Measure theory

1

1.1 Measure and probability

Let $\mathbb{S}$ be an arbitrary nonempty set. Recall, a family of $\mathcal{S} \subseteq \mathcal{P}(\mathbb{S})$ is called *σ–algebra* if

i) $\emptyset \in \mathcal{S}$,

ii) $A \in \mathcal{S}$ implies $A^c \in \mathcal{S}$,

iii) $A_i \in \mathcal{S}$, $i \in \mathbb{N}$, implies $\cup_{i=1}^{\infty} A_i \in \mathcal{S}$.

If $\mathcal{S}$ satisfies i),ii) and iii'): $A_i \in \mathcal{S}$, $i = 1, \ldots, n$, implies $\cup_{i=1}^{n} A_i \in \mathcal{S}$, we say that $\mathcal{S}$ is an *algebra.*

For a class of sets $\mathcal{C} \subseteq \mathcal{P}(\mathbb{S})$, by $\sigma(\mathcal{C})$ we denote the smallest σ–algebra containing $\mathcal{C}$. Using exercise 1.1, one can check that

$$\sigma(\mathcal{C}) = \cap_{\mathcal{C} \subseteq \mathcal{F}_\alpha} \mathcal{F}_\alpha,$$

where in the intersection above $\mathcal{F}_\alpha$ is assumed to be an arbitrary σ–algebra on $\mathbb{S}$ containing $\mathcal{C}$.

Exercise 1.1 Suppose $(\mathcal{F}_\alpha)_{\alpha \in I}$ is an arbitrary collection of σ–algebras on the same set $\mathbb{S}$, then $\cap_{\alpha \in I} \mathcal{F}_\alpha$ is a σ–algebra as well.

When $\mathbb{S}$ is a metric space (e.g. $\mathbb{R}^d$), the most frequently used σ-algebra is the *Borel* σ-algebra generated by the topology, i.e. $\mathcal{B} = \mathcal{B}(\mathbb{S}) := \sigma(\{C : C \text{ open set}\})$.

Exercise 1.2 a) For $\mathbb{S} = \mathbb{R}$, the Borel σ-algebra can be generated as $\sigma(\mathcal{C})$ for the following classes of sets: i) $\mathcal{C} = \{(-\infty, x] : x \in x \in \mathbb{R}\})$, ii) $\mathcal{C} = \{(-\infty, x) : x \in \mathbb{R}\})$, iii) $\mathcal{C} = \{(x, y) : x < y \in \mathbb{R}\})$, iv) $\mathcal{C} = \{(x, y] : x < y \in \mathbb{R}\})$. Can you suggest a countable class $\mathcal{C}$ generating $\mathcal{B}$?
b) In a general separable metric space $\mathbb{S}$ Borel σ-algebra can be generated by all open balls, i.e. $\mathcal{B} = \sigma(\{B(x, r) : x \in \mathbb{S}, r > 0\})$.

Suppose the pair $(\mathbb{S}, \mathcal{S})$ denotes a nonempty set and a σ–algebra on that set, the pair is called a measurable space, and a function $\mu : \mathcal{S} \to [0, \infty] = [0, \infty) \cup \{\infty\}$ is called a *measure* if it satisfies

i) $\mu(\emptyset) = 0$,

ii) for disjoint $A_i \in \mathcal{S}$, $i \in \mathbb{N}$ we have

$$\mu(\cup_i A_i) = \sum_i \mu(A_i).$$

The second property above is called σ–additivity. If in addition μ also satisfies

iii) $\mu(\mathbb{S}) = 1\,,$

we say that μ is a *probability*. The triple $(\mathbb{S}, \mathcal{S}, \mu)$ is then called a *probability space*.

To emphasize the difference with other spaces and measures, the *probability space* is often denoted by symbols $(\Omega, \mathcal{F}, \mathbb{P})$.

Example 1.1 a) $\mu(A) = 0$ is called a trivial measure, it is well defined on $\mathcal{P}(\mathbb{S})$ for any set $\mathbb{S}$.
b) if μ is a measure, for $c \geq 0$, $c \cdot \mu$ is also a measure.
c) if μ_i, $i \in \mathbb{N}$ is a sequence of measures, then $\mu(A) = \sum_i \mu_i(A)$ is also a measure.
d) for a fixed $x \in \mathbb{S}$ $\mu(A) = \delta_x(A) = \mathbb{1}_A(x)$ is called a Dirac measure concentrated in x, it is well defined on $\mathcal{P}(\mathbb{S})$.
e) for $\sum_i \alpha_i = 1$ and (x_i) a sequence in $\mathbb{S}$, $\mu = \sum_i \alpha_i \delta_{x_i}$ is *discrete probability measure.*
f) on $\mathbb{R}$, $\mathbb{R}_+$ or $[0,1]$, it may be useful to construct a measure such that $\lambda(a,b] = b - a$, for any $a \leq b$, however, it is not clear what one should take as a σ–algebra. One of the basic results of measure theory says that $\mathcal{S} = \mathcal{P}(\mathbb{R})$ does not work, i.e. there is no measure on a σ–algebra so large, which takes length of intervals as their measure.

Remark 1.1 In the finite or a countable space $\mathbb{S} = \{x_i : i \in \mathbb{N}\}$, probability measure on $\mathcal{S}$ can be always extended to $\mathcal{P}(\mathbb{S})$ and written as (x_i), $\mu = \sum_i \alpha_i \delta_{x_i}$ with $\sum_i \alpha_i = 1$.

All the properties of the next lemma hold in any probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and are rather easy to show. Properties a), c) and d) also hold for if we substitute $\mathbb{P}$ with a general measure μ. The last one holds as well provided that $\mu(A_1) < \infty$.

Lemma 1.1 (Properties of probability measures)**.** *a) For disjoint $A_1, \dots, A_n \in \mathcal{S}$, we have $\mathbb{P}(\cup_i A_i) = \sum_i \mathbb{P}(A_i)\,.$*

b) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ for each $A \in \mathcal{S}$.

c) For any $A_i \in \mathcal{S}$, $i \in \mathbb{N}$ we have $\mu(\cup_i A_i) \leq \sum_i \mu(A_i)\,.$

d) For any $A_1 \subseteq A_2 \subseteq \ldots \in \mathcal{S}$, and $A = \cup_i A_i$ (i.e. $A_i \nearrow A$) we have $\mathbb{P}(A) = \lim_n \mathbb{P}(A_n)$.

e) For any $A_1 \supseteq A_2 \supseteq \ldots \in \mathcal{S}$, and $A = \cap_i A_i$ (i.e. $A_i \searrow A$) we have $\mathbb{P}(A) = \lim_n \mathbb{P}(A_n)$.

We say that the property of elements $\omega \in A \in \mathcal{F}$ holds almost surely (or that A holds a.s.) if $\mathbb{P}(A) = 1$ or equivalently if $\mathbb{P}(A^c) = 0$. For a general measure μ and $A \in \mathcal{S}$ we say that A is μ almost everywhere (μ-a.e.) set if $\mu(A^c) = 0$. In such a case we also say that the set A^c is negligible.

