Advanced Probability

Alexander Sokol

Anders Rønn-Nielsen

Department of Mathematical Sciences
University of Copenhagen

Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen

ISBN 978-87-7078-999-8

Contents

Preface

The purpose of this monograph is to present a detailed introduction to selected fundamentals of modern probability theory. The focus is in particular on discrete-time and continuous-time processes, including the law of large numbers, Lindeberg's central limit theorem, martingales, the martingale convergence theorem and the martingale central limit theorem, as well as basic results on Brownian motion. The reader is assumed to have a reasonable grasp of basic analysis and measure theory, as can be obtained through Hansen (2009), Carothers (2000) or Ash (1972), for example.

We have endeavoured throughout to present the material in a logical fashion, with detailed proofs allowing the reader to perceive not only the big picture of the theory, but also to understand the finer elements of the methods of proof used. Exercises are given at the end of each chapter, with hints for the exercises given in the appendix. The exercises form an important part of the monograph. We strongly recommend that any reader wishing to acquire a sound understanding of the theory spends considerable time solving the exercises.

While we share the responsibility for the ultimate content of the monograph and in particular any mistakes therein, much of the material is based on books and lecture notes from other individuals, in particular "Videregående sandsynlighedsregning" by Martin Jacobsen, the lecture notes on weak convergence by Søren Tolver Jensen, "Sandsynlighedsregning på Målteoretisk Grundlag" by Ernst Hansen as well as supplementary notes by Ernst Hansen, in particular a note on the martingale central limit theorem. We are also indebted to Ketil Biering Tvermosegaard, who diligently translated the lecture notes by Martin Jacobsen and thus eased the migration of their contents to their present form in this monograph.

We would like to express our gratitude to our own teachers, particularly Ernst Hansen, Martin Jacobsen and Søren Tolver Jensen, who taught us measure theory and probability theory.

Also, many warm thanks go to Henrik Nygaard Jensen, who meticulously read large parts of the manuscript and gave many useful comments.

Alexander Sokol
Anders Rønn-Nielsen
København, August 2012

Since the previous edition of the book, a number of misprints and errors have been corrected, and various other minor amendments have been made. We are grateful to the many students who have contributed to the monograph by identifying mistakes and suggesting improvements.

Alexander Sokol
Anders Rønn-Nielsen
København, June 2013

Chapter 1

Sequences of random variables

In this chapter, we will consider sequences of random variables and the basic results on such sequences, in particular the strong law of large numbers, which formalizes the intuitive notion that averages of independent and identically distributed events tend to the common mean.

We begin in Section 1.1 by reviewing the measure-theoretic preliminaries for our later results. In Section 1.2, we discuss modes of convergence for sequences of random variables. The results given in this section are fundamental to much of the remainder of this monograph, as well as modern probability in general. In Section 1.3, we discuss the concept of independence for families of σ-algebras, and as an application, we prove the Kolmogorov zero-one law, which shows that for sequences of independent variables, events which, colloquially speaking, depend only on the tail of the sequence either have probability zero or one. In Section 1.4, we apply the results of the previous sections to prove criteria for the convergence of sums of independent variables. Finally, in Section 1.5, we prove the strong law of large numbers, arguably the most important result of this chapter.

1.1 Measure-theoretic preliminaries

As noted in the introduction, we assume given a level of familiarity with basic real analysis and measure theory. Some of the main results assumed to be well-known in the following are reviewed in Appendix A. In this section, we give an independent review of some basic results, and review particular notation related to probability theory.

We begin by recalling some basic definitions. Let Ω be some set. A σ-algebra $\mathcal{F}$ on Ω is a set of subsets of Ω with the following three properties: $\Omega \in \mathcal{F}$, if $F \in \mathcal{F}$ then $F^c \in \mathcal{F}$ as well, and if $(F_n)_{n\geq 1}$ is a sequence of sets with $F_n \in \mathcal{F}$ for $n \geq 1$, then $\cup_{n=1}^\infty F_n \in \mathcal{F}$ as well. We refer to the second condition as $\mathcal{F}$ being stable under complements, and we refer to the third condition as $\mathcal{F}$ being stable under countable unions. From these stability properties, it also follows that if $(F_n)_{n\geq 1}$ is a sequence of sets in $\mathcal{F}$, $\cap_{n=1}^\infty F_n \in \mathcal{F}$ as well. We refer to the pair $(\Omega, \mathcal{F})$ as a measurable space, and we refer to the elements of $\mathcal{F}$ as events. A probability measure P on $(\Omega, \mathcal{F})$ is a mapping $P : \mathcal{F} \to [0, 1]$ such that $P(\emptyset) = 0$, $P(\Omega) = 1$ and whenever (F_n) is a sequence of disjoint sets in $\mathcal{F}$, it holds that $\sum_{n=1}^\infty P(F_n)$ is convergent and $P(\cup_{n=1}^\infty F_n) = \sum_{n=1}^\infty P(F_n)$. We refer to the latter property as the σ-additivity of the probability measure P. We refer to the triple $(\Omega, \mathcal{F}, P)$ as a probability space.

Next, assume given a measure space $(\Omega, \mathcal{F})$, and let $\mathbb{H}$ be a set of subsets of Ω. We may then form the set $\mathfrak{A}$ of all σ-algebras on Ω containing $\mathbb{H}$, this is a subset of the power set of the power set of Ω. We may then define $\sigma(\mathbb{H}) = \cap_{\mathcal{F}\in\mathfrak{A}}\mathcal{F}$, the intersection of all σ-algebras in $\mathfrak{A}$, that is, the intersection of all σ-algebras containing $\mathbb{H}$. This is a σ-algebra as well, and it is the smallest σ-algebra on Ω containing $\mathbb{H}$ in the sense that for any σ-algebra $\mathcal{G}$ containing $\mathbb{H}$, we have $\mathcal{G} \in \mathfrak{A}$ and therefore $\sigma(\mathbb{H}) = \cap_{\mathcal{F}\in\mathfrak{A}}\mathcal{F} \subseteq \mathcal{G}$. We refer to $\sigma(\mathbb{H})$ as the σ-algebra generated by $\mathbb{H}$, and we say that $\mathbb{H}$ is a generating family for $\sigma(\mathbb{H})$.

Using this construction, we may define a particular σ-algebra on the Euclidean spaces: The Borel σ-algebra $\mathcal{B}_d$ on $\mathbb{R}^d$ for $d \geq 1$ is the smallest σ-algebra containing all open sets in $\mathbb{R}^d$. We denote the Borel σ-algebra on $\mathbb{R}$ by $\mathcal{B}$.

Next, let $(F_n)_{n\geq 1}$ be a sequence of sets in $\mathcal{F}$. If $F_n \subseteq F_{n+1}$ for all $n \geq 1$, we say that $(F_n)_{n\geq 1}$ is increasing. If $F_n \supseteq F_{n+1}$ for all $n \geq 1$, we say that $(F_n)_{n\geq 1}$ is decreasing. Assume that $\mathbb{D}$ is a set of subsets of Ω such that the following holds: $\Omega \in \mathbb{D}$, if $F, G \in \mathbb{D}$ with $F \subseteq G$ then $G \setminus F \in \mathbb{D}$ and if $(F_n)_{n\geq 1}$ is an increasing sequence of sets in $\mathbb{D}$ then $\cup_{n=1}^\infty F_n \in \mathbb{D}$. If this is the case, we say that $\mathbb{D}$ is a Dynkin class. Furthermore, if $\mathbb{H}$ is a set of subsets of Ω such that whenever $F, G \in \mathbb{H}$ then $F \cap G \in \mathbb{H}$, then we say that $\mathbb{H}$ is stable under finite intersections. These two concepts combine in the following useful manner, known as Dynkin's lemma: Let $\mathbb{D}$ be a Dynkin class on Ω, and $\mathbb{H}$ be a set of subsets of Ω which is stable under finite intersections. If $\mathbb{H} \subseteq \mathbb{D}$, then $\sigma(\mathbb{H}) \subseteq \mathbb{D}$.

Dynkin's lemma is useful when we desire to show that some property holds for all sets $F \in \mathcal{F}$. A consequence of Dynkin's lemma is that if P and Q are two probability measures on $\mathcal{F}$ which are equal on a generating family for $\mathcal{F}$ which is stable under finite intersections, then P and Q are equal on all of $\mathcal{F}$.

Assume given a probability space $(\Omega, \mathcal{F}, P)$. The probability measure satisfies that for any pair of events $F, G \in \mathcal{F}$ with $F \subseteq G$, $P(G \setminus F) = P(G) - P(F)$. Also, if (F_n) is an increasing sequence in $\mathcal{F}$, then $P(\cup_{n=1}^{\infty} F_n) = \lim_{n\to\infty} P(F_n)$, and if (F_n) is a decreasing sequence in $\mathcal{F}$, then $P(\cap_{n=1}^{\infty} F_n) = \lim_{n\to\infty} P(F_n)$. These two properties are known as the upwards and downwards continuity of probability measures, respectively.

Given a mapping $X : \Omega \to \mathbb{R}$, we say that X is $\mathcal{F}$-$\mathcal{B}$ measurable if it holds for all $A \in \mathcal{B}$ that $X^{-1}(B) \in \mathcal{F}$, where we use the notation $X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\}$. Letting the σ-algebras involved be implicit, we may simply say that X is measurable. A measurable mapping $X : \Omega \to \mathbb{R}$ is referred to as a random variable. For convenience, we also write $(X \in B)$ instead of $X^{-1}(B)$ when $B \subseteq \mathbb{R}$. Measurability of X ensures that whenever $B \in \mathcal{B}$, the subset $(X \in B)$ of Ω is $\mathcal{F}$ measurable, such that $P(X \in B)$ is well-defined. Furthermore, the integral $\int |X| \,\mathrm{d}P$ is well-defined. In the case where it is finite, we say that X is integrable, and the integral $\int X \,\mathrm{d}P$ is then well-defined and finite. We refer to this as the mean of X and write $EX = \int X \,\mathrm{d}P$. In the case where $|X|^p$ is integrable for some $p > 0$, we say that X has p'th moment and write $EX^p = \int X^p \,\mathrm{d}P$.

Also, if $(X_i)_{i\in I}$ is a family of variables, we denote by $\sigma((X_i)_{i\in I})$ the σ-algebra generated by $(X_i)_{i\in I}$, meaning the smallest σ-algebra on Ω making X_i measurable for all $i \in I$, or equivalently, the smallest σ-algebra containing $\mathbb{H}$, where $\mathbb{H}$ is the class of subsets $(X_i \in B)$ for $i \in I$ and $B \in \mathcal{B}$. Also, for families of variables, we write $(X_i)_{i\in I}$ and (X_i) interchangably, understanding that the index set is implicit in the latter case.

1.2 Convergence of sequences of random variables

We are now ready to introduce sequences of random variables and consider their modes of convergence. For the remainder of the chapter, we work within the context of a probability space $(\Omega, \mathcal{F}, P)$.

Definition 1.2.1. *A sequence of random variables $(X_n)_{n\geq 1}$ is a sequence of mappings from Ω to $\mathbb{R}$ such that each X_n is a random variable.*

If $(X_n)_{n\geq 1}$ is a sequence of random variables, we also refer to $(X_n)_{n\geq 1}$ as a discrete-time stochastic process, or simply a stochastic process. These names are interchangeable. For brevity, we also write (X_n) instead of $(X_n)_{n\geq 1}$. In Definition 1.2.1, all variables are assumed to take values in $\mathbb{R}$, in particular ruling out mappings taking the values ∞ or $-\infty$ and ruling out variables with values in $\mathbb{R}^d$. This distinction is made solely for convenience, and if need be, we will also refer to sequences of random variables with values in $\mathbb{R}^d$ or other measure spaces as sequences of random variables.

A natural first question is when sequences of random variables exist with particular distributions. For example, does there exists a sequence of variables (X_n) such that $(X_1, \ldots, X_n)$ are independent for all $n \geq 1$ and such that for each $n \geq 1$, X_n has some particular given distribution? Such questions are important, and will be relevant for our later construction of examples and counterexamples, but are not our main concern here. For completeness, results which will be sufficient for our needs are given in Appendix A.3.

The following fundamental definition outlines the various modes of convergence of random variables to be considered in the following.

Definition 1.2.2. *Let (X_n) be a sequence of random variables, and let X be some other random variable.*

(1). X_n converges in probability to X if for all $\varepsilon > 0$, $\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$.

(2). X_n converges almost surely to X if $P(\lim_{n\to\infty} X_n = X) = 1$.

(3). X_n converges in $\mathcal{L}^p$ to X for some $p \geq 1$ if $\lim_{n\to\infty} E|X_n - X|^p = 0$.

(4). X_n converges in distribution to X if for all bounded, continuous mappings $f : \mathbb{R} \to \mathbb{R}$, $\lim_{n\to\infty} Ef(X_n) = Ef(X)$.

In the affirmative, we write $X_n \xrightarrow{P} X$, $X_n \xrightarrow{\text{a.s.}} X$, $X_n \xrightarrow{\mathcal{L}^p} X$ and $X_n \xrightarrow{\mathcal{D}} X$, respectively.

Definition 1.2.2 defines four modes of convergence: Convergence in probability, almost sure convergence, convergence in $\mathcal{L}^p$ and convergence in distribution. Convergence in distribution of random variables is also known as convergence in law. Note that convergence in $\mathcal{L}^p$ as given in Definition 1.2.2 is equivalent to convergence in $\|\cdot\|_p$ in the seminormed vector space $\mathcal{L}^p(\Omega, \mathcal{F}, P)$, see Section A.2. In the remainder of this section, we will investigate the connections between these modes of convergence. A first question regards almost sure convergence. The statement that $P(\lim_{n\to\infty} X_n = X) = 1$ is to be understood as that the set $\{\omega \in \Omega \mid X_n(\omega) \text{ converges to } X(\omega)\}$ has probability one. For this to make sense, it is necessary that this set is measurable. The following lemma ensures that this is always the case. For the proof of the lemma, we recall that for any family $(F_i)_{i\in I}$ of subsets of Ω, it holds that

$$\cap_{i\in I}F_i = \{\omega \in \Omega \mid \forall\, i \in I : \omega \in F_i\} \tag{1.1}$$

$$\cup_{i\in I}F_i = \{\omega \in \Omega \mid \exists i \in I : \omega \in F_i\}, \tag{1.2}$$

demonstrating the connection between set intersection and the universal quantifier and the connection between set union and the existential quantifier.

Lemma 1.2.3. *Let (X_n) be a sequence of random variables, and let X be some other variable. The subset F of Ω given by $F = \{\omega \in \Omega \mid X_n(\omega)$ converges to $X(\omega)\}$ is $\mathcal{F}$ measurable. In particular, it holds that*

$$F = \cap_{m=1}^\infty \cup_{n=1}^\infty \cap_{k=n}^\infty (|X_k - X| \le \tfrac{1}{m}). \tag{1.3}$$

Proof. We first prove the equality (1.3), and to this end, we first show that for any sequence (x_n) of real numbers and any real x, it holds that x_n converges to x if and only if

$$\forall\, m \in \mathbb{N}\, \exists\, n \in \mathbb{N}\, \forall\, k \ge n : |x_k - x| \le \tfrac{1}{m}. \tag{1.4}$$

To this end, recall that x_n converges to x if and only if

$$\forall\, \varepsilon > 0\, \exists\, n \in \mathbb{N}\, \forall\, k \ge n : |x_k - x| \le \varepsilon. \tag{1.5}$$

It is immediate that (1.5) implies (1.4). We prove the converse implication. Therefore, assume that (1.4) holds. Let $\varepsilon > 0$ be given. Pick a natural $m \ge 1$ so large that $\frac{1}{m} \le \varepsilon$. Using (1.4), take a natural $n \ge 1$ such that for all $k \ge n$, $|x_k - x| \le \frac{1}{m}$. It then also holds that for $k \ge n$, $|x_k - x| \le \varepsilon$. Therefore, (1.5) holds, and so (1.5) and (1.4) are equivalent.

We proceed to prove (1.3). Using what we already have shown, we obtain

$$\begin{aligned} F &= \{\omega \in \Omega \mid X_n(\omega) \text{ converges to } X(\omega)\} \\ &= \{\omega \in \Omega \mid \forall\, \varepsilon > 0\, \exists\, n \in \mathbb{N}\, \forall\, k \ge n : |X_n(\omega) - X(\omega)| \le \varepsilon\} \\ &= \{\omega \in \Omega \mid \forall\, m \in \mathbb{N}\, \exists\, n \in \mathbb{N}\, \forall\, k \ge n : |X_n(\omega) - X(\omega)| \le \tfrac{1}{m}\}, \end{aligned}$$

and applying (1.1) and (1.2), this yields

$$\begin{aligned} F &= \cap_{m=1}^\infty \{\omega \in \Omega \mid \exists\, n \in \mathbb{N}\, \forall\, k \ge n : |X_n(\omega) - X(\omega)| \le \tfrac{1}{m}\} \\ &= \cap_{m=1}^\infty \cup_{n=1}^\infty \{\omega \in \Omega \mid \forall\, k \ge n : |X_n(\omega) - X(\omega)| \le \tfrac{1}{m}\} \\ &= \cap_{m=1}^\infty \cup_{n=1}^\infty \cap_{k=n}^\infty \{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| \le \tfrac{1}{m}\} \\ &= \cap_{m=1}^\infty \cup_{n=1}^\infty \cap_{k=n}^\infty (|X_k - X| \le \tfrac{1}{m}), \end{aligned}$$

as desired. We have now proved (1.3). Next, as X_k and X both are $\mathcal{F}$ measurable mappings, $|X_k - X|$ is $\mathcal{F}$ measurable as well, so set $(|X_k - X| \leq \frac{1}{m})$ is in $\mathcal{F}$. As a consequence, we obtain that $\cap_{m=1}^\infty \cup_{n=1}^\infty \cap_{k=n}^\infty (|X_k - X| \leq \frac{1}{m})$ is an element of $\mathcal{F}$. We conclude that $F \in \mathcal{F}$, as desired. □

Lemma 1.2.3 ensures that the definition of almost sure convergence given in Definition 1.2.2 is well-formed. A second immediate question regards convergence in probability: Does it matter whether we consider the limit of $P(|X_n - X| \geq \varepsilon)$ or $P(|X_n - X| > \varepsilon)$? The following lemma shows that this is not the case.

Lemma 1.2.4. *Let (X_n) be a sequence of random variables, and let X be some other variable. It holds that X_n converges in probability to X if and only if it holds that for each $\varepsilon > 0$, $\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$.*

Proof. First assume that for each $\varepsilon > 0$, $\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$. We need to show that X_n converges in probability to X, meaning that for each $\varepsilon > 0$, $\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$. To prove this, first fix $\varepsilon > 0$. We then obtain

$$\limsup_{n\to\infty} P(|X_n - X| \geq \varepsilon) \leq \limsup_{n\to\infty} P(|X_n - X| > \tfrac{\varepsilon}{2}) = 0,$$

so $\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$, as desired. Conversely, if it holds for all $\varepsilon > 0$ that $\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$, we find for any $\varepsilon > 0$ that

$$\limsup_{n\to\infty} P(|X_n - X| > \varepsilon) \leq \limsup_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0,$$

which proves the other implication. □

Also, we show that limits for three of the modes of convergence considered are almost surely unique.

Lemma 1.2.5. *Let (X_n) be a sequence of random variables and let X and Y be two other variables. Assume that X_n converges both to X and to Y in probability, almost surely or in $\mathcal{L}^p$ for some $p \geq 1$. Then X and Y are almost surely equal.*

Proof. First assume that $X_n \xrightarrow{P} X$ and $X_n \xrightarrow{P} Y$. Fix $\varepsilon > 0$. Note that if $|X - X_n| \leq \varepsilon/2$ and $|X_n - Y| \leq \varepsilon/2$, we have $|X - Y| \leq \varepsilon$. Therefore, we also find that $|X - Y| > \varepsilon$ implies that either $|X - X_n| > \varepsilon/2$ or $|X_n - Y| > \varepsilon/2$. Hence, we obtain

$$\begin{aligned} P(|X-Y| \geq \varepsilon) &\leq P(|X-X_n| + |X_n - Y| \geq \varepsilon) \\ &\leq P((|X-X_n| \geq \tfrac{\varepsilon}{2}) \cup (|X_n - Y| \geq \tfrac{\varepsilon}{2})) \\ &\leq P(|X-X_n| \geq \tfrac{\varepsilon}{2}) + P(|X_n - Y| \geq \tfrac{\varepsilon}{2}), \end{aligned}$$

so that $P(|X-Y| \geq \varepsilon) \leq \limsup_{n\to\infty} P(|X-X_n| \geq \frac{\varepsilon}{2}) + P(|X_n - Y| \geq \frac{\varepsilon}{2}) = 0$. As $(|X-Y| > 0) = \cup_{n=1}^{\infty}(|X-Y| \geq \frac{1}{n})$ and a union of null sets again is a null sets, we conclude that $(|X-Y| > 0)$ is a null set, such that X and Y are almost surely equal.

In the case where $X_n \xrightarrow{\text{a.s.}} X$ and $X_n \xrightarrow{\text{a.s.}} Y$, the result follows since limits in $\mathbb{R}$ are unique. If $X_n \xrightarrow{\mathcal{L}^p} X$ and $X_n \xrightarrow{\mathcal{L}^p} Y$, we obtain $\|X-Y\|_p \leq \limsup_{n\to\infty} \|X - X_n\|_p + \|X_n - Y\|_p = 0$, so $E|X-Y|^p = 0$, yielding that X and Y are almost surely equal. Here, $\|\cdot\|_p$ denotes the seminorm on $\mathcal{L}^p(\Omega, \mathcal{F}, P)$. □

Having settled these preliminary questions, we next consider the question of whether some of the modes of convergence imply another mode of convergence. Before proving our basic theorem on this, we show a few lemmas of independent interest. In the following lemma, $f(X)$ denotes the random variable defined by $f(X)(\omega) = f(X(\omega))$.

Lemma 1.2.6. *Let (X_n) be a sequence of random variables, and let X be some other variable. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function. If X_n converges almost surely to X, then $f(X_n)$ converges almost surely to $f(X)$. If X_n converges in probability to X, then $f(X_n)$ converges in probability to $f(X)$.*

Proof. We first consider the case of almost sure convergence. Assume that X_n converges almost surely to X. As f is continuous, we find for each ω that if $X_n(\omega)$ converges to $X(\omega)$, $f(X_n(\omega))$ converges to $f(X(\omega))$ as well. Therefore,

$$P(f(X_n) \text{ converges to } f(X)) \geq P(X_n \text{ converges to } X) = 1,$$

proving the result. Next, we turn to the more difficult case of convergence in probability. Assume that X_n converges in probability to X, we need to prove that $f(X_n)$ converges in probability to $f(X)$. Let $\varepsilon > 0$, we thus need to show $\lim_{n\to\infty} P(|f(X_n) - f(X)| > \varepsilon) = 0$. To this end, let $m \geq 1$. As $[-(m+1), m+1]$ is compact, f is uniformly continuous on this set. Choose $\delta > 0$ such that δ parries ε for this uniform continuity of f. We may assume without loss of generality that $\delta \leq 1$. We then have that for x and y in $[-(m+1), m+1]$, $|x-y| \leq \delta$ implies $|f(x) - f(y)| \leq \varepsilon$. Now assume that $|f(x) - f(y)| > \varepsilon$. If $|x - y| \leq \delta$ and $|x| \leq m$, we obtain $x, y \in [-(m+1), m+1]$ and thus a contradiction with $|f(x) - f(y)| > \varepsilon$.

Therefore, when $|f(x) - f(y)| > \varepsilon$, it must either hold that $|x - y| > \delta$ or $|x| > m$. This yields

$$\begin{aligned} P(|f(X_n) - f(X)| > \varepsilon) &\leq P((|X_n - X| > \delta) \cup (|X| > m)) \\ &\leq P(|X_n - X| > \delta) + P(|X| > m). \end{aligned}$$

Note that while δ depends on m, neither δ nor m depends on n. Therefore, as X_n converges in probability to X, the above estimate allows us to conclude

$$\limsup_{n\to\infty} P(|f(X_n) - f(X)| > \varepsilon) \leq \limsup_{n\to\infty} P(|X_n - X| > \delta) + P(|X| > m) = P(|X| > m).$$

As m was arbitrary, we then finally obtain

$$\limsup_{n\to\infty} P(|f(X_n) - f(X)| > \varepsilon) \leq \lim_{m\to\infty} P(|X| > m) = 0,$$

by downwards continuity. This shows that $f(X_n)$ converges in probability to $f(X)$. □

Lemma 1.2.7. *Let X be a random variable, let $p > 0$ and let $\varepsilon > 0$. It then holds that $P(|X| \geq \varepsilon) \leq \varepsilon^{-p} E|X|^p$.*

Proof. We simply note that $P(|X| \geq \varepsilon) = E1_{(|X|\geq\varepsilon)} \leq \varepsilon^{-p} E|X|^p 1_{(|X|\geq\varepsilon)} \leq \varepsilon^{-p} E|X|^p$, which yields the result. □

Theorem 1.2.8. *Let (X_n) be a sequence of random variables, and let X be some other variable. If X_n converges in $\mathcal{L}^p$ to X for some $p \geq 1$, or if X_n converges almost surely to X, then X_n also converges in probability to X. If X_n converges in probability to X, then X_n also converges in distribution to X.*

Proof. We need to prove three implications. First assume that X_n converges in $\mathcal{L}^p$ to X for some $p \geq 1$, we want to show that X_n converges in probability to X. By Lemma 1.2.7, it holds for any $\varepsilon > 0$ that

$$\limsup_{n\to\infty} P(|X_n - X| \geq \varepsilon) \leq \limsup_{n\to\infty} \varepsilon^{-p} E|X_n - X|^p = 0,$$

so $P(|X_n - X| \geq \varepsilon)$ converges as n tends to infinity, and the limit is zero. Therefore, X_n converges in probability to X. Next, assume that X_n converges almost surely to X. Again, we wish to show that X_n converges in probability to X. Fix $\varepsilon > 0$. Using that

$(|X_n - X| \geq \varepsilon) \subseteq \cup_{k=n}^{\infty}(|X_k - X| \geq \varepsilon)$ and that the sequence $(\cup_{k=n}^{\infty}(|X_k - X| \geq \varepsilon))_{n\geq 1}$ is decreasing, we find

$$\begin{aligned}\limsup_{n\to\infty} P(|X_n - X| \geq \varepsilon) &\leq \lim_{n\to\infty} P(\cup_{k=n}^{\infty}|X_k - X| \geq \varepsilon)\\ &= P(\cap_{n=1}^{\infty}\cup_{k=n}^{\infty}|X_k - X| \geq \varepsilon)\\ &\leq P(X_n \text{ does not converge to } X) = 0,\end{aligned}$$

so X_n converges in probability to X, as desired. Finally, we need to show that if X_n converges in probability to X, then X_n also converges in distribution to X. Assume that X_n converges in probability to X, and let $f : \mathbb{R} \to \mathbb{R}$ be bounded and continuous. Let $c \geq 0$ be such that $|f(x)| \leq c$ for all x. Applying the triangle inequality, $|f(X_n) - f(X)| \leq |f(X_n)| + |f(X)| \leq 2c$, and so we obtain for any $\varepsilon > 0$ that

$$\begin{aligned}&|Ef(X_n) - Ef(X)| \leq E|f(X_n) - f(X)|\\ &= E|f(X_n) - f(X)|1_{(|f(X_n)-f(X)|>\varepsilon)} + E|f(X_n) - f(X)|1_{(|f(X_n)-f(X)|\leq\varepsilon)}\\ &\leq 2cP(|f(X_n) - f(X)| > \varepsilon) + \varepsilon. \qquad (1.6)\end{aligned}$$

By Lemma 1.2.6, $f(X_n)$ converges in probability to $f(X)$. Therefore, (1.6) shows that $\limsup_{n\to\infty} |Ef(X_n) - Ef(X)| \leq \varepsilon$. As $\varepsilon > 0$ was arbitrary, this allows us to conclude $\limsup_{n\to\infty} |Ef(X_n) - Ef(X)| = 0$, and as a consequence, $\lim_{n\to\infty} Ef(X_n) = Ef(X)$. This proves the desired convergence in distribution of X_n to X. □

Theorem 1.2.8 shows that among the four modes of convergence defined in Definition 1.2.2, convergence in $\mathcal{L}^p$ and almost sure convergence are the strongest, convergence in probability is weaker than both, and convergence in distribution is weaker still. There is no general simple relationship between convergence in $\mathcal{L}^p$ and almost sure convergence. Note also an essential difference between convergence in distribution and the other three modes of convergence: While both convergence in $\mathcal{L}^p$, almost sure convergence and convergence in probability depend on the multivariate distribution of (X_n, X), convergence in distribution merely depends on the marginal laws of X_n and X. For this reason, the theory for convergence in distribution is somewhat different than the theory for the other three modes of convergence. In the remainder of this chapter and the next, we only consider the other three modes of convergence.

Example 1.2.9. Let $\xi \in \mathbb{R}$, let $\sigma > 0$ and let (X_n) be a sequence of random variables such that for all $n \geq 1$, X_n is normally distributed with mean ξ and variance σ^2. Assume furthermore that $X_1, \ldots, X_n$ are independent for all $n \geq 1$. Put $\hat{\xi}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$. We claim that $\hat{\xi}_n$ converges in $\mathcal{L}^p$ to ξ for all $p \geq 1$.

To prove this, note that by the properties of normal distributions, $\frac{1}{n}\sum_{k=1}^n X_k$ is normally distributed with mean ξ and variance $\frac{1}{n}\sigma^2$. Therefore, $\sqrt{n}\sigma^{-1}(\xi - \frac{1}{n}\sum_{k=1}^n X_k)$ is standard normally distributed. With m_p denoting the p'th absolute moment of the standard normal distribution, we thus obtain

$$E|\xi - \hat{\xi}_n|^p = E\left|\xi - \frac{1}{n}\sum_{k=1}^n X_k\right|^p = \frac{\sigma^p}{n^{p/2}}E\left|\sqrt{n}\sigma^{-1}\left(\xi - \frac{1}{n}\sum_{k=1}^n X_k\right)\right|^p = \frac{\sigma^p m_p}{n^{p/2}},$$

which converges to zero, proving that $\frac{1}{n}\sum_{k=1}^n X_k \xrightarrow{\mathcal{L}^p} \xi$ for all $p \geq 1$. ∘

The following lemma shows that almost sure convergence and convergence in probability enjoy strong stability properties.

Lemma 1.2.10. *Let (X_n) and (Y_n) be sequences of random variables, and let X and Y be two other random variables. If X_n converges in probability to X and Y_n converges in probability to Y, then $X_n + Y_n$ converges in probability to $X + Y$, and X_nY_n converges in probability to XY. Also, if X_n converges almost surely to X and Y_n converges almost surely to Y, then $X_n + Y_n$ converges almost surely to $X + Y$, and X_nY_n converges almost surely to XY.*

Proof. We first show the claims for almost sure convergence. Assume that X_n converges almost surely to X and that Y_n converges almost surely to Y. Note that as addition is continuous, we have that whenever $X_n(\omega)$ converges to $X(\omega)$ and $Y_n(\omega)$ converges to $Y(\omega)$, it also holds that $X_n(\omega) + Y_n(\omega)$ converges to $X(\omega) + Y(\omega)$. Therefore,

$$P(X_n + Y_n \text{ converges to } X + Y) \geq P((X_n \text{ converges to } X) \cap (Y_n \text{ converges to } Y)) = 1,$$

and by a similar argument, we find

$$P(X_nY_n \text{ converges to } XY) \geq P((X_n \text{ converges to } X) \cap (Y_n \text{ converges to } Y)) = 1,$$

since the intersection of two almost sure sets also is an almost sure set. This proves the claims on almost sure convergence. Next, assume that X_n converges in probability to X and that Y_n converges in probability to Y. We first show that $X_n + Y_n$ converges in probability to $X + Y$. Let $\varepsilon > 0$ be given. We then obtain

$$\begin{aligned}\limsup_{n\to\infty} P(|X_n + Y_n - (X + Y)| \geq \varepsilon) &\leq \limsup_{n\to\infty} P((|X_n - X| \geq \tfrac{\varepsilon}{2}) \cup (|Y_n - Y| \geq \tfrac{\varepsilon}{2}))\\ &\leq \limsup_{n\to\infty} P(|X_n - X| \geq \tfrac{\varepsilon}{2}) + P(|Y_n - Y| \geq \tfrac{\varepsilon}{2}) = 0,\end{aligned}$$

proving the claim. Finally, we show that X_nY_n converges in probability to XY. This will follow if we show that $X_nY_n - XY$ converges in probability to zero. To this end, we note the relationship $X_nY_n - XY = (X_n - X)(Y_n - Y) + (X_n - X)Y + (Y_n - Y)X$, so by what we already have shown, it suffices to show that each of these three terms converge in probability to zero. For the first term, recall that for all $x, y \in \mathbb{R}$, $x^2+2xy+y^2 \geq 0$ and $x^2-2xy+y^2 \geq 0$, so that $|xy| \leq \frac{1}{2}x^2 + \frac{1}{2}y^2$. Therefore, we obtain for all $\varepsilon > 0$ that

$$\begin{aligned} P(|(X_n - X)(Y_n - Y)| \geq \varepsilon) &\leq P(\tfrac{1}{2}(X_n - X)^2 + \tfrac{1}{2}(Y_n - Y)^2 \geq \varepsilon) \\ &\leq P(\tfrac{1}{2}(X_n - X)^2 \geq \tfrac{1}{2}\varepsilon) + P(\tfrac{1}{2}(Y_n - Y)^2 \geq \tfrac{1}{2}\varepsilon) \\ &= P(|X_n - X| \geq \sqrt{\varepsilon}) + P(|Y_n - Y| \geq \sqrt{\varepsilon}). \end{aligned}$$

Taking the limes superior, we conclude $\lim_{n\to\infty} P(|(X_n - X)(Y_n - Y)| \geq \varepsilon) = 0$, and thus $(X_n - X)(Y_n - Y)$ converges in probability to zero. Next, we show that $(X_n - X)Y$ converges in probability to zero. Again, let $\varepsilon > 0$. Consider also some $m \geq 1$. We then obtain

$$\begin{aligned} &P(|(X_n - X)Y| \geq \varepsilon) \\ &= P((|(X_n - X)Y| \geq \varepsilon) \cap (|Y| \leq m)) + P((|(X_n - X)Y| \geq \varepsilon) \cap (|Y| > m)) \\ &\leq P(m|X_n - X| \geq \varepsilon) + P(|Y| > m) = P(|X_n - X| \geq \tfrac{1}{m}\varepsilon) + P(|Y| > m). \end{aligned}$$

Therefore, we obtain $\limsup_{n\to\infty} P(|(X_n - X)Y| \geq \varepsilon) \leq P(|Y| > m)$ for all $m \geq 1$, from which we conclude $\limsup_{n\to\infty} P(|(X_n - X)Y| \geq \varepsilon) \leq \lim_{m\to\infty} P(|Y| > m) = 0$, by downwards continuity. This shows that $(X_n - X)Y$ converges in probability to zero. By a similar argument, we also conclude that $(Y_n - Y)X$ converges in probability to zero. Combining our results, we conclude that X_nY_n converges in probability to XY, as desired. □

Lemma 1.2.10 could also have been proven using a multidimensional version of Lemma 1.2.6 and the continuity of addition and multiplication. Our next goal is to prove another connection between two of the modes of convergence, namely that convergence in probability implies almost sure convergence along a subsequence, and use this to show completeness properties of each of our three modes of convergence, in the sense that we wish to argue that Cauchy sequences are convergent for both convergence in $\mathcal{L}^p$, almost sure convergence and convergence in probability.

We begin by showing the Borel-Cantelli lemma, a general result which will be useful in several contexts. Let (F_n) be a sequence of events. We then define

$$\begin{aligned} (F_n \text{ i.o.}) &= \{\omega \in \Omega \mid \omega \in F_n \text{ infinitely often } \} \\ (F_n \text{ evt.}) &= \{\omega \in \Omega \mid \omega \in F_n \text{ eventually } \} \end{aligned}$$

Note that $\omega \in F_n$ for infinitely many n if and only if for each $n \geq 1$, there exists $k \geq n$ such that $\omega \in F_k$. Likewise, it holds that $\omega \in F_n$ eventually if and only if there exists $n \geq 1$

such that for all $k \geq n$, $\omega \in F_k$. Therefore, we also have $(F_n \text{ i.o.}) = \cap_{n=1}^\infty \cup_{k=n}^\infty F_k$ and $(F_n \text{ evt.}) = \cup_{n=1}^\infty \cap_{k=n}^\infty F_k$. This shows in particular that the sets $(F_n \text{ i.o.})$ and $(F_n \text{ evt.})$ are measurable. Also, we obtain the equality $(F_n \text{ i.o.})^c = (F_n^c \text{ evt.})$. It is customary also to write $\limsup_{n\to\infty} F_n$ for $(F_n \text{ i.o.})$ and $\liminf_{n\to\infty} F_n$ for $(F_n \text{ evt.})$, although this is not a notation which we will be using. The main useful result about events occurring infinitely often is the following.

Lemma 1.2.11 (Borel-Cantelli)**.** *Let (F_n) be a sequence of events. If $\sum_{n=1}^\infty P(F_n)$ is finite, then $P(F_n \text{ i.o.}) = 0$.*

Proof. As the sequence of sets $(\cup_{k=n}^\infty F_k)_{n\geq 1}$ is decreasing, we obtain by the downward continuity of probability measures that

$$P(F_n \text{ i.o.}) = P(\cap_{n=1}^\infty \cup_{k=n}^\infty F_k) = \lim_{n\to\infty} P(\cup_{k=n}^\infty F_k) \leq \lim_{n\to\infty} \sum_{k=n}^\infty P(F_k) = 0,$$

with the final equality holding since the tail sum of a convergent series always tends to zero. □

Lemma 1.2.12. *Let (X_n) be a sequence of random variables, and let X be some other variable. Assume that for all $\varepsilon > 0$, $\sum_{n=1}^\infty P(|X_n - X| \geq \varepsilon)$ is finite. Then X_n converges almost surely to X.*

Proof. Recalling (1.3), it suffices to show that

$$P(\cap_{m=1}^\infty \cup_{n=1}^\infty \cap_{k=n}^\infty (|X_k - X| \leq \tfrac{1}{m})) = 1.$$

Fix $\varepsilon > 0$. By Lemma 1.2.11, we find that the set $(|X_n - X| \geq \varepsilon \text{ i.o.})$ has probability zero. As $(|X_n - X| < \varepsilon \text{ evt.})^c = (|X_n - X| \geq \varepsilon \text{ i.o.})$, we obtain $P(|X_n - X| < \varepsilon \text{ evt.}) = 1$. As $\varepsilon > 0$ was arbitrary, we in particular obtain $P(|X_n - X| \leq \frac{1}{m} \text{ evt.}) = 1$ for all $m \geq 1$. As the intersection of a countable family of almost sure events again is an almost sure event, this yields

$$P(\cap_{m=1}^\infty \cup_{n=1}^\infty \cap_{k=n}^\infty (|X_k - X| \leq \tfrac{1}{m})) = P(\cap_{m=1}^\infty (|X_k - X| \leq \tfrac{1}{m} \text{ evt.})) = 1,$$

as desired. □

Lemma 1.2.13. *Let (X_n) be a sequence of random variables, and let X be some other variable. Assume that X_n converges in probability to X. There is a subsequence (X_{n_k}) converging almost surely to X.*

Proof. Let $(\varepsilon_k)_{k\geq 1}$ be a sequence of positive numbers decreasing to zero. For each k, it holds that $\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon_k) = 0$. In particular, for any $k, n^* \geq 1$, we may always pick $n > n^*$ such that $P(|X_n - X| \geq \varepsilon_k) \leq 2^{-k}$. Therefore, we may recursively define a strictly increasing sequence of indicies $(n_k)_{k\geq 1}$ such that for each k, $P(|X_{n_k} - X| \geq \varepsilon_k) \leq 2^{-k}$. We claim that the sequence $(X_{n_k})_{k\geq 1}$ satisfies the criterion of Lemma 1.2.12. To see this, let $\varepsilon > 0$. As $(\varepsilon_k)_{k\geq 1}$ decreases to zero, there is m such that for $k \geq m$, $\varepsilon_k \leq \varepsilon$. We then obtain

$$\sum_{k=m}^{\infty} P(|X_{n_k} - X| \geq \varepsilon) \leq \sum_{k=m}^{\infty} P(|X_{n_k} - X| \geq \varepsilon_k) \leq \sum_{k=m}^{\infty} 2^{-k},$$

which is finite. Hence, $\sum_{k=1}^{\infty} P(|X_{n_k} - X| \geq \varepsilon)$ is also finite, and Lemma 1.2.12 then shows that X_{n_k} converges almost surely to X. □

We are now almost ready to introduce the concept of being Cauchy with respect to each of our three modes of convergence and show that being Cauchy implies convergence.

Lemma 1.2.14. *Let (X_n) be a sequence of random variables. It then holds that*

$$(X_n \textit{ is Cauchy}) = \cap_{m=1}^{\infty} \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} (|X_n - X_k| \leq \tfrac{1}{m}), \tag{1.7}$$

and in particular, $(X_n$ is Cauchy$)$ is measurable.

