Probability Theory
Lecture Notes

Marcelo Gallardo

Semester 2024-1

Pontificia Universidad Católica del Perú

marcelo.gallardo@pucp.edu.pe

Abstract

These notes accompany the course *Probability and Statistics 1*, taught at the Pontificia Universidad Católica del Perú during the 2024-1 semester. They give an introduction to probability theory from the measure-theoretic point of view: probability spaces, random variables and random vectors, integration, independence, the classical convergence theorems, L^p spaces, characteristic functions and conditional expectation.

Conventions. Throughout, *r.v.* abbreviates *random variable* (or *random vector*, depending on context), *a.s.* abbreviates *almost surely*, and *i.i.d.* abbreviates *independent and identically distributed*. The symbol $\biguplus$ denotes a union of pairwise disjoint sets. We write $A - B$ and $A \setminus B$ interchangeably for set difference. The monotone and dominated convergence theorems are abbreviated MCT and DCT, respectively.

Contents

1 Lecture 1

Remark. As motivation for the axiomatic construction that follows, it is worth keeping in mind **Bertrand's paradox**: the statement "a chord is chosen at random in a circle" does not determine a probability, because different reasonable ways of formalizing "at random" lead to different answers. See [Le Gall(2022), p. 144]. The moral is that a probabilistic model is not determined by the sample space alone: one must also specify the measure.

1. Ω will denote the sample space. It may be finite, infinite, etc.; the only requirement is that $\Omega \neq \emptyset$. For instance, for the roll of a die, $\Omega = \{1, \ldots, 6\}$.

2. $\mathbb{P} : 2^\Omega \to [0, 1]$ (when the sample space is finite or countable). This function (a probability measure, or simply a probability) satisfies a number of properties.

   (a) $\mathbb{P}(\Omega) = 1$

   (b) $\mathbb{P}(\emptyset) = 0$

   (c) Given $A \subset \Omega$, $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$

   (d) (σ-additivity) If $A_1, A_2, A_3, \ldots$ in 2^Ω are pairwise disjoint, then

   $$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

   The first two properties together with σ-additivity are the axioms; the remaining ones are consequences.

Remark. In general it is not always possible to work on 2^Ω (the largest σ-algebra one can consider on Ω). Indeed, already for the Lebesgue measure on $[0, 1]$ we cannot work with all subsets of $[0, 1]$. We therefore consider

$$\mathbb{P} : \mathscr{F} \to [0, 1]$$

where $\mathscr{F}$ is a σ-algebra on Ω.

Properties:

- If $A_1, \ldots, A_n \in \mathscr{F}$ are pairwise disjoint, then
$$\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \mathbb{P}(A_k).$$
This follows from the countably infinite case by taking $A_j = \emptyset$ for $j \geq n+1$.

- If $A, B \in \mathscr{F}$ and $A \subset B$, then
$$\mathbb{P}(A) + \mathbb{P}(B - A) = \mathbb{P}(B).$$
In particular, $\mathbb{P}(A) \leq \mathbb{P}(B)$.

- If $A, B \in \mathscr{F}$, then
$$\mathbb{P}(A \cup B) + \mathbb{P}(A \cap B) = \mathbb{P}(A) + \mathbb{P}(B).$$
This follows from
$$\begin{aligned} \mathbb{P}(A) + \mathbb{P}(B - A) &= \mathbb{P}(A \cup B) \\ \mathbb{P}(A \cap B) + \mathbb{P}(B - A) &= \mathbb{P}(B). \end{aligned}$$

- If $A_1, A_2, \dots \in \mathscr{F}$, then
$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} \mathbb{P}(A_k).$$
Proof. Consider the sets $B_1 = A_1$, $B_2 = A_2 - A_1$, $\ldots$, $B_k = A_k - \bigcup_{j=1}^{k-1} A_j$, and so on. Then $B_i \cap B_j = \emptyset$ for $i \neq j$. Moreover, $\bigcup_{j=1}^{k} A_j = \bigcup_{j=1}^{k} B_j$ (by induction on k). Hence
$$\bigcup_{j=1}^{\infty} A_j = \bigcup_{j=1}^{\infty} B_j,$$

and therefore

$$\begin{aligned}\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) &= \mathbb{P}\left(\bigcup_{j=1}^{\infty} B_j\right)\\ &= \sum_{j=1}^{\infty}\mathbb{P}(B_j)\\ &\le \sum_{j=1}^{\infty}\mathbb{P}(A_j),\end{aligned}$$

where the last inequality follows from $B_j \subset A_j$. $\square$

- If $A_1, A_2, \ldots, A_n \in \mathscr{F}$, then

$$\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) \le \sum_{k=1}^{n}\mathbb{P}(A_k).$$

- If $A_1, A_2, \dots \in \mathscr{F}$ and $A_n \uparrow A$[1], then $\lim_{n\to\infty}\mathbb{P}(A_n) = \mathbb{P}(A)$.

Proof. Set $B_1 = A_1$, $B_2 = A_2 - A_1$, $\ldots$, $B_k = A_k - A_{k-1}$, and so on. Clearly $B_i \cap B_j = \emptyset$ for $i \ne j$, and $\bigcup_{j=1}^{n} A_j = \bigcup_{j=1}^{n} B_j$ for every $n \in \mathbb{N}$. Therefore,

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}\left(\bigcup_{j=1}^{\infty} B_j\right)\\ &= \sum_{j=1}^{\infty}\mathbb{P}(B_j)\\ &= \lim_{n\to\infty}\sum_{j=1}^{n}\mathbb{P}(B_j)\\ &= \lim_{n\to\infty}\mathbb{P}(A_n).\end{aligned}$$

$\square$

[1] That is, $A_1 \subset A_2 \subset \cdots \subset A_{n-1} \subset A_n$ and $\bigcup_{j=1}^{\infty} A_j = A$.

Exercise 1.1. Let $A_1, A_2, \dots \in \mathscr{F}$ with $A_n \downarrow A$[2]. Show that $\lim_n \mathbb{P}(A_n) = \mathbb{P}(A)$. Note that, because we work with probability measures, there is no need to add the hypothesis $\mathbb{P}(A_1) < \infty$.

Proof. Set $B_n = A_1 - A_n$. Then:

- $B_n \cap A_n = \emptyset$;
- $A_1 = B_n \cup A_n$;
- $\mathbb{P}(A_1) = \mathbb{P}(B_n) + \mathbb{P}(A_n)$;
- $\mathbb{P}(A_1) - \mathbb{P}(B_n) = \mathbb{P}(A_n)$;
- $B_n \subset B_{n+1}$;
- since $B_n \subset B_{n+1}$, $\mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) = \lim_{n\to\infty} \mathbb{P}(B_n)$;
- moreover,
$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} (A_1 - A_n) = A_1 - \bigcap_{n=1}^{\infty} A_n.$$
Hence,
$$\begin{aligned} \lim_n \mathbb{P}(B_n) &= \mathbb{P}(A_1) - \mathbb{P}\left(\bigcap_{n\in\mathbb{N}} A_n\right) \\ \mathbb{P}\left(\bigcap_{n\in\mathbb{N}} A_n\right) &= \mathbb{P}(A_1) - \lim_n \mathbb{P}(B_n) \\ &= \lim_n \mathbb{P}(A_n). \end{aligned}$$

□

Inclusion–exclusion principle. Let $A_1, A_2, \dots, A_n \in \mathscr{F}$. We want to

[2] This means that $A_{j+1} \subset A_j$ and $\bigcap_{j\geq 1} A_j = A$.

compute the probability of the union:

$$\mathbb{P}\left(\bigcup_{j=1}^{n} A_j\right) = \sum_{j=1}^{n} \mathbb{P}(A_j) - \sum_{1\le i<j\le n} \mathbb{P}(A_i \cap A_j)$$
$$+ \sum_{1\le i<j<k\le n} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1}\mathbb{P}\left(\bigcap_{j=1}^{n} A_j\right).$$

Proof. The identity certainly holds for $n = 1, 2$. Assume it holds for $n = k$:

$$\mathbb{P}\left(\bigcup_{j=1}^{k} A_j\right) = \sum_{j=1}^{k} \mathbb{P}(A_j) - \sum_{1\le i<j\le k} \mathbb{P}(A_i \cap A_j)$$
$$+ \cdots + (-1)^{k+1}\mathbb{P}(A_1 \cap \cdots \cap A_k).$$

On the one hand, writing $B = \bigcup_{j=1}^{k} A_j$,

$$\mathbb{P}(B \cup A_{k+1}) = \mathbb{P}(B) + \mathbb{P}(A_{k+1}) - \mathbb{P}(B \cap A_{k+1}).$$

On the other hand,

$$\begin{aligned}
\mathbb{P}(B \cap A_{k+1}) &= \mathbb{P}\left(\bigcup_{j=1}^{k} (A_j \cap A_{k+1})\right) \\
&= \sum_{j=1}^{k} \mathbb{P}(A_j \cap A_{k+1}) + \cdots \\
&+ (-1)^{k+1}\mathbb{P}((A_1 \cap A_{k+1}) \cap (A_2 \cap A_{k+1}) \cap \cdots \cap (A_k \cap A_{k+1})) \\
&= \sum_{j=1}^{k} \mathbb{P}(A_j \cap A_{k+1}) + \cdots + (-1)^{k+1}\mathbb{P}\left(A_1 \cap \cdots \cap A_{k+1}\right).
\end{aligned}$$

Therefore,

$$
\begin{aligned}
\mathbb{P}(B \cup A_{k+1}) &= \mathbb{P}(B) + \mathbb{P}(A_{k+1}) \\
&\quad - \left[\sum_{j=1}^{k} \mathbb{P}(A_j \cap A_{k+1}) + \cdots + (-1)^{k+1} \mathbb{P}\left(A_1 \cap \cdots \cap A_{k+1}\right) \right] \\
&= \sum_{j=1}^{k} \mathbb{P}(A_j) - \sum_{1 \le i < j \le k} \mathbb{P}(A_i \cap A_j) + \cdots \\
&\quad + (-1)^{k+1} \mathbb{P}(A_1 \cap \cdots \cap A_k) + \mathbb{P}(A_{k+1}) \\
&\quad - \left[\sum_{j=1}^{k} \mathbb{P}(A_j \cap A_{k+1}) + \cdots + (-1)^{k+1} \mathbb{P}\left(A_1 \cap \cdots \cap A_{k+1}\right) \right] \\
&= \sum_{j=1}^{k+1} \mathbb{P}(A_j) + \cdots + (-1)^{k+2} \mathbb{P}(A_1 \cap \cdots \cap A_{k+1}).
\end{aligned}
$$

□

Remark. The inclusion–exclusion principle can be rewritten as

$$
\mathbb{P}\left(\bigcup_{i=1}^{n} A_i \right) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{I \subset \{1,\ldots,n\},\ |I|=k} \mathbb{P}(A_I)
$$

with $A_I = \bigcap_{i \in I} A_i$. If we write

$$
q_k = \sum_{I \subset \{1,\ldots,n\},\ |I|=k} \mathbb{P}(A_I), \qquad k = 1, \ldots, n,
$$

we obtain the *Bonferroni inequalities*: truncating the alternating sum at a positive term gives an upper bound, and truncating it at a negative term gives a lower bound. That is,

$$
\mathbb{P}\left(\bigcup_{i=1}^{n} A_i \right) \le q_1, \qquad \mathbb{P}\left(\bigcup_{i=1}^{n} A_i \right) \ge q_1 - q_2, \qquad \mathbb{P}\left(\bigcup_{i=1}^{n} A_i \right) \le q_1 - q_2 + q_3,
$$

and so on, with equality reached at step n.

2 Lecture 2

2.1 σ-algebras and probability measures

Definition 2.1. Let $\Omega \neq \emptyset$. We say that $\mathscr{F} \subset 2^\Omega$ is a σ-algebra on Ω if

1. $\Omega \in \mathscr{F}$;
2. $A \in \mathscr{F}$ implies $A^c \in \mathscr{F}$;
3. $A_1, A_2, \dots \in \mathscr{F}$ implies $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{F}$.

Remark. If $A, B \in \mathscr{F}$, then $A \cap B \in \mathscr{F}$, since $A \cap B = (A^c \cup B^c)^c$.[3]

Definition 2.2. A probability space is a triple $(\Omega, \mathscr{F}, \mathbb{P})$ consisting of

1. a nonempty set Ω, called the sample space;
2. a σ-algebra $\mathscr{F}$ on Ω[4], whose elements are called events;
3. a probability measure $\mathbb{P}$ on $(\Omega, \mathscr{F})$, in the sense of the next definition.

Definition 2.3. A probability measure is a function $\mathbb{P} : \mathscr{F} \to \mathbb{R}$ such that

1. $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$;
2. $0 \leq \mathbb{P}(A) \leq 1$ for all $A \in \mathscr{F}$;
3. if $A_1, A_2, \dots \in \mathscr{F}$ are pairwise disjoint, then
$$\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(A_k).$$

Example 2.4. Let $\Omega = \{\omega_1, \dots, \omega_n\}$ or $\Omega = \{\omega_1, \omega_2, \dots\}$. Take $\mathscr{F} = 2^\Omega$ and $p_1, \dots, p_n \in [0, 1]$ with $\sum_{k=1}^n p_k = 1$ (respectively $\sum_{k=1}^\infty p_k = 1$ in the second case). Then $\mathbb{P} : \mathscr{F} \to \mathbb{R}$ given by $\mathbb{P}(A) = \sum_{\omega_k \in A} p_k$ is a probability measure.

---

[3]Indeed, $A^c, B^c \in \mathscr{F}$ (because A and B are), so $A^c \cup B^c \in \mathscr{F}$ by property (3), and one more application of (2) gives $(A^c \cup B^c)^c \in \mathscr{F}$.

[4]A nonempty family such that $A \in \mathscr{F} \implies A^c \in \mathscr{F}$ and, if $A_1, A_2, \dots \in \mathscr{F}$, then $\bigcup_{n\in\mathbb{N}} A_n \in \mathscr{F}$.

Proof. We check the conditions one by one (for the infinite case; the finite case is entirely analogous):

1. $\mathbb{P}(\Omega) = \sum_{\omega_k \in \Omega} p_k = \sum_{k=1}^{\infty} p_k = 1$.
2. $\mathbb{P}(\emptyset) = \sum_{\omega_k \in \emptyset} p_k = 0$ (empty sum).
3. If $A = \biguplus_{n=1}^{\infty} A_n$, then

$$\sum_{\omega_k \in A} p_k = \sum_{n=1}^{\infty} \sum_{\omega_k \in A_n} p_k = \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

□

Remark. Note that, for every A, the series

$$\sum_{\omega_k \in A} p_k$$

converges absolutely, since $\sum_{\omega_k \in A} p_k \leq \sum_{\omega_k \in \Omega} p_k = 1$. This allows us to rearrange its terms. More generally, a series with nonnegative terms can always be rearranged.

Example 2.5. Let $\Omega = \biguplus_{k=1}^{n} A_k$ (the A_k pairwise disjoint with union Ω). Consider

$$\mathscr{F} = \left\{ A_J = \bigcup_{j \in J} A_j : \; J \subset \{1, \dots, n\} \right\}$$

and its countable analogue, for $\Omega = \biguplus_{k \in \mathbb{N}} A_k$,

$$\mathscr{F} = \left\{ A_J = \bigcup_{j \in J} A_j : \; J \subset \mathbb{N} \right\}.$$

First of all, these are σ-algebras on Ω. Next, let $p_1, \dots, p_n \geq 0$ with $\sum_{k=1}^{n} p_k = 1$ (analogously $p_1, p_2, \dots \geq 0$ with $\sum_{k \in \mathbb{N}} p_k = 1$). Then

$$\mathbb{P}(A_J) = \sum_{j \in J} p_j$$

is a probability measure on $(\Omega, \mathscr{F})$.

Proof. We first show that $\mathscr{F}$ is a σ-algebra:

1. Taking $J = \mathbb{N}$ (or $J = \{1, \dots, n\}$), the corresponding union is Ω. Hence $\Omega \in \mathscr{F}$.

2. Let $B \in \mathscr{F}$; we may write
$$B = \bigcup_{j \in J} A_j.$$
Since the A_j form a partition of Ω,
$$B^c = \bigcup_{j \notin J} A_j = \bigcup_{j \in J^c} A_j \in \mathscr{F},$$
because $J^c \subset \{1, \dots, n\}$ (respectively $J^c \subset \mathbb{N}$).

3. Let $B_1, B_2, \dots \in \mathscr{F}$ with $B_i = \bigcup_{j \in J_i} A_j$. Setting $\tilde{J} = \bigcup_{i \geq 1} J_i$,
$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} \bigcup_{j \in J_i} A_j = \bigcup_{j \in \tilde{J}} A_j \in \mathscr{F}.$$

Finally, checking that $\mathbb{P}$ is a probability measure is straightforward. Indeed, $J \subset \mathbb{N}$ and $p_j \geq 0$, so $0 \leq \mathbb{P}(A_J) \leq \sum_k p_k = 1$. The σ-additivity follows from the fact that the index sets of a disjoint union of elements of $\mathscr{F}$ partition the index set, together with the fact that a series of nonnegative terms may be rearranged freely. $\square$

Example 2.6. Let $(\Omega, \mathscr{F})$ be a measurable space and $x_0 \in \Omega$. The function $\delta_{x_0} : \mathscr{F} \to \mathbb{R}$ given by
$$\delta_{x_0}(A) = \begin{cases} 1, & \text{if } x_0 \in A, \\ 0, & \text{otherwise}, \end{cases}$$
is a probability measure on $(\Omega, \mathscr{F})$, known as the Dirac measure on $(\Omega, \mathscr{F})$ concentrated at the point x_0.

Proof. We check the properties:

1. $x_0 \notin \emptyset$, hence $\delta_{x_0}(\emptyset) = 0$.
2. $x_0 \in \Omega$, hence $\delta_{x_0}(\Omega) = 1$.
3. By definition, $\delta_{x_0}(A) \in [0, 1]$.
4. Let $A_1, A_2, \dots \in \mathscr{F}$ be pairwise disjoint. If $x_0 \in \biguplus_{n=1}^{\infty} A_n$, there is a unique n_0 with $x_0 \in A_{n_0}$, and $x_0 \notin A_n$ for $n \neq n_0$. Thus

$$\delta_{x_0}\left(\biguplus_{n=1}^{\infty} A_n\right) = 1 = \delta_{x_0}(A_{n_0}) = \sum_{n=1}^{\infty} \delta_{x_0}(A_n).$$

   If $x_0 \notin \biguplus_{n=1}^{\infty} A_n$, both sides equal 0.

□

Example 2.7. If P, Q are probability measures on $(\Omega, \mathscr{F})$ and $\alpha \in (0, 1)$, then $\alpha P + (1-\alpha)Q$ is also a probability measure on $(\Omega, \mathscr{F})$. More generally, if $P_1, \dots, P_n$ are probability measures and $t_1, \dots, t_n \in [0, 1]$ satisfy $\sum_{i=1}^{n} t_i = 1$, then $\sum_{i=1}^{n} t_i P_i$ is a probability measure as well. The same holds in the countable case.

Proof. We prove all items in the countably infinite case (which subsumes the others):

1. $\left(\sum_{i=1}^{\infty} t_i P_i\right)(\emptyset) = \sum_{i=1}^{\infty} t_i \cdot 0 = 0$.
2. $\left(\sum_{i=1}^{\infty} t_i P_i\right)(\Omega) = \sum_{i=1}^{\infty} t_i P_i(\Omega) = \sum_{i=1}^{\infty} t_i = 1$.

3. If $A = \biguplus_{j=1}^{\infty} A_j$, then

$$\begin{aligned}\left(\sum_{i=1}^{\infty} t_i P_i\right)(A) &= \left(\sum_{i=1}^{\infty} t_i P_i\right)\left(\biguplus_{j=1}^{\infty} A_j\right)\\ &= \sum_{i=1}^{\infty} t_i \left[\sum_{j=1}^{\infty} P_i(A_j)\right]\\ &= \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} t_i P_i(A_j)\\ &= \sum_{j=1}^{\infty}\left(\sum_{i=1}^{\infty} t_i P_i\right)(A_j),\end{aligned}$$

where the interchange of sums is legitimate because all terms are non-negative.

□

2.2 Random variables and random vectors

Definition 2.8. We say that $X : \Omega \to \mathbb{R}^n$ is

1. a *random variable*, if $n = 1$;
2. a *random vector*, if $n \geq 2$,

provided it is measurable with respect to the corresponding σ-algebras. When the distinction is immaterial we shall also say "random vector" in the case $n = 1$.

Definition 2.9. *Generated σ-algebra.* Let Ω be a nonempty set and $\mathscr{C}$ a family of subsets of Ω. Let $\{\mathscr{F}_\lambda\}_{\lambda\in\Lambda}$ be the family of all σ-algebras on Ω containing $\mathscr{C}$; it is nonempty, since 2^Ω is one of them. We define

$$\sigma(\mathscr{C}) = \bigcap_{\lambda\in\Lambda} \mathscr{F}_\lambda,$$

the smallest σ-algebra on Ω containing $\mathscr{C}$.

Remark. Note that $\sigma(\mathscr{C})$ is indeed a σ-algebra, since an intersection of a family of σ-algebras is a σ-algebra, and it contains $\mathscr{C}$. Moreover, if $\mathscr{C}_1 \subset \mathscr{C}_2$, then $\sigma(\mathscr{C}_1) \subset \sigma(\mathscr{C}_2)$.

Borel sets.

$$\mathscr{B}(\mathbb{R}) = \sigma(\text{open intervals of } \mathbb{R})$$
$$\mathscr{B}(\mathbb{R}^n) = \sigma(\text{open sets of } \mathbb{R}^n).$$

More generally, if $(E, \mathcal{T})$ is a topological space, then

$$\mathscr{B}(E) = \sigma(\{\text{open sets of } E\}).$$

Proposition 2.10. Let

1. $\mathcal{C}_1 = \{(-\infty, a] : a \in \mathbb{R}\}$ and $\mathcal{C}_1^\circ = \{(-\infty, a] : a \in \mathbb{Q}\}$,
2. $\mathcal{C}_2 = \{(-\infty, a) : a \in \mathbb{R}\}$,
3. $\mathcal{C}_3 = \{(a, \infty) : a \in \mathbb{R}\}$,
4. $\mathcal{C}_4 = \{[a, \infty) : a \in \mathbb{R}\}$.

Then

$$\sigma(\mathcal{C}_1^\circ) = \sigma(\mathcal{C}_1) = \sigma(\mathcal{C}_2) = \sigma(\mathcal{C}_3) = \sigma(\mathcal{C}_4) = \mathscr{B}(\mathbb{R}).$$

On the other hand, if for $a = (a_1, \dots, a_n) \in \mathbb{R}^n$ we write $(-\infty, a] = \prod_{i=1}^n (-\infty, a_i]$ and set

$$\mathcal{C} = \{(-\infty, a] : \ a \in \mathbb{R}^n\},$$

then

$$\sigma(\mathcal{C}) = \mathscr{B}(\mathbb{R}^n).$$

2.3 Inverse image (preimage) of a set under a function

Let $f : A \to B$ be a function and $C \subset B$. The inverse image of C under f is the set

$$f^{-1}(C) = \{x \in A : f(x) \in C\} \subset A.$$

Let C and C_λ, $\lambda \in \Lambda$, be subsets of B. Then

1. $f^{-1}\left(\bigcup_{\lambda\in\Lambda} C_\lambda\right) = \bigcup_{\lambda\in\Lambda} f^{-1}(C_\lambda)$;
2. $f^{-1}\left(\bigcap_{\lambda\in\Lambda} C_\lambda\right) = \bigcap_{\lambda\in\Lambda} f^{-1}(C_\lambda)$;
3. $f^{-1}(C^c) = [f^{-1}(C)]^c$;
4. $f^{-1}(\emptyset) = \emptyset$ and $f^{-1}(B) = A$.

3 Lecture 3

Definition 3.1. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ a function. We say that X is a random variable when $X^{-1}(A) \in \mathscr{F}$ for every $A \in \mathscr{B}(\mathbb{R})$; in symbols,

$$X : (\Omega, \mathscr{F}) \to (\mathbb{R}, \mathscr{B}_{\mathbb{R}}).$$

This extends to $\mathbb{R}^n$ in the obvious way.

Notation. Let $X : \Omega \to \mathbb{R}^n$ be a random vector and $A \in \mathscr{B}(\mathbb{R}^n)$. We write

$$\{X \in A\} = \{\omega \in \Omega : X(\omega) \in A\} = X^{-1}(A).$$

Notation. Let $X : \Omega \to \mathbb{R}$ be a random variable and $a, b \in \mathbb{R}$. Then

$$\begin{aligned}
\{X = a\} &= \{\omega \in \Omega : X(\omega) = a\} = X^{-1}(\{a\}) \\
\{X < a\} &= \{\omega \in \Omega : X(\omega) < a\} = X^{-1}((-\infty, a)) \\
\{X \leq a\} &= \{\omega \in \Omega : X(\omega) \leq a\} = X^{-1}((-\infty, a]) \\
\{X > a\} &= \{\omega \in \Omega : X(\omega) > a\} = X^{-1}((a, \infty)) \\
\{X \geq a\} &= \{\omega \in \Omega : X(\omega) \geq a\} = X^{-1}([a, \infty)) \\
\{a \leq X \leq b\} &= \{\omega \in \Omega : a \leq X(\omega) \leq b\} = X^{-1}([a, b]).
\end{aligned}$$

Example 3.2. Let $\Omega = \{a, b, c, d\}$ and $\mathscr{F} = \{\emptyset, \Omega, \{a, b\}, \{c, d\}\}$. Let $X, Y : \Omega \to \mathbb{R}$ be given by

$$X(a) = X(b) = 1, \quad X(c) = X(d) = 2$$

$$Y(a) = Y(c) = -1, \quad Y(b) = Y(d) = 0.$$

Then X is a random variable but Y is not. Indeed, for $A \in \mathscr{B}(\mathbb{R})$,

$$X^{-1}(A) = \begin{cases} \Omega & \text{if } 1, 2 \in A, \\ \{a, b\} & \text{if } 1 \in A,\ 2 \notin A, \\ \{c, d\} & \text{if } 1 \notin A,\ 2 \in A, \\ \emptyset & \text{if } 1, 2 \notin A. \end{cases}$$

On the other hand, $Y^{-1}(\{0\}) = \{b, d\} \notin \mathscr{F}$. Note that if we change the σ-algebra, taking for instance $\{\emptyset, \Omega, \{a, c\}, \{b, d\}\}$, then Y becomes a random variable and X no longer is one.