Example 1.2 By finite additivity, for any $C \subseteq A$, we have $\mathbb{P}(C) + \mathbb{P}(A \setminus C) = \mathbb{P}(A)$, i.e. $\mathbb{P}(C) = \mathbb{P}(A) - \mathbb{P}(A \setminus C)$. Also for any A, B measurable, $\mathbb{P}(A \cap B) = \mathbb{P}(A \setminus (A \cap B)) + \mathbb{P}(B \setminus (A \cap B)) + \mathbb{P}(A \cap B)$ producing

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)\,.$$

Inductively, it follows that for any $A_1, \dots, A_n$ measurable, we have inclusion-exclusion formula

$$\mathbb{P}(\cup_1^n A_i) = \sum_1^n \mathbb{P}(A_i) - \sum_{1\le i<j\le n} \mathbb{P}(A_i \cap A_j) + \sum_{1\le i<j<k\le n} \mathbb{P}(A_i \cap A_j \cap A_k)$$
$$- \dots + (-1)^{n+1}\mathbb{P}(A_1 \cap A_2 \cap \ldots \cap A_n).$$

If a probability (or a measure) $\mathbb{P}$ in the triple $(\Omega, \mathcal{F}, \mathbb{P})$ has the property that $\mathbb{P}(A) = 0$ and $A' \subseteq A \in \mathcal{F}$ imply that $A' \in \mathcal{F}$, we say that the (probability) measure $\mathbb{P}$ is complete. It is not too hard to see that for any $(\Omega, \mathcal{F}, \mathbb{P})$ we can extend $\mathbb{P}$ from σ-algebra $\mathcal{F}$ to a possibly larger σ-algebra $\mathcal{F}'$ so that $\mathbb{P}$ in $(\Omega, \mathcal{F}', \mathbb{P})$ becomes complete. It is sometimes desirable to extend a probability $\mathbb{P}$ defined on a class of sets to the smallest σ-algebra containing that class. If the given class is an algebra, this can be always done and in a unique way.

Theorem 1.2 (Caratheodory)**.** *Assume $\mathbb{P}$ is a probability on an algebra $\mathcal{F}'$, then there exists a unique extension of $\mathbb{P}$ on $\sigma(\mathcal{F}')$.*

Idea of the proof. Observe, we actually demand that $\mathbb{P}$ is σ-additive on $\mathcal{F}'$ and that $\mathbb{P}(\Omega) = 1$. To construct the extension, one can introduce the so-called outer measure on $\mathcal{P}(\Omega)$ given by

$$\mathbb{P}^*(A) = \inf_{A\subseteq \cup_i A_i; A_i \in \mathcal{F}'} \sum_1^k \mathbb{P}(A_i)\,.$$

Note, the unions above are possibly countably infinite, that is $k \in \mathbb{N}_0 \cup \{\infty\}$ in general. Denote by $\mathcal{F}^*$ a family of the so called $\mathbb{P}^*$-measurable sets, that is

$$\mathcal{F}^* = \{A \subseteq \Omega : \mathbb{P}^*(A \cap E) + \mathbb{P}^*(A^c \cap E) = \mathbb{P}^*(E) \text{ for every } E \subseteq \Omega\}\,.$$

Clearly $\emptyset, \Omega \in \mathcal{F}^*$. It is straightforward to show that: $\mathbb{P}^*(\emptyset) = 0$; $\mathbb{P}^*(A) \ge 0$ for all A; $A \subseteq B$ implies $\mathbb{P}^*(A) \le \mathbb{P}^*(B)$. Moreover, for any $A_i \subseteq \Omega$, $\mathbb{P}^*(\cup_i A_i) \le \sum_i \mathbb{P}^*(A_i)$, i.e. $\mathbb{P}^*$ is σ-subadditive. To show the last claim, observe that for any $\varepsilon > 0$ and A_i there exist $B_{i,n}$, $A_i \subseteq \cup_n B_{i,n}$ and $\sum_n \mathbb{P}(B_{i,n}) < \mathbb{P}^*(A_i) + \varepsilon/2^i$. Therefore $\cup_i A_i \subseteq \cup_i \cup_n B_{i,n}$ and thus

$$\mathbb{P}^*(\cup_i A_i) \le \sum_i \sum_n \mathbb{P}(B_{i,n}) \le \sum_i \mathbb{P}^*(A_i) + \varepsilon.$$

Since ε was arbitrary, the claim follows. Moreover, this implies that $\mathbb{P}^*(A\cap E)+\mathbb{P}^*(A^c\cap E) \ge \mathbb{P}^*(E)$ for any A and E, therefore

$$A \in \mathcal{F}^* \text{ if and only if } \mathbb{P}^*(A \cap E) + \mathbb{P}^*(A^c \cap E) \le \mathbb{P}^*(E) \text{ for every } E. \tag{1.1}$$

Using this one can show the following facts

i) $\mathcal{F}^*$ is σ-algebra.

ii) $\mathcal{F}^* \supseteq \mathcal{F}'$

iii) $\mathbb{P}^*$ is σ-additive on $\mathcal{F}^*$, and a measure therefore.

iv) $\mathbb{P}^*|_{\mathcal{F}'} = \mathbb{P}$

The first three properties are subjects of exercises 1.5 and 1.6, they imply that $(\Omega, \mathcal{F}^*, \mathbb{P}^*)$ is a probability space, and ii) in particular implies $\mathcal{F}^* \supseteq \sigma(\mathcal{F}')$. The last one follows by the exercise 1.7.

It remains to show that this extension is unique. This will follow from the π/λ–theorem of the next section. $\square$

Remark 1.2 In general the σ-algebra $\mathcal{F}^*$ in the proof above is strictly larger that $\sigma(\mathcal{F}')$. Thus $(\Omega, \mathcal{F}^*, \mathbb{P}^*)$ and $(\Omega, \sigma(\mathcal{F}'), \mathbb{P})$ are two probability spaces extending $\mathbb{P}$, in the latter case we write $\mathbb{P}$ instead of $\mathbb{P}^*|_{\sigma(\mathcal{F}')}$

Exercise 1.3 Prove that the sets in $\mathcal{F}^*$ form an algebra. (Hint: to show finite additivity use (1.1)).

Exercise 1.4 Suppose (A_n) is a finite or countable sequence of disjoint sets in $\mathcal{F}^*$. Prove that for any $E \subseteq \Omega$ we have $\mathbb{P}^*(E \cap (\cup_i A_i)) = \sum_i \mathbb{P}^*(E \cap A_i)$. (Hint: subadditivity helps to show $\leq$, for the other direction start with finite unions).

Exercise 1.5 Show that sets in $\mathcal{F}^*$ form a σ-algebra, and $\mathbb{P}^*$ on $\mathcal{F}^*$ is σ–additive. (Hint: show first that for disjoint A_i's, $\cup_i A_i \in \mathcal{F}^*$, for the second part use the exercise above).

Exercise 1.6 Show $\mathcal{F}' \subseteq \mathcal{F}^*$.

Exercise 1.7 Show $\mathbb{P}^*(A) = \mathbb{P}(A)$ for each $A \in \mathcal{F}'$.

Lebesgue measure

Denote by $\mathcal{B}'$ the family of subsets of $(0,1]$ of the form $\cup_{i=1}^k (a_i, b_i]$ where $0 \leq a_1 < b_1 < \ldots a_k < b_k \leq 1$, $k \geq 0$. Thus, unless $k = 0$, $A \in \mathcal{B}'$ is a finite union of disjoint intervals of the type $I = (a, b]$. It is not difficult to see that $\mathcal{B}'$ is an algebra. Moreover, the length of the interval $|I| = b - a$ allows one to produce a function $\lambda : \mathcal{B}' \to [0,1]$ such that

$$\lambda(\cup_{i=1}^k (a_i, b_i]) = \sum_{i=1}^k (b_i - a_i)\,.$$

It is known that λ is σ-additive and therefore a probability on $\mathcal{B}'$. This is not too difficult, but a bit technical to show, see theorems 1.3 and 2.2 in [1]. By previous theorem, it has a unique extension to a measure on Borel σ-algebra $\mathcal{B} = \sigma(\mathcal{B}')$.

In the proof of Caratheodory theorem we introduced another σ–algebra $\mathcal{M} := \mathcal{F}^*$ containing $\mathcal{B}'$. We say that it consists of Lebesgue measurable sets. It is known that $\mathcal{B} \subsetneqq \mathcal{M} \subsetneqq \mathcal{P}(\mathbb{S})$. It turns out that λ is not complete on $\mathcal{B}$, and that $\mathcal{M}$ is exactly the completion of $\mathcal{B}$ with respect to λ. Of course, $((0,1], \mathcal{B}, \lambda)$ and $((0,1], \mathcal{M}, \lambda)$ are both probability spaces.

