}\{T = n\} \cdot \mathbb{P}\{U = 2r\}, \quad n \geq r > 0. \quad \square$$

EXERCISES 3.4

- Let $S_n = \sum_1^n X_i$, $n \geq 1$, where $\{X_n, n \geq 1\}$ are i.i.d. with $\mathbb{P}\{X_1 = 1\} = \mathbb{P}\{X_1 = -1\} = \frac{1}{2}$. If $T_i = \inf\{n \geq 1 : S_n = i\}$, $i = 1, 2$, then
 - $\mathbb{P}\{T_1 < \infty\} = 1$, $\sum_{n=1}^{\infty} \mathbb{P}\{T_1 > n\} = \infty$.
 - T_1 and $T_2 - T_1$ are i.i.d.
- Show that (a) and (b) of Exercise 1 hold if rather

$$T_1 = \inf\{n \geq 1 : S_n = 0\} \quad \text{and} \quad T_2 = \inf\{n > T_1 : S_n = 0\}.$$
- Let $S_n = \sum_1^n a_i X_i$, where $\{a_i, i \geq 1\}$ are constants and $\{X_n, n \geq 1\}$ are as in Exercise 1. If $S_n \xrightarrow{a.s.} S$, where $|S| \leq M < \infty$, a.e., prove via Lévy's inequality that (i) $\sup_{n \geq 1} |S_n| \leq M$, a.e., and, moreover, by an extension of Exercise 3.3.12 that (ii) $\sum_1^{\infty} |a_n| \leq M$.

4. Let $\{X_n, n \geq 1\}$ be i.i.d. with $P\{X_1 = 1\} = P\{X_1 = -1\} = \frac{1}{2}$ and $S_n = \sum_1^n X_i$, $n \geq 1$. For positive integers i and j , let T be the smallest positive integer for which $S_n = i$ or $-j$. Then

$$i P\{S_T = i\} - j P\{S_T = j\} = 0, \quad i^2 P\{S_T = i\} + j^2 P\{S_T = j\} = i \cdot j,$$

where $S_T = S_m$ on the set $\{T = m\}$, $m \geq 1$.

5. If in Bernoulli trials with parameter $\frac{1}{2}$, $T_k = \inf\{n \geq 1 : S_n = k\}$, prove that

$$P\{T_k = n\} = 2^{-n} \left\{ \binom{n-1}{[(n-k)/2]} - \binom{n-1}{[(n-k-1)/2]} \right\} \text{ for } n \geq k.$$

Hint: Apply (2) to $P\{T_k \leq n\} = P\{\max_{1 \leq j \leq n} S_j \geq k\}$.

6. If S_n is as in Exercise 1, prove for $N = 0, 1, \dots$ that

$$\begin{aligned} P\left\{\max_{j \leq n} S_j \leq N, S_{n+1} > N\right\} &= 2P\{S_{n+1} \geq N+1\} - P\{S_{n+1} = N+1\} \\ &\quad - 2P\{S_n \geq N+1\} + P\{S_n = N+1\}. \end{aligned}$$

7. If $S_n = \sum_1^n X_j$, $n \geq 1$ where $\{X_n, n \geq 1\}$ are Bernoulli trials with parameter p , verify that $P\{\max_{j \leq n} (S_j - jp) > x\} \leq c P\{S_n - np > x\}$ where $c > 0$.

8. Let $\{A_n, n \geq 1\}$ and $\{B_n, n \geq 1\}$ be events such that for all large m and $n > m$, A_n is independent of $B_n(\bigcup_{j=m}^{n-1} A_j B_j)$. Then $\sum_{n=1}^{\infty} P\{A_n\} = \infty$ implies $P\{A_n B_n, \text{ i.o.}\} \geq \lim_{n \rightarrow \infty} P\{B_n\}$. *Hint:* Use the techniques of Theorem 3.2.1 and Lemma 3.3.3.

References

- O. Barndorff-Nielsen, "On the rate of growth of the partial maxima of a sequence of independent, identically distributed random variables," *Math. Scand.* **9** (1961), 383–394.
- L. Baum, M. Katz, and H. Stratton, "Strong Laws for Ruled Sums," *Ann. Math. Stat.* **42** (1971), 625–629.
- F. Cantelli, "Su due applicazioni di un teorema di G. Boole," *Rend. Accad. Naz. Lincei* **26** (1917).
- K. L. Chung, "The strong law of large numbers," *Proc. 2nd Berkeley Symp. Stat. and Prob.* (1951), 341–352.
- K. L. Chung, *Elementary Probability Theory with Stochastic Processes*, Springer-Verlag, Berlin, New York, 1974.
- J. L. Doob, *Stochastic Processes*, Wiley, New York, 1953.
- W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd ed., Wiley, New York, 1950.
- A. Kolmogorov, *Foundations of Probability*, (Nathan Morrison, translator), Chelsea, New York, 1950.
- P. Lévy, *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, Paris, 1937; 2nd ed., 1954.
- M. Loève, *Probability Theory*, 3rd ed. Van Nostrand, Princeton, 1963; 4th ed., Springer-Verlag, Berlin and New York, 1977–1978.
- R. von Mises, "Über aufteilungs und Besetzungs Wahrscheinlichkeiten," *Revue de la Faculté des Sciences de l'Université d'Istanbul*, N.S. **4** (1939), 145–163.
- A. Renyi, *Foundations of Probability*, Holden-Day, San Francisco, 1970.

4 Integration in a Probability Space

4.1 Definition, Properties of the Integral, Monotone Convergence Theorem

There are two basic avenues to integration. In the modern approach the integral is introduced first for simple functions—as a weighted average of the values of the function—and then defined for any nonnegative measurable function f as a limit of the integrals of simple nonnegative functions increasing to f . Conceptually this is extremely simple, but a certain price is paid in terms of proofs. The alternative classical approach, while employing a less intuitive definition, achieves considerable simplicity in proofs of elementary properties.

If X is a (real) measurable function on a probability space (Ω, \mathcal{F}, P) , the **integral** of X over Ω with respect to P is denoted by $\int_{\Omega} X dP$, abbreviated by $E X$ or $E[X]$, referred to as the **expectation** or **mean** of X and defined (when possible) by:

- i. If $X \geq 0$, then $E X = \infty$ when $P\{X = \infty\} > 0$, while if $P\{X = \infty\} = 0$,

$$E X = \lim_{n \rightarrow \infty} \sum_{i=1}^{\infty} \frac{i}{2^n} P \left\{ \frac{i}{2^n} < X \leq \frac{i+1}{2^n} \right\}. \quad (1)$$

- ii. For general X , if either $E X^+ < \infty$ or $E X^- < \infty$, then

$$E X = E X^+ - E X^-, \quad (2)$$

where $X^+ = \max(X, 0)$, $X^- = \max(-X, 0)$. In this case, the expectation of X is said to **exist** and $E X \in [-\infty, \infty]$, denoted by $|E X| \leq \infty$. If $|E X| < \infty$, X is called **integrable**.

- iii. If $E X^+ = E X^- = \infty$, $E X$ is undefined.

It is not difficult to see that the limit in (1) always exists, since setting

$$p_{n,i} = P\left\{\frac{i}{2^n} < X \leq \frac{i+1}{2^n}\right\}, \quad s_n = \sum_{i=1}^{\infty} \frac{i}{2^n} p_{n,i}, \quad (3)$$

additivity of P guarantees $p_{n,i} = p_{n+1,2i} + p_{n+1,2i+1}$, whence

$$s_n = \sum_{i=1}^{\infty} \frac{2i}{2^{n+1}} (p_{n+1,2i} + p_{n+1,2i+1}) \leq s_{n+1},$$

and so $\lim_{n \rightarrow \infty} s_n$ exists. Furthermore, $E X \geq s_n, n \geq 1$, for $X \geq 0$.

It is trivial to check that

$$\begin{aligned} E[1] &= 1, \quad 0 \leq E X \leq \infty \quad \text{when } X \geq 0, \\ X &\stackrel{\text{a.s.}}{=} 0 \quad \text{if } E|X| = 0, \quad P\{|X| < \infty\} = 1 \quad \text{if } E|X| < \infty, \\ E X &= E Y \text{ if } P\{X = Y\} = 1, \quad E[-X] = -E X \quad \text{if } |E X| \leq \infty, \end{aligned} \quad (4)$$

and easy to verify that if $X \geq 0$ and $P\{X = \infty\} = 0$,

$$\begin{aligned} E X &= \lim_{n \rightarrow \infty} \sum_{i=0}^{\infty} \frac{i+1}{2^n} P\left\{\frac{i}{2^n} < X \leq \frac{i+1}{2^n}\right\} \\ &= \lim_{n \rightarrow \infty} \sum_{i=0}^{\infty} \frac{i+1}{2^n} P\left\{\frac{i}{2^n} \leq X < \frac{i+1}{2^n}\right\} \\ &= \lim_{n \rightarrow \infty} \sum_{i=1}^{\infty} \frac{i}{2^n} P\left\{\frac{i}{2^n} \leq X < \frac{i+1}{2^n}\right\}. \end{aligned} \quad (5)$$

For example, the last line of (5) equals

$$\begin{aligned} &\lim_{n \rightarrow \infty} \left[\sum_{i=1}^{\infty} \frac{i}{2^n} P\left\{\frac{i}{2^n} < X < \frac{i+1}{2^n}\right\} + \sum_{i=1}^{\infty} \frac{i-1}{2^n} P\left\{X = \frac{i}{2^n}\right\} \right] \\ &= \lim_{n \rightarrow \infty} \left[\sum_{i=1}^{\infty} \frac{i}{2^n} P\left\{\frac{i}{2^n} < X < \frac{i+1}{2^n}\right\} + \sum_{i=0}^{\infty} \frac{i}{2^n} P\left\{X = \frac{i+1}{2^n}\right\} \right], \end{aligned}$$

which, in turn, coincides with the definition of $E X$.

Every integrable, symmetric r.v. has mean zero as follows from (2) and Exercise 3.3.3(ii), whereas the expectation of a nonintegrable symmetric r.v. is undefined.

A measurable function X is called **elementary** if for some sequence $\{A_n, n \geq 1\}$ of disjoint sets of \mathcal{F}

$$X = \sum_{n=1}^{\infty} x_n I_{A_n}, \quad (6)$$

where $-\infty \leq x_n \leq \infty$. Clearly, simple functions (Exercise 1.4.3) are elementary. An elementary function X always engenders a partition $\{B_n\}$ of Ω in \mathcal{F} (i.e., $B_n \in \mathcal{F}$, B_n disjoint, and $\bigcup_1^{\infty} B_n = \Omega$). It suffices to note that

$$X = \sum_{n=1}^{\infty} y_n I_{B_n}, \quad (6')$$

where $B_1 = (\bigcup_{n=1}^{\infty} A_n)^c$, $y_1 = 0$, $y_{n+1} = x_n$, $B_{n+1} = A_n$, $n \geq 1$.

The basic properties of the integral—linearity, order preservation, and monotone convergence—are embodied in

Theorem 1. Let $X, Y, X_n, Y_n, n \geq 1$, be measurable functions on a probability space (Ω, \mathcal{F}, P) .

- If $X = \sum_{n=1}^{\infty} x_n I_{A_n}$ is a nonnegative elementary function, then

$$E X = \sum_{n=1}^{\infty} x_n P\{A_n\}. \quad (7)$$

- If $X \geq 0$, a.c., there exist elementary functions X_n with $0 \leq X_n \uparrow X$ a.c. and $E X_n \uparrow E X$; moreover, if $P\{X = \infty\} = 0$, then X_n may be chosen so that $P\{X - X_n \leq 2^{-n}, n \geq 1\} = 1$.
- If $X \geq Y \geq 0$, a.c., then $E X \geq E Y \geq 0$ and for any a in $(0, \infty)$,

$$P\{X \geq a\} \leq \frac{1}{a} E X \quad (\text{Markov inequality}). \quad (8)$$

- If $0 \leq X_n \uparrow X$ a.c., then $E X_n \uparrow E X$ (monotone convergence theorem).
- If $E X, E Y, E X + E Y$ are all defined, then $E[X + Y]$ exists and

$$E[X + Y] = E X + E Y \quad (\text{additivity}). \quad (9)$$

- If $X \geq Y$, a.c., and $E X, E Y$ exist, then

$$E X \geq E Y \quad (\text{order preservation}). \quad (10)$$

- If $E X$ exists and Y is integrable, then for all finite real a, b , $E[aX + bY]$ exists and

$$E[aX + bY] = a E X + b E Y \quad (\text{linearity}). \quad (11)$$

PROOF. (i) If $P\{X = \infty\} > 0$, then $x_n = \infty$ for some n for which $P\{A_n\} > 0$ so that (7) obtains. If, rather, $P\{X = \infty\} = 0$, then $x_m = \infty$ for some m requires that $P\{A_m\} = 0$, whence $x_m P\{A_m\} = 0$. By setting $x'_n = 0$ on any $A_n \subset \{X = \infty\}$ and $x'_n = x_n$ otherwise, and defining $X' \equiv X \cdot I_{\{X < \infty\}} = \sum_{n=1}^{\infty} x'_n I_{A_n}$, it is evident that $P\{X' = X\} = 1$, $E X' = E X$, and $X'(\omega) < \infty$ for all ω . Let

$$I_{ni} = \left(\frac{i}{2^n}, \frac{i+1}{2^n} \right], \quad p_{ni} = P\{X' \in I_{ni}\} = \sum_{j: x'_j \in I_{ni}} P\{A_j\}$$

and note that for all $n \geq 1$

$$\sum_{i=1}^{\infty} \frac{i}{2^n} p_{ni} \leq \sum_{i=1}^{\infty} \sum_{j: x'_j \in I_{ni}} x'_j P\{A_j\} \leq \sum_{j=1}^{\infty} x'_j P\{A_j\}, \quad (12)$$

$$\sum_{i=0}^{\infty} \frac{i+1}{2^n} p_{ni} \geq \sum_{i=0}^{\infty} \sum_{j: x'_j \in I_{ni}} x'_j P\{A_j\} = \sum_{j=1}^{\infty} x'_j P\{A_j\}, \quad (13)$$

so that by prior observations

$$\mathbb{E} X = \mathbb{E} X' = \sum_1^{\infty} x'_j P\{A_j\} = \sum_1^{\infty} x_j P\{A_j\}.$$

To prove (ii), set

$$\varphi_{ni} = \frac{i}{2^n} I_{[i/2^n < X \leq (i+1)/2^n]}$$

and define the simple functions $X_n^{(1)}$ and elementary functions $X_n^{(2)}$ by

$$X_n^{(1)} = nI_{\{X>n\}} + \sum_{i=1}^{n \cdot 2^n - 1} \varphi_{ni}, \quad X_n^{(2)} = \sum_{i=1}^{\infty} \varphi_{ni}. \quad (14)$$

Then $X_n^{(1)} \uparrow X$, $X_n^{(2)} \uparrow X'$, and $X_n^{(j)} \leq X_{n+1}^{(j)}$, $j = 1, 2$; furthermore, via (i) $\mathbb{E} X_n^{(2)} = s_n \leq s_{n+1} = \mathbb{E} X_{n+1}^{(2)}$. Consequently, if $P\{X = \infty\} = 0$, $\mathbb{E} X_n^{(2)} \uparrow \mathbb{E} X$ according to the definition (1). Moreover, under the same proviso $P\{X = \infty\} = 0$, subadditivity guarantees

$$P\left\{\bigcup_{n=1}^{\infty} [X - X_n^{(2)} > 2^{-n}]\right\} \leq \sum_{n=1}^{\infty} P\{X - X_n^{(2)} > 2^{-n}\} = 0,$$

which is tantamount to the last assertion in (ii). On the other hand, if $P\{X = \infty\} > 0$, then $\mathbb{E} X_n^{(1)} \geq n P\{X = \infty\} \rightarrow \infty = \mathbb{E} X$ and it will follow from the initial portion of the proof of (iii) that $\mathbb{E} X_n^{(1)}$ is monotone.

Apropos of (iii), suppose first that $X = \sum_{i=1}^{\infty} x_i I_{A_i}$, $Y = \sum_{j=1}^{\infty} y_j I_{B_j}$, where $\{A_i\}$ and $\{B_j\}$ are each partitions of Ω in \mathcal{F} . Then

$$X = \sum_{i,j} x_i I_{A_i B_j}, \quad Y = \sum_{i,j} y_j I_{A_i B_j},$$

and $P\{X \geq Y \geq 0\} = 1$ entails $x_i \geq y_j \geq 0$ if $P\{A_i B_j\} > 0$. By (i)

$$\mathbb{E} X = \sum_{i,j} x_i P\{A_i B_j\} \geq \sum_{i,j} y_j P\{A_i B_j\} = \mathbb{E} Y.$$

For the general case it may be supposed that $P\{X = \infty\} = 0$, whence also $P\{Y = \infty\} = 0$. If $X_n = X_n^{(2)}$ as in (14) and Y_n is defined analogously via Y , then these elementary functions satisfy $P\{X_n \geq Y_n \geq 0\} = 1$, so that by the part already proved $\mathbb{E} X_n \geq \mathbb{E} Y_n$. Hence, by (ii), $\mathbb{E} X \geq \mathbb{E} Y$.

Consequently, the Markov inequality follows directly from $X \geq aI_{\{X \geq a\}}$ and (i).

To prove the monotone convergence theorem, note first in case $P\{X = \infty\} > 0$ that by the Markov inequality and Corollary 1.5.2

$$\mathbb{E} X_n \geq a P\{X_n > a\} \xrightarrow{n \rightarrow \infty} a P\{X > a\} \geq a P\{X = \infty\} \xrightarrow{n \rightarrow \infty} \infty,$$

which is fitting since $\mathbb{E} X = \infty$.

If rather $P\{X = \infty\} = 0$, note that by this same corollary as $m \rightarrow \infty$

$$\begin{aligned} P\left\{\frac{i}{2^n} < X_m \leq \frac{i+1}{2^n}\right\} &= P\left\{X_m > \frac{i}{2^n}\right\} - P\left\{X_m > \frac{i+1}{2^n}\right\} \\ &\rightarrow P\left\{X > \frac{i}{2^n}\right\} - P\left\{X > \frac{i+1}{2^n}\right\} \\ &= P\left\{\frac{i}{2^n} < X \leq \frac{i+1}{2^n}\right\}. \end{aligned} \quad (15)$$

For any $a < E X$, the definition of $E X$ ensures the existence of positive integers n, k such that

$$\sum_{i=1}^k \frac{i}{2^n} P\left\{\frac{i}{2^n} < X \leq \frac{i+1}{2^n}\right\} > a,$$

whence, via (15), for all large m

$$E X_m \geq \sum_{i=1}^k \frac{i}{2^n} P\left\{\frac{i}{2^n} < X_m \leq \frac{i+1}{2^n}\right\} > a. \quad (16)$$

By (iii), $E X_m \leq E X_{m+1} \leq E X$, which in conjunction with (16) yields $E X_m \uparrow E X$.

Apropos of (v), if $X = \sum_1^\infty x_i I_{A_i} \geq 0$, $Y = \sum_1^\infty y_j I_{B_j} \geq 0$, where $\{A_i\}$ and $\{B_j\}$ are partitions of Ω in \mathcal{F} , then $X + Y = \sum_{i,j} (x_i + y_j) I_{A_i B_j}$, yielding via (7) and σ -additivity

$$\begin{aligned} E[X + Y] &= \sum_{i,j} (x_i + y_j) P\{A_i B_j\} \\ &= \sum_i x_i P\{A_i\} + \sum_j y_j P\{B_j\} = E X + E Y. \end{aligned} \quad (17)$$

In conjunction with (ii), (17) yields additivity for $X \geq 0$, $Y \geq 0$.

In general, if $E X = \infty$, then $E X^+ = \infty$, $E X^- < \infty$, $E Y > -\infty$, $E Y^- < \infty$, so that $P\{X > -\infty, Y > -\infty\} = 1$. Now (Exercise 1.4.4), $(X + Y)^- \leq X^- + Y^-$ and $X^+ \leq (X + Y)^+ + Y^-$, whence by (iii) and the part already proved

$$E(X + Y)^- \leq E(X^- + Y^-) = E X^- + E Y^- < \infty,$$

$$E(X + Y)^+ + E Y^- = E[(X + Y)^+ + Y^-] \geq E X^+ = \infty,$$

implying $E[X + Y] = \infty$ and hence (9). Similarly (9) holds if $E X = -\infty$ or $E Y = -\infty$ or $E Y = \infty$.

Lastly, if $|E X| + |E Y| < \infty$, then $E X^+ < \infty$, $E X^- < \infty$, implying $E|X| = E X^+ + E X^- < \infty$ by the portion already proved. Similarly, $E|Y| < \infty$, whence $P\{|X| < \infty, |Y| < \infty\} = 1$ and

$$E|X + Y| \leq E|X| + E|Y| < \infty.$$

Thus, from

$$X^+ + Y^+ = (X + Y)^+ + [X^- + Y^- - (X + Y)^-],$$

by the part already proved

$$\mathbb{E} X^+ + \mathbb{E} Y^+ = \mathbb{E}[X^+ + Y^+] = \mathbb{E}(X + Y)^+ + \mathbb{E}[X^- + Y^- - (X + Y)^-],$$

yielding again via the additivity or subtractivity already proved

$$\mathbb{E} X + \mathbb{E} Y = \mathbb{E}(X + Y).$$

To dispatch (vi) in the nontrivial case $\mathbb{E} X < \infty$, $\mathbb{E} Y > -\infty$, note that via (v) and (iii)

$$\mathbb{E} X = \mathbb{E} Y + \mathbb{E}[X - Y] \geq \mathbb{E} Y.$$

Finally, apropos of (vii), let $0 < a < \infty$ and $X \geq 0$. If $\mathbb{P}\{X = \infty\} > 0$, then $\mathbb{E}[aX] = \infty = a \mathbb{E} X$. If $\mathbb{P}\{X = \infty\} = 0$, then, recalling (ii), $X_n^{(2)} = \sum_{i=1}^{\infty} \varphi_{ni} \uparrow X$, a.c., whence $0 \leq aX_n^{(2)} \uparrow aX$, a.c., and by (i) and (iv)

$$\mathbb{E}[aX] = \lim_n \mathbb{E}[aX_n^{(2)}] = \lim_n \mathbb{E}\left[\sum_{i=1}^{\infty} a\varphi_{ni}\right] = a \mathbb{E} X.$$

In general, for $0 < a < \infty$ and $|\mathbb{E} X| \leq \infty$

$$\mathbb{E}[aX] = \mathbb{E}(aX)^+ - \mathbb{E}(aX^-) = a \mathbb{E} X^+ - a \mathbb{E} X^- = a \mathbb{E} X.$$

Since $\mathbb{E}[0 \cdot X] = 0 \cdot \mathbb{E} X = 0$ and $\mathbb{E}[-X] = -\mathbb{E} X$,

$$\mathbb{E} aX = a \mathbb{E} X \tag{18}$$

for any finite a whenever $\mathbb{E} X$ exists. Finally, (11) follows via (9) and (18). \square

Note that if X, Z are measurable functions with $X \leq Z$ (respectively, $X \geq Z$) where $\mathbb{E} Z^+ < \infty$ (respectively, $\mathbb{E} Z^- < \infty$) then $\mathbb{E} X$ exists. A fortiori if X is bounded above or below by an integrable random variable, then $\mathbb{E} X$ exists.

Corollary 1. *A measurable function X on $(\Omega, \mathcal{F}, \mathbb{P})$ is integrable iff $|X|$ is integrable, and in such a case*

$$\mathbb{P}\{|X| \geq a\} \leq \frac{1}{a} \mathbb{E}|X|, \quad a > 0 \quad (\text{Markov inequality}) \tag{19}$$

and

$$|\mathbb{E} X| \leq \mathbb{E}|X|. \tag{20}$$

If X is a discrete r.v. with values $\{x_j, 1 \leq j \leq n\}$ and p.d.f. $\{p_j, 1 \leq j \leq n\}$, where $1 \leq n \leq \infty$, Theorem 1(i) ensures that

$$\mathbb{E} X = \sum_{j=1}^n x_j p_j \tag{21}$$

when $x_j \geq 0$. Consequently, (21) holds for any discrete r.v. provided that $\sum_{j=1}^{\infty} x_j^+ p_j$ or $\sum_{j=1}^{\infty} x_j^- p_j$ is finite when $n = \infty$. On the other hand, if X is an

absolutely continuous r.v. with density g and $E X$ exists,

$$E X = \int_{-\infty}^{\infty} t g(t) dt \quad (22)$$

according to Corollary 6.5.4.

In the case of an infinite series of nonnegative measurable functions, the operations of expectation and summation may be interchanged as follows directly from Lebesgue's monotone convergence theorem (Theorem 1(iv)):

Corollary 2. *If $\{X_n, n \geq 1\}$ are nonnegative measurable functions on (Ω, \mathcal{F}, P) , then*

$$E \sum_{n=1}^{\infty} X_n = \sum_{n=1}^{\infty} E X_n. \quad (23)$$

Without the nonnegativity proviso, (23) is, in general, false even when $\sum_{i=1}^{\infty} X_i$ converges, a.c.

EXAMPLE 1. Let $\{Y_n, n \geq 1\}$ be i.i.d. random variables with $P\{Y_n = \pm 1\} = \frac{1}{2}$ and define $T = \inf\{n \geq 1 : \sum_{i=1}^n Y_i = 1\}$ where $\inf\{\emptyset\} = \infty$. Then $T < \infty$, a.c. (Exercise 3.4.1), and, setting $X_n = Y_n I_{[T \geq n]}$,

$$\sum_{n=1}^{\infty} X_n = \sum_{n=1}^{\infty} Y_n I_{[T \geq n]} = \sum_{n=1}^T Y_n = 1$$

by definition of T , and so $E \sum_{n=1}^{\infty} X_n = 1$. However, since the event $\{T \geq n\} \in \sigma(Y_1, \dots, Y_{n-1})$, the r.v.s Y_n and $I_{[T \geq n]}$ are independent by Corollary 3.2.2, whence it follows from Theorem 4.3.3 that

$$E X_n = E Y_n E I_{[T \geq n]} = 0, \quad n \geq 1,$$

so that (23) fails.

Definition. Given any nonnegative constants $\{b_n, n \geq 0\}$, a continuous function $b(\cdot)$ on $[0, \infty)$ is called an **extension** of $\{b_n\}$ to $[0, \infty)$ if $b(n) = b_n$, $n \geq 0$. Moreover, when $\{b_n\}$ is strictly monotone, $b(\cdot)$ is a **strictly monotone extension** of $\{b_n\}$ if it is both strictly monotone and an extension of $\{b_n\}$.

Corollary 3. *Let $\{b_n, n \geq 0\}$ be a strictly increasing sequence with $0 \leq b_n \uparrow \infty$ and let $b(\cdot)$ be a strictly monotone extension of $\{b_n\}$ to $[0, \infty)$. Then for any r.v. $X \geq 0$, a.c.*

$$\sum_{n=1}^{\infty} P\{X \geq b_n\} \leq E b^{-1}(X) \leq \sum_{n=0}^{\infty} P\{X > b_n\}. \quad (24)$$

In particular, for any $r > 0$ and any r.v. X ,

$$\sum_{n=1}^{\infty} P\{|X| \geq n^{1/r}\} \leq E|X|^r \leq \sum_{n=0}^{\infty} P\{|X| > n^{1/r}\}. \quad (25)$$

$$\mathbb{E}|X|^r = r \int_0^\infty t^{r-1} \mathbb{P}\{|X| > t\} dt. \quad (26)$$

PROOF. Set $\varphi(x) = b^{-1}(x)$,

$$Y = \sum_1^\infty j I_{\{j \leq \varphi(X) < j+1\}}, \quad Z = \sum_0^\infty (j+1) I_{\{j < \varphi(X) \leq j+1\}},$$

whence $Y \leq \varphi(X) \leq Z$, a.c., since $\mathbb{P}\{X < \infty\} = 1 = \mathbb{P}\{\varphi(X) < \infty\}$. Thus,

$$\mathbb{E} Y \leq \mathbb{E} \varphi(X) \leq \mathbb{E} Z.$$

But

$$\begin{aligned} \sum_1^\infty \mathbb{P}\{X \geq b_n\} &= \sum_1^\infty \mathbb{P}\{\varphi(X) \geq n\} = \sum_{n=1}^\infty \sum_{j=n}^\infty \mathbb{P}\{j \leq \varphi(X) < j+1\} \\ &= \sum_{j=1}^\infty j \mathbb{P}\{j \leq \varphi(X) < j+1\} = \mathbb{E} Y \end{aligned}$$

and

$$\begin{aligned} \sum_0^\infty \mathbb{P}\{X > b_n\} &= \sum_{n=0}^\infty \sum_{j=n}^\infty \mathbb{P}\{j < \varphi(X) \leq j+1\} \\ &= \sum_{j=0}^\infty (j+1) \mathbb{P}\{j < \varphi(X) \leq j+1\} = \mathbb{E} Z, \end{aligned}$$

completing the proof of (24).

Clearly, (25) follows from (24) with $b(x) = x^r$, $r > 0$. Finally, it suffices to verify (26) when $r = 1$ since the statement for $r > 0$ then follows by a change of variable. Now, for any $a > 0$, (25) ensures

$$\begin{aligned} \int_1^\infty \mathbb{P}\{|aX| \geq u\} du &\leq \sum_{n=1}^\infty \mathbb{P}\{|aX| \geq n\} \leq a \cdot \mathbb{E}|X| \leq 1 + \sum_{n=1}^\infty \mathbb{P}\{|aX| > n\} \\ &\leq 1 + \int_0^\infty \mathbb{P}\{|aX| > u\} du \end{aligned}$$

whence setting $t = u/a$

$$\int_{a^{-1}}^\infty \mathbb{P}\{|X| \geq t\} dt \leq \mathbb{E}|X| \leq a^{-1} + \int_0^\infty \mathbb{P}\{|X| > t\} dt$$

and the conclusion follows as $a \rightarrow \infty$. \square

EXERCISES 4.1

- If X is a geometric r.v. with parameter p , that is, $\mathbb{P}\{X = k\} = pq^k$, $k = 0, 1, \dots$, $0 < p < 1$ verify that $\mathbb{E}X = q/p$, $\mathbb{E}(X-1) = 2(q/p)^2$, $\mathbb{E}(X-1)(X-2) = 6(q/p)^3$, $\mathbb{E}X^3 = 6(q/p)^3 + 6(q/p)^2 + q/p$ and $\mathbb{E}X^2 - (\mathbb{E}X)^2 = (q/p)^2 + q/p$.

2. If X is an integrable r.v., for every $\varepsilon > 0$ there is a simple function X_ε with $E|X - X_\varepsilon| < \varepsilon$.
3. Utilize Exercise 1.2.6 to give an alternative proof of Corollary 2.1.3.
4. (i) If $P\{X = 1\} = p$, $P\{X = 0\} = q = 1 - p$, prove that $E(X - EX)^k = pq[q^{k-1} - (-p)^{k-1}]$, $k = 1, 2, \dots$. (ii) if X is a r.v. with $\sum_{j=0}^{\infty} P\{X = j\} = 1$, then $E X = \sum_{j=1}^{\infty} P\{X \geq j\}$.
5. If X is an integrable r.v., $n P\{|X| \geq n\} = o(1)$, but the converse does not hold.
6. Construct a sequence of discrete r.v.s X_n such that $X_n \xrightarrow{P} 0$ but $E X_n \not\rightarrow 0$.
7. If $\{X, X_n, n \geq 1\}$ is a sequence of r.v.s on some probability space such that

$$\sum_{n=1}^{\infty} E|X_n - X|^r < \infty,$$

some $r > 0$, then $X_n \xrightarrow{a.e.} X$.

8. If (i) $\{X_n, n \geq 1\}$ is a sequence of nonnegative, integrable r.v.s with $S_n = X_1 + \dots + X_n$ and (ii) $\sum_{n=1}^{\infty} E X_n < \infty$, then S_n converges a.c. Hence, if (i) obtains, $E S_n > 0$, $n \geq 1$, and $\sum_{n=1}^{\infty} E X_n/E S_n < \infty$, then $S_n/E S_n$ converges a.c.
9. If X and Y are measurable functions on (Ω, \mathcal{F}, P) with $E XI_A = E YI_A$ for all $A \in \mathcal{F}$, then $X = Y$, a.c. Hint: Consider $A_{r,s} = \{\omega: \infty \geq X(\omega) \geq r > s \geq Y(\omega) \geq -\infty\}$.
10. If $\{X_n, n \geq 1\}$ are i.i.d. random variables then (i) $1/n \max_{1 \leq j \leq n} |X_j| \xrightarrow{a.e.} 0$ iff $E|X_1| < \infty$. (ii) $1/n \max_{1 \leq j \leq n} |X_j| \xrightarrow{P} 0$ iff $n \cdot P\{|X_1| > n\} \rightarrow 0$.

4.2 Indefinite Integrals, Uniform Integrability, Mean Convergence

For any measurable function X on a probability space (Ω, \mathcal{F}, P) whose expectation $E X$ exists, define

$$v\{A\} = E X \cdot I_A = \int_A X dP, \quad A \in \mathcal{F}. \quad (1)$$

The set function $v\{A\}$ defined on \mathcal{F} by (1) is called the **indefinite integral of X** .

Moreover, any nonnegative measurable function X on (Ω, \mathcal{F}, P) generates a new measure v on \mathcal{F} via (1), and if $E X = 1$, this will be a probability measure.

Lemma 1. *The indefinite integral v of any measurable function X on (Ω, \mathcal{F}, P) whose expectation exists is a σ -additive set function on \mathcal{F} , that is, for disjoint $A_j \in \mathcal{F}, j \geq 1$,*

$$v\left\{\bigcup_j A_j\right\} = \sum_j v\{A_j\}.$$

PROOF. By Corollary 4.1.2,

$$\begin{aligned}\sum_{j=1}^{\infty} \int_{A_j} X^+ dP &= \sum_{j=1}^{\infty} E X^+ I_{A_j} = E \sum_{j=1}^{\infty} X^+ I_{A_j} \\ &= E X^+ I_{\cup_{j=1}^{\infty} A_j} = \int_{\cup_{j=1}^{\infty} A_j} X^+ dP\end{aligned}$$

and, similarly,

$$\sum_{j=1}^{\infty} \int_{A_j} X^- dP = \int_{\cup_{j=1}^{\infty} A_j} X^- dP.$$

Thus, since $E X$ exists,

$$\begin{aligned}\sum_{j=1}^{\infty} \int_{A_j} X dP &= \sum_{j=1}^{\infty} \left(\int_{A_j} X^+ dP - \int_{A_j} X^- dP \right) \\ &= \int_{\cup_{j=1}^{\infty} A_j} X^+ dP - \int_{\cup_{j=1}^{\infty} A_j} X^- dP = \int_{\cup_{j=1}^{\infty} A_j} X dP. \quad \square\end{aligned}$$

Corollary 1. If X is a nonnegative measurable function on (Ω, \mathcal{F}, P) , its indefinite integral is a measure on \mathcal{F} .

The integrability of a measurable function X can be characterized via its indefinite integral according to

Lemma 2. A measurable function X is integrable iff for every $\varepsilon > 0$ there corresponds a $\delta > 0$ such that $A \in \mathcal{F}, P\{A\} < \delta$ implies

$$\int_A |X| dP < \varepsilon, \quad E|X| \leq \frac{1}{\delta}. \quad (2)$$

PROOF. If X is integrable and $X_k = |X| I_{\{|X| \leq k\}}$, then $X_k \uparrow |X|$, a.c., whence by Theorem 4.1.1(iv), $E X_k \uparrow E|X|$, which entails $E|X| I_{\{|X| > K\}} \rightarrow 0$ as $K \rightarrow \infty$. If K is a positive integer for which $E|X| I_{\{|X| > K\}} < \varepsilon/2$, set $\delta = \min(\varepsilon/2K, 1/E|X|)$. Then for $A \in \mathcal{F}$ with $P\{A\} < \delta$,

$$\int_A |X| dP = \int_A |X| I_{\{|X| > K\}} dP + \int_A |X| I_{\{|X| \leq K\}} dP \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

and so (2) holds.

Conversely, (2) implies $|X|$ and therefore X integrable. \square

This suggests the following

Definition. A sequence of r.v.s $\{X_n, n \geq 1\}$ is called **uniformly integrable** (abbreviated u.i.) if for every $\varepsilon > 0$ there corresponds a $\delta > 0$ such that

$$\sup_{n \geq 1} \int_A |X_n| dP < \varepsilon \quad (3)$$

whenever $P\{A\} < \delta$ and, in addition,

$$\sup_{n \geq 1} E|X_n| < \infty. \quad (4)$$

Furthermore, $\{X_n\}$ is said to be **u.i. from above or below** according as $\{X_n^+\}$ or $\{X_n^-\}$ is u.i.

Some immediate consequences of the definition are:

- i. $\{X_n\}$ is u.i. iff $\{|X_n|\}$ is u.i.
- ii. If $\{X_n\}$ and $\{Y_n\}$ are each u.i., so is $\{X_n + Y_n\}$.
- iii. If $\{X_n, n \geq 1\}$ is u.i., so is any subsequence of $\{X_n\}$.
- iv. $\{X_n\}$ is u.i. iff it is u.i. from above and from below.
- v. If $|X_n| \leq Y$ with $EY < \infty$, then $\{X_n\}$ is u.i.

An alternative characterization of uniform integrability appears in

Theorem 1 (u.i. criterion). *A sequence of r.v.s $\{X_n, n \geq 1\}$ is u.i. iff*

$$\lim_{a \rightarrow \infty} \sup_{n \geq 1} \int_{\{|X_n| > a\}} |X_n| dP = 0. \quad (5)$$

PROOF. If $\{X_n\}$ is u.i., then $\sup E|X_n| \leq C < \infty$, whence for any $\varepsilon > 0$, choosing δ as in (3), the Markov inequality ensures that

$$P\{|X_n| > a\} \leq a^{-1} E|X_n| \leq \frac{C}{a} < \delta, \quad n \geq 1,$$

provided $a > C/\delta$. Consequently, from (3) for $a > C/\delta$,

$$\sup_{n \geq 1} \int_{\{|X_n| > a\}} |X_n| dP < \varepsilon, \quad (6)$$

which is tantamount to (5). Conversely, for any $\varepsilon > 0$, choosing a sufficiently large to activate (6),

$$E|X_n| \leq a + \int_{\{|X_n| > a\}} |X_n| dP \leq a + \varepsilon, \quad n \geq 1,$$

yielding (4). Moreover, selecting $\delta = \varepsilon/a$, for any A with $P\{A\} < \delta$ and all $n \geq 1$

$$\begin{aligned} \int_A |X_n| dP &= \int_{A \cap \{|X_n| \leq a\}} |X_n| dP + \int_{A \cap \{|X_n| > a\}} |X_n| dP \\ &\leq a P\{A\} + \int_{\{|X_n| > a\}} |X_n| dP \leq \varepsilon + \varepsilon = 2\varepsilon, \end{aligned}$$

so that (3) holds. \square

The importance of uniform integrability will become apparent from Theorem 3.

Associated with any probability space (Ω, \mathcal{F}, P) are the \mathcal{L}_p spaces of all measurable functions X (necessarily r.v.s) for which $E|X|^p < \infty$, denoted by \mathcal{L}_p or $\mathcal{L}_p(\Omega, \mathcal{F}, P)$, $p > 0$. Random variables in \mathcal{L}_p will be dubbed \mathcal{L}_p r.v.s.

The inequalities of Section 1 show that a r.v. $X \in \mathcal{L}_p$ iff the series $\sum_1^\infty P\{|X| > n^{1/p}\}$ converges.

Definition. A sequence $\{X_n, n \geq 1\}$ of \mathcal{L}_p r.v.s is said to converge in mean of order p (to a r.v. X) if $E|X_n - X|^p \rightarrow 0$ as $n \rightarrow \infty$. This will be denoted by $X_n \xrightarrow{\mathcal{L}_p} X$. Convergence in mean of orders one and two are called convergence in mean and convergence in mean square (or quadratic mean) respectively.

Convergence in mean of order p for any $p > 0$ implies convergence in probability as follows from the Markov inequality via

$$P\{|X_n - X| \geq \varepsilon\} = P\{|X_n - X|^p \geq \varepsilon^p\} \leq \varepsilon^{-p} E|X_n - X|^p.$$

Moreover, convergence of X_n to X in mean of order p entails $X \in \mathcal{L}_p$ in view of the inequality

$$E|X|^p \leq 2^p [E|X_n|^p + E|X_n - X|^p],$$

the latter being an immediate consequence of

Lemma 3. If X and Y are nonnegative r.v.s and $p > 0$, then

$$E(X + Y)^p \leq 2^p [E X^p + E Y^p]. \quad (7)$$

PROOF. For $a > 0, b > 0$, $(a + b)^p \leq [2 \max(a, b)]^p \leq 2^p [a^p + b^p]$ and (7) follows.

In particular, $X \in \mathcal{L}_p, Y \in \mathcal{L}_p$ imply $X + Y \in \mathcal{L}_p$. \square

Among the most important and frequently used results of integration are Lebesgue's dominated convergence theorem (Corollary 3) and monotone convergence theorem, and Fatou's lemma. The second will be obtained by first verifying the latter.

Theorem 2 (i) (Monotone convergence theorem). If the r.v.s $\{X_n, n \geq 1\}$ are u.i. from below and $X_n \uparrow X$, a.c., then $E X^- < \infty$ and $E X_n \uparrow E X$.

(ii) (Fatou's lemma). If the r.v.s $\{X_n, n \geq 1\}$ are u.i. from below and $E \liminf X_n$ exists, then

$$E \liminf_{n \rightarrow \infty} X_n \leq \liminf_{n \rightarrow \infty} E X_n. \quad (8)$$

The most typical usage of Fatou's lemma and (i) would be as in

Corollary 2. If the r.v.s $\{X_n, n \geq 0\}$ satisfy $X_n \geq X_0$, a.c., $n \geq 1$, where X_0 is integrable (a fortiori, if $X_n \geq 0$, a.c., $n \geq 1$) then $E \lim_{n \rightarrow \infty} X_n$ exists, (8) holds, and, moreover, if $X_n \uparrow X$, a.c., then $E X_n \uparrow E X$.

PROOF OF THEOREM 2. (i) $X_n \uparrow X$ a.c. entails $X_n^- \downarrow X^-$, a.c., whence $\sup E X_n^- < \infty$ implies $E X^- < \infty$. Then, $0 \leq X_n + X_1^- \uparrow X + X_1^-$ whence, by Theorem 4.1 (iv), $E(X_n + X_1^-) \uparrow E(X + X_1^-)$ implying that $E X_n \uparrow E X$.

To prove (ii), define, for $K > 0$,

$$X'_n = X_n \wedge (-K), \quad X''_n = X_n - X'_n \quad (9)$$

and note that, for any $\varepsilon > 0$ and sufficiently large K ,

$$\inf_{n \geq 1} E X''_n = \inf_{n \geq 1} E(X_n + K) I_{\{X_n \leq -K\}} \geq \inf_{n \geq 1} E(-X_n^- I_{\{X_n^- \geq K\}}) \geq -\varepsilon. \quad (10)$$

Hence, setting $Y_n = \inf_{i \geq n} X'_i$, it follows from $-K \leq Y_n \uparrow \lim_{m \rightarrow \infty} X'_m$ and (10) that

$$E \lim_{m \rightarrow \infty} X'_m = \lim_{n \rightarrow \infty} E Y_n \leq \lim_{n \rightarrow \infty} E X'_n. \quad (11)$$

Now $X_n \leq X'_n$ and $E \lim_{n \rightarrow \infty} X_n$ exists by hypothesis, whence $E \lim_{n \rightarrow \infty} X_n \leq E \lim_{n \rightarrow \infty} X'_n$. Thus, by (9), (10), and (11)

$$\begin{aligned} \lim E X_n &= \lim E(X'_n + X''_n) \geq \lim E X'_n - \varepsilon \\ &\geq E \lim X'_n - \varepsilon \geq E \lim X_n - \varepsilon, \end{aligned}$$

and (8) obtains as $\varepsilon \rightarrow 0$. \square

The next lemma, especially when $A_n = B_n$, may be regarded as an extension of the Borel–Cantelli theorem which relaxes the independence requirement therein.

Lemma 4. If $\{A_n, n \geq 1\}$ and $\{B_n, n \geq 1\}$ are sequences of events satisfying

$$P\left\{A_i \cap \bigcup_{j=1}^{\infty} A_{i+jk}\right\} \leq P\{A_i\} P\left\{\bigcup_{j=k}^{\infty} B_j\right\} \quad (12)$$

for all large i and some positive integer k and

$$\sum_{i=1}^{\infty} P\{A_i\} = \infty, \quad (13)$$

then

$$P\left\{\bigcup_{j=k}^{\infty} B_j\right\} = 1. \quad (14)$$

Moreover, if (12) holds for infinitely many integers k and (13) obtains, then

$$P\{B_n, \text{i.o.}\} = 1. \quad (15)$$

PROOF. Replacing i by $i + nk$ and then setting $A_n^* = A_{n+1}^* = A_{i+nk}$ in (12),

$$P\left\{\bigcup_{j=k}^{\infty} B_j\right\} \geq P\left\{\bigcup_{j=1}^{\infty} A_{i+(n+jk)} | A_{i+nk}\right\} = P\left\{\bigcup_{h=n+1}^{\infty} A_h^* | A_n^*\right\}$$

for all large n , whence (14) follows from Theorem 3.1.3, noting that

$$\sum_{i=1}^k \sum_{n=1}^{\infty} P\{A_{i+nk}\} = \infty = \sum_{n=1}^{\infty} P\{A_{n,i}^*\}$$

for some i . The final statement is a consequence of the monotonicity in the left side of (14).

EXAMPLE 1. If $\{S_n = \sum_{i=1}^n X_i, n \geq 1\}$ where $\{X_n, n \geq 1\}$ are i.i.d. random variables, then for any $\varepsilon \geq 0$, $P\{|S_n| \leq \varepsilon, \text{i.o.}\} = 1$ or 0 according as

$$\sum_{n=1}^{\infty} P\{|S_n| \leq \varepsilon\}$$

diverges or converges.

PROOF. In view of the Borel–Cantelli lemma, it suffices to consider the case of divergence and here it may be supposed that $\sum_{n=1}^{\infty} P\{0 \leq S_n \leq \varepsilon\} = \infty$ since otherwise $\sum_{n=1}^{\infty} P\{0 \geq S_n \geq -\varepsilon\} = \infty$ whence X_n may be replaced by $-X_n$. Setting $A_i = \{0 \leq S_i \leq \varepsilon\}$, $B_j = \{|S_j| \leq \varepsilon\}$, $C_{ij} = \{|S_j - S_i| \leq \varepsilon\}$, clearly $A_i \cup_{j=i+k}^{\infty} A_j \subset A_i \cup_{j=i+k}^{\infty} C_{ij}$, implying via independence and Theorem 1.6.1 that for all $k \geq 1$

$$P\left\{A_i \cup_{j=i+k}^{\infty} A_j\right\} \leq P\{A_i\} P\left\{\bigcup_{j=i+k}^{\infty} C_{ij}\right\} = P\{A_i\} P\left\{\bigcup_{j=i+k}^{\infty} B_{j-i}\right\}.$$

Thus, $P\{B_n, \text{i.o.}\} = 1$ by Lemma 4. \square

Lemma 5. If $\{S_n = \sum_{i=1}^n X_i, n \geq 1, S_0 = 0\}$, where $\{X_n, n \geq 1\}$ are i.i.d. random variables, then for any $\delta > 0$ and positive integers k, N

$$\sum_{i=0}^N P\{|S_i| < k\delta\} \leq 2k \sum_{i=0}^N P\{|S_i| < \delta\}. \quad (16)$$

PROOF. If $T_j = \inf\{n \geq 0 : j\delta \leq S_n < (j+1)\delta\}$, where $\inf\{\emptyset\} = \infty$, then

$$\begin{aligned} \sum_{i=0}^N P\{j\delta \leq S_i < (j+1)\delta\} &= \sum_{i=0}^N \sum_{n=0}^i P\{T_j = n, j\delta \leq S_i < (j+1)\delta\} \\ &\leq \sum_{n=0}^N \sum_{i=n}^N P\{T_j = n\} P\{|S_i - S_n| < \delta\} \leq \sum_{n=0}^N P\{|S_n| < \delta\}, \end{aligned}$$

and so (16) follows by summing on j from $-k$ to $k-1$. \square

A sequence of partial sums $\{S_n, n \geq 1\}$ of i.i.d. random variables $\{X_n, n \geq 1\}$ is called a **random walk** and it is customary to set $S_0 = 0$. The origin and the random walk itself are said to be **recurrent** if $P\{|S_n| < \varepsilon, \text{i.o.}\} = 1$, for all $\varepsilon > 0$. Otherwise, the random walk is **nonrecurrent** or transient. Thus, Example 1 furnishes a criterion for a random walk to be recurrent. However,

here as in virtually all questions concerning random walks, criteria involving the underlying distribution F of X are far preferable to those involving S_n for the simple reason that the former are much more readily verifiable (see Example 5.2.1).

The definition of a recurrent random walk applies also to sums S_n of i.i.d. random vectors where $|S_n|$ signifies Euclidean distance from the origin. According to Example 2 and Exercise 4.2.10, there is an abrupt change in the behavior of simple random walks in going from the plane to three dimensions. This is likewise true for general random walks (Chung and Fuchs, 1951).

EXAMPLE 2 (Polya). Let $\{S_n, n \geq 0\}$ be a **simple random walk** on $(-\infty, \infty)^m$ whose initial position is at the origin, that is, $S_n = \sum_{i=1}^m X_i$, where $\{X_n, n \geq 1\}$ are i.i.d. random vectors with

$$P\{X_n = (e_1, \dots, e_m)\} = \frac{1}{2^m} \quad \text{where } e_i = 0, 1, -1 \quad \text{and} \quad \sum_{i=1}^m e_i^2 = 1.$$

When $m = 1$, $\{X_n, n \geq 1\}$ constitute Bernoulli trials with parameter $p = \frac{1}{2}$ and it follows immediately from Theorem 3.4.3 that $\{S_n, n \geq 0\}$ is recurrent in this case. When $m = 2$, setting $A_n = \{S_{2n} = (0, 0)\}$ and recalling Exercises 2.1.3, 2.3.3,

$$P\{A_n\} = 4^{-2n} \sum_{j=0}^n \frac{(2n)!}{[j!(n-j)!]^2} = \binom{2n}{n} 4^{-2n} \sum_{j=0}^n \binom{n}{j}^2 = \left[\binom{2n}{n} 2^{-2n} \right]^2 \sim \frac{1}{\pi n},$$

and so $\sum_1^\infty P\{A_n\} = \infty$. Moreover, $A_i A_j = A_i C_{ij}$ for $i < j$ where $C_{ij} = \{S_{2j} - S_{2i} = (0, 0)\}$, implying

$$P\left\{A_i \bigcup_{j=i+k}^\infty A_j\right\} = P\{A_i\} P\left\{\bigcup_{j=i+k}^\infty C_{ij}\right\} = P\{A_i\} P\left\{\bigcup_{j=i+k}^\infty A_{j-i}\right\}.$$

Thus, Lemma 4 applies with $B_i = A_i$, revealing that a simple random walk in the plane is recurrent. On the other hand, for $m \geq 3$ the origin and random walk are nonrecurrent (Exercise 4.2.10).

Lemma 6. Let $\{X_n, n \geq 1\}$ be independent, symmetric r.v.s and $\{a_n, n \geq 1\}$, $\{c_n, n \geq 1\}$ sequences of positive numbers with $a_n \rightarrow \infty$. If $X'_n = X_n I_{\{|X_n| \leq c_n\}}$, $S'_n = \sum_{j=1}^n X'_j$, $S_n = \sum_{j=1}^n X_j$, then

$$P\left\{\overline{\lim}_{n \rightarrow \infty} \frac{S'_n}{a_n} > 1\right\} = 1 \quad \text{implies} \quad P\left\{\overline{\lim}_{n \rightarrow \infty} \frac{S_n}{a_n} \geq 1\right\} = 1. \quad (17)$$

PROOF. The hypothesis ensures that $N_m = \inf\{j \geq m: S'_j > a_j\}$ ($= \infty$ otherwise) is a bona fide r.v. for all $m \geq 1$. If whenever $n \geq m$

$$P\{S_n \geq S'_n, N_m = n\} = P\{S_n \leq S'_n, N_m = n\}, \quad (18)$$

then

$$\begin{aligned} \mathbb{P}\left\{\bigcup_{j=m}^{\infty} [S_j > a_j]\right\} &\geq \mathbb{P}\left\{\bigcup_{j=m}^{\infty} [S'_j > a_j, S_j \geq S'_j]\right\} \geq \mathbb{P}\{S_{N_m} \geq S'_{N_m}, N_m < \infty\} \\ &= \sum_{n=m}^{\infty} \mathbb{P}\{S_n \geq S'_n, N_m = n\} \geq \frac{1}{2}, \end{aligned}$$

implying

$$\mathbb{P}\left\{\overline{\lim}_{n \rightarrow \infty} \frac{S_n}{a_n} \geq 1\right\} \geq \mathbb{P}\{S_n > a_n, \text{i.o.}\} \geq \frac{1}{2}$$

and hence also the conclusion of (17) by the Kolmogorov zero-one law.

To verify (18), set $X_j^* = X_j I_{\{|X_j| \leq c_j\}} - X_j I_{\{|X_j| > c_j\}}$ and note that by symmetry and independence the joint distributions of (X_1, \dots, X_n) and (X_1^*, \dots, X_n^*) are identical for all n . Hence, if $n > m$,

$$\begin{aligned} &\mathbb{P}\{S_n \geq S'_n, N_m = n\} \\ &= \mathbb{P}\left\{\sum_1^n X_j I_{\{|X_j| > c_j\}} \geq 0, \sum_1^n X_j I_{\{|X_j| \leq c_j\}} > a_n, S'_i \leq a_i, m \leq i < n\right\} \\ &= \mathbb{P}\left\{\sum_1^n X_j^* I_{\{|X_j^*| > c_j\}} \geq 0, \sum_1^n X_j^* I_{\{|X_j^*| \leq c_j\}} > a_n, \sum_1^i X_j^* I_{\{|X_j^*| \leq c_j\}} \leq a_i, m \leq i < n\right\} \\ &= \mathbb{P}\{S_n \leq S'_n, N_m = n\}, \end{aligned}$$

and equality also holds for $n = m$, *mutatis mutandis*. \square

The next result reveals the extent to which mean convergence is stronger than convergence in probability and provides a Cauchy criterion for the former.

Theorem 3 (i) (Mean Convergence Criterion). If the r.v.s $\{|X_n|^p, n \geq 1\}$ are u.i. for some $p > 0$ and $X_n \xrightarrow{p} X$, then $X \in \mathcal{L}_p$ and $X_n \xrightarrow{\mathcal{L}_p} X$.

Conversely, if $X_n, n \geq 1$, are \mathcal{L}_p r.v.s with $X_n \xrightarrow{\mathcal{L}_p} X$, then $X \in \mathcal{L}_p$, $X_n \xrightarrow{p} X$, and $\{|X_n|^p, n \geq 1\}$ are u.i.

(ii) (Cauchy Convergence Criterion). If $\{X_n, n \geq 1\}$ are \mathcal{L}_p r.v.s with $\sup_{n > m} \mathbb{E}|X_m - X_n|^p = o(1)$ as $n \rightarrow \infty$, there exists a r.v. $X \in \mathcal{L}_p$ such that $X_n \xrightarrow{\mathcal{L}_p} X$ and conversely.

PROOF. If $X_n \xrightarrow{p} X$, by Lemma 3.3.2 there exists a subsequence X_{n_k} with $X_{n_k} \xrightarrow{\text{a.e.}} X$. By Fatou's lemma (Corollary 2)

$$\mathbb{E}|X|^p = \mathbb{E}\left[\varliminf_{k \rightarrow \infty} |X_{n_k}|^p\right] \leq \varliminf_{k \rightarrow \infty} \mathbb{E}|X_{n_k}|^p \leq \sup_{m \geq 1} \mathbb{E}|X_m|^p < \infty$$

since $\{|X_n|^p\}$ is u.i. Again employing the latter, for any $\varepsilon > 0$, $\delta > 0$ may be chosen such that $\mathbb{P}\{A\} < \delta$ implies

$$\sup_{n \geq 1} \mathbb{E}\{|X_n|^p I_A\} < \varepsilon, \quad \mathbb{E}\{|X|^p I_A\} < \varepsilon, \quad (19)$$

the second part holding by Lemma 2. Moreover, $X_n \xrightarrow{p} X$ ensures that $n \geq N$ entails

$$P\{|X_n - X| > \varepsilon\} < \delta. \quad (20)$$

Consequently, for $n \geq N$, by (19), (20) and Lemma 3

$$\begin{aligned} E|X_n - X|^p &= E[|X_n - X|^p(I_{\{|X_n - X| \leq \varepsilon\}} + I_{\{|X_n - X| > \varepsilon\}})] \\ &\leq \varepsilon^p + 2^p E[I_{\{|X_n - X| > \varepsilon\}}(|X_n|^p + |X|^p)] < 2^{p+1}(\varepsilon + \varepsilon^p), \end{aligned}$$

so that $X_n \xrightarrow{p} X$.

Conversely, if $X_n \xrightarrow{p} X$, then, as noted earlier, $X_n \xrightarrow{p} X \in \mathcal{L}_p$ and by Lemma 3

$$\sup_{n \geq 1} E|X_n|^p I_A \leq 2^p \sup_{n \geq 1} E[|X_n - X|^p I_A + |X|^p I_A] < \infty \quad (21)$$

for all $A \in \mathcal{F}$ and, in particular, for $A = \Omega$. Since

$$\begin{aligned} \sup E|X_n - X|^p I_{\{|X_n - X| > K\}} &\leq \max_{n \leq m} E|X_n - X|^p I_{\{|X_n - X| > K\}} + \sup_{n > m} E|X_n - X|^p \\ &\xrightarrow{K \rightarrow \infty} \sup_{n > m} E|X_n - X|^p \xrightarrow{m \rightarrow \infty} 0, \end{aligned}$$

$\{|X_n - X|^p, n \geq 1\}$ is u.i. whence (21) guarantees that $\{|X_n|^p, n \geq 1\}$ is u.i. Apropos of (ii), if the Cauchy criterion holds, the Markov inequality ensures that for any $\varepsilon > 0$

$$\sup_{m > n} P\{|X_m - X_n| > \varepsilon\} \leq \varepsilon^{-p} \sup_{m > n} E|X_m - X_n|^p = o(1)$$

and Lemma 3.3.2 guarantees a subsequence n_k with $X_{n_k} \xrightarrow{a.s.}$ some r.v. X . By Corollary 2 (Fatou's lemma), $E|X_m - X|^p \leq \liminf_{k \rightarrow \infty} E|X_m - X_{n_k}|^p$, implying

$$0 \leq \lim_{m \rightarrow \infty} E|X_m - X|^p \leq \lim_{m \rightarrow \infty} \liminf_{k \rightarrow \infty} E|X_m - X_{n_k}|^p = 0,$$

so that $X_n \xrightarrow{p} X$. Conversely, if $X_n \xrightarrow{p} X$, by Lemma 3

$$\sup_{m > n} E|X_m - X_n|^p \leq 2^p \left[\sup_{m > n} E|X_m - X|^p + E|X - X_n|^p \right] = o(1)$$

and the Cauchy criterion obtains. \square

Corollary 3 (Lebesgue Dominated Convergence Theorem). *Let $\{X, X_n, n \geq 1\}$ be a sequence of r.v.s with $X_n \xrightarrow{p} X$. If $E[\sup_{n \geq 1} |X_n|] < \infty$, then $E|X_n - X| \rightarrow 0$ and a fortiori $E X_n \rightarrow E X$.*

PROOF. The hypothesis ensures that $\{X_n, n \geq 1\}$ is u.i., whence the conclusion follows from Theorem 3(i) with $p = 1$. \square

Corollary 4. *If $\{X, X_n, n \geq 1\}$ is a sequence of nonnegative \mathcal{L}_1 r.v.s with $X_n \xrightarrow{p} X$, then $E X_n \rightarrow E X$ iff $E|X_n - X| \rightarrow 0$ iff $\{X_n, n \geq 1\}$ is u.i.*

PROOF. Sufficiency is immediate from $|E(X_n - E(X))| = |E(X_n - X)| \leq E|X_n - X|$. Apropos of necessity, since $0 \leq (X - X_n)^+ \leq X$ and $(X - X_n)^+ \xrightarrow{P} 0$ by Corollary 3.3.3, dominated convergence (Corollary 3) guarantees $E(X - X_n)^+ \rightarrow 0$. By hypothesis $E(X - X_n) \rightarrow 0$, whence $E(X - X_n)^- \rightarrow 0$, and so $E|X - X_n| \rightarrow 0$. \square

Among other things, Corollary 4 underlines the importance of the concept of u.i.

Corollary 5. If $X_n, n \geq 1$, are \mathcal{L}_p r.v.s with $X_n \xrightarrow{\mathcal{L}_p} X$ for some $p > 0$, then $X \in \mathcal{L}_p$ and $E|X_n|^p \rightarrow E|X|^p$.

PROOF. The hypothesis implies that $X_n \xrightarrow{P} X$ and Corollary 3.3.3 ensures that $|X_n|^p \xrightarrow{P} |X|^p$, whence $E|X_n|^p \rightarrow E|X|^p$ by Corollary 4. \square

In the proof of Theorem 2.2.2 it was shown for binomial r.v.s S_n that $E|S_n - np|^4 = O(n^2)$ and so $E|(S_n/n) - p|^4 = o(1)$. Thus, it follows directly from Corollary 5 that $E(S_n/n)^4 \rightarrow p^4$. More generally, since $S_n/n \xrightarrow{P} p$ and $|S_n/n| \leq 1$ Corollaries 3 and 5 ensure that $E|S_n/n|^\beta \rightarrow p^\beta$ for every $\beta > 0$.

EXERCISES 4.2

1. Improve the inequality of Lemma 3 by showing that

$$(a + b)^p \leq (a^p + b^p) \cdot \max(1, 2^{p-1}) \quad \text{for } a > 0, b > 0, p > 0;$$

also, if $a_i \geq 0$, then $(\sum_{i=1}^n a_i)^p \geq \sum_{i=1}^n a_i^p$ (resp. \leq) for $p \geq 1$ (resp. \leq).

2. Prove that if r.v.s $X_n \xrightarrow{\mathcal{L}_p} X$ and $X_n \xrightarrow{\mathcal{L}_p} Y$, then $X = Y$, a.c. Construct r.v.s $X, X_n, n \geq 1$, such that (i) $X_n \xrightarrow{\mathcal{L}_p} X$ but $X_n \not\xrightarrow{\mathcal{L}_p} X$ for any $p > 0$, (ii) $X_n \xrightarrow{\mathcal{L}_p} X$ for all $p > 0$ but $X_n \not\xrightarrow{\mathcal{L}_p} X$.
3. Let $\{A_n, n \geq 1\}$ be events with $\sum_{n=1}^{\infty} P(A_n) = \infty$. If there exist events $\{B_i, i \geq 1\}$ and $\{D_{ij}, j > i, i \geq 1\}$ such that for all large i and some positive integer k (resp. infinitely many $k > 0$)
 - i. $A_i A_j \subset D_{ij}, i < j$,
 - ii. $P(\bigcup_{j=k+i}^{\infty} D_{ij}) = P(\bigcup_{j=k+i}^{\infty} B_{j-i})$,
 - iii. the classes $\{A_i\}$ and $\{D_{i,i+k}, D_{i,i+k+1} D_{i,i+k}^c, D_{i,i+k+2} D_{i,i+k+1}^c D_{i,i+k}^c, \dots\}$ are independent,
 then $P(\bigcup_{j=-k}^{\infty} B_j) = 1$ (resp. $P(B_n, i.o.) = 1$).
4. If X_1 and X_2 are independent r.v.s and $X_1 + X_2 \in \mathcal{L}_p$ for some $p \in (0, \infty)$, then $X_i \in \mathcal{L}_p, i = 1, 2$. Hint: For all large $\lambda > 0$,

$$P\{|X_1| > \lambda\} \leq 2 P\left\{|X_1| > \lambda, |X_2| < \frac{\lambda}{2}\right\} \leq 2 P\left\{|X_1 + X_2| > \frac{\lambda}{2}\right\}.$$

5. Let $P\{X_n = a_n > 0\} = 1/n = 1 - P\{X_n = 0\}, n \geq 1$. Is $\{X_n, n \geq 1\}$ u.i. if
 - i. $a_n = o(n)$,
 - ii. $a_n = cn > 0$?

6. If $\{X_n, n \geq 1\}$ are r.v.s with $\sup_{n \geq 1} E|X_n|^\beta < \infty$ for some $\beta > 0$, then $\{|X_n|^\alpha, n \geq 1\}$ is u.i. for $0 < \alpha < \beta$.
7. If the r.v.s $X_n, n \geq 1$, are u.i., so are $S_n/n, n \geq 1$, where $S_n = \sum_{i=1}^n X_i$; in particular, if $X_n, n \geq 1$, are identically distributed \mathcal{L}_1 random variables, then $\{S_n/n, n \geq 1\}$ is u.i.
8. Show that (i) if the Poisson r.v.s S_n have p.d.f. $p(k; n\lambda), n \geq 1$, then $E|(S_n/n) - \lambda| \rightarrow 0$.
Hint: Recall Exercise 2.2.2. (ii) Any sequence of r.v.s $Y_n \xrightarrow{P} 0$ iff $E(|Y_n|/(1 + |Y_n|)) = o(1)$.
9. (i) Construct a sequence of r.v.s that is u.i. from below and for which the expectation of $\liminf X_n$ does not exist. (ii) Show that the hypothesis $\sup_{n \geq 1} |X_n| \in \mathcal{L}_1$ of Corollary 3 is equivalent to $|X_n| \leq Y \in \mathcal{L}_1, n \geq 1$.
10. Let $\{X_n, n \geq 1\}$ be a simple symmetric random walk in R^k , that is, $\{X_n\}$ are i.i.d. random vectors such that $P\{X_n = (e_1, \dots, e_k)\} = 1/2k$, where $e_j = 0, 1$, or -1 and $\sum_j e_j^2 = 1$. Prove that
- $$P\left\{\sum_1^n X_i \text{ returns to its origin, i.o.}\right\} = 0 \quad \text{for } k = 3.$$

11. If $\{X, X_n, n \geq 1\}$ are r.v.s on (Ω, \mathcal{F}, P) , show that the indefinite integrals $\int_A X_n dP \rightarrow \int_A X dP$, finite, uniformly for all $A \in \mathcal{F}$ iff X_n converges to X in mean.
12. If the two sequences of integrable r.v.s $\{X_n\}, \{Y_n\}$ satisfy $P\{X_n \geq Y_n \geq 0\} = 1$, $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$, and $E|X_n - Y_n| \rightarrow 0$, then $E|Y_n - Y| \rightarrow 0$.
13. Let $\{X_n, n \geq 1\}$ be a sequence of \mathcal{L}_p r.v.s, $p > 0$ with $\sup[\int_A |X_n|^p dP; n \geq 1]$ and $P\{A\} < \delta\} = o(1)$ as $\delta \rightarrow 0$. Then $X_n \xrightarrow{P} X$ iff $X_n \xrightarrow{Z_p} X$.
14. If $X \in \mathcal{L}_1(\Omega, \mathcal{F}, P)$ and $\sigma(X), \mathcal{G}$ are independent classes of events, prove that $E XI_A = E X \cdot P\{A\}$, all $A \in \mathcal{G}$.

15. (Kochen–Stone) If $\{Z_n, n \geq 1\}$ is a sequence of r.v.s with $0 < E Z_n^2 < \infty$, $E Z_n \neq 0$, and $\overline{\lim}_{n \rightarrow \infty} (E Z_n)^2/E Z_n^2 > 0$, then $P\{\overline{\lim}_{n \rightarrow \infty} Z_n/E Z_n \geq 1\} > 0$. *Hint:* If $Y_n = Z_n/E Z_n$, there is a subsequence $\{n'\}$ with $E Y_{n'}^2 < K < \infty$. Replacing $\{n'\}$ by $\{n\}$ for notational simplicity, $E \overline{\lim} Y_n^2 \leq K$ by Corollary 2. Since $-Y_n \leq 1 + Y_n^2$, necessarily $E \overline{\lim} (-Y_n)$ exists. Then Theorem 2(ii) ensures (since $\{Y_n\}$ is u.i.) that $E \overline{\lim} Y_n \geq \overline{\lim} E Y_n = 1$.

16. (Kochen–Stone) If $\{A_n, n \geq 1\}$ is a sequence of events such that for some $c > 0$

- i. $P\{A_i A_j\} \leq c P\{A_i\} [P\{A_{j-i}\} + P\{A_j\}], i < j$
- ii. $\sum_{n=1}^{\infty} P\{A_n\} = \infty$,

then $P\{A_n, \text{i.o.}\} > 0$.

Hint: If $Z_n = \sum_{i=1}^n I_{A_i}$, note that $E Z_n^2 \leq E Z_n + 4c(\sum_i P\{A_i\})^2 \leq (1 + 4c)(E Z_n)^2$ for all large n since $E Z_n \rightarrow \infty$ and, via Exercise 15,

$$P\{A_n, \text{i.o.}\} \geq P\{\overline{\lim} Z_n/E Z_n \geq 1\} > 0.$$

4.3 Jensen, Hölder, Schwarz Inequalities

A finite real function g on an interval $J \subset (-\infty, \infty)$ is called **convex** on J if whenever $x_1, x_2 \in J$ and $\lambda \in [0, 1]$

$$g(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda g(x_1) + (1 - \lambda)g(x_2). \quad (1)$$

Geometrically speaking, the value of a convex function at any point on the line segment joining x_1 to x_2 lies on or below the line segment joining $g(x_1)$ and $g(x_2)$. Since $t = u((t - s)/(u - s)) + s((u - t)/(u - s))$,

$$g(t) \leq \frac{t - s}{u - s} g(u) + \frac{u - t}{u - s} g(s), \quad s < t < u, \quad (2)$$

or equivalently

$$\frac{g(t) - g(s)}{t - s} \leq \frac{g(u) - g(t)}{u - t}, \quad s < t < u. \quad (3)$$

If g is convex on an open interval J_0 , it follows from (2) that $\overline{\lim}_{s < t \rightarrow s} g(t) \leq g(s)$, $\underline{\lim}_{t > s \rightarrow t} g(s) \geq g(t)$, $\underline{\lim}_{t < u \rightarrow t} g(u) \geq g(t)$, and $\overline{\lim}_{u > t \rightarrow u} g(t) \leq g(u)$, whence g is continuous on J_0 . Furthermore, as a consequence of (3), a differentiable function g is convex on J_0 iff g' is nondecreasing on J_0 . Thus, if g is convex on J_0 and twice differentiable, $g'' \geq 0$. Conversely, if $g'' \geq 0$ on J_0 , a two-term Taylor expansion yields, setting $q_1 = \lambda$, $q_2 = 1 - \lambda$,

$$q_i g(x_i) \geq q_i [g(q_1 x_1 + q_2 x_2) + q_{3-i}(x_i - x_{3-i})g'(q_1 x_1 + q_2 x_2)], \quad i = 1, 2,$$

and summing, (1) holds, that is, g is convex on J_0 . Moreover, it is shown in Hardy *et al.* (1934, pp. 91–95) that

- i. If g is convex on an open interval J_0 , it has left and right derivatives g'_l and g'_r at every point of J_0 , with $g'_l \leq g'_r$, each derivative being nondecreasing,
- ii. If g is convex on an interval J , at each interior point $\xi \in J$,

$$g(t) \geq g(\xi) + (t - \xi)g'_r(\xi), \quad t \in J. \quad (4)$$

Theorem 1. *If X is an integrable r.v., c is a finite constant, and g is a convex function on $(-\infty, \infty)$, then*

$$\mathbb{E} g^-(X - \mathbb{E} X + c) < \infty. \quad (5)$$

Moreover, if $\alpha(t)$ and $t - \alpha(t)$ are nondecreasing on $(-\infty, \infty)$, then $\mathbb{E}|\alpha(X)| < \infty$, $\mathbb{E} g^-(\alpha(X) - \mathbb{E} \alpha(X) + c) < \infty$, and

$$\mathbb{E} g(X - \mathbb{E} X + c) \geq \mathbb{E} g(\alpha(X) - \mathbb{E} \alpha(X) + c). \quad (6)$$

PROOF. By (4), $g(t) \geq g(0) + tg'_r(0)$ for $t \in (-\infty, \infty)$, whence (5) holds. Since monotonicity ensures $t^+ + \alpha(0) \geq \alpha(t^+) \geq \alpha(t) \geq \alpha(-t^-) \geq -t^- + \alpha(0)$, the hypothesis implies $|\alpha(X)| \leq |X| + |\alpha(0)|$ and so $\alpha(X)$ is integrable. Consequently, (4) yields $\mathbb{E} g^-(\alpha(X) - \mathbb{E} \alpha(X) + c) < \infty$. Set

$$\beta(t) = t - \alpha(t) - \mathbb{E} X + \mathbb{E} \alpha(X), \quad t \in (-\infty, \infty).$$

Then $\mathbb{E}|\beta(X)| < \infty$ and $\mathbb{E} \beta(X) = 0$. If $P\{\beta(X) = 0\} = 1$, (6) holds trivially. Otherwise, $\beta(t_1) < 0$, $\beta(t_2) > 0$ for some $t_1, t_2 \in (-\infty, \infty)$. If $t_0 = \inf\{t : \beta(t) > 0\}$, then $t_1 \leq t_0 \leq t_2$ by monotonicity of $t - \alpha(t)$, and

$$t \geq t_0 \quad \text{if } \beta(t) > 0, \quad t \leq t_0 \quad \text{if } \beta(t) < 0. \quad (7)$$

Again employing (4),

$$g(X - \mathbb{E} X + c) \geq g(\alpha(X) - \mathbb{E} \alpha(X) + c) + \beta(X)g'(\alpha(X) - \mathbb{E} \alpha(X) + c). \quad (8)$$

By (7), $X \geq t_0$ when $\beta(X) > 0$ and $X \leq t_0$ for $\beta(X) < 0$. Since both g' and α are nondecreasing, necessarily

$$\beta(X)g'(\alpha(X) - \mathbb{E} \alpha(X) + c) \geq \beta(X)g'(\alpha(t_0) - \mathbb{E} \alpha(X) + c). \quad (9)$$

Taking expectations in (8) and (9), the conclusion (6) follows by recalling that $\mathbb{E} \beta(X) = 0$. \square

Corollary 1. If g is a convex function on $(-\infty, \infty)$, for any \mathcal{L}_1 r.v. X and any finite constant c

$$\mathbb{E} g(X - \mathbb{E} X + c) \geq g(c) \quad (10)$$

and, in particular,

$$\mathbb{E} g(X) \geq g(\mathbb{E} X) \quad (\text{Jensen's inequality}). \quad (11)$$

Corollary 2. If X is an \mathcal{L}_1 r.v., then for $1 \leq p < \infty$

$$\mathbb{E}|X - \mathbb{E} X|^p \geq \mathbb{E}|Y - \mathbb{E} Y|^p, \quad (12)$$

where for some choice of $-\infty \leq a < b \leq \infty$

$$Y = XI_{[a \leq X \leq b]} + aI_{[X < a]} + bI_{[X > b]}.$$

PROOF. Take $c = 0$, $\alpha(t) = \max[a, \min(t, b)]$, $g(t) = |t|^p$, $p \geq 1$ in (6). \square

In particular, abbreviating $(\mathbb{E}|X|)^p$ by $\mathbb{E}^p|X|$, Jensen's inequality (11) yields for $1 \leq p < \infty$

$$\mathbb{E}|X|^p \geq \mathbb{E}^p|X| \quad \text{or} \quad \mathbb{E}^{1/p}|X|^p \geq \mathbb{E}|X|. \quad (13)$$

Replacing p and $|X|$ respectively by p'/p and $|X|^p$ in (13),

$$\mathbb{E}^{1/p}|X|^p \leq \mathbb{E}^{1/p'}|X|^{p'}, \quad 0 < p < p' < \infty \quad (14)$$

and so convergence in mean of order p implies convergence in mean of any order less than p . A convenient, widespread notation is to set

$$\|X\|_p = \mathbb{E}^{1/p}|X|^p, \quad p > 0, \quad (15)$$

and it is customary to refer to $\|X\|_p$ as the **p -norm** of X . According to (14), $\|X\|_p \leq \|X\|_{p'}$ for $0 < p < p'$. Moreover, $\|X\|_p$ satisfies the triangle inequality for $p \geq 1$ as noted in Exercise 10 and $\|cX\|_p = |c| \cdot \|X\|_p$.

Theorem 2 (Hölder Inequality). If X, Y are measurable functions on a probability space (Ω, \mathcal{F}, P) , then for $p > 1$, $p' > 1$ with $(1/p) + (1/p') = 1$

$$\mathbb{E}|XY| \leq \|X\|_p \cdot \|Y\|_{p'}. \quad (16)$$

PROOF. In proving (16), it may be supposed that $0 < \|X\|_p \|Y\|_{p'} < \infty$ since (16) is trivial otherwise. Set

$$U = \frac{|X|}{\|X\|_p}, \quad V = \frac{|Y|}{\|Y\|_{p'}},$$

entailing $\|U\|_p = 1 = \|V\|_{p'}$. Now, $-\log t$ is a convex function on $(0, \infty)$, whence, via (1), for $a, b > 0$

$$-\log\left(\frac{a^p}{p} + \frac{b^{p'}}{p'}\right) \leq -\frac{1}{p}\log a^p - \frac{1}{p'}\log b^{p'} = -\log ab,$$

or equivalently

$$ab \leq \frac{a^p}{p} + \frac{b^{p'}}{p'}, \quad 0 \leq a, b \leq \infty.$$

Thus,

$$\mathbb{E} UV \leq \frac{1}{p} \mathbb{E} U^p + \frac{1}{p'} \mathbb{E} V^{p'} = \frac{1}{p} + \frac{1}{p'} = 1,$$

which is tantamount to (16). \square

Corollary 3 (Schwarz Inequality). *For any \mathcal{L}_2 random variables X and Y ,*

$$|\mathbb{E} XY| \leq \mathbb{E}|XY| \leq (\mathbb{E} X^2)^{1/2}(\mathbb{E} Y^2)^{1/2}. \quad (17)$$

Corollary 4 (Liapounov). *If X is a non-negative \mathcal{L}_p r.v., all $p > 0$, and*

$$g(p) = \log \mathbb{E} X^p, \quad 0 \leq p < \infty, \quad (18)$$

then g is convex on $[0, \infty)$.

PROOF. For $0 \leq p_1, p_2 < \infty$ and $q_1, q_2 > 0$ with $q_1 + q_2 = 1$, noting that $1/q_i > 1$, $i = 1, 2$, Hölder's inequality yields

$$\begin{aligned} g(q_1 p_1 + q_2 p_2) &= \log \mathbb{E}(X^{q_1 p_1} \cdot X^{q_2 p_2}) \leq \log(\|X^{q_1 p_1}\|_{1/q_1} \cdot \|X^{q_2 p_2}\|_{1/q_2}) \\ &= \log(\mathbb{E}^{q_1} X^{p_1} \cdot \mathbb{E}^{q_2} X^{p_2}) = q_1 g(p_1) + q_2 g(p_2). \end{aligned} \quad \square$$

If X is a r.v. on a probability space, $\mathbb{E}|X|^p$, $p > 0$ is called the **p th absolute moment** of X (or its distribution), while for any positive integer k , $\mathbb{E} X^k$ (if it exists) is termed the **k th moment** of X .

For any \mathcal{L}_1 r.v. X , the **variance** of X is defined by

$$\sigma_X^2 = \sigma^2(X) = \mathbb{E}(X - \mathbb{E} X)^2, \quad (19)$$

while its positive square root is the **standard deviation** of X . Clearly, for every finite constant c , $\sigma^2(X + c) = \sigma^2(X)$ and $\sigma^2(cX) = c^2\sigma^2(X)$. The variance or standard deviation of X provides information about the extent to which the distribution of X clusters about its mean and this is reflected in the simple but extremely useful **Tchebychev inequality**

$$\mathbb{P}\{|X - \mathbb{E} X| \geq a\} \leq \frac{\sigma_X^2}{a^2}, \quad a > 0, \quad (20)$$

which follows from a direct application of the Markov inequality to the r.v. $(X - \mathbb{E} X)^2$.

A basic tool of probability theory is **truncation**. Two alternative methods of truncating a r.v. X are

$$Y' = XI_{\{a \leq X \leq c\}}, \quad (i)$$

$$Y = XI_{\{a \leq X \leq c\}} + aI_{\{X < a\}} + cI_{\{X > c\}}, \quad (ii)$$

where a, c are constants such that $-\infty \leq a < c \leq \infty$. One or both equality signs in the set of the indicator function of (i) may be deleted. Whenever both a and c are finite, Y and Y' are bounded r.v.s and hence have moments of all orders. For $X \in \mathcal{L}_1$, Corollary 2 reveals that $\sigma_Y^2 \leq \sigma_X^2$, whereas no comparable inequality between σ_Y^2 and σ_X^2 exists (Exercise 1).

If X, Y are r.v.s with $0 < \sigma(X), \sigma(Y) < \infty$, the **correlation coefficient** between X and Y or simply the correlation of X and Y is given by

$$\rho_{X,Y} = \rho(X, Y) = \frac{\mathbb{E}(X - \mathbb{E} X)(Y - \mathbb{E} Y)}{\sigma(X) \cdot \sigma(Y)}. \quad (21)$$

If $\rho(X, Y) = 0$, the r.v.s X and Y are said to be **uncorrelated**.

It follows directly from the Schwarz inequality (17) that $|\rho(X, Y)| \leq 1$. The correlation coefficient $\rho_{X,Y}$ indicates the extent to which there is a linear relationship between the r.v.s X and Y (Exercise 7).

\mathcal{L}_2 r.v.s $X_n, n \geq 1$, are called **uncorrelated** if X_n and X_m are uncorrelated for each pair of distinct indices n, m . Independent \mathcal{L}_2 r.v.s are necessarily uncorrelated as follows from

Theorem 3. If X and Y are independent \mathcal{L}_1 r.v.s, then $X \cdot Y \in \mathcal{L}_1$ and

$$\mathbb{E} XY = \mathbb{E} X \cdot \mathbb{E} Y. \quad (22)$$

PROOF. To prove (22) suppose first that $X \geq 0, Y \geq 0$. For $m \geq 1$ and $j = 1, 2, \dots$, set $m_j = j/2^m$, $Y_{m,j} = I_{\{m_j < Y \leq m_{j+1}\}}$, and $Y_m = \sum_{j=1}^{\infty} m_j Y_{m,j}$. Then $0 \leq Y_m \uparrow Y$, a.c., and $0 \leq XY_m \uparrow XY$, a.c. By the definition of expectation and independence, taking n_i analogous to m_j

$$\begin{aligned}
E XY_{m,j} &= \lim_n \sum_{i=1}^{\infty} n_i P\{n_i < XY_{m,j} \leq n_{i+1}\} \\
&= \lim_n \sum_{i=1}^{\infty} n_i P\{n_i < X \leq n_{i+1}, m_j < Y \leq m_{j+1}\} \\
&= \lim_{n \rightarrow \infty} \sum_{i=1}^{\infty} n_i P\{n_i < X \leq n_{i+1}\} \cdot P\{m_j < Y \leq m_{j+1}\} \\
&= P\{m_j < Y \leq m_{j+1}\} \cdot E X.
\end{aligned}$$

By the monotone convergence theorem

$$\begin{aligned}
E XY &= \lim_m E XY_m = \lim_m E \sum_{j=1}^{\infty} m_j XY_{m,j} = \lim_m \sum_{j=1}^{\infty} m_j E XY_{m,j} \\
&= \lim_m \sum_{j=1}^{\infty} m_j P\{m_j < Y \leq m_{j+1}\} E X = E Y \cdot E X.
\end{aligned}$$

Thus, in the general case, recalling Exercise 3.1.5,

$$E|XY| = E|X| \cdot E|Y| \quad \text{and} \quad E X^{\pm} Y^{\pm} = E X^{\pm} \cdot E Y^{\pm}$$

for all four pairings of positive and negative parts. Consequently,

$$\begin{aligned}
E X \cdot E Y &= (E X^+ - E X^-)(E Y^+ - E Y^-) \\
&= E(X^+ - X^-)(Y^+ - Y^-) = E XY. \quad \square
\end{aligned}$$

Corollary 5. If $\{X_j, 1 \leq j \leq n\}$ are uncorrelated \mathcal{L}_2 r.v.s, (a fortiori, independent \mathcal{L}_2 r.v.s), then

$$\sigma^2 \left(\sum_{j=1}^n X_j \right) = \sum_{j=1}^n \sigma^2(X_j). \quad (23)$$

PROOF. Since $E \sum_{j=1}^n X_j = \sum_{j=1}^n E X_j$ and $\sigma^2(X + c) = \sigma^2(X)$ for every finite constant c , it may be supposed that $E X_j = 0, 1 \leq j \leq n$. Then

$$\sigma^2 \left(\sum_{j=1}^n X_j \right) = E \left(\sum_{j=1}^n X_j \right)^2 = \sum_{j=1}^n E X_j^2 + \sum_{i \neq j} E X_i X_j = \sum_{j=1}^n \sigma^2(X_j). \quad \square$$

Even though Tchebychev's inequality is an easy consequence of the order-preserving property of expectations, it is quite useful, especially in verifying convergence in probability (Exercise 4).

Let $\{X_n, n \geq 1\}$ be a sequence of r.v.s with $E|X_n|^{\beta} \leq C < \infty$ for some $\beta \geq 1$ and $S_n = \sum_1^n X_i$. By Hölder's inequality (Exercise 4.3.5) or Jensen's inequality

$$E \left| \frac{S_n}{n} \right|^{\beta} \leq \frac{1}{n} \sum_{j=1}^n E|X_j|^{\beta} \leq C,$$

and so according to Exercise 4.2.6 $\{|S_n/n|^\alpha, n \geq 1\}$ is u.i. for $0 < \alpha < \beta$. If, moreover, $\{|X_n|^\beta, n \geq 1\}$ is u.i., so is $\{|S_n/n|^\beta, n \geq 1\}$, generalizing Exercise 4.2.7.

EXAMPLE 1. If $\{|X_n|^\beta, n \geq 1\}$ is u.i. for some $\beta \geq 1$ and $S_n = \sum_{i=1}^n X_i$, then $|S_n/n|^\beta$ is u.i. In particular, if $\{X_n, n \geq 1\}$ are identically distributed \mathcal{L}_β r.v.s for $\beta \geq 1$, then $\{|S_n/n|^\beta, n \geq 1\}$ is u.i.

PROOF. Since the case $\beta = 1$ is obvious (Exercise 4.2.7), suppose that $\beta > 1$. By Hölder's inequality, $|S_n/n|^\beta \leq \frac{1}{n} \sum_{i=1}^n |X_i|^\beta$, whence uniform integrability follows easily. \square

EXAMPLE 2. Let $S_n = \sum_{j=1}^n X_j, n \geq 1$ where $\{X_n, n \geq 1\}$ are independent r.v.s with $E X_n = 0, E X_n^2 = 1$. If $\{X_n^2, n \geq 1\}$ is u.i., then $\{S_n^2/n, n \geq 1\}$ is u.i.

PROOF. Define $Y_n = X_n I_{\{|X_n| \leq K\}} - E X_n I_{\{|X_n| \leq K\}}$, $K > 0$ and $Z_n = X_n - Y_n, n \geq 1$. Then $\{Y_n, n \geq 1\}$ are independent r.v.s with $E Y_n = 0$ and $E Y_n^2 \leq 1$. The same statement likewise holds for $\{Z_n, n \geq 1\}$. Hence, if $T_n = \sum_{j=1}^n Y_j, W_n = \sum_{j=1}^n Z_j$,

$$\begin{aligned} E(T_n/n^{1/2})^4 &= \frac{1}{n^2} \left(\sum_{j=1}^n E Y_j^4 + 2 \sum_{i < j} E Y_i^2 \cdot E Y_j^2 \right) \\ &\leq \frac{1}{n^2} \left[nK^4 + 2 \binom{n}{2} \right] \leq K^4 + 1 \end{aligned}$$

and so $\{(T_n/n^{1/2})^2, n \geq 1\}$ is u.i. by Exercise 4.2.6. Hence, for any $\varepsilon > 0$, there is a $\delta > 0$ such that $P\{A\} < \delta$ entails

$$\sup_{n \geq 1} E \left(\frac{T_n^2}{n} I_A \right) < \frac{\varepsilon}{4}.$$

On the other hand, via u.i. of $\{X_n^2, n \geq 1\}$, $K = K_\varepsilon > 0$ may be chosen so that $\sup_{n \geq 1} E X_n^2 I_{\{|X_n| > K\}} < \varepsilon/4$ whence

$$E(W_n^2/n) = \frac{1}{n} \sum_{j=1}^n E Z_j^2 \leq \frac{1}{n} \sum_{j=1}^n E X_j^2 I_{\{|X_j| > K\}} < \frac{\varepsilon}{4}.$$

Consequently, for $P\{A\} < \delta$ and all $n \geq 1$,

$$E \left(\frac{S_n^2}{n} I_A \right) \leq 2 E \left(\frac{T_n^2 + W_n^2}{n} I_A \right) < \varepsilon.$$

Clearly, $E(S_n^2/n) = 1, n \geq 1$ and so $\{S_n^2/n, n \geq 1\}$ and a fortiori $\{S_n/n^{1/2}, n \geq 1\}$ is u.i. \square

A generalization of Example 2 appears in Corollary 11.3.2.

EXAMPLE 3. If $\{Y_n, n \geq 1\}$ are non-negative r.v.s satisfying $\lim_{n \rightarrow \infty} E Y_n = 1 = \lim_{n \rightarrow \infty} E Y_n^p$ for some $p > 1$ then $Y_n \xrightarrow{\mathcal{L}_p} 1$.

PROOF. Let $r = \min(1, p/2)$. If $p < 2$, then $r < 1 < p$ and so by Exercise 4.3.8

$$\left(\liminf_{n \rightarrow \infty} \mathbb{E} Y_n^r\right)^{(p-1)/(p-r)} \geq \liminf_{n \rightarrow \infty} \frac{\mathbb{E} Y_n}{(\mathbb{E} Y_n^p)^{(1-r)/(p-r)}} = 1$$

implying

$$0 \leq \overline{\lim}_{n \rightarrow \infty} \sigma_{Y_n}^2 = \overline{\lim}_{n \rightarrow \infty} \mathbb{E} (Y_n^{2r} - \mathbb{E}^2 Y_n^r) \leq 0. \quad (24)$$

Thus, $\sigma_{Y_n}^2 \rightarrow 0$ and $\mathbb{E} Y_n^r \rightarrow 1$ so that $Y_n^r \xrightarrow{P} 1$ and hence $Y_n^p \xrightarrow{P} 1$. Since $\mathbb{E} Y_n^p \rightarrow 1$, $\{Y_n^p, n \geq 1\}$ is u.i. which, in turn, ensures $\mathbb{E} |Y_n - 1|^p \rightarrow 0$.

If rather $p \geq 2$, then $r = 1$ and $\mathbb{E} Y_n^{2r} = \mathbb{E} Y_n^2 \rightarrow 1$ since $(\mathbb{E} Y_n^q)^{1/q}$ is an increasing function of q . Thus, (24) holds with equality throughout and the conclusion follows as before. \square

EXERCISES 4.3

1. (i) Show for any \mathcal{L}_1 r.v. X that $\sigma_X^2 = \mathbb{E} X^2 - (\mathbb{E} X)^2$. (ii) If $P\{X = 1\} = p \in (0, 1)$ and $P\{X = 0\} = P\{X = 2\} = (1 - p)/2$, then, setting $Z = XI_{\{X > 1\}}$, $W = XI_{\{X \leq 1\}}$, necessarily $\sigma_W^2 < \sigma_X^2 < \sigma_Z^2$.
2. Calculate the mean and variance of U_n , the number of empty cells in the random assignment of r balls into n cells.
3. Prove (i) for any r.v.s $\{S_n, n \geq 1\}$ with $\sigma^2(S_n) = o(n^2)$ that $n^{-1}(S_n - \mathbb{E} S_n) \xrightarrow{P} 0$, (ii) for any r.v. X and positive numbers a, t that necessarily $P\{X \geq a\} \leq e^{-at} \mathbb{E} e^{tX}$, (iii) if m is a median of $X \in \mathcal{L}_1$, then $|m - \mathbb{E} X| \leq \mathbb{E}|X - \mathbb{E} X|$ and hence $|m - \mathbb{E} X| \leq \sigma$ for $X \in \mathcal{L}_2$ with variance σ^2 , (iv) for \mathcal{L}_1 r.v.s X and Y with $\mathbb{E} XY$ finite, their covariance, denoted $\text{Cov}[X, Y]$, is defined by $\text{Cov}[X, Y] = \mathbb{E}[X - \mathbb{E} X][Y - \mathbb{E} Y]$. Prove that if f and g are non-decreasing functions and X is an r.v. with $\mathbb{E} f(X), \mathbb{E} g(X), \mathbb{E} f(X) \cdot g(X)$ finite, then $\text{Cov}[f(X), g(X)] \geq 0$, (v) for any \mathcal{L}_2 r.v.s $X_j, 1 \leq j \leq n$, check that

$$\sigma^2\left(\sum_{j=1}^n X_j\right) = \sum_{j=1}^n \sigma^2(X_j) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j).$$

4. For arbitrary real numbers $a_i, b_i, 1 \leq i \leq n$, prove that

$$\sum_1^n |a_i b_i| \leq \left(\sum_1^n |a_i|^p\right)^{1/p} \left(\sum_1^n |b_i|^q\right)^{1/q},$$

provided $(1/p) + (1/q) = 1$, $p > 0$, $q > 0$. Hint: Apply Hölder's inequality to suitable r.v.s X, Y .

5. If $S_n = \sum_{j=1}^n X_j$ where $\{X_j, j \geq 1\}$ are independent r.v.s with $\mathbb{E} X_j = 0$, $\mathbb{E} X_j^2 = \sigma_j^2$, $s_n^2 = \sum_{j=1}^n \sigma_j^2$, then

$$\mathbb{P}\{S_n \geq x\} \geq \left(1 - \frac{2s_n^2}{x^2}\right) \sum_{j=1}^n \mathbb{P}\{X_j \geq 2x\}, \quad x > 0.$$

Hint: $\mathbb{P}\{S_n \geq x, X_j \geq x, \max_{i < j} X_i < x\} \geq \mathbb{P}\{X_j \geq 2x\} \cdot \mathbb{P}\{S_n - X_j \geq -x, \max_{i < j} X_i < x\}$.

6. (i) If X is an \mathcal{L}_1 r.v., $Y = Y(a, b)$ is as in Corollary 2, and $Z = Y(a', b')$, where $a \leq a' < b' \leq b$, then $E|Y - EY|^p \geq E|Z - EZ|^p$ for any $p \geq 1$. (ii) For $g(t), \alpha(t)$ as in Theorem 1, $Eg(X) \geq Eg(\alpha(X)) - E\alpha(X) + EX \geq g(EX)$ and $Eg(X) \geq Eg(cX + (1 - c)EX) \geq g(EX)$, $0 \leq c \leq 1$.
7. (i) If $X' = aX + b$, $Y' = cX + d$, verify that $\rho(X', Y') = \pm\rho(X, Y)$ according as $ac > 0$ or $ac < 0$. (ii) if X, Y are r.v.s with $0 < \sigma(X), \sigma(Y) < \infty$, then $\rho(X, Y) = 1$ iff $(X - EX)/\sigma(X) = (Y - EY)/\sigma(Y)$, a.c. (iii) if $X = \sin Z$, $Y = \cos Z$, where $P\{Z = \pm 1\} = \frac{1}{2} = P\{Z = \pm 2\}$, then $\rho(X, Y) = 0$ despite X, Y being functionally related.
8. Verify for $0 < a < b < d$ and any nonnegative r.v. Y that

$$EY^b \leq (EY^a)^{(d-b)/(d-a)}(EY^d)^{(b-a)/(d-a)}.$$

Utilize this to conclude for any positive r.v.s $\{Y_n, n \geq 1\}$ and positive constants $\{c_n, n \geq 1\}$ that if $\sum_{n=1}^{\infty} c_n^{\alpha} EY_n^{\alpha} < \infty$ for $\alpha = \alpha_1$ and α_2 , where $\alpha_i > 0, i = 1, 2$, then the series converges for all α in $[\alpha_1, \alpha_2]$.

9. The **moment-generating function** (m.g.f.) of a r.v. X is the function $\varphi(h) = Ee^{hX}$. If $\varphi(h)$ is finite for $h = h_0 > 0$ and $h = -h_0$, verify that (i) $Ee^{h|X|} \leq \varphi(h) + \varphi(-h) < \infty$, $0 < h \leq h_0$ (ii) $\log \varphi(h)$ is convex in $[-h_0, h_0]$ and strictly convex if X is non-degenerate (iii) if $EX = 0$, then $\varphi(h) \geq 1$ for $|h| \leq h_0$. (iv) The k th moment of X is finite for every positive integer k and equals $\varphi^{(k)}(0)$, (v) if X_1 and X_2 are independent r.v.s with (finite) m.g.f.s $\varphi_1(h), \varphi_2(h)$ (say) for $|h| < h_0$, then the m.g.f. of $X_1 + X_2$ is $\varphi_1(h) \cdot \varphi_2(h)$ in this interval, (vi) if $\varphi(h)$ is finite in $[0, h_0]$, all moments of X^+ are finite.

10. Prove Minkowski's inequality, that is, if $X_1 \in \mathcal{L}_p, X_2 \in \mathcal{L}_p$, then

$$\|X_1 + X_2\|_p \leq \|X_1\|_p + \|X_2\|_p, \quad p \geq 1.$$

Hint: Apply Hölder's inequality to $E|X_1||X_1 + X_2|^{p-1}$ when $p > 1$.

11. Let $\{X_n, n \geq 1\}$ be r.v.s with (i) $E|X_n|^2 \leq 1$ and (ii) $E((X_1 + \dots + X_n)/n)^2 = O(n^{-\alpha})$ for some $\alpha > 0$. Then $(1/n)\sum_{j=1}^n X_j \xrightarrow{a.s.} 0$. Note that (i) \Rightarrow (ii) if $E X_i X_j = 0, i \neq j$. *Hint:* Choose $n_m = \text{smallest integer } \geq m^{2/\alpha}, m = 1, 2, \dots$. Then $(1/n_m)\sum_{j=1}^{n_m} X_j \xrightarrow{a.s.} 0$ as $m \rightarrow \infty$ and

$$E \left\{ \max_{1 \leq k \leq n_{m+1}-n_m} \left| \frac{1}{n_m+k} \sum_{j=1}^k X_{n_m+j} \right|^2 \right\} \leq E \frac{n_{m+1}-n_m}{n_m^2} \sum_1^{n_{m+1}-n_m} X_{n_m+j}^2 = O(m^{-2}).$$

12. For any sequence $\{A_n, n \geq 1\}$ of events, define

$$Y_n = \frac{1}{n} \sum_{j=1}^n (I_{A_j} - P(A_j)), \quad p_1(n) = \frac{1}{n} \sum_{j=1}^n P(A_j),$$

$$p_2(n) = \frac{2}{n(n-1)} \sum_{1 \leq j < k \leq n} P(A_j A_k), \quad d_n = p_2(n) - p_1^2(n).$$

Prove that (i) $EY_n^2 = d_n + n^{-1}[p_1(n) - p_2(n)]$, (ii) $Y_n \xrightarrow{P} 0$ iff $EY_n^2 \rightarrow 0$ iff $d_n \rightarrow 0$, (iii) $Y_n \xrightarrow{a.s.} 0$ if $d_n = O(n^{-\alpha})$ for some $\alpha > 0$. *Hint:* Utilize Exercise 11.

13. For events $\{A_n, n \geq 1\}$ and $\varepsilon > 0$, there exist distinct indices j, k in $[1, n]$, where $n > 1/\varepsilon$, such that $P(A_j A_k) \geq p_1^2(n) - \varepsilon$, where $p_1(n)$ is as in Exercise 12. *Hint:* $p_2(n) \geq p_1^2(n) - 1/n$.

14. **Bernstein's Inequality.** If $S_n = \sum_{j=1}^n X_j$ where $\{X_j, 1 \leq j \leq n\}$ are independent r.v.s with $E X_j = 0$, $E X_j^2 = \sigma_j^2$, $s_n^2 = \sum_{j=1}^n \sigma_j^2 > 0$ which satisfy (i) $E|X_j|^k \leq (k!/2)\sigma_j^2 c^{k-2}$ for $k > 2$, $1 \leq j \leq n$, $0 < c < \infty$ (a fortiori if (ii)' $P\{|X_j| \leq c\} = 1$, $1 \leq j \leq n$), then

$$P\{S_n > x\} \leq \exp\left\{\frac{-x^2}{2(s_n^2 + cx)}\right\}, \quad x > 0.$$

Hint: $\exp\{tX_j\} \leq 1 + tX_j + \sum_{k=2}^{\infty} t^k |X_j|^k / k!$ valid for $t > 0$ implies $E e^{tX_j} \leq \exp\{\sigma_j^2 t^2 / 2(1 - tc)\}$, $0 < tc < 1$ and choosing $t = x/(s_n^2 + cx)$ yields $E \exp\{xS_n/(s_n^2 + cx)\} \leq \exp\{x^2/2(s_n^2 + cx)\}$, $x > 0$. Now apply Exercise 4.3.3(ii).

15. If $S_n = \sum_{j=1}^n X_j$ where $\{X_j, 1 \leq j \leq n\}$ are independent r.v.s with $E X_j = 0$, $E X_j^2 = \sigma_j^2$, $s_n^2 = \sum_{j=1}^n \sigma_j^2 > 0$, then for $\lambda > \gamma > 1$,

$$P\left\{\max_{1 \leq j \leq n} |S_j| \geq \lambda s_n\right\} \leq \frac{\gamma^2}{\gamma^2 - 1} P\{|S_n| \geq (\lambda - \gamma)s_n\}.$$

Hint: If $T = \inf\{1 \leq j \leq n: |S_j| \geq \lambda s_n\}$ and $T = n + 1$ otherwise, then for $1 < \gamma < \lambda$

$$P\left\{\max_{1 \leq j \leq n} |S_j| \geq \lambda s_n\right\} \leq P\{|S_n| \geq (\lambda - \gamma)s_n\} + \sum_{j=1}^{n-1} P\{T = j\} \cdot P\{|S_n - S_j| \geq \gamma s_n\}.$$

Now apply Tchebychev's inequality.

16. If $\{S_n, X_n, n \geq 1\}$ are as in exercise 15 and $c_n \rightarrow \infty$, then $(1/c_n s_n) \max_{1 \leq j \leq n} |S_j| \xrightarrow{P} 0$.
17. For $X, Y \in \mathcal{L}_p$, $p > 1$, define $d(X, Y) = \|X - Y\|_p$, and show that d has all the attributes of a metric except one. For $X, Y \in \mathcal{L}_p$, write $X \sim Y$ if $X = Y$, a.e., and let \mathcal{L}_p^* be the space of equivalence classes of \mathcal{L}_p . Show that \mathcal{L}_p^* is a Banach space, i.e., a normed, linear, complete space. When $p = 2$, \mathcal{L}_p^* is a Hilbert space under the inner product $(X, Y) = E XY$.
18. Prove that if $S_n = \sum_{i=1}^n X_i$, where $\{X_i, i \geq 1\}$ are independent r.v.'s with $E X_i = 0$, $1 \leq i \leq n$, then for any $\varepsilon > 0$,

$$P\left\{\max_{1 \leq j \leq n} |S_j| > \varepsilon\right\} \leq 2P\{S_n \geq \varepsilon - E|S_n|\}. \quad (24)$$

Hint. Use the Feller-Chung lemma with $A_j = \{S_j > \varepsilon\}$, $B_j = \{S_n - S_j > -E|S_n|\}$.

References

- Y. S. Chow and W. J. Studden, "Monotonicity of the variance under truncation and variations of Jensen's inequality," *Ann. Math. Stat.* **40** (1969), 1106–1108.
 K. L. Chung and W. H. J. Fuchs, "On the distribution of values of sums of random variables," *Mem. Amer. Soc.* **6** (1951).
 J. L. Doob, *Stochastic Processes*, Wiley, New York, 1953.
 P. R. Halmos, *Measure Theory*, Van Nostrand, Princeton, 1950; Springer-Verlag, Berlin and New York, 1974.
 P. Hall, "On the \mathcal{L}_p convergence of random variables," *Proc. Cambridge Philos. Soc.* **82** (1977), 439–446.
 G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*, Cambridge Univ. Press, London, 1934.

- S. B. Kochen and C. J. Stone, "A note on the Borel–Cantelli lemma," *Ill. Jour. Math.* **8** (1964), 248–251.
- A. Liapounov, "Nouvelle forme du théorème sur la limite de probabilité," *Mem. Acad. Sc. St. Petersbourg* **12** (1905), No. 5.
- M. Loève, *Probability Theory*, 3rd ed., Van Nostrand, Princeton, 1963; 4th ed., Springer-Verlag, Berlin and New York, 1977–1978.
- G. Polya, "Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Strassennetz," *Math. Ann.* **84** (1921), 149–160.
- S. Saks, *Theory of the Integral* (L. C. Young, translation), Stechert–Hafner, New York, 1937.
- H. Teicher, "On the law of the iterated logarithm," *Ann. Prob.* **2** (1974), 714–728.

5 Sums of Independent Random Variables

Of paramount concern in probability theory is the behavior of sums $\{S_n, n \geq 1\}$ of independent random variables $\{X_i, i \geq 1\}$. The case where the $\{X_i\}$ are i.i.d. is of especial interest and frequently lends itself to more incisive results. The sequence of sums $\{S_n, n \geq 1\}$ of i.i.d. r.v.s $\{X_n\}$ is alluded to as a random walk; in the particular case when the component r.v.s $\{X_n\}$ are nonnegative, the random walk is referred to as a renewal process.

5.1 Three Series Theorem

The first question to be dealt with apropos of sums of independent r.v.s is when such sums converge a.c. A partial answer is given next, and the ensuing lemmas culminate in the Kolmogorov three series theorem.

Theorem 1 (Khintchine–Kolmogorov Convergence Theorem). *Let $\{X_n, n \geq 1\}$ be independent \mathcal{L}_2 r.v.s with $E X_n = 0, n \geq 1$. If $\sum_{j=1}^{\infty} E X_j^2 < \infty$, then $\sum_1^{\infty} X_j$ converges a.c. and in quadratic mean and, moreover, $E(\sum_1^{\infty} X_j)^2 = \sum_{j=1}^{\infty} E X_j^2$.*

PROOF. If $S_n = \sum_1^n X_j$, by Corollary 4.3.5

$$E(S_m - S_n)^2 = E\left(\sum_{n+1}^m X_j\right)^2 = \sum_{n+1}^m E X_j^2 \rightarrow 0 \quad (1)$$

as $m > n \rightarrow \infty$, whence, according to Theorem 4.2.3, $S_n \xrightarrow{\mathcal{L}_2}$ some r.v. S , denoted by $S = \sum_1^{\infty} X_j$. *A fortiori*, $S_n \xrightarrow{P} \sum_1^{\infty} X_j$ and so by Lévy's theorem

(Theorem 3.3.1) $S_n \xrightarrow{a.s.} \sum_1^\infty X_j$. The remainder follows from

$$E\left(\sum_1^\infty X_j\right)^2 = \lim_{n \rightarrow \infty} E S_n^2 = \lim_{n \rightarrow \infty} \sum_{j=1}^n E X_j^2 = \sum_1^\infty E X_j^2 \quad (2)$$

via Corollaries 4.2.5 and 4.3.5. \square

The first lemma involves “summation by parts” and is as useful as its integral counterpart.

Lemma 1 (Abel). If $\{a_n\}$, $\{b_n\}$ are sequences of real numbers and $A_n = \sum_{j=0}^n a_j$, $n \geq 0$, then for $n \geq 1$

$$\sum_{j=1}^n a_j b_j = A_n b_n - A_0 b_1 - \sum_{j=1}^{n-1} A_j (b_{j+1} - b_j); \quad (3)$$

if $\sum_{j=1}^\infty a_j$ converges and $A_n^* = \sum_{j=n+1}^\infty a_j$, then for $n \geq 1$

$$\sum_{j=1}^n a_j b_j = A_0^* b_1 - A_n^* b_n + \sum_{j=1}^{n-1} A_j^* (b_{j+1} - b_j); \quad (4)$$

moreover, if $a_n \geq 0$, $b_{n+1} \geq b_n \geq 0$, $A_n^* = \sum_{j=n+1}^\infty a_j < \infty$, then

$$\sum_{j=1}^\infty a_j b_j = A_0^* b_1 + \sum_{j=1}^\infty A_j^* (b_{j+1} - b_j). \quad (5)$$

PROOF.

$$\sum_1^n a_j b_j = \sum_1^n (A_j - A_{j-1}) b_j = \sum_1^n A_j b_j - \sum_1^n A_{j-1} b_j,$$

yielding (3). Take $a_0 = -\sum_{j=1}^\infty a_j = -A_0^*$ in (3) to obtain (4). Next, assuming $a_n \geq 0$, $b_{n+1} \geq b_n \geq 0$, if $\overline{\lim}_n A_n^* b_n > 0$, then $\sum_{n+1}^\infty a_j b_j \geq A_n^* b_n$ implies $\sum_1^\infty a_j b_j = \infty$. By (4), $A_0^* b_1 + \sum_1^\infty A_j^* (b_{j+1} - b_j) \geq \sum_1^\infty a_j b_j = \infty$, so that (5) holds. If, rather, $\overline{\lim} A_n^* b_n = 0$, then (5) obtains by letting $n \rightarrow \infty$ in (4). \square

The following **Kronecker lemma** is a *sine qua non* for probability theory as will become apparent in the next section.

Lemma 2 (Kronecker). If $\{a_n\}$, $\{b_n\}$ are sequences of real numbers with $0 < b_n \uparrow \infty$, $\sum_1^\infty (a_j/b_j)$ converging, then

$$\frac{1}{b_n} \sum_{j=1}^n a_j \rightarrow 0. \quad (6)$$

PROOF. This will be demonstrated in the alternative equivalent form that convergence of $\sum_1^\infty a_j$ entails $\sum_1^\infty a_j b_j = o(b_n)$. By (4),

$$\frac{1}{b_n} \sum_{j=1}^n a_j b_j = -A_n^* + A_0^* \frac{b_1}{b_n} + \frac{1}{b_n} \sum_{j=1}^{n-1} A_j^* (b_{j+1} - b_j). \quad (7)$$

For any $\varepsilon > 0$, choose the integer m so that $|A_j^*| < \varepsilon$ for $j \geq m$. Then

$$-\varepsilon \leq \liminf_n \frac{1}{b_n} \sum_{j=m}^{n-1} A_j^*(b_{j+1} - b_j) \leq \limsup_n \frac{1}{b_n} \sum_{j=m}^{n-1} A_j^*(b_{j+1} - b_j) \leq \varepsilon,$$

whence from (7)

$$-\varepsilon \leq \liminf_n \frac{1}{b_n} \sum_{j=1}^n a_j b_j \leq \limsup_n \frac{1}{b_n} \sum_{j=1}^n a_j b_j \leq \varepsilon$$

and $(1/b_n) \sum_{j=1}^n a_j b_j \rightarrow 0$. \square

Lemma 3. Let $\{X_n\}$ be independent r.v.s with $E X_n = 0$, $S_n = \sum_1^n X_i$, $n \geq 1$, and $E \sup_n |X_n|^2 < \infty$. If $P\{\sup_n |S_n| < \infty\} > 0$, then S_n converges a.c. and in quadratic mean and

$$\sum_1^\infty E |X_n|^2 < \infty. \quad (8)$$

PROOF. It suffices to prove (8) since Theorem 1 guarantees the rest. To this end, set $Z^2 = \sup_n |X_n|^2$ and choose $K > 0$ sufficiently large so that

$$P\left\{\sup_n |S_n| < K\right\} > 0.$$

Define $T = \inf\{n: |S_n| \geq K\}$ and note that T is positive integer valued with $P\{T = \infty\} > 0$. Since $\{T \geq n\} = \{|S_j| < K, 1 \leq j < n\} \in \sigma(X_1, \dots, X_{n-1})$ for $n \geq 2$, the r.v.s X_n and $I_{\{T \geq n\}}$ are independent for $n \geq 1$. Let $U_n = \sum_{j=1}^n X_j I_{\{T \geq j\}}$ and observe that U_n is $\sigma(X_1, \dots, X_n)$ -measurable and

$$\begin{aligned} U_n &= S_{\min(T, n)}, \\ U_n^2 &= |S_{\min(T-1, n-1)} + X_{\min(T, n)}|^2 \leq 2(K^2 + Z^2), \\ E U_n^2 &\leq 2(K^2 + E Z^2) = C < \infty. \end{aligned} \quad (9)$$

Now, setting $U_0 = 0$, for $j \geq 1$

$$U_j^2 = U_{j-1}^2 + 2U_{j-1} X_j I_{\{T \geq j\}} + X_j^2 I_{\{T \geq j\}},$$

and so by independence and Theorem 4.3.3

$$E U_j^2 - E U_{j-1}^2 = P\{T \geq j\} E X_j^2.$$

Summing over $1 \leq j \leq n$,

$$C \geq E U_n^2 = \sum_{j=1}^n P\{T \geq j\} E X_j^2 \geq P\{T = \infty\} \sum_{j=1}^n E X_j^2,$$

which yields (8) when $n \rightarrow \infty$. \square

Lemma 4. If $\{X_n\}$ are independent r.v.s with $E \sup_n |X_n| < \infty$ and $S_n = \sum_1^n X_i$ converges a.c., then $\sum_1^\infty E X_n$ converges.

PROOF. Define K, T, U_n as in Lemma 3, which is permissible since $S_n \xrightarrow{\text{a.c.}} S$ some r.v. S . Now, $\min[T, n] \rightarrow T$ and so (9) ensures that

$$U_n = S_{\min(T, n)} \xrightarrow{\text{a.c.}} S_T, \quad (10)$$

where $S_T = S_n$ on $\{T = n\}, n \geq 1$, and $S_T = S$ on $\{T = \infty\}$. A more extensive discussion of S_T occurs in Section 3 of this chapter, where it is pointed out that S_T is a bona fide r.v. As in (9),

$$\mathbb{E} \sup_n |U_n| \leq K + \mathbb{E} \sup_n |X_n| < \infty,$$

which, in conjunction with (10) and the Lebesgue dominated convergence theorem, ensures

$$\lim_{n \rightarrow \infty} \mathbb{E} U_n = \mathbb{E} S_T, \quad \text{finite.} \quad (11)$$

By the independence observed in Lemma 3,

$$\mathbb{E} U_n = \mathbb{E} \sum_{j=1}^n X_j I_{\{T \geq j\}} = \sum_1^n \mathbb{P}\{T \geq j\} \mathbb{E} X_j,$$

whence

$$\mathbb{E} X_n = \frac{\mathbb{E} U_n - \mathbb{E} U_{n-1}}{\mathbb{P}\{T \geq n\}}.$$

Employing Lemma 1 with $b_j = 1/\mathbb{P}\{T \geq j\}$, $a_j = \mathbb{E} U_j - \mathbb{E} U_{j-1}$, $j \geq 1$, $a_0 = A_0 = 0$,

$$\sum_{j=1}^n \mathbb{E} X_j = \frac{\mathbb{E} U_n}{\mathbb{P}\{T \geq n\}} - \sum_{j=1}^{n-1} \left[\frac{1}{\mathbb{P}\{T > j\}} - \frac{1}{\mathbb{P}\{T \geq j\}} \right] \mathbb{E} U_j$$

and so, recalling (11) and $\mathbb{P}\{T = \infty\} > 0$, $\sum_1^\infty \mathbb{E} X_j$ converges. \square

Corollary 1. If $\{X_n, n \geq 1\}$ are independent r.v.s which are uniformly bounded, i.e., $\mathbb{P}\{|X_n| \leq C < \infty, n \geq 1\} = 1$, and moreover, $S_n = \sum_1^n X_i \xrightarrow{\text{a.c.}} S$ finite, then $\sum_1^\infty \mathbb{E} X_j$ and $\sum_{j=1}^\infty \sigma^2(X_j)$ converge.

PROOF. The series of means converges by Lemma 4, whence $\sum_1^\infty (X_i - \mathbb{E} X_i)$ converges a.c. and Lemma 3 applies. \square

Definition. Two sequences of r.v.s $\{X_n\}$ and $\{Y_n\}$ will be called **equivalent** if $\sum_1^\infty \mathbb{P}\{X_n \neq Y_n\} < \infty$.

If $\{X_n\}$, $\{Y_n\}$ are equivalent sequences of r.v.s, the Borel-Cantelli lemma ensures that $\mathbb{P}\{X_n \neq Y_n, \text{ i.o.}\} = 0$. Hence, $\mathbb{P}\{\sum X_i \text{ converges}\} = 1$ iff $\mathbb{P}\{\sum Y_i \text{ converges}\} = 1$.

The way is now paved for presentation of

Theorem 2 (Kolmogorov Three Series Theorem). *If $\{X_n\}$ are independent r.v.s, then $\sum_1^\infty X_i$ converges a.c. iff*

- i. $\sum_1^\infty P\{|X_n| > 1\} < \infty$,
- ii. $\sum_1^\infty E[X'_n]$ converges,
- iii. $\sum_1^\infty \sigma_{X'_n}^2 < \infty$,

where $X'_n = X_n I_{\{|X_n| \leq 1\}}$, $n \geq 1$.

PROOF. Sufficiency: If the three series converge, then $\sum_1^\infty (X'_n - E X'_n)$ converges a.c. by Theorem 1, whence (ii) implies that $\sum_1^\infty X'_n$ converges a.c. According to (i), $\{X_n\}$, $\{X'_n\}$ are equivalent sequences of r.v.s and so $\sum X_i$ converges a.c.

Conversely, if $\sum X_i$ converges a.c., then $X_n \xrightarrow{a.c.} 0$, implying $P\{|X_n| > 1, i.o.\} = 0$, whence (i) holds by the Borel-Cantelli theorem (Theorem 3.1.1). Also, $\{X_n\}$, $\{X'_n\}$ are equivalent sequences, so that necessarily $\sum_1^\infty X'_n$ converges a.c. The remaining series, (ii) and (iii), now converge by Corollary 1. \square

Corollary 2. *If $\{X_n\}$ are independent r.v.s satisfying $E X_n = 0$, $n \geq 1$, and*

$$\sum_1^\infty E[X_n^2 I_{\{|X_n| \leq 1\}} + |X_n| I_{\{|X_n| > 1\}}] < \infty, \quad (12)$$

then $\sum_1^\infty X_n$ converges a.c.

PROOF. Since $E X_n = 0$,

$$\sum_1^\infty |E X_n I_{\{|X_n| \leq 1\}}| = \sum_1^\infty |E X_n I_{\{|X_n| > 1\}}| \leq \sum_1^\infty E|X_n| I_{\{|X_n| > 1\}} < \infty.$$

Moreover, by the Markov inequality (Theorem 4.1.1(iii))

$$P\{|X_n| > 1\} = P\{|X_n| I_{\{|X_n| > 1\}} > 1\} \leq E|X_n| I_{\{|X_n| > 1\}},$$

whence the corollary flows from Theorem 2. \square

Corollary 3 (Loève). *If $\{X_n\}$ are independent r.v.s and for some constants $0 < \alpha_n \leq 2$, $\sum_1^\infty E|X_n|^{\alpha_n} < \infty$, where $E X_n = 0$ when $1 \leq \alpha_n \leq 2$, then $\sum_1^\infty X_n$ converges a.c.*

PROOF. It suffices to consider separately the cases $1 \leq \alpha_n \leq 2$, $n \geq 1$, and $0 < \alpha_n < 1$, $n \geq 1$. In the former instance, $X_n^2 I_{\{|X_n| \leq 1\}} + |X_n| I_{\{|X_n| > 1\}} \leq |X_n|^{\alpha_n}$, whence (12) obtains. In the latter,

$$\sum_1^\infty E(X_n^2 + |X_n|) I_{\{|X_n| \leq 1\}} \leq 2 \sum_1^\infty E|X_n|^{\alpha_n} < \infty,$$

and in both cases

$$\sum_1^\infty P[|X_n| \geq 1] \leq \sum_1^\infty E|X_n|^{\alpha_n} < \infty,$$

so that the three series of Theorem 2 converge. \square

In much of probability theory, integration is with respect to a single probability measure P , and so it seems natural to write $\int_A X$ as an abbreviation for $\int_A X dP$. Abundant use will be made of this concise notation.

Turning to the i.i.d. case:

Theorem 3 (Marcinkiewicz-Zygmund). *If $\{X_n\}$ are i.i.d. with $E|X_1|^p < \infty$ for some p in $(0, 2)$, then*

$$\sum_1^{\infty} \left(\frac{X_n}{n^{1/p}} - E Y_n \right)$$

converges a.c., where

$$Y_n = n^{-1/p} X_n I_{\{|X_n| \leq n^{1/p}\}}.$$

Furthermore, if either (i) $0 < p < 1$ or (ii) $1 < p < 2$ and $E X_1 = 0$, then $\sum_1^{\infty} X_n/n^{1/p}$ converges a.c.

PROOF. Set $A_j = \{(j-1)^{1/p} < |X_1| \leq j^{1/p}\}$, $j \geq 1$. Then for $\alpha > p > 0$

$$\begin{aligned} \sum_1^{\infty} E|Y_n|^2 &= \sum_{n=1}^{\infty} \sum_{j=1}^n n^{-\alpha/p} \int_{A_j} |X_1|^2 = \sum_{j=1}^{\infty} \sum_{n=j}^{\infty} n^{-\alpha/p} \int_{A_j} |X_1|^2 \\ &\leq \sum_{j=1}^{\infty} \left(j^{-\alpha/p} + \frac{p}{\alpha-p} j^{(\alpha-p)/p} \right) \int_{A_j} |X_1|^2 \\ &\leq \sum_{j=1}^{\infty} \left(\frac{1}{j} + \frac{p}{\alpha-p} \right) \int_{A_j} |X_1|^p \leq \frac{\alpha}{\alpha-p} E|X_1|^p < \infty, \quad (13) \end{aligned}$$

whence ($\alpha = 2$) $\sum_1^{\infty} (Y_n - E Y_n)$ converges a.c. by Theorem 1. Since, recalling Corollary 4.1.3,

$$\sum_1^{\infty} P \left\{ \frac{X_n}{n^{1/p}} \neq Y_n \right\} = \sum_1^{\infty} P \{ |X_1| > n^{1/p} \} \leq E|X_1|^p < \infty,$$

the sequences $\{X_n/n^{1/p}\}$, $\{Y_n\}$ are equivalent, and so $\sum_1^{\infty} (X_n/n^{1/p} - E Y_n)$ converges a.c.

In case (i), where $0 < p < 1$, $\sum_{n=1}^{\infty} |E Y_n| < \infty$ via (13) with $\alpha = 1$, and this same series converges in case (ii), $1 < p < 2$ and $E X_1 = 0$, since

$$\begin{aligned} \sum_1^{\infty} |E Y_n| &\leq \sum_1^{\infty} n^{-1/p} \int_{\{|X_n| > n^{1/p}\}} |X_n| = \sum_{n=1}^{\infty} \sum_{j=n+1}^{\infty} n^{-1/p} \int_{A_j} |X_1| \\ &= \sum_{j=2}^{\infty} \sum_{n=1}^{j-1} n^{-1/p} \int_{A_j} |X_1| \leq \frac{p}{p-1} \sum_{j=1}^{\infty} (j-1)^{(p-1)/p} \int_{A_j} |X_1| \\ &\leq \frac{p}{p-1} \sum_1^{\infty} \int_{A_j} |X_1|^p = \frac{p}{p-1} E|X_1|^p < \infty. \end{aligned}$$

Thus, the second part of Theorem 3 follows from the first. \square

EXAMPLE 1. Let $\{X_n, n \geq 1\}$ be i.i.d. r.v.s with $E|X_1| < \infty$. If $\{a_n, n \geq 1\}$ are real numbers such that $a_n = O(1/n)$ and $\sum_1^\infty a_n$ converges, then $\sum_{n=1}^\infty a_n X_n$ converges a.c.

PROOF. By considering X_n^+ and X_n^- separately, it may and will be supposed that $X_1 \geq 0$. Set

$$X'_n = X_n I_{\{X_n \leq n\}}, \quad Y_n = X'_n - E X'_n.$$

By the Borel–Cantelli theorem, $\{X'_n\}$ and $\{X_n\}$ are equivalent sequences, and from (4) it follows that a convergent series remains convergent when its terms are multiplied by any convergent, monotone sequence (Abel's test). Thus, $E X'_n \uparrow E X_1$ implies $\sum a_n E X'_n$ converges. Consequently, to prove $\sum a_n X_n$ converges a.c., it suffices to prove $\sum a_n Y_n$ converges a.c. By hypothesis, $n^2 a_n^2 \leq A < \infty$, and so, setting $A_j = \{j-1 < X_1 \leq j\}$,

$$\begin{aligned} \sum_1^\infty a_n^2 E Y_n^2 &\leq \sum_1^\infty a_n^2 E X_n'^2 \leq A \sum_{n=1}^\infty n^{-2} \int_{[X_1 \leq n]} X_1^2 \\ &\leq A \sum_{n=1}^\infty \sum_{j=1}^n n^{-2} \int_{A_j} X_1^2 \leq 2A E X_1, \end{aligned}$$

where the last inequality follows via (13) with $\alpha = 2, p = 1$. Hence, $\sum a_n Y_n$ converges a.c. by the Khintchine–Kolmogorov convergence theorem. \square

EXAMPLE 2. If $\{b_n, n \geq 1\}$ are constants satisfying $0 < b_n \uparrow \infty$ and

$$b_n^2 \sum_{j=n}^\infty b_j^{-2} = O(n) \tag{14}$$

and $\{X, X_n, n \geq 1\}$ are i.i.d. r.v.s with $\sum_{n=1}^\infty P\{|X| > b_n\} < \infty$, then

$$\sum_{n=1}^\infty b_n^{-1} (X_n - E X I_{\{|X| \leq b_n\}}) \tag{15}$$

converges a.c. Moreover, if $E X = 0$ and

$$b_n \sum_{j=1}^n b_j^{-1} = O(n), \tag{16}$$

then

$$\sum_{n=1}^\infty \frac{X_n}{b_n} \tag{17}$$

converges a.c.

Remark. If $b_n/n \uparrow$ or $b_j/b_n \geq A(j/n)^\delta$ for $j \geq n$, where $\delta > \frac{1}{2}, A > 0$, then (14) holds; if $b_j/b_n \geq A(j/n)^\delta$ for $j \leq n$, where $0 < \delta < 1, A > 0$, then (16) obtains.

PROOF. Set $b_0 = 0$, $A_n = \{b_{n-1} < |X| \leq b_n\}$, and $Y_n = X_n I_{\{|X_n| \leq b_n\}}$. Then

$$\infty > \sum_{n=0}^{\infty} P\{|X_n| > b_n\} = \sum_{n=0}^{\infty} \sum_{j=n+1}^{\infty} P\{A_j\} = \sum_{j=1}^{\infty} j P\{A_j\} \quad (18)$$

whence for some C in $(0, \infty)$,

$$\begin{aligned} \sum_{j=1}^{\infty} b_j^{-2} \sigma^2(Y_j) &\leq \sum_{j=1}^{\infty} b_j^{-2} E Y_j^2 \leq \sum_{j=1}^{\infty} b_j^{-2} \sum_{n=1}^j \int_{A_n} X_j^2 \\ &= \sum_{n=1}^{\infty} \sum_{j=n}^{\infty} b_j^{-2} \int_{A_n} X^2 \leq \sum_{n=1}^{\infty} b_n^2 \sum_{j=n}^{\infty} b_j^{-2} P\{A_n\} \\ &\leq C \sum_{n=1}^{\infty} n P\{A_n\} < \infty \end{aligned}$$

via (14) and (18). By Theorem 5.1.1, $\sum_1^{\infty} b_j^{-1}(Y_j - E Y_j)$ converges a.c., and so, again employing (18), $\sum_1^{\infty} b_j^{-1}(X_j - E Y_j)$ converges a.c., yielding (15).

Moreover, if $E X = 0$,

$$\begin{aligned} \sum_{j=1}^{\infty} b_j^{-1} |E Y_j| &\leq \sum_{j=1}^{\infty} b_j^{-1} E |X| I_{\{|X| > b_j\}} = \sum_{j=1}^{\infty} b_j^{-1} \sum_{n=j}^{\infty} \int_{A_{n+1}} |X| \\ &= \sum_{n=1}^{\infty} \sum_{j=1}^n b_j^{-1} \int_{A_{n+1}} |X| \leq \sum_{n=1}^{\infty} b_{n+1} \sum_{j=1}^n b_j^{-1} P\{A_{n+1}\} \\ &\leq C \sum_{n=1}^{\infty} (n+1) P\{A_{n+1}\} < \infty \end{aligned}$$

via (16) and (18). Consequently, (17) follows via (15). \square

Theorem 4. Let $\{X_n\}$ be independent r.v.s such that for some positive ε, δ

$$\inf_{n \geq 1} P\{|X_n| > \varepsilon\} = \delta, \quad (19)$$

and suppose that $\sum_1^{\infty} a_n X_n$ converges a.c. for some sequence $\{a_n\}$ of real numbers. If either (i) $X_n, n \geq 1$, are i.i.d. and nondegenerate or (ii) $E X_n = 0, n \geq 1$, and $\{X_n\}$ is u.i., then

$$\sum_1^{\infty} a_n^2 < \infty. \quad (20)$$

PROOF. Since $\sum_1^{\infty} a_n X_n$ converges a.c., necessarily $a_n X_n \xrightarrow{P} 0$, implying via (19) that $a_n \rightarrow 0$. In proving (20), it may be supposed that $a_n \neq 0, n \geq 1$. If $Y_n = X_n I_{\{|a_n X_n| \leq 1\}}$, the three series theorem requires that

$$\sum_1^{\infty} a_n^2 \sigma_{Y_n}^2 < \infty. \quad (21)$$

In case (i), $\sigma_{Y_n}^2 = E(X_1 I_{\{|a_n X_1| \leq 1\}} - E Y_n)^2$. If $\sigma_{Y_{n_j}}^2 \rightarrow 0$ for some subsequence n_j , then $X_1 I_{\{|a_{n_j} X_1| \leq 1\}} - E Y_{n_j} \xrightarrow{P} 0$, implying X_1 degenerate, contrary to the hypothesis of (i). Thus $\lim_{n \rightarrow \infty} \sigma_{Y_n}^2 > 0$, whence (21) implies (20).

Under (ii), since $E X_n = 0$,

$$|E Y_n| = \left| \int_{\{|X_n| > |a_n|^{-1}\}} X_n \right| \leq \int_{\{|X_n| > |a_n|^{-1}\}} |X_n| = o(1)$$

by uniform integrability and $a_n \rightarrow 0$. Hence,

$$\begin{aligned} \liminf_n \sigma_{Y_n}^2 &= \liminf_n (E Y_n^2 - E^2 Y_n) = \liminf_n E Y_n^2 \geq \liminf_n E^2 |Y_n| \\ &= \liminf_n \left(E |X_n| - \int_{\{|a_n X_n| > 1\}} |X_n| \right)^2 = \liminf_n E^2 |X_n| \geq \varepsilon^2 \delta^2 \end{aligned}$$

by (19), and once more (20) flows from (21). \square

Corollary 4 (Marcinkiewicz-Zygmund). *If $\{X_n\}$ are independent r.v.s with $E X_n = 0$, $E X_n^2 = 1$, $n \geq 1$, and $\inf_n E|X_n| > 0$, then a.c. convergence of $\sum a_n X_n$ for some real numbers a_n entails convergence of $\sum a_n^2$.*

PROOF. Uniform integrability of $\{X_n\}$ is implied by $E X_n^2 = 1$, $n \geq 1$, while $\inf_n E|X_n| > 0$ and $E X_n^2 = 1$ ensure (19) (Exercise 5). \square

Definition. A series $\sum_{n=1}^{\infty} X_n$ of r.v.s will be said to **converge a.c. unconditionally** if $\sum_{k=1}^{\infty} X_{n_k}$ converges a.c. for every rearrangement (n_1, n_2, \dots) of $(1, 2, \dots)$. More specifically, a rearrangement is a one-to-one map of the positive integers onto the positive integers.

In the case of degenerate r.v.s X_n , $n \geq 1$, the series $\sum_1^{\infty} X_n$ converges unconditionally iff $\sum |X_n|$ converges (Weierstrass). However, the analogue for nondegenerate independent random variables is invalid. In fact, if X_n , $n \geq 1$, are independent r.v.s with $P\{X_n = \pm 1/n\} = \frac{1}{2}$, then $\sum X_n$ converges a.c. unconditionally by Theorem 1, but $\sum |X_n| = \infty$, a.c.

Lemma 5. *If $\{X_n\}$ are independent r.v.s with $E X_n = 0$, $n \geq 1$, and $\sum_1^{\infty} E X_n^2 < \infty$, then $\sum X_n$ converges a.c. unconditionally and $\sum_{j=1}^{\infty} X_{n_j} = \sum_1^{\infty} X_j$, a.c., for every rearrangement $\{n_j\}$ of $\{j\}$.*

PROOF. Theorem 1 ensures that $\sum X_n$ converges a.c. unconditionally. Moreover, for any fixed rearrangement $\{n_j\}$, define $S'_m = \sum_{j=1}^m X_{n_j}$, $S_m = \sum_1^m X_j$. Then, setting $Q = \{n_1, \dots, n_m\} \Delta \{1, 2, \dots, m\}$,

$$E(S'_m - S_m)^2 = \sum_{k \in Q} E X_k^2.$$

Now, if $\{n_1, \dots, n_m\} \supset \{1, 2, \dots, j\}$, then

$$E(S'_m - S_m)^2 \leq \sum_{j+1}^{\infty} E X_k^2 = o(1) \quad \text{as } j \rightarrow \infty.$$

Hence, $S'_m - S_m \xrightarrow{P} 0$ as $m \rightarrow \infty$, implying $\sum_{j=1}^{\infty} X_{n_j} = \sum_1^{\infty} X_j$ a.c. \square

Theorem 5. A series $\sum_1^\infty X_n$ of independent r.v.s X_n converges a.c. unconditionally iff for $Y_n = X_n I_{\{|X_n| \leq 1\}}$

- i'. $\sum_1^\infty P\{|X_n| > 1\} < \infty$,
- ii'. $\sum_1^\infty |E Y_n| < \infty$,
- iii'. $\sum_1^\infty E Y_n^2 < \infty$,

and if so, $\sum_1^\infty X_{n_j} = \sum_1^\infty X_j$ a.c. for every rearrangement $\{n_j\}$ of $\{j\}$. Moreover, a series of independent r.v.s $\{X_n\}$ converges absolutely a.c. iff (i'), (iii'), and

- ii''. $\sum_1^\infty E|Y_n| < \infty$

hold.

PROOF. Since the series appearing in (i'), (ii'), (iii') are independent of the order of summation, the three series theorem (Theorem 2) guarantees that $\sum_1^\infty X_n$ converges a.c. unconditionally, and by Lemma 5

$$\sum_1^\infty (Y_{n_j} - E Y_{n_j}) = \sum_1^\infty (Y_j - E Y_j), \quad \text{a.c.}$$

Then, in view of (ii'), $\sum_1^\infty Y_{n_j} = \sum_1^\infty Y_j$ a.c., whence (i') ensures that $\sum_1^\infty X_{n_j} = \sum_1^\infty X_j$ a.c.

Conversely, if $\sum X_n$ converges a.c. unconditionally, (i), (ii), (iii) of Theorem 2 are valid for every rearrangement $\{n_j\}$ of $\{j\}$. By the Weierstrass theorem (ii') holds, and hence also $\sum_1^\infty E^2 Y_n < \infty$. But this latter and (iii) entail (iii').

The proof of the final statement is similar. \square

Corollary 5. If the series $\sum_1^\infty X_n$ of independent r.v.s X_n converges a.c., then $\sum (X_n - c_n)$ converges a.c. unconditionally, where

$$c_n = E X_n I_{\{|X_n| \leq 1\}}. \quad (22)$$

PROOF. Set $Y_n = X_n I_{\{|X_n| \leq 1\}}$. By the three series theorem

$$\sum_1^\infty \sigma_{Y_n}^2 < \infty, \quad \sum_1^\infty P\{X_n \neq Y_n\} < \infty, \quad (23)$$

whence Lemma 5 guarantees that $\sum_1^\infty (Y_n - E Y_n)$ converges a.c. unconditionally. Then by (23), $\sum_1^\infty (X_n - E Y_n)$ converges a.c. unconditionally. \square

EXERCISES 5.1

1. Let $S_n = \sum_1^n X_i$, where $\{X_n, n \geq 1\}$ are independent r.v.s. (i) If $\sum_{n=1}^\infty P\{|X_n| > c\} = \infty$, all $c > 0$, then $\overline{\lim}_{n \rightarrow \infty} |S_n| = \infty$, a.c. (ii) If $\{b_n, n \geq 1\}$ are positive constants for which $\underline{\lim}_{n \rightarrow \infty} P\{S_{n-1} > -\delta b_n\} > 0$ for all $\delta > 0$, then $\overline{\lim} S_n/b_n \leq C < \infty$, a.c., implies $\sum_{n=1}^\infty P\{X_n > \varepsilon b_n\} < \infty$ for all $\varepsilon > C$. Hint: Recall Lemma 3.3.4.
2. Any series $\sum_{i=1}^\infty X_i$ of r.v.s converges absolutely a.c. if $\sum_{n=1}^\infty E|X_n|r^n < \infty$ for some sequence r_n in $(0, 1]$.