Example 3.3. Indicator function. Let $(\Omega, \mathscr{F})$ be a measurable space and $A \subset \Omega$. The function $\mathbf{1}_A : \Omega \to \mathbb{R}$ given by

$$\mathbf{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A, \end{cases}$$

is called the indicator function of A. One has

$$\mathbf{1}_A \text{ is a random variable } \iff A \in \mathscr{F}.$$

3.1 Law or distribution of a random vector

Definition 3.4. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}^n$ a random vector. The law, or distribution, of X is the probability measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ given by

$$\begin{aligned} \mathbb{P}_X : \ & \mathscr{B}(\mathbb{R}^n) \to \mathbb{R} \\ & A \mapsto \mathbb{P}(X^{-1}(A)). \end{aligned}$$

Exercise 3.5. Verify that $\mathbb{P}_X$ is a probability measure.

Example 3.6. Let $X \sim B(n, p)$. Then

$$\mathbb{P}_X(\{k\}) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n.$$

Formally,

$$\begin{aligned} \mathbb{P}_X : \ & \mathscr{B}(\mathbb{R}) \to \mathbb{R} \\ & A \mapsto \sum_{k \in A} \binom{n}{k} p^k (1-p)^{n-k}. \end{aligned}$$

Example 3.7. Let $X \sim \mathcal{N}(0,1)$. Then

$$\underbrace{\mathbb{P}(a < X \leq b)}_{=\mathbb{P}_X((a,b])} = \int_a^b \varphi(x)\,dx, \qquad \forall\ a < b,$$

where

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

To obtain the measure of every Borel set:

$$\begin{aligned} \mathbb{P}_X :\ & \mathscr{B}(\mathbb{R}) \to \mathbb{R} \\ & A \mapsto \underbrace{\int_A \varphi(x)\,dx}_{\text{Lebesgue integral}} . \end{aligned}$$

3.2 Conditional probability

Definition 3.8. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $B \in \mathscr{F}$ with $\mathbb{P}(B) > 0$. For each $A \in \mathscr{F}$ we define the conditional probability of A given the event B by

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Remark. The function

$$\begin{aligned} \mathbb{P}_B &: \mathscr{F} \to \mathbb{R} \\ & A \mapsto \mathbb{P}(A|B) \end{aligned}$$

is a probability measure on $(\Omega, \mathscr{F})$.

Suppose $A_1, A_2, \ldots, A_n \in \mathscr{F}$ and $\Omega = \biguplus_{k=1}^n A_k$. If $A \in \mathscr{F}$, then

$$\mathbb{P}(A) = \sum_{k=1}^n \mathbb{P}(A \cap A_k).$$

If, in addition, $\mathbb{P}(A_1), \ldots, \mathbb{P}(A_n) > 0$, then (law of total probability)

$$\mathbb{P}(A) = \sum_{k=1}^n \mathbb{P}(A|A_k)\mathbb{P}(A_k).$$

3.3 Discrete random vectors

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}^n$ a random vector. We say that X is a discrete random vector when there exists a countable set $A \subset \mathbb{R}^n$ such that

$$\mathbb{P}(X \in A) = 1.$$

Definition 3.9. Probability mass function. The probability mass function of a discrete random vector is defined by

$$\begin{aligned} \mathbb{P}_X : \;\; & A \to \mathbb{R} \\ & x \mapsto \mathbb{P}\{X = x\} \end{aligned}$$

or, equivalently,

$$\begin{aligned} \mathbb{P}_X : \;\; & \mathbb{R}^n \to \mathbb{R} \\ & x \mapsto \mathbb{P}\{X = x\}. \end{aligned}$$

3.4 Mean or expectation

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and X a discrete random variable, with $A \subset \mathbb{R}$ countable such that $\mathbb{P}(X \in A) = 1$:

1. if $A = \{x_1, \dots, x_n\}$, then $\mathbb{E}[X] = \sum_{k=1}^{n} x_k \mathbb{P}_X(x_k)$;
2. if $A = \{x_1, x_2, \dots\}$, then $\mathbb{E}[X] = \sum_{k=1}^{\infty} x_k \mathbb{P}_X(x_k)$.

If all the x_k are ≥ 0, or all are ≤ 0, the expectation is well defined (possibly with an infinite value). In general, set

$$\begin{aligned} I &= \{k \in \mathbb{Z}_+ : \; x_k \geq 0\} \\ J &= \{k \in \mathbb{Z}_+ : \; x_k \leq 0\} \\ S_I &= \sum_{k \in I} \underbrace{x_k \mathbb{P}_X(x_k)}_{\geq 0} \\ S_J &= \sum_{k \in J} \underbrace{x_k \mathbb{P}_X(x_k)}_{\leq 0}. \end{aligned}$$

If at least one of S_I, S_J is finite, we put $\mathbb{E}[X] = S_I + S_J$. If $S_I = +\infty$ and $S_J = -\infty$, the expectation $\mathbb{E}[X]$ is not defined. We say that X is integrable when both are finite, that is, when $\sum_k |x_k| \mathbb{P}_X(x_k) < \infty$.

Proposition 3.10. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, X, Y discrete random variables and $\alpha \in \mathbb{R}$. Then:

1. αX is a discrete r.v. and $\mathbb{E}[\alpha X] = \alpha \mathbb{E}[X]$.
2. $X + Y$ is a discrete r.v. and $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.
3. If $X \leq Y$, then $\mathbb{E}[X] \leq \mathbb{E}[Y]$, provided both expectations are defined.
4. If $g : \mathbb{R} \to \mathbb{R}$ is Borel measurable, then $g(X)$ is a discrete r.v. and $\mathbb{E}[g(X)] = \sum_k g(x_k) \mathbb{P}_X(x_k)$, the sum being finite or infinite as the case may be.

Proof. Item by item:

1. If $\alpha \neq 0$, then $\mathbb{E}[\alpha X] = \sum_{k=1}^{\infty} (\alpha x_k) p_k = \alpha \sum_{k=1}^{\infty} x_k p_k = \alpha \mathbb{E}[X]$. If $\alpha = 0$, then $\alpha X \equiv 0$ and $\mathbb{E}[0] = 0 = 0 \cdot \mathbb{E}[X]$.
2. Summing twice:

$$\begin{aligned}
\mathbb{E}[X + Y] &= \sum_x \sum_y (x + y) p(x, y) \\
&= \sum_x \sum_y \left[x p(x, y) + y p(x, y) \right] \\
&= \sum_x \sum_y x p(x, y) + \sum_x \sum_y y p(x, y) \\
&= \sum_x x \sum_y p(x, y) + \sum_y y \sum_x p(x, y) \\
&= \sum_x x p_X(x) + \sum_y y p_Y(y) \\
&= \mathbb{E}[X] + \mathbb{E}[Y].
\end{aligned}$$

Here,

$$p(x,y) = \mathbb{P}(X = x, Y = y)$$
$$p_X(x) = \mathbb{P}(X = x) = \sum_y p(x,y)$$
$$p_Y(y) = \mathbb{P}(Y = y) = \sum_x p(x,y).$$

3. If $X \leq Y$, then $Y - X \geq 0$ and hence $(y - x)p(x,y) \geq 0$ for every pair (x,y) with $p(x,y) > 0$. Therefore,

$$0 \leq \sum_x \sum_y (y - x)p(x,y) = \mathbb{E}[Y] - \mathbb{E}[X].$$

4. By definition, together with Proposition 4.2(4), proved in the next lecture.

□

4 Lecture 4

Lemma 4.1. Let X be a discrete r.v. with finite mean, written as

$$X = \sum_{i=1}^{\infty} c_i \mathbf{1}_{A_i}.$$

Then

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} c_i \mathbb{P}(A_i).$$

Proof. Here the A_i are pairwise disjoint with $\bigcup_i A_i = \Omega$, so that X takes values in the countable set $\{c_1, c_2, \dots\}$ and $A_i = X^{-1}(\{c_i\})$. Hence $p_i := \mathbb{P}_X(c_i) = \mathbb{P}(A_i)$ and

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} c_i p_i = \sum_{i=1}^{\infty} c_i \mathbb{P}(A_i).$$

This follows from

$$X^{-1}\{c_i\} = A_i \implies \mathbb{P} \circ X^{-1}\{c_i\} = P_X(c_i) = \mathbb{P}(A_i).$$

□

Proposition 4.2. (This is Proposition 3.10; we restate it because we now prove it again, this time starting from the representation of a discrete r.v. as a combination of indicator functions.) Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, X, Y discrete random variables and $\alpha \in \mathbb{R}$. Then:

1. αX is a discrete r.v. and $\mathbb{E}[\alpha X] = \alpha \mathbb{E}[X]$.
2. $X + Y$ is a discrete r.v. and $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.
3. If $X \leq Y$, then $\mathbb{E}[X] \leq \mathbb{E}[Y]$, provided both expectations are defined.
4. If $g : \mathbb{R} \to \mathbb{R}$ is Borel measurable, then $g(X)$ is a discrete r.v. and $\mathbb{E}[g(X)] = \sum_k g(x_k)\mathbb{P}_X(x_k)$, the sum being finite or infinite as the case may be.

Proof. Item by item:

1. Write $X = \sum_{k=1}^{\infty} x_k \mathbf{1}_{A_k}$ and $Y = \sum_{j=1}^{\infty} y_j \mathbf{1}_{B_j}$, where $A_k = X^{-1}(\{x_k\})$ and $B_j = Y^{-1}(\{y_j\})$. Then
$$\begin{aligned} \alpha X &= \sum_{k=1}^{\infty} (\alpha x_k)\mathbf{1}_{A_k}, \\ \mathbb{E}[\alpha X] &= \sum_{k=1}^{\infty} (\alpha x_k)\mathbb{P}(A_k) \\ &= \alpha \sum_{k=1}^{\infty} x_k \mathbb{P}(X^{-1}\{x_k\}) = \alpha \sum_{k=1}^{\infty} x_k \mathbb{P}_X(x_k) = \alpha \mathbb{E}[X]. \end{aligned}$$
2. We have $X + Y = \sum_{k=1}^{\infty} x_k \mathbf{1}_{A_k} + \sum_{j=1}^{\infty} y_j \mathbf{1}_{B_j}$, hence
$$\begin{aligned} \mathbb{E}[X+Y] &= \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} (x_k + y_j)\mathbb{P}(X = x_k, Y = y_j) \\ &= \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} x_k \mathbb{P}(X = x_k, Y = y_j) + \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} y_j \mathbb{P}(X = x_k, Y = y_j) \\ &= \sum_{k=1}^{\infty} x_k \mathbb{P}(A_k) + \sum_{j=1}^{\infty} y_j \mathbb{P}(B_j) \\ &= \mathbb{E}[X] + \mathbb{E}[Y]. \end{aligned}$$

3. Since $\sum_{j\geq 1} \mathbf{1}_{B_j} = 1$ and $\sum_{k\geq 1} \mathbf{1}_{A_k} = 1$, we may always write

$$X = \left(\sum_{k=1}^{\infty} x_k \mathbf{1}_{A_k}\right) \cdot \sum_{j=1}^{\infty} \mathbf{1}_{B_j} = \sum_{k,j\geq 1} x_k \mathbf{1}_{A_k \cap B_j}$$

and, in the same way,

$$Y = \sum_{k,j\geq 1} y_j \mathbf{1}_{A_k \cap B_j},$$

where in both sums it suffices to consider the pairs (k, j) with $A_k \cap B_j \neq \emptyset$. Picking $\omega \in A_k \cap B_j$ gives $x_k = X(\omega) \geq Y(\omega) = y_j$ (for the reverse inequality, swap the roles of X and Y), whence

$$\sum_{k,j\geq 1} y_j \mathbb{P}(A_k \cap B_j) \leq \sum_{k,j\geq 1} x_k \mathbb{P}(A_k \cap B_j).$$

4. From $X = \sum_{k=1}^{\infty} x_k \mathbf{1}_{A_k}$ we get $g(X) = \sum_{k=1}^{\infty} g(x_k)\mathbf{1}_{A_k}$, so that

$$\mathbb{E}[g(X)] = \sum_{k=1}^{\infty} g(x_k)\mathbb{P}(A_k).$$

$\square$

Definition 4.3. Let $X : \Omega \to \mathbb{R}^n$, $n \geq 1$, be a random vector. We define the moment generating function

$$M_X : \mathbb{R}^n \to \mathbb{R}, \qquad M_X(t) = \mathbb{E}[e^{\langle t,X\rangle}],$$

and the characteristic function

$$\Phi_X : \mathbb{R}^n \to \mathbb{C}, \qquad \Phi_X(t) = \mathbb{E}[e^{i\langle t,X\rangle}].$$

Theorem 4.4. Let $X : \Omega \to \mathbb{R}$ be a random variable. Then

1. $M_X^{(k)}(0) = \mathbb{E}[X^k]$ for $k = 1, 2, 3, \ldots$, provided M_X is finite on a neighbourhood of 0;

2. $\Phi_X^{(k)}(t) = \mathbb{E}[(iX)^k e^{itX}]$ for $k = 1, 2, \ldots$ and $t \in \mathbb{R}$, provided $\mathbb{E}[|X|^k] < \infty$.

Definition 4.5. Let $X : \Omega \to \mathbb{R}$ be a discrete r.v. with $\mathbb{E}[X^2] < \infty$. The variance of X is defined by

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2],$$

and the standard deviation of X by

$$\sigma_X = \sqrt{\mathrm{Var}(X)}.$$

Proposition 4.6. Let X be a discrete r.v. with $\mathbb{E}[X^2] < \infty$.

1. If $c \in \mathbb{R}$, then $\mathrm{Var}(cX) = c^2 \,\mathrm{Var}(X)$.
2. If $c \in \mathbb{R}$, then $\mathrm{Var}(X + c) = \mathrm{Var}(X)$.
3. $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Proof. Item by item.

1. Since $\mathbb{E}[cX] = c\mathbb{E}[X]$,
$$\mathbb{E}[(cX - \mathbb{E}[cX])^2] = \mathbb{E}[(c(X - \mathbb{E}[X]))^2] = c^2\,\mathbb{E}[(X - \mathbb{E}[X])^2].$$
2. Since $\mathbb{E}[X + c] = \mathbb{E}[X] + c$, we get
$$\mathbb{E}[(X + c - \mathbb{E}[X + c])^2] = \mathbb{E}[(X - \mathbb{E}[X])^2].$$
3. Finally,
$$\begin{aligned}\mathbb{E}[(X - \mathbb{E}[X])^2] &= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] \\ &= \mathbb{E}[X^2] - \mathbb{E}[2X\mathbb{E}[X]] + \mathbb{E}[\mathbb{E}[X]^2] \\ &= \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[X] + \mathbb{E}[X]^2 \\ &= \mathbb{E}[X^2] - \mathbb{E}[X]^2.\end{aligned}$$

□

4.1 Independence of discrete random vectors

Let $X_\lambda : \Omega \to \mathbb{R}^{n_\lambda}$, $\lambda \in \Lambda$, be discrete random vectors. We say that they are independent when

$$\mathbb{P}(X_{\lambda_1} = x_1, \ldots, X_{\lambda_k} = x_k) = \prod_{i=1}^{k} \mathbb{P}\{X_{\lambda_i} = x_i\},$$

for all $k \geq 2$, all $\lambda_1, \ldots, \lambda_k \in \Lambda$ pairwise distinct and all $x_j \in \mathbb{R}^{n_{\lambda_j}}$.

Definition 4.7. Bernoulli distribution. Let $p \in (0, 1)$. We write $X \sim Be(p)$ if $\mathbb{P}\{X = 1\} = p$ and $\mathbb{P}\{X = 0\} = 1 - p$; that is, $\mathbb{P}_X(1) = p$ and $\mathbb{P}_X(0) = 1 - p$. Then

$$\begin{aligned}
\mathbb{E}[X] &= p \\
\mathrm{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = p - p^2 = p(1-p) \\
M_X(s) &= \mathbb{E}[e^{sX}] = pe^s + (1-p).
\end{aligned}$$

Definition 4.8. Binomial distribution. Let $n \in \mathbb{N}$ and $p \in (0, 1)$. Let X be the number of successes in n independent Bernoulli trials, each with success probability p. We write $X \sim B(n, p)$, and

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad k = 0, 1, \ldots, n.$$

Proposition 4.9. One has

1. $\mathbb{E}[X] = np$;
2. $\mathrm{Var}(X) = np(1-p)$;
3. $M_X(s) = (pe^s + 1 - p)^n$.

Proof. First,

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} \frac{n!}{(n-k)!(k-1)!} p^k (1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^{j} (1-p)^{(n-1)-j} \\
&= np,
\end{aligned}$$

the last equality being the binomial theorem. Next,

$$\begin{aligned}
\mathbb{E}[X^2] &= \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} kn \binom{n-1}{k-1} p^k (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{k=1}^{n} k \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{j=0}^{n-1} (j+1) \binom{n-1}{j} p^{j} (1-p)^{(n-1)-j} \\
&= np \left(\sum_{j=0}^{n-1} j \binom{n-1}{j} p^{j} (1-p)^{(n-1)-j} + \sum_{j=0}^{n-1} \binom{n-1}{j} p^{j} (1-p)^{(n-1)-j} \right) \\
&= np \left((n-1)p + 1 \right) \\
&= n^2 p^2 + np(1-p),
\end{aligned}$$

where in the penultimate step we used that the first sum is the mean of a $B(n-1, p)$ variable, namely $(n-1)p$, and that the second sum equals 1. Hence

$$\mathrm{Var}(X) = np(1-p) + n^2p^2 - (np)^2 = np(1-p).$$

Finally,

$$\begin{aligned}
M_X(s) &= \mathbb{E}[e^{sX}] \\
&= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} e^{sk} \\
&= \sum_{k=0}^{n} \binom{n}{k} (e^s p)^k (1-p)^{n-k} \\
&= (e^s p + (1-p))^n.
\end{aligned}$$

□

Definition 4.10. Geometric distribution. If the discrete random variable X models the total number of trials up to and including the first success in a sequence of independent Bernoulli trials, each with success probability p, then the probability mass function of $X \sim Geom(p)$ is

$$\mathbb{P}\{X = x\} = p(1-p)^{x-1}, \qquad x = 1, 2, 3, \ldots$$

The variable $Y = X - 1$, counting the number of *failures before* the first success, satisfies $\mathbb{P}\{Y = k\} = p(1-p)^k$ for $k = 0, 1, 2, \ldots$; we write $Y \sim G(p)$. Both conventions are in common use and it pays to keep them apart: in what follows, $Geom(p)$ counts trials and $G(p)$ counts failures.

Proposition 4.11. If $X \sim Geom(p)$, then

1. $\mathbb{E}[X] = 1/p$;
2. $\mathrm{Var}(X) = \frac{1-p}{p^2}$;
3. $M_X(t) = \dfrac{pe^t}{1-(1-p)e^t}$, for $t < -\ln(1-p)$.

Proof. Write $q = 1 - p \in (0, 1)$. Then

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=1}^{\infty} xpq^{x-1} \\
&= p \sum_{x=1}^{\infty} \frac{\mathrm{d}}{\mathrm{d}q}(q^x) \\
&= p \frac{\mathrm{d}}{\mathrm{d}q}\left(\sum_{x=1}^{\infty} q^x\right) \\
&= p \frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{q}{1-q}\right) \\
&= \frac{p}{(1-q)^2} = \frac{1}{p},
\end{aligned}$$

where interchanging the derivative and the series is legitimate because the power series converges uniformly on compact subsets of $(-1, 1)$.

For the variance we use the factorial moment. First,

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - \mathbb{E}[X]^2,$$

by linearity of the expectation. To compute $\mathbb{E}[X(X-1)]$, with $q = 1 - p$,

$$\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{x \geq 1} x(x-1)pq^{x-1} = pq \sum_{x \geq 2} x(x-1)q^{x-2} \\
&= pq \sum_{x \geq 2} \frac{\mathrm{d}^2}{\mathrm{d}q^2}(q^x) = pq \frac{\mathrm{d}^2}{\mathrm{d}q^2}\left(\sum_{x \geq 0} q^x\right) \\
&= pq \frac{\mathrm{d}^2}{\mathrm{d}q^2}\left(\frac{1}{1-q}\right) = pq \frac{\mathrm{d}}{\mathrm{d}q}\left(\frac{1}{(1-q)^2}\right) \\
&= pq \frac{2}{(1-q)^3} = \frac{2(1-p)}{p^2}.
\end{aligned}$$

Therefore,

$$\mathrm{Var}(X) = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2(1-p) + p - 1}{p^2} = \frac{1-p}{p^2}.$$

Lastly, for $t < -\ln(1-p)$ we have $(1-p)e^t < 1$ and

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{tX}] &= \sum_{x=1}^{\infty} p(1-p)^{x-1}e^{tx} \\ &= \frac{p}{1-p}\sum_{x=1}^{\infty}\big((1-p)e^t\big)^x \\ &= \frac{p}{1-p}\cdot\frac{(1-p)e^t}{1-(1-p)e^t} \\ &= \frac{pe^t}{1-(1-p)e^t}. \end{aligned}$$

$\square$

Definition 4.12. Negative binomial distribution. Let $r \in \mathbb{N}$ and $p \in (0,1)$. Let X be the number of *failures* accumulated before the r-th success, in a sequence of independent Bernoulli trials with parameter p. We write $X \sim NB(r,p)$, and

$$\mathbb{P}(X=k) = \binom{r+k-1}{k} p^r(1-p)^k, \qquad k = 0,1,2,\ldots$$

With this convention the total number of trials performed is $X+r$, and the case $r=1$ recovers the failure-counting geometric distribution: $NB(1,p) = G(p)$.

Remark. The negative binomial distribution can be seen as a generalization of the geometric distribution. Consider a random experiment with two possible outcomes: success, occurring with probability p, and failure, with probability $1-p$. For instance, rolling a die repeatedly, obtaining a 6 counts as a success. If we count the failures before the first success, the random variable follows a geometric distribution with parameter p, denoted $G(p)$, whose mass function is

$$\mathbb{P}(X=k) = p(1-p)^k$$

for $k \in \{0,1,2,\ldots\}$. If we aim for n successes and count the failures before the n-th success, the random variable follows a negative binomial distribution with parameters n and p, denoted $NB(n,p)$. The geometric distribution is the particular case $n = 1$, that is, $NB(1,p) = G(p)$. The mass function of the negative binomial, for $k \in \{0, 1, 2, \ldots\}$, is

$$\mathbb{P}(X = k) = \binom{k+n-1}{k} p^n (1-p)^k.$$

This distribution can be generalized to non-integer values of n by replacing the factorials with the Gamma function.

Proposition 4.13. If $X \sim NB(r,p)$, then

1. $\mathbb{E}[X] = \frac{r(1-p)}{p}$;
2. $\mathrm{Var}(X) = \frac{r(1-p)}{p^2}$;
3. $M_X(s) = p^r \left(1 - (1-p)e^s\right)^{-r}$, for $s < -\ln(1-p)$.

Proof. For (1) and (2) it suffices to note that if $X \sim NB(r,p)$, then $X = \sum_{i=1}^r Y_i$ with $Y_1, \ldots, Y_r$ independent and $Y_i \sim G(p)$; since $\mathbb{E}[Y_i] = (1-p)/p$ and $\mathrm{Var}(Y_i) = (1-p)/p^2$, the claim follows from linearity of the expectation and additivity of the variance over independent summands.

As for the moment generating function, we use the identity

$$\binom{-r}{y} = (-1)^y \binom{r+y-1}{y}, \qquad y = 0, 1, 2, \ldots$$

Then

$$\begin{aligned} M_X(t) &= \sum_{y=0}^{\infty} e^{ty} \binom{r+y-1}{y} (1-p)^y p^r \\ &= p^r \sum_{y=0}^{\infty} \binom{r+y-1}{y} \left(e^t(1-p)\right)^y \\ &= p^r \sum_{y=0}^{\infty} \binom{-r}{y} \left(-e^t(1-p)\right)^y. \end{aligned}$$

If $t < -\ln(1-p)$, then $|e^t(1-p)| < 1$, and the binomial series $(1+x)^{-r} = \sum_{y=0}^{\infty} \binom{-r}{y} x^y$, valid for $|x| < 1$, gives

$$M_X(t) = \frac{p^r}{\left(1 - (1-p)e^t\right)^r}.$$

□

5 Lecture 5

Suppose we have N numbered balls: $1, 2, \ldots, N$. We are going to draw n of them sequentially. Then

$$\Omega = \{(a_1, a_2, \ldots, a_n) : a_k = 1, \ldots, N, \text{ for } k = 1, \ldots, n\}.$$

	Ordered	Unordered
With replacement	$\lvert\Omega\rvert = N^n$	$\lvert\Omega\rvert = \binom{N+n-1}{N-1} = \binom{N+n-1}{n}$
Without replacement	$\lvert\Omega\rvert = \frac{N!}{(N-n)!}$	$\lvert\Omega\rvert = \frac{N!}{(N-n)!n!} = \binom{N}{n}$

Table 1: Cardinality of Ω.

Remark. The only formula that is not immediate is the one for unordered selections with replacement. To understand it one uses the "stars and bars" argument: a selection is determined by the multiplicities $x_1, \ldots, x_N \geq 0$ with $\sum_i x_i = n$, and each multiplicity vector is encoded as a row of n boxes ($\square$) and $N-1$ separators ($\bullet$). There are $\binom{N+n-1}{n}$ ways to choose the positions of the boxes among the $N+n-1$ available slots.

$$\begin{gathered} 1 \to x_1 \\ \vdots \\ N \to x_N \end{gathered}$$

with $x_1, \ldots, x_N \geq 0$ such that $\sum_i x_i = n$:

$$\underbrace{\underbrace{\square \cdots \square}_{x_1} \bullet \underbrace{\square \cdots \square}_{x_2} \bullet \cdots \bullet \underbrace{\square \cdots \square}_{x_N}}_{N+n-1}$$

Along a sequence (a tree),

$$\mathbb{P}(A_1 \cap B_2 \cap C_2 \cap D_2) = \mathbb{P}(A_1)\mathbb{P}(B_2|A_1)\mathbb{P}(C_2|A_1 \cap B_2)\mathbb{P}(D_2|A_1 \cap B_2 \cap C_2).$$

Example 5.1. Birthday problem. In a classroom with n people, what is the probability that two of them share a birthday? Assuming 365 equally likely days and independence across people,

$$p_n = \frac{365^n - 365 \cdot 364 \cdots (366-n)}{365^n} = 1 - \prod_{j=1}^{n-1} \left(1 - \frac{j}{365}\right).$$

For instance,

n	22	23	40	64
p_n	0.476	0.507	0.891	0.997

5.1 Multinomial distribution

Consider r possible types of outcome in an experiment. We perform the experiment n consecutive times, independently. Let

$$X = (X_1, \dots, X_r),$$

where X_k is the number of outcomes of type k obtained in the n experiments. Then

$$\mathbb{P}\{X = (n_1, \dots, n_r)\} = \frac{n!}{n_1! \cdots n_r!} p_1^{n_1} \cdots p_r^{n_r},$$

where $n_1, \dots, n_r \geq 0$ and $\sum_i n_i = n$. Here p_k is the probability of obtaining an outcome of type k in a single trial, and $\frac{n!}{n_1! \cdots n_r!}$ is the multinomial coefficient, which counts the number of ways in which n trials can be distributed among r outcome categories, correcting for indistinguishable permutations within each category.

