MATH5825
Measure, Integration and Probability

Peter Straka

September 23, 2014

Contents

Chapter 1

Introduction to Measures

1.1 Lebesgue's Problem of measure

(This section is taken from Hendrik Grundling's 2011 lecture.) Assigning a measure (length, area, volume, ...) to a subset S of $\mathbb{R}^d$ is necessary in mathematics, e.g. for any type of integral. A reasonable measure $m(S)$ of S should satisfy the following requirements:

1. If $S \subset \mathbb{R}^d$ is congruent (after shifts, rotations & reflections) to $T \subset \mathbb{R}^d$, then $m(S) = m(T)$
2. If $S = \bigcup_{i=1}^{\infty} S_i$ where $S_i \cap S_j = \varnothing$ for $i \neq j$, then $m(S) = \sum_{i=1}^{\infty} m(S)$ (countable additivity)
3. $m(I) = 1$ for $I = [0, 1]^d \subset \mathbb{R}^d$ (normedness)
4. $m(S) \geqslant 0$

However this humble attempt is still too much to ask as the following shall show. Below, we construct a set S which cannot satisfy all four requirements.

For $x, y \in [0, 1]$ define $x \sim y$ if and only if $x - y \in \mathbb{Q}$. This is an equivalence relation, so $[0, 1]$ is partitioned into (disjoint) equivalence classes. Define the "Vitali set" $S \subset [0, 1]$ by choosing one point from each equivalence class (this assumes the Axiom of Choice).

Lemma 1.1. If $r, q \in \mathbb{Q} \cap [0, 1]$, $r \neq q$ then $(S + r) \cap (S + q) = \varnothing$.

Proof. If $x \in (S + r) \cap (S + q)$ then $x = p + r = t + q$ for $p, t \in S$. Then $p - t = q - r \neq 0$ and $q - r \in \mathbb{Q}$. So $p \sim t$, $p \neq t$. Since $p, t \in S$, this violates the definition of S. $\square$

Lemma 1.2. $[0, 1] \subseteq \bigcup\{S + r | r \in \mathbb{Q} \cap [-1, 1]\} =: T \subseteq [-1, 2]$.

Proof. Let $x \in [0,1)$. Then $x \sim p$ for some $p \in S$. Thus $x - p =: r \in \mathbb{Q}$ and as $x, p \in [0,1]$ it follows that $r \in \mathbb{Q} \cap [-1,1]$. So $x \in S + r$, $r \in \mathbb{Q} \cap [-1,1]$. $\square$

Proposition 1.3. In $\mathbb{R}$, Lebesgue's problem of measure has no solution.

Proof. Let S be as above. We show a contradiction to the requirement $m(S + r) = m(S)$ for all r.

$$\begin{aligned} 1 = m([0,1]) = m(T) - m(T \setminus [0,1]) \leqslant m(T) \\ = \sum_{r \in \mathbb{Q} \cap [-1,1]} m(S+r) = \sum_{r \in \mathbb{Q} \cap [-1,1]} m(S) \end{aligned}$$

$m(S)$ can't be 0, hence the right-hand side must be $+\infty$. But this contradicts

$$\begin{aligned} \infty = m(T) = m([-1,2]) - m([-1,2) \setminus T) \leqslant m([-1,2]) \\ = m([-1,0]) + m([0,1]) + m([1,2]) - m(\{0,1\}) \leqslant 3m([0,1]) = 3 \end{aligned}$$

$\square$

In higher dimensions, things are even worse:

Theorem 1.4 (Banach-Tarski)**.** Let $S, T \subset \mathbb{R}^d$ be bounded with non-empty interiors, $d \geqslant 3$. Then there is a $k \in N$ and partitions $\{E_i, \dots, E_k\}$, $\{F_i, \dots F_k\}$ of S, T respectively such that E_i is congruent to F_i $\forall i$. (That is, S and T are equidecomposable.)

Thus we can take e.g. a unit sphere S, cut it up into finitely many pieces, and reassemble it to a set T of two unit spheres. If Lebesgue's problem of measure had a solution in $\mathbb{R}^d$, $d \geqslant 3$, then by congruence $m(S) = m(T)$ which contradicts $m(T) = 2m(S)$ and $m(S) > 0$.

A way out of this situation is to restrict attention to a smaller set of subsets, on which Lebesgue's problem of measure does have a solution.

1.2 σ-algebras

Definition 1.5. Let X be any non-empty set. A system $\mathcal{A}$ of subsets of X is called an algebra on X if it has the following properties:

1. $X \in \mathcal{A}$,
2. If $\{A, B\} \subset \mathcal{A}$ then $A \cup B \in \mathcal{A}$
3. If $A \in \mathcal{A}$ then $A^\complement \in \mathcal{A}$ (where $A^\complement = X \setminus A$).

If additionally the following condition holds, then $\mathcal{A}$ is called a σ-algebra:

4. If $\{A_1, A_2, \ldots\} \subset \mathcal{A}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

A measurable space is any pair $(X, \mathcal{A})$ where $\mathcal{A}$ is a σ-algebra on X.

Algebras are closed under finite intersections: If $\{A, B\} \subset \mathcal{A}$, then

$$(A \cap B)^{\complement} = A^{\complement} \cup B^{\complement} \in \mathcal{A}.$$

Similarly, σ-algebras are closed under countably infinite intersections.

Definition 1.6. Let X be a non-empty set. A system $\mathcal{C} \subset 2^X$ of subsets of X is called a ring if $\varnothing \in \mathcal{C}$ and $A \cup B, A \backslash B \in \mathcal{C}$ for any $A, B \in \mathcal{C}$.

Every σ-algebra is an algebra; every finite algebra is a σ-algebra.

Example 1.7. Let X be any non-empty set, and consider the following subsets of 2^X:

1. $\mathcal{A} = \{\varnothing, X\}$. This is the "coarsest", "smallest" or "trivial" σ-algebra on X (and thus also an algebra and a ring).

2. $\mathcal{A} = \{\varnothing, A, A^{\complement}, X\}$, where $A \subset X$. This is a σ-algebra, an algebra and a ring.

3. The power set 2^X, i.e. the family of all subsets of X, where X is any set (the "finest" or "largest" σ-algebra on X).

4. Let $X = \mathbb{R}$. Then

$$\mathcal{C} := \left\{ \bigcup_{k=1}^{n} (a_k, b_k] : n \in \mathbb{N}, a_k, b_k \in \mathbb{R}, a_k \leqslant b_k, b_k \leqslant a_{k+1} \right\} \subset 2^{\mathbb{R}}. \quad (1.8)$$

is a ring, but not an algebra. $\mathcal{C} \cup \mathbb{R}$ is an algebra, but not a σ-algebra. We shall later see that the σ-algebra generated by $\mathcal{C}$ is the Borel-σ-algebra on $\mathbb{R}$. □

If two algebras (or σ-algebras) $\mathcal{A}_1, \mathcal{A}_2$ are defined on the same set X, then their intersection (not necessarily their union) is again an algebra (or σ-algebra) (try a simple example). This even holds for *arbitrary* (uncountably infinite) intersections.

Lemma 1.9. Let X be any set, and let $\mathcal{E}$ be a system of subsets of X. Then there exist

1. a smallest algebra $\alpha(\mathcal{E}) \supset \mathcal{E}$
2. a smallest σ-algebra $\sigma(\mathcal{E}) \supset \mathcal{E}$

Proof. Let $\mathcal{S}_\alpha$ be the set of algebras which contain $\mathcal{E}$, and let $\mathcal{S}_\sigma$ be the set of σ-algebras which contain $\mathcal{E}$. Then

$$2^X \in \mathcal{S}_\sigma \subset \mathcal{S}_\alpha,$$

where 2^X is the power set of X (the set of all possible subsets). In particular neither $\mathcal{S}_\alpha$ nor $\mathcal{S}_\sigma$ are empty. Then

$$\alpha(\mathcal{E}) = \bigcap_{\mathcal{A}\in\mathcal{S}_\alpha} \mathcal{A} \qquad\qquad \sigma(\mathcal{E}) = \bigcap_{\mathcal{A}\in\mathcal{S}_\sigma} \mathcal{A}$$

are as required. □

The algebra and σ-algebra $\alpha(\mathcal{E})$ and $\sigma(\mathcal{E})$ are said to be generated by $\mathcal{E}$.

1.3 Measures

Definition 1.10. A countably additive function μ from a σ-algebra $\mathcal{A}$ of subsets of X into $[0, \infty]$ is called a measure. Then $(X, \mathcal{A}, \mu)$ is called a measure space.

Example 1.11.

1. For a set X with σ-algebra $\mathcal{A}$, define

$$\begin{aligned} \mu : \mathcal{A} &\to [0, \infty] \\ A &\mapsto \mu(A) = \text{ number of elements in A} \end{aligned}$$

   Such μ is called the *counting measure*.

2. For a σ-algebra $\mathcal{A}$ of X, fix $x \in X$ and define

$$\begin{aligned} \delta_x : \mathcal{A} &\to [0, \infty] \\ A &\mapsto \delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \end{aligned}$$

   Then δ_x is called the *Dirac measure* concentrated at x.

3. For a σ-algebra $\mathcal{A}$ on X let $\mu(\varnothing) = 0$ and $\mu(A) = \infty$ if $A \neq \varnothing$, $A \in \mathcal{A}$. Then μ is called the trivial measure.

The aim of this section is to construct a measure on a suitable σ-algebra of $\mathbb{R}$ which solves Lebesgue's problem of measure. Other measurable spaces which are important in probability theory are:

$(\mathbb{R}^\infty, \mathcal{B}(\mathbb{R}^\infty))$: This space is used for probabilistic models of experiments with infinitely many steps, that is, stochastic processes in discrete time steps.

$(D(\mathbb{R}^d), \mathcal{D})$: Here $D(\mathbb{R}^d)$ denotes the set of càdlàg (right-continuous with left limits) paths (mappings from $[0, \infty)$ to $\mathbb{R}^d$) with countably many jumps. This set can be equipped with a metric and hence a topology. The corresponding Borel-σ-algebra is $\mathcal{D}$. A probability measure ($\mu(D(\mathbb{R}^d)) = 1$) on this measurable space then governs the behaviour of a stochastic process in continuous time.

Theorem 1.12 (Basic properties of measures)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space. Then

1. $\mu(\varnothing) = 0$
2. $A \subset B \Rightarrow \mu(A) \leqslant \mu(B)$
3. If $A_1 \subset A_2 \subset A_3 \subset \ldots$, then
$$\lim_{n\to\infty} \mu(A_n) = \mu\left(\bigcup_{n=1}^{\infty} A_n\right)$$
4. If $A_1 \supset A_2 \supset A_3 \supset \ldots$ where $\mu(A_1) < \infty$, then
$$\lim_{n\to\infty} \mu(A_n) = \mu\left(\bigcap_{n=1}^{\infty} A_n\right)$$

Outer Measure

Definition 1.13. Let X be a non-empty set. Let $\mathcal{A} \subset 2^X$, $\varnothing \in \mathcal{A}$, and for any $E \in 2^X$, there exists a covering $\{A_n\}_{n=1}^{\infty} \subset \mathcal{A}$ so that $E \subset \bigcup_{n=1}^{\infty} A_n$. Let $\mu : \mathcal{A} \to [0, \infty]$ be any map such that $\mu(\varnothing) = 0$. Then

$$\begin{aligned} \mu^* : 2^X &\to [0, \infty] \\ E &\mapsto \inf\left\{\sum_{n=1}^{\infty} \mu(A_n) | A_n \in \mathcal{A}, E \subset \bigcup_{n=1}^{\infty} A_n\right\} \end{aligned}$$

is called the outer measure defined by μ and $\mathcal{A}$. (A μ as above is called a *pre-measure*.)

If $A \in 2^X$ satisfies

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \backslash A) \quad \forall B \in 2^X,$$

then A is called μ^*-measurable.

Lemma 1.14 (Subadditivity of μ^*)**.** For any sets E and $E_n \subset X$, if $E \subset \bigcup_n E_n$, then $\mu^*(E) \leqslant \sum_n \mu^*(E_n)$.

Proof. WLOG assume the right-hand side is not $+\infty$. Let $\varepsilon > 0$, and suppose $A_{nm} \in \mathcal{A}$ are such that $\bigcup_m A_{nm} \supset E_n$, and

$$\sum_m \mu(A_{nm}) < \mu^*(E_n) + \frac{\varepsilon}{2^n}.$$

Then $E \subset \bigcup_n \bigcup_m A_{nm}$, and

$$\mu^*(E) \leqslant \sum_n \sum_m \mu(A_{nm}) < \sum_n \left(\mu^*(E_n) + \frac{\varepsilon}{2^n}\right) = \sum_n \mu^*(E_n) + \varepsilon$$

Since the above holds for arbitrarily small ε, the statement follows. □

Theorem 1.15 (Carathéodory)**.** The μ^*-measurable sets $\mathcal{M}(\mu^*)$ form a σ-algebra on X, and μ^* is a measure on $\mathcal{M}(\mu^*)$.

Proof. Check that $F \in \mathcal{M}(\mu^*)$ iff $F^\complement \in \mathcal{M}(\mu^*)$. Suppose that $A, B \in \mathcal{M}(\mu^*)$. For any $E \in 2^X$, we have

$$\begin{aligned}
\mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \backslash A) \\
&= \mu^*(E \cap A \cap B) + \mu^*(E \cap A \backslash B) + \mu^*(E \backslash A) \\
&= \mu^*(E \cap (A \cap B)) + \mu^*(E \backslash (A \cap B)),
\end{aligned}$$

using that $A \in \mathcal{M}(\mu^*)$ in equality 1 & 3 and that $B \in \mathcal{M}(\mu^*)$ in equality 2. Thus $\mathcal{M}(\mu^*)$ is an algebra (right?). Now let $E_n \in \mathcal{M}(\mu^*)$ for $n = 1, 2, \ldots$, $F := \bigcup_{j=1}^\infty E_j$, and $F_n := \bigcup_{j=1}^n E_j \in \mathcal{M}(\mu^*)$. Since $E_n \backslash \bigcup_{j<n} E_j \in \mathcal{M}(\mu^*)$ for all n, we may assume E_n disjoint in proving F measurable. For any $E \subset X$ we have

$$\begin{aligned}
\mu^*(E) &= \mu^*(E \backslash F_n) + \mu^*(E \cap F_n) \\
&= \mu^*(E \backslash F_n) + \mu^*(E \cap F_n \cap E_n) + \mu^*(E \cap F_n \backslash E_n) \\
&= \mu^*(E \backslash F_n) + \mu^*(E \cap E_n) + \mu^*(E \cap F_{n-1})
\end{aligned}$$

by μ^*-measurability of F_n and E_n. We have shown

$$\mu^*(E \cap F_n) = \mu^*(E \cap E_n) + \mu^*(E \cap F_{n-1})$$

which inductively means $\mu^*(E \cap F_n) = \sum_{j=1}^n \mu^*(E \cap E_j)$. Thus

$$\mu^*(E) = \mu^*(E \backslash F_n) + \sum_{j=1}^n \mu^*(E \cap E_j) \geqslant \mu^*(E \backslash F) + \sum_{j=1}^n \mu^*(E \cap E_j).$$

As this holds for every n, we have

$$\mu^*(E) \geqslant \mu^*(E \backslash F) + \sum_{j=1}^{\infty} \mu^*(E \cap E_j) \geqslant \mu^*(E \backslash F) + \mu^*(E \cap F) \tag{1.16}$$

using Lemma 1.14. Again using Lemma 1.14, we have $\mu^*(E) = \mu^*(E \backslash F) + \mu^*(E \cap F)$, and thus $F \in \mathcal{M}(\mu^*)$. Hence $\mathcal{M}(\mu^*)$ is a σ-algebra. Setting $E = F$ in (1.16) shows $\mu^*(F) = \sum_{j=1}^{n} \mu^*(F \cap E_j)$ and hence that μ^* is a measure on F. □

1.4 Lebesgue Measure

Consider the ring of subsets

$$\mathcal{C} = \{(a, b] : a, b \in \mathbb{R}, a \leqslant b\} \tag{1.17}$$

and the pre-measure $\gamma : \mathcal{C} \to [0, \infty)$ given by

$$\gamma((a, b]) = b - a. \tag{1.18}$$

Then $\mathcal{C}$ and γ define an outer measure λ^* on $X = \mathbb{R}$. This outer measure is called the Lebesgue outer measure. The Lebesgue measure can then be defined as the measure λ on the measurable space $(\mathbb{R}, \mathcal{M}(\lambda^*))$, which is constructed as in Theorem 1.15. The elements of $\mathcal{M}(\lambda^*)$ are called the Lebesgue-measurable sets.

Below we define the Borel σ-algebra $\mathcal{B}(\mathbb{R})$, the "standard" σ-algebra on $\mathbb{R}$. Its elements are the Borel-measurable sets, or Borel sets, or simply "measurable" sets. It is possible to develop an intuition for the Borel-σ-algebra, since it is generated by the open intervals: We can construct Borel sets through countable unions and intersections of open and closed intervals. As it turns out (in this section and the next), $\mathcal{M}(\lambda^*)$ is almost equal to $\mathcal{B}(\mathbb{R})$.

Definition 1.19 (Topology)**.** Let X be a non-empty set, and let $\mathcal{T} \subset 2^X$ be a system of subsets of X with the following properties:

1. $\varnothing \in \mathcal{T}$ and $X \in \mathcal{T}$.
2. If $A, B \in \mathcal{T}$, then $A \cap B \in \mathcal{T}$.
3. If $A_i \in \mathcal{T}$ for all $i \in I$ (I is not necessarily countable), then $\bigcup_{i \in I} A_i \in \mathcal{T}$

Then $\mathcal{T}$ is called a topology on X. The pair $(X, \mathcal{T})$ is called a topological space, and the elements of $\mathcal{T}$ are called the open sets in X. A set $A \subset X$ is called closed if $A^\complement$ is open. The σ-algebra $\sigma(\mathcal{T})$ is called the Borel σ-field on X. The elements of $\sigma(\mathcal{T})$ are called Borel sets.

Example 1.20. Let $X = \mathbb{R}$, and let $\mathcal{T}$ be the system of countably infinite unions of intervals $\{(a, b) : a, b \in \mathbb{R} \cup \{-\infty, +\infty\}\}$. Then $(\mathbb{R}, \mathcal{T})$ is a topological space, and $\sigma(\mathcal{T})$ is the Borel σ-field $\mathcal{B}(\mathbb{R})$ on $\mathbb{R}$. The Borel σ-algebra is generated by any of the following systems:

$$\mathcal{E}_1 = \{(a, b) : a, b \in \mathbb{R}\}, \quad \mathcal{E}_2 = \{[a, b] : a, b \in \mathbb{R}\},$$
$$\mathcal{E}_3 = \{(a, b] : a, b \in \mathbb{R}\}, \quad \mathcal{E}_4 = \{[a, b) : a, b \in \mathbb{R}\}$$

Theorem 1.21. Borel sets are Lebesgue-measurable: $\mathcal{B}(\mathbb{R}) \subset \mathcal{M}(\lambda^*)$.

Proof. Since we know that $\mathcal{M}(\mu^*)$ is a σ-algebra, it suffices to show that it contains a generator of $\mathcal{B}(\mathbb{R})$, such as the ring $\mathcal{C}$ of half-open intervals $(a, b]$ (1.17). That is, we show

$$\mu^*(B) \geqslant \mu^*(B \cap I) + \mu^*(B \backslash I)$$

for any interval $I = (a, b] \in \mathcal{C}$ and any subset $B \in 2^X$.

Let $\{I_k\}_{k \in \mathbb{N}} \subset \mathcal{C}$ be a covering of B such that $\sum_k \gamma(I_k) \leqslant \mu^*(B) + \varepsilon$, where γ is the pre-measure (1.18) and ε is any positive small number. For each k, let $\{E_{k\ell}\}_{\ell \in \mathbb{N}} \subset \mathcal{C}$ and $\{F_{k\ell}\}_{\ell \in \mathbb{N}} \subset \mathcal{C}$ be a covering of $B \cap I$ and $B \backslash I$, respectively, such that

$$\sum_{\ell} \gamma(E_{k\ell}) \leqslant \mu^*(B \cap I_k) + 2^{-k}\varepsilon \qquad \sum_{\ell} \gamma(F_{k\ell}) \leqslant \mu^*(B \backslash I_k) + 2^{-k}\varepsilon.$$

Then

$$\begin{aligned}
\mu^*(B \cap I) + \mu^*(B \backslash I) &\leqslant \sum_k \sum_{\ell} [\gamma(E_{k\ell}) + \gamma(F_{k\ell})] \\
&\leqslant \sum_k \left[\gamma(I_k \cap I) + 2^{-k}\varepsilon + \gamma(I_k \backslash I) + 2^{-k}\varepsilon\right] \\
&= \sum_k \left[\gamma(I_k) + 2^{-k+1}\right] \leqslant \mu^*(B) + 2\varepsilon.
\end{aligned}$$

and as the above inequality holds for arbitrary $\varepsilon > 0$, the result follows. □

In the above proof, the decisive step was the equality $\gamma(I_k) = \gamma(I_k \cap I) + \gamma(I_k \backslash I)$, which relies on the assumption that $\mathcal{C}$ is a ring; the remaining steps work for any system $\mathcal{C}$ and any pre-measure γ.

The Lebesgue measure $\lambda(A)$ is defined for all Lebesgue sets $A \in \mathcal{M}(\mu^*)$; by restricting the domain of λ from $\mathcal{M}(\mu^*)$ to $\mathcal{B}(\mathbb{R})$, we define the Lebesgue measure for the Borel sets. This is a different measure (since the domains are different), but we still denote it by λ.

Uniqueness of Lebesgue Measure

To uniquely determine a measure μ on a measurable space $(X, \mathcal{A})$, it is not necessary to know all values $\{\mu(A) : A \in \mathcal{A}\}$. As it turns out, it suffices to determine $\mu(A)$ only for A from a much smaller subset of $\mathcal{A}$. This result is proved using the *monotone class theorem* due to Dynkin. It is a powerful tool for the characterisation of measures, and we will use it to prove the uniqueness of Lebesgue measure λ.

Definition 1.22. Let X be a non-empty set. A d-class on X is a system of subsets of X, $\mathcal{D} \subset 2^X$, which satisfies the following:

1. $X \in \mathcal{D}$
2. If $A, B \in \mathcal{D}$ such that $B \subset A$, then $A \backslash B \in \mathcal{D}$
3. For any increasing sequence $A_1 \subset A_2 \subset \ldots$ where $A_j \in \mathcal{D}$, one has $\bigcup_{j=1}^{\infty} A_j \in \mathcal{D}$

A π-class on X is a system of subsets of X, $\pi \subset 2^X$, which is closed with respect to finite intersections. That is, it satisfies $A, B \in \pi \Rightarrow A \cap B \in \pi$.

Any σ-algebra is a d-class. Similarly to algebras and σ-algebras, the intersection of arbitrarily many d-classs on X is again a d-class. Hence we can say that a d-class $\mathcal{D} = d(\mathcal{C})$ is generated by $\mathcal{C} \subset 2^X$ if $\mathcal{D}$ is the smallest d-class containing $\mathcal{C}$.

Theorem 1.23 (Monotone Class Theorem, Dynkin)**.** Let X be a non-empty set, and let $\mathcal{C}$ be a π-class on X. Then $\sigma(\mathcal{C}) = d(\mathcal{C})$.

