Probability Theory

An update made by Thomas Walker on notes scribed by Ivan Kirev and Samuel Lam (4)
from a lecture series on Probability Theory given by Professor Igor Krasovsky.

Contents

Part I. Measure Theory and Random Variables

1 Events, Probability and Random Variables

In developing an abstract mathematical framework to describe the likelihood of events happening in a random experiment we can formalise large-sample results, including the Law of Large Numbers and the Central Limit Theorem. An experiment can be described in the following way.

1. Let the possible outcomes ω of the experiment be the sample space, Ω.
2. Let events A be subsets of the sample space Ω that we may observe. The collection of such subsets is denoted as $\mathscr{F}$.
3. Assign a value $\mathbb{P}(A) \in [0,1]$ to each of the subsets $A \in \mathscr{F}$ to quantify how likely the event is to occur.

With this construction, we have a few problems to resolve.

- What should the collection of events $\mathscr{F}$ include?
- How should we assign values to the events in $\mathscr{F}$?

It should be the case that we can either observe noting or something, that is, we should let $\mathscr{F}_* := \{\varnothing, \Omega\} \subseteq \mathscr{F}$. However, any useful experiment should be able to differentiate different observed outcomes, and so $\mathscr{F}$ should include more than just the whole set Ω as a potential result.
One could suggest letting $\mathscr{F}$ contain all possible subsets of Ω. We denote this $\mathscr{F} = \mathscr{F}^* := 2^\Omega$, and refer to it as the power set of Ω. For Ω a countable set, this is fine, and it will often be the case that $\mathscr{F} = 2^\Omega$. However, issues arise if Ω is uncountable, for example, if $\Omega = \mathbb{R}$.
Next, there are certain properties that $\mathbb{P}$ ought to satisfy to make sense. For example, it should be finitely additive. That is, if $A, B \in \mathscr{F}$ are disjoint outcomes, then

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

When one extends this property to countably many disjoint events, potential contradictions may arise if $\mathscr{F}$ does not have a certain structure. In the uncountable case, 2^Ω will often not possess such a structure. Therefore, we want to choose $\mathscr{F}$ such that it contains more information than $\mathscr{F}_*$ but is strictly smaller than 2^Ω.
At this point, one may be wondering why we need to extend the finite additivity property to the countable additivity property. Well, often one is interested in the long-term behaviour of a random experiment, such as the expected value of a dynamical system as it continues to run forward in time. Hence, questions regarding limits arise naturally, for which one needs to reason about countably many events rather than just a finite number.
Mathematicians, therefore, attempted to find a suitable criterion for $\mathscr{F}$ and $\mathbb{P}$ so that they would not give rise to contradictions, but still allow $\mathscr{F}$ to be large enough to be useful. One of the most successful attempts was made by Andrey Kolmogorov in 1933 when he devised the axioms of probability in his Foundations of the Theory of Probability. His work led to the development of the measure theory, which forms the foundations of modern probability theory.

1.1 Algebras and σ-algebras

Let Ω be a set of points ω.

Definition 1.1.1 A nonempty system of subsets of Ω is called an algebra $\mathscr{A}$ if

- $\Omega \in \mathscr{A}$,
- $A, B \in \mathscr{A}$ implies that $A \cup B \in A$, and
- $A \in \mathscr{A}$ implies that $A^c \in \mathscr{A}$.

If in addition, all countable unions $\bigcup_{n=1}^{\infty} A_n \in \mathscr{A}$ whenever $A_1, A_2, \cdots \in \mathscr{A}$, then $\mathscr{A}$ is a σ-algebra.

Remark 1.1.2 Note that we can consider the complements of events to show that a σ-algebra (algebra) is also closed under countable (finite) intersections.

Definition 1.1.3

- A set function $\mu : \mathscr{A} \longrightarrow [0, \infty]$ is finitely additive if for any disjoint $A, B \in \mathscr{A}$ we have
$$\mu(A \cup B) = \mu(A) + \mu(B).$$
- Let $\mathscr{F}$ be a σ-algebra. A set function $\mu : \mathscr{F} \to [0, \infty]$ is σ-additive if for any disjoint $A_1, A_2, \cdots \in \mathscr{F}$, it follows that
$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n).$$
Such a μ is called a measure on $\mathscr{F}$. A measure μ is a probability measure if $\mu(\Omega) = 1$.
- A measure is σ-finite if there exists a partition $\Omega = \bigcup_{k=1}^{\infty} \Omega_k$, where the Ω_k are pairwise disjoint, such $\mu(\Omega_k) < \infty$ for $k \in \mathbb{N}$.

Definition 1.1.4 A probability space is a triple $(\Omega, \mathscr{F}, \mathbb{P})$, where Ω is a set called the sample space, $\mathscr{F}$ is a σ-algebra of subsets of Ω, and $\mathbb{P}$ is a probability measure on $\mathscr{F}$. An element of $\mathscr{F}$ is called an event.

Proposition 1.1.5 Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. The probability measure $\mathbb{P}$ satisfies the following.

1. $\mathbb{P}(\varnothing) = 0$.
2. If $A, B \in \mathscr{F}$ then
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$
3. If $A, B \in \mathscr{F}$ and $B \subseteq A$ then
$$\mathbb{P}(B) \leq \mathbb{P}(A).$$
4. If $A_1, A_2, \cdots \in \mathscr{F}$, then
$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

Proof.

1. As $\varnothing \in \mathscr{F}$ is disjoint from itself, it follows by finite additivity that
$$\mathbb{P}(\varnothing) = \mathbb{P}(\varnothing \cup \varnothing) = 2\mathbb{P}(\varnothing).$$
Hence, $\mathbb{P}(\varnothing) = 0$.
2. As $A \cap B^c \in \mathscr{F}$ is disjoint from B, it follows that
$$\mathbb{P}(A \cup B) = \mathbb{P}\left(B \cup (A \cap B^c)\right) = \mathbb{P}(B) + \mathbb{P}\left(A \cap B^c\right).$$
Similarly,
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}\left(A^c \cap B\right).$$
Adding these together it follows that
$$2\mathbb{P}(A \cup B) = P(A) + P(B) + \mathbb{P}\left(A^c \cap B\right) + \mathbb{P}\left(A \cap B^c\right).$$

Noting that

$$A \cup B = (A \cap B) \cup (A^c \cap B) \cup (A \cap B^c)$$

where all individual intersections of the right-hand side are disjoint, it follows by finite additivity that

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

3. As $A = (A \cap B) \cup (A \cap B^c)$, $A \cap B = B$, and $\mathbb{P}$ is non-negative, it follows that

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) = \mathbb{P}(B) + \mathbb{P}(A \cap B^c) \geq \mathbb{P}(B).$$

4. Let $\tilde{A}_1 := A_1$, and $\tilde{A}_n = A_n \setminus \left(\bigcup_{k=1}^{n-1} A_k \right)$ for $n \geq 2$. Then $\left(\tilde{A}_n\right)_{n \in \mathbb{N}}$ are disjoint with

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \tilde{A}_n.$$

Moreover, by statement 3 we have $\mathbb{P}\left(\tilde{A}_n\right) \leq \mathbb{P}(A_n)$. Therefore, using σ-additivity we deduce that

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mathbb{P}\left(\bigcup_{n=1}^{\infty} \tilde{A}_n\right) \\ &= \sum_{n=1}^{\infty} \mathbb{P}\left(\tilde{A}_n\right) \\ &\leq \sum_{n=1}^{\infty} \mathbb{P}(A_n). \end{aligned}$$

■

Proposition 1.1.6 Let $\mathbb{P}$ be a finitely additive set function defined over an algebra $\mathscr{A}$, with $\mathbb{P}(\Omega) = 1$. Then the following statements are equivalent.

1. $\mathbb{P}$ is σ-additive, and hence a probability measure.
2. $\mathbb{P}$ is continuous from below. That is, for sets $A_1, A_2, \dots \in \mathscr{A}$ such that $A_1 \subseteq A_2 \subseteq \dots$ and $\bigcup_{n=1}^{\infty} A_n \in \mathscr{A}$, we have
$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \to \infty} \mathbb{P}(A_n).$$
3. $\mathbb{P}$ is continuous from above. That is, for sets $B_1, B_2, \dots \in \mathscr{A}$ such that $B_1 \supseteq B_2 \supseteq \dots$ and $\bigcap_{n=1}^{\infty} B_n \in \mathscr{A}$, we have
$$\lim_{n \to \infty} \mathbb{P}(B_n) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n\right).$$
4. $\mathbb{P}$ is continuous at $\varnothing$. That is, for sets $B_1, B_2, \dots \in \mathscr{A}$ such that $B_1 \supseteq B_2 \supseteq \dots$ and $\bigcap_{n=1}^{\infty} B_n = \varnothing$, we have
$$\lim_{n \to \infty} \mathbb{P}(B_n) = 0.$$

Proof. (1) $\Rightarrow$ (2). Consider the sets $\tilde{A}_1 = A_1$, and $\tilde{A}_n = A_n \setminus A_{n-1}$ for $n \geq 2$. Then the sets $\left(\tilde{A}_n\right)_{n \in \mathbb{N}}$ are a collection of disjoint sets. Moreover,

$$\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \tilde{A}_n.$$

Therefore,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} \tilde{A}_n\right)$$

$$
\begin{aligned}
&\overset{(1)}{=} \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{P}\left(\tilde{A}_k\right) \\
&= \lim_{n\to\infty} \left(\mathbb{P}\left(A_1\right) + \left(\mathbb{P}(A_2) - \mathbb{P}(A_1)\right) + \cdots + \left(\mathbb{P}(A_n) - \mathbb{P}(A_{n-1})\right)\right) \\
&= \lim_{n\to\infty} \mathbb{P}(A_n).
\end{aligned}
$$

$(2) \Rightarrow (3)$. Let $n \geq 1$ and consider the events $\tilde{B}_n = B_1 \setminus B_n$. Then

$$\mathbb{P}\left(\tilde{B}_n\right) = \mathbb{P}(B_1) - \mathbb{P}\left(B_n\right).$$

The sequence $\left(\tilde{B}_n\right)_{n\in\mathbb{N}}$ is an increasing sequence of events with

$$\bigcup_{n=1}^{\infty} \tilde{B}_n = B_1 \setminus \bigcap_{n=1}^{\infty} B_n.$$

By (2) it follows that

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} \tilde{B}_n\right) = \lim_{n\to\infty} \mathbb{P}\left(\tilde{B}_n\right).$$

Therefore,

$$
\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(B_n) &= \mathbb{P}(B_1) - \lim_{n\to\infty} \mathbb{P}\left(\tilde{B}_n\right) \\
&= \mathbb{P}(B_n) - \mathbb{P}\left(\bigcup_{n=1}^{\infty} \tilde{B}_n\right) \\
&= \mathbb{P}(B_1) - \mathbb{P}\left(B_1 \setminus \bigcap_{n=1}^{\infty} B_n\right) \\
&= \mathbb{P}(B_1) - \mathbb{P}(B_1) + \mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n\right) \\
&= \mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n\right).
\end{aligned}
$$

$(3) \Rightarrow (4)$. As $\varnothing \in \mathscr{A}$, it follows that

$$\lim_{n\to\infty} \mathbb{P}(B_n) \overset{(3)}{=} \mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n\right) = \mathbb{P}(\varnothing) = 0.$$

$(4) \Rightarrow (1)$. Let $A_1, A_2, \dots \in \mathscr{A}$ be pairwise disjoint with

$$A := \bigcup_{n=1}^{\infty} A_n \in \mathscr{A},$$

and

$$B_n := \bigcup_{i=n+1}^{\infty} A_i.$$

Note that $\bigcup_{i=1}^{n} A_i$ and B_n are disjoint sets such that

$$A = \bigcup_{i=1}^{n} A_i \cup B_n.$$

Therefore, by finite additivity we have

$$\mathbb{P}(A) = \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) + \mathbb{P}(B_n).$$

The sequence of sets $(B_n)_{n\in\mathbb{N}}$ is decreasing and such that $\bigcap_{n=1}^{\infty} B_n = \varnothing$. Therefore,

$$\begin{aligned}
\sum_{i=1}^{\infty} \mathbb{P}(A_i) &= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(A_i) \\
&= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \\
&= \lim_{n\to\infty} \left(\mathbb{P}\left(A\right) - \mathbb{P}(B_n)\right) \\
&= \mathbb{P}\left(A\right) - \lim_{n\to\infty} \mathbb{P}(B_n) \\
&\overset{(4)}{=} \mathbb{P}(A).
\end{aligned}$$

■

Proposition 1.1.7 Let μ be a finitely additive measure on an algebra $\mathscr{A}$ and let the sets $A_1, A_2, \dots \in \mathscr{A}$ be pairwise disjoint with $A = \bigcup_{i=1}^{\infty} A_i \in \mathscr{A}$. Then

$$\sum_{i=1}^{\infty} \mu(A_i) \leq \mu(A).$$

Proof. Note that $\bigcup_{i=1}^{n} A_i \subseteq A$ for any $n \in \mathbb{N}$. Therefore, by finite additivity and statement 3 of Proposition 1.1.5 we have

$$\mu(A) \geq \mu\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \mu(A_i).$$

Taking the limit as $n \to \infty$ preserves the inequality and so

$$\mu(A) \geq \sum_{i=1}^{\infty} \mu(A_i)$$

as required. ■

Example 1.1.8 Let Ω be a sample space. Then,

1. $\mathscr{F}_* = \{\varnothing, \Omega\}$, and
2. $\mathscr{F}^* = \{A : A \subseteq \Omega\} = 2^{\Omega}$

are σ-algebras.

Lemma 1.1.9 For any collection $\mathscr{E}$ of subsets of Ω there exists a minimal algebra $a(\mathscr{E})$ and a minimal σ-algebra $\sigma(\mathscr{E})$ that contains all elements of $\mathscr{E}$. Equivalently, $a(\mathscr{E})$ $(\sigma(\mathscr{E}))$ is the intersection of all algebras (σ-algebras) that contain $\mathscr{E}$.

Proof. Intersection, countable or uncountable, of algebras (σ-algebras) containing $\mathscr{E}$ is an algebra (σ-algebra) containing $\mathscr{E}$. ■

Remark 1.1.10 In the context of Lemma 1.1.9 we say that $\sigma(\mathscr{E})$ is generated by $\mathscr{E}$.

Exercise 1.1.11 Let $\mathscr{D} = \{D_1, D_2, \dots\}$ be a countable partition of Ω such that $D_i \cap D_j = \varnothing$ for $i \neq j$

and $\Omega = \bigcup_{j=1}^{\infty} D_j$. Show that

$$\sigma(\mathscr{D}) = \left\{ \bigcup_{j \in I} D_j : I \subseteq \mathbb{N} \right\}.$$

1.2 Measurable Spaces

Definition 1.2.1 A measurable space is a pair $(E, \mathscr{E})$, where E is a set and $\mathscr{E}$ is a σ-algebra on E.

Let $\mathbb{R} = (-\infty, \infty)$ be the real line and

$$(a, b] = \{x \in \mathbb{R} : a < x \leq b\},$$

for $-\infty \leq a < b < \infty$. Let $\mathscr{A}$ be the algebra of subsets of $\mathbb{R}$ such that $A \in \mathscr{A}$ if for some $n < \infty$ we have

$$A = \bigcup_{i=1}^{n} (a_i, b_i].$$

Let $\mathscr{B}(\mathbb{R})$ be the smallest σ-algebra containing $\mathscr{A}$. Then for $a < b$ we observe that

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

Thus, in addition to containing intervals of the form $(a, b]$, the Borel σ-algebra $\mathscr{B}(\mathbb{R})$ also contains singleton sets $\{a\}$ and intervals of the form

- (a, b),
- $[a, b]$,
- $[a, b)$,
- $(-\infty, b)$,
- $(-\infty, b]$, and
- (a, ∞).

Exercise 1.2.2 Show that $\mathscr{B}(\mathbb{R})$ can be generated by the collection of

1. open intervals of the form (a, b),
2. closed intervals $[a, b]$,
3. half intervals,
4. intervals of the form $(-\infty, a]$ or $[a, \infty)$,
5. open sets with respect to the Euclidean metric, or
6. closed sets with respect to the Euclidean metric.

Let $\mathbb{R}^n = \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{n}$. That is, the set of ordered n-tuples $x = (x_1, \ldots, x_n)$, where $x_k \in \mathbb{R}$ for $k = 1, \ldots, n$. A rectangle then refers to a set of the form

$$I = I_1 \times \cdots \times I_n = \{x \in \mathbb{R}^n : x_k \in I_k,\ k = 1, \ldots, n\}$$

where $I_k = (a_k, b_k]$ is known as a side of the rectangle. Let $\mathscr{I}$ be the collection of all rectangles I. The smallest σ-algebra $\sigma(\mathscr{I})$ generated by the system $\mathscr{I}$ is the Borel σ-algebra of subsets of $\mathbb{R}^n$ denoted $\mathscr{B}(\mathbb{R}^n)$.
Instead of the rectangles $I = I_1 \times \cdots \times I_n$ let us consider the rectangles $B = B_1 \times \cdots \times B_n$ with Borel sides. That is, B_k is a Borel subset of the real line that appears in the k^{th} place in the direct product $\mathbb{R} \times \cdots \times \mathbb{R}$. The smallest σ-algebra containing all rectangles with Borel sides is denoted by

$$\mathscr{B}(\mathbb{R}) \otimes \cdots \otimes \mathscr{B}(\mathbb{R})$$

and called the direct product of the σ-algebras $\mathscr{B}(\mathbb{R})$. In fact,

$$\mathscr{B}(\mathbb{R}^n) = \mathscr{B}(\mathbb{R}) \otimes \cdots \otimes \mathscr{B}(\mathbb{R}). \tag{1.1}$$

In other words, the σ-algebra generated by the rectangles $I = I_1 \times \cdots \times I_n$ coincides with the σ-algebra generated by rectangles $B = B_1 \times \cdots \times B_n$ with Borel sides. We will now justify this.

Lemma 1.2.3 Let $\mathscr{E}$ be a collection of subsets of Ω and let $B \subseteq \Omega$. Consider

$$\mathscr{E} \cap B = \{A \cap B : A \in \mathscr{E}\}.$$

Then

$$\sigma(\mathscr{E} \cap B) = \sigma(\mathscr{E}) \cap B,$$

where on the left-hand side the σ-algebra is over B, whereas on the right-hand side the σ-algebra is over Ω.

Proof. Step 1: Show that $\sigma(\mathscr{E} \cap B) \subseteq \sigma(\mathscr{E}) \cap B$.
Clearly, $B = \Omega \cap B \in \sigma(\mathscr{E}) \cap B$. Moreover, for any $A \in \sigma(\mathscr{E})$ we have

$$B \setminus (A \cap B) = (\Omega \setminus A) \cap B$$

and so $B \setminus (A \cap B) \in \sigma(\mathscr{E}) \cap B$ as $\Omega \setminus A \in \sigma(\mathscr{E})$. Note that for $A_1, A_2, \dots \in \sigma(\mathscr{E})$ we have

$$\bigcup_{i=1}^{\infty} (A_i \cap B) = \left(\bigcup_{i=1}^{\infty} A_i \right) \cap B \in \sigma(\mathscr{E}) \cap B,$$

where the membership follows as $\bigcup_{i=1}^{\infty} A_i \in \sigma(\mathscr{E})$. Therefore, $\sigma(\mathscr{E}) \cap B$ is a σ-algebra over B. Hence, as $\mathscr{E} \cap B \subseteq \sigma(\mathscr{E}) \cap B$ it follows that

$$\sigma(\mathscr{E} \cap B) \subseteq \sigma(\sigma(\mathscr{E}) \cap B) = \sigma(\mathscr{E}) \cap B.$$

Step 2: Show that $\sigma(\mathscr{E}) \cap B \subseteq \sigma(\mathscr{E} \cap B)$.
Let

$$\mathscr{G} := \{A : A \cap B \in \sigma(\mathscr{E} \cap B)\}.$$

Clearly, $\Omega \cap B \in \sigma(\mathscr{E} \cap B)$ and so $\Omega \in \mathscr{G}$. Let $A \in \mathscr{G}$, then note that

$$(\Omega \setminus A) \cap B = B \setminus (A \cap B).$$

By assumption $A \cap B \in \sigma(\mathscr{E} \cap B)$, which is a σ-algebra over B and so $B \setminus (A \cap B) \in \sigma(\mathscr{E} \cap B)$. Meaning $\Omega \setminus A \in \mathscr{G}$. Next let $A_1, A_2, \dots \in \mathscr{G}$, then as

$$\left(\bigcup_{i=1}^{\infty} A_i \right) \cap B = \bigcup_{i=1}^{\infty} (A_i \cap B)$$

it follows that $\bigcup_{i=1}^{\infty} A_i \in \mathscr{G}$. Therefore, $\mathscr{G}$ is a σ-algebra over Ω. Hence, as $\mathscr{E} \subseteq \mathscr{G}$ it follows that $\sigma(\mathscr{E}) \subseteq \mathscr{G}$. So for $A \in \sigma(\mathscr{E})$ we have $A \in \mathscr{G}$ which implies that $A \cap B \in \sigma(\mathscr{E} \cap B)$, and so $\sigma(\mathscr{E}) \cap B \subseteq \sigma(\mathscr{E} \cap B)$. ■

It is clear that for $n = 1$, the σ-algebras $\mathscr{B}(\mathbb{R}^n)$ and $\mathscr{B}(\mathbb{R}) \otimes \cdots \otimes \mathscr{B}(\mathbb{R})$ are the same.

Lemma 1.2.4 For $n = 2$, equation (1.1) holds.

Proof. ($\subseteq$). For any open set, A we can write

$$A \subseteq \bigcup_{x \in A \cap \mathbb{Q}^2} R(x, \tau(x))$$

where $R(x, \tau(x))$ is the open square centered at x and of side length $\tau(x)$. As $A \cap \mathbb{Q}^2$ is countable and $R(x, \tau(x)) \in \mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R})$, it follows that $A \in \mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R})$.
($\supseteq$). Suffices to check that $B_1 \times B_2 \in \mathscr{B}\left(\mathbb{R}^2\right)$ for any Borel sets B_1, B_2. Note that $B_1 \times \mathbb{R} \in \mathscr{B}\left(\mathbb{R}^2\right)$ since

$$B_1 \times \mathbb{R} \in \sigma(\{\text{open subsets of } \mathbb{R}\}) \times \mathbb{R} = \sigma(\{\text{open subsets of } \mathbb{R} \times \mathbb{R}\}).$$

Similarly, $\mathbb{R} \times B_2 \in \mathscr{B}(\mathbb{R}^2)$, and so $B_1 \times B_2 = (B_1 \times \mathbb{R}) \cap (\mathbb{R} \times B_2) \in \mathscr{B}(\mathbb{R}^2)$. ■

The case for any $n > 2$ can be discussed similarly to Lemma 1.2.4. The space $((\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$, on the other hand, requires a different approach. However, it is useful to outline this as $((\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$ is consistently utilised for constructing probabilistic models of experiments with infinitely many steps. Let $\mathbb{R}^\infty = \{x = (x_1, x_2, \dots), x_k \in \mathbb{R}\}$.

Definition 1.2.5 A set $C \subseteq \mathbb{R}^\infty$ is called cylindrical if it is of the form

$$C = \left\{x \in \mathbb{R}^\infty : (x_1, \dots, x_n) \in \tilde{C}_n\right\}$$

for some $n \geq 1$ and $\tilde{C}_n \in \mathscr{B}(\mathbb{R}^n)$.

Exercise 1.2.6 Show that the cylindrical sets form an algebra.

The σ-algebra generated by cylindrical sets is called the cylindrical σ-algebra and is denoted $\mathscr{B}(\mathbb{R}^\infty)$. One can verify that

$$\mathscr{B}(\mathbb{R}^\infty) = \sigma\left(\{A_1 \times A_2 \times \dots \subseteq \mathbb{R}^\infty,\ A_k \in \mathscr{B}(\mathbb{R})\}\right).$$

Example 1.2.7 For $c \in \mathbb{R}$, let

$$A = \left\{x \in \mathbb{R}^\infty : \limsup_{n\to\infty}(x_n) = \inf_{n\in\mathbb{N}} \sup_{k>n}(x_k) > c\right\}.$$

Then $A \in \mathscr{B}(\mathbb{R}^\infty)$ because

$$A = \bigcap_{n=1}^{\infty} \bigcup_{k=n+1}^{\infty} \{x \in \mathbb{R}^\infty : x_k > c\}.$$

Similarly, letting

$$B = \left\{x \in \mathbb{R}^\infty : \liminf_{n\to\infty}(x_n) = \sup_{n\in\mathbb{N}} \inf_{k>n}(x_k) > c\right\},$$

we have that $B \in \mathscr{B}(\mathbb{R}^\infty)$ because

$$B = \bigcap_{n=1}^{\infty} \bigcup_{k=n+1}^{\infty} \{x \in \mathbb{R}^\infty : x_k > c\}.$$

Exercise 1.2.8 For $c \in \mathbb{R}$, show that $D = \{x \in \mathbb{R}^\infty : \lim_{n\to\infty}(x_n) = c\} \in \mathscr{B}(\mathbb{R}^\infty)$.

1.3 Probability Distributions

Throughout this section, we will consider the importance of non-decreasing functions for describing probability measures on measurable spaces.

Lemma 1.3.1 A non-decreasing function $g(x)$ on $\mathbb{R}$ is continuous up to possibly countably many discontinuities of the first kind. That is, for $\epsilon \searrow 0$ the limits $g(x+\epsilon)$ and $g(x-\epsilon)$ exists but are distinct.

Proof. Note that it must be the case that $\lim_{\epsilon \searrow 0}(g(x+\epsilon) - g(x-\epsilon)) > 0$ due to the non-decreasing property of g. Thus we can form open balls at the right limit of the function at each jump. These open sets are distinct and contain at least one rational number. As the rational numbers are countable on $\mathbb{R}$, it follows that there are at most countably many such open balls, and hence countably many such jumps. ■

Lemma 1.3.1 is a positive result, as by construction our probability measures behave well in the countable domain. Moreover, we deduce that the derivative of a non-decreasing function, $g(x)$, denoted $g'(x)$ exists Lebesgue almost everywhere.

Exercise 1.3.2 Let $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mathbb{P})$ be a probability space and let $F(x) := \mathbb{P}((-\infty, x])$ for $x \in \mathbb{R}$. Show that,

- $F(x)$ is non-decreasing,
- $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$, and
- $F(x)$ is right-continuous for all $x \in \mathbb{R}$.

Definition 1.3.3 A function $F : \mathbb{R} \to [0,1]$ satisfying the properties of Exercise 1.3.2 is called a distribution function on $\mathbb{R}$.

Thus to every probability measure $\mathbb{P}$ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ there corresponds a distribution function. The opposite is also true and there exists a one-to-one correspondence between distribution functions and probability measures.

Theorem 1.3.4 Let $F = F(x)$ be a distribution function on $\mathbb{R}$. Then there exists a unique probability measure $\mathbb{P}$ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that

$$\mathbb{P}((a,b]) = F(b) - F(a),$$

for all $-\infty \le a < b < \infty$.

This relies on the following fundamental result in measure theory.

Theorem 1.3.5 — Caratheodory Theorem. Let μ_0 be a σ-additive (pre-)measure on $(\Omega, \mathscr{A})$, where $\mathscr{A}$ is an algebra of subsets of Ω. Then there exists a measure μ on $(\Omega, \sigma(\mathscr{A}))$, such that

$$\mu(A) = \mu_0(A)$$

for all $A \in \mathscr{A}$. If μ_0 is additionally σ-finite, then the measure μ is unique.

Definition 1.3.6 A measure μ on a σ-algebra Σ on Ω is complete if any subset of a set of measure zero (null sets) is measurable. That is, if $A \in \Sigma$ is such that $\mu(A) = 0$, then for any $B \subseteq A$ we have that $B \in \Sigma$ and $\mu(B) = 0$.

Requiring completeness helps avoid any caveats in proving results relating to measures. The space $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mathbb{P})$ with $\mathbb{P}$ constructed from Theorem 1.3.4 is not complete as there are subsets of Borel sets that are not themselves Borel sets. Fortunately, one can enlarge the σ-algebra to include null sets. A measure μ on Σ can be completed by extending Σ to

$$\overline{\Sigma} = \sigma(\Sigma \cup \{B \in \Omega : B \subseteq A \in \Sigma,\ \mu(A) = 0\}),$$

with $\mu(B) = 0$ for any B a subset of a null set. The completion of the measure obtained in Theorem 1.3.4

is called the Lebesgue-Stiltjes measure. In particular, the distribution function $F(x) = x$ corresponds to the Lebesgue measure on $\mathbb{R}$.

1.4 Measures on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$

1.4.1 Discrete Measures

A discrete measure $\mathbb{P}_{\text{disc}}$ has a piece-wise constant distribution $F_{\text{disc}} = F(x)$. The function F has jumps at the points $x_1, x_2, \ldots$, that is $\Delta F(x_i) > 0$, where $\Delta F(x) = F(x) - F(x^-)$.

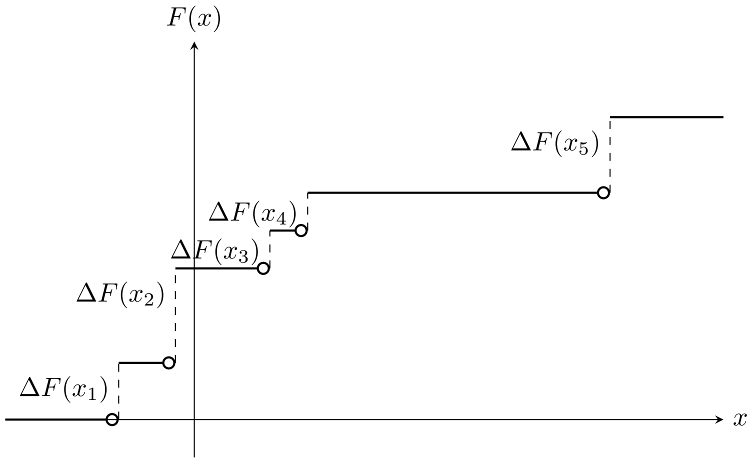


Figure 1: CDF of a discrete measure.

A discrete measure is concentrated at the points $x_1, x_2, \ldots$, known as atoms. Letting

$$p_k := \mathbb{P}(\{x_k\}) = \Delta F(x_k) > 0,$$

it follows by properties of a distribution function that

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

In particular,

$$F_{\text{disc}}(x) = \sum_{x_k \le x} p_k,$$

and for $A \subseteq \mathbb{N}$ we have

$$\mathbb{P}(A) = \sum_{k \in A} p_k.$$

We refer to $(p_1, p_2, \ldots)$ as the discrete probability distribution.

Example 1.4.1

- The Discrete Uniform distribution for $n \in \mathbb{N}$ has $p_k = \frac{1}{n}$ for $k = 1, \ldots, n$.
- The Bernoulli distribution for $\mathsf{B}(1, p)$ has $p_1 = p$ and $p_2 = 1 - p$ for $0 \le p \le 1$.
- The Binomial distribution $\mathsf{B}(n, p)$ has $p_k = \binom{n}{k} p^k (1-p)^k$ for $k = 0, \ldots, n$ and $0 \le p \le 1$.
- The Poisson distribution $\mathsf{Po}(\lambda)$ has $p_k = e^{-\lambda} \frac{\lambda^k}{k!}$ for $\lambda > 0$ and $k = 0, 1, \ldots$.

1.4.2 Absolutely Continuous Measures

Proposition 1.4.2 Let f be an integrable non-negative function such that

$$F(x) = F_{\text{ac}}(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$$

with respect to the Lebesgue measure. Then the set function $\mathbb{P}_{\text{ac}}(A) = \int_A f(t)\,\mathrm{d}t$ for $A \in \mathscr{F}$ is a measure. In particular, f is a density of $\mathbb{P}_{\text{ac}}$.

Proof. For half-open intervals let $\mathbb{P}_{\text{ac}}((a,b]) = \int_a^b f(t)\,\mathrm{d}t$. Then use Theorem 1.3.5 to extend the measure to the σ-algebra. ∎

The measure $\mathbb{P}_{\text{ac}}$ is absolutely continuous with respect to the Lebesgue measure μ. In the sense that if $\mu(A) = 0$ then $\mathbb{P}_{\text{ac}}(A) = 0$.

Theorem 1.4.3 — Radon-Nikodym. If $\mathbb{P}$ is a measure such that $\mu(A) = 0$ implies $\mathbb{P}(A) = 0$, then $\mathbb{P}$ has a density.

Note that there is a connection between the absolute continuity of measures and the absolute continuity of functions. If $\mathbb{P}$ is an absolutely continuous measure then $F_{\text{ac}}(x)$ is an absolutely continuous function, with $F'_{\text{ac}}(x) = f(x)$ almost everywhere.

Example 1.4.4

- The Uniform distribution on $[a,b]$ has density

$$f(x) = \frac{1}{b-a}.$$

- The Normal or Gaussian distribution on $\mathbb{R}$ has density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x-m)^2}{2\sigma^2}\right)$$

  for $m \in \mathbb{R}$ and $\sigma > 0$.

- The Gamma distribution on $[0,\infty)$ has

$$f(x) = \frac{x^{\alpha-1}\exp\left(-\frac{x}{\beta}\right)}{\Gamma(\alpha)\beta^\alpha}$$

  for $a, \beta > 0$.

1.4.3 Singular Continuous Measures

Definition 1.4.5 A measure ν is concentrated on a measurable set A if $\nu(E) = 0$ for any $E \subseteq \mathbb{R} \setminus A$.

Singular continuous measures are those whose distribution functions are continuous but have all their points of increase on sets of zero Lebesgue measure. More specifically, $F(x) = F_{\text{sc}}(x)$ is continuous at any x and $\mathbb{P}_{\text{sc}}$ is concentrated on a set of Lebesgue measure zero. In particular, the distribution has no atoms. For x in this set, $F'_{\text{sc}}(x) \neq 0$ or does not exist. Thus $F'_{\text{sc}}(x) = 0$ almost everywhere and by continuity we have that $\mathbb{P}_{\text{sc}}(\{x\}) = 0$ for each point $x \in \mathbb{R}$.

Example 1.4.6 Consider the interval $[0,1]$ and construct $F(x)$ by the following procedure formulated by Cantor. Divide $[0,1]$ into thirds and let

$$F_1(x) = \begin{cases} \frac{3}{2}x & x \in \left[0, \frac{1}{3}\right] \\ \frac{1}{2} & x \in \left[\frac{1}{3}, \frac{2}{3}\right] \\ \frac{3}{2}x - \frac{1}{2} & x \in \left[\frac{2}{3}, 1\right]. \end{cases}$$

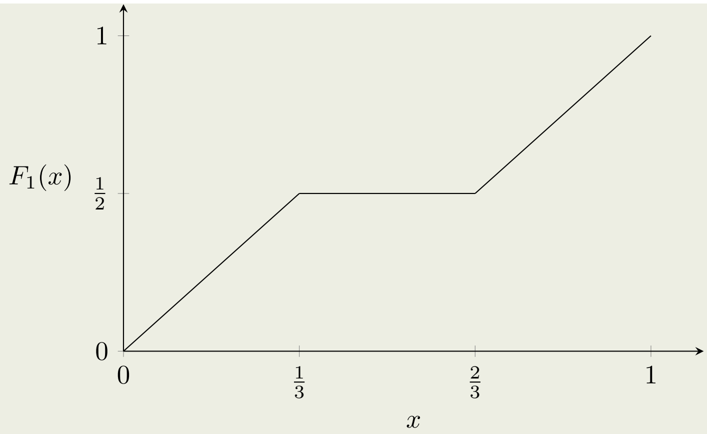


Figure 2: First step of constructing the devil staircase.

Next divide the intervals $[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$ into thirds and let

$$F_2(x) = \begin{cases} 0 & x = 0 \\ \frac{1}{2} & x \in \left(\frac{1}{3}, \frac{2}{3}\right) \\ \frac{1}{4} & x \in \left(\frac{1}{9}, \frac{2}{9}\right) \\ \frac{3}{4} & x \in \left(\frac{7}{9}, \frac{8}{9}\right) \\ 1 & x = 1, \end{cases}$$

defining it in the intermediate intervals by linear interpolation.

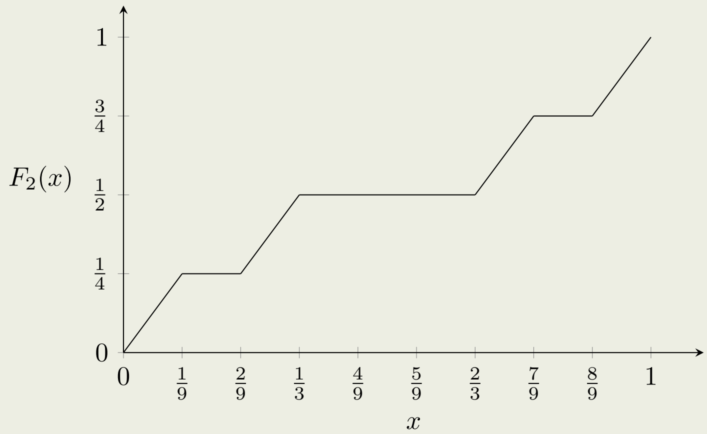


Figure 3: Second step of constructing the Cantor's construction.

Continuing, we construct a sequence of functions $F_n(x)$ for $n = 1, 2, \ldots$ which converge to a non-decreasing continuous function $F(x)$, named the Cantor function, whose points of increase form a set of Lebesgue measure zero. From the construction of $F(x)$ we see that the total length of the intervals $\left(\frac{1}{3}, \frac{2}{3}\right), \left(\frac{1}{9}, \frac{2}{9}\right), \left(\frac{7}{9}, \frac{8}{9}\right), \ldots$ on which the function is constant is

$$\frac{1}{3} + \frac{2}{9} + \frac{4}{27} + \cdots = 1. \tag{1.2}$$

Let N be the set of points of increase of the Cantor function $F(x)$. It follows from (1.2) that $\mathsf{Leb}(N) = 0$. At the same time, if μ is the measure corresponding to the Cantor function $F(x)$, we have $\mu(N) = 1$. That is, μ singular with respect to the Lebesgue measure Leb.

Theorem 1.4.7 — Hahn decomposition. Any probability distribution has a representation of the form

$$F(x) = a_1 F_{\text{disc}}(x) + a_2 F_{\text{ac}}(x) + a_3 F_{\text{sc}}(x)$$

for $a_1 + a_2 + a_3 = 1$.

1.5 Measures on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$

Distribution functions on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ are defined similarly. For example, when $n = 2$ we have

$$F(x, y) = \mathbb{P}((-\infty, x] \times (-\infty, y]).$$

For probability spaces $(\Omega_1, \mathscr{F}_1, \mathbb{P}_1)$, $(\Omega_2, \mathscr{F}_2, \mathbb{P}_2)$ the product measure on $(\Omega_1 \times \Omega_2, \mathscr{F}_1 \otimes \mathscr{F}_2)$ is defined as follows.

1. Set
$$\mathbb{P}_0(A_1 \times A_2) = \mathbb{P}_1(A_1)\mathbb{P}_2(A_2)$$
for $A_1 \in \mathscr{F}_1$ and $A_2 \in \mathscr{F}_2$.
2. Then extend $\mathbb{P}_0$ to the algebra generated by $A_1 \times A_2$, and show that $\mathbb{P}_0$ is a σ-additive measure on this algebra.
3. Apply Theorem 1.3.5 to obtain the extension.

This extension is the product measure and is denoted $\mathbb{P}_1 \otimes \mathbb{P}_2$.

1.6 Measures on $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$

On $\mathbb{R}^n$ for $n \geq 1$, probability measures were constructed in the following way.

1. It was first defined for elementary sets of the form $(a, b]$.
2. The definition was then extended to sets of the form $A = \sum_{i=1}^n (a_i, b_i]$.
3. The extension to sets in $\mathscr{B}(\mathbb{R}^n)$ was provided by Theorem 1.3.5.

A similar procedure of constructing probability measures also works for the space $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$. Let

$$\mathscr{I}_n(B) = \{x \in \mathbb{R}^\infty : (x_1, \ldots, x_n) \in B\}$$

denote the cylinder set in $\mathbb{R}^\infty$ with base $B \in \mathscr{B}(\mathbb{R}^n)$. It is natural to take the cylinder sets as the elementary sets in $\mathbb{R}^\infty$ whose probabilities enable us to determine the probability measure on the sets of $\mathscr{B}(\mathbb{R}^\infty)$.

Definition 1.6.1 A sequence $(\mathbb{P}_n)_{n\in\mathbb{N}}$ of probability measures, where $\mathbb{P}_n$ is a probability measure on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$, is consistent if for all $n = 1, 2, \ldots$ and $B \in \mathscr{B}(\mathbb{R}^n)$ we have

$$\mathbb{P}_{n+1}(B \times \mathbb{R}) = \mathbb{P}_n(B).$$

Theorem 1.6.2 — Kolmogorov Extension Theorem. For any consistent sequence $(\mathbb{P}_n)_{n\in\mathbb{N}}$, there exists a unique probability measure $\mathbb{P}$ on $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$ such that

$$\mathbb{P}(\mathscr{I}_n(B)) = \mathbb{P}_n(B),$$

for $B \in \mathscr{B}(\mathbb{R}^n)$ and $n \in \mathbb{N}$.

1.7 Random Variables

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space.

Definition 1.7.1 A real function $\xi : \Omega \to \mathbb{R}$ is an $\mathscr{F}$-measurable function, or a random variable if

$$\xi^{-1}(B) = \{\omega : \xi(\omega) \in B\} \in \mathscr{F}$$

for every $B \in \mathscr{B}(\mathbb{R})$. Equivalently, $\xi^{-1}(B)$ is a measurable set in Ω. When $(\Omega, \mathscr{F}) = (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$, the $\mathscr{B}(\mathbb{R}^n)$-measurable functions are called Borel functions.

Random variables are used to summarise the abstract outcomes $\omega \in \Omega$ with a real number or vector.

Exercise 1.7.2 The experiment of throwing two independent fair six-faced dice can be represented by the probability space $(\Omega_1, \mathscr{F}_1, \mathbb{P}_1) \otimes (\Omega_2, \mathscr{F}_2, \mathbb{P}_2)$, where for $i = 1, 2$, $\Omega_i = \{1, 2, 3, 4, 5, 6\}$ is the outcome from dice i, $\mathscr{F}_i = 2^{\Omega_i}$ and $\mathbb{P}_i(\{j\}) \equiv \frac{1}{6}$ for each $j \in \{1, 2, 3, 4, 5, 6\}$. The function $X : \Omega_1 \times \Omega_2 \to \mathbb{R}$ given by

$$X(\omega_1, \omega_2) = \omega_1 + \omega_2,$$

summarises the outcome of the dice by their sum.

1. What is the range of X?
2. Check that X is a measurable function by determining its possible pre-images.

Lemma 1.7.3 Let $\mathscr{D}$ be a collection of subsets on $\mathbb{R}$ such that $\sigma(\mathscr{D}) = \mathscr{B}(\mathbb{R})$. A necessary and sufficient condition that a function $\xi = \xi(\omega)$ is a random variable is that

$$\xi^{-1}(D) = \{\omega : \xi(\omega) \in D\} \in \mathscr{F}$$

for all $D \in \mathscr{D}$.

Proof. Let

$$\mathscr{G} = \left\{B : \xi^{-1}(B) \in \mathscr{F}\right\} \subseteq \mathscr{B}(\mathbb{R}).$$

Clearly, $\varnothing \in \mathscr{G}$ as $\xi^{-1}(\varnothing) = \varnothing \in \mathscr{F}$. For $B \in \mathscr{G}$ it follows that

$$\xi^{-1}\left(B^c\right) = \left(\xi^{-1}(B)\right)^c \in \mathscr{F}$$

and so $B^c \in \mathscr{G}$. Moreover, for $B_1, B_2, \dots \in \mathscr{G}$ we have

$$\xi^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right) = \bigcup_{i=1}^{\infty} \xi^{-1}(B_i) \in \mathscr{F}$$

and so $\bigcup_{i=1}^{\infty} B_i \in \mathscr{G}$ which implies that $\mathscr{G}$ is a σ-algebra. Note that $\mathscr{D} \subseteq \mathscr{G}$ with $\sigma(\mathscr{D}) = \mathscr{B}(\mathbb{R})$ so that

$$\mathscr{B}(\mathbb{R}) = \sigma(\mathscr{D}) \subseteq \sigma(\mathscr{G}) = \mathscr{G} \subseteq \mathscr{B}(\mathbb{R}),$$

therefore, $\mathscr{G} = \mathscr{B}(\mathbb{R})$ meaning ξ is a random variable. Conversely, if ξ is a random variable then $\mathscr{G} = \mathscr{B}(\mathbb{R})$, which implies that $\mathscr{D} \subseteq \mathscr{G}$. Hence, $\xi^{-1}(D) \in \mathscr{F}$ for all $D \in \mathscr{D}$. ■

Corollary 1.7.4 A necessary and sufficient condition for $\xi = \xi(\omega)$ to be a random variable is that

$$\{\omega : \xi(\omega) < x\} \in \mathscr{F}$$

for every $x \in \mathbb{R}$, or that

$$\{\omega : \xi(\omega) \leq x\} \in \mathscr{F}$$

for every $x \in \mathbb{R}$.

Proof. These follow from Lemma 1.7.3 as the collections

$$\mathscr{D}_1 := \{(-\infty, x) : x \in \mathbb{R}\}$$

and

$$\mathscr{D}_2 := \{(-\infty, x] : x \in \mathbb{R}\}$$

are such that $\sigma(\mathscr{D}_1) = \mathscr{B}(\mathbb{R})$ and $\sigma(\mathscr{D}_2) = \mathscr{B}(\mathbb{R})$. ■

Exercise 1.7.5 Let ξ be a random variable. Show that

$$\mathscr{F}_\xi := \left\{\xi^{-1}(B) : B \in \mathscr{B}(\mathbb{R})\right\} \subseteq \mathscr{F}$$

is a σ-algebra.

Lemma 1.7.6 Let $\varphi = \varphi(x)$ be a Borel function and $\xi = \xi(\omega)$ a random variable. Then the composition $\eta = \varphi \circ \xi$ is also a random variable. In particular, η is $\mathscr{F}_\xi$- measurable.

Proof. For $B \in \mathscr{B}(\mathbb{R})$ we have

$$\begin{aligned}\{\omega : \eta(\omega) \in B\} &= \{\omega : \varphi(\xi(\omega)) \in B\} \\ &= \left\{\omega : \xi(\omega) \in \varphi^{-1}(B)\right\} \\ &\in \mathscr{F}_\xi \subseteq \mathscr{F},\end{aligned}$$

where the membership to $\mathscr{F}_\xi$ follows by the fact that $\varphi^{-1}(B) \in \mathscr{B}(\mathbb{R})$. Thus, we conclude that η is $\mathscr{F}$-measurable, meaning it is a random variable. ■

Example 1.7.7 If ξ is a random variable and $f : \mathbb{R} \to \mathbb{R}$ is a continuous function then $f(\xi)$ is a random variable. Therefore, if ξ is a random variable then the following are also random variables.

1. ξ^n.
2. $\xi^+ := \max(\xi, 0)$.
3. $\xi^- := \max(-\xi, 0)$.
4. $|\xi|$.

Lemma 1.7.8 Let ξ and η be random variables. Then, when the operations are well-defined, the following are random variables.