The factorial $n!$ represents all the ways in which n events could be ordered if they were distinct. However, since we are interested in outcomes with n_k indistinguishable results of type k, we must correct this count by dividing by $n_k!$ for each k, the number of ways of ordering n_k identical events. This prevents overcounting sequences that are indistinguishable from one another.

The product $p_1^{n_1} \cdots p_r^{n_r}$ is the joint probability of obtaining one specific sequence of outcomes, under the assumption that the trials are independent. Since the multinomial distribution generalizes the binomial distribution to more than two categories, this product is the natural extension of the probability of a sequence of Bernoulli trials.

Indeed, the multinomial coefficient admits the telescoping factorization

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-\cdots-n_{r-1}}{n_r}=\frac{n!}{n_1!\cdots n_r!},$$

which corresponds to first choosing the n_1 trials of type 1, then the n_2 of type 2 among the remaining ones, and so on. Each specific sequence has probability $p_1^{n_1}\cdots p_r^{n_r}$.

Remark. One has

$$\sum_{k=0}^{n}\binom{n}{k}=2^n=(1+1)^n$$

and

$$(a+b)^n=\sum_{k=0}^{n}\underbrace{\binom{n}{k}}_{\text{ways to place } k \text{ letters } b \text{ and } n-k \text{ letters } a \text{ in a row of length } n}a^{n-k}b^k.$$

Exercise 5.2. Prove that

$$\sum_{k=0}^{n}\binom{n}{k}^2=\binom{2n}{n},$$

$$\sum_{k=0}^{n}k(k-1)\binom{n}{k}=n(n-1)2^{n-2}.$$

Proof. For the first identity, note that

$$\begin{aligned}(1+x)^{2n}&=\sum_{k=0}^{2n}\binom{2n}{k}x^k\\&=\binom{2n}{0}+\cdots+\binom{2n}{n}x^n+\cdots+\binom{2n}{2n}x^{2n}.\end{aligned}$$

On the other hand,

$$(1+x)^n(1+x)^n=\left[\sum_{k=0}^{n}\binom{n}{k}x^k\right]^2.$$

Expanding the product, the coefficient of x^n is

$$\sum_{k=0}^{n} \binom{n}{k} \binom{n}{n-k} = \sum_{k=0}^{n} \binom{n}{k}^2.$$

This can also be seen combinatorially: split a set of $2n$ elements into two blocks of n elements and choose n elements in total. Taking k from the first block and $n-k$ from the second, for $k = 0, \ldots, n$, yields $\sum_{k=0}^{n} \binom{n}{k}\binom{n}{n-k}$ selections.

For the second identity,

$$\begin{aligned}
(1+x)^n &= \sum_{k=0}^{n} \binom{n}{k} x^k \\
n(1+x)^{n-1} &= \sum_{k=1}^{n} k \binom{n}{k} x^{k-1} \\
(n-1)n(1+x)^{n-2} &= \sum_{k=2}^{n} k(k-1) \binom{n}{k} x^{k-2} \\
(n-1)n2^{n-2} &= \sum_{k=2}^{n} k(k-1) \binom{n}{k},
\end{aligned}$$

where the last step is obtained by evaluating at $x = 1$. $\square$

6 Lecture 6

Vandermonde's identity. For $m, n, t \in \mathbb{N}$,

$$\sum_{k=\max\{0,t-n\}}^{\min\{t,m\}} \binom{m}{k}\binom{n}{t-k} = \binom{m+n}{t}.$$

Interpretation: to choose t elements from a set of $m+n$ split into two blocks of sizes m and n, choose k from the first and $t-k$ from the second. With the convention $\binom{\alpha}{k} = \frac{\alpha(\alpha-1)\cdots(\alpha-k+1)}{k!}$, the identity remains valid for $m, n \in \mathbb{R}$.

Proposition 6.1. Markov's inequality. Let X be a nonnegative r.v. and $\varepsilon > 0$. Then

$$\mathbb{P}\{X \geq \varepsilon\} \leq \frac{\mathbb{E}[X]}{\varepsilon}.$$

Proof. One has

$$\int_\Omega X \, d\mathbb{P} = \int_{\{X \geq \varepsilon\}} X \, d\mathbb{P} + \int_{\{X < \varepsilon\}} X \, d\mathbb{P} \geq \varepsilon \underbrace{\int_{\{X \geq \varepsilon\}} d\mathbb{P}}_{=\mathbb{P}\{X \geq \varepsilon\}}.$$

$\square$

Corollary 6.2. Chebyshev's inequality. One has

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}.$$

It suffices to apply Markov's inequality to the nonnegative r.v. $(X - \mathbb{E}[X])^2$ at level ε^2, noting that $\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathrm{Var}(X)$ and that $\{|X - \mathbb{E}[X]| \geq \varepsilon\} = \{(X - \mathbb{E}[X])^2 \geq \varepsilon^2\}$.

Example 6.3. Let $X_1, \ldots, X_n$ be independent r.v.'s with distribution $Be(p)$. Then $S_n = \sum_{i=1}^n X_i \sim B(n,p)$ and

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \varepsilon\right) = \mathbb{P}\left(|S_n - \underbrace{np}_{=\mathbb{E}[S_n]}| \geq n\varepsilon\right) \leq \frac{\mathrm{Var}(S_n)}{n^2\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2}.$$

This is the weak law of large numbers for the Bernoulli scheme.

Weierstrass approximation theorem (via Bernstein polynomials). Let $f : [0,1] \to \mathbb{R}$ be continuous. For each $n \in \mathbb{Z}_+$ set

$$B_n : [0,1] \to \mathbb{R}, \qquad B_n(p) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) \binom{n}{k} p^k (1-p)^{n-k}.$$

We claim that $B_n \to f$ uniformly on $[0,1]$. Since f is continuous on the compact set $[0,1]$, there exists $M > 0$ with $|f(x)| \le M$ for all $x \in [0,1]$, and f is uniformly continuous. Since $\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1$,

$$|f(p) - B_n(p)| = \left| \sum_{k=0}^{n} \left[f(p) - f\left(\frac{k}{n}\right) \right] \binom{n}{k} p^k (1-p)^{n-k} \right|.$$

Fix $\varepsilon > 0$. By uniform continuity there is $\delta > 0$ such that, for all $x, y \in [0,1]$,

$$|x - y| < \delta \implies |f(x) - f(y)| < \varepsilon.$$

Write $A = \{k \in \{0, \dots, n\} : |p - k/n| < \delta\}$. Then

$$\begin{aligned}
|f(p) - B_n(p)| &\le \sum_{k \in A} \left| f(p) - f\left(\frac{k}{n}\right) \right| \binom{n}{k} p^k (1-p)^{n-k} \\
&\quad + \sum_{k \in A^c} \left| f(p) - f\left(\frac{k}{n}\right) \right| \binom{n}{k} p^k (1-p)^{n-k} \\
&\le \varepsilon \sum_{k \in A} \binom{n}{k} p^k (1-p)^{n-k} + 2M \sum_{k \in A^c} \binom{n}{k} p^k (1-p)^{n-k} \\
&\le \varepsilon + 2M\, \mathbb{P}\left(\left| \frac{S_n}{n} - p \right| \ge \delta \right) \\
&\le \varepsilon + \frac{2Mp(1-p)}{n\delta^2} \\
&\le \varepsilon + \frac{M}{2n\delta^2},
\end{aligned}$$

where $S_n \sim B(n,p)$, the penultimate step uses Chebyshev's inequality (previous example) and the last one uses $p(1-p) \le 1/4$. Since the bound does not depend on p and $\varepsilon > 0$ was arbitrary, $B_n \to f$ uniformly on $[0,1]$.

6.1 (Absolutely) continuous random vectors

Let $X : \Omega \to \mathbb{R}^n$ be a random vector. We say that X is an absolutely continuous random vector when there exists a function $f : \mathbb{R}^n \to [0, \infty)$ such that

$$\mathbb{P}\{X \leq a\} = \int_{-\infty}^{a_1} dx_1 \int_{-\infty}^{a_2} dx_2 \cdots \int_{-\infty}^{a_n} dx_n \, f(x_1, \ldots, x_n), \qquad \forall\, a \in \mathbb{R}^n.$$

Remark. The function f is called the density function of the random vector X.

$$\mathbb{P}_X(A) = \mathbb{P}\{X \in A\} = \int_A f(x)\, dx, \qquad \forall\, A \in \mathscr{B}(\mathbb{R}^n).$$

Notation. We write

$$d\mathbb{P}_X = f(x)\, dx, \qquad f(x) = \frac{d\mathbb{P}_X}{dx},$$

and say that f is the Radon–Nikodym derivative of $\mathbb{P}_X$ with respect to Lebesgue measure dx.

Remark. One has

$$\int_{\mathbb{R}^n} f(x)\, dx = 1.$$

Let $X : \Omega \to \mathbb{R}$ be an absolutely continuous r.v. with density f. When the integral $\int_{\mathbb{R}} |x| f(x)\, dx$ is finite, we say that X is integrable and define

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f(x)\, dx.$$

Proposition 6.4. If $X : \Omega \to \mathbb{R}$ is a continuous r.v. with density f and $g : \mathbb{R} \to \mathbb{R}$ is Borel measurable with $\int_{\mathbb{R}} |g(x)| f(x)\, dx < \infty$, then

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f(x)\, dx.$$

6.2 Classical (absolutely) continuous distributions

1. Normal distribution: $X \sim \mathcal{N}(\mu, \sigma^2)$, $f(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$, $\mathbb{E}[X] = \mu$, $\mathrm{Var}(X) = \sigma^2$ and $M(s) = e^{\mu s + \frac{\sigma^2 s^2}{2}}$.

2. Uniform distribution: $X \sim U([a,b])$,

$$f(t) = \begin{cases} \frac{1}{b-a}, & \text{if } t \in [a,b], \\ 0, & \text{otherwise,} \end{cases}$$

$\mathbb{E}[X] = \frac{a+b}{2}$, $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$, $M(s) = \frac{e^{sb}-e^{sa}}{s(b-a)}$.

3. Exponential distribution: $X \sim \exp(\lambda)$,

$$f(t) = \begin{cases} \lambda e^{-\lambda t}, & \text{if } t \geq 0, \\ 0, & \text{if } t < 0, \end{cases}$$

$\mathbb{E}[X] = \frac{1}{\lambda}$, $\mathrm{Var}(X) = \frac{1}{\lambda^2}$ and $M(s) = \frac{\lambda}{\lambda - s}$ for $s < \lambda$.

6.3 Almost sure occurrence

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and consider a property $\mathscr{P}(\omega)$ that each $\omega \in \Omega$ may or may not satisfy. We say that $\mathscr{P}$ holds almost surely (a.s.) when there exists $A \in \mathscr{F}$ such that $\mathscr{P}(\omega)$ holds for every $\omega \in A$ and $\mathbb{P}(A^c) = 0$ (equivalently, $\mathbb{P}(A) = 1$).

Example 6.5. Let $X, Y : \Omega \to \mathbb{R}$ be random variables. We say that $X \geq Y$ a.s. if there exists $A \in \mathscr{F}$ with $\mathbb{P}(A) = 1$ such that $X(\omega) \geq Y(\omega)$ for all $\omega \in A$. Note that

$$\{\omega \in \Omega : X(\omega) \geq Y(\omega)\} = (X - Y)^{-1}([0, \infty)) \in \mathscr{F}.$$

7 Lecture 7

7.1 Distribution function

Let $X : \Omega \to \mathbb{R}^n$ be a random vector. The distribution function of X is the function $F : \mathbb{R}^n \to \mathbb{R}$ given by

$$F_X(a) = \mathbb{P}\{X \leq a\} = P_X((-\infty, a]),$$

where $\{X \leq a\} = \{\omega \in \Omega : X(\omega) \leq a\}$.

Proposition 7.1. Let X be a random variable. Then

1. $0 \leq F_X(t) \leq 1$ for all $t \in \mathbb{R}$;

2. F_X is non-decreasing; in particular, left-hand limits of F_X exist;

3. F_X is right-continuous;

4. $\lim_{t\to-\infty} F_X(t) = 0$ and $\lim_{t\to\infty} F_X(t) = 1$.

Proof. Item by item:

1. By definition, since $\mathbb{P}(A) \in [0, 1]$ for every A.

2. If $s < t$, then

$$F_X(s) = P_X((-\infty, s]) \leq P_X((-\infty, t]) = F_X(t).$$

3. Fix $a \in \mathbb{R}$. We want to show that $\lim_{t\to a^+} F_X(t) = F_X(a)$. It suffices to show that if $t_n \downarrow a$ (t_n decreasing with $t_n > a$), then $\lim_{n\to\infty} F_X(t_n) = F_X(a)$. We have

$$F_X(t_n) = P_X((-\infty, t_n])$$
$$F_X(a) = P_X((-\infty, a]).$$

Put $A_n = (-\infty, t_n]$. Then $A = (-\infty, a] = \bigcap_{n\in\mathbb{N}} A_n$ and, since $A_{n+1} \subset A_n$, we have $A_n \downarrow A$. As $\lim_n P_X(A_n) = P_X(A)$, the claim follows. It is worth asking why the "left-hand" argument fails: if we take $t_n \uparrow a$ and $A_n = (-\infty, t_n]$, then $\bigcup_{n\in\mathbb{N}} A_n = (-\infty, a) \neq A$.

4. We have $\lim_{t\to-\infty} F_X(t) = 0$ if and only if $\lim_n F_X(t_n) = 0$ for every sequence $t_n \to -\infty$. We may assume t_n decreasing; with $B_n = (-\infty, t_n]$ we get $B_n \downarrow \bigcap_{n\in\mathbb{N}} B_n = \emptyset$, so $\lim_n P_X(B_n) = P_X(\emptyset) = 0$. The argument for $\lim_{t\to\infty} F_X(t) = 1$ is analogous, using increasing sets whose union is $\mathbb{R}$.

$\square$

Remark. The law of the random vector X determines F_X. Moreover, F_X characterizes the law P_X of X.

Remark. If the random vectors $X, Y : \Omega \to \mathbb{R}^n$ have the same distribution function, then they have the same distribution (law).

$$\underbrace{\{\text{Distributions on } (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))\}}_{\text{probability measures}} \longleftrightarrow \underbrace{\{\text{Distribution functions on } \mathbb{R}^n\}}_{\text{with the properties above}}.$$

7.2 Independence of events

Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and let A_λ, $\lambda \in \Lambda$, be events. We say that the events are independent when

$$\mathbb{P}\left(\bigcap_{k=1}^{n} A_{\lambda_k}\right) = \prod_{k=1}^{n} \mathbb{P}(A_{\lambda_k}),$$

for every $n \in \mathbb{Z}_+$ and all pairwise distinct $\lambda_1, \ldots, \lambda_n \in \Lambda$.

Example 7.2. Let $A, B \in \mathscr{F}$. Then A and B are independent if and only if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Example 7.3. Let $A, B, C \in \mathscr{F}$. Then A, B and C are independent if and only if

1. $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$;
2. $\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C)$;
3. $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$;

4. $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$.

Note that the first three conditions alone (pairwise independence) do not imply the fourth.

7.3 Limit superior and limit inferior of a sequence of sets

Let $A_1, A_2, A_3, \ldots$ be a sequence of sets. We define

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \quad \text{and} \quad \liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k.$$

1. If $\omega \in \limsup A_n$, then $\omega \in \bigcup_{k \geq n} A_k$ for every $n \in \mathbb{N}$. Hence $\omega \in A_k$ for infinitely many k.

2. If $\omega \in \liminf A_n$, then $\omega \in \bigcap_{k \geq m} A_k$ for some $m \in \mathbb{N}$; that is, $\omega \in A_j$ for every $j \geq m$. Hence $\omega \notin A_k$ for at most finitely many indices k.

Proposition 7.4. One has

1. $\liminf A_n \subset \limsup A_n$;

2. if $B_n = \bigcup_{k \geq n} A_k$, then $B_n \downarrow \limsup A_n$;

3. if $C_n = \bigcap_{k \geq n} A_k$, then $C_n \uparrow \liminf A_n$;

4. $\limsup A_n = \{\omega \in \Omega : \ \omega \text{ belongs to infinitely many } A_k\}$;

5. $\liminf A_n = \{\omega \in \Omega : \ \omega \notin A_k \text{ for at most finitely many } k\}$.

Definition 7.5. We say that the sequence A_n is convergent if $\liminf A_n = \limsup A_n$.

Example 7.6. Let $\Omega = \mathbb{R}$ and $A_n = \left(-\frac{1}{n}, n\right]$. Then

$$\limsup A_n = [0, \infty).$$

Indeed, $\bigcup_{k \geq n} A_k = (-1/n, \infty)$ and the intersection over n is $[0, \infty)$. Moreover, $\bigcap_{k \geq n} A_k = [0, n]$, so that

$$\liminf A_n = [0, \infty).$$

That is,

$$\liminf A_n = \limsup A_n.$$

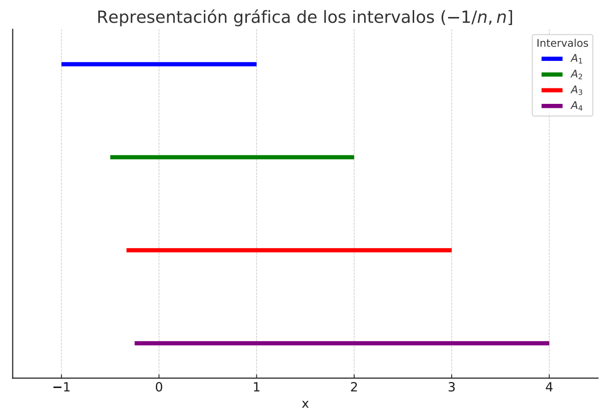

Figure 1: The intervals A_n.

Example 7.7. Let $\Omega = \mathbb{R}$ and

$$A_n = \begin{cases} [0, n), & \text{if } n \text{ is odd,} \\ (-n, 0), & \text{if } n \text{ is even.} \end{cases}$$

Then $\limsup A_n = \mathbb{R}$ and $\liminf A_n = \emptyset$.

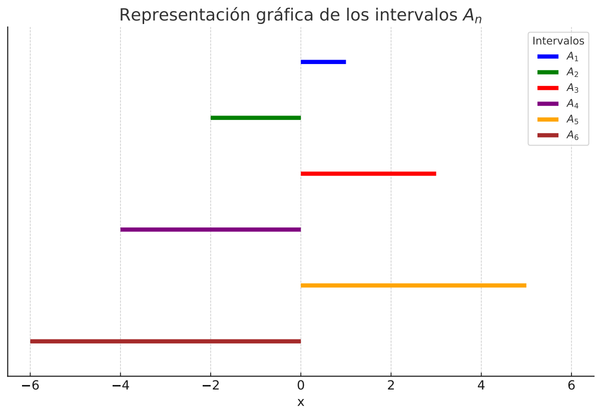

Figure 2: The intervals A_n of the second example.

8 Lecture 8

Lemma 8.1. First Borel–Cantelli lemma. Let $A_1, A_2, \ldots$ be a sequence of events with $\sum_n \mathbb{P}(A_n) < \infty$. Then

$$\mathbb{P}\left(\limsup A_n\right) = 0.$$

Proof. By definition,

$$\mathbb{P}\left(\limsup A_n\right) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} \underbrace{\bigcup_{k=n}^{\infty} A_k}_{=B_n}\right) = \lim_n \mathbb{P}(B_n),$$

since $B_n \downarrow \limsup A_n$. Now

$$0 \leq \mathbb{P}(B_n) = \mathbb{P}\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \sum_{k=n}^{\infty} \mathbb{P}(A_k) \xrightarrow[n\to\infty]{} 0,$$

as this is the tail of a convergent series. Hence

$$\mathbb{P}\left(\limsup A_n\right) = 0.$$

$\square$

Lemma 8.2. Second Borel–Cantelli lemma. Let $A_1, A_2, \ldots$ be a sequence of independent events with $\sum_n \mathbb{P}(A_n) = \infty$. Then

$$\mathbb{P}\left(\limsup A_n\right) = 1.$$

Proof. Let $B_n = \bigcup_{k \geq n} A_k$, so that $B_n \downarrow \limsup A_n$ and $\mathbb{P}(\limsup A_n) = \lim_n \mathbb{P}(B_n)$. We shall prove that $\mathbb{P}(B_n) = 1$ for every $n \geq 1$, which amounts to proving that $\mathbb{P}(B_n^c) = 0$ for every $n \geq 1$. By definition,

$$\mathbb{P}(B_n^c) = \mathbb{P}\left(\bigcap_{k=n}^{\infty} A_k^c\right).$$

Set

$$D_m = \bigcap_{k=n}^{m} A_k^c \implies D_m \downarrow \bigcap_{k=n}^{\infty} A_k^c.$$

Then, by continuity of the measure and by Exercise 8.3,

$$\begin{aligned}\mathbb{P}(B_n^c) &= \lim_{m\to\infty} \mathbb{P}(D_m) = \lim_{m\to\infty} \prod_{k=n}^{m} \mathbb{P}(A_k^c) \\ &= \lim_{m\to\infty} \prod_{k=n}^{m} [1-\mathbb{P}(A_k)] \\ &= \prod_{k=n}^{\infty} [1-\mathbb{P}(A_k)].\end{aligned}$$

Since $\sum_{k\geq n} \mathbb{P}(A_k) = \infty$ for every n, Lemma 8.4 gives $\prod_{k=n}^{\infty}[1-\mathbb{P}(A_k)] = 0$. Hence $\mathbb{P}(B_n) = 1$ for every n, and

$$\mathbb{P}\left(\limsup A_n\right) = \lim_{n} \mathbb{P}(B_n) = 1.$$

□

Exercise 8.3. Prove that if the events A_λ, $\lambda \in \Lambda$, are independent, then so are the events A_λ^c, $\lambda \in \Lambda$.

Lemma 8.4. Let $x_k \in [0,1]$, $k = 1, 2, \dots$. If $\sum_{k=1}^{\infty} x_k = \infty$, then $\prod_{k=1}^{\infty}(1 - x_k) = 0$.

Proof. On the one hand, $\sum_{k=1}^{\infty} x_k$ is well defined because $x_k \geq 0$. On the other hand, the sequence

$$\pi_m = \prod_{k=1}^{m} (1 - x_k)$$

is decreasing and bounded below by 0, hence it has a limit. Now, for every $t \in \mathbb{R}$,

$$e^t \geq t + 1,$$

so that

$$e^{-x_k} \geq 1 - x_k \geq 0$$

and

$$\underbrace{\prod_{k=1}^{m} e^{-x_k}}_{=e^{-\sum_{k=1}^{m} x_k}} \geq \prod_{k=1}^{m} (1 - x_k) \geq 0.$$

Since

$$\lim_{m\to\infty} e^{-\sum_{k=1}^{m} x_k} = 0,$$

the conclusion follows. $\square$

Lemma 8.5. Let $x_k \in [0,1)$, $k = 1, 2, \ldots$. If $\prod_{k=1}^{\infty}(1 - x_k) = 0$, then

$$\sum_{k=1}^{\infty} x_k = \infty.$$

Combining with the previous lemma: for $x_k \in [0,1)$ one has $\prod_{k\geq 1}(1-x_k) = 0$ if and only if $\sum_{k\geq 1} x_k = \infty$.

8.1 Independence of σ-algebras

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $\mathscr{F}_\lambda$, $\lambda \in \Lambda$, a family of σ-algebras on Ω contained in $\mathscr{F}$.

Definition 8.6. We say that the σ-algebras $\mathscr{F}_\lambda$, $\lambda \in \Lambda$, are independent when

$$\mathbb{P}\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} \mathbb{P}(A_k),$$

for

- every $n \in \mathbb{Z}_+$ and all pairwise distinct $\lambda_1, \ldots, \lambda_n \in \Lambda$,
- all $A_1 \in \mathscr{F}_{\lambda_1}, \ldots, A_n \in \mathscr{F}_{\lambda_n}$.

Remark. Note that:

1. If the σ-algebras $\mathscr{F}_\lambda$, $\lambda \in \Lambda$, are independent and $A_\lambda \in \mathscr{F}_\lambda$ for every $\lambda \in \Lambda$, then the events A_λ, $\lambda \in \Lambda$, are independent.
2. If the σ-algebras $\mathscr{F}_\lambda$, $\lambda \in \Lambda$, are independent and $J \subset \Lambda$, then the σ-algebras $\mathscr{F}_\lambda$, $\lambda \in J$, are independent.
3. If the σ-algebras $\mathscr{F}_\lambda$, $\lambda \in \Lambda$, are independent and, for each $\lambda \in \Lambda$, $\mathscr{G}_\lambda \subset \mathscr{F}_\lambda$ is a σ-algebra on Ω, then the σ-algebras $\mathscr{G}_\lambda$ are independent.

Definition 8.7. Let Ω be a nonempty set and $\mathscr{C}$ a nonempty family of subsets of Ω. We say that $\mathscr{C}$ is a π-system on Ω when

$$A \cap B \in \mathscr{C} \quad \text{for all } A, B \in \mathscr{C}.$$

Remark. If $\mathscr{C}$ is a π-system on Ω, then

$$\bigcap_{k=1}^{n} A_k \in \mathscr{C} \quad \text{for all } n \in \mathbb{Z}_+ \text{ and all } A_1, \ldots, A_n \in \mathscr{C}.$$

Theorem 8.8. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let $\mathscr{C}_\lambda$, $\lambda \in \Lambda$, be π-systems on Ω such that

$$\mathbb{P}\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} \mathbb{P}(A_k),$$

for every $n \in \mathbb{Z}_+$, all pairwise distinct $\lambda_1, \ldots, \lambda_n \in \Lambda$ and all $A_i \in \mathscr{C}_{\lambda_i}$. Then the σ-algebras $\sigma(\mathscr{C}_\lambda)$, $\lambda \in \Lambda$, are independent.