3. If $X_n, n \geq 1$, are independent r.v.s and $Y_n^c = X_n I_{\{|X_n| \leq c\}}$, then a.c. convergence of $\sum X_n$ ensures convergence of $\sum P\{|X_n| > c\}$, $\sum E Y_n^c$, $\sum \sigma_{Y_n}^2$ for every $c > 0$. Conversely, convergence of these series for some $c > 0$ guarantees a.c. convergence of $\sum X_n$.
4. If $\{X_n\}, \{Y_n\}$ are equivalent sequences of r.v.s, prove that
 - i. $P\{\sum X_n \text{ converges}\} = 1$ iff $P\{\sum Y_n \text{ converges}\} = 1$,
 - ii. if $0 < b_n \uparrow \infty$, then $P\{\sum X_i = o(b_n)\} = 1$ iff $P\{\sum Y_i = o(b_n)\} = 1$.
5. Let $\{X_n\}$ be u.i.r.v.s. Then $\inf P[|X_n| > \varepsilon] > 0$, for some $\varepsilon > 0$, iff $\inf_n E|X_n| > 0$.
6. If $X_n, n \geq 1$, are i.i.d. \mathcal{L}_1 r.v.s, then $\sum (X_n/n)$ converges a.c. if either (i) X_1 is symmetric or (ii) $E|X_1| \log^+ |X_1| < \infty$ and $E X_1 = 0$.
7. Let $\{a_n\}$ be a sequence of positive numbers with $\sum a_n = \infty$. Then, if $p > 0$, there exist independent r.v.s X_n with $E X_n = 0$, $E|X_n|^p = a_n$ such that $\sum X_n$ diverges a.c., thereby furnishing a partial converse to Corollary 3 when $0 < p \leq 2$.
8. If $X_n, n \geq 1$, are independent r.v.s with $P\{X_n = 1\} = P\{X_n = -1\} = \frac{1}{2}$, then $\sum X_n/\sqrt{n}$ diverges a.c. although $\sum E|X_n/\sqrt{n}|^p < \infty$, all $p > 2$. Thus, the restriction of exponents in Corollary 2 is essential.
9. If $\{A_n, n \geq 1\}$ are independent events with $P\{A_n\} > 0, n \geq 1$ and $\sum_{n=1}^{\infty} P\{A_n\} = \infty$ then $\sum_{j=1}^n I_{A_j}/\sum_{j=1}^n P\{A_j\} \xrightarrow{a.e.} 1$. Hint: If $a_n = \sum_{j=1}^n P\{A_j\}$ and $X_j = (I_{A_j} - P\{A_j\})/a_j$ then $\{X_j, j \geq 1\}$ are independent with $E X_j = 0$, $E X_j^2 \leq P\{A_j\}/a_j^2 \leq 1/a_{j-1} - 1/a_j$, $j > 1$.
10. If $X_n, n \geq 1$, are independent r.v.s with $p_n = P\{X_n = a_n\} = 1 - P\{X_n = -a_n\}$, characterize the sequences (a_n, p_n) for which $\sum X_n$ converges a.c.; specialize to $p_n \equiv \frac{1}{2}$, to $a_n \equiv a$, and to $a_n = n^{-x}$ ($x > 0$).
11. If, in the three series theorem, the alternative truncation $Z_n = \min[1, \max(X_n, -1)]$ is employed, then convergence of the two series $\sum E Z_n, \sum \sigma_{Z_n}^2$ is equivalent to the a.c. convergence of $\sum X_n$.
12. For any sequence of r.v.s $\{S_n\}$, it is always possible to find constants $0 < a_n \uparrow \infty$ for which $S_n/a_n \xrightarrow{a.e.} 0$. Hint: If $0 < \epsilon_n \downarrow 0$, choose $a_n > a_{n-1}$ such that $P\{|S_n| > a_n \epsilon_n\} < 2^{-n}$.
13. (Chung) Let Ψ be a positive, even function with $x^{-2}\Psi(x) \downarrow, |x|^{-1}\Psi(x) \uparrow$ as $|x| \uparrow$. If $0 < b_n \uparrow \infty$, $\{X_n\}$ are independent with $E X_n = 0$, $\sum (E\Psi(X_n)/\Psi(b_n)) < \infty$, then $\sum (X_n/b_n)$ converges a.c. Hint: Apply Corollary 2 with X_n replaced by X_n/b_n .
14. If $\{X_n, n \geq 1\}$ are i.i.d. r.v.s with $E|X_1| < \infty$, prove that $\sum_{n=1}^{\infty} X_n(\sin nt)/n$ converges a.c. for every $t \in (-\infty, \infty)$. Conversely, a.c. convergence of this series for some $t \neq k\pi, k$ an integer and i.i.d. $\{X_n, n \geq 1\}$ implies $E|X_1| < \infty$. Hint: For $m = 1, 2, \dots$, choose integers n_m so that $n_m t \in (2m\pi + (\pi/4), 2m\pi + (\pi/2))$ for $t = \pi/4$.
15. If $\{b_n, n \geq 1\}$ are finite constants with $0 < b_n \uparrow \infty$ and (i) $b_n^2 \sum_{j=n}^{\infty} b_j^{-2} = O(d(b_n))$, where d is a nondecreasing mapping of $[0, \infty)$ into $[0, \infty)$ for which (ii) $d(b_n) \geq cn > 0, n \geq 1$, and (iii) $x^2/d(|x|) \uparrow$ as $|x| \uparrow$, then for any i.i.d. r.v.s $\{X, X_n, n \geq 1\}$ with (iv) $E d(|X|) < \infty$, the conclusion (15) obtains. Moreover, if $E X = 0$ and (v) $|x|/d(|x|) \downarrow$ as $|x| \uparrow$ and (vi) $b_n \sum_{j=n}^{\infty} b_j^{-1} = O(d(b_n))$, then (17) holds.
16. If $\{X, X_n, n \geq 1\}$ are i.i.d. r.v.s with $E X = 0, E X^2(1 + \log^+ |X|)^{-2\delta} < \infty$ for some $\delta > 0$, then $\sum n^{-1/2} (\log n)^{-(1/2)-\delta} X_n$ converges a.c.

5.2 Laws of Large Numbers

In a sense, a.c. convergence of a series of independent r.v.s X_n is atypical and in the nondegenerate i.i.d. case it is nonexistent. Thus, the issue rather becomes one of the magnitude of the partial sums $S_n = \sum_1^n X_i$. When $P\{X_n = 1\} = p = 1 - P\{X_n = 0\}$, so that S_n is a binomial r.v. with parameters n and p , it was proved in Theorem 2.2.3 that

$$\frac{S_n - np}{n} = \frac{1}{n} \sum_{i=1}^n (X_i - E X_i) \xrightarrow{\text{a.c.}} 0.$$

Definition. A sequence $\{X_n\}$ of \mathcal{L}_1 r.v.s is said to obey the **classical strong law of large numbers (SLLN)** if

$$(a) \quad \frac{1}{n} \sum_{i=1}^n (X_i - E X_i) \xrightarrow{\text{a.c.}} 0.$$

If, merely,

$$(b) \quad \frac{1}{n} \sum_{i=1}^n (X_i - E X_i) \xrightarrow{P} 0,$$

the sequence $\{X_n\}$ satisfies the **classical weak law of large numbers (WLLN)**.

From a wider vista, n may not reflect the real magnitude and the expectations need not exist. Thus, there is occasion to consider the more general **strong and weak laws of large numbers**

$$\frac{1}{a_n} \sum_{i=1}^n (X_i - b_i) \xrightarrow{\text{a.c.}} 0,$$

where $0 < a_n \uparrow \infty$. Here, the smaller the order of magnitude of a_n , the more precise the SLLN becomes; the fuzzy notion of an optimal choice of a_n impinges on the law of the iterated logarithm (Chapter 10). Note, in this context, Exercise 5.1.12.

The first SLLN may be reaped as a direct application of Kronecker's lemma to Corollary 3 of the three series theorem (Theorem 5.1.2), thereby obtaining Loève's generalization of a result of Kolmogorov ($\alpha_n \equiv 2$).

Theorem 1. If $\{X_n\}$ are independent r.v.s satisfying

$$\sum_1^\infty \frac{E|X_n|^{\alpha_n}}{n^{\alpha_n}} < \infty \tag{1}$$

for some choice of α_n in $(0, 2]$, where $E X_n = 0$ whenever $1 \leq \alpha_n \leq 2$, then

$$\frac{1}{n} \sum_{j=1}^n X_j \xrightarrow{\text{a.c.}} 0. \tag{2}$$

Corollary 1. Let $\{X_n\}$ be independent \mathcal{L}_2 r.v.s with $E X_n = 0$, $\sum_1^\infty E X_n^2/n^2 < \infty$. Then (2) holds.

According to Corollary 1, independent r.v.s with means zero and variances $n(\log n)^{-\delta}$ obey the classical SLLN if $\delta > 1$. This remains true for any $\delta > 0$ by Corollary 10.1.4. provided (as is necessary) $\sum_{n=1}^\infty P\{|X_n| \geq nz\} < \infty$, $z > 0$.

In the i.i.d. case, the next theorem gives a generalization due to Marcinkiewicz and Zygmund of a classical SLLN ($p = 1$) of Kolmogorov.

Theorem 2 (Marcinkiewicz–Zygmund). If $\{X_n\}$ are i.i.d. r.v.s and $S_n = \sum_1^n X_i$, then for any p in $(0, 2)$

$$\frac{S_n - nc}{n^{1/p}} \xrightarrow{\text{a.c.}} 0 \quad (3)$$

for some finite constant c iff $E|X_1|^p < \infty$, and if so, $c = E X_1$ when $1 \leq p < 2$ while c is arbitrary (and hence may be taken as zero) for $0 < p < 1$.

PROOF. If (3) holds, then

$$\frac{X_n}{n^{1/p}} = \frac{S_n - nc}{n^{1/p}} - \left(\frac{n-1}{n}\right)^{1/p} \frac{S_{n-1} - nc}{(n-1)^{1/p}} \xrightarrow{\text{a.c.}} 0,$$

whence by the Borel–Cantelli theorem $\sum_1^\infty P\{|X_1| \geq n^{1/p}\} < \infty$. Thus, $E|X_1|^p < \infty$ by Corollary 4.1.3. Conversely, if $E|X_1|^p < \infty$, by Theorem 5.1.3 the following series converge a.c.:

$$\begin{aligned} \sum_1^\infty \frac{X_n - E X_n}{n^{1/p}}, & \quad 1 < p < 2; \\ \sum_1^\infty \frac{X_n - E X_n I_{\{|X_n| \leq n\}}}{n^{1/p}}, & \quad p = 1; \\ \sum_1^\infty \frac{X_n}{n^{1/p}}, & \quad 0 < p < 1. \end{aligned} \quad (4)$$

For $p \neq 1$, (3) obtains (with $c = E X_1$ or 0) by Kronecker's lemma. When $p = 1$, since $E X_n I_{\{|X_n| \leq n\}} = E X_1 I_{\{|X_1| \leq n\}} \rightarrow E X_1$ as $n \rightarrow \infty$, Kronecker's lemma, in conjunction with (4), yields

$$\lim_n \frac{S_n}{n} - E X_1 = \lim_n \frac{1}{n} \left(S_n - \sum_{j=1}^n E X_j I_{\{|X_j| \leq j\}} \right) = 0, \quad \text{a.c.} \quad \square$$

Corollary 2 (Kolmogorov). If $\{X_n\}$ are i.i.d. r.v.s, $S_n = \sum_1^n X_i$ and $E X_1$ exists, then $S_n/n \xrightarrow{\text{a.c.}} E X_1$.

PROOF. If $|\mathbb{E} X_1| < \infty$, the conclusion is contained in Theorem 2. If $\mathbb{E} X_1 = -\infty$, define $Y_n = \max(X_n, -K)$, where $K > 0$. Then $\{Y_n\}$ are i.i.d. with $\mathbb{E}|Y_1| < \infty$. By Theorem 2

$$\overline{\lim}_{n \rightarrow \infty} \frac{S_n}{n} \leq \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n Y_i \stackrel{\text{a.c.}}{\longrightarrow} \mathbb{E} Y_1 \stackrel{K \rightarrow \infty}{\longrightarrow} \mathbb{E} X_1.$$

Hence, $S_n/n \xrightarrow{\text{a.c.}} -\infty$. Analogously, $S_n/n \xrightarrow{\text{a.c.}} \infty$ when $\mathbb{E} X_1 = \infty$. \square

EXAMPLE 1. If $\{X_n, n \geq 1\}$ are i.i.d. \mathcal{L}_p random variables for some p in $(0, 2)$, then (i) $\sum_{n=1}^{\infty} n^{-2/p} X_n^2 < \infty$, a.c. Moreover, if $\mathbb{E} X = 0$ whenever $1 \leq p < 2$, (ii) $n^{-2/p} \sum_{j=2}^n X_j S_{j-1} \xrightarrow{\text{a.c.}} 0$.

PROOF. According to the proof of Theorem 5.1.3, $Y_n = n^{-1/p} X_n I_{\{|X_n| \leq n^{1/p}\}}$, $n \geq 1$ and $n^{-1/p} X_n$, $n \geq 1$ are equivalent sequences and $\sum_{n=1}^{\infty} \mathbb{E} Y_n^2 < \infty$ implying (i). Then, by Kronecker's lemma, $n^{-2/p} \sum_{j=1}^n X_j^2 \xrightarrow{\text{a.c.}} 0$, and so Theorem 2 ensures that

$$n^{-2/p} \sum_{j=2}^n X_j S_{j-1} = \frac{1}{2} n^{-2/p} \left[S_n^2 - \sum_{j=1}^n X_j^2 \right] \xrightarrow{\text{a.c.}} 0. \quad \square$$

If $\{X_n, n \geq 1\}$ are i.i.d. r.v.s, $S_n = \sum_1^n X_i$ and $\{b_n, n \geq 1\}$ are constants with $0 < b_n \uparrow \infty$, the relationship between $X_n/b_n \xrightarrow{\text{a.c.}} 0$ and $(S_n - C_n)/b_n \xrightarrow{\text{a.c.}} 0$ for some sequence of constants $\{C_n, n \geq 1\}$ is described by

Theorem 3. Let $\{X, X_n, n \geq 1\}$ be i.i.d. r.v.s with partial sums $\{S_n, n \geq 1\}$ and $\{b_n, n \geq 1\}$ constants such that $0 < b_n \uparrow \infty$ and

$$\sum_{n=1}^{\infty} \mathbb{P}\{|X_n| > b_n\} < \infty. \quad (5)$$

(i) (Feller) If (α) $b_n/n \uparrow \infty$ or (β) $b_n/n \downarrow$, the first half of (6) holds and $\mathbb{E} X = 0$ or (γ) $\mathbb{E} X = 0$ and

$$b_n^2 \sum_{j=n}^{\infty} b_j^{-2} = O(n), \quad b_n \sum_{j=1}^n b_j^{-1} = O(n), \quad (6)$$

then

$$\frac{S_n}{b_n} \xrightarrow{\text{a.c.}} 0. \quad (7)$$

(ii) If $\{a_n, n \geq 1\}$ are positive constants with $A_n = \sum_1^n a_j \rightarrow \infty$ such that $b_n = A_n/a_n \uparrow \infty$ and satisfies the first half of (6), then

$$\frac{1}{A_n} \sum_{j=1}^n a_j (X_j - \mathbb{E} X I_{\{|X| \leq A_j/a_j\}}) \xrightarrow{\text{a.e.}} 0 \quad (8)$$

and, moreover, if $\mathbb{E} X = 0$,

$$\frac{1}{A_n} \sum_{j=1}^n a_j X_j \xrightarrow{\text{a.e.}} 0. \quad (9)$$

PROOF. According to Example 5.1.2, condition (5), the initial portion of (6) and Kronecker's lemma ensure

$$\frac{1}{b_n} \sum_{j=1}^n (X_j - \mathbb{E} X I_{\{|X| \leq b_j\}}) \xrightarrow{\text{a.e.}} 0. \quad (10)$$

The conclusion (8) in case (ii) follows likewise.

Now $b_n/n \uparrow \infty$ implies the first half of (6), hence (10) and also

$$\begin{aligned} \left| \frac{1}{b_n} \sum_{j=1}^n \mathbb{E} X I_{\{|X| \leq b_j\}} \right| &\leq \frac{1}{b_n} \sum_{j=1}^n \mathbb{E} |X| (I_{\{|X| \leq b_N\}} + I_{\{b_N < |X| \leq b_j\}}) \\ &\leq \frac{nb_N}{b_n} + \frac{n}{b_n} \int_{\{b_N < |X| \leq b_n\}} |X| \\ &\leq \frac{nb_N}{b_n} + \sum_{j=N}^n \frac{j}{b_j} \int_{\{b_{j-1} < |X| \leq b_j\}} |X| \\ &\leq \frac{nb_N}{b_n} + \sum_{j=N}^{\infty} j \mathbb{P}\{b_{j-1} < |X| \leq b_j\} \\ &\xrightarrow{n \rightarrow \infty} \sum_{j=N}^{\infty} j \mathbb{P}\{b_{j-1} < |X| \leq b_j\} \xrightarrow{N \rightarrow \infty} 0 \end{aligned}$$

by Example 5.1.2 (18), and so (7) follows from (10), thereby proving (α) of (i). In case (β) since $\mathbb{E} X = 0$,

$$\begin{aligned} \left| \frac{1}{b_n} \sum_{j=1}^n \mathbb{E} X I_{\{|X| \leq b_j\}} \right| &\leq \frac{1}{b_n} \sum_{j=1}^{N-1} \mathbb{E} |X| I_{\{|X| \leq b_j\}} + \frac{1}{b_n} \left| \sum_{j=N}^n \mathbb{E} X I_{\{|X| > b_j\}} \right| \\ &\leq O\left(\frac{N}{b_n}\right) + \frac{1}{b_n} \sum_{j=N}^n \int_{\{b_j < |X| \leq b_n\}} |X| + \frac{n}{b_n} \int_{\{|X| > b_n\}} |X| \\ &\leq O\left(\frac{N}{b_n}\right) + \sum_{j=N}^{\infty} \mathbb{P}\{|X| > b_j\} + \frac{n}{b_n} \sum_{j=n+1}^{\infty} b_j \mathbb{P}\{b_{j-1} < |X| \leq b_j\} \end{aligned}$$

$$\begin{aligned}
&= O\left(\frac{N}{b_n}\right) + \sum_{j=N}^{\infty} P\{|X| > b_j\} + \sum_{n+1}^{\infty} j P\{b_{j-1} < |X| \leq b_j\} \\
&= o(1)
\end{aligned}$$

as n , and then $N \rightarrow \infty$ by Example 5.1.2 (18).

In case (γ), (7) is a direct consequence of Kronecker's lemma applied to (17) of Example 5.1.2.

To prove the remaining portion of (ii) it suffices, via (8), to note that $\sum_{j=1}^n a_j E X I_{\{|X| \leq A_j/a_j\}} = o(A_n)$ in view of $E X = 0$ and

$$\left| \frac{1}{A_n} \sum_{j=1}^n a_j E X I_{\{|X| > A_j/a_j\}} \right| \leq \frac{1}{A_n} \sum_{j=1}^n a_j E |X| I_{\{|X| > A_j/a_j\}} = o(1). \quad \square$$

Corollary 3. Let $\{a_n, n \geq 1\}$ be positive constants with $A_n = \sum_1^n a_j \rightarrow \infty$ and

$$\frac{a_n}{A_n} \downarrow 0, \quad \left(\frac{A_n}{a_n} \right)^2 \sum_{j=n}^{\infty} \left(\frac{a_j}{A_j} \right)^2 = O(n). \quad (11)$$

For any i.i.d. r.v.s $\{X_n, n \geq 1\}$, there exist constants C_n for which

$$\frac{1}{A_n} \sum_{j=1}^n a_j (X_j - C_j) \xrightarrow{a.e.} 0 \quad (12)$$

iff

$$\frac{a_n X_n}{A_n} \xrightarrow{a.e.} 0. \quad (13)$$

PROOF. If (13) obtains, so does (12) with $C_j = E X I_{\{|X| \leq A_j/a_j\}}$ by Theorem 3. Conversely, since $a_n = o(A_n)$ guarantees $A_n/A_{n-1} \rightarrow 1$, (12) ensures

$$a_n (X_n - C_n)/A_n \xrightarrow{a.e.} 0.$$

Moreover, $a_n X_n/A_n \xrightarrow{P} 0$ via $a_n = o(A_n)$. Then $a_n C_n/A_n \rightarrow 0$, which, in turn, ensures (13). \square

Although necessary and sufficient conditions for the WLLN are available in the case of independent r.v.s, a complete discussion (see Section 10.1) requires the notion of symmetrization. The i.i.d. situation, however, is amenable to methods akin to those already employed.

Theorem 4 (Feller). If $\{X_n\}$ are i.i.d. r.v.s and $S_n = \sum_{i=1}^n X_i$, then

$$\frac{S_n - C_n}{n} \xrightarrow{P} 0 \quad (14)$$

for some choice of real numbers C_n iff

$$n \mathbb{P}\{|X_1| > n\} \rightarrow 0, \quad (15)$$

and if so, $C_n/n = \mathbb{E} X_1 I_{\{|X_1| \leq n\}} + o(1)$.

PROOF. Sufficiency: Set $X'_j = X_j I_{\{|X_j| \leq n\}}$ for $1 \leq j \leq n$ and $S'_n = \sum_{j=1}^n X'_j$. Then, for each $n \geq 2$, $\{X'_j, 1 \leq j \leq n\}$ are i.i.d. and for $\varepsilon > 0$, $\mathbb{P}\{|(S_n/n) - (S'_n/n)| > \varepsilon\} \leq \mathbb{P}\{S_n \neq S'_n\} = \mathbb{P}\{\bigcup_{j=1}^n [X_j \neq X'_j]\} \leq n \mathbb{P}\{|X_1| > n\}$, so that (15) entails $(S'_n/n) - (S_n/n) \xrightarrow{P} 0$. Thus, to prove (14) it suffices to verify that

$$\frac{S'_n - \mathbb{E} S'_n}{n} = \frac{S'_n}{n} - \mathbb{E} X_1 I_{\{|X_1| \leq n\}} \xrightarrow{P} 0. \quad (16)$$

By Lemma 5.1.1 (4), Corollary 4.3.5, and (15),

$$\begin{aligned} \mathbb{E}(S'_n - \mathbb{E} S'_n)^2 &= \sum_1^n \sigma^2(X'_j) \leq \sum_1^n \mathbb{E}(X'_j)^2 = n \mathbb{E}(X'_1)^2 \\ &= n \sum_{j=1}^n \int_{\{j-1 < |X_1| \leq j\}} X_1^2 \\ &\leq n \sum_{j=1}^n j^2 [\mathbb{P}\{|X_1| > j-1\} - \mathbb{P}\{|X_1| > j\}] \\ &= n \left[\mathbb{P}\{|X_1| > 0\} - n^2 \mathbb{P}\{|X_1| > n\} \right. \\ &\quad \left. + \sum_{j=1}^{n-1} ((j+1)^2 - j^2) \mathbb{P}\{|X_1| > j\} \right] \\ &\leq 3n \left[1 + \sum_{j=1}^{n-1} j \mathbb{P}\{|X_1| > j\} \right] = o(n^2), \end{aligned} \quad (17)$$

which implies (16) and hence (14) with $C_n = n \mathbb{E} X_1 I_{\{|X_1| \leq n\}}$.

Conversely, if (14) holds, setting $c_n = C_n - C_{n-1}$, $n \geq 1$, $C_0 = 0$,

$$\frac{X_n - c_n}{n} = \frac{S_n - C_n}{n} - \frac{n-1}{n} \left(\frac{S_{n-1} - C_{n-1}}{n-1} \right) \xrightarrow{P} 0,$$

whence $(X_n - c_n)/n \xrightarrow{P} 0$, necessitating $c_n = o(n)$. By Lévy's inequality (Lemma 3.3.5), for any $\varepsilon > 0$

$$\begin{aligned} &\mathbb{P} \left\{ \max_{1 \leq j \leq n} |S_j - C_j - m(S_j - C_j - S_n + C_n)| \geq \frac{n\varepsilon}{2} \right\} \\ &\leq 2 \mathbb{P} \left\{ |S_n - C_n| \geq \frac{n\varepsilon}{2} \right\} = o(1); \end{aligned} \quad (18)$$

but, taking $X_j = S_j - C_j$ in Exercise 3.3.7,

$$\max_{1 \leq j \leq n} |m(S_j - C_j) - S_n + C_n| = o(n). \quad (19)$$

Thus, from (18) and (19), for all $\varepsilon > 0$

$$\lim_n P\left\{ \max_{1 \leq j \leq n} |S_j - C_j| < n\varepsilon \right\} = 1. \quad (20)$$

Moreover, for $\max_{1 \leq j \leq n} |c_j| < nc$, and hence for all large n ,

$$\begin{aligned} P\left\{ \max_{1 \leq j \leq n} |S_j - C_j| < n\varepsilon \right\} &\leq P\left\{ \max_{1 \leq j \leq n} |X_j - c_j| < 2n\varepsilon \right\} \\ &\leq P\left\{ \max_{1 \leq j \leq n} |X_j| < 3n\varepsilon \right\}, \end{aligned}$$

which, in conjunction with (20), yields

$$P^n\{|X_1| < 3n\varepsilon\} = P\left\{ \max_{1 \leq j \leq n} |X_j| < 3n\varepsilon \right\} \rightarrow 1$$

or, equivalently, for all $\varepsilon > 0$

$$n \log[1 - P\{|X_1| \geq 3n\varepsilon\}] \rightarrow 0 \quad (21)$$

as $n \rightarrow \infty$. Since $\log(1 - x) = -x + o(x)$ as $x \rightarrow 0$, (21) entails (15).

The final characterization of C_n/n results from the fact that (14) entails (15), which, in turn, implies (14) with $C_n/n = E X_1 I_{\{|X_1| \leq n\}}$. \square

EXAMPLE 2 (Chung–Ornstein). If $\{X, X_n, n \geq 1\}$ are i.i.d. with $n P\{|X| > n\} = o(1)$ and $E X I_{\{|X| \leq n\}} = o(1)$ (*a fortiori* if $E X = 0$) then the random walk $\{S_n = \sum_1^n X_i, n \geq 1\}$ is recurrent. If, rather, $E X > 0$ or $E X < 0$, the random walk is nonrecurrent.

PROOF. Take $N = k \cdot m$ in Lemma 4.2.5, where m is an arbitrary integer. Then for any $\varepsilon > 0$

$$\sum_{n=0}^{km} P\{|S_n| < \varepsilon\} \geq \frac{1}{2k} \sum_{n=0}^{km} P\{|S_n| < k\varepsilon\} \geq \frac{1}{2k} \sum_{n=0}^{km} P\left\{ \left| \frac{S_n}{n} \right| < \frac{n\varepsilon}{m} \right\}.$$

According to Theorem 4, $S_n/n \xrightarrow{P} 0$, and so

$$\sum_{n=0}^{\infty} P\{|S_n| < \varepsilon\} \geq \frac{m}{2} \lim_{k \rightarrow \infty} \frac{1}{km} \sum_{n=0}^{km} P\left\{ \left| \frac{S_n}{n} \right| < \frac{\varepsilon}{m} \right\} = \frac{m}{2}.$$

Since m is arbitrary, the series on the left diverges for all $\varepsilon > 0$ and the conclusion follows from Example 4.2.1. The final remark stems from the strong law of large numbers. \square

In Kolmogorov's classical SLLN (Corollary 2) involving identically distributed \mathcal{L}_1 random variables $\{X_n, n \geq 1\}$, independence can be weakened to *pairwise independence* (X_i independent of X_j for all $i \neq j$). One way of demonstrating this is via strong laws for non-negative random variables.

Lemma 1 (Etemadi). *If $S_n = \sum_{j=1}^n X_j$, $n \geq 1$, where $\{X_n, n \geq 1\}$ are non-negative \mathcal{L}_2 random variables satisfying*

$$\sup_{n \geq 1} E X_n = B < \infty \quad (22)$$

$$\sum_{n=1}^{\infty} \sigma_{X_n}^2 / n^2 < \infty \quad (23)$$

$$E X_m X_n \leq E X_m \cdot E X_n, \quad n > m \geq 1, \quad (24)$$

then

$$\frac{S_n - E S_n}{n} \xrightarrow{\text{a.s.}} 0. \quad (25)$$

PROOF. Let $k_n = [\alpha^n]$, $\alpha > 1$. Since all pairs of r.v.'s have non-positive covariance, for all $\varepsilon > 0$ and some C in $(0, \infty)$

$$\begin{aligned} \varepsilon^2 \sum_{n=1}^{\infty} P\{|S_{k_n} - E S_{k_n}| > \varepsilon k_n\} &\leq \sum_{n=1}^{\infty} \sigma_{S_{k_n}}^2 / k_n^2 \leq \sum_{n=1}^{\infty} \sum_{j=1}^{k_n} \sigma_{X_j}^2 / k_n^2 \\ &= \sum_{j=1}^{\infty} \sum_{n=(\log j)/\log \alpha}^{\infty} k_n^{-2} \sigma_{X_j}^2 \leq C \sum_{j=1}^{\infty} \sigma_{X_j}^2 / j^2 < \infty, \end{aligned}$$

whence the Borel–Cantelli lemma ensures

$$(S_{k_n} - E S_{k_n}) / k_n \xrightarrow{\text{a.s.}} 0.$$

Via monotonicity of S_k , for any $k \in [k_n, k_{n+1})$,

$$\frac{|S_k - E S_k|}{k} \leq \frac{k_{n+1}}{k_n} \left| \frac{S_{k_{n+1}} - E S_{k_{n+1}}}{k_{n+1}} \right| + \left| \frac{S_{k_n} - E S_{k_n}}{k_n} \right| + \frac{E(S_{k_{n+1}} - S_{k_n})}{k_n},$$

and so, since $k_n^{-1} E(S_{k_{n+1}} - S_{k_n}) \leq k_n^{-1} (k_{n+1} - k_n) B \leq (\alpha - 1) B$,

$$\overline{\lim}_{k \rightarrow \infty} \left| \frac{S_k - E S_k}{k} \right| \leq (\alpha - 1) B \xrightarrow{\text{a.s.}} 0,$$

yielding (25). \square

Corollary 4. *If $S_n = \sum_{j=1}^n X_j$, $n \geq 1$, where $\{X_n, n \geq 1\}$ are non-negative, pairwise independent random variables such that*

$$\sup_{n \geq 1} E X_n < \infty \quad (26)$$

$$\sum_{n=1}^{\infty} n^{-2} E X_n^2 I_{[X_n \leq n]} < \infty \quad (27)$$

$$\sum_{n=1}^{\infty} \mathbf{P}\{X_n > n\} < \infty, \quad (28)$$

then

$$\frac{S_n - \mathbf{E} S_n}{n} \xrightarrow{\text{a.s.}} 0. \quad (29)$$

PROOF. The non-negative, pairwise independent, \mathcal{L}_2 r.v.'s $Y_n = X_n I_{\{X_n \leq n\}}$, $n \geq 1$ clearly satisfy (24) and also (22) and (23) in view of (26) and (27). Thus, by Lemma 1,

$$\frac{1}{n} \left(\sum_{i=1}^n Y_i - \sum_{i=1}^n \mathbf{E} Y_i \right) \xrightarrow{\text{a.s.}} 0. \quad (30)$$

Now, via (26) and (27),

$$0 \leq \frac{1}{n} \mathbf{E} X_n I_{\{X_n > n\}} = \frac{1}{n} (\mathbf{E} X_n - \mathbf{E} X_n I_{\{X_n \leq n\}}) \leq \frac{B}{n} - \frac{1}{n^2} \mathbf{E} X_n^2 I_{\{X_n \leq n\}} = o(1)$$

implying as $n \rightarrow \infty$ that

$$\frac{1}{n} \left(\mathbf{E} S_n - \sum_{i=1}^n \mathbf{E} Y_i \right) = \frac{1}{n} \sum_{i=1}^n \mathbf{E} X_i I_{\{X_i > i\}} \rightarrow 0. \quad (31)$$

Since (28) ensures that $\{X_n, n \geq 1\}$ and $\{Y_n, n \geq 1\}$ are equivalent sequences, (30) holds with $\sum_{i=1}^n Y_i$ replaced by S_n , and this together with (31) yields (29). \square

Theorem 5 (Etemadi). Let $S_n = \sum_{i=1}^n X_i$, $n \geq 1$, where $\{X_n, n \geq 1\}$ are pairwise independent, identically distributed \mathcal{L}_1 random variables. Then $S_n/n \xrightarrow{\text{a.s.}} \mathbf{E} X_1$.

PROOF. Evidently, $\{X_n^+, n \geq 1\}$ (likewise $\{X_n^-, n \geq 1\}$) satisfies (26), (28), and also (27) since

$$\begin{aligned} \sum_{n=1}^{\infty} n^{-2} \mathbf{E}(X_n^+)^2 I_{\{X_n^+ \leq n\}} &= \sum_{n=1}^{\infty} n^{-2} \sum_{j=1}^n \mathbf{E}(X^+)^2 I_{[j-1 < X^+ \leq j]} \\ &= \sum_{j=1}^{\infty} \sum_{n=j}^{\infty} n^{-2} \mathbf{E}(X^+)^2 I_{[j-1 < X^+ \leq j]} \\ &\leq C \sum_{j=1}^{\infty} j^{-1} \mathbf{E}(X^+)^2 I_{[j-1 < X^+ \leq j]} \\ &\leq C \mathbf{E} \sum_{j=1}^{\infty} X^+ I_{[j-1 < X^+ \leq j]} \\ &= C \mathbf{E} X^+ < \infty, \end{aligned}$$

whence Corollary 4 guarantees

$$S_n/n = \frac{1}{n} \sum_{j=1}^n (X_j^+ - X_j^-) \xrightarrow{\text{a.s.}} \mathbf{E} X_1^+ - \mathbf{E} X_1^- = \mathbf{E} X_1. \quad \square$$

An extremely useful tool in probability theory is the Kolmogorov inequality, which, as will be seen in Section 7.4, holds in a much broader context.

Theorem 6 (Kolmogorov inequality). *If $\{X_j, 1 \leq j \leq n\}$ are independent \mathcal{L}_2 r.v.s with $E X_j = 0, S_j = \sum_1^n X_i, 1 \leq j \leq n$, then for $\varepsilon > 0$*

$$P \left\{ \max_{1 \leq j \leq n} |S_j| \geq \varepsilon \right\} \leq \frac{1}{\varepsilon^2} \sum_{j=1}^n E X_j^2. \quad (32)$$

PROOF. Define $T = \text{smallest integer } j \text{ in } [1, n] \text{ for which } |S_j| \geq \varepsilon$ if such exists and $T = n + 1$ otherwise. Then $\{T \geq j\} \in \sigma(X_1, \dots, X_{j-1})$ for $1 \leq j \leq n + 1$, where the σ -algebra in question is $\{\emptyset, \Omega\}$ when $j = 1$. Thus, X_j and $I_{[T \geq j]}$ are independent r.v.s for $1 \leq j \leq n$. Moreover, since $S_T = S_j$ on $\{T = j\}, 1 \leq j \leq n$, and $S_{\min[T, n]} = \sum_{j=1}^n X_j I_{[T \geq j]}$,

$$\begin{aligned} P \left\{ \max_{1 \leq j \leq n} |S_j| \geq \varepsilon \right\} &= P\{T \leq n\} = \sum_{j=1}^n P\{T = j\} \\ &\leq \frac{1}{\varepsilon^2} \sum_{j=1}^n \int_{[T=j]} S_j^2 = \frac{1}{\varepsilon^2} \int_{[T \leq n]} S_{\min[T, n]}^2 \\ &\leq \frac{1}{\varepsilon^2} E \left(\sum_{j=1}^n X_j I_{[T \geq j]} \right)^2 \\ &= \frac{1}{\varepsilon^2} E \left(\sum_{j=1}^n X_j^2 I_{[T \geq j]} + 2 \sum_{1 \leq i < j \leq n} X_i X_j I_{[T \geq j]} \right) \\ &= \frac{1}{\varepsilon^2} \left(\sum_{j=1}^n P\{T \geq j\} E X_j^2 + 2 \sum_{1 \leq i < j \leq n} E(X_i I_{[T \geq j]}) \cdot E X_j \right) \\ &\leq \frac{1}{\varepsilon^2} \sum_{j=1}^n E X_j^2. \end{aligned} \quad \square$$

The versatility of the Kolmogorov inequality will be exemplified in extending Theorem 4 to the case of random indices.

Theorem 7. *If $\{X_n, n \geq 1\}$ are i.i.d. r.v.s obeying*

$$n P\{|X_1| > n\} = o(1) \quad (33)$$

and $\{T_n, n \geq 1\}$ are positive integer-valued r.v.s satisfying

$$\frac{T_n}{n} \xrightarrow{P} c, \quad \text{where } 0 < c < \infty, \quad (34)$$

then, setting $S_n = \sum_1^n X_i$,

$$(S_{T_n}/T_n) - E X_1 I_{[|X_1| \leq n]} \xrightarrow{P} 0. \quad (35)$$

PROOF. Define $X'_j = X_j I_{\{|X_j| \leq n\}}$, $S'_j = \sum_{i=1}^j X'_i$, $M'_j = E S'_j$, and $c_n = [c \cdot n] = \text{largest integer } \leq c \cdot n$. Then,

$$\begin{aligned} P\{S_{T_n} \neq S'_{T_n}, T_n \leq 2c_n\} &\leq P\left\{\bigcup_{j=1}^{2c_n} [|X_j| > n]\right\} \\ &\leq \sum_{j=1}^{2c_n} P\{|X_j| > n\} \leq 2c_n P\{|X_1| > n\} = o(1) \end{aligned}$$

by (33), whence, taking cognizance of (34),

$$P\{S_{T_n} \neq S'_{T_n}\} \leq P\{S_{T_n} \neq S'_{T_n}, T_n \leq 2c_n\} + P\{T_n > 2c_n\} = o(1),$$

that is,

$$S_{T_n} - S'_{T_n} \xrightarrow{P} 0. \quad (36)$$

Now the prior calculation of (17) shows that $E X'_1{}^2 = o(n)$, whence, defining

$$B_j = \{|S'_j - M'_j| > n\varepsilon\}, \quad D_n = \bigcup_{j=1}^{2c_n} B_j,$$

by Kolmogorov's inequality (Theorem 6)

$$P\{D_n\} \leq \frac{1}{n^2\varepsilon^2} \sum_{j=1}^{2c_n} E X'_j{}^2 \leq \frac{2c}{n\varepsilon^2} E X'_1{}^2 = o(1).$$

Thus,

$$\begin{aligned} P\{B_{T_n}\} &\leq P\{T_n \leq 2c_n, B_{T_n}\} + P\{T_n > 2c_n\} \\ &\leq P\{D_n\} + P\{T_n > 2c_n\} = o(1), \end{aligned}$$

so that

$$\frac{S_{T_n} - M'_{T_n}}{n} \xrightarrow{P} 0.$$

Hence, by (36) and (34)

$$\frac{S_{T_n} - M'_{T_n}}{T_n} = \frac{n}{T_n} \left(\frac{S_{T_n} - M'_{T_n}}{n} \right) \xrightarrow{P} 0.$$

But $M'_{T_n} = \sum_{j=1}^{T_n} E X'_j = T_n E X'_1 = T_n E X_1 I_{\{|X_1| \leq n\}}$, whence

$$\frac{S_{T_n}}{T_n} - E X_1 I_{\{|X_1| \leq n\}} \xrightarrow{P} 0. \quad \square$$

For any sequence of r.v.s $\{X_n, n \geq 1\}$ and nonnegative numbers u, v define

$$S_{u,v} = \sum_{1 \leq j \leq v} X_{[u]+j}, \quad S_v = S_{0,v}, \quad S_0 = 0. \quad (37)$$

The r.v.s $S_{u,v}$ are called **delayed sums** (of the sequence X_n). If $\sum_{n=1}^{\infty} P\{|S_n| > n\varepsilon\} < \infty$ for every $\varepsilon > 0$, then by the Borel–Cantelli lemma, $S_n/n \xrightarrow{a.c.} 0$. The converse is false even when the $\{X_n\}$ are i.i.d. since then by the classical SLLN (Theorem 5.2.2, $p = 1$), $S_n/n \xrightarrow{a.c.} 0$ iff $E X_1 = 0$, whereas according to a theorem of Hsu–Robbins–Erdos (Chapter 10) $\sum_{n=1}^{\infty} P\{|S_n| > n\varepsilon\} < \infty$, all $\varepsilon > 0$ iff $E X_1 = 0$, $E X_1^2 < \infty$. However, in the i.i.d. case $S_n/n \xrightarrow{a.c.} 0$ does imply that $\sum_{n=1}^{\infty} (1/n)P\{|S_n| > n\varepsilon\} < \infty$ for every $\varepsilon > 0$ by a result of Spitzer, and the next theorem due to Baum–Katz extends this.

For any sequence of r.v.s $\{X_n, n \geq 1\}$ define via (37).

$$S_{u,v}^* = \max_{1 \leq j \leq v} |S_{u,j}|, \quad S_v^* = \max_{1 \leq j \leq v} |S_j|, \quad u \geq 0, v \geq 0. \quad (38)$$

Theorem 8 (Baum–Katz–Spitzer). *Let $\{X_n, n \geq 1\}$ be i.i.d. r.v.s with $E|X_1|^p < \infty$ for some p in $(0, 2)$ and $E X_1 = 0$ if $1 \leq p < 2$. If $\alpha p \geq 1$ (hence $\alpha > \frac{1}{2}$), then for every $\varepsilon > 0$*

$$\sum_{n=1}^{\infty} n^{\alpha p - 2} P\{S_n^* > n^{\alpha} \varepsilon\} < \infty. \quad (39)$$

PROOF. By Theorem 5.2.2, $S_n/n^{1/p} \xrightarrow{a.c.} 0$, implying $S_n^*/n^{1/p} \xrightarrow{a.c.} 0$ and hence

$$\frac{S_{n,n}^*}{n^{1/p}} \leq \frac{S_n^* + S_{2n}^*}{n^{1/p}} \xrightarrow{a.c.} 0. \quad (40)$$

Suppose first that $\alpha p = 1$. Since $\{S_{2^n, 2^n}^*, n \geq 1\}$ are independent r.v.s, (40) and the Borel–Cantelli theorem (Theorem 3.2.1) imply that for every $\varepsilon > 0$ and $c = (\log 2)^{-1}$

$$\begin{aligned} \infty &> \sum_{n=1}^{\infty} P\{S_{2^n, 2^n}^* \geq 2^{\alpha n} \varepsilon\} = \sum_{n=1}^{\infty} P\{S_{2^n}^* \geq 2^{\alpha n} \varepsilon\} \\ &\geq \int_0^{\infty} P\{S_{2^t}^* \geq 2^{\alpha(t+1)} \varepsilon\} dt > c \int_1^{\infty} x^{-1} P\{S_x^* \geq 2^{\alpha} \varepsilon x^{\alpha}\} dx, \end{aligned}$$

and so for all $\varepsilon' > 0$

$$\sum_{n=1}^{\infty} \frac{1}{n} P\{S_n^* \geq n^{\alpha} \varepsilon'\} < \infty.$$

Suppose next that $\alpha p > 1$. Since for $m \geq 1$

$$(m+1)^{\alpha p/(\alpha p-1)} \geq m^{\alpha p/(\alpha p-1)} + \frac{\alpha p}{\alpha p-1} m^{1/(\alpha p-1)} \geq m^{\alpha p/(\alpha p-1)} + m^{1/(\alpha p-1)},$$

the r.v.s

$$\{S_{m^{\alpha p/(\alpha p-1)}, m^{1/(\alpha p-1)}}^*, m \geq 1\}$$

are independent. Moreover, (40) implies

$$m^{-\alpha/(\alpha p-1)} S_{m^{\alpha p/(\alpha p-1)}, m^{1/(\alpha p-1)}}^* \leq m^{-\alpha/(\alpha p-1)} S_{m^{\alpha p/(\alpha p-1)}, m^{\alpha p/(\alpha p-1)}}^* \xrightarrow{a.c.} 0$$

as $m \rightarrow \infty$, whence for all $\varepsilon > 0$

$$\begin{aligned} \infty &> \sum_{m=1}^{\infty} P\{S_{m^{1/(xp-1)}, m^{1/(xp-1)}}^* \geq m^{x/(xp-1)}\varepsilon\} \\ &= \sum_{m=1}^{\infty} P\{S_{m^{1/(xp-1)}}^* \geq m^{x/(xp-1)}\varepsilon\} \\ &\geq \int_1^{\infty} P\{S_t^* \geq t^{x/(xp-1)}\varepsilon\} dt \\ &\geq (xp-1) \int_1^{\infty} x^{xp-2} P\{S_x^* \geq \varepsilon' x^x\} dx \end{aligned}$$

for $\varepsilon' = 2^{x/(xp-1)}\varepsilon$ and (39) follows. \square

The converse to Theorem 8 appears in Section 6.4.

Corollary 5. If $\{X_n, n \geq 1\}$ are i.i.d. r.v.s with $E X_1 = 0$, $E|X_1|^p < \infty$ for some p in $[1, 2)$, then

$$\sum_{n=1}^{\infty} n^{p-2} P\{|S_n| > n\varepsilon\} < \infty, \quad \varepsilon > 0. \quad (41)$$

The convergence of series such as (39), (41) is enhanced when an explicit bound $C(\varepsilon, p)$ is obtained for the sum since uniform integrability of stopping or last times related to S_n may be deducible therefrom.

EXERCISES 5.2

- Prove for i.i.d. r.v.s $\{X_n\}$ with $S_n = \sum_1^n X_i$ that $(S_n - C_n)/n \xrightarrow{a.s.} 0$ for some sequence of constants C_n iff $E|X_1| < \infty$.
- If $\{X_n\}$ are i.i.d. with $E|X_1|^p = \infty$ for some $p \in (0, \infty)$, then $P\{\overline{\lim}_n |S_n|/n^{1/p} = \infty\} = 1$.
- Demonstrate for i.i.d. r.v.s $\{X_n\}$ that $E \sup_n |X_n/n| < \infty$ iff $E|X_1| \log^+ |X_1| < \infty$.
- If $S_n = \sum_1^n X_i$, where $\{X_n\}$ are i.i.d. \mathcal{L}_p r.v.s for some $p \geq 1$, then $E|S_n/n|^p \rightarrow |E X|^p$. Hint: Recall Example 4.3.1.
- If $\{X_n\}$ are i.i.d. r.v.s with $E X_1 = 1$ and $\{a_n\}$ are bounded real numbers, then $(1/n) \sum_1^n a_j \rightarrow 1$ iff $(1/n) \sum_{j=1}^n a_j X_j \xrightarrow{a.s.} 1$.
- Let $\{X_n\}$ be i.i.d. r.v.s with $(S_n/n) - C_n \xrightarrow{P} 0$ for some sequence of constants $\{C_n\}$. Prove that
 - $\int_{\{x_1 < |X_1| < \beta n\}} |X_1| = o(1)$ whenever $0 < \alpha < \beta < \infty$,
 - $2 \int_{\{|X_1| \leq 2n\}} X_1 - \int_{\{|X_1 + X_2| \leq 2n\}} (X_1 + X_2) = o(1)$.
- Let $\{X, X_n, n \geq 1\}$ be i.i.d. \mathcal{L}_p r.v.s for some p in $(0, 2)$ where $E X = 0$ whenever $1 \leq p < 2$. If $\{a_n, n \geq 1\}$ are positive constants satisfying $a_n^p / \sum_{j=1}^n a_j^p = O(1/n)$, $\sum_{j=1}^n a_j^p \rightarrow \infty$, then

$$\frac{1}{\left(\sum_{j=1}^n a_j^p\right)^{1/p}} \sum_{j=1}^n a_j X_j \xrightarrow{\text{a.e.}} 0. \quad (*)$$

Conversely, if $a_n^p / \sum_{j=1}^n a_j^p \sim Cn$, then (*) ensures $X \in \mathcal{L}_p$ (and also $E X = 0$ if $p = 1$).
Hint: if $Y_n = a_n (\sum_{j=1}^n a_j^p)^{-1/p} X_j \cdot I_{\{|X_j| \leq n^{1/p}\}}$, then $E|Y_n|^p \leq Cn^{-p} E|X|^p I_{\{|X| \leq n^{1/p}\}}$ whence $\sum_{n=1}^{\infty} E|Y_n|^p < \infty$ for $\alpha > p$ and $\sum_{n=1}^{\infty} |E Y_n| < \infty$ for $1 < p < 2$ as in Theorem 5.1.3. N.B. Exercise 7 with $p = 1$ and Theorem 5.2.3(ii) are related to summability.

8. For $r > 0$ and any r.v. X , prove that $X \in \mathcal{L}_r$ iff

$$\sum_{n=1}^{\infty} n^{r-1} (\log n)^r P\{|X| \geq n \log n\} < \infty$$

Hint: Employ the techniques of Theorem 7.

9. (Klass–Teicher) If $\{X, X_n, n \geq 1\}$ are i.i.d. r.v.s and $\{b_n, n \geq 1\}$ are constants with (i) $b_n/n \uparrow$ or (ii) $b_n/n \downarrow, b_n/n^{1/2} \rightarrow \infty$, $\sum_n (b_j/j)^2 = O(b_n^2/n)$ and $b_n \uparrow$, then

$$\frac{1}{b_n} \sum_{i=1}^n X_i - \frac{n}{b_n} E X I_{\{|X| \leq b_n\}} \xrightarrow{P} 0 \text{ iff } n P\{|X| > b_n\} = o(1).$$

10. Prove that if $\{X_n, n \geq 1\}$ are independent r.v.s with $E X_n = 0$, $E X_n^2 = \sigma_n^2$, $s_n^2 = \sum_i \sigma_i^2 \rightarrow \infty$, then $s_n^{-1} (\log s_n^2)^{-\alpha} \sum_i X_i \xrightarrow{\text{a.e.}} 0$ for $\alpha > \frac{1}{2}$.

11. (Feller–Chung) Let $P(x)$ be nonnegative and nonincreasing on $(0, \infty)$ and suppose that the positive, nondecreasing sequence $\{b_n, n \geq 1\}$ satisfies (*) $\lim_{n \rightarrow \infty} b_n/b_n > c > 1$ for some integer r or *a fortiori* either (**) $b_n/n^\beta \uparrow$ for some $\beta > 0$ or (***) $b_n^2 \sum_{j=n}^{\infty} b_j^{-2} = O(n)$. Then $\sum_{n=1}^{\infty} P(xb_n)$ either diverges for all $x > 0$ or converges for all $x > 0$. (*Hint:* Any integer m satisfies $r^k \leq m < r^{k+1}$, where $k = k(m) \rightarrow \infty$ as $m \rightarrow \infty$, whence $b_m \geq c^k b_n$, all large n). Thus, if $\{X, X_n\}$ are i.i.d. and $\sum_{n=1}^{\infty} P\{|X| > Cb_n\} = \infty$ for some $C > 0$ and $\{b_n\}$ as stipulated, $\overline{\lim}_{n \rightarrow \infty} |X_n|/b_n = \infty = \overline{\lim}_{n \rightarrow \infty} |S_n|/b_n$.

12. (Heyde–Rogozin) If $\{X, X_n, n \geq 1\}$ are i.i.d. with (*) $\lim_{x \rightarrow \infty} x^2 P\{|X| > x\}/E X^2 I_{\{|X| \leq x\}} > \alpha > 0$, then for every sequence $\{b_n\}$ satisfying $0 < b_n \uparrow \infty$ either $\overline{\lim}_{n \rightarrow \infty} |S_n|/b_n = \infty$ or $(1/b_n)(S_n - \sum_i E X I_{\{|X| \leq b_n\}}) \xrightarrow{\text{a.e.}} 0$. In particular, for symmetric i.i.d. r.v.s satisfying (*), $S_n/b_n \rightarrow 0$ or $\overline{\lim} S_n/b_n = \infty$. *Hint:* $P\{|X| > x\} \leq (1 + \delta^2 x^{-1}) P\{|X| > \delta x\}$, $\delta > 1$, $x > x_0$, whence the series of Exercise 11 either converges for all $C > 0$ or diverges for all $C > 0$.

13. (Komlós–Révész) Let $\{X_n, n \geq 1\}$ be independent r.v.s with means $E X_n$ and positive variances σ_n^2 satisfying $\lim E X_n = c$ and $\sum_{n=1}^{\infty} \sigma_n^{-2} = \infty$. Then

$$\frac{\sum_{j=1}^n X_j/\sigma_j^2}{\sum_{j=1}^n \sigma_j^{-2}} \xrightarrow{\text{a.e.}} c.$$

Hint: $\sum_{j=1}^n [(X_j - E X_j)/\sigma_j^2 \sum_{i=1}^n \sigma_i^{-2}]$ converges a.e.

14. (Heyde) Let $\{X_n, n \geq 1\}$ be i.i.d. \mathcal{L}_p r.v.s, $S_n = \sum_i X_i$, $\bar{S}_n = \max_{1 \leq j \leq n} S_j$. (i) If $1 \leq p < 2$, then $n^{-1/p} (\bar{S}_n - n\mu) \xrightarrow{\text{a.e.}} 0$ or $n^{-1/p} \bar{S}_n \xrightarrow{\text{a.e.}} 0$ according as $\mu = E X_1 \geq 0$ or $\mu < 0$. (ii) If $0 < p < 1$, then $n^{-1/p} \bar{S}_n \xrightarrow{\text{a.e.}} 0$.

15. (Derman-Robbins) $\{X_n\}$ i.i.d. with $E X_1$ nonexistent does not preclude $P\{S_n/n \rightarrow \infty\} = 1$. Hint: Let $\lim_{x \rightarrow \infty} x^\beta P\{X > x\} > 0$ and $E(X^-)^\beta < \infty$ for $1 > \beta > x > 0$. The latter implies $n^{-1/\beta} \sum_1^n X_i^- \xrightarrow{a.s.} 0$, while the former entails $P\{\sum_1^n X_i^+ \leq Cn^{1/\beta}\} \leq P\{\max_{1 \leq i \leq n} X_i^+ \leq Cn^{1/\beta}\} \leq \exp(-Cn^\gamma)$, $\gamma = 1 - x\beta^{-1} > 0$, $C > 0$.

16. (Klass) Let $\{X_n\}$ be i.i.d. with $E X_1 = 0$ and $\sum_{n=1}^{\infty} P\{|X_n| > b_n\} < \infty$, where

$$b_n > 0, n^{-2}b_n^2 \downarrow, n^{-1}b_n^2 \uparrow \infty.$$

Then $E|S_n| = o(b_n)$. Hint: $X_j = Y_{n,j} + Z_{n,j}$, where $Y_{n,j} = X_j I_{\{|X_j| \leq b_n\}}$. In particular, X_n i.i.d. with $E X_1 = 0$, $E|X_1|^p < \infty$, $p \in [1, 2]$, implies $E|S_n| = o(n^{1/p})$.

17. **Strong Laws for Arrays.** Let $\{X_{ni}, 1 \leq i \leq n\}$ be an array of r.v.'s that are identically distributed and *rowwise independent*, i.e., $\{X_{n1}, \dots, X_{nn}\}$ are independent for $n \geq 2$. If $E|X_{11}|^q < \infty$, $0 < q < 2$, and $E X_{11} = 0$ whenever $1 \leq q < 2$, then $n^{-2/q} \sum_{i=1}^n X_{ni} \xrightarrow{a.s.} 0$. (For an extension to $2 \leq q < 4$, see Example 10.4.1.)

5.3 Stopping Times, Copies of Stopping Times, Wald's Equation

Let (Ω, \mathcal{F}, P) be a probability space and $\{\mathcal{F}_n, n \geq 1\}$ an increasing sequence of sub- σ -algebras of \mathcal{F} , that is, $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \dots \subset \mathcal{F}$. A measurable function $T = T(\omega)$ taking values $1, 2, \dots, \infty$ is called a **stopping time** relative to $\{\mathcal{F}_n\}$ or simply an $\{\mathcal{F}_n\}$ -time if $\{T = j\} \in \mathcal{F}_j$, $j = 1, 2, \dots$. Clearly, If $T \equiv m$, then T is an \mathcal{F}_m -time. If T is an \mathcal{F}_n -time, then setting $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_\infty = \sigma(\bigcup_{n=1}^{\infty} \mathcal{F}_n)$,

$$\{T \geq n\} = \Omega - \bigcup_1^{n-1} \{T = j\} \in \mathcal{F}_{n-1}, \quad 1 \leq n \leq \infty. \quad (1)$$

Moreover, since $\{T = \infty\} = \Omega - \{T < \infty\}$, a stopping time is completely determined by the sets $\{T = n\}$, $1 \leq n < \infty$.

A stopping time T is said to be **finite** if $P\{T = \infty\} = 0$ and **defective** if $P\{T = \infty\} > 0$. A finite stopping time is also called a **stopping rule** or **stopping variable**. When $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, $n \geq 1$, for some sequence of r.v.s $\{X_n\}$, an \mathcal{F}_n -time will generally be alluded to as an $\{X_n\}$ -time or a stopping time relative to $\{X_n\}$. Stopping times and rules have already appeared incognito in Theorem 5.2.5, Lemmas 5.1.3, 5.1.4, Theorem 3.4.1, and Lemma 3.3.5.

The notion of a stopping time derives from gambling. Its definition is simply the mathematical formulation of the fact that, since an honest gambler cannot peer into the future, his decision to stop gambling at any time n must be based solely upon the outcomes X_1, \dots, X_n up to that time and not on subsequent outcomes $X_j, j > n$.

Let $\{X_n, n \geq 1\}$ constitute a sequence of r.v.s and $\{\mathcal{F}_n, n \geq 1\}$ an increasing sequence of sub- σ -algebras of \mathcal{F} such that X_n is \mathcal{F}_n -measurable for each $n \geq 1$. Then $\{X_n, \mathcal{F}_n, n \geq 1\}$ will be called a **stochastic sequence**.

Clearly, for any sequence of r.v.s $\{X_n\}$, $\{X_n, \sigma(X_1, \dots, X_n), n \geq 1\}$ is a stochastic sequence.

For any stochastic sequence $\{X_n, \mathcal{F}_n, n \geq 1\}$ and \mathcal{F}_n -time T , define

$$X_T = X_{T(\omega)}(\omega), \quad \text{where } X_\infty(\omega) = \overline{\lim}_{n \rightarrow \infty} X_n(\omega), \quad (2)$$

$$\mathcal{F}_\infty = \sigma\left(\bigcup_1^\infty \mathcal{F}_n\right), \quad \mathcal{F}_T = \{A: A \in \mathcal{F}_\infty, A \cdot \{T = n\} \in \mathcal{F}_n, n \geq 1\}. \quad (3)$$

It is easy to verify that (i) \mathcal{F}_T is a sub- σ -algebra of \mathcal{F}_∞ , (ii) X_T and T are \mathcal{F}_T -measurable, and (iii) \mathcal{F}_T coincides with \mathcal{F}_m when $T \equiv m$.

Since

$$\mathbf{P}\{|X_T| = \infty\} = \mathbf{P}\{T = \infty, |X_\infty| = \infty\},$$

X_T is a r.v. if either T is finite or $\overline{\lim} X_n$ is finite, a.c.

For $\{\mathcal{F}_n\}$ -times T_1 and T_2 with $T_1 \leq T_2$,

$$\mathcal{F}_{T_1} \subset \mathcal{F}_{T_2} \quad (4)$$

since if $A \in \mathcal{F}_{T_1}$, then $A\{T_2 \leq n\} = A\{T_1 \leq n\} \cdot \{T_2 \leq n\} \in \mathcal{F}_n$, $1 \leq n < \infty$. If T is an $\{\mathcal{F}_n\}$ -time, so is $T + m$ for every integer $m \geq 1$, and since $T \leq T + m$, necessarily $\mathcal{F}_T \subset \mathcal{F}_{T+1} \subset \dots \subset \mathcal{F}_\infty$. Hence, $\mathcal{F}_\infty = \sigma\left(\bigcup_1^\infty \mathcal{F}_n\right) \subset \sigma\left(\bigcup_1^\infty \mathcal{F}_{T+n}\right) \subset \mathcal{F}_\infty$, that is, $\sigma\left(\bigcup_1^\infty \mathcal{F}_{T+n}\right) = \mathcal{F}_\infty$. Consequently, the sequence $\mathcal{G}_n = \mathcal{F}_{T+n}$, $n \geq 1$, may be utilized to define a new stopping time. Note, incidentally, that $\mathcal{F}_{T+n} = \mathcal{F}_{n+m}$ when $T \equiv m$.

Suppose that T_1 is an $\{\mathcal{F}_n, n \geq 1\}$ -time and T_2 is a $\{\mathcal{G}_n, n \geq 1\}$ -time, where $\mathcal{G}_n = \mathcal{F}_{T_1+n}$, $n \geq 1$. Then $T = T_1 + T_2$ is an $\{\mathcal{F}_n, n \geq 1\}$ -time and $\mathcal{F}_T = \mathcal{G}_{T_2}$. Since for $1 \leq m < \infty$

$$\{T = m\} = \bigcup_{j=1}^m \{T_1 = j, T_2 = m - j\} \in \mathcal{F}_m,$$

T is an $\{\mathcal{F}_n, n \geq 1\}$ -time. To prove $\mathcal{F}_T = \mathcal{G}_{T_2}$, let $A \in \mathcal{G}_{T_2}$. Then for $m = 1, 2, \dots$ and $j = 1, \dots, m-1$,

$$A\{T_2 = m-j\} \in \mathcal{G}_{m-j} = \mathcal{F}_{T_1+m-j},$$

$$A\{T_1 = j, T_2 = m-j\} = A\{T_2 = m-j\}\{T_1 + m-j = m\} \in \mathcal{F}_m,$$

$$A\{T = m\} = \bigcup_{j=1}^{m-1} A\{T_1 = j, T_2 = m-j\} \in \mathcal{F}_m,$$

which implies that $A \in \mathcal{F}_T$. Conversely, let $A \in \mathcal{F}_T$. Then for $r = 1, 2, \dots$, and $m = 0, 1, \dots$,

$$A\{T = r+m\} \in \mathcal{F}_{r+m},$$

$$A\{T_2 = m\} \cdot \{T_1 + m = r+m\} = A\{T = r+m\}\{T_1 = r\} \in \mathcal{F}_{r+m},$$

$$A\{T_2 = m\} \in \mathcal{F}_{T_1+m} = \mathcal{G}_m,$$

which implies that $A \in \mathcal{G}_{T_2}$.

On the other hand, if T_1, T_2 are $\{\mathcal{F}_n, n \geq 1\}$ -times with $T_1 < T_2$ and

$$\tau = \begin{cases} T_2 - T_1 & \text{if } T_1 < \infty \\ \infty & \text{if } T_1 = \infty, \end{cases} \quad (5)$$

then $T_2 = T_1 + \tau$ and τ is a $\{\mathcal{G}_n, n \geq 1\}$ -time, where $\mathcal{G}_n = \mathcal{F}_{T_1+n}$. It suffices to note that when $1 \leq n < \infty$

$$\{\tau = n\} \{T_1 + n = j + n\} = \{T_1 = j\} \{T_2 = j + n\} \in \mathcal{F}_{j+n}, \quad 1 \leq j < \infty,$$

implies $\{\tau = n\} \in \mathcal{F}_{T_1+n} = \mathcal{G}_n, n \geq 1$.

Lemma 1. *If T is an $\{X_n\}$ -time for some sequence $\{X_n\}$ of r.v.s, there exists a sequence $\{C_n\}$ of disjoint Borel cylinder sets of $(R^\infty, \mathcal{B}^\infty)$ whose corresponding bases B_n are n -dimensional Borel sets, $n \geq 1$, such that*

$$\{\omega: T = n\} = \{\omega: (X_1, \dots, X_n, \dots) \in C_n\}, \quad n = 1, 2, \dots \quad (6)$$

Conversely, given any sequence $\{C_n, n \geq 1\}$ of disjoint Borel cylinder sets with n -dimensional Borel bases, an $\{X_n\}$ -time T is defined by (6) and $\{T = \infty\} = \Omega - \bigcup_1^\infty \{T = n\}$.

PROOF. If T is an $\{X_n\}$ -time, then $\{T = n\} \in \sigma(X_1, \dots, X_n), n \geq 1$, whence by Theorem 1.4.4 there exists an n -dimensional Borel set B_n for which

$$\{T = n\} = \{(X_1, \dots, X_n) \in B_n\}.$$

For each $n \geq 1$, let C'_n be the cylinder set in R^∞ with base B_n . Then

$$\{T = n\} = \{(X_1, \dots, X_n, \dots) \in C'_n\}, \quad n \geq 1.$$

Moreover, $C_n = C'_n - \bigcup_{j=1}^{n-1} C'_j, n \geq 1$, are disjoint Borel cylinder sets with n -dimensional Borel bases. Since $\{T = m\} \cdot \{T = n\} = \emptyset$ for $m \neq n$, (6) follows.

Conversely, given a sequence of disjoint cylinders $C_n \in \mathcal{B}^\infty$ with n -dimensional Borel bases $B_n, n \geq 1$, if T is defined as stipulated, then

$$\{T = n\} = \{(X_1, \dots, X_n) \in B_n\} \in \sigma(X_1, \dots, X_n), \quad n \geq 1,$$

so that T is an $\{X_n\}$ -time. \square

Lemma 2. *If $\{X_n, n \geq 1\}$ are i.i.d. r.v.s and T is a finite $\{\mathcal{F}_n\}$ -time where \mathcal{F}_n and $\sigma(X_j, j > n)$ are independent, $n \geq 1$, then \mathcal{F}_T and $\sigma(X_{T+1}, X_{T+2}, \dots)$ are independent and $\{X_{T+n}, n \geq 1\}$ are i.i.d. with the same distribution as X_1 .*

PROOF. If $\lambda_1, \dots, \lambda_n$ are real numbers and $A \in \mathcal{F}_T$,

$$\begin{aligned} P\left\{A \cdot \bigcap_{i=1}^n [X_{T+i} < \lambda_i]\right\} &= \sum_{j=1}^{\infty} P\left\{A \cdot [T = j] \bigcap_{i=1}^n [X_{j+i} < \lambda_i]\right\} \\ &= \sum_{j=1}^{\infty} P\{A \cdot [T = j]\} \cdot P\left\{\bigcap_{i=1}^n [X_{j+i} < \lambda_i]\right\} \end{aligned}$$

since $A \cdot [T = j] \in \mathcal{F}_j$. Hence,

$$\begin{aligned} P\left\{A \cdot \bigcap_{i=1}^n [X_{T+i} < \lambda_i]\right\} &= \sum_{j=1}^{\infty} P\{A \cdot [T = j]\} \prod_{i=1}^n P\{X_{j+i} < \lambda_i\} \\ &= \sum_{j=1}^{\infty} P\{A \cdot [T = j]\} \prod_{i=1}^n P\{X_i < \lambda_i\} \\ &= P\{A\} \prod_{i=1}^n P\{X_i < \lambda_i\}. \end{aligned} \quad (7)$$

Hence, taking $A = \Omega$,

$$P\{X_{T+i} < \lambda_i, 1 \leq i \leq n\} = \prod_{i=1}^n P\{X_i < \lambda_i\}$$

and, in particular, $P\{X_{T+j} < \lambda_j\} = P\{X_j < \lambda_j\} = P\{X_1 < \lambda_j\}, 1 \leq j \leq n$. Thus, since n is an arbitrary positive integer, $\{X_{T+n}, n \geq 1\}$ are i.i.d. with the same distribution as X_1 . Consequently, from (7)

$$P\left\{A \cdot \bigcap_{i=1}^n [X_{T+i} < \lambda_i]\right\} = P\{A\} \prod_{i=1}^n P\{X_{T+i} < \lambda_i\}, \quad A \in \mathcal{F}_T,$$

and therefore, in view of the arbitrariness of $\lambda_1, \dots, \lambda_n$ and n , the classes \mathcal{F}_T and $\sigma(X_{T+1}, X_{T+2}, \dots)$ are independent. \square

Corollary 1. $\sigma(T)$ and $\sigma(X_{T+1}, X_{T+2}, \dots)$ are independent.

PROOF. It suffices to recall that T is \mathcal{F}_T -measurable. \square

Next, let $\{X_n, n \geq 1\}$ be i.i.d. r.v.s and T a finite $\{X_n\}$ -time. Then, by Lemma 1 there exist disjoint cylinder sets $\{C_n, n \geq 1\}$ in \mathcal{B}^∞ with n -dimensional Borel bases such that

$$\{T = n\} = \{(X_1, X_2, \dots) \in C_n\}, \quad 1 \leq n < \infty.$$

Define $T^{(1)} = T_1 = T$ and $T^{(j+1)}, j \geq 1$, via $T_j = \sum_{i=1}^j T^{(i)}$, by

$$\{T^{(j+1)} = n\} = \{(X_{T_j+1}, X_{T_j+2}, \dots) \in C_n\}, \quad 1 \leq n < \infty.$$

Then, as noted earlier for $j = 1$, $T^{(j+1)}$ is a finite $\{X_{T_j+n}, n \geq 1\}$ -time, $j \geq 1$. The stopping variables $\{T^{(j+1)}, j \geq 1\}$ or $\{T^{(j)}, j \geq 1\}$ will be called **copies of T** . Moreover, as follows from earlier discussion, $T_m = \sum_{j=1}^m T^{(j)}$ is a (finite) $\{X_n\}$ -time and $\mathcal{F}_{T_m} \subset \mathcal{F}_{T_{m+1}}, m \geq 1$.

In a sequence of Bernoulli trials with parameter $p = \frac{1}{2}$, i.e., $S_n = \sum_{i=1}^n X_i$, where $\{X_n\}$ are i.i.d. r.v.s with $P\{X_i = \pm 1\} = \frac{1}{2}$, let $T = T_1 = T^{(1)} = \inf\{n \geq 1 : S_n = 0\}$ and $T^{(j+1)} = \inf\{n \geq 1 : S_{T_j+n} = 0\}$. Then $T_j = \sum_{i=1}^j T^{(i)}$, $j \geq 1$, are the return times to the origin and $T^{(j)}$ is the time between the $(j-1)$ st and j th return. If, rather, $T = \inf\{n \geq 1 : S_n = 1\}$, then T_j is the **first passage time** through the barrier at j , whereas $T^{(j)}$ is the amount of time required to pass from $j-1$ to j . Either choice of T yields a finite $\{X_n\}$ -time with infinite expectation (Exercises 3.4.1, 3.4.2).

Lemma 3. Let $\{X_n, n \geq 1\}$ be i.i.d. r.v.s and T a finite $\{X_n\}$ -time. If $T_0 = 0$, $T^{(1)} = T$, and $\{T^{(j)}, j > 1\}$ are copies of T , then, setting $T_m = \sum_{j=1}^m T^{(j)}$, the random vectors

$$V_m = (T^{(m)}, X_{T_{m-1}+1}, X_{T_{m-1}+2}, \dots, X_{T_m}), \quad m \geq 1,$$

are i.i.d.

PROOF. As already noted, T_m is a finite $\{X_n\}$ -time. Moreover, it is easy to see that V_m and hence also (V_1, \dots, V_m) is \mathcal{F}_{T_m} -measurable. By Lemma 2, $\sigma(X_{T_m+1}, X_{T_m+2}, \dots)$ is independent of \mathcal{F}_{T_m} , $m \geq 1$, and, since $T^{(m+1)}$ and $(X_{T_m+1}, \dots, X_{T_{m+1}})$ are $\sigma(X_{T_m+1}, X_{T_m+2}, \dots)$ -measurable, $\sigma(V_{m+1})$ and \mathcal{F}_{T_m} are independent for $m \geq 1$. Thus V_{m+1} is independent of (V_1, \dots, V_m) , $m \geq 1$, which is tantamount to independence of $\{V_m, m \geq 1\}$.

Furthermore, for all real λ_i , $1 \leq i \leq n$, and $m, n \geq 1$, if $\{C_n, n \geq 1\}$ are the Borel cylinder sets of (6) defining $T^{(1)}$,

$$\begin{aligned} q_m &= P\{T^{(m)} = n, X_{T_{m-1}+1} < \lambda_1, X_{T_{m-1}+2} < \lambda_2, \dots, X_{T_m} < \lambda_{T^{(m)}}\} \\ &= P\{T^{(m)} = n, X_{T_{m-1}+1} < \lambda_1, \dots, X_{T_{m-1}+n} < \lambda_n\} \\ &= P\{(X_{T_{m-1}+1}, X_{T_{m-1}+2}, \dots, X_{T_{m-1}+n}, \dots) \in C_n, \\ &\quad X_{T_{m-1}+1} < \lambda_1, \dots, X_{T_{m-1}+n} < \lambda_n\}, \\ &= P\{(X_1, X_2, \dots) \in C_n, X_1 < \lambda_1, \dots, X_n < \lambda_n\} = q_1 \end{aligned}$$

by Lemma 2, since T_m is a finite $\{X_n\}$ -time. Thus, $\{V_m, m \geq 1\}$ are i.i.d. random vectors. \square

Corollary 2. If T is a finite $\{X_n\}$ -time, where $\{X_n, n \geq 1\}$ are i.i.d. random variables, then the copies $\{T^{(n)}, n \geq 1\}$ of T are i.i.d. random variables.

If $\{X_n, n \geq 1\}$ are r.v.s and T is an $\{X_n\}$ -time, then the expectation of $|X_T|$ is

$$\begin{aligned} E|X_T| &= \int_{\{T < \infty\}} |X_T| + \int_{\{T = \infty\}} |\overline{\lim} X_n| \\ &= \sum_{j=1}^{\infty} \int_{\{T=j\}} |X_j| + \int_{\{T = \infty\}} |\overline{\lim} X_n|. \end{aligned} \quad (8)$$

Analogously, if the expectation of X_T exists,

$$E X_T = E(X_T)^+ - E(X_T)^- = \sum_{j=1}^{\infty} \int_{\{T=j\}} X_j + \int_{\{T=\infty\}} \overline{\lim} X_n. \quad (9)$$

As customary, X_T is called integrable if $|E X_T| < \infty$. If a stopping time T is integrable, then necessarily $P\{T \geq n\} = o(n^{-1})$ and *a fortiori* T is a finite stopping time or stopping variable.

Consider a sum of a random number of r.v.s $S_T = \sum_1^T X_i$, where T is an $\{X_n, n \geq 1\}$ -time. If T is integrable and the $\{X_n\}$ are i.i.d. \mathcal{L}_1 r.v.s, the expectation of S_T has the natural form (10).

Theorem 1 (Wald's Equation). *Let $\{X_n, n \geq 1\}$ be i.i.d. r.v.s on (Ω, \mathcal{F}, P) , $S_n = \sum_1^n X_i$, $n \geq 1$, and let $\{\mathcal{F}_n, n \geq 1\}$ constitute an increasing sequence of sub- σ -algebras of \mathcal{F} with (i) \mathcal{F}_n and $\sigma(X_{n+1})$ independent, $n \geq 1$. If $E X_1$ exists and T is an $\{\mathcal{F}_n\}$ -time with $E T < \infty$, then*

$$E S_T = E X_1 \cdot E T. \quad (10)$$

PROOF. Suppose first that $E|X_1| < \infty$. Then by Theorem 4.3.3

$$E \sum_1^\infty |X_n| I_{\{T \geq n\}} = \sum_1^\infty E|X_n| I_{\{T \geq n\}} = \sum_1^\infty E|X_n| P\{T \geq n\} = E T \cdot E|X_1|,$$

whence by the dominated convergence theorem

$$E S_T = E \sum_1^\infty X_n \cdot I_{\{T \geq n\}} = \sum_1^\infty P\{T \geq n\} \cdot E X_n = E X_1 \cdot E T.$$

If, rather, $E X_1 = \infty$, then $E X_1^- < \infty$, $E X_1^+ = \infty$, whence by what has just been proved

$$E S_T^- \leq E \sum_1^\infty X_n^- I_{\{T \geq n\}} = E T \cdot E X_1^- < \infty.$$

Therefore, $E S_T$ exists and

$$\begin{aligned} E S_T &= E \sum_1^\infty (X_n^+ - X_n^-) I_{\{T \geq n\}} = E \sum_1^\infty X_n^+ I_{\{T \geq n\}} - E \sum_1^\infty X_n^- I_{\{T \geq n\}} \\ &= E T \cdot E X_1^+ - E T \cdot E X_1^- = E T \cdot E X_1. \end{aligned}$$

The proof for $E X_1 = -\infty$ is analogous. \square

Corollary 3. *If $\{X_n\}$ are i.i.d. r.v.s for which $E X_1$ exists and T is an $\{X_n\}$ -time with $E T < \infty$, then, setting $S_n = \sum_1^n X_i$, (10) obtains.*

The next result is a refinement of Theorem 1, with (i) due to Robbins–Samuel.

Theorem 2. *Let $\{X_n\}$ be i.i.d. r.v.s, let $S_n = \sum_1^n X_i$, and let T be a finite $\{X_n\}$ -time for which $E S_T$ exists. (i) If $E X_1$ exists and either $E X_1 \neq 0$ or $E T < \infty$, then (10) holds. (ii) If $P\{|X_1| > n\} = o(n^{-1})$ and $E T < \infty$, then*

$$\frac{E S_T}{E T} = \lim_{n \rightarrow \infty} E X_1 I_{\{|X_1| \leq n\}}, \quad (11)$$

and when S_T is integrable,

$$\frac{S_n}{n} \xrightarrow{P} \frac{E S_T}{E T}. \quad (12)$$

PROOF. Let $T^{(j)}$, $j \geq 1$, be the copies of T and set $T_0 = 0$, $T_m = \sum_1^m T^{(j)}$. Then according to Corollary 2 and Lemma 3, $T^{(j)}$, $j \geq 1$, are i.i.d. and so are

$S^{(j)} = X_{T_{j-1}+1} + \cdots + X_{T_j}$ for $j \geq 1$. By Corollary 5.2.2

$$\frac{S_{T_m}}{m} = \frac{S^{(1)} + \cdots + S^{(m)}}{m} \xrightarrow{\text{a.s.}} \mathbb{E} S_T, \quad \frac{T_m}{m} \xrightarrow{\text{a.s.}} \mathbb{E} T. \quad (13)$$

If $\mathbb{E} X_1$ exists, then by this same corollary and exercise 3.3.10

$$\frac{S_{T_m}}{T_m} \xrightarrow{\text{a.s.}} \mathbb{E} X_1,$$

which, in conjunction with (13), yields (10). On the other hand, if $\mathbb{P}\{|X_1| > n\} = o(n^{-1})$, then by Theorem 5.2.6

$$\frac{S_{T_m}}{T_m} - c_m \xrightarrow{P} 0, \quad (14)$$

where $c_m = \mathbb{E} X_1 I_{\{|X_1| \leq m\}}$. Hence, any subsequence $\{n'\}$ of the positive integers has itself a subsequence, say $\{n'\}$, for which by (14) and Lemma 3.3.2

$$\frac{S_{T_{n'}}}{T_{n'}} - c_{n'} \xrightarrow{\text{a.s.}} 0. \quad (15)$$

Then, via (15) and (13), recalling that $\mathbb{E} T < \infty$,

$$c_{n'} = \left(c_{n'} - \frac{S_{T_{n'}}}{T_{n'}} \right) + \frac{S_{T_{n'}}}{n'} \cdot \frac{n'}{T_{n'}} \xrightarrow{\text{a.s.}} \frac{\mathbb{E} S_T}{\mathbb{E} T}.$$

Thus, every subsequence of $\{c_n\}$ has itself a subsequence whose limit is independent of the subsequences, yielding (11). Finally, (12) follows from (11) and Theorem 5.2.4. \square

Corollary 4. Let $\{X_n, n \geq 1\}$ be i.i.d. r.v.s, let $S_n = \sum_1^n X_i$ and let T be an integrable $\{X_n\}$ -time. If $n \mathbb{P}(|X_1| > n) = o(1)$ and $\mathbb{E} X_1 I_{\{|X_1| \leq n\}}$ has no limit as $n \rightarrow \infty$, then $\mathbb{E} S_T$ does not exist.

The next theorem is the second moment analogue of Wald's equation.

Theorem 3. If $\{X_n\}$ are independent r.v.s with $\mathbb{E} X_n = 0$, $\mathbb{E} X_n^2 = \sigma^2 < \infty$, $S_n = \sum_1^n X_i$, $n \geq 1$ and T is an $\{\mathcal{F}_n\}$ -time with $\mathbb{E} T < \infty$ where (i) $\mathcal{F}_n \supset \sigma(X_1 \cdots X_n)$ and (ii) \mathcal{F}_n and $\sigma(X_{n+1})$ are independent, $n \geq 1$, then

$$\mathbb{E} S_T^2 = \sigma^2 \mathbb{E} T. \quad (16)$$

PROOF. If $T(n) = \min[T, n]$, then $T(n)$ is a finite stopping time and

$$\begin{aligned} \mathbb{E} S_{T(n)}^2 &= \mathbb{E} \left(\sum_1^n X_j I_{\{T \geq j\}} \right)^2 \\ &= \mathbb{E} \left(\sum_1^{n-1} X_j I_{\{T \geq j\}} \right)^2 + \mathbb{E} X_n^2 I_{\{T \geq n\}} + 2 \mathbb{E} X_n I_{\{T \geq n\}} \sum_1^{n-1} X_j I_{\{T \geq j\}}. \end{aligned}$$

Since $I_{\{T \geq n\}} \sum_1^{n-1} X_j I_{\{T \geq j\}}$ is \mathcal{F}_{n-1} -measurable and $\sigma(X_n)$ is independent of

\mathcal{F}_{n-1} ,

$$\mathbb{E} X_n I_{\{T \geq n\}} \sum_{j=1}^{n-1} X_j I_{\{T \geq j\}} = \mathbb{E} X_n \mathbb{E} \sum_{j=1}^{n-1} X_j I_{\{T \geq j\}} I_{\{T \geq n\}} = 0,$$

whence

$$\mathbb{E} \left(\sum_{j=1}^n X_j I_{\{T \geq j\}} \right)^2 - \mathbb{E} \left(\sum_{j=1}^{n-1} X_j I_{\{T \geq j\}} \right)^2 = \mathbb{E} X_n^2 I_{\{T \geq n\}} = \sigma^2 \mathbb{P}[T \geq n],$$

and summing,

$$\mathbb{E} S_{T(n)}^2 = \mathbb{E} \left(\sum_{j=1}^n X_j I_{\{T \geq j\}} \right)^2 = \sigma^2 \sum_{j=1}^n \mathbb{P}\{T \geq j\} = \sigma^2 \mathbb{E} T(n).$$

Since $T(n) \uparrow T$,

$$\lim_{n \rightarrow \infty} \mathbb{E} S_{T(n)}^2 = \sigma^2 \mathbb{E} T < \infty. \quad (17)$$

Moreover, in completely analogous fashion,

$$\begin{aligned} \mathbb{E}(S_{T(n)} - S_{T(m)})^2 &= \mathbb{E} \left(\sum_{j=m+1}^n X_j I_{\{T \geq j\}} \right)^2 = \sigma^2 [\mathbb{E} T(n) - \mathbb{E} T(m)] \\ &= o(1) \quad \text{as } n > m \rightarrow \infty. \end{aligned}$$

Thus, $S_{T(n)} \xrightarrow{\mathcal{D}_2} S$, and so by Corollary 4.2.5 and (17), $\mathbb{E} S^2 = \sigma^2 \mathbb{E} T$. It remains to identify S with S_T , and this follows from the existence of a subsequence of $S_{T(n)}$ converging a.c. to S and $T(n) \xrightarrow{a.s.} T$. \square

Corollary 5. If $\{X_n\}$ are i.i.d. r.v.s with $\mathbb{E} X_1 = 0$, $\mathbb{E} X_1^2 = \sigma^2 < \infty$ and T is an $\{X_n\}$ -time with $\mathbb{E} T < \infty$, then (16) holds.

EXAMPLE 1. Let $\{X_n, n \geq 1\}$ be independent r.v.s with $\sup_{n \geq 1} \mathbb{E}|X_n|^r \leq M < \infty$, where either $r = 2$, $\mathbb{E} X_n = 0$, or $0 < r \leq 1$. If $\{a_n, n \geq 1\}$ are positive constants with $a_n \downarrow$, $A_n = \sum_{j=1}^n a_j$ and $S_n = \sum_{i=1}^n X_i$, then for any finite $\{X_n\}$ -time T

$$\mathbb{E} a_T |S_T|^r \leq M \mathbb{E} A_T \quad (18)$$

and, moreover, for any α in $(0, 1)$

$$\mathbb{E}|S_T|^{r\alpha} \leq (M/\alpha)^\alpha \mathbb{E} T^\alpha. \quad (19)$$

PROOF. If T is a bounded stopping time, $S_0 = 0$, and $\delta_{r,2} = 2$ or 0 according as $r = 2$ or not, then via independence

$$\begin{aligned} \mathbb{E} a_T |S_T|^r &= \sum_{j=1}^{\infty} \int_{\{T=j\}} a_j |S_j|^r \\ &= \sum_{j=1}^{\infty} \sum_{n=1}^j \int_{\{T=j\}} (a_n |S_{n-1}| + X_n)^r - a_{n-1} |S_{n-1}|^r \end{aligned}$$

$$\begin{aligned} &\leq \sum_{n=1}^{\infty} \int_{\{T \geq n\}} a_n (|X_n|^r + \delta_{r,2} X_n S_{n-1}) \leq M \sum_{n=1}^{\infty} a_n \mathbb{P}\{T \geq n\} \\ &= M \sum_{j=1}^{\infty} \sum_{n=1}^j a_n \mathbb{P}\{T = j\} = M \mathbb{E} A_T. \end{aligned}$$

Hence, for any finite stopping time T , (18) holds for $T(n) = \min[T, n]$, yielding (18) as $n \rightarrow \infty$ by Fatou's lemma and monotone convergence. To prove (19), note that by Hölder's inequality

$$\mathbb{E}|S_T|^{\alpha} = \mathbb{E} \frac{|S_T|^{\alpha}}{T^{\alpha(1-\alpha)}} \cdot T^{\alpha(1-\alpha)} \leq \left(\mathbb{E} \frac{|S_T|^r}{T^{1-\alpha}} \right)^{\alpha} (\mathbb{E} T^{\alpha})^{1-\alpha},$$

and so, employing (18) with $a_n = 1/n^{1-\alpha}$, $A_n \leq n^{\alpha}/\alpha$, (19) follows. \square

As an application of Example 1, if $\{X_n, n \geq 1\}$ are independent r.v.s that obey the Central Limit Theorem with $\mathbb{E} X_n = 0$, $\mathbb{E} X_n^2 = 1$, $S_n = \sum_1^n X_i$, $n \geq 1$ and

$$\begin{aligned} T_c &= \inf\{n \geq 1 : |S_n| \geq cn^{1/2}\}, \quad c > 0, \\ U_m &= \inf\{n \geq m : |S_n| \geq c_n\}, \quad \frac{c_n}{n^{1/2}} \rightarrow \infty, \end{aligned} \tag{20}$$

then $\mathbb{E} T_c^2 = \infty$ for $c^2 > 1/\alpha$, $0 < \alpha < 1$, and $\mathbb{E} U_m^2 = \infty$ for $m > m_\alpha$, all $\alpha > 0$. The latter statement follows from the former which, in turn, results from (19) via

$$c^{2\alpha} \mathbb{E} T_c^2 \leq \mathbb{E}|S_{T_c}|^{2\alpha} \leq \left(\frac{1}{\alpha}\right)^\alpha \mathbb{E} T_c^2.$$

The same conclusion holds if $\{X_n, n \geq 1\}$ are independent r.v.s with mean zero and variance one which obey the central limit theorem (Chapter 9), since T_c is a finite stopping time (Exercise 3.2.12).

Lemma 4. Let $\{X_n, n \geq 1\}$ be i.i.d. random variables and $T = \inf\{n \geq 1 : X_n \in B\}$, where B is a linear Borel set such that $0 < \mathbb{P}\{X_1 \in B\} < 1$. If $T_0 = 0$, $T_n = \sum_{j=1}^n T^{(j)}$, $n \geq 1$, where $\{T^{(j)}, j \geq 1\}$ are copies of T , then setting

$$Y_n = X_{T_n}, \quad Z_n = \sum_{T_{n-1} < j < T_n} X_j, \quad n \geq 1, \tag{21}$$

$\{Y_1, Z_1, Y_2, Z_2, \dots\}$ is a sequence of independent variables.

PROOF. If $n = 1$, Y_1 and Z_1 are independent according to exercise 5.3.12. For $n \geq 2$, via Lemma 3,

$$\begin{aligned} &\mathbb{P} \left\{ \bigcap_{i=1}^n [Y_i < \mu_i, Z_i < \lambda_i] \right\} \\ &= \sum_{j=1}^{\infty} \mathbb{P} \left\{ T^{(n)} = j, X_{T_{n-1}+j} < \mu_n, \sum_{T_{n-1}+1}^{T_{n-1}+j-1} X_i < \lambda_n, \bigcap_{i=1}^{n-1} [Y_i < \mu_i, Z_i < \lambda_i] \right\} \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^{\infty} P \left\{ T^{(n)} = j, X_{T_{n-1}+j} < \mu_n, \sum_{i=T_{n-1}+1}^{T_{n-1}+j-1} X_i < \lambda_n \right\} \\
&\quad \cdot P \left\{ \bigcap_{i=1}^{n-1} [Y_i < \mu_i, Z_i < \lambda_i] \right\} \\
&= P\{Y_n < \mu_n, Z_n < \lambda_n\} P \left\{ \bigcap_{i=1}^{n-1} [Y_i < \mu_i, Z_i < \lambda_i] \right\} \\
&= P\{Y_n < \mu_n\} P\{Z_n < \lambda_n\} P \left\{ \bigcap_{i=1}^{n-1} [Y_i < \mu_i, Z_i < \lambda_i] \right\}
\end{aligned}$$

since the joint distributions of (Y_n, Z_n) and (Y_1, Z_1) are identical. The conclusion then follows by induction. \square

If $\{X, X_n, n \geq 1\}$ are i.i.d. random variables with $E X = 0, E X^2 < \infty$, then $\overline{\lim}_{n \rightarrow \infty} S_n/n^{1/2} = \infty$, a.c., as noted in exercise 3.2.12. This remains true even when $E X^2 = \infty$, according to

Theorem 4 (Stone). If $S_n = \sum_{j=1}^n X_j, n \geq 1$, where $\{X_n, n \geq 1\}$ are i.i.d. random variables with $E X_1 = 0, E|X_1| > 0$, then

$$\overline{\lim}_{n \rightarrow \infty} S_n/n^{1/2} = \infty = -\underline{\lim}_{n \rightarrow \infty} S_n/n^{1/2}, \quad \text{a.s.} \quad (22)$$

PROOF. It suffices to prove the first half of (22). To this end, choose finite constants a, b such that $a < 0 < b$ and $P\{a < X_1 < 0\}P\{0 < X_1 < b\} > 0$. Then $T = \inf\{n \geq 1: a < X_n < b\}$ is an $\{X_n\}$ -time with finite expectation. Since, as noted earlier, the theorem holds when $E X_1^2 < \infty$, it may be supposed that $E X_1^2 = \infty$, whence $0 < P\{a < X_1 < b\} < 1$. Set $T_0 = 0, T_n = \sum_{j=1}^n T^{(j)}, n \geq 1$, where $\{T^{(j)}, j \geq 1\}$ are copies of T and define

$$Y_n = X_{T_n}, \quad Z_n = \sum_{T_{n-1} < j < T_n} X_j. \quad (23)$$

According to Lemma 3, $\{Y_j, j \geq 1\}$ and $Z_j, j \geq 1\}$ are each i.i.d. sequences and, moreover, via Lemma 4, $\{Y_j, j \geq 1\}$ is independent of $\{Z_j, j \geq 1\}$. By Wald's equation, $E Y_1 + E Z_1 = E \sum_1^T X_j = 0$, whence

$$S_{T_n} = \sum_{j=1}^n (Y_j + Z_j) = \sum_{j=1}^n (Y_j - E Y_j) + \sum_{j=1}^n (Z_j - E Z_j). \quad (24)$$

By example 5.2.2, $\{Z_j - E Z_j, j \geq 1\}$ is a recurrent sequence, and so $P\{|\sum_{j=1}^n (Z_j - E Z_j)| < M, \text{i.o.}\} = 1$ for some M in $(0, \infty)$. Define $\tau_m = \inf\{n \geq m: |\sum_{j=1}^n (Z_j - E Z_j)| < M\}$, whence $\tau_m < \infty$, a.s. and $\tau_m \uparrow \infty$ as $m \rightarrow \infty$. Now, $\sigma^2 = E(Y_1 - E Y_1)^2 \in (0, \infty)$ whence

$$M + S_{T_{\tau_m}} \geq \sum_{j=1}^{\tau_m} (Y_j - E Y_j) = \left[\frac{1}{\sigma \tau_m^{1/2}} \sum_{j=1}^{\tau_m} (Y_j - E Y_j) \right] \sigma \tau_m^{1/2} \quad (25)$$

via (24), and since τ_m is independent of $\{Y_j, j \geq 1\}$,

$$\begin{aligned} & P \left\{ \frac{1}{\sigma \tau_m^{1/2}} \sum_{j=1}^{\tau_m} (Y_j - E Y_j) > x \right\} \\ &= \sum_{k=m}^{\infty} P \left\{ \frac{1}{\sigma k^{1/2}} \sum_{j=1}^k (Y_j - E Y_j) > x \right\} P\{\tau_m = k\} \xrightarrow{m \rightarrow \infty} 1 - \Phi(x) \end{aligned}$$

for all x via the Central Limit Theorem for i.i.d. random variables (Corollary 9.1.2).

Consequently, for any $x > 0$, as $n \rightarrow \infty$,

$$P \left\{ S_{T_{\tau_m}} > \frac{x}{2} \sigma \tau_m^{1/2} \right\} \geq 1 - \Phi(x) + o(1). \quad (26)$$

Now, $T_{\tau_m}/\tau_m = \frac{1}{\tau_m} \sum_{j=1}^{\tau_m} T^{(j)} \xrightarrow{\text{a.e.}} E T < \infty$, implying via (26) that, for $x > 0$, as $m \rightarrow \infty$,

$$P \left\{ S_{T_{\tau_m}} > \frac{x\sigma}{3} (T_{\tau_m}/E T)^{1/2} \right\} \geq 1 - \Phi(x) + o(1),$$

and so

$$\begin{aligned} & P \left\{ S_{T_{\tau_m}}/T_{\tau_m}^{1/2} > \frac{x\sigma}{3} E^{-1/2} T, \text{i.o.} \right\} \\ &= \lim_{m \rightarrow \infty} P \left\{ \bigcup_{j=m}^{\infty} \left[S_{T_{\tau_j}}/T_{\tau_j}^{1/2} > \frac{x\sigma}{3} E^{-1/2} T \right] \right\} \\ &\geq 1 - \Phi(x) \end{aligned}$$

whence for all $x > 0$

$$Q_x = P \left\{ \overline{\lim}_{n \rightarrow \infty} S_n/n^{1/2} \geq \frac{x\sigma}{3} E^{-1/2} T \right\} \geq 1 - \Phi(x).$$

Thus, $Q_x = 1$ by the Kolmogorov zero-one law and (22) follows as $x \rightarrow \infty$. \square

Lemma 5. Let $S_0 = 0$, $S_n = \sum_{i=1}^n X_i$, $\bar{S}_n = \max_{0 \leq j \leq n} S_j$, $n \geq 1$, where $\{X, X_n, n \geq 1\}$ are i.i.d. random variables with $E X = 0$, $E|X| > 0$. Then, for $x > 0$ and any integer $k \geq 1$,

$$P\{\bar{S}_n \geq 2kx\} \leq nP\{X > x\} + P^k\{\bar{S}_n \geq x\}.$$

PROOF. If $X_0 = 0$, $\bar{X}_n = \max_{0 \leq j \leq n} X_j$, and $A_k = \{\bar{S}_n \geq 2kx, \bar{X}_n \leq x\}$, it suffices to prove that $P\{A_k\} \leq P^k\{\bar{S}_n \geq x\}$. To this end, define $T_0 = 0$, $T_i = \sum_{j=1}^i T^{(j)}$, $1 \leq i \leq k$, where

$$T^{(i)} = \inf\{j \geq 1: S_{T_{i-1}+j} - S_{T_{i-1}} \geq x\}, \quad 1 \leq i \leq k.$$

By Corollary 5.4.2, T_1 is a finite stopping time and, by Corollary 5.3.2, $\{T^{(i)}, 1 \leq i \leq k\}$ are i.i.d. Clearly, $T_1 \leq n$ on the set A_1 since otherwise $A_1 \cdot \{\bar{S}_n < x\} \neq \emptyset$, a contradiction. For $k \geq 2$, suppose, inductively, that $T_i \leq n$ on A_i for $1 \leq i \leq m < k$. Then, on A_{m+1} ,

$$\bar{S}_{T_{m+1}-1} \leq \sum_{i=1}^{m+1} \max_{1 \leq j \leq T^{(i)}} (S_{T_{i-1}+j} - S_{T_{i-1}}) + \sum_{i=1}^m X_{T^{(i)}} < 2(m+1)x,$$

implying that $n > T_{m+1} - 1$, that is, $T_{m+1} \leq n$ on A_{m+1} . Thus, for $m = k - 1$,

$$P\{A_k\} \leq P\left\{\sum_{i=1}^k T^{(i)} \leq n\right\} = P^k\{\bar{S}_n \geq x\}. \quad \square$$

EXERCISES 5.3

1. Verify that if T is an $\{\mathcal{F}_n\}$ -time, X_{T_m+j} is \mathcal{F}_{T_m+j} -measurable, $j \geq 1, m \geq 1$.
2. If T_1 and T_2 are \mathcal{F}_n -times, so are $T_1 T_2$, $\max(T_1, T_2)$, $\min(T_1, T_2)$, and kT_1 , where k is a positive integer.
3. If T is an integrable $\{X_n\}$ -time, where $\{X_n\}$ are independent r.v.s with $E X_n = 0$, $E|X_n| \leq C < \infty, n \geq 1$, then Wald's equation holds.
4. If $(X_i, Y_i), i \geq 1$, are i.i.d. \mathcal{L}_2 random vectors with $E X_1 = E Y_1 = 0$, and $\mathcal{F}_n = \sigma(X_1, Y_1, \dots, X_n, Y_n)$, $S_n = \sum_{i=1}^n X_i$, $U_n = \sum_{i=1}^n Y_i$, then for any integrable \mathcal{F}_n -time T , the identity $E S_T U_T = E T \cdot EX_1 Y_1$ holds.
5. Let $S_n = \sum_{i=1}^n X_i$ where $\{X_n, n \geq 1\}$ are i.i.d. r.v.s with $E X_1 = \mu > 0$ and $N = N_p$ an $\{X_n\}$ -time (or a r.v. independent of $\{X_n, n \geq 1\}$) having the geometric distribution. Then $\lim_{p \rightarrow 0} P\{S_N/E S_N < x\} = 1 - e^{-x}, x > 0$.
6. Show that the condition $E T < \infty$ cannot be dropped in Corollary 3. Hint: Consider $P\{X_n = 1\} = P\{X_n = -1\} = \frac{1}{2}$ and $T = \inf\{n \geq 1: S_n > 0\}$.
7. If $\{X_n, n \geq 1\}$ are independent random variables with $E X_n = 0, E X_n^2 = 1$ and T^* (resp. T_*) = $\inf\{n \geq 1: |S_n| > (\text{resp. } <)cn^{1/2}\}$, where $\inf \emptyset = \infty$, then $E T^* = \infty, c \geq 1$ and $E T_* = \infty, c \leq 1$.
8. If $\{X_n, n \geq 1\}$ are independent r.v.s with $E X_n = 0, E X_n^2 = \sigma_n^2, n \geq 1$ and T is an \mathcal{F}_n -time for which $E \sum_1^T \sigma_j^2 < \infty$, where (i) $\mathcal{F}_n \supset \sigma(X_1, \dots, X_n)$, (ii) \mathcal{F}_n and $\sigma(X_{n+1})$ are independent, $n \geq 1$, then $E(\sum_1^T X_j)^2 = E \sum_1^T \sigma_j^2$.
9. (Yadrenko) If $\{X_n, n \geq 1\}$ are i.i.d. r.v.s uniformly distributed on $[0, 1]$ and $T = \inf\{n > 1: S_n \geq 1\}$, where $S_n = \sum_1^n X_j$, prove that $E T = e = 2 E S_T$.
10. Utilize the method of Example 1 to give an alternative proof of Theorem 3.
11. Let $\{X_n, n \geq 1\}$ be i.i.d. with $P\{X_1 > M\} > 0$, all M in $(0, \infty)$. If $T_c = \inf\{n \geq 1: X_n \geq c\}$, where $|c| < \infty$, prove that $E T_c^m < \infty, m \geq 1$, and $E X_{T_c} \leq E X_{T_d}$ for $c < d$ and $E X_{T_c} = P^{-1}\{X \geq c\} \int_{\{X \geq c\}} X$.
12. If $\{X, X_n, n \geq 1\}$ are i.i.d., B is a linear Borel set with $P\{X \in B\} > 0$, and $T = \inf\{n \geq 1: X_n \in B\}$, then X_T and $\sum_1^{T-1} X_i$ are independent r.v.s. Moreover, if $E|X| < \infty$, then $E X_T = E X I_{\{X \in B\}} \cdot E T$.

13. Show that $\limsup_{n \rightarrow \infty} n^{-1/2} |\sum_{i=1}^n X_i| = \infty$, a.c. for any sequence of i.i.d. random variables $\{X_n, n \geq 1\}$ with $E|X_1| > 0$.
14. Let $S_n = \sum_{i=1}^n X_i$, $n \geq 1$, where $\{X, X_n, n \geq 1\}$ are i.i.d. random variables with $E X = 0$, $E|X| > 0$. For $x > 0$, $\beta > 0$, and $k = 1, 2, \dots$, show that

$$q_k \equiv P \left\{ \max_{1 \leq j \leq n} S_j/j^\beta \geq 2kx \right\} \leq P \left\{ \max_{1 \leq j \leq n} X_j/j^\beta \geq x \right\} + P^k \left\{ \max_{1 \leq j \leq n} S_j/j^\beta \geq x \right\},$$

and if $1 \leq r \leq 2$ and $\beta < 1/r$, there exists a universal constant $C_{r,\beta}$ such that

$$q_k \leq P \left\{ \max_{1 \leq j \leq n} X_j/j^\beta \geq x \right\} + [C_{r,\beta}(2x)^{-r} n^{1-r\beta} E|X|^r]^k.$$

Hint: This generalizes Lemma 5.3.5.

5.4 Chung–Fuchs Theorem, Elementary Renewal Theorem, Optimal Stopping

An instance when the generality of Theorem 5.3.1 (as opposed to the specificity of its corollary) is desirable is furnished by

Theorem 1. *Let $\{X_n\}$ be i.i.d. r.v.s with $E|X_1| > 0$, $S_n = \sum_1^n X_i$, and define*

$$T_+ = \inf\{n \geq 1 : S_n \geq 0\}, \quad T_c = \inf\{n \geq 1 : S_n > c > 0\}. \quad (1)$$

Then:

- i. $P\{T_+ < \infty\} = 1$ iff $\overline{\lim} S_n \stackrel{a.e.}{=} \infty$, in which case $E S_{T_+} > 0$;
- ii. $E T_+ < \infty$ iff $E T_c < \infty$ for all c in $(0, \infty)$.

PROOF. If under the hypothesis of (i), T_+ is a finite $\{X_n\}$ -time, then S_{T_+} is a r.v. and $E S_{T_+} \geq E X_1^+ \geq 0$. Let $T^{(j)}$, $j \geq 1$, be copies of T_+ and set $T_0 = 0$, $T_n = \sum_1^n T^{(j)}$, $n \geq 1$. By Lemma 5.3.3, the r.v.s $S^{(m)} = X_{T_{m-1}+1} + \dots + X_{T_m}$, $m \geq 1$, are i.i.d., whence by Corollary 5.2.2

$$\frac{S_{T_n}}{n} = \frac{S^{(1)} + \dots + S^{(n)}}{n} \xrightarrow{a.e.} E S_{T_+} \geq E X_1^+.$$

Hence, $\overline{\lim} S_n \geq \overline{\lim} n E X_1^+ \geq 0$, a.c., and, moreover, $\overline{\lim} S_n = +\infty$, a.c., since $E X_1^+ > 0$ in view of the fact that the only permissible alternative, namely, $E X_1 < 0$, would imply (same corollary) that $S_n/n \xrightarrow{a.e.} E X_1 < 0$, contradicting $\overline{\lim} S_n \geq 0$, a.c. The converse is trivial.

Apropos of (ii), since via (i)

$$\frac{S^{(1)} + \dots + S^{(n)}}{n} \xrightarrow{a.e.} E S_{T_+} > 0,$$

for any $c > 0$ there exists an integer k such that $P\{\sum_1^k S^{(j)} > c\} \geq \frac{1}{2}$. Setting $Z_n = \sum_{(n-1)k+1}^{nk} S^{(j)}$, the r.v.s $\{Z_n, n \geq 1\}$ are i.i.d. by Lemma 5.3.3 and clearly $Z_n \geq 0$, a.c., with $P\{Z_1 > c\} \geq \frac{1}{2}$. Define $\tau = \inf\{n \geq 1 : Z_n > c\}$. Then

$$P\{\tau \geq n\} = P\{Z_1 \leq c, \dots, Z_{n-1} \leq c\} = P^{n-1}\{Z_1 \leq c\},$$

whence

$$E\tau = \sum_{n=1}^{\infty} P\{\tau \geq n\} = \sum_{n=1}^{\infty} P^{n-1}\{Z_1 \leq c\} = \frac{1}{P\{Z_1 > c\}} \leq 2.$$

Moreover, since

$$c < \sum_1^{\tau} Z_i = \sum_1^{k\tau} S^{(j)} = \sum_{j=1}^{k\tau} \sum_{T_{j-1}+1}^{T_j} X_i = \sum_1^{T_{k\tau}} X_i = S_{T_{k\tau}},$$

necessarily $T'_c \leq T_{k\tau}$, whence

$$E T'_c \leq E T_{k\tau} = E \sum_1^{k\tau} T^{(j)} = E(k\tau) \cdot E T_+ < \infty$$

by Wald's equation since $k\tau$ is a \mathcal{G}_n -time, where $\mathcal{G}_n = \sigma(T^{(1)}, \dots, T^{(n)}, S^{(1)}, \dots, S^{(n)})$, i.e., $\{k\tau = kn\} \in \mathcal{G}_{nk}$ and $\sigma(T^{(n+1)})$ is independent of \mathcal{G}_n (recall Lemma 5.3.3). Again, the converse is trivial since $T_+ \leq T'_c$. \square

The stopping rule T_+ and its partial analogue T'_{-} , defined in (2) below, are interconnected as in

Theorem 2. Let $\{X_n\}$ be i.i.d. r.v.s, $S_n = \sum_1^n X_i$, and define

$$\begin{aligned} T_+ &= \inf\{n \geq 1 : S_n \geq 0\}, & T_- &= \inf\{n \geq 1 : S_n \leq 0\}, \\ T'_{+} &= \inf\{n \geq 1 : S_n > 0\}, & T'_{-} &= \inf\{n \geq 1 : S_n < 0\}. \end{aligned} \tag{2}$$

Then, (i)

$$E T_+ = \frac{1}{P\{T'_- = \infty\}}, \quad E T'_{+} = \frac{1}{P\{T_- = \infty\}}, \tag{3}$$

$$P\{T_- = \infty\} = (1 - \xi)P\{T'_- = \infty\}, \tag{4}$$

where

$$\xi = \sum_1^{\infty} P\{S_1 > 0, \dots, S_{n-1} > 0, S_n = 0\} = P\{S_{T_-} = 0, T_- < \infty\}. \tag{5}$$

Moreover, if $E|X_1| > 0$,

ii. $\xi < 1$,

iii. T_- is defective iff $S_n \xrightarrow{a.e.} \infty$.

PROOF. (i) For $k > 1$ and $1 \leq n \leq k$, define

$$A_1^k = \{S_1 \leq S_2, \dots, S_1 \leq S_k\},$$

$$A_n^k = \{S_1 > S_n, \dots, S_{n-1} > S_n \leq S_{n+1}, \dots, S_n \leq S_k\}.$$

Then, if $k \geq n \geq 1$,

$$\begin{aligned} P\{A_n^k\} &= P\{X_n < 0, \dots, X_2 + \dots + X_n < 0\} \\ &\quad \cdot P\{X_{n+1} \geq 0, \dots, X_{n+1} + \dots + X_k \geq 0\} \\ &= P\{T_+ \geq n\}P\{T'_- > k - n\}, \end{aligned}$$

and so when $k \geq 1$

$$1 = \sum_{n=1}^k P\{T_+ \geq n\}P\{T'_- > k - n\}, \quad (6)$$

yielding

$$1 \geq P\{T'_- = \infty\}E T_+, \quad (7)$$

whence $P\{T'_- = \infty\} > 0$ implies $E T_+ < \infty$. Conversely, if $E T_+ < \infty$, via (6)

$$1 \leq \sum_{n=1}^j P\{T_+ \geq n\}P\{T'_- > k - j\} + \sum_{n=j}^k P\{T_+ \geq n\};$$

letting $k \rightarrow \infty$,

$$1 \leq \sum_{n=1}^j P\{T_+ \geq n\}P\{T'_- = \infty\} + \sum_{n=j}^{\infty} P\{T_+ \geq n\},$$

and then letting $j \rightarrow \infty$,

$$1 \leq P\{T'_- = \infty\}E T_+,$$

implying $P\{T'_- = \infty\} > 0$. Consequently, $P\{T'_- = \infty\}E T_+ = 1$ if either $P\{T'_- = \infty\} > 0$ or $E T_+ < \infty$, and so, without qualification,

$$E T_+ = \frac{1}{P\{T'_- = \infty\}}.$$

Similarly,

$$E T'_- = \frac{1}{P\{T_- = \infty\}},$$

establishing (3). Next,

$$\begin{aligned} P\{T'_- = \infty\} - P\{T_- = \infty\} &= P\{T'_- = \infty, T_- < \infty\} \\ &= \sum_{n=1}^{\infty} P\{T_- = n, T'_- = \infty\} \end{aligned}$$

$$\begin{aligned}
&= \sum_{n=1}^{\infty} P\{S_1 > 0, \dots, S_{n-1} > 0, S_n = 0, T'_- = \infty\} \\
&= \sum_{n=1}^{\infty} P\{S_1 > 0, \dots, S_{n-1} > 0, S_n = 0, S_{n+1} \geq 0, S_{n+2} \geq 0, \dots\} \\
&= \sum_{n=1}^{\infty} P\{S_1 > 0, \dots, S_{n-1} > 0, S_n = 0, X_{n+1} \geq 0, X_{n+1} + X_{n+2} \geq 0, \dots\} \\
&= \sum_{n=1}^{\infty} P\{S_1 > 0, \dots, S_{n-1} > 0, S_n = 0\} P\{S_1 \geq 0, S_2 \geq 0, \dots\} \\
&= \xi P\{T'_- = \infty\},
\end{aligned}$$

yielding

$$(1 - \xi)P\{T'_- = \infty\} = P\{T_- = \infty\},$$

which completes the proof of (i).

(ii) Suppose that $E|X_1| > 0$. If $\xi = 1$, then $T_- < \infty$, a.c., and, replacing $\{X_n\}$ by $\{-X_n\}$ in Theorem 1(i), $E S_{T_-} < 0$. On the other hand, $\xi = 1$ entails $S_{T_-} = 0$, a.c., via (5), a flagrant contradiction.

(iii) If T_- and hence also T'_- is defective, $E T_+ < \infty$ by (3) and *a fortiori* $T_+ < \infty$, a.c., so that according to Theorem 1(i)

$$\overline{\lim}_{n \rightarrow \infty} S_n \stackrel{\text{a.c.}}{\equiv} \infty. \quad (8)$$

Thus, $T = \inf\{n \geq 1 : S_n > K\} < \infty$, a.c. for all $K > 0$ and, via Lemma 5.3.2,

$$P\{\underline{\lim}_{n \rightarrow \infty} (S_{T+n} - S_T) \geq 0\} = P\{\underline{\lim}_{n \rightarrow \infty} S_n \geq 0\} = \varepsilon > 0.$$

Consequently, with probability at least ε ,

$$\underline{\lim}_{n \rightarrow \infty} S_n = \underline{\lim}_{n \rightarrow \infty} S_{T+n} = S_T + \underline{\lim}_{n \rightarrow \infty} (S_{T+n} - S_T) > K,$$

implying that $P\{\underline{\lim}_{n \rightarrow \infty} S_n = \infty\} \geq \varepsilon > 0$ whence, by the Kolmogorov zero-one law, $\underline{\lim}_{n \rightarrow \infty} S_n = \infty$, a.c., that is, $S_n \stackrel{\text{a.c.}}{\rightarrow} \infty$.

Conversely, if $\underline{\lim}_{n \rightarrow \infty} S_n \stackrel{\text{a.c.}}{\equiv} \infty$, T cannot be finite since Theorem 1 (with $X \rightarrow -X$) then would entail $\underline{\lim}_{n \rightarrow \infty} S_n \stackrel{\text{a.c.}}{\equiv} -\infty$. \square

A combination of Theorems 1 and 2 yields

Corollary 1. *If $\{X_n\}$ are i.i.d. r.v.s, $S_n = \sum_1^n X_i$, and*

$$T_+ = \inf\{n \geq 1 : S_n \geq 0\}, \quad T_- = \inf\{n \geq 1 : S_n \leq 0\},$$

$$T'_c = \inf\{n \geq 1 : S_n > c > 0\}, \quad T'_{-c} = \inf\{n \geq 1 : S_n < -c < 0\},$$

then either $S_n = 0$, $n \geq 1$, a.c., or one of the following holds:

- T_- is defective, $\underline{\lim}_{n \rightarrow \infty} S_n \stackrel{\text{a.c.}}{\equiv} \infty$, $E T_+ < \infty$, $E T'_c < \infty$, $c > 0$;*

- ii. T_+ is defective, $\lim S_n \stackrel{a.c.}{=} -\infty$, $E T_- < \infty$, $E T'_{-c} < \infty$, $c > 0$;
- iii. T_+ and T_- are finite, $\overline{\lim} S_n \stackrel{a.c.}{=} \infty$, $\underline{\lim} S_n \stackrel{a.c.}{=} -\infty$, $E T_+ = E T_- = \infty$.

PROOF. Supposing $\{S_n, n \geq 1\}$ nondegenerate, necessarily $E|X_1| > 0$.

(i) If T_- is defective, T'_{-c} is defective *a fortiori*, whence by Theorem 2 $S_n \xrightarrow{a.c.} \infty$ and $E T_+ < \infty$. Moreover, by Theorem 1(ii), $E T'_c < \infty$ for $c > 0$. Similarly for (ii).

If, as in (iii), both T_+ and T_- are finite, then Theorem 1(i) ensures that

$$\overline{\lim} S_n = \infty \quad a.c., \quad \underline{\lim} S_n = -\infty \quad a.c.;$$

moreover, according to Theorem 2(ii) and (4), T'_{-c} and (analogously) T'_c are finite, whence by Theorem 2(i), $E T_+ = E T_- = \infty$. \square

The following provides a criterion for category (iii) of Corollary 1.

Corollary 2 (Chung–Fuchs). *Let $\{X_n, n \geq 1\}$ be i.i.d. r.v.s with $E X_1 = 0$ and $E|X_1| > 0$ and let $S_n = \sum_1^n X_j$. Then*

$$\overline{\lim}_{n \rightarrow \infty} S_n = \infty \quad a.c., \quad \underline{\lim}_{n \rightarrow \infty} S_n = -\infty \quad a.c.$$

PROOF. By Corollary 1, it suffices to prove that $E T_+ = E T_- = \infty$. Suppose, on the contrary, that $E T_+$ (say) $< \infty$. Then, by Wald's equation,

$$0 \leq E(X_1^+) \leq E S_{T_+} = E T_+ \cdot E X_1 = 0,$$

implying $E X_1^+ = 0$. Since $E X_1 = 0$, $E X_1^- = 0$ and therefore $E|X_1| = 0$, a contradiction. \square

The next theorem, which extends Corollary 2, asserts that the same conclusion holds if $S_n/n \xrightarrow{P} 0$ and $E|X_1| > 0$.

Theorem 3 (Chung–Ornstein). *If $\{X_n\}$ are i.i.d. r.v.s with $E|X_1| > 0$, $S_n = \sum_1^n X_i$, and*

$$\frac{S_n}{n} \xrightarrow{P} 0 \tag{9}$$

then T_+ and T_- are finite and

$$P\left\{\overline{\lim}_n S_n = \infty\right\} = 1 = P\left\{\underline{\lim}_n S_n = -\infty\right\}. \tag{10}$$

PROOF. According to Corollary 1 it suffices to prove that T_+ , T_- are finite and, by symmetry, merely that T_- is finite. Suppose contrariwise that T_- is defective, whence $E T_+ < \infty$ by Theorem 2. Now the WLLN (Theorem 5.2.4) together with (9) implies

$$n P\{|X_1| > n\} = o(1), \quad E X_1 I_{\{|X_1| \leq n\}} = o(1). \tag{11}$$

Since $E S_{T_+}$ exists and $E T_+ < \infty$ as noted above, Theorem 5.3.2(ii) ensures

$E S_{T_+} = 0$. Thus, $E X_1^+ \leq E S_{T_+} = 0$, entailing $E X_1 I_{\{X_1 \leq 0\}} \rightarrow -E X_1^- < 0$, in contradiction of (11). Consequently T_+ is finite. \square

The same proof yields

Corollary 3. Let $\{X_n\}$ be i.i.d. r.v.s with $E|X_1| > 0$, $S_n = \sum_i^n X_i$, and $n P\{|X_1| > n\} \rightarrow 0$. (i) If either $\lim E X_1 I_{\{X_1 \leq n\}}$ does not exist or $\lim E X_1 I_{\{X_1 \leq n\}} \geq 0$, then $T_+ < \infty$, a.c. (ii) If, moreover, $E X_1^+ = E X_1^- = \infty$, and if $\lim E X_1 \times I_{\{X_1 \leq n\}} = c$ finite or $E X_1 I_{\{X_1 \leq n\}}$ has no limit, then $P\{\overline{\lim} S_n = \infty\} = P\{\underline{\lim} S_n = -\infty\} = 1$.

Corollary 1 implies that apart from degeneracy only three possible modes of behavior exist for sums S_n of i.i.d. r.v.s $\{X_n\}$. If, moreover, $E X_1^+ = E X_1^- = \infty$, the same trichotomy will be shown to exist for the averages S_n/n .

Lemma 1. (i) If $\{X_n, n \geq 1\}$ are i.i.d. r.v.s with $S_n = \sum_i^n X_i$, $n \geq 1$, then $S_n \xrightarrow{a.s.} \infty$ iff there exists an $\{X_n\}$ -time T for which $E T < \infty$ and $E S_T > 0$. (ii) Moreover, when $E X_1^+ = \infty$, then $(S_n/n) \xrightarrow{a.s.} \infty$ iff there exists an $\{X_n\}$ -time T with $E T < \infty$ and $E S_T > -\infty$.

PROOF. Under (i), if $S_n \xrightarrow{a.s.} \infty$, then Corollary 1 ensures that $E T_+ < \infty$, where $T_+ = \inf\{n \geq 1 : S_n \geq 0\}$, and clearly $E S_{T_+} \geq E X_1^+ > 0$. Conversely, if T is an $\{X_n\}$ -time with $E T < \infty$ and $E S_T > 0$, let $T^{(j)}, j \geq 1$, be copies of T with $T_n = \sum_i^n T^{(j)}$, $T_0 = 0$. By Lemma 5.3.3, $V_n = (T^{(1)}, S^{(n)})$, $n \geq 1$, are i.i.d. random vectors, where $S^{(n)} = X_{T_{n-1}+1} + \dots + X_{T_n}$. Define a $\{V_n\}$ -time τ by

$$\tau = \inf\{n \geq 1 : S_{T_n} \geq 0\} = \inf\left\{n \geq 1 : \sum_1^n S^{(j)} \geq 0\right\}.$$

Since $\sum_1^n S^{(j)} \xrightarrow{a.s.} \infty$, Corollary 1 guarantees that $E \tau < \infty$ and since $S_{T_\tau} \geq 0$, $T_+ \leq T_\tau$. Hence, via Wald's equation

$$E T_+ \leq E T_\tau = E \sum_1^\tau T^{(j)} = E \tau \cdot E T^{(1)} < \infty$$

and, invoking Corollary 1 once more, $S_n \xrightarrow{a.s.} \infty$.

Apropos of (ii), let T be an $\{X_n\}$ -time with $E T < \infty$ and $E S_T > -\infty$. For $K > 0$, define $X'_n = X_n - K$, $S'_n = \sum_i^n X'_j$, and

$$T' = T I_{\{X'_1 \leq 0\}} + I_{\{X'_1 > 0\}}.$$

Then $T' \leq T$, $E(S'_T)^+ \geq E(X'_1)^+ = \infty$, and

$$E(S'_T)^- \leq E(S'_T)^- \leq E(S_T^- + KT) < \infty.$$

Consequently, by (i) $S_n - nK = S'_n \xrightarrow{\text{a.e.}} \infty$ for all $K > 0$, implying $S_n/n \xrightarrow{\text{a.e.}} \infty$. The remainder of (ii) follows trivially from (i). \square

Theorem 4 (Kesten). *If $\{X_n, n \geq 1\}$ are i.i.d. r.v.s with $E X_1^+ = E X_1^- = \infty$ and $S_n = \sum_1^n X_i, n \geq 1$, then one of the following holds:*

- i. $S_n/n \xrightarrow{\text{a.e.}} \infty$;
- ii. $S_n/n \xrightarrow{\text{a.e.}} -\infty$;
- iii. $\overline{\lim}(S_n/n) \xrightarrow{\text{a.e.}} \infty$, $\underline{\lim}(S_n/n) \xrightarrow{\text{a.e.}} -\infty$.

PROOF. (i) If $S_n \xrightarrow{\text{a.e.}} \infty$, Corollary 1 ensures $E T_+ < \infty$ and $E S_{T_+} \geq E X_1^+ = \infty$, where $T_+ = \inf\{n \geq 1 : S_n \geq 0\}$, so that $S_n/n \xrightarrow{\text{a.e.}} \infty$ by Lemma 1. Similarly, $S_n \xrightarrow{\text{a.e.}} -\infty$ guarantees (ii). Since the hypothesis precludes degeneracy, only the alternative $\overline{\lim} S_n = \infty$, a.e., and $\underline{\lim} S_n = -\infty$, a.e., remains, in which case by (ii) of Lemma 1 no $\{X_n\}$ -time T with $E T < \infty$ and $|E S_T| \leq \infty$ exists. Hence, if $X'_n = X_n - K$ and $S'_n = \sum_1^n X'_1$, no $\{X_n\}$ -time T' with $E T' < \infty$ and $|E S'_{T'}| \leq \infty$ exists for any finite constant K . Again invoking Lemma 1 and Corollary 1, $\overline{\lim} S'_n = \infty$, a.e., and $\underline{\lim} S'_n = -\infty$, a.e. In view of the arbitrariness of K , (iii) follows. \square

Corollary 4. *If $\{X_n\}$ are i.i.d. random variables, then, for every α in $(0, \frac{1}{2}]$, one of the following holds:*

- (i) $S_n/n^\alpha \xrightarrow{\text{a.e.}} \infty$.
- (ii) $S_n/n^\alpha \xrightarrow{\text{a.e.}} -\infty$.
- (iii) $\overline{\lim} S_n/n^\alpha \xrightarrow{\text{a.e.}} \underline{\lim} S_n/n^\alpha \xrightarrow{\text{a.e.}} -\infty$.

PROOF. If $\mu = E X$ exists, then $S_n/n \xrightarrow{\text{a.e.}} \mu \in [-\infty, \infty]$, and if $\mu \neq 0$, either (i) or (ii) holds. If, rather, $\mu = 0$, Theorem 5.3.4 asserts that (iii) obtains for $\alpha = 1/2$ (a fortiori, for $\alpha < \frac{1}{2}$). Finally, if $E X^+ = E X^- = \infty$, Corollary 4 follows from Theorem 4. \square

Renewal theory is concerned with the so-called renewal function $E N_c$, where

$$N_c = \max \left\{ j : S_j = \sum_1^j X_i \leq c \right\}, \quad c > 0,$$

the r.v.s $\{X_n, n \geq 1\}$ being i.i.d. with $\mu = E X_1 \in (0, \infty]$. Although, N_c is not a stopping time, when $X_1 \geq 0$, a.e. $N_c + 1 = T'_c$. Thus, when $X_1 \geq 0$, a.e., (13) and (14) below hold with N_c replacing T'_c and the former is known as the **elementary renewal theorem**. A stronger result (due to Blackwell) asserts that $E N_{c+x} - E N_c \rightarrow \alpha/\mu$ as $c \rightarrow \infty$ (modification being necessary when $\{X_n\}$ are lattice r.v.s).

Stopping times may be utilized to obtain the elementary renewal theorem. A first step in this direction is

Lemma 2 (Gundy-Siegmund). *Let $\{X_n, n \geq 1\}$ be independent, nonnegative \mathcal{L}_1 r.v.s and let $\{T_m, m \geq 1\}$ be a sequence of $\{X_n\}$ -times satisfying $E T_m < \infty$, $m \geq 1$, and $\lim_{m \rightarrow \infty} E T_m = \infty$. If*

$$\sum_{j=1}^n \int_{\{X_j > \varepsilon j\}} X_j = o(n), \quad \varepsilon > 0,$$

then $E X_{T_n} = o(E T_n)$ as $n \rightarrow \infty$.

PROOF. For $\varepsilon > 0$, choose $N \geq 1$ such that $\sum_{j=1}^n \int_{\{X_j > \varepsilon j\}} X_j < n\varepsilon$ for $n \geq N$. Then, if $n \geq N$,

$$\begin{aligned} E X_{T_n} &= \int_{\{X_{T_n} \leq \varepsilon T_n\}} X_{T_n} + \sum_{j=1}^{\infty} \int_{\{T_n = j, X_j > \varepsilon j\}} X_j \\ &\leq \varepsilon E T_n + \sum_{j=1}^{N-1} E X_j + \sum_{j=N}^{\infty} E X_j I_{\{X_j > \varepsilon j, T_n \geq j\}} \\ &= \varepsilon E T_n + O(1) + \sum_{j=N}^{\infty} P\{T_n \geq j\} E X_j I_{\{X_j > \varepsilon j\}} \\ &= \varepsilon E T_n + O(1) + \sum_{k=N}^{\infty} \sum_{j=N}^k E X_j I_{\{X_j > \varepsilon j\}} P\{T_n = k\} \\ &\leq \varepsilon E T_n + O(1) + \sum_{k=N}^{\infty} \varepsilon k P\{T_n = k\} \leq 2\varepsilon E T_n + O(1), \end{aligned}$$

and so $E X_{T_n} = o(E T_n)$. \square

Theorem 5 (Elementary Renewal Theorem). *If $\{X_n, n \geq 1\}$ are i.i.d. with $\mu = E X_1 \in (0, \infty]$, $S_n = \sum_1^n X_i$, and*

$$T'_c = \inf\{n \geq 1 : S_n > c\}, \quad c > 0, \quad (12)$$

then (i) $E T'_c < \infty$ and

$$\lim_{c \rightarrow \infty} \frac{E T'_c}{c} = \frac{1}{E X_1}. \quad (13)$$

Moreover, (ii) if $\sigma_{X_1}^2 = \sigma^2 < \infty$, then

$$\lim_{c \rightarrow \infty} \frac{\sigma^2(T'_c)}{c} = \frac{\sigma^2}{\mu^3} \quad (14)$$

PROOF. By Corollary 5.2.2, $S_n/n \xrightarrow{a.s.} E X_1 > 0$, whence $S_n \xrightarrow{a.s.} \infty$. Thus, Corollary 5.4.1 guarantees that the $\{X_n\}$ -time T'_c has finite expectation.

For any m in $(0, \mu)$, choose the positive constant K sufficiently large to ensure $\mathbb{E} X_1 I_{[X_1 \leq K]} > m$ and define

$$X'_n = X_n I_{[X_n \leq K]}, \quad S'_n = \sum_1^n X'_i, \quad V = \inf\{n \geq 1 : S'_n > c\}.$$

Then $\{X'_n\}$ are i.i.d. and, as earlier, $\mathbb{E} V < \infty$. By Wald's equation

$$K + c \geq \mathbb{E} S'_V = \mathbb{E} V \cdot \mathbb{E} X'_1,$$

and so

$$\frac{\mathbb{E} T'_c}{K + c} \leq \frac{\mathbb{E} V}{K + c} \leq \frac{1}{\mathbb{E} X'_1} \leq \frac{1}{m},$$

whence

$$\varlimsup_{c \rightarrow \infty} \frac{\mathbb{E} T'_c}{c} \leq \frac{1}{m}$$

and ($m \rightarrow \mu$)

$$\varlimsup_{c \rightarrow \infty} \frac{\mathbb{E} T'_c}{c} \leq \frac{1}{\mu}. \quad (15)$$

If $\mu = \infty$, (15) is tantamount to (13). If, rather, $\mu < \infty$, then by Wald's equation $\mu \mathbb{E} T'_c = \mathbb{E} S_{T'_c} > c$, implying

$$\varliminf_{c \rightarrow \infty} \frac{\mathbb{E} T'_c}{c} \geq \frac{1}{\mu},$$

which, in conjunction with (15), yields (13).

Apropos of (ii), setting $T = T'_c$, by Theorem 5.3.3

$$\mathbb{E}(S_T - \mu T)^2 = \sigma^2 \mathbb{E} T < \infty. \quad (16)$$

Since $T = T'_c \uparrow \infty$ as $c \uparrow \infty$ and $\int_{[X_1^2 > \varepsilon n]} X_1^2 = o(1)$, $\varepsilon > 0$, by Lemma 2 and part (i)

$$\mathbb{E} X_T^2 = o(\mathbb{E} T) = o(c) \quad \text{as } c \rightarrow \infty.$$

Hence,

$$\mathbb{E}(S_T - c)^2 \leq \mathbb{E}(X_T^+)^2 \leq \mathbb{E} X_T^2 = o(c) \quad \text{as } c \rightarrow \infty, \quad (17)$$

so that

$$\begin{aligned} \mathbb{E} S_T^2 &< \infty, \quad \mathbb{E}|X_T| = o(\sqrt{c}), \\ c \leq \mu \mathbb{E} T = \mathbb{E} S_T &\leq c + \mathbb{E}|X_T| = c + o(\sqrt{c}), \end{aligned} \quad (18)$$

whence $\mathbb{E} T^2 < \infty$ via (16) and (18).

From (17) and (18)

$$\mathbb{E}(S_T - \mu \mathbb{E} T)^2 \leq 2[\mathbb{E}(S_T - c)^2 + (c - \mu \mathbb{E} T)^2] = o(c), \quad (19)$$

and from (16), (19), and Schwarz's inequality

$$\mu^2 E(T - E T)^2 = E(\mu T - S_T + S_T - \mu E T)^2 = \sigma^2 E T + o(c).$$

Hence, by (13)

$$\sigma_T^2 = E(T - E T)^2 = \frac{\sigma^2}{\mu^2} E T + o(c) = \frac{\sigma^2}{\mu^3} c + o(c),$$

which is (14). \square

When $X_n, n \geq 1$ are i.i.d. with $E X_1 = \mu > 0, \sigma^2 = E X_1^2 \in (0, \infty), (\mu^3/c\sigma^2)^{1/2}$, $(T_c - (c/\mu))$ has a limiting normal distribution as $c \rightarrow \infty$ according to Theorem 9.4.2.

EXAMPLE 1. If $\{X, X_n, n \geq 1\}$ are nonnegative i.i.d. r.v.s with $P\{X > 0\} > 0$ and $T_x = \inf\{n \geq 1 : S_n = \sum_i^n X_i > x\}$,

$$\frac{1}{2} E T_x \leq \frac{x}{E \min(X, x)} \leq E T_x, \quad x > 0. \quad (20)$$

Remark. Since, setting $S_0 = 0$,

$$E T_x = \sum_{n=1}^{\infty} P\{T_x \geq n\} = \sum_{n=0}^{\infty} P\{S_n \leq x\},$$

and since if F is the d.f. of X and

$$a(x) = \frac{x}{\int_0^x [1 - F(y)] dy}, \quad (21)$$

according to Corollary 6.2.1 and Example 6.2.2

$$\frac{x}{a(x)} = \int_0^x y dF(y) + x[1 - F(x)] = E \min(X, x),$$

it follows that the inequalities of (20) may be transcribed as

$$\frac{1}{2} \sum_{n=0}^{\infty} P\{S_n \leq x\} \leq a(x) \leq \sum_{n=0}^{\infty} P\{S_n \leq x\}, \quad x > 0. \quad (22)$$

PROOF. Since $T_x \geq n$ iff $S_{n-1} \leq x$, the series expression for $E T_x$ is immediate (see Exercise 4.1.4). Set

$$X'_n = \min(X_n, x), \quad S'_n = \sum_1^n X'_i,$$

and note that (omitting subscripts) $x \leq \min(S_T, x) \leq S'_T$, implying

$$x \leq E S'_T \leq x + E X'_T \leq 2x.$$

Moreover, by Wald's equation

$$\mathbb{E} S'_T = \mathbb{E} T \cdot \mathbb{E} X'_1 = \mathbb{E} T \cdot \mathbb{E} \min(X, x),$$

yielding (20). \square

Consider the plight of a gambler who makes a series of plays with outcomes $X_n, n \geq 1$, where the $X_n, n \geq 1$, are i.i.d. r.v.s, and who has the option of stopping at any time $n \geq 1$ with the fortune $Y_n = \max_{1 \leq i \leq n} X_i - cn$, $c > 0$. Since the gambler is not clairvoyant his choice of a rule for cessation of play is a stopping time, i.e., his decision to stop at any specific time n must be based only on X_1, \dots, X_n (and not on X_j with $j > n$). Does there exist an **optimal stopping rule**, i.e., one which maximizes $\mathbb{E} Y_T$ over the class of all finite stopping times?

Theorem 6. Let $\{X_n, n \geq 1\}$ be i.i.d. r.v.s with $\mathbb{E}|X_1| < \infty$ and

$$Y_n = \max_{1 \leq i \leq n} X_i - cn, \quad n \geq 1, \quad (23)$$

where c is a positive constant. If T is the finite $\{X_n\}$ -time defined by

$$T = \inf\{n \geq 1 : X_n \geq \beta\}, \quad (24)$$

where β is the unique solution of

$$\mathbb{E}(X_1 - \beta)^+ = c, \quad (25)$$

then $\mathbb{E} T < \infty$ and T is an optimal stopping rule, that is,

$$\mathbb{E} Y_T = \beta \geq \mathbb{E} Y_{T'} \quad (26)$$

for every finite $\{X_n\}$ -time T' for which $\mathbb{E} Y_{T'}$ exists.

PROOF. Since $f(\beta) = \mathbb{E}(X_1 - \beta)^+$ is a decreasing continuous function of β with $\lim_{\beta \rightarrow \infty} f(\beta) = 0$, $\lim_{\beta \rightarrow -\infty} f(\beta) = \infty$, and is strictly decreasing when $f(\beta) > 0$, necessarily (25) has a unique solution β . Set $p = P\{X_1 \geq \beta\}$ and $q = 1 - p$. By (25), $p > 0$. Then

$$P\{T = n\} = pq^{n-1}, \quad P\{T < \infty\} = \sum P\{T = n\} = \frac{p}{1-q} = 1,$$

$$\mathbb{E} T = \sum_1^\infty npq^{n-1} = p \frac{d}{dq} \left(\sum_0^\infty q^n \right) = \frac{p}{(1-q)^2} = \frac{1}{p},$$

and, since $X_j \geq \beta$ on $\{T = j\}$,

$$\mathbb{E} X_T = \sum_1^\infty \mathbb{E} X_j I_{\{T=j\}}$$

$$\begin{aligned}
 &= \sum_1^{\infty} \mathbb{E} X_j I_{\{T \geq j\}} I_{\{X_j \geq \beta\}} \\
 &= \sum_1^{\infty} \mathbb{P}\{T \geq j\} \mathbb{E} X_1 I_{\{X_1 \geq \beta\}} \\
 &= \sum_1^{\infty} [\mathbb{E}(X_1 - \beta)^+ + \beta p] \mathbb{P}\{T \geq j\} \\
 &= (c + \beta p) \mathbb{E} T = \frac{c}{p} + \beta.
 \end{aligned}$$

Moreover,

$$\mathbb{E} Y_T^- = \mathbb{E}(X_T - cT)^- \leq \mathbb{E} X_T^- + c \mathbb{E} T \leq \beta^- + c \mathbb{E} T < \infty,$$

whence $\mathbb{E} Y_T$ exists and

$$\mathbb{E} Y_T = \mathbb{E} \max_{1 \leq j \leq T} X_j - c \mathbb{E} T = \mathbb{E} X_T - \frac{c}{p} = \beta,$$

yielding the first part of (26). Moreover, if T' is any finite $\{X_n\}$ -time for which $\mathbb{E} Y_{T'}$ exists, it may be supposed without loss of generality that $\mathbb{E} Y_{T'} > -\infty$. Since for any $b > \beta$

$$Y_n = \max_{1 \leq i \leq n} X_i - cn \leq b + \sum_1^n [(X_i - b)^+ - c]$$

and $\mathbb{E}[(X_1 - b)^+ - c] < 0$ by (25), Theorem 5.3.2 ensures that

$$\mathbb{E} Y_{T'} \leq b + \mathbb{E} \sum_1^{T'} [(X_i - b)^+ - c] < b.$$

Hence, $\mathbb{E} Y_{T'} \leq \beta$, which is the second portion of (26). \square

The stopping rule T_c in (28) is of especial interest since the finiteness of its moments depends on the value of the parameter c . According to Exercise 5.3.7 (also the application of Corollary 5.3.5) $\mathbb{E} T_c = \infty$ for $c \geq 1$ whereas Example 2 demonstrates under the Lindeberg condition (27) (see Chapter 9 for a more general form) that $\mathbb{E} T_c < \infty$ for $0 < c < 1$. For $\{X_n, n \geq 1\}$ as in Example 2 there exist constants $c_k \downarrow 0$ (c_k is the smallest positive root of the Hermite polynomial of degree $2k$) such that $\mathbb{E} T_c^k(m) < \infty$ for $c < c_k$, all $m \geq 1$ while $\mathbb{E} T_c^k(m) = \infty$ for all large m when $c > c_k$ [B. Brown].

EXAMPLE 2. Let $S_n = \sum_{j=1}^n X_j$ where $\{X_j, j \geq 1\}$ are independent r.v.s with $\mathbb{E} X_n = 0$, $\mathbb{E} X_n^2 = 1$, $n \geq 1$ obeying

$$\sum_{j=1}^n \int_{\{X_j^2 > ej\}} X_j^2 = o(n) \tag{27}$$

and define

$$T_c = T_c(1), T_c(m) = \inf\{n \geq m : |S_n| > cn^{1/2}\}, c > 0, m = 1, 2, \dots \quad (28)$$

Then for all $m \geq 1$, $E T_c(m) < \infty$ for all c in $(0, 1)$.

PROOF. The argument for $m > 1$ requires only minor modifications from that for $m = 1$ which will therefore be supposed. Let $c \in (0, 1)$ and assume that $E T_c = \infty$. If $V = \min(T_c, n)$, clearly $E V < \infty$ and $E V \rightarrow \infty$ as $n \rightarrow \infty$. Thus, $E X_V^2 = o(E V)$ by Lemma 2 and via Theorem 5.3.3

$$\begin{aligned} E V &= E S_V^2 = E S_{V-1}^2 + 2 E S_{V-1} X_V + E X_V^2 \\ &\leq c^2 E V + 2c[E V \cdot E X_V^2]^{1/2} + E X_V^2 \\ &= c^2 E V + o(E V). \end{aligned}$$

Hence, $(1 - c^2) E V \leq o(E V)$ yielding a contradiction as $n \rightarrow \infty$. Thus, $E T_c < \infty$ for $0 < c < 1$. \square

EXERCISES 5.4

1. Verify the second equality of (3).
2. Prove that the stopping rule T of (24) remains optimal when $Y_n = X_n - cn$, $c > 0$.
3. Prove Corollary 3.
4. Prove that if $\{X_n\}$ are i.i.d. with $E X_1 = \mu \in (0, \infty]$ and $N_c = \sup\{n \geq 1 : S_n \leq c\}$, then $N_c/c \xrightarrow{\text{a.s.}} 1/\mu$.
5. Let $\{X_n\}$ be i.i.d. r.v.s with $E X_1 = \mu > 0$ and

$$T = T'_c = \inf\{n \geq 1 : S_n > c\} \quad \text{for } c > 0.$$

Prove that $T/c \xrightarrow{\text{a.s.}} 1/\mu$, a.c. If, moreover, $E X_1^2 < \infty$, then

$$\frac{S_T - \mu E T}{\sqrt{T}} \xrightarrow{P} 0 \quad \text{as } c \rightarrow \infty.$$

6. Prove that if $\{X_n\}$ are i.i.d. with $E X_1 = \mu \in (0, \infty]$ and $T_c = \inf\{n \geq 1 : S_n > cn^\alpha\}$, $c > 0$, $0 < \alpha < 1$, then

$$\frac{T_c^{1-\alpha}}{c} \xrightarrow{\text{a.s.}} \frac{1}{\mu} \quad \text{as } c \rightarrow \infty \quad \text{and } E T_c \sim \left(\frac{c}{\mu}\right)^{1/(1-\alpha)}.$$

7. (Chow-Robbins) Let $\{Y_n, n \geq 1\}$ be positive r.v.s with $\lim Y_n = 1$, a.c., and $\{a_n, n \geq 1\}$ positive constants with $a_n \rightarrow \infty$, $a_n/a_{n-1} \rightarrow 1$. For $c > 0$ define $N = N_c = \inf\{n \geq 1 : Y_n \leq a_n/c\}$. Prove that $P(N < \infty, \lim_c a_N/c = 1) = 1$ and, if $E \sup_{n \geq 1} Y_n < \infty$, then $\lim_{c \rightarrow \infty} E a_N/c = 1$.

8. If $S_n = \sum_1^n X_i$, where $\{X_n, n \geq 1\}$ are nondegenerate i.i.d. r.v.s, $T = \inf\{j \geq 1 : S_j < -a < 0 \text{ or } S_j > b > 0\}$ is of interest in sequential analysis (Wald). Prove that T is a stopping variable with finite moments of all orders. Hint: $P(T > rn) \leq P(|S_{r(j-1)} - S_{r(j-1)}| < a + b, 1 \leq j \leq n)$ for all integers $r > 0$.

