Measure Theory for Engineers

Manish Agarwal

Oct 2020

Based on the the measure theory book by Terence Tao. There are a very few mathematicians who can explain mathematics as intuitively as Terence Tao.

Contents

0.1 Preliminaries

We now layout some definitions from background material, that we use throughout.

- A **metric space** (X, d) is a space X of objects, called points, together with a distance function or metric $d : X \times X \to [0, \infty)$, which associates to each pair x, y of points in X a non-negative real number $d(x, y) \geq 0$ and satisfies the four axioms: Identity, Positivity, Symmetry, and Triangle inequality. Let $\boldsymbol{R}$ be the real numbers, and let $d : \boldsymbol{R} \times \boldsymbol{R} \to [0, \infty)$ be the metric $d(x, y) := |x - y|$. Then $(\boldsymbol{R}, d)$ is a metric space, referred to as **standard metric** in $\boldsymbol{R}$.

- **Metric balls** Let (X, d) be a metric space, let x_0 be a point in X, and let $r > 0$. We define the ball $B_{(X,d)(x_0,r)}$ in X, centered at x_0, and with radius r, in the metric d, to be the set $B_{(X,d)(x_0,r)} := \{x \in X : d(x, x_0) < r\}$. When the metric space is apparent we denote it by $B(x_0, r)$.

- **Interior, exterior, boundary** Let (x, d) be a metric space, let E be a subset of X, and let x_0 be a point in X. We say that x_0 is an interior point of E if there exists a radius $r > 0$ such that $B(x_0, r) \subseteq E$. We say that x_0 is an exterior point of E if there exists a radius $r > 0$ such that $B(x_0, r) \cap E = \emptyset$. We say that x_0 is a boundary point of E if it is neither an interior or exterior point of E. $int(E)$ denote the set of all interior of E and $ext(E)$ denote the set of all exterior of E. Finally ∂E denote the set of all boundary of E.

- **Closure** Let (X, d) be a metric space, let E be a subset of X, and let x_0 be a point in X We say that x_0 is an adherent point of E if for every radius $r > 0$, the ball $B(x_0, r)$ has a non-empty intersection with E. The set of all adherent points of E is called the closure of E and is denoted $\overline{E}$.

- **Open and closed sets** Let (X, d) be a metric space, and let E be a subset of X. We say that E is closed if it contains all of its boundary points. We say that E is open if it contains non of its boundary points. If E contains some of its boundary points but not others, then it is neither open nor closed.

- **Complete metric spaces** A metric space (X, d) is said to be complete iff every Cauchy sequence in (X, d) is in fact convergent in (X, d).

- **Compactness** A metric space (X, d) is said to be compact iff every sequence in (X, d) has at least one convergent subsequence. A subset Y of a metric space X is said to be compact if the subspace $(Y, d|_{Y \times Y})$ is compact.

- **Bounded sets** Let (X, d) be a metric space, and let Y be a subset of X. We say that Y is bounded iff there exists a ball $B(x, r)$ in X which contains Y.

- **Heine-Borel theorem** Let $(\boldsymbol{R}^n, d)$ be a Euclidean space with either the Euclidean metric, the taxicab metric, or the sup norm metric. Let E be a subset of $\boldsymbol{R}^n$. Then E is compact iff it is closed and bounded.

We work with extended real numbers. However, the laws of cancellation do not apply once some variables are allowed to be infinite. We also note that once one adopts the convention $\infty.0 = 0.\infty = 0$, then multiplication becomes **upward continuous**, in the sense that whenever $x_n \in [0, +\infty]$ increases to $x \in [0, +\infty]$, and

$y_n \in [0,+\infty]$ increases to $y \in [0,+\infty]$, then $x_n y_n$ increases to xy. But it is not downward continuous, e.g., $1/n \to 0$ but $1/n.\infty \not\to 0.\infty$. This asymmetry will ultimately cause us to define integration from below rather than from above, which leads to other asymmetries, e.g. monotone convergence theorem applies for monotone increasing functions, but not necessarily for monotone decreasing one.

If one wants to keep as many useful laws of algebra as one can, then one can add in infinity, or have negative numbers, but it is difficult to have both at the same time. Because of this tradeoff, we will see two overlapping types of measure and integration theory: the non-negative theory, which involves quantities taking values in $[0,+\infty]$, and the absolutely integrable theory, which involves quantities taking values in $(-\infty,+\infty)$ or $\boldsymbol{C}$. For instance, the fundamental convergence theorem for the former theory is the monotone convergence theorem, while the fundamental convergence theorem for the latter is the dominated convergence theorem.

One feature of the extended non-negative real axis is that all sums are convergent, and and infinite sum is the supremum of all finite subsums $\sum_{n=1}^{\infty} x_n = \sup_{F \subset \boldsymbol{N}, F \text{ finite}} \sum_{n \in F} x_n$, e.g., $\sum_{n=1}^{N} x_n$. Thus fo any collection $(a_\alpha)_{\alpha \in A}$ of numbers $x_\alpha \in [0,+\infty]$ indexed by an arbitrary set A, finite or infinite, countable or uncountable, we can define the sum $\sum_{\alpha \in A} x_\alpha$ by the formula $\sum_{\alpha \in A} x_\alpha = \sup_{F \subset A, F \text{ finite}} \sum_{\alpha \in F} x_\alpha$. From this definition one can relabel the collection in an arbitrary fashion without affecting the sum, i.e. given any bijection $\phi : B \to A$, one has the change of variables formula $\sum_{\alpha \in A} x_\alpha = \sum_{\beta \in B} x_{\phi(\beta)}$. When the sums are of signed elements, the above rearrangement identity can fail when the series is not absolutely convergent. This is called **Riemann rearrangement theorem**. **Tonelli's theorem for series**: Let $(x_{n,m})_{n,m \in \boldsymbol{N}}$ be a doubly infinite sequence of extended non-negative reals $x_{n,m} \in [0,+\infty]$. Then $\sum_{(n,m) \in \boldsymbol{N}^2} x_{n,m} = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} x_{n,m} = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} x_{n,m}$. In the signed case one needs an additional assumption of absolute convergence of series to interchange sums; this is **Fubini's theorem for series**.

Axiom of choice: Let $(E_\alpha)_{\alpha \in A}$ be a family of non-empty sets E_α, indexed by an index set A. Then we can find a family $(x_\alpha)_{\alpha \in A}$ of elements x_α of E_α, indexed by the same set A. When A is infinite, one cannot deduce this axiom from the other axioms of set theory. We note a theorem of **Godel** that states that any statement that can be phrased in the first-order language of Peano arithmetic, and which is proven with the axiom of choice, can also be proven without the axiom of choice. In other words, it is only when asking questions about infinite objects that are beyond the scope of Peano arithmetic that one can encounter statements that are provable using axiom of choice, but are not provable without it.

1 The problem of measure

One of the most fundamental concept in Euclidean geometry is that of the measure $m(E)$ of a solid body E in one ore more dimensions - length, area, or volume of E. In the classical approach to geometry, the measure of a body was often computed by partitioning that body into finitely many components, moving around each component by a rigid motion, and then reassembling those components to form a simpler body which presumably has the same measure. With the advent of analytical geometry, Euclidean geometry became reinterpreted as the study of Cartesian products $\boldsymbol{R}^d$ of the real line $\boldsymbol{R}$. It was no longer intuitively obvious how to define the measure $m(E)$ of a general subset E of $\boldsymbol{R}^d$. This is called the *problem of measure*. For example, set $A = [0,1]$ and $B = [0,2]$ can be mapped $A \to B$ using a bijection. So even though there is a one-to-one mapping the measures are obviously different. One can point to the infinite number of components as the cause of this breakdown of intuition. But Banach-Tarski paradox (relying on axiom of choice shows a unit ball in $\boldsymbol{R}^3$ can be disassembled into just five pieces and reassembled, after translating and rotating, to form two disjoint copies of the ball B) and similar examples forced to abandon the aim of measuring every possible subset E of $\boldsymbol{R}^d$, and instead to settle for only measuring a certain subclass of non-pathological subsets of $\boldsymbol{R}^d$ called *measurable sets*.

The first attempt is the **Jordan measure**, closely related to Riemann integral (or Barboux integral) which suffices for measuring most of the ordinary sets. However for more complicated sets in analysis and limits of other sets, Jordan measure turns out to be inadequate, and must be extended to the more general notion of Lebesgue measurability, with the corresponding **Lebesgue measure** that extends Jordan measure. Monotone and dominated convergence theorems which do not hold for Jordan measure are examples of this extension.

1.1 Elementary measure

To measure a very simple class of sets, namely the elementary sets (finite unions of boxes) we introduce the elementary measure.

Definition 1.1. *An interval is a subset of* $\boldsymbol{R}$ *of the form* $[a,b] := \{x \in \boldsymbol{R} : a \leq x \leq b\}$, $[a,b) := \{x \in \boldsymbol{R} : a \leq x < b\}$, $(a,b] := \{x \in \boldsymbol{R} : a < x \leq b\}$, *or* $(a,b) := \{x \in \boldsymbol{R} : a < x < b\}$, *where* $a \leq b$ *are real numbers. We define the length* $|I|$ *of an interval* $I = [a,b], [a,b), (a,b], (a,b)$ *to be* $|I| := b - a$. *A box in* $\boldsymbol{R}^d$ *is a Cartesian product* $B := I_1 \times \ldots \times I_d$ *of* d *intervals* $I_1, \ldots, I_d$. *The volume* $|B|$ *of such a box* B *is defined as* $|B| := |I_1| \times \ldots \times |I_d|$. *An elementary set is any subset of* $\boldsymbol{R}^d$ *which is the union of a finite number of boxes.*

Lemma 1.1. *Let* $E \subset \boldsymbol{R}^d$ *be an elementary set*

1. *E can be expressed as the finite union of disjoint boxes.*
2. *If* E *is partitioned as the finite union* $B_1 \cup \ldots \cup B_k$ *of disjoint boxes, then the quantity* $m(E) := |B_1| + \ldots + |B_k|$ *is independent of the partition.*

We refer to $m(E)$ *as the elementary measure of* E.

For any interval I, the length of I can be recovered by the limiting formula

$$|I| = \lim_{N\to\infty} \frac{1}{N}(I \cap \frac{1}{N}\boldsymbol{Z})$$

where $\frac{1}{N}\boldsymbol{Z} := \{\frac{n}{N} : n \in \boldsymbol{Z}\}$ and A denotes the cardinality of a finite set A. Taking Cartesian products, we see that

$$|B| = \lim_{N\to\infty} \frac{1}{N^d}(B \cap \frac{1}{N}\boldsymbol{Z}^d)$$

This makes precise an important intuition, that the continuous concept of measure can be viewed as a limit of the discrete concept of cardinality. We might be tempted to now define measure $m(E)$ of an arbitrary set $E \subset \boldsymbol{R}^d$ by the formula

$$m(E) := \lim_{N\to\infty} \frac{1}{N^d}(E \cap \frac{1}{N}\boldsymbol{Z}^d)$$

This would be valid for all Jordan measurable sets. But it is not satisfactory, in general, for two reasons. Firstly, the limit may not exist. Secondly, even if the limit exists, it may not obey reasonable properties such as translation invariance. For example, if $d = 1$ and $E := \boldsymbol{Q} \cap [0,1] : \{x \in \boldsymbol{Q} : 0 \leq x \leq 1\}$, then this definition would give E a measure of 1, but would give the translation $E + \sqrt{2} := \{x + \sqrt{2} : x \in \boldsymbol{Q}; 0 \leq x \leq 1\}$ a measure of zero.

From the definitions, it is clear that $m(E)$ is non-negative real number for every elementary set E, and that $m(E \cup F) = m(E) + m(F)$ whenever E and F are disjoint elementary sets. We refer to this property as *finite additivity*; by induction it also implies that $m(E_1 \cup \ldots \cup E_k) = m(E_1) + \ldots + m(E_k)$ whenever $E_1, \ldots, E_k$ are disjoint elementary sets. We also have the obvious degenerate case $m(\emptyset) = 0$. Finally, elementary measure clearly extends the notion of volume, in the sense that $m(B) = |B|$ for all boxes B. From non-negativity and finite additivity we conclude the monotonicity property $m(E) \leq m(F)$ whenever $E \subset F$ are nested elementary sets. From this and finite additivity we easily obtain the *finite subadditivity* property $m(E \cup F) \leq m(E) + m(F)$, whenever E, F are elementary set which are not necessarily disjoint. Again, by induction one has $m(E_1 \cup \ldots \cup E_k) \leq m(E_1) + \ldots + m(E_k)$ whenever $E_1, \ldots, E_k$ are elementary sets, not necessarily disjoint. It is also clear from the definition that we have the *translation invariance* $m(E + x) = m(E)$ for all elementary sets E and $x \in \boldsymbol{R}^d$. In fact these properties define a unique elementary measure up to normalization.

1.2 Jordan measure

The elementary sets are very restrictive and can not measure even a rotated box. However such sets E can be approximated from within and without by elementary sets $A \subset E \subset B$, such that the inscribing elementary set a and the circumscribing elementary set B can be used to give lower and upper bounds on the measure of E. With finer sets A and B, one can hope that these two bounds eventually match.

Definition 1.2. *(Jordan measure) Let $E \subset \boldsymbol{R}^d$ be a bounded set. The Jordan inner measure of E is defined as $m_{*,(J)}(E) := \sup_{A \subset E,\ A \text{ elementary}} m(A)$. The Jordan outer measure of E is defined as $m^{*,(J)}(E) := \inf_{B \supset E,\ B \text{ elementary}} m(B)$. If $m_{*,(J)}(E) = m^{*,(J)}(E)$, then we say that E is Jordan measurable, and call it the Jordan measure of E.*

We can write characterization of Jordan measurability as follows. Let $E \subset \boldsymbol{R}^d$ be bounded. By convention, we do not consider unbounded sets to be Jordan measurable. Then we can show that E is Jordan measurable $\iff$ for every $\varepsilon > 0$, there exit elementary sets $A \subset E \subset B$ such that $m(B \backslash A) \leq \varepsilon \iff$ for every $\varepsilon > 0$, there exists an elementary set A such that $m^{*,(J)}(A \Delta E) \leq \varepsilon$. Every elementary set E is Jordan measurable, and the Jordan measure and elementary measure coincide for such sets. Jordan measure also follows the properties of Boolean closure, non-negativity, finite additivity, monotonicity, finite subadditivity and translation invariance. It is also unique up to a normalization constant.

Unbounded sets are not Jordan measurable, but not every bounded subset of $\boldsymbol{R}^d$ is Jordan measurable either. The bullet-riddled square $[0,1]^2 \backslash \boldsymbol{Q}^2$, and the set of bullets $[0,1]^2 \cap \boldsymbol{Q}^2$, both have Jordan inner measure zero and Jordan outer measure one. In particular, both sets are not Jordan measurable. Informally, any set with a lot of holes or very fractal boundary, is unlikely to be Jordan measurable. In order to measure such sets we will need to develop Lebesgue measure. The following *Caratheodory* type property will come handy later: Let $E \subset \boldsymbol{R}^d$ be a bounded set and $F \subset \boldsymbol{R}^d$ be an elementary set. Then, $m^{*,(J)}(E) = m^{*,(J)}(E \cap F) + m^{*,(J)}(E \backslash F)$.

1.3 Connection with the Riemann integral

Definition 1.3. *(Riemann integrability) Let $[a,b]$ be an interval of positive length, and $f : [a,b] \to \boldsymbol{R}$ be a function. A tagged partition $\mathcal{P} = ((x_0, x_1, \ldots, x_n), (x_1^*, \ldots, x_n^*))$ of $[a,b]$ is a finite sequence of real numbers $a = x_0 < x_1 < \ldots < x_n = b$, together with additional numbers $x_{i-1} \leq x_i^* \leq x_i$ for each $i = 1, \ldots, n$. We abbreviate $x_i - x_{i-1}$ as δx_i. The quantity $\Delta(\mathcal{P}) := \sup_{1 \leq i \leq n} \delta x_i$ will be called the norm of the tagged partition. The Riemann sum $\mathcal{R}(f, \mathcal{P})$ of f with respect to the tagged partition $\mathcal{P}$ is defined as $\mathcal{R}(f, \mathcal{P}) := \sum_{i=1}^{n} f(x_i^*) \delta x_i$. We say that f is Riemann integrable on $[a,b]$ if there exists a real number, denoted $\int_a^b f(x)dx$ and referred to as the Riemann integral of f on $[a,b]$, for which we have $\int_a^b f(x)dx = \lim_{\Delta(\mathcal{P}) \to 0} \mathcal{R}(f, \mathcal{P})$ by which we mean that for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|\mathcal{R}(f, \mathcal{P}) - \int_a^b f(x)dx| \leq \varepsilon$ for every tagged partition $\mathcal{P}$ with $\Delta(\mathcal{P}) \leq \delta$.*

If $[a,b]$ is an interval of zero length, we adopt the convention that every function $f : [a,b] \to \boldsymbol{R}$ is Riemann integrable, with a Riemann integral of zero. Note that unbounded functions cannot be Riemann integrable. A more convenient formulation of the Riemann integral can be formulated using the machinery of piecewise constant functions via the Darboux formulation.

Definition 1.4. *(Piecewise constant functions) Let $[a,b]$ be an interval. A piecewise constant function $f : [a,b] \to \boldsymbol{R}$ is a function for which there exists a partition $[a,b]$ into finitely many intervals $I_1, \ldots, I_n$, such that f is equal to a constant c_i on each of the intervals I_i. Further, if f is piecewise constant, then $\sum_{i=1}^{n} c_i |I_i|$ is independent of the choice of partition used. We denote this quantity by p.c.$\int_a^b f(x)dx$, and refer to it as the piecewise constant integral of f on $[a,b]$.*

Lemma 1.2. *(basic properties of the piecewise constant integral) Let $[a,b]$ be an interval, and let $f, g : [a,b] \to \boldsymbol{R}$ be piecewise constant functions. Then,*

1. *Linearity: For any real number c, cf and $f+g$ are piecewise constant, with p.c.$\int\limits_a^b f(x)+g(x)dx =$ p.c.$\int\limits_a^b f(x)dx +$ p.c.$\int\limits_a^b g(x)dx$.*

2. *Monotonicity: If $f \leq g$ pointwise, i.e. $f(x) \leq g(x), \forall x \in [a,b]$ then p.c.$\int\limits_a^b f(x) \leq$ p.c.$\int\limits_a^b g(x)dx$.*

3. *Indicator: If E is an elementary subset of $[a,b]$, then the indicator function $1_E : [a,b] \to \boldsymbol{R}$ defined by $1_E := \begin{cases} 1 & x \in E \\ 0 & \text{otherwise} \end{cases}$ is piecewise constant, and p.c.$\int\limits_a^b 1_E(x)dx = m(E)$.*

Definition 1.5. *(Darboux integral) Let $[a,b]$ be an interval, and $f : [a,b] \to \boldsymbol{R}$ be a bounded function. Then lower Darboux integral of f on $[a,b]$ is defined as $\underline{\int\limits_a^b} f(x)dx := \sup\limits_{g \leq f, p.c.}$ p.c.$\int\limits_a^b g(x)dx$, where g ranges over all piecewise constant functions that are pointwise bounded above by f. Similarly, we define the upper Darboux integral of f on $[a,b]$ by the formula $\overline{\int\limits_a^b} f(x)dx := \inf\limits_{h \geq f, p.c.}$ p.c.$\int\limits_a^b h(x)dx$. Clearly, $\underline{\int\limits_a^b} f(x)dx \leq \overline{\int\limits_a^b} f(x)dx$. If these two quantities are equal, we say that f is Darboux integrable, and refer to this quantity as the Darboux integral of f on $[a,b]$.*

For an interval $[a,b]$, and a bounded function $f : [a,b] \to \boldsymbol{R}$, f is Riemann integrable iff it is Darboux integrable, in which case the Riemann integral and Darboux integrals are equal. Further, any continuous function $f : [a,b] \to \boldsymbol{R}$ is Riemann integrable. More generally, any bounded, piecewise continuous function $f : [a,b] \to \boldsymbol{R}$ is Riemann integrable. Now we connect the Riemann integral to Jordan measure.

Lemma 1.3. *(Basic properties of the Riemann integral) Let $[a,b]$ be an interval, and let $f, g : [a,b] \to \boldsymbol{R}$ be Riemann integrable. Then*

- *Linearity: For any real number c, cf and $f+g$ are Riemann integrable, with $\int\limits_a^b cf(x)dx = c\int\limits_a^b f(x)dx$ and $\int\limits_a^b (f(x)+g(x))\,dx = \int\limits_a^b f(x)dx + \int_a^b g(x)dx$.*

- *Monotonicity: If $f \leq g$ pointwise, i.e. $f(x) \leq g(x)$ for all $x \in [a,b]$ then $\int\limits_a^b f(x)dx \leq \int\limits_a^b g(x)dx$.*

- *Indicator: If E is a Jordan measurable on $[a,b]$, then the indicator function $1_E : [a,b] \to \boldsymbol{R}$, defined by setting $1_E(x) := 1$ when $x \in E$ and $1_E(x) := 0$ otherwise, is Riemann integrable, and $\int\limits_a^b 1_E(x)dx = m(E)$.*

These properties uniquely define the Riemann integral, in the sense that that functional $f \mapsto \int\limits_a^b f(x)dx$ is the only map from the space of Riemann integrable functions on $[a,b]$ to $\boldsymbol{R}$ which obeys all three of the above properties. We can extend this to higher dimensions. For two dimensions we have the area interpretation of the Riemann integral. Let $[a,b]$ be an interval, and let $f : [a,b] \to \boldsymbol{R}$ be a bounded function. f is Riemann integrable iff the sets $E_+ := \{(x,t) : x \in [a,b]; 0 \leq t \leq f(x)\}$ and $E_- := \{(x,t) : x \in [a,b]; f(x) \leq t \leq 0\}$ are both Jordan measurable in $\boldsymbol{R}^2$, in which case one has $\int\limits_a^b f(x)dx = m^2(E_+) - m^2(E_-)$, where m^2 denotes the two-dimensional Jordan measure.

2 Lebesgue measure

Not all sets are Jordan measurable, even if one restricts attention to bounded sets. There even exist compact (bounded open) sets that are not Jordan measurable. Another class that it fails to cover is countable unions or intersections of sets that are already known to be measurable. This creates problems with Riemann integrability and point-wise limits.