Clearly, we can make analogous construction on each interval $(k-1, k]$, $k \in \mathbb{Z}$, and obtain Lebesgue measure on the whole real line by summing all the corresponding measures.

1.2 Uniqueness and π/λ-systems

A family of subsets $\mathcal{G}$ of Ω is called π-system if it satisfies

π) $A, B \in \mathcal{G}$ implies $A \cap B \in \mathcal{G}$.

A family of subsets $\mathcal{L}$ of Ω is called λ-system (or a Dynkin system) if it satisfies

λ_1) $\Omega \in \mathcal{L}$,

λ_2) $A \in \mathcal{L}$ implies $A^c \in \mathcal{L}$,

λ_3) $A_1, A_2, \ldots \in \mathcal{L}$ disjoint implies $\cup_n A_n \in \mathcal{L}$.

Clearly, any σ-algebra is λ-system and π-system.

Exercise 1.8 Find λ-system on $\{1, 2, 3, 4, \}$ which is not σ-algebra.

Exercise 1.9 If a class $\mathcal{L}$ satisfies λ_1) and λ_3), then λ_2) holds if and only if the following holds

$$\lambda_2')\quad A, B \in \mathcal{L}, A \subseteq B \text{ implies } B \setminus A \in \mathcal{L}.$$

Exercise 1.10 If a class $\mathcal{L}$ satisfies λ_1 and λ_2', then λ_3 holds if and only if the following holds

$$\lambda_3')\quad A_1 \subseteq A_2 \subseteq \ldots \text{ and } A_n \in \mathcal{L} \text{ for all } n, \text{ implies } \cup_n A_n \in \mathcal{L}.$$

Lemma 1.3. *Suppose $\mathcal{G}$ is a π-system and λ-system, then it is also σ-algebra.*

Proof. Note, $\mathcal{G}$ is closed for finite intersections and taking complements. Since $A \cup B = (A^c \cap B^c)^c$ it closed for finite unions. It is then also closed for countable unions of $A_n \in \mathcal{G}$, since for $B_n = A_n \setminus (A_1 \cup \ldots \cup A_{n-1}) = A_n \cap A_1^c \cap \ldots \cap A_{n-1}^c \in \mathcal{G}$, by λ_3) we have

$$\cup_n A_n = \cup_n B_n \in \mathcal{G}\,.$$

□

Theorem 1.4 (Dynkin/Sierpiński). *Suppose $\mathcal{G}$ is a π-system and $\mathcal{L}$ is a λ-system. If $\mathcal{G} \subseteq \mathcal{L}$ then also $\sigma(\mathcal{G}) \subseteq \mathcal{L}$.*

Proof. Denote by $\mathcal{L}_{\mathcal{G}} = \cap_{\mathcal{L}_\alpha \supseteq \mathcal{G}} \mathcal{L}_\alpha$ with intersection running over all λ-systems $\mathcal{L}_\alpha$ containing $\mathcal{G}$. Clearly, $\mathcal{L}_{\mathcal{G}} \subseteq \mathcal{L}$ and $\mathcal{L}_{\mathcal{G}}$ is λ-system. It is sufficient to show it is a π-system, it will follow that it is σ-algebra as well.

Take, $A, B \in \mathcal{L}_{\mathcal{G}}$, we need to show $A \cap B \in \mathcal{L}_{\mathcal{G}}$. For $A \subseteq \Omega$ define

$$\mathcal{L}_A = \{B : A \cap B \in \mathcal{L}_{\mathcal{G}}\}.$$

Note first that $\mathcal{L}_A$ is λ-system for $A \in \mathcal{L}_{\mathcal{G}}$. Indeed: λ_1): $\Omega \in \mathcal{L}_A$ is obvious, λ_2'): $B_1, B_2 \in \mathcal{L}_A$, $B_1 \supseteq B_2$ implies $A \cap B_1$,$A \cap B_2 \in \mathcal{L}_{\mathcal{G}}$, since $\mathcal{L}_{\mathcal{G}}$ is λ -system, it follows $A \cap B_1 \setminus (A \cap B_2) = A \cap (B_1 \setminus B_2) \in \mathcal{L}_{\mathcal{G}}$. For λ_3) take disjoint $B_n \in \mathcal{L}_A$, thus $A \cap B_n$ are also disjoint elements of $\mathcal{L}_{\mathcal{G}}$ which is a λ-system. Thus $\cup_n A \cap B_n = A \cap (\cup_n B_n) \in \mathcal{L}_{\mathcal{G}}$, thus $\cup_n B_n \in \mathcal{L}_{\mathcal{G}}$.

By assumption, $\mathcal{G}$ is π-system, thus $A, B \in \mathcal{G}$ implies $A \cap B \in \mathcal{G} \subseteq \mathcal{L}_{\mathcal{G}}$. In particular conclude: 1) if $A \in \mathcal{G}$, $\mathcal{L}_A \supseteq \mathcal{G}$ and therefore $\mathcal{L}_A \supseteq \mathcal{L}_{\mathcal{G}}$. Now observe 2) if $A \in \mathcal{L}_{\mathcal{G}}$ and $B \in \mathcal{G}$, by symmetry $A \cap B \in \mathcal{L}_{\mathcal{G}}$, in other words $\mathcal{L}_A \supseteq \mathcal{G}$ and therefore $\mathcal{L}_A \supseteq \mathcal{L}_{\mathcal{G}}$ again. Therefore, 3) $A, B \in \mathcal{L}_{\mathcal{G}}$ implies $B \in \mathcal{L}_A$ (and vice versa, of course), i.e. $A \cap B \in \mathcal{L}_{\mathcal{G}}$.

Hence $\mathcal{L}_{\mathcal{G}}$ is a π-system, λ-system and therefore σ-algebra by the previous lemma. Hence, it contains $\sigma(\mathcal{G})$. □

Theorem 1.5. *Suppose $\mathcal{G}$ is a π-system, and $\mathbb{P}_1$ and $\mathbb{P}_2$ are two probability measures on $\sigma(\mathcal{G})$. If $\mathbb{P}_1|_{\mathcal{G}} = \mathbb{P}_2|_{\mathcal{G}}$ then*

$$\mathbb{P}_1|_{\sigma(\mathcal{G})} = \mathbb{P}_2|_{\sigma(\mathcal{G})}$$

Proof. Suppose $\mathbb{P}_1(A) = \mathbb{P}_2(A)$ for every $A \in \mathcal{G}$. Therefore, $\mathcal{L} = \{A \in \sigma(\mathcal{G}) : \mathbb{P}_1(A) = \mathbb{P}_2(A) \supseteq \mathcal{G}$. Clearly $\Omega \in \mathcal{L}$ and $A \in \mathcal{L}$ implies $A^c \in \mathcal{L}$. Assume disjoint $A_1, A_2, \ldots \in \mathcal{L}$ then $\mathbb{P}_1(\cup_n A_n) = \sum_n \mathbb{P}_1(A_n) = \sum_n \mathbb{P}_2(A_n) = \mathbb{P}_2(\cup_n A_n)$. Therefore, $\mathcal{L}$ is λ-system. By the previous theorem $\mathcal{L} \supseteq \sigma(\mathcal{G})$. □

Remark 1.3 Theorem above implies uniqueness in Caratheodory theorem 1.2

Exercise 1.11 Suppose that $\mathbb{P}_1$ and $\mathbb{P}_2$ are two probability measures on $(\mathbb{R}, \mathcal{B})$. a) Show $\mathbb{P}_1(A) = \mathbb{P}_2(A)$ for $A \in \mathcal{G} = \{(-\infty, x] : x \in \mathbb{R}\}$ implies $\mathbb{P}_1 = \mathbb{P}_2$. b) Show that the same holds for $\mathcal{G} = \{(a, b] : a < b \in \mathbb{Q}\}$. Hint: recall $\mathcal{B} = \sigma\{(-\infty, x] : x \in \mathbb{R}\}$.