9. (Alternative proof that $T_+ < \infty$, a.c., implies $E T_- = \infty$.) Let $\{T^{(j)}, j \geq 1\}$ be copies of T_+ and set $T_n = \sum_{j=1}^n T^{(j)}$. Then $Z = \sum_{i=1}^{\infty} I_{\{T_i < \infty\}}$ = number of times S_n exceeds S_i , $i < n$ (take $S_0 = 0$) and $P\{T_- \geq n + 1\} = P\{S_n > S_i, 0 \leq i < n\} = P\{\bigcup_{i=1}^n [T_i = n]\}$, whence $E T_- - 1 = E Z = \infty$.
10. If $\{X_n, n \geq 1\}$ are i.i.d. with $S_n = \sum_1^n X_i$, then (i) $\overline{\lim} S_n = \infty$, a.c., iff there is a finite $\{X_n\}$ -time T with $E S_T > 0$. (ii) Moreover, when $E X_1^+ = \infty$ $\overline{\lim} S_n/n = \infty$, a.c., iff there exists a finite $\{X_n\}$ -time T with $E S_T > -\infty$.
11. If $\{X_n, n \geq 1\}$ are i.i.d. r.v.s with $E X_1 = \mu > -\infty$ and $S_n = \sum_1^n X_i$, $T_c = \inf\{n \geq 1: S_n > cn\}$, then $E T_c < \infty$ for $c < \mu$ and $E T_c = \infty$ for $c \geq \mu$.
12. Let $\{S_n = \sum_1^n X_i, n \geq 1, S_0 = 0\}$ be a random walk on the line. A real number c is said to be **recurrent** if $P\{|S_n - c| < \varepsilon, \text{i.o.}\} = 1$ for all $\varepsilon > 0$ and is called **possible** if for every $\varepsilon > 0$ there is an integer n such that $P\{|S_n - c| < \varepsilon\} > 0$. Prove that if c is possible and b is recurrent, then $b - c$ is recurrent. Since every recurrent value is clearly possible, the set Q of recurrent values is an additive group. Show that Q is closed. Thus, if Q is nonempty, $Q = (-\infty, \infty)$ or $Q = \{nc: c \neq 0, n = 0, \pm 1, \pm 2, \dots\}$ or $Q = \{0\}$ (the latter only if X is degenerate at 0).

References

- J. H. Abbott and Y. S. Chow, "Some necessary conditions for a.s. convergence of sums of independent r.v.s.", *Bull. Institute Math. Academia Sinica* **1** (1973), 1-7.
- L. E. Baum and M. Katz, "Convergence rates in the law of large numbers," *Trans. Amer. Math. Soc.* **120** (1965), 108-123.
- B. Brown, "Moments of a stopping rule related to the central limit theorem," *Ann. Math. Stat.* **40** (1969), 1236-1249.
- D. L. Burkholder, "Independent sequences with the Stein property," *Ann. Math. Stat.* **39** (1968), 1282-1288.
- Y. S. Chow, "Local convergence of martingales and the law of large numbers," *Ann. Math. Stat.* **36** (1965), 552-558.
- Y. S. Chow, "Delayed sums and Borel summability of independent, identically distributed random variables," *Bull. Inst. Math., Academia Sinica* **1** (1973), 207-220.
- Y. S. Chow and H. Robbins, "On the asymptotic theory of fixed-width sequential confidence intervals for the mean," *Ann. Math. Stat.* **36** (1965), 457-462.
- Y. S. Chow and H. Teicher, "Almost certain summability of i.i.d. random variables," *Ann. Math. Stat.* **42** (1971), 401-404.
- Y. S. Chow, H. Robbins, and D. Siegmund, *Great Expectations: The Theory of Optimal Stopping*, Houghton Mifflin, Boston, 1972.
- Y. S. Chow, H. Robbins, and H. Teicher, "Moments of randomly stopped sums," *Ann. Math. Stat.* **36** (1965), 789-799.
- K. L. Chung, "Note on some strong laws of large numbers," *Amer. Jour. Math.* **69** (1947), 189-192.
- K. L. Chung, *A Course in Probability Theory*, Harcourt Brace, New York, 1968; 2nd ed., Academic Press, New York, 1974.
- K. L. Chung and W. H. J. Fuchs, "On the distribution of values of sums of random variables," *Mem. Amer. Math. Soc.* **6** (1951).
- K. L. Chung and D. Ornstein, "On the recurrence of sums of random variables," *Bull. Amer. Math. Soc.* **68** (1962), 30-32.

- C. Derman and H. Robbins, "The SLLN when the first moment does not exist," *Proc. Nat. Acad. Sci. U.S.A.* **41** (1955), 586–587.
- J. L. Doob, *Stochastic Processes*, Wiley, New York, 1953.
- K. B. Erickson, "The SLLN when the mean is undefined," *Trans. Amer. Math. Soc.* **185** (1973), 371–381.
- W. Feller, "Über das Gesetz der grossen Zahlen," *Acta Univ. Szeged., Sect. Sci. Math.* **8** (1937), 191–201.
- W. Feller, "A limit theorem for random variables with infinite moments," *Amer. Jour. Math.* **68** (1946), 257–262.
- W. Feller, *An Introduction to Probability Theory and its applications*, Vol. 2, Wiley, New York, 1966.
- R. Gundy and D. Siegmund, "On a stopping rule and the central limit theorem," *Ann. Math. Stat.* **38** (1967), 1915–1917.
- C. C. Heyde, "Some renewal theorems with applications to a first passage problem," *Ann. Math. Stat.* **37** (1966), 699–710.
- T. Kawata, *Fourier Analysis in Probability Theory*, Academic Press, New York, 1972.
- H. Kesten, "The limit points of a random walk," *Ann. Math. Stat.* **41** (1970), 1173–1205.
- A. Khintchine and A. Kolmogorov, "Über Konvergenz von Reichen, deren Glieder durch den Zufall bestimmt werden," *Rec. Math. (Mat. Sbornik)* **32** (1924), 668–677.
- A. Kolmogorov, "Über die Summen durch den Zufall bestimmter unabhängiger Größen," *Math. Ann.* **99** (1928), 309–319; **102** (1930), 484–488.
- M. J. Klass, "Properties of optimal extended-valued stopping rules," *Ann. Prob.* **1** (1973), 719–757.
- M. Klass and H. Teicher, "Iterated logarithm laws for random variables barely with or without finite mean," *Ann. Prob.* **5** (1977), 861–874.
- K. Knopp, *Theory and Application of Infinite Series*, Stechert–Hafner, New York, 1928.
- P. Lévy, *Theorie de l'addition des variables aléatoires*, Gauthier–Villars, Paris, 1937; 2nd ed., 1954.
- M. Loève, "On almost sure convergence," *Proc. Second Berkeley Symp. Math. Stat. Prob.*, pp. 279–303, Univ. of California Press, 1951.
- M. Loève, *Probability Theory*, 3rd ed., Van Nostrand, Princeton, 1963; 4th ed., Springer–Verlag, Berlin and New York, 1977–1978.
- J. Marcinkiewicz and A. Zygmund, "Sur les fonctions indépendantes," *Fund. Math.* **29** (1937), 60–90.
- P. Revesz, *The Laws of Large Numbers*, Academic Press, New York, 1968.
- H. Robbins and E. Samuel, "An extension of a lemma of Wald," *J. Appl. Prob.* **3** (1966), 272–273.
- F. Spitzer, "A combinatorial lemma and its applications to probability theory," *Trans. Amer. Math. Soc.* **82** (1956), 323–339.
- C. J. Stone, "The growth of a recurrent random walk," *Ann. Math. Stat.* **37** (1966), 1040–1041.
- H. Teicher, "Almost certain convergence in double arrays," *Z. Wahr. verw. Gebiete* **69** (1985), 331–345.
- A. Wald, "On cumulative sums of random variables," *Ann. Math. Stat.* **15** (1944), 283–296.

6 Measure Extensions, Lebesgue–Stieltjes Measure, Kolmogorov Consistency Theorem

6.1 Measure Extensions, Lebesgue–Stieltjes Measure

A salient underpinning of probability theory is the one-to-one correspondence between distribution functions on R^n and probability measures on the Borel subsets of R^n . Verification of this correspondence involves the notion of measure extension.

Recall that a measure μ on a class \mathcal{A} of subsets of a space Ω is σ -finite if $\Omega = \bigcup_{n=1}^{\infty} \Omega_n$ with $\Omega_n \in \mathcal{A}$, $\mu(\Omega_n) < \infty$, $n \geq 1$. Moreover, if μ and v are set functions on classes \mathcal{G} and \mathcal{H} respectively with $\mathcal{G} \subset \mathcal{H}$ and $\mu(A) = v(A)$ for each $A \in \mathcal{G}$, then v is dubbed an extension of μ to \mathcal{H} while μ is called the restriction of v to \mathcal{G} , the latter being denoted by $\mu = v|_{\mathcal{G}}$.

Theorem 1. *If μ is a measure on a semi-algebra \mathcal{S} of subsets of Ω , there exists a measure extension v of μ to $\sigma(\mathcal{S})$. Moreover, if μ is a probability or σ -finite measure, then v is likewise and the extension is unique.*

PROOF. For each subset $A \subset \Omega$, define

$$v\{A\} = \inf \left\{ \sum_{n=1}^{\infty} \mu\{S_n\}: \bigcup_{n=1}^{\infty} S_n \supset A, S_n \in \mathcal{S}, n \geq 1 \right\}. \quad (1)$$

$$\mathcal{M} = \{D \subset \Omega: v(D \cdot A) + v(D^c \cdot A) = v(A) \text{ for all } A \subset \Omega\}. \quad (2)$$

Clearly, for $A \subset B \subset \Omega$, $0 = v(\emptyset) \leq v(A) \leq v(B)$; moreover, $\Omega \in \mathcal{M}$, and $D^c \in \mathcal{M}$ whenever $D \in \mathcal{M}$.

(i) $\mu = v|_{\mathcal{G}}$.

If $S \in \mathcal{S}$, then by (1), $v\{S\} \leq \mu\{S\} + \mu\{\emptyset\} + \mu\{\emptyset\} + \dots = \mu\{S\}$, while if $\bigcup_1^\infty S_n \supseteq S$, where $S_n \in \mathcal{S}$, $n \geq 1$, then, since μ is a measure on \mathcal{S} , by Corollary 1.5.1 $\mu\{S\} = \mu\{\bigcup_1^\infty S_n\} \leq \sum_1^\infty \mu\{S_n\}$, and so via (1), $\mu\{S\} \leq v\{S\}$.

(ii) v is subadditive, that is, for $A_0 \subset \bigcup_1^\infty A_n \subset \Omega$

$$v\{A_0\} \leq \sum_1^\infty v\{A_n\}. \quad (3)$$

In proving (3) it may be supposed that $v\{A_n\} < \infty$, $n \geq 1$. For any $\varepsilon > 0$ and $n \geq 1$, choose $S_{n,m} \in \mathcal{S}$, $m \geq 1$, with

$$A_n \subset \bigcup_1^\infty S_{n,m}, \quad \sum_1^\infty \mu\{S_{n,m}\} \leq v\{A_n\} + \frac{\varepsilon}{2^n}.$$

Since $\bigcup_{n,m} S_{n,m} \supseteq \bigcup_n A_n \supseteq A_0$, by (1)

$$v\{A_0\} \leq \sum_{n,m} \mu\{S_{n,m}\} \leq \sum_1^\infty v\{A_n\} + \varepsilon,$$

and, letting $\varepsilon \rightarrow 0$, (3) obtains.

(iii) $\mathcal{S} \subset \mathcal{M}$.

Let $S \in \mathcal{S}$ and $A \subset \Omega$. By (ii) and the definition of \mathcal{M} it suffices to verify that

$$v\{A\} \geq v\{S \cdot A\} + v\{S^c \cdot A\}, \quad (4)$$

and in so doing it may clearly be supposed that $v\{A\} < \infty$.

For any $\varepsilon > 0$, choose $S_n \in \mathcal{S}$, $n \geq 1$, such that $\bigcup_1^\infty S_n \supseteq A$ and

$$v\{A\} + \varepsilon \geq \sum_1^\infty \mu\{S_n\}. \quad (5)$$

Since \mathcal{S} is a semi-algebra, for each $n \geq 1$ there exists a finite partition of $S^c S_n$ in \mathcal{S} , say $(S_{n,m}, m = 1, \dots, m_n)$. Then $(S \cdot S_n, S_{n,m}, m = 1, 2, \dots, m_n)$ is a partition of S_n in \mathcal{S} , and by additivity of μ on \mathcal{S}

$$\mu\{S_n\} = \mu\{S \cdot S_n\} + \sum_{m=1}^{m_n} \mu\{S_{n,m}\}, \quad n \geq 1. \quad (6)$$

But $\bigcup_{n=1}^\infty \bigcup_{m=1}^{m_n} S_{n,m} = \bigcup_{n=1}^\infty S^c \cdot S_n \supseteq A \cdot S^c$ and $\bigcup_1^\infty S S_n \supseteq SA$, whence by (1)

$$\sum_{n=1}^\infty \sum_{m=1}^{m_n} \mu\{S_{n,m}\} \geq v\{A \cdot S^c\}, \quad \sum_{n=1}^\infty \mu\{S \cdot S_n\} \geq v\{S \cdot A\}. \quad (7)$$

Consequently, from (5), (6) and (7)

$$v\{A\} + \varepsilon \geq \sum_1^\infty \mu\{S \cdot S_n\} + \sum_{n=1}^\infty \sum_{m=1}^{m_n} \mu\{S_{n,m}\} \geq v\{S \cdot A\} + v\{S^c \cdot A\},$$

which is tantamount to (4) as $\varepsilon \rightarrow 0$.

(iv) \mathcal{M} is an algebra and for every $D \in \mathcal{M}$ and $A \subset \Omega$ and finite partition $\{D_n, n = 1, 2, \dots, m\}$ of D in \mathcal{M} ,

$$v\{A \cdot D\} = \sum_1^m v\{A \cdot D_n\}. \quad (8)$$

Now for $A \subset \Omega$ and $D_i \in \mathcal{M}, i = 1, 2$,

$$\begin{aligned} v\{A\} &= v\{A \cdot D_1\} + v\{A \cdot D_1^c\} \\ &= v\{A \cdot D_1 D_2\} + v\{AD_1 D_2^c\} + v\{AD_1^c D_2\} \\ &\quad + v\{AD_1^c D_2^c\}. \end{aligned} \quad (9)$$

Replacing A by $A(D_1 \cup D_2)$ in (9),

$$v\{A(D_1 \cup D_2)\} = v\{A \cdot D_1 D_2\} + v\{AD_1 D_2^c\} + v\{AD_1^c D_2\}, \quad (10)$$

whence from (9) and (10)

$$v\{A\} = v\{A(D_1 \cup D_2)\} + v\{AD_1^c D_2^c\},$$

and so, via the definition, $D_1 \cup D_2 \in \mathcal{M}$. Since, as noted at the outset, \mathcal{M} is closed under complements, \mathcal{M} is an algebra, and, moreover, if $D_1 D_2 = \emptyset$, (10) yields

$$v\{A(D_1 \cup D_2)\} = v\{AD_1\} + v\{AD_2\},$$

which is precisely (8) when $m = 2$. The general statement of (8) follows by induction.

(v) \mathcal{M} is a σ -algebra, and for every $D \in \mathcal{M}$ and $A \subset \Omega$ and σ -partition $\{D_n, n \geq 1\}$ of D in \mathcal{M} ,

$$v\{A \cdot D\} = \sum_1^\infty v\{A \cdot D_n\}. \quad (11)$$

Let $\{D_n, n \geq 1\}$ be the stated σ -partition of D and set $E_n = \bigcup_1^n D_i$. By (iv), $E_n \in \mathcal{M}$ for every positive integer n , whence for any $A \subset \Omega$,

$$\begin{aligned} v\{A\} &= v\{A E_n\} + v\{A E_n^c\} \geq v\{A \cdot E_n\} + v\{AD^c\} \\ &= \sum_1^n v\{A \cdot D_i\} + v\{A \cdot D^c\}. \end{aligned}$$

Hence, by (3)

$$v\{A\} \geq \sum_1^\infty v\{A \cdot D_i\} + v\{A \cdot D^c\} \geq v\{A \cdot D\} + v\{A \cdot D^c\}, \quad (12)$$

and so equality holds throughout (12), yielding (11) upon replacement of A by AD . Moreover, if $\{D_j, j \geq 1\}$ is any sequence of disjoint sets of \mathcal{M} and $D = \bigcup_1^\infty D_j$, (12) remains intact, whence it is clear that (see Exercise 1.3.2) \mathcal{M} is a σ -algebra. Clearly, if μ is finite or σ -finite, v inherits the characteristic.

It follows directly via (i), (iii), and (v) that (vi) $\mathcal{M} \supset \sigma(\mathcal{S})$, v is a measure on \mathcal{M} , and $\mu = v|_{\mathcal{S}}$.

Finally, to prove uniqueness, let v^* be any measure extension of μ to $\sigma(\mathcal{S})$ and define

$$\mathcal{E} = \{E : E \in \sigma(\mathcal{S}) \text{ and } v(E) = v^*(E)\}.$$

If $\mu\{\Omega\} < \infty$, then $v\{\Omega\} = v^*\{\Omega\} < \infty$ and it is easily verified that \mathcal{E} is a λ -class containing the π -class \mathcal{S} , whence $v = v^*$ on $\sigma(\mathcal{S})$. If μ is merely σ -finite on \mathcal{S} , there exist sets $\Omega_n \in \mathcal{S}$ with $\bigcup_1^\infty \Omega_n = \Omega$ and $\mu\{\Omega_n\} < \infty$, $n \geq 1$. Then, as just seen, $v = v^*$ on $\Omega_n \cap \sigma(\mathcal{S})$, $n \geq 1$, and so $v = v^*$ on $\sigma(\mathcal{S})$. \square

The set function v defined by (1) on the σ -algebra of all subsets of Ω is called the **outer measure induced by μ** , while the σ -algebra \mathcal{M} described in (2) is termed the **σ -algebra of v -measurable sets**.

Any measure μ on a σ -algebra \mathcal{A} is called **complete** if $B \subset A \in \mathcal{A}$ and $\mu\{A\} = 0$ imply $B \in \mathcal{A}$ (and necessarily $\mu\{B\} = 0$). A **complete measure space** is a measure space whose measure is complete.

The outer measure v stipulated in (1) defines a complete measure extension of μ to \mathcal{M} , the σ -algebra of all v -measurable sets (Exercise 2).

As an important application of Theorem 1, the special cases of Lebesgue-Stieltjes and Lebesgue measure on the line will be considered. Let $R = [-\infty, \infty]$ and \mathcal{S} denote the semi-algebra of all intervals of the form $[a, b)$, $-\infty \leq a \leq b \leq \infty$ and $R, \{\infty\}$ in addition. A class of measures known as Lebesgue-Stieltjes measures will be generated on \mathcal{S} via monotone functions.

For any mapping f of $R = [-\infty, \infty]$ into R , set

$$f(t+) = \lim_{s \rightarrow t+} f(s) = \lim_{t < s \rightarrow t} f(s), \quad t \in [-\infty, \infty], \quad f(\infty+) = f(\infty),$$

$$f(t-) = \lim_{s \rightarrow t-} f(s) = \lim_{t > s \rightarrow t} f(s), \quad t \in (-\infty, \infty], \quad f(-\infty-) = f(-\infty),$$

when these limits exist. If $f(t) = f(t-)$, then f is said to be continuous from the left or **left continuous at $t \in R$** . A function which is left continuous at all points of some set $T \subset R$ is called left continuous on T and when $T = R$, simply **left continuous**. Similarly, $f(t) = f(t+)$ defines right continuity at t and the analogous terms are employed.

If f is a function with $f(t-)$ existing for each $t \in R$, then $g(t) = f(t-)$ is left continuous (Exercise 3). In particular, if f is a monotone function on R , $f(t-)$ exists and $g(t) = f(t-)$ is left continuous. Since the set of discontinuities of a monotone function is countable and two left-continuous functions are identical if they coincide except for a countable set, every monotone function m on R defines a unique left-continuous function $F = F_m$ via $F(t) = m(t-)$.

Lemma 1. *Let F be a nondecreasing, left continuous function on $R = [-\infty, \infty]$ with $F(-\infty) = F(-\infty+)$ and $|F(t)| < \infty$, $|t| < \infty$. If*

$$\begin{aligned} \mu([a, b]) &= F(b) - F(a), \quad -\infty \leq a \leq b \leq \infty, \\ \mu(\{\infty\}) &= 0 = \mu(\emptyset), \quad \mu(R) = F(\infty) - F(-\infty), \end{aligned} \tag{13}$$

then μ is a measure on the semi-algebra \mathcal{S} .

PROOF. Clearly, μ is nonnegative and additive on \mathcal{S} . To verify σ -additivity, consider $S \in \mathcal{S}$ with $S \neq \{\infty\}$ or \emptyset and let $\{S_n, n \geq 1\}$ be a σ -partition of S in \mathcal{S} . By Corollary 1.5.1

$$\mu\{S\} \geq \sum_1^{\infty} \mu\{S_n\}. \quad (14)$$

In proving the reverse inequality, since $\mu\{R\} = \mu\{[-\infty, \infty)\}$, it may be supposed that $S = [a, b]$, $-\infty \leq a < b \leq \infty$. Moreover, in view of the hypothesized equality $F(-\infty) = F(-\infty+)$ and left continuity, it suffices to establish that

$$\mu\{[c, d]\} \leq \sum_1^{\infty} \mu\{S_n \cdot [c, d]\} \quad \text{whenever } a \leq c < d < b \leq \infty, c \neq -\infty. \quad (15)$$

Since $\infty \notin [a, b]$, necessarily $S_n \neq \{\infty\}$, $n \geq 1$ whence $S_n \cdot [c, d] = [a_n, b_n]$, where $c \leq a_n \leq b_n \leq d$, $n \geq 1$.

For any $\varepsilon > 0$, set $J_n = (a_n - \delta_n, b_n)$ where $\delta_n > 0$ satisfies $F(a_n) - F(a_n - \delta_n) < \varepsilon/2^n$, $n \geq 1$ via left continuity. Then $[c, d] = \bigcup_1^{\infty} [a_n, b_n] \subset \bigcup_1^{\infty} J_n$ whence by the Heine–Borel theorem $[c, d] \subset \bigcup_1^m J_{n_k} \subset \bigcup_1^m [a_{n_k} - \delta_{n_k}, b_{n_k})$ for some finite set of integers n_1, \dots, n_m . By Corollary 1.5.1

$$\begin{aligned} \mu\{[c, d]\} &\leq \sum_1^m (F(b_{n_k}) - F(a_{n_k} - \delta_{n_k})) \leq \sum_1^m (F(b_{n_k}) - F(a_{n_k}) + \varepsilon 2^{-n_k}) \\ &\leq \sum_1^m (F(b_{n_k}) - F(a_{n_k})) + \varepsilon \leq \sum_1^{\infty} \mu\{S_k\} + \varepsilon. \end{aligned}$$

Thus, (15) obtains as $\varepsilon \rightarrow 0$, and μ is a measure on \mathcal{S} . \square

Theorem 2. Any nondecreasing, finite function m on $(-\infty, \infty)$ determines a complete measure v_m on the σ -algebra \mathcal{M}_m of all v_m -measurable subsets of $R = [-\infty, \infty]$ with

$$\begin{aligned} v_m\{[a, b]\} &= m(b-) - m(a-), \quad -\infty < a \leq b < \infty, \\ v_m(\{\infty\}) &= v_m(\{-\infty\}) = 0. \end{aligned} \quad (16)$$

Moreover, $\mathcal{M}_m \supset \mathcal{B}$ the σ -algebra \mathcal{B} of all Borel sets of R and v_m is unique on \mathcal{B} .

PROOF. Set $F(t) = m(t-)$ for $-\infty < t \leq \infty$ and $F(-\infty) = m(-\infty+)$. Then F is defined and left continuous on R with $F(-\infty+) = F(-\infty)$. By Lemma 1, the set function μ defined on \mathcal{S} by (13) is a measure thereon, whence Theorem 1 guarantees the existence of a measure extension \tilde{v}_m of μ to $\mathcal{M}_m \supset \mathcal{B} = \sigma(\mathcal{S})$, where \tilde{v}_m and \mathcal{M}_m are defined by (1) and (2) respectively. According to Exercise 2, \tilde{v}_m is complete on \mathcal{M}_m .

Hence, $v_m\{A\} = \tilde{v}_m\{A - \{-\infty\}\}$, $A \in \mathcal{M}_m$ is a complete measure on \mathcal{M}_m satisfying (16).

To verify uniqueness on \mathcal{B} , let v^* be any other measure on \mathcal{B} satisfying

(16). Then

$$\begin{aligned} v^*(\{\infty\}) &= v^*(\{-\infty\}) = v_m(\{\infty\}) = v_m(\{-\infty\}) \\ v^*([a, b]) &= m(b-) - m(a-) = v_m([a, b]), \quad -\infty < a \leq b < \infty, \end{aligned}$$

and for $-\infty < b < \infty$

$$\begin{aligned} v^*([-\infty, b]) &= v^*((-\infty, b]) = \lim_{a \rightarrow -\infty} v^*([a, b]) = \lim v_m([a, b]) \\ &= v_m([-\infty, b]); \end{aligned}$$

similarly, for $-\infty < a < \infty$

$$v^*([a, \infty)) = v_m([a, \infty))$$

Thus, taking $a = b$,

$$v^*(R) = v^*([-\infty, \infty)) = v_m([-\infty, \infty)) = v_m(R).$$

Hence, v_m and v^* coincide on \mathcal{S} whence by Theorem 1 these σ -finite measures also coincide on $\mathcal{B} = \sigma(\mathcal{S})$. \square

For any finite, nondecreasing function m on $(-\infty, \infty)$ the corresponding measure v_m is called the **Lebesgue–Stieltjes measure determined by m** . Similarly, the complete measure space (R, \mathcal{M}_m, v_m) is referred to as the **Lebesgue–Stieltjes measure space determined by m** .

In the important special case $m(t) = t$ for $t \in (-\infty, \infty)$, the corresponding measure $v = v_m$ is the renowned **Lebesgue measure**, generalizing the notion of length; the sets of $\mathcal{M} = \mathcal{M}_m$ are called **Lebesgue-measurable sets** and (R, \mathcal{M}, v) is the **Lebesgue measure space** (of R).

EXERCISES 6.1

- Let μ be a measure on a σ -algebra \mathcal{A} and define $\bar{\mathcal{A}} = \{A \Delta N : A \in \mathcal{A}, N \subset B \in \mathcal{A}, \mu\{B\} = 0\}$. Then $\bar{\mathcal{A}}$ is a σ -algebra, $\bar{\mathcal{A}} \supset \mathcal{A}$, and $\bar{\mu}$ is a complete measure on $\bar{\mathcal{A}}$, where $\bar{\mu}(A \Delta N) = \mu\{A\}$ for $A \Delta N \in \bar{\mathcal{A}}$.
- Prove that the extension v (as defined in Theorem 1) of a measure μ on \mathcal{S} to the σ -algebra \mathcal{M} of all v -measurable sets is complete.
- Let f map $R = [-\infty, \infty]$ into R with $f(t-)$ existing for every $t \in R$. Then $g(t) = f(t-)$, $t \in R$, is left continuous, i.e., $g(t) = g(t-)$, $t \in R$.
- If v_m is the Lebesgue–Stieltjes measure determined by a finite nondecreasing function m on $(-\infty, \infty)$ and $F(t) = m(t-)$, $-\infty \leq t \leq \infty$, prove that for $-\infty < a < b < \infty$

$$\begin{aligned} v([a, b]) &= F(b-) - F(a-), & v((a, b]) &= F(b+) - F(a+), \\ v([a, b]) &= F(b+) - F(a-), & v((a, b)) &= F(b) - F(a+), \\ v(\{\infty\}) &= v(\{-\infty\}) = 0, & v(\{a\}) &= F(a+) - F(a-). \end{aligned}$$

- If $m(t)$ is as in Exercise 4 and $G(t) = m(t+)$, $-\infty \leq t \leq \infty$, then the nondecreasing, right continuous function G determines a measure $\tilde{\mu}$ on the semi-algebra $\tilde{\mathcal{S}}$ of all

- finite or infinite intervals of the form $(a, b]$, $a < b$, and also $\emptyset, R = [-\infty, \infty], \{-\infty\}$ via $\tilde{\mu}\{(a, b]\} = G(b) - G(a)$, $\tilde{\mu}\{(-\infty)\} = \mu\{\emptyset\} = 0$, $\tilde{\mu}\{R\} = G(\infty) - G(-\infty)$.
6. There is a 1-1 correspondence between d.f.s and probability measures on the Borel sets of the line.
 7. If (R, \mathcal{M}, v) is the Lebesgue measure space (of R), $\Omega = [0, 1]$, $\mathcal{A} = \Omega \cdot \mathcal{M}$, and $P = v|_{\mathcal{A}}$, then (Ω, \mathcal{A}, P) is a probability space.
 8. If X is a r.v. with d.f. F , then $P\{X^+ = 0\} = F(0+)$, $P\{X^+ < x\} = F(x)$, $x > 0$, and $P\{X^- = 0\} = 1 - F(0)$, $P\{X^- < x\} = 1 - F(-x+)$, $x > 0$. Find the d.f. of $|X|$.
 9. Give an example to show that the uniqueness assertion of Theorem 1 is not true without the restriction of σ -finiteness on \mathcal{S} . Hint: Take $\Omega = \{r: r \text{ rational}, 0 \leq r < 1\}$, $\mathcal{S} = \{\Omega \cdot [a, b]: 0 \leq a \leq b \leq 1\}$, $\mu(\emptyset) = 0$ and $\mu(A) = \infty$ if $\emptyset \neq A \in \mathcal{S}$, $v(A) = \text{number of elements in } A$ for $A \in \sigma(\mathcal{S})$, $v^*(A) = 2v(A)$ for $A \in \sigma(\mathcal{S})$.
 10. If (R, \mathcal{M}, v) is the Lebesgue measure space of the real line and E is a Lebesgue measurable set, so is $E + x = \{y + x: y \in E\}$ and, moreover, $v\{E + x\} = v\{E\}$ for every $x \in (-\infty, \infty)$.
 11. For any real x, y consider the equivalence relation $x \sim y$ if $x - y = r = \text{rational}$. Let the subset E of $[0, 1]$ contain exactly one point of each equivalence class. Then E is a non-Lebesgue-measurable set. Hint: (i) if $x \in (0, 1)$, then $x \in E + r$ for some r in $(-1, 1)$. (ii) $(E + r) \cap (E + s) = \emptyset$ for distinct rationals r, s . Thus, if E is Lebesgue measurable so is $S = \bigcup_{r \in F}(E + r)$, where F is the set of rationals in $(-1, 1)$ via Exercise 10. Then $S \subset (-1, 2)$, whence $3 \geq v\{S\} = \sum_1^\infty v\{E\}$, implying $v\{S\} = \sum v\{E\} = 0$. However, by (i), $[0, 1] \subset \bigcup(E + r) = S$, so that $v\{S\} \geq 1$, a contradiction.

6.2 Integration in a Measure Space

Let (S, Σ, μ) constitute a measure space, that is, S is an arbitrary nonempty set, Σ is a σ -algebra of subsets of S , and μ is a measure on Σ . In the event that μ is σ -finite, (S, Σ, μ) will be termed a σ -finite measure space. Any property that obtains except on a set of μ -measure zero will be said to hold **almost everywhere** (abbreviated a.e.).

Let X be a Σ -measurable function, that is, a mapping from S to $R = [-\infty, \infty]$ for which $X^{-1}(\mathcal{B}) \subset \Sigma$, where \mathcal{B} is the class of Borel subsets of R . Then, paralleling the approach in Chapter 4, the integral

$$\mathbb{E} X = \int_S X d\mu$$

may be defined as follows:

(i) If $X \geq 0$, a.e., and $\mu\{X = \infty\} > 0$, then $\mathbb{E} X = \infty$, while if $X \geq 0$, a.e., and $\mu\{X = \infty\} = 0$,

$$\mathbb{E} X = \lim_{n \rightarrow \infty} \sum_{i=1}^{\infty} \frac{i}{2^n} \mu \left\{ \frac{i}{2^n} < X \leq \frac{i+1}{2^n} \right\}. \quad (1)$$

(ii) In general, if either $E X^+ < \infty$ or $E X^- < \infty$,

$$E X = E X^+ - E X^-, \quad (2)$$

in which case $E X$ is said to exist, denoted by $|E X| \leq \infty$. If $|E X| < \infty$, that is, $E X$ exists and is finite, X is called an **integrable function**, or, simply, **integrable**.

If $|E X| \leq \infty$, the indefinite integral of X is defined by

$$\int_A X d\mu = E X I_A, \quad A \in \Sigma.$$

As in Chapter 4, it is easy to verify that the limit in (1) always exists and that:

$$\begin{aligned} 0 \leq E X \leq \infty &\quad \text{if } X \geq 0, \text{ a.e.,} & E 1 = \mu(S); \\ |X| < \infty, \text{ a.e.,} &\quad \text{if } E|X| < \infty; & X = 0, \text{ a.e.,} \quad \text{iff } E|X| = 0; \\ E X = E \cdot Y &\quad \text{if } X = Y, \text{ a.e., and } |E X| \leq \infty. \end{aligned}$$

From (1) and (2) it is readily seen that if (S, Σ_i, μ_i) , $i = 1, 2$, are measure spaces with $\mu_1 = \mu_2|_{\Sigma_1}$, and X is a Σ_1 -measurable function, then

$$\int_S X d\mu_2 = \int_S X d\mu_1 \quad (3)$$

in the sense that if one of the integrals exists, so does the other and the two are equal.

Associated with any measure space (S, Σ, μ) are the spaces $\mathcal{L}_p = \mathcal{L}_p(S, \Sigma, \mu)$, $p > 0$, of all measurable functions X for which $E|X|^p < \infty$. For any measurable function X , and especially when $X \in \mathcal{L}_p$, the \mathcal{L}_p -norm of X is defined by

$$\|X\|_p = E^{1/p}|X|^p. \quad (4)$$

Let $\{X_n, n \geq 1\}$ be a sequence of measurable functions on the measure space (S, Σ, μ) . If $\mu\{\lim X_n \neq \overline{\lim}_n X\} = 0$, set $X = \overline{\lim}_n X_n$ whence $\lim X_n = X$, a.e., denoted by $X_n \xrightarrow{\text{a.e.}} X$. If X is finite a.e., X_n will be said to **converge a.e.**, denoted by $X_n \xrightarrow{\text{a.e.}} X$, finite. Alternatively, if X_n is finite a.e., $n \geq 1$, then X_n **converges in measure** to X , written $X_n \xrightarrow{\mu} X$, if $\lim \mu\{|X_n - X| > \varepsilon\} = 0$ for all $\varepsilon > 0$. These are obvious analogues of a.c. convergence and convergence in probability on a probability space, but the correspondence is not without limitation (see Exercise 6.2.4). In case $\|X_n - X\|_p = o(1)$, the earlier notation $X_n \xrightarrow{\mathcal{F}_p} X$ will be employed.

In dealing with the basic properties of the integral in Chapter 4, the proof of Theorem 4.1.1(i) utilized the fact that for any nonnegative integrable r.v. X on a probability space (Ω, \mathcal{F}, P)

$$E X = \lim_{n \rightarrow \infty} \sum_{i=0}^{\infty} \frac{i+1}{2^n} P\left\{\frac{i}{2^n} < X \leq \frac{i+1}{2^n}\right\}. \quad (5)$$

The counterpart of (5) in the case of a nonnegative, integrable function X on a measure space (S, Σ, μ) is

$$E X = \lim_{n \rightarrow \infty} \sum_{i=1}^{\infty} \frac{i+1}{2^n} \mu \left\{ \frac{i}{2^n} < X \leq \frac{i+1}{2^n} \right\}. \quad (6)$$

To prove (6), note that, setting $s_n = \sum_{i=1}^{\infty} (i/2^n) \mu(i/2^n < X \leq (i+1)/2^n)$,

$$s_{2n} - s_n \geq \sum_{i=1}^{2^n-1} \frac{i}{2^{2n}} \mu \left\{ \frac{i}{2^{2n}} < X \leq \frac{i+1}{2^{2n}} \right\} \geq 2^{-2n} \mu \{ 2^{-2n} < X \leq 2^{-n} \} = o(1)$$

since $E X = \lim_{n \rightarrow \infty} s_n$. Moreover, $2^{-2n} \mu \{ X > 2^{-n} \} \leq 2^{-n} E X = o(1)$, so that

$$2^{-2n} \mu \{ X > 2^{-2n} \} = o(1),$$

whence

$$2^{-(2n-1)} \mu \{ X > 2^{-(2n-1)} \} \leq 2 \cdot 2^{-2n} \mu \{ X > 2^{-2n} \} = o(1).$$

Consequently, the difference of the right and left sides of (6), viz., $\lim_{n \rightarrow \infty} 2^{-n} \mu \{ X > 2^{-n} \}$, is zero.

It may be noted via (6) or (1) that if X is a nonnegative, integrable function on (S, Σ, μ) , then μ is σ -finite on $S \{ X > 0 \}$.

Theorem 1. Let X, Y, X_n, Y_n be measurable functions on the measure space (S, Σ, μ) .

- i. If $X = \sum_1^{\infty} x_n I_{A_n} \geq 0$, where $\{A_n, n \geq 1\}$ are disjoint measurable sets, then

$$E X = \sum_1^{\infty} x_n \mu \{ A_n \} \quad \text{where } 0 \cdot \infty = \infty \cdot 0 = 0. \quad (7)$$

- ii. (a) If $X \geq 0$, a.e., there exist elementary functions X_n with $0 \leq X_n \uparrow X$, a.e., and $E X_n \uparrow E X$. (b) If, moreover, $X < \infty$, a.e., then $X - X_n \leq 2^{-n}$, $n \geq 1$, a.e., is attainable.

- iii. If $X \geq Y \geq 0$, a.e., then $E X \geq E Y$ and for $0 < a < \infty$

$$E X \geq a \mu \{ X \geq a \}, \quad (8)$$

- iv. If $0 \leq X_n \uparrow X$, a.e., then $E X_n \uparrow E X$ (monotone convergence).

- v. If $E X, E Y$, and $E X + E Y$ are defined, then so is $E(X + Y)$ and

$$E(X + Y) = E X + E Y.$$

- vi. If $X \geq Y$, a.e., $|E X| \leq \infty$, $|E Y| \leq \infty$, then $E X \geq E Y$.

- vii. If $E X, E Y, a E X + b E Y$ are defined for finite a, b , then

$$E(aX + bY) = a E X + b E Y. \quad (9)$$

- viii. X is integrable iff $|X|$ is integrable and if $X \geq 0$, $Y \geq 0$, a.e., then $E(X + Y)^p \leq 2^p(E X^p + E Y^p)$, $p > 0$.

PROOF. The argument follows that of Theorem 4.1.1 with two modifications. Firstly, in proving (i) write $I_{ni} = (i/2^n, (i+1)/2^n]$, $\mu_{ni} = \mu\{X' \in I_{ni}\}$, and replace (13) of Theorem 4.1.1 by

$$\sum_{i=1}^{\infty} \frac{i+1}{2^n} \mu_{ni} \geq \sum_{i=1}^{\infty} \sum_{j: x'_j \in I_{ni}} x'_j \mu\{A_j\} = \sum_{j: x'_j > 2^{-n}} x'_j \mu\{A_j\} \xrightarrow{n \rightarrow \infty} \sum_{j=1}^{\infty} x'_j \mu\{A_j\}.$$

Consequently, $E X = \sum_{j=1}^{\infty} x'_j \mu\{A_j\}$ via (6).

Secondly, in the proof of (iv), i.e., monotone convergence, replace (15), (16) of Section 4.1 by

$$\lim_{m \rightarrow \infty} E X_m \geq \sum_{i=1}^k \frac{i}{2^n} \lim_{m \rightarrow \infty} \mu\{X_m \in I_{ni}\} \geq \sum_{i=1}^k \frac{i}{2^n} \mu\{X \in I_{ni}\} > a,$$

utilizing Exercise 1.5.2. \square

The next theorem incorporates analogues of Lemma 4.2.1, Corollaries 4.2.2 and 4.2.3, and Theorem 4.3.2.

Theorem 2. Let $\{X_n, n \geq 1\}$ be a sequence of measurable functions on the measure space (S, Σ, μ) .

i. If $X_n \geq 0$, a.e., $n \geq 1$, then

$$E \liminf_{n \rightarrow \infty} X_n \leq \liminf_{n \rightarrow \infty} E X_n \quad (\text{Fatou}). \quad (10)$$

ii. If $|X_n| \leq Y$, a.e., where $E Y < \infty$ and either $X_n \xrightarrow{a.e.} X$ or $X_n \xrightarrow{a.s.} X$, then

$$E|X_n - X| \rightarrow 0 \quad (\text{Lebesgue dominated convergence theorem}). \quad (11)$$

iii. If $|X|^p \in \mathcal{L}_1$, $|Y|^{p'} \in \mathcal{L}_1$, where $p > 1$, $p' > 1$, $(1/p) + (1/p') = 1$, then $XY \in \mathcal{L}_1$ and

$$E|XY| \leq \|X\|_p \|Y\|_{p'} \quad (\text{Hölder inequality}). \quad (12)$$

iv. If $|E X| \leq \infty$, the indefinite integral $v\{A\} = \int_A X d\mu$ is a σ -additive set function on Σ .

PROOF. The proofs of Hölder's inequality and (iv) are identical with those of Chapter 4, while the argument in Fatou's lemma is precisely that used for X'_n in Theorem 4.2.2(ii). Apropos of (ii), suppose first that $X_n \xrightarrow{a.e.} X$. The hypothesis ensures that $X_n \in \mathcal{L}_1$, $X \in \mathcal{L}_1$, whence by (10)

$$E(Y \pm X) = E \liminf_{n \rightarrow \infty} (Y \pm X_n) \leq \liminf_{n \rightarrow \infty} E(Y \pm X_n),$$

implying $E X \leq \liminf_{n \rightarrow \infty} E X_n \leq \overline{\lim}_{n \rightarrow \infty} E X_n \leq E X$, that is, $|X_n| \leq Y \in \mathcal{L}_1$ and $X_n \xrightarrow{a.e.} X$ imply

$$E X = \lim_{n \rightarrow \infty} E X_n. \quad (13)$$

Since $0 \leq |X_n - X| \leq 2Y$, (13) ensures that $E|X_n - X| \rightarrow 0$.

On the other hand, if $X_n \xrightarrow{P} X$, then (Exercise 3.3.13) every subsequence of X_n , say X'_n , satisfies $\sup_{m > n} \mu\{|X'_m - X'_n| > \varepsilon\} = o(1)$, all $\varepsilon > 0$, and hence has a further subsequence, say X_{n_k} , with $X_{n_k} \xrightarrow{a.e.} Y_0$, finite. By the portion already proved, $E|X_{n_k} - Y_0| = o(1)$, whence $X_{n_k} \xrightarrow{P} Y_0$ via (8). Thus $X = Y_0$, a.e., and $E|X_{n_k} - X| = o(1)$. Consequently, every subsequence of $E|X_n - X|$ has a further subsequence converging to zero and so $E|X_n - X| = o(1)$. \square

Although most properties of a probability space carry over to a measure space, Jensen's inequality and the usefulness of uniform integrability are conspicuous exceptions (see Exercises 12, 13, and 4).

Let $(R, \mathcal{M}_m, v_m^*)$ be the Lebesgue-Stieltjes measure space determined by a finite, nondecreasing function m on $(-\infty, \infty)$ and let X be a Borel function on R with $|E X| \leq \infty$. Then

$$\int_A X dm \quad \text{or} \quad \int_A X(t) dm(t), \quad A \in \mathcal{M}_m,$$

is just an alternative notation for the indefinite Lebesgue-Stieltjes integral $\int_A X dv_m^*$, $A \in \mathcal{M}_m$, that is, for the indefinite integral on the Lebesgue-Stieltjes measure space determined by m . According to (3), if $v_m = v_m^*|_{\mathcal{B}}$, then

$$\int_A X dm = \int_A X dv_m^* = \int_A X dv_m, \quad A \in \mathcal{B},$$

since X is a Borel function. Consequently, when dealing with Borel sets B , the Lebesgue-Stieltjes integral $\int_B X dm$ may be envisaged as being defined either in the Lebesgue-Stieltjes measure space $(R, \mathcal{M}_m, v_m^*)$ or in the Borel measure space (R, \mathcal{B}, v_m) . Since $v_m\{\{\infty\}\} = v_m\{\{-\infty\}\} = 0$, the integral $\int_{-\infty}^{\infty} X dm$ may be unambiguously defined as

$$\int_{(-\infty, \infty)} X dm = \int_{(-\infty, \infty)} X dm = \int_{(-\infty, \infty)} X dm = \int_{(-\infty, \infty)} X dm.$$

However, if $-\infty < a < b < \infty$, $\int_a^b X dm$ is not, in general, well defined since

$$\begin{aligned} \int_{[a, b]} X dm &= [m(a+) - m(a-)]X(a) \\ &\quad + [m(b+) - m(b-)]X(b) + \int_{(a, b)} X dm \end{aligned}$$

via additivity of the indefinite integral and Exercise 6.1.4. On the other hand, if a and b are **continuity points** of m , i.e., $m(b+) = m(b-)$ and $m(a+) = m(a-)$, then $\int_a^b X dm$ may be interpreted as the common value

$$\int_{[a, b]} X dm = \int_{(a, b)} X dm = \int_{(a, b)} X dm = \int_{[a, b]} X dm.$$

Thus, only when a and b are continuity points (including $\pm\infty$) of m will the notation $\int_a^b X dm$ be utilized for a Lebesgue-Stieltjes integral.

As in the case of Riemann integrals, if $a < b$,

$$\int_b^a X dm = - \int_a^b X dm$$

by fiat (under the prior proviso).

In the important special case $m(t) = t$, $-\infty < t < \infty$, the Lebesgue-Stieltjes integral $\int_A X dm$ is denoted by $\int_A X(t)dt$ and is called the **Lebesgue integral**.

Let X be a measurable transformation from a measure space (S, Σ, μ) to a measurable space (T, \mathcal{A}) . Then, as in Section 1.6, the measure μ induces a measure v on \mathcal{A} via

$$v\{A\} = \mu\{X^{-1}(A)\}, \quad A \in \mathcal{A}. \quad (14)$$

In fact, $v\{\emptyset\} = 0$, $v\{T\} = \mu\{S\}$, and v is σ -additive since if $\{A_n, n \geq 1\}$ are disjoint sets of \mathcal{A} , then $X^{-1}(A_n)$, $n \geq 1$, are disjoint sets of Σ and $X^{-1}(\bigcup_1^\infty A_n) = \bigcup_1^\infty X^{-1}(A_n)$. The measure v , induced by μ and the measurable transformation X , will be denoted by μX^{-1} . The next result might well be called the change of variable theorem since it justifies a technique immortalized by integral calculus.

Theorem 3. *Let X be a measurable transformation from a measure space (S, Σ, μ) to a measurable space (T, \mathcal{A}) and let $v = \mu X^{-1}$ be the induced measure on \mathcal{A} . If g is a real \mathcal{A} -measurable function on T , then*

$$\int_T g dv = \int_S g(X) d\mu \quad (15)$$

in the sense that if either integral exists, so does the other and the two are equal.

PROOF. Note that $g(X)$ is a real Σ -measurable function. Since $g = g^+ - g^-$ and g^\pm are measurable functions on (T, \mathcal{A}) , it suffices to prove that every nonnegative \mathcal{A} -measurable function is in \mathcal{G} , where

$$\mathcal{G} = \{g: g \text{ is a nonnegative } \mathcal{A}\text{-measurable function for which (15) holds}\}.$$

Now by monotone convergence and linearity, \mathcal{G} is a monotone system (see Section 1.4), and if $g = I_A$ for any $A \in \mathcal{A}$, then

$$\int_T g dv = \int_A dv = \mu\{X^{-1}(A)\} = \int_S I_A(X) d\mu.$$

Thus, \mathcal{G} contains all indicator functions of sets of the σ -algebra \mathcal{A} , whence by Theorem 1.4.3, \mathcal{G} contains all nonnegative \mathcal{A} -measurable functions. \square

As an application of Theorem 3, let (Ω, \mathcal{F}, P) be a probability space and X a r.v. thereon. Then X is a measurable transformation from (Ω, \mathcal{F}, P) to (R, \mathcal{B}) (as usual, $R = [-\infty, \infty]$) and in conjunction with P induces a probability measure $v_X = P X^{-1}$ on \mathcal{B} via

$$v_X\{B\} = P\{X^{-1}(B)\}, \quad B \in \mathcal{B}. \quad (16)$$

In particular, for $-\infty < a < b < \infty$

$$v_X([a, b]) = P\{a \leq X < b\} = F_X(b) - F_X(a),$$

where F_X is the d.f. of X . Since F_X is left continuous, Theorem 6.1.2 ensures that the measure v_X coincides with the restriction to \mathcal{B} of the Lebesgue-Stieltjes measure determined by F_X . Actually, if v_X is completed to $\bar{\mathcal{B}}$ as in Exercise 6.1.1, then v_X on $\bar{\mathcal{B}}$ is precisely the Lebesgue-Stieltjes measure on \mathcal{M}_F determined by F_X (for a proof, see Halmos (1950, p. 56)).

Corollary 1. *If X is a r.v. on a probability space (Ω, \mathcal{F}, P) with d.f. F_X and g is a finite Borel function,*

$$E g(X) = \int_{-\infty}^{\infty} g(t) dF_X(t) = \int_{-\infty}^{\infty} t dF_{g(X)}(t) \quad (17)$$

in the sense that if either of the integrals exists, so does the other and the two are equal.

PROOF. Let v_X be defined on \mathcal{B} by (16). Then, as noted earlier, v_X is the restriction to \mathcal{B} of the Lebesgue-Stieltjes measure determined by F_X . By Theorem 3

$$E g(X) = \int_{\Omega} g(X) dP = \int_{\mathcal{R}} g d v_X = \int_{-\infty}^{\infty} g(t) dF_X(t) \quad (18)$$

in the sense delineated therein. Replacing $g(t)$ by t and X by $g(X)$ in (18) yields

$$E g(X) = \int_{-\infty}^{\infty} t dF_{g(X)}(t), \quad (19)$$

again in the aforementioned sense. Thus (17) and hence the corollary is proved. \square

It may be noted that $E X$ has so far been calculated only when X is an elementary function. Corollary 1 says explicitly (and Theorem 3 implicitly) that $E X$ may be replaced by a Lebesgue-Stieltjes integral, but this does not resolve the problem of evaluation. Theorem 4 asserts that under modest conditions the Lebesgue-Stieltjes and Riemann-Stieltjes integrals coincide, and fortunately the latter is susceptible of calculation via (21)(x), (vii), etc. (see below), as illustrated by Lemma 1.

A finite nondecreasing function m on $(-\infty, \infty)$ elicits, in addition to the Lebesgue-Stieltjes integral $\int_A X dm$ (defined for Borel functions X and Borel sets A whenever $\int_R X dm$ exists), also the Riemann-Stieltjes integral

$$\int_a^b f(t) dm(t), \quad -\infty < a < b < \infty.$$

(This will never be denoted by $\int_D f(t) dm(t)$, where D is a closed, open, or half-open interval with end points a and b , such notation being reserved for Lebesgue-Stieltjes integrals).

If $f(t)$ and $m(t)$ are bounded functions on $[a, b]$ the **Riemann-Stieltjes integral** of f with respect to m from a to b is defined by

$$\int_a^b f(t)dm(t) = \lim \sum_{k=1}^n f(\xi_k)[m(t_k) - m(t_{k-1})] \quad (20)$$

provided the limit, which is taken as $\max_{1 \leq k \leq n} (t_k - t_{k-1}) \rightarrow 0$, exists independently of the choice of ξ_k in $[t_{k-1}, t_k]$, where

$$a = t_0 < t_1 < \cdots < t_n = b, \quad t_{k-1} \leq \xi_k \leq t_k, \quad 1 \leq k \leq n.$$

When f is continuous and m is finite and nondecreasing on $[a, b]$, both $\int_a^b f(t)dm(t)$ and $\int_a^b m(t)df(t)$ exist (see Widder (1961, Chapter 5)). In the special case $m(t) = t$, the integral defined by (20) coalesces to the **Riemann integral** of f from a to b and is denoted $\int_a^b f(t)dt$.

Let f, f_1, f_2 be continuous functions and m, m_1, m_2 finite, nondecreasing functions on the finite interval $[a, b]$. Then (Widder (1961, Chapter 5)), if c denotes a finite constant,

- i. $\int_a^b cf(t)dm(t) = c \int_a^b f(t)dm(t), \quad \int_a^b f(t)d(m(t) + c) = \int_a^b f(t)dm(t),$
- ii. $\int_a^b dm(t) = m(b) - m(a),$
 $\int_a^b (f_1(t) + f_2(t))dm(t) = \int_a^b f_1(t)dm(t) + \int_a^b f_2(t)dm(t),$
- iii. $\int_a^b f(t)d(m_1(t) + m_2(t)) = \int_a^b f(t)dm_1(t) + \int_a^b f(t)dm_2(t),$
- iv. $\int_a^b f(t)dm(t) = \int_a^c f(t)dm(t) + \int_c^b f(t)dm(t), \quad a < c < b,$
- v. $\int_a^b f_1(t)dm(t) \leq \int_a^b f_2(t)dm(t) \text{ if } f_1(t) \leq f_2(t) \text{ on } [a, b],$
- vi. $\left| \int_a^b f(t)dm(t) \right| \leq \int_a^b |f(t)|dm(t) \leq [m(b) - m(a)] \max_{a \leq t \leq b} |f(t)|,$
- vii. $\int_a^b f(t)dm(t) + \int_a^b m(t)df(t) = m(b)f(b) - m(a)f(a),$
- viii. $\int_a^b m(t)df(t) = m(a) \int_a^\xi df(t) + m(b) \int_\xi^b df(t) \text{ for some } \xi \text{ in } [a, b],$
- ix. $\int_a^b f(t)m(t)dt = m(a) \int_a^\xi f(t)dt + m(b) \int_\xi^b f(t)dt \text{ for some } \xi \text{ in } [a, b],$
- x. $\int_a^b f(t)dm(t) = \int_a^b f(t)m'(t)dt,$

provided m has a continuous derivative m' on $[a, b]$.

The Riemann-Stieltjes integral $\int_a^b f(t)dm(t)$ has been defined for finite closed intervals $[a, b]$ and exists if f is continuous and m is finite and non-decreasing on $[a, b]$. Moreover, if f is continuous and m is finite and non-decreasing on $[a, \infty)$ for some finite constant a , the definition may be extended via

$$\int_a^\infty f(t)dm(t) = \lim_{b \rightarrow \infty} \int_a^b f(t)dm(t) \quad (22)$$

provided the limit on the right exists. Analogously, for any finite constant b , if $f(t)$ is continuous and $m(t)$ is finite and nondecreasing on $(-\infty, b]$ (resp. on $(-\infty, \infty)$) define

$$\begin{aligned} \int_{-\infty}^b f(t)dm(t) &= \lim_{a \rightarrow -\infty} \int_a^b f(t)dm(t), \\ \int_{-\infty}^\infty f(t)dm(t) &= \lim_{\substack{b \rightarrow \infty \\ a \rightarrow -\infty}} \int_a^b f(t)dm(t) \end{aligned} \quad (23)$$

provided the limits on the right exist (independently of the manner in which $b \rightarrow \infty$, $a \rightarrow -\infty$). If $a = -\infty$ or $b = \infty$ or both, $\int_a^b f(t)dm(t)$ is frequently alluded to as an improper Riemann-Stieltjes integral.

The relationship between the Riemann-Stieltjes integral $\int_a^b f(t)dm(t)$ and the Lebesgue-Stieltjes integral $\int_{[a, b]} f(t)dm(t)$ is embodied in

Theorem 4. *If $f(t)$ is a continuous function and $m(t)$ is a finite nondecreasing function on $(-\infty, \infty)$, then if $m(a) = m(a-)$ and $m(b) = m(b-)$,*

$$\int_{[a, b]} f(t)dm(t) = \int_a^b f(t)dm(t), \quad -\infty < a < b < \infty; \quad (24)$$

moreover, if $|\int_{(-\infty, \infty)} f(t)dm(t)| \leq \infty$, then the Riemann-Stieltjes integral $\int_{-\infty}^\infty f(t)dm(t)$ exists and

$$\begin{aligned} \int_{(-\infty, \infty)} f(t)dm(t) &= \int_{-\infty}^\infty f(t)dm(t), \\ \int_{(-\infty, b]} f(t)dm(t) &= \int_{-\infty}^b f(t)dm(t) \quad \text{if } m(b) = m(b-), \\ \int_{[a, \infty)} f(t)dm(t) &= \int_a^\infty f(t)dm(t) \quad \text{if } m(a-) = m(a). \end{aligned} \quad (25)$$

PROOF. Choose $t_j^{(n)}$ such that $m(t_j^{(n)}) = m(t_j^{(n)}-)$, $1 \leq j \leq n$,

$$a = t_0^{(n)} < t_1^{(n)} < \dots < t_n^{(n)} = b, \quad \max_{1 \leq k \leq n} (t_k^{(n)} - t_{k-1}^{(n)}) < \frac{2(b-a)}{n}.$$

On $[a, b]$ define

$$f_n(t) = \sum_{k=1}^n f(t_{k-1}^{(n)}) I_{[t_{k-1}^{(n)}, t_k^{(n)}]}(t),$$

and note that $f_n \rightarrow f$ on $[a, b]$ by the continuity of f .

Since the Lebesgue-Stieltjes measure $v = v_m$ determined by m is finite on the finite interval $[a, b]$ and $\max_{a \leq t \leq b} |f(t)| < \infty$, by the Lebesgue dominated convergence theorem

$$\begin{aligned} \int_{[a, b]} f(t) dm(t) &= \int_{[a, b]} f dv = \lim_n \int_{[a, b]} f_n dv \\ &= \lim_n E \sum_{k=1}^n f(t_{k-1}^{(n)}) I_{[t_{k-1}^{(n)}, t_k^{(n)}]} \\ &= \lim_n \sum_{k=1}^n f(t_{k-1}^{(n)}) [m(t_k^{(n)}) - m(t_{k-1}^{(n)})] = \int_a^b f(t) dm(t) \end{aligned}$$

since the Riemann-Stieltjes integral exists.

To prove (25), note that via the portion already proved

$$\begin{aligned} \int_{[a, \infty)} f^\pm(t) dm(t) &= \lim_{\substack{b \rightarrow \infty \\ m(b) = m(b-)}} \int_{[a, b]} f^\pm(t) dm(t) = \lim_{\substack{b \rightarrow \infty \\ m(b) = m(b-)}} \int_a^b f^\pm(t) dm(t) \\ &= \int_a^\infty f^\pm(t) dm(t), \end{aligned}$$

and the comparable statement with $[a, \infty)$ replaced by $(-\infty, b)$ follows analogously. The remaining portion of (25) obtains by letting $a \rightarrow -\infty$ subject to $m(a) = m(a-)$. \square

The Lebesgue-Stieltjes and Riemann-Stieltjes integrals can be extended to nonincreasing functions m in an obvious fashion by defining

$$\begin{aligned} \int_B f(t) dm(t) &= - \int_B f(t) d(-m(t)), \quad B \in \mathcal{B}, \\ \int_a^b f(t) dm(t) &= - \int_a^b f(t) d(-m(t)), \quad -\infty \leq a < b \leq \infty, \end{aligned} \tag{26}$$

whenever the integrals on the right are defined. In the special case where $m = 1 - F$ with F a d.f., the latter may also be expressed as

$$\int_a^b f(t) d[1 - F(t)] = - \int_a^b f(t) dF(t). \tag{27}$$

It may be noted via the definition that whenever $\int_a^b f(t) dm(t)$ exists for $-\infty < a < b < \infty$, so does $\int_{-b}^{-a} f(-t) dm(-t)$ and the two are equal.

The absolute moment $E|X|^r$, $r > 0$ always exists for any r.v. X , and its

finiteness has been shown equivalent to convergence of the series $\sum_1^\infty P\{|X| \geq n^{1/r}\}$ in Corollary 4.1.2. It may be evaluated as a Riemann integral involving the tails of the distribution of X as follows from Corollary 2 of

Lemma 1. *If X is a r.v. with d.f. F and $m(t)$ is a continuous nondecreasing function on $(-\infty, \infty)$ with $m(0) = 0$, then*

- i. $E|m(X)| = \int_{-\infty}^0 F(t)dm(t) + \int_0^\infty [1 - F(t)]dm(t),$
- ii. if $|E m(X)| < \infty$, $E m(X) = \int_0^\infty [1 - F(t)] dm(t) - \int_{-\infty}^0 F(t) dm(t).$
- iii. $E m(|X|) = \int_0^\infty [1 - F(t) + F(-t)]dm(t).$

PROOF. By Theorem 4, the discussion following, and integration by parts ((21)(vii)), if $C > 0$ and $G = 1 - F$,

$$\begin{aligned} E|m(X)| &= \int_{-\infty}^\infty |m(t)|dF(t) = \int_0^\infty m(t)dF(t) - \int_{-\infty}^0 m(t)dF(t) \\ &= \lim_{C \rightarrow \infty} \left[- \int_0^C m(t)d[1 - F(t)] - \int_{-C}^0 m(t)dF(t) \right] \\ &= \lim_{C \rightarrow \infty} \left[m(-C)F(-C) + \int_{-C}^0 F(t)dm(t) - m(C)G(C) \right. \\ &\quad \left. + \int_0^C G(t)dm(t) \right] \leq \int_{-\infty}^0 F(t)dm(t) + \int_0^\infty G(t)dm(t). \quad (28) \end{aligned}$$

Thus, in proving (i) it may be supposed that $E|m(X)| < \infty$, and so, since

$$\begin{aligned} 0 \leq m(C)G(C) - m(-C)F(-C) &\leq \int_C^\infty m(t)dF(t) - \int_{-\infty}^{-C} m(t)dF(t) \\ &= \int_{\{|t| \geq C\}} |m(t)|dF(t) = o(1) \end{aligned}$$

as $C \rightarrow \infty$, (i) follows from (28).

Part (ii) follows by applying (i) separately to m^+ and m^- . Likewise (i) implies (iii) in view of $1 - F_{|X|}(t) = P\{|X| \geq t\} = G(t) + F(-t+)$ and Exercise 6.2.6.

Corollary 2. *For any r.v. X with d.f. F ,*

$$E|X|^r = r \int_0^\infty t^{r-1} [1 - F(t) + F(-t)]dt, \quad r > 0, \quad (29)$$

and if $E X^{2k+1} < \infty$ for some $k = 0, 1, \dots$, then

$$E X^{2k+1} = (2k+1) \int_0^\infty t^{2k} [1 - F(t) - F(-t)]dt. \quad (30)$$

PROOF. The first statement is an immediate consequence of (iii) and (x) of (21), while the second follows similarly from (ii). \square

The following combines Example 2.1.1 and Corollary 4.2.4:

EXAMPLE 1. Let $\{X_n, n \geq 0\}$ be nonnegative functions with $X_n \in \mathcal{L}_1(S, \Sigma, \mu)$ for $n \geq 0$ and $E X_n \rightarrow E X_0$. If either $X_n \xrightarrow{a.s.} X_0$ or $X_n \xrightarrow{a.e.} X_0$, then $E|X_n - X_0| \rightarrow 0$.

PROOF. If $Y_n = (X_0 - X_n)^+$, then $Y_n \leq X_0$, $n \geq 1$. Since either $Y_n \xrightarrow{a.s.} 0$ or $Y_n \xrightarrow{a.e.} 0$, by Lebesgue's dominated convergence theorem $E(X_0 - X_n)^+ = E Y_n \rightarrow 0$. Then $E(X_0 - X_n) \rightarrow 0$ implies $E(X_0 - X_n)^- \rightarrow 0$, and so $E|X_0 - X_n| \rightarrow 0$. \square

Two alternative types of truncation of r.v.s have already been utilized in Theorem 4.2.2 and Theorem 5.1.3, 5.2.4, etc., and it is of interest to compare the corresponding expectations.

EXAMPLE 2. If X is a nonnegative r.v. with d.f. F and

$$Y = \min(X, s), \quad Z = XI_{\{X < s\}}, \quad (31)$$

then for any choice of r and s in $(0, \infty)$

$$E Y^r = r \int_0^s t^{r-1} [1 - F(t)] dt, \quad (32)$$

$$E Z^r = E Y^r - s^r [1 - F(s)] = \int_0^s t^r dF(t). \quad (33)$$

PROOF. If G is the d.f. of Y , then

$$G(x) = F(x)I_{\{x < s\}} + I_{\{x \geq s\}},$$

whence by Corollary 2

$$E Y^r = r \int_0^\infty t^{r-1} [1 - G(t)] dt = r \int_0^s t^{r-1} [1 - F(t)] dt.$$

Since $Z = YI_{\{X < s\}}$, via (21)(vii)

$$\begin{aligned} E Z^r &= E Y^r I_{\{X < s\}} = E Y^r - E Y^r I_{\{X \geq s\}} \\ &= r \int_0^s t^{r-1} [1 - F(t)] dt - s^r [1 - F(s)] = \int_0^s t^r dF(t). \end{aligned} \quad \square$$

EXERCISES 6.2

- If μ_1, μ_2 are measures on (Ω, \mathcal{F}) and $\int X d(\mu_1 + \mu_2)$ exists so does $\int X d\mu_i$, $i = 1, 2$, and $\int X d(\mu_1 + \mu_2) = \int X d\mu_1 + \int X d\mu_2$.
- The integral of a nonnegative measurable function over a set of measure zero has the value 0; also, if $\int_A g d\mu = 0$ for every measurable set A , then $g = 0$, a.e.

3. If S is the set of positive integers, Σ is the class of all subsets of S , and $\mu(A) = \text{number of integers in } A \in \Sigma$, then (S, Σ, μ) is a nonfinite measure space and convergence in measure is equivalent to uniform convergence everywhere.
4. Demonstrate in a nonfinite measure space (S, Σ, μ) that $X_n \xrightarrow{\text{a.e.}} X$ does not necessarily imply $X_n \xrightarrow{\text{a.s.}} X$. Hint: Utilize Exercise 3. If $X_n \xrightarrow{\text{a.s.}} X$, does $X_n Y \xrightarrow{\text{a.s.}} XY$?
5. If $X_n \in \mathcal{L}_p(S, \Sigma, \mu)$ and $X_n \xrightarrow{L_p} X$ for some $p > 0$, then $X \in \mathcal{L}_p$ and $E|X_n|^p \rightarrow E|X|^p$.
6. If f is a finite, nondecreasing function and m is a continuous nondecreasing function on $[a, b]$, where $-\infty < a < b < \infty$, then

$$\int_a^b f(t) dm(t) = \int_a^b f(t+) dm(t) = \int_a^b f(t-) dm(t).$$

7. Prove **Minkowski's inequality**: If $X_i \in \mathcal{L}_p$, $i = 1, 2$, and $p \geq 1$, $\|X_1 + X_2\|_p \leq \|X_1\|_p + \|X_2\|_p$.
8. If $F_{|X|}$ is the d.f. of $|X|$, verify that for every $c > 0$

$$\int_c^\infty t dF_{|X|}(t) = c P\{|X| \geq c\} + \int_c^\infty P\{|X| \geq t\} dt.$$

- (i) Show that a sequence of r.v.'s $\{X_n, n \geq 1\}$ is u.i. iff $\sup_{n \geq 1} \int_c^\infty P\{|X_n| \geq t\} dt \rightarrow 0$ as $c \rightarrow \infty$.
- A r.v. X is said to be **stochastically larger** than a r.v. Y if $P\{X \geq x\} \geq P\{Y \geq x\}$ for all x .
- (ii) If the r.v.s X_n , $n \geq 1$, are u.i. and $|X_n|$ is stochastically larger than $|Y_n|$, $n \geq 1$, then $\{Y_n, n \geq 1\}$ is u.i.
- If $\{X_n, n \geq 1\}$ are \mathcal{L}_1 r.v.s with a common distribution, then $E \max_{1 \leq i \leq n} |X_i| = o(n)$. Hint: Use Exercise 8 to establish u.i.
- Show that the analogue of (21)(iv) for Lebesgue-Stieltjes integrals is not true in general. Construct an example for which the Riemann-Stieltjes integral over a finite interval $[a, b]$ fails to exist.
- Establish that $g(t) = (\sin t)/t$ is Riemann but not Lebesgue integrable over $(-\infty, \infty)$ and find a function $g(t)$ which is Lebesgue integrable but not Riemann integrable.
- Let $S = \{1, 2\}$, $\Sigma = \{\{1\}, \{2\}, \emptyset, S\}$ and μ = counting measure. If $X(s) \equiv 1$, Jensen's inequality fails for the convex function X^2 .
- Let $S = \{1, 2, \dots\}$, $\Sigma = \{A : A \subset S\}$, μ = counting measure on Σ . If $X_n(s) = n^{-1} I_{\{1 \leq s \leq n\}}$, then $X_n \xrightarrow{\mu} 0$ and $E X_n \equiv 1$ despite the fact that for $\mu\{A\} < 1$, $E X_n I_A \equiv 0$. Thus, Theorem 4.2.3(i) may fail in a σ -finite measure space.
- If f, m are finite and nondecreasing on $(-\infty, \infty)$ with f continuous, prove that $\int_{[a, b]} f(t) dm(t) + \int_{[a, b]} m(t) df(t) = f(b)m(b+) - f(a)m(a-)$.
- Let $p_n \in (0, 1)$, $q_n = 1 - p_n$, where $np_n \rightarrow \lambda \in (0, \infty)$. Let μ be counting measure on the class of all subsets of $\Omega = \{1, 2, \dots\}$. If $X(j) = \lambda^j e^{-\lambda}/j!$ and $X_n(j) = \binom{n}{j} p_n^j q_n^{n-j}$, prove that $X_n \xrightarrow{\mu} X$ and $X_n \xrightarrow{L_p} X$, $p \geq 1$. Hint: Apply Example 1 or Example 2.1.1.
- (Erickson) In Example 5.4.1 the function $a(x) = x/\int_0^x [1 - F(y)] dy$ was encountered, where F is the d.f. of a nonnegative r.v. X . (i) Show that $a(x)$ is nondecreasing. (ii) Prove that $E X = \infty$ implies $E a(X) = \infty$. Hint: $E a(X) < \infty$ entails $a(x) = o((1 - F(x))^{-1})$ and hence $E X / \int_0^x y dF(y) < \infty$, contradicting the Abel-Dini theorem.

17. Random variables $\{|X_n|^p, n \geq 1\}$ are u.i. iff $\sup_{n \geq 1} \int_K^\infty t^{p-1} P\{|X_n| > t\} dt \rightarrow 0$ as $K \rightarrow \infty$.
18. (Young's inequality) Let φ be a continuous, strictly increasing function on $[0, \infty)$ with $\varphi(0) = 0$. If ψ is the inverse function (which therefore satisfies the same conditions) then for any $a \geq 0, b \geq 0$

$$ab \leq \int_0^a \varphi(x) dx + \int_0^b \psi(x) dx$$

and equality holds iff $b = \varphi(a)$.

6.3 Product Measure, Fubini's Theorem, n -Dimensional Lebesgue–Stieltjes Measure

Let $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, 2$, denote two measure spaces. Ignoring for a moment the measures, the spaces engender (Section 1.3) a product measurable space (Ω, \mathcal{F}) , where $\Omega = \bigtimes_{i=1}^2 \Omega_i = \Omega_1 \times \Omega_2$ and $\mathcal{F} = \bigtimes_{i=1}^2 \mathcal{F}_i = \mathcal{F}_1 \times \mathcal{F}_2 = \sigma(\{A_1 \times A_2 : A_i \in \mathcal{F}_i, i = 1, 2\})$.

For any set $A \in \mathcal{F}$, the sections

$$A^{(1)}(\omega_1) = \{\omega_2 : (\omega_1, \omega_2) \in A\}, \quad A^{(2)}(\omega_2) = \{\omega_1 : (\omega_1, \omega_2) \in A\}$$

are \mathcal{F}_2 - and \mathcal{F}_1 -measurable respectively according to Theorem 1.4.2. Thus, $\mu_2\{A^{(1)}(\omega_1)\}$ and $\mu_1\{A^{(2)}(\omega_2)\}$ are well-defined real functions, the first on Ω_1 and the second on Ω_2 . For notational simplicity these will be denoted by $\mu_2\{A^{(1)}\}$ and $\mu_1\{A^{(2)}\}$ respectively.

Now if μ_1 and μ_2 are finite measures,

$$\mathcal{A} = \{A \in \mathcal{F} : \mu_{3-i}\{A^{(i)}\} \text{ is } \mathcal{F}_i\text{-measurable}, i = 1, 2\}$$

is a λ -class containing all rectangles with measurable sides, whence $\mathcal{A} \supset \mathcal{F}$, that is, $\mu_{3-i}\{A^{(i)}\}$ is \mathcal{F}_i -measurable $i = 1, 2$; this carries over to the case where μ_i is σ -finite, $i = 1, 2$, since if $\Omega_i = \bigcup_{j=1}^\infty B_{ij}$, where B_{ij} are disjoint sets of \mathcal{F}_i , $\mu_i\{B_{ij}\} < \infty, j \geq 1, i = 1, 2$, then $\Omega_1 \times \Omega_2 = \bigcup_{j_1, j_2} (B_{1j_1} \times B_{2j_2})$ and every measurable set $A = \bigcup_{j_1, j_2} A(B_{1j_1} \times B_{2j_2})$ with $\mu_i\{A^{(3-i)}\} = \sum_{j_1, j_2} \mu_i\{A(B_{1j_1} \times B_{2j_2})\}^{(3-i)}$ being \mathcal{F}_{3-i} -measurable, $i = 1, 2$. Thus, when $(\Omega_i, \mathcal{F}_i, \mu_i)$ are σ -finite measure spaces, $i = 1, 2$, and $A \in \mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$, $\mu_{3-i}\{A^{(i)}\}$ is \mathcal{F}_i -measurable, $i = 1, 2$.

Theorem 1. If $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, 2$, are σ -finite measure spaces and $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$, then there exists a σ -finite measure space $(\Omega, \mathcal{F}, \mu)$ for which

$$\mu\{B_1 \times B_2\} = \mu_1\{B_1\} \cdot \mu_2\{B_2\} \quad \text{if } B_i \in \mathcal{F}_i, \quad i = 1, 2; \quad (1)$$

moreover, the measure μ is uniquely determined by (1) and, furthermore, for every $A \in \mathcal{F}$, $\mu_{3-i}\{A^{(i)}\}$ is \mathcal{F}_i -measurable, $i = 1, 2$, and

$$\mu\{A\} = \int_{\Omega_2} \mu_1\{A^{(2)}\} d\mu_2 = \int_{\Omega_1} \mu_2\{A^{(1)}\} d\mu_1. \quad (2)$$

PROOF. For $A \in \mathcal{F}$, define μ by the first equality in (2). Then $\mu \geq 0$ and $\mu\{\emptyset\} = 0$. Since $(\bigcup_1^\infty A_n)^{(2)} = \bigcup_1^\infty A_n^{(2)}$ for every sequence $\{A_n\}$ of \mathcal{F} sets, and disjointness of $\{A_n\}$ entails that of $\{A_n^{(2)}\}$, the set function μ so defined is a measure on \mathcal{F} , whence $(\Omega, \mathcal{F}, \mu)$ is a measure space. Moreover, if $A = B_1 \times B_2$, where $B_i \in \mathcal{F}_i$, $i = 1, 2$, then

$$\mu\{A\} = \int_{\Omega_2} \mu_1\{A^{(2)}\} d\mu_2 = \int_{B_2} \mu_1\{A^{(2)}\} d\mu_2 = \int_{B_2} \mu_1\{B_1\} d\mu_2 = \prod_{i=1}^2 \mu_i\{B_i\},$$

so that (1) holds. Now (Exercise 1.5.3),

$$\mathcal{G} = \{B_1 \times B_2 : B_i \in \mathcal{F}_i, i = 1, 2\}$$

is a semi-algebra and, furthermore, since μ_1 and μ_2 are σ -finite, $\tilde{\mu} = \mu|_{\mathcal{G}}$ is a σ -finite measure on \mathcal{G} .

By Theorem 6.1.1 the extension μ of $\tilde{\mu}$ is σ -finite on $\mathcal{F} = \sigma(\mathcal{G})$ and uniquely determined by $\tilde{\mu}$. Finally, if

$$\mu^*\{A\} = \int_{\Omega_1} \mu_2\{A^{(1)}\} d\mu_1, \quad A \in \mathcal{F},$$

then, by analogy with the preceding, μ^* is a measure on \mathcal{F} satisfying (1), whence by uniqueness $\mu = \mu^*$, i.e., (2) holds. \square

The measure μ defined in (2) is called the **product measure** of the σ -finite measures μ_1 and μ_2 and is denoted by $\mu_1 \times \mu_2$. The σ -finite measure space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, \mu_1 \times \mu_2)$ is referred to as the **product measure space** (of the σ -finite measure spaces $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, 2$).

In situations where more than one measure is floating around, almost everywhere statements must be qualified. Thus, **a.e.** [μ_1] abbreviates the statement "except for a set of μ_1 -measure zero."

Corollary 1. If $(\Omega, \mathcal{F}, \mu)$ is the product measure space of the σ -finite measure spaces $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, 2$, and $\mu\{A\} = 0$ for some $A \in \mathcal{F}$, then $\mu_1\{A^{(2)}(\omega_2)\} = 0$, a.e. [μ_2], and $\mu_2\{A^{(1)}(\omega_1)\} = 0$, a.e. [μ_1].

If $X = X(\omega_1, \omega_2)$ is a function on the product measurable space $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$, then the functions $X_{\omega_1}^{(1)}(\omega_2)$ defined for each $\omega_1 \in \Omega_1$ by

$$X_{\omega_1}^{(1)}(\omega_2) = X(\omega_1, \omega_2)$$

are called **sections of X at ω_1** . Analogously, the functions $X_{\omega_2}^{(2)}(\omega_1)$ are **sections of X at ω_2** . It follows directly from Theorem 1.4.2(ii) that every section at ω_i of an $\mathcal{F}_1 \times \mathcal{F}_2$ -measurable function X is an \mathcal{F}_{3-i} -measurable function, $i = 1, 2$. Define for $i = 1, 2$

$$A_{3-i} = \left\{ \omega_{3-i} : \int_{\Omega_1} X^+(\omega_1, \omega_2) d\mu_i = \int_{\Omega_1} X^-(\omega_1, \omega_2) d\mu_i = \infty \right\}. \quad (3)$$

Theorem 2 (Fubini). If $(\Omega, \mathcal{F}, \mu)$ is the product measure space of the σ -finite measure spaces $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, 2$, and X is an \mathcal{F} -measurable function whose integral exists, then

$$\int_{\Omega} X(\omega_1, \omega_2) d\mu_i \equiv \left[\int_{\Omega_1} X^+(\omega_1, \omega_2) d\mu_i - \int_{\Omega_1} X^-(\omega_1, \omega_2) d\mu_i \right] I_{A_{3-i}^c}$$

is \mathcal{F}_{3-i} -measurable, $i = 1, 2$, and

$$\int_{\Omega} X d\mu = \int_{\Omega_1} d\mu_1 \int_{\Omega_2} X(\omega_1, \omega_2) d\mu_2 = \int_{\Omega_2} d\mu_2 \int_{\Omega_1} X(\omega_1, \omega_2) d\mu_1; \quad (4)$$

moreover, in the case where X is integrable so are almost all sections of X at ω_i for $i = 1, 2$.

PROOF. The general case follows from consideration of X^+ , X^- , and so $X \geq 0$ may be so supposed; moreover, μ_1 and μ_2 may be assumed finite. Let $\mathcal{H} = \{X : X \geq 0, X \text{ is } \mathcal{F}\text{-measurable}, \int_{\Omega_i} X(\omega_1, \omega_2) d\mu_i \text{ is } \mathcal{F}_{3-i}\text{-measurable, } i = 1, 2, \text{ and (4) holds}\}$. Then \mathcal{H} contains all indicator functions of sets of \mathcal{F} by Theorem 1. Since \mathcal{H} is a monotone system, \mathcal{H} contains all nonnegative \mathcal{F} -measurable functions by Theorem 1.4.3. Moreover, if $X \geq 0$ is integrable, it is clear from (4) that so are $\int_{\Omega_2} X(\omega_1, \omega_2) d\mu_2$ and $\int_{\Omega_1} X(\omega_1, \omega_2) d\mu_1$ for almost all ω_1 and ω_2 respectively. \square

If $(\Omega_i, \mathcal{F}_i, \mu_i)$, $1 \leq i \leq n$, are σ -finite measure spaces and $(\Omega, \mathcal{F}) = \bigtimes_{i=1}^n (\Omega_i, \mathcal{F}_i)$ is the product measurable space, it can be proved inductively that there is one and only one measure μ on $\bigtimes_{i=1}^n \mathcal{F}_i$ such that for every measurable rectangle $B_1 \times \cdots \times B_n$

$$\mu(B_1 \times \cdots \times B_n) = \prod_{i=1}^n \mu_i(B_i).$$

Then μ is called the product measure and denoted by $\mu_1 \times \cdots \times \mu_n$ or $\bigtimes_{i=1}^n \mu_i$. Associativity of ordinary multiplication together with Theorem 6.1.1 guarantees that $\bigtimes_{i=1}^n \mu_i = (\bigtimes_{i=1}^m \mu_i) \times (\bigtimes_{m+1}^n \mu_i)$ for any integer m in $[1, n]$. The measure space $(\Omega, \mathcal{F}, \mu) = (\bigtimes_1^n \Omega_i, \bigtimes_{i=1}^n \mathcal{F}_i, \bigtimes_{i=1}^n \mu_i)$ is alluded to as the (σ -finite) **product measure space** of $(\Omega_i, \mathcal{F}_i, \mu_i)$, $1 \leq i \leq n$, and is denoted by $\bigtimes_{i=1}^n (\Omega_i, \mathcal{F}_i, \mu_i)$. The extension of Fubini's theorem is immediate, and so the integral of any nonnegative \mathcal{F} -measurable or μ -integrable function on the product measure space $\bigtimes_{i=1}^n (\Omega_i, \mathcal{F}_i, \mu_i)$ may be evaluated by iterated integrals **in any order** (i.e., via an n -fold iteration of integrals analogous to (4), where $n = 2$).

In the important special case where $\Omega_i = R = [-\infty, \infty]$, $\mathcal{F}_i = \mathcal{B}$ = the class of all (linear) Borel measurable sets and $\mu_i = \mu = \text{Lebesgue measure}$, $1 \leq i \leq n$, the corresponding product measure space $(R^n, \mathcal{B}^n, \mu^n)$ is the **n -dimensional Borel measure space**.

In Section 1.6 it was observed that the d.f. F_X of a r.v. X on a probability

space (Ω, \mathcal{F}, P) is a d.f. in the sense that it is a nondecreasing, left-continuous function on R with

$$F(-\infty) = F(-\infty+) = 0, \quad F(\infty) = F(\infty-) = 1.$$

The converse was mentioned there but its proof need be deferred no longer. If G is an arbitrary d.f. on the line, there always exists a r.v. X on some probability space with $F_X = G$. It suffices to take $\Omega = R$, $\mathcal{F} = \mathcal{B}$ = linear Borel sets, $X(\omega) = \omega$ (the coordinate r.v.) and to define $P\{[a, b]\} = G(b) - G(a)$, $-\infty \leq a \leq b \leq \infty$. By Theorem 6.1.2 this ensures a unique probability measure P on \mathcal{B} , and, clearly,

$$F_X(x) = P\{X(\omega) < x\} = P\{\omega : \omega < x\} = G(x).$$

This statement and indeed the notions of Lebesgue-Stieltjes measure, Lebesgue-Stieltjes integral, etc., extend readily to R^n . An **n -dimensional distribution function** or **d.f. on R^n** is a function F satisfying

$$\lim_{x_j \rightarrow -\infty} F(x_1, \dots, x_n) = F(x_1, \dots, x_{j-1}, -\infty, x_{j+1}, \dots, x_n) = 0, \quad 1 \leq j \leq n \quad (5, i)$$

$$\lim_{x_j > y_j \rightarrow x_j} F(x_1, \dots, x_{j-1}, y_j, x_{j+1}, \dots, x_n) = F(x_1, \dots, x_j, \dots, x_n), \quad 1 \leq j \leq n, \quad (5, ii)$$

$$\begin{aligned} \Delta_n^{a,b} F &\equiv F(b_1, \dots, b_n) - \sum_{j=1}^n F(b_1, \dots, b_{j-1}, a_j, b_{j+1}, \dots, b_n) \\ &+ \sum_{1 \leq j < k \leq n} F(b_1, \dots, b_{j-1}, a_j, b_{j+1}, \dots, b_{k-1}, a_k, b_{k+1}, \dots, b_n) - \dots \\ &+ (-1)^n F(a_1, \dots, a_n) \geq 0, \end{aligned} \quad (5, iii)$$

$$\lim_{\substack{x_j \rightarrow \infty \\ 1 \leq j \leq n}} F(x_1, \dots, x_n) = F(\infty, \dots, \infty) = 1, \quad (5, iv)$$

whenever $-\infty \leq a_i \leq b_i \leq \infty$, $1 \leq i \leq n$.

Let \mathcal{S} be the class of all finite or infinite intervals $[a, b)$ together with \emptyset , R , $\{\infty\}$ and define

$$\mathcal{S}^n = \{S : S = S_1 \times S_2 \times \dots \times S_n, S_i \in \mathcal{S}, 1 \leq i \leq n\}$$

and

$$P(S) = \begin{cases} \Delta_n^{a,b} F & \text{when } S = \bigtimes_{i=1}^n S_i \text{ with either } S_i = [a_i, b_i] \text{ or} \\ & S_i = [a_i, b_i], \quad b_i = -a_i = \infty \\ 0 & \text{if } S = \bigtimes_{i=1}^n S_i \text{ with some } S_i = \emptyset \text{ or } \{\infty\}. \end{cases} \quad (6)$$

where F satisfies (5, i)–(5, iii) and is finite on $(-\infty, \infty)^n$. As in Lemma 6.1.1, P is a measure on the semi-algebra \mathcal{S}^n , and so by Theorem 6.1.1, P can be extended to \mathcal{M}^n , the class of all P -measurable sets, the extension being unique on the class \mathcal{B}^n of n -dimensional Borel sets. The measure P and the measure space (R^n, \mathcal{M}^n, P) are called respectively the **n -dimensional Lebesgue-Stieltjes measure** and the **n -dimensional Lebesgue-Stieltjes measure space determined by F** . The same terms respectively are used for the restriction of P to \mathcal{B}^n and (R^n, \mathcal{B}^n, P) . If F is a d.f., that is, satisfies (5, iv) also, then (R^n, \mathcal{B}^n, P) is a probability space.

If X_1, \dots, X_n are r.v.s on a probability space (Ω, \mathcal{F}, P) , their joint d.f.

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = P\{X_1 < x_1, \dots, X_n < x_n\}$$

is readily verified to be a d.f. on R^n in the sense of (5). Conversely, given any n -dimensional d.f. F as in (5, i)–(5, iv), there always exist n r.v.s X_1, \dots, X_n on some probability space (Ω, \mathcal{F}, P) whose joint d.f. is the preordained F . It suffices to choose $\Omega = R^n$, $\mathcal{F} = \mathcal{B}^n$, $X_i(\omega) = i$ th coordinate of ω , $1 \leq i \leq n$, and to define P on \mathcal{S}^n via (6). According to the prior discussion, P is uniquely determined on the Lebesgue-Stieltjes measure space (R^n, \mathcal{B}^n, P) . Moreover,

$$\begin{aligned} P\{X_1 < x_1, \dots, X_n < x_n\} &= P\{\omega : \omega_i < x_i, 1 \leq i \leq n\} \\ &= \lim_{a \rightarrow -\infty} \Delta_a^{x_i} F = F(x_1, \dots, x_n). \end{aligned}$$

Thus, there is a one-to-one correspondence between distribution functions on R^n and probability measures on (R^n, \mathcal{B}^n) .

If $Y = Y(\omega_1, \dots, \omega_n)$ is an integrable function on the Lebesgue-Stieltjes measure space (R^n, \mathcal{M}^n, P) , the integral of $Y \cdot I_A$ for any $A \in \mathcal{B}^n$ is denoted by $\int_A Y I_A dP$ or by

$$\int \cdots \int Y(\omega_1, \dots, \omega_n) dF(\omega_1, \dots, \omega_n) \quad (7)$$

and abbreviated (when the context is clear) by $\int_A Y dF$.

The analogue of the first part of Corollary 6.2.1 is

Theorem 3. *If $X = (X_1, \dots, X_m)$ is a random vector with d.f. F_X and g is a Borel function on R^m ,*

$$E g(X) = \int_{R^m} g(t_1, \dots, t_m) dF_X(t_1, \dots, t_m)$$

in the sense that if either of the integrals exists, so does the other and the two are equal.

It should be noted that the iterated integrals (4) of Fubini's theorem may also be written as in

$$\int_{\Omega_1} \int_{\Omega_2} X(\omega_1, \omega_2) d\mu_2(\omega_2) d\mu_1(\omega_1), \quad \int_{\Omega_1} \int_{\Omega_2} X d\mu_2 d\mu_1. \quad (8)$$

Definition. **Convolution** of d.f.s is a binary operation denoted by $*$, which associates with every pair F_1, F_2 of d.f.s, a d.f. F , denoted by $F_1 * F_2$, and defined by

$$F(x) = (F_1 * F_2)(x) = \int_{-\infty}^{\infty} F_1(x - y)dF_2(y). \quad (9)$$

The dominated convergence theorem permits limits to be taken inside the integral and thus the convolution $F_1 * F_2$ of any two d.f.s is again a d.f.

The primary interest of convolution for probability theory stems from

Theorem 4. *If X_1 and X_2 are independent r.v.s on some probability space (Ω, \mathcal{F}, P) with corresponding d.f.s F_1 and F_2 , then their sum $X_1 + X_2$ has the d.f. $F_1 * F_2$.*

PROOF. Since, via the independence of X_1 and X_2 , the mapping $\omega \rightarrow (X_1(\omega), X_2(\omega))$ takes P -measure on \mathcal{F} sets into the product $(F_1 \times F_2)$ -measure on the Borel sets of the plane, by Theorem 3 and Fubini's theorem

$$\begin{aligned} P\{X_1 + X_2 < x\} &= \int_{\Omega} I_{[X_1 + X_2 < x]}(\omega) dP(\omega) = \int_{[z+y < x]} d(F_1 \times F_2)(z, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_{[z+y < x]}(z, y) dF_1(z) dF_2(y) \\ &= \int_{-\infty}^{\infty} F_1(x - y) dF_2(y), \end{aligned}$$

and so the d.f. of $X_1 + X_2$ is $F_1 * F_2$. □

The convolution formula (9) may be paralleled by a convolution of the measures $F_1\{\cdot\}$, $F_2\{\cdot\}$ determined by the d.f.s F_1 , F_2 . To this end, let the translation of the Borel set B by the amount y be denoted by $B - y = \{x - y : x \in B\}$ and define

$$F\{B\} = \int_{-\infty}^{\infty} F_1\{B - y\} dF_2(y). \quad (10)$$

It is a simple matter to check that F is a probability measure on the Borel sets of the line. Moreover, when $B = (-\infty, x)$, (10) reduces to (9). Then by Theorem 6.1.1 the measure $F\{\cdot\}$ of (10) is that determined by the d.f. $F_1 * F_2$ of (9), thereby justifying the common name F . Furthermore, for any Borel function g integrable relative to $F_1 * F_2$,

$$\int_{-\infty}^{\infty} g(z) d(F_1 * F_2)(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x + y) dF_1(x) dF_2(y). \quad (11)$$

When g is the indicator function I_B of some Borel set B , then $g(x + y) = I_{B-y}$, for every real y , whence (11) coalesces to (10). Hence, by Theorem 1.4.3, (11) holds for any $(F_1 * F_2)$ -integrable function.

EXAMPLE 1. Let (Γ, \mathcal{F}) and (Λ, \mathcal{G}) be measurable spaces with v a measure on \mathcal{G} and $\mu_\lambda(A)$ a function on $\mathcal{F} \times \Lambda$ such that $\mu_\lambda(\cdot)$ is a measure on \mathcal{F} for almost all $\lambda \in \Lambda$ and moreover $\mu_\lambda(A)$ is \mathcal{G} -measurable for every $A \in \mathcal{F}$. Define

$$\mu(A) = \int_{\Lambda} \mu_\lambda(A) dv(\lambda), \quad A \in \mathcal{F}$$

and for any \mathcal{F} -measurable function f

$$\int_{\Gamma} f(y) d\mu_\lambda(y) = \left[\int_{\Gamma} f^+(y) d\mu_\lambda(y) - \int_{\Gamma} f^-(y) d\mu_\lambda(y) \right] I_{\{\int_{\Gamma} f^+ d\mu_\lambda = \infty = \int_{\Gamma} f^- d\mu_\lambda\}}$$

Then μ is a measure on \mathcal{F} , $\int_{\Gamma} f(y) d\mu_\lambda(y)$ is \mathcal{G} -measurable and

$$\int_{\Gamma} f(y) d\mu(y) = \int_{\Lambda} \left(\int_{\Gamma} f(y) d\mu_\lambda(y) \right) dv(\lambda) \quad (12)$$

if either integral exists.

PROOF. Let $\Lambda' = \{\lambda: \mu_\lambda \text{ is a measure on } \mathcal{F}\}$. Then $\mu(\Lambda - \Lambda') = 0$ whence $\mu(A) = \int_{\Lambda'} \mu_\lambda(A) dv(\lambda)$. Via the monotone convergence theorem, μ is a measure on \mathcal{F} . As usual, in verifying (12), it suffices to consider non-negative f . Let $\mathcal{H} = \{f \geq 0: f \text{ is } \mathcal{F}\text{-measurable}, \int_{\Gamma} f(y) d\mu_\lambda(y) \text{ is } \mathcal{G}\text{-measurable and (12) holds}\}$. Then $I_A \in \mathcal{H}$ for all $A \in \mathcal{F}$ by definition and moreover by Theorem 6.2.1 \mathcal{H} is a monotone system. Thus, by Theorem 1.4.3, \mathcal{H} contains all non-negative \mathcal{F} -measurable functions. \square

EXERCISES 6.3

- Let $(\Omega_i, \mathcal{F}_i, P_i)$ be a probability space, where $\Omega_i = (-\infty, \infty)$ and $\mathcal{F}_i = \{B: B \text{ is a Borel subset of } (-\infty, \infty)\}$, $i = 1, 2$. If $(\Omega, \mathcal{F}, P) = (\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2, P_1 \times P_2)$ and if $X_i(\omega) = \omega_i$, $i = 1, 2$, for $\omega = (\omega_1, \omega_2) \in \Omega$, then X_1 and X_2 are independent r.v.s on (Ω, \mathcal{F}, P) .
- If $\Omega_1 = \Omega_2 = [0, 1]$, $\mathcal{F}_1 = \mathcal{F}_2 = [0, 1] \cdot \mathcal{B}$, $\mu_1 = \text{Lebesgue measure}$, and $\mu_2(A) = \text{number of points in } A$, then $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, 2$, are measure spaces. The set $A = \{(\omega_1, \omega_2): \omega_1 = \omega_2\} \in \mathcal{F}_1 \times \mathcal{F}_2$, but

$$\int_{\Omega_2} \mu_1(A^{(2)}) d\mu_2 = 0 = 1 - \int_{\Omega_1} \mu_2(A^{(1)}) d\mu_1,$$

whence Theorem 1 as well as the uniqueness part of Theorem 6.1.1 is invalid if σ -finiteness is omitted.

- For any (not necessarily σ -finite) measure spaces $(\Omega_i, \mathcal{F}_i, \mu_i)$, $i = 1, 2$, and any integrable function X on the product measure space $(\Omega, \mathcal{F}, \mu)$,

$$\int_{\Omega} X d\mu = \int_{\Omega_1} d\mu_1 \int_{\Omega_2} X d\mu_2 = \int_{\Omega_2} d\mu_2 \int_{\Omega_1} X d\mu_1.$$

Hint: Consider $X \geq 0$ and utilize the remark just prior to Theorem 6.2.1.

4. For $j \geq 1$, define $a_{i,2j-1} = (2^j - 1)^{-1}/2^j$ and $a_{i,2j} = -a_{i,2j+1}$, $i \geq 1$. The row sums of the double series $\sum_j a_{ij}$ are simply the elements of the first column. What are the column sums? Taking μ_1 and μ_2 as counting measure, does this example contradict Fubini's theorem?
5. Show that $g(\omega_1, \omega_2) = e^{-\omega_1 \omega_2} - 2e^{-2\omega_1 \omega_2}$ is Lebesgue integrable (i) over $\Omega_1 = [1, \infty)$ for each ω_2 and (ii) over $\Omega_2 = (0, 1]$ for each ω_1 , but that Fubini's theorem fails. Why?
6. An alternative construction of a r.v. X on a probability space with a preassigned d.f. F is to take $\Omega = [0, 1]$, \mathcal{F} = Borel subsets of $[0, 1]$, P = Lebesgue measure on $[0, 1]$, and $X(\omega) = F^{-1}(\omega)$, where $F^{-1}(\omega) = \sup\{x: F(x) < \omega\}$.
7. Prove that the random vectors $X = (X_1, \dots, X_m)$ and $Y = (Y_1, \dots, Y_n)$ on (Ω, \mathcal{F}, P) are independent of one another iff their joint d.f. $F_{X,Y} = F_X \cdot F_Y$ and conclude that X and Y independent entails $(R^{m+n}, \mathcal{B}^{m+n}, v_{X,Y}) = (R^m, \mathcal{B}^m, v_X) \times (R^n, \mathcal{B}^n, v_Y)$, where v_X , v_Y , and $v_{X,Y}$ are the Lebesgue-Stieltjes measures determined by F_X , F_Y , $F_{X,Y}$ respectively.
8. Random variables X_1, X_2 have a **bivariate Poisson distribution** if

$$P\{X_1 = j, X_2 = k\} = e^{-(a_1 + a_2 + a_{12})} \sum_{i=0}^{\min(j,k)} \frac{a_{12}^i a_1^{j-i} a_2^{k-i}}{i!(j-i)!(k-i)!}$$

for any pair of nonnegative integers (j, k) , where a_1, a_2, a_{12} are nonnegative parameters. Define a probability space and r.v.s X_1, X_2 on it whose joint distribution is bivariate Poisson and show that X_j is a Poisson r.v. with mean $a_j + a_{12}$. Prove that the correlation $\rho(X_1, X_2) \geq 0$ and that X_1 and X_2 are independent iff $a_{12} = 0$.

9. Random variables X_1, \dots, X_k have a **multinomial distribution** if $P\{X_i = x_i, 1 \leq i \leq k\} = n! \prod_{i=1}^k (p_i^{x_i} / (x_i!))$ for any nonnegative integers x_i , $1 \leq i \leq k$, with $\sum_i x_i = n$ and zero otherwise. Here, n is a positive integer and $\sum_i p_i = 1$, $p_i > 0$, $1 \leq i \leq k$. Prove that if $\{A_i, 1 \leq i \leq k\}$ is a partition of Ω in \mathcal{F} with $p_i = P\{A_i\}$ and $X_i = \text{number of occurrences of } A_i \text{ in } n \text{ independent trials}$, $1 \leq i \leq k$, then X_1, \dots, X_k have a multinomial distribution. Show that X_i is a binomial r.v. with parameters p_i , n and that $\rho(X_i, X_j) < 0$.
10. Random variables $\{X_i, 1 \leq i \leq n\}$ are called **interchangeable** if their joint d.f. is a symmetric function, i.e., invariant under permutations, and r.v.s $\{X_n, n \geq 1\}$ are interchangeable if every finite subset is interchangeable. Prove that if $\{X_n, n \geq 1\}$ are \mathcal{L}_2 interchangeable r.v.s, then $\rho(X_1, X_2) \geq 0$. Hint: $\sigma^2(\sum_i X_i) \geq 0$.

6.4 Infinite-Dimensional Product Measure Space, Kolmogorov Consistency Theorem

In determining a product measure on an infinite-dimensional product measurable space (Section 1.3) it is well to keep in mind the distinguished role played by the number 1 in infinite (and even finite) products.

Let $(\Omega_i, \mathcal{F}_i, P_i)$, $i \geq 1$, be a sequence of probability spaces and define

$$\begin{aligned}\Omega &= \bigtimes_{i=1}^{\infty} \Omega_i, & \Omega'_n &= \bigtimes_{i=1}^n \Omega_i, & \Omega''_n &= \bigtimes_{i=n+1}^{\infty} \Omega_i, \\ \mathcal{F} &= \bigtimes_{i=1}^{\infty} \mathcal{F}_i, & \mathcal{F}'_n &= \bigtimes_{i=1}^n \mathcal{F}_i, & \mathcal{F}''_n &= \bigtimes_{i=n+1}^{\infty} \mathcal{F}_i, \\ \mathcal{G}_n &= \{A : A = B_n \times \Omega''_n \text{ where } B_n \in \mathcal{F}'_n\}, & \mathcal{G} &= \bigcup_{n=1}^{\infty} \mathcal{G}_n.\end{aligned}$$

Then for each $n \geq 1$, $(\Omega''_n, \mathcal{F}''_n)$ is a measurable space, $\{\mathcal{G}_n\}$ is an increasing sequence of sub- σ -algebras of \mathcal{F} , and \mathcal{G} is the algebra of cylinders of \mathcal{F} . For $A = B_n \times \Omega''_n \in \mathcal{G}_n$, define

$$P\{A\} = (P_1 \times P_2 \times \cdots \times P_n)\{B_n\}.$$

Note that P is well defined on \mathcal{G} , since if $A = B_n \times \Omega''_n = B_m \times \Omega''_m \in \mathcal{G}_n \cdot \mathcal{G}_m$ where, say $m > n$, then necessarily $B_m = B_n \times \Omega_{n+1} \times \cdots \times \Omega_m$, whence

$$\begin{aligned}(P_1 \times \cdots \times P_m)\{B_m\} &= (P_1 \times \cdots \times P_n)\{B_n\} \cdot \prod_{i=n+1}^m P_i\{\Omega_i\} \\ &= (P_1 \times \cdots \times P_n)\{B_n\}.\end{aligned}$$

Clearly, P is σ -additive on \mathcal{G}_n , $n \geq 1$, and additive on the algebra \mathcal{G} , and, moreover, $P\{\Omega\} = 1$.

Now $\Omega = \Omega'_n \times \Omega''_n$, $\mathcal{F} = \mathcal{F}'_n \times \mathcal{F}''_n$, $n \geq 1$, that is, if $\omega \in \Omega$, $\omega = (\omega'_n, \omega''_n)$, where $\omega'_n \in \Omega'_n$, $\omega''_n \in \Omega''_n$ for $n \geq 1$. For any $A \subset \Omega$, set

$$A(\omega'_n) = \{\omega''_n : (\omega'_n, \omega''_n) \in A\}, \quad A(\omega''_n) = \{\omega'_n : (\omega'_n, \omega''_n) \in A\}.$$

By Theorem 1.4.2, if $A \in \mathcal{F}$, then $A(\omega'_n)$ is \mathcal{F}''_n -measurable and if $A \in \mathcal{G}$, say $A = B_m \times \Omega''_m \in \mathcal{G}_m$, where $m > n$, then

$$\begin{aligned}A(\omega'_n) &= \{\omega''_n : (\omega'_n, \omega''_n) \in B_m \times \Omega''_m\} \\ &= \{(\omega_{n+1}, \dots, \omega_m) : (\omega_1, \dots, \omega_m) \in B_m\} \times \Omega''_m \\ &= B_m(\omega'_n) \times \Omega''_m\end{aligned}\tag{1}$$

is an $(m - n)$ -dimensional cylinder in $(\Omega''_n, \mathcal{F}''_n)$ so that if $P^{(n)}$ is the analogue of P for the space $(\Omega''_n, \mathcal{F}''_n)$

$$P^{(n)}\{A(\omega'_n)\} = (P_{n+1} \times \cdots \times P_m)\{B_m(\omega'_n)\}.$$

Hence, via Theorem 6.3.1 and Fubini's theorem,

$$\begin{aligned}P\{A\} &= (P_1 \times \cdots \times P_m)\{B_m\} \\ &= \int_{\Omega'_n} (P_{n+1} \times \cdots \times P_m)\{B_m(\omega'_n)\} d(P_1 \times \cdots \times P_n) \\ &= \int_{\Omega_n} \cdots \int_{\Omega_1} P^{(n)}\{A(\omega'_n)\} dP_1(\omega_1) \cdots dP_n(\omega_n).\end{aligned}\tag{2}$$

Theorem 1. If $(\Omega_i, \mathcal{F}_i, P_i)$, $i \geq 1$, is a sequence of probability spaces, there exists a unique probability measure P on the product σ -algebra $\bigtimes_{i=1}^{\infty} \mathcal{F}_i$ such that for every cylinder $A = B_n \times \Omega_{n+1} \times \Omega_{n+2} \times \dots$ with $B_n \in \bigtimes_{i=1}^n \mathcal{F}_i$,

$$P\{A\} = (P_1 \times \dots \times P_n)\{B_n\}. \quad (3)$$

PROOF. Clearly, P , as defined by (3), is nonnegative and additive on the algebra \mathcal{G} with $P\{\Omega\} = 1$. In view of Theorem 6.1.1 it is sufficient to verify σ -additivity of P on \mathcal{G} and thus by Theorem 1.5.2(iv) to prove that $A_n \in \mathcal{G}$, $n \geq 1$, with $A_n \downarrow$ and $\inf_{n \geq 1} P\{A_n\} = \varepsilon > 0$ imply $\bigcap_{n=1}^{\infty} A_n \neq \emptyset$. To this end, set $D_n = \{\omega_1 : P^{(1)}\{A_n(\omega_1)\} > \varepsilon/2\}$, noting that the prime in (2) is superfluous since $\Omega'_1 = \Omega_1$. Since

$$\begin{aligned} \varepsilon &\leq P\{A_n\} = \int_{\Omega_1} P^{(1)}\{A_n(\omega_1)\} dP_1 \\ &= \int_{D_n} P^{(1)}\{A_n(\omega_1)\} dP_1 + \int_{D_n^c} P^{(1)}\{A_n(\omega_1)\} dP_1 \\ &\leq P_1\{D_n\} + \frac{\varepsilon}{2}, \end{aligned}$$

necessarily $P_1\{D_n\} \geq \varepsilon/2$, $n \geq 1$. Now $\{D_n\}$ is a decreasing sequence of \mathcal{F}_1 sets, whence, since P_1 is a probability measure on \mathcal{F}_1 , Theorem 1.5.2(iv) ensures the existence of a point $\omega_1^* \in \Omega_1$ with $\omega_1^* \in \bigcap_{n=1}^{\infty} D_n$. Thus, $A_n(\omega_1^*)$, $n \geq 1$, is a decreasing sequence of $\mathcal{G}^{(1)}$ sets with $P^{(1)}\{A_n(\omega_1^*)\} \geq \varepsilon/2$, $n \geq 1$ and the argument just offered with Ω_1 , $\{A_n\}$, ε applies to Ω_2' , $\{A_n(\omega_1^*)\}$, $\varepsilon/2$, and yields a point $\omega_2^* \in \Omega_2$ with $P^{(2)}\{A_n(\omega_1^*, \omega_2^*)\} \geq \varepsilon/4$, $n \geq 1$. Continuing in this fashion a sequence $(\omega_1^*, \omega_2^*, \dots)$ is obtained with $\omega_m^* \in \Omega_m$ and

$$P^{(m)}\{A_n(\omega_1^*, \dots, \omega_m^*)\} \geq \frac{\varepsilon}{2^m}, \quad m = 1, 2, \dots; n = 1, 2, \dots \quad (4)$$

To prove that $\omega^* = (\omega_1^*, \omega_2^*, \dots) \in \bigcap_{n=1}^{\infty} A_n$, note that since for each $n \geq 1$, $A_n = B_m \times \Omega_m'$ for some $m \geq 1$, necessarily

$$A_n(\omega_1^*, \dots, \omega_m^*) = \begin{cases} \Omega_m' & \text{if } (\omega_1^*, \dots, \omega_m^*) \in B_m \\ \emptyset & \text{if } (\omega_1^*, \dots, \omega_m^*) \notin B_m. \end{cases}$$

But (4) ensures that $A_n(\omega_1^*, \dots, \omega_m^*) \neq \emptyset$. Thus, $(\omega_1^*, \dots, \omega_m^*) \in B_m$, whence $(\omega_1^*, \dots, \omega_m^*, \omega_{m+1}^*, \dots) \in B_m \times \Omega_m' = A_n$, $n \geq 1$. \square

Theorem 1 establishes that a sequence of probability spaces $(\Omega_i, \mathcal{F}_i, P_i)$, $i \geq 1$, engenders an **infinite-dimensional product measure space** $(\bigtimes_{i=1}^{\infty} \Omega_i, \bigtimes_{i=1}^{\infty} \mathcal{F}_i, \bigtimes_{i=1}^{\infty} P_i)$ such that for every $n = 1, 2, \dots$, denoting $\bigtimes_{i=1}^n P_i$ by P ,

$$P\left\{A_1 \times \dots \times A_n \times \left(\bigtimes_{n+1}^{\infty} \Omega_i\right)\right\} = \prod_{i=1}^n P_i\{A_i\}, \quad A_i \in \mathcal{F}_i, 1 \leq i \leq n.$$

The proof is based on the fact that if $\mathcal{G}_n = \{A : A = B_n \times \Omega_n'\}$ is the class of cylinders with n -dimensional bases $B_n \in \bigtimes_{i=1}^n \mathcal{F}_i$ and

$$\hat{P}_n\{A\} = (P_1 \times \dots \times P_n)\{B_n\},$$

then \hat{P}_n is a probability measure on the σ -algebra \mathcal{G}_n with $\hat{P}_n = \hat{P}_{n+1}|_{\mathcal{G}_n}$ and, moreover, if $P\{A\} = \lim_n \hat{P}_n\{A\}$ for $A \in \mathcal{G} = \bigcup_1^\infty \mathcal{G}_n$, then P is σ -additive on the algebra \mathcal{G} , whence there is a unique extension to $\sigma(\mathcal{G}) = \bigcup_1^\infty \mathcal{F}_i$.

The following question then poses itself: If $(\Omega, \mathcal{F}_i, P_i)$, $i \geq 1$, is a sequence of probability spaces with $\mathcal{F}_i \subset \mathcal{F}_{i+1}$ and $P_i = P_{i+1}|_{\mathcal{F}_i}$, is $P\{A\} = \lim_n P_n\{A\}$, $A \in \mathcal{G} = \bigcup_1^\infty \mathcal{F}_i$, necessarily σ -additive on the algebra \mathcal{G} ? The answer is, in general, negative; see Example 3. However, if $\Omega = R^\infty$ and \mathcal{F}_n is the class of cylinders of $(R^\infty, \mathcal{B}^\infty)$ with n -dimensional bases, the answer becomes affirmative.

Theorem 2. Let $(R^n, \mathcal{B}^n, P_n)$, $n \geq 1$, be a sequence of probability spaces with

$$P_{n+1}\{A_n \times R\} = P_n\{A_n\}, \quad A_n \in \mathcal{B}^n, n \geq 1, \quad (5)$$

and let $\tilde{\mathcal{B}}^n$ be the σ -algebra of cylinders in $(R^\infty, \mathcal{B}^\infty)$ with n -dimensional Borel bases. If $\mathcal{G} = \bigcup_1^\infty \tilde{\mathcal{B}}^n$ and for each $A = A_n \times \bigcup_{n+1}^\infty R$ with $A_n \in \mathcal{B}^n$,

$$P\{A\} = P_n\{A_n\}, \quad (6)$$

then P is σ -additive on the algebra \mathcal{G} and determines a unique probability measure extension \tilde{P} to \mathcal{B}^∞ .

PROOF. In view of Theorem 6.1.1, it suffices to prove that P is well defined and a probability measure on \mathcal{G} . The former is an immediate consequence of (5) and, clearly, P is nonnegative and additive on \mathcal{G} with $P\{R^\infty\} = 1$.

Let \mathcal{D}_n and $\tilde{\mathcal{D}}_n$ denote the classes of all sets of the form $J_1 \times \cdots \times J_n$ and $J_1 \times \cdots \times J_n \times R \times R \times \cdots$ respectively, where J_i signifies an interval of R , $1 \leq i \leq n$, i.e., $J_i = [a_i, b_i]$, $[a_i, b_i)$, (a_i, b_i) , or $(a_i, b_i]$ for $-\infty \leq a_i \leq b_i \leq \infty$. Then the classes \mathcal{H}_n and $\tilde{\mathcal{H}}_n$ of all finite unions of sets of \mathcal{D}_n and $\tilde{\mathcal{D}}_n$ respectively are algebras. If $\tilde{\mathcal{H}} = \bigcup_1^\infty \tilde{\mathcal{H}}_n$, then $\tilde{\mathcal{H}}$ is a subalgebra of \mathcal{G} , whence P is defined and additive on $\tilde{\mathcal{H}}$.

To check σ -additivity on $\tilde{\mathcal{H}}$, let $\{\tilde{A}_n, n \geq 1\}$ be a decreasing sequence of sets of $\tilde{\mathcal{H}}$ with $\inf_n P\{\tilde{A}_n\} = \varepsilon > 0$, whence $\tilde{A}_n = A_{m_n} \times R \times R \times \cdots$ for some $A_{m_n} \in \mathcal{H}_{m_n}$, and $m_{n+1} > m_n, n \geq 1$. Since P_n is a probability measure on \mathcal{H}_n , every interval $J_1 \times \cdots \times J_n$ of \mathcal{D}_n contains a closed subinterval whose P_n -measure is arbitrarily close to that of $J_1 \times \cdots \times J_n$. Thus, there is a closed set B_n (which is a finite union of closed intervals) with

$$P_{m_n}\{A_{m_n} - B_n\} < \frac{\varepsilon}{2^{n+1}}, \quad B_n \subset A_{m_n}. \quad (7)$$

Let $\tilde{B}_n = B_n \times R \times R \times \cdots$, whence (7) holds with $\tilde{A}_n, \tilde{B}_n, P$ replacing A_{m_n}, B_n, P_{m_n} respectively. Consequently, if $\tilde{C}_n = \tilde{B}_1 \cdot \tilde{B}_2 \cdots \tilde{B}_n$,

$$P\{\tilde{A}_n - \tilde{C}_n\} = P\left\{\bigcup_{j=1}^n (\tilde{A}_n - \tilde{B}_j)\right\} \leq P\left\{\bigcup_{j=1}^n (\tilde{A}_j - \tilde{B}_j)\right\} < \frac{\varepsilon}{2},$$

whence

$$P\{\tilde{C}_n\} > P\{\tilde{A}_n\} - \frac{\varepsilon}{2} \geq \frac{\varepsilon}{2}$$

and $\tilde{C}_n \neq \emptyset$. Let $\omega^{(n)} = (\omega_1^{(n)}, \omega_2^{(n)}, \dots) \in \tilde{C}_n$. Since $\tilde{C}_n \downarrow$, necessarily $\omega^{(n+p)} \in \tilde{C}_n \subset \tilde{B}_n$, implying $(\omega_1^{(n+p)}, \dots, \omega_{m_n}^{(n+p)}) \in B_n$, $p = 0, 1, \dots$. Choose a subsequence $\{n_{1k}\}$ of the positive integers for which $\omega_1^{(n_{1k})} \rightarrow$ a limit $\omega_1^{(0)}$ (finite or infinite) as $k \rightarrow \infty$. Likewise, there is a subsequence $\{n_{2k}\}$ of $\{n_{1k}\}$ with $\omega_2^{(n_{2k})} \rightarrow \omega_2^{(0)}$ as $k \rightarrow \infty$, etc. Then

$$\omega^{(n_{kk})} = (\omega_1^{(n_{kk})}, \omega_2^{(n_{kk})}, \dots) \rightarrow (\omega_1^{(0)}, \omega_2^{(0)}, \dots) \equiv \omega^{(0)}$$

and so as $k \rightarrow \infty$

$$(\omega_1^{(n_{kk})}, \dots, \omega_{m_n}^{(n_{kk})}) \rightarrow (\omega_1^{(0)}, \dots, \omega_{m_n}^{(0)}) \in B_n, \quad n \geq 1.$$

Therefore $\omega^{(0)} \in \tilde{B}_n \subset \tilde{A}_n$, $n \geq 1$, so that $\bigcap_1^\infty \tilde{A}_n \neq \emptyset$. Consequently, P is σ -additive on the algebra $\tilde{\mathcal{H}}$ with $P\{R^\infty\} = 1$, whence by Theorem 6.1.1, P has a unique extension \tilde{P} to $\mathcal{B}^\infty = \sigma(\tilde{\mathcal{H}})$. Clearly, \tilde{P} is σ -additive on \mathcal{G} , and since $\tilde{P} = P$ on $\tilde{\mathcal{H}}_n$, $\tilde{P} = P$ on $\sigma(\tilde{\mathcal{H}}_n) = \mathcal{B}_n$, $n \geq 1$, whence $\tilde{P} = P$ on $\mathcal{G} = \bigcup_1^\infty \mathcal{B}_n$. \square

Is it always possible to define a sequence of r.v.s $\{X_n, n \geq 1\}$ on some probability space (Ω, \mathcal{F}, P) such that the joint d.f.s of all finite subsets of r.v.s X_1, \dots, X_n coincide with d.f.s $F_{1, \dots, n}$ assigned *a priori*? The answer is yes provided the assignment is not internally contradictory.

A family $\{F_{1, \dots, n}(x_1, \dots, x_n)\}$ of n -dimensional d.f.s defined for all $n \geq 1$ will be called **consistent** if for all $n \geq 1$

$$F_{1, \dots, n}(x_1, \dots, x_n) = \lim_{x_{n+1} \rightarrow +\infty} F_{1, \dots, n+1}(x_1, \dots, x_{n+1}). \quad (8)$$

Theorem 3 (Kolmogorov Consistency Theorem). *If $\{F_{1, \dots, n}, n \geq 1\}$ is a consistent family of d.f.s, then there exists a probability measure P on $(R^\infty, \mathcal{B}^\infty)$ such that the d.f.s of the coordinate r.v.s X_1, \dots, X_n on $(R^\infty, \mathcal{B}^\infty, P)$ coincide with the preassigned d.f.s $F_{1, \dots, n}$, that is, such that if*

$$X_k(\omega) = \omega_k, \quad k = 1, 2, \dots, \quad \text{for } \omega = (\omega_1, \omega_2, \dots) \in R^\infty \quad (9)$$

then for all $n \geq 1$,

$$P\{X_1 < x_1, \dots, X_n < x_n\} = F_{1, \dots, n}(x_1, \dots, x_n). \quad (10)$$

PROOF. If A_n is an n -dimensional set of \mathcal{B}^n , define

$$P_n(A_n) = \int \cdots \int dF_{1, \dots, n}(t_1, \dots, t_n), \quad (11)$$

whence $(R^n, \mathcal{B}^n, P_n)$ is a probability space, $n \geq 1$. Employing the notation of (6) of Section 3, for all pairs of real numbers $a_i < b_i$, $1 \leq i \leq n$,

$$\begin{aligned} P_{n+1}\left\{\bigtimes_{i=1}^n [a_i, b_i] \times R\right\} &= \lim_{\substack{a_{n+1} \rightarrow -\infty \\ b_{n+1} \rightarrow \infty}} P_{n+1}\left\{\bigtimes_{i=1}^{n+1} [a_i, b_i]\right\} \\ &= \lim_{\substack{a_{n+1} \rightarrow -\infty \\ b_{n+1} \rightarrow \infty}} \Delta_{n+1}^{a, b} F_{1, \dots, n+1} = \Delta_n^{a, b} F_{1, \dots, n} = P_n\left\{\bigtimes_{i=1}^n [a_i, b_i]\right\} \end{aligned}$$

in view of (8). Hence, $P_{n+1}\{A_n \times R\} = P_n\{A_n\}$ for all $A_n \in \mathcal{B}^n$, $n \geq 1$, and so by Theorem 2 there is a probability measure P on $(R^\infty, \mathcal{B}^\infty)$ such that for all $A_n \in \mathcal{B}^n$

$$P\{A_n \times R \times R \times \dots\} = P_n\{A_n\}, \quad n \geq 1.$$

The coordinate functions defined by (9) are r.v.s and by (9) and (11)

$$\begin{aligned} P\{X_1 < x_1, \dots, X_n < x_n\} &= P\{\omega: \omega_1 < x_1, \dots, \omega_n < x_n\} \\ &= F_{1, \dots, n}(x_1, \dots, x_n). \end{aligned} \quad \square$$

Corollary 1. If $\{X_n, n \geq 1\}$ is a sequence of r.v.s on some probability space (Ω, \mathcal{F}, P) , there exists a sequence of coordinate r.v.s $\{X'_n, n \geq 1\}$ on the probability space $(R^\infty, \mathcal{B}^\infty, P')$ such that the joint d.f.s of X_1, \dots, X_n and X'_1, \dots, X'_n are identical for all $n \geq 1$.

Theorem 4. If $\{X_n, n \geq 1\}$ are r.v.s on some probability space (Ω, \mathcal{F}, P) , there exist two sequences of coordinate r.v.s $\{X''_n, n \geq 1\}$ and $\{X'_n, n \geq 1\}$ on $(R^\infty, \mathcal{B}^\infty)$ each having the same joint distributions as $\{X_n, n \geq 1\}$, i.e.,

$$F_{X_1, \dots, X_n} = F_{X'_1, \dots, X'_n} = F_{X''_1, \dots, X''_n}, \quad n \geq 1, \quad (12)$$

such that the stochastic processes

$$\{X''_n, n \geq 1\} \quad \text{and} \quad \{X'_n, n \geq 1\} \quad (13)$$

are independent of one another.

PROOF. For $\omega = (\omega_1, \omega_2, \dots) \in R^\infty$, define $Y_j(\omega) = \omega_j$, $j \geq 1$, and \hat{F}_n , \hat{P} by

$$\begin{aligned} \hat{F}_{2n} &= \hat{P}\{Y_1 < y_1, \dots, Y_{2n} < y_{2n}\} = F_{X_1, X_2, \dots, X_n}(y_1, y_3, \dots, y_{2n-1}) \\ &\quad \cdot F_{X_1, X_2, \dots, X_n}(y_2, y_4, \dots, y_{2n}), \\ \hat{F}_{2n-1} &= \hat{P}\{Y_1 < y_1, \dots, Y_{2n-1} < y_{2n-1}\} \\ &= F_{X_1, X_2, \dots, X_n}(y_1, y_3, \dots, y_{2n-1}) \\ &\quad \cdot F_{X_1, X_2, \dots, X_{n-1}}(y_2, y_4, \dots, y_{2n-2}) \end{aligned}$$

for $n \geq 1$. Then $\{\hat{F}_n, n \geq 1\}$ is a consistent sequence of d.f.s and hence determines a probability measure \hat{P} on $(R^\infty, \mathcal{B}^\infty)$ which clearly satisfies (12) and (13) if $X''_n = Y_{2n-1}$, $X'_n = Y_{2n}$, $n \geq 1$. \square

The r.v.s $X_n^*, n \geq 1$, defined by

$$X_n^* = X''_n - X'_n = Y_{2n-1} - Y_{2n}, \quad n \geq 1,$$

are called the **symmetrized** X_n . Given any sequence of r.v.s $\{X_n, n \geq 1\}$, in subsequent allusion to the symmetrized X_n , namely X_n^* , the distinction between X_n and X_n^* (which are probabilistically indistinguishable) will be glossed over and X_n^* will be written

$$X_n^* = X_n - X_n', \quad n \geq 1,$$

where $\{X_n', n \geq 1\}$ is independent of $\{X_n, n \geq 1\}$ and possesses the same joint distributions.

Since the $\{X_n^*, n \geq 1\}$ are symmetric r.v.s, it is frequently easier (especially for independent $\{X_n, n \geq 1\}$) to prove a result for $\{X_n^*\}$ rather than attempt a direct argument with $\{X_n\}$ (see Chapter 10).

The following examples complement the three series theorem, the second furnishing a typical exploitation of symmetrization. For any r.v.s $\{X_n, n \geq 1\}$ and positive constant c , define

$$s_n^2(c) = \sum_{j=1}^n \sigma^2(X_j I_{\{|X_j| \leq c\}}) = \sum_{j=1}^n (\mathbb{E} X_j^2 I_{\{|X_j| \leq c\}} - \mathbb{E}^2 X_j I_{\{|X_j| \leq c\}}). \quad (14)$$

EXAMPLE 1. Let $S_n = \sum_{i=1}^n X_i$, $n \geq 1$, where $\{X_n, n \geq 1\}$ are independent, symmetric r.v.s and let $s_n(c)$ be as in (14).

If for some $c > 0$, $s_n^2(c) \rightarrow \infty$, then

$$\overline{\lim}_{n \rightarrow \infty} \frac{S_n}{s_n(c)} \stackrel{\text{a.c.}}{=} \infty \stackrel{\text{a.c.}}{=} -\underline{\lim}_{n \rightarrow \infty} \frac{S_n}{s_n(c)}. \quad (15)$$

If $\sum_{n=1}^{\infty} \mathbb{P}\{|X_n| > c\} = \infty$, all $c > 0$, then

$$\overline{\lim}_{n \rightarrow \infty} S_n \stackrel{\text{a.c.}}{=} \infty \stackrel{\text{a.c.}}{=} -\underline{\lim}_{n \rightarrow \infty} S_n. \quad (16)$$

If $s_n^2(c) = O(1)$, all $c > 0$ and $\sum_{n=1}^{\infty} \mathbb{P}\{|X_n| > c\} < \infty$, some $c > 0$ then

$$S_n \text{ converges a.c.} \quad (17)$$

PROOF. Define

$$X_n^c = X_n I_{\{|X_n| \leq c\}}. \quad (18)$$

Then symmetry guarantees that $\mathbb{E} X_n^c = 0$, $n \geq 1$, and the hypothesis of (15) ensures that the uniformly bounded r.v.s $\{X_n^c, n \geq 1\}$ obey a central limit theorem (Theorem 9.1.1 or Exercise 9.1.1). Thus if Φ denotes the normal d.f., for all $b > 0$

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \bigcup_{j=n}^{\infty} \left[\sum_{i=1}^j X_i^c > b s_n(c) \right] \right\} \geq \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \sum_{i=1}^n X_i^c > b s_n(c) \right\} = 1 - \Phi(b) > 0,$$

implying by the Kolmogorov zero-one law that for all $b > 0$

$$\mathbb{P} \left\{ \overline{\lim}_{n \rightarrow \infty} \frac{1}{s_n(c)} \sum_{i=1}^n X_i^c > \frac{b}{2} \right\} \geq \mathbb{P} \left\{ \overline{\lim}_{n \rightarrow \infty} \frac{1}{s_n(c)} \sum_{i=1}^n X_i^c \geq b \right\} = 1.$$

Since $s_n(c) \rightarrow \infty$, Lemma 4.2.6 ensures

$$P\left\{\overline{\lim}_{n \rightarrow \infty} \frac{S_n}{s_n(c)} \geq \frac{b}{2}\right\} = 1, \quad (19)$$

and so, in view of the arbitrariness of b , (19) obtains with $b = \infty$. The analogous statement for the lower limit follows by symmetry.

Apropos of (16), via $|X_n| \leq |S_n| + |S_{n-1}|$ and the Borel-Cantelli theorem

$$2 \overline{\lim}_{n \rightarrow \infty} |S_n| \geq \overline{\lim}_{n \rightarrow \infty} |X_n| = \infty, \text{ a.c.,}$$

and by symmetry $\overline{\lim}_{n \rightarrow \infty} S_n = \infty = -\lim_{n \rightarrow \infty} S_n$, a.c. The final statement, (17), follows immediately from the three series theorem (Theorem 5.1.2). \square

Corollary 2. For independent, symmetric r.v.s $\{X_n\}$ either $\sum_1^n X_i$ converges a.c. or $\overline{\lim}_{n \rightarrow \infty} \sum_1^n X_i = \infty = -\lim_{n \rightarrow \infty} \sum_1^n X_i$, a.c.

EXAMPLE 2. Let $\{X_n, n \geq 1\}$ be independent r.v.s and $S_n = \sum_1^n X_i, n \geq 1$.

If $\sum_{n=1}^{\infty} P\{|X_n| > c\} = \infty$, all $c > 0$, then

$$\overline{\lim}_{n \rightarrow \infty} |S_n| = \infty, \quad \text{a.c.} \quad (20)$$

If for some $c > 0$, $\sum_{n=1}^{\infty} P\{|X_n| > c\} < \infty$, and $s_n^2(c) \rightarrow \infty$, then

$$\overline{\lim}_{n \rightarrow \infty} \frac{|S_n|}{s_n(c)} \stackrel{\text{a.c.}}{\longrightarrow} \infty. \quad (21)$$

Suppose that for some $c > 0$,

$$\sum_{n=1}^{\infty} P\{|X_n| > c\} < \infty \quad \text{and} \quad s_n^2(c) = O(1). \quad (22)$$

If $\overline{\lim}_{n \rightarrow \infty} \sum_{j=1}^n E X_j^c = \infty$ (resp. $\underline{\lim}_{n \rightarrow \infty} \sum_{j=1}^n E X_j^c = -\infty$), then $\overline{\lim}_{n \rightarrow \infty} S_n = \infty$ (resp. $\underline{\lim}_{n \rightarrow \infty} S_n = -\infty$), a.c. If $-\infty < \underline{\lim}_{n \rightarrow \infty} \sum_{j=1}^n E X_j^c \leq \overline{\lim}_{n \rightarrow \infty} \sum_{j=1}^n E X_j^c < \infty$, then $\overline{\lim}_{n \rightarrow \infty} |S_n| < \infty$, a.c.

PROOF. Since the proof of (16) did not utilize symmetry, (20) is immediate. To establish (21), let $\{X_n'', n \geq 1\}$ and $\{X_n', n \geq 1\}$ be i.i.d. stochastic processes on some probability space whose joint distributions (of X_1, \dots, X_n or X_1'', \dots, X_n'') are identical with those of X_1, \dots, X_n . The double prime will be deleted in what follows. Since, in the notation of (18),

$$\begin{aligned} 2\sigma_{X_n^c}^2 &= \sigma_{X_n^c - X_n'}^2 = E[X_n^c - X_n']^2 = \int_{\{|X_n| \leq c, |X_n'| \leq c\}} (X_n - X_n')^2 \\ &\quad + 2 \int_{\{|X_n| \leq c, |X_n'| > c\}} X_n^2 \end{aligned}$$

$$\begin{aligned}
&= \int_{\{|X_n - X'_n| \leq 2c\}} (X_n - X'_n)^2 + \int_{\{|X_n - X'_n| \leq 2c, |X'_n| > c\}} (X_n - X'_n)^2 \\
&\quad + 2 \int_{\{|X_n| \leq c, |X'_n| > c\}} X_n^2 + \int_{\{|X_n - X'_n| \leq 2c, |X'_n| \leq c, |X_n| > c\}} (X_n - X'_n)^2
\end{aligned}$$

necessarily

$$2\sigma_{X_n}^2 \leq \sigma_{(X_n - X'_n)^2}^2 + 2c^2 P\{|X_n| > c\},$$

$$\sigma_{(X_n - X'_n)^2}^2 \leq 2\sigma_{X_n}^2 + 8c^2 P\{|X_n| > c\},$$

and so $\sum_{j=1}^n \sigma_{(X_j - X'_j)^2}^2 \sim 2s_n^2(c)$. It follows from (15) that $\overline{\lim}_{n \rightarrow \infty} |S_n - S'_n|/s_n(c) = \infty$, a.c., and this, in turn, ensures the conclusion of (21).

Finally, the hypothesis of (22) guarantees that $\sum_{j=1}^n (X_j - E X_j)$ converges a.c. The final conclusion of (22) follows from the identity

$$\sum_{j=1}^n X_j = \sum_{j=1}^n (X_j - E X_j) + \sum_{j=1}^n E X_j,$$

while the prior ones follow upon replacing n in this identity by suitable subsequences $\{n_i\}$. \square

Since Lévy's inequality acquired a very simple form for symmetric r.v.s. (Corollary 3.35), symmetrization is especially useful in proving the following converse to Theorem 5.2.8.

Theorem 5. Let $\alpha p \geq 1$ and $S_n = \sum_{i=1}^n X_i$, $n \geq 1$, where $\{X, X_n, n \geq 1\}$ are non-degenerate i.i.d. random variables. If, for some $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} n^{\alpha p - 2} P\{|S_n| \geq \varepsilon n^{\alpha}\} < \infty, \quad (23)$$

then $\alpha > 0$ and $E|X|^p < \infty$. Moreover, $E X = 0$ if either $\alpha < 1$ or $\alpha = 1$ and (23) holds for all $\varepsilon > 0$.

PROOF. By Theorem 4, there exists a sequence $\{X', X'_n, n \geq 1\}$ such that $\{X, X_n, n \geq 1\}$ and $\{X', X'_n, n \geq 1\}$ are i.i.d. Set $Y = X - X'$, $Y_n = X_n - X'_n$, and $T_n = \sum_{j=1}^n Y_j$, $n \geq 1$. Then, for some $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} n^{\alpha p - 2} P\{|T_n| \geq 2\varepsilon n^{\alpha}\} < \infty, \quad (24)$$

so that, as $m \rightarrow \infty$,

$$\sum_{n=m}^{2m} n^{\alpha p - 2} P\{|T_n| \geq 2\varepsilon(1 + 2^{\alpha})m^{\alpha}\} \leq \sum_{n=m}^{2m} n^{\alpha p - 2} P\{|T_n| \geq 2\varepsilon n^{\alpha}\} = o(1). \quad (25)$$

Now, if $T_n^* = \max_{1 \leq j \leq n} |T_j|$ and $Y_n^* = \max_{1 \leq j \leq n} |Y_j|$, for any $\delta > 0$, via Lévy's inequality,

$$2P\{|T_n| \geq \delta\} \geq P\{T_n^* \geq \delta\} \geq P\{Y_n^* \geq 2\delta\} = 1 - P^n\{|Y| < 2\delta\}$$

$$\begin{aligned}
 &= P\{|Y| \geq 2\delta\} \sum_{j=0}^{n-1} P^j\{|Y| < 2\delta\} \\
 &\leq P\{|Y| \geq 2\delta\} \sum_{j=0}^{n-1} [1 - 2P\{|T_j| \geq \delta\}].
 \end{aligned} \tag{26}$$

Consequently, choosing $\delta = \eta n^\alpha$, where $\eta = 2\varepsilon$, (24) implies that

$$\sum_{n=1}^{\infty} n^{\alpha p - 2} P\{|Y| \geq 2\eta n^\alpha\} \sum_{j=0}^{n-1} [1 - 2P\{|T_j| \geq \eta n^\alpha\}] < \infty, \tag{27}$$

while (25) and the initial inequality of (26) ensure that, as $m \rightarrow \infty$,

$$m^{\alpha p - 1} P\{T_m^* \geq \eta(1 + 2^\alpha)m^\alpha\} \leq \sum_{n=m}^{2m} n^{\alpha p - 2} P\{T_n^* \geq \eta(1 + 2^\alpha)m^\alpha\} = o(1).$$

Since $\alpha p \geq 1$, as $m \rightarrow \infty$,

$$P\{T_m^* \geq \eta' m^\alpha\} = o(1), \tag{28}$$

where $\eta' = \eta(1 + 2^\alpha) = 2\varepsilon(1 + 2^\alpha)$. In view of $P\{T_m^* > 0\} \geq P\{T_1^* > 0\} = P\{|Y| \neq 0\} > 0$, necessarily $\alpha \geq 0$, so that $\alpha p \geq 1$ guarantees $\alpha > 0$.

Clearly, (23) holds when ε is replaced by $\varepsilon(1 + 2^\alpha)$, whence (27) likewise holds with η replaced by η' . Thus, since $P\{|T_j| \geq \eta' n^\alpha\} \leq P\{T_n^* \geq \eta' n^\alpha\}$ for $1 \leq j \leq n$, (28) ensures that the second sum in (27) with η replaced by η' is $n + o(n)$, implying that

$$\sum_{n=1}^{\infty} n^{\alpha p - 1} P\{|Y| \geq 2\eta' n^\alpha\} < \infty,$$

and hence also that

$$\infty > \int_1^{\infty} t^{\alpha p - 1} P\{|Y| \geq 2\eta' t^\alpha\} dt = C \int_{2\eta'}^{\infty} u^{\alpha p - 1} P\{|Y| \geq u\} du$$

for some positive constant C , yielding $E|Y|^\alpha < \infty$. Hence, via independence of X and X' ,

$$\infty > E|X - X'|^\alpha \geq E[(X - K)^+]^\alpha P\{X' < K\},$$

and choosing K large enough so that $P\{X' < K\} > 0$, necessarily $E(X^+)^\alpha < \infty$. Similarly, $E(X^-)^\alpha < \infty$ and so $E|X|^\alpha < \infty$.

When $\alpha \leq 1$, clearly $p \geq 1$, so that $E|X| < \infty$. By Theorem 5.2.8, for all $\varepsilon' > 0$,

$$\sum_{n=1}^{\infty} n^{\alpha p - 2} P\{|S_n - nE X| \geq \varepsilon' n^\alpha\} < \infty,$$

which, in conjunction with (23), yields

$$\sum_{n=1}^{\infty} n^{\alpha p - 2} P\{|E X| \geq 2\varepsilon n^{\alpha-1}\} < \infty.$$

Thus, $E X = 0$ if either $\alpha < 1$ or $\alpha = 1$ and (23) holds for all $\varepsilon > 0$. \square

Actually, $\alpha > 1/2$ in Theorem 5, since the alternative, $\alpha \leq 1/2$, requires $p \geq 2$, whence the Central Limit Theorem for i.i.d. random variables (Corollary 9.1.2) implies that

$$P\{|T_n| \geq 2\epsilon n^{\alpha}\} \geq P\{|T_n| \geq 2\epsilon n^{1/2}\} \rightarrow 2[1 - \Phi(2\epsilon)] > 0,$$

which is incompatible with (24).

EXAMPLE 3. Let $\Omega = \{1, 2, \dots\}$, $\mathcal{F}_n = \sigma(\{1\}, \{2\}, \dots, \{n\})$, $P_n(A) = 0$ or 1 according as A is a finite set or not. Since no two infinite sets of \mathcal{F}_n are disjoint, P_n is a probability measure on \mathcal{F}_n and moreover, $P_{n+1}|_{\mathcal{F}_n} = P_n$. However, $P\{A\} = \lim P_n\{A\}$ is not σ -additive since

$$P\{\{m\}\} = \lim_{n \rightarrow \infty} P_n\{\{m\}\} = 0, \quad m = 1, 2, \dots$$

whereas

$$P\{\Omega\} = \lim_{n \rightarrow \infty} P_n\{\Omega\} = 1 \neq \sum_{n=1}^{\infty} P\{\{n\}\}$$

EXERCISES 6.4

- Verify that it is possible to define a sequence of independent r.v.s $\{X_n\}$ with specified d.f.s F_n on a probability space.
- Let $(\Omega, \mathcal{B}, \nu)$ be the probability space consisting of Lebesgue measure on the Borel subsets of $[0, 1]$. Each ω in $[0, 1]$ may be written in binary expansion as $\omega = X_1 X_2 \dots = \sum_1^{\infty} 2^{-n} X_n$, where $X_n = X_n(\omega) = 0$ or 1 and this expansion is unique except for a set of ω of the form $m/2^n$ which is countable and hence has probability (Lebesgue measure) zero. For definiteness regarding such ω , only the expansion consisting of infinitely many ones will be used. Show that for all $n \geq 1$, $\{\omega: X_n(\omega) = 1\} \in \mathcal{B}$ and that $\{X_n, n \geq 1\}$ is a sequence of independent r.v.s. Describe the r.v.s Y_n where $\omega = \sum_{n=1}^{\infty} r^{-n} Y_n$ and r is an integer > 2 .
- In Exercise 2, define $Z_n = Z_n(\omega) = 1$ or -1 according as the integer i for which $(i-1)/2^n \leq \omega < i/2^n$ is odd or even. Show that any two but not all three of $Z_1, Z_2, Z_1 \cdot Z_2$ are independent. What is the relation between Z_n and X_n ? The Z_n , $n \geq 1$, are known as the **Rademacher** functions.
- Let $\Omega_0 = \{0, 1\}$, $\mathcal{F}_0 = \text{class of subsets of } \Omega_0$, $\mu_0(\{1\}) = \frac{1}{2} = \mu_0(\{0\})$, and define $\mu_i \equiv \mu_0$, $i \geq 1$, and $(\Omega, \mathcal{F}, \mu) = (\Omega_0^{\infty}, \mathcal{F}_0^{\infty}, \bigtimes_1^{\infty} \mu_i)$. Prove the following. (i) Each point of Ω is an \mathcal{F} -set of μ -measure zero. (ii) The set D of all points of Ω with only finitely many coordinates equal to 1 has μ -measure 0. (iii) Define $\Omega' = \Omega - D$, $\mathcal{F}' = \mathcal{F}\Omega'$, $\mu'\{A \cap \Omega'\} = \mu\{A\}$, $A \in \mathcal{F}$. (iv) For each $\omega \in \Omega'$, $z(\omega) = \sum_{i=1}^{\infty} \omega_i 2^{-i}$ is a 1-1 map of Ω' onto $Z = (0, 1)$. (v) If $C = \{z: 0 \leq z \leq b \leq 1\}$, $A = \{\omega: z(\omega) \in C\}$, then A is measurable and $\mu'\{A\} = b - a$. Hint: It suffices to take binary rational a, b . (vi) For any Borel subset B of Z , the set $A = \{\omega: z(\omega) \in B\}$ is measurable and $\mu'\{A\} = \text{Lebesgue measure of } B$.
- Let $T = [a, b]$, $-\infty \leq a < b \leq \infty$, be an index set, $\Omega = R^T = \{\omega: \omega = \omega(t), t \in T\}$ = space of real functions on T . If B_n is a Borel set of R^n , then $A = \{\omega(t): (\omega(t_1), \dots, \omega(t_n)) \in B_n\}$ is a Borel cylinder set for any choice of distinct t_1, \dots, t_n in T . The class

- \mathcal{A}^T of Borel cylinder sets is an algebra. Define $\mathcal{B}^T = \sigma(\mathcal{A}^T)$ and let $T = [a, b]$, $-\infty \leq a < b \leq \infty$. Do the sets $\{\omega(t): \omega(t) \text{ is bounded on } T\}$, $\{\omega(t): \omega(t) \text{ is continuous on } T\}$ belong to \mathcal{B}^T ? If $A^* = \{\omega(t): \omega(t_i) \in B_i, i = 1, 2, \dots\}$, and \mathcal{A}^* is the class of all such set A^* (as t_i and B_i vary), set $\mathcal{B}^* = \sigma(\mathcal{A}^*)$. Is $\mathcal{B}^T = \mathcal{B}^*$?
6. If $S_n = \sum_1^n X_i$, where $\{X_i, X_n, n \geq 1\}$ are i.i.d. then $\sum_1^\infty n^{-1} P\{|S_n| > \varepsilon n\} < \infty$ all $\varepsilon > 0$ iff $E X = 0$. Hint: Sufficiency is contained in Theorem 5.2.7. For necessity, define $S_n^* = \sum_1^n X_j^*$ where $\{X_j^*, n \geq 1\}$ is the symmetrized sequence. The hypothesis ensures convergence of $\sum_1^\infty n^{-1} P\{|S_n^*| > n\varepsilon\}$ and hence also $\sum_1^\infty n^{-1} P\{\max_{1 \leq j \leq n} |X_j^*| > n\varepsilon\}$, $\varepsilon > 0$.