The Jordan outer measure is $m^{*,(J)}(E) := \inf_{B \supset E;\ B \text{ elementary}} m(B)$ of a set $E \subset \boldsymbol{R}^d$. We adopt the convention that $m^{*,(J)}(E) = +\infty$ if E is unbounded. Observe that from the finite additivity and subadditivity of elementary measure we can also write the Jordan outer measure as $m^{*,(J)}(E) := \inf_{B_1 \cup \ldots \cup B_k \supset E;\ B_1,\ldots,B_k \text{ boxes}} |B_1| + \ldots + |B_k|$ i.e. the Jordan outer measure is the infimal cost required to cover E by a finite union of boxes. We now modify this by replacing the finite union of boxes by a countable union of boxes, leading to the Lebesgue outer measure, $m^*(E) := \inf_{\bigcup_{n=1}^{\infty} B_n \supset E;\ B_1, B_2, \ldots \text{ boxes}} \sum_{n=1}^{\infty} |B_n|$, thus the Lebesgue outer measure is the infimal cost required to cover E by a countable union of boxes. This countable sum could be infinite and so the Lebesgue outer measure $m^*(E)$ could well equal $+\infty$. Clearly, we always have $m^*(E) \leq m^{*,(J)}(E)$.

Example 2.1. In one dimension, $m^{*,(J)}(\boldsymbol{Q})$ is infinite, and $m^{*,(J)}(\boldsymbol{Q} \cap [-R, R]) = m^{*,(J)}([-R, R]) = 2R$. One the other hand, all countable sets E have Lebesgue outer measure zero. Indeed one simply cover $E = \{x_1, x_2, \ldots\} \subset \boldsymbol{R}^d$ by the degenerate boxes $\{x_1\}, \{x_2\}, \ldots$ of sidelength and volume zero. This is precisely because we allow countable number of sets in Lebesgue measure but only allow finite number of sets in Jordan measure.

$\boldsymbol{\varepsilon/2^n}$ **trick**: Alternatively, one can cover each x_n by a cube B_n of sidelength $\varepsilon/2^n$ for some arbitrary $\varepsilon > 0$, leading to total cost of $\sum_{n=1}^{\infty} (\varepsilon/2^n)^d$, which converges to $C_d \varepsilon^d$ for some absolute constant C_d. A ε can be arbitrarily small, we see that the Lebesgue out measure must be zero. □

From this example we see that a set may be unbounded while still having Lebesgue outer measure zero, in contrast to Jordan outer measure. In analogy to Jordan theory, we would also like to define a concept of 'Lebesgue inner measure' to complement that of outer measure. The previous example makes it clear that there is asymmetry here between the inner Jordan and a possibly equivalent inner Lebesgue measure, even when we are willing to replace finite unions of boxes with countable ones. This highlights the fundamental issue with Jordan measure. We need new inspiration.

One can get a sort of Lebesgue inner measure by taking complements. Let $E \subset \boldsymbol{R}^d$ be a bounded set. We can define Lebesgue inner measure $m_*(E) := m(A) - m^*(A \backslash E)$ for any elementary set A containing E, i.e. if A and A' are two elementary sets containing E, then $m(A) - m^*(A \backslash E) = m(A') - m^*(A' \backslash E)$. In general, $m_*(E) \leq m^*(E)$ and the equality holds if and only if E is Lebesgue measurable. This leads to one possible definition of Lebesgue measurability, namely the Caratheodory criterion for Lebesgue measurability. Which states that: E is Lebesgue measurable $\iff$ For every elementary set A, one has $m(A) = m^*(A \cap E) + m^*(A \backslash E)$ $\iff$ For every box B, one has $|B| = m^*(B \cap E) + m^*(B \backslash E)$.

Example 2.2. Continuing with the example of $E = \boldsymbol{Q} \cap [-R, R]$, we choose $A = [-R, R]$, so that $A \supseteq E$ is satisfied. This implies $A \backslash E = (\boldsymbol{R} \backslash \boldsymbol{Q}) \cap [-R, R]$. Now, since A is an elementary set we have $m(A) = 2R$. Further we have $m^*(A \backslash E) = m^*((\boldsymbol{R} \backslash \boldsymbol{Q}) \cap [-R, R]) = 2R$ as well. This is evident from the definition of Lebesgue outer measure. Hence, we have by definition of Lebesgue inner measure $m_*(E) = 0$, which matches the outer measure ($m^*(E) = 0$) and hence the Lebesgue measure is defined on this set and this set is Lebesgue measurable. □

However this is not the most intuitive formulation of this concept to work with, and we will instead use a different, but logically equivalent, definition of Lebesgue measurability. We start with the observation that the Jordan measurable sets can be efficiently contained in elementary sets, with an error that has small Jordan outer measure. In a similar vein, we will define Lebesgue measurable sets to be sets that can be efficiently contained in open sets, with an error that has small Lebesgue outer measure.

Definition 2.1. *(Lebesgue measurability) A set $E \subset \boldsymbol{R}^d$ is said to be Lebesgue measurable if, for every $\varepsilon > 0$, there exists an open set $U \subset \boldsymbol{R}^d$ containing E such that $m^*(U \backslash E) \leq \varepsilon$. If E is Lebesgue measurable, we refer to $m(E) := m^*(E)$ as the Lebesgue measure of E, which may be equal to $+\infty$.*

The intuition that measurable sets are almost open is also know as Littlewood's first principle. Lebesgue measure extends Jordan measure, in the sense that every Jordan measurable set is Lebesgue measurable, and the Lebesgue measure and Jordan measure of a Jordan measurable set are always equal. Some pathological sets constructed using axiom of choice are still not Lebesgue measurable.

2.1 Properties of Lebesgue outer measure

We study the Lebesgue outer measure m^*, which takes values in $[0, +\infty]$.

Axiom 2.1. *(The outer measure axioms)*

- *Empty set:* $m^*(\emptyset) = 0$.
- *Monotonicity: If* $E \subset F \subset \boldsymbol{R}^d$*, then* $m^*(E) \leq m^*(F)$.
- *Countable subadditivity: If* $E_1, E_2, \ldots \subset \boldsymbol{R}^d$ *is a countable sequence of sets, then* $m^*(\bigcup\limits_{n=1}^{\infty} E_n) \leq \sum\limits_{n=1}^{\infty} m^*(E_n)$.

Countable subadditivity, when combined with the empty set axiom, gives finite subadditivity property. Lebesgue outer measure is a model example of an abstract outer measure on a general set X, which is an assignment $E \mapsto m^*(E)$ of elements of $[0, +\infty]$ to arbitrary subsets E of a space X that obeys the above three axioms. Jordan outer measure will not be an abstract outer measure as it is only finitely subadditive rather than countably subadditive. For example, the rationals $\boldsymbol{Q}$ have infinite Jordan outer measure, despite being the countable union of points, each of which have a Jordan outer measure of zero. One *cannot hope to upgrade countable subadditivity to uncountable subadditivity*: $\boldsymbol{R}^d$ is an uncountable union of points, each of which has Lebesgue outer measure zero, but $\boldsymbol{R}^d$ has infinite Lebesgue outer measure.

Lemma 2.1. *(Finite additivity for disjoint sets) Let* $E, F \subset \boldsymbol{R}^d$ *be such that* $dist(E, F) > 0$*, where* $dist(E, F) := inf\{|x - y| : x \in E, y \in F\}$ *is the distance between* E *and* F*. Then* $m^*(E \cup F) = m^*(E) + m^*(F)$.

We now calculate the Lebesgue outer measure of some other sets.

Lemma 2.2. *(Outer measure of elementary sets) Let* E *be an elementary set. Then the Lebesgue outer measure* $m^*(E)$ *of* E *is equal to the elementary measure* $m(E)$ *of* $E : m^*(E) = m(E)$.

To prove this lemma we make use of Heine-Borel theorem, which ultimately exploits the important fact that the reals are complete. This allows us to compute the Lebesgue outer measure of a finite union of boxes. From this and monotonicity we conclude that the Lebesgue outer measure of any set is bounded below by its Jordan inner measure. As it is also bounded above by the Jordan outer measure, we have $m_{*,(J)}(E) \leq m^*(E) \leq m^{*,(J)}(E)$ for every $E \subset \boldsymbol{R}^d$. We can now explain why not every bounded open set or compact set is Jordan measurable. Consider the countable set $\boldsymbol{Q} \cap [0, 1]$ enumerated as $\{q_1, q_2, q_3, \ldots\}$. For $\varepsilon > 0$ we consider the set $U := \bigcup\limits_{n=1}^{\infty} (q_n - \varepsilon/2^n, q_n + \varepsilon/2^n)$. This is the union of open sets and is thus open. By countable subadditivity, on has $m^*(U) \leq \sum\limits_{n=1}^{\infty} 2\varepsilon/2^n = 2\varepsilon$. Finally, as U is dense in $[0, 1]$, i.e. $\overline{U}$ contains $[0, 1]$, we have $m^{*,(J)}(U) = m^{*,(J)}(\overline{U}) \geq m^{*,(J)}([0, 1]) = 1$. For small enough ε we see that the Lebesgue outer measure and Jordan outer measure of U disagree. We, thus, conclude that that the bounded open set U is not Jordan measurable. This also implies that the complement of U in, say $[-2, 2]$ is also not Jordan measurable, despite being a compact set.

Turning to countable union of boxes, it is convenient to introduce the notion of almost disjoint boxes if their interiors are disjoint. Since a box has the same elementary measure as its interior, we see that the finite additivity property $m(B)1 \cup \ldots, \cup B_k) = |B_1| + \ldots + |B_k|$ holds for almost disjoint boxes $B_1, \ldots, B_k$ and not just for disjoint boxes.

Lemma 2.3. *(Outer measure of countable unions of almost disjoint boxes) Let* $E = \bigcup\limits_{n=1}^{\infty} B_n$ *be a countable union of almost disjoint boxes* $B_1, B_2, \ldots$*. Then* $m^*(E) = \sum\limits_{n=1}^{\infty} |B_n|$.

For example, $\boldsymbol{R}^d$ itself has an infinite outer measure. If $E = \bigcup\limits_{n=1}^{\infty} B_n = \bigcup\limits_{n=1}^{\infty} B'_n$ can be decomposed into different ways as the countable union of almost disjoint boxes, then $\sum\limits_{n=1}^{\infty} |B_n| = \sum\limits_{n=1}^{\infty} |B'_n|$. Here, we also have the Lebesgue outer measure of E equal to the Jordan inner measure $m^*(E) = m_{*,(J)}(E)$. However, not every set can be expressed as the countable union of almost disjoint boxes, e.g. the set $\boldsymbol{R} \backslash \boldsymbol{Q}$, which contain no boxes other than the singleton sets. However, there is an important class of sets are called **open sets**.

Lemma 2.4. *Let $E \subset \boldsymbol{R}^d$ be an open set. Then E can be expressed as the countable union of almost disjoint boxes, and in fact, as the countable union of almost disjoint closed cubes.*

This can be shown using the construction of **dyadic cube** of the form $Q = \left[\frac{i_1}{2^n}, \frac{i_1+1}{2^n}\right] \times \ldots \times \left[\frac{i_d}{2^n}, \frac{i_d+1}{2^n}\right]$, for some integers $n, i_1, \ldots, i_d$, with side lengths at most 1 by restricting n to be non-negative integers. And using the dyadic nesting property: given any two closed dyadic cubes, either they are almost disjoint, or one of them is contained in the other - we can prove this lemma.

We now have a formula for the Lebesgue outer measure of any open set: it is exactly equal to the Jordan inner measure of that set, or the total volume of any partitioning of that set into almost disjoint boxes. We, thus, have a formula for the Lebesgue outer measure of any arbitrary set:

Lemma 2.5. *(Outer regularity) Let $E \subset \boldsymbol{R}^d$ be an arbitrary set. Then one has $m^*(E) = \inf_{E \subset U, U \text{ open}} m^*(U)$.*

Note that the reverse statement $m^*(E) = \sum_{U \subset E, U \text{ open}} m^*(U)$ is false.

2.2 Lebesgue measurability

Lemma 2.6. *(Existence of Lebesgue measurable sets)*

- *Every open set is Lebesgue measurable.*
- *Every closed set is Lebesgue measurable.*
- *Every set of Lebesgue outer measure zero is measurable, called null sets.*
- *The empty set $\emptyset$ is Lebesgue measurable.*
- *If $E \subset \boldsymbol{R}^d$ is Lebesgue measurable, then so its complement $\boldsymbol{R}^d \backslash E$.*
- *If $E_1, E_2, E_3, \ldots \subset \boldsymbol{R}^d$ are a sequence of Lebesgue measurable sets, then the union $\bigcup_{n=1}^{\infty} E_n$ is Lebesgue measurable.*
- *If $E_1, E_2, E_3, \ldots \subset \boldsymbol{R}^d$ are a sequence of Lebesgue measurable sets, then the intersection $\bigcap_{n=1}^{\infty} E_n$ is Lebesgue measurable.*

Informally, if one starts with such basic subsets of $\boldsymbol{R}^d$ as open or closed sets and then takes at most countably many boolean operations, one will always end up with a Lebesgue measurable set. Nevertheless, using the axiom of choice one can construct sets that are not Lebesgue measurable. As a consequence, we cannot generalize the countable closure properties here to uncountable closure properties. The properties $4, 5, 6$ in the previous lemma assert that the collection of Lebesgue measurable subsets of $\boldsymbol{R}^d$ form a σ-algebra, which is a strengthening of the more classical concept of a boolean algebra. Also, note that this lemma is significantly stronger than the counterpart for Jordan measurability, in particular by allowing countably many boolean operations instead of just finitely many. The following criteria for measurability applies for $E \subset \boldsymbol{R}^d$, which are equivalent

- E is Lebesgue measurable.
- Outer approximation by open: For every $\varepsilon > 0$, one can contain E in an open set U with $m^*(U \backslash E) \leq \varepsilon$.
- Almost open: For every $\varepsilon > 0$, one can find an open set U such that $m^*(U \Delta E) \leq \varepsilon$.
- Inner approximation by closed: For every $\varepsilon > 0$, one can find a closed set F contained in E with $m^*(E \backslash F) \leq \varepsilon$.
- Almost closed: For every $\varepsilon > 0$, one can find a closed set F such that $m^*(F \Delta E) \leq \varepsilon$.
- Almost measurable: For every $\varepsilon > 0$, one can find a Lebesgue measurable set E_ε such that $m^*(E_\varepsilon \Delta E) \leq \varepsilon$.

Now we look at the Lebesgue measure $m(E)$ of a Lebesgue measurable set E, which is defined to equal its Lebesgue outer measure $m^*(E)$. If E is Jordan measurable, then the Lebesgue measure and Jordan measure of E coincide. Thus Lebesgue measure is an extension of Jordan measure, which is an extension of elementary measure. Lebesgue measure obeys significantly better properties than Lebesgue outer measure, when restricted to Lebesgue measurable sets.

Lemma 2.7. *(The measure axioms)*

- *Empty set:* $m(\emptyset) = 0$.
- *Countable additivity: If* $E_1, E_2, \ldots \subset \boldsymbol{R}^d$ *is a countable sequence of disjoint Lebesgue measurable sets, then* $m(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} m(E_n)$.

Theorem 2.1. *(Monotone convergence theorem for measurable sets)*

- *Upward monotone convergence: Let* $E_1 \subset E_2 \ldots \subset \boldsymbol{R}^n$ *be a countable non-decreasing sequence of Lebesgue measurable sets. Then,* $m(\bigcup_{n=1}^{\infty} E_n) = \lim_{n\to\infty} m(E_n)$.
- *Downward monotone convergence: Let* $\boldsymbol{R}^d \supset E_1 \supset E_2 \supset \ldots$ *be a countable non-increasing sequence of Lebesgue measurable sets. Then, if one of the* $m(E_n)$ *is finite, then* $m(\bigcap_{n=1}^{\infty} E_n) = \lim_{n\to\infty} m(E_n)$.

We say that a sequence of sets in $\boldsymbol{R}^d$ converges pointwise to another set E in $\boldsymbol{R}^d$ if the indicator functions 1_{E_n} converge pointwise to 1_E. If E_n are all Lebesgue measurable, and converge pointwise to E, then E is Lebesgue measurable too.

Theorem 2.2. *(Dominated convergence theorem) Suppose that the* E_n *are all contained in another Lebesgue measurable set* F *of finite measure. Then* $m(E_n)$ *converges to* $m(E)$.

Later we will generalize the monotone and dominated convergence theorems to measurable functions instead of measurable sets. The *inner regularity* property of Lebesgue measurable set $E \subset \boldsymbol{R}^d$ can be seen from the definition $m(E) = \sup_{k \subset E, K \text{ compact}} m(K)$. The inner and outer regularity properties of measure can be used to define the concept of *Radon measure*.

Theorem 2.3. *(Caratheodory criterion) Let* $E \subset \boldsymbol{R}^d$. *The following are equivalent:*

- E *is Lebesgue measurable.*
- *For every elementary set* A, *one has* $m(A) = m^*(A \cap E) + m^*(A \backslash E)$.
- *For every box* B, *one has* $|B| = m^*(B \cap E) + m^*(B \backslash E)$.

Definition 2.2. *(Inner measure) Let* $E \subset \boldsymbol{R}^d$ *be a bounded set. We define the Lebesgue inner measure* $m_*(E)$ *of* E *by the formula* $m_*(E) := m(A) - m^*(A \backslash E)$ *for any elementary set* A *containing* E.

We can show that $m_*(E) \leq m^*(E)$, and the equality holds iff E is Lebesgue measurable. We can also show that if $E \subset \boldsymbol{R}^d$ is Lebesgue measurable, then $E + x$ is Lebesgue measurable for any $x \in \boldsymbol{R}^d$ and that $m(E+x) = m(E)$, which means it is translation invariant.

Theorem 2.4. *(Uniqueness of Lebesgue measure) Lebesgue measure* $E \mapsto m(E)$ *is the only map from Lebesgue measurable sets to* $[0, +\infty]$ *that obeys the following axioms:*

- *Empty set:* $m(\emptyset) = 0$.
- *Countable additivity: If* $E_1, E_2, \ldots \subset \boldsymbol{R}^d$ *is a countable sequence of disjoint Lebesgue measurable sets, then* $m(\bigcup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} m(E_n)$
- *Translation invariance: If* E *is Lebesgue measurable and* $x \in \boldsymbol{R}^d$, *then* $m(E + x) = m(E)$.
- *Normalization:* $m([0,1]^d) = 1$

2.3 Non-measurable sets

A famous theorem of Solovay asserts that, if one is willing to drop the axiom of choice, there exist models of set theory in which all subsets of $\boldsymbol{R}^d$ are measurable. So any demonstration of the existence of non-measurable sets must use the axiom of choice in some essential way.

Proposition 2.1. *There exists a subset $E \subset [0,1]$ which is not Lebesgue measurable.*

Proof: We create the quotient group $\boldsymbol{R}/\boldsymbol{Q} := \{x + \boldsymbol{Q} : x \in \boldsymbol{R}\}$. Each coset C of $\boldsymbol{R}/\boldsymbol{C}$ is dense in $\boldsymbol{R}$, and so has a non-empty intersection with $[0,1]$. Applying axiom of choice, we may thus find an element $x_C \in C \cap [0,1]$ for each $C \in \boldsymbol{R}/\boldsymbol{Q}$. We then let $E := \{x_C : C \in \boldsymbol{R}/\boldsymbol{Q}\}$ be the collection of all these coset representatives. By construction $E \subset [0,1]$.

Let y be an element of $[0,1]$. Then it must lie in some coset C of $\boldsymbol{R}/\boldsymbol{Q}$, and thus differs from x_C by some rational number in $[-1,1]$. In other words, we have $[0,1] \subset \bigcup\limits_{q \in \boldsymbol{Q} \cap [-1,1]} (E+q)$. On the other hand, we clearly have $\bigcup\limits_{q \in \boldsymbol{Q} \cap [-1,1]} (E+q) \subset [-1,2]$. Also, the different translates $E+q$ are disjoint, because E contains only on element from each coset of $\boldsymbol{Q}$. This set E is claimed not to be Lebesgue measurable.

By contradiction, suppose that E was Lebesgue measurable. Then the translates $E+q$ would also be Lebesgue measurable. By countable additivity, we thus have $m(\bigcup\limits_{q \in \boldsymbol{Q} \cap [-1,1]} (E+q)) = \sum\limits_{q \in \boldsymbol{Q} \cap [-1,1]} m(E+q)$, and thus by translation invariance $1 \leq \sum\limits_{q \in \boldsymbol{Q} \cap [-1,1]} m(E) \leq 3$. On the other hand, the sum $\sum\limits_{q \in \boldsymbol{Q} \cap [-1,1]} m(E)$ is either zero (if $m(E) = 0$) or infinite (if $m(E) > 0$), leading to desired contradiction. $\square$

Thus, in the presence of the axiom of choice, one cannot hope to extend Lebesgue measure to arbitrary subsets of $\boldsymbol{R}$ while retaining both the countable additivity and the translation invariance properties.

3 The Lebesgue integral

We now use Lebesgue measure $m(E)$ to define the Lebesgue integral $\int\limits_{\boldsymbol{R}^d} f(x)dx$ of functions $f : \boldsymbol{R}^d \to \boldsymbol{C} \cup \{\infty\}$. The function need to be Lebesgue measurable to be Lebesgue integrable. Further, the function either need to be unsigned (taking values $[0,+\infty]$), or absolutely integrable. The idea of infinite summation $\sum\limits_{n=1}^{\infty} c_n$ with as sequence of numbers c_n can be viewed as a discrete analogue of the Lebesgue integral. If c_n lie in the extended non-negative real axis $[0,+\infty]$ we have a case of *unsigned infinite sum* and the sum can be defined as a limit of the partial sums $\lim\limits_{N \to \infty} \sum_{n=1}^{N} c_n$ or equivalently as a supremum of arbitrary finite partial sums $\sup\limits_{A \subset \boldsymbol{N}, A \text{ finite}} \sum\limits_{n \in A} c_n$. The unsigned infinite sum always exists, though its value may be infinite even when each term is individually finite. On the other hand, if c_n lie in the complex plane $\boldsymbol{C}$ and obey the absolute summability condition $\sum\limits_{n=1}^{\infty} |c_n| < \infty$, we have *absolutely summable infinite series.* Here, the partial sums $\sum\limits_{n=1}^{N} c_n$ converge to a limit and the infinite sum now takes values in $\boldsymbol{C}$ rather than $[0,+\infty]$. The value of an absolutely convergent sum is unchanged if one rearranges the terms in the series in an arbitrary fashion and we can define the absolutely summable infinite series in terms of the unsigned infinite sums by using $\sum\limits_{n=1}^{\infty} c_n = \sum\limits_{n=1}^{\infty} \Re(c_n) + i \sum\limits_{n=1}^{\infty} \Im(C_n)$ for complex absolutely summable c_n, and $\sum\limits_{n=1}^{\infty} c_n = \sum\limits_{n=1}^{\infty} c_n^- \sum\limits_{n=1}^{\infty} c_n^-$ for real absolutely summable c_n, where $c_n^+ := max(c_n, 0)$ and $c_n^- := max(-c_n, 0)$ are the magnitudes of the positive and negative parts of c_n.