Proof. Since $\sigma(\mathcal{C})$ is a d-class, we have $\sigma(\mathcal{C}) \supset d(\mathcal{C})$. It remains to show $\sigma(\mathcal{C}) \subset d(\mathcal{C})$, which follows if we can show that $d(\mathcal{C})$ is a σ-algebra. We check that $d(\mathcal{C})$ indeed satisfies the first two properties required of a σ-algebra. For the last property, let $E_1, E_2, \ldots$ be a sequence (not necessarily increasing) of sets in $d(\mathcal{C})$. Since

$$\bigcup_{j=1}^{\infty} E_j = \bigcup_{k=1}^{\infty} F_k$$

where $F_k = \bigcup_{j=1}^{k} E_j$, we can write the countable union of E_j as an *increasing* countable union of F_j. Since $d(\mathcal{C})$ is a d-class, $\bigcup_{k=1}^{\infty} F_k$ will lie in $d(\mathcal{C})$; but we need to check that $F_k \in d(\mathcal{C})$. Since $F_k^{\complement} = \bigcap_{j=1}^{k} E_j^{\complement}$, it hence suffices to show that $d(\mathcal{C})$ is closed under *finite* intersections. First, define

$$\mathcal{D}_1 := \{E \in d(\mathcal{C}) : E \cap C \in d(\mathcal{C}) \, \forall C \in \mathcal{C}\}$$

By definition of $\mathcal{C}$, we have $\mathcal{C} \subset \mathcal{D}_1$. The identities

$$(A \backslash B) \cap C = (A \cap C) \backslash (B \cap C)$$
$$\left(\bigcup_{n=1}^{\infty} A_n \right) \cap C = \bigcup_{n=1}^{\infty} (A_n \cap C)$$

show that $\mathcal{D}_1$ is closed with respect to proper set differences and countable unions of increasing sets, that is, $\mathcal{D}_1$ is a d-class containing $\mathcal{C}$ and $\mathcal{D}_1 \supset d(\mathcal{C})$. By definition, $\mathcal{D}_1 \subset d(\mathcal{C})$, and so $\mathcal{D}_1 = d(\mathcal{C})$. Next, define

$$\mathcal{D}_2 := \{E \in d(\mathcal{C}) : E \cap F \in d(\mathcal{C}) \, \forall F \in d(\mathcal{C})\}$$

We see that $\mathcal{C} \subset \mathcal{D}_2$ and $X \in \mathcal{D}_2$. By the same arguments as above, $\mathcal{D}_2 = d(\mathcal{C})$. This shows that $d(\mathcal{C})$ is closed with respect to finite intersections. □

Definition 1.24. Let $(X, \mathcal{A}, \mu)$ be a measure space. If there is a sequence $\{E_n\}_{n \in \mathbb{N}} \subset \mathcal{A}$ such that $X = \bigcup_n E_n$ and $\mu(E_n) < \infty$ for all n, then the measure μ is said to be σ-finite.

For instance, Lebesgue measure is σ-finite on the measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$; See this by setting $E_n = (-n, n]$. Moreover, any finite measure space with $\mu(X) < \infty$ is σ-finite.

Theorem 1.25. Let $(X, \mathcal{A})$ be a measurable space, and let $\mathcal{C}$ be a π-system on X such that $\mathcal{A} = \sigma(\mathcal{C})$. If μ and ν are measures on $(X, \mathcal{A})$ that agree on $\mathcal{C}$ and if there is an increasing sequence $\{C_n\}$ of sets that belong to $\mathcal{C}$, have finite measure under μ and ν, and satisfy $\bigcup_n C_n = X$, then $\mu = \nu$.

Proof. First, let's assume that μ and ν are finite measures, that is, $\mu(X) < \infty$ and $\nu(X) < \infty$. Define $\mathcal{D} = \{A \in \mathcal{A} : \mu(A) = \nu(A)\}$. We see $\mathcal{C} \subset \mathcal{D}$. Then $\mathcal{D}$ is a d-class because:

1. $\mu(X) = \nu(X) \Rightarrow X \in \mathcal{D}$

2. Let $A, B \in \mathcal{D}$ such that $A \subset B$. Then $\mu(A \backslash B) = \mu(A) - \mu(B) = \nu(A) - \nu(B) = \nu(A \backslash B)$, and thus $A \backslash B \in \mathcal{D}$. (This step assumes that the measures are finite, since $\infty - \infty$ is to be avoided.)

3. Let $E_n \in \mathcal{D}$ be an increasing sequence in $\mathcal{D}$. By Theorem 1.12,

$$\mu \left(\bigcup_{n=1}^{\infty} E_n \right) = \lim_{n \to \infty} \mu(E_n) = \lim_{n \to \infty} \nu(E_n) = \nu \left(\bigcup_{n=1}^{\infty} E_n \right)$$

and hence $\bigcup_{n=1}^{\infty} E_n \in \mathcal{D}$.

By Theorem 1.23, $\mathcal{A} = \mathcal{D}$, thus $\mu = \nu$.

Now assume that μ and ν are σ-finite with respect to the sequences $\{E_n\}_{n\in\mathbb{N}}$ and $\{F_n\}_{n\in\mathbb{N}}$; then both are σ-finite with respect to the the sequence $\{C_n\}_{n\in\mathbb{N}}$ where $C_n = E_n \cap F_n$. For each n, define measures μ_n and ν_n on $(X, \mathcal{A})$ by $\mu_n(A) = \mu(A \cap C_n)$ and $\nu_n(A) = \nu(A \cap C_n)$. From the first part of the proof, we have $\mu_n = \nu_n$, and thus using Th 1.12

$$\mu(A) = \lim_n \mu_n(A) = \lim_n \nu_n(A) = \nu(A)$$

for every $A \in \mathcal{A}$. $\square$

Theorem 1.26. The Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is the unique measure which satisfies $\lambda((a, b]) = b - a$ for all $a, b \in \mathbb{R}$, $a \leqslant b$.

Proof. First, we show $\lambda((a, b]) = b - a$. Indeed, consider an interval $I = (a, b]$. It is covered by the sequence $I \cup \varnothing \cup \varnothing \cup \ldots$ of elements in the ring $\mathcal{C}$, which shows $\lambda^*(I) \leqslant \gamma(I) = b - a$. It is intuitively clear (but technically requires a proof) that this covering is optimal. That is: *any* countable covering of I with sets $\{C_n\}_{n\in\mathbb{N}} \subset \mathcal{C}$ is such that $\sum_n \gamma(C_n) \geqslant b - a$. But this means $\lambda^*(I) \geqslant b - a$, and so $\lambda^*(I) = b - a$. I is a Borel set, and thus also a Lebesgue set, hence $\lambda(I) = \lambda^*(I) = b - a$. We know that λ is σ-finite; any other measure which maps $(a, b]$ to $b - a$ is also σ-finite. Th 1.25 completes the proof. $\square$

Since the pre-measure γ and thus the outer Lebesgue-measure λ^* are translation invariant, it follows that Lebesgue measure is translation invariant. In fact, up to a multiplicative constant, Lebesgue measure is the only translation invariant measure on $\mathbb{R}$:

Theorem 1.27. Let ν be a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ which is finite on bounded sets and translation invariant (that is, $\nu(A + x) = \nu(A)\,\forall A \in \mathcal{B}(\mathbb{R}), x \in \mathbb{R}$). Then ν is a multiple of the Lebesgue measure.

Proof. The interval $I = (0, 1]$ is the pairwise disjoint union of n intervals $((k - 1)/n, k/n]$, $k = 1, \ldots, n$. These are all translates of each other and hence all have the same measure as any fixed one, say B. Set $c := \nu(I)$, then

$$n\nu(B) = c = c \times 1 = c\lambda(I) = c\, n\, \lambda(B),$$

and $\nu(B) = c\,\lambda(B)$. Any $C \in \mathcal{C}$ can be represented as a countable union of intervals of length $1/n$ ($n \in \mathbb{N}$). The family of such intervals hence generates $\mathcal{B}(\mathbb{R})$. Hence by Th 1.25, $\nu = c\lambda$. $\square$

Finally, here are a few properties of Lebesgue measure:

Theorem 1.28. Lebesgue measure on $\mathbb{R}$ has the following properties:

1. If $A \in 2^X$ is countable, then it is a Borel set, and $\lambda(A) = 0$
2. If $A \in \mathcal{M}(\lambda^*)$ is bounded, then $\lambda(A) < \infty$
3. For every $A \in \mathcal{M}(\lambda^*)$ we have $\lambda(A) = \inf\{\lambda(U) : U \text{ open }, U \supset A\}$
4. For every $A \in \mathcal{M}(\lambda^*)$ we have $\lambda(A) = \sup\{\lambda(K) : K \text{ compact }, K \subset A\}$. (Recall that a set $K \in 2^X$ is compact iff it is both closed and bounded.)

Proof. Properties 1. and 2. are left as an exercise. For 3. and 4, note that monotonicity of measures ($A \subset B \Rightarrow \lambda(A) \leqslant \lambda(B)$) implies

$$\lambda(A) \leqslant \inf\{\lambda(U) : U \text{ open }, U \supset A\},$$
$$\lambda(A) \geqslant \sup\{\lambda(K) : K \text{ compact }, K \subset A\}$$

and it remains to show the reverse inequalities. For part 3, we may assume $\lambda(A) < \infty$. Let $\varepsilon > 0$, and find a covering $\bigcup C_k$ of A where $C_k \in \mathcal{C}$ are half open intervals, such that

$$\sum_k \gamma(C_k) < \lambda(A) + \varepsilon.$$

Replace $C_k = (a_k, b_k]$ by $R_k = (a_k, b_k + 2^{-k}\varepsilon]$ to get a covering such that

$$\sum_k \gamma(R_k) < \lambda(A) + 2\varepsilon.$$

The open set $U = \bigcup_k (a_k, b_k + 2^{-k}\varepsilon)$ contains A, and $\lambda(U) \leqslant \lambda(A) + 2\varepsilon$. As ε was arbitrary, 3. is proved.

For part 4., assume first that A is bounded. Let C be a closed and bounded set that includes A, and let ε be an arbitrary positive number. Use part 3. to choose an open set U that includes $C \backslash A$ and satisfies

$$\lambda(U) < \lambda(C \backslash A) + \varepsilon.$$

Let $K = C \backslash U$. (Draw a picture!) Then K is a closed and bounded (and hence compact) subset of A; it satisfies $C \subset K \cup U$, and so satisfies

$$\lambda(C) \leqslant \lambda(K) + \lambda(U)$$

The two inequalities (together with $\lambda(C \backslash A) = \lambda(C) - \lambda(A)$ now imply that $\lambda(A) - \varepsilon < \lambda(K)$. Since ε was arbitrary, part 4 is proved in the case where A is bounded. Finally, consider the case where A is not bounded. Suppose that

b is a real number that satisfies $b < \lambda(A)$; we shall produce a compact subset K of A such that $b < \lambda(K)$. Let $\{A_j\}$ be an increasing sequence of bounded measurable subsets of A such that $A = \bigcup_j A_j$ (for example, A_j could be defined to be $A \cap [-j, +j]$. By Th 1.12, $\lambda(A) = \lim_j \lambda(A_j)$, and so we can choose j_0 so that $\lambda(A_{j_0}) > b$. Now apply to A_{j_0} what was shown above for bounded A, obtaining a compact subset of A_{j_0} such that $\lambda(K) > b$. Since $K \subset A$ and since b was an arbitrary number less than $\lambda(A)$, the proof is complete. □

1.5 Completion of Measures

In this section, we establish a result which shows that Lebesgue sets can be very well approximated by Borel sets: For every Lebesgue set, there is a Borel subset and a Borel superset of the same measure. We show that the Lebesgue σ-algebra results from the Borel σ-algebra by completion. Completed probability spaces are of interest in probability theory.

Definition 1.29. Let $(X, \mathcal{A}, \mu)$ be a measure space.

1. A subset $B \in 2^X$ is called μ-null if there is $A \in \mathcal{A}$ such that $B \subset A$ and $\mu(A) = 0$.

2. If $\mathcal{A}$ contains all μ-null sets, then μ and $(X, \mathcal{A}, \mu)$ are called complete.

3. A property holds μ-almost everywhere if the set of $x \in X$ for which it doesn't hold is a μ-null set.

A measure space $(X, \mathcal{A}, \mu)$ which is not complete can be completed, in a unique well-defined way. First, define the completion of $\mathcal{A}$ under μ as the system $\mathcal{A}_\mu$ of all subsets $A \subset X$ for which there exist $E, F \in \mathcal{A}$ such that $E \subset A \subset F$ and $\mu(F \backslash E) = 0$. Members of $\mathcal{A}_\mu$ are sometimes called μ-measurable.

To define a candidate for a measure $\bar{\mu}$ on $\mathcal{A}_\mu$, let $A \in \mathcal{A}_\mu$, where $E, F \in \mathcal{A}$ are such that $E \subset A \subset F$, and set $\bar{\mu}(A) := \mu(E)$. This does not depend on the choice of E of F, hence $\bar{\mu}$ is well defined.

Proposition 1.30. Let $(X, \mathcal{A}, \mu)$ be a measure space. Then $\mathcal{A}_\mu$ is a σ-algebra on X, and $\bar{\mu}$ is a measure on $(X, \mathcal{A}_\mu)$. The restriction of $\bar{\mu}$ from $\mathcal{A}_\mu$ to $\mathcal{A}$ is μ.

Proof. This is left as an exercise. □

Proposition 1.31. Suppose $(X, \mathcal{M}(\mu^*), \mu)$ is the measure space obtained from an outer measure by the Carathéodory construction (Th 1.15). That is, μ is the restriction of μ^* from 2^X to $\mathcal{M}(\mu^*)$. Then the measure μ is complete.

Proof. If $B \in 2^X$ is μ-null, then $B \subset A$ for some $A \in \mathcal{M}(\mu^*)$ with $\mu^*(A) = 0$. Outer measures are monotone, thus $\mu^*(B) = 0$. For arbitrary $C \in 2^X$,

$$\mu^*(C) \leqslant \mu^*(C \cap B) + \mu^*(C \backslash B) \leqslant \mu^*(C \backslash B) \leqslant \mu^*(C);$$

the first inequality follows from subadditivity of outer measures, the second from $\mu^*(C \cap B) \leqslant \mu^*(B) = 0$ and the third from $C \backslash B \subset C$. It follows that $B \in \mathcal{M}(\mu^*)$. □

Lemma 1.32. Let A be a Lebesgue set in $\mathbb{R}$. Then there exist Borel subsets E and F such that $E \subset A \subset F$ and $\lambda(F \backslash E) = 0$.

Proof. First suppose that A is a Lebesgue set such that $\lambda(A) < \infty$. For each positive integer n use Th 1.28 to choose a compact set K_n that satisfies $K_n \subset A$ and $\lambda(A) - 1/n < \lambda(K_n)$ and an open set U_n that satisfies $A \subset U_n$ and $\lambda(U_n) \leqslant \lambda(A) + 1/n$. Let $E = \bigcup_n K_n$ and $F = \bigcap_n U_n$. Then E and F belong to $\mathcal{B}(\mathbb{R})$ and satisfy $E \subset A \subset F$. The relation

$$\lambda(F \backslash E) \leqslant \lambda(U_n \backslash K_n) = \lambda(U_n \backslash A) + \lambda(A \backslash K_n) \leqslant 2/n$$

holds for each n, and so $\lambda(F \backslash E) = 0$. Thus the lemma is proved in the case where $\lambda(A) < \infty$. If A is an arbitrary Lebesgue set, then A is the union of a sequence $\{A_n\}$ of Lebesgue measurable sets each of which satisfies $\lambda(A_n) < \infty$. For each n we can choose Borel sets E_n and F_n such that $E_n \subset A_n \subset F_n$ and $\lambda(F_n \backslash E_n) = 0$. The sets E and F defined by $E = \bigcup_n E_n$ and $F = \bigcup_n F_n$ then satisfy $E \subset A \subset F$ and $\lambda(F \backslash E) = 0$ (note that $F \backslash E \subset \bigcup_n (F_n \backslash E_n)$). □

Theorem 1.33. The measure space $(\mathbb{R}, \mathcal{M}(\lambda^*), \lambda)$ is the completion of the measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$.

Proof. Let us write λ for the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $\bar{\lambda}$ for its completion. We want to show that $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \bar{\lambda}) = (\mathbb{R}, \mathcal{M}(\lambda^*), \lambda_m)$, where we have written λ_m for the Lebesgue measure on $(\mathbb{R}, \mathcal{M}(\lambda^*))$ to be able to discern it. Lemma 1.32 says that if $A \in \mathcal{M}(\lambda^*)$, then $A \in \mathcal{B}(\mathbb{R})_\lambda$, hence $\mathcal{M}(\lambda^*) \subset \mathcal{B}(\mathbb{R})_\lambda$. Moreover, the restriction of $\bar{\lambda}$ from $\mathcal{B}(\mathbb{R})_\lambda$ to $\mathcal{M}(\lambda^*)$ equals λ_m. To see this, let $E \subset A \subset F$ for some Borel sets E and F with $\lambda(F \backslash E)$; then by definition of the completion $\bar{\lambda}(A) = \lambda(E)$. Since Borel sets are Lebesgue sets, $\lambda(E) = \lambda_m(E)$, and $\lambda_m(A) = \lambda_m(E) + \lambda_m(A \backslash E) = \lambda_m(E)$, that is $\bar{\lambda}(A) = \lambda_m(A)$ as required.

It remains to check the inclusion $\mathcal{B}(\mathbb{R})_\lambda \subset \mathcal{M}(\lambda^*)$; since we already know that $\bar{\lambda}$ and λ_m coincide on $\mathcal{M}(\lambda^*)$, the theorem will follow. Let $A \in \mathcal{B}(\mathbb{R})_\lambda$. Then there are Borel sets E and F such that $E \subset A \subset F$ and $\lambda(F \backslash E) = 0$. Since $A \backslash E \subset F \backslash E$ and $\lambda_m(F \backslash E) = \lambda(F \backslash E) = 0$, the completeness of Lebesgue measure on $\mathcal{M}(\lambda^*)$ (Prop 1.31) implies that $A \backslash E \in \mathcal{M}(\lambda^*)$. Thus $A = A \backslash E \cup E \in \mathcal{M}(\lambda^*)$. □

Remark 1.34. This completes our discussion of Lebesgue measure. We have only constructed and studied it on $\mathbb{R}^d$ for $d = 1$, to keep notation simple. The entire theory above remains correct if the ring $\mathcal{C}$ of half-open intervals $(a, b]$ in $\mathbb{R}$ is replaced by the ring of "blocks" $(a_1, b_1] \times (a_2, b_2] \times \ldots \times (a_d, b_d]$ in $\mathbb{R}^d$, and the function $\gamma((a, b]) = b - a$ is replaced by $\gamma((a_1, b_1] \times (a_2, b_2] \times \ldots \times (a_d, b_d]) = \prod_{k=1}^{d}(b_k - a_k)$. This results in Lebesgue measure on $\mathbb{R}^d$, Lebesgue sets $\mathcal{M}(\lambda^*)$ in $\mathbb{R}^d$ and Borel sets $\mathcal{B}(\mathbb{R}^d)$ in $\mathbb{R}^d$. The same results regarding completeness remain true.

Chapter 2

Lebesgue Integration

Starting from a measure space $(X, \mathcal{A}, \mu)$ we want to define integrals by generalising Riemann sums. The key point will be that for a function $f : X \to \mathbb{R}$ we split up the *range* space into intervals $\left(\frac{j-1}{k}, \frac{j}{k}\right]$ and then approximate an integral by sums: $\int f \, d\mu \approx \sum_{j=1}^{\infty} \left(\frac{j-1}{k}\right) \mu(E_j)$ where

$$E_j = f^{-1}\left(\left[\frac{j-1}{k}, \frac{j}{k}\right)\right) = \left\{x \in X \;\middle|\; f(x) \in \left[\frac{j-1}{k}, \frac{j}{k}\right)\right\}$$

For a *measurable function*, E_j will lie in $\mathcal{A}$.

2.1 Measurable functions

Definition 2.1. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be two measurable spaces. A map $f : X \to Y$ is called $\mathcal{A}/\mathcal{B}$-measurable, or simply measurable, if

$$\forall\, B \in \mathcal{B} : \quad f^{-1}(B) \in \mathcal{A}.$$

If f is real-valued, it is implicit that measurability statements are with respect to the Borel σ-algebra, unless otherwise specified, and f is then called $\mathcal{A}$-measurable. Recall that for a function $f : X \to Y$, the pre-image is a function

$$\begin{aligned} f^{-1} &: 2^Y \to 2^X \\ &B \mapsto f^{-1}(B) = \{x \in X : f(x) \in B\} \end{aligned}$$

One can verify that compositions of measurable maps are measurable.

Example 2.2. Let f map from $(\mathbb{R}, \mathcal{B}(\mathbb{R})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R})$ as follows:

$$f(x) = \begin{cases} 1 & \text{if } x \in [0, 1] \\ 0 & \text{else} \end{cases}$$

Then as B ranges over all possible subsets $\in \mathcal{B}(\mathbb{R}^d)$, the pre-image $f^{-1}(B)$ is always one of the sets $\varnothing, \mathbb{R}, [0,1]$ or $[0,1]^{\complement}$. These are Borel sets, and hence f is measurable.

Theorem 2.3. Let $(X, \mathcal{F})$ and $(Y, \mathcal{G})$ be two measurable spaces, and assume that $\mathcal{G}_0 \subset 2^Y$ is such that $\sigma(\mathcal{G}_0) = \mathcal{G}$. Then $f : X \to Y$ is measurable iff

$$f^{-1}(G) \in \mathcal{F} \quad \forall G \in \mathcal{G}_0. \tag{2.4}$$

Proof. The above is a necessary condition for measurability of f; we only need to show that it is sufficient. Hence assume that (2.4) holds and define

$$\mathcal{D} = \{G \in \mathcal{G} : f^{-1}(G) \in \mathcal{F}\}.$$

For any collection $\{G_i\}_{i \in I} \subset \mathcal{G}$, we have the relations

$$f^{-1}\left(\bigcup_{i \in I} G_i\right) = \bigcup_{i \in I} f^{-1}(G_i), \qquad f^{-1}\left(\bigcap_{i \in I} G_i\right) = \bigcap_{i \in I} f^{-1}(G_i),$$
$$\left(f^{-1}(G_i)\right)^{\complement} = f^{-1}(G_i^{\complement}).$$

This shows that $\mathcal{D}$ is a σ-algebra on Y, and by definition $\mathcal{D} \subset \mathcal{G}$. If (2.4) holds, then $\mathcal{G}_0 \subset \mathcal{D}$, and then

$$\mathcal{G} = \sigma(\mathcal{G}_0) \subset \sigma(\mathcal{D}) = \mathcal{D} \subset \mathcal{G}$$

which shows $\mathcal{D} = \mathcal{G}$, that is, f is measurable. □

As an example, suppose $Y = \mathbb{R}$ above. For f to be measurable, it suffices to show (2.4) for $\mathcal{G}_0$ being all open intervals, or all intervals of the form $(-\infty, b)$, ...

Let $(X, \mathcal{T}_1)$ and $(Y, \mathcal{T}_2)$ be two *topological* spaces. The ε-δ definition of continuity of mappings in $\mathbb{R}$ is equivalent to the following more general definition of continuity:

$$f : X \to Y \text{ is continuous iff } \forall U \in \mathcal{T}_2 : \quad f^{-1}(U) \in \mathcal{T}_1,$$

that is, pre-images of open sets are open. We can hence say:

Corollary 2.5. Continuous functions are measurable.