1. $\xi + \eta$.
2. $\xi - \eta$.
3. $\xi\eta$.
4. $\frac{\xi}{\eta}$.
5. $\max(\xi, \eta)$
6. $\min(\xi, \eta)$.

Proof.

1. Let $x \in \mathbb{R}$. Suppose $r, s \in \mathbb{Q}$ are such that $r + s < x$. If $\omega \in \Omega$ is such that $\xi(\omega) < r$ and $\eta(\omega) < s$ then $\xi(\omega) + \eta(\omega) < x$. On the other hand, if $\xi(\omega) + \eta(\omega) < x$, then there exists some $q \in \mathbb{Q}$ such that $\xi(\omega) + \eta(\omega) < q < x$. In particular, there exists $r, s \in \mathbb{Q}$ such that $r + s = q$ where $\xi(\omega) < r$ and $\eta(\omega) < s$. Therefore,

$$\{\omega : \xi(\omega) + \eta(\omega) < x\} = \bigcup_{r+s<x,(r,s)\in\mathbb{Q}^2} \{\omega : \xi(\omega) < r\} \cap \{\omega : \eta(\omega) < s\}.$$

   As $\mathbb{Q}^2$ is countable and ξ and η are random variables, it follows that $\{\omega : \xi(\omega) + \eta(\omega) < x\} \in \mathscr{F}$. Thus, we conclude that $\xi + \eta$ is a random variable using Corollary 1.7.4.

2. This follows from statement 1 by noting that $-\eta$ is a random variable.

3. Note that
$$\xi\eta = \frac{1}{2}\left((\xi+\eta)^2 - \xi^2 - \eta^2\right).$$
Thus, we conclude that $\xi\eta$ is a random variable as all the terms on the right-hand side are random variables.

4. This follows from statement 3 by noting that $\frac{1}{\eta}$ is a random variable, provided it is well-defined.

5. Note that
$$\max(\xi,\eta) = \frac{1}{2}\left(|\xi-\eta| + \xi + \eta\right).$$
As the right-hand side is a random variable, $\max(\xi,\eta)$ is a random variable.

6. Noting $\min(\xi,\eta) = \max(-\xi,-\eta)$, we deduce that $\min(\xi,\eta)$ is a random variable by using statement 5.

■

Example 1.7.9 An alternative way of arriving at statement 4 of Example 1.7.7 is by using statement 1 from Lemma 1.7.8 with statements 2 and 3 from Example 1.7.7.

Lemma 1.7.10 Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of random variables. Provided they exist, the following are also random variables.

1. $\sup_{n\in\mathbb{N}}(\xi_n)$.
2. $\inf_{n\in\mathbb{N}}(\xi_n)$.
3. $\lim_{n\to\infty}(\xi_n)$.

Proof.

1. For $x \in \mathbb{R}$ we have
$$\left\{\omega : \sup_{n\in\mathbb{N}}(\xi_n(\omega)) > x\right\} = \bigcup_{n=1}^{\infty}\{\omega : \xi_n(\omega) > x\}.$$
The right-hand side is a countable union of sets in $\mathscr{F}$, and so $\{\sup_{n\in\mathbb{N}}(\xi_n) > x\} \in \mathscr{F}$. Therefore, $\sup_{n\in\mathbb{N}}(\xi_n)$ is a random variable by Corollary 1.7.4.

2. As $\inf_{n\in\mathbb{N}}(\xi_n) = -\sup_{n\in\mathbb{N}}(-\xi_n)$, by statement 1 it follows that $\inf_{n\in\mathbb{N}}(\xi_n)$ is a random variable.

3. If $\lim_{n\in\mathbb{N}}(\xi_n)$ exists then $\liminf_{n\to\infty}(\xi_n) = \lim_{n\to\infty}(\xi_n)$. As $\liminf_{n\to\infty} = \sup_{n\in\mathbb{N}}\inf_{k\geq n}(\xi_n)$ we can use statements 1 and 2 to deduce that $\lim_{n\to\infty}(\xi_n)$ is a random variable.

■

Definition 1.7.11 A random variable ξ is called simple if
$$\xi(\omega) = \sum_{j=1}^{n} x_j \chi_{D_j}(\omega)$$
for some $n \geq 1$, with $D_1, \ldots, D_n$ being a partition of Ω consisting of measurable sets and
$$\chi_D(\omega) = \begin{cases} 1, & \omega \in D \\ 0, & \text{otherwise.} \end{cases}$$

Lemma 1.7.12

1. For every random variable $\xi = \xi(\omega)$ there is a sequence of simple random variables $(\xi_n)_{n\in\mathbb{N}}$, such that $|\xi_n| \leq |\xi|$ and $\xi_n(\omega) \to \xi(\omega)$ as $n \to \infty$, for all $\omega \in \Omega$.

2. For any random variable $\xi(\omega) \geq 0$ there exists a pointwise non-decreasing sequence of simple random variables $\xi_1(\omega) \leq \xi_2(\omega) \leq \cdots \leq \xi(\omega)$ such that

$$\lim_{n\to\infty} \xi_n(\omega) = \xi(\omega)$$

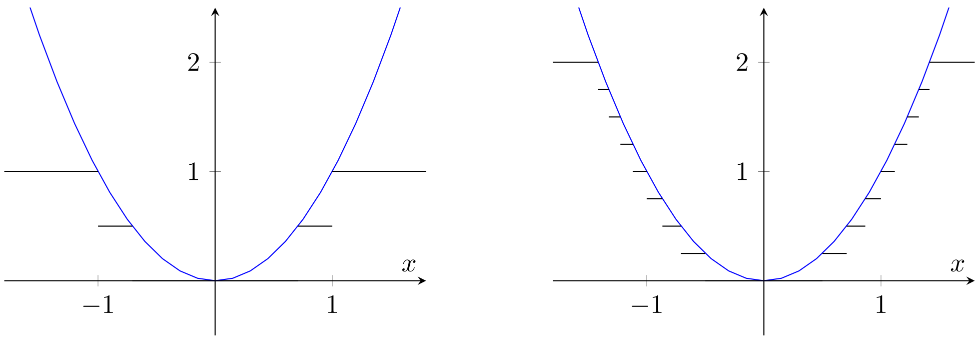
for all $\omega \in \Omega$. Such a sequence is usually denoted $\xi_n \nearrow \xi$.

Proof. Consider statement 2. For $n = 1, 2, \ldots$, let

$$\xi_n(\omega) = \sum_{j=0}^{n2^n-1} \frac{j}{2^n} \chi_{\left\{\omega : \frac{j}{2^n} \leq \xi(\omega) < \frac{j+1}{2^n}\right\}} + n\chi_{\{\omega:\xi(\omega)\geq n\}}.$$

Then the sequence $\xi_n(\omega)$ is such that $\xi_n \nearrow \xi$ for all $\omega \in \Omega$. For statement 1, we observe that ξ can be represented in the form $\xi = \xi^+ - \xi^-$, where $\xi^+ = \max(\xi, 0)$ and $\xi^- = \max(-\xi, 0)$. ■

One can explicitly build a simple function approximation to the function $f(x) = x^2$ for all $x \in \mathbb{R}$.



(a) $n = 1$ **(b)** $n = 2$

Figure 4: Approximation of x^2.

A common strategy to show results relating to random variables is the four-step proof.

1. Prove the statement for indicator functions.
2. Extend the statement for simple random variables by using linearity.
3. Extend the statement to non-negative random variables by taking limits.
4. Extend the statement to arbitrary random variables by considering their positive and negative parts.

Lemma 1.7.13 Consider a measurable space $(\Omega, \mathscr{F})$ and a disjoint decomposition $\mathscr{D} = \{D_1, D_2, \ldots\}$ of Ω. Let $\xi = \xi(\omega)$ be a $\sigma(\mathscr{D})$-measurable random variable. Then ξ is representable in the form

$$\xi(\omega) = \sum_{k=1}^{\infty} \alpha_k \chi_{D_k}(\omega),$$

where $\alpha_k \in \mathbb{R}$. That is, $\xi(\omega)$ is constant on the elements D_k of the decomposition.

Proof. Suppose that $\xi = \chi_D$ for $D \in \sigma(\mathscr{D})$. Then using Exercise 1.1.11 we know that $D = \bigcup_{k\in I} D_k$ for some $I \subseteq \mathbb{N}$. Therefore,

$$\xi = \sum_{k=1}^{\infty} \alpha_k \chi_{D_k}$$

with

$$\alpha_k = \begin{cases} 1 & k \in I \\ 0 & \text{otherwise.} \end{cases}$$

Using linearity this is extended to ξ a simple function. When ξ is a non-negative random variable there exists a sequence of simple functions $(\xi_n)_{n\in\mathbb{N}}$ such that $\xi_n \nearrow \xi$. As ξ is a random variable its image does not include infinity. Letting

$$\xi_n = \sum_{k=1}^{\infty} \alpha_k^{(n)} \chi_{D_k}(\omega)$$

it follows that $\alpha_k := \lim_{n\to\infty} \left(\alpha_k^{(n)}\right)$ is well-defined. As each ξ_n is a non-negative random variable, we can apply the monotone convergence to deduce that

$$\begin{aligned}
\xi &= \lim_{n\to\infty} (\xi_n) \\
&= \lim_{n\to\infty} \left(\sum_{k=1}^{\infty} \alpha_k^{(n)} \chi_{D_k} \right) \\
&= \sum_{k=1}^{\infty} \left(\lim_{n\to\infty} \alpha_k^{(n)} \right) \chi_{D_k} \\
&= \sum_{k=1}^{\infty} \alpha_k \chi_{D_k}.
\end{aligned}$$

For ξ a general random variable, we can extend the result by using the decomposition $\xi = \xi^+ - \xi^-$. ■

1.8 Distributions of random variables

Definition 1.8.1 The probability distribution of a random variable ξ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is $\mathbb{P}_\xi$ where

$$\mathbb{P}_\xi(B) = \mathbb{P}\left(\{\omega : \xi(\omega) \in B\}\right)$$

for $B \in \mathscr{B}(\mathbb{R})$.

Definition 1.8.2 The function

$$F_\xi(x) := \mathbb{P}_\xi((-\infty, x]) = \mathbb{P}(\{\omega : \xi(\omega) \le x\}),$$

for $x \in \mathbb{R}$ is the distribution function of ξ.

Example 1.8.3 From Exercise 1.7.2, one can verify that the probability distribution of X on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ satisfies

$$\mathbb{P}_X(\{j\}) = \frac{6 - |7-j|}{36},$$

for $j \in \{2, 3, \ldots, 12\}$. Using this, we can extend $\mathbb{P}_X$ to other sets in $\mathscr{B}(\mathbb{R})$.

Notice that there are multiple random variables, on potentially different probability spaces, which give the same distribution function. Indeed, we can always construct a probability measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ given a distribution function F_ξ. Therefore, for any random variable ξ on a probability $(\Omega, \mathscr{F}, \mathbb{P})$, the identity random variable $I(\omega) = \omega$ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mathbb{P}_\xi)$ has the same distribution as ξ. For example, consider the space $\left(\Omega, 2^\Omega, Q\right)$, with $\Omega = \{2, 3, \ldots, 12\}$ and $Q(\{j\}) = \mathbb{P}_X(\{j\})$. Then the random variable $\xi(j) = j$ for all $j \in \Omega \subseteq \mathbb{R}$ is such that $Q_\xi(A) \equiv \mathbb{P}_X(A)$ for all $A \in \mathscr{B}(\mathbb{R})$.

Definition 1.8.4 The vector function

$$\xi = (\xi_1, \ldots, \xi_n) : \Omega \to \mathbb{R}^n$$

is a random vector if for any $B \subseteq \mathscr{B}(\mathbb{R}^n)$, we have $\xi^{-1}(B) \in \mathscr{F}$. As before, we construct $\mathbb{P}_\xi$ and say that $\mathbb{P}_\xi = \mathbb{P}_{(\xi_1,\dots,\xi_n)}$ is the joint distribution of $\xi_1, \dots, \xi_n$ given by

$$F_\xi(x_1, \dots, x_n) = \mathbb{P}(\xi_1 \le x_1, \dots, \xi_n \le x_n).$$

Exercise 1.8.5 Show that the vector $\xi = (\xi_1, \dots, \xi_n)$ is a random variable if and only if $\xi_1, \dots, \xi_n$ are random variables.

For $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$ we can similarly define random sequences $\xi = (\xi_1, \xi_2, \dots)$.

1.9 Solution to Exercises

Exercise 1.1.11

Solution. Let $\mathscr{C} = \left\{\bigcup_{i\in I} D_i : I \subseteq \mathbb{N}\right\}$.

- As $\Omega = \bigcup_{i\in\mathbb{N}} D_i$ we have that $\Omega \in \mathscr{C}$.
- For $A \in \mathscr{C}$, there exists $I \subseteq \mathbb{N}$ such that $A = \bigcup_{i\in I} D_i$. Let $I' = \mathbb{N} \setminus I$ and let $B = \bigcup_{i\in I'} D_i \in \mathscr{C}$. As $\mathscr{D}$ forms a partition, $B = A^c$ and so $A^c \in \mathscr{C}$.
- Let $A_1, A_2, \dots \in \mathscr{C}$. Then each A_i is a countable union of elements from $\mathscr{D}$. Therefore, $A = \bigcup_{i\in\mathbb{N}} A_i$ is also a countable union of elements from $\mathscr{D}$ and so $A \in \mathscr{C}$.

The above show that $\mathscr{C}$ is a σ-algebra. Therefore, as $\mathscr{D} \subseteq \mathscr{C}$ and $\mathscr{C} \subseteq \sigma(\mathscr{D})$ it follows that

$$\sigma(\mathscr{D}) \subseteq \sigma(\mathscr{C}) = \mathscr{C} \subseteq \sigma(\mathscr{D}).$$

Hence, $\sigma(\mathscr{D}) = \mathscr{C}$. ■

Exercise 1.2.2

Solution.

1. As $(a, b] = \bigcap_{n=1}^{\infty} \left(a, b + \frac{1}{n}\right)$ it follows that the open intervals also generate $\mathscr{B}(\mathbb{R})$.
2. As $(a, b] = \bigcap_{n=1}^{\infty} \left[a + \frac{1}{n}, b\right]$ it follows that the closed intervals generate $\mathscr{B}(\mathbb{R})$.
3. As $(a, b] = \bigcap_{n=1}^{\infty} \left[a + \frac{1}{n}, b + \frac{1}{n}\right)$ it follows that the right-open half intervals generate $\mathscr{B}(\mathbb{R})$.
4. As $(a, b] = (-\infty, a]^c \cap (-\infty, b]$ it follows that intervals of the form $(-\infty, a]$ or $[a, \infty)$ generate $\mathscr{B}(\mathbb{R})$.
5. As (a, b) is open it follows from statement 1 that open sets generate $\mathscr{B}(\mathbb{R})$.
6. As $[a, b]$ are closed sets it follows from statement 2 that closed sets generate $\mathscr{B}(\mathbb{R})$.

■

Exercise 1.2.6

Solution. Let $\mathscr{C}$ be the collection of cylindrical sets.

- Note $\{x \in \mathbb{R}^\infty : (x_1) \in \mathbb{R}\} = \mathbb{R}^\infty$ and $\{x \in \mathbb{R}^\infty : (x_1) \in \emptyset\} = \emptyset$. Hence, $\emptyset, \mathbb{R}^\infty \in \mathscr{C}$.
- Let $C = \left\{x \in \mathbb{R}^\infty : (x_1, \dots, x_n) \in \tilde{C}_n\right\} \in \mathscr{C}$ with $\tilde{C}_n \in \mathscr{B}(\mathbb{R}^n)$. Then as $\left(\tilde{C}_n\right)^c \in \mathscr{B}(\mathbb{R}^n)$ it follows that $C^c = \left\{x \in \mathbb{R}^\infty : (x_1, \dots, x_n) \in \left(\tilde{C}_n\right)^c\right\} \in \mathscr{C}$.
- Consider $C_1, \dots, C_m \in \mathscr{C}$ where $C_i = \left\{x \in \mathbb{R}^\infty : (x_1, \dots, x_{n_i}) \in \tilde{C}_{n_i}\right\}$ for $\tilde{C}_{n_i} \in \mathscr{B}(\mathbb{R}^{n_i})$. Let $n = \max_{i=1,\dots,m}(n_i)$, then for each $\tilde{C}_{n_i}$ we can let $\tilde{D}_i = \tilde{C}_{n_i} \times \mathbb{R}^{n-n_i} \in \mathscr{B}(\mathbb{R}^n)$. We note that

  $$C_i = \left\{x \in \mathbb{R}^\infty : (x_1, \dots, x_{n_i}, \dots, x_n) \in \tilde{D}_i\right\}.$$

  Moreover,

  $$\bigcup_{i=1}^{m} C_i = \left\{x \in \mathbb{R}^\infty : (x_1, \dots, x_n) \in \bigcup_{i=1}^{m} \tilde{D}_i\right\}$$

  where $\bigcup_{i=1}^{m} \tilde{D}_i \in \mathscr{B}(\mathbb{R}^n)$ and so $\bigcup_{i=1}^{m} C_i \in \mathscr{C}$.

Therefore, the cylindrical sets form an algebra. ■

Exercise 1.2.8

Solution. Note $x \in D$ if and only if for all $\epsilon > 0$ there exists an $N \in \mathbb{N}$ such that for $n \geq N$ we have $|x_n - c| < \epsilon$. Therefore,

$$D = \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{ x \in \mathbb{R}^{\infty} : |x_n - c| < \frac{1}{k} \right\}.$$

Note that

$$\bigcap_{n=N}^{\infty} \left\{ x \in \mathbb{R}^{\infty} : |x_n - c| < \frac{1}{k} \right\} = \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{N-1} \times \left(c - \frac{1}{k}, c + \frac{1}{k} \right) \times \left(c - \frac{1}{k}, c + \frac{1}{k} \right) \times \cdots \in \mathscr{B}(\mathbb{R}^{\infty}).$$

Therefore, $D \in \mathscr{B}(\mathbb{R}^{\infty})$. ■

Exercise 1.3.2

Solution.

1. Let $x \leq y$. Then $(-\infty, x] \subseteq (-\infty, y]$ so that $\mathbb{P}((-\infty, x]) \leq \mathbb{P}((-\infty, y])$, which implies that $F(x) \leq F(y)$.

2. Let $(x_n)_{n \in \mathbb{N}}$ be a sequence such that $x_{n+1} \leq x_n$ and $x_n \to -\infty$ as $n \to \infty$. Then the sets $A_n = (-\infty, x_n]$ are a decreasing sequence of events. Therefore,

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

   Note that $\bigcap_{n=1}^{\infty} A_n = \emptyset$ so that $\lim_{n \to \infty} F(x_n) = 0$. We can conclude that $\lim_{x \to -\infty} F(x) = 0$. Similarly, let $(x_n)_{n \in \mathbb{N}}$ be a sequence such that $x_n \leq x_{n+1}$ and $x_n \to \infty$ as $n \to \infty$. Then the sets $A_n = (x_n, \infty)$ are a sequence of decreasing events. Therefore,

$$0 = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n \right) = \lim_{n \to \infty} \mathbb{P}(A_n).$$

   As $\mathbb{P}(A_n^c) = F(x_n)$, it follows that $1 = \lim_{n \to \infty} F(x_n)$ from which we conclude that $\lim_{x \to \infty} F(x) = 1$.

3. Let $x \in \mathbb{R}$ and $x_n \searrow x$ monotonically. Then the sets $A_n = (-\infty, x_n]$ are a sequence of decreasing events with the property that $\bigcap_{n=1}^{\infty} A_n = (-\infty, x]$. Therefore,

$$F(x) = \mathbb{P}((-\infty, x]) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n \right) = \lim_{n \to \infty} F(x_n).$$

   More generally we see that $F(x) = \lim_{y \searrow x} F(y)$ and so $F(x)$ is right-continuous.

■

Exercise 1.7.2

Solution. The possible outcomes of X are $X(\omega_1, \omega_2) = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$. The possible pre-images are summarised in Table 1.

Image	Pre-Image
2	$\{(1,1)\}$
3	$\{(1,2)(2,1)\}$
4	$\{(1,3),(2,2),(3,1)\}$
5	$\{(1,4),(2,3),(3,2),(4,1)\}$
6	$\{(1,5),(2,4),(3,3),(4,2),(5,1)\}$
7	$\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$
8	$\{(2,6),(3,5),(4,4),(5,3),(6,2)\}$
9	$\{(3,6),(4,5),(5,4),(6,3)\}$
10	$\{(4,6),(5,5),(6,4)\}$
11	$\{(5,6),(6,5)\}$
12	$\{(6,6)\}$

Table 1: Pre-images of X.

■

Exercise 1.7.5

Solution.

- As $\xi^{-1}(\varnothing) = \varnothing$ it follows that $\varnothing \in \mathscr{F}_\xi$.
- Let $A \in \mathscr{F}_\xi$ such that $A = \xi^{-1}(B)$ for some $B \in \mathscr{B}(\mathbb{R})$. Then $A^c = \left(\xi^{-1}(B)\right)^c = \xi^{-1}(B^c)$. As $B^c \in \mathscr{B}(\mathbb{R})$ it follows that $A^c \in \mathscr{F}_\xi$.
- Let $A_1, A_2, \dots \in \mathscr{F}_\xi$ with $A_i = \xi^{-1}(B_i)$ for some $B_i \in \mathscr{B}(\mathbb{R})$. Then

$$\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} \xi^{-1}(B_i) = \xi^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right).$$

  As $\bigcup_{i=1}^{\infty} B_i \in \mathscr{B}(\mathbb{R})$ it follows that $\bigcup_{i=1}^{\infty} A_i \in \mathscr{F}_\xi$.

We conclude that $\mathscr{F}_\xi$ is a σ-algebra. ■

Exercise 1.8.5

Solution. $(\Rightarrow)$. Without loss of generality consider ξ_1 and $B \in \mathscr{B}(\mathbb{R})$. Then,

$$\begin{aligned}\xi_1^{-1}(B) &= \{\omega : \xi_1(\omega) \in B\} \\ &= \{\omega : \xi(\omega) \in B \times \mathbb{R} \times \cdots \times \mathbb{R}\}.\end{aligned}$$

As $B \times \mathbb{R} \times \cdots \times \mathbb{R} \in \mathscr{B}(\mathbb{R}^n)$ and ξ is a random variable we conclude that $\xi_1^{-1}(B) \in \mathscr{F}$. Hence, ξ_1 is a random variable.
$(\Leftarrow)$. For $B \in \mathscr{B}(\mathbb{R}^n)$, we can write $B = B_1 \times \cdots \times B_n$ for $B_i \in \mathscr{B}(\mathbb{R})$. Therefore,

$$\xi^{-1}(B) = \{\omega : \xi_1(\omega) \in B_1, \dots, \xi_n(\omega) \in B_n\} = \bigcap_{i=1}^{n} \xi_i^{-1}(B_i).$$

Then as each ξ_i is a random variable we have that $\xi_i^{-1}(B_i) \in \mathscr{F}$, which implies that $\xi^{-1}(B) \in \mathscr{F}$. Therefore, ξ is a random variable. ■

2 Expectation and Integrals

2.1 The Lebesgue Integral

The expectation of a simple random variable $\xi = \sum_{j=1}^{n} x_j \chi_{D_j}$ is

$$\mathbb{E}(\xi) = \sum_{j=1}^{n} x_j \mathbb{P}(D_j),$$

where the sets D_j form a partition of Ω. For an arbitrary non-negative random variable $\xi = \xi(\omega)$ we can construct a sequence of simple non-negative random variables $(\xi_n)_{n\in\mathbb{N}}$ such that $\xi_n(\omega) \nearrow \xi(\omega)$, as $n \to \infty$ for each $\omega \in \Omega$. We then set $\mathbb{E}(\xi) = \lim_{n\to\infty} \mathbb{E}(\xi_n)$, which exists since $\mathbb{E}(\xi_n) \leq \mathbb{E}(\xi_{n+1})$, possibly taking infinite value.

Definition 2.1.1 The expectation $\mathbb{E}(\xi)$ of a non-negative random variable ξ is the Lebesgue integral with respect to $\mathbb{P}$ given by

$$\mathbb{E}(\xi) := \lim_{n\to\infty} \mathbb{E}(\xi_n) = \int_\Omega \xi \, \mathrm{d}\mathbb{P} = \int_\Omega \xi(\omega) \, \mathbb{P}(\mathrm{d}\omega).$$

To see that Definition 2.1.1 is consistent, one has to show it is independent of the choice of $\xi_n \nearrow \xi$.

Definition 2.1.2 An arbitrary random variable ξ is integrable if $\mathbb{E}(|\xi|) < \infty$.

Remark 2.1.3 An integrable random variable may not be non-negative. However, its expectation is well-defined as $|\xi|$ is a non-negative random variable.

Definition 2.1.4 The expectation of an integrable random variable is $\mathbb{E}(\xi) = \mathbb{E}(\xi^+) - \mathbb{E}(\xi^-)$.

Definition 2.1.4 is well-defined as $\mathbb{E}(\xi^-) < \infty$ by assumption.

2.2 Properties

Proposition 2.2.1 Let ξ and η be integrable random variables and let c be a constant. Then

1. $\mathbb{E}(c) = c$,
2. $\mathbb{E}(c\xi) = c\mathbb{E}(\xi)$,
3. $\xi + \eta$ is integrable with $\mathbb{E}(\xi + \eta) = \mathbb{E}(\xi) + \mathbb{E}(\eta)$,
4. $\xi \leq \eta$ implies that $\mathbb{E}(\xi) \leq \mathbb{E}(\eta)$,
5. if $\xi = \eta$ almost everywhere with respect to $\mathbb{P}$, that is the equality holds up to sets of zero $\mathbb{P}$-measure, then $\mathbb{E}(\xi) = \mathbb{E}(\eta)$, and
6. if $\xi \geq 0$ is such that $\mathbb{E}(\xi) = 0$, then $\xi = 0$ almost everywhere.

2.2.1 Exchanging limits and expectations

Theorem 2.2.2 — Monotone Convergence Theorem. Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of random variables such that $0 \leq \xi_1 \leq \xi_2 \leq \ldots$. Then

$$\lim_{n\to\infty} \mathbb{E}(\xi_n) = \mathbb{E}\left(\lim_{n\to\infty} \xi_n\right)$$

exists, potentially being infinite.

Remark 2.2.3

- Note that Theorem 2.2.2 also holds for a sequence of random variables $(\xi_n)_{n\in\mathbb{N}}$ where $\eta \le \xi_1 \le \dots$ for $\mathbb{E}(\eta) > -\infty$. Just consider $\xi_n - \eta$ instead of ξ_n in the original formulation of Theorem 2.2.2.
- Similarly, if $(\xi_n)_{n\in\mathbb{N}}$ is a sequence of random variables such that $\dots \le \xi_2 \le \xi_1 \le \eta$, for $\mathbb{E}(\eta) < \infty$, then just consider $\eta - \xi_n$ in the original formulation of Theorem 2.2.2.

Corollary 2.2.4 Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of non-negative random variables. Then

$$\mathbb{E}\left(\sum_{n=1}^{\infty}\xi_n\right) = \sum_{n=1}^{\infty}\mathbb{E}(\xi_n).$$

Proof. Let $\eta_n = \sum_{k=1}^{n}\xi_k$. Then as each ξ_k is non-negative, we have that $0 \le \eta_1 \le \eta_2 \le \dots$. Therefore

$$\begin{aligned}
\sum_{k=1}^{\infty}\mathbb{E}(\xi_k) &= \lim_{n\to\infty}\sum_{k=1}^{n}\mathbb{E}(\xi_k) \\
&= \lim_{n\to\infty}\mathbb{E}\left(\sum_{k=1}^{n}\xi_k\right) \\
&= \lim_{n\to\infty}\mathbb{E}(\eta_n) \\
&\overset{\text{Thm } 2.2.2}{=} \mathbb{E}\left(\lim_{n\to\infty}\eta_n\right) \\
&= \mathbb{E}\left(\sum_{k=1}^{\infty}\xi_k\right).
\end{aligned}$$

■

Theorem 2.2.5 — Fatou's Lemma. Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of non-negative random variables. Then

$$\mathbb{E}\left(\liminf_{n\to\infty}(\xi_n)\right) \le \liminf_{n\to\infty}\mathbb{E}(\xi_n).$$

Proof. Let $f_n = \inf_{m\ge n}\xi_n$. Then, $0 \le f_n \le f_{n+1}$ for all $n\in\mathbb{N}$. Therefore, by Theorem 2.2.2 we have that

$$\begin{aligned}
\mathbb{E}\left(\liminf_{n\in\mathbb{N}}(\xi_n)\right) &= \mathbb{E}\left(\lim_{n\to\infty}f_n\right) \\
&= \lim_{n\to\infty}\left(\mathbb{E}\left(f_n\right)\right) \\
&= \liminf_{n\to\infty}\left(\mathbb{E}\left(f_n\right)\right) \\
&\le \liminf_{n\to\infty}\left(\mathbb{E}\left(\xi_n\right)\right).
\end{aligned}$$

■

Remark 2.2.6

- In Theorem 2.2.5, $\xi_n \ge 0$ can be replaced by $\xi_n \ge \eta$, if $\mathbb{E}(\eta) > -\infty$.
- In Theorem 2.2.5, if $\xi_n < \eta$ and $\mathbb{E}(\eta) < \infty$, then the statement holds for $\limsup$ instead.

Theorem 2.2.7 — Dominated Convergence Theorem. Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of random variables such that $\xi_n \to \xi$ almost surely. If there exists an integrable random variable η such that $|\xi_n| \le \eta$ for all $n\in\mathbb{N}$, then ξ is integrable, with

$$\mathbb{E}(\xi_n) \to \mathbb{E}(\xi),$$

and

$$\mathbb{E}\left(|\xi_n - \xi|\right) \to 0$$

as $n \to \infty$.

Corollary 2.2.8 Let $\eta, \xi, (\xi_n)_{n\in\mathbb{N}}$ be random variables such that $|\xi_n| \leq \eta$, $\xi_n \to \xi$ almost everywhere and $\mathbb{E}\left(\eta^p\right) < \infty$ for some $p > 0$. Then $\mathbb{E}\left(|\xi|^p\right) < \infty$ and $\mathbb{E}\left(|\xi - \xi_n|^p\right) \to 0$ as $n \to \infty$.

Proof. Note that $|\xi_n|^p \to |\xi|^p$ almost surely, $|\xi_n|^p \leq \eta^p$ and $\mathbb{E}\left(\eta^p\right) < \infty$ Thus, by Theorem 2.2.7 it follows that

$$\mathbb{E}\left(|\xi_n|^p\right) \to \mathbb{E}\left(|\xi|^p\right).$$

As the inequality $\mathbb{E}\left(|\xi_n|^p\right) \leq \mathbb{E}\left(\eta^p\right)$ is preserved under the limit it follows that $\mathbb{E}\left(|\xi|^p\right) \leq \mathbb{E}\left(\eta^p\right) < \infty$. Next, consider $\mu_n := \xi - \xi_n$. It is clear that $\mu_n \to 0$ almost surely and $|\mu_n| \leq |\xi| + \eta$ with

$$\mathbb{E}\left(||\xi| + \eta|^p\right) \leq \mathbb{E}\left(|\xi|^p\right) + \mathbb{E}\left(|\eta|^p\right) \leq 2\mathbb{E}\left(|\eta|^p\right) < \infty.$$

So by Theorem 2.2.7 it follows that

$$\mathbb{E}\left(|\xi - \xi_n|^p\right) = \mathbb{E}\left(|\mu_n|^p\right) \to 0$$

as $n \to \infty$. ■

Remark 2.2.9 In all the above theorems, the integral over Ω can be replaced by the integral over any measurable $A \subseteq \Omega$.

2.2.2 Change of variables

Theorem 2.2.10 Let $\xi : \mathscr{F} \to \mathbb{R}$ be a random variable with probability distribution $\mathbb{P}_\xi$. If $g = g(x)$ is a Borel function, then for all $A \in \mathscr{B}(\mathbb{R})$ we have

$$\int_A g(x)\,\mathrm{d}\mathbb{P}_\xi = \int_{\xi^{-1}(A)} g(\xi(\omega))\,\mathrm{d}\mathbb{P},$$

where the integrals exist or do not exist simultaneously. In particular, for $A = \mathbb{R}$ we obtain

$$\mathbb{E}(g(\xi(\omega))) = \int_\Omega g(\xi(\omega))\,\mathrm{d}\mathbb{P} = \int_{-\infty}^{\infty} g(x)\,\mathrm{d}\mathbb{P}_\xi \equiv \int_{-\infty}^{\infty} g(x)\,\mathrm{d}F_\xi.$$

Proof. We use the four-step proof. The result clearly holds for $g = \chi_B(x)$ for $B \in \mathscr{B}(\mathbb{R})$. Therefore, it also holds for simple $g(x)$ by the linearity of the integral. For any non-negative measurable function g we can consider a sequence of simple functions $g_n \nearrow g$. The result for g then follows from the monotone convergence theorem. For arbitrary measurable $g(x)$ we use $g(x) = g_+ - g_-$. ■

Remark 2.2.11 Theorem 2.2.10 guarantees that the expectation only depends on the probability distribution, and not on the underlying probability space.

1. If ξ is discrete, so that $\mathscr{F}_\xi$ is discrete, taking values $x_1, x_2, \ldots$ with probabilities $p_1, p_2, \ldots$ then

$$\mathbb{E}(g(\xi)) = \sum_{n\in\mathbb{N}} g(x_n)p_n.$$

2. If ξ is absolutely continuous, so that F_ξ is absolutely continuous, with density $f(x)$, then

$$\mathbb{E}(g(\xi)) = \int_{-\infty}^{\infty} g(x)f(x)\,\mathrm{d}x.$$

This provides a way to calculate expectations of $g(\xi)$ without being conscious of the actual distribution of $g(\xi)$. Thus we can make sense of the expectation of probability distributions without specifying its underlying probability space. So is there any point in specifying the underlying probability space of a random variable instead of assuming it to be $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mathbb{P}_\xi)$? Unfortunately, the answer can be no, since there is a way to develop probability theory, and notions of random variables, without going through Kolmogorov's constructions. Indeed, the underlying probability space is insignificant in most applications; nevertheless, this formulation is still helpful in understanding more complicated random variables like stochastic processes.

2.3 Exchanging the Order of Integration

Suppose $(X_1, \mathscr{M}_1, \mu_1)$ and $(X_2, \mathscr{M}_2, \mu_2)$ are measure spaces. Consider the product measure $\mu_1 \times \mu_2$ on the space

$$X = X_1 \times X_2 = \{(x_1, x_2) : x_1 \in X_1,\, x_2 \in X_2\}\,.$$

Assume the measure spaces are complete and σ-finite. Given a set E in $\mathscr{M}$ let

1. $E_{x_1} = \{x_2 \in X_2 : (x_1, x_2) \in E\}$, and
2. $E^{x_2} = \{x_1 \in X_1 : (x_1, x_2) \in E\}$.

Theorem 2.3.1 — Fubini's Theorem. In the setting above, suppose that $f(x_1, x_2)$ is an integrable function on $(X_1 \times X_2, \mu_1 \times \mu_2)$. Then the following hold.

- For almost every $x_2 \in X_2$, the slice $f^{x_2}(x_1) = f(x_1, x_2)$ is integrable on (X_1, μ_1).
- The function $\int_{X_1} f(x_1, x_2)\, \mathrm{d}\mu_1$ is integrable on X_2.
- We can exchange integrals as follows

$$\int_{X_2} \left(\int_{X_1} f(x_1, x_2)\, \mathrm{d}\mu_1 \right) \mathrm{d}\mu_2 = \int_{X_1} \left(\int_{X_2} f(x_1, x_2)\, \mathrm{d}\mu_2 \right) \mathrm{d}\mu_1 = \int_{X_1 \times X_2} f \, \mathrm{d}\,(\mu_1 \times \mu_2)\,.$$

Remark 2.3.2 In general, the product space $(X, \mathscr{M}, \mu)$ is not complete. One can construct the completion of this space $\overline{\mathscr{M}}$, as the collection of sets of the form $E \cup Z$, where $E \in \mathscr{M}$ and $Z \subseteq F$ with $F \in \mathscr{M}$ and $\mu(F) = 0$. One then extends the measure with $\overline{\mu}(E \cup Z) := \mu(E)$.

- $\overline{\mathscr{M}}$ is the smallest σ-algebra containing $\mathscr{M}$ and all subsets of elements of $\mathscr{M}$ of measure zero.
- The function $\overline{\mu}$ is a measure on $\overline{\mathscr{M}}$,

Theorem 2.3.1 continues to hold in this completed space.

In Theorem 2.3.1, we assume that the function f is integrable over the product space. We can relax this condition, by instead assuming that f is a non-negative measurable function.

Theorem 2.3.3 — Tonelli's Theorem. Suppose that $f(x_1, x_2) : X_1 \times X_2 \to [0, \infty]$ is a non-negative measurable function on $(X_1 \times X_2, \mu_1 \times \mu_2)$. Then

$$\int_{X_2} \left(\int_{X_1} f(x_1, x_2)\, \mathrm{d}\mu_1 \right) \mathrm{d}\mu_2 = \int_{X_1} \left(\int_{X_2} f(x_1, x_2)\, \mathrm{d}\mu_2 \right) \mathrm{d}\mu_1 = \int_{X_1 \times X_2} f \, \mathrm{d}\,(\mu_1 \times \mu_2)\,.$$

Theorem 2.3.4 — Fubini-Tonelli Theorem. If f is a measurable function, then

$$\int_{X_1} \left(\int_{X_2} |f(x_1, x_2)|\, \mathrm{d}\mu_2 \right) \mathrm{d}\mu_1 = \int_{X_2} \left(\int_{X_1} |f(x_1, x_2)|\, \mathrm{d}\mu_1 \right) \mathrm{d}\mu_2 = \int_{X_1 \times X_2} |f| \, \mathrm{d}\,(\mu_1 \times \mu_2)\,.$$

Besides, if any one of these integrals is finite, then

$$\int_{X_1} \left(\int_{X_2} f(x_1, x_2)\, \mathrm{d}\mu_2 \right) \mathrm{d}\mu_1 = \int_{X_2} \left(\int_{X_1} f(x_1, x_2)\, \mathrm{d}\mu_1 \right) \mathrm{d}\mu_2 = \int_{X_1 \times X_2} f \, \mathrm{d}\,(\mu_1 \times \mu_2)\,.$$

The absolute value of f in the conditions of Theorem 2.3.4 can be replaced by either the positive or the negative part of f. Theorem 2.3.4 is a generalisation of Theorem 2.3.3 as one can take the negative part of a non-negative function to be zero whilst maintaining a finite integral. Informally all these conditions say that the double integral of f is well-defined, though possibly infinite. The advantage of the Theorem 2.3.4 over Theorem 2.3.1 is that the repeated integrals of the absolute value of $|f|$ may be easier to study than the double integral. As in Theorem 2.3.1, the single integrals may fail to be defined on a zero-measure set.

Proposition 2.3.5 Let ξ be a non-negative integrable random variable. Then

$$\mathbb{E}(\xi) = \int_{[0,\infty)} \mathbb{P}(\xi \geq x)\,\mathrm{d}x. \tag{2.1}$$

Proof. We have

$$\begin{aligned}\mathbb{E}(\xi) &= \int_{[0,\infty)} x\,\mathrm{d}\mathbb{P}_\xi \\ &= \int_{[0,\infty)} \left(\int_0^x \mathrm{d}t\right) \mathrm{d}\mathbb{P}_\xi \\ &= \int_{[0,\infty)} \mathbb{P}(\xi \geq t)\,\mathrm{d}t,\end{aligned}$$

where we have applied Theorem 2.3.1 to

$$g(t,x) = \begin{cases} 1 & 0 \leq t \leq x \\ 0 & \text{otherwise.} \end{cases}$$

■

Exercise 2.3.6 Generalise the proof of Proposition 2.3.5 to show that if $\xi \geq 0$ and $p \geq 1$ then

$$\mathbb{E}\left(\xi^p\right) = \int_0^\infty p y^{p-1} \mathbb{P}(\xi \geq y)\,\mathrm{d}y. \tag{2.2}$$

2.4 Jensen's Inequality and L^p Spaces

2.4.1 Convex Functions and Jensen Inequality

Definition 2.4.1

- A set $\Omega \subseteq \mathbb{R}^n$ is convex if for all $x, y \in \Omega$ and $\lambda \in [0,1]$ the point $(1-\lambda)x + \lambda y \in \Omega$.
- Let $E \subseteq \mathbb{R}^n$ be a convex set. A function $g : E \to \mathbb{R}$ is convex if for all $x, y \in \Omega$ and $\lambda \in [0,1]$ we have
$$g((1-\lambda)x + \lambda y) \leq (1-\lambda)g(x) + \lambda g(y).$$
- Let $E \subseteq \mathbb{R}^n$ be a convex set. A function $g : E \to \mathbb{R}$ is concave if $-g$ is convex.

Proposition 2.4.2 Let $g : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Then for $x_0 \in \mathbb{R}^n$ there is a vector $v \in \mathbb{R}^n$, dependent on x_0, such that

$$g(x) \geq g(x_0) + v^\top (x - x_0), \tag{2.3}$$

for all $x \in \mathbb{R}^n$.

Remark 2.4.3 Any vector v satisfying (2.3) is called a subgradient of g at x_0.

Theorem 2.4.4 — Jensen's Inequality. Let ξ be an integrable random variable and let $g : \mathbb{R} \to \mathbb{R}$ be a measurable convex function. Then
$$g(\mathbb{E}(\xi)) \leq \mathbb{E}(g(\xi)).$$

Proof. If $g(x)$ is convex then for each $x_0 \in \mathbb{R}$ there is a $v \in \mathbb{R}$ such that
$$g(x) \geq g(x_0) + v(x - x_0)$$
for all $x \in \mathbb{R}$. Putting $x = \xi$ and $x_0 = \mathbb{E}(\xi)$, we find that
$$g(\xi) \geq g(\mathbb{E}(\xi)) + v(\xi - \mathbb{E}(\xi)).$$
Taking the expectation of both sides we get $g(\mathbb{E}(\xi)) \leq \mathbb{E}(g(\xi))$. ■

Exercise 2.4.5 Prove that the composition of a convex non-decreasing function and a convex function is convex. Conclude that $f(x) = |x|^r$ is a convex function.

Corollary 2.4.6 — Lyapunov's Inequality. If $0 < p < q < \infty$, then
$$\left(\mathbb{E}\left(|\xi|^p\right)\right)^{\frac{1}{p}} \leq \left(\mathbb{E}\left(|\xi|^q\right)\right)^{\frac{1}{q}}. \tag{2.4}$$

Proof. Let $r = \frac{q}{p}$. Then letting $\eta = |\xi|^p$ and using Exercise 2.4.5 we can apply Jensen's inequality to $g(x) = |x|^r$ to get that $|\mathbb{E}(\eta)|^r \leq \mathbb{E}(|\eta|^r)$. That is,
$$\left(\mathbb{E}\left(|\xi|^p\right)\right)^{\frac{q}{p}} \leq \mathbb{E}\left(|\xi|^q\right)$$
from which (2.4) follows. Consequently, if $\mathbb{E}\left(|\xi|^q\right) < \infty$ then $\mathbb{E}\left(|\xi|^p\right) = \left(\mathbb{E}\left(|\xi|^q\right)\right)^{\frac{p}{q}} < \infty$. ■

The following chain of inequalities among absolute moments
$$\mathbb{E}(|\xi|) \leq \left(\mathbb{E}\left(|\xi|^2\right)\right)^{\frac{1}{2}} \leq \cdots \leq \left(\mathbb{E}\left(|\xi|^n\right)\right)^{\frac{1}{n}} \leq \ldots$$
is a consequence of Lyapunov's inequality.

Remark 2.4.7 As a warning, Lyapunov inequality is only true when $\mathbb{P}$ is a finite measure, which is certainly the case for probability measures.

Definition 2.4.8 Let ξ be a random variable with $\mathbb{E}\left(|\xi|^p\right) < \infty$. For integers $0 \leq k \leq p$, the k^{th} moment of ξ is $\mathbb{E}\left(\xi^k\right)$.

Definition 2.4.9 A sequence $(\xi_n)_{n\in\mathbb{N}}$ of random variables converges in L^p to the random variable ξ if
$$\left(\mathbb{E}\left(|\xi_n - \xi|^p\right)\right)^{\frac{1}{p}} \to 0$$
as $n \to \infty$.

Often one uses the notation $\|\xi\|_{L^p} = \left(\mathbb{E}\left(|\xi|^p\right)\right)^{\frac{1}{p}}$. This is done to allude to the fact that $\|\cdot\|_{L^p}$ defines a norm on the vector space of functions with finite p^{th} moments, up to sets of $\mathbb{P}$-measure zero. More formally, we say that $\xi \in \mathscr{L}^p$ if $\mathbb{E}\left(|\xi|^p\right) < \infty$ with $\xi \sim \eta$ if $\xi = \eta$ almost everywhere. Then $\|\cdot\|_{L^p}$ defines a norm on $L^p = \mathscr{L}^p / \sim$. This result is facilitated by the following inequalities.

Proposition 2.4.10 — Hölder's Inequality. Let $p \in [1, \infty]$ and let $q \in [1, \infty]$ be such that $\frac{1}{p} + \frac{1}{q} = 1$ with $q = \infty$ if $p = 1$ and vice-versa. If $\xi \in L^p$ and $\eta \in L^q$, then
$$\|\xi\eta\|_{L^1(\Omega)} := \mathbb{E}(|\xi\eta|) \leq \|\xi\|_{L^p} \|\eta\|_{L^q}.$$

Proposition 2.4.11 — Minkowski's Inequality. If $\xi, \eta \in L^p$ for $1 \leq p \leq \infty$, then $\xi + \eta \in L^p$ and

$$\left(\mathbb{E}\left(|\xi+\eta|^p\right)\right)^{\frac{1}{p}} \leq \left(\mathbb{E}\left(|\xi|^p\right)\right)^{\frac{1}{p}} + \left(\mathbb{E}\left(|\eta|^p\right)\right)^{\frac{1}{p}}.$$

Remark 2.4.12 If the sequence $(\xi_n)_{n\in\mathbb{N}}$ converges in L^p to ξ, then

$$0 \leq \left| \|\xi_n\|_{L^p} - \|\xi\|_{L^p} \right| \leq \|\xi_n - \xi\|_{L^p} \to 0$$

as $n \to \infty$. That is, $\|\xi_n\|_{L^p} \to \|\xi\|_{L^p}$.

2.5 Tail Bounds

Most large sample results concern extreme events, for example, whether the value of a random variable deviates from its mean. This section builds the necessary tools to derive upper bounds for the probability of such events. These bounds are usually called tail bounds since they correspond to the tail of the densities of random variables. In particular, we will see how the tail bounds are related to the integrability of the random variables.

Example 2.5.1

1. Let ξ_1 have a normal distribution $\mathsf{N}(0,1)$ with density

$$f_1(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

for $x \in \mathbb{R}$, such that

$$\begin{aligned}
\mathbb{P}(\xi_1 > c) &= \int_c^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \\
&\leq \int_c^\infty \frac{1}{\sqrt{2\pi}} \frac{x}{c} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \\
&= \frac{1}{c\sqrt{2\pi}} \exp\left(-\frac{c^2}{2}\right) \\
&=: r_1(c).
\end{aligned}$$

2. Let ξ_2 have a double exponential (Laplace) distribution with density

$$f_2(x) = \frac{1}{2} e^{-|x|}$$

for $x \in \mathbb{R}$, such that

$$\begin{aligned}
\mathbb{P}(\xi_2 > c) &= \int_c^\infty \frac{1}{2} \exp\left(-x\right) \mathrm{d}x \\
&= \frac{1}{2} \exp(-c) \\
&=: r_2(c)
\end{aligned}$$

for $c > 0$.