Exercise 1.12 Denote by $\mathcal{G} = \{A \subseteq (0, 1] : A$ and $(A + r \pmod 1)$ are both Lebesgue measurable and $\lambda(A) = \lambda(A + r \pmod 1)$ for all $r \in (0, 1]\}$, where λ denotes the Lebesgue measure on $(0, 1]$. Show that $\mathcal{G}$ is a) a λ-system and b) it contains all intervals $(a, b] \subseteq (0, 1]$. Conclude that $\mathcal{G} \supseteq \mathcal{B}$.

Exercise 1.13 Consider the following relation on $(0, 1]$: $x \sim y$ if there exists $r \in \mathbb{Q} \cap (0, 1]$ such that $x + r \pmod 1 = y$. Show that a) this is equivalence relation on $(0, 1]$; b) if H is a set which consists of exactly one representative $x \in (0, 1]$ in each of the equivalence classes, and if $r \neq s$, $r, s \in \mathbb{Q} \cap (0, 1]$, then $H_r \cap H_s = \emptyset$ where $H_r = H + r \pmod 1$; c) show finally that $(0, 1] = \cup_{r \in \mathbb{Q} \cap (0,1]} H_r$.

The set H from the last exercise is called Vitali set. Using the last two exercises we can conclude that it is not a Borel set. If it was, it would be Lebesgue measurable and satisfy $\lambda(H) = \lambda(H_r)$ for each $r \in \mathbb{Q} \cap (0, 1]$. In that case, we would have $1 = \lambda(0, 1] = \lambda(\cup_{r \in \mathbb{Q} \cap (0,1]} H_r) = \infty \cdot \lambda(H)$ which is not possible for any value of $\lambda(H)$. One can show that any Lebesgue measurable set also has the property $\lambda(A) = \lambda(A + r \pmod 1)$, therefore H is not Lebesgue measurable either.

1.3 Integration

Measurable functions

Suppose $(\mathbb{S}, \mathcal{S})$ and $(\mathbb{T}, \mathcal{T})$ are two measurable spaces, a function $f : \mathbb{S} \to \mathbb{T}$ is called measurable (or $\mathcal{S}/\mathcal{T}$-measurable more precisely) if $f^{-1}(A) \in \mathcal{S}$ for each $A \in \mathcal{T}$. In the case of real functions $f : \mathbb{S} \to \mathbb{R}$ (or $\overline{\mathbb{R}}$) we say f is measurable if it is $\mathcal{S}/\mathcal{B}$-measurable, i.e. $f^{-1}(B) \in \mathcal{S}$ for each Borel set B.

Exercise 1.14 Suppose $g : \mathbb{S} \to \mathbb{T}$ is continuous, and $(\mathbb{S}, \mathcal{S})$ and $(\mathbb{T}, \mathcal{T})$ are such that $\mathcal{S}$, $\mathcal{T}$ are Borel σ-algebras on metric spaces $\mathbb{S}$ and $\mathbb{T}$. Show that g is measurable as well.

The following lemma is very easy to prove.

Lemma 1.6. *Suppose $f : \Omega \to \mathbb{S}$ and $g : \mathbb{S} \to \mathbb{T}$ are both measurable, where $(\Omega, \mathcal{F})$, $(\mathbb{S}, \mathcal{S})$ and $(\mathbb{T}, \mathcal{T})$ are three measurable spaces, Their composition is $g \circ f : \Omega \to \mathbb{T}$ is also measurable.*

Lemma 1.7. *Suppose $g : \mathbb{S} \to \mathbb{T}$ and $\mathcal{T} = \sigma(\mathcal{C})$ for some class of sets $\mathcal{C}$ in $\mathbb{T}$, then g is measurable if and only if $g^{-1}(C) \in \mathcal{S}$ for every $C \in \mathcal{C}$.*

Proof. Necessity is obvious. Suppose $g^{-1}(C) \in \mathcal{S}$ for every $C \in \mathcal{C}$, and denote

$$\mathcal{G} = \{B \in \mathcal{T} : g^{-1}(B) \in \mathcal{S}\}\,.$$

By assumption $\mathcal{G} \supseteq \mathcal{C}$, we need to show $\mathcal{G} \supseteq \mathcal{T}$. However note: i) $g^{-1}(\mathbb{T}) = \mathbb{S}$, ii) $g^{-1}(B) \in \mathcal{S}$ implies $g^{-1}(B^c) = (g^{-1}(B))^c \in \mathcal{S}$. And finally iii) $g^{-1}(\cup_i B_i) = \cup_i g^{-1}(B_i)$. Therefore $\mathcal{G}$ is σ-algebra, and therefore $\mathcal{G} \supseteq \mathcal{T} = \sigma(\mathcal{C})$

□

Corollary 1.8. *A mapping $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B})$ is a random variable if and only if $X^{-1}(-\infty, c] = \{X \le c\} \in \mathcal{F}$ for every $c \in \mathbb{R}$.*

Corollary 1.9. *Suppose $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}^k$, $Y : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}^l$ are two random vectors then the same holds for $(X, Y) : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}^{k+l}$.*

Corollary 1.10. *Suppose $X, Y : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ are random variables, then the same holds for $X \cdot Y$, $X + Y$ and cX for any $c \in \mathbb{R}$.*

Example 1.3 On $(\mathbb{S}, \mathcal{S})$, for any $A \in \mathcal{S}$, function $f(s) = \mathbb{1}_A(s)$ is measurable, the same is true for linear combinations of such functions by the exercise and the lemma above.

Suppose that $(\mathbb{S}, \mathcal{S})$ denotes a measurable space and that $f : \mathbb{S} \to \mathbb{R}$ is a function which has a finite range, i.e. the range $f(\mathbb{S})$ is a set of finite cardinality. Such a function is called *simple*. In that case $f(\mathbb{S}) = \{a_1, \ldots, a_n\}$ for some natural number n, and one can write $f = \sum_{i=1}^n a_i \mathbb{1}_{A_i}$ where $A_i = f^{-1}(a_i) \in \mathcal{S}$, and f is always measurable. Denote the set of simple functions by $\mathbb{K}_s = \{f : \mathbb{S} \to \mathbb{R} :\ f \text{ measurable and } |f(\mathbb{S})| = n \in \mathbb{N}\}$.

It turns out that any measurable function $f : \mathbb{S} \to \overline{\mathbb{R}}_+$ can be approximated by a sequence (f_n) in $\mathbb{K}_s$, so that $f_n \to f$ in all points of $\mathbb{S}$.

Lemma 1.11. *Assume function $f : \mathbb{S} \to \overline{\mathbb{R}}_+$ is measurable. Then there exists a sequence (f_n) in $\mathbb{K}_s$, such that $f_n(x) \to f(x)$ for each $x \in \mathbb{S}$.*

Proof. Fix an integer $n \in \mathbb{N}$ and let

$$a_k = a_{n,k} = \frac{k}{2^n}\,, \quad k = 0, 1, \ldots, 2^n n\,.$$

Let

$$f_n(x) = \begin{cases} a_k = \frac{k}{2^n} & \text{for } x \text{ such that } f(x) \in [a_k, a_{k+1}), k < 2^n n, \\ n & \text{for } x \text{ such that } f(x) \geq n. \end{cases}$$

Observe $(f_n(x))$ is nondecreasing real sequence for every x. It is smaller or equal than $f(x)$. On the other hand, if $f(x) < n$ we have $|f_n(x) - f(x)| \leq 1/2^n$. If $f(x) = \infty$, clearly $f_n(x) = n \nearrow \infty$. Thus $f_n(x) \nearrow f(x)$ in all $x \in \mathbb{S}$. □

Remark 1.4 In particular, the lemma claims that every nonnegative random variable can be approximated by a sequence of simple simple random variables.