Proposition 2.6. Let $(X, \mathcal{A}, \mu)$ be a measure space, and let f and g be real-valued $\mathcal{A}$-measurable functions.

1. Then the functions $f + g$, αf for every $\alpha \in \mathbb{R}$, fg, f/g, $f \wedge g(x) = \min(f(x), g(x))$ and $f \vee g(x) = \max(f(x), g(x))$ are measurable functions.

2. The sets $\{x : f(x) < g(x)\}$, $\{x : f(x) \leqslant g(x)\}$ and $\{x : f(x) = g(x)\}$ are measurable.

3. Finally, if f_n is a sequence of $\mathcal{A}$-measurable functions, then $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$ and $\liminf_n f_n$ are measurable functions.

Here the sup function is defined by $\sup_n f_n(x) = \sup\{f_n(x) : n \in \mathbb{N}\}$, and the inf function is defined similarly. The limsup function is defined by

$$\limsup_n f_n(x) = \limsup\{f_n(x) : n \in \mathbb{N}\} = \inf_k \sup_{n \geqslant k} f_n(x),$$

and the liminf function is defined similarly:

$$\liminf_n f_n(x) = \liminf\{f_n(x) : n \in \mathbb{N}\} = \sup_k \inf_{n \geqslant k} f_n(x).$$

Proof. We only give hints: $f(x) + g(x) < t \Leftrightarrow f(x) < r$ and $g(x) < t - r$ for some $r \in \mathbb{Q}$. Show that f^2 is measurable, then use $fg = (f+g)^2 - f^2 - g^2)/2$. For the sets, note that $f(x) < g(x)$ iff there exists a rational r such that $f(x) < r < g(x)$. Take complements for the second statement about sets, and differences for the third. For f/g, see that $\{x : g(x) > 0\}$ is measurable, and see that $f(x)/g(x) < t \Leftrightarrow f(x) < tg(x)$. For min and max, note that $f \wedge g(x) > t \Leftrightarrow f(x) > t$ and $g(x) > t$, and $f \vee g(x) < t \Leftrightarrow f(x) < t$ and $g(x) < t$. Proceed similarly for the inf and sup of countably many functions. □

Exercise 2.7. Show that any non-decreasing function $f : \mathbb{R} \to \mathbb{R}$ is (Borel-) measurable.

2.2 Integrals of simple functions

Definition 2.8. Let X be a non-empty set. Given $A \subset X$, the indicator function for A is the function χ_A defined on X by $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ if $x \notin A$. A function f defined on X is called a simple function if it is a linear combination of indicator functions: that is, there exist $n \in \mathbb{N}$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ such that $f(x) = \sum_{i=1}^{n} \alpha_i \chi_{A_i}(x)$

Different linear combinations can give the same simple function; but any simple function can be represented by a "canonical" linear combination as follows. If f is simple, then it has finite range, say $\{y_1, \ldots, y_n\}$. Then we can write

$$f(x) = \sum_{i=1}^{n} y_i \chi_{A_i}(x)$$

where $A_i = f^{-1}(\{y_i\})$. Note that the A_i are disjoint. Moreover, if X is endowed with the σ-field $\mathcal{A}$, then f is $\mathcal{A}$-measurable iff $\forall i : A_i \in \mathcal{A}$.

Definition 2.9 (Integral of a simple function)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space, and let $f = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$ be a measurable simple function on X. The Lebesgue Integral of f is defined as

$$\int f\, d\mu = \int f(x)\mu(dx) = \sum_{i=1}^{n} \alpha_i \mu(A_i)$$

whenever the above linear combination is well defined.

Recall that we have defined $0 \cdot \infty := 0$, but we have left $\infty - \infty$ undefined. This simple definition of the Lebesgue integral can be extended to a larger class (larger than Riemann-integrable) of functions, see below. But first, we note a few properties of the Lebesgue integral:

Proposition 2.10. The Lebesgue integral is linear and monotonic, that is:

1. $\int \alpha f d\mu = \alpha \int f d\mu$ for any simple function f and $\alpha \in \mathbb{R}$
2. $\int (f+g) d\mu = \int f d\mu + \int g d\mu$ for any two simple functions f and g
3. If f and g are simple functions such that $f(x) \leqslant g(x)$ holds at each $x \in X$, then $\int f d\mu \leqslant \int g d\mu$.

Proof. Suppose $f = \sum_i a_i \chi_{A_i}$ and $g = \sum_j b_j \chi_{B_j}$, in canonical representation. Then

$$\int \alpha f d\mu = \sum_i \alpha a_i \mu(A_i) = \alpha \sum_i a_i \mu(A_i) = \alpha \int f d\mu.$$

For the sum, see that

$$f + g = \sum_i \sum_j (a_i + b_j) \chi_{A_i \cap B_j}$$

and so

$$\begin{aligned}
\int (f+g) d\mu &= \sum_i \sum_j (a_i + b_j) \mu(A_i \cap B_j) \\
&= \sum_i \sum_j a_i \mu(A_i \cap B_j) + \sum_i \sum_j b_j \mu(A_i \cap B_j) \\
&= \sum_i a_i \mu(A_i) + \sum_j b_j \mu(B_j) = \int f d\mu + \int g d\mu
\end{aligned}$$

due to the additivity property of the measure μ. Finally, $f \leqslant g$, then $g - f$ is a non-negative simple function, and $\int (g - f) d\mu \geqslant 0$. But then

$$\int g d\mu = \int (f + (g - f)) d\mu = \int f d\mu + \int (g - f) d\mu \geqslant \int f d\mu.$$

□

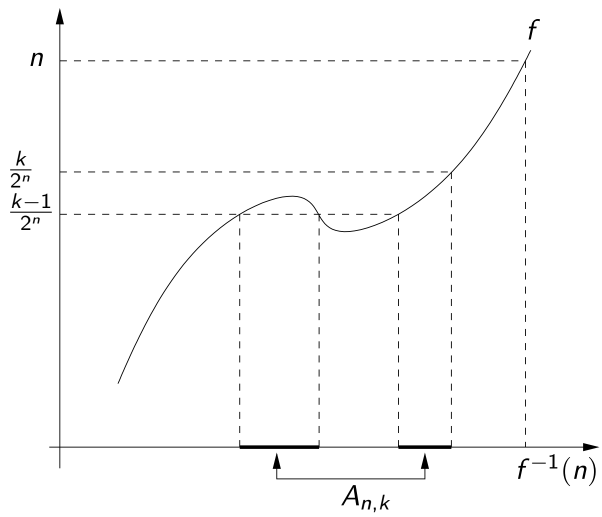

Figure 2.1: Each $A_{n,k}$ for $n \in \mathbb{N}$ and $k \in \{1, 2, \ldots, n \cdot 2^n\}$

Theorem 2.11. Let $(X, \mathcal{A}, \mu)$ be a measure space, and let f be any measurable function defined on X with values in $[0, \infty]$. Then there exists a sequence of simple functions $s_1, s_2, \ldots$ on X such that

1. $0 \leqslant s_1 \leqslant s_2 \leqslant \ldots$
2. $\forall x : \lim\limits_{n\to\infty} s_n(x) = f(x)$.

Proof. For each $n \in \mathbb{N}$ and $k \in \{1, 2, \ldots, n2^n\}$ define

$$A_{n,k} := f^{-1}\left(\left(\frac{k-1}{2^n}, \frac{k}{2^n}\right]\right), \qquad A_n := f^{-1}((n, \infty])$$

(see Figure 2.1). Then $A_{n,k} \in \mathcal{A}$ as f is measurable. Define

$$s_n := \sum_{k=1}^{n2^n-1} \frac{k-1}{2^n} \cdot \chi_{A_{n,k}} + n \cdot \chi_{A_n}.$$

Then on A_{nk}, we have $s_n(x) = (k-1)/2^n$. As n increases by 1, A_{nk} is divided into two disjoint halves:

$$A_{n,k} = A_{n+1,2k-1} \cup A_{n+1,2k}$$

and thus $s_{n+1}(x) = (2k-1-1)/2^{n+1} = s_n(x)$ or $s_{n+1}(x) = (2k-1)/2^{n+1} > s_n(x)$ for $x \in A_{nk}$. Similarly, one sees that $s_n \leqslant s_{n+1}$ on E_{n+1} and on $f^{-1}((n, n+1])$, and the first statement is proven.

Now fix $x \in X$ and $n > x$, and let n and k be such that $x \in A_{nk}$. Then $f(x) \in ((k-1)/2^n, k/2^n]$ and $s_n(x) = (k-1)/2^n$, thus $f(x) - s_n(x) \leqslant 2^{-n}$. This shows the second statement. □

Corollary 2.12. Let $(X, \mathcal{A}, \mu)$ be a measure space. A function $f : X \to [0, \infty]$ is measurable iff it is the pointwise limit of simple functions.

Proof. This follows from Prop 2.6 and Th 2.11. □

2.3 Integrals of positive functions

Definition 2.13 (Integral of a positive function)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space and $f : X \to [0, \infty]$ measurable. The Lebesgue Integral of f with respect to μ is

$$\int f d\mu := \sup \left\{ \int s d\mu : s \text{ is simple, measurable and } 0 \leqslant s \leqslant f \right\}.$$

For $E \in \mathcal{A}$ define

$$\int_E f \, d\mu := \int \chi_E \cdot f \, d\mu.$$

The supremum is not taken over the empty set, and thus $\int f d\mu$ is well defined; though it may well equal ∞.

Proposition 2.14. Let $(X, \mathcal{A}, \mu)$ be a measure space and f and g non-negative measurable functions on X such that $f \leqslant g$. Then

$$\int_E f \, d\mu \leqslant \int_E g \, d\mu \qquad \forall\, E \in \mathcal{A}.$$

Proof. Since $0 \leqslant f \leqslant g$ any simple measurable non-negative function s such that $0 \leqslant s \leqslant f$ also satisfies $0 \leqslant s \leqslant g$, hence as the integral is the supremum over these,

$$\int f d\mu \leqslant \int g d\mu.$$

Since $\chi_E f \leqslant \chi_E g \ \ \forall\, E \in \mathcal{A}$ the claim follows. □

Note: If $f(x) \geqslant m > 0 \ \ \forall\, x \in E$ then $g := m \cdot \chi_E \leqslant f \cdot \chi_E$ so

$$\int_E f \, d\mu \geqslant \int_E m \cdot \chi_E \, d\mu = m\, \mu(E).$$

Theorem 2.15 (Monotone Convergence Theorem, Beppo Levi)**.** For a measure space $(X, \mathcal{A}, \mu)$ let f_n be non-negative measurable functions ($n \in \mathbb{N}$) such that $0 \leqslant f_1 \leqslant f_2 \leqslant \ldots$ and $f(x) := \lim_n f_n(x)$ exists for every $x \in X$. Then the thusly defined function f is also non-negative measurable, and

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

That is, limit and integration may be interchanged.

Proof. That f is measurable follows from Prop 2.6 and the fact that $f = \limsup_n f_n$ $(= \liminf_n f_n = \sup_n f_n)$. Prop 2.14 implies

$$0 \leqslant \int f_1 \, d\mu \leqslant \int f_2 \, d\mu \leqslant \dots \text{ and } \int f_n d\mu \leqslant \int f \, d\mu \, \forall n \in \mathbb{N}$$

and hence

$$\alpha := \lim_{n\to\infty} \int f_n \, d\mu \leqslant \int f \, d\mu. \tag{2.16}$$

It remains to show the opposite inequality "$\geqslant$" in (2.16). Let $c \in (0,1)$ and s be a simple function such that $0 \leqslant s \leqslant f$. Define

$$A_n := \{x \in X : f_n(x) \geqslant cs(x)\},$$

then note that $A_n \in \mathcal{A}$ due to Prop 2.6. Since $f_n \leqslant f_{n+1}$, we have $A_1 \subseteq A_2 \subseteq \dots$. Since $c < 1$, we have $cs \leqslant cf \leqslant f = \lim_n f_n$, and so $\bigcup_n A_n = X$. Then $f_n \geqslant f_n \chi_{A_n} \geqslant cs\chi_{A_n}$, and so

$$\int f_n \, d\mu \geqslant \int_{A_n} f_n \, d\mu \geqslant \int_{A_n} cs \, d\mu. \tag{2.17}$$

Suppose that s has the representation $s = \sum_{i=1}^{k} \alpha_i \chi_{B_i}$; then

$$\lim_n \int_{A_n} cs \, d\mu = \lim_n \sum_{i=1}^{k} c\alpha_i \mu(B_i \cap A_n) = \sum_{i=1}^{k} c\alpha_i \mu(B_i) = c \int s \, d\mu$$

using linearity of sums and Th 1.12 part 3. The rightmost term in (2.17) hence converges to $c \int s \, d\mu$, and the leftmost term to α. This shows $\alpha \geqslant c \int s \, d\mu$. This inequality holds for every $c \in (0,1)$, and hence $\alpha \geqslant \int s \, d\mu$. The latter inequality holds for every simple measurable s, and thus by Def 2.13, $\alpha \geqslant \int f \, d\mu$. □

Using Beppo Levi's theorem, we derive the usual properties of the Lebesgue Integral:

Theorem 2.18. Let $(X, \mathcal{A}, \mu)$ be a measure space, and let f and g be non-negative measurable functions on X. Then

(1) $\int c \cdot f \, d\mu = c \int f \, d\mu \qquad \forall c \in [0, \infty)$;

(2) $\int (f + g) \, d\mu = \int f \, d\mu + \int g \, d\mu$;

(3) $\int f \, d\mu = 0$ iff $f = 0$ μ-a.e.;

(4) The map $\nu : \mathcal{A} \to [0, \infty]$ by $\nu(A) := \int_A f \; d\mu \quad \forall A \in \mathcal{A}$ defines a measure ν. (The notation $d\nu = f \, d\mu$ is then common.)

Proof. (1) and (2) are left as an exercise. For (3), let $A_n = \{f > 1/n\}$ which is shorthand for $\{x \in X : f(x) > 1/n\}$. We have

$$\chi_{A_n} \leqslant n f$$

(check this by plugging in $x \in A_n$ and $x \notin A_n$). Then

$$\mu(A_n) = \int \chi_{A_n} d\mu \leqslant n \int f d\mu$$

by Prop 2.14. Suppose now that $\int f \, d\mu = 0$. Then $\mu(A_n) = 0$ for every n, and note that $\bigcup_n A_n = \{f > 0\}$; hence by subadditivity

$$\mu(\{f > 0\}) = \mu\left(\bigcup_n A_n\right) \leqslant \sum_n \mu(A_n) = \sum_n 0 = 0,$$

that is, $f = 0$ μ-a.e. On the other hand, suppose that $\int f \, d\mu > 0$. Then by Def 2.13 there exists a simple, non-negative measurable s such that $0 < \int s \, d\mu$. Suppose this s has representation $\sum_k \alpha_k \chi_{B_k}$. Since $\sum_k \alpha_k \mu(B_k)$ is positive, for at least one pair (α_k, B_k) we have both $\alpha_k > 0$ and $\mu(B_k) > 0$. But then f is not equal to 0 μ-a.e. since $f(x) \geqslant \alpha_k > 0$ for all $x \in B_k$ which has positive measure.

For (4), let $\nu : \mathcal{A} \to [0, \infty]$ by $\nu(A) := \int_A f \; d\mu$. Since $\nu(\varnothing) = 0$ we only need to check that ν satisfies countable additivity. Let $A_n \in \mathcal{A}$ be pairwise disjoint and $A := \bigcup\limits_{n=1}^{\infty} A_n \in \mathcal{A}$. Then $\chi_A \cdot f = \sum\limits_{n=1}^{\infty} \chi_{A_n} \cdot f$. Let $f_k := \sum\limits_{n=1}^{k} \chi_{A_n} f$; then $0 \leqslant f_1 \leqslant f_2 \leqslant \dots$ and $\lim\limits_{k \to \infty} f_k(x) = f(x)\chi_A(x)$. So

$$\begin{aligned}\nu(A) = \int_A f \; d\mu = \int \chi_A f \; d\mu = \int \lim_k f_k \, d\mu = \lim_k \int f_k \; d\mu \\ = \lim_{k \to \infty} \sum_{n=1}^{k} \int_{A_n} f \, d\mu = \sum_{n=1}^{\infty} \nu(A_n).\end{aligned}$$

□

We have established the Lebesgue integral of positive measurable functions. In order to extend our work to functions with positive and negative values, we just one more step:

Definition 2.19 (Lebesgue integral of measurable functions)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space and $f : X \to \mathbb{R} \cup \{-\infty, +\infty\}$ a measurable function. The positive and negative parts of f are given by $f_+ = f \vee 0$ and $f^- = -(f \wedge 0)$. If $\int f^+ \, d\mu < \infty$ *or* $\int f^- \, d\mu < \infty$, then the Lebesgue Integral of f with respect to μ is

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu \in \mathbb{R} \cup \{-\infty, +\infty\}.$$

We write $f \in L^1(\mu)$ whenever $\int |f| d\mu < \infty$.

We collect a few properties of the Lebesgue integral:

Proposition 2.20. Let $(X, \mathcal{A}, \mu)$ be a measure space and f and g two measurable functions $X \to \mathbb{R} \cup \{-\infty, +\infty\}$. The following statements hold whenever the integrals exist:

(1) For every $\alpha \in \mathbb{R}$, $\int \alpha f \, d\mu = \alpha \int f \, d\mu$.

(2) $\int f + g \, d\mu = \int f \, d\mu + \int g \, d\mu$ (recall that $+\infty - \infty$ and $-\infty + \infty$ are undefined).

(3) $\left|\int f \, d\mu\right| \leqslant \int |f| \, d\mu$.

(4) $f \leqslant g$ μ-a.e. iff $\int_A f \, d\mu \leqslant \int_A g \, d\mu$ for every $A \in \mathcal{A}$.

(5) $f = g$ μ-a.e. iff $\int_A f \, d\mu = \int_A g \, d\mu$ for every $A \in \mathcal{A}$.

Proof. (1) and (2) follow from Th 2.18 and the decompositions $f = f^+ - f^-$, $g = g^+ - g^-$; details are left as an exercise. For (3), see that

$$\begin{aligned}\left|\int f \, d\mu\right| &= \left|\int f^+ \, d\mu - \int f^- \, d\mu\right| \leqslant \left|\int f^+ \, d\mu\right| + \left|\int f^- \, d\mu\right| \\ &= \int f^+ \, d\mu + \int f^- \, d\mu = \int f^+ + f^- \, d\mu = \int |f| \, d\mu\end{aligned}$$

by the triangle inequality. For (4), first note that

$$\begin{aligned}f \leqslant g \quad \mu\text{-a.e.} &\Rightarrow g - f = (g - f)^+ \quad \mu\text{-a.e.} \\ &\Rightarrow \int g - f \, d\mu \geqslant 0 \Rightarrow \int g \, d\mu \geqslant \int f \, d\mu\end{aligned}$$

The reverse implication "$A \Leftarrow B$" is equivalent to "not $A \Rightarrow$ not B", which we show. Let $h := (f - g)^+$ and $A := \{h > 0\}$. If $\mu(A) > 0$, then $\int h \, d\mu > 0$ by Th 2.18(3) (negate both sides). But this means $\int_A f \, d\mu > \int_A g \, d\mu$. (5) follows from (4), applied to $f \leqslant g$ and $g \leqslant f$. $\square$

In $X = \mathbb{R}^d$, the Lebesgue integral with μ being Lebesgue measure generalises the Riemann integral. The main advantage however is that it extends e.g. to any measure space and in particular any (separable) metric space, which is needed for many applications analysis and probability.

2.4 Dominated Convergence Theorem

Recall that for a sequence $\{a_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$, $\liminf_n a_n$ denotes the smallest of all its limit points. A point $a \in \mathbb{R}$ is a limit point if $\forall \varepsilon > 0$ the neighbourhood $(a-\varepsilon, a+\varepsilon)$ contains infinitely many members of the sequence. There may well be infinitely many limit points; but it can be shown that the set of limit points is closed, that is, the infimum of all limit points is again a limit point. Hence it is valid to say that $\liminf_n a_n$ is "the" smallest limit point. Moreover, the following is a valid characterisation of $a := \liminf_n a_n$:

$$\begin{aligned} \forall \varepsilon > 0 : a_n < a - \varepsilon &\text{ for at most finitely many } n, \text{ and} \\ a_n < a + \varepsilon &\text{ for infinitely many } n. \end{aligned}$$

Exercise 2.21. Let $\{a_n\}_{n\in\mathbb{N}} \subset \mathbb{R}$, and let a satisfy the line above. Show that

$$a = \sup_n \inf_{k\geqslant n} a_k.$$

Lemma 2.22 (Fatou)**.** For any sequence of non-negative measurable functions $\{f_n\}_{n\in\mathbb{N}}$ on a measure space $(X, \mathcal{A}, \mu)$,

$$\int \liminf_n f_n \, d\mu \leqslant \liminf_n \int f_n \, d\mu.$$

Proof. Define the sequence of functions $g_n := \inf_{k\geqslant n} f_k$ by $g_n(x) := \inf_{k\geqslant n} f_k(x)$, and note that each g_n is measurable (Prop 2.6). Also note that $g_n(x)$ is non-decreasing for each $x \in X$, and $\lim_n g_n = \sup_n g_n = \liminf_n f_n$. Then $\forall k \geqslant n$ we have $g_n \leqslant f_k$, and thus $\int g_n \, d\mu \leqslant \int f_k \, d\mu$, and thus

$$\int g_n \, d\mu \leqslant \inf_{k\geqslant n} \int f_k \, d\mu \quad \forall n.$$

Using Beppo Levi's theorem, we find for the limit of the left-hand side

$$\lim_n \int g_n \, d\mu = \int \lim_n g_n \, d\mu = \int \liminf_n f_n \, d\mu.$$

The right-hand side is non-decreasing in n, and hence

$$\lim_n \inf_{k\geqslant n} \int f_k \, d\mu = \sup_n \inf_{k\geqslant n} \int f_k \, d\mu = \liminf_n \int f_k \, d\mu.$$

The inequality must be preserved after taking limits, which proves the lemma. □

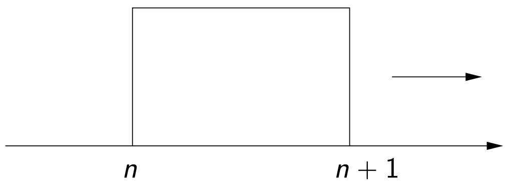

Figure 2.2: Characteristic function on the interval $(n, n+1]$

Example 2.23 (Moving Hump)**.** We know that if the convergence is monotone, then $\liminf_n f_n = \lim_n f_n$, and by Beppo Levi's theorem equality holds in Fatou's lemma. To see that we can have a strict inequality, consider the non-monotone sequence let $f_n = \chi_{(n,n+1]}$ (see Figure 2.2).

Then $\lim_n f_n(x) = 0 \;\; \forall x$ and $\int_{\mathbb{R}} f_n \, d\mu = 1$. So

$$0 = \int 0 \, d\mu = \int_{\mathbb{R}} \liminf_n f_n \, d\mu = 0 < \liminf_n \int_{\mathbb{R}} f_n \, d\mu = 1.$$

Theorem 2.24 (Dominated Convergence Theorem)**.** For a measure space $(X, \mathcal{A}, \mu)$, let $f \in L^1(\mu)$ and $\{f_n\}_{n\in\mathbb{N}} \subset L^1(\mu)$ be a sequence of integrable functions such that

(1) $\lim_n f_n(x)$ exists μ-a.e., and $\lim_n f_n(x) = f(x)$ where it exists

(2) for some $g \in \mathcal{L}^1(\mu)$ we have $|f_n(x)| \leqslant g(x)$ μ-a.e. for every n (that is, f_n is dominated by g).

