Prerequisite materials for Stochastic Processes III

Mo Dick Wong

Last update: October 3, 2024

Abstract

This set of notes has been prepared for *Stochastic Processes III* which I taught in Durham University in Michaelmas 2022-2023, and contains mostly materials from first year probability module. All typos are mine and please send any corrections to mo-dick.wong@durham.ac.uk.

Contents

1 Elementary set theory

1.1 Basic set operations

Let Ω be a (non-empty) set.
If $A \subseteq \Omega$ is a subset of Ω, we often denote by $A^c := \Omega \setminus A$ its complement (with respect to Ω).

rcise:eqSets

Exercise 1.1. *Show that two sets A, B are equal to each other iff $A \subseteq B$ and $B \subseteq A$.*

Exercise 1.2 (De Morgan's laws)**.** *Let $\{A_t, t \in \mathcal{T}\}$ be a collection of subsets of Ω ($\mathcal{T}$ could be finite, or countably infinite, or even uncountable). Show that the following are true.*

- $\left(\bigcup_{t\in\mathcal{T}} A_t\right)^c = \bigcap_{t\in\mathcal{T}} A_t^c$.
- $\left(\bigcap_{t\in\mathcal{T}} A_t\right)^c = \bigcup_{t\in\mathcal{T}} A_t^c$.

Exercise 1.3. *Let Ω_1, Ω_2 be two non-empty sets and $f : \Omega_1 \to \Omega_2$ be a function. If $O_2 \subseteq \Omega_2$ is a subset of Ω_2, we define $f^{-1}(O_2) := \{\omega_1 \in \Omega_1 : f(\omega_1) \in O_2\}$, and with abuse of notation write $f^{-1}(\omega_2) := f^{-1}(\{\omega_2\})$ for any $\omega_2 \in \Omega_2$.*

Suppose $\{B_t, t \in \mathcal{T}\}$ is a collection of subsets of Ω_2. Show that

$$f^{-1}\left(\bigcup_{t\in\mathcal{T}} B_t\right) = \bigcup_{t\in\mathcal{T}} f^{-1}(B_t) \qquad \text{and} \qquad f^{-1}\left(\bigcap_{t\in\mathcal{T}} B_t\right) = \bigcap_{t\in\mathcal{T}} f^{-1}(B_t).$$

1.2 Infinite intersection/union

rcise:eqSets

Exercise 1.4. *Using or otherwise, show that for any $a < b$ and $\delta := 0.001(b-a)$, we have*

- $[a, b] = \bigcap_{n=1}^{\infty} \left(a - \frac{\delta}{n}, b + \frac{\delta}{n}\right)$.
- $(a, b) = \bigcup_{n=1}^{\infty} \left[a + \frac{\delta}{n}, b - \frac{\delta}{n}\right]$.

(The moral of the story in the above exercise is that closed intervals can be formed by countable union of open intervals, whereas open intervals can be formed by countable intersection of closed intervals; try to experiment wth half-open/closed intervals as well.)

Recall that any set formed by arbitrary union of open sets is again open.

Exercise 1.5. *Use De Morgan's law or otherwise, explain why arbitrary (and in particular countable) intersection of closed sets gives rise to another closed set.*

1.3 Partition

We say $\Pi := \{B_t, t \in \mathcal{T}\}$ is a partition of a set Ω if $B_t \subseteq \Omega$ for each $t \in \mathcal{T}$ and the following holds:

(i) For any distinct $s, t \in \mathcal{T}$, $B_s \cap B_t = \emptyset$ (mutually exclusive).

(ii) $\cup_{t\in\mathcal{T}} B_t = \Omega$ (i.e. Π forms a cover of Ω).

Exercise 1.6. *Let $\Omega = [0, 1]$. Which of the following is not a partition of Ω? (And why?)*

(1) $\Pi_1 = \{\{x\} : x \in \Omega\}$.

(2) $\Pi_2 = \{(0, \frac{1}{2}], (\frac{1}{2}, 1]\}$.

(3) $\Pi_3 = \{[0, \frac{1}{2}], [\frac{1}{2}, 1]\}$.

(4) $\Pi_4 = \{(2^{-n}, 2^{-(n-1)}] : n \in \mathbb{N}\}$.

Solution:

- Π_2 *is not a partition because* $\cup_{A\in\Pi_2} A = (0, 1]$ *which does not contain* 0.
- Π_3 *is not a partition because the two intervals share the common point* $\frac{1}{2}$.
- Π_4 *is not a partition of* $[0, 1]$ *but a partition of* $(0, 1]$.

2 Probability space

When we speak of a probability space, we mean a triple $(\Omega, \mathcal{F}, \mathbb{P})$:

2.1 Sample space

The (non-empty) set Ω is known as the **sample space**. This is the set of all possible outcomes in an experiment. We say a sample space is **discrete** when it is countable (finite or not).

Example 2.1. *In an experiment involving* 5 *coin tosses, the sample space would be given by*

$$\Omega := \{\omega = (\omega_1, \dots, \omega_5) \in \{\mathrm{H}, \mathrm{T}\}^5\},$$

i.e. each sample is of the form $\omega = (\omega_1, \dots, \omega_5)$ *where* $\omega_i \in \{\mathrm{H}, \mathrm{T}\}$ *is the result in the i-th coin toss.*

Exercise 2.2. *What would be a sample space for an experiment with two dice throw?*

Solution: $\Omega := [6] \times [6] = \{(i, j) : i, j \in \{1, \dots, 6\}\}$.

2.2 σ-algebra

$\mathcal{F}$ is a **σ-algebra** on Ω, i.e. a collection of events under consideration that satisfies the following compatibility conditions:

(SA-1) $\emptyset \in \mathcal{F}$.

(SA-2) If $A \in \mathcal{F}$, then $A^c := \Omega \setminus A \in \mathcal{F}$.

(SA-3) If $A_n \in \mathcal{F}$ for all $n \in \mathbb{N}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Exercise 2.3. *Show that (SA-3) can be replaced by*

(SA-4) If $A_n \in \mathcal{F}$ *for all* $n \in \mathbb{N}$, *then* $\bigcap_{n=1}^{\infty} A_n \in \mathcal{F}$.

Solution: this follows from De Morgan's law: $\bigcap_{n=1}^{\infty} A_n = (\bigcup_{n=1}^{\infty} A_n^c) \in \mathcal{F}$.

When the underlying sample space Ω is discrete, we usually set $\mathcal{F} = 2^{\Omega}$ = the collection of all subsets of Ω (including itself and the emptyset $\emptyset$).

Exercise 2.4. *Let* $\Omega = \{\mathrm{H}, \mathrm{T}\}$. *What is* 2^{Ω}*?*

Solution: $2^{\Omega} = \{\emptyset, \Omega, \{\mathrm{H}\}, \{\mathrm{T}\}\}$.