Integrals

Suppose that $(\mathbb{S}, \mathcal{S}, \mu)$ denotes a measure space and that $f : \mathbb{S} \to \mathbb{R}$ is a function, we aim to define $\int f d\mu$. Consider first a simple function $f \in \mathbb{K}_s$, since $f(\mathbb{S}) = \{a_1, \ldots, a_n\}$ for some n, they have a representation

$$f = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$$

where $A_i = f^{-1}(a_i)$. When such a function is also nonnegative, i.e. if all $a_i \geq 0$ above, we write $f \in \mathbb{K}_{s+}$ and define

$$\int f d\mu := \sum_{i=1}^{n} a_i \mu(A_i) \,.$$

To ease the notation, or to specify the space and the variable of integration, we sometimes write $\int f d\mu = \mu f = \int_{\mathbb{S}} f(s)\mu(ds)$. One can show that if $f = \sum_{j=1}^{m} b_j \mathbb{1}_{B_j}$ for some other choice of values b_j and sets B_j, then also $\sum_{i=1}^{n} a_i \mu(A_i) = \sum_{j=1}^{m} b_j \mu(B_j)$, this is very easy to show if B_j's are disjoint, but holds in general as well. As a consequence, definition of $\int f d\mu$ does not depend on the representation.

One can further show that this integral has several nice properties, which are useful in the sequel: suppose $f, g \in \mathbb{K}_{s+}$, $c \geq 0$ then $\int cf d\mu = c \int f d\mu$, $\int (f+g) d\mu = \int f d\mu + \int g d\mu$, and

$$\text{if } f \leq g \text{ then } \int f d\mu \leq \int g d\mu. \tag{1.2}$$

Moreover, for a fixed $f \in \mathbb{K}_{s+}$ a function on $\mathcal{S}$ given by

$$C \mapsto \int f \mathbb{1}_C d\mu \tag{1.3}$$

is a new measure on $(\mathbb{S}, \mathcal{S})$.

Exercise 1.15 Prove that (1.2) holds and that (1.3) defines a measure.

Denote next $\mathbb{K}_+ = \{f : \mathbb{S} \to \overline{\mathbb{R}}_+ : \ f \text{ measurable}\}$. For $f \in \mathbb{K}_+$ we denote

$$\int f d\mu := \sup\left\{\int \varphi d\mu : \varphi \in \mathbb{K}_{s+}, \varphi \le f\right\}.$$

Check that the integrals of simple nonnegative functions coincide in two definitions. Moreover, we again have $\int f d\mu \le \int g d\mu$ if $f \le g$ and $\int cf d\mu = c\int f d\mu$ for $c \ge 0$.

Theorem 1.12 (Monotone convergence theorem). *Suppose $f_n \in \mathbb{K}_+, n \in \mathbb{N}$, if $f_n \nearrow f$, then*

$$\int f d\mu = \lim_n \int f_n d\mu\,.$$

Proof. Note, as a limit of nonnegative measurable functions $f \in \mathbb{K}_+$. Clearly $\int f_n d\mu \le \int f d\mu$, showing $\int f d\mu \ge \lim_n \int f_n d\mu$. Fix now any $\varphi \in \mathbb{K}_{s+}$ and $\alpha \in (0,1)$, $\varphi \le f$, let $F_n = \{x : f_n(x) \ge \alpha\varphi(x)\}$. Note, $F_n \in \mathcal{S}$ and $F_n \nearrow \mathbb{S}$

$$\int f_n d\mu \ge \int f_n \mathbb{1}_{F_n} d\mu \ge \alpha \int \varphi \mathbb{1}_{F_n} d\mu.$$

By (1.3) $\alpha \int \varphi \mathbb{1}_{F_n} d\mu \nearrow \alpha \int \varphi \mathbb{1}_{\mathbb{S}} d\mu = \alpha \int \varphi d\mu$. In particular,

$$\lim_n \int f_n d\mu \ge \alpha \int f d\mu$$

since $\alpha \in (0,1)$ was arbitrary $\int f d\mu \le \lim_n \int f_n d\mu$. □

MCT is useful because it provides a way of calculating integrals if we cannot determine the supremum in the definition of the integral directly. It also gives the following.

Theorem 1.13. *Suppose (f_n) is a finite or countable sequence in $\mathbb{K}_+$ and $f = \sum_n f_n$, then*

$$\int f d\mu = \sum_n \int f_n d\mu\,.$$

Proof. Take two sequences $\varphi_n^j \in \mathbb{K}_{s+}$ $j = 1,2$ such that $\varphi_n^j \nearrow f_j$. By the additivity of integrals of simple functions and previous theorem

$$\int (f_1 + f_2) d\mu = \lim_n \int (\varphi_n^1 + \varphi_n^2) d\mu = \lim_n \int \varphi_n^1 d\mu + \lim_n \int \varphi_n^2 d\mu = \int f_1 d\mu + \int f_2 d\mu.$$

Use the induction and $\sum_{n=1}^\infty f_n = \lim_N \sum_{n=1}^N f_n$ to finish the proof. □

Proposition 1.14. *Suppose $f \in \mathbb{K}_+$. Then $\int f d\mu = 0$ if and only if $f = 0$ μ–almost everywhere.*

Proof. Recall $f = 0$ μ-a.e. means $\mu\{x : f(x) \neq 0\} = 0$. For $f = \sum_i a_i \mathbb{1}_{A_i} \in \mathbb{K}_{s+}$, $\int f d\mu = 0 = \sum_i a_i \mu(A_i)$ if and only if $a_i\mu(A_i) = 0$ for every i, which gives the claim.

Sufficiency: suppose $f = 0$ μ a.e., then for any $\varphi \in \mathbb{K}_{s+}$, $\varphi \leq f$ also $\varphi = 0$ μ-a.e. Thus $\int f d\mu = \sup_{\varphi \leq f} \int \varphi d\mu = 0$.

Necessity: Let $F_n = \{x : f(x) > 1/n\} \nearrow F = \{x : f(x) > 0\} = \cup_n F_n$. If $\mu(F) > 0$ then also $\mu(F_n) > 0$ for some n. However, then

$$\int f d\mu \geq \frac{1}{n}\mu(F_n) > 0$$

giving a contradiction.

□

Corollary 1.15. *Suppose $f, g \in \mathbb{K}_+$ satisfy $f = g$ μ-a.e. then $\int f d\mu = \int g d\mu$.*

Proof. Note $E = \{x : f(x) \neq g(x)\} \in \mathcal{S}$, and $f = f1_E + f1_{E^c}$, where both summands are in $\mathbb{K}_{s+}$. By theorem 1.13 and the proposition above

$$\int f d\mu = \int f1_E d\mu + \int f1_{E^c} d\mu = \int g1_E d\mu = \int g d\mu.$$

□

Lemma 1.16 (Fatou)**.** *Suppose (f_n) is a sequence in $\mathbb{K}_+$ then $\int \liminf_n f_n d\mu \leq \liminf_n \int f_n d\mu$.*

Proof. Recall, for a nonnegative sequence (a_n) we have $\liminf_n a_n = \sup_n \inf_{k\geq n} a_k = \lim_n \inf_{k\geq n} a_k$. Observe, $\inf_{k\geq n} f_k \leq f_j$ for every $j \geq n$ and thus $\int \inf_{k\geq n} f_k d\mu \leq \int f_j d\mu$ for such j. Therefore

$$\int \inf_{k\geq n} f_k d\mu \leq \inf_{j\geq n} \int f_j d\mu$$

Using MCT, this yields

$$\begin{aligned}\int \liminf_n f_n d\mu &= \int \lim_n \inf_{k\geq n} f_k d\mu = \lim_n \int \inf_{k\geq n} f_k d\mu \\ &\leq \lim_n \inf_{k\geq n} \int f_k d\mu = \liminf_n \int f_n d\mu\end{aligned}$$

□

Exercise 1.16 Suppose that μ and ν are two measures on $(\mathbb{S}, \mathcal{S})$, we have seen that $\mu + \nu$ is also a measure. If $f \in \mathbb{K}_+$, show in detail that $\int f d(\mu + \nu) = \int f d\mu + \int f d\nu$.

Exercise 1.17 Suppose $f \in \mathbb{K}_+$ and $\int f d\mu < \infty$, then $\mu\{x : f(x) = \infty\} = 0$ and $\mu\{x : f(x) > 1/n\} < \infty$ for any $n \in \mathbb{N}$.