Then

$$\lim_n \int |f_n - f| \, d\mu = 0 \quad \text{and} \quad \lim_n \int f_n \, d\mu = \int f \, d\mu.$$

Proof. Let N_1 be the set of x such that $\lim_n f_n(x)$ does not exist; and let $N_n^{(2)}$ be the set of x such that $|f_n(x)| > g(x)$ These sets are all μ-null, and hence $N := N_1 \cup \bigcup_n N_n^{(2)}$ is μ-null. On $X \backslash N$, $|f| \leqslant g$, and so from the triangle inequality we know that $|f_n - f| \leqslant 2g$ holds on $X \backslash N$. Thus $2g - |f_n - f|$ is non-negative on $X \backslash N$. By Fatou's Lemma,

$$\int_{X\backslash N} \liminf_n (2g - |f_n - f|) \, d\mu \leqslant \liminf_n \int_{X\backslash N} (2g - |f_n - f|) \, d\mu \qquad (2.25)$$

Noting that $\liminf_n (2g - |f_n - f|)\chi_{X\backslash N} = \lim_n (2g - |f_n - f|)\chi_{X\backslash N} = 2g\chi_{X\backslash N}$, we see that the left-hand side above equals $2\int_{X\backslash N} g \, d\mu$. For the right-hand side, first note that if $\{a_n\}_{n\in\mathbb{N}}$ is a sequence in $\mathbb{R}$, then $\liminf_n (a_n + c) = c + \liminf a_n$

for every $c \in \mathbb{R}$ and $\liminf_n(-a_n) = -\limsup_n a_n$. Hence

$$\begin{aligned}\liminf_n \int_{X\setminus N} 2g - |f_n - f|\, d\mu &= \liminf_n \left(2\int_{X\setminus N} g\, d\mu - \int_{X\setminus N} |f_n - f|\, d\mu\right)\\ &= 2\int_{X\setminus N} g\, d\mu - \limsup_n \int_{X\setminus N} |f_n - f|\, d\mu.\end{aligned}$$

Inequality (2.25) now reads

$$2\int_{X\setminus N} g\, d\mu \leqslant 2\int_{X\setminus N} g\, d\mu - \limsup_n \int_{X\setminus N} |f_n - f|\, d\mu,$$

which implies $\limsup_n \int_{X\setminus N} |f_n - f|\, d\mu = 0$. The sequence $\int_{X\setminus N} |f_n - f|\, d\mu$ can't have any negative limit points, and so 0 must be the only limit point, that is, $\lim_n \int_{X\setminus N} |f_n - f|\, d\mu = 0$. The first assertion follows since $\int_N |f_n - f|\, d\mu = 0$. Finally, note that

$$0 \leqslant \left|\int f_n\, d\mu - \int f\, d\mu\right| = \left|\int f_n - f\, d\mu\right| \leqslant \int |f_n - f|\, d\mu \to 0$$

as $n \to \infty$, which shows the second assertion. □

Example 2.26. Let $E_n \in \mathcal{A}$ and $E_1 \supset E_2 \supset \ldots,\ E := \bigcap_{i=1}^{\infty} E_i$. Let $f \in L^1(\mu)$ and define $f_n = \chi_{E_n} \cdot f$. Then

$$\chi_E \cdot f = \lim_{n\to\infty} f_n \qquad \text{and} \qquad |f_n| \leqslant |f| \in L^1(\mu) \qquad \forall\, n.$$

So by DCT

$$\lim_{n\to\infty} \int_X f_n\ d\mu = \lim_{n\to\infty} \int_{E_n} f\ d\mu = \int_E f\ d\mu.$$

Example 2.27. The "dominated" part is important. Consider the moving hump example $f_n = \chi_{[n,n+1]}$ from Figure 2.2. Then $\lim_n f_n(x) = 0$ for every $x \in X$, and the smallest function which dominates all f_n is $\chi_{[0,\infty)}$, which is not in $L^1(\mu)$. We have

$$\lim_{n\to\infty} \int_{\mathbb{R}} f_n\ d\mu = 1 \neq 0 = \int_{\mathbb{R}} \lim_{n\to\infty} f_n\ d\mu.$$

Corollary 2.28 (Exchanging the order of integral and limit/derivative)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space, and let f be a function $f : X \times [a, b] \to \mathbb{R}$. Define $f_t(x) = f(x, t)$, and assume that $f_t \in L^1(\mu)$ for all $t \in [a, b]$.

(1) If there is a $g \in L^1(\mu)$ such that $|f_t| \leqslant g$ on X for every $t \in [a, b]$, and if $\lim_{t \to t_0} f(x, t) = f(x, t_0)$ for every $x \in X$, then

$$\lim_{t \to t_0} \int_X f(x, t)\ d\mu(x) = \int_X f(x, t_0)\ d\mu(x),$$

and in particular $F(t) := \int f(x, t)\ d\mu(x)$ is continuous at t_0.

(2) If the partial derivative $\partial f/\partial t$ exists for all $(x, t) \in X \times [a, b]$, and if there is a function $g \in L^1(\mu)$ such that $|\partial f/\partial t| \leqslant g$ for every $x \in X$ and $t \in (a, b)$, then F is differentiable and

$$F'(t) = \int \frac{\partial f}{\partial t}(x, t)\ d\mu(x) \qquad \text{for } t \in (a, b).$$

Proof.

(1) Apply DCT to $f_n(x) := f(x, t_n)$ where t_n is any sequence which converges to t_0.

(2) Let

$$h_n(x) := \frac{f(x, t_n) - f(x, t_0)}{t_n - t_0} \qquad \text{where } t_n \to t_0.$$

Then

$$\frac{\partial f}{\partial t}(x, t_0) = \lim_{n \to \infty} h_n(x).$$

So $\partial f/\partial t$ is measurable with respect to x. By the Mean Value Theorem with respect to t we have

$$|h_n(x)| \leqslant \sup_{t \in (a,b)} \left| \frac{\partial f}{\partial t}(x, t) \right| \leqslant g(x).$$

But then

$$F'(t_0) = \lim_{n \to \infty} \frac{F(t_n) - F(t_0)}{t_n - t_0} = \lim_{n \to \infty} \int_X h_n(x)\ \mu(x) = \int_X \frac{\partial f}{\partial t}(x, t_0)\ d\mu(x)$$

by the dominated convergence theorem.

□

Chapter 3

Product measures

3.1 Construction

Definition 3.1. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be σ-algebras. The product σ-algebra of $\mathcal{A}$ and $\mathcal{B}$ is

$$\mathcal{A} \otimes \mathcal{B} := \sigma(\mathcal{A} \times \mathcal{B}).$$

A set $A \times B \in \mathcal{A} \times \mathcal{B}$ is called a (measurable) rectangle.

Example 3.2. We have $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. Recall that $\mathcal{B}(\mathbb{R}^2)$ is generated by 2-blocks of the form $(s_1, t_1] \times (s_2, t_2]$; Since these lie in $\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$, we have $\mathcal{B}(\mathbb{R}^2) \subset \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$. Conversely, let $P_i : \mathbb{R}^2 \to \mathbb{R}$ be the projections $P_1(x, y) = x$ and $P_2(x, y) = y$. These are continuous and hence Borel measurable. So if $A, B \in \mathcal{B}(\mathbb{R})$ then

$$A \times B = (A \times \mathbb{R}) \cap (\mathbb{R} \times B) = P_1^{-1}(A) \cap P_2^{-1}(B) \in \mathcal{B}(\mathbb{R}^2).$$

(see Figure 3.1). Thus we obtain the reverse inclusion, and

$$\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}).$$

Definition 3.3. For a map $f : X \times Y \to Z$ define the x-slices and y-slices f_x, f^y on Y, X respectively by

$$f_x(y) := f(x, y) =: f^y(x)$$

By identifying a set $E \in X \times Y$ with its indicator function χ_E, x- and y-slices extend to sets (see Figure 3.2).

Theorem 3.4. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces.

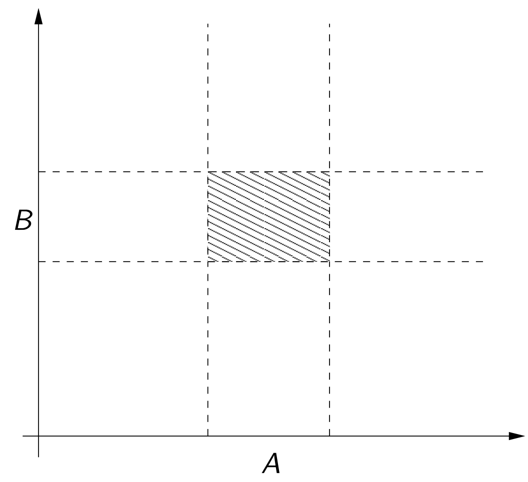

Figure 3.1: The intersection of A and B

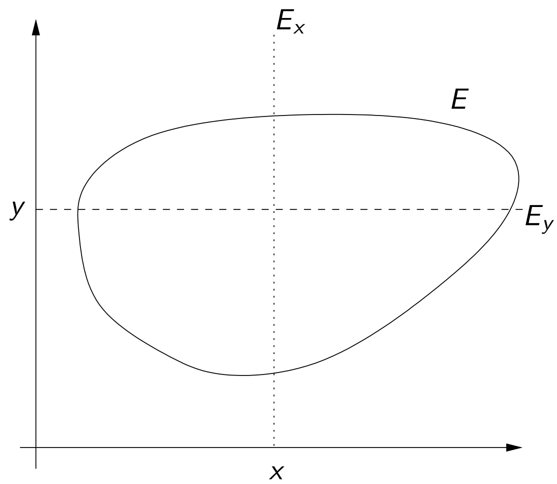

Figure 3.2: The slices E_x and E_y

(1) If $E \in \mathcal{A} \otimes \mathcal{B}$ then

$$E_x := \{y \in Y : (x, y) \in E\} \in \mathcal{B} \qquad \forall x \in X$$

(see Figure 3.2).

(2) $f : X \times Y \to \mathbb{R}$ is $\mathcal{A} \otimes \mathcal{B}$-measurable iff f_x is $\mathcal{B}$-measurable for every $x \in X$.

Proof. (1) Let $x \in X$ and

$$\mathcal{F} := \{E \in \mathcal{A} \otimes \mathcal{B} : E_x \in \mathcal{B}\}.$$

Then $\mathcal{A} \times \mathcal{B} \subset \mathcal{F} \subset \mathcal{A} \otimes \mathcal{B}$. Since

$$(E^{\complement})_x = (E_x)^{\complement} \qquad \text{and} \qquad \left(\bigcup_{n=1}^{\infty} E_n \right)_x = \bigcup_{n=1}^{\infty} (E_n)_x,$$

$\mathcal{F}$ is closed with respect to complements and countable unions. Thus $\mathcal{F}$ is a σ-algebra, so

$$\mathcal{A} \otimes \mathcal{B} = \sigma(\mathcal{A} \times \mathcal{B}) \subset \sigma(\mathcal{F}) = \mathcal{F} \subset \mathcal{A} \otimes \mathcal{B}.$$

which means $\mathcal{F} = \mathcal{A} \otimes \mathcal{B}$.

(2) Since $(f_x)^{-1}((a, b]) = (f^{-1}((a, b]))_x$ this follows from part (1) and Th 2.3.

□

Theorem 3.5. Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces. For $E \in \mathcal{A} \otimes \mathcal{B}$, the function $\varphi^E : X \to [0, \infty]$ given by $\varphi^E(x) := \nu(E_x)$ is $\mathcal{A}$-measurable.

Proof. Assume first that ν is finite and observe that due to Th 3.4, $\varphi^E(x) = \nu(E_x)$ is well defined. Define

$$\mathcal{F} := \{E \in \mathcal{A} \otimes \mathcal{B} : \varphi^E \text{ is } \mathcal{A}\text{-measurable}\}.$$

This contains all rectangles $A \times B \in \mathcal{A} \times \mathcal{B}$ because $\nu((A \times B)_x) = \nu(B) \cdot \chi_A(x)$ which is $\mathcal{A}$-measurable. Now $\mathcal{F}$ is a Dynkin Class on $X \times Y$ because

- $X \times Y \in \mathcal{F}$;
- If $E, F \in \mathcal{F}$ with $E \subset F$ then

$$\nu((F \backslash E)_x) = \nu(F_x \backslash E_x) = \nu(F_x) - \nu(E_x) \Rightarrow F \backslash E \in \mathcal{F}.$$

- If $\{E_n\}_{n=1}^{\infty} \subset \mathcal{F}$ is an increasing sequence then so is $\{(E_n)_x\}_{n=1}^{\infty} \subset \mathcal{B}$. Thus

$$\varphi^{\bigcup_n E_n}(x) = \nu\left(\left(\bigcup_n E_n \right)_x \right) = \nu\left(\bigcup_n (E_n)_x \right) = \lim_n \nu\left((E_n)_x\right) = \lim_n \varphi^{E_n}(x)$$

for all $x \in X$, where the third equality is due to Th 1.12. By Prop 2.6, the last term is seen to be $\mathcal{A}$-measurable.

Now $\mathcal{A} \times \mathcal{B}$ is a π-system, since $(A \times B) \cap (C \times D) = (A \cap C) \times (B \cap D)$. So since $\mathcal{A} \times \mathcal{B} \subset \mathcal{F}$, we have

$$\mathcal{A} \otimes \mathcal{B} = \sigma(\mathcal{A} \times \mathcal{B}) = d(\mathcal{A} \times \mathcal{B}) \subset d(\mathcal{F}) = \mathcal{F} \subset \mathcal{A} \otimes \mathcal{B},$$

where the second equality follows from the monotone class theorem, Th 1.23. But this means $\mathcal{F} = \mathcal{A} \otimes \mathcal{B}$.

Now let ν be σ-finite, so $Y = \bigcup_n D_n$ with $\nu(D_n) < \infty$. Define the sequence of finite measures $\nu_n(B) := \nu(B \cap D_n)$. We may assume D_n increasing. Then

$$\begin{aligned}\varphi^E(x) = \nu(E_x) = \nu\left(E_x \cap \bigcup_n D_n\right) &= \nu\left(\bigcup_n (E_x \cap D_n)\right) \\ &= \lim_n \nu(E_x \cap D_n) = \lim_n \nu_n(E_x)\end{aligned}$$

where the fourth equality is due to Th 1.12. By what has been shown above, the last term is a measurable function of x, and hence by Prop 2.6 it is measurable. □

Theorem 3.6. Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces. Then there is a unique measure $\mu \times \nu : \mathcal{A} \otimes \mathcal{B} \to [0, \infty]$ such that

$$(\mu \times \nu)(A \times B) = \mu(A)\nu(B) \qquad \forall\, A \in \mathcal{A},\ B \in \mathcal{B}.$$

Moreover, for any $E \in \mathcal{A} \otimes \mathcal{B}$ we have

$$(\mu \times \nu)(E) = \int_X \nu(E_x)\, d\mu(x) = \int_Y \mu(E^y)\, d\nu(y).$$

The measure $\mu \times \nu$ is called the product measure of μ and ν.

Proof. From Th 3.5 we know that the maps $x \mapsto \nu(E_x)$ and $y \mapsto \mu(E^y)$ are measurable. Now define

$$(\mu \times \nu)_1(E) := \int_Y \mu(E^y)\, d\nu(y) \qquad \text{and} \qquad (\mu \times \nu)_2(E) := \int_X \mu(E_x)\, d\mu(x).$$

Then

$$\begin{aligned}(\mu \times \nu)_1(A \times B) &= \int_Y \mu(A) \cdot \chi_B(y)\, d\nu(y) = \mu(A)\nu(B) \\ &= \int_X \nu(B) \cdot \chi_A(x)\, d\mu(x) = (\mu \times \nu)_2(A \times B).\end{aligned} \tag{3.7}$$

We can check that $(\mu \times \nu)_1(\varnothing) = 0 = (\mu \times \nu)_2(\varnothing)$. We check countable additivity of $(\mu \times \nu)_1$. Let $\{E_n\}_{n \in \mathbb{N}} \subset \mathcal{A} \otimes \mathcal{B}$ be a disjoint sequence such that

$E = \bigcup_n E_n$. Then $\{(E_n)^y\}_{n\in\mathbb{N}} \subset \mathcal{A}$ is a disjoint sequence: Indeed, suppose there is an $x \in (E_m)^y \cap (E_n)^y$ for $n \neq m$. Then $(x, y) \in E_m$ and $(x, y) \in E_n$, which contradicts $E_n \cap E_m = \emptyset$. So we have

$$\mu(E^y) = \mu\left(\left(\bigcup_n E_n\right)^y\right) = \mu\left(\bigcup_n (E_n)^y\right) = \sum_{n=1}^{\infty} \mu((E_n)^y) \tag{3.8}$$

We hence obtain

$$\begin{aligned}(\mu \times \nu)_1(E) = \int_Y \mu(E^y)\ d\nu(y) &= \int_Y \sum_{n=1}^{\infty} \mu((E_n)^y)\ d\nu(y) \\ &= \sum_{n=1}^{\infty} \int_Y \mu((E_n)^y)\ d\nu(y) = \sum_{n=1}^{\infty} (\mu \times \nu)_1(E_n),\end{aligned}$$

where the third equality follows from the sequence $g_n := \sum_{k=1}^{n} \mu((E_k)^y)$ defining an increasing sequence of measurable functions with limit $g = \sum_{k=1}^{\infty} \mu((E_k)^y)$ and Beppo Levi's theorem Th 2.15. Thus $(\mu \times \nu)_1$ is a measure, and likewise one sees that $(\mu \times \nu)_2$ is a measure. We now want to apply Th 1.25 to show that $(\mu \times \nu)_1 = (\mu \times \nu)_2$. It remains to check that they are both σ-finite with respect to the same exhausting sequence on which they coincide and which lies in $\mathcal{A} \times \mathcal{B}$: Let

$$X = \bigcup_{k=1}^{\infty} A_k, \quad \mu(A_k) < \infty \qquad \text{and} \qquad Y = \bigcup_{j=1}^{\infty} B_j, \quad \mu(B_j) < \infty.$$

Then

$$(\mu \times \nu)_1(A_k \times B_j) = (\mu \times \nu)_2(A_k \times B_j) = \mu(A_k)\nu(B_j) < \infty$$

for all k and j, and

$$X \times Y = \bigcup_{(j,k)\in\mathbb{N}^2} A_k \times B_j$$

is a countable union. $\square$

Example 3.9. We have $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$. Write λ_n for the Lebesgue measure on $\mathbb{R}^n$; then

$$\lambda_2((a, b] \times (c, d]) = (b - a)(d - c) = (\lambda_1 \times \lambda_1)((a, b] \times (c, d])$$

hence $\lambda_2 = \lambda_1 \times \lambda_1$.

3.2 Fubini's Theorem

Theorem 3.10 (Tonelli-Fubini)**.** Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be σ-finite measure spaces.

(1) (Tonelli) If $f : X \times Y \to [0, \infty]$ is a $\mathcal{A} \otimes \mathcal{B}$-measurable function, then the functions g, h are measurable where

$$g(x) := \int_Y f_x \, d\nu \quad \forall\, x \in X \qquad \text{and} \qquad h(y) := \int_X f^y \, d\mu \quad \forall\, y \in Y.$$

Moreover

$$\begin{aligned} \int_{X\times Y} f \, d(\mu \times \nu) &= \int_X \left[\int_Y f(x, y) \, d\nu(y) \right] \, d\mu(x) \\ &= \int_Y \left[\int_X f(x, y) \, d\mu(x) \right] \, d\nu(y) \end{aligned} \tag{3.11}$$

(2) (Fubini) If $f \in L^1(\mu \times \nu)$ then $f_x \in L^1(\nu)$ for μ-a.e. x, and $f^y \in L^1(\mu)$ for ν-a.e. y. Moreover, if $g(x) = \int_Y f_x \, d\nu$ then $g \in L^1(\mu)$. If $h(y) = \int_X f^y \, d\mu$ then $h \in L^1(\nu)$ and Equation (3.11) holds.

Proof. If f is a characteristic function χ_E, then (1) follows from Th 3.6. If f is a simple function, then (1) follows from the linearity of the integral. For f non-negative and measurable on $X \times Y$, let s_n be simple functions as in Th 2.11 such that

$$0 \leqslant s_1 \leqslant s_2 \leqslant \dots \qquad \text{and} \qquad \lim_{n\to\infty} s_n(x, y) = f(x, y) \quad \forall\, x, y.$$

Then $(s_n)_x$ and $(s_n)^y$ are increasing sequences of functions, for every $x \in X$ (resp. $y \in Y$). Their limiting functions are $\lim_n (s_n)_x = f_x$ and $\lim_n (s_n)^y = f^y$. Let

$$g_n(x) := \int_Y (s_n)_x(y) \, d\nu(y) \qquad \text{and} \qquad h_n(y) := \int_X (s_n)^y(x) \, d\mu(x).$$

Then by Beppo-Levi's Th 2.15,

$$\lim_n g_n(x) = \int_Y f_x \, d\nu(y) \quad \forall x \qquad \text{and} \qquad \lim_n h_n(y) = \int_X f^y \, d\mu(x) \quad \forall y.$$

Denote $g := \lim_n g_n$ and $h := \lim_n h_n$; we know from Prop 2.6 that g and h are measurable. Then

$$\begin{aligned} \int_X g \, d\mu = \lim_n \int_X g_n \, d\mu &= \lim_n \int_X \left(\int_Y (s_n)_x \, d\nu \right) \, d\mu \\ &= \lim_n \int_{X\times Y} s_n \, d(\mu \times \nu) = \int_{X\times Y} f \, d(\mu \times \nu) \end{aligned}$$

holds, where the first and last equality follows from Beppo-Levi's Th 2.15. Similarly, one sees $\int_Y h\,d\nu = \int_{X\times Y} f\,d(\mu\times\nu)$. This establishes (1) (Tonelli's Theorem).

It also shows that if $f \in L^1(\mu\times\nu)$ is non-negative, then (3.11) holds, and

$$\int_{X\times Y} f\ d(\mu\times\nu) < \infty \Rightarrow g \in L^1(\mu) \text{ and } h \in L^1(\nu).$$

Hence $g < \infty$ μ-a.e. But this means that for μ-a.e. $x \in X$, we have $\int_Y f_x\,d\nu(y) < \infty$. Since f_x is non-negative, this means $f_x \in L^1(\nu)$ for μ-a.e. x. Symmetrically, one sees that $f^y \in L^1(\mu)$ for ν-a.e. $y \in Y$.

Finally, suppose that $f \in L^1(\mu\times\nu)$ has both a non-zero positive part f^+ and a non-zero negative part f^-. Then $f^+ \in L^1(\mu\times\nu)$ and $f^- \in L^1(\mu\times\nu)$ are both non-negative. By the above paragraph, $(f^+)_x$, $(f^-)_x$, $(f^+)^y$ and $(f^-)^y$ are all in $L^1(\nu)$ for μ-a.e. x (resp. in $L^1(\mu)$ for ν-a.e. y). But then $f_x = (f^+)_x - (f^-)_x \in L^1(\nu)$ for μ-a.e. x and $f^y = (f^+)^y - (f^-)^y \in L^1(\mu)$ for ν-a.e. y. Equation (3.11) then still holds since it holds for positive integrands and the integral is linear. □

Example 3.12. Let $\{a_{n,m}\}_{(n,m)\in\mathbb{N}^2} \subset \mathbb{R}$ be a doubly indexed sequence of real numbers such that

$$\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} |a_{n,m}| < \infty. \tag{3.13}$$

Then $\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} a_{n,m} = \sum_{m=1}^{\infty}\sum_{n=1}^{\infty} a_{n,m}$.