Corollary 8.9. Let $\mathscr{F}_\lambda$, $\lambda \in \Lambda$, be independent σ-algebras and let I_j, $j \in J$, be pairwise disjoint subsets of Λ. Then the σ-algebras

$$\sigma\left(\bigcup_{\lambda \in I_j} \mathscr{F}_\lambda\right), \qquad j \in J,$$

are independent.

Proof. For each $j \in J$, set

$$\mathscr{C}_j = \left\{\bigcap_{k=1}^{n} A_k : \ n \in \mathbb{Z}_+, \ \lambda_1, \ldots, \lambda_n \in I_j, \ A_i \in \mathscr{F}_{\lambda_i}\right\}.$$

If $A, B \in \mathscr{C}_j$, then $A = \bigcap_{k=1}^{n} A_k$ and $B = \bigcap_{i=1}^{m} B_i$, with $A_k \in \mathscr{F}_{\lambda_k}$ and $B_i \in \mathscr{F}_{\mu_i}$, where $\lambda_k, \mu_i \in I_j$. Then

$$A \cap B = \bigcap_{k=1}^{n} A_k \ \cap \ \bigcap_{i=1}^{m} B_i \in \mathscr{C}_j,$$

so $\mathscr{C}_j$ is a π-system for every $j \in J$. If $m \in \mathbb{N}$, $j_1, \ldots, j_m$ are pairwise distinct and $A_1 \in \mathscr{C}_{j_1}, \ldots, A_m \in \mathscr{C}_{j_m}$, write

$$A_1 = \bigcap_{i=1}^{n_1} A_1^i,\ n_1 \in \mathbb{N}, \lambda_1^1, \ldots, \lambda_{n_1}^1 \in I_1, A_1^1 \in \mathscr{F}_{\lambda_1^1}, \ldots, A_1^{n_1} \in \mathscr{F}_{\lambda_{n_1}^1}$$

$$\vdots$$

$$A_m = \bigcap_{i=1}^{n_m} A_m^i,\ n_m \in \mathbb{N}, \lambda_1^m, \ldots, \lambda_{n_m}^m \in I_m, A_m^1 \in \mathscr{F}_{\lambda_1^m}, \ldots, A_m^{n_m} \in \mathscr{F}_{\lambda_{n_m}^m}.$$

Then

$$\begin{aligned}\mathbb{P}\left(\bigcap_{k=1}^m A_k\right) &= \mathbb{P}\left(\bigcap_{1\le k\le m,\ 1\le i\le n_k} A_k^i\right) = \prod_{1\le k\le m,\ 1\le i\le n_k} \mathbb{P}(A_k^i) \\ &= \prod_{k=1}^m \prod_{i=1}^{n_k} \mathbb{P}(A_k^i) = \prod_{k=1}^m \mathbb{P}\left(\bigcap_{i=1}^{n_k} A_k^i\right) = \prod_{k=1}^m \mathbb{P}(A_k).\end{aligned}$$

By the previous theorem, the σ-algebras $\sigma(\mathscr{C}_j)$ are independent. Set $\mathscr{G}_j = \bigcup_{\lambda\in I_j} \mathscr{F}_\lambda$. Since $\mathscr{G}_j \subset \mathscr{C}_j \subset \sigma(\mathscr{G}_j)$ (the first inclusion by taking $n = 1$, the second because $\mathscr{C}_j$ consists of finite intersections of elements of $\mathscr{G}_j$), it follows that

$$\sigma(\mathscr{C}_j) = \sigma(\mathscr{G}_j) = \sigma\left(\bigcup_{\lambda\in I_j} \mathscr{F}_\lambda\right),$$

which is what we wanted. □

Exercise 8.10. Let $\mathscr{F}_1, \mathscr{F}_2, \ldots$ be independent σ-algebras. Then

1. $\sigma\left(\bigcup_{k\ge 1} \mathscr{F}_{2k}\right)$ and $\sigma\left(\bigcup_{k\ge 0} \mathscr{F}_{2k+1}\right)$ are independent.
2. $\mathscr{F}_1$, $\sigma(\mathscr{F}_2 \cup \mathscr{F}_3)$, $\sigma(\mathscr{F}_4 \cup \mathscr{F}_5 \cup \mathscr{F}_6)$, ... are independent.
3. If $A_1 \in \mathscr{F}_1$, $A_2 \in \mathscr{F}_2$ and $A_3 \in \mathscr{F}_3$, then $A_1 \cup A_2$ and A_3 are independent. Indeed, $A_1 \cup A_2 \in \sigma(\mathscr{F}_1 \cup \mathscr{F}_2)$.

8.2 σ-algebras generated by random vectors

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. If $X : \Omega \to \mathbb{R}^n$ is a random vector, the σ-algebra generated by X is defined by

$$\sigma(X) = \underbrace{\{X^{-1}(B) : \ B \in \mathscr{B}(\mathbb{R}^n)\}}_{\text{already a } \sigma\text{-algebra}}.$$

If $X_\lambda : \Omega \to \mathbb{R}^{n_\lambda}$, $\lambda \in \Lambda$, are random vectors, the σ-algebra generated by $\{X_\lambda : \lambda \in \Lambda\}$ is defined by

$$\sigma(\{X_\lambda : \lambda \in \Lambda\}) = \sigma\left(\{X_\lambda^{-1}(B) : \ \lambda \in \Lambda, \ B \in \mathscr{B}(\mathbb{R}^{n_\lambda})\}\right).$$

8.3 Independence of random vectors

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $X_\lambda : \Omega \to \mathbb{R}^{n_\lambda}$, $\lambda \in \Lambda$, random vectors. We say that the X_λ, $\lambda \in \Lambda$, are independent when the σ-algebras $\sigma(X_\lambda)$, $\lambda \in \Lambda$, are independent.

Remark. Let $X_\lambda : \Omega \to \mathbb{R}^{m_\lambda}$, $\lambda \in \Lambda$, be random vectors. The following are equivalent:

a) X_λ, $\lambda \in \Lambda$, are independent random vectors.

b) For every $n \in \mathbb{Z}_+$, all pairwise distinct $\lambda_1, \dots, \lambda_n \in \Lambda$ and all $B_1 \in \mathscr{B}(\mathbb{R}^{m_{\lambda_1}}), \dots, B_n \in \mathscr{B}(\mathbb{R}^{m_{\lambda_n}})$,

$$\mathbb{P}(X_{\lambda_1} \in B_1, \dots, X_{\lambda_n} \in B_n) = \prod_{k=1}^{n} \mathbb{P}(X_{\lambda_k} \in B_k).$$

c) For every $n \in \mathbb{Z}_+$, all pairwise distinct $\lambda_1, \dots, \lambda_n \in \Lambda$ and all $a_1 \in \mathbb{R}^{m_{\lambda_1}}, \dots, a_n \in \mathbb{R}^{m_{\lambda_n}}$,

$$\mathbb{P}(X_{\lambda_1} \leq a_1, \dots, X_{\lambda_n} \leq a_n) = \prod_{k=1}^{n} \mathbb{P}(X_{\lambda_k} \leq a_k).$$

Proposition 8.11. If the random vectors X_λ, $\lambda \in \Lambda$, are independent and $B_\lambda \in \mathscr{B}(\mathbb{R}^{m_\lambda})$ for each $\lambda \in \Lambda$, then the events

$$\{X_\lambda \in B_\lambda\}, \qquad \lambda \in \Lambda,$$

are independent.

Proposition 8.12. If the random vectors X_λ, $\lambda \in \Lambda$, are independent and $J \subset \Lambda$, then the random vectors X_λ, $\lambda \in J$, are independent.

Proposition 8.13. If the random vectors X_λ, $\lambda \in \Lambda$, are independent and the functions $f_\lambda : \mathbb{R}^{m_\lambda} \to \mathbb{R}$ are Borel measurable, then the random variables $f_\lambda(X_\lambda)$, $\lambda \in \Lambda$, are independent. This is because

$$\sigma(f_\lambda \circ X_\lambda) \subset \sigma(X_\lambda).$$

9 Lecture 9

If the random variables

$$\begin{gathered}X_1^{(1)}, X_2^{(1)}, \cdots, X_{n_1}^{(1)}\\X_1^{(2)}, X_2^{(2)}, \cdots, X_{n_2}^{(2)}\\\vdots\\X_1^{(k)}, X_2^{(k)}, \cdots, X_{n_k}^{(k)}\end{gathered}$$

are independent and the functions $f_j : \mathbb{R}^{n_j} \to \mathbb{R}^{m_j}$ are Borel measurable, then the random vectors

$$g_j = f_j(X_1^{(j)}, \ldots, X_{n_j}^{(j)})$$

are independent. Indeed,

$$\begin{aligned}X^{(j)} &= (X_1^{(j)}, \ldots, X_{n_j}^{(j)})\\g_j &= f_j(X^{(j)})\\\sigma(g_j) \subset \sigma(X^{(j)}) &= \sigma\left((X_1^{(j)}, \ldots, X_{n_j}^{(j)})\right)\\&= \sigma(\{X_1^{(j)}, \ldots, X_{n_j}^{(j)}\}).\end{aligned}$$

We know that the σ-algebras

$$\sigma(\{X_1^{(j)}, \ldots, X_{n_j}^{(j)}\}), \qquad j = 1, \ldots, k,$$

are independent, whence

$$\begin{aligned}&\Longrightarrow \sigma(g_j),\ j = 1, \ldots, k, \text{ are independent}\\&\Longrightarrow g_j,\ j = 1, \ldots, k, \text{ are independent.}\end{aligned}$$

Example 9.1. Suppose that

$$X, Y, Z, U, V, W, R, S, T$$

are independent random variables. Then the random vectors

$$X + Y + S + T, \quad (Z^2, e^Z), \ \frac{UV}{U^2 + V^2 + 1}$$

are independent, since they are measurable functions of disjoint blocks of the original variables.

9.1 Integration

Let $(\Omega, \mathscr{F}, \mu)$ be a measure space, that is,

$$\mu : \mathscr{F} \to [0, \infty]$$

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

$$A_1, A_2, \dots \in \mathscr{F}$$

$$A_i \cap A_j = \emptyset \;\; \forall\, i \neq j : \;\; \mu\left(\bigcup_{k \in \mathbb{N}} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k).$$

Definition 9.2. A function $f : \Omega \to \mathbb{R}$ is measurable if $f^{-1}(B) \in \mathscr{F}$ for every $B \in \mathscr{B}_{\mathbb{R}}$.

Definition 9.3. We say that a measurable function f is simple when f takes finitely many values; that is, when

$$f(\Omega) = \{f(\omega) : \; \omega \in \Omega\}$$

is a finite set.

Definition 9.4. We say that f is a nonnegative function when $f(\omega) \geq 0$ for every $\omega \in \Omega$.

Let $f : \Omega \to \mathbb{R}$ be measurable, simple and nonnegative. Then

$$f(\Omega) = \{x_1, \dots, x_n\},$$

and setting $A_k = f^{-1}(\{x_k\})$,

$$\Omega = \biguplus_{k=1}^{n} A_k, \qquad A_1, \dots, A_n \text{ pairwise disjoint,}$$

with

$$f(\omega) = \begin{cases} x_1, & \text{if } \omega \in A_1, \\ x_2, & \text{if } \omega \in A_2, \\ & \vdots \\ x_n, & \text{if } \omega \in A_n. \end{cases}$$

Hence

$$f = \sum_{k=1}^{n} x_k \mathbf{1}_{A_k}.$$

We define[5]

$$\int_\Omega f\, d\mu = \sum_{k=1}^{n} x_k \mu(A_k).$$

Remark. We adopt the usual measure-theoretic conventions:

$$\begin{aligned} a \cdot \infty &= \infty \quad \text{if } a > 0, \\ a \cdot \infty &= 0 \quad \text{if } a = 0. \end{aligned}$$

Proposition 9.5. Let $y_1, \dots, y_m \geq 0$ and let $B_1, \dots, B_m \in \mathscr{F}$ be pairwise disjoint with $\biguplus_j B_j = \Omega$, and set

$$g = \sum_{j=1}^{m} y_j \mathbf{1}_{B_j}.$$

Then g is measurable (a linear combination of measurable functions), simple (it takes at most m values) and nonnegative ($y_j \geq 0$), and

$$\int_\Omega g\, d\mu = \sum_{j=1}^{m} y_j \mu(B_j).$$

Moreover, the formula holds even if the B_j are not disjoint: the value of the integral does not depend on the chosen representation of g.

[5] The definition mirrors the interpretation of the integral as a weighted average:

$$\frac{1}{b-a}\int_a^b f(x)dx \simeq \underbrace{\frac{1}{b-a}\sum_{k=1}^{n} f(x_k^*)(t_k - t_{k-1})}_{\text{weighted average}},$$

where the weights add up to one:

$$\frac{1}{b-a}\sum_{k=1}^{n}(t_k - t_{k-1}) = 1 = \sum_{k=1}^{n} \lambda_k.$$

Let $f : \Omega \to \overline{\mathbb{R}}$ be a nonnegative measurable function, where

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$$
$$\mathscr{B}(\overline{\mathbb{R}}) = \{A = B \cup C : \ B \in \mathscr{B}(\mathbb{R}) \ \wedge \ C \subset \{-\infty, \infty\}\}.$$

Definition 9.6. We define

$$\int_\Omega f\, d\mu = \sup \left\{ \int_\Omega g\, d\mu : \ g \text{ measurable, simple, nonnegative, and } g \leq f \right\}.$$

Notation. We extend the supremum to $\overline{\mathbb{R}}$: given $A \subset \mathbb{R}$,

$$\sup\ A = \begin{cases} \text{the usual supremum} & \text{if } A \neq \emptyset \text{ and } A \text{ is bounded above,} \\ +\infty & \text{if } A \neq \emptyset \text{ and } A \text{ is not bounded above,} \\ -\infty & \text{if } \ A = \emptyset. \end{cases}$$

Proposition 9.7. Let $f : \Omega \to \overline{\mathbb{R}}$ be a nonnegative measurable function. For each $n \geq 1$, set

$$f_n(\omega) = \sum_{k=0}^{n2^n - 1} \frac{k}{2^n} \mathbf{1}_{f^{-1}\left(\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right)\right)}(\omega) + n\, \mathbf{1}_{f^{-1}([n,\infty])}(\omega).$$

Then $f_n \uparrow f$, that is,

$$\begin{cases} f_1(\omega) \leq f_2(\omega) \leq \cdots & \forall\, \omega \in \Omega, \\ \lim_{n\to\infty} f_n(\omega) = f(\omega) & \forall\, \omega \in \Omega. \end{cases}$$

Proposition 9.8. Let $f, g : \Omega \to \overline{\mathbb{R}}$ be nonnegative measurable functions.

1. If $\alpha \geq 0$, then αf is nonnegative measurable and $\int \alpha f\, d\mu = \alpha \int f\, d\mu$.

2. $f + g$ is nonnegative measurable and $\int (f+g)\, d\mu = \int f\, d\mu + \int g\, d\mu$.

3. If $f \leq g$, then $\int f\, d\mu \leq \int g\, d\mu$.

Lemma 9.9. For f and g measurable, simple and nonnegative, $\int (f+g)\, d\mu = \int f\, d\mu + \int g\, d\mu$.

Proof. Write

$$f = \sum_{i=1}^{r} x_i \mathbf{1}_{A_i}, \qquad g = \sum_{j=1}^{s} y_j \mathbf{1}_{B_j},$$

where $\{A_i\}_{i=1}^r$ and $\{B_j\}_{j=1}^s$ are measurable partitions of Ω. The sets $A_i \cap B_j$ form a common refinement: they are pairwise disjoint and their union is Ω. On $A_i \cap B_j$ the function $f + g$ is constant, equal to $x_i + y_j$, so that

$$\begin{aligned}
\int (f+g)\, d\mu &= \sum_{i=1}^{r} \sum_{j=1}^{s} (x_i + y_j) \mu(A_i \cap B_j) \\
&= \sum_{i=1}^{r} x_i \sum_{j=1}^{s} \mu(A_i \cap B_j) + \sum_{j=1}^{s} y_j \sum_{i=1}^{r} \mu(A_i \cap B_j) \\
&= \sum_{i=1}^{r} x_i \mu(A_i) + \sum_{j=1}^{s} y_j \mu(B_j) = \int f\, d\mu + \int g\, d\mu,
\end{aligned}$$

using the finite additivity of μ. □

Proof. Item (3) is immediate from the definition of the integral as a supremum, since every nonnegative simple function $h \le f$ also satisfies $h \le g$. Item (1) follows in the same way, noting that $h \mapsto \alpha h$ is a bijection between the nonnegative simple functions dominated by f and those dominated by αf (for $\alpha > 0$; the case $\alpha = 0$ is trivial).

For (2): there exist sequences (f_n) and (g_n) of nonnegative measurable simple functions with $f_n \uparrow f$ and $g_n \uparrow g$. By Lemma 9.9,

$$\int (f_n + g_n)\, d\mu = \int f_n\, d\mu + \int g_n\, d\mu, \qquad n \ge 1.$$

Since $f_n \uparrow f$, $g_n \uparrow g$ and $f_n + g_n \uparrow f + g$, Theorem 9.10 lets us pass to the limit on both sides and conclude that $\int (f+g)\, d\mu = \int f\, d\mu + \int g\, d\mu$. □

Theorem 9.10. Monotone convergence theorem (MCT). Let $f, f_1, f_2, \ldots$ be nonnegative measurable functions. If $f_n \uparrow f$, then $\int f_n\, d\mu \to \int f\, d\mu$.

10 Lecture 10

Lemma 10.1. Fatou's lemma. Let $(\Omega, \mathscr{F}, \mu)$ be a measure space and $f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}$ nonnegative measurable functions. Then

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

Remark. Set $x_n = \int_\Omega f_n \, d\mu \in [0, \infty]$. Then

$$\liminf x_n = \sup_{n \ge 1} \inf\{x_k : k \ge n\} = \lim_{n \to \infty} \inf\{x_k : k \ge n\},$$

so the $\liminf$ of the integrals is well defined. On the other hand,

$$h(\omega) = \liminf_n f_n(\omega) \in [0, \infty]$$

is a well-defined measurable function.

Exercise 10.2. Let $f : \Omega \to \overline{\mathbb{R}}$ be a nonnegative measurable function. Show that

$$\int_\Omega f \, d\mu = 0 \implies f = 0 \text{ a.s.}$$

Example 10.3. On the measure space $(\mathbb{R}, \mathscr{B}_{\mathbb{R}}, m)$, let $f_n = \mathbf{1}_{[n, n+1)}$. Then

$$\lim_n f_n = 0,$$

but

$$\lim_n \int f_n = \lim_n 1 = 1 \ne 0 = \int \lim_n f_n.$$

Example 10.4. Now take $([0,1], \mathscr{B}_{[0,1]}, m)$ and $f_n = n\mathbf{1}_{(0,1/n)}$, $n \ge 1$. Then

$$\lim_n f_n(x) = 0 \quad \text{pointwise on } [0,1],$$

but

$$\lim_n \underbrace{\int_{[0,1]} f_n \, dx}_{= n\, m((0,1/n)) = 1} = 1 \ne 0 = \int_{[0,1]} 0 \, dx.$$

Both examples show that the inequality in Fatou's lemma may be strict.

Theorem 10.5. Fix $(\Omega, \mathscr{F}, \mu)$ and let $f : \Omega \to \overline{\mathbb{R}}$ be a nonnegative measurable function. Then

a) the function $\nu : \mathscr{F} \to [0, \infty]$ given by

$$\nu(A) = \int_\Omega f \cdot \mathbf{1}_A \, d\mu$$

is a measure on $(\Omega, \mathscr{F})$;

b) if $g : \Omega \to \overline{\mathbb{R}}$ is a nonnegative measurable function, then

$$\int_\Omega g \, d\nu = \int_\Omega g \cdot f \, d\mu.$$

Proof. Item by item:

a) Clearly $\nu(\emptyset) = \int_\Omega f \cdot \mathbf{1}_\emptyset \, d\mu = 0$. If $A_1, A_2, \dots \in \mathscr{F}$ are pairwise disjoint and $A = \biguplus_k A_k$, then the partial sums $f \cdot \mathbf{1}_{\bigcup_{k \le n} A_k}$ increase to $f \cdot \mathbf{1}_A$, so by the MCT and finite additivity of the integral,

$$\nu(A) = \int_\Omega f \, \mathbf{1}_A \, d\mu = \lim_n \sum_{k=1}^n \int_\Omega f \, \mathbf{1}_{A_k} \, d\mu = \sum_{k=1}^\infty \nu(A_k).$$

Hence ν is a measure.

b) Suppose first that g is simple and nonnegative: $g = \sum_{k=1}^n x_k \mathbf{1}_{A_k}$. Then

$$\int_\Omega g \, d\nu = \sum_{k=1}^n x_k \nu(A_k) = \sum_{k=1}^n x_k \int_\Omega f \mathbf{1}_{A_k} \, d\mu = \int_\Omega f \sum_{k=1}^n x_k \mathbf{1}_{A_k} \, d\mu.$$

Now suppose g is nonnegative. There is a sequence $g_n \uparrow g$ of nonnegative simple functions, and

$$\int_\Omega g_n \, d\nu = \int_\Omega g_n f \, d\mu.$$

Since $g_n \uparrow g$ and $g_n f \uparrow gf$, the monotone convergence theorem gives

$$\int_\Omega g \, d\nu = \int_\Omega g f \, d\mu.$$

$\square$

Theorem 10.6. Change of variables. Let $(\Omega_1, \mathscr{F}_1)$ and $(\Omega_2, \mathscr{F}_2)$ be measurable spaces, $f : \Omega_1 \to \Omega_2$ a measurable function and $g : \Omega_2 \to \overline{\mathbb{R}}$ a nonnegative measurable function.

a) If μ is a measure on $(\Omega_1, \mathscr{F}_1)$, then the function $\nu = \mu f^{-1} : \mathscr{F}_2 \to [0, \infty]$, $A \mapsto \mu(f^{-1}(A))$, is a measure on $(\Omega_2, \mathscr{F}_2)$.

b) $\int_{\Omega_2} g \, d\nu = \int_{\Omega_1} g \circ f \, d\mu$.

Proof. Item by item:

a) $\nu(\emptyset) = \mu(f^{-1}(\emptyset)) = \mu(\emptyset) = 0$, and if the $A_k \in \mathscr{F}_2$ are pairwise disjoint, so are their preimages $f^{-1}(A_k)$, with $f^{-1}(\biguplus_k A_k) = \biguplus_k f^{-1}(A_k)$; the σ-additivity of ν follows from that of μ.

b) Suppose $g = \sum_{k=1}^n x_k \mathbf{1}_{A_k}$. Then

$$\int_{\Omega_2} g \, d\nu = \sum_{k=1}^n x_k \nu(A_k) = \sum_{k=1}^n x_k \mu(f^{-1}(A_k))$$

and

$$\int_{\Omega_1} \underbrace{g \circ f}_{=\sum_{k=1}^n x_k \mathbf{1}_{f^{-1}(A_k)}} d\mu = \sum_{k=1}^n x_k \mu(f^{-1}(A_k)).$$

Now suppose g is nonnegative. There exists a sequence (g_n) of nonnegative measurable simple functions with $g_n \uparrow g$, and

$$\int_{\Omega_2} g_n \, d\nu = \int_{\Omega_1} g_n \circ f \, d\mu, \qquad n \geq 1.$$

By the MCT, $\int_{\Omega_2} g_n \, d\nu \to \int_{\Omega_2} g \, d\nu$ and $\int_{\Omega_1} g_n \circ f \, d\mu \to \int_{\Omega_1} g \circ f \, d\mu$. Hence

$$\int_{\Omega_2} g \, d\nu = \int_{\Omega_1} g \circ f \, d\mu.$$

□

10.1 Integration of measurable functions

Let $f : \Omega \to \overline{\mathbb{R}}$ be a measurable function. We define the positive part and the negative part of f,

$$f^+ : \Omega \to \overline{\mathbb{R}}, \quad f^- : \Omega \to \overline{\mathbb{R}},$$

by

$$f^+(\omega) = \begin{cases} f(\omega), & \text{if } f(\omega) \geq 0, \\ 0, & \text{if } f(\omega) < 0, \end{cases}$$

and

$$f^-(\omega) = \begin{cases} 0, & \text{if } f(\omega) > 0, \\ -f(\omega), & \text{if } f(\omega) \leq 0. \end{cases}$$

Remark. The functions $\varphi : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ and $\psi : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ given by

$$\varphi(x) = x^+ = \begin{cases} x, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0, \end{cases} \quad \text{and} \quad \psi(x) = x^- = \begin{cases} 0, & \text{if } x > 0, \\ -x, & \text{if } x \leq 0, \end{cases}$$

are measurable, and $f^+ = \varphi \circ f$, $f^- = \psi \circ f$. In particular, f^+ and f^- are nonnegative measurable functions.

Remark. One has $x = x^+ - x^-$ and $|x| = x^+ + x^-$ for every $x \in \overline{\mathbb{R}}$. In particular, $f = f^+ - f^-$ and $|f| = f^+ + f^-$.

When at least one of the integrals $\int f^+$, $\int f^-$ is finite, we say that the integral of f with respect to μ is well defined, with value

$$\int_\Omega f \, d\mu = \int_\Omega f^+ \, d\mu - \int_\Omega f^- \, d\mu.$$

When both $\int f^+$ and $\int f^-$ are finite, we say that f is μ-integrable.

Proposition 10.7. Let $f : \Omega \to \overline{\mathbb{R}}$ be a measurable function. Then

$$f \text{ is integrable} \iff \int_\Omega |f| \, d\mu < \infty.$$

Proposition 10.8. Let f and g be integrable measurable functions.

1. If $c \in \mathbb{R}$, then cf is integrable and

$$\int cf \, d\mu = c \int f \, d\mu.$$

2. $f + g$ is integrable and $\int (f+g) \, d\mu = \int f \, d\mu + \int g \, d\mu$.

3. If $f \leq g$, then $\int f \, d\mu \leq \int g \, d\mu$. In particular, $\left| \int f \, d\mu \right| \leq \int |f| \, d\mu$.