Analogously we first define an unsigned Lebesgue integral $\int\limits_{\boldsymbol{R}^d} f(x)dx$ of measurable unsigned functions $f : \boldsymbol{R}^d \to [0,+\infty]$, and then use that to define the absolutely convergent Lebesgue integral $\int\limits_{\boldsymbol{R}^d} f(x)dx$ of absolutely integrable function $f : \boldsymbol{R}^d \to \boldsymbol{C} \cup \{\infty\}$. In contrast to absolutely integrable series which cannot have any infinite terms, absolutely integrable functions will be allowed to occasionally become infinite, however only on a set of Lebesgue measure zero. To define unsigned Lebesgue integral we use the idea analogous to lower Darboux integral $\int_a^b f(x) = \underline{\int_a^b} f(x)dx := \sup_{g \leq f; g \text{ p.c.}} p.c. \int_a^b g(x)dx$. The integral $p.c. \int_a^b g(x)$ is a piecewise constant integral, formed by breaking up the piecewise constant functions g, h into finite linear combinations of indicator functions 1_I of intervals I, and then measuring the length of each interval. Similar definition allows use to define a lower Lebesgue integral $\int_{\boldsymbol{R}^d} f(x)dx$ of any unsigned function $f : \boldsymbol{R}^d \to [0, +\infty]$, simply by replacing intervals by more general class of Lebesgue measurable sets, i.e. replacing p.c. functions with simple functions. If the function is Lebesgue measurable, then we refer to the lower Lebesgue integral simply as the Lebesgue integral. If obeys the properties of monotonicity and additivity and behaves quite well with respect to limits as we will show via Fatou's lemma and monotone convergence theorem. After having dealt with unsigned Lebesgue integral, we define absolute convergent Lebesgue integral. This obeys the fundamentally important dominated convergence theorem.

The Lebesgue integral and the Lebesgue measure can be viewed as completions of the Riemann integral and Jordan measure respectively. Lebesgue theory extends the Riemann theory: every Jordan measurable set is Lebesgue measurable and every Riemann integrable function is Lebesgue measurable with the two measures and integrals compatible. The Lebesgue theory can be approximated by the Riemann theory: every Lebesgue measurable set can be approximated by simpler sets (open sets or elementary sets) and Lebesgue measurable functions can be approximated by nicer functions (Riemann integrable or continuous functions). The Lebesgue theory is complete in various ways as hinted by the convergence theorems. *Egorov's theorem* asserts that a pointwise converging sequence of functions can be approximated as a locally uniformly converging sequence of functions. These facts are captured by Littlewood's three principles of real analysis.

3.1 Integration of simple functions

Definition 3.1. *(Simple functions) A complex valued, simple function $f : \boldsymbol{R}^d \to \boldsymbol{C}$ is a finite linear combination $f = c_1 \mathbf{1}_{E_1} + \ldots + c_k \mathbf{1}_{E_k}$ of indicator functions $\mathbf{1}_{E_i}$ of Lebesgue measurable sets $E_i \subset \boldsymbol{R}^d$ for $i = 1, \ldots, k$, where $k \geq 0$ is a natural number and $c_1, \ldots, c_k \in \boldsymbol{C}$ are complex numbers. An unsigned simple function $f : \boldsymbol{R}^d \to [0, +\infty]$ is define similarly with c_i taking values in $[0, +\infty]$ rather than $\boldsymbol{C}$.*

The space $Simp(\boldsymbol{R}^d)$ of complex valued simple functions forms a complex vector space and is closed under pointwise product $f, g \mapsto fg$ and complex conjugation $f \mapsto \overline{f}$, thus forming a *commutative $\star$-algebra.* The space $Simp^+(\boldsymbol{R}^d)$ of unsigned simple functions is a $[0, +\infty]$-module, is closed under addition, and scalar multiplication by elements in $[0, +\infty]$. The set $E_1, \ldots, E_k$ may not be disjoint but we can use boolean algebra to make them so. For any k subsets $E_1, \ldots, E_k$ of $\boldsymbol{R}^d$ partition $\boldsymbol{R}^d$ into 2^k disjoint sets, each of which is an intersection of E_i or the complement $\boldsymbol{R}^d \backslash E_i$ for $i = 1, \ldots, k$ and is measurable. The complex or unsigned simple function is constant on each of these sets, and so can easily be decomposed as a linear combination of the indicator function on these sets. If f is a complex valued simple function, then its absolute value $|f| : x \mapsto |f(x)|$ is an unsigned simple function on this decomposition.

We, intuitively, define $\int_{\boldsymbol{R}^d} \mathbf{1}_E dx$ of an indicator function of a measurable set E to equal $m(E)$.

Definition 3.2. *(Integral of a unsigned simple function) If $f = c_1 \mathbf{1}_{E_1} + \ldots + \mathbf{1}_{E_k}$ is an unsigned simple function, the integral $Simp(\int_{\boldsymbol{R}^d} f(x)dx)$ is defined by the formula $Simp \int_{\boldsymbol{R}^d} f(x)dx := c_1 m(E) + \ldots + c_k m(E_k)$, thus $Simp(\int_{\boldsymbol{R}^d} f(x)dx)$ will take values in $[0, +\infty]$.*

Lemma 3.1. *(Well-definedness of simple integral) Let $k, k' \geq 0$ be natural numbers, $c_1, \ldots, c_k, c_1', , c_k' \in [0, +\infty]$, and let $E_1, \ldots, E_k, E_1', \ldots, E_{k'}' \subset \boldsymbol{R}^d$ be Lebesgue measurable sets such that the identity $c_1 \mathbf{1}_{E_1} + \ldots + c_k \mathbf{1}_{E_k} = c_1' \mathbf{1}_{E_1'} + \ldots + c_{k'}' \mathbf{1}_{E_{k'}'}$ holds identically on $\boldsymbol{R}^d$. Then one has $c_1 m(E_1) + \ldots + c_k m(E_k) = c_1' m(E_1') + \ldots + c_k' m(E_{k'}')$.*

Definition 3.3. *(Almost everywhere and support) A property $P(x)$ of a point $x \in \boldsymbol{R}^d$ is said to hold almost everywhere in $\boldsymbol{R}^d$, or for almost every point $x \in \boldsymbol{R}^d$, if the set of $x \in \boldsymbol{R}^d$ for which $P(x)$ fails has Lebesgue measure zero, i.e. P is true outside of a null set. This is abbreviated as a.e.. Two functions $f, g : \boldsymbol{R}^d \to Z$ into an arbitrary range Z are said to agree almost everywhere if one has $f(x) = g(x)$ for almost every $x \in \boldsymbol{R}^d$. The support of a function $f : \boldsymbol{R}^d \to \boldsymbol{C}$ or $f : \boldsymbol{R}^d \to [0, +\infty]$ is defined to be the set $\{x \in \boldsymbol{R}^d : f(x) \neq 0\}$ where f is non-zero.*

If $P(x)$ holds for almost every x, and $P(x) \implies Q(x)$, then $Q(x)$ holds for almost every x. Also, if $P_1(x), P_s(x), \ldots$ are an at most countable family of properties, each of which individually holds for almost every x, then they will simultaneously be true for almost every x, because the countable union of null sets is still a null set. The property of agreeing almost everywhere is an equivalence relation, referred to as almost everywhere equivalence. The closure of the support is called *closed support.*

Proposition 3.1. *(Basic properties of the simple unsigned integral) Let $f, g : \mathbf{R}^d \to [0, +\infty]$ be simple unsigned functions.*

- *Unsigned linearity: We have $Simp \int_{\mathbf{R}^d} f(x)+g(x)dx = Simp \int_{\mathbf{R}^d} f(x)dx + Simp \int_{\mathbf{R}^d} g(x)dx$ and $Simp \int_{\mathbf{R}^d} cf(x)dx = cSimp \int_{\mathbf{R}^d} f(x)dx$ for all $c \in [0, +\infty]$.*
- *Finiteness: We have $Simp \int_{\mathbf{R}^d} f(x)dx < \infty$ iff f is finite almost everywhere, and its support has finite measure.*
- *Vanishing: We have $Simp \int_{\mathbf{R}^d} f(x)dx = 0$ iff f is zero almost everywhere.*
- *Equivalence: If f and g agree almost everywhere, then $Simp \int_{\mathbf{R}^d} f(x)dx = Simp \int_{\mathbf{R}^d} g(x)dx$.*
- *Monotonicity: If $f(x) \leq g(x)$ for almost every $x \in \mathbf{R}^d$, then $Simp \int_{\mathbf{R}^d} f(x)dx \leq Simp \int_{\mathbf{R}^d} g(x)dx$.*
- *Compatibility with Lebesgue measure: For any Lebesgue measurable E, one has $Simp \int_{\mathbf{R}^d} 1_E(x)dx = m(E)$.*
- *Uniqueness: The simple unsigned integral $f \mapsto Simp \int_{\mathbf{R}^d} f(x)dx$ is the only map from the space $Simp^+(\mathbf{R}^d)$ of unsigned simple functions to $[0, +\infty]$ that obeys all the above properties.*

Definition 3.4. *(Absolutely convergent simple integral) A complex valued simple function $f : \mathbf{R}^d \to \mathbf{C}$ is said to be absolutely integrable if $Simp \int_{\mathbf{R}^d} |f(x)|dx < \infty$. If f is absolutely integrable, the integral $Simp \int_{\mathbf{R}^d} f(x)dx$ is defined for real signed f by the formula $Simp \int_{\mathbf{R}^d} f(x)dx := Simp \int_{\mathbf{R}^d} f_+(x)dx - Simp \int_{\mathbf{R}^d} f_-(x)dx$ where $f_+(x) := max(f(x), 0)$ and $f_-(x) := max(-fx(), 0)$, these functions have finite integrals, and for complex-valued f by the formula $Simp \int_{\mathbf{R}^d} f(x)dx := Simp \int_{\mathbf{R}^d} \Re f(x)dx + iSimp \int_{\mathbf{R}^d} \Im f(x)dx$.*

From the previous proposition we note that a complex-valued simple function f is absolutely integrable iff it has finite measure support. In particular, the space $Simp^{abs}(\mathbf{R}^d)$ of absolutely integrable simple functions is a complex vector space.

Proposition 3.2. *(Basic properties of the complex-values simple integral) Let $f, g : \mathbf{R}^d \to \mathbf{C}$ be absolutely integrable simple functions.*

- *Linearity: We have $Simp \int_{\mathbf{R}^d} f(x) + g(x)dx = Simp \int_{\mathbf{R}^d} f(x)dx + Simp \int_{\mathbf{R}^d} g(x)dx$ and $Simp \int_{\mathbf{R}^d} cf(x)dx = cSimp \int_{\mathbf{R}^d} f(x)dx$ for all $c \in \mathbf{C}$. Also we have $Simp \int_{\mathbf{R}^d} \overline{f(x)}dx = \overline{Simp \int_{\mathbf{R}^d} f(x)dx}$.*
- *Equivalence: If f and g agree almost everywhere, then $Simp \int_{\mathbf{R}^d} f(x)dx = Simp \int_{\mathbf{R}^d} g(x)dx$.*
- *Compatibility with Lebesgue measure: For any Lebesgue measurable E, on has $Simp \int_{\mathbf{R}^d} 1_E(x)dx = m(E)$.*
- *Uniqueness: The complex-values simple integral $f \mapsto Simp \int_{\mathbf{R}^d} f(x)dx$ is the only map from the space $Simp^{abs}(\mathbf{R}^d)$ to absolutely integrable simple functions to $\mathbf{C}$ that obeys all the above properties.*

Simple functions that agree almost everywhere, have the same integral. One can have 'noise' or 'errors' in in the function $f(x)$ on a null set, and this will not affect the final value of the integral. One can, then, even integrate functions f that are not defined everywhere on $\boldsymbol{R}^d$, but merely defined almost everywhere on $\boldsymbol{R}^d$, simply by extending f to all of $\boldsymbol{R}^d$, say by setting f equal to zero in all other places. Functions like $sin(x)/x$ can't be evaluated at every point, but can be integrated using this extension. In functional analysis this abstraction is used to improve the properties of various function spaces. The 'Lebesgue philosophy' that one is willing to lose control on sets of measure zero to gain access to the powerful tool of the Lebesgue integral, distinguishes it from descriptive set theory where one needs to control a function at absolutely every point.

3.2 Measurable functions

We can extend the class of unsigned simple functions to the larger class of unsigned Lebesgue measurable functions.

Definition 3.5. *(Unsigned measurable function) An unsigned function $f : \boldsymbol{R}^d \to [0, +\infty]$ is unsigned Lebesgue measurable, or measurable for short, if it is the pointwise limit of unsigned simple functions, i.e. if there exists a sequence $f_1, f_2, \ldots : \boldsymbol{R}^d \to [0, +\infty]$ of unsigned simple functions such that $f_n(x) \to f(x)$ for every $x \in \boldsymbol{R}^d$.*

Lemma 3.2. *(Equivalent notions of measurability) Let $f : \boldsymbol{R}^d \to [0, +\infty]$ be an unsigned function. Then the following are equivalent:*

- *f is unsigned Lebesgue measurable.*
- *f is the pointwise limit of unsigned simple functions f_n, thus the limit $\lim_{n\to\infty} f_n(x)$ exists and is equal to $f(x)$ for all $x \in \boldsymbol{R}^d$.*
- *f is the pointwise almost everywhere limit of unsigned simple functions f_n, thus the limit $\lim_{n\to\infty} f_n(x)$ exists and is equal to $f(x)$ for almost every $x \in \boldsymbol{R}^d$.*
- *f is the supremum $f(x) = \sup_n f_n(x)$ of an increasing sequence $0 \le f_1 \le f_2 \le \ldots$ of unsigned simple functions f_n, each of which are bounded with finite measure support.*
- *For every $\lambda \in [0, +\infty]$, the set $\{x \in \boldsymbol{R}^d : f(x) > \lambda\}$ is Lebesgue measurable.*
- *For every $\lambda \in [0, +\infty]$, the set $\{x \in \boldsymbol{R}^d : f(x) \ge \lambda\}$ is Lebesgue measurable.*
- *For every $\lambda \in [0, +\infty]$, the set $\{x \in \boldsymbol{R}^d : f(x) < \lambda\}$ is Lebesgue measurable.*
- *For every $\lambda \in [0, +\infty]$, the set $\{x \in \boldsymbol{R}^d : f(x) \le \lambda\}$ is Lebesgue measurable.*
- *For every interval $I \subset [0, +\infty)$, the set $f^{-1}(I) := \{x \in \boldsymbol{R}^d : f(x) \in I\}$ is Lebesgue measurable.*
- *For every relatively open set $U \subset [0, +\infty)$, the set $f^{-1}(U) := \{x \in \boldsymbol{R}^d : f(x) \in U\}$ is Lebesgue measurable.*
- *For every relatively closed set $K \subset [0, +\infty)$, the set $f^{-1}(K) := \{x \in \boldsymbol{R}^d : f(x) \in K\}$ is Lebesgue measurable.*

These equivalence can now be used to defined plenty of measurable functions. The inverse image of a Lebesgue measurable set by a measurable function need not remain Lebesgue measurable. However, with a slightly stronger measurability property of Borel measurability we will see that $f^{-1}(E)$ would be measurable as well.

Definition 3.6. *(complex measurability) An almost everywhere defined complex-valued function $f : \boldsymbol{R}^d \to \boldsymbol{C}$ is Lebesgue measurable, or measurable for short, if it is the pointwise almost everywhere limit of complex-valued simple functions.*

Like before, there are several equivalent definitions. The notion of complex-valued measurability and unsigned measurability are compatible when applied to a function that takes values in $[0, +\infty) = [0, +\infty] \cap \boldsymbol{C}$ everywhere or almost everywhere.

3.3 Unsigned Lebesgue integrals

Definition 3.7. *(Lower unsigned Lebesgue integral) Let $f : \boldsymbol{R}^d \to [0, +\infty]$ be an unsigned function, no necessarily measurable. We define the lower unsigned Lebesgue integral $\underline{\int_{\boldsymbol{R}^d}} f(x)dx$ to be the quantity $\underline{\int_{\boldsymbol{R}^d}} := \sup_{0 \leq g \leq f; g \text{ simple}} Simp \int_{\boldsymbol{R}^d} g(x)dx$ where g ranges over all unsigned simple functions $g : \boldsymbol{R}^d \to [0, +\infty]$ that are pointwise bounded by f. Similarly on defines the upper unsigned Lebesgue integral $\overline{\int_{\boldsymbol{R}^d}} f(x)dx := \inf_{h|geqf; h \text{ simple}} Simp \int_{\boldsymbol{R}^d} h(x)d$ The upper Lebesgue integral is always at least as large as the lower Lebesgue integral.*

We can prove all the properties of the lower Lebesgue integral aw we did for the integral of simple functions.

Definition 3.8. *(Unsigned Lebesgue integral) If $f : \boldsymbol{R}^d \to [0, +\infty]$ is measurable, we define the unsigned Lebesgue integral $\int_{\boldsymbol{R}^d} f(x)dx$ of f to equal the lower unsigned Lebesgue integral $\underline{\int_{\boldsymbol{R}^d}} f(x)dx$. For non-measurable functions, we leave the unsigned Lebesgue integral undefined.*

The lower and upper Lebesgue integrals of measurable, bounded, and vanishing outside a set of finite measures are equal.

Lemma 3.3. *Finite additivity of the Lebesgue integral Let $f, g : \boldsymbol{R}^d \to [0, +\infty]$ be measurable. Then $\int_{\boldsymbol{R}^d} f(x) + g(x)dx = \int_{\boldsymbol{R}^d} f(x)dx + \int_{\boldsymbol{R}^d} g(x)dx$.*

This property when enhanced to countable additivity is also called the monotone convergence theorem and will be introduced shortly.

Lemma 3.4. *(Uniqueness of the Lebesgue integral) The Lebesgue integral $f \mapsto \int_{\boldsymbol{R}^d} f(x)dx$ is the only map from measurable unsigned functions $f : \boldsymbol{R}^d \to [0, \infty]$ to $[0, +\infty]$ that obeys the following properties for measurable $f, g : \boldsymbol{R}^d \to [0, +\infty]$:*

- *Compatibility with the simple integral: If f is simple, then $\int_{\boldsymbol{R}^d} f(x)dx = Simp \int_{\mathbf{R}^d} f(x)dx$.*
- *Finite additivity: $\int_{\boldsymbol{R}^d} f(x) + g(x)dx = \int_{\boldsymbol{R}^d} f(x)dx + \int_{\boldsymbol{R}^d} g(x)dx$.*
- *Horizontal truncation: As $n \to \infty$, $\int_{\boldsymbol{R}^d} \min(f(x), n)dx$ converges to $\int_{\boldsymbol{R}^d} f(x)dx$.*
- *Vertical truncation: As $n \to \infty$, $\int_{\boldsymbol{R}^d} f(x)1_{|x| \leq n}dx$ converges to $\int_{\boldsymbol{R}^d} f(x)dx$.*

If $f : \boldsymbol{R}^d \to [0, +\infty]$ is a measurable function, then there is translation invariance as well, $\int_{\boldsymbol{R}^d} f(x + y)dx = \int_{\boldsymbol{R}^d} f(x)dx$ for any $y \in \boldsymbol{R}^d$.

Lemma 3.5. *(compatibility with Riemann integral) Let $f : [a, b] \to [0, +\infty]$ be Riemann integrable. If we extend f to $\boldsymbol{R}$ by declaring f to equal zero outside of $[a, b]$ then $\int_{\boldsymbol{R}^d} f(x)dx = \int_a^b f(x) = dx$.*

Lemma 3.6. *Let $f : \boldsymbol{R}^d \to [0, +\infty]$ be measurable. Then for any $0 < \lambda < \infty$, one has $m(\{x \in \boldsymbol{R}^d : f(x) \geq \lambda\}) \leq \frac{1}{\lambda} \int_{\boldsymbol{R}^d} f(x)dx$.*

This implies that if $\int_{\boldsymbol{R}^d} f(x)dx < \infty$ then f is finite almost everywhere. Also, if $\int_{\boldsymbol{R}^d} f(x)dx = 0$ iff f is zero almost everywhere. The use of integral $\int_{\boldsymbol{R}^d} f(x)dx$ to control the distribution of f is known as *the moment method*. One can also control this distribution using higher moments such as $\int_{\boldsymbol{R}^d} |f(x)|^p dx$ for various values of p, or exponential moments such as $\int_{\boldsymbol{R}^d} e^{tf(x)}dx$ or the Fourier moments $\int_{\boldsymbol{R}^d} e^{itf(x)}dx$ for various values of t; such moment methods are fundamental to probability theory.

3.4 Absolute integrability

Definition 3.9. *(Absolute integrability) An almost everywhere defined measurable function* $f : \boldsymbol{R}^d \to \boldsymbol{C}$ *is said to be absolutely integrable if the unsigned integral* $\|f\|_{L^1(\boldsymbol{R}^d)} := \int_{\boldsymbol{R}^d} |f(x)|dx$ *is finite. We refer to this quantity* $\|f\|_{L^1(\boldsymbol{R}^d)}$ *as the* $L^1(\boldsymbol{R}^d)$ *norm of* f*, and use* $L^1(\boldsymbol{R}^d)$ *or* $L^1(\boldsymbol{R}^d \to \boldsymbol{C})$ *to denote the space of absolutely integrable functions. If* f *is real-valued and absolutely integrable, we define the Lebesgue integral* $\int_{\boldsymbol{R}^d} f(x)dx$ *by the formula* $\int_{\boldsymbol{R}^d} f(x)dx := \int_{\boldsymbol{R}^d} f_+(x)dx + \int_{\boldsymbol{R}^d} f_-(x)dx$ *where* $f_+ := \max(f, 0)$ *and* $f_- := \min(-f, 0)$ *are the magnitudes of the positive and negative components of* f*. If* f *is complex valued and absolutely integrable, we define the Lebesgue integral* $\int_{\boldsymbol{R}^d} f(x)dx$ *by the formula* $\int_{\boldsymbol{R}^d} f(x)dx := \int_{\boldsymbol{R}^d} \Re f(x)dx + i \int_{\boldsymbol{R}^d} \Im f(x)dx$ *where the two integrals on the right are interpreted as real-valued absolutely integrable Lebesgue integrals.*

The absolutely integrable Lebesgue integral extends the absolutely integrable simple integral. Attempts to define integrals for non-absolutely integrable functions, analogous to the improper integrals in Riemannian theory like $\int_{-\infty}^{\infty} f(x)dx := \lim_{R\to\infty} \int_{-R}^{R} f(x)dx$ lead to poor behavior with respect to various important operations, such as change of variables or exchanging limits and integrals. We instead deal with such exotic integrals on an ad hoc basis.

from the pointwise triangle inequality $|f(x) + g(x)| \leq |f(x)| + |g(x)|$, we conclude the L^1 triangle inequality $\|f + g\|_{L^1(\boldsymbol{R}^d)} \leq \|f\|_{L^1(\boldsymbol{R}^d)} + \|g\|_{L^1(\boldsymbol{R}^d)}$ for any almost everywhere defined measurable $f, g : \boldsymbol{R}^d \to \boldsymbol{C}$. Also, for any complex number c we have $\|cf\|_{L^1(\boldsymbol{R}^d)} = |c|\|f\|_{L^1(\boldsymbol{R}^d)}$. As such, we see that $L^1(\boldsymbol{R}^d) \to \boldsymbol{C}$ is a complex vector space. A function $f \in L^1(\boldsymbol{R}^d \to \boldsymbol{C})$ has zero L^1 norm, $\|f\|_{L^1(\boldsymbol{R}^d)} = 0$, iff f is zero almost everywhere. Given two functions $f, g \in L^1(\boldsymbol{R}^d \to \boldsymbol{C})$, we can define the L^1 distance $d_{L^1}(f, g)$ between them by the formula $d_{L^1}(f, g) := \|f - g\|_{L^1(\boldsymbol{R}^d)}$. Due to triangle's inequality the distance obeys almost all the axioms of a metric on $L^1(\boldsymbol{R}^d)$, with one exception: it is possible for two different functions $f, g \in L^1(\boldsymbol{R}^d \to \boldsymbol{C})$ to have a zero L^1 distance, if they agree almost everywhere. As such, d_{L^1} is only a semi-metric (or pseudo-metric) rather than a metric. However, if one adopts the convention that any two functions that agree almost everywhere are considered equivalent, then one recovers a genuine metric. More formally, one works in the quotient space of $L^1(\boldsymbol{R}^d)$ by the equivalence relation of almost everywhere agreement, which by abuse of notation is also denoted $L^1(\boldsymbol{R}^d)$. We will later show that this quotient space is a complete metric space, a fact known as the L^1 Riesz-Fischer theorem; this completeness is one of the main reasons why we spend so much effort setting up Lebesgue integration theory in the first place.