Proof. Recall that a sequence (x_n) in $\mathbb{R}$ is Cauchy if and only if

$$\forall\, \varepsilon > 0\; \exists\, n \in \mathbb{N}\; \forall\, k, i \geq n : |x_k - x_i| \leq \varepsilon \tag{1.8}$$

We will first argue that this is equivalent to

$$\forall\, m \in \mathbb{N}\; \exists\, n \in \mathbb{N}\; \forall\, k \geq n : |x_k - x_n| \leq \tfrac{1}{m}. \tag{1.9}$$

To this end, first assume that (1.8) holds. Let $m \in \mathbb{N}$ be given and choose $\varepsilon > 0$ so small that $\varepsilon \leq \frac{1}{m}$. Using (1.8), take $n \in \mathbb{N}$ so that $|x_k - x_i| \leq \varepsilon$ whenever $k, i \geq n$. Then it holds in particular that $|x_k - x_n| \leq \frac{1}{m}$. Thus, (1.9) holds. To prove the converse implication, assume that (1.9) holds. Let $\varepsilon > 0$ be given and take $m \in \mathbb{N}$ so large that $\frac{1}{m} \leq \varepsilon/2$. Using (1.9), take $n \in \mathbb{N}$ so that for all $k \geq n$, $|x_k - x_n| \leq \frac{1}{m}$. We then obtain that for all $k, i \geq n$, it holds that $|x_k - x_i| \leq |x_k - x_n| + |x_i - x_n| \leq \frac{2}{m} \leq \varepsilon$. We conclude that (1.8) holds. We have now shown that (1.8) and (1.9) are equivalent.

Using this result, we obtain

$$\begin{aligned}(X_n \text{ is Cauchy}) &= \{\omega \in \Omega \mid\; \forall\, \varepsilon > 0\; \exists\, n \in \mathbb{N}\; \forall\, k, i \geq n : |X_k(\omega) - X_n(\omega)| \leq \varepsilon\} \\ &= \{\omega \in \Omega \mid\; \forall\, m \in \mathbb{N}\; \exists\, n \in \mathbb{N}\; \forall\, k \geq n : |X_k(\omega) - X_n(\omega)| \leq \tfrac{1}{m}\} \\ &= \cap_{m=1}^{\infty} \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} (|X_n - X_k| \leq \tfrac{1}{m}),\end{aligned}$$

as desired. As a consequence, the set (X_n is Cauchy) is $\mathcal{F}$ measurable. □

We are now ready to define what it means to be Cauchy with respect to each of our modes of convergence. In the definition, we use the convention that a double sequence $(x_{nm})_{n,m\geq 1}$ converges to x as n and m tend to infinity if it holds that for all $\varepsilon > 0$, there is $k \geq 1$ such that $|x_{nm} - x| \leq \varepsilon$ whenever $n, m \geq k$. In particular, a sequence $(x_n)_{n\geq 1}$ is Cauchy if and only if $|x_n - x_m|$ tends to zero as n and m tend to infinity.

Definition 1.2.15. *Let (X_n) be a sequence of random variables. We say that X_n is Cauchy in probability if it holds for any $\varepsilon > 0$ that $P(|X_n - X_m| \geq \varepsilon)$ tends to zero as m and n tend to infinity. We say that X_n is almost surely Cauchy if $P((X_n)$ is Cauchy$) = 1$. Finally, we say that X_n is Cauchy in $\mathcal{L}^p$ for some $p \geq 1$ if $E|X_n - X_m|^p$ tends to zero as m and n tend to infinity.*

Note that Lemma 1.2.14 ensures that the definition of being almost surely Cauchy is well-formed, since (X_n is Cauchy) is measurable.

Theorem 1.2.16. *Let (X_n) be a sequence of random variables. If X_n is Cauchy in probability, there exists a random variable X such that X_n converges in probability to X. If X_n is almost surely Cauchy, there exists a random variable X such that X_n converges almost surely to X. If (X_n) is a sequence in $\mathcal{L}^p$ which is Cauchy in $\mathcal{L}^p$, there exists a random variable X in $\mathcal{L}^p$ such that X_n converges to X in $\mathcal{L}^p$.*

Proof. The result on sequences which are Cauchy in $\mathcal{L}^p$ is immediate from Fischer's completeness theorem, so we merely need to show the results for being Cauchy in probability and being almost surely Cauchy.

Consider the case where X_n is almost surely Cauchy. As $\mathbb{R}$ equipped with the Euclidean metric is complete, (X_n is convergent) = (X_n is Cauchy), so in particular, by Lemma 1.2.14, the former is a measurable almost sure set. Define X by letting $X = \lim_{n\to\infty} X_n$ when the limit exists and zero otherwise. Then X is measurable, and we have

$$P(\lim_{n\to\infty} X_n = X) = P(X_n \text{ is Cauchy}) = 1,$$

so X_n converges almost surely to X, proving the result for being almost surely Cauchy.

Finally, assume that X_n is Cauchy in probability. For each k, $P(|X_n - X_m| \geq 2^{-k})$ tends to zero as m and n tend to infinity. In particular, we find that for each k, there is n^*

such that for $n, m \geq n^*$, it holds that $P(|X_n - X_m| \geq 2^{-k}) \leq 2^{-k}$. Therefore, we may pick a sequence of strictly increasing indicies (n_k) such that $P(|X_n - X_m| \geq 2^{-k}) \leq 2^{-k}$ for $n, m \geq n_k$. We then obtain in particular that $P(|X_{n_{k+1}} - X_{n_k}| \geq 2^{-k}) \leq 2^{-k}$ for all $k \geq 1$. From this, we find that $\sum_{k=1}^\infty P(|X_{n_{k+1}} - X_{n_k}| \geq 2^{-k})$ is finite, so by Lemma 1.2.11, $P(|X_{n_{k+1}} - X_{n_k}| \geq 2^{-k} \text{ i.o.}) = 0$, leading to $P(|X_{n_{k+1}} - X_{n_k}| < 2^{-k} \text{ evt.}) = 1$. In particular, it holds almost surely that $\sum_{k=1}^\infty |X_{n_{k+1}} - X_{n_k}|$ is finite. For any $k > i \geq 1$, we have

$$|X_{n_k} - X_{n_i}| \leq \sum_{j=i}^{k-1} |X_{n_{j+1}} - X_{n_j}| \leq \sum_{j=i}^{\infty} |X_{n_{j+1}} - X_{n_j}|.$$

Now, as the tail sums of convergent sums tend to zero, the above shows that on the almost sure set where $\sum_{k=1}^\infty |X_{n_{k+1}} - X_{n_k}|$ is finite, $(X_{n_k})_{k\geq 1}$ is Cauchy. In particular, $(X_{n_k})_{k\geq 1}$ is almost surely Cauchy, so by what was already shown, there exists a variable X such that X_{n_k} converges almost surely to X. In order to complete the proof, we will argue that (X_n) converges in probability to X. To this end, fix $\varepsilon > 0$. Let $\delta > 0$. As (X_n) is Cauchy in probability, there is n^* such that for $m, n \geq n^*$, $P(|X_n - X_m| \geq \frac{\varepsilon}{2}) \leq \delta$. And as X_{n_k} converges almost surely to X, X_{n_k} also converges in probability to X by Theorem 1.2.8. Therefore, for k large enough, $P(|X_{n_k} - X| \geq \frac{\varepsilon}{2}) \leq \delta$. Let k be so large that this holds and simultaneously so large that $n_k \geq n^*$. We then obtain for $n \geq n^*$ that

$$\begin{aligned} P(|X_n - X| \geq \varepsilon) &\leq P(|X_n - X_{n_k}| + |X_{n_k} - X| \geq \varepsilon) \\ &\leq P(|X_n - X_{n_k}| \geq \tfrac{\varepsilon}{2}) + P(|X_{n_k} - X| \geq \tfrac{\varepsilon}{2}) \leq 2\delta. \end{aligned}$$

Thus, for n large enough, $P(|X_n - X| \geq \varepsilon) \leq 2\delta$. As δ was arbitrary, we conclude that $\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0$, showing that X_n converges in probability to X. This concludes the proof. □

This concludes our preliminary investigation of convergence of sequences of random variables.

1.3 Independence and Kolmogorov's zero-one law

In this section, we generalize the classical notion of independence of random variables and events to a notion of independence of σ-algebras. This general notion of independence encompasses all types of independence which will be relevant to us.

Definition 1.3.1. *Let I be some set and let $(\mathcal{F}_i)_{i\in I}$ be a family of σ-algebras. We say that the family of σ-algebras is independent if it holds for any finite sequence of distinct indicies*

$i_1, \ldots, i_n \in I$ and any $F_1 \in \mathcal{F}_{i_1}, \ldots, F_n \in \mathcal{F}_{i_n}$ that

$$P(\cap_{k=1}^n F_k) = \prod_{k=1}^n P(F_k). \tag{1.10}$$

The abstract definition in Definition 1.3.1 will allow us considerable convenience as regards matters of independence. The following lemma shows that when we wish to prove independence, it suffices to prove the equality (1.10) for generating families which are stable under finite intersections.

Lemma 1.3.2. *Let I be some set and let $(\mathcal{F}_i)_{i \in I}$ be a family of σ-algebras. Assume that for each i, $\mathcal{F}_i = \sigma(\mathbb{H}_i)$, where $\mathbb{H}_i$ is a set family which is stable under finite intersections. If it holds for any finite sequence of distinct indicies $i_1, \ldots, i_n \in I$ that $P(\cap_{k=1}^n F_k) = \prod_{k=1}^n P(F_k)$, where $F_1 \in \mathbb{H}_{i_1}, \ldots, F_n \in \mathbb{H}_{i_n}$, then $(\mathcal{F}_i)_{i \in I}$ is independent.*

Proof. We apply Dynkin's lemma and an induction proof. We wish to show that for each n, it holds for all sequences of n distinct indicies $i_1, \ldots, i_n \in I$ and all finite sequences of sets $F_1 \in \mathcal{F}_{i_1}, \ldots, F_n \in \mathcal{F}_{i_n}$ that $P(\cap_{k=1}^n F_k) = \prod_{k=1}^n P(F_k)$. The induction start is trivial, so it suffices to show the induction step. Assume that the result holds for n, we wish to prove it for $n+1$. Fix a finite sequence of $n+1$ distinct indicies $i_1, \ldots, i_{n+1} \in I$. We wish to show that

$$P(\cap_{k=1}^{n+1} F_k) = \prod_{k=1}^{n+1} P(F_k) \tag{1.11}$$

for $F_1 \in \mathcal{F}_{i_1}, \ldots, F_{n+1} \in \mathcal{F}_{i_{n+1}}$. To this end, let $k \leq n+1$, and let $F_j \in \mathcal{F}_{i_j}$ for $j \neq k$. Define

$$\mathbb{D} = \left\{ F_k \in \mathcal{F}_{i_k} \,\middle|\, P(\cap_{j=1}^{n+1} F_j) = \prod_{j=1}^{n+1} P(F_j) \right\}. \tag{1.12}$$

We claim that $\mathbb{D}$ is a Dynkin class. To see this, we need to prove that $\Omega \in \mathbb{D}$, that $B \setminus A \in \mathbb{D}$ whenever $A \subseteq B$ and $A, B \in \mathbb{D}$, and that whenever (A_n) is an increasing sequence in $\mathbb{D}$, $\cup_{n=1}^\infty A_n \in \mathbb{D}$ as well. By our induction assumption, $\Omega \in \mathbb{D}$. Let $A, B \in \mathbb{D}$ with $A \subseteq B$. We then obtain

$$\begin{aligned}
P((B \setminus A) \cap \cap_{j \neq k} F_j) &= P(B \cap A^c \cap \cap_{j \neq k} F_j) = P((B \cap \cap_{j \neq k} F_j) \cap A^c) \\
&= P((B \cap \cap_{j \neq k} F_j) \cap (A \cap \cap_{j \neq k} F_j)^c) \\
&= P(B \cap \cap_{j \neq k} F_j) - P(A \cap \cap_{j \neq k} F_j) \\
&= P(B) \prod_{j \neq k} P(F_j) - P(A) \prod_{j \neq k} P(F_j) = P(B \setminus A) \prod_{j \neq k} P(F_j),
\end{aligned}$$

so that $B \setminus A \in \mathbb{D}$. Finally, let (A_n) be an increasing sequence of sets in $\mathbb{D}$. We then obtain

$$P((\cup_{n=1}^\infty A_n) \cap \cap_{j\neq k} F_j) = P(\cup_{n=1}^\infty A_n \cap \cap_{j\neq k} F_j) = \lim_{n\to\infty} P(A_n \cap \cap_{j\neq k} F_j)$$
$$= \lim_{n\to\infty} P(A_n) \prod_{j\neq k} P(F_j) = P(\cup_{n=1}^\infty A_n) \prod_{j\neq k} P(F_j),$$

so $\cup_{n=1}^\infty A_n \in \mathbb{D}$. This shows that $\mathbb{D}$ is a Dynkin class.

We are now ready to argue that (1.11) holds. Note that by our assumption, we know that (1.11) holds for $F_1 \in \mathbb{H}_{i_1}, \ldots, F_{n+1} \in \mathbb{H}_{i_{n+1}}$. Consider $F_2 \in \mathbb{H}_{i_2}, \ldots, F_{n+1} \in \mathbb{H}_{i_{n+1}}$. The family $\mathbb{D}$ as defined in (1.12) then contains $\mathbb{H}_{i_1}$, and so Dynkin's lemma yields $\mathcal{F}_{i_1} = \sigma(\mathbb{H}_{i_1}) \subseteq \mathbb{D}$. This shows that (1.11) holds when $F_1 \in \mathcal{F}_{i_1}$ and $F_2 \in \mathbb{H}_{i_2}, \ldots, F_{n+1} \in \mathbb{H}_{i_{n+1}}$. Next, let $F_1 \in \mathcal{F}_{i_1}$ and consider a finite sequence of sets $F_3 \in \mathbb{H}_{i_3}, \ldots, F_{n+1} \in \mathbb{H}_{i_{n+1}}$. Then $\mathbb{D}$ as defined in (1.12) contains $\mathbb{H}_{i_2}$, and therefore by Dynkin's lemma contains $\sigma(\mathbb{H}_{i_2}) = \mathcal{F}_{i_2}$, proving that (1.11) holds when $F_1 \in \mathcal{F}_{i_1}$, $F_2 \in \mathcal{F}_{i_2}$ and $F_3 \in \mathbb{H}_{i_3}, \ldots, F_{n+1} \in \mathbb{H}_{i_{n+1}}$. By a finite induction argument, we conclude that (1.11) in fact holds when $F_1 \in \mathcal{F}_{i_1}, \ldots, F_{n+1} \in \mathcal{F}_{i_{n+1}}$, as desired. This proves the induction step and thus concludes the proof. □

The following definition shows how we may define independence between families of variables and families of events from Definition 1.3.1.

Definition 1.3.3. *Let I be some set and let $(X_i)_{i\in I}$ be a family of random variables. We say that the family is independent when the family of σ-algebras $(\sigma(X_i))_{i\in I}$ is independent. Also, if $(F_i)_{i\in I}$ is a family of events, we say that the family is independent when the family of σ-algebras $(\sigma(1_{F_i}))_{i\in I}$ is independent.*

Next, we show that Definition 1.3.3 agrees with our usual definitions of independence.

Lemma 1.3.4. *Let I be some set and let $(X_i)_{i\in I}$ be a family of random variables. The family is independent if and only if it holds for any finite sequence of distinct indicies $i_1, \ldots, i_n \in I$ and any $A_1, \ldots, A_n \in \mathcal{B}$ that $P(\cap_{k=1}^n (X_{i_k} \in A_k)) = \prod_{k=1}^n P(X_{i_k} \in A_k)$.*

Proof. From Definition 1.3.3, we have that $(X_i)_{i\in I}$ is independent if and only if $(\sigma(X_i))_{i\in I}$ is independent, which by Definition 1.3.1 is the case if and only if for any finite sequence of distinct indicies $i_1, \ldots, i_n \in I$ and any $F_1 \in \sigma(X_{i_1}), \ldots, F_n \in \sigma(X_{i_n})$ it holds that $P(\cap_{k=1}^n F_k) = \prod_{k=1}^n P(F_k)$. However, we have $\sigma(X_i) = \{(X_i \in A) \mid A \in \mathcal{B}\}$ for all $i \in I$, so the condition is equivalent to requiring that $P(\cap_{k=1}^n (X_{i_k} \in A_k)) = \prod_{k=1}^n P(X_{i_k} \in A_k)$ for any finite sequence of distinct indicies $i_1, \ldots, i_n \in I$ and any $A_1, \ldots, A_n \in \mathcal{B}$. This proves the claim. □

Lemma 1.3.5. *Let I be some set and let $(F_i)_{i\in I}$ be a family of events. The family is independent if and only if it holds for any finite sequence of distinct indicies $i_1,\ldots,i_n \in I$ that $P(\cap_{k=1}^n F_{i_k}) = \prod_{k=1}^n P(F_{i_k})$.*

Proof. From Definition 1.3.3, $(F_i)_{i\in I}$ is independent if and only if $(\sigma(1_{F_i}))_{i\in I}$ is independent. Note that for all $i \in I$, $\sigma(1_{F_i}) = \{\Omega, \emptyset, F_i, F_i^c\}$, so $\sigma(1_{F_i})$ is generated by $\{F_i\}$. Therefore, Lemma 1.3.2 yields the conclusion. □

We will also have need of the following properties of independence.

Lemma 1.3.6. *Let I be some set and let $(\mathcal{F}_i)_{i\in I}$ be a family of σ-algebras. Let $(\mathcal{G}_i)_{i\in I}$ be another family of σ-algebras, and assume that $\mathcal{G}_i \subseteq \mathcal{F}_i$ for all $i \in I$. If $(\mathcal{F}_i)_{i\in I}$ is independent, so is $(\mathcal{G}_i)_{i\in I}$.*

Proof. This follows immediately from Definition 1.3.1. □

Lemma 1.3.7. *Let I be some set and let $(X_i)_{i\in I}$ be a family of independent variables. For each i, let $\psi_i : \mathbb{R} \to \mathbb{R}$ be some measurable mapping. Then $(\psi_i(X_i))_{i\in I}$ is also independent.*

Proof. As $\sigma(\psi_i(X_i)) \subseteq \sigma(X_i)$, this follows from Lemma 1.3.6. □

Lemma 1.3.8. *Let I be some set and let $(\mathcal{F}_i)_{i\in I}$ be an independent family of σ-algebras. Let $J, J' \subseteq I$ and assume that J and J' are disjoint. Then, the σ-algebras $\sigma((\mathcal{F}_i)_{i\in J})$ and $\sigma((\mathcal{F}_i)_{i\in J'})$ are independent.*

Proof. Let $\mathcal{G} = \sigma((\mathcal{F}_i)_{i\in J})$ and $\mathcal{G}' = \sigma((\mathcal{F}_i)_{i\in J'})$. We define

$$\mathbb{H} = \{\cap_{k=1}^n F_k \mid n \geq 1, i_1,\ldots,i_n \in J \text{ and } F_1 \in \mathcal{F}_{i_1},\ldots,F_n \in \mathcal{F}_{i_n}\}$$
$$\mathbb{H}' = \{\cap_{k=1}^{n'} G_k \mid n' \geq 1, i'_1,\ldots,i'_n \in J' \text{ and } G_1 \in \mathcal{F}_{i'_1},\ldots,G_n \in \mathcal{F}_{i'_n}\}.$$

Then $\mathbb{H}$ and $\mathbb{H}'$ are generating families for $\mathcal{G}$ and $\mathcal{G}'$, respectively, stable under finite intersections. Now let $F \in \mathbb{H}$ and $G \in \mathbb{H}'$. Then, there exists $n, n' \geq 1$, $i_1,\ldots,i_n \in J$ and $i'_1,\ldots,i'_n \in J'$ and $F_1 \in \mathcal{F}_{i_1},\ldots,F_n \in \mathcal{F}_{i_n}$ and $G_1 \in \mathcal{F}_{i'_1},\ldots,G_n \in \mathcal{F}_{i'_{n'}}$ such that $F = \cap_{k=1}^n F_k$ and $G = \cap_{k=1}^{n'} G_k$. Since J and J' are disjoint, the sequence $i_1,\ldots,i_n,i'_1,\ldots,i'_{n'}$ consists of distinct indicies. As $(\mathcal{F}_i)_{i\in I}$ is independent, we then obtain

$$P(F \cap G) = P((\cap_{k=1}^n F_k) \cap (\cap_{k=1}^{n'} G_k)) = \left(\prod_{k=1}^n P(F_k)\right)\left(\prod_{k=1}^{n'} P(G_k)\right) = P(F)P(G).$$

Therefore, Lemma 1.3.2 shows that $\mathcal{G}$ and $\mathcal{G}'$ are independent, as desired. □

Before ending the section, we show some useful results where independence is involved.

Definition 1.3.9. *Let (X_n) be a sequence of random variables. The tail σ-algebra of (X_n) is defined as the σ-algebra $\cap_{n=1}^{\infty}\sigma(X_n, X_{n+1}, \ldots)$.*

Colloquially speaking, the tail σ-algebra of (X_n) consists of events which only depend on the tail properties of (X_n). For example, as we will see shortly, the set where (X_n) is convergent is an element in the tail σ-algebra.

Theorem 1.3.10 (Kolmogorov's zero-one law)**.** *Let (X_n) be a sequence of independent variables. Let $\mathcal{J}$ be the tail σ-algebra of (X_n). For each $F \in \mathcal{J}$, it holds that either $P(F) = 0$ or $P(F) = 1$.*

Proof. Let $F \in \mathcal{J}$ and define $\mathbb{D} = \{G \in \mathcal{F} \mid P(G \cap F) = P(G)P(F)\}$, the family of sets in $\mathcal{F}$ independent of F. We claim that $\mathbb{D}$ contains $\sigma(X_1, X_2, \ldots)$. To prove this, we use Dynkin's Lemma. We first show that $\mathbb{D}$ is a Dynkin class. Clearly, $\Omega \in \mathbb{D}$. If $A, B \in \mathbb{D}$ with $A \subseteq B$, we obtain

$$\begin{aligned} P((B \setminus A) \cap F) &= P(B \cap A^c \cap F) = P((B \cap F) \cap (A \cap F)^c) \\ &= P(B \cap F) - P(A \cap F) = P(B)P(F) - P(A)P(F) = P(B \setminus A)P(F), \end{aligned}$$

so $B \setminus A \in \mathbb{D}$ as well. And if (B_n) is an increasing sequence in $\mathbb{D}$, we obtain

$$\begin{aligned} P((\cup_{n=1}^{\infty} B_n) \cap F) &= P(\cup_{n=1}^{\infty} B_n \cap F) = \lim_{n\to\infty} P(B_n \cap F) \\ &= \lim_{n\to\infty} P(B_n)P(F) = P(\cup_{n=1}^{\infty} B_n)P(F), \end{aligned}$$

proving that $\cup_{n=1}^{\infty} B_n \in \mathbb{D}$. We have now shown that $\mathbb{D}$ is a Dynkin class. Now fix $n \geq 2$. As $F \in \mathcal{J}$, it holds that $F \in \sigma(X_n, X_{n+1}, \ldots)$. Since the sequence (X_n) is independent, Lemma 1.3.8 shows that $\sigma(X_n, X_{n+1}, \ldots)$ is independent of $\sigma(X_1, \ldots, X_{n-1})$. Therefore, $\sigma(X_1, \ldots, X_{n-1}) \in \mathbb{D}$ for all $n \geq 2$. As the family $\cup_{n=1}^{\infty}\sigma(X_1, \ldots, X_n)$ is a generating family for $\sigma(X_1, X_2, \ldots)$ which is stable under finite intersections, Dynkin's lemma allows us to conclude $\sigma(X_1, X_2, \ldots) \subseteq \mathbb{D}$. From this, we obtain $\mathcal{J} \subseteq \mathbb{D}$, so $F \in \mathbb{D}$. Thus, for any $F \in \mathcal{J}$, it holds that $P(F) = P(F \cap F) = P(F)^2$, yielding that $P(F) = 0$ or $P(F) = 1$. □

Example 1.3.11. Let (X_n) be a sequence of independent variables. Recalling Lemma 1.2.14, we have for any $k \geq 1$ that

$$\begin{aligned}((X_n)_{n\geq 1} \text{ is convergent}) &= ((X_n)_{n\geq 1} \text{ is Cauchy}) = ((X_n)_{n\geq k} \text{ is Cauchy}) \\ &= \cap_{m=1}^\infty \cup_{n=1}^\infty \cap_{i=n}^\infty (|X_{k+n-1} - X_{k+i-1}| \leq \tfrac{1}{m}),\end{aligned}$$

which is in $\sigma(X_k, X_{k+1}, \ldots)$. As k was arbitrary, we find that $((X_n)_{n\geq 1}$ is convergent) is in the tail σ-algebra of (X_n). Thus, Theorem 1.3.10 allows us to conclude that the probability of (X_n) being convergent is either zero or one. ◦

Combining Theorem 1.3.10 and Lemma 1.2.11, we obtain the following useful result.

Lemma 1.3.12 (Second Borel-Cantelli)**.** *Let (F_n) be a sequence of independent events. Then $P(F_n$ i.o.$)$ is either zero or one, and the probability is zero if and only if $\sum_{n=1}^\infty P(F_n)$ is finite.*

Proof. Let $\mathcal{J}$ be the tail-σ-algebra of the sequence (1_{F_n}) of variables, Theorem 1.3.10 then shows that $\mathcal{J}$ only contains sets of probability zero or one. Note that for any $m \geq 1$, we have $(F_n \text{ i.o.}) = \cap_{n=1}^\infty \cup_{k=n}^\infty F_k = \cap_{n=m}^\infty \cup_{k=n}^\infty F_k$, so $(F_n$ i.o.) is in $\mathcal{J}$. Hence, Theorem 1.3.10 shows that $P(F_n$ i.o.) is either zero or one.

As regards the criterion for the probability to be zero, note that from Lemma 1.2.11, we know that if $\sum_{n=1}^\infty P(F_n)$ is finite, then $P(F_n \text{ i.o.}) = 0$. We need to show the converse, namely that if $P(F_n \text{ i.o.}) = 0$, then $\sum_{n=1}^\infty P(F_n)$ is finite. This is equivalent to showing that if $\sum_{n=1}^\infty P(F_n)$ is infinite, then $P(F_n \text{ i.o.}) \neq 0$. And to prove this, it suffices to show that if $\sum_{n=1}^\infty P(F_n)$ is infinite, then $P(F_n \text{ i.o.}) = 1$.

Assume that $\sum_{n=1}^\infty P(F_n)$ is infinite. As it holds that $(F_n \text{ i.o.})^c = (F_n^c \text{ evt.})$, it suffices to show $P(F_n^c \text{ evt.}) = 0$. To do so, we note that since the sequence (F_n) is independent, Lemma 1.3.7 shows that the sequence (F_n^c) is independent as well. Therefore,

$$\begin{aligned}P(F_n^c \text{ evt.}) &= P(\cup_{n=1}^\infty \cap_{k=n}^\infty F_k^c) = \lim_{n\to\infty} P(\cap_{k=n}^\infty F_k^c) \\ &= \lim_{n\to\infty}\lim_{i\to\infty} P(\cap_{k=n}^i F_k^c) = \lim_{n\to\infty}\lim_{i\to\infty} \prod_{k=n}^i P(F_k^c) = \lim_{n\to\infty} \prod_{k=n}^\infty P(F_k^c),\end{aligned}$$

since the sequence $(\cap_{k=n}^i F_k^c)_{i\geq 1}$ is decreasing. Next, note that for $x \geq 0$, we have

$$-x = \int_0^x (-1)\,\mathrm{d}y \leq \int_0^x (-\exp(-y))\,\mathrm{d}y = \int_0^x \frac{\mathrm{d}}{\mathrm{d}y}\exp(-y)\,\mathrm{d}y = \exp(-x) - 1,$$

which implies $1 - x \leq \exp(-x)$. This allows us to conclude

$$\lim_{n\to\infty} \prod_{k=n}^{\infty} P(F_k^c) = \lim_{n\to\infty} \prod_{k=n}^{\infty} (1 - P(F_k)) \leq \lim_{n\to\infty} \prod_{k=n}^{\infty} \exp(-P(F_k))$$
$$= \lim_{n\to\infty} \exp\left(-\sum_{k=n}^{\infty} P(F_k)\right) = 0,$$

finally yielding $P(F_n^c \text{ evt.}) = 0$ and so $P(F_n \text{ i.o.}) = 1$, as desired. □

1.4 Convergence of sums of independent variables

In this section, we consider a sequence of independent variables (X_n) and investigate when the sum $\sum_{k=1}^n X_k$ converges as n tends to infinity. During the course of this section, we will encounter sequences (x_n) such that $\sum_{k=1}^n x_k$ converges, while $\sum_{k=1}^n |x_k|$ may not converge, that is, series which are convergent but not absolutely convergent. In such cases, $\sum_{k=1}^\infty x_k$ is not always well-defined. However, for notational convenience, we will apply the following convention: For a sequence (x_n), we say that $\sum_{k=1}^\infty x_k$ converges when $\lim_{n\to\infty} \sum_{k=1}^n x_k$ exists, and say that $\sum_{k=1}^\infty x_k$ diverges when $\lim_{n\to\infty} \sum_{k=1}^n x_k$ does not exist, and in the latter case, $\sum_{k=1}^\infty x_k$ is undefined. With these conventions, we can say that we in this section seek to understand when $\sum_{n=1}^\infty X_n$ converges for a sequence (X_n) of independent variables.

Our first result is an example of a maximal inequality, that is, an inequality which yields bounds on the distribution of a maximum of random variables. We will use this result to prove a sufficient criteria for a sum of variables to converge almost surely and in $\mathcal{L}^2$. Note that in the following, just as we write EX for the expectation of a random variable X, we write VX for the variance of X.

Theorem 1.4.1 (Kolmogorov's maximal inequality)**.** *Let $(X_k)_{1\leq k\leq n}$ be a finite sequence of independent random variables with mean zero and finite variance. It then holds that*

$$P\left(\max_{1\leq k\leq n}\left|\sum_{i=1}^k X_i\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon^2} V \sum_{k=1}^n X_k.$$

Proof. Define $S_k = \sum_{i=1}^k X_i$, we may then state the desired inequality as

$$P(\max_{1\leq k\leq n} |S_k| \geq \varepsilon) \leq \varepsilon^{-2} V S_n. \tag{1.13}$$

Let $T = \min\{1 \le k \le n \mid |S_k| \ge \varepsilon\}$, with the convention that the minimum of the empty set is ∞. Colloquially speaking, T is the first time where the sequence $(S_k)_{1\le k\le n}$ takes an absolute value equal to or greater than ε. Note that T takes its values in $\{1,\ldots,n\}\cup\{\infty\}$. And for each $k \le n$, it holds that $(T \le k) = \cup_{i=1}^k (|S_i| \ge \varepsilon)$, so in particular T is measurable. Now, $(\max_{1\le k\le n}|S_k| \ge \varepsilon) = \cup_{k=1}^n(|S_k|\ge\varepsilon) = (T\le n)$. Also, whenever T is finite, it holds that $|S_T| \ge \varepsilon$, so that $1 \le \varepsilon^{-2}S_T^2$. Therefore, we obtain

$$\begin{aligned} P(\max_{1\le k\le n}|S_k|\ge\varepsilon) &= P(T\le n) = E1_{(T\le n)} \le \varepsilon^{-2}ES_T^2 1_{(T\le n)} \\ &= \varepsilon^{-2}ES_{T\wedge n}^2 1_{(T\le n)} \le \varepsilon^{-2}ES_{T\wedge n}^2 = \varepsilon^{-2}E\left(\sum_{k=1}^n X_k 1_{(T\ge k)}\right)^2. \end{aligned} \tag{1.14}$$

Expanding the square, we obtain

$$\begin{aligned} E\left(\sum_{k=1}^n X_k 1_{(T\ge k)}\right)^2 &= \sum_{k=1}^n EX_k^2 1_{(T\ge k)} + 2\sum_{k=1}^{n-1}\sum_{i=k+1}^n EX_kX_i 1_{(T\ge k)}1_{(T\ge i)} \\ &\le \sum_{k=1}^n EX_k^2 + 2\sum_{k=1}^{n-1}\sum_{i=k+1}^n EX_kX_i 1_{(T\ge k)}1_{(T\ge i)}. \end{aligned} \tag{1.15}$$

Now, as $(T \ge k) = (T > k-1) = (T \le k-1)^c = \cap_{i=1}^{k-1}(|S_i| \ge \varepsilon)^c$ for any $2 \le k \le n$, we find that $(T \ge k)$ is $\sigma(X_1,\ldots,X_{k-1})$ measurable. In particular, for $1\le k\le n-1$ and $k+1\le i\le n$, we obtain that X_k, $(T\ge k)$ and $(T\ge i)$ all are $\sigma(X_1,\ldots,X_{i-1})$ measurable. As $\sigma(X_1,\ldots,X_{i-1})$ is independent of $\sigma(X_i)$, this allows us to conclude

$$EX_kX_i1_{(T\ge k)}1_{(T\ge i)} = E(X_i)EX_k1_{(T\ge k)}1_{(T\ge i)} = 0, \tag{1.16}$$

since X_i has mean zero. Collecting our conclusions from (1.14), (1.15) and (1.16), we obtain $P(\max_{1\le k\le n}|S_k|\ge\varepsilon) \le \varepsilon^{-2}\sum_{k=1}^n EX_k^2 = \varepsilon^{-2}\sum_{k=1}^n VX_k = \varepsilon^{-2}VS_n$, as desired. □

Theorem 1.4.2 (Khinchin-Kolmogorov convergence theorem)**.** *Let (X_n) be a sequence of independent variables with mean zero and finite variances. If it holds that $\sum_{n=1}^\infty VX_n$ is finite, then $\sum_{n=1}^\infty X_n$ converges almost surely and in $\mathcal{L}^2$.*

Proof. For any sequence (x_n) in $\mathbb{R}$, it holds that (x_n) is Cauchy if and only if for each $m\ge1$, there is $n\ge1$ such that whenever $k\ge n+1$, it holds that $|x_k-x_n|<\frac{1}{m}$. Put $S_n=\sum_{k=1}^n X_k$. We show that S_n is almost surely convergent. We have

$$\begin{aligned} P(S_n \text{ is convergent}) &= P(S_n \text{ is Cauchy}) \\ &= P(\cap_{m=1}^\infty \cup_{n=1}^\infty \cap_{k=n+1}^\infty (|S_k-S_n|\le \tfrac{1}{m})) \\ &= P(\cap_{m=1}^\infty \cup_{n=1}^\infty (\sup_{k\ge n+1}|S_k-S_n|\le \tfrac{1}{m})). \end{aligned}$$

As the intersection of a countable family of almost sure sets again is an almost sure set, we find that in order to show almost sure convergence of S_n, it suffices to show that for each $m \geq 1$, $\cup_{n=1}^{\infty}(\sup_{k\geq n+1}|S_k - S_n| \leq \frac{1}{m})$ is an almost sure set. However, we have $P(\cup_{n=1}^{\infty}(\sup_{k\geq n+1}|S_k - S_n| \leq \frac{1}{m})) \geq P(\sup_{k\geq i+1}|S_k - S_i| \leq \frac{1}{m})$ for all $i \geq 1$, yielding $P(\cup_{n=1}^{\infty}(\sup_{k\geq n+1}|S_k - S_n| \leq \frac{1}{m})) \geq \liminf_{n\to\infty} P(\sup_{k\geq n+1}|S_k - S_n| \leq \frac{1}{m})$. Combining our conclusions, we find that in order to show the desired almost sure convergence of S_n, it suffices to show $\lim_{n\to\infty} P(\sup_{k\geq n+1}|S_k - S_n| \leq \frac{1}{m}) = 1$ for all $m \geq 1$, which is equivalent to showing

$$\lim_{n\to\infty} P(\sup_{k\geq n+1}|S_k - S_n| > \tfrac{1}{m}) = 0 \tag{1.17}$$

for all $m \geq 1$. We wish to apply Theorem 1.4.1 to show (1.17). To do so, we first note that

$$\begin{aligned} P(\sup_{k\geq n+1}|S_k - S_n| > \tfrac{1}{m}) &= P(\cup_{k=n+1}^{\infty}(\max_{n+1\leq i\leq k}|S_i - S_n| > \tfrac{1}{m})) \\ &= \lim_{k\to\infty} P(\max_{n+1\leq i\leq k}|S_i - S_n| > \tfrac{1}{m}), \end{aligned} \tag{1.18}$$

since the sequence $(\max_{n+1\leq i\leq k}|S_i - S_n| > \frac{1}{m})_{k\geq n+1}$ is increasing in k. Applying Theorem 1.4.1 to the independent variables $X_{n+1}, \ldots, X_k$ with mean zero, we find, for $k \geq n+1$,

$$P(\max_{n+1\leq i\leq k}|S_i - S_n| > \tfrac{1}{m}) = P\left(\max_{n+1\leq i\leq k}\left|\sum_{j=n+1}^{i} X_j\right| > \frac{1}{m}\right) \leq (1/m)^{-2} V \sum_{i=n+1}^{k} X_i.$$

Therefore, recalling (1.18) and using independence, we conclude

$$\begin{aligned} P(\sup_{k\geq n+1}|S_k - S_n| > \tfrac{1}{m}) &\leq \lim_{k\to\infty} (1/m)^{-2} V \sum_{i=n+1}^{k} X_i \\ &= \lim_{k\to\infty} (1/m)^{-2} \sum_{i=n+1}^{k} V X_i = (1/m)^{-2} \sum_{i=n+1}^{\infty} V X_i. \end{aligned}$$

As the series $\sum_{n=1}^{\infty} V X_n$ is assumed convergent, the tail sums converge to zero and we finally obtain $\lim_{n\to\infty} P(\sup_{k\geq n+1}|S_k - S_n| > \frac{1}{m}) = 0$, which is precisely (1.17). Thus, by our previous deliberations, we may now conclude that S_n is almost surely convergent. It remains to prove convergence in $\mathcal{L}^2$. Let S_∞ be the almost sure limit of S_n, we will show that S_n also converges in $\mathcal{L}^2$ to S_∞. By an application of Fatou's lemma, we get

$$\begin{aligned} E(S_n - S_\infty)^2 &= E \liminf_{k\to\infty}(S_n - S_k)^2 \\ &\leq \liminf_{k\to\infty} E(S_n - S_k)^2 = \liminf_{k\to\infty} E\left(\sum_{i=n+1}^{k} X_i\right)^2. \end{aligned} \tag{1.19}$$

Recalling that the sequence (X_n) consists of independent variables with mean zero, we obtain

$$E\left(\sum_{i=n+1}^{k} X_i\right)^2 = E\sum_{i=n+1}^{k}\sum_{j=n+1}^{k} X_iX_j = \sum_{i=n+1}^{k} EX_i^2 = \sum_{i=n+1}^{k} VX_i^2. \tag{1.20}$$

Combining (1.19) and (1.20), we get $E(S_n - S_\infty)^2 \leq \sum_{i=n+1}^{\infty} VX_i^2$. As the series is convergent, the tail sums converge to zero, so we conclude $\lim_{n\to\infty} E(S_n - S_\infty)^2 = 0$. This proves convergence in $\mathcal{L}^2$ and so completes the proof. □

Theorem 1.4.3 (Kolmogorov's three-series theorem)**.** *Let (X_n) be a sequence of independent variables. Let $\varepsilon > 0$. Then $\sum_{n=1}^{\infty} X_n$ converges almost surely if the following three series are convergent:*

$$\sum_{n=1}^{\infty} P(|X_n| > \varepsilon), \quad \sum_{n=1}^{\infty} EX_n 1_{(|X_n|\leq\varepsilon)} \text{ and } \sum_{n=1}^{\infty} VX_n 1_{(|X_n|\leq\varepsilon)}.$$

Proof. First note that as $\sum_{n=1}^{\infty} P(|X_n| > \varepsilon)$ is finite, we have $P(|X_n| > \varepsilon \text{ i.o.}) = 0$ by Lemma 1.2.11, which allows us to conclude $P(X_n \leq \varepsilon \text{ evt.}) = P((X_n > \varepsilon \text{ i.o.})^c) = 1$. Thus, almost surely, the sequences (X_n) and $(X_n 1_{(|X_n|\leq\varepsilon)})$ are equal from a point onwards. Therefore, $\sum_{k=1}^{n} X_k$ converges almost surely if and only if $\sum_{k=1}^{n} X_n 1_{(|X_n|\leq\varepsilon)}$ converges almost surely, so in order to prove the theorem, it suffices to show that $\sum_{k=1}^{n} X_n 1_{(|X_n|\leq\varepsilon)}$ converges almost surely. To this end, define $Y_n = X_n 1_{(|X_n|\leq\varepsilon)} - E(X_n 1_{(|X_n|\leq\varepsilon)})$. As the sequence (X_n) is independent, so is the sequence (Y_n). Also, Y_n has mean zero and finite variance, and by our assumptions, $\sum_{n=1}^{\infty} VY_n$ is finite. Therefore, by Theorem 1.4.2, it holds that $\sum_{k=1}^{n} Y_n$ converges almost surely as n tends to infinity. Thus, $\sum_{k=1}^{n} X_n 1_{(|X_n|\leq\varepsilon)} - E(X_n 1_{(|X_n|\leq\varepsilon)})$ and $\sum_{k=1}^{n} EX_k 1_{(|X_k|\leq\varepsilon)}$ converge almost surely, allowing us to conclude that $\sum_{k=1}^{n} X_n 1_{(|X_n|\leq\varepsilon)}$ converges almost surely. This completes the proof. □

1.5 The strong law of large numbers

In this section, we prove the strong law of large numbers, a key result in modern probability theory. Let (X_n) be a sequence of independent, identically distributed integrable variables with mean μ. Intuitively speaking, we would expect that $\frac{1}{n}\sum_{k=1}^{n} X_k$ in some sense converges to μ. The strong law of large numbers shows that this is indeed the case, and that the convergence is almost sure. In order to demonstrate the result, we first show two lemmas which will help us to prove the general statement by proving a simpler statement. Both lemmas consider the case of nonnegative variables. Lemma 1.5.1 establishes that in order to prove $\frac{1}{n}\sum_{k=1}^n X_k \xrightarrow{\text{a.s.}} \mu$, it suffices to prove $\frac{1}{n}\sum_{k=1}^n X_k 1_{(X_k\le k)} \xrightarrow{\text{a.s.}} \mu$, reducing to the case of bounded variables. Lemma 1.5.2 establishes that in order to prove $\frac{1}{n}\sum_{k=1}^n X_k 1_{(X_k\le k)} \xrightarrow{\text{a.s.}} \mu$, it suffices to prove $\lim_{k\to\infty}\frac{1}{n_k}\sum_{i=1}^{n_k}(X_i 1_{(X_i\le i)} - EX_i 1_{(X_i\le i)}) = 0$ for particular subsequences $(n_k)_{k\ge 1}$, reducing to a subsequence, and allowing us to focus our attention on bounded variables with mean zero.