3. Let ξ_3 have a standard Cauchy distribution with density

$$f_3(x) = \frac{1}{\pi(1+x^2)}$$

for $x \in \mathbb{R}$, such that

$$\begin{aligned}\mathbb{P}(\xi_3 > c) &= \frac{1}{2} - \frac{1}{\pi}\arctan c\\ &= \frac{1}{\pi}\arctan\frac{1}{c}\\ &\leq \frac{1}{\pi c}\\ &=: r_3(c).\end{aligned}$$

Then $\mathbb{P}(X > c)$ decays faster for the normal distribution than the double exponential distribution, and the Cauchy distribution admits the slowest decay. In particular, we have $r_3(c) \gg r_2(c) \gg r_1(c)$ in the sense that $\frac{r_1(c)}{r_2(c)} \to 0$ and $\frac{r_2(c)}{r_3(c)} \to 0$ when $c \to \infty$.

Exercise 2.5.2 In the context of Example 2.5.1, verify the following observations.

1. ξ_1 has zero odd moments, and has $(2k)^{\text{th}}$ moments

   $$m_{1,k} = (2k-1)!! := (2k-1) \times \cdots \times 3 \times 1 = \frac{(2k)!}{2^k k!}$$

   for all $k \in \mathbb{Z}_{\geq 1}$.
2. ξ_2 has zero odd moments, and has $(2k)^{\text{th}}$ moments $m_{2,k} = (2k)!$.
3. ξ_3 has $(2k)^{\text{th}}$ moments $m_{3,k} = \infty$.

Therefore, we see that $\infty = m_{3,k} \gg m_{2,k} \gg m_{1,k}$ as $k \to \infty$, in the sense that as $k \to \infty$ we have $\frac{m_{1,k}}{m_{2,k}} \to 0$. We therefore suspect that there is a connection between the tail bounds and the growth of moments.

To standardise the discussion of random variables, one often centralises the moments in the following way.

Definition 2.5.3 The k^{th} central moment, for $k \in \mathbb{N}$, of a random variable ξ is $\mathbb{E}\left((\xi - \mathbb{E}(\xi))^k\right)$ whenever $\mathbb{E}\left(|\xi|^k\right) < \infty$. In particular, the first central moment is zero. The main central moments of interest are the following.

- The 2^{nd} central moment $\mathbb{V}(\xi) := \mathbb{E}\left((\xi - \mathbb{E}(\xi))^2\right)$ is called the variance.
- The 3^{rd} central moment $\mathbb{E}\left((\xi - \mathbb{E}(\xi))^3\right)$ is called the skewness.
- The 4^{th} central moment $\mathbb{E}\left((\xi - \mathbb{E}(\xi))^4\right)$ is called the kurtosis.

Theorem 2.5.4 — Markov's Inequality. Let ξ be a non-negative integrable random variable and $c > 0$ a constant. Then

$$\mathbb{P}(\xi \geq c) \leq \frac{\mathbb{E}(\xi)}{c}.$$

Proof. This follows from

$$\mathbb{E}(\xi) \geq \mathbb{E}\left(\xi \cdot \chi_{\xi \geq c}\right) \geq c\mathbb{E}\left(\chi_{\xi \geq c}\right) = c\mathbb{P}(\xi \geq c).$$

■

Remark 2.5.5 A generalisation of the Markov inequality considers ξ a random variable and g a non-negative Borel function. Let $c > 0$ be a constant and suppose that $\mathbb{E}(g(\xi))$ exists, then

$$\mathbb{P}(g(\xi) \geq c) \leq \frac{\mathbb{E}(g(\xi))}{c}. \tag{2.5}$$

Let us interrupt our discussion of tail bounds by proving an interesting result regarding L^p norms. How powerful Markov's inequality (2.5) is for proving tail bounds for specific distributions depends on the integrability of ξ.

Corollary 2.5.6 Let $\xi \in L^p(\Omega)$ for $p \geq 1$. Then for all $\varepsilon > 0$, we have

$$\mathbb{P}(|\xi - \mathbb{E}(\xi)| \geq \varepsilon) = \mathbb{P}\left(|\xi - \mathbb{E}(\xi)|^p \geq \varepsilon^p\right) \leq \frac{\mathbb{E}\left(|\xi - \mathbb{E}(\xi)|^p\right)}{\varepsilon^p}. \tag{2.6}$$

When $p = 2$ we obtain Chebyshev's inequality, which states that

$$\mathbb{P}(|\xi - \mathbb{E}(\xi)| \geq \varepsilon) \leq \frac{\mathbb{V}(\xi)}{\varepsilon^2}.$$

2.5.1 Chernoff Bound and Moment Generating Function

For the case when $\xi \in L^\infty(\Omega)$ and the k^{th} moment does not grow too quickly, one may choose an optimal p such that the right-hand side of (2.6) is minimised. This is rarely done in practice. Instead, we consider the moment-generating function.

Definition 2.5.7 — Moment Generating Function. The moment generating function of a random variable ξ is

$$M_\xi(t) = \mathbb{E}(\exp(tX)) = \int_{-\infty}^{\infty} e^{tx}\,\mathrm{d}F_\xi(x).$$

A moment-generating function does not necessarily exist for all values of $t \in \mathbb{R}$. For example, a random variable ξ with Cauchy distribution has $M_\xi(t) = \infty$ for all $t \neq 0$, and is equal to 1 for $t = 0$. However, if we can show that $M_\xi(t) < \infty$ for a small neighbourhood of zero, say $t \in (-h, h)$, then we have the following result.

Proposition 2.5.8 — Chernoff Inequality. Let ξ be a non-negative random variable, then for all $\varepsilon > 0$ and $t \in (0, h)$ we have

$$\mathbb{P}(\xi \geq \varepsilon) \leq \frac{M_X(t)}{e^{t\varepsilon}}. \tag{2.7}$$

Proof. Noting that $e^{t\xi}$ is a non-negative random variable, we can use Theorem 2.5.4 to deduce that

$$\mathbb{P}(\xi \geq \varepsilon) = \mathbb{P}\left(e^{t\xi} \geq e^{t\varepsilon}\right) \leq \frac{\mathbb{E}\left(e^{t\xi}\right)}{e^{t\varepsilon}} = \frac{M_X(t)}{e^{t\varepsilon}}.$$

■

Example 2.5.9 Let ξ follow a standard normal distribution $\mathbf{N}(0, 1)$. The $(2k)^{\text{th}}$ moments of ξ are given by $(2k-1)!!$. Thus, for $k \in \mathbb{N}$ and $c > 0$ we have

$$\mathbb{P}(X > c) \leq \frac{(2k-1)!!}{c^{2k}}. \tag{2.8}$$

Even though using larger k will lead to a faster rate of decay as $c \to \infty$, the numerator is also larger, so it is harder to use the bound for practical applications. We can obtain a sharper bound than (2.8) by using the Chernoff bounds. Observe that

$$\begin{aligned}
\mathbb{E}\left(e^{t\xi}\right) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx - \frac{x^2}{2}}\,\mathrm{d}x \\
&= \exp^{\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{2}}\,\mathrm{d}x \\
&= \exp^{\frac{t^2}{2}}
\end{aligned}$$

so that by (2.7) we have

$$\begin{aligned}\mathbb{P}(\xi > c) &\leq \exp\left(\frac{t^2}{2} - ct\right) \\ &= \exp\left(-\frac{c^2}{2} + \frac{1}{2}(t-c)^2\right). \end{aligned} \tag{2.9}$$

Since (2.9) holds for all $t > 0$, we can choose the optimal t such that the right-hand side is minimised. In our case, we choose $c > 0$ to obtain

$$\begin{aligned}\mathbb{P}(\xi > c) &\leq \exp\left(\frac{t^2}{2} - ct\right) \\ &= \exp\left(-\frac{c^2}{2}\right). \end{aligned}$$

2.6 Solution to Exercises

Exercise 2.3.6

Solution. Note that

$$\mathbb{E}\left(\xi^p\right) = \int_0^\infty t^p \,\mathrm{d}\mathbb{P}_\xi(t) = \int_0^\infty \int_0^t p y^{p-1} \,\mathrm{d}y \,\mathrm{d}\mathbb{P}_\xi(t).$$

Applying Fubini's theorem to

$$g(y,t) = p y^{p-1} \chi_{\{0 \leq y \leq t\}}$$

allows us to interchange the integrals to conclude that

$$\begin{aligned}\mathbb{E}\left(\xi^p\right) &= \int_0^\infty \int_y^\infty p y^{p-1} \,\mathrm{d}\mathbb{P}_\xi(t) \,\mathrm{d}y \\ &= \int_0^\infty p y^{p-1} \mathbb{P}(\xi \geq y) \,\mathrm{d}y. \end{aligned}$$

■

Exercise 2.4.5

Solution. Let $f = g \circ h$, where g is a non-decreasing convex function and h is a convex function. Then for $x, y \in \Omega$ and $\lambda \in [0,1]$ we have

$$\begin{aligned}(g \circ h)((1-\lambda)x + \lambda y) &= g(h((1-\lambda)x + \lambda y)) \\ &\overset{(1)}{\leq} g((1-\lambda)h(x) + \lambda h(y)) \\ &\overset{(2)}{\leq} (1-\lambda)(g \circ h)(x) + \lambda (g \circ h)(y), \end{aligned}$$

where in (1) we have used the fact that h is convex and g is non-decreasing, and in (2) we have used the fact that g is convex. Therefore, $g \circ h$ is a convex function. Hence, with $g(x) = x^r$ for $x \geq 0$ and $h(x) = |x|$, we deduce that $f(x) = |x|^r$ is convex. ■

Exercise 2.5.2

Solution.

1. The odd moments are given by
$$\int_{-\infty}^\infty \frac{x^{2k+1}}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \,\mathrm{d}x$$
which is zero as the integrand is an odd function. For the even moments, we can proceed by induction.
   - Note that $\mathbb{E}\left(\xi^2\right) = \mathbb{V}(\xi) + \mathbb{E}(\xi) = 1 = (2(1)-1)!!$.

- For $k \geq 1$ we have

$$\begin{aligned}
\mathbb{E}\left(\xi^{2k}\right) &= \int_{-\infty}^{\infty} \frac{x^{2k}}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \\
&= 2\int_{0}^{\infty} \frac{x^{2k}}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \\
&\overset{(1)}{=} 2\left(\left[-x^{2k-1} \exp\left(-\frac{x^2}{2}\right)\right]_0^\infty + \int_0^\infty (2k-1)x^{2k-2} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x\right) \\
&= (2k-1)\int_{-\infty}^{\infty} x^{2k-2} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \\
&\overset{\text{Ind Hyp.}}{=} (2k-1)(2k-3)!! \\
&= (2k-1)!!,
\end{aligned}$$

where in (1) we have performed integration by parts with $u = x^{2k-1}$ and $\frac{\mathrm{d}v}{\mathrm{d}x} = x\exp\left(-\frac{x^2}{2}\right)$.

2. The odd moments are given by

$$\int_{-\infty}^{\infty} \frac{x}{2} \exp(-x)\, \mathrm{d}x$$

which is zero as the integrand is an odd function. For the even moments, we can proceed by induction.

- The case for $k = 0$ is clear.
- For $k \geq 1$ we have

$$\begin{aligned}
\mathbb{E}\left(\xi^{2k}\right) &= \int_{-\infty}^{\infty} \frac{x^{2k}}{2} \exp(-x)\, \mathrm{d}x \\
&= \int_{0}^{\infty} \frac{x^{2k}}{2} \exp(-x)\, \mathrm{d}x \\
&= 2\left(\left[-x^{2k} \exp(-x)\right]_0^\infty + \int_0^\infty 2kx^{2k-1} \exp(-x)\, \mathrm{d}x\right) \\
&= 2(2k)\int_0^\infty x^{2k-1} \exp(-x)\, \mathrm{d}x \\
&= 2(2k)\left(\left[-x^{2k-1} \exp(-x)\right]_0^\infty + \int_0^\infty (2k-1)x^{2k-2} \exp(-x)\, \mathrm{d}x\right) \\
&= 2(2k)(2k-1)\frac{(2k-2)!}{2} \\
&= (2k)!.
\end{aligned}$$

3. Recalling that $\mathbb{E}(|\xi|) = \infty$, we can use Corollary 2.4.6 to conclude that all higher-order moments are also infinite.

■

3 More on Random Variables

3.1 Transformation of Random Variables

Let us consider the problem of determining the distribution function of a random variable which is the function of other random variables. That is, let ξ be a random variable with distribution function $F_\xi(x)$ (and density $f_\xi(x)$ if it exists), and let $\varphi = \varphi(x)$ be a Borel function such that $\eta = \varphi(\xi)$. Then we would like to determine the distribution function of η. Proceeding directly we get that

$$F_\eta(y) = \mathbb{P}(\eta \le y) = \mathbb{P}\left(\xi \in \varphi^{-1}(-\infty, y]\right) = \int_{\varphi^{-1}(-\infty,y]} \mathrm{d}F_\xi,$$

which expresses the distribution function $F_\eta(y)$ in terms of $F_\xi(x)$ and φ.

Example 3.1.1

1. Let $\eta = a\xi + b$ with $a > 0$. Then

$$F_\eta(y) = \mathbb{P}(\eta \le y) = \mathbb{P}\left(\xi \le \frac{y-b}{a}\right) = F_\xi\left(\frac{y-b}{a}\right).$$

2. Let $\eta = \xi^2$. Then it is evident that $F_\eta(y) = 0$ if $y < 0$. While for $y \ge 0$, we have

$$\begin{aligned} F_\eta(y) &= \mathbb{P}(\xi^2 \le y) \\ &= \mathbb{P}\left(-\sqrt{y} \le \xi \le \sqrt{y}\right) \\ &= \mathbb{P}_\xi\left((-\infty, \sqrt{y}]\right) - \mathbb{P}_\xi\left((-\infty, -\sqrt{y})\right) \\ &= F_\xi\left(\sqrt{y}\right) - F_\xi\left(-\sqrt{y}\right) + \mathbb{P}\left(\xi = -\sqrt{y}\right). \end{aligned}$$

Proposition 3.1.2 — Probability Integral Transform. Let ξ be a random variable on $(\Omega, \mathscr{F}, \mathbb{P})$ with distribution function $F_\xi(x)$. Let U be a uniformly distributed random variable on $([0,1], \mathscr{B}([0,1]), \mathrm{Leb})$, where Leb denotes the Lebesgue measure. Define the right inverse of F_ξ on $[0,1]$ as

$$F_\xi^{-1}(y) = \sup\left(\{x : F_\xi(x) < y\}\right), \tag{3.1}$$

and extend it so that $F_\xi^{-1}(0) = -\infty$ and $F_\xi^{-1}(1) = \infty$. Then ξ has the same distribution function as $F_\xi^{-1}(U)$. In such a case we say that $F_\xi^{-1}(U)$ is equally distributed as ξ, or $\xi \overset{d}{=} F_\xi^{-1}(U)$.

Proof. Note that if F_ξ is invertible we have

$$F_{F_\xi^{-1}(U)}(y) = \mathrm{Leb}\left(F_\xi^{-1}(U) \le y\right) \overset{(1)}{=} \mathrm{Leb}\left(U \le F_\xi(y)\right) = F_\xi(y),$$

where in (1) we have used the fact that F_ξ is invertible, and hence bijective, to deduce that $F_\xi^{-1}(u) \le y$ if and only if $u \le F_\xi(y)$. Consequently, $F_\xi^{-1}(U) \overset{d}{=} \xi$. Thus, to complete the proof, it suffices to show that (1) holds for general distribution functions F_ξ.
$(\Leftarrow)$ Assume $u \le F_\xi(y)$. Then clearly whenever $F_\xi(x) < u$ we have $F_\xi(x) < F_\xi(y)$ so that $x \le y$. Therefore, y is an upper bound of the set $\{x : F_\xi(x) < u\}$ and so

$$F_\xi^{-1}(u) = \sup(\{x : F_\xi(x) < u\}) \le y.$$

$(\Rightarrow)$ Assume $F_\xi^{-1}(u) \le y$ but for contradiction that $u > F_\xi(y)$. Then $y \in \{x : F_\xi(x) < u\}$ so that $y \le F_\xi^{-1}(u)$, which implies that $y = F_\xi^{-1}(u)$. However, consider a sequence $(x_n)_{n\in\mathbb{N}}$ that monotonically converges to y from above without ever equalling y. Then since F_ξ is a right-continuous decreasing function it follows that there exists an $n \in \mathbb{N}$ such that $F_\xi(y) < F_\xi(x_n) < u$. Therefore, $x_n \in \{x : F_\xi(x) < u\}$ which implies that $x_n \le F_\xi^{-1}(u) = y$ which is a contradiction. ■

Example 3.1.3 Let F be the distribution function for η which has a uniform distribution on $\left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$ with an atom at $\frac{2}{3}$.

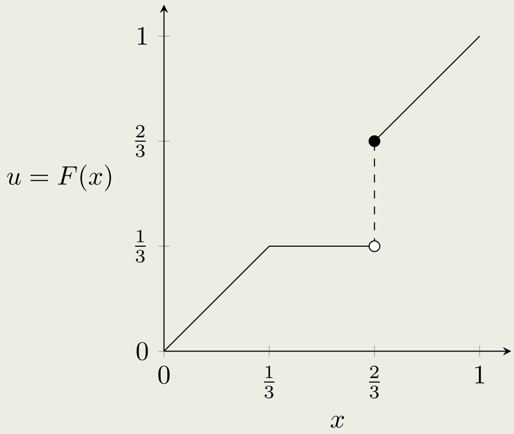


Figure 5: An example of a distribution function.

The inverse $F^{-1}(u)$ is given in Figure 6.

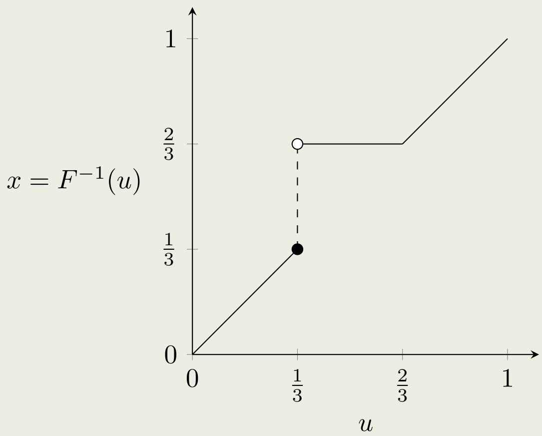


Figure 6: The pseudo-inverse of the previous example.

Now we turn to the problem of determining $f_\eta(y)$. Let us suppose that the range of ξ is a (finite or infinite) open interval $I = (a, b)$. Moreover, suppose that $\varphi = \varphi(x)$, with domain (a, b), is a strictly increasing or decreasing continuously differentiable function. We also suppose that $\varphi'(x) \neq 0$ for $x \in I$ so that we can write $h(y) = \varphi^{-1}(y)$. For definiteness suppose that φ is strictly increasing so that for $y \in \varphi(I)$ it follows that,

$$\begin{aligned}
F_\eta(y) &= \mathbb{P}(\eta \leq y) \\
&= \mathbb{P}\left(\varphi(\xi) \leq y\right) \\
&= \mathbb{P}\left(\xi \leq \varphi^{-1}(y)\right) \\
&= \mathbb{P}(\xi \leq h(y)) \\
&= \int_{-\infty}^{h(y)} f_\xi(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{y} f_\xi(h(z))h'(z)\,\mathrm{d}z.
\end{aligned}$$

Therefore,

$$f_\eta(y) = f_\xi(h(y))h'(y).$$

On the other hand, if $\varphi(x)$ is strictly decreasing, then

$$f_\eta(y) = f_\xi(h(y))\left(-h'(y)\right).$$

In either case

$$f_\eta(y) = f_\xi(h(y))\,|h'(y)|\,.$$

Example 3.1.4 For $\eta = a\xi + b$ with $a \neq 0$ we have

$$h(y) = \frac{y-b}{a}$$

and

$$f_\eta(y) = \frac{1}{|a|} f_\xi\left(\frac{y-b}{a}\right).$$

Lemma 3.1.5 Let $\varphi = \varphi(x)$, defined on the set $\sum_{k=1}^n [a_k, b_k]$, be continuously differentiable and either strictly increasing or strictly decreasing on each open interval $I_k = (a_k, b_k)$, with $\varphi'(x) \neq 0$ for $x \in I_k$. Let $h_k = h_k(y)$ be the inverse of $\varphi(x)$ for $x \in I_k$. Then

$$f_\eta(y) = \sum_{k=1}^n f_\xi(h_k(y))\,|h_k'(y)| \cdot \chi_{D_k}(y),$$

where D_k is the domain of $h_k(y)$.

Example 3.1.6

1. Let $\eta = \xi^2$ with $I_1 = (-\infty, 0)$ and $I_2 = (0, \infty)$. Observe that $h_1(y) = -\sqrt{y}$ and $h_2(y) = \sqrt{y}$, so that
$$f_\eta(y) = \begin{cases} \frac{1}{2\sqrt{y}}\left(f_\xi\left(\sqrt{y}\right) + f_\xi\left(-\sqrt{y}\right)\right) & y > 0 \\ 0 & y \leq 0. \end{cases}$$
In particular, if $\xi \sim \mathsf{N}(0,1)$ we have
$$f_{\xi^2}(y) = \begin{cases} \frac{1}{\sqrt{2\pi y}} e^{-\frac{y}{2}} & y > 0 \\ 0 & y \leq 0. \end{cases}$$

2. Let $\eta = |\xi|$, then
$$f_{|\xi|}(y) = \begin{cases} f_\xi(y) + f_\xi(-y) & y > 0 \\ 0 & y \leq 0. \end{cases}$$

3. Let $\eta = \sqrt{|\xi|}$, then
$$f_{\sqrt{|\xi|}}(y) = \begin{cases} 2y\left(f_\xi\left(y^2\right) + f_\xi\left(-y^2\right)\right) & y > 0 \\ 0 & y \leq 0. \end{cases}$$

3.2 Independent Random Variables

Definition 3.2.1 Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a measure space.

- A finite collection of events $\{A_1, \ldots, A_n\}$ is independent if $\mathbb{P}\left(\cap_{i=1}^n A_i\right) = \prod_{i=1}^n \mathbb{P}(A_i)$. An infinite collection $\{A_1, A_2, \ldots\}$ is (mutually) independent if any finite sub-collection of events is independent.
- A finite collection of sub-σ-algebras $\{\mathscr{F}_1, \ldots, \mathscr{F}_n\}$ of $\mathscr{F}$ is (mutually) independent if for any $A_1 \in \mathscr{F}_1, \ldots, A_n \in \mathscr{F}_n$, we have $\mathbb{P}\left(\cap_{i=1}^n A_i\right) = \prod_{i=1}^n \mathbb{P}(A_i)$. An infinite collection $\{\mathscr{F}_1, \mathscr{F}_2, \ldots\}$ of sub-σ-algebras of $\mathscr{F}$ is (mutually) independent if any finite sub-collection is independent.
- A finite collection $\{\xi_1, \ldots, \xi_n\}$ of random variables on $(\Omega, \mathscr{F}, \mathbb{P})$ is (mutually) independent if

the collection of corresponding sub-σ-algebras $\{\sigma(\xi_1), \dots, \sigma(\xi_n)\}$ is independent. In particular if $B_1, \dots, B_n \in \mathscr{B}(\mathbb{R})$, then

$$\mathbb{P}(\xi_1 \in B_1, \dots, \xi_n \in B_n) = \prod_{i=1}^{n} \mathbb{P}(\xi_i \in B_i) = \prod_{i=1}^{n} \mathbb{P}_{\xi_i}(B_i).$$

An infinite collection $\{\xi_1, \xi_2, \dots\}$ of random variables on $(\Omega, \mathscr{F}, \mathbb{P})$ is (mutually) independent if the corresponding collection of sub-σ-algebras $\{\sigma(\xi_1), \sigma(\xi_2), \dots\}$ is (mutually) independent.

Remark 3.2.2 Another notion of independence says that the collection of events $\{A_1, \dots, A_n\}$, is pairwise independent if for all i, j with $i \neq j$ we have $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$. Mutual independence implies pairwise independence but pairwise independence does not imply mutual independence.

Proposition 3.2.3 A necessary and sufficient condition for the random variables $\xi_1, \dots, \xi_n$ to be independent is that

$$F_\xi(x_1, \dots, x_n) = F_{\xi_1}(x_1) \dots F_{\xi_n}(x_n)$$

for all $(x_1, \dots, x_n) \in \mathbb{R}^n$.

Corollary 3.2.4 If $\xi = (\xi_1, \dots, \xi_n)$ has a density f_ξ, then each ξ_i has a density f_{ξ_i}. Furthermore, $\xi_1, \dots, \xi_n$ are independent if and only if

$$f_\xi(x_1, \dots, x_n) = f_{\xi_1}(x_1) \cdots f_{\xi_n}(x_n)$$

for all $(x_1, \dots, x_n) \in \mathbb{R}^n$ except possibly for a Borel subset of $\mathbb{R}^n$ with Lebesgue measure zero.

Corollary 3.2.5 If $\xi_1, \dots, \xi_n$ are independent and ξ_i has density f_{ξ_i}, for $i = 1, \dots, n$, then ξ has a density f_ξ given by

$$f_\xi(x_1, \dots, x_n) = f_{\xi_1}(x_1) \cdots f_{\xi_n}(x_n).$$

Remark 3.2.6 Even if $\xi_1, \dots, \xi_n$ each has a density, it does not follow that $(\xi_1, \dots, \xi_n)$ has a density.

The outcome of independent random variables does not depend on the outcome of the other variables. Therefore, it seems as though if we were to combine independent random variables, we should be able to construct a distribution function for this combination, provided we know the distribution functions of the individual random variables.

Proposition 3.2.7 Let ξ and η be independent random variables. Then the distribution function $F_{\xi+\eta}$ of their sum is given by the convolution of their distribution functions. That is,

$$F_{\xi+\eta}(z) = (F_\xi \star F_\eta)(z) = \int_{-\infty}^{\infty} F_\xi(z-y)\,\mathrm{d}F_\eta(y) = \int_{-\infty}^{\infty} F_\eta(z-x)\,\mathrm{d}F_\xi(x).$$

Proof. Using Proposition 3.2.3 we note that $F_{(\xi,\eta)}(x,y) = F_\xi(x)F_\eta(y)$. Then

$$\begin{aligned}
F_{\xi+\eta}(z) &= \int_{\{x,y:x+y\le z\}} \mathrm{d}F_\xi(x) \cdot \mathrm{d}F_\eta(y) \\
&= \int_{\mathbb{R}^2} \chi_{x+y\le z}\,\mathrm{d}F_\xi(x) \cdot \mathrm{d}F_\eta(y) \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} \chi_{x+y\le z}\,\mathrm{d}F_\xi(x) \right) \mathrm{d}F_\eta(y) \\
&= \int_{-\infty}^{\infty} F_\xi(z-y)\,\mathrm{d}F_\eta(y).
\end{aligned}$$

As this argument is symmetric we also get that

$$F_{\xi+\eta}(z) = \int_{-\infty}^{\infty} F_\eta(z-x)\,\mathrm{d}F_\xi(x).$$

■

Corollary 3.2.8 If ξ and η are independent absolutely continuous random variables, then the density of $\xi+\eta$ is given by the convolution of the densities,

$$f_{\xi+\eta}(z) = (f_\xi \star f_\eta)(z) = \int_{-\infty}^{\infty} f_\xi(z-y) f_\eta(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} f_\eta(z-x) f_\xi(x)\,\mathrm{d}x.$$

Example 3.2.9

1. Let $\xi \sim \mathsf{N}\left(m_1, \sigma_1^2\right)$ and $\eta \sim \mathsf{N}\left(m_2, \sigma_2^2\right)$, so that

$$f_\xi(x) = \frac{1}{\sigma_1}\varphi\left(\frac{x-m_1}{\sigma_1}\right)$$

and

$$f_\eta(x) = \frac{1}{\sigma_2}\varphi\left(\frac{x-m_2}{\sigma_2}\right),$$

where

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

Then

$$\begin{aligned} f_{\xi+\eta}(z) &= \int_{-\infty}^{\infty} f_\eta(z-x) f_\xi(x)\,\mathrm{d}x \\ &= \frac{1}{\sqrt{\sigma_1^2+\sigma_2^2}}\varphi\left(\frac{z-(m_1+m_2)}{\sqrt{\sigma_1^2+\sigma_2^2}}\right). \end{aligned}$$

Therefore, $\xi+\eta \sim \mathsf{N}\left(m_1+m_2, \sigma_1^2+\sigma_2^2\right)$.

2. Let $\xi_1, \ldots, \xi_n$ be independent $\mathsf{N}(0,1)$ random variables. Then

$$f_{\xi_1^2+\cdots+\xi_n^2}(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & x > 0, \\ 0 & x \le 0. \end{cases} \tag{3.2}$$

The random variable $\xi_1^2+\cdots+\xi_n^2$ is usually denoted by χ_n^2 and its distribution is the χ^2-distribution with n degrees of freedom. To show this one can proceed by induction.

- For $n=1$ we can use statement 1 of Example 3.1.6.
- Suppose $f_{\xi_1^2+\ldots\xi_{n-1}^2}(x)$ has the form of (3.2). Then,

$$\begin{aligned} f_{\xi_1^2+\cdots+\xi_n}(x) &= \int_{-\infty}^{\infty} f_{\xi_n^2}(x-z) f_{\xi_1^2+\cdots+\xi_{n-1}^2}(z)\,\mathrm{d}z \\ &= \int_0^x \frac{1}{\sqrt{2\pi(x-z)}} e^{-\frac{x-z}{2}} \frac{1}{2^{\frac{n-1}{2}}\Gamma\left(\frac{n-1}{2}\right)} z^{\frac{n-1}{2}-1} e^{-\frac{z}{2}}\,\mathrm{d}z \\ &= \frac{e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)} \int_0^x \frac{z^{\frac{n-1}{2}-1}}{\sqrt{x-z}}\,\mathrm{d}z \\ &\overset{z=tx}{=} \frac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)} \int_0^1 \frac{t^{\frac{n-1}{2}-1}}{\sqrt{1-t}}\,\mathrm{d}t \end{aligned}$$

$$\begin{aligned}
&\stackrel{t=y^2}{=} \frac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)}\int_0^1 \frac{2y^{n-1}}{\sqrt{1-y^2}}\,\mathrm{d}y\\
&\stackrel{y=\sin(\theta)}{=} \frac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)}\int_0^{\frac{\pi}{2}} 2\sin^{n-2}(\theta)\,\mathrm{d}\theta\\
&= \frac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\sqrt{\pi}\Gamma\left(\frac{n-1}{2}\right)}\begin{cases}\frac{(n-3)(n-5)\dots(1)}{(n-2)(n-4)\dots(2)}\frac{\pi}{2} & n \text{ even}\\ \frac{(n-3)(n-1)\dots(2)}{(n-2)(n-4)\dots(3)} & n \text{ odd}\end{cases}\\
&= \frac{x^{\frac{n}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)},
\end{aligned}$$

where in the last equality we use the identity

$$\Gamma(n-1)\Gamma\left(n-\frac{1}{2}\right) = 2^{3-2n}\sqrt{\pi}\Gamma(2n-2)$$

to make the simplification.

Proposition 3.2.10 Let ξ and η be independent random variables with $\mathbb{E}(\xi) < \infty$ and $\mathbb{E}(\eta) < \infty$. Then $\mathbb{E}(\xi\eta) < \infty$ with $\mathbb{E}(\xi\eta) = \mathbb{E}(\xi)\mathbb{E}(\eta)$.

Proof. Suppose that ξ and η are non-negative. Let

- $\xi_n = \sum_{k=0}^{\infty} \frac{k}{n}\chi_{\left\{\frac{k}{n}\le\xi(\omega)<\frac{k+1}{n}\right\}}$, and
- $\eta_n = \sum_{k=0}^{\infty} \frac{k}{n}\chi_{\left\{\frac{k}{n}\le\eta(\omega)<\frac{k+1}{n}\right\}}$.

Then $\xi_n \le \xi$ and $\eta_n \le \eta$ with $|\xi-\xi_n| \le \frac{1}{n}$ and $|\eta-\eta_n| \le \frac{1}{n}$ for all $n \in \mathbb{N}$. Since ξ and η are integrable we can apply the dominated convergence theorem to deduce that

- $\lim_{n\to\infty}\mathbb{E}(\xi_n) = \mathbb{E}(\xi)$, and
- $\lim_{n\to\infty}\mathbb{E}(\eta_n) = \mathbb{E}(\eta)$.

Hence,

$$\begin{aligned}
\mathbb{E}(\xi_n\eta_n) &\stackrel{(1)}{=} \sum_{j,k\ge 0}\frac{jk}{n^2}\mathbb{E}\left(\chi_{\left\{\frac{j}{n}\le\xi<\frac{j+1}{n}\right\}}\chi_{\left\{\frac{k}{n}\le\eta<\frac{k+1}{n}\right\}}\right)\\
&\stackrel{(2)}{=} \sum_{j,k\ge 0}\frac{jk}{n^2}\mathbb{E}\left(\chi_{\left\{\frac{j}{n}\le\xi<\frac{j+1}{n}\right\}}\right)\mathbb{E}\left(\chi_{\left\{\frac{k}{n}\le\eta<\frac{k+1}{n}\right\}}\right)\\
&= \mathbb{E}(\xi_n)\mathbb{E}(\eta_n),
\end{aligned}$$

where (1) is an application of the monotone convergence theorem, and (2) follows from independence. Moreover,

$$\begin{aligned}
|\mathbb{E}(\xi\eta)-\mathbb{E}(\xi_n\eta_n)| &\le \mathbb{E}(|\xi\eta-\xi_n\eta_n|)\\
&= \mathbb{E}(|\xi(\eta-\eta_n)+\eta_n(\xi-\xi_n)|)\\
&\le \frac{1}{n}\mathbb{E}(|\xi|)+\frac{1}{n}\mathbb{E}\left(|\eta|+\frac{1}{n}\right)\\
&\to 0
\end{aligned}$$

as $n\to\infty$. Therefore,

$$\mathbb{E}(\xi\eta) = \lim_{n\to\infty}\mathbb{E}(\xi_n\eta_n) = \lim_{n\to\infty}\mathbb{E}(\xi_n)\lim_{n\to\infty}\mathbb{E}(\eta_n) = \mathbb{E}(\xi)\mathbb{E}(\eta),$$

and $\mathbb{E}(\xi\eta) < \infty$. For general random variables use the representations

- $\xi = \xi^+ - \xi^-$, and
- $\eta = \eta^+ - \eta^-$.

∎

Proposition 3.2.11 Integrable random variables ξ and η are independent if and only if for all Borel-measurable functions f and g we have $\mathbb{E}(f(\xi)g(\eta)) = \mathbb{E}(f(\xi))\mathbb{E}(g(\eta))$.

Proof. ($\Rightarrow$). Step 1: Let $f = \chi_{B_1}$ and $g = \chi_{B_2}$ for $B_1, B_2 \in \mathscr{B}(\mathbb{R})$.
In this case,

$$\begin{aligned}\mathbb{E}(f(\xi)g(\eta)) &= \mathbb{E}(\chi_{\{\xi\in B_1\}}\chi_{\{\eta\in B_2\}})\\ &= \mathbb{E}(\chi_{\{\xi\in B_1,\eta\in B_2\}})\\ &= \mathbb{P}(\xi\in B_1,\eta\in B_2)\\ &= \mathbb{P}(\xi\in B_1)\mathbb{P}(\eta\in B_2)\\ &= \mathbb{E}(\chi_{\{\xi\in B_1\}})\mathbb{E}(\chi_{\{\eta\in B_2\}})\\ &= \mathbb{E}(f(\xi))\mathbb{E}(g(\eta)).\end{aligned}$$

Step 2: Let f and g be simple functions.
By the linearity of the expectations it follows from step 1 that $\mathbb{E}(f(\xi)g(\eta)) = \mathbb{E}(f(\xi))\mathbb{E}(g(\eta))$.
Step 3: Let f and g be non-negative measurable functions.
There exist sequences of simple functions $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$ such that $f_n(x)\to f(x)$ and $g_n(x)\to g(x)$ with $f_n(x)\le f_{n+1}(x)$ and $g_n(x)\le g_{n+1}(x)$. Therefore, by monotone convergence and step 2 it follows that

$$\begin{aligned}\mathbb{E}(f(\xi)g(\eta)) &= \mathbb{E}(\lim_{n\to\infty} f_n(\xi)g_n(\eta))\\ &= \lim_{n\to\infty}\mathbb{E}(f_n(\xi)g_n(\eta))\\ &= \lim_{n\to\infty}\mathbb{E}(f_n(\xi))\mathbb{E}(g_n(\eta))\\ &= \mathbb{E}(f(\xi))\mathbb{E}(g(\eta)).\end{aligned}$$

Step 4: Let f and g be arbitrary measurable functions.
Using the decompositions $f = f^+ - f^-$ and $g = g^+ - g^-$, and the linearity of expectation, it follows from step 3 that $\mathbb{E}(f(\xi)g(\eta)) = \mathbb{E}(f(\xi))\mathbb{E}(g(\eta))$.
($\Leftarrow$). For any $B_1, B_2 \in \mathscr{B}(\mathbb{R})$ let $f = \chi_{B_1}$ and $g = \chi_{B_2}$. Then

$$\mathbb{P}(\xi\in B_1,\eta\in B_2) = \mathbb{E}(f(\xi)g(\eta)) = \mathbb{E}(f(\xi))\mathbb{E}(g(\eta)) = \mathbb{P}(\xi\in B_1)\mathbb{P}(\eta\in B_2)$$

Therefore, ξ and η are independent. ■

3.3 Correlation

Definition 3.3.1 Let ξ and η be random variables defined on the same probability space. Provided that their expectations exist, their covariance is

$$\mathsf{Cov}(\xi,\eta) := \mathbb{E}((\xi-\mathbb{E}(\xi))(\eta-\mathbb{E}(\eta))).$$

Remark 3.3.2 Note that

$$\mathbb{V}(\xi+\eta) = \mathbb{V}(\xi) + \mathbb{V}(\eta) + 2\mathsf{Cov}(\xi,\eta).$$

Hence, $\mathsf{Cov}(\xi,\eta) = 0$ implies that $\mathbb{V}(\xi+\eta) = \mathbb{V}(\xi)+\mathbb{V}(\eta)$.

Definition 3.3.3 Random variables ξ and η are uncorrelated if

$$\mathsf{Cov}(\xi,\eta) = 0.$$

Corollary 3.3.4 Independent random variables are uncorrelated.

Proof. Using Proposition 3.2.10 we deduce that

$$\mathsf{Cov}(\xi,\eta) = \mathbb{E}(\xi\eta) - \mathbb{E}(\xi)\mathbb{E}(\eta) = 0.$$

■

Remark 3.3.5 The converse of Corollary 3.3.4 is not true.

Example 3.3.6 Consider a random variable α which takes the values $\{0, \frac{\pi}{2}, \pi\}$ uniformly. Then $\xi = \sin(\alpha)$ and $\eta = \cos(\alpha)$ are uncorrelated since

$$\begin{aligned}\text{Cov}(\xi, \eta) &= \mathbb{E}(\xi\eta) - \mathbb{E}(\xi)\mathbb{E}(\eta) \\ &= 0 - \left(\frac{1}{3}\right)(0) \\ &= 0.\end{aligned}$$

However, they are not independent since

$$\mathbb{P}(\xi = 1, \eta = 1) = 0 \neq \frac{1}{9} = \mathbb{P}(\xi = 1)\mathbb{P}(\eta = 1).$$

On the other hand, the random variables ξ and η^2 are correlated since

$$\begin{aligned}\text{Cov}\left(\xi, \eta^2\right) &= \mathbb{E}\left(\xi\eta^2\right) - \mathbb{E}(\xi)\mathbb{E}\left(\eta^2\right) \\ &= 0 - \left(\frac{1}{3}\right)\left(\frac{2}{3}\right) \\ &= -\frac{2}{9}.\end{aligned}$$

Part II. Concepts of Convergence

4 Convergence in Probability

We now have sufficient tools from measure theory to get into the first serious topic of probability, limiting theorems. Given a random sequence $(\xi_1, \xi_2, \dots)$ with each ξ_n independent and identically distributed, we would like to study the deviation between the empirical mean $\frac{S_n}{n}$, where $S_n = \xi_1 + \dots + \xi_n$, and the expectation $\mathbb{E}(\xi_1)$ as $n \to \infty$.

4.1 Definition and Properties

We have already encountered one form of convergence, namely L^p convergence. Here we introduce an alternative notion of convergence, known as convergence in probability.

Definition 4.1.1 A sequence $(\xi_n)_{n\in\mathbb{N}}$ of random variables from a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ to $\mathbb{R}$ converges in probability, or in measure $\mathbb{P}$, to the random variable ξ, denoted by $\xi_n \xrightarrow{p} \xi$, if for every $\varepsilon > 0$ we have

$$\mathbb{P}(|\xi_n - \xi| > \varepsilon) \to 0,$$

as $n \to \infty$.

Proposition 4.1.2 Let ξ and $(\xi_n)_{n\in\mathbb{N}}$ be random variables on $(\Omega, \mathscr{F}, \mathbb{P})$. If $\xi_n \xrightarrow{L^p} \xi$ for $p \geq 1$, then $\xi_n \xrightarrow{p} \xi$.

Proof. The random variable $|\xi_n - \xi|^p$ is non-negative. So for any $\epsilon > 0$, by Markov's inequality, it follows that

$$\begin{aligned}
\mathbb{P}\left(|\xi_n - \xi| \geq \epsilon\right) &= \mathbb{P}\left(|\xi_n - \xi|^p \geq \epsilon^p\right) \\
&\leq \frac{\mathbb{E}\left(|\xi_n - \xi|^p\right)}{\epsilon^p} \\
&\xrightarrow{n\to\infty} 0.
\end{aligned}$$

Therefore, $\xi_n \xrightarrow{p} \xi$. ■

Example 4.1.3 The converse of Proposition 4.1.2 is not true. Consider $\xi_n = n\chi_{(0,\frac{1}{n}]}(\omega)$. Then

$$\mathbb{P}(|\xi_n| > \epsilon) = \frac{1}{n} \xrightarrow{n\to\infty} 0.$$

Therefore, $\xi_n \xrightarrow{p} 0$. However,

$$\mathbb{E}\left(|\xi_n|^p\right) = n^{p-1} \overset{n\to\infty}{\not\to} 0.$$

Therefore, $\xi_n \overset{L^p}{\not\to} 0$.

Exercise 4.1.4 Let $(\xi_n)_{n\in\mathbb{N}}$ and $(\eta_n)_{n\in\mathbb{N}}$ be sequences of random variables on $(\Omega, \mathscr{F}, \mathbb{P})$, and let ξ, η be random variables on $(\Omega, \mathscr{F}, \mathbb{P})$.

1. Check that the limit of convergence in probability is almost surely unique. That is, if ξ_n converges in probability to ξ and ξ' then $\xi = \xi'$ almost surely.

2. Prove that if $\xi_n \xrightarrow{p} \xi$ and $\eta_n \xrightarrow{p} \eta$ then for all real numbers a, b we have $a\xi_n + b\eta_n \xrightarrow{p} a\xi + b\eta$.

3. Prove that if $\xi_n \xrightarrow{p} \xi$ and $\eta_n \xrightarrow{p} \eta$ then $\xi_n\eta_n \xrightarrow{p} \xi\eta$. Notice this is not necessarily true if we have L^p convergence. What's wrong with the argument, and can we refine the statement?

4. Show that if $\xi_n \xrightarrow{p} \xi$, $\eta_n \xrightarrow{p} \eta$ and $\varphi(x, y)$ is a continuous function, then $\varphi(\xi_n, \eta_n) \xrightarrow{p} \varphi(\xi, \eta)$.

4.2 Coin Flipping Example

We can motivate the notion of convergence in probability by considering flipping n independent coins. Firstly, consider the flipping of just one coin. Assume the outcomes are zero and one with a probability $p \in (0,1)$ of getting a one. We aim to express this as a $\{0,1\}$-valued random variable ξ on a suitable probability space $(\Omega, \mathscr{A}, \mathbb{P})$, such that $\mathbb{P}_\xi(\{0\}) = 1-p$ and $\mathbb{P}_\xi(\{1\}) = p$, or in other words,

$$\mathbb{P}_\xi(\{x\}) = p^x(1-p)^{1-x}, \tag{4.1}$$

for $x = 0, 1$. For the probability space, there are several choices.

- The natural choice would be
  - $\Omega = \{0, 1\}$,
  - $\mathscr{A} = 2^\Omega$, and
  - $\mathbb{P}(\{\omega\}) = p^\omega(1-p)^{1-\omega}$.

  In this case, our desired random variable would be $\eta(\omega) = \omega$.
- A more complicated choice would be
  - $\Omega = [0, 1]$,
  - $\mathscr{A} = \mathscr{B}([0, 1])$, and
  - $\mathbb{P}(E) = \mathsf{Leb}(E)$.

  In this case, our desired random variable would be $\xi(\omega) = \chi_{(p,1]}(\omega)$.

Remark 4.2.1 The more complicated formulation represents how a computer simulates the flipping of a biased coin. First, it generates a random number $r \in [0,1]$ from a uniform distribution (for instance, by using the `numpy.random.rand` function in Python), then returns zero if $r < 1-p$ and one otherwise.

In both cases we see that $\mathbb{P}_\xi(\{0\}) = 1-p$ and $\mathbb{P}_\xi(\{1\}) = p$. In fact, from Theorem 2.2.10 we know that the distribution functions, and therefore the expectation, will not depend on our choice of probability spaces and random variables, as long as $\mathbb{P}_\xi$ satisfies (4.1). How can we extend the experiment to the flipping of n coins? It would be wrong to define $\xi_1, \dots, \xi_n$ on the same sample space $(\Omega, \mathscr{A}, \mathbb{P})$ such that $\xi_1 = \cdots = \xi_n$. This would correspond to the flipping of a single coin once and recording the result n times. It is clear that in such a construction the random variables would not be independent. It will be hard to write down a large number of independent random variables defined on any of the sample spaces $(\Omega, \mathscr{A}, \mathbb{P})$ in the above example. A standard way of describing n independent coin flips (or n independent trials in general) is to assume that the random variables $\xi_1, \dots, \xi_n$ lie in different probability spaces. That is, ξ_1 is defined on $(\Omega_1, \mathscr{A}_1, \mathbb{P}_1)$, ξ_2 is defined on $(\Omega_2, \mathscr{A}_2, \mathbb{P}_2)$ and so on. Proceeding in this way requires us to operate in these different spaces simultaneously. Thus we need to consider the product space $\left(\Omega^{(n)}, \mathscr{A}^{(n)}, \mathbb{P}^{(n)}\right) = \otimes_{i=1}^n (\Omega_i, \mathscr{A}_n, \mathbb{P}_n)$. The sample space of this product space is

$$\Omega^{(n)} = \Omega_1 \times \cdots \times \Omega_n = \{\omega = (\omega_1, \dots, \omega_n) : \omega_i \in \Omega_i, \text{ for } i = 1, \dots, n\},$$

the collection of events are

$$\mathscr{A}^{(n)} = \sigma\left(\{A_1 \times \cdots \times A_n : A_i \in \mathscr{A}_i, \text{ for } i = 1, \dots, n\}\right),$$

and the probability measure $\mathbb{P}^{(n)}$ satisfies

$$\mathbb{P}^{(n)}(A_1 \times \cdots \times A_n) = \prod_{i=1}^n \mathbb{P}_i(A_i).$$

Consequently, we define the family of projection functions onto the i^{th} component, $\text{proj}_i^{(n)} : \left(\Omega^{(n)}, \mathscr{A}^{(n)}\right) \to (\Omega_i, \mathscr{A}_i)$ such that $\text{proj}_i^{(n)}(\omega_1, \dots, \omega_n) = \omega_i$. For convenience, we drop the superscript (n) if there is no ambiguity. Notice that the projection functions are measurable since the pre-image of any set in $\mathscr{A}_i$ is

$$\text{proj}_i^{-1}(A_i) = \Omega_1 \times \cdots \times \Omega_{i-1} \times A_i \times \Omega_{i+1}, \cdots \times \Omega_n \in \mathscr{A}^{(n)}.$$

We can define the random variables $\tilde{\xi}_i : \left(\Omega^{(n)}, \mathscr{A}^{(n)}, \mathbb{P}^{(n)}\right) \to \{0, 1\}$ such that $\tilde{\xi}_i(\omega) = \xi_i(\text{proj}_i(\omega))$. These random variables are an accurate description of flipping n coins.