Proof. Let $(X,\mathcal{A},\mu) = (\mathbb{N},\mathcal{P}(\mathbb{N}),\epsilon) = (Y,\mathcal{B},\nu)$ where ϵ denotes the counting measure. Then $\int_{\mathbb{N}} f(n)\ d\epsilon(n) = \sum_{n=1}^{\infty} f(n)$, and

$$\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} a_{n,m} = \int_{\mathbb{N}}\int_{\mathbb{N}} a_{n,m}\ d\epsilon(m)\ d\epsilon(n).$$

Equation (3.13) implies that the mapping $a : (n,m) \mapsto a_{n,m}$ lies in $L^1(\epsilon\times\epsilon)$. The statement now follows from Fubini's theorem. □

Chapter 4

Absolute Continuity

4.1 Signed Measures

Definition 4.1. Let $(X, \mathcal{A})$ be a measurable space. A signed measure is a map $\nu : \mathcal{A} \to [-\infty, \infty]$ such that

(1) $\nu(\varnothing) = 0$;

(2) for every countable disjoint sequence $\{A_j\}_{j\in\mathbb{N}} \subset \mathcal{A}$

$$\nu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \nu(A_j) \qquad \text{(Countable addivitivity);}$$

(3) $+\infty$ and $-\infty$ cannot both be attained.

A signed measure ν is finite if $\nu(A) \notin \{-\infty, +\infty\}$ for any $A \in \mathcal{A}$.

Example 4.2.

(1) Any positive measure is a signed measure.

(2) If μ_1, μ_2 are positive measures, one of which is finite, then $\nu = \mu_1 - \mu_2$ is a signed measure.

The following result can be proven just as parts 3. and 4. in Th 1.12.

Theorem 4.3. Let $(X, \mathcal{A}, \nu)$ be a signed measure space, then

(1) If $A_j \in \mathcal{A}$ with $A_1 \subset A_2 \subset \dots$ then

$$\nu\left(\bigcup_{j=1}^{\infty} A_j\right) = \lim_{j\to\infty} \nu(A_j)$$

(2) If $A_j \in \mathcal{A}$ with $A_1 \supset A_2 \supset \ldots$ and if $\nu(A_1) < \infty$ then

$$\nu\left(\bigcap_{j=1}^{\infty} A_j\right) = \lim_{j\to\infty} \nu(A_j).$$

Similar to the decomposition $f = f_+ - f_-$ for functions, we want to decompose any signed measure into a positive and negative part.

Definition 4.4. Let $(X, \mathcal{A}, \nu)$ be a signed measure space. A set $A \in \mathcal{A}$ is called positive (respectively negative, null) if $\forall B \in \mathcal{A}$ with $B \subset A$ we have $\nu(B) \geqslant 0$ (respectively $\nu(B) \leqslant 0$, $\nu(B) = 0$).

Note that subsets of positive sets are positive, and subsets of negative sets are negative.

Lemma 4.5. Let $(X, \mathcal{A}, \nu)$ be a signed measure space such that $\nu(A) \neq +\infty$ for any $A \in \mathcal{A}$. For any $E \in \mathcal{A}$ with $0 < \nu(E) < \infty$, there is a positive set $F \subset E$, such that $\nu(F) > 0$. In short: sets of positive measure have positive subsets of positive measure.

Proof. If E is positive, then $F = E$ is as required. If not, it has subsets of negative measure. Let $n_1 \in \mathbb{N}$ be the smallest number such that there is a measurable set $E_1 \subset E$ with $\nu(E_1) < -1/n_1$. Proceed: if $E \backslash E_1$ is not positive, let $n_2 \in \mathbb{N}$ be the smallest number such that there is a measurable set $E_2 \subset E \backslash E_1$ with $\nu(E_2) < -1/n_2$. Continue inductively, to obtain a sequence of sets $E_k \subset E \backslash \bigcup_{j=1}^{k-1} E_j$ of negative measure. If the sequence terminates after N steps, then $F = E \backslash \bigcup_{j=1}^{N} E_j$ is as required, that is, it is positive and $\nu(F) = \nu(E) - \sum_j \nu(E_j) > 0$.

Assume now that the sequence does not terminate. Let $F := E \backslash \bigcup_{j=1}^{\infty} E_j$, so we have a disjoint union $E = F \cup \bigcup_{j=1}^{\infty} E_j$. Hence

$$0 < \nu(E) = \nu(F) + \sum_{j=1}^{\infty} \nu(E_j)$$

All $\nu(E_j)$ are negative, and by assumption $\nu(F) < \infty$. So we have

$$0 < \sum_j |\nu(E_j)| < \nu(F) < \infty.$$

This shows $\nu(F) > 0$, and it remains to show that F is positive. By construction, $1/n_j < |\nu(E_j)|$, and so $\sum_j 1/n_j < \infty$, which implies $1/n_j \to 0$ and $n_j \to \infty$. Let

$L \subset F$, $L \in \mathcal{A}$ and fix $\varepsilon > 0$. Choose n_j with $1/(n_j - 1) < \varepsilon$. Recall that n_j is the *smallest* number in $\mathbb{N}$ such that

$$\exists\, E_j \subset E \backslash \bigcup_{k=1}^{j-1} E_j, \quad \nu(E_j) < -\frac{1}{n_j}.$$

Since $L \subset F \subset E \backslash \bigcup_{k=1}^{j-1} E_k$ this means $\nu(L) \geqslant -1/(n_j - 1) \geqslant -\varepsilon$. Since $\varepsilon > 0$ is arbitrary, $\nu(L) \geqslant 0$. □

4.2 Decomposition of Signed Measures

Theorem 4.6 (Hahn Decomposition Theorem)**.** For any signed measure space $(X, \mathcal{A}, \nu)$ there exists a positive set P and a negative set N such that $X = P \cup N$ and $P \cap N = \varnothing$. If P', N' is another such partition, then $P \Delta P' = N \Delta N'$ is null.

Proof. Without loss of generality assume $+\infty$ is not attained by ν (else replace ν by $-\nu$). Let $\beta := \sup\{\nu(A) : A \text{ is positive}\}$. Then $\beta \geqslant 0$, since $\varnothing$ is positive. Approximate β via a sequence $\{A_k\}_{k\in\mathbb{N}} \subset \mathcal{A}$ of positive sets such that $\beta = \lim_k \nu(A_k)$. Let $P := \bigcup_k A_k$. We show that P is positive. Disjointify the sequence A_k; that is, let $B_1 := A_1$ and $B_n := A_n \backslash \bigcup_{k=1}^{n-1} A_k$ if $n \geqslant 2$. Then the B_j are disjoint and $P = \bigcup_n B_n$. So if $E \subset P$, $E \in \mathcal{A}$ then $B_n \cap E \subset A_n$ are disjoint and $\nu(B_n \cap E) \geqslant 0$ as A_n is positive. So

$$\nu(E) = \nu\left(\bigcup_{n=1}^{\infty}(B_n \cap E)\right) = \sum_{n=1}^{\infty} \nu(B_n \cap E) \geqslant 0$$

by countable additivity. Thus P is positive. Moreover, $\nu(P \backslash A_k) \geqslant 0 \;\; \forall\, k$. Hence

$$\nu(P) = \nu(A_k) + \nu(P \backslash A_k) \geqslant \nu(A_k),$$

and so $\nu(P) \geqslant \lim_k \nu(A_k) = \beta$.

Next we show that $N = X \backslash P$ is negative. Let $E \subset N$, $E \in \mathcal{A}$. Suppose $\nu(E)$ were > 0, then by Lem 4.5 there would be a positive $F \subset E$, $F \in \mathcal{A}$ with $\nu(F) > 0$. But then $F \cup P$ would also be positive, so

$$\beta \geqslant \nu(F \cup P) = \nu(F) + \nu(P) \geqslant \nu(F) + \beta.$$

This would contradict $\nu(F) > 0$. Thus $\nu(E) < 0$, and N is negative.

Finally, if N', P' is another partition $X = P' \cup N'$, then $P \backslash P' \subset P$ and $P \backslash P' \subset N'$. So $P \backslash P'$ is both positive and negative, that is, null. □

Definition 4.7. Two signed measures μ, ν on $(X, \mathcal{A})$ are mutually singular (denoted $\mu \perp \nu$) if there is a partition $X = E \cup F$ with $E \cap F = \varnothing$ and $E, F \in \mathcal{A}$ such that E is null for μ and F is null for ν.

Theorem 4.8 (Jordan Decomposition)**.** Let $(X, \mathcal{A}, \nu)$ be a signed measure space. Then there are two unique *positive* measures ν^+ and ν^- such that $\nu = \nu^+ - \nu^-$ and $\nu^+ \perp \nu^-$.

Proof. Let $X = P \cup N$ for the Hahn decomposition of ν. Then the positive measures on $\mathcal{A}$ given by $\nu^+(A) = \nu(A \cap P)$ and $\nu^-(A) = -\nu(A \cap N)$ are as required.

Suppose there is another such decomposition $\nu = \mu^+ - \mu^-$ with $\mu^+ \perp \mu^-$. Then there are $E, F \in \mathcal{A}$ with $X = E \cup F$, $E \cap F = \varnothing$ and $\mu^+(F) = 0 = \mu^-(E)$. The decomposition $X = E \cup F$ is another Hahn decomposition for ν, so by Th 4.6 $P \Delta E$ is ν-null. Thus

$$\mu^+(A) = \mu^+(A \cap E) = \nu(A \cap E) = \nu(A \cap P) \text{ (as } P\Delta E \text{ is } \nu\text{-null)} \ = \nu^+(A).$$

Likewise $\mu^- = \nu^-$. □

Definition 4.9. For a signed measure space $(X, \mathcal{A}, \nu)$ we define:

(1) $\mathcal{L}^1(\nu) := \mathcal{L}^1(\nu^+) \cap \mathcal{L}^1(\nu^-)$, $\int f \, d\nu := \int f \, d\nu^+ - \int f \, d\nu^-$ for $f \in \mathcal{L}^1(\nu)$.

(2) the total variation of ν is the positive measure $|\nu| := \nu^+ + \nu^-$.

(3) ν is finite (respectively σ-finite if $|\nu|$ is finite (respectively σ-finite).

(4) Let μ be a *positive* measure on $(X, \mathcal{A})$, and suppose $\mu(A) = 0$ implies $\nu(A) = 0$ for all $A \in \mathcal{A}$. Then ν is absolutely continuous with respect to μ, written $\nu \ll \mu$.

(5) If two positive measures λ and μ satisfy $\lambda \ll \mu$ and $\mu \ll \lambda$, then they are equivalent, written $\lambda \sim \mu$.

Exercise 4.10. Let μ be a positive measure and ν a signed measure. Prove the following:

(1) $A \in \mathcal{A}$ is ν-null if and only if $|\nu|(A) = 0$.

(2) If $\nu \ll \mu$ and $\nu \perp \mu$ then $\nu = 0$.

(3) $\nu \ll \mu$ if and only if $|\nu| \ll \mu$ if and only if $\nu^+ \ll \mu$ and $\nu^- \ll \mu$.

(4) If $f \in \mathcal{L}^1(\mu)$, μ a positive measure and $\nu(A) = \int_A f \, d\mu$ then $\nu \ll \mu$.

4.3 The Radon-Nikodym Theorem

Theorem 4.11. Let ν be a σ-finite signed measure on $(X, \mathcal{A})$ and let μ be a σ-finite positive measure on $(X, \mathcal{A})$.

(1) (Lebesgue decomposition) There are unique σ-finite signed measures λ, ρ such that $\nu = \lambda + \rho$ with $\lambda \perp \mu$ and $\rho \ll \mu$.

(2) (Radon-Nikodym Theorem) There is a unique $f \in L^1(\mu)$ such that

$$d\rho = f\ d\mu \qquad \text{i.e. } \rho(A) = \int_A f\ d\mu \qquad \forall\, A \in \mathcal{A}.$$

Proof. First assume μ, ν are both finite and positive. Define

$$\mathcal{F} := \left\{ f : X \to [0, \infty] \;\middle|\; f \text{ is measurable and } \int_A f\ d\mu \leqslant \nu(A) \;\; \forall\, A \in \mathcal{A} \right\}$$

($0 \in \mathcal{F}$, where 0 is the zero function).

Claim. If $f, g \in \mathcal{F}$, then $h := \max(f, g) \in \mathcal{F}$.

Proof. Let $B := \{x \in X \mid f(x) > g(x)\}$ and $h = \chi_B \cdot f + \chi_{B^c} \cdot g$. Then $\forall\, A \in \mathcal{A}$,

$$\int_A h\ d\mu = \int_{A \cap B} f\ d\mu + \int_{A \cap B^c} g\ d\mu \leqslant \nu(A \cap B) + \nu(A \cap B^c) = \nu(A).$$

Thus $h \in \mathcal{F}$. □

Now let $a = \sup\{\int f\ d\mu : f \in \mathcal{F}\}$ and note that $0 \leqslant a \leqslant \nu(X) < \infty$. Choose a sequence $\{g_n\}_{n \in \mathbb{N}} \subset \mathcal{F}$ which approximates a, that is, $\int_X g_n\ d\mu \to a$. Then the sequence $f_n = \max(g_1, \ldots, g_n) \in \mathcal{F}$ satisfies $f_n \geqslant g_n$ and so $\int f_n\ d\mu \to a$. Moreover, $f_{n+1} \geqslant f_n$. Put $f := \lim_n f_n$. By Beppo Levi's Theorem, $\int_A f\ d\mu = \lim_n \int_A f_n\ d\mu \leqslant \nu(A) \;\; \forall\, A \in \mathcal{A}$. Thus $f \in \mathcal{F}$ and moreover

$$\int f\, d\mu = \lim_n \int f_n\, d\mu = a < \infty. \tag{4.12}$$

Since $f \geqslant 0$, we have $f \in L^1(\mu)$. Define a measure $d\lambda := d\nu - f\ d\mu$, that is,

$$\lambda(A) = \nu(A) - \int_A f\, d\mu \;\; \forall\, A \in \mathcal{A}.$$

As $f \in \mathcal{F}$ we see λ is a positive measure.

Claim. If $\lambda \perp \mu$ were not true, then there is $\varepsilon > 0$ and $E \in \mathcal{A}$ such that $\mu(E) > 0$ and E is a positive set for $\lambda - \varepsilon\mu$.

Proof. $\lambda - \frac{1}{n}\mu$ is a signed measure with Hahn decomposition $X = P_n \cup N_n$ with $P_n \cap N_n = \varnothing$. Define $P := \bigcup_{n=1}^{\infty} P_n$ and $N = P^c = \bigcap_{n=1}^{\infty} N_n$. Then N is a negative set for $\lambda - \frac{1}{n}\mu \quad \forall\, n$ i.e.

$$0 \leqslant \lambda(N) \leqslant \frac{1}{n}\mu(N) < \infty \qquad \forall\, n$$

$\Rightarrow \lambda(N) = 0$. If $\mu(P) = 0$ then $\lambda \perp \mu$. Otherwise $\mu(P) > 0$, hence $\mu(P_n) > 0$ for some n. Now P_n is a positive set for $\lambda - \frac{1}{n}\mu$, so $E := P_n$ and $\varepsilon := \frac{1}{n}$ are as required. ☐

So if λ and μ were *not* mutually singular, there is $\varepsilon > 0$ and $E \in \mathcal{A}$ such that $\mu(E) > 0$ and $\lambda(B) \geqslant \varepsilon\mu(B)$ for all $B \subset E$, $B \in \mathcal{A}$. This means that for any $A \in \mathcal{A}$

$$\varepsilon\mu(E \cap A) \leqslant \lambda(E \cap A) \leqslant \lambda(A) = \nu(A) - \int_A f\, d\mu.$$

Rearrange to get

$$\int_A (f + \varepsilon \cdot \chi_E)\, d\mu \leqslant \nu(A) \qquad \forall\, A \in \mathcal{A}$$

This implies $f + \varepsilon \cdot \chi_E \in \mathcal{F}$ and by (4.12) the contradiction

$$a \geqslant \int f + \varepsilon \cdot \chi_E\, d\mu = a + \varepsilon\mu(E) > a.$$

It follows that $\lambda \perp \mu$, and f and $d\rho := f\, d\mu$ are as required.

We will now prove uniqueness. If $d\nu = d\lambda' + f'\, d\mu$ is another such decomposition, then

$$d\lambda - d\lambda' = (f' - f)\, d\mu. \tag{4.13}$$

But $(\lambda - \lambda') \perp \mu$; to see this, first note that $\lambda \perp \mu$ means a decomposition $X = E \cup F$ such that $\lambda(E) = 0 = \mu(F)$. Similarly, $\lambda' \perp \mu$ means $X = E' \cup F'$ such that $\lambda'(E') = 0 = \mu(F')$. Then X is the disjoint union of $E \cap E'$ and $F \cup F'$, and $(\lambda - \lambda')(E \cap E') = 0$ and $\mu(F \cup F') = 0$.

Now since $(f' - f)\, d\mu \ll d\mu$, the measures in (4.13) are both singular and absolutely continuous to μ, and hence $= 0$; moreover, $f = f'$ μ-a.e. by Th 2.18(3).

Extending the proof to the σ-finite case is not spectacular: Let μ, ν be σ-finite that is, $X = \bigcup_n X_n = \bigcup_k Y_k$ with $\mu(X_n) < \infty$, $\nu(Y_k) < \infty \quad \forall\, n, k$ and $Z_{n,k} = X_n \cap Y_k$. Relabel $Z_{n,k}$ to obtain Z_j, $j \in \mathbb{N}$ such that

$$X = \bigcup_j Z_j \qquad \text{with } \mu(Z_j) < \infty \text{ and } \nu(Z_j) < \infty.$$

Disjointify; that is, let $W_n := \bigcup_{k=1}^{n} Z_k \setminus \bigcup_{k=1}^{n-1} Z_k$. Define $\mathcal{A}_j := \{A \cap W_j : A \in \mathcal{A}\}$, $\mu_j(A) := \mu(A \cap W_j)$ and $\nu_j(A) := \nu(A \cap W_j)$ for all $A \in \mathcal{A}$ and apply the theorem to the finite measure spaces $(W_j, \mathcal{A}_j, \nu_j)$ and μ_j. Then we have $d\nu_j = d\lambda_j + f_j\, d\mu_j$ where $\lambda_j \perp \mu_j$. This means that $W_j = E_j \cup F_j$ where $E_j \cap F_j = \emptyset$ and $\lambda_j(E_j) = 0 = \mu_j(F_j)$. We define $f_j(x) = 0$ if $x \notin E_j$. Let

$$\lambda = \sum_{j=1}^{\infty} \lambda_j \qquad \text{and} \qquad f = \sum_{j=1}^{\infty} f_j.$$

Then let $E = \bigcup E_j$, $F = \bigcup F_j$. We have the disjoint union $X = E \cup F$, and $\lambda(E) = 0 = \mu(F)$, hence $\lambda \perp \mu$. Moreover, $d\nu = d\lambda + f\, d\mu$. For uniquness, the same argument as above works. Finally, if $\nu = \nu^+ - \nu^-$ is a σ-finite *signed* measure, apply what has been shown above to ν^+ and ν^-. □

Definition 4.14. Let $(X, \mathcal{A})$ be a measure space, ν a signed measure and μ a positive measure such that $\nu \ll \mu$. Then by Th 4.11, $d\nu = f\, d\mu$ for a μ-a.e. unique $f \in L^1(\mu)$. This f is called the Radon-Nikodym derivative of ν with respect to μ and is denoted by $f = d\nu/d\mu$.

Theorem 4.15 (Chain Rule)**.** Let μ, λ, ν be σ-finite measures on $(X, \mathcal{A})$ with ν a signed measure, λ and μ are positive measures such that $\nu \ll \mu$ and $\mu \ll \lambda$. Then

(1) $g\frac{d\nu}{d\mu} \in L^1(\mu) \quad \forall g \in L^1(\nu)$ and

$$\int g\, d\nu = \int g \frac{d\nu}{d\mu}\, d\mu \tag{4.16}$$

(2) $\nu \ll \lambda$ and

$$\frac{d\nu}{d\lambda} = \frac{d\nu}{d\mu}\frac{d\mu}{d\lambda}, \qquad \lambda\text{-a.e.}$$

Proof. By considering ν^+ and ν^- separately, we may assume ν is positive. By definition

$$\nu(A) = \int_A \frac{d\nu}{d\mu}\, d\mu = \int \chi_A \cdot \frac{d\nu}{d\mu}\, d\mu \qquad A \in \mathcal{A}.$$

So Equation (4.16) holds for all functions χ_A with $A \in \mathcal{A}$ and $\nu(A) < \infty$. By linearity of the Lebesgue integral, it holds for all positive simple functions in $L^1(\nu)$. By Beppo Levi's Th 2.15 and Th 2.11, it holds for all non-negative functions in $L^1(\nu)$. By subtraction of positive and negative part and linearity of integral, (4.16) holds for all $g \in L^1(\nu)$. Finally,

$$\nu(E) = \int_E \frac{d\nu}{d\mu}\, d\mu = \int_E \frac{d\nu}{d\mu} \cdot \frac{d\mu}{d\lambda}\, d\lambda \qquad \forall E \in \mathcal{A}$$

by applying Equation (4.16) to $\mu \ll \lambda$ setting $g = \chi_E \cdot (d\nu/d\mu)$. □

If $\mu \sim \nu$ and μ, ν are positive then

$$\left(\frac{d\mu}{d\nu}\right) \cdot \left(\frac{d\nu}{d\mu}\right) = 1 \text{ a.e. on } X.$$

On $\mathbb{R}^n$, the Radon-Nikodym derivatives can be calculated explicitly.

Theorem 4.17 (Lebesgue Differentiation Theorem)**.** Let $\mu =$ Lebesgue measure on $\mathbb{R}^n$, let ν be a signed Radon measure on $\mathbb{R}^n$ with $d\nu = d\lambda + f\ d\mu$ its Lebesgue decomposition. Then

$$f(x) = \lim_{r \to 0} \left[\frac{\nu(B_r(x))}{\mu(B_r(x))} \right] \qquad \mu\text{-a.e.}$$

where $B_r(x) := \{y \in \mathbb{R}^n : \|x - y\| < r\}$. In particular, if $\nu \ll \mu$, then

$$\frac{d\nu}{d\mu}(x) = \lim_{r \to 0} \frac{\nu(B_r(x))}{\mu(B_r(x))} \qquad \mu\text{-a.e.}.$$

The proof is very technical. See Real Analysis, Modern Techniques and Their applications, Second Edition (1999) - Gerald B. Folland pp 95–98.

Chapter 5

Conditional Probabilities

5.1 Random Elements

Definition 5.1. (1) A probability space is a measure space $(\Omega, \mathcal{F}, \mathbf{P})$ with $\mathbf{P}(\Omega) = 1$.

(2) Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and let $(E, \mathcal{E})$ be a measurable space. A measurable function $X : \Omega \to E$ is called random element with values in E. If $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then X is called a random variable; if $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, then X is called a random vector.

(3) We will write $X \in B$ for the event $\{\omega \in \Omega : X(\omega) \in B\}$.