The linearity properties of the unsigned integral induce analogous linearity properties of the absolutely convergent Lebesgue integral. We can localize the absolutely convergent integral to any measurable subset E of $\boldsymbol{R}^d$. If $f : E \to \boldsymbol{C}$ is a function, we say that f is measurable if its extension $\tilde{f} : \boldsymbol{R}^d \to \boldsymbol{C}$ is measurable, where $\tilde{f}(x)$ is defined to be equal to $f(x)$ when $x \in E$ and zero otherwise. We then define $\int_E f(x)dx := \int_{\boldsymbol{R}^d} \tilde{f}(x)dx$. As a special case of the more general Lebesgue integration theory on abstract measure spaces the triangle inequality is presented here.

Lemma 3.7. *Let* $f \in L^1(\boldsymbol{R}^d \to \boldsymbol{C})$*. Then* $|\int_{\boldsymbol{R}^d} f(x)dx| \leq \int_{\boldsymbol{R}^d} |f(x)|dx$*.*

3.5 Littlewood's three principles

Littlewood's three principles are informal heuristics that convey much of the basic intuition behind the measure theory of Lebesgue. Briefly, the three principles are

- Every measurable set is nearly a finite sum of intervals.
- Every absolutely integrable function is nearly continuous.

- Every pointwise convergent sequence of functions is nearly uniformly convergent.

Various manifestations of the first principle were given before. We turn to the second principle. Define a step function to be a finite linear combination of indicator functions 1_B of boxes B.

Theorem 3.1. *(Approximation of L^1 functions) Let $f \in L^1(\boldsymbol{R}^d)$ and $\epsilon > 0$*

- *There exists an absolutely integrable simple function g such that $\|f - g\|_{L^1(\boldsymbol{R}^d)} \leq \epsilon$.*
- *There exists a step function g such that $\|f - g\|_{L^1(\boldsymbol{R}^d)} \leq \epsilon$.*
- *There exists a continuous, compactly supported g such that $\|f - g\|_{L^1(\boldsymbol{R}^d)} \leq \epsilon$.*

In other words, the absolutely integrable simple functions the step functions, and the continuous, compactly supported functions are all dense subsets of $L^1(\boldsymbol{R}^d)$ with respect to the $L^1(\boldsymbol{R}^d)$ metric. Turning to the third principle we recall three basic ways in which a sequence $f_n : \boldsymbol{R}^d \to \boldsymbol{C}$ of functions can converge to a limit of $f : \boldsymbol{R}^d \to \boldsymbol{C}$:

- Pointwise convergence $f_n(x) \to f(x)$ for every $x \in \boldsymbol{R}^d$.
- Pointwise almost everywhere convergence $f_n(x) \to f(x)$ for almost every $x \in \boldsymbol{R}^d$.
- Uniform convergence For every $\epsilon > 0$, there exists N such that $|f_n(x) - f(x)| \leq \epsilon$ for all $n \geq N$ and all $x \in \boldsymbol{R}^d$.

Uniform convergence implies pointwise convergence, which in turn implies pointwise almost everywhere convergence. We now add a fourth mode of convergence, that is weaker than uniform convergence but stronger than pointwise convergence. A property P is said to hold locally on some domain X if, for every point x_0 in that domain, there is an open neighbourhood of x_0 in X on which P holds.

Definition 3.10. *(Locally uniform convergence) A sequence of functions $f_n : \boldsymbol{R}^d \to \boldsymbol{C}$ converges locally uniformly to limit $f : \boldsymbol{R}^d \to \boldsymbol{C}$ if, for every bounded subset E of $\boldsymbol{R}^d$, f_n converges uniformly to f on E, i.e., for every bounded $E \subset \boldsymbol{R}^d$ and every $\epsilon > 0$, there exists $N > 0$ such that $|f_n(x) - f(x)| \leq \epsilon$ for all $n \geq N$ and $x \in E$.*

Using Heine-Borel theorem, we can say for $\boldsymbol{R}^d$, f_n converges locally uniformly to f if, for every point $x_0 \in \boldsymbol{R}^d$, there exits an open neighbourhood U of x_0 such that f_n converges uniformly to f on U. However, on domains on which the Heine-Borel theorem does not hold, the bounded-set notion of local uniform convergence is not equivalent to the open-set notion of local uniform convergence. For example, the function $x \mapsto x/n$ on $\boldsymbol{R}$ for $n = 1, 2, \dots$ converge locally uniformly, and hence pointwise, to zero on $\boldsymbol{R}$, but do not converge uniformly. Similarly, the partial sums $\sum_{n=0}^{N} \frac{x^n}{n!}$ of the Taylor series $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ converges to e^x locally uniformly, and hence pointwise, on $\boldsymbol{R}$, but not uniformly. Finally, the functions $f_n(x) := \frac{1}{nx} 1_{x>0}$ for $n = 1, 2, \dots$ with convention that $f_n(0) = 0$ converge pointwise everywhere to zero, but do not converge locally uniformly.

The examples show that pointwise convergence, either everywhere or almost everywhere, is a weaker concept than local uniform convergence. Nevertheless, a remarkable theorem of Egorov, which demonstrates Littlewood's third principle, asserts that one can recover local uniform convergence as long as one is willing to delete a set of small measure:

Theorem 3.2. *(Egorov's theorem) Let $f_n : \boldsymbol{R}^d \to \boldsymbol{C}$ be a sequence of measurable functions that converge pointwise almost everywhere to another function $f : \boldsymbol{R}^d \to \boldsymbol{C}$, and let $\epsilon > 0$. Then there exists a Lebesgue measurable set A of measure at most ϵ, such that f_n converges locally uniformly to f outside of A.*

Unfortunately, one cannot upgrade local uniform convergence to uniform convergence in Egorov's theorem. For example, for a moving bump $f_n := 1_{n,n+1}$ on $\boldsymbol{R}$, converges pointwise, and locally uniformly, to the zero function $f = 0$. However, for any $0 < \epsilon < 1$ and any n, we have $|f_n(x) - f(x)| > \epsilon$ on a set of measure 1, namely on the interval $[n, n+1]$. Thus, if one wanted f_n to converge uniformly to f outside of a set A, then that set A has to contain a set of measure 1. In fact, it must contain the intervals $[n, n+1]$ for all sufficiently large n and must therefore have infinite measure. However, if all the f_n and f were supported on a fixed set E of finite measure, then the above "escape to horizontal infinity" cannot occur, and one can recover uniform convergence outside the set of arbitrary small measure.

Theorem 3.3. *(Lusin's theorem) Let $f : \boldsymbol{R}^d \to \boldsymbol{C}$ be absolutely integrable, and let $\epsilon > 0$. Then there exists a Lebesgue measurable set $E \subset \boldsymbol{R}^d$ of measure at most ϵ such that the restriction of f to the complementary set $\boldsymbol{R}^d$ is continuous on that set.*

This theorem does not imply that the unrestricted function f is continuous on $\boldsymbol{R}^d$. For instance, the absolutely integrable function $1_{\boldsymbol{Q}} : \boldsymbol{R} \to \boldsymbol{C}$ is nowhere continuous, so is certainly not continuous on $\boldsymbol{R}$ for any E of finite measure; but if one deletes the measure zero set $E := \boldsymbol{Q}$ from the reals, then the restriction of f to $\boldsymbol{R}$ is identically zero and thus continuous.

4 Abstract measure spaces

We now generalize from the Euclidean space $\boldsymbol{R}^d$ to a more general space X. For this purposes, apart from the set X we also need

- A collection $\mathcal{B}$ of subsets of X that one is allowed to measure
- The measure $m(E) \in [0, +\infty]$ one assigns to each measurable set $E \in \mathcal{B}$.

For example, Lebesgue measure theory covers the case when X is a Euclidean space $\boldsymbol{R}^d$, $\mathcal{B}$ is the collection of $\mathcal{L}[\boldsymbol{R}^d]$ all Lebesgue measurable subsets of $\boldsymbol{R}^d$, and $m(E)$ is the Lebesgue measure $\mu(E)$ of E. The collection $\mathcal{B}$ has to obey a number of axioms, e.g. being closed with respect to countable unions, that make it a σ-algebra, which is a stronger variant of the more well-known concept of boolean algebra. Similarly, the measure μ has to obey a number of axioms, most notably a countable additivity axiom, in order to obtain a measure and integration theory comparable to the Lebesgue theory on Euclidean spaces. When all these axioms are satisfied, the triple $(X, \mathcal{B}, \mu)$ is known as a **measure space**. On any measure space, one can set up the unsigned and absolutely convergent integrals in almost exactly the same way as was done for Lebesgue integral on Euclidean spaces. Although the approximation theorems are largely unavailable at this level of generality due to the lack of such concepts as "elementary set" or "continuous function", one does have the fundamental convergence theorems for the subject, namely Fatou's lemma, the monotone convergence theorem and the dominated convergence theorem. These measures are constructed using *Riesz representation theorem*, which we don't cover here.

4.1 Boolean algebras

Definition 4.1. *(Boolean algebras) Let X be a set. A Boolean algebra on X is a collection $\mathcal{B}$ of X which obeys the following properties*

- *Empty set: $\emptyset \in \mathcal{B}$.*
- *Complement: If $E \in \mathcal{B}$, then the complement $E^c := X \backslash E \in \mathcal{B}$.*
- *Finite unions: If $E, F \in \mathcal{B}$, then $E \cup F \in \mathcal{B}$.*

We sometimes say that E is $\mathcal{B}$-measurable, if $E \in \mathcal{B}$. Given two Boolean algebras $\mathcal{B}, \mathcal{B}'$ on X, we say that $\mathcal{B}'$ is finer than, a sub-algebra of, or a refinement of $\mathcal{B}$, or that $\mathcal{B}$ is coarser than or a coarsening of $\mathcal{B}'$, if $\mathcal{B} \subset \mathcal{B}'$. By using de Morgan's law it can be shown that Boolean algebra is also closed under other Boolean algebra operations such as intersection $E \cap F$, set difference $E \backslash F$ and symmetric difference $E \Delta F$. Given any set X, the coarsest Boolean algebra is the trivial algebra $\{\emptyset, X\}$, in which the only measurable sets are the empty set and the whole set. The finest Boolean algebra is the discrete algebra $2^X := \{E : E \subset X\}$, in which every set is measurable. All other Boolean algebras are intermediate between these two extremes: finer than the trivial algebra, but coarser than the discrete one.

Let $\overline{\mathcal{E}[\boldsymbol{R}^d]}$ be the collection of those sets $E \subset \boldsymbol{R}^d$ that are either elementary sets, or co-elementary sets, i.e. complement of an elementary sets. Then $\overline{\mathcal{E}[\boldsymbol{R}^d]}$ is a Boolean algebra called *elementary Boolean algebra* of $\boldsymbol{R}^d$. Let $\overline{\mathcal{J}[\boldsymbol{R}^d]}$ be the collection of subsets of $\boldsymbol{R}^d$ that are either Jordan measurable or co-Jordan measurable, i.e. complement of a Jordan measurable set. Then $\overline{\mathcal{J}[\boldsymbol{R}^d]}$ called *Jordan algebra* on $\boldsymbol{R}^d$ and is a Boolean algebra that is fine than the elementary algebra. Let $L(\boldsymbol{R}^d)$ be a collection of Lebesgue measurable subsets of $\boldsymbol{R}^d$. Then $\mathcal{L}[\boldsymbol{R}^d]$ is a Boolean algebra, called the *Lebesgue algebra* on $\boldsymbol{R}^d$ and is finer than the Jordan algebra. Let $\mathcal{N}(\boldsymbol{R}^d)$

be the collection of subsets of $\boldsymbol{R}^d$ that are either Lebesgue null sets or Lebesgue co-null sets, i.e. complement of null sets. Then $\mathcal{N}(\boldsymbol{R}^d)$ is a Boolean algebra called the *null algebra* on $\boldsymbol{R}^d$, and is coarser than the Lebesgue algebra.

Let $\mathcal{B}$ be a Boolean algebra on a set X, and let Y be a subset of X, not necessarily $\mathcal{B}$-measurable. The *restriction* $\mathcal{B}|_Y := \{E \cap Y : E \in \mathcal{B}\}$ of $\mathcal{B}$ to Y is a Boolean algebra on Y. If Y is $\mathcal{B}$-measurable then $\mathcal{B}|_Y = \mathcal{B} \cap 2^Y = \{E \subset Y : E \in \mathcal{B}\}$. Let X be partitioned into a union $X = \cup_{\alpha \in I} A_\alpha$ of disjoint sets A_α, which we refer to as atoms. Then this partition generates a Boolean algebra $\mathcal{A}((A_\alpha)_{\alpha \in I})$, defined as the collection of all the sets E of the form $E = \cup_{\alpha \in J} A_\alpha$ for some $J \subset I$. This is called *atomic algebra* with atoms $(A_\alpha)_{\alpha \in I}$. The trivial algebra corresponds to the trivial partition $X = X$ into a single atom; the discrete algebra corresponds to the discrete partition $X = \cup_{x \in X\{x\}}$ into singleton atoms. Finer partitions lead to finer atomic algebra. Empty atoms have no impact, so without loss of generality on can delete all empty atoms.

Let n be an integer The *dyadic algebra* $\mathcal{D}_n(\boldsymbol{R}^d)$ at scale 2^{-n} in $\boldsymbol{R}^d$ is defined to the atomic algebra generated by the half-open dyadic cubes $\left[\frac{i_1}{2^n}, \frac{i_1+1}{2^n}\right) \times \ldots \times \left[\frac{i_d}{2^n}, \frac{i_d+1}{2^n}\right)$ of length 2^{-n}. These are Boolean algebras which are increasing in n: $\mathcal{D}_{n+1} \supset \mathcal{D}_n$. Let $(\mathcal{B}_\alpha)_{\alpha \in I}$ be a family of Boolean algebras on a set X, indexed by a, possibly infinite or uncountable, label set I. The intersection $\cap_{\alpha \in I} \mathcal{B}_\alpha$ of these algebras is still a Boolean algebra, and is the finest Boolean algebra that is coarser than all of the $\mathcal{B}_\alpha$.

Definition 4.2. *(Generation of algebra) Let $\mathcal{F}$ be any family of sets in X. We define $(\mathcal{F})_{bool}$ to be the intersection of all the Boolean algebras that contain $\mathcal{F}$, which is again a Boolean algebra. Equivalently, $(\mathcal{F})_{bool}$ is the coarsest Boolean algebra that contains $\mathcal{F}$. We say that $(\mathcal{F})_{bool}$ is the Boolean algebra generated by $\mathcal{F}$.*

$\mathcal{F}$ is a Boolean algebra iff $(\mathcal{F})_{bool} = \mathcal{F}$; thus each Boolean algebra is generated by itself. Elementary algebra $\mathcal{E}(\boldsymbol{R}^d)$ is generated by the collection of boxes in $\boldsymbol{R}^d$. The Boolean algebra $(\mathcal{F})_{bool}$ can be described explicitly in terms of $\mathcal{F}$, which is a collection of sets in a set X. We define $\mathcal{F}_0 := \mathcal{F}$ and for each $n \geq 1$ we define $\mathcal{F}_n$ to be the collection of all sets that either the union of a finite number of sets in $\mathcal{F}_{n-1}$, including the empty union $\emptyset$, or the complement of such a union. Thus $(\mathcal{F})_{bool} = \cup_{n=0}^{\infty} \mathcal{F}_n$.

4.2 σ-algebras and measurable spaces

We need to improve Boolean algebra to a countable union axiom.

Definition 4.3. *(Sigma algebras) Let X be a set. A σ-algebra on X is a collection $\mathcal{B}$ of X which obeys the following properties:*

- *Empty set: $\emptyset \in \mathcal{B}$.*
- *Complement: If $E \in \mathcal{B}$, then the complement $E^c := X \backslash E \in B$.*
- *Countable unions: If $E_1, E_2, \ldots \in \mathcal{B}$, then $\cup_{n=1}^{\infty} E + n \in \mathcal{B}$.*

we refer to the pair $(X, \mathcal{B})$ of a set X together with a σ-algebra on that set as a measurable space. From de Morgan's law, which is just as valid for infinite unions and intersections as it is for finite ones, we see that σ-algebras are closed under countable intersections as well as countable unions. By padding a finite union into a countable union by using the empty set, we see that every σ-algebra is automatically a Boolean algebra. Lebesgue and null algebras are σ-algebras, but the elementary and Jordan algebras are not. Any restriction $\mathcal{B}|_Y$ of a σ-algebra $\mathcal{B}$ to a subspace Y of X is again a σ-algebra on the subspace Y. The intersection $\cap_{\alpha \in I} \mathcal{B}_\alpha$ of an arbitrary, and possibly infinite or uncountable, number of σ-algebras $\mathcal{B}_\alpha$ is again a σ-algebra, and is the finest σ-algebra that is coarser than all of the $\mathcal{B}_\alpha$.

Definition 4.4. *(Generation of σ-algebras) Let $\mathcal{F}$ be any family of sets in X. We define $(\mathcal{F})$ be to the intersection of all the -algebras that contain $\mathcal{F}$, which is again a σ-algebra. Equivalently, $(\mathcal{F})$ is the coarsest σ-algebra that contains $\mathcal{F}$. We say that $(\mathcal{F})$ is the σ-algebra generated by $\mathcal{F}$.*

We have the trivial inclusion $(\mathcal{F})_{bool} \subset (\mathcal{F})$. The equality holds iff $(\mathcal{F})_{bool}$ is a σ-algebra. We have the following principle, analogous to principle of induction: if $\mathcal{F}$ is a family of sets in X, and $P(E)$ is a property of sets $E \subset X$ which obeys the following axioms:

- $P(\emptyset)$ is true.
- $P(E)$ is true for all $E \in \mathcal{F}$.
- If $P(E)$ is true for some $E \subset X$, then $P(X \backslash E)$ is also true.
- If $E_1, E_2, \ldots \subset X$ are such that $P(E_n)$ is true for all n, then $P(\cup_{n=1}^{\infty} E_n)$ is also true.

Then one can conclude that $P(E)$ is true for all $E \in (\mathcal{F})$.

Definition 4.5. *(Borel σ-algebra) Let X be a metric space, or more generally a topological space. The Borel σ-algebra $\mathcal{B}[X]$ of X is defined to be the σ-algebra generated by the open subsets of X. Elements of $\mathcal{B}[X]$ will be called Borel measurable.*

The Borel σ-algebra contains the open sets, the closed sets, which are complement of the open sets, the countable unions of closed sets (known as F_σ sets), the countable intersections of open sets (known as G_δ sets), the countable intersections of F_σ sets, and so forth. In $\boldsymbol{R}^d$, every open set is Lebesgue measurable, and so we see that the Borel σ-algebra is coarser than the Lebesgue σ-algebra. We define Borel σ-algebra to be generated by the open sets. However they are also generated by several other sets, like closed subsets of $\boldsymbol{R}^d$.

Let $\mathcal{F}$ be a collection of sets in a set X, and let ω_1 be the first uncountable ordinal. Define the sets $\mathcal{F}_\alpha$ for every countable ordinal $\alpha \in \omega_1$ via transfinite induction as follows:

- $\mathcal{F}_\alpha := \mathcal{F}$.
- For each countable successor ordinal $\alpha = \beta + 1$, we define $\mathcal{F}_\alpha$ to be the collection of all sets that either the union of an at most countable number of sets in $\mathcal{F}_{n-1}$, including the empty union $\emptyset$, or the complement of such a union.
- For each countable limit ordinal $\alpha = \sup\limits_{\beta<\alpha} \beta$, we define $\mathcal{F}_\alpha := \cup_{\beta<\alpha} \mathcal{F}_\beta$.

Then $(\mathcal{F}) = \cup_{\alpha \in \omega_1} \mathcal{F}_\alpha$. In the case where $\mathcal{F}$ is the collection of open sets in a topological space, so that $(\mathcal{F})$, then the sets $\mathcal{F}_\alpha$ are essentially the Borel hierarchy, which starts at the open and closed sets, then moves on to the F_σ and G_δ sets, and so forth.

There can be Lebesgue measurable subsets that are not Borel measurable. A large majority of the explicitly constructible sets that one actually encounters in practice tend to be Borel measurable.

4.3 Countably additive measures and measure spaces

We now endow the σ-algebra with a measure. We begin with finite additive theory and then supplant it by the countably additive theory.