4. If $A_1, \dots, A_n \in \mathscr{F}$ are pairwise disjoint, then

$$\int_{\bigcup_{k=1}^n A_k} f \, d\mu = \sum_{k=1}^n \int_{A_k} f \, d\mu.$$

5. The same holds for a countably infinite family:

$$\int_{\bigcup_{k=1}^\infty A_k} f \, d\mu = \sum_{k=1}^\infty \int_{A_k} f \, d\mu.$$

6. If $A_1, A_2, \dots \in \mathscr{F}$ and either $A_n \downarrow A$ or $A_n \uparrow A$, then

$$\lim_n \int_{A_n} f \, d\mu = \int_A f \, d\mu.$$

Remark. To prove Proposition 10.8 we use:

- for $c > 0$: $(cf)^+ = cf^+$ and $(cf)^- = cf^-$;
- for $c < 0$: $(cf)^+ = -cf^-$ and $(cf)^- = -cf^+$;
- $(f+g)^+ \leq f^+ + g^+$ and $(f+g)^- \leq f^- + g^-$;
- if $f \leq g$, then $f^+ \leq g^+$ and $g^- \leq f^-$.

Exercise 10.9. Let $f, g : \Omega \to \overline{\mathbb{R}}$ be measurable functions. Show that

1. if $f = 0$ a.s., then $\int f \, d\mu = 0$;

2. if f is integrable and $g = f$ a.s., then g is integrable and $\int f \, d\mu = \int g \, d\mu$.

Theorem 10.10. Let $f, g, f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}$ be measurable functions such that

1. $\lim_n f_n = f$ (it suffices that this hold a.s.);
2. $|f_n| \leq g$ for every $n \geq 1$ (it suffices that this hold a.s.);
3. g is integrable.

Then f is integrable and $\lim_n \int f_n \, d\mu = \int f \, d\mu$. This is the **dominated convergence theorem (DCT)**.

11 Lecture 11

Corollary 11.1. Let $f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}$ be measurable functions such that

$$\sum_{k=1}^{\infty} \int |f_k|\, d\mu < \infty.$$

Then $f = \sum_{k=1}^{\infty} f_k$ is defined a.s. (the series $\sum_{k=1}^{\infty} f_k(\omega)$ converges absolutely for almost every ω), is finite almost surely and integrable, and

$$\int f\, d\mu = \sum_{k=1}^{\infty} \int f_k\, d\mu.$$

Proof. This follows from the dominated convergence theorem. By the MCT, $\int \sum_{k\geq 1} |f_k|\, d\mu = \sum_{k\geq 1} \int |f_k|\, d\mu < \infty$, so that $g := \sum_{k\geq 1} |f_k|$ is integrable and, in particular, finite a.s.; hence the series $\sum_k f_k(\omega)$ converges absolutely for almost every ω. Consider the partial sums

$$S_N = \sum_{k=1}^{N} f_k.$$

Then $|S_N| \leq \sum_{k=1}^{N} |f_k| \leq g$, which is integrable. Therefore

$$\lim_N \int S_N\, d\mu = \int \lim_N S_N\, d\mu,$$

which gives

$$\sum_{k=1}^{\infty} \int f_k\, d\mu = \lim_N \underbrace{\int S_N\, d\mu}_{=\sum_{k=1}^{N} \int f_k\, d\mu} = \int \sum_{k=1}^{\infty} f_k\, d\mu.$$

□

Corollary 11.2. Bounded convergence theorem. Let $(\Omega, \mathscr{F}, \mu)$ be a finite measure space, $f, f_1, f_2, \dots : \Omega \to \overline{\mathbb{R}}$ measurable functions and $M > 0$ a constant such that

1. $|f_n| \leq M$ a.s. for every $n \geq 1$;

2. $f_n \to f$ a.s.

Then

$$\int f_n \, d\mu \to \int f \, d\mu.$$

Proof. A consequence of the DCT with $g \equiv M$, since $\int_\Omega |g| \, d\mu = M\mu(\Omega) < \infty$. $\square$

Theorem 11.3. Change of measure. Let $(\Omega, \mathscr{F}, \mu)$ be a measure space, $f : \Omega \to \overline{\mathbb{R}}$ a nonnegative measurable function, $d\nu = f \, d\mu$, and $g : \Omega \to \overline{\mathbb{R}}$ measurable. Then

a) g is ν-integrable $\iff$ gf is μ-integrable;

b) if g is ν-integrable, then $\int_\Omega g \, d\nu = \int_\Omega gf \, d\mu$.

Theorem 11.4. Change of variables. Let $(\Omega_1, \mathscr{F}_1)$ and $(\Omega_2, \mathscr{F}_2)$ be measurable spaces, μ a measure on $(\Omega_1, \mathscr{F}_1)$, $f : \Omega_1 \to \Omega_2$ measurable, $\nu = \mu f^{-1}$, and $g : \Omega_2 \to \overline{\mathbb{R}}$ measurable. Then

1. g is ν-integrable $\iff$ $g \circ f$ is μ-integrable;

2. if g is ν-integrable, then $\int_{\Omega_2} g \, d\nu = \int_{\Omega_1} g \circ f \, d\mu$.

11.1 Outer measure

Let Ω be a nonempty set. An outer measure on Ω is a function $\mu^* : 2^\Omega \to [0, \infty]$ such that

1. $\mu^*(\emptyset) = 0$;

2. if $A \subset B \subset \Omega$, then $\mu^*(A) \leq \mu^*(B)$;

3. if $A_1, A_2, \dots \subset \Omega$, then $\mu^*\left(\bigcup_{k=1}^\infty A_k\right) \leq \sum_{k=1}^\infty \mu^*(A_k)$.

Theorem 11.5. Let $\mathscr{C}$ be a family of subsets of Ω and $\mu : \mathscr{C} \to [0, \infty]$ such that

1. $\emptyset \in \mathscr{C}$;

2. $\mu(\emptyset) = 0$.

Then the function $\mu^* : 2^\Omega \to [0, \infty]$ defined by

$$\mu^*(A) = \inf \left\{ \sum_{k \geq 1} \mu(C_k) : \ C_1, C_2, \dots \in \mathscr{C} \ \wedge \ \bigcup_{k \geq 1} C_k \supset A \right\}$$

is an outer measure on Ω.

Theorem 11.6. Let Ω be a nonempty set and $\mu^* : 2^\Omega \to [0, \infty]$ an outer measure on Ω. We say that $E \subset \Omega$ is μ^*-measurable when

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c), \qquad \forall \ A \subset \Omega.$$

Then

1. $\mathcal{M}_{\mu^*} = \{E \subset \Omega : E \text{ is } \mu^*\text{-measurable}\}$ is a σ-algebra on Ω;

2. $\overline{\mu} = \mu^*\big|_{\mathcal{M}_{\mu^*}} : \mathcal{M}_{\mu^*} \to [0, \infty]$ is a measure on $(\Omega, \mathcal{M}_{\mu^*})$.

Theorem 11.7. Carathéodory extension theorem. Suppose that

1. $\mathscr{C}$ is a semialgebra on Ω; that is, $\emptyset, \Omega \in \mathscr{C}$; if $A_1, A_2 \in \mathscr{C}$, then $A_1 \cap A_2 \in \mathscr{C}$; and if $A \in \mathscr{C}$, then $\Omega - A$ is a finite union of pairwise disjoint elements of $\mathscr{C}$;

2. $\mu(\emptyset) = 0$;

3. if $A_1, A_2, \dots \in \mathscr{C}$ are pairwise disjoint and $A = \bigcup_{k=1}^\infty A_k \in \mathscr{C}$, then $\mu(A) = \sum_{k=1}^\infty \mu(A_k)$.

Then

a) $\overline{\mu}$ is a complete measure on $(\Omega, \mathcal{M}_{\mu^*})$;

b) $\overline{\mu}$ extends μ: $\mathscr{C} \subset \mathcal{M}_{\mu^*}$ and $\overline{\mu}(A) = \mu(A)$ for every $A \in \mathscr{C}$.

11.2 Product measure

Let $(\Omega_1, \mathscr{F}_1, \mu_1)$ and $(\Omega_2, \mathscr{F}_2, \mu_2)$ be measure spaces. Write $\Omega = \Omega_1 \times \Omega_2 = \{(x,y) : x \in \Omega_1,\ y \in \Omega_2\}$,

$$\mathcal{R} = \{A \times B :\ A \in \mathscr{F}_1,\ B \in \mathscr{F}_2\},$$

$$\mu_0 : \mathcal{R} \to [0,\infty], \qquad A \times B \mapsto \mu_1(A)\,\mu_2(B),$$

and $\mathscr{F} = \mathscr{F}_1 \otimes \mathscr{F}_2 = \sigma(\mathcal{R})$. One has

1. $\mathcal{R}$ is a semialgebra on Ω;
2. $\mu_0(\emptyset) = \mu_1(\emptyset)\mu_2(\emptyset) = 0$;
3. μ_0 is σ-additive on $\mathcal{R}$,

so the Carathéodory extension theorem applies and yields a measure on $(\Omega, \mathscr{F})$ extending μ_0; it is unique if μ_1 and μ_2 are σ-finite. The measure $\mu = \mu_1 \times \mu_2$ is called the product measure of μ_1 and μ_2, and $(\Omega, \mathscr{F}, \mu)$ is the product measure space of $(\Omega_1, \mathscr{F}_1, \mu_1)$ and $(\Omega_2, \mathscr{F}_2, \mu_2)$.

11.3 Tonelli's theorem

Let $(\Omega_1, \mathscr{F}_1, \mu_1)$ and $(\Omega_2, \mathscr{F}_2, \mu_2)$ be σ-finite measure spaces[6] and let $f : \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}$ be a nonnegative $\mathscr{F}_1 \otimes \mathscr{F}_2$-measurable function. Then:

1. For each $x \in \Omega_1$, the function $f_x : \Omega_2 \to \overline{\mathbb{R}}$, $y \mapsto f(x,y)$, is $\mathscr{F}_2$-measurable and nonnegative.
2. The function $\varphi : \Omega_1 \to \overline{\mathbb{R}}$, $x \mapsto \int_{\Omega_2} f(x,y)\, d\mu_2(y)$, is $\mathscr{F}_1$-measurable.
3. For each $y \in \Omega_2$, the function $f^y : \Omega_1 \to \overline{\mathbb{R}}$, $x \mapsto f(x,y)$, is $\mathscr{F}_1$-measurable and nonnegative.
4. The function $\psi : \Omega_2 \to \overline{\mathbb{R}}$, $y \mapsto \int_{\Omega_1} f^y(x)\, d\mu_1(x)$, is $\mathscr{F}_2$-measurable.

[6] That is, μ_i is a measure on Ω_i and there exist $\{A^i_k\}_{k\geq 1} \subset \mathscr{F}_i$ with $\mu_i(A^i_k) < \infty$ for every k and $\bigcup_{k=1}^{\infty} A^i_k = \Omega_i$.

5. Finally,

$$\int_{\Omega_1 \times \Omega_2} f(x,y)\, d(\mu_1 \times \mu_2)(x,y) = \int_{\Omega_1} \varphi(x)\, d\mu_1(x) = \int_{\Omega_2} \psi(y)\, d\mu_2(y).$$

12 Lecture 12

Theorem 12.1. Fubini's theorem. Let $(\Omega_1, \mathscr{F}_1, \mu_1)$ and $(\Omega_2, \mathscr{F}_2, \mu_2)$ be σ-finite measure spaces and let $f : \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}$ be an integrable $\mathscr{F}_1 \otimes \mathscr{F}_2$-measurable function. Then:

1. For each $x \in \Omega_1$, the function $f_x : \Omega_2 \to \overline{\mathbb{R}}$, $y \mapsto f(x, y)$, is $\mathscr{F}_2$-measurable.

2. f_x is μ_2-integrable for μ_1-a.e. x.

3. The function $\varphi : \Omega_1 \to \overline{\mathbb{R}}$, $x \mapsto \int_{\Omega_2} f_x(y)\, d\mu_2(y)$, defined μ_1-a.e., is $\mathscr{F}_1$-measurable and μ_1-integrable.

4. For each $y \in \Omega_2$, the function $f^y : \Omega_1 \to \overline{\mathbb{R}}$, $x \mapsto f(x, y)$, is $\mathscr{F}_1$-measurable.

5. f^y is μ_1-integrable for μ_2-a.e. y.

6. The function $\psi : \Omega_2 \to \overline{\mathbb{R}}$, $y \mapsto \int_{\Omega_1} f(x, y)\, d\mu_1(x)$, defined μ_2-a.e., is $\mathscr{F}_2$-measurable and μ_2-integrable.

7. $\int_{\Omega_1 \times \Omega_2} f(x, y)\, d(\mu_1 \times \mu_2)(x, y) = \int_{\Omega_1} \varphi(x)\, d\mu_1(x) = \int_{\Omega_2} \psi(y)\, d\mu_2(y)$.

Remark. Recall that f is integrable if and only if $\int |f|\, d\mu < \infty$. Hence, to decide whether Fubini's theorem applies, one usually applies Tonelli's theorem to $|f|$ first.

Remark. If $(\Omega_1, \mathscr{F}_1, \mu_1)$ and $(\Omega_2, \mathscr{F}_2, \mu_2)$ are probability spaces, then $(\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2, \mu_1 \times \mu_2)$ is a probability space as well.

Remark. If $E \in \mathscr{F}_1 \otimes \mathscr{F}_2$, then

$$(\mu_1 \times \mu_2)(E) = \inf \left\{ \sum_{k \geq 1} \mu_1(A_k)\mu_2(B_k) \ : A_k \in \mathscr{F}_1,\ B_k \in \mathscr{F}_2, \right.$$
$$\left. E \subset \bigcup_k A_k \times B_k \right\}.$$

Remark. Both Tonelli's and Fubini's theorems generalize to finitely many measure spaces, not just two.

Example 12.2. Let $\Omega_1 = \Omega_2 = [0,1]$, $\mathscr{F}_1 = \mathscr{F}_2 = \mathscr{B}[0,1]$ and let $D = \{(x,y) \in [0,1] \times [0,1] : \ x = y\}$ be the diagonal. We compute $(\mu_1 \times \mu_2)(D)$ in the following cases:

a) μ_1 and μ_2 are both the counting measure;

b) μ_1 and μ_2 are both the Lebesgue measure;

c) μ_1 is the Lebesgue measure and μ_2 is the counting measure.

One has:

a) If $z_1, z_2, \dots$ are distinct points of D, then $(\mu_1 \times \mu_2)(D) \geq \sum_{k=1}^n (\mu_1 \times \mu_2)(\{z_k\}) = n$ for every n, since each point has product measure 1. Hence $(\mu_1 \times \mu_2)(D) = \infty$.

b) Covering D with the n squares $[\frac{k-1}{n}, \frac{k}{n}]^2$, $k = 1, \dots, n$, gives $(\mu_1 \times \mu_2)(D) \leq \sum_{k=1}^n \frac{1}{n^2} = \frac{1}{n}$ for every n. Hence $(\mu_1 \times \mu_2)(D) = 0$.

c) We claim that $(\mu_1 \times \mu_2)(D) = \infty$. Write $\mu = \mu_1$ (Lebesgue) and $\nu = \mu_2$ (counting). It suffices to show that if $D \subset \bigcup_k A_k \times B_k$ with $A_k, B_k \in \mathscr{B}[0,1]$, then $\sum_{k=1}^\infty \mu(A_k)\nu(B_k) = \infty$.

Indeed, in case (c) the set D is covered by the $A_k \times B_k$, $k \in \mathbb{N}$. Let $M \subset \mathbb{N}$ be the set of indices with $\mu(A_k) = 0$, and let $S = \bigcup_{k \in M} A_k$; then $\mu(S) = 0$, and therefore $[0,1] \setminus S$ is uncountable. The set $F = \{(x,x) : x \in [0,1] \setminus S\}$ is contained in D and covered by the sets $A_k \times B_k$ with $k \in \mathbb{N} \setminus M$. Hence, for each $x \in [0,1] \setminus S$ there is $k \in \mathbb{N} \setminus M$ with $(x,x) \in A_k \times B_k$ and, in particular, $x \in B_k$. Thus $[0,1] \setminus S \subset \bigcup_{k \notin M} B_k$. Since $[0,1] \setminus S$ is uncountable, at least one B_{k_0} is uncountable, and then $\nu(B_{k_0}) = \infty$. But $k_0 \notin M$, so $\mu(A_{k_0}) > 0$ and therefore $\mu(A_{k_0})\nu(B_{k_0}) = \infty$. Hence

$$\sum_k \mu(A_k)\nu(B_k) = \infty,$$

as claimed. Note that (b) and (c) show that the σ-finiteness hypothesis in Tonelli's and Fubini's theorems is not superfluous.

12.1 Radon–Nikodym

Let μ and ν be measures on the measurable space $(\Omega, \mathscr{F})$. We say that ν is absolutely continuous with respect to μ when

$$\nu(A) = 0 \quad \text{for every } A \in \mathscr{F} \text{ with } \mu(A) = 0.$$

We denote this by $\nu \ll \mu$, and we also say that μ dominates ν.

Definition 12.3. We say that μ and ν are equivalent when $\nu \ll \mu$ and $\mu \ll \nu$.

Example 12.4. Let ν and m be, respectively, the counting and the Lebesgue measures on $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$. Then

$$m \ll \nu, \qquad \nu \not\ll m.$$

Indeed, if $\nu(A) = 0$, then $A = \emptyset$ and $m(\emptyset) = 0$. On the other hand, $\{0\} \in \mathscr{B}(\mathbb{R})$ satisfies $m(\{0\}) = 0$ but $\nu(\{0\}) = 1 \neq 0$.

Remark. If $(\Omega, \mathscr{F})$ is a measurable space, μ a measure on $(\Omega, \mathscr{F})$, $f : \Omega \to \overline{\mathbb{R}}$ a nonnegative measurable function and $d\nu = f\, d\mu$, then $\nu \ll \mu$.

Proof. If $A \in \mathscr{F}$ and $\mu(A) = 0$, then

$$\nu(A) = \int_A f\, d\mu = \int_\Omega f \mathbf{1}_A\, d\mu = 0.$$

□

Remark. Does there exist a nonnegative measurable $f : \mathbb{R} \to \overline{\mathbb{R}}$ with $dm = f\, d\nu$? **No.** Suppose such an f existed. Then

$$0 = m(\{a\}) = \int_{\{a\}} f\, d\nu = f(a)\nu(\{a\}) = f(a), \qquad \forall\, a \in \mathbb{R},$$

and therefore

$$1 = m([0,1]) = \int_{[0,1]} f\, d\nu = \int_{[0,1]} 0\, d\nu = 0,$$

which is absurd.

Theorem 12.5. Radon–Nikodym theorem. Let μ and ν be σ-finite measures on the measurable space $(\Omega, \mathscr{F})$ with $\nu \ll \mu$. Then

1. there exists a nonnegative measurable $f : \Omega \to \overline{\mathbb{R}}$ with $d\nu = f\, d\mu$, written $f = \frac{d\nu}{d\mu}$;
2. f is unique μ-a.e.: if g is another function with the same property, then $g = f$ μ-a.e.

12.2 Expectation of a random variable

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \overline{\mathbb{R}}$ a random variable. The mean, or expectation, of X is defined by

$$\mathbb{E}[X] = \int_\Omega X\, d\mathbb{P}.$$

Proposition 12.6. Let $X : \Omega \to \mathbb{R}^n$ be a random vector and $g : \mathbb{R}^n \to \mathbb{R}$ a Borel measurable function. Then

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}^n} g(x)\, dP_X(x).$$

Definition 12.7. Let $(\Omega_1, \mathscr{F}_1, \mathbb{P}^1)$ and $(\Omega_2, \mathscr{F}_2, \mathbb{P}^2)$ be probability spaces, and $X : \Omega_1 \to \mathbb{R}^n$, $Y : \Omega_2 \to \mathbb{R}^n$ random vectors. We say that X and Y are identically distributed when $\mathbb{P}^1_X = \mathbb{P}^2_Y$. We denote this by $X \sim Y$.

Proposition 12.8. If $X \sim Y$, then:

1. if $g : \mathbb{R}^n \to \mathbb{R}$ is Borel measurable, then $g(X) \sim g(Y)$;
2. if $n = 1$ and X is integrable, then so is Y, and $\mathbb{E}[X] = \mathbb{E}[Y]$.

Proposition 12.9. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \overline{\mathbb{R}}$ a nonnegative r.v. Then

$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X > x)\, dx = \int_0^\infty \mathbb{P}(X \geq x)\, dx.$$

Proof. We have

$$\int_0^\infty \mathbb{P}(X > x)\,dx = \int_0^\infty \left[\int_\Omega \mathbf{1}_{X>x}(\omega)\,d\mathbb{P}(\omega)\right] dx,$$

where

$$\mathbf{1}_{X>x}(\omega) = \begin{cases} 1, & \text{if} \quad \underbrace{X(\omega) > x}_{\text{that is, } x \in [0, X(\omega))}, \\ 0, & \text{otherwise.} \end{cases}$$

By Tonelli's theorem,

$$\begin{aligned} \int_0^\infty \left[\int_\Omega \mathbf{1}_{X>x}(\omega)\,d\mathbb{P}(\omega)\right] dx &= \int_\Omega \left[\int_0^\infty \mathbf{1}_{[0,X(\omega))}(x)\,dx\right] d\mathbb{P}(\omega) \\ &= \int_\Omega X(\omega)\,d\mathbb{P}(\omega) \\ &= \mathbb{E}[X]. \end{aligned}$$

□

Remark. Recall that $\mathbb{P}(X \le x) = F_X(x)$. Hence

$$\mathbb{E}[X] = \int_0^\infty [1 - F_X(x)]\,dx.$$

13 Lecture 13

Theorem 13.1. Let $X : \Omega \to \mathbb{R}$ be a nonnegative random variable. Then

$$\sum_{n=1}^{\infty} \mathbb{P}(X \geq n) \leq \mathbb{E}[X] \leq 1 + \sum_{n=1}^{\infty} \mathbb{P}(X \geq n).$$

Proof. We prove the second inequality first. For $n \geq 1$, set $A_n = \{n-1 \leq X < n\}$ and $B_n = \{X \geq n\}$[7]. Then

$$\begin{aligned}
\mathbb{E}[X] &= \int_{\Omega} X \, d\mathbb{P} \\
&= \sum_{n=1}^{\infty} \int_{A_n} X \, d\mathbb{P} \\
&\leq \sum_{n=1}^{\infty} \int_{A_n} n \, d\mathbb{P} \\
&= \sum_{n=1}^{\infty} n \, \mathbb{P}(A_n) \\
&= \sum_{n=1}^{\infty} n[\mathbb{P}(B_{n-1}) - \mathbb{P}(B_n)] \\
&= \lim_n \sum_{k=1}^{n} k[\mathbb{P}(B_{k-1}) - \mathbb{P}(B_k)] \\
&= \lim_n \left[\mathbb{P}(B_0) + \sum_{k=1}^{n-1} \mathbb{P}(B_k) - n \, \mathbb{P}(B_n) \right] \\
&\leq 1 + \lim_n \sum_{k=1}^{n-1} \mathbb{P}(B_k) \\
&= 1 + \sum_{k=1}^{\infty} \mathbb{P}(B_k),
\end{aligned}$$

using Abel summation by parts, $\mathbb{P}(B_0) = \mathbb{P}(X \geq 0) = 1$ and $n \, \mathbb{P}(B_n) \geq 0$. □

Remark. We are using that $\lim_n \sum_{k=1}^{n} \mathbb{P}(B_k)$ exists in $[0, \infty]$: the terms

[7] Note that $B_{n-1} = A_n \uplus B_n$.

are nonnegative, so the partial sums increase (the limit may be infinite). Without that information one has to work with lim inf or lim sup.

Proof. As for the first inequality, arguing as above but bounding $X \geq n-1$ on A_n gives

$$\mathbb{E}[X] \geq \lim_m \left[\sum_{n=1}^{m} \mathbb{P}(B_n) - m\,\mathbb{P}(B_m) \right].$$

If $\mathbb{E}[X] = \infty$ there is nothing to prove. If $\mathbb{E}[X] < \infty$, then, since $B_m \downarrow \emptyset$, the measure ν given by $d\nu = X\,d\mathbb{P}$ is finite[8] and, by continuity, $\nu(B_m) \to \nu(\emptyset) = 0$. Now

$$\begin{aligned} \nu(B_m) = \int_{B_m} X\,d\mathbb{P} \geq \int_{B_m} m\,d\mathbb{P} = m\,\mathbb{P}(B_m) \geq 0 \\ \implies \lim_m m\,\mathbb{P}(B_m) = 0. \end{aligned}$$

Hence

$$\mathbb{E}[X] \geq \sum_{n=1}^{\infty} \mathbb{P}(B_n).$$

□

Corollary 13.2. If the random variable X takes values in $\{0, 1, 2, \dots\}$, then

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} \mathbb{P}(X \geq n).$$

Corollary 13.3. Let $A_1, A_2, \dots$ be events. The following are equivalent:

1. $A_1, A_2, \dots$ are independent;
2. $\sigma(A_1), \sigma(A_2), \dots$ are independent σ-algebras;
3. $\mathbf{1}_{A_1}, \mathbf{1}_{A_2}, \dots$ are independent random variables.

Remark. One has

$$\sigma(\mathbf{1}_{A_i}) = \{\emptyset, \Omega, A_i, A_i^c\} = \sigma(A_i).$$

[8] Because $\nu(\Omega) = \int_\Omega X\,d\mathbb{P} < \infty$.

Remark. In Corollary 13.3, the family of events need not be countable.

Example 13.4. If $A_1, A_2, \dots$ are independent events, then

1. $B_1, B_2, \dots$ are independent, where for each $k \geq 1$ we choose either $B_k = A_k$ or $B_k = A_k^c$;
2. $\bigcup_{k=1}^{\infty} A_{2k}$, $A_1^c \cap A_3$ and $(A_5 \Delta A_9)^c \cup A_{15}$ are independent;
3. $\sigma\left(\bigcup_{k=1}^{\infty} \sigma(A_{2k})\right)$, $\sigma(\sigma(A_1) \cup \sigma(A_3))$ and $\sigma(\sigma(A_5) \cup \sigma(A_9) \cup \sigma(A_{15}))$ are independent σ-algebras.