Lemma 1.5.1. *Let (X_n) be a sequence of independent, identically distributed variables with common mean μ. Assume that $X_n \ge 0$ for all $n \ge 1$. Then $\frac{1}{n}\sum_{k=1}^n X_k$ converges almost surely if and only if $\frac{1}{n}\sum_{k=1}^n X_k 1_{(X_k\le k)}$ converges almost surely, and in the affirmative, the limits are the same.*

Proof. Let ν denote the common distribution of the X_n. Applying Tonelli's theorem, we find

$$\begin{aligned}\sum_{n=1}^\infty P(X_n \ne X_n 1_{(X_n\le n)}) &= \sum_{n=1}^\infty P(X_n > n) = \sum_{n=1}^\infty \int 1_{(x>n)}\,\mathrm{d}\nu(x)\\ &= \sum_{n=1}^\infty\sum_{k=n}^\infty \int 1_{(k<x\le k+1)}\,\mathrm{d}\nu(x) = \sum_{k=1}^\infty\sum_{n=1}^k \int 1_{(k<x\le k+1)}\,\mathrm{d}\nu(x)\\ &= \sum_{k=1}^\infty \int k1_{(k<x\le k+1)}\,\mathrm{d}\nu(x) \le \sum_{k=1}^\infty \int x1_{(k<x\le k+1)}\,\mathrm{d}\nu(x)\\ &\le \int_0^\infty x\,\mathrm{d}\nu(x) = \mu. \end{aligned} \tag{1.21}$$

Thus, $\sum_{n=1}^\infty P(X_n \ne X_n 1_{(X_n\le n)})$ is finite, and so Lemma 1.2.11 allows us to conclude that $P(X_n \ne X_n 1_{(X_n\le n)} \text{ i.o.}) = 0$, which then implies that $P(X_n = X_n 1_{(X_n\le n)} \text{ evt.}) = 1$. Hence, almost surely, X_n and $X_n 1_{(X_n\le n)}$ are equal from a point N onwards, where N is stochastic. For $n \ge N$, we therefore have

$$\frac{1}{n}\sum_{k=1}^n (X_k - X_k 1_{(X_k\le k)}) = \frac{1}{n}\sum_{k=1}^N (X_k - X_k 1_{(X_k\le k)}),$$

and by rearrangement, this yields that almost surely, for $n \ge N$,

$$\frac{1}{n}\sum_{k=1}^n X_k = \frac{1}{n}\sum_{k=1}^n X_k 1_{(X_k\le k)} + \frac{1}{n}\sum_{k=1}^N (X_k - X_k 1_{(X_k\le k)}).$$

As the last term on the right-hand side tends almost surely to zero, the conclusions of the lemma follows. □

Lemma 1.5.2. *Let (X_n) be a sequence of independent, identically distributed variables with common mean μ. Assume that for all $n \ge 1$, $X_n \ge 0$. For $\alpha > 1$, define $n_k = [\alpha^k]$, with $[\alpha^k]$*

denoting the largest integer which is less than or equal to α^k. It if holds for all $\alpha > 1$ that

$$\lim_{k\to\infty} \frac{1}{n_k}\sum_{i=1}^{n_k}(X_i 1_{(X_i\leq i)} - EX_i 1_{(X_i\leq i)}) = 0$$

almost surely, then $\frac{1}{n}\sum_{k=1}^n X_k 1_{(X_k\leq k)}$ converges to μ almost surely.

Proof. First note that as $\alpha > 1$, we have $n_k = [\alpha^k] \leq [\alpha^{k+1}] = n_{k+1}$. Therefore, (n_k) is increasing. Also, as $[\alpha^k] > \alpha^k - 1$, n_k tends to infinity as k tends to infinity. Define a sequence (Y_n) by putting $Y_n = X_n 1_{(X_n\leq n)}$. Our assumption is then that for all $\alpha > 1$, it holds that $\lim_{k\to\infty}\frac{1}{n_k}\sum_{i=1}^{n_k} Y_i - EY_i = 0$ almost surely, and our objective is to demonstrate that $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n Y_i = \mu$ almost surely. Let ν be the common distribution of the X_n. Note that by the dominated convergence theorem,

$$\lim_{n\to\infty} EY_n = \lim_{n\to\infty} EX_n 1_{(X_n\leq n)} = \lim_{n\to\infty}\int_0^\infty 1_{(x\leq n)} x \,\mathrm{d}\nu(x) = \int_0^\infty x\,\mathrm{d}\nu(x) = \mu.$$

As convergence of a sequence implies convergence of the averages, this allows us to conclude that $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n EY_k = \mu$ as well. And as convergence of a sequence implies convergence of any subsequence, we obtain $\lim_{k\to\infty}\frac{1}{n_k}\sum_{i=1}^{n_k} EY_i = \mu$ from this. Therefore, we have

$$\lim_{k\to\infty}\frac{1}{n_k}\sum_{i=1}^{n_k} Y_i = \lim_{k\to\infty}\frac{1}{n_k}\sum_{i=1}^{n_k} Y_i - EY_i + \lim_{k\to\infty}\frac{1}{n_k}\sum_{i=1}^{n_k} EY_i = \mu,$$

almost surely. We will use this to prove that $\frac{1}{n}\sum_{k=1}^n Y_k$ converges to μ. To do so, first note that since $\alpha^n - 1 < [\alpha^n] \leq \alpha^n$, it holds that

$$\alpha - \alpha^{-n} = \frac{\alpha^{n+1}-1}{\alpha^n} \leq \frac{n_{k+1}}{n_k} \leq \frac{\alpha^{n+1}}{\alpha^n - 1} = \frac{\alpha}{1-\alpha^{-n}},$$

from which it follows that $\lim_{k\to\infty}\frac{n_{k+1}}{n_k} = \alpha$. Now fix $m \geq 1$ and define a sequence $(k(m))_{m\geq 1}$ by putting $k(m) = \sup\{i \geq 1 \mid n_i \leq m\}$. As the sequence $(\{i \geq 1 \mid n_i \leq m\})_{m\geq 1}$ is increasing in m, $k(m)$ is increasing as well. And as $n_{k(m)+1} > m$, we find that $k(m)$ tends to infinity as m tends to infinity. Finally, $n_{k(m)} \leq m \leq n_{k(m)+1}$, by the properties of the supremum. As $Y_i \geq 0$, we thus find that

$$\frac{1}{n_{k(m)+1}}\sum_{i=1}^{n_{k(m)}} Y_i \leq \frac{1}{m}\sum_{i=1}^m Y_i \leq \frac{1}{n_{k(m)}}\sum_{i=1}^{n_{k(m)+1}} Y_i.$$

We therefore obtain, using that $k(m)$ tends to infinity as m tends to infinity,

$$\begin{aligned}\frac{1}{\alpha}\mu &= \liminf_{m\to\infty} \frac{n_{k(m)}}{n_{k(m)+1}} \frac{1}{n_{k(m)}} \sum_{i=1}^{n_{k(m)}} Y_i = \liminf_{m\to\infty} \frac{1}{n_{k(m)+1}} \sum_{i=1}^{n_{k(m)}} Y_i \\ &\leq \liminf_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} Y_i \leq \limsup_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} Y_i \leq \limsup_{m\to\infty} \frac{1}{n_{k(m)}} \sum_{i=1}^{n_{k(m)+1}} Y_i \\ &\leq \limsup_{m\to\infty} \frac{n_{k(m)+1}}{n_{k(m)}} \frac{1}{n_{k(m)+1}} \sum_{i=1}^{n_{k(m)+1}} Y_i = \alpha\mu.\end{aligned}$$

In conclusion, we have now shown for all $\alpha > 1$ that

$$\frac{1}{\alpha}\mu \leq \liminf_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} Y_i \leq \limsup_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} Y_i \leq \alpha\mu.$$

Letting α tend to one strictly from above, we obtain

$$\liminf_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} Y_i = \limsup_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} Y_i = \mu,$$

almost surely, proving that $\frac{1}{m}\sum_{i=1}^{m} Y_i$ is almost surely convergent with limit μ. This concludes the proof of the lemma. □

Theorem 1.5.3 (The strong law of large numbers)**.** *Let (X_n) be a sequence of independent, identically distributed variables with common mean μ. It then holds that $\frac{1}{n}\sum_{k=1}^{n} X_k$ converges almost surely to μ.*

Proof. We first consider the case where $X_n \geq 0$ for all $n \geq 1$. Combining Lemma 1.5.1 and Lemma 1.5.2, we find that in order to prove the result, it suffices to show that

$$\lim_{k\to\infty} \frac{1}{n_k} \sum_{i=1}^{n_k} (Y_i - EY_i) = 0 \tag{1.22}$$

almost surely, where $Y_n = X_n 1_{(X_n \leq n)}$, $n_k = [\alpha^k]$ and $\alpha > 1$. In order to do so, by Lemma 1.2.12, it suffices to show that for any $\varepsilon > 0$, $\sum_{k=1}^{\infty} P(|\frac{1}{n_k}\sum_{i=1}^{n_k}(Y_i - EY_i)| \geq \varepsilon)$ is finite. Using Lemma 1.2.7, we obtain

$$\begin{aligned}\sum_{k=1}^{\infty} P\left(\left|\frac{1}{n_k}\sum_{i=1}^{n_k}(Y_i - EY_i)\right| \geq \varepsilon\right) &\leq \frac{1}{\varepsilon^2}\sum_{k=1}^{\infty} E\left(\frac{1}{n_k}\sum_{i=1}^{n_k}(Y_i - EY_i)\right)^2 \\ &= \frac{1}{\varepsilon^2}\sum_{k=1}^{\infty}\sum_{i=1}^{n_k}\frac{1}{n_k^2} VY_i.\end{aligned} \tag{1.23}$$

Now, as all terms in the above are nonnegative, we may apply Tonelli's theorem to obtain

$$\begin{aligned}\sum_{k=1}^{\infty}\sum_{i=1}^{n_k}\frac{1}{n_k^2}VY_i &= \sum_{k=1}^{\infty}\sum_{i=1}^{\infty}1_{(i\leq n_k)}\frac{1}{n_k^2}VY_i \\ &= \sum_{i=1}^{\infty}\sum_{k=1}^{\infty}1_{(i\leq n_k)}\frac{1}{n_k^2}VY_i = \sum_{i=1}^{\infty}VY_i\sum_{k:n_k\geq i}\frac{1}{n_k^2}. \end{aligned} \tag{1.24}$$

We wish to identify a bound for the inner sum as a function of i. To this end, note that for $x \geq 2$, we have $[x] \geq x - 1 \geq \frac{x}{2}$, and for $1 \leq x < 2$, $[x] = 1 \geq \frac{x}{2}$ as well. Thus, for all $x \geq 1$, we have $[x] \geq \frac{x}{2}$. Let $m_i = \inf\{k \geq 1 \mid n_k \geq i\}$, we then have

$$\sum_{k:n_k\geq i}\frac{1}{n_k^2} = \sum_{k=m_i}^{\infty}\frac{1}{[\alpha^k]^2} \leq \sum_{k=m_i}^{\infty}\frac{1}{(\alpha^k/2)^2} = 4\sum_{k=m_i}^{\infty}\alpha^{-2k} = \frac{4\alpha^{-2m_i}}{1-\alpha^{-2}},$$

where we have applied the formula for summing a geometric series. Noting that m_i satisfies that $\alpha^{m_i} \geq [\alpha^{m_i}] = n_{m_i} \geq i$, we obtain $\alpha^{-2m_i} = (\alpha^{-m_i})^2 \leq i^{-2}$, resulting in the estimate $\sum_{k:n_k\geq i}\frac{1}{n_k^2} \leq 4(1-\alpha^{-2})^{-1}i^{-2}$. Combining this with (1.23) and (1.24), we find that in order to show almost sure convergence of $\frac{1}{n_k}\sum_{i=1}^{n_k}(Y_i - EY_i)$ to zero, it suffices to show that $\sum_{i=1}^{\infty}\frac{1}{i^2}VY_i$ is finite. To this end, let ν denote the common distribution of the X_n. We then apply Tonelli's theorem to obtain

$$\begin{aligned}\sum_{i=1}^{\infty}\frac{1}{i^2}VY_i \leq \sum_{i=1}^{\infty}\frac{1}{i^2}EX_i^2 1_{(X_i\leq i)} &= \sum_{i=1}^{\infty}\frac{1}{i^2}\sum_{j=1}^{i}EX_i^2 1_{(j-1<X_i\leq j)} \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{i}\frac{1}{i^2}\int_{j-1}^{j}x^2\,\mathrm{d}\nu(x) = \sum_{j=1}^{\infty}\int_{j-1}^{j}x^2\,\mathrm{d}\nu(x)\sum_{i=j}^{\infty}\frac{1}{i^2}. \end{aligned} \tag{1.25}$$

Now, for $j \geq 2$ it holds that $j + 2 \leq 2j$, leading to $j \leq 2(j-1)$ and therefore

$$\sum_{i=j}^{\infty}\frac{1}{i^2} \leq \sum_{i=j}^{\infty}\frac{1}{i(i-1)} = \sum_{i=j}^{\infty}\frac{1}{i-1} - \frac{1}{i} = \frac{1}{j-1} \leq \frac{2}{j}, \tag{1.26}$$

and the same equality for $j = 1$ follows as $\sum_{i=1}^{\infty}\frac{1}{i^2} = 1 + \sum_{i=2}^{\infty}\frac{1}{i^2} \leq 1 + 1 = 2$. Combining (1.25) and (1.26), we obtain

$$\sum_{i=1}^{\infty}\frac{1}{i^2}VY_i \leq 2\sum_{j=1}^{\infty}\frac{1}{j}\int_{j-1}^{j}x^2\,\mathrm{d}\nu(x) \leq 2\sum_{j=1}^{\infty}\int_{j-1}^{j}x\,\mathrm{d}\nu(x) = 2\int_0^{\infty}x\,\mathrm{d}\nu(x),$$

which is finite, since the X_i has finite mean. We have now shown that $\sum_{i=1}^{\infty}\frac{1}{i^2}VY_i$ is convergent, and therefore we may now conclude that $\frac{1}{n_k}\sum_{i=1}^{n_k}(Y_i - EY_i)$ converges to zero almost surely, proving (1.22). Lemma 1.5.1 and Lemma 1.5.2 now yields that $\frac{1}{n}\sum_{k=1}^{n}X_k$ converges almost surely to μ.

It remains to extend the result to the case where X_n is not assumed to be nonnegative. Therefore, we now let (X_n) be any sequence of independent, identically distributed variables with mean μ. With $x^+ = \max\{0, x\}$ and $x^- = \max\{0, -x\}$, Lemma 1.3.7 shows that the sequences (X_n^+) and (X_n^-) each are independent and identically distributed with finite means $\int x^+ \, \mathrm{d}\nu(x)$ and $\int x^- \, \mathrm{d}\nu(x)$, and both sequences consist only of nonnegative variables. Therefore, from what we already have shown, we obtain

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k^+ - \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X_k^- = \int x^+ \, \mathrm{d}\nu(x) - \int x^- \, \mathrm{d}\nu(x) = \mu,$$

as desired. □

In fact, the convergence in Theorem 1.5.3 holds not only almost surely, but also in $\mathcal{L}^1$. In Chapter 2, we will obtain this as a consequence of a more general convergence theorem. Before concluding the chapter, we give an example of a simple statistical application of Theorem 1.5.3.

Example 1.5.4. Consider a measurable space $(\Omega, \mathcal{F})$ endowed with a sequence of random variables (X_n). Assume given a parameter set Θ and a set of probability measures $(P_\theta)_{\theta\in\Theta}$ such that for the probability space $(\Omega, \mathcal{F}, P_\theta)$, (X_n) consists of independent and identically distributed variables with second moment. Assume further that the first moment is ξ_θ and that the second moment is σ_θ^2. Natural estimators of the mean and variance parameter functions based on n samples are then

$$\hat{\xi}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \quad \text{and} \quad \hat{\sigma}_n^2 = \frac{1}{n}\sum_{k=1}^{n}\left(X_k - \frac{1}{n}\sum_{i=1}^{n} X_i\right)^2.$$

Theorem 1.5.3 allows us to conclude that under P_θ, it holds that $\hat{\xi}_n \xrightarrow{\text{a.s.}} \xi_\theta$, and furthermore, by two further applications of Theorem 1.5.3, as well as Lemma 1.2.6 and Lemma 1.2.10, we obtain

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \frac{2}{n}X_k\sum_{i=1}^{n} X_i + \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 = \frac{1}{n}\sum_{k=1}^{n} X_k^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2 \xrightarrow{\text{a.s.}} \sigma_\theta^2.$$

The strong law of large numbers thus allows us to conclude that the natural estimators $\hat{\xi}_n$ and $\hat{\sigma}_n^2$ converge almost surely to the true mean and variance. ∘

1.6 Exercises

Exercise 1.1. Let X be a random variable, and let (a_n) be a sequence of real numbers converging to zero. Define $X_n = a_n X$. Show that X_n converges almost surely to zero. ∘

Exercise 1.2. Give an example of a sequence of random variables (X_n) such that (X_n) converges in probability but does not converge almost surely to any variable. Give an example of a sequence of random variables (X_n) such that (X_n) converges in probability but does not converge in $\mathcal{L}^1$ to any variable. ∘

Exercise 1.3. Let (X_n) be a sequence of random variables such that X_n is Poisson distributed with parameter $1/n$. Show that X_n converges in $\mathcal{L}^1$ to zero. ∘

Exercise 1.4. Let (X_n) be a sequence of random variables such that X_n is Gamma distributed with shape parameter n^2 and scale parameter $1/n$. Show that X_n does not converge in $\mathcal{L}^1$ to any integrable variable. ∘

Exercise 1.5. Consider a probability space $(\Omega, \mathcal{F}, P)$ such that Ω is countable and such that $\mathcal{F}$ is the power set of Ω. Let (X_n) be a sequence of random variables (X_n) on $(\Omega, \mathcal{F}, P)$, and let X be another variable. Show that if $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{\text{a.s.}} X$. ∘

Exercise 1.6. Let (X_n) be a sequence of random variables and let X be another variable. Let (F_n) be a sequence of sets in $\mathcal{F}$. Assume that for all k, $X_n 1_{F_k} \xrightarrow{P} X 1_{F_k}$, and assume that $\lim_{k\to\infty} P(F_k^c) = 0$. Show that $X_n \xrightarrow{P} X$. ∘

Exercise 1.7. Let (X_n) be a sequence of random variables and let X be another variable. Let $(\varepsilon_k)_{k\geq 1}$ be a sequence of nonnegative real numbers converging to zero. Show that X_n converges in probability to X if and only if $\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon_k) = 0$ for all $k \geq 1$. ∘

Exercise 1.8. Let (X_n) be an sequence of random variables, and let X be some other variable. Show that $X_n \xrightarrow{\text{a.s.}} X$ if and only if $\sup_{k\geq n} |X_k - X| \xrightarrow{P} 0$. ∘

Exercise 1.9. Let X and Y be two variables, and define

$$d(X, Y) = E\frac{|X - Y|}{1 + |X - Y|}.$$

Show that d is a pseudometric on the space of real stochastic variables, in the sense that $d(X, Y) \leq d(X, Z) + d(Z, Y)$, $d(X, Y) = d(Y, X)$ and $d(X, X) = 0$ for all X, Y and Z. Show that $d(X,Y)=0$ if and only if X and Y are almost surely equal. Let (X_n) be a sequence of random variables and let X be some other variable. Show that $X_n \xrightarrow{P} X$ if and only if $\lim_{n\to\infty} d(X_n,X)=0$. ∘

Exercise 1.10. Let (X_n) be a sequence of random variables. Show that there exists a sequence of positive constants (c_n) such that $c_n X_n \xrightarrow{\text{a.s.}} 0$. ∘

Exercise 1.11. Let (X_n) be a sequence of i.i.d. variables with mean zero. Assume that X_n has fourth moment. Show that for all $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k\right| \geq \varepsilon\right) \leq \frac{4EX_1^4}{\varepsilon^4}\sum_{n=1}^{\infty}\frac{1}{n^2}.$$

Use this to prove the following result: For a sequence (X_n) of i.i.d. variables with fourth moment and mean μ, it holds that $\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{\text{a.s.}} \mu$. ∘

Exercise 1.12. Let (X_n) be a sequence of random variables and let X be some other variable. Assume that there is $p>1$ such that $\sup_{n\geq 1} E|X_n|^p$ is finite. Show that if $X_n \xrightarrow{P} X$, then $E|X|^p$ is finite and $X_n \xrightarrow{\mathcal{L}^q} X$ for $1 \leq q < p$. ∘

Exercise 1.13. Let (X_n) be a sequence of random variables, and let X be some other variable. Assume that $X_n \xrightarrow{\text{a.s.}} X$. Show that for all $\varepsilon > 0$, there exists $F \in \mathcal{F}$ with $P(F^c) \leq \varepsilon$ such that

$$\lim_{n\to\infty} \sup_{\omega\in F} |X_n(\omega) - X(\omega)| = 0,$$

corresponding to X_n converging uniformly to X on F. ∘

Exercise 1.14. Let (X_n) be a sequence of random variables. Let X be some other variable. Let $p>0$. Show that if $\sum_{n=1}^{\infty} E|X_n - X|^p$ is finite, then $X_n \xrightarrow{\text{a.s.}} X$. ∘

Exercise 1.15. Let (X_n) be a sequence of random variables, and let X be some other variable. Assume that almost surely, the sequence (X_n) is increasing. Show that if $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{\text{a.s.}} X$. ∘

Exercise 1.16. Let (X_n) be a sequence of random variables and let (ε_n) be a sequence of nonnegative constants. Show that if $\sum_{n=1}^{\infty} P(|X_{n+1} - X_n| \geq \varepsilon_n)$ and $\sum_{n=1}^{\infty} \varepsilon_n$ are finite, then (X_n) converges almost surely to some random variable. ∘

Exercise 1.17. Let (U_n) be a sequence of i.i.d. variables with common distribution being the uniform distribution on the unit interval. Define $X_n = \max\{U_1,\ldots,U_n\}$. Show that X_n

converges to 1 almost surely and in $\mathcal{L}^p$ for $p \geq 1$. ∘

Exercise 1.18. Let (X_n) be a sequence of random variables. Show that if there exists $c > 0$ such that $\sum_{n=1}^{\infty} P(X_n > c)$ is finite, then $\sup_{n \geq 1} X_n$ is almost surely finite. ∘

Exercise 1.19. Let (X_n) be a sequence of independent random variables. Show that if $\sup_{n \geq 1} X_n$ is almost surely finite, there exists $c > 0$ such that $\sum_{n=1}^{\infty} P(X_n > c)$ is finite. ∘

Exercise 1.20. Let (X_n) be a sequence of i.i.d. random variables with common distribution being the standard exponential distribution. Calculate $P(X_n/\log n > c \text{ i.o.})$ for all $c > 0$ and use the result to show that $\limsup_{n\to\infty} X_n/\log n = 1$ almost surely. ∘

Exercise 1.21. Let (X_n) be a sequence of random variables, and let $\mathcal{J}$ be the corresponding tail-σ-algebra. Let $B \in \mathcal{B}$. Show that $(X_n \in B \text{ i.o.})$ and $(X_n \in B \text{ evt.})$ are in $\mathcal{J}$. ∘

Exercise 1.22. Let (X_n) be a sequence of random variables, and let $\mathcal{J}$ be the corresponding tail-σ-algebra. Let $B \in \mathcal{B}$ and let (a_n) be a sequence of real numbers. Show that if $\lim_{n\to\infty} a_n = 0$, then $(\lim_{n\to\infty} \sum_{k=1}^{n} a_{n-k+1} X_k \in B)$ is in $\mathcal{J}$. ∘

Exercise 1.23. Let (X_n) be a sequence of independent random variables concentrated on $\{0, 1\}$ with $P(X_n = 1) = p_n$. Show that $X_n \xrightarrow{P} 0$ if and only if $\lim_{n\to\infty} p_n = 0$, and show that $X_n \xrightarrow{\text{a.s.}} 0$ if and only if $\sum_{n=1}^{\infty} p_n$ is finite. ∘

Exercise 1.24. Let (X_n) be a sequence of nonnegative random variables. Show that if $\sum_{n=1}^{\infty} EX_n$ is finite, then $\sum_{k=1}^{n} X_k$ is almost surely convergent. ∘

Exercise 1.25. Let (X_n) be an i.i.d. sequence of independent random variables such that $P(X_n = 1)$ and $P(X_n = -1)$ both are equal to $\frac{1}{2}$. Let (a_n) be a sequence of real numbers. Show that the sequence $\sum_{k=1}^{n} a_k X_k$ either is almost surely divergent or almost surely convergent. Show that the sequence is almost surely convergent if $\sum_{n=1}^{\infty} a_n^2$ is finite. ∘

Exercise 1.26. Give an example of a sequence (X_n) of independent variables with first moment such that $\sum_{k=1}^{n} X_k$ converges almost surely while $\sum_{k=1}^{n} EX_k$ diverges. ∘

Exercise 1.27. Let (X_n) be a sequence of independent random variables with $EX_n = 0$. Assume that $\sum_{n=1}^{\infty} E(X_n^2 1_{(|X_n| \leq 1)} + |X_n| 1_{(|X_n| > 1)})$ is finite. Show that $\sum_{k=1}^{n} X_k$ is almost surely convergent. ∘

Exercise 1.28. Let (X_n) be an sequence of independent and identically distributed random variables. Show that $E|X_1|$ is finite if and only if $P(|X_n| > n \text{ i.o.}) = 0$. ∘

Exercise 1.29. Let (X_n) be an sequence of independent and identically distributed random variables. Assume that there is c such that $\frac{1}{n}\sum_{k=1}^n X_k \xrightarrow{\text{a.s.}} c$. Show that $E|X_1|$ is finite and that $EX_1 = c$. ∘

Chapter 2

Ergodicity and stationarity

In Section 1.5, we proved the strong law of large numbers, which shows that for a sequence (X_n) of integrable, independent and identically distributed variables, the empirical means converge almost surely to the true mean. A reasonable question is whether such a result may be extended to more general cases. Consider a sequence (X_n) where each X_n has the same distribution ν with mean μ. If the dependence between the variables is sufficiently weak, we may hope that the empirical means still converge to the true mean.

One fruitful case of sufficiently weak dependence turns out to be embedded in the notion of a stationary stochastic process. The notion of stationarity is connected with the notion of measure-preserving mappings. Our plan for this chapter is as follows. In Section 2.1, we investigate measure-preserving mappings, in particular proving the ergodic theorem, which is a type of law of large numbers. Section 2.2 investigates sufficient criteria for the ergodic theorem to hold. Finally, in Section 2.3, we apply our results to stationary processes and prove versions of the law of large numbers for such processes.

2.1 Measure preservation, invariance and ergodicity

As in the previous chapter, we work in the context of a probability space $(\Omega, \mathcal{F}, P)$. Our main interest of this section will be a particular type of measurable mapping $T : \Omega \to \Omega$. Recall that for such a mapping T, the image measure $T(P)$ is the measure on $\mathcal{F}$ defined by that for any $F \in \mathcal{F}$, $T(P)(F) = P(T^{-1}(F))$.

Definition 2.1.1. *Let $T : \Omega \to \Omega$ be measurable. We say that T is P-measure preserving, or measure preserving for P, or simply measure preserving, if the image measure $T(P)$ is equal to P.*

Another way to state Definition 2.1.1 is thus that T is measure preserving precisely when $P(T^{-1}(F)) = P(F)$ for all $F \in \mathcal{F}$.

Definition 2.1.2. *Let $T : \Omega \to \Omega$ be measurable. The T-invariant σ-algebra, or simply the invariant σ-algebra, is defined by $\mathcal{I}_T = \{F \in \mathcal{F} \mid T^{-1}(F) = F\}$.*

As the operation of taking the preimage $T^{-1}(F)$ is stable under complements and countable unions, the set family $\mathcal{I}_T$ in Definition 2.1.2 is in fact a σ-algebra.

Definition 2.1.3. *Let $T : \Omega \to \Omega$ be measurable and measure preserving. The mapping T is said to be P-ergodic, or to be ergodic for P, or simply ergodic, if $P(F)$ is either zero or one for all $F \in \mathcal{I}_T$.*

We have now introduced three concepts: measure preservation of a mapping T, the invariant σ-algebra for a mapping T and ergodicity for a mapping T. These will be the main objects of study for this section. Before proceeding, we introduce a final auxiliary concept. Recall that $\circ$ denotes function composition, in the sense that if $T : \Omega \to \Omega$ and $X : \Omega \to \mathbb{R}$, $X \circ T$ denotes the mapping from Ω to $\mathbb{R}$ defined by $(X \circ T)(\omega) = X(T(\omega))$.

Definition 2.1.4. *Let $T : \Omega \to \Omega$ be measurable, and let X be a random variable. X is said to be T-invariant, or simply invariant, if $X \circ T = X$.*

We are now ready to begin preparations for the main result of this section, the ergodic theorem. Note that for $T : \Omega \to \Omega$, it is sensible to consider $T \circ T$, denoted T^2, which is defined by $(T \circ T)(\omega) = T(T(\omega))$, and more generally, T^n for some $n \geq 1$. In the following, T denotes some measurable mapping from Ω to Ω. The ergodic theorem states that if T is measure preserving and ergodic, it holds for any variable X with p'th moment, $p \geq 1$, that the average $\frac{1}{n}\sum_{k=1}^{n} X \circ T^{k-1}$ converges almost surely and in $\mathcal{L}^p$ to the mean EX. In order to show the result, we first need a few lemmas.

Lemma 2.1.5. *Let X be a random variable. It holds that X is invariant if and only if X is $\mathcal{I}_T$ measurable.*

Proof. First assume that X is invariant, and consider $A \in \mathcal{B}$. We need to prove that $(X \in A)$ is in $\mathcal{I}_T$, which is equivalent to showing $T^{-1}(X \in A) = (X \in A)$. To obtain this, we simply note that as $X \circ T = X$,

$$\begin{aligned} T^{-1}(X \in A) &= \{\omega \in \Omega \mid T(\omega) \in (X \in A)\} = \{\omega \in \Omega \mid X(T(\omega)) \in A\} \\ &= (X \circ T \in A) = (X \in A). \end{aligned}$$

Thus, X is $\mathcal{I}_T$ measurable. Next, assume that X is $\mathcal{I}_T$ measurable, we wish to demonstrate that X is invariant. Fix some $x \in \mathbb{R}$. As $\{x\} \in \mathcal{B}$, we have $(X = x) \in \mathcal{I}_T$, yielding $(X = x) = T^{-1}(X = x) = (X \circ T = x)$. Next, fix $\omega \in \Omega$. We wish to show that $X(\omega) = (X \circ T)(\omega)$. From what we just proved, it holds that $X(T(\omega)) = x$ if and only if $X(\omega) = x$. In particular, $X(T(\omega)) = X(\omega)$ if and only if $X(\omega) = X(\omega)$, and the latter is trivially true. Thus, $(X \circ T)(\omega) = X(\omega)$, so $X \circ T = X$. Hence, X is invariant. □

Lemma 2.1.6. *Let T be P-measure preserving. Let X be an integrable random variable. Define $S_n = \sum_{k=1}^n X \circ T^{k-1}$. It then holds that $EX1_{(\sup_{n\geq 1} \frac{1}{n} S_n > 0)} \geq 0$.*

Proof. Fix n and define $M_n = \max\{0, S_1, \ldots, S_n\}$. Note that $\sup_{n\geq 1} \frac{1}{n} S_n > 0$ if and only if there exists n such that $\frac{1}{n} S_n > 0$, which is the case if and only if there exists n such that $M_n > 0$. As the sequence of sets $((M_n > 0))_{n\geq 1}$ is increasing, the dominated convergence theorem then shows that

$$EX1_{(\sup_{n\geq 1} \frac{1}{n} S_n > 0)} = EX1_{\cup_{n=1}^\infty (M_n > 0)} = E \lim_{n\to\infty} X1_{(M_n>0)} = \lim_{n\to\infty} EX1_{(M_n>0)},$$

and so it suffices to prove that $EX1_{(M_n>0)} \geq 0$ for each n. To do so, fix n. Note that as T is measure preserving, so is T^n for all n. As M_n is nonnegative, we then have

$$0 \leq EM_n \leq E\sum_{i=1}^n |S_i| \leq E\sum_{i=1}^n \sum_{k=1}^i |X| \circ T^{k-1} = \sum_{i=1}^n \sum_{k=1}^i E|X| = \frac{n(n+1)}{2} E|X|,$$

which shows that M_n is integrable. As $E(M_n \circ T)1_{(M_n>0)} \leq E(M_n \circ T) = EM_n$ by the measure preservation property of T, $(M_n \circ T)1_{(M_n>0)}$ is also integrable, and we have

$$\begin{aligned} EX1_{(M_n>0)} &= E(X + M_n \circ T)1_{(M_n>0)} - E(M_n \circ T)1_{(M_n>0)} \\ &\geq E(X + M_n \circ T)1_{(M_n>0)} - EM_n \\ &= E(X + M_n \circ T)1_{(M_n>0)} - EM_n 1_{(M_n>0)}. \end{aligned}$$

Therefore, it suffices to show that $(X + M_n \circ T)1_{(M_n>0)} \geq M_n 1_{(M_n>0)}$. To do so, note that for $1 \leq k \leq n-1$, it holds that $X + M_n \circ T \geq X + S_k \circ T = X + \sum_{i=1}^k X \circ T^i = S_{k+1}$, and also $X + M_n \circ T \geq X = S_1$. Therefore, $X + M_n \circ T \geq \max\{S_1, \ldots, S_n\}$. From this, it follows that $(X + M_n \circ T)1_{(M_n>0)} \geq \max\{S_1, \ldots, S_n\}1_{(M_n>0)} = M_n 1_{(M_n>0)}$, as desired. □

Theorem 2.1.7 (Birkhoff-Khinchin ergodic theorem)**.** *Let $p \geq 1$, let X be a variable with p'th moment and let T be a mapping which is measure preserving and ergodic. It then holds that $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X \circ T^{k-1} = EX$ almost surely and in $\mathcal{L}^p$.*

Proof. We first consider the case where X has mean zero. Define $S_n = \sum_{k=1}^{n} X \circ T^{k-1}$. We need to show that $\lim_{n\to\infty} \frac{1}{n} S_n = 0$ almost surely and in $\mathcal{L}^p$. Put $Y = \limsup_{n\to\infty} \frac{1}{n} S_n$, we will show that almost surely, $Y \leq 0$. If we can obtain this, a symmetry argument will then allow us to obtain the desired conclusion. In order to prove that $Y \leq 0$ almost surely, we first take $\varepsilon > 0$ and show that $P(Y > \varepsilon) = 0$. To this end, we begin by noting that

$$Y \circ T = \limsup_{n\to\infty} \frac{1}{n}(S_n \circ T) = \limsup_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} X \circ T^k = \limsup_{n\to\infty} \frac{1}{n}(S_{n+1} - X) = Y,$$

so Y is T-invariant. Therefore, by Lemma 2.1.5, $(Y > \varepsilon)$ is in $\mathcal{I}_T$. As T is ergodic, it therefore holds that $P(Y > \varepsilon)$ either is zero or one, our objective is to show that the probability in fact is zero. We will obtain this by applying Lemma 2.1.6 to a suitably chosen random variable. Define $X' : \Omega \to \mathbb{R}$ by $X' = (X - \varepsilon)1_{(Y>\varepsilon)}$ and put $S'_n = \sum_{k=1}^{n} X' \circ T^{k-1}$, we will eventually apply Lemma 2.1.6 to X'. First note that

$$\begin{aligned} S'_n &= \sum_{k=1}^{n} X' \circ T^{k-1} = \sum_{k=1}^{n} (X - \varepsilon)1_{(Y>\varepsilon)} \circ T^{k-1} \\ &= \sum_{k=1}^{n} X1_{(Y>\varepsilon)} \circ T^{k-1} - \varepsilon \sum_{k=1}^{n} 1_{(Y>\varepsilon)} \circ T^{k-1} \\ &= 1_{(Y>\varepsilon)} \left(\sum_{k=1}^{n} X \circ T^{k-1} - n\varepsilon \right) = 1_{(Y>\varepsilon)}(S_n - n\varepsilon), \end{aligned}$$

allowing us to conclude

$$\begin{aligned} (Y > \varepsilon) &= (Y > \varepsilon) \cap (\sup_{n\geq 1} \tfrac{1}{n} S_n > \varepsilon) = (Y > \varepsilon) \cap \cup_{n=1}^{\infty} (\tfrac{1}{n} S_n > \varepsilon) \\ &= \cup_{n=1}^{\infty} (Y > \varepsilon) \cap (S_n - n\varepsilon > 0) = \cup_{n=1}^{\infty} (S'_n > 0) \\ &= \cup_{n=1}^{\infty} (\tfrac{1}{n} S'_n > 0) = (\sup_{n\geq 1} \tfrac{1}{n} S'_n > 0). \end{aligned} \tag{2.1}$$

This relates the event $(Y > \varepsilon)$ to the sequence (S'_n). Applying Lemma 2.1.6 and recalling (2.1), we obtain $E1_{(Y>\varepsilon)}X' \geq 0$, which implies

$$\varepsilon P(Y > \varepsilon) \leq E1_{(Y>\varepsilon)}X. \tag{2.2}$$

Finally, recall that by ergodicity of T, $P(Y > \varepsilon)$ is either zero or one. If $P(Y > \varepsilon)$ is one, (2.2) yields $\varepsilon \leq 0$, a contradiction. Therefore, we must have that $P(Y > \varepsilon)$ is zero. We now use this to complete the proof of almost sure convergence. As $P(Y > \varepsilon)$ is zero for all $\varepsilon > 0$, we conclude that $P(Y > 0) = P(\cup_{n=1}^\infty (Y > \frac{1}{n})) = 0$, so $\limsup_{n\to\infty} \frac{1}{n} S_n = Y \leq 0$ almost surely. Next note that

$$-\liminf_{n\to\infty} \frac{1}{n} S_n = -\liminf_{n\to\infty} \frac{1}{n} \sum_{k=1}^n X \circ T^{k-1} = \limsup_{n\to\infty} \frac{1}{n} \sum_{k=1}^n (-X) \circ T^{k-1},$$

so applying the same result with $-X$ instead of X, we also obtain $-\liminf_{n\to\infty} \frac{1}{n} S_n \leq 0$ almost surely. All in all, this shows that $0 \leq \liminf_{n\to\infty} \frac{1}{n} S_n \leq \limsup_{n\to\infty} \frac{1}{n} S_n \leq 0$ almost surely, so $\lim_{n\to\infty} \frac{1}{n} S_n = 0$ almost surely, as desired. Finally, considering the case where EX is nonzero, we may use our previous result with the variable $X - EX$ to obtain $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n X \circ T^{k-1} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^n (X - EX) \circ T^{k-1} + EX = EX$, completing the proof of almost sure convergence in the general case.