Definition 4.6. *(Finitely additive measure) Let $\mathcal{B}$ be a Boolean algebra on a space X. An unsigned finitely additive measure μ on $\mathcal{B}$ is a map $\mu : \mathcal{B} \to [0, +\infty]$ that obeys the following axioms:*

- *Empty set: $mu(\emptyset) = 0$.*
- *Finite additivity: Whenever $E, F \in \mathcal{B}$ are disjoint, then $\mu(E \cup F) = \mu(E) + \mu(F)$.*

The empty set axiom is needed to rule out the degenerate situation in which every set, including the empty set, has infinite measure. As an example, let $x \in \mathcal{B}$ be an arbitrary Boolean algebra on X. Then the Dirac measure δ_x at x, defined by setting $\delta_x(E) := 1_E(x)$, is finitely additive. The zero measure $0 : E \mapsto 0$ is a finitely additive measure on any Boolean algebra. Linear combination of finite measures are also finite measures. Further, if $\mathcal{B}$ is a Boolean algebra on X, $\mu : \mathcal{B} \to [0, +\infty]$ is a finitely additive measure, and Y is a $\mathcal{B}$-measurable subset of X, then the restriction $\mu|_Y : \mathcal{B}|_{Y \to [0,+\infty]}$ of $\mathcal{B}$ of Y, defined by setting $\mu|_Y(E) := \mu(E)$ whenever $E \in \mathcal{B}|_Y$, i.e. if $E \in \mathcal{B}$ and $E \subset Y$, is also a finitely additive measure. If $\mathcal{B}$ is a Boolean algebra on X, then the function $: \mathcal{B} \to [0, +\infty]$ defined by setting (E) to be the cardinality of E if E is finite, and $(E) := +\infty$ if E is infinite, is a finitely additive measure, known as *counting measure.*

Let $\mu : \mathcal{B} \to [0, +\infty]$ be a finitely additive measure on a Boolean σ-algebra $\mathcal{B}$. Then we have the following properties

- Monotonicity: If E, F are $\mathcal{B}$-measurable and $E \subset F$, then $\mu(E) \leq \mu(F)$.
- Finite additivity: If k is a natural number, and $E_1, , E_k$ are $\mathcal{B}$-measurable and disjoint, then $\mu(E_1 \cup \ldots \cup E_k) = \mu(E_1) + +\mu(E_k)$.
- Finite subadditivity: If k is a natural number, and $E_1, \ldots, E_k$ are $\mathcal{B}$-measurable, then $\mu(E_1 \cup \ldots \cup E_k) \leq \mu(E_1) + \ldots + \mu(E_k)$.
- Inclusion-exclusion for two sets: If E, F are $\mathcal{B}$-measurable, then $\mu(E \cup F) + \mu(E \cap F) = \mu(E) + \mu(F)$.

One can characterise measures completely for any finite algebra uniquely. This is about the limit of what one can say about finitely additive measures at this level of generality. We now specialize to countably additive measures on σ-algebra.

Definition 4.7. *(Countably additive measure) Let $(X, \mathcal{B})$ be a measure space. An unsigned countably additive measure μ on $\mathcal{B}$, or measure for short, is a map $\mu : \mathcal{B} \to [0, +\infty]$ that obeys the following axioms:*

- *Empty set: $\mu(\emptyset) = 0$.*
- *Countable additivity: Whenever $E_1, E_2, \ldots \in \mathcal{B}$ are a countable sequence of disjoint measurable sets, then $\mu(\cup_{n=1}^{\infty} E_n) = \sum\limits_{n=1}^{\infty} \mu(E_n)$.*

The triplet $(X, \mathcal{B}, \mu)$, where $(X, \mathcal{B})$ is a *measurable space* and $\mu : \mathcal{B} \to [0, +\infty]$ is a countably additive measure, is known as a measure space. Thus, a measurable space equipped with a measure is called a measure space. Lebesgue measure is a countably additive measure on the Lebesgue σ-algebra, and hence on every sub σ-algebra such as the Borel σ-algebra. The Dirac and counting measures are also countably additive. Any restriction on a countably additive measure to a measurable subspace is again countably additive. Countable combinations of a measure is also countably additive measure. Countably additivity measures are necessarily finitely additive, and so countably additive measures inherit all the properties of finitely additive properties, such as monotonicity and finite subadditivity. There are some additional properties:

- Countable subadditivity: If $E_1, E_2, \ldots$ are $\mathcal{B}$-measurable, then $\mu(\cap_{n=1}^{\infty} E_n) \leq \sum\limits_{n=1}^{\infty} \mu(E_n)$.
- Upwards monotone convergence: If $E_1 \subset E_2 \subset \ldots$ are $\mathcal{B}$-measurable, then $\mu(\cup_{n=1}^{\infty} E_n) = \lim\limits_{n \to \infty} \mu(E_n) = \sup\limits_{n} \mu(E_n)$.
- Downwards monotone convergence: If $E_1 \supset E_2 \supset \ldots$ are $\mathcal{B}$-measurable, and $\mu(E_n) < \infty$ for at least one n, then $\mu(\cap_{n=1}^{\infty} E_n) = \lim\limits_{n \to \infty} \mu(E_n) = \inf\limits_{n} \mu(E_n)$.

Let $(X, \mathcal{B}, \mu)$ be a measure space. Let $E_1, E_2, \ldots$ be a sequence of $\mathcal{B}$-measurable sets that converge to another set E, in the sense that 1_{E_n} converges pointwise to 1_E. Then E is also $\mathcal{B}$-measurable. If there exists a $\mathcal{B}$-measurable set F of finite measure, i.e. $\mu(F) < \infty$, that contains all of the E_n, then $\lim\limits_{n \to \infty} \mu(E_n) = \mu(E)$. This is called *dominated convergence for sets.*

Let X be an at most countable set with the discrete σ-algebra. Then every measure μ on this measurable space can be uniquely represented in the form $\mu = \sum\limits_{x \in X} c_x \delta_x$ for some $x_x \in [0, +\infty]$, thus $\mu(E) = \sum\limits_{x \in E} c_x$ for all $E \subset X$. This is not true for uncountable case.

Definition 4.8. *(Completeness) A null set of a measure space $(X, \mathcal{B}, \mu)$ is defined to be a $\mathcal{B}$-measurable set of measure zero. A sub-null set is any subset of a null set A measure space is said to be complete if every sub-null set is a null set.*

Thus, for instance, the Lebesgue measure space $(\boldsymbol{R}^d, \mathcal{L}[\boldsymbol{R}^d], m)$ is complete, but the Borel measure measure space $(\boldsymbol{R}^d, \mathcal{B}[\boldsymbol{R}^d], m)$ is not. It is fairly easy to modify any measure space to be complete. Let $(X, \mathcal{B}, \mu)$ be a measure space. There exists a unique refinement $(X, \overline{\mathcal{B}}, \overline{\mu})$, known as the completion of $(X, \mathcal{B}, \mu)$, which is the coarsest refinement of $(X, \mathcal{B}, \mu)$ that is complete. Furthermore, $\overline{\mathcal{B}}$ consists precisely of those sets that differ from a $\mathcal{B}$-measurable set by a $\mathcal{B}$-subnull set. In fact, the Lebesgue measure space $(\boldsymbol{R}^d, \mathcal{L}[\boldsymbol{R}^d], m)$ is the completion of the Borel measure space $(\boldsymbol{R}^d, \mathcal{B}[\boldsymbol{R}^d], m)$.

4.4 Measurable functions, and integration on a measure

We now define the notion of measurable function, which is analogous to that of a continuous function in topology. Recall that a function $f : X \to Y$ between two topological spaces X, Y is continuous if the inverse image $f^{-1}(U)$ of any open set is open.

Definition 4.9. *Let $(X, \mathcal{B})$ be a measurable space, and let $f : X \to [0, +\infty]$ for $f : X \to \mathcal{C}$ be an unsigned or complex-valued function. We say that f is measurable if $f^{-1}(U)$ if $\mathcal{B}$-measurable for every open subset U of $[0, +\infty]$ or $\mathcal{C}$.*

This generalizes the notion of Lebesgue measurable function. Like we did before - first a simple integral, then an unsigned integral, and then finally an absolutely convergent integral - we follow the same three stages here.

Definition 4.10. *(Simple integral) Let $(X, \mathcal{B}, \mu)$ be a measure space with $\mathcal{B}$ finite. X is partitioned into a finite number of atoms $A_1, \dots, A_n$. If $f : X \to [0, +\infty]$ is measurable, then it has a unique representation of the form $f = \sum\limits_{i=1}^{n} c_i 1_{A_i}$ for some $c_1, \dots, c_n \in [0, +\infty]$. We then define the simple integral by the formula $Simp \int\limits_X f d\mu := \sum\limits_{i=1}^{n} c_i \mu(A_i)$.*

With this definition, it is clear that one has the monotonicity property $Simp \int\limits_X f d\mu \leq Sim \int\limits_X g d\mu$ whenever $f \leq g$ are unsigned measurable, as well as the linearity properties $Simp \int\limits_X f + g d\mu = Simp \int\limits_X f d\mu + Simp \int\limits_X g d\mu$ and $Simp \int\limits_X df d\mu = c \times Simp \int_X f d\mu$ for unsigned measurable f, g and $c \in [0, +\infty]$.

These simple integrals are unaffected by refinements. Let (X, B, μ) be a measure space, and let $(X, \mathcal{B}', \mu')$ be a refinement of $(X, \mathcal{B}, \mu)$, which means that $\mathcal{B}'$ contains $\mathcal{B}$ and $\mu' : \mathcal{B}' \to [0, +\infty]$ agrees with $\mu : \mathcal{B} \to [0, +\infty]$ on $\mathcal{B}$. For both $\mathcal{B}$ and $\mathcal{B}'$ finite, and $f : \mathcal{B} \to [0, +\infty]$ being measurable we have $Simp \int\limits_X f d\mu = Simp \int\limits_X f d\mu'$.

This allows use to extend the simple integral to simple functions.

Definition 4.11. *(Integral of simple functions) An unsigned simple function $f : X \to [0, +\infty]$ on a measurable space $(X, \mathcal{B})$ is a measurable function that takes on finitely many values $a_1, \dots, a_k$. This function is automatically measurable with respect to at least one finite sub-σ-algebra $\mathcal{B}'$ and $\mathcal{B}$, namely the σ-algebra $\mathcal{B}'$ generated by the pre images $f^{-1}(\{a_1\}), \dots, f^{-1}(\{a_k\})$ of $a_1, \dots, a_k$. We then define the simple integral as $Simp \int\limits_X f d\mu := Simp \int\limits_X f d\mu|_{\mathcal{B}'}$, where $\mu|_{\mathcal{B}'} : \mathcal{B}' \to [0, +\infty]$ is the restriction of $\mu : \mathcal{B} \to [0, +\infty]$ to $\mathcal{B}'$.*

Lemma 4.1. *(Basic properties of the simple integral) Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f, g : X \to [0, +\infty]$ be simple functions.*

- *Monotonicity: If $f \leq g$ pointwise, then $Simp \int_X f d\mu \leq Simp \int_X g d\mu$.*
- *Compatibility with measure: For every $\mathcal{B}$-measurable set E, we have $Simp \int_X 1_E d\mu = \mu(E)$.*
- *Homogeneity: For every $c \in [0, +\infty]$, one has $Simp \int_X cf d\mu = c \times Simp \int_X f d\mu$.*
- *Finite additivity: $Simp \int_X (f + g) d\mu = Simp \int_X f d\mu + Simp \int_X g d\mu$.*
- *Insensitivity to refinement: If $(X, \mathcal{B}', \mu')$ is a refinement of $(X, \mathcal{B}, \mu)$, then $Simp \int_X f d\mu = Simp \int_X f d\mu'$.*
- *Almost everywhere equivalence: If $f(x) = g(x)$ for μ-almost every $x \in X$, then $Simp \int_X f d\mu = Simp \int_X g d\mu$.*

- *Finiteness: Simp* $\int_X f d\mu < \infty$ *iff* f *is finite almost everywhere, and is supported on a set of finite measure.*

- *Vanishing: Simp* $\int_X f d\mu = 0$ *iff* f *is zero almost everywhere.*

A consequence of these properties is the *Inclusion-exclusion principle.* Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $A_1, \ldots, A_n$ be $\mathcal{B}$-measurable sets of finite measure. Then,

$$\mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{J \subset \{1,\ldots,n\}: J \neq \emptyset} (-1)^{|J|-1} \mu\left(\bigcap_{i \in J} A_i\right)$$

From a simple integral, we can now define the unsigned integral, in analogy to how the unsigned Lebesgue integral was constructed before.

Definition 4.12. *Let* $(X, \mathcal{B}, \mu)$ *be a measure space, and let* $f : X \to [0, +\infty]$ *be measurable. Then we define the unsigned integral* $\int_X f d\mu$ *of* f *by the formula* $\int_X f d\mu := \sup_{0 \leq g \leq f; g \text{ simple}} \text{Simp} \int_X g d\mu$.

If $f : \boldsymbol{R}^d \to [0, +\infty]$ is Lebesgue measurable, then $\int_{\boldsymbol{r}^d} f(x) dx = \int_{\boldsymbol{R}^d} f dm.$

Lemma 4.2. *(Basic properties of unsigned integral) Let* $(X, \mathcal{B}, \mu)$ *be a measure space, and let* $f, g : X \to [0, +\infty]$ *be measurable.*

- *Almost everywhere equivalence: If* $f = g$ μ*-almost everywhere, then* $\int_x f d\mu = \int_X g d\mu$.

- *Monotonicity: If* $f \leq g$ μ*-almost everywhere, then* $\int_X f d\mu \leq \int_X g d\mu$.

- *Homogeneity: We have* $\int_X c f d\mu = c \int_X f d\mu$ *for every* $c \in [0, +\infty]$.

- *Superadditivity: We have* $\int_X (f + g) d\mu \geq \int_X f d\mu + \int_X g d\mu$.

- *Compatibility with the simple integral: If* f *is simple, then* $\int_X f d\mu = \text{Simp} \int_X f d\mu$.

- *Markov's inequality: For any* $0 < \lambda < \infty$, *one has* $\mu(\{x \in X : f(x) \geq \lambda\}) \leq \frac{1}{\lambda} \int_X f d\mu$. *In particular if* $\int_X f d\mu < \infty$, *then the sets* $\{x \in X : f(x) \geq \lambda\}$ *have finite measure for each* $\lambda > 0$.

- *Finiteness: If* $\int_X f d\mu < \infty$, *then* $f(X)$ *is finite for* μ*-almost every* x.

- *Vanishing: If* $\int_X f d\mu = 0$, *then* $f(x)$ *is zero for* μ*-almost every* x.

- *Vertical truncation: We have* $\lim_{n \to \infty} \int_x \min(f, n) d\mu = \int_X f d\mu$.

- *Horizontal truncation: If* $E_1 \subset E_2 \subset \ldots$ *is an increasing sequence of* $\mathcal{B}$*-measurable sets, then* $\lim_{n \to \infty} \int_X f 1_{E_n} d\mu = \int_X f 1_{\cup_{n=1}^{\infty} E_n} d\mu$.

- *Restriction: If* Y *is a measurable subset of* X, *then* $\int_X f 1_Y d\mu = \int_Y f|_Y d\mu|_Y$, *where* $f|_Y : Y \to [0, +\infty]$ *is the restriction of* $f : X \to [0, +\infty]$ *to* Y, *and the restriction* $\mu|_Y$ *is as defined before.*

Theorem 4.1. *Let* $(X, \mathcal{B}, \mu)$ *be a measure space, and let* $f, g : X \to [0, +\infty]$ *be measurable. Then* $\int_X (f+g) d\mu = \int_X f d\mu + \int_X g d\mu$.

The linearity of μ also holds as seen from $\int_X fd(c\mu) = c\int_X fd\mu$ for every $c \in [0, +\infty]$ and if $\mu_1, \mu_2, \ldots$ are a sequence of measures on $\mathcal{B}$, then $\int_X fd\sum_{n=1}^{\infty}\mu_n = \sum_{n=1}^{\infty}\int_X fd\mu_n$. Once one has the unsigned integral, one can define the absolutely convergent integral exactly as in the Lebesgue case.

Definition 4.13. *(Absolutely convergent integral) Let $(X, \mathcal{B}, \mu)$ be a measure space. A measurable function $f : X \to \boldsymbol{C}$ is said to be absolutely integrable if the unsigned integral $\|f\|_{L^1(X,\mathcal{B},\mu)} := \int_X |f|d\mu$ is finite, and use $L^1(X, \mathcal{B}, \mu)$, $L^1(X)$, or $L^1(\mu)$ to denote the space of absolutely integrable functions. If f is real-valued and absolutely integrable, we define the integral $\int_X fd\mu$ by the formula $\int_X fd\mu := \int_X f_+d\mu - \int_X f_{d\mu}$ where $f_+ := max(f, 0)$, $f_{:=max(-f,0)}$ are the magnitudes of the positive and negative components of f. If f is complex-valued and absolutely integrable, we define the integral $\int_X fd\mu$ by the formula $\int_X fd\mu := \int_X \Re fd\mu + i\int_X \Im fd\mu$ where the two integrals on the right are interpreted as real-valued integrals.*

Proposition 4.1. *(Properties of absolutely convergent integral) Let $(X, \mathcal{B}, \mu)$ be a measure space.*

- *$L^1(X, \mathcal{B}, \mu)$ is a complex vector space.*
- *The integration map $f \mapsto \int_X fd\mu$ is a complex linear map from $L^1(X, \mathcal{B}, \mu)$ to $\boldsymbol{C}$.*
- *Triangle inequality: $\|f + g\|_{L^1(\mu)} \leq \|f\|_{L^1(\mu)} + \|g\|_{L^1(\mu)}$.*
- *Homogeneity: $\|cf\|_{L^1(\mu)} = |c|\|f\|_{L^1(\mu)}$ for all $f, g \in L^1(X, \mathcal{B}, \mu)$ and $c \in \boldsymbol{C}$.*
- *If $f, g \in L^1(X, \mathcal{B}, \mu)$ are such that $f(x) = g(x)$ for μ-almost every $x \in X$, then $\int_X fd\mu = \int_X gd\mu$.*
- *If $f \in L^1(X, \mathcal{B}, \mu)$ and $(X, \mathcal{B}', \mu')$ is a refinement of $(X, \mathcal{B}, \mu)$, then $f \in L^1(X, \mathcal{B}', \mu')$ and $\int_X fd\mu' = \int_X fd\mu$.*
- *If $f \in L^1(X, \mathcal{B}, \mu)$, then $\|f\|_{L^1(\mu)} = 0$ iff f is zero μ-almost everywhere.*
- *If $Y \subset X$ is $\mathcal{B}$-measurable and $f \in L^1(X, \mathcal{B}, \mu)$, then $f|_{Y \in L^1(X, \mathcal{B}_Y, \mu_Y)}$ and $\int_Y f|_Y d\mu|_Y = \int_X f1_Y d\mu$.*

4.5 The convergence theorems

Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f_1, f_2, \ldots : X \to [0, +\infty]$ be a sequence of measurable functions. Suppose that as $n \to \infty$, $f_n(x)$ converges pointwise either everywhere, or μ-almost everywhere, to a measurable limit f. When would such a pointwise convergence imply convergence of the integral, or in other words when can we ensure that one can interchange integrals and limits, i.e $\lim_{n\to\infty}\int_X f_n d\mu \stackrel{?}{=} \int_X \lim_{n\to\infty} f_n d\mu$. We give three important examples to see why this might fail.

- Escape to horizontal infinity: Let X be the real line with Lebesgue measure, and let $f_n := 1_{[n,n+1]}$. Then f_n converges pointwise to $f := 0$, but $\int_{\boldsymbol{R}} f_n(x)dx = 1$ does not converge to $\int_{\boldsymbol{R}} f(x)dx = 0$. Somehow, all the mass in the f_n has escaped by moving off to infinity in a horizontal direction, leaving none behind for the pointwise limit f.
- Escape to width infinity: Let X be the real line with Lebesgue measure, and let $f_n := \frac{1}{n}1_{[0,n]}$. Then f_n now converges uniformly $f := 0$, but $\int_{\boldsymbol{R}} f_n(x)dx = 1$ does not converge to $\int_{\boldsymbol{R}} f(x)dx = 0$. The increasing wide nature of the support of the f_n prevents this from happening.
- Escape to vertical infinity: Let X be the unit interval $[0, 1]$ with Lebesgue measure, restricted from $\boldsymbol{R}$, and let $f_n : n1_{[\frac{1}{n},\frac{2}{n}]}$. Now, we have finite measure, and f_n converges pointwise to f, but no uniform convergence. And again, $\int_{[0,1]} f_n(x)dx = 1$ is not converging to $\int_{[0,1]} f(x)dx = 0$. This time the mass has escaped vertically, through the increasingly large values of f_n.

These are the three ways in which the phase space fails to be compact for the time-frequency functions - escape to spatial infinity, escape to zero frequency, and escape to infinite frequency. Once we shut down these avenues of escape to infinity, one can recover the convergence of the integral. There are two major ways to accomplish this. One is to enforce monotonicity, which prevents each f_n from abandoning the location where the mass of the preceding $f_1, \dots, f_{n-1}$ was concentrated.

Theorem 4.2. *Monotone convergence theorem Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $0 \leq f_1 \leq f_2 \leq \dots$ be a monotone nondecreasing sequence of unsigned measurable functions of X. Then we have $\lim_{n\to\infty} \int_X f_n d\mu = \int_X \lim_{n\to\infty} f_n d\mu$.*

The result still holds if the monotonicity $f_n \leq f_{n+1}$ only holds almost everywhere rather than everywhere.

Lemma 4.3. *(Tonelli's theorem for sums and integrals) Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f_1, f_2, \dots : X \to [0, +\infty]$ be a sequence of unsigned measurable functions. Then one has $\int_X \sum_{n=1}^{\infty} f_n d\mu = \sum_{n=1}^{\infty} \int_X f_n d\mu$.*

Lemma 4.4. *(Borel-Cantelli lemma) Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $E_1, E_2, E_3, \dots$ be a sequence of $\mathcal{B}$-measurable sets such that $\sum_{n=1}^{\infty} \mu(E_n) < \infty$. Then, almost every $x \in X$ is contained in at most finitely many of the E_n, i.e. $\{n \in \mathcal{N} : x \in E_n\}$ is finite for almost every $x \in X$.*

When one does not have monotonicity, one can at least obtain an important inequality, known as Fatou's lemma.

Lemma 4.5. *(Fatou's lemma) Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f_1, f_2, \dots : X \to [0, +\infty]$ be a sequence of unsigned measurable functions. Then $\int_X \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int_X f_n d\mu$.*

Fatou's lemma tells us that when taking the pointwise limit of unsigned functions f_n, that mass $\int_X f_n d\mu$ can be destroyed in the limit, but it cannot be created in the limit. The unsigned hypothesis is necessary here. Finally, the other major way to shut down loss of mass via escape to infinity, is to dominate all of the functions involved by an absolutely convergent one. This result is known as the dominated convergence theorem.