6.5 Absolute Continuity of Measures, Distribution Functions; Radon–Nikodym Theorem

Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and T an arbitrary nonempty set. The **essential supremum** g of a family $\{g_t, t \in T\}$ of measurable functions from Ω into $R = [-\infty, \infty]$, denoted by $\text{esup}_{t \in T} g_t$, is defined by the properties:

- i. g is measurable,
- ii. $g \geq g_t$, a.e., for each $t \in T$,
- iii. For any h satisfying (i) and (ii), $h \geq g$, a.e.

Clearly, if such a g exists, it is unique in the sense that two such essential suprema of the same family are equal a.e.

Lemma 1. *Let $(\Omega, \mathcal{F}, \mu)$ be a σ -finite measure space and $\{g_t, t \in T\}$ a nonempty family of real, measurable functions. Then there exists a countable subset $T_0 \subset T$ such that*

$$\sup_{t \in T_0} g_t = \text{esup}_{t \in T} g_t.$$

PROOF. Since μ is σ -finite, it suffices to prove the theorem when μ is finite; moreover, by considering $\tan^{-1} g_t$, if necessary, it may be supposed that $|g_t| \leq C < \infty$ for all $t \in T$. Let \mathcal{J} signify the class of all countable subsets $I \subset T$ and set

$$\alpha = \sup_{I \in \mathcal{J}} E \left(\sup_{t \in I} g_t \right),$$

whence α is finite. Choose $I_n \in \mathcal{J}$, $n \geq 1$ for which $\alpha = \sup_{n \geq 1} E(\sup_{t \in I_n} g_t)$ and let $T_0 = \bigcup_1^\infty I_n$. Then T_0 is a countable subset of T and clearly $\alpha = E[\sup_{t \in T_0} g_t]$. The measurable function $g = \sup_{t \in T_0} g_t$ satisfies (ii) since otherwise for some $t \in T$ necessarily $\alpha < E \max(g, g_t) \leq \alpha$. Obviously, (iii) holds, and so $g = \text{esup}_{t \in T} g_t$. \square

Definition. If $(\Omega, \mathcal{F}, \mu_i)$, $i = 1, 2$, are two measure spaces and $\mu_1\{A\} = 0$ whenever $\mu_2\{A\} = 0$, then μ_1 is said to be **absolutely continuous with respect to μ_2** or simply **μ_2 -continuous**. If, rather $\mu_1\{N^c\} = 0$ for some set $N \in \mathcal{F}$ with $\mu_2\{N\} = 0$, then μ_1 is called **μ_2 -singular** (or the pair μ_1, μ_2 is dubbed singular).

If g is a nonnegative integrable function on $(\Omega, \mathcal{F}, \mu)$, the indefinite integral $v_g\{A\} = \int_A g d\mu$ is absolutely continuous relative to μ . This is tantamount to saying that the integral of a nonnegative function g over a set A of measure zero has the value zero (Exercise 6.2.2). The Radon–Nikodym theorem asserts under modest assumptions that the indefinite integral v_g is the prototype of a measure absolutely continuous with respect to μ . The crucial step in proving this is

Lemma 2. Let $(\Omega, \mathcal{F}, \mu)$ be a σ -finite measure space and v a σ -finite measure on \mathcal{F} , and let \mathcal{H} denote the family of all measurable functions $h \geq 0$ satisfying $\int_A h d\mu \leq v\{A\}$, $A \in \mathcal{F}$. Then

$$v\{A\} = \psi\{A\} + \int_A g d\mu, \quad A \in \mathcal{F} \quad (1)$$

where ψ is a μ -singular measure and

$$g = \operatorname{esup}_{h \in \mathcal{H}} h. \quad (2)$$

PROOF. Since μ and v are σ -finite, it suffices by a standard argument to consider the case where both are finite and, furthermore, the trivial case $\mu \equiv 0$ may be eliminated. According to Lemma 1, there exists a sequence $h_n \in \mathcal{H}$, $n \geq 1$, for which $g = \operatorname{esup}_{h \in \mathcal{H}} h = \sup_{n \geq 1} h_n$. Now if $h_1, h_2 \in \mathcal{H}$, then $h = \max(h_1, h_2) \in \mathcal{H}$ since

$$\int_A h d\mu = \int_{A[h_1 \geq h_2]} h_1 d\mu + \int_{A[h_1 < h_2]} h_2 d\mu \leq v\{A\}, \quad A \in \mathcal{F},$$

and so it may be supposed that $h_n \leq h_{n+1}$, $n \geq 1$. Then $g = \lim_n h_n$, whence by the monotone convergence theorem

$$\int_A g d\mu \leq v\{A\}, \quad A \in \mathcal{F}.$$

Consequently, ψ as defined by (1) is a measure.

Next, for $n \geq 1$ and $A \in \mathcal{F}$ with $\mu\{A\} > 0$,

$$\mathcal{D}_n(A) = \left\{ B \in \mathcal{F} : B \subset A, \psi\{B\} < \frac{1}{n} \mu\{B\} \right\}$$

is nonempty; otherwise, the choice $h_0 = (1/n)I_A$ would guarantee for all $B \in \mathcal{F}$ that

$$\int_B h_0 d\mu = \frac{1}{n} \mu\{AB\} \leq \psi\{AB\} \leq \psi\{B\} = v\{B\} - \int_B g d\mu,$$

implying $h_0 + g \in \mathcal{X}$ and thus violating $g = \operatorname{esup}_{h \in \mathcal{H}} h$.

Choose $B_{1,n} \in \mathcal{D}_n(\Omega)$ with

$$\mu\{B_{1,n}\} \geq \frac{1}{2} \sup\{\mu\{B\}: B \in \mathcal{D}_n(\Omega)\}$$

$= \alpha_{1,n}$ (say). If $\mu\{B_{1,n}^c\} = 0$, stop; otherwise, choose $B_{2,n} \in \mathcal{D}_n(B_{1,n}^c)$ with

$$\mu\{B_{2,n}\} \geq \frac{1}{2} \sup\{\mu\{B\}: B \in \mathcal{D}_n(B_{1,n}^c)\}$$

$= \alpha_{2,n}$ (say). If $\mu\{B_{1,n}^c \cdot B_{2,n}^c\} = 0$, stop; otherwise, choose $B_{3,n} \in \mathcal{D}_n(B_{1,n}^c \cdot B_{2,n}^c)$ with

$$\mu\{B_{3,n}\} \geq \frac{1}{2} \sup\{\mu\{B\}: B \in \mathcal{D}_n(B_{1,n}^c \cdot B_{2,n}^c)\} = \alpha_{3,n}$$

and so on. If the process terminates at some finite step k_n , set $B_{j,n} = \emptyset$, $j > k_n$.

Since $\mathcal{D}_n(A_1) \subset \mathcal{D}_n(A_2)$ for $A_1 \subset A_2$, necessarily $B_{j,n} \in \mathcal{D}_n(\Omega)$ for $j \geq 1$ if $B_{j,n} \neq \emptyset$, and since $\mathcal{D}_n(A)$ is closed under countable disjoint unions, $M_n = \bigcup_{j=1}^{\infty} B_{j,n} \in \mathcal{D}_n(\Omega)$, $n \geq 1$. Now, if $\mu\{M_n^c\} > 0$, for some $n \geq 1$, there exists some $D \in \mathcal{D}_n(M_n^c)$, whence $\mu\{D\} > 0$. Moreover, $\alpha_{m,n} \rightarrow 0$ as $m \rightarrow \infty$ via disjointness of $\{B_{j,n}, j \geq 1\}$ and finiteness of μ . However, for all m

$$\begin{aligned} 2\alpha_{m,n} &= \sup \left\{ \mu\{B\}: B \in \mathcal{D}_n \left(\bigcap_{j=1}^{m-1} B_{j,n}^c \right) \right\} \\ &\geq \sup \{ \mu\{B\}: B \in \mathcal{D}_n(M_n^c) \} \geq \mu\{D\} > 0, \end{aligned}$$

a contradiction. Thus, $\mu\{M_n^c\} = 0$, $n \geq 1$, and $\psi\{M_n\} < (1/n)\mu\{M_n\} = (1/n)\mu\{\Omega\}$. Consequently,

$$\begin{aligned} \psi \left\{ \bigcap_{n=1}^{\infty} M_n \right\} &\leq \lim_{n \rightarrow \infty} \psi\{M_n\} = 0 \\ \mu \left\{ \left(\bigcap_{n=1}^{\infty} M_n \right)^c \right\} &= \mu \left\{ \bigcup_{n=1}^{\infty} M_n^c \right\} \leq \sum_{n=1}^{\infty} \mu\{M_n^c\} = 0. \quad \square \end{aligned}$$

Corollary 1 (Lebesgue Decomposition Theorem). *If μ, v are σ -finite measures on a measurable space (Ω, \mathcal{F}) , there exist two uniquely determined measures λ_1, λ_2 such that $v = \lambda_1 + \lambda_2$, where λ_2 is μ -continuous and λ_1 is μ -singular.*

PROOF. It suffices to verify uniqueness. Let $\lambda_1 + \lambda_2 = \lambda'_1 + \lambda'_2$, where λ_1, λ'_1 are μ -singular and λ_2, λ'_2 are μ -continuous. If $\lambda_1 \neq \lambda'_1$, there exists $A \in \mathcal{F}$ with $\mu\{A\} = 0$ and $\lambda_1\{A\} \neq \lambda'_1\{A\}$. But then $\lambda_2\{A\} \neq \lambda'_2\{A\}$, violating absolute continuity. \square

The Lebesgue decomposition may be used to prove

Theorem 1 (Radon–Nikodym). *If v_1, v_2, μ are σ -finite measures on a measurable space (Ω, \mathcal{F}) with v_i being μ -continuous, $i = 1, 2$, and if $v = v_1 - v_2$ is well defined on \mathcal{F} (i.e., $v_1\{\Omega\}$ and $v_2\{\Omega\}$ are not both ∞) then there exists an \mathcal{F} -*

measurable function g , finite a.e. $[\mu]$, such that

$$v\{A\} = \int_A g \, d\mu, \quad A \in \mathcal{F}, \quad (3)$$

and g is unique up to sets of μ -measure zero.

PROOF. Let g_i and ψ_i be defined as in Lemma 2, $i = 1, 2$. Then both v_i and $\int_A g_i \, d\mu$ are μ -continuous and hence also ψ_i , $i = 1, 2$. Since according to Lemma 2, ψ_i is μ -singular, $i = 1, 2$, necessarily (Exercise 1) $\psi_i \equiv 0$, whence

$$v_i\{A\} = \int_A g_i \, d\mu, \quad A \in \mathcal{F}, \quad i = 1, 2.$$

Moreover, $g = g_1 - g_2$ is \mathcal{F} -measurable and so

$$v\{A\} = v_1\{A\} - v_2\{A\} = \int_A g \, d\mu, \quad A \in \mathcal{F},$$

which is (3). In proving uniqueness, it may be assumed that μ is finite. If g^* is any other \mathcal{F} -measurable function satisfying (3), then for any $C > 0$

$$A = \{C > g^* > g > -C\} \in \mathcal{F},$$

whence

$$\int_A g^* \, d\mu = v\{A\} = \int_A g \, d\mu,$$

necessitating $\mu\{A\} = 0$, all $C > 0$ and hence $\mu\{g^* > g\} = 0$. Analogously, $\mu\{g^* < g\} = 0$ and so $g^* = g$, a.e. Finally, when v is finite, g is μ -integrable and hence finite a.e. $[\mu]$, whence the latter also obtains when v is σ -finite. \square

Corollary 2. $|v\{A\}| < \infty$ for all $A \in \mathcal{F}$ iff g is μ -integrable and v is a measure iff $g \geq 0$ a.e. $[\mu]$.

A function g defined by (3) is called the **Radon–Nikodym derivative of v with respect to μ** and is denoted, in suggestive fashion, by $dv/d\mu$. Thus, if v is a (well-defined) difference of two μ -continuous, σ -finite measures, (3) may be restated as

$$v\{A\} = \int_A \frac{dv}{d\mu} \, d\mu, \quad A \in \mathcal{F}. \quad (3')$$

Theorem 2. Let μ be a σ -finite measure and v a μ -continuous, σ -finite measure on the measurable space (Ω, \mathcal{F}) . If X is an \mathcal{F} -measurable function whose integral $\int_{\Omega} X \, dv$ exists, then for every $A \in \mathcal{F}$,

$$\int_A X \, dv = \int_A X \frac{dv}{d\mu} \, d\mu \quad (4)$$

PROOF. It may be supposed that μ is finite and (via $X = X^+ - X^-$) that $X \geq 0$. Let \mathcal{H} be the class of nonnegative \mathcal{F} -measurable functions for which (4) obtains. Then \mathcal{H} is a monotone system which contains the indicator functions of all sets in \mathcal{F} . By Theorem 1.4.3, \mathcal{H} contains all nonnegative \mathcal{F} -measurable functions. \square

Corollary 3. If v, μ, λ are σ -finite measures on a measurable space (Ω, \mathcal{F}) with v being μ -continuous and μ being λ -continuous, then v is λ -continuous and

$$\frac{dv}{d\lambda} = \frac{dv}{d\mu} \cdot \frac{d\mu}{d\lambda}, \quad \text{a.e. } [\lambda].$$

PROOF. Clearly, v is λ -continuous and $dv/d\mu$ is \mathcal{F} -measurable with $\int_{\Omega} (dv/d\mu)d\mu$ extant. Thus, by Theorem 2, for all $A \in \mathcal{F}$

$$\int_A \frac{dv}{d\mu} \frac{d\mu}{d\lambda} d\lambda = \int_A \frac{dv}{d\mu} d\mu = v(A) = \int_A \frac{dv}{d\lambda} d\lambda,$$

which is tantamount to the conclusion. \square

If F is a d.f. on R for which there exists a Borel function f on $(-\infty, \infty)$ with

$$F(x) = \int_{(-\infty, x]} f(t)dt = \int_{-\infty}^x f(t)dt, \quad -\infty < x < \infty \quad (5)$$

where dt signifies Lebesgue measure on (R, \mathcal{B}) , then, as noted in Section 1.6, F is said to be absolutely continuous and f is called the density function of F . In particular, if F is the d.f. attached to some r.v. X on a probability space, then f is also called the density of X . Clearly, when the f of (5) exists, it is unique to within sets of Lebesgue measure zero.

Corollary 4. If X is a r.v. on a probability space (Ω, \mathcal{F}, P) with d.f. F and density f , and g is a Borel function on $R = [-\infty, \infty]$ such that $E g(X)$ exists, then for any linear Borel set B

$$\int_{[X \in B]} g(X) dP = \int_B g(t)f(t)dt \quad (6)$$

and, in particular,

$$P\{X \in B\} = \int_B f(t)dt. \quad (7)$$

PROOF. Let μ denote the restriction to \mathcal{B} of the Lebesgue-Stieltjes measure determined by F . From (5), for $-\infty < a < b < \infty$

$$\int_{[a, b]} f(t)dt = F(b) - F(a) = F(b-) - F(a-),$$

and so by the uniqueness of the restriction of Lebesgue–Stieltjes measure to \mathcal{B}

$$\mu\{B\} = \int_B f(t)dt, \quad B \in \mathcal{B}.$$

Thus, μ is absolutely continuous with respect to Lebesgue measure, whence by Theorem 2

$$\begin{aligned} \int_B g(t)f(t)dt &= \int_B g(t)d\mu(t) = \int_{-\infty}^{\infty} g(t)I_B(t)d\mu(t) \\ &= E I_B(X)g(X) = \int_{\{X \in B\}} g(X) dP, \end{aligned}$$

recalling Corollary 6.2.1. \square

Analogously, if (i) F is an n -dimensional d.f. or (ii) F is the joint d.f. of r.v.s X_1, \dots, X_n on some probability space and there exists a Borel function f on R^n with

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} f(t_1, \dots, t_n)dt_1 \cdots dt_n, \quad (8)$$

where $dt_1 \cdots dt_n$ signifies Lebesgue measure on (R^n, \mathcal{B}^n) , then F is declared absolutely continuous and f is called the density function of F in case (i) and the joint density of the r.v.s X_1, \dots, X_n in case (ii). Again, f is unique to within sets of n -dimensional Lebesgue measure zero. Under (ii), if g is any Borel function on R^n for which $E g(X_1, \dots, X_n)$ exists, then for any $B \in \mathcal{B}^n$

$$\int_{\{(X_1, \dots, X_n) \in B\}} g(X_1, \dots, X_n) dP = \int_B g(t_1, \dots, t_n) f(t_1, \dots, t_n) dt_1 \cdots dt_n \quad (9)$$

and, in particular,

$$P\{(X_1, \dots, X_n) \in B\} = \int_B f(t_1, \dots, t_n) dt_1 \cdots dt_n. \quad (10)$$

EXERCISES 6.5

- If ψ and μ are measures such that ψ is both μ -continuous and μ -singular, then $\psi \equiv 0$.
- Two measures μ, ν are called **equivalent**, denoted $\mu \equiv \nu$, if each is absolutely continuous with respect to the other. Verify that this is an equivalence relation. If $(\Omega, \mathcal{F}, \mu)$ is a probability space, X_i is a nonnegative \mathcal{L}_1 random variable and μ_i is the indefinite integral of X_i , $i = 1, 2$, then, if $\mu([X_1 = 0] \Delta [X_2 = 0]) = 0$, the two indefinite integrals are equivalent measures.
- If $(\Omega, \mathcal{F}, \mu_i)$ is a measure space, $i = 1, 2$, then μ_1 is absolutely continuous relative to $\mu_1 + \mu_2$.
- If $F(x; a, b) = (b - a)^{-1}(x - a)$, $a \leq x \leq b$, the corresponding measure $F\{\cdot\}$ is absolutely continuous relative to Lebesgue measure with Radon–Nikodym derivative $f(x) = (b - a)^{-1}I_{[a \leq x \leq b]}$ and F is called the **uniform** distribution on

- [a, b]. When $a = 0, b = 1$, $F(\cdot)$ coincides with Lebesgue measure on $[0, 1]$ and a r.v. with d.f. F is said to be **uniformly distributed**. Show that if X_1, \dots, X_n are i.i.d. uniformly distributed r.v.s, the measure determined by the d.f. of $X = (X_1, \dots, X_n)$ is the n -dimensional Lebesgue measure on the hypercube $0 \leq x_i \leq 1, 1 \leq i \leq n$, of \mathbb{R}^n .
5. A completely additive set function v on a measurable space (Ω, \mathcal{F}) which assumes at most one of the values $+\infty$ and $-\infty$ and satisfies $v(\emptyset) = 0$ is sometimes called a **signed measure**. If v is a signed measure on a measurable space (Ω, \mathcal{F}) and $v^+(A) = \sup\{v(B): A \supset B \in \mathcal{F}\}$, then v^+ is a measure satisfying $v^+(A) \geq v(A), A \in \mathcal{F}$. Likewise, $v^-(A) = -\inf\{v(B): A \supset B \in \mathcal{F}\}$ is a measure with $v^-(A) \geq -v(A)$; the measures v^+ and v^- are called the upper and lower variations respectively and the representation $v = v^+ - v^-$ is the **Jordan decomposition** of v . If v is σ -finite, so are v^+ and v^- .
 6. If $v = v^+ - v^-$ is the Jordan decomposition (Exercise 5) of the signed measure v , then $\bar{v} = v^+ + v^-$ is a measure called the **total variation** of v . Clearly, $|v(A)| \leq \bar{v}(A), A \in \mathcal{F}$.
 7. If v is a signed measure on a measurable space (Ω, \mathcal{F}) with total variation \bar{v} (see Exercise 6) and X is integrable relative to \bar{v} , one may define $\int X dv = \int X dv^+ - \int X dv^-$. Prove that if v is finite, $\bar{v}(A) = \sup\{|\int_A X dv|: X \text{ is measurable and } |X| \leq 1\}$.
 8. Let V denote a linear functional on $\mathcal{L}_p(\Omega, \mathcal{F}, \mu)$, i.e., range $V = (-\infty, \infty)$ and $V(af + bg) = aV(f) + bV(g)$ for all $f, g \in \mathcal{L}_p$. It is continuous if $V(f_n) \rightarrow V(f)$ whenever $\|f_n - f\|_p \rightarrow 0$, and V is bounded if $|V(f)| \leq C_p \|f\|_p$ for all $f \in \mathcal{L}_p$, where C_p is a finite constant. Prove that a continuous linear functional is bounded. *Hint:* Otherwise, there exist $f_n \in \mathcal{L}_p$ with $|V(f_n)| > n \|f_n\|_p$, whence if $g_n = f_n/(n \|f_n\|_p)$, $\|g_n\|_p = o(1)$ but $|V(g_n)| > 1$.
 9. If $(\Omega, \mathcal{F}, \mu)$ is a σ -finite measure space and V is a continuous linear functional on $\mathcal{L}_p(\Omega, \mathcal{F}, \mu)$, $p > 1$, there exists $g \in \mathcal{L}_q$, where $(1/p) + (1/q) = 1$, such that $V(f) = \int f \cdot g d\mu$ for all $f \in \mathcal{L}_p$. This is known as the (Riesz) representation theorem. *Hint:* Let μ be finite. Since $I_A \in \mathcal{L}_p(\Omega, \mathcal{F}, \mu)$ for all $A \in \mathcal{F}$, a set function on \mathcal{F} is defined by $v(A) = V(I_A)$. It is finitely additive and, moreover, σ -additive by continuity of V and the fact that $V(0) = 0$. Further, v is finite since V is bounded (Exercise 8); v is absolutely continuous with respect to μ . By the Radon–Nikodym theorem, there is an \mathcal{F} -measurable g , finite a.e., with $V(I_A) = v(A) = \int_A g d\mu$.
 10. Set functions $\{v_n, n \geq 1\}$ on $(\Omega, \mathcal{F}, \mu)$ are **uniformly absolutely continuous** relative to μ if for all $\varepsilon > 0$, $\mu(A) < \delta_\varepsilon$ implies $|v_n(A)| < \varepsilon$ for all $n \geq 1$. The sequence $\{v_n, n \geq 1\}$ is **equicontinuous from above at \emptyset** if for all $\varepsilon > 0$ and $A_m \downarrow \emptyset$, $|v_n(A_m)| < \varepsilon$ for all $n \geq 1$ whenever $m \geq m_\varepsilon$. Prove that if measures $\{v_n, n \geq 1\}$ are equicontinuous from above at \emptyset and also absolutely continuous relative to μ , then $\{v_n, n \geq 1\}$ are uniformly absolutely continuous relative to μ .
 11. If $f_n \in \mathcal{L}_p(\Omega, \mathcal{F}, \mu)$, $n \geq 1$, then $\|f_n - f_m\|_p = o(1)$ as $n, m \rightarrow \infty$ iff (i) $f_n - f_m \xrightarrow{p} 0$ as, $n, m \rightarrow \infty$ and (ii) $\int_A |f_n|^p d\mu, n \geq 1$, are equicontinuous from above at \emptyset .
 12. Random variables X_1, \dots, X_n have a (nonsingular) joint **normal distribution** if their d.f. is absolutely continuous with density defined by $f(x_1, \dots, x_n) = (2\pi)^{-n/2}$

$\sqrt{|A|} \exp\{-\frac{1}{2} \sum_{i,j=1}^n a_{ij}(x_i - \theta_i)(x_j - \theta_j)\}$, where $A = \{a_{ij}\}$ is a positive definite matrix of order n and $|A|$ signifies the determinant of A . Here, $\theta = (\theta_1, \dots, \theta_n)$ is a real vector. Verify that this yields a bona fide probability measure and that $E X_i = \theta_i$, $\rho(X_i, X_j) = \sigma_{ij}/\sigma_i \sigma_j$, where $\{\sigma_{ij}\}$ is the inverse matrix of A .

References

- J. L. Doob, *Stochastic Processes*, Wiley, New York, 1953.
- P. R. Halmos, *Measure Theory*, Van Nostrand, Princeton, 1950; Springer-Verlag, Berlin and New York, 1974.
- G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*, Cambridge Univ. Press, London, 1934.
- A. N. Kolmogorov, Foundations of Probability (Nathan Morrison, translator), Chelsea, New York, 1950.
- M. Loève, *Probability Theory*, 3rd ed., Van Nostrand, Princeton, 1963; 4th ed., Springer-Verlag, Berlin and New York, 1977–1978.
- E. J. McShane, *Integration*, Princeton Univ. Press, Princeton, 1944.
- M. E. Monroe, *Introduction to Measure and Integration*, Addison-Wesley, Cambridge, Mass., 1953.
- H. Robbins, "Mixture of Distributions," *Ann. Math. Statist.* **19** (1948), 360–369.
- S. Saks, *Theory of the Integral* (L. C. Young, translator), Stechert-Hafner, New York, 1937.
- J. L. Snell, "Applications of martingale system theorems," *Trans. Amer. Math. Soc.* **73** (1952), 293–312.
- D. V. Widder, *Advanced Calculus*, 2nd ed., Prentice-Hall, Englewood Cliffs, New Jersey, 1961.

7 Conditional Expectation, Conditional Independence, Introduction to Martingales

7.1 Conditional Expectations

From a theoretical vantage point, conditioning is a useful means of exploiting auxiliary information. From a practical vantage point, conditional probabilities reflect the change in unconditional probabilities due to additional knowledge.

The latter is represented by a sub- σ -algebra \mathcal{G} of the basic σ -algebra \mathcal{F} of events occurring in the underlying probability space (Ω, \mathcal{F}, P) . Associated with any measurable function X on Ω whose integral is defined, i.e., $|E X| \leq \infty$, is a function Y on Ω with $|E Y| \leq \infty$ satisfying

- i. Y is \mathcal{G} -measurable,
- ii. $\int_A Y dP = \int_A X dP$, all $A \in \mathcal{G}$.

Such a function Y is called the conditional expectation of X given \mathcal{G} and is denoted by $E\{X|\mathcal{G}\}$. In view of (i) and (ii), any \mathcal{G} -measurable function Z which differs from $Y = E\{X|\mathcal{G}\}$ on a set of probability zero also qualifies as $E\{X|\mathcal{G}\}$. In other words, the conditional expectation $E\{X|\mathcal{G}\}$ is only defined to within an equivalence, i.e., is any representative of a class of functions whose elements differ from one another only on sets of probability measure zero—an unpleasant feature.

To establish the existence of $E\{X|\mathcal{G}\}$ for any \mathcal{F} -measurable function X with $|E X| \leq \infty$, define the set functions $\lambda, \lambda^+, \lambda^-$ on \mathcal{F} by

$$\lambda = \lambda^+ - \lambda^-, \quad \lambda^\pm\{A\} = \int_A X^\pm dP, \quad A \in \mathcal{F}. \quad (1)$$

The measures λ^\pm are P -continuous on \mathcal{F} and so, if their restrictions $\lambda_\#^\pm \equiv \lambda^\pm|_{\#}$ are σ -finite, the Radon–Nikodym theorem (Theorem 6.5.1) ensures the existence of $Y = d\lambda_\#^\pm/dP_\#$ satisfying

$$\int_A Y dP = \lambda^\pm(A) = \int_A X dP, \quad A \in \mathcal{G}.$$

Thus, it suffices when $\lambda_\#^\pm$ are σ -finite to equate

$$E\{X|\mathcal{G}\} = \frac{d\lambda_\#}{dP_\#} \quad (2)$$

and to recall that the Radon–Nikodym derivative of (2) is unique to within sets of measure zero.

The second of the following lemmas shows that a similar procedure may be employed even when σ -finiteness is lacking.

Lemma 1. *If v is a P -continuous measure on \mathcal{F} , there exists a set $E \in \mathcal{F}$ such that v is σ -finite on $\mathcal{F} \cap E$ and for each $A \in \mathcal{F} \cap E^c$*

$$v(A) = 0 = P\{A\} \quad \text{or} \quad v(A) = \infty > P\{A\} > 0. \quad (3)$$

PROOF. Set

$$\mathcal{D} = \{D: D \in \mathcal{F}, v \text{ is } \sigma\text{-finite on } \mathcal{F} \cap D\}$$

and then choose $D_n \in \mathcal{D}$, $n \geq 1$, with $\sup_{n \geq 1} P\{D_n\} = \sup_{D \in \mathcal{D}} P\{D\} = \alpha$ (say). Clearly, $E = \bigcup_{n=1}^{\infty} D_n \in \mathcal{D}$, whence $P\{E\} = \alpha$. Moreover, for $D \in \mathcal{F} \cap E^c$ either $v\{D\} < \infty$, implying $D \cup E \in \mathcal{D}$ and hence $\alpha \geq P\{D \cup E\} = P\{D\} + \alpha$, that is, $P\{D\} = 0 = v\{D\}$, or alternatively $v\{D\} = \infty$, whence $P\{D\} > 0$ by the P -continuity of v . \square

Lemma 2. *If v is a P -continuous measure on \mathcal{F} , there exists an \mathcal{F} -measurable function $dv/dP \geq 0$, a.c., with*

$$v\{A\} = \int_A \frac{dv}{dP} dP, \quad A \in \mathcal{F}. \quad (4)$$

Moreover, if v is σ -finite, then dv/dP is a random variable.

PROOF. Choose E as in Lemma 1 and set $v' = v|_{\mathcal{F} \cap E}$ and $P' = P|_{\mathcal{F} \cap E}$. Then v' is σ -finite and P' -continuous, whence by the Radon–Nikodym theorem (Theorem 6.5.1) dv'/dP' exists on E . Define

$$\frac{dv}{dP} = \begin{cases} \frac{dv'}{dP'} & \text{on } E \\ \infty & \text{on } E^c. \end{cases}$$

Then dv/dP is \mathcal{F} -measurable and (4) holds for $A \in \mathcal{F} \cap E$. If, rather, $A \in \mathcal{F} \cap E^c$, then (3) ensures that (4) still obtains. Finally, if dv/dP were infinite on a set B of positive P -measure, then for every measurable subset $A \subset B$

$$v\{A\} = \int_A \frac{dv}{dP} dP = \infty \quad \text{or} \quad 0$$

and v would not be σ -finite. In other words, if v is σ -finite, dv/dP is a r.v.

□

Theorem 1. *If X is an \mathcal{F} -measurable function with $|E X| \leq \infty$ and \mathcal{G} is a sub- σ -algebra of \mathcal{F} , then (i) there exists a \mathcal{G} -measurable function $E\{X|\mathcal{G}\}$ unique to within sets of measure zero, called the **conditional expectation of X relative to \mathcal{G}** , for which*

$$\int_A E\{X|\mathcal{G}\} dP = \int_A X dP \quad \text{for all } A \in \mathcal{G}. \quad (5)$$

(ii) *If X is integrable and Z is an integrable, \mathcal{G} -measurable random variable such that for some π -class \mathcal{D} with $\sigma(\mathcal{D}) = \mathcal{G}$*

$$E Z = E X, \quad \int_A Z dP = \int_A X dP, \quad A \in \mathcal{D}, \quad (6)$$

then $Z = E\{X|\mathcal{G}\}$, a.c.

PROOF. (i) For $A \in \mathcal{G}$, define the measures $v^\pm\{A\} = \int_A X^\pm dP$. By Lemma 2, $dv^\pm/dP_{\mathcal{G}}$ exists, where $P_{\mathcal{G}} = P|_{\mathcal{G}}$. Since $|E X| \leq \infty$, at least one of the pair $E X^+$, $E X^-$ is finite, and so at least one of $dv^\pm/dP_{\mathcal{G}}$ is integrable. Thus, if

$$Y = \begin{cases} \frac{dv^+}{dP_{\mathcal{G}}} - \frac{dv^-}{dP_{\mathcal{G}}} & \text{on } \min\left(\frac{dv^+}{dP_{\mathcal{G}}}, \frac{dv^-}{dP_{\mathcal{G}}}\right) < \infty \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

the function Y is \mathcal{G} -measurable, $|E Y| \leq \infty$, and for $A \in \mathcal{G}$

$$\begin{aligned} \int_A Y dP &= \int_A \frac{dv^+}{dP_{\mathcal{G}}} dP_{\mathcal{G}} - \int_A \frac{dv^-}{dP_{\mathcal{G}}} dP_{\mathcal{G}} \\ &= \int_A X^+ dP - \int_A X^- dP = \int_A X dP, \end{aligned}$$

so that (5) holds with $E\{X|\mathcal{G}\} = Y$. As for uniqueness, let Y_1 and Y_2 be \mathcal{G} -measurable and satisfy (5). Then $P[Y_1 > r > Y_2] = 0$ for every finite rational number r . Hence, $P\{Y_1 > Y_2\} = 0$. Similarly $P\{Y_2 > Y_1\} = 0$ and so $Y_1 = Y_2$, a.c.

A propos of (ii), if

$$\mathcal{A} = \left\{ A : A \in \mathcal{F}, \int_A Z dP = \int_A X dP \right\},$$

then $\Omega \in \mathcal{A} \supset \mathcal{D}$ by (6). Since X and Z are integrable, \mathcal{A} is a λ -class, whence by Theorem 1.3.2 $\mathcal{A} \supset \sigma(\mathcal{D}) = \mathcal{G}$ and so $Z = E(X|\mathcal{G})$, a.c. \square

An immediate consequence of Theorem 1 is

Corollary 1. *Let $\mathcal{G}_1, \mathcal{G}_2$ be σ -algebras of events and let X, Y be \mathcal{L}_1 random variables. If $\sigma(X)$ and \mathcal{G}_1 are independent, then $E\{X|\mathcal{G}_1\} = E X$ a.c. and if*

$$\int_{A_1 A_2} X dP = \int_{A_1 A_2} Y dP, \quad A_i \in \mathcal{G}_i, i = 1, 2,$$

then $E\{X|\sigma(\mathcal{G}_1 \cup \mathcal{G}_2)\} = E\{Y|\sigma(\mathcal{G}_1 \cup \mathcal{G}_2)\}$, a.c.

A concrete construction of a conditional expectation appears in

Corollary 2. *Let the random vectors $X = (X_1, \dots, X_m)$ and $Y = (Y_1, \dots, Y_n)$ on (Ω, \mathcal{F}, P) be independent of one another and let f be a Borel function on $R^m \times R^n$ with $|E f(X, Y)| \leq \infty$. If, for $x \in R^m$,*

$$g(x) = \begin{cases} E f(x, Y) & \text{if } |E f(x, Y)| \leq \infty \\ 0, & \text{otherwise,} \end{cases} \quad (8)$$

then g is a Borel function on R^m with

$$g(X) = E\{f(X, Y)|\sigma(X)\}, \quad \text{a.c.} \quad (9)$$

PROOF. Let F_X, F_Y , and $F_{X,Y}$ be the joint distribution functions of the random vectors, X , Y and (X, Y) respectively, and denote the corresponding Lebesgue–Stieltjes measures by v_X, v_Y , and $v_{X,Y}$. Since f is a Borel function on R^{m+n} , Theorem 1.4.2 ensures that $f(x, y)$ is a Borel function on R^n for each fixed $x \in R^m$, and so by Theorem 6.3.3 and Fubini's theorem

$$g_{\pm}(x) \equiv E f^{\pm}(x, Y) = \int_{R^n} f^{\pm}(x, y) dv_Y(y)$$

are Borel functions on R^m . Thus

$$D = \{x \in R^m : g_+(x) = \infty = g_-(x)\}$$

is a Borel set and $g(x) = [g_+(x) - g_-(x)]I_D(x)$ is a Borel function whence by Theorem 1.4.4, $g(X)$ is $\sigma(X)$ -measurable.

By independence $(R^{m+n}, \mathcal{B}^{m+n}, v_{X,Y}) = (R^m, \mathcal{B}^m, v_X) \times (R^n, \mathcal{B}^n, v_Y)$. If $A \in \sigma(X)$, Theorem 1.4.4 guarantees the existence of $B \in \mathcal{B}^n$ such that $A = \{X \in B\}$, and once more via Theorem 6.3.3 and Fubini's theorem

$$\begin{aligned}
 \int_A g_{\pm}(X) dP &= E I_A g_{\pm}(X) = \int_B g_{\pm}(x) dv_X(x) \\
 &= \int_B \int_{R^n} f^{\pm}(x, y) dv_Y(y) dv_X(x) \\
 &= \int_{B \times R^n} f^{\pm} dv_{X,Y} \\
 &= \int_A f^{\pm}(X, Y) dP.
 \end{aligned}$$

Since $|E f(X, Y)| \leq \infty$, either $E g_+(X)$ or $E g_-(X)$ is finite whence $g_+(X) < \infty$, a.c. or $g_-(X) < \infty$, a.c. Hence, $I_B(X) = 0$, a.c. Thus, for $A \in \sigma(X)$,

$$\begin{aligned}
 \int_A g(X) dP &= \int_A [g_+(X) - g_-(X)] I_{B^c} dP = \int_A g_+(X) dP - \int_A g_-(X) dP \\
 &= \int_A f^+(X, Y) dP - \int_A f^-(X, Y) dP = \int_A f(X, Y) dP,
 \end{aligned}$$

and so (9) holds by Theorem 1(i). \square

As will be seen in Section 2, the conditional probability of an event A given \mathcal{G} , denoted $P\{A|\mathcal{G}\}$, may be defined as $E\{I_A|\mathcal{G}\}$. A typical exploitation of conditional expectation appears in

EXAMPLE 1 (Kesten). If $\{S_n = \sum_i^n X_i, n \geq 1\}$, where $\{X_i, X_n \geq 1\}$ are non-negative i.i.d. r.v.s with $E X = \infty$, then $\overline{\lim}_{n \rightarrow \infty} S_n / S_{n-1} = \infty$, a.c.

PROOF. Set $A_i = \{S_{i-1} \leq \varepsilon X_i\}$, $\varepsilon > 0$, where $S_0 = 0$, and note that $A_i A_j \subset A_i \{\sum_{k=i+1}^{j-1} X_k \leq \varepsilon X_j\}$ for $i < j$, implying for $k \geq 1$ via independence and identical distributions

$$P\left\{A_i \bigcup_{j=i+k}^{\infty} A_j\right\} \leq P\{A_i\} P\left\{\bigcup_{j=k}^{\infty} A_j\right\},$$

so that (12) of Lemma 4.2.4 obtains with $B_j = A_j$. Moreover, if F is the d.f. of X and $a(0) = 0$,

$$a(x) = \frac{x}{\int_0^x [1 - F(y)] dy}, \quad x > 0,$$

then, clearly, $a(x)/x \downarrow$ and (Exercise 6.2.16) $a(x) \uparrow$ and $E a(X) = \infty$. It follows that $E a(\varepsilon X) = \infty$, all $\varepsilon > 0$. However, via Corollary 2 and Examples 5.4.1, 6.2.2

$$\begin{aligned}
 \sum_{n=1}^{\infty} P\{A_n\} &= E \sum_{n=1}^{\infty} P\{S_{n-1} < \varepsilon X_n | X_n\} = \int_0^{\infty} \sum_{n=1}^{\infty} P\{S_{n-1} \leq ex\} dF(x) \\
 &\geq \int_0^{\infty} a(ex) dF(x) = E a(\varepsilon X),
 \end{aligned}$$

and so Lemma 4.2.4 ensures $P\{A_n, i.o.\} = 1$, all $\varepsilon > 0$, which is tantamount to the conclusion of Example 1.

It follows immediately from this example that for any i.i.d. r.v.s $\{X, X_n, n \geq 1\}$ with $E|X| = \infty$,

$$\overline{\lim}_{n \rightarrow \infty} \frac{|X_n|}{|\sum_{i=1}^{n-1} X_i|} = \infty, \quad \text{a.c.} \quad \square$$

Let X be an \mathcal{F} -measurable function with $|E X| \leq \infty$ and $\{Y_\lambda, \lambda \in \Lambda\}$, $\{\mathcal{G}_\lambda, \lambda \in \Lambda\}$, nonempty families of random variables and σ -algebras of events respectively. It is customary to define

$$\begin{aligned} E\{X | Y_\lambda, \lambda \in \Lambda\} &= E\{X | \sigma(Y_\lambda, \lambda \in \Lambda)\}, \\ E\{X | \mathcal{G}_\lambda, \lambda \in \Lambda\} &= E\{X | \sigma(\mathcal{G}_\lambda, \lambda \in \Lambda)\}, \end{aligned} \quad (10)$$

and, in particular,

$$\begin{aligned} E\{X | Y_1, \dots, Y_n\} &= E\{X | \sigma(Y_1, \dots, Y_n)\}, \\ E\{X | Y\} &= E\{X | \sigma(Y)\}. \end{aligned} \quad (11)$$

Since, by definition $E\{X | Y_1, \dots, Y_n\}$ is $\sigma(Y_1, \dots, Y_n)$ -measurable, Theorem 1.4.4 guarantees that for some Borel function g on R^n

$$E\{X | Y_1, \dots, Y_n\} = g(Y_1, \dots, Y_n).$$

Conversely, if g is a Borel function on R^n such that $|E g(Y_1, \dots, Y_n)| \leq \infty$ and for every $A \in \sigma(Y_1, \dots, Y_n)$

$$\int_A g(Y_1, \dots, Y_n) dP = \int_A X dP,$$

then $g(Y_1, \dots, Y_n) = E\{X | Y_1, \dots, Y_n\}$, a.c.

In particular, if $Y = I_A$ for some $A \in \mathcal{F}$, then $\sigma(Y) = \{\emptyset, A, A^c, \Omega\}$ and every version of $E\{X | Y\}$ must be constant on each of the sets A, A^c , necessitating

$$E\{X | Y\}(\omega) = \begin{cases} \frac{1}{P\{A\}} \int_A X dP & \text{if } \omega \in A \\ \frac{1}{P\{A^c\}} \int_{A^c} X dP & \text{if } \omega \in A^c, \end{cases} \quad (12)$$

where either of the constants on the right can be construed as any number in $[-\infty, \infty]$ when the corresponding set A or A^c has probability zero.

More generally, if $\{A_n, n \geq 1\}$ is a σ -partition of Ω in \mathcal{F} with $P\{A_n\} > 0$, $n \geq 1$, and $\mathcal{G} = \sigma(A_n, n \geq 1)$, then for any measurable function X with $|E X| \leq \infty$

$$E\{X | \mathcal{G}\} = \sum_1^\infty \left(\frac{1}{P\{A_n\}} \int_{A_n} X dP \right) I_{A_n}, \quad \text{a.c.} \quad (13)$$

Moreover, this remains valid even if $P\{A_m\} = 0$ for certain indices m , the quantity in the parenthesis being then interpreted as an arbitrary constant.

Some simple consequences of the definition of conditional expectation are

$$E\{1|\mathcal{G}\} = 1, \text{ a.c., } E\{X|\mathcal{G}\} \geq 0, \text{ a.c. if } X \geq 0, \text{ a.c.} \quad (14, i)$$

$$\begin{aligned} E\{cX|\mathcal{G}\} &= c E\{X|\mathcal{G}\}, \text{ a.c., if } |E X| \leq \infty \text{ and} \\ &\quad c \text{ is a finite constant,} \end{aligned} \quad (14, ii)$$

$$\begin{aligned} E\{X + Y|\mathcal{G}\} &= E\{X|\mathcal{G}\} + E\{Y|\mathcal{G}\}, \text{ a.c., if } E(X^- + Y^-) < \infty \text{ or} \\ &\quad E(X^+ + Y^+) < \infty. \end{aligned} \quad (14, iii)$$

These properties assert roughly that if $TX = E\{X|\mathcal{G}\}$, then T is linear, order preserving (monotone), and $T1 = 1$.

$$E\{X|\mathcal{G}\} = X, \text{ a.c. if } X \text{ is } \mathcal{G}\text{-measurable.} \quad (14, iv)$$

If $\mathcal{G}_1, \mathcal{G}_2$ are σ -algebras with $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$ and $|E X| \leq \infty$, then

$$E\{E\{X|\mathcal{G}_2\}|\mathcal{G}_1\} = E\{X|\mathcal{G}_1\} = E\{E\{X|\mathcal{G}_1\}|\mathcal{G}_2\}, \text{ a.c.} \quad (14, v)$$

PROOF. Since $E\{E\{X|\mathcal{G}_2\}|\mathcal{G}_1\}$ is \mathcal{G}_1 -measurable and for $A \in \mathcal{G}_1$

$$\int_A E\{E\{X|\mathcal{G}_2\}|\mathcal{G}_1\} dP = \int_A E\{X|\mathcal{G}_2\} dP = \int_A X dP,$$

the first equality of (14, v) follows. Since $E\{X|\mathcal{G}_1\}$ is \mathcal{G}_i -measurable for $i = 1$ and hence $i = 2$, the second equality of (14, v) follows from (14, iv). \square

Theorem 2. Let $\{X_n, n \geq 1\}$ and Y be random variables with $E|Y| < \infty$ and \mathcal{G} a σ -algebra of events.

- i. (Monotone convergence theorem for conditional expectations). If $Y \leq X_n \uparrow X$, a.c., then $E\{X_n|\mathcal{G}\} \uparrow E\{X|\mathcal{G}\}$, a.c.
- ii. (Fatou lemma for conditional expectations). If $Y \leq X_n, n \geq 1$, a.c., then $E\{\liminf X_n|\mathcal{G}\} \leq \liminf E\{X_n|\mathcal{G}\}$, a.c.
- iii. (Lebesgue dominated convergence theorem for conditional expectations). If $X_n \xrightarrow{a.e.} X$ and $|X_n| \leq |Y|, n \geq 1$, a.c., then $E\{X_n|\mathcal{G}\} \xrightarrow{a.e.} E\{X|\mathcal{G}\}$.

PROOF. (i) By the monotone property of conditional expectations, $E\{Y|\mathcal{G}\} \leq E\{X_n|\mathcal{G}\} \uparrow$ some function Z , a.c. For $A \in \mathcal{G}$, by ordinary monotone convergence

$$\int_A Z dP = \lim_n \int_A E\{X_n|\mathcal{G}\} dP = \lim_n \int_A X_n dP = \int_A X dP,$$

and, since Z is \mathcal{G} -measurable, $Z = E\{X|\mathcal{G}\}$, a.c.

Apropos of (ii), set $Y_n = \inf_{m \geq n} X_m$. Then $Y \leq Y_n \uparrow \lim_{m \rightarrow \infty} X_m$, whence $\underline{\lim}_n E\{X_n|\mathcal{G}\} \geq \underline{\lim}_n E\{Y_n|\mathcal{G}\} = E\{\lim_n X_n|\mathcal{G}\}$ by (i). Finally, (iii) is a consequence of (ii) via

$$\pm E\{X|\mathcal{G}\} = E\{\pm X|\mathcal{G}\} \leq \underline{\lim} E\{\pm X_n|\mathcal{G}\} = \underline{\lim}(\pm E\{X_n|\mathcal{G}\}), \quad \text{a.c. } \square$$

An extremely useful fact about conditional expectations is furnished by

Theorem 3. *Let X be a random variable with $|E X| \leq \infty$ and \mathcal{G} a σ -algebra of events. If Y is a finite-valued \mathcal{G} -measurable random variable such that $|E XY| \leq \infty$, then*

$$E\{XY|\mathcal{G}\} = Y E\{X|\mathcal{G}\}, \quad \text{a.c.} \quad (15)$$

PROOF. By separate consideration of X^\pm and Y^\pm it may be supposed that $X \geq 0$ and $Y \geq 0$. Moreover, by the monotone convergence theorem for conditional expectations it may even be assumed that X and Y are bounded r.v.s. Set

$$v\{A\} = \int_A XY dP, \quad \mu\{A\} = \int_A X dP, \quad A \in \mathcal{F}.$$

Then both μ and v are finite, P -continuous measures on \mathcal{F} and, denoting as usual the restrictions of P , v , μ to \mathcal{G} by $P_\mathcal{G}$, $v_\mathcal{G}$, $\mu_\mathcal{G}$,

$$\frac{dv_\mathcal{G}}{dP_\mathcal{G}} = E\{XY|\mathcal{G}\}, \quad \frac{d\mu_\mathcal{G}}{dP_\mathcal{G}} = E\{X|\mathcal{G}\}, \quad \frac{d\mu}{dP} = X, \quad \text{a.c.}$$

For $A \in \mathcal{G}$, by Theorem 6.5.2

$$\int_A Y d\mu_\mathcal{G} = \int_A Y \frac{d\mu_\mathcal{G}}{dP_\mathcal{G}} dP_\mathcal{G} = \int_A Y E\{X|\mathcal{G}\} dP_\mathcal{G},$$

and for $A \in \mathcal{F}$

$$\int_A Y d\mu = \int_A Y \frac{d\mu}{dP} dP = \int_A YX dP.$$

Consequently,

$$\int_A Y E\{X|\mathcal{G}\} dP = \int_A XY dP, \quad A \in \mathcal{G},$$

and since $Y E\{X|\mathcal{G}\}$ is \mathcal{G} -measurable, (15) follows. \square

Theorem 4 (Jensen's Inequality for Conditional Expectations). *Let Y be \mathcal{F} -measurable with $|E Y| \leq \infty$ and g any finite convex function on $(-\infty, \infty)$ with $|E g(Y)| \leq \infty$. If for some σ -algebra \mathcal{G} of events, (i) $X = E\{Y|\mathcal{G}\}$, a.c. or (ii) X is \mathcal{F} -measurable with $X \leq E\{Y|\mathcal{G}\}$, a.c. and $g \uparrow$, then*

$$g(X) \leq E\{g(Y)|\mathcal{G}\}, \quad \text{a.c.} \quad (16)$$

PROOF. Since (ii) follows directly from (i), it suffices to prove the latter. To this end, define

$$g^*(t) = \lim_{s \rightarrow t^-} \frac{g(s) - g(t)}{s - t},$$

whence g^* is a finite, nondecreasing function on $(-\infty, \infty)$. Now the secant line of a convex function is always above the one-sided tangent line, that is,

$$g(t) \geq g(s) + (t - s)g^*(s), \quad -\infty < s, t < \infty, \quad (17)$$

whence if $A = \{|X| \leq M, g^*(X) \geq 0\}$, $0 < M < \infty$ and $B = \{|X| < \infty, g^*(X) \geq 0\}$ both $g(X)$ and $g^*(X)$ are bounded on $A \in \mathcal{G}$, so by (17)

$$I_A g(Y) \geq I_A g(X) + I_A(Y - X)g^*(X), \quad \text{a.c.}$$

Since $|\mathbb{E} I_A(Y - X)g^*(X)| \leq \infty$, by Theorem 3,

$$I_A \mathbb{E}\{g(Y)|\mathcal{G}\} = \mathbb{E}\{I_A g(Y)|\mathcal{G}\} \geq I_A g(X), \quad \text{a.c.}$$

As $M \rightarrow \infty$, $I_A \xrightarrow{\text{a.s.}} I_B$, so (16) holds on B . Similarly, for $\{g^*(X) \leq 0, |X| < \infty\}$.

Consider next $D = \{X = \infty\}$. If $g^*(s) > 0$, some s in $(-\infty, \infty)$, then (17) ensures

$$I_D \mathbb{E}\{g(Y)|\mathcal{G}\} \geq g(s)I_D + I_D \mathbb{E}\{Y - s|\mathcal{G}\}g^*(s) = \infty = g(X) \cdot I_D$$

If $g^*(s) \leq 0$, all s in $(-\infty, \infty)$, then $g \downarrow$ whence $g(\infty) \leq g(X)$, a.c. and (16) holds on D since

$$I_D \mathbb{E}\{g(Y)|\mathcal{G}\} \geq I_D g(\infty) = I_D g(X).$$

Let $D' = \{X = -\infty\}$. If $g^*(s) > 0$, all s in $(-\infty, \infty)$, then $g \uparrow$ whence

$$I_{D'} \mathbb{E}\{g(Y)|\mathcal{G}\} \geq I_{D'} g(-\infty) = g(X)I_{D'}$$

whereas if $g^*(s) < 0$ for some s in $(-\infty, \infty)$, then via (17) $g(Y) \geq g(s) + (Y - s)g^*(s)$ implying

$$\begin{aligned} \mathbb{E}\{g(Y)|\mathcal{G}\} &\geq g(s) + [\mathbb{E}\{Y|\mathcal{G}\} - s]g^*(s) \\ &= g(s) + [X - s]g^*(s) = \infty, \quad \text{a.c. on } D' \end{aligned}$$

and (16) holds on D' . □

Corollary 3. For X and \mathcal{G} as in the theorem, with probability one

$$|\mathbb{E}\{X|\mathcal{G}\}| \leq \mathbb{E}\{|X||\mathcal{G}\}, \quad \mathbb{E}^r\{|X||\mathcal{G}\} \leq \mathbb{E}\{|X|^r|\mathcal{G}\}, \quad r \geq 1,$$

$$\mathbb{E}\{\max(a, X)|\mathcal{G}\} \geq \max\{a, \mathbb{E}\{X|\mathcal{G}\}\}, \quad -\infty < a < \infty.$$

Theorem 5 (Extended Fatou Lemma For Conditional Expectations). Let $\{X_n, n \geq 1\}$ be random variables with $|\mathbb{E} X_n| \leq \infty$, $n \geq 1$, and \mathcal{G} a σ -algebra of events. If $|\mathbb{E} \underline{\lim}_n X_n| \leq \infty$ and

$$\sup_{n \geq 1} \mathbb{E}\{X_n^- I_{\{X_n > k\}}|\mathcal{G}\} \xrightarrow{\text{a.s.}} 0 \quad \text{as } k \rightarrow \infty,$$

then $\mathbb{E}\{\underline{\lim}_n X_n|\mathcal{G}\} \leq \underline{\lim}_n \mathbb{E}\{X_n|\mathcal{G}\}$, a.c.

PROOF. If $Y_k = \sup_{n \geq 1} E\{X_n^- I_{\{X_n > k\}} | \mathcal{G}\}$, $k > 0$, then $Y_k \xrightarrow{\text{a.e.}} 0$ by hypothesis. Since with probability one

$$E\{X_n | \mathcal{G}\} = E\{X_n(I_{\{X_n \leq k\}} + I_{\{X_n > k\}}) | \mathcal{G}\} \geq E\{X_n I_{\{X_n \leq k\}} | \mathcal{G}\} = Y_k,$$

it follows via Theorem 2(ii) that for all $k > 0$

$$\begin{aligned} \varliminf_n E\{X_n | \mathcal{G}\} &\geq \varliminf_n E\{X_n I_{\{X_n \leq k\}} | \mathcal{G}\} = Y_k \\ &\geq E\{\varliminf_n X_n I_{\{X_n \leq k\}} | \mathcal{G}\} = Y_k \\ &\geq E\{\varliminf_n X_n | \mathcal{G}\} = Y_k, \quad \text{a.e.}, \end{aligned}$$

which yields the theorem upon letting $k \rightarrow \infty$. \square

Corollary 4. Let \mathcal{G} be a σ -algebra of events, $\{X_n, n \geq 1\}$ random variables with $|E X_n| \leq \infty$, $n \geq 1$, and $\lim_{k \rightarrow \infty} E\{|X_n| I_{\{|X_n| > k\}} | \mathcal{G}\} = 0$, uniformly in n with probability one. If $X_n \xrightarrow{\text{a.e.}} X$, where $|E X| \leq \infty$, then $E\{X_n | \mathcal{G}\} \xrightarrow{\text{a.e.}} E\{X | \mathcal{G}\}$.

PROOF. Applying the theorem to X_n and $-X_n$,

$$E\{X | \mathcal{G}\} \leq \varliminf_n E\{X_n | \mathcal{G}\} \leq \varlimsup_n E\{X_n | \mathcal{G}\} \leq E\{-X | \mathcal{G}\}, \quad \text{a.e.} \quad \square$$

Corollary 5. Let \mathcal{G} be a σ -algebra of events and $\{X_n, n \geq 1\}$ random variables with $|E X_n| \leq \infty$, $n \geq 1$. If $X_n \xrightarrow{\text{a.e.}} X_0$, where $|E X_0| \leq \infty$ and for some $r > 1$

$$\sup_{n \geq 1} E\{|X_n|^r | \mathcal{G}\} < \infty, \quad \text{a.e.},$$

then $E\{X_n | \mathcal{G}\} \xrightarrow{\text{a.e.}} E\{X_0 | \mathcal{G}\}$.

PROOF. Let $X'_n = X_n I_A$, $n \geq 0$, where $A = \{\sup_{n \geq 1} E\{|X_n|^r | \mathcal{G}\} < C < \infty\}$. Since $E|X'_n|^r = E[E\{|X'_n|^r | \mathcal{G}\}] \leq C$, Fatou's lemma ensures that $E|X'_0|^r \leq C$. Moreover, for $K > 0$,

$$E\{|X'_n| I_{\{|X'_n| > K\}} | \mathcal{G}\} \leq K^{1-r} E\{|X'_n|^r | \mathcal{G}\} \leq CK^{1-r}$$

uniformly in n with probability one, whence $E\{X'_n | \mathcal{G}\} \xrightarrow{\text{a.e.}} E\{X'_0 | \mathcal{G}\}$ by Corollary 4 and the conclusion follows as $C \rightarrow \infty$. \square

EXAMPLE 2. Let $S_n = \sum_{j=1}^n X_j$, where $\{X_n, n \geq 1\}$ are independent r.v.s and let $\{\alpha_n, n \geq 1\}$ be constants such that $P\{S_n < \alpha_n\} > 0$, $n \geq 1$. Then

$$P\left\{\bigcup_{n=1}^{\infty} [S_n \geq \alpha_n]\right\} = 1 \quad (18)$$

entails

$$P\{S_n \geq \alpha_n, \text{i.o.}\} = 1. \quad (19)$$

PROOF. Set $A_N = \bigcup_{n=N}^{\infty} [S_n \geq \alpha_n]$, $N \geq 1$, and suppose that $P\{A_N\} = 1$ for

some $N \geq 1$. Then $1 = E P\{A_N | S_N\}$, implying $P\{A_N | S_N\} = 1$, a.c., and so on the set $\{S_N < \alpha_N\}$

$$P\{A_{N+1} | S_N\} = P\{A_N | S_N\} = 1, \quad \text{a.c.} \quad (20)$$

Next, if $h(x) = P\{\bigcup_{n=N+1}^{\infty} [S_n - S_N \geq \alpha_n - x]\}$, then $h \uparrow$ and by Corollary 7.1.2

$$h(S_N) = P\{A_{N+1} | S_N\}, \quad \text{a.c.} \quad (21)$$

According to (20), $h(S_N) = 1$, a.c., on $\{S_N < \alpha_N\}$. Thus

$$\int_{(-\infty, \alpha_N)} [1 - h(x)] dP\{S_N < x\} = P\{S_N < \alpha_N\} - \int_{\{S_N < \alpha_N\}} h(S_N) dP = 0,$$

and so $h(x) = 1$ for every $x < \alpha_N$ that is a point of increase of $P\{S_N < x\}$. Since $P\{S_N < \alpha_N\} > 0$, there must be at least one such x . Then monotonicity of h guarantees $h(x) = 1$ for all $x \geq \alpha_N$ and (21) ensures

$$P\{A_{N+1} | S_N\} = 1, \quad \text{a.c.}$$

on $[S_N \geq \alpha_N]$. Consequently, recalling (20), $P\{A_{N+1} | S_N\} = 1$, a.c., and so $P\{A_{N+1}\} = 1$. Since $P\{A_1\} = 1$ by hypothesis, it follows inductively that $P\{A_N\} = 1$ for all $N \geq 1$, which, in turn, yields (19). \square

Corollary 6. If $S_n = \sum_1^n X_i$, where $\{X_n, n \geq 1\}$ are independent r.v.s with $P\{X_n < 0\} > 0$, $n \geq 1$, then for nonnegative, constants $\{\alpha_n, n \geq 1\}$, (18) implies (19).

PROOF. $P\{S_n < \alpha_n\} \geq P\{S_n < 0\} \geq P\{X_1 < 0\} \cdots P\{X_n < 0\} > 0$.

Remark. Clearly, both equality signs can be deleted from (18) and (19) provided $P\{S_n < \alpha_n\} > 0$, $n \geq 1$, is modified to $P\{S_n \leq \alpha_n\} > 0$, $n \geq 1$.

EXERCISES 7.1

- If X is a r.v. with $|E X| \leq \infty$ and \mathcal{G} a σ -algebra of events with $\sigma(X)$ and \mathcal{G} independent classes, then $E\{X | \mathcal{G}\} = E X$, a.c. In particular, if $\{X_n\}$ are independent r.v.s with $|E X_n| \leq \infty$, $n \geq 1$, then $E\{X_n | X_1, \dots, X_{n-1}\} = E\{X_n | X_{n+1}, X_{n+2}, \dots\} = E X_n$, a.c.
- Let $\{A_n, n \geq 1\}$ be σ -partition of Ω in \mathcal{F} . Verify that if $\mathcal{G} = \sigma(A_n, n \geq 1)$ and $|E X| \leq \infty$, then

$$E\{X | \mathcal{G}\} = \sum_1^{\infty} I_{A_n} \left(\frac{\int_{A_n} X dP}{P\{A_n\}} \right).$$

where the parenthetical quantity is an arbitrary real number if $P\{A_n\} = 0$.

- Show that Corollary 5 remains true if $E|X| < \infty$ and $C = \infty$.

4. Let \mathcal{D} be a semi-algebra of events and X, Y r.v.s with $E|X| < \infty, E|Y| < \infty$. If $\int_D X dP \leq \int_D Y dP, D \in \mathcal{D}$, then $E\{X|\sigma(\mathcal{D})\} \leq E\{Y|\sigma(\mathcal{D})\}$, a.c.
5. Let (X_1, X_2) be jointly normally distributed (Exercise 6.5.12) with p.d.f.

$$[2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}]^{-1} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{x_1^2}{\sigma_1^2} - \frac{2\rho x_1 x_2}{\sigma_1 \sigma_2} + \frac{x_2^2}{\sigma_2^2}\right]\right\}.$$

Find $E\{X_1 | X_2\}$ and the conditional variance.

6. Prove that
- X is an \mathcal{L}_2 r.v. and Y is a r.v. such that $E\{X|Y\} = Y$, a.c. and $E\{Y|X\} = X$, a.c., then $X = Y$, a.c.
 - Let $\mathcal{G}_1, \mathcal{G}_2$ be σ -algebras of events and X an \mathcal{L}_1 r.v. If $X_1 = E\{X|\mathcal{G}_1\}, X_2 = E\{X_1|\mathcal{G}_2\}$, and $X = X_2$, a.c., then $X_1 = X_2$, a.c.
 - If X, Y are \mathcal{L}_1 r.v.s with $E\{X|Y\} = Y$, a.c., and $E\{Y|X\} = X$, a.c., then $X = Y$, a.c.
7. Prove that the set E of lemma 1 and the function dv/dP are unique to within an equivalence.
8. Let X be an \mathcal{L}_2 r.v. and \mathcal{G} a σ -algebra of events. Prove that (i) $\sigma^2(E\{X|\mathcal{G}\}) \leq \sigma^2(X)$, (ii) if for any α in $(-\infty, \infty)$, $Y = \min(X, \alpha)$,

$$E\{[X - E\{X|\mathcal{G}\}]^2|\mathcal{G}\} \geq E\{[Y - E\{Y|\mathcal{G}\}]^2|\mathcal{G}\}, \quad \text{a.c.}$$

(iii) $E(X - E\{X|\mathcal{G}\})^2 \leq E(X - Y)^2$ for any \mathcal{G} -measurable r.v. Y .

9. Let \mathcal{G} be an σ -algebra of events and $(X_n, n \geq 1)$ r.v.s. If for some $p \geq 1$, $X_n \xrightarrow{\mathcal{F}_p} X$, then $E\{X_n|\mathcal{G}\} \xrightarrow{\mathcal{F}_p} E\{X|\mathcal{G}\}$.
10. Let \mathcal{G} be a σ -algebra of events and X a nonnegative r.v. Prove that $E\{X|\mathcal{G}\} = \text{esup}\{h: h \text{ is } \mathcal{G}\text{-measurable, } h \geq 0, \text{ a.c., and } \int_A h dP \leq \int_A X dP, \text{ all } A \in \mathcal{G}\}$.
11. Show via an example that in Theorem 2(iii) $X_n \xrightarrow{P} X$ cannot replace $X_n \xrightarrow{\text{a.c.}} X$.

12. If $\{X_n, n \geq 1\}$ are \mathcal{L}_1 interchangeable r.v.s (Exercise 6.3.10) and $\mathcal{F}_n = \sigma(\sum_{i=1}^n X_i, j \geq n)$, prove that $E\{X_i|\mathcal{F}_n\} = (1/n) \sum_j^n X_j$, a.c., $1 \leq i \leq n$. More generally, if $\{X_n, n \geq 1\}$ are interchangeable r.v.s, ϕ is a symmetric Borel function on R^m with $E|\phi(X_1, \dots, X_m)| < \infty$, and $\mathcal{F}_n = \sigma(U_{m,j}, j \geq n)$, where

$$\binom{n}{m} U_{m,n} = \sum_{1 \leq i_1 < \dots < i_m \leq n} \phi(X_{i_1}, \dots, X_{i_m}), \quad n \geq m,$$

then for $1 \leq i_1 < \dots < i_m \leq n+1$

$$E\{\phi(X_{i_1}, \dots, X_{i_m})|\mathcal{F}_{n+1}\} = E\{\phi(X_1, \dots, X_m)|\mathcal{F}_{n+1}\}, \quad \text{a.c.}$$

13. (Chow–Robbins) Let $S_n = \sum_1^n X_i$, $n \geq 1$, where $\{X, X_n, n \geq 1\}$ are i.i.d. with $E|X| = \infty$. Then for any positive constants $\{b_n\}$, either $P\{\lim_{n \rightarrow \infty} b_n^{-1}|S_n| = 0\} = 1$ or (*) $P\{\overline{\lim}_{n \rightarrow \infty} b_n^{-1}|S_n| = \infty\} = 1$. Hint: If (*) fails, by the zero-one law $\overline{\lim} b_n^{-1}|S_n| < \infty$, a.c., whence $P\{\overline{\lim} b_n^{-1}|S_{n-1}| < \infty\} = P\{\overline{\lim} b_n^{-1}|\sum_{i=2}^n X_i| < \infty\} = 1$, entailing $P\{\overline{\lim} b_n^{-1}|X_{n+1}| < \infty\} \geq P\{\overline{\lim} b_n^{-1}(|S_n| + |S_{n-1}|) < \infty\} = 1$. Now apply $\overline{\lim} b_n^{-1}|S_n| \leq \overline{\lim}|S_n|/(1 + |X_{n+1}|) \cdot \overline{\lim}(1 + |X_{n+1}|)/b_n$ and Example 1.

7.2 Conditional Probabilities, Conditional Probability Measures

Let (Ω, \mathcal{F}, P) be a fixed probability space and \mathcal{G} a sub- σ -algebra of \mathcal{F} . For $A, B \in \mathcal{F}$, define

$$P\{A | \mathcal{G}\} = E\{I_A | \mathcal{G}\}, \quad P\{A | B\} = E\{I_A | I_B\}. \quad (1)$$

The former, $P\{A | \mathcal{G}\}$, is called the **conditional probability of the event A given \mathcal{G}** and according to the prior section is a \mathcal{G} -measurable function on Ω satisfying

$$\int_G P\{A | \mathcal{G}\} dP = \int_G I_A dP = P\{A \cdot G\}, \quad G \in \mathcal{G}. \quad (2)$$

The latter, $P\{A | B\}$, is called the **conditional probability of the event A given the event B** and according to 7.1(12), if $P\{B\} > 0$,

$$P\{A | B\} = \frac{1}{P\{B\}} \int_B I_A dP = \frac{P\{AB\}}{P\{B\}}, \quad \omega \in B. \quad (3)$$

The properties 7.1(14) of conditional expectation and monotone convergence yield:

$$0 \leq P\{A | \mathcal{G}\} \leq 1, \text{ a.c.}; \quad (4, i)$$

$$P\{A | \mathcal{G}\} = 0, \text{ a.c., iff } P\{A\} = 0, \quad P\{A | \mathcal{G}\} = 1, \text{ a.c., iff } P\{A\} = 1; \quad (4, ii)$$

If $\{A_n, n \geq 1\}$ are disjoint sets of \mathcal{F} , then

$$P\left\{\bigcup_1^\infty A_n \middle| \mathcal{G}\right\} = \sum_{n=1}^\infty P\{A_n | \mathcal{G}\}, \quad \text{a.c.}; \quad (4, iii)$$

If $A_n \in \mathcal{F}$, $n \geq 1$, and $\lim_{n \rightarrow \infty} A_n = A$, then

$$\lim_{n \rightarrow \infty} P\{A_n | \mathcal{G}\} = P\{A | \mathcal{G}\}, \quad \text{a.c.} \quad (4, iv)$$

Property (iii) asserts that for each sequence $\{A_n\}$ of disjoint events

$$P\left\{\bigcup_1^\infty A_n \middle| \mathcal{G}\right\}(\omega) = \sum_{n=1}^\infty P\{A_n | \mathcal{G}\}(\omega)$$

except for ω in a null set which may well depend on the particular sequence $\{A_n\}$. It does not stipulate that there exists a fixed null set N such that

$$P\left\{\bigcup_1^\infty A_n \middle| \mathcal{G}\right\}(\omega) = \sum_{n=1}^\infty P\{A_n | \mathcal{G}\}(\omega), \quad \omega \in N^c$$

for every sequence $\{A_n\}$ of disjoint events. In fact, the later is false (Halmos, 1950, p. 210).

Definition. Let $\mathcal{F}_1, \mathcal{G}$ be σ -algebras of events. A **conditional probability measure** or **regular conditional probability** on \mathcal{F}_1 given \mathcal{G} is a function $P(A, \omega)$ defined for $A \in \mathcal{F}_1$ and $\omega \in \Omega$ such that

- for each $\omega \in \Omega$, $P(A, \omega)$ is a probability measure on \mathcal{F}_1 ,
- for each $A \in \mathcal{F}_1$, $P(A, \omega)$ is a \mathcal{G} -measurable function on Ω coinciding with the conditional probability of A given \mathcal{G} , i.e., $P(A, \omega) = P\{A | \mathcal{G}\}(\omega)$, a.c.

One advantage of regular conditional probabilities is that conditional expectations may then be envisaged as ordinary expectations relative to the conditional probability measure, as stated in

Theorem 1. If for some pair $\mathcal{F}_1, \mathcal{G}$ of σ -algebras of events, $P_\omega\{A\} \equiv P(A, \omega)$ is a regular conditional probability on \mathcal{F}_1 given \mathcal{G} and X is an \mathcal{F}_1 -measurable function with $|E X| \leq \infty$, then

$$E\{X | \mathcal{G}\}(\omega) = \int_{\Omega} X dP_\omega, \quad \text{a.c.} \quad (5)$$

PROOF. By separate consideration of X^+ and X^- , it may be supposed that $X \geq 0$. Let

$$\mathcal{H} = \{X : X \geq 0, X \text{ is } \mathcal{F}_1\text{-measurable, and (5) holds}\}.$$

By definition of conditional probability measure, $I_A \in \mathcal{H}$ for $A \in \mathcal{F}_1$. Since, as is readily verified, \mathcal{H} is a monotone system, Theorem 1.4.3 ensures that \mathcal{H} contains all nonnegative \mathcal{F}_1 -measurable functions. \square

In general, a regular conditional probability measure does not exist (Doob, 1953, p. 624). A simple case where it does is

EXAMPLE 1. Let $\mathcal{F}_1 = \mathcal{F}$, $\mathcal{G} = \sigma(A_n, n \geq 1)$, where $\{A_n, n \geq 1\}$ is a σ -partition of Ω in \mathcal{F} . Then, if

$$P(A, \omega) = \begin{cases} P\{AA_n\}/P\{A_n\}, & \omega \in A_n, P\{A_n\} > 0 \\ P\{A\}, & \omega \in A_n, P\{A_n\} = 0, \end{cases}$$

$P(A, \omega) = P\{A | \mathcal{G}\}$ a.c. (as in Exercise 7.1.2), and so $P(A, \omega)$ is a regular conditional probability relative to \mathcal{F} given \mathcal{G} .

A more interesting and important illustration is that of

EXAMPLE 2. Let X_1, X_2 be coordinate r.v.s on a probability space (R^2, \mathcal{B}^2, P) with an absolutely continuous d.f. $F(x_1, x_2)$, i.e., for some Borel function f

$$F(x_1, x_2) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(s, t) ds dt, \quad (x_1, x_2) \in R^2. \quad (6)$$

Let $\mathcal{F}_1 = \mathcal{B}^2 = \sigma(X_1, X_2)$, $\mathcal{G} = \sigma(X_2) = \mathbb{R} \times \mathcal{B} = \{\mathbb{R} \times B : B \in \mathcal{B}\}$, and for $(x_1, x_2) \in \mathbb{R}^2$ define

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, t) dt, \quad f_2(x_2) = \int_{-\infty}^{\infty} f(s, x_2) ds, \quad (7)$$

$$f_1(x_1 | x_2) = \begin{cases} f(x_1, x_2)/f_2(x_2) & \text{if } f_2(x_2) > 0 \\ f_1(x_1) & \text{if } f_2(x_2) = 0. \end{cases} \quad (8)$$

By Fubini's theorem, $f_i(x_i)$ is a Borel function on \mathbb{R}^1 for $i = 1, 2$, and so $f_1(x_1 | x_2)$ is a Borel function on \mathbb{R}^2 . For $B \in \mathcal{B}^2$ and $x = (x_1, x_2) \in \mathbb{R}^2$ define

$$P(B, x) = \int_{\{s : (s, x_2) \in B\}} f_1(s | x_2) ds. \quad (9)$$

Then for each $x \in \mathbb{R}^2$, $P(B, x)$ is a probability measure on \mathcal{B}^2 , and for each $B \in \mathcal{B}^2$, $P(B, x)$ is a Borel function in x_2 and hence $\sigma(X_2)$ -measurable. Moreover, for $B \in \mathcal{B}^2$ and $A_2 = \mathbb{R} \times B_2 \in \sigma(X_2)$

$$\begin{aligned} \int_{A_2} P(B, x) dP &= \int_{B_2} \int_{-\infty}^{\infty} P(B, (s, t)) f(s, t) ds dt \\ &= \int_{B_2} \int_{-\infty}^{\infty} \left[\int_{\{u : (u, t) \in B\}} f_1(u | t) du \right] f(s, t) ds dt \\ &= \int_{B_2} \int_{\{u : (u, t) \in B\}} f_1(u | t) f_2(t) du dt \\ &= \int_{B_2} \int_{\{u : (u, t) \in B\}} f(u, t) du dt = \int_{B_2} \int_{-\infty}^{\infty} I_B f(u, t) du dt \\ &= P\{B \cdot A_2\}, \end{aligned}$$

and so according to (2), $P(B, x) = P\{B | X_2\}(x)$, a.c., for each $B \in \mathcal{B}^2$. Consequently, $P(B, x)$ is a regular conditional probability measure on \mathcal{B}^2 given $\sigma(X_2)$. Hence, by Theorem 1 for any Borel function h on \mathbb{R}^2 with

$$|\mathbb{E} h(X_1, X_2)| \leq \infty$$

$$\mathbb{E}\{h(X_1, X_2) | X_2\} = \int_{-\infty}^{\infty} h(s, X_2) f_1(s | X_2) ds, \quad \text{a.c.} \quad (10)$$

The Borel function $f_1(x_1 | x_2)$ is called the **conditional density** of X_1 given $X_2 = x_2$, while $\int_{-\infty}^{\infty} h(s, x_2) f_1(s | x_2) ds$ is termed the conditional expectation of $h(X_1, X_2)$ given that $X_2 = x_2$ and is denoted by

$$\mathbb{E}\{h(X_1, X_2) | X_2 = x_2\}.$$

Analogously, $f_2(x_2 | x_1)$ is the conditional density of X_2 given $X_1 = x_1$.

Moreover,

$$F_1(x_1 | x_2) = \int_{[u < x_1]} f_1(u | x_2) du, \quad F_2(x_2 | x_1) = \int_{[u < x_2]} f_2(u | x_1) du$$

are the **conditional d.f.s** of X_1 given $X_2 = x_2$ and of X_2 given $X_1 = x_1$ respectively, and from (10)

$$\mathbb{P}\{X_1 < x_1 | X_2\} = F_1(x_1 | X_2), \quad \mathbb{P}\{X_2 < x_2 | X_1\} = F_2(x_2 | X_1).$$

As noted in Chapter 1, any random vector $X = (X_1, \dots, X_n)$ on (Ω, \mathcal{F}, P) induces a probability measure P_X on the class \mathcal{B}^n of n -dimensional Borel sets via $P_X\{B\} = P\{X^{-1}(B)\}$, $B \in \mathcal{B}^n$. For any σ -algebra $\mathcal{G} \subset \mathcal{F}$ if $P_X\{B | \mathcal{G}\}$ is defined by $P_X\{B | \mathcal{G}\} = P\{X^{-1}(B) | \mathcal{G}\}$, $B \in \mathcal{B}^n$, then $P_X\{B | \mathcal{G}\}$ is a \mathcal{G} -measurable r.v. for each $B \in \mathcal{B}^n$. Moreover, for any sequence of disjoint sets $B_m \in \mathcal{B}^n$, $m \geq 1$, there exists a null set $N \in \mathcal{G}$ depending in general on $\{B_m, m \geq 1\}$ such that

$$P_X\left\{\bigcup_{m=1}^{\infty} B_m \mid \mathcal{G}\right\}(\omega) = \sum_{m=1}^{\infty} P_X\{B_m | \mathcal{G}\}(\omega), \quad \omega \in N^c. \quad (11)$$

Can a version of $P_X\{B | \mathcal{G}\}$ be chosen for each $B \in \mathcal{B}^n$ so that on the complement of a single null set $N \in \mathcal{G}$, (11) holds for all disjoint sequences $\{B_m, m \geq 1\} \subset \mathcal{B}^n$? An affirmative answer to this case has been given by Doob in Theorem 2 (below)

Definition. Let $X = (X_1, \dots, X_n)$ or (X_1, X_2, \dots) be a stochastic process on (Ω, \mathcal{F}, P) and \mathcal{G} a σ -algebra of events. A function $P_X(B, \omega)$ on $\mathcal{B}^n \times \Omega$, $1 \leq n \leq \infty$, is termed a **regular conditional distribution for X given \mathcal{G}** provided that for each $\omega \in \Omega$, $P_X(B, \omega)$ is a probability measure on \mathcal{B}^n and that for $B \in \mathcal{B}^n$, $P_X(B, \omega)$ is \mathcal{G} -measurable with

$$P_X(B, \omega) = P\{X^{-1}(B) | \mathcal{G}\}, \quad \text{a.c.} \quad (12)$$

Theorem 2 (Doob). If $X = (X_1, \dots, X_n)$ is a random vector on a probability space (Ω, \mathcal{F}, P) and \mathcal{G} is a σ -algebra of events, there exists a regular conditional distribution for X given \mathcal{G} .

PROOF. For any rational numbers λ_i , $1 \leq i \leq n$, define

$$F_n^\omega(\lambda_1, \dots, \lambda_n) = P\left\{\bigcap_{i=1}^n [X_i < \lambda_i] \mid \mathcal{G}\right\}(\omega). \quad (13)$$

By the properties enumerated in (4) there is a null set $N \in \mathcal{G}$ such that for $\omega \in N^c$ and all rational numbers $\lambda_i, \lambda'_i, r_{im}$

$$F_n^\omega(\lambda_1, \dots, \lambda_n) \geq F_n^\omega(\lambda'_1, \dots, \lambda'_n) \quad \text{if } \lambda_i > \lambda'_i, 1 \leq i \leq n, \quad (i)$$

$$F_n^\omega(\lambda_1, \dots, \lambda_n) = \lim_{\substack{r_{im} \uparrow \lambda_i \\ 1 \leq i \leq n}} F_n^\omega(r_{1m}, \dots, r_{nm}), \quad (ii)$$

$$\lim_{\lambda_i \rightarrow -\infty} F_n^\omega(\lambda_1, \dots, \lambda_n) = 0, \quad 1 \leq i \leq n, \quad \lim_{\substack{\lambda_i \rightarrow \infty \\ 1 \leq i \leq n}} F_n^\omega(\lambda_1, \dots, \lambda_n) = 1, \quad (\text{iii})$$

$$\Delta_n^{\lambda, \lambda'} F_n^\omega \geq 0 \quad \text{if } \lambda \leq \lambda', \quad (\text{iv})$$

where the notation in (iv) is that of (5) of Section 6.3; also, $\lambda \leq \lambda'$ signifies that the inequality holds for each coordinate. Define for any real numbers μ_i and $\omega \in N^c$

$$F_n^\omega(\mu_1, \dots, \mu_n) = \lim_{\substack{\lambda_i \rightarrow \mu_i^- \\ i \leq i \leq n}} F_n^\omega(\lambda_1, \dots, \lambda_n), \quad \lambda_i \text{ rational}, 1 \leq i \leq n, \quad (14)$$

while for $\omega \in N$, set

$$F_n^\omega(\mu_1, \dots, \mu_n) = P \left\{ \bigcap_{i=1}^n [X_i < \mu_i] \right\}. \quad (15)$$

Then for each $\omega \in \Omega$, $F_n^\omega(\mu_1, \dots, \mu_n)$ is an n -dimensional d.f. and hence determines a Lebesgue-Stieltjes measure μ_ω on \mathcal{B}^n with $\mu_\omega(R^n) = 1$. For $B \in \mathcal{B}^n$ and $\omega \in \Omega$ define

$$P_X(B, \omega) = \mu_\omega(B). \quad (16)$$

If

$$\mathcal{H} = \{B: B \in \mathcal{B}^n, P_X(B, \omega) \text{ is } \mathcal{G}\text{-measurable}, P_X(B, \omega) = P\{X^{-1}(B)|\mathcal{G}\}, \text{a.c.}\},$$

$$\mathcal{D} = \left\{ B: B = \bigcup_{i=1}^n (-\infty, \lambda_i], \text{ rational } \lambda_i \in (-\infty, \infty), 1 \leq i \leq n \right\}$$

then \mathcal{H} is a λ -class, \mathcal{D} is a π -class, and by (13) through (16) $\mathcal{H} \supset \mathcal{D}$. Hence, by Theorem 1.3.2, $\mathcal{H} \supset \sigma(\mathcal{D}) = \mathcal{B}^n$, that is, $P_X(B, \omega)$ is a regular conditional distribution for X given \mathcal{G} . \square

Corollary 1. If $X = (X_1, X_2, \dots)$ is a (countable) stochastic process on (Ω, \mathcal{F}, P) and \mathcal{G} is a σ -algebra of events, there exists a regular conditional distribution for X given \mathcal{G} .

PROOF. For all $n \geq 1$ and rational λ_i , $1 \leq i \leq n$, choose $F_n^\omega(\lambda_1, \dots, \lambda_n)$ as in (13). Select the null set $N \in \mathcal{G}$ such that in addition to properties (i)–(iv), for $\omega \in N^c$

$$\lim_{\lambda_{n+1} \rightarrow \infty} F_{n+1}^\omega(\lambda_1, \dots, \lambda_{n+1}) = F_n^\omega(\lambda_1, \dots, \lambda_n), \quad n \geq 1. \quad (v)$$

Define $F_n^\omega(\mu_1, \dots, \mu_n)$ as in (14), (15). Then for each $\omega \in \Omega$, $\{F_n^\omega, n \geq 1\}$ is a consistent family of d.f.s and hence by Theorem 6.4.3 determines a measure μ_ω on $(R^\infty, \mathcal{B}^\infty)$. Define $P_X(B, \omega) = \mu_\omega(B)$, $B \in \mathcal{B}^\infty$, and

$$\mathcal{H} = \{B: B \in \mathcal{B}^\infty, P_X(B, \omega) \text{ is } \mathcal{G}\text{-measurable},$$

$$P_X(B, \omega) = P\{X^{-1}(B)|\mathcal{G}\}, \text{a.c.},$$

$$\mathcal{D} = \bigcup_{n=1}^\infty \left\{ B: B \in \mathcal{B}^\infty, B = \bigcup_{i=1}^n (-\infty, \lambda_i] \times R \times R \times \dots \text{ rational } \lambda_i \in (-\infty, \infty) \right\},$$

whence \mathcal{H} is a λ -class, \mathcal{D} is a π -class, and $\mathcal{H} \supset \mathcal{D}$, so that by Theorem 1.3.2 $\mathcal{H} \supset \sigma(\mathcal{D}) = \mathcal{B}^\infty$. \square

Corollary 2. Let $X = (X_1, \dots, X_n)$ or (X_1, X_2, \dots) be a stochastic process on (Ω, \mathcal{F}, P) and \mathcal{G} a σ -algebra of events. If $P_X^\omega(B) = P_X(B, \omega)$ is a regular conditional distribution for X given \mathcal{G} and h is a Borel function on R^n , $1 \leq n \leq \infty$, with $|E h(X)| \leq \infty$, then

$$E\{h(X) | \mathcal{G}\}(\omega) = \int_{R^n} h(x) dP_X^\omega(x), \quad \text{a.c.,}$$

where $P_X^\omega(x) = P_X^\omega(\bigcup_{i=1}^n (-\infty, x_i])$ for $x = (x_1, x_2, \dots, x_n)$.

PROOF. As in Theorem 1. \square

The question naturally poses itself whether a regular conditional distribution for X given \mathcal{G} engenders a corresponding regular conditional probability on $\sigma(X)$ given \mathcal{G} . The answer is positive under the proviso of

Theorem 3. Let $X = (X_1, \dots, X_n)$ or (X_1, X_2, \dots) be a stochastic process on (Ω, \mathcal{F}, P) and \mathcal{G} a σ -algebra of events. If the range of X , i.e., $\{X(\omega); \omega \in \Omega\}$ is a Borel set, then there exists a regular conditional probability on $\sigma(X)$ given \mathcal{G} .

PROOF. Let $P_X(B, \omega)$ be a regular conditional distribution for X given \mathcal{G} and set $S = \{X(\omega); \omega \in \Omega\}$. Then

$$P_X(S, \omega) = P\{X^{-1}(S) | \mathcal{G}\}(\omega) = P\{\Omega | \mathcal{G}\}(\omega) = 1$$

except for $\omega \in N \in \mathcal{G}$ with $P(N) = 0$. If for $A \in \sigma(X)$ there are two sets $B_1, B_2 \in \mathcal{B}^\infty$ with $A = X^{-1}(B_1) = X^{-1}(B_2)$, then $B_1 \Delta B_2 \subset S^c$, implying

$$P_X(B_1, \omega) = P_X(B_1 B_2, \omega) = P_X(B_2, \omega), \quad \omega \in N^c.$$

Hence, for $A \in \sigma(X)$ it is permissible to define

$$P(A, \omega) = P_X(B, \omega), \quad A = X^{-1}(B), \quad \omega \in N^c,$$

and for any fixed $\omega_0 \in N^c$ set

$$P(A, \omega) = P_X(B, \omega_0), \quad \omega \in N.$$

Then for each $\omega \in \Omega$, $P(A, \omega)$ is a probability measure on $\sigma(X)$, and for each $A \in \sigma(X)$, $P(A, \omega)$ is \mathcal{G} -measurable with

$$P(A, \omega) = P_X(B, \omega) = P\{A | \mathcal{G}\}(\omega), \quad \text{a.c.} \quad \square$$

The preceding notions will be utilized to prove the conditional Hölder inequality.

Theorem 4. If X, Y are random variables on (Ω, \mathcal{F}, P) , \mathcal{G} is a σ -algebra of events and $1 < p < \infty$, $(1/p) + (1/q) = 1$, then

$$E\{|XY| | \mathcal{G}\} \leq E^{1/p}\{|X|^p | \mathcal{G}\} \cdot E^{1/q}\{|Y|^q | \mathcal{G}\}, \quad \text{a.c.} \quad (17)$$

PROOF. For $B \in \mathcal{B}^2$ and $\omega \in \Omega$, let $\hat{P}_\omega(B) = P(B, \omega)$ be a regular conditional distribution for (X, Y) given \mathcal{G} . By Corollary 2

$$\begin{aligned} E\{|XY||\mathcal{G}\}(\omega) &= \int_{R^2} |x_1 x_2| d\hat{P}_\omega, \quad \text{a.c.,} \\ E^{1/p}\{|X|^p|\mathcal{G}\}(\omega) &= \left(\int_{R^2} |x_1|^p d\hat{P}_\omega \right)^{1/p}, \quad \text{a.c.,} \\ E^{1/q}\{|Y|^q|\mathcal{G}\}(\omega) &= \left(\int_{R^2} |x_2|^q d\hat{P}_\omega \right)^{1/q}, \quad \text{a.c.} \end{aligned}$$

Since for each $\omega \in \Omega$, \hat{P}_ω is a probability measure on \mathcal{B}^2 , (17) follows from the ordinary Hölder inequality. \square

EXERCISES 7.2

1. If the events $\{B_j\}$ of (Ω, \mathcal{F}, P) constitute a finite or σ -partition of Ω in \mathcal{F} and have positive probabilities, prove **Bayes' theorem**, i.e., for any event A

$$P\{B_i | A\} = \frac{P\{A | B_i\}P\{B_i\}}{\sum_j P\{A | B_j\}P\{B_j\}}.$$

2. If in Example 2, $F_i(x_i) = \lim_{x_{j+1} \rightarrow \infty} F(x_1, x_2)$, $i = 1, 2$, prove that the d.f. F_1 is a mixture of the family $F(x_1 | x_2)$ in the sense of Exercise 8.2.3.
3. If $X = (X_1, \dots, X_n)$ is a random vector whose d.f. F is absolutely continuous relative to Lebesgue measure with density f find the conditional density of (X_1, \dots, X_j) given $X_{j+1} = x_{j+1}, \dots, X_n = x_n$, where $1 \leq j < n$ and verify the existence of a regular conditional probability measure on \mathcal{B}^n given $\sigma(X_{j+1}, \dots, X_n)$.
4. If r.v.s X and Y have a joint normal distribution (Exercise 6.5.12), show that the conditional distributions of X given Y and Y given X are both normal. Is the converse true?
5. If X_1, \dots, X_k have a multinomial distribution (Exercise 6.3.9) find the conditional distribution of X_1 given X_2 , the conditional mean $E\{X_1 | X_2\}$, and the conditional variance $\sigma^2(X_1 | X_2)$.
6. (D. G. Kendall) If A_1, \dots, A_n are interchangeable events on (Ω, \mathcal{F}, P) and $S_n =$ number of events A_1, \dots, A_n that occur, prove for $1 \leq i_1 < \dots < i_j \leq n$ and $1 \leq j \leq N \leq n$ that
- $$P\{A_{i_1} \cdots A_{i_j} | S_n = N\} = \binom{n-j}{N-j} / \binom{n}{N}.$$
7. Let (Ω, \mathcal{F}, P) be a probability space and \mathcal{G} a σ -algebra of events. If $A \in \mathcal{F}$ and $B = \{\omega : P\{A | \mathcal{G}\} > 0\}$, verify that (i) $B \in \mathcal{G}$, (ii) $P\{A - B\} = 0$, and, moreover, if B' satisfies (i) and (ii) then $P\{B - B'\} = 0$ (i.e., B is a \mathcal{G} -measurable cover of A).
8. Prove the conditional Minkowski inequality (Exercise 6.2.7) and also the conditional Markov inequality $E\{|Y||\mathcal{G}\} \geq cP\{|Y \geq c| \mathcal{G}\}$ for any $c > 0$, $Y \in \mathcal{L}_1$, and sub- σ -algebra \mathcal{G} .

7.3 Conditional Independence, Interchangeable Random Variables

The notion of conditional independence parallels that of unconditional or ordinary independence. As usual, all events and random variables pertain to the basic probability space (Ω, \mathcal{F}, P) .

Definition. Let \mathcal{G} be a σ -algebra of events and $\{\mathcal{G}_n, n \geq 1\}$ a sequence of classes of events. Two such classes \mathcal{G}_1 and \mathcal{G}_2 are said to be **conditionally independent given \mathcal{G}** if for all $A_i \in \mathcal{G}_i, i = 1, 2$,

$$P\{A_1 A_2 | \mathcal{G}\} = P\{A_1 | \mathcal{G}\} \cdot P\{A_2 | \mathcal{G}\}, \quad \text{a.c.} \quad (1)$$

More generally, the sequence $\{\mathcal{G}_n, n \geq 1\}$ is declared **conditionally independent given \mathcal{G}** if for all choices of $A_m \in \mathcal{G}_{k_m}$, where $k_i \neq k_j$ for $i \neq j$, $m = 1, 2, \dots, n$, and $n = 2, 3, \dots$,

$$P\{A_1 A_2 \cdots A_n | \mathcal{G}\} = \prod_{i=1}^n P\{A_i | \mathcal{G}\}, \quad \text{a.c.} \quad (2)$$

Finally, a sequence $\{X_n, n \geq 1\}$ of random variables is called **conditionally independent given \mathcal{G}** if the sequence of classes $\mathcal{G}_n = \sigma(X_n), n \geq 1$, is conditionally independent given \mathcal{G} .

If $\mathcal{G} = \{\emptyset, \Omega\}$, then conditional independence given \mathcal{G} coalesces to ordinary (unconditional) independence, while if $\mathcal{G} = \mathcal{F}$, then every sequence of classes of events is conditionally independent given \mathcal{G} .

Independent r.v.s $\{X_n\}$ may lose their independence under conditioning; for example, if $\{X_n\}$ are Bernoulli trials with parameter p and $S_n = \sum_1^n X_i$, then $P\{X_i = 1 | S_2\} > 0, i = 1, 2$ for $S_2 = 0$ or 2, whereas $P\{X_1 = 1, X_2 = 1 | S_2\} = 0$ when $S_2 = 0$.

On the other hand, dependent r.v.s may gain independence under conditioning, i.e., become conditionally independent. Thus, if $\{X_n\}$ are independent integer valued r.v.s (see Exercise 7.3.3) and $S_n = \sum_1^n X_i$, clearly the r.v.s $\{S_n, n \geq 1\}$ are dependent. However, given that the event $\{S_2 = k\}$ (of positive probability) occurs,

$$\begin{aligned} P\{S_1 = i, S_3 = j | S_2\} &= \frac{P\{S_1 = i, S_2 = k, S_3 = j\}}{P\{S_2 = k\}} \\ &= \frac{P\{S_1 = i\}P\{X_2 = k - i\}P\{X_3 = j - k\}}{P\{S_2 = k\}} \\ &= P\{S_1 = i | S_2\} \frac{P\{X_3 = j - k, S_2 = k\}}{P\{S_2 = k\}} \\ &= P\{S_1 = i | S_2\}P\{S_3 = j | S_2\}. \end{aligned}$$

If the subscript n of S_n is interpreted as "time" and the r.v.s S_1, S_2, S_3 are envisaged as chance occurrences of the past, present and future times respectively, then the prior relationship may be picturesquely stated as, "The past and future are conditionally independent given the present." In fact, designating the r.v. S_n as "the present," the r.v.s S_1, \dots, S_{n-1} as "the past," and the r.v.s S_{n+1}, \dots, S_{n+m} as "the future," it may be verified for any $n \geq 2, m > 0$, that the past and the future are conditionally independent given the present. This property holds not only for sums S_n of independent r.v.s but more generally when the r.v.s $\{S_n, n \geq 1\}$ constitute a Markov chain (Exercise 7.3.1).

Theorem 1. If $\mathcal{G}_i, i = 1, 2, 3$, are σ -algebras of events, then conditional independence of \mathcal{G}_1 and \mathcal{G}_2 given \mathcal{G}_3 is equivalent to any one of the following:

- i. For all $A_1 \in \mathcal{G}_1$,

$$P\{A_1 | \sigma(\mathcal{G}_2 \cup \mathcal{G}_3)\} = P\{A_1 | \mathcal{G}_3\}, \quad \text{a.c.} \quad (3)$$

- ii. For every $D \in \mathcal{D}$ where \mathcal{D} is a π -class with $\mathcal{G}_1 = \sigma(\mathcal{D})$,

$$P\{D | \sigma(\mathcal{G}_2 \cup \mathcal{G}_3)\} = P\{D | \mathcal{G}_3\}, \quad \text{a.c.} \quad (4)$$

- iii. For every $D_i \in \mathcal{D}_i$ where \mathcal{D}_i is a π -class with $\mathcal{G}_i = \sigma(\mathcal{D}_i), i = 1, 2$,

$$P\{D_1 D_2 | \mathcal{G}_3\} = P\{D_1 | \mathcal{G}_3\} \cdot P\{D_2 | \mathcal{G}_3\}, \quad \text{a.c.} \quad (5)$$

- iv. For every $\sigma(\mathcal{G}_1 \cup \mathcal{G}_3)$ -measurable function X with $|E X| \leq \infty$,

$$E\{X | \sigma(\mathcal{G}_2 \cup \mathcal{G}_3)\} = E\{X | \mathcal{G}_3\}, \quad \text{a.c.} \quad (6)$$

PROOF. (i) $(1) \Rightarrow (3)$: Let $A_i \in \mathcal{G}_i, i = 1, 2, 3$. By the definition of conditional expectation, (1), and Theorem 7.1.3,

$$\begin{aligned} \int_{A_2 A_3} I_{A_1} &= \int_{A_3} I_{A_1 A_2} = \int_{A_3} P\{A_1 A_2 | \mathcal{G}_3\} = \int_{A_3} P\{A_1 | \mathcal{G}_3\} \cdot P\{A_2 | \mathcal{G}_3\} \\ &= \int_{A_3} I_{A_2} P\{A_1 | \mathcal{G}_3\} = \int_{A_2 A_3} P\{A_1 | \mathcal{G}_3\}, \end{aligned}$$

whence (3) obtains by Theorem 7.1.1(ii). Conversely, by 7.1, (14)(v), Theorem 7.1.3, and (3),

$$\begin{aligned} P\{A_1 A_2 | \mathcal{G}_3\} &= E\{P\{A_1 A_2 | \sigma(\mathcal{G}_2 \cup \mathcal{G}_3)\} | \mathcal{G}_3\} \\ &= E\{I_{A_2} P\{A_1 | \sigma(\mathcal{G}_2 \cup \mathcal{G}_3)\} | \mathcal{G}_3\} \\ &= E\{I_{A_2} P\{A_1 | \mathcal{G}_3\} | \mathcal{G}_3\} = P\{A_1 | \mathcal{G}_3\} P\{A_2 | \mathcal{G}_3\}, \quad \text{a.c.} \end{aligned}$$

Apropos of (ii), since $(3) \Rightarrow (4)$ trivially, it suffices to prove the reverse implication. If

$$\mathcal{A} = \{A : A \in \mathcal{F}, P\{A | \sigma(\mathcal{G}_2 \cup \mathcal{G}_3)\} = P\{A | \mathcal{G}_3\}, \text{a.c.}\},$$

then \mathcal{A} is a λ -class $\supset \mathcal{D}$ via (4), and so by Theorem 1.3.2, $\mathcal{A} \supset \sigma(\mathcal{D}) = \mathcal{G}_1$, which is tantamount to (3).

Concerning (iii), it suffices to prove (5) \Rightarrow (1), the converse being trivial. Since

$$\mathcal{A} = \{A_2 \in \mathcal{F} : P\{D_1 A_2 | \mathcal{G}_3\} = P\{D_1 | \mathcal{G}_3\} \cdot P\{A_2 | \mathcal{G}_3\}, \text{ a.c., all } D_1 \in \mathcal{D}_1\},$$

$$\mathcal{A}^* = \{A_1 \in \mathcal{F} : P\{A_1 A_2 | \mathcal{G}_3\} = P\{A_1 | \mathcal{G}_3\} \cdot P\{A_2 | \mathcal{G}_3\}, \text{ a.c., all } A_2 \in \mathcal{G}_2\}$$

are both λ -classes and $\mathcal{A} \supset \mathcal{D}_2$ by (5), Theorem 1.3.2 ensures that $\mathcal{A} \supset \sigma(\mathcal{D}_2) = \mathcal{G}_2$. Then a reapplication of this theorem yields $\mathcal{A}^* \supset \sigma(\mathcal{D}_1) = \mathcal{G}_1$ and (1) obtains.

Finally, to verify (iv) note that, taking $X = I_{A_1}$, (6) \Rightarrow (3), and hence it suffices to prove (3) \Rightarrow (6). For $A_i \in \mathcal{G}_i$, $i = 1, 3$ via (3),

$$\begin{aligned} P\{A_1 A_3 | \sigma(\mathcal{G}_2 \cup \mathcal{G}_3)\} &= I_{A_3} P\{A_1 | \sigma(\mathcal{G}_2 \cup \mathcal{G}_3)\} = I_{A_3} P\{A_1 | \mathcal{G}_3\} \\ &= P\{A_1 A_3 | \mathcal{G}_3\}, \quad \text{a.c.} \end{aligned}$$

If \mathcal{A} denotes the class of non-negative \mathcal{F} -measurable functions for which (6) holds, then \mathcal{A} is a λ -system which, as just shown, contains the indicator functions of all sets of a π -class generating $\sigma(\mathcal{G}_1 \cup \mathcal{G}_3)$. Hence, by Theorem 1.4.3, \mathcal{A} contains all non-negative $\sigma(\mathcal{G}_1 \cup \mathcal{G}_3)$ -measurable functions and therefore (6) holds. \square

Corollary 1. Random variables X_1, \dots, X_n on (Ω, \mathcal{F}, P) are conditionally independent given a σ -algebra \mathcal{G} of events iff for all $(x_1, \dots, x_n) \in R^n$

$$P\left\{\bigcap_{i=1}^n [X_i < x_i] \mid \mathcal{G}\right\} = \prod_{i=1}^n P\{X_i < x_i | \mathcal{G}\}, \quad \text{a.c.}$$

PROOF. This follows easily from (iii). \square

A direct application of (6) yields

Corollary 2. If random variables X_1 and X_2 are conditionally independent given a σ -algebra \mathcal{G} of events and $|\mathbb{E} X_1| \leq \infty$, then

$$\mathbb{E}\{X_1 | \sigma(\mathcal{G} \cup \sigma(X_2))\} = \mathbb{E}\{X_1 | \mathcal{G}\}, \quad \text{a.c.}$$

Corollary 3. Let \mathcal{G}_i be a σ -algebra of events, $i = 1, 2, 3$. If $\sigma(\mathcal{G}_1 \cup \mathcal{G}_3)$ is independent of \mathcal{G}_2 , then \mathcal{G}_1 and \mathcal{G}_2 are conditionally independent given \mathcal{G}_3 (as well as unconditionally independent).

PROOF. Since (3) with the subscripts 1 and 2 interchanged holds via Exercise 7.1.1, this follows directly from Theorem 1. \square

Recall that r.v.s $\{X_n, n \geq 1\}$ are called **interchangeable** if the joint distribution of every finite subset of k of these r.v.s depends only upon k and not the particular subset, $k \geq 1$.

Definition. A mapping $\pi = (\pi_1, \pi_2, \dots)$ from the set N of all positive integers onto itself is called a **finite permutation** if π is one-to-one and $\pi_n = n$ for all but a finite number of integers. Let Q denote the set of all finite permutations π and let \mathcal{B}^∞ be the class of Borel subsets of $R^\infty = R \times R \times \dots$ and $X = (X_1, X_2, \dots)$ a sequence of r.v.s on (Ω, \mathcal{F}, P) . Define $\pi X = (X_{\pi_1}, X_{\pi_2}, \dots)$ for $\pi = (\pi_1, \pi_2, \dots)$. Then

$$\mathcal{S} = \{X^{-1}(B) : B \in \mathcal{B}^\infty, P\{X^{-1}(B) \Delta (\pi X)^{-1}(B)\} = 0, \text{all } \pi \in Q\}, \quad (7)$$

is called the **σ -algebra of permutable events** (of X).

Theorem 2. Random variables $X_n, n = 1, 2, \dots$, on (Ω, \mathcal{F}, P) are interchangeable iff they are conditionally independent and identically distributed given some σ -algebra \mathcal{G} of events. Moreover, \mathcal{G} can be taken to be either the σ -algebra \mathcal{S} of permutable events or the tail σ -algebra \mathcal{T} , and

$$P\{X_1 < x_1 | \mathcal{S}\} = P\{X_1 < x_1 | \mathcal{T}\}, \quad \text{a.c.} \quad (8)$$

PROOF. Sufficiency is immediate since if

$$P\{X_1 < x_1, \dots, X_m < x_m | \mathcal{G}\} = \prod_{j=1}^m P\{X_j < x_j | \mathcal{G}\}, \quad \text{a.c., } m \geq 1,$$

it follows upon taking expectations that $\{X_n, n \geq 1\}$ are interchangeable.

Apropos of necessity, for $n \geq 1$ and any real x , define

$$\xi_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i < x\}}. \quad (9)$$

Then

$$E[\xi_n(x) - \xi_m(x)]^2 = \frac{|m-n|}{m \cdot n} (P\{X_1 < x\} - P\{X_1 < x, X_2 < x\}) \rightarrow 0$$

as $m, n \rightarrow \infty$, whence $\xi_n(x) \xrightarrow{P}$ some \mathcal{T} -measurable r.v. $\xi(x)$, implying (Exercise 3.3.1(iv))

$$\prod_{j=1}^m \xi_n(x_j) \xrightarrow{P} \prod_{j=1}^m \xi(x_j), \quad m \geq 1. \quad (10)$$

Let $\alpha_1, \dots, \alpha_m$ denote positive integers and set

$$M = \{\alpha : \alpha = (\alpha_1, \dots, \alpha_m), \text{ where } 1 \leq \alpha_i \leq n \text{ for } 1 \leq i \leq m\},$$

$$N = \{\alpha : \alpha \in M, \alpha_i \neq \alpha_j, i \neq j, 1 \leq i, j \leq m\},$$

$$I(\alpha) = I_{\{X_{\alpha_1} < x_1, \dots, X_{\alpha_m} < x_m\}}.$$

Whenever $n > m$, taking cognizance of (9),

$$\begin{aligned} \prod_{j=1}^m \xi_n(x_j) &= \frac{1}{n^m} \prod_{j=1}^m \sum_{i=1}^n I_{\{X_i < x_j\}} = \frac{1}{n^m} \sum_{\alpha \in M} I(\alpha) \\ &= \frac{1}{n^m} \left(\sum_{\alpha \in M - N} + \sum_{\alpha \in N} \right) I(\alpha). \end{aligned} \quad (11)$$

Since as $n \rightarrow \infty$

$$\frac{1}{n^m} \sum_{x \in M - N} I(x) \leq \frac{1}{n^m} \sum_{x \in M - N} 1 = \frac{n^m - n(n-1) \cdots (n-m+1)}{n^m} \rightarrow 0$$

for each element $A = X^{-1}(B) \in \mathcal{S}$, (11) entails

$$\int_A \prod_{j=1}^m \xi_n(x_j) dP = n^{-m} \sum_{x \in N} \int_{X^{-1}(B)} I(x) dP + o(1). \quad (12)$$

Now for any $\alpha \in N$, if $\pi \in Q$ is chosen so that $\pi_i = \alpha_i$, $1 \leq i \leq m$, then by definition of \mathcal{S} and interchangeability

$$\begin{aligned} \int_{X^{-1}(B)} I(\alpha) dP &= \int_{(\pi X)^{-1}(B)} I_{\{X_{\pi_1} < x_1, \dots, X_{\pi_m} < x_m\}} dP \\ &= \int_{X^{-1}(B)} I_{\{X_1 < x_1, \dots, X_m < x_m\}} dP. \end{aligned}$$

Thus, from (12)

$$\begin{aligned} \int_A \prod_{j=1}^m \xi_n(x_j) dP &= \frac{n(n-1) \cdots (n-m+1)}{n^m} \\ &\times \int_{X^{-1}(B)} I_{\{X_1 < x_1, \dots, X_m < x_m\}} dP + o(1) \\ &\rightarrow \int_A I_{\{X_1 < x_1, \dots, X_m < x_m\}} dP. \end{aligned} \quad (13)$$

On the other hand, the left side of (13) tends to $\int_A \prod_{j=1}^m \xi(x_j) dP$ by (10) and dominated convergence. Hence for any $A \in \mathcal{S}$, all real x_1, \dots, x_m , and positive integers m ,

$$\int_A \prod_{j=1}^m \xi(x_j) dP = \int_A I_{\{X_1 < x_1, \dots, X_m < x_m\}} dP,$$

and therefore

$$E \left\{ \prod_{j=1}^m \xi(x_j) \mid \mathcal{S} \right\} = P\{X_1 < x_1, \dots, X_m < x_m \mid \mathcal{S}\}, \quad \text{a.c.} \quad (14)$$

Since $\mathcal{S} \supset \mathcal{F}$ and $\xi(x)$ is \mathcal{F} -measurable for every x , (14) entails

$$P\{X_1 < x_1, \dots, X_m < x_m \mid \mathcal{S}\} = \prod_{j=1}^m \xi(x_j), \quad \text{a.c.} \quad (15)$$

$$P\{X_1 < x_1, \dots, X_m < x_m \mid \mathcal{F}\} = E \left\{ \prod_{j=1}^m \xi(x_j) \mid \mathcal{F} \right\} = \prod_{j=1}^m \xi(x_j), \quad \text{a.c.} \quad (16)$$

In particular, for $m = 1$, $\xi(x_1) = P\{X_1 < x_1 \mid \mathcal{S}\} = P\{X_1 < x_1 \mid \mathcal{F}\}$, a.c., and by (15), (16)

$$\begin{aligned} \mathbf{P}\{X_1 < x_1, \dots, X_m < x_m \mid \mathcal{S}\} &= \prod_1^m \mathbf{P}\{X_1 < x_i \mid \mathcal{S}\} = \prod_1^m \mathbf{P}\{X_1 < x_i \mid \mathcal{T}\} \\ &= \mathbf{P}\{X_1 < x_1, \dots, X_m < x_m \mid \mathcal{T}\}, \quad \text{a.c.} \end{aligned} \tag{17}$$

Finally, for any $j = 2, \dots, m$, letting $x_i \rightarrow \infty$ for $i \neq j$, $1 \leq i \leq m$, in (17),

$$\begin{aligned} \mathbf{P}\{X_j < x_j \mid \mathcal{S}\} &= \mathbf{P}\{X_1 < x_j \mid \mathcal{S}\} = \mathbf{P}\{X_1 < x_j \mid \mathcal{T}\} \\ &= \mathbf{P}\{X_j < x_j \mid \mathcal{T}\}, \quad \text{a.c.} \end{aligned} \tag{18}$$

Together, (17) and (18) assert that $\{X_n, n \geq 1\}$ are conditionally i.i.d. given either \mathcal{S} or \mathcal{T} and that (8) holds. \square

Corollary 4 (de Finetti). *If $\{X_n, n \geq 1\}$ are interchangeable r.v.s, there exists a σ -algebra \mathcal{G} of events such that for all $m \geq 1$*

$$\mathbf{P}\{X_1 < x_1, \dots, X_m < x_m\} = \int \prod_{j=1}^m \mathbf{P}\{X_1 < x_j \mid \mathcal{G}\} d\mathbf{P}.$$

Corollary 5. *If r.v.s $\{X_n, n \geq 1\}$ are conditionally independent given some σ -algebra \mathcal{G} of events, there exists a regular conditional distribution, say \mathbf{P}^ω , for $X = (X_1, X_2, \dots)$ given \mathcal{G} such that for each $\omega \in \Omega$ the coordinate r.v.s $\{\xi_n, n \geq 1\}$ of the probability space $(\mathbb{R}^\infty, \mathcal{B}^\infty, \mathbf{P}^\omega)$ are independent. Moreover, if $X_j \in \mathcal{L}_1$, $1 \leq j \leq n$, and $\mathbf{E}[X_1 X_2 \cdots X_n]$ exists,*

$$\mathbf{E}[X_1 X_2 \cdots X_n \mid \mathcal{G}] = \prod_{i=1}^n \mathbf{E}[X_i \mid \mathcal{G}], \quad \text{a.c., } n \geq 2. \tag{19}$$

PROOF. For any rational λ , define $F_i^\omega(\lambda) = \mathbf{P}\{X_i < \lambda \mid \mathcal{G}\}(\omega)$, $i \geq 1$. There exists a null set $N \in \mathcal{G}$ such that for $\omega \in N^c$, all positive integers n , and all rational $\lambda, \lambda', r, r_1, \dots, r_n$:

$$\begin{aligned} \mathbf{P}\left\{\bigcap_{i=1}^n [X_i < r_i] \mid \mathcal{G}\right\}(\omega) &= \prod_{i=1}^n F_i^\omega(r_i); \\ F_i^\omega(\lambda) &\geq F_i^\omega(\lambda'), \quad \lambda > \lambda'; \quad F_i^\omega(\lambda) = \lim_{r \uparrow \lambda} F_i^\omega(r), \quad 1 \leq i \leq n; \\ \lim_{\lambda \rightarrow -\infty} F_i^\omega(\lambda) &= 0, \quad \lim_{\lambda \rightarrow \infty} F_i^\omega(\lambda) = 1, \quad 1 \leq i \leq n. \end{aligned}$$

Define for $\omega \in N^c$, rational r_1, \dots, r_n , and all $n \geq 2$,

$$F_{n,\omega}(r_1, \dots, r_n) = \prod_{i=1}^n F_i^\omega(r_i).$$

Then properties (i)–(v) of Theorem 7.2.2 and Corollary 7.2.1 hold, so that defining $F_{n,\omega} = \prod_{j=1}^n F_{X_j}$ for $\omega \in N$ and continuing as in the proof of Corollary 7.2.1, existence is ensured of a regular conditional distribution \mathbf{P}^ω for X given \mathcal{G} .

Let $\{\xi_n, n \geq 1\}$ be coordinate r.v.s on $(R^\infty, \mathcal{B}^\infty, P^\omega)$. Since for all $n \geq 2$ and real x_1, \dots, x_n

$$P^\omega\{\xi_1 < x_1, \dots, \xi_n < x_n\} = \prod_{i=1}^n F_i^\omega(x_i-) = \prod_{i=1}^n P^\omega\{\xi_i < x_i\}$$

for $\omega \in N^c$, while

$$\begin{aligned} P^\omega\{\xi_1 < x_1, \dots, \xi_n < x_n\} &= F_{n,\omega}(x_1, \dots, x_n) = \prod_{i=1}^n P\{X_i < x_i\} \\ &= \prod_{i=1}^n P^\omega\{\xi_i < x_i\} \end{aligned}$$

for $\omega \in N$, it follows that $\{\xi_n, n \geq 1\}$ are independent r.v.s on $(R^\infty, \mathcal{B}^\infty, P^\omega)$ for each $\omega \in \Omega$.

Apropos of (19), if P_X^ω is a regular conditional distribution for $X = (X_1, X_2, \dots)$ given \mathcal{G} , then since P_X is a product measure, say, $P_1 \times P_2 \times \dots$ on \mathcal{B}^∞ , via Corollary 7.2.2 and Fubini's theorem

$$\begin{aligned} E\{X_1 X_2 \cdots X_n | \mathcal{G}\} &= \int_{R^\infty} x_1 x_2 \cdots x_n dP_X^\omega. \\ &= \prod_{i=1}^n \int_R x_i dP_i(x_i) = \prod_{i=1}^n E\{X_i | \mathcal{G}\}, \quad \text{a.c.} \quad \square \end{aligned}$$

EXAMPLE 1. Let $\{X_n, n \geq 1\}$ be interchangeable r.v.s on (Ω, \mathcal{F}, P) and $\mathcal{G} = \mathcal{F}$ or \mathcal{S} (as in Theorem 2). (i) If $X_1 \in \mathcal{L}_1$, then $1/n \sum_{i=1}^n X_i \xrightarrow{a.e.} E\{X_1 | \mathcal{G}\}$ (ii) If $n \cdot P\{|X_1| > n\} = o(1)$, then $1/n \sum_{i=1}^n X_i - E\{X_1 | \mathcal{G}\} \rightarrow 0$.

PROOF. Let P^ω be a regular conditional distribution for $X = (X_1, X_2, \dots)$ given \mathcal{G} such that the coordinate r.v.s $\{\xi_n, n \geq 1\}$ of the probability space $(R^\infty, \mathcal{B}^\infty, P^\omega)$ are i.i.d. for each $\omega \in \Omega$. If $S_n = \sum_{i=1}^n X_i$ and $E^\omega \xi_1 = \int_{R^\infty} \xi_1 dP^\omega$, then by Corollary 7.2.2 $E\{X_1 | \mathcal{G}\} = E^\omega \xi_1$, a.c. By Corollaries 7.2.2 and 7.3.5 for any $\varepsilon > 0$

$$\begin{aligned} \lim_{m \rightarrow \infty} P\left\{\bigcup_{n=m}^{\infty} \left[\left| \frac{1}{n} S_n - E\{X_1 | \mathcal{G}\} \right| > \varepsilon \right] \right\} &= \lim_{m \rightarrow \infty} P\left\{\bigcup_{n=m}^{\infty} \left[\left| \frac{1}{n} S_n(\omega) - E^\omega \xi_1 \right| > \varepsilon \right] \right\} \\ &= \lim_{m \rightarrow \infty} E P^\omega\left\{\bigcup_{n=m}^{\infty} \left[\left| \frac{1}{n} \sum_{i=1}^n \xi_i - E^\omega \xi_1 \right| > \varepsilon \right] \right\} = 0. \quad \square \end{aligned}$$

Apropos of (ii), set $S'_n = \sum_{j=1}^n X'_j$, $T'_n = \sum_{j=1}^n \xi_j$, where $X'_j = X_j I_{\{X_j \leq n\}}$, $\xi'_j = \xi_j I_{\{\xi_j \leq n\}}$, $1 \leq j \leq n$. Then $P\{S_n \neq S'_n\} \leq \sum_{j=1}^n P\{|X_j| > n\} = o(1)$ and

$$\begin{aligned} E(S'_n - n E\{X'_1 | \mathcal{G}\})^2 &= E E^\omega(T'_n - n E^\omega \xi'_1)^2 = E E^\omega \left[\sum_{j=1}^n (\xi'_j - E^\omega \xi'_1) \right]^2 \\ &= n E E^\omega (\xi'_1 - E^\omega \xi'_1)^2 \leq n E \xi'_1^2 = o(n^2) \end{aligned}$$

according to (17) of Theorem 5.2.4. Thus, $S'_n/n - E\{X'_1 | \mathcal{G}\} \xrightarrow{P} 0$, and so $S_n/n - E\{X_1 | \mathcal{G}\} \xrightarrow{P} 0$. \square

In view of Corollaries 4 and 5 it seems plausible that many results valid for sequences of i.i.d. random variables carry over to sequences of interchangeable r.v.s (see Chapter 9).

Theorem 3. Let $\{X_n, n \geq 1\}$ be conditionally independent given a σ -algebra \mathcal{G} of events and let $\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(X_j, j \geq n)$ denote the tail σ -algebra. Then for any $T \in \mathcal{T}$ there exists some $G \in \mathcal{G}$ with $P\{G \Delta T\} = 0$.

PROOF. Via Corollary 1, $\sigma(X_1, \dots, X_n)$ and $\sigma(X_j, j > n)$ are conditionally independent given \mathcal{G} for any $n \geq 1$. Then $\bigcup_{n=1}^{\infty} \sigma(X_1, \dots, X_n)$ and \mathcal{T} are conditionally independent given \mathcal{G} , whence by Theorem 1(iii), $\sigma(X_1, X_2, \dots)$ and \mathcal{T} are conditionally independent given \mathcal{G} . Hence, for $T \in \mathcal{T}$, $P\{T | \mathcal{G}\} = P^2\{T | \mathcal{G}\}$, a.c., implying $P\{P\{T | \mathcal{G}\} = 0 \text{ or } 1\} = 1$. Thus,

$$P\{T | \mathcal{G}\} = I_G, \quad \text{a.c.},$$

for some $G \in \mathcal{G}$, implying

$$P\{T \cdot G\} = \int_G P\{T | \mathcal{G}\} dP = P\{G\}$$

or $P\{G \cdot T^c\} = 0$. Analogously, $P\{T \cdot G^c\} = 0$, and so $P\{G \Delta T\} = 0$. \square

Theorem 2 states that interchangeable r.v.s $\{X_n, n = 1, 2, \dots\}$ are conditionally i.i.d. given either the tail σ -algebra \mathcal{T} or the σ -algebra \mathcal{S} of permutable events. What is the relationship between \mathcal{S} and \mathcal{T} ? More generally, if $\{X_n, n \geq 1\}$ are conditionally i.i.d. given \mathcal{G} and also given \mathcal{H} , what is the connection between \mathcal{G} and \mathcal{H} ?

Definition. If (Ω, \mathcal{F}, P) is a probability space and \mathcal{G} is a sub- σ -algebra of \mathcal{F} , the **completion** \mathcal{G}^* of \mathcal{G} is a sub- σ -algebra of \mathcal{F} (Exercise 1.5.9) defined by

$$\mathcal{G}^* = \{G \Delta N : G \in \mathcal{G}, N \in \mathcal{F} \text{ with } P\{N\} = 0\}.$$

This definition of completion differs slightly from the customary one which replaces \mathcal{G}^* by $\mathcal{G}_* = \{G \Delta M : G \in \mathcal{G}, M \subset N \in \mathcal{F}, P\{N\} = 0\}$. However, \mathcal{G}_* may not be a sub- σ -algebra of \mathcal{F} , whereas $\mathcal{G}^* \subset \mathcal{F}$.

Lemma 1. Let \mathcal{G} be a σ -algebra of events and \mathcal{G}^* its completion. If X^* is any \mathcal{G}^* -measurable function, there exists a \mathcal{G} -measurable function X such that $X = X^*$, a.c.

PROOF. Let

$$\mathcal{H} = \{X^* : X^* \geq 0, X^* \text{ is } \mathcal{G}^*\text{-measurable, and } X^* = X, \text{ a.c.}, \\ \text{for some } \mathcal{G}\text{-measurable function } X\}.$$

Then \mathcal{H} is a monotone system and $I_G \in \mathcal{H}$ for all $G \in \mathcal{G}^*$. Hence, \mathcal{H} contains all nonnegative \mathcal{G}^* -measurable functions. The general case now follows from $X^* = X^{*+} - X^{*-}$. \square

6. If a r.v. X is independent of the i.i.d. r.v.s $\{X_n, n \geq 1\}$, show that $Y_n = X_n + X$, $n \geq 1$, are interchangeable r.v.s. Express that joint d.f. of (Y_1, \dots, Y_n) in terms of the d.f.s of X and X_1 .
7. For any $n \geq 2$, find n interchangeable r.v.s $\{X_1, \dots, X_n\}$ which cannot be embedded in a collection of $n+1$ interchangeable r.v.s $\{X_1, \dots, X_n, X_{n+1}\}$. Hint: Recall Exercise 6.3.10.
8. If $\{X_n, n \geq 1\}$ are \mathcal{L}_1 interchangeable r.v.s with $\rho = E(X_1 - E X_1)(X_2 - E X_1)$, then $E X_1 X_2 = 0$ iff $E X_1 = 0$ and $\rho = 0$.
9. If $\{X_n, n \geq 1\}$ are \mathcal{L}_2 interchangeable r.v.s and $\{Y_n, n \geq 1\}$ is defined by $X_n = Y_n + E\{X_1 | \mathcal{S}\}$, then $\{Y_n, n \geq 1\}$ constitute uncorrelated interchangeable r.v.s.
10. Prove that if $\{X_n\}$ are \mathcal{L}_1 interchangeable r.v.s then

$$E \left| \frac{1}{n} \sum_{j=1}^n X_j \right| \leq E \left| \frac{1}{n-1} \sum_{j=1}^{n-1} X_j \right|, \quad n \geq 2.$$

11. Prove that if the r.v.'s $\{X_n, n \geq 1\}$ of Corollary 5 are conditionally i.i.d. given \mathcal{G} , then the coordinate r.v.'s $\{\xi_n, n \geq 1\}$ are i.i.d.

7.4 Introduction to Martingales

The indebtedness of probability theory to gambling is seldom as visible as in the concept and development of martingales. The underlying notion is that of a fair game in which regardless of the whims of chance in assigning outcomes to the past and present, one's (conditional) total expected future fortune is precisely one's current aggregate. Analogously, submartingales and supermartingales correspond to favorable and unfavorable games, respectively. Although martingales were first discussed by Lévy, the realization of their potential, the fundamental development of the subject and indeed most of the theorems presented are due to Doob.

Let (Ω, \mathcal{F}, P) be a probability space and N some subset of the integers and $\pm \infty$, that is, $N \subset \{-\infty, \dots, -2, -1, 0, 1, 2, \dots, \infty\}$. A sequence \mathcal{F}_n of sub- σ -algebras of \mathcal{F} with indices in N will be called a stochastic basis if it is increasing, i.e., $\mathcal{F}_m \subset \mathcal{F}_n$ for $m < n$. If $\{\mathcal{F}_n, n \in N\}$ is a stochastic basis and S_n is an \mathcal{F}_n -measurable function for each $n \in N$, then $\{S_n, \mathcal{F}_n, n \in N\}$ is called a **stochastic sequence**. An \mathcal{L}_p **stochastic sequence**, $p > 0$, is one for which $E|S_n|^p < \infty$, $n \in N$, and correspondingly an \mathcal{L}_p **bounded** stochastic sequence is one satisfying $\sup_{n \in N} E|S_n|^p < \infty$.

Definition. A **submartingale** is a stochastic sequence $\{S_n, \mathcal{F}_n, n \in N\}$ with $|E S_n| \leq \infty$, $n \in N$, such that for all $m, n \in N$ with $m < n$

$$E\{S_n | \mathcal{F}_m\} \geq S_m, \quad a.c. \tag{1}$$

If $\{-S_n, \mathcal{F}_n, n \in N\}$ is a submartingale, then $\{S_n, \mathcal{F}_n, n \in N\}$ is called a **supermartingale**. Moreover, if $\{S_n, \mathcal{F}_n, n \in N\}$ is both a submartingale and a supermartingale, it is termed a **martingale**.

Definition. If $\{\mathcal{F}_n, n \in N\}$ is a stochastic basis, a **stopping time** or \mathcal{F}_n -time T is a function from Ω to $N \cup \{\infty\}$ such that $\{T = n\} \in \mathcal{F}_n, n \in N$, and

$$\mathbb{P}\{[T \in N] \cup [T = \infty]\} = 1.$$

When $N = \{1, 2, \dots, \infty\}$, the preceding coincides with the definition of Section 5.3 and T is a **finite stopping time** or **stopping variable** if

$$\mathbb{P}\{T = \infty\} = 0.$$

If $\{S_n, \mathcal{F}_n, n \in N\}$ is a submartingale and there exists a measurable function R with $|\mathbb{E} R| \leq \infty$ (resp. an $\bigcap_{n \in N} \mathcal{F}_n$ -measurable function L with $|\mathbb{E} L| \leq \infty$) such that for each $n \in N$

$$\mathbb{E}\{R | \mathcal{F}_n\} \geq S_n \quad \left(\text{resp. } \mathbb{E}\left\{S_n \middle| \bigcap_{n \in N} \mathcal{F}_n\right\} \geq L\right), \quad \text{a.c.,} \quad (2)$$

then $\{S_n, \mathcal{F}_n, N\}$ is said to be **closed on the right** (resp. left). A **closed submartingale** $\{S_n, \mathcal{F}_n, N\}$ is one which is closed both on the right and left. A submartingale $\{S_n, \mathcal{F}_n, N\}$ is declared to have a **last** (resp., first) **element** if N has a maximum (resp. minimum). Obviously, a submartingale with a last (resp. first) element is closed on the right (resp. left). A martingale $\{S_n, \mathcal{F}_n, N\}$ is said to be **closed on the right** by R (resp. on left by L) if (2) holds with equality.

The submartingale $\{S_n, \mathcal{F}_n, n \geq 1\}$ will be closed iff it is closed on the right since it is automatically closed on the left by S_1 . An analogous statement holds for $\{S_n, \mathcal{F}_n, n \leq -1\}$.

The important special case of a submartingale or martingale $\{S_n, \mathcal{F}_n, N\}$ with $\mathcal{F}_n = \sigma(S_m, m \leq n, m \in N)$, $n \in N$, will be denoted by $\{S_n, N\}$ or $\{S_n, n \in N\}$.

Simple properties of conditional expectations readily entail the following:

- i. The stochastic sequence $\{S_n, \mathcal{F}_n, -\infty < n < \infty\}$ is a submartingale iff for every finite integer n , $|\mathbb{E} S_n| \leq \infty$ and $\mathbb{E}\{S_{n+1} | \mathcal{F}_n\} \geq S_n$, a.c.
- ii. If $\{S_n, \mathcal{F}_n, N\}$ is a submartingale and $\{G_n, \mathcal{G}_n, n \in N\}$ is a stochastic sequence with $\mathcal{G}_n \subset \mathcal{F}_n, n \in N$, then $\{S_n, \mathcal{G}_n, N\}$ is a submartingale.
- iii. If both $\{S_n, \mathcal{F}_n, N\}$ and $\{S'_n, \mathcal{F}_n, N\}$ are submartingales and $\mathbb{E} S_n + \mathbb{E} S'_n$ exists for each $n \in N$, then $\{S_n + S'_n, \mathcal{F}_n, N\}$ is a submartingale.
- iv. An \mathcal{L}_1 stochastic sequence $\{S_n = \sum_1^n X_i, \mathcal{F}_n, -\infty < n < \infty\}$ is a submartingale (resp. martingale) iff for $-\infty < n < \infty$

$$\mathbb{E}\{X_{n+1} | \mathcal{F}_n\} \geq 0 \quad (\text{resp. } \mathbb{E}\{X_{n+1} | \mathcal{F}_n\} = 0), \quad \text{a.c.};$$

the r.v.s $\{X_n\}$ are called **martingale differences** in the latter case.

- v. If $\{S_n = \sum_{i=1}^n X_i, \mathcal{F}_n, n \geq 1\}$ is an \mathcal{L}_2 martingale, then

$$\mathbb{E} S_n^2 = \sum_{j=1}^n \mathbb{E} X_j^2, \quad n \geq 1.$$

The condition of (iv) corroborates the view of S_n as an aggregate of outcomes X_n of favorable or fair games.

If, as in Example 3, $\{S_n, \mathcal{F}_n, n \geq m\}$ is a stochastic sequence such that $\{S_n^*, \mathcal{F}_n^*, n \leq -m\}$ is a martingale, where $S_n^* = S_{-n}$, $\mathcal{F}_n^* = \mathcal{F}_{-n}$, $n \leq -m$, then $\{S_n, \mathcal{F}_n, n \geq m\}$ is sometimes alluded to as a downward (or reverse) martingale, in contradistinction to the more standard upward martingale $\{S_n, \mathcal{F}_n, n \geq m\}$ of Example 1.

The following examples attest to the pervasiveness of martingales.

EXAMPLE 1. Let $\{X_n, n \geq 1\}$ be independent \mathcal{L}_1 random variables with $S_n = \sum_{i=1}^n X_i$. Then $\{S_n, n \geq 1\}$ is a submartingale if $E X_n \geq 0$, $n > 1$, and a martingale if $E X_n = 0$, $n > 1$.

EXAMPLE 2. Let $\{X_n, n \geq 1\}$ be independent \mathcal{L}_1 random variables with $E X_n = 0$, $n \geq 1$. For any integer $k \geq 1$, if

$$S_{k,n} = \sum_{1 \leq i_1 < \dots < i_k \leq n} X_{i_1} X_{i_2} \cdots X_{i_k}, \quad n \geq k,$$

then $\{S_{k,n}, n \geq k\}$ is an \mathcal{L}_1 martingale. The important special case $k = 1$ is subsumed in Example 1. More generally, if $\{X_n, n \geq 1\}$ are \mathcal{L}_k random variables with $E\{X_{n+1} | \mathcal{F}_n\} = 0$, $n \geq 1$, where $\{X_n, \mathcal{F}_n, n \geq 1\}$ is a stochastic sequence, then $\{S_{k,n}, \mathcal{F}_n, n \geq k\}$ is an \mathcal{L}_1 martingale with $E S_{k,k} = 0$.

EXAMPLE 3. Let $\{X_n, n \geq 1\}$ be interchangeable random variables and $\varphi(\cdot)$, a symmetric Borel function on R^m with $E|\varphi(X_1, \dots, X_m)| < \infty$. A sequence of so-called U -statistics is defined for any integer $m \geq 1$ by

$$U_{m,n} = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \dots < i_m \leq n} \varphi(X_{i_1}, \dots, X_{i_m}), \quad n \geq m.$$

If $\mathcal{F}_n = \sigma(U_{m,j}, j \geq n)$, then for $1 \leq i_1 < \dots < i_m \leq n + 1$ and $B \in \mathcal{F}_{n+1}$

$$\int_B \varphi(X_{i_1}, \dots, X_{i_m}) dP = \int_B \varphi(X_1, \dots, X_m) dP$$

via symmetry and interchangeability (Exercise 7.1.12), implying

$$E\{\varphi(X_{i_1}, \dots, X_{i_m}) | \mathcal{F}_{n+1}\} = E\{\varphi(X_1, \dots, X_m) | \mathcal{F}_{n+1}\}, \quad \text{a.c.},$$

for $1 \leq i_1 < \dots < i_m \leq n + 1$, and hence

$$E\{U_{m,n} | \mathcal{F}_{n+1}\} = E\{U_{m,n+1} | \mathcal{F}_{n+1}\} = U_{m,n+1}, \quad \text{a.c.}$$

Consequently, if $U_n^* = U_{m,-n}$ and $\mathcal{F}_n^* = \mathcal{F}_{-n}$, $n \leq -m$, then $\{U_n^*, \mathcal{F}_n^*, n \leq -m\}$ is a martingale closed on the right. The important special cases $m = 1$, $\varphi(x) = x$ and $m = 2$, $\varphi(x_1, x_2) = (x_1 - x_2)^2/2$ yield the arithmetic mean $U_{1,n} = (1/n) \sum_{i=1}^n X_i = \bar{X}_n$ (say) and the sample variance $U_{2,n} = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ respectively.

in n . Since $|S'_n| \leq |S_n| + \sum_{j=2}^n E\{X_j | X_1 \cdots X_{j-1}\}$,

$$E|S'_n| \leq E|S_n| + \sum_{j=2}^n E X_j = E|S_n| + E S_n - E X_1 \leq 3 \sup_{n \geq 1} E|S_n|.$$

Hence, every \mathcal{L}_1 submartingale $\{S_n, n \geq 1\}$ can be expressed as $S_n = S'_n + S''_n$, $n \geq 1$, where $\{S'_n, n \geq 1\}$ is an \mathcal{L}_1 martingale and $0 \leq S''_n \uparrow$ a.c. Moreover, if $\{S_n, n \geq 1\}$ is \mathcal{L}_1 bounded so are $S'_n, n \geq 1$, and $S''_n, n \geq 1$.

The first question to be explored is convergence of submartingales.

Theorem 1. If $\{S_n, \mathcal{F}_n, n \geq 1\}$ is a submartingale such that for every $\{\mathcal{F}_n\}$ -time T

$$\int_{[T < \infty]} S_T dP \neq \infty,$$

then $\lim_{n \rightarrow \infty} S_n$ exists a.c.

PROOF. If $\mu_n\{A\} = \int_A S_n dP$, $A \in \mathcal{F}_n$, then by the submartingale property, $\mu_{n+1}\{A\} \geq \mu_n\{A\}$ for $A \in \mathcal{F}_n$. Suppose that for some pair of real numbers α, β

$$V = \left\{ \overline{\lim}_{n \rightarrow \infty} S_n > \alpha > \beta > \underline{\lim}_{n \rightarrow \infty} S_n \right\}$$

has positive probability, say $P\{V\} > \delta > 0$. It may and will be supposed, replacing S_n by $(\alpha - \beta)^{-1}(S_n - \beta)$, that $\alpha = 1$, $\beta = 0$.

Set $m_0 = 1$, $B_0 = \Omega$, $V_0 = B_0 V$ and $C_n^0 = B_0 \{S_n > 1, S_j \leq 1, \text{ for } m_0 \leq j < n\}$. Define $A_1 = \bigcup_{n=m_0}^{n_1} C_n^0$, where n_1 is large enough to ensure $P\{V_0 - A_1\} < \delta/4$.

Next, define $D_n^0 = A_1 \{S_n < 0, S_j \geq 0, m_1 \leq j < n\}$, $B_1 = \bigcup_{n_1}^{m_1} D_n^0$, where m_1 is large enough to guarantee $P\{V_0 - B_1\} < \delta/4$ and note that

$$\begin{aligned} \mu_{m_1}\{A_1 - B_1\} &\geq \mu_{m_1}\{A_1 - B_1\} + \sum_{j=n_1}^{m_1} \mu_j\{D_j^0\} \\ &= \mu_{m_1}\left\{A_1 - \bigcup_{n_1}^{m_1} D_j^0\right\} + \sum_{j=n_1}^{m_1} \mu_j\{D_j^0\} \\ &= \mu_{m_1}\left\{A_1 - \bigcup_{n_1}^{m_1-1} D_j^0\right\} + \sum_{j=n_1}^{m_1-1} \mu_j\{D_j^0\} \\ &\geq \mu_{m_1-1}\left\{A_1 - \bigcup_{n_1}^{m_1-1} D_j^0\right\} + \sum_{j=n_1}^{m_1-1} \mu_j\{D_j^0\} \geq \dots \geq \\ &\geq \mu_{n_1}\{A_1\} = \sum_{n=m_0}^{n_1} \mu_{n_1}\{C_n^0\} \\ &\geq P\{A_1\} \geq P\{V_0\} - P\{V_0 - A_1\} > 3\delta/4. \end{aligned}$$

Lemma 2. If $\{S_n, \mathcal{F}_n, n \geq 1\}$ is a submartingale with $\sup_{n \geq 1} E S_n^+ = M < \infty$, then for any stopping time T

$$\int_{[T < \infty]} S_T^+ dP \leq M, \quad \int_{[T < \infty]} |S_T| dP \leq 2M - E S_1 \leq 3 \sup_{n \geq 1} E |S_n|. \quad (5)$$

PROOF. Set $T' = \min(T, n), n \geq 1$. Since $\{S_n^+, \mathcal{F}_n, n \geq 1\}$ is a submartingale by Lemma 1,

$$E S_{T'}^+ = \sum_{j=1}^n \int_{[T=j]} S_j^+ + \int_{[T>n]} S_n^+ \leq \sum_{j=1}^n \int_{[T=j]} S_n^+ + \int_{[T>n]} S_n^+ = E S_n^+,$$

so that $E S_{T'}^+ \leq M$, whence the first part of (5) follows from Fatou's lemma. Next,

$$\begin{aligned} E S_1 &= \int_{[T=1]} S_1 + \int_{[T>1]} S_1 \leq \int_{[T=1]} S_1 + \int_{[T>1]} S_2 \\ &\leq \int_{[T=1]} S_1 + \int_{[T=2]} S_2 + \int_{[T>2]} S_2 \leq \dots \\ &\leq \int_{[1 \leq T \leq n]} S_T + \int_{[T>n]} S_n = E S_{T'}, \end{aligned}$$

implying

$$E |S_{T'}| = 2 E S_{T'}^+ - E S_{T'}^- \leq 2M - E S_1,$$

and the remainder of (5) follows once more by Fatou. \square

The special case $X_n \equiv X_0$ of the next corollary, due to Doob (1940), may be considered a milestone in the development of martingales. When X_0 is bounded, this had been obtained by Lévy (1937).