It remains to show convergence in $\mathcal{L}^p$, meaning that we wish to prove the convergence of $E|EX - \frac{1}{n} \sum_{k=1}^n X \circ T^{k-1}|^p$ to zero as n tends to infinity. With $\|\cdot\|_p$ denoting the seminorm on $\mathcal{L}^p(\Omega, \mathcal{F}, P)$ given by $\|X\|_p = (EX^p)^{1/p}$, we will show that $\|EX - \frac{1}{n} S_n\|_p$ tends to zero. To this end, we fix $m \geq 1$ and define $X' = X 1_{(|X| \leq m)}$ and $S_n' = \sum_{k=1}^n X' \circ T^{k-1}$. By the triangle inequality, we obtain

$$\|EX - \tfrac{1}{n} S_n\|_p \leq \|EX - EX'\|_p + \|EX' - \tfrac{1}{n} S_n'\|_p + \|\tfrac{1}{n} S_n' - \tfrac{1}{n} S_n\|_p.$$

We consider the each of the three terms on the right-hand side. For the first term, it holds that $\|EX - EX'\|_p = |EX - EX'| = |EX 1_{(|X|>m)}| \leq E|X| 1_{(|X|>m)}$. As for the second term, the results already proven show that $\frac{1}{n} S_n'$ converges almost surely to EX'. As $|S_n'| = |\frac{1}{n} \sum_{k=1}^n X' \circ T^{k-1}| \leq m$, the dominated convergence theorem allows us to conclude $\lim_{n\to\infty} E|X' - \frac{1}{n} S_n'|^p = E \lim_{n\to\infty} |X' - \frac{1}{n} S_n'|^p = 0$, which implies that we have $\lim_{n\to\infty} \|EX' - \frac{1}{n} S_n\|_p = 0$. Finally, we may apply the triangle inequality and the measure preservation property of T to obtain

$$\begin{aligned} \|\tfrac{1}{n} S_n' - \tfrac{1}{n} S_n\|_p &= \left\| \frac{1}{n} \sum_{k=1}^n X' \circ T^{k-1} - \frac{1}{n} \sum_{k=1}^n X \circ T^{k-1} \right\|_p \leq \frac{1}{n} \sum_{k=1}^n \|X \circ T^{k-1} - X' \circ T^{k-1}\|_p \\ &= \frac{1}{n} \sum_{k=1}^n \left(\int |(X - X') \circ T^{k-1}|^p \, \mathrm{d}P \right)^{1/p} = \frac{1}{n} \sum_{k=1}^n \left(\int |(X - X')|^p \, \mathrm{d}P \right)^{1/p} \\ &= \|X - X'\|_p = (E|X|^p 1_{(|X|>m)})^{1/p}. \end{aligned}$$

Combining these observations, we obtain

$$\begin{aligned} \limsup_{n\to\infty} \|EX - \tfrac{1}{n} S_n\|_p &\leq \limsup_{n\to\infty} \|EX - EX'\|_p + \|EX' - \tfrac{1}{n} S_n'\|_p + \|\tfrac{1}{n} S_n' - \tfrac{1}{n} S_n\|_p \\ &\leq E|X| 1_{(|X|>m)} + (E|X|^p 1_{(|X|>m)})^{1/p} \end{aligned} \tag{2.3}$$

By the dominated convergence theorem, both of these terms tend to zero as m tends to infinity. As the bound in (2.3) holds for all m, we conclude $\limsup_{n\to\infty} \|EX - \frac{1}{n}S_n\|_p = 0$, which yields convergence in $\mathcal{L}^p$. □

Theorem 2.1.7 shows that for any variable X with p'th moment and any measure preserving and ergodic transformation T, a version of the strong law of large numbers holds for the process $(X \circ T^{k-1})_{k\geq 1}$ in the sense that $\frac{1}{n}\sum_{k=1}^n X \circ T^{k-1}$ converges almosts surely and in $\mathcal{L}^p$ to EX. Note that in this case, the measure preservation property of T shows that X and $X \circ T^{k-1}$ have the same distribution for all $k \geq 1$. Therefore, Theorem 2.1.7 is a type of law of large numbers for processes of identical, but not necessarily independent variables.

2.2 Criteria for measure preservation and ergodicity

To apply Theorem 2.1.7, we need to be able to show measure preservation and ergodicity. In this section, we prove some sufficient criteria which will help make this possible in practical cases. Throughout this section, T denotes a measurable mapping from Ω to Ω.

First, we consider a simple lemma showing that in order to prove that T is measure preserving, it suffices to check the claim only for a generating family which is stable under finite intersections.

Lemma 2.2.1. *Let $\mathbb{H}$ be a generating family for $\mathcal{F}$ which is stable under finite intersections. If $P(T^{-1}(F)) = P(F)$ for all $F \in \mathbb{H}$, then T is P-measure preserving.*

Proof. As both P and $T(P)$ are probability measures, this follows from the uniqueness theorem for probability measures. □

Next, we consider the somewhat more involved problem of showing that a measure preserving mapping is ergodic. A simple first result is the following.

Theorem 2.2.2. *Let T be measure preserving. Then T is ergodic if and only if every invariant random variable is constant almost surely.*

Proof. First assume that T is ergodic. Let X be an invariant random variable. By Lemma 2.1.5, X is $\mathcal{I}_T$ measurable, so in particular $(X \leq x) \in \mathcal{I}_T$ for all $x \in \mathbb{R}$. As T is ergodic, all events in $\mathcal{I}_T$ have probability zero or one, so we find that $P(X \leq x)$ is zero or one for all $x \in \mathbb{R}$.

We claim that this implies that X is constant almost surely. To this end, we define c by putting $c = \sup\{x \in \mathbb{R} \mid P(X \leq x) = 0\}$. As we cannot have $P(X \leq x) = 1$ for all $x \in \mathbb{R}$, $\{x \in \mathbb{R} \mid P(X \leq x) = 0\}$ is nonempty, so c is not minus infinity. And as we cannot have $P(X \leq x) = 0$ for all $x \in \mathbb{R}$, $\{x \in \mathbb{R} \mid P(X \leq x) = 0\}$ is not all of $\mathbb{R}$. As $x \mapsto P(X \leq x)$ is increasing, this implies that $\{x \in \mathbb{R} \mid P(X \leq x) = 0\}$ is bounded from above, so c is not infinity. Thus, c is finite.

Now, by definition, c is the least upper bound of the set $\{x \in \mathbb{R} \mid P(X \leq x) = 0\}$. Therefore, any number strictly smaller than c is not an upper bound. From this we conclude that for $n \geq 1$, there is c_n with $c - \frac{1}{n} < c_n$ such that $P(X \leq c_n) = 0$. Therefore, we must also have $P(X \leq c - \frac{1}{n}) \leq P(X \leq c_n) = 0$, and so $P(X < c) = \lim_{n\to\infty} P(X \leq c - \frac{1}{n}) = 0$. On the other hand, as c is an upper bound for the set $\{x \in \mathbb{R} \mid P(X \leq x) = 0\}$, it holds for any $\varepsilon > 0$ that $P(X \leq c + \varepsilon) \neq 0$, yielding that for all $\varepsilon > 0$, $P(X \leq c + \varepsilon) = 1$. Therefore, $P(X \leq c) = \lim_{n\to\infty} P(X \leq c + \frac{1}{n}) = 1$. All in all, we conclude $P(X = c) = 1$, so X is constant almost surely. This proves the first implication of the theorem.

Next, assume that every invariant random variable is constant almost surely, we wish to prove that T is ergodic. Let $F \in \mathcal{I}_T$, we have to show that $P(F)$ is either zero or one. Note that 1_F is $\mathcal{I}_T$ measurable and so invariant by Lemma 2.1.5. Therefore, by our assumption, 1_F is almost surely constant, and this implies that $P(F)$ is either zero or one. This proves the other implication and so concludes the proof. □

Theorem 2.2.2 is occasionally useful if the T-invariant random variables are easy to characterize. The following theorem shows a different avenue for proving ergodicity based on a sort of asymptotic independence criterion.

Theorem 2.2.3. *Let T be P-measure preserving. T is ergodic if and only if it holds for all $F, G \in \mathcal{F}$ that $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n P(F \cap T^{-(k-1)}(G)) = P(F)P(G)$.*

Proof. First assume that T is ergodic. Fix $F, G \in \mathcal{F}$. Applying Theorem 2.1.7 with the integrable variable 1_G, and noting that $1_G \circ T^{k-1} = 1_{T^{-(k-1)}(G)}$, we find that $\frac{1}{n}\sum_{k=1}^n 1_{T^{-(k-1)}(G)}$ converges almost surely to $P(G)$. Therefore, $\frac{1}{n}\sum_{k=1}^n 1_F 1_{T^{-(k-1)}(G)}$ converges almost surely to $1_F P(G)$. As this sequence of variables is bounded, the dominated convergence theorem yields

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}P(F\cap T^{-(k-1)}(G)) &= \lim_{n\to\infty}E\frac{1}{n}\sum_{k=1}^{n}1_F 1_{T^{-(k-1)}(G)}\\ &= E\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}1_F 1_{T^{-(k-1)}(G)}\\ &= E1_F P(G) = P(F)P(G),\end{aligned}$$

proving the first implication. Next, we consider the other implication. Assume that for all $F, G \in \mathcal{F}$, $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}P(F\cap T^{-(k-1)}(G)) = P(F)P(G)$. We wish to show that T is ergodic. Let $F \in \mathcal{I}_T$, we then obtain

$$P(F) = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}P(F\cap T^{-(k-1)}(F)) = P(F)^2,$$

so that $P(F)$ is either zero or one, and thus T is ergodic. □

Definition 2.2.4. *If* $\lim_{n\to\infty}P(F\cap T^{-n}(G)) = P(F)P(G)$ *for all* $F, G \in \mathcal{F}$, *we say that* T *is mixing. If* $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}|P(F\cap T^{-(k-1)}(G)) - P(F)P(G)| = 0$ *for all* $F, G \in \mathcal{F}$, *we say that* T *is weakly mixing.*

Theorem 2.2.5. *Let* T *be measure preserving. If* T *is mixing, then* T *is weakly mixing. If* T *is weakly mixing, then* T *is ergodic.*

Proof. First assume that T is mixing. Let $F, G \in \mathcal{F}$. As T is mixing, we find that $\lim_{n\to\infty}P(F\cap T^{-n}(G)) = P(F)P(G)$, and so $\lim_{n\to\infty}|P(F\cap T^{-n}(G)) - P(F)P(G)| = 0$. As convergence of a sequence implies convergence of the averages, this implies that we have $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}|P(F\cap T^{-(k-1)}(G)) - P(F)P(G)| = 0$, so T is weakly mixing. Next, assume that T is weakly mixing, we wish to show that T is ergodic. Let $F, G \in \mathcal{F}$. As $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}|P(F\cap T^{-(k-1)}(G)) - P(F)P(G)| = 0$, we also obtain

$$\begin{aligned}&\limsup_{n\to\infty}\left|\frac{1}{n}\sum_{k=1}^{n}P(F\cap T^{-(k-1)}(G)) - P(F)P(G)\right|\\ \le&\limsup_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}|P(F\cap T^{-(k-1)}(G)) - P(F)P(G)|,\end{aligned}$$

which is zero, so $\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}P(F\cap T^{-(k-1)}(G)) = P(F)P(G)$, and Theorem 2.2.3 shows that T is ergodic. This proves the theorem. □

Lemma 2.2.6. *Let T be measure preserving, and let $\mathbb{H}$ be a generating family for $\mathcal{F}$ which is stable under finite intersections. Assume that one of the following holds:*

(1). $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n P(F\cap T^{-(k-1)}(G)) = P(F)P(G)$ *for all* $F,G\in\mathbb{H}$.

(2). $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n |P(F\cap T^{-(k-1)}(G)) - P(F)P(G)| = 0$ *for all* $F,G\in\mathbb{H}$.

(3). $\lim_{n\to\infty} P(F\cap T^{-n}(G)) = P(F)P(G)$ *for all* $F,G\in\mathbb{H}$.

Then, the corresponding statement also holds for all $F,G\in\mathcal{F}$.

Proof. The proofs for the three cases are similar, so we only argue that the third claim holds. Fix $F\in\mathbb{H}$ and define

$$\mathbb{D} = \left\{ G\in\mathcal{F} \,\middle|\, \lim_{n\to\infty} P(F\cap T^{-n}(G)) = P(F)P(G) \right\}.$$

We wish to argue that $\mathbb{D}$ is a Dynkin class. To this end, note that since $T^{-1}(\Omega)=\Omega$, it holds that $\Omega\in\mathbb{D}$. Take $A,B\in\mathbb{D}$ with $A\subseteq B$. We then also have $T^{-n}(A)\subseteq T^{-n}(B)$, yielding

$$\begin{aligned}\lim_{n\to\infty} P(F\cap T^{-n}(B\setminus A)) &= \lim_{n\to\infty} P(F\cap (T^{-n}(B)\setminus T^{-n}(A)))\\ &= \lim_{n\to\infty} P(F\cap T^{-n}(B)) - \lim_{n\to\infty} P(F\cap T^{-n}(A))\\ &= P(F)P(B) - P(F)P(A) = P(F)P(B\setminus A),\end{aligned}$$

and so $B\setminus A\in\mathbb{D}$. Finally, let (A_n) be an increasing sequence in $\mathbb{D}$ and let $A=\cup_{n=1}^\infty A_n$. As $\lim_{m\to\infty} P(A_m)=P(A)$, we obtain $\lim_{m\to\infty} P(A\setminus A_m)=0$. Pick $\varepsilon>0$ and let m be such that for $i\geq m$, $P(A\setminus A_i)\leq\varepsilon$. Note that $T^{-n}(A_i)\subseteq T^{-n}(A)$. As T is measure preserving, we obtain for all $n\geq 1$ and $i\geq m$ that

$$\begin{aligned}0 &\leq P(F\cap T^{-n}(A)) - P(F\cap T^{-n}(A_i))\\ &= P(F\cap (T^{-n}(A)\setminus T^{-n}(A_i))) = P(F\cap (T^{-n}(A\setminus A_i)))\\ &\leq P(T^{-n}(A\setminus A_i)) = P(A\setminus A_i)\leq\varepsilon.\end{aligned}$$

From this we find that for all $n\geq 1$ and $i\geq m$,

$$P(F\cap T^{-n}(A_i)) - \varepsilon \leq P(F\cap T^{-n}(A)) \leq P(F\cap T^{-n}(A_i)) + \varepsilon,$$

and therefore, for $i\geq m$,

$$\begin{aligned}P(F)P(A_i)-\varepsilon &= \lim_{n\to\infty} P(F\cap T^{-n}(A_i)) - \varepsilon \leq \liminf_{n\to\infty} P(F\cap T^{-n}(A))\\ &\leq \limsup_{n\to\infty} P(F\cap T^{-n}(A)) \leq \lim_{n\to\infty} P(F\cap T^{-n}(A_i)) + \varepsilon\\ &= P(F)P(A_i)+\varepsilon. \end{aligned} \tag{2.4}$$

As (2.4) holds for all $i \geq m$, we in particular conclude that

$$\begin{aligned} P(F)P(A) - \varepsilon &= \lim_{i\to\infty} P(F)P(A_i) - \varepsilon \leq \liminf_{n\to\infty} P(F \cap T^{-n}(A)) \\ &\leq \limsup_{n\to\infty} P(F \cap T^{-n}(A)) = \lim_{i\to\infty} P(F)P(A_i) + \varepsilon \\ &= P(F)P(A) + \varepsilon. \end{aligned} \tag{2.5}$$

And as $\varepsilon > 0$ is arbitrary in (2.5), we conclude $\lim_{n\to\infty} P(F \cap T^{-n}(A)) = P(F)P(A)$, so that $A \in \mathbb{D}$. Thus, $\mathbb{D}$ is a Dynkin class. By assumption, $\mathbb{D}$ contains $\mathbb{H}$, and therefore, by Dynkin's lemma, $\mathcal{F} = \sigma(\mathbb{H}) \subseteq \mathbb{D}$. This proves $\lim_{n\to\infty} P(F \cap T^{-n}(G)) = P(F)P(G)$ when $F \in \mathbb{H}$ and $G \in \mathcal{F}$. Next, we extend this to all $F \in \mathcal{F}$. To do so, fix $G \in \mathcal{F}$ and define

$$\mathbb{E} = \left\{ F \in \mathcal{F} \,\middle|\, \lim_{n\to\infty} P(F \cap T^{-n}(G)) = P(F)P(G) \right\}.$$

Similarly to our earlier arguments, we find that $\mathbb{E}$ is a Dynkin class. As $\mathbb{E}$ contains $\mathbb{H}$, Dynkin's lemma yields $\mathcal{F} = \sigma(\mathbb{H}) \subseteq \mathbb{E}$. This shows $\lim_{n\to\infty} P(F \cap T^{-n}(G)) = P(F)P(G)$ when $F, G \in \mathcal{F}$ and so proves the first claim. By similar arguments, we obtain the two remaining claims. □

Combining Theorem 2.2.5 and Lemma 2.2.6, we find that in order to show ergodicity of T, it suffices to show that T is mixing or weakly mixing for events in a generating system for $\mathcal{F}$ which is stable under finite intersections. This is in several cases a viable method for proving ergodicity.

2.3 Stationary processes and the law of large numbers

We will now apply the results from Section 2.1 and Section 2.2 to obtain laws of large numbers for the class of processes known as stationary processes. In order to do so, we first need to investigate in what sense we can consider the simultaneous distribution of an entire process (X_n). Once we have done so, we will be able to obtain our main results by applying the ergodic theorem to this simultaneous distribution.

The results require some formalism. By $\mathbb{R}^n$ for $n \geq 1$, we denote the n-fold product of $\mathbb{R}$, the set of n-tuples with elements from $\mathbb{R}$. Analogously, we define $\mathbb{R}^\infty$ as the set of all sequences of real numbers, in the sense that $\mathbb{R}^\infty = \{(x_n)_{n\geq 1} \mid x_n \in \mathbb{R} \text{ for all } n \geq 1\}$. Recall that the Borel σ-algebra on $\mathbb{R}^n$, defined as the smallest σ-algebra containing all open sets, also is given as the smallest σ-algebra making all coordinate projections measurable. In analogy with this, we make the following definition of the Borel σ-algebra on $\mathbb{R}^\infty$. By $\hat{X}_n : \mathbb{R}^\infty \to \mathbb{R}$, we denote the n'th coordinate projection of $\mathbb{R}^\infty$, $\hat{X}_n(x) = x_n$, where $x = (x_n)_{n\geq 1}$.

Definition 2.3.1. *The infinite-dimensional Borel σ-algebra, $\mathcal{B}_\infty$, is the smallest σ-algebra making $\hat{X}_n$ measurable for all $n \geq 1$.*

In detail, Definition 2.3.1 states the following. Let $\mathfrak{A}$ be the family of all σ-algebras $\mathcal{G}$ on $\mathbb{R}^\infty$ such that for all $n \geq 1$, $\hat{X}_n$ is $\mathcal{G}$-$\mathcal{B}$ measurable. $\mathcal{B}_\infty$ is then the smallest σ-algebra in the set $\mathfrak{A}$ of σ-algebras, explicitly constructed as $\mathcal{B}_\infty = \cap_{\mathcal{G}\in\mathfrak{A}}\mathcal{G}$.

In the following lemmas, we prove some basic results on the measure space $(\mathbb{R}^\infty, \mathcal{B}_\infty)$. In Lemma 2.3.2, a generating family which is stable under finite intersections is identified, and in Lemma 2.3.3, the mappings which are measurable with respect to $\mathcal{B}_\infty$ are identified. In Lemma 2.3.4, we show how we can apply $\mathcal{B}_\infty$ to describe and work with stochastic processes.

Lemma 2.3.2. *Let $\mathbb{K}$ be a generating family for $\mathcal{B}$ which is stable under finite intersections. Define $\mathbb{H}$ as the family of sets $\{x \in \mathbb{R}^\infty \mid x_1 \in B_1, \ldots, x_n \in B_n\}$, where $n \geq 1$ and $B_1 \in \mathbb{K}, \ldots, B_n \in \mathbb{K}$. $\mathbb{H}$ is then a generating family for $\mathcal{B}_\infty$ which is stable under finite intersections.*

Proof. It is immediate that $\mathbb{H}$ is stable under finite intersections. Note that if F is a set such that $F = \{x \in \mathbb{R}^\infty \mid x_1 \in B_1, \ldots, x_n \in B_n\}$ for some $n \geq 1$ and $B_1 \in \mathbb{K}, \ldots, B_n \in \mathbb{K}$, we also have

$$\begin{aligned}
F &= \{x \in \mathbb{R}^\infty \mid x_1 \in B_1, \ldots, x_n \in B_n\} \\
&= \{x \in \mathbb{R}^\infty \mid \hat{X}_1(x) \in B_1, \ldots, \hat{X}_n(x) \in B_n\} \\
&= \{x \in \mathbb{R}^\infty \mid x \in \hat{X}_1^{-1}(B_1), \ldots, x \in \hat{X}_n^{-1}(B_n)\} \\
&= \cap_{k=1}^n \hat{X}_n^{-1}(B_k).
\end{aligned}$$

Therefore, $\mathbb{H} \subseteq \mathcal{B}_\infty$, and so $\sigma(\mathbb{H}) \subseteq \mathcal{B}_\infty$. It remains to argue that $\mathcal{B}_\infty \subseteq \sigma(\mathbb{H})$. To this end, fix $n \geq 1$ and note that $\mathbb{H}$ contains $\hat{X}_n^{-1}(B)$ for all $B \in \mathbb{K}$. Therefore, $\sigma(\mathbb{H})$ contains $\hat{X}_n^{-1}(B)$ for all $B \in \mathbb{K}$. As $\{B \in \mathcal{B} \mid \hat{X}_n^{-1}(B) \in \sigma(\mathbb{H})\}$ is a σ-algebra which contains $\mathbb{K}$, we conclude that it also contains $\mathcal{B}$. Thus, $\sigma(\mathbb{H})$ contains $\hat{X}_n^{-1}(B)$ for all $B \in \mathcal{B}$, and so $\sigma(\mathbb{H})$ is a σ-algebra on $\mathbb{R}^\infty$ making all coordinate projections measurable. As $\mathcal{B}_\infty$ is the smallest such σ-algebra, we conclude $\mathcal{B}_\infty \subseteq \sigma(\mathbb{H})$. All in all, we obtain $\mathcal{B}_\infty = \sigma(\mathbb{H})$, as desired. □

Lemma 2.3.3. *Let $X : \Omega \to \mathbb{R}^\infty$. X is $\mathcal{F}$-$\mathcal{B}_\infty$ measurable if and only if $\hat{X}_n \circ X$ is $\mathcal{F}$-$\mathcal{B}$ measurable for all $n \geq 1$.*

Proof. First assume that X is $\mathcal{F}$-$\mathcal{B}_\infty$ measurable. As $\hat{X}_n$ is $\mathcal{B}_\infty$-$\mathcal{B}$ measurable by definition, we find that $\hat{X}_n \circ X$ is $\mathcal{F}$-$\mathcal{B}$ measurable. Conversely, assume that $\hat{X}_n \circ X$ is $\mathcal{F}$-$\mathcal{B}$ measurable for all $n \geq 1$, we wish to show that X is $\mathcal{F}$-$\mathcal{B}_\infty$ measurable. To this end, it suffices to show that $X^{-1}(A) \in \mathcal{F}$ for all A in a generating family for $\mathcal{B}_\infty$. Define $\mathbb{H}$ by putting $\mathbb{H} = \{\hat{X}_n^{-1}(B) \mid n \geq 1, B \in \mathcal{B}\}$, $\mathbb{H}$ is then a generating family for $\mathcal{B}_\infty$. For any $n \geq 1$ and $B \in \mathcal{B}$, we have $X^{-1}(\hat{X}_n^{-1}(B)) = (\hat{X}_n \circ X)^{-1}(B) \in \mathcal{F}$ by our assumptions. Thus, X is $\mathcal{F}$-$\mathcal{B}_\infty$ measurable, as was to be proven. □

Lemma 2.3.4. *Let (X_n) be a stochastic process. Defining a mapping $X : \Omega \to \mathbb{R}^\infty$ by putting $X(\omega) = (X_n(\omega))_{n\geq 1}$, it holds that X is $\mathcal{F}$-$\mathcal{B}_\infty$ measurable.*

Proof. As $\hat{X}_n \circ X = X_n$ and X_n is $\mathcal{F}$-$\mathcal{B}$ measurable by assumption, the result follows from Lemma 2.3.3. □

Letting $(X_n)_{n\geq 1}$ be a stochastic process, Lemma 2.3.4 shows that with $X : \Omega \to \mathbb{R}^\infty$ defined by $X(\omega) = (X_n(\omega))_{n\geq 1}$, X is $\mathcal{F}$-$\mathcal{B}_\infty$ measurable, and therefore, the image measure $X(P)$ is well-defined. This motivates the following definition of the distribution of a stochastic process.

Definition 2.3.5. *Letting $(X_n)_{n\geq 1}$ be a stochastic process. The distribution of $(X_n)_{n\geq 1}$ is the probability measure $X(P)$ on $\mathcal{B}_\infty$.*

Utilizing the above definitions and results, we can now state our plan for the main results to be shown later in this section. Recall that one of our goals for this section is to prove an extension of the law of large numbers. The method we will apply is the following. Consider a stochastic process (X_n). The introduction of the infinite-dimensional Borel-σ-algebra and the measurability result in Lemma 2.3.4 have allowed us in Definition 2.3.5 to introduce the concept of the distribution of a process. In particular, we have at our disposal a probability space $(\mathbb{R}^\infty, \mathcal{B}_\infty, X(P))$. If we can identify a suitable transformation $T : \mathbb{R}^\infty \to \mathbb{R}^\infty$ such that T is measure preserving and ergodic for $X(P)$, we will be able to apply Theorem 2.1.7 to obtain a type of law of large numbers with $X(P)$ almost sure convergence and convergence in $\mathcal{L}^p(\mathbb{R}^\infty, \mathcal{B}_\infty, X(P))$. If we afterwards succeed in transfering the results from the probability space $(\mathbb{R}^\infty, \mathcal{B}_\infty, X(P))$ back to the probability space $(\Omega, \mathcal{F}, P)$, we will have achieved our goal.

Lemma 2.3.6. *Let (X_n) be a stochastic process. Define $X : \Omega \to \mathbb{R}^\infty$ by $X(\omega) = (X_n(\omega))_{n\geq 1}$. The image measure $X(P)$ is the unique probability measure on $\mathcal{B}_\infty$ such that for all $n \geq 1$*

and all $B_1 \in \mathcal{B}, \ldots, B_n \in \mathcal{B}$, it holds that

$$P(X_1 \in B_1, \ldots, X_n \in B_n) = X(P)(\cap_{k=1}^n \hat{X}_k^{-1}(B_k)). \tag{2.6}$$

Proof. Uniqueness follows from Lemma 2.3.2 and the uniqueness theorem for probability measures. It remains to show that $X(P)$ satisfies (2.6). To this end, we note that

$$\begin{aligned} X(P)(\cap_{k=1}^n \hat{X}_k^{-1}(B_k)) &= P(X^{-1}(\cap_{k=1}^n \hat{X}_k^{-1}(B_k))) = P(\cap_{k=1}^n X^{-1}(\hat{X}_k^{-1}(B_k))) \\ &= P(\cap_{k=1}^n (\hat{X}_k \circ X)^{-1}(B_k))) = P(\cap_{k=1}^n (X_k \in B_k)) \\ &= P(X_1 \in B_1, \ldots, X_n \in B_n). \end{aligned}$$

This completes the proof. □

Lemma 2.3.6 may appear rather abstract at a first glance. A clearer statement might be obtained by noting that $\cap_{k=1}^n \hat{X}_k^{-1}(B_k) = B_1 \times \cdots \times B_k \times \mathbb{R}^\infty$. The lemma then states that the distribution $X(P)$ is the only probability measure on $\mathcal{B}_\infty$ such that the $X(P)$-measure of a "finite-dimensional rectangle" of the form $B_1 \times \cdots \times B_k \times \mathbb{R}^\infty$ has the same measure as $P(X_1 \in B_1, \ldots, X_n \in B_n)$, a property reminiscent of the characterizing feature of the distribution of an ordinary finite-dimensional random variable.

Using the above, we may now formalize the notion of a stationary process. First, we define $\theta : \mathbb{R}^\infty \to \mathbb{R}^\infty$ by putting $\theta((x_n)_{n\geq 1}) = (x_{n+1})_{n\geq 1}$. We refer to θ as the shift operator. Note that by Lemma 2.3.3, θ is $\mathcal{B}_\infty$-$\mathcal{B}_\infty$ measurable. The mapping θ will play the role of the measure preserving and ergodic transformation in our later use of Theorem 2.1.7.

Definition 2.3.7. *Let (X_n) be a stochastic process. We say that (X_n) is a stationary process, or simply stationary, if it holds that θ is measure preserving for the distribution of (X_n). We say that a stationary process is ergodic if θ is ergodic for the distribution of (X_n).*

According to Definition 2.3.7, the property of being stationary is related to the measure preservation property of the mapping θ on $\mathcal{B}_\infty$ in relation to the measure $X(P)$ on $\mathcal{B}_\infty$, and the property of being ergodic is related to the invariant σ-algebra of θ, which is a sub-σ-algebra of $\mathcal{B}_\infty$. It is these conceptions of stationarity and ergodicity we will be using when formulating our laws of large numbers. However, for practical use, it is convenient to be able to express stationarity and ergodicity in terms of the probability space $(\Omega, \mathcal{F}, P)$ instead of $(\mathbb{R}^\infty, \mathcal{B}_\infty, X(P))$. The following results will allow us to do so.

Lemma 2.3.8. *Let (X_n) be a stochastic process. The following are equivalent.*

(1). (X_n) *is stationary.*

(2). $(X_n)_{n\geq 1}$ *and* $(X_{n+1})_{n\geq 1}$ *have the same distribution.*

(3). For all $k \geq 1$, $(X_n)_{n\geq 1}$ *and* $(X_{n+k})_{n\geq 1}$ *have the same distribution.*

Proof. We first prove that (1) implies (3). Assume that (X_n) is stationary and fix $k \geq 1$. Define a process Y by setting $Y = (X_{n+k})_{n\geq 1}$, we then also have $Y = \theta^k \circ X$. As (X_n) is stationary, θ is $X(P)$-measure preserving. By an application of Theorem A.2.13, this yields $Y(P) = (\theta^k \circ X)(P) = \theta^k(X(P)) = X(P)$, showing that $(X_n)_{n\geq 1}$ and $(X_{n+k})_{n\geq 1}$ have the same distribution, and so proving that (1) implies (3).

As it is immediate that (3) implies (2), we find that in order to complete the proof, it suffices to show that (2) implies (1). Therefore, assume that $(X_n)_{n\geq 1}$ and $(X_{n+1})_{n\geq 1}$ have the same distribution, meaning that $X(P)$ and $Y(P)$ are equal, where $Y = (X_{n+1})_{n\geq 1}$. We then obtain $\theta(X(P)) = (\theta \circ X)(P) = Y(P) = X(P)$, so θ is $X(P)$-measure preserving. This proves that (2) implies (1), as desired. □

An important consequence of Lemma 2.3.8 is the following.

Lemma 2.3.9. *Let* (X_n) *be a stationary stochastic process. For all* $k \geq 1$ *and* $n \geq 1$, $(X_1, \ldots, X_n)$ *has the same distribution as* $(X_{1+k}, \ldots, X_{n+k})$.

Proof. Fix $k \geq 1$ and $n \geq 1$. Let $Y = (X_{n+k})_{n\geq 1}$. By Lemma 2.3.8, it holds that $(X_n)_{n\geq 1}$ and $(Y_n)_{n\geq 1}$ have the same distribution. Let $\varphi : \mathbb{R}^\infty \to \mathbb{R}^n$ denote the projection onto the first n coordinates of $\mathbb{R}^\infty$. Using Theorem A.2.13, we then obtain

$$\begin{aligned}(X_1, \ldots, X_n)(P) &= \varphi(X)(P) = \varphi(X(P)) = \varphi(Y(P)) \\ &= \varphi(Y)(P) = (Y_1, \ldots, Y_n)(P) = (X_{1+k}, \ldots, X_{n+k})(P),\end{aligned}$$

proving that $(X_1, \ldots, X_n)$ has the same distribution as $(X_{1+k}, \ldots, X_{n+k})$, as was to be shown. □

Next, we consider a more convenient formulation of ergodicity for a stationary process.

Definition 2.3.10. *Let* (X_n) *be a stationary process. The invariant* σ*-algebra* $\mathcal{I}(X)$ *for the process is defined by* $\mathcal{I}(X) = \{X^{-1}(B) \mid B \in \mathcal{B}_\infty, B \text{ is invariant for } \theta\}$.

Lemma 2.3.11. *Let (X_n) be a stationary process. (X_n) is ergodic if and only if it holds that for all $F \in \mathcal{I}(X)$, $P(F)$ is either zero or one.*

Proof. First assume that (X_n) is ergodic, meaning that θ is ergodic for $X(P)$. This means that with $\mathcal{I}_\theta$ denoting the invariant σ-algebra for θ on $\mathcal{B}_\infty$, $X(P)(B)$ is either zero or one for all $B \in \mathcal{I}_\theta$. Now let $F \in \mathcal{I}(X)$, we then have $F = (X \in B)$ for some $B \in \mathcal{I}_\theta$, so we obtain $P(F) = P(X \in B) = X(P)(B)$, which is either zero or one. This proves the first implication. Next, assume that for all $F \in \mathcal{I}(X)$, $P(F)$ is either zero or one. We wish to show that (X_n) is ergodic. Let $B \in \mathcal{I}_\theta$. We then obtain $X(P)(B) = P(X^{-1}(B))$, which is either zero or one as $X^{-1}(B) \in \mathcal{I}(X)$. Thus, (X_n) is ergodic. □

Lemma 2.3.8 and Lemma 2.3.11 shows how to reformulate the definitions in Definition 2.3.7 more concretely in terms of the probability space $(\Omega, \mathcal{F}, P)$ and the process (X_n). We are now ready to use the ergodic theorem to obtain a law of large numbers for stationary processes.

Theorem 2.3.12 (Ergodic theorem for ergodic stationary processes)**.** *Let (X_n) be an ergodic stationary process, and let $f : \mathbb{R}^\infty \to \mathbb{R}$ be some $\mathcal{B}_\infty$-$\mathcal{B}$ measurable mapping. If $f((X_n)_{n\geq 1})$ has p'th moment, $\frac{1}{n}\sum_{k=1}^n f((X_i)_{i\geq k})$ converges almost surely and in $\mathcal{L}^p$ to $Ef((X_i)_{i\geq 1})$.*

Proof. We first investigate what may be obtained by using the ordinary ergodic theorem of Theorem 2.1.7. Let $\hat{P} = X(P)$, the distribution of (X_n). By our assumptions, θ is $\hat{P}$-measure preserving and ergodic. Also, f is a random variable on the probability space $(\mathbb{R}^\infty, \mathcal{B}_\infty, \hat{P})$, and

$$\int |f|^p \,\mathrm{d}\hat{P} = \int |f|^p \,\mathrm{d}X(P) = \int |f \circ X|^p \,\mathrm{d}P = E|f(X)|^p = E|f((X_n)_{n\geq 1})|^p,$$

which is finite by our assumptions. Thus, considered as a random variable on $(\mathbb{R}^\infty, \mathcal{B}_\infty, \hat{P})$, f has p'th moment. Letting $\mu = Ef((X_n)_{n\geq 1})$, Theorem 2.1.7 yields

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n f \circ \theta^{k-1} = \int f \,\mathrm{d}\hat{P} = Ef((X_n)_{n\geq 1}) = \mu,$$

in the sense of $\hat{P}$ almost sure convergence and convergence in $\mathcal{L}^p(\mathbb{R}^\infty, \mathcal{B}_\infty, \hat{P})$. These are limit results on the probability space $(\mathbb{R}^\infty, \mathcal{B}_\infty, \hat{P})$. We would like to transfer these results to our original probability space $(\Omega, \mathcal{F}, P)$. We first consider the case of almost sure convergence. We wish to argue that $\frac{1}{n}\sum_{k=1}^n f((X_i)_{i\geq k})$ converges P-almost surely to μ. To do so, first note that

$$\begin{aligned}\left(\frac{1}{n}\sum_{k=1}^n f((X_i)_{i\geq k}) \text{ converges to } \mu\right) &= \left\{\omega\in\Omega \,\middle|\, \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n f((X_i)_{i\geq k}(\omega)) = \mu\right\}\\ &= \left\{\omega\in\Omega \,\middle|\, \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n f((X_i(\omega))_{i\geq k}) = \mu\right\}\\ &= \left\{\omega\in\Omega \,\middle|\, \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n f(\theta^{k-1}(X(\omega))) = \mu\right\},\end{aligned}$$

and this final set is equal to $X^{-1}(A)$, where

$$A = \left\{x\in\mathbb{R}^\infty \,\middle|\, \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n f(\theta^{k-1}(x)) = \mu\right\},$$

or, with our usual probabilistic notation, $A = (\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n f\circ\theta^{k-1} = \mu)$. Therefore, we obtain

$$\begin{aligned}P\left(\frac{1}{n}\sum_{k=1}^n f((X_i)_{i\geq k}) \text{ converges to } \mu\right) &= P(X^{-1}(A)) = X(P)(A)\\ &= \hat{P}(A) = \hat{P}\left(\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^n f\circ\theta^{k-1} = \mu\right),\end{aligned}$$

and the latter is equal to one by the $\hat{P}$-almost sure convergence of $\frac{1}{n}\sum_{k=1}^n f\circ\theta^{k-1}$ to μ. This proves P-almost sure convergence of $\frac{1}{n}\sum_{k=1}^n f((X_i)_{i\geq k})$ to μ. Next, we consider convergence in $\mathcal{L}^p$. Here, we need $\lim_{n\to\infty} E|\mu - \frac{1}{n}\sum_{k=1}^n f((X_i)_{i\geq k})|^p = 0$. To obtain this, we note that for any $\omega\in\Omega$, it holds that

$$\begin{aligned}\left(\left|\mu - \frac{1}{n}\sum_{k=1}^n f((X_i)_{i\geq k})\right|^p\right)(\omega) &= \left|\mu - \frac{1}{n}\sum_{k=1}^n f((X_i(\omega))_{i\geq k})\right|^p\\ &= \left|\mu - \frac{1}{n}\sum_{k=1}^n f(\theta^{k-1}(X(\omega)))\right|^p\\ &= \left(\left|\mu - \frac{1}{n}\sum_{k=1}^n f\circ\theta^{k-1}\right|^p\right)(X(\omega)),\end{aligned}$$

yielding

$$\left|\mu - \frac{1}{n}\sum_{k=1}^n f((X_i)_{i\geq k})\right|^p = \left(\left|\mu - \frac{1}{n}\sum_{k=1}^n f\circ\theta^{k-1}\right|^p\right)\circ X,$$

and so the change-of-variables formula allows us to conclude that

$$
\begin{aligned}
\lim_{n\to\infty} E\left|\mu-\frac{1}{n}\sum_{k=1}^{n} f((X_i)_{i\geq k})\right|^p &= \lim_{n\to\infty}\int\left|\mu-\frac{1}{n}\sum_{k=1}^{n} f((X_i)_{i\geq k})\right|^p \,\mathrm{d}P \\
&= \lim_{n\to\infty}\int\left(\left|\mu-\frac{1}{n}\sum_{k=1}^{n} f((\hat{X}_i)_{i\geq k})\right|^p\right)\circ X\,\mathrm{d}P \\
&= \lim_{n\to\infty}\int\left(\left|\mu-\frac{1}{n}\sum_{k=1}^{n} f((\hat{X}_i)_{i\geq k})\right|^p\right)\mathrm{d}X(P) \\
&= \lim_{n\to\infty}\int\left(\left|\mu-\frac{1}{n}\sum_{k=1}^{n} f((\hat{X}_i)_{i\geq k})\right|^p\right)\mathrm{d}\hat{P} = 0,
\end{aligned}
$$

by the $\mathcal{L}^p(\mathbb{R}^\infty, \mathcal{B}_\infty, \hat{P})$-convergence of $\frac{1}{n}\sum_{k=1}^n f((\hat{X}_i)_{i\geq k})$ to μ. This demonstrates the desired convergence in $\mathcal{L}^p$ and so concludes the proof of the theorem. □

Theorem 2.3.12 is the main theorem of this section. As the following corollary shows, a simpler version of the theorem is obtained by applying the theorem to a particular type of function from $\mathbb{R}^\infty$ to $\mathbb{R}$.

Corollary 2.3.13. *Let (X_n) be an ergodic stationary process, and let $f : \mathbb{R} \to \mathbb{R}$ be some Borel measurable mapping. If $f(X_1)$ has p'th moment, $\frac{1}{n}\sum_{k=1}^n f(X_k)$ converges almost surely and in $\mathcal{L}^p$ to $Ef(X_1)$.*

Proof. Define $g : \mathbb{R}^\infty \to \mathbb{R}$ by putting $g((x_n)_{n\geq 1}) = f(x_1)$. Then $g = f \circ \hat{X}_1$, so as f is $\mathcal{B}$-$\mathcal{B}$ measurable and $\hat{X}_1$ is $\mathcal{B}_\infty$-$\mathcal{B}$ measurable, g is $\mathcal{B}_\infty$-$\mathcal{B}$ measurable. Also, $g((X_i)_{i\geq 1}) = f(X_1)$, which has p'th moment by assumption. Therefore, Theorem 2.3.12 allows us to conclude that $\frac{1}{n}\sum_{k=1}^n g((X_i)_{i\geq k})$ converges almost surely and in $\mathcal{L}^p$ to $Ef(X_1)$. As $g((X_i)_{i\geq k}) = f(X_k)$, this yields the desired conclusion. □

Theorem 2.3.12 and Corollary 2.3.13 yields powerful convergence results for stationary and ergodic processes. Next, we show that our results contain the strong law of large numbers for independent and identically distributed variables as a special case. In addition, we also obtain $\mathcal{L}^p$ convergence of the empirical means. To show this result, we need to prove that sequences of independent and identically distributed variables are stationary and ergodic.