Theorem 4.3. *(Dominated convergence theorem) Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f_1, f_2, \dots : X \to \boldsymbol{C}$ be a sequence of measurable functions that converge pointwise μ-almost everywhere to a measurable limit $f : X \to \boldsymbol{C}$. Suppose that there is an unsigned absolutely integrable function $G : X \to [0, +\infty]$ such that $|f_n|$ are pointwise μ-almost everywhere bounded by G for each n. Then we have $\lim_{n\to\infty} \int_X f_n d\mu = \int_X f d\mu$.*

This statement fails if there is no absolutely integrable dominating function G. The monotone convergence theorem is, in some sense, a defining property of the unsigned integral.

Proposition 4.2. *(Characterisation of the unsigned integral) Let $(X, \mathcal{B})$ be a measurable space. $I : f \mapsto I(f)$ be a map from the space $\mathcal{U}(X, \mathcal{B})$ of unsigned measurable functions $f : X \to [0, +\infty]$ to $[0, +\infty]$ that obeys the following axioms*

- *Homogeneity: For every $f \in \mathcal{U}(X, \mathcal{B})$ and $c \in [0, +\infty]$, one has $I(cf) = cI(f)$.*
- *Finite additivity: For every $f, g \in \mathcal{U}(X, \mathcal{B})$, one has $I(f + g) = I(f) + I(g)$.*
- *Monotone convergence: If $0 \leq f_1, \leq f_2 \leq \dots$ are a non-decreasing sequence of unsigned measurable functions, then $I(\lim_{n\to\infty} f_n) = \lim_{n\to\infty} I(f_n)$.*

Then there exists a unique measure μ on $(X, \mathcal{B})$ such that $I(f) = \int_X f d\mu$ for all $f \in \mathcal{U}(X, \mathcal{B})$. Furthermore, μ is given by the formula $\mu(E) := I(1_E)$ for all $\mathcal{B}$-measurable set E.

5 Modes of convergence

For a sequence of d-dimensional vectors it is unambiguous what it means for that sequence to converge to a limit. To be concrete, for a sequence $v_1, v_2, v_3, \dots$ converging to v ($\in \boldsymbol{R}^d$ or $\in \boldsymbol{C}^d$) for every $\varepsilon > 0$, there exists an N such that $\|v_n - v\| \leq \varepsilon$ for all $n \geq N$. For the purpose of convergence, all the norms are equivalent - Euclidean norm $\|v\|_2$, supremum norm $\|v\|_\infty$. If one however has a sequence $f_1, f_2, f_3, \dots$ of functions $f_n : X \to \boldsymbol{R}$ or $f_n : X \to \boldsymbol{C}$ on a common domain X, and a limit $f : X \to \boldsymbol{R}$ or $f : X \to \boldsymbol{C}$, there can be many different ways in which the sequence f_n may or may not converge to the limit f. This is because once X becomes infinite, the function f_n acquire an infinite number of degrees of freedom, and this allows them to approach f in any number of nonequivalent ways.

There are two basic *modes of convergence*:

- We say that f_n converges to f pointwise if, for every $x \in X$, $f_n(x)$ converges to $f(x)$, i.e. for every $\varepsilon > 0$ and $x \in X$, there exists N (that depends on both ε and x) such that $|f_n(x) - f(x)| \leq \varepsilon$ whenever $n \geq N$.
- We say that f_n converges to f uniformly if, for every $\varepsilon > 0$, there exists N such the for every $n \geq N$, $|f_n(x) - f(x)| \leq \varepsilon$ for every $x \in X$. Here N is not permitted to depend on x, but must be chosen uniformly in x.

Uniform convergence implies pointwise convergence, but not conversely, e.g. $f_n(x) := \frac{x}{n}$ converges pointwise to $f(x) := 0$, but not uniformly. However, there are other modes of convergence. We discuss some of the modes of convergence that arise from measure theory, when the domain X is equipped with the structure of a measure space $(X, \mathcal{B}, \mu)$, and the functions f_n, and their limit f, are measurable with respect to this space.

- We say that f_n converges to f pointwise almost everywhere if, for μ-almost everywhere $x \in X$, $f_n(x)$ converges to $f(x)$.
- We say that f_n converges to f uniformly almost everywhere or in L^∞ norm if, for every $\varepsilon > 0$, there exists N such that for every $n \geq N$, $|f_n(x) - f(x)| \leq \varepsilon$ for μ-almost every $x \in X$.
- We say that f_n converges to f almost uniformly if, for every $\varepsilon > 0$, there exists an exceptional set $E \in \mathcal{B}$ of measure $\mu(E) \leq \varepsilon$ such that f_n converges uniformly to f on the complement of E.
- We say that f_n converges to f in L^1 norm if the quantity $\|f_n - f\|_{L^1(\mu)} = \int_X |f_n(x) - f(x)| d\mu$ converges to 0 as $n \to \infty$.
- We say that f_n converges to f in measure if, for every $\varepsilon > 0$, the measures $\mu(\{x \in X : |f_n(x) - f(x)| \geq \varepsilon\})$ converge to zero as $n \to \infty$.

Each of these five modes of convergence is unaffected if one modifies f_n or f on a set of measure zero. In contrast, the pointwise and uniform modes of convergence can be affected if one modifies f_n or f even on a single point. The L^1 and L^∞ modes of converges are special cases of the L^p mode of convergence. In context of probability theory where f_n and f are interpreted as random variables, convergence in L^1 norm is often referred to as convergence in mean, pointwise convergence almost everywhere is often referred to as almost sure convergence, and convergence in measure is often referred to as convergence in probability.

Proposition 5.1. *Let $(X, \mathcal{B}, \mu)$ be a measure space, and let $f_n : X \to \boldsymbol{C}$ and $f : X \to \boldsymbol{C}$ be measurable functions.*

- *If f_n converges to f uniformly, then f_n converges to f pointwise.*
- *If f_n converges to f uniformly, then f_n converges to f in L^∞ norm. Conversely, if f_n converges to f in L^∞ norm, then f_n converges to f uniformly outside of a null set.*
- *If f_n converges to f in L^∞ norm, then f_n converges to f almost uniformly.*
- *If f_n converges to f almost uniformly, then f_n converges to f pointwise almost everywhere.*
- *If f_n converges to f pointwise, then f_n converges to f pointwise almost everywhere.*

- *If f_n converges to f in L^1 norm, then f_n converges to f in measure.*
- *If f_n converges to f almost uniformly, then f_n converges to f in measure.*

We give examples to distinguish between these modes, in the case when X is the real line $\boldsymbol{R}$ with Lebesgue measure.

- Escape to horizontal infinity: Let $f_n := 1_{[n,n+1]}$. Then f_n converges to zero pointwise, thus pointwise everywhere, but not uniformly, in L^∞ norm, almost uniformly in L^1 norm, or in measure.
- Escape to width infinity: Let $f_n : \frac{1}{n}1_{[0,n]}$. Then f_n converges to zero uniformly, thus pointwise, pointwise almost everywhere, in L^∞ norm, almost uniformly, in measure, but not in L^1 norm.
- Escape to vertical infinity: Let $f_n := n1_{[\frac{1}{n},\frac{2}{n}]}$. Then f_n converges to zero pointwise, thus pointwise almost everywhere, almost uniformly, hence in measure, but not uniformly, not in L^∞, not in L^1 norm.
- Typewriter sequence: Let $f_n := 1_{[\frac{n-fk}{2^k},\frac{n-2^k+1}{2^k}]}$ whenever $k \geq 0$ and $2^k \leq n < 2^{k+1}$. This is a sequence of indicator functions of intervals of decreasing length, marching across the unit interval $[0,1]$ over and over again. Then f_n converges to zero in measure and in L^1 norm, but not pointwise almost everywhere, hence also not pointwise, not almost uniformly, nor in L^∞ norm, nor uniformly.

The L^∞ norm $\|f\|_{L^\infty(\mu)}$ of a measurable function $f : X \to \boldsymbol{C}$ is defined to the infimum of all the quantities $M \in [0,+\infty]$ that are essential upper bounds for f in the sense that $|f(x)| \leq M$ for almost every x. Then f_n converges to f in L^∞ norm iff $\|f_n - f\|_{L^\infty(\mu)} \to 0$ as $n \to \infty$. Then L^∞ and L^1 norms are part of the the larger family of L^p norms. One particular advantage of L^1 convergence is that, in the case when f_n are absolutely integrable, it implies convergence of the integrals $\int_X f_n d\mu \to \int_X f d\mu$, via triangle inequality. None of the other modes of convergence automatically imply this convergence of the integral. Also note that one cannot rank these modes in a linear order from strongest to weakest. If one imposes some additional assumptions to shut down one or more of these escape to infinity scenarios, such as *finite measure hypothesis* $\mu(X) < \infty$ or a *uniform integrability hypothesis*, then one can obtain some additional implications between the different modes.

5.1 Uniqueness

Even though the modes of convergence all differ form each other, they are all compatible in the sense that they never disagree about which function f a sequence of functions f_n converges to, outside of a set of measure zero.

Proposition 5.2. *Let $f_n : X \to \boldsymbol{C}$ be a sequence of measurable functions, and let $f, g : X \to \boldsymbol{C}$ be two additional measurable functions. Suppose that f_n converges to f along one of the seven modes of convergence defined above, and f_n converges to g along another of the seven modes of convergence. Then f and g agree almost everywhere.*

5.2 The case of a step function

Consider the case of $f = 0$ and each of the f_n a step function, i.e. $f_n = A_n 1_{E_n}$ of a measurable set E_n, with $A_n > 0$ assumed to be positive reals and E_n has a positive measure $\mu(E_n) > 0$. We assume that either A_n converges to zero or is bounded away from zero, i.e. $\exists c > 0$ such that $A_n \geq c$, $\forall n$. We now ask, what it means for this sequence f_n to converge to zero for the seven modes of convergence? The answer depends on three quantities: the *height* A_n of the nth function f_n, the *width* $\mu(E_n)$ of the nth function f_n, and the Nth tail support $E_N^* := \cup_{n \geq N} E_n$ for the sequence $f_1, f_2, \ldots$. In fact,

- f_n converges uniformly to zero iff $A_n \to 0$ as $n \to \infty$.
- f_n converges in L^∞ norm to zero iff $A_n \to 0$ as $n \to \infty$.
- f_n converges almost uniformly to zero iff $A_n \to 0$ as $n \to \infty$, or $\mu(E_N^*) \to 0$ as $N \to \infty$.
- f_n converges pointwise to zero iff $A_n \to 0$ as $n \to \infty$, or $\cap_{N=1}^\infty E_N^* = \emptyset$.
- f_n converges pointwise almost everywhere to zero iff $A_n \to 0$ as $n \to \infty$, or $\cap_{N=1}^\infty E_N^*$ is a null set.

- f_n converges in measure to zero iff $A_n \to 0$ as $n \to \infty$, or $\mu(E_n) \to 0$ as $n \to \infty$.

- f_n converges in L^1 norm iff $A_n\mu(E_n) \to 0$ as $n \to \infty$.

Informally, when the height goes to zero, then one has convergence to zero in all modes except possibly for L^1 convergence, which requires that the product of the height and the width goes to zero. If instead the height is bounded away from zero and the width is positive, then we never have uniform or L^∞ convergence, but we have convergence in measure if the width goes to zero, we have almost uniform convergence if the tail support has measure that goes to zero, we have pointwise almost everywhere convergence if the tail support shrinks to a null set, and the pointwise convergence if the tail support shrinks to the empty set.

Comparing it to the four examples given before we see:

- In the escape to horizontal infinity scenario, the height and width do not shrink to zero, but the tail set shrinks to the empty set.

- In the escape to the width infinity scenario, the height goes to zero, but the width and tail support go to infinity, causing the L^1 norm to stay bounded away from zero.

- In the escape to vertical infinity, the height goes to infinity, but the width and tail support go to zero or the empty set, causing the L^1 norm to stay bounded away from zero.

- In the typewriter example, the width goes to zero, but the height and the tail support stay fixed and thus bounded away from zero.

$f_n = A_n 1_{E_n}$ are monotone increasing iff $A_n \leq A_{n+1}$ and $E_n \subset E_{n+1}$ for each n. In such cases, f_n converges pointwise to $f := A1_E$, where $A := \lim\limits_{n\to\infty} A_n$ and $E := \cup_{n=1}^{\infty} E_n$. The monotone convergence theorem then asserts that $A_n\mu(E_n) \to A\mu(E)$ as $n \to \infty$.

5.3 Finite measure spaces

The situation simplifies somewhat in the space X has finite measure, e.g. in a probability space. This shuts down two of the four escapes: escape to horizontal infinity or width infinity, creating a few more equivalences.

Theorem 5.1. *(Egorov's theorem) Let X have finite measure, and let $f_n : X \to \boldsymbol{C}$ and $f : X \to \boldsymbol{C}$ be measurable functions. Then f_n converges to f pointwise almost everywhere iff f_n converges to f almost uniformly.*

Another feature of the finite measure case is that L^∞ convergence implies L^1 convergence.

5.4 Fast convergence

The typewriter example shows that L^1 convergence is not strong enough to force almost uniform or pointwise almost everywhere convergence. However, this can be rectified if one assumes that the L^1 convergence is sufficiently fast:

Proposition 5.3. *(Fast L^1 convergence) Suppose that $f_n, f : X \to \boldsymbol{C}$ are measurable functions such that $\sum\limits_{n=1}^{\infty} \|f_n - f\|_{L^1(\mu)} < \infty$; thus not only do the quantities $\|f_n - f\|_{L^1(\mu)}$ go to zero, which would mean L^1 convergence, but they converge in an absolutely summable fashion.*

- *f_n converges pointwise almost everywhere to f.*

- *f_n converges almost uniformly to f.*

Thus, L^1 convergence implies almost uniform or pointwise almost everywhere convergence if we are allowed to pass to a subsequence.

Proposition 5.4. *Suppose that $f_n : X \to \boldsymbol{C}$ are a sequence of measurable functions that converge in L^1 norm to a limit f. Then there exists a subsequence f_{n_j} that converges almost uniformly, and hence pointwise almost everywhere, to f, while remaining convergence in L^1 norm.*

One can strengthen the corollary a bit by relaxing L^1 convergence to convergence in measure:

Proposition 5.5. *Suppose that $f_n : X \to \boldsymbol{C}$ are a sequence of measurable functions that converge in measure to a limit f. Then there exists a subsequence f_{n_j} that converges almost uniformly, and hence pointwise almost everywhere, to f.*

5.5 Domination and uniform integrability

Now we turn to the reverse question, of whether almost uniform convergence, pointwise almost everywhere convergence, or convergence in measure can imply L^1 convergence. In general, the answer is no. However, one can do better if one places some domination hypothesis on the f_n that shut the vertical and width infinity escape routes. We say that a sequence $f_n : X \to \boldsymbol{C}$ is dominated if there exists an absolutely integrable function $g : X \to C$ such that $|f_n(x)| \leq g(x)$ for all n and almost every x. For instance, if X has finite measure and the f_n are uniformly bounded, then they are dominated. The dominated convergence theorem then asserts that if f_n converges to f pointwise almost everywhere, then it necessarily converges to f in L^1 norm, and hence also in measure.

There is a more general notion than domination, known as uniform integrability, which serves as a substitute for domination in many, but not all, contexts.

Definition 5.1. *(Uniform integrability) A sequence $f_n : X \to \boldsymbol{C}$ of absolutely integrable functions is said to be uniformly integrable if the following three statements hold:*

- *Uniform bound on L^1 norm: One has $\sup_n \|f_n\|_{L^1(\mu)} = \sup_n \int_X |f_n| d\mu < +\infty$.*
- *No escape to vertical infinity: One has $\sup_n \int_{|f_n| \geq M} |f_n| d\mu \to 0$ as $M \to +\infty$.*
- *No escape to width infinity: One has $\sup_n \int_{|f_n| \leq \delta} |f_n| d\mu \to 0$ as $\delta \to 0$.*

For the case of step function $f_n = A_n 1_{E_n}$, the uniform bound on the L^1 norm then asserts that $A_n \mu(E_n)$ stays bounded. The lack of escape to vertical infinity means that along any subsequence for which $A_n \to \infty$, $A_n \mu(E_n)$ must go to zero. Similarly, lack of escape to width infinity means that along any subsequence for which $A_n \to 0$, $A_n \mu(E_n)$ must go to zero.

In case of finite measure space, there is no escape to width infinity, and the criterion for uniform integrability simplifies to just that of excluding vertical infinity. The dominated convergence theorem does not have an analogue in the uniformly integrable settings. However,

Theorem 5.2. *(Uniform integrable convergence in measure) Let $f_n : X \to \boldsymbol{C}$ be a uniformly integrable sequence of functions, and let $f : X \to \boldsymbol{C}$ be another function. Then f_n converges in L^1 norm to f iff f_n converges to f in measure.*

Let X be a probability space. Given any real-values measurable function $f : X \to \boldsymbol{R}$, we define the cumulative distribution function $F : \boldsymbol{R} \to [0, 1]$ of f to be the function $F(\lambda) := \mu(\{x \in X; f(x) \leq \lambda\})$. Given another sequence $f_n : X \to \boldsymbol{R}$ of real-valued measurable functions, we say that f_n *converges in distribution* to f if the cumulative distribution function $F_n(\lambda)$ of f_n converges pointwise to the cumulative distribution function $F(\lambda)$ of f at all $\lambda \in \boldsymbol{R}$ for which F is continuous. If f_n converges to f in any of the seven senses discussed, then it converges in distribution to f. Even if f_n converges to f in distribution, it may not converge in the above seven senses. If f_n converges to f in distribution, and g_n converges to g, then $f_n + g_n$ need not converge to $f + g$, and hence is not linear. Finally, f_n can converge in distribution to two different limits f, g which are not equal almost everywhere. Thus convergence in distribution, commonly used in probability, is quite a weak notion of convergence.

6 Differentiation theorems

Let $[a, b]$ be a compact interval of positive length, $-\infty < a < b < +\infty$. The function $F : [a, b] \to \boldsymbol{R}$ is said to be differentiable at a point $x \in [a, b]$ if the limit $F'(x) := \lim_{y \to x; y \in [a,b] \setminus x} \frac{F(y) - f(x)}{y - x}$ exists. This is called *classical derivative.* of F at x. We say F is everywhere differentiable if it is differentiable at all points $x \in [a, b]$. If F is differentiable everywhere and its derivative F' is continuous, then we say that F is continuously differentiable. If $F : [a, b] \to \boldsymbol{R}$ is everywhere differentiable, then F is continuous and F' is measurable. If F is almost everywhere differentiable, then the almost everywhere defined function F' is measurable, but F need not be continuous.

In single-variable calculus, the operations of integration and differentiation are connected by a number of basic theorems, starting with Rolle's theorem.

Theorem 6.1. *(Rolle's theorem) Let $[a, b]$ be a compact interval of positive length, and let $F : [a, b] \to \boldsymbol{R}$ be a differentiable function such that $F(a) = F(b)$. Then there exists $x \in (a, b)$ such that $F'(x) = 0$.*

Rolle's theorem can fail if f is merely assumed to be almost everywhere differentiable, even if one adds the additional hypothesis that f is continuous. Thus, everywhere differentiability is a significantly stronger property than almost everywhere differentiability. Rolle's theorem only works in the real scalar case when F is real valued, as it relies on the least upper bound property for the domain $\boldsymbol{R}$. If, for instance, we consider complex-valued scalar functions, or finite-dimensional vector space functions like $\boldsymbol{R}^n$, then the theorem can fail. One can easily amplify Rolle's theorem to the mean value theorem:

Theorem 6.2. *(Mean value theorem) Let $[a, b]$ be a compact interval of positive length, and let $F : [a, b] \to \boldsymbol{R}$ be a differentiable function. Then there exists $x \in (a, b)$ such that $F'(x) = \frac{F(b) - F(a)}{b - a}$.*

Mean value theorem is also only applicable to real scalar functions. Let $[a, b]$ be a compact interval of positive length, and let $F"[a, b] \to \boldsymbol{R}$ and $G : [a, b] \to \boldsymbol{R}$ be differentiable functions. Then $F'(x) = G'(x)$ for every $x \in [a, b]$ iff $F(x) = G(x) + C$ for some constant $C \in \boldsymbol{R}$ and all $x \in [a, b]$. We can use mean value theorem to deduce the second fundamental theorems of calculus:

Theorem 6.3. *(fundamental theorems of calculus)*

- *First fundamental theorem of calculus: Let $[a, b]$ be a compact interval of positive length. Let $f : [a, b] \to \boldsymbol{C}$ be a continuous function, and let $F : [a, b] \to \boldsymbol{C}$ be the indefinite integral $F(x) := \int_a^x f(t)dt$. Then F is differentiable on $[a, b]$, with derivative $F'(x) = f(x)$ for all $x \in [a, b]$. In particular, F is continuously differentiable.*
- *Second fundamental theorem of calculus: Let $F : [a, b] \to \boldsymbol{R}$ be a differentiable function, such that F' is Riemann integrable. Then the Riemann integral $\int_a^b F'(x)dx$ of F' is equal to $F(b) - F(a)$. In particular, we have $\int_a^b F'(x)dx = F(b) - F(a)$ when ever F is continuously differentiable.*

Even though the mean value theorem only holds for real scalar functions, the fundamental theorem of calculus holds for complex or vector-valued functions, as one can simply apply that theorem to each component of that function separately.

Corollary 6.1. *(Differentiation theorem for continuous functions) Let $f : [a, b] \to \boldsymbol{C}$ be a continuous function on a compact interval. Then we have $\lim_{h \to 0^+} \frac{1}{h} \int_{[x, x+h]} f(t)dt = f(x)$ for all $x \in [a, b)$, $\lim_{h \to 0^+} \frac{1}{h} \int_{[x-h, x]} f(t)dt = f(x)$ for all $x \in (a, b]$, and thus $\lim_{h \to 0^+} \frac{1}{2h} \int_{[x-h, x+h]} f(t)dt = f(x)$ for all $x \in (a, b)$.*

In this chapter we explore the question of the extent to which these theorems continue to hold when the differentiability or integrability conditions on the various functions F, F', f are relaxed. Among the results proven in this chapter are:

- The Lebesgue differentiation theorem, which roughly speaking asserts that Differentiation theorem for continuous functions continue to hold for almost every x if f is merely absolutely integrable, rather than continuous.

- A number of differentiation theorems, which assert for instance that monotone, Lipschitz, or bounded variation functions in one dimension are almost everywhere differentiable.
- The second fundamental theorem of calculus for absolutely continuous functions.

6.1 The Lebesgue differentiation theorem in one dimension

We want to build up the following theorem in this section.

Theorem 6.4. *(Lebesgue differentiation theorem in one-dimension) Let $f : \boldsymbol{R} \to \boldsymbol{C}$ be an absolutely integrable function, and let $F : \boldsymbol{R} \to \boldsymbol{C}$ be the definite integral $F(x) := \int_{[-\infty,x]} f(t)dt$. Then F is continuous and almost everywhere differentiable, and $F'(x) = f(x)$ for almost every $x \in \boldsymbol{R}$.*

This can be viewed as a variant of differentiation theorem for continuous functions; the hypotheses are weaker because f is only assumed to be absolutely integrable, rather than continuous. The conclusion is weaker too because F is only found to be almost everywhere differentiable, rather than everywhere differentiable.