Exercise 2.5. *Which of the following is a σ-algebra of the sample space for two coin tosses* $\Omega := \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$*?*

(1) $\mathcal{F}_1 := \{\emptyset, \Omega\}$.

(2) $\mathcal{F}_2 := \{\emptyset, \Omega, \{\mathrm{HH}\}, \{\mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}\}$.

(3) $\mathcal{F}_3 := \{\emptyset, \Omega, \{\mathrm{HH}, \mathrm{HT}\}, \{\mathrm{TH}, \mathrm{TT}\}\}$.

(4) $\mathcal{F}_4 := \{\emptyset, \Omega, \{\mathrm{HH}\}, \{\mathrm{TT}\}, \{\mathrm{HT}, \mathrm{TH}\}\}$.

Solution: all of them except $\mathcal{F}_4$, *since* $\{\mathrm{HH}\} \cup \{\mathrm{TT}\} = \{\mathrm{HH}, \mathrm{TT}\} \notin \mathcal{F}_4$.

Exercise 2.6. *Following the last example, what is the smallest σ-algebra containing* $\{\mathrm{HH}\}$ *and* $\{\mathrm{HT}\}$*?*

Solution: $\{\emptyset, \Omega, \{\mathrm{HH}\}, \{\mathrm{HT}\}, \{\mathrm{HH}, \mathrm{HT}\}, \{\mathrm{TH}, \mathrm{TT}\}, \{\mathrm{HH}, \mathrm{TH}, \mathrm{TT}\}, \{\mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}\}$.

Exercise 2.7. *Consider* $\Omega = \mathbb{R}$. *Let* $\mathcal{F}$ *be the Borel σ-algebra of* $\mathbb{R}$, *i.e. the smallest σ-algebra containing all open sets of* $\mathbb{R}$, *and* $\mathcal{G}$ *be the smallest σ-algebra containing all singletons* $\{x\}$. *Show that (i)* $\mathcal{G} \subseteq \mathcal{F}$ *and (ii)* $\mathcal{F} \not\subseteq \mathcal{G}$.

Solution:

(i) We want to show that $\{x\} \in \mathcal{F}$ *for each* $x \in \mathbb{R}$. *But then* $\{x\}^c = (-\infty, x) \cup (x, \infty)$ *which is an open set that is contained in* $\mathcal{F}$ *by definition. Using (SA-2), it follows that* $\{x\} = (\{x\}^c)^c \in \mathcal{F}$.

(ii) *Let us consider $\mathcal{G}' := \{A \subseteq \Omega : \text{only one of } A, A^c \text{ is uncountable}\}$, which is a σ-algebra (check that (SA-1)-(SA-3) are indeed satisfied). Our claim is that $\mathcal{G} = \mathcal{G}'$, from which we can deduce the result because e.g. $(0,1), (0,1)^c \in \mathcal{F}$ but both of these sets are uncountable, i.e. $(0,1) \notin \mathcal{G}$.*

To verify the claim, first observe that $\mathcal{G} \subseteq \mathcal{G}'$, since all singleton sets are automaticaly contained in $\mathcal{G}'$ by definition. On the other hand, if $A \in \mathcal{G}'$ is countable, then $A = \bigcup_{x \in A}\{x\} \in \mathcal{G}$ by (SA-3); if $A \in \mathcal{G}'$ then we instead establish $A^c \in \mathcal{G}$ using the same argument, from which we deduce $A \in \mathcal{G}$ by (SA-2). Thus $\mathcal{G}' \subseteq \mathcal{G}$ and we are done.

2.3 Probability measure

$\mathbb{P}$ is known as the **probability measure**, which is a set-valued function $\mathbb{P} : \mathcal{F} \to [0,1]$ that assigns values to events consistently (known as **probability axioms**):

(PM-1) $\mathbb{P}(A) \geq 0$ for any $A \in \mathcal{F}$.

(PM-2) $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$.

(PM-3) (**Countable additivity.**) If $A_n \in \mathcal{F}$ are mutually disjoint, i.e. $A_j \cap A_k = \emptyset$ for any $j \neq k$, then $\mathbb{P}\left(\bigcup_{n=1}^\infty A_n\right) = \sum_{n=1}^\infty \mathbb{P}(A_n)$.

Exercise 2.8. *Show that the assumption $\mathbb{P}(\emptyset) = 0$ in (PM-2) is redundant.*

Solution: since $\emptyset, \Omega \in \mathcal{F}$ are mutually disjoint, we see that $1 = \mathbb{P}(\Omega) = \mathbb{P}(\Omega \cup \emptyset) = \mathbb{P}(\Omega) + \mathbb{P}(\emptyset) = 1 + \mathbb{P}(\emptyset)$ by (PM-2) and (PM-3), from which we obtain $\mathbb{P}(\emptyset) = 0$. Alternatively, $\mathbb{P}(\emptyset) = \mathbb{P}(\emptyset \cup \emptyset) = 2\mathbb{P}(\emptyset)$.

There are a few basic equalities/inequalities regarding probability measures. We will list few of them below and leave them as revision exercise.

exo:basicPM

Exercise 2.9. *Prove the following claims.*

(1) *Inclusion-exclusion principle: if $A, B \in \mathcal{F}$, then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$. In partcular, if $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.*

(2) *Continuity of probability measure: suppose $A_n \in \mathcal{F}$ is a monotonically increasing sequence of events, i.e.*

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$$

then $\mathbb{P}\left(\bigcup_{n=1}^\infty A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n)$. Hence, establish an analogous claim for monotonically decreasing sequence of events, i.e.

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots \qquad \Rightarrow \qquad \mathbb{P}\left(\bigcap_{n=1}^\infty A_n\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

(3) *Union bound: if $A_n \in \mathcal{F}$ is a sequence of events, then $\mathbb{P}\left(\bigcup_{n=1}^\infty A_n\right) \leq \sum_{n=1}^\infty \mathbb{P}(A_n)$.*

Solution: we only sketch some arguments here but details can be found in your first year probablity notes.