Corollary 2.3.14. *Let (X_n) be a sequence of independent, identically distributed variables. Then (X_n) is stationary and ergodic. Assume furthermore that X_n has p'th moment for some $p \geq 1$, and let μ be the common mean. Then $\frac{1}{n}\sum_{k=1}^n X_k$ converges to μ almost surely and in $\mathcal{L}^p$.*

Proof. We first show that (X_n) is stationary. Let ν denote the common distribution of the X_n. Let $X = (X_n)_{n\geq 1}$ and $Y = (X_{n+1})_{n\geq 1}$. Fix $n \geq 1$ and $B_1, \ldots, B_n \in \mathcal{B}$, we then obtain

$$\begin{aligned} Y(P)(\cap_{k=1}^n \hat{X}_k^{-1}(B_k)) &= P(X_2 \in B_1, \ldots, X_{n+1} \in B_n) = \prod_{i=1}^n \nu(B_i) \\ &= P(X_1 \in B_1, \ldots, X_n \in B_n) = X(P)(\cap_{k=1}^n \hat{X}_k^{-1}(B_k)), \end{aligned}$$

so by Lemma 2.3.2 and the uniqueness theorem for probability measures, we conclude that $Y(P) = X(P)$, and thus (X_n) is stationary. Next, we show that (X_n) is ergodic. Let $\mathcal{I}(X)$ denote the invariant σ-algebra for (X_n), and let $\mathcal{J}$ denote the tail-σ-algebra for (X_n), see Definition 1.3.9. Let $F \in \mathcal{I}(X)$, we then have $F = (X \in B)$ for some $B \in \mathcal{I}_\theta$, where $\mathcal{I}_\theta$ is the invariant σ-algebra on $\mathbb{R}^\infty$ for the shift operator. Therefore, for any $n \geq 1$, we obtain

$$\begin{aligned} (X \in B) &= (X \in \theta^{-n}(B)) = (\theta^n(X) \in B) \\ &= ((X_{n+1}, X_{n+2}, \ldots) \in B) \in \sigma(X_{n+1}, X_{n+2}, \ldots). \end{aligned} \tag{2.7}$$

As n is arbitrary in (2.7), we conclude $(X \in B) \in \mathcal{J}$, and as a consquence, $\mathcal{I}(X) \subseteq \mathcal{J}$. Now recalling from Theorem 1.3.10 that $P(F)$ is either zero or one for all $F \in \mathcal{J}$, we obtain that whenever $F \in \mathcal{I}(X)$, $P(F)$ is either zero or one as well. By Lemma 2.3.11, this shows that (X_n) is ergodic. Corollary 2.3.13 yields the remaining claims of the corollary. □

In order to apply Theorem 2.3.12 and Corollary 2.3.13 in general, we need results on how to prove stationarity and ergodicity. As the final theme of this section, we show two such results.

Lemma 2.3.15. *Let (X_n) be stationary. Assume that for all $m, p \geq 1$, $A_1, \ldots, A_m \in \mathcal{B}$ and $B_1, \ldots, B_p \in \mathcal{B}$:*

(1). With $F = \cap_{i=1}^m (X_i \in A_i)$ and $G_k = \cap_{i=1}^p (X_{i+k-1} \in B_i)$ for $k \geq 1$, it holds that $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n P(F \cap G_k) = P(F)P(G_1)$.

(2). With $F = \cap_{i=1}^m (X_i \in A_i)$ and $G_k = \cap_{i=1}^p (X_{i+k-1} \in B_i)$ for $k \geq 1$, it holds that $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n |P(F \cap G_k) - P(F)P(G_1)| = 0$.

(3). With $F = \cap_{i=1}^m (X_i \in A_i)$ and $G_n = \cap_{i=1}^p (X_{i+n} \in B_i)$ for $n \geq 1$, it holds that $\lim_{n\to\infty} P(F \cap G_n) = P(F)P(G_1)$.

Then (X_n) is ergodic.

Proof. We only prove the result in the case where the third convergence holds, as the other two cases follow similarly. Therefore, assume that the first criterion holds, such that for all $m, p \geq 1$, $A_1, \ldots, A_m \in \mathcal{B}$ and $B_1, \ldots, B_p \in \mathcal{B}$, it holds that

$$\lim_{n\to\infty} P(\cap_{i=1}^m (X_i \in A_i) \cap \cap_{i=1}^p (X_{i+n} \in B_i)) = P(\cap_{i=1}^m (X_i \in A_i))P(\cap_{i=1}^p (X_i \in B_i)). \quad (2.8)$$

We wish to show that (X_n) is ergodic. Recall that by Definition 2.3.7 that since (X_n) is stationary, θ is measure preserving for $\hat{P}$, where $\hat{P} = X(P)$. Also recall from Definition 2.3.7 that in order to show that (X_n) is ergodic, we must show that θ is ergodic for $\hat{P}$. We will apply Lemma 2.2.6 and Theorem 2.2.5 to the probability space $(\mathbb{R}^\infty, \mathcal{B}_\infty, \hat{P})$ and the transformation θ. Note that as θ is measure preserving for $\hat{P}$, Lemma 2.2.6 and Theorem 2.2.5 are applicable.

Define $\mathbb{H}$ as the family of sets $\{x \in \mathbb{R}^\infty \mid x_1 \in B_1, \ldots, x_n \in B_n\}$, where $n \geq 1$ and $B_1 \in \mathcal{B}, \ldots, B_n \in \mathcal{B}$. By Lemma 2.3.2, $\mathbb{H}$ is then a generating family for $\mathcal{B}_\infty$ which is stable under finite intersections. By Lemma 2.2.6 and Theorem 2.2.5, θ is ergodic for $\hat{P}$ if it holds that for all $F, G \in \mathbb{H}$,

$$\lim_{n\to\infty} \hat{P}(F \cap \theta^{-n}(G)) = \hat{P}(F)\hat{P}(G). \quad (2.9)$$

However, for any $F, G \in \mathbb{H}$, we have that there is $m, p \geq 1$ such that $F = \cap_{i=1}^m \hat{X}_i^{-1}(A_i)$ and $G = \cap_{i=1}^p \hat{X}_i^{-1}(B_i)$, and so

$$\begin{aligned}
\hat{P}(F \cap \theta^{-n}(G)) &= X(P)(\cap_{i=1}^m \hat{X}_i^{-1}(A_i) \cap \theta^{-n}(\cap_{i=1}^p \hat{X}_i^{-1}(B_i))) \\
&= X(P)(\cap_{i=1}^m \hat{X}_i^{-1}(A_i) \cap \cap_{i=1}^p \hat{X}_{i+n}^{-1}(B_i)) \\
&= P(\cap_{i=1}^m X_i^{-1}(A_i) \cap \cap_{i=1}^p X_{i+n}^{-1}(B_i)) \\
&= P(\cap_{i=1}^m (X_i \in A_i) \cap \cap_{i=1}^p (X_{i+n} \in B_i)),
\end{aligned}$$

and similarly, we obtain $\hat{P}(F) = P(\cap_{i=1}^m (X_i \in A_i))$ and $\hat{P}(G) = P(\cap_{i=1}^p (X_i \in A_i))$. Thus, for $F, G \in \mathbb{H}$ with $F = \cap_{i=1}^m \hat{X}_i^{-1}(A_i)$ and $G = \cap_{i=1}^p \hat{X}_i^{-1}(B_i)$, (2.9) is equivalent (2.8). As we have assumed that (2.8) holds for all $m, p \geq 1$, $A_1, \ldots, A_m \in \mathcal{B}$ and $B_1, \ldots, B_p \in \mathcal{B}_m$, we conclude that (2.9) holds for all $F, G \in \mathbb{H}$. Lemma 2.2.6 then allows us to conclude that (2.9) holds for all $F, G \in \mathcal{B}_\infty$, and Theorem 2.2.5 then allows us to conclude that θ is ergodic for $\hat{P}$, so that (X_n) is ergodic, as desired. □

Lemma 2.3.16. *Let (X_n) be a sequence of random variables. Let $\varphi : \mathbb{R}^\infty \to \mathbb{R}$ be measurable, and define a sequence of random variables (Y_n) by putting $Y_n = \varphi(X_n, X_{n+1}, \ldots)$. If (X_n) is stationary, then (Y_n) is stationary. And if (X_n) is both stationary and ergodic, then (Y_n) is both stationary and ergodic.*

Proof. We first derive a formal expression for the sequence (Y_n) in terms of (X_n). Define a mapping $\Phi : \mathbb{R}^\infty \to \mathbb{R}^\infty$ by putting, for $k \geq 1$, $\Phi((x_i)_{i\geq 1})_k = \varphi((x_i)_{i\geq k})$. Equivalently, we also have $\Phi((x_i)_{i\geq 1})_k = (\varphi \circ \theta^{k-1})((x_i)_{i\geq 1})$. As θ is $\mathcal{B}_\infty$-$\mathcal{B}_\infty$ measurable by Lemma 2.3.3 and φ is $\mathcal{B}_\infty$-$\mathcal{B}$ measurable, Φ has $\mathcal{B}_\infty$ measurable coordinates, and so is $\mathcal{B}_\infty$-$\mathcal{B}_\infty$ measurable, again by Lemma 2.3.3. And we have $(Y_n) = \Phi((X_n)_{n\geq 1})$.

Now assume that (X_n) is stationary. Let $\hat{P}$ be the distribution of (X_n), and let $\hat{Q}$ be the distribution of (Y_n). By Definition 2.3.7, our assumption that (X_n) is stationary means that θ is measure preserving for $\hat{P}$, and in order to show that (Y_n) is stationary, we must show that θ is measure preserving for $\hat{Q}$. To do so, we note that for all $k \geq 1$, it holds that

$$\theta(\Phi((x_i)_{i\geq 1}))_k = \Phi((x_i)_{i\geq 1})_{k+1} = \varphi(\theta^k((x_i)_{i\geq 1})) = \varphi(\theta^{k-1}(\theta((x_i)_{i\geq 1}))) = \Phi(\theta((x_i)_{i\geq 1}))_k,$$

which means that $\theta \circ \Phi = \Phi \circ \theta$, and so, since θ is measure preserving for $\hat{P}$,

$$\theta(\hat{Q}) = \theta(\Phi(\hat{P})) = (\theta \circ \Phi)(\hat{P}) = (\Phi \circ \theta)(\hat{P}) = \Phi(\hat{P}) = \hat{Q},$$

proving that θ also is measure preserving for $\hat{Q}$, so (Y_n) is stationary. Next, assume that (X_n) is ergodic. By Definition 2.3.7, this means that all elements of the invariant σ-algebra $\mathcal{I}_\theta$ of θ has $\hat{P}$ measure zero or one. We wish to show that (Y_n) is ergodic, which means that we need to show that all elements of $\mathcal{I}_\theta$ has $\hat{Q}$ measure zero or one. Let $A \in \mathcal{I}_\theta$, such that $\theta^{-1}(A) = A$. We then have $\hat{Q}(A) = \hat{P}(\Phi^{-1}(A))$, so it suffices to show that $\Phi^{-1}(A)$ is invariant for θ, and this follows as

$$\theta^{-1}(\Phi^{-1}(A)) = (\Phi \circ \theta)^{-1}(A) = (\theta \circ \Phi)^{-1}(A) = \Phi^{-1}(\theta^{-1}(A)) = \Phi^{-1}(A).$$

Thus, $\Phi^{-1}(A)$ is invariant for θ. As θ is ergodic for $\hat{P}$, $\hat{P}(\Phi^{-1}(A))$ is either zero or one, and so $\hat{Q}(A)$ is either zero or one. Therefore, θ is ergodic for $\hat{Q}$. This shows that (Y_n) is ergodic, as desired. □

We end the section with an example showing how to apply the ergodic theorem to obtain limit results for empirical averages for a practical case of a process consisting of variables which are not independent.

Example 2.3.17. Let (X_n) be a sequence of independent and identically distributed variables concentrated on $\{0, 1\}$ with $P(X_n = 1) = p$. The elements of the sequence $(X_n X_{n+1})_{n\geq 1}$ then have the same distribution for each $n \geq 1$, but they are not independent. We will use the results of this section to examine the behaviour of $\frac{1}{n}\sum_{k=1}^n X_k X_{k+1}$. By Corollary 2.3.14, (X_n) is stationary and ergodic. Define a mapping $f : \mathbb{R}^\infty \to \mathbb{R}$ by putting $f((x_n)_{n\geq 1}) = x_1 x_2$, f is then $\mathcal{B}_\infty$-$\mathcal{B}$ measurable, and $f((X_i)_{i\geq 1}) = X_1 X_2$. Noting that $EX_1X_2 = p^2$ and that X_1X_2 has moments of all orders, Theorem 2.3.12 shows that $\frac{1}{n}\sum_{k=1}^n X_kX_{k+1}$ converges to p^2 almost surely and in $\mathcal{L}^p$ for all $p \geq 1$. ∘

2.4 Exercises

Exercise 2.1. Consider the probability space $([0,1), \mathcal{B}_{[0,1)}, P)$ where P is the Lebesgue measure. Define $T(x) = 2x - [2x]$ and $S(x) = x + \lambda - [x+\lambda]$, $\lambda \in \mathbb{R}$. Here, $[x]$ is the unique integer satisfying $[x] \leq x < [x] + 1$. Show that T and S are P-measure preserving. ∘

Exercise 2.2. Define $T : [0,1) \to [0,1)$ by letting $T(x) = \frac{1}{x} - [\frac{1}{x}]$ for $x > 0$ and zero otherwise. Show that T is Borel measurable. Define P as the nonnegative measure on $([0,1), \mathcal{B}_{[0,1)})$ with density $t \mapsto \frac{1}{\log 2}\frac{1}{1+t}$ with respect to the Lebesgue measure. Show that P is a probability measure, and show that T is measure preserving for P. ∘

Exercise 2.3. Define $T : [0,1] \to [0,1]$ by putting $T(x) = \frac{1}{2}x$ for $x > 0$ and one otherwise. Show that there is no probability measure P on $([0,1], \mathcal{B}_{[0,1]})$ such that T is measure preserving for P. ∘

Exercise 2.4. Consider the probability space $([0,1), \mathcal{B}_{[0,1)}, P)$ where P is the Lebesgue measure. Define $T : [0,1) \to [0,1)$ by $T(x) = x + \lambda - [x+\lambda]$. T is then P-measure preserving. Show that if λ is rational, T is not ergodic. ∘

Exercise 2.5. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let T be measure preserving. Let X be an integrable random variable and assume that $X \circ T \leq X$ almost surely. Show that $X = X \circ T$ almost surely. ∘

Exercise 2.6. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $T : \Omega \to \Omega$ be measurable. Assume that T is measure preserving. Show that if T^2 is ergodic, T is ergodic as well. ∘

Exercise 2.7. Give an example of a probability space $(\Omega, \mathcal{F}, P)$ and a measurable mapping $T : \Omega \to \Omega$ such that T^2 is measure preserving but T is not measure preserving. ∘

Exercise 2.8. Let $(\Omega, \mathcal{F}, P)$ be a probability space and let T be measurable and measure preserving. We may then think of T as a random variable with values in $(\Omega, \mathcal{F})$. Let $F \in \mathcal{F}$.

(1). Show that $(T^n \in F \text{ i.o.})$ is invariant.

(2). Show that for $n \geq 1$, $(\cup_{k=0}^{\infty} T^{-k}(F)) \setminus (\cup_{k=n}^{\infty} T^{-k}(F))$ is a null set.

(3). Show that $F \cap (T^k \in F^c \text{ evt.})$ is a null set.

(4). Assume that $P(F) > 0$. Show that if T is ergodic, $P(T^n \in F \text{ i.o.}) = 1$.

∘

Exercise 2.9. Let $(\Omega, \mathcal{F})$ be a measurable space and let $T : \Omega \to \Omega$ be measurable. Assume that T is measure preserving. Show that the mapping T is ergodic if and only if it holds for all random variables X and Y such that X is integrable and Y is bounded that $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} EY(X \circ T^{k-1}) = (EY)(EX)$. ∘

Exercise 2.10. Consider the probability space $([0,1), \mathcal{B}_{[0,1)}, P)$ where P is the Lebesgue measure. Define $T : [0,1) \to [0,1)$ by $T(x) = 2x - [2x]$. T is then P-measure preserving. Show that T is mixing. ∘

Exercise 2.11. Let $(\Omega_1, \mathcal{F}_1, P_1)$ and $(\Omega_2, \mathcal{F}_2, P_2)$ be two probability spaces. Consider two measurable mappings $T_1 : \Omega_1 \to \Omega_1$ and $T_2 : \Omega_2 \to \Omega_2$. Assume that T_1 is P_1-measure preserving and that T_2 is P_2-measure preserving. Define a probability space $(\Omega, \mathcal{F}, P)$ by putting $(\Omega, \mathcal{F}, P) = (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, P_1 \otimes P_2)$. Define a mapping $T : \Omega \to \Omega$ by putting $T(\omega_1, \omega_2) = (T_1(\omega_1), T_2(\omega_2))$.

(1). Show that T is P-measure preserving.

(2). Let $\mathcal{I}_{T_1}$, $\mathcal{I}_{T_2}$ and $\mathcal{I}_T$ be the invariant σ-algebras for T_1, T_2 and T. Show that the inclusion $\mathcal{I}_{T_1} \otimes \mathcal{I}_{T_2} \subseteq \mathcal{I}_T$ holds.

(3). Argue that if T is ergodic, both T_1 and T_2 are ergodic.

(4). Argue that T is mixing if and only if both T_1 and T_2 are mixing.

∘

Exercise 2.12. Let (X_n) be a stationary process. Fix $B \in \mathcal{B}$. Show that $(X_n \in B \text{ i.o.})$ is in $\mathcal{I}(X)$. ∘

Exercise 2.13. Let (X_n) and (Y_n) be two stationary processes. Let U be a random variable concentrated on $\{0,1\}$ with $P(U = 1) = p$, and assume that U is independent of X and independent of Y. Define $Z_n = X_n 1_{(U=0)} + Y_n 1_{(U=1)}$. Show that (Z_n) is stationary. ∘

Exercise 2.14. We say that a process (X_n) is weakly stationary if it holds that X_n has second moment for all $n \geq 1$, $EX_n = EX_k$ for all $n, k \geq 1$ and $\text{Cov}(X_n, X_k) = \gamma(|n-k|)$ for some $\gamma : \mathbb{N}_0 \to \mathbb{R}$. Now assume that (X_n) is some process such that X_n has second moment for all $n \geq 1$. Show that if (X_n) is stationary, (X_n) is weakly stationary. ◦

Exercise 2.15. We say that a process (X_n) is Gaussian if all of its finite-dimensional distributions are Gaussian. Let (X_n) be some Gaussian process. Show that (X_n) is stationary if and only if it is weakly stationary in the sense of Exercise 2.14. ◦

Chapter 3

Weak convergence

In Chapter 1, in Definition 1.2.2, we introduced four modes of convergence for a sequence of random variables: Convergence in probability, almost sure convergence, convergence in $\mathcal{L}^p$ and convergence in distribution. Throughout most of the chapter, we concerned ourselves solely with the first three modes of convergence. In this chapter, we instead focus on convergence in distribution and the related notion of weak convergence of probability distributions.

While our main results in Chapter 1 and Chapter 2 were centered around almost sure and $\mathcal{L}^p$ convergence of $\frac{1}{n}\sum_{k=1}^n X_k$ for various classes of processes (X_n), the theory of weak convergence covered in this chapter will instead allow us to understand the asymptotic distribution of $\frac{1}{n}\sum_{k=1}^n X_k$, particularly through the combined results of Section 3.5 and Section 3.6.

The chapter is structured as follows. In Section 3.1, we introduce weak convergence of probability measures, and establish that convergence in distribution of random variables and weak convergence of probability measures essentially are the same. In Section 3.2, Section 3.3 and Section 3.4, we investigate the fundamental properties of weak convergence, in the first two sections outlining conncetions with cumulative distribution functions and convergence in probability, and in the third section introducing the characteristic function and proving the major result that weak convergence of probability measures is equivalent to pointwise convergence of characteristic functions.

After this, in Section 3.5, we prove several versions of the central limit theorem which in its simplest form states that under certain regularity conditions, the empirical mean $\frac{1}{n}\sum_{k=1}^n X_k$ of independent and identically distributed random variables can for large n be approximated by a normal distribution with the same mean and variance as $\frac{1}{n}\sum_{k=1}^n X_k$. This is arguably the main result of the chapter, and is a result which is of great significance in practical statistics. In Section 3.6, we introduce the notion of asymptotic normality, which provides a convenient framework for understanding and working with the results of Section 3.5. Finally, in Section 3.7, we state without proof some multidimensional analogues of the results of the previous sections.

3.1 Weak convergence and convergence of measures

Recall from Definition 1.2.2 that for a sequence of random variables (X_n) and another random variable X, we say that X_n converges in distribution to X and write $X_n \xrightarrow{\mathcal{D}} X$ when $\lim_{n\to\infty} Ef(X_n) = Ef(X)$ for all bounded, continuous mappings $f : \mathbb{R} \to \mathbb{R}$. Our first results of this section will show that convergence in distribution of random variables in a certain sense is equivalent to a related mode of convergence for probability measures.

Definition 3.1.1. *Let (μ_n) be a sequence of probability measures on $(\mathbb{R}, \mathcal{B})$, and let μ be another probability measure. We say that μ_n converges weakly to μ and write $\mu_n \xrightarrow{wk} \mu$ if it holds for all bounded, continuous mappings $f : \mathbb{R} \to \mathbb{R}$ that $\lim_{n\to\infty} \int f \,\mathrm{d}\mu_n = \int f \,\mathrm{d}\mu$.*

Lemma 3.1.2. *Let (X_n) be a sequence of random variables and let X be another random variable. Let μ denote the distribution of X, and for $n \geq 1$, let μ_n denote the distribution of X_n. Then $X_n \xrightarrow{\mathcal{D}} X$ if and only if $\mu_n \xrightarrow{wk} \mu$.*

Proof. We have $Ef(X_n) = \int f \circ X_n \,\mathrm{d}P = \int f \,\mathrm{d}X_n(P) = \int f \,\mathrm{d}\mu_n$, and by similar arguments, $Ef(X) = \int f \,\mathrm{d}\mu$. From these observations, the result follows. □

Lemma 3.1.2 clarifies that convergence in distribution of random variables is a mode of convergence depending only on the marginal distributions of the variables involved. In particular, we may investigate the properties of convergence in distribution of random variables by investigating the properties of weak convergence of probability measures on $(\mathbb{R}, \mathcal{B})$. Lemma 3.1.2 also allows us to make sense of convergence of random variables to a probability measure in the following manner: We say that X_n converges in distribution to μ for a sequence of random variables (X_n) and a probability measure μ, and write $X_n \xrightarrow{\mathcal{D}} \mu$, if it holds that $\mu_n \xrightarrow{wk} \mu$, where μ_n is the distribution of X_n.

The topic of weak convergence of probability measures in itself provides ample opportunities for a rich mathematical theory. However, there is good reason for considering both convergence in distribution of random variables and weak convergence of probability measures, in spite of the apparent equivalence of the two concepts: Many results are formulated most naturally in terms of random variables, particularly when transformations of the variables are involved, and furthermore, expressing results in terms of convergence in distribution for random variables often fit better with applications.

In the remainder of this section, we will prove some basic properties of weak convergence of probability measures. Our first interest is to prove that weak limits of probability measures are unique. By $C_b(\mathbb{R})$, we denote the set of bounded, continuous mappings $f : \mathbb{R} \to \mathbb{R}$, and by $C_b^u(\mathbb{R})$, we denote the set of bounded, uniformly continuous mappings $f : \mathbb{R} \to \mathbb{R}$. Note that $C_b^u(\mathbb{R}) \subseteq C_b(\mathbb{R})$.

Lemma 3.1.3. *Assume given two intervals $[a, b] \subseteq (c, d)$. There exists a function $f \in C_b^u(\mathbb{R})$ such that $1_{[a,b]}(x) \leq f(x) \leq 1_{(c,d)}(x)$ for all $x \in \mathbb{R}$.*

Proof. As $[a, b] \subseteq (c, d)$, we have that $a \leq x \leq b$ implies $c < x < d$. In particular, $c < a$ and $b < d$. Then, to obtain the desired mapping, we simply define

$$f(x) = \begin{cases} 0 & \text{for } x \leq c \\ \frac{x-c}{a-c} & \text{for } c < x < a \\ 1 & \text{for } a \leq x \leq b \\ \frac{d-x}{d-b} & \text{for } b < x < d \\ 0 & \text{for } d \leq x \end{cases},$$

and find that f possesses the required properties. □

The mappings whose existence are proved in Lemma 3.1.3 are known as Urysohn functions, and are also occasionally referred to as bump functions, although this latter name in general is reserved for functions which have continuous derivatives of all orders. Existence results of this type often serve to show that continuous functions can be used to approximate other types of functions. Note that if $[a, b] \subseteq (c, d)$ and $1_{[a,b]}(x) \leq f(x) \leq 1_{(c,d)}(x)$ for all $x \in \mathbb{R}$, it then holds for $x \in [a, b]$ that $1 = 1_{[a,b]}(x) \leq f(x) \leq 1_{(c,d)}(x) = 1$, so that $f(x) = 1$. Likewise, for $x \notin (c, d)$, we have $0 = 1_{[a,b]}(x) \leq f(x) \leq 1_{(c,d)}(x) = 0$, so $f(x) = 0$.

In the following lemma, we apply Lemma 3.1.3 to show a useful criterion for two probability measures to be equal, from which we will obtain as an immediate corollary the uniqueness of limits for weak convergence of probability measures.

Lemma 3.1.4. *Let μ and ν be two probability measures on $(\mathbb{R},\mathcal{B})$. If $\int f\,\mathrm{d}\mu = \int f\,\mathrm{d}\nu$ for all $f \in C_b^u(\mathbb{R})$, then $\mu = \nu$.*

Proof. Let μ and ν be probability measures on $(\mathbb{R},\mathcal{B})$ and assume that $\int f\,\mathrm{d}\mu = \int f\,\mathrm{d}\nu$ for all $f \in C_b^u(\mathbb{R})$. By the uniqueness theorem for probability measures, we find that in order to prove that $\mu = \nu$, it suffices to show that $\mu((a,b)) = \nu((a,b))$ for all $a < b$. To this end, let $a < b$ be given. Now pick $n \in \mathbb{N}$ so large that $a + 1/n < b - 1/n$. By Lemma 3.1.3, there then exists a mapping $f_n \in C_b^u(\mathbb{R})$ such that $1_{[a+1/n,b-1/n]} \leq f \leq 1_{(a,b)}$. By our assumptions, we then have $\int f_n\,\mathrm{d}\mu = \int f_n\,\mathrm{d}\nu$.

Now, for $x \notin (a,b)$, we have $x \notin [a+1/n, b-1/n]$ as well, so $f_n(x) = 0$. And for $x \in (a,b)$, it holds that $x \in [a+1/n, b-1/n]$ for n large enough, yielding $f_n(x) = 1$ for n large enough. Thus, $\lim_{n\to\infty} f_n(x) = 1_{(a,b)}(x)$ for all $x \in \mathbb{R}$. By the dominated convergence theorem, we then obtain

$$\mu((a,b)) = \int 1_{(a,b)}\,\mathrm{d}\mu = \lim_{n\to\infty}\int f_n\,\mathrm{d}\mu = \lim_{n\to\infty}\int f_n\,\mathrm{d}\nu = \int 1_{(a,b)}\,\mathrm{d}\nu = \nu((a,b)),$$

and the result follows. □

Lemma 3.1.5. *Let (μ_n) be a sequence of probability measures on $(\mathbb{R},\mathcal{B})$, and let μ and ν be two other such probability measures. If $\mu_n \xrightarrow{wk} \mu$ and $\mu_n \xrightarrow{wk} \nu$, then $\mu = \nu$.*

Proof. For all $f \in C_b(\mathbb{R})$, we obtain $\int f\,\mathrm{d}\nu = \lim_{n\to\infty} f\,\mathrm{d}\mu_n = \int f\,\mathrm{d}\mu$. In particular, this holds for $f \in C_b^u(\mathbb{R})$. Therefore, by Lemma 3.1.4, it holds that $\nu = \mu$. □

Lemma 3.1.5 shows that limits for weak convergence of probability measures are uniquely determined. Note that this is not the case for convergence in distribution of variables. To understand the issue, note that combining Lemma 3.1.2 and Lemma 3.1.5, we find that if $X_n \xrightarrow{\mathcal{D}} X$, then we also have $X_n \xrightarrow{\mathcal{D}} Y$ if and only if X and Y have the same distribution. Thus, for example, if $X_n \xrightarrow{\mathcal{D}} X$, where X is normally distributed with mean zero, then $X_n \xrightarrow{\mathcal{D}} -X$ as well, since X and $-X$ have the same distribution, even though it holds that $P(X = -X) = P(X = 0) = 0$.

In order to show weak convergence of μ_n to μ, we need to prove $\lim_{n\to\infty}\int f\,\mathrm{d}\mu_n = \int f\,\mathrm{d}\mu$ for all $f \in C_b(\mathbb{R})$. A natural question is whether it suffices to prove this limit result for a smaller class of mappings than $C_b(\mathbb{R})$. We now show that it in fact suffices to consider elements of $C_b^u(\mathbb{R})$. For $f : \mathbb{R} \to \mathbb{R}$ bounded, we denote by $\|f\|_\infty$ the uniform norm of f, meaning that

$\|f\|_\infty = \sup_{x\in\mathbb{R}} |f(x)|$. Before obtaining the result, we prove the following useful lemma. Sequences of probability measures satisfying the property (3.1) referred to in the lemma are said to be tight.

Lemma 3.1.6. *Let (μ_n) be a sequence of probability measures on $(\mathbb{R}, \mathcal{B})$, and let μ be some other probability measure. If $\lim_{n\to\infty} \int f \,\mathrm{d}\mu_n = \int f \,\mathrm{d}\mu$ for all $f \in C_b^u(\mathbb{R})$, it holds that*

$$\lim_{M\to\infty} \sup_{n\geq 1} \mu_n([-M, M]^c) = 0. \tag{3.1}$$

In particular, (3.1) holds if μ_n is weakly convergent.

Proof. Fix $\varepsilon > 0$. We will argue that there is $M > 0$ such that $\mu_n([-M, M]^c) \leq \varepsilon$ for $n \geq 1$. To this end, let $M^* > 0$ be so large that $\mu([-M^*/2, M^*/2]^c) < \varepsilon$. By Lemma 3.1.3, we find that there exists a mapping $g \in C_b^u(\mathbb{R})$ with $1_{[-M^*/2,M^*/2]}(x) \leq g(x) \leq 1_{(-M^*,M^*)}(x)$ for $x \in \mathbb{R}$. With $f = 1 - g$, we then also obtain $1_{(-M^*,M^*)^c}(x) \leq f(x) \leq 1_{[-M^*/2,M^*/2]^c}(x)$. As $f \in C_b^u(\mathbb{R})$ as well, this yields

$$\begin{aligned}\limsup_{n\to\infty} \mu_n([-M^*, M^*]^c) &\leq \limsup_{n\to\infty} \int 1_{(-M^*,M^*)^c} \,\mathrm{d}\mu_n \leq \limsup_{n\to\infty} \int f \,\mathrm{d}\mu_n \\ &= \int f \,\mathrm{d}\mu \leq \int 1_{[-M^*/2,M^*/2]^c} \,\mathrm{d}\mu = \mu([-M^*/2, M^*/2]^c) < \varepsilon,\end{aligned}$$

and thus $\mu_n([-M^*, M^*]^c) < \varepsilon$ for n large enough, say $n \geq m$. Now fix $M_1, \ldots, M_m > 0$ such that $\mu_n([-M_n, M_n]^c) < \varepsilon$ for $n \leq m$. Putting $M = \max\{M^*, M_1, \ldots, M_m\}$, we obtain that $\mu_n([-M, M]^c) \leq \varepsilon$ for all $n \geq 1$. This proves (3.1). □

Theorem 3.1.7. *Let (μ_n) be a sequence of probability measures on $(\mathbb{R}, \mathcal{B})$, and let μ be some other probability measure. Then $\mu_n \xrightarrow{wk} \mu$ if and only if $\lim_{n\to\infty} \int f \,\mathrm{d}\mu_n = \int f \,\mathrm{d}\mu$ for $f \in C_b^u(\mathbb{R})$.*

Proof. As $C_b^u(\mathbb{R}) \subseteq C_b(\mathbb{R})$, it is immediate that if $\mu_n \xrightarrow{wk} \mu$, then $\lim_{n\to\infty} \int f \,\mathrm{d}\mu_n = \int f \,\mathrm{d}\mu$ for $f \in C_b^u(\mathbb{R})$. We need to show the converse. Therefore, assume that for $f \in C_b^u(\mathbb{R})$, we have $\lim_{n\to\infty} \int f \,\mathrm{d}\mu_n = \int f \,\mathrm{d}\mu$. We wish to show that $\lim_{n\to\infty} \int f \,\mathrm{d}\mu_n = \int f \,\mathrm{d}\mu$ for all $f \in C_b(\mathbb{R})$.

Using Lemma 3.1.6, take $M > 0$ such that $\mu_n([-M, M]^c) \leq \varepsilon$ for all $n \geq 1$ and such that $\mu([-M, M]^c) \leq \varepsilon$ as well. Note that for any $h \in C_b(\mathbb{R})$ such that $\|h\|_\infty \leq \|f\|_\infty$ and

$f(x) = h(x)$ for $x \in [-M, M]$, we then have

$$\begin{aligned}\left|\int f\,\mathrm{d}\mu - \int h\,\mathrm{d}\mu\right| &= \left|\int f 1_{[-M,M]^c}\,\mathrm{d}\mu - \int h 1_{[-M,M]^c}\,\mathrm{d}\mu\right| \\ &\le \left|\int f 1_{[-M,M]^c}\,\mathrm{d}\mu\right| + \left|\int h 1_{[-M,M]^c}\,\mathrm{d}\mu\right| \\ &\le \|f\|_\infty \mu([-M,M]^c) + \|h\|_\infty \mu([-M,M]^c) \le 2\varepsilon\|f\|_\infty, \qquad (3.2)\end{aligned}$$

and similarly for μ_n instead of μ. To complete the proof, we now take $f \in C_b(\mathbb{R})$, we wish to show $\lim_{n\to\infty} \int f\,\mathrm{d}\mu_n = \int f\,\mathrm{d}\mu$. To this end, we locate $h \in C_b^u(\mathbb{R})$ agreeing with f on $[-M, M]$ with $\|h\|_\infty \le \|f\|_\infty$ and apply (3.2). Define $h : \mathbb{R} \to \mathbb{R}$ by putting

$$h(x) = \begin{cases} f(-M)\exp(M+x) & \text{for } x < -M \\ f(x) & \text{for } -M \le x \le M \\ f(M)\exp(M-x) & \text{for } x > M \end{cases}.$$

Then $\|h\|_\infty \le \|f\|_\infty$. We wish to argue that h is uniformly continuous. Note that as continuous functions are uniformly continuous on compact sets, f is uniformly continuous on $[-M, M]$, and thus h also is uniformly continuous on $[-M, M]$. Furthermore, for $x, y > M$ with $|x - y| \le \delta$, the mean value theorem allows us to obtain

$$|h(x) - h(y)| \le |f(M)||\exp(M-x) - \exp(M-y)| \le |f(M)||x-y|,$$

and similarly, $|h(x) - h(y)| \le |f(-M)||x - y|$ for $x, y < -M$. We conclude that h is a continuous function which is uniformly continuous on $(-\infty, -M)$, on $[-M, M]$ and on (M, ∞). Hence, h is uniformly continuous on $\mathbb{R}$. Furthermore, h agrees with f on $[-M, M]$. Collecting our conclusions, we now obtain by (3.2) that

$$\begin{aligned}\left|\int f\,\mathrm{d}\mu_n - \int f\,\mathrm{d}\mu\right| &\le \left|\int f\,\mathrm{d}\mu_n - \int h\,\mathrm{d}\mu_n\right| + \left|\int h\,\mathrm{d}\mu_n - \int h\,\mathrm{d}\mu\right| + \left|\int h\,\mathrm{d}\mu - \int f\,\mathrm{d}\mu\right| \\ &\le 4\varepsilon\|f\|_\infty + \left|\int h\,\mathrm{d}\mu_n - \int h\,\mathrm{d}\mu\right|,\end{aligned}$$

leading to $\limsup_{n\to\infty} |\int f\,\mathrm{d}\mu_n - \int f\,\mathrm{d}\mu| \le 4\varepsilon\|f\|_\infty$. As $\varepsilon > 0$ was arbitrary, this shows $\lim_{n\to\infty} \int f\,\mathrm{d}\mu_n = \int f\,\mathrm{d}\mu$, proving $\mu_n \xrightarrow{wk} \mu$. □

Before turning to a few examples, we prove some additional basic results on weak convergence. Lemma 3.1.8 and Lemma 3.1.9 give results which occasionally are useful for proving weak convergence.

Lemma 3.1.8. *Let (μ_n) be a sequence of probability measures on $(\mathbb{R}, \mathcal{B})$, and let μ be some other probability measure on $(\mathbb{R}, \mathcal{B})$. Let $h : \mathbb{R} \to \mathbb{R}$ be a continuous mapping. If $\mu_n \xrightarrow{wk} \mu$, it then also holds that $h(\mu_n) \xrightarrow{wk} h(\mu)$.*

Proof. Let $f \in C_b(\mathbb{R})$. Then $f \circ h \in C_b(\mathbb{R})$ as well, and we obtain

$$\lim_{n\to\infty} \int f(x)\,\mathrm{d}h(\mu_n)(x) = \lim_{n\to\infty} \int f(h(x))\,\mathrm{d}\mu_n(x) = \int f(h(x))\,\mathrm{d}\mu(x) = \int f(x)\,\mathrm{d}h(\mu)(x),$$

proving that $h(\mu_n) \xrightarrow{wk} h(\mu)$, as desired. □

Lemma 3.1.9 (Scheffé's lemma)**.** *Let* (μ_n) *be a sequence of probability measures on* $(\mathbb{R}, \mathcal{B})$*, and let* μ *be another probability measure on* $(\mathbb{R}, \mathcal{B})$*. Assume that there is a measure* ν *such that* $\mu_n = g_n \cdot \nu$ *for* $n \geq 1$ *and* $\mu = g \cdot \nu$*. If* $\lim_{n\to\infty} g_n(x) = g(x)$ *for* ν*-almost all* x*, then* $\mu_n \xrightarrow{wk} \mu$*.*

Proof. To prove the result, we first argue that $\lim_{n\to\infty} \int |g_n - g|\,\mathrm{d}\nu = 0$. To this end, with $x^+ = \max\{0, x\}$ and $x^- = \max\{0, -x\}$, we first note that since both μ_n and μ are probability measures, we have

$$0 = \int g_n\,\mathrm{d}\nu - \int g\,\mathrm{d}\nu = \int g_n - g\,\mathrm{d}\nu = \int (g_n - g)^+\,\mathrm{d}\nu - \int (g_n - g)^-\,\mathrm{d}\nu,$$

which implies $\int (g_n - g)^+\,\mathrm{d}\nu = \int (g_n - g)^-\,\mathrm{d}\nu$ and therefore

$$\int |g_n - g|\,\mathrm{d}\nu = \int (g_n - g)^+\,\mathrm{d}\nu + \int (g_n - g)^-\,\mathrm{d}\nu = 2\int (g_n - g)^-\,\mathrm{d}\nu.$$

It therefore suffices to show that this latter tends to zero. To do so, note that

$$(g_n - g)^-(x) = \max\{0, -(g_n(x) - g(x))\} = \max\{0, g(x) - g_n(x)\} \leq g(x), \tag{3.3}$$

and furthermore, since $x \mapsto x^-$ is continuous, $(g_n - g)^-$ converges almost surely to 0. As $0 \leq (g_n - g)^- \leq g$ by (3.3), and g is integrable with respect to ν, the dominated convergence theorem yields $\lim_{n\to\infty} \int (g_n - g)^-\,\mathrm{d}\nu = 0$. Thus, we obtain $\lim_{n\to\infty} \int |g_n - g|\,\mathrm{d}\nu = 0$. In order to obtain the desired weak convergence from this, let $f \in C_b(\mathbb{R})$, we then have

$$\begin{aligned}\limsup_{n\to\infty} \left| \int f(x)\,\mathrm{d}\mu_n(x) - \int f(x)\,\mathrm{d}\mu(x) \right| &\leq \limsup_{n\to\infty} \int |f(x)||g_n(x) - g(x)|\,\mathrm{d}\nu(x) \\ &\leq \|f\|_\infty \limsup_{n\to\infty} \int |g_n(x) - g(x)|\,\mathrm{d}\nu(x) = 0,\end{aligned}$$

proving $\lim_{n\to\infty} \int f\,\mathrm{d}\mu_n = \int f\,\mathrm{d}\mu$ and hence $\mu_n \xrightarrow{wk} \mu$. □

Lemma 3.1.8 shows that weak convergence is conserved under continuous transformations, a result similar in spirit to Lemma 1.2.6. Lemma 3.1.9 shows that for probability measures which have densities with respect to the same common measure, almost sure convergence of the densities is sufficient to obtain weak convergence. This is in several cases a very useful observation.

This concludes our preliminary investigation of weak convergence of probability measures. We end the section with some examples where weak convergence naturally occur.

Example 3.1.10. Let (x_n) be a sequence of real numbers and consider the corresponding Dirac measures (ε_{x_n}), that is, ε_{x_n} is the probability measure which accords probability one to the set $\{x_n\}$ and zero to all Borel subsets in the complement of $\{x_n\}$. We claim that if x_n converges to x for some $x \in \mathbb{R}$, then ε_{x_n} converges weakly to ε_x. To see this, take $f \in C_b(\mathbb{R})$. By continuity, we then have

$$\lim_{n\to\infty} \int f \,\mathrm{d}\varepsilon_{x_n} = \lim_{n\to\infty} f(x_n) = f(x) = \int f \,\mathrm{d}\varepsilon_x,$$

yielding weak convergence of ε_{x_n} to ε_x. ∘

Example 3.1.11. Let μ_n be the uniform distribution on $\{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}\}$. We claim that μ_n converges weakly to the uniform distribution on $[0,1]$. To show this, let $f \in C_b(\mathbb{R})$, we then have $\int f \,\mathrm{d}\mu_n = \frac{1}{n}\sum_{k=1}^n f((k-1)/n)$. Now define a mapping $f_n : [0,1] \to \mathbb{R}$ by putting $f_n(x) = \sum_{k=1}^n f((k-1)/n) 1_{[(k-1)/n, k/n)}(x)$, we then obtain $\int f \,\mathrm{d}\mu_n = \int_0^1 f_n(x)\,\mathrm{d}x$. As f is continuous, we have $\lim_{n\to\infty} f_n(x) = f(x)$ for all $0 \le x < 1$. As f is bounded, the dominated convergence theorem then allows us to conclude that $\lim_{n\to\infty} \int f \,\mathrm{d}\mu_n = \int_0^1 f(x)\,\mathrm{d}x$, which shows that μ_n converges weakly to the uniform distribution on $[0,1]$. ∘

Note that the measures (μ_n) in Example 3.1.11 are discrete in nature, while the limit measure is continuous in nature. This shows that qualities such as being discrete or continuous in nature are not preserved by weak convergence.