Exercise 4.2.2 Verify the following.

1. The marginal distribution of ξ_i, defined as the measure $A \mapsto \mathbb{P}_{\tilde{\xi}_i}(\Omega_1 \times \cdots \times A \times \cdots \times \Omega_n)$ satisfies (4.1).
2. The family $\left(\tilde{\xi}_i\right)_{i=1}^n$ of random variables is independent.

Next, we want to extend the above construction to infinity. More specifically, we want to consider the space $(\Omega, \mathscr{A}) = \otimes_{i=1}^{\infty}(\Omega_i, \mathscr{A}_i)$ with a suitable probability measure $\mathbb{P}$, so that we can discuss large-sample theorems. It should be the case that this probability measure satisfies

$$\mathbb{P}\left(A_1 \times \cdots \times A_n \times \Omega_{n+1} \times \dots\right) = \prod_{i=1}^{n} \mathbb{P}_i(A_i).$$

If we use $\Omega_i \equiv \Omega$ and $\mathscr{A}_i \equiv \mathscr{A}$ in our above examples and assume the natural choice, then we can safely set

$$\mathbb{P}\left(\{(\omega_1, \omega_2, \dots)\}\right) = \prod_{i=1}^{\infty} p^{\omega_i}(1-p)^{1-\omega_i} = p^{\sum_{i \in \mathbb{N}} \omega_i}(1-p)^{\sum_{i \in \mathbb{N}}(1-\omega_i)},$$

since the probability measure is well-defined for all singletons $\{(\omega_1, \omega_2, \dots)\}$. If we use the example when $\Omega_i = [0, 1]$, we can check that our sequence of measures $\left(\mathbb{P}^{(n)}\right)_{n \in \mathbb{N}}$ is consistent and we can apply Theorem 1.6.2 to construct $\mathbb{P}$.

The above shows that we need not worry about specifying a single probability space to describe a sequence of independent experiments. If we want to describe an infinite sequence of experiments with an underlying distribution $\mathbb{P}_\xi$, we can consider the infinite product space $(\mathbb{R}^\infty, \mathscr{B}(\mathbb{R}^\infty))$ equipped with the probability measure $\mathbb{P}$ as determined by Theorem 1.6.2. The projections onto the i^{th} component proj_i are random variables with distribution $\mathbb{P}_\xi$. From now on, we abuse notation by not mentioning the underlying probability space, dropping the tilde sign above ξ and interpreting any operations in the above sense.

4.3 Bernoulli's Law of Large numbers

With ξ_k a $\{0, 1\}$-valued random variable, taking value one with probability $p \in (0, 1)$, let $S_n = \xi_1 + \cdots + \xi_n$. Then

$$\mathbb{E}(S_n) = \sum_{k=1}^{n} \mathbb{E}(\xi_k) = \sum_{k=1}^{n} \left(1 \cdot \mathbb{P}_{\xi_k}(\xi_k = 1) + 0 \cdot \mathbb{P}_{\xi_k}(\xi_k = 0)\right) = np.$$

Thus the mean value of $\frac{S_n}{n}$ is equal to p. The question now is what does $|\frac{1}{n}S_n(\omega) - p|$ converge to for large n? Moreover, in what sense does this convergence occur? It cannot be that

$$\left|\frac{S_n(\omega)}{n} - p\right| \to 0$$

uniformly or pointwise in ω, because there is always an ω such that $\xi_k(\omega) = 1$ for all $k \in \mathbb{N}$, so $\frac{1}{n}S_n(\omega) \equiv 1 \nrightarrow p$. Therefore, we must consider a weaker notion of convergence, as illustrated in the following exercise.

Exercise 4.3.1 Verify that $S_n \sim \mathsf{B}(n,p)$. Hence, show that

$$\left\|\frac{S_n}{n} - p\right\|_{L^2}^2 \to 0$$

as $n \to \infty$.

With a more general analysis, we have L^p convergence for any $p \in [1,\infty)$. Moreover, by Chebyshev's inequality, it follows for any fixed $\epsilon > 0$ that

$$\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \epsilon\right) \leq \frac{\mathbb{V}\left(\frac{S_n}{n}\right)}{\epsilon^2} = \frac{p(1-p)}{n\epsilon^2}. \tag{4.2}$$

Hence, we can always make the probability $\mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > \epsilon\right)$ arbitrarily small. That is, we have convergence in probability. Stating the above formally gives us the Bernoulli Law of Large numbers.

Theorem 4.3.2 — Bernoulli Law of Large Numbers. Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed Bernoulli random variables, with parameter $p \in (0,1)$. Then $\frac{S_n}{n}$ converges in probability to p.

4.4 Weak Law of Large Numbers

We can generalise the ideas of the previous section. Note, from Markov's inequality, if $\xi_n \to \xi$ in L^p for $p \geq 1$, then we must have $\xi_n \xrightarrow{p} \xi$ because

$$\mathbb{P}(|\xi_n - \xi| > \varepsilon) \leq \frac{\|\xi_n - \xi\|_{L^p}^p}{\varepsilon^p} \to 0.$$

Let $\xi_1, \dots, \xi_n$ be random variables and let

$$S_n^{(c)} = \sum_{k=1}^n (\xi_k - \mathbb{E}(\xi_k)).$$

Observe that $\mathbb{E}\left(S_n^{(c)}\right) = 0$. How can we use Chebyshev's inequality to make the weakest assumptions on the $\xi_1, \dots, \xi_n$ such that $S_n^{(c)}$ converges? For simplicity, we can first assume that the $\xi_k \in L^2$ with $\mathbb{V}(\xi_k) \leq C$ for some constant C that is independent of $k \in \mathbb{N}$. Then, if we assume that the $\xi_1, \dots, \xi_n$ are pairwise uncorrelated, which is a weaker assumption than independence, we have

$$\mathbb{V}(S_n) = \sum_{k=1}^n \mathbb{V}(\xi_k) \leq Cn.$$

Theorem 4.4.1 — L^2 Weak Law of Large Numbers. Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of uncorrelated L^2 random variables such that $\mathbb{V}(\xi_n) \leq C$ for some $C > 0$ and every $n \in \mathbb{N}$. Then

$$\frac{S_n^{(c)}}{n} \xrightarrow{p} 0.$$

Proof. Using Chebyshev's inequality and the fact the random variables are uncorrelated, it follows for all $\epsilon > 0$ that

$$\mathbb{P}\left(\left|\frac{S_n^{(c)}}{n}\right| > \varepsilon\right) \leq \frac{\mathbb{V}\left(\frac{S_n^{(c)}}{n}\right)}{\varepsilon^2} \leq \frac{C}{n\varepsilon^2} \xrightarrow{n\to\infty} 0.$$

■

Corollary 4.4.2 Let $(\xi_n)_{n\in\mathbb{N}}$ be integrable, independent and identically distributed random variables,

such that $\mathbb{V}(\xi_n) < \infty$. Then
$$\frac{S_n}{n} \xrightarrow{p} \mathbb{E}(\xi_1).$$

Example 4.4.3 Consider f a measurable function on $[0,1]$ with
$$C := \int_0^1 |f(x)|^2 \, \mathrm{d}x < \infty.$$
The integral
$$\theta = \int_0^1 f(x) \, \mathrm{d}x,$$
is often intractable to compute in practice. However, if U is a random variable on some probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with uniform distribution $\mathsf{Unif}[0,1]$, then $\theta = \mathbb{E}(f(U))$. For example, one could take $U : \omega \in ([0,1], \mathscr{B}[0,1], \mathsf{Leb}) \mapsto \omega$. Therefore, an approach to estimating θ would be to use an empirical mean approach.

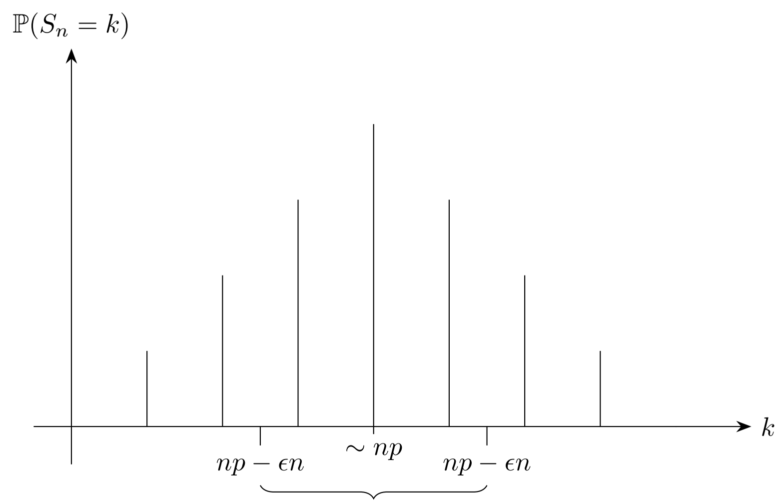
1. Let $(U_n(\omega))_{n\in\mathbb{N}}$ be an independent and identically distributed sample from $U \sim \mathsf{Unif}[0,1]$.
2. Then evaluate the empirical mean
$$\hat{\theta}_n(\omega) := \frac{f(U_1(\omega)) + \cdots + f(U_n(\omega))}{n}.$$

It is clear that $\mathbb{E}\left(\hat{\theta}_n\right) = \theta$. Moreover, since the random variables $f(U_i)$ are independent and identically distributed with finite variance
$$\mathbb{V}(f(U_i)) = \mathbb{E}\left((f(U_i))^2\right) - \theta^2 = C - \theta^2 < \infty,$$
Theorem 4.4.1 tells us that $\hat{\theta}_n(\omega) \xrightarrow{n\to\infty} \theta$ in probability.

Exercise 4.4.4 — L^2 Weak Law of Large Number for weakly correlated random variables. Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence random variables on a common probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with $\mathbb{E}(\xi_i) = 0$ and $\mathbb{E}(\xi_i \xi_j) = r_{|i-j|}$, where $(r_k)_{k\in\mathbb{N}}$ is a sequence of real numbers such that $r_k \xrightarrow{k\to\infty} 0$. Let $S_n = \sum_{i=1}^n \xi_i$. Show that $\frac{S_n}{n} \xrightarrow{n\to\infty} 0$ in probability.

4.5 Local and Central Limit Theorem

We return to the coin-flipping scenario. Let $S_n = \xi_1 + \cdots + \xi_n$, where $\xi_k \sim \mathsf{B}(1,p)$ are independent and identically distributed. As discussed in the previous section, S_n tends to be close to np for large n.



Specifically, let us take some interval $\mathscr{I}_n = (n(p-\varepsilon), n(p+\varepsilon))$. If we pick some sufficiently large ε, say $\varepsilon = n^\alpha$ where $\alpha > \frac{1}{2}$, then by Chebyshev's inequality (4.2) we know that

$$\mathbb{P}(S_n \in \mathscr{I}_n^c) = \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| > n^{\alpha-1}\right) \leq \frac{p(1-p)}{n^{1+2\alpha-2}} = \frac{p(1-p)}{n^{2\alpha-1}} \xrightarrow{n\to\infty} 0.$$

These decaying bounds of $\mathbb{P}(S_n \in \mathscr{I}_n^c)$ no longer exist when $\alpha \leq \frac{1}{2}$. How much do we know about $\mathbb{P}(S_n \in \mathscr{I}_n^c)$ for this case? Will it tend to some non-trivial constant in $(0,1)$, or will it increase and tend to one? We will see that the central limit theorem suggests that at the boundary $\alpha = \frac{1}{2}$, the quantity $\mathbb{P}(S_n \in \mathscr{I}_n^c)$ tends to some non-trivial constant in $(0,1)$. This suggests that the rescaled mean $\frac{S_n}{\sqrt{n}}$ will converge in some way to a non-trivial distribution.

Definition 4.5.1 Consider sequences $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$.

- Big O notation. We say that $g_n = O(f_n)$ as $n \to \infty$ if $\left|\frac{g_n}{f_n}\right|$ is bounded for sufficiently large n. That is, there exists constant $C > 0$ and $N \in \mathbb{N}$ such that for all $n \geq N$ we have $|g_n| \leq C|f_n|$.
- Small o notation. We say that $g_n = o(f_n)$ as $n \to \infty$ if $\left|\frac{g_n}{f_n}\right| \to 0$ as $n \to \infty$. In other words, for all $\epsilon > 0$, there exists $N := N(\epsilon) \in \mathbb{N}$ such that for all $n \geq N$ we have $|g_n| \leq \epsilon|f_n|$. We sometimes write $g_n \ll f_n$ or $f_n \gg g_n$.
- Asymptotic equivalence. We write $g_n \sim f_n$ if $\left|\frac{g_n}{f_n}\right| \to 1$ as $n \to \infty$. Equivalently, we have $g_n = (1+o(1))f_n$.
- Order. We write $g_n = \Theta(f_n) = \mathrm{ord}(f_n)$ if $g_n = O(f_n)$ but g_n is not $o(f_n)$.

Remark 4.5.2

- The use of equal sign is an abuse of notation.
- The statements of Definition 4.5.1 can be extended to any functions $f(x)$ defined on real or complex numbers, in such case, we can assume x tends to some point x_0 including ∞.
- We can also consider order notation for sequences of functions. Let $g_n = g_n(\alpha)$ and $f_n = f_n(\alpha)$. We say $g_n(\alpha) = O(|f_n(\alpha)|)$ uniformly if the statement 1 of Definition 4.5.1 holds for constants $C > 0$ and $N \in \mathbb{N}$ independent of α. We also have analogous definition for $g_n(\alpha) = o(|f_n(\alpha)|)$.

Lemma 4.5.3 — Stirling's Approximation. As $n \to \infty$, we have

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \left(1 + O\left(\frac{1}{n}\right)\right).$$

Using Lemma 4.5.3 we can sketch an asymptotic analysis of the binomial coefficient. Suppose $n, k, n-k$ tend to infinity, for example let $k = np$ for $p \in (0,1)$. With Lemma 4.5.3 we obtain

$$\begin{aligned}
\binom{n}{k} &= \frac{n!}{k!(n-k)!} \\
&= \frac{\sqrt{n}}{\sqrt{2\pi k(n-k)}} \frac{\left(\frac{n}{e}\right)^n}{\left(\frac{k}{e}\right)^k \left(\frac{n-k}{e}\right)^{n-k}} \frac{1 + O\left(\frac{1}{n}\right)}{\left(1 + O\left(\frac{1}{n}\right)\right)\left(1 + O\left(\frac{1}{n}\right)\right)} \\
&= \frac{\sqrt{n}}{\sqrt{2\pi k(n-k)}} \exp\left(n\log(n) - k\log(k) - (n-k)\log(n-k)\right) \frac{1 + O\left(\frac{1}{n}\right)}{\left(1 + O\left(\frac{1}{n}\right)\right)\left(1 + O\left(\frac{1}{n}\right)\right)}.
\end{aligned}$$

To gain intuition, one can treat $O\left(\frac{1}{n}\right)$ as being exactly equal to $\frac{1}{n}$. That is,

$$\left(1 + O\left(\frac{1}{n}\right)\right)^{-2} = \left(1 + \frac{1}{n}\right)^{-2} = 1 - \frac{2}{n} + \dots,$$

and so,

$$\frac{1 + O\left(\frac{1}{n}\right)}{\left(1 + O\left(\frac{1}{n}\right)\right)\left(1 + O\left(\frac{1}{n}\right)\right)} = \left(1 + \frac{1}{n}\right)\left(1 - \frac{2}{n} + \dots\right) = 1 - \frac{1}{n} + \dots = 1 + O\left(\frac{1}{n}\right).$$

Putting this together, we have

$$\binom{n}{k} = \frac{\sqrt{n}}{\sqrt{2\pi k(n-k)}} \exp\left(n\log(n) - k\log(k) - (n-k)\log(n-k)\right)\left(1 + O\left(\frac{1}{n}\right)\right). \tag{4.3}$$

Exercise 4.5.4 Show that $\mathbb{P}(S_n = k)$ is monotone in k below and above its point of maximum.

With this, we can prove the local limit theorem, which specifies the local asymptotics of a probability mass distribution at the point $S_n = k$.

Theorem 4.5.5 — Local Limit Theorem. For any $0 < p < 1$, we have

$$\max_{0 \le k \le n}\left|\mathbb{P}(S_n = k) - \frac{1}{\sqrt{2\pi p(1-p)}\sqrt{n}} e^{-\frac{x^2}{2p(1-p)}}\right| = o\left(\frac{1}{\sqrt{n}}\right),$$

as $n \to \infty$ and where $x = x_{k,n} := \dfrac{k - np}{\sqrt{n}}$.

Proof. Consider k such that

$$|x_{k,n}| \le \frac{A_n}{\sqrt{n}}$$

where $A_n = n^\epsilon$ for $\epsilon \in (0,1)$. Then as $k = np + x\sqrt{n}$ we have that

$$k = np\left(1 + O\left(\frac{A_n}{n}\right)\right)$$

which implies that

$$n - k = n(1-p)\left(1 + O\left(\frac{A_n}{n}\right)\right).$$

Hence, k and $n-k$ tend to infinity as $n \to \infty$, which means we can utilise (4.3) to get that

$$\mathbb{P}(S_n = k) = \underbrace{\frac{\sqrt{n}}{\sqrt{2\pi k(n-k)}}}_{\text{(A)}} \underbrace{\exp\left(n\log(n) - k(\log(k) - \log(p)) - (n-k)(\log(n-k) - \log(1-p))\right)}_{\text{(B)}}\left(1 + O\left(\frac{1}{n}\right)\right).$$

Analysing (A) we notice that

$$\begin{aligned}
\text{(A)} &= \frac{\sqrt{n}}{\sqrt{2\pi k(n-k)}} \\
&= \frac{1}{\sqrt{2\pi np(1-p)\left(1+O\left(\frac{A_n}{n}\right)\right)\left(1+O\left(\frac{A_n}{n}\right)\right)}} \\
&= \frac{1}{\sqrt{2\pi np(1-p)}}\left(1+O\left(\frac{A_n}{n}\right)\right).
\end{aligned}$$

Similarly, analysing (B) we notice that

$$\begin{aligned}
\text{(B)} &= \exp\left(n\log(n) - k\left(\log(n)+\log\left(1+\frac{x}{p\sqrt{n}}\right)\right) - (n-k)\left(\log(n)+\log\left(1-\frac{x}{(1-p)\sqrt{n}}\right)\right)\right) \\
&= \exp\left(-\left((np+x\sqrt{n})\log\left(1+\frac{x}{p\sqrt{n}}\right) + (n(1-p)-x\sqrt{n})\log\left(1-\frac{x}{(1-p)\sqrt{n}}\right)\right)\right) \\
&= \exp\Bigg(-\Bigg(np\left(\frac{x}{p\sqrt{n}} - \frac{x^2}{2p^2n} + O\left(\frac{x^3}{n^{\frac{3}{2}}}\right)\right) + \frac{x^2}{p} + O\left(\frac{x^3}{n^{\frac{1}{2}}}\right)\right) \\
&\quad + n(1-p)\left(-\frac{x}{(1-p)\sqrt{n}} - \frac{x^2}{2(1-p)^2n} + O\left(\frac{x^3}{n^{\frac{3}{2}}}\right)\right) + \frac{x^2}{(1-p)} + O\left(\frac{x^3}{n^{\frac{1}{2}}}\right)\Bigg)\Bigg) \\
&= \exp\left(-\frac{x^2}{2p(1-p)} + O\left(\frac{x^3}{\sqrt{n}}\right)\right) \\
&= \exp\left(-\frac{x^2}{2p(1-p)} + O\left(\frac{A_n^3}{n^2}\right)\right) \\
&= \exp\left(-\frac{x^2}{2p(1-p)}\right)\left(1+O\left(\frac{A_n^3}{n^2}\right)\right).
\end{aligned}$$

Hence,

$$\mathbb{P}(S_n = k) = \frac{1}{\sqrt{2\pi p(1-p)n}}\exp\left(-\frac{x^2}{2p(1-p)}\right)\left(1+O\left(\frac{A_n}{n}\right)+O\left(\frac{A_n^3}{n^2}\right)\right).$$

Letting $\varepsilon = \frac{7}{12}$, we get $\frac{A_n^3}{n^2} = n^{-\frac{1}{4}}$ and $\frac{A_n}{n} = n^{-\frac{5}{12}}$, so that

$$\max_{|x|\leq\frac{A_n}{\sqrt{n}}} \mathbb{P}(S_n = k) = \frac{1}{\sqrt{2\pi p(1-p)n}}\exp\left(-\frac{x^2}{2p(1-p)}\right)\underbrace{\left(1+O\left(\frac{1}{n^{\frac{5}{12}}}\right)\right)}_{=1+o\left(\frac{1}{\sqrt{n}}\right)}. \tag{4.4}$$

Now consider $|x| > \frac{A_n}{\sqrt{n}}$. Observe that

$$\begin{aligned}
&\max_{|x|>\frac{A_n}{\sqrt{n}}}\left|\mathbb{P}(S_n = k) - \frac{1}{\sqrt{2\pi p(1-p)n}}\exp\left(-\frac{x^2}{2p(1-p)}\right)\right| \\
&\quad\leq \max_{|x|>\frac{A_n}{\sqrt{n}}}|\mathbb{P}(S_n = k)| + \max_{|x|>\frac{A_n}{\sqrt{n}}}\left|\frac{1}{\sqrt{2\pi p(1-p)n}}\exp\left(-\frac{x^2}{2p(1-p)}\right)\right| \\
&\quad\leq \max|\mathbb{P}(S_n = \lfloor np + A_n\rfloor), \mathbb{P}(S_n = \lceil np - A_n\rceil)| + \frac{1}{\sqrt{2\pi p(1-p)n}}\exp\left(-\frac{A_n^2}{2np(1-p)}\right).
\end{aligned}$$

Where the first bound is an application Exercise 4.5.4. With $A_n = n^{\frac{7}{12}}$, the second term is $o\left(\frac{1}{\sqrt{n}}\right)$. Furthermore, as

$$n^{\frac{1}{12}} - n^{-\frac{1}{2}} = \frac{np + A_n - np - 1}{\sqrt{n}} \leq \frac{\lfloor np + A_n\rfloor - np}{\sqrt{n}} \leq \frac{np + A_n - np}{\sqrt{n}} = n^{\frac{1}{12}}$$

we note that $x_{\lfloor np+A_n\rfloor,k} \sim n^{\frac{1}{12}}$. A similar result holds for $x_{\lceil np+A_n\rceil,k}$, and so the first term is also $o\left(\frac{1}{\sqrt{n}}\right)$. Combining these results with (4.4) completes the proof. ■

Theorem 4.5.5 says that

$$\mathbb{P}\left(\frac{S_n - np}{\sqrt{n}} = x\right) = \frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{2\pi p(1-p)}}e^{-\frac{x^2}{2p(1-p)}} + o(1)\right)$$

as $n \to \infty$. At first glance, you may find this result not useful, as it only tells us that the probability decays to zero at a rate of $O\left(\frac{1}{\sqrt{n}}\right)$. However, since $\dfrac{S_n - np}{np(1-p)}$ seems to converge to a continuous distribution, we really should look at the density by ignoring the $\frac{1}{\sqrt{n}}$. The things inside the square bracket suggest that the density function of $n^{-\frac{1}{2}}(S_n - np)$ converges to the normal distribution $\mathsf{N}(0, p(1-p))$, which is equivalent to the distribution of $\frac{S_n - np}{np(1-p)}$ converging to the standard normal $\mathsf{N}(0,1)$. The above heuristics can be formalised by adding the local probabilities and considering the cumulative distribution function. We therefore arrive at the central limit theorem.

Theorem 4.5.6 — de Moivre-Laplace CLT. For any $0 < p < 1$ and $x \in \mathbb{R}$ we have,

$$\lim_{n\to\infty}\left(\mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \le x\right)\right) = \Phi(x),$$

where

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, \mathrm{d}y,$$

is the density of $\mathsf{N}(0,1)$.

Proof Sketch. Note

$$\begin{aligned}\mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \le x\right) &= \mathbb{P}\left(S_n \le np + x\sqrt{np(1-p)}\right) \\ &= \sum_{k=0}^{\left\lfloor np - n^{\frac{7}{12}}\right\rfloor - 1} \mathbb{P}(S_n = k) + \sum_{k=\left\lfloor np - n^{\frac{7}{12}}\right\rfloor}^{\left\lfloor np + x\sqrt{np(1-p)}\right\rfloor} \mathbb{P}(S_n = k).\end{aligned}$$

The first term is a sum of a polynomial number of terms of order $O\left(\exp\left(-n^{\frac{1}{2}}\right)\right)$ and so will vanish as $n \to \infty$. The second term is a Riemann sum. Writing

$$T_n = \left\{k : \left\lfloor np - n^{\frac{7}{12}}\right\rfloor \le k \le \left\lfloor np + x\sqrt{np(1-p)}\right\rfloor\right\}$$

we have

$$\begin{aligned}\sum_{k\in T_n} \mathbb{P}(S_n = k) &= \sum_{k \in T_n} \frac{1}{\sqrt{n}} \frac{1}{\sqrt{2\pi p(1-p)}} \exp\left(-\frac{1}{2}\left(\frac{k - np}{\sqrt{np(1-p)}}\right)^2\right) \\ &= \sum_{k\in T_n - np} \frac{1}{\sqrt{n}} \frac{1}{\sqrt{2\pi p(1-p)}} \exp\left(-\frac{1}{2}\left(\frac{\frac{k}{\sqrt{n}}}{\sqrt{p(1-p)}}\right)^2\right).\end{aligned}$$

This is almost a Riemann sum on a partition of $(-\infty, x]$ with a mesh size of $\frac{1}{\sqrt{n}}$, however, it is missing some boundary terms. One can show that the boundary terms lead to an $o(1)$ contribution to conclude that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \le x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, \mathrm{d}y.$$

We omit the details here. ∎

Theorem 4.5.6 demonstrates that for a sequence of independent Bernoulli random variables the quantity $\sqrt{n}(\frac{S_n}{n} - p)$ converges in distribution to a random variable with normal distribution $\mathsf{N}(0, p(1-p))$.

Exercise 4.5.7 Using Theorem 4.5.6 prove Theorem 4.3.2.

Remark 4.5.8 Regarding $S_n \sim \mathsf{Bin}(n, p)$, we can consider the behaviour of S_n if we let p be a function of n. Specifically, we let $p = p(n)$ be such that $p(n) \to 0$ and $p(n) \cdot n \to \lambda > 0$ as $n \to \infty$. Then for fixed $k \in \mathbb{N}$ we have

$$\begin{aligned}\mathbb{P}(S_n = k) &= \frac{n!}{k!(n-k)!} p(n)^k (1-p(n))^k \\ &= \frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^k \left(1 - \frac{\lambda}{n} + o\left(\frac{1}{n}\right)\right)^{n-k}.\end{aligned}$$

Therefore,

$$\mathbb{P}(S_n = k) \to \frac{1}{k!}\lambda^k e^{-\lambda},$$

and so $S_n \sim \mathsf{Bin}(n, p(n))$ converges to a $\mathsf{Po}(\lambda)$ distribution.

4.6 Solution to Exercises

Exercise 4.1.4

Solution.

1. Consider the sets $A_k = \{\omega : |\xi(\omega) - \xi'(\omega)| > \frac{1}{k}\}$. Note that $\{\omega : \xi(\omega) \neq \xi'(\omega)\} = \bigcup_{k=1}^{\infty} A_k$. Moreover,

$$\begin{aligned}\mathbb{P}(A_k) &\overset{(1)}{\leq} \mathbb{P}\left(|\xi - \xi_n| + |\xi_n - \xi'| > \frac{1}{k}\right) \\ &\overset{(2)}{\leq} \mathbb{P}\left(|\xi - \xi_n| > \frac{1}{2k}\right) + \mathbb{P}\left(|\xi_n - \xi'| > \frac{1}{2k}\right) \\ &\overset{n\to\infty}{\longrightarrow} 0,\end{aligned}$$

where (1) is an application of the triangle inequality, and in (2) we use the fact that

$$\left\{|\xi - \xi_n| + |\xi_n - \xi'| > \frac{1}{k}\right\} \subseteq \left\{|\xi - \xi_n| > \frac{1}{2k}\right\} \cup \left\{|\xi_n - \xi'| > \frac{1}{2k}\right\}.$$

Therefore, $\mathbb{P}(A_k) = 0$ which implies that

$$\begin{aligned}\mathbb{P}(\xi \neq \xi') &= \mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) \\ &= \lim_{n\to\infty} \mathbb{P}(A_k) \\ &= 0,\end{aligned}$$

where we have used the continuity of the measure $\mathbb{P}$, as $(A_k)_{k\in\mathbb{N}}$ is a sequence of increasing events. We conclude that $\xi = \xi'$ almost everywhere.

2. Using the triangle inequality we have that

$$|a\xi_n + b\eta_n - a\xi - b\eta| \leq |a||\xi_n - \xi| + |b||\eta_n - \eta|.$$

Therefore,

$$\mathbb{P}(|a\xi_n + b\eta_n - a\xi - b\eta| \geq \epsilon) \leq \mathbb{P}(|a||\xi_n - \xi| + |b||\eta_n - \eta| \geq \epsilon).$$

If ω satisfies the inequality on the right-hand side, it must be the case that either

   (a) $|a||\xi_n - \xi| \geq \frac{\epsilon}{2}$, or

   (b) $|b||\eta_n - \eta| \geq \frac{\epsilon}{2}$.

Hence,

$$\mathbb{P}(|a||\xi_n - \xi| + |b||\eta_n - \eta| \geq \epsilon) \leq \mathbb{P}\left(|\xi_n - \xi| \geq \frac{\epsilon}{2|a|}\right) + \mathbb{P}\left(|\eta_n - \eta| \geq \frac{\epsilon}{2|b|}\right).$$

Both terms on the right-hand side tend to zero as $n \to \infty$ by assumption. Therefore,

$$\mathbb{P}(|a\xi_n + b\eta_n - a\xi - b\eta| \geq \epsilon) \overset{n\to\infty}{\longrightarrow} 0,$$

which implies that $a\xi_n + b\eta_n \xrightarrow{p} a\xi + b\eta$.

3. Observe that

$$\mathbb{P}(|\xi_n\eta_n - \xi\eta| \geq \epsilon) \leq \mathbb{P}\left(|\xi_n||\eta_n - \eta| \geq \frac{\epsilon}{2}\right) + \mathbb{P}\left(|\eta||\xi_n - \xi| \geq \frac{\epsilon}{2}\right). \tag{4.5}$$

Consider the first term on the right-hand side, note that given an $M > 0$ it follows that

$$\left\{|\xi_n||\eta_n - \eta| \geq \frac{\epsilon}{2}\right\} \subseteq \{|\xi_n - \xi| \geq 1\} \cup \{|\xi| \geq M\} \cup \left\{|\eta_n - \eta| \geq \frac{\epsilon}{2(M+1)}\right\}.$$

This follows by considering the event than no of the events on the right-hand side hold, then

$$|\xi_n||\eta_n - \eta| \leq (|\xi_n - \xi| + |\xi|)\,|\eta_n - \eta| < (M+1)\frac{\epsilon}{2(M+1)} = \frac{\epsilon}{2}.$$

Therefore,

$$\begin{aligned}\mathbb{P}\left(|\xi_n||\eta_n - \eta| \geq \frac{\epsilon}{2}\right) &\leq \mathbb{P}(|\xi_n - \xi| \geq 1) + \mathbb{P}(|\xi| \geq M) + \mathbb{P}\left(|\eta_n - \eta| \geq \frac{\epsilon}{2(M+1)}\right)\\ &\overset{n\to\infty}{\longrightarrow} \mathbb{P}(|\xi| \geq M)\\ &\overset{M\to\infty}{\longrightarrow} 0.\end{aligned}$$

Similarly, for the second term on the right-hand side of (4.5) we have

$$\begin{aligned}\mathbb{P}\left(|\eta||\xi_n - \xi| \geq \frac{\epsilon}{2}\right) &\leq \mathbb{P}(|\eta| \geq M) + \mathbb{P}\left(|\eta_n - \eta| \geq \frac{\epsilon}{2M}\right)\\ &\overset{n\to\infty}{\longrightarrow} \mathbb{P}(|\eta| \geq M)\\ &\overset{M\to\infty}{\longrightarrow} 0.\end{aligned}$$

Therefore, we can conclude that

$$\mathbb{P}(|\xi_n\eta_n - \xi\eta| \geq \epsilon) \overset{n\to\infty}{\longrightarrow} 0$$

which implies that $\xi_i\eta_i \xrightarrow{p} \xi\eta$.

4. Let $M \in \mathbb{R}$. On $[-M, M]^2$ the function $\varphi(x, y)$ is uniformly continuous. Therefore, given an $\epsilon > 0$ there exists a $\delta > 0$ such that for all $(x, y), (x', y') \in [-M, M]^2$ with $|(x, y) - (x', y')| < \delta$ we have

$$|\varphi(x, y) - \varphi(x', y')| < \epsilon.$$

Therefore,

$$\begin{aligned}\mathbb{P}\Big(&|\varphi(\xi_n, \eta_n) - \varphi(\xi, \eta)| \geq \epsilon \Big| (\xi_n, \eta_n), (\xi, \eta) \in [-M, M]^2\Big)\\ &\leq \mathbb{P}\left(|(\xi_n, \eta_n) - (\xi, \eta)| \geq \delta \Big| (\xi_n, \eta_n), (\xi, \eta) \in [-M, M]^2\right)\\ &\leq \mathbb{P}\left(|\xi_n - \xi| \geq \frac{\delta}{2}\Big| \xi_n, \xi \in [-M, M]\right) + \mathbb{P}\left(|\eta_n - \eta| \geq \frac{\delta}{2}\Big| \eta_n, \eta \in [-M, M]\right)\\ &\overset{n\to\infty}{\longrightarrow} 0.\end{aligned}$$

Sending $M \to \infty$ completes the proof.

■

Exercise 4.2.2

Solution.

1. It is clear that

$$\begin{aligned}\mathbb{P}_{\xi_i}(\{x\}) &= \mathbb{P}_{\tilde{\xi}_i}(\Omega_1 \times \cdots \times \{x\} \times \cdots \times \Omega_n)\\ &= \mathbb{P}^{(n)}(\Omega_1 \times \cdots \times \{x\} \times \cdots \times \Omega_n)\\ &= \mathbb{P}_i(\{x\})\\ &= p^x(1-p)^{1-x}\end{aligned}$$

for $x = 0, 1$.

2. By construction

$$\begin{aligned}\mathbb{P}\left(\tilde{\xi}_1 \in A_1, \ldots, \tilde{\xi}_n \in A_n\right) &= \mathbb{P}^{(n)}(A_1 \times \cdots \times A_n)\\ &= \prod_{i=1}^{n} \mathbb{P}_i(A_i).\end{aligned}$$

Therefore, the $\left(\tilde{\xi}_i\right)$ are independent.

■

Exercise 4.3.1

Solution. As $S_n \sim \mathsf{B}(n,p)$ we know that $\mathbb{E}(S_n) = np$ and $\mathbb{V} = np(1-p)$. Therefore,

$$\begin{aligned}\mathbb{E}\left(\left|\frac{S_n}{n} - p\right|^2\right) &= \mathbb{E}\left(\left|\frac{S_n - np}{n}\right|^2\right)\\ &= \frac{1}{n^2}\mathbb{E}\left((S_n - np)^2\right)\\ &= \frac{1}{n^2}\mathbb{V}(S_n)\\ &= \frac{p(1-p)}{n}\\ &\overset{n\to\infty}{\longrightarrow} 0.\end{aligned}$$

■

Exercise 4.4.4

Solution. Note that as $(r_k)_{k\in\mathbb{N}}$ converges to zero, the sequence is bounded. Suppose that $r_k \leq M$ for all $k \in \mathbb{N}$. Moreover, for any $\delta > 0$ we can find an $N_\delta \in \mathbb{N}$ such that $r_k \leq \delta$ for all $k \geq N_\delta$. Therefore, for an $\epsilon > 0$ and $n > N_\delta$ we have

$$\begin{aligned}\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq \epsilon\right) &\leq \frac{\mathbb{V}(S_n)}{n^2\epsilon^2}\\ &= \frac{\sum_{i,j=1}^{n} \mathbb{E}(\xi_i\xi_j)}{n^2\epsilon^2}\\ &= \frac{(nr_0 + 2(n-1)r_1 + \cdots + 2(n-N_\delta)r_{N_\delta}) + (2(n-N_\delta-1)r_{N_\delta+1} + \cdots + 2(1)r_n)}{n^2\epsilon^2}\\ &\leq \frac{2nN_\delta M + 2\delta(1 + \cdots + n - N_\delta - 1)}{n^2\epsilon^2}\\ &\leq \frac{2nN_\delta M + 2\delta\frac{1}{2}n(n-1)}{n^2\epsilon^2}\\ &\overset{n\to\infty}{\longrightarrow} \frac{\delta}{\epsilon^2}.\end{aligned}$$

As $\delta > 0$ we arbitrary, for a fixed $\epsilon > 0$ we conclude that

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq \epsilon\right) \xrightarrow{n\to\infty} 0.$$

■

Exercise 4.5.4

Solution. Consider the quotient

$$q = \frac{\binom{n}{k+1}p^{k+1}(1-p)^{n-k-1}}{\binom{n}{k}p^k(1-p)^{n-k}} = \frac{(n-k)p}{(k+1)(1-p)}.$$

When $q < 1$ deduce that the binomial probability density function (PDF) is decreasing and when $q > 1$ we deduce that the binomial PDF is increasing. Therefore,

- the PDF is increasing for $k < p(n+1) - 1$, and
- the PDF is decreasing for $k > p(n+1) - 1$.

When $p(n+1)$ is an integer the PDF is maximal for both $(n+1)p$ and $(n+1)p - 1$. ■

Exercise 4.5.7

Solution. Given a $\delta > 0$ there exists an $X_\delta > 0$ such that for all $x > X_\delta$ we have

$$\Phi(x) - \Phi(-x) = \int_{-x}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \geq 1 - \delta.$$

From Theorem 4.5.6 we have that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq x\right) = \Phi(x)$$

which implies that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n}{n} - p \leq \frac{x\sqrt{p(1-p)}}{\sqrt{n}}\right) = \Phi(x).$$

For fixed $\epsilon > 0$, fix $x > X_\delta$ and choose $N \geq \frac{x^2 p(1-p)}{\epsilon^2}$. It follows for $n \geq N$ that

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| < \epsilon\right) &\geq \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \leq \frac{x\sqrt{p(1-p)}}{\sqrt{n}}\right) \\
&= \lim_{n\to\infty} \left(\mathbb{P}\left(\frac{S_n}{n} - p \leq \frac{x\sqrt{p(1-p)}}{\sqrt{n}}\right) - \mathbb{P}\left(\frac{S_n}{n} - p \leq -\frac{x\sqrt{p(1-p)}}{\sqrt{n}}\right)\right) \\
&= \Phi(x) - \Phi(-x) \\
&\geq 1 - \delta.
\end{aligned}$$

Hence,

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{S_n}{n} - p\right| \geq \epsilon\right) < \delta.$$

We conclude that $\frac{S_n}{n} \xrightarrow{p} 0$. ■

5 Almost Sure Convergence

5.1 Definition

It will be useful now to briefly introduce almost sure convergence, and identify how it relates to convergence in probability and L^p convergence. The full theory of almost sure convergence will be discussed in Chapter 7.

Definition 5.1.1 A sequence $(\xi_n)_{n\in\mathbb{N}}$ of random variables on probability space $(\Omega, \mathscr{F}, \mathbb{P})$ converges $\mathbb{P}$-almost surely to the random variable ξ, denoted by $\xi_n \xrightarrow{\text{a.s.}} \xi$, if

$$\mathbb{P}\left(\left\{\omega : \xi_n(\omega) \overset{n\to\infty}{\nrightarrow} \xi(\omega)\right\}\right) = 0.$$

Further discussions will follow, for which the aim is to determine the following implications of convergence.

- Almost sure convergence and convergence in L^p both imply convergence in probability.
- Convergence in probability implies convergence in distribution.

$$\begin{array}{ccccc} & & \xrightarrow{L^p} & & \\ & & \Downarrow & & \\ \xrightarrow{\text{a.s.}} & \Longrightarrow & \xrightarrow{p} & \Longrightarrow & \xrightarrow{d} \end{array}$$

5.2 Connection to Convergence in Probability

5.2.1 Borel-Cantelli Lemma

Definition 5.2.1 Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of events.

- Let
$$\limsup_{n\to\infty}(A_n) = \bigcap_{n=1}^{\infty}\bigcup_{k\geq n} A_k.$$
If $\omega \in \limsup_{n\to\infty}(A_n)$, we say that A_n occurs infinitely often.
- Let
$$\liminf_{n\to\infty}(A_n) = \bigcup_{n=1}^{\infty}\bigcap_{k\geq n} A_k.$$
If $\omega \in \liminf_{n\to\infty}(A_n)$, we say that A_n occurs eventually.

Exercise 5.2.2

1. Show that
$$\mathbb{P}\left(\liminf_{n\to\infty} A_n\right) \leq \liminf_{n\to\infty} \mathbb{P}(A_n).$$
2. Describe the complement of $\limsup_{n\to\infty}(A_n)$.

Theorem 5.2.3 — Borel-Cantelli Lemma. Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of events.

1. If $\sum_{n=1}^{\infty}\mathbb{P}(A_n) < \infty$ then $\mathbb{P}(A_n \text{ i.o.}) = 0$.
2. If $\sum_{n=1}^{\infty}\mathbb{P}(A_n) = \infty$ and A_n are mutually independent then $\mathbb{P}(A_n \text{ i.o.}) = 1$.

Proof.

1. By continuity of the measure we have

$$\mathbb{P}(A_n \text{ i.o.}) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k\geq n} A_k\right) \leq \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathbb{P}(A_k) = 0.$$

2. Observe that $\{A_n \text{ i.o.}\}^c = \{A_n^c \text{ ev.}\} = \bigcup_{n=1}^{\infty} \bigcap_{k\geq n} A_k^c$. Hence, we have

$$1 - \mathbb{P}(A_n \text{ i.o.}) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k\geq n} A_k^c\right).$$

By independence we have that

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

Note that $\log(1-x) \leq -x$ for $x \in [0,1)$, and thus

$$\begin{aligned}\log\left(\mathbb{P}\left(\bigcap_{k\geq n} A_k^c\right)\right) &= \log\left(\prod_{k\geq n}(1-\mathbb{P}(A_k))\right)\\ &\leq -\sum_{k\geq n}\mathbb{P}(A_k)\\ &= -\infty,\end{aligned}$$

That is, $\mathbb{P}\left(\bigcap_{k\geq n} A_k^c\right) = 0$ for all $n \in \mathbb{N}$. Therefore,

$$1 - \mathbb{P}\left(A_n \text{ i.o.}\right) = \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k\geq n} A_k^c\right) = 0$$

so that $\mathbb{P}(A_n \text{ i.o.}) = 1$.

■

Remark 5.2.4 Theorem 5.2.3 is an example of a zero-one law.

Example 5.2.5 For real numbers in $[0,1]$, we consider the event that its binary expansion contains a finite string of $\{0,1\}$ infinitely many times. Assume the desired string to be $(x_1,\ldots,x_m)$. We consider the sequence of events

$$A_n := \{\omega : \xi_{nm+1}(\omega) = x_1, \ldots, \xi_{nm+m}(\omega) = x_m\}$$

for $n \geq 0$ on $([0,1], \mathscr{B}([0,1]), \mathsf{Leb})$, where $\xi_k(\omega) = \omega_k$. The event A_n occurs when the desired string appears starting from digit $nm+1$. These are mutually independent, given that the ξ_k are independent. Moreover, they are identically distributed as $\mathbb{P}(A_n) = \frac{1}{2^m}$ for every $n \in \mathbb{N}$. Therefore, as $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, it follows from Theorem 5.2.3 that $\mathbb{P}(A_n \text{ i.o.}) = 1$. In other words, the finite string $(x_1,\ldots,x_m) \in \{0,1\}^m$ appears infinitely often in the binary expansion of almost every real number in the interval $[0,1]$.

5.2.2 Applications of the Borel-Cantelli Lemma

Proposition 5.2.6 A sequence $(\xi_n)_{n\in\mathbb{N}}$ converges $\mathbb{P}$-almost surely to ξ if and only if

$$\mathbb{P}\left(\sup_{k\geq n} |\xi_k - \xi| \geq \varepsilon\right) \overset{n\to\infty}{\longrightarrow} 0,$$

for every $\varepsilon > 0$.