(2) *Note that the analogous result for monotonically decreasing sequence of events follows directly using De Morgan's law:*

$$\mathbb{P}\left(\bigcap_{n=1}^\infty A_n\right) = 1 - \mathbb{P}\left(\bigcup_{n=1}^\infty A_n^c\right) = 1 - \lim_{n\to\infty} \mathbb{P}\left(A_n^c\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

(3) *The union bound follows easily from countable additivity. If we write $B_n = \bigcup_{k=1}^n A_k$ and $C_n := A_n \setminus B_n$ (with the convention that $C_1 := A_1$), then $C_n \subseteq A_n$ and are mutually disjoint and $\bigcup_{n=1}^\infty C_n = \bigcup_{n=1}^\infty A_n$. Hence*

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

3 Random variables

Let $\mathcal{S}$ be a set. We say X is a $\mathcal{S}$-valued random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ if it is a "nice" function $X : \Omega \to \mathcal{S}$. We are typically in one of the following scenarios:

- $\mathcal{S} = \mathbb{R}$: in which case we say X is a real-valued random variable.
- $\mathcal{S} = \mathbb{R}^d$: in which case we often say X is a (d-dimensional real) random vector, since we can write $X = (X_1, \ldots, X_d)$ where each of $X_i : \Omega \to \mathbb{R}$ is a random variable,
- $X(\Omega)$ (or $\mathcal{S}$) is countable: in which case we say X is a **discrete** random variable.

3.1 "Nice-ness": measurable functions

Those of you who have done Probability II would have seen the proper definition of "nice-ness", i.e. measurability, at least for real-valued random variables. This concept was also briefly mentioned in Probability I but was not examinable.

Let us now briefly discuss this concept with some examples/exercise. We will also go through this concept during the first few weeks of our lectures in Michaelmas when we review basic probability theory, but you are strongly encouraged to go through the explanations below before the start of the term.

Real-valued random variables. In Probability II, the definition of a random variable probably goes like this: *a (real-valued) random variable X on a probablity space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ such that*

$$X^{-1}((-\infty, x]) := \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \qquad \forall x \in \mathbb{R}. \tag{3.1}$$

{eq:realRVold}

This definition may not look very motivated, and so we will consider the following definition instead and see why this is equivalent.

Definition 3.1. *We say X is a real-valued random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ if $X : \Omega \to \mathbb{R}$ is a function such that*

$$X^{-1}(O) := \{\omega \in \Omega : X(\omega) \in O\} \in \mathcal{F} \qquad \text{for all open subsets } O \subseteq \Omega. \tag{3.2}$$

{eq:realRV}

Exercise 3.2. *Show that the two definitions of real-valued random variable are equivalent.*

Solution:

- *Let us first show that* (3.2) *implies* (3.1).

  Suppose X is a random variable under the definition (3.2). *We want to establish $X^{-1}((-\infty, x]) \in \mathcal{F}$, but*

  $$X^{-1}\left(\bigcap_{n=1}^{\infty} (-\infty, x + n^{-1})\right) = \bigcap_{n=1}^{\infty} X^{-1}\left((-\infty, x + n^{-1})\right) \in \mathcal{F}$$

  which follows from axiom (SA-4) of σ-algebra (which followed from De Morgan's law and the equivalent (SA-3) axiom).

- *Let us show* (3.1) *implies* (3.2). *First, we show that $X^{-1}((a, b]) \in \mathcal{F}$ for any $a < b$. This is true because*

  $$\begin{aligned} X^{-1}\left((a, b]\right) &= X^{-1}\left((-\infty, b]\right) \setminus X^{-1}\left((-\infty, a]\right) \\ &= X^{-1}\left((-\infty, b]\right) \cap X^{-1}\left((-\infty, a]\right)^c \end{aligned}$$

  and we just need to recall the fact that $X^{-1}\left((-\infty, a]\right)^c \in \mathcal{F}$ by (SA-2), and hence $X^{-1}\left((-\infty, b]\right) \cap X^{-1}\left((-\infty, a]\right)^c \in \mathcal{F}$ by (SA-3').

Next, we need to use a fact about open sets in $\mathbb{R}$: any open sets $O \subseteq \mathbb{R}$ can be written as a countable union of open intervals $O = \bigcup_{n=1}^{\infty}(a_n, b_n)$. Then

$$X^{-1}(O) = \bigcup_{n=1}^{\infty} X^{-1}((a_n, b_n))$$

which is in $\mathcal{F}$ if we know $X^{-1}((a_n, b_n)) \in \mathcal{F}$ for every $n \in \mathbb{N}$ because of axiom (SA-3) of σ-algebras. But each (a_n, b_n) can again be written as a countable union e.g. $\cup_{k=1}^{\infty}(a_n, b_n - \frac{1}{k}]$ and so the claim follows from another application of (SA-3).

General $\mathcal{S}$-valued random variables. In order to talk about (Borel) random variables taking values in general set $\mathcal{S}$, we need to equip $\mathcal{S}$ with a topology, i.e. which subsets $O \subseteq \mathcal{S}$ are defined as open sets of $\mathcal{S}$. Then:

Definition 3.3. *An $\mathcal{S}$-valued random variable X on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathcal{S}$ such that*

$$X^{-1}(O) := \{\omega \in \Omega : X(\omega) \in O\} \in \mathcal{F} \qquad \text{for all open subsets } O \subseteq \mathcal{S}. \tag{3.3}$$ {eq:genRV}

One way to understand the definition of general random variables is to view it as generalisation of the concept of continuity: we say $f : \Omega \to \mathcal{S}$ is a continuous function if

$$f^{-1}(O) := \{\omega \in \Omega : f(\omega) \in O\} \quad \text{is an open subset of } \Omega. \tag{3.4}$$ {eq:continuous}

Discrete random variables. When $\mathcal{S}$ is countable, we equip it with the discrete topology, i.e. all subsets of $\mathcal{S}$ (and in particular all singleton sets) are declared open in $\mathcal{S}$. Under such scenario, $X : \Omega \to \mathcal{S}$ is a random variable if and only if $X^{-1}(s) = \{\omega \in \Omega : X(\omega) = s\} \in \mathcal{F}$ for all $s \in \mathcal{S}$.

Note that this definition is compatible with the situation where $\mathcal{S} = \mathbb{R}$ but $X(\Omega)$ is countable:

Exercise 3.4. *Suppose $X : \Omega \to \mathbb{R}$ is a function on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathcal{X} := \{X(\omega) : \omega \in \Omega\}$ is countable.*

(i) Show that if $X^{-1}(x) \in \mathcal{F}$ for every $x \in \mathbb{R}$, then X is a random variable in the sense of (3.2).

(ii) Show that if X is a random variable in the sense of (3.2), *then $X^{-1}(x) \in \mathcal{F}$ for every $x \in \mathbb{R}$.*

Solution:

- *Note that for any open subset $O \subseteq \mathbb{R}$, the set $O \cap \mathcal{X}$ is countable and thus*

$$X^{-1}(O) = \bigcup_{x \in O} X^{-1}(x) = \bigcup_{x \in O \cap \mathcal{X}} X^{-1}(x) \in \mathcal{F}.$$

- *Note that $X^{-1}(x) = \left(X^{-1}(x)^c\right)^c = \left(X^{-1}((-\infty, x)) \cup X^{-1}((x, \infty))\right)^c \in \mathcal{F}$.*

Exercise 3.5. *If $X = (X_1, \ldots, X_d)$ is a $\mathbb{R}^d$-valued random variable in the sense of* (3.3), *show that each of X_i is a real-valued random variable in the sense of* (3.2).