Example 13.5. If (X, Y), (Z, W) and (U, R, S) are independent, then

$$\mathbf{1}_{\{X>Y\}}, \quad Z + W, \quad \mathbf{1}_{\{U^2+R^2=S^2\}}$$

are independent random variables. Indeed,

$$\begin{aligned} \sigma(\mathbf{1}_{\{X>Y\}}) &= \sigma(\{X > Y\}) \subset \sigma((X, Y)) \\ \sigma(Z + W) &\subset \sigma((Z, W)) \\ \sigma(\mathbf{1}_{\{U^2+R^2=S^2\}}) &\subset \sigma((U, R, S)). \end{aligned}$$

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $X_1, X_2, \dots, X_n$ random variables. Let $X = (X_1, X_2, \dots, X_n) : \Omega \to \mathbb{R}^n$ be the associated random vector.

- P_X is a probability measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$.
- $P_{X_1}, \dots, P_{X_n}$ are probability measures on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, and $P_{X_1} \times \cdots \times P_{X_n}$ is a probability measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}) \otimes \cdots \otimes \mathscr{B}(\mathbb{R})) = (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$.

Theorem 13.6. The following are equivalent:

1. $X_1, \dots, X_n$ are independent random variables;
2. $P_{(X_1,\dots,X_n)} = P_{X_1} \times \cdots \times P_{X_n}$.

Remark. Given a probability space $(\Omega, \mathscr{F}, \mathbb{P})$:

- let $\mathscr{F}_1, \dots, \mathscr{F}_n$ be σ-algebras on Ω contained in $\mathscr{F}$. Then $\mathscr{F}_1, \dots, \mathscr{F}_n$ are independent if and only if $\mathbb{P}\left(\bigcap_{k=1}^n A_k\right) = \prod_{k=1}^n \mathbb{P}(A_k)$ for all $A_1 \in \mathscr{F}_1, \dots, A_n \in \mathscr{F}_n$;

- $X_1, \dots, X_n$ are independent random variables if and only if

$$\mathbb{P}\left(\{X_1 \in A_1\} \cap \cdots \cap \{X_n \in A_n\}\right) = \prod_{k=1}^n \mathbb{P}(X_k \in A_k), \qquad \forall\, A_i \in \mathscr{B}(\mathbb{R}).$$

Lemma 13.7. Let $(\Omega, \mathscr{F})$ be a measurable space and P, Q probability measures on it. If

- $\mathscr{C}$ is a π-system on Ω,
- $\sigma(\mathscr{C}) = \mathscr{F}$,
- $P(A) = Q(A)$ for every $A \in \mathscr{C}$,

then $P = Q$ on $(\Omega, \mathscr{F})$.

Proof. We prove Theorem 13.6.
(2) $\implies$ (1). Let $A_1, \dots, A_n \in \mathscr{B}(\mathbb{R})$ and $B_k = \{X_k \in A_k\} \in \mathscr{F}$ for $k = 1, \dots, n$. Then

$$\begin{aligned} P_{(X_1,\dots,X_n)}(A_1 \times \cdots \times A_n) &= \mathbb{P}(\{\omega \in \Omega : (X_1, \dots, X_n)(\omega) \in A_1 \times \cdots \times A_n\}) \\ &= \mathbb{P}(\{\omega \in \Omega : X_1(\omega) \in A_1, \dots, X_n(\omega) \in A_n\}) \\ &= \mathbb{P}(X_1 \in A_1, \dots, X_n \in A_n). \end{aligned}$$

On the other hand, by hypothesis,

$$\begin{aligned} P_{(X_1,\dots,X_n)}(A_1 \times \cdots \times A_n) &= (P_{X_1} \times \cdots \times P_{X_n})(A_1 \times \cdots \times A_n) \\ &= \prod_{k=1}^n P_{X_k}(A_k) \\ &= \prod_{k=1}^n \mathbb{P}(X_k \in A_k). \end{aligned}$$

Hence

$$\mathbb{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \prod_{k=1}^{n} \mathbb{P}(X_k \in A_k),$$

that is, the random variables are independent.

(1) $\implies$ (2). We must show that if

$$P_{(X_1,\ldots,X_n)}(A_1 \times \cdots \times A_n) = (P_{X_1} \times \cdots \times P_{X_n})(A_1 \times \cdots \times A_n), \quad \forall\, A_i \in \mathscr{B}(\mathbb{R}),$$

then the measures $P_{(X_1,\ldots,X_n)}$ and $P_{X_1} \times \cdots \times P_{X_n}$ coincide. For this we use Lemma 13.7: since

$$\mathscr{C} = \{A_1 \times \cdots \times A_n : \ A_1, \ldots, A_n \in \mathscr{B}(\mathbb{R})\}$$

is a π-system generating $\mathscr{B}(\mathbb{R}^n)$, the conclusion follows. $\square$

14 Lecture 14

Proposition 14.1. If $X_1, \dots, X_n$ are independent and integrable random variables, then

$$\mathbb{E}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} \mathbb{E}[X_i].$$

Proof. Consider $X : (\Omega, \mathscr{F}) \to (\mathbb{R}^n, \mathscr{B}_{\mathbb{R}^n})$ given by $X = (X_1, \dots, X_n)$, and the functions $g, \tilde{g} : \mathbb{R}^n \to \mathbb{R}$ defined by $g(x_1, \dots, x_n) = x_1 \cdots x_n$ and $\tilde{g}(x_1, \dots, x_n) = |x_1 \cdots x_n|$. Then $X_1 \cdots X_n = g \circ X$ and

$$\begin{aligned}
\mathbb{E}[|X_1 \cdots X_n|] &= \int_\Omega \tilde{g}(X)\, d\mathbb{P} \\
&= \int_{\mathbb{R}^n} \tilde{g}(x)\, d\mathbb{P}_X(x) \\
&= \int_{\mathbb{R}^n} |x_1 \cdots x_n|\, d(\mathbb{P}_{X_1} \times \cdots \times \mathbb{P}_{X_n})(x) \\
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \cdots \left[\int_{\mathbb{R}} |x_1| \cdots |x_n|\, d\mathbb{P}_{X_n}(x_n)\right] \cdots \right] d\mathbb{P}_{X_1}(x_1) \\
&= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \cdots |x_1| \cdots |x_{n-1}| \left[\int_{\mathbb{R}} |x_n|\, d\mathbb{P}_{X_n}(x_n)\right] \cdots \right] d\mathbb{P}_{X_1}(x_1) \\
&= \mathbb{E}[|X_n|] \cdots \mathbb{E}[|X_2|]\, \mathbb{E}[|X_1|] < \infty,
\end{aligned}$$

using Theorem 13.6 and Tonelli's theorem. Hence $X_1 \cdots X_n$ is integrable, and

$$\mathbb{E}[X_1 \cdots X_n] = \int_\Omega g(X)\, d\mathbb{P} = \int_{\mathbb{R}^n} g(x)\, d\mathbb{P}_X(x) = \cdots = \mathbb{E}[X_n] \cdots \mathbb{E}[X_1],$$

repeating the computation above with g in place of $\tilde{g}$, now justified by Fubini's theorem. □

Proposition 14.2. Suppose X, Y are independent random variables and $E \in \mathscr{B}(\mathbb{R}^2)$. Then

$$\mathbb{P}((X,Y) \in E) = \int_{\mathbb{R}} \mathbb{P}_X(E^y)\, d\mathbb{P}_Y(y) = \int_{\mathbb{R}} \mathbb{P}_Y(E^x)\, d\mathbb{P}_X(x),$$

where $E^y = \{x \in \mathbb{R} : (x,y) \in E\}$ and $E^x = \{y \in \mathbb{R} : (x,y) \in E\}$.

Proof. On the one hand,

$$\begin{aligned}\mathbb{P}((X,Y)\in E) &= \int_{\mathbb{R}^2} \mathbf{1}_E(x,y)\, d\mathbb{P}_{(X,Y)}(x,y)\\ &= \int_{\mathbb{R}^2} \mathbf{1}_E(x,y)\, d(\mathbb{P}_X\times\mathbb{P}_Y)(x,y)\\ &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \mathbf{1}_E(x,y)\, d\mathbb{P}_X(x)\right] d\mathbb{P}_Y(y)\\ &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} \mathbf{1}_{E^y}(x)\, d\mathbb{P}_X(x)\right] d\mathbb{P}_Y(y),\end{aligned}$$

where $E^y = \{x\in\mathbb{R} : (x,y)\in E\}$. Hence

$$\int_{\mathbb{R}} \left[\int_{\mathbb{R}} \mathbf{1}_{E^y}(x)\, d\mathbb{P}_X(x)\right] d\mathbb{P}_Y(y) = \int_{\mathbb{R}} \mathbb{P}_X(E^y)\, d\mathbb{P}_Y(y).$$

Analogously,

$$\mathbb{P}((X,Y)\in E) = \int_{\mathbb{R}} \mathbb{P}_Y(E^x)\, d\mathbb{P}_X(x)$$

with $E^x = \{y\in\mathbb{R} : (x,y)\in E\}$. ☐

Example 14.3. The case of the diagonal:

$$\mathbb{P}(X=Y) = \mathbb{P}((X,Y)\in D) = \int_{\mathbb{R}} \mathbb{P}_X(D^y)\, d\mathbb{P}_Y(y) = \int_{\mathbb{R}} \mathbb{P}(X=y)\, d\mathbb{P}_Y(y).$$

Example 14.4. We compute

$$\mathbb{P}(X+Y\le a) = \mathbb{P}((X,Y)\in E) = \int_{\mathbb{R}} \mathbb{P}_X(E^y)\, d\mathbb{P}_Y(y)$$

with $E^y = (-\infty, a-y]$. That is,

$$\int_{\mathbb{R}} \mathbb{P}_X((-\infty,a-y])\, d\mathbb{P}_Y(y) = \int_{\mathbb{R}} \mathbb{P}(X\le a-y)\, d\mathbb{P}_Y(y),$$

which is the convolution formula.

14.1 Distribution functions

A distribution function is a function $F:\mathbb{R}\to\mathbb{R}$ such that

1. $0\le F(x)\le 1$ for all $x\in\mathbb{R}$;

2. F is non-decreasing;

3. F is right-continuous;

4. $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

Remark. If F is a distribution function, then there exist a probability space and a random variable X on it with $F_X = F$.

Remark. We say that F is discrete when there exist a countable set $A \subset \mathbb{R}$ and numbers $p_x \geq 0$, $x \in A$, such that

$$F(t) = \sum_{x \leq t} p_x.$$

F is absolutely continuous when there exists a nonnegative Borel measurable $g : \mathbb{R} \to \mathbb{R}$ with

$$F(t) = \int_{-\infty}^{t} g(u)\, du.$$

F is singular when F is continuous and F' exists and equals 0 almost everywhere.

Example 14.5. The Cantor function. One constructs $F : [0,1] \to [0,1]$ as the uniform limit of the sequence F_n defined by: $F_0(x) = x$ and, at each step, on every interval where F_n is affine, one subdivides the interval into three equal parts and redefines F_{n+1} to be constant (equal to the mean value) on the middle third, keeping it affine on the two outer thirds.

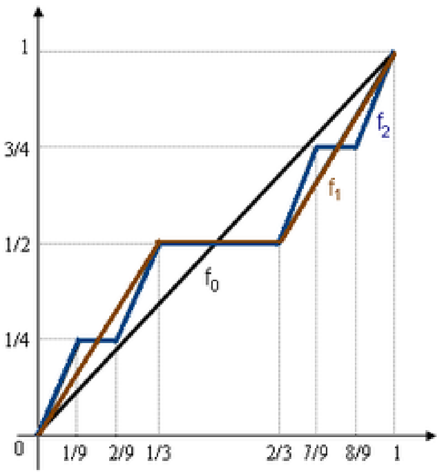

Figure 3: The Cantor function.

One checks that $(F_n(x))_n$ is Cauchy, with $|F_n(x) - F_m(x)| \leq 2^{-n}$ for all $m > n \geq 1$ and all $x \in [0,1]$. Hence the limit F exists and the convergence is uniform, since $|F_n(x) - F(x)| \leq 2^{-n}$. In particular, F is continuous and, since $F' = 0$ outside the Cantor set (which has Lebesgue measure zero), F is singular.

Theorem 14.6. If F is a distribution function, then there exist weights $\alpha_1, \alpha_2, \alpha_3 \geq 0$ with $\alpha_1 + \alpha_2 + \alpha_3 = 1$ and distribution functions F_1 discrete, F_2 absolutely continuous and F_3 singular such that

$$F = \alpha_1 F_1 + \alpha_2 F_2 + \alpha_3 F_3.$$

The decomposition is unique.

14.2 Distributions and distribution functions on $(\mathbb{R}, \mathscr{B}_\mathbb{R})$

A distribution on $(\mathbb{R}, \mathscr{B}_\mathbb{R})$ is a probability measure on $(\mathbb{R}, \mathscr{B}_\mathbb{R})$. If μ is a distribution on $(\mathbb{R}, \mathscr{B}_\mathbb{R})$, then the function

$$\begin{aligned} F_\mu : \mathbb{R} &\to \mathbb{R} \\ t &\mapsto \mu((-\infty, t]) \end{aligned}$$

is a distribution function.

Remark. If F is a distribution function, then there exists a unique distribution μ_F on $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ such that

$$\mu_F((-\infty, t]) = F(t), \qquad \forall\, t \in \mathbb{R}.$$

In particular, the correspondences

$$\begin{aligned} \{\text{Distributions on } (\mathbb{R}, \mathscr{B}_{\mathbb{R}})\} &\longleftrightarrow \{\text{Distribution functions on } \mathbb{R}\} \\ \mu &\longmapsto F_\mu \\ F &\longmapsto \mu_F \end{aligned}$$

are inverse to each other.

14.3 Variance of a random variable

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ a random variable with $\mathbb{E}[X^2] < \infty$. The variance of X is defined by

$$\operatorname{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Proposition 14.7. Let X be a random variable with $\mathbb{E}[X^2] < \infty$. Then

a) $\operatorname{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$;

b) $\operatorname{Var}(X) \geq 0$; in particular, $\mathbb{E}[X^2] \geq \mathbb{E}[X]^2$;

c) if $\operatorname{Var}(X) = 0$, then $X = \mathbb{E}[X]$ a.s.;

d) $\operatorname{Var}(cX) = c^2 \operatorname{Var}(X)$ for every $c \in \mathbb{R}$, and $\operatorname{Var}(X + c) = \operatorname{Var}(X)$;

e) if $X_1, \ldots, X_n$ are independent with $\mathbb{E}[X_j^2] < \infty$ for $j = 1, \ldots, n$, then

$$\operatorname{Var}(X_1 + \cdots + X_n) = \sum_{j=1}^{n} \operatorname{Var}(X_j).$$

15 Lecture 15

Proposition 15.1. Chebyshev's inequality. Let X be a random variable with $\mathbb{E}[X^2] < \infty$ and let $t > 0$. Then

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}.$$

Proof. Write $Y = X - \mathbb{E}[X]$. Then

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[Y^2] \\
&= \int_\Omega Y^2 \, d\mathbb{P} \\
&= \int_{\{|Y| \geq t\}} Y^2 \, d\mathbb{P} + \int_{\{|Y| < t\}} Y^2 \, d\mathbb{P} \\
&\geq \int_{\{|Y| \geq t\}} t^2 \, d\mathbb{P} + \int_{\{|Y| < t\}} Y^2 \, d\mathbb{P} \\
&\geq \int_{\{|Y| \geq t\}} t^2 \, d\mathbb{P} \\
&= t^2 \mathbb{P}(\{|Y| \geq t\}).
\end{aligned}$$

□

Proposition 15.2. Markov's inequality. Let X be a random variable and $t > 0$. Then

$$\mathbb{P}(\{|X| \geq t\}) \leq \frac{\mathbb{E}[|X|]}{t}.$$

Proof. We have

$$\begin{aligned}
\mathbb{E}[|X|] &= \int_\Omega |X| \, d\mathbb{P} \\
&= \int_{\{|X| \geq t\}} |X| \, d\mathbb{P} + \int_{\{|X| < t\}} |X| \, d\mathbb{P} \\
&\geq \int_{\{|X| \geq t\}} t \, d\mathbb{P} = t \, \mathbb{P}\{|X| \geq t\}.
\end{aligned}$$

□

Theorem 15.3. Let Z be a nonnegative random variable, $\varphi : [0,\infty) \to [0,\infty)$ non-decreasing, and $t > 0$ with $\varphi(t) > 0$. Then

$$\mathbb{P}(Z \geq t) \leq \frac{\mathbb{E}[\varphi(Z)]}{\varphi(t)}.$$

Proof. We have

$$\begin{aligned}
\mathbb{E}[\varphi(Z)] &= \int_{\{Z\geq t\}} \varphi(Z)\, d\mathbb{P} + \int_{\{Z<t\}} \varphi(Z)\, d\mathbb{P} \\
&\geq \int_{\{Z\geq t\}} \varphi(t)\, d\mathbb{P} \\
&= \varphi(t)\, \mathbb{P}(Z \geq t).
\end{aligned}$$

□

Definition 15.4. Let $\varphi : I \to \mathbb{R}$. We say that φ is convex when

$$a, b \in I,\ a < b,\ t \in [0,1] \implies \varphi((1-t)a + tb) \leq (1-t)\varphi(a) + t\varphi(b).$$

Definition 15.5. Let $\varphi : I \to \mathbb{R}$. We say that φ is strictly convex when

$$a, b \in I,\ a < b,\ t \in (0,1) \implies \varphi((1-t)a + tb) < (1-t)\varphi(a) + t\varphi(b).$$

Proposition 15.6. Finite Jensen inequality. If $\varphi : I \to \mathbb{R}$ is convex, $a_1, \ldots, a_n \in I$ and $t_1, \ldots, t_n \in (0,1)$ with $\sum_i t_i = 1$, then

$$\varphi\left(\sum_{i=1}^{n} t_i a_i\right) \leq \sum_{i=1}^{n} t_i \varphi(a_i).$$

Proposition 15.7. If $\varphi : I \to \mathbb{R}$ is of class C^2 and convex, then $\varphi''(x) \geq 0$ for every $x \in I$.

Proposition 15.8. If $\varphi : I \to \mathbb{R}$ is of class C^2 and $\varphi''(x) > 0$, then φ is strictly convex.

Proposition 15.9. Let $\varphi : I \to \mathbb{R}$ be convex and let $a \in I$ be an interior point. Then the function $\psi : I \setminus \{a\} \to \mathbb{R}$,

$$\psi(t) = \frac{\varphi(t) - \varphi(a)}{t - a},$$

is non-decreasing.

Proposition 15.10. The one-sided derivatives $D^+\varphi(a) = \lim_{t\to a^+} \psi(t)$ and $D^-\varphi(a) = \lim_{t\to a^-} \psi(t)$ exist, and $D^-\varphi(a) \le D^+\varphi(a)$.

Proposition 15.11. If $\theta \in [D^-\varphi(a), D^+\varphi(a)]$, then

$$\varphi(t) \ge \varphi(a) + \theta(t-a), \qquad \forall\ t \in I.$$

Theorem 15.12. Jensen's inequality. Let X be a random variable and $\varphi : \mathbb{R} \to \mathbb{R}$ a convex function such that X and $\varphi(X)$ are integrable. Then

$$\mathbb{E}[\varphi(X)] \ge \varphi(\mathbb{E}[X]).$$

Proof. Take $a = \mathbb{E}[X]$. There exists $\theta \in \mathbb{R}$ such that

$$\varphi(t) \ge \varphi(a) + \theta(t-a), \qquad \forall\ t \in \mathbb{R},$$

hence

$$\varphi(X(\omega)) \ge \varphi(a) + \theta(X(\omega) - a), \qquad \forall\ \omega \in \Omega.$$

Taking expectations,

$$\mathbb{E}[\varphi(X)] \ge \mathbb{E}[\varphi(a) + \theta(X-a)],$$

and by linearity,

$$\mathbb{E}[\varphi(X)] \ge \underbrace{\mathbb{E}[\varphi(a)]}_{=\varphi(a)} + \theta \underbrace{\mathbb{E}[X-a]}_{=0} = \varphi(a) = \varphi(\mathbb{E}[X]).$$

$\square$

16 Lecture 16

16.1 Characteristic and moment generating functions

Let X be a random variable. The moment generating function of X is defined by

$$\psi_X(t) = \mathbb{E}[e^{tX}], \qquad \forall\, t \in \mathbb{R},$$

and the characteristic function of X by

$$\varphi_X(t) = \mathbb{E}[e^{itX}] = \mathbb{E}[\cos(tX) + i\sin(tX)].$$

Theorem 16.1. If $\mathbb{E}[e^{\delta|X|}] < \infty$ for some $\delta > 0$, then X^k is integrable and $\mathbb{E}[X^k] = \psi_X^{(k)}(0)$ for every $k \in \mathbb{Z}_+$.

Theorem 16.2. If $\mathbb{E}[|X|^r] < \infty$ for some $r \in \mathbb{Z}_+$, then φ_X is of class C^r and

$$\varphi_X^{(r)}(t) = \mathbb{E}[(iX)^r e^{itX}], \qquad \forall\, t \in \mathbb{R}.$$

Proof. Suppose first that $r = 1$. Then

$$\begin{aligned}
\frac{\varphi(t+h) - \varphi(t)}{h} &= \frac{\mathbb{E}[e^{i(t+h)X}] - \mathbb{E}[e^{iXt}]}{h} \\
&= \frac{1}{h}\left[\int_{\mathbb{R}} e^{i(t+h)x}\, dP_X(x) - \int_{\mathbb{R}} e^{itx}\, dP_X(x)\right] \\
&= \int_{\mathbb{R}} \frac{e^{i(t+h)x} - e^{itx}}{h}\, dP_X(x) \\
&= \int_{\mathbb{R}} \frac{e^{itx}(e^{ihx} - 1)}{h}\, dP_X(x) \\
&= \int_{\mathbb{R}} ix\, \frac{e^{itx}(e^{ihx} - 1)}{ixh}\, dP_X(x).
\end{aligned}$$

Letting $h \to 0$ and using the DCT,

$$\lim_{h \to 0} \frac{\varphi(t+h) - \varphi(t)}{h} = \int_{\mathbb{R}} ixe^{itx}\, dP_X(x).$$

Indeed, for $h \in (-1, 1)$ and $x \neq 0$,

$$
\begin{aligned}
\left| ixe^{itx} \frac{e^{ihx} - 1}{ihx} \right| &\leq \left| ixe^{itx} \frac{-1 + \cos(hx) + i \sin(hx)}{hx} \right| \\
&\leq |ixe^{itx}| \cdot \left(\underbrace{\left| \frac{1 - \cos(hx)}{hx} \right|}_{\leq \theta_0} + \underbrace{\left| \frac{\sin(hx)}{hx} \right|}_{\leq 1} \right) \\
&\leq \theta_1 |x|.
\end{aligned}
$$

The bound $\theta_1 |x|$ does not depend on h and is integrable, since

$$\int_{\mathbb{R}} \theta_1 |x| \, dP_X(x) = \theta_1 \mathbb{E}[|X|] < \infty.$$

Therefore, by the DCT,

$$\lim_{h \to 0} \frac{\varphi(t + h) - \varphi(t)}{h} = \int_{\mathbb{R}} ixe^{itx} \, dP_X(x) = \mathbb{E}[iXe^{itX}].$$

To show that φ is of class C^1 we must check that φ' is continuous:

$$\varphi'(t + h) - \varphi'(t) = \int_{\mathbb{R}} ixe^{itx} \left[e^{ihx} - 1 \right] dP_X(x).$$

Since

$$\left| ixe^{itx} \left[e^{ihx} - 1 \right] \right| \leq 2|x|,$$

which is integrable, the DCT gives

$$\lim_{h \to 0} \left[\varphi'(t + h) - \varphi'(t) \right] = \int_{\mathbb{R}} 0 \, dP_X(x) = 0.$$

Now let $r \in \mathbb{Z}_+$. First, $\mathbb{E}[|X|^{r+1}] < \infty$ implies $\mathbb{E}[|X|^r] < \infty$. Indeed, in general, if $0 < \alpha < \beta$, then $\mathbb{E}[|X|^\beta] < \infty \implies \mathbb{E}[|X|^\alpha] < \infty$. It suffices to prove this for $\alpha = 1$, since we may always set $Y = |X|^\alpha$ and consider Y^γ with $\gamma = \beta/\alpha > 1$. Then

$$
\begin{aligned}
\mathbb{E}[Y] &= \int_{\{Y \geq 1\}} Y \, d\mathbb{P} + \int_{\{Y < 1\}} Y \, d\mathbb{P} \\
&\leq \int_{\{Y \geq 1\}} Y^\gamma \, d\mathbb{P} + \int_{\{Y < 1\}} 1 \, d\mathbb{P} \\
&\leq \mathbb{E}[Y^\gamma] + 1 < \infty.
\end{aligned}
$$

An alternative argument uses Jensen's inequality: $t \mapsto t^\gamma$ is convex on $[0, \infty)$ for $\gamma > 1$. Applying it to the truncations $Y_N = \min\{Y, N\}$, which are bounded and hence integrable,

$$\mathbb{E}[Y_N^\gamma] \geq (\mathbb{E}[Y_N])^\gamma.$$

Since $Y_N \uparrow Y$ and $Y_N^\gamma \uparrow Y^\gamma$, letting $N \to \infty$ the MCT gives

$$\mathbb{E}[Y^\gamma] \geq (\mathbb{E}[Y])^\gamma.$$

Now,

$$\frac{\varphi^{(r)}(t+h) - \varphi^{(r)}(t)}{h} = \int_{\mathbb{R}} (ix)^r e^{itx} \left[\frac{e^{ihx} - 1}{h}\right] dP_X(x),$$

and

$$\begin{aligned} (ix)^r e^{itx} \left[\frac{e^{ihx} - 1}{h}\right] &\to (ix)^{r+1} e^{itx}, \\ \left|(ix)^r e^{itx} \left[\frac{e^{ihx} - 1}{h}\right]\right| &\leq \theta_1 |x|^{r+1}, \\ \int_{\mathbb{R}} \theta_1 |x|^{r+1} \, dP_X(x) &= \theta_1 \mathbb{E}[|X|^{r+1}] < \infty. \end{aligned}$$

By the DCT,

$$\begin{aligned} \lim_{h \to 0} \frac{\varphi^{(r)}(t+h) - \varphi^{(r)}(t)}{h} &= \int_{\mathbb{R}} (ix)^{r+1} e^{itx} \, dP_X(x) \\ &= \mathbb{E}[(iX)^{r+1} e^{itX}]. \end{aligned}$$

The continuity of $\varphi^{(r+1)}$ is proved analogously. $\square$

Proposition 16.3. Let φ be the characteristic function of the random variable X. Then

1. $|\varphi(t)| \leq \varphi(0) = 1$ for every $t \in \mathbb{R}$;

2. φ is uniformly continuous;

3. if $a, b \in \mathbb{R}$, then $\varphi_{aX+b}(t) = \varphi(at)e^{itb}$;

4. $\varphi(-t) = \overline{\varphi(t)}$;

5. φ takes real values if and only if X is a symmetric random variable, that is, $P_X(B) = P_X(-B)$ for every $B \in \mathscr{B}(\mathbb{R})$ (equivalently, $X \sim -X$).