Proof. Note that $\xi_n(\omega) \not\to \xi(\omega)$ if and only if there exists an $\epsilon > 0$ such that

$$|\xi_n(\omega) - \xi(\omega)| \geq \epsilon$$

infinitely often. So let $A_n^\epsilon = \{\omega : |\xi_n(\omega) - \xi(\omega)| \geq \epsilon\}$ and $A^\epsilon = \limsup_{n\to\infty} (A_n^\epsilon)$. Then

$$\{\omega : \xi_n(\omega) \not\to \xi(\omega)\} = \bigcup_{\varepsilon \geq 0} A^\varepsilon.$$

As the sets A^ϵ are nested, one can restrict ϵ to the form $\epsilon = \frac{1}{m}$ for some positive integer m, so that

$$\{\omega : \xi_n(\omega) \not\to \xi(\omega)\} = \bigcup_{m=1}^{\infty} A^{\frac{1}{m}}.$$

Hence, $\mathbb{P}(\{\omega : \xi_n(\omega) \not\to \xi(\omega)\}) = 0$ if and only if $\mathbb{P}\left(\bigcup_{m=1}^\infty A^{\frac{1}{m}}\right) = 0$. If this holds it follows for all $m \geq 1$ that

$$\mathbb{P}\left(A^{\frac{1}{m}}\right) \leq \mathbb{P}\left(\bigcup_{m=1}^{\infty} A^{\frac{1}{m}}\right) = 0.$$

Conversely, if for $m \geq 1$ we have $\mathbb{P}\left(A^{\frac{1}{m}}\right) = 0$ then by the continuity of the measure it follows that $\mathbb{P}\left(\bigcup_{m=1}^\infty A^{\frac{1}{m}}\right) = 0$. Now $\mathbb{P}\left(A^{\frac{1}{m}}\right) = 0$ for all $m \geq 1$ happens if and only if $\mathbb{P}(A^\epsilon) = 0$ for all $\epsilon > 0$. Therefore, as

$$\begin{aligned}
\mathbb{P}(A^\varepsilon) &= \mathbb{P}\left(\bigcap_{n\geq 1} \bigcup_{k \geq n} A_k^\varepsilon\right) \\
&= \lim_{n\to\infty} \left(\mathbb{P}\left(\bigcup_{k\geq n} A_k^\varepsilon\right)\right) \\
&= \mathbb{P}\left(\sup_{k\geq n} |\xi_k - \xi| \geq \varepsilon\right),
\end{aligned}$$

we complete the proof. ■

Corollary 5.2.7 Convergence almost surely implies convergence in probability.

$$\begin{array}{ccccc}
 & & \xrightarrow{L^p} & & \\
 & & \Downarrow & & \\
\xrightarrow{\text{a.s.}} & \Longrightarrow & \xrightarrow{p} & \Longrightarrow & \xrightarrow{d}
\end{array}$$

Proof. Suppose that the sequence $(\xi_n)_{n\in\mathbb{N}}$ converges $\mathbb{P}$-almost surely to ξ. Let $\epsilon > 0$. Then for any $\delta > 0$ we can use Proposition 5.2.6 to find an $N \in \mathbb{N}$ such that

$$\mathbb{P}\left(\sup_{k\geq n} |\xi_k - \xi| \geq \epsilon\right) \leq \delta$$

for all $n \geq N$. It follows that

$$\mathbb{P}(|\xi_n - \xi| < \epsilon) \geq 1 - \delta$$

for all $n \geq N$. In particular,

$$\lim_{n\to\infty} \mathbb{P}(|\xi_n - \xi| < \epsilon) = 1$$

which implies that

$$\lim_{n\to\infty} \mathbb{P}(|\xi_n - \xi| \geq \epsilon) = 0.$$

Meaning $\xi_n \to \xi$ in probability. ■

Exercise 5.2.8 Show that the sequence $(\xi_n)_{n\in\mathbb{N}}$ converges almost surely if and only if, for all $\varepsilon > 0$, we have

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{k,l\geq n} |\xi_k - \xi_l| \geq \varepsilon\right) = 0,$$

or equivalently,

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{k\geq 0} |\xi_{n+k} - \xi_n| \geq \varepsilon\right) = 0.$$

Example 5.2.9 Consider the sequence $f_n := \chi_{A_n}$ for

$$A_n = \left[\frac{n}{2^k} - 1, \frac{n+1}{2^k} - 1\right]$$

whenever $k \geq 0$ and $2^k \leq n < 2^{k+1}$. Namely,

- $A_1 = [0, 1]$,
- $A_2 = \left[0, \frac{1}{2}\right]$,
- $A_3 = \left[\frac{1}{2}, 1\right]$,
- $A_4 = \left[0, \frac{1}{4}\right]$,
- $A_5 = \left[\frac{1}{4}, \frac{1}{2}\right]$,
- $\vdots$

Plotting these indicator functions one observes that they move from left to right over $[0, 1]$, half their width and repeat. Therefore, the functions converge in probability to zero. However, given that the indicator function moves from left to right infinitely many times, for all $\omega \in [0, 1]$, we have $f_n(\omega) = 1$ infinitely often and so $f_n(\omega)$ does not converge almost surely.

There is a partial converse to the implication that almost sure converge implies convergence in probability. More specifically, we will see that if a sequence converges in probability, then we can extract a subsequence that converges almost surely.

Lemma 5.2.10 A sufficient condition for a sequence of random variables $(\xi_n)_{n\in\mathbb{N}}$ to converge almost surely to ξ is that

$$\sum_{n=1}^{\infty} \mathbb{P}(|\xi_n - \xi| \geq \varepsilon) < \infty$$

is satisfied for all $\varepsilon > 0$.

Proof. Let $A_n^\varepsilon = \{\omega : |\xi_n(\omega) - \xi(\omega)| \geq \varepsilon\}$. Since $\sum_{n=1}^{\infty} \mathbb{P}(A_n^\epsilon) < \infty$, by Theorem 5.2.3 we know that $\mathbb{P}(A^\varepsilon) := \mathbb{P}(A_n^\epsilon \text{ i.o.}) = 0$ for all $\epsilon > 0$. Following the arguments in Proposition 5.2.6 we know that $\mathbb{P}(\{\omega : \xi_n(\omega) \not\to \xi(\omega)\}) = 0$ as desired. ∎

Corollary 5.2.11 Let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence of positive numbers such that $\varepsilon_n \searrow 0$ as $n \to \infty$. If $(\xi_n)_{n\in\mathbb{N}}$ and ξ are such that

$$\sum_{n=1}^{\infty} \mathbb{P}(|\xi_n - \xi| \geq \varepsilon_n) < \infty,$$

then $\xi_n \xrightarrow{\text{a.s.}} \xi$.

Proof. Fix $\epsilon > 0$. Choose $N \in \mathbb{N}$ such that for all $n \geq N$ we have $\epsilon_n < \epsilon$. Then

$$\sum_{n=1}^{\infty} \mathbb{P}(|\xi_n - \xi| \geq \epsilon) \leq \underbrace{\sum_{n=1}^{N-1} \mathbb{P}(|\xi_n - \xi| \geq \epsilon)}_{\leq N-1 < \infty} + \underbrace{\sum_{n=N}^{\infty} \mathbb{P}(|\xi_n - \xi| \geq \epsilon_n)}_{<\infty},$$

so by Lemma 5.2.10 we have $\xi_n \xrightarrow{\text{a.s.}} \xi$. ■

Theorem 5.2.12 If $\xi_n \xrightarrow{\mathbb{P}} \xi$, then there exists a subsequence such that $\xi_{n_k} \xrightarrow{\text{a.s.}} \xi$.

Proof. Since $\lim_{n\to\infty} \left(\mathbb{P}\left(|\xi_n - \xi| > \frac{1}{k}\right)\right) = 0$ for all $k \in \mathbb{N}$, we can choose a subsequence $(\xi_{n_k})_{k\in\mathbb{N}}$ such that

$$\mathbb{P}\left(|\xi_{n_k} - \xi| > \frac{1}{k}\right) \leq \frac{1}{2^k}$$

for all $k \in \mathbb{N}$. Since $\sum_{k=1}^{\infty} \frac{1}{2^k}$ converges, by Corollary 5.2.11 we have $\xi_{n_k} \xrightarrow{\text{a.s.}} \xi$. ■

Corollary 5.2.13 If $\xi_1 \geq \xi_2 \geq \cdots \geq 0$ are random variables such that $\xi_n \xrightarrow{\mathbb{P}} 0$, then $\xi_n \xrightarrow{\text{a.s.}} 0$.

Proof. Let $\varepsilon > 0$ and let $A_n = \{\omega : \xi_n(\omega) > \varepsilon\}$. Then by continuity,

$$\mathbb{P}\left(\limsup_{n\in\mathbb{N}}(\xi_n) > \varepsilon\right) \leq \mathbb{P}(\xi_n > \varepsilon \text{ i.o.}) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{k\geq n} A_k\right).$$

Since the sequence of events $(A_n)_{n\in\mathbb{N}}$ is non-increasing, the right-hand side equals $\lim_{n\to\infty}(\mathbb{P}(A_n))$ which is zero since $\xi_n \xrightarrow{\mathbb{P}} 0$. Thus,

$$\mathbb{P}\left(\limsup_{n\in\mathbb{N}}(\xi_n) > \varepsilon\right) = 0$$

for all $\epsilon > 0$ and hence

$$\begin{aligned}
\mathbb{P}(\xi_n \not\to \xi) &= \mathbb{P}\left(\bigcup_{m=1}^{\infty} \bigcap_{N=1}^{\infty} \bigcup_{n\geq N} \left\{\xi_n > \frac{1}{m}\right\}\right) \\
&= \mathbb{P}\left(\bigcup_{m=1}^{\infty} \limsup_{n\in\mathbb{N}}(\xi_n) > \frac{1}{m}\right) \\
&\leq \sum_{m=1}^{\infty} \mathbb{P}\left(\limsup_{n\in\mathbb{N}}(\xi_n) > \frac{1}{m}\right) \\
&= 0.
\end{aligned}$$

■

5.3 Connection to L^p convergence

Example 4.1.3 shows that almost sure convergence does not guarantee L^p convergence. However, if we have almost sure convergence and convergence in mean, then we have L^1 convergence.

Theorem 5.3.1 Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of non-negative random variables such that $\xi_n \xrightarrow{\text{a.s.}} \xi$ and $\mathbb{E}(\xi_n) \to \mathbb{E}(\xi) < \infty$. Then $\xi_n \xrightarrow{L^1} \xi$, that is,

$$\mathbb{E}(|\xi_n - \xi|) \to 0$$

as $n \to \infty$.

Proof. For sufficiently large $n \in \mathbb{N}$ we have $\mathbb{E}(|\xi_n|) < \infty$. Hence,

$$\begin{aligned}\mathbb{E}(|\xi_n - \xi|) &= \mathbb{E}\left((\xi - \xi_n)\chi_{\xi \geq \xi_n}\right) + \mathbb{E}\left((\xi_n - \xi)\chi_{\xi_n > \xi}\right) \\ &= 2\mathbb{E}\left((\xi - \xi_n)\chi_{\xi \geq \xi_n}\right) + \mathbb{E}(\xi_n - \xi).\end{aligned}$$

The term $\mathbb{E}(\xi_n - \xi)$ tends to zero by assumption. Note that $0 \leq (\xi - \xi_n)\chi_{\xi \geq \xi_n} \leq \xi$, so by the dominated convergence theorem it follows that $\mathbb{E}(\xi - \xi_n)\chi_{\xi \geq \xi_n} \to 0$. Therefore, $\mathbb{E}(|\xi_n - \xi|) \to 0$ as $n \to \infty$. ■

Example 5.3.2 Let $(\xi_n)_{n \in \mathbb{N}}$ be independent $\{0,1\}$-valued random variables on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$, with $\mathbb{P}(\xi_n = 1) = \frac{1}{n}$. Then

$$\mathbb{E}\left(|\xi_n - 0|^p\right) = \frac{1}{n} \to 0,$$

so $\xi_n \xrightarrow{L^p} 0$. However,

$$\begin{aligned}\{\omega : \xi_n \to 0\} &= \{\xi_n = 0 \text{ eventually}\} \\ &= \bigcup_{n=1}^{\infty} \bigcap_{k \geq n} \{\xi_k = 0\},\end{aligned}$$

where the inner intersections are an increasing sequence of sets. Therefore,

$$\begin{aligned}\mathbb{P}(\xi_n \to 0) &= \lim_{n \to \infty} \mathbb{P}\left(\bigcap_{k \geq n} \{\xi_k = 0\}\right) \\ &= \lim_{n \to \infty} \prod_{k \geq n} \mathbb{P}(\xi_k = 0) \\ &= \lim_{n \to \infty} \prod_{k \geq n} \left(1 - \frac{1}{k}\right) \\ &= 0.\end{aligned}$$

Indeed,

$$\begin{aligned}\prod_{k \geq n} \left(1 - \frac{1}{k}\right) &= \lim_{N \to \infty} \prod_{k=n}^{N} \frac{k-1}{k} \\ &= \lim_{N \to \infty} \frac{n-1}{n} \frac{n}{n+1} \cdots \frac{N-1}{N} \\ &= 0.\end{aligned}$$

Thus, the sequence $(\xi_n)_{n \in \mathbb{N}}$ does not converge almost surely to zero. Hence, we conclude that L^p convergence does not imply almost sure convergence.

5.4 Strong Law of Large Numbers

Definition 5.4.1 Let $(\xi_n)_{n \in \mathbb{N}}$ be a sequence of integrable random variables and let $S_n = \xi_1 + \cdots + \xi_n$. Then the sequence $(\xi_n)_{n \in \mathbb{N}}$ satisfies the strong law of large numbers if

$$\frac{S_n - \mathbb{E}(S_n)}{n} \xrightarrow{\text{a.s.}} 0.$$

Note that the strong law of large numbers properties is stronger than the weak law of large numbers, as now we require convergence almost surely which is stronger than convergence in probability. Recall that for independent and identically distributed random variables $(\xi_n)_{n \in \mathbb{N}}$ with $\mathbb{V}(\xi_1) < \infty$ we showed the L^2 weak law of large numbers. Imposing a stronger moment assumption we arrive at a strong law of large numbers.

Proposition 5.4.2 — Cantelli's Strong Law of Large Numbers. Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables with $\mathbb{E}\left(\xi_1^4\right) < \infty$. Then

$$\frac{S_n - \mathbb{E}(S_n)}{n} \xrightarrow{\text{a.s.}} 0.$$

Proof. Without loss of generality, we centralise the random variables by subtracting their mean, that is $\mathbb{E}(\xi_1) = 0$. Note by Chebyshev's inequality we have

$$\mathbb{P}\left(\left|\frac{S_n}{n}\right| > \epsilon\right) < \frac{\mathbb{E}\left(\left|\frac{S_n}{n}\right|^4\right)}{\epsilon^4}.$$

In the expansion of forth moments, the only non-vanishing terms are terms of the form $\mathbb{E}\left(\xi_j^4\right)$ and $\mathbb{E}\left(\xi_i^2\xi_j^2\right) = \mathbb{E}\left(\xi_i^2\right)\mathbb{E}\left(\xi_j^2\right)$ with $i \neq j$, noticing that the odd moments of ξ_i vanish. We therefore have the expansion

$$\mathbb{E}\left(S_n^4\right) = \mathbb{E}\left(\sum_{k=1}^n \xi_k^4 + \binom{4}{2,2}\sum_{j,k=1, j<k} \xi_j^2\xi_k^2\right)$$

where $\binom{4}{2,2} = \frac{4!}{(2!)^2} = 6$ comes from the multinomial theorem. There are $\frac{n(n-1)}{2}$ unique ways to choose the indices (j,k) such that $j<k$, therefore with the independent and identically distributed assumption we have

$$\mathbb{E}\left(\left|\frac{S_n}{n}\right|^4\right) = \frac{1}{n^4}\left(n\mathbb{E}\left(\xi_1^4\right) + 3n(n-1)\mathbb{E}\left(\xi_1^2\right)^2\right).$$

From Corollary 2.4.6 we have that $\mathbb{E}\left(\xi_1^2\right)^{\frac{1}{2}} \leq \mathbb{E}\left(\xi_1^4\right)^{\frac{1}{4}}$, and so

$$\mathbb{E}\left(\left|\frac{S_n}{n}\right|^4\right) \leq \frac{3n^2-2n}{n^4}\mathbb{E}\left(\xi_1^4\right) \lesssim \frac{1}{n^2}.$$

Therefore $\sum_{n\in\mathbb{N}} \mathbb{P}\left(\left|\frac{S_n}{n}\right| \geq \epsilon\right) \lesssim \sum_{n\in\mathbb{N}} \frac{1}{n^2} < \infty$. Thus we can conclude by applying Lemma 5.2.10. ■

We will now work towards Kolmogorov's strong law of large numbers.

Proposition 5.4.3 — Kolmogorov's Maximal Inequality. Let $(\xi_n)_{n\in\mathbb{N}}$ be independent random variables with finite variances. Then for all $n \geq 1$ and $x > 0$, we have

$$\mathbb{P}\left(\max_{1\leq k\leq n} |S_k - \mathbb{E}(S_k)| \geq x\right) \leq \frac{\mathbb{V}(S_n)}{x^2}.$$

Proof. Without loss of generality, we suppose $\mathbb{E}(\xi_1) = 0$. Consider the event

$$A = \left\{\omega : \max_{1\leq k\leq n} |S_k(\omega)| \geq x\right\}$$

and the events

$$A_k = \{\omega : |S_j| < x \text{ for } j = 1,\dots,k-1 \text{ and } |S_k| \geq x\}$$

for $k = 1,\dots,n$. The sets A_k are mutually disjoint and such that $A = \bigcup_{k=1}^n A_k$. Therefore,

$$\begin{aligned}
\mathbb{E}\left(S_n^2\right) &\geq \mathbb{E}\left(S_n^2\chi_A\right) \\
&= \sum_{k=1}^n \mathbb{E}\left(S_n^2\chi_{A_k}\right) \\
&= \sum_{k=1}^n \mathbb{E}\left((S_k + \xi_{k+1} + \dots \xi_n)^2\chi_{A_k}\right) \\
&= \sum_{k=1}^n \underbrace{\mathbb{E}\left(S_k^2\chi_{A_k}\right)}_{\geq x^2\mathbb{P}(A_k)} + \underbrace{2\mathbb{E}\left((S_k\chi_{A_k})(\xi_{k+1}+\dots+\xi_n)\right)}_{=2\mathbb{E}(S_k\chi_{A_k})\mathbb{E}(\xi_{k+1}+\dots+\xi_n)=0} + \underbrace{\mathbb{E}\left((\xi_{k+1}+\dots+\xi_n)^2\right)}_{\geq 0}
\end{aligned}$$

$$\begin{aligned} &\geq x^2 \sum_{k=1}^{n} \mathbb{P}(A_k) \\ &= x^2 \mathbb{P}(A). \end{aligned}$$

Therefore,

$$\mathbb{V}(S_n) = \mathbb{E}\left(S_n^2\right) \geq x^2 \mathbb{P}(A).$$

■

Remark 5.4.4 Note that Proposition 5.4.3 is stronger than Chebyshev's inequality since Chebyshev's inequality only gives

$$\mathbb{P}\left(|S_k - \mathbb{E}(S_k)| \geq x\right) \leq \frac{\mathbb{V}(S_n)}{x^2}$$

for $k - 1, \ldots, n$. Which when combined with a union bound gives

$$\mathbb{P}\left(\max_{1\leq k\leq n} |S_k - \mathbb{E}(S_k)| \geq x\right) = \mathbb{P}\left(\bigcup_{k=1}^{n} \{|S_k - \mathbb{E}(S_k)| \geq x\}\right) \leq \frac{n\mathbb{V}(S_n)}{x^2}.$$

Hence, Proposition 5.4.3 removes the factor of n in the bound.

Lemma 5.4.5 Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of real-valued independent random variables with $\mathbb{E}(\xi_n) = 0$ for all $n \in \mathbb{N}$. If $\sum_{n=1}^{\infty} \mathbb{V}(\xi_n) < \infty$, then $\sum_{n=1}^{\infty} \xi_n$ converges almost surely.

Proof. Note that

$$\begin{aligned} 0 \leq \sup_{m,n\geq k} |S_n - S_m| &\leq \sup_{m,n\geq k} \left(|S_n - S_k| + |S_k - S_m|\right) \\ &= 2\sup_{n\geq k} |S_n - S_k| =: 2\sigma_k. \end{aligned}$$

By Proposition 5.4.3, we have that

$$\begin{aligned} \mathbb{P}(\sigma_k \geq x) &= \lim_{m\to\infty} \mathbb{P}\left(\max_{m\geq n\geq k} |S_n - S_k| \geq x\right) \\ &\leq \frac{1}{x^2} \lim_{m\to\infty} \sum_{n=k+1}^{m} \mathbb{V}(\xi_n) \\ &\leq \frac{1}{x^2} \sum_{n=k+1}^{\infty} \mathbb{V}(\xi_n) \\ &\xrightarrow{k\to\infty} 0. \end{aligned}$$

Therefore,

$$\lim_{k\to\infty} \mathbb{P}\left(\sup_{n,m\geq k} |S_n - S_m| \geq \epsilon\right) = 0$$

and so we conclude by using the result of Exercise 5.2.8. ■

Theorem 5.4.6 Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent random variables. If $\sum_{n=1}^{\infty} \mathbb{E}(\xi_n)$ and $\sum_{n=1}^{\infty} \mathbb{V}(\xi_n)$ converge, then $\sum_{n=1}^{\infty} \xi_n$ converges almost surely.

Proof. Consider

$$\sum_{n=1}^{\infty} \xi_n = \sum_{i=n}^{\infty} (\xi_n - \mathbb{E}(\xi_n)) + \sum_{n=1}^{\infty} \mathbb{E}(\xi_n).$$

By assumption we know that $\sum_{i=n}^{\infty} \mathbb{E}(\xi_n)$ converges almost surely, and $\sum_{n=1}^{\infty} (\xi_n - \mathbb{E}(\xi_n))$ converges almost surely by Lemma 5.4.5. Therefore, $\sum_{n=1}^{\infty} \xi_n$ converges almost surely. ■

In our aim to prove Kolmogorov's strong law of large numbers, we require some results regarding the convergence of weighted averages.

Lemma 5.4.7 — Toeplitz. Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of non-negative numbers. Let $b_n = \sum_{i=1}^n a_i$ so that $b_1 = a_1 > 0$, and $b_n \nearrow \infty$ as $n \to \infty$. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of numbers converging to x. Then

$$\frac{1}{b_n}\sum_{j=1}^n a_j x_j \to x.$$

In particular, if $a_n = 1$, then

$$\frac{x_1 + \cdots + x_n}{n} \to x.$$

Proof. Fix $\epsilon > 0$. Choose $N_0 := N_0(\epsilon)$ such that for all $n \geq N_0$ we have $|x_j - x| < \frac{\epsilon}{2}$. Choose $N_1 > N_0$ such that $\frac{1}{b_{N_1}}\sum_{j=1}^{N_0}|a_j||x_j - x| < \frac{\epsilon}{2}$, which exists since $|x_j - x|$ is bounded for $j = 1, \ldots, N_0$. Then for any $n > N_1$, we have

$$\begin{aligned}\left|\frac{1}{b_n}\sum_{j=1}^n a_j x_j - x\right| &\leq \left|\frac{1}{b_n}\sum_{j=1}^{N_0} a_j(x_j - x)\right| + \left|\frac{1}{b_n}\sum_{j=N_0+1}^{n} a_j(x_j - x)\right| \\ &\leq \frac{1}{b_{N_1}}\left|\sum_{j=1}^{N_0} a_j(x_j - x)\right| + \left|\frac{1}{b_n}\sum_{j=N_0+1}^{n} a_j(x_j - x)\right| \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2}\underbrace{\left(\frac{1}{b_n}\sum_{j=N_0+1}^{n} a_j\right)}_{\leq 1} \leq \epsilon.\end{aligned}$$

■

Exercise 5.4.8 Suppose $(\xi_k)_{k\in\mathbb{N}}$ is a sequence of independent random variables with common mean m and variance $\mathbb{V}(\xi_k) = k\eta(k)$ with the condition that $\mathbb{V}(\xi_k) \to \infty$, $\eta(k) > 0$ and $\eta(k) \searrow 0$ as $k \to \infty$. Using Lemma 5.4.7 prove that the sequence satisfies the weak law of large numbers. That is, show that $n^{-1}\sum_{i=1}^n \xi_i \to m$ as $n \to \infty$ in L^2 and in probability.

Lemma 5.4.9 Let $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$ be as in Lemma 5.4.7 and let $(x_n)_{n\in\mathbb{N}}$ be a sequence of numbers such that $\sum_{n=1}^\infty x_n$ converges. Then

$$\frac{1}{b_n}\sum_{j=1}^n b_j x_j \to 0$$

as $n \to \infty$. In particular, if $b_n = n$, $x_n = \frac{y_n}{n}$ and $\sum_{j=1}^\infty \frac{y_j}{j}$ converges, then

$$\frac{y_1 + \cdots + y_n}{n} \to 0$$

as $n \to \infty$.

Proof. Let $b_0 = S_0 = 0$ and $S_n = \sum_{j=1}^n x_j$. Then

$$\begin{aligned}\sum_{j=1}^n b_j x_j &= \sum_{j=1}^n b_j(S_j - S_{j-1}) \\ &= b_n S_n - b_0 S_0 - \sum_{j=1}^n S_j(b_j - b_{j-1}) \\ &= b_n S_n - b_0 S_0 - \sum_{j=1}^n a_j S_j.\end{aligned}$$

Dividing by b_n gives

$$\frac{1}{b_n}\sum_{j=1}^{n} b_j x_j = S_n - \underbrace{\frac{b_0 S_0}{b_n}}_{\to 0} - \frac{1}{b_n}\sum_{j=1}^{n} a_j S_j.$$

So when $n \to \infty$, we see that $b_n^{-1}\sum_{j=1}^{n} b_j x_j \xrightarrow{n\to\infty} 0$ by Lemma 5.4.7. ■

Theorem 5.4.10 Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent random variables with $\mathbb{E}\left(\xi_n^2\right) < \infty$ for all $n \in \mathbb{N}$. Let $(b_n)_{n\in\mathbb{N}}$ be a sequence of positive numbers such that $b_n \nearrow \infty$ and

$$\sum_{n=1}^{\infty} \frac{\mathbb{V}(\xi_n)}{b_n^2} < \infty.$$

Then

$$\frac{S_n - \mathbb{E}(S_n)}{b_n} \xrightarrow{\text{a.s.}} 0.$$

When $b_n = n$ we obtain a strong law of large numbers.

Proof. Observe that

$$\frac{S_n - \mathbb{E}(S_n)}{b_n} = \frac{1}{b_n}\sum_{i=1}^{n} b_k \frac{\xi_k - \mathbb{E}(\xi_k)}{b_k}. \tag{5.1}$$

Moreover,

$$\begin{aligned}\mathbb{V}\left(\sum_{k=1}^{n} \frac{\xi_k - \mathbb{E}(\xi_k)}{b_k}\right) &= \sum_{k=1}^{n} \mathbb{V}\left(\frac{\xi_k - \mathbb{E}(\xi_k)}{b_k}\right) \\ &= \sum_{k=1} \frac{\mathbb{V}(\xi_k)}{b_k^2} \\ &< \infty.\end{aligned}$$

Therefore, by Theorem 5.4.6 the sum $\sum_{k=1}^{\infty} \frac{\xi_k - \mathbb{E}(\xi_k)}{b_k}$ converges almost surely. Hence, applying Lemma 5.4.9 to (5.1) completes the proof. ■

Exercise 5.4.11 For ξ an integrable non-negative random variable, show that

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

Theorem 5.4.12 — Kolmogorov's Strong Law of Large Numbers. Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables with $\mathbb{E}(|\xi_1|) < \infty$. Then

$$\frac{S_n}{n} \xrightarrow{\text{a.s}} \mathbb{E}(\xi_1)$$

as $n \to \infty$.

Proof. Without loss of generality assume that $\mathbb{E}(\xi_1) = 0$. By Exercise 5.4.11 it follows that $\sum_{n\in\mathbb{N}} \mathbb{P}(|\xi_n| \geq n) \leq \mathbb{E}(|\xi_1|) < \infty$. By statement 1 of Theorem 5.2.3, we know that $\mathbb{P}(|\xi_n| \geq n \text{ i.o.}) = 0$. That is, $|\xi_n| < n$ eventually $\mathbb{P}$-almost everywhere. So letting $\tilde{\xi}_n = \xi_n \chi_{|\xi_n|<n}$, we have

$$\frac{S_n}{n} \xrightarrow{\text{a.s}} 0$$

if and only if

$$\frac{1}{n}\sum_{i=1}^{n} \tilde{\xi}_i \xrightarrow{\text{a.s}} 0.$$

By dominated convergence theorem it follows that $\mathbb{E}\left(\tilde{\xi}_n\right) \xrightarrow{n\to\infty} \mathbb{E}(\xi_1) = 0$. Therefore, using Lemma 5.4.7 with $x_n = \mathbb{E}\left(\tilde{\xi}_n\right)$ it follows that $\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(\tilde{\xi}_i\right) \to 0$ as $n \to \infty$. Therefore

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} 0$$

if and only if

$$\frac{1}{n}\sum_{i=1}^n \left(\tilde{\xi}_i - \mathbb{E}\left(\tilde{\xi}_i\right)\right) \xrightarrow{\text{a.s.}} 0. \tag{5.2}$$

By Lemma 5.4.9 we know that (5.2) holds if $\sum_{n=1}^\infty \frac{\tilde{\xi}_n - \mathbb{E}(\tilde{\xi}_n)}{n}$ converges. Using Theorem 5.4.6 this is the case if $\sum_{n=1}^\infty \frac{\mathbb{V}(\tilde{\xi}_n - \mathbb{E}(\tilde{\xi}_n))}{n^2}$ converges. Observe that

$$\begin{aligned}
\mathbb{V}\left(\sum_{n=1}^\infty \frac{\tilde{\xi}_n - \mathbb{E}\left(\tilde{\xi}_n\right)}{n}\right) &\leq \sum_{n=1}^\infty \frac{\mathbb{E}\left(\xi_1^2 \chi_{|\xi_1|<n}\right)}{n^2} \\
&= \sum_{n=1}^\infty \frac{1}{n^2}\sum_{k=1}^n \mathbb{E}\left(\xi_n^2 \chi_{\{|\xi_1|\in[k-1,k)\}}\right) \\
&= \sum_{k=1}^\infty \mathbb{E}\left(\xi_n^2 \chi_{|\xi_1|\in[k-1,k)}\right) \underbrace{\left(\sum_{n=k}^\infty \frac{1}{n^2}\right)}_{\leq 2/k} \\
&= 2\sum_{k=1}^\infty \mathbb{E}\left(|\xi| \underbrace{\frac{|\xi|}{k}}_{\leq 1} \chi_{\{|\xi_1|\in[k-1,k)\}}\right) \\
&\leq 2\mathbb{E}(|\xi_1|) \\
&< \infty.
\end{aligned}$$

Hence, (5.2) holds which completes the proof. ■

Example 5.4.13 Consider the probability space $([0,1), \mathscr{B}([0,1)), \mathsf{Leb})$. The binary representation of $\omega \in [0,1)$ is

$$\omega = \frac{\omega_1}{2} + \frac{\omega_2}{2^2} + \cdots = 0.\omega_1\omega_2\ldots$$

where $\omega_j \in \{0,1\}$. Let $\xi_j(\omega) = \omega_j$. Then for $(x_1,\ldots,x_n) \in \{0,1\}^n$ consider

$$\begin{aligned}
A_{(x_1,\ldots,x_n)} &= \{\omega : \xi_1 = x_1, \ldots, \xi_n = x_n\} \\
&= \left\{\omega : \frac{x_1}{2} + \cdots + \frac{x_n}{2^n} \leq \omega < \frac{x_1}{2} + \cdots + \frac{x_n}{2^n} + \frac{1}{2^n}\right\}
\end{aligned}$$

so that $\mathbb{P}\left(A_{(x_1,\ldots,x_n)}\right) = \frac{1}{2^n}$. Therefore, the $(\xi_n)_{n\in\mathbb{N}}$ are independent and identically distributed Bernoulli random variables with $\mathbb{P}(\xi_1 = 1) = \frac{1}{2}$, and so by the strong law of large numbers we conclude that

$$\frac{1}{n}\sum_{k=1}^n \xi_k \xrightarrow{\text{a.s}} \mathbb{E}(\xi_1) = \frac{1}{2}.$$

That is, for almost every number in $[0,1)$ the proportion of zeros and ones in its binary expansion tends to $\frac{1}{2}$. We call such numbers normal.

5.5 Kolmogorov's 0-1 Law

Definition 5.5.1 Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space.

1. Let $(\mathscr{F}_n)_{n\in\mathbb{N}}$ be a sequence of sub-σ-algebras of $\mathscr{F}$. Define the σ-algebra generated by their union as
$$\bigvee_{i=1}^{n} \mathscr{F}_i = \sigma\left(\bigcup_{i=1}^{n} \mathscr{F}_i\right),$$
with the natural extension for when $n = \infty$.

2. Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of random variables defined on $(\Omega, \mathscr{F}, \mathbb{P})$. Then
$$\sigma(\xi_1, \ldots, \xi_n) = \bigvee_{i=1}^{n} \sigma(\xi_i),$$
with the natural extension for when $n = \infty$.

Definition 5.5.2 Under setting of Definition 5.5.1, let $\mathscr{F}_n^p = \sigma(\xi_n, \ldots, \xi_p)$ for $p \geq n$ and $\mathscr{F}_n^\infty = \sigma(\xi_n, \ldots)$. For a sequence of random variables $(\xi_n)_{n\in\mathbb{N}}$, let
$$\mathscr{T} = \bigcap_{n=1}^{\infty} \mathscr{F}_n^\infty$$
be referred to as the tail σ-algebra. Events of $\mathscr{T}$ are called tail events.

Example 5.5.3

- For $B \in \mathscr{B}(\mathbb{R})$, note that
$$\{\xi_n \in B \text{ i.o.}\} = \bigcap_{n=1}^{\infty} \bigcup_{k\geq n} \{\xi_k \in B\} \in \mathscr{T}.$$
- Similarly, $\{\sum_{n=1}^{\infty} \xi_n \text{ converges}\} \in \mathscr{T}$.
- The event $\{\xi_{10} \in B\}$ may not be in $\mathscr{T}$, since its occurrence may be affected by changing a finite number of ξ_n, namely changing ξ_{10}. Similarly, $\{\omega : \xi_n \notin \mathscr{B} \text{ for all } n \in \mathbb{N}\}$ is not in $\mathscr{T}$ since its occurrence may be affected by just a single ξ_n.

Lemma 5.5.4 If $\mathscr{A}, \mathscr{B} \subset \mathscr{F}$ are independent, that is for all $A \in \mathscr{A}$ and $B \in \mathscr{B}$ we have $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, then $\sigma(\mathscr{A})$ and $\sigma(\mathscr{B})$ are independent.

Proof. Consider $A \in \mathscr{A}$ and the measures

- $\mathbb{P}_A^{(1)}(B) = \mathbb{P}(AB)$, and
- $\mathbb{P}_A^{(2)}(B) = \mathbb{P}(A)\mathbb{P}(B)$.

These measures coincide on $\mathscr{B}$, and so by Theorem 1.3.5 they coincide on $\sigma(\mathscr{B})$. That is, $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$ for all $A \in \mathscr{A}$ and $B \in \sigma(\mathscr{B})$. Now let $B \in \sigma(\mathscr{B})$ and consider the measures

- $Q_B^{(1)}(A) = \mathbb{P}(AB)$, and
- $Q_B^{(2)}(A) = \mathbb{P}(A)\mathbb{P}(B)$.

Similarly, we conclude using Theorem 1.3.5 that for all $A \in \sigma(\mathscr{A})$ and $B \in \sigma(\mathscr{B})$ we have $\mathbb{P}(AB) = \mathbb{P}(A)\mathbb{P}(B)$. ■

Lemma 5.5.5 Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent random variables. Then $\mathscr{T}$ is independent of itself.

Proof. Note that $\mathscr{F}_1^n$ is independent of $\mathscr{F}_{n+1}^{n+k}$ for all $k \in \mathbb{N}$ as the random variables ξ_i are independent for all $i \in \mathbb{N}$. Hence, $\mathscr{F}_1^n$ is independent with $\bigcup_{k=1}^{\infty} \mathscr{F}_{n+1}^{n+k}$. So by using Lemma 5.5.4 we deduce that $\mathscr{F}_1^n$ is independent with $\mathscr{F}_{n+1}^{\infty}$. As $\mathscr{T} \subset \mathscr{F}_{n+1}^{\infty}$ it follow that $\mathscr{F}_1^n$ is independent of $\mathscr{T}$. Hence, $\bigcup_{n=2}^{\infty} \mathscr{F}_1^n$ is independent with $\mathscr{T}$, and so by using Lemma 5.5.4 we have that $\mathscr{F}_1^{\infty}$ is independent with $\mathscr{T}$. However, as $\mathscr{T} \subset \mathscr{F}_1^{\infty}$ we get that $\mathscr{T}$ is independent of $\mathscr{T}$. ∎

Theorem 5.5.6 — Kolmogorov's Zero-One Law. Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent random variables and let $A \in \mathscr{T}$. Then $\mathbb{P}(A) \in \{0, 1\}$.

Proof. Using Lemma 5.5.5 we have $\mathbb{P}(A) = \mathbb{P}(A \cap A) = \mathbb{P}(A)^2$. Hence, $\mathbb{P}(A)$ must have a value of zero or one. ∎

Example 5.5.7 Given independent sequence $(\xi_n)_{n\in\mathbb{N}}$ of random variables and $B_1, B_2, \dots \in \mathscr{B}(\mathbb{R})$, we have

$$\{\xi_n \in B_n \text{ i.o.}\} = \bigcap_{n=1}^{\infty} \bigcup_{k\geq n} \{\xi_k \in B_k\} \in \mathscr{T}.$$

Likewise, for the independent events $(A_n)_{n\in\mathbb{N}}$, the sequence of random variables $(\chi_{A_n})_{n\in\mathbb{N}}$ are independent, so that the event $\limsup_{n\to\infty}(A_n) := \{\chi_{A_n} = 1 \text{ i.o.}\}$ is a tail event. Theorem 5.5.6 then says that $\limsup_{n\to\infty}(A_n)$ must have probability zero or one, which coincides with observations made in Theorem 5.2.3.

5.6 Law of Iterated Logarithms

Definition 5.6.1

1. A function $\varphi^*(n)$ is called upper for S_n if $S_n \leq \varphi^*(n)$ for all $n \geq n_0$, for some $n_0 \in \mathbb{N}$, with probability one.
2. A function $\varphi_*(n)$ is called lower for S_n if $S_n > \varphi_*(n)$ for infinitely many n with probability one.

Remark 5.6.2 If a function $\psi(n)$ is such that for all $\epsilon > 0$ the function $(1+\epsilon)\psi(n)$ is upper for S_n and the function $(1-\epsilon)\psi(n)$ is lower for S_n, then the function $\psi(n)$ is an optimal rate of convergence for S_n.

Example 5.6.3 Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent Bernoulli random variables with $\mathbb{P}(\xi_n = 1) = \mathbb{P}(\xi_n = -1) = \frac{1}{2}$ and let $S_n = \xi_1 + \cdots + \xi_n$. We know that

$$\frac{S_n}{n} \xrightarrow{\text{a.s}} 0.$$

Moreover, since $\sum_{n=1}^{\infty} \frac{1}{(n(\log n)^{2\epsilon+1})} < \infty$ for all $\epsilon > 0$, it follows that

$$\frac{S_n}{\sqrt{n(\log(n))^{1+2\epsilon}}} \xrightarrow{\text{a.s.}} 0$$

for all $\epsilon > 0$ by Theorem 5.4.10. However, by Theorem 4.5.6 we know that

$$\frac{S_n}{\sqrt{n}} \overset{d}{\not\to} 0,$$

instead it converges in distribution to a normal random variable.

Consider some $\varphi(n)$.

- Then
$$\left\{\limsup_{n\to\infty}\left(\frac{S_n}{\varphi(n)}\right)\leq 1\right\}=\left\{\lim_{n\to\infty}\sup_{m\geq n}\left(\frac{S_m}{\varphi(m)}\right)\leq 1\right\}$$
which means that for all $\epsilon>0$ there exists an $n_1\in\mathbb{N}$ such that $\sup_{m\geq n}\left(\frac{S_m}{\varphi(m)}\right)\leq 1+\varepsilon$ for all $n\geq n_1$. Equivalently, $S_m\leq(1+\varepsilon)\varphi(m)$ for all $m\geq n_1$. Therefore, if $\mathbb{P}\left(\limsup_{n\to\infty}\left(\frac{S_n}{\varphi(n)}\right)\leq 1\right)=1$, it follows that $(1+\varepsilon)\varphi(n)$ is upper for S_n for all $\varepsilon>0$.
- Similarly,
$$\left\{\limsup_{n\to\infty}\left(\frac{S_n}{\varphi(n)}\right)\geq 1\right\}=\left\{\lim_{n\to\infty}\sup_{m\geq n}\left(\frac{S_m}{\varphi(m)}\right)\geq 1\right\}$$
means that for all $\epsilon>0$ there exists an $n_1\in\mathbb{N}$ such that $\sup_{m\geq n}\left(\frac{S_m}{\varphi(m)}\right)\geq 1-\varepsilon$ for all $n>n_1$. Equivalently, $S_m\geq(1-\varepsilon)\varphi(m)$ for infinitely many m. So if $\mathbb{P}\left(\limsup_{n\to\infty}\left(\frac{S_n}{\varphi(n)}\right)\geq 1\right)=1$ then $(1-\varepsilon)\varphi(n)$ is lower for S_n for all $\varepsilon>0$.

The following propositions will be useful for proving Theorem 5.6.6. The proof of Proposition 5.6.4 is beyond the scope of these notes, for its proof refer to Theorem 9.4 in [2].

Proposition 5.6.4 Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables with $\mathbb{E}(\xi_1)=0$ and $\mathbb{E}\left(\xi_1^2\right)=1$. Let $(a_n)_{n\in\mathbb{N}}\subseteq\mathbb{R}$ be such that $a_n\to\infty$ and $\frac{a_n}{\sqrt{n}}\to 0$. Furthermore, let $(b_n)_{n\in\mathbb{N}}\subseteq\mathbb{R}$ be a sequence such that $b_n\to 0$. Then
$$\mathbb{P}\left(S_n\geq a_n\sqrt{n}\right)=\exp\left(-\frac{a_n^2(1+b_n)}{2}\right).$$

Proposition 5.6.5 Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent random variables with $\mathbb{E}(\xi_i)=0$ and $\mathbb{E}\left(\xi_i^2\right)=1$. Then for $\alpha\geq\sqrt{2}$ we have
$$\mathbb{P}\left(\frac{M_n}{\sqrt{n}}\geq\alpha\right)\leq 2\mathbb{P}\left(\frac{S_n}{\sqrt{n}}\geq\alpha-\sqrt{2}\right)$$
where $M_n:=\max(S_0,S_1,\ldots,S_n)$ for $S_0=0$.

Proof. Let
$$A_j=\left\{M_{j-1}\leq\alpha\sqrt{n}\leq M_j\right\}.$$
Then
$$\mathbb{P}\left(\frac{M_n}{\sqrt{n}}\geq\alpha\right)\leq\mathbb{P}\left(\frac{S_n}{\sqrt{n}}\geq\alpha-\sqrt{2}\right)+\sum_{j=1}^{n-1}\mathbb{P}\left(A_j\cap\left\{\frac{S_n}{\sqrt{n}}\leq\alpha-\sqrt{2}\right\}\right).$$
Note that the variance of S_n-S_j is $n-j$, so using independence and Chebyshev's inequality it follows that
$$\begin{aligned}\mathbb{P}\left(A_j\cap\left\{\frac{|S_n-S_j|}{\sqrt{n}}>\sqrt{2}\right\}\right)&=\mathbb{P}(A_j)\mathbb{P}\left(\frac{|S_n-S_j|}{\sqrt{n}}>\sqrt{2}\right)\\&\leq\mathbb{P}(A_j)\frac{n-j}{2n}\\&\leq\frac{1}{2}\mathbb{P}(A_j).\end{aligned}$$
Since $\bigcup_{j=1}^{n-1}A_j\subseteq\{M_n\geq\alpha\sqrt{n}\}$ we deduce that
$$\mathbb{P}\left(\frac{M_n}{\sqrt{n}}\geq\alpha\right)\leq\mathbb{P}\left(\frac{S_n}{\sqrt{n}}\geq\alpha-\sqrt{2}\right)+\frac{1}{2}\mathbb{P}\left(\frac{M_n}{\sqrt{n}}+\alpha\right),$$

which upon rearrangement gives the desired inequality. ■

Theorem 5.6.6 — Law of Iterated Logarithm. Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent identically distributed random variables with $\mathbb{E}(\xi_1) = 0$ and $\mathbb{E}\left(\xi_1^2\right) = \sigma^2 > 0$. Then

$$\mathbb{P}\left(\limsup_{n\to\infty}\left(\frac{S_n}{\psi(n)}\right) = 1\right) = 1, \tag{5.3}$$

where

$$\psi(n) = \sqrt{2\sigma^2 n \log\left(\log(n)\right)}.$$

That is, for all $\varepsilon > 0$, the function $(1+\varepsilon)\psi$ is upper and the function $(1-\varepsilon)\psi$ is lower for S_n.

Proof. For simplicity, we will assume that $\sigma^2 = 1$. Note that (5.3) is equivalent to

$$\mathbb{P}\left(S_n \geq (1+\epsilon)\psi(n) \text{ i.o.}\right) = 0 \tag{5.4}$$

and

$$\mathbb{P}\left(S_n \geq (1-\epsilon)\psi(n) \text{ i.o.}\right) = 1 \tag{5.5}$$

for all $\epsilon > 0$.

Step 1: Show (5.4).

Given an $\epsilon > 0$, choose $\theta > 1$ such that $\theta^2 < 1+\epsilon$. Let $n_k = \lfloor \theta^k \rfloor$ and $x_k = \theta\sqrt{2\log(\log(n_k))}$. Using Proposition 5.6.4 and Proposition 5.6.5, it follows that

$$\mathbb{P}\left(\frac{M_{n_k}}{\sqrt{n_k}} \geq x_k\right) \leq 2\exp\left(-\frac{1}{2}\left(x_k - \sqrt{2}\right)^2(1+b_k)\right)$$

where $b_k \to 0$. Note that $\frac{1}{2}\left(x_k - \sqrt{2}\right)(1+b_k) = O\left(\theta^2\log(k)\right)$, and so for large k we have

$$\mathbb{P}\left(\frac{M_{n_k}}{\sqrt{n_k}} \geq x_k\right) \leq \frac{2}{k^\theta}.$$

Since $\theta > 1$, it follows by statement 1 of Theorem 5.2.3 that

$$\mathbb{P}\left(M_{n_k} \geq \theta\psi(n_k) \text{ i.o.}\right) = 0.$$

Suppose that $n_{k-1} < n \leq n_k$ and $S_n > (1+\epsilon)\psi(n)$. Then $\psi(n) \geq \psi(n_{k-1}) \sim \frac{1}{\sqrt{\theta}}\psi(n_k)$. Hence, by our choice of θ it follows that

$$(1+\epsilon)\psi(n) > \theta\psi(n_k)$$

for large enough k. Thus for k sufficiently large we have that $S_n > (1+\epsilon)\psi(n)$ implies $M_{n_k} \geq \theta\psi(n_k)$, and so we have show (5.4).

Step 2: Show (5.5).