It is easy to show the continuity part. Let $f : \boldsymbol{R} \to \boldsymbol{C}$ be an absolutely integrable function, and let $F : \boldsymbol{R} \to \boldsymbol{C}$ be the definite integral $F(x) := \int_{[-\infty,x]} f(t)dt$. Then F is continuous. The main difficulty is to show that $F'(x) = f(x)$ for almost every $x \in R$. This can be shown from the following theorem:

Theorem 6.5. *Let $f : \boldsymbol{R} \to \boldsymbol{C}$ be an absolutely integrable function. Then $\lim_{h\to 0^+} \frac{1}{h}\int_{[x,x+h]} f(t)dt = f(x)$ for almost every $x \in \boldsymbol{R}$, and $\lim_{h\to 0^+} \frac{1}{h}\int_{[x-h,x]} f(t)dt = f(x)$ for almost every $x \in \boldsymbol{R}$.*

This is a convergence theorem and is generally proved using *density argument*. We will need the quantitative estimate given by the special case of the Hardy-Littlewood maximal inequality.

Lemma 6.1. *(One sided Hardy-Littlewood maximal inequality) Let $f : \boldsymbol{R} \to \boldsymbol{C}$ be an absolutely integrable function, and let $\lambda > 0$. Then $m(\{x \in \boldsymbol{R} : \sup_{h>0} \frac{1}{h}\int_{[x,x+h]} |f(t)|dt \geq \lambda\}) \leq \frac{1}{\lambda}\int_{\boldsymbol{R}} |f(t)|dt$*

This can now be used to prove the Lebesgue differentiation theorem.

6.2 The Lebesgue differentiation theorem in higher dimension

Now we extend the Lebesgue differentiation theorem to higher dimensions.

Theorem 6.6. *(Lebesgue differentiation theorem in general dimension) Let $f : \boldsymbol{R} \to \boldsymbol{C}$ be an absolutely integrable function. Then for almost every $x \in \boldsymbol{R}^d$, one has $\lim_{r\to 0} \frac{1}{m(B(x,r))}\int_{B(x,r)} |f(y) - f(x)|dy = 0$ and $\lim_{r\to 0} \frac{1}{m(B(x,r))}\int_{B(x,r)} f(y)dy = f(x)$, where $B(x,r) := \{y \in \boldsymbol{R}^d : |x-y| < r\}$ is the open ball of radius r centered at x.*

A point x for which the theorem holds is called a *Lebesgue point* of f; thus, for an absolutely integrable function f, almost every point in $\boldsymbol{R}^d$ will be Lebesgue point for $\boldsymbol{R}^d$. To prove the theorem we use the density argument again.

Given a Lebesgue measurable set $E \subset \boldsymbol{R}^d$, call a point $x \in \boldsymbol{R}^d$ a point of density for E if $\frac{m(E\cap B(x,r))}{m(B(x,r))} \to 1$ as $r \to 0$. Thus, for instance, if $E = [-1,1]\{0\}$, then every point in $(-1,1)$, including 0, is a point of density for E, but the endpoints $-1, 1$, as well as the exterior of E, are not points of density.

6.3 Almost everywhere differentiability

The function $f(x) := |x|$ is continuous but not differentiable everywhere. However, it is still almost everywhere differentiable. We can construct continuous functions that are in fact nowhere differentiable. Let $F : \boldsymbol{R} \to \boldsymbol{R}$ be the function $F(x) := \sum_{n=1}^{\infty} 4^{-n} \sin(8^n \pi x)$. F is well defined, bounded and continuous. However, it is not differentiable at any point $x \in \boldsymbol{R}$. It is called Weierstrass function. There difficulty here is that the continuous function can still contain a large amount of oscillation, which can lead to breakdown of differentiability. If one can limit the amount of oscillation present, then one can recover a fair bit of differentiability.

Theorem 6.7. *(Monotone differentiation theorem) Any function $f : \boldsymbol{R} \to \boldsymbol{R}$ which is monotone is differentiable almost everywhere.*

To understand the differentiability of F, we introduce the four *Dini derivatives* of F at x.

- The upper right derivative $\overline{D^+}F(x) := \limsup_{h\to 0^+} \frac{F(x+h)-F(x)}{h}$.
- The lower right derivative $\underline{D^+}F(x) := \liminf_{h\to 0^+} \frac{F(x+h)-F(x)}{h}$.
- The upper left derivative $\overline{D^-}F(x) := \limsup_{h\to 0^-} \frac{F(x+h)-F(x)}{h}$.
- The lower left derivative $\underline{D^-}F(x) := \liminf_{h\to 0^-} \frac{F(x+h)-F(x)}{h}$.

Regardless of whether F is differentiable or not, or even whether F is continuous or not, the four Dini derivatives always exist and take values in the extended real line $[-\infty, \infty]$. A function F is differentiable at x precisely when the four derivatives are equal and finite. The one-sided Hardy-Littlewood maximal inequality has an analogue in this setting:

Lemma 6.2. *(One-sided Hardy-Littlewood inequality) Let $F : [a, b] \to \boldsymbol{R}$ be a continuous monotone non-decreasing function, and let $\lambda > 0$. Then we have $m(\{x \in [a, b] : \overline{D^+}F(x) \geq \lambda\}) \leq \frac{F(b)-F(a)}{\lambda}$. Similarly for the other three Dini derivatives of F. If F is not assumed to be continuous, then we have the weaker inequality $m(\{x \in [a, b] : \overline{D^+}F(x) \geq \lambda\}) \leq C\frac{F(b)-F(a)}{\lambda}$ for some absolute constant $C > 0$.*

Sending $\lambda \to \lambda$ in the above lemma, and sending $[a, b]$ to $\boldsymbol{R}$, we can conclude that all the four Dini derivatives of a continuous monotone non-decreasing function are finite almost everywhere. We also need Jump functions to remove the requirement of continuity in the Monotone differentiation theorem.

Definition 6.1. *(Jump function) A basic jump function J is a function of the form* $J(x) = \begin{cases} 0 & \text{when } x < x_0 \\ \theta & \text{when } x = x_0 \\ 1 & \text{when } x > x_0 \end{cases}$ *for some real number $x_o \in \boldsymbol{R}$ and $0 \leq \theta \leq 1$; x_0 is called the point of discontinuity for J and θ the fraction. This is a monotone non-decreasing function but have discontinuity. A jump function is any absolutely convergent combination of basic jump functions of the from $F = \sum_n c_n J_n$, where n ranges over an at most countable set, each J_n is a basic jump function, and the c_n are positive reals with $\sum_n c_n < \infty$. If there are only finitely many n involved, we say that F is a piecewise constant jump function.*

Thus, for instance if $q_1, q_2, \ldots$ is any enumeration of the rationals, then $\sum_{n=1}^{\infty} 2^{-n} 1_{[q_n,+\infty]}$ is a jump function. All jump functions are monotone non-decreasing. The key fact is that these functions, together with the continuous monotone functions, essentially generate all monotone functions in the bounded case.

Lemma 6.3. *(Continuous-singular decomposition for monotone functions) Let $F : \boldsymbol{R} \to \boldsymbol{R}$ be a monotone non-decreasing function.*

- *The only discontinuities of F are jump discontinuities. More precisely, if x is a point where F is discontinuous, then the limits $\lim_{y\to x^-} F(y)$ and $\lim_{y\to x^+} F(y)$ both exist, but are unequal, with $\lim_{y\to x^-} F(y) < \lim_{y\to x^-} F(y)$.*
- *There are at most countably many discontinuities of F.*

- *If F is bounded, then F can be expressed as the sum of a continuous monotone non-decreasing function F_c and a jump function F_{pp}.*

These constructions can now be used to prove the Monotone differentiation theorem. Just as the integration theory of unsigned functions can be used to develop the integration theory of the absolutely convergent functions, the differentiation theory of monotone functions can be used to develop the a parallel differentiation theory for the class of functions of *bounded variation*:

Definition 6.2. *(Bounded variations) Let $F : \boldsymbol{R} \to \boldsymbol{R}$ be a function. The total variation $\|F\|_{TV(\boldsymbol{R})}$ or $||F||_{TV}$ for short, of F is defined to be the supremum $\|F\|_{TV(\boldsymbol{R})} := \sup\limits_{x_0<\ldots<x_n} \sum\limits_{i=1}^{n} |F(x_i) - F(x_{i+1})|$ where the supremum ranges over all finite increasing sequences $x_0, \ldots, x_n$ of real numbers with $n \geq 0$; this is a quantity in $[0, +\infty]$. We say that F has bounded variation on $\boldsymbol{R}$ if $\|F\|_{TV(\boldsymbol{R})}$ is finite.*

Given any interval $[a, b]$ the definition is the same, but the points $x_0, \ldots, x_n$ are restricted to lie in $[a, b]$. Much as an absolutely integrable function can be expressed as the difference of its positive and negative parts, a bounded variation function can be expressed as the difference of two bounded monotone functions:

Proposition 6.1. *A function $F : \boldsymbol{R} \to \boldsymbol{R}$ is of bounded variation iff it is the difference of two bounded monotone functions.*

We define the positive variation by $F^+(x) := \sup\limits_{x_0<\ldots<x_n\leq x} \sum\limits_{i=1}^{n} max(F(x_{i+1}) - F(x_i), 0)$ and the negative variation $F^-(x) := \sup\limits_{x_0<\ldots<x_n\leq x} \sum\limits_{i=1}^{n} max(-F(x_{i+1}) + F(x_i), 0)$. Further, $F(x) = F(-\infty + F^+(x) - F^-(x)$, $\|F\|_{TV[a,b]} = F^+(b) - F^+(a) + F^-(b) - F^-(a)$, and $\|F\|_{TV} = F^+(+\infty) + F^-(+\infty)$, for every interval $[a, b]$, where $F(-\infty) := \lim\limits_{x\to-\infty} F(x)$, $F^+(+\infty) = \lim\limits_{x\to+\infty} F^+(x)$, and $F^-(+\infty) = \lim\limits_{x\to+\infty} F^-(x)$.

Theorem 6.8. *(BV differentiation theorem) Every bounded variation function is differentiable almost everywhere.*

A function is called locally of bounded variation if it is of bounded variation on every compact interval $[a, b]$, and is differentiable almost everywhere.

Theorem 6.9. *(Lipschitz differentiation theorem) A function $f : \boldsymbol{R} \to R$ is said to be Lipschitz continuous if there exists a constant $C > 0$ such that $|f(x)-f(y)| \leq C|x-y|$ for all $x, y \in \boldsymbol{R}$; the smallest C with this property is known as the Lipschitz constant of f. Every Lipschitz continuous function F is locally of bounded variation, and hence differentiable almost everywhere. The derivative F', when it exists, is bounded in magnitude by the Lipschitz constant of F.*

The same result in higher dimensions is known as *Radamacher differentiation theorem*, which is proven using Fubini-Tonelli theorem, which is a general mechanism to deduce higher-dimensional results in analysis from lower-dimensional ones.

A function $f : \boldsymbol{R} \to \boldsymbol{R}$ is said to be convex if one has $f(1-t)x + ty) \leq (1-t)f(x) + tf(y)$ for all $x < y$ and $0 < t < 1$. If f is convex, then it is continuous and almost everywhere differentiable, and its derivative f' is equal almost everywhere to a monotone non-decreasing function, so is itself almost everywhere differentiable. Thus convex functions are almost everywhere twice differentiable.

6.4 The second fundamental theorem of calculus

We now look at the second fundamental theorem of calculus in the cases where F is not assumed to be continuously differentiable. We begin with the case when $F[a, b] \to \boldsymbol{R}$ is monotone non-decreasing. By monotone differentiation theorem, F is differentiable almost everywhere in $[a, b]$, so F' is defined almost everywhere; from monotonicity we see that F' is non-negative whenever it is defined and is measurable. One half of the second fundamental theorem is easy:

Theorem 6.10. *(Upper bound for second fundamental theorem) Let $F : [a, b] \to \boldsymbol{R}$ be monotone non-decreasing. Then $\int_{[a,b]} F'(x)dx \leq F(b) - F(a)$. In particular, F' is absolutely integrable.*

Thus, any function of bounded variation has an almost everywhere defined derivative that is absolutely integrable. In the Lipschitz case, one can do better. If $F : [a, b] \to \boldsymbol{R}$ be Lipschitz continuous, then $\int_{[a,b]} F'(x)dx = F(b) - F(a)$. For the monotone case, to recover the equality is problematic. All the variation of F may be concentrated in a set of measure zero, and thus undetectable by the Lebesgue integral of F'. For example the heaviside function $F := 1_{[0,+\infty)}$ vanishes almost everywhere but $F(b) - F(a)$ is not equal to $\int_{[a,b]} F'(x)dx$ if b and a lie on the opposite sides of the discontinuity at 0. This is a problem with all jump functions. It is possible for the jumps to be in uncountable number of locations, e.g. in Cantor function also called *Devil's staircase function*. Essentially, the classical derivative $F'(x) := \lim_{h\to 0; h\neq 0} \frac{F(x+h)-F(x)}{h}$ of a function has some defects; it cannot see some of the variation of a continuous monotone function such as the Contor function. The Lebesgue-Stieltjes integral is possible way to capture all the variation of a monotone function, and which is related to the classical derivative via the *Lebesgue-Radon-Nikodym theorem*.

We add an additional hypothesis of absolute continuity to the continuous monotone non-increasing function F before we can recover the second fundamental theorem. A function $F : \boldsymbol{R} \to \boldsymbol{R}$ is *continuous* if, for every $\epsilon > 0$ and $x_0 \in \boldsymbol{R}$, there exists a $\delta > 0$ such that $|F(b) - F(a)| \leq \epsilon$ whenever (a, b) is an interval of length at most δ that contains x_0. A function $F : \boldsymbol{R} \to \boldsymbol{R}$ is **uniformly continuous** if, for every $\epsilon > 0$, there exists a $\delta > 0$ such that $|F(b) - F(a)| \leq \epsilon$ whenever (a, b) is an interval of length at most δ.

Definition 6.3. *A function $F : \boldsymbol{R} \to \boldsymbol{R}$ is said to be absolutely continuous if, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that $\sum_{j=1}^{n} |F(b_j) - F(a_j)| \leq \varepsilon$ whenever $(a_1, b_1), \ldots, (a_n, b_n)$ is a finite collection of disjoint intervals of total length $\sum_{j=1}^{n} b_j - a_j$ at most δ.*

We define absolute continuity for a function $F : [a, b] \to \boldsymbol{R}$ defined on an interval $[a, b]$ similarly, with the requirement that the intervals $[a_j, b_j]$ like in the domain $[a, b]$ of F. For absolutely continuous functions, we can recover the second fundamental theorem of calculus:

Theorem 6.11. *(Second fundamental theorem of absolutely continuous functions) Let $F : [a, b] \to \boldsymbol{R}$ be absolutely continuous. Then $\int_{[a,b]} F'(x)dx = F(b) - F(a)$.*

A function $F : [a, b] \to \boldsymbol{R}$ is absolutely continuous iff it takes the form $F(x) = \int_{[a,b]} f(y)dy + C$ for some absolutely integrable $f : [a, b] \to \boldsymbol{R}$ and a constant C.

7 Outer measure, pre-measures, and product measure

We have focused primarily on one specific example of a countably additive measure, namely *Lebesgue measure*. This measure was constructed from a more primitive concept of *Lebesgue outer measure*, which in turn was constructed from the even more primitive concept of *elementary measure*. Both these constructions can be abstracted. We give the *Caratheodory extension theorem*, which constructs a countably additive measure from any abstract outer measure; this generalizes the construction of Lebesgue measure from Lebesgue outer measure. One can in turn construct outer measures from another concept known as a *pre-measure*, of which elementary measure is a typical example. With these tools one can start constructing many more measures, like *Lebesgue-Stieltjes measures, product measures*, and *Hausdorff measures*. Which can then help establish *Kolmogorov extension theorem*, which allows one to construct variety of measures on infinite-dimensional spaces, particularly important for probability theory.

7.1 Outer measure and the Caratheodory extension theorem

Definition 7.1. *(Abstract outer measure) Let X be a set. An abstract outer measure is a map $\mu^* : 2^X \to [0, +\infty]$ that assigns an unsigned extended real number $\mu^*(E) \in [0, +\infty]$ to every set $E \subset X$ which obeys the following axioms:*

- *Empty set:* $\mu^*(\emptyset) = 0$.

- *Monotonicity: If $E \subset F$, then $\mu^*(E) \le \mu^*(F)$.*
- *Countable subadditivity: If $E_1, E_2, \ldots \subset X$ is a countable sequence of subsets of X, then $\mu^*(\bigcup\limits_{n=1}^{\infty} E_n) \le \sum\limits_{n=1}^{\infty} \mu^*(E_n)$.*

Outer measures are also known as exterior measures. For instance, Lebesgue outer measure m^* is an outer measure. However since Jordan outer measure $m^{*,(J)}$ is only finitely subadditive it not not an outer measure. Outer measures are weaker than measures in the sense that they are merely countably subadditive, rather than countably additive. On the other hand, they are able to measure all subsets of X, whereas measures can only measure a σ-algebra of measurable sets.

We used Lebesgue outer measure together with the notion of an open set to define the concept of Lebesgue measurability. We don not have the notion of open set in the abstract setting. We, thus, introduce the abstract definition of measurability here:

Definition 7.2. *(Caratheodory measurability) Let μ^* be an outer measure on a set X. A set $E \subset X$ is said to be Caratheodory measurable with respect to μ^* if one has $\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \backslash E)$ for every set $A \subset X$.*

If E is a null set for an outer measure μ^*, i.e. $\mu^*(E) = 0$, then E is Caratheodory measurable with respect to μ^*. The construction of Lebesgue measure can then be abstracted as follows:

Theorem 7.1. *(Caratheodory extension theorem) Let $\mu^* : 2^X \to [0, +\infty]$ be an outer measure on a set X, let $\mathcal{B}$ be the collection of all subsets of X that are Caratheodory measurable with respect to μ^*, and let $\mu : \mathcal{B} \to [0, +\infty]$ be the restriction of μ^* to $\mathcal{B}$, thus $\mu(E) := \mu^*(E)$ whenever $E \in B$. Then $\mathcal{B}$ is a σ-algebra, and μ is a measure.*

The measure constructed by the Caratheodory extension theorem is complete.

7.2 Pre-measures

Finitely additive measures, like elementary or Jordan measure, could be extended to a countably additive measure, namely Lebesgue measure. Given a finitely additive measure $\mu_0 : \mathcal{B}_0 \to [0, +\infty]$ on a Boolean algebra $\mathcal{B}_0$, is it possible to find a σ-algebra $\mathcal{B}$ refining $\mathcal{B}_0$, and a countably additive measure $\mu : \mathcal{B} \to [0, +\infty]$ that extends μ_0? There is a necessary condition: μ_0 already has to be countably additive within $\mathcal{B}_0$, i.e. if $E_1, E_2, \ldots \in \mathcal{B}_0$ were disjoint sets such that their union $\bigcup\limits_{n=1}^{\infty} E_n$ is also in $\mathcal{B}_0$. Then, in order for μ_0 to be extendible to a countably additive measure, it is clearly necessary that $\mu_0(\bigcup\limits_{n=1}^{\infty}) = \sum\limits_{n=1}^{\infty} \mu_0(E_n)$. Using Caratheodory extension theorem, it can be shown that this necessary condition is also sufficient.

Definition 7.3. *(Pre-measure) A pre-measure on a Boolean algebra $\mathcal{B}_0$ is a finitely additive measure $\mu_0 : \mathcal{B}_0 \to [0, +\infty]$ with the property that $\mu_0(\bigcup\limits_{n=1}^{\infty} E_n) = \sum\limits_{n=1}^{\infty} \mu_0(E_n)$ whenever $E_1, E_2, E_3, \ldots \in \mathcal{B}_0$ are disjoint sets such that $\bigcup\limits_{n=1}^{\infty} E_n$ is in $\mathcal{B}_0$.*

Theorem 7.2. *(Hahn-Kolmogorov theorem) Every pre-measure $\mu_0 : \mathcal{B}_0 \to [0, +\infty]$ on a Boolean algebra $\mathcal{B}_0$ in X can be extended to a countably additive measure $\mu : \mathcal{B} \to [0, +\infty]$.*

The Hahn-Kolmogorov extension of elementary measure is Lebesgue measure. This is not quite the unique extension of μ_0 to a countably additive measure, though. On could, for instance, restrict Lebesgue measure to the Borel σ-algebra, and this would still be a countably additive extension of elementary measure. However, the extension is unique within its own σ-algebra. This theorem can now be used to construct a variety of measures.

7.3 Lebesgue-Stieltjes measure

Theorem 7.3. *(Existence of Lebesgue-Stieltjes measure) Let $F : \boldsymbol{R} \to \boldsymbol{R}$ be a monotone non-decreasing function, and define the left and right limits $F_-(x) := \sup\limits_{y<x} F(y)$; $F_+(x) := \inf\limits_{y>x} F(y)$, thus one has $F_-(x) \le F(x) \le F_+(x)$ for all x. Let $\mathcal{B}[\boldsymbol{R}]$ be the Borel σ-algebra on $\boldsymbol{R}$. Then there exists a unique Borel measure $\mu_F : \mathcal{B}[\boldsymbol{R}] \to [0, +\infty]$ such that $\mu_F([a, b]) = F_+(b) - F_-(a)$, $\mu_F([x, b)) = F_-(b) - F_-(a)$, $\mu_f([a, b)) = F_+(b) - F_+(a)$, $\mu_F((a, b)) = F_-(b) - F_+(a)$ for all $-\infty < b < a < \infty$, and $\mu_F(\{a\}) = F_+(a) - F_-(a)$ for all $a \in \boldsymbol{R}$.*

The measure μ_F is known as the Lebesgue-Stieltjes measure μ_F of F. In the special case when $F_+(-\infty) = 0$ and $F_-(+\infty) = 1$, then μ_F is a probability measure, and $F_+(x) = \mu_F((-\infty, x])$ is known as the *cumulative distribution function* of μ_F.

7.4 Product measure

Given two sets X and Y, one can form their Cartesian product $X \times Y = \{(x, y) : x \in X, y \in Y\}$. This set is naturally equipped with the coordinate projection maps $\pi_X : X \times Y \to X$ and $\pi_Y : XY \to Y$ defined by setting $\pi_X(x, y) := x$ and $\pi_Y(x, y) := y$.