Proof. Item by item:

1. $|\varphi(t)| = |\mathbb{E}[e^{itX}]| \leq \mathbb{E}[|e^{itX}|] = 1 = \varphi(0)$ for every $t \in \mathbb{R}$.

2. $|\varphi(t+h) - \varphi(t)| = \left|\int_{\mathbb{R}} e^{itx}(e^{ihx} - 1)\, dP_X(x)\right| \leq \int_{\mathbb{R}} |e^{ihx} - 1|\, dP_X(x)$, and the bound does not depend on t; the DCT shows that it tends to 0 as $h \to 0$.

3. $\varphi_{aX+b}(t) = \mathbb{E}[e^{it(aX+b)}] = e^{itb}\mathbb{E}[e^{itaX}] = e^{itb}\varphi(ta)$.

4. $\varphi(-t) = \mathbb{E}[e^{-itX}] = \mathbb{E}[\overline{e^{itX}}] = \overline{\mathbb{E}[e^{itX}]} = \overline{\varphi(t)}$.

5. If X is symmetric, then $X \sim -X$ and $\varphi_X(t) = \varphi_{-X}(t) = \mathbb{E}[e^{it(-X)}] = \varphi_X(-t) = \overline{\varphi_X(t)}$ for every $t \in \mathbb{R}$, so $\varphi_X(t) \in \mathbb{R}$. Conversely, if φ_X is real-valued, then

$$\varphi_{-X}(t) = \varphi_X(-t) = \overline{\varphi_X(t)} = \varphi_X(t), \qquad \forall\, t \in \mathbb{R},$$

and since the characteristic function determines the law, $-X \sim X$.

$\square$

Proposition 16.4. If $\mathbb{E}[|X|^r] < \infty$ for some $r \in \mathbb{Z}_+$, then, as $t \to 0$,

a) $\varphi(t) = \displaystyle\sum_{j=0}^{r} \frac{i^j \mathbb{E}[X^j]}{j!} t^j + o(|t|^r)$;

b) $\varphi(t) = \displaystyle\sum_{j=0}^{r-1} \frac{i^j \mathbb{E}[X^j]}{j!} t^j + \theta_r \frac{\mathbb{E}[|X|^r]}{r!} t^r$, where $\theta_r = \theta_r(t)$ satisfies $|\theta_r| \leq 2$.

17 Lecture 17

Inversion formula.

Let $F : \mathbb{R} \to \mathbb{R}$ be a distribution function, μ_F the associated distribution and φ its characteristic function. Then

1. $$F(b) - F(a) - \tfrac{1}{2}\mu_F(\{b\}) + \tfrac{1}{2}\mu_F(\{a\}) = \lim_{c\to\infty} \frac{1}{2\pi} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t)\, dt$$

   for all $a < b$.

2. If $\int_{-\infty}^{\infty} |\varphi(t)|\, dt < \infty$, then F has a density given by

   $$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t)\, dt, \qquad \forall\, x \in \mathbb{R}.$$

Proof. We have

$$\int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t)\, dt = \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \int_{\mathbb{R}} e^{itx}\, d\mu_F(x)\, dt.$$

Fubini's theorem applies on $[-c, c] \times \mathbb{R}$ (which has finite product measure, since μ_F is a probability measure) because the integrand is bounded:

$$\frac{e^{-ita} - e^{-itb}}{it} = \int_a^b e^{-its}\, ds \quad \Longrightarrow \quad \left| \frac{e^{-ita} - e^{-itb}}{it} e^{itx} \right| \le b - a.$$

Thus, writing

$$I_c = \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t)\, dt,$$

Fubini's theorem gives

$$\int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} \int_{\mathbb{R}} e^{itx}\, d\mu_F(x)\, dt = \int_{\mathbb{R}} \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} e^{itx}\, dt\, d\mu_F(x).$$

Now,

$$\begin{aligned}
B(c,x) &= \int_{-c}^{c} \frac{e^{-ita} - e^{-itb}}{it} e^{itx}\, dt \\
&= \int_{-c}^{c} \frac{1}{it} \big[\cos(t(x-a)) + i \sin(t(x-a)) \\
&\qquad\qquad - \cos(t(x-b)) - i \sin(t(x-b)) \big]\, dt \\
&= 2 \int_{0}^{c} \left[\frac{\sin(t(x-a))}{t} - \frac{\sin(t(x-b))}{t} \right] dt.
\end{aligned}$$

Since

$$\int_0^\infty \frac{\sin(t\alpha)}{t}\, dt = \frac{\pi}{2} \operatorname{sgn}(\alpha),$$

we have, on the one hand, $|B(c,x)| \leq K$ for some constant $K > 0$ independent of c and x, and, on the other hand,

$$\lim_{c\to\infty} B(c,x) = \begin{cases} 0, & \text{if } x < a, \\ \pi, & \text{if } x = a, \\ 2\pi, & \text{if } a < x < b, \\ \pi, & \text{if } x = b, \\ 0, & \text{if } x > b. \end{cases}$$

Hence, writing $h(x) = \lim_{c\to\infty} B(c,x)$, the DCT gives

$$\lim_{c\to\infty} I_c = \int_{\mathbb{R}} h(x)\, d\mu_F(x) = \pi\, \mu_F(\{a\}) + 2\pi\, \mu_F((a,b)) + \pi\, \mu_F(\{b\}).$$

Dividing by 2π and noting that $\mu_F((a,b)) = F(b) - F(a) - \mu_F(\{b\})$, the claim follows. $\square$

Let $X : \Omega \to \mathbb{R}^n$ be a random vector. The characteristic function of X is defined by $\varphi_X : \mathbb{R}^n \to \mathbb{C}$, $t \mapsto \mathbb{E}[e^{i\langle t, X\rangle}]$.

Proposition 17.1. Let φ be the characteristic function of the random vector $X : \Omega \to \mathbb{R}^n$. Then

1. $|\varphi(t)| \leq \varphi(0) = 1$;

2. φ is uniformly continuous;

3. if $a \in \mathbb{R}$ and $b \in \mathbb{R}^n$, then

$$\varphi_{aX+b}(t) = \varphi(at)e^{i\langle t,b\rangle}, \qquad \forall\, t \in \mathbb{R}^n;$$

4. if $X = (X_1, \ldots, X_n)$, the random variables $X_1, \ldots, X_n$ are independent if and only if $\varphi(t) = \prod_{k=1}^n \varphi_{X_k}(t_k)$ for every $t = (t_1, \ldots, t_n) \in \mathbb{R}^n$;

5. if $Y : \tilde{\Omega} \to \mathbb{R}^n$ is a random vector with $\varphi_Y = \varphi$, then $Y \sim X$.

Proposition 17.2. Weak law of large numbers. Let $X_1, X_2, \ldots$ be i.i.d. integrable random variables with

$$\mathbb{E}[X_1] = m.$$

Then, with $S_n = \sum_{k=1}^n X_k$,

$$\frac{S_n}{n} \longrightarrow m$$

in distribution and hence, the limit being constant, also in probability.

Proof. We compute

$$\begin{aligned}
\varphi_{S_n/n}(t) &= \mathbb{E}[e^{it\frac{S_n}{n}}] \\
&= \mathbb{E}\left[e^{\sum_{k=1}^n \frac{itX_k}{n}}\right] \\
&= \mathbb{E}\left[\prod_{k=1}^n e^{\frac{itX_k}{n}}\right] \\
&= \prod_{k=1}^n \mathbb{E}\left[e^{\frac{itX_k}{n}}\right] \\
&= \left(\mathbb{E}\left[e^{\frac{itX_1}{n}}\right]\right)^n \\
&= \varphi_{X_1}\left(\frac{t}{n}\right)^n.
\end{aligned}$$

Using the first-order expansion of the characteristic function, valid because $\mathbb{E}[|X_1|] < \infty$,

$$\varphi_{X_1}\left(\frac{t}{n}\right)^n = \left[1 + im\frac{t}{n} + o\left(\frac{1}{n}\right)\right]^n.$$

Letting $n \to \infty$, for each fixed t,

$$\lim_{n\to\infty} \varphi_{S_n/n}(t) = e^{itm},$$

which is the characteristic function of the constant random variable $Y \equiv m$. We conclude by Lévy's continuity theorem: pointwise convergence of the characteristic functions to a function continuous at 0 implies convergence in distribution. $\square$

Proposition 17.3. Central limit theorem. Let $X_1, X_2, \ldots$ be i.i.d. random variables with $\mathbb{E}[X_1^2] < \infty$, $\mathbb{E}[X_1] = m$ and $\mathrm{Var}(X_1) = \sigma^2 > 0$. Then

$$\frac{S_n - mn}{\sigma\sqrt{n}} \longrightarrow \mathcal{N}(0,1)$$

in distribution.

Proof. Let $Y_k = X_k - \mathbb{E}[X_k] = X_k - m$ and $T_n = \sum_{k=1}^n Y_k = S_n - mn$. Then

$$\begin{aligned}
\varphi_{\frac{T_n}{\sigma\sqrt{n}}}(t) &= \prod_{k=1}^n \mathbb{E}\left[e^{\frac{itY_k}{\sigma\sqrt{n}}}\right] \\
&= \left(\mathbb{E}\left[e^{\frac{itY_1}{\sigma\sqrt{n}}}\right]\right)^n \\
&= \left(\varphi_{Y_1}\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n.
\end{aligned}$$

Since $\mathbb{E}[Y_1] = 0$ and $\mathbb{E}[Y_1^2] = \sigma^2$, the second-order expansion of the characteristic function gives, as $n \to \infty$ with t fixed,

$$\begin{aligned}
\varphi_{Y_1}\left(\frac{t}{\sigma\sqrt{n}}\right) &= 1 + i \cdot 0 \cdot \frac{t}{\sigma\sqrt{n}} + \frac{i^2}{2}\sigma^2\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + o\left(\frac{1}{n}\right) \\
&= 1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right).
\end{aligned}$$

Therefore

$$\varphi_{\frac{T_n}{\sigma\sqrt{n}}}(t) = \left[1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right]^n \longrightarrow e^{-\frac{t^2}{2}} = \varphi_Z(t),$$

where $Z \sim \mathcal{N}(0,1)$. We conclude by Lévy's continuity theorem. $\square$

18 Lecture 18

18.1 L^p spaces

Let $(\Omega, \mathscr{F}, \mu)$ be a measure space and $p > 0$. We define

$$L^p(\Omega, \mathscr{F}, \mu) = \left\{ f : \Omega \to \mathbb{R} \text{ measurable} : \int_\Omega |f|^p \, d\mu < \infty \right\}.$$

Remark. We write $\ell^p(E) = L^p(E, 2^E, \nu)$, where ν is the counting measure.

Theorem 18.1. The space $L^p(\Omega, \mathscr{F}, \mu)$ is a vector space.

Proof. Since $L^p(\Omega, \mathscr{F}, \mu)$ is a subset of the vector space of measurable functions $f : \Omega \to \mathbb{R}$, we only need to check that $f + \lambda g \in L^p$ whenever $f, g \in L^p$ and $\lambda \in \mathbb{R}$. Pointwise,

$$\begin{aligned}
|f + \lambda g|^p &\leq (|f| + |\lambda g|)^p \\
&\leq (2 \max\{|f|, |\lambda g|\})^p \\
&= 2^p \max\{|f|^p, |\lambda g|^p\} \\
&\leq 2^p \left(|f|^p + |\lambda|^p \cdot |g|^p\right),
\end{aligned}$$

whence

$$\int |f + \lambda g|^p \, d\mu \leq 2^p \left[\int |f|^p \, d\mu + |\lambda|^p \int |g|^p \, d\mu \right] < \infty,$$

and therefore $f + \lambda g \in L^p$. □

Lemma 18.2. Young's inequality. Let $p, q > 1$ be conjugate exponents, that is, $\frac{1}{p} + \frac{1}{q} = 1$. If $x, y \geq 0$, then

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}.$$

Proof. If $x = 0$ or $y = 0$ this is immediate. If $x, y > 0$, set $a = \ln x$ and $b = \ln y$. Since $u \mapsto e^u$ is convex and $\frac{1}{p} + \frac{1}{q} = 1$,

$$\frac{x^p}{p} + \frac{y^q}{q} = \frac{1}{p} e^{ap} + \frac{1}{q} e^{bq} \geq e^{\frac{ap}{p} + \frac{bq}{q}} = e^{a+b} = xy.$$

□

Remark. It also holds that, if $p_1, \ldots, p_k > 1$ satisfy $\sum_{i=1}^k \frac{1}{p_i} = 1$ and $x_1, \ldots, x_k \geq 0$, then

$$\prod_{i=1}^k x_i \leq \sum_{i=1}^k \frac{x_i^{p_i}}{p_i}.$$

Lemma 18.3. Hölder's inequality. Let $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$ and let $f, g : \Omega \to \mathbb{R}$ be measurable. Then

$$\int |fg| \, d\mu \leq \left(\int |f|^p \, d\mu \right)^{1/p} \left(\int |g|^q \, d\mu \right)^{1/q}.$$

Moreover, when $f \in L^p$ and $g \in L^q$, *equality* holds if and only if there exist constants $a, b \geq 0$ with $a^2 + b^2 \neq 0$ such that $b|f|^p = a|g|^q$ a.s.

Proof. First, if $\int |f|^p \, d\mu = 0$, then $f = 0$ a.s., so $|fg| = 0$ a.s. and both sides vanish; similarly if $\int |g|^q \, d\mu = 0$. If $\int |f|^p \, d\mu = \infty$ or $\int |g|^q \, d\mu = \infty$ (and the other factor is nonzero), the right-hand side is ∞ and the inequality is trivial.

In the case $\int |f|^p \, d\mu = \int |g|^q \, d\mu = 1$, Young's inequality applied pointwise to $x = |f(\omega)|$, $y = |g(\omega)|$ gives

$$1 = \frac{1}{p} \int |f|^p \, d\mu + \frac{1}{q} \int |g|^q \, d\mu \geq \int |fg| \, d\mu.$$

If $\int |f|^p \, d\mu$ and $\int |g|^q \, d\mu$ both lie in $(0, \infty)$, set

$$\tilde{f} = \frac{f}{\left(\int |f|^p \, d\mu \right)^{1/p}}, \qquad \tilde{g} = \frac{g}{\left(\int |g|^q \, d\mu \right)^{1/q}}.$$

Then $\int |\tilde{f}|^p \, d\mu = \int |\tilde{g}|^q \, d\mu = 1$ and, by the previous case, $\int |\tilde{f}\tilde{g}| \, d\mu \leq 1$, which is exactly the desired inequality. □

Remark. It also holds that, given $p_1, \ldots, p_k > 1$ with $\frac{1}{p_1} + \cdots + \frac{1}{p_k} = 1$ and measurable $f_1, \ldots, f_k : \Omega \to \mathbb{R}$,

$$\prod_{i=1}^k \left(\int |f_i|^{p_i} \, d\mu \right)^{\frac{1}{p_i}} \geq \int |f_1 \cdots f_k| \, d\mu.$$

Theorem 18.4. Minkowski's inequality. Let $p \geq 1$ and $f, g \in L^p$. Then

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p.$$

Proof. For $p = 1$ this is an immediate consequence of the pointwise triangle inequality. For $p > 1$, set $q = \frac{p}{p-1}$ and $\theta = \frac{p}{q} = p - 1$. Since $f + g \in L^p$ (the space is a vector space), $\big\| \, |f+g|^\theta \, \big\|_q = \|f+g\|_p^\theta$. By Hölder's inequality,

$$\|f\|_p \cdot \big\| \, |f+g|^\theta \, \big\|_q \geq \int |f| \cdot |f+g|^\theta \, d\mu,$$
$$\|g\|_p \cdot \big\| \, |f+g|^\theta \, \big\|_q \geq \int |g| \cdot |f+g|^\theta \, d\mu,$$

and, adding,

$$(\|f\|_p + \|g\|_p) \, \|f+g\|_p^\theta \geq \int (|f|+|g|)|f+g|^\theta \, d\mu \geq \int |f+g|^p \, d\mu = \|f+g\|_p^p.$$

If $\|f+g\|_p = 0$ there is nothing to prove; otherwise $\|f+g\|_p^p < \infty$ and we divide by $\|f+g\|_p^\theta = \|f+g\|_p^{p-1}$ to obtain the result. $\square$

Is $\|\cdot\|_p$ a norm on L^p? We have the triangle inequality (Minkowski), homogeneity and the implication $f = 0 \implies \|f\|_p = 0$. However, $\|f\|_p = 0$ does not imply $f = 0$: we can only conclude that $f = 0$ a.s. For this reason $\|\cdot\|_p$ is merely a seminorm on L^p, and a genuine norm is obtained by passing to the quotient by the relation of a.s. equality; this is the convention normally in force when one speaks of "the normed space L^p".

Theorem 18.5. For $1 \leq p < \infty$, the set of integrable simple functions is dense in L^p.

19 Lecture 19

Definition 19.1. Let (X_n) be a sequence of random variables and X a random variable. We say that X_n converges a.s. to X when

$$\mathbb{P}\{\omega \in \Omega : \lim_n X_n(\omega) = X(\omega)\} = 1.$$

We write $X_n \to X$ a.s.

Definition 19.2. We say that X_n converges to X in probability when

$$\lim_n \mathbb{P}(|X_n - X| \geq \varepsilon) = 0 \qquad \forall\, \varepsilon > 0,$$

equivalently,

$$\lim_n \mathbb{P}(|X_n - X| < \varepsilon) = 1 \qquad \forall\, \varepsilon > 0.$$

We write $X_n \to X$ in probability, or $X_n \xrightarrow{\mathbb{P}} X$.

Definition 19.3. We say that X_n converges to X in L^p, with $p > 0$, when

1. $X, X_1, X_2, \dots \in L^p$;
2. $\lim_n \mathbb{E}(|X_n - X|^p) = 0$.

We write $X_n \to X$ in L^p.

Definition 19.4. We say that X_n converges to X in distribution (or in law) when one (and hence all) of the following conditions hold:

1. $\lim_n F_{X_n}(t) = F_X(t)$ for every t that is a continuity point of F_X;
2. $\lim_n \varphi_{X_n}(t) = \varphi_X(t)$ for every $t \in \mathbb{R}$;
3. $\lim_n \mathbb{E}[g(X_n)] = \mathbb{E}[g(X)]$ for every continuous bounded $g : \mathbb{R} \to \mathbb{R}$.

Theorem 19.5. Let (X_n) be a sequence of random variables and X a random variable. Then:

1. if $X_n \to X$ a.s., then $X_n \to X$ in probability;

2. if $X_n \to X$ in L^p, then $X_n \to X$ in probability;

3. if $X_n \to X$ in probability, then $X_n \to X$ in distribution.

Proof. For (1), fix $\varepsilon > 0$. For each $n \in \mathbb{Z}_+$, set $A_n = \{|X_n - X| < \varepsilon\} \in \mathscr{F}$ and $B_n = \bigcap_{k \geq n} A_k \subset A_n$. Then $B_N \uparrow \bigcup_{N=1}^{\infty} B_N$. On the other hand,

$$\mathbb{P}\underbrace{\{\lim_n X_n = X\}}_{=C} = 1.$$

If $\omega \in C$, there is $N \in \mathbb{Z}_+$ such that $|X_n(\omega) - X(\omega)| < \varepsilon$ for all $n \geq N$; that is, there is N with $\omega \in B_N$, hence $\omega \in \bigcup_{N=1}^{\infty} B_N$. Therefore $C \subset \bigcup_{N=1}^{\infty} B_N$ and

$$1 = \mathbb{P}(C) \leq \mathbb{P}\left(\bigcup_{N=1}^{\infty} B_N\right) = \lim_N \mathbb{P}(B_N) \leq \liminf_N \mathbb{P}(A_N) \leq 1,$$

so $\lim_N \mathbb{P}(A_N) = 1$, that is, $\lim_N \mathbb{P}(|X_N - X| \geq \varepsilon) = 0$.
For (2), fix $\varepsilon > 0$. By Markov's inequality applied to $|X_n - X|^p$,

$$0 \leq \mathbb{P}(|X_n - X| \geq \varepsilon) \leq \frac{\mathbb{E}[|X_n - X|^p]}{\varepsilon^p}.$$

Letting $n \to \infty$ gives $\lim_n \mathbb{P}\{|X_n - X| \geq \varepsilon\} = 0$.
Finally, for (3), let t be a continuity point of $F = F_X$ and let $\varepsilon > 0$. We know that

$$\lim_n \mathbb{P}\{|X_n - X| \geq \varepsilon\} = 0.$$

Then

$$
\begin{aligned}
\mathbb{P}\{X_n \le t\} &= \mathbb{P}\{X_n \le t, |X_n - X| \ge \varepsilon\} \\
&\quad + \mathbb{P}\{X_n \le t, |X_n - X| < \varepsilon\} \\
&\le \mathbb{P}\{|X_n - X| \ge \varepsilon\} + \mathbb{P}\{X < t + \varepsilon\} \\
\limsup_n \mathbb{P}\{X_n \le t\} &\le \mathbb{P}\{X < t + \varepsilon\} = F((t+\varepsilon)^-) \\
\limsup_n \mathbb{P}\{X_n \le t\} &\le F(t) \quad (\text{letting } \varepsilon \downarrow 0, \text{ by continuity at } t) \\
\mathbb{P}\{X \le t - \varepsilon\} &= \mathbb{P}\{X \le t - \varepsilon, |X_n - X| \ge \varepsilon\} \\
&\quad + \mathbb{P}\{X \le t - \varepsilon, |X_n - X| < \varepsilon\} \\
&\le \mathbb{P}\{|X_n - X| \ge \varepsilon\} + \mathbb{P}\{X_n \le t\} \\
\mathbb{P}\{X \le t - \varepsilon\} &\le \liminf_n \mathbb{P}\{X_n \le t\} \\
F(t^-) &\le \liminf_n \mathbb{P}\{X_n \le t\} \\
F(t) &\le \liminf_n \mathbb{P}\{X_n \le t\} \\
&\le \limsup_n \mathbb{P}\{X_n \le t\} \le F(t).
\end{aligned}
$$

□

Example 19.6. On the space $([0,1], \mathscr{B}([0,1]), m)$, let $X_n = n^{1/p}\mathbf{1}_{(0,1/n)}$. Then $X_n \to 0$ a.s., but

$$\mathbb{E}[|X_n - 0|^p] = (n^{1/p})^p \cdot \frac{1}{n} = 1.$$

Hence $X_n \not\to 0$ in L^p.

Example 19.7. On $([0,1], \mathscr{B}([0,1]), m)$, consider the typewriter sequence: $X_1 = 1$, $X_2 = \mathbf{1}_{[0,1/2)}$, $X_3 = \mathbf{1}_{[1/2,1]}$, $X_4 = \mathbf{1}_{[0,1/3)}$, $X_5 = \mathbf{1}_{[1/3,2/3)}$, $X_6 = \mathbf{1}_{[2/3,1]}$, ... Since

$$|X_n| \le 1 \implies X_n \in L^p,$$

we have $X_n \to 0$ in L^p but $X_n \not\to 0$ a.s. Indeed,

$$\{x \in [0,1] : \lim_n X_n(x) = 0\} = \emptyset.$$

Theorem 19.8. If $X_n \to X$ in probability, then there is a subsequence $(X_{n_k})_k$ of (X_n) with $X_{n_k} \to X$ a.s.

Corollary 19.9. Let $X_1, X_2, \dots$ be random variables. The following are equivalent:

1. $X_n \to X$ in probability;
2. every subsequence of (X_n) has a further subsequence converging to X a.s.

Proof. (1) $\implies$ (2). If $X_n \to X$ in probability, every subsequence $(X_{n_k})_k$ also converges to X in probability; by the previous theorem, $(X_{n_k})_k$ has in turn a subsequence converging to X a.s.
(2) $\implies$ (1). Suppose $X_n \not\to X$ in probability. Then there exist $\varepsilon > 0$, $\theta > 0$ and a subsequence $(X_{n_k})_k$ such that

$$\mathbb{P}\{|X_{n_k} - X| \geq \varepsilon\} \geq \theta, \qquad \forall\, k \geq 1.$$

By hypothesis, $(X_{n_k})_k$ has a subsequence $(X_{n_{k_j}})_j$ converging to X a.s., hence in probability, so that $\mathbb{P}\{|X_{n_{k_j}} - X| \geq \varepsilon\} \to 0$. This contradicts the previous inequality. $\square$

20 Lecture 20

We return to the theorem stated at the end of the previous lecture, now with its proof.

Theorem 20.1. If $X_n \to X$ in probability, then there is a subsequence $(X_{n_k})_k$ of $(X_n)_n$ with $X_{n_k} \to X$ a.s.

Proof. For each $k \in \mathbb{Z}_+$ we have $\lim_n \mathbb{P}\left\{|X_n - X| \geq \frac{1}{k}\right\} = 0$. We may therefore choose $n_k \in \mathbb{Z}_+$ with

$$\mathbb{P}\left\{|X_{n_k} - X| \geq \frac{1}{k}\right\} < \frac{1}{2^k},$$

and we may take the n_k so that $n_1 < n_2 < \cdots$. Then

$$\sum_{k=1}^{\infty} \mathbb{P}\left\{|X_{n_k} - X| \geq \frac{1}{k}\right\} \leq 1 < \infty.$$

By the first Borel–Cantelli lemma,

$$\mathbb{P}\underbrace{\left\{|X_{n_k} - X| \geq \tfrac{1}{k} \text{ for infinitely many } k\right\}}_{=A} = 0.$$

If $\omega \in A^c$, there is $k_0 = k_0(\omega)$ such that

$$0 \leq |X_{n_k}(\omega) - X(\omega)| < \frac{1}{k}, \qquad k \geq k_0,$$

and hence

$$\lim_{k\to\infty} X_{n_k}(\omega) = X(\omega).$$

Therefore $X_{n_k} \to X$ a.s. $\square$

Corollary 20.2. $X_n \to X$ in probability if and only if every subsequence of (X_n) has a further subsequence converging to X a.s.