Given $\epsilon > 0$, choose $\theta \in \mathbb{N}$ such that $\frac{3}{\sqrt{\theta}} < \epsilon$. Let $n_k = \theta^k$ so that $n_k - n_{k-1} \to \infty$. Let $x_k = \left(1+\frac{1}{\theta}\right)\psi(n_k)$, and apply Proposition 5.6.4 with $a_n = \frac{x_k}{\sqrt{n_k - n_{k-1}}}$ to deduce that

$$\mathbb{P}\left(S_{n_k} - S_{n_{k-1}} \geq x_k\right) = \mathbb{P}\left(S_{n_k - n_{k-1}} \geq x_k\right) = \exp\left(-\frac{x_k^2}{2\left(n_k - n_{k-1}\right)}(1+b_k)\right),$$

where $b_k \to 0$. Note that

$$\frac{x_k^2}{2\left(n_k - n_{k-1}\right)}(1+b_k) = O\left(\left(1+\frac{1}{\theta}\right)\log(k)\right)$$

and so for k large enough we have that

$$\mathbb{P}\left(S_{n_k} - S_{n_{k-1}} \geq x_k\right) \geq \frac{1}{k}.$$

As the events are independent we can apply statement 2 of Theorem 5.2.3 to deduce that

$$\mathbb{P}\left(S_{n_k} - S_{n_{k-1}} \geq x_k \text{ i.o.}\right) = 1.$$

On the other hand, one can apply (5.4) to $(-\xi_n)_{n\in\mathbb{N}}$ to deduce that

$$\mathbb{P}\left(-S_{n_{k-1}} \leq 2\psi\left(n_{k-1}\right) \text{ e.v}\right) = 1.$$

As $2\psi\left(n_{k-1}\right) \leq \frac{2}{\sqrt{\theta}}\psi\left(n_k\right)$ it follows that

$$\mathbb{P}\left(S_{n_k} \geq x_k - \frac{2}{\sqrt{\theta}}\psi(n_k) \text{ i.o.}\right) = 1.$$

Noting that $x_k - \frac{2}{\sqrt{\theta}}\psi(n_k) > (1-\epsilon)\psi\left(n_k\right)$, by our choice of θ, (5.5) follows. ■

5.7 Solution to Exercises

Exercise 5.2.2

Solution.

- Recall that

  $$\liminf_{n\to\infty}(A_n) = \bigcup_{n=1}^{\infty}\bigcap_{k\geq n} A_k.$$

  Note that

  $$\bigcap_{k\geq n} A_k \subseteq \bigcap_{k\geq n+1} A_k$$

  so that by the continuity of the measure it follows that

  $$\mathbb{P}\left(\liminf_{n\to\infty}(A_n)\right) = \lim_{n\to\infty}\mathbb{P}\left(\bigcap_{k\geq n} A_k\right).$$

  Moreover, as $\bigcap_{k\geq n} A_k \subseteq A_k$ for all $k \geq n$ it follows that

  $$\mathbb{P}\left(\bigcap_{k\geq n} A_k\right) \leq \inf_{k\geq n}\mathbb{P}(A_k)$$

  for all $m \geq n$. Therefore,

  $$\mathbb{P}\left(\liminf_{n\to\infty} A_n\right) \leq \liminf_{n\to\infty}\mathbb{P}\left(A_n\right).$$

- Using de Morgan's law it follows that

  $$\left(\limsup_{n\to\infty}(A_n)\right)^c = \left(\bigcap_{n=1}^{\infty}\bigcup_{k\geq n} A_k\right)^c = \bigcup_{n=1}^{\infty}\bigcap_{k\geq n} A_k^c = \liminf_{n\to\infty}\left(A_n^c\right).$$

■

Exercise 5.2.8

Solution. For $n \in \mathbb{N}$ and $\epsilon > 0$ consider the set

$$B_n^{\epsilon} = \left\{\omega : \sup_{k,l\geq n}|\xi_k(\omega) - \xi_l(\omega)| \geq \epsilon\right\}.$$

If $\xi_n(\omega) \not\to \xi(\omega)$ then there exists an $\epsilon > 0$ such that for all $n \in \mathbb{N}$ there exists $k, l \geq n$ such that $|\xi_k(\omega) - \xi_l(\omega)| \geq \epsilon$. Otherwise, $(\xi_n(\omega))_{n\in\mathbb{N}}$ would be Cauchy and therefore convergent. Hence,

$$\{\omega : \xi_n(\omega) \not\to \xi(\omega)\} \subseteq \bigcup_{m=1}^{\infty}\bigcap_{n=1}^{\infty} B_n^{\frac{1}{m}}.$$

On the other hand, if $\omega \in \bigcap_{n=1}^{\infty} B_n^{\frac{1}{m}}$ for some $m \in \mathbb{N}$ then for all $n \in \mathbb{N}$ there exists $k, l \geq n$ such that $|\xi_k(\omega) - \xi_l(\omega)| \geq \frac{1}{2m}$. Hence, $\xi_n(\omega) \not\to \xi_(\omega)$. Therefore,

$$\bigcup_{m=1}^{\infty} \bigcap_{n=1}^{\infty} B_n^{\frac{1}{m}} \subseteq \{\omega : \xi_n(\omega) \not\to \xi(\omega)\}.$$

Note that for any $m' \in \mathbb{N}$ we have

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n^{\frac{1}{m'}}\right) \leq \mathbb{P}\left(\bigcup_{m=1}^{\infty} \bigcap_{n=1}^{\infty} B_n^{\frac{1}{m}}\right) \leq \sum_{m=1}^{\infty} \mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n^{\frac{1}{m}}\right).$$

So that $\mathbb{P}\left(\bigcup_{m=1}^{\infty} \bigcap_{n=1}^{\infty} B_n^{\frac{1}{m}}\right) = 0$ if and only if $\mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n^{\frac{1}{m}}\right) = 0$ for all $m \in \mathbb{N}$. As the sets $B_n^{\frac{1}{m}}$ are decreasing in n we know that

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n^{\frac{1}{m}}\right) = \lim_{n\to\infty} \mathbb{P}\left(B_n^{\frac{1}{m}}\right) = \lim_{n\to\infty} \mathbb{P}\left(\sup_{k,l\geq n} |\xi_k - \xi_l| \geq \frac{1}{m}\right)$$

for all $m \in \mathbb{N}$. Therefore, $\xi_n \to \xi$ almost surely if and only if

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{k,l\geq n} |\xi_k - \xi_l| \geq \epsilon\right) = 0$$

for all $\epsilon > 0$. ■

Exercise 5.4.8

Solution. Let $S_n = \frac{1}{n} \sum_{k=1}^{n} \xi_k$. Then

$$\mathbb{V}(S_n) = \frac{1}{n^2} \sum_{k=1}^{n} k\eta(k). \tag{5.6}$$

In the notation of Lemma 5.4.7 let $a_k = k - 1$ and $x_k = \eta(k)$ so that

$$b_n = \sum_{k=1}^{n} a_k = \frac{1}{2} n(n-1) \leq \frac{1}{2}(n+1)^2.$$

Consequently,

$$\frac{2}{n^2} \sum_{k=1}^{n} k\eta(k) \leq \frac{1}{b_n} \sum_{k=1}^{n} k\eta(k) \overset{n\to\infty}{\longrightarrow} 0.$$

Therefore, using (5.6) we have that $\mathbb{V}(S_n) \to 0$ as $n \to \infty$, hence, $S_n \to m$ in L^2. Moreover for $\epsilon > 0$, using Chebyshev's inequality we observe that

$$\mathbb{P}(|S_n - m| \geq \epsilon) \leq \frac{\mathbb{V}(S_n)}{\epsilon^2}.$$

Hence, $S_n \to m$ in probability. ■

Exercise 5.4.11

Solution. Proceeding directly from $\sum_{n=1}^{\infty} \mathbb{P}(\xi \geq n)$ it follows that

$$\begin{aligned}
\sum_{n=1}^{\infty} \mathbb{P}(\xi \geq n) &= \sum_{n=1}^{\infty} \sum_{k\geq n} \mathbb{P}(k \leq \xi < k+1) \\
&= \sum_{k=1}^{\infty} k\mathbb{P}(k \leq \xi < k+1) \\
&= \sum_{k=0}^{\infty} \mathbb{E}\left(k\chi_{[k,k+1)}\right)
\end{aligned}$$

$$
\begin{aligned}
&\le \sum_{k=0}^{\infty} \mathbb{E}\left(\xi \chi_{[k,k+1)}\right) \\
&= \mathbb{E}(\xi) \\
&\le \sum_{k=0}^{\infty} \mathbb{E}\left((k+1)\chi_{[k,k+1)}\right) \\
&= 1 + \sum_{n=1}^{\infty} \mathbb{P}(\xi \ge n).
\end{aligned}
$$

■

6 Convergence in Distribution

In this section, we focus on the weak convergence of measures defined $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

6.1 Weak Convergence

Definition 6.1.1 For $n \in \mathbb{N}$, let $\xi_n : (\Omega_n, \mathscr{F}_n, \mathbb{P}_n) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ be a random variable. Then $\xi_n \to \xi$ weakly as $n \to \infty$ if

$$\lim_{n\to\infty} (\mathbb{E}_{\mathbb{P}_n}(f(\xi_n))) = \mathbb{E}_{\mathbb{P}}(f(\xi))$$

for all $f \in \mathscr{C}_b(\mathbb{R})$.

It is important to note that we do not need to specify the probability space of ξ_n when establishing convergence in distribution, what matters is the distribution of ξ_n. Weak convergence is often taken to be the definition of convergence in distribution, however, convergence in distribution has also been defined differently. Specifically, we say that $\xi_n \to \xi$ in distribution if the distribution function $F_{\xi_n}(x) \to F_\xi(x)$ pointwise for all x where F_ξ is continuous. To show that weak convergence is equivalent to our usual definition of convergence in distribution, we have to show that weak convergence can be formulated on a single probability space.

Theorem 6.1.2 Suppose μ and $(\mu_n)_{n\in\mathbb{N}}$ are probability measures such that the corresponding distribution functions F_n converge pointwise to F at the points of continuity of F. Then there exists random variables ξ and $(\xi_n)_{n\in\mathbb{N}}$ on a probability space $(\Omega', \mathscr{F}', \mathbb{P}')$ such that ξ has distribution μ, ξ_n has distribution μ_n, and $\xi_n \to \xi$ almost surely with respect to $\mathbb{P}'$.

Proof. Let $(\Omega', \mathscr{F}', \mathbb{P}') = ([0,1], \mathscr{B}([0,1]), \mathsf{Leb})$. Then μ and μ_n induce distribution functions F_n and F. Let F^{-1} and F_n^{-1} be their right inverses as defined in equation (3.1). Then by Proposition 3.1.2 the random variables $F^{-1}(U)$ and $F_n^{-1}(U)$, where U is the uniformly distributed random variables on $([0,1], \mathscr{B}([0,1]), \mathsf{Leb})$, have the same distribution as μ and μ_n respectively. It remains to show that $F_n^{-1}(u) \to F^{-1}(u)$ almost surely as $n \to \infty$. Note that if $F^{-1}(\{u\})$ is not finite, then u is a point of discontinuity of F^{-1}. Hence, there are only countably many such points as F^{-1} is a non-decreasing right-continuous function. Thus, it suffices to show that the limit holds for points u whose preimage under F is finite. Let u be such a point and observe the following.

- If $x < F^{-1}(u)$, then using the argument in the proof of Proposition 3.1.2 we have $F(x) < u$. If x is a point of continuity of F, then $F_n(x) \to F(x)$ by assumption, which implies that $F_n(x) < u$ for sufficiently large n. For such n we have that $x \le F_n^{-1}(u)$ which implies that $x \le \liminf_{n\to\infty} \left(F_n^{-1}(u)\right)$. Therefore, as we can choose a sequence $(x_k)_{k\in\mathbb{N}}$ of points of continuity of F such that $x_k \nearrow F^{-1}(u)$, we obtain $F^{-1}(u) \le \liminf_{n\to\infty} \left(F_n^{-1}(u)\right)$.
- If $x > F^{-1}(u)$, then $F(x) \ge u$. In fact we must have $F(x) > u$, for if $F(x) = u$ then $F(y) = u$ for any $y \in [F^{-1}(u), x]$, contradicting the assumption that the preimage of singleton $\{u\}$ is finite. Repeating the above arguments yields $F^{-1}(u) \ge \limsup_{n\to\infty} \left(F_n^{-1}(u)\right)$.

Combining the observations its follows that $F_n^{-1}(u) \to F^{-1}(u)$. ■

Theorem 6.1.3 Let ξ and $(\xi_n)_{n\in\mathbb{N}}$ be integrable random variables. Then the following are equivalent.

1. $\xi_n \to \xi$ weakly.
2. $\limsup_{n\to\infty} (\mathbb{P}(\xi_n \in E)) \le \mathbb{P}(\xi \in E)$ for any closed set $E \subseteq \mathbb{R}$.
3. $\liminf_{n\to\infty} (\mathbb{P}(\xi_n \in O)) \ge \mathbb{P}(\xi \in O)$ for any open set $O \subseteq \mathbb{R}$.
4. $\lim_{n\to\infty} (\mathbb{P}(\xi_n \in C)) = \mathbb{P}(\xi \in C)$ for any C such that $\mathbb{P}(\xi \in \partial C) = 0$.
5. Let $F_{\xi_n}(x)$ be the distribution function of ξ_n and similarly for $F_\xi(x)$. Then $F_{\xi_n}(x) \to F_\xi(x)$ pointwise at any point of continuity of $F_\xi(x)$.
6. For all bounded Lipschitz functions f, it follows that $\mathbb{E}(f(\xi_n)) \to \mathbb{E}(f(\xi))$.[a]

[a]This statement is not examinable for the lecture series, therefore its proof is omitted.

Proof. $(1) \Rightarrow (2)$. Let $E \subseteq \mathbb{R}$ be a closed set and consider the function $f(x) = \chi_E(x)$. Let

$$g(t) = \begin{cases} 1 & t \leq 0 \\ 1-t & 0 \leq t \leq 1 \\ 0 & t \geq 1. \end{cases}$$

Then consider

$$f_\epsilon(x) = g\left(\frac{1}{\epsilon}\rho(x, E)\right)$$

where

$$\rho(x, E) = \inf\{|x-y| : y \in E\}.$$

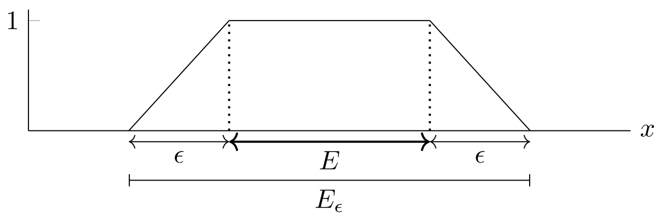
Note that $E_\epsilon := \{x : \rho(x, E) < \epsilon\}$ forms a decreasing sequence of sets as $\epsilon \searrow 0$ such that $E_\epsilon \searrow E$. Observe that,

$$\mathbb{P}_n(\xi_n \in E) = \int f \,\mathrm{d}\mathbb{P}_n \leq \int f_\epsilon \,\mathrm{d}\mathbb{P}_n.$$

Consequently,

$$\begin{aligned} \limsup_{n\to\infty} \left(\mathbb{P}_n(\xi_n \in E)\right) &\leq \limsup_{n\to\infty} \left(\int f_\epsilon \,\mathrm{d}\mathbb{P}_n\right) \\ &\overset{(\star)}{=} \int f_\epsilon \,\mathrm{d}\mathbb{P} \\ &\leq \mathbb{P}(\xi \in E_\epsilon) \\ &\xrightarrow{\epsilon\searrow 0} \mathbb{P}(\xi \in E), \end{aligned}$$

where $(\star)$ follows from the fact that $\xi_n \to \xi$ weakly.



$(2) \Rightarrow (3)$. Let $O \in \mathbb{R}$ open, then $E = \mathbb{R} \setminus O$ is closed. Therefore,

$$\begin{aligned} \liminf_{n\to\infty} \left(\mathbb{P}(\xi_n \in O)\right) &= \liminf_{n\to\infty} \left(1 - \mathbb{P}(\xi_n \in E)\right) \\ &= 1 - \limsup_{n\to\infty} \left(\mathbb{P}\left(\xi_n \in E\right)\right) \\ &\overset{(2)}{\geq} 1 - \mathbb{P}(\xi \in E) \\ &\mathbb{P}(\xi \in O). \end{aligned}$$

$(3) \Rightarrow (4)$. Recall that $\bar{C} = C \cup \partial C$ and $\mathring{C} = C \setminus \partial C$. As $\mathbb{P}(\xi \in \partial C) = 0$ it follows that

- $\limsup_{n\to\infty} \left(\mathbb{P}_n(\xi_n \in C)\right) \leq \limsup_{n\to\infty} \left(\mathbb{P}_n\left(\xi_n \in \bar{C}\right)\right) \leq \mathbb{P}\left(\xi \in \bar{C}\right) = \mathbb{P}(\xi \in C)$, and
- $\liminf_{n\to\infty} \left(\mathbb{P}_n(\xi_n \in C)\right) \geq \liminf_{n\to\infty} \left(\mathbb{P}_n\left(\xi_n \in \mathring{C}\right)\right) \geq \mathbb{P}\left(\xi \in \mathring{C}\right) = \mathbb{P}(\xi \in C)$.

Therefore, $\lim \mathbb{P}_n(\xi_n \in C) = \mathbb{P}(\xi \in C)$.
$(4) \Rightarrow (5)$. Let $x \in \mathbb{R}$ be a point of continuity for F_ξ. Then $\mathbb{P}(\xi = x) = 0$. Therefore,

$$\begin{aligned} F_{\xi_n}(x) &= \mathbb{P}(\xi_n \in (-\infty, x]) \\ &\xrightarrow{(4)} \mathbb{P}(\xi \in (-\infty, x]) \end{aligned}$$

$$= F_\xi(x).$$

$(5) \Rightarrow (1)$. Let f be a bounded continuous function. Let $X, X_1, X_2, \ldots$ be the random variables given by Theorem 6.1.2, defined on the probability space $(\Omega', \mathscr{F}', \mathbb{P}')$. It follows that $f(X_n) \to f(X)$ almost surely with respect to $\mathbb{P}'$. Therefore, using the dominated convergence theorem we conclude that

$$\mathbb{E}_n(f(\xi_n)) = \mathbb{E}(f(X_n)) \to \mathbb{E}(f(X)) = \mathbb{E}(f(\xi)).$$

Hence, $\xi_n \to \xi$ weakly. ■

Remark 6.1.4 Theorem 6.1.3 establishes equivalent notions of weak convergence, also known as convergence in distribution due to statement 5 of Theorem 6.1.3. In any case, we will use $\xi_n \xrightarrow{d} \xi$ to indicate that a sequence of random variables $(\xi_n)_{n\in\mathbb{N}}$ converges weakly to ξ.

Corollary 6.1.5 Suppose that the random variables ξ_n, ξ have densities $f_n(x)$, $f(x)$, respectively. If $f_n(x) \to f(x)$ for any x then $\xi_n \xrightarrow{d} \xi$.

Proof. It suffices to show that

$$F_n(x) = \int_{-\infty}^{x} f_n(y)\,\mathrm{d}y \to F(x) = \int_{-\infty}^{x} f(y)\,\mathrm{d}y$$

for any x. Using

$$|F_n(x) - F(x)| \le \int_{-\infty}^{x} |f_n(y) - f(y)|\,\mathrm{d}y,$$

we just need to check that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} |f_n(y) - f(y)|\,\mathrm{d}y = 0.$$

Note that $a = a_+ - a_-$ and $|a| = a_+ + a_-$ for any real a. Since f_n and f are densities, they integrate to one, so for each $n \in \mathbb{N}$ we have

$$0 = \int_{-\infty}^{\infty} f(y) - f_n(y)\,\mathrm{d}y = \int_{-\infty}^{\infty} (f(y) - f_n(y))_+ - (f(y) - f_n(y))_-\,\mathrm{d}y.$$

Hence,

$$\int_{-\infty}^{\infty} |f(y) - f_n(y)|\,\mathrm{d}y = 2\int_{\infty}^{\infty} (f(y) - f_n(y))_+\,\mathrm{d}y.$$

Therefore, as

$$0 \le (f(y) - f_n(y))_+ \le f(y),$$

we use the dominated convergence theorem to deduce that

$$\int_{-\infty}^{\infty} |f(y) - f_n(y)|\,\mathrm{d}y = 2\int_{\infty}^{\infty} (f(y) - f_n(y))_+\,\mathrm{d}y \overset{n\to\infty}{\longrightarrow} 0.$$

■

Exercise 6.1.6

1. Let F be continuous and suppose $F_n(x) \to F(x)$ pointwise. Show that F_n converges uniformly to F. That is,

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \overset{n\to\infty}{\longrightarrow} 0.$$

2. Give an example of distribution functions $F_n(x)$, $F(x)$ such that $F_n(x) \xrightarrow{d} F(x)$, but

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \not\to 0,$$

as $n \to \infty$.

3. Give an example of probability measures $\mathbb{P}, \mathbb{P}_n$ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that $\mathbb{P}_n \xrightarrow{d} \mathbb{P}$, but convergence $\mathbb{P}_n(B) \to \mathbb{P}(B)$ does not hold for all $B \in \mathscr{B}(\mathbb{R})$.

6.2 Connection to Convergence in Probability

We will show convergence in probability is a strictly stronger notion than convergence in distribution. However, there does exist a partial converse that we will explore.

Theorem 6.2.1 Let ξ and $(\xi_n)_{n\in\mathbb{N}}$ be random variables from $(\Omega, \mathscr{F}, \mathbb{P})$ to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. If $\xi_n \xrightarrow{p} \xi$ then $\xi_n \xrightarrow{d} \xi$.

$$\begin{array}{ccccc} & & \xrightarrow{L^p} & & \\ & & \Downarrow & & \\ \xrightarrow{\text{a.s.}} & \Longrightarrow & \xrightarrow{p} & \Longrightarrow & \xrightarrow{d} \end{array}$$

Proof. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function, such that

$$|f(x)| \le C$$

for all $x \in \mathbb{R}$. Choose M such that $\mathbb{P}(|\xi| > M) \le \frac{\epsilon}{6C}$. Note that for $x \in [-M, M]$ the function f is uniformly continuous. Therefore, there exists a $\delta > 0$ such that for $|x - y| < \delta$, with $x, y \in [-M, M]$, we have

$$|f(x) - f(y)| \le \frac{\epsilon}{3}.$$

Moreover, there exists an $N \in \mathbb{N}$ such that

$$\mathbb{P}(|\xi_n - \xi| > \delta) < \frac{\epsilon}{6C}$$

for $n \ge N$. Hence for $n \ge N$ it follows that,

$$\begin{aligned} \mathbb{E}(|f(\xi_n) - f(\xi)|) \le& \mathbb{E}\left(|f(\xi_n) - f(\xi)|\big||\xi_n - \xi| \le \delta, |\xi| \le M\right) \\ &+ \mathbb{E}\left(|f(\xi_n) - f(\xi)|\big||\xi_n - \xi| \le \delta, |\xi| > M\right) \\ &+ \mathbb{E}\left(|f(\xi_n) - f(\xi)|\big||\xi_n - \xi| > \delta\right)\mathbb{P}(|\xi_n - \xi| > \delta) \\ \le& \frac{\epsilon}{3} + 2C\frac{\epsilon}{6C} + 2C\frac{\epsilon}{6C} \\ =& \epsilon. \end{aligned}$$

■

Proposition 6.2.2 Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of random variables such that $\xi_n \to c$ in distribution, where c is a constant. Then $\xi_n \to c$ in probability.

Proof. Given an $\epsilon > 0$, the set $E_\epsilon = \mathbb{R} \setminus B_\epsilon(c)$ is closed. Therefore, by Theorem 6.1.3 we conclude that

$$\begin{aligned} \lim_{n\to\infty} (\mathbb{P}(|\xi_n - c| \ge \epsilon)) &= \lim_{n\to\infty} (\mathbb{P}(\xi_n \in E_\epsilon)) \\ &= \limsup_{n\to\infty} (\mathbb{P}(\xi_n \in E_\epsilon)) \\ &\le \mathbb{P}(c \in E_\epsilon) \\ &= 0. \end{aligned}$$

Therefore, $\xi_n \to c$ in probability. ■

Example 6.2.3 Note that convergence in distribution does not imply convergence in probability. Consider a real-valued random variable X that is symmetric about zero, such as $\xi \sim \mathsf{N}(0,1)$. Then the sequence $\xi_n := (-1)^{n+1}\xi$ converges in distribution, but not in probability.

Proposition 6.2.4 — Continuous Mapping Theorem. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a measurable function with points of discontinuity U_φ. Let $\xi, (\xi_n)_{n\in\mathbb{N}} : (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ be random variables where $\mathbb{P}(\xi \in U_\varphi) = 0$ and $\xi_n \xrightarrow{d} \xi$. Then $\varphi(\xi_n) \xrightarrow{d} \varphi(\xi)$.

Proof. Let $E \subseteq \mathbb{R}$ be a closed set. Let $x \in \overline{\varphi^{-1}(E)} \setminus U_\varphi$. By construction, there exists a sequence $(x_n)_{n\in\mathbb{N}} \subseteq \varphi^{-1}(E)$ such that $x_n \to x$. As φ is continuous at x it follows that $\varphi(x_n) \to \varphi(x)$. As $(\varphi(x_n))_{n\in\mathbb{N}} \subseteq E$ and E is closed, it follows that $\varphi(x) \in E$ which implies that $x \in \varphi^{-1}(E)$. Therefore,

$$\overline{\varphi^{-1}(E)} \subseteq \varphi^{-1}(E) \cup U_\varphi.$$

Hence,

$$\begin{aligned}
\limsup_{n\to\infty} \left(\mathbb{P}(\varphi(\xi_n) \in E)\right) &= \limsup_{n\to\infty} \left(\mathbb{P}\left(\xi_n \in \varphi^{-1}(E)\right)\right) \\
&\overset{(1)}{\leq} \limsup_{n\to\infty} \left(\mathbb{P}\left(\xi_n \in \overline{\varphi^{-1}(E)}\right)\right) \\
&\overset{\text{Stat 2 Thm 6.1.3}}{\leq} \mathbb{P}\left(\xi \in \overline{\varphi^{-1}(E)}\right) \\
&\leq \mathbb{P}\left(\xi \in \varphi^{-1}(E)\right) + \mathbb{P}(U_\varphi) \\
&= \mathbb{P}\left(\xi \in \varphi^{-1}(E)\right) \\
&= \mathbb{P}(\varphi(\xi) \in E),
\end{aligned}$$

where (1) is justified as $\varphi^{-1}(E) \subseteq \overline{\varphi^{-1}(E)}$. Therefore, by statement 5 of Theorem 6.1.3 we conclude that $\varphi(\xi_n) \to \varphi(\xi)$ in distribution. ∎

Theorem 6.2.5 Let $\xi, (\xi_n)_{n\in\mathbb{N}} : (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ be random variables such that $\xi_n \to \xi$ in distribution. Moreover, let $(\eta_n)_{n\in\mathbb{N}} : (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ be random variables such that $\mathbb{P}(|\xi_n - \eta_n| > \epsilon) \to 0$ as $n \to \infty$ for all $\epsilon > 0$. Then $\eta_n \to \xi$ in distribution.

The proof Theorem 6.2.5 relies on point 6. of Theorem 6.1.3, and so we will omit it here.

Corollary 6.2.6 Consider the random variables $\xi, (\xi_n)_{n\in\mathbb{N}}$ and $(\eta_n)_{n\in\mathbb{N}}$ from $(\Omega, \mathscr{F}, \mathbb{P})$ to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Assume $\xi_n \overset{n\to\infty}{\longrightarrow} \xi$ in distribution and $\eta_n \overset{n\to\infty}{\longrightarrow} c$ in probability, with c being a constant. For the random variables $T_n : \omega \in \Omega \mapsto (\xi_n(\omega), \eta_n(\omega))$ and $T : \omega \in \Omega \mapsto (\xi_n(\omega), c)$ we have that $T_n \overset{n\to\infty}{\longrightarrow} T$ in distribution.

One can refer to [3] for proof of these results.

Theorem 6.2.7 — Slutsky's Theorem. Consider the random variables ξ, $(\xi_n)_{n\in\mathbb{N}}$ and $(\eta_n)_{n\in\mathbb{N}}$ from $(\Omega, \mathscr{F}, \mathbb{P})$ to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. If $\xi_n \xrightarrow{d} \xi$ and $\eta_n \xrightarrow{p} c$, with c being a constant, then the following hold.

1. $\xi_n + \eta_n \xrightarrow{d} \xi + c$.
2. $\xi_n \eta_n \xrightarrow{d} c\xi$.
3. If $c \neq 0$, then $\frac{\xi_n}{\eta_n} \xrightarrow{d} \frac{\xi}{c}$.

Proof. Using Corollary 6.2.6 we can apply Proposition 6.2.4 to $\varphi(x,y) = x+y$, $\varphi(x,y) = xy$ and $\varphi(x,y) = \frac{x}{y}$ respectively. ∎

Example 6.2.8 Consider a real-valued random variable ξ, that is symmetric about zero. Let $\xi_n = \xi$ and $\eta_n = (-1)^{n+1}\xi$. Then $\xi_n + \eta_n$ takes the form $(2\xi, 0, 2\xi, 0, \dots)$ which does not converge in distribution. Hence, it is not true in general that $\xi_n \xrightarrow{d} \xi$ and $\eta_n \xrightarrow{d} \eta$ implies $\xi_n + \eta_n \xrightarrow{d} \xi + \eta$.

6.3 Relative Compactness and Tightness

Soon we will prove the central limit theorem by looking at characteristic functions. To establish the relationship between characteristic functions and measures, we need some tools relating to the collections of measures and clarify the meaning of relatively compact and tight collections. These concepts are also useful for working with stochastic processes and ergodic theory.

Definition 6.3.1 A family of probability measures $\mathscr{P} = (\mathbb{P}_\alpha)_{\alpha \in A}$ with the corresponding family of distribution functions $(F_\alpha)_{\alpha \in A}$, is relatively compact if every sequence of measures from $\mathscr{P}$ contains a subsequence that weakly converges to a probability measure.

Remark 6.3.2 We emphasise that in this definition the limit measure needs to be a probability measure, although it need not belong to the original class $\mathscr{P}$. In fact, $\mathscr{P}$ is relatively compact if its closure with respect to the Levi-Prokhorov metric is compact.

Example 6.3.3 The collection consisting of weakly convergent sequences of measures is relatively compact.

Lemma 6.3.4 Let $\mathbb{P}$ be a probability measure and $(\mathbb{P}_n)_{n\in\mathbb{N}}$ a family of probability measures. Then $\mathbb{P}_n \xrightarrow{d} \mathbb{P}$ if and only if every subsequence $\left(\mathbb{P}_{n'_k}\right)_{k\in\mathbb{N}}$ of $(\mathbb{P}_n)_{n\in\mathbb{N}}$ contains a subsequence $\left(\mathbb{P}_{n''_k}\right)_{k\in\mathbb{N}}$ such that $\mathbb{P}_{n''_k} \xrightarrow{d} \mathbb{P}$.

Proof. $(\Rightarrow)$. For $\mathbb{P}_n \xrightarrow{d} \mathbb{P}$ it is the case that

$$\int f(x)\,\mathrm{d}\mathbb{P}_n(x) \to \int f(x)\,\mathrm{d}\mathbb{P}(x)$$

for all bounded and continuous functions f. Clearly, if this holds then any subsequence $\left(\mathbb{P}_{n'_k}\right)_{k\in\mathbb{N}}$ converges in distribution to $\mathbb{P}$. Hence, one can take the subsequence $n''_k = n'_k$.

$(\Leftarrow)$. Suppose for contradiction that $\mathbb{P}_n \overset{d}{\not\to} \mathbb{P}$. Then for some $\epsilon > 0$ there exists a bounded and continuous function f such that

$$\left|\int f(x)\,\mathrm{d}\mathbb{P}_k(x) - \int f(x)\,\mathrm{d}\mathbb{P}\right| \geq \epsilon$$

for infinitely many $k \in \mathbb{N}$. Hence, we can extract a subsequence $\left(\mathbb{P}_{n'_k}\right)_{k\in\mathbb{N}} \subseteq (\mathbb{P}_n)_{n\in\mathbb{N}}$ such that

$$\left|\int f(x)\,\mathrm{d}\mathbb{P}_{n'_k}(x) - \int f(x)\,\mathrm{d}\mathbb{P}\right| \geq \epsilon$$

for all $k \in \mathbb{N}$. However, this has no subsequence converging to $\mathbb{P}$. Hence, we get a contradiction. ■

Example 6.3.5 A given family of probability measures $\mathscr{P}$ is not necessarily relatively compact. Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of real-valued random variables and suppose $F_{\xi_n}(x) \to F(x)$ for all $x \in \mathbb{R}\setminus U_F$. Then $F(x)$ is not necessarily a distribution function. Hence, the family $(\xi_n)_{n\in\mathbb{N}}$ is not relatively compact.

1. The variables could run away to infinity. Let ξ_n be $U[n, n+1]$, then $F(x) \equiv 0$.

2. The variables could spread across infinity. Let ξ_n be $U[-n, n]$, then $F(x) \equiv \frac{1}{2}$.

Let us denote the collection of non-decreasing, right-continuous functions with an image in $[0, 1]$ as $\mathscr{G} = \{F : \mathbb{R} \to [0, 1]\}$, and refer to it as the collection of generalised distribution functions. Distribution functions form a subset of $\mathscr{G}$ for which $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

Theorem 6.3.6 — Helly's Selection Theorem. The collection $\mathscr{G}$ of generalised distribution functions is sequentially compact. That is, for any sequence $(F_n)_{n\in\mathbb{N}} \subseteq \mathscr{G}$, there exists a function $F \in \mathscr{G}$ and a subsequence $(F_{n_k})_{k\in\mathbb{N}} \subseteq (F_n)_{n\in\mathbb{N}}$ such that $F_{n_k}(x) \to F(x)$ for every point $x \in \mathbb{R} \setminus U_F$, where U_F is the set of discontinuities of F.

Proof. Let $(q_k)_{k\in\mathbb{N}}$ be an enumeration of $\mathbb{Q}$, that is a bijection from $\mathbb{N}$ to $\mathbb{Q}$.

- The sequence $(F_n(q_1))_{n\in\mathbb{N}}$ is bounded, so by the Bolzano-Weierstrass theorem it has a subsequence $\left(F_{n_k^{(1)}}(q_1)\right)_{k\in\mathbb{N}}$ which converges to some $G(q_1) \in [0, 1]$.
- Consider the sequence $\left(F_{n_k^{(1)}}(q_2)\right)_{k\in\mathbb{N}}$. This is also bounded and so by the Bolzano-Weierstrass theorem it has a subsequence $\left(F_{n_k^{(2)}}(q_2)\right)_{k\in\mathbb{N}}$ which converges to some $G(q_2) \in [0, 1]$.
- We repeat to extract further subsequences.

We can illustrate the above procedure with the following diagram.

$$
\begin{array}{cccccccccccccc}
F_1 & F_2 & F_3 & F_4 & F_5 & F_6 & F_7 & F_8 & F_9 & F_{10} & F_{11} & F_{12} & F_{13} & \dots \\
 & \downarrow & & \downarrow & & \downarrow & \downarrow & & \downarrow & \downarrow & & \downarrow & & \\
 & \boxed{F_{n_1^{(1)}}} & & F_{n_2^{(1)}} & & F_{n_3^{(1)}} & F_{n_4^{(1)}} & & F_{n_5^{(1)}} & F_{n_6^{(1)}} & & F_{n_7^{(1)}} & & \dots \\
 & & & \downarrow & & & \downarrow & & \downarrow & \downarrow & & \downarrow & & \\
 & & & F_{n_1^{(2)}} & & & \boxed{F_{n_2^{(2)}}} & & F_{n_3^{(2)}} & F_{n_4^{(2)}} & & F_{n_5^{(2)}} & & \dots \\
 & & & \downarrow & & & & & \downarrow & & & \downarrow & & \\
 & & & F_{n_1^{(3)}} & & & & & F_{n_2^{(3)}} & & & \boxed{F_{n_3^{(3)}}} & & \dots \\
 & & & & & & & & \downarrow & & & \downarrow & & \\
 & & & & & & & & \vdots & & & \vdots & &
\end{array}
$$

We now extract the diagonal subsequence, that is we let $n_k = n_k^{(k)}$ for all $k \in \mathbb{N}$. Note that for any $m \in \mathbb{N}$ we have $(n_k)_{k\in\mathbb{N}} \subseteq \left(n_k^{(m)}\right)_{k\in\mathbb{N}}$. Hence, $F_{n_k}(q_m) \to G(q_m)$ as $\left(F_{n_k^{(m)}}\right)_{k\in\mathbb{N}}$ is a converging sequence, and any subsequence of a converging sequence converges to the same limit as the main sequence. Therefore, $F_{n_k}(q) \to G(q)$ for all $q \in \mathbb{Q}$. Moreover, for any $p, q \in \mathbb{Q}$ we have $G(p) \le G(q)$ by the monotonicity of limits. Let

$$F(x) = \inf\{F(q) : q \in \mathbb{Q}, q > x\} = \lim_{q \searrow x} G(q).$$

It is clear that $F \in \mathscr{G}$. Moreover, for all $q \in \mathbb{Q}$ we have $F(q) \ge G(q)$ and for $x < q$ we have $F(x) \le G(q)$. It remains to show that $F_{n_k}(x) \to F(x)$ for all $x \in \mathbb{R} \setminus U_F$, so let us fix $x \in \mathbb{R} \setminus U_F$ and some arbitrary $\epsilon > 0$. One can choose $y < r < x < q$ with $r, q \in \mathbb{Q}$, such that

$$F(x) - \epsilon < F(y) \le F(r) = G(r) \le F(x) \le F(q) = G(q) < F(x) + \epsilon,$$

and so for sufficiently large k, it follows that $F_{n_k}(r), F_{n_k}(q) \in (F(x) - \epsilon, F(x) + \varepsilon)$ which implies that $F_{n_k}(x) \in (F(x) - \epsilon, F(x) + \varepsilon)$. Since $\varepsilon > 0$ is arbitrary, the proof is complete. ∎

It turns out that the tightness provides a necessary and sufficient condition for a family of probability measures (or finite measures) to be relatively compact.

Definition 6.3.7 A family of probability measures $\mathscr{P} = (\mathbb{P}_\alpha)_{\alpha\in A}$ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ is tight if for all $\varepsilon > 0$

there exists a compact set $K \subseteq \mathbb{R}$ such that

$$\sup_{\alpha \in A} (\mathbb{P}_\alpha(\mathbb{R} \backslash K)) \leq \varepsilon.$$

Theorem 6.3.8 — Prokhorov's theorem. A family of probability measures $\mathscr{P}$ is tight if and only if it is relatively compact.

Proof. ($\Leftarrow$). Let us suppose that $\mathscr{P} := (\mathbb{P}_\alpha)_{\alpha \in A}$ is relatively compact but not tight. Then there exists a $\epsilon > 0$ such that for any compact $K \subseteq \mathbb{R}$ we have $\sup_{\alpha \in A} (\mathbb{P}_\alpha(\mathbb{R} \setminus K)) > \epsilon$. Hence, for any $n \in \mathbb{N}$ there is a $\mathbb{P}_{\alpha_n}$ such that

$$\mathbb{P}_{\alpha_n}(\mathbb{R} \setminus (-n, n)) > \epsilon. \tag{6.1}$$

By relative compactness, there is a subsequence $(\mathbb{P}_{\alpha_{n_k}})_{k \in \mathbb{N}}$ such that $\mathbb{P}_{\alpha_{n_k}} \xrightarrow{k \to \infty} Q$ weakly, where Q is a probability measure. Therefore, by Theorem 6.1.3, it follows that

$$\limsup_{k \to \infty} \left(\mathbb{P}_{\alpha_{n_k}}(\mathbb{R} \setminus (-n, n)) \right) \leq Q(\mathbb{R} \setminus (-n, n)).$$

But the right-hand side tends to zero as $n \to \infty$, which contradicts (6.1). Hence $\mathscr{P}$ must be tight.
($\Rightarrow$). Now let $\mathscr{P}$ be tight, and $(\mathbb{P}_n)_{n \in \mathbb{N}}$ be a sequence of elements in $\mathscr{P}$ with corresponding distribution functions $(F_n)_{n \in \mathbb{N}}$. By Theorem 6.3.6, there exists a subsequence $(F_{n_k})_{k \in \mathbb{N}}$ such that $F_{n_k}(x) \to F(x)$ for $x \in \mathbb{R} \setminus U_F$ where F is some generalised distribution function. We now check that $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$. Fix $\epsilon > 0$, then from tightness there is an $I = (a, b]$ such that,

$$\sup_{n \in \mathbb{N}} (\mathbb{P}_n(\mathbb{R} \setminus I)) < \epsilon.$$

Consequently, $1 - \inf_{n \in \mathbb{N}} (\mathbb{P}_n(I)) > \epsilon$ which implies that $\mathbb{P}_n(I) > 1 - \epsilon$ for all $n \in \mathbb{N}$. Now let $a' < a < b < b'$ where $a', b' \in \mathbb{R} \setminus U_F$, then

$$\begin{aligned}
1 - \epsilon &< \mathbb{P}_{n_k}((a, b]) \\
&< \mathbb{P}_{n_k}((a', b']) \\
&= F_{n_k}(b') - F_{n_k}(a') \\
&\xrightarrow{k \to \infty} F(b') - F(a').
\end{aligned}$$

This implies that $F(+\infty) - F(-\infty) \geq 1$. Since, $F : \mathbb{R} \to [0, 1]$ we must have $F(+\infty) = 1$ and $F(-\infty) = 0$. Therefore, F is a distribution function which implies that $\mathscr{P}$ is relatively compact. ■

Remark 6.3.9

1. From Theorem 6.3.8, one sees that for a family of random variables $(\xi_n)_{n \in \mathbb{N}}$, if

$$F_{\xi_n}(x) \to F(x)$$

for all points of continuity of F, where F is a distribution function, then $(\xi_n)_{n \in \mathbb{N}}$ is tight.

2. Theorem 6.3.8 remains true for measures on $\mathbb{R}^n$, $\mathbb{R}^\infty$ and more generally on any complete separable metric space with a Borel σ-algebra of sets.

Exercise 6.3.10 Let $(\xi_n)_{n \in \mathbb{N}}$ be a family of random variables such that $\mathbb{E}(|\xi_n|) \leq M$ for all $n \in \mathbb{N}$. Show that $(\xi_n)_{n \in \mathbb{N}}$ is tight.

6.4 Solution to Exercises

Exercise 6.1.6

Solution.

1. Fix $\epsilon > 0$. As $\lim_{x\to\infty} F(x) = 1$, $\lim_{x\to-\infty} F(x) = 0$ and $F(x)$ is non-decreasing, there exists an $M \in \mathbb{R}$ such that for $x > M$ we have $F(x) \in \left(1 - \frac{\epsilon}{2}, 1\right]$ and for $x < -M$ we have $F(x) \in [0, \frac{\epsilon}{2})$. Let $x_\epsilon > M$, then there exists an $N_1 \in \mathbb{N}$ such that $|F_n(x_\epsilon) - F(x_\epsilon)| < \frac{\epsilon}{2}$ for $n \geq N_1$. Therefore, for all $x \geq x_\epsilon$ and $n \geq N_1$ it follows that $F_n(x) \in (1-\epsilon, 1]$, so that $|F_n(x) - F(x)| < \epsilon$. Similarly, there exists an $N_2 \in \mathbb{N}$ such that for $x \leq -x_\epsilon$ and $n \geq N_2$ we have $|F_n(x) - F(x)| < \epsilon$. As F is continuous it is uniformly continuous on $[-x_\epsilon, x_\epsilon]$ so that we can choose

$$-x_\epsilon = x_0 < x_1 < \cdots < x_k = x_\epsilon$$

such that $|F(x_{i+1}) - F(x_i)| < \frac{\epsilon}{5}$ for each $i \in \{0, \ldots, k\}$. For each $i \in \{0, \ldots, k\}$ let $\tilde{N}_i \in \mathbb{N}$ be such that

$$|F_n(x_i) - F(x_i)| \leq \frac{\epsilon}{5}$$

for all $n \geq \tilde{N}_i$. For $n \geq \tilde{N} := \max_{i\in\{0,\ldots,k\}} \left(\tilde{N}_i\right)$ it follows that

$$\begin{aligned}
|F_n(x_{i+1}) - F_n(x_i)| &\leq |F_n(x_{i+1}) - F(x_{i+1})| + |F(x_{i+1}) - F(x_i)| + |F(x_i) - F_n(x_i)| \\
&\leq \frac{\epsilon}{5} + \frac{\epsilon}{5} + \frac{\epsilon}{5} \\
&\leq \frac{3\epsilon}{5}.
\end{aligned}$$

Note that for any $x \in [-x_\epsilon, x_\epsilon)$ we have $x_i \leq x < x_{i+1}$ for some $i \in \{0, \ldots, k-1\}$. By the non-decreasing property of F we know that $F(x_i) \leq F(x) \leq F(x_{i+1})$ and similarly for each F_n. Therefore,

$$\begin{aligned}
|F_n(x) - F(x)| &\leq |F_n(x) - F_n(x_i)| + |F_n(x_i) - F(x_i)| + |F(x_i) - F(x)| \\
&\leq |F_n(x_{i+1}) - F_n(x_i)| + |F_n(x_i) - F(x_i)| + |F(x_i) - F(x_{i+1})| \\
&\leq \frac{3\epsilon}{5} + \frac{\epsilon}{5} + \frac{\epsilon}{5} \\
&< \epsilon.
\end{aligned}$$

Therefore, for $n \geq N := \max\left(N_1, N_2, \tilde{N}\right)$ we have that

$$|F_n(x) - F(x)| < \epsilon$$

for all $x \in \mathbb{R}$. Hence, $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to 0$ which is equivalent to uniform convergence.

2. Consider the random variable $\xi : \Omega \to \mathbb{R}$ where $\mathbb{P}(\xi = 1) = 1$. The distribution function of ξ is given by

$$F(x) = \begin{cases} 0 & x < 1 \\ 1 & x \geq 1. \end{cases}$$

For $n \in \mathbb{N}$, let $\xi_n : \Omega \to \mathbb{R}$ be the random variable where $\mathbb{P}\left(\xi_n = 1 - \frac{1}{n}\right) = 1$. Similarly, the distribution function of ξ_n is given by

$$F_n(x) = \begin{cases} 0 & x < 1 - \frac{1}{n} \\ 1 & x \geq 1 - \frac{1}{n}. \end{cases}$$

Then for $x > 1$ it is clear that $F_n(x) = F(x) = 1$ for all $n \in \mathbb{N}$ which implies that $\lim_{n\to\infty} F_n(x) = F(x)$. For $x < 1$, there exists a $N \in \mathbb{N}$ such that $\frac{1}{N} < 1 - x$. Therefore, for $n \geq N$ we have that $F_n(x) = F(x) = 0$ and so $\lim_{n\to\infty} F_n(x) = F(x)$. As F is only continuous for $\mathbb{R} \setminus \{0\}$ we conclude that $F_n \xrightarrow{d} F$. However, $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 1$ for all $n \in \mathbb{N}$.

3. Consider the setting of statement 2 and let $B = (0, 1) \in \mathscr{B}(\mathbb{R})$. Then $\mathbb{P}(\xi_n \in B) = 1$ for all $n \in \mathbb{N}$, however, $\mathbb{P}(\xi \in (0, 1)) = 0$.

■

Exercise 6.3.10

Solution. For $\epsilon > 0$ let $K = \left[-\frac{2M}{\epsilon}, \frac{2M}{\epsilon}\right]$. Then using Markov's inequality it follows that

$$\begin{aligned}
\mathbb{P}\left(\xi_n \in \mathbb{R} \setminus K\right) &= \mathbb{P}\left(|\xi_n| > \frac{2M}{\epsilon}\right) \\
&\leq \frac{\mathbb{E}(|\xi_n|)}{\frac{2M}{\epsilon}} \\
&\leq \frac{\epsilon}{2M} M \\
&< \epsilon.
\end{aligned}$$

Therefore, the family $(\xi_n)_{n\in\mathbb{N}}$ is tight. ■

7 Convergence of Characteristic Functions

In this chapter, we look at the characteristic function of measures of $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$.