Now suppose that $(X, \mathcal{B}_X)$ and $(Y, \mathcal{B}_Y)$ are measurable spaces. Then we can still form the Cartesian production $X \times Y$ and the projection maps $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$. But now we can also form the *pullback* σ-algebras $\pi_X^*(\mathcal{B}_X) := \{\pi_X^{-1}(E) : E \in \mathcal{B}_X\} = \{E \times Y : E \in \mathcal{B}_X\}$ and $\pi_Y^*(\mathcal{B}_Y) := \{\pi_Y^{-1}(E) : E \in \mathcal{B}_Y\} = \{X \times E : E \in \mathcal{B}_Y\}$. We then define the *product σ-algebra* $\mathcal{B}_X \times \mathcal{B}_Y$ to be the σ-algebra generated by the union of these two σ-algebras: $\mathcal{B}_X \times \mathcal{B}_Y := \langle \pi_X^*(\mathcal{B}_Y) \cup \pi_Y^*(\mathcal{B}_Y) \rangle$.

Now suppose that we have two measure spaces $(X, \mathcal{B}_X, \mu_X)$ and $(Y, \mathcal{B}_Y, \mu_Y)$. It is natural to expect that we should be able to multiply the two measures $\mu_X : \mathcal{B}_X \to [0, +\infty]$ and $\mu_Y : \mathcal{B}_Y \to [0, +\infty]$ to form a product measure $\mu_X \times \mu_Y : \mathcal{B}_X \times \mathcal{B}_Y \to [0, +\infty]$, with the expectation $\mu_X \times \mu_Y(E \times F) = \mu_X(E)\mu_Y(F)$ whenever $E \in \mathcal{B}_X$ and $F \in \mathcal{B}_Y$. To construct this measure, it is convenient to make the assumption that both spaces are σ-finite.

Definition 7.4. *A measure space $(X, \mathcal{B}, \mu)$ is σ-finite if X cab be expressed as the countable union of sets of finite measure.*

$\boldsymbol{R}^d$ with Lebesgue measure is σ-finite, $\boldsymbol{R}^d$ with counting measure is not σ-finite. Most measure spaces we encounter in analysis, including all probability spaces, are σ-finite. As long as we restrict to the σ-finite case, product measures always exists and is unique:

Proposition 7.1. *(Existence and uniqueness of product measure) Let $(X, \mathcal{B}_X, \mu_X)$ and $(Y, \mathcal{B}_Y, \mu_Y)$ be σ-finite measure spaces. Then there exists a unique measure $\mu_X \times \mu_Y$ on $\mathcal{B}_X \times \mathcal{B}_Y$ that obeys $\mu_X \times \mu_Y(E \times F) = \mu_X(E)\mu_Y(F)$ whenever $E \in \mathcal{B}_X$ and $F \in \mathcal{B}_Y$.*

The product $m^d \times m^{d'}$ of the Lebesgue measures $m^d, m^{d'}$ on $(\boldsymbol{R}^d, \mathcal{L}[\boldsymbol{R}^d])$ and $(\boldsymbol{R}^d, \mathcal{L}[\boldsymbol{R}^{d'}])$ respectively will agree with Lebesgue measure $m^{d+d'}$ on the production space $\mathcal{L}[\boldsymbol{R}^d] \times L[\boldsymbol{R}^{d'}]$, which is a subalgebra of $\mathcal{L}[\boldsymbol{R}^{d+d'}]$. After taking the completion $\overline{m^d \times m^{d'}}$ of this production measure, one obtains the full Lebesgue measure $m^{d+d'}$. Now we integrate using this product measure. For this we define a *monotone class* in X as a collection $\mathcal{B}$ of subsets of X with the following two closure properties:

- If $E_1 \subset E_2 \subset \ldots$ are a countable increasing sequence of sets in $\mathcal{B}$, then $\bigcup_{n=1}^{\infty} E_n \in \mathcal{B}$.
- If $E_1 \supset E_2 \supset \ldots$ are a countable decreasing sequence of sets in $\mathcal{B}$, then $\bigcap_{n=1}^{\infty} E_n \in \mathcal{B}$.

Lemma 7.1. *(Monotone class lemma) Let $\mathcal{A}$ be a Boolean algebra on X. Then $\langle A \rangle$, the generated σ-algebra, is the smallest monotone class that contains $\mathcal{A}$.*

Theorem 7.4. *(Tonelli's theorem, incomplete version) Let $(X, \mathcal{B}_X, \mu_X)$ and $(Y, \mathcal{B}_Y, \mu_Y)$ be σ-finite measure spaces, and let $f : X \times Y \to [0, +\infty]$ be measurable with respect to $\mathcal{B}_X \times \mathcal{B}_Y$. Then:*

- *The functions $x \mapsto \int_Y f(x, y) d\mu_Y(y)$ and $y \int_X f(x, y) d\mu_X(x)$ are measurable with respect to $\mathcal{B}_X$ and $\mathcal{B}_Y$ respectively.*
- *We have* $\int_{X \times Y} f(x, y) d\mu_X \times \mu_Y(x, y) = \int_X \left(\int_Y f(x, y) d\mu_Y(y) \right) d\mu_X(x) = \int_Y \left(\int_X f(x, y) d\mu_X(x) \right) d\mu_Y(y).$

Now let $(X, \mathcal{B}_X, \mu_X)$ and $(Y, \mathcal{B}_Y, \mu_Y)$ be σ-finite measure spaces, and let $E \in \mathcal{B}_X \times \mathcal{B}_Y$ be null set with respect to $\mu_X \times \mu_Y$. Then for μ_Y-almost every $x \in X$, the set $E_x := \{y \in Y : (x, y) \in E\}$ is μ_Y-null set; and similarly, for μ_Y-almost every $y \in Y$, the set $E_y = \{x \in X : (x, y) \in E\}$ is a μ_X-null set. With this corollary, we can extend Tonelli's theorem to the completion $(X \times Y, \overline{\mathcal{B}_X \times \mathcal{B}_Y}, \overline{\mu_X \times \mu_Y})$ of the product space $(X \times Y, \mathcal{B}_X \times \mathcal{B}_Y, \mu_X \times \mu_Y)$, as constructed before. We can now extend Tonelli's theorem to this context:

Theorem 7.5. *(Tonelli's theorem, complete version) Let $(X, \mathcal{B}_X, \mu_X)$ and $(Y, \mathcal{B}_Y, \mu_Y)$ be complete σ-finite measure spaces, and let $f : X \times Y \to [0, +\infty]$ be measurable with respect to $\overline{\mathcal{B}_X \times \mathcal{B}_Y}$. Then:*

- *For μ_X-almost every $x \in X$, the function $y \mapsto f(x, y)$ is $\mathcal{B}_Y$-measurable, and in particular $\int_Y f(x, y) d\mu_Y(y)$ exists. Furthermore, the μ_X-almost everywhere defined map $\mapsto \int_Y f(x, y) d\mu_Y$ is $\mathcal{B}_Y$-measurable.*
- *For μ_Y-almost every $y \in Y$, the function $x \mapsto f(x, y)$ is $\mathcal{B}_X$-measurable, and in particular $\int_X f(x, y) d\mu_X(x)$ exists. Furthermore, the μ_Y-almost everywhere defined map $y \mapsto \int_X f(x, y) d\mu_X$ is $\mathcal{B}_Y$-measurable.*
- *We have $\int_{X \times Y} f(x, y) d\overline{\mu_X \times \mu_Y}(x, y) = \int_X \left(\int_Y f(x, y) d\mu_Y(y) \right) d\mu_X(x) = \int_X \left(\int_Y f(x, y) d\mu_Y(y) \right) d\mu_X(x)$.*

Tonelli's Theorem is for the unsigned integral, but it leads to an important analogue for the absolute integral, known as Fubini's theorem:

Theorem 7.6. *(Fubini's theorem) Let $(X, \mathcal{B}_X, \mu_X)$ and $(Y, \mathcal{B}_Y, \mu_Y)$ be complete σ-finite measure spaces, and let $f : X \times Y \to \boldsymbol{C}$ be absolutely integrable with respect to $\overline{\mathcal{B}_X \times \mathcal{B}_Y}$. Then:*

- *For μ_X-almost every $x \in X$, the function $y \mapsto f(x, y)$ is absolutely integrable with respect to μ_Y, and in particular $\int_Y f(x, y) d\mu_Y(y)$ exists. Furthermore, the μ_X-almost everywhere defined map $\mapsto \int_Y f(x, y) d\mu_Y$ is absolutely integrable with respect to μ_X.*
- *For μ_Y-almost every $y \in Y$, the function $x \mapsto f(x, y)$ is absolutely integrable with respect to μ_X, and in particular $\int_X f(x, y) d\mu_X(x)$ exists. Furthermore, the μ_Y-almost everywhere defined map $y \mapsto \int_X f(x, y) d\mu_X$ is absolutely integrable with respect to μ_Y.*
- *We have $\int_{X \times Y} f(x, y) d\overline{\mu_X \times \mu_Y}(x, y) = \int_X \left(\int_Y f(x, y) d\mu_Y(y) \right) d\mu_X(x) = \int_X \left(\int_Y f(x, y) d\mu_Y(y) \right) d\mu_X(x)$.*

Informally, Fubini's theorem allows one to always interchange the order of two integrals, as long as the integrand is absolutely integrable in the product space or its completion.

Theorem 7.7. Fubini-Tonelli theorem *Let $(X, \mathcal{B}_X, \mu_X)$ and $(Y, \mathcal{B}_Y, \mu_Y)$ be complete σ-finite measure spaces, and let $f : X \times Y \to \boldsymbol{C}$ be measurable with respect to $\overline{\mathcal{B}_X \times \mathcal{B}_Y}$. If $\int_X \left(\int_Y |f(x, y)| d\mu_Y(y) \right) d\mu_X(x) < \infty$ then f is absolutely integrable with respect to $\overline{\mathcal{B}_X \times \mathcal{B}_Y}$, and in particular the conclusions of Fubini's theorem hold. Similarly if we use the integral other way round.*

8 Related topics

8.1 The Rademacher differentiation theorem

The Fubini-Tonelli theorem is often used in extending lower-dimensional results to higher-dimensional ones. We illustrate this by extending the one-dimensional Lipschitz differentiation theorem to higher dimensions, obtaining the Radamacher differentiation theorem.

Definition 8.1. *(Lipschitz continuity) A function $f : X \to Y$ from one metric space (X, d_X) to another (Y, d_Y) is said to be Lipschitz continuous if there exists a constant $C > 0$ such that $d_Y(f(x), f(x')) \leq C d_X(x, x')$ for all $x, x' \in X$.*

Lipschitz continuous functions are uniformly continuous, and hence continuous.

Definition 8.2. *(Differentiability) Let $f : \boldsymbol{R}^d \to \boldsymbol{R}$ be a function, and let $x_0 \in \boldsymbol{R}^d$. For any $v \in \boldsymbol{R}^d$, we say that f is directionally differentiable at x_0 in the direction v if the limit $D_v f(x_0) := \lim_{h \to 0; h \in \boldsymbol{R} \setminus \{0\}} \frac{f(x_0 + hv) - f(x_0)}{h}$ exists, in which case we call $D_v f(x_0)$ the directional derivative of f at x_0 in this direction. If $v = e_i$ is one of the standard basis vectors $e_1, \dots, e_d$ or $\boldsymbol{R}^d$, we write $D_v f(x_0)$ as $\frac{\partial f}{\partial x_i}(x_0)$, and refer to this as the partial derivative of f at x_0 in the e_i direction. We say that f is totally differentiable at x_0 if there exists a vector $\nabla f(x_0) \in \boldsymbol{R}^d$ with the property that $\lim_{h \to 0; h \to \boldsymbol{R}^d \setminus \{0\}} \frac{f(x_0 + h) - f(x_0) - h . \nabla f(x_0)}{|h|} = 0$, where $v.w$ is the usual dot product on $\boldsymbol{R}^d$. We refer to $\nabla f(x_0)$ as the gradient of f at x_0.*

Total differentiability implies directional and partial differentiability, but not conversely.

Theorem 8.1. *(Rademacher differentiation theorem) Let $f : \boldsymbol{R}^d \to \boldsymbol{R}$ be Lipschitz continuous. Then f is totally differentiable at x_0 for almost every $x_0 \in \boldsymbol{R}^d$.*

8.2 Probability spaces

Definition 8.3. *(Probability space) A probability space is a measure space $(\Omega, \mathcal{F}, \boldsymbol{P})$ of total measure $1 : \boldsymbol{P}(\Omega) = 1$. The measure $\boldsymbol{P}$ is known as a probability measure.*

Ω is called the sample space, and is interpreted as the set of all possible outcomes $\omega \in \Omega$ that a random system could be in. The σ-algebra F is know as the event space, and is interpreted as the set of all possible events $E \in \mathcal{F}$ that one can measure. The measure $\boldsymbol{P}(E)$ of an event is known as the probability of that event. The various axioms of a probability space then formalize the *Kolmogorov axioms of probability.*

Many concepts in measure theory are of importance in probability theory, although with a different terminology. The notion of a property holding almost everywhere is called *almost surely.* A measurable function is now referred to as *random variable* denoted by X, and its integral on the probability space is known as expectation of that random variable denoted by $\boldsymbol{E}(X)$. Borel-Cantelli lemma now reads as follows: given any sequence $E_1, E_2, \dots$ of events such that $\sum_{n=1}^{\infty} P(E_n) < \infty$, it is almost surely true that at most finitely many of these events hold. Markov's inequality becomes the assertion that $\boldsymbol{P}(X \geq \lambda) \leq \frac{1}{\lambda}\boldsymbol{E}X$ for any non-negative random variable X and any $0 < \lambda < \infty$.

8.2.1 Uniform probability measure

Given any measure space $(X, \mathcal{B}, \mu)$ with $0 < \mu(X) < +\infty$, the space $(X, \mathcal{B}, \frac{1}{\mu(X)}\mu)$ is a probability space. For instance, if Ω is a non-empty finite set with the discrete σ-algebra 2^Ω and the counting measure $\#$, then the *normalized counting measure* $\frac{1}{\#(\Omega)}\#$ is a probability measure known as the *discrete uniform probability measure* on ω, and $(\Omega, 2^\Omega, \frac{1}{\#(\Omega)}\#)$ is a probability space. This space models the act of drawing an element of the discrete set Ω uniformly at random.

Similarly, if $\Omega \subset \boldsymbol{R}^d$ is a Lebesgue measurable set of positive finite Lebesgue measure, $0 < m(\Omega) < \infty$, then $(\Omega, \mathcal{L}[\boldsymbol{R}^d]|_\Omega, \frac{1}{m(\Omega)}m|_\Omega)$ is a probability space. The probability measure $\frac{1}{m(\Omega)}m|_\Omega$ is known as the continuous uniform probability measure on Ω. This models the act of drawing an element of the continuous set Ω uniformly at random.

8.2.2 General probability measures

If Ω is a, possibly infinite, non-empty set with the discrete σ-algebra 2^Ω, and if $(p_\omega)_{\omega \in \Omega}$ are a collection of real numbers on $[0, 1]$ with $\sum_{\omega \in \Omega} p_\omega = 1$, then the probability measure $\boldsymbol{P}$ defined by $\boldsymbol{P} := \sum_{\omega \Omega} p_\omega \delta_\omega$, or in other words $\boldsymbol{P}(E) := \sum_{\omega \in E} p_\omega$, is indeed a probability measure, and $(\Omega, 2^\Omega, \boldsymbol{P})$ is a probability space. This function $\omega \mapsto p_\omega$ is known as the *discrete probability distribution* of the state variable ω.

Similarly, if Ω is a Lebesgue measurable subset of $\boldsymbol{R}^d$ of positive and possibly infinite measure, and $f : \Omega \to [0, +\infty]$ is a Lebesgue measurable function on Ω, where we restrict the Lebesgue measure space on $\boldsymbol{R}^d$ to Ω, with $\int_\Omega f(x)dx = 1$, then $(\Omega, \mathcal{L}[\boldsymbol{R}^d]|_\Omega, \boldsymbol{P})$ is a probability space, where $\boldsymbol{P} := m_f$ is the measure $P(E) := \int_\Omega 1_E(x) f(x)dx = \int_E f(x)dx$. The function f is known as the *continuous probability density* of the state variable ω. This density is not quite unique, since one can modify it on a set of probability zero, but it is well defined up to this ambiguity.

8.3 Infinite product spaces and the Kolmogorov extension theorem

We considered the product of two sets, measurable spaces or σ-finite measure spaces. We now consider how to generalise this concept to products of more than two such spaces. The axiom of choice in set theory allows us to form a Cartesian product $X_A := \prod_{\alpha\in A} X_\alpha$ of any family $(X_\alpha)_{\alpha\in A}$ of sets indexed by another set A, which consists of the space of all tuples $x_A = (x_\alpha)_{\alpha\in A}$ indexed by A, for which $x_\alpha \in X_\alpha$ for all $\alpha \in A$. For any $\beta \in A$, we have the coordinate projection maps $\pi_\beta : X_A \to X_\beta$ defined by $\pi_\beta((x_\alpha)_{\alpha\in A}) := x_\beta$. As before, given any σ-algebra $\mathcal{B}_\beta$ on X_β, we can pull it back by π_β to create a σ-algebra $\pi_\beta^*(\mathcal{B}_\beta) := \{\pi_\beta^{-1}(E_\beta) \in \mathcal{B}_\beta\}$ on X_A, which is a σ-algebra. Informally, $\pi_\beta^*(\mathcal{B}_\beta)$ describes those sets or events that depend only on the x_β coordinate of the state $x_A = (x_\alpha)_{\alpha\in A}$, and whose dependence on x_β is $\mathcal{B}_\beta$-measurable. We can define the product σ-algebra $\prod_{\beta\in A} \mathcal{B}_\beta := \langle \bigcup_{\beta\in A} \pi_\beta^*(\mathcal{B}_\beta)\rangle$.

Now we consider the problem of constructing a measure μ_A on the product space X_A. Any such measure will induce pushforward measure $\mu_B := (\pi_B)_*\mu_A$ on X_B, thus $\mu_B(E_B) := \mu_A(\pi_B^{-1}(E_B))$ for all $E_B \in \mathcal{B}_B$. These measures obey the compatibility relation $(\pi_{C\leftarrow B})_*\mu_B = \mu_C$ whenever $C \subset B \subset A$. One can then ask whether one can reconstruct μ_A from just the projections μ_B to finite subsets B. This is possible in the important special case when the μ_B and hence μ_A are probability measures, provided one imposes an additional inner regularity hypothesis on the measures μ_B.

Definition 8.4. *(Inner regularity) A metrisable inner regular measure space $(X, \mathcal{B}, \mu, d)$ is a measure space $(X, \mathcal{B}, \mu)$ equipped with a metric d such that every compact set is measurable; and one has $\mu(E) = \sup_{K\subset E, K \text{ compact}} \mu(K)$ for all measurable E. We say that μ is inner regular if it is associated to an inner regular measure space.*

Thus for instance Lebesgue measure is inner regular, as are Dirac measures and counting measures. Indeed, most of the measures that one encounters in application will be inner regular.

Theorem 8.2. *(Kolmogorov extension theorem) Let $((X_\alpha, \mathcal{B}_\alpha), \mathcal{F}_\alpha)_{\alpha\in A}$ be a family of measurable spaces $(X_\alpha, \mathcal{B}_\alpha)$, equipped with a topology $\mathcal{F}_\alpha$. For each finite $B \subset A$, let μ_B be an inner regular probability measure on $\mathcal{B}_B := \prod_{\alpha\in B} \mathcal{B}_\alpha$ with the product topology $\mathcal{F}_B := \prod_{\alpha\in B} \mathcal{F}_\alpha$, obeying the compatibility condition $(\pi_{C\leftarrow B})_*\mu_B = \mu_C$ whenever $C \subset B \subset A$ are two nested finite subsets of A. Then there exists a unique probability measure μ_A on $\mathcal{B}_A$ with the property that $(\pi_B)_*\mu_A = \mu_B$ for all finite $B \subset A$.*

This theorem is a fundamental tool in probability theory allowing one to construct a probability space to hold a variety of random processes $(X_t)_{t\in T}$, both in discrete case and in the continuous case. It can be used to rigorously construct a process for Brownian motion known as the Wiener process.

Theorem 8.3. *(Existence of product measures) Let A be an arbitrary set. For each $\alpha \in A$, let $(X_\alpha, \mathcal{B}_\alpha, \mu_\alpha)$ be a probability space in which X_α is a locally compact, σ-compact metric space, with $\mathcal{B}_\alpha$ being its Borel σ-algebra, i.e. the σ-algebra generated by the open sets. Then there exists a unique probability measure $\mu_A = \prod_{\alpha\in A} \mu_\alpha$ on $(X_A, \mathcal{B}_A) := (\prod_{\alpha\in A} X_\alpha, \prod_{\alpha\in A} \mathcal{B}_\alpha)$ with the property that $\mu_A(\prod_{\alpha\in A} E_\alpha) = \prod_{\alpha\in A} \mu_\alpha(E_\alpha)$ whenever $E_\alpha \in \mathcal{B}_\alpha$ for each $\alpha \in A$, and one has $E_\alpha = X_\alpha$ for all but finitely many of the α.*

Let $A := \boldsymbol{N}$, and for each $\alpha \in A$, let $(X_\alpha, \mathcal{B}_\alpha, \mu_\alpha)$ be the two-element set $X_\alpha = \{0, 1\}$ with the discrete metric and thus discrete σ-algebra and the uniform probability measure μ_α. Then there exists a probability measure μ on the infinite discrete cube $X_A := \{0, 1\}^{\boldsymbol{N}}$, known as the *uniform Bernoulli measure* on this cube. The coordinate functions $\pi_\alpha : X_A \to \{0, 1\}$ can then be interpreted as a countable sequence of random variables taking values in $\{0, 1\}$. From the properties of product measure one can easily check that these random variables are uniformly distributed on $\{0, 1\}$ and are jointly independent. Informally, Bernoulli measure allows one to model an infinite number of 'coin flips'. One can replace the natural numbers here by any other index set, and have a similar construction.

Repeating the previous example, replacing $\{0, 1\}$ with unit interval $[0, 1]$, with the usual metric, the Borel σ-algebra, and the uniform probability measure gives a probability measure on the infinite continuous cube $[0, 1]^{\boldsymbol{N}}$, and the coordinate functions $\pi_\alpha : X_A \to [0, 1]$ can now be interpreted as jointly independent random variables, each having the uniform distribution on $[0, 1]$.

Repeating again the previous example, this time replacing $[0, 1]$ with $\boldsymbol{R}$, with the usual metric, and Borel σ-algebra, and the normal probability distribution $d\mu_\alpha = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$, thus $\mu_\alpha(E) = \int_E \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$ for every Borel set E, gives the probability space that supports a countable sequence of jointly independent Gaussian random variables π_α.