Corollary 1. Let $\{\mathcal{F}_n, n \geq 1\}$ be a stochastic basis and $\{X_n, n \geq 0\}$ a sequence of \mathcal{L}_1 r.v.s with $X_n \xrightarrow{a.s.} X_0$ and $E \sup_{n \geq 1} |X_n| < \infty$. Then, if

$$\mathcal{F}_\infty = \sigma \left(\bigcup_{n=1}^{\infty} \mathcal{F}_n \right),$$

$$E\{X_n | \mathcal{F}_n\} \xrightarrow{a.s.} E\{X_0 | \mathcal{F}_\infty\}.$$

PROOF. (i) It will first be demonstrated for any integrable r.v. X_0 that

$$E\{X_0 | \mathcal{F}_n\} \xrightarrow{a.s.} E\{X_0 | \mathcal{F}_\infty\}.$$

By considering X_0^+ and X_0^- separately, it may and will be supposed that $X_0 \geq 0$. Then, setting $S_n = E\{X_0 | \mathcal{F}_n\}$, $\{S_n, \mathcal{F}_n, n \geq 1\}$ is a nonnegative martingale with $\sup_{n \geq 1} E|S_n| = E S_1 = E X_0 < \infty$. Hence, by Theorem 1 and Lemma 2, $\lim_{n \rightarrow \infty} S_n = S_\infty$ exists, a.c., and, moreover, S_∞ is an integrable r.v. by Fatou's lemma. Since

$$\mathbb{P}\{S_n > C\} \leq C^{-1} \mathbb{E} S_n = C^{-1} \mathbb{E} X_0 \rightarrow 0 \text{ as } C \rightarrow \infty,$$

$$\int_{\{S_n > C\}} S_n = \int_{\{S_n > C\}} \mathbb{E}\{X_0 | \mathcal{F}_n\} = \int_{\{S_n > C\}} X_0 \rightarrow 0$$

as $C \rightarrow \infty$, and so $\{S_n, n \geq 1\}$ is u.i., whence $S_n \xrightarrow{\mathbb{P}} S_\infty$ by Corollary 4.2.4. Thus, for all $n \geq 1$ and $A \in \mathcal{F}_n$, if $m > n$,

$$\int_A X_0 = \int_A S_n = \int_A S_m \xrightarrow{m \rightarrow \infty} \int_A S_\infty,$$

implying $\int_A X_0 = \int_A S_\infty$ for all $A \in \bigcup_1^\infty \mathcal{F}_n$, so that $\mathbb{E}\{X_0 | \mathcal{F}_\infty\} = S_\infty = \lim S_n$, a.c.

Next, set $Y_m = \sup_{n \geq m} |X_n - X_0|$. For every integer m and all $n \geq m$

$$\begin{aligned} D_n &\equiv |\mathbb{E}\{X_n | \mathcal{F}_n\} - \mathbb{E}\{X_0 | \mathcal{F}_\infty\}| \\ &\leq |\mathbb{E}\{(X_n - X_0) | \mathcal{F}_n\}| + |\mathbb{E}\{X_0 | \mathcal{F}_n\} - \mathbb{E}\{X_0 | \mathcal{F}_\infty\}| \\ &\leq \mathbb{E}\{Y_m | \mathcal{F}_n\} + |\mathbb{E}\{X_0 | \mathcal{F}_n\} - \mathbb{E}\{X_0 | \mathcal{F}_\infty\}|, \end{aligned}$$

whence for every integer m

$$\overline{\lim}_{n \rightarrow \infty} D_n \leq \mathbb{E}\{Y_m | \mathcal{F}_\infty\}$$

by the part (i) already proved. Since $Y_m \xrightarrow{\text{a.s.}} 0$ and $|Y_m| \leq 2 \sup |X_n| \in \mathcal{L}_1$, it follows that $\mathbb{E}\{Y_m | \mathcal{F}_\infty\} \xrightarrow{\text{a.s.}} 0$ as $m \rightarrow \infty$ by Theorem 7.1.2, whence $D_n \xrightarrow{\text{a.s.}} 0$. \square

Theorem 2. Let $\{S_n, \mathcal{F}_n, n \geq 1\}$ be a submartingale and $\mathcal{F}_\infty = \sigma(\bigcup_1^\infty \mathcal{F}_n)$.

- i. If $\sup_{n \geq 1} \mathbb{E} S_n^+ < \infty$, then $S_\infty = \lim S_n$ exists a.c. with S_∞ finite a.c. on $\{S_1 > -\infty\}$. Moreover, if $\sup_{n \geq 1} \mathbb{E}|S_n| < \infty$, then $\mathbb{E}|S_\infty| < \infty$.
- ii. If $\{S_n^+, n \geq 1\}$ are u.i., then $S_\infty = \lim S_n$ exists a.c. and $\{S_n, \mathcal{F}_n, 1 \leq n \leq \infty\}$ is a submartingale.
- iii. If $\{S_n, \mathcal{F}_n, n \geq 1\}$ is a submartingale closed by some \mathcal{F} -measurable function S with $\mathbb{E} S^+ < \infty$, then $\{S_n^+, n \geq 1\}$ are u.i., so that

$$S_\infty = \lim_{n \rightarrow \infty} S_n$$

exists a.c. and, moreover, $\{S_n, \mathcal{F}_n, 1 \leq n \leq \infty\}$ is a submartingale closed by S .

- iv. The r.v.s $\{S_n, n \geq 1\}$ are u.i. iff $\{S_n, \mathcal{F}_n, 1 \leq n \leq \infty\}$ is an \mathcal{L}_1 submartingale with $\lim_{n \rightarrow \infty} \mathbb{E} S_n = \mathbb{E} S_\infty$ iff $S_n \xrightarrow{\mathbb{P}} S_\infty$, where $S_\infty = \lim S_n$.

PROOF. (i) By Lemma 2 and Theorem 1, $S_n \xrightarrow{\mathbb{P}} S_\infty$. Moreover, if $\sup_{n \geq 1} \mathbb{E}|S_n| < \infty$, Fatou's lemma guarantees $\mathbb{E}|S_\infty| < \infty$. Next, for any $k > 0$ set $S'_n = S_n I_{\{S_1 > -k\}}$. Then $\{S'_n, \mathcal{F}_n, n \geq 1\}$ is a submartingale with $\sup_{n \geq 1} \mathbb{E} S'_n^+ < \infty$ and $\mathbb{E} S'_1 \geq -k$. Lemma 2 with $T \equiv n$ ensures $\sup_{n \geq 1} \mathbb{E}|S'_n| < \infty$, whence S_∞ is finite a.c. on $\{S_1 > -k\}$. Letting $k \rightarrow \infty$, the remaining portion of (i) obtains.

- iii. If the martingale $\{S_n, \mathcal{F}_n, n \geq 1\}$ is closed by some r.v. $S \in \mathcal{L}_1$, then $\{S_n, n \geq 1\}$ is u.i., so that $S_n \xrightarrow{\text{a.e.}} S_\infty \in \mathcal{L}_1$ and, moreover, $E\{S | \mathcal{F}_n\} = S_n$, a.e., $1 \leq n \leq \infty$.
- iv. The r.v.s $\{S_n, n \geq 1\}$ are u.i. iff $\{S_n, \mathcal{F}_n, 1 \leq n \leq \infty\}$ is an \mathcal{L}_1 martingale with $\lim E S_n = E S_\infty$ iff $S_n \xrightarrow{\mathcal{L}_1} S_\infty$, where $S_\infty = \lim S_n$.

PROOF. Parts (i), (ii), (iii), (iv) follow directly from their counterparts in Theorem 2 (applied in the latter three cases to both S_n and $-S_n$). \square

Corollary 2. If $\{S_n, \mathcal{F}_n, n \geq 1\}$ is a positive (or negative) \mathcal{L}_1 martingale, $S_n \xrightarrow{\text{a.e.}} S_\infty \in \mathcal{L}_1$.

The next theorem illustrates how global martingale convergence in conjunction with stopping times yields local martingale convergence.

Theorem 4. If

$$\left\{ S_n = \sum_{i=1}^n X_i, \mathcal{F}_n, n \geq 1 \right\}$$

is a submartingale with $E \sup_{n \geq 1} X_n^+ < \infty$, then S_n converges a.e. on the set $\{X_1 > -\infty, \sup_{n \geq 1} S_n < \infty\}$.

PROOF. For any $c > 0$, define $T = T_c = \inf\{n \geq 1 : S_n > c\}$. Then T is a stopping time and $\{T_c = \infty\} = \{\sup_{n \geq 1} S_n \leq c\} \rightarrow \{\sup_{n \geq 1} S_n < \infty\}$ as $c \rightarrow \infty$. As seen in Example 5, $\{U_n = \sum_{j=1}^n X_j I_{\{T \geq j\}}, \mathcal{F}_n, n \geq 1\}$ is a submartingale and

$$\begin{aligned} E U_n^+ &= E \left(\sum_{j=1}^n X_j I_{\{T \geq j\}} \right)^+ \leq E \left(\sum_{j=1}^n X_j I_{\{T > j\}} \right)^+ + E \left(\sum_{j=1}^n X_j I_{\{T=j\}} \right)^+ \\ &\leq C + E \sup X_n^+ < \infty. \end{aligned}$$

Hence, by Theorem 2, U_n converges a.e. on $\{X_1 > -\infty\}$. Therefore, $S_n = \sum_{j=1}^n X_j$ converges a.e. on $\{X_1 > -\infty, T_c = \infty\}$. Consequently, letting $c \rightarrow \infty$, S_n converges a.e. on $\{X_1 > -\infty, \sup_{n \geq 1} S_n < \infty\}$. \square

An issue of considerable importance and unquestionable utility in probability theory is the effect on expected values of randomly stopping a stochastic sequence.

Theorem 5. Let $\{S_n, \mathcal{F}_n, n \geq 1\}$ be a submartingale.

- i. If T is a finite $\{\mathcal{F}_n\}$ -time with

$$|E S_T| \leq \infty, \quad \lim_{n \rightarrow \infty} \int_{[T > n]} S_n^+ = 0, \quad (6)$$

then for $n \geq 1$

$$E\{S_T | \mathcal{F}_n\} \geq S_n, \text{ a.e. on } \{T \geq n\} \quad \text{and} \quad E S_T \geq E S_1. \quad (7)$$

and $T = -1$ if no such j exists, then T is a stopping rule or finite $\{\mathcal{F}_j\}$ -time and, moreover, a bounded r.v. Since by definition $Y_T = 1$ on $\bigcup_1^n \{S_j \geq j\}$ and zero elsewhere, Corollary 3 implies that on the set $\{S_n < n\}$

$$\mathbb{P}\left\{\bigcup_1^n [S_j \geq j] \mid S_n\right\} = \mathbb{E}\{Y_T \mid \mathcal{F}_{n-1}\} = Y_{n-1} = \frac{S_n}{n},$$

which is tantamount to the proposition. \square

Setting $X_j = 2$ or 0 according as the j th vote goes to the loser or his rival, note that if r_j (resp. s_j) of the first j votes are cast for the loser (resp. rival), then given that $S_n = 2r$ the event $S_j < j$, $1 \leq j \leq n = r + s$, is tantamount to $r_j < s_j$, $1 \leq j \leq n$.

Theorem 6. Let $\{S_n = \sum_1^n X_j, \mathcal{F}_n, n \geq 1\}$ be a submartingale with $\mathbb{E} X_n^+ < \infty$, $n \geq 1$ and let T be a finite $\{\mathcal{F}_n\}$ -time and $\mathcal{F}_0 = (\emptyset, \Omega)$. If (i)

$$\mathbb{E} \sum_{n=1}^T \mathbb{E}\{X_n^+ \mid \mathcal{F}_{n-1}\} < \infty \quad (14)$$

or (ii) $\{S_n^+, n \geq 1\}$ are uniformly integrable, then for $n \geq 1$

$$\mathbb{E}\{S_T \mid \mathcal{F}_n\} \geq S_n, \text{ a.c. on } \{T \geq n\}, \text{ and } \mathbb{E} S_T \geq \mathbb{E} S_1. \quad (15)$$

PROOF. Under (i)

$$\begin{aligned} \mathbb{E} \sum_1^T X_n^+ &= \mathbb{E} \sum_1^\infty X_n^+ I_{\{T \geq n\}} = \sum_1^\infty \mathbb{E}[I_{\{T \geq n\}} \mathbb{E}\{X_n^+ \mid \mathcal{F}_{n-1}\}] \\ &= \mathbb{E} \sum_1^\infty I_{\{T \geq n\}} \mathbb{E}\{X_n^+ \mid \mathcal{F}_{n-1}\} = \mathbb{E} \sum_1^T \mathbb{E}\{X_n^+ \mid \mathcal{F}_{n-1}\} < \infty. \end{aligned}$$

Hence, $\mathbb{E} S_T^+ \leq \mathbb{E} \sum_1^T X_n^+ < \infty$ and, moreover,

$$\int_{\{T > n\}} S_n^+ \leq \int_{\{T > n\}} \sum_{j=1}^n X_j^+ \leq \int_{\{T > n\}} \sum_{j=1}^T X_j^+ = o(1)$$

as $n \rightarrow \infty$. Thus, (6) and consequently the conclusion (7) of Theorem 5 hold.

Under (ii), $\sup_{n \geq 1} \mathbb{E} S_n^+ < \infty$ whence $\mathbb{E} S_T^+ < \infty$ by Lemma 2. Since $\mathbb{P}\{T > n\} = o(1)$, the remainder of (6) and consequently (15) follow from u.i. \square

Corollary 5. If $\{S_n = \sum_1^n X_i, \mathcal{F}_n, n \geq 1\}$ is an \mathcal{L}_1 martingale and T is a finite $\{\mathcal{F}_n\}$ -time satisfying

$$\mathbb{E} \sum_{n=1}^T \mathbb{E}\{|X_n| \mid \mathcal{F}_{n-1}\} < \infty \quad (16)$$

(in particular, if T is a bounded r.v.) or if $\{S_n, \mathcal{F}_n, n \geq 1\}$ is u.i., then for any

$n \geq 1$

$$\mathbb{E}\{S_T | \mathcal{F}_n\} = S_n, \text{ a.e. on } \{T \geq n\}, \text{ and } \mathbb{E} S_T = \mathbb{E} S_1. \quad (17)$$

Corollary 6. If $\{S_n = \sum_{j=1}^n X_j, \mathcal{F}_n, n \geq 1\}$ is an \mathcal{L}_1 martingale and T is a finite $\{\mathcal{F}_n\}$ -time satisfying (12) or (16), then for any $r \geq 1$

$$\mathbb{E}\{|S_T|^r | \mathcal{F}_n\} \geq |S_n|^r, \text{ a.e. on } \{T \geq n\}, \text{ and } \mathbb{E}|S_T|^r \geq \mathbb{E}|S_1|^r. \quad (18)$$

PROOF. By Lemma 1, $\{|S_n|, \mathcal{F}_n, n \geq 1\}$ is a submartingale, and so, by Theorem 6

$$(|S_n|I_{[T \geq n]})^r \leq \mathbb{E}'\{|S_T|I_{[T \geq n]} | \mathcal{F}_n\} \leq \mathbb{E}\{|S_T|^r I_{[T \geq n]} | \mathcal{F}_n\},$$

recalling the conditional Jensen inequality. \square

Next, martingale generalizations of Theorem 5.3.1 (Wald's equation) and Theorem 5.3.3 will be obtained. Let $\{S_n = \sum_{j=1}^n X_j, \mathcal{F}_n, n \geq 1\}$ be an \mathcal{L}_2 martingale. As noted in Example 2, $\{S_n^2 - \sum_{j=1}^n X_j^2 = 2 \sum_{1 \leq i < j \leq n} X_i X_j, \mathcal{F}_n, n \geq 2\}$ is an \mathcal{L}_1 martingale with expectation zero, whence for any stopping variable T , setting $T(n) = \min(T, n)$,

$$\mathbb{E} S_{T(n)}^2 = \mathbb{E} \sum_1^{T(n)} X_j^2 \quad (19)$$

by Corollary 4. Since $T(n) \uparrow T$, it therefore follows from Fatou's lemma that

$$\mathbb{E} S_T^2 \leq \mathbb{E} \sum_1^T X_j^2. \quad (20)$$

Thus, if $\mathbb{E} S_T^2 = \infty$, equality holds in (20). In the contrary case $\mathbb{E} S_T^2 < \infty$, in view of (20), equality will hold therein recalling (19) if,

$$\mathbb{E} S_T^2 \geq \mathbb{E} S_{T(n)}^2 \quad (21)$$

Hence, it suffices to verify (21) to establish equality in (20).

Lemma 3. If $\{S_n, \mathcal{F}_n, n \geq 1\}$ is an \mathcal{L}_2 martingale and T is any finite $\{\mathcal{F}_n\}$ -time, then (20) obtains. Moreover, if

$$\lim_{n \rightarrow \infty} \int_{[T > n]} S_n^2 < \infty \quad (22)$$

or

$$\lim_{n \rightarrow \infty} \int_{[T > n]} |S_n| = 0 \quad (23)$$

holds, then $\mathbb{E} S_T^2 = \mathbb{E} \sum_1^T X_j^2$.

PROOF. In view of the prior discussion, it suffices to show that (22) \Rightarrow (23)

\Rightarrow (21). Suppose that $\underline{\lim} \int_{[T>n]} |S_n| = c > 0$. For any k in $(0, \infty)$

$$\underline{\lim}_{n \rightarrow \infty} \int_{[T>n]} S_n^2 \geq k \underline{\lim}_{n \rightarrow \infty} \int_{[T>n, |S_n|>k]} |S_n| = k \cdot c \rightarrow \infty$$

as $k \rightarrow \infty$, and so (22) \Rightarrow (23). Next, supposing, as is permissible, that $E S_T^2 < \infty$, necessarily $E|S_T| < \infty$, and so via Theorem 5, (23) implies that on $\{T \geq n\}$

$$E\{|S_T| \mid \mathcal{F}_n\} \geq |S_n|,$$

whence on $\{T \geq n\}$

$$E\{S_T^2 \mid \mathcal{F}_n\} \geq S_n^2 = S_{T(n)}^2.$$

Since $E\{S_T^2 \mid \mathcal{F}_n\} = E\{S_{T(n)}^2 \mid \mathcal{F}_n\} = S_{T(n)}^2$ on $\{T < n\}$, (21) follows. \square

Theorem 7. If $\{S_n = \sum_1^n X_j, \mathcal{F}_n, n \geq 1\}$ is an \mathcal{L}_2 martingale and T is any finite $\{\mathcal{F}_n\}$ -time, then $E S_T^2 \leq E \sum_1^T X_j^2$. Moreover, if any one of the four conditions

$$\begin{aligned} \underline{\lim}_{n \rightarrow \infty} \int_{[T>n]} |S_n| &= 0, & \underline{\lim}_{n \rightarrow \infty} \int_{[T>n]} S_n^2 &< \infty, \\ E \sum_{n=1}^T X_n^2 &< \infty, & E \sum_{n=1}^T |X_n| &< \infty, \end{aligned} \tag{24}$$

holds, then, setting $\mathcal{F}_0 = \{\emptyset, \Omega\}$,

$$E S_T^2 = E \sum_{j=1}^T X_j^2 = E \sum_{j=1}^T E\{X_j^2 \mid \mathcal{F}_{j-1}\}. \tag{25}$$

Either of the last two conditions of (24) implies $E S_T = E X_1$. If $\{S_n = \sum_1^n X_i, \mathcal{F}_n, n \geq 1\}$ is merely an \mathcal{L}_1 martingale, the last condition of (24) entails $E S_T = E X_1$.

PROOF. First, $E \sum_1^T X_j^2 < \infty \Rightarrow$ (22) since, recalling (19),

$$\int_{[T>n]} S_n^2 \leq E S_{T \wedge n}^2 = E \sum_1^{T \wedge n} X_j^2 \leq E \sum_1^T X_j^2. \tag{26}$$

and so in view of Lemma 3 and Corollary 4 it suffices to note that the final condition of (24) ensures that

$$E|S_T| \leq E \sum_1^T |X_n| < \infty$$

and

$$\mathbb{E}|S_n|I_{\{T>n\}} \leq \mathbb{E} \sum_1^n |X_j|I_{\{T>n\}} \leq \mathbb{E} \sum_1^T |X_j|I_{\{T>n\}} = o(1). \quad \square$$

Corollary 7. If $\{X_n, n \geq 1\}$ are independent r.v.s with $\mathbb{E} X_n = 0, n \geq 1$, and T is a finite $\{\mathcal{F}_n\}$ -time, then, setting $S_n = \sum_{i=1}^n X_i$,

$$\mathbb{E} \sum_1^T |X_j| < \infty \quad (27)$$

implies $\mathbb{E} S_T = 0$. If $\sigma_n^2 = \mathbb{E} X_n^2 < \infty, n \geq 1$, then either (27) or

$$\mathbb{E} \sum_1^T \sigma_j^2 < \infty \quad (28)$$

implies

$$\mathbb{E} S_T^2 = \mathbb{E} \sum_1^T X_j^2 = \mathbb{E} \sum_1^T \sigma_j^2. \quad (29)$$

PROOF. This follows directly from Theorem 7. \square

A useful device in obtaining Doob's maximal inequality (33) and a martingale generalization (36) of Kolmogorov's inequality is

Lemma 4. Let $\{S_n = \sum_{j=1}^n X_j, \mathcal{F}_n, n \geq 1\}$ be an \mathcal{L}_1 stochastic sequence and $\{v_n, \mathcal{F}_{n-1}, n \geq 1\}$ a stochastic sequence with $v_n \in \mathcal{L}_\infty, n \geq 1$. Then for any bounded $\{\mathcal{F}_n\}$ -time T ,

$$\mathbb{E} v_T S_T = \mathbb{E} \sum_1^T [v_j \mathbb{E}\{X_j | \mathcal{F}_{j-1}\} + (v_j - v_{j-1})S_{j-1}], \quad S_0 = 0, \mathcal{F}_0 = \mathcal{F}_1, \quad (30)$$

and, moreover, if $(v_{j+1} - v_j)S_j \leq 0$, a.c., $j \geq 0$,

$$\mathbb{E} v_T S_T \leq \mathbb{E} \sum_1^T v_j \mathbb{E}\{X_j | \mathcal{F}_{j-1}\}. \quad (31)$$

PROOF. If $U_n = v_n S_n - \sum_1^n [v_j \mathbb{E}\{X_j | \mathcal{F}_{j-1}\} + (v_j - v_{j-1})S_{j-1}]$, then $\{U_n, \mathcal{F}_n, n \geq 1\}$ is a martingale and (30) follows from Corollary 4 or 5. Then (31) follows directly from (30). \square

Corollary 8 (Dubins–Freedman). If $\{S_n = \sum_1^n X_j, \mathcal{F}_n, n \geq 1\}$ is an \mathcal{L}_2 martingale with $\mathbb{E} X_1 = 0$ and $Y_n = \mathbb{E}\{X_n^2 | \mathcal{F}_{n-1}\}, n \geq 1$, where $\mathcal{F}_0 = \{\emptyset, \Omega\}$, then for any stopping time T and real numbers a, b with $b > 0$

$$\int_{\{T < \infty\}} \left(\frac{a + S_T}{b + Y_1 + \dots + Y_T} \right)^2 \leq \frac{a^2 + Y_1}{(b + Y_1)^2} + \frac{1}{b + Y_1}. \quad (32)$$

PROOF. If $v_n^{-1} = (b + Y_1 + \dots + Y_n)^2$, then $\{v_n, \mathcal{F}_{n-1}, n \geq 1\}$ is a stochastic sequence with $v_{n+1} \leq v_n$. Since $\{(a + S_n)^2, \mathcal{F}_n, n \geq 1\}$ is a nonnegative submartingale, by (31) (with $S_0 = -a$)

$$\begin{aligned} \int_{[T \leq n]} \left(\frac{a + S_T}{b + Y_1 + \dots + Y_T} \right)^2 &\leq \sum_{j=1}^n \mathbb{E} v_j \mathbb{E} \{(a + S_j)^2 - (a + S_{j-1})^2 | \mathcal{F}_{j-1}\} \\ &= v_1(a^2 + Y_1) + \mathbb{E} \sum_{j=2}^n v_j Y_j. \end{aligned}$$

Since

$$v_j Y_j = \frac{Y_j}{(b + Y_1 + \dots + Y_j)^2} \leq \frac{1}{b + Y_1 + \dots + Y_{j-1}} - \frac{1}{b + Y_1 + \dots + Y_j},$$

the conclusion (32) follows as $n \rightarrow \infty$. \square

Theorem 8. If $\{S_n = \sum_{j=1}^n X_j, \mathcal{F}_n, n \geq 1\}$ is a nonnegative \mathcal{L}_1 submartingale and $\{v_n, \mathcal{F}_{n-1}, n \geq 1\}$ a stochastic sequence of \mathcal{L}_∞ r.v.s with $v_n \geq v_{n+1} \geq 0$, a.c., then for any $\lambda > 0$, (i)

$$\lambda \mathbb{P} \left\{ \max_{1 \leq j \leq n} v_j S_j \geq \lambda \right\} + \int_{[\max v_j S_j < \lambda]} v_n S_n \leq \sum_{j=1}^n \mathbb{E} v_j X_j \quad (33)$$

and (ii) (Doob Inequalities)

$$\mathbb{P} \left\{ \max_{1 \leq j \leq n} S_j \geq \lambda \right\} \leq \frac{1}{\lambda} \int_{[\max S_j \geq \lambda]} S_n, \quad (34)$$

whence

$$\begin{aligned} \|S_n\|_p &\leq \left\| \max_{1 \leq j \leq n} S_j \right\|_p \leq \frac{p}{p-1} \|S_n\|_p, \quad p > 1, \\ \left\| \max_{1 \leq j \leq n} S_j \right\|_p &< \frac{e}{e-1} (1 + \|S_n \log^+ S_n\|_p), \quad p = 1. \end{aligned} \quad (35)$$

(iii) (Hájek–Rényi Inequality) If $\{U_n = \sum_{j=1}^n u_j, \mathcal{F}_n, n \geq 1\}$ is an \mathcal{L}_2 martingale and $\{b_n, n \geq 1\}$ is a positive, nondecreasing real sequence, then for any $\lambda > 0$

$$\mathbb{P} \left\{ \max_{1 \leq j \leq n} \left| \frac{U_j}{b_j} \right| \geq \lambda \right\} \leq \frac{1}{\lambda^2} \sum_{j=1}^n \frac{\mathbb{E} u_j^2}{b_j^2}. \quad (36)$$

PROOF. Let $T_i = \inf\{n \geq 1 : v_n S_n \geq \lambda^i\}$ and $T'_i = \min(T_i, n), n \geq 1, i = 1, 2$. From (31) of Lemma 4

$$\begin{aligned} \lambda^i \mathbb{P} \left\{ \max_{1 \leq j \leq n} v_j S_j \geq \lambda^i \right\} + \int_{[\max v_j S_j < \lambda^i]} v_n S_n \\ \leq \int_{[T_i \leq n]} v_{T_i} S_{T_i} + \int_{[T_i > n]} v_n S_n = \mathbb{E} v_{T_i} S_{T_i} \\ \leq \mathbb{E} \sum_1^{T_i} v_j \mathbb{E} \{X_j | \mathcal{F}_{j-1}\} \leq \sum_{j=1}^n \mathbb{E} v_j X_j, \end{aligned} \quad (37)$$

yielding (33) for $i = 1$. To obtain (36), take $i = 2$, $v_j = b_j^{-2}$ in (37), noting that $S_n = U_n^2$, $n \geq 1$, is a nonnegative \mathcal{L}_1 submartingale. Since (34) is an immediate consequence of (33), it remains to prove (35). Setting $S_n^* = \max_{1 \leq j \leq n} S_j$, if $p > 1$ it follows from Corollary 6.2.2 and (34) that

$$\begin{aligned}\mathbb{E} S_n^{*p} &= p \int_0^\infty \lambda^{p-1} \mathbb{P}\{S_n^* \geq \lambda\} d\lambda \leq p \int_0^\infty \lambda^{p-2} \int_{[S_n^* \geq \lambda]} S_n d\mathbb{P} d\lambda \\ &= p \mathbb{E} S_n \int_0^{S_n^*} \lambda^{p-2} d\lambda = \frac{p}{p-1} \mathbb{E} S_n (S_n^*)^{(p-1)}.\end{aligned}$$

Hence, if $(p-1)q = p$, by Hölder

$$\mathbb{E} S_n^{*p} \leq \frac{p}{p-1} \|S_n\|_p \|S_n^{*(p-1)}\|_q = \frac{p}{p-1} \|S_n\|_p \mathbb{E}^{1/q} S_n^{*p},$$

yielding the first part of (35). If, rather, $p = 1$, again via (34)

$$\begin{aligned}\mathbb{E} S_n^* - 1 &\leq \mathbb{E}(S_n^* - 1)^+ = \int_0^\infty \mathbb{P}\{S_n^* - 1 \geq \lambda\} d\lambda \\ &\leq \int_0^\infty \frac{1}{\lambda + 1} \int_{[S_n^* \geq \lambda + 1]} S_n d\mathbb{P} d\lambda \\ &= \mathbb{E} S_n \int_0^{(S_n^*-1)^+} \frac{d\lambda}{\lambda + 1} = \mathbb{E} S_n \log^+ S_n^*.\end{aligned}$$

Since for constants $a \geq 0$, $b > 0$ necessarily $a \log b \leq a \log^+ a + be^{-1}$,

$$\mathbb{E} S_n^* - 1 \leq \mathbb{E} S_n \log^+ S_n + e^{-1} \mathbb{E} S_n^*,$$

from which the second portion of (35) is immediate. \square

EXAMPLE 9. If $\{S_n, \mathcal{F}_n, n \geq 1\}$ is a submartingale and h is any nonnegative, increasing, convex function, then for any positive t and real x

$$\mathbb{P}\left\{\max_{1 \leq j \leq n} S_j \geq x\right\} \leq \frac{\mathbb{E} h(tS_n)}{h(tx)} \quad (38)$$

and, in particular,

$$\mathbb{P}\left\{\max_{1 \leq j \leq n} S_j \geq x\right\} \leq e^{-tx} \mathbb{E} e^{tS_n}, \quad t > 0. \quad (39)$$

PROOF. Since $\{h(tS_j), \mathcal{F}_j, 1 \leq j \leq n\}$ is a nonnegative submartingale via Lemma 1, (34) ensures that

$$\begin{aligned}\mathbb{P}\left\{\max_{1 \leq j \leq n} S_j \geq x\right\} &\leq \mathbb{P}\left\{h\left(\max_{1 \leq j \leq n} tS_j\right) \geq h(tx)\right\} = \mathbb{P}\left\{\max_{1 \leq j \leq n} h(tS_j) \geq h(tx)\right\} \\ &\leq \frac{\mathbb{E} h(tS_n)}{h(tx)}. \quad \square\end{aligned}$$

The next example generalizes Example 5.2.1.

EXAMPLE 10. If $S_n = \sum_{j=1}^n X_j$, where $\{X_n, n \geq 1\}$ are i.i.d. \mathcal{L}_p r.v.'s for some p in $(0, 2)$ with $E X_1 = 0$ whenever $1 \leq p < 2$, then

$$\sum_{j=2}^{\infty} j^{-2/p} X_j S_{j-1} \text{ converges a.c.} \quad (40)$$

PROOF. If $Y_n = n^{-1/p} X_n I_{\{|X_n| \leq n^{1/p}\}}$, $n \geq 1$, and $T_n = n^{-1/p} S_{n-1} I_{\{|S_{n-1}| \leq n^{1/p}\}}$, $n \geq 2$, then $\{T_n(Y_n - E Y_n), n \geq 1\}$ is a martingale difference sequence with

$$\sum_{n=2}^{\infty} E T_n^2 (Y_n - E Y_n)^2 \leq \sum_{n=2}^{\infty} E(Y_n - E Y_n)^2 \leq \sum_{n=2}^{\infty} E Y_n^2 < \infty$$

by (13) of Theorem 5.1.3. Then $\sum_{n=2}^{\infty} T_n(Y_n - E Y_n)$ converges a.c. via Exercise 7.4.1. Moreover, $\sum_{n=2}^{\infty} E|T_n E Y_n| \leq \sum_{n=2}^{\infty} |E Y_n| < \infty$, and $\{Y_n\}$ and $\{n^{-1/p} X_n\}$ are equivalent sequences according to Theorem 5.1.3. Hence, $\sum_{n=2}^{\infty} T_n E Y_n$ converges a.c., and so $\sum_{n=2}^{\infty} T_n Y_n$ converges a.c.

By the Marcinkiewicz-Zygmund strong law of large numbers, $P\{|S_{n-1}| > n^{1/p}, \text{i.o.}\} = 0$ whence $\{T_n Y_n, n \geq 1\}$ and $\{S_{n-1}/n^{1/p} \cdot X_n/n^{1/p}, n \geq 1\}$ are equivalent sequences and (40) follows. \square

EXERCISES 7.4

- If $\{S_n, n \geq 1\}$ is a martingale satisfying (i) $\sum_{j=1}^{\infty} E X_j^2 < \infty$ or (ii) $\sum_{j=1}^{\infty} E(|X_j| I_{\{|X_j|>1\}} + X_j^2 I_{\{|X_j|\leq 1\}}) < \infty$ or (iii) $\sum_{j=1}^{\infty} E|X_j|^p < \infty$ for some p in $[1, 2]$, then S_n converges a.c. *Hint:* For (ii) consider $X'_j = X_j I_{\{|X_j|\leq 1\}} - E\{X_j I_{\{|X_j|\leq 1\}}|X_1, \dots, X_{j-1}\}$ and $X''_j = X_j - X'_j$.
- In statistical problems, likelihood ratios $U_n = g_n(X_1, \dots, X_n)/f_n(X_1, \dots, X_n)$ are encountered, where f_n, g_n are densities, each being a candidate for the actual density of r.v.s X_1, \dots, X_n . If $\{X_n, n \geq 1\}$ are coordinate r.v.s on $(R^{\infty}, \mathcal{B}^{\infty}, P)$ and g_n vanishes whenever f_n does, show that $\{U_n, n \geq 1\}$ is a martingale when f_n is the true density.
- There exist martingales $\{S_n, \mathcal{F}_n, n \geq 1\}$ and stopping variables T for which (13) fails. Let $\{X_n, n \geq 1\}$ be i.i.d. with $E X_1 = 0$ and set $T = \inf\{n \geq 1 : S_n = \sum_1^n X_i > 0\}$. Then for $n = 1$, $E\{S_T - S_n | \mathcal{F}_n\} > 0$ on $[T > 1]$.
- If $\{S_n, \mathcal{F}_n, n \geq 1\}$ is a martingale or positive submartingale, then for any stopping time T , $E|S_T| \leq \lim_{n \rightarrow \infty} E|S_n|$. In particular, if $\{S_n, \mathcal{F}_n, n \geq 1\}$ is \mathcal{L}_1 bounded, $E|S_T| < \infty$ for every stopping time T .
- If $E|S_T| < \infty$, it is spurious generality to have $\lim_{n \rightarrow \infty}$ rather than $\lim_{n \rightarrow \infty}$ in (6) or (12). *Hint:* If $V_n = S_n^+$ in the first case and $|S_n|$ in the second, then, as in the proof of Theorem 5,

$$Z \sum_{A(T \geq n)} V_n \leq \int_{A(n \leq T \leq m)} V_T + \int_{A(T > m)} V_m.$$

- (i) Find a positive \mathcal{L}_1 martingale which is not u.i. (ii) If $Y_n, n \geq 1$, are r.v.s on (Ω, \mathcal{F}, P) and $A \in \sigma(Y_1, Y_2, \dots)$, then $P\{A | Y_1, \dots, Y_n\} \xrightarrow{a.c.} I_A$. *Hint:* Apply Corollary 1. (iii) If $\{\mathcal{F}_n, n \geq 1\}$ is a stochastic basis and $\{X_n, n \geq 1\}$ are r.v.'s with

$E \sup_{n \geq 1} |X_n| < \infty$, then

$$E\{\liminf_{n \rightarrow \infty} X_n | \mathcal{F}_\infty\} \leq \liminf_{n \rightarrow \infty} E\{X_n | \mathcal{F}_\infty\}.$$

7. If $\{A_n, n \geq 1\}$ is a sequence of events and $A_n \uparrow \mathcal{F}_\infty$, prove the following version of the Borel–Cantelli theorem:

$$P\left\{A_n, \text{i.o. } \Delta \left[\sum_{n=1}^{\infty} P(A_{n+1} | \mathcal{F}_n) = \infty \right] \right\} = 0.$$

8. If $\{S_n = \sum_{i=1}^n X_i, n \geq 1\}$ is an \mathcal{L}_p martingale for some p in $[1, 2]$ and $0 < b_n \uparrow \infty$ then $\lim S_n/b_n = 0$, a.c. on the set A where $\sum_{n=2}^{\infty} b_n^{-p} E\{|X_n|^p | \mathcal{F}_{n-1}\} < \infty$. In particular, $S_n/b_n \xrightarrow{a.s.} 0$ if $P\{A\} = 1$. Hint: Set $T_n = T \wedge n$ where $T = \inf\{n \geq 1 : \sum_{j=1}^{n+1} b_j^{-p} E\{|X_j|^p | \mathcal{F}_{j-1}\} > K > 0$ and apply Exercise 1 to $\sum_1^{T_n} X_j/b_j$.
9. Let Y be an \mathcal{L}_1 r.v. and $\{\mathcal{G}_n, -\infty < n < \infty\}$ a stochastic basis. If $\mathcal{G}_\infty = \sigma(\bigcup_{n=-\infty}^{\infty} \mathcal{G}_n)$, $\mathcal{G}_{-\infty} = \bigcap_{n=-\infty}^{\infty} \mathcal{G}_n$, and $U_n = E\{Y | \mathcal{G}_n\}, -\infty \leq n \leq \infty$, then $(U_n, \mathcal{G}_n, -\infty \leq n \leq \infty)$ is a martingale.
10. Let $\{S_n = \sum_{i=1}^n X_i, \mathcal{F}_n, n \geq 1\}$ be a martingale with bounded conditional variance, i.e., $E\{S_n - E\{S_n | \mathcal{F}_{n-1}\}\}^2 | \mathcal{F}_{n-1} = E\{X_n^2 | \mathcal{F}_{n-1}\} \leq \sigma_n^2, n \geq 1$, where σ_n^2 is a finite constant. If $U_{k,n}$ is as defined in Example 2, show that $E(U_{k,n} - U_{k,n-1})^2 \leq \sigma_n^2 E U_{k-1,n-1}^2$, and hence that $E U_{k,n}^2 \leq \sum_{j=k}^{n-1} \sigma_{j+1}^2 + \sum_{i=k}^j \sigma_i^2$. If $\sigma_n^2 = \sigma^2$, $E U_{k,n}^2 \leq (2k)\sigma^{2k}$ and $E(U_{k,n} - U_{k,n-1})^2 \leq n^{k-1}\sigma^{2k}, k \geq 1$.
11. Show, if $\{S_n, \mathcal{F}_n, n \geq 1\}$ is a nonnegative martingale with $E S_1 = 1$, that $P\{S_n \geq \lambda, \text{some } n \geq 1\} \leq 1/\lambda$.
12. Prove that an \mathcal{L}_1 stochastic sequence $(\sum_{i=1}^n X_i, \sigma(X_1, \dots, X_n), n \geq 1)$ is a martingale iff $E X_{n+1} \phi(X_1, \dots, X_n) = 0$ for all $n \geq 1$ and all bounded Borel functions ϕ , whereas the \mathcal{L}_1 r.v.s $\{X_n, n \geq 1\}$ are independent iff for all $n \geq 1$

$$E \psi(X_{n+1}) \phi(X_1, \dots, X_n) = 0$$

for all bounded Borel functions ϕ, ψ with $E \psi(X_{n+1}) = 0$.

13. (Harris) A branching process is a sequence $\{Z_n, n \geq 0\}$ of nonnegative integer-valued r.v.s with $Z_0 = 1$ and such that the conditional distribution of Z_{n+1} given (Z_0, \dots, Z_n) is that of a sum of Z_n i.i.d. r.v.s each with the distribution of Z_1 . If $E Z_1 = m \in (0, \infty)$, verify that $\{W_n = Z_n/m^n, n \geq 1\}$ is a convergent martingale.

14. (Breiman) Let A, B be linear Borel sets and $\{X_n, n \geq 1\}$ r.v.s. If

$$P\left\{\bigcup_{j=n}^{\infty} [X_j \in B] \mid X_1, \dots, X_n\right\} \geq \delta I_{\{X_n \in A\}}, \delta > 0,$$

prove that $P\{X_n \in A, \text{i.o.}\} \leq P\{X_n \in B, \text{i.o.}\}$. Hint: If $F_N = \bigcup_{j=N}^{\infty} [X_j \in B]$, $\lim_{n \rightarrow \infty} P\{F_N | X_1, \dots, X_n\} = I_{F_N}$.

15. (Doob) If $\{X_n, n \geq 1\}$ are independent r.v.s with $E X_n = 0, n \geq 1$, and $S_n = \sum_{i=1}^n X_i$, $S_n^* = \max_{1 \leq i \leq n} |S_i|$, then $E S_n^* \leq 8 E|S_n|$, thereby improving (33) for $p = 1$. Hint: Via Ottaviani's inequality (Exercise 3.3.16)

$$\int_0^{\infty} P\{S_n^* > 2y\} dy \leq 2 E|S_n| + \int_{2E|S_n|}^{\infty} 2 P\{|S_n| > y\} dy.$$

and, for $r \geq 1$,

$$f_1(X_r) = E\{f(X_r, X_{r+1}, \dots, X_{r+k-1})|X_r\} - Ef$$

are i.i.d. random variables.

Lemma 2. *The functions $f_j = f_j(X_1, \dots, X_j)$ defined in (2) are completely degenerate for $2 \leq j \leq k$ and*

$$\begin{aligned} f(X_1, \dots, X_k) &= Ef + \sum_{i=1}^k f_1(X_i) \\ &\quad + \sum_{1 \leq i_1 < i_2 \leq k} f_2(X_{i_1} X_{i_2}) + \dots \\ &\quad + \sum_{1 \leq i_1 < \dots < i_{k-1} \leq k} f_{k-1}(X_{i_1}, \dots, X_{i_{k-1}}) \\ &\quad + f_k(X_1, \dots, X_k). \end{aligned}$$

PROOF. Since $Q_j(I - Q_j)f_j^* = 0$ and the operators commute,

$$E\{f_j|X_1, \dots, X_{j-1}\} = Q_j \cdots Q_k f_j = Q_j \cdots Q_k \prod_{l=1}^j (I - Q_l) f_j^* = 0, \quad \text{a.c.}$$

whence $f_j(X_1, \dots, X_j)$ is completely degenerate, $2 \leq j \leq k$. The final statement follows from the representation

$$f = \prod_{i=1}^k [(I - Q_i) + Q_i] f,$$

noting via Lemma 1 that

$$Q_1 \cdots Q_k f = E\{f|X_\alpha, \alpha \neq 1, \dots, k\} = Ef. \quad \square$$

Lemma 3. *Let $f \in R^k$, $k \geq 2$, with $|Ef| \leq \infty$ and let $1 \leq i_1 < \dots < i_k \leq n$ and $1 \leq \beta_1 < \dots < \beta_m \leq n$ be two sets of integers. (i) If $\{i_1, \dots, i_k\} \cap \{\beta_1, \dots, \beta_m\} = \{\delta_1, \dots, \delta_j\}$, where $1 \leq j \leq k \wedge m$, then*

$$E\{f(X_{i_1}, \dots, X_{i_k})|X_{\beta_1}, \dots, X_{\beta_m}\} = E\{f(X_{i_1}, \dots, X_{i_k})|X_{\delta_1}, \dots, X_{\delta_j}\}.$$

(ii) If $i_k \notin \{\beta_1, \dots, \beta_m\}$, then

$$\begin{aligned} E\{f(X_{i_1}, \dots, X_{i_k})|X_{\beta_1}, \dots, X_{\beta_m}, X_{i_1}, \dots, X_{i_{k-1}}\} \\ = E\{f(X_{i_1}, \dots, X_{i_k})|X_{i_1}, \dots, X_{i_{k-1}}\}. \end{aligned}$$

PROOF. Let $\{\alpha_{j+1}, \dots, \alpha_k\} = \{i_1, \dots, i_k\} - \{\delta_1, \dots, \delta_j\}$ and $\{\gamma_{j+1}, \dots, \gamma_m\} = \{\beta_1, \dots, \beta_m\} - \{\delta_1, \dots, \delta_j\}$. Set $\mathcal{G}_1 = \sigma(X_{\alpha_{j+1}}, \dots, X_{\alpha_k})$, $\mathcal{G}_2 = \sigma(X_{\gamma_{j+1}}, \dots, X_{\gamma_m})$, and $\mathcal{G}_3 = \sigma(X_{\delta_1}, \dots, X_{\delta_j})$. Then $\sigma(\mathcal{G}_1 \cup \mathcal{G}_3)$ is independent of \mathcal{G}_2 whence \mathcal{G}_1 and \mathcal{G}_2 are conditionally independent given \mathcal{G}_3 by Corollary 7.3.3. Since $f(X_{i_1}, \dots, X_{i_k})$ is $\sigma(\mathcal{G}_1 \cup \mathcal{G}_3)$ -measurable, (i) follows from Theorem 7.3.1 (iv).

Apropos of (ii), if A is the set of distinct integers among $\beta_1, \dots, \beta_m, i_1, \dots, i_{k-1}$, then $\{i_1, \dots, i_{k-1}\} = A \cap \{i_1, \dots, i_k\}$, so that (ii) follows from (i). \square

Corollary 1. If h is a completely degenerate kernel on R^k , $k \geq 2$, and $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ then, for any constants $\{a_j, j \geq 1\}$,

$$\left\{ \sum_{j=k}^n a_j \sum_{1 \leq i_1 < \dots < i_{k-1} < j} h(X_{i_1}, \dots, X_{i_{k-1}}, X_j), \mathcal{F}_n, n \geq k \right\}$$

is a martingale and, in particular ($a_j = 1, j \geq 1$), so is $\{S_{k,n}(h), \mathcal{F}_n, n \geq k\}$.

An important aspect of non-degenerate U -statistics is their decomposition into an average of i.i.d. random variables plus a finite linear combination of completely degenerate U -statistics. This is an immediate consequence of

Theorem 1 (Hoeffding's decomposition). If $\{X_n, n \geq 1\}$ are i.i.d. random variables and h is a symmetric function on R^k , $k \geq 2$, with $E h(X_1, \dots, X_k) = 0$, then (i)

$$U_{k,n}(h) = \sum_{j=1}^k \binom{k}{j} U_{j,n}(h_j), \quad (3)$$

where h_j is as in (2). Moreover, if h is degenerate of order $i - 1$, where $2 \leq i \leq k$, the first $i - 1$ terms of the sum in (3) vanish. Furthermore, (ii) if $S_{j,n}(h_j) = \binom{n}{j} U_{j,n}(h_j)$, $1 \leq j \leq k$, then $\{S_{j,n}(h_j), \mathcal{F}_n = \sigma(X_1, \dots, X_n), n \geq j\}$ is a martingale, $1 \leq j \leq k$, and, if $E h^2 < \infty$, then $E S_{l,n}(h_l) S_{m,n}(h_m) = 0$ for $m \neq l$.

PROOF. Replacing f, X_1, \dots, X_k by $h, X_{i_1}, \dots, X_{i_k}$ in Lemma 2 and summing,

$$\binom{n}{k} U_{k,n}(h) = S_{k,n}(h_k) + \sum_{j=1}^{k-1} \left[\sum_{1 \leq i_1 < \dots < i_k \leq n} \sum_{1 \leq i_1 < \dots < i_j \leq k} h(X_{i_1}, \dots, X_{i_j}) \right].$$

Each of the $\binom{n}{j}$ terms $h_j(X_{i_1}, \dots, X_{i_j})$ appears the same number of times in the bracketed double sum, and all together there are $\binom{n}{k} \binom{k}{j}$ terms. Thus, $h_j(X_{i_1}, \dots, X_{i_j})$ is repeated $\binom{n}{k} \binom{k}{j} / \binom{n}{j}$ times. Hence,

$$\binom{n}{k} U_{k,n}(h) = S_{k,n}(h_k) + \sum_{j=1}^{k-1} \frac{\binom{n}{k} \binom{k}{j}}{\binom{n}{j}} S_{j,n}(h_j) = \binom{n}{k} \sum_{j=1}^k \binom{k}{j} U_{j,n}(h_j),$$

which is tantamount to (3). According to Lemma 2, h_j is completely degenerate for $j \geq 2$, whereas $U_{1,n}(h_1) = 1/n \sum_{r=1}^n E\{h|X_r\}$ is an average of i.i.d. random variables. Moreover, if h is degenerate of order $i - 1$, where $2 \leq i \leq k$, then $h_j^* = 0 = h_j$ for $1 \leq j \leq i - 1$ via (2), whence the first $i - 1$ terms of (3) vanish.

Furthermore, since h_j is completely degenerate for $j \geq 2$, Corollary 1 ensures that $\{S_{j,n}(h_j), \mathcal{F}_n, n \geq k\}$ is a martingale when $j \geq 2$, and clearly the same conclusion holds for $j = 1$.

Finally, if $l \neq m$, it follows from Lemma 4(i) that

$$E h_l(X_{i_1}, \dots, X_{i_l}) h_m(X_{\beta_1}, \dots, X_{\beta_m}) = 0,$$

implying that $E S_{l,n}(h_l) S_{m,n}(h_m) = 0$ for $l \neq m$. \square

Lemma 4. Let $1 \leq i_1 < \dots < i_k \leq n$ and $1 \leq \beta_1 < \dots < \beta_m \leq n$ be two sets of integers with $\{i_1, \dots, i_k\} \cap \{\beta_1, \dots, \beta_m\} = \{i_1, \dots, i_j\}$, where $1 \leq j < k$. (i) If f is a completely degenerate function on R^k , $k \geq 2$, and g is a function on R^m with $|E g| \leq \infty$, $|E fg| \leq \infty$, then

$$E f(X_{i_1}, \dots, X_{i_k})g(X_{\beta_1}, \dots, X_{\beta_m}) = 0.$$

(ii) If $m = k$ and h is a symmetric function on R^k , $k \geq 2$, with $E h^2 < \infty$, then

$$E h(X_{i_1}, \dots, X_{i_k})h(X_{\beta_1}, \dots, X_{\beta_k}) = E[E\{h(X_1, \dots, X_k)|X_1, \dots, X_j\}]^2,$$

and the right side is an increasing function of j .

PROOF. Employing Lemma 3(i),

$$\begin{aligned} & E\{f(X_{i_1}, \dots, X_{i_k})g(X_{\beta_1}, \dots, X_{\beta_m})|X_{i_1}, \dots, X_{i_j}\} \\ &= E\{f(X_{i_1}, \dots, X_{i_k})E\{g(X_{\beta_1}, \dots, X_{\beta_m})|X_{i_1}, \dots, X_{i_k}\}|X_{i_1}, \dots, X_{i_j}\} \\ &= E\{f(X_{i_1}, \dots, X_{i_k})E\{g(X_{\beta_1}, \dots, X_{\beta_m})|X_{i_1}, \dots, X_{i_j}\}|X_{i_1}, \dots, X_{i_j}\} \\ &= E\{g(X_{\beta_1}, \dots, X_{\beta_m})|X_{i_1}, \dots, X_{i_j}\}E\{f(X_{i_1}, \dots, X_{i_k})|X_{i_1}, \dots, X_{i_j}\}, \end{aligned} \tag{4}$$

and the last term equals zero a.s. when f is completely degenerate, so that (i) is an immediate consequence of (4).

Apropos of (ii), taking $m = k$, $g = f = h$ in (4),

$$\begin{aligned} & E\{h(X_{i_1}, \dots, X_{i_k})h(X_{\beta_1}, \dots, X_{\beta_k})|X_{i_1}, \dots, X_{i_j}\} \\ &= E\{h(X_{i_1}, \dots, X_{i_k})|X_{i_1}, \dots, X_{i_j}\}E\{h(X_{\beta_1}, \dots, X_{\beta_k})|X_{i_1}, \dots, X_{i_j}\} \\ &= E^2\{h(X_1, \dots, X_k)|X_1, \dots, X_j\} \end{aligned}$$

since $\{X_n, n \geq 1\}$ are i.i.d. whence (ii) follows upon taking expectations. Moreover, if $Y_j = E\{h(X_1, \dots, X_k)|X_1, \dots, X_j\}$, $1 \leq j \leq k$ and $\mathcal{F}_j = \sigma(X_1, \dots, X_j)$, then $\{Y'_j, \mathcal{F}_j, 1 \leq j \leq k\}$ is a martingale for $r = 1$ and hence a submartingale for $r = 2$. Thus, $E Y_j^2$, $1 \leq j \leq k$ is an increasing sequence. \square

Corollary 2. Let $E h^2 < \infty$. If the U-statistic $U_{k,n}(h)$ is degenerate of order $i - 1$, where $2 \leq i \leq k$ or non-degenerate ($i = 1$) with $E h = 0$, then its variance is

$$\sigma_{U_{k,n}(h)}^2 = \binom{n}{k}^{-1} \sum_{j=i}^k \binom{k}{j} \binom{n-k}{k-j} E[E^2\{h(X_1, \dots, X_k)|X_1, \dots, X_j\}],$$

and so

$$\sigma_{U_{n,k}(h)}^2 = \frac{i! \binom{k}{i}^2 e^i}{n^i} E[E^2\{h(X_1, \dots, X_k)|X_1, \dots, X_i\}] + O\left(\frac{1}{n^{i+1}}\right).$$

PROOF. Set $q_j = E[h^2(X_1, \dots, X_k) | X_1, \dots, X_j]$. Since the number of pairs $(i_1, \dots, i_k), (\beta_1, \dots, \beta_k)$ with exactly j integers in common in $\binom{n}{k} \binom{k}{j} \binom{n-k}{k-j}$, Lemma 4 yields

$$\begin{aligned}\sigma_{U_{k,n}(h)}^2 &= \binom{n}{k}^{-2} \sum_{1 \leq i_1 < \dots < i_k \leq n} \sum_{1 \leq \beta_1 < \dots < \beta_k \leq n} E[h(X_{i_1}, \dots, X_{i_k}) h(X_{\beta_1}, \dots, X_{\beta_k})] \\ &= \binom{n}{k}^{-1} \sum_{j=1}^k \binom{k}{j} \binom{n-k}{k-j} q_j,\end{aligned}$$

and, if h is degenerate of order $i-1$, that first $i-1$ terms of the sum vanish.

For large n , the dominant term of the sum is

$$\begin{aligned}\binom{k}{i} \binom{n-k}{k-i} \binom{n}{k}^{-1} q_i &= i! \binom{k}{i}^2 q_i \frac{[(n-k)!]^2}{n!(n-2k+i)!} \\ &= i! \binom{k}{i}^2 q_i \left(\frac{e}{n}\right)^i \left(1 + O\left(\frac{1}{n}\right)\right)\end{aligned}$$

via Stirling's formula (Lemma 2.3.1). \square

Since (Example 7.4.3) $\{U_{k,n}(h), \sigma(U_{k,j}), j \geq n, n \geq k\}$ is a reverse martingale for $h \in \mathcal{L}_1$, it follows (see the discussion after Theorem 11.1.1) under this proviso that $U_{k,n}(h) \xrightarrow{\text{a.s.}} E h$ or, equivalently, $[1/\binom{n}{k}] S_{k,n}(h) \xrightarrow{\text{a.s.}} E h$. The next theorem provides an analogue of the Marcinkiewicz–Zygmund strong law of large numbers for U -statistics.

Define

$$p_j = \frac{jp}{k - (k-j)p}, \quad 1 \leq j \leq k. \quad (5)$$

Then $p_k = p$ and for p in $(1, 2k/(2k-j))$, the sequence $\{p_j, 1 \leq j \leq k\}$ is decreasing, and $p_j \in (1, 2)$.

Theorem 2. Let $\{X_n, n \geq 1\}$ be i.i.d. random variables on some probability space (Ω, \mathcal{F}, P) and h a symmetric function on R^k , $k \geq 2$. If (i) h is degenerate of order $i-1$ and $1 < p < 2k/(2k-i)$, where $2 \leq i \leq k$, then

$$n^{k(p-1)/p} |U_{k,n}(h) - E h| \xrightarrow{\text{a.s.}} 0, \quad \text{that is, } n^{-k/p} S_{k,n}(h - E h) \xrightarrow{\text{a.s.}} 0 \quad (6)$$

provided

$$E\{h(X_1, \dots, X_k) | X_1, \dots, X_j\} \in \mathcal{L}_{p_j} \quad i \leq j \leq k. \quad (7)$$

(ii) Alternatively, (6) holds for non-degenerate h if $1 < p < k/(k-1)$ and (7) is satisfied for $i = 1$. (iii) Moreover, if $0 < p < 1$ and $h \in \mathcal{L}_p$, then $n^{-k/p} S_{k,n}(h) \xrightarrow{\text{a.s.}} 0$.

PROOF. Suppose initially that $i = k$ and $h \in \mathcal{L}_p$, whence, according to (i), $1 < p < 2$. Hence, $\alpha = k/p \in (k/2, k)$. Set

$$\begin{aligned} g_j(X_1, \dots, X_k) &= h(X_1, \dots, X_k) I_{[|h| \leq j^p]} \\ &\quad - E\{h(X_1, \dots, X_k) I_{[|h| \leq j^p]} | X_1, \dots, X_{k-1}\} \end{aligned}$$

and define

$$\begin{aligned} \tilde{S}_{k,n}(h) &= \sum_{j=k}^n j^{-\alpha} \sum_{1 \leq i_1 < \dots < i_{k-1} < j} h(X_{i_1}, \dots, X_{i_{k-1}}, X_j), \\ T_{k,n}(h) &= \sum_{j=k}^n j^{-\alpha} \sum_{1 \leq i_1 < \dots < i_{k-1} < j} g_j(X_{i_1}, \dots, X_{i_{k-1}}, X_j). \end{aligned}$$

If $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, Corollary 1 asserts that $\{\tilde{S}_{k,n}(h), \mathcal{F}_n, n \geq k\}$ is a martingale and the same is true of $\{T_{k,n}(h), \mathcal{F}_n, n \geq k\}$ since, via Lemma 3(i),

$$\begin{aligned} E\{g_n(X_{i_1}, \dots, X_{i_{k-1}}, X_n) | \mathcal{F}_{n-1}\} \\ = E\{g_n(X_{i_1}, \dots, X_{i_{k-1}}, X_n) | X_{i_1}, \dots, X_{i_{k-1}}\} \xrightarrow{\text{a.s.}} 0. \end{aligned}$$

Hence, employing Lemma 4(i) and setting $A_n = \{(n-1) < |h|^{1/\alpha} \leq n\}, n \geq 1$,

$$\begin{aligned} E[T_{k,n}(h)]^2 &= \sum_{j=k}^n j^{-2\alpha} \binom{j-1}{k-1} E g_j^2 \leq \sum_{j=k}^{\infty} j^{k-1-2\alpha} E h^2 I_{[|h| \leq j^p]} \\ &= \sum_{j=k}^{\infty} j^{k-1-2\alpha} \sum_{n=1}^j \int_{A_n} h^2 dP \leq \sum_{n=1}^{\infty} \left(\frac{1}{n} + \frac{1}{2\alpha-k}\right) n^{k-2\alpha} \int_{A_n} h^2 dP \\ &\leq \left(\frac{2\alpha-k+1}{2\alpha-k}\right) \sum_{n=1}^{\infty} \int_{A_n} |h|^{k/\alpha} dP = \left(\frac{2\alpha-k+1}{2\alpha-k}\right) E|h|^p < \infty. \end{aligned}$$

Moreover, since h is completely degenerate,

$$\begin{aligned} \sup_{n \geq k} E|\tilde{S}_{k,n}(h) - T_{k,n}(h)| &\leq 2 \sum_{j=k}^{\infty} j^{-\alpha} \binom{j-1}{k-1} E|h| I_{[|h| > j^p]} \\ &\leq 2E\left(|h| \sum_{j \leq |h|^{1/\alpha}} j^{k-1-\alpha}\right) \leq CE|h|^{k/\alpha} < \infty. \end{aligned}$$

Thus, in view of

$$\sup_{n \geq k} E|\tilde{S}_{k,n}(h)| \leq \sup_{n \geq k} E|\tilde{S}_{k,n}(h) - T_{k,n}(h)| + \sup_{n \geq k} E|T_{k,n}(h)|,$$

the martingale $\{\tilde{S}_{k,n}(h), \mathcal{F}_n, n \geq k\}$ is \mathcal{L}_1 -bounded and hence convergent so that Kronecker's lemma yields

$$n^{-k/p} S_{k,n-1}(h) = n^{-\alpha} \sum_{j=k}^n \sum_{1 \leq i_1 < \dots < i_{k-1} < j} h(X_{i_1}, \dots, X_{i_{k-1}}, X_j) \xrightarrow{\text{a.s.}} 0. \quad (8)$$

Therefore, (8) holds whenever h is a completely degenerate kernel belonging

Now,

$$\begin{aligned}
 & \sum_{j=k}^{\infty} P\{D_{k-1,j}(h) \neq D_{k-1,j}(h_j)\} \\
 & \leq \sum_{j=k}^{\infty} P\{h(X_{i_1}, \dots, X_{i_{k-1}}, X_j) > j^{\alpha} \\
 & \quad \text{for some choice of } 1 \leq i_1 < \dots < i_{k-1} < j\} \\
 & \leq \sum_{j=k}^{\infty} \binom{j-1}{k-1} P\{h(X_1, \dots, X_k) > j^{\alpha}\} \leq \sum_{j=k}^{\infty} j^{k-1} \sum_{n=j+1}^{\infty} P\{A_n\} \\
 & \leq \sum_{n=1}^{\infty} (n-1)^k E I_{A_n} \leq E \sum_{n=1}^{\infty} h^{k/\alpha} I_{A_n} = E h^p < \infty.
 \end{aligned}$$

Hence, $P\{D_{k-1,n}(h) \neq D_{k-1,n}(h_n), \text{i.o.}\} = 0$. Moreover, since $\alpha > k$,

$$\begin{aligned}
 E \sum_{j=k}^n j^{-\alpha} D_{k-1,j}(h_j) &= \sum_{j=k}^n j^{-\alpha} \binom{j-1}{k-1} E h_j \leq \sum_{j=k}^{\infty} j^{k-1-\alpha} \sum_{n=1}^j E h I_{A_n} \\
 &\leq \frac{(\alpha-k+1)}{\alpha-k} \sum_{n=1}^{\infty} \frac{1}{n^{2k-1}} E h I_{A_n} \leq \left(\frac{\alpha-k+1}{\alpha-k} \right) E h^{k/\alpha} < \infty.
 \end{aligned}$$

Thus, the series $\sum_{j=k}^{\infty} j^{-\alpha} D_{k-1,j}(h_j)$ and hence also $\sum_{j=k}^{\infty} j^{-\alpha} D_{k-1,j}(h)$ converges a.s. whence $n^{-k/p} S_{k,n}(h) \xrightarrow{\text{a.s.}} 0$. \square

Corollary 3. Let $h \in \mathcal{L}_p$ with h degenerate of order $i-1$ and $E\{h(X_1, \dots, X_k)|X_1, \dots, X_{k-1}\} \in \mathcal{L}_{p_i}$ where p_i is as in (5). If (i) $1 < p < 2k/(2k-i)$ when $i \geq 2$ or (ii) $1 < p < k/(k-1)$ when $i=1$ and $E h=0$, then $n^{-k/p} S_{k,n}(h) \xrightarrow{\text{a.s.}} 0$.

Corollary 4. If $\{U_{k,n}(h), n \geq k \geq 2\}$ is a sequence of completely degenerate U-statistics with $E|h|^{2k/(2k-1)} < \infty$, then, as $n \rightarrow \infty$,

$$n^{1/2} U_{k,n}(h) \xrightarrow{\text{a.s.}} 0. \quad (10)$$

PROOF. For completely degenerate $h \in \mathcal{L}_p$, $1 < p < 2$, Theorem 2 ensures that $n^{k(p-1)/p} U_{k,n}(h) \xrightarrow{\text{a.s.}} 0$. Since $k(p-1)/p = 1/2$ for $p = 2k/(2k-1)$, (10) follows. \square

Lemma 5. Let $\{U_{k,n}(h), n \geq k \geq 2\}$ be a sequence of non-degenerate U-statistics with $E h = 0$. If

$$E\{h(X_1, \dots, X_k)|X_1, \dots, X_j\} \in \mathcal{L}_{2j/(2j-1)}, \quad 2 \leq j \leq k \quad (11)$$

(a fortiori, if $E|h|^{4/3} < \infty$), then

$$n^{1/2} U_{k,n}(h) = \frac{k}{n^{1/2}} \sum_{r=1}^n E\{h(X_r, X_{r+1}, \dots, X_{r+k-1})|X_r\} + o(1), \quad \text{a.s.} \quad (12)$$

PROOF. Via Hoeffding's decomposition,

$$n^{1/2} U_{k,n}(h) = \frac{k}{n^{1/2}} \sum_{r=1}^n E\{h(X_r, \dots, X_{r+k-1})|X_r\} + n^{1/2} \sum_{j=2}^k \binom{k}{j} U_{j,n}(h_j).$$

As in the proof of Theorem 2, (11) implies that $h_j \in \mathcal{L}_{2j/(2j-1)}$, $2 \leq j \leq k$ whence (12) follows from Corollary 4. Clearly, $E|h|^{4/3} < \infty$ ensures (11). \square

A Central Limit Theorem and Law of the Iterated Logarithm for non-degenerate U -statistics will be proved in Sections 9.1 and 10.2.

EXERCISE 7.5

1. If $\tilde{S}_{k,n} = \sum_{1 \leq i_1 < \dots < i_k \leq n} X_{i_1} \cdots X_{i_k}$, $n \geq k \geq 1$ and $\tilde{S}_{0,n} = 1$, $\tilde{S}_{1,n} = \sum_{i=1}^n X_i$, and $S_n^{(j)} = \sum_{i=1}^n X_i^j$, $j \geq 1$, prove that (i)

$$(k+1)\tilde{S}_{k+1,n} = \sum_{j=0}^k (-1)^j \tilde{S}_{k-j,n} S_n^{(j+1)}, \quad k \geq 1,$$

and (ii) if $\{X, X_n, n \geq 1\}$ are i.i.d. with $E X = 0$, then $\tilde{S}_{k,n}/\binom{n}{k}$ is a completely degenerate U -statistic for $k \geq 2$.

2. Let $\tilde{S}_{k,n}$, $n \geq k$ and $S_n^{(j)}$, $j \geq 1$ be as in Exercise 1, where $\{X, X_n, n \geq 1\}$ are i.i.d., and let $\{b_n, n \geq 1\}$ be constants such that $0 < b_n/n^{1/2} \uparrow \infty$. If (i) $(1/b_n) \sum_1^n X_i \xrightarrow{a.s.} 0$ and (ii) $\sum_{n=1}^{\infty} P\{|X| > b_n\} < \infty$, prove that $\tilde{S}_{k,n}/b_n^k \xrightarrow{a.s.} 0$, $k \geq 2$. In particular, if $X \in \mathcal{L}_p$, $0 < p < 2$ and $E X = 0$ whenever $p \geq 1$, then $\tilde{S}_{k,n}/n^{k/p} \xrightarrow{a.s.} 0$. Hint: Employ the Newton identities

$$k\tilde{S}_{k,n} = (-1)^k \sum \prod_{i=1}^k (S_n^{(j)})^{m_j} / j^{m_j} m_j!,$$

where the sum is over all non-negative integers m_j , $1 \leq j \leq k$ satisfying $\sum_{j=1}^k jm_j = k$.

3. Find a non-degenerate kernel h such that $E|h|^p = \infty$, some $p > 1$, but $E\{h(X_1, \dots, X_k)|X_1\} \in \mathcal{L}_q$, all $q > 0$.
4. Show that if $E h(X_1, \dots, X_k) = 0$, for $2 \leq j \leq k$

$$\begin{aligned} E\{h(X_1, \dots, X_k)|X_1, \dots, X_j\} &= h_j(X_1, \dots, X_j) + \sum_{i=1}^j h_i(X_i) \\ &\quad + \sum_{1 \leq i_1 < i_2 \leq j} h_2(X_{i_1}, X_{i_2}) + \dots \\ &\quad + \sum_{1 \leq i_1 < \dots < i_{j-1} \leq j} h_{j-1}(X_{i_1}, \dots, X_{i_{j-1}}). \end{aligned}$$

5. (Kemperman) For any countable partition of $(-\infty, \infty)$ into Borel sets B_j , $j = 1, 2, \dots$, define the non-negative, symmetric function h by $h(X_1, \dots, X_k) = \sum_{j=1}^{\infty} u_j \prod_{i=1}^k I_{\{X_i \in B_j\}}$, $k \geq 2$ and let $\{X, X_n, n \geq 1\}$ be i.i.d. random variables with $P\{X \in B_j\} = \pi_j > 0$, $j \geq 1$, where $\sum_1^{\infty} \pi_j = 1$. Set $\pi_j = c/j^{\alpha}$, $u_j = j^{\beta}$ with $\alpha > 1$, $\beta > 1$. Prove that if $k\alpha - \beta > 1 \geq k\alpha - p\beta$ and $\alpha[1 + (k-1)p] > 1 + p\beta$, then $h \in \mathcal{L}_1$.

$h \notin \mathcal{L}_p$, $E\{h(X_1, \dots, X_k) | X_1\} \in \mathcal{L}_p$. These inequalities hold, for example, if $\beta = k - 1$ and $1 < \alpha < [1 + (k - 1)p]/k$.

6. If $1 < p < (2k/2k - i)$, there exist degenerate kernels of order $i = 1, 2 \leq i \leq k$ such that $n^{k(p-1)/p} U_{k,n} \xrightarrow{\text{a.s.}} 0$ despite $E|E\{h(X_1, \dots, X_k) | X_1, \dots, X_i\}|^{p_i} = \infty$, where p_i is as in (5). Hint: Let $\{X_n, n \geq 1\}$ be symmetric i.i.d. random variables with $P\{|X| > t\} = ct^{-p}(\log t)^{-1}, t \geq 2$, where $1 < p < 2$. Then

$$g_i = \sum_{1 \leq x_1 < \dots < x_{k-i} \leq k} \left(\prod_{r=1}^{k-i} |x_{x_r}| \right) \prod_{s \neq x_1, \dots, x_{k-i}} x_s$$

is degenerate of order $i - 1, 2 \leq i \leq k$. Hint: Set $X'_l = X_l I_{\{|X_l| \leq e^{2^{l-1}p}\}}, l \geq 1, \alpha > 0$.

References

- Y. V. Borovskich and V. S. Korolyuk, *Theory of U-Statistics*, Kluwer Academic Publishers, Boston, 1994.
- L. Breiman, *Probability*, Addison-Wesley, Reading, Mass., 1968.
- H. Bühlman, "Austauschbare stochastiche Variablen und ihre Grenzwertsätze," *Univ. of California Publications in Statistics*, **3** (1960), 1–36.
- Y. S. Chow, "A martingale inequality and the law of large numbers," *Proc. Amer. Math. Soc.* **11** (1960), 107–111.
- Y. S. Chow, H. Robbins, and D. Siegmund, *Great Expectations: The Theory of Optimal Stopping*, Houghton Mifflin, Boston, 1972.
- Y. S. Chow, H. Robbins, and H. Teicher, "Moments of randomly stopped sums," *Ann. Math. Stat.* **36** (1965), 789–799.
- K. L. Chung, *A course in Probability Theory*, Harcourt Brace, New York, 1968; 2nd ed., Academic Press, New York, 1974.
- J. L. Doob, "Regular properties of certain families of chance variables," *Trans. Amer. Math. Soc.* **47** (1940), 455–486.
- J. L. Doob, *Stochastic Processes*, Wiley, New York, 1953.
- L. E. Dubins and D. A. Freedman, "A sharper form of the Borel-Cantelli lemma and the strong law," *Ann. Math. Stat.* **36** (1965), 800–807.
- E. B. Dynkin and A. Mandelbaum, "Symmetric statistics, Poisson point processes and multiple Wiener integrals," *Ann. Statist.* **11** (1983), 739–745.
- B. de Finetti, La prévision; ses lois logiques, ses sources subjectives," *Annales de l'Institut Henri Poincaré* **7** (1937), 1–68.
- E. Giné and J. Zinn, "Marcinkiewicz-type laws of large numbers and convergence of moments for U-statistics," *Probability in Banach Spaces* **8** (1992), 273–291, Birkhäuser, Boston.
- J. Hájek and A. Rényi, "Generalization of an inequality of Kolmogorov," *Acta Math. Acad. Sci. Hung.* **6** (1955), 281–283.
- P. R. Halmos, *Measure Theory*, Van Nostrand, Princeton, N. J., 1950; Springer-Verlag, Berlin and New York, 1974.
- E. Hewitt and L. J. Savage, "Symmetric measures on Cartesian products," *Trans. Amer. Math. Soc.* **80** (1955), 470–501.
- W. Hoeffding, "The strong law of large numbers for U-statistics," Institute of Statistics Mimeo Series 302 (1961), University of North Carolina, Chapel Hill, NC.
- D. G. Kendall, "On finite and infinite sequences of exchangeable events," *Studia Scient. Math. Hung.* **2** (1967), 319–327.

- M. J. Klass, "Properties of optimal extended-valued stopping rules," *Ann. Prob.* **1** (1973), 719–757.
- K. Krickeberg, *Probability Theory*, Addison-Wesley, Reading, Mass., 1965.
- P. Lévy, *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, Paris, 1937; 2nd ed., 1954.
- M. Loève, *Probability Theory*, 3rd ed., Van Nostrand, Princeton, 1963; 4th ed., Springer-Verlag, Berlin and New York, 1977–1978.
- P. K. Sen, "On \mathcal{L}_p convergence of *U*-statistics," *Ann. Inst. Statist. Math.* **26** (1974), 55–60.
- R. J. Serfling, *Approximation Theorems of Mathematical Statistics*, Wiley, New York.
- H. Teicher, "On the Marcinkiewicz-Zygmund strong law for *U*-statistics," *J. Theoret. Probability* **10** (1997),

8

Distribution Functions and Characteristic Functions

8.1 Convergence of Distribution Functions, Uniform Integrability, Helly–Bray Theorem

Distribution functions are mathematical artifacts with properties that are independent of any probabilistic setting. Notwithstanding, most of the theorems of interest are geared to d.f.s of r.v.s and the majority of proofs are simpler and more intuitive when couched in terms of r.v.s having, or probability measures determined by, the given d.f.s. Since r.v.s possessing pre-assigned d.f.s can always be defined on some probability space, the language of r.v.s and probability will be utilized in many of the proofs without further ado.

Recall that a d.f. on the line is a nondecreasing, left-continuous function on $R = [-\infty, \infty]$ with $F(\infty) = \lim_{x \rightarrow \infty} F(x) = 1$, $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$. A discrete d.f. was defined in Section 1.6 as roughly tantamount to a “step function” with a finite or denumerable number of jumps. As such, it determines and is determined by a probability density function (p.d.f.), say f , and a nonempty countable subset S of $(-\infty, \infty)$ with f positive on S and vanishing on S^c . Absolutely continuous d.f.s were encountered in Section 6.5. A d.f. F is absolutely continuous iff $F(x) = \int_{-\infty}^x f(t)dt$, $-\infty < x < \infty$, for some Borel function $f \geq 0$, a.e., with $\int_{-\infty}^{\infty} f(t)dt = 1$. A d.f. F is termed **singular** if it is continuous and its corresponding probability measure is singular with respect to Lebesgue measure (Exercise 5).

The first proposition states that any d.f. on R is a convex linear combination of these three types.

Also, a d.f. is **degenerate** or **improper** if it has only a single point of increase (Exercise 1.6.4) and otherwise **nondegenerate** or **proper**.

Proposition 1. If F is an arbitrary d.f. on $R = [-\infty, \infty]$, then $F = \alpha_1 F_1 + \alpha_2 F_2 + \alpha_3 F_3$ where $\sum_{i=1}^3 \alpha_i = 1$, $\alpha_i \geq 0$, $i = 1, 2, 3$, and F_1, F_2, F_3 are discrete, absolutely continuous and singular d.f.s respectively.

PROOF. If F is discrete, $F = F_1$, $\alpha_1 = 1$, while if F is continuous, F coincides with F^* in what follows. Set $S_1 = \{x : F(x+) - F(x) > 0\}$ so that if F is neither discrete nor continuous, $\alpha_1 = P\{S_1\} \in (0, 1)$ where P is the measure induced by F . Hence, if P_1 is the probability measure determined by

$$P_1\{\{x\}\} = \frac{1}{\alpha_1} P\{\{x\}\}, \quad x \in S_1,$$

$$P_1\{B\} = 0, \quad B \in S_1^c \cdot \mathcal{B},$$

then the d.f. corresponding to P_1 , say F_1 , is discrete. Moreover, $P^* = [1/(1 - \alpha_1)](P - \alpha_1 P_1)$ is a probability measure vanishing on all one-point sets, whence its corresponding d.f. $F^* = [1/(1 - \alpha_1)](F - \alpha_1 F_1)$ is continuous. If P^* is absolutely continuous (resp. singular) relative to Lebesgue measure, its d.f. F^* may be taken as F_2 (resp. F_3) and $\alpha_3 = 0$ (resp. $\alpha_2 = 0$). Otherwise, by Corollary 6.5.1, $F^* = \beta F_2 + (1 - \beta)F_3$, $0 < \beta < 1$, where F_2 is absolutely continuous and, moreover, F_3 is singular. Thus, $F - \alpha_1 F_1 = (1 - \alpha_1)F^* = \beta(1 - \alpha_1)F_2 + (1 - \beta)(1 - \alpha_1)F_3$ is the asserted decomposition. \square

The **support** (Exercise 1.6.4) or **spectrum** of an arbitrary d.f. F is the closed set S defined by

$$S = \{x : F(x + \varepsilon) - F(x - \varepsilon) > 0, \text{ all } \varepsilon > 0\}.$$

and the elements of S are called **points of increase**.

An instance of convergence of a sequence of d.f.s to a d.f. occurred in Corollary 2.3.1, but the situation there was too specialized to furnish clues to the general problem.

For any real function G , let $C(G)$ denote the **set of continuity points** of G , that is, $C(G) = \{x : -\infty < x < \infty, G(x-) = G(x+) = G(x)\}$. Note that if G is monotone, $C(G)$ is the complement of a countable set and *a fortiori* dense in $(-\infty, \infty)$.

Definition. A sequence of nondecreasing functions G_n on $(-\infty, \infty)$ is said to **converge weakly** to a nondecreasing function G on $(-\infty, \infty)$, denoted by $G_n \xrightarrow{*} G$, if $\lim_{n \rightarrow \infty} G_n(x) = G(x)$ for all $x \in C(G)$. If, in addition, $G_n(\infty) \rightarrow G(\infty)$ and $G_n(-\infty) \rightarrow G(-\infty)$ where, as usual, $G(\pm\infty) = \lim_{x \rightarrow \pm\infty} G(x)$, then $\{G_n\}$ is said to **converge completely** to G , denoted by $G_n \xrightarrow{s} G$.

In the special case of d.f.s F_n , complete convergence of $\{F_n\}$ guarantees that the “limit function” F , if left continuous (as may and will be supposed via Lemma 8.2.1 even when merely $F_n \xrightarrow{*} F$), is a d.f.

If $\{X_n, n \geq 1\}$ is a sequence of r.v.s on some probability space (Ω, \mathcal{F}, P) with d.f.s F_{X_n} that converge completely to F , the r.v.s $\{X_n\}$ are said to **converge in distribution or law**, symbolized by $X_n \xrightarrow{d} X_F$. Here X_F is, in general, a **fictitious r.v.** with d.f. F . It is not asserted that any such "r.v." exists on (Ω, \mathcal{F}, P) , but, of course, one can always define a r.v. X with d.f. F on another probability space; rather, $X_n \xrightarrow{d} X_F$ is simply a convenient alternative notation for $F_{X_n} \xrightarrow{c} F$. Clearly, convergence in distribution is a property of the d.f.s of the r.v.s in question and not of the r.v.s themselves.

However, if $X_n \xrightarrow{P} X$ (*a fortiori*, if $X_n \xrightarrow{ac} X$ or $X_n \xrightarrow{d} X$), then the following Corollary 1 asserts that a bona fide r.v. X_F on (Ω, \mathcal{F}, P) does exist and coincides with X . Such a case may be denoted simply by $X_n \xrightarrow{d} X$, that is, $F_{X_n} \xrightarrow{c} F_X$.

Theorem 1 (Slutsky). If $\{X_n, n \geq 1\}$ and $\{Y_n, n \geq 1\}$ are r.v.s on some probability space with $X_n - Y_n \xrightarrow{P} 0$ and $Y_n \xrightarrow{d} X_F$, then $X_n \xrightarrow{d} X_F$.

PROOF. Let $x, x \pm \varepsilon \in C(F)$, where $\varepsilon > 0$ and $x \in (-\infty, \infty)$. Then

$$\begin{aligned} P\{X_n < x\} &= P\{X_n < x, |X_n - Y_n| < \varepsilon\} + P\{X_n < x, |X_n - Y_n| \geq \varepsilon\} \\ &\leq F_{Y_n}(x + \varepsilon) + P\{|X_n - Y_n| \geq \varepsilon\} \end{aligned}$$

and, analogously,

$$F_{Y_n}(x - \varepsilon) \leq P\{X_n < x\} + P\{|X_n - Y_n| \geq \varepsilon\}.$$

Thus,

$$F(x - \varepsilon) \leq \varliminf_n F_{X_n}(x) \leq \varlimsup_n F_{X_n}(x) \leq F(x + \varepsilon),$$

and letting $\varepsilon \rightarrow 0$ subject to $x \pm \varepsilon$ in $C(F)$, the conclusion follows. \square

Corollary 1. If $\{X, X_n, n \geq 1\}$ are r.v.s on some probability space with $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{d} X$.

Corollary 2. If $\{X_n\}, \{Y_n\}, \{Z_n\}$ are sequences of r.v.s on (Ω, \mathcal{F}, P) with $X_n \xrightarrow{d} X_F$, $Y_n \xrightarrow{P} a$, $Z_n \xrightarrow{P} b$, where a, b are finite constants, then

$$X_n Y_n + Z_n \xrightarrow{d} aX_F + b.$$

Note. Here, $aX_F + b$ is a fictitious r.v. whose distribution coincides with that of $aX + b$ when X is a bona fide r.v. with d.f. F .

PROOF. By the theorem it suffices to prove $X_n Y_n + b \xrightarrow{d} aX_F + b$ or equivalently that $X_n Y_n \xrightarrow{d} aX_F$. Since it is obvious that $aX_n \xrightarrow{d} aX_F$, applying the theorem once more it suffices to prove that $X_n(Y_n - a) \xrightarrow{P} 0$ or, renotating, that $X_n U_n \xrightarrow{P} 0$ if $X_n \xrightarrow{d} X_F$, $U_n \xrightarrow{P} 0$. To this end, for any $\delta > 0$, choose $\pm h \in C(F)$

such that $F(h) - F(-h) \geq 1 - \delta$. Then, for all sufficiently large n and any $\varepsilon > 0$,

$$\begin{aligned} P\{|U_n X_n| > \varepsilon\} &\leq P\{|U_n X_n| > \varepsilon, 0 < |X_n| \leq h\} + 2\delta \\ &\leq P\{|U_n| > \varepsilon/h\} + 2\delta \xrightarrow{n \rightarrow \infty} 2\delta \end{aligned}$$

and the result follows as $\delta \rightarrow 0$. \square

Corollary 3. If $\{a, b, a_n, b_n, n \geq 1\}$ are finite constants with $a_n \rightarrow a, b_n \rightarrow b$ and the r.v.s $X_n \xrightarrow{d} X_F$, then $a_n X_n + b_n \xrightarrow{d} aX_F + b$.

If $\{X_n\}$ is a sequence of r.v.s and b_n is a sequence of positive constants such that $X_n/b_n \xrightarrow{P} 0$, it is natural, paralleling the classical notation, to designate this by $X_n = o_p(b_n)$. Analogously, $X_n = O_p(b_n)$ will signify that X_n/b_n is bounded in probability, i.e., for every $\varepsilon > 0$, there are constants $C_\varepsilon, N_\varepsilon$ such that

$$P\{|X_n| \geq C_\varepsilon b_n\} \leq \varepsilon$$

for $n > N_\varepsilon$. In this notation, Theorem 1 says that if $X_n \xrightarrow{d} X_F$, then $X_n + o_p(1) \xrightarrow{d} X_F$. A calculus paralleling that of o and O exists for o_p and O_p . For example, the Taylor expansion

$$f(x) = \sum_{j=0}^k \frac{(x-c)^j}{j!} f^{(j)}(c) + o(|x-c|^k),$$

valid as $x \rightarrow c$ under the hypothesis below (Cramér, 1946, p. 290) leads directly to

Corollary 4. If $f(x)$ has k derivatives at $x = c$ and the r.v.s X_n satisfy $X_n = c + o_p(b_n)$, where $b_n = 1$ or $b_n = o(1)$, then

$$f(X_n) = \sum_{j=0}^k \frac{(X_n - c)^j}{j!} f^{(j)}(c) + o_p(b_n^k).$$

It will be advantageous to prove the ensuing for nondecreasing functions G_n on $(-\infty, \infty)$. In the special case where the G_n are d.f.s, the condition which conjoined with weak convergence yields complete convergence is (iii)(γ) in

Lemma 1. Let $\{G_n, n \geq 0\}$ be finite, nondecreasing functions on $(-\infty, \infty)$ with $G_n \xrightarrow{P} G_0$. Set $\Delta G_n = G_n(\infty) - G_n(-\infty)$, $n \geq 0$, where $G(\infty) = G(\infty -)$ and $G(-\infty) = G(-\infty +)$. Then

- i. $\overline{\lim}_{n \rightarrow \infty} G_n(-\infty) \leq G_0(-\infty) \leq G_0(\infty) \leq \underline{\lim}_{n \rightarrow \infty} G_n(\infty)$,
- ii. $\Delta G_0 \leq \underline{\lim}_{n \rightarrow \infty} \Delta G_n$.

Moreover, if $\Delta G_n(a) = G_n(a) - G_n(-a)$ for $n \geq 0$, $0 < a < \infty$, and if $\Delta G_n < \infty$ for $n \geq 1$, then

- iii. (α) $\lim G_n(\pm \infty) = G_0(\pm \infty)$, finite iff (β) $\lim_{n \rightarrow \infty} \Delta G_n = \Delta G_0 < \infty$ iff (γ) $\sup_{n \geq 1} [\Delta G_n - \Delta G_n(a)] = o(1)$ as $a \rightarrow \infty$.

Then

$$\begin{aligned}
 H_n &\equiv \int_a^b g \, dF_n - \int_a^b g \, dF \\
 &= \sum_{i=1}^k \left\{ \left[\int_{x_i}^{x_{i+1}} g(x) dF_n(x) - \int_{x_i}^{x_{i+1}} g(x_i) dF_n(x) \right] \right. \\
 &\quad + \left[\int_{x_i}^{x_{i+1}} g(x_i) dF_n(x) - \int_{x_i}^{x_{i+1}} g(x_i) dF(x) \right] \\
 &\quad \left. + \left[\int_{x_i}^{x_{i+1}} g(x_i) dF(x) - \int_{x_i}^{x_{i+1}} g(x) dF(x) \right] \right\} \\
 &= \sum_{i=1}^k \left\{ \int_{x_i}^{x_{i+1}} [g(x) - g(x_i)] dF_n(x) + \int_{x_i}^{x_{i+1}} [g(x_i) - g(x)] dF(x) \right. \\
 &\quad \left. + g(x_i) [F_n(x_{i+1}) - F_n(x_i) - F(x_{i+1}) + F(x_i)] \right\}.
 \end{aligned}$$

Hence,

$$|H_n| \leq \varepsilon + \varepsilon + \sum_{i=1}^k |g(x_i)| |F_n(x_{i+1}) - F_n(x_i) - F(x_{i+1}) + F(x_i)| \rightarrow 2\varepsilon$$

as $n \rightarrow \infty$. Since ε is arbitrary, (1) follows. \square

Lemma 3. (i) If $\{G_n, n \geq 0\}$ are finite, nondecreasing functions on $(-\infty, \infty)$ with $\lim_{n \rightarrow \infty} G_n(x) = G_0(x)$ for $x \in$ some dense subset D of $(-\infty, \infty)$, then $G_n \xrightarrow{\text{w}} G_0$.

(ii) Let $\{F_n, n \geq 1\}$ be d.f.s with $F_n \xrightarrow{\text{w}} F_0$ and g a nonnegative continuous function on $(-\infty, \infty)$. For $n \geq 0$, $a \in C(F_0)$, and $x \in [-\infty, \infty]$, define

$$G_n(x) = \int_a^x g \, dF_n.$$

Then G_n is finite and nondecreasing on $(-\infty, \infty)$, $n \geq 0$, and

- (a) $G_n \xrightarrow{\text{w}} G_0$,
- (b) $\liminf \int_a^\infty g \, dF_n \geq \int_a^\infty g \, dF_0$, $\limsup \int_{-\infty}^a g \, dF_n \geq \int_{-\infty}^a g \, dF_0$.

PROOF. (i) If $x \in C(G_0)$ and $\varepsilon > 0$, choose $\delta > 0$ such that $|G_0(y) - G_0(x)| < \varepsilon$ for $|x - y| < \delta$. Select $x_i \in D$, $i = 1, 2$, with $x - \delta < x_1 < x < x_2 < x + \delta$. Then

$$G_0(x) - \varepsilon < G_0(x_1) \leftarrow G_n(x_1) \leq G_n(x) \leq G_n(x_2) \rightarrow G_0(x_2) < G_0(x) + \varepsilon,$$

whence $\lim G_n(x) = G_0(x)$ for $x \in C(G_0)$. Apropos of (ii), note that by (the Helly-Bray) Lemma 2 and part (i) of the current lemma $G_n \xrightarrow{\text{w}} G_0$. Then (b) follows directly from Lemma 1(i). \square

Definition. If $\{F_n, n \geq 1\}$ is a sequence of d.f.s on R , and g is a real, continuous function on $(-\infty, \infty)$, then g is called **uniformly integrable (u.i.) relative to $\{F_n\}$** if

$$\sup_{n \geq 1} \int_{\{|y| \geq a\}} |g(y)| dF_n(y) = o(1) \quad \text{as } a \rightarrow \infty. \quad (2)$$

Furthermore, $\{F_n, n \geq 1\}$ is said to be **tight** if the function 1 is u.i. relative to $\{F_n\}$.

Clearly, (i) if f and g are u.i. relative to $\{F_n\}$, so are f^+ and $af + bg$ for any finite real numbers a, b . (ii) if f, g are continuous, $|f| \leq |g|$, and g is u.i. relative to $\{F_n\}$, so is f .

Thus, in the case of d.f.s, Lemma 1(iii) may be rephrased as follows:

If the d.f.s $F_n \xrightarrow{\text{w}} F$, then $F_n \xrightarrow{\text{c}} F$ iff $\{F_n\}$ is tight iff every bounded continuous function g is u.i. relative to $\{F_n\}$.

Theorem 2. *If $\{F_n, n \geq 1\}$ is a sequence of d.f.s on R with $F_n \xrightarrow{\text{w}} F$ and g is a nonnegative, continuous function on $(-\infty, \infty)$ for which $\int_{-\infty}^{\infty} g dF_n < \infty$, $n \geq 1$, then*

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} g dF_n = \int_{-\infty}^{\infty} g dF < \infty \quad (3)$$

iff g is u.i. relative to $\{F_n\}$.

PROOF. For any $a \in C(F)$ and $x \in [-\infty, \infty]$ define $G_n(x) = \int_a^x g dF_n$, $G(x) = \int_a^x g dF$. By Lemma 3, $G_n \xrightarrow{\text{w}} G$. If g is u.i. relative to $\{F_n\}$, then (iii)(γ) of Lemma 1 holds, whence by (iii)(α) thereof

$$G_n(\pm \infty) \rightarrow G(\pm \infty), \quad \text{finite},$$

which is virtually (3).

Conversely, if (3) holds, so does (iii)(β) of Lemma 1 for G, G_n as defined, whence by (iii)(γ), g is u.i. relative to $\{F_n\}$. \square

Corollary 5. *If the d.f.s $F_n \xrightarrow{\text{w}} F$ and g is a continuous function on $(-\infty, \infty)$ which is u.i. relative to $\{F_n\}$, then (3) holds and $\int_{-\infty}^{\infty} |g| dF < \infty$.*

Corollary 6 (Helly–Bray Theorem). (i) *If the d.f.s $F_n \xrightarrow{\text{c}} F$ and g is a bounded, continuous function on $(-\infty, \infty)$, then*

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} g dF_n = \int_{-\infty}^{\infty} g dF. \quad (4)$$

(ii) *If the d.f.s $F_n \xrightarrow{\text{w}} F$ and g is continuous on $(-\infty, \infty)$ with $\lim_{y \rightarrow \pm \infty} g(y) = 0$, then Eq. (4) holds.*

PROOF. Since $|g| \leq M < \infty$ and 1 is u.i. relative to $\{F_n\}$ by (iii)(γ) of Lemma 1, necessarily g is u.i. relative to $\{F_n\}$, and the conclusion follows from Corollary 5.

In case (ii), for any $\varepsilon > 0$ and sufficiently large a , $|g(y)| < \varepsilon$ for $|y| \geq a$ and so g is u.i. relative to $\{F_n\}$. \square

Corollary 7. If d.f.s $F_n \xrightarrow{\text{w}} F$ and $\{\int |x|^s dF_n(x), n \geq 1\}$ is a bounded sequence for some $s > 0$, then

- i. $F_n \xrightarrow{\text{s}} F$,
- ii. $\int |x|^r dF_n(x) \rightarrow \int |x|^r dF(x)$, $0 \leq r < s$, and
- iii. $\int x^k dF_n(x) \rightarrow \int x^k dF(x)$, $k = 1, 2, \dots, [s]$, $k \neq s$.

PROOF. This follows from Corollary 5 since for $0 \leq r < s$ and some C in $(0, \infty)$

$$\int_{\{|x| \geq a\}} |x|^r dF_n(x) \leq \frac{1}{a^{s-r}} \int |x|^s dF_n(x) < \frac{C}{a^{s-r}}, \quad n \geq 1. \quad \square$$

The Helly–Bray theorem (Corollary 6(i)) is extremely useful and clearly complements the earlier Helly–Bray lemma.

The notion of a function u.i. relative to d.f.s $\{F_n\}$ is redolent of that of uniform integrability of r.v.s $\{X_n\}$ encountered in Chapter 4. The connection between these is elucidated in

Proposition 2. Let g be a continuous function on $(-\infty, \infty)$ and let $\{X_n\}$ be r.v.s on a probability space (Ω, \mathcal{F}, P) with d.f.s $\{F_n\}$. If g is u.i. relative to $\{F_n\}$, then the r.v.s $\{g(X_n)\}$ are u.i. Conversely, if the r.v.s $\{g(X_n)\}$ are u.i. and either (i) $|g(t)| \rightarrow \infty$ as $|t| \rightarrow \infty$ or (ii) $\{F_n\}$ is tight, then g is u.i. relative to $\{F_n\}$.

PROOF. Throughout, in addition to any other requirements choose $a > 0$ so that $\pm a \in \bigcap_1^\infty C(F_n)$. If g is u.i. relative to $\{F_n\}$ and $\varepsilon > 0$, select a so that for $n \geq 1$ it also satisfies the first inequality of

$$\varepsilon > \int_{\{|y| \geq a\}} |g(y)| dF_n = \int_{\{|X_n| \geq a\}} |g(X_n)| dP \geq \int_{\{|g(X_n)| > b\}} |g(X_n)| dP, \quad (5)$$

where $b = \max\{|g(y)| : |y| \leq a\}$ and the equality holds via Theorem 6.2.4 and Corollary 6.2.1. Thus $\{g(X_n)\}$ are u.i.

Conversely, in case (i), as $a \rightarrow \infty$ there exists $K = K_a \rightarrow \infty$ such that

$$\int_{\{|y| \geq a\}} |g(y)| dF_n = \int_{\{|X_n| \geq a\}} |g(X_n)| dP \leq \int_{\{|g(X_n)| \geq K\}} |g(X_n)| dP,$$

whence u.i. of $\{g(X_n)\}$ implies that of g relative to $\{F_n\}$. Under (ii), for any $\varepsilon > 0$ choose $b > 0$ such that

$$\sup_{n \geq 1} \int_{\{|g(X_n)| > b\}} |g(X_n)| dP < \varepsilon$$

and then select $a > 0$ so that $\sup_{n \geq 1} P\{|X_n| \geq a\} < \varepsilon/b$. Then, for $n \geq 1$

$$\int_{\{|y| \geq a\}} |g(y)| dF_n(y) = \int_{\{|X_n| \geq a\}} |g(X_n)| dP \leq \varepsilon + b P\{|X_n| \geq a\} < 2\varepsilon,$$

whence g is u.i. relative to $\{F_n\}$. \square

Proposition 2 in conjunction with Theorem 2 yields the following improvement of Theorem 4.2.3(i):

Corollary 8. *For some $p > 0$, let $\{X_n, n \geq 1\}$ be \mathcal{L}_p r.v.s on (Ω, \mathcal{F}, P) with $X_n \xrightarrow{d} X_F$. Then $E|X_n|^p \rightarrow E|X_F|^p$, finite iff $\{|X_n|^p, n \geq 1\}$ is u.i.*

If \mathcal{F}^* denotes the class of all d.f.s on R , many distances may be defined on \mathcal{F}^* . One prominent choice is $d^*[F, G] = \sup_{x \in R} |F(x) - G(x)|$ (see Exercise 2). The Lévy distance $d[F, G]$ corresponds to the maximum distance between F and G measured along lines of slope -1 (in contemplating this, draw vertical lines connecting F and also G at any discontinuities) multiplied by the factor $1/\sqrt{2}$. Formally,

$$d[F_n, F] = \inf\{h > 0: F(x - h) - h \leq F_n(x) \leq F(x + h) + h, \text{all } x\}. \quad (6)$$

Theorem 3. *Let $\{F, F_n, n \geq 1\}$ be d.f.s. Then (i) $F_n \xrightarrow{d} F$ iff (ii) $\int g dF_n \rightarrow \int g dF$ for every bounded, continuous function g iff (iii) $d[F_n, F] \rightarrow 0$ iff (iv) $\overline{\lim} F_n(C) \leq F(C)$, $\underline{\lim} F_n(V) \geq F(V)$ for all closed sets C and open sets V , where $F_n(\cdot)$, $F(\cdot)$ are the probability measures determined by F_n , F respectively.*

PROOF. That (i) implies (ii) is the Helly–Bray theorem, Corollary 6. To show that (i) implies (iii), for any $\varepsilon > 0$ choose $a, b \in C(F)$ such that $\varepsilon/2$ exceeds both $F(a)$ and $1 - F(b)$ and then select $a_j \in C(F)$, $0 \leq j \leq m$, with $a_0 = a < a_1 < \dots < a_m = b$ and $|a_j - a_{j-1}| < \varepsilon/2$, $1 \leq j \leq m$. Determine N_j , $0 \leq j \leq m$, so that $n \geq N_j$ entails $|F_n(a_j) - F(a_j)| < \varepsilon/2$ and set $N = \max_{0 \leq j \leq m} N_j$. Let $n > N$. If $x \leq a_0$,

$$F_n(x) \leq F_n(a_0) < F(a_0) + \frac{\varepsilon}{2} \leq \varepsilon + F(x + \varepsilon),$$

$$F_n(x) \geq 0 > F(x) - \frac{\varepsilon}{2} \geq F(x - \varepsilon) - \varepsilon,$$

and, analogously, $F(x - \varepsilon) - \varepsilon \leq F_n(x) \leq F(x + \varepsilon) + \varepsilon$ for $x \geq a_m$. Moreover, if $a_{j-1} \leq x \leq a_j$ for some j , $1 \leq j \leq m$,

$$F_n(x) \leq F_n(a_j) < F(a_j) + \frac{\varepsilon}{2} \leq F(x + \varepsilon) + \varepsilon,$$

$$F_n(x) \geq F_n(a_{j-1}) > F(a_{j-1}) - \frac{\varepsilon}{2} \geq F(x - \varepsilon) - \varepsilon.$$

Combining these, $d[F_n, F] < \varepsilon$ and (iii) follows.

To verify that (iii) implies (i), for any $x_0 \in C(F)$ and $\varepsilon > 0$ choose $\delta > 0$ such that $|x - x_0| \leq \delta$ entails $|F(x) - F(x_0)| < \varepsilon$. Set $h = \min(\varepsilon, \delta)$ and select N so that $d[F_n, F] < h$ when $n \geq N$. Then, for $n \geq N$ from (6).

$$\begin{aligned} F_n(x_0) &\leq F(x_0 + h) + h \leq F(x_0 + \delta) + \varepsilon \leq F(x_0) + 2\varepsilon, \\ F_n(x_0) &\geq F(x_0 - h) - h \geq F(x_0 - \delta) - \varepsilon \geq F(x_0) - 2\varepsilon, \end{aligned}$$

and (i) follows.

To obtain (i) from (ii), retain the choice of x_0, δ, ε ; define

$$h(x) = \begin{cases} 1, & x \leq x_0 - \delta \\ \frac{x_0 - x}{\delta}, & x_0 - \delta \leq x \leq x_0 \\ 0, & x > x_0, \end{cases}$$

and set $h_1(x) = h(x)$, $h_2(x) = h(x - \delta)$. For any d.f. G

$$\begin{aligned} G(x_0 - \delta) &= \int_{-\infty}^{x_0 - \delta} h_1 \, dG \leq \int_{-\infty}^{\infty} h_1 \, dG \leq \int_{-\infty}^{x_0} \, dG \leq G(x_0) \\ G(x_0) &= \int_{-\infty}^{x_0} h_2 \, dG \leq \int_{-\infty}^{\infty} h_2 \, dG \leq \int_{-\infty}^{x_0 + \delta} \, dG \leq G(x_0 + \delta), \end{aligned}$$

and so, taking $G = F$ and then $G = F_n$,

$$\begin{aligned} \int h_1 \, dF - \int h_1 \, dF_n &\geq F(x_0 - \delta) - F_n(x_0) \geq F(x_0) - F_n(x_0) - \varepsilon, \\ \int h_2 \, dF - \int h_2 \, dF_n &\leq F(x_0 + \delta) - F_n(x_0) \leq F(x_0) - F_n(x_0) + \varepsilon, \end{aligned}$$

whence via (ii), for all sufficiently large n

$$|F_n(x_0) - F(x_0)| \leq \varepsilon + \sum_{i=1}^2 \left| \int h_i \, dF - \int h_i \, dF_n \right| < 3\varepsilon.$$

It remains to establish the equivalence of (i) and (iv). Under the latter, for any $a, x \in C(F)$ with $a < x$

$$\begin{aligned} \varliminf_{n \rightarrow \infty} F_n(x) &\geq \varliminf_{n \rightarrow \infty} [F_n(x) - F_n(a)] \geq \varliminf_{n \rightarrow \infty} F_n((a, x)) \geq F((a, x)) \\ &= F(x) - F(a) \xrightarrow{x \rightarrow -a} F(x), \\ 1 - \varlimsup_{n \rightarrow \infty} F_n(a) &= \varlimsup_{n \rightarrow \infty} [1 - F_n(a)] \geq \varlimsup_{n \rightarrow \infty} [F_n(x) - F_n(a)] \\ &\geq F(x) - F(a) \xrightarrow{x \rightarrow a} 1 - F(a). \end{aligned}$$

Hence,

$$F(x) \geq \varlimsup_{n \rightarrow \infty} F_n(x) \geq \varliminf_{n \rightarrow \infty} F_n(x) \geq F(x) \quad \text{for } x \in C(F).$$

PROOF. Note at the outset that if H is any nondegenerate d.f. with $H(a_1x + b_1) = H(a_2x + b_2)$ for all x , where a_j and b_j are real, $j = 1, 2$, and $a_1 + a_2 \neq 0$, then $a_1 = a_2$, $b_1 = b_2$. For if $a_1x + b_1 \neq a_2x + b_2$, then

$$x' = \frac{1}{2}[(a_1 + a_2)x + b_1 + b_2]$$

cannot be a point of increase of $H(x)$. Thus, if x' is a point of increase, $a_1x + b_1 = a_2x + b_2$ and, since there are two distinct points of increase, necessarily $a_1 = a_2$, $b_1 = b_2$.

The hypothesis (and proof) is easily couched in terms of r.v.s, namely, $X_n \xrightarrow{d} X_F$ (nondegenerate) and $Y_n = a_n^{-1}(X_n - b_n) \xrightarrow{d} Y_G$ (nondegenerate), where $a_n > 0$. Let a be a finite or infinite limit point of $\{a_n\}$ and $\{n'\}$ a subsequence of the positive integers $\{n\}$ with $a_{n'} \rightarrow a$; let b be a finite or infinite limit point of $\{b_{n'}\}$ and $\{n''\}$ a subsequence of $\{n'\}$ with $b_{n''} \rightarrow b$.

If $a = \infty$, $X_{n''}/a_{n''} \xrightarrow{P} 0$, whence by Theorem 8.1.1 $-b_{n''}/a_{n''} = Y_{n''} - (X_{n''}/a_{n''}) \xrightarrow{d} Y_G$, which is impossible since Y_G is nondegenerate. Likewise, $a = 0$ is precluded since this would entail $b_{n''} = X_{n''} - a_{n''} Y_{n''} \xrightarrow{d} X_F$. Thus, $0 < a < \infty$ and, since $(a_{n''} - a)Y_{n''} \xrightarrow{P} 0$,

$$aY_{n''} + b_{n''} = X_{n''} - (a_{n''} - a)Y_{n''} \xrightarrow{d} X_F,$$

which ensures b finite in view of $aY_{n''} \xrightarrow{d} aY_G$. Thus, by Corollary 8.1.3

$$Y_G \xrightarrow{d} Y_{n''} = a_{n''}^{-1}(X_{n''} - b_{n''}) \xrightarrow{d} a^{-1}(X_F - b)$$

or $G(x) = F(ax + b)$. By the remark at the outset, no other subsequence of $\{b_{n''}\}$ can have an alternative limit point b' , that is, $\lim b_{n''} = b$. Analogously, if $\{n^*\}$ is a subsequence of $\{n\}$ with $a_{n^*} \rightarrow a^*$, $b_{n^*} \rightarrow b^*$, the prior reasoning requires $F(ax + b) = F(a^*x + b^*)$, which, by the initial comment, entails $a = a^*$, $b = b^*$. Thus $a_n \rightarrow a$, $b_n \rightarrow b$. \square

Corollary 2. If $F_n(a_nx + b_n) \xrightarrow{c} F(x)$, $F_n(\alpha_nx + \beta_n) \xrightarrow{c} G(x)$, where F, G are nondegenerate, $a_n, \alpha_n > 0$, then F and G are related by (1) and $\alpha_n/a_n \rightarrow a$, $(\beta_n - b_n)/a_n \rightarrow b$. In particular, if $G = F$, $\alpha_n \sim a_n$, $\beta_n - b_n = o(a_n)$.

As seen in Section 6.3, the class of distribution functions on R is closed under the convolution operation $*$. It is likewise closed under the more general operation of mixture (Exercise 3).

EXERCISES 8.2

1. Give an example showing that Theorem 3 is invalid without the nondegeneracy hypothesis.
2. Verify that the class \mathcal{F}^* of all d.f.s on R is a complete metric space under the Levy distance.
3. Let Λ be a Borel subset of \tilde{R}^m where $\tilde{R} = (-\infty, \infty)$, $m \geq 1$, and for every $\lambda \in \Lambda$ let $F(x; \lambda)$ be a d.f. on R such that $F(x; \lambda)$ is a Borel function on R^{m+1} . If G is any d.f. whose support $\subset \Lambda$, show that $H(x) = \int_{\Lambda} F(x; \lambda) dG(\lambda)$ is a d.f. on R . It is called a

G -mixture of the family $\mathcal{F} = \{F(x; \lambda), \lambda \in \Lambda\}$ or simply a **mixture**. Convolution is the special case $F(x; \lambda) = F(x - \lambda)$, $m = 1$. If E_H and E_{F_λ} denote integration relative to H and $F(x; \lambda)$ respectively, show for any Borel function φ with $E_H|\varphi| < \infty$ that $E_H[\varphi] = \int E_{F_\lambda}[\varphi]dG(\lambda)$.

4. Let Q be the planar region bounded by the quadrilateral with vertices $(0, 0), (1, 1), (0, \frac{1}{2}), (\frac{1}{2}, 1)$ and the triangle with vertices $(\frac{1}{2}, 0), (1, 0), (1, \frac{1}{2})$. If X, Y are jointly uniformly distributed on Q , that is, P is Lebesgue measure on the Borel subsets of Q , prove that $F_{X+Y} = F_X * F_Y$ despite the nonindependence of X and Y .
5. The convolution of two discrete distributions $F_j, j = 1, 2$, is discrete, and if the support $S(F_j)$ contains n_j points, $j = 1, 2$, $S(F_1 * F_2)$ contains at most $n_1 \cdot n_2$ and at least $n_1 + n_2 - 1$ points.
6. The convolution of two d.f.s is absolutely continuous if at least one of the components is absolutely continuous. The converse is false (Exercise 8.4.6).
7. If $\{X_n, n \geq 1\}$ are i.i.d. with a uniform distribution on $[0, 1]$, that is, $F(x) = x$ for $0 \leq x \leq 1$, show that $n(\max_{1 \leq i \leq n} X_i - 1) \xrightarrow{d} X_G$.
8. The Lévy **concentration function** of a d.f. F is defined by

$$Q(a) = \sup_{x \in \mathbb{R}} [F((x + a)_+) - F(x)], \quad a \geq 0.$$

Demonstrate that if $F = F_1 * F_2$, $Q_1(a/2) \cdot Q_2(a/2) \leq Q(a) \leq \min_{i=1, 2} Q_i(a)$, and deduce that $F_1 * F_2$ is continuous iff F_1 or F_2 is continuous.

9. Let F_i be a discrete d.f. with maximum jump q_i , $i \geq 1$, and suppose that $G_n = F_1 * F_2 * \dots * F_n \xrightarrow{d} G$. Prove that if G is continuous, $\prod_{i=1}^n q_i = o(1)$.
10. If f_1, f_2 are densities, their convolution is defined by

$$f = f_1 * f_2 = \int f_1(x - y)f_2(y)dy.$$

Verify that if F_j is an absolutely continuous d.f. with density f_j , $j = 1, 2$, then $F = F_1 * F_2$ is absolutely continuous with density $f = f_1 * f_2$.

11. (Chow–Robbins) Prove that if F is a d.f. with $F(0) = 0$, then $G(x) = \prod_{i=1}^x F(x + n)$ is a d.f. iff $\int x dF(x) < \infty$. More generally, G has a finite k th moment iff F has a finite $(k + 1)$ st moment.
12. If X and Y are independent r.v.s such that $X + Y$ and X have identical d.f.s, then $Y = 0$, a.c. *Hint:* Employ the Lévy concentration function of Exercise 8.

8.3 Characteristic Functions, Inversion Formula, Lévy Continuity Theorem

Any r.v. X on a probability space (Ω, \mathcal{F}, P) has an associated characteristic function (abbreviated c.f.) $\phi_X(t)$ defined for all real t by

$$\phi(t) = \phi_X(t) = E e^{itX} = E \cos tX + i E \sin tX. \quad (1)$$

On the other hand, any d.f. F on \mathbb{R} has a corresponding Fourier-Stieltjes transform $\varphi_F(t)$ defined for all real t by

$$\varphi(t) = \varphi_F(t) = \int_{-\infty}^{\infty} e^{itx} dF(x) = \int_{-\infty}^{\infty} \cos tx dF(x) + i \int_{-\infty}^{\infty} \sin tx dF(x), \quad (1')$$

where the integrals are of the improper Riemann-Stieltjes variety.

When $F = F_X$ is the d.f. of a r.v. X on some probability space (Ω, \mathcal{F}, P) , then according to Theorem 6.2.4 the two integrals coincide. Thus, the probability integral of (1) may be envisaged as the Riemann-Stieltjes (or Lebesgue-Stieltjes) integral of (1') with $F = F_X$ and vice versa in view of oft-repeated comments about construction of a r.v. X on (Ω, \mathcal{F}, P) with a preassigned d.f. F . Thus, the term c.f. may also be used in conjunction with d.f.

Clearly, $\varphi(0) = 1$, $|\varphi(t)| \leq 1$, $\varphi(t) = \overline{\varphi(-t)}$, where $\bar{\varphi}$ signifies the complex conjugate of φ , and φ is uniformly continuous on $(-\infty, \infty)$ since by the dominated convergence theorem

$$|\varphi(t+h) - \varphi(t)| = |\mathbb{E} e^{ithX} (e^{ihX} - 1)| \leq \mathbb{E} |e^{ihX} - 1| \xrightarrow{h \rightarrow 0} 0$$

independently of t . Moreover, for any constants a, b

$$\varphi_{aX+b}(t) = e^{ibt} \varphi_X(at). \quad (2)$$

The importance of c.f.s in probability theory stems on the one hand from the one-to-one correspondence between c.f.s and d.f.s, stated as a corollary to the Lévy inversion formula below, and on the other from the ease of operation with c.f.s, due largely to the parallelism between convolution of d.f.s and multiplication of (the corresponding) c.f.s as formulated in Theorem 2 (below). In delving into probabilistic problems one may, therefore, operate interchangeably with d.f.s or their c.f.s.

Theorem 1 (Lévy Inversion Formula). *If X is a r.v. with c.f. φ , then for $-\infty < a < b < \infty$*

$$\begin{aligned} & \lim_{C \rightarrow \infty} \frac{1}{2\pi} \int_{-C}^C \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt \\ &= \mathbb{P}\{a < X < b\} + \frac{\mathbb{P}\{X = a\} + \mathbb{P}\{X = b\}}{2}. \end{aligned} \quad (3)$$

If $a, b \in C(F)$, where $F = F_X$, the right side of (3) reduces to $F(b) - F(a)$. The integrand of (3) is defined at $t = 0$ by continuity and is not, in general, absolutely integrable over $(-\infty, \infty)$.

PROOF. Set

$$I(C) = \frac{1}{2\pi} \int_{-C}^C \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt = \frac{1}{2\pi} \int_{-C}^C \frac{e^{-ita} - e^{-itb}}{it} \mathbb{E} e^{itX} dt.$$

Since $[(e^{-ita} - e^{-itb})/it]e^{itX}$ is bounded for all ω and t , by Fubini's theorem

$$\begin{aligned} I(C) &= \frac{1}{2\pi} E \int_{-C}^C \frac{e^{it(X-a)} - e^{it(X-b)}}{it} dt \\ &= \frac{1}{\pi} E \int_0^C \frac{\sin t(X-a) - \sin t(X-b)}{t} dt \\ &= \frac{1}{\pi} E \left[\int_0^{C(X-a)} \frac{\sin t}{t} dt - \int_0^{C(X-b)} \frac{\sin t}{t} dt \right] = E J_C(X), \end{aligned}$$

where

$$J_C(u) = \frac{1}{\pi} \int_{C(u-b)}^{C(u-a)} \frac{\sin t}{t} dt \xrightarrow{C \rightarrow \infty} \begin{cases} 1, & a < u < b \\ \frac{1}{2}, & u = a, b \\ 0, & u < a, u > b. \end{cases}$$

Since $|J_C(u)| \leq 2$ for all $-\infty < u, C < \infty$, by dominated convergence

$$\begin{aligned} \lim_{C \rightarrow \infty} I(C) &= \lim_{C \rightarrow \infty} E J_C(X) = E \lim_{C \rightarrow \infty} J_C(X) \\ &= E[\frac{1}{2}I_{\{X=a \text{ or } b\}} + I_{\{a < X < b\}}] = \frac{P\{X=a\} + P\{X=b\}}{2} + P\{a < X < b\}. \end{aligned}$$

□

Corollary 1. *There is a one-to-one correspondence between d.f.s on R and their c.f.s.*

PROOF. The very definition (1') shows that identical d.f.s have identical c.f.s, while the converse follows from

$$F(b) = \lim_{a \rightarrow -\infty} \lim_{C \rightarrow \infty} \frac{1}{2\pi} \int_{-C}^C \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt, \quad b \in C(F). \quad \square$$

Corollary 2. *If $\int_{-\infty}^{\infty} |\varphi(t)| dt < \infty$, then for $-\infty < a < b < \infty$*

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt &= P\{a < X < b\} + \frac{P\{X=a\} + P\{X=b\}}{2} \\ &= F(b) - F(a), \end{aligned} \quad (3')$$

where the corresponding d.f. F is absolutely continuous with a bounded, continuous density

$$f(x) = F'(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt. \quad (4)$$

PROOF. In view of the hypothesis, (3) transcribes as the first equality in (3'), whence, letting $b \rightarrow a+$, Lebesgue's dominated convergence theorem ensures $P\{X=a\} = 0$. Hence, F is continuous, yielding the second equality. Since for b approaching a the integrand of (3') divided by $b-a$ tends to $e^{-ita}\varphi(t)$, (4) follows by dominated convergence. Boundedness of f is apparent from

(4), and continuity of f follows once more by dominated convergence. To verify absolute continuity, note that via Fubini's theorem for $-\infty < a < b < \infty$

$$\begin{aligned} \int_a^b f(x)dx &= \frac{1}{2\pi} \int_a^b dx \int_{-\infty}^{\infty} e^{-itx} \varphi(t)dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(\int_a^b e^{-itx} dx \right) \varphi(t)dt \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t)dt = F(b) - F(a) \end{aligned}$$

by (3'). Since f is continuous, $f(x) \geq 0$ on $(-\infty, \infty)$ and, letting $b \rightarrow \infty$, $a \rightarrow -\infty$, $f \in \mathcal{L}_1(-\infty, \infty)$ and $\|f\| = 1$. \square

Theorem 2. If F_i is a d.f. with corresponding c.f. φ_i , $i = 0, 1, 2$, then $F_0 = F_1 * F_2$ iff $\varphi_0 = \varphi_1 \cdot \varphi_2$.

PROOF. Suppose $F_0 = F_1 * F_2$. Since e^{itz} is bounded and continuous, by (11) of Section 6.3

$$\begin{aligned} \varphi_0(t) &= \int e^{itz} d(F_1 * F_2)(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{itz+x+y} dF_1(x)dF_2(y) \\ &= \varphi_1(t) \cdot \varphi_2(t). \end{aligned}$$

Conversely, if $\varphi_0 = \varphi_1 \cdot \varphi_2$ and $F = F_1 * F_2$ has c.f. φ , then by the portion of the theorem just proved $\varphi = \varphi_1 \cdot \varphi_2 = \varphi_0$, whence by Corollary 1, $F_0 = F = F_1 * F_2$. \square

Corollary 3. If X_1, X_2 are independent r.v.s on some probability space with c.f.s $\varphi_{X_1}, \varphi_{X_2}$, then the c.f. of their sum is

$$\varphi_{X_1+X_2} = \varphi_{X_1} \cdot \varphi_{X_2}. \quad (5)$$

PROOF. This is an immediate consequence of Theorem 2 and Theorem 6.3.4. \square

Corollary 4. Convolution is associative and commutative and so the class \mathcal{F}^* of all d.f.s on R is an Abelian semigroup.

What is the analogue vis-à-vis c.f.s of complete convergence of d.f.s? This and more is answered by

Theorem 3 (Levy continuity theorem). Let $\{F_n\}$ be a sequence of d.f.s on R with corresponding c.f.s $\{\varphi_n\}$. If $F_n \xrightarrow{s} F$, then $\lim_{n \rightarrow \infty} \varphi_n(t) = \varphi_F(t)$ uniformly in $|t| \leq T$ for all $T > 0$. Conversely, if $\varphi_n(t)$ converges to a limit $g(t)$ on $(-\infty, \infty)$ which is continuous at $t = 0$, then g is the c.f. of some d.f. F and $F_n \xrightarrow{s} F$.

5. Prove that a real-valued c.f. ϕ satisfies $\phi(t+h) \leq \phi(h)$ for $t > 0$ and sufficiently small $h > 0$.
6. Verify that the c.f. corresponding to a uniform distribution on $(-\alpha, \alpha)$ is $\phi_\alpha(t) = (\sin \alpha t)/\alpha$, $\alpha > 0$, and that $\lim \phi_n(t)$ exists but F_n fails to a d.f.
7. Find the c.f. ϕ_X if $P\{X = \alpha\} = P\{X = -\alpha\} = \frac{1}{2}$. Show by iteration of an elementary trigonometric identity that

$$\frac{\sin t}{t} = \frac{\sin t/2^n}{t/2^n} \prod_{j=1}^n \cos \frac{t}{2^j} \rightarrow \prod_{j=1}^{\infty} \cos \frac{t}{2^j}$$

and utilize this to state a result on convergence in law of sums of independent r.v.s.

8. If ϕ is the c.f. of the d.f. F , prove that

$$\lim_{C \rightarrow \infty} \frac{1}{2C} \int_{-C}^C e^{-itx} \phi(t) dt = F(x+) - F(x),$$

and if X is an integer valued r.v., show that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-itj} \phi_X(t) dt = P(X = j).$$

Give an analogous formula for $S_n = \sum_{i=1}^n X_i$, where $\{X_i\}$ are i.i.d. integer-valued r.v.s.

9. Prove that $\lim_{C \rightarrow \infty} (1/2C) \int_{-C}^C |\phi(t)|^2 dt = \sum_{x \in R} [F(x+) - F(x)]^2$. Hint: if X_1, X_2 are independent with common d.f. F , apply Exercise 8 at $x = 0$ to $X_1 - X_2$.
10. Prove that

$$F(x+) + F(x-) = 1 - \frac{1}{\pi} \int \frac{e^{-itx} \phi(t)}{it} dt$$

where

$$\int = \lim_{\substack{\epsilon \rightarrow 0 \\ \epsilon \rightarrow \infty}} \left(\int_{-\epsilon}^{-\pi} + \int_{\pi}^{\epsilon} \right).$$

and hence that

$$F(x) = \frac{1}{2} - \frac{1}{2\pi} \int \frac{e^{-itx} \phi(t)}{it} dt$$

for F continuous.

11. Utilize c.f.s to give an alternative proof of Corollary 8.1.10.
12. Prove for any d.f. with c.f. ϕ that for all real x and $h > 0$

$$\frac{1}{2h} \int_x^{x+2h} F(y) dy - \frac{1}{2h} \int_{x-h}^x F(y) dy = \frac{1}{\pi} \int_{-\infty}^x \left(\frac{\sin u}{u} \right)^2 e^{-iux/h} \phi(u/h) du.$$

Hint: Apply the Levy inversion formula to $F * G_h$, where G_h is the uniform distribution on $[-h, h]$.

13. Utilize Exercise 12 to prove the converse portion of Theorem 3.

14. Let $H_1(x) = \int_{\Lambda} F(x; \lambda) dG_1(\lambda)$ be a G_1 -mixture (Exercise 8.2.3) of the additively closed family $\mathcal{F} = \{F(x; \lambda), \lambda \in \Lambda \subset R^m\}$ (Exercise 8.4.5). Verify that the convolution $H_1 * H_2$ is a $(G_1 * G_2)$ -mixture of \mathcal{F} . Hint: Utilize Exercise 3.
15. If, in Exercise 14, $m = 1$ and $\Lambda = [0, \infty)$ or $[0, 1, 2, \dots]$ or {nonnegative rationals}, then the mapping of $\mathcal{G} = \{G: G(\Lambda) = 1\}$ onto $\mathcal{H} = \{H: H(x) = \int_{\Lambda} F(x; \lambda) dG(\lambda), G \in \mathcal{G}\}$ is one-to-one. Then \mathcal{H} is said to be **identifiable**. Hint: $\psi(z; G) = \int_{\Lambda} z^{\lambda} dG(\lambda)$ is analytic in $0 < |z| < 1$ and $\varphi(t; \lambda) = \varphi^{\lambda}(t)$, $\lambda \in \Lambda$ where $\varphi(t) = \varphi(t; 1)$.

8.4 The Nature of Characteristic Functions, Analytic Characteristic Functions, Cramér–Lévy Theorem

In view of the crucial importance of c.f.s in probability theory, it is desirable to amass some information concerning their nature. A first step in this direction is a tally of the more popular d.f.s and the corresponding c.f.s.

Distribution	Density	Support	c.f.
Degenerate		{x}	e^{iz}
Symmetric Bernoulli	$\frac{1}{2}$	{1, -1}	$\cos t$
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$	$x \in \{0, 1, \dots, n\}$	$(pe^t + 1 - p)^n, 0 < p < 1$
Poisson	$\lambda^x \frac{e^{-\lambda}}{x!}$	$x \in \{0, 1, \dots\}$	$\exp\{\lambda(e^t - 1)\}, \lambda > 0$
Normal	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\}$	$(-\infty, \infty)$	$e^{i\theta t - \sigma^2 t^2/2}, \sigma > 0$
Symmetric Uniform	$\frac{1}{2x}$	$[-x, x]$	$\frac{\sin xt}{xt}, x > 0$
Triangular	$\frac{1}{x} \left(1 - \frac{ x }{x}\right)$	$[-x, x]$	$\frac{2(1 - \cos xt)}{x^2 t^2}, x > 0$
Inverse Triangular	$\frac{1 - \cos \alpha x}{\pi \alpha x^2}$	$(-\infty, \infty)$	$\left[1 - \frac{ t }{\alpha}\right]^+, \alpha > 0$
Gamma	$\frac{\alpha^\lambda}{\Gamma(\lambda)} x^{\lambda-1} e^{-\alpha x}$	$(0, \infty)$	$\left(1 - \frac{it}{\alpha}\right)^{-\lambda}, \lambda > 0, \alpha > 0$
Cauchy	$\frac{\alpha}{\pi(\alpha^2 + x^2)}$	$(-\infty, \infty)$	$e^{-\alpha t }, \alpha > 0$

Characteristic functions and moments are intimately related, as will be seen via the preliminary

Lemma 1. For $t \in (-\infty, \infty)$ and nonnegative integers n

$$\begin{aligned} e^t - \sum_{j=0}^n \frac{(it)^j}{j!} &= \frac{(it)^{n+1}}{n!} \int_0^1 e^{itu} (1-u)^n du \\ &= i^{n+1} \int_0^t dt_{n+1} \int_0^{t_{n+1}} dt_n \cdots \int_0^{t_2} e^{itu} dt_1, \end{aligned} \quad (1)$$

and for any δ in $[0, 1]$

$$\left| e^t - \sum_{j=0}^n \frac{(it)^j}{j!} \right| \leq \frac{2^{1-\delta} |t|^{n+\delta}}{(1+\delta)(2+\delta) \cdots (n+\delta)}, \quad (2)$$

where the denominator of the right side of (2) is unity for $n = 0$.

PROOF. Since for $n \geq 1$

$$\frac{(it)^{n+1}}{n!} \int_0^1 e^{itu} (1-u)^n du = \frac{-(it)^n}{n!} + \frac{(it)^n}{(n-1)!} \int_0^1 e^{itu} (1-u)^{n-1} du,$$

the first part of (1) follows by summing while the second is obtained inductively.

To prove (2), let I_n denote the left side of (1). Since $|e^t - 1| = 2|\sin t/2| \leq 2^{1-\delta}|t|^\delta$ for $0 \leq \delta \leq 1$, let $n \geq 1$, whence from (1)

$$\begin{aligned} |I_n| &\leq \int_0^{|t|} \int_0^{t_{n+1}} \cdots \int_0^{t_2} |e^{itu} - 1| dt_2 \cdots dt_{n+1} \\ &\leq 2^{1-\delta} \int_0^{|t|} \int_0^{t_{n+1}} \cdots \int_0^{t_2} t_2^\delta dt_2 \cdots dt_{n+1} = \frac{2^{1-\delta} |t|^{n+\delta}}{(1+\delta) \cdots (n+\delta)}. \quad \square \end{aligned}$$

Corollary 1. $|e^z - 1| \leq (e^{|z|} - 1)$ for any complex z .

PROOF. (2) with $\delta = 1$, $n = 0$ yields the bound $2(e^{|z|} - 1)$. However, directly

$$|e^z - 1| = \left| \sum_1^\infty z^j/j! \right| \leq \sum_1^\infty |z|^j/j! = e^{|z|} - 1. \quad \square$$

Theorem 1. If X is a r.v. with c.f. φ and $E|X|^{n+\delta} < \infty$ for some nonnegative integer n and some δ in $[0, 1]$, then φ has continuous derivatives $\varphi^{(k)}$ of orders $k \leq n$ and

$$\varphi^{(k)}(t) = i^k E X^k e^{itX}, \quad \varphi^{(k)}(0) = i^k E X^k, \quad 1 \leq k \leq n, \quad (3)$$

$$\varphi(t) = \sum_{j=0}^n \frac{(it)^j E X^j}{j!} + O(|t|^{n+\delta}), \quad O(|t|^{n+\delta}) \leq \frac{2^{1-\delta} E|X|^{n+\delta} |t|^{n+\delta}}{(1+\delta) \cdots (n+\delta)}, \quad (4)$$

$$\varphi(t) = \sum_{j=0}^n \frac{(it)^j E X^j}{j!} + o(|t|^n) \quad \text{as } t \rightarrow 0. \quad (5)$$

Conversely, if $\varphi^{(2k)}(0)$ exists and is finite for some $k = 1, 2, \dots$, then

$$\mathbb{E} X^{2k} < \infty.$$

PROOF. (4) follows easily from (2). To prove (3), note that via (1)

$$e^{itX} - \sum_{j=0}^{k-1} \frac{(itX)^j}{j!} = (iX)^k \int_0^t \int_0^{t_k} \cdots \int_0^{t_2} e^{it_1 X} dt_1 \cdots dt_k, \quad (6)$$

so that

$$\psi_k(t) \equiv \varphi(t) - \sum_{j=0}^{k-1} \frac{(it)^j \mathbb{E} X^j}{j!} = i^k \mathbb{E} X^k \int_0^t \int_0^{t_k} \cdots \int_0^{t_2} e^{it_1 X} dt_1 \cdots dt_k$$

and

$$\frac{\psi_k(t+h) - \psi_k(t)}{h} = \mathbb{E}(iX)^k \frac{1}{h} \int_t^{t+h} \int_0^{t_k} \cdots \int_0^{t_2} e^{it_1 X} dt_1 \cdots dt_k.$$

Since

$$\left| \frac{1}{h} \int_t^{t+h} \int_0^{t_k} \cdots \int_0^{t_2} e^{it_1 X} dt_1 \cdots dt_k \right| \leq (|t| + |h|)^{k-1} = O(1) \quad \text{as } h \rightarrow 0,$$

by the dominated convergence theorem, for $1 \leq k \leq n$

$$\psi'_k(t) = i^k \mathbb{E} X^k \int_0^t \int_0^{t_{k-1}} \cdots \int_0^{t_2} e^{it_1 X} dt_1 \cdots dt_{k-1} \quad (7)$$

and, in particular,

$$\varphi'(t) = \psi'_1(t) = i \mathbb{E} X e^{itX}.$$

Repeating the previous argument, it follows via (7) that for $2 \leq k \leq n$

$$\psi_k^{(2)}(t) = \frac{d^2 \psi_k(t)}{dt^2} = i^k \mathbb{E} X^k \int_0^t \int_0^{t_{k-2}} \cdots \int_0^{t_2} e^{it_1 X} dt_1 \cdots dt_{k-2},$$

and, continuing in this fashion, for $k \leq n$

$$\psi_k^{(n-1)}(t) = i^k \mathbb{E} X^k \int_0^t e^{it_1 X} dt_1,$$

$$\varphi^{(n)}(t) = \psi_k^{(n)}(t) = i^n \mathbb{E} X^n e^{itX},$$

which is (3). Now by (3),

$$\varphi^{(n)}(t+h) - \varphi^{(n)}(t) = i^n \mathbb{E} X^n e^{ithX} (e^{ithX} - 1),$$

and since $|e^{ithX} - 1| \leq 2$, once more by dominated convergence

$$\lim_{h \rightarrow 0} \varphi^{(n)}(t+h) = \varphi^{(n)}(t)$$

and φ has a continuous n th derivative, whence (5) follows from Taylor's theorem.

To prove the converse, define for any finite function g for $-\infty < x, h < \infty$

$$\Delta_h^{(1)} g(x) = g(x + h) - g(x - h),$$

$$\Delta_h^{(n)} g(x) = \Delta_h^{(n-1)} g(x + h) - \Delta_h^{(n-1)} g(x - h), \quad n = 2, 3, \dots$$

Then $\Delta_h^{(n)}[ag_1(x) + bg_2(x)] = a\Delta_h^{(n)}g_1(x) + b\Delta_h^{(n)}g_2(x)$ for $-\infty < a, b < \infty$ and

$$(2h)^{-2n} \Delta_h^{(2n)} \varphi(t) = E(2h)^{-2n} \Delta_h^{(2n)} e^{itX} = E e^{itX} \left(\frac{e^{ihX} - e^{-ihX}}{2h} \right)^{2n}.$$

If $g(x)$ is a polynomial of degree n , $\Delta_h g(x)$ is a polynomial of degree at most $n - 1$. Hence $\Delta_h^{(m)} g(x) = 0$ if g is a polynomial of degree $< m$ and, moreover, $\Delta_h^{(2m)} x^{2m} = (2h)^{2m} (2m)!$ Since $\varphi^{(2n)}(0)$ is finite, as $t \rightarrow 0$ (Hardy, 1952, p. 290)

$$\varphi(t) = \varphi(0) + t\varphi'(0) + \dots + \frac{t^{2n}}{(2n)!} [\varphi^{(2n)}(0) + o(1)],$$

so that $\varphi^{(2n)}(0) = \lim_{h \rightarrow 0} (2h)^{-2n} \Delta_h^{(2n)} \varphi(t)|_{t=0}$, whence

$$|\varphi^{(2n)}(0)| = \left| \lim_{h \rightarrow 0} E \left(\frac{e^{ihX} - e^{-ihX}}{2h} \right)^{2n} \right| = \lim_{h \rightarrow 0} E \left(\frac{\sin hX}{h} \right)^{2n} \geq E X^{2n}$$

by Fatou's lemma, completing the proof. \square

In Section 3.3, a r.v. X was defined to be symmetric if X and $-X$ have the same distribution. The property in question belongs to the d.f. and is tantamount to saying that a d.f. is symmetric if $F(x) = 1 - F(-x+)$, $x \in R$.

Proposition 1. A c.f. $\varphi(t)$ is real valued (for real t) iff its d.f. F is symmetric.

PROOF. The proof is most easily couched in terms of r.v.s. If F is symmetric, that is, X and $-X$ are identically distributed,

$$\varphi(t) = \varphi_X(t) = \varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi(t)},$$

so that φ is real. Conversely, if φ is real,

$$\varphi_X(t) = \overline{\varphi_X(t)} = \varphi_X(-t) = \varphi_{-X}(t),$$

and so by the one-to-one correspondence between c.f.s and d.f.s, X and $-X$ are identically distributed, that is, F is symmetric. \square

A function $\varphi(t)$ on the real line is said to be **periodic** with a period t_0 if $\varphi(t + t_0) = \varphi(t)$ for all real t .

Proposition 2. If φ is the c.f. of the d.f. F and $|\varphi(t_0)| = 1$ for some $t_0 \neq 0$, then F is discrete with support S consisting of a subset of points in arithmetic progression and conversely. In this case, $S \subset \{(2\pi j + \theta_0)/t_0, j = 0, \pm 1, \dots\}$ for some real θ_0 and $e^{-it\theta_0/t_0}\varphi(t)$ is a periodic function with period t_0 .

PROOF. Since $|\varphi(t_0)| = 1$, $\varphi(t_0) = e^{i\theta_0}$ for some real θ_0 , whence $1 = e^{-i\theta_0}\varphi(t_0) = \int e^{i(t_0x - \theta_0)} dF(x) = \int \cos(t_0x - \theta_0) dF(x)$. In r.v. parlance,

$$E[1 - \cos(t_0 X - \theta_0)] = 0,$$

necessitating

$$\begin{aligned} & P\{t_0 X - \theta_0 = 2j\pi, j = 0, \pm 1, \dots\} \\ &= P\left\{X = \frac{2j\pi + \theta_0}{t_0}, j = 0, \pm 1, \dots\right\} = 1. \end{aligned}$$

Conversely, the latter entails, setting $p_j = P\{X = (2j\pi + \theta_0)/t_0\} \geq 0$,

$$\varphi(t) = \sum_j p_j e^{it(2j\pi + \theta_0)/t_0} = e^{i\theta_0 t/t_0} \sum_j p_j e^{2j\pi it/t_0}, \quad (8)$$

whence $|\varphi(t_0)| = 1$ and $e^{-i\theta_0 t_0} \varphi(t)$ is periodic with a period t_0 . \square

Corollary 2. A c.f. $\varphi(t)$ satisfies either (i) $|\varphi(t)| < 1$ for all $t \neq 0$, (ii) $|\varphi(t)| \equiv 1$, or (iii) $|\varphi(t)| = 1$ for countably many isolated values of t .

PROOF. It suffices to prove that if $|\varphi(t_n)| = 1$ for $t_n \rightarrow s$ (finite), $0 \neq t_n \neq s$, then $|\varphi(t)| \equiv 1$. Now, via (8) with t_n replacing t_0 and then t_m replacing t_0 , for $k, m, n = 1, 2, \dots$

$$|\varphi(k(t_m - t_n))| = |\varphi(kt_m)| = 1.$$

Since $t_m - t_n \rightarrow 0$, every interval (a, b) contains a point c with $|\varphi(c)| = 1$, whence by continuity $|\varphi(t)| \equiv 1$. \square

Corollary 3. If a c.f. φ satisfies $|\varphi(t_0)| = |\varphi(\alpha t_0)| = 1$ for some $t_0 \neq 0$ and irrational α , then φ is degenerate, that is, $\varphi(t) = e^{i\theta t}$ for some real θ .

PROOF. If φ is nondegenerate, by Proposition 2, $P\{X = (2\pi j_i + \theta_0)/t_0\} > 0$ for some θ_0 and integers $j_1 \neq j_2$. Then for some real θ_1 and integers $k_1 \neq k_2$

$$\frac{2\pi j_i + \theta_0}{t_0} = \frac{2\pi k_i + \theta_1}{\alpha t_0}, \quad i = 1, 2,$$

$$2\pi\left(j_1 - \frac{k_1}{\alpha}\right) = \frac{\theta_1}{\alpha} - \theta_0 = 2\pi\left(j_2 - \frac{k_2}{\alpha}\right),$$

whence $k_1 - k_2 = \alpha(j_1 - j_2)$, a contradiction. \square

Corollary 2 in combination with Theorem 8.3.3. elicits some interesting conclusions about convergence in distribution of normalized sums of i.i.d. r.v.s.

Theorem 2. If $\{X_n\}$ are i.i.d. with nondegenerate d.f. F , $S_n = \sum_1^n X_i$, and for some positive constants a_n and real b_n and nondegenerate d.f. G

$$\frac{S_n - b_n}{a_n} \xrightarrow{d} S_G,$$

Then

$$y_j(t) = -t \sum_{i=j}^m \frac{c_i}{\lambda_i} + \sum_{i=j}^m c_i,$$

and so $y_j(t) = y_{j+1}(t)$ entails $t = \lambda_j = t_j$ for $j < m$. Moreover,

$$\psi_{j+1} + \psi_{j-1} \geq 2\psi_j = \left(\frac{t_{j+1} + t_{j-1}}{t_j} \right) \psi_j$$

by convexity, implying $y_j(0) \geq y_{j+1}(0)$. Thus, $c_j = y_j(0) - y_{j+1}(0) \geq 0$ and $\sum_{j=1}^m c_j = y_1(0) = 1$. Furthermore, $\psi_m(t) \equiv \sum_{j=1}^m c_j (1 - (|t|/\lambda_j))^+$ coincides with $y_j(|t|)$ for $t_{j-1} \leq |t| \leq t_j$, $1 \leq j \leq m$, and $\psi_m(t)$ is a c.f., $m \geq 1$, via Exercise 8.3.3 and the table of c.f.s. By hypothesis $\psi(0) = \psi(0+)$, whence $\psi(t) = \lim_{n \rightarrow \infty} \psi_n(t)$, and so ψ is a c.f. by Theorem 8.3.3. \square

Corollary 4. *Two different c.f.s may coincide in any fixed finite interval $(-T, T)$, $T > 0$.*

PROOF. If suffices to change the c.f. $e^{-|t|}$ in $(T, T+1)$ and $(-T-1, -T)$ by replacing its arcs therein by line segments preserving continuity. According to Proposition 3, the modified function is still a c.f. \square

A necessary and sufficient condition for a continuous complex function φ on $(-\infty, \infty)$ with $\varphi(0) = 1$ to be a c.f. exists (Exercise 16) but is too unwieldy to constitute a working criterion.