Example 3.1.12. Let (ξ_n) and (σ_n) be two real sequences with limits ξ and σ, respectively, where we assume that $\sigma > 0$. Let μ_n be the normal distribution with mean ξ_n and variance σ_n^2. We claim that μ_n converges to μ, where μ denotes the normal distribution with mean ξ and variance σ^2. To demonstrate this result, define mappings g_n for $n \ge 1$ and g by putting $g_n(x) = \frac{1}{\sigma_n\sqrt{2\pi}}\exp(-\frac{1}{2\sigma_n^2}(x-\xi_n)^2)$ and $g(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{1}{2\sigma^2}(x-\xi)^2)$. Then, μ_n has density g_n with respect to the Lebesgue measure, and μ has density g with respcet to the Lebesgue measure. As g_n converges pointwisely to g, Lemma 3.1.9 shows that μ_n converges to μ, as desired. ∘

3.2 Weak convergence and distribution functions

In this section, we investigate the connection between weak convergence of probability measures and convergence of the corresponding cumulative distribution functions. We will show that weak convergence is not in general equivalent to pointwise convergence of cumulative distribution functions, but is equivalent to pointwise convergence on a dense subset of $\mathbb{R}$.

Lemma 3.2.1. *Let (μ_n) be a sequence of probability measures on $(\mathbb{R}, \mathcal{B})$, and let μ be some other probability measure. Assume that μ_n has cumulative distribution function F_n for $n \geq 1$, and assume that μ has cumulative distribution function F. If $\mu_n \stackrel{wk}{\longrightarrow} \mu$, then it holds that $\lim_{n\to\infty} F_n(x) = F(x)$ whenever F is continuous at x.*

Proof. Assume that $\mu_n \stackrel{wk}{\longrightarrow} \mu$ and let x be such that F is continuous at x. Let $\varepsilon > 0$. By Lemma 3.1.3, there exists $h \in C_b(\mathbb{R})$ such that $1_{[x-2\varepsilon,x-\epsilon]}(y) \leq h(y) \leq 1_{(x-3\varepsilon,x)}(y)$ for $y \in \mathbb{R}$. Putting $f(y) = h(y)$ for $y \geq x - \varepsilon$ and $f(y) = 1$ for $y < x - \varepsilon$, we find that $0 \leq f \leq 1$, $f(y) = 1$ for $y \leq x - \varepsilon$ and $f(y) = 0$ for $y > x$. Thus, $1_{(-\infty,x-\varepsilon]}(y) \leq f(y) \leq 1_{(-\infty,x]}(y)$ for $y \in \mathbb{R}$. This implies $F(x-\varepsilon) \leq \int f \,\mathrm{d}\mu$ and $\int f \,\mathrm{d}\mu_n \leq F_n(x)$, from which we conclude

$$F(x-\varepsilon) \leq \int f \,\mathrm{d}\mu = \lim_{n\to\infty} \int f \,\mathrm{d}\mu_n \leq \liminf_{n\to\infty} F_n(x). \tag{3.4}$$

Similarly, there exists $g \in C_b(\mathbb{R})$ such that $0 \leq g \leq 1$, $g(y) = 1$ for $y \leq x$ and $g(y) = 0$ for $y > x + \varepsilon$, implying $F_n(x) \leq \int g \,\mathrm{d}\mu_n$ and $\int g \,\mathrm{d}\mu \leq F(x+\varepsilon)$ and allowing us to obtain

$$\limsup_{n\to\infty} F_n(x) \leq \lim_{n\to\infty} \int g \,\mathrm{d}\mu_n = \int g \,\mathrm{d}\mu \leq F(x+\varepsilon). \tag{3.5}$$

Combining (3.4) and (3.5), we conclude that for all $\varepsilon > 0$, it holds that

$$F(x-\varepsilon) \leq \liminf_{n\to\infty} F_n(x) \leq \limsup_{n\to\infty} F_n(x) \leq F(x+\varepsilon).$$

Since F is continuous at x, we may now let ε tend to zero and obtain that $\liminf_{n\to\infty} F_n(x)$ and $\limsup_{n\to\infty} F_n(x)$ are equal, and the common value is $F(x)$. Therefore, $F_n(x)$ converges and $\lim_{n\to\infty} F_n(x) = F(x)$. This completes the proof. □

The following example shows that in general, weak convergence does not imply convergence of the cumulative distribution functions in all points. After the example, we prove the general result on the correspondence between weak convergence and convergence of cumulative distribution functions.

Example 3.2.2. For each $n \geq 1$, let μ_n be the Dirac measure in $\frac{1}{n}$, and let μ be the Dirac measure at 0. According to Example 3.1.10, $\mu_n \xrightarrow{wk} \mu$. But with F_n being the cumulative distribution function for μ_n and with F being the cumulative distribution function for μ, we have $F_n(0) = 0$ for all $n \geq 1$, while $F(0) = 1$, so that $\lim_{n\to\infty} F_n(0) \neq F(0)$. ◦

Theorem 3.2.3. *Let (μ_n) be a sequence of probability measures on $(\mathbb{R}, \mathcal{B})$, and let μ be some other probability measure. Assume that μ_n has cumulative distribution function F_n for $n \geq 1$, and assume that μ has cumulative distribution function F. Then $\mu_n \xrightarrow{wk} \mu$ if and only if there exists a dense subset A of $\mathbb{R}$ such that $\lim_{n\to\infty} F_n(x) = F(x)$ for $x \in A$.*

Proof. First assume that $\mu_n \xrightarrow{wk} \mu$, we wish to identify a dense subset of $\mathbb{R}$ such that we have pointwise convergence of the cumulative distribution functions on this set. Let B be the set of discontinuity points of F, B is then countable, so B^c is dense in $\mathbb{R}$. By Lemma 3.2.1, $\lim_{n\to\infty} F_n(x) = F(x)$ whenever $x \in B^c$, and so B^c satisfies the requirements.

Next, assume that there exists a dense subset A of $\mathbb{R}$ such that $\lim_{n\to\infty} F_n(x) = F(x)$ for $x \in A$. We wish to show that $\mu_n \xrightarrow{wk} \mu$. To this end, let $f \in C_b(\mathbb{R})$, we need to prove $\lim_{n\to\infty} \int f \,\mathrm{d}\mu_n = \int f \,\mathrm{d}\mu$. Fix $\varepsilon > 0$. Recall that $F(x)$ tends to zero and one as x tends to minus infinity and infinity, respectively. Therefore, we may find $a, b \in A$ with $a < b$ such that $\lim_{n\to\infty} F_n(a) = F(a)$, $\lim_{n\to\infty} F_n(b) = F(b)$, $F(a) < \varepsilon$ and $F(b) > 1 - \varepsilon$. For n large enough, we then also obtain $F_n(a) < \varepsilon$ and $F_n(b) > 1 - \varepsilon$. Applying these properties, we obtain

$$\begin{aligned}\limsup_{n\to\infty} \left| \int f1_{(-\infty,a]} \,\mathrm{d}\mu_n - \int f1_{(-\infty,a]} \,\mathrm{d}\mu \right| &\leq \limsup_{n\to\infty} \|f\|_\infty (\mu_n((-\infty, a]) + \mu((-\infty, a])) \\ &= \limsup_{n\to\infty} \|f\|_\infty (F_n(a) + F(a)) \leq 2\varepsilon \|f\|_\infty\end{aligned}$$

and similarly,

$$\begin{aligned}\limsup_{n\to\infty} \left| \int f1_{(b,\infty)} \,\mathrm{d}\mu_n - \int f1_{(b,\infty)} \,\mathrm{d}\mu \right| &\leq \limsup_{n\to\infty} \|f\|_\infty (\mu_n((b, \infty)) + \mu((b, \infty))) \\ &= \limsup_{n\to\infty} \|f\|_\infty (1 - F_n(b) + (1 - F(b))) \leq 2\varepsilon \|f\|_\infty.\end{aligned}$$

As a consequence, we obtain the bound

$$\limsup_{n\to\infty} \left| \int f \,\mathrm{d}\mu_n - \int f \,\mathrm{d}\mu \right| \leq 4\varepsilon \|f\|_\infty + \limsup_{n\to\infty} \left| \int f1_{(a,b]} \,\mathrm{d}\mu_n - \int f1_{(a,b]} \,\mathrm{d}\mu \right|. \tag{3.6}$$

Now, f is uniformly continuous on $[a, b]$. Pick $\delta > 0$ parrying ε for this uniform continuity. Using that A is dense in $\mathbb{R}$, pick a partition $a = x_0 < x_1 < \cdots < x_m = b$ of elements in

A such that $|x_i - x_{i-1}| \leq \delta$ for all $i \leq m$. We then have $|f(x) - f(x_{i-1})| \leq \varepsilon$ whenever $x_{i-1} < x \leq x_i$, and so

$$\begin{aligned}\left|\int f1_{(a,b]} \,\mathrm{d}\mu_n - \int \sum_{i=1}^m f(x_{i-1})1_{(x_{i-1},x_i]} \,\mathrm{d}\mu_n\right| &= \left|\sum_{i=1}^m \int (f(x) - f(x_{i-1}))1_{(x_{i-1},x_i]}(x) \,\mathrm{d}\mu_n(x)\right| \\ &\leq \sum_{i=1}^m \int |f(x) - f(x_{i-1})|1_{(x_{i-1},x_i]}(x) \,\mathrm{d}\mu_n(x) \\ &\leq \varepsilon\mu_n((a,b]) \leq \varepsilon,\end{aligned}$$

leading to $|\int f1_{(a,b]} \,\mathrm{d}\mu_n - \sum_{i=1}^m f(x_{i-1})(F_n(x_i) - F_n(x_{i-1}))| \leq \varepsilon$. By a similar argument, we obtain the same bound with μ instead of μ_n. Combining these conclusions, we obtain

$$\begin{aligned}\left|\int f1_{(a,b]} \,\mathrm{d}\mu_n - \int f1_{(a,b]} \,\mathrm{d}\mu\right| &\leq 2\varepsilon + \left|\sum_{i=1}^m f(x_{i-1})(F_n(x_i) - F_n(x_{i-1}) - (F(x_i) - F(x_{i-1})))\right| \\ &\leq 2\varepsilon + \|f\|_\infty \sum_{i=1}^m |F_n(x_i) - F(x_i)| + |F_n(x_{i-1}) - F(x_{i-1})|.\end{aligned}$$

As each x_i is in A, this yields

$$\limsup_{n\to\infty} |\int f1_{(a,b]} \,\mathrm{d}\mu_n - \int f1_{(a,b]} \,\mathrm{d}\mu| \leq 2\varepsilon. \tag{3.7}$$

Combining (3.6) and (3.7), we obtain $\limsup_{n\to\infty} |\int f \,\mathrm{d}\mu_n - \int f \,\mathrm{d}\mu| \leq 4\varepsilon\|f\|_\infty + 2\varepsilon$. As $\varepsilon > 0$ was arbitrary, this yields $\lim_{n\to\infty} \int f \,\mathrm{d}\mu_n = \int f \,\mathrm{d}\mu$. As a consequence, $\mu_n \xrightarrow{wk} \mu$, as was to be shown. □

3.3 Weak convergence and convergence in probability

In this section, we will investigate the connections between convergence in distribution of random variables and convergence in probability. In general, these two modes of convergence do not work well together, but in the case where we have convergence in distribution of one sequence and convergence in probability towards a constant of another, we may obtain useful results. As we in this section work with random variables instead of measures, we assume given some background probability space $(\Omega, \mathcal{F}, P)$. In general, given two sequences (X_n) and (Y_n), our results in this section will only involve the distributions of (X_n, Y_n), and in principle each (X_n, Y_n) could be defined on separate probability spaces $(\Omega_n, \mathcal{F}_n, P_n)$. This, however, is mostly a theoretical distinction and is of little practical importance. When formulating our results, we wil for clarity in general not mention the background probability space explicitly.

We begin with a simple equivalence. Note that for $x \in \mathbb{R}$, statements such as $X_n \xrightarrow{P} x$ and $X_n \xrightarrow{\mathcal{D}} x$ are equivalent to the statements that $\lim_{n\to\infty} P(|X_n - x| > \varepsilon) = 0$ for $\varepsilon > 0$ and $\lim_{n\to\infty} Ef(X_n) = f(x)$ for $f \in C_b(\mathbb{R})$, respectively, and so in terms of stochasticity depend only on the distributions of X_n for each $n \geq 1$. In the case of convergence in probability, this is not the typical situation, as we in general have that $X_n \xrightarrow{P} X$ is a statement depending on the multivariate distributions (X_n, X).

Lemma 3.3.1. *Let (X_n) be a sequence of random variables, and let $x \in \mathbb{R}$. Then $X_n \xrightarrow{P} x$ if and only if $X_n \xrightarrow{\mathcal{D}} x$.*

Proof. By Theorem 1.2.8, we know that if $X_n \xrightarrow{P} x$, then $X_n \xrightarrow{\mathcal{D}} x$ as well. In order to prove the converse, assume that $X_n \xrightarrow{\mathcal{D}} x$, we wish to show that $X_n \xrightarrow{P} x$. Take $\varepsilon > 0$. By Lemma 3.1.3, there exists $g \in C_b(\mathbb{R})$ such that $1_{[x-\varepsilon/2, x+\varepsilon/2]}(y) \leq g(y) \leq 1_{(x-\varepsilon, x+\varepsilon)}(y)$ for $y \in \mathbb{R}$. With $f = 1 - g$, we then also obtain $1_{(x-\varepsilon,x+\varepsilon)^c}(y) \leq f(y) \leq 1_{[x-\varepsilon/2,x+\varepsilon/2]^c}(y)$, and in particular, $f(x) = 0$. By weak convergence, we may then conclude

$$\begin{aligned}\limsup_{n\to\infty} P(|X_n - x| \geq \varepsilon) &= \limsup_{n\to\infty} E1_{(x-\varepsilon,x+\varepsilon)^c}(X_n) \\ &\leq \limsup_{n\to\infty} Ef(X_n) = Ef(x) = 0,\end{aligned}$$

so $X_n \xrightarrow{P} x$, as desired. □

Lemma 3.3.2 (Slutsky's Lemma)**.** *Let (X_n, Y_n) be a sequence of random variables, and let X be some other variable. If $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{P} 0$, then $X_n + Y_n \xrightarrow{\mathcal{D}} X$.*

Proof. Applying Theorem 3.1.7, we find that in order to obtain the result, it suffices to prove that $\lim_{n\to\infty} Ef(X_n + Y_n) = Ef(X)$ for $f \in C_b^u(\mathbb{R})$. Fix $f \in C_b^u(\mathbb{R})$. Note that

$$\begin{aligned}&\limsup_{n\to\infty} |Ef(X_n + Y_n) - Ef(X)| \\ &\leq \limsup_{n\to\infty} |Ef(X_n + Y_n) - Ef(X_n)| + \limsup_{n\to\infty} |Ef(X_n) - Ef(X)| \\ &= \limsup_{n\to\infty} |Ef(X_n + Y_n) - Ef(X_n)|, \end{aligned} \tag{3.8}$$

so it suffices to show that the latter is zero. To this end, take $\varepsilon > 0$, and pick $\delta > 0$ parrying ε for the uniform continuity of f. We then obtain in particular that for $x \in \mathbb{R}$ and $|y| \leq \delta$, $|f(x + y) - f(x)| \leq \varepsilon$. We then obtain

$$|Ef(X_n + Y_n) - Ef(X_n)| \leq E|f(X_n + Y_n) - f(X_n)| \leq \varepsilon + 2\|f\|_\infty P(|Y_n| > \delta),$$

and as $Y_n \xrightarrow{P} 0$, this implies $\limsup_{n\to\infty} |Ef(X_n + Y_n) - Ef(X_n)| \leq \varepsilon$. As $\varepsilon > 0$ was arbitrary, this yields $\limsup_{n\to\infty} |Ef(X_n + Y_n) - Ef(X_n)| = 0$. Combining this with (3.8), we obtain $\lim_{n\to\infty} Ef(X_n + Y_n) = Ef(X)$, as desired. □

Theorem 3.3.3. *Let (X_n, Y_n) be a sequence of random variables, let X be some other variable and let $y \in \mathbb{R}$. Let $h : \mathbb{R}^2 \to \mathbb{R}$ be a continuous mapping. If $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{P} y$, then $h(X_n, Y_n) \xrightarrow{\mathcal{D}} h(X, y)$.*

Proof. First note that $h(X_n, Y_n) = h(X_n, Y_n) - h(X_n, y) + h(X_n, y)$. Define $h_y : \mathbb{R} \to \mathbb{R}$ by $h_y(x) = h(x, y)$. The distribution of $h(X_n, y)$ is then $h_y(X_n(P))$ and the distribution of $h(X, y)$ is $h_y(X(P))$. As we have assumed that $X_n \xrightarrow{\mathcal{D}} X$, Lemma 3.1.2 yields that $X_n(P) \xrightarrow{wk} X(P)$. Therefore, as h_y is continuous, $h_y(X_n(P)) \xrightarrow{wk} h_y(X(P))$ by Lemma 3.1.8, which by Lemma 3.1.2 implies that $h(X_n, y) \xrightarrow{wk} h(X, y)$. Therefore, by Lemma 3.3.2, it suffices to prove that $h(X_n, Y_n) - h(X_n, y)$ converges in probability to zero.

To this end, let $\varepsilon > 0$, we have to show $\lim_{n\to\infty} P(|h(X_n, Y_n) - h(X_n, y)| > \varepsilon) = 0$. We have assumed that h is continuous. Equipping $\mathbb{R}^2$ with the metric $d : \mathbb{R}^2 \times \mathbb{R}^2 \to [0, \infty)$ given by $d((a_1, a_2), (b_1, b_2)) = |a_1 - b_1| + |a_2 - b_2|$, h is in particular continuous with respect to this metric on $\mathbb{R}^2$. Now let $M, \eta > 0$ and note that h is uniformly continous on the compact set $[-M, M] \times [y - \eta, y + \eta]$. Therefore, we may pick $\delta > 0$ parrying ε for this uniform continuity, and we may assume that $\delta \leq \eta$. We then have

$$\begin{aligned}(|X_n| \leq M) \cap (|Y_n - y| \leq \delta) &\subseteq (|X_n| \leq M) \cap (|Y_n - y| \leq \delta) \cap (d((X_n, Y_n), (X_n, y)) \leq \delta) \\ &\subseteq (|h(X_n, Y_n) - h(X_n, y)| \leq \varepsilon),\end{aligned}$$

which yields $(|h(X_n, Y_n) - h(X_n, y)| > \varepsilon) \subseteq (|X_n| > M) \cup (|Y_n - y| > \delta)$ and thus

$$\begin{aligned}\limsup_{n\to\infty} P(|h(X_n, Y_n) - h(X_n, y)| > \varepsilon) &\leq \limsup_{n\to\infty} P(|X_n| > M) + P(|Y_n - y| > \delta) \\ &\leq \sup_{n\geq 1} P(|X_n| > M).\end{aligned}$$

By Lemma 3.1.6, the latter tends to zero as M tends to infinity. We therefore conclude that $\limsup_{n\to\infty} P(|h(X_n, Y_n) - h(X_n, y)| > \varepsilon) = 0$, so $h(X_n, Y_n) - h(X_n, y)$ converges in probability to zero and the result follows. □

3.4 Weak convergence and characteristic functions

Let $\mathbb{C}$ denote the complex numbers. In this section, we will associate to each probability measure μ on $(\mathbb{R}, \mathcal{B})$ a mapping $\varphi : \mathbb{R} \to \mathbb{C}$, called the characteristic function of μ. We will see that the characteristic function determines the probability measure uniquely, in the sense that two probability measures with equal characteristic functions in fact are equal. Furthermore, we will show, and this will be the main result of the section, that weak convergence of probability measures is equivalent to pointwise convergence of characteristic functions. As characteristic functions in general are pleasant to work with, both from theoretical and practical viewpoints, this result is of considerable use.

Before we introduce the characteristic function, we recall some results from complex analysis. For $z \in \mathbb{C}$, we let $\Re(z)$ and $\Im(z)$ denote the real and imaginary parts of z, and with i denoting the imaginary unit, we always have $z = \Re(z) + i\Im(z)$. $\Re$ and $\Im$ are then mappings from $\mathbb{C}$ to $\mathbb{R}$. Also, for $z \in \mathbb{C}$ with $z = \Re(z) + i\Im(z)$, we define $\overline{z} = \Re(z) - i\Im(z)$ and refer to $\overline{z}$ as the complex conjugate of z.

Also recall that we may define the complex exponential by its Taylor series, putting

$$e^z = \sum_{n=0}^\infty \frac{z^n}{n!}$$

for any $z \in \mathbb{C}$, where the series is absolutely convergent. We then also obtain the relationship $e^{iz} = \cos z + i \sin z$, where the complex cosine and the complex sine functions are defined by their Taylor series,

$$\cos z = \sum_{n=0}^\infty \frac{(-1)^n z^{2n}}{(2n)!} \text{ and } \sin z = \sum_{n=0}^\infty \frac{(-1)^n z^{2n+1}}{(2n+1)!}.$$

In particular, for $t \in \mathbb{R}$, we obtain $e^{it} = \cos t + i \sin t$, where $\cos t$ and $\sin t$ here are the ordinary real cosine and sine functions. This shows that the complex exponential of a purely imaginary argument yields a point on the unit circle corresponding to an angle of t measured in radians.

Let $(E, \mathcal{E}, \mu)$ be a measure space and let $f : E \to \mathbb{C}$ be a complex valued function defined on E. Then $f(z) = \Re(f(z)) + i\Im(f(z))$. We refer to the mappings $z \mapsto \Re(f(z))$ and $z \mapsto \Im(f(z))$ as the real and imaginary parts of f, and denote them by $\Re f$ and $\Im f$, respectively. Endowing $\mathbb{C}$ with the σ-algebra $\mathcal{B}_\mathbb{C}$ generated by the open sets, it also holds that $\mathcal{B}_\mathbb{C}$ is the smallest σ-algebra making $\Re$ and $\Im$ measurable. We then obtain that for any $f : E \to \mathbb{C}$, f is $\mathcal{E}$-$\mathcal{B}_\mathbb{C}$ measurable if and only if both the real and imaginary parts of f are $\mathcal{E}$-$\mathcal{B}$ measurable.

Definition 3.4.1. *Let $(E, \mathcal{E}, \mu)$ be a measure space. A measurable function $f : E \to \mathbb{C}$ is said to be integrable if both $\Re f$ and $\Im f$ are integrable, and in the affirmative, the integral of f is defined by $\int f \,\mathrm{d}\mu = \int \Re f \,\mathrm{d}\mu + i \int \Im f \,\mathrm{d}\mu$.*

The space of integrable complex functions is denoted $\mathcal{L}_{\mathbb{C}}(E, \mathcal{E}, \mu)$ or simply $\mathcal{L}_{\mathbb{C}}$. Note that we have the inequalities $|\Re f| \le |f|$, $|\Im f| \le |f|$ and $|f| \le |\Re f| + |\Im f|$. Therefore, f is integrable if and only if $|f|$ is integrable.

Example 3.4.2. Let $\gamma \neq 0$ be a real number. Since $|e^{i\gamma t}| = 1$ for all $t \in \mathbb{R}$, $t \mapsto e^{i\gamma t}$ is integrable with respect to the Lebesgue measure on all compact intervals $[a, b]$. As it holds that $e^{i\gamma t} = \cos \gamma t + i \sin \gamma t$, we obtain

$$\begin{aligned}
\int_a^b e^{i\gamma t} \,\mathrm{d}t &= \int_a^b \cos \gamma t \,\mathrm{d}t + i \int_a^b \sin \gamma t \,\mathrm{d}t \\
&= \frac{\sin \gamma b - \sin \gamma a}{\gamma} + i \frac{-\cos \gamma b + \cos \gamma a}{\gamma} \\
&= \frac{-i}{\gamma}(\cos \gamma b + i \sin \gamma b - \cos \gamma a - i \sin \gamma a) = \frac{e^{i\gamma b} - e^{i\gamma a}}{i\gamma},
\end{aligned}$$

extending the results for the real exponential function to the complex case. ∘

Lemma 3.4.3. *Let $(E, \mathcal{E}, \mu)$ be a measure space. If $f, g \in \mathcal{L}_{\mathbb{C}}$ and $z, w \in \mathbb{C}$, it then holds that $zf + wg \in \mathcal{L}_{\mathbb{C}}$ and $\int zf + wg \,\mathrm{d}\mu = z \int f \,\mathrm{d}\mu + w \int g \,\mathrm{d}\mu$.*

Proof. We first show that for f integrable and $z \in \mathbb{C}$, it holds that zf is integrable and $\int zf \,\mathrm{d}\mu = z \int f \,\mathrm{d}\mu$. First off, note that $\int |zf| \,\mathrm{d}\mu = \int |z||f| \,\mathrm{d}\mu = |z| \int |f| \,\mathrm{d}\mu < \infty$, so zf is integrable. Furthermore,

$$\begin{aligned}
\int zf \,\mathrm{d}\mu &= \int (\Re(z) + i\Im(z))(\Re f + i\Im f) \,\mathrm{d}\mu \\
&= \int \Re(z)\Re f - \Im(z)\Im f + i(\Re(z)\Im f + \Im(z)\Re f) \,\mathrm{d}\mu \\
&= \int \Re(z)\Re f - \Im(z)\Im f \,\mathrm{d}\mu + i \int \Re(z)\Im f + \Im(z)\Re f \,\mathrm{d}\mu \\
&= \Re(z) \int \Re f \,\mathrm{d}\mu - \Im(z) \int \Im f \,\mathrm{d}\mu + i \left(\Re(z) \int \Im f \,\mathrm{d}\mu + \Im(z) \int \Re f \,\mathrm{d}\mu \right) \\
&= (\Re(z) + i\Im(z)) \left(\int \Re f \,\mathrm{d}\mu + i \int \Im f \,\mathrm{d}\mu \right) = z \int f \,\mathrm{d}\mu,
\end{aligned}$$

as desired. Next, we show that for $f, g \in \mathcal{L}_{\mathbb{C}}$, $f + g \in \mathcal{L}_{\mathbb{C}}$ and $\int f + g \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu + \int g \,\mathrm{d}\mu$. First, as $|f + g| \le |f| + |g|$, it follows that $f + g \in \mathcal{L}_{\mathbb{C}}$. In order to obtain the desired identity for the integrals, we note that

$$\begin{aligned}\int f+g\,\mathrm{d}\mu &= \int \Re(f+g)\,\mathrm{d}\mu + i\int \Im(f+g)\,\mathrm{d}\mu\\ &= \int \Re f + \Re g\,\mathrm{d}\mu + i\int \Im f + \Im g\,\mathrm{d}\mu\\ &= \int \Re f + i\int \Im f\,\mathrm{d}\mu + \int \Re g\,\mathrm{d}\mu + i\int \Im g\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu,\end{aligned}$$

as desired. Collecting our conclusions, we obtain the desired result. □

Lemma 3.4.4. *Let $(E,\mathcal{E},\mu)$ be a measure space. If $f\in\mathcal{L}_{\mathbb{C}}$, then $|\int f\,\mathrm{d}\mu| \le \int |f|\,\mathrm{d}\mu$.*

Proof. Recall that for $z\in\mathbb{C}$, there exists $\theta\in\mathbb{R}$ such that $z=|z|e^{i\theta}$. Applying this to the integral $\int f\,\mathrm{d}\mu$, we obtain $|\int f\,\mathrm{d}\mu| = e^{-i\theta}\int f\,\mathrm{d}\mu = \int e^{-i\theta} f\,\mathrm{d}\mu$ by Lemma 3.4.3. As the left hand side is real, the right hand side must be real as well. Hence, we obtain

$$\begin{aligned}\left|\int f\,\mathrm{d}\mu\right| &= \Re\left(\int e^{-i\theta} f\,\mathrm{d}\mu\right) = \int \Re(e^{-i\theta}f)\,\mathrm{d}\mu\\ &\le \int |\Re(e^{-i\theta}f)|\,\mathrm{d}\mu \le \int |e^{-i\theta}f|\,\mathrm{d}\mu = \int |f|\,\mathrm{d}\mu,\end{aligned}$$

as desired. □

Next, we state versions of the dominated convergence theorem and Fubini's theorem for complex mappings.

Theorem 3.4.5. *Let $(E,\mathcal{E},\mu)$ be a measure space, and let (f_n) be a sequence of measurable mappings from E to $\mathbb{C}$. Assume that the sequence (f_n) converges μ-almost everywhere to some mapping f. Assume that there exists a measurable, integrable mapping $g: E\to[0,\infty)$ such that $|f_n|\le g$ μ-almost everywhere for all n. Then f_n is integrable for all $n\ge 1$, f is measurable and integrable, and*

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

Proof. As f_n converges μ-almost everywhere to f, we find that $\Re f_n$ converges μ-almost everywhere to $\Re f$ and $\Im f_n$ converges μ-almost everywhere to $\Im f$. Furthermore, we have $|\Re f_n|\le g$ and $|\Im f_n|\le g$ μ-almost everywhere. Therefore, the dominated convergence theorem for real-valued mappings yields

$$\begin{aligned}\lim_{n\to\infty}\int f_n\,\mathrm{d}\mu &= \lim_{n\to\infty}\int \Re f_n\,\mathrm{d}\mu + i\int \Im f_n\,\mathrm{d}\mu\\ &= \int \lim_{n\to\infty}\Re f_n\,\mathrm{d}\mu + i\int \lim_{n\to\infty}\Im f_n\,\mathrm{d}\mu = \int \lim_{n\to\infty} f_n\,\mathrm{d}\mu,\end{aligned}$$

as desired. □

Theorem 3.4.6. *Let $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$ be two σ-finite measure spaces, and assume that $f : E \times F \to \mathbb{C}$ is $\mathcal{E} \otimes \mathcal{F}$ measurable and $\mu \otimes \nu$ integrable. Then $y \mapsto f(x, y)$ is integrable with respect to ν for μ-almost all x, the set where this is the case is measurable, and it holds that*

$$\int f(x,y)\,\mathrm{d}(\mu\otimes\nu)(x,y) = \int\int f(x,y)\,\mathrm{d}\nu(y)\,\mathrm{d}\mu(x).$$

Proof. As f is $\mu \otimes \nu$ integrable, both $\Re f$ and $\Im f$ are $\mu \otimes \nu$ integrable as well. Therefore, the Fubini theorem for real-valued mappings yields that $y \mapsto \Re f(x, y)$ and $y \mapsto \Im f(x, y)$ are integrable μ-almost surely, and the sets where this is the case are measurable. As a consequence, the set where $y \mapsto f(x, y)$ is integrable is measurable and is a μ-almost sure set. The Fubini theorem for real-valued mappings furthermore yields that

$$\begin{aligned}
\int f(x,y)\,\mathrm{d}(\mu\otimes\nu)(x,y) &= \int \Re f(x,y)\,\mathrm{d}(\mu\otimes\nu)(x,y) + i\int \Im f(x,y)\,\mathrm{d}(\mu\otimes\nu)(x,y)\\
&= \int\int \Re f(x,y)\,\mathrm{d}\nu(y)\,\mathrm{d}\mu(x) + i\int\int \Im f(x,y)\,\mathrm{d}\nu(y)\,\mathrm{d}\mu(x)\\
&= \int\int \Re f(x,y)\,\mathrm{d}\nu(y)\,\mathrm{d}\mu(x) + \int i\int \Im f(x,y)\,\mathrm{d}\nu(y)\,\mathrm{d}\mu(x)\\
&= \int\int \Re f(x,y)\,\mathrm{d}\nu(y) + i\int \Im f(x,y)\,\mathrm{d}\nu(y)\,\mathrm{d}\mu(x)\\
&= \int\int f(x,y)\,\mathrm{d}\nu(y)\,\mathrm{d}\mu(x),
\end{aligned}$$

as was to be proven. □

We are now ready to introduce the characteristic function of a probability measure on $(\mathbb{R}, \mathcal{B})$.

Definition 3.4.7. *Let μ be a probability measure on $(\mathbb{R}, \mathcal{B})$. The characteristic function for μ is the function $\varphi : \mathbb{R} \to \mathbb{C}$ defined by $\varphi(\theta) = \int e^{i\theta x}\,\mathrm{d}\mu(x)$.*

Since $|e^{i\theta x}| = 1$ for all values of θ and x, the integral $\int e^{i\theta x}\,\mathrm{d}\mu(x)$ in Definition 3.4.7 is always well-defined. For a random variable X with distribution μ, we also introduce the characteristic function φ of X as the characteristic function of μ. The characteristic function φ of X may then also be expressed as

$$\varphi(\theta) = \int e^{i\theta x}\,\mathrm{d}\mu(x) = \int e^{i\theta x}\,\mathrm{d}X(P)(x) = \int e^{i\theta X(\omega)}\,\mathrm{d}P(\omega) = Ee^{i\theta X}.$$

The following lemmas demonstrate some basic properties of characteristic functions.

Lemma 3.4.8. *Let μ be a probability measure on $(\mathbb{R},\mathcal{B})$ and assume that φ is the characteristic function of μ. The mapping φ has the following properties.*

(1). $\varphi(0)=1$.

(2). For all $\theta\in\mathbb{R}$, $|\varphi(\theta)|\leq 1$.

(3). For all $\theta\in\mathbb{R}$, $\varphi(-\theta)=\overline{\varphi(\theta)}$.

(4). φ *is uniformly continuous.*

Furthermore, for $n\geq 1$, if $\int |x|^n \,\mathrm{d}\mu(x)$ is finite, then φ is n times continuously differentiable, and with $\varphi^{(n)}$ denoting the n'th derivative, we have $\varphi^{(n)}(\theta)=i^n\int x^n e^{i\theta x}\,\mathrm{d}\mu(x)$. In particular, in the affirmative, $\varphi^{(n)}(0)=i^n\int x^n\,\mathrm{d}\mu(x)$.

Proof. The first claim follows as $\varphi(0)=\int e^{i0x}\,\mathrm{d}\mu(x)=\mu(\mathbb{R})=1$, and the second claim follows as $|\varphi(\theta)|=|\int e^{i\theta x}\,\mathrm{d}\mu(x)|\leq\int |e^{i\theta x}|\,\mathrm{d}\mu(x)=1$. Also, the third claim follows since

$$\begin{aligned}\varphi(-\theta)&=\int e^{i(-\theta)x}\,\mathrm{d}\mu(x)=\int \cos(-\theta x)+i\sin(-\theta x)\,\mathrm{d}\mu(x)\\&=\int\cos(-\theta x)\,\mathrm{d}\mu(x)+i\int\sin(-\theta x)\,\mathrm{d}\mu(x)\\&=\int\cos(\theta x)\,\mathrm{d}\mu(x)-i\int\sin(\theta x)\,\mathrm{d}\mu(x)=\overline{\varphi(\theta)}.\end{aligned}$$

To obtain the fourth claim, let $\theta\in\mathbb{R}$ and let $h>0$. We then have

$$\begin{aligned}|\varphi(\theta+h)-\varphi(\theta)|&=\left|\int e^{i(\theta+h)x}-e^{i\theta x}\,\mathrm{d}\mu(x)\right|=\left|\int e^{i\theta x}(e^{ihx}-1)\,\mathrm{d}\mu(x)\right|\\&\leq\int|e^{i\theta x}||(e^{ihx}-1)|\,\mathrm{d}\mu(x)=\int|(e^{ihx}-1)|\,\mathrm{d}\mu(x),\end{aligned}$$

where $\lim_{h\to 0}\int|(e^{ihx}-1)|\,\mathrm{d}\mu(x)=0$ by the dominated convergence theorem. In order to use this to obtain uniform continuity, let $\varepsilon>0$. Choose $\delta>0$ so that for $0\leq h\leq\delta$, $\int|(e^{ihx}-1)|\,\mathrm{d}\nu(x)\leq\varepsilon$. We then find for any $x,y\in\mathbb{R}$ with $x<y$ and $|x-y|\leq\delta$ that $|\varphi(y)-\varphi(x)|=|\varphi(x+(y-x))-\varphi(y)|\leq\varepsilon$, and as a consequence, we find that φ is uniformly continuous.

Next, we prove the results on the derivative. We apply an induction argument, and wish to show for $n\geq 0$ that if $\int|x|^n\,\mathrm{d}\mu(x)$ is finite, then φ is n times continuously differentiable with $\varphi^{(n)}=i^n\int x^n e^{i\theta x}\,\mathrm{d}\mu(x)$. Noting that the induction start holds, it suffices to prove the induction step. Assume that the result holds for n, we wish to prove it for $n+1$. Assume that $\int |x|^{n+1}\,\mathrm{d}\mu(x)$ is finite. Fix $\theta \in \mathbb{R}$ and $h > 0$. We then have

$$\begin{aligned}\frac{\varphi^{(n)}(\theta+h)-\varphi^{(n)}(\theta)}{h} &= \frac{1}{h}\left(i^n \int x^n e^{i(\theta+h)x}\,\mathrm{d}\mu(x) - i^n \int x^n e^{i\theta x}\,\mathrm{d}\mu(x)\right)\\ &= i^n \int x^n e^{i\theta x}\frac{e^{ihx}-1}{h}\,\mathrm{d}\mu(x). \qquad (3.9)\end{aligned}$$

We wish to apply the dominated convergence theorem to calculate the limit of the above as h tends to zero. First note that by l'Hôpital's rule, we have

$$\lim_{h\to 0}\frac{e^{ihx}-1}{h} = \lim_{h\to 0}\frac{\cos hx - 1}{h} + i\frac{\sin hx}{h} = \lim_{h\to 0} -x\sin hx + ix\cos hx = ix,$$

so the integrand in the final expression of (3.9) converges pointwise to $ix^{n+1}e^{i\theta x}$. Note furthermore that since $|\cos x - 1| = |\int_0^x \sin y\,\mathrm{d}y| \le |x|$ and $|\sin x| = |\int_0^x \cos y\,\mathrm{d}y| \le |x|$ for all $x \in \mathbb{R}$, we have

$$\left|\frac{e^{ihx}-1}{h}\right| = \left|\frac{\cos hx - 1}{h} + i\frac{\sin hx}{h}\right| \le 2|x|,$$

yielding that $|x^n e^{i\theta x}\frac{e^{ihx}-1}{h}| \le 2|x|^{n+1}$. As we have assumed that the latter is integrable with respect to μ, the dominated convergence theorem applies and allows us to conclude that

$$\begin{aligned}\lim_{h\to 0}\frac{\varphi^{(n)}(\theta+h)-\varphi^{(n)}(\theta)}{h} &= \lim_{h\to 0} i^n \int x^n e^{i\theta x}\frac{e^{ihx}-1}{h}\,\mathrm{d}\mu(x)\\ &= i^n \int \lim_{h\to 0} x^n e^{i\theta x}\frac{e^{ihx}-1}{h}\,\mathrm{d}\mu(x)\\ &= i^{n+1}\int x^{n+1}\,\mathrm{d}\mu(x),\end{aligned}$$

as desired. This proves that φ is $n+1$ times differentiable, and yields the desired expression for $\varphi^{(n+1)}$. By another application of the dominated convergence theorem, we also obtain that $\varphi^{(n+1)}$ is continuous. This completes the induction proof. As a consequence of this latter result, it also follows that when $\int |x|^n\,\mathrm{d}\mu(x)$ is finite, $\varphi^{(n)}(0) = i^n \int x^n\,\mathrm{d}\mu(x)$. This completes the proof of the lemma. □

Lemma 3.4.9. *Assume that X is a random variable with characteristic function φ, and let $\alpha, \beta \in \mathbb{R}$. The variable $\alpha + \beta X$ has characteristic function ϕ given by $\phi(\theta) = e^{i\theta\alpha}\varphi(\beta\theta)$.*

Proof. Noting that $\phi(\theta) = Ee^{i\theta(\alpha+\beta X)} = e^{i\theta\alpha}Ee^{i\beta\theta X} = e^{i\theta\alpha}\varphi(\beta\theta)$, the result follows. □

Next, we show by example how to calculate the characteristic functions of a few distributions.