Take now an arbitrary function $f : \mathbb{S} \to \overline{\mathbb{R}}$, if it is measurable, the same holds for $f^+ = f \vee 0$ and $f^- = (-f) \vee 0$. Clearly $f = f^+ - f^-$ and $|f| = f^+ + f^-$. Observe, integrals $\int f^+ d\mu$ and $\int f^- d\mu$ are always well defined. Provided that at least one of them is finite, we define

$$\int f d\mu := \int f^+ d\mu - \int f^- d\mu.$$

They are both finite if and only if $\int |f| d\mu < \infty$, denote the class of such functions by

$$L^1 = L^1(\mathbb{S}, \mathcal{S}, \mu) = \{f : \mathbb{S} \to \overline{\mathbb{R}} \text{ measurable } : \int |f| d\mu < \infty\}$$

Proposition 1.17. *The set L^1 is a real vector space with the standard addition and scalar multiplication. Moreover, $\int (af + bg) d\mu = a \int f d\mu + b \int g d\mu$, for all $f, g \in L^1$ and all $a, b \in \mathbb{R}$.*

Proof. Homogeneity of the integral is clear from our earlier arguments. To show additivity, note $h = f + g$ satisfies $h = h^+ - h^- = f^+ - f^- + g^+ - g^-$, so $h^+ + f^- + g^- = h^- + f^+ + g^+$, with all six functions in $\mathbb{K}_+$, therefore we can apply the integrals and rearrange them again to get $\int h^+ d\mu - \int h^- d\mu = \int f^+ d\mu - \int f^- d\mu + \int g^+ d\mu - \int g^- d\mu$. □

If $f \in L^1$, then $|\int f d\mu| = |\int f^+ d\mu - \int f^- d\mu| \le \int f^+ d\mu + \int f^- d\mu = \int |f| d\mu$, yielding.

Lemma 1.18. *If $f \in L^1$ then $|\int f d\mu| \le \int |f| d\mu$.*

Exercise 1.18 For $f \in L^1$, we can write $\{x : f(x) \neq 0\} = \cup_n B_n$ with $\mu(B_n) < \infty$ (hint: $B_n = \{x : |f(x)| > 1/n\}$).

Proposition 1.19. *Suppose $f, g \in L^1$, then the following three claims are equivalent: i) $\int_E f d\mu = \int_E g d\mu$ for all $E \in \mathcal{S}$; ii) $\int_E |f - g| d\mu = 0$ and iii) $f = g$ μ–a.e.*

Proof. Equivalence between ii) and iii) is a consequence of Proposition 1.14. To show ii) implies i), note

$$\left| \int_E f d\mu - \int_E g d\mu \right| = \left| \int_E (f - g) d\mu \right| \le \int_E |f - g| d\mu \le \int |f - g| d\mu = 0.$$

Finally, to prove i) implies iii) suppose $f = g$ μ–a.e. does not hold. Denote $u = f - g$, then for some E, $\mu(E) > 0$, $u^+ > 0$ and $u^- = 0$ on E, or vice versa. Assume the first case (without loss of generality), then

$$\int_E (f - g) d\mu = \int (f - g) \mathbb{1}_E d\mu = \int_E f d\mu - \int_E g d\mu > 0$$

which is a contradiction with i).

□

A particularly useful result of the integration theory is the following.

Theorem 1.20 (Dominated convergence theorem)**.** *Suppose that for $(f_n) \in L^1$ and some measurable function f we have: a) $f_n \to f$ μ–a.e. and b) there exists $g \geq 0 \in L^1$ such that $|f_n| \leq g$ μ–a.e. for all $n \in \mathbb{N}$, then $f \in L^1$ and*

$$\int f d\mu = \lim_n \int f_n d\mu.$$

Proof. Check that the assumption b) implies that there exists E_1 with $\mu(E_1^c) = 0$ such that $|f_n| \leq g$ on E_1 for all $n \in \mathbb{N}$ simultaneously. Also, there exists E_2 with $\mu(E_2^c) = 0$ and $f_n \to f$ on E_2. Observe that $E = E_1 \cap E_2$ satisfies $\mu(E^c) = 0$ as well. On E, $g + f_n$ and $g - f_n$ are both nonnegative. By the last proposition above, Fatou lemma and linearity of the integral on L^1

$$\begin{aligned}\int (g+f)d\mu &= \int (g+f)\mathbb{1}_E d\mu \leq \liminf_n \int (g+f_n)\mathbb{1}_E d\mu \\ &= \liminf_n \int (g+f_n)d\mu = \int g d\mu + \liminf_n \int f_n d\mu.\end{aligned}$$

Thus $\int f d\mu \leq \liminf_n \int f_n d\mu$. Applying the same reasoning on $(g - f)$, we obtain the second inequality below

$$\liminf_n \int f_n d\mu \geq \int f d\mu \geq \limsup_n \int f_n d\mu,$$

to end the argument. □

Exercise 1.19 Consider measurable space $((0,1], \mathcal{B}((0,1]), \lambda)$ and functions $X_n(u) = n^2 \mathbb{1}_{(0,\frac{1}{n}]}(u)$. Determine $X(u) = \lim_{n\to\infty} X_n(u)$ for all $u \in (0,1]$ together with values $\int X d\lambda$ and $\lim_{n\to\infty} \int X_n d\lambda$.

Exercise 1.20 Provided that a measure μ is complete on the space $(\mathbb{S}, \mathcal{S})$, one can remove the requirement that f is measurable in the theorem above.

Recall the notion of Lebesgue measure λ on $\mathbb{R}$ or $(0,1]$ say. There are two σ–algebras we introduced in that setting: $\mathcal{B}$, consisting of Borel sets or $\mathcal{M}$, consisting of Lebesgue measurable sets. It is known that $\mathcal{B} \subsetneqq \mathcal{M} \subsetneqq \mathcal{P}(\mathbb{S})$. Furthermore, λ is not complete on the first one, and not well defined on the last one. It turns out that $\mathcal{M}$ is exactly the completion of $\mathcal{B}$ with respect to λ. For nice bounded functions $f : [a,b] \to \mathbb{R}$ we are familiar with the notion of Riemann integral, it is interesting that whenever such an integral exists it coincides with Lebesgue integral thus we write

$$\int_{[a,b]} f d\lambda = \int_a^b f(x)dx\,.$$

The equality of two integrals stems from the fact that the so-called Darboux sums in the definition of the Riemann integral can be viewed as integrals of simple functions with respect to the Lebesgue measure. For a precise proof: assume $f : [a,b] \to \mathbb{R}$ is a bounded, Riemann integrable function. Then there exists a sequence of finite *nested* partitions

$\{t_i^n\} \subseteq \{t_i^{n+1}\}$ of $[a,b]$ such that $\max_i |t_i^n - t_{i-1}^n| \to 0$, $a = t_0^n < t_1^n < \ldots < t_n^n = b$, so that for $I_j^n = (t_i^n, t_{i+1}^n]$ and $m_j^n = \inf_{I_j^n} f$, $M_j^n = \sup_{I_j^n} f$

$$s^n = \sum_j m_j^n |I_j^n| \nearrow J \quad \text{and} \quad S^n = \sum_j M_j^n |I_j^n| \searrow J\,,$$

where $J = \int_a^b f(s)ds$ in the Riemann sense, and $|I| = \lambda(I)$ denotes the length of the corresponding partition interval. In the standard definition, one typically takes m_j^n and M_j^n to be infimum and supremum of f over the closure of I_j^n. However, such infimum/supremum can be only smaller/larger, and since they converge to J, the same holds for the sequences $(s_n), (S_n)$ above.