Solution: let $\pi_j : \mathbb{R}^d \to \mathbb{R}, x = (x_1, \ldots, x_d) \mapsto \pi(x) := x_j$. Then π_j is obviously a continuous function and $X_j = \pi_j \circ X$. Thus for any open set $O \subseteq \mathbb{R}$, the set $\pi_j^{-1}(O)$ is an open set in $\mathbb{R}^d$ and hence

$$X_j^{-1}(O) = X^{-1}\left(\pi_j^{-1}(O)\right) \in \mathcal{F}.$$

3.2 Independence

We say two random variables X_1, X_2 are **independent** of each other if

$$\mathbb{P}(X_1 \in A_1, X_2 \in A_2) = \mathbb{P}(X_1 \in A_1)\mathbb{P}(X_2 \in A_2) \tag{3.5}$$ {eq:indep}

for all "nice" subsets A_1 and A_2. When X_1 and X_2 are both discrete, (3.5) can be reduced to

$$\mathbb{P}(X_1 = x_2, X_2 = x_2) = \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2) \qquad \forall (x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2. \tag{3.6}$$ {eq:indepDis}

Exercise 3.6. *Show the equivalence of* (3.5) *and* (3.6) *when the random variables X_1, X_2 are discrete.*

Solution: it is obvious that (3.5) *is stronger than* (3.6) *(since we can always take $A_1 = \{x_1\}$ and $A_2 = \{x_2\}$), so we only need to show that* (3.6) *implies* (3.5) *in the discrete case, but*

$$\begin{aligned}\mathbb{P}(X_1 \in A_1, X_2 \in A_2) &= \sum_{x_1 \in A_1, x_2 \in A_2} \mathbb{P}(X_1 = x_1, X_2 = x_2) \\ &= \sum_{x_1 \in A_1, x_2 \in A_2} \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2) = \mathbb{P}(X_1 \in A_1)\mathbb{P}(X_2 \in A_2)\end{aligned}$$

where the first equality follows from countable addivitiy of probability measure, and the second equality from independence in the sense of (3.6).

Observe that the notion of independence of random variables generalises that of independent events: $E_1, E_2 \in \mathcal{F}$ are said to be **independent** if $\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2)$.

Exercise 3.7. *Let $E_1, E_2 \in \mathcal{F}$ be two events and $X_i := 1_{E_i}$ the associated indicators. Show that:*

(i) the independence of X_1 and X_2 implies that of E_1 and E_2; and

(ii) the independence of E_1 and E_2 implies that of X_1 and X_2.

Solution:

(i) Suppose X_1 is independent of X_2, then

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(X_1 = 1, X_2 = 1) = \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1) = \mathbb{P}(E_1)\mathbb{P}(E_2),$$

i.e. E_1 is independent of E_2.

(ii) Suppose E_1 is independent of E_2. Since $X_1, X_2 \in \{0, 1\}$ are discrete random variables, we just have to establish

$$\mathbb{P}(X_1 = i_1, X_2 = i_2) = \mathbb{P}(X_1 = i_1)\mathbb{P}(X_2 = i_2) \qquad \forall i_1, i_2 \in \{0, 1\}.$$

This can be verified in a straightforward way.

Exercise 3.8. *Can an event $A \in \mathcal{F}$ be independent of itself?*

Solution: suppose A is independent of itself. Then

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)\mathbb{P}(A) \Rightarrow \mathbb{P}(A) \in \{0, 1\}.$$

This shows that $\mathbb{P}(A) \in \{0, 1\}$ is a necessary and sufficient condition for "self-independence". Two examples of such sets are: $\emptyset$ and Ω.

Mutual independence. Let $\{X_i, i \in \mathcal{I}\}$ be a collection of random variables (on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$). This collection may be finite, countably infinite, or even uncountably infinite. We say the random variables X_i are **mutually independent** if any finite subcollection is mutually independent, i.e. for any $n \in \mathbb{N}$, and any distinct $i_1, \ldots, i_n \in \mathcal{I}$, we have

$$\mathbb{P}\left(X_{i_1} \in A_{i_1}, \ldots, X_{i_n} \in A_{i_n}\right) = \mathbb{P}(X_{i_1} \in A_{i_1}) \cdots \mathbb{P}(X_{i_n} \in A_{i_n})$$

for any "nice" subsets $A_{i_1}, \ldots, A_{i_n}$.

Note that *pairwise* independence is weaker than mutual independence, as the following exercise shows.

Exercise 3.9. *Let X_1, X_2 be independent random variables with $\mathbb{P}(X_i = \pm 1) = \frac{1}{2}$, and let $Y := X_1 X_2$. Show that*

- *Y is independent of X_i for each i separately; but*
- *X_1, X_2 and Y are not mutually independent.*

Solution: first observe that $\mathbb{P}(Y = \pm 1) = \frac{1}{2}$. Then

- *it is straightforward to check that e.g. $\mathbb{P}(X_i = j, Y = k) = \mathbb{P}(X_i = j)\mathbb{P}(Y = k)$ for $i \in \{1, 2\}$ and any $j, k \in \{\pm 1\}$; but*
- *$\mathbb{P}(X_1 = 1, X_2 = 1, Y = -1) = 0 \neq \frac{1}{8} = \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1)\mathbb{P}(Y = -1)$.*

3.3 Distributions

Let X be a general $\mathcal{S}$-valued random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The distribution of $\mathcal{X}$ (with respect to $\mathbb{P}$) is a set-valued function

$$\mathbb{P}_X : A \mapsto \mathbb{P}_X(A) := \mathbb{P}(X \in A) \qquad \text{for all "nice" subsets } A \subseteq \mathcal{S}.$$

When $\mathcal{S} = \mathbb{R}^d$ and $X = (X_1, \ldots, X_d)$, we also define the **(joint) cumulative distribution function (c.d.f.)** of X as

$$F_X(x) = F_{X_1,\ldots,X_d}(x_1, \ldots, x_d) := \mathbb{P}\left(X_1 \leq x_1, \ldots, X_d \leq x_d\right) \qquad \forall x = (x_1, \ldots, x_d) \in \mathbb{R}^d. \tag{3.7}$$

{eq:jointcdf}

istributions

Exercise 3.10. *Do you recognise the following distributions?*

- Bernoulli(p) *and* Binomial(n, p).
- Geometric(p) *and* NegativeBinomial(n, p).
- Poisson(λ).
- Exp(λ) *and* Gamma(n, λ).
- Uniform$[0, 1]$.
- $\mathcal{N}(\mu, \sigma^2)$.