(4) The probability measure[1] $\mathbf{P}_X$ on $(E, \mathcal{E})$ given by

$$\mathbf{P}_X(A) = \mathbf{P}(X^{-1}(A))$$

is called the distribution or probability law of the random element X. The measure space $(E, \mathcal{E}, \mathbf{P}_X)$ is said to be the canonical probability space for X.

Recall that $\mathcal{B}(\mathbb{R}^d)$ is defined as $\sigma(\mathcal{C}^d)$ where $\mathcal{C}^d$ are the d-dimensional blocks $\times_{k=1}^d (a_k, b_k] := (a_1, b_1] \times \ldots \times (a_d, b_d]$. We define the d-dimensional product of Borel-σ-algebras as

$$\bigotimes_{k=1}^d \mathcal{B}(\mathbb{R}) := \mathcal{B}(\mathbb{R}) \otimes \ldots \otimes \mathcal{B}(\mathbb{R}) := \sigma(\mathcal{B}(\mathbb{R}) \times \ldots \times \mathcal{B}(\mathbb{R}))$$

Just as in Ex 3.2, it can be shown that $\bigotimes_{k=1}^d \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^d)$.

[1] Check that this is indeed a probability measure!

It follows that if X_k are random variables for $k = 1, \ldots d$, then $X := (X_1, \ldots, X_d)$ is a random vector; to show this, by Th 2.3 it suffices to check that $X^{-1}(\times_{k=1}^{d}(a_k, b_k]) \in \mathcal{F}$. Conversely, if $X = (X_1, \ldots, X_d)$ is a random vector, then each X_k is a random variable, since $X_k = \pi_k \circ X$ and the projections $\pi_k : \mathbb{R}^d \to \mathbb{R}$ given by $(x_1, \ldots, x_d) \mapsto x_k$ are measurable.

Definition 5.2. Let X be a random variable. The expectation of X is the Lebesgue integral $\mathbf{E}(X) = \int X \, d\mathbf{P}$. If X is a random vector, then the expectation of X is the vector $\mathbf{E}(X) = [\mathbf{E}(X_1), \ldots, \mathbf{E}(X_d)]$.

If two random variables X and Y are equal $\mathbf{P}$-a.e., we say they are equal $\mathbf{P}$-almost surely ($\mathbf{P}$-a.s., or just a.s.). Recall that a random variable X lies in $L^1(\mathbf{P})$ if $\mathbf{E}(X_+) < \infty$ and $\mathbf{E}(X_-) < \infty$, hence if $\mathbf{E}(|X|) = \mathbf{E}(X_+) + \mathbf{E}(X_-) < \infty$. Almost sure equality defines an equivalence relation on $L^1(\Omega)$. This means that $L^1(\Omega)$ is the disjoint union of equivalence classes, where within each equivalence class all functions are pairwise equal a.s. Below, the conditional expectation and conditional probability with respect to a σ-algebra is defined as an $L^1(\Omega)$ random variable; it is only unique up to this sort of equivalence.

The following result is for later use:

Lemma 5.3 (Change of Variable in Lebesgue integral)**.** Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and X a random element with values in $(E, \mathcal{E})$. Then for any $f : E \to \mathbb{R}$ which is in $L^1(\mathbf{P}_X)$, $f \circ X$ is in $L^1(\mathbf{P})$, and

$$\int_B f \, d\mathbf{P}_X = \int_{X^{-1}(B)} f \circ X \, d\mathbf{P}.$$

Proof. Algebraic induction on f. $\square$

In particular, this means that expectations can be calculated as

$$\mathbf{E}(X) = \int_{\mathbb{R}} \mathrm{id}_X \, \mathbf{P}_X = \int_{\mathbb{R}} x \, \mathbf{P}_X(dx)$$

where $\mathrm{id}_X : \mathbb{R} \to \mathbb{R}$, $x \mapsto x$.

5.2 Independence

Definition 5.4 (Independence of Events)**.** Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space. A family of events $\{A_n\}_{n \in I} \subset \mathcal{F}$ is independent if for any <u>finite</u> subset $J \subset I$

$$\mathbf{P}\left(\bigcap_{j \in J} A_j\right) = \prod_{j \in J} \mathbf{P}(A_j). \tag{5.5}$$

(Recall that sets have distinct elements, e.g. $\{1, 2, 2\}$ is not a set. Set inclusion is not strict, i.e. $A \subset A$.)

Proposition 5.6. Let $\{A_n\}_{n\in\mathcal{I}}$ be a family of independent events on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Define $A_n^0 := A_n$ and $A_n^1 := A_n^\complement$, and let $\{e_n\}_{n\in\mathcal{I}} \subset \{0,1\}$. Then the family $\{A_n^{e_n}\}_{n\in\mathcal{I}}$ is independent.

Proof. Pick any $s \in \mathcal{I}$ such that $e_s = 1$. We show that the family which results from replacing A_s by $A_s^\complement$ is independent. Let $J \subset \mathcal{I}$ be a finite subset. If $s \notin J$, then (5.5) holds. If $s \in J$, then

$$\mathbf{P}\left(A_s^\complement \cap \bigcap_{j\in J,\, j\neq s} A_j\right) = \mathbf{P}\left(\bigcap_{j\in J,\, j\neq s} A_j \setminus \bigcap_{j\in J} A_j\right) = \mathbf{P}\left(\bigcap_{j\in J,\, j\neq s} A_j\right) - \mathbf{P}\left(\bigcap_{j\in J} A_j\right)$$
$$= \prod_{j\in J,\, j\neq s} \mathbf{P}(A_j) - \prod_{j\in J} \mathbf{P}(A_j)\mathbf{P}(A_j) = \mathbf{P}(A_s^\complement) \prod_{j\in J,\, j\neq s} \mathbf{P}(A_j)$$

and (5.5) still holds.

Consider now the family $\{A_n^{e_n}\}_{n\in\mathcal{I}}$ and pick any finite subset $J \subset \mathcal{I}$. If (5.5) is satisfied for $\{A_n^{e_n}\}_{n\in\mathcal{I}}$, we are done. But by proceeding inductively as above, we see that the family $\{A_n^{e_n}\}_{n\in J}$ is independent. Hence for this family, (5.5) holds. As the equation does not involve any terms indexed by $\mathcal{I}\setminus J$, it holds also for $\{A_n^{e_n}\}_{n\in\mathcal{I}}$. □

Lemma 5.7 (Borel-Cantelli Lemma)**.** Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and let $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$.

(1) If $\sum_n \mathbf{P}(A_n) < \infty$, then $\mathbf{P}(\limsup_n A_n) = 0$.

(2) Assume that $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ is a sequence of independent events. Then $\sum_n \mathbf{P}(A_n) = \infty$ implies $\mathbf{P}(\limsup_n A_n) = 1$.

Proof. Let $B_n = \bigcup_{j=n}^\infty A_j$. Then $B_{n+1} \subset B_n$ and

$$\limsup_n A_n = \bigcap_{n=1}^\infty \bigcup_{j=n}^\infty A_j = \bigcap_{n=1}^\infty B_n.$$

The subadditivity of the measure $\mathbf{P}$ gives $\mathbf{P}(B_n) \leqslant \sum_{j=n}^\infty \mathbf{P}(A_j)$ and so

$$\mathbf{P}(\limsup_n A_n) = \mathbf{P}\left(\bigcap_{n=1}^\infty B_n\right) = \lim_n \mathbf{P}(B_n) \leqslant \lim_n \sum_{j=n}^\infty \mathbf{P}(A_j)$$
$$= \lim_n \left(\sum_{j=1}^\infty \mathbf{P}(A_j) - \sum_{j=1}^{n-1} \mathbf{P}(A_j)\right) = 0$$

which proves (1). To prove (2), note that by the De Morgan Laws we may equivalently show

$$0 = \mathbf{P}\left(\bigcup_{n=1}^{\infty}\bigcap_{j=n}^{\infty} A_j^{\complement}\right) = \mathbf{P}(\liminf_n A_n^{\complement}).$$

It suffices to show that $C_n := \bigcap_{j=n}^{\infty} A_j^{\complement}$ is a null set for every n. Using independence and the inequality $1 + x \leqslant e^x$, we obtain

$$\begin{aligned}\mathbf{P}(C_n) = \mathbf{P}\left(\bigcap_{N=n}^{\infty}\bigcap_{j=n}^{N} A_j^{\complement}\right) &= \lim_N \mathbf{P}\left(\bigcap_{j=n}^{N} A_j^{\complement}\right) = \lim_N \prod_{j=n}^{N} \mathbf{P}(A_j^{\complement}) \\ &\leqslant \lim_N \prod_{j=n}^{N} \exp(-\mathbf{P}(A_j)) = \lim_N \exp\left(-\sum_{j=n}^{N} \mathbf{P}(A_j)\right) = 0.\end{aligned}$$

□

Definition 5.8 (Independence of families of events)**.** Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, let I be a set, and for each $i \in I$ let $\mathcal{G}_i \subset \mathcal{F}$ be a family of events. Then the family of families of sets $\{\mathcal{G}_i\}_{i\in I}$ is independent if any family of events $\{A_i\}_{i\in I}$ which satisfies $A_i \in \mathcal{G}_i$ is independent.

Lemma 5.9. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, let $\mathcal{C} \subset \mathcal{F}$ be a π-system, and let $\mathcal{E} \subset \mathcal{F}$ be another family of events. Then $\sigma(\mathcal{C})$ and $\mathcal{E}$ are independent iff $\mathcal{C}$ and $\mathcal{E}$ are independent.

Proof. Apply the Monotone Class Theorem 1.23: Let

$$\mathcal{N} = \{A \in \sigma(\mathcal{C}) : \mathbf{P}(A \cap E) = \mathbf{P}(A)\mathbf{P}(E)\, \forall E \in \mathcal{E}\}.$$

Then $\mathcal{C} \subset \mathcal{N}$. We show that $\mathcal{N}$ is a d-system, then $\sigma(C) = d(\mathcal{C}) \subset d(\mathcal{N}) = \mathcal{N}$ and we are done. By Pr 5.6, $\mathcal{N}$ is closed under complements. Let $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{N}$ be such that $A_n \subset A_{n+1}$ for all n. Then

$$\begin{aligned}\mathbf{P}\left(\left(\bigcup_n A_n\right) \cap E\right) &= \mathbf{P}\left(\bigcup_n A_n \cap E\right) \\ &= \lim_n \mathbf{P}(A_n \cap E) = \lim_n \mathbf{P}(A_n)\mathbf{P}(E) = \mathbf{P}\left(\bigcup_n A_n\right)\mathbf{P}(E)\end{aligned}$$

and hence $\bigcup_n A_n \in \mathcal{N}$, and $\mathcal{N}$ is a d-system. □

Definition 5.10. Let X be a random element on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with values in $(E, \mathcal{E})$. The σ-algebra generated by X is

$$\sigma(X) := \{X^{-1}(B) : B \in \mathcal{E}\}.$$

It is the smallest σ-algebra on Ω which renders X measurable. (Check that it really is a σ-algebra.)

In probability theory, σ-algebras $\mathcal{F}$ are interpreted as collections of events $F \in \mathcal{F}$. In a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, $\mathcal{F}$ is the collection of *all* events. In contrast, the sub-σ-algebra $\sigma(X)$ may be interpreted as the smaller collection of events *which may be expressed in terms of X.*

Definition 5.11 (Independence of Random Elements)**.** A family of random elements $\{X_i\}_{i\in I}$ on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is independent if the family of σ-algebras $\{\sigma(X_i)\}_{i\in I}$ is independent.

Lemma 5.12. Let $X_1, \ldots, X_N$ be random elements with value in some measurable spaces $(E_1, \mathcal{E}_1), \ldots, (E_N, \mathcal{E}_N)$ and distributions $\mathbf{P}_{X_1}, \ldots, \mathbf{P}_{X_N}$. Then the X_n are independent iff $X := (X_1, \ldots, X_N)$ has distribution $\mathbf{P}_X = \mathbf{P}_{X_1} \otimes \ldots \otimes \mathbf{P}_{X_N}$.

Proof. If the X_n are independent,

$$\begin{aligned}
&\mathbf{P}_X(B_1 \times \ldots \times B_N) = \mathbf{P}(X^{-1}(B_1 \times \ldots \times B_N)) \\
&= \mathbf{P}(\{\omega \in \Omega : X_1 \in B_1, \ldots, X_N \in B_N\}) = \mathbf{P}\left(\bigcap_{n=1}^{N} \{X_n \in B_n\}\right) \\
&= \prod_{n=1}^{N} \mathbf{P}(X_n \in B_n) = \prod_{n=1}^{N} \mathbf{P}_{X_n}(B_n) = \mathbf{P}_{X_1} \otimes \ldots \otimes \mathbf{P}_{X_N}(B_1 \times \ldots \times B_N).
\end{aligned}$$

Since sets of the form $B_1 \times \ldots \times B_N$ generate $\mathcal{E}_1 \otimes \ldots \otimes \mathcal{E}_N$, this shows $\mathbf{P}_X = \mathbf{P}_{X_1} \otimes \ldots \otimes \mathbf{P}_{X_N}$.

On the other hand, assume $\mathbf{P}_X = \mathbf{P}_{X_1} \otimes \ldots \otimes \mathbf{P}_{X_N}$, and let $A_n \in \sigma(X_n)$ for $n = 1, \ldots, N$. By definition of $\sigma(X_n)$, there are $B_n \in \mathcal{E}_n$ such that $A_n = X_n^{-1}(B_n)$. Then

$$\begin{aligned}
&\mathbf{P}\left(\bigcap_{n=1}^{N} A_n\right) = \mathbf{P}\left(\bigcap_{n=1}^{N} X_n^{-1}(B_n)\right) = \mathbf{P}(\{\omega \in \Omega : X_n(\omega) \in B_n, n = 1, \ldots, N\}) \\
&= \mathbf{P}(X \in B_1 \times \ldots \times B_N) = \mathbf{P}_X(B_1 \times \ldots \times B_N) \\
&= \mathbf{P}_{X_1} \otimes \ldots \otimes \mathbf{P}_{X_N}(B_1 \times \ldots \times B_N) = \prod_{n=1}^{N} \mathbf{P}_{X_n}(B_n) = \prod_{n=1}^{N} \mathbf{P}(A_n),
\end{aligned}$$

which shows independence of $X_1, \ldots, X_N$. $\square$

Lemma 5.13 (Conditioning)**.** Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and X and Y independent random variables with values in some measurable spaces $(E_1, \mathcal{E}_1)$ and $(E_2, \mathcal{E}_2)$. Let $f : E_1 \times E_2 \to \mathbb{R}$ be measurable such that $\mathbf{E}(|f(X, Y)|) < \infty$. Then

$$\mathbf{E}(f(X, Y)) = \mathbf{E}(\mathbf{E}(f^y(X))|_{y=Y}).$$

In particular, for any independent random variables $X_1, \dots, X_N$,

$$\mathbf{E}\left(\prod_{n=1}^{N} X_n\right) = \prod_{n=1}^{N} \mathbf{E}(X_n).$$

Proof. Using Lem 5.12, and Fubini's Theorem,

$$\begin{aligned}\mathbf{E}(f(X, Y)) &= \int f \circ (X, Y)\, d\mathbf{P} = \int f\, d\mathbf{P}_{(X,Y)} = \int f\, d(\mathbf{P}_X \otimes \mathbf{P}_Y) \\ &= \int_{E_2} \int_{E_1} f(x, y)\mathbf{P}_X(dx)\mathbf{P}_Y(dy) = \int_{E_2} \int_{E_1} f^y(x)\mathbf{P}_X(dx)\mathbf{P}_Y(dy) \\ &= \int_{E_2} \mathbf{E}(f^y(X))\mathbf{P}_Y(dy) = \mathbf{E}(\mathbf{E}(f^y(X))|_{y=Y}).\end{aligned}$$

Set $f(x_1, \dots, x_N) = \prod_{n=1}^{N} x_n$ and use induction on N to get the remaining statement. ☐

5.3 Conditional Expectation

Definition 5.14. Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, and let $\mathcal{A} \subset \mathcal{F}$ be a sub-σ-algebra. Any $\mathcal{A}$-measurable random variable Y which satisfies

$$\int_A Y\, d\mathbf{P} = \int_A X\, d\mathbf{P}, \quad A \in \mathcal{A}$$

is called the conditional expectation of X with respect to $\mathcal{A}$. We write $Y = E(X|\mathcal{A})$.

Theorem 5.15. For any $X \in L^1(\mathbf{P})$ and $\mathcal{A} \subset \mathcal{F}$, a conditional expectation $E(X|\mathcal{A})$ exists, and any two conditional expectations Y and Z given $\mathcal{A}$ are equal a.s.

Proof. Let $\mu(A) := \int_A X\, d\mathbf{P}$ for all $A \in \mathcal{A}$. Then μ is a measure by Th 2.18(4). If $\mathbf{P}(N) = 0$, then $\mu(N) = \int \chi_N X\, d\mathbf{P} = 0$ by Th 2.18(3), since $\chi_N X = 0$ a.s. So $\mu \ll \mathbf{P}$. By Radon-Nikodym Th 4.11, there is $Y \in L^1(\mathbf{P})$ such that

$\mu(A) = \int_A Y\, d\mathbf{P}$ for all $A \in \mathcal{A}$. Such Y satisfies the definition of conditional expectation. If Z also satisfies it, then

$$\int_A Y - Z\, d\mathbf{P} = 0, \quad A \in \mathcal{A}. \tag{5.16}$$

Since Y and Z are $\mathcal{A}$-measurable, by Prop 2.20(5) this means $Y = Z$ $\mathbf{P}$-a.s. $\square$

Theorem 5.17 (Properties of Conditional Expectation)**.** Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, $X, Y \in L^1(\mathbf{P})$, $c \in \mathbb{R}$ and let $\mathcal{C} \subset \mathcal{A} \subset \mathcal{F}$ be sub-σ-algebras. The following properties hold:

(1) $\mathbf{E}(cX + Y|\mathcal{A}) = c\mathbf{E}(X|\mathcal{A}) + \mathbf{E}(Y|\mathcal{A})$

(2) $\mathbf{E}(X|\mathcal{C}) = \mathbf{E}(\mathbf{E}(X|\mathcal{A})|\mathcal{C})$.

(3) $\mathbf{E}(X|\{\varnothing, \Omega\}) = \mathbf{E}(X)$

(4) $\mathbf{E}(X|\mathcal{F}) = X$

(5) If $X \leqslant Y$ a.s., then $\mathbf{E}(X|\mathcal{A}) \leqslant \mathbf{E}(Y|\mathcal{A})$ a.s.

(6) $|\mathbf{E}(X|\mathcal{A})| \leqslant \mathbf{E}(|X||\mathcal{A})$ a.s. In particular, $\mathbf{E}(X|\mathcal{A}) \in L^1(\mathbf{P})$.

(7) If $X_n \to X$ a.s., $|X_n| \leqslant Y$ and $Y \in L^1(\mathbf{P})$, then $\mathbf{E}(X_n|\mathcal{A}) \to \mathbf{E}(X|\mathcal{A})$ and $\mathbf{E}(|X_n - X||\mathcal{A}) \to 0$.

(8) If $XY \in L^1(\mathbf{P})$ and Y is $\mathcal{A}$-measurable, then $\mathbf{E}(XY|\mathcal{A}) = Y\mathbf{E}(X|\mathcal{A})$.

(9) If $\sigma(X)$ and $\mathcal{A}$ are independent, then $\mathbf{E}(X|\mathcal{A}) = \mathbf{E}(X)$.

Proof. (5) By definition of conditional expectation and Prop 2.20(4) we have for every $A \in \mathcal{A}$

$$\int_A \mathbf{E}(X|\mathcal{A})\, d\mathbf{P} = \int_A X\, d\mathbf{P} \leqslant \int_A Y\, d\mathbf{P} = \int_A \mathbf{E}(Y|\mathcal{A})\, d\mathbf{P}.$$

Since $\mathbf{E}(X|\mathcal{A})$ and $\mathbf{E}(Y|\mathcal{A})$ are $\mathcal{A}$-measurable, the statement follows from Prop 2.20(4).
(6) follows from (5) with $-|X| \leqslant X \leqslant |X|$.
(7) Let $Z_n = \sup_{k \geqslant n} |X_k - X|$. Then $Z_n \downarrow 0$ a.s., and by (5) and (6)

$$|\mathbf{E}(X_n|\mathcal{A}) - \mathbf{E}(X|\mathcal{A})| = |\mathbf{E}(X_n - X|\mathcal{A})| \leqslant \mathbf{E}(|X_n - X||\mathcal{A}) \leqslant \mathbf{E}(Z_n|\mathcal{A}).$$

Since $\mathbf{E}(Z_{n+1}|\mathcal{A}) \leqslant \mathbf{E}(Z_n|\mathcal{A})$ a.s., the limit $h = \lim_n \mathbf{E}(Z_n|\mathcal{A})$ exists a.s. Then

$$0 \leqslant \int h\, d\mathbf{P} \leqslant \int \mathbf{E}(Z_n|\mathcal{A})\, d\mathbf{P} = \int Z_n\, d\mathbf{P} \to 0, \quad n \to \infty,$$

where the last statement follows from the DCT Th **??**, since $0 \leqslant Z_n \leqslant 2Y$ and $\mathbf{E}(Y) < \infty$. Consequently $\int h \, d\mathbf{P} = 0$ and thus $h = 0$ a.s. by Th 2.18(3).
(8) Algebraic induction: First, let $Y = \chi_B$ where $B \in \mathcal{A}$. Then for every $A \in \mathcal{A}$

$$\int_A XY \, d\mathbf{P} = \int_{A \cap B} X \, d\mathbf{P} = \int_{A \cap B} \mathbf{E}(X|\mathcal{A}) \, d\mathbf{P} = \int_A Y\mathbf{E}(X|\mathcal{A}) \, d\mathbf{P}.$$

By part (1), the statement remains valid if Y is a simple function $\sum_{k=1}^n y_k \chi_{B_k}$ where $B_k \in \mathcal{A}$. Now let Y be any $\mathcal{A}$-measurable non-negative random variable with $\mathbf{E}(|Y|) < \infty$, and let Y_n be a sequence of simple functions such that $Y_n \uparrow Y$. By what has been shown, $\mathbf{E}(XY_n|\mathcal{A}) = Y_n\mathbf{E}(X|\mathcal{A})$. As $|XY_n| \leqslant |XY| < \infty$ and $XY_n \to XY$, by part (7) $\mathbf{E}(XY_n|\mathcal{A}) \to \mathbf{E}(XY|\mathcal{A})$. Then by (6) $\mathbf{E}(X|\mathcal{A}) \in L^1(\mathbf{P})$ we have $\mathbf{E}(X|\mathcal{A}) \neq \infty$ a.s., and so $Y_n\mathbf{E}(X|\mathcal{A}) \to Y\mathbf{E}(X|\mathcal{A})$. Finally, for general $Y \in L^1(\mathbf{P})$, we have

$$\begin{aligned} \mathbf{E}(XY|\mathcal{A}) = \mathbf{E}(X(Y_+ + Y_-)|\mathcal{A}) &= \mathbf{E}(XY_+|\mathcal{A}) + \mathbf{E}(XY_-|\mathcal{A}) \\ &= Y_+\mathbf{E}(X|\mathcal{A}) + Y_-\mathbf{E}(X|\mathcal{A}) = Y\mathbf{E}(X|\mathcal{A}). \end{aligned}$$

(9) Algebraic induction on X. In the last step, use (7). □

Our next aim is to define a conditional expectation of a random variable X, given the value of another random element Y.