Consider now

$$g^n = \sum_j m_j^n \mathbb{1}_{I_j^n} \le f \le G^n = \sum_j M_j^n \mathbb{1}_{I_j^n}, \tag{1.4}$$

as functions on the complete measure space $((a,b], \mathcal{F}^*, \lambda)$. By the boundedness and monotonicity, at a fixed point $x \in (a,b]$, both $g^n(x)$ and $G^n(x)$ converge to limits, denote them by $g(x)$ and $G(x)$. By the DCT

$$\int G d\lambda = \lim \int G^n d\lambda = \lim S^n = J = \lim s^n = \lim \int g^n d\lambda = \int g d\lambda$$

In particular $\int (G-g)d\lambda = 0$, since $G \ge g$, we have $\lambda(E) = 0$ for $E = \{x : G(x) > g(x)\}$. Since $g \le f \le G$, $f = G$ on E^c, because λ is complete on $\mathcal{F}^*$, $f|_{(a,b]}$ is measurable, and therefore the same holds for $f : [a,b] \to \mathbb{R}$. Finally by (1.4)

$$\int_{[a,b]} f(s)\lambda(ds) = \int_{(a,b]} f(s)\lambda(ds) = J = \int_a^b f(x)dx\,.$$

Example 1.4 Recall $f = \mathbb{1}_{\mathbb{Q}}$ is bounded nonnegative, but not Riemann integrable on $[0,1]$, still it is equal to 0 λ–a.e., so its Lebesgue integral exists and is equal to 0 as well.

Exercise 1.21 Think of a function on $[0,\infty)$ whose (improper) Riemann integral $\int_0^\infty f(x)dx$ exists, but $\int_{[0,\infty)} f d\lambda$ does not.

Expectations

Suppose $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B})$ is a random variable (i.e. measurable function). For $X^+ = X \vee 0$ and $X^- = (-X) \vee 0$, the following integrals

$$\int X^+ d\mathbb{P} \quad \text{and} \quad \int X^- d\mathbb{P}$$

are well defined. If at least one of them is finite, we define the *expectation* (or mean) of X as

$$\mathbb{E}X = \int X(\omega)\mathbb{P}(d\omega) = \int X^+ d\mathbb{P} - \int X^- d\mathbb{P}\,.$$

Note, X is integrable if both summands above are finite, or if $\mathbb{E}|X| = \int |X(\omega)|\mathbb{P}(d\omega) = \int X^+ d\mathbb{P} + \int X^- d\mathbb{P} < \infty$.

Exercise 1.22 Suppose $X : \Omega \to \overline{\mathbb{R}}_+$ is a random variable such that $\mathbb{E}X < \infty$, then $X < \infty$ almost surely, see Exercise 1.17.

Exercise 1.23 Suppose $X_1, X_2, \ldots : \Omega \to \overline{\mathbb{R}}_+$ are random variables, then $X = \sum_n X_n$ is also a random variable. Moreover $\mathbb{E} \sum_n X_n = \sum_n \mathbb{E}X_n$.

Exercise 1.24 Suppose $X_1, X_2, \ldots : \Omega \to \overline{\mathbb{R}}_+$ are random variables such that $\sum_n \mathbb{E}X_n < \infty$, then the random series $\sum_n X_n < \infty$ almost surely.

Exercise 1.25 Suppose $X_1, X_2, \ldots : \Omega \to \mathbb{R}$ are random variables such that $\sup_n \mathbb{E}|X_n| = M < \infty$, and suppose $|\varrho| < 1$, then the random series $\sum_n \varrho^n X_n$ converges almost surely.

Change of variable and change of measure

Suppose $(\mathbb{S}, \mathcal{S})$ and $(\mathbb{T}, \mathcal{T})$ are two measurable spaces, and $g : \mathbb{S} \to \mathbb{T}$ is a measurable function. Together with a measure μ on $(\mathbb{S}, \mathcal{S})$, g induces a measure on $(\mathbb{T}, \mathcal{T})$ as in

$$\mu_g(B) = \mu(g^{-1}(B)) .$$

Example 1.5 (Distribution) If $g = X : \Omega \to \mathbb{R}$ is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$ then the relation

$$P_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B)$$

induces a probability measure on $(\mathbb{R}, \mathcal{B})$, which we sometimes call the distribution (or law) of X.

Note now, if we have induced measure μ_g and a measurable function $f : \mathbb{T} \to \mathbb{R}$ we can try to calculate two integrals $\int (f \circ g) d\mu$ and $\int f d\mu_g$.

Lemma 1.21. *For any measurable f, g as above*

$$\int (f \circ g) d\mu = \int f d\mu_g .$$

Observe, the lemma also claims that if one of the integrals above is well defined, then the same holds for the other one.

Proof. Set $f = \mathbb{1}_A$, for $A \in \mathcal{T}$. Clearly

$$\mu(\mathbb{1}_A \circ g) = \int_{\mathbb{S}} \mathbb{1}_A \circ g(s) \mu(ds) = \int_{\mathbb{S}} \mathbb{1}_{g^{-1}(A)}(s) \mu(ds) = \mu(g^{-1}(A)) = \int_{\mathbb{T}} \mathbb{1}_A \mu_g(dt).$$

Thus, the integrals coincide for all simple functions f, and therefore for all nonnegative functions (by the MCT) and finally for all measurable functions by definition. □

Example 1.6 Consider $g = X : \Omega \to \mathbb{R}$, and $f(x) = x$, from the lemma above we get

$$\mathbb{E}X = \int_{\Omega} X d\mathbb{P} = \int_{\mathbb{R}} s dP_X(s) .$$

For the same reason, with $g = X$ and $f(x) = |x|^p$, we have for any $p \geq 0$

$$\mathbb{E}|X|^p = \int_{\Omega} X d\mathbb{P} = \int_{\mathbb{R}} |s|^p dP_X(s) .$$

Check that: if $f : \mathbb{S} \to \mathbb{R}_+$ is measurable on $(\mathbb{S}, \mathcal{S}, \mu)$, then

$$f \cdot \mu(A) = \int f \mathbb{1}_A d\mu$$

defines a new measure on $(\mathbb{S}, \mathcal{S})$. We write $\nu_f := f \cdot \mu$ and say that ν^f is constructed by the *change of measure*. Suppose that some other measure ν satisfies $\nu = f \cdot \mu$, show that the function f is μ-a.e. unique. In such a case, we have

$$\mu(A) = 0 \quad \text{implies} \quad \nu(A) = 0 \quad \text{for all } A \in \mathcal{S}\,. \tag{1.5}$$

Whenever measures μ and ν satisfy (1.5) we say that ν is *absolutely continuous* with respect to μ and write $\nu \ll \mu$.

Example 1.7 (Change of probability measure) Suppose $X : \Omega \to \mathbb{R}_+$ is a random variable such that $\mathbb{E}X = 1$. Then $\nu_X(A) = \mathbb{E}(X\mathbb{1}_A)$ for $A \in \mathcal{F}$, defines a new probability on $(\Omega, \mathcal{F})$.

If one can find a set $A \in \mathcal{S}$ such that $\mu(A) = 0$ and $\nu(A^c) = 0$, we say that μ and ν are *mutually singular* and write $\nu \perp \mu$.

Recall now that a measure μ on $(\mathbb{S}, \mathcal{S})$ is σ-finite, if there is a sequence (B_i) in $\mathcal{S}$, such that $\cup B_i = \mathbb{S}$ and $\mu(B_i) < \infty$ for each i.

Theorem 1.22 (Lebesgue/Radon-Nikodym). *Suppose that measures μ and ν on $(\mathbb{S}, \mathcal{S})$ are σ-finite, then there exist measures ν_a and ν_s such that*

i) $\nu_a \ll \mu$,

ii) $\nu_s \perp \mu$,

iii) $\nu = \nu_a + \nu_s$.

Moreover, for some μ-a.e. unique measurable function $f : \mathbb{S} \to \mathbb{R}$ we have $\nu_a = f \cdot \mu$.

In the case $\nu \ll \mu$, one can take $\nu_s = 0$ in the statement of the previous theorem. In that case, one can sometimes refer to f of the theorem as a *density* or *derivative* of ν with respect to μ, so we also write $f = d\nu/d\mu$.