Exercise 3.11. *Show that joint c.d.f.s are left-continuous with right limits, i.e.*

$$\lim_{x_j \to c^{\pm}} F_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = F_{X_1,\ldots,X_d}(x_1, \ldots, x_{j-1}, c, x_{j+1}, \ldots, x_d)$$

and $\lim_{x_j \to c^-} F_{X_1,\ldots,X_d}(x_1, \ldots, x_d)$ exists (but not necessarily equal to the right limit).

Solution: the claim follows from continuity of probability measure in the sense of Exercise 2.9(2), since $\lim_{x_j \to c^{\pm}} F_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = \lim_{n \to \infty} \mathbb{P}(A_n^{\pm})$ where

$$A_n^{\pm} := \left\{ X_k \leq x_k \quad \forall k \neq j \text{ and } X_j \leq c \pm \frac{1}{n} \right\}.$$

which satisfies $A_1^+ \supseteq A_2^+ \supseteq \cdots$ and $A_1^- \subseteq A_2^- \subseteq \cdots$ with

$$\bigcap_{n \geq 1} A_n^+ = \{X_k \leq x_k \quad \forall k \neq j \text{ and } X_j \leq c\},$$

$$\text{and} \qquad \bigcup_{n \geq 1} A_n^- = \{X_k \leq x_k \quad \forall k \neq j \text{ and } X_j < c\}.$$

Discrete real-valued random variables. Let $\mathcal{S} = \mathbb{R}^d$ and $X = (X_1, \ldots, X_d)$ be discrete, i.e. $\mathcal{X}_j := \{X_j(\omega) : \omega \in \Omega\}$ is countable for each $j = 1, \ldots, d$. We often speak of the **(joint) probability mass function (p.m.f.)**, i.e. the function

$$\begin{aligned} p_{X_1,\ldots,X_d} &: \mathbb{R}^d \to \mathbb{R}_+ \\ &(x_1, \ldots, x_d) \mapsto p_{X_1,\ldots,X_d}(x_1, \ldots, x_d) := \mathbb{P}(X_1 = x_1, \ldots, X_d = x_d). \end{aligned}$$

Continuous real-valued random variables. We say $X = (X_1, \ldots, X_d)$ has jointly continuous distribution if (3.7) is **continuous** as a function of $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$. Furthermore, the joint distribution is said to be **absolutely continuous**[1] if it has a density function $f_X(x) = f_{X_1,\ldots,X_d}(x_1, \ldots, x_d)$, i.e.

$$\mathbb{P}((X_1, \ldots, X_d) \in A) = \int_A f_{X_1,\ldots,X_d}(x_1, \ldots, x_d) dx_1 \cdots dx_d.$$

for any "nice" subset $A \subseteq \mathbb{R}^d$, and in particular, with $A := (-\infty, x_1] \times \cdots \times (-\infty, x_d]$,

$$F_{X_1,\ldots,X_d}(x_1, \ldots, x_d) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} f_{X_1,\ldots,X_d}(x_1, \ldots, x_d) dx_1 \cdots dx_d.$$

Exercise 3.12. *Given any monotonically increasing function $f : \mathbb{R} \to \mathbb{R}$ (i.e. $f(x) \geq f(y)$ whenever $x \geq y$), the generalised inverse $\overleftarrow{f}$ is defined as*

$$\overleftarrow{f}(y) := \inf\{x \in \mathbb{R} : f(x) \geq y\} \qquad \forall y \in I_f := \{f(u) : u \in \mathbb{R}\}.$$

Let X be a real-valued random variable with c.d.f. F_X and U be a Uniform$[0, 1]$ *random variable. What is the distribution of the random variable $\overleftarrow{F_X}(U)$?*

Solution: note that

$$\mathbb{P}\left(\overleftarrow{F}_X(U) \leq u\right) = \mathbb{P}(U \leq F_X(u)) = \int_0^{F_X(u)} ds = F_X(u)$$

by the definition of $\overleftarrow{F}_X$ (and the fact that the c.d.f. F_X is right-continuous).

Exercise 3.13. *Let X_1, X_2 be two independent* Exp(2) *random variables, i.e. $\mathbb{P}(X_1 \leq x) = \mathbb{P}(X_2 \leq x) = 1 - e^{-2x}$ for any $x \geq 0$. Find:*

(i) the value of $\mathbb{P}(X_1 < X_2)$.

(ii) the probability density function of $\min(X_1, X_2)$*;*

(iii) the probability density function of $X_1 + X_2$;

Solution:

(i) Since $\mathbb{P}(X_1 = X_2) = 0$, we have $\mathbb{P}(X_1 < X_2) = \mathbb{P}(X_2 < X_1) = \frac{1}{2}$ by symmetry.

(ii) Note that $Y := \min(X_1, X_2)$ satisfies

$$\mathbb{P}(Y > y) = \mathbb{P}(\min(X_1, X_2) > y) = \mathbb{P}(X_1 > y, X_2 > y) = \mathbb{P}(X_1 > y)\mathbb{P}(X_2 > y) = e^{-4y}$$

for any $y \geq 0$, i.e. Y is distributed according to Exp(4).

(iii) We derive the c.d.f. of $Y := X_1 + X_2$ by integration: since $f_{X_i}(x_i) := 2e^{-2x_i} 1_{\{x_i \geq 0\}}$, we have

$$\begin{aligned} \mathbb{P}(Y > y) &\overset{y \geq 0}{=} \int_{\{x_1 + x_2 \leq y\}} 4e^{-2x_1 - 2x_2} dx_1 dx_2 \\ &= \int_0^\infty \int_{(y - x_2)_+}^\infty 4e^{-2x_1 - 2x_2} dx_1 dx_2 \qquad (\textit{here } (y - x_2)_+ := \max(y - x_2, 0)) \\ &= \int_0^\infty 2e^{-2x_2} e^{-2(y - x_2)_+} dx_2 = 2ye^{-2y} - e^{-2y}. \end{aligned}$$

Equivalently $f_Y(y) = 4ye^{-2y} 1_{\{y \geq 0\}}$, i.e. $Y \sim$ Gamma(2, 2) *random variable.*

[1] With respect to Lebesgue measure.

4 Expectation

4.1 Real-valued random variables

Let us recall the definition of **expectation** or **expected value** from first year probability: let X be a real-valued random variable and $g : \mathbb{R} \to \mathbb{R}$ is some "nice" function, then

$$\mathbb{E}\left[g(X)\right] := \begin{cases} \displaystyle\sum_{x \in \mathcal{X}} g(x) p_X(x) & \text{if } \mathcal{X} := \{X(\omega) : \omega \in \Omega\} \text{ is discrete/countable} \\ \displaystyle\int_{\mathbb{R}} g(x) f_X(x) dx & \text{if } X \text{ is absolutely continuous with p.d.f. } f_X(\cdot) \end{cases} \tag{4.1}$$

{eq:expDef}

provided that the sum/integral is well-defined, i.e. $\mathbb{E}[|g(X)|] < \infty$, in which case we say $g(X)$ is **integrable**.