Theorem 5.18. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, X a random element with values in $(E, \mathcal{E})$, and Y a random variable. If Y is $\sigma(X)$-measurable, then there is a Borel function $f : E \to \mathbb{R}$ such that $Y = f \circ X$ (draw a diagram).

Proof. Define the following two classes of random variables:

$$\begin{aligned} \Phi^X &= \{\text{all } \sigma(X)\text{-measurable random variables}\}, \\ \tilde{\Phi}^X &= \{\text{all } \sigma(X)\text{-measurable random variables of the form } Y = f \circ X\} \end{aligned}$$

Clearly, $\tilde{\Phi}^X \subset \Phi^X$; the statement of the theorem is $\tilde{\Phi}^X \supset \Phi^X$. Again, proof by algebraic induction. Assume first that $Y \in \Phi^X$ is an indicator function χ_A where $A \in \sigma(X)$. By definition of $\sigma(X)$, there is a set $B \in \mathcal{E}$ such that $A = X^{-1}(B)$. But then $\chi_A = \chi_B \circ X \in \tilde{\Phi}^X$. Next, let $Y = \sum_{k=1}^n y_k \chi_{A_k}$. Then let B_k satisfy $X^{-1}(B_k) = A_k$, and Y has the representation

$$Y = \sum_{k=1}^n y_k (\chi_{B_k} \circ X) = \left(\sum_{k=1}^n y_k \chi_{B_k} \right) \circ X \in \tilde{\Phi}^X$$

Finally, if Y is any random variable defined on Ω, then there are sequences of simple functions $g_n^+ \uparrow Y_+$ and $g_n^- \uparrow Y_-$, and hence the sequence $g_n := g_n^+ + g_n^-$

converges to Y. By what has been shown, there is a sequence of measurable f_n such that $g_n = f_n \circ X$. Now define $B = \{x \in E : \lim_n f_n(x) \text{ exists}\}$. Since $B = \{x \in E : \limsup_n f_n(x) = \liminf_n f_n(x)\}$, we see that $B \in \mathcal{E}$. Then let

$$f(x) = \begin{cases} \lim_n f_n(x) & x \in B \\ 0 & x \notin B \end{cases}$$

and see that $Y(\omega) = \lim_n g_n(\omega) = \lim_n f_n(X(\omega))$. Hence $X(\omega) \in B$ for every $\omega \in \Omega$, and $f \circ X = Y$. □

Corollary 5.19. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, X a random element with values in $(E, \mathcal{E})$, and Y a random variable. Then there exists a measurable function $g : E \to \mathbb{R}$ such that

$$\mathbf{E}(Y|X) := \mathbf{E}(Y|\sigma(X)) = g \circ X.$$

We will write $g(x) = \mathbf{E}(Y|X = x)$ for all $x \in E$, which may be interpreted as the conditional expectation of Y given that $X = x$.

Definition 5.20 (Conditional Distribution)**.** Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, $\mathcal{A} \subset \mathcal{F}$ a sub-σ-algebra, and $F \in \mathcal{F}$.

(1) The conditional probability of $F \in \mathcal{F}$ given $\mathcal{A}$ is the $\mathcal{A}$-measurable random variable

$$\mathbf{P}(F|\mathcal{A}) := \mathbf{E}(\chi_F|\mathcal{A}).$$

(2) The conditional probability of $F \in \mathcal{F}$ given that $X = x$ is the measurable function

$$\mathbf{P}(F|X = x) := g(x)$$

where $g : E \to \mathbb{R}$ satisfies $g \circ X = \mathbf{P}(F|\mathcal{A})$ (see Cor 5.19).

(3) A regular conditional probability given $\mathcal{A}$ is a function

$$\mathcal{F} \times \Omega \ni (F, \omega) \mapsto \mathbf{P}(F|\mathcal{A})(\omega) \in [0, 1]$$

denoted by $\mathbf{P}(\cdot|\mathcal{A})(\cdot)$ or $\mathbf{P}(\cdot|\mathcal{A})$, which satisfies the following conditions:

(a) for each $F \in \mathcal{F}$, the mapping $\omega \mapsto \mathbf{P}(F|\mathcal{A})(\omega)$ is the conditional probability of F given $\mathcal{A}$ as defined in (1)

(b) for $\mathbf{P}$-a.e. $\omega \in \Omega$, the mapping $F \mapsto \mathbf{P}(F|\mathcal{A})(\omega)$ is a probability measure on $\mathcal{F}$.

(4) A regular conditional probability given X is a function

$$\mathcal{F} \times E \ni (F, x) \mapsto \mathbf{P}(F|X = x) \in [0, 1]$$

denoted by $\mathbf{P}(\cdot|X = \cdot)$, which satisfies the following conditions:

(a) for each fixed $F \in \mathcal{F}$, the mapping $x \mapsto \mathbf{P}(F|X = x)$ is as in (2),

(b) for $\mathbf{P}_X$-a.e. $x \in E$, the mapping $F \mapsto \mathbf{P}(F|X = x)$ is a probability measure on $\mathcal{F}$.

(5) Let Y be a random element with values in $(G, \mathcal{G})$. A regular conditional distribution of Y given $\mathcal{A}$ is a function

$$\mathcal{G} \times \Omega \ni (C, \omega) \mapsto \mathbf{P}_{Y|\mathcal{A}}(C, \omega) \in [0, 1]$$

which satisfies $\mathbf{P}_{Y|\mathcal{A}}(C, \cdot) = \mathbf{P}(Y^{-1}(C)|\mathcal{A})(\cdot)$ $\mathbf{P}$-a.s. where $\mathbf{P}(\cdot|\mathcal{A})(\cdot)$ is as in (3).

(6) A regular conditional distribution of Y given X is a function

$$\mathcal{G} \times E \ni (C, x) \mapsto \mathbf{P}_{Y|X}(C|x) \in [0, 1]$$

which satisfies $\mathbf{P}_{Y|X}(C|\cdot) = \mathbf{P}(Y^{-1}(C)|X = \cdot)$ where $\mathbf{P}(\cdot|X = \cdot)$ is as in (4).

Exercise 5.21. In the setting of the above definition, show the following statements:

(1) $\mathbf{P}(F|\mathcal{A}) \in [0, 1]$, $\mathbf{P}$-a.s.

(2) For every $B \in \mathcal{E}$ and $C \in \mathcal{G}$,

$$\mathbf{P}(X \in B, Y \in C) := \mathbf{P}(X^{-1}(B) \cap Y^{-1}(C)) = \int_B \mathbf{P}(Y \in C|X = x)\mathbf{P}_X(dx) \tag{5.22}$$

(3) Let $C = \bigcup_n C_n \in \mathcal{G}$ be a countably infinite disjoint union. Then

$$\mathbf{P}\left(Y \in \bigcup_n C_n|X = x\right) = \sum_n \mathbf{P}(Y \in C_n|X = x)$$

holds for $\mathbf{P}_X$-a.e. $x \in E$.

Technical issues. Note that statement (3) above together with $\mathbf{P}(Y \in \emptyset | X = x) = 0$ $\mathbf{P}_X$-a.s. do *not* mean that the mapping

$$C \mapsto \mathbf{P}(Y \in C | X = x)$$

defines a probability measure on $(G, \mathcal{G})$, for $\mathbf{P}_X$-a.e. x. The $\mathbf{P}_X$ null set in which the statement may not hold depends on the set C and even on the sequence C_n. Simply excluding these sets does not work, as there are uncountably many of them, and their union could be as big as the whole space. In the simple setting of discrete and continuous random elements, these issues do not cause a problem:

Conditioning on discrete X. Let X be discrete, that is $X : \Omega \to E$ where $\mathbf{P}(X \in E_1) = 1$ for some countable set $E_1 = \{x_1, x_2, \dots\} \subset E$. Then for every x_n,

$$\mathbf{P}(F | X = x_n) = \frac{\mathbf{P}(F \cap \{X = x_n\})}{\mathbf{P}(X = x_n)}$$

defines a regular conditional probability on $(\Omega, \mathcal{F})$ given X. In particular, if Y is another discrete random element and $\mathbf{P}_{(X,Y)}(x_n, y_m) = \mathbf{P}(X = x_n, Y = y_m)$ is the distribution of (X, Y), then

$$\mathbf{P}(Y = y_m | X = x_n) = \frac{\mathbf{P}_{(X,Y)}(x_n, y_m)}{\mathbf{P}_X(x_n)}$$

defines a regular conditional distribution of Y given X.

Proof. Verify

$$\frac{\mathbf{P}(F \cap \{X = \cdot\})}{\mathbf{P}(X = \cdot)} \circ X = \mathbf{P}(F | X)$$

and replace F by $\{Y = y_m\}$. □

Conditioning continuous Y on continuous X. Suppose that X and Y are random variables whose joint law is absolutely continuous wrt Lebesgue measure "$dx\, dy$", with joint density (i.e. Radon-Nikodym derivative)

$$\mathbf{P}_{(X,Y)}(dx, dy) = f(x, y)\, dx\, dy.$$

Then the distribution $\mathbf{P}_X$ of X is absolutely continuous wrt Lebesgue measure with density

$$f_X(x) = \int_{\mathbb{R}} f(x, y)\, dy$$

and

$$\mathbf{P}_{Y|X}(B|x) := \mathbf{P}(Y \in B|X = x) = \int_B \frac{f(x,y)}{f_X(x)}\,dy$$

defines a regular conditional probability on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ given X. For Lebesgue-almost every $x \in \mathbb{R}$, the measure $B \mapsto \mathbf{P}_{Y|X}(B|x)$ is absolutely continuous wrt Lebesgue measure on $\mathbb{R}$, with Radon-Nikodym derivative (density)

$$y \mapsto f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}.$$

Proof. Verify

$$\mathbf{P}(Y \in B|X) = \int_B \frac{f(X,y)}{f_X(X)}\,dy$$

□

Conditioning discrete Y on continuous X (also see "mixtures.")

Suppose that X is continuous and Y discrete, and write $\pi_n := \mathbf{P}(Y = y_n)$ for $n \in \mathbb{N}$. Further, suppose that

$$\mathbf{P}(X \in B|Y = y_n) = \int_B f_n(x)\,dx$$

for a sequence of probability densities f_n. (The distributions f_n are mixed, with mixing probabilities π_n.) Then the joint distribution of X and Y is defined by

$$\mathbf{P}_{(X,Y)}(B,C) = \sum_{y_n \in C} \pi_n \int_B f_n(x)\,dx,$$

and

$$\mathbf{P}(Y = y_n|X = x) = \frac{\pi_n f_n(x)}{\sum_m \pi_m f_m(x)}$$

defines a regular conditional distribution of Y given x.

Proof. Verify

$$\mathbf{P}(Y = y_n|X) = \frac{\pi_n f_n(X)}{\sum_m \pi_m f_m(X)}.$$

□

Conditional probabilities do not exist in the most general setting, and we need more assumptions to guarantee existence.

Definition 5.23. A topological space $(S, \mathcal{T})$ is called a Polish space if

(1) there exists a metric $d : S \times S \to [0, \infty)$ such that $\mathcal{T}$ is generated by all open balls $B_x(\varepsilon) := \{y \in S : d(x, y) < \varepsilon\}$ ($\mathcal{T}$ is "metrizable")

(2) the metric space (S, d) is complete, that is, for every sequence $\{x_n\}_{n\in\mathbb{N}} \subset S$ which satisfies $\forall\varepsilon > 0\, \exists N \in \mathbb{N} : m, n \geqslant N \Rightarrow d(x_n, x_m) < \varepsilon$, the limit $\lim_n x_n$ exists and lies in S

(3) the metric space (S, d) is separable, that is, there exists a sequence $\{x_n\}_{n\in\mathbb{N}} \subset S$ such that $\forall x\, \forall\varepsilon > 0\, \exists n : x_n \in B_x(\varepsilon)$.

Theorem 5.24. Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and $\mathcal{A} \subset \mathcal{F}$ be a sub-σ-algebra. Let $(S, \mathcal{T})$ any Polish space with Borel-σ-algebra $\mathcal{B} = \sigma(\mathcal{T})$, and $Y : \Omega \to S$ any random element with values in S. Then a regular conditional distribution of Y given $\mathcal{A}$,

$$\mathcal{B} \times \Omega \ni (B, \omega) \mapsto \mathbf{P}_{Y|\mathcal{C}}(B, \omega) \in [0, 1]$$

exists. It is unique in the sense that if $\mathbf{P}'_{Y|\mathcal{C}}$ is another conditional distribution of Y given $\mathcal{A}$, then for $\mathbf{P}$-almost all $\omega \in \Omega$, the probability measures $\mathbf{P}'_{Y|\mathcal{C}}(\cdot, \omega)$ and $\mathbf{P}_{Y|\mathcal{C}}(\cdot, \omega)$ are identical.

Proof. See e.g. Dudley, "Real Analysis and Probability," Ch 10.2. □

5.4 Disintegration of Probability Measures

Using the product measure, we can construct higher dimensional random vectors from a collection of (independent) random variables (see Lem 5.12). A construction in which the components of the higher dimensional random vector are not independent was hinted in the mixture example on page 58. In general, let X and Y denote random elements with values in $(E, \mathcal{E})$ and $(G, \mathcal{G})$. Let K be a probability kernel from E to F; that is, $K : E \times \mathcal{G} \to [0, 1]$ is such that $K(\cdot, C)$ is measurable for every $C \in \mathcal{G}$, and $K(x, \cdot)$ is a probability measure for ($\mathbf{P}_X$-almost) every $x \in E$. Then we may define a joint distribution of (X, Y) via

$$\mathbf{P}_{(X,Y)}(B \times C) = \int_B K(x, C)\, \mathbf{P}_X(dx). \tag{5.25}$$

As an example, let X be an exponentially distributed random variable with Lebesgue density $\lambda e^{-\lambda x}$, and let $K(x, C) = \int_C x^{-1}\chi_{(0,x)}(y)\, dy$ (uniform density on the interval $(0, x)$.) Then (5.25) defines a joint law of (X, Y) where X is exponentially distributed and the law of Y is uniform on $(0, X)$.

Can *all* higher-dimensional distributions $\mathbf{P}_{(X,Y)}$ of random elements (X, Y) be deconstructed into a distribution $\mathbf{P}_X$ of X (the "marginal") and a conditional distribution $\mathbf{P}_{Y|X}$ of Y given X? Surprisingly, yes, whenever a regular conditional distribution of Y given X exists:

Theorem 5.26. Let X and Y be two random elements on measurable spaces $(E, \mathcal{E})$ and $(G, \mathcal{G})$, such that a regular conditional distribution $\mathbf{P}_{Y|X} : \mathcal{G} \times E$ exists. Let f be a measurable function on $E \times G$ such that $\mathbf{E}|f(X, Y)| < \infty$. Then

$$\mathbf{E}(f(X, Y)|X) = \int f(X, y)\, d\mathbf{P}_{Y|X}(dy, X)$$

holds $\mathbf{P}$-a.s.

Proof. First, we show the equality

$$\mathbf{E}\left(\int f(X, y)\, d\mathbf{P}_{Y|X}(dy|X)\right) = \mathbf{E}(f(X, Y)). \tag{5.27}$$

For $f = \chi_{B\times C}$, we have

$$\begin{aligned}
&\mathbf{E}\left(\int \chi_B(x)\chi_C(y) d\mathbf{P}_{Y|X}(dy, X)\right) = \mathbf{E}(\chi_B(X)\mathbf{P}_{Y|X}(C|X)) \\
&= \int_B \mathbf{P}_{Y|X}(C, x)\, \mathbf{P}_X(dx) = \int_B \mathbf{P}(Y^{-1}(C)|X = x)\mathbf{P}_X(dx) \\
&= \int_{X^{-1}(B)} \mathbf{P}(Y^{-1}(C)|X)\, d\mathbf{P} = \int_{X^{-1}(B)} \chi_{Y^{-1}(C)}\, d\mathbf{P} \\
&= \mathbf{P}(X \in B, Y \in C) = \mathbf{E}(\chi_{B\times C}(X, Y))
\end{aligned}$$

and (5.27) holds. Next we show that (5.27) holds for arbitrary $\mathcal{E}\otimes\mathcal{G}$-measurable indicator functions. The class of functions

$$\mathcal{D} := \{D \in \mathcal{E} \otimes \mathcal{G} : \chi_D \text{ satisfies } (5.27)\}$$

is closed under complements (use $\chi_{D^c} = 1 - \chi_D$) and monotone unions $D = \bigcup_n D_n$ where $D_1 \subset D_2 \subset \ldots$ (use monotone convergence) and hence a d-system. As $\mathcal{E}\times\mathcal{G} \subset \mathcal{D}$ by what was shown above, $\mathcal{E}\otimes\mathcal{G} = \sigma(\mathcal{E}\times\mathcal{G}) \subset \sigma(\mathcal{D}) = d(\mathcal{D}) = \mathcal{D}$, (5.27) indeed holds for arbitrary indicator functions. By linearity of the integral, (5.27) holds for arbitrary simple functions. By monotone convergence, (5.27) holds for arbitrary measurable $f \geqslant 0$. Let $f \geqslant 0$ be such that $\mathbf{E}(f(X, Y)) < \infty$. Then we may replace $f(x, y)$ in (5.27) by $f(x, y)\chi_B(x)$ and the statement of the theorem follows for $f \geqslant 0$ measurable. By differencing $f = f^+ - f^-$, the general result follows for arbitrary f. $\square$

Chapter 6

Limits of Probability Measures

6.1 Characteristic Functions

Definition 6.1. Let $(S, \mathcal{T})$ be a topological space, and write $C_b(S)$ for all real valued bounded and continuous functions. A sequence of laws $\{\mathbf{P}_n\}_{n\in\mathbb{N}}$ converges (weakly)[1] to a law $\mathbf{P}$, written $\mathbf{P}_n \to \mathbf{P}$, if for every $f \in C_b(S)$, $\int f \, d\mathbf{P}_n \to \int f \, d\mathbf{P}$ as $n \to \infty$.

Example 6.2. Show that $\delta_{1/n} \to \delta_0$, but $\int f \, d\delta_{1/n} = \int f \, d\delta_0$ does not hold for all measurable f.

The following Lemma ensures that limits are unique:

Lemma 6.3. If (S, d) is a metric space, μ and ν are two laws on S, and $\int f \, d\mu = \int f \, d\nu$ for all $f \in C_b(S)$, then $\mu = \nu$.

Proof. See Dudley, 9.3.2. □

Below, we will need the Lebesgue integral of a complex-valued function. By relabelling the axes, $\mathbb{C}$ may be viewed as $\mathbb{R}^2$, giving an immediate representation of open sets in $\mathbb{C}$. We endow $\mathbb{C}$ with its Borel-σ-algebra.

For measurable $f : \mathbb{R}^d \to \mathbb{C}$, we can write $f(x) = u(x) + iv(x)$, where u and v are measurable and real-valued. Then we define

$$\int f \, d\mu := \int u \, d\mu + i \int v \, d\mu.$$

[1] This is actually weak* convergence in the Banach space $C_b(S)$.

Definition 6.4. Let X be a random vector with distribution μ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. The characteristic function of X (or of μ) is the function

$$\begin{aligned}\hat{\mu}(k) : \mathbb{R}^d &\to \mathbb{C} \\ k &\mapsto \int \exp(i\langle x, k\rangle)\, d\mathbf{P}_X(x) = \mathbf{E}(\exp(i\langle X, k\rangle))\end{aligned}$$

Here $\langle a, b\rangle := \sum_{k=1}^d a_k b_k$ denotes the usual inner product in $\mathbb{R}^d$. Note that $\hat{\mu}$ is well-defined since the measure space is finite and $|\exp(i\langle x, k\rangle)| \leqslant 1$ for every $k \in \mathbb{R}^d$.

Lemma 6.5. If μ is a probability measure on $\mathbb{R}^d$, then $\hat{\mu}$ is continuous at 0 with $\hat{\mu}(0) = 1$.

Proof. Both real and imaginary part of $e^{i\langle k,x\rangle}$ are dominated by 1. Apply the DCT. □

Lemma 6.6. If $\mu_n \to \mu$, then $\hat{\mu}_n(k) \to \hat{\mu}(k)$ for all $k \in \mathbb{R}^d$.

Proof. This follows since the real and imaginary parts of e^{ikx} are in $C_b(\mathbb{R})$. □

Definition 6.7. Given two probability measures μ and ν on $\mathbb{R}^d$, their convolution is defined as the measure

$$\mu \star \nu(B) = \int_{\mathbb{R}^d} \nu(B - x)\mu(dx)$$

where $B - x := \{b - x : b \in B\}$. (Why is the integrand $\nu(B - x)$ measurable?)

Theorem 6.8. Let X and Y be two independent random vectors in $\mathbb{R}^d$ with distributions μ and ν. Then the distribution of of $X + Y$ is $\mu \star \nu$.

Proof. By independence, $\mathbf{P}_{(X,Y)} = \mu \otimes \nu$. Then by Tonelli's theorem

$$\begin{aligned}\mathbf{P}_{X+Y}(B) &= \int \chi_B d\mathbf{P}_{X+Y} = \int \chi_B(x + y)\mathbf{P}_{(X,Y)}(d(x, y)) \\ &= \iint \chi_B(x + y)\, \nu(dy)\, \mu(dx) = \iint \chi_{B-x}(y)\, \nu(dy)\, \mu(dx) \\ &= \mu \star \nu(B)\end{aligned}$$

□

In particular, this shows $\mu \star \nu = \nu \star \mu$, and that $\mu \star \nu$ is a probability measure if both μ and ν are. Moreover, convolution is associative:

Theorem 6.9. Given probability measures μ, ν and ρ on $\mathbb{R}^d$,

$$(\mu \star \nu) \star \rho = \mu \star (\nu \star \rho).$$

Proof. Verify that both $(\mu \star \nu) \star \rho(A)$ and $\mu \star (\nu \star \rho)(A)$ equal

$$\mu \otimes \nu \otimes \rho(\{(x, y, z) \in \mathbb{R}^{3d} : x + y + z \in A\})$$

$\square$

Lemma 6.10. If μ and ν are probability laws on $\mathbb{R}^d$ and μ has a density f, then $\mu \star \nu$ has a density $h(x) = \int f(x - y)\nu(dy)$. If ν also has a density, g, then $h(x) = \int f(x - y)g(y)\, dy$.