One of the most beautiful and striking theorems of probability theory, conjectured by Lévy and proved by Cramér, asserts that if a normal d.f. is the convolution of two nondegenerate d.f.s, each must be normal.

To place this in perspective, an algebraic viewpoint is helpful. If d.f.s F , F_1, F_2 are in the relation $F = F_1 * F_2$, it is natural to call F_1 and F_2 factors or divisors of F . The corresponding c.f.s φ_1 and φ_2 are literal divisors of φ . If J_α denotes the d.f. degenerate at α , then J_α is a factor of every d.f. F since $\varphi_F(t) = e^{i\alpha t} [e^{-i\alpha t} \varphi_F(t)]$, but these are trivial and unwanted factors. Clearly, all divisors of discrete d.f.s are discrete, but the analogous statement for continuous d.f.s is false, as revealed by the trigonometric identity

$$\frac{\sin 2t}{2t} = \frac{\sin t}{t} \cos t.$$

Definition. A family \mathcal{F}' of d.f.s is **factor closed** if for all $F \in \mathcal{F}'$ the convolution relation $F = G_1 * G_2$ with G_i a nondegenerate d.f. implies that $G_i \in \mathcal{F}'$, $i = 1, 2$.

The theorem of Cramér–Lévy may be rephrased by saying that the family of normal d.f.s is factor closed. The only known method of proof is via c.f.s and requires a discussion of analytic c.f.s.

The next proposition asserts that a c.f. φ is r -analytic iff $\varphi(z)$ is defined and analytic in the strip $|\mathcal{I}(z)| = |y| < r$ of the complex plane. Consequently, the statement that φ is r -analytic may also be interpreted as asserting that $\varphi(z)$ as defined by (10) is analytic in $|\mathcal{I}(z)| < r$.

Proposition 5. A c.f. φ with d.f. F is r -analytic iff $\varphi(z)$ (in (10)) is an analytic function in $|\mathcal{I}(z)| < r$, and in this case for $k = 0, 1, 2, \dots$,

$$\varphi^{(k)}(z) = \sum_{j=k}^{\infty} \frac{i^j \alpha_j z^{j-k}}{(j-k)!}, \quad |z| < r, \quad (11)$$

where α_j is the j th moment of F and $\varphi^{(0)} = \varphi$.

PROOF. If φ is r -analytic and $C(F)$ is the continuity set of F , for $a, b \in C(F)$ by uniform convergence of the exponential series

$$g(z) = \int_a^b e^{izx} dF(x) = \sum_{j=0}^{\infty} \frac{(iz)^j}{j!} \int_a^b x^j dF(x),$$

and since the latter series converges uniformly in $|z| \leq M$ for all $0 < M < \infty$, $g(z)$ is entire and for $k = 0, 1, \dots$

$$g^{(k)}(z) = \sum_{j=k}^{\infty} \frac{i^j z^{j-k}}{(j-k)!} \int_a^b x^j dF(x) = \int_a^b (ix)^k e^{izx} dF(x).$$

Choose $\pm a_n \in C(F)$ with $0 < a_n \uparrow \infty$, $n \geq 1$, and define

$$g_1(z) = \int_{-a_1}^{a_1} e^{izx} dF(x), \quad g_n(z) = \int_{[a_{n-1} \leq |x| < a_n]} e^{izx} dF(x), \quad n > 1.$$

Then by the preceding, $g_n(z)$ is entire, $n \geq 1$, and for $|\mathcal{I}(z)| \leq s < r$

$$\sum_{n=m}^{\infty} |g_n(z)| \leq \int_{\{|u| \geq a_{m-1}\}} e^{|u\mathcal{I}(z)|} dF(u) \leq \int_{\{|u| \geq a_{m-1}\}} e^{s|u|} dF(u) < \infty$$

by Proposition 4. Thus, $\sum_1^{\infty} g_n(z)$ converges uniformly in $|\mathcal{I}(z)| \leq s < r$, whence (Titchmarsh, 1932, p. 95) $\varphi(z) = \int_{-\infty}^{\infty} e^{izu} dF(u) = \sum_1^{\infty} g_n(z)$ is analytic in $|\mathcal{I}(z)| < r$. Moreover, for $|\mathcal{I}(z)| < r$ and $k = 0, 1, \dots$, setting $a_0 = 0$,

$$\begin{aligned} \varphi^{(k)}(z) &= \sum_{n=1}^{\infty} g_n^{(k)}(z) = \sum_{n=1}^{\infty} \sum_{j=k}^{\infty} \frac{i^k (iz)^{j-k}}{(j-k)!} \int_{[a_{n-1} \leq |x| < a_n]} x^j dF(x) \\ &= \sum_{n=1}^{\infty} \int_{[a_{n-1} \leq |x| < a_n]} (ix)^k e^{izx} dF(x) = E(ix)^k e^{izX} \end{aligned}$$

and so, via (and in the notation of) Proposition 4, for $|z| \leq s < r$

$$\begin{aligned} & \sum_{n=1}^{\infty} \sum_{j=k}^{\infty} \left| \frac{i^k (iz)^{j-k}}{(j-k)!} \int_{[a_{n-1} \leq |u| < a_n]} u^j dF(u) \right| \\ & \leq \sum_{j=k}^{\infty} \frac{s^{j-k}}{(j-k)!} \beta_j = \frac{d^k}{ds^k} \sum_{j=0}^{\infty} \frac{\beta_j s^j}{j!} < \infty, \end{aligned}$$

whence

$$\varphi^{(k)}(z) = \sum_{j=k}^{\infty} \frac{i^k (iz)^{j-k}}{(j-k)!} \alpha_j, \quad |z| < r.$$

Conversely, if $\varphi(z)$ is defined in $|\mathcal{J}(z)| < r$, then (iv) of Proposition 4 holds whence φ is r -analytic. \square

Proposition 6. (i) If F is a d.f. whose c.f. φ is r -analytic for some $r > 0$, then F is uniquely determined by its moment sequence $\{\alpha_j = \int x^j dF(x), j \geq 1\}$.

(ii) A c.f. φ with moment sequence $\{\alpha_j, j \geq 1\}$ is r -analytic for some $r > 0$ iff

$$\limsup_{n \rightarrow \infty} \frac{(\alpha_{2n})^{1/2n}}{2n} < \infty. \quad (12)$$

PROOF. (i) Denote F, φ by F_1, φ_1 and suppose for $j \geq 1$ that $\alpha_j = \int x^j dF_k(x)$, $k = 1, 2$, where F_2 is a d.f. with c.f. φ_2 . By hypothesis and Proposition 4, φ_2 is r -analytic, whence by Proposition 5, $\varphi_k(z)$ is analytic in $|\mathcal{J}(z)| < r$, $k = 1, 2$, and

$$\varphi_1(z) = \sum_{j=0}^{\infty} \frac{\alpha_j (iz)^j}{j!} = \varphi_2(z), \quad |z| < r.$$

Therefore (Titchmarsh, 1932, p. 88), $\varphi_1(z) \equiv \varphi_2(z)$ in $|\mathcal{J}(z)| < r$, whence $\varphi_1(t) = \varphi_2(t)$ for $-\infty < t < \infty$. Then $F_1 = F_2$ by Corollary 8.3.1.

(ii) By Stirling's formula (Section 2.3)

$$\frac{1}{n} \log n! = \frac{1}{n} \left[\left(n + \frac{1}{2} \right) \log n - n + C + o(1) \right] = \log n + O(1)$$

whence, if $b_n = (|\alpha_n|/n!)^{-1/n}$, necessarily $\log b_n = -\log(|\alpha_n|^{1/n}/n) + O(1)$. Since the radius of convergence of a power series $\sum c_n z^n$ is $\liminf |c_n|^{-1/n}$ (Titchmarsh, 1932, p. 213), (ii) follows from Proposition 4. Conversely, (12) and $|\alpha_{2n-1}|^{1/(2n-1)} \leq \beta_{2n-1}^{1/(2n-1)} \leq \alpha_{2n}^{1/2n}$ imply (11) for $k = 0$ and some $r > 0$, so that *a fortiori* (ii) of Proposition (4) holds. \square

On the other hand a d.f. may be uniquely determined by its moment sequence without having an analytic c.f. To verify this, it is necessary to invoke the celebrated Carleman criterion (Shohat and Tamarkin, 1943, p. 19) which asserts that a d.f. F is uniquely determined by its moment sequence $\{\alpha_j, j \geq 1\}$ if

$$\sum_{j=1}^{\infty} \frac{1}{(\alpha_{2j})^{1/2j}} = \infty. \quad (13)$$

EXAMPLE 1. Let X be a random variable with $P\{X = n\} = Ce^{-n/\log n}$, $n \geq 3$, where C is such that $P\{X \geq 3\} = 1$. Then for $x \geq 3$ and $m \geq 4$, set

$$h(x) = x^m e^{-x/\log x}.$$

Now

$$\frac{h'(x)}{h(x)} = \frac{m}{x} - \frac{1}{\log x} + \frac{1}{\log^2 x} = 0$$

iff $x = x_m$, where

$$m = \frac{x_m}{\log x_m} \left(1 - \frac{1}{\log x_m}\right) \leq \frac{x_m}{\log x_m}.$$

As $m \rightarrow \infty$, $x_m \rightarrow \infty$, and therefore $m \log x_m/x_m \rightarrow 1$, whence $(m/2)\log m \leq x_m \leq 2m \log m$ for $m \geq m_0$. Define

$$M_m = \max_{x \geq 3} x^m e^{-x/\log x} = x_m^m e^{-x_m/\log x_m}.$$

Now for $m \geq m_0$

$$\begin{aligned} \alpha_{2m} &= C \sum_{n=3}^{\infty} n^{2m} e^{-n/\log n} \leq CM_{2m+2} \sum_{n=3}^{\infty} \frac{1}{n^2} \\ &\leq CM_{2m+2} \leq C[(4m+4)\log(2m+2)]^{2m+2} e^{-2m-2}, \end{aligned}$$

and therefore for all large m

$$(\alpha_{2m})^{1/2m} \leq 4C^{1/2m}[(2m+2)\log(2m+2)]^{1+(1/m)} \leq 8(2m+2)\log(2m+2).$$

Hence, $\sum_1^\infty (\alpha_{2j})^{-1/2j} = \infty$, and moreover $\overline{\lim}_{m \rightarrow \infty} (\alpha_{2m})^{1/2m}/2m = \infty$ since

$$(\alpha_{2m})^{1/2m} > (CM_{2m})^{1/2m} > e^{-1} \left(\frac{m}{2} \log m \right) \text{ for all large } m.$$

Thus φ is not r -analytic for any $r > 0$ by Proposition 6, whereas the Carleman criterion ensures that the distribution in question is uniquely determined by its moments. \square

Proposition 7. If $\varphi, \varphi_1, \varphi_2$ are c.f.s with $\varphi(t) = \varphi_1(t)\varphi_2(t)$ for all real t and φ is r_0 -analytic, then φ_i is r_0 -analytic, $i = 1, 2$, and $\varphi(z) = \varphi_1(z) \cdot \varphi_2(z)$ for $|z| < r_0$.

PROOF. If F, F_1, F_2 are the corresponding d.f.s, by Proposition 4

$$\int_{-\infty}^{\infty} e^{rx} dF(x) < \infty$$

for $0 < r < r_0$, whence by Section 6.3, (11),

$$\int_{-\infty}^{\infty} e^{ru} dF(u) = \int_{-\infty}^{\infty} e^{rx} dF_1(x) \int_{-\infty}^{\infty} e^{ry} dF_2(y), \quad |r| < r_0.$$

$E X_j^2 = -\varphi_j''(0) = -2a_{2j}$ implies $\varphi_j(t) = e^{-\sigma_j^2 t^2/2}$, where $\sigma_j^2 = E X_j^2 \geq 0$ and $\sigma_1^2 + \sigma_2^2 = 1$. \square

Lemma 2 (Titchmarsh, 1932, p. 247). *If $\varphi(z)$ is an entire, nonvanishing function, then $\varphi(z) = \exp\{g(z)\}$, where $g(z)$ is entire.*

PROOF. If $h(z) = \varphi'(z)/\varphi(z)$, then h is entire, whence $g(z) = \int_0^z h(w)dw$ is entire. Since

$$\frac{d}{dz} [\varphi(z)e^{-g(z)}] = e^{-g(z)} \left[\varphi'(z) - \varphi(z) \cdot \frac{\varphi'(z)}{\varphi(z)} \right] = 0$$

necessarily, $\varphi(z) = Ce^{g(z)} = e^{g(z)}$ upon noting that $g(0) = 0$. \square

Lemma 3 (Titchmarsh, 1932, p. 176). *If $\varphi(z) = \sum_0^\infty a_n z^n$ is analytic in $|z| < r_0$ and $A(r) = \max_{|z|=r} \Re\{\varphi(z)\}$, $0 < r < r_0$, then $|a_n|r^n \leq 4A^+(r) - 2\Re\{\varphi(0)\}$, $n > 0$.*

PROOF. Set $z = re^{i\theta}$, $0 < r < r_0$, and $a_n = \alpha_n + i\beta_n$, $n \geq 0$, whence $\varphi(z) = \sum_0^\infty a_n z^n = u(r, \theta) + iv(r, \theta)$ say. Since

$$u(r, \theta) = \sum_{n=0}^{\infty} (\alpha_n \cos n\theta - \beta_n \sin n\theta)r^n$$

converges uniformly in θ , for $n > 0$,

$$\frac{1}{\pi} \int_0^{2\pi} u(r, \theta) \cos n\theta d\theta = \alpha_n r^n, \quad \frac{1}{\pi} \int_0^{2\pi} u(r, \theta) \sin n\theta d\theta = -\beta_n r^n$$

and, recalling that $a_n = \alpha_n + i\beta_n$,

$$|a_n r^n| = \left| \frac{1}{\pi} \int_0^{2\pi} u(r, \theta) e^{-in\theta} d\theta \right| \leq \frac{1}{\pi} \int_0^{2\pi} |u(r, \theta)| d\theta, \quad n > 0.$$

Now $\Re\{\varphi(0)\} = \alpha_0 = (1/2\pi) \int_0^{2\pi} u(r, \theta) d\theta$, implying for $r > 0$ that $|a_n r^n| + 2\alpha_0 \leq 1/\pi \int_0^{2\pi} [u(r, \theta) + |u(r, \theta)|] d\theta \leq 4A^+(r)$, which is Lemma 3. \square

A theorem of Raikov–Ottaviani asserts that the Poisson family is factor closed (to within translations), but the proof closely parallels that of Theorem 3 (see Exercise 12).

Theorem 4. *The family of binomial distributions is factor closed (to within translations).*

PROOF. Suppose that $\varphi_1(t) \cdot \varphi_2(t) = \varphi(t) = (pe^{it} + q)^n$, $0 < p < 1$, $q = 1 - p$, n = positive integer. Clearly, the “translations” $e^{i\alpha t}\varphi_1(t)$ and $e^{-i\alpha t}\varphi_2(t)$ are also factors, but the theorem asserts that for some choice of α these are both binomial. Since $1 = |\varphi(2\pi)| = \prod_{j=1}^2 |\varphi_j(2\pi)| \leq 1$, Proposition 2 guarantees that F_j is discrete with support $S_j \subset \{c'_j + k : k \text{ an integer}, c'_j \text{ real}\}$, $j = 1, 2$. Clearly (Exercise 8.2.5), S_j is bounded since the support of the binomial

10. If $F_j, j = 1, 2, 3$, are d.f.s with $F_1 * F_3 = F_2 * F_3$, it does not follow that $F_1 = F_2$.
Hint: If φ is the c.f. of a r.v. X with $P\{X = 0\} = \frac{1}{2}$, $P\{X = \pm k\pi\} = 2/(k^2\pi^2)$, $k = 1, 3, 5, \dots$, then φ is a periodic c.f. coinciding with $\varphi_1(t) = (1 - |t|)I_{|t| \leq 1}$ in $[-1, 1]$.
11. There exist uncountably many absolutely continuous d.f.s with support $(0, \infty)$ and identical moments $v_k, k = 1, 2, \dots$. *Hint:* For $0 < x < \frac{1}{2}, r > 0$, and a nonnegative integer k , set $c = (k + 1)/x$, $b = r + ix$ in the well-known formula $b^{-c}\Gamma(c) = \int_0^\infty y^{c-1}e^{-by} dy$. Convert this to a similar example with support $(-\infty, \infty)$.
12. Prove the Raikov–Ottaviani theorem that the family of Poisson d.f.s is factor closed.
Hint: It may be supposed, without loss of generality, that the support S_j (of any factor $F_j \subset \{0, 1, 2, \dots\}$) utilize the corresponding probability generating functions $\psi(w) = e^{kw-1}$, $\psi_j(w) = E w^{X_j} = \sum_{i=0}^{\infty} P\{X_j = i\}w^i, j = 1, 2$.
13. If c.f.s φ_1, φ_2 satisfy $\prod_{j=1}^2 [\varphi_j(t)]^{n_j} = (pe^t + q)^n$, where $\alpha_j > 0, j = 1, 2$, then $\varphi(t) = e^{itc} (pe^t + q)^{n_j}$ with n_1, n_2 positive integers such that $\alpha_1 n_1 + \alpha_2 n_2 = n$; also $c_1 \alpha_1 + c_2 \alpha_2 = 0$.
14. If $1 - \varphi(t) = O(|t|^x)$ as $t \rightarrow 0$ for some x in $(0, 2]$, then $P\{|X| > c\} = O(c^{-x})$ as $c \rightarrow \infty$. *Hint:* Integrate $\int_{\{|x| > c\}} (1 - \cos tx) dF(x) \leq kt^x$ over $(0, c^{-1})$ or use Lemma 8.3.1.
15. Prove that

$$\frac{2}{\pi} \int_{-\infty}^{\infty} \frac{1 - \Re\{\varphi(t)\}}{t^2} dt = \int_{-\infty}^{\infty} |x| dF(x).$$

16. Show that a c.f. is **positive definite** i.e., $\sum_{j,k=1}^n \varphi(t_j - t_k) \rho_j \bar{\rho}_k \geq 0$ for any $n \geq 1$, real t_1, \dots, t_n and complex ρ_1, \dots, ρ_n . Conversely, a continuous, positive definite function φ on $(-\infty, \infty)$ with $\varphi(0) = 1$ is a c.f. This is **Bochner's theorem**.

8.5 Remarks on k -Dimensional Distribution Functions and Characteristic Functions

Multidimensional d.f.s (Section 6.3) lack the simplicity of their one-dimensional counterparts and it is frequently easier to deal with the corresponding measure or c.f. The c.f. of a random vector $X = (X_1, \dots, X_k)$ on a probability space (Ω, \mathcal{F}, P) is defined by

$$\varphi(t) = \varphi(t_1, \dots, t_k) = E \exp\left(i \sum_1^k t_j X_j\right) \quad (1)$$

and is also expressible as a Fourier–Stieltjes transform

$$\varphi(t_1, \dots, t_k) = \int \exp\left(i \sum_1^k t_j x_j\right) dF(x_1, \dots, x_k),$$

where F is the d.f. of X .

It is easy to verify that a joint d.f. $F(x_1, \dots, x_k)$ is continuous at a point (x_1, \dots, x_k) if its **marginal d.f.**

$$F_j(y) = \lim_{\substack{y_i \rightarrow \infty \\ i \neq j}} F(y_1, \dots, y_{j-1}, y, y_{j+1}, \dots, y_k)$$

is continuous at $y = x_j$ for each $j = 1, 2, \dots, k$. However, in contradistinction to the case $k = 1$, F may be discontinuous even though the corresponding measure $F\{\cdot\} = P(X^{-1})$ assigns probability zero to every point of R^k . For example F may allocate probability one to some hyperplane or hypersurface of R^m , $1 \leq m < k$, without assigning positive probability to any point therein (see Exercise 1).

For any bounded increasing function F on R^k , define

$$C(F) = \{x = (x_1, \dots, x_k) : F_j(x_j+) = F_j(x_j-), 1 \leq j \leq k\},$$

where F_j is the "marginal function" of the prior paragraph. As noted above, $C(F)$ need not coincide with the set of continuity points of F when $k > 1$.

Definition. A sequence $\{F_n, n \geq 1\}$ of d.f.s on R^k converges weakly to a function F on R^k , denoted by $F_n \xrightarrow{*} F$, if $\lim F_n(x) = F(x)$ for all $x \in C(F)$. If, moreover, F is a d.f., then F_n is said to converge completely to F , denoted by $F_n \xrightarrow{c} F$.

A straightforward generalization of Theorem 8.3.1 yields

Theorem 1. If $X = (X_1, \dots, X_k)$ is a random vector with c.f. φ and d.f. F , then

$$\begin{aligned} & \lim_{\epsilon \rightarrow \infty} \frac{1}{(2\pi)^k} \int_{-\epsilon}^{\epsilon} \cdots \int_{-\epsilon}^{\epsilon} \prod_{j=1}^k \frac{e^{-it_j a_j} - e^{-it_j b_j}}{it_j} \varphi(t_1, \dots, t_k) dt_1 \cdots dt_k \\ &= E \prod_{j=1}^k [\tfrac{1}{2} I_{[X_j = a_j \text{ or } b_j]} + I_{[a_j < X_j < b_j]}] \end{aligned}$$

and the right side reduces to $P\{a_j < X_j < b_j, 1 \leq j \leq k\}$ when $a = (a_1, \dots, a_k)$ and $b = (b_1, \dots, b_k)$ are in $C(F)$.

The reformulation and proof of Theorem 8.3.2 for R^k are immediate. The statement of the Lévy continuity theorem (Theorem 8.3.3) carries over verbatim to R^k and the proof generalizes readily.

It is worthy of note that independence may be characterized in terms of c.f.s as well as d.f.s.

Theorem 2. Random variables X_1, \dots, X_k on some probability space are independent iff their joint c.f. is the product of their marginal c.f.s, i.e.,

$$\varphi_{X_1, \dots, X_k}(t_1, \dots, t_k) = \prod_{j=1}^k \varphi_{X_j}(t_j). \quad (2)$$

4. If F_i , $1 \leq i \leq k$, are d.f.s on R^1 , show that for any α in $[-1, 1]$, $F(x_1, \dots, x_k) = [1 + \alpha \prod_{i=1}^k (1 - F_i(x_i))] \prod_{i=1}^k F_i(x_i)$ is a d.f. with the given marginal d.f.s.
5. Prove that $X = (X_1, \dots, X_k)$ has a multivariate normal distribution with mean vector $\theta = (\theta_1, \dots, \theta_k)$ and covariance matrix $\Sigma = \{\sigma_{ij}\}$ iff every linear combination $cX' = \sum_{j=1}^k c_j X_j$ has a normal distribution on R^1 with mean $c\theta'$ and variance $c\Sigma c'$.
6. If $X_n = (X_{n1}, \dots, X_{nk})$, $n \geq 1$, is a sequence of random vectors for which every linear combination $cX_n' \xrightarrow{d} N_{c\mu', c\Sigma c'}$, where $N_{\mu, \Sigma}$ is a fictitious normal r.v. with mean μ and variance Σ , prove that the d.f. of X_n converges to the normal d.f. on R^k with mean vector μ and covariance matrix Σ .
7. Prove the Cramér-Lévy theorem (Theorem 8.4.3) in R^k . Hint: Use the result for $k = 1$.
8. Generalize Theorem 8.4.4 to the multinomial distribution.
9. Prove Theorem 8.5.1 and deduce the one-to-one correspondence between d.f.s and c.f.s on R^k .
10. Prove the k -dimensional analogue of the continuity theorem (Theorem 8.3.3).
11. Verify that if $F_n \xrightarrow{d} F$, the marginal d.f.s $F_{n,j} \xrightarrow{d} F_j$, $1 \leq j \leq k$.
12. Let the random vectors $X_n \xrightarrow{d} X_0$, where $X_0 = (X_{01}, \dots, X_{0k})$ is a possibly fictitious random vector with d.f. F . If $\{Y_n, n \geq 0\}$ are k -dimensional random vectors whose i th component is $g_i(X_{n,1}, \dots, X_{n,k})$, $1 \leq i \leq k$, $n \geq 0$, where $\{g_i, 1 \leq i \leq k\}$ are continuous functions on R^k , then $Y_n \xrightarrow{d} Y_0$.

References

- H. E. Bray, "Elementary properties of the Stieltjes integral," *Ann. Math.* **20** (1919), 177-186.
- F. P. Cantelli, "Una teoria astratta del calcolo delle probabilità," *Ist. Ital. Attuari* **3** (1932).
- H. Chernoff, "Large sample theory: Parametric case," *Ann. Math. Stat.* **27** (1956), 1-22.
- Y. S. Chow and H. Robbins, "On optimal stopping rules," *Z. Wahr.* **2** (1963), 33-49.
- K. L. Chung, *A Course in Probability Theory*, Harcourt Brace, New York, 1968; 2nd ed., Academic Press, New York, 1974.
- H. Cramér, "Über eine Eigenschaft der normalen Verteilungsfunktion," *Math. Z.* **41** (1936), 405-414.
- H. Cramér, *Mathematical Methods of Statistics*, Princeton Univ. Press, Princeton, 1946.
- H. Cramér, *Random Variables and Probability Distributions*, Cambridge Tracts Math. No. 36, Cambridge Univ. Press, London, 1937; 3rd ed., 1970.
- J. L. Doob, *Stochastic Processes*, Wiley, New York, 1953.
- W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2, Wiley, New York, 1966.
- M. Fréchet and J. Shohat, "A proof of the generalized second limit theorem in the theory of probability," *Trans. Amer. Math. Soc.* **33** (1931).
- J. Glivenko, *Stieltjes Integral*, 1936 [in Russian].

- B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables* (K. L. Chung, translator), Addison-Wesley, Reading, Mass., 1954.
- G. H. Hardy, *A course of Pure Mathematics*, 10th ed., Cambridge Univ. Press, New York, 1952.
- E. Helly, "Über lineare Funktionaloperationen," *Sitz. Nat. Kais. Akad. Wiss.* **121**, IIa (1921), 265–277.
- P. Lévy, *Calcul des probabilités*, Gauthier-Villars, Paris, 1925.
- P. Lévy, *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, Paris, 1937; 2nd ed., 1954.
- M. Loève, *Probability Theory*, 3rd ed., Van Nostrand, Princeton, 1963; 4th ed., Springer-Verlag, Berlin and New York, 1977–1978.
- E. Lukacs, *Characteristic Functions*, 2nd ed., Haffner, New York, 1970.
- G. Polya, "Remarks on characteristic functions," *Proc. 1st Berkeley Symp. Stat. and Prob.*, 1949, 115–123.
- D. A. Raikov, "On the decomposition of Gauss and Poisson laws," *Izv. Akad. Nauk. USSR (Ser. Mat.)* **2** (1938a), 91–124 [in Russian].
- D. A. Raikov, "Un théorème de la théorie des fonctions caractéristiques analytiques," *Izvest. Fak. Mat. Mek. Univ. Tomsk NII* **2** (1938b), 8–11.
- H. Robbins, "Mixture of distributions," *Ann. Math. Stat.* **19** (1948), 360–369.
- S. Saks, *Theory of the Integral* (L. C. Young, translator), Stechert-Hafner, New York, 1937.
- N. A. Sapogov, "The stability problem for a theorem of Cramer," *Izv. Akad. Nauk. USSR (ser. Mat.)* **15** (1951), 205–218. [See also selected translations, *Math. Stat. and Prob.* **1**, 41–53, Amer. Math. Soc.]
- H. Scheffé, "A useful convergence theorem for probability distributions," *Ann. Math. Stat.* **18** (1947), 434–438.
- J. A. Shohat and J. D. Tamarkin, "The problem of moments," Math. Survey No. 1, Amer. Math. Soc., New York, 1943.
- E. Slutsky, "Über stochastiche Asymptoten und Grenzwerte," *Metron* **5** (1925), 1–90.
- H. Teicher, "On the factorization of distributions," Ph.D. Thesis, 1950. [See also *Ann. Math. Stat.* **25** (1954), 769–774.]
- H. Teicher, "Sur les puissances de fonctions caractéristiques," *Comptes Rendus* **246** (1958), 694–696.
- H. Teicher, "On the mixture of distributions," *Ann. Math. Stat.* **31** (1960), 55–73.
- H. Teicher, "Identifiability of mixtures," *Ann. Math. Stat.* **32** (1961), 244–248.
- E. C. Titchmarsh, *The Theory of Functions*, Oxford Univ. Press, 1932; 2nd ed., 1939.

9

Central Limit Theorems

Central limit theorems have played a paramount role in probability theory starting—in the case of independent random variables—with the DeMoivre–Laplace version and culminating with that of Lindeberg–Feller. The term “central” refers to the pervasive, although nonunique, role of the normal distribution as a limit of d.f.s of normalized sums of (classically independent) random variables. Central limit theorems also govern various classes of dependent random variables and the cases of martingales and interchangeable random variables will be considered.

9.1 Independent Components

Consider at the outset a sequence $\{X_n, n \geq 1\}$ of independent random variables with finite variances $\{\sigma_n^2, n \geq 1\}$. No generality is lost and much convenience is gained by supposing (as will be done) that $E X_n = 0, n \geq 1$. Set

$$S_n = \sum_{i=1}^n X_i, \quad s_n^2 = E S_n^2 = \sum_{i=1}^n \sigma_i^2, \quad n \geq 1. \quad (1)$$

The problem, of analytic rather than measure-theoretic character, is to determine when S_n , suitably normalized (to S_n/s_n), converges in law to the standard normal distribution. The solution is linked to the following

Definition. Random variables $\{X_n, n \geq 1\}$ with $E X_n = 0, E X_n^2 = \sigma_n^2 < \infty$, and d.f.s F_n are said to obey the **Lindeberg condition** if $s_n^2 = \sum_1^n \sigma_i^2 > 0$ for some n and

$$\sum_{j=1}^n \int_{\{|x| > \alpha_j\}} x^2 dF_j(x) = o(s_n^2), \quad \text{all } \varepsilon > 0. \quad (2)$$

Condition (2) requires that $s_n \rightarrow \infty$ and is equivalent to the classical form of the Lindeberg condition, viz.,

$$\sum_{j=1}^n \int_{\{|x| > \varepsilon s_n\}} x^2 dF_j(x) = o(s_n^2) \quad \text{for all } \varepsilon > 0. \quad (2')$$

Monotonicity of s_n yields $(2) \Rightarrow (2')$, while the reverse implication follows by noting that for all $\varepsilon > 0$ and arbitrarily small $\delta > 0$

$$\begin{aligned} s_n^{-2} \sum_{j=1}^n \int_{\{|x| > \varepsilon s_j\}} x^2 dF_j(x) &= s_n^{-2} \left[\sum_{j:s_j \leq \delta s_n} + \sum_{j:s_j > \delta s_n} \right] \\ &\leq \delta^2 + s_n^{-2} \sum_{j=1}^n \int_{\{|x| > \varepsilon \delta s_n\}} x^2 dF_j \rightarrow \delta^2. \end{aligned}$$

Despite their equivalence, (2) is structurally simpler than (2'). Moreover, (2) or (2') entails

$$\max_{1 \leq j \leq n} \frac{\sigma_j^2}{s_n^2} = o(1), \quad (3)$$

since for arbitrary $\varepsilon > 0$

$$\max_j \sigma_j^2 s_n^{-2} \leq \max_j s_n^{-2} \left[\varepsilon^2 s_n^2 + \int_{\{|x| > \varepsilon s_n\}} x^2 dF_j \right] = \varepsilon^2 + o(1).$$

If $\{X_n\}$ are independent with $E X_n = 0$, $E X_n^2 = \sigma_n^2$, either $s_n^2 = \sum_i \sigma_i^2 \rightarrow \infty$ or $s_n^2 \uparrow s^2 < \infty$. The latter contingency is devoid of interest in the current context since, if N_{μ, σ^2} denotes a fictitious normal random variable with mean μ , variance σ^2 , and both $S_n/s_n \xrightarrow{d} N_{0, 1}$ and $s_n \uparrow s$, then $S_n \xrightarrow{d} N_{0, s^2}$ by Corollary 8.1.3. In terms of the characteristic function of X_j , say $\varphi_j(t)$, this entails

$$\varphi_1(t) \cdot \prod_{j=2}^{\infty} \varphi_j(t) = \lim_{n \rightarrow \infty} \prod_{j=1}^n \varphi_j(t) = \exp \left\{ - \frac{s^2 t^2}{2} \right\}.$$

By the Cramér–Lévy theorem (Theorem 8.4.3) both of the c.f.s $\varphi_1(t)$ and $\prod_{j=2}^{\infty} \varphi_j(t)$ must be normal. Isolating $\varphi_2(t)$ analogously, it follows that X_2 and eventually all X_n are normally distributed.

Theorem 1 (Lindeberg). *If $\{X_n, n \geq 1\}$ are independent random variables with zero means, variances $\{\sigma_n^2\}$, and distribution functions $\{F_n\}$ satisfying (2), then the distribution functions of the normalized sums S_n/s_n tend to the standard normal. Conversely (Feller), convergence of these distribution functions to the standard normal and*

$$\frac{\sigma_n}{s_n} \rightarrow 0, \quad s_n \rightarrow \infty \quad (4)$$

imply (2).

PROOF. Let t, ε be fixed but arbitrary real numbers, the latter being positive. Set

$$Y_j(t) = e^{itX_j} - 1 - itX_j + \frac{t^2 X_j^2}{2}, \quad a_j(t) = e^{-\sigma_j^2 t^2/2} - 1 + \frac{\sigma_j^2 t^2}{2}$$

and note that $|Y_j(t)| \leq \min[t^2 X_j^2, (1/6)|tX_j|^3]$ by Lemma 8.4.1 with $\delta = 1$ and $n = 1, 2$. Consequently, recalling that $E X_j = 0$,

$$\begin{aligned} |E e^{itX_j/s_n} - e^{-\sigma_j^2 t^2/2s_n^2}| &= \left| E Y_j\left(\frac{t}{s_n}\right) - a_j\left(\frac{t}{s_n}\right) \right| \\ &\leq E \left[\frac{t^2 X_j^2}{s_n^2} I_{\{|X_j| > \varepsilon s_n\}} + \left| \frac{tX_j}{s_n} \right|^3 I_{\{|X_j| \leq \varepsilon s_n\}} \right] + \frac{\sigma_j^4 t^4}{8s_n^4}. \end{aligned} \quad (5)$$

Thus, for $1 \leq j \leq n$, setting $S_0 = 0 = s_0$ and utilizing independence and (5),

$$\begin{aligned} &\left| E \exp \left\{ it \left(\frac{S_j}{s_n} \right) + \frac{s_j^2 t^2}{2s_n^2} \right\} - E \exp \left\{ it \left(\frac{S_{j-1}}{s_n} \right) + \frac{s_{j-1}^2 t^2}{2s_n^2} \right\} \right| \\ &= \left| E \exp \left\{ it \left(\frac{S_{j-1}}{s_n} \right) + \frac{s_{j-1}^2 t^2}{2s_n^2} \right\} E \left(\exp \left\{ \frac{itX_j}{s_n} \right\} - \exp \left\{ -\frac{\sigma_j^2 t^2}{2s_n^2} \right\} \right) \right| \\ &\leq e^{t^2/2} \left| E \exp \left\{ \frac{itX_j}{s_n} \right\} - \exp \left\{ -\frac{\sigma_j^2 t^2}{2s_n^2} \right\} \right| \\ &\leq e^{t^2/2} E \left[\frac{t^2 X_j^2}{s_n^2} I_{\{|X_j| > \varepsilon s_n\}} + \varepsilon |t|^3 \frac{X_j^2}{s_n^2} I_{\{|X_j| \leq \varepsilon s_n\}} + \frac{t^4 \sigma_j^2}{s_n^2} \max_{1 \leq j \leq n} \frac{\sigma_j^2}{s_n^2} \right] \end{aligned} \quad (6)$$

Finally, via (6) and (3)

$$\begin{aligned} &|E e^{itS_n/s_n} - e^{-t^2/2}| \\ &= \left| e^{-t^2/2} \sum_{j=1}^n E \left(\exp \left\{ it \left(\frac{S_j}{s_n} \right) + \frac{s_j^2 t^2}{2s_n^2} \right\} - \exp \left\{ it \left(\frac{S_{j-1}}{s_n} \right) + \frac{s_{j-1}^2 t^2}{2s_n^2} \right\} \right) \right| \\ &\leq \frac{t^2}{s_n^2} \sum_{j=1}^n E X_j^2 I_{\{|X_j| > \varepsilon s_n\}} + \varepsilon |t|^3 + o(1), \end{aligned} \quad (7)$$

and, since ε is arbitrary, the conclusion follows from the hypothesis and Theorem 8.3.3.

Conversely, (4) entails (3) via

$$\max_{1 \leq j \leq n} \frac{\sigma_j}{s_n} \leq \max_{1 \leq j \leq m} \frac{\sigma_j}{s_n} + \max_{m < j \leq n} \frac{\sigma_j}{s_j} \rightarrow 0$$

as first n and then $m \rightarrow \infty$. This, in turn, yields as $n \rightarrow \infty$

$$\begin{aligned} \sum_{j=1}^n \left| \log \varphi_j\left(\frac{t}{s_n}\right) - \left(\varphi_j\left(\frac{t}{s_n}\right) - 1 \right) \right| &\leq \sum_{j=1}^n \left| \varphi_j\left(\frac{t}{s_n}\right) - 1 \right|^2 \\ &= o(1) \sum_{j=1}^n \left| \varphi_j\left(\frac{t}{s_n}\right) - 1 \right| = o(1) \quad (8) \end{aligned}$$

since Lemma 8.4.1 guarantees $|\varphi_j(t/s_n) - 1| \leq t^2 \sigma_j^2 / 2s_n^2$, which in conjunction with (3) implies

$$\max_{1 \leq j \leq n} \left| \varphi_j\left(\frac{t}{s_n}\right) - 1 \right| = o(1), \quad \sum_{j=1}^n \left| \varphi_j\left(\frac{t}{s_n}\right) - 1 \right| \leq \frac{t^2}{2}.$$

The hypothesis ensures $\prod_{j=1}^n \varphi_j(t/s_n) \rightarrow e^{-t^2/2}$, whence $\sum_{j=1}^n \log \varphi_j(t/s_n) \rightarrow -t^2/2$ and, taking cognizance of (8), $-t^2/2 = \sum_{j=1}^n [\varphi_j(t/s_n) - 1] + o(1)$ or

$$\begin{aligned} \frac{t^2}{2} - \sum_{j=1}^n \int_{\{|x| \leq \epsilon s_n\}} \left(1 - \cos \frac{tx}{s_n}\right) dF_j(x) \\ = \sum_{j=1}^n \int_{\{|x| > \epsilon s_n\}} \left(1 - \cos \frac{tx}{s_n}\right) dF_j(x) + o(1). \end{aligned}$$

Since the integrand on the right $\leq 2 \leq 2x^2/\epsilon^2 s_n^2$, while that on the left $\leq t^2 x^2 / 2s_n^2$, it follows that as $n \rightarrow \infty$

$$1 - \frac{1}{s_n^2} \sum_{j=1}^n \int_{\{|x| \leq \epsilon s_n\}} x^2 dF_j(x) \leq \frac{4}{t^2 \epsilon^2} + o(1),$$

$$\lim_{n \rightarrow \infty} s_n^{-2} \sum_{j=1}^n \int_{\{|x| \leq \epsilon s_n\}} x^2 dF_j(x) \geq 1 - \frac{4}{t^2 \epsilon^2} \xrightarrow{t \rightarrow \infty} 1,$$

which is tantamount to (2'). □

Corollary 1 (Liapounov). *If $\{X_n\}$ are independent with $E X_n = 0$ and $\sum_{j=1}^n E|X_j|^{2+\delta} = o(s_n^{2+\delta})$ for some $\delta > 0$, then*

$$\lim_{n \rightarrow \infty} P\left\{\frac{S_n}{s_n} < x\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du. \quad (9)$$

PROOF. Take $q = 2$ and $r = 2 + \delta$ in

$$\begin{aligned} 0 &\leq \frac{1}{s_n^q} \sum_{j=1}^n \int_{\{|x| > \epsilon s_n\}} |x|^q dF_j(x) \\ &\leq \frac{1}{\epsilon^{r-q} s_n^r} \sum_{j=1}^n \int_{\{|x| > \epsilon s_n\}} |x|^r dF_j = o(1), \quad 0 < q < r. \quad \square \quad (10) \end{aligned}$$

Corollary 2. If $\{X_n\}$ are i.i.d. with $E X_n = \mu$, $E(X_n - \mu)^2 = \sigma^2 \in (0, \infty)$, then

$$\lim_{n \rightarrow \infty} P \left\{ \frac{S_n - n\mu}{\sigma\sqrt{n}} < x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du. \quad (11)$$

If the $\{X_n\}$ have finite moments of order k , what additional assumptions will ensure moment convergence, i.e., that for any positive integer $k > 2$

$$\lim_{n \rightarrow \infty} E \left(\frac{S_n}{s_n} \right)^k = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^k e^{-x^2/2} dx. \quad (12)$$

The answer is tied to the next

Definition. Random variables $\{X_n, n \geq 1\}$ with $E X_n = 0$, $E X_n^2 = s_n^2$ are said to obey a Lindeberg condition of order $r > 0$ if

$$\sum_{j=1}^n \int_{\{|x| > \varepsilon s_j\}} |x|^r dF_j(x) = o(s_n^r), \quad \text{all } \varepsilon > 0. \quad (13)$$

For $r = 2$, this is just the ordinary Lindeberg condition. Surprisingly, for $r > 2$, (13) is equivalent to

$$\sum_{j=1}^n E |X_j|^r = o(s_n^r) \quad (13')$$

and also to (13'') (defined as (13) with s_j replaced by s_n). Clearly (13') \Rightarrow (13) \Rightarrow (13'') and so to establish equivalence for $r > 2$ it suffices to verify that (13'') \Rightarrow (13'). The latter follows from the fact that for $r > 2$ and all $\varepsilon > 0$

$$\begin{aligned} \sum_{j=1}^n E |X_j|^r &\leq \sum_{j=1}^n E \{(es_n)^{-2} X_j^2 I_{\{|X_j| \leq es_n\}} + |X_j|^r I_{\{|X_j| > es_n\}}\} \\ &\leq \varepsilon^{r-2} s_n^r + o(s_n^r). \end{aligned}$$

According to (10), a Lindeberg condition of order $r > 2$ implies that of order q , for all q in $[2, r]$; in particular, a Lindeberg condition of order $r > 2$ implies the central limit theorem (9).

Theorem 2. Let $\{X_n, n \geq 1\}$ be independent random variables with $E X_n = 0$, $E X_n^2 = s_n^2$. If $\{X_n\}$ satisfies a Lindeberg condition of order r for some integer $r \geq 2$, then (12) obtains for $k = 1, 2, \dots, r$.

Corollary 3. If $\{X_n, n \geq 1\}$ are i.i.d. with $E X_1 = 0$, $E X_1^{2k} \in (0, \infty)$ for some positive integer k , then $\lim_{n \rightarrow \infty} E(S_n/s_n)^{2j} = (2j)!/2^j j!$, $E(S_n/s_n)^{2j+1} = o(1)$, $1 \leq j \leq k$.

PROOF OF THEOREM 2. Since $E(S_n/s_n)^2 = 1$, $n \geq 1$, the theorem is clearly valid for $r = 2$. In the case $r > 2$, suppose inductively that the theorem holds for

Lemma 2. If, in addition to the hypotheses of Lemma 1, G is of bounded variation on $(-\infty, \infty)$ and $F - G \in \mathcal{L}_1$, then for every $T > 0$

$$\sup_x |F(x) - G(x)| \leq \frac{2}{\pi} \int_0^T \left| \frac{\varphi_F(t) - \varphi_G(t)}{t} \right| dt + \frac{24M}{\pi T}, \quad (18)$$

where φ_F, φ_G are Fourier-Stieltjes transforms of F, G .

PROOF. In the nontrivial case where the integral is finite,

$$\varphi_F(t) - \varphi_G(t) = \int_{-\infty}^{\infty} e^{itx} d[F(x) - G(x)] = -it \int_{-\infty}^{\infty} [F(x) - G(x)] e^{itx} dx,$$

whence

$$\frac{\varphi_F(t) - \varphi_G(t)}{-it} e^{-itc} = \int_{-\infty}^{\infty} H_c(x) e^{itx} dx.$$

Since the above right side is bounded, via Fubini

$$\begin{aligned} \int_{-T}^T \frac{\varphi_F(t) - \varphi_G(t)}{-it} e^{-itc} [T - |t|] dt &= \int_{-T}^T \int_{-\infty}^{\infty} H_c(x) e^{itx} (T - |t|) dx dt \\ &= \int_{-\infty}^{\infty} \int_{-T}^T e^{itx} (T - |t|) H_c(x) dt dx \\ &= \int_{-\infty}^{\infty} \frac{2(1 - \cos Tx)}{x^2} H_c(x) dx \\ &= 2T \int_{-\infty}^{\infty} \frac{\sin^2 x}{x^2} H_c\left(\frac{2x}{T}\right) dx \end{aligned}$$

and so

$$\begin{aligned} \left| \int_{-\infty}^{\infty} \left(\frac{\sin x}{x} \right)^2 H_c\left(\frac{2x}{T}\right) dx \right| &= \frac{1}{2T} \left| \int_{-T}^T \frac{\varphi_F - \varphi_G}{-it} e^{-itc} (T - |t|) dt \right| \\ &\leq \frac{1}{2} \int_{-T}^T \left| \frac{\varphi_F - \varphi_G}{t} \right| dt = \int_0^T \left| \frac{\varphi_F(t) - \varphi_G(t)}{t} \right| dt, \end{aligned}$$

whence by Lemma 1

$$\int_0^T \left| \frac{\varphi_F(t) - \varphi_G(t)}{t} \right| dt \geq 2M\delta \left(\frac{\pi}{2} - \frac{6}{T\delta} \right),$$

yielding (18). \square

which differ from the choices of (25), (24) only in the sense of Corollary 8.2.2. Thus, it may and will be supposed that $E X^2 = \infty$.

By the dominated convergence theorem

$$c^{-2} \int_{\{|x| < c\}} x^2 dF(x) = o(1)$$

as $c \rightarrow \infty$ and so, defining

$$B_n = \sup \left\{ c : c^{-2} \int_{\{|x| < c\}} x^2 dF(x) \geq \frac{1}{n} \right\}, \quad (25)$$

necessarily $B_n \uparrow \infty$ and via continuity

$$\frac{n}{B_n^2} \int_{\{|x| < B_n\}} x^2 dF(x) = 1. \quad (26)$$

Under the hypothesis of (23), if $0 < \varepsilon \leq 1$,

$$\frac{\int_{\{cB_n \leq |x| \leq B_n\}} x^2 dF(x)}{\int_{\{|x| < B_n\}} x^2 dF(x)} \leq \frac{B_n^2 \int_{\{|x| \geq cB_n\}} dF}{\int_{\{|x| < cB_n\}} x^2 dF(x)} = o(1),$$

whence, recalling (26), for $0 < \varepsilon \leq 1$

$$\frac{n}{B_n^2} \int_{\{|x| < cB_n\}} x^2 dF(x) = \frac{n}{B_n^2} \int_{\{|x| < B_n\}} x^2 dF(x) \left[1 - \frac{\int_{\{cB_n \leq |x| \leq B_n\}} x^2 dF(x)}{\int_{\{|x| < B_n\}} x^2 dF(x)} \right] \rightarrow 1. \quad (27)$$

Thus, via (27) and (23)

$$\lim_{n \rightarrow \infty} n P\{|X_1| > \varepsilon B_n\} = \lim_{n \rightarrow \infty} \frac{n \int_{\{|x| > \varepsilon B_n\}} dF}{n B_n^{-2} \int_{\{|x| < \varepsilon B_n\}} x^2 dF} = 0 \quad (28)$$

for $0 < \varepsilon \leq 1$ and hence for all $\varepsilon > 0$. Likewise (27) holds for all $\varepsilon > 0$.

Define $X'_j = X_j I_{\{|X_j| < B_n\}}$, $1 \leq j \leq n$, and $S'_n = \sum_{j=1}^n X'_j$. Now (26) entails $n^{1/2} = o(B_n)$ so that if $Y_n = (n^{1/2}/B_n) X_1 I_{\{|X_1| < \varepsilon B_n\}}$, then $Y_n \xrightarrow{P} 0$ and, moreover, $E Y_n^2 \rightarrow 1$ by (27). Thus, $\{Y_n, n \geq 1\}$ is uniformly integrable, whence $(n^{1/2}/B_n) \int_{\{|x| < \varepsilon B_n\}} x dF = E Y_n = o(1)$ for all $\varepsilon > 0$, and so (27) implies

$$\frac{n}{B_n^2} \left[\int_{\{|x| < \varepsilon B_n\}} x^2 dF - \left(\int_{\{|x| < \varepsilon B_n\}} x dF \right)^2 \right] \rightarrow 1, \quad \varepsilon > 0. \quad (29)$$

Now for $\varepsilon = 1$, (29) asserts that $s_n^2 \equiv \sigma^2(S'_n) \sim B_n^2$ and, moreover, $E Y_n = o(1)$ implies $E X'_1 = o(B_n n^{-1/2})$. Thus, for $0 < \varepsilon \leq 1$ and all large n

$$\begin{aligned} & \frac{1}{s_n^2} \sum_{j=1}^n E[X'_j - E X'_j]^2 I_{\{|X'_j - E X'_j| > \varepsilon s_n\}} = \frac{n}{s_n^2} E[X'_1 - E X'_1]^2 I_{\{|X'_1 - E X'_1| > \varepsilon s_n\}} \\ & \leq \frac{2n}{s_n^2} \{E(X'_1)^2 I_{\{|X'_1| > \varepsilon s_n/2\}} + (E X'_1)^2\} = \frac{2n(1 + o(1))}{B_n^2} E X_1^2 I_{\{(e/2)s_n < |X_1| < B_n\}} + o(1) \\ & \leq 2n(1 + o(1)) P\left\{|X_1| > \frac{\varepsilon}{4} B_n\right\} + o(1) = o(1). \end{aligned}$$

according to (28). Consequently, $(S_n' - E S_n')/s_n \xrightarrow{d} N_{0,1}$ by Exercise 9.1.2 and,

$$P\{S_n \neq S_n'\} \leq n P\{|X_1| \geq B_n\} = o(1),$$

it follows that

$$\frac{S_n}{B_n} - A_n = \frac{S_n - E S_n'}{B_n} \xrightarrow{d} N_{0,1}.$$

Conversely, if (22) prevails, then (28) and (29) hold for all $\varepsilon > 0$ by Corollary 12.2.3, whence

$$\frac{(\varepsilon B_n)^2 \int_{\{|x| > B_n \varepsilon\}} dF}{\int_{\{|x| < B_n \varepsilon\}} x^2 dF(x)} \leq \frac{\varepsilon^2 n P\{|X_1| > B_n \varepsilon\}}{n B_n^{-2} [\int_{\{|x| < B_n \varepsilon\}} x^2 dF - (\int_{\{|x| < \varepsilon B_n\}} x dF)^2]} = o(1)$$

for all $\varepsilon > 0$, which is tantamount to (23). \square

EXAMPLE 1 (Friedman, Katz, Koopmans). If $\{X, X_n, n \geq 1\}$ are i.i.d. r.v.s with $E X = 0$, $E X^2 = \sigma^2 \in (0, \infty)$, $\sigma_n^2 = \sigma^2(X'_n)$ where $X'_n = (X \wedge n^{1/2}) \vee (-n^{1/2})$, $S_n = \sum_{j=1}^n X_j$, then

$$\sum_{n=1}^{\infty} n^{-1} \sup_x |P\{S_n/n^{1/2} < x\} - \Phi(x/\sigma_n)| < \infty. \quad (30)$$

PROOF. Since $\sigma_n \rightarrow \sigma$, it may and will be supposed that $\sigma_n > 0$, $n \geq 1$ and that $\sigma = 1$. Let S'_n be the sum of n i.i.d. r.v.s having the common distribution of X'_n and set $\mu_n = E X'_n$. In view of

$$|P\{S_n < xn^{1/2}\} - P\{S'_n < xn^{1/2}\}| \leq n P\{|X| > n^{1/2}\}$$

it suffices to prove (30) with S_n replaced by S'_n . Now

$$\begin{aligned} P\{S'_n < xn^{1/2}\} - \Phi(x/\sigma_n) &= \left[P\left\{\frac{S'_n - n\mu_n}{n^{1/2}\sigma_n} < \frac{xn^{1/2} - n\mu_n}{n^{1/2}\sigma_n}\right\} - \Phi\left(\frac{x - n^{1/2}\mu_n}{\sigma_n}\right) \right] \\ &\quad + \left[\Phi\left(\frac{x - n^{1/2}\mu_n}{\sigma_n}\right) - \Phi\left(\frac{x}{\sigma_n}\right) \right] = A_n + B_n \text{ (say).} \end{aligned}$$

If C_j , $j \geq 1$ are positive constants, the Berry-Esseen theorem ensures that if $A'_n = \sup_x |A_n|$ and $B'_n = \sup_x |B_n|$

$$\sum_{n=1}^{\infty} n^{-1} A'_n \leq \sum_{n=1}^{\infty} C_1 n^{-3/2} E|X'_n - \mu_n|^3 / \sigma_n^3 \leq C_2 \sum_{n=1}^{\infty} n^{-3/2} E|X'_n|^3 < \infty$$

since

$$\begin{aligned} \sum_{n=1}^{\infty} n^{-3/2} E|X'_n|^3 &= 3 \sum_{n=1}^{\infty} n^{-3/2} \int_0^{n^{1/2}} t^2 P\{|X| > t\} dt \\ &= 3 \int_0^{\infty} t^2 P\{|X| > t\} \sum_{n \geq (t \vee 1)^2} n^{-3/2} dt \\ &= C_3 \int_1^{\infty} t P\{|X| > t\} dt < \infty. \end{aligned}$$

On the other hand, since $\sigma_n \rightarrow \sigma$, via integration by parts, $E X = 0$ and

4. Failure of the Lindeberg condition does not preclude asymptotic normality. Let $\{Y_n\}$ be i.i.d. with $E Y_n = 0, E Y_n^2 = 1$; let $\{Z_n\}$ be independent with $P\{Z_n = \pm n\} = 1/2n^2, P\{Z_n = 0\} = 1 - (1/n^2)$ and $\{Z_n\}$ independent of $\{Y_n\}$. Prove that if $X_n = Y_n + Z_n, S_n = \sum_i X_i$, then $S_n/\sqrt{n} \xrightarrow{d} N_{0,1}$ and the Lindeberg condition cannot hold. Explain why this does not contravene Theorem 1.
5. Let $\{Y_n, n \geq 1\}$ be i.i.d. r.v.s with finite variance σ^2 (say $\sigma^2 = 1$) and let $\{\sigma_n^2, n \geq 1\}$ be nonzero constants with $s_n^2 = \sum_i \sigma_i^2 \rightarrow \infty$. Show that the **weighted i.i.d. r.v.s** $\{\sigma_n Y_n, n \geq 1\}$ obey the central limit theorem, i.e., $(1/s_n) \sum_i \sigma_i Y_i \xrightarrow{d} N_{0,1}$ if $\sigma_n = o(s_n)$ and $E Y = 0$.
6. Let $\{X_n\}$ be independent with

$$P\{X_n = \pm 1\} = \frac{1}{2a}, P\{X_n = \pm n\} = \frac{1}{2}\left(1 - \frac{1}{a}\right)\frac{1}{n^2},$$

$$P\{X_n = 0\} = \left(1 - \frac{1}{a}\right)\left(1 - \frac{1}{n^2}\right), n \geq 1, a > 1.$$

Again, S_n/\sqrt{n} has a limiting normal distribution despite the Lindeberg condition being vitiated.

7. Prove that the Liapounov condition of Corollary 9.1.1 is more stringent the larger the value of δ .
8. (i) For what positive values of a , if any, does a C.L.T. hold for i.i.d. symmetric random variables with $F(x) = 1 - 1/(2x^a), x \geq 1, F(x) = \frac{1}{2}, 0 \leq x \leq 1$.
(ii) Does a C.L.T. hold for independent $\{X_n\}$ with $P\{X_n = \pm 1\} = \frac{1}{4}, P\{X_n = \pm n\} = 1/(2n^3), P\{X_n = 0\} = (1/2) - (1/n^3)$? Hint: Apply Theorems 4 and 1.
9. C.L.T. for U -statistics: Let $\{X_n\}$ be i.i.d. and $\varphi(x_1, \dots, x_m)$ a real, symmetric function of its m arguments. If $E \varphi^2(X_1, \dots, X_m) < \infty$ and $E \varphi(X_1, \dots, X_m) = \theta$, then $n^{1/2}(U_n - \theta) \xrightarrow{d} N_{0,\sigma^2}$, where $U_n = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \dots < i_m \leq n} \varphi(X_{i_1}, \dots, X_{i_m})$ and $\sigma^2 = m^2 E[\varphi^2(X_1, \dots, X_m) | X_1] - \theta^2$. Hint: A C.L.T. applies to

$$V_n = \frac{m}{n^{1/2}} \sum_{i=1}^n [E\{\varphi(X_i, X_2, \dots, X_m) | X_i\} - \theta]$$

since the components of the sum are i.i.d. with mean 0 and variance σ^2/m^2 , while $U_n - V_n \xrightarrow{d} 0$ via $E(U_n - V_n)^2 = o(1)$.

10. Let $X_n = (X_{n1}, \dots, X_{nk}), n \geq 1$, be i.i.d. random vectors with $E X_{nj} = \mu_j$,

$$\text{Cov}(X_{ni}, X_{nj}) = \sigma_{ij}, 1 \leq i \leq j \leq k.$$

If $S_{nj} = \sum_{i=1}^n X_{ij}, n \geq 1, 1 \leq j \leq k$, prove that

$$(n^{-1/2}(S_{n1} - n\mu_1), \dots, n^{-1/2}(S_{nk} - n\mu_k))$$

converges in distribution to the k -dimensional normal distribution with mean vector zero and covariance matrix $\{\sigma_{ij}\}$. Hint: Recall Exercise 8.5.6.

11. Let $\{X_n\}$ be independent r.v.s with $E X_n = 0, E X_n^2 = 1$ which obey the Lindeberg condition. If $n_j = [jn/k] = \text{greatest integer} \leq jn/k$ for $j = 0, 1, \dots, k$ and $S_n = \sum_{i=1}^n X_i$, prove that $((k/n)^{1/2}S_{n1}, \dots, (k/n)^{1/2}S_{nk})$ converges in distribution to the k -dimensional normal with mean vector zero and covariance matrix $\{\sigma_{ij}\}$, where $\sigma_{ij} = \min(i, j)$. Conclude that

$$\lim_{n \rightarrow \infty} P\left\{ \max_{1 \leq j \leq k} S_{n_j} < xn^{1/2} \right\} = P\left\{ \frac{1}{\sqrt{k}} \max_{1 \leq j \leq k} Y_j < x \right\},$$

where $\{Y_j, 1 \leq j \leq k\}$ are i.i.d. r.v.s with the standard normal d.f.

12. Prove that if $\{X_n, n \geq 1\}$ are independent r.v.s with $E X_n = 0, E X_n^2 = \sigma_n^2$ which obey central limit theorem and $\lim_{n \rightarrow \infty} E(s_n^{-1/2} \sum_1^n X_i)^{2k} = (2k)!/2^k k!$, then $\{X_n\}$ obeys a Lindeberg condition of order $2k$.
13. Let $\{X_n, n \geq 1\}$ be i.i.d. with symmetric density $f(x) = |x|^{-3} I_{\{|x| > 1\}}$. Show that $(n \log n)^{-1/2} \sum_1^n X_i \xrightarrow{d} N_{0,1}$.

9.2 Interchangeable Components

For an **infinite** sequence of interchangeable random variables $\{X_n\}$, Corollary 7.3.4 exhibits the joint distributions as mixtures of distributions corresponding to i.i.d. random variables and thereby provides a tool for proving central and other limit theorems.

Initially it will be assumed that $E X_1^2 < \infty$, whence no further generality is lost in supposing $E X_1 = 0, E X_1^2 = 1$.

Theorem 1. *Let $\{X_n, n \geq 1\}$ be interchangeable random values. If $E X = 0, E X^2 = 1$, and*

$$\text{Cov}(X_1, X_2) = 0 = \text{Cov}(X_1^2, X_2^2), \quad (1)$$

then $(1/\sqrt{n}) \sum_1^n X_i \xrightarrow{d} N_{0,1}$. Conversely, if $E X^2 < \infty$ and $(1/\sqrt{n}) \sum_1^n X_i \xrightarrow{d} N_{0,1}$, then (1) holds.

PROOF. According to Theorem 7.3.2 the r.v.s $\{X_n\}$ are conditionally i.i.d. given the σ -algebra \mathcal{G} of permutable events, and according to Corollary 7.3.5 there is a regular conditional distribution P^ω for X given \mathcal{G} such that the coordinate r.v.s $\{\xi_n, n \geq 1\}$ are i.i.d. on the probability space $(R^\omega, \mathcal{B}^\omega, P^\omega)$. Moreover, by this same corollary, for $i \neq j$

$$\begin{aligned} E X_i X_j &= E[E(X_i X_j | \mathcal{G})] = E[E^2(X_i | \mathcal{G})], \\ E[X_i^2 - 1][X_j^2 - 1] &= E[E^2(X_i^2 - 1 | \mathcal{G})] \end{aligned} \quad (2)$$

and so (1) is equivalent to

$$E\{X_i | \mathcal{G}\} = 0, \quad E\{X_i^2 | \mathcal{G}\} = 1, \quad i \geq 1, \quad \text{a.c.} \quad (1')$$

Consequently, for all ω

$$E_\omega(\xi_1) \equiv \int_{-\infty}^{\infty} \xi_1 dP^\omega = 0, \quad \sigma_\omega^2(\xi_1) \equiv \int_{-\infty}^{\infty} \xi_1^2 dP^\omega = 1,$$

and if $S_n = \sum_1^n X_i, T_n = \sum_1^n \xi_i$, then via Corollaries 7.2.2, 7.3.4, 7.3.5, and

dominated convergence

$$\begin{aligned} \lim_{n \rightarrow \infty} P\left\{\frac{S_n}{\sqrt{n}} < x\right\} &= \lim_{n \rightarrow \infty} \int P\left\{\frac{S_n}{\sqrt{n}} < x \mid \mathcal{G}\right\} dP = \lim_{n \rightarrow \infty} \int P^{\omega}\left\{\frac{T_n}{\sqrt{n}} < x\right\} dP \\ &= \int \lim_{n \rightarrow \infty} P^{\omega}\left\{\frac{T_n}{\sqrt{n}} < x\right\} dP, \end{aligned} \quad (3)$$

and so sufficiency follows from Corollary 9.1.2.

Conversely, if $S_n/\sqrt{n} \xrightarrow{d} N_{0,1}$, then $S_n/n \xrightarrow{P} 0$. By Example 1 of 7.3.1, $S_n/n \xrightarrow{a.s.} E\{X|\mathcal{G}\}$ and the first half of (1)' follows. Apropos of the second half,

$$\begin{aligned} e^{-t^2/2} &= \lim_{n \rightarrow \infty} E[E\{e^{itS_n/n^{1/2}}|\mathcal{G}\}] = \lim_{n \rightarrow \infty} EE_{\omega} e^{itT_n/n^{1/2}} \\ &= E e^{-\sigma_{\omega}^2(\xi)t^2/2} \end{aligned}$$

so that

$$E e^{-(1-\sigma_{\omega}^2(\xi))t^2/2} = 1, \quad \text{all } t \in (-\infty, \infty),$$

implying ($t \rightarrow \infty$) that $\sigma_{\omega}^2(\xi) \leq 1$, a.c. Consequently, $\sigma_{\omega}^2(\xi) = 1 = E\{X^2|\mathcal{G}\}$, a.c., whence (1)' and hence (1) is obtained. \square

If $S_n = \sum_1^n X_i$, $s_n^2 = E S_n^2$, it is quite possible that $S_n/s_n \xrightarrow{d} N_{0,1}$ despite violation of (1). The point is that if $\text{Cov}(X_1, X_2) = \rho^2 > 0$, the order of magnitude of S_n is n rather than $n^{1/2}$. Recall that X_F denotes a fictitious r.v. with d.f. F .

Theorem 2. If $\{X_n, n \geq 1\}$ are \mathcal{L}_1 interchangeable random variables with partial sums $S_n, n \geq 1$, then

- (i) $S_n/n \xrightarrow{d} S_G$, where G is the d.f. of $E\{X_1|\mathcal{G}\}$ and \mathcal{G} is the σ -algebra of permutable events,
- (ii) if F is any distribution function uniquely determined by its moments $\alpha_k, k \geq 1$, and $E X_1 X_2 \cdots X_k$ exists, all $k \geq 1$, then $S_n/n \xrightarrow{d} S_F$ iff $E X_1 X_2 \cdots X_k = \alpha_k, k \geq 1$.

Corollary 1. If $\{X_n, n \geq 1\}$ are \mathcal{L}_1 interchangeable random variables with $\text{Cov}(X_1, X_2) = \rho^2 > 0$ and $E X_1 X_2 \cdots X_k$ exists for all $k \geq 1$ then $(1/\rho n) \sum_1^n X_i \xrightarrow{d} N_{0,1}$ iff $E X_1 X_2 \cdots X_k$ vanishes for odd integers and equals $1 \cdot 3 \cdots (k-1)\rho^k$ for even integers k .

PROOF OF THEOREM 2. (i) is an immediate consequence of the strong law of large numbers for interchangeable r.v.s (Example 7.3.1). To prove (ii) note via Corollary 7.3.5 that, for all $k \geq 1$,

$$E X_1 \cdots X_k = E[E\{X_1 \cdots X_k|\mathcal{G}\}] = E[E\{X_1|\mathcal{G}\}]^k. \quad (4)$$

Thus, if $S_n/n \xrightarrow{d} S_F$, where F has finite moments $\alpha_k, k \geq 1$, part (i) ensures that the right and hence the left sides of (4) are α_k .

Conversely, if the left and hence the right sides of (4) equal α_k , then (i) and the Frechet-Shohat theorem 8.2.1 guarantee that $S_n/n \xrightarrow{d} S_F$. \square

Unfortunately, interchangeable r.v.s encountered in practical situations of interest are likely to be finite in number and not embeddable in an infinite sequence, so that Corollary 7.3.4 and prior results are inapplicable. A case of this sort occurred in Chapter 3.1 with the random distribution of balls into cells.

Suppose that $N = N_n$ balls are distributed at random into n cells and set $X_{nj} = 1$ or 0 according as the j th cell, $1 \leq j \leq n$, is or is not empty. Then $\{X_{nj}, 1 \leq j \leq n, n \geq 1\}$ constitute a double sequence of r.v.s which, for each $n > 1$, form a finite collection of interchangeable r.v.s with

$$\text{Cov}(X_{n1}, X_{n2}) = \left(1 - \frac{2}{n}\right)^N - \left(1 - \frac{1}{n}\right)^{2N} < 0. \quad (5)$$

Thus, recalling Exercises 7.3.6 and 6.3.10, $\{X_{nj}, 1 \leq j \leq n\}$ is not embeddable in an infinite sequence of interchangeable r.v.s. Nonetheless, asymptotic normality of the distribution of the number of empty cells, i.e., $\sum_{j=1}^n X_{nj}$, can be proved by more *ad hoc* methods.

By way of preliminaries, set $U = U_n = \sum_1^n X_{nj}$ and note that

$$\mathbb{E} U = \sum_{j=1}^n \mathbb{E} X_{nj} = n \left(1 - \frac{1}{n}\right)^N \quad (6)$$

and from (5)

$$\begin{aligned} \sigma_U^2 &= n \left[\left(1 - \frac{1}{n}\right)^N - \left(1 - \frac{1}{n}\right)^{2N} \right] + n(n-1) \left[\left(1 - \frac{2}{n}\right)^N - \left(1 - \frac{1}{n}\right)^{2N} \right] \\ &= n \left[\left(1 - \frac{1}{n}\right)^N + (n-1) \left(1 - \frac{2}{n}\right)^N - n \left(1 - \frac{1}{n}\right)^{2N} \right] \leq n \left(1 - \frac{1}{n}\right)^N. \end{aligned} \quad (7)$$

Let $S_j = S_j^{(n)}$ denote the waiting time until the occupation of the j th new cell by a ball. Set $S_0 = 0$ and $Y_{n,j} = S_j - S_{j-1}$, $j \geq 1$. Clearly, $\{Y_{n,j}, 1 \leq j \leq n\}$ are independent with $Y_{n,1} = S_1 = 1$ and

$$\mathbb{P}\{Y_{n,j} = i\} = \left(\frac{j-1}{n}\right)^{i-1} \left(1 - \frac{j-1}{n}\right), \quad j \geq 2, \quad i \geq 1. \quad (8)$$

That is, $\{Y_{n,j} - 1, 2 \leq j \leq n\}$ are independent geometric r.v.s.

At most $n-k$ empty cells, i.e., at least k occupied cells in the random casting of N balls into n cells is tantamount to the k th new cell being entered by the N th or a prior ball. Thus, for $2 \leq k \leq n$

$$\mathbf{P}\{S_k \leq N\} = \mathbf{P}\{U_n \leq n - k\}, \quad (9)$$

and the possibility arises of shunting the asymptotic normality of a sum U_n of interchangeable r.v.s to one S_k of independent r.v.s.

Theorem 3. Let $U = U_n$ designate the number of empty cells engendered by distributing $N = N_n$ balls at random into n cells, and set $a = a_n = N/n$, $b = b_n = (e^a - 1 - a)^{1/2}$, and $\sigma = \sigma_{U_n}$.

(i) If $N \rightarrow \infty$ and $a/n \rightarrow 0$ as $n \rightarrow \infty$, then

$$\sigma^2 = ne^{-2a}b^2(1 + o(1)). \quad (10)$$

(ii) $\sigma \rightarrow \infty$ iff $ne^{-2a}b^2 \rightarrow \infty$, in which case $a = o(n)$, $N \rightarrow \infty$, (10) holds, and

$$\frac{U_n - \mathbf{E} U_n}{\sigma} \xrightarrow{d} N_{0,1}, \quad (11)$$

$$\frac{U_n - ne^{-a}}{\sqrt{ne^{-a}b}} \xrightarrow{d} N_{0,1}. \quad (12)$$

PROOF. (i) Since for $-\infty < \alpha < \infty$

$$\left(1 - \frac{\alpha}{n}\right)^n = \exp\left\{n\left(-\frac{\alpha}{n} - \frac{\alpha^2}{2n^2} + O(n^{-3})\right)\right\} = \exp\left\{-\alpha - \frac{\alpha^2}{2n} + O(n^{-2})\right\}, \quad (13)$$

it follows via (7) that

$$\begin{aligned} \sigma^2 &= n \left\{ \left(1 - \frac{1}{n}\right)^{na} + (n-1) \left(1 - \frac{2}{n}\right)^{na} - n \left(1 - \frac{1}{n}\right)^{2na} \right\} \\ &= ne^{-2a} \left\{ e^a \exp\left(-\frac{a}{2n}(1 + o(1))\right) + n \left[\exp\left(-\frac{2a}{n} + O(an^{-2})\right) \right. \right. \\ &\quad \left. \left. - \exp\left(-\frac{a}{n} + O(an^{-2})\right) \right] - \exp\left(-\frac{2a}{n} + O(an^{-2})\right) \right\} \\ &= ne^{-2a} \left[e^a - a - 1 - \frac{ae^a}{2n}(1 + o(1)) + O(an^{-1}) + O(an^{-2}) \right] \\ &= ne^{-2a}b^2 - \frac{ae^{-a}}{2}(1 + o(1)) + O(ae^{-2a}) \\ &= ne^{-2a}b^2 + O(ae^{-a}) \end{aligned} \quad (14)$$

under the assumptions that $n \rightarrow \infty$ and $a = o(n)$. Moreover,

Thus, the Liapounov and hence also the Lindeberg condition of Exercise 9.1.2 holds. \square

Remarks. (i) The denominator in (12) can be replaced by $\sqrt{ne^{-a}}(e^a - 1 - \alpha)^{1/2}$ if $a \rightarrow \alpha \in (0, \infty)$, by $a(n/2)^{1/2}$ if $a \rightarrow 0$, and by $(ne^{-a})^{1/2}$ if $a \rightarrow \infty$. (ii) If $a/n > \varepsilon > 0$, then $\sigma^2 \leq ne^{-a} \rightarrow 0$ via (7); if $N = na \leq C < \infty$, then $n^{1/2}a = o(1)$ and $\sigma^2 = (na^2/2)(1 + o(1)) + O(a) \rightarrow 0$ via (14). (iii) If $\sigma^2 \rightarrow \sigma_0^2 \in (0, \infty)$ and $a \rightarrow \infty$, then via (10) $a = \log n - \log \sigma_0^2 + o(1)$; if $\sigma^2 \rightarrow \sigma_0^2 \in (0, \infty)$ and $a \rightarrow 0$, then $a^2 = (2\sigma_0^2/n)(1 + o(1))$. Here, the possible limiting distributions can be completely determined (see Theorem 4, Theorem 3.1.4, and Exercise 6). (iv) If $\sigma^2 \rightarrow 0$, either $a \rightarrow 0$ or $a \rightarrow \infty$ and the limit distributions are degenerate (Exercise 6).

Theorem 4. If U_n is the number of empty cells in distributing N_n balls at random into n cells and $N_n^2/n = N_n a_n \rightarrow \delta^2 \in [0, \infty)$, then

$$\lim_{n \rightarrow \infty} P\{U_n - (n - N_n) = j\} = \frac{(\delta^2/2)^j e^{-\delta^2/2}}{j!}, \quad j = 0, 1, \dots \quad (26)$$

PROOF. According to (9) and (8) with $N = N_{n,j}$

$$P\{U_n \leq n - N + k\} = P\{S_{N-k} \leq N\} = P\{S_{N-k} - (N - k) \leq k\},$$

where $S_{N-k} - (N - k) = \sum_{j=1}^{N-k} (Y_j - 1)$ and $\{Y_j - 1, j \geq 2\}$ are independent, geometric r.v.s with success probabilities $1 - (j-1)/n$. Hence, the characteristic function of $S_{N-k} - (N - k)$, say $\phi(t)$, is given by

$$\begin{aligned} \phi(t) &= \prod_{j=1}^{N-k} \left(1 - \frac{j-1}{n}\right) \left[1 - \frac{j-1}{n} e^{it}\right]^{-1} \\ &= \prod_{j=1}^{N-k} \exp \left\{ -\frac{j-1}{n} + \frac{j-1}{n} e^{it} + O\left(\frac{j^2}{n^2}\right) \right\} \\ &= \exp \left\{ \frac{N_n^2}{2n} (e^{it} - 1) + O\left(\frac{N_n^3}{n^2}\right) \right\} \\ &\rightarrow \exp \left\{ \frac{\delta^2}{2} (e^{it} - 1) \right\}, \end{aligned}$$

and Corollary 8.3.5 guarantees that the d.f. of $U_n - (n - N_n)$ tends to a Poisson d.f. with parameter $\delta^2/2$, yielding (26). \square

EXERCISES 9.2

1. For an arbitrary d.f. G , concoct a sequence of interchangeable r.v.s for which $P\{S_n < x\sqrt{n}\} \rightarrow \int \Phi(x/y) dG(y)$.
2. If $\{X_n, n \geq 1\}$ are interchangeable r.v.s with $E X_1 = 0$, $E X_1^2 = 1$, $E|X_1|^3 < \infty$, $\text{Cov}(X_1, X_2) = 0 = \text{Cov}(X_1^2, X_2^2)$, then Corollary 9.1.4 holds.
3. Suppose that for each $n = 1, 2, \dots$, $\{X_{ni}, i = 1, 2, \dots\}$ constitute interchangeable random variables with $E X_{n1} = 0$, $E X_{n1}^2 = 1$, $E|X_{n1}|^3 < \infty$ and set

$$m_n(\omega) = E\{X_{n1} | \mathcal{G}_n\}, \sigma_n^2(\omega) = E[(X_{n1} - m_n)^2 | \mathcal{G}_n], \quad \alpha_n(\omega) = E\{|X_{n1} - m_n|^3 | \mathcal{G}_n\}$$

when \mathcal{G}_n is the σ -algebra relative to which $\{X_{ni}, i \geq 1\}$ are conditionally i.i.d. If $\sqrt{n}m_n(\omega) \xrightarrow{P} 0$, $\sigma_n(\omega) \xrightarrow{P} 1$, $\alpha_n(\omega)/\sqrt{n}\sigma_n^3(\omega) \xrightarrow{P} 0$, and $S_n = \sum_{i=1}^n X_{ni}$, then $S_n/\sqrt{n} \xrightarrow{d} N_{0,1}$.

4. If for each $n \geq 1$ $\{X_{ni}, i \geq 1\}$ are interchangeable r.v.s with $E X_{n1} = 0$, $E X_{n1}^2 = 1$, $E|X_{n1}|^3 < \infty$ and $E X_{n1} X_{n2} = o(n^{-1})$, $E X_{n1}^2 X_{n2}^2 \rightarrow 1$, $E|X_{n1}|^3 = o(n^{1/2})$, then $S_n/\sqrt{n} \xrightarrow{d} N_{0,1}$.
5. If two of the last three conditions of Exercise 4 obtain but the third is violated, construct a sequence $\{X_{ni}, i \geq 1\}$ of interchangeable processes for which $S_n/\sqrt{n} \not\xrightarrow{d} N_{0,1}$.
6. Apropos of Theorems 3 and 4, prove that (i) if $a_n - \log n \rightarrow \infty$, $P\{\sum_i^n X_{ni} = 0\} \rightarrow 1$, (ii) if $Na_n \rightarrow 0$, $P\{\sum_i^n X_{ni} = n - N\} \rightarrow 1$, (iii) if $a_n - \log n \rightarrow \delta$ finite and P_λ designates a Poisson r.v. with mean λ , $\sum_{i=1}^n X_{ni} \xrightarrow{d} P_{\exp\{-\delta\}}$ (see Theorem 3.1.2).
7. Let the i.i.d. r.v.s $\{Z_n, n \geq 1\}$ be independent of Y , where Y is uniformly distributed on $(0, 1)$ and $P\{Z_n = \pm 1\} = \frac{1}{2}$. Then $X_n = Y^{-1/2}Z_n$, $n \geq 1$, are \mathcal{L}_1 interchangeable r.v.s with $(1/n)\sum_i^n X_i \xrightarrow{a.s.} 0$ but $E X_1 X_2$ does not exist.

9.3 The Martingale Case

If the component r.v.s $\{X_n, n \geq 1\}$ are **martingale differences**, that is,

$$E\{X_{n+1} | \mathcal{F}_n\} = 0, n \geq 1,$$

and, moreover, obey the Lindeberg condition, a central limit theorem will hold provided the conditional variances are sufficiently well behaved.

Theorem 1. If $\{X_n, \mathcal{F}_n, n \geq 1\}$ is a stochastic sequence with $E\{X_{n+1} | \mathcal{F}_n\} = 0$, $n \geq 0$ (with $\mathcal{F}_0 = \{\emptyset, \Omega\}$), such that $\{X_n\}$ obey the Lindeberg condition and, setting $\sigma_n^2 = E X_n^2$, $s_n^2 = \sum_{j=1}^n \sigma_j^2$,

$$\sum_{j=1}^n E|E\{X_j^2 | \mathcal{F}_{j-1}\} - \sigma_j^2| = o(s_n^2), \quad (1)$$

then $(1/s_n)\sum_{j=1}^n X_j \xrightarrow{d} N_{0,1}$.

PROOF. The proof of Theorem 9.1.1 may easily be adapted to handle the current situation. In fact, let $Y_j(t)$, $a_j(t)$ be exactly as defined there and set

$$b_j(t) = \frac{t^2}{2} [E\{X_j^2 | \mathcal{F}_{j-1}\} - \sigma_j^2].$$

Then 9.1(5) holds provided (i) all expectations E therein are changed to $E\{\cdot | \mathcal{F}_{j-1}\}$, (ii) the term $b_j(t/s_n)$ is added to the expression within the absolute value signs on the right, and (iii) the term $|b_j(t/s_n)|$ is added within the brackets

on the extreme right. Analogously, (6) obtains if the alteration (iii) is effected and E is replaced by $E\{\cdot|\mathcal{F}_{j-1}\}$ in the second expectation of the second expression (where independence was used) and in all succeeding expressions of (6). Finally, (7) holds if $E\sum_{j=1}^n |b_j(t/s_n)|$ is appended to the extreme right side and so, recalling (1), the theorem follows. \square

A version of the Berry-Esseen theorem also carries over to martingales.

Theorem 2. If $\{S_n = \sum_1^n X_i, n \geq 1\}$ is an \mathcal{L}_3 martingale with $E S_n = 0$, $E S_n^2 = s_n^2 = \sum_1^n \sigma_i^2$, $n \geq 1$, there exists an absolute constant c_0 such that

$$\left| P\left\{ \frac{S_n}{s_n} < x \right\} - \Phi(x) \right| \leq c_0 \left[\frac{1}{s_n^3} \sum_1^n E|X_j|^3 \right]^{1/4} + \left[c_0 - \frac{1}{\sqrt{2\pi}} \right] \frac{\sum_1^n E|E\{X_j^2|X_1, \dots, X_{j-1}\} - \sigma_j^2|}{s_n^{1/2} \sqrt{\sum_1^n E|X_j|^3}}.$$

Lemma 1. If f is a real function with $f^{(n-1)}$ absolutely continuous on $[a, b]$, then, setting $f^{(0)} = f$,

$$f(b) = \sum_{j=0}^{n-1} \frac{(b-a)^j}{j!} f^{(j)}(a) + \frac{1}{(n-1)!} \int_a^b (b-x)^{n-1} f^{(n)}(x) dx. \quad (2)$$

PROOF. Set

$$F(x) = f(b) - \sum_{j=0}^{n-1} \frac{(b-x)^j}{j!} f^{(j)}(x), \quad (3)$$

whence $F'(x) = [-(b-x)^{n-1}/(n-1)!] f^{(n)}(x)$, a.e. on $[a, b]$. Thus, F is absolutely continuous on $[a, b]$ and

$$F(a) = F(b) - \int_a^b F'(x) dx = \frac{1}{(n-1)!} \int_a^b (b-x)^{n-1} f^{(n)}(x) dx,$$

which is tantamount to (2) in view of (3). \square

Lemma 2. If $\{S_n = \sum_1^n X_i, n \geq 1\}$ is as in Theorem 2, then for any function f with absolutely continuous second derivative on every bounded interval and some absolute constant C in $(0, \frac{2}{3}]$

$$\begin{aligned} & \left| E f\left(\frac{S_n}{s_n}\right) - E f(Y_0) \right| \\ & \leq C \left[\frac{\|f^{(3)}\|}{s_n^3} \sum_{j=1}^n E|X_j|^3 + \frac{\|f^{(2)}\|}{s_n^2} \sum_{j=1}^n E|E\{X_j^2|X_1, \dots, X_{j-1}\} - \sigma_j^2| \right], \end{aligned} \quad (4)$$

where Y_0 is normal with mean 0, variance 1, and $\|f^{(j)}\| = \inf[M : \mu\{|f^{(j)}(x)| \geq M\} = 0]$, $1 \leq j \leq 3$, with μ denoting Lebesgue measure on the real line.

PROOF. Let $\{Y_n, n \geq 0\}$ be independent normal r.v.s with means zero and variances σ_n^2 ($\sigma_0^2 = 1$) with $\{Y_n, n \geq 0\}$ independent of $\{X_n, n \geq 1\}$.

Set $Q_{j,n} = \sum_{i=1}^{j-1} X_i + \sum_{i=j+1}^n Y_i$, $1 \leq j \leq n$, and note that $Q_{j,n} + X_j = Q_{j+1,n} + Y_{j+1}$. By Lemma 1

$$\left| f(a+h) - f(a) - hf'(a) - \left(\frac{h^2}{2}\right) f^{(2)}(a) \right| \leq \frac{|h|^3}{6} \|f^{(3)}\|,$$

whence for some r.v.s θ_j with $|\theta_j| \leq 1$, $j = 1, 2$,

$$\begin{aligned} E f\left(\frac{S_n}{s_n}\right) - E f(Y_0) &= \sum_{j=1}^n E \left[f\left(\frac{Q_{j,n} + X_j}{s_n}\right) - f\left(\frac{Q_{j,n} + Y_j}{s_n}\right) \right] \\ &= \sum_{j=1}^n E \left[\frac{X_j - Y_j}{s_n} f'\left(\frac{Q_{j,n}}{s_n}\right) + \frac{X_j^2 - Y_j^2}{2s_n^2} f^{(2)}\left(\frac{Q_{j,n}}{s_n}\right) \right. \\ &\quad \left. + \frac{\theta_j(|X_j|^3 + |Y_j|^3)}{6s_n^3} \|f^{(3)}\| \right]. \end{aligned} \quad (5)$$

Since $E|Y_j|^3 = c_1 \sigma_j^3 = c_1 (E X_j^2)^{3/2} \leq c_1 E|X_j|^3$ and

$$\begin{aligned} E(X_j^i - Y_j^i) f^{(i)}\left(\frac{Q_{j,n}}{s_n}\right) \\ = \begin{cases} E \left\{ f^{(i)}\left(\frac{Q_{j,n}}{s_n}\right) E\{X_j^i - Y_j^i | X_1, \dots, X_{j-1}, Y_{j+1}, \dots, Y_n\} \right\} = 0, & i = 1 \\ E f^{(2)}\left(\frac{Q_{j,n}}{s_n}\right) [E\{X_j^2 | X_1, \dots, X_{j-1}\} - \sigma_j^2], & i = 2, \end{cases} \end{aligned}$$

the conclusion of the lemma follows from (5). \square

PROOF OF THEOREM 2. Let $\varepsilon > 0$ and define $h(-t) = -h(t)$, where

$$h(t) = \begin{cases} \frac{16t^3}{3\varepsilon^3} - \frac{2t}{\varepsilon}, & 0 \leq t \leq \frac{\varepsilon}{4} \\ \frac{16}{3\varepsilon^3} \left(\frac{\varepsilon}{2} - t\right)^3 - \frac{1}{2}, & \frac{\varepsilon}{4} \leq t \leq \frac{\varepsilon}{2} \\ -\frac{1}{2}, & t \geq \frac{\varepsilon}{2}. \end{cases} \quad (6)$$

Then h is a nonincreasing odd function with h'' absolutely continuous and $\|h''\| = 8\varepsilon^{-2}$, $\|h'''\| = 32\varepsilon^{-3}$, and consequently

$$f(t) = h\left(t - x - \frac{\varepsilon}{2}\right) + \frac{1}{2} \quad (7)$$

9.4 Miscellaneous Central Limit Theorems

The first central limit theorem concerns sums of random numbers, say t_n , of independent random variables $\{X_n, n \geq 1\}$ and permits t_n to be highly dependent upon the sequence $\{X_n\}$.

Theorem 1 (Doeblin–Anscombe). *Let $\{X_n, n \geq 1\}$ be independent random variables with $E X_n = 0$, $E X_n^2 = 1$, and $\{t_n, n \geq 1\}$ positive integer-valued random variables with $t_n/b_n \xrightarrow{P} c$, where $\{c, b_n, n \geq 1\}$ are positive, finite constants such that $b_n \uparrow \infty$. If $n^{-1/2} \sum_{j=1}^{t_n} X_j \xrightarrow{d} N_{0,1}$, then*

$$\frac{1}{t_n^{1/2}} \sum_{j=1}^{t_n} X_j \xrightarrow{d} N_{0,1}.$$

PROOF. Set $S_n = \sum_1^n X_j$ and $k_n = [cb_n] = \text{greatest integer } \leq cb_n$.

Now,

$$\frac{S_{t_n}}{t_n^{1/2}} = \left(\frac{k_n}{t_n} \right)^{1/2} \left[\frac{S_{k_n}}{k_n^{1/2}} + \frac{S_{t_n} - S_{k_n}}{k_n^{1/2}} \right] \quad (1)$$

and according to the hypothesis the first factor on the right side of (1) converges in probability to one. Moreover, $S_{k_n}/k_n^{1/2}$ as a subsequence of $S_n/n^{1/2}$ converges in distribution to $N_{0,1}$. Thus, to prove the theorem it suffices to establish that

$$\frac{S_{t_n} - S_{k_n}}{k_n^{1/2}} \xrightarrow{P} 0. \quad (2)$$

To this end, note that for any positive numbers ε, δ

$$\begin{aligned} P\{|S_{t_n} - S_{k_n}| > \varepsilon k_n^{1/2}\} &\leq P\{|S_{t_n} - S_{k_n}| > \varepsilon k_n^{1/2}, |t_n - k_n| \leq \delta k_n\} \\ &\quad + P\{|t_n - k_n| > \delta k_n\}. \end{aligned} \quad (3)$$

Since the event corresponding to the first term on the right implies $A_n^+ \cup A_n^-$, where

$$\begin{aligned} A_n^+ &= \left\{ \max_{k_n \leq j \leq (1+\delta)k_n} |S_j - S_{k_n}| > \varepsilon k_n^{1/2} \right\}, \\ A_n^- &= \left\{ \max_{k_n(1-\delta) \leq j \leq k_n} |S_j - S_{k_n}| > \varepsilon k_n^{1/2} \right\}, \end{aligned}$$

and since by Kolmogorov's inequality

$$P(A_n^\pm) \leq (\varepsilon^2 k_n)^{-1} E(S_{k_n \pm \delta k_n} - S_{k_n})^2 \leq (\varepsilon^2 k_n)^{-1} \delta k_n = \varepsilon^{-2} \delta,$$

it follows from (3) that

$$P\{|S_{t_n} - S_{k_n}| > \varepsilon k_n^{1/2}\} \leq 2\varepsilon^{-2} \delta + o(1). \quad (4)$$

For arbitrary $\varepsilon > 0$, the first term on the right in (4) tends to zero with δ and the theorem follows. \square

Actually, Theorem 1 holds even if t_n/n converges in probability to a positive random variable (Blum *et al.*, 1962–1963; Mogyorodi, 1962).

If r.v.s $\{X_n, n \geq 1\}$ obey the hypothesis of the ensuing theorem, then $N_c = \sup\{n \geq 1 : \sum_{i=1}^n X_i \leq c\}$, is a bona fide r.v. In the special case of classical renewal theory (Section 5.4), where $X_n > 0, n \geq 1$, implies $N_c + 1 = T_c(0)$ [see (5)], asymptotic normality of N_c is an immediate consequence of Theorem 2 below. On the other hand, if merely $E X_n \equiv \mu > 0$, then only the inequality $N_c \geq T_c(0) - 1$ may be inferred. Nonetheless, N_c is still asymptotically normal under the hypothesis of Corollary 2.

Theorem 2. If $\{X, X_n, n \geq 1\}$ are i.i.d. r.v.s with $E X = \mu > 0$ and $\sigma^2 = \sigma_X^2 \in (0, \infty)$, define $S_n = \sum_1^n X_i$ and

$$T = T_c = T_c(\alpha) = \inf\{n \geq 1 : S_n > cn^\alpha\}, \quad -\infty < \alpha < 1, c > 0. \quad (5)$$

Then, as $c \rightarrow \infty$

$$\frac{\mu(1-\alpha)[T_c(\alpha) - (c/\mu)^{1/(1-\alpha)}]}{\sigma(c/\mu)^{1/2(1-\alpha)}} \xrightarrow{d} N_{0,1}. \quad (6)$$

PROOF. For simplicity, take $\mu = 1$. Since via the strong law of large numbers $T_c/c^{1/(1-\alpha)} \xrightarrow{a.s.} 1$ as $c \rightarrow \infty$ (Exercise 5.4.6), by Theorems 1 and 8.1.1

$$\left(\frac{T}{c^{1/(1-\alpha)}}\right)^{1/2} \frac{(S_T - T)}{\sigma\sqrt{T}} = \frac{S_T - cT^\alpha}{\sigma c^{1/2(1-\alpha)}} + \frac{cT^\alpha - T}{\sigma c^{1/2(1-\alpha)}} \xrightarrow{d} N_{0,1}. \quad (7)$$

Since $E X_1^2 < \infty$, $X_n/\sqrt{n} \xrightarrow{a.s.} 0$, implying $X_T/\sqrt{T} \xrightarrow{a.s.} 0$ as $c \rightarrow \infty$, whence

$$0 \leq \frac{S_T - cT^\alpha}{\sigma c^{1/2(1-\alpha)}} \leq \frac{X_T}{\sqrt{T}} \frac{\sqrt{T}}{\sigma c^{1/2(1-\alpha)}} \xrightarrow{a.s.} 0,$$

and (7) ensures

$$\frac{cT^\alpha - T}{\sigma c^{1/2(1-\alpha)}} \xrightarrow{d} N_{0,1}.$$

Now $Z_c = 1 + o(1)$ entails $Z_c^{1-\alpha} = 1 + (Z_c - 1)(1 - \alpha + o(1))$, so that

$$\begin{aligned} cT^\alpha - T &= T([T^{-1}c^{1/(1-\alpha)}]^{1-\alpha} - 1) \\ &= T[T^{-1}c^{1/(1-\alpha)} - 1](1 - \alpha + o(1)) \\ &= -(1 - \alpha)[T - c^{1/(1-\alpha)}](1 + o(1)), \end{aligned}$$

and the theorem follows. \square

Corollary 1. Under the hypothesis of Theorem 2, for $-\infty < \alpha < 1$

$$\frac{\max_{1 \leq j \leq n}(S_j/j^\alpha) - \mu n^{1-\alpha}}{\sigma n^{(1/2)-\alpha}} \xrightarrow{d} N_{0,1}. \quad (8)$$

PROOF. For $x \neq 0$, define

$$c = c_x = n^{1-\alpha}\mu - x\sigma n^{(1/2)-\alpha}. \quad (9)$$

$\{X_n\}$ is independent of the underlying distributions F_n of X_n , $n \geq 1$, (ii) evaluating this limit distribution for an expeditious choice of $\{F_n\}$.

By combining their technique with a sequence of stopping times, it is possible to eliminate step (ii) in

Theorem 3. If $\{X_n, n \geq 1\}$ are independent r.v.s with $E X_n = 0$, $E X_n^2 = \sigma^2 \in (0, \infty)$, $n \geq 1$, which satisfy the Lindeberg condition, $S_j = \sum_{i=1}^j X_i$, and $T_c^* = \inf\{j \geq 1 : S_j > c\}$, then as $c \rightarrow \infty$, $(\sigma^2/c^2)T_c^*$ converges in distribution to the positive stable distribution of characteristic exponent $\frac{1}{2}$ or equivalently $(1/\sigma n^{1/2})\max_{1 \leq j \leq n} S_j$ converges to the positive normal distribution, i.e., for $x > 0, y > 0$

$$\lim_{c \rightarrow \infty} P\left\{T_c^* > \frac{c^2 y}{\sigma^2}\right\} = \lim_{n \rightarrow \infty} P\left\{\max_{1 \leq j \leq ny} S_j \leq x \sigma n^{1/2}\right\} = 2\Phi\left(\frac{x}{y^{1/2}}\right) - 1. \quad (10)$$

Note. For $y = 1$, the right side of (10) is a d.f. in x , namely, the positive normal distribution, while for $x = 1$ it is one minus a d.f. in y ; the latter, $2[1 - \Phi(y^{-1/2})]$, $y > 0$, is the so-called positive stable distribution of characteristic exponent $\frac{1}{2}$ (Chapter 12).

PROOF. Without loss of generality, take $\sigma = 1 = y$ and let $x > 0$. In view of the Lindeberg condition, for every $\delta > 0$

$$\sum_{j=1}^n P\{|X_j| > \delta \sqrt{n}\} \leq \frac{1}{n \delta^2} \sum_{j=1}^n E X_j^2 I_{\{|X_j| > \delta \sqrt{n}\}} = o(1). \quad (11)$$

For any positive integer k , set $n_j = [jn/k]$, $j = 0, 1, \dots, k$ and $n = k, k+1, \dots$. If

$$Y_j = Y_{n_j, j} = X_{n_j+1} + \dots + X_{n_{j+1}}, \quad j = 0, 1, \dots, k-1,$$

then $S_n = S_{n_k} = \sum_{i=0}^{k-1} Y_i$. Moreover, Y_0, \dots, Y_{k-1} are independent r.v.s and, furthermore, for fixed $j = 0, 1, \dots, k-1$ as $n \rightarrow \infty$

$$\frac{1}{\sigma^2(Y_j)} \sum_{i=n_j+1}^{n_{j+1}} E X_i^2 I_{\{|X_i| > \delta \sigma(Y_j)\}} \leq \frac{2k}{n} \sum_{i=n_j+1}^{n_{j+1}} E X_i^2 I_{\{|X_i| > (\delta/2)\sqrt{n/k}\}} = o(1). \quad (12)$$

Consequently, as $n \rightarrow \infty$ the r.v.s $Y_{n,j}/\sqrt{n} \xrightarrow{d} N_{0, 1/k}$ for $j = 0, 1, \dots, k-1$.

Next, for each $i = 1, 2, \dots, n$ let $m(i)$ be the integer for which $n_{m(i)-1} < i \leq n_{m(i)}$ and note that $0 < m(i) \leq k$. For any $\varepsilon > 0$, setting $A_i = A_{i,n}(\varepsilon) = \{|S_{n_{m(i)}} - S_i| \geq \varepsilon \sqrt{n}\}$, $1 \leq i \leq n$, and omitting the * in T_c^*

$$\begin{aligned} P\{T_{x\sqrt{n}} \leq n\} - P\{S_n > x\sqrt{n}\} &= P\left\{S_n \leq x\sqrt{n}, \max_{1 \leq m \leq n} S_m > x\sqrt{n}\right\} \\ &\leq \sum_{i=1}^{n-1} P\{T_{x\sqrt{n}} = i, A_i\} + \sum_{i=1}^{n-1} P\{S_n \leq x\sqrt{n}, T_{x\sqrt{n}} = i, A_i^c\} \\ &= D + B \text{ (say). Now, via Kolmogorov's inequality} \end{aligned} \quad (13)$$

$$\begin{aligned}
D &= \sum_{i=1}^{n-1} P\{T_{x\sqrt{n}} = i\} P\{A_i\} \\
&\leq \sum_{i=1}^{n-1} \frac{1}{n\varepsilon^2} (n_{m(i)} - n_{m(i)-1}) P\{T_{x\sqrt{n}} = i\} \leq \frac{1}{n\varepsilon^2} \frac{2n}{k} = \frac{2}{k\varepsilon^2}. \quad (14)
\end{aligned}$$

On the other hand, recalling that $n_k = n$ and $S_n/\sqrt{n} \xrightarrow{d} N_{0,1}$,

$$\begin{aligned}
B &\leq \sum_{i=1}^{n-1} [P\{S_n \leq (x - \varepsilon)\sqrt{n}, T_{x\sqrt{n}} = i, A_i^c\} \\
&\quad + P\{(x - \varepsilon)\sqrt{n} < S_n \leq x\sqrt{n}, T_{x\sqrt{n}} = i\}] \\
&\leq \sum_{i=1}^{n-1} P\{T_{x\sqrt{n}} = i, S_{n_{m(i)}} - S_n > 0\} + P\{(x - \varepsilon)\sqrt{n} < S_n \leq x\sqrt{n}\} \\
&= \sum_{i=1}^{n-1} P\{T_{x\sqrt{n}} = i\} P\{S_{n_{m(i)}} - S_n > 0\} + O(\varepsilon). \quad (15)
\end{aligned}$$

By virtue of $n^{-1/2} Y_{n,j} \xrightarrow{d} N_{0,1/k}$, $j = 0, 1, \dots, k-1$, and independence, for fixed k as $n \rightarrow \infty$ necessarily

$$n^{-1/2}(S_{n_k} - S_{n_{m(i)}}) = n^{-1/2} \sum_{j=m(i)}^{k-1} Y_j \xrightarrow{d} N_{0,(k-m(i))/k} \text{ for } i = 1, 2, \dots, n_{k-1}.$$

Noting that $1 \leq m(i) \leq k$ (despite $1 \leq i < n \rightarrow \infty$)

$$P\{S_{n_{m(i)}} - S_{n_k} > 0\} \rightarrow \frac{1}{2} \text{ uniformly for } 1 \leq i \leq n_{k-1}. \quad (16)$$

On the other hand, for $n_{k-1} < i < n$ necessarily $m(i) = k$, whence the left side of (16) is zero. Consequently, from (15) and (16)

$$\begin{aligned}
B &\leq \frac{1}{2} \sum_{i=1}^{n-1} P\{T_{x\sqrt{n}} = i\} + o(1) + O(\varepsilon) \\
&\leq \frac{1}{2} P\{T_{x\sqrt{n}} \leq n\} + o(1) + O(\varepsilon), \quad (17)
\end{aligned}$$

and so, combining (13), (14), and (17),

$$\frac{1}{2} \lim_{n \rightarrow \infty} P\{T_{x\sqrt{n}} \leq n\} \leq 1 - \Phi(x) + \frac{2}{k\varepsilon^2} + O(\varepsilon). \quad (18)$$

To obtain the reverse inequality for the lower limit, observe that $A_n = \emptyset$, whence via (14)

$$\begin{aligned}
&P\{T_{x\sqrt{n}} \leq n\} = P\{S_n > (x + 2\varepsilon)\sqrt{n}\} \\
&= P\{S_n \leq (x + 2\varepsilon)\sqrt{n}, \max S_j > x\sqrt{n}\} \\
&\geq \sum_{i=1}^n P\{T_{x\sqrt{n}} = i, S_n < S_{n_{m(i)}}, |X_i| \leq \varepsilon\sqrt{n}, A_i^c\} \\
&\geq \sum_{i=1}^n P\{T_{x\sqrt{n}} = i, S_n < S_{n_{m(i)}}, |X_i| \leq \varepsilon\sqrt{n}\} - \sum_{i=1}^n P\{T_{x\sqrt{n}} = i, A_i\}
\end{aligned}$$

$$\leq E(1 + X''_{n,x})e^{U''_{n,x}^2} \leq e^c[1 + E(X''_n)^2]$$

implying via (4)" that $\{T_n, n \geq 1\}$ is u.i.

Define $r(x)$ by $r(0) = 0$ and (see Lemma 12.1.1)

$$e^{ix} = (1 + ix)e^{-(x^2/2) + r(x)} \quad (6)$$

whence $1 = \exp\{r(x) + \sum_{j=3}^{\infty} (-ix)^j/j\}$ for $|x| < 1$ implying $r(x) = x^4a(x) + ix^3b(x)$, where

$$0 < a(x) = 1/4 - x^2/6 + x^4/8 \dots < 1/4$$

$$0 < b(x) = 1/3 - x^2/5 + x^4/7 \dots < 1/3$$

so that $|r(x)| < |x|^3$ for $|x| < 1$. Now (6) entails

$$e^{iS'_n} = T_n W'_n = (T_n - 1)W_n + (T_n - 1)(W'_n - W_n) + W'_n \quad (7)$$

where

$$W'_n = \exp\{-\frac{1}{2}U''_n^2 + \sum_j r(X''_{n,j})\}, \quad W_n = \exp\{-\frac{1}{2}U''_n^2\}. \quad (8)$$

Since on $\{X''_n < 1\}$, $|\sum_{j=1}^{k_n} r(X''_{n,j})| < \sum_j |X''_{n,j}|^3 \leq X''_n U''_n^2 \xrightarrow{P} 0$, (3)" guarantees

$$W'_n \xrightarrow{d} e^{-U''_n^2/2} \quad (9)$$

while (1)" and (8) ensure $W'_n/W_n \xrightarrow{P} 1$. Thus, $0 \leq W_n \leq 1$ entails $W'_n - W_n \xrightarrow{P} 0$ and so recalling that $\{T_n\}$ is u.i. (and a fortiori, tight)

$$(T_n - 1)(W'_n - W_n) \xrightarrow{P} 0. \quad (10)$$

Next, since $\{X''_{n,j}, \mathcal{F}_{n,j}, 1 \leq j \leq k_n\}$ are martingale differences,

$$\begin{aligned} E_1 T_n &= E_1 \prod_{j=1}^{k_n-1} (1 + iX''_{n,j}) E_{k_n-1} (1 + iX''_{n,k_n}) \\ &= E \prod_{j=1}^{k_n-1} (1 + iX''_{n,j}) = \dots = 1 + iX''_{n,1} \end{aligned}$$

and so, recalling that $0 \leq W_n \leq 1$

$$|E(T_n - 1)W_n| = |E W_n E_1(T_n - 1)| \leq E|E_1(T_n - 1)| \leq E|X''_{n,1}| \rightarrow 0 \quad (11)$$

via (2)", (4)" and dominated convergence.

Now in view of (7)

$$e^{iS'_n} - W_n(T_n - 1) = (T_n - 1)(W'_n - W_n) + W'_n \quad (12)$$

while (9) and (10) ensure that the right and hence left side of (12) $\xrightarrow{d} e^{-U''_n^2/2}$. Since the left side is u.i., Corollary 8.1.8 in conjunction with (11) guarantees that

$$E e^{iS'_n} \rightarrow E e^{-U''_n^2/2}. \quad (13)$$

Consequently,

$$\begin{aligned} |\mathbb{E}(e^{iS_n'} - e^{-\eta^2/2})| &\leq |\mathbb{E}(e^{iS_n'} - e^{iS_n''})| + |\mathbb{E} e^{iS_n''} - e^{\eta^2/2}| + |\mathbb{E}(e^{-\eta^2/2} - e^{-\eta^2/2})| \\ &\leq 2P\{U_n'^2 > c\} + |\mathbb{E}(e^{iS_n''} - e^{-\eta^2/2})| + P\{\eta^2 > c\} \end{aligned}$$

which, in conjunction with (13) implies that for any $\varepsilon > 0$

$$\overline{\lim} |\mathbb{E}(e^{iS_n'} - e^{-\eta^2/2})| \leq 3\varepsilon$$

provided $c \geq C_\varepsilon$. Replacing S_n' by tS_n for $t \neq 0$,

$$\lim_{n \rightarrow \infty} \mathbb{E} e^{itS_n} = \mathbb{E} e^{-t^2\eta^2/2}.$$

Thus $S_n' \xrightarrow{d} S_G$ where the c.f. of S_G is $\mathbb{E} e^{-t^2\eta^2/2}$. Finally, $S_n \xrightarrow{d} S_G$ via (5).

Corollary 1 (McLeish). If for each $n \geq 1$, $\{S_{n,j} = \sum_{i=1}^j X_{n,i}, \mathcal{F}_{n,j}, 1 \leq j \leq k_n < \infty\}$ is an \mathcal{L}_2 stochastic sequence on (Ω, \mathcal{F}, P) satisfying (2), (4), (5) and

$$U_n^2 \xrightarrow{P} \eta^2 \quad (1')$$

for some non-negative $\bigcap_{n=1}^{\infty} \mathcal{F}_{n,1}$ -measurable r.v. η^2 then the conclusion of Theorem 1 holds.

In order to replace conditions (2), (4) by a conditional Lindeberg condition (22) and substitute $V_n^2 = \sum_{j=1}^n \mathbb{E}\{X_{nj}^2 | \mathcal{F}_{nj-1}\}$ for U_n^2 , several lemmas will be needed.

Lemma 1 (Dvoretzky). If the events $A_j \in \mathcal{F}_j$, $1 \leq j \leq n$ where $\{\mathcal{F}_j, j \geq 0\}$ is an increasing sequence of sub- σ -algebras of the basic σ -algebra \mathcal{F} , then

$$P\left\{\bigcup_{j=1}^n A_j\right\} \leq \varepsilon + P\left\{\sum_{j=1}^n P\{A_j | \mathcal{F}_{j-1}\} > \varepsilon\right\}, \quad \varepsilon > 0. \quad (14)$$

In particular, for any non-negative stochastic sequence $\{Y_{n,j}, \mathcal{F}_{n,j}, 1 \leq j \leq k_n\}$, if $\sum_{j=1}^k \mathbb{E}\{Y_{n,j} I_{\{Y_{n,j} \geq \varepsilon\}} | \mathcal{F}_{n,j-1}\} \xrightarrow{P} 0$, all $\varepsilon > 0$, then

$$\max_{1 \leq j \leq k_n} Y_{n,j} \xrightarrow{P} 0. \quad (15)$$

PROOF. Setting $\mu_k = \sum_{j=1}^k P\{A_j | \mathcal{F}_{j-1}\}$, $1 \leq k \leq n$

$$P\left\{\bigcup_{j=1}^n A_j | \mu_n \leq \varepsilon\right\} \leq P\left\{\bigcup_{j=1}^n A_j | \mu_j \leq \varepsilon\right\} = E \sum_{j=1}^n P\{A_j | \mathcal{F}_{j-1}\} I_{\{\mu_j \leq \varepsilon\}} \leq \varepsilon$$

so that

$$P\left\{\bigcup_{j=1}^n A_j\right\} \leq P\left\{\bigcup_{j=1}^n A_j | \mu_n \leq \varepsilon\right\} + P\{\mu_n > \varepsilon\} \leq \varepsilon + P\{\mu_n > \varepsilon\}. \quad \square$$

Lemma 2 (Dvoretzky). Let W be any \mathcal{G} -measurable, \mathcal{L}_2 r.v. where \mathcal{G} is a sub- σ -algebra of \mathcal{F} . Then for any $\varepsilon > 0$ and any \mathcal{L}_2 r.v. Y for which $E\{Y | \mathcal{G}\} \stackrel{a.s.}{=} 0$, with probability one

$$E\{Y^2 I_{\{Y^2 > \varepsilon\}} | \mathcal{G}\} \leq E\{(2Y + W)^2 I_{\{(2Y + W)^2 > \varepsilon\}} | \mathcal{G}\}. \quad (16)$$

PROOF. Without loss of generality suppose $\varepsilon = 1$. Define

$$A = \{W^2 \geq 1\}, \quad Z = (2Y + W)^2 - Y^2$$

$$Q(D) = E(Y^2 + Z)I_{\{Y^2 + Z > 1\}D} - E(Y^2)I_{\{Y^2 > 1\}D}, \quad D \in \mathcal{F}.$$

It suffices to prove for any $G \in \mathcal{G}$ that $Q(G) \geq 0$. Since $E(WYI_{AG}) = E(WI_{AG})E(Y|\mathcal{G}) = 0$ implies

$$E(ZI_{AG}) = 3E(Y^2I_{AG}) + E(W^2I_{AG})$$

and moreover

$$E(ZI_{\{Y^2 + Z \leq 1\}AG}) \leq E(ZI_{\{0 \leq Z \leq 1\}AG}) \leq P\{AG\} \leq E(W^2I_{AG})$$

necessarily

$$Q(AG) = E(ZI_{AG}) - E(ZI_{\{Y^2 + Z \leq 1\}AG}) + E(Y^2I_{\{Y^2 + Z > 1\}AG}) - E(Y^2)I_{\{Y^2 > 1\}AG} \geq 0.$$

On A^c , $Y^2 + Z \leq 1$ iff $(2Y + W + 1)(2Y + W - 1) \leq 0$ iff $-(1 + W)/2 \leq Y \leq (1 - W)/2$. Hence $A^c[Y^2 + Z \leq 1] \subset [Y^2 \leq 1]$ so that $A^c[Y^2 + Z \leq 1] \cdot [Y^2 > 1] = \emptyset$. Furthermore, since $Z < 0$ iff $|2Y + W| < |Y|$ which, in turn, implies $|Y| < |W|$, necessarily $[Z < 0]A^c \subset [Y^2 < 1]$ so that $Z \geq 0$ on $A^c[Y^2 \geq 1]$. Thus

$$\begin{aligned} Q(A^cG) &= E(Y^2 + Z)I_{\{Y^2 + Z > 1 \geq Y^2\}A^cG} - E(Y^2)I_{\{Y^2 + Z \leq 1 < Y^2\}A^cG} \\ &\quad + E(ZI_{\{Y^2 + Z > 1, Y^2 > 1\}A^cG}) \geq 0 \end{aligned}$$

whence

$$Q(G) = Q(AG) + Q(A^cG) \geq 0. \quad \square$$

Corollary 2. If $E|X|^2 < \infty$ and $\varepsilon > 0$, with probability one

$$E\{|X - E\{X|\mathcal{G}\}|^2 I_{\{|X - E\{X|\mathcal{G}\}| > 2\varepsilon\}}|\mathcal{G}\} \leq 4E\{X^2 I_{\{|X| > \varepsilon\}}|\mathcal{G}\}. \quad (17)$$

PROOF. Take $W = 2E\{X|\mathcal{G}\}$ and $Y = X - E\{X|\mathcal{G}\}$ \square

Lemma 3 (McLeish). For each $n \geq 1$ let $\{Y_{n,j}, \mathcal{F}_{n,j}, 1 \leq j \leq k_n < \infty\}$ be a non-negative \mathcal{L}_1 stochastic sequence and let

$$S_{n,j} = \sum_{i=1}^j Y_{n,i}, \quad \mu_{nj} = \sum_{i=1}^j E\{Y_{nj}|\mathcal{F}_{n,j-1}\}, \quad 1 \leq j \leq k_n,$$

$$S_n = S_{n,k_n}, \quad \mu_n = \mu_{n,k_n}.$$

If

$$\sum_{i=1}^{k_n} E\{Y_{n,j}I_{\{Y_{n,j} > \varepsilon\}}|\mathcal{F}_{n,j-1}\} \xrightarrow{P} 0, \quad \varepsilon > 0 \quad (18)$$

and

$$\{\mu_n, n \geq 1\} \text{ is tight} \quad (19)$$

then

$$\max_{1 \leq j \leq k_n} |S_{n,j} - \mu_{n,j}| \xrightarrow{P} 0. \quad (20)$$

PROOF. Via Lemma 1,

$$Y_n^* = \max_{j \leq k_n} Y_{n,j} \xrightarrow{P} 0.$$

Let $Y'_{n,j} = Y_{n,j} I_{\{Y_{n,j} \leq K^{-2}, \mu_{n,j} \leq K\}}$, $K > 0$.

Now

$$\left\{ \sum_{j=1}^{k_n} |Y_{n,j} - Y'_{n,j}| > 0 \right\} \subset \{\mu_n > K\} \cup \{Y_n^* > K^{-2}\}$$

implying

$$\overline{\lim}_{n \rightarrow \infty} P \left\{ \sum_{j=1}^{k_n} |Y_{n,j} - Y'_{n,j}| > 0 \right\} \leq \sup_{n \geq 1} P\{\mu_n > K\} + \overline{\lim}_{n \rightarrow \infty} P\{Y_n^* > K^{-2}\} \xrightarrow{K \rightarrow \infty} 0$$

whence defining $S'_{n,j}$, S'_n , $\mu'_{n,j}$, μ'_n analogously

$$\max_{1 \leq j \leq k_n} |S_{n,j} - S'_{n,j}| \leq \sum_{j=1}^{k_n} |Y_{n,j} - Y'_{n,j}| \xrightarrow{P} 0.$$

Moreover,

$$\begin{aligned} \max_{j \leq k_n} |\mu_{n,j} - \mu'_{n,j}| &\leq \sum_{j=1}^{k_n} E\{|Y_{n,j} - Y'_{n,j}| \mid \mathcal{F}_{n,j-1}\} \\ &\leq \sum_{j=1}^{k_n} (E\{Y_{n,j} I_{\{Y_{n,j} > K^{-2}\}} \mid \mathcal{F}_{n,j-1}\} + E\{Y_{n,j} \mid \mathcal{F}_{n,j-1}\} I_{\{\mu_{n,j} > K\}}) \\ &\leq \sum_{j=1}^{k_n} E\{Y_{n,j} I_{\{Y_{n,j} > K^{-2}\}} \mid \mathcal{F}_{n,j-1}\} + \mu_n I_{\{\mu_n > K\}} \end{aligned}$$

whence, in view of (18) and (19), for any $\delta > 0$

$$\begin{aligned} P \left\{ \max_{1 \leq j \leq k_n} |\mu_{n,j} - \mu'_{n,j}| > \delta \right\} &\leq P \left\{ \sum_{j=1}^{k_n} E\{Y_{n,j} I_{\{Y_{n,j} > K^{-2}\}} \mid \mathcal{F}_{n,j-1}\} > \delta/2 \right\} \\ &\quad + P\{\mu_n I_{\{\mu_n > K\}} > \delta/2\} \leq P \left\{ \sum_{j=1}^{k_n} E\{Y_{n,j} I_{\{Y_{n,j} > K^{-2}\}} \mid \mathcal{F}_{n,j-1}\} > \delta/2 \right\} \\ &\quad + \sup_{n \geq 1} P\{\mu_n > K\} \xrightarrow{n \rightarrow \infty} \sup_{n \geq 1} P\{\mu_n > K\} \xrightarrow{K \rightarrow \infty} 0. \end{aligned}$$

Furthermore, by Theorem 7.4.8 for any $\delta > 0$

$$\begin{aligned} (\delta^2/4) P \left\{ \max_j |S'_{n,j} - \mu'_{n,j}| > \delta \right\} &\leq \frac{1}{4} E \max_j (S'_{n,j} - \mu'_{n,j})^2 \leq E(S'_n - \mu'_n)^2 \\ &\leq \sum_{j=1}^{k_n} E Y'_{n,j}^2 \leq K^{-2} E \sum_{j=1}^{k_n} Y'_{n,j} = K^{-2} E \sum_{j=1}^{k_n} E\{Y'_{n,j} \mid \mathcal{F}_{n,j-1}\} \\ &\leq K^{-2} E \sum_{j=1}^{k_n} E\{Y_{n,j} \mid \mathcal{F}_{n,j-1}\} I_{\{\mu_{n,j} \leq K\}} \leq K^{-1} \xrightarrow{K \rightarrow \infty} 0. \end{aligned}$$

Finally,

$$\begin{aligned} \max_j |S_{n,j} - \mu_{n,j}| \\ \leq \max_j |S_{n,j} - S'_{n,j}| + \max_j |S'_{n,j} - \mu'_{n,j}| + \max_j |\mu'_{n,j} - \mu_{n,j}| \xrightarrow{P} 0. \quad \square \end{aligned}$$

As earlier, $E_{j-1}\{\cdot\}$ abbreviates $E\{\cdot|\mathcal{F}_{n,j-1}\}$. Condition (22) is a conditional Lindeberg condition and hence less stringent than the classical version.

Theorem 2. For each $n \geq 1$, let $\{S_{n,j} = \sum_{i=1}^j X_{ni}, \mathcal{F}_{n,j}, 1 \leq j \leq k_n < \infty\}$ be an \mathcal{L}_2 stochastic sequence such that for some $\mathcal{F}_{n,1}$ -measurable r.v. u_n^2

$$\sum_{j=1}^{k_n} [E_{j-1}\{X_{n,j}^2\} - E_{j-1}^2\{X_{n,j}\}] - u_n^2 \xrightarrow{P} 0 \quad (21)$$

$$\sum_{j=1}^{k_n} E_{j-1}\{X_{nj}^2 I_{\{|X_{n,j}|>\varepsilon\}}\} \xrightarrow{P} 0, \quad \varepsilon > 0 \quad (22)$$

$$\sum_{j=1}^{k_n} [E_{j-1}\{X_{n,j}^2\} - E_{j-1}^2\{X_{n,j}\}] \xrightarrow{d} \eta_F^2 \quad (23)$$

$$\sum_{j=1}^{k_n} E_{j-1}\{X_{n,j}\} \xrightarrow{P} 0, \quad \sum_{j=1}^{k_n} E_{j-1}^2\{X_{n,j}\} \xrightarrow{P} 0. \quad (24)$$

where $\mathcal{F}_{n,0}$ is the trivial σ -algebra. Then $S_n \equiv S_{n,k_n} \xrightarrow{d} S_G$ where $E e^{itS_G} = E e^{-\eta_F^2 t^2/2}$.

PROOF. For $n \geq 1$, define $X'_{n,j} = X_{n,j} - E_{j-1}\{X_{n,j}\}$. By Corollary 2, $\{X'_{n,j}\}$ satisfies (22) and hence also condition (2) of Theorem 1, via Lemma 1. Set $V_n^2 = \sum_{j=1}^{k_n} E_{j-1}\{X_{n,j}^2\}$, $U_n^2 = \sum_{j=1}^{k_n} X_{nj}^2$ and define $V'_n{}^2$ and $U'_n{}^2$ analogously. Via (23)

$$V'_n{}^2 \xrightarrow{d} \eta_F^2$$

The latter ensures that $V'_n{}^2$ is tight which, in conjunction with (22) for $\{X'_{n,j}\}$ guarantees that

$$U'_n{}^2 - V'_n{}^2 \xrightarrow{P} 0$$

by Lemma 3. Consequently, (1) and (3) of Theorem 1 hold for $\{X'_{n,j}\}$.

Next, define the martingale difference sequence

$$X''_{n,j} = X'_{n,j} I_{\{\sum_{i=1}^j E_{i-1}\{X'_{ni} I_{\{|X'_{n,i}|>\varepsilon\}}\} \leq 1\}}$$

and note that (2) and (5) hold trivially for $\{X''_{n,j}\}$. Likewise (1) and (3) obtain since $U'_n{}^2 - U''_n{}^2 \xrightarrow{P} 0$ in view of

$$U'_n{}^2 - U''_n{}^2 \leq U'_n{}^2 I_{\{\sum_{n=1}^{k_n} E_{j-1}\{X'_{nj} I_{\{|X'_{n,j}|>\varepsilon\}}\} > 1\}}$$

and the conditional Lindeberg condition.

Moreover,

$$\begin{aligned} E \max_{j \leq k_n} X''_{n,j}^2 &\leq \varepsilon^2 + E \sum_{j=1}^{k_n} X'_{nj}^2 I_{\{|X'_{nj}|>\varepsilon, \sum_{i=1}^j E_{i-1}\{X'_{ni} I_{\{|X'_{n,i}|>\varepsilon\}}\} \leq 1\}} \\ &\leq \varepsilon^2 + E \sum_{j=1}^{k_n} E_{j-1}\{X'_{nj}^2 I_{\{|X'_{nj}|>\varepsilon\}}\} I_{\{\sum_{i=1}^j E_{i-1}\{X'_{ni} I_{\{|X'_{n,i}|>\varepsilon\}}\} \leq 1\}} \\ &\leq \varepsilon^2 + 1 \end{aligned}$$

References

- F. Anscombe, "Large sample theory of sequential estimation," *Proc. Cambr. Philos. Soc.* **48** (1952), 600–607.
- S. Bernstein, "Several comments concerning the limit theorem of Liapounov," *Dokl. Akad. Nauk. SSSR* **24** (1939), 3–7.
- A. C. Berry, "The accuracy of the Gaussian approximation to the sum of independent variates," *Trans. Amer. Math. Soc.* **49** (1941), 122–136.
- J. Blum, D. Hanson, and J. Rosenblatt, "On the CLT for the sum of a random number of random variables," *Z. Wahr. Verw. Geb.* **1** (1962–1963), 389–393.
- J. Blum, H. Chernoff, M. Rosenblatt, and H. Teicher, "Central limit theorems for interchangeable processes," *Can. Jour. Math.* **10** (1958), 222–229.
- K. L. Chung, *A Course in Probability Theory*, Harcourt Brace, New York, 1968; 2nd ed., Academic Press, New York, 1974.
- W. Doeblin, "Sur deux problèmes de M. Kolmogorov concernant les chaînes dénombrables," *Bull. Soc. Math. France* **66** (1938), 210–220.
- J. L. Doob, *Stochastic Processes*, Wiley New York, 1953.
- A. Dvoretzky, "Asymptotic normality for sums of dependent random variables," *Proc. Sixth Berkeley Symp. on Stat. and Prob.* 1970, 513–535.
- P. Erdos and M. Kac, "On certain limit theorems of the theory of probability," *Bull. Amer. Math. Soc.* **52** (1946), 292–302.
- C. Esseen, "Fourier analysis of distribution functions," *Acta Math.* **77** (1945), 1–125.
- W. Feller, "Über den Zentralen Grenzwertsatz der wahrscheinlichkeitsrechnung," *Math. Zeit.* **40** (1935), 521–559.
- N. Friedman, M. Katz, and L. Koopmans, "Convergence rates for the central limit theorem," *Proc. Nat. Acad. Sci.* **56** (1966), 1062–1065.
- P. Hall and C. C. Heyde, *Martingale Limit Theory and its Application*, Academic Press, New York, 1980.
- K. Knopp, *Theory and Application of Infinite Series*, Stechert–Hafner, New York, 1928.
- P. Lévy, *Théorie de l'addition des variables aléatoires*, Gauthier–Villars, Paris, 1937; 2nd ed., 1954.
- J. Lindeberg, "Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung," *Math. Zeit.* **15** (1922), 211–225.
- D. L. McLeish, "Dependent Central Limit Theorems and invariance principles," *Ann. Prob.* **2** (1974), 620–628.
- J. Mogyorodi, "A CLT for the sum of a random number of independent random variables," *Magyar. Tud. Akad. Mat. Kutató Int. Közl.* **7** (1962), 409–424.
- A. Renyi, "Three new proofs and a generalization of a theorem of Irving Weiss," *Magyar. Tud. Akad. Mat. Kutató Int. Közl.* **7** (1962), 203–214.
- A. Renyi, "On the CLT for the sum of a random number of independent random variables," *Acta Math. Acad. Sci. Hung.* **11** (1960), 97–102.
- B. Rosen, "On the asymptotic distribution of sums of independent, identically distributed random variables," *Arkiv för Mat.* **4** (1962), 323–332.
- D. Siegmund, "On the asymptotic normality of one-sided stopping rules," *Ann. Math. Stat.* **39** (1968), 1493–1497.
- H. Teicher, "On interchangeable random variables," *Studi di Probabilità Statistica e Ricerca Operativa in Onore di Giuseppe Pompilj*, pp. 141–148, Gubbio, 1971.
- H. Teicher, "A classical limit theorem without invariance or reflection," *Ann. Math. Stat.* **43** (1973), 702–704.
- P. Van Beek, "An application of the Fourier method to the problem of sharpening the Berry–Esseen inequality," *Z. Wahr.* **23** (1972), 187–197.
- I. Weiss, "Limit distributions in some occupancy problems," *Ann. Math. Stat.* **29** (1958), 878–884.
- V. Zolotarev, "An absolute estimate of the remainder term in the C.L.T.," *Theor. Prob. and its Appl.* **11** (1966), 95–105.

10 Limit Theorems for Independent Random Variables

10.1 Laws of Large Numbers

Prior discussion of the strong and weak laws of large numbers centered around the i.i.d. case. Necessary and sufficient conditions for the weak law are available when the underlying random variables are merely independent and have recently been obtained for the strong law as well. Unfortunately, the practicality of the latter conditions leaves much to be desired.

A few words are in order on a method of considerable utility in probability theory, namely, that of symmetrization. In Chapter 6 it was pointed out that, given a sequence of r.v.s $\{X_n, n \geq 1\}$ on (Ω, \mathcal{F}, P) , a symmetrized sequence of r.v.s $\{X_n^*, n \geq 1\}$ can be defined—if necessary by constructing a new probability space. The joint distributions of

$$X_n^* = X_n - \bar{X}_n, \quad n \geq 1,$$

are determined by the fact that $\{X_n^*, n \geq 1\}$ is independent of $\{X_n, n \geq 1\}$ and possesses the same joint d.f.s, that is, $\{X_n, n \geq 1\}$ and $\{X_n^*, n \geq 1\}$ are i.i.d. stochastic processes. In particular, if the initial r.v.s $X_n, n \geq 1$, are independent, so are the symmetrized $X_n^*, n \geq 1$, namely, the $X_n^*, n \geq 1$.

The salient point concerning symmetrization is that the X_n^* are symmetric about zero while the magnitude of the sum of the two tails of the corresponding d.f. F_n^* is roughly the same as that of F_n . The relation between the distributions and moments is stated explicitly in

Lemma 1. If $\{X_j, 1 \leq j \leq n\}$ and $\{X'_j, 1 \leq j \leq n\}$ are i.i.d. stochastic processes with medians m_j and $X_j^* = X_j - m_j$, then for any $n \geq 1, \varepsilon > 0$, and real a

$$\frac{1}{2} P \left\{ \max_{1 \leq j \leq n} (X_j - m_j) \geq \varepsilon \right\} \leq P \left\{ \max_{1 \leq j \leq n} X_j^* \geq \varepsilon \right\}, \quad (1)$$

$$\begin{aligned} \frac{1}{2} P \left\{ \max_{1 \leq j \leq n} |X_j - m_j| \geq \varepsilon \right\} &\leq P \left\{ \max_{1 \leq j \leq n} |X_j^*| \geq \varepsilon \right\} \\ &\leq 2 P \left\{ \max_{1 \leq j \leq n} |X_j - a| \geq \frac{\varepsilon}{2} \right\}, \end{aligned} \quad (2)$$

$$\frac{1}{2} E \left(\max_{1 \leq j \leq n} |X_j - m_j| \right)^p \leq E \left(\max_{1 \leq j \leq n} |X_j^*| \right)^p \leq 2 E \left(2 \max_{1 \leq j \leq n} |X_j - a| \right)^p, \quad (2')$$

where $p > 0$. Moreover, if $E X_1 = 0$,

$$E|X_1|^p \leq E|X_1^*|^p \leq 2^p E|X_1|^p, \quad p \geq 1. \quad (3)$$

PROOF. Set $A_j = \{X_j - m_j \geq \varepsilon\}$, $B_j = \{X'_j - m_j \leq 0\}$, $C_j = \{X_j^* \geq \varepsilon\}$. Then $A_j \cdot B_j \subset C_j$ and by Lemma 3.3.3

$$\frac{1}{2} P \left\{ \bigcup_{j=1}^n A_j \right\} \leq P \left\{ \bigcup_{j=1}^n A_j B_j \right\} \leq P \left\{ \bigcup_{j=1}^n C_j \right\},$$

which is (1). To prove (2), set

$$T = \inf \{j \geq 1 : |X_j - m_j| \geq \varepsilon\}.$$

Then via independence,

$$\begin{aligned} P\{T \leq n\} &\leq 2 \sum_{j=1}^n [P\{T = j, X_j - m_j \geq \varepsilon, X'_j \leq m_j\} \\ &\quad + P\{T = j, X_j - m_j \leq -\varepsilon, X'_j \geq m_j\}] \\ &\leq 2 \sum_{j=1}^n P\{T = j, |X_j^*| \geq \varepsilon\} = 2 \sum_{j=1}^n P\{T = j, |X_j^*| \geq \varepsilon\} \\ &\leq 2 P \left\{ \max_{1 \leq j \leq n} |X_j^*| \geq \varepsilon \right\}. \end{aligned}$$

This yields the first inequality of (2) and the second follows from

$$\begin{aligned} P \left\{ \max_{1 \leq j \leq n} |X_j^*| \geq \varepsilon \right\} &\leq P \left\{ \max_{1 \leq j \leq n} |X_j - a| \geq \frac{\varepsilon}{2} \text{ or } \max_{1 \leq j \leq n} |X'_j - a| \geq \frac{\varepsilon}{2} \right\} \\ &\leq 2 P \left\{ 2 \max_{1 \leq j \leq n} |X_j - a| \geq \varepsilon \right\}. \end{aligned}$$

Moreover, (2) in turn yields (2') via Corollary 6.2.2.

Apropos of (3), since $E\{X_1 - X'_1 | X_1\} = X_1$, the conditional Jensen inequality ensures $E|X_1^*|^p \geq E|X_1|^p$. The remaining portion follows trivially from Exercise 4.2.1 and does not require that $E X_1 = 0$. \square

Theorem 1 (Weak law). *For each $n \geq 1$, let $\{X_{nj}, 1 \leq j \leq k_n \rightarrow \infty\}$ be independent r.v.s, $S_n = \sum_{j=1}^{k_n} X_{nj}$, and let m_{nj} denote a median of X_{nj} . Then for some real numbers A_n*

$$S_n - A_n \xrightarrow{P} 0, \quad (4)$$

$$\max_{1 \leq j \leq k_n} |m_{nj}| \rightarrow 0, \quad (5)$$

as $n \rightarrow \infty$ iff

$$\sum_{j=1}^{k_n} P\{|X_{nj}| \geq \varepsilon\} \rightarrow 0, \quad \varepsilon > 0, \quad (6)$$

$$\sum_{j=1}^{k_n} \sigma^2(X_{nj} I_{\{|X_{nj}| < 1\}}) \rightarrow 0, \quad (7)$$

in which case

$$A_n - \sum_{j=1}^{k_n} E X_{nj} I_{\{|X_{nj}| < 1\}} \rightarrow 0. \quad (8)$$

PROOF. (6) implies (5) trivially. To prove (4), set

$$Y_{nj} = X_{nj} I_{\{|X_{nj}| < 1\}}, \quad V_n = \sum_{j=1}^{k_n} Y_{nj},$$

and note that via independence (7) ensures $V_n - E V_n \xrightarrow{P} 0$. Since by (6)

$$P\{V_n \neq S_n\} \leq \sum_{j=1}^{k_n} P\{|X_{nj}| \geq 1\} = o(1), \quad (9)$$

$S_n - E V_n \xrightarrow{P} 0$, yielding (4) and (8).

Conversely, if (4) and (5) hold, let $(X'_{n1}, \dots, X'_{nk_n})$ and $(X_{n1}, \dots, X_{nk_n})$ be i.i.d. random vectors for each $n \geq 1$ and set $X_{nj}^* = X_{nj} - X'_{nj}$, $S_n^* = \sum_{j=1}^{k_n} X_{nj}^*$. Then (4) entails $S_n^* \xrightarrow{P} 0$ and so by Levy's inequality (Corollary 3.3.5) for any $\varepsilon > 0$

$$P\left\{\max_{j \leq k_n} |X_{nj}^*| \geq \varepsilon\right\} \leq P\left\{\max_{k \leq k_n} \left|\sum_1^k X_{nj}^*\right| \geq \frac{\varepsilon}{2}\right\} \leq 2 P\left\{|S_n^*| \geq \frac{\varepsilon}{2}\right\} = o(1),$$

whence

$$\exp\left\{-\sum_1^{k_n} P\{|X_{nj}^*| \geq \varepsilon\}\right\} \geq \prod_{j=1}^{k_n} P\{|X_{nj}^*| < \varepsilon\} = P\left\{\max_{j \leq k_n} |X_{nj}^*| < \varepsilon\right\} \rightarrow 1$$

as $n \rightarrow \infty$, implying for all $\varepsilon > 0$ that

$$\sum_{j=1}^{k_n} P\{|X_{nj}^*| \geq \varepsilon\} = o(1). \quad (10)$$

$$\sum_{j=1}^{k_n} E(X_{nj} - m_{nj})^2 I_{\{|X_{nj} - m_{nj}| < 1\}} = o(1), \quad (14)$$

in which case A_n may be chosen to be $\sum_{j=1}^{k_n} \{m_{nj} + E(X_{nj} - m_{nj})I_{\{|X_{nj} - m_{nj}| < 1\}}\}$.