Example 3.4.10. Let φ be the characteristic function of the standard normal distribution, we wish to obtain a closed-form expression for φ. We will do this by proving that φ satisfies a particular differential equation. To this end, let f be the density of the standard normal distribution, $f(x) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}x^2)$. Note that for any $\theta \in \mathbb{R}$, we have by Lemma 3.4.8 that

$$\varphi(\theta) = \int_{-\infty}^{\infty} e^{i\theta x} f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} e^{-i\theta x} f(-x)\,\mathrm{d}x = \int_{-\infty}^{\infty} e^{-i\theta x} f(x)\,\mathrm{d}x = \varphi(-\theta) = \overline{\varphi(\theta)}.$$

As a consequence, $\Im\varphi(\theta) = 0$, so $\varphi(\theta) = \int_{-\infty}^{\infty}\cos(\theta x)f(x)\,\mathrm{d}x$. Next, note that

$$\left|\frac{\mathrm{d}}{\mathrm{d}\theta}\cos(\theta x)f(x)\right| = \left|-x\sin(\theta x)f(x)\right| \le |x|f(x),$$

which is integrable with respect to the Lebesgue measure. Therefore, $\varphi(\theta)$ is differentiable for all $\theta \in \mathbb{R}$, and the derivative may be calculated by an exchange of limits. Recalling that $f'(x) = -xf(x)$, we obtain

$$\begin{aligned}\varphi'(\theta) &= \frac{\mathrm{d}}{\mathrm{d}\theta}\int_{-\infty}^{\infty}\cos(\theta x)f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty}\frac{\mathrm{d}}{\mathrm{d}\theta}\cos(\theta x)f(x)\,\mathrm{d}x\\ &= -\int_{-\infty}^{\infty} x\sin(\theta x)f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty}\sin(\theta x)f'(x)\,\mathrm{d}x.\end{aligned}$$

Partial integration then yields

$$\begin{aligned}\varphi'(\theta) &= \lim_{M\to\infty}\int_{-M}^{M}\sin(\theta x)f'(x)\,\mathrm{d}x\\ &= \lim_{M\to\infty}\sin(\theta M)f(M) - \sin(-\theta M)f(-M) - \int_{-M}^{M}\theta\cos(\theta x)f(x)\,\mathrm{d}x\\ &= -\lim_{M\to\infty}\int_{-M}^{M}\theta\cos(\theta x)f(x)\,\mathrm{d}x = -\theta\varphi(\theta),\end{aligned}$$

since $\lim_{M\to\infty} f(M) = \lim_{M\to\infty} f(-M) = 0$. Thus, φ satisfies $\varphi'(\theta) = -\theta\varphi(\theta)$. All the solutions to this differential equation are of the form $\theta \mapsto c\exp(-\frac{1}{2}\theta^2)$ for some $c \in \mathbb{R}$, so we conclude that there exists $c \in \mathbb{R}$ such that $\varphi(\theta) = c\exp(-\frac{1}{2}\theta^2)$ for all $\theta \in \mathbb{R}$. As $\varphi(0) = 1$, this implies $\varphi(\theta) = \exp(-\frac{1}{2}\theta^2)$.

By Lemma 3.4.9, we then also obtain as an immediate corollary that the characteristic function for the normal distribution with mean ξ and variance σ^2, where $\sigma > 0$, is given by $\theta \mapsto \exp(i\xi\theta - \frac{1}{2}\sigma^2\theta^2)$. ∘

Example 3.4.11. In this example, we derive the characteristic function for the standard exponential distribution. Let φ denote the characteristic function, we then have

$$\varphi(\theta) = \int_0^\infty \cos(\theta x)e^{-x}\,\mathrm{d}x + i\int_0^\infty \sin(\theta x)e^{-x}\,\mathrm{d}x,$$

and we need to evaluate both of these intgrals. In order to do so, fix $a, b \in \mathbb{R}$ and note that

$$\frac{d}{dx}\left(a\cos(\theta x)e^{-x} + b\sin(\theta x)e^{-x}\right) = (b\theta - a)\cos(\theta x)e^{-x} - (a\theta + b)\sin(\theta x)e^{-x}.$$

Next, note that the pair of equations $b\theta - a = c$ and $-(a\theta + b) = d$ have unique solutions in a and b given by $a = (-c - d\theta)/(1+\theta^2)$ and $b = (c\theta - d)/(1+\theta^2)$, such that we obtain

$$\frac{d}{dx}\left(\frac{-c - d\theta}{1+\theta^2}\cos(\theta x)e^{-x} + \frac{c\theta - d}{1+\theta^2}\sin(\theta x)e^{-x}\right) = c\cos(\theta x)e^{-x} + d\sin(\theta x)e^{-x}. \tag{3.10}$$

Using (3.10) with $c = 1$ and $d = 0$, we conclude that

$$\int_0^\infty \cos(\theta x)e^{-x}\,\mathrm{d}x = \lim_{M\to\infty}\left[\frac{-\cos(\theta x)}{1+\theta^2}e^{-x} + \frac{\theta\sin(\theta x)}{1+\theta^2}e^{-x}\right]_0^M = \frac{1}{1+\theta^2}, \tag{3.11}$$

and likewise, using (3.10) with $c = 0$ and $d = 1$, we find

$$\int_0^\infty \sin(\theta x)e^{-x}\,\mathrm{d}x = \lim_{M\to\infty}\left[\frac{-\theta\cos(\theta x)}{1+\theta^2}e^{-x} + \frac{-\sin(\theta x)}{1+\theta^2}e^{-x}\right]_0^M = \frac{\theta}{1+\theta^2}. \tag{3.12}$$

Combining (3.11) and (3.12), we conclude

$$\varphi(\theta) = \frac{1}{1+\theta^2} + i\frac{\theta}{1+\theta^2} = \frac{1+i\theta}{1+\theta^2} = \frac{1+i\theta}{(1+i\theta)(1-i\theta)} = \frac{1}{1-i\theta}.$$

By Lemma 3.4.9, we then also obtain that the exponential distribution with mean λ for $\lambda > 0$ has characteristic function $\theta \mapsto \frac{1}{1-i\lambda\theta}$. ∘

Example 3.4.12. We wish to derive the characteristic function for the Laplace distribution. Denote by φ this characteristic function. Using the relationships $\sin(-\theta x) = -\sin(\theta x)$ and $\cos(-\theta x) = \cos(\theta x)$ and recalling (3.10), we obtain

$$\begin{aligned}
\varphi(\theta) &= \int_{-\infty}^\infty \cos(\theta x)\frac{1}{2}e^{-|x|}\,\mathrm{d}x + i\int_{-\infty}^\infty \sin(\theta x)\frac{1}{2}e^{-|x|}\,\mathrm{d}x \\
&= \int_{-\infty}^\infty \cos(\theta x)\frac{1}{2}e^{-|x|}\,\mathrm{d}x = \int_0^\infty \cos(\theta x)e^{-x}\,\mathrm{d}x \\
&= \lim_{M\to\infty}\left[\frac{-\cos(\theta x)}{1+\theta^2}e^{-x} + \frac{\theta\sin(\theta x)}{1+\theta^2}e^{-x}\right]_0^M = \frac{1}{1+\theta^2}.
\end{aligned}$$

∘

Next, we introduce the convolution of two probability measures and argue that characteristic functions interact in a simple manner with convolutions.

Definition 3.4.13. *Let μ and ν be two probability measures on $(\mathbb{R}, \mathcal{B})$. The convolution $\mu * \nu$ of μ and ν is the probability measure $h(\mu \otimes \nu)$ on $(\mathbb{R}, \mathcal{B})$, where $h : \mathbb{R}^2 \to \mathbb{R}$ is given by $h(x, y) = x + y$.*

The following lemma gives an important interpretation of the convolution of two probability measures.

Lemma 3.4.14. *Let X and Y be two independent random variables X and Y defined on the same probability space. Assume that X has distribution μ and that Y has distribution ν. Then $X + Y$ has distribution $\mu * \nu$.*

Proof. As X and Y are independent, it holds that $(X, Y)(P) = \mu \otimes \nu$. With $h : \mathbb{R}^2 \to \mathbb{R}$ defined by $h(x, y) = x + y$, we then have, by the theorem on successive transformations, that

$$(X + Y)(P) = h(X, Y)(P) = h((X, Y)(P)) = h(\mu \otimes \nu) = \mu * \nu,$$

so $\mu * \nu$ is the distribution of $X + Y$. □

Lemma 3.4.15. *Let μ and ν be probability measures on $(\mathbb{R}, \mathcal{B})$ with characteristic functions φ and ϕ. Then $\mu * \nu$ has characteristic function $\theta \mapsto \varphi(\theta)\phi(\theta)$.*

Proof. Let ψ be the characteristic function of $\mu * \nu$. Fix $\theta \in \mathbb{R}$, we need to demonstrate that $\psi(\theta) = \varphi(\theta)\phi(\theta)$. Let $h : \mathbb{R}^2 \to \mathbb{R}$ be given by $h(x, y) = x + y$. Using Fubini's theorem, we obtain

$$\begin{aligned}
\psi(\theta) &= \int e^{i\theta z} \, \mathrm{d}(\mu * \nu)(z) = \int e^{i\theta z} \, \mathrm{d}h(\mu \otimes \nu)(z) \\
&= \int e^{i\theta h(x,y)} \, \mathrm{d}(\mu \otimes \nu)(x, y) = \int e^{i\theta(x+y)} \, \mathrm{d}(\mu \otimes \nu)(x, y) \\
&= \int \int e^{i\theta x} e^{i\theta y} \, \mathrm{d}\mu(x) \, \mathrm{d}\nu(y) = \int e^{i\theta y} \int e^{i\theta x} \, \mathrm{d}\mu(x) \, \mathrm{d}\nu(y) \\
&= \int e^{i\theta x} \, \mathrm{d}\mu(x) \int e^{i\theta y} \, \mathrm{d}\nu(y) = \varphi(\theta)\phi(\theta),
\end{aligned}$$

proving the desired result. □

As mentioned earlier, two of our main objectives in this section is to prove that probability measures are uniquely determined by characteristic functions, and to prove that weak convergence is equivalent to pointwise convergence of characteristic functions. To show these results, we will employ a method based on convolutions with normal distributions.

We will need three technical lemmas. Lemma 3.4.16 shows that convoluting a probability measure with a normal distribution approximates the original probability measure when the mean in the normal distribution is zero and the variance is small. Lemma 3.4.17 will show that if we wish to prove weak convergence of some sequence (μ_n) to some μ, it suffices to prove weak convergence when both the sequence and the limit are convoluted with a normal distribution with mean zero and small variance. Intuitively, this is not a surprising result. Its usefulness is clarified by Lemma 3.4.18, which states that the convolution of any probability measure μ with a particular normal distribution has density with respect to the Lebesgue measure, and the density can be obtained in closed form in terms of the characteristic function of the measure μ. This is a frequently seen feature of convolutions: The convolution of two probability measures in general inherits the regularity properties of each of the convoluted measures, in this particular case, the regularity property of having a density with respect to the Lebesgue measure. Summing up, Lemma 3.4.16 shows that convolutions with small normal distributions are close to the original probability measure, Lemma 3.4.17 shows that in order to prove weak convergence, it suffices to consider probability measures convoluted with normal distributions, and Lemma 3.4.18 shows that such convolutions possess good regularity properties.

Lemma 3.4.16. *Let μ be a probability measure on $(\mathbb{R},\mathcal{B})$. Let ξ_k be the normal distribution with mean zero and variance $\frac{1}{k}$. Then $\mu * \xi_k \xrightarrow{wk} \mu$.*

Proof. Consider a probability space endowed with two independent random variables X and Y, where X has distribution μ and Y follows a standard normal distribution. Define $Y_k = \frac{1}{\sqrt{k}}Y$, then Y_k is independent of X and has distribution ξ_k. As a consequence, we also obtain $P(|Y_k| > \delta) \leq \delta^{-2}E|Y_k|^2 = \delta^{-2}/k$ by Lemma 1.2.7, so Y_k converges in probability to 0. Therefore, Lemma 3.3.2 yields $X + Y_k \xrightarrow{\mathcal{D}} X$. However, by Lemma 3.4.14, $X + Y_k$ has distribution $\mu * \xi_k$. Thus, we conclude that $\mu * \xi_k \xrightarrow{wk} \mu$. □

Lemma 3.4.17. *Let (μ_n) be a sequence of probability measures on $(\mathbb{R},\mathcal{B})$, and let μ be some other probability measure. Let ξ_k be the normal distribution with mean zero and variance $\frac{1}{k}$. If it holds for all $k \geq 1$ that $\mu_n * \xi_k \xrightarrow{wk} \mu * \xi_k$, then $\mu_n \xrightarrow{wk} \mu$ as well.*

Proof. According to Theorem 3.1.7, it suffices to show that $\lim_{n\to\infty} \int f \,\mathrm{d}\mu_n = \int f \,\mathrm{d}\mu$ for $f \in C_b^u(\mathbb{R})$. To do so, let $f \in C_b^u(\mathbb{R})$. Fix $n, k \geq 1$. For convenience, assume given a probability space with independent random variables X_n, Y_k and X, such that X_n has distribution μ_n, Y_k has distribution ξ_k and X has distribution μ. We then have

$$\begin{aligned}\left|\int f\,\mathrm{d}\mu_n - \int f\,\mathrm{d}\mu\right| &= |Ef(X_n) - Ef(X)| \\ &\le |Ef(X_n) - Ef(X_n + Y_k)| + |Ef(X_n + Y_k) - Ef(X + Y_k)| \\ &+ |Ef(X + Y_k) - Ef(X)|. \end{aligned} \tag{3.13}$$

We will prove that the limes superior of the left-hand side is zero by bounding the limes superior of each of the three terms on the right-hand side. First note that by our assumptions, $\lim_{n\to\infty} |Ef(X_n + Y_k) - Ef(X + Y_k)| = \lim_{n\to\infty} |\int f\,\mathrm{d}(\mu_n * \xi_k) - \int f\,\mathrm{d}(\mu * \xi_k)| = 0$. Now consider some $\varepsilon > 0$. Pick δ parrying ε for the uniform continuity of f. We then obtain

$$\begin{aligned}|Ef(X_n) - Ef(X_n + Y_k)| &\le E|f(X_n) - f(X_n + Y_k)| \\ &\le \varepsilon + E|f(X_n) - f(X_n + Y_k)|1_{(|Y_k|>\delta)} \\ &\le \varepsilon + 2\|f\|_\infty P(|Y_k| > \delta),\end{aligned}$$

and similarly, $|Ef(X) - Ef(X + Y_k)| \le \varepsilon + 2\|f\|_\infty P(|Y_k| > \delta)$ as well. Combining these observations with (3.13), we get $\limsup_{n\to\infty} |\int f\,\mathrm{d}\mu_n - \int f\,\mathrm{d}\mu| \le 2\varepsilon + 4\|f\|_\infty P(|Y_k| > \delta)$. By Lemma 1.2.7, $P(|Y_k| > \delta) \le \delta^{-2}E|Y_k|^2 = \delta^{-2}/k$, so $\lim_{k\to\infty} P(|Y_k| > \delta) = 0$. All in all, this yields $\limsup_{n\to\infty} |\int f\,\mathrm{d}\mu_n - \int f\,\mathrm{d}\mu| \le 2\varepsilon$. As $\varepsilon > 0$ was arbitrary, this proves $\lim_{n\to\infty} \int f\,\mathrm{d}\mu_n = \int f\,\mathrm{d}\mu$, and thus $\mu_n \xrightarrow{wk} \mu$ by Theorem 3.1.7. □

Lemma 3.4.18. *Let μ be some probability measure, and let ξ_k be the normal distribution with mean zero and variance $\frac{1}{k}$. Let φ be the characteristic function for μ. The probability measure $\mu * \xi_k$ then has density f with respect to the Lebesgue measure, and the density is given by*

$$f(u) = \frac{1}{2\pi}\int \varphi(x)\exp\left(-\frac{1}{2k}x^2\right)e^{-iux}\,\mathrm{d}x.$$

Proof. Let $x \in \mathbb{R}$. By Tonelli's theorem and the change of variable $u = y + z$, we obtain

$$\begin{aligned}(\mu * \xi_k)((-\infty, x]) &= \int 1_{(y+z\le x)}\,\mathrm{d}(\mu\otimes\xi_k)(y,z) = \int\int 1_{(y+z\le x)}\,\mathrm{d}\xi_k(z)\,\mathrm{d}\mu(y) \\ &= \int\int 1_{(y+z\le x)}\frac{\sqrt{k}}{\sqrt{2\pi}}\exp\left(-\frac{k}{2}z^2\right)\mathrm{d}z\,\mathrm{d}\mu(y) \\ &= \int\int 1_{(u\le x)}\frac{\sqrt{k}}{\sqrt{2\pi}}\exp\left(-\frac{k}{2}(u-y)^2\right)\mathrm{d}u\,\mathrm{d}\mu(y) \\ &= \int_{-\infty}^{x}\int \frac{\sqrt{k}}{\sqrt{2\pi}}\exp\left(-\frac{k}{2}(y-u)^2\right)\mathrm{d}\mu(y)\,\mathrm{d}u.\end{aligned}$$

This implies that $\mu_n * \xi_k$ has density with respect to the Lebesgue measure, and the density f is given by $f(u) = \int \frac{\sqrt{k}}{\sqrt{2\pi}} \exp\left(-\frac{k}{2}(y-u)^2\right) \,\mathrm{d}\mu(y)$. By Example 3.4.10, $\exp(-\frac{k}{2}(y-u)^2)$ is the characteristic function of the normal distribution with mean zero and variance k, evaluated in $y-u$. Therefore, we have

$$\exp\left(-\frac{k}{2}(y-u)^2\right) = \int e^{i(y-u)x} \frac{1}{\sqrt{2\pi k}} \exp\left(-\frac{x^2}{2k}\right) \,\mathrm{d}x.$$

Substituting this in our expression for the density and applying Fubini's theorem, we obtain

$$\begin{aligned} f(u) &= \int \frac{\sqrt{k}}{\sqrt{2\pi}} \int e^{i(y-u)x} \frac{1}{\sqrt{2\pi k}} \exp\left(-\frac{x^2}{2k}\right) \,\mathrm{d}x \,\mathrm{d}\mu(y) \\ &= \frac{1}{2\pi} \int\int e^{i(y-u)x} \exp\left(-\frac{x^2}{2k}\right) \,\mathrm{d}x \,\mathrm{d}\mu(y) \\ &= \frac{1}{2\pi} \int\int e^{iyx} \,\mathrm{d}\mu(y) \exp\left(-\frac{1}{2k}x^2\right) e^{-iux} \,\mathrm{d}x \\ &= \frac{1}{2\pi} \int \varphi(x) \exp\left(-\frac{1}{2k}x^2\right) e^{-iux} \,\mathrm{d}x, \end{aligned}$$

proving the lemma. □

With Lemma 3.4.17 and Lemma 3.4.18 in hand, our main results on characteristic functions now follow without much difficulty.

Theorem 3.4.19. *Let μ and ν be probability measures on $(\mathbb{R}, \mathcal{B})$ with characteristic functions φ and ϕ, respectively. Then μ and ν are equal if and only if φ and ϕ are equal.*

Proof. If μ and ν are equal, we obtain for $\theta \in \mathbb{R}$ that

$$\varphi(\theta) = \int e^{i\theta x} \,\mathrm{d}\mu(x) = \int e^{i\theta x} \,\mathrm{d}\nu(x) = \phi(\theta),$$

so φ and ϕ are equal. Conversely, assume that φ and ϕ are equal. Let ξ_k be the normal distribution with mean zero and variance $\frac{1}{k}$. By Lemma 3.4.18, $\mu * \xi_k$ and $\nu * \xi_k$ both have densities with respect to the Lebesgue measure, and the densities f_k and g_k are given by

$$\begin{aligned} f_k(u) &= \frac{1}{2\pi} \int \varphi(x) \exp\left(-\frac{1}{2k}x^2\right) e^{-iux} \,\mathrm{d}x \quad \text{and} \\ g_k(u) &= \frac{1}{2\pi} \int \phi(x) \exp\left(-\frac{1}{2k}x^2\right) e^{-iux} \,\mathrm{d}x, \end{aligned}$$

respectively. As φ and ϕ are equal, f_k and g_k are equal, and so $\mu * \xi_k$ and $\nu * \xi_k$ are equal. By Lemma 3.4.16, $\mu * \xi_k \xrightarrow{wk} \mu$ and $\nu * \xi_k \xrightarrow{wk} \nu$. As $\mu * \xi_k$ and $\nu * \xi_k$ are equal by our above observations, we find that $\mu * \xi_k \xrightarrow{wk} \mu$ and $\mu * \xi_k \xrightarrow{wk} \nu$, yielding $\mu = \nu$ by Lemma 3.1.5. □

Theorem 3.4.20. *Let* (μ_n) *be a sequence of probability measures on* $(\mathbb{R}, \mathcal{B})$*, and let* μ *be some other probability measure. Assume that* μ_n *has characteristic function* φ_n *and that* μ *has characteristic function* φ*. Then* $\mu_n \xrightarrow{wk} \mu$ *if and only if* $\lim_{n\to\infty} \varphi_n(\theta) = \varphi(\theta)$ *for all* $\theta \in \mathbb{R}$*.*

Proof. First assume that $\mu_n \xrightarrow{wk} \mu$. Fix $\theta \in \mathbb{R}$. Since $x \mapsto \cos(\theta x)$ and $x \mapsto \sin(\theta x)$ are in $C_b(\mathbb{R})$, we obtain

$$\begin{aligned}\lim_{n\to\infty} \varphi_n(\theta) &= \lim_{n\to\infty} \int \cos(\theta x)\,\mathrm{d}\mu_n(x) + i \int \sin(\theta x)\,\mathrm{d}\mu_n(x) \\ &= \int \cos(\theta x)\,\mathrm{d}\mu(x) + i \int \sin(\theta x)\,\mathrm{d}\mu(x) = \varphi(\theta),\end{aligned}$$

as desired. This proves one implication. It remains to prove that if the characteristic functions converge, the probability measures converge weakly.

In order to do so, assume that $\lim_{n\to\infty} \varphi_n(\theta) = \varphi(\theta)$ for all $\theta \in \mathbb{R}$. We will use Lemma 3.4.17 and Lemma 3.4.18 to prove the result. Let ξ_k be the normal distribution with mean zero and variance $\frac{1}{k}$. By Lemma 3.4.18, $\mu_n * \xi_k$ and $\mu * \xi_k$ both have densities with respect to the Lebesgue measure, and the densities f_{nk} and f_k are given by

$$\begin{aligned}f_{nk}(u) &= \frac{1}{2\pi} \int \varphi_n(x) \exp\left(-\frac{1}{2k}x^2\right) e^{-iux}\,\mathrm{d}x \quad \text{and} \\ f_k(u) &= \frac{1}{2\pi} \int \varphi(x) \exp\left(-\frac{1}{2k}x^2\right) e^{-iux}\,\mathrm{d}x,\end{aligned}$$

respectively. Since $|\varphi_n|$ and $|\varphi|$ are bounded by one, the dominated convergence theorem yields $\lim_{n\to\infty} f_{nk}(u) = f_k(u)$ for all $k \geq 1$ and $u \in \mathbb{R}$. By Lemma 3.1.9, we may then conclude $\mu_n * \xi_k \xrightarrow{wk} \mu * \xi_k$ for all $k \geq 1$, and Lemma 3.4.17 then shows that $\mu_n \xrightarrow{wk} \mu$, as desired. □

For the following corollary, we introduce $C_b^\infty(\mathbb{R})$ as the set of continuous, bounded functions $f : \mathbb{R} \to \mathbb{R}$ which are differentiable infinitely often with bounded derivatives.

Corollary 3.4.21. *Let* (μ_n) *be a sequence of probability measures on* $(\mathbb{R}, \mathcal{B})$*, and let* μ *be some other probability measure. Then* $\mu_n \xrightarrow{wk} \mu$ *if and only if* $\lim_{n\to\infty} \int f\,\mathrm{d}\mu_n = \int f\,\mathrm{d}\mu$ *for* $f \in C_b^\infty(\mathbb{R})$*.*

Proof. As $C_b^\infty(\mathbb{R}) \subseteq C_b(\mathbb{R})$, it is immediate that if $\mu_n \xrightarrow{wk} \mu$, then $\lim_{n\to\infty} \int f\,\mathrm{d}\mu_n = \int f\,\mathrm{d}\mu$ for $f \in C_b^\infty(\mathbb{R})$. To show the converse implication, assume that $\lim_{n\to\infty} \int f\,\mathrm{d}\mu_n = \int f\,\mathrm{d}\mu$ for

$f \in C_b^\infty(\mathbb{R})$. In particular, it holds for $\theta \in \mathbb{R}$ that $\lim_{n\to\infty} \int \sin(\theta x)\,\mathrm{d}\mu_n(x) = \int \sin(\theta x)\,\mathrm{d}\mu(x)$ and $\lim_{n\to\infty} \int \cos(\theta x)\,\mathrm{d}\mu_n(x) = \int \cos(\theta x)\,\mathrm{d}\mu(x)$. Letting φ_n and φ denote the characteristic functions for μ_n and μ, respectively, we therefore obtain

$$\begin{aligned}\lim_{n\to\infty} \varphi_n(\theta) &= \lim_{n\to\infty} \int \cos(\theta x)\,\mathrm{d}\mu_n(x) + \int \sin(\theta x)\,\mathrm{d}\mu_n(x) \\ &= \int \cos(\theta x)\,\mathrm{d}\mu(x) + \int \sin(\theta x)\,\mathrm{d}\mu(x) = \varphi(\theta)\end{aligned}$$

for all $\theta \in \mathbb{R}$, so that Theorem 3.4.20 yields $\mu_n \xrightarrow{wk} \mu$, as desired. □

3.5 Central limit theorems

In this section, we use our results from Section 3.4 to prove Lindeberg's central limit theorem, which gives sufficient requirements for a normalized sum of independent variables to be approximated by a normal distribution in a weak convergence sense. This is one of the main classical results in the theory of weak convergence.

The proof relies on proving pointwise convergence of characteristic functions and applying Theorem 3.4.20. In order to prove such pointwise convergence, we will be utilizing some finer properties of the complex exponential, as well as a particular inequality for complex numbers. We begin by proving these auxiliary results, after which we prove the central limit theorem for the case of independent and identically distributed random variables. This result is weaker than the Lindeberg central limit theorem to be proven later, but the arguments applied illustrate well the techniques to be used in the more difficult proof of Lindeberg's central limit theorem, which is given afterwards.

Lemma 3.5.1. *Let $z_1, \ldots, z_n$ and $w_1, \ldots, w_n$ be complex numbers with $|z_i| \leq 1$ and $|w_i| \leq 1$ for all $i = 1, \ldots, n$. It then holds that $|\prod_{i=1}^n z_i - \prod_{i=1}^n w_i| \leq \sum_{i=1}^n |z_i - w_i|$.*

Proof. For $n \geq 2$, we have

$$\begin{aligned}\left|\prod_{i=1}^n z_i - \prod_{i=1}^n w_i\right| &\leq \left|\prod_{i=1}^n z_i - \left(\prod_{i=1}^{n-1} z_i\right) w_n\right| + \left|\left(\prod_{i=1}^{n-1} z_i\right) w_n - \prod_{i=1}^n w_i\right| \\ &= \left(\prod_{i=1}^{n-1} |z_i|\right) |z_n - w_n| + |w_n| \left|\prod_{i=1}^{n-1} z_i - \prod_{i=1}^{n-1} w_i\right| \\ &\leq |z_n - w_n| + \left|\prod_{i=1}^{n-1} z_i - \prod_{i=1}^{n-1} w_i\right|,\end{aligned}$$

and the desired result then follows by induction. □

Lemma 3.5.2. *It holds that*

(1). $|e^x - (1+x)| \leq \frac{1}{2}x^2$ *for all* $x \leq 0$.

(2). $\left|e^{ix} - 1 - ix + \frac{x^2}{2}\right| \leq \frac{3}{2}x^2$ *for all* $x \in \mathbb{R}$.

(3). $\left|e^{ix} - 1 - ix + \frac{x^2}{2}\right| \leq \frac{1}{3}|x|^3$ *for all* $x \in \mathbb{R}$.

Proof. To prove the first inequality, we apply a first order taylor expansion of the exponential mapping around zero. Fix $x \in \mathbb{R}$, by Taylor's theorem we then find that there exists $\xi(x)$ between zero and x such that $\exp(x) = 1 + x + \frac{1}{2}\exp(\xi(x))x^2$, which for $x \leq 0$ yields $|\exp(x) - (1+x)| \leq \frac{1}{2}|\exp(\xi(x))x^2| \leq \frac{1}{2}x^2$. This proves the first inequality.

Considering the second inequality, recall that $e^{ix} = \cos x + i \sin x$. We therefore obtain

$$\begin{aligned}|e^{ix} - 1 - ix + \tfrac{1}{2}x^2| = |e^{ix} - 1 - ix| + \tfrac{1}{2}x^2 &= |\cos x - 1 + i(\sin x - x)| + \tfrac{1}{2}x^2 \\ &\leq |\cos x - 1| + |\sin x - x| + \tfrac{1}{2}x^2.\end{aligned}$$

Recalling that $\cos' = -\sin$, $\cos'' = -\cos$, $\sin' = \cos$ and $\sin'' = -\sin$, first order Taylor expansions around zero yield the existence of $\xi^*(x)$ and $\xi^{**}(x)$ between zero and x such that

$$\cos x = 1 - \tfrac{1}{2}\cos(\xi^*(x))x^2 \quad \text{and} \quad \sin x = x - \tfrac{1}{2}\sin(\xi^{**}(x))x^2,$$

which yields $|\cos x - 1| \leq \frac{1}{2}|\cos(\xi^*(x))x^2| \leq \frac{1}{2}x^2$ and $|\sin x - x| \leq \frac{1}{2}|\sin(\xi^{**}(x))x^2| \leq \frac{1}{2}x^2$. Combining our three inequalities, we obtain $|e^{ix} - 1 - ix + \frac{1}{2}x^2| \leq \frac{3}{2}x^2$, proving the second inequality. Finally, we demonstrate the third inequality. Second order Taylor expansions around zero yield the existence of $\eta^*(x)$ and $\eta^{**}(x)$ such that

$$\cos x = 1 - \tfrac{1}{2}x^2 + \tfrac{1}{6}\sin(\eta^*(x))x^3 \quad \text{and} \quad \sin x = x - \tfrac{1}{6}\cos(\eta^{**}(x))x^3,$$

allowing us to obtain

$$\begin{aligned}|e^{ix} - 1 - ix + \tfrac{1}{2}x^2| &= |\cos x - 1 + \tfrac{1}{2}x^2 + i(\sin x - x)| \\ &\le |\cos x - 1 + \tfrac{1}{2}x^2| + |\sin x - x| \\ &\le \tfrac{1}{6}|\sin(\eta^*(x))x^3| + \tfrac{1}{6}|\cos(\eta^{**}(x))x^3| \le \tfrac{1}{3}|x|^3,\end{aligned}$$

as desired. This proves the third inequality. □

The combination of Lemma 3.5.1, Lemma 3.5.2 and Theorem 3.4.20 is sufficient to obtain the following central limit theorem for independent and identically distributed variables.

Theorem 3.5.3 (Classical central limit theorem)**.** *Let (X_n) be a sequence of independent and identically distributed random variables with mean ξ and variance σ^2, where $\sigma > 0$. It then holds that*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} \frac{X_k - \xi}{\sigma} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

where $\mathcal{N}(0,1)$ denotes the standard normal distribution.

Proof. It suffices to consider the case where $\xi = 0$ and $\sigma^2 = 1$. In this case, we have to argue that $\frac{1}{\sqrt{n}}\sum_{k=1}^n X_k \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$. Denote by φ the common characteristic function of X_n for $n \ge 1$, and denote by φ_n the characteristic function of $\frac{1}{\sqrt{n}}\sum_{k=1}^n X_k$. Lemma 3.4.15 and Lemma 3.4.9 show that $\varphi_n(\theta) = \varphi(\theta/\sqrt{n})^n$. Recalling from Example 3.4.10 that the standard normal distribution has characteristic function $\theta \mapsto \exp(-\frac{1}{2}\theta^2)$, Theorem 3.4.20 yields that in order to prove the result, it suffices to show for all $\theta \in \mathbb{R}$ that

$$\lim_{n\to\infty} \varphi(\theta/\sqrt{n})^n = e^{-\theta^2/2}. \tag{3.14}$$

To do so, first note that by Lemma 3.4.8 and Lemma 3.5.1 we obtain

$$\begin{aligned}|\varphi(\theta/\sqrt{n})^n - \exp(-\tfrac{1}{2}\theta^2)| &= |\varphi(\theta/\sqrt{n})^n - \exp(-\tfrac{1}{2n}\theta^2)^n| \\ &\le n|\varphi(\theta/\sqrt{n}) - \exp(-\tfrac{1}{2n}\theta^2)|. \end{aligned}\tag{3.15}$$

Now, as the variables (X_n) have second moment, we have from Lemma 3.4.8 that φ is two times continuously differentiable with $\varphi(0) = 1$, $\varphi'(0) = 0$ and $\varphi''(0) = -1$. Therefore, a first-order Taylor expansion shows that for each $\theta \in \mathbb{R}$, there exists $\xi(\theta)$ between 0 and θ such that $\varphi(\theta) = \varphi(0) + \varphi'(0)\theta + \frac{1}{2}\varphi''(\xi(\theta))\theta^2 = 1 + \frac{1}{2}\theta^2\varphi''(\xi(\theta))$. In particular, this yields

$$\begin{aligned}\varphi(\theta/\sqrt{n}) &= 1 + \tfrac{1}{2n}\theta^2\varphi''(\xi(\theta/\sqrt{n})) \\ &= 1 - \tfrac{1}{2n}\theta^2 + \tfrac{1}{2n}\theta^2(1 + \varphi''(\xi(\theta/\sqrt{n}))). \end{aligned}\tag{3.16}$$

Combining (3.15) and (3.16) and applying the first inequality of Lemma 3.5.2, we obtain

$$\begin{aligned}|\varphi(\theta/\sqrt{n})^n - \exp(-\tfrac{1}{2}\theta^2)| &\le n|1 - \tfrac{1}{2n}\theta^2 + \tfrac{1}{2n}\theta^2(1+\varphi''(\xi(\theta/\sqrt{n}))) - \exp(-\tfrac{1}{2n}\theta^2)| \\ &\le n|1 - \tfrac{1}{2n}\theta^2 - \exp(-\tfrac{1}{2n}\theta^2)| + \tfrac{1}{2}\theta^2|1+\varphi''(\xi(\theta/\sqrt{n}))| \\ &\le \tfrac{n}{2}(\tfrac{1}{2n}\theta^2)^2 + \tfrac{1}{2}\theta^2|1+\varphi''(\xi(\theta/\sqrt{n}))| \\ &= \tfrac{1}{8n}\theta^4 + \tfrac{1}{2}\theta^2|1+\varphi''(\xi(\theta/\sqrt{n}))|. \end{aligned} \tag{3.17}$$

Now, as n tends to infinity, $\theta/\sqrt{n}$ tends to zero, and so $\xi(\theta(\sqrt{n}))$ tends to zero. As φ'' by Theorem 3.4.8 is continuous with $\varphi''(0) = -1$, this implies $\lim_{n\to\infty} \varphi''(\xi(\theta/\sqrt{n})) = -1$. Therefore, we obtain from (3.17) that $\limsup_{n\to\infty} |\varphi(\theta/\sqrt{n})^n - \exp(-\frac{1}{2}\theta^2)| = 0$, proving (3.14). As a consequence, Theorem 3.4.20 yields $\frac{1}{\sqrt{n}}\sum_{k=1}^n X_k \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$. This concludes the proof. □

Theorem 3.5.3 and its proof demonstrates that in spite of the apparently deep nature of the central limit theorem, the essential ingredients in its proof are simply first-order Taylor expansions, bounds on the exponential function and Theorem 3.4.20. Next, we will show how to extend Theorem 3.5.3 to the case where the random variables are not necessarily identically distributed. The suitable framework for the statement of such more general results is that of triangular arrays.

Definition 3.5.4. *A triangular array is a double sequence $(X_{nk})_{n\ge k\ge 1}$ of random variables.*

Let $(X_{nk})_{n\ge k\ge 1}$ be a triangular array. We think of $(X_{nk})_{n\ge k\ge 1}$ as ordered in the shape of a triangle as follows:

$$\begin{array}{llll} X_{11} & & & \\ X_{21} & X_{22} & & \\ X_{31} & X_{32} & X_{33} & \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

We may then define the row sums by putting $S_n = \sum_{k=1}^n X_{nk}$, and we wish to establish conditions under which S_n converges in distribution to a normally distributed limit. In general, we will consider the case where $(X_{nk})_{k\le n}$ is independent for each $n \ge 1$, where $EX_{nk} = 0$ for all $n \ge k \ge 1$ and where $\lim_{n\to\infty} VS_n = 1$. In this case, it is natural to hope that under suitable regularity conditions, S_n converges in distribution to a standard normal distribution. The following example shows how the case considered in Theorem 3.5.3 can be put in terms of a triangular array.

Example 3.5.5. Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with mean ξ and variance σ^2, where $\sigma > 0$. For $1 \le k \le n$, we then define $X_{nk} = \frac{1}{\sqrt{n}} \frac{X_k - \xi}{\sigma}$. Ordering the variables in the shape of a triangle, we have

$$\begin{array}{llll}
\frac{1}{\sqrt{1}} \frac{X_1-\xi}{\sigma} & & & \\
\frac{1}{\sqrt{2}} \frac{X_1-\xi}{\sigma} & \frac{1}{\sqrt{2}} \frac{X_2-\xi}{\sigma} & & \\
\frac{1}{\sqrt{3}} \frac{X_1-\xi}{\sigma} & \frac{1}{\sqrt{3}} \frac{X_2-\xi}{\sigma} & \frac{1}{\sqrt{3}} \frac{X_3-\xi}{\sigma} & \\
\vdots & \vdots & \vdots & \ddots
\end{array}$$

The row sums of the triangular array are then $S_n = \sum_{k=1}^n X_{nk} = \frac{1}{\sqrt{n}} \sum_{k=1}^n \frac{X_k - \xi}{\sigma}$, which is the same as the expression considered in Theorem 3.5.3. ◦

Theorem 3.5.6 (Lindeberg's central limit theorem)**.** *Let $(X_{nk})_{n \ge k \ge 1}$ be a triangular array of variables with second moment. Assume that for each $n \ge 1$, the family $(X_{nk})_{k \le n}$ is independent and assume that $EX_{nk} = 0$ for all $n \ge k \ge 1$. With $S_n = \sum_{k=1}^n X_{nk}$, assume that $\lim_{n\to\infty} VS_n = 1$. Finally, assume that for all $c > 0$,*

$$\lim_{n\to\infty} \sum_{k=1}^n E1_{(|X_{nk}|>c)} X_{nk}^2 = 0. \tag{3.18}$$

It then holds that $S_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$, where $\mathcal{N}(0,1)$ denotes the standard normal distribution.

Proof. We define $\sigma_{nk}^2 = VX_{nk}$ and $\eta_n^2 = \sum_{k=1}^n \sigma_{nk}^2$. Our strategy for the proof will be similar to that for the proof of Theorem 3.5.3. Let φ_{nk} be the characteristic function of X_{nk}, and let φ_n be the characteristic function of S_n. As $(X_{nk})_{k \le n}$ is independent for each $n \ge 1$, Lemma 3.4.15 shows that $\varphi_n(\theta) = \prod_{k=1}^n \varphi_{nk}(\theta)$. Recalling Example 3.4.10, we find that by Theorem 3.4.20, in order to prove the theorem, it suffices to show for all $\theta \in \mathbb{R}$ that

$$\lim_{n\to\infty} \prod_{k=1}^n \varphi_{nk}(\theta) = \exp(-\tfrac{1}{2}\theta^2). \tag{3.19}$$

First note that by the triangle inequality, Lemma 3.4.8 and Lemma 3.5.1, we obtain

$$\begin{aligned}
\left| \prod_{k=1}^n \varphi_{nk}(\theta) - \exp(-\tfrac{1}{2}\theta^2) \right| &\le \left| \exp(-\tfrac{1}{2}\eta_n^2\theta^2) - \exp(-\tfrac{1}{2}\theta^2) \right| + \left| \prod_{k=1}^n \varphi_{nk}(\theta) - \exp(-\tfrac{1}{2}\eta_n^2\theta^2) \right| \\
&\le \left| \exp(-\tfrac{1}{2}\eta_n^2\theta^2) - \exp(-\tfrac{1}{2}\theta^2) \right| + \sum_{k=1}^n \left| \varphi_{nk}(\theta) - \exp(-\tfrac{1}{2}\sigma_{nk}^2\theta^2) \right|.
\end{aligned}$$

where the former term tends to zero, since $\lim_{n\to\infty} \eta_n = 1$ by our assumptions. We wish to show that the latter term also tends to zero. By Lemma 3.5.2, we have

$$\begin{aligned}
|\varphi_{nk}(\theta) - \exp(-\tfrac{1}{2}\sigma_{nk}^2\theta^2)| &\le |\varphi_{nk}(\theta) - (1 - \tfrac{1}{2}\sigma_{nk}^2\theta^2)| + |\exp(-\tfrac{1}{2}\sigma_{nk}^2\theta^2) - (1 - \tfrac{1}{2}\sigma_{nk}^2\theta^2)| \\
&\le |\varphi_{nk}(\theta) - (1 - \tfrac{1}{2}\sigma_{nk}^2\theta^2)| + \tfrac{1}{2}(\tfrac{1}{2}\sigma_{nk}^2\theta^2)^2,
\end{aligned}$$

such that

$$\begin{aligned}\sum_{k=1}^n \left|\varphi_{nk}(\theta) - \exp(-\tfrac{1}{2}\sigma_{nk}^2\theta^2)\right| &\le \sum_{k=1}^n |\varphi_{nk}(\theta) - (1 - \tfrac{1}{2}\sigma_{nk}^2\theta^2)| + \sum_{k=1}^n \tfrac{1}{2}|\tfrac{1}{2}\sigma_{nk}^2\theta^2|^2 \\ &= \sum_{k=1}^n |\varphi_{nk}(\theta) - (1 - \tfrac{1}{2}\sigma_{nk}^2\theta^2)| + \frac{\theta^4}{8}\sum_{k=1}^n \sigma_{nk}^4.\end{aligned}$$

Combining our conclusions, we find that (3.19) follows if only we can show

$$\lim_{n\to\infty}\sum_{k=1}^n |\varphi_{nk}(\theta) - (1 - \tfrac{1}{2}\sigma_{nk}^2\theta^2)| = 0 \quad \text{and} \quad \lim_{n\to\infty}\sum_{k=1}^n \sigma_{nk}^4 = 0. \tag{3.20}$$

Consider the first limit in (3.20). Fix $c > 0$. As $EX_{nk} = 0$ and $EX_{nk}^2 = \sigma_{nk}^2$, we may apply the two final inequalities of Lemma 3.5.2 to obtain

$$\begin{aligned}\sum_{k=1}^n |\varphi_{nk}(\theta) - (1 - \tfrac{1}{2}\sigma_{nk}^2\theta^2)| &= \sum_{k=1}^n |Ee^{i\theta X_{nk}} - 1 - iEX_{nk} + \tfrac{1}{2}\theta^2 EX_{nk}^2| \\ &\le \sum_{k=1}^n E|e^{i\theta X_{nk}} - 1 - iX_{nk} + \tfrac{1}{2}\theta^2 X_{nk}^2| \\ &\le \sum_{k=1}^n E1_{(|X_{nk}|\le c)}\tfrac{1}{3}|\theta X_{nk}|^3 + E1_{(|X_{nk}|>c)}\tfrac{3}{2}|\theta X_{nk}|^2 \\ &\le \frac{c|\theta|^3}{3}\eta_n^2 + \frac{3\theta^2}{2}\sum_{k=1}^n E1_{(|X_{nk}|>c)}X_{nk}^2.\end{aligned} \tag{3.21}$$

Now, by our asumption (3.18), $\lim_{n\to\infty}\sum_{k=1}^n E1_{(|X_{nk}|>c)}X_{nk}^2 = 0$, while we also have $\lim_{n\to\infty}(1/3)c|\theta|^3\eta_n^2 = (1/3)c|\theta|^3$. Applying these results with the bound (3.21), we obtain

$$\limsup_{n\to\infty}\sum_{k=1}^n |\varphi_{nk}(\theta) - (1 - \tfrac{1}{2}\sigma_{nk}^2\theta^2)| \le \frac{c|\theta|^3}{3},$$

and as $c > 0$ was arbitrary, this yields $\lim_{n\to\infty}\sum_{k=1}^n |\varphi_{nk}(\theta) - (1 - \frac{1}{2}\sigma_{nk}^2\theta^2)| = 0$, as desired. For the second limit in (3.20), we note that for all $c > 0$, it holds that

$$\begin{aligned}\sum_{k=1}^n \sigma_{nk}^4 &\le \left(\max_{k\le n}\sigma_{nk}^2\right)\sum_{k=1}^n \sigma_{nk}^2 = \eta_n^2 \max_{k\le n} EX_{nk}^2 \\ &= \eta_n^2 \max_{k\le n}(E1_{(|X_{nk}|\le c)}X_{nk}^2 + E1_{(|X_{nk}|>c)}X_{nk}^2) \\ &\le \eta_n^2 c^2 + \eta_n^2\sum_{k=1}^n E1_{(|X_{nk}|>c)}X_{nk}^2,\end{aligned}$$

so by (3.18), $\limsup_{n\to\infty}\sum_{k=1}^n \sigma_{nk}^4 \le c^2$ for all $c > 0$. Again, as $c > 0$ was arbitrary, this yields $\lim_{n\to\infty}\sum_{k=1}^n \sigma_{nk}^4 = 0$. Thus, both of the limit result in (3.20) hold. Therefore, (3.19) holds, and so Theorem 3.4.20 allows us to conclude the proof. □

The conditions given in Theorem 3.5.6 are in many cases sufficient to obtain convergence in distribution to the standard normal distribution. The main important condition (3.18) is known as Lindeberg's condition. The condition, however, is not always easy to check. The following result yields a central limit theorem where the conditions are less difficult to verify. Here, the condition (3.22) is known as Lyapounov's condition.