Proof. Use the substitution $x + y \to z$ in the integral

$$\mu \star \nu(A) = \iint \chi_A(x + y)\, \mu(dx)\, \nu(dy) = \iint \chi_A(x + y)f(x)dx\, \nu(dy).$$

$\square$

Definition 6.11. The Gaussian probability distribution on $\mathbb{R}$ with mean 0 and variance $\sigma^2 > 0$ is the law $N(0, \sigma^2)$ with Lebesgue density

$$\varphi_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-x^2/(2\sigma^2))$$

Moreover, we define $N(0, \sigma^2) \otimes \ldots \otimes N(0, \sigma^2) =: N(\mathbf{0}, \sigma^2 I)$

Lemma 6.12. The characteristic function of the law $N(0, \sigma^2)$ on $\mathbb{R}$ is

$$\hat{\varphi}_\sigma(u) = \int_{\mathbb{R}} e^{ixu}\varphi_\sigma(x)\, dx = \exp(\sigma^2 u^2/2) = \frac{\sqrt{2\pi}}{\sigma}\varphi_{1/\sigma}(u)$$

Proof. See Dudley 9.4.2. $\square$

Lemma 6.13. Let μ be a finite measure on $\mathbb{R}^d$ with characteristic function $\hat{\mu}(k) = \int e^{i\langle x,k\rangle}\, \mu(dx)$. Then the measure $\mu^{(\sigma)} := \mu \star N(\mathbf{0}, \sigma^2 I)$ has a Lebesgue density $f^{(\sigma)}$ which satisfies

$$f^{(\sigma)}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int \hat{\mu}(k)e^{-ikx}\varphi_{1/\sigma}(k)\, dk \tag{6.14}$$

Proof. By Lem 6.12, for $d = 1$, $\varphi_\sigma(x) = \hat{\varphi}_{1/\sigma}(x)/(\sigma\sqrt{2\pi})$. By Lem 6.10,

$$\begin{aligned} f^{(\sigma)}(x) &= \int \varphi_\sigma(x-y)\mu(dy) = \int \varphi_\sigma(y-x)\mu(dy) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int \hat{\varphi}_{1/\sigma}(y-x)\mu(dy) = \frac{1}{\sigma\sqrt{2\pi}} \iint e^{ik(y-x)}\varphi_{1/\sigma}(k)dk\,\mu(dy) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \iint e^{iky}\,\mu(dy)e^{-ikx}\varphi_{1/\sigma}(k)\,dk = \frac{1}{\sigma\sqrt{2\pi}} \int \hat{\mu}(k)e^{-ikx}\varphi_{1/\sigma}(k)\,dk \end{aligned}$$

For $d > 1$, note that the law $N(\mathbf{0}, \sigma^2 I)$ on $\mathbb{R}^d$ also has a density, namely

$$\varphi_\sigma(x_1, \ldots, x_d) = \frac{1}{(\sigma\sqrt{2\pi})^d} \prod_{i=1}^{d} \exp(-x_i^2/(2\sigma^2)).$$

Then replace xk by $\langle x, k\rangle$, k^2 by $\|k\|^2$ and $(2\pi)^{-1}$ by $(2\pi)^{-d}$. □

Lemma 6.15. For any law μ on $\mathbb{R}^d$, as $\sigma \downarrow 0$, we have $\mu^{(\sigma)} \to \mu$.

Proof. Let X be a random vector with law μ and let Y be an independent random vector with law $N(0, I)$. Then by Th 6.8 $X + \sigma Y$ has law $\mu^{(\sigma)}$. As $\sigma \downarrow 0$, we have $X + \sigma Y \to X$ a.s. Hence by dominated convergence, for $f \in C_b(\mathbb{R}^d)$ we find $\int f\,d\mu^{(\sigma)} = \mathbf{E}(f(X + \sigma Y)) \to \mathbf{E}(f(X)) = \int f\,d\mu$ as $\sigma \downarrow 0$. □

Theorem 6.16 (Uniqueness Theorem)**.** Let μ and ν be distributions on $\mathbb{R}^d$ with the same characteristic function. Then $\mu = \nu$.

Proof. Write $g = \hat{\mu} = \hat{\nu}$ for the characteristic function of both μ and ν. By Lem 6.13, $\mu^{(\sigma)}$ and $\nu^{(\sigma)}$ both have the same density, hence $\mu^{(\sigma)} = \nu^{(\sigma)}$. Then by Lem 6.15, $\mu \leftarrow \mu^{(\sigma)} = \nu^{(\sigma)} \to \nu$, and $\mu = \nu$. □

6.2 Lévy's Continuity Theorem

Lemma 6.17. If μ_n and μ are laws such that for every subsequence $\mu_{n(k)}$ there is a subsubsequence $\mu_{n(k(r))} \to \mu$, then $\mu_n \to \mu$.

Proof. If not, then for some $f \in C_b(\mathbb{R}^d)$, $\int f\,d\mu_n \not\to \int f\,d\mu$. Then for some $\varepsilon > 0$ and sequence $n(k)$, $|\int f\,d\mu_{n(k)} - \int f\,d\mu| > \varepsilon$ for all k. Then $\mu_{n(k(r))} \to \mu$ gives a contradiction. □

Lemma 6.18. Let $\{a_n\}_{n\in\mathbb{N}}$ be a sequence of points in $\mathbb{R}$, and let $\{u_n\}_{n\in\mathbb{N}}$ be a sequence of functions $\mathbb{R} \to \mathbb{R}$. Then there is a subsequence $u_{n_1}, u_{n_2}, \ldots$ that converges at all points a_n.

Proof. G. Cantor's diagonal method: The sequence $u_n(a_1)$ must have at least one limit point $\in \mathbb{R} \cup \{-\infty, +\infty\}$. Pick a subsequence of $u_n(a_1)$ which converges to this limit point. Extract a further subsequence for which both $u_n(a_1)$ and $u_n(a_2)$ converge. Proceed inductively to construct a sequence of sequences $u_n^{(k)}$. Then the "diagonal sequence" $u_n^{(n)}$ must converge at every a_k. □

Lemma 6.19. Let μ be a finite measure on $\mathbb{R}$ and μ_n a sequence of finite measures on $\mathbb{R}$. Write $F_n(x) = \mu_n((-\infty, x])$ and $F(x)$ for the cumulative distribution functions. Then $\mu_n \to \mu$ iff $F_n(x) \to F(x)$ at every point of continuity x of F.

Proof. "$\Rightarrow$" Suppose $\mu_n \to \mu$ and let x be a point of continuity of F. Let $\varepsilon > 0$, then there is $\delta > 0$ such that $F(x+\delta) - F(x-\delta) < \varepsilon$. Let $f \in C_b(\mathbb{R})$ be such that $f(t) = 1$ on $(-\infty, x-\delta)$, $f(t)$ decreasing on $[x-\delta, x]$ and $f(t) = 0$ on (x, ∞). Then $\int f d\mu_n \leqslant F_n(x) \leqslant \int g d\mu_n$, and hence $\int f d\mu \leqslant \liminf F_n(x) \leqslant \limsup F_n(x) \leqslant \int g d\mu$. But $\int g d\mu - \int f d\mu \leqslant \int \chi_{(x-\delta, x+\delta]} d\mu < \varepsilon$, and since ε is arbitrary, $\liminf F_n(x) = \limsup F_n(x)$. Since also $\int f d\mu \leqslant F(x) \leqslant \int g d\mu$, $\lim F_n(x) = F(x)$.

"$\Leftarrow$" Let $f \in C_b(\mathbb{R})$ with $-M \leqslant f(x) \leqslant M$. We put

$$D = \{t \in \mathbb{R} : \mu\{x : f(x) = t\} \neq 0\}$$

and consider a decomposition $T_k = (t_0, t_1, \ldots, t_k)$ of $[-M, M]$:

$$-M = t_0 < t_1 < \ldots < t_k = M, \quad k \geqslant 1,$$

with $t_i \notin D$, $i = 1, \ldots, k$. (Observe that D is at most countable since the sets $f^{-1}(\{t\})$ are disjoint and μ is finite.) Let $B_i := f^{-1}((t_i, t_{i+1}])$. Since $f(x)$ is continuous and therefore the set $f^{-1}(t_i, t_{i+1})$ is open, we have that the boundary $\partial B_i \subset f^{-1}(\{t_i\}) \cup f^{-1}(\{t_{i+1}\})$. The points $t_i, t_{i+1} \notin D$; therefore $\mu(\partial B_i) = 0$ and hence by assumption $\mu_n(B_i) \to \mu(B_i)$. But

$$\begin{aligned}
&\left|\int f d\mu_n - \int f d\mu\right| \\
&\leqslant \left|\int f d\mu_n - \sum_{i=0}^{k-1} t_i \mu_n(B_i)\right| + \left|\sum_{i=0}^{k-1} t_i \mu_n(B_i) - \sum_{i=0}^{k-1} t_i \mu(B_i)\right| + \left|\sum_{i=0}^{k-1} t_i \mu(B_i) - \int f d\mu\right| \\
&\qquad \leqslant 2 \max\{t_{i+1} - t_i : 0 \leqslant i \leqslant k-1\} + \left|\sum_{i=0}^{k-1} t_i \mu_n(B_i) - \sum_{i=0}^{k-1} t_i \mu(B_i)\right|
\end{aligned}$$

and since the t_i are arbitrary, $\left|\int f d\mu_n - \int f d\mu\right| \to 0$. □

Proposition 6.20 (Selection Theorem)**.** For any sequence of probability laws $\mu_1, \mu_2, \ldots$ on $\mathbb{R}$, there is a subsequence $\mu_{n_1}, \mu_{n_2}, \ldots$ which converges weakly to a finite measure μ with $\mu(\mathbb{R}) \leqslant 1$.

Proof. Write $F_n(x) := \mu_n((-\infty, x])$ (cumulative distribution function). Note that each F_n is increasing (not strictly) from 0 to 1 and right-continuous with left-hand limits (rcll). Let a_n be an enumeration of $\mathbb{Q}$ (or any other countable set which is dense in $\mathbb{R}$). Then by Lem 6.18 there is a subsequence $F_{n_1}, F_{n_2}, \ldots$ which converges at every a_n. Denote the limits by $G(a_n)$. For $x \notin \{a_n\}_{n\in\mathbb{N}}$, define $G(x) := \inf\{G(a_n) : a_n > x\}$. Note that G is also increasing. Now let $F(x) = G(x)$ where is G is continuous at x, and $F(x) = G(x+)$ where G has a jump. Then F is rcll. We show that $F_n(x) \to F(x)$ at points of continuity x. If x is such a point, then there are two points $a_i < x < a_j$ such that

$$G(a_j) - G(a_i) < \varepsilon, \qquad G(a_i) \leqslant F(x) \leqslant G(a_j). \tag{6.21}$$

As the F_n are monotone, we have $F_{n_k}(a_i) \leqslant F_{n_k}(x) \leqslant F_{n_k}(a_j)$. Letting $k \to \infty$, we see from (6.21) that no limit point of the sequence $F_{n_k}(x)$ can differ from $F(x)$ by more than ε, and so $F_{n_k}(x) \to F(x)$ at all points of continuity. By Lem 6.19, $\mu_{n_k} \to \mu$. Note that we only know that $\lim_{x\to-\infty} F(x) \geqslant 0$ and $\lim_{x\to+\infty} F(x) \leqslant 1$. □

Lemma 6.22. Let μ_n be a sequence of probability laws with characteristic function $\hat{\mu}_n$. If $\hat{\mu}_n(k) \to \hat{\mu}(k)$ for all k, then there exists a finite law μ with $\mu(\mathbb{R}) \leqslant 1$ such that $\mu_n \to \mu$.

Proof. Pick any subsequence $n(k)$. By Th 6.20, there is a subsubsequence $n(k(r))$ and a finite measure μ with $\mu(\mathbb{R}) \leqslant 1$ such that $\mu_{n(k(r))} \to \mu$. Apply (6.14) to each pair $(\mu_{n(k(r))}, \hat{\mu}_{n(k(r))})$. In the limit as $n \to \infty$, we again recover (6.14): the LHS converges to the density of $\mu^{(\sigma)}$, since $\varphi_\sigma \in C_b(\mathbb{R})$, and the RHS converges by the DCT with $\varphi_{1/\sigma}$ as a dominating function. Since $\hat{\mu}_n$ converges, the limit of the RHS is independent of the choice of subsequence $n(k)$. Hence the limit $\mu^{(\sigma)}$ of the LHS must also be independent of the choice of $n(k)$. By Lem 6.17, $\mu_n^{(\sigma)} \to \mu^{(\sigma)}$. Then

$$\begin{aligned}
&\left|\int f\,d\mu_n - \int f\,d\mu\right| \\
&\leqslant \left|\int f\,d\mu_n - \int f\,d\mu_n^{(1/k)}\right| + \left|\int f\,d\mu_n^{(1/k)} - \int f\,d\mu^{(1/k)}\right| + \left|\int f\,d\mu^{(1/k)} - \int f\,d\mu\right|
\end{aligned}$$

for all n, k. Pick a subsequence $n(m)$, and then a "diagonal" subsubsequence for which the above RHS tends to 0. By Lem 6.17 again, the above LHS tends to 0. □

Theorem 6.23 (Lévy Continuity Theorem)**.** In order that a sequence μ_n of laws on $\mathbb{R}$ converges weakly to a law μ, it is necessary and sufficient that the sequence of characteristic functions $\hat{\mu}_n(k)$ converge for all k to some $\hat{\mu}(k)$, and that $\hat{\mu}$ is continuous in some neighbourhood of the origin. In this case $\hat{\mu}$ is the characteristic function of μ.

Proof. "$\Rightarrow$" follows from Def 6.1, Lem 6.6 and Lem 6.5.
"$\Leftarrow$" By Lem 6.22, the limit $\mu_n \to \mu$ exists, where $\mu(\mathbb{R}) \leqslant 1$. Rewrite (6.14) as

$$\sigma\sqrt{2\pi}\int \varphi_\sigma(x-y)\,\mu(dy) = \int \hat{\mu}(k)e^{-ikx}\varphi_{1/\sigma}(k)\,dk.$$

We let $\sigma \to +\infty$. The RHS equals $\mathbf{E}(\hat{\mu}(Y)e^{-iYx/\sigma})$ where $Y \sim N(0,1)$. Since $Y/\sigma \to 0$ a.s. and since $\hat{\mu}$ is continuous at 0, by the DCT this expectation converges to $\hat{\mu}(0)$. But by assumption $\hat{\mu}(0) = \lim \hat{\mu}_n(0) = \lim 1 = 1$. On the other hand, note that $\sigma\sqrt{2\pi}\varphi_\sigma(x) \leqslant 1$, and integrating this inequality against μ shows $\mu(\mathbb{R}) \geqslant 1$. Hence $\mu(\mathbb{R}) = 1$, that is, μ is a law. $\square$

We close this chapter by noting that, besides a.s. convergence, there are two other main notions of convergence in probability. We list:

Definition 6.24. Let X_n be a sequence of $\mathbb{R}^d$-valued random variables (random vectors) and X be an $\mathbb{R}^d$-valued random variable.

(1) X_n converges to X almost surely if $\mathbf{P}(\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}) = 1$.

(2) X_n converges to X in probability if, for all $a > 0$, $\lim_n \mathbf{P}(|X_n - X| > a) = 0$.

(3) X_n converges to X in L^p $(1 \leqslant p < \infty)$ if $\lim_n \mathbf{E}(|X_n - X|^p) = 0$.
The case $p = 2$ is often called "convergence in mean square."

(4) X_n converges to X weakly or in distribution if $\lim_n \mathbf{E}(f(X_n)) = \mathbf{E}(f(X))$ for all $f \in C_b(\mathbb{R}^d)$.

Note that X_n converges to X weakly iff $\mathbf{P}_{X_n} \to \mathbf{P}_X$ (weakly). The following relations are important:

$$(1) \Rightarrow (2) \Rightarrow (4);$$
$$(3) \Rightarrow (2) \Rightarrow (4).$$

The Levy Continuity theorem makes (4) a very versatile notion of convergence in probability theory, as will become clear in the next chapter.

Chapter 7

Infinite Divisibility

7.1 Definition

Definition 7.1. (1) Write $M_1(\mathbb{R}^d)$ for the set of all probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

(2) We write $\mu^{*n} = \mu * \ldots * \mu$ (n times) and say that μ has a convolution nth root if there exists a measure $\mu^{1/n} \in \mathcal{M}_1(\mathbb{R}^d)$ for which $(\mu^{1/n})^{*n} = \mu$.

Exercise 7.2. If $\mu, \nu \in M_1(\mathbb{R}^d)$, show that the characteristic function of $\mu * \nu$ is $\hat{\mu}\hat{\nu}$.

Exercise 7.3. Let X have a gamma distribution in $\mathbb{R}$ with parameters $n \in \mathbb{N}$ and $\lambda > 0$, so that X has the density

$$f(x) = \frac{\lambda^n x^{n-1} e^{-\lambda x}}{(n-1)!} \quad \text{for } x > 0.$$

Show that X has a convolution nth root given by the exponential distribution with parameter λ and density $\lambda e^{-\lambda x}$.

Definition 7.4. A random vector X is infinitely divisible if, for all $n \in \mathbb{N}$, there exist i.i.d. random vectors $Y_1^{(n)}, \ldots, Y_n^{(n)}$ such that

$$\mathbf{P}_X = \mathbf{P}_{Y_1^{(n)} + \ldots + Y_n^{(n)}}.$$

Note that the last line is more typically written as $X \stackrel{d}{=} Y_1^{(n)} + \ldots + Y_n^{(n)}$. A probability law is called infinitely divisible if it is the law of an infinitely divisible random vector.

Proposition 7.5. The following are equivalent:

(1) X is infinitely divisible;

(2) $\mathbf{P}_X$ has a convolution nth root that is itself the law of a random vector, for each $n \in \mathbb{N}$;

(3) $\hat{\mathbf{P}}_X$ has an nth root that is itself the characteristic function of a random variable, for each $n \in \mathbb{N}$.

Proof. Exercise. □

We can hence say that $\mu \in M_1(\mathbb{R}^d)$ is infinitely divisible if it has a convolution nth root in $M_1(\mathbb{R}^d)$ for each $n \in \mathbb{N}$.

7.2 Examples

Gaussian random vectors. If a random vector X has a density of the form

$$\varphi_{m,A} = (2\pi)^{-d/2}(\det(A))^{-1/2} \exp\left[-\langle x - m, A^{-1}(x - m)\rangle/2\right]$$

where $m \in \mathbb{R}^d$ and $A \in \mathbb{R}^{d \times d}$ is positive definite, then we say that X is (non-degenerate) Gaussian. Its law is often denoted as $N(m, A)$. The vector m and matrix A are then called mean and covariance, since $\mathbf{E}(X) = m$ and $\mathbf{E}\left[(X - m)(X - m)^\top\right] = A$. One can then calculate

$$\hat{\mathbf{P}}_X(u) = \exp\left[i\langle m, u\rangle - \langle u, Au\rangle/2\right]$$

and hence

$$\left[\hat{\mathbf{P}}_X(u)\right]^{1/n} = \exp\left[i\langle m/n, u\rangle - \langle u, n^{-1}Au\rangle/2\right]$$

which shows that X is infinitely divisible with $Y_j^{(n)} \sim N(m/n, (1/n)A)$ for each $1 \leqslant j \leqslant n$.

Poisson random variables. A random variable X is Poisson if it takes values in $\mathbb{N} \cup \{0\}$ and there exists $c > 0$ such that

$$\mathbf{P}(X = n) = \frac{c^n}{n!}e^{-c}.$$

We write $X \sim \pi(c)$). We have $\mathbf{E}(X) = \mathrm{Var}(X) = c$. One can verify that

$$\hat{\mathbf{P}}_X(u) = \exp\left[c(e^{iu} - 1)\right],$$

from which we deduce that X is infinitely divisible with each $Y_j^{(n)} \sim \pi(c/n)$, for $1 \leqslant j \leqslant n$, $n \in \mathbb{N}$.

Compound Poisson random variables. Suppose that $\{Z_n\}_{n\in\mathbb{N}}$ is an i.i.d. (independent and identically distributed) sequence of random vectors with common law $\mathbf{P}_Z$ and let $N \sim \pi(c)$ be a Poisson random variable that is independent of all the Z_n. The compound Poisson random vector X is defined as $X = Z_1 + \ldots + Z_N$. (Why is it measurable?) We may think of X as a random walk with a random number of steps, controlled by the Poisson random variable N.

Proposition 7.6. The characteristic function of a compound Poisson random variable is given by

$$\hat{\mathbf{P}}_X(u) = \exp\left[\int_{\mathbb{R}^d} \left(e^{i\langle u,y\rangle} - 1\right) c\mathbf{P}_Z(dy)\right].$$

Proof. By conditioning and using independence we find that

$$\begin{aligned}
\hat{\mathbf{P}}_X(u) &= \sum_{n=0}^{\infty} \mathbf{E}(\exp[i\langle u, Z(1) + \ldots + Z(N)\rangle]|N = n)\mathbf{P}(N = n) \\
&= \sum_{n=0}^{\infty} \mathbf{E}(\exp[i\langle u, Z(1) + \ldots + Z(n)\rangle])e^{-c}\frac{c^n}{n!} \\
&= e^{-c}\sum_{n=0}^{\infty} \frac{[c\hat{\mathbf{P}}_Z(u)]^n}{n!} = \exp[c(\hat{\mathbf{P}}_Z(u) - 1)],
\end{aligned}$$

and the result follows on writing $\hat{\mathbf{P}}_Z(u) = \int_{\mathbb{R}^d} e^{i\langle u,y\rangle}\mathbf{P}_Z(dy)$. $\square$

If X is compound Poisson as above, we write $X \sim \pi(c, \mathbf{P}_Z)$. It is clearly infinitely divisible with each $Y_n^{(n)} \sim \pi(c/n, \mathbf{P}_Z)$, for $1 \leqslant j \leqslant n$. The random variable X will have a finite mean iff each Z_n does. This can be shown e.g. by differentiating $\hat{\mathbf{P}}_X(u)$: In general, using the DCT we have

$$\left.\frac{\partial}{\partial u_k}\hat{\mathbf{P}}_X(u)\right|_{u=0} = i\mathbf{E}(X_k)$$

which in this case equals $c\mathbf{E}(Z)$. Similar remarks apply for the moments $\mathbf{E}(X_k X_\ell)$ and higher-order moments.

Exercise 7.7. (1) Verify that the sum of two independent infinitely divisible random vectors is itself infinitely divisible.

(2) Using the Lévy Continuity Theorem, show that the weak limit of a sequence of infinitely divisible probability measures is itself infinitely divisible.

7.3 The Lévy Khintchine Formula

Definition 7.8. Let ν be a Borel measure with $\nu(\{0\}) = 0$. We say that it is a Lévy measure if

$$\int (\|y\|^2 \wedge 1)\nu(dy) < \infty.$$

Note that ν may be an infinite measure, but $\nu(B(0,\varepsilon)^{\complement}) < \infty$ for all ε, where $B(x,\varepsilon) := \{y \in \mathbb{R}^d : \|y - x\| < \varepsilon\}$.

Exercise 7.9. (1) Any Lévy measure is σ-finite.

(2) A Borel measure is a Lévy measure iff $\nu(\{0\}) = 0$ and

$$\int \frac{\|y\|^2}{1 + \|y\|^2}\nu(dy) < \infty.$$

Theorem 7.10 (Lévy-Khintchine)**.**

(1) $\mu \in M_1(\mathbb{R}^d)$ is infinitely divisible if there exists a vector $b \in \mathbb{R}^d$, a positive definite (symmetric) $d \times d$ matrix A and a Lévy measure ν such that, for all $u \in \mathbb{R}^d$,

$$\hat{\mu}(u) = \exp\left(i\langle b, u\rangle - \frac{\langle u, Au\rangle}{2} + \int \left[e^{i\langle u,y\rangle} - 1 - i\langle u, y\rangle \chi_{B(0,1)}(y)\right] \nu(dy)\right).$$

(2) Conversely, any mapping of the above form is the characteristic function of an infinitely divisible probability measure on $\mathbb{R}^d$.

Proof. We can only prove part (2). First, we show that the right-hand side defines a characteristic function. Let $\{\alpha_n\}_{n\in\mathbb{N}} \subset \mathbb{R}^d$ be such that $\alpha_n^{(k)} \downarrow 0$ for each $k = 1, \ldots, d$. Then define the box $B_n := \{y \in \mathbb{R}^d : -\alpha_n^{(k)} \leqslant y \leqslant \alpha_n^{(k)}\}$, and define the function

$$\varphi_n(u) := \exp\left(i\left\langle b - \int_{B(0,1)\cap B_n} y\,\nu(dy), u\right\rangle - \frac{1}{2}\langle u, Au\rangle + \int_{B_n^{\complement}}\right)$$

□