Exercise 4.1. *Let $A_1, A_2 \in \mathcal{F}$ and $X_i := 1_{A_i}$ be the associated indicator functions. What is the value of $\mathbb{E}[X_1^p X_2^q]$ for $p, q > 0$?*

Solution: $\mathbb{E}[X_1^p X_2^q] = \mathbb{E}[1_{A_1}^p 1_{A_2}^q] = \mathbb{P}(A_1 \cap A_2)$.

Some common terminologies for a real-valued random variable X:

- the **mean** of X is $\mathbb{E}[X]$
- for $p \in \mathbb{R}$ we call $\mathbb{E}[X^p]$ the **p-th moment** of X, and $\mathbb{E}[|X|^p]$ the **p-th absolute moment**.
- the **variance** of X is defined as $\mathrm{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2]$.

Exercise 4.2. *Compute the mean and variance of distributions in Exercise 3.10.*

More generally, if one has a collection of real-valued random variables $(X_1, \ldots, X_d)$ and $g : \mathbb{R}^d \to \mathbb{R}$ is "nice", then

$$\mathbb{E}[g(X_1, \ldots, X_d)]$$
$$:= \begin{cases} \displaystyle\sum_{x_1 \in \mathcal{X}_1, \ldots, x_d \in \mathcal{X}_d} g(x_1, \ldots, x_d) p_{X_1, \ldots, X_d}(x_1, \ldots, x_d) & \text{if } \mathcal{X}_i := \{X_i(\omega) : \omega \in \Omega\} \text{ are countable} \\ \displaystyle\int_{\mathbb{R}^d} g(x_1, \ldots, x_d) f_{X_1, \ldots, X_d}(x_1, \ldots, x_d) dx_1 \cdots dx_d & \text{if } (X_1, \ldots, X_d) \text{ are jointly absolutely continuous} \end{cases} \tag{4.2}$$

{eq:expMultiple}

provided that the sum/integral is well-defined (i.e. integrable).

Exercise 4.3. *Let X_1, X_2 be two independent* $\mathrm{Exp}(1)$ *random variables. Find* $\mathbb{E}[\min(X_1, X_2)^2]$ *by:*

(i) direct integration; and

(ii) using the distribution of $Y = \min(X_1, X_2)$.

Solution: we omit (i) here which can be done by direct computation. Let us consider (ii), and note that $Y \sim \mathrm{Exp}(2)$ since $\mathbb{P}(Y > y) = \mathbb{P}(X_1 > y)\mathbb{P}(X_2 > y) = e^{-2y}$ for any $y \geq 0$ by independence. Therefore

$$\mathbb{E}[\min(X_1, X_2)^2] = \mathbb{E}[Y^2] = \int_0^\infty y^2 \cdot 2e^{-2y} dy = \frac{1}{2}.$$

Given two real-valued random variables X_1 and X_2, the **covariance** $\mathrm{Cov}(X_1, X_2)$ is defined as $\mathbb{E}[(X_1 - \mathbb{E}[X_1])(X_2 - \mathbb{E}[X_2])$. Notice that this generalises the concept of variance, since $\mathrm{Var}(X) = \mathrm{Cov}(X, X)$.

4.2 Properties of expectation

Given the definition of expected value (4.2) as a summation/integration, it is not hard to see that $\mathbb{E}[\cdot]$ is a linear map: if $X_1, \ldots, X_d$ are a collection of real-valued random variables and $a_1, \ldots, a_d \in \mathbb{R}$ are constants, then

$$\mathbb{E}\left[\sum_{k=1}^{d} a_k X_k\right] = \sum_{k=1}^{d} a_k \mathbb{E}[X_k] \tag{4.3}$$ {eq:linearity}

provided that all the expectations appearing in the equality exist. This may be extended to a countable collection of random variables if certain conditions are satisfied (using monotone/dominated convergence theorems).

Exercise 4.4. *Using linearity, prove the following properties of covariance: if X, Y, Z are random variables with finite second moments, and $a, b, c \in \mathbb{R}$ are constants, then*

(i) $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$.

(ii) $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$. *In particular,* $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.

(iii) $\mathrm{Cov}(aX, bY + cZ) = ab\mathrm{Cov}(X, Y) + ac\mathrm{Cov}(X, Z)$.

Solution: *Let us just focus on (iii) and use (ii). Then*

$$\begin{aligned}\mathrm{Cov}(aX, bY + cZ) &= \mathbb{E}[(aX)(bY + cZ)] - \mathbb{E}[aX]\mathbb{E}[bY + cZ] \\ &= ab\left(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\right) + ac\left(\mathbb{E}[XZ] - \mathbb{E}[X]\mathbb{E}[Z]\right)\end{aligned}$$

which is the desired claim.

Another important property of expectation is factorisation when working with independent random variables. More precisely, if X_1, X_2 are two independent (real-valued) random variables, then for any "nice" functions $g_i : \mathbb{R} \to \mathbb{R}$,

$$\mathbb{E}[g_1(X_1)g_2(X_2)] = \mathbb{E}[g_1(X_1)]\mathbb{E}[g_2(X_2)]$$

provided that the expected values appearing on both sides on the equality are well-defined.

Exercise 4.5. *Prove the following formula for independent random variables $X_1, \ldots, X_d$ with finite second moments: if $a_1, \ldots, a_d \in \mathbb{R}$ are some arbitrary constants, then*

$$\mathrm{Var}\left(a_1 X_1 + \cdots + a_d X_d\right) = a_1^2 \mathrm{Var}(X_1) + \cdots + a_d^2 X_d \mathrm{Var}(X_d).$$

Solution: *Using the previous exercise, we see that*

$$\mathrm{Var}\left(\sum_{k=1}^{d} a_k X_k\right) = \sum_{j,k} a_j a_k \mathrm{Cov}(X_j, X_k)$$

where

$$\mathrm{Cov}(X_j, X_k) = \mathbb{E}[X_j X_k] - \mathbb{E}[X_j]\mathbb{E}[X_k] = \begin{cases} 0 & \text{if } j = k \text{ (by independence)} \\ \mathrm{Var}(X_j) & \text{if } j = k \end{cases}$$

and we are done.