PROOF. Since $Z_{nj} = X_{nj} - m_{nj}$ has zero as a median, (12) follows from (13) and (14) by Theorem 1, noting (9). Conversely, under (12), setting $U_n = \sum_{j=1}^{k_n} Z_{nj}$, $B_n = A_n - \sum_{j=1}^{k_n} m_{nj}$,

$$U_n - B_n \xrightarrow{P} 0, \quad (15)$$

whence (13) holds by Theorem 1. To prove (14), set

$$Y_{nj} = Z_{nj} I_{\{|Z_{nj}| < 1\}}$$

and let $(Y'_{n1}, \dots, Y'_{nk_n})$ and $(Y_{n1}, \dots, Y_{nk_n})$ be i.i.d. random vectors. It follows from (15) and (13) that

$$\sum_{j=1}^{k_n} \frac{(Y_{nj} - Y'_{nj})}{3} \xrightarrow{P} 0,$$

and since $|(Y_{nj} - Y'_{nj})/3| < 1$ and (Exercise 3.3.3) $m(Y_{nj}) = 0$, Theorem 1 and Lemma 1 ensure that

$$o(1) = \sum_{j=1}^{k_n} \sigma^2 \left(\frac{Y_{nj} - Y'_{nj}}{3} \right) = \frac{1}{9} \sum_{j=1}^{k_n} E(Y_{nj} - Y'_{nj})^2 \geq \frac{1}{18} \sum_{j=1}^{k_n} E Y_{nj}^2,$$

which is tantamount to (14). \square

Corollary 2. If for each $n \geq 1$, $\{Y_{nj}, 1 \leq j \leq k_n \rightarrow \infty\}$ are independent, positive r.v.s, then $S_n = \sum_{j=1}^{k_n} Y_{nj} \xrightarrow{P} 1$ and $\max_{1 \leq j \leq k_n} m(Y_{nj}) \rightarrow 0$ iff for all $\varepsilon > 0$

$$\sum_{j=1}^{k_n} P\{Y_{nj} \geq \varepsilon\} = o(1), \quad (16)$$

$$\sum_{j=1}^{k_n} E Y_{nj} I_{\{Y_{nj} < 1\}} \rightarrow 1. \quad (17)$$

Moreover, if $\sum_{j=1}^{k_n} E Y_{nj} = 1$, then $S_n \xrightarrow{P} 1$ and $\max_{1 \leq j \leq k_n} m(Y_{nj}) \rightarrow 0$ iff

$$\sum_{j=1}^{k_n} E Y_{nj} I_{\{Y_{nj} \geq \varepsilon\}} = o(1), \quad \varepsilon > 0. \quad (18)$$

PROOF. Necessity of (16) and (17) is an immediate consequence of (6) and (8) of Theorem 1 with $A_n \equiv 1$. Sufficiency likewise follows from Theorem 1 once

where $1 < \alpha \leq 2$ and $A_{n+1,n}/A_{n,n} \leq \gamma < \infty$, all $n \geq 1$. If, for some β in $[0, 1/\alpha)$ and positive δ, c ,

$$\sum_{n=1}^{\infty} P\{|X_n| > \delta A_n (\log_2 A_n)^{1-\beta}\} < \infty \quad (36)$$

$$\sum_{n=1}^{\infty} \frac{1}{A_n^2 (\log_2 A_n)^{2(1-\beta)}} E X_n^2 I_{[c A_n (\log_2 A_n)^{1-\beta} < |X_n| \leq \delta A_n (\log_2 A_n)^{1-\beta}]} < \infty \quad (37)$$

then

$$\frac{S_n}{A_n (\log_2 A_n)^{1-\beta}} \xrightarrow{a.s.} 0. \quad (38)$$

PROOF. Set $Y_n = X_n I_{[|X_n| \leq c A_n (\log_2 A_n)^{1-\beta}]}$, $W_n = X_n I_{[|X_n| > \delta A_n (\log_2 A_n)^{1-\beta}]}$, $V_n = X_n - Y_n - W_n$. Now since $\beta < 1/\alpha$,

$$\begin{aligned} \left| \sum_{i=1}^n E W_i \right| &\leq \sum_{i=1}^n (E |X_i| I_{[\delta A_i (\log_2 A_i)^{1-\beta} < |X_i| \leq A_i (\log_2 A_i)^{1-\beta}} \\ &\quad + E |X_i| I_{[|X_i| > A_i (\log_2 A_i)^{1-\beta}]}) \\ &\leq A_n (\log_2 A_n)^{-\beta} \sum_{i=1}^n P\{|X_i| > \delta A_i (\log_2 A_i)^{1-\beta}\} \\ &\quad + \frac{(\log_2 A_n)^{\beta(\alpha-1)}}{A_n^{\alpha-1}} \sum_{i=1}^n E |X_i|^{\alpha} I_{[|X_i| > A_n (\log_2 A_n)^{-\beta}]} \\ &= o(A_n (\log_2 A_n)^{1-\beta}) \end{aligned}$$

and so, in view of (36),

$$\frac{1}{A_n (\log_2 A_n)^{1-\beta}} \sum_{i=1}^n (W_i - E W_i) \xrightarrow{a.s.} 0.$$

Secondly, (37) and Kronecker's lemma guarantee

$$\frac{1}{A_n (\log_2 A_n)^{1-\beta}} \sum_{i=1}^n (V_i - E V_i) \xrightarrow{a.s.} 0.$$

Thus, since $E X_n = 0$, it suffices to verify that

$$\frac{1}{A_n (\log_2 A_n)^{1-\beta}} \sum_{i=1}^n (Y_i - E Y_i) \xrightarrow{a.s.} 0. \quad (39)$$

To this end, note that if $n_k = \inf\{n \geq 1 : A_n \geq \gamma^k\}$,

$$A_{n_k} \leq \gamma A_{n_k-1} < \gamma^{k+1} \leq A_{n_{k+1}},$$

and so $\{n_k, k \geq 1\}$ is strictly increasing. Moreover, for all $k \geq 1$

$$\frac{A_{n_k}}{A_{n_{k-1}}} < \frac{\gamma^{k+1}}{\gamma^{k-1}} = \gamma^2 > 1.$$

Therefore, setting $U_n = \sum_{i=1}^n (Y_i - E Y_i)$, for all $\varepsilon > 0$

$$\begin{aligned} P\{U_n > 2\gamma^2 \varepsilon A_n (\log_2 A_n)^{1-\beta}, \text{i.o.}\} \\ &\leq P\left\{\max_{n_{k-1} < n \leq n_k} U_n > 2\gamma^2 \varepsilon A_{n_{k-1}} (\log_2 A_{n_{k-1}})^{1-\beta}, \text{i.o.}\right\} \\ &\leq P\left\{\max_{1 \leq n \leq n_k} U_n > \varepsilon A_{n_k} (\log_2 A_{n_k})^{1-\beta}, \text{i.o.}\right\}. \end{aligned} \quad (40)$$

Now, $|Y_n - E Y_n| \leq c_n t_n$, where $c_n = 2c A_n (\log_2 A_n)^{-\beta}/t_n$ and

$$\begin{aligned} t_n^2 = E U_n^2 &\leq \sum_{i=1}^n E X_i^2 I_{[|X_i| \leq c A_i (\log_2 A_i)^{-\beta}]} \\ &\leq \frac{c^{2-\alpha} A_n^{2-\alpha}}{(\log_2 A_n)^{\beta(2-\alpha)}} \sum_{i=1}^n E |X_i|^\alpha = \frac{c^{2-\alpha} A_n^2}{(\log_2 A_n)^{\beta(2-\alpha)}}. \end{aligned} \quad (41)$$

Consequently, setting

$$\lambda_n = \frac{\varepsilon A_n^2 (\log_2 A_n)^{1-2\beta}}{t_n^2}, \quad x_n = \frac{t_n (\log_2 A_n)^\beta}{A_n}$$

it follows that $c_n \cdot x_n = 2c$. Since (41) ensures that $2c \lambda_n x_n \geq 2c^{\alpha-1} \varepsilon (\log_2 A_n)^{1-\beta\alpha} \rightarrow \infty$ and $h(x) \geq (x/2) \log(1+x)$ as $x \rightarrow \infty$, for all large n (see 10.2(1))

$$x_n^2 h(2c \lambda_n) \geq c \lambda_n x_n^2 \log(1 + 2c \lambda_n) = \varepsilon (\log_2 A_n) \log(1 + 2c \lambda_n) > 8c^2 \log_2 A_n.$$

Thus, via (3) of Lemma 10.2.1

$$\begin{aligned} P\left\{\max_{1 \leq n \leq n_k} U_n > \varepsilon A_{n_k} (\log_2 A_{n_k})^{1-\beta}\right\} &= P\left\{\max_{1 \leq n \leq n_k} U_n > \lambda_{n_k} x_{n_k} t_{n_k}\right\} \\ &\leq \exp\{-\frac{1}{2} x_{n_k}^2 h(2c \lambda_{n_k})/4c^2\} \leq \exp\{-2 \log_2 A_{n_k}\} \leq \frac{1}{(k \log \gamma)^2} \end{aligned}$$

and so (40) and the Borel–Cantelli lemma ensure

$$\overline{\lim}_{n \rightarrow \infty} \frac{U_n}{A_n (\log_2 A_n)^{1-\beta}} \leq 0, \text{ a.c.} \quad (42)$$

Since $\{-(Y_n - E Y_n), n \geq 1\}$ have the same bounds and variances as $\{Y_n - E Y_n, n \geq 1\}$, (42) likewise obtains with $-U_n$ replacing U_n thereby proving (39) and the theorem. \square

3. Let $\{X_n, n \geq 1\}$ be independent r.v.s with $E X_n = 0$, $E X_n^2 = \sigma_n^2 < \infty$, $s_n^2 = \sum_1^n \sigma_i^2 \rightarrow \infty$, and $\sigma_n = o(s_n)$. Prove the result of Raikov that

$$\frac{1}{s_n} \sum_{j=1}^n X_j \xrightarrow{d} N_{0,1} \quad \text{iff} \quad \frac{1}{s_n^2} \sum_{j=1}^n X_j^2 \xrightarrow{P} 1.$$

Hint: Apply Corollary 2 to $Y_{nj} = X_j^2/s_n^2$, $1 \leq j \leq n$,

4. Necessary and sufficient conditions for the strong law for independent $\{X_n\}$ with $E X_n = 0$, $E X_n^2 = \sigma_n^2 < \infty$ cannot be framed solely in terms of variances. *Hint:* Consider

$$P\{X_n = \pm cn\} = \frac{1}{2c^2 n \log n} = \frac{1}{2} [1 - P\{X_n = 0\}], \quad n \geq 2,$$

and

$$P\left\{X_n = \pm \frac{n}{\log n}\right\} = \frac{\log n}{2n} = \frac{1}{2} [1 - P\{X_n = 0\}], \quad n \geq 2.$$

5. Define a sequence of r.v.s for which $(1/n) \sum_1^n X_i \xrightarrow{a.s.} 0$ but $\sum (X_n/n)$ diverges a.c., revealing limitations to the "Kronecker lemma approach."
6. Let $\{X_n, n \geq 1\}$ be independent r.v.s with $E X_n = 0$, $E X_n^2 = \sigma_n^2$, satisfying (28).
- (i) Show that if $\sigma_n^2 \sim n/(\log n)^\delta$, $\delta > 0$, then $(*) (1/n) \sum_1^n X_i \xrightarrow{a.s.} 0$.
 - (ii) If, rather, $\sigma_n^2 \sim n/(\log \log n)^\delta$, gives necessary and sufficient conditions for $(*)$ in terms of δ when $|X_n| = O(n/\log \log n)$, a.c.
7. Let $\{X_n, n \geq 1\}$ be independent r.v.s with $E X_n = 0$, $E X_n^2 = \sigma_n^2$, $s_n^2 = \sum_1^n \sigma_i^2$, and $\sigma_n^2 s_{n-1}^2 \leq E X_n^4$, where $(*) \sum_{n=1}^{\infty} n^{-4} E X_n^4 < \infty$. Prove that the classical SLLN holds. Compare $(*)$ with (33) when $r = 2$.
8. Show that Theorem 2 extends to the case where $\{X_n, n \geq 1\}$ are martingale differences with constant conditional variances σ_n^2 .
9. If $\{X_n\}$ are independent r.v.s with variances σ_n^2 , $s_n^2 = \sum_1^n \sigma_i^2 = o(b_n^2)$, and $\sum_1^{\infty} (E X_n^4/b_n^4) < \infty$, where $b_n \uparrow \infty$, then $(1/b_n^2) \sum_{i=1}^n X_i^2 \xrightarrow{a.s.} 0$.
10. Let $X_n = b^n Y_n$, $n \geq 1$, $b > 1$, where $\{Y_n\}$ are bounded i.i.d. random variables. Prove that $(1/b_n) \sum_{i=1}^n X_i \xrightarrow{a.s.} 0$ provided $b_n/b^n \rightarrow \infty$. Compare with Exercise 5.2.8.
11. Let $\{X_n, n \geq 1\}$ be independent, symmetric r.v.s and $X'_n = X_n I_{\{|X_n| < b_n\}}$, where $0 < b_n < \infty$. If $X_n^* = X_n I_{\{|X_n| \leq b_n\}} - X_n I_{\{|X_n| > b_n\}}$, then $\{X_n^*, n \geq 1\}$ are independent, symmetric r.v.s with the same joint d.f.s as $\{X_n, n \geq 1\}$. Let S_n, S'_n, S_n^* be the corresponding partial sums and $\varepsilon_n > 0$. Then

$$[S'_n > \varepsilon_n] \subset [S_n > \varepsilon_n] \cup [S_n^* > \varepsilon_n]$$

and

$$P[S'_n > \varepsilon_n, \text{i.o.}] \leq 2 P[S_n > \varepsilon_n, \text{i.o.}].$$

12. Let $X_n = \sigma_n Y_n$, $n \geq 1$, where $\{Y_n\}$ are bounded i.i.d. r.v.s and

$$\sigma_n^2 = (\log n)^{-1} \exp\left\{\frac{2\lambda n}{\log n}\right\}, \quad n \geq 1, \lambda > 0.$$

Show that $s_n^2 = \sum_1^n \sigma_i^2 \sim (2\lambda)^{-1} \exp(2\lambda n/\log n)$ and that $1/b_n \sum_{i=1}^n X_i \xrightarrow{a.s.} 0$ whenever $s_n^2 \log \log s_n^2 = o(b_n^2)$.

10.2 Law of the Iterated Logarithm

One of the most beautiful and profound discoveries of probability theory is the celebrated iterated logarithm law. This law, due in the case of suitably bounded independent random variables to Kolmogorov, was the culmination of a series of strides by mathematicians of the caliber of Hausdorff, Hardy, Littlewood, and Khintchine. The crucial instruments in the proof are sharp exponential bounds for the probability of large deviations of sums of independent r.v.s with vanishing means.

The next lemmas are generalized analogues of Kolmogorov's exponential bounds. The probabilistic inequality (3) is of especial interest in the cases (i) $\lambda_n \equiv \lambda$, $x_n \rightarrow \infty$, $c_n \rightarrow 0$, and (ii) $\lambda_n \rightarrow \infty$, $c_n x_n = a > 0$.

Define

$$\begin{aligned} h(x) &= (1+x)\log(1+x) - x, \quad x \geq 0 \\ g(x) &= x^{-2}(e^x - 1 - x), \quad -\infty < x < \infty \end{aligned} \tag{1}$$

Lemma 1. Let $S_n = \sum_{j=1}^n X_j$ where $\{X_j, 1 \leq j \leq n\}$ are independent r.v.s with $E X_j = 0$, $E X_j^2 = \sigma_j^2$, $s_n^2 = \sum_{j=1}^n \sigma_j^2 > 0$.

(i) If $P\{X_j \leq c_n s_n\} = 1$, $1 \leq j \leq n$, then

$$E e^{tS_n/s_n} \leq e^{t^2 g(c_n t)}, \quad t > 0, \tag{2}$$

and if, in addition, $c_n x_n \leq a_n$, then for all λ_n and x_n such that $a_n \cdot x_n > 0$, $a_n \lambda_n > 0$

$$P\left\{\max_{1 \leq j \leq n} S_j \geq \lambda_n x_n s_n\right\} \leq e^{-h(a_n \lambda_n) x_n^2 / a_n^2} \tag{3}$$

(ii) If $P\{X_j \geq -c_n s_n\} = 1$, $1 \leq j \leq n$, then

$$E e^{tS_n/s_n} \geq \exp\left\{t^2 g(-c_n t) \left[1 - \frac{t^2 g(-tc_n)}{s_n^4} \sum_{j=1}^n \sigma_j^4\right]\right\}, \quad t > 0, \tag{4}$$

and if, in addition, $\sigma_j \leq c_n s_n$, $1 \leq j \leq n$,

$$\begin{aligned} E e^{tS_n/s_n} &\geq \exp\{t^2 g(-tc_n)[1 - t^2 c_n^2 g(-tc_n)]\} \\ &\geq \exp\{t^2 g(-tc_n)[1 - (t^2 c_n^2 / 2)]\}, \quad t > 0. \end{aligned} \tag{5}$$

PROOF. The representation $g(x) = \int_0^1 \int_0^x e^{uy} dy du$ shows that g is non-negative, increasing and convex. The point of departure for (2), (3), (4) is the simple observation

$$E e^{tX_j/s_n} = 1 + E\left[e^{tX_j/s_n} - 1 - \frac{tX_j}{s_n}\right] = 1 + t^2 E \frac{X_j^2}{s_n^2} g\left(\frac{tX_j}{s_n}\right). \tag{6}$$

Hence, monotonicity ensures under (i) that if $t > 0$,

$$E e^{tX_j/s_n} \leq 1 + \frac{t^2 \sigma_j^2}{s_n^2} g(tc_n) \leq \exp\{t^2 g(tc_n) \sigma_j^2 / s_n^2\} \tag{7}$$

and (1) follows via independence.

If, rather, (ii) obtains, then (6) in conjunction with the elementary inequality $(1+u)e^{u^2} > e^u$, $u > 0$, yields for $t > 0$

$$\mathbb{E} e^{tX_j/s_n} \geq 1 + \frac{t^2 \sigma_j^2}{s_n^2} g(-tc_n) \geq \exp \left\{ \frac{t^2 \sigma_j^2}{s_n^2} g(-tc_n) - \frac{t^4 \sigma_j^4}{s_n^4} g^2(-tc_n) \right\},$$

whence (4) is immediate. Under the additional hypothesis $\sigma_j \leq c_n s_n$, $1 \leq j \leq n$, necessarily $s_n^{-4} \sum_{j=1}^n \sigma_j^4 \leq c_n^2$, and (5) follows from (4) in view of $g(0) = \frac{1}{2}$.

To establish (3), note that via Example 7.4.9 and (1), for $t > 0$

$$\mathbb{P} \left\{ \max_{1 \leq j \leq n} S_j \geq \lambda x_n s_n \right\} \leq e^{-\lambda x_n s_n} \mathbb{E} e^{\lambda S_n} \leq \exp \{-\lambda x_n s_n + t^2 s_n^2 g(c_n s_n t)\}, \quad (8)$$

and so, setting $a = a_n$ and $t = bx_n/s_n$ where $bx_n > 0$,

$$\begin{aligned} \mathbb{P} \left\{ \max_{1 \leq j \leq n} S_j \geq \lambda x_n s_n \right\} &\leq \exp \{-x_n^2 [\lambda b - b^2 g(c_n x_n b)]\} \\ &\leq \exp \{-x_n^2 [\lambda b - b^2 g(ab)]\} \end{aligned}$$

Employing the definition (1) of g , the prior exponent is minimized at $b = a^{-1} \log(1 + a\lambda)$ with a value of

$$\frac{-x_n^2}{a^2} [(1 + a\lambda) \log(1 + a\lambda) - a\lambda] = \frac{-x_n^2}{a^2} h(a\lambda)$$

Clearly, nothing precludes λ or a from being a function of n . \square

The simple inequality of (3) yields a generalization of the easier half of the law of the iterated logarithm. In what follows $\log_2 n$ will abbreviate $\log \log n$.

Corollary 1. Let $\{X_n, n \geq 1\}$ be independent r.v.s with $\mathbb{E} X_n = 0$, $\mathbb{E} X_n^2 = \sigma_n^2$, $s_n^2 = \sum_1^n \sigma_i^2 \rightarrow \infty$.

(i) If $\mathbb{P}\{X_n \leq d_n\} = 1$, $n \geq 1$ where $d_n > 0$, then with probability one

$$\overline{\lim}_{n \rightarrow \infty} \frac{S_n}{s_n (\log_2 s_n^2)^{1/2}} \leq \sqrt{2} \quad \text{or} \quad \overline{\lim}_{n \rightarrow \infty} \frac{S_n}{s_n (\log_2 s_n^2)^{1/2}} \leq \frac{1}{a} h^{-1}(a^2) \quad (9)$$

according as $(\log_2 s_n^2)^{1/2} d_n / s_n \rightarrow 0$ or $(\log_2 s_n^2)^{1/2} d_n / s_n \rightarrow a > 0$.

(ii) If $\mathbb{P}\{|X_n| \leq d_n\} = 1$, $n \geq 1$ where $(\log_2 s_n^2)^{1/2} d_n / s_n \rightarrow a$, then with probability one,

$$\overline{\lim}_{n \rightarrow \infty} \frac{|S_n|}{s_n (\log_2 s_n^2)^{1/2}} \leq \frac{1}{a} h^{-1}(a^2). \quad (10)$$

PROOF. Let $b_n = s_n (\log_2 s_n^2)^{-1/2}$ and suppose $d_n/b_n \rightarrow a \geq 0$. Since $0 < b_n \uparrow \infty$, it follows that $b_n^{-1}(\max_{1 \leq i \leq n} d_i) \rightarrow a$ and so d_n may be taken to be increasing.

For any $\alpha > 1$, define $n_0 = \inf\{n \geq 1: s_n \geq \alpha\}$ and $n_k = \inf\{n > n_{k-1}: s_n \geq \alpha s_{n_{k-1}}\}$, $k \geq 1$. Then $s_{n_{k-1}} \leq s_{n_k} < \alpha s_{n_{k-1}}$ so that $s_{n_k} \geq \alpha^k$ and $\log_2 s_{n_k}^2 \sim \log_2 s_{n_{k-1}}^2$. Hence,

$$\begin{aligned} \mathbb{P}\{S_n > \lambda\alpha^2 s_n (\log_2 s_n^2)^{1/2}, \text{i.o.}\} &\leq \mathbb{P}\left\{\max_{n_{k-1} \leq n < n_k} S_n > \lambda\alpha^2 s_{n_{k-1}} (\log_2 s_{n_{k-1}}^2)^{1/2}, \text{i.o.}\right\} \\ &\leq \mathbb{P}\left\{\max_{1 \leq j < n_k} S_j > \lambda s_{n_{k-1}} (\log_2 s_{n_{k-1}}^2)^{1/2}, \text{i.o.}\right\} \quad (11) \end{aligned}$$

For any $\gamma > 0$, setting $\lambda = (1/u)h^{-1}(\gamma)$, $x_n = (\log_2 s_n^2)^{1/2}$, $n = n_k - 1$ and noting for any $u > a$ and all large n that $c_n(\log_2 s_n^2)^{1/2} = (\log_2 s_n^2)^{1/2}d_n/s_n < u$, Lemma 1 ensures that for all large k

$$\begin{aligned} \mathbb{P}\left\{\max_{1 \leq j < n_k} S_j > \frac{1}{u} h^{-1}(\gamma) s_{n_{k-1}} (\log_2 s_{n_{k-1}}^2)^{1/2}\right\} &\leq \exp\left\{-\frac{1}{u^2} h(h^{-1}(\gamma)) \log_2 s_{n_{k-1}}^2\right\} \\ &\leq (2k \log \alpha)^{-\gamma/u^2} \end{aligned}$$

Hence, via (11) and the Borel–Cantelli lemma, for all $\gamma > u^2$

$$\mathbb{P}\left\{S_n > \frac{\alpha^2}{u} h^{-1}(\gamma) s_n (\log_2 s_n^2)^{1/2}, \text{i.o.}\right\} = 0$$

whence with probability one for $a > 0$,

$$\overline{\lim}_{n \rightarrow \infty} \frac{S_n}{s_n (\log_2 s_n^2)^{1/2}} \leq \frac{\alpha^2}{a} \inf_{\gamma > u^2} h^{-1}(\gamma) \downarrow \frac{h^{-1}(a^2)}{a} \quad (12)$$

proving the second part of (i). If rather, $a = 0$ then (12) holds for arbitrarily small a . Since $h(a) \sim a^2/2$ as $a \rightarrow 0$, necessarily $h^{-1}(a) \sim (2a)^{1/2}$ whence $(1/a)h^{-1}(a^2) \rightarrow \sqrt{2}$ as $a \rightarrow 0$ yielding the first portion of (i). Finally, under (ii), (12) holds for both S_n and $-S_n$ so that (10) obtains. \square

For any positive integer n , let $\{X_{n,j}, 1 \leq j \leq n\}$ constitute independent r.v.s with d.f.s $F_{n,j}$ and finite moment generating functions $\varphi_{n,j}(t) = \exp\{\psi_{n,j}(t)\}$ for $0 \leq t < t_0$. Suppose that $S_{n,n} = \sum_{j=1}^n X_{n,j}$ has d.f. F_n . For any t in $[0, t_0]$, define **associated d.f.s** $F_{n,j}^{(t)}$ by

$$F_{n,j}^{(t)}(x) = \frac{1}{\varphi_{n,j}(t)} \int_{-\infty}^x e^{iy} dF_{n,j}(y)$$

and let $\{X_{n,j}(t), 1 \leq j \leq n\}$ be (fictitious) independent r.v.s with d.f.s $\{F_{n,j}^{(t)}, 1 \leq j \leq n\}$. Since the c.f. of $X_{n,j}(t)$ is $\varphi_{n,j}(t + iu)/\varphi_{n,j}(t)$, setting $\psi_n(t) = \sum_{j=1}^n \psi_{n,j}(t)$, the c.f. of $S_n(t) = \sum_{j=1}^n X_{n,j}(t)$ is given by

$$\mathbb{E} e^{iuS_n(t)} = \prod_{j=1}^n \frac{\varphi_{n,j}(t + iu)}{\varphi_{n,j}(t)} = \exp\{\psi_n(t + iu) - \psi_n(t)\}.$$

Thus, the mean and variance of $S_n(t)$ are $\psi'_n(t)$ and $\psi''_n(t)$ respectively and, moreover, the d.f. of $S_n(t)$ is

$$F_n^{(t)}(x) = e^{-\psi_n(t)} \int_{-\infty}^x e^{iy} dF_n(y),$$

$|g(x) - \frac{1}{2}| < (|x|/6)(1 - (|x|/4))^{-1}$ for $0 < |x| < 4$, if $\lim t_n c_n = 0$,

$$\psi_n''(t_n) = \sum_{j=1}^n \psi_{n,j}''(t_n) = 1 + O(t_n c_n). \quad (16)$$

Thus, via (5), (15), (16) and $g(0) = \frac{1}{2}$, $g_1(0) = 1$, for any γ in $(0, 1)$ and all large n

$$\psi_n(t_n) - t_n \psi_n'(t_n) \geq t_n^2 [g(-t_n c_n) - t_n^2 c_n^2 g^2(-t_n c_n) - g_1(t_n c_n)] \geq \frac{-t_n^2}{2} (1 + \gamma),$$

$$v_n \equiv \frac{(1 - \gamma)t_n - \psi_n(t_n)}{\sqrt{\psi_n''(t_n)}} = -\gamma t_n (1 + o(1)) \leq \frac{-\gamma t_n}{2}.$$

Consequently, taking $u = (1 - \gamma)t_n$ in (13),

$$\begin{aligned} P\{S_n > (1 - \gamma)s_n t_n\} &\geq \exp\{\psi_n(t_n) - t_n \psi_n'(t_n)\} \\ &\times \int_{s_n}^0 \exp\{-ty\sqrt{\psi_n''(t_n)}\} dF_n^{(t_n)}(y\sqrt{\psi_n''(t_n)} + \psi_n'(t_n)) \\ &\geq \exp\{-t_n^2/2(1 + \gamma)\} \int_{-\gamma t_n/2}^0 dF_n^{(t_n)}(y\sqrt{\psi_n''(t_n)} + \psi_n'(t_n)) \\ &\geq C_\gamma \exp\{(-t_n^2/2)(1 + \gamma)\} \end{aligned} \quad (17)$$

since

$$\frac{S_n(t_n) - \psi_n'(t_n)}{\sqrt{\psi_n''(t_n)}} = \sum_{j=1}^n Z_{n,j} \xrightarrow{d} N_{0,1}$$

by Exercise 9.1.2 or Corollary 12.2.2 in view of

$$\begin{aligned} E Z_{n,j} &= 0, \quad \sum_{j=1}^n E Z_{n,j}^2 = 1, \quad \text{and} \quad |Z_{n,j}| = \left| \frac{X_{n,j}(t_n) - E X_{n,j}(t_n)}{\sqrt{\psi_n''(t_n)}} \right| \\ &\leq \frac{2c_n}{\sqrt{\psi_n''(t_n)}} = o(1). \end{aligned}$$

Finally, set $t_n = (1 - \gamma)x_n$ in (17) to obtain (14). \square

Remark. If $x_n \rightarrow \infty$, then for every γ in $(0, 1)$ the constant $C_\gamma > \frac{1}{2} - \varepsilon$ for all $\varepsilon > 0$ provided $n \geq$ some integer N_ε .

The strong law asserts under certain conditions that with probability one sums S_n of independent r.v.s with zero means are $o(n)$. In the symmetric Bernoulli case, Hausdorff proved in 1913 that $S_n \xrightarrow{sc} O(n^{(1/2)+\varepsilon})$, $\varepsilon > 0$. The order of magnitude was improved to $O(\sqrt{n \log n})$ by Hardy and Littlewood in 1914 and to $O(\sqrt{n \log_2 n})$ by Khintchine in 1923. (Here, as elsewhere, $\log_2 n$ denotes $\log \log n$ and $\log_{k+1} n = \log \log_k n$, $k \geq 1$). One year later Khintchine obtained the iterated logarithm law for the special case in question and in 1929 Kolmogorov proved

Theorem 1 (Law of the Iterated Logarithm). *Let $\{X_n, n \geq 1\}$ be independent r.v.s with $E X_n = 0$, $E X_n^2 = \sigma_n^2$, $s_n^2 = \sum_1^n \sigma_i^2 \rightarrow \infty$. If $|X_n| \leq d_n$, a.c., where the constant $d_n = o(s_n / (\log_2 s_n)^{1/2})$ as $n \rightarrow \infty$, then, setting $S_n = \sum_{i=1}^n X_i$,*

$$P \left\{ \overline{\lim}_{n \rightarrow \infty} \frac{S_n}{s_n (\log_2 s_n)^{1/2}} = \sqrt{2} \right\} = 1. \quad (18)$$

PROOF. Choose the integers n_k , $k \geq 1$ such that $s_{n_k} \leq \alpha^k < s_{n_{k+1}}$ and note that $\sigma_n^2 / s_n^2 = o(1)$ whence $s_{n_k} \sim \alpha^k$. According to Corollary 1,

$$\overline{\lim}_{n \rightarrow \infty} \frac{S_n}{s_n \sqrt{\log_2 s_n}} \leq \sqrt{2}, \quad \text{a.c.} \quad (19)$$

To establish the reverse inequality, choose γ in $(0, 1)$ and define independent events

$$A_k = \{S_{n_k} - S_{n_{k-1}} > (1 - \gamma)^2 g_k h_k\}, \quad k \geq 1,$$

where, since $s_{n_k} \sim \alpha^k$, $\alpha > 1$,

$$\begin{aligned} g_k^2 &\equiv s_{n_k}^2 - s_{n_{k-1}}^2 \sim s_{n_k}^2 \left(1 - \frac{1}{\alpha^2}\right) \\ h_k^2 &\equiv 2 \log_2 g_k \sim 2 \log_2 s_{n_k} < (1 + \gamma)(2 \log k) \end{aligned} \quad (20)$$

for all large k . Thus, taking $x_{n_k} = h_k$ in Lemma 2, noting (20) and that $d_{n_k} h_k / g_k = o(1)$,

$$\begin{aligned} P\{A_k\} &\geq C_\gamma \exp\{-h_k^2(1 - \gamma)(1 - \gamma^2)/2\} \geq C_\gamma \exp\{-(1 - \gamma^2)^2 \log k\} \\ &= \frac{C_\gamma}{k^{(1 - \gamma^2)^2}} \end{aligned}$$

for all large k , whence by the Borel-Cantelli theorem

$$P\{A_k, \text{i.o.}\} = 1. \quad (21)$$

Next, choose α so large that $(1 - \gamma)^2(1 - \alpha^{-2})^{1/2} - (2/\alpha) > (1 - \gamma)^3$ and set $t_n = \sqrt{2 \log_2 s_n}$, implying for all large k that

$$\begin{aligned} (1 - \gamma)^2 g_k h_k - 2s_{n_{k-1}} t_{n_{k-1}} &\sim [(1 - \gamma)^2(1 - \alpha^{-2})^{1/2} - 2\alpha^{-1}] s_{n_k} t_{n_k} \\ &> (1 - \gamma)^3 s_{n_k} t_{n_k}. \end{aligned}$$

Hence, setting $B_k = \{|S_{n_{k-1}}| \leq 2s_{n_{k-1}} t_{n_{k-1}}\}$,

$$A_k B_k \subset \{S_{n_k} > (1 - \gamma)^2 g_k h_k - 2s_{n_{k-1}} t_{n_{k-1}}\} \subset \{S_{n_k} > (1 - \gamma)^3 s_{n_k} t_{n_k}\}$$

again for all large k . However, (ii) of Corollary 1 guarantees $P\{B_k^c, \text{i.o.}\} = 0$, which, in conjunction with (21), entails

$$P\{S_{n_k} > (1 - \gamma)^3 s_{n_k} t_{n_k}, \text{i.o.}\} \geq P\{A_k \cdot B_k, \text{i.o.}\} = 1.$$

Thus, with probability one

$$\varlimsup_{n \rightarrow \infty} \frac{S_n}{s_n t_n} \geq \varlimsup_{k \rightarrow \infty} \frac{S_{n_k}}{s_{n_k} t_{n_k}} \geq (1 - \gamma)^3,$$

and letting $\gamma \downarrow 0$ the reverse inequality of (19) is proved. \square

Corollary 2. Under the conditions of Theorem 1

$$P\left\{\varliminf_{n \rightarrow \infty} \frac{S_n}{s_n \sqrt{\log_2 s_n}} = -\sqrt{2}\right\} = 1 = P\left\{\varlimsup_{n \rightarrow \infty} \frac{|S_n|}{s_n \sqrt{\log_2 s_n}} = \sqrt{2}\right\}.$$

To extend the law of the iterated logarithm (LIL) from bounded to unbounded r.v.s without losing ground, a refined truncation is necessary. This means that the truncation constants, far from being universal, should (as first realized by Hartman and Wintner in the i.i.d. case) depend upon the tails of the distributions of the r.v.s involved.

Let $\{X_n, n \geq 1\}$ denote independent random variables with $E X_n = 0$, $E X_n^2 = \sigma_n^2$, $s_n^2 = \sum_{i=1}^n \sigma_i^2 \rightarrow \infty$. Then $\{X_n, n \geq 1\}$ obeys the LIL if (17) obtains.

Theorem 2. If $\{X_n, n \geq 1\}$ are independent r.v.s with $E X_n = 0$, $E X_n^2 = \sigma_n^2$, $s_n^2 = \sum_{i=1}^n \sigma_i^2 \rightarrow \infty$, in order that $\{X_n, n \geq 1\}$ obey the LIL it is necessary that

$$\sum_{n=1}^{\infty} P\{X_n > \delta s_n (\log_2 s_n^2)^{1/2}\} < \infty, \quad \delta > \sqrt{2}. \quad (22)$$

PROOF. If $b_n^2 = 2s_n^2 \log_2 s_n^2$ and $S_n = \sum_{i=1}^n X_i$, $S_0 = 0$, then $\overline{\lim}_{n \rightarrow \infty} S_n/b_n \stackrel{a.s.}{=} 1$. Now $S_{n-1}/b_n \xrightarrow{P} 0$ since $\sigma^2(S_{n-1}) = s_{n-1}^2 = o(b_n^2)$, and clearly S_{n-1} is independent of (X_n, X_{n+1}, \dots) for all $n \geq 1$. Hence, by Lemma 3.3.4(ii) $\lim X_n/b_n \leq 1 + \varepsilon$, a.c. and (22) follows by the Borel–Cantelli theorem. \square

Corollary 3. Under the hypothesis of Theorem 2, in order that both $\{X_n\}$ and $\{-X_n\}$ obey the LIL, it is necessary that

$$\sum_{n=1}^{\infty} P\{|X_n| > \delta s_n (\log_2 s_n^2)^{1/2}\} < \infty, \quad \delta > \sqrt{2}. \quad (23)$$

The next result stipulates two conditions which, conjoined with (23) for a fixed δ , are sufficient for the LIL. One of these, (25), clearly implies the Lindeberg criterion and hence the asymptotic normality of $\sum_{i=1}^n X_i/s_n$.

Theorem 3. If $\{X_n, n \geq 1\}$ are independent r.v.s with $E X_n = 0$, $E X_n^2 = \sigma_n^2$, $s_n^2 = \sum_{i=1}^n \sigma_i^2 \rightarrow \infty$, and d.f.s $\{F_n, n \geq 1\}$ satisfying for some $\delta > 0$

$$\sum_{n=1}^{\infty} P\{|X_n| > \delta s_n (\log_2 s_n^2)^{1/2}\} < \infty, \quad (24)$$

and set

$$\begin{aligned} X'_n &= X_n I_{\{|X_n| \leq b_n\}}, & X''_n &= X_n I_{\{|X_n| > \delta s_n (\log_2 s_n^2)^{1/2}\}}, \\ X''_n &= X_n - X'_n - X'''_n, & S'_n &= \sum_1^n X'_j, & S''_n &= \sum_1^n X''_j, & S'''_n &= \sum_1^n X'''_j. \end{aligned} \quad (30)$$

Now

$$\sigma_n^2 - \sigma_{X_n}^2 = E X_n^2 I_{\{|X_n| > b_n\}} + E^2 X_n I_{\{|X_n| \leq b_n\}} \leq 2 E X_n^2 I_{\{|X_n| > b_n\}},$$

recalling that $E X_n = 0$, and so (28) ensures $\sigma_{S'_n}^2 \sim \sigma_{S_n}^2$. Thus Theorem 1 yields

$$\lim_{n \rightarrow \infty} \frac{S'_n - E S'_n}{s_n (\log_2 s_n^2)^{1/2}} \xrightarrow{\text{a.c.}} 1. \quad (31)$$

Secondly, Kronecker's lemma and (29) guarantee that

$$\frac{S''_n - E S''_n}{s_n (\log_2 s_n^2)^{1/2}} \xrightarrow{\text{a.c.}} 0. \quad (32)$$

Thirdly, (24) implies that $S'''_n = O(1)$ with probability one, and, furthermore,

$$\begin{aligned} |E S'''_n| &\leq \sum_{j=1}^n \int_{\{|x| > \delta s_j (\log_2 s_j^2)^{1/2}\}} |x| dF_j(x) \\ &\leq \sum_{j=1}^n \int_{[\delta s_j (\log_2 s_j^2)^{1/2} < |x| \leq s_n (\log_2 s_n^2)^{-1/2}]} |x| dF_j(x) \\ &\quad + \sum_{j=1}^n \int_{[|x| > s_n (\log_2 s_n^2)^{-1/2}]} |x| dF_j(x) \\ &\leq \frac{s_n}{(\log_2 s_n^2)^{1/2}} \sum_{j=1}^n P\{|X_j| > \delta s_j (\log_2 s_j^2)^{1/2}\} \\ &\quad + \frac{(\log_2 s_n^2)^{1/2}}{s_n} \sum_{j=1}^n \int_{[|x| > s_n (\log_2 s_n^2)^{-1/2}]} x^2 dF_j(x) = o(s_n (\log_2 s_n^2)^{1/2}) \end{aligned}$$

via (24) and (25).

The first portion of the theorem is an immediate consequence of (30), (31), (32), and the assertion just after (32).

In the alternative case, note that since $\gamma_n(\varepsilon)$ and hence the series of (27) is decreasing in ε , there exists, as earlier, a sequence $\varepsilon_n = o(1)$ such that (25) and (27) hold with ε replaced by ε_i and ε_n respectively. Define

$$b_n = \varepsilon_n s_n (\log_2 s_n^2)^{-1/2}$$

and X'_n, X''_n, X'''_n as in (30), but now with $\delta = 1$ and the new choice of b_n . The only link in the prior chain of argument requiring modification is that used to establish (32).

PROOF. For all $\varepsilon > 0$,

$$\begin{aligned} \sum_{n=k}^{\infty} \frac{\gamma_n}{(s_n^2 \log_2 s_n^2)^k} \sum_{j_{k-1}=k-1}^{n-1} \gamma_{j_{k-1}} \cdots \sum_1^{j_2-1} \gamma_{j_1} &\leq \sum_{n=k}^{\infty} \frac{\gamma_n}{\sum_1^n \gamma_j} \left(\frac{\sum_1^n \gamma_j}{s_n^2 \log_2 s_n^2} \right)^k \\ &\leq C \sum_{n=k}^{\infty} \frac{\gamma_n}{(\sum_1^n \gamma_j)(\log \sum_1^n \gamma_j)^{pk}} < \infty \end{aligned}$$

for $k > 1/p$. \square

All the prior conditions for the LIL simplify greatly in the special case of weighted i.i.d. random variables. Define such a class Q by

$$Q = \{\sigma_n Y_n, n \geq 1 : Y_n, n \geq 1, \text{ are i.i.d. random variables with mean 0, variance } \sigma_Y^2 \text{ in } (0, \infty] \text{ and } \sigma_n, n \geq 1 \text{ are nonzero constants satisfying } s_n^2 = \sum_1^n \sigma_i^2 \rightarrow \infty\} \quad (36)$$

and let F denote the common d.f. of $\{Y_n\}$.

To obtain the classical Hartman-Wintner theorem governing the i.i.d. case only part (i) of the following theorem is needed.

Theorem 4. Let $\{\sigma_n Y_n\} \in Q$ with $\sigma_n^2 = o(s_n^2/\log_2 s_n^2)$ and $\sigma_Y^2 = 1$. If either (i) for some α in $(0, 2]$

$$\sum_{n=1}^{\infty} \left(\frac{\sigma_n^2}{s_n^2 \log_2 s_n^2} \right)^{\alpha/2} \int_{\{y^2 \geq \varepsilon s_n^2/\sigma_n^2 \log_2 s_n^2\}} |y|^2 dF(y) < \infty \quad \text{for all } \varepsilon > 0 \quad (37)$$

or (ii)

$$\sum_{n=1}^{\infty} P \left\{ Y_1^2 > \frac{\delta s_n^2 \log_2 s_n^2}{\sigma_n^2} \right\} < \infty \quad \text{for all } \delta > 0 \quad (38)$$

and for some $\rho > 0$,

$$q_s(\varepsilon) \equiv \int_{[\varepsilon s_n^2/\sigma_n^2 \log_2 s_n^2 < y^2 \leq (s_n^2 \log_2 s_n^2)/\sigma_n^2]} y^2 dF(y) = O \left(\frac{\log_2 s_n^2}{(\log s_n^2)^\rho} \right) \quad \text{for all } \varepsilon > 0 \quad (39)$$

then the LIL holds for $\{\sigma_n Y_n\}$, that is,

$$P \left\{ \lim_{n \rightarrow \infty} \frac{\sum_{j=1}^n \sigma_j Y_j}{s_n (2 \log_2 s_n^2)^{1/2}} = 1 \right\} = 1. \quad (40)$$

PROOF. In the weighted i.i.d. case, condition (25) becomes

$$\frac{1}{s_n^2} \sum_{j=1}^n \sigma_j^2 \int_{\{y^2 > \varepsilon s_j^2/\sigma_j^2 \log_2 s_j^2\}} y^2 dF(y) = o(1), \quad \varepsilon > 0, \quad (41)$$

and is automatic whenever the integral therein is $o(1)$, that is, whenever

$\sigma_n^2 = o(s_n^2/\log_2 s_n^2)$. The first part of the theorem thus follows from Corollary 4 since (37) is just a transcription of (34).

Likewise, (38) is a transliteration of (24), and so the second portion will follow from Corollary 5 once (35) is established. To this end, note that $\sigma_n^2 = o(s_n^2/\log_2 s_n^2)$ entails $\lim \sigma_n^2/s_{n-1}^2 = 0$, whence

$$a_n \equiv \frac{\log(1 + (\sigma_n^2/s_{n-1}^2))}{\log s_{n-1}^2} = \frac{O(\sigma_n^2/s_{n-1}^2)}{\log s_{n-1}^2} = \frac{o(1)}{\log s_{n-1}^2}$$

and

$$\log s_n^2 = (1 + a_n) \log s_{n-1}^2, \quad (42)$$

so that,

$$\begin{aligned} & s_n^2 (\log s_{n-1}^2)^\rho \log_2 s_n^2 \\ &= (s_{n-1}^2 + \sigma_n^2)[(1 + a_n)^{-1} \log s_n^2]^\rho [\log_2 s_{n-1}^2 + \log(1 + a_n)] \\ &= \left(1 + \frac{\sigma_n^2}{s_{n-1}^2}\right) \left[1 + \rho \frac{o(1)}{\log s_{n-1}^2}\right] \left[1 + \frac{o(1)}{(\log s_{n-1}^2) \log s_n^2}\right] \\ &\quad \times s_{n-1}^2 (\log s_n^2)^\rho \log_2 s_{n-1}^2 \\ &= s_{n-1}^2 (\log s_n^2)^\rho \log_2 s_{n-1}^2 + (1 + o(1)) \sigma_n^2 (\log s_n^2)^\rho \log_2 s_{n-1}^2. \end{aligned}$$

Hence, if $q_n \equiv q_n(\varepsilon) = C_\varepsilon (\log_2 s_n^2 / (\log s_n^2)^\rho)$, noting that

$$\log_i s_{n-1}^2 = (1 + o(1)) \log_i s_n^2, \quad i = 1, 2,$$

$$\frac{s_n^2 \log_2 s_n^2}{(\log s_n^2)^\rho} - \frac{s_{n-1}^2 \log_2 s_{n-1}^2}{(\log s_{n-1}^2)^\rho} = \frac{(1 + o(1)) \sigma_n^2 \log_2 s_{n-1}^2}{(\log s_{n-1}^2)^\rho} = C_\varepsilon^{-1} (1 + o(1)) \sigma_n^2 q_n,$$

implying for all large n that

$$\sum_{j=1}^n \sigma_j^2 q_j \leq 2C_\varepsilon \frac{s_n^2 \log_2 s_n^2}{(\log s_n^2)^\rho} \leq \frac{2C_\varepsilon s_n^2 \log_2 s_n^2}{(\log \sum_1^n \sigma_j^2 q_j)^\rho} \quad (43)$$

Consequently, $q_n(\varepsilon) = C_\varepsilon (\log_2 s_n^2 / (\log s_n^2)^\rho)$ (and *a fortiori* $q_n(\varepsilon) = O(\log_2 s_n^2 / (\log s_n^2)^\rho)$) entails (43). But (43) is precisely (35) in the weighted i.i.d. case since then $\gamma_n(\varepsilon) = \sigma_n^2 q_n(\varepsilon)$. \square

The status of the LIL in Q is conveniently described in terms of

$$\gamma_n = n \frac{\sigma_n^2}{s_n^2}, \quad n \geq 1. \quad (44)$$

Note that $\gamma_1 = 1$, $0 < \gamma_n < n$, $n > 1$, and

$$\frac{s_n^2}{s_1^2} = \prod_{j=2}^n \left(1 - \frac{\gamma_j}{j}\right)^{-1}. \quad (45)$$

Consequently, under the hypothesis

$$\frac{\gamma_n}{n} = \frac{\sigma_n^2}{s_n^2} \leq 1 - \frac{1}{\delta} \quad \text{for some } \delta > 1, \quad (46)$$

it follows from (45) that for some $c > 0$

$$s_n^2 \leq c\delta^n, \quad \log_2 s_n^2 \leq (1 + o(1))\log n. \quad (47)$$

Lemma 3. If $\{\sigma_n, n \geq 1\}$ satisfies (46), $s_n^2 \rightarrow \infty$ and

$$\gamma_n = o((\log s_n^2)\log_2 s_n^2), \quad (48)$$

then for every $\mu_1 > 0$ and real μ_2 necessarily $n^{\mu_1}/(\log_2 s_n^2)^{\mu_2} \uparrow \infty$ (all large n) and

$$\sum_{j=1}^n \frac{1}{j^{1-\mu_1}(\log_2 s_j^2)^{\mu_2}} = O\left(\frac{n^{\mu_1}}{(\log_2 s_n^2)^{\mu_2}}\right). \quad (49)$$

PROOF. Under (46), recalling (45) and employing

$$\left(1 - \frac{\gamma_n}{n}\right)^{-1} = 1 + \frac{\gamma_n}{n} \left(1 - \frac{\gamma_n}{n}\right)^{-1} \leq \exp\{\delta\gamma_n/n\},$$

there follows

$$\log s_n^2 = \log s_{n-1}^2 - \log\left(1 - \frac{\gamma_n}{n}\right) \leq \left(1 + \frac{\delta\gamma_n}{n \log s_{n-1}^2}\right) \log s_{n-1}^2,$$

implying

$$\log_2 s_n^2 \leq \left(1 + \frac{\delta\gamma_n}{n(\log s_{n-1}^2)\log_2 s_{n-1}^2}\right) \log_2 s_{n-1}^2.$$

Therefore, noting that these entail $\log_i s_n^2 = (1 + o(1))\log_{i-1} s_{n-1}^2$, $i = 1, 2$,

$$\begin{aligned} 1 - \left(\frac{n-1}{n}\right)^{\mu_1} \left(\frac{\log_2 s_n^2}{\log_2 s_{n-1}^2}\right)^{\mu_2} &\geq 1 - \left(1 - \frac{1}{n}\right)^{\mu_1} \left[1 + \frac{\delta\gamma_n}{n(\log s_{n-1}^2)\log_2 s_{n-1}^2}\right]^{\mu_2} \\ &\geq 1 - \left[1 - \frac{\mu_1}{n} + O\left(\frac{1}{n^2}\right)\right] \left[1 + o\left(\frac{1}{n}\right)\right] \\ &= \frac{\mu_1 + o(1)}{n} \end{aligned}$$

for $\mu_2 \geq 0$; the same conclusion is obvious when $\mu_2 < 0$, so that for all μ_2

$$\frac{n^{\mu_1}}{(\log_2 s_n^2)^{\mu_2}} - \frac{(n-1)^{\mu_1}}{(\log_2 s_{n-1}^2)^{\mu_2}} \geq \frac{\mu_1(1 + o(1))}{n^{1-\mu_1}(\log_2 s_n^2)^{\mu_2}}, \quad (50)$$

whence for all large n

$$\sum_{j=1}^n \frac{1}{j^{1-\mu_1} (\log_2 s_j^2)^{\mu_2}} \leq \frac{2}{\mu_1} \frac{n^{\mu_1}}{(\log_2 s_n^2)^{\mu_2}},$$

which is tantamount to (49). Moreover, (50) ensures that $n^{\mu_1} (\log_2 s_n^2)^{-\mu_2}$ is increasing for all large n . When $\mu_2 > 0$, (47) guarantees that it tends to ∞ as $n \rightarrow \infty$, whereas this is obvious if $\mu_2 \leq 0$. \square

Theorem 5. If $\{\sigma_n Y_n\} \in Q$, where $\sigma_n^2 = o(s_n^2/\log_2 s_n^2)$ and $\gamma_n = O((\log_2 s_n^2)^\beta)$ for some $\beta < 1$, then the LIL holds for $\{\sigma_n Y_n, n \geq 1\}$ provided $E Y^2 < \infty$.

PROOF. According to Theorem 4 it suffices to verify (37) for some α in $(0, 2]$. Now the hypotheses entail $\gamma_n = o(n)$, thus *a fortiori* (46) and also $\gamma_n \leq K(\log_2 s_n^2)^\beta$, whence Lemma 3 is applicable. Setting

$$r_j \equiv \frac{\varepsilon_j}{\gamma_j \log_2 s_j^2} \geq \frac{\varepsilon_j}{K(\log_2 s_j^2)^{1+\beta}} \equiv q_j,$$

this lemma guarantees $q_j \uparrow \infty$ all large j (and for convenience this will be supposed for all $j \geq 1$); the lemma also certifies for any α in $[0, 2]$ that

$$\sum_{j=1}^n \left(\frac{\gamma_j}{j \log_2 s_j^2} \right)^{\alpha/2} \leq K \sum_{j=1}^n j^{-\alpha/2} (\log_2 s_j^2)^{(\alpha/2)(\beta-1)} \leq K_1 n^{1-(\alpha/2)} (\log_2 s_n^2)^{(\alpha/2)(\beta-1)}.$$

Consequently, for any $\varepsilon > 0$ and some constant K_ε in $(0, \infty)$

$$\begin{aligned} & \sum_{j=1}^{\infty} \left(\frac{\sigma_j^2}{s_j^2 \log_2 s_j^2} \right)^{\alpha/2} \int_{[y^2 \geq r_j]} |y|^\alpha dF(y) \leq \sum_{j=1}^{\infty} \left(\frac{\gamma_j}{j \log_2 s_j^2} \right)^{\alpha/2} \int_{[y^2 \geq q_j]} |y|^\alpha dF(y) \\ &= \sum_{j=1}^{\infty} \left(\frac{\gamma_j}{j \log_2 s_j^2} \right)^{\alpha/2} \sum_{n=j}^{\infty} \int_{[q_n \leq y^2 < q_{n+1}]} |y|^\alpha dF(y) \\ &\leq \sum_{n=1}^{\infty} \frac{K_1 n^{1-(\alpha/2)}}{(\log_2 s_n^2)^{(1-\beta)\alpha/2}} \int_{[q_n \leq y^2 < q_{n+1}]} |y|^\alpha dF(y) \\ &\leq K_\varepsilon \sum_{n=1}^{\infty} (\log_2 s_n^2)^{1+\beta-\alpha} \int_{[q_n \leq y^2 < q_{n+1}]} y^2 dF(y) < \infty \end{aligned}$$

provided $1 + \beta \leq \alpha < 2$. Thus, (37) obtains and the theorem is proved. \square

Corollary 6. If $s_n^2 \rightarrow \infty$, $\gamma_n = O(1)$, and $\{Y, Y_n, n \geq 1\}$ are i.i.d. random variables with $E Y = 0$, the LIL holds for $\{\sigma_n Y_n\}$ and $\{-\sigma_n Y_n\}$ iff $E Y^2 < \infty$.

PROOF. The hypothesis implies (46), whence (47) ensures

$$\frac{\sigma_n^2 \log_2 s_n^2}{s_n^2} = \frac{\gamma_n \log_2 s_n^2}{n} \leq \frac{C \log n}{n} = o(1),$$

and so the conclusions follow from Theorems 5 and 6. \square

where $\alpha_1, \dots, \alpha_j$ are positive integers with $\sum_{i=1}^j \alpha_i = 2k$, $A_{\alpha_1, \dots, \alpha_j} = (\alpha_1 + \dots + \alpha_j)!/\alpha_1! \cdots \alpha_j!$, and i_1, \dots, i_j are distinct integers in $[1, n]$. Since $E X_{i_1}^{\alpha_1} \cdots X_{i_j}^{\alpha_j} = 1$ when $\alpha_1, \dots, \alpha_j$ are all even and zero otherwise,

$$E S_n^{2k} = \sum A_{2\beta_1, \dots, 2\beta_j} c_{i_1}^{2\beta_1} \cdots c_{i_j}^{2\beta_j},$$

β_1, \dots, β_j being positive integers with $\sum_{i=1}^j \beta_i = k$. Hence

$$\begin{aligned} E S_n^{2k} &= \sum \frac{A_{2\beta_1, \dots, 2\beta_j}}{A_{\beta_1, \dots, \beta_j}} \cdot A_{\beta_1, \dots, \beta_j} c_{i_1}^{2\beta_1} \cdots c_{i_j}^{2\beta_j} \\ &\leq B_{2k}^{2k} s_n^{2k}, \end{aligned}$$

where $s_n^2 = \sum_{i=1}^n c_i^2$ and

$$\begin{aligned} B_{2k}^{2k} &= \sup \frac{A_{2\beta_1, \dots, 2\beta_j}}{A_{\beta_1, \dots, \beta_j}} = \sup \frac{(2k)!}{(2\beta_1)! \cdots (2\beta_j)!} \frac{\beta_1! \cdots \beta_j!}{k!} \\ &\leq \sup \frac{2k(2k-1) \cdots (k+1)}{\prod_{i=1}^j 2\beta_i(2\beta_i-1) \cdots (\beta_i+1)} \leq \frac{2k(2k-1) \cdots (k+1)}{2^{\beta_1+\cdots+\beta_j}} \\ &= \frac{2k(2k-1) \cdots (k+1)}{2^k} \leq k^k. \end{aligned}$$

Thus, when $p = 2k$ the upper inequality of (1) holds with $\bar{B}_{2k} \leq k^{1/2}$. Since $\|S_n\|_p$ is increasing in p , $\|S_n\|_p \leq \|S_n\|_{2k} \leq \bar{B}_{2k} s_n$ for $p \leq 2k$, whence the upper inequality of (1) obtains with $\bar{B}_p \leq k^{1/2}$, where k is the smallest integer $\geq p/2$.

It suffices to establish the lower inequality for $0 < p < 2$ since $\|S_n\|_p \geq \|S_n\|_2 = s_n$ for $p \geq 2$. Recalling the logarithmic convexity of the \mathcal{L}_p norm established in Section 4.3 and choosing $r_1, r_2 > 0$ such that $r_1 + r_2 = 1$, $pr_1 + 4r_2 = 2$,

$$s_n^2 = \|S_n\|_2^2 \leq \|S_n\|_p^{pr_1} \|S_n\|_4^{4r_2} \leq \|S_n\|_p^{pr_1} (2^{1/2} s_n)^{4r_2},$$

whence

$$\|S_n\|_p^{pr_1} \geq 4^{-r_2} s_n^{2-4r_2} = 4^{-r_2} s_n^{pr_1},$$

$$\|S_n\|_p \geq 4^{-r_2/pr_1} s_n$$

Hence, the lower inequality holds for $0 < p < 2$ with $\bar{A}_p \geq 4^{-r_2/pr_1} = 2^{-(2-p)/p}$ and for $p \geq 2$ with $\bar{A}_p \geq 1$. \square

Corollary 1. Under the hypothesis of Theorem 1, if $s^2 = \sum_1^\infty c_j^2 < \infty$, then (i) $S_n = \sum c_i X_i \xrightarrow{a.s.} S$, (ii) $\|S_n\|_p \leq k^{1/2}s$, where k is the smallest integer $\geq p/2$, (iii) $E e^{tS^2} < \infty$ for all $t > 0$.

PROOF. Theorem 5.1.2 guarantees (i) while (ii) follows from Khintchine's inequality and Fatou's lemma. Apropos of (iii),

$$E e^{tS^2} = \sum_{j=0}^\infty \frac{t^j}{j!} E S^{2j} \leq \sum_{j=0}^\infty \frac{t^j}{j!} (J^{1/2} s)^{2j} = \sum_{j=0}^\infty \frac{t^j}{j!} (ts^2)^j < \sum_{j=0}^\infty (ts^2 e)^j$$

since $j^j/j! < \sum_{n=0}^\infty j^n/n! = e^j$. Thus, $E e^{tS^2} < \infty$ for $ts^2 e < 1$. Finally, since

$s_n \rightarrow s$, for any $t > 0$ the integer n may be chosen so that $2te(s^2 - s_n^2) < 1$. Then

$$\mathbb{E} e^{tS^2} = \mathbb{E} e^{t(S - S_n + S_n)^2} \leq \mathbb{E}[e^{2tS_n^2} \cdot e^{2t(S - S_n)^2}] < \infty$$

since S_n^2 is a bounded r.v. for fixed n . \square

Theorem 2 (Marcinkiewicz-Zygmund Inequality). *If $\{X_n, n \geq 1\}$ are independent r.v.s with $\mathbb{E} X_n = 0$, then for every $p \geq 1$ there exist positive constants A_p, B_p depending only upon p for which*

$$A_p \left\| \left(\sum_1^n X_j^2 \right)^{1/2} \right\|_p \leq \left\| \sum_1^n X_j \right\|_p \leq B_p \left\| \left(\sum_{j=1}^n X_j^2 \right)^{1/2} \right\|_p. \quad (2)$$

PROOF. Clearly (Exercise 4.2.4), $\sum_1^n X_j \in \mathcal{L}_p$ iff $X_j \in \mathcal{L}_p$, $1 \leq j \leq n$, iff $(\sum_1^n X_j^2)^{1/2} \in \mathcal{L}_p$, whence the latter may be supposed. Let $\{X_n^*, n \geq 1\}$ be the symmetrized $\{X_n, n \geq 1\}$, that is, $X_n^* = X_n - X'_n$, $n \geq 1$, where $\{X'_n, n \geq 1\}$ is independent of and identically distributed with $\{X_n, n \geq 1\}$. Moreover, let $\{V_i, n \geq 1\}$ constitute a sequence of i.i.d. r.v.s independent of $\{X_n, X'_n, n \geq 1\}$ with $\mathbb{P}\{V_1 = 1\} = \mathbb{P}\{V_1 = -1\} = \frac{1}{2}$. Since

$$\mathbb{E} \left\{ \sum_1^n V_i (X_i - X'_i) \mid V_1, \dots, V_n, X_1, \dots, X_n \right\} = \sum_1^n V_i X_i,$$

it follows that for any integer $n > 0$, $\{\sum_1^n V_i X_i, \sum_1^n V_i (X_i - X'_i)\}$ is a two-term martingale, leading to the first inequality of

$$\begin{aligned} \mathbb{E} \left| \sum_1^n V_i X_i \right|^p &\leq \mathbb{E} \left| \sum_1^n V_i X_i^* \right|^p \leq 2^{p-1} \mathbb{E} \left\{ \left| \sum_1^n V_i X_i \right|^p + \left| \sum_1^n V_i X_i' \right|^p \right\} \\ &= 2^p \mathbb{E} \left| \sum_1^n V_i X_i \right|^p. \end{aligned} \quad (3)$$

Since Khintchine's inequality (1) is applicable to $\mathbb{E}\{\|\sum_1^n V_i X_i\|^p \mid X_1, X_2, \dots\}$, necessarily

$$\bar{A}_p^p \mathbb{E} \left(\sum_1^n X_i^2 \right)^{p/2} \leq \mathbb{E} \left| \sum_1^n V_i X_i \right|^p \leq \bar{B}_p^p \mathbb{E} \left(\sum_1^n X_i^2 \right)^{p/2},$$

which, in conjunction with (3), yields

$$\bar{A}_p^p \mathbb{E} \left(\sum_1^n X_i^2 \right)^{p/2} \leq \mathbb{E} \left| \sum_1^n V_i X_i^* \right|^p \leq 2^p \bar{B}_p^p \mathbb{E} \left(\sum_1^n X_i^2 \right)^{p/2}. \quad (4)$$

However, in view of the symmetry of $\{X_j^*, 1 \leq j \leq n\}$

$$\mathbb{E} \left| \sum_1^n V_j X_j^* \right|^p = \mathbb{E} \left| \sum_1^n X_j^* \right|^p \leq 2^p \mathbb{E} \left| \sum_1^n X_j \right|^p$$

whence, recalling Lemma 10.1.1 (or repeating the earlier two-term martingale

Thus, to complete the proof of sufficiency for symmetric $\{X_n\}$ it suffices via Lemma 4.2.3 to prove that $E(\sup c_n |S'_n|)^r < \infty$, and this will be done for all $r > 0$, supposing only that $E X = 0$, $E X^2 = 1$. To this end, set $n_k = [3^k]$, whence by Levy's inequality and symmetry

$$\begin{aligned} P\left\{\sup_{n>c^*} c_n |S'_n| \geq u\right\} &\leq \sum_{k=1}^{\infty} P\left\{c_{n_k} \sup_{n_k \leq n < n_{k+1}} |S'_n| \geq u\right\} \\ &\leq 4 \sum_{k=1}^{\infty} P\{c_{n_k} S'_{n_{k+1}} \geq u\}. \end{aligned} \quad (15)$$

It follows from Example 7.4.9 and (1) of Lemma 10.2.1, noting $g(1) < 1$, that for $0 < tb_n \leq 1$

$$P\{S'_n \geq x\} \leq e^{-tx} \prod_{j=1}^n E e^{tX_j} \leq \exp\left(-tx + t^2 \sum_1^n E X_j'^2\right) \leq e^{-tx + mt^2}.$$

Hence, taking $t = ((\log_2 n_{k+1})/n_{k+1})^{1/2}$, $c = b_{n_{k+1}}$,

$$\begin{aligned} \log P\{S'_{n_{k+1}} \geq c_{n_k}^{-1} u\} &\leq -u(n_k \log_2 n_k)^{1/2} \left(\frac{\log_2 n_{k+1}}{n_{k+1}}\right)^{1/2} + \log_2 n_{k+1} \\ &\leq -(\alpha u - 1) \log_2 n_{k+1} \end{aligned}$$

for some positive constant α . Therefore, choosing u_0 such that $\alpha u_0 > 5$, it follows via (15) that

$$\begin{aligned} \int_{u_0}^{\infty} u'^{-1} P\{\sup c_n |S'_n| \geq u\} du &\leq 4 \sum_{k=1}^{\infty} \int_{u_0}^{\infty} u'^{-1} \exp\{-(\alpha u - 1) \log(k+1)\} du \\ &< C \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty, \end{aligned}$$

so that $E(\sup c_n |S'_n|)^r < \infty$ by Corollary 6.2.2.

In the case of general $\{X_n\}$, if $X_n^* = X_n - X_n'$, $n \geq 1$, are the symmetrized r.v.s,

$$\begin{aligned} E \sup c_n' |S_n'|^r &= E \sup c_n' |E\{S_n^* | X_1, X_2, \dots\}|^r \\ &\leq E \sup c_n' E\{|S_n^*|^r | X_1, X_2, \dots\} \\ &\leq E \sup c_n' |S_n^*|^r < \infty. \end{aligned}$$

For $r > 2$ the converse follows from the necessity part of Theorem 3. When $r = 2$, Theorem 3 merely yields $E X_1^2 < \infty$. However,

$$E \sup_{n \geq c^*} \frac{X_n^2}{n \log_2 n} \leq 2 E \sup_{n \geq c^*} \frac{S_n^2 + S_{n-1}^2}{n \log_2 n} < \infty,$$

10.4 Maxima of Random Walks

In Section 5.2 moments of i.i.d. r.v.s $\{X_n, n \geq 1\}$ and hence behavior of the corresponding random walk $\{S_n = \sum_i^n X_i, n \geq 1\}$ were linked to convergence of series involving $|S_n|$ and $S_n^* = \max_{1 \leq i \leq n} |S_i|$. Here, analogous results for the one-sided case involving explicit bounds for the series will be obtained. For any positive v , set

$$\begin{aligned} X_0 &= 0, & \bar{X}_v &= \max_{0 \leq j \leq v} X_j, & S_v &= \sum_{i=0}^{[v]} X_i, & \bar{S}_v &= \max_{0 \leq j \leq v} S_j, \\ X_v^* &= \max_{0 \leq j \leq v} |X_j|, & S_v^* &= \max_{0 \leq j \leq v} |S_j|. \end{aligned} \quad (1)$$

Theorem 1. *If $\{X, X_n, n \geq 1\}$ are i.i.d. r.v.s with $E X = 0$ and p, α, γ are constants satisfying $1 \leq \gamma \leq 2, \alpha\gamma > 1, \alpha p > 1$, there exists a constant $C = C_{p, \alpha, \gamma} \in (0, \infty)$ such that*

$$\sum_{n=1}^{\infty} n^{\alpha p - 2} P\{\bar{S}_n \geq n^\alpha\} \leq C [E(X^+)^p + (E|X|^\gamma)^{(\alpha p - 1)/(\alpha\gamma - 1)}]. \quad (2)$$

PROOF. Suppose without loss of generality that $E|X| > 0, E(X^+)^p < \infty$, and $E|X|^\gamma < \infty$. Clearly, for any $k > 0$, by Lemma 5.3.5,

$$P\{\bar{S}_n \geq n^\alpha\} \leq n P\{X > n^\alpha/2k\} + P^k\{\bar{S}_n \geq n^\alpha/2k\}. \quad (3)$$

Now via Theorem 7.4.8 and the Marcinkiewicz–Zygmund inequality

$$\begin{aligned} P\left\{\bar{S}_n \geq \frac{n^\alpha}{2k}\right\} &\leq P\left\{S_n^* \geq \frac{n^\alpha}{2k}\right\} \leq C_\gamma \left(\frac{n^\alpha}{2k}\right)^{-\gamma} E|S_n|^\gamma \\ &\leq C_\gamma \left(\frac{n^\alpha}{2k}\right)^{-\gamma} E\left(\sum_1^n X_j^2\right)^{\gamma/2} \leq C_{\gamma, k} n^{1-\alpha\gamma} E|X|^\gamma \end{aligned} \quad (4)$$

Set $k = 1 + [(xp - 1)/(\alpha\gamma - 1)]$. Then $\lambda = k - (\alpha p - 2)/(\alpha\gamma - 1) > 1/(\alpha\gamma - 1)$, and if $E|X|^\gamma \geq 1$, from (4)

$$\begin{aligned} I &\equiv \sum_1^{\infty} n^{\alpha p - 2} P^k\left\{\bar{S}_n \geq \frac{n^\alpha}{2k}\right\} = \sum_{n^{\alpha\gamma-1} > E|X|^\gamma} + \sum_{n^{\alpha\gamma-1} \leq E|X|^\gamma} \\ &\leq C_1 [\sum n^{\alpha p - 2 + k(1 - \alpha\gamma)} E^k |X|^\gamma + \sum n^{\alpha p - 2}] \\ &= C_2 [(E|X|^\gamma)^{(1 - \lambda(\alpha\gamma - 1))/(\alpha\gamma - 1) + k} + (E|X|^\gamma)^{(\alpha p - 1)/(\alpha\gamma - 1)}] \\ &= C(E|X|^\gamma)^{(\alpha p - 1)/(\alpha\gamma - 1)}. \end{aligned} \quad (5)$$

On the other hand, if $E|X|^\gamma < 1$, again via (4)

$$I \leq C_1 \sum_{n=1}^{\infty} n^{-\lambda(\alpha\gamma - 1)} E^k |X|^\gamma \leq C_2 E^k |X|^\gamma \leq C(E|X|^\gamma)^{(\alpha p - 1)/(\alpha\gamma - 1)}. \quad (6)$$

Hence, from (2), (5), and (6)

$$\begin{aligned} \sum_{n=1}^{\infty} n^{xp-2} P\{\bar{S}_n \geq n^x\} &\leq \sum_{k=1}^{\infty} \left[n^{xp-1} P\left\{X > \frac{n^x}{2k}\right\} + n^{xp-2} P^k \left\{\bar{S}_n \geq \frac{n^x}{2k}\right\} \right] \\ &\leq C [E(X^+)^p + (E|X|^\gamma)^{(xp-1)/(x\gamma-1)}]. \end{aligned} \quad \square$$

Taking $\alpha = 1, \gamma = p$ in Theorem 1 yields

Corollary 1. If $\{X, X_n, n \geq 1\}$ are i.i.d. r.v.s with $E X = 0$ and $1 < p \leq 2$, then for some constant $C_p \in (0, \infty)$

$$\begin{aligned} \sum_{n=1}^{\infty} n^{p-2} P\{\bar{S}_n \geq n\} &\leq C_p E|X|^p, \\ \sum_{n=1}^{\infty} n^{p-2} P\{S_n^* \geq n\} &\leq 2C_p E|X|^p. \end{aligned} \quad (7)$$

Corollary 2 (Hsu–Robbins). If $\{X, X_n, n \geq 1\}$ are i.i.d. r.v.s, then

$$\sum_{n=1}^{\infty} P\{|S_n| \geq n\varepsilon\} < \infty, \quad \varepsilon > 0, \quad (8)$$

iff $E X = 0, E X^2 < \infty$.

PROOF. (8) follows immediately from Corollary 1 applied to X/ε while the converse is the special case $\alpha = 1, p = 2$ of Theorem 6.4.5. \square

EXAMPLE 1. Strong Laws for Arrays. Let $\{X_{ni}, 1 \leq i \leq n, n \geq 1\}$ be an array of random variables that are identically distributed and *rowwise independent*, i.e., $\{X_{n1}, \dots, X_{nn}\}$ are independent for $n \geq 2$. If $E|X_{11}|^q < \infty, 0 < q < 4$, and $E X_{11} = 0$ whenever $1 \leq q < 4$, then $n^{-2/q} \sum_{i=1}^n X_{ni} \xrightarrow{a.s.} 0$.

PROOF. The case $0 < q < 2$ is covered in Exercise 5.2.17. Set $S_{nn} = \sum_{i=1}^n X_{ni}$, $n \geq 1$. When $1 \leq q < 4$, replacing p by q and choosing $\gamma = 2, \alpha = 2/q$ in Theorem 1, it follows that

$$\sum_{n=1}^{\infty} P\{S_{nn} > \varepsilon n^{2/q}\} < \infty, \quad \varepsilon > 0. \quad (9)$$

Since X_i may be replaced by $-X_i$ in Theorem 1 (when $E|X|^p < \infty$), (9) holds with S_{nn} replaced by $-S_{nn}$, and so

$$\sum_{n=1}^{\infty} P\{|S_{nn}| > \varepsilon n^{2/q}\} < \infty, \quad \varepsilon > 0, \quad (10)$$

whence the Borel–Cantelli lemma implies that $n^{-2/q} S_{nn} \xrightarrow{a.s.} 0$. \square

For any r.v.s $\{Y_n, n \geq 1\}$ and nonnegative v , set $Y_0 = 0$,

$$Y_v = Y_{[v]}, \quad \bar{Y}_v = \max_{0 \leq j \leq v} Y_j, \quad (11)$$

and for $p > 0, \alpha > 0$ define (as usual, $\sup \emptyset = 0$)

$$\begin{aligned} M(\varepsilon) &= M(\varepsilon, \alpha) = \sup_{n \geq 0} (Y_n - \varepsilon n^\alpha) = \sup_{n \geq 0} (\bar{Y}_n - \varepsilon n^\alpha), \\ \bar{L}(\varepsilon) &= \bar{L}(\varepsilon, \alpha) = \sup\{n \geq 1 : \bar{Y}_n \geq \varepsilon n^\alpha\}, \\ I(\varepsilon) &= I(\varepsilon, p, \alpha) = \int_0^\infty v^{\alpha p - 2} P\left\{\sup_{j \geq v} j^{-\alpha} \bar{Y}_j \geq \varepsilon\right\} dv, \\ J(\varepsilon) &= J(\varepsilon, p, \alpha) = \int_0^\infty v^{\alpha p - 2} P\{\bar{Y}_v \geq \varepsilon v^\alpha\} dv. \end{aligned} \quad (12)$$

Clearly, $J(\varepsilon) \leq I(\varepsilon)$.

Lemma 1. For $\alpha p > 1$,

$$I(2\varepsilon) \leq \frac{1}{\alpha p - 1} \varepsilon^{(1-\alpha p)/\alpha} E[M(\varepsilon)]^{(\alpha p - 1)/\alpha} \leq (2^{(\alpha p - 1)/\alpha} - 1)^{-1} J\left(\frac{\varepsilon}{2}\right), \quad (13)$$

$$J(\varepsilon) \leq \max(2^{\alpha p - 2}, 1) \sum_{n=1}^\infty n^{\alpha p - 2} P\{\bar{Y}_n \geq \varepsilon n^\alpha\}, \quad (14)$$

$$E[\bar{L}(\varepsilon)]^{\alpha p - 1} \leq (\alpha p - 1) I(\varepsilon). \quad (15)$$

PROOF. (14) is trivial. For (13), set $\beta = (\alpha p - 1)/\alpha$. Then

$$\begin{aligned} \varepsilon^{-\beta} E[M(\varepsilon)]^\beta &= \int_0^\infty P\left\{\sup_{n \geq 1} (Y_n - \varepsilon n^\alpha) \geq \varepsilon v^{1/\beta}\right\} dv \\ &\leq \sum_{k=1}^\infty \int_0^\infty P\left\{\max_{\{(2^{k-1}-1)v^{1/\beta} < u^\alpha \leq 2^k v^{1/\beta}\}} (Y_u - \varepsilon u^\alpha) \geq \varepsilon v^{1/\beta}\right\} dv \\ &\leq \sum_{k=1}^\infty \int_0^\infty P\{\bar{Y}_{(2^k v^{1/\beta})^{1/\alpha}} \geq 2^{k-1} \varepsilon v^{1/\beta}\} dv \\ &= \sum_{k=1}^\infty 2^{-k\beta} (\alpha p - 1) \int_0^\infty u^{\alpha p - 2} P\{\bar{Y}_u \geq \varepsilon u^\alpha/2\} du \\ &= (\alpha p - 1)(2^\beta - 1)^{-1} J\left(\frac{\varepsilon}{2}\right), \end{aligned}$$

establishing the second half of (13). Moreover,

$$\begin{aligned} I(2\varepsilon) &= \int_0^\infty v^{\alpha p - 2} P\left\{\sup_{j \geq v} j^{-\alpha} \bar{Y}_j \geq 2\varepsilon\right\} dv \\ &\leq \int_0^\infty v^{\alpha p - 2} P\left\{\sup_{n \geq 0} (\bar{Y}_n - \varepsilon n^\alpha) \geq \varepsilon v^\alpha\right\} dv \end{aligned}$$

$$= \int_0^\infty v^{\alpha p - 2} P\left\{\left(\frac{M(\varepsilon)}{\varepsilon}\right)^{1/\alpha} \geq v\right\} dv = \frac{1}{\alpha p - 1} E\left[\frac{M(\varepsilon)}{\varepsilon}\right]^{(\alpha p - 1)/\alpha},$$

which is the first half of (13). Apropos of (15),

$$\begin{aligned} P\{\bar{L}(\varepsilon) \geq v\} &= P\{\bar{Y}_n \geq \varepsilon n^\alpha \text{ for some } n \geq v\} \\ &\leq P\left\{\sup_{n \geq v} n^{-\alpha} \bar{Y}_n \geq \varepsilon\right\}, \end{aligned}$$

so that

$$E[\bar{L}(\varepsilon)]^{\alpha p - 1} = (\alpha p - 1) \int_0^\infty v^{\alpha p - 2} P\{\bar{L}(\varepsilon) \geq v\} dv \leq (\alpha p - 1) J(\varepsilon). \quad \square$$

The combination of Lemma 1 and Theorem 1 yields

Theorem 2. If $\{X_i, X_n, n \geq 1\}$ are i.i.d. r.v.s with $E X = 0$, $S_n = \sum_1^n X_i$, $S_0 = 0$, and $1 \leq \gamma \leq 2$, $\alpha\gamma > 1$, $\alpha p > 1$, then for some constant $C = C_{p, \alpha, \gamma} \in (0, \infty)$

$$E\left[\sup_{n \geq 0} (S_n - n^\alpha)\right]^{(\alpha p - 1)/\alpha} \leq C \left[E(X^+)^p + (E|X|^\gamma)^{(\alpha p - 1)/(\alpha\gamma - 1)} \right]. \quad (16)$$

Lemma 2. Let $\{X_i, X_n, n \geq 1\}$ be i.i.d. r.v.s with $E|X|^{1/\alpha} < \infty$, where $\alpha > \frac{1}{2}$ and $E X = 0$ if $\alpha \leq 1$. If $S_n = \sum_1^n X_i$ and

$$L(\varepsilon) = L(\varepsilon, \alpha) = \sup\{n \geq 1 : S_n \geq \varepsilon n^\alpha\}, \quad (17)$$

$$L_1(\varepsilon) = L_1(\varepsilon, \alpha) = \sup\{n \geq 1 : X_n \geq \varepsilon n^\alpha\}, \quad (18)$$

where $\varepsilon > 0$ and $\sup \emptyset = 0$, then $E L^\gamma(\varepsilon) < \infty$ implies $E L_1^\gamma(2\varepsilon) < \infty$, $\gamma > 0$.

PROOF. Set $A_j = \{X_j \geq 2\varepsilon j^\alpha\}$, $B_j = \{|S_{j-1}| \leq \varepsilon j^\alpha\}$. Now $n^{-\alpha} S_n \xrightarrow{a.s.} 0$ by Theorem 5.2.2, and so $P\{B_n\} \rightarrow 1$ as $n \rightarrow \infty$. Since for $n \geq 1$ the classes $\{B_n\}$ and $\{A_n, A_n A_{n+1}^c, \dots\}$ are independent, by Lemma 3.3.3 for $n \geq n_0$

$$\begin{aligned} P\{L(\varepsilon) \geq n\} &\geq P\left\{\bigcup_{j=n}^\infty A_j B_j\right\} \geq P\left\{\bigcup_{j=n}^\infty A_j\right\}_{j \geq n} \inf P\{B_j\} \\ &\geq \frac{1}{2} P\{L_1(2\varepsilon) \geq n\}, \end{aligned}$$

and the lemma follows. \square

Lemma 3. Let $\{X_i, X_n, n \geq 1\}$ be i.i.d., $\alpha > 0$, $\alpha p > 1$, $\varepsilon > 0$, and

$$J_1(\varepsilon) = \int_0^\infty v^{\alpha p - 2} P\{\bar{X}_v \geq \varepsilon v^\alpha\} dv. \quad (19)$$

Then, if $L_1(\varepsilon)$ is as in (18),

$$(\alpha p - 1) J_1(\varepsilon) \leq E[L_1(2^{-\alpha}\varepsilon)]^{\alpha p - 1}, \quad \alpha p e^p J_1(\varepsilon) \leq E(X^+)^p, \quad (20)$$

$$J_1(1) < \infty \Rightarrow E(X^+)^p < \infty \quad \text{and} \quad \frac{E(X^+)^p - 1}{2\alpha p[1 + E(X^+)^{1/p}]} \leq J_1(1). \quad (21)$$

PROOF. The second inequality of (20) follows immediately from

$$P\{\bar{X}_v \geq \varepsilon v^x\} \leq v P\{X \geq \varepsilon v^x\}.$$

As for the first,

$$\begin{aligned} (\alpha p - 1)J_1(\varepsilon) &= (\alpha p - 1) \int_0^\infty v^{\alpha p - 2} P\left\{ \max_{v < n \leq 2v-1} X_n \geq \varepsilon v^x \right\} dv \\ &\leq (\alpha p - 1) \int_0^\infty v^{\alpha p - 2} P\left\{ \bigcup_{n=v}^\infty \left[X_n \geq \varepsilon \left(\frac{n}{2}\right)^x \right] \right\} dv \\ &= (\alpha p - 1) \int_0^\infty v^{\alpha p - 2} P\{L_1(2^{-x}\varepsilon) \geq v\} dv = E[L_1(2^{-x}\varepsilon)]^{\alpha p - 1}. \end{aligned}$$

To prove (21), note that for $v \geq 1$

$$\begin{aligned} P\{\bar{X}_v \geq v^x\} &= \sum_{j=1}^{[v]} P\{X_j \geq v^x\} P^{j-1}\{X < v^x\} \\ &\geq [v] P\{X < v^x\} P\{X \geq v^x\} \\ &= [v] P\{X \geq v^x\} [1 - P\{\bar{X}_v \geq v^x\}], \end{aligned}$$

whence for $v \geq 1$

$$\begin{aligned} [v] P\{X \geq v^x\} &\leq (1 + [v] P\{X \geq v^x\}) P\{\bar{X}_v \geq v^x\} \\ &\leq [1 + E(X^+)^{1/x}] P\{\bar{X}_v \geq v^x\} \end{aligned}$$

by the Markov inequality. Hence

$$\begin{aligned} E(X^+)^p &= \alpha p \int_0^\infty v^{\alpha p - 1} P\{X \geq v^x\} dv \\ &\leq 1 + 2\alpha p \int_1^\infty v^{\alpha p - 2} P\{\bar{X}_v \geq v^x\} dv \cdot [1 + E(X^+)^{1/x}]. \quad (22) \end{aligned}$$

Finally, if $X' = \max(X, c)$, $c > 0$, then (22) holds with X^+ replaced by X' . If $E(X^+)^p = \infty$, then, since $\alpha p > 1$, $E(X')^{1/x} = o(E(X^+)^p)$ as $c \rightarrow \infty$, implying $J_1(1) = \infty$. Thus, (21) obtains. \square

Theorem 3. Let $\{X, X_n, n \geq 1\}$ be i.i.d. r.v.s, $S_n = \sum_i^n X_i$, $S_0 = X_0 = 0$, and $p > 1/\alpha > 0$, $\varepsilon > 0$. Set

$$M(\varepsilon) = \sup_{n \geq 0} (S_n - \varepsilon n^x), \quad M_1(\varepsilon) = \sup_{n \geq 0} (X_n - \varepsilon n^x),$$

$$L(\varepsilon) = \sup\{n \geq 1 : S_n \geq \varepsilon n^x\}, \quad L_1(\varepsilon) = \sup\{n \geq 1 : X_n \geq \varepsilon n^x\}.$$

Then, (i)

$$\begin{aligned} \mathbb{E}(X^+)^p < \infty &\text{ iff } J_1(1) = \int_0^\infty v^{zp-2} \mathbb{P}\{\bar{X}_v \geq v^z\} dv < \infty \\ &\text{ iff } \mathbb{E}[M_1(\varepsilon)]^{(zp-1)/z} < \infty; \quad \varepsilon > 0, \\ &\text{ iff } \mathbb{E}[L_1(\varepsilon)]^{zp-1} < \infty, \quad \varepsilon > 0. \end{aligned}$$

(ii) Suppose $\mathbb{E}(X^+)^p < \infty$ for $p \geq 1$, $\mathbb{E} X = 0$, and $\mathbb{E}|X|^\gamma < \infty$ for some $\gamma \in (1/\alpha, 2]$ when $\frac{1}{2} < \alpha < 1$. Then for any $\alpha > \frac{1}{2}$ and all $\varepsilon > 0$

$$\mathbb{E}[M(\varepsilon)]^{(zp-1)/z} < \infty, \quad \mathbb{E}[L(\varepsilon)]^{zp-1} < \infty, \quad (23)$$

and $J(\varepsilon) \leq I(\varepsilon) < \infty$, where the latter are as in (12) but with S_v replacing \bar{Y}_v .

(iii) Let $\alpha > \frac{1}{2}$, $\mathbb{E}|X|^{1/z} < \infty$, and $\mathbb{E} X = 0$ if $\alpha \leq 1$. If either of the conditions of (23) holds or $I(\varepsilon) < \infty$ or $J(\varepsilon) < \infty$, then $\mathbb{E}(X^+)^p < \infty$.

PROOF. By Lemma 3, $\mathbb{E}(X^+)^p < \infty$ iff $J_1(1) < \infty$ iff $J_1(\varepsilon) < \infty$ for all $\varepsilon > 0$. Then Lemma 1 ensures $\mathbb{E}[M_1(\varepsilon)]^{(zp-1)/z} < \infty$, $I(2\varepsilon) < \infty$, $\varepsilon > 0$, and hence $\mathbb{E}[L_1(\varepsilon)]^{zp-1} < \infty$ since $L_1(\varepsilon) \leq \bar{L}_1(\varepsilon)$. Conversely, by Lemma 1

$$\mathbb{E}[M_1(\varepsilon)]^{(zp-1)/z} < \infty$$

implies $J_1(\varepsilon) < \infty$, $\varepsilon > 0$, and hence $J_1(\varepsilon) < \infty$, $\varepsilon > 0$. Moreover, if $\mathbb{E}[L_1(\varepsilon)]^{zp-1} < \infty$, $\varepsilon > 0$, then by Lemma 3, $J_1(\varepsilon) < \infty$, $\varepsilon > 0$.

Apropos of (ii), if $\frac{1}{2} < \alpha < 1$, the first half of (23) follows from Theorem 2. Then by Lemma 1, $J(\varepsilon) \leq I(\varepsilon) < \infty$ and $\mathbb{E}[L(\varepsilon)]^{zp-1} \leq \mathbb{E}[\bar{L}(\varepsilon)]^{zp-1} < \infty$. If, rather, $\alpha \geq 1$, define $X'_n = X_n I_{\{X_n \geq -C\}}$, $C > 0$. Then $S'_n = \sum_1^n X'_j \geq S_n$. Since $\mathbb{E}(X^+)^p < \infty$ for some $p \geq 1$ ($p > 1$ if $\alpha = 1$), necessarily $\gamma \equiv \min(p, 2) \in [1, 2]$ and $\mathbb{E}|X'_1|^\gamma < \infty$. Hence, by Theorem 2

$$\begin{aligned} &\mathbb{E} \left[\sup_{n \geq 0} (S_n - n^z(\varepsilon + \mathbb{E} X'_1)) \right]^{(zp-1)/z} \\ &\leq \mathbb{E} \left[\sup_{n \geq 0} \left(\sum_{j=1}^n (X'_j - \mathbb{E} X'_1) - \varepsilon n^z \right) \right]^{(zp-1)/z} < \infty \end{aligned}$$

by Theorem 1, and since $\mathbb{E} X'_1 = o(1)$ as $C \rightarrow \infty$, the first half of (23) is established. The remainder of (23) and (ii) follow from $L'(\varepsilon) \geq L(\varepsilon)$, $I'(\varepsilon) \geq I(\varepsilon)$.

To prove (iii), note that by Lemma 1

$$J(\varepsilon) < \infty \Rightarrow \mathbb{E}[M(\varepsilon)]^{(zp-1)/z} < \infty \Rightarrow I(\varepsilon) < \infty \Rightarrow \mathbb{E}[\bar{L}(\varepsilon)]^{zp-1} < \infty,$$

whence $\mathbb{E}[L(\varepsilon)]^{zp-1} < \infty$. Then by Lemma 2, $\mathbb{E}[L_1(\varepsilon)]^{zp-1} < \infty$, implying $\mathbb{E}(X^+)^p < \infty$ by part (i). \square

Corollary 3 (Kiefer–Wolfowitz). *If $\{X, X_n, n \geq 1\}$ are i.i.d. with $\mathbb{E} X < 0$ and $S_n = \sum_1^n X_i$, $S_0 = 0$, $p > 1$, then $\mathbb{E}(\sup_{n \geq 0} S_n)^{p-1} < \infty$ iff $\mathbb{E}(X^+)^p < \infty$.*

PROOF. Apply Theorem 3 with $\alpha = 1$ and $\varepsilon = -E X$ to $\{X_n - E X, n \geq 1\}$.

Corollary 4. Let $\{X, X_n, n \geq 1\}$ be i.i.d. r.v.s with $E X = 0$, $S_n = \sum_1^n X_i$, and define $\hat{L}(\varepsilon) = \sup\{n \geq 1 : |S_n| \geq n\varepsilon\}$, $\varepsilon > 0$. Then for $p > 1$, $E[\hat{L}(\varepsilon)]^{p-1} < \infty$ for all $\varepsilon > 0$ iff $E|X|^p < \infty$.

PROOF. Clearly, $\hat{L}(\varepsilon) = \max[L^+(\varepsilon), L^-(\varepsilon)] \leq L^+(\varepsilon) + L^-(\varepsilon)$, where $L^-(\varepsilon) = \sup\{n \geq 1 : S_n \leq -n\varepsilon\}$ and $L^+(\varepsilon) = \sup\{n \geq 1 : S_n \geq n\varepsilon\}$. \square

Corollary 5. If $\{X, X_n, n \geq 1\}$ are i.i.d. with $E X = \mu > 0$, $E(X^-)^2 < \infty$ and $N_0 = \sup\{n \geq 0 : S_n \leq 0\}$, where $S_n = \sum_1^n X_i$, then $E N_0 < \infty$ and

$$\sum_{n=1}^{\infty} P\left\{\inf_{j \geq n} S_j \leq 0\right\} < \infty.$$

PROOF. Set $Y_n = \mu - X_n$, whence

$$\begin{aligned} \sum_{n=1}^{\infty} P\left\{\inf_{j \geq n} S_j \leq 0\right\} &= \sum_{n=1}^{\infty} P\left\{\sup_{j \geq n} \sum_{i=1}^j (Y_i - \mu) \geq 0\right\} \\ &\leq \sum_{n=1}^{\infty} P\left\{\sup_{j \geq n} \sum_{i=1}^j \left(Y_i - \frac{\mu}{2}\right) \geq \frac{n\mu}{2}\right\} \\ &\leq \sum_{n=1}^{\infty} P\left\{\sup_{j \geq 1} \sum_{i=1}^j \left(Y_i - \frac{\mu}{2}\right) > \frac{n\mu}{2}\right\} < \infty \end{aligned}$$

by Corollary 3. Since $\{N_0 \geq n\} = \bigcup_{j=n}^{\infty} \{S_j \leq 0\} \subset \{\inf_{j \geq n} S_j \leq 0\}$, necessarily $E N_0 < \infty$. \square

For any r.v.s Y, Z let the relation $Y \sim Z$ signify that Y and Z have identical distributions.

Theorem 4. Let $\{X, X_n, n \geq 1\}$ be i.i.d. r.v.s with $S_n = \sum_1^n X_i$, $n \geq 1$, $S_0 = 0$, $\bar{S}_n = \max_{0 \leq j \leq n} S_j$, $\bar{S} = \sup_{n \geq 0} S_n$. Then

- i. $\bar{S}_n \sim (\bar{S}_{n-1} + X)^+$, $n \geq 1$,
- ii. if $S_n \xrightarrow{a.s.} -\infty$, then $\bar{S} \sim (\bar{S} + X)^+$,
- iii. If $E X \in (-\infty, 0)$ and $E X^2 < \infty$, then $E \bar{S} = [\sigma_X^2 - \sigma_{(\bar{S}+X)^+}^2]/(-2 E X)$.

PROOF. If $W_0 = 0$, $W_n = (W_{n-1} + X_n)^+$, $n \geq 1$, then

$$W_n = \max[S_n, S_n - S_1, \dots, S_n - S_{n-1}, 0], \quad n \geq 1. \quad (24)$$

Clearly, (24) obtains for $n = 1$ and, assuming its validity for $n - 1 \geq 1$,

$$\begin{aligned} W_n &= (\max[S_{n-1}, S_{n-1} - S_1, \dots, S_{n-1} - S_{n-2}, 0] + X_n)^+ \\ &= (\max[S_n, S_n - S_1, \dots, S_n - S_{n-1}])^+ \\ &= \max[S_n, S_n - S_1, \dots, S_n - S_{n-1}, 0]. \end{aligned}$$

whence (24) holds for $n \geq 1$. Thus,

$$W_n \sim \max[S_n, S_{n-1}, \dots, S_1, 0] = \bar{S}_n, \quad n \geq 1,$$

and (i) follows.

Next, if $S_n \xrightarrow{\text{a.s.}} -\infty$, then $\bar{S}_n \xrightarrow{\text{a.s.}} \bar{S}$ (finite), whence (ii) is an immediate consequence of (i).

Under the hypothesis of (iii), $E \bar{S} < \infty$ by Corollary 3. Moreover, assuming temporarily that $E(X^+)^3 < \infty$, this same corollary ensures $E \bar{S}^2 < \infty$. By (ii),

$$E \bar{S} = E(\bar{S} + X)^+ = E(\bar{S} + X) + E(\bar{S} + X)^-,$$

implying

$$E X = -E(\bar{S} + X)^-. \quad (25)$$

Similarly,

$$\begin{aligned} E[(\bar{S} + X)^-]^2 - E X^2 &= E[(\bar{S} + X)^2 - 2(\bar{S} + X)(\bar{S} + X)^+ \\ &\quad + (\bar{S} + X)^{+2}] - E X^2 \\ &= E \bar{S}^2 + 2 E \bar{S} X - E[(\bar{S} + X)^+]^2 \\ &= 2 E \bar{S} E X, \end{aligned}$$

and combining this with (25) yields (iii) when $E(X^+)^3 < \infty$. In general, set $X'_n = \min(X_n, C)$ and define \bar{S}' , X' analogously via X'_n , whence

$$-2 E \bar{S}' E X' = \sigma_{X'}^2 - \sigma_{(\bar{S}' + X')^-}^2. \quad (26)$$

Since $(\bar{S}' + X')^- \leq (X')^- = X^-$, $\bar{S}' \leq \bar{S}$, and $\lim_{C \rightarrow \infty} \bar{S}' = \bar{S}$, by Lebesgue's dominated convergence theorem $E \bar{S}' \rightarrow E \bar{S}$ and $\sigma_{(\bar{S}' + X')^-}^2 \rightarrow \sigma_{(\bar{S} + X)^-}^2$ and (iii) follows from (26). \square

Corollary 6. If $\{X, X_n, n \geq 1\}$ are i.i.d. r.v.s with $E X = 0$, $\sigma^2 = E X^2 < \infty$, $S_n = \sum_1^n X_i$, $S_0 = 0$, and

$$M = M(\varepsilon) = \sup_{n \geq 0} (S_n - n\varepsilon) = \sup_{n \geq 0} \sum_{i=1}^n (X_i - \varepsilon), \quad (27)$$

where $\varepsilon > 0$, then

$$\lim_{\varepsilon \rightarrow 0} \varepsilon E M(\varepsilon) = \frac{\sigma^2}{2}. \quad (28)$$

PROOF. To avoid triviality, suppose $\sigma^2 > 0$. By Theorem 4(iii)

$$2\varepsilon E M(\varepsilon) = \sigma_{X-\varepsilon}^2 - \sigma_{(M+X-\varepsilon)^-}^2. \quad (29)$$

Since $(M + X - \varepsilon)^- \leq X^- + \varepsilon$ and

$$\lim_{\varepsilon \rightarrow 0} M(\varepsilon) \geq \lim_{\varepsilon \rightarrow 0} \sup_{0 < j \leq n} (S_j - j\varepsilon) = \sup_{0 \leq j < n} S_j \xrightarrow{n \rightarrow \infty} \infty,$$

$\sigma_{(M+X-\varepsilon)^-}^2 \rightarrow 0$ by dominated convergence and (28) follows from (29). \square

$$\begin{aligned}
 &= \sum_{j=1}^{\infty} P\{U^{(1)} = j\} E W_{U_j} I_{[U_j < \infty]} \\
 &= P\{U_+ < \infty\} E W_{U_+} I_{[U_+ < \infty]}
 \end{aligned}$$

and similarly for $n > 2$, it follows from (32) that

$$\begin{aligned}
 E M(\varepsilon) &= \sum_{n=1}^{\infty} P^{n-1}\{U_+ < \infty\} E W_{U_n} I_{[U_n < \infty]} \\
 &= \frac{1}{P\{U_+ = \infty\}} E W_{U_+} I_{[U_+ < \infty]}.
 \end{aligned}$$

To prove (ii), note that via Wald's equation, the initial notation of (31), and part (i),

$$\begin{aligned}
 E(-W_{U_-}) \cdot E W_{U_+} I_{[U_+ < \infty]} &= \varepsilon E U_- \cdot E W_{U_+} I_{[U_+ < \infty]} \\
 &= \varepsilon E M(\varepsilon) \rightarrow \frac{\sigma^2}{2}
 \end{aligned}$$

as $\varepsilon \rightarrow 0$ by Corollary 6. Hence, from Fatou's lemma

$$\begin{aligned}
 E(-S_{T_-}) \cdot E S_{T_+} &\leq \liminf_{\varepsilon \rightarrow 0} E(-W_{U_-}) \cdot \liminf_{\varepsilon \rightarrow 0} E W_{U_+} I_{[U_+ < \infty]} \\
 &\leq \liminf_{\varepsilon \rightarrow 0} E(-W_{U_-}) E W_{U_+} I_{[U_+ < \infty]} = \frac{\sigma^2}{2}.
 \end{aligned}$$

Since both $E(-S_{T_-})$ and $E S_{T_+}$ are positive, each is therefore finite and (ii) is proved.

Finally, as already observed, $E U_- = 1/P\{U_+ = \infty\} < \infty$, whence (iii) follows from (i) and Corollary 3.

References

- L. E. Baum and M. Katz, "Convergence rates in the law of large numbers," *Trans. Amer. Math. Soc.* **120** (1965), 108–123.
- H. D. Brunk, "The strong law of large numbers," *Duke Math. J.* **15** (1948), 181–195.
- Y. S. Chow, "A martingale inequality and the law of large numbers," *Proc. Amer. Math. Soc.* **11** (1960), 107–111.
- Y. S. Chow, "On a strong law of large numbers for martingales," *Ann. Math. Stat.* **38** (1967), 610.
- Y. S. Chow, "Delayed sums and Borel summability of independent, identically distributed random variables," *Bull. Inst. Math., Academia Sinica* **1** (1973), 207–220.

- Y. S. Chow and T. L. Lai, "Some one-sided theorems on the tail distribution of sample sums with applications to the last time and largest excess of boundary crossings," *Trans. Amer. Math. Soc.* (1975).
- Y. S. Chow, H. Robbins, and D. Siegmund, *Great Expectations: The Theory of Optimal Stopping*, Houghton Mifflin, Boston, 1972.
- K. L. Chung, "Note on some strong laws of large numbers," *Amer. J. Math.* **69** (1947), 189–192.
- K. L. Chung, "The strong law of large numbers," *Proc. 2nd Berkeley Symp. Stat. and Prob.* (1951), 341–352.
- J. L. Doob, *Stochastic Processes*, Wiley, New York, 1953.
- V. A. Egorov, "On the strong law of large numbers and the law of the iterated logarithm for sequences of independent random variables," *Theor. Prob. Appl.* **15** (1970), 509–514.
- W. Feller, *An Introduction to Probability Theory and Its Applications*, Vol. 2, Wiley, New York, 1966.
- W. Feller, "An extension of the law of the iterated logarithm to variables without variance," *Journ. Math. and Mech.* **18** (1968), 343–356.
- B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables* (K. L. Chung, translator), Addison-Wesley, Reading, Mass., 1954.
- P. Hartman and A. Wintner, "On the law of the iterated logarithm," *Amer. Jour. Math.* **63** (1941), 169–176.
- P. L. Hsu and H. Robbins, "Complete convergence and the law of large numbers," *Proc. Nat. Acad. Sci. U.S.A.* **33** (1947), 25–31.
- A. Khintchine, "Über dyadische Brüche," *Math. Zeit.* **18** (1923), 109–116.
- A. Khintchine, "Über einen Satz der Wahrscheinlichkeitsrechnung," *Fund. Math.* **6** (1924), 9–20.
- J. Kiefer and J. Wolfowitz, "On the characteristics of the general queueing process with applications to random walk," *Ann. Math. Stat.* **27** (1956), 147–161.
- J. F. C. Kingman, "Some inequalities for the queue G1/G/1," *Biometrika* **49** (1962), 315–324.
- A. Kolmogorov, "Über der Gesetz des Iterierten Logarithmus," *Math. Annalen* **101** (1929), 126–135.
- M. Loève, *Probability Theory*, 3rd ed., Van Nostrand, Princeton, 1963; 4th ed., Springer-Verlag, Berlin and New York, 1977–1978.
- J. Marcinkiewicz and A. Zygmund, "Sur les fonctions indépendantes," *Fund. Math.* **29** (1937), 60–90.
- J. Marcinkiewicz and A. Zygmund, "Quelques théorèmes sur les fonctions indépendantes," *Studia Math.* **7** (1938), 104–120.
- S. V. Nagaev, "On necessary and sufficient conditions for the strong law of large numbers," *Theor. Prob. Appl.* **17** (1972), 573–581.
- Y. V. Prohorov, "The strong law of large numbers," *Izv. Akad. Nauk. Ser. Mat.* **14** (1950), 523–536 [in Russian].
- Y. V. Prohorov, "Some remarks on the strong law of large numbers," *Theor. Prob. Appl.* **4** (1959), 201–208.
- D. A. Raikov, "On a connection between the central limit theorem in the theory of probability and the law of large numbers," *Izv. Nauk USSR, Sov. Math.* (1938), 323–338.
- D. Siegmund, "On moments of the maximum of normed sums," *Ann. Math. Stat.* **40** (1969), 527–531.
- F. Spitzer, "A combinatorial lemma and its applications to probability theory," *Trans. Amer. Math. Soc.* **82** (1956), 323–339.
- F. Spitzer, "The Wiener-Hopf equation whose kernel is a probability density," *Duke Math. Jour.* **24** (1957), 327–343.

- V. Strassen, "A converse to law of the iterated logarithm," *Z. Wahr.* **4** (1966), 265–268.
H. M. Taylor, "Bounds for stopped partial sums," *Ann. Math. Stat.* **43** (1972), 733–747.
H. Teicher, "A dominated ergodic type theorem," *Z. Wahr.* **8** (1967), 113–116.
H. Teicher, "Some new conditions for the strong law," *Proc. Nat. Acad. Sci. U.S.A.* **59** (1968), 705–707.
H. Teicher, "Completion of a dominated ergodic theorem," *Ann. Math. Stat.* **42** (1971), 2156–2158.
H. Teicher, "On interchangeable random variables," *Studi di Probabilità, Statistica e Ricerca Operativa in Onore di Giuseppe Pompilj*, pp. 141–148, Gubbio, 1971.
H. Teicher, "On the law of the iterated logarithm," *Ann. Prob.* **2** (1974), 714–728.
H. Teicher, "A necessary condition for the iterated logarithm law," *Z. Wahr. Verw. Geb.* (1975), 343–349.
H. Teicher, "Generalized exponential bounds, iterated logarithm and strong laws," *Z. Wahr. Verw. Geb.* (1979), 293–307.
N. Wiener, "The ergodic theorem," *Duke Math. Jour.* **5** (1938) 1–18.
A. Zygmund, *Trigonometric Series*, Vol. I, Cambridge, 1959.

11 Martingales

An introduction to martingales appeared in Section 7.4, where convergence theorems for submartingales $\{S_n, \mathcal{F}_n, n \geq 1\}$ (relating to differentiation theory) were discussed. Here, emphasis will fall upon convergence theorems for martingales $\{S_{-n}, \mathcal{F}_{-n}, n \leq -1\}$ (relating to ergodic theorems). In demarcating the two cases, it is natural to refer to a martingale $\{S_n, \mathcal{F}_n, n \geq 1\}$ as an upward martingale and to allude to a martingale $\{S_{-n}, \mathcal{F}_{-n}, n \leq -1\}$ when written $\{S_n, \mathcal{F}_n, n \geq 1\}$ as a downward or reverse martingale. Martingale and stochastic inequalities will also be dealt with.

11.1 Upcrossing Inequality and Convergence

The convergence approach of Section 7.4 does not carry over to downward martingales since the formal analogue of a "first time" in the former case is a "last time" in the latter, whereas a genuine first time is not well defined. The ensuing upcrossing inequality provides an avenue of approach to downward martingales.

Definition. If $r_1 < r_2$ and a_1, \dots, a_n are finite, real numbers, the number of upcrossings of the interval $[r_1, r_2]$ by the sequence a_1, a_2, \dots, a_n is defined as the number of times the elements a_j pass from "on or below r_1 " to "on or above r_2 ." More precisely, let α_1 be the smallest integer (if any) for which $a_{\alpha_1} \leq r_1$, and in general for $j \geq 2$ let α_j be the smallest integer (if any) exceeding α_{j-1} for which

$$a_{\alpha_j} \begin{cases} \geq r_2 & \text{if } j \text{ is even} \\ \leq r_1 & \text{if } j \text{ is odd.} \end{cases}$$

If $2u$ is the largest even integer j for which α_j is defined then u is called the number of upcrossings (if α_2 is undefined, then $u = 0$).

Lemma 1 (Panzone). *If $\{S_j, \mathcal{F}_j \equiv \sigma(S_i, 1 \leq i \leq j), 1 \leq j \leq n\}$ is a nonnegative submartingale, r is a positive \mathcal{F}_1 -measurable random variable and U is the number of upcrossings of $[0, r]$ by (S_1, \dots, S_n) , then*

$$\mathbb{E} rU + \mathbb{E} S_1 \leq \int_{\{S_{n-1} > 0\}} S_n + \int_{\{S_{n-1} = 0, S_n \geq r\}} S_n \leq \mathbb{E} S_n. \quad (1)$$

PROOF. For $n = 2$, (1) is confirmed by

$$\begin{aligned} \mathbb{E} rU + \mathbb{E} S_1 &= \int_{\{S_1 = 0, S_2 \geq r\}} r + \int_{\{S_1 > 0\}} S_1 \leq \int_{\{S_1 = 0, S_2 \geq r\}} S_2 \\ &\quad + \int_{\{S_1 > 0\}} S_2 \leq \mathbb{E} S_2. \end{aligned}$$

Suppose inductively that (1) holds with n replaced by $n - 1$ for all submartingales, and set

$$\begin{aligned} T_j &= S_j, \quad 1 \leq j \leq n - 2 \\ T_{n-1} &= \begin{cases} S_n & \text{if } 0 < S_{n-1} < r \\ S_{n-1} & \text{otherwise.} \end{cases} \end{aligned}$$

For $A \in \sigma(T_j, 1 \leq j \leq n - 2) = \sigma(S_j, 1 \leq j \leq n - 2)$

$$\begin{aligned} \int_A T_{n-1} &= \int_A S_{n-1} \leq \int_A S_{n-2} \\ &\leq \int_{A \{S_{n-1} \geq r\}} S_{n-1} + \int_{A \{0 < S_{n-1} < r\}} S_n = \int_A T_{n-1}. \end{aligned}$$

Hence, $\mathbb{E}\{T_{n-1} | T_j, 1 \leq j \leq n - 2\} \geq T_{n-2}$, a.c. Clearly, for $2 \leq m < n - 1$ $\mathbb{E}\{T_m | T_j, 1 \leq j \leq m - 1\} \geq T_{m-1}$, a.c., and so $\{T_j, \mathcal{F}_j, 1 \leq j \leq n - 1\}$ is a nonnegative submartingale.

Let V be the number of upcrossings of $[0, r]$ by (T_1, \dots, T_{n-1}) . Then

$$U = V + I_{\{S_{n-1} = 0, S_n \geq r\}},$$

whence by the induction hypothesis

$$\begin{aligned} \mathbb{E} rU + \mathbb{E} S_1 &= \mathbb{E} rV + \mathbb{E} rI_{\{S_{n-1} = 0, S_n \geq r\}} + \mathbb{E} S_1 \\ &\leq \mathbb{E} T_{n-1} + \mathbb{E} rI_{\{S_{n-1} = 0, S_n \geq r\}} \\ &= \int_{\{0 < S_{n-1} < r\}} S_n + \int_{\{S_{n-1} \geq r\}} S_{n-1} + \mathbb{E} rI_{\{S_{n-1} = 0, S_n \geq r\}} \\ &\leq \int_{\{0 < S_{n-1} < r\}} S_n + \int_{\{S_{n-1} \geq r\}} S_n + \int_{\{S_{n-1} = 0, S_n \geq r\}} S_n \\ &\leq \int_{\{S_{n-1} > 0\}} S_n + \int_{\{S_{n-1} = 0, S_n \geq r\}} S_n \leq \mathbb{E} S_n. \quad \square \end{aligned}$$

for large a since $P\{|S_n| \geq a\} \leq a^{-1} E|S_n| \rightarrow 0$ as $a \rightarrow \infty$. The conclusion of (2) clearly also holds for the finitely many integers n in $[m, 0]$, and so uniform integrability is established.

Next, for $A \in \mathcal{F}_{-\infty}$ and every $m \leq 0$

$$\int_A S_{-\infty} = \lim_{n \rightarrow -\infty} \int_A S_n \leq \int_A S_m,$$

whence $E\{S_m | \mathcal{F}_{-\infty}\} \geq S_{-\infty}$, a.c. for all m , concluding the proof of (i).

Apropos of (ii), $\{S_n^p, \mathcal{F}_n, -\infty < n \leq 0\}$ is a nonnegative submartingale which is easily seen to be u.i., whence the conclusion of (ii) follows. \square

Since U -statistics

$$U_{m,n} = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \dots < i_m \leq n} \varphi(X_{i_1}, \dots, X_{i_m}), \quad n \geq m,$$

constitute a downward martingale (Example 3, Section 7.4), Theorem 1 yields directly a.c. and \mathcal{L}_1 convergence of $U_{m,n}$ as $n \rightarrow \infty$. Moreover, when the $\{X_n\}$ are i.i.d., the Hewitt–Savage zero–one law (Corollary 7.3.7) ensures that the limit is a constant which clearly must coincide with $E\varphi(X_1, \dots, X_m)$.

Corollary 2. If $\{S_n, \mathcal{F}_n, -\infty < n < \infty\}$ is an \mathcal{L}_1 submartingale with $\sup_{-\infty < n < \infty} E|S_n| < \infty$ and $X_n = S_n - S_{n-1}$, then

$$\sum_{-\infty}^{\infty} X_i = \lim_{m,n \rightarrow \infty} \sum_{-m}^n X_i = S_{\infty} - S_{-\infty}, \quad \text{a.c.}$$

where $S_{\pm\infty} = \lim_{n \rightarrow \pm\infty} S_n$ and, moreover, $E|S_{\infty} - S_{-\infty}| < \infty$.

PROOF. By Theorem 1, $\sum_{-m}^0 X_j = S_0 - S_{-m-1} \xrightarrow{\text{a.c.}} S_0 - S_{-\infty}$ with $E|S_{-\infty}| < \infty$ by Fatou. Moreover, by Theorem 7.4.2, $\sum_1^m X_j = S_m - S_0 \xrightarrow{\text{a.c.}} S_{\infty} - S_0$ with $E|S_{\infty}| < \infty$. \square

Corollary 3. If $\{S_n, \mathcal{F}_n, -\infty \leq n \leq 0\}$ is a submartingale with $E S_0^+ < \infty$ and $\lim_{n \rightarrow -\infty} E S_n = K > -\infty$, then $S_n \xrightarrow{\text{a.c.}} S$ and $E\{S | \mathcal{F}_{-\infty}\} \geq S_{-\infty}$, a.c.

PROOF. By Theorem 1, $S_n \xrightarrow{\text{a.c.}} S$ as $n \rightarrow \infty$. For $A \in \mathcal{F}_{-\infty}$

$$\int_A S_{-\infty} \leq \int_A S_n \rightarrow \int_A S$$

by uniform integrability and the corollary follows. \square

Corollary 4. Let $\{Y_n, -\infty < n < \infty\}$ be a sequence of r.v.s on a probability space (Ω, \mathcal{F}, P) and $\{\mathcal{F}_n, -\infty < n < \infty\}$ a sequence of increasing sub- σ -algebras of \mathcal{F} with $\mathcal{F}_{\infty} = \sigma(\bigcup_0^{\infty} \mathcal{F}_n)$ and $\mathcal{F}_{-\infty} = \bigcap_{-\infty}^{-1} \mathcal{F}_n$. If $\lim_{n \rightarrow \infty} Y_n = Y_{\infty}$, a.c., $\lim_{n \rightarrow -\infty} Y_n = Y_{-\infty}$, a.c., and $E \sup_{-\infty < n < \infty} |Y_n| < \infty$, then

$$\begin{aligned} \lim_{n \rightarrow \infty} E\{Y_n | \mathcal{F}_n\} &= E\{Y_{\infty} | \mathcal{F}_{\infty}\}, \quad \text{a.c.,} \\ \lim_{n \rightarrow -\infty} E\{Y_n | \mathcal{F}_n\} &= E\{Y_{-\infty} | \mathcal{F}_{-\infty}\}, \quad \text{a.c.,} \end{aligned} \tag{3}$$

Moreover, since $\int_{\{T \geq j\}} S_{j-1} X_j = 0$,

$$\begin{aligned} - \sum_{j=-\infty}^n \int_{\{T > j\}} S_{j-1} (S_j - S_{j-1}) &= \sum_{-\infty}^n \int_{\{T = j\}} (S_{j-1} S_j - S_{j-1}^2) \\ &\leq \int_{\{T \leq n\}} (C|S_T| - S_{T-1}^2) \\ &\leq CK - \int_{\{T \leq n\}} S_{T-1}^2, \end{aligned}$$

recalling Exercise 7.4.4. Consequently, (5) ensures that for all $n > -\infty$

$$\int_{\{T = \infty\}} \sum_{-\infty}^n X_j^2 \leq C^2 + 2CK.$$

Thus, $\sum_{-\infty}^{\infty} X_j^2 < \infty$, a.c. on the set $\{T = \infty\}$. By Corollary 2, $P\{T = \infty\} = P\{\sup_{n \geq -\infty} |S_n| \leq C\} \rightarrow 1$ as $C \rightarrow \infty$, and so $\sum_{-\infty}^{\infty} X_j^2$ converges a.c. \square

Corollary 5. If $\{S_n = \sum_1^n X_j, \mathcal{F}_n, 1 \leq n < \infty\}$ is an \mathcal{L}_1 bounded martingale, then for every integer $k \geq 2$, $U_{k,n} = \sum_{1 \leq i_1 < \dots < i_k \leq n} X_{i_1} X_{i_2} \cdots X_{i_k}$ converges a.c. as $n \rightarrow \infty$.

PROOF. Since for $k \geq 2$, $\sum_{j=1}^n |X_j|^k \leq (\sum_{j=1}^n X_j^2)^{k/2}$, $n > 1$, Theorem 2 ensures that $\sum_{j=1}^n |X_j|^k$ converges a.c. as $n \rightarrow \infty$, $k \geq 2$, while Theorem 7.4.3 guarantees that S_n converges a.c. The corollary now follows from the identity of Lemma 10.1.2. \square

In view of the identity $U_{k,n} - U_{k,n-1} = X_n U_{k-1,n-1}$, $n \geq k > 1$, the prior corollary asserts that if $S_n = \sum_1^n X_j$, $n \geq 1$, is an \mathcal{L}_1 bounded martingale, then $U_{k,n} = \sum_{j=k}^n X_j U_{k-1,j-1}$ converges a.c. Note that in the formation of $U_{k,n}$ the martingale differences X_j have been multiplied by an \mathcal{F}_{j-1} -measurable function, viz., $U_{k-1,j-1}$. A more general result of this nature appears in part (ii) of the ensuing theorem.

For any sequence of r.v.s $\{Y_n\}$, the generic notation

$$Y^* = \sup_n |Y_n| \tag{6}$$

will be employed for the **maximal function** $\sup_n |Y_n|$.

Lemma 2. Let $\{S_n = \sum_1^n X_j, \mathcal{F}_n, n \geq 1\}$ be an \mathcal{L}_1 martingale and $V_r = (\sum_1^{\infty} |X_n|^r)^{1/r}$, $r \geq 1$. Then for every $K > 0$

$$X_n = X_n^{(1)} + X_n^{(2)} + X_n^{(3)}, \tag{7}$$

where for $j = 1, 2, 3$, $\{X_n^{(j)}, \mathcal{F}_n, n \geq 1\}$ are martingale difference sequences satisfying

$$E \sum_1^{\infty} |X_n^{(1)}|^r \leq c_r E \min(V_r, K)^r \tag{8}$$

2. If $\{X_n, n \geq 1\}$ are \mathcal{L}_1 interchangeable r.v.s, prove that $1/n \sum_1^n X_i \xrightarrow{\text{a.s.}}$ some r.v. Y (also in \mathcal{L}_1). Verify that $Y = E\{X_1 | \bigcap_1^\infty \mathcal{G}_n\}$, where $\mathcal{G}_n = \sigma(\sum_1^j X_i, j \geq n)$. Hint: $n^{-1} \sum_1^n X_i = E\{n^{-1} \sum_1^n X_i | \mathcal{G}_n\} = E\{X_1 | \mathcal{G}_n\}$.
3. Let Y be a r.v. with $E|Y| < \infty$ and $E|Y|\log(1 + |Y|) < \infty$. If $\{\mathcal{G}_n, n \geq 1\}$ is a sequence of independent σ -algebras of events, then $E\{Y | \mathcal{G}_n\} \xrightarrow{\text{a.s.}} 0$ as $n \rightarrow \infty$. Hint: Apply Corollary 4 and Theorem 7.4.8.

11.2 Martingale Extension of Marcinkiewicz–Zygmund Inequalities

Let (Ω, \mathcal{F}, P) be a probability space and let $\{\emptyset, \Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_\infty \subset \mathcal{F}$ be a sequence of σ -algebras with $\mathcal{F}_\infty = \sigma(\bigcup_1^\infty \mathcal{F}_n)$. For any stochastic sequence $\{f_n, \mathcal{F}_n, n \geq 1\}$ define $f_0 = 0$ and

$$\begin{aligned} f &= \{f_n, n \geq 1\}, \quad f^* = \sup_{n \geq 1} |f_n|, \quad \|f\|_p = \sup_{n \geq 1} \|f_n\|_p, \quad f_\infty = \overline{\lim}_{n \rightarrow \infty} f_n, \\ f_n^* &= \max_{1 \leq j \leq n} |f_j|, \quad d_n = f_n - f_{n-1}, \quad n \geq 1, \quad d = \{d_n, n \geq 1\}, \\ S(f) &= S_\infty(f) = \left(\sum_1^\infty d_j^2 \right)^{1/2}, \quad S_n(f) = \left(\sum_1^n d_j^2 \right)^{1/2}. \end{aligned}$$

Recall that f is \mathcal{L}_p bounded if $\|f\|_p < \infty$. Moreover, f will be called a submartingale or martingale whenever $\{f_n, \mathcal{F}_n, n \geq 1\}$ is such. For any real numbers a, b let $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$.

Lemma 1. *If f is an \mathcal{L}_1 bounded martingale or nonnegative \mathcal{L}_1 bounded submartingale and $T = \inf\{n \geq 1 : |f_n| > \lambda\}$, $\lambda > 0$, then*

$$E S_{T-1}^2(f) + E f_{T-1}^2 \leq 2 E f_T f_{T-1} \leq 2\lambda \|f\|. \quad (1)$$

PROOF. Since $|f_{T-1}| \leq \lambda$, and $f_\infty = \overline{\lim} f_n = \underline{\lim} f_n$, a.c., by Fatou

$$E|f_T f_{T-1}| \leq \lambda E|f_T| \leq \lambda \lim_{n \rightarrow \infty} E|f_n| = \lambda \|f\|,$$

and so it remains to prove the first inequality in (1). Setting $T_n = T \wedge n$ for $n \geq 1$,

$$\begin{aligned} S_{n-1}^2(f) + f_{n-1}^2 &= 2 \sum_{1 \leq j \leq k \leq n-1} d_j d_k = 2 \sum_{j=1}^{n-1} d_j (f_{n-1} - f_{j-1}) \\ &= 2f_{n-1}^2 - 2 \sum_1^{n-1} f_{j-1} d_j = 2 \left[f_n f_{n-1} - \sum_1^n f_{j-1} d_j \right], \end{aligned}$$

whence (employing Corollary 7.4.5 in the martingale case)

$$E[S_{T_n-1}^2(f) + f_{T_n-1}^2] \leq 2 E f_{T_n} f_{T_n-1}.$$

so that the first inequality of (7) fails for $p = 1$. However, the second inequality does hold. More precisely, Corollary 1 obtains when $p = 1$, as will be shown in the next theorem.

Lemma 4. *Let $f = \{f_n \equiv \sum_1^n d_j, n \geq 1\}$ be an \mathcal{L}_1 martingale with $|d_j| \leq V_j$, where $\{V_n, \mathcal{F}_{n-1}, n \geq 1\}$ is a stochastic sequence. If $\lambda > 0, \beta > 1$ and $0 < \delta < \beta - 1$ then*

$$\mathbb{P}\{f^* > \beta\lambda, S(f) \vee V^* \leq \delta\lambda\} \leq \frac{2\delta^2}{(\beta - \delta - 1)^2} \mathbb{P}\{f^* > \lambda\}, \quad (15)$$

$$\mathbb{P}\{S(f) > \beta\lambda, f^* \vee V^* \leq \delta\lambda\} \leq \frac{9\delta^2}{\beta^2 - \delta^2 - 1} \mathbb{P}\{S(f) > \lambda\}. \quad (16)$$

PROOF. Set $S_0(f) = 0$ and define

$$\mu = \inf\{n \geq 1 : |f_n| > \lambda\}, \quad v = \inf\{n \geq 1 : |f_n| > \beta\lambda\},$$

$$\sigma = \inf\{n \geq 0 : S_n(f) \vee V_{n+1} > \delta\lambda\}.$$

If

$$h_n = \sum_{j=1}^n d_j I_{[\mu < j \leq v \wedge \sigma]}$$

then $h = \{h_n, n \geq 1\}$ is a martingale with $S(h) = 0$ on $\{\mu = \infty\} = \{f^* \leq \lambda\}$.

Moreover, recalling that $|d_j| \leq V_j$ by hypothesis,

$$\begin{aligned} S^2(h) &\leq S_\sigma^2(f) = [S_{\sigma-1}^2(f) + d_\sigma^2]I_{\{\sigma < \infty\}} + S_\sigma^2(f)I_{\{\sigma = \infty\}} \\ &\leq 2\delta^2\lambda^2. \end{aligned}$$

Hence,

$$\|h\|_2^2 = \mathbb{E} S^2(h)I_{\{f^* > \lambda\}} \leq 2\delta^2\lambda^2 \mathbb{P}\{f^* > \lambda\},$$

implying via the Kolmogorov inequality (Theorem 7.4.8) that

$$\begin{aligned} &\mathbb{P}\{f^* > \beta\lambda, S(f) \vee V^* \leq \delta\lambda\} \\ &= \mathbb{P}\{\sigma = \infty, v < \infty, h_n = f_{v \wedge n} - f_{\mu \wedge n}, n \geq 1\} \\ &\leq \mathbb{P}\{h^* > \beta\lambda - (1 + \delta)\lambda\} \leq [(\beta - 1 - \delta)\lambda]^{-2} \|h\|_2^2 \\ &\leq \frac{2\delta^2}{(\beta - \delta - 1)^2} \mathbb{P}\{f^* > \lambda\}. \end{aligned}$$

To obtain (16), define analogously

$$\begin{aligned} \mu' &= \inf\{n \geq 1 : S_n(f) > \lambda\}, \quad v' = \inf\{n \geq 1 : S_n(f) > \beta\lambda\} \\ \sigma' &= \inf\{n \geq 0 : f_n^* \vee V_{n+1} > \delta\lambda\}. \end{aligned}$$

If

$$g_n = \sum_{j=1}^n d_j I_{[\mu' < j \leq v' \wedge \sigma']}$$

then $g = \{g_n, n \geq 1\}$ is a martingale with $g^* = 0$ on $\{\mu' \geq \sigma'\} \supset \{S(f) \leq \lambda\}$, whence

$$\begin{aligned} g^* &\leq (f_{\sigma'}^* + f_{\mu'}^*)I_{\{\mu' < \sigma' < \infty\}} + 2f^*I_{\{\mu' < \sigma' = \infty\}} \\ &\leq 3\delta\lambda, \end{aligned}$$

implying $E(g^*)^2 \leq 9\delta^2\lambda^2 P\{S(f) > \lambda\}$. Thus,

$$\begin{aligned} &P\{S(f) > \beta\lambda, f^* \vee V^* \leq \delta\lambda\} \\ &= P\{v' < \infty = \sigma', S_n^2(g) = S_{v' \wedge n}^2(f) - S_{\mu' \wedge n}^2(f), n \geq 1\} \\ &\leq P\{S^2(g) > [\beta^2 - (1 + \delta^2)]\lambda^2\} \leq [(\beta^2 - 1 - \delta^2)\lambda^2]^{-1} E S^2(g) \\ &\leq [(\beta^2 - 1 - \delta^2)\lambda^2]^{-1} E(g^*)^2 \leq \frac{9\delta^2}{\beta^2 - \delta^2 - 1} P\{S(f) > \lambda\}. \quad \square \end{aligned}$$

Lemma 5. If $f = \{f_n = \sum_1^n d_i, n \geq 1\}$ is an \mathcal{L}_1 martingale and

$$g_n = \sum_1^n a_j, \quad a_j = d_j I_{\{|d_j| \leq 2d_{j-1}^*\}} - E\{d_j I_{\{|d_j| \leq 2d_{j-1}^*\}} | \mathcal{F}_{j-1}\}, \quad (17)$$

$$h_n = \sum_1^n b_j, \quad b_j = d_j I_{\{|d_j| > 2d_{j-1}^*\}} - E\{d_j I_{\{|d_j| > 2d_{j-1}^*\}} | \mathcal{F}_{j-1}\}, \quad (18)$$

then $g = \{g_n, n \geq 1\}$ and $h = \{h_n, n \geq 1\}$ are \mathcal{L}_1 martingales with $f_n = g_n + h_n, n \geq 1$ and

$$|a_n| \leq 4d_{n-1}^*, \quad (19)$$

$$\sum_{n=1}^{\infty} |d_n I_{\{|d_n| > 2d_{n-1}^*\}}| \leq 2d^*, \quad (20)$$

$$\sum_{n=1}^{\infty} E|b_n| \leq 4 E d^*. \quad (21)$$

PROOF. The validity of (19) is clear. On the set $\{|d_j| > 2d_{j-1}^*\}, |d_j| + 2d_{j-1}^* \leq 2|d_j| \leq 2d_j^*$, implying

$$\sum_1^{\infty} |d_j I_{\{|d_j| > 2d_{j-1}^*\}}| \leq 2 \sum_1^{\infty} (d_j^* - d_{j-1}^*) = 2d^*,$$

which is (20). This, in turn, yields (21) via

$$\begin{aligned} \sum_1^{\infty} E|b_j| &\leq E \sum_1^{\infty} |d_j I_{\{|d_j| > 2d_{j-1}^*\}}| + \sum_1^{\infty} E|E\{d_j I_{\{|d_j| > 2d_{j-1}^*\}} | \mathcal{F}_{j-1}\}| \\ &\leq 2 E d^* + 2 E d^* = 4 E d^*. \quad \square \end{aligned}$$

Theorem 2 (Davis). There exist constants $0 < A < B < \infty$ such that, for any \mathcal{L}_1 martingale $f = \{f_n, n \geq 1\}$,

$$A E S(f) \leq E f^* \leq B E S(f). \quad (22)$$

PROOF. Writing $f_n = g_n + h_n$ as in Lemma 5, it follows therefrom that

$$\mathbb{E} f^* \leq \mathbb{E}(g^* + h^*) \leq \mathbb{E} g^* + \sum_1^\infty \mathbb{E}|b_j| \leq \mathbb{E} g^* + 4 \mathbb{E} d^*, \quad (23)$$

$$\mathbb{E} S(f) \leq \mathbb{E}[S(g) + S(h)] \leq \mathbb{E} S(g) + \sum_1^\infty \mathbb{E}|b_j| \leq \mathbb{E} S(g) + 4 \mathbb{E} d^*. \quad (24)$$

Since $g_n = \sum_1^n a_j$ is a martingale with $|a_n| \leq 4d_{n-1}^*$, Lemma 4 ensures that for $\lambda > 0$, $\beta > \delta$, and $0 < \delta < 1$

$$\begin{aligned} \mathbb{P}\{g^* > \beta\lambda, S(g) \vee 4d^* \leq \delta\lambda\} &\leq \frac{2\delta^2}{(\beta - \delta - 1)^2} \mathbb{P}\{g^* > \lambda\}, \\ \mathbb{P}\{S(g) > \beta\lambda, g^* \vee 4d^* \leq \delta\lambda\} &\leq \frac{9\delta^2}{\beta^2 - \delta^2 - 1} \mathbb{P}\{S(g) > \lambda\}. \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{P}\{g^* > \beta\lambda\} &\leq \mathbb{P}\{S(g) > \delta\lambda\} + \mathbb{P}\{4d^* > \delta\lambda\} + \frac{2\delta^2}{(\beta - \delta - 1)^2} \mathbb{P}\{g^* > \lambda\}, \\ \mathbb{P}\{S(g) > \beta\lambda\} &\leq \mathbb{P}\{g^* > \delta\lambda\} + \mathbb{P}\{4d^* > \delta\lambda\} + \frac{9\delta^2}{\beta^2 - \delta^2 - 1} \mathbb{P}\{S(g) > \lambda\}, \end{aligned}$$

implying

$$\beta^{-1} \mathbb{E} g^* \leq \delta^{-1} \mathbb{E} S(g) + 4\delta^{-1} \mathbb{E} d^* + \frac{2\delta^2}{(\beta - \delta - 1)^2} \mathbb{E} g^*, \quad (25)$$

$$\beta^{-1} \mathbb{E} S(g) \leq \delta^{-1} \mathbb{E} g^* + 4\delta^{-1} \mathbb{E} d^* + \frac{9\delta^2}{\beta^2 - \delta^2 - 1} \mathbb{E} S(g). \quad (26)$$

Now, recalling Lemma 5,

$$\mathbb{E} g^* \leq \mathbb{E}(f^* + h^*) \leq \mathbb{E} f^* + \sum_1^\infty \mathbb{E}|b_j| \leq \mathbb{E} f^* + 4 \mathbb{E} d^*, \quad (27)$$

$$\mathbb{E} S(g) \leq \mathbb{E}[S(f) + S(h)] \leq \mathbb{E} S(f) + \sum_1^\infty \mathbb{E}|b_j| \leq \mathbb{E} S(f) + 4 \mathbb{E} d^*. \quad (28)$$

In order to subtract in (25) and (26), rewrite these for the martingale $\{g_1, g_2, \dots, g_n, g_n, g_n, \dots\}$, obtaining via (27) and (28)

$$\left[\beta^{-1} - \frac{2\delta^2}{(\beta - \delta - 1)^2} \right] \mathbb{E} g_n^* \leq \delta^{-1} \mathbb{E} S_n(f) + 8\delta^{-1} \mathbb{E} d_n^* \leq 9\delta^{-1} \mathbb{E} S(f), \quad (29)$$

$$\left[\beta^{-1} - \frac{9\delta^2}{\beta^2 - \delta^2 - 1} \right] \mathbb{E} S_n(g) \leq \delta^{-1} \mathbb{E} f_n^* + 8\delta^{-1} \mathbb{E} d_n^* \leq 17\delta^{-1} \mathbb{E} f^*. \quad (30)$$

For small δ , the coefficients on the left in (29), (30) are positive, and so, letting $n \rightarrow \infty$ and recalling (23), (24), there exist constants B_1, B, A_1, A for which

$$\mathbb{E} f^* \leq B_1 \mathbb{E} S(f) + 4 \mathbb{E} d^* \leq B \mathbb{E} S(f),$$

$$\mathbb{E} S(f) \leq A_1 \mathbb{E} f^* + 4 \mathbb{E} d^* \leq A^{-1} \mathbb{E} f^*. \quad \square$$

Corollary 4. *If the stochastic sequence $\{|X_n|^p, \mathcal{F}_n, n \geq 1\}$ is uniformly integrable for some $p \in (0, 2)$ and $S_n = \sum_1^n X_i$, then*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}|S_n - a_n|^p = 0, \quad (31)$$

where $a_n \equiv 0$ for $0 < p < 1$ and $a_n = \sum_{j=1}^n \mathbb{E}\{X_j | \mathcal{F}_{j-1}\}$ if $1 \leq p < 2$.

PROOF. Define $Y_n = X_n$ if $0 < p < 1$ and $Y_n = X_n - \mathbb{E}\{X_n | \mathcal{F}_{n-1}\}$ when $1 \leq p < 2$. Then $\{|Y_n|^p, n \geq 1\}$ is u.i., whence for any $\varepsilon > 0$ there exists a constant $M > 1$ with

$$\sup_{n \geq 1} \int_{\{|Y_n| > M\}} |Y_n|^p < \varepsilon.$$

Set

$$Y'_n = Y_n I_{\{|Y_n| \leq M\}}, \quad Y''_n = Y_n - Y'_n.$$

Then, if $0 < p < 1$,

$$\begin{aligned} \mathbb{E}|S_n|^p &= \mathbb{E} \left| \sum_1^n (Y'_j + Y''_j) \right|^p \leq \mathbb{E} \left(\sum_1^n |Y'_j| \right)^p + \mathbb{E} \left(\sum_1^n |Y''_j| \right)^p \\ &\leq (nM)^p + \sum_1^n \mathbb{E}|Y''_j|^p \leq (nM)^p + n\varepsilon \end{aligned}$$

and (31) follows. If, rather, $1 \leq p < 2$, then Theorems 1 and 2 guarantee the existence of a constant C with

$$\begin{aligned} \mathbb{E}|S_n - a_n|^p &= \mathbb{E} \left| \sum_1^n Y_j \right|^p \leq C \mathbb{E} \left(\sum_1^n Y_j^2 \right)^{p/2} \\ &\leq C \mathbb{E} \left[\left(\sum_1^n Y_j^2 \right)^{p/2} + \left(\sum_1^n Y_j^2 \right)^{p/2} \right] \\ &\leq C \left[(nM^2)^{p/2} + \mathbb{E} \sum_1^n |Y_j|^p \right] \leq C[(nM^2)^{p/2} + n\varepsilon], \end{aligned}$$

again yielding (31). \square

EXERCISES 11.2

- Let $\{X_n, n \geq 1\}$ be independent r.v.s with $\mathbb{E} X_n = 0, \mathbb{E} X_n^2 = 1$ for $n \geq 1$. If $T = \inf\{n \geq 1 : \sum_1^n X_j > 0\}$, then $\mathbb{P}\{T < \infty\} = 1$ and $\mathbb{E} T^{1/2} = \infty$.
- Let $\{X_n, n \geq 1\}$ be interchangeable r.v.s with $\mathbb{E} X_1 = 0 = \mathbb{E} X_1 X_2, \mathbb{E} X_1^2 = 1$. If $T = \inf\{n \geq 1 : \sum_1^n X_1 > 0\}$, then $\mathbb{P}\{T < \infty\} = 1$ and $\mathbb{E} T^{1/2} = \infty$.

PROOF. Set $W_0 = Z_0 = 0$ and define for $n \geq 1$

$$\begin{aligned} Z_n &= \sum_1^n z_j, & Z &= Z_\infty = \sum_1^\infty z_j, \\ W_n &= \sum_1^n E\{z_j | \mathcal{F}_{j-1}\}, & W &= W_\infty = \sum_1^\infty E\{z_j | \mathcal{F}_{j-1}\}. \end{aligned}$$

In verifying (11) it may be supposed that $E Z > 0$ and $E \Phi(Z) < \infty$. Then convexity and $\Phi(\infty-) = \infty$ entail $E Z < \infty$, so that $E W < \infty$. For $\lambda \geq 0$ define the \mathcal{F}_n -time $T = \inf\{n \geq 0 : W_{n+1} > \lambda\}$, whence

$$\begin{aligned} \int_{[0 \leq T \leq n]} (W - Z - \lambda) &= \sum_{j=0}^n \int_{[T=j]} (W_j - \lambda + W - W_j - Z) \\ &\leq \sum_{j=0}^n \int_{[T=j]} E\{W - W_j - Z | \mathcal{F}_j\} \leq 0 \end{aligned}$$

or

$$\int_{[0 \leq T \leq n]} (W - \lambda) \leq \int_{[0 \leq T \leq n]} Z,$$

and so, letting $n \rightarrow \infty$,

$$E(W - \lambda)^+ = \int_{[0 \leq T < \infty]} (W - \lambda) \leq \int_{[0 \leq T < \infty]} Z = \int_{[W > \lambda]} Z. \quad (12)$$

Next, setting $d = c - 1$ and applying Lemma 2,

$$t\varphi(t) \leq \int_t^{2t} \varphi(u) du = \Phi(2t) - \Phi(t) \leq d\Phi(t). \quad (13)$$

Since φ is nonnegative, nondecreasing on $[0, \infty)$, and bounded on every finite interval $[0, M]$, necessarily φ is a function of bounded variation. Hence, via integration by parts (for Riemann-Stieltjes integrals), for $t \in (0, \infty)$

$$\begin{aligned} \int_0^t \varphi(u) du &= t\varphi(t) - \int_0^t u d\varphi(u) \\ &= \int_0^t (t-u) d\varphi(u) + t\varphi(0) = \int_0^\infty (t-u)^+ d\varphi(u) + t\varphi(0), \end{aligned}$$

and so by Theorem 6.2.4 and Lemma 2,

$$\Phi(t) = \int_0^\infty (t-u)^+ d\varphi(u) + t\varphi(0), \quad (14)$$

where the integral is in the Lebesgue–Stieltjes sense. From (14), recalling (12),

$$\begin{aligned} \mathbb{E} \Phi(W_n) &= \mathbb{E} \left[\int_0^\infty (W_n - u)^+ d\varphi(u) + W_n \varphi(0) \right] \\ &= \int_0^\infty \mathbb{E}(W_n - u)^+ d\varphi(u) + \varphi(0) \mathbb{E} W_n \\ &\leq \int_0^\infty \int_{[W_n > u]} Z_n d\mathbb{P} d\varphi(u) + \varphi(0) \mathbb{E} Z_n \\ &= \mathbb{E} Z_n \int_{[0, W_n)} d\varphi(u) + \varphi(0) \mathbb{E} Z_n = \mathbb{E} Z_n \varphi(W_n), \end{aligned}$$

implying

$$a \mathbb{E} \Phi\left(\frac{W_n}{a}\right) \leq \mathbb{E} Z_n \varphi\left(\frac{W_n}{a}\right) \quad \text{for } a \in (0, \infty). \quad (15)$$

Let ψ on $[0, \infty)$ be an inverse function of φ , that is, $\psi(\varphi(u)) = u = \varphi(\psi(u))$ for $u \in (0, \infty)$, and set

$$\Psi(t) = \int_0^t \psi(u) du.$$

Then, by Young's inequality (Exercise 6.2.15) $uv \leq \Phi(u) + \Psi(v)$ for $u, v \in (0, \infty)$ with equality holding for $v = \varphi(u)$. Consequently,

$$\frac{Z_n}{b} \varphi\left(\frac{W_n}{a}\right) \leq \Phi\left(\frac{Z_n}{b}\right) + \Psi\left(\varphi\left(\frac{W_n}{a}\right)\right),$$

and so from (15)

$$a \mathbb{E} \Phi\left(\frac{W_n}{a}\right) \leq b \left[\mathbb{E} \Phi\left(\frac{Z_n}{b}\right) + \mathbb{E} \Psi\left(\varphi\left(\frac{W_n}{a}\right)\right) \right]. \quad (16)$$

Next, take $v = \varphi(u)$ in Young's inequality, yielding via (13)

$$\Psi(\varphi(t)) = t\varphi(t) - \Phi(t) \leq (d-1)\Phi(t), \quad d \geq 1,$$

and, combining this with (16),

$$[a - b(d-1)] \mathbb{E} \Phi\left(\frac{W_n}{a}\right) \leq b \mathbb{E} \Phi\left(\frac{Z_n}{b}\right). \quad (17)$$

From (13), $\varphi(u) \leq (d/u)\Phi(u)$, whence for $r \in [1, \infty)$

$$\log \frac{\Phi(ru)}{\Phi(u)} = \int_u^{ru} \frac{\varphi(t)}{\Phi(t)} dt \leq d \int_u^{ru} \frac{dt}{t} = \log r^d$$

or

$$\Phi(ru) \leq r^d \Phi(u), \quad r \geq 1. \quad (18)$$

Taking $a = d$ and $b = 1$ in (17),

$$\mathbb{E} \Phi\left(\frac{W_n}{d}\right) \leq \mathbb{E} \Phi(Z_n),$$

whence, recalling (18),

$$\mathbb{E} \Phi(W_n) \leq \mathbb{E} \Phi(dZ_n) \leq d^d \mathbb{E} \Phi(Z_n), \quad (19)$$

and (11) follows with $B = d^d$ by monotone convergence. \square

Corollary 1 (Garsia). *For every sequence $\{z_n, n \geq 1\}$ of nonnegative measurable functions on (Ω, \mathcal{F}, P) and any $p > 1$*

$$\mathbb{E} \left[\sum_1^\infty \mathbb{E}\{z_j | \mathcal{F}_{j-1}\} \right]^p \leq p^p \mathbb{E} \left(\sum_1^\infty z_j \right)^p. \quad (20)$$

PROOF. When $\Phi(t) = t^p$, (13) yields $d = p$, whence (20) follows directly from (19).

Theorem 1 (Burkholder–Davis–Gundy). *Let Φ be a nondecreasing function on $[0, \infty]$, finite and convex on $[0, \infty)$ with $\Phi(0) = 0$, $\Phi(\infty) = \Phi(\infty -)$, and $\Phi(2\lambda) \leq c\Phi(\lambda)$ for all $\lambda > 0$ and some c in $(0, \infty)$. Then there exist constants $0 < A_c < B_c < \infty$ such that for any martingale f*

$$A_c \mathbb{E} \Phi(S(f)) \leq \mathbb{E} \Phi(f^*) \leq B_c \mathbb{E} \Phi(S(f)), \quad (21)$$

where $S(f)$ and f^* are as in Section 2.