Example 20.3. One has:

1. if $X_n \to X$ and $Y_n \to Y$ a.s., then $X_n + Y_n \to X + Y$ a.s. and $X_n Y_n \to XY$ a.s.;

2. if $X_n \to X$ and $Y_n \to Y$ in probability, then $X_n + Y_n \to X + Y$ and $X_n Y_n \to XY$ in probability;

3. item (2) follows from (1) and Corollary 20.2: given a subsequence, extract from it a further subsequence along which $X_n \to X$ a.s.; from that one extract another along which also $Y_n \to Y$ a.s. Along this last subsequence, (1) gives a.s. convergence of the product, and the corollary brings us back to convergence in probability of $X_n Y_n$.

Theorem 20.4. Let C be a constant. If $X_n \to C$ in distribution, then $X_n \to C$ in probability.

Proof. Fix $\varepsilon > 0$. We have

$$\begin{aligned}\mathbb{P}\{X_n \geq C + \varepsilon\} &= 1 - \mathbb{P}\{X_n < C + \varepsilon\} \\ &= 1 - F_n((C+\varepsilon)^-) \\ &\leq 1 - F_n(C + \varepsilon/2).\end{aligned}$$

As $n \to \infty$, $1 - F_n(C + \varepsilon/2) \to 1 - F(C + \varepsilon/2) = 0$, since $C + \varepsilon/2$ is a continuity point of F and

$$F(t) = \begin{cases} 0, & \text{if } t < C, \\ 1, & \text{if } t \geq C. \end{cases}$$

Hence

$$\mathbb{P}\{X_n \geq C + \varepsilon\} \to 0.$$

On the other hand,

$$\mathbb{P}\{X_n \leq C - \varepsilon\} = F_n(C - \varepsilon) \to F(C - \varepsilon) = 0.$$

Therefore

$$\mathbb{P}\{|X_n - C| \geq \varepsilon\} \to 0.$$

□

20.1 The space L^∞

Let $(\Omega, \mathscr{F}, \mu)$ be a measure space. For each function $f : \Omega \to \mathbb{R}$ we define the essential supremum of $|f|$ by

$$\|f\|_\infty = \inf\{a \in \mathbb{R} : \mu(\{|f| > a\}) = 0\}.$$

The space $L^\infty = L^\infty(\Omega, \mathscr{F}, \mu)$ is defined as the set of measurable functions $f : \Omega \to \mathbb{R}$ with $\|f\|_\infty < \infty$.

Theorem 20.5. One has:

1. if $f, g : \Omega \to \mathbb{R}$ are measurable, then $\|f\|_1 \cdot \|g\|_\infty \geq \|fg\|_1$;
2. $\|\cdot\|_\infty$ is a norm on L^∞;
3. if $f, f_1, f_2, \cdots : \Omega \to \mathbb{R}$ are measurable, then $\|f_n - f\|_\infty \to 0$ if and only if there exists $E \in \mathscr{F}$ with $\mu(E^c) = 0$ and $f_n \to f$ uniformly on E;
4. $(L^\infty, \|\cdot\|_\infty)$ is a Banach space;
5. the set of simple functions in L^∞ is dense in L^∞.

Theorem 20.6. If $0 < p < q < r \leq \infty$, then

1. $L^q \subset L^p + L^r$;
2. $L^p \cap L^r \subset L^q$ and $\|f\|_q \leq \|f\|_p^\lambda \|f\|_r^{1-\lambda}$, with $\lambda = \frac{q^{-1} - r^{-1}}{p^{-1} - r^{-1}}$.

Proof. For (1), write $f = f\mathbf{1}_{\{|f| \leq 1\}} + f\mathbf{1}_{\{|f| > 1\}} = g + h$. Then

$$\begin{aligned}
\int |g|^r &= \int |f|^r \mathbf{1}_{\{|f| \leq 1\}} \leq \int |f|^q \mathbf{1}_{\{|f| \leq 1\}} < \infty \\
&\implies g \in L^r, \\
\int |h|^p &= \int |f|^p \mathbf{1}_{\{|f| > 1\}} \\
&\leq \int |f|^q \mathbf{1}_{\{|f| > 1\}} < \infty \\
&\implies h \in L^p.
\end{aligned}$$

For (2), set $\alpha = \frac{r-p}{r-q}$, $\beta = \frac{r-p}{q-p}$, $a = \frac{p(r-q)}{r-p}$, $b = \frac{r(q-p)}{r-p}$. By Hölder's inequality,

$$\| \, |f|^q \, \|_1 \leq \| \, |f|^a \, \|_\alpha \, \| \, |f|^b \, \|_\beta.$$

Therefore

$$\begin{aligned}
\int |f|^q &\leq \left(\int |f|^p \right)^{\frac{r-q}{r-p}} \left(\int |f|^r \right)^{\frac{q-p}{r-p}} \\
&= \|f\|_p^{\frac{p(r-q)}{r-p}} \|f\|_r^{\frac{r(q-p)}{r-p}} \\
\|f\|_q &\leq \|f\|_p^{\frac{p(r-q)}{q(r-p)}} \|f\|_r^{\frac{r(q-p)}{q(r-p)}},
\end{aligned}$$

and one checks that $\lambda = \frac{p(r-q)}{q(r-p)}$. □

Theorem 20.7. If μ is a finite measure and $0 < p < q \leq \infty$, then $L^q \subset L^p$ and $\|f\|_p \leq \|f\|_q \, \mu(\Omega)^{\frac{1}{p} - \frac{1}{q}}$ for every measurable $f : \Omega \to \mathbb{R}$.

Proof. Set $\alpha = \frac{q}{p}$ and $\beta = \frac{q}{q-p}$, so that $\frac{1}{\alpha} + \frac{1}{\beta} = 1$. Then

$$\begin{aligned}
\left\| \, |f|^p \, \right\|_1 &\leq \left\| \, |f|^p \, \right\|_\alpha \left\| \, 1 \, \right\|_\beta \\
\int |f|^p \, d\mu &\leq \left(\int |f|^q \, d\mu \right)^{\frac{p}{q}} \left(\int 1 \, d\mu \right)^{\frac{q-p}{q}} \\
\|f\|_p &\leq \|f\|_q \, \mu(\Omega)^{\frac{q-p}{qp}} = \|f\|_q \, \mu(\Omega)^{\frac{1}{p} - \frac{1}{q}}.
\end{aligned}$$

□

20.2 The dual space of L^p

Let $p, q \in [1, \infty]$ with $\frac{1}{p} + \frac{1}{q} = 1$. For each $g \in L^q$, the map $\varphi_g : L^p \to \mathbb{R}$ defined by

$$\varphi_g(f) = \int_\Omega fg \, d\mu$$

is a linear functional on L^p:

1. $\varphi_g(\alpha f) = \alpha \varphi_g(f)$ for all $\alpha \in \mathbb{R}$ and $f \in L^p$;

2. $\varphi_g(f + h) = \varphi_g(f) + \varphi_g(h)$ for all $f, h \in L^p$.

Remark. Given $\varphi : E \to \mathbb{R}$,

$$\|\varphi\| = \sup_{u\in E,\, \|u\|=1} |\varphi(u)|.$$

Theorem 20.8. If $q \in [1, \infty)$, then

$$\|\varphi_g\| = \|g\|_q.$$

If $q = \infty$ and μ is σ-finite, then also

$$\|\varphi_g\| = \|g\|_\infty.$$

21 Lecture 21

Theorem 21.1. If $p \in (1,\infty)$, the map $g \mapsto \varphi_g$ is a bijective isometry between L^q and $(L^p)^*$, where $\frac{1}{q}+\frac{1}{p}=1$. The same holds for $p=1$ when μ is σ-finite.

Definition 21.2. We define

$$E^* = \{\text{bounded linear functionals on } E\}.$$

For $\varphi \in E^*$ we set $\|\varphi\| = \sup_{\|v\|=1,\, v\in E} |\varphi(v)|$. An element $\psi \in (E^*)^*$ is a bounded linear functional $\psi : E^* \to \mathbb{R}$. Each $x \in E$ defines one:

$$\psi_x : E^* \to \mathbb{R}, \qquad \varphi \mapsto \varphi(x).$$

We say that E is *reflexive* when the canonical map $x \mapsto \psi_x$ is a bijection from E onto $(E^*)^*$.

Corollary 21.3. If $p \in (1,\infty)$, then L^p is reflexive.

Proof. We have

$$\begin{aligned}(L^p)^* &\simeq L^q\\ ((L^p)^*)^* &\simeq (L^q)^* \simeq L^p.\end{aligned}$$

□

21.1 Conditional expectation

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and $B \in \mathscr{F}$ with $\mathbb{P}(B) > 0$. For each event A, the conditional probability of A given B is defined by

$$\mathbb{P}_B(A) = \mathbb{P}(A|B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)}.$$

The function $\mathbb{P}_B : \mathscr{F} \to \mathbb{R}$ is a probability measure on $(\Omega, \mathscr{F})$.

Now, given an integrable random variable X and a sub-σ-algebra $\mathscr{G}$ of $\mathscr{F}$, we look for a random variable Y such that

1. Y is $\mathscr{G}$-measurable;

2. $\int_A Y\, d\mathbb{P} = \int_A X\, d\mathbb{P}$ for every $A \in \mathscr{G}$.

The restriction $\mathbb{P}|_{\mathscr{G}} : \mathscr{G} \to \mathbb{R}$ is a probability measure on $(\Omega, \mathscr{G})$. Let $\nu_1 : \mathscr{G} \to \mathbb{R}$, $A \mapsto \int_A X^+\, d\mathbb{P}$, and $\nu_2 : \mathscr{G} \to \mathbb{R}$, $A \mapsto \int_A X^-\, d\mathbb{P}$; both are finite measures, since X is integrable. If $A \in \mathscr{G}$ satisfies $\mathbb{P}(A) = 0$, then $\nu_1(A) = \nu_2(A) = 0$, so $\nu_1, \nu_2 \ll \mathbb{P}|_{\mathscr{G}}$. By the Radon–Nikodym theorem there exist $\mathscr{G}$-measurable Y_1, Y_2 with $d\nu_1 = Y_1\, d\mathbb{P}|_{\mathscr{G}}$ and $d\nu_2 = Y_2\, d\mathbb{P}|_{\mathscr{G}}$. Set $Y = Y_1 - Y_2$, which is $\mathscr{G}$-measurable. If $A \in \mathscr{G}$,

$$\nu_1(A) = \int_A Y_1\, d\mathbb{P}|_{\mathscr{G}} = \int_A Y_1\, d\mathbb{P},$$

and analogously for ν_2. Subtracting,

$$\int_A X\, d\mathbb{P} = \int_A Y\, d\mathbb{P}, \qquad \forall\ A \in \mathscr{G},$$

which is the required property. As for uniqueness: if Z is $\mathscr{G}$-measurable and satisfies the same, then $\int_A (Y - Z)\, d\mathbb{P} = 0$ for every $A \in \mathscr{G}$; taking $A = \{Y > Z\}$ and $A = \{Y < Z\}$, both in $\mathscr{G}$, gives $Y = Z$ $\mathbb{P}$-a.s.

Theorem 21.4. Let X be an integrable random variable and $\mathscr{G}$ a sub-σ-algebra of $\mathscr{F}$. Then

a) there exists a random variable Y such that

- Y is $\mathscr{G}$-measurable;
- $\int_A Y\, d\mathbb{P} = \int_A X\, d\mathbb{P}$ for every $A \in \mathscr{G}$;

b) if Z is a random variable satisfying those two conditions, then $Z = Y$ a.s.

Y is called the conditional expectation of X given $\mathscr{G}$ and is denoted by $\mathbb{E}[X|\mathscr{G}]$.

Remark. The distribution of X is $P_X(A) = \mathbb{P}\{X \in A\}$. Analogously, the law of X conditioned on the event B is $P_X^B(A) = \mathbb{P}_B\{X \in A\}$.

Example 21.5. If $X = c$ is constant, then $\mathbb{E}[c|\mathscr{G}] = c$.

Example 21.6. If $\mathscr{G} = \{\Omega, \emptyset\}$, then $\mathbb{E}[X|\mathscr{G}] = \mathbb{E}[X]$.

Example 21.7. If $\Omega = \biguplus_{k=1}^{n} B_k$ with $B_1, \dots, B_n \in \mathscr{F}$ and $\mathbb{P}(B_k) > 0$ for every k, and $\mathscr{G} = \sigma(\{B_1, \dots, B_n\})$, then

$$\mathbb{E}[X|\mathscr{G}] = \sum_{1 \le k \le n} \left(\frac{1}{\mathbb{P}(B_k)} \int_{B_k} X \, d\mathbb{P} \right) \mathbf{1}_{B_k},$$

where each summand is the average of X over the atom B_k.

Example 21.8. If $\Omega = \biguplus_{k=1}^{\infty} B_k$ with $B_k \in \mathscr{F}$ and $\mathbb{P}(B_k) > 0$ for every k, and $\mathscr{G} = \sigma(\{B_k : \ k \ge 1\})$, then

$$\mathbb{E}[X|\mathscr{G}] = \sum_{k \ge 1} \left(\frac{1}{\mathbb{P}(B_k)} \int_{B_k} X \, d\mathbb{P} \right) \mathbf{1}_{B_k}.$$

Example 21.9. Let $A \in \mathscr{F}$ and let $\mathscr{G} \subset \mathscr{F}$ be a sub-σ-algebra. The conditional probability of A given $\mathscr{G}$ is defined by

$$\mathbb{P}[A|\mathscr{G}] = \mathbb{E}[\mathbf{1}_A|\mathscr{G}].$$

Proposition 21.10. Let X, Z be integrable random variables and $\mathscr{G}$ a sub-σ-algebra of $\mathscr{F}$. Then

1. if $\alpha \in \mathbb{R}$, then $\mathbb{E}[\alpha X|\mathscr{G}] = \alpha \mathbb{E}[X|\mathscr{G}]$;
2. $\mathbb{E}[X + Z|\mathscr{G}] = \mathbb{E}[X|\mathscr{G}] + \mathbb{E}[Z|\mathscr{G}]$;
3. if $X \le Z$, then $\mathbb{E}[X|\mathscr{G}] \le \mathbb{E}[Z|\mathscr{G}]$;
4. $\mathbb{E}(\mathbb{E}[X|\mathscr{G}]) = \mathbb{E}[X]$;
5. if $\mathscr{H} \subset \mathscr{G}$, then
$$\mathbb{E}[\mathbb{E}[X|\mathscr{G}]|\mathscr{H}] = \mathbb{E}[X|\mathscr{H}].$$

Proof. Write $X_{\mathscr{G}} = \mathbb{E}[X|\mathscr{G}]$ and $Z_{\mathscr{G}} = \mathbb{E}[Z|\mathscr{G}]$. In each case it suffices to check that the proposed candidate is $\mathscr{G}$-measurable and has the same integrals as the left-hand side over every $A \in \mathscr{G}$; a.s. uniqueness does the rest.

1. $\alpha X_{\mathscr{G}}$ is $\mathscr{G}$-measurable and, by linearity of the integral,

$$\int_A \alpha X_{\mathscr{G}}\, d\mathbb{P} = \alpha \int_A X\, d\mathbb{P} = \int_A \alpha X\, d\mathbb{P}, \qquad \forall\ A \in \mathscr{G}.$$

2. $X_{\mathscr{G}} + Z_{\mathscr{G}}$ is $\mathscr{G}$-measurable and, if $A \in \mathscr{G}$, then

$$\int_A (X_{\mathscr{G}} + Z_{\mathscr{G}})\, d\mathbb{P} = \int_A X\, d\mathbb{P} + \int_A Z\, d\mathbb{P} = \int_A (X+Z)\, d\mathbb{P}.$$

3. If $X \le Z$, let $A = \{X_{\mathscr{G}} > Z_{\mathscr{G}}\} \in \mathscr{G}$. Then

$$\int_A X_{\mathscr{G}}\, d\mathbb{P} = \int_A X\, d\mathbb{P} \le \int_A Z\, d\mathbb{P} = \int_A Z_{\mathscr{G}}\, d\mathbb{P},$$

   which forces $\mathbb{P}(A) = 0$; that is, $X_{\mathscr{G}} \le Z_{\mathscr{G}}$ a.s.

4. Taking $A = \Omega \in \mathscr{G}$: $\underbrace{\int_\Omega X_{\mathscr{G}}\, d\mathbb{P}}_{=\mathbb{E}[\mathbb{E}[X|\mathscr{G}]]} = \underbrace{\int_\Omega X\, d\mathbb{P}}_{=\mathbb{E}[X]}.$

5. If $\mathscr{H} \subset \mathscr{G}$ and $A \in \mathscr{H}$, then $A \in \mathscr{G}$ and

$$\int_A \mathbb{E}[X|\mathscr{G}]\, d\mathbb{P} = \int_A X\, d\mathbb{P} = \int_A \mathbb{E}[X|\mathscr{H}]\, d\mathbb{P};$$

   since $\mathbb{E}[X|\mathscr{H}]$ is $\mathscr{H}$-measurable, it agrees a.s. with $\mathbb{E}[\mathbb{E}[X|\mathscr{G}] \,|\, \mathscr{H}]$.

$\square$

22 Lecture 22

Proposition 22.1. Let X, Z be integrable random variables and $\mathscr{G}$ a sub-σ-algebra of $\mathscr{F}$. Then

1. if Z is $\mathscr{G}$-measurable and ZX is integrable, then
$$\mathbb{E}[ZX|\mathscr{G}] = Z\,\mathbb{E}[X|\mathscr{G}];$$
2. if X is independent of $\mathscr{G}$, then
$$\mathbb{E}[X|\mathscr{G}] = \mathbb{E}[X].$$

Proposition 22.2. Conditional Jensen inequality. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function, X a random variable such that X and $\varphi(X)$ are integrable, and $\mathscr{G}$ a sub-σ-algebra of $\mathscr{F}$. Then

$$\varphi\left(\mathbb{E}[X|\mathscr{G}]\right) \leq \mathbb{E}[\varphi(X)|\mathscr{G}] \qquad \text{a.s.}$$

Proposition 22.3. (Conditional versions of the convergence theorems.) Let (X_n) be a sequence of integrable random variables and $\mathscr{G}$ a sub-σ-algebra of $\mathscr{F}$. All equalities and inequalities hold a.s.

1. If $X_n \uparrow X$ with X integrable, then $\mathbb{E}[X_n|\mathscr{G}] \uparrow \mathbb{E}[X|\mathscr{G}]$.
2. If $X_n \downarrow X$ with X integrable, then $\mathbb{E}[X_n|\mathscr{G}] \downarrow \mathbb{E}[X|\mathscr{G}]$.
3. (Conditional Fatou) If $X_n \geq 0$ for every $n \geq 1$, then
$$\mathbb{E}[\liminf_n X_n|\mathscr{G}] \leq \liminf_n \mathbb{E}[X_n|\mathscr{G}].$$
4. (Conditional DCT) If $|X_n| \leq Z$ for every $n \geq 1$ with Z integrable and $X_n \to X$ a.s., then $\mathbb{E}[X_n|\mathscr{G}] \to \mathbb{E}[X|\mathscr{G}]$.

23 Lecture 23

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and X an integrable random variable. If Y is a random variable, we define

$$\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)].$$

If Y_i, $i \in I$, are random variables, we define

$$\mathbb{E}[X|Y_i,\ i \in I] = \mathbb{E}[X|\sigma(\{Y_i :\ i \in I\})].$$

Theorem 23.1. Let Y, Z be random variables. The following are equivalent:

a) $\sigma(Z) \subset \sigma(Y)$;

b) Z is $\sigma(Y)$-measurable;

c) there exists a Borel measurable function $g : \mathbb{R} \to \mathbb{R}$ with $Z = g(Y)$.

Proof. $(a) \Leftrightarrow (b)$. If $\sigma(Z) \subset \sigma(Y)$ and $A \in \mathscr{B}(\mathbb{R})$, then $Z^{-1}(A) \in \sigma(Z) \subset \sigma(Y)$, so Z is $\sigma(Y)$-measurable. Conversely, if Z is $\sigma(Y)$-measurable, then $\sigma(Z) = \{Z^{-1}(A) : A \in \mathscr{B}(\mathbb{R})\} \subset \sigma(Y)$.
$(c) \implies (b)$ is immediate, since a composition of measurable functions is measurable.
$(b) \implies (c)$. Let first Z be simple, measurable and nonnegative, say

$$Z = \sum_{j=1}^{n} z_j \mathbf{1}_{A_j}.$$

Since $A_j \in \sigma(Y)$,[9]

$$Z = \sum_{j=1}^{n} z_j \mathbf{1}_{B_j}(Y) = g(Y).$$

Now suppose Z is measurable and nonnegative. There is a sequence (Z_n) of nonnegative simple random variables with $Z_n \uparrow Z$, each Z_n being $\sigma(Z)$-measurable, so that

$$\sigma(Z_n) \subset \sigma(Z) \subset \sigma(Y).$$

[9] That is, $A_j = Y^{-1}(B_j)$ with $B_j \in \mathscr{B}(\mathbb{R})$.

By the first part, for each $n \geq 1$ there is a Borel measurable $g_n : \mathbb{R} \to \mathbb{R}$ with $Z_n = g_n(Y)$. Let $\tilde{g} = \sup_{n\geq 1} g_n : \mathbb{R} \to [0, \infty]$. For $\omega \in \Omega$ we have $g_n(Y(\omega)) = Z_n(\omega) \uparrow Z(\omega)$, hence $\tilde{g}(Y(\omega)) = Z(\omega)$. Set $A = \tilde{g}^{-1}(\{\infty\}) \in \mathscr{B}(\mathbb{R})$ and

$$g = \tilde{g} \cdot \mathbf{1}_{A^c} : \ \mathbb{R} \to [0, \infty),$$

which is Borel measurable. For every $\omega \in \Omega$, $Z(\omega) = \tilde{g}(Y(\omega)) = g(Y(\omega))$. Finally, if Z is measurable of arbitrary sign, apply the above to Z^+ and Z^- (which are nonnegative and $\sigma(Y)$-measurable) to obtain

$$Z = Z^+ - Z^- = g_1(Y) - g_2(Y) = g(Y), \qquad g = g_1 - g_2.$$

□

Corollary 23.2. If X is an integrable random variable and Y is a random variable, then

$$\mathbb{E}[X|Y] = g(Y), \tag{1}$$

where $g : \mathbb{R} \to \mathbb{R}$ is a Borel measurable function.

Remark. If $y \in \mathbb{R}$ and g satisfies (1), we define

$$\mathbb{E}[X|Y = y] = g(y).$$

Theorem 23.3. Let X be a nonnegative integrable random variable and Y a random variable. Let P_Y be the law of Y and $\nu : \mathscr{B}(\mathbb{R}) \to \mathbb{R}$ be given by

$$A \mapsto \int_{Y^{-1}(A)} X \, d\mathbb{P}.$$

Then

1. ν is a measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$;

2. $\mathbb{E}[X|Y = y] = \frac{d\nu}{dP_Y}(y)$.

Proof. Item (1) is proved exactly as item (a) of the change-of-measure theorem: ν is the measure with density X with respect to $\mathbb{P}$, transported to $\mathbb{R}$ by Y.

For (2), let $f : \mathbb{R} \to \mathbb{R}$ be Borel measurable with $f(Y) = \mathbb{E}[X|Y]$; such an f exists by the previous corollary. By the definition of conditional expectation, $f(Y)$ is $\sigma(Y)$-measurable and

$$\int_A f(Y)\,d\mathbb{P} = \int_A X\,d\mathbb{P}, \qquad \forall\ A \in \sigma(Y).$$

Every $A \in \sigma(Y)$ is of the form $A = Y^{-1}(B)$ with $B \in \mathscr{B}(\mathbb{R})$. Setting $h(y) = \mathbf{1}_B(y)f(y)$ and using the change-of-variables theorem,

$$\begin{aligned}
\int_A f(Y)\,d\mathbb{P} &= \int_{Y^{-1}(B)} f(Y)\,d\mathbb{P} \\
&= \int_\Omega f(Y)\mathbf{1}_{Y^{-1}(B)}\,d\mathbb{P} \\
&= \int_\Omega h(Y)\,d\mathbb{P} \\
&= \int_\mathbb{R} h(y)\,dP_Y(y) \\
&= \int_\mathbb{R} \mathbf{1}_B(y)f(y)\,dP_Y(y).
\end{aligned}$$

On the other hand, this value equals $\int_{Y^{-1}(B)} X\,d\mathbb{P} = \nu(B)$. Since this holds for every $B \in \mathscr{B}(\mathbb{R})$, we conclude that $d\nu = f\,dP_Y$, that is, $f = \frac{d\nu}{dP_Y}$ P_Y-a.s. □

Remark. If g_1 and g_2 satisfy (1), then $g_1 = g_2$ P_Y-a.s.; in particular, $\mathbb{E}[X|Y = y]$ is well defined for P_Y-almost every y.

Remark. If $y_0 \in \mathbb{R}$ is such that $\mathbb{P}(Y = y_0) > 0$, then

$$\mathbb{E}[X|Y = y_0] = \mathbb{E}[X|\{Y = y_0\}] = \frac{1}{\mathbb{P}\{Y = y_0\}} \int_{\{Y=y_0\}} X\,d\mathbb{P}.$$

Remark. Recall that

$$\mathbb{P}_B(A) = \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Then

$$\mathbb{E}_B[X] = \mathbb{E}[X|B] = \int_\Omega X\,d\mathbb{P}_B = \frac{1}{\mathbb{P}(B)} \int_B X\,d\mathbb{P}.$$

References

[Billingsley(1995)] P. Billingsley. *Probability and Measure*. 3rd ed., Wiley, 1995.

[Durrett(2019)] R. Durrett. *Probability: Theory and Examples*. 5th ed., Cambridge University Press, 2019.

[Folland(1999)] G. B. Folland. *Real Analysis: Modern Techniques and Their Applications*. 2nd ed., Wiley, 1999.

[Le Gall(2022)] J.-F. Le Gall. *Measure Theory, Probability, and Stochastic Processes*. Springer, 2022.

[Williams(1991)] D. Williams. *Probability with Martingales*. Cambridge University Press, 1991.