4.3 Elementary conditional expectations

Let X be an integrable random variable and $B \in \mathcal{F}$ be an event with strictly positive probability $\mathbb{P}(B) > 0$. We define the **conditional expectation** of X given B as

$$\mathbb{E}[X|B] := \frac{\mathbb{E}\left[X 1_B\right]}{\mathbb{P}(B)}.$$

In particular, if $A \in \mathcal{F}$ and $X = 1_A$, then

$$\mathbb{E}[X|B] = \frac{\mathbb{E}[1_A 1_B]}{\mathbb{P}(B)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A|B).$$

Partition theorem. More generally, if $\Pi := \{B_i \in \mathcal{I}\}$ is a partition of the sample space Ω with at most countably many parts B_i, then the partition theorem says

$$\mathbb{E}[X] = \sum_{i \in \mathcal{I}} \mathbb{E}[X|B_i]\mathbb{P}(B_i)$$

provided that we interpret $\mathbb{E}[X|B_i]\mathbb{P}(B_i) = 0$ as soon as $\mathbb{P}(B_i) = 0$. This generalises the partition theorem for probability: if we take $X = 1_A$ for some event $A \in \mathcal{F}$, then X is obviously integrable and

$$\mathbb{P}(A) = \mathbb{E}[X] = \sum_{i \in \mathcal{I}} \mathbb{E}[1_A|B_i]\mathbb{P}(B_i) = \sum_{i \in \mathcal{I}} \mathbb{P}(A|B_i)\mathbb{P}(B_i).$$

se:thinning

Exercise 4.6. *Let* $X \sim \mathrm{Binomial}(N, p)$ *where* N *is an independent* $\mathrm{Poisson}(\lambda)$ *variable. Find* $\mathbb{E}[X]$ *in terms of* p *and* λ.

Solution: using the partition theorem, we get

$$\mathbb{E}[X] = \sum_{n \geq 0} \mathbb{E}[X|N = n]\mathbb{P}(N = n) = \sum_{n \geq 0} np\mathbb{P}(N = n) = p\mathbb{E}[N] = p\lambda.$$

4.4 Generating functions

Let X be a real-valued random variable.

- The moment generating function (m.g.f.) $M_X(\cdot)$ is defined as $M_X(t) := \mathbb{E}[e^{tX}]$ for values of $t \in \mathbb{R}$ such that e^{tX} is integrable.
- When X only takes values in $\{0, 1, \ldots,\} =: \mathbb{N}_0$, we also speak of the probability generating function

  $$G_X(t) := \mathbb{E}[t^X] = \sum_{n=0}^{\infty} p_X(n)t^n$$

  for all values of $t \in \mathbb{R}$ such that the above sum converges. Note that probability generating function always exists for $|t| < 1$ for such a random variable X.

Exercise 4.7. *What are the generating functions for the distrbutions in Exercise 3.10?*

One important fact about generating functions is that they **characterise** the distributions of the underlying random variables: e.g. if $M_X(t)$ exists for t inside some (non-empty) open interval, then there is only one probability distribution that gives rise to this moment generating function. The case for probability generating function is similar.[2]

Exercise 4.8. *What are the connections between*

(i) $\mathrm{Bernoulli}(p)$ *and* $\mathrm{Binomial}(n, p)$*?*

(ii) $\mathrm{Geometric}(p)$ *and* $\mathrm{NegativeBinomial}(n, p)$*?*

(iii) $\mathrm{Exp}(\lambda)$ *and* $\Gamma(n, \lambda)$*?*

Solution:

The latter are sums of n i.i.d. copies of random variables with the former distributions. This can be easily verified by comparing moment generating functions, for instance.

Exercise 4.9. *What is the distribution of the random variable X in Exercise 4.6?*

Solution:

Let us identify the distribution by means of probability generating function:

$$\mathbb{E}\left[t^X\right] = \sum_{n \geq 0} \mathbb{E}\left[t^X|N = n\right] e^{-\lambda}\frac{\lambda^n}{n!} = e^{-\lambda}\sum_{n \geq 0} \left[(1-p) + pt\right]^n \frac{\lambda^n}{n!} = e^{p\lambda(t-1)},$$

i.e. $X \sim \mathrm{Poisson}(p\lambda)$.

[2]And always holds if X is non-negative because it always exists for $|t| < 1$.

Note that these generating functions get their names because we can extract information about moments/probabilities by performing derivatives: provided that we can interchange expectations and differentiations, we have (for any $n \in \mathbb{N}_0$)

$$\mathbb{E}[X^n] = \left.\frac{d^n}{dt^n}\right|_{t=0} \mathbb{E}\left[e^{tX}\right]$$

and for discrete random variables $X \in \mathbb{N}_0$,

$$p_X(n) = \mathbb{P}(X = n) = \frac{1}{n!} \left.\frac{d^n}{dt^n}\right|_{t=0} G_X(t).$$

Note also the connections between the two generating functions: $G_X(e^t) = M_X(t)$.

Exercise 4.10. *Find a formula for positive integer moments of $\mathcal{N}(0, \sigma^2)$ random variables.*

Solution: *Let $X \sim \mathcal{N}(0, \sigma^2)$. Then*

$$M_X(t) = e^{\frac{t^2}{2}\sigma^2} = \sum_{n \geq 0} \frac{1}{n!} \left(\frac{t^2}{2}\sigma^2\right)^n$$

and thus

$$\mathbb{E}[X^n] = \begin{cases} 0 & \text{if } n \text{ is odd} \\ \sigma^n \dfrac{n!}{(n/2)! 2^{n/2}} = \sigma^n (n-1)!! & \text{if } n \text{ is even.} \end{cases}$$

Multivariate generating functions. Given a collection of real-valued random variables $(X_1, \ldots, X_d)$, one can define analogous objects known as joint m.g.f.s and joint p.g.f.s, e.g.

$$M_{X_1, \ldots, X_d}(t_1, \ldots, t_d) := \mathbb{E}\left[\exp\left(\sum_{k=1}^{d} t_k X_k\right)\right]$$

$$\text{and} \qquad G_{X_1, \ldots, X_d}(t_1, \ldots, t_d) := \mathbb{E}\left[\prod_{k=1}^{d} t_k^{X_k}\right]$$

and similar properties hold but we do not discuss them further in this document.