Suppose $X : \Omega \to \mathbb{R}^d$ is a random vector, whose distribution P_X is absolutely continuous with respect to the Lebesgue measure, i.e. it has a density f_X say, then we call X a *continuous* random vector (or variable).

1.4 Product measures and Fubini theorem

Suppose $(\mathbb{S}, \mathcal{S})$ and $(\mathbb{T}, \mathcal{T})$ are two measurable spaces, on $\mathbb{S} \times \mathbb{T}$ we introduce the following σ-algebra generated by rectangles

$$\mathcal{S} \times \mathcal{T} = \sigma\{B \times C : B \in \mathcal{S},\ C \in \mathcal{T}\}\,.$$

We say that that $f : \mathbb{S} \times \mathbb{T} \to \mathbb{R}_+$ is measurable if it is $\mathcal{S} \times \mathcal{T}/\mathcal{B}$ measurable.

Lemma 1.23. *Suppose that a measure μ on $(\mathbb{S}, \mathcal{S})$ is σ-finite, and that $f : \mathbb{S} \times \mathbb{T} \to \mathbb{R}_+$ is measurable, then*

i) $s \mapsto f(s,t)$ is $\mathcal{S}$ measurable function from $\mathbb{S} \to \mathbb{R}_+$ for every fixed t.

ii) $t \mapsto \int_{\mathbb{S}} f(s,t)\mu(ds)$ is $\mathcal{T}$ measurable function from $\mathbb{T} \to \overline{\mathbb{R}}_+$.

Proof. Suppose μ is finite (without loss of generality, otherwise, we repeat the argument on each set B_i such that $\cup B_i = \mathbb{S}$ and $\mu(B_i) < \infty$ for each i). Suppose $f = \mathbb{1}_{B\times C}$, clearly: i) $f(s,t) = \mathbb{1}_B(s)\mathbb{1}_C(t)$, thus for $t \in C$, $f(s,t) = \mathbb{1}_B(s)$, otherwise $f(s,t) = 0$, in either case $s \mapsto f(s,t)$ is measurable. To show ii) note $\int \mathbb{1}_B(s)\mathbb{1}_C(t)\mu(ds) = \mathbb{1}_C(t)\mu(B)$ which is measurable as a function of t. Thus, i) and ii) hold for simple functions which are constant on rectangles, adapting the proof of Lemma 1.11, one can show that any nonnegative functions can be approximated by such simple functions. Therefore i) and ii) hold for all nonnegative functions. □

Theorem 1.24 (Lebesgue/Fubini/Tonelli)**.** *Suppose that measures μ and ν in $(\mathbb{S}, \mathcal{S}, \mu)$ and $(\mathbb{T}, \mathcal{T}, \nu)$ are σ-finite, then*

i) There exists a unique measure $\mu \times \nu$ on $(\mathbb{S} \times \mathbb{T}, \mathcal{S} \times \mathcal{T})$ such that $\mu \times \nu(B \times C) = \mu(B)\nu(C)$ for all $B \in \mathcal{S}$, $C \in \mathcal{T}$.

ii) For every measurable $f : \mathbb{S} \times \mathbb{T} \to \mathbb{R}_+$

$$\mu \times \nu(f) = \int_{\mathbb{S}} \mu(ds) \int_{\mathbb{T}} f(s,t)\nu(dt) = \int_{\mathbb{T}} \nu(dt) \int_{\mathbb{S}} f(s,t)\mu(ds)\,. \tag{1.6}$$

iii) Suppose $f : \mathbb{S} \times \mathbb{T} \to \mathbb{R}$ is measureble and $\mu \times \nu|f| < \infty$, then (1.6) still holds.

Note formula (1.6) claims that all three integrals therein are actually equal, we typically use the equality of the second and the third.

Proof. i) Define for an arbitrary $A \in \mathcal{S} \times \mathcal{T}$,

$$\mu \times \nu(A) = \int_{\mathbb{S}} \mu(ds) \int_{\mathbb{T}} \mathbb{1}_A(s,t)\nu(dt) \tag{1.7}$$

This is well defined measure, that satisfies property i) in the statement of the theorem. Suppose μ and ν are finite, observe that rectangles $B \times C$ form a π-system, there could be only one measure satisfying this property (otherwise, since the measures are σ-finite, we can write $\mathbb{S} = \cup S_i$, $\mu(S_i) < \infty$ and $\mathbb{T} = \cup T_i$, $\nu(T_i) < \infty$, $\mathbb{S} \times \mathbb{T} = \cup_{i,j} S_i \times T_j$ we can repeat the argument for the sets $S_i \times T_j$).

ii) By (1.7) and the linearity of integral, equality in (1.6) holds for all simple functions constant on rectangles, and since any measurable nonnegative function can be approximated by such simple functions as in Lemma 1.11, the MCT yields (1.6) for all nonnegative functions too.

iii) In general, if $\mu \times \nu|f| < \infty$, then

$$\begin{aligned}\int f d(\mu \times \nu) &= \int f^+ d(\mu \times \nu) - \int f^- d(\mu \times \nu) \\ &= \int_{\mathbb{S}} \mu(ds) \int_{\mathbb{T}} f^+(s,t)\nu(dt) - \int_{\mathbb{S}} \mu(ds) \int_{\mathbb{T}} f^-(s,t)\nu(dt) \\ &= \int_{\mathbb{S}} \mu(ds) \int_{\mathbb{T}} f(s,t)\nu(dt)\end{aligned}$$

if we interpret the integral at the end as the difference of the previous two integrals. Observe, the integrals $\int_{\mathbb{T}} f^+(s,t)\nu(dt)$ and $\int_{\mathbb{T}} f^-(s,t)\nu(dt)$ can be simultaneously infinite for some s. However, $\mu \times \nu|f| < \infty$ implies $\mu\{s : \int_{\mathbb{T}} f^+(s,t)\nu(dt) = \int_{\mathbb{T}} f^-(s,t)\nu(dt) = \infty\} = \mu\{s : \int_{\mathbb{T}} f^+(s,t)\nu(dt) + \int_{\mathbb{T}} f^-(s,t)\nu(dt) = \infty\} =: \mu(N_{\mathbb{S}}) = 0$. Therefore, the final integral in our calculation one could and should read as $\int_{N_{\mathbb{S}}^c} \mu(ds) \int_{\mathbb{T}} f(s,t)\nu(dt)$.

$\square$

Exercise 1.26 Suppose that $\mathbb{S} = \mathbb{S}' = (0,1]$ and $\mathcal{S} = \mathcal{S}' = \mathcal{B}$. Assume that $\mu =$ Lebesgue measure and that $\nu(\cdot) = \mathrm{card}(\cdot)$ represents the counting measure. Consider the diagonal set $D = \{(s,s) : s \in \mathbb{S}\}$. Show that D is measurable) and that for the indicator function $f(s,t) = \mathbb{1}_D(s,t)$ all three integrals in (1.6) take different value.

Exercise 1.27 Suppose that $\mathbb{S} = \mathbb{S}' = \mathbb{N}$ and $\mathcal{S} = \mathcal{S}' = \mathcal{P}(\mathbb{N})$. Assume that $\mu = \nu$ are both counting measures. Consider the function: $f(i,j) = 1$ if $i = j$; $f(i,j) = -1$ if $i = j + 1$; and $f(i,j) = 0$ otherwise. Show that $\mu \times \nu|f| = +\infty$, moreover the second and the third integral in (1.6) both exist, but are not equal.

Bibliography

[1] Billingsley, P., 1995. Probability and measure. Wiley.

[2] Durrett, R., 2019. Probability: theory and examples (Vol. 49). Cambridge University Press.

[3] Folland, G.B., 1999. Real analysis: modern techniques and their applications. Wiley.

[4] Kallenberg, O., 2021. Foundations of modern probability (3rd ed). Springer.

[5] Klenke, A., 2013. Probability theory: a comprehensive course. Springer.

[6] Sarapa, N., 2002. Teorija vjerojatnosti. Školska knjiga.