7.1 Characteristic Function

Definition 7.1.1 The characteristic function of a random variable $\xi : \Omega \to \mathbb{R}$ is

$$\varphi(t) := \varphi_\xi(t) \equiv \mathbb{E}\left(e^{it\xi}\right) := \int_\Omega e^{it\xi(\omega)}\,\mathbb{P}(\mathrm{d}\omega)$$

for $t \in \mathbb{R}$.

Remark 7.1.2 We may generalise Definition 7.1.1 to random variables on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. The characteristic function of a random vector $\xi := (\xi_1, \dots, \xi_n)$ is

$$\varphi_\xi(t_1, \dots, t_n) := \mathbb{E}\left(\exp\left(i\sum_{k=1}^n t_k\xi_k\right)\right).$$

The characteristic function of a random variable only depends on its distribution. If $F(x)$ has density $f(x)$, with respect to the Lebesgue measure, then

$$\varphi(t) = \int_{\mathbb{R}^n} e^{it^\top x} f(x)\,\mathrm{d}x.$$

Proposition 7.1.3

1. If ξ is a random variable and $\eta = a\xi + b$ for constants a, b, then $\varphi_\eta(t) = e^{itb}\varphi_\xi(at)$.
2. For a characteristic function φ we have $|\varphi(t)| \le \varphi(0) = 1$.
3. Let ξ be a random variable. Then $\varphi_\xi(t)$ is uniformly continuous on $\mathbb{R}$.
4. If $\xi_1, \dots, \xi_n$ are independent random variables and $S = \xi_1 + \dots + \xi_n$, then
$$\varphi_S(t) = \prod_{j=1}^n \varphi_{\xi_j}(t).$$

Proof.

1. From the linearity of the expectation,
$$\begin{aligned}\varphi_\eta(t) &= \mathbb{E}\left(e^{it\eta}\right)\\ &= \mathbb{E}\left(e^{it(a\xi+b)}\right)\\ &= \mathbb{E}\left(e^{iat\xi}e^{itb}\right)\\ &= e^{itb}\mathbb{E}\left(e^{iat\xi}\right)\\ &= \varphi_\xi(at).\end{aligned}$$
2. Observe that
$$|\varphi(t)| = \left|\int_\Omega e^{it\xi(\omega)}\,\mathbb{P}(\mathrm{d}\omega)\right| \le \int_\Omega \left|e^{it\xi(\omega)}\right|\,\mathbb{P}(\mathrm{d}\omega) \le \int_\Omega \mathbb{P}(\mathrm{d}\omega) = 1.$$
3. Note that
$$|\varphi(t+h) - \varphi(t)| = \left|\mathbb{E}\left(e^{it\xi}\left(e^{ih\xi} - 1\right)\right)\right| \le \mathbb{E}\left(\left|e^{ih\xi} - 1\right|\right).$$
By dominated convergence theorem we know that $\mathbb{E}\left(\left|e^{ih\xi} - 1\right|\right) \to 0$ as $h \to 0$. Therefore, φ is uniformly continuous.

4. As the random variables are independent, we can use Proposition 3.2.11 to deduce that

$$\begin{aligned}\varphi_S(t) &= \mathbb{E}\left(e^{it\left(\sum_{j=1}^n \xi_j\right)}\right)\\ &= \mathbb{E}\left(\prod_{j=1}^n e^{it\xi_j}\right)\\ &= \prod_{j=1}^n \mathbb{E}\left(e^{it\xi_j}\right)\\ &= \prod_{j=1}^n \varphi_{\xi_j}(t).\end{aligned}$$

■

Example 7.1.4 Let $\varphi(t)$ be a characteristic function.

- From statement 2 of Proposition 7.1.3 we note that $\psi(t) := \mathrm{Im}(\varphi(t))$ cannot be a characteristic function as $\psi(0) = 0 \neq 1$.
- Using statement 1 of Proposition 7.1.3 we note that $\psi(t) := \varphi(-t)$ is a characteristic function. Therefore, as $|\varphi(t)|^2 = \varphi(t)\overline{\varphi(t)} = \varphi(t)\psi(t)$, it follows from statement 2 of Proposition 7.1.3 that $|\varphi(t)|^2$ is a characteristic function.

The moment-generating function also shares properties 1 and 4 of Proposition 7.1.3, but the lack of properties 2 and 3 means it is preferable to use characteristic functions to establish weak convergence.

Exercise 7.1.5

1. Let $\xi \sim \mathsf{B}(n,p)$. Then
$$\varphi_\xi(t) = \left(pe^{it} + (1-p)\right)^n.$$
2. Let $\xi \sim \mathsf{N}(m,\sigma^2)$. Then
$$\varphi_\xi(t) = \exp\left(itm - \frac{t^2\sigma^2}{2}\right).$$
    - Let X and Y be independent identically distributed random variables with zero mean and unit variance. Prove using characteristic functions that if the distribution F of $\frac{X+Y}{\sqrt{2}}$ is the same as that of X and Y, then F is the standard normal distribution.
3. Let $\xi \sim \mathsf{Po}(\lambda)$. Then
$$\varphi_\xi(t) = e^{-\lambda+\lambda e^{it}}.$$

7.2 Obtaining Moments

The existence of moments for a real-valued random variable is determined by the smoothness of its characteristic function at zero.

Proposition 7.2.1 Let ξ be a random variable with a characteristic function φ and distribution function F.

1. If $\mathbb{E}\left(|\xi|^n\right) < \infty$ for some $n \geq 1$ then $\varphi^{(\tau)}(t)$ exists for any $0 \leq \tau \leq n$ and the following hold.
    (a) $\varphi^{(\tau)}(t) = \int_{-\infty}^{\infty} (ix)^\tau e^{itx}\,\mathrm{d}F(x)$.
    (b) $\mathbb{E}\left(\xi^\tau\right) = \frac{\varphi^{(\tau)}(0)}{i^\tau}$.
    (c) $\varphi(t) = \sum_{\tau=0}^{n-1} \frac{(it)^\tau}{\tau!}\mathbb{E}\left(\xi^\tau\right) + \frac{(it)^n}{n!}\varepsilon_n(t)$, where $|\varepsilon_n(t)| \leq 3\mathbb{E}\left(|\xi|^n\right)$ and $\varepsilon_n(t) \to 0$ as $t \to 0$.

2. If $\mathbb{E}\left(|\xi|^n\right) < \infty$ for all $n \geq 1$ and

$$\limsup_{n\to\infty} \left(\frac{\mathbb{E}\left(|\xi|^n\right)^{\frac{1}{n}}}{n} \right) = \frac{1}{e \cdot R} < \infty,$$

for $R > 0$, then

$$\varphi(t) = \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \mathbb{E}\left(\xi^n\right)$$

converges for all $|t| < R$.

Proof.

1. Since $\mathbb{E}\left(|\xi|^n\right) < \infty$, we have $\mathbb{E}\left(|\xi|^r\right) < \infty$ for any $r \leq n$ by Corollary 2.4.6. Consider the difference quotient

$$\frac{\varphi(t+h) - \varphi(t)}{h} = \mathbb{E}\left(e^{it\xi} \left(\frac{e^{ih\xi} - 1}{h} \right) \right).$$

Note that

$$\begin{aligned} \left| e^{it\xi} \frac{e^{ih\xi} - 1}{h} \right| &= \left| \frac{e^{ih\xi} - 1}{h} \right| \\ &= \left| \frac{1 + ihx + O\left(h^2\right) - 1}{h} \right| \\ &\leq |\xi|. \end{aligned}$$

So, if $\mathbb{E}\left(|\xi|\right) < \infty$ it follows from the dominated convergence theorem that

$$\begin{aligned} \varphi'(t) &= \lim_{h\to 0} \frac{\varphi(t+h) - \varphi(t)}{h} \\ &= \lim_{h\to 0} \mathbb{E}\left(e^{it\xi} \left(\frac{e^{ih\xi} - 1}{h} \right) \right) \\ &= \mathbb{E}\left(e^{it\xi} \lim_{h\to 0} \left(\frac{e^{ih\xi} - 1}{h} \right) \right) \\ &= \mathbb{E}\left(i\xi e^{it\xi} \right) \\ &= i \int_{-\infty}^{\infty} x e^{itx} \, \mathrm{d}F(x). \end{aligned}$$

The existence of the derivatives $\varphi^{(r)}(t)$ for $1 < r \leq n$ follows by induction. Note that (b) follows immediately from (a). To establish (c), consider the Taylor expansion

$$e^{iy} = \sum_{k=0}^{n-1} \frac{(iy)^k}{k!} + \frac{(iy)^n}{n!} (\cos(\theta_1 y) + i \sin(\theta_2 y))$$

for $y \in \mathbb{R}$, $|\theta_1| \leq 1$, and $|\theta_2| \leq 1$. Then,

$$e^{it\xi} = \sum_{k=0}^{n-1} \frac{(it\xi)^k}{k!} + \frac{(it\xi)^n}{n!} (\cos(\theta_1 \xi) + i \sin(\theta_2 \xi))$$

so that

$$\mathbb{E}\left(e^{it\xi}\right) = \sum_{k=0}^{n-1} \frac{(it)^k}{k!} \mathbb{E}\left(\xi^k\right) + \frac{(it)^n}{n!} \left(\mathbb{E}\left(\xi^n\right) + \varepsilon_n(t)\right),$$

where

$$\varepsilon_n(t) = \mathbb{E}\left(\xi^n (\cos(\theta_1 t\xi) + i \sin(\theta_2 t\xi) - 1)\right).$$

It is clear that $|\varepsilon_n(t)| \leq 3\mathbb{E}\left(|\xi^n|\right)$. Then dominated convergence shows that $\varepsilon_n(t) \to 0$ as $t \to 0$.

2. Let $0 < t_0 < T$. Then
$$\limsup_{n\to\infty}\left(\frac{\left(\mathbb{E}\left(|\xi|^n\right)\right)^{\frac{1}{n}}}{n}\right) \le \frac{1}{et_0}$$
implies that
$$\limsup_{n\to\infty}\left(\frac{\left(\mathbb{E}\left(|\xi|^n e^n t_0^n\right)\right)^{\frac{1}{n}}}{n}\right) \le 1.$$
Thus
$$\limsup_{n\to\infty}\left(\left(\frac{\mathbb{E}\left(|\xi|^n e^n t_0^n\right)}{n^n}\right)^{\frac{1}{n}}\right) < 1.$$
By Stirling's formula we have that
$$\limsup_{n\to\infty}\left(\left(\frac{\mathbb{E}\left(|\xi|^n t_0^n\right)}{n!}\right)^{\frac{1}{n}}\right) < \limsup_{n\to\infty}\left(\left(\frac{\mathbb{E}\left(|\xi|^n e^n t_0^n\right)}{n^n}\right)^{\frac{1}{n}}\right) < 1.$$
Consequently, the series $\sum_{n=0}^{\infty}\mathbb{E}\left(\frac{|\xi|^n t_0^n}{n!}\right)$ converges by the root test and so the series $\sum_{n=0}^{\infty}\frac{(it)^n}{n!}\mathbb{E}\left(\xi^n\right)$ converges for $|t| \le t_0$. By statement 1, for $n \ge 1$ we know that
$$\varphi(t) = \sum_{r=0}^{n}\frac{(it)^r}{r!}\mathbb{E}\left(\xi^r\right) + R_n(t),$$
where $|R_n(t)| \le \frac{3|t|^n}{n!}\mathbb{E}\left(|\xi|^n\right)$. Therefore
$$\varphi(t) = \sum_{r=0}^{\infty}\frac{(it)^r}{r!}\mathbb{E}\left(\xi^r\right)$$
for all $|t| < T$.

■

Remark 7.2.2 Statement 2 of Proposition 7.2.1 gives a sufficient condition for the moments $\mathbb{E}\left(\xi^n\right)$ to determine $\varphi(t)$ uniquely. Indeed, they already determine $\varphi(t)$ for $-R < t < R$. Take s such that $|s| < \frac{R}{2}$ and follow the proof to get
$$\varphi(t) = \sum_{k=0}^{\infty} i^k\frac{(t-s)^k}{k!}\varphi^{(k)}(s),$$
where
$$\varphi^{(k)}(s) = \mathbb{E}\left(\xi^k e^{is\xi}\right),$$
for $-\frac{R}{2} < s < \frac{R}{2}$. Therefore, the moments uniquely determine $\varphi(t)$ for $|t| < \frac{3}{2}R$. Proceed analogously to increase the domain in which $\varphi(t)$ is defined.

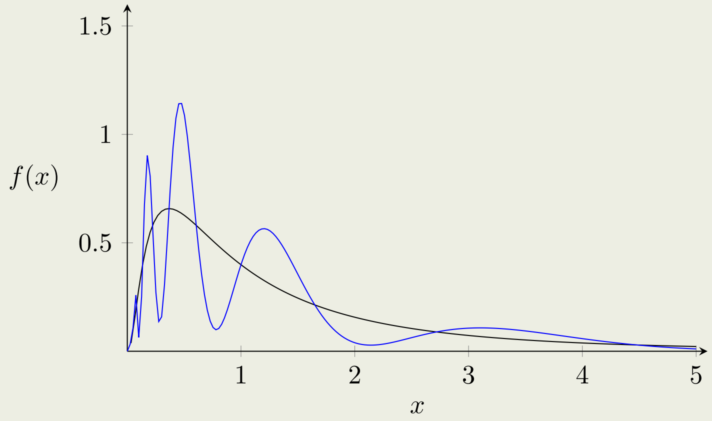
Theorem 7.2.3 — Carleman's test. A sufficient condition for the unique determination of the characteristic function $\varphi(t)$ is
$$\sum_{n=0}^{\infty}\frac{1}{\left(\mathbb{E}\left(\xi^{2n}\right)\right)^{\frac{1}{2n}}} = \infty.$$

Example 7.2.4 If $\mathbb{E}\left(\xi^n\right)$ grows too fast, there may be multiple characteristic functions $\varphi(t)$ with the same moments. Consider a random variable distributed as a standard log-normal distribution,
$$f(x) = \frac{1}{x\sqrt{2\pi}}\exp\left(-\frac{(\log(x))^2}{2}\right),$$

for $x \geq 0$. Consider another random variable with density

$$f_a(x) = f(x) \times (1 + a\sin(2\pi\log(x))),$$

for $x \geq 0$ and $a \in [-1, 1]$.



These seemingly different random variables have the same r^{th} moment. To see this, it suffices to evaluate the integral

$$\int_0^\infty x^{r-1}\exp\left(-\frac{(\log(x))^2}{2}\right)\sin(2\pi\log(x))\,\mathrm{d}x,$$

for $r = 0, 1, \ldots$. With a variable substitution of $s = \log(x)$, the integral becomes

$$\int_{-\infty}^\infty \exp((r-1)s)\exp\left(-\frac{s^2}{2}\right)\sin(2\pi s)\,\mathrm{d}x.$$

The integrand is an L^1 function multiplied by $\sin(2\pi s)$. Therefore, by the Riemann-Lebesgue lemma, the integral is zero, and the random variables have the same moments.

Exercise 7.2.5

1. Verify that if ξ has a standard normal distribution, then $\exp(\xi)$ has a standard log-normal distribution.
2. Show that the r^{th} moment of $\exp(\xi)$ is equal to $\exp\left(\frac{r^2}{2}\right)$.

Notice that the moments grow too fast for the characteristic function to be analytic.

7.3 Inversion Formula

Theorem 7.3.1 — Inversion formula I. Let ξ be a random variable on $(\Omega, \mathscr{F}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with distribution $F(x)$ and characteristic function $\varphi(t)$. If $a < b$ are points of continuity of F then

$$F(b) - F(a) = \lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it}\varphi(t)\,\mathrm{d}t.$$

To prove Theorem 7.3.1 let us introduce

$$S(T) = \int_0^T \frac{\sin x}{x}\,\mathrm{d}x.$$

Note $S(T)$ is a differentiable function with $S(T) > 0$ whenever $T > 0$. Moreover, it can be shown with standard calculus techniques[1] that

$$\lim_{T\to\infty} S(T) = \int_0^\infty \frac{\sin x}{x}\,\mathrm{d}x = \frac{\pi}{2}.$$

[1] https://www.wikihow.com/Integrate-the-Sinc-Function

We therefore know that $S(T)$ is a bounded function and $\sup_{T>0} S(T)$ exists. We also note that

$$\int_0^T \frac{\sin(kx)}{x}\,\mathrm{d}x = \int_0^T \frac{\sin(kx)}{kx}\,\mathrm{d}(kx) = S(kT)$$

for $k > 0$ and when $k < 0$, we have

$$\int_0^T \frac{\sin(kx)}{x}\,\mathrm{d}x = -\int_0^T \frac{\sin(|k|x)}{x}\,\mathrm{d}x = -S(|k|T).$$

Equivalently, we can write

$$\int_0^T \frac{\sin(kx)}{x}\,\mathrm{d}x = \operatorname{sgn}(k)S(|k|T).$$

Moreover, as the integrand is even we know that

$$\int_{-T}^T \frac{\sin(kx)}{x}\,\mathrm{d}x = 2\operatorname{sgn}(k)S(|k|T).$$

Proof. (Theorem 7.3.1). For fixed $T > 0$, let

$$I_T = \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it}\varphi(t)\,\mathrm{d}t = \int_{-T}^T \int_{-\infty}^\infty \frac{e^{-ita} - e^{-itb}}{it} e^{itx}\,\mathrm{d}F(x)\,\mathrm{d}t$$

Note that the integrand is bounded uniformly in (t, x) by

$$\left|\frac{e^{-ita} - e^{-itb}}{it} e^{itx}\right| \le \left|\int_a^b e^{-its}\,\mathrm{d}s\right| \le |b-a|, \tag{7.1}$$

and

$$\int_{-T}^T \int_{-\infty}^\infty |b-a|\,\mathrm{d}F(x)\,\mathrm{d}t = 2T|b-a| < \infty.$$

Therefore, by Fubini's theorem we may exchange the order of integration so that

$$I_T = \int_{-\infty}^\infty \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} e^{itx}\,\mathrm{d}t\,\mathrm{d}F(x) = \int_{-\infty}^\infty \int_{-T}^T \frac{e^{it(x-a)} - e^{it(x-b)}}{it}\,\mathrm{d}t\,\mathrm{d}F(x).$$

Since the domain of integration of the inner integral is symmetric, we can ignore the odd parts of the integrand,

$$I_T = \int_{-\infty}^\infty \int_{-T}^T \frac{\sin(t(x-a)) - \sin(t(x-b))}{t}\,\mathrm{d}t\,\mathrm{d}F(x).$$

Now let

$$J_{T,x} := \int_{-T}^T \frac{\sin(t(x-a)) - \sin(t(x-b))}{t}\,\mathrm{d}t$$

so that

$$I_T = \int_{-\infty}^\infty J_{T,x}\,\mathrm{d}F(x).$$

Note that

$$|J_{T,x}| \le \left|\int_{-T}^T \frac{\sin(t(x-a))}{t}\,\mathrm{d}t\right| + \left|\int_{-T}^T \frac{\sin(t(x-b))}{t}\,\mathrm{d}t\right| \le 4\sup_{T>0} S(T) < \infty,$$

which is integrable with respect to F. Therefore, by the dominated convergence theorem we have that

$$I_\infty := \lim_{T\to\infty} I_T = \int_{-\infty}^\infty \lim_{T\to\infty} J_{T,x}\,\mathrm{d}F(x).$$

Note that

$$\lim_{T\to\infty} J_{T,x} = \begin{cases} 0 & x \notin [a,b] \\ \pi & x \in \{a,b\} \\ 2\pi & x \in (a,b). \end{cases}$$

Therefore,

$$I_\infty = 2\pi(F(b-) - F(a)) - \pi\left(F(a) - F(a-) - (F(b-) - F(b))\right),$$

but as a and b are points of continuity we conclude that

$$F(b) - F(a) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t)\, \mathrm{d}t.$$

■

Corollary 7.3.2 There is a one-to-one correspondence between probability distributions on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ and characteristic functions.

Proof. Let F and G be probability distribution functions on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with the same characteristic function. By Theorem 7.3.1 we note that $F(b) - F(a) = G(b) - G(a)$ for any $a < b$ that are points of continuity of F and G, which are dense in $\mathbb{R}$. Since the collection of open intervals $\{(a, b) : a < b\}$ generates $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$, we must have $F = G$. ■

Exercise 7.3.3 Let X and Y be independent identically distributed random variables with zero mean and unit variance. Prove using characteristic functions that if the distribution F of $\frac{X+Y}{\sqrt{2}}$ is the same as that of X and Y, then F is the standard normal distribution.

Proposition 7.3.4 — Inversion formula II. Let ξ be a random variable on $(\Omega, \mathscr{F}, \mathbb{P}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with distribution $F(x)$ and characteristic function $\varphi(t)$. If $\int_{-\infty}^{\infty} |\varphi(t)|\, \mathrm{d}t < \infty$, then $F(x)$ is absolutely continuous with density $f(x)$, and

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t)\, \mathrm{d}t. \tag{7.2}$$

Proof. Let $f(x)$ be as given by (7.2). Then for $|h| > 0$ consider

$$f(x+h) - f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left(e^{-it(x+h)} - e^{-itx}\right) \varphi(t)\, \mathrm{d}t.$$

Note that

$$\left|e^{-it(x+h)} - e^{-itx}\right| |\varphi(t)| \le 2|\varphi(t)|$$

and

$$\left(e^{-it(x+h)} - e^{-itx}\right) \varphi(t) \overset{|h|\to 0}{\longrightarrow} 0.$$

Therefore, as

$$\int_{-\infty}^{\infty} 2|\varphi(t)|\, \mathrm{d}t = 2 \int_{-\infty}^{\infty} |\varphi(t)|\, \mathrm{d}t < \infty$$

we deduce from the dominated convergence theorem that

$$|f(x+h) - f(x)| \overset{|h|\to 0}{\longrightarrow} 0,$$

which implies that f is continuous. Similarly, one can show that f is differentiable. Hence, f is integrable on $[a, b]$ for $a < b$. Therefore,

$$\begin{aligned}
\int_a^b f(x)\, \mathrm{d}x &= \int_a^b \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-itx} \varphi(t)\, \mathrm{d}t\, \mathrm{d}x \\
&\overset{\text{Fubini.}}{=} \int_{-\infty}^{\infty} \frac{1}{2\pi} \frac{e^{-ita} - e^{-itb}}{it} \varphi(t)\, \mathrm{d}t \\
&\overset{\text{Thm 7.3.1}}{=} F(a) - F(b)
\end{aligned}$$

where $F(x) = \int_{-\infty}^{x} f(y)\,\mathrm{d}y$ for all $x \in \mathbb{R}$ and so F is absolutely continuous. Suppose F has density g, then

$$\varphi(t) = \int_{-\infty}^{\infty} e^{itx} g(x)\,\mathrm{d}x.$$

As $\varphi(t)$ is integrable we can apply the inversion formula for Fourier transforms to deduce that

$$g(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\varphi(t)\,\mathrm{d}t = f(x).$$

Therefore, the density of F is f. ■

7.4 Central Limit Theorems

Exercise 7.4.1 Let ξ be an integer-valued random variable with characteristic function $\varphi_\xi(t)$. Show that

$$\mathbb{P}(\xi = k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ikt}\varphi_\xi(t)\,\mathrm{d}t,$$

for $k \in \mathbb{Z}$.

Example 7.4.2 Characteristic functions can be used to understand the asymptotics of sequences of random variables. For example, consider the random variable $S_n = X_n - X_n'$, where X_n and X_n' are independent and identically binomial random variables with parameters $n \in \mathbb{N}$ and $p \in (0,1)$. Then using statement 1 of Exercise 7.1.5 we have

$$\begin{aligned}
\varphi_{S_n}(t) &= \left(\frac{1}{2} + \frac{1}{2}e^{it}\right)^n \left(\frac{1}{2} + \frac{1}{2}e^{-it}\right)^n \\
&= \frac{1}{4^n}\left(2 + e^{it} + e^{-it}\right)^n \\
&= \frac{1}{2^n}\left(1 + \cos(t)\right)^n \\
&= \left(1 - \sin^2\left(\frac{t}{2}\right)\right)^n.
\end{aligned}$$

Hence, using Exercise 7.4.1 we get that

$$\begin{aligned}
\mathbb{P}(S_n = 0) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(1 - \sin^2\left(\frac{t}{2}\right)\right)^n \mathrm{d}t \\
&= \frac{1}{\pi}\int_{0}^{\pi}\left(1 - \sin^2\left(\frac{t}{2}\right)\right)^n \mathrm{d}t.
\end{aligned}$$

Therefore, with the understanding that the main contribution from an increasingly small neighbourhood of zero, we deduce that

$$\begin{aligned}
\mathbb{P}(S_n = 0) &= \frac{1}{\pi}\int_{0}^{\epsilon}\left(1 - \sin^2\left(\frac{t}{2}\right)\right)^n \mathrm{d}t + \frac{1}{\pi}\int_{\epsilon}^{\pi}\left(1 - \sin^2\left(\frac{t}{2}\right)\right)^n \mathrm{d}t \\
&\le \frac{1}{\pi\sqrt{n}}\int_{0}^{\epsilon\sqrt{n}}\left(1 - \sin^2\left(\frac{s}{2\sqrt{n}}\right)\right)^n \mathrm{d}t + \left(1 - \sin^2\left(\frac{\epsilon}{2}\right)\right)^n \\
&= \frac{1}{\pi\sqrt{n}}\int_{0}^{\infty}\chi_{\{s\le\epsilon\sqrt{n}\}}\left(1 - \sin^2\left(\frac{s}{2\sqrt{n}}\right)\right)^n \mathrm{d}t + o\left(\frac{1}{\sqrt{n}}\right) \\
&\sim \frac{1}{\pi\sqrt{n}}\int_{0}^{\infty}\left(1 - \left(\frac{s}{2\sqrt{n}}\right)^2\right)^n \mathrm{d}t + o\left(\frac{1}{\sqrt{n}}\right) \\
&\sim \frac{1}{\pi\sqrt{n}}\int_{0}^{\infty} e^{-\frac{s^2}{4}}\,\mathrm{d}s + o\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{\pi\sqrt{n}}} + o\left(\frac{1}{\sqrt{n}}\right) \\
&= \frac{1}{\sqrt{\pi n}}\left(1+o(1)\right).
\end{aligned}
$$

The following lemma will be useful for proving Theorem 7.4.5.

Lemma 7.4.3 If $\mathbb{P}$ is a probability measure on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with characteristic function $\varphi(t)$, then

$$
\mathbb{P}\left(|x| \geq \frac{2}{\epsilon}\right) \leq \frac{1}{\epsilon}\int_{-\epsilon}^{\epsilon} 1-\varphi(t)\,\mathrm{d}t
$$

for all $\epsilon > 0$.

Proof. Note that for $x \neq 0$, we have

$$
\int_{-\epsilon}^{\epsilon} 1-e^{itx}\,\mathrm{d}t = 2\epsilon - \frac{e^{it\epsilon}-e^{-it\epsilon}}{ix} = 2\epsilon\left(1-\frac{\sin\epsilon x}{\epsilon x}\right).
$$

Therefore,

$$
\begin{aligned}
\frac{1}{\epsilon}\int_{-\epsilon}^{\epsilon} 1-\varphi(t)\,\mathrm{d}t &= \frac{1}{\epsilon}\int_{-\epsilon}^{\epsilon}\left(1-\int_{\mathbb{R}} e^{itx}\,\mu(\mathrm{d}x)\right)\,\mathrm{d}t \\
&= \frac{1}{\epsilon}\int_{-\epsilon}^{\epsilon}\int_{\mathbb{R}} 1-e^{itx}\,\mu(\mathrm{d}x)\,\mathrm{d}t \\
&\overset{\text{Fubini.}}{=} \int_{\mathbb{R}}\left(\frac{1}{\epsilon}\int_{-\epsilon}^{\epsilon} 1-e^{itx}\,\mathrm{d}t\right)\,\mathbb{P}(\mathrm{d}x) \\
&= \int_{\mathbb{R}} 2\underbrace{\left(1-\frac{\sin\epsilon x}{\epsilon x}\right)}_{\geq 0}\,\mathbb{P}(\mathrm{d}x) \\
&\geq 2\int_{-\infty}^{-\frac{2}{\epsilon}}\left(1-\frac{\sin\epsilon x}{\epsilon x}\right)\,\mathbb{P}(\mathrm{d}x) + 2\int_{\frac{2}{\epsilon}}^{\infty}\left(1-\frac{\sin\epsilon x}{\epsilon x}\right)\,\mathbb{P}(\mathrm{d}x).
\end{aligned}
$$

Note that

$$
1-\frac{\sin\epsilon x}{\epsilon x} \geq \frac{1}{2}
$$

for $|x| \geq \frac{2}{\epsilon}$. Therefore,

$$
\begin{aligned}
\frac{1}{\epsilon}\int_{-\epsilon}^{\epsilon} 1-\varphi(t)\,\mathrm{d}t &\geq \int_{-\infty}^{-\frac{2}{\epsilon}}\mathbb{P}(\mathrm{d}x) + \int_{\frac{2}{\epsilon}}^{\infty}\mathbb{P}(\mathrm{d}x) \\
&= \mathbb{P}\left(|x| \geq \frac{2}{\epsilon}\right).
\end{aligned}
$$

■

Remark 7.4.4 Lemma 7.4.3 shows that the tail of the measure $\mathbb{P}$, hence the existence of moments, is determined by the smoothness of φ at zero.

Theorem 7.4.5 — Levi's Continuity theorem. Let $(\varphi_n)_{n\in\mathbb{N}}$ be the characteristic functions corresponding to the sequence of distribution functions $(F_n)_{n\in\mathbb{N}}$.

1. If $F_n \to F$ weakly, where F is a distribution function, then $\varphi_n(t) \to \varphi(t)$ pointwise for all $t \in \mathbb{R}$, where φ is the characteristic function of F.

2. If $\varphi(t) := \lim_{n\to\infty} \varphi_n(t)$ exists for all $t \in \mathbb{R}$, and is continuous at $t = 0$, then $\varphi(t)$ is a characteristic function of some distribution function F. Moreover, $F_n \to F$ weakly.
3. If $\varphi_n(t)$ is a characteristic function corresponding to a distribution function F_n and $\varphi(t)$ is a characteristic function corresponding to some distribution function F. Then $\varphi_n(t) \to \varphi(t)$ for all $t \in \mathbb{R}$ if and only if $F_n \to F$ weakly.

Proof.

1. Recall that $F_n \to F$ weakly means that
$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f(x)\,\mathrm{d}F_n(x) = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}F(x)$$
for every bounded and real continuous functions f. As $\mathrm{Re}\left(e^{itx}\right)$ and $\mathrm{Im}\left(e^{itx}\right)$ are bounded and real continuous functions, it follows that
$$\begin{aligned}
\lim_{n\to\infty} \varphi_n(t) &= \lim_{n\to\infty} \int_{-\infty}^{\infty} e^{itx}\,\mathrm{d}F_n(x) \\
&= \lim_{n\to\infty} \left(\int_{-\infty}^{\infty} \mathrm{Re}\left(e^{itx}\right)\,\mathrm{d}F_n(x) + \int_{-\infty}^{\infty} \mathrm{Im}\left(e^{itx}\right)\,\mathrm{d}F_n(x) \right) \\
&= \int_{-\infty}^{\infty} \mathrm{Re}\left(e^{itx}\right)\,\mathrm{d}F(x) + \int_{-\infty}^{\infty} \mathrm{Im}\left(e^{itx}\right)\,\mathrm{d}F(x) \\
&= \int_{-\infty}^{\infty} e^{itx}\,\mathrm{d}F(x) \\
&= \varphi(t).
\end{aligned}$$
2. As φ is continuous at zero we know that $\varphi(0) = 1$ as $\varphi_n(0) = 1$ for all $n \in \mathbb{N}$. Therefore, for all $\epsilon > 0$ there is $u > 0$ such that
$$1 - \varphi(t) \leq \frac{\epsilon}{4}$$
for all $t \in [-u, u]$. Hence, using the dominated convergence theorem it follows that
$$\frac{\epsilon}{2} \geq \frac{1}{u} \int_{-u}^{u} 1 - \varphi(t)\,\mathrm{d}t = \lim_{n\to\infty} \frac{1}{u} \int_{-u}^{u} 1 - \varphi_n(t)\,\mathrm{d}t.$$
Using Lemma 7.4.3, there is an $n_0 \in \mathbb{N}$ such that
$$\mathbb{P}_n\left(|x| \geq \frac{2}{u}\right) \leq \frac{1}{u} \int_{-u}^{u} 1 - \varphi_n(t)\,\mathrm{d}t \leq \epsilon \tag{7.3}$$
for all $n \geq n_0$. In particular, one can choose u such that (7.3) holds for all $n \in \mathbb{N}$. Thus, $(\mathbb{P}_n)_{n\in\mathbb{N}}$ is tight. Consequently, using Theorem 6.3.8, we have that for any subsequence $(F_{n_k})_{k\in\mathbb{N}} \subseteq (F_n)_{n\in\mathbb{N}}$, there is a further subsequence that converges weakly to some F. Using statement 1 it follows that $\varphi(t)$ is the characteristic function of this F. In particular, this shows that the limiting distribution function is the same regardless of the subsequence we choose. Suppose that $F_n \not\to F$ weakly. Then there is a point $y \in \mathbb{R} \setminus U_F$, such that a subsequence $\left(F_{n'_k}\right)_{k\geq 1}$ exists with the property that
$$\left|F_{n'_k}(y) - F(y)\right| \geq \epsilon$$
for all $k \in \mathbb{N}$. However, by the above arguments, there is a further subsequence $(F_{n_{k_j}})_{j\in\mathbb{N}}$ which converges to F, giving rise to a contradiction. Therefore, we conclude that $F_n \to F$ weakly.
3. $(\Leftarrow)$. This is statement 1.
$(\Rightarrow)$. As $\varphi(t)$ is a characteristic function, it is continuous at zero so we can use statement 2 to deduce that $F_n \to F$ weakly.

■

Remark 7.4.6 The utility of statement 2 in Theorem 7.4.5 is that it does not assume a priori that $\varphi(t)$ is a characteristic function. Instead, it shows that with the additional requirement that $\varphi(t)$ is continuous at zero, then $\varphi(t)$ must be a characteristic function.

Theorem 7.4.7 — Central Limit Theorem. Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed non-degenerate random variables with $\mathbb{E}\left(\xi_1^2\right) < \infty$ and $S_n = \xi_1 + \cdots + \xi_n$. Then

$$\mathbb{P}\left(\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathbb{V}(S_n)}} \leq x\right) \to \Phi(x) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}}\,\mathrm{d}u$$

as $n \to \infty$ for all $x \in \mathbb{R}$.

Exercise 7.4.8 Show that Theorem 7.4.7 can be written as

$$\frac{S_n - \mathbb{E}\left(S_n\right)}{\sqrt{\mathbb{V}\left(S_n\right)}} \xrightarrow{d} \mathsf{N}(0,1).$$

Theorem 7.4.9 — Lindeberg's Central Limit Theorem. Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent random variables on the $(\Omega, \mathscr{F}, \mathbb{P})$ with $\mathbb{E}\left(\xi_n^2\right) < \infty$ for every $n \in \mathbb{N}$. Let

1. $m_k = \mathbb{E}(\xi_k)$,
2. $\sigma_k^2 = \mathbb{V}(\xi_k) > 0$,
3. $S_n = \xi_1 + \cdots + \xi_n$, and
4. $D_n^2 = \sum_{k=1}^{n} \sigma_k^2$.

Moreover, suppose that

$$\frac{1}{D_n^2} \sum_{k=1}^{n} \mathbb{E}\left(|\xi_k - m_k|^2 \chi_{\{|\xi_k - m_k| \geq \varepsilon D_n\}}\right) \overset{n\to\infty}{\longrightarrow} 0 \tag{7.4}$$

for every $\varepsilon > 0$. Then

$$\frac{S_n - \mathbb{E}(S_n)}{D_n} \xrightarrow{d} \mathsf{N}(0,1).$$

Remark 7.4.10 The condition given in equation (7.4) is known as the Lindeberg condition.

We focus on some special cases in which the Lindeberg condition is satisfied and consequently, the central limit theorem is valid. One of the most prominent is the Lyapunov condition.

Corollary 7.4.11 — Lyapunov's Central Limit Theorem. Assume the conditions of Theorem 7.4.9 and in addition assume that the sequence $(\xi_n)_{n\in\mathbb{N}}$ is such that

$$\frac{1}{D_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}\left(|\xi_k - m_k|^{2+\delta}\right) \to 0 \tag{7.5}$$

for some $\delta > 0$ as $n \to \infty$. Then the sequence $(\xi_n)_{n\in\mathbb{N}}$ satisfies the Lindeberg condition and so the conclusions of Theorem 7.4.7 hold.

Proof. Let $\varepsilon > 0$. Then

$$\begin{aligned}
\mathbb{E}\left(|\xi_k - m_k|^{2+\delta}\right) &\geq \mathbb{E}\left(|\xi_k - m_k|^{2+\delta} \chi_{\{|\xi_k - m_k| \geq \varepsilon D_n\}}\right) \\
&\geq \varepsilon^{\delta} D_n^{\delta} \mathbb{E}\left(|\xi_k - m_k|^{2} \chi_{\{|\xi_k - m_k| \geq \varepsilon D_n\}}\right),
\end{aligned}$$

which implies that

$$\frac{1}{D_n^2}\sum_{k=1}^{n}\mathbb{E}\left(|\xi_k-m_k|^2\chi_{\{|\xi_k-m_k|\geq\epsilon D_n\}}\right)\leq\frac{1}{\varepsilon^\delta}\frac{1}{D_n^{2+\delta}}\sum_{k=1}^{n}\mathbb{E}\left(|\xi_k-m_k|^{2+\delta}\right)$$
$$\overset{n\to\infty}{\longrightarrow}0.$$

Therefore, $(\xi_k)_{k\in\mathbb{N}}$ satisfies the Lindeberg condition. ■

Remark 7.4.12 The condition of equation (7.5) is known as the Lyapunov condition.

Exercise 7.4.13 Under the settings of Corollary 7.4.11, suppose that there exists K such that

$$|\xi_n|\leq K<\infty$$

for all $n\in\mathbb{N}$, and that $D_n\to\infty$ as $n\to\infty$. Show that the sequence $(\xi_n)_{n\in\mathbb{N}}$ satisfies the Lindeberg condition.

Theorem 7.4.7 does not hold when $\mathbb{E}\left(\xi_1^2\right)=\infty$. Let $(\xi_n)_{n\in\mathbb{N}}$ be independent and identically distributed with the Cauchy distribution, that is they have the density

$$f(x)=\frac{\theta}{\pi\left(x^2+\theta^2\right)}$$

for $\theta>0$. One can see that

$$\varphi_{\xi_1}(t)=e^{-\theta|t|}$$

for $t\in\mathbb{R}$, which implies that

$$\varphi_{\frac{S_n}{n}}(t)=\left(\exp\left(\frac{-\theta|t|}{n}\right)\right)^n=e^{-\theta|t|},$$

Therefore, $\frac{S_n}{n}$ also has a Cauchy distribution.

7.5 Berry-Esseen Inequality

Theorem 7.4.7 implies that if $F_n(x)$ is the distribution function of the random variable $\frac{S_n-\mathbb{E}(S_n)}{\sqrt{\mathbb{V}(S_n)}}$, then

$$\sup_{x\in\mathbb{R}}\left(|F_n(x)-\Phi(x)|\right)\overset{n\to\infty}{\longrightarrow}0.$$

At what rate does the left-hand side decay? If $(\xi_n)_{n\in\mathbb{N}}$ are independent and identically distributed with $\xi_1\in L^3$, then we get bounds on the rate of decay.

Theorem 7.5.1 — Berry-Esseen Inequality. Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables with $\mathbb{E}\left(|\xi_1|^3\right)<\infty$. Then

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(\frac{S_n-\mathbb{E}(S_n)}{\sqrt{\mathbb{V}(S_n)}}\leq x\right)-\Phi(x)\right|\leq C\frac{\mathbb{E}\left(|\xi_1-\mathbb{E}(\xi_1)|^3\right)}{\sigma^3\sqrt{n}},$$

where the constant C satisfies

$$\frac{1}{\sqrt{2\pi}}\leq C\leq\frac{1}{2}.$$

Although we do not provide a proof of Theorem 7.5.1, we note that the rate $O\left(\frac{1}{\sqrt{n}}\right)$ is optimal.

Remark 7.5.2 Let $(\xi_n)_{n\in\mathbb{N}}$ be independent and identically distributed Bernoulli random variables with

$\mathbb{P}(\xi_n = 1) = \mathbb{P}(\xi_n = -1) = \frac{1}{2}$. By symmetry, we know that,

$$2\mathbb{P}\left(\sum_{k=1}^{2n} \xi_k < 0\right) + \mathbb{P}\left(\sum_{k=1}^{2n} \xi_k = 0\right) = 1.$$

Therefore,

$$\begin{aligned}\left|\mathbb{P}\left(\sum_{k=1}^{2n} \xi_k < 0\right) - \frac{1}{2}\right| &= \frac{1}{2}\mathbb{P}\left(\sum_{k=1}^{2n} \xi_k = 0\right) \\ &= \frac{1}{2}\binom{2n}{n}\frac{1}{2^{2n}} \\ &\sim \frac{1}{2\sqrt{\pi n}} \\ &= \frac{1}{\sqrt{2\pi}\sqrt{2n}}.\end{aligned}$$

Then $\mathbb{E}\left(|\xi_1|^3\right) = 1 = \sigma$, and Theorem 7.5.1 cannot be improved in terms of $O\left(\frac{1}{\sqrt{n}}\right)$ and $C \geq \frac{1}{\sqrt{2\pi}}$.

Exercise 7.5.3 For a sequence $(\xi_n)_{n\in\mathbb{N}}$ of independent and identically distributed Bernoulli random variables with parameter $p \in (0,1)$, show that there exists $C_1, C_2 > 0$ such that

$$\frac{C_1}{\sqrt{n}} \leq \sup_{x\in\mathbb{R}} |F_{S_n} - \Phi(x)| \leq \frac{C_2}{\sqrt{n}}$$

for sufficiently large n.

7.6 Constructing Characteristic Functions

The following theorems determine whether a function φ is a characteristic function of some measure on $\mathbb{R}$, and if so, whether we can easily construct the underlying measure. The constructions are usually difficult and therefore are not usually covered in great detail. Nevertheless, the proofs in this section serve as great examples of using tools developed in the previous chapter. For further discussions refer to [1].

Exercise 7.6.1

1. For $k \in \mathbb{N}$ let $\varphi_k(t)$ be a characteristic function and let $(\lambda_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}_{>0}$ satisfy $\sum_{k=1}^{\infty} \lambda_k = 1$. Show that $\sum_{k=1}^{\infty} \lambda_k \varphi_k(t)$ is a characteristic function.
2. Let $\varphi(t)$ be a characteristic function. Show that $\mathrm{Re}(\varphi(t))$ is a characteristic function.
3. Let $\varphi(t)$ be a characteristic function. Show that $e^{\lambda(\varphi(t)-1)}$, for $\lambda > 0$, is a characteristic function.

Theorem 7.6.2 — Bochner-Khinchin. Let $\varphi(t)$ be continuous, $t \in \mathbb{R}$, with $\varphi(0) = 1$. A necessary and sufficient condition that $\varphi(t)$ is a characteristic function is that it is positive semi-definite. That is, for all $t_1, \ldots, t_n \in \mathbb{R}$, $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$, and $n = 1, 2, \ldots$ we have

$$\sum_{j,k=1}^{n} \varphi(t_j - t_k)\lambda_j\overline{\lambda}_k \geq 0.$$

To show necessity, we note that if φ is a characteristic function of a real-valued random variable ξ, then

$$\sum_{j,k=1}^{n} \varphi(t_j - t_k)\lambda_j\overline{\lambda}_k = \mathbb{E}(\eta\overline{\eta}) = \mathbb{E}\left(|\eta|^2\right) \geq 0,$$

where $\eta = \sum_{j=1}^{n} \lambda_j e^{it_j \xi}$.

Exercise 7.6.3 Let $\varphi(t)$ be a characteristic function. Show that $\int_0^1 \varphi(ut)\, du$ is a characteristic function.

7.6.1 Polya's Criterion

Theorem 7.6.4 — Polya's criterion. Let a continuous even real-valued function $\varphi(t)$ satisfy $\varphi(t) \geq 0$, $\varphi(0) = 1$, $\varphi(t) \to 0$ as $t \to \infty$ and be convex on $0 \leq t < \infty$. Then $\varphi(t)$ is a characteristic function.

As an observation, we note that the function $\varphi(t)$ must be strictly decreasing over $[0, \infty)$. To see this, we let $0 < r < s$, and consider $t_0 > s$ such that $0 < f(t_0) < \frac{f(r)}{2}$. By convexity, we have

$$f(s) \leq \frac{f(t_0) - f(r)}{t_0 - r}(s - r) + f(r) < f(r).$$

7.6.2 Marcinkiewicz Theorem

Theorem 7.6.5 — Marcinkiewicz's Theorem. If a characteristic function $\varphi(t)$ is of the form $e^{p(t)}$, where $p(t)$ is a polynomial, then the degree of $p(t)$ is at most 2.

Example 7.6.6 As a simple example, e^{-t^4} is not a characteristic function of any real-valued random variables.

7.6.3 Cumulants

Definition 7.6.7 If an expansion

$$\log\left(\varphi_\xi(t)\right) = \sum_{k=0}^{n} \frac{(it)^k}{k!} s_k + o\left(|t|^n\right),$$

exists as $t \to 0$, then the coefficients s_k are called the cumulants of ξ.

Exercise 7.6.8 Show that $\mathbb{E}(\xi) = s_1$, and $\mathbb{V}(\xi) = s_2$.

Remark 7.6.9 If $\xi \sim \mathsf{N}(m, \sigma^2)$ then

- $s_1 = m$,
- $s_2 = \sigma^2$, and
- $s_k = 0$ for $k \geq 3$.

In general, by Theorem 7.6.5, if for a random variable ξ there exists $n \in \mathbb{N}$ such that $s_k = 0$, for all $k \geq n$, then $s_k = 0$ for all $k \geq 3$ and $\xi \sim \mathsf{N}(s_1, s_2)$.

7.6.4 Degenerate distributions

The following theorem shows that a property of the characteristic function of a random variable can lead to a non-trivial conclusion about the nature of the random variable.

Theorem 7.6.10 Let $\varphi(t)$ be a characteristic function of a random variable ξ. If $|\varphi(t_0)| = 1$ for some

$t_0 \neq 0$, then ξ is concentrated at the points $a + nh$ for some $a \in \mathbb{R}$ and where $h = \frac{2\pi}{t_0}$. That is,

$$\sum_{n=-\infty}^{\infty} \mathbb{P}(\xi = a + nh) = 1.$$

Proof. If $|\varphi(t_0)| = 1$ for some $t_0 \neq 0$ then $\varphi(t_0) = e^{it_0 a}$ for some $a \in \mathbb{R}$. Therefore,

$$e^{it_0 a} = \int_{-\infty}^{\infty} e^{it_0 x} \, \mathrm{d}F(x)$$

which implies that

$$1 = \int_{-\infty}^{\infty} e^{it_0(x-a)} \, \mathrm{d}F(x).$$

Equating real parts we see that

$$1 = \int_{-\infty}^{\infty} \cos\left(t_0(x-a)\right) \, \mathrm{d}F(x)$$

which then implies that

$$\int_{-\infty}^{\infty} 1 - \cos\left(t_0(x-a)\right) \, \mathrm{d}F(x) = 0.$$

Since $1 - \cos\left(t_0(x-a)\right) \geq 0$, it follows that $1 = \cos\left(t_0(x-a)\right)$ for $\mathbb{P}$-almost every x. That is, $\mathbb{P}$ is concentrated at the points $x = a + n\left(\frac{2\pi}{t_0}\right)$ for $n \in \mathbb{Z}$. ■

7.7 Solution to Exercises

Exercise 7.1.5

Solution.