Exercise 4.11. *Let $X_1, \ldots, X_d$ be a collection of real-valued random variables and suppose their joint m.g.f. $M_{X_1, \ldots, X_d}(t_1, \ldots, t_d)$ exists everywhere on $\mathbb{R}^d$. Assuming that you can interchange differentiation and expectation, explain how you may obtain the cross moments*

$$\mathbb{E}\left[\prod_{k=1}^{d} X_k^{n_k}\right] \qquad \text{where} \qquad (n_1, \ldots, n_d) \in \mathbb{N}_0^d.$$

Solution:

We have

$$\mathbb{E}\left[\prod_{k=1}^{d} X_k^{n_k}\right] = \left.\frac{\partial^{n_1}}{\partial t_1^{n_1}}\right|_{t_1=0} \cdots \left.\frac{\partial^{n_d}}{\partial t_d^{n_d}}\right|_{t_d=0} M_{X_1, \ldots, X_d}(t_1, \ldots, t_d)$$

5 Miscellaneous topics

5.1 Markov inequality

Suppose X is a non-negative random variable with finite mean, i.e. $\mathbb{P}(X \geq 0) = 1$ and $\mathbb{E}[X] < \infty$. Markov's inequality says that

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t} \qquad \forall t > 0.$$

Exercise 5.1. *Establish the Chebyshev's inequality, i.e. if X has finite second moment, then*

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

Solution: this follows from Markov's inequality with $Y := |X - \mathbb{E}[X]|^2$, i.e.

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

Exercise 5.2. *Let $X \sim \mathcal{N}(0,1)$. Show that*

$$\mathbb{P}(X > t) \leq e^{-\frac{t^2}{2}}.$$

Solution: this can be obtained by direct computation of $\mathbb{P}(X > t)$ with a bit of analysis, but we shall proceed with the method of Chernoff bound. Indeed, for any $u > 0$, Markov's inequality implies that

$$\mathbb{P}(X > t) = \mathbb{P}(e^{uX} > e^{ut}) \leq e^{-ut}\mathbb{E}[e^{uX}] = e^{-ut}e^{\frac{u^2}{2}}.$$

The claim now follows by choosing $u = t$.

5.2 Modes of convergence

Let $X_1, X_2, \dots$ be a sequence of real-valued random variables. We would like to recall two notions of convergence of random variables from first year probability.

Convergence in probability. We say X_n converges to some random variable X **in probability** if the following holds: for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \epsilon\right) = 0.$$

Exercise 5.3. *Let $X_1, \dots$ be a sequence of i.i.d. random variables with zero mean and finite variance. Show that the sample mean $\overline{X}_n := n^{-1}\sum_{k=1}^n X_k$ converges in probability to 0.*

Solution:

We use Chebyshev's inequality: for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(|\overline{X}_n - 0| > \epsilon\right) \leq \lim_{n\to\infty} \frac{\mathrm{Var}(\overline{X}_n)}{\epsilon^2} = \lim_{n\to\infty} \frac{n^{-1}\mathrm{Var}(X_1)}{\epsilon^2} = 0.$$

Convergence in distrbution. We say X_n converges to some random variable X **in distribution** if

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t) \qquad \text{for any continuity points } t \text{ of } F_X(\cdot). \tag{5.1}$$ {eq:convDis}

In particular, if the random variable X has a continuous distribution, then (5.1) has to hold everywhere on $\mathbb{R}$.

Exercise 5.4. *Suppose X_n converges in probability to some random variable X. Show that X_n also converges in distribution to X, i.e. convergence in probability is a stronger notion than convergence in distribution.*

Solution: suppose $t \in \mathbb{R}$ is a continuity point of F_X, i.e. for any $\epsilon > 0$ there exists $\delta = \delta(\epsilon, t) > 0$ such that $|F_X(s) - F_X(t)| \leq \epsilon$ whenever $|s-t| < \delta$. By checking that

$$F_{X_n}(t) := \mathbb{P}(X_n \leq t) \begin{cases} \leq \mathbb{P}\left(X \leq t + \dfrac{\delta}{2}\right) + \mathbb{P}\left(|X_n - X| > \dfrac{\delta}{2}\right) \\ \geq \mathbb{P}\left(X \leq t - \dfrac{\delta}{2}\right) + \mathbb{P}\left(|X_n - X| > \dfrac{\delta}{2}\right) \end{cases}$$

which means

$$\limsup_{n\to\infty} |F_{X_n}(t) - F_X(t)| \leq \max\left(|F_X(t + \frac{\delta}{2}) - F_X(t)|, F_X(t - \frac{\delta}{2}) - F_X(t)|\right) \leq \epsilon.$$

Since $\epsilon > 0$ is arbitrary, we have established $\lim_{n\to\infty} F_{X_n}(t) = F_X(t)$ for every continuity point t of F_X, i.e. X_n converges in distribution to X.

Exercise 5.5. *Suppose X_n converges in distribution to some constant $c \in \mathbb{R}$. Show that X_n converges in probability to c as well.*

Solution: for any $\epsilon > 0$,

$$\begin{aligned} \mathbb{P}(|X_n - c| > \epsilon) &= \mathbb{P}(X_n > c + \epsilon) + \mathbb{P}(X_n < c - \epsilon) \\ &= 1 - F_{X_n}(c + \epsilon) + F_{X_n}(c - \epsilon) \xrightarrow{n\to\infty} 1 - 1 + 0 = 0. \end{aligned}$$

since $c \pm \epsilon$ are continuity points of the c.d.f. of the constant random variable c.

Exercise 5.6. *Show that convergence in distribution does not necessarily imply convergence in probability.*

Solution: consider $X \sim \mathcal{N}(0,1)$ and define $X_n := (-1)^n X$. Then $X_n \sim \mathcal{N}(0,1)$ but it does not converge in probability to any random variable.

Characterisation of convergence in distribution. We take for granted the following fact regarding generating functions. If X_n is a sequence of random variables and X is another random variable, and suppose there exists a non-empty open interval $I \subseteq \mathbb{R}$ on which our sequence of moment generating functions converges, i.e.

$$\lim_{n\to\infty} M_{X_n}(t) = M_X(t) \qquad \forall t \in I,$$

then X_n converges in distribution to X. An analogous result also holds for probability generating functions.

Exercise 5.7 (Method of moments)**.** *Let X_n be a sequence of uniformly bounded random variables, i.e. there exists some $M > 0$ such that $|X_n| \leq M$ for all n. Suppose there exists some random variable X such that $\lim_{n\to\infty} \mathbb{E}[X_n^k] \to \mathbb{E}[X^k]$ for every positive integers k. Show that X_n converges to X in distribution.*

Solution: let us try to establish the claim with moment generating function. First observe that the moment generating functions for all these random variables exist everywhere on $\mathbb{R}$: indeed

$$M_X(t) := \mathbb{E}\left[e^{tX}\right] = \mathbb{E}\left[\sum_{k\geq 0} \frac{(tX)^k}{k!}\right] = \sum_{k\geq 0} \mathbb{E}\left[X^k\right] \frac{t^k}{k!}$$

which is absolutely summable since $\mathbb{E}[X^k] \leq M^k$. The same holds true for each of X_n. By the assumption of convergence in moments, we see that $M_{X_n}(t) \to M_X(t)$ for every $t \in \mathbb{R}$. The claim follows by the characterisation of convergence in distribution with generating functions.