PROOF. By Lemma 11.2.5, $f = g + h$, where the martingales g, h are defined by $g_n = \sum_{j=1}^n a_j$, $h_n = \sum_{j=1}^n b_j$ with

$$\begin{aligned} a_n &= d_n I_{\{|d_n| \leq 2d_{n-1}^*\}} - \mathbb{E}\{d_n I_{\{|d_n| \leq 2d_{n-1}^*\}} | \mathcal{F}_{n-1}\}, \\ b_n &= d_n I_{\{|d_n| > 2d_{n-1}^*\}} + \mathbb{E}\{d_n I_{\{|d_n| \leq 2d_{n-1}^*\}} | \mathcal{F}_{n-1}\}, \\ Z &\equiv \sum_1^\infty |d_j I_{\{|d_j| > 2d_{j-1}^*\}}| \leq 2d^*, \quad |a_n| \leq 4d_{n-1}^*, \\ W &\equiv \sum_1^\infty |\mathbb{E}\{d_j I_{\{|d_j| \leq 2d_{j-1}^*\}} | \mathcal{F}_{j-1}\}| \leq \sum_1^\infty \mathbb{E}\{|d_j| I_{\{|d_j| > 2d_{j-1}^*\}} | \mathcal{F}_{j-1}\}. \end{aligned} \quad (22)$$

Then

$$\begin{aligned} f^* &\leq g^* + h^* \leq g^* + \sum_1^\infty |b_j| \leq g^* + Z + W, \\ S(g) &\leq S(f) + S(h) \leq S(f) + \sum_1^\infty |b_j| \leq S(f) + Z + W, \end{aligned} \quad (23)$$

and

$$\begin{aligned} S(f) &\leq S(g) + S(h) \leq S(g) + Z + W, \quad d^* \leq S(f) \\ g^* &\leq f^* + h^* \leq f^* + Z + W, \quad d^* \leq 2f^*. \end{aligned} \quad (24)$$

By Lemmas 11.2.4 and 1 there are finite, positive constants $B_j = B_j(c)$, $j = 1, 2$ such that

$$\begin{aligned} \mathbb{E} \Phi(g^*) &\leq B_1 \mathbb{E} \Phi(S(g) \vee 4d^*) \leq B_1 \mathbb{E}[\Phi(S(g)) + \Phi(4d^*)], \\ \mathbb{E} \Phi(S(g)) &\leq B_2 \mathbb{E} \Phi(g^* \vee 4d^*) \leq B_2 \mathbb{E}[\Phi(g^*) + \Phi(4d^*)]. \end{aligned} \quad (25)$$

Moreover, by Lemma 3, for some $B_3 = B_3(c) \in (0, \infty)$

$$\mathbb{E} \Phi(W) \leq B_3 \mathbb{E} \Phi(Z), \quad (26)$$

whence via (23)

$$\begin{aligned} \mathbb{E} \Phi(f^*) &\leq \mathbb{E} \Phi(g^* + Z + W) \leq \mathbb{E}[\Phi(3g^*) + \Phi(3Z) + \Phi(3W)] \\ &\leq c^2 \mathbb{E}[\Phi(g^*) + \Phi(Z) + \Phi(W)] \\ &\leq c^2 B_1 \mathbb{E}[\Phi(S(g)) + \Phi(4d^*)] + c^2 \mathbb{E}[\Phi(Z) + \Phi(W)] \quad \text{by (25)} \\ &\leq c^2 B_1 \mathbb{E} \Phi(S(f) + Z + W) + c^2 B_1 \mathbb{E} \Phi(4d^*) \\ &\quad + c^2 \mathbb{E}[\Phi(Z) + \Phi(W)] \quad \text{by (23)} \\ &\leq c^4 B_1 \mathbb{E} \Phi(S(f)) + c^2 B_1 \mathbb{E} \Phi(4d^*) \\ &\quad + (c^4 B_1 + c^2) \mathbb{E}[\Phi(Z) + \Phi(W)] \\ &\leq c^4 B_1 \mathbb{E} \Phi(S(f)) + c^2 B_1 \mathbb{E} \Phi(4d^*) \\ &\quad + (c^4 B_1 + c^2)(1 + B_3) \mathbb{E} \Phi(Z) \quad \text{by (26)} \\ &\leq c^4 B_1 \mathbb{E} \Phi(S(f)) + c^4 B_1 \mathbb{E} \Phi(d^*) \\ &\quad + (c^5 B_1 + c^3)(1 + B_3) \mathbb{E} \Phi(d^*) \quad \text{by (22)} \\ &\leq [2c^4 B_1 + (1 + B_3)(c^5 B_1 + c^3)] \mathbb{E} \Phi(S(f)) \quad \text{by (24),} \end{aligned}$$

yielding the upper inequality in (21). Similarly, for some finite, positive constants $A_j = A_j(c)$, $1 \leq j \leq 7$,

$$\begin{aligned} \mathbb{E} \Phi(S(f)) &\leq \mathbb{E} \Phi(S(g) + Z + W) \quad \text{by (24)} \\ &\leq A_2 \mathbb{E}[\Phi(g^*) + \Phi(4d^*) + \Phi(Z) + \Phi(W)] \quad \text{by (25)} \\ &\leq A_3 \mathbb{E}[\Phi(f^* + Z + W) + \Phi(4d^*) + \Phi(Z) + \Phi(W)] \quad \text{by (24)} \\ &\leq A_4 \mathbb{E}[\Phi(f^*) + \Phi(d^*) + \Phi(Z) + \Phi(W)] \\ &\leq A_5 \mathbb{E}[\Phi(f^*) + \Phi(d^*) + \Phi(Z)] \quad \text{by (26)} \\ &\leq A_6 \mathbb{E}[\Phi(f^*) + \Phi(d^*)] \quad \text{by (22)} \\ &\leq A_7 \mathbb{E} \Phi(f^*), \end{aligned}$$

completing the proof of (21).

As will be demonstrated shortly, it is useful to have a version of Theorem 1 with $S(f) = (\sum_1^\infty d_j^2)^{1/2}$ replaced by $s(f) = (\sum_{j=1}^\infty \mathbb{E}\{d_j^2 | \mathcal{F}_{j-1}\})^{1/2}$, and for this an analogue of Lemma 11.2.4 is needed.

Lemma 4. *If f is a martingale, $1 \leq \alpha \leq 2$, and*

$$s(f) = s_\alpha(f) = \left(\sum_1^\infty \mathbb{E}\{|d_j|^\alpha | \mathcal{F}_{j-1}\} \right)^{1/\alpha}, \quad (27)$$

then for any $\lambda > 0$, $\beta > 1$, and $\delta \in (0, \beta - 1)$ there exists a finite, positive constant B_α such that

$$\mathbb{P}\{f^* > \beta\lambda, s(f) \vee d^* \leq \delta\lambda\} \leq \frac{B_\alpha \delta^\alpha}{(\beta - 1 - \delta)^\alpha} \mathbb{P}\{f^* > \lambda\}. \quad (28)$$

PROOF. Define $s_n(f) = (\sum_{j=1}^n \mathbb{E}\{|d_j|^2 | \mathcal{F}_{j-1}\})^{1/2}$ and

$$\begin{aligned} \mu &= \inf\{n \geq 1 : |f_n| > \lambda\}, \\ \nu &= \inf\{n \geq 1 : |f_n| > \beta\lambda\}, \\ \sigma &= \inf\{n \geq 0 : |d_n| \vee s_{n+1}(f) > \delta\lambda\} \end{aligned}$$

and

$$h_n = \sum_{j=1}^n d_j I_{[\mu < j \leq \nu \wedge \sigma]} \equiv \sum_{j=1}^n a_j \quad (\text{say}).$$

Then $\{h_n, \mathcal{F}_n, n \geq 1\}$ is a martingale with $S(h) = 0$ on $\{\mu = \infty\}$ and

$$s^2(h) = s_\sigma^2(h) I_{[\mu < \infty]} \leq \delta^2 \lambda^2 I_{\{f^* > \lambda\}}.$$

Hence, by Theorems 11.2.1 and 11.2.2, for some $B = B_\alpha \in (0, \infty)$

$$\begin{aligned} \|h\|_\alpha^2 &= \sup_{n \geq 1} \mathbb{E}|h_n|^2 \leq \mathbb{E} h^{*2} \leq B_\alpha \mathbb{E} \left(\sum_1^\infty a_j^2 \right)^{\alpha/2} \leq B_\alpha \mathbb{E} \sum_1^\infty |a_j|^2 \\ &= B_\alpha \mathbb{E} s^2(h) \leq B_\alpha \delta^2 \lambda^2 \mathbb{P}\{f^* > \lambda\}, \end{aligned}$$

whence

$$\begin{aligned} &\mathbb{P}\{f^* > \beta\lambda, s(f) \vee d^* \leq \delta\lambda\} \\ &= \mathbb{P}\{\nu < \infty, \sigma = \infty, h_n = f_{\nu \wedge n} - f_{\mu \wedge n} \quad \text{for all } n \geq 1\} \\ &\leq \mathbb{P}\{h^* > \beta\lambda - \lambda - \delta\lambda\} \\ &\leq B_\alpha \left(\frac{\delta}{\beta - 1 - \delta} \right)^\alpha \mathbb{P}\{f^* > \lambda\}. \end{aligned} \quad \square$$

Theorem 2. Let Φ be a nondecreasing function on $[0, \infty]$, continuous on $[0, \infty)$ with $\Phi(0) = 0$, $\Phi(\infty) = \Phi(\infty -)$, and $\Phi(2\lambda) \leq c\Phi(\lambda)$ for all $\lambda > 0$ and some $c \in (0, \infty)$. Then, for every α in $[1, 2]$ there exists a finite positive constant $B = B_{c,\alpha}$ such that for any martingale f

$$\mathbb{E} \Phi(f^*) \leq B \mathbb{E} \Phi(s(f)) + B \mathbb{E} \Phi(d^*), \quad (29)$$

where $s(f) = (\sum_{j=1}^\infty \mathbb{E}\{|d_j|^2 | \mathcal{F}_{j-1}\})^{1/2}$.

PROOF. Choose $\beta = 3$ and $\delta \in (0, 1)$ such that $B_\alpha \delta^\alpha c^3 < 1$, where B is the constant in (28). Then by Lemmas 1 and 4 there exists $B = B_{c,\alpha} \in (0, \infty)$ such that

$$\mathbb{E} \Phi(f^*) \leq B \mathbb{E} \Phi(s(f) \vee d^*) \leq B \mathbb{E} [\Phi(s(f)) + \Phi(d^*)]. \quad \square$$

Theorem 3. If Φ is a nondecreasing function on $[0, \infty]$, finite and concave on $(0, \infty)$ with $\Phi(0) = 0$, $\Phi(\infty) = \Phi(\infty -)$ and $\alpha \in [1, 2]$, there exists a constant $A = A_\alpha \in (0, \infty)$ such that for any martingale f

$$\mathbb{E} \Phi(f^{*\alpha}) \leq A \mathbb{E} \Phi(s^\alpha(f)), \quad (35)$$

where $s^\alpha(f) = \sum_1^\infty \mathbb{E}\{|d_j|^\alpha | \mathcal{F}_{j-1}\}$.

PROOF. By Theorems 11.2.1 and 11.2.2, for some constant $C = C_\alpha \in (0, \infty)$

$$\mathbb{E}(f^*)^\alpha \leq C \mathbb{E} \left(\sum_1^\infty d_j^2 \right)^{\alpha/2} \leq C \mathbb{E} \sum_1^\infty |d_j|^\alpha = C \mathbb{E} s^\alpha(f). \quad (36)$$

For $u > 0$, define

$$T = \inf \left\{ n \geq 0 : s_{n+1}^\alpha(f) \equiv \sum_1^{n+1} \mathbb{E}\{|d_j|^\alpha | \mathcal{F}_{j-1}\} > u \right\}.$$

Then $s_T^\alpha(f) \leq s^\alpha(f) \wedge u$ and

$$(f^*)^\alpha \wedge u \leq u I_{[T < \infty]} + \sup_{n \geq 1} |f_{T \wedge n}|^\alpha.$$

Hence, via (36),

$$\begin{aligned} \mathbb{E}[(f^*)^\alpha \wedge u] &\leq u \mathbb{P}\{s^\alpha(f) > u\} + C \mathbb{E} s^\alpha(f) \\ &\leq (C + 1) \mathbb{E}[s^\alpha(f) \wedge u]. \end{aligned}$$

Consequently, from Lemma 5 and (36)

$$\begin{aligned} \mathbb{E} \Phi(f^{*\alpha}) &= - \int_0^\infty \mathbb{E}(f^{*\alpha} \wedge u) d\varphi(u) + \varphi(\infty) \mathbb{E} f^{*\alpha} \\ &\leq -(C + 1) \int_0^\infty \mathbb{E}[s^\alpha(f) \wedge u] d\varphi(u) + C \varphi(\infty) \mathbb{E} s^\alpha(f) \\ &\leq (C + 1) \left[- \int_0^\infty \mathbb{E}[s^\alpha(f) \wedge u] d\varphi(u) + \varphi(\infty) \mathbb{E} s^\alpha(f) \right] \\ &= (C + 1) \mathbb{E} \Phi(s^\alpha(f)). \end{aligned} \quad \square$$

Corollary 3. If $0 < p < \alpha$ and $1 \leq \alpha \leq 2$, there exists a constant $A = A_\alpha \in (0, \infty)$ such that, for any martingale f ,

$$\mathbb{E}(f^*)^p \leq A \mathbb{E} \left(\sum_1^\infty \mathbb{E}\{|d_j|^\alpha | \mathcal{F}_{j-1}\} \right)^{p/\alpha}. \quad (37)$$

PROOF. $\Phi(t) = t^{p/\alpha}$, $t \geq 0$, satisfies the requirements of Theorem 3, whence

$$\mathbb{E}(f^*)^p = \mathbb{E} \Phi(f^{*\alpha}) \leq A \mathbb{E} \Phi(s^\alpha(f)) = A \mathbb{E} s^\alpha(f). \quad \square$$

Lemma 7. If $\{X_n, \mathcal{F}_n, n \geq 1\}$ is a stochastic sequence and T is an \mathcal{F}_n -time satisfying

$$\mathbb{E} \sum_{i=1}^T \mathbb{E}\{|X_i|^\gamma | \mathcal{F}_{i-1}\} < \infty, \quad \mathbb{E} \left(\sum_1^T \mathbb{E}\{|X_i| | \mathcal{F}_{i-1}\} \right)^\gamma < \infty \quad (38)$$

for some $\gamma \geq 1$, then

$$\mathbb{E} \left(\sum_1^T |X_i| \right)^\gamma < \infty. \quad (39)$$

PROOF. When $\gamma = 1$, the first two lines of the proof of Theorem 7.4.6 (with X_n^+ replaced by $|X_n|$) yield (39). For $\gamma > 1$, set $V_{\alpha,n} = \sum_{i=1}^n \mathbb{E}\{|X_i|^\alpha | \mathcal{F}_{i-1}\}$, $\alpha > 0$, $n \geq 1$, and $Y_i = |X_i| - \mathbb{E}\{|X_i| | \mathcal{F}_{i-1}\}$, $i \geq 1$, and consider the martingale $f_n = \sum_{i=1}^n Y_i I_{\{T \geq i\}}$, $n \geq 1$. Since $\sum_{i=1}^{\infty} \mathbb{E}\{|Y_i| I_{\{T \geq i\}} | \mathcal{F}_{i-1}\} = \sum_1^T \mathbb{E}\{|Y_i| | \mathcal{F}_{i-1}\} \leq 2V_{1,T}$, it follows, taking $\alpha = 1$, $\Phi(x) = |x|^\gamma$ in Theorem 11.3.3, that

$$\begin{aligned} \mathbb{E} \left| \sum_1^{T \wedge n} Y_i \right|^\gamma &\leq \mathbb{E} \sup_{n \geq 1} \left| \sum_1^n Y_i I_{\{T \geq i\}} \right|^\gamma \\ &\leq B \cdot \mathbb{E} \left[(2V_{1,T})^\gamma + \left(\sup_{i \geq 1} |Y_i| I_{\{T \geq i\}} \right)^\gamma \right] \\ &\leq B \cdot \mathbb{E} \left[(2V_{1,T})^\gamma + \sum_{i=1}^{\infty} |Y_i|^\gamma I_{\{T \geq i\}} \right] \\ &= B \cdot \left[\mathbb{E}(2V_{1,T})^\gamma + \mathbb{E} \sum_{i=1}^{\infty} I_{\{T \geq i\}} \mathbb{E}\{|Y_i|^\gamma | \mathcal{F}_{i-1}\} \right] \\ &\leq 2^{\gamma+1} B [\mathbb{E} V_{1,T}^\gamma + \mathbb{E} V_{\gamma,T}]. \end{aligned}$$

Consequently,

$$\mathbb{E} \left(\sum_1^{T \wedge n} |X_i| \right)^\gamma \leq 2^\gamma \left[\mathbb{E} \left| \sum_1^{T \wedge n} Y_i \right|^\gamma + \mathbb{E} V_{\gamma,T}^\gamma \right] = O(1),$$

and so monotone convergence yields (39). \square

Theorem 4. If $\{S_n = \sum_{i=1}^n X_i, \mathcal{F}_n, n \geq 1\}$ is a martingale and T is an \mathcal{F}_n -time with

$$\mathbb{E} \sum_1^T \mathbb{E}\{|X_i|^\gamma | \mathcal{F}_{i-1}\} < \infty, \quad \mathbb{E} \left(\sum_1^T \mathbb{E}\{X_i^2 | \mathcal{F}_{i-1}\} \right)^{\gamma/2} < \infty \quad (40)$$

for some $\gamma \geq 2$, then

$$\mathbb{E}|S_T|^\gamma < \infty. \quad (41)$$

PROOF. By Theorem 11.2.1,

$$\mathbb{E}|S_{T \wedge n}|^\gamma \leq O(1) \cdot \mathbb{E} \left(\sum_1^{T \wedge n} X_i^2 \right)^{\gamma/2} \leq O(1) \mathbb{E} \left(\sum_1^T X_i^2 \right)^{\gamma/2} = O(1)$$

via Lemma 7 with $|X|$ replaced by X^2 , whence the conclusion follows by Fatou's lemma. \square

Corollary 4. If $S_n = \sum_{j=1}^n X_j$, $n \geq 1$, where $\{X_n, n \geq 1\}$ are independent random variables with $E X_n = 0$, $n \geq 1$, and $\sup_{n \geq 1} E|X_n|^\gamma < \infty$ for some $\gamma \geq 2$, then $E|S_T|^\gamma < \infty$ for any $\{X_n\}$ -time T with $E T^{\gamma/2} < \infty$.

EXERCISE 11.3

1. Prove Corollary 11.2.3 by applying Corollary 11.3.3.

11.4 Stochastic Inequalities

Throughout this section the generic notation

$$\bar{a} = 1 - a \quad (1)$$

will be adopted and the conditional expectation $E\{U|\mathcal{G}\}$ and conditional variance will be denoted $E_{\mathcal{G}} U$ and $\sigma_{\mathcal{G}}^2(U)$ respectively. Integrability requirements in many of the lemmas of this section can clearly be weakened to a.c. finiteness of the conditional expectations therein.

Lemma 1. If \mathcal{G} is a σ -algebra of measurable sets, β is a \mathcal{G} -measurable r.v. at most equal to one a.c., and U is an L_2 random variable satisfying

$$E_{\mathcal{G}} U \leq -\alpha \sigma_{\mathcal{G}}^2(U), \quad P\{\beta + U \leq 1\} = 1, \quad (2)$$

for some $\alpha > 0$, then

$$E_{\mathcal{G}} \frac{\beta + U}{1 + \alpha\beta + U} \leq \frac{\beta}{1 + \alpha\beta}.$$

PROOF.

$$\begin{aligned} \frac{1}{1 + \alpha\beta + U} - \frac{1}{1 + \alpha\bar{\beta}} &= \frac{1}{1 + \alpha\beta + U} \frac{\alpha U}{1 + \alpha\bar{\beta}} \\ &= \frac{\alpha U}{1 + \alpha\bar{\beta}} \left(\frac{1}{1 + \alpha\bar{\beta}} + \frac{1}{1 + \alpha\beta + U} \frac{\alpha U}{1 + \alpha\bar{\beta}} \right) \leq \frac{\alpha U + \alpha^2 U^2}{(1 + \alpha\bar{\beta})^2}, \end{aligned} \quad (3)$$

and so via (2)

$$\begin{aligned} \frac{1}{1 + \alpha\beta + E_{\mathcal{G}} U} - \frac{1}{1 + \alpha\bar{\beta}} &= \frac{1}{1 + \alpha\beta + E_{\mathcal{G}} U} \frac{\alpha E_{\mathcal{G}} U}{1 + \alpha\bar{\beta}} \\ &\leq \frac{-\alpha^2 \sigma_{\mathcal{G}}^2(U)}{(1 + \alpha\bar{\beta} + E_{\mathcal{G}} U)^2}. \end{aligned} \quad (4)$$

it is trivially true since $Y_0 = 0$. Hence, Lemma 2 ensures that $\{V_n/(1 + \alpha\bar{V}_n), \mathcal{F}_n, n \geq 1\}$ is a supermartingale. Consequently, noting that $V_T = 1$, setting $T(n) = \min[T, n]$, and employing Theorem 7.4.5.

$$\begin{aligned} \mathbb{P}\{S_n \geq 1, \text{some } n \geq 1\} &= \mathbb{P}\{V_n = 1, \text{some } n \geq 1\} = \int_{[T < \infty]} \frac{V_T}{1 + \alpha\bar{V}_T} \\ &\leq \liminf_{n \rightarrow \infty} \mathbb{E} \frac{V_{T(n)}}{1 + \alpha\bar{V}_{T(n)}} \leq \mathbb{E} \frac{Y_1}{1 + \alpha\bar{Y}_1}. \end{aligned}$$

Since $X_1 = Y_1$, (7) follows. \square

Theorem 2. If $\{S_n = \sum_1^n X_i, \mathcal{F}_n, n \geq 1\}$ is an \mathcal{L}_2 martingale with $\mathbb{E} S_1 = 0$ and $\mathcal{F}_0 = (\emptyset, \Omega)$, then for any positive constants a, b

$$\mathbb{P}\left\{S_n \geq a \sum_1^n \mathbb{E}\{X_j^2 | \mathcal{F}_{j-1}\} + b, \text{some } n \geq 1\right\} \leq (1 + ab)^{-1}. \quad (9)$$

PROOF. It suffices to prove the theorem when $a = 1$. Define $S_0 = X_0 = 0$, $\mathcal{F}_0 = \mathcal{F}_{-1}$, and for $K > 0$ set

$$U_n = \frac{S_n + K - \sum_{j=0}^n \mathbb{E}_{\mathcal{F}_{j-1}} X_j^2}{K + b}, \quad n \geq 0,$$

whence $u_n = U_n - U_{n-1} = (K + b)^{-1}[X_n - \mathbb{E}_{\mathcal{F}_{n-1}} X_n^2]$, $n \geq 1$, and $u_0 = K(K + b)^{-1} = U_0$. Note that for $n \geq 1$

$$\mathbb{E}_{\mathcal{F}_{n-1}} u_n = \frac{-1}{K + b} \mathbb{E}_{\mathcal{F}_{n-1}} X_n^2$$

and

$$\sigma_{\mathcal{F}_{n-1}}^2(u_n) = (K + b)^{-2} \mathbb{E}_{\mathcal{F}_{n-1}} X_n^2 = -(K + b)^{-1} \mathbb{E}_{\mathcal{F}_{n-1}} u_n, \quad n \geq 1,$$

so that for $\alpha = (K + b)$

$$\mathbb{E}_{\mathcal{F}_{n-1}} u_n = -\alpha \sigma_{\mathcal{F}_{n-1}}^2(u_n). \quad (10)$$

Then, via Theorem 1

$$\begin{aligned} \mathbb{P}\left\{S_n \geq \sum_1^n \mathbb{E}_{j-1} X_j^2 + b, \text{some } n \geq 1\right\} &= \mathbb{P}\{U_n \geq 1, \text{some } n \geq 0\} \\ &\leq \frac{U_0}{1 + \alpha U_0} = \frac{K}{K + b} \frac{1}{1 + b} \xrightarrow{\alpha \rightarrow \infty} \frac{1}{1 + b}. \end{aligned} \quad \square$$

Corollary 1. If $\{X_n, n \geq 1\}$ are independent r.v.s with $\mathbb{E} X_n = 0$, $\mathbb{E} X_n^2 = 1$, $n \geq 1$, and $S_n = \sum_1^n X_i$, then for positive a, b

$$\mathbb{P}\{S_n \geq an + b, \text{some } n \geq 1\} \leq \frac{1}{1 + ab}.$$

Notation. In the remainder of this section, the generic notation

$$a' = e^a - 1 - a \quad (11)$$

will be employed.

Next, analogues of Lemmas 1 and 2 will yield a counterpart of Theorem 1.

Lemma 3. If U, V, Y are r.v.s with $Y \geq 0, |U| \leq 1, E U = 0$, and both Y and V are \mathcal{G} -measurable for some σ -algebra \mathcal{G} of measurable sets, then

$$E_{\mathcal{G}}[\cosh(Y(U + V))] \leq \exp(Y' E_{\mathcal{G}} U^2) \cosh(YV).$$

PROOF. $(YU)' = \sum_{j=2}^{\infty} Y^j U^j / j! \leq U^2 Y'$, whence

$$E_{\mathcal{G}} e^{YU} \leq 1 + Y' E_{\mathcal{G}} U^2 \leq \exp(Y' E_{\mathcal{G}} U^2). \quad (12)$$

Applying (12) to $-U$ also,

$$E_{\mathcal{G}} \cosh(Y(V + U)) = \frac{1}{2} E_{\mathcal{G}}(e^{Y(V+U)} + e^{-Y(V+U)}) \leq \exp(Y' E_{\mathcal{G}} U^2) \cosh(YV).$$

□

Lemma 4. Let $\{S_n = \sum_1^n X_j, \mathcal{F}_n, n \geq 1\}$ be a martingale with $|X_n| \leq 1, n \geq 1$, and set $u_n = E_{\mathcal{F}_{n-1}} X_n^2, n \geq 1$, where $\mathcal{F}_0 = (\emptyset, \Omega)$. For any real y and positive numbers λ, u

$$\left\{ \exp \left\{ - \sum_{j=1}^n u_j \left(\frac{\lambda}{u + \sum_1^j u_i} \right)' \right\} \cosh \lambda \left(\frac{y + S_n}{u + \sum_1^n u_j} \right), \mathcal{F}_n, n \geq 1 \right\}$$

is a supermartingale.

PROOF. Designate the putative supermartingale by $\{V_n, \mathcal{F}_n, n \geq 1\}$ and set $u_0 = u$. By Lemma 3, for $n \geq 2$

$$\begin{aligned} E_{\mathcal{F}_{n-1}} V_n &= \exp \left\{ - \sum_1^n u_j \left(\frac{\lambda}{u + \sum_1^j u_i} \right)' \right\} E_{\mathcal{F}_{n-1}} \cosh \frac{\lambda(y + S_{n-1} + X_n)}{\sum_0^n u_j} \\ &\leq \exp \left\{ - \sum_{j=1}^n u_j \left(\frac{\lambda}{\sum_0^j u_i} \right)' + u_n \left(\frac{\lambda}{\sum_0^n u_j} \right)' \right\} \cosh \frac{\lambda(y + S_{n-1})}{\sum_0^n u_j} \leq v_{n-1}. \end{aligned}$$

Setting $V_0 = \cosh(\lambda y/u)$, the same argument shows that if $E X_1 = 0, \{V_n, \mathcal{F}_n, n \geq 0\}$ is likewise a supermartingale. □

Theorem 3. If $\{S_n = \sum_1^n X_j, \mathcal{F}_n, n \geq 1\}$ is an \mathcal{L}_2 martingale with $E S_1 = 0$ and T is a stopping variable for which $X_n^2 \leq 1$ on $\{T \geq n\}, n \geq 1$, then for any real y and positive numbers λ, u

$$E \cosh \frac{\lambda(y + S_T)}{u + \sum_1^T E_{\mathcal{F}_{j-1}} X_j^2} \leq e^u \left(\frac{\lambda}{u} \right)' \cosh \frac{\lambda y}{u}. \quad (13)$$

$$\leq \frac{1-\alpha}{2} e^\theta + \frac{1+\alpha}{2} e^{-\theta} \leq 1 \quad \text{for } 0 \leq \theta \leq \log \frac{1+\alpha}{1-\alpha}$$

$= \theta_0$ (say). Consequently, for $n \geq 1$

$$E_{\mathcal{F}_{n-1}} e^{\theta_0 S_n} = e^{\theta_0 S_{n-1}} E_{\mathcal{F}_{n-1}} e^{\theta_0 X_n} \leq e^{\theta_0 S_{n-1}},$$

whence $\{e^{\theta_0 S_n}, \mathcal{F}_n, n \geq 0\}$ is a supermartingale. Hence, for any stopping rule T , setting $T(n) = \min[T, n]$,

$$E e^{\theta_0 S_T} \leq \liminf_n E e^{\theta_0 S_{T(n)}} \leq E e^{\theta_0 S_0} = 1.$$

Finally, if $T = \inf\{n \geq 1 : S_n \geq \lambda > 0\}$,

$$e^{\theta_0 \lambda} P\{S_n \geq \lambda \text{ for some } n \geq 1\} = e^{\theta_0 \lambda} P\{T < \infty\} \leq E e^{\theta_0 S_T} \leq 1,$$

which is tantamount to (16). \square

Suppose that a gambler with initial capital $f_0 \in (0, 1)$ wins or loses (at each play) whatever amount he stakes with probabilities p and q respectively, where $0 < p \leq \frac{1}{2}$, the successive plays being independent. What is the best gambling scheme for boosting his capital to one dollar? It will be shown that a bold strategy is optimal, namely, whenever his capital $f_n \leq \frac{1}{2}$, the entire capital should be staked, whereas whenever $f_n > \frac{1}{2}$, only the amount $1 - f_n$ (needed to obtain his goal) should be bet.

Let $\{X_n, n \geq 1\}$ be i.i.d. random variables with

$$P\{X_n = 1\} = 1 - P\{X_n = -1\} = p \leq \frac{1}{2}$$

and let $\bar{a} = \min(a, 1 - a)$ whenever $0 \leq a \leq 1$.

For any constant $0 \leq f = f_0 \leq 1$, define for $n \geq 1$,

$$f_n = f_{n-1} + \tilde{f}_{n-1} X_n = \begin{cases} f_{n-1}(1 + X_n) & \text{if } 0 \leq f_{n-1} \leq \frac{1}{2} \\ f_{n-1}(1 - X_n) + X_n & \text{if } \frac{1}{2} \leq f_{n-1} \leq 1, \end{cases} \quad (17)$$

and note that $\{f_n, n \geq 1\}$ represent the successive (random) fortunes associated with the bold strategy just described. Set

$$\begin{aligned} p_n(f) &= P\left\{\max_{0 \leq j \leq n} f_j \geq 1\right\}, \quad 0 \leq f \leq 1, n = 0, 1, \dots, \\ p_n(f) &= 1 \quad \text{if } f \geq 1, \quad p_n(f) = 0 \quad \text{if } f \leq 0. \end{aligned} \quad (18)$$

Lemma 5. For any constants $0 \leq f - s \leq f$, setting $q = 1 - p$,

$$p_{n+1}(f) \geq pp_n(f+s) + qp_n(f-s), \quad n = 0, 1, \dots \quad (19)$$

PROOF. For every n , by (17) $p_n(f) \uparrow$ in f . If $A_n(f) = \{\max_{0 \leq j \leq n} f_j \geq 1\}$, then

$$\begin{aligned} p_{n+1}(\frac{1}{2}f) &= p P\{A_{n+1}(\frac{1}{2}f) | X_1 = 1\} + q P\{A_{n+1}(\frac{1}{2}f) | X_1 = -1\} \\ &= p P\{A_n(f)\} = pp_n(f) \end{aligned}$$

and

$$\begin{aligned} p_{n+1}\left(\frac{1+f}{2}\right) &= p \mathbb{P}\left\{A_{n+1}\left(\frac{1+f}{2}\right) \mid X_1 = 1\right\} \\ &\quad + q \mathbb{P}\left\{A_{n+1}\left(\frac{1+f}{2}\right) \mid X_1 = -1\right\} \\ &= p + q \mathbb{P}\{A_n(f)\} = p + qp_n(f). \end{aligned}$$

Hence, for $0 \leq f \leq 1$ and $n \geq 0$

$$p_{n+1}(f) = pp_n(2f) + qp_n(2f - 1), \quad (20)$$

and obviously (20) holds for $f \leq 0$ or $f \geq 1$. Consequently, (20) obtains for all $f \in (-\infty, \infty)$ and all $n \geq 0$.

To prove (19), define for $0 \leq f - s \leq f$ and $n \geq 0$

$$\Delta_n(f, s) = p_{n+1}(f) - pp_n(f + s) - qp_n(f - s). \quad (21)$$

If $f + s < 1$, then $\Delta_0(f, s) = p_1(f) \geq 0$, while if $f + s \geq 1$, monotonicity of p_n and (20) ensure $\Delta_0(f, s) \geq 0$. Hence, for $0 \leq f - s \leq f$ and $n = 0$

$$\Delta_n(f, s) \geq 0. \quad (22)$$

Now from (20) and (21)

$$\Delta_n(f, s) = p[p_n(2f) - p_n(f + s)] + q[p_n(2f - 1) - p_n(f - s)],$$

and employing (20) once more

$$\begin{aligned} \Delta_{n+1}(f, s) &= p^2[p_n(4f) - p_n(2f + 2s)] + q^2[p_n(4f - 3) - p_n(2f - 2s - 1)] \\ &\quad + pq[p_n(4f - 1) - p_n(2f + 2s - 1) \\ &\quad + p_n(4f - 2) - p_n(2f - 2s)], \end{aligned} \quad (23)$$

and so

$$\begin{aligned} \Delta_{n+1}(f, s) &= p\Delta_n(2f, 2s) + q\{p[p_n(4f - 2) - p_n(2f + 2s - 1)] \\ &\quad + q[p_n(4f - 3) - p_n(2f - 2s - 1)]\}. \end{aligned} \quad (24)$$

To verify (22) for $n \geq 1$, suppose inductively that it holds for a fixed but arbitrary integer n . If (i) $f - s \geq \frac{1}{2}$, then $2f - 1 \geq 2s$, whence by (24) $\Delta_{n+1}(f, s) = p\Delta_n(2f, 2s) + q\Delta_n(2f - 1, 2s) \geq 0$; if (ii) $f + s \leq \frac{1}{2}$, then $2f - 2s \leq 2f + 2s \leq 1$ and again via (24), $\Delta_{n+1}(f, s) \geq p\Delta_n(2f, 2s) \geq 0$; finally, if (iii) $0 \leq f - s < \frac{1}{2} < f + s$, then $\frac{1}{4} < f$, and by (23)

$$\Delta_{n+1}(f, s) \geq pq[p_n(4f - 1) - p_n(2f + 2s - 1) + p_n(4f - 2) - p_n(2f - 2s)]. \quad (25)$$

Since $q \geq p$, if $s > \frac{1}{4}$, then from (25)

$$\Delta_{n+1}(f, s) \geq p\{[p[p_n(4f - 1) - p_n(2f + 2s - 1)]$$

$$\begin{aligned} & + q[p_n(4f - 2) - p_n(2f - 2s)]\} \\ & = p\Delta_n(2f - \frac{1}{2}, 2s - \frac{1}{2}) \geq 0, \end{aligned}$$

and if $s \leq \frac{1}{4}$, again via (25)

$$\begin{aligned} \Delta_{n+1}(f, s) & \geq p\{p[p_n(4f - 1) - p_n(2f - 2s)] \\ & + q[p_n(4f - 2) - p_n(2f + 2s - 1)]\} \\ & = p\Delta_n(2f - \frac{1}{2}, \frac{1}{2} - 2s) \geq 0, \end{aligned}$$

completing the induction. \square

Lemma 6. Let $\{X_n, \mathcal{F}_n, 1 \leq n \leq N < \infty\}$ be an \mathcal{L}_1 stochastic sequence, let C_N be the class of all stopping rules T with $P\{T \leq N\} = 1$ and define

$$\begin{aligned} \gamma_N & = \gamma_N^N = X_N, \\ \gamma_n & = \gamma_n^N = \max [X_n, E\{\gamma_{n+1} | \mathcal{F}_n\}], \quad 1 \leq n < N. \end{aligned} \tag{26}$$

Then, if $\sigma = \inf\{n \geq 1 : X_n = \gamma_n\}$,

$$\sup_{T \in C_N} E X_T = E \gamma_1 = E X_\sigma. \tag{27}$$

PROOF. Clearly, $\{\gamma_n, \mathcal{F}_n, 1 \leq n \leq N\}$ is a supermartingale. Thus, for any $T \in C_N$,

$$E X_T = \sum_{j=1}^N \int_{[T=j]} X_j \leq \sum_{j=1}^N \int_{[T=j]} \gamma_j = E \gamma_T \leq E \gamma_1.$$

Moreover,

$$\begin{aligned} E \gamma_1 & = \int_{\{\sigma=1\}} \gamma_1 + \int_{\{\sigma>1\}} \gamma_1 = \int_{\{\sigma=1\}} \gamma_1 + \int_{\{\sigma>1\}} E\{\gamma_2 | \mathcal{F}_1\} \\ & = \int_{\{\sigma=1\}} \gamma_1 + \int_{\{\sigma>1\}} \gamma_2 = \int_{\{\sigma=1\}} \gamma_1 + \int_{\{\sigma=2\}} \gamma_2 + \int_{\{\sigma>2\}} \gamma_2 \\ & = \dots = \sum_{j=1}^N \int_{\{\sigma=j\}} \gamma_j = E \gamma_\sigma = E X_\sigma, \end{aligned}$$

and (27) follows. \square

Lemma 7. For any random variables Y_1, \dots, Y_N setting $X_n = I_{\{Y_n \geq 1\}}$ and $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$, $1 \leq n \leq N$,

$$P\left\{\max_{1 \leq n \leq N} Y_n \geq 1\right\} = E \gamma_1 \tag{28}$$

where γ_1 is as in (26).

$$= p \cdot p_{N-n-1}(h_n + h_n g_n) + q \cdot p_{N-n-1}(h_n - h_n g_n) \\ \leq p_{N-n}(h_n),$$

recalling Lemma 5. This completes the (backward) induction, whence for $0 \leq n \leq N$

$$d_n \leq p_{N-n}(h_n),$$

and, in particular, recalling (33),

$$\mathbf{P} \left\{ \max_{0 \leq j \leq N} h_j \geq 1 \right\} = \mathbf{E} d_0 \leq p_N(h_0) = p_N(f). \quad \square$$

References

- D. G. Austin, "A sample property of martingales," *Ann. Math. Stat.* **37** (1966), 1396–1397.
- D. Blackwell, "On optimal systems," *Ann. Math. Stat.* **25** (1954), 394–397.
- B. M. Brown, "A note on convergence of moments," *Ann. Math. Stat.* **42** (1971), 777–779.
- D. L. Burkholder, "Martingale transforms," *Ann. Math. Stat.* **37** (1966), 1494–1504.
- D. L. Burkholder, "Distribution function inequalities for martingales," *Ann. Probability* **1** (1973), 19–42.
- D. L. Burkholder and R. F. Gundy, "Extrapolation and interpolation of quasi-linear operators on martingales," *Acta Math.* **124** (1970), 249–304.
- D. L. Burkholder, B. J. Davis, and R. F. Gundy, "Inequalities for convex functions of operators on martingales," *Proc. Sixth Berkeley Symp. Math. Stat. Prob.* **2** (1972), 223–240.
- Y. S. Chow, "On a strong law of large numbers for martingales," *Ann. Math. Stat.* **38** (1967), 610.
- Y. S. Chow, "Convergence of sums of squares of martingale differences," *Ann. Math. Stat.* **39** (1968), 123–133.
- Y. S. Chow, "On the L_p -convergence for $n^{-1/p}S_n$, $0 < p < 2$," *Ann. Math. Stat.* **42** (1971), 393–394.
- K. L. Chung, *A Course in Probability Theory*, Harcourt Brace, New York, 1968; 2nd ed., Academic Press, New York, 1974.
- B. Davis, "A comparison test for martingale inequalities," *Ann. Math. Stat.* **40** (1969), 505–508.
- J. L. Doob, *Stochastic Processes*, Wiley, New York, 1953.
- L. E. Dubins and D. A. Freedman, "A sharper form of the Borel–Cantelli lemma and the strong law," *Ann. Math. Stat.* **36** (1965), 800–807.
- L. E. Dubins and L. J. Savage, *How to Gamble If You Must*, McGraw–Hill, New York, 1965.
- A. M. Garsia, "On a convex function inequality for martingales," *Ann. Probability* **1** (1973), 171–174.
- R. F. Gundy, "A decomposition for \mathcal{L}_1 -bounded martingales," *Ann. Math. Stat.* **39** (1968), 134–138.
- J. Neveu, *Martingales à temps discrets*, Masson, Paris, 1972.

- R. Panzone, "Alternative proofs for certain upcrossing inequalities," *Ann. Math. Stat.* **38** (1967), 735-741.
- E. M. Stein, *Topics in Harmonic Analysis Related to the Littlewood-Paley Theory*, Princeton Univ. Press, Princeton, 1970.
- H. Teicher, "Moments of randomly stopped sums-revisited," *Jour. Theor. Prob.* **9** (1995), 779-793.
- A. Zygmund, *Trigonometric Series*, Vol. I, Cambridge, 1959.

12 Infinitely Divisible Laws

Row sums $\sum_{i=1}^{k_n} X_{ni}$ of arrays of random variables $\{X_{ni}, 1 \leq i \leq k_n \rightarrow \infty, n \geq 1\}$ that are rowwise independent have been considered briefly with respect to the Marcinkiewicz–Zygmund type strong laws of large numbers (Example 10.4.1). In this same context, non-Iterated Logarithm laws and generalizations thereof have been dealt with by H. Cramér and C. Esseen (see references at the end of this chapter). Here, limit distributions of row sums of the variables in such an array will be treated.

It is a remarkable fact that the class of limit distributions of normed sums of i.i.d. random variables is severely circumscribed. If the underlying r.v.s, say $\{X_n, n \geq 1\}$ have merely absolute moments of order r , then for $r \geq 2$ only the normal distribution can arise as a limit, while if $0 < r \leq 2$, the limit law belongs to a class called stable distributions. If the basic r.v.s are merely independent (and infinitesimal when normed cf. (1) of Section 2), a larger class of limit laws, the so-called class \mathcal{L} emerges. But even the class \mathcal{L} does not contain a distribution of such crucial importance as the Poisson. A perusal of the derivation (Chapter 2) of the Poisson law as a limit of binomial laws B_n reveals that the success probability associated with B_n is a function of n . Thus, if B_{n-1} is envisaged as the distribution of the sum of i.i.d. random variables Y_1, \dots, Y_{n-1} , then B_n must be the distribution of the sum of n different i.i.d. random variables which, therefore, may as well be labeled $X_{n,1}, X_{n,2}, \dots, X_{n,n}$. In other words, to obtain the Poisson law as a limit of distributions of sums of i.i.d. (or even independent) random variables, a double sequence schema $\{X_{nj}, j = 1, \dots, k_{n \rightarrow \infty}\}$ must be employed (with $X_{n,1}, \dots, X_{n,k_n}$ independent within rows for each $n = 1, 2, \dots$). Under one further proviso, the class of limit laws of (row) sums of such r.v.s coincides with the class of infinitely divisible laws.

12.1 Infinitely Divisible Characteristic Functions

It should be borne in mind that the notion of infinite divisibility as presented below is a distribution concept requiring no mention or consideration of r.v.s. In fact, an attempt to define it via r.v.s can lead to unnecessary difficulty and complication (see Gnedenko and Kolmogorov (1954)).

Definition. A d.f. F is called **infinitely divisible** (i.d.) if for every integer $n \geq 1$ there exists a d.f. F_n such that $F = F_n * F_n * \cdots * F_n = (F_n)^n$ or equivalently if its c.f. φ (also called i.d.) is the n th power of a c.f. φ_n for every integer $n \geq 1$.

Clearly, the normal, Poisson, and degenerate distributions are i.d. Moreover, if $F(x)$ is i.d., all distributions of the same type $F(ax + b)$ are i.d.

A number of useful facts about i.d. distributions and c.f.s will be amassed in the propositions which follow.

Proposition 1. *A d.f. F with bounded support is i.d. iff it is degenerate.*

PROOF. Although the proof can be couched solely in terms of d.f.s, it is more easily intuited, and hence will be presented, in terms of r.v.s. Thus, if X is a r.v. with d.f. F , the hypothesis implies that with probability one, $|X| \leq C < \infty$. Without loss of generality suppose $E X = 0$. Then, since $F = (F_n)^n$, $n \geq 1$, if $\{X_{ni}, 1 \leq i \leq n\}$ are (fictitious) i.i.d. random variables with d.f. F_n , necessarily $E X_{n1} = 0$ and $|X_{n1}| \leq C/n$ with probability one. Consequently, if σ^2 denotes the variance of X , $0 \leq \sigma^2 = \sum_{i=1}^n E(X_{ni} - E X_{n1})^2 = n E X_{n1}^2 \leq n(C/n)^2 = o(1)$, whence $\sigma^2 = 0$, implying $P\{X = E X\} = 1$, i.e., F degenerate. \square

Proposition 2. *An i.d. c.f. $\varphi(t)$ does not vanish (for real t).*

PROOF. By hypothesis, $\varphi = \varphi_n^n$, $n \geq 1$, with φ_n a c.f. Then $\psi = |\varphi|^2$ and $\psi_n = |\varphi_n|^2$ are real-valued c.f.s and the positive real function ψ has a unique real, positive n th root, say $\psi^{1/n}$, $n \geq 1$. Since necessarily $\psi = \psi_n^n$, $n \geq 1$, the positive real function ψ_n must coincide with $\psi^{1/n}$. Thus, $0 \leq \psi \leq 1$ implies $\Gamma(t) = \lim_{n \rightarrow \infty} \psi_n(t) = 1$ or 0 according as $\psi(t) > 0$ or $\psi(t) = 0$. Then, $\psi(0) = 1$ and continuity of ψ imply $\Gamma(t) = 1$ throughout some neighborhood of $t = 0$. Theorem 8.3.3 ensures that $\Gamma(t)$ is a c.f., whence continuity dictates that Γ is nonvanishing. Hence, ψ and therefore also φ is nonvanishing. \square

If φ is an i.d. c.f. so that $\varphi = \varphi_n^n$, $n \geq 1$, it seems plausible that $\varphi_n = \varphi^{1/n}$. But how can one choose a continuous version of $\varphi^{1/n}$? The following lemma asserts that a continuous logarithm (hence also n th root) of a continuous, nonvanishing complex function on $(-\infty, \infty)$ can be defined.

Lemma 1. *If $f(t)$ is a continuous, nonvanishing complex function on $[-T, T]$ with $f(0) = 1$, there is a unique (single-valued) continuous function $\lambda(t)$ on*

$[-T, T]$ with $\lambda(0) = 0$ and $f(t) = e^{\lambda(t)}$. Moreover, $[-T, T]$ is replaceable by $(-\infty, \infty)$.

PROOF. If $\rho_T = \inf_{[-T, T]} |f(t)|$, then $0 < \rho_T \leq 1$. Since f is uniformly continuous on $[-T, T]$, there exists δ_T in $(0, \rho_T)$ such that $|t' - t| \leq \delta_T$ implies $|f(t') - f(t)| < \rho_T/2 \leq \frac{1}{2}$. Choose points $\{t_j\}$ with $t_0 = 0$ such that $-T = t_{-m} < \dots < t_{-1} < t_0 < t_1 < \dots < t_m = T$ and $t_{j+1} - t_j = t_1 - t_0 \leq \delta_T$. Define

$$L(z) = \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{j} (z-1)^j, \quad |z-1| < 1.$$

Then $L(z)$ is the unique determination (principal value) of $\log z$ vanishing at $z = 1$. For $t \in [t_{-1}, t_1]$, $|f(t) - 1| = |f(t) - f(t_0)| \leq \frac{1}{2}$, and so $L(f(t))$ is defined. Set

$$\lambda(t) = L(f(t)), \quad t \in [t_{-1}, t_1].$$

Then $\lambda(0) = L(1) = 0$ and $\lambda(t)$ is continuous with $\exp\{\lambda(t)\} = f(t)$ in $[t_{-1}, t_1]$. Since for $t \in [t_k, t_{k+1}]$, $|f(t)/f(t_k) - 1| \leq (\rho_T/2)/\rho_T = \frac{1}{2}$, for any k the definition of λ may be extended from $[t_{-k}, t_k]$ to $[t_k, t_{k+1}]$ by $\lambda(t) = \lambda(t_k) + L((f(t)/f(t_k)))$; analogously, replacing t_k by t_{-k} , the definition extends to $[t_{-k-1}, t_{-k}]$. Then $\lambda(t)$ is defined and continuous in $[-T, T]$, and for $t \in [t_k, t_{k+1}]$, $k \geq 1$,

$$\begin{aligned} e^{\lambda(t)} &= \exp\left(L\left(\frac{f(t)}{f(t_k)}\right) + \lambda(t_k)\right) = \exp\left(L\left(\frac{f(t)}{f(t_k)}\right)\right. \\ &\quad \left.+ \sum_{j=0}^{k-1} L\left(\frac{f(t_{j+1})}{f(t_j)}\right)\right) = f(t). \end{aligned}$$

A similar statement holds in $[t_{-k-1}, t_{-k}]$. Next, given λ in $[-T, T]$, it may be extended by the prior method to $[-T-1, T+1]$, and hence by induction to $(-\infty, \infty)$. Finally, if two such functions λ and λ' exist, $e^{\lambda(t)} = e^{\lambda'(t)}$, whence $\lambda(t) - \lambda'(t) = 2\pi i k(t)$ with $k(t)$ an integer. Since $k(t)$ is continuous with $k(0) = 0$, necessarily $k(t) = 0$ and λ is unique.

Definition. The function $\lambda(t)$ defined by Lemma 1 is called the **distinguished logarithm** of $f(t)$ and is denoted by $\text{Log } f(t)$. Also $\exp\{(1/n)\lambda(t)\}$ is called the **distinguished n th root** of $f(t)$ and is denoted by $f^{1/n}(t)$.

Note. Clearly, if $\psi(t)$ is a continuous complex function on $(-\infty, \infty)$ with $\psi(0) = 0$, then $\text{Log } e^{\psi(t)} = \psi(t)$. Moreover, for f, g as in Lemma 1, $\text{Log } f \cdot g = \text{Log } f + \text{Log } g$, $\text{Log}(f/g) = \text{Log } f - \text{Log } g$, and $\text{Log } f = L(f)$ for $|t| \leq T$ whenever $\sup_{|t| \leq T} |f(t) - 1| < 1$. Thus, for k an integer, $\text{Log}(e^{ait+2k\pi i}) = ait$ and

$$(e^{ait+2k\pi i})^{1/n} = e^{ait/n}.$$

Lemma 2. Let f, f_k , $k \geq 1$ be as in Lemma 1. If $f_k \rightarrow f$ uniformly in $[-T, T]$, then $\text{Log } f_k - \text{Log } f \rightarrow 0$ uniformly in $[-T, T]$.

PROOF. Since $\min_{|t| \leq T} |f(t)| > 0$ and $f_k \rightarrow f$ uniformly in $[-T, T]$, $\sup_{|t| \leq T} |(f_k(t)/f(t)) - 1| \leq \frac{1}{2}$ for $k \geq K_0$. Then

$$\text{Log } f_k - \text{Log } f = \text{Log } \frac{f_k}{f} = L\left(\frac{f_k(t)}{f(t)}\right) \rightarrow L(1) = 0, \text{ uniformly in } [-T, T].$$

□

Proposition 3. A c.f. $\varphi(t)$ is i.d. iff its distinguished n th root $\varphi^{1/n}(t) = e^{(1/n)\text{Log } \varphi(t)}$ is a c.f. for every positive integer.

PROOF. If φ is i.d., $\varphi = \varphi_n^n$, $n \geq 1$, and so by Proposition 2, φ and hence also φ_n is nonvanishing, whence their distinguished n th roots and logarithms are well defined by Lemma 1. Moreover, $e^{\text{Log } \varphi} = \varphi = \varphi_n^n = e^{n \text{Log } \varphi_n}$, so that $\text{Log } \varphi(t) = n \text{Log } \varphi_n(t) + 2\pi i k(t)$ with $k(t)$ an integer. Since $\text{Log } \varphi$ and $\text{Log } \varphi_n$ are continuous and vanish at zero, $k(t) = 0$, whence $\text{Log } \varphi_n = (1/n)\text{Log } \varphi$, which is tantamount to $\varphi_n = \varphi^{1/n}$.

Conversely if the distinguished n th root of φ exists and is a c.f. for every $n \geq 1$, $\varphi = e^{\text{Log } \varphi} = (e^{(1/n)\text{Log } \varphi})^n$ shows that φ is i.d. □

Proposition 4. A finite product of i.d. c.f.s is i.d. Moreover, if i.d. c.f.s $\varphi_k \rightarrow \varphi$, a c.f., then this limit c.f. φ is i.d.

PROOF. Clearly, if $\varphi = \varphi_n^n$, $\psi = \psi_n^n$, $n \geq 1$, then $\varphi \cdot \psi = [\varphi_n \cdot \psi_n]^n$, $n \geq 1$, shows that a product of two and hence any finite number of i.d. c.f.s is i.d. Suppose next that the i.d. c.f.s $\varphi_k \rightarrow \varphi$, a c.f. Then, the i.d. c.f.s $\psi_k(t) = |\varphi_k(t)|^2 = \varphi_k(t) \cdot \varphi_k(-t) \rightarrow$ the c.f. $\psi(t) = |\varphi(t)|^2$. Consequently, $\psi_k^{1/n}$ as the positive n th root of the positive function ψ_k tends as $k \rightarrow \infty$ to the nonnegative n th root $\psi^{1/n}$ of the nonnegative function ψ . Since for $n \geq 1$, $\psi_k^{1/n}$ is a sequence of c.f.s whose limit $\psi^{1/n}$ is continuous at $t = 0$, $\psi^{1/n}$ is a c.f. for $n \geq 1$. Thus, ψ is i.d. and hence nonvanishing. Consequently, φ is nonvanishing, whence $\text{Log } \varphi$ is defined by Lemma 1. By Lemma 2,

$$\varphi_k^{1/n} - \varphi^{1/n} = \varphi^{1/n}[e^{(1/n)(\text{Log } \varphi_k - \text{Log } \varphi)} - 1] \rightarrow 0, n \geq 1,$$

as $k \rightarrow \infty$, and since $\varphi^{1/n}$ is continuous at $t = 0$, it is a c.f. for all $n \geq 1$, so that φ is i.d. by Proposition 3. □

Since c.f.s $\exp\{\lambda(e^{itw} - 1) + it\theta\}$, $\lambda > 0$, of Poisson type are i.d., it follows from Proposition 4 that $\exp\{\sum_{j=1}^k [\lambda_j(e^{itw_j} - 1) + it\theta_j]\}$ and hence also $\exp\{it\theta + \int_{-\infty}^{\infty} (e^{itw} - 1)dG(u)\}$ with G a bounded, increasing function is i.d. The latter comes close but does not quite exhaust the class of i.d. c.f.s.

Proposition 5. The class of i.d. laws coincides with the class of distribution limits of finite convolutions of distributions of Poisson type.

PROOF. That every such limit is i.d. follows directly from Proposition 4. Conversely, if φ is i.d., so that $\varphi = \varphi_n^n$, $n \geq 1$, then

$$n[\varphi_n(t) - 1] = n[e^{(1/n)\text{Log } \varphi} - 1] \rightarrow \text{Log } \varphi,$$

that is

$$\lim_{n \rightarrow \infty} \exp\{n[\varphi_n(t) - 1]\} = \varphi(t).$$

Now,

$$n[\varphi_n(t) - 1] = \int_{-\infty}^{\infty} n(e^{itu} - 1)dF_n(u), \quad n \geq 1,$$

and a net $-\infty < -M_n = u_{n,1} < u_{n,2} < \dots < u_{n,k_n+1} = M_n < \infty$ may be chosen whose points are continuity points of F_n and such that

$$\max_j (u_{n,j+1} - u_{n,j}) \leq \frac{1}{2n^3}$$

and

$$\int_{\{|u| \geq M_n\}} dF_n(u) \leq \frac{1}{4n^2}.$$

Then for $|t| \leq n$, choosing $\lambda_{n,j} = n[F_n(u_{n,j+1}) - F_n(u_{n,j})]$,

$$\begin{aligned} & \left| \int_{u_{n,j}}^{u_{n,j+1}} (e^{itu} - 1)n dF_n(u) - \lambda_{n,j}(e^{itu_{n,j}} - 1) \right| \\ &= \left| \int_{u_{n,j}}^{u_{n,j+1}} (e^{itu} - e^{itu_{n,j}})n dF_n(u) \right| \\ &\leq \int_{u_{n,j}}^{u_{n,j+1}} |1 - e^{it(u_{n,j} - u)}|n dF_n(u) \leq \int_{u_{n,j}}^{u_{n,j+1}} |t(u_{n,j} - u)|n dF_n(u) \\ &\leq \frac{1}{2n} \int_{u_{n,j}}^{u_{n,j+1}} dF_n(u). \end{aligned}$$

Hence, for $|t| \leq n$, summing over $1 \leq j \leq k_n$,

$$\begin{aligned} & \left| \int_{-\infty}^{\infty} (e^{itu} - 1)n dF_n(u) - \sum_{j=1}^{k_n} \lambda_{n,j}(e^{itu_{n,j}} - 1) \right| \\ &\leq 2n \int_{\{|u| \geq M_n\}} dF_n(u) + \sum_{j=1}^{k_n} \frac{1}{2n} \int_{u_{n,j}}^{u_{n,j+1}} dF_n(u) \leq \frac{1}{n}. \end{aligned}$$

Consequently, for $|t| \leq n$ and sufficiently large n

$$\begin{aligned} & \left| e^{n(\varphi_n(t) - 1)} - \prod_{j=1}^{k_n} \exp(\lambda_{n,j}(e^{itu_{n,j}} - 1)) \right| \\ &= \left| \exp\left(\sum_1^{k_n} \lambda_{n,j}(e^{itu_{n,j}} - 1)\right) - \exp(n(\varphi_n(t) - 1)) \right| \\ &= |e^{n(\varphi_n(t) - 1)}| \left| \exp\left(\sum_1^{k_n} \lambda_{n,j}(e^{itu_{n,j}} - 1) - \int_{-\infty}^{\infty} (e^{itu} - 1)n dF_n(u)\right) - 1 \right| \\ &\leq 2|\varphi(t)|2(e^{1/n} - 1) = o(1), \end{aligned}$$

recalling Corollary 8.4.1. Consequently, for all real t

$$\varphi(t) = \lim_{n \rightarrow \infty} e^{i\psi_n(t) - 1} = \lim_{n \rightarrow \infty} \prod_{j=1}^{k_n} \exp(\lambda_{n,j}(e^{itu_{n,j}} - 1)). \quad \square$$

For any real γ and nondecreasing, left-continuous $G(u)$ with $G(-\infty) = 0$, $G(\infty) < \infty$, set

$$\psi(t) = \psi(t; \gamma, G) = i\gamma t + \int_{-\infty}^{\infty} \left(e^{itu} - 1 - \frac{itu}{1+u^2} \right) \left(\frac{1+u^2}{u^2} \right) dG(u), \quad (1)$$

where the integrand, say $h(t, u)$, is defined at $u = 0$ by continuity (whence $h(t, 0) = -t^2/2$). Since $e^{itu} - 1 = O(1)$ as $|u| \rightarrow \infty$ and $e^{itu} - 1 = itu + O(u^2)$ as $u \rightarrow 0$, dominated convergence ensures that $\psi(t)$ is continuous, and, clearly, $\psi(0) = 0$ and $\psi(t) = \text{Log } e^{\psi(t)}$.

Theorem 1. $\varphi(t) = \exp\{\psi(t; \gamma, G)\}$ as defined by (1) is an i.d. c.f. for every real γ and G as stipulated. Moreover, φ uniquely determines γ and G .

PROOF. The integrand $h(t, u)$ of ψ satisfies $|h(t, u)| \leq C < \infty$ for $|t| \leq T$ and all real u . Choose $-M_n = u_{n,1} < u_{n,2} < \dots < u_{n,k_n+1} = M_n$ to be nonzero continuity points of G for which

$$\int_{\{|u| \geq M_n\}} dG(u) \leq \frac{1}{2Cn}, \quad \max_{\substack{u_{n,j} \leq u \leq u_{n,j+1} \\ 1 \leq j \leq k_n, |t| \leq T}} |h(t, u_{n,j}) - h(t, u)| \leq \frac{1}{2nG(\infty)}.$$

Then, for $|t| \leq T$

$$\begin{aligned} & \left| \int_{-\infty}^{\infty} h(t, u) dG(u) - \sum_{j=1}^{k_n} h(t, u_{n,j}) [G(u_{n,j+1}) - G(u_{n,j})] \right| \\ & \leq C \int_{\{|u| \geq M_n\}} dG(u) + \sum_{j=1}^{k_n} \frac{G(u_{n,j+1}) - G(u_{n,j})}{2nG(\infty)} \leq \frac{1}{n}. \end{aligned}$$

Thus, setting $\lambda_{n,j} = [(1+u_{n,j}^2)/u_{n,j}^2][G(u_{n,j+1}) - G(u_{n,j})]$ and

$$\begin{aligned} a_{n,j} &= \frac{\gamma}{k_n} - \frac{G(u_{n,j+1}) - G(u_{n,j})}{u_{n,j}}, \\ e^{\psi(t)} &= \lim_{n \rightarrow \infty} \exp \left\{ i\gamma t + \sum_{j=1}^{k_n} h(t, u_{n,j}) [G(u_{n,j+1}) - G(u_{n,j})] \right\} \\ &= \lim_{n \rightarrow \infty} \prod_{j=1}^{k_n} \exp(it a_{n,j} + \lambda_{n,j}(e^{itu_{n,j}} - 1)), \end{aligned}$$

and so $\varphi(t)$ is a c.f. and i.d. by Proposition 4.

Apropos of uniqueness, define

$$\begin{aligned} -V(t) &= \int_{t-1}^{t+1} \psi(w) dw - 2\psi(t) = \int_{t-1}^{t+1} \int_{-\infty}^{\infty} h(w, u) dG(u) dw + 2iyt - 2\psi(t) \\ &= \int_{-\infty}^{\infty} \left[e^{itu} \left(\frac{e^{iu} - e^{-iu}}{iu} \right) - 2 - \frac{2itu}{1+u^2} \right] \left(\frac{1+u^2}{u^2} \right) dG(u) \\ &\quad - 2 \int_{-\infty}^{\infty} h(t, u) dG(u) \\ &= -2 \int_{-\infty}^{\infty} e^{itu} \left(1 - \frac{\sin u}{u} \right) \left(\frac{1+u^2}{u^2} \right) dG(u) = - \int_{-\infty}^{\infty} e^{itu} dH(u), \end{aligned}$$

where

$$H(u) = 2 \int_{-\infty}^u \left(1 - \frac{\sin x}{x} \right) \left(\frac{1+x^2}{x^2} \right) dG(x).$$

Clearly, H is nondecreasing and left continuous with $H(-\infty) = 0$, $H(\infty) = C_1 < \infty$. Thus, ψ determines V , which, in turn, determines H (by Theorem 8.3.1), which by Theorem 6.5.2 determines G (hence, also γ). \square

Let $\{G, G_n, n \geq 1\}$ be nondecreasing, left-continuous functions of bounded variation with $G(-\infty) = G_n(-\infty) = 0$, $n \geq 1$. Recall as in Section 8.1 that $G_n \xrightarrow{s} G$ iff $G_n \xrightarrow{s} G$ and $G_n(\infty) \rightarrow G(\infty)$, $G_n(-\infty) \rightarrow G(-\infty)$.

Theorem 2. Let $\{\gamma, \gamma_n, n \geq 1\}$ be finite real numbers and $\{G, G_n, n \geq 1\}$ nondecreasing left-continuous functions of bounded variation which vanish at $-\infty$. If $\gamma_n \rightarrow \gamma$ and $G_n \xrightarrow{s} G$, then $\psi(t, \gamma_n, G_n) \rightarrow \psi(t, \gamma, G)$ for all real t , where ψ is as in (1). Conversely, if $\psi(t; \gamma_n, G_n)$ tends to a continuous function $g(t)$ as $n \rightarrow \infty$, then necessarily $g(t) = \psi(t, \gamma, G)$ and $\gamma_n \rightarrow \gamma$, $G_n \xrightarrow{s} G$.

PROOF. If $G(\infty) = 0$, then $G_n(\infty) \rightarrow 0$ by complete convergence, whence $\psi(t; \gamma_n, G_n) \rightarrow \psi(t; \gamma, G) = iyt$, recalling that the integrand $h(t, u)$ of (1) is bounded in modulus for fixed t . If $G(\infty) > 0$, then $G_n(\infty) > 0$ for all large n by Lemma 8.1.1, whence $(1/G_n(\infty))\psi(t; \gamma_n, G_n) \rightarrow (1/G(\infty))\psi(t; \gamma, G)$ by the Helly–Bray theorem, and so $\psi(t; \gamma_n, G_n) \rightarrow \psi(t; \gamma, G)$.

Apropos of the final assertion, Theorem 1 ensures that the i.d. c.f.s $e^{\psi_n(t)} = e^{\psi(t; \gamma_n, G_n)} \rightarrow e^{\psi(t)}$, continuous. Thus, $e^{\psi(t)}$ is a c.f. and i.d. by Theorem 1 and Proposition 4. Define $\alpha(t) = \text{Log } e^{\psi(t)}$ and $\alpha_n(t) = \text{Log } e^{\psi_n(t)} = \psi_n(t)$. By Theorem 8.3.3 $e^{\psi_n(t)} \rightarrow e^{\psi(t)}$ uniformly in $|t| \leq T$ for all $T \in (0, \infty)$, whence by Lemma 2, $\psi_n(t) \rightarrow \alpha(t)$ uniformly in $|t| \leq T$ and $\alpha(t)$ is continuous. Hence, recalling the proof and notation of the last part of Theorem 1 and defining

$$\begin{aligned} V_n(t) &= 2\psi_n(t) - \int_{t-1}^{t+1} \psi_n(y) dy = \int_{-\infty}^{\infty} e^{iy} dH_n(y), \\ V(t) &= 2\alpha(t) - \int_{t-1}^{t+1} \alpha(y) dy, \end{aligned}$$

it follows that $V_n(t) \rightarrow V(t)$, continuous, and, in particular, $H_n(\infty) = V_n(0) \rightarrow V(0)$, whence $V(0) \geq 0$. If $V(0) = 0$, then

$$H_n(\infty) = 2 \int_{-\infty}^{\infty} \left(1 - \frac{\sin x}{x}\right) \left(\frac{1+x^2}{x^2}\right) dG_n(x) \rightarrow 0,$$

implying $G_n(\infty) \rightarrow 0$, whence $G(u) \equiv 0$ and necessarily γ_n tends to a finite limit γ . If $V(0) > 0$, the d.f.s $H_n(u)/H_n(\infty)$ (whose c.f.s $V_n(t)/V_n(0) \rightarrow V(t)/V(0)$) converge to a limit d.f., say $H(u)/V(0)$. Thus, $H_n \xrightarrow{d} H$, and by the Helly–Bray theorem for any continuity point u of H , recalling Theorem 6.5.2,

$$\begin{aligned} 2G_n(u) &= \int_{-\infty}^u \left(1 - \frac{\sin y}{y}\right)^{-1} \frac{y^2}{1+y^2} dH_n(y) \\ &\rightarrow \int_{-\infty}^u \left(1 - \frac{\sin y}{y}\right)^{-1} \frac{y^2}{1+y^2} dH(y). \end{aligned} \quad (2)$$

Define $G(u)$ to be the integral on the right side of (2). Since the continuity points of G and H are identical, $G_n \xrightarrow{d} G$. Hence, γ_n tends to a finite limit γ . Clearly, $\psi(t, \gamma, G) = g(t)$. \square

From the preceding, a canonical form for i.d. c.f.s known as the Lévy–Khintchine representation follows readily.

Theorem 3 (Lévy–Khintchine representation). *A c.f. $\varphi(t)$ is i.d. iff*

$$\varphi(t) = \exp \left\{ i\gamma t + \int_{-\infty}^{\infty} \left(e^{itu} - 1 - \frac{itu}{1+u^2} \right) \left(\frac{1+u^2}{u^2} \right) dG(u) \right\}, \quad (3)$$

where γ, G are as stipulated in (1).

PROOF. Theorem 1 asserts that $\varphi(t) = e^{\psi(t; \gamma, G)}$ as above is i.d., and so it suffices to prove the converse. If $\varphi = \varphi_n$, $n \geq 1$, as in the proof of Proposition 5, $n[\varphi_n(t) - 1] \rightarrow \text{Log } \varphi(t)$. Now,

$$\begin{aligned} n[\varphi_n(t) - 1] &= \int_{-\infty}^{\infty} n(e^{itu} - 1) dF_n(u) \\ &= it \int_{-\infty}^{\infty} \frac{u}{1+u^2} n dF_n(u) \\ &\quad + \int_{-\infty}^{\infty} \left(e^{itu} - 1 - \frac{itu}{1+u^2} \right) \left(\frac{1+u^2}{u^2} \right) \frac{u^2}{1+u^2} n dF_n(u). \end{aligned}$$

Set

$$\begin{aligned} \gamma_n &= \int_{-\infty}^{\infty} \frac{nu}{1+u^2} dF_n(u), \quad G_n(u) = \int_{-\infty}^u \frac{ny^2}{1+y^2} dF_n(y), \\ \psi_n &= \psi(t; \gamma_n, G_n). \end{aligned}$$

As noted above, $\psi_n(t) \rightarrow \text{Log } \varphi(t)$, which is continuous. Thus, by Theorem 2 $\gamma_n \rightarrow \gamma$, $G_n \xrightarrow{\text{d}} G$, and

$$\varphi(t) = \lim_{n \rightarrow \infty} e^{\psi_n(t)} = e^{\psi(t; \gamma, G)}. \quad \square$$

In the case of distributions with finite variance, the canonical form in (3) admits considerable simplification.

Theorem 4 (Kolmogorov). *A function $\varphi(t)$ is the c.f. of an i.d. distribution with finite variance iff for some real γ^* and nondecreasing, left-continuous G^* with $G^*(-\infty) = 0$, $G^*(\infty) < \infty$*

$$\varphi(t) = \exp \left\{ i\gamma^* t + \int_{-\infty}^{\infty} (e^{itu} - 1 - itu) \frac{1}{u^2} dG^*(u) \right\} \quad (4)$$

and, moreover, φ uniquely determines γ^* and G^* .

PROOF. If the d.f. corresponding to the i.d. c.f. $\varphi(t) = e^{\psi(t)}$ has a finite second moment, ψ has a finite second derivative at zero and *a fortiori* a generalized second derivative at zero. Now, via (3)

$$\begin{aligned} \frac{\psi(2t) - 2\psi(0) + \psi(-2t)}{(2t)^2} &= \int_{-\infty}^{\infty} \frac{e^{2itu} - 2 + e^{-2itu}}{-(2it)^2} \frac{1+u^2}{u^2} dG(u) \\ &= - \int_{-\infty}^{\infty} \left(\frac{e^{itu} - e^{-itu}}{2it} \right)^2 \frac{1+u^2}{u^2} dG(u) = - \int_{-\infty}^{\infty} \left(\frac{\sin tu}{tu} \right)^2 (1+u^2) dG(u), \end{aligned}$$

whence

$$-\psi''(0) \geq \lim_{t \rightarrow 0} \int_{|u| \leq 1/t} \left(\frac{\sin tu}{tu} \right)^2 (1+u^2) dG(u) \geq (\sin 1)^2 \int_{-\infty}^{\infty} (1+u^2) dG(u).$$

Thus, $G^*(u) = \int_{-\infty}^u (1+y^2) dG(y)$ has the asserted properties i.e., $G^*(\infty) < \infty$ and

$$\begin{aligned} \int \left(e^{itu} - 1 - \frac{itu}{1+u^2} \right) \left(\frac{1+u^2}{u^2} \right) dG(u) &= \int \left(e^{itu} - 1 - \frac{itu}{1+u^2} \right) \frac{1}{u^2} dG^*(u) \\ &= \int (e^{itu} - 1 - itu) \frac{1}{u^2} dG^*(u) + it \int \left(u - \frac{u}{1+u^2} \right) \frac{1}{u^2} dG^*(u), \end{aligned}$$

whence (4) holds with $\gamma^* = \gamma + \int u dG$.

Conversely, if $\varphi(t) = \exp\{\Gamma(t)\}$ is as in (4), then, since

$$\frac{1}{u^2} \left(u - \frac{u}{1+u^2} \right) = O(1),$$

necessarily $\Gamma(t) = \psi(t; \gamma, G) + itc$ for some constant c , where

$$G(u) = \int_{-\infty}^u (1+y^2)^{-1} dG^*(y),$$

Now, noting via (2) that $|a| < \tau$,

$$\begin{aligned} \int_{\{|x| < \tau\}} (x - a)^2 dF &\leq [1 + (\tau + |a|)^2] \int_{\{|x| < \tau\}} \frac{(x - a)^2}{1 + (x - a)^2} dF, \\ \left| \int_{\{|x| < \tau\}} (x - a) dF \right| &= |a| \int_{\{|x| \geq \tau\}} dF \\ &\leq \frac{|a| [1 + (\tau + |a|)^2]}{(\tau - |a|)^2} \int_{\{|x| \geq \tau\}} \frac{(x - a)^2}{1 + (x - a)^2} dF, \end{aligned} \tag{4}$$

and so it follows from (3) that

$$\begin{aligned} \sup_{|t| \leq T} |1 - \tilde{\phi}_{n,k}(t)| &\leq [1 + (\tau + |a|)^2] \left(\frac{T^2}{2} + \frac{2 + |a|T}{(\tau - |a|)^2} \right) \int \frac{x^2}{1 + x^2} d\tilde{F}_{n,k} \\ &= \frac{1}{d_1} \int \frac{x^2}{1 + x^2} d\tilde{F}_{n,k}, \end{aligned}$$

where, noting that (i) entails $\max_k |a_{n,k}| < \tau/2$ for all large n ,

$$\begin{aligned} \frac{1}{d_1} &= [1 + (\tau + |a|)^2] \left[\frac{2 + T|a|}{(\tau - |a|)^2} + \frac{T^2}{2} \right] \\ &\leq \left(1 + \frac{9\tau^2}{4} \right) \left[\frac{2 + \tau T}{\tau^2/4} + \frac{T^2}{2} \right] = \frac{1}{c_1}, \end{aligned}$$

yielding the lower bound. Next, if F^* denotes the d.f. of a r.v. X^* with c.f. $|\varphi(t)|^2$, from the elementary inequality

$$\begin{aligned} \left(1 - \frac{\sin Tx}{Tx} \right) \left(\frac{1 + x^2}{x^2} \right) &\geq c(T) > 0, \\ \int_0^T (1 - |\varphi(u)|^2) du &= \int_{-\infty}^{\infty} \int_0^T (1 - \cos ux) du dF^* \\ &= T \int_{-\infty}^{\infty} \left(1 - \frac{\sin Tx}{Tx} \right) dF^*(x) \\ &\geq T c(T) \int_{-\infty}^{\infty} \frac{x^2}{1 + x^2} dF^*(x). \end{aligned} \tag{5}$$

For any r.v. X with d.f. F and median m , let X^* denote the symmetrized X and define

$$\begin{aligned} F^m(x) &= \mathbb{P}\{X - m < x\}, \quad q^m(x) = \mathbb{P}\{|X - m| \geq x\}, \\ q^*(x) &= \mathbb{P}\{|X^*| \geq x\}. \end{aligned}$$

In view of $1 - |\varphi|^2 \leq -\log|\varphi|^2 = -2\log|\varphi|$, the upper bound follows from (8) with $c_2 = 2d_2(\tau)/Tc(T)$, noting that for sufficiently large n

$$\begin{aligned} d_2 &= [1 + (\tau + |m|)^2] \left(1 + \frac{1 + 2\tau(\tau + |m|)}{(\tau - |m|)^2} \right) \\ &\leq (1 + 4\tau^2) \left(1 + \frac{1 + 4\tau^2}{(\tau/2)^2} \right) = d_2(\tau). \end{aligned} \quad \square$$

Lemma 3. If $\{X_{n,k}\}$ are infinitesimal r.v.s with c.f.s $\varphi_{n,k}$ for which

$$\lim_{n \rightarrow \infty} \prod_{k=1}^{k_n} |\varphi_{n,k}(t)| = f(t)$$

exists and is continuous at $t = 0$, then for any τ in $(0, \infty)$ there exists a constant C depending on τ and $\{\varphi_{n,k}\}$ such that

$$\sum_{k=1}^{k_n} \int \frac{x^2}{1+x^2} d\tilde{F}_{n,k}(x) \leq C.$$

PROOF. Since $\prod_{k=1}^{k_n} |\varphi_{n,k}(t)|^2 \rightarrow f^2(t)$, continuous at zero, f^2 is a c.f., whence T may be chosen so that $f^2 > \frac{3}{4}$ for $|t| \leq T$. Then, by uniform convergence $\prod_{k=1}^{k_n} |\varphi_{n,k}(t)|^2 > \frac{1}{2}$ for $n \geq N_T$ and $|t| \leq T$ and $\sum_{k=1}^{k_n} \log|\varphi_{n,k}(t)| \rightarrow \log f(t)$ uniformly in $|t| \leq T$. Hence from Lemma 2

$$\begin{aligned} \sum_{k=1}^{k_n} \int_{-\infty}^{\infty} \frac{x^2}{1+x^2} d\tilde{F}_{n,k}(x) &\leq -c_2 \sum_{k=1}^{k_n} \int_0^T \log|\varphi_{n,k}(t)| dt \\ &\rightarrow -c_2 \int_0^T \log f(t) dt. \end{aligned} \quad \square$$

The next lemma indicates that the c.f. of a sum of infinitesimal rowwise independent random variables behaves like that of a related i.d. c.f.

Lemma 4. If $\{X_{n,k}\}$ are infinitesimal r.v.s such that for some τ in $(0, \infty)$ there exists a C in $(0, \infty)$ with $\sum_{k=1}^{k_n} \int [x^2/(1+x^2)] d\tilde{F}_{n,k}(x) \leq C$, $n \geq 1$, then $\sum_{k=1}^{k_n} [\text{Log } \tilde{\varphi}_{n,k}(t) - (\tilde{\varphi}_{n,k}(t) - 1)] = o(1)$ for all real t . Moreover, for any constants A_n and all real t

$$\text{Log } e^{-iA_n} \prod_{k=1}^{k_n} \varphi_{n,k}(t) - \psi(t; \gamma_n, G_n) = o(1), \quad (9)$$

where

$$\gamma_n = -A_n + \sum_{k=1}^{k_n} \left(a_{n,k} + \int \frac{x}{1+x^2} d\tilde{F}_{n,k}(x) \right),$$

$$G_n(u) = \sum_{k=1}^{k_n} \int_{-\infty}^u \frac{x^2}{1+x^2} d\tilde{F}_{n,k}(x)$$

and ψ is as in (1) of Section 1.

PROOF. Any i.d. law characterized by (γ, G) is obtainable as a limit of distributions of row sums of independent, infinitesimal r.v.s $X_{n,k}$. It suffices to choose $k_n = n$ and take as the c.f. of $X_{n,k}$ the i.d. c.f. characterized by $(\gamma/k_n, (1/k_n)G)$ since such $X_{n,k}$ are clearly infinitesimal. Next, if for some constants A_n , $e^{-itA_n} \prod_{k=1}^{k_n} \varphi_{n,k}(t) \rightarrow g(t)$, a c.f., then Lemma 3 applies with $f = |g|$, and hence also Lemma 4, so that by Theorem 12.1.2, $\gamma_n \rightarrow \gamma$, $G_n \xrightarrow{d} G$, and $g = \exp\{\psi(t; \gamma, G)\}$. Finally, if $\gamma_n \rightarrow \gamma$, $G_n \xrightarrow{d} G$, Theorem 12.1.2 ensures $\psi(t; \gamma_n, G_n) \rightarrow \psi(t; \gamma, G)$, and $\sum_{k=1}^k \int [x^2/(1+x^2)] d\tilde{F}_{n,k} = G_n(\infty) \rightarrow G(\infty) < \infty$ whence Lemma 4 guarantees that $e^{-itA_n} \prod_{k=1}^{k_n} \varphi_{n,k}(t) \rightarrow e^{\psi(t; \gamma, G)}$. \square

Corollary 1. *The only admissible choices of the constants A_n are*

$$A_n = \sum_{k=1}^{k_n} \left(a_{n,k} + \int \frac{x}{1+x^2} d\tilde{F}_{n,k} \right) - \gamma + o(1)$$

for some constants γ and $\tau > 0$.

The next question that poses itself is under what conditions on $F_{n,k}$ a particular i.d. limit is obtained.

Theorem 2. *If $\{X_{n,k}, 1 \leq k \leq k_n \rightarrow \infty, n \geq 1\}$ are rowwise independent, infinitesimal r.v.s, then for any constants A_n , $e^{-itA_n} \prod_{k=1}^{k_n} \varphi_{n,k}(t) \rightarrow$ the i.d. c.f. $\exp\{\psi(t; \gamma, G)\}$ iff*

$$\begin{aligned} \sum_1^{k_n} F_{nk}(u) &\rightarrow \int_{-\infty}^u \frac{1+x^2}{x^2} dG(x), \quad 0 > u \in C(G) \\ \sum_1^{k_n} [1 - F_{nk}(u)] &\rightarrow \int_u^\infty \frac{1+x^2}{x^2} dG(x), \quad 0 < u \in C(G) \end{aligned} \quad (11)$$

$$\lim_{\varepsilon \rightarrow 0} \overline{\lim}_{n \rightarrow \infty} \sum_{k=1}^{k_n} \left[\int_{|x|<\varepsilon} x^2 dF_{nk}(x) - \left(\int_{|x|<\varepsilon} x dF_{nk}(x) \right)^2 \right] = G(0+) - G(0-), \quad (12)$$

$$-A_n + \sum_1^{k_n} \int_{|x|<\varepsilon} x dF_{nk}(x) \rightarrow \gamma + \int_{|x|<\varepsilon} x dG(x) - \int_{|x|\geq\varepsilon} \frac{1}{x} dG(x) \quad (13)$$

for some fixed $\varepsilon > 0$ for which $\pm\varepsilon$ are continuity points of G .

PROOF. By Theorem 1 it suffices to show that (11), (12), (13) are equivalent to

$$\sum_{k=1}^{k_n} \int_{-\infty}^u \frac{x^2}{1+x^2} d\tilde{F}_{nk}(x) = G_n(u) \xrightarrow{d} G(u) \quad (14)$$

and

$$-A_n + \sum_1^{k_n} \left(a_{nk} + \int \frac{x}{1+x^2} d\tilde{F}_{nk}(x) \right) = \gamma_n \rightarrow \gamma. \quad (15)$$

This will be effected by proving $(14) \Leftrightarrow (14') \Leftrightarrow (11'), (12') \Leftrightarrow (11), (12)$, and, moreover, that when (14) obtains $(13) \Leftrightarrow (15)$. In this schema, (11') is (11)

with $F_{n,k}$ replaced by $\tilde{F}_{n,k}$, and (12') and (14') are defined by

$$\lim_{\varepsilon \rightarrow 0^+} \varlimsup_n [G_n(\varepsilon) - G_n(-\varepsilon)] = G(0+) - G(0-) \quad (12')$$

(where G_n is as in (14)), and

$$\begin{aligned} G_n(x) &\rightarrow G(x), & 0 > x \in C(G), \\ G_n(\infty) - G_n(x) &\rightarrow G(\infty) - G(x), & 0 < x \in C(G), \\ \lim_{\varepsilon \rightarrow 0} \varlimsup_{n \rightarrow \infty} [G_n(\varepsilon) - G_n(-\varepsilon)] &= G(0+) - G(0-). \end{aligned} \quad (14')$$

Since (14) implies $\lim_{\varepsilon \rightarrow 0} \varlimsup_n [G_n(\varepsilon) - G_n(-\varepsilon)] = \lim_{\varepsilon \rightarrow 0} [G(\varepsilon) - G(-\varepsilon)] = G(0+) - G(0-)$, it is apparent that (14) \Rightarrow (14'). Conversely, under (14'), taking $\pm x \in C(G)$,

$$\begin{aligned} G(0+) - G(0-) &= \lim_{x \rightarrow 0^+} \varlimsup_n [G_n(x) - G_n(-x)] \\ &= \lim_{x \rightarrow 0^+} \varlimsup_n [G_n(x) - G_n(\infty) - G_n(-x) + G_n(\infty)] \\ &= \lim_{x \rightarrow 0^+} [G(x) - G(\infty) - G(-x) + \varlimsup_n G_n(\infty)] \\ &= G(0+) - G(0-) - G(\infty) + \varlimsup_n G_n(\infty), \end{aligned}$$

so that $\varlimsup_n G_n(\infty) = G(\infty)$, whence (14) holds. Of course, (14) \Rightarrow (12') trivially and (14') \Rightarrow (11') since for continuity points x , by the Helly–Bray theorem,

$$\begin{aligned} \sum_1^{k_n} \tilde{F}_{nk}(x) &= \sum_1^{k_n} \int_{-\infty}^x \frac{1+u^2}{u^2} \frac{u^2}{1+u^2} d\tilde{F}_{nk}(u) \\ &= \int_{-\infty}^x \frac{1+u^2}{u^2} dG_n(u) \rightarrow \int_{-\infty}^x \frac{1+u^2}{u^2} dG(u), \quad x < 0, \end{aligned}$$

$$\sum_1^{k_n} [1 - \tilde{F}_{nk}(x)] = \int_x^\infty \frac{1+u^2}{u^2} dG_n(u) \rightarrow \int_x^\infty \frac{1+u^2}{u^2} dG(u), \quad x > 0.$$

Conversely, (11'), (12') \Rightarrow (14') since for continuity points u , by the Helly–Bray theorem,

$$\begin{aligned} G_n(u) &= \int_{-\infty}^u dG_n(x) = \int_{-\infty}^u \frac{x^2}{1+x^2} d\left(\sum_1^{k_n} \tilde{F}_{nk}(x)\right) \\ &\rightarrow \int_{-\infty}^u \frac{x^2}{1+x^2} \frac{1+x^2}{x^2} dG(x) = G(u), \quad u < 0, \\ G_n(\infty) - G_n(u) &= \int_u^\infty \frac{x^2}{1+x^2} d\left(\sum_1^{k_n} \tilde{F}_{nk}(x)\right) \\ &= - \int_u^\infty \frac{x^2}{1+x^2} d\left(\sum_1^{k_n} [1 - \tilde{F}_{nk}(x)]\right) \\ &\rightarrow \int_u^\infty dG(x) = G(\infty) - G(u), \quad u > 0 \end{aligned}$$

and

$$\begin{aligned}
& \left| \int_{\{|x| < \varepsilon\}} (x - a_{nk})^2 dF_{nk} - \int_{\{|x| < \varepsilon\}} x^2 dF_{nk} + \left(\int_{\{|x| < \varepsilon\}} x dF_{nk} \right)^2 \right| \\
&= \left| \int_{\{|x| < \varepsilon\}} (-2ax + a^2) dF + \left(\int_{\{|x| < \varepsilon\}} x dF \right)^2 \right| \\
&= \left| \left(\int_{\{|x| < \varepsilon\}} x dF - a \right)^2 - a^2 \int_{\{|x| \geq \varepsilon\}} dF \right| \\
&= \left| \left(\int_{\{\varepsilon \leq |x| < \tau\}} x dF \right)^2 - a^2 \int_{\{|x| \geq \varepsilon\}} dF \right| \\
&\leq a_n \int_{[\varepsilon \leq |x| < \tau]} \tau dF + a_n^2 \int_{\{|x| \geq \varepsilon\}} dF \\
&\leq a_n(a_n + \tau) \int_{\{|x| \geq \varepsilon\}} dF_{nk},
\end{aligned}$$

implying

$$J_n(\varepsilon) \leq \sum_{k=1}^{k_n} \int_{\{\varepsilon - a_n \leq |x| \leq \varepsilon + a_n\}} (\varepsilon + a_n)^2 dF_{nk} + a_n(a_n + \tau) \sum_{k=1}^{k_n} \int_{\{|x| \geq \varepsilon\}} dF_{nk} = o(1)$$

since, choosing η in $(0, \varepsilon)$, where $\pm \varepsilon, \varepsilon \pm \eta$ are continuity points of G , and noting that $a_n < \eta$ for $n \geq N_\eta$ (and recalling that $(11) \Leftrightarrow (11')$)

$$\begin{aligned}
\sum_{k=1}^{k_n} \int_{\{|x| \geq \varepsilon\}} dF_{nk} &= \sum_k F_{nk}(-\varepsilon) + \sum_k [1 - F_{nk}(\varepsilon)] \rightarrow \int_{-\infty}^{-\varepsilon} \frac{1+x^2}{x^2} dG(x) \\
&\quad + \int_{\varepsilon}^{\infty} \frac{1+x^2}{x^2} dG(x) < \infty, \\
\sum_{k=1}^{k_n} \int_{\{\varepsilon - a_n \leq |x| \leq \varepsilon + a_n\}} dF_{nk}(x) &\leq \sum_k \int_{\{\varepsilon - \eta \leq |x| \leq \varepsilon + \eta\}} dF_{nk}(x) \\
&\rightarrow \int_{\{\varepsilon - \eta \leq |x| \leq \varepsilon + \eta\}} \frac{1+x^2}{x^2} dG(x) \quad (16) \\
&= o(1) \quad \text{as } \eta \rightarrow 0.
\end{aligned}$$

To complete the proof of the theorem it remains to show that (13) \Leftrightarrow (15) under (14). Since for all k

$$\begin{aligned}
 \left| \int_{\{|x| < \tau\}} x d\tilde{F}_{nk}(x) \right| &= \left| \int_{\{|x-a| < \tau\}} (x-a)dF \right| \\
 &= \left| \int_{\{|x-a| < \tau\}} (x-a)dF - \int_{\{|x| < \tau\}} (x-a)dF + a \int_{\{|x| \geq \tau\}} dF \right| \\
 &\leq \left| \int_{\{|x-a| < \tau \leq |x|\}} (x-a)dF - \int_{\{|x| < \tau \leq |x-a|\}} (x-a)dF \right| \\
 &\quad + |a| \int_{\{|x| \geq \tau\}} dF \\
 &\leq \tau \int_{\{\tau \leq |x| < \tau + |a|\}} dF \\
 &\quad + (\tau + |a|) \int_{\{\tau - |a| \leq |x| < \tau\}} dF + |a| \int_{\{|x| \geq \tau\}} dF \\
 &\leq \int_{\{\tau - a_n \leq |x| \leq \tau + a_n\}} (2\tau + a_n)dF + a_n \int_{\{|x| \geq \tau\}} dF, \\
 \left| \sum_{k=1}^{k_n} \int_{\{|x| < \tau\}} x d\tilde{F}_{nk}(x) \right| &\leq \sum_{k=1}^{k_n} \int_{\{\tau - a_n \leq |x| \leq \tau + a_n\}} (2\tau + a_n)dF_{nk} \\
 &\quad + a_n \sum_{k=1}^{k_n} \int_{\{|x| \geq \tau\}} dF_{nk} = o(1)
 \end{aligned}$$

via (16), and so, recalling that $\pm\tau$ are continuity points of G , it follows from

$$\begin{aligned}
 \sum_{k=1}^{k_n} \int \frac{x}{1+x^2} d\tilde{F}_{nk}(x) &= \sum_{k=1}^{k_n} \left(\int_{\{|x| < \tau\}} x d\tilde{F}_{nk} - \int_{\{|x| < \tau\}} \frac{x^3}{1+x^2} d\tilde{F}_{nk}(x) \right. \\
 &\quad \left. + \int_{\{|x| \geq \tau\}} \frac{x}{1+x^2} d\tilde{F}_{nk}(x) \right) \\
 &= o(1) - \int_{\{|x| < \tau\}} x dG_n(x) + \int_{\{|x| \geq \tau\}} \frac{1}{x} dG_n(x)
 \end{aligned}$$

that (13) \Leftrightarrow (15) under (14). \square

Theorem 3. Let $\{X_{n,k}, 1 \leq k \leq k_n \rightarrow \infty, n \geq 1\}$ be rowwise independent r.v.s.

(i) If $\{X_{n,k}\}$ are infinitesimal and $\sum_{k=1}^{k_n} X_{n,k} - A_n$ converges in distribution for some choice of constants A_n , the limit is normal iff

$$\max_{1 \leq k \leq k_n} |X_{n,k}| \xrightarrow{P} 0 \tag{17}$$

or equivalently

$$\sum_{k=1}^{k_n} \int_{\{|x| \geq \varepsilon\}} dF_{nk} = o(1), \quad \varepsilon > 0. \quad (18)$$

(ii) $\{X_{n,k}\}$ are infinitesimal and $\sum_{k=1}^{k_n} (X_{n,k} - a_{n,k})$ has a limiting normal distribution necessarily $N(0, \sigma^2)$ iff for all $\varepsilon > 0$, (18) holds and

$$\lim_{n \rightarrow \infty} \sum_{k=1}^{k_n} \left[\int_{\{|x| < \varepsilon\}} x^2 dF_{nk}(x) - \left(\int_{\{|x| < \varepsilon\}} x dF_{nk}(x) \right)^2 \right] = \sigma^2. \quad (19)$$

PROOF. (i) If $\sum X_{n,k} - A_n$ has a limiting distribution, then (11), (12), and (13) hold. Moreover, if the limit is normal, since $\exp\{\psi(t, \gamma, G)\} = e^{-(\sigma^2 t^2/2) + it\theta}$ iff $\gamma = 0$, $G(u) = 0$ for $u < 0$ and $G(u) = \sigma^2 u + \theta$, $u > 0$, (11) of Theorem 2 requires $\sum_1^{k_n} F_{nk}(-\varepsilon) = o(1) = \sum_1^{k_n} [1 - F_{nk}(\varepsilon)]$, $\varepsilon > 0$, which is tantamount to (18) and also to $\prod_1^{k_n} (1 - \int_{\{|x| > \varepsilon\}} dF_{nk}(x)) \rightarrow 1$, $\varepsilon > 0$, in view of $1 - \sum p_i \leq \prod (1 - p_i) \leq e^{-\sum p_i} \leq 1$ (where $0 \leq p_i \leq 1$). It is therefore also equivalent to (17) via $P\{\max_{1 \leq k \leq k_n} |X_{nk}| \geq \varepsilon\} = 1 - \prod_{k=1}^{k_n} (1 - P\{|X_{n,k}| \geq \varepsilon\})$.

Apropos of (ii), (18) and (19) ensure (11) and (12) with G as above and $\sigma^2 = G(0+) - G(0-)$, while (13) is satisfied with $A_n = \sum_{k=1}^{k_n} a_{n,k}$, $\gamma = 0$. Thus, $\sum_{k=1}^{k_n} (X_{n,k} - a_{n,k})$ has a limiting normal d.f. $N(0, \sigma^2)$ by Theorem 2. As for the converse, (18) holds via (i), and therefore it suffices to verify that (12) and (18) entail (19). Now, for $0 < \varepsilon' < \varepsilon$

$$\begin{aligned} & \sum_{k=1}^{k_n} \left\{ \int_{\{|x| < \varepsilon\}} x^2 dF_{nk} - \left(\int_{\{|x| < \varepsilon\}} x dF_{nk} \right)^2 \right\} \\ &= \sum_{k=1}^{k_n} \left\{ \int_{\{|x| < \varepsilon'\}} x^2 dF_{nk} - \left(\int_{\{|x| < \varepsilon'\}} x dF_{nk} \right)^2 \right\} \\ &+ \sum_{k=1}^{k_n} \left\{ \int_{[\varepsilon' \leq |x| < \varepsilon]} x^2 dF_{nk} - \left(\int_{[\varepsilon' \leq |x| < \varepsilon]} x dF_{nk} \right)^2 \right\} \\ &- 2 \sum_{k=1}^{k_n} \left(\int_{\{|x| < \varepsilon'\}} x dF_{nk} \right) \left(\int_{[\varepsilon' \leq |x| < \varepsilon]} x dF_{nk} \right), \end{aligned} \quad (20)$$

and in view of (18),

$$\begin{aligned} 0 &\leq \sum_{k=1}^{k_n} \left\{ \int_{[\varepsilon' \leq |x| < \varepsilon]} x^2 dF_{nk} - \left(\int_{[\varepsilon' \leq |x| < \varepsilon]} x dF_{nk} \right)^2 \right\} \\ &\leq \sum_{k=1}^{k_n} \int_{[\varepsilon' \leq |x| < \varepsilon]} x^2 dF_{nk} \leq \varepsilon^2 \sum_{k=1}^{k_n} \int_{\{|x| \geq \varepsilon'\}} dF_{nk} = o(1), \\ 0 &\leq 2 \sum_{k=1}^{k_n} \left| \int_{\{|x| < \varepsilon'\}} x dF_{nk} \right| \left| \int_{[\varepsilon' \leq |x| < \varepsilon]} x dF_{nk} \right| \\ &\leq 2\varepsilon\varepsilon' \sum_{k=1}^{k_n} \int_{\{|x| \geq \varepsilon'\}} dF_{nk} = o(1). \end{aligned}$$

Consequently, from (20), for all positive $\varepsilon, \varepsilon'$

$$\begin{aligned} & \overline{\lim}_{n \rightarrow \infty} \sum_{k=1}^{k_n} \left\{ \int_{\{|x| < \varepsilon\}} x^2 dF_{nk} - \left(\int_{\{|x| < \varepsilon\}} x dF_{nk} \right)^2 \right\} \\ &= \overline{\lim}_{n \rightarrow \infty} \sum_{k=1}^{k_n} \left\{ \int_{\{|x| < \varepsilon'\}} x^2 dF_{nk} - \left(\int_{\{|x| < \varepsilon'\}} x dF_k \right)^2 \right\}. \end{aligned}$$

Thus, the prior upper and lower limits are independent of ε , whence (12) ensures (19). \square

Corollary 2. If $\{X_{nk}, 1 \leq k \leq k_n \rightarrow \infty, n \geq 1\}$ are infinitesimal rowwise independent r.v.s with zero means and variances σ_{nk}^2 satisfying $\sum_{k=1}^{k_n} \sigma_{nk}^2 = 1$, $n \geq 1$, then $\sum_{k=1}^{k_n} X_{nk}$ has a limiting standard normal distribution iff for all $\varepsilon > 0$

$$\sum_{k=1}^{k_n} \int_{\{|x| \geq \varepsilon\}} x^2 dF_{nk}(x) = o(1). \quad (21)$$

PROOF. In view of

$$\begin{aligned} & \sum_{k=1}^{k_n} \int_{\{|x| \geq \varepsilon\}} x^2 dF_{nk}(x) \\ & \geq \max \left[\sum_{k=1}^{k_n} \left(\int_{\{|x| \geq \varepsilon\}} x dF_{nk}(x) \right)^2, \varepsilon \sum_{k=1}^{k_n} \int_{\{|x| \geq \varepsilon\}} |x| dF_{nk}(x), \varepsilon^2 \sum_{k=1}^{k_n} \int_{\{|x| \geq \varepsilon\}} dF_{nk}(x) \right], \end{aligned} \quad (22)$$

(21) implies (18) and also $\sum_{k=1}^{k_n} (\int_{\{|x| < \varepsilon\}} x dF_{nk})^2 = o(1)$ for all $\varepsilon > 0$, noting that $E X_{nk} = 0$. Consequently (21), which is equivalent to $\sum_{k=1}^{k_n} \int_{\{|x| < \varepsilon\}} x^2 dF_{nk}(x) \rightarrow 1$ via $\sum_{k=1}^{k_n} E X_{nk}^2 = 1$, also implies (19). Then, Theorem 3 guarantees a limiting standard normal d.f. for $\sum_{k=1}^{k_n} X_{nk}$ since $\sum_{k=1}^{k_n} a_{nk} = o(1)$ via (22). Conversely, if $\sum_{k=1}^{k_n} X_{nk}$ is asymptotically $N(0, 1)$, then taking $\gamma = 0 = A_n$ in Theorem 2, (11) and (13) ensure $\sum_{k=1}^{k_n} a_{nk} = o(1)$ so that for all $\varepsilon > 0$,

$$1 = \lim_{n \rightarrow \infty} \sum_{k=1}^{k_n} E X_{nk}^2 \geq \lim_{n \rightarrow \infty} \sum_{k=1}^{k_n} \int_{\{|x| < \varepsilon\}} x^2 dF_{nk}(x) \geq 1$$

via (19), implying $\sum_{k=1}^{k_n} \int_{\{|x| < \varepsilon\}} x^2 dF_{nk}(x) \rightarrow 1$ for all $\varepsilon > 0$, and, as already noted, this is tantamount to (21). \square

Corollary 3. If $\{X_n, n \geq 1\}$ are independent r.v.s with d.f.s $\{F_n, n \geq 1\}$, then $(1/B_n) \sum_1^n X_i - A_n$ has a limiting standard normal d.f. and $\{X_k/B_n, 1 \leq k \leq n\}$ are infinitesimal for some sequence $\{B_n\}$ of positive constants tending to ∞ iff for all $\varepsilon > 0$

$$\sum_{k=1}^n \int_{\{|x| \geq \varepsilon B_n\}} dF_k(x) = o(1), \quad (23)$$

Since $G(\infty) - G(0+) < \infty$, necessarily $0 < \alpha < 2$ and moreover, from (7)

$$e^{-\alpha(x+h_1)} + e^{-\alpha(x+h_2)} = e^{-\alpha(x+h)} \text{ or}$$

$$b_1^\alpha + b_2^\alpha = b^\alpha. \quad (9)$$

Proceeding in similar fashion with $J^-(x)$ if $G(0-) > 0$, it follows that

$$dG(x) = \frac{c_2|x|^{1-\alpha}}{1+x^2} dx, \quad x < 0, \quad c_2 = \alpha_0 J^-(0) > 0, \quad (10)$$

and again via (7)

$$b_1^{\alpha_0} + b_2^{\alpha_0} = b^{\alpha_0}. \quad (11)$$

Setting $b_1 = b_2 = 1$ in (9) and (11) reveals that $\alpha_0 = \alpha$.

Summarizing, if $G \not\equiv 0$, either $G(\infty) - G(0+) = 0 = c_1$ and $G(0-) = 0 = c_2$, whence $G(0+) - G(0-) = \sigma^2 > 0$, entailing $b^2 = b_1^2 + b_2^2$ via (6) and $\varphi(t)$ normal, i.e., $\alpha = 2$ or alternatively $0 < \alpha < 2$ and $G(0+) = G(0-)$ via (6), (9) and (11) with $b_1 = b_2 = 1$. In the latter case, from (4), (8), (10)

$$\begin{aligned} \psi(t) = i\gamma t + c_1 \int_{0+}^{\infty} \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) \frac{dx}{x^{1+\alpha}} \\ + c_2 \int_{-\infty}^{0-} \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) \frac{dx}{|x|^{1+\alpha}}. \end{aligned} \quad (12)$$

Next, (12) will be evaluated in terms of elementary functions.

(i) $0 < \alpha < 1$.

$$\begin{aligned} \psi(t) = it \left[\gamma - c_1 \int_0^\infty \frac{dx}{x^\alpha(1+x^2)} + c_2 \int_{-\infty}^0 \frac{dx}{|x|^\alpha(1+x^2)} \right] \\ + c_1 \int_{0+}^\infty (e^{itx} - 1) \frac{dx}{x^{1+\alpha}} + c_2 \int_{-\infty}^{0-} (e^{itx} - 1) \frac{dx}{|x|^{1+\alpha}}. \end{aligned} \quad (13)$$

By contour integration

$$\begin{aligned} 0 = \int_Q (e^{iz} - 1) \frac{dz}{z^{1+\alpha}} = \int_{R_1}^{R_2} (e^{iv} - 1) \frac{dv}{v^{1+\alpha}} + \int_{R_2}^{R_1} (e^{-iu} - 1) \frac{du}{u^{1+\alpha}} \\ + \sum_{j=1}^2 (-1)^j \int_0^{\pi/2} (\exp(iR_j e^{i\theta}) - 1) \frac{R_j e^{i\theta} id\theta}{R_j^{1+\alpha} e^{i\theta(1+\alpha)}}. \end{aligned}$$

Now, since $v \equiv |\exp(iRe^{i\theta}) - 1| \leq 2R$ for $0 < R < 1$ and $v \leq e^{-R \sin \theta} + 1 \leq 2$ for $0 \leq \theta \leq \pi/2$,

$$\left| \int_0^{\pi/2} (\exp(iRe^{i\theta}) - 1) \frac{id\theta}{R^\alpha e^{i\theta\alpha}} \right| \leq \begin{cases} 2 \int_0^{\pi/2} R^{1-\alpha} d\theta = o(1) & \text{as } R \rightarrow 0 \\ 2 \int_0^{\pi/2} \frac{d\theta}{R^\alpha} = o(1) & \text{as } R \rightarrow \infty, \end{cases}$$

it follows that

$$\begin{aligned} \int_0^\infty (e^{iv} - 1) \frac{dv}{v^{1+\alpha}} &= -i^{-\alpha} \lim_{\substack{R_1 \rightarrow 0 \\ R_2 \rightarrow \infty}} \int_{R_2}^{R_1} (e^{-u} - 1) \frac{du}{u^{1+\alpha}} = i^{-\alpha} \int_0^\infty (e^{-u} - 1) \frac{du}{u^{1+\alpha}} \\ &= \frac{-i^{-\alpha}}{\alpha} \int_0^\infty u^{-\alpha} e^{-u} du = \frac{-i^{-\alpha}}{\alpha} \Gamma(1-\alpha) = i^{-\alpha} \Gamma(-\alpha) \end{aligned}$$

via integration by parts and the recursion formula for the Γ function. Thus, if $t > 0$,

$$\int_0^\infty (e^{itx} - 1) \frac{dx}{x^{1+\alpha}} = t^\alpha \int_0^\infty (e^{iv} - 1) \frac{dv}{v^{1+\alpha}} = t^\alpha i^{-\alpha} \Gamma(-\alpha) = t^\alpha e^{-i\pi\alpha/2} \Gamma(-\alpha).$$

Since

$$\begin{aligned} \int_{-\infty}^0 (e^{itx} - 1) \frac{dx}{|x|^{1+\alpha}} &= \int_0^\infty (e^{-itx} - 1) \frac{dx}{x^{1+\alpha}} = \overline{\int_0^\infty (e^{itx} - 1) \frac{dx}{x^{1+\alpha}}} \\ &= t^\alpha e^{i\pi\alpha/2} \Gamma(-\alpha), \end{aligned}$$

setting

$$\gamma' = \gamma - c_1 \int_0^\infty \frac{dx}{x^2(1+x^2)} + c_2 \int_{-\infty}^0 \frac{dx}{|x|^2(1+x^2)}$$

and

$$c = -(c_1 + c_2) \cdot \Gamma(-\alpha) \cos \frac{\alpha\pi}{2} \geq 0,$$

for $t > 0$, from (13),

$$\begin{aligned} \psi(t) &= it\gamma' + \Gamma(-\alpha)t^\alpha \left[(c_1 + c_2) \cos \frac{\pi}{2}\alpha + i(c_2 - c_1) \sin \frac{\pi}{2}\alpha \right] \\ &= it\gamma' + \alpha \Gamma(-\alpha)t^\alpha \left([J(0) + J^-(0)] \cos \frac{\pi}{2}\alpha + [J^-(0) - J(0)] \sin \frac{\pi}{2}\alpha \right) \\ &= it\gamma' - ct^\alpha \left\{ 1 + i\beta \tan \left(\frac{\pi}{2}\alpha \right) \right\}, \end{aligned}$$

where

$$\beta = \frac{c_2 - c_1}{c_2 + c_1} = \frac{J^-(0) - J(0)}{J^-(0) + J(0)} \in [-1, 1]. \quad (14)$$

For $t < 0$,

$$\begin{aligned}\psi(t) &= \overline{\psi(-t)} = -i\gamma'(-t) - c(-t)^{\alpha} \left\{ 1 - i\beta \tan\left(\frac{\pi}{2}\right)\alpha \right\} \\ &= i\gamma't - c|t|^{\alpha} \left\{ 1 + i\beta \frac{t}{|t|} \tan\left(\frac{\pi}{2}\right)\alpha \right\},\end{aligned}$$

which dovetails with (3).

(ii) $1 < \alpha < 2$. Since $x/(1+x^2) = x - [x^3/(1+x^2)]$, it follows from (12) that for suitable γ''

$$\begin{aligned}\psi(t) &= it\gamma'' + c_1 \int_0^\infty (e^{itx} - 1 - itx) \frac{dx}{x^{1+\alpha}} \\ &\quad + c_2 \int_{-\infty}^0 (e^{itx} - 1 - itx) \frac{dx}{|x|^{1+\alpha}}.\end{aligned}\tag{15}$$

By the same contour integration

$$\begin{aligned}\int_0^\infty (e^{iv} - 1 - iv) \frac{dv}{v^{1+\alpha}} &= i^{-\alpha} \int_0^\infty (e^{-y} - 1 + y) \frac{dy}{y^{1+\alpha}} \\ &= i^{-\alpha} \alpha^{-1} \int_0^\infty (1 - e^{-y}) y^{-\alpha} dy \\ &= i^{-\alpha} M(\alpha) \text{ say, where } 0 < M(\alpha) < \infty,\end{aligned}$$

and so, setting

$$c = -M(\alpha)(c_1 + c_2) \cos \frac{\pi}{2}\alpha \geq 0,$$

for $t > 0$, from (15)

$$\begin{aligned}\psi(t) &= it\gamma'' - ct^{\alpha} \left\{ 1 + i\beta \tan\left(\frac{\pi}{2}\right)\alpha \right\} \\ &= it\gamma'' - c|t|^{\alpha} \left\{ 1 + i\frac{t}{|t|} \beta \tan\left(\frac{\pi}{2}\right)\alpha \right\},\end{aligned}$$

and exactly as in case (i) the above also holds for $t < 0$.

(iii) $\alpha = 1$. Since

$$\int_0^\infty \frac{1 - \cos u}{u^2} du = \int_0^\infty \frac{\sin u}{u} du = \frac{\pi}{2}$$

and

$$\lim_{\epsilon \rightarrow 0+} \int_\epsilon^\infty \frac{\sin v}{v^2} dv = \lim_{\epsilon \rightarrow 0+} \int_\epsilon^\infty \frac{v + o(v^2)}{v^2} dv = \log u, \quad u > 0,$$

if $t > 0$,

$$\begin{aligned}
& \int_0^\infty \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) \frac{dx}{x^2} \\
&= \int_0^\infty \frac{\cos tx - 1}{x^2} dx + i \int_0^\infty \left(\sin tx - \frac{tx}{1+x^2} \right) \frac{dx}{x^2} \\
&= -\frac{\pi}{2} t + i \lim_{\epsilon \rightarrow 0^+} \left[-t \int_\epsilon^\infty \frac{\sin v}{v^2} dv + t \int_\epsilon^\infty \left(\frac{\sin v}{v^2} - \frac{1}{v(1+v^2)} \right) dv \right] \\
&= -\frac{\pi}{2} t - it \log t + it\gamma_0,
\end{aligned}$$

noting

$$\frac{\sin v}{v^2} - \frac{1}{v(1+v^2)} = \frac{v + O(v^3)}{v^2} - \frac{1}{v} + \frac{v}{1+v^2} = O(v) \quad \text{as } v \rightarrow 0.$$

Thus, setting $c = (\pi/2)(c_1 + c_2) \geq 0$, for $t > 0$, from (12)

$$\begin{aligned}
\psi(t) &= it\tilde{\gamma} + c_1 \left[-\frac{\pi}{2}t - it \log t \right] + c_2 \left[-\frac{\pi}{2}t + it \log t \right] \\
&= it\tilde{\gamma} - ct \left[1 + i\beta \frac{2}{\pi} \log t \right] = it\tilde{\gamma} - c|t| \left[1 + i\beta \frac{t}{|t|} \frac{2}{\pi} \log |t| \right],
\end{aligned}$$

which coincides with (3) for $t > 0$ and also for $t < 0$. Clearly, $|\beta| \leq 1$ from (14).

Conversely, suppose that $\varphi(t)$ is defined by (3). If $\alpha = 2$, then $\varphi(t) = \exp\{i\tilde{\gamma}t - ct^2/2\}$, so that φ is a normal c.f. and hence a stable c.f. If, rather, $0 < \alpha < 2$, let $J^-(0)$ and $J^+(0)$ be determined by

$$\frac{J^-(0) - J(0)}{J^-(0) + J(0)} = \beta, \quad (16)$$

$$c = \begin{cases} -\alpha\Gamma(-\alpha)[J^-(0) + J(0)]\cos\frac{\pi}{2}\alpha & \text{if } \alpha \neq 1 \\ \frac{\pi\alpha}{2}[J^-(0) + J(0)] & \text{if } \alpha = 1. \end{cases} \quad (17)$$

Set

$$G(x) = \begin{cases} \alpha J^-(0) \int_{-\infty}^x \frac{dy}{|y|^{\alpha-1}(1+y^2)}, & x < 0 \\ \alpha J(0) \int_0^x \frac{dy}{y^{\alpha-1}(1+y^2)} + G(0), & x > 0 \end{cases} \quad (18)$$

and define for arbitrary $\gamma' \in (-\infty, \infty)$

$$\psi(t) = it\gamma' + \int_{-\infty}^{\infty} \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) \frac{1+x^2}{x^2} dG(x). \quad (19)$$

Then

$$\begin{aligned} \psi(t) &= it\gamma' + \alpha J(0) \int_0^{\infty} \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) \frac{dx}{x^{1+\alpha}} \\ &\quad + \alpha J^-(0) \int_{-\infty}^0 \left(e^{itx} - 1 - \frac{itx}{1+x^2} \right) \frac{dx}{|x|^{1+\alpha}} \end{aligned}$$

and, from the computations following (12), for some $\gamma'' \in (-\infty, \infty)$

$$\psi(t) = it(\gamma'' + \gamma') - c|t|^2 \left\{ 1 + i\beta \frac{t}{|t|} w(t, \alpha) \right\}. \quad (20)$$

Hence, from (3), (19), and (20), choosing $\gamma' = \gamma - \gamma''$,

$$\varphi(t) = \exp\{\psi(t)\},$$

whence φ is an i.d. c.f. by Theorem 12.1.3.

Moreover, $\varphi(t)$ is stable, since setting

$$s(t) = it\gamma - c|t|^2 \left[1 + i\beta \frac{t}{|t|} w(t, \alpha) \right],$$

for $\alpha \neq 1, 2$ and positive $b_i, i = 1, 2$,

$$s(b_1 t) + s(b_2 t) = s(bt) + ita,$$

where $b = b_1 + b_2 - a > 0$ and $b^2 = b_1^2 + b_2^2$. If, rather, $\alpha = 1$ then (21) obtains with $b = b_1 + b_2$ and

$$a = \frac{2}{\pi} c\beta [b \log b - b_1 \log b_1 - b_2 \log b_2].$$

Thus, φ is a stable c.f. □

EXERCISES 12.3

1. Show that the mass of a proper stable d.f. is confined to $(0, \infty)$ (resp. $(-\infty, 0)$) iff $\beta = -1$ (resp. $\beta = 1$).
2. Prove that if X has a stable distribution with characteristic exponent $\alpha \in (0, 2)$, then $E|X|^\beta < \infty$ for $\beta < \alpha$ and $E|X|^\beta = \infty$ for $\beta > \alpha$. Hint: If $X = X_1 + X_2$, where X_1, X_2 are i.i.d., then $E|X|^\beta < \infty$ for $\beta < \alpha$ by Exercise 8.4.11. If $\beta > \alpha$, then $E|X|^\beta = \infty$ by Theorem 8.4.1 (4).

3. Prove that all proper stable distributions are continuous and infinitely differentiable.
4. A d.f. F (with c.f. ϕ) is said to be in the **domain of attraction** (resp. domain of normal attraction) of a (stable) distribution G (with c.f. ψ) if for suitable constants A_n, B_n (resp. for some A_n and $B_n = bn^{1/\alpha}$), $\lim e^{itA_n}\phi^n(t/B_n) = \psi(t)$. Show that every stable distribution belongs to its own domain of (normal) attraction.
5. In coin tossing, prove that the probability that the m th return to equilibrium (equal number of heads and tails) occurs before m^2x tends to $2[1 - \Phi(x^{-1/2})]$ as $m \rightarrow \infty$. As usual, Φ is the standard normal d.f.
6. The limit distribution of Exercise 5 has density $f(x) = (2\pi x^3)^{-1/2}e^{-1/2x}$, $x > 0$. This is actually the stable density function corresponding to $\alpha = \frac{1}{2}$, $\beta = -1$, $\gamma = 0$, $c = 1$.
7. If $S_n = \sum_1^n X_i$, $n \geq 1$ is a random walk with X_i having a symmetric stable distribution of characteristic exponent α , show that $S_n/n^{1/\alpha}$ also has this distribution. Hence, if $1 \leq r < \alpha < 2$, $E|S_n| = Cn^{1/\alpha}$ for some C in $(0, \infty)$ whence the conclusion of Theorem 10.3.4 fails for $1 \leq r < 2$.

References

- K. L. Chung, *A Course in Probability Theory*, Harcourt Brace, New York, 1968; 2nd ed., Academic Press, New York, 1974.
- H. Cramér, "Su un teorema relativo alla legge uniforme dei grande numeri," *Giornale dell'Istituto degli Attuari* 5 (1934), 1–13.
- C. Esseen, "Fourier analysis of distribution functions," *Acta Math* 77 (1945), 79.
- B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, Reading, Mass., 1954.
- P. Lévy, *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, Paris, 1937; 2nd ed., 1954.
- M. Loève, *Probability Theory*, 3rd ed., Van Nostrand, Princeton, 1963; 4th ed., Springer-Verlag, Berlin and New York, 1977–1978.

Index

- \mathcal{A} measurable function, 15
- Abbott, 163
- Abel, 114, 119
- Absolutely continuous
 - distribution function, 27, 270
 - set function, 203, 205
 - random variables, 27
- Additive (set function), 18
 - σ -, 19
 - see also* Subadditive set function
- Algebra, 6
 - degenerate σ -, (of events), 238
 - product σ -, 8
 - semi-, 20
 - σ -, 6
 - see also* σ -algebra generated by;
 σ -algebra of permutable
events
 - tail σ -, 63
- Almost certainly (surely), 20
- Almost everywhere, 171
- Anscombe, 340, 353
- At random, 56
- Austin, 408, 442
- Ballot problem, 250
- Banach space, 111
- Barndorff-Nielsen, 65, 83
- Baum, 83, 135, 163, 401
- Baum-Katz-Spitzer theorem, 135
 - converse, 199
- Bayes theorem, 228
- Bernoulli, 53, 75
 - trials with parameter p , 75–81
 - trials with success probability p , 56
 - weak law of large numbers, 39
- Bernstein, 53, 353
 - inequality, 111
 - polynomial, 42
- Berry, 318, 322, 353
 - Esseen theorem, 322, 337
- Binomial
 - distribution function, 31, 294
 - negative, 38, 60
 - random variable, 31, 50, 56
- Bivariate Poisson distribution, 191
- Blackwell, 156, 437, 442
- Blum, 341, 353
- Bochner theorem, 308
- Bonferroni inequalities, 38
- Borel, 53, 75
 - Cantelli lemma, 42, 44, 101, 102
 - Cantelli theorem, 61, 96, 258

- Borel (*cont.*)
 (measurable) function, 14
 line, 11
 set, 11
 space, 11, 186
 strong law of large numbers, 42
 zero-one criterion, 61
- Bounded in probability, 273
- Branching process, 258
- Bray, 274, 276, 311
- Breiman, 258, 268
- Brown, B., 161, 163, 428, 442
- Brunk, 363, 401
 –Chung strong law of large numbers, 363
- Buhlman, 268
- Burkholder, 163, 414, 422, 425, 442
 –Davis-Gundy inequality, 425
 inequality, 414
- Cantelli, 42, 44, 61, 83, 96, 258, 284, 311
- Carleman criterion, 303
- Cauchy
 convergence criterion, 99
 distribution function, 294
- Central limit theorem, 48, 313, 345
 De Moivre-Laplace, 46, 47
- Doeblin-Anscombe, 340
 for martingales, 327, 336, 345
 for Poisson random variables, 52
 for sums of interchangeable random variables, 328, 329
 for U-statistics, 326
- Liapounov, 316
- Lindeberg-Feller, 314
- Characteristic exponent, 469
- Characteristic function, 286
r-analytic, 301
 entire, 301
- Chernoff, 281, 311, 353
- Chow, 111, 163, 221, 268, 286, 311, 401, 402, 442
- Chung, 70, 71, 83, 98, 111, 123, 130, 137, 154, 163, 268, 311, 353, 402, 477
- Class, 1
 λ , 7
 monotone, 6
 π , 7
- Complement, 4
- Complete
 compactness, 282
 convergence of distribution functions, 271
 convergence of random variables, 43
- Completion (of), 25, 236
- Conditional
 density, 224
 distribution function, 225
 expectation, 210, 212, 213, 216–218
 independence, 229–232
 probability, 59, 222, 223
 probability measure (regular conditional probability), 223
see also Regular conditional distribution, 225
- Consistent
 family of distribution functions, 195
- Continuity point, 175, 271
- Convergence
 almost certainly (a.c.) or almost surely (a.s.), 43, 66
 a.c. unconditionally, 121
 almost everywhere (a.e.), 172, 173
 complete (for distribution functions), 271, 309
 complete (for random variables), 44
 in distribution (law), 272
 in mean of order p , 95, 99, 102, 104, 109
 in measure, 172
 in probability, 43, 66, 72–74
 moment, 317
 weak, 271
- Convex function, 102–104, 110
 inequality for martingales, 421, 422, 425

- Convolution, 189
Coordinate random variable, 57, 187, 196
Copies (of a stopping time), 141, 142, 146, 147, 150, 155
Correlation coefficient, 106
Countable set, 1
Counting measure, 24
Covariance, 109
Cramer, 273, 311, 477
–Levy theorem, 305
- Davis, 418, 422, 425, 442
Defective stopping time, 138
de Finetti, 33, 53, 234, 268
Degenerate
distribution function, 31, 70, 294
 σ -algebra (of events), 238
random variable, 31, 64
U-statistic, 259
- Delayed sum, 135
De Moivre, 53, 75
–Laplace (central limit) theorem, 46–48
Density (function), 27, 31, 46, 206
Discrete
distribution function, 27
random variable, 27
- Disjoint
class, 4
- Distinguished
logarithm, 446
 n th root, 446
- Distribution function (d.f.), 25, 26, 28, 30, 270, 271
absolutely continuous, 270
associated, 370
binomial, 31
bivariate Poisson, 191
Cauchy, 294
conditional, 225
degenerate, 31, 270, 294
discrete, 27
exponential, 60
gamma, 294
geometric, 60
- hypergeometric, 39
infinitely divisible, 445
inverse triangular, 294
joint, 26, 187
joint normal, 208, 221, 228
marginal, 309
multinomial, 191
 n -dimensional, 187
negative binomial, 38, 60
normal, 31, 53, 294
Poisson, 31, 294
positive normal, 78, 343
positive stable, 343
sample (empirical), 284
singular, 270
stable, 468
symmetric Bernoulli, 294
triangular, 294
- Doeblin, 340, 353
–Anscombe central limit theorem, 340
- Doob, 29, 83, 111, 164, 209, 223, 225, 239, 245, 268, 291, 311, 353, 402, 442
maximal inequalities, 255
upcrossing inequality, 405
- Domain of attraction, 477
Dubins, 254, 268, 433–437
Dvoretzky, 339, 348, 352, 353, 441
Dynkin, 29, 268
- Egorov, D. H., theorem, 75
Egorov, V. A., 402
Elementary function, 85
Equicontinuous, 208
Equivalence
of measures, 207
of sequences of random variables, 116
- Erdos, 135, 342, 353
Erickson, 164, 183
Esseen, 318, 322, 353
Essential supremum, 202
Etemadi, 131, 132
Exchangeable (see interchangeable)
random variables

- Event, 19
- Expectation (mean), 84
 - existence of, 84
- Exponential
 - distribution, 60
 - random variable, 60, 65
- Extension
 - of a sequence, 90
 - of a set function, 20, 165
- Factor closed, 300
- Fatou lemma, 95, 174
 - for conditional expectations, 216
 - extended, 218
- Feller, 53, 61, 70, 83, 126, 128, 137, 164, 311, 313, 314, 353, 402
 - Chung lemma, 70
 - weak law of large numbers, 128
- Fictitious random variable, 272
- Finite
 - measure space, 23
 - partition, 18
 - permutation, 232
 - real line, 11
 - set, 2
 - set function, 18
 - stopping time (rule, variable), 138, 240
- Finitely additive set function, 18
- First passage time, 141
- Frechet, 283, 311
 - Shohat theorem, 283
- Freedman, 254, 268, 442
- Friedman, 325
- Fubini theorem, 186, 191, 213
- Fuchs, 98, 111, 154, 163
- Function
 - σ -measurable, 15
 - absolutely continuous, 27
 - additive, 18
 - Borel (measurable), 14
 - convex, 102–104
 - density, 27, 31, 46, 206
 - discrete distribution, 27
 - distribution, 26
- elementary, 85
- finite set, 18
- integrable, 84, 89, 172
- joint distribution, 26, 187
- left continuous, 25, 26
- moment generating, 110
- monotone set, 19
- probability density, 27
- set, 18
- σ -additive (countably additive) set, 19
- σ -finite set, 18
- simple, 18
- subadditive set, 19
- subtractive set, 24
- tail, 64
- Gambler's ruin problem, 81, 250
- Gamma distribution, 294
- Garsia, 425, 442
- Geometric distribution, 60, 91
- Glivenko, 284, 311
 - Cantelli theorem, 284, 311
- Gnedenko, 312, 402, 477
- Gundy, 157, 164, 422, 425, 442
- Haag, 33, 53
- Hajek, 255, 268
 - Renyi inequality, 255
- Halmos, 29, 111, 177, 209, 222, 268
- Hanson, 353
- Hardy, 53, 103, 111, 209, 297, 312, 368, 372
- Harris, 258
- Hartman, 374, 378, 402
 - Wintner law of the iterated logarithm, 382
- Hausdorff, 29, 368, 372
- Helly, 274, 312
 - Bray lemma, 274
 - Bray theorem, 276
- Hewitt, 238, 268
- Heyde, 137, 164, 346, 353
- Hoeffding, 268
 - decomposition, 261

- Holder inequality, 104, 109, 174, 227
Hsu, 44, 53, 135, 393, 402
Hypergeometric distribution, 39
- Identifiable, 294
Independent
 classes, 54
 conditionally, 229–232
 events, 54
 families of random variables, 55
 identically distributed (i.i.d.)
 random variables, 55
Indicator, 4
Induced measure, 25, 176
Infinite dimensional product
 measurable space, 10
 measure space 191–193
- Infinitely
 often (i.o.), 2
 divisible (distribution), 445
- Inequality
 Burkholder, 414
 Burkholder-Davis-Gundy, 422, 425
 Doob maximal, 255
 Doob upcrossing, 404–406
 Hajek-Renyi, 255
 Holder, 104, 109, 174, 227
 Jensen, 104, 217
 Khintchine, 384
 Kolmogorov, 133, 255
 Levy, 72
 Marcinkiewicz-Zygmund, 386
 Markov, 86, 89, 173
 Minkowski, 110, 183
 Ottaviani, 75
 Schwarz, 105
 Tchebychev, 40, 106
 Young, 184, 424
- Integrable, 84, 93, 172
 function, 172
 uniformly, 93
- Integral, 84, 171
 indefinite, 92, 93
 Lebesgue, 176
- Lebesgue-Stieltjes, 175, 180
Riemann, 178
Riemann-Stieltjes, 178–180
- Interchangeable (exchangeable)
 events, 33
 random variables, 191, 231–234, 238, 241
- Inverse triangular distribution, 294
- Inversion formula, 287
- Jensen inequality, 104
 for conditional expectations, 217
- Joint
 distribution function, 26, 187
 normal distribution, 208
 probability density function, 55
- Jordan decomposition, 208
- Kac, 342, 353
Katz, 83, 135, 163, 325, 401
Kawata, 164
Kemperman, 267
Kendall, 228, 268
Kesten, 156, 164, 214
Khintchine, 113, 164, 368, 372, 402, 451
 inequality, 384
 –Kolmogorov convergence theorem, 113
Kiefer, 397, 402
Kingman, 402
Klass, 137, 138, 164, 269
Knopp, 164, 353
Kochen, 102, 112
Kolmogorov, 83, 113, 124, 164, 208, 312, 368, 372, 402, 445, 452, 477
 consistency theorem, 195
 inequality, 133, 255
 law of the iterated logarithm, 373
 strong law of large numbers, 125
 three series theorem, 117
 zero-one law, 64
- Komlos, 137

- Koopmans, 325
 Krickeberg, 269
 Kronecker lemma, 114
- Lai, 402
 Laplace, 46, 47, 53, 75
 Law of the iterated logarithm (LIL)
 independent random variables, 373–382
 U-statistics, 382
 Hartman-Wintner, 382
 Kolmogorov, 373
- Law of large numbers, Strong
 (SLLN), Weak (WLLN)
 definition, 124
 Bernoulli (weak), 39
 Borel (strong), 42
 Brunk-Chung (strong), 363
 Etemadi, 132
 for arrays, 138, 393
 for interchangeable random
 variables, 235
 for U-statistics, 263
 Feller (weak), 128
 generalized SLLN, 126–128,
 361–366
 generalized WLLN, 356–359, 467
 Kolmogorov (strong), 125
 Marcinkiewicz-Zygmund (strong),
 125
- Lebesgue
 decomposition theorem, 204
 dominated convergence theorem,
 100, 174
 for conditional expectations,
 216
 integral, 176
 measurable set, 170
 measure, 170
 measure space, 170
 monotone convergence theorem,
 86, 90, 95, 173
 for conditional expectations,
 216
 –Stieltjes measure, 170, 188
- Stieltjes measure space, 170, 175,
 188
 see also Non-Lebesgue measurable set, 171
- Left continuous, 25, 168
 Levy, 71, 72, 83, 164, 239, 245, 269,
 305, 312, 353, 451, 477
 class \mathcal{L} (of distributions), 468
 concentration function, 286
 continuity theorem, 289
 distance, 278
 inequality, 72
 inversion formula, 287
 –Khintchine representation, 451
 theorem, 72
- Liapounov, 105, 112
 central limit theorem, 316
- Likelihood ratio, 257
- Lindeberg, 313, 353
 condition, 313, 317
 –Feller central limit theorem, 314
- Linear Borel set, 11
- Littlewood, 111, 209, 368, 372
- Loeve, 83, 112, 117, 124, 164, 209,
 269, 312, 402, 477
- \mathcal{L}_p random variable, 95
 \mathcal{L}_p space, 95
 Lukacs, 312
- Marcinkiewicz, 118, 121, 125, 164,
 386, 387, 402
 –Zygmund inequality, 387
 –Zygmund strong law of large
 numbers, 125, 138, 263,
 393
- Marginal distribution, 309
- Markov
 chain, 238
 inequality, 86, 89, 173
- Martingale, 239
 central limit theorems, 336, 345,
 346, 351
 convergence theorems, 247, 248,
 406–408, 411
 differences, 241, 336

- inequalities, 255, 256, 409, 414, 418, 425, 427, 430, 434
moments of stopped (martingales), 250, 253, 431
reversed (downward), 241
Wald equation for, 250–253, 415
- Match, 38
- Maximal inequalities, 255, 256, 368
- McShane, 209
- McLeish, 348, 349, 353
- Mean convergence criterion, 99
- Measurable cover, 228
function, 14
rectangle, 9
set, 8
see also \mathcal{A} -measurable function; Lebesgue or v -measurable set
- Measure, 19
complete, 168
conditional probability, 222
convergence in, 172
counting, 24
extension, 165
finite, 23
induced, 25
Lebesgue, (and measure space), 170
Lebesgue-Stieltjes, (and measure space), 170
 n -dimensional Lebesgue-Stieltjes, 188
outer, 168
product, 184
restriction of a, 165
 σ -finite, 165
signed, 208
space, 19
see also Infinite (and n)-dimensional product measure space, 185, 193
- Median, 71, 109
- Minkowski inequality, 110, 183
- Mixture, 190, 286, 292, 294, 345, 453
- Mogyorodi, 341, 353
- Moment, 105
convergence, 277, 317, 442
generating function, 110, 281
of randomly stopped sums, 250, 253, 431, 432
- Monotone class, 6
convergence theorem, 86, 90, 95, 173
convergence theorem for conditional expectations, 216
sequence of sets, 3
set function, 19
system 15
- Monroe, 209
- Multinomial distribution, 191
- Nagaev, 366, 402
- n -dimensional Borel measure space, 186
distribution function, 187
- Negative binomial distribution 38, 60
part, 15
- Neveu, 442
- Newman, 61
- Nikodym, 204, 205
- Non-Lebesgue measurable set, 171
- Normal distribution, 31, 294
positive, 78, 343
- Normal random variable, 31
- v -measurable set, 168
- Null event, 20
- Number of upcrossings, 404–406
- Optimal stopping rule, 160
- Ornstein, 130, 154, 163
- Ottaviani, 305
inequality, 75
- Outer measure, 168
- p -norm, 104
- Panzone, 405, 443
- Parameter, 31
- Periodic, 297

- Permutable events, 232
- Poincare formula, 33
- Point
 - of continuity, 175, 271
 - of increase, 28, 271
- Poisson
 - distribution, 31, 52, 294
 - random variable, 31, 65
 - theorem, 32
- Polya, 98, 112, 299, 312
- Positive
 - definite, 308
 - normal distribution, 78, 343
 - part, 15
 - stable distribution, 343
- Probability, 19
 - conditional, 59, 222, 223
 - density function, 27
 - space, 19
 - success probability, 56
- Product
 - measurable space, 8
 - measure, 184
 - measure space, 184, 185
 - σ -algebra, 8
 - space, 8
- Prohorov, 402
- Rademacher functions, 201
- Radon
 - Nikodym derivative, 205
 - Nikodym theorem, 204
- Raikov, 312, 402, 467
 - Ottaviani theorem, 308
- Random
 - allocation of balls into cells, 56, 57, 330–335
 - vector, 26
 - walk, 76, 392–401
 - see also Simple random walk
- Random variable, 20
 - absolutely continuous, 27
 - binomial, 31
 - coordinate, 57, 196, 234, 235
 - degenerate, 31
- discrete, 27
- exponential, 65
- fictitious, 272
- independent, identically distributed (i.i.d.), 55
- interchangeable (exchangeable), 191, 231–234, 238, 241
 - \mathcal{L}_p , 95
- normal, 31, 65
- Poisson, 31, 65
- symmetric, 72, 74
- symmetrized, 197
- Real line, 11
- Rectangle, 9
- Recurrent, 97, 98, 130, 163
- Regular conditional distribution, 225
- Renyi, 52, 53, 268, 353
- Renewal theorem (elementary), 157–159
- Restriction, 20, 165
- Revesz, 137, 164
- Riemann
 - integral, 178
 - Lebesgue lemma, 307
 - Stieltjes integral, 178–180
- Riesz representation theorem, 208
- Robbins, 45, 53, 135, 138, 160, 164, 209, 221, 286, 312, 375, 384
- Rogozin, 137
- Rosenblatt, J., 353
- Rosenblatt, M., 353
- Rowwise independence, 138, 393, 454, 459, 460
- Saks, 29, 112, 209, 312
- Same type, 284
- Sample (empirical)
 - distribution, 284
 - space, 19
- Samuel, 143, 164
- Samuels, 82
- Sapogov, 312
- Savage, 238, 269, 433–437
- Scheffé, 312

- Schwarz inequality, 105
Second moment analogue of Wald's equation, 144, 253, 254
Section
 of a function, 185
 of a set, 13
Semi-algebra, 20, 24
Set
 Borel, 11
 function, 19
 measurable, 8
 operation, 2
Shohat, 283, 303, 312
Siegmund, 164, 402
 σ -additive (countably additive), 19
 σ -algebra, 6
 generated by, 7, 16
 of permutable events, 232
 σ -finite, 18
 measure, 165
 partition, 18
Signed measure, 208
Simple
 function, 18
 random walk, 98
Singular
 distribution, 270
 μ -singular, 203
Slutsky, 272, 312
Snell, 209
Space, 4, 8, 11, 19, 95, 111, 168, 185
Spitzer, 135, 164, 402
Stable distribution, 468–477
Standard
 deviation, 106
 normal, 31
Stein, E., 402
Stirling, 53
 formula, 45, 49
Stochastic
 larger, 183
 \mathcal{L}_p -bounded stochastic sequence, 239
 matrix, 238
 sequence, 138, 239
Stone, 102, 112, 147, 164
Stopping time, 138, 240
 $\{X_n\}$ -time, 138
Strassen, 403
Stratton, 83
Strong law of large numbers (SLLN), 124
independent random variables, 42, 124–128, 356–359, 361–366
interchangeable random variables, 235
martingale differences, 258, 415
pairwise independent random variables, 132
U-statistics, 263
arrays, 138, 393
Studden, 111
Subadditive set function, 19
Submartingale, 239
 closed, 240
 convergence theorems, 246, 248, 406, 407
Subtractive, 24
Success probability, 56
Supermartingale, 239
Support (spectrum), 28, 271
Symmetric
 Bernoulli distribution, 294
 distribution, 72, 297
 random variable, 72
Symmetrized random variable, 197
System
 λ -, 15
 monotone, 15
Tail
 event, 64
 function, 64
 of a distribution, 49
 σ -algebra
Tamarikin, 283, 303, 312
Taylor, 403
Tchebychev inequality, 40, 106

- Teicher, [53](#), [112](#), [137](#), [163](#), [164](#), [268](#),
[269](#), [312](#), [353](#), [403](#), [443](#)
- Three series theorem, [117](#), [123](#)
- Tight, [276](#)
- Titchmarsh, [302](#), [303](#), [306](#), [312](#)
- Total variation, [208](#)
- Triangular distribution, [294](#)
- Truncation, [106](#), [110](#), [182](#), [213](#)
- Uncorrelated, [106](#)
- Uniform distribution, [39](#), [293](#)
- Uniformly
- absolutely continuous, [208](#)
 - bounded random variables, [116](#)
 - integrable ([u.i.](#)) random variables, [93](#), [94](#)
 - integrable relative to distribution functions, [276](#), [277](#)
- U-statistics, [241](#), 259–268
- Central limit theorem, [326](#)
 - degenerate, [259](#)
 - decomposition of, [261](#)
 - Strong law of large numbers, [263](#)
 - Law of the iterated logarithm, [382](#)
- Van Beek, [323](#), [353](#)
- Variance, [105](#)
- Von Mises, [58](#), [83](#)
- Wald, [158](#), [162](#), [164](#)
- equation, [143](#), [144](#), [253](#), [254](#), [415](#)
- Weak
- compactness (sequential), [282](#)
 - law of large numbers, [124](#), [128](#),
[235](#), [356](#)–[359](#)
- Weierstrass, [121](#)
- approximation theorem, [42](#)
- Weiss, [L](#), [353](#)
- Widder, [178](#), [209](#)
- Wiener dominated ergodic theorem, [387](#), [403](#)
- Wintner, [374](#), [378](#), [402](#)
- Wolfowitz, [347](#), [402](#)
- Yadrenko, [149](#)
- Young, L. C., [112](#)
- Young, W. H., inequality, [184](#), [424](#)
- Zero-one law, [97](#)
- Hewitt-Savage, [238](#)
 - Kolmogorov, [64](#)
 - for interchangeable random variables, [238](#)
 - see also* Borel zero-one criterion
- Zolotarev, [353](#)
- Zygmund, [118](#), [121](#), [125](#), [164](#), [386](#),
[387](#), [402](#), [422](#), [443](#)

This is a text comprising the major theorems of probability theory and the measure theoretical foundations of the subject. The main topics treated are independence, interchangeability, and martingales; particular emphasis is placed upon stopping times, both as tools in proving theorems and as objects of interest themselves. No prior knowledge of measure theory is assumed, and a unique feature of the book is the combined presentation of measure and probability. It is easily adapted for graduate students familiar with measure theory as indicated by the guidelines in the Preface. Special features include:

- An up-to-date treatment of *U*-statistics
- A comprehensive treatment of the law of the iterated logarithm
- The Marcinkiewicz-Zygmund inequality, its extension to martingales and applications thereof
- Development and applications of the second moment analogue of Wald's equation
- Limit theorems for martingale arrays; the Central Limit Theorem for the interchangeable and martingale cases; moment convergence in the Central Limit Theorem
- Complete discussion, including Central Limit Theorem, of the random casting of r_n balls into n cells
- Recent martingale inequalities
- Cramér-Lévy theory and factor-closed families of distributions

This edition includes a section dealing with *U*-statistic, adds additional theorems and examples, and includes simpler versions of some proofs.

Yuan Shih Chow is Professor of Statistics at Columbia University.
Henry Teicher is Professor of Statistics at Rutgers University.



Springer

ISBN 0-387-40607-7
www.springer-ny.com

ISBN 0-387-40607-7

9 780387 406077 >

Copyrighted material