1. If $\xi_1 = \mathsf{B}(1, p)$, then
$$\varphi_{\xi_1} = \mathbb{E}\left(e^{it\xi_1}\right) = e^0(1-p) + e^{it}p.$$
Therefore, by statement 4 of Proposition 7.1.3 we deduce that
$$\varphi_\xi(t) = \left(pe^{it} + (1-p)\right)^n.$$

2. Let $\eta = \frac{\xi - m}{\sigma}$. Then $\eta \sim \mathsf{N}(0,1)$ has density
$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$
It is then sufficient to show that $\varphi_\eta(t) = e^{-\frac{t^2}{2}}$. Observe that
$$\begin{aligned}
\varphi_\eta(t) &= \mathbb{E}\left(e^{it\eta}\right) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx - \frac{x^2}{2}} \, \mathrm{d}x \\
&= e^{-\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-it)^2} \, \mathrm{d}x \\
&= e^{-\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty - it}^{\infty - it} e^{-\frac{z^2}{2}} \, \mathrm{d}z \\
&\overset{(1)}{=} e^{-\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} \, \mathrm{d}x \\
&= e^{-\frac{t^2}{2}}.
\end{aligned}$$
We justify (1) using contour integration. Let
$$I_R = \oint_\gamma e^{-\frac{z^2}{2}} \, dz$$
where $\gamma = \gamma_1 \cup \gamma_2 \cup \gamma_3 \cup \gamma_4$ for

- $\gamma_1 := \{z = -u : -R \le u \le R\}$,
- $\gamma_2 := \{z = -R - itu : 0 \le u \le 1\}$,
- $\gamma_3 := \{z = u - it : -R \le u \le R\}$, and
- $\gamma_4 = \{z = R - it(1-u) : 0 \le u \le 1\}$.

Note that

$$\begin{aligned}\int_{\gamma_2} \exp\left(-\frac{z^2}{2}\right) \mathrm{d}z &= \int_0^1 \exp\left(-\frac{(-R-itu)^2}{2}\right)(-it)\,\mathrm{d}u \\ &= \int_0^1 \exp\left(-\frac{R^2}{2}\right)\exp\left(-Rtui\right)\exp\left(\frac{t^2u^2}{2}\right)(-it)\,\mathrm{d}u.\end{aligned}$$

Hence, $\int_{\gamma_2} \exp\left(-\frac{z^2}{2}\right) \mathrm{d}z \overset{R\to\infty}{\longrightarrow} 0$. Similarly, $\int_{\gamma_4} \exp\left(-\frac{z^2}{2}\right) \mathrm{d}z \overset{R\to\infty}{\longrightarrow} 0$. On the other hand,

$$\int_{\gamma_1} \exp\left(-\frac{z^2}{2}\right) \mathrm{d}z = -\int_{-R}^{R} \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u$$

and

$$\begin{aligned}\int_{\gamma_3} \exp\left(-\frac{z^2}{2}\right) &= \int_{-R}^{R} \exp\left(-\frac{(u-it)^2}{2}\right) \mathrm{d}u \\ &\overset{v=u-it}{=} \int_{-R-it}^{R-it} \exp\left(-\frac{v^2}{2}\right) \mathrm{d}v.\end{aligned}$$

Therefore, as $e^{-\frac{z^2}{2}}$ is analytic in the region defined by γ we deduce that

$$0 = \int_{\gamma} \exp\left(-\frac{z^2}{2}\right) \mathrm{d}z = \left(\int_{\gamma_1} + \int_{\gamma_2} + \int_{\gamma_3} + \int_{\gamma_4}\right) \exp\left(-\frac{z^2}{2}\right) \mathrm{d}z.$$

Consequently, as $R \to \infty$ we deduce that

$$0 = -\int_{-\infty}^{\infty} \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u + \int_{-\infty-it}^{\infty-it} \exp\left(-\frac{v^2}{2}\right) \mathrm{d}v$$

which implies that

$$\int_{-\infty}^{\infty} \exp\left(-\frac{u^2}{2}\right) \mathrm{d}u = \int_{-\infty-it}^{\infty-it} \exp\left(-\frac{v^2}{2}\right) \mathrm{d}v.$$

3. Proceeding directly we see that

$$\begin{aligned}\varphi_\xi(t) &= \mathbb{E}\left(e^{it\xi}\right) \\ &= e^{-\lambda}\sum_{k=0}^{\infty} e^{itk}\frac{\lambda^k}{k!} \\ &= e^{-\lambda}\sum_{k=0}^{\infty} \frac{\left(e^{it}\lambda\right)^k}{k!} \\ &= e^{-\lambda}e^{\lambda e^{it}} \\ &= e^{-\lambda+\lambda e^{it}}.\end{aligned}$$

■

Exercise 7.2.5

Solution. Let $\eta = \exp(\xi)$. Then as $\exp(\cdot)$ is a strictly increasing function we note that

$$f_\eta(x) = f_\xi\left(\log(x)\right)\frac{1}{x} = \frac{1}{x\sqrt{2\pi}}\exp\left(-\frac{(\log(x))^2}{2}\right)$$

for $x > 0$. Therefore, η has a standard log-normal distribution. Moreover,

$$\begin{aligned}\mathbb{E}(\eta^r) &= \mathbb{E}(\exp(\xi)^r) \\ &= \int_{-\infty}^{\infty} \exp(rx)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right) \mathrm{d}x \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-r)^2}{2}+\frac{r^2}{2}\right) \mathrm{d}x \\ &= \exp\left(\frac{r^2}{2}\right)\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(x-r)^2}{2}\right) \mathrm{d}x \\ &= \exp\left(\frac{r^2}{2}\right).\end{aligned}$$

■

Exercise 7.3.3

Solution. Let $\varphi(t)$ be the characteristic function of F. Then

$$\varphi(t) = \varphi_{\frac{X+Y}{\sqrt{2}}}(t) = \varphi_X\left(\frac{t}{\sqrt{2}}\right)\varphi_Y\left(\frac{t}{\sqrt{2}}\right) = \varphi\left(\frac{t}{\sqrt{2}}\right)^2.$$

Let $\psi(t) = \log(\varphi(t))$ so that $\psi(t) = 2\psi\left(\frac{t}{\sqrt{2}}\right)$. Consequently,

- $\psi'(t) = \sqrt{2}\psi'\left(\frac{t}{\sqrt{2}}\right)$, and
- $\psi'' = \psi''\left(\frac{t}{\sqrt{2}}\right)$.

As $\psi''(t)$ is continuous at zero we must have that $\psi''(t)$ is a constant. In particular, it follows that

$$\psi''(t) = \psi''(0) = \frac{\varphi(0)\varphi''(0) - (\varphi'(0))^2}{\varphi(0)^2} = \frac{i^2\mathbb{V}(X)}{1} = -1.$$

Therefore, $\psi'(t) = -t + c$. As $\psi'(0) = 0$ we know that $c = 0$ and $\psi(t) = -\frac{t^2}{2}$. Hence, $\varphi(t) = \exp\left(-\frac{t^2}{2}\right)$ and so, by Corollary 7.3.2, F is a standard normal distribution. ■

Exercise 7.4.1

Solution. Note that,

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ikt}\varphi_\xi(t)\,\mathrm{d}t = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ikt}\sum_{m\in\mathbb{Z}} e^{imt}\mathbb{P}(\xi = m)\,\mathrm{d}t.$$

As $\sum_{m\in\mathbb{Z}} e^{imt}\mathbb{P}(\xi = m)$ is absolutely convergent we can interchange the integral and the sum to deduce that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ikt}\varphi_\xi(t)\,\mathrm{d}t = \frac{1}{2\pi}\sum_{m\in\mathbb{Z}}\int_{-\pi}^{\pi} e^{i(m-k)t}\mathbb{P}(\xi = m)\,\mathrm{d}t.$$

Then as

$$\int_{-\pi}^{\pi} e^{i(m-k)t}\,\mathrm{d}t = \begin{cases} 2\pi & m = k \\ 0 & \text{otherwise}\end{cases}$$

we conclude that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ikt}\varphi_\xi(t)\,\mathrm{d}t = \mathbb{P}(\xi = k).$$

■

Exercise 7.4.8

Solution. Set $m = \mathbb{E}(\xi_1)$, $\sigma^2 = \mathbb{V}(\xi_1)$, and $\varphi(t) = \mathbb{E}\left(e^{it(\xi_1 - m)}\right)$. Then by independence, we have

$$\begin{aligned}\varphi_n(t) &= \mathbb{E}\left(\exp\left(it\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathbb{V}(S_n)}}\right)\right)\\ &= \left(\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n.\end{aligned}$$

Since $\mathbb{E}\left(\xi_1^2\right) < \infty$, we have that

$$\varphi(t) = 1 - \frac{\sigma^2 t^2}{2} + o\left(t^2\right)$$

as $t \to 0$. So

$$\varphi_n(t) = \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \to e^{-\frac{t^2}{2}}$$

for all $t \in \mathbb{R}$. This is the characteristic function of $\mathsf{N}(0,1)$ and so the result follows by Theorem 7.4.5. ■

Exercise 7.4.13

Solution. Note that

$$|m_k| = |\mathbb{E}(\xi_k)| \leq \mathbb{E}(|\xi_k|) \leq K$$

for all $k \in \mathbb{N}$. Therefore, $|\xi_k - m_k| \leq 2K$ for all $k \in \mathbb{N}$. As $D_n \to \infty$ monotonically, it follows that for some $N \in \mathbb{N}$ we have

$$\chi_{\{|\xi_k - m_k| \geq \epsilon D_n\}} = 0$$

for $n \geq N$. Therefore, for $n \geq N$ we have

$$\begin{aligned}\frac{1}{D_n^2}\sum_{k=1}^{n}\mathbb{E}\left(|\xi_k - m_k|^2\chi_{\{|\xi_k - m_k|\geq\epsilon D_n\}}\right) &= \frac{1}{D_n^2}\sum_{k=1}^{N-1}\mathbb{E}\left(|\xi_k - m_k|^2\chi_{\{|\xi_k - m_k|\geq\epsilon D_n\}}\right)\\ &\quad + \frac{1}{D_n^2}\sum_{k=N}^{n}\mathbb{E}\left(|\xi_k - m_k|^2\chi_{\{|\xi_k - m_k|\geq\epsilon D_n\}}\right)\\ &= \frac{1}{D_n^2}\sum_{k=1}^{N-1}\mathbb{E}\left(|\xi_k - m_k|^2\chi_{\{|\xi_k - m_k|\geq\epsilon D_n\}}\right)\\ &\leq \frac{4K^2(N-1)}{D_n^2}\\ &\overset{n\to\infty}{\longrightarrow} 0.\end{aligned}$$

Hence, Linderberg's condition is satisfied. ■

Exercise 7.5.3

Solution. From Theorem 7.5.1 we know that $\sup_{x\in\mathbb{R}}|F_{S_n}(x) - \Phi(x)|$ is $O\left(\frac{1}{\sqrt{n}}\right)$ as $\mathbb{E}\left(|\xi_1|^3\right) < \infty$. On the other hand, we note that the continuous approximation $\Phi(x)$ of the discontinuous function $F_{S_n}(x)$, can be at best half the size of the discontinuous. That is,

$$\begin{aligned}\sup_{x\in\mathbb{R}}|F_{S_n}(x) - \Phi(x)| &\geq \frac{1}{2}\sup_{x\in\mathbb{R}}\left|F_{S_n}(x) - F_{S_n}\left(x^-\right)\right|\\ &= \frac{1}{2}\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left(S_n = np + x\sqrt{p(1-p)n}\right)\right|.\end{aligned}$$

From Theorem 4.5.5 we know that that the right-hand side is of order $\frac{1}{\sqrt{n}}$. ■

Exercise 7.6.1

Solution.

1. For $k \in \mathbb{N}$ let ζ_k be a random variable with characteristic function $\varphi_k(t)$. Let η be the random variable where $\mathbb{P}(\eta = k) = \lambda_k$ for $k \in \mathbb{N}$. Note that $\zeta_\eta(\omega) = \sum_{k=1}^{n} \zeta_k \mathbf{1}_{\{\eta=k\}}(\omega)$ is the sum of the product of random variables and hence a random variable. Therefore, we can consider its characteristic function. Namely,

$$\begin{aligned}\varphi_{\zeta_\eta}(t) &= \mathbb{E}\left(e^{it\zeta_\eta}\right)\\ &= \sum_{k=1}^{\infty} \mathbb{E}\left(e^{it\zeta_k}\right)\mathbb{P}(\eta = k)\\ &= \sum_{k=1}^{\infty} \lambda_k \varphi_k(t).\end{aligned}$$

Therefore, $\sum_{n=1}^{\infty} \lambda_k \varphi_k(t)$ is a characteristic function.

2. From statement 1 of Proposition 7.1.3, we note that $\varphi(-t)$ is a characteristic function. Therefore, as $\mathrm{Re}(\varphi(t)) = \frac{1}{2}\varphi(t) + \frac{1}{2}\varphi(-t)$, we can use statement 1 to deduce that $\mathrm{Re}(\varphi(t))$ is a characteristic function.

3. Let $(\xi_k)_{k\in\mathbb{N}}$ be a sequence of independent and identically distributed random variables with the characteristic function $\varphi(t)$, and let η be a random variable independent of the sequence $(\xi_k)_{k\in\mathbb{N}}$ with Po(λ) distribution, for $\lambda \geq 0$. Let $X = \sum_{k=0}^{\eta} \xi_k$, then

$$\begin{aligned}\varphi_X(t) &= \mathbb{E}\left(\exp\left(itX\right)\right)\\ &= \mathbb{E}\left(\exp\left(\sum_{k=0}^{\eta} it\xi_k\right)\right)\\ &= \sum_{n=0}^{\infty} \mathbb{E}\left(\prod_{k=0}^{n} \exp\left(it\xi_k\right)\right)\mathbb{P}(\eta = n)\\ &\overset{(1)}{=} \sum_{n=0}^{\infty}\prod_{k=0}^{n} \mathbb{E}\left(\exp\left(it\xi_k\right)\right)\frac{e^{-\lambda}\lambda^n}{n!}\\ &= e^{-\lambda}\sum_{n=0}^{\infty} \frac{\varphi(t)^n\lambda^n}{n!}\\ &= e^{-\lambda}e^{\lambda\varphi(t)}\\ &= e^{\lambda(\varphi(t)-1)}\end{aligned}$$

where in (1) we have used the independence assumption of the sequence $(\xi_k)_{k\in\mathbb{N}}$. Therefore, $e^{\lambda(\varphi(t)-1)}$ is a characteristic function.

■

Exercise 7.6.3

Proof. Let $\psi(t) = \int_0^1 \varphi(ut)\,\mathrm{d}u$. Then $\psi(0) = \int_0^1 1\,\mathrm{d}u = 1$. Moreover for all $t_1, \dots, t_n \in \mathbb{R}$, $\lambda_1, \dots, \lambda_n \in \mathbb{C}$ and $n = 1, 2, \dots$ we have that

$$\begin{aligned}\sum_{j,k=1}^{n} \psi(t_j - t_k)\lambda_j\overline{\lambda}_k &= \sum_{j,k=1}^{n}\left(\int_0^1 \varphi(u(t_j - t_k))\,\mathrm{d}u\right)\lambda_j\overline{\lambda}_k\\ &= \int_0^1 \sum_{j,k=1}^{n} \varphi(t'_j - t'_k)\lambda_j\overline{\lambda}_k\,\mathrm{d}u\end{aligned}$$

where $t'_i = ut_i \in \mathbb{R}$. Therefore, as $\varphi(t)$ is a characteristic function we know that

$$\sum_{j,k=1}^{n} \varphi\left(t'_j - t'_k\right)\lambda_j\overline{\lambda}_k \geq 0$$

by Theorem 7.6.2. This implies that

$$\sum_{j,k=1}^{n} \psi(t_j - t_k)\lambda_j\overline{\lambda}_k \geq 0$$

and so by Theorem 7.6.2 $\psi(t) = \int_0^t \varphi(ut)\,\mathrm{d}u$ is a characteristic function. ■

Exercise 7.6.8

Solution. On the one hand,

$$\frac{\mathrm{d}}{\mathrm{d}t}\log(\varphi_\xi(t)) = \frac{\varphi_\xi'(t)}{\varphi_\xi(t)}$$

and on the other hand,

$$\frac{\mathrm{d}}{\mathrm{d}t}\log(\varphi_\xi(t)) = \sum_{k=1}^{n} \frac{i(it)^{k-1}}{(k-1)!} s_k + o\left(|t|^{n-1}\right).$$

Substituting $t = 0$ gives

$$is_1 = \frac{i\mathbb{E}(\xi)}{(1)} = i\mathbb{E}(\xi)$$

and so $\mathbb{E}(\xi) = s_1$. Similarly,

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\log(\varphi_\xi(t)) = \frac{\varphi_\xi(t)\varphi_\xi''(t) - \left(\varphi_\xi'(t)\right)^2}{(\varphi_\xi(t))^2}$$

and

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\log(\varphi_\xi(t)) = \sum_{k=2}^{n} \frac{i^2(it)^{k-2}}{(k-2)!} s_k + o\left(|t|^{n-2}\right).$$

Substituting $t = 0$ gives

$$i^2 s_2 = \frac{(1)\left(i^2\mathbb{E}\left(\xi^2\right)\right) - (i\mathbb{E}(\xi))^2}{1^2} = i^2\mathbb{V}(\xi)$$

and so $\mathbb{V}(\xi) = s_2$. ■

Part III. Introduction to Stochastic Analysis

8 Conditional Expectation

When studying stochastic processes $(\xi_\alpha)_{\alpha \in A}$, it is natural to determine how different random variables are related. In particular, we want to know if observing one random variable will give more information on the other random variables in the process. For this, we need the notion of conditional probability and conditional expectation.

8.1 Preliminary Measure Theory

To ensure that our notions of conditional probability and conditional expectation are well-defined it will be useful to make note of the following result in measure theory.

Theorem 8.1.1 — Radon-Nikodym Theorem. Let μ be a finite measure on the measure space $(\Omega, \mathscr{F})$. Let λ be a measure on $\mathscr{F}$ that is absolutely continuous with respect to μ. That is, $\lambda(A) = 0$ whenever $\mu(A) = 0$. Then there exists an $\mathscr{F}$-measurable function f such that

$$\lambda(A) = \int_A f \, \mathrm{d}\mu$$

for all $A \in \mathscr{F}$. Moreover, f is determined uniquely up to sets of measure zero. Consequently, f is called the derivative of λ with respect to μ and is often denoted $f = \frac{\mathrm{d}\lambda}{\mathrm{d}\mu}$.

If ξ is a non-negative random variable on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, and $\mathscr{G} \subseteq \mathscr{F}$ is a σ-algebra, then the measure defined by $Q(G) = \int_G \xi \, \mathrm{d}\mathbb{P}$ for all $G \in \mathscr{G}$ is an absolutely continuous measure with respect to $\mathbb{P}$. Hence, by Theorem 8.1.1 there exists a $\mathscr{G}$-measurable function, f, such that $Q(G) = \int_G f \, \mathrm{d}\mathbb{P}$ for all $G \in \mathscr{G}$. We can extend this naturally to general random variables ξ that are such that $\max\left(\mathbb{E}\left(\xi^+|\mathscr{G}\right), \mathbb{E}\left(\xi^-|\mathscr{G}\right)\right) < \infty$ almost surely.

8.2 Conditional Expectation and Probability

Definition 8.2.1 Let ξ be an integrable random variable on $(\Omega, \mathscr{F}, \mathbb{P})$ and let $\mathscr{G} \subseteq \mathscr{F}$ be a σ-algebra. Then there exists a random variable $\mathbb{E}(\xi|\mathscr{G})$, referred to as the conditional expectation of ξ given $\mathscr{G}$, that satisfies the following.

- $\mathbb{E}(\xi|\mathscr{G})$ is $\mathscr{G}$-measurable and integrable.
- For every $G \in \mathscr{G}$ we have

$$\mathbb{E}(\chi_G \mathbb{E}(\xi|\mathscr{G})) = \int_G \mathbb{E}(\xi|\mathscr{G}) \, \mathrm{d}\mathbb{P} = \int_G \xi \, \mathrm{d}\mathbb{P} = \mathbb{E}(\chi_G \xi).$$

Remark 8.2.2 Theorem 8.1.1 ensures that the conditional expectation of an integrable random variable is unique up to sets of measure zero.

Definition 8.2.3 Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. Then the conditional probability of $B \in \mathscr{F}$ with respect to a σ-algebra $\mathscr{G} \subseteq \mathscr{F}$ is

$$\mathbb{P}(B|\mathscr{G}) = \mathbb{E}(\chi_B|\mathscr{G}).$$

Note that for a fixed $B \in \mathscr{F}$, the conditional probability $\mathbb{P}(B|\mathscr{G})$ is a $\mathscr{G}$-measurable random variable such that

$$\int_G \mathbb{P}(B|\mathscr{G}) \, \mathrm{d}\mathbb{P} = \int_G \chi_B \, \mathrm{d}\mathbb{P} = \mathbb{P}(G \cap B)$$

for all $G \in \mathscr{G}$ as we would expect from traditional notions of conditional probabilities. Moreover, let ξ be a random variable on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and let $\mathscr{G} = \sigma(\{D_1, D_2, \dots\})$ where $\{D_1, D_2, \dots\}$ forms a disjoint partition of Ω. That is,

$$\Omega = \bigcup_{i=1}^{\infty} D_i$$

with each D_i disjoint. Moreover, suppose that $\mathbb{P}(D_i) > 0$ for $i = 1, 2, \dots$. Then all $\mathscr{G}$-measurable functions have the form

$$f(\omega) = \sum_{i=1}^{\infty} c_i \chi_{D_i}(\omega)$$

and thus are constant on each D_i. As $\mathbb{E}(\xi|\mathscr{G})$ is $\mathscr{G}$ measurable it must be of this form. In particular, suppose that $\mathbb{E}(\xi|\mathscr{G}) = \sum_{i=1}^{\infty} c_i \chi_{D_i}(\omega)$, then

$$\begin{aligned}\mathbb{E}\left(\xi \chi_{D_i}\right) &= \int_{D_i} \xi \, d\mathbb{P} \\ &= \int_{D_i} \mathbb{E}(\xi|\mathscr{G}) \, d\mathbb{P} \\ &= c_i \mathbb{P}(D_i),\end{aligned}$$

hence

$$\mathbb{E}(\xi|D_i) := c_i = \frac{\mathbb{E}(\xi \chi_{D_i})}{\mathbb{P}(D_i)}.$$

Similarly,

$$\begin{aligned}\mathbb{P}(B|D_i) &= \mathbb{E}\left(\chi_B | D_i\right) \\ &= \frac{\mathbb{E}\left(\chi_B \chi_{D_i}\right)}{\mathbb{P}(D_i)} \\ &= \frac{\mathbb{P}(B \cap D_i)}{\mathbb{P}(D_i)}.\end{aligned}$$

Hence,

$$\mathbb{P}(B|\mathscr{G}) = \sum_{i=1}^{\infty} \mathbb{P}(B|D_i) \chi_{D_i}.$$

In particular,

$$\mathbb{P}(B|\{\varnothing, \Omega\}) = \mathbb{P}(B).$$

8.3 Properties of Conditional Expectation

Proposition 8.3.1 Let ξ and η be random variables on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and let $\mathscr{G} \subseteq F$ be a σ-algebra. Then,

$$\mathbb{E}(a\xi + b\eta + c \,|\, \mathscr{G}) = a\mathbb{E}(\xi \,|\, \mathscr{G}) + b\mathbb{E}(\eta \,|\, \mathscr{G}) + c$$

almost surely.

Proof. For all $G \in \mathscr{G}$, we have

$$\begin{aligned}\mathbb{E}\left(\chi_G \mathbb{E}(a\xi + b\eta + c \,|\, \mathscr{G})\right) &= \mathbb{E}(\chi_G(a\xi + b\eta + c)) \\ &= a\mathbb{E}(\chi_G \xi) + b\mathbb{E}(\chi_G \eta) + c\mathbb{E}(\chi_G) \\ &= a\mathbb{E}\left(\chi_G \mathbb{E}(\xi \,|\, \mathscr{G})\right) + b\mathbb{E}(\chi_G \mathbb{E}(\eta \,|\, \mathscr{G})) + c\mathbb{E}(\chi_G) \\ &= \mathbb{E}\left(\chi_G (a\mathbb{E}(\xi \,|\, \mathscr{G}) + b\mathbb{E}(\eta \,|\, \mathscr{G}) + c)\right).\end{aligned}$$

As $a\mathbb{E}(\xi|\mathscr{G}) + b\mathbb{E}(\eta|\mathscr{G}) + c$ is $\mathscr{G}$-measurable, it follows that

$$\mathbb{E}(a\xi + b\eta + c \,|\, \mathscr{G}) = a\mathbb{E}(\xi \,|\, \mathscr{G}) + b\mathbb{E}(\eta \,|\, \mathscr{G}) + c.$$

∎

Proposition 8.3.2 Let ξ and η be random variables on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ such that $\xi \leq \eta$ almost surely and let $\mathscr{G} \subseteq \mathscr{F}$ be a σ-algebra. Then,

$$\mathbb{E}(\xi|\mathscr{G}) \leq \mathbb{E}(\eta|\mathscr{G})$$

almost surely.

Proof. Consider the event $G = \{\mathbb{E}(\xi|\mathscr{G}) > \mathbb{E}(\eta|\mathscr{G})\}$. We know $G \in \mathscr{G}$ as $\mathbb{E}(\xi|\mathscr{G})$ and $\mathbb{E}(\eta|\mathscr{G})$ are $\mathscr{G}$-measurable. Thus,

$$\begin{aligned}\int_G \xi \, \mathrm{d}\mathbb{P} &= \int_G \mathbb{E}(\xi|\mathscr{G}) \, \mathrm{d}\mathbb{P} \\ &> \int_G \mathbb{E}(\eta|\mathscr{G}) \, \mathrm{d}\mathbb{P} \\ &= \int_G \eta \, \mathrm{d}\mathbb{P}.\end{aligned}$$

Hence, $\mathbb{P}(G) = 0$ as to not contradict $\xi \leq \mu$ almost surely. Therefore, $\mathbb{E}(\xi|\mathscr{G}) \leq \mathbb{E}(\eta|\mathscr{G})$ almost surely. ■

Corollary 8.3.3 Let η be a random variable on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with $\eta \geq 0$ almost surely. Then for a σ-algebra $\mathscr{G} \subseteq \mathscr{F}$ we have

$$\mathbb{E}(\eta \,|\, \mathscr{G}) \geq 0$$

almost surely.

Proof. Letting $\xi = 0$ in the context of Proposition 8.3.2, it follows that

$$\mathbb{E}(\eta|\mathscr{G}) \geq \mathbb{E}(\xi|\mathscr{G}) = 0$$

almost surely. ■

Exercise 8.3.4 Let ξ and η be random variables defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and let $\mathscr{G} \subseteq \mathscr{F}$ be a σ-algebra.

1. Show that $\mathbb{E}(\xi|\{\varnothing, \Omega\}) = \mathbb{E}(\xi)$.
2. Show that $\mathbb{E}(\xi|\mathscr{F}) = \xi$ almost everywhere.
3. Suppose that ξ is independent of $\mathscr{G}$, which means that ξ and χ_B are independent for all $B \in \mathscr{G}$, then $\mathbb{E}(\xi|\mathscr{G}) = \mathbb{E}(\xi)$.

Theorem 8.3.5 Suppose that $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space and ξ is an integrable random variable taking values in an open interval $I \subseteq \mathbb{R}$. Let $g : I \to \mathbb{R}$ be convex and let $\mathscr{G} \subseteq \mathscr{F}$ be a σ-algebra. If $\mathbb{E}(|g(\xi)|) < \infty$, then

$$\varphi(\mathbb{E}(\xi|\mathscr{G})) \leq \mathbb{E}(\varphi(\xi)|\mathscr{G})$$

almost surely.

Corollary 8.3.6 Suppose that $(\Omega, \mathscr{F}, \mathbb{P})$ is a probability space and ξ is an integrable random variable. Then

$$|\mathbb{E}(\xi|\mathscr{G})| \leq \mathbb{E}(|\xi||\mathscr{G})$$

almost surely.

Proof. One can apply Theorem 8.3.5 with $\varphi(x) = |x|$. Or one can proceed as follows. Note that for any $G \in \mathscr{G}$ we have

$$\int_G |\mathbb{E}(\xi|\mathscr{G})| \, \mathrm{d}\mathbb{P} \leq \int_G \mathbb{E}\left(\xi^+|\mathscr{G}\right) \, \mathrm{d}\mathbb{P} + \int_G \mathbb{E}\left(\xi^-|\mathscr{G}\right) \, \mathrm{d}\mathbb{P}$$

$$
\begin{aligned}
&= \int_G \xi^+ + \xi^- \, \mathrm{d}\mathbb{P} \\
&= \int_G |\xi| \, \mathrm{d}\mathbb{P} \\
&= \int_G \mathbb{E}(|\xi||\mathscr{G}) \, \mathrm{d}\mathbb{P}.
\end{aligned}
$$

As this holds for all $G \in \mathscr{G}$ it follows that $\mathbb{E}(|\xi||\mathscr{G}) \geq |\mathbb{E}(\xi|\mathscr{G})|$ almost surely. ■

Proposition 8.3.7 Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space, ξ an integrable random variable and $\mathscr{F}_1, \mathscr{F}_2 \subseteq \mathscr{F}$ σ-algebras with $\mathscr{F}_1 \subseteq \mathscr{F}_2$. Then,

$$
\mathbb{E}(\mathbb{E}(\xi|\mathscr{F}_2)|\mathscr{F}_1) = \mathbb{E}(\xi|\mathscr{F}_1) = \mathbb{E}(\mathbb{E}(\xi|\mathscr{F}_1)|\mathscr{F}_2) \tag{8.1}
$$

almost surely.

Proof. Let $G \in \mathscr{F}_1$. On the one hand,

$$
\int_G \mathbb{E}(\xi|\mathscr{F}_1) \, \mathrm{d}\mathbb{P} = \int_G \xi \, \mathrm{d}\mathbb{P}.
$$

On the other hand, as $G \in \mathscr{F}_1 \subseteq \mathscr{F}_2$, we have

$$
\int_G \mathbb{E}\left(\mathbb{E}\left(\xi|\mathscr{F}_2\right)|\mathscr{F}_1\right) \, \mathrm{d}\mathbb{P} = \int_G \mathbb{E}(\xi|\mathscr{F}_2) \, \mathrm{d}\mathbb{P} = \int_G \xi \, \mathrm{d}\mathbb{P}.
$$

Hence, by the uniqueness of the conditional expectation of ξ with respect to $\mathscr{F}_1$ we deduce that $\mathbb{E}(\mathbb{E}(\xi|\mathscr{F}_2)|\mathscr{F}_1) = \mathbb{E}(\xi|\mathscr{F}_1)$. Now let $G \in \mathscr{F}_2$. As $\mathbb{E}(\xi|\mathscr{F}_1)$ is $\mathscr{F}_1$-measurable it is also $\mathscr{F}_2$-measurable. Hence, as

$$
\int_G \mathbb{E}(\xi|\mathscr{F}_1) \, \mathrm{d}\mathbb{P} = \int_G \mathbb{E}\left(\mathbb{E}(\xi|\mathscr{F}_1)|\mathscr{F}_2\right) \, \mathrm{d}\mathbb{P}
$$

it follows that $\mathbb{E}(\xi|\mathscr{F}_1) = \mathbb{E}(\mathbb{E}(\xi|\mathscr{F}_1)|\mathscr{F}_2)$. ■

Corollary 8.3.8 For ξ an integrable random variable on the probability space $(\Omega, \mathscr{F}, \mathbb{P})$ we have

$$
\mathbb{E}(\mathbb{E}(\xi|\mathscr{G})) = \mathbb{E}(\xi).
$$

Proof. Take $\mathscr{F}_1 = \{\varnothing, \Omega\}$ and $\mathscr{F}_2 = \mathscr{G}$ in Proposition 8.3.7 and then use statement 1 of Exercise 8.3.4 to conclude. ■

Proposition 8.3.9 Let $(\xi_n)_{n \in \mathbb{N}}$ be a sequence of random variables. Suppose $|\xi_n| \leq \eta$ where $\mathbb{E}(\eta) < \infty$, and $\xi_n \to \xi$ almost surely. Then,

$$
\mathbb{E}(\xi_n|\mathscr{G}) \xrightarrow{\text{a.s.}} \mathbb{E}(\xi|\mathscr{G})
$$

and

$$
\mathbb{E}(|\xi_n - \xi||\mathscr{G}) \xrightarrow{\text{a.s.}} 0.
$$

Proof. Let $\zeta_n = \sup_{m \geq n} |\xi_m - \xi|$. Then $0 \leq |\zeta_n| \leq 2\eta$ and $\zeta_n \to 0$ almost surely, so by the dominated convergence theorem we have $\mathbb{E}(\zeta_n) \xrightarrow{n\to\infty} 0$. By Corollary 8.3.6 it follows that

$$
0 \leq |\mathbb{E}(\xi_n|\mathscr{G}) - \mathbb{E}(\xi|\mathscr{G})| \leq \mathbb{E}(|\xi_n - \xi||\mathscr{G}) \leq \mathbb{E}(\zeta_n|\mathscr{G}). \tag{8.2}
$$

Since the sequence $(\mathbb{E}(\zeta_n|\mathscr{G})(\omega))_{n\in\mathbb{N}}$ is decreasing in n for fixed ω, its limit exists ω-almost surely. In particular,

$$
0 \leq \mathbb{E}\left(\lim_{n\to\infty} \mathbb{E}(\zeta_n|\mathscr{G})\right) \leq \lim_{n\to\infty} \mathbb{E}\left(\mathbb{E}(\zeta_n|\mathscr{G})\right) = \lim_{n\to\infty} \mathbb{E}(\zeta_n) = 0.
$$

Hence, $\lim_{n\to\infty} \mathbb{E}(\zeta_n|\mathscr{G}) = 0$ almost everywhere, so from (8.2) it follows that

$$
\mathbb{E}(\xi_n|\mathscr{G}) \xrightarrow{\text{a.s}} \mathbb{E}(\xi|\mathscr{G})
$$

and

$$
\mathbb{E}(|\xi_n - \xi||\mathscr{G}) \xrightarrow{\text{a.s}} 0.
$$

■

Corollary 8.3.10 Let ξ and η be random variables on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with ξ, η and $\xi\eta$ integrable. Let $\mathscr{G} \subseteq \mathscr{F}$ be a σ-algebra and suppose that η is $\mathscr{G}$-measurable. Then

$$\mathbb{E}(\xi\eta|\mathscr{G}) = \eta\mathbb{E}(\xi|\mathscr{G})$$

almost everywhere.

Proof. Note that $\mathbb{E}(\xi|\mathscr{G})$ and η are $\mathscr{G}$-measurable.
Step 1. Consider $\eta = \chi_A$ for $A \in \mathscr{G}$.
Let $B \in \mathscr{G}$. On the one hand,

$$\begin{aligned}\mathbb{E}\left(\chi_B \mathbb{E}\left(\xi\eta|\mathscr{G}\right)\right) &= \mathbb{E}\left(\chi_B \xi\eta\right) \\ &= \mathbb{E}\left(\chi_{A\cap B}\xi\right).\end{aligned}$$

On the other hand,

$$\begin{aligned}\mathbb{E}\left(\chi_B \eta\mathbb{E}(\xi|\mathscr{G})\right) &= \mathbb{E}\left(\chi_{A\cap B}\mathbb{E}\left(\xi|\mathscr{G}\right)\right) \\ &= \mathbb{E}\left(\chi_{A\cap B}\xi\right),\end{aligned}$$

where we have used that $A \cap B \in \mathscr{G}$. Thus, $\mathbb{E}(\xi\eta|\mathscr{G}) = \eta\mathbb{E}(\xi|\mathscr{G})$.
Step 2. Consider η to be a simple random variable.
We can extend the result of step 1 to simple random variables by using Proposition 8.3.1.
Step 3. Consider η a general integrable random variable.
Any integrable random variable η can be approximated by simple functions $(\eta_n)_{n\in\mathbb{N}}$ with $|\eta_n| \le \eta$. Moreover, $\eta_n\xi \to \eta\xi$ almost surely, with $|\eta_n\xi| \le |\eta\xi|$. Therefore, as $\mathbb{E}(\eta\xi) < \infty$ we can apply statement 1 of Proposition 8.3.9 to deduce that $\mathbb{E}(\eta_n\xi|\mathscr{G}) \to \mathbb{E}(\eta\xi|\mathscr{G})$. From step 2 we know that $\mathbb{E}(\eta_n\xi|\mathscr{G}) = \eta_n\mathbb{E}(\xi|\mathscr{G})$. Therefore, as $\eta_n\mathbb{E}(\xi|\mathscr{G}) \to \eta\mathbb{E}(\xi|\mathscr{G})$ we deduce that $\mathbb{E}(\eta\xi|\mathscr{G}) = \eta\mathbb{E}(\xi|\mathscr{G})$. ∎

8.4 Conditioning on a Random Variable

Definition 8.4.1 The conditional expectation of a random variable ξ with respect to a random variable η is

$$\mathbb{E}(\xi|\eta) := \mathbb{E}(\xi|\sigma(\eta)),$$

where $\sigma(\eta)$ is the σ-algebra generated by η.

Theorem 8.4.2 Let μ and η be random variables such that μ is $\sigma(\eta)$-measurable. Then there exists a Borel-measurable function $f : \mathbb{R} \to \mathbb{R}$ such that

$$\mu = f(\eta).$$

In particular, there exists a Borel-measurable function $g : \mathbb{R} \to \mathbb{R}$ such that

$$\mathbb{E}(\mu|\eta) = g(\eta).$$

Proof. Step 1. Consider $\mu = \sum_{i=1}^n c_j \chi_{A_j}$, with $(A_j)_{j=1}^n$ partitioning Ω.
As μ is $\sigma(\eta)$-measurable it must be the case that $A_j \in \sigma(\eta)$ for $j = 1, \dots, n$. Hence, for $j = 1, \dots, n$ there exists $B_j \in \mathscr{B}(\mathbb{R})$ such that $\eta^{-1}(B_j) = A_j$. It is clear that $(B_j)_{j=1}^n$ partitions $\eta(\Omega)$. Hence, set

$$f(x) = \begin{cases} \sum_{i=1}^n c_j \chi_{B_j}(x) & x \in \bigcup_{j=1}^n B_j \\ 0 & \text{otherwise,} \end{cases}$$

so that $f(\eta(\omega)) = \mu(\omega)$ as required.
Step 2. Consider μ a general random variable.
We can approximate μ with a sequence of simple random variables $(\mu_n)_{n\in\mathbb{N}}$ such that $\mu_n(\omega) \to \mu(\omega)$ for $\omega \in \Omega$. By step 1 we have Borel-measurable functions f_n such that $\mu_n = f_n(\eta)$. Set

$$f(x) = \begin{cases} \lim_{n\to\infty}(f_n(x)) & \text{if it exists on } \eta(\Omega) \\ 0 & \text{otherwise.} \end{cases}$$

Then $f(x)$ is Borel-measurable and

$$\mu(\omega) = \lim_{n\to\infty}(\mu_n(\omega)) = \lim_{n\to\infty}(f_n(\eta(\omega))) = f(\eta(\omega))$$

as required. ■

Example 8.4.3 Consider real-valued random variables X and Y defined on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Assume the random vector (X, Y) has a continuous joint density $f_{X,Y}(x,y) > 0$. Recall that X has density $f_X(x) = \int_\Omega f_{X,Y}(x,y)\,\mathrm{d}y$ and Y has density $f_Y(y) = \int_\Omega f_{X,Y}(x,y)\,\mathrm{d}x$. Assume $f_X(x), f_Y(y) > 0$ almost everywhere in $\mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ is a Borel measurable function with $\mathbb{E}(|h(X)|) < \infty$. By Theorem 8.4.2, we know that $\mathbb{E}(h(X) \mid Y) = \phi(Y)$ for some unique Borel-measurable ϕ almost everywhere. That is,

$$\mathbb{E}(\chi_A \phi(Y)) = \mathbb{E}(\chi_A h(X))$$

for all $A \in \sigma(Y)$. Since $A \in \sigma(Y) \subseteq \mathscr{F}$, we know that $A = Y^{-1}(B)$ for some $B \in \mathscr{B}(\mathbb{R})$. Then,

$$\begin{aligned}
\mathbb{E}(\chi_A h(X)) &= \mathbb{E}(\chi_B(Y) h(X)) \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x)\chi_B(y) f_{X,Y}(x,y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x)\chi_B(y)\frac{f_{X,Y}(x,y)}{f_Y(y)} f_Y(y)\,\mathrm{d}y\,\mathrm{d}x \\
&\overset{\text{Fubini}}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x)\chi_B(y)\frac{f_{X,Y}(x,y)}{f_Y(y)} f_Y(y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \chi_B(y)\left(\int_{-\infty}^{\infty} h(x)\frac{f_{X,Y}(x,y)}{f_Y(y)}\,\mathrm{d}x\right) f_Y(y)\,\mathrm{d}y.
\end{aligned}$$

So by the uniqueness of ϕ we deduce that

$$\phi(y) = \int_{-\infty}^{\infty} h(x)\frac{f_{X,Y}(x,y)}{f_Y(y)}\,\mathrm{d}x.$$

Exercise 8.4.4

1. Let X and Y be random variables taking values in $\mathbb{N}$, with joint mass $p_{X,Y}(x,y)$ for $x, y \in \mathbb{N}$. Assume $h : \mathbb{N} \to \mathbb{R}$ is such that $\mathbb{E}(|h(X)|) < \infty$. Verify that $\mathbb{E}(h(X) \mid Y) = \phi(Y)$ where

$$\phi(y) = \sum_{x\in\mathbb{N}} h(x)\frac{p_{X,Y}(x,y)}{p_Y(y)}$$

for $p_Y(y) \neq 0$.

2. Consider random variables Z_1, Z_2 on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with $Z_1 \sim \text{Po}(\lambda_1)$ and $Z_2 \sim \text{Po}(\lambda_2)$. Assuming $p = \frac{\lambda_1}{\lambda_1+\lambda_2}$, show that

$$\mathbb{P}(Z_1 = k \mid Z_1 + Z_2 = n) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Proposition 8.4.5 If $\mathbb{E}(\xi^2) < \infty$, then

$$\min_f \mathbb{E}\left((\xi - f(\eta))^2\right) = \mathbb{E}\left((\xi - \mathbb{E}(\xi|\eta))^2\right),$$

where the minimum is taken over all $\sigma(\eta)$-measurable functions such that $\mathbb{E}\left(f^2(\eta)\right) < \infty$.

8.5 Solution to Exercises

Exercise 8.3.4

Solution.

1. As $\mathbb{E}(\xi)$ is a constant it is $\{\varnothing, \Omega\}$-measurable. Moreover,
$$\int_{\varnothing} \xi \,\mathrm{d}\mathbb{P} = 0 = \int_{\varnothing} \mathbb{E}(\xi) \,\mathrm{d}\mathbb{P}$$
and
$$\int_{\Omega} \xi \,\mathrm{d}\mathbb{P} = \mathbb{E}(\xi) = \mathbb{E}(\xi) \int_{\Omega} \mathrm{d}\mathbb{P} = \int_{\Omega} \mathbb{E}(\xi) \,\mathrm{d}\mathbb{P}.$$
Therefore, $\mathbb{E}\left(\xi | \{\varnothing, \Omega\}\right) = \mathbb{E}(\xi)$.
2. As ξ is $\mathscr{F}$-measurable it follows that $\mathbb{E}(\xi|\mathscr{F}) = \xi$.
3. Let $B \in \mathscr{G}$, then
$$\begin{aligned}\int_{B} \xi \,\mathrm{d}\mathbb{P} &= \mathbb{E}\left(\xi \chi_B\right) \\ &= \mathbb{E}(\xi)\mathbb{E}\left(\chi_B\right) \\ &= \mathbb{E}(\xi) \int_{B} \mathrm{d}\mathbb{P} \\ &= \int_{B} \mathbb{E}(\xi) \,\mathrm{d}\mathbb{P}.\end{aligned}$$

■

Exercise 8.4.4

Solution.

1. It suffices to consider singleton sets $\{Y = y\}$ as Y takes values in $\mathbb{N}$. Proceeding as in Example 8.4.3 we see that
$$\begin{aligned}\mathbb{E}\left(h(X)\chi_{\{Y=y\}}\right) &= \sum_{x\in\mathbb{N}} h(x) p_{X,Y}(x,y) \\ &= \sum_{x\in\mathbb{N}} h(x) \frac{p_{X,Y}(x,y)}{p_Y(y)} p_Y(y) \\ &= \mathbb{E}\left(\phi(y)\chi_{\{Y=y\}}\right).\end{aligned}$$
2. Recall that $Z_1 + Z_2 \sim \mathrm{Po}(\lambda_1 + \lambda_2)$. Let $h = \chi_{\{Z_1 = k\}}$, then
$$\mathbb{P}(Z_1 = k | Z_1 + Z_2 = n) = \mathbb{E}\left(h(Z_1) | Z_1 + Z_2\right)(n).$$
Hence,
$$\begin{aligned}\mathbb{P}(Z_1 = k | Z_1 + Z_2 = n) &= \sum_{x\in\mathbb{N}} h(x) \frac{\mathbb{P}(Z_1 = x, Z_1 + Z_2 = n)}{\mathbb{P}(Z_1 + Z_2 = n)} \\ &= \frac{\mathbb{P}(Z_1 = k)\mathbb{P}(Z_2 = n-k)}{\mathbb{P}(Z_1 + Z_2 = n)} \\ &= \frac{\left(\frac{e^{-\lambda_1}\lambda_1^k}{k!}\right)\left(\frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}\right)}{\left(\frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}\right)} \\ &= \frac{n!}{k!(n-k)!}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k \left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}.\end{aligned}$$

■

References

[1] Eugene Lukacs. “A Survey of the Theory of Characteristic Functions”. In: **Advances in Applied Probability** 4.1 (1972), pp. 1–38. ISSN: 00018678. URL: http://www.jstor.org/stable/1425805 (visited on 08/19/2022).

[2] Patrick Billingsley. **Probability and Measure**. 3rd ed. Wiley Series in Probability and Mathematical Statistics. New York: Wiley, 1995. 593 pp. ISBN: 978-0-471-00710-4.

[3] Wikipedia contributors. **Proofs of convergence of random variables — Wikipedia, The Free Encyclopedia**. 2023. URL: https://en.wikipedia.org/w/index.php?title=Proofs_of_convergence_of_random_variables&oldid=1180638140.

[4] Samuel Lam Ivan Kirev. **Imperial Probability Theory**. 2023. URL: https://github.com/Samuel-CHLam/Imperial-Probability-Theory.