Theorem 3.5.7 (Lyapounov's central limit theorem)**.** *Let $(X_{nk})_{n\geq k\geq 1}$ be a triangular array of variables with third moment. Assume that for each $n \geq 1$, the family $(X_{nk})_{k\leq n}$ is independent and assume that $EX_{nk} = 0$ for all $n \geq k \geq 1$. With $S_n = \sum_{k=1}^n X_{nk}$, assume that $\lim_{n\to\infty} VS_n = 1$. Finally, assume that there is $\delta > 0$ such that*

$$\lim_{n\to\infty} \sum_{k=1}^n E|X_{nk}|^{2+\delta} = 0. \tag{3.22}$$

It then holds that $S_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$, where $\mathcal{N}(0,1)$ denotes the standard normal distribution.

Proof. We note that for $c > 0$, it holds that $|X_{nk}| > c$ implies $1 \leq |X_{nk}|^\delta/c^\delta$ and so

$$\sum_{k=1}^n E1_{(|X_{nk}|>c)}X_{nk}^2 \leq \sum_{k=1}^n E1_{(|X_{nk}|>c)}\tfrac{1}{c^\delta}|X_{nk}|^{2+\delta} \leq \frac{1}{c^\delta}\sum_{k=1}^n E|X_{nk}|^{2+\delta},$$

so Lyapounov's condition (3.22) implies Lindeberg's condition (3.18). Therefore, the result follows from Theorem 3.5.6. □

In order to apply Theorem 3.5.7, we require that the random variables in the triangular array have third moments. In many cases, this requirement is satisfied, and so Lyapounov's central limit theorem is frequently useful. However, the moment condition is too strong to obtain the classical central limit theorem of Theorem 3.5.3 as a corollary. As the following example shows, this theorem in fact does follow as a corollary from the stronger Lindeberg's central limit theorem.

Example 3.5.8. Let (X_n) be a sequence of independent and identically distributed random variables with mean ξ and variance σ^2, where $\sigma > 0$. As in Example 3.5.5, we define a triangular array by putting $X_{nk} = \frac{1}{\sqrt{n}}\frac{X_k-\xi}{\sigma}$ for $n \geq k \geq 1$. The elements of each row are then independent, with $EX_{nk} = 0$, and the row sums of the triangular array are given by $S_n = \sum_{k=1}^n X_{nk} = \frac{1}{\sqrt{n}}\sum_{k=1}^n \frac{X_k-\xi}{\sigma}$ and satisfy $VS_n = 1$. We obtain for $c > 0$ that

$$\begin{aligned}\lim_{n\to\infty}\sum_{k=1}^n E1_{(|X_{nk}|>c)}X_{nk}^2 &= \lim_{n\to\infty}\sum_{k=1}^n E1_{(|X_k-\xi|>c\sigma\sqrt{n})}\frac{(X_k-\xi)^2}{\sigma^2 n}\\ &= \lim_{n\to\infty}\frac{1}{\sigma^2}E1_{(|X_1-\xi|>c\sigma\sqrt{n})}(X_1-\xi)^2 = 0,\end{aligned}$$

by the dominated convergence theorem, since X_1 has second moment. Thus, we conclude that the triangular array satisfies Lindeberg's condition, and therefore, Theorem 3.5.6 applies and yields $\frac{1}{\sqrt{n}}\sum_{k=1}^n \frac{X_k-\xi}{\sigma} = S_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$, as in Theorem 3.5.3. ∘

3.6 Asymptotic normality

In Section 3.5, we saw examples of particular normalized sums of random variables converging to a standard normal distribution. The intuitive interpretation of these results is that the non-normalized sums approximate normal distributions with nonstandard parameters. In order to easily work with this idea, we in this section introduce the notion of asymptotic normality.

Definition 3.6.1. *Let (X_n) be a sequence of random variables, and let ξ and σ be real with $\sigma > 0$. We say that X_n is asymptotically normal with mean ξ and variance $\frac{1}{n}\sigma^2$ if it holds that $\sqrt{n}(X_n - \xi) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2)$, where $\mathcal{N}(0,\sigma^2)$ denotes the normal distribution with mean zero and variance σ^2. If this is the case, we write*

$$X_n \overset{as}{\sim} \mathcal{N}\left(\xi, \frac{1}{n}\sigma^2\right). \tag{3.23}$$

The results of Theorem 3.5.3 can be restated in terms of asymptotic normality as follows. Assume that (X_n) is a sequence of independent and identically distributed random variables with mean ξ and variance σ^2, where $\sigma > 0$. Theorem 3.5.3 then states that $\frac{1}{\sqrt{n}}\sum_{k=1}^n \frac{X_k-\xi}{\sigma} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$. By Lemma 3.1.8, this implies $\frac{1}{\sqrt{n}}\sum_{k=1}^n (X_k - \xi) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2)$, and so

$$\sqrt{n}\left(\left(\frac{1}{n}\sum_{k=1}^n X_k\right) - \xi\right) = \frac{1}{\sqrt{n}}\sum_{k=1}^n (X_k - \xi) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2),$$

which by Definition 3.6.1 corresponds to $\frac{1}{n}\sum_{k=1}^n X_k \overset{as}{\sim} \mathcal{N}(\xi, \frac{1}{n}\sigma^2)$. The intuitive content of this statement is that as n tends to infinity, the average $\frac{1}{n}\sum_{k=1}^n X_k$ is approximated by a normal distribution with the same mean and variance as the empirical average, namely ξ and $\frac{1}{n}\sigma^2$.

We next show two properties of asymptotic normality, namely that for asymptotically normal sequences (X_n), X_n converges in probability to the mean, and we show that asymptotic normality is preserved by transformations with certain mappings. These results are of considerable practical importance when analyzing the asymptotic properties of estimators based on independent and identically distributed samples.

Lemma 3.6.2. *Let (X_n) be a sequence of random variables, and let ξ and σ be real with $\sigma > 0$. Assume that X_n is asymptotically normal with mean ξ and variance $\frac{1}{n}\sigma^2$. It then holds that $X_n \xrightarrow{P} \xi$.*

Proof. Fix $\varepsilon > 0$. As X_n is asymptotically normal, $\sqrt{n}(X_n - \xi) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$, so Lemma 3.1.6 yields $\lim_{M\to\infty} \sup_{n\geq 1} P(\sqrt{n}|X_n - \xi| \geq M) = 0$. Now let $M > 0$, we then have

$$\begin{aligned}\limsup_{n\to\infty} P(|X_n - \xi| \geq \varepsilon) &= \limsup_{n\to\infty} P(\sqrt{n}|X_n - \xi| \geq \sqrt{n}\varepsilon)\\ &\leq \limsup_{n\to\infty} P(\sqrt{n}|X_n - \xi| \geq M)\\ &\leq \sup_{n\geq 1} P(\sqrt{n}|X_n - \xi| \geq M),\end{aligned}$$

and as $M > 0$ was arbitrary, this implies $\limsup_{n\to\infty} P(|X_n - \xi| \geq \varepsilon) = 0$. As $\varepsilon > 0$ was arbitrary, we obtain $X_n \xrightarrow{P} \xi$. □

Theorem 3.6.3 (The delta method)**.** *Let (X_n) be a sequence of random variables, and let ξ and σ be real with $\sigma > 0$. Assume that X_n is asymptotically normal with mean ξ and variance $\frac{1}{n}\sigma^2$. Let $f : \mathbb{R} \to \mathbb{R}$ be measurable and differentiable in ξ. Then $f(X_n)$ is asymptotically normal with mean $f(\xi)$ and variance $\frac{1}{n}\sigma^2 f'(\xi)^2$.*

Proof. By our assumptions, $\sqrt{n}(X_n - \xi) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$. Our objective is to demonstrate that $\sqrt{n}(f(X_n) - f(\xi)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2 f'(\xi)^2)$. Note that when defining $R : \mathbb{R} \to \mathbb{R}$ by putting $R(x) = f(x) - f(\xi) - f'(\xi)(x - \xi)$, we obtain $f(x) = f(\xi) + f'(\xi)(x - \xi) + R(x)$, and in particular

$$\begin{aligned}\sqrt{n}(f(X_n) - f(\xi)) &= \sqrt{n}(f'(\xi)(X_n - \xi) + R(X_n))\\ &= f'(\xi)\sqrt{n}(X_n - \xi) + \sqrt{n}R(X_n) \qquad (3.24)\end{aligned}$$

As $\sqrt{n}(X_n - \xi) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$, Lemma 3.1.8 shows that $f'(\xi)\sqrt{n}(X_n - \xi) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2 f'(\xi)^2)$. Therefore, by Lemma 3.3.2, the result will follow if we can prove $\sqrt{n}R(X_n) \xrightarrow{P} 0$. To this end, let $\varepsilon > 0$. Note that as f is differentiable at ξ, we have

$$\lim_{x\to\xi} \frac{R(x)}{x - \xi} = \lim_{x\to\xi} \frac{f(x) - f(\xi)}{x - \xi} - f'(\xi) = 0.$$

Defining $r(x) = R(x)/(x - \xi)$ when $x \neq \xi$ and $r(\xi) = 0$, we then find that r is measurable and continuous at ξ, and $R(x) = (x - \xi)r(x)$. In particular, there exists $\delta > 0$ such that whenever $|x - \xi| < \delta$, we have $|r(x)| < \varepsilon$. It then also holds that if $|r(x)| \geq \varepsilon$, we have $|x - \xi| \geq \delta$. From this and Lemma 3.6.2, we get $\limsup_{n\to\infty} P(|r(X_n)| \geq \varepsilon) \leq \limsup_{n\to\infty} P(|X_n - \xi| \geq \delta) = 0$, so $r(X_n) \xrightarrow{P} 0$. As the multiplication mapping $(x, y) \mapsto xy$ is continuous, we obtain by Theorem 3.3.3 that $\sqrt{n}R(X_n) = \sqrt{n}(X_n - \xi)r(X_n) \xrightarrow{\mathcal{D}} 0$, and so by Lemma 3.3.1, we get $\sqrt{n}R(X_n) \xrightarrow{P} 0$. Combining our conclusions with (3.24), Lemma 3.3.2 now shows that $\sqrt{n}(f(X_n) - f(\xi)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2 f'(\xi)^2)$, completing the proof. □

Using the preceeding results, we may now give an example of a practical application of the central limit theorem and asymptotic normality.

Example 3.6.4. As in Example 1.5.4, consider a measurable space $(\Omega, \mathcal{F})$ endowed with a sequence of random variables (X_n). Assume given for each $\xi \in \mathbb{R}$ a probability measure P_ξ such that for the probability space $(\Omega, \mathcal{F}, P_\xi)$, (X_n) consists of independent and identically distributed variables with mean ξ and unit variance. We may then define an estimator of the mean by putting $\hat{\xi}_n = \frac{1}{n}\sum_{k=1}^n X_k$. As the variables have second moment, Theorem 3.5.3 shows that $\hat{\xi}_n$ is asymptotically normal with mean ξ and variance $\frac{1}{n}$.

This intuitively gives us some information about the distribution of $\hat{\xi}_n$ for large $n \geq 1$. In order to make practical use of this, let $0 < \gamma < 1$. We consider the problem of obtaining a confidence interval for the parameter ξ with confidence level approximating γ as n tends to infinity. The statement that $\hat{\xi}_n$ is asymptotically normal with the given parameters means that $\sqrt{n}(\hat{\xi}_n - \xi) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$. With Φ denoting the cumulative distribution function for the standard normal distribution, we obtain $\lim_{n\to\infty} P_\xi(\sqrt{n}(\hat{\xi}_n - \xi) \leq x) = \Phi(x)$ for all $x \in \mathbb{R}$ by Lemma 3.2.1. Now let z_γ be such that $\Phi(-x) = (1-\gamma)/2$, meaning that we have $z_\gamma = -\Phi^{-1}((1-\gamma)/2)$. As $(1-\gamma)/2 < 1/2$, $z_\gamma > 0$. Also, $\Phi(z_\gamma) = 1 - \Phi(-z_\gamma) = 1 - (1-\gamma)/2$, and so we obtain

$$\lim_{n\to\infty} P_\xi(-z_\gamma \leq \sqrt{n}(\hat{\xi}_n - \xi) \leq z_\gamma) = \Phi(z_\gamma) - \Phi(-z_\gamma) = \gamma.$$

However, we also have

$$\begin{aligned} P_\xi(-z_\gamma \leq \sqrt{n}(\hat{\xi}_n - \xi) \leq z_\gamma) &= P_\xi(-z_\gamma/\sqrt{n} \leq \hat{\xi}_n - \xi \leq z_\gamma/\sqrt{n}) \\ &= P_\xi(\hat{\xi}_n - z_\gamma/\sqrt{n} \leq \xi \leq \hat{\xi}_n + z_\gamma/\sqrt{n}), \end{aligned}$$

so if we define $I_\gamma = (\hat{\xi}_n - z_\gamma/\sqrt{n}, \hat{\xi}_n + z_\gamma/\sqrt{n})$, we have $\lim_{n\to\infty} P_\xi(\xi \in I_\gamma) = \gamma$ for all $\xi \in \mathbb{R}$. This means that asymptotically speaking, there is probability γ that I_γ contains ξ. In particular, as $\Phi(-1.96) \approx 2.5\%$, we find that $(\hat{\xi}_n - 1.96/\sqrt{n}, \hat{\xi}_n + 1.96/\sqrt{n})$ is a confidence interval which a confidence level approaching a number close to 95% as n tends to infinity. ∘

3.7 Higher dimensions

Throughout this chapter, we have worked with weak convergence of random variables with values in $\mathbb{R}$, as well as probability measures on $(\mathbb{R}, \mathcal{B})$. Among our most important results are the results that weak convergence is equivalent to convergence of characteristic functions, the interplay between convergence in distribution and convergence in probability, the central limit theorems and our results on asymptotic normality. The theory of weak convergence and all of its major results can be extended to the more general context of random variables with values in $\mathbb{R}^d$ and probability measures on $(\mathbb{R}^d, \mathcal{B}_d)$ for $d \geq 1$, and to a large degree, it is these multidimensional results which are most useful in practice. In this section, we state the main results from the multidimensional theory of weak convergence without proof.

In the following, $C_b(\mathbb{R}^d)$ denotes the set of continuous, bounded mappings $f : \mathbb{R}^d \to \mathbb{R}$.

Definition 3.7.1. *Let (μ_n) be a sequence of probability measures on $(\mathbb{R}^d, \mathcal{B}_d)$, and let μ be another probability measure. We say that μ_n converges weakly to μ and write $\mu_n \xrightarrow{wk} \mu$ if it holds for all $f \in C_b(\mathbb{R}^d)$ that $\lim_{n\to\infty} \int f \,\mathrm{d}\mu_n = \int f \,\mathrm{d}\mu$.*

As in the univariate case, the limit measure is determined uniquely. Also, we say that a sequence of random variables (X_n) with values in $\mathbb{R}^d$ converges in distribution to a random variable X with values in $\mathbb{R}^d$ or a probability measure μ on $(\mathbb{R}^d, \mathcal{B}_d)$ if the distributions converge weakly. The following analogue of Lemma 3.1.8 then holds.

Lemma 3.7.2. *Let (μ_n) be a sequence of probability measures on $(\mathbb{R}^d, \mathcal{B}_d)$, and let μ be another probability measure. Let $h : \mathbb{R}^d \to \mathbb{R}^p$ be some continuous mapping. If it holds that $\mu_n \xrightarrow{wk} \mu$, then it also holds that $h(\mu_n) \xrightarrow{wk} h(\mu)$.*

An important result which relates multidimensional weak convergence to one-dimensional weak convergence is the following result. In Theorem 3.7.3, θ^t denotes transpose, and the mapping $x, y \mapsto x^t y$ for $x, y \in \mathbb{R}^d$ thus corresponds to the ordinary inner product on $\mathbb{R}^d$.

Theorem 3.7.3 (Cramér-Wold's device)**.** *Let (X_n) be a sequence of random variables with values in $\mathbb{R}^d$, and let X be some other such variable. Then $X_n \xrightarrow{\mathcal{D}} X$ if and only if it holds for all $\theta \in \mathbb{R}^d$ that $\theta^t X_n \xrightarrow{\mathcal{D}} \theta^t X$.*

Letting $(X_n)_{n\geq 1}$ be a sequence of random variables with values in $\mathbb{R}^d$ and letting X be some other such variable, we may define a multidimensional analogue of convergence in probability by saying that X_n converges in probability to X and writing $X_n \xrightarrow{P} X$ when

$\lim_{n\to\infty} P(\|X_n - X\| \geq \varepsilon) = 0$ for all $\varepsilon > 0$, where $\|\cdot\|$ is some norm on $\mathbb{R}^d$. We then have that $X_n \xrightarrow{P} x$ if and only if $X_n \xrightarrow{\mathcal{D}} x$, and the following multidimensional version of Theorem 3.3.3 holds.

Theorem 3.7.4. *Let (X_n, Y_n) be a sequence of random variables with values in $\mathbb{R}^d$ and $\mathbb{R}^p$, respectively, let X be some other variable with values in $\mathbb{R}^d$ and let $y \in \mathbb{R}^p$. Consider a continuous mapping $h : \mathbb{R}^d \times \mathbb{R}^p \to \mathbb{R}^m$. If $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{P} y$, then it holds that $h(X_n, Y_n) \xrightarrow{\mathcal{D}} h(X, y)$.*

We may also define characteristic functions in the multidimensional setting. Let μ be a probability measure on $(\mathbb{R}^d, \mathcal{B}_d)$. We define the characteristic function for μ to be the mapping $\varphi : \mathbb{R}^d \to \mathbb{C}$ defined by $\varphi(\theta) = \int e^{i\theta^t x} \, \mathrm{d}\mu(x)$. As in the one-dimensional case, the characteristic function determines the probability measure uniquely, and weak convergence is equivalent to pointwise convergence of probability measures.

The central limit theorem also holds in the multidimensional case.

Theorem 3.7.5. *Let (X_n) be a sequence of independent and identically distributed random variables with values in $\mathbb{R}^d$ with mean vector ξ and positive semidefinite variance matrix Σ. It then holds that*

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} (X_k - \xi) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma),$$

where $\mathcal{N}(0, \Sigma)$ denotes the normal distribution with mean zero and variance matrix Σ.

As in the one-dimensional case, we may introduce a notion of asymptotic normality. For a sequence of random variables (X_n) with values in $\mathbb{R}^d$, we say that X_n is asymptotically normal with mean ξ and variance $\frac{1}{n}\Sigma$ if $\sqrt{n}(X_n - \xi) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma)$, and in this case, we write $X_n \overset{as}{\sim} \mathcal{N}(\xi, \frac{1}{n}\Sigma)$. If $X_n \overset{as}{\sim} \mathcal{N}(\xi, \frac{1}{n}\Sigma)$, it also holds that $X_n \xrightarrow{P} \xi$. Also, we have the following version of the delta method in the multidimensional case.

Theorem 3.7.6. *Let (X_n) be a sequence of random variables with values in $\mathbb{R}^d$, and assume that X_n is asymptotically normal with mean ξ and variance $\frac{1}{n}\Sigma$. Let $f : \mathbb{R}^d \to \mathbb{R}^p$ be measurable and differentiable in ξ, then $f(X_n)$ is asymptotically normal with mean $f(\xi)$ and variance $\frac{1}{n}Df(\xi)\Sigma Df(\xi)^t$, where $Df(\xi)$ is the Jacobian of f at ξ, that is, the $p \times d$ matrix consisting of the partial derivatives of f at ξ.*

Note that Theorem 3.7.6 reduces to Theorem 3.6.3 for $d = p = 1$, and in the one-dimensional case, the products in the expression for the asymptotic variance commute, leading to a simpler expression in the one-dimensional case than in the multidimensional case.

To show the strengh of the multidimensional theory, we give the following example, extending Example 3.6.4.

Example 3.7.7. As in Example 3.6.4, consider a measurable space $(\Omega, \mathcal{F})$ endowed with a sequence of random variables (X_n). Let $\Theta = \mathbb{R} \times (0, \infty)$. Assume for each $\theta = (\xi, \sigma^2)$ that we are given a probability measure P_θ such that for the probability space $(\Omega, \mathcal{F}, P_\theta)$, (X_n) consists of independent and identically distributed variables with fourth moment, and with mean ξ and variance σ^2. As in Example 1.5.4, we may then define estimators of the mean and variance based on n samples by putting

$$\hat{\xi}_n = \frac{1}{n}\sum_{k=1}^n X_k \quad \text{and} \quad \hat{\sigma}_n^2 = \frac{1}{n}\sum_{k=1}^n X_k^2 - \left(\frac{1}{n}\sum_{k=1}^n X_k\right)^2.$$

Now note that the variables (X_n, X_n^2) also are independent and identically distributed, and with ρ denoting $\operatorname{Cov}(X_n, X_n^2)$ and η^2 denoting VX_n^2, we have

$$E\begin{pmatrix} X_n \\ X_n^2 \end{pmatrix} = \begin{pmatrix} \xi \\ \sigma^2 + \xi^2 \end{pmatrix} \quad \text{and} \quad V\begin{pmatrix} X_n \\ X_n^2 \end{pmatrix} = \begin{pmatrix} \sigma^2 & \rho \\ \rho & \eta^2 \end{pmatrix}. \tag{3.25}$$

Let μ and Σ denote the mean and variance, respectively, in (3.25). By $\overline{X}_n$ and $\overline{X_n^2}$, we denote $\frac{1}{n}\sum_{k=1}^n X_k$ and $\frac{1}{n}\sum_{k=1}^n X_k^2$, respectively. Using Theorem 3.7.5, we then obtain that $(\overline{X}_n, \overline{X_n^2})$ is asymptotically normal with parameters $(\mu, \frac{1}{n}\Sigma)$.

We will use this multidimensional relationship to find the asymptotic distributions of $\hat{\xi}_n$ and $\hat{\sigma}_n^2$, and we will do so by applying Theorem 3.7.6. To this end, we first consider the mapping $f : \mathbb{R}^2 \to \mathbb{R}$ given by $f(x, y) = x$. Note that we have $Df(x, y) = (1 \;\; 0)$. As $\hat{\xi}_n = f(\overline{X}_n, \overline{X_n^2})$, Theorem 3.7.6 yields that $\hat{\xi}_n$ is asymptotically normal with mean $f(\mu) = \xi$ and variance

$$\frac{1}{n}Df(\mu)\Sigma Df(\mu)^t = \frac{1}{n}\begin{pmatrix} 1 & 0 \end{pmatrix}\begin{pmatrix} \sigma^2 & \rho \\ \rho & \eta^2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{n}\sigma^2,$$

in accordance with what we would have obtained by direct application of Theorem 3.5.3. Next, we consider the variance estimator. Define $g : \mathbb{R}^2 \to \mathbb{R}$ by putting $g(x, y) = y - x^2$. We then have $Dg(x, y) = (-2x \;\; 1)$. As $\hat{\sigma}_n^2 = g(\overline{X}_n, \overline{X_n^2})$, Theorem 3.7.6 shows that $\hat{\sigma}_n^2$ is asymptotically normal with mean $g(\mu) = \sigma^2$ and variance

$$\frac{1}{n}Dg(\mu)\Sigma Dg(\mu)^t = \frac{1}{n}\begin{pmatrix} -2\xi & 1 \end{pmatrix}\begin{pmatrix} \sigma^2 & \rho \\ \rho & \eta^2 \end{pmatrix}\begin{pmatrix} -2\xi \\ 1 \end{pmatrix} = \frac{1}{n}(4\xi^2\sigma^2 - 4\xi\rho + \eta^2).$$

Thus, applying Theorem 3.7.5 and Theorem 3.7.6, we have proven that both $\hat{\xi}_n$ and $\hat{\sigma}_n^2$ are asymptotically normal, and we have identified the asymptotic parameters.

Next, consider some $0 < \gamma < 1$. We will show how to construct a confidence interval for ξ which has a confidence level approximating γ as n tends to infinity. Note that this was already accomplished in Example 3.6.4 in the case where the variance was known and equal to one. In this case, we have no such assumptions. Now, we already know that $\hat{\xi}_n$ is asymptotically normal with parameters $(\xi, \frac{1}{n}\sigma^2)$, meaning that $\sqrt{n}(\hat{\xi}_n - \xi) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$. Next, note that as $\hat{\sigma}_n^2$ ias asymptotically normal with mean σ^2, Lemma 3.6.2 shows that $\hat{\sigma}_n^2 \xrightarrow{P} \sigma^2$. Therefore, using Theorem 3.3.3, we find that $\sqrt{n}(\hat{\xi} - \xi)/\sqrt{\hat{\sigma}_n^2} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$. We may now proceed as in Example 3.6.4 and note that with Φ denoting the cumulative distribution function for the standard normal distribution, $\lim_{n\to\infty} P(\sqrt{n}(\hat{\xi}_n - \xi)/\sqrt{\hat{\sigma}_n^2} \le x) = \Phi(x)$ for all $x \in \mathbb{R}$ by Lemma 3.2.1. Putting $z_\gamma = -\Phi^{-1}((1-\gamma)/2)$, we then obtain $z_\gamma > 0$ and $\Phi(z_\gamma) - \Phi(-z_\gamma) = \gamma$, and if we define $I_\gamma = (\hat{\xi}_n - z_\gamma\sqrt{\hat{\sigma}_n^2/n}, \hat{\xi}_n + z_\gamma\sqrt{\hat{\sigma}_n^2/n})$, we then obtain

$$\begin{aligned}
\lim_{n\to\infty} P_\theta(\xi \in I_\gamma) &= \lim_{n\to\infty} P_\theta(\hat{\xi}_n - \sqrt{\hat{\sigma}_n^2}z_\gamma/\sqrt{n} \le \xi \le \hat{\xi}_n + \sqrt{\hat{\sigma}_n^2}z_\gamma/\sqrt{n}) \\
&= \lim_{n\to\infty} P_\theta(-\sqrt{\hat{\sigma}_n^2}z_\gamma/\sqrt{n} \le \hat{\xi}_n - \xi \le \sqrt{\hat{\sigma}_n^2}z_\gamma/\sqrt{n}) \\
&= \lim_{n\to\infty} P_\theta(-z_\gamma \le \sqrt{n}(\hat{\xi}_n - \xi)/\sqrt{\hat{\sigma}_n^2} \le z_\gamma) \\
&= \Phi(z_\gamma) - \Phi(-z_\gamma) = \gamma,
\end{aligned}$$

so I_γ is a confidence interval for ξ such that asymptotically speaking, there is probability γ that I_γ contains ξ. ◦

3.8 Exercises

Exercise 3.1. Let (θ_n) be a sequence of positive numbers. Let μ_n denote the uniform distribution on $[0, \theta_n]$. Show that μ_n converges weakly if and only if θ_n is convergent. In the affirmative case, identify the limiting distribution. ◦

Exercise 3.2. Let (μ_n) be a sequence of probability measures concentrated on $\mathbb{N}_0$, and let μ be another such probability measure. Show that $\mu_n \xrightarrow{wk} \mu$ if and only if it holds that $\lim_{n\to\infty} \mu_n(\{k\}) = \mu(\{k\})$ for all $k \ge 0$. ◦

Exercise 3.3. Let μ_n denote the Student's t-distribution with n degrees of freedom, that is, the distribution with density f_n given by $f_n(x) = \frac{\Gamma(n+\frac{1}{2})}{\sqrt{2n\pi}\Gamma(n)}(1 + \frac{x^2}{2n})^{-(n+\frac{1}{2})}$. Show that μ_n converges weakly to the standard normal distribution. ◦

Exercise 3.4. Let (p_n) be a sequence in $(0,1)$, and let μ_n be the binomial distribution with success probability p_n and length n. Assume that $\lim_{n\to\infty} np_n = \lambda$ for some $\lambda \ge 0$. Show that if $\lambda > 0$, then μ_n converges weakly to the Poisson distribution with parameter λ. Show that if $\lambda = 0$, then μ_n converges weakly to the Dirac measure at zero. ◦

Exercise 3.5. Let X_n be a random variable which is Beta distributed with shape parameters (n, n). Define $Y_n = \sqrt{8n}(X_n - \frac{1}{2})$. Show that Y_n has density with respect to the Lebesgue measure. Show that the densities converge pointwise to the density of the standard normal distribution. Argue that Y_n converges in distribution to the standard normal distribution. ◦

Exercise 3.6. Let μ be a probability measure on $(\mathbb{R}, \mathcal{B})$ with cumulative distribution function F. Let $q : (0,1) \to \mathbb{R}$ be a quantile function for μ, meaning that for all $0 < p < 1$, it holds that $F(q(p)-) \leq p \leq F(q(p))$. Let μ_n be the probability measure on $(\mathbb{R}, \mathcal{B})$ given by putting $\mu_n(B) = \frac{1}{n}\sum_{k=1}^n 1_{(q(k/(n+1)) \in B)}$ for $B \in \mathcal{B}$. Show that μ_n converges weakly to μ. ◦

Exercise 3.7. Let (ξ_n) and (σ_n) be sequences in $\mathbb{R}$, where $\sigma_n > 0$. Let μ_n denote the normal distribution with mean ξ_n and variance σ_n^2. Show that μ_n converges weakly if and only if ξ_n and σ_n both converge. In the affirmative case, identify the limiting distribution. ◦

Exercise 3.8. Let (μ_n) be a sequence of probability measures on $(\mathbb{R}, \mathcal{B})$ such that μ_n has cumulative distribution function F_n. Let μ be some other probability measure with cumulative distribution function F. Assume that F is continuous and assume that μ_n converges weakly to μ. Let (x_n) be a sequence of real numbers converging to some point x. Show that $\lim_{n\to\infty} F_n(x_n) = F(x)$. ◦

Exercise 3.9. Let μ_n be the measure on $(\mathbb{R}, \mathcal{B})$ concentrated on $\{k/n \mid k \geq 1\}$ such that $\mu_n(\{k/n\}) = \frac{1}{n}(1 - \frac{1}{n})^{k-1}$ for each $k \in \mathbb{N}$. Show that μ_n is a probability measure and that μ_n converges weakly to the standard exponential distribution. ◦

Exercise 3.10. Calculate the characteristic function of the binomial distribution with success parameter p and length n. ◦

Exercise 3.11. Calculate an explicit expression for the characteristic function of the Poisson distribution with parameter λ. ◦

Exercise 3.12. Consider a probability space endowed with two independent variables X and Y with distributions μ and ν, respectively, where μ has characteristic function φ and ν has characteristic function ϕ. Show that the variable XY has characteristic function ψ given by $\psi(\theta) = \int \varphi(\theta y) \, d\nu(y)$. ◦

Exercise 3.13. Consider a probability space endowed with four independent variables X, Y, Z and W, all standard normally distributed. Calculate the characteristic function of $XY - ZW$ and argue that $XY - ZW$ follows a Laplace distribution. ◦

Exercise 3.14. Assume (X_n) is a sequence of independent random variables. Assume that there exists $\beta > 0$ such that $|X_n| \leq \beta$ for all $n \geq 1$. Define $S_n = \sum_{k=1}^n X_k$. Prove that if it holds that $\sum_{n=1}^\infty VX_n$ is infinite, then $(S_n - ES_n)/\sqrt{VS_n}$ converges in distribution to the standard normal distribution. ◦

Exercise 3.15. Let (X_n) be a sequence of independent random variables. Let $\varepsilon > 0$. Show that if $\sum_{k=1}^n X_k$ converges almost surely as n tends to infinity, then the following three series are convergent: $\sum_{n=1}^\infty P(|X_n| > \varepsilon)$, $\sum_{n=1}^\infty EX_n 1_{(|X_n| \leq \varepsilon)}$ and $\sum_{n=1}^\infty VX_n 1_{(|X_n| \leq \varepsilon)}$. ◦

Exercise 3.16. Consider a measurable space $(\Omega, \mathcal{F})$ endowed with a sequence (X_n) of random variables as well as a family of probability measures $(P_\lambda)_{\lambda>0}$ such that under P_λ, (X_n) consists of independent and identically distributed variables such that X_n follows a Poisson distribution with mean λ for some $\lambda > 0$. Let $\overline{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$. Find a mapping $f : (0, \infty) \to (0, \infty)$ such that for each $\lambda > 0$, it holds that under P_λ, $f(\overline{X}_n)$ is asymptotically normal with mean $f(\lambda)$ and variance $\frac{1}{n}$. ◦

Exercise 3.17. Let (X_n) be a sequence of independent random variables such that X_n has mean ξ and unit variance. Put $S_n = \sum_{k=1}^n X_k$. Let $\alpha > 0$. Show that $(S_n - n\xi)/n^\alpha \xrightarrow{P} 0$ if and only if $\alpha > 1/2$. ◦

Exercise 3.18. Let $\theta > 0$ and let (X_n) be a sequence of independent and identically distributed random variables such that X_n follows a normal distribution with mean θ and variance θ. The maximum likelihood estimator for estimation of θ based on n samples is $\hat{\theta}_n = -\frac{1}{2} + (\frac{1}{4} + \frac{1}{n}\sum_{k=1}^n X_k^2)^{1/2}$. Show that $\hat{\theta}_n$ is asymptotically normal with mean θ and variance $\frac{1}{n}\frac{4\theta^3+2\theta^2}{4\theta^2+2\theta+1}$. ◦

Exercise 3.19. Let $\mu > 0$ and let (X_n) be a sequence of independent and identically distributed random variables such that X_n follows an exponential distribution with mean $1/\mu$. Let $\overline{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$. Show that $\overline{X}_n$ and $\overline{X}_n^{-1}$ are asymptotically normal and identify the asymptotic parameters. Define $Y_n = \frac{1}{\log n}\sum_{k=1}^n \frac{X_k}{k}$. Show that $Y_n \xrightarrow{P} 1/\mu$. ◦

Exercise 3.20. Let $\theta > 0$ and let (X_n) be a sequence of independent and identically distributed random variables such that X_n follows a uniform distribution on $[0, \theta]$. Define $\overline{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$. Show that $\overline{X}_n$ is asymptotically normal with mean μ and variance $\frac{1}{n}\mu^2$.

Next, put $Y_n = \frac{4}{n^2}\sum_{k=1}^n kX_k$. Demonstrate that $Y_n \xrightarrow{P} \theta$. Use Lyapounov's central limit theorem to show that $(Y_n - \theta)/(\sqrt{4\theta^2/9n})$ converges to a standard normal distribution. ∘

Exercise 3.21. Let (X_n, Y_n) be a sequence of independent and identically distributed variables such that for each $n \geq$, X_n and Y_n are independent, where X_n follows a standard normal distribution and Y_n follows an exponential distribution with mean α for some $\alpha > 0$. Define $S_n = \frac{1}{n}\sum_{k=1}^n X_k + Y_k$ and $T_n = \frac{1}{n}\sum_{k=1}^n X_k^2$. Show that (S_n, T_n) is asymptotically normally distributed and identify the asymptotic parameters. Show that $S_n/\sqrt{T_n}$ is asymptotically normally distributed and identify the asymptotic parameters. ∘

Exercise 3.22. Let (X_n) be a sequence of independent and identically distributed variables such that X_n follows a normal distribution with mean μ and variance σ^2 for some $\sigma > 0$. Assume that $\mu \neq 0$. Define $\overline{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$ and $S_n^2 = \frac{1}{n}\sum_{k=1}^n (X_k - \overline{X}_n)^2$. Show that $S_n/\overline{X}_n$ is asymptotically normally distributed and identify the asymptotic parameters. ∘

Chapter 4

Signed measures and conditioning

In this chapter, we will consider two important but also very distinct topics: Decompositions of signed measures and conditional expectations. The topics are only related by virtue of the fact that we will use results from the first section to prove the existence of the conditional expectations to be defined in the following section. In the first section, the framework is a measurable space that will be equipped with a so–called signed measure. In the rest of the chapter, the setting will a probability space endowed with a random variable.

4.1 Decomposition of signed measures

In this section, we first introduce a generalization of bounded measures, namely bounded, signed measures, where negative values are allowed. We then show that a signed measure can always be decomposed into a difference between two positive, bounded measures. Afterwards we prove the main result of the section, stating how to decompose a bounded, signed measure with respect to a bounded, positive measure. Finally we will show the Radon–Nikodym theorem, which will be crucial in Section 4.2 in order to prove the existence of conditional expectations.

In the rest of the section, we let $(\Omega, \mathcal{F})$ be a measurable space. Recall that $\mu : \mathcal{F} \to [0, \infty]$ is a measure on $(\Omega, \mathcal{F})$ if $\mu(\emptyset) = 0$ and for all disjoint sequences $F_1, F_2, \ldots$ it holds that

$$\mu\Big(\bigcup_{n=1}^{\infty} F_n\Big) = \sum_{n=1}^{\infty} \mu(F_n)\,.$$

If $\mu(\Omega) < \infty$ we say that μ is a finite measure. However, in the context of this section, we shall most often use the name *bounded, positive measure*. A natural generalisation of a bounded, positive measure is to allow negative values. Hence we consider the following definition:

Definition 4.1.1. *A bounded, signed measure ν on $(\Omega, \mathcal{F})$ is a map $\nu : \mathcal{F} \to \mathbb{R}$ such that*

(1) $\sup\{|\nu(F)| \mid F \in \mathcal{F}\} < \infty$,

(2) $\nu(\bigcup_{n=1}^{\infty} F_n) = \sum_{n=1}^{\infty} \nu(F_n)$ *for all pairwise disjoint $F_1, F_2, \ldots \in \mathcal{F}$.*

Note that condition (2) is similar to the σ–additivity condition for positive measures. Condition (1) ensures that ν is bounded.

A bounded, signed measure has further properties that resemble properties of positive measures:

Theorem 4.1.2. *Assume that ν is a bounded, signed measure on $(\Omega, \mathcal{F})$. Then*

(1) $\nu(\emptyset) = 0\,.$

(2) ν is finitely additive: If $F_1, \ldots, F_N \in \mathcal{F}$ are disjoint sets, then

$$\nu\big(\bigcup_{n=1}^{N} F_n\big) = \sum_{n=1}^{N} \nu(F_n)\,.$$

(3) ν is continuous: If $F_n \uparrow F$ or $F_n \downarrow F$, with $F_1, F_2, \ldots \in \mathcal{F}$, then

$$\nu(F_n) \to \nu(F)$$

Proof. To prove (1) let $F_1 = F_2 = \cdots = \emptyset$ in the σ–additivity condition. Then we can utilize the simple fact $\emptyset = \bigcup_{n=1}^{\infty} \emptyset$ such that

$$\nu(\emptyset) = \nu\big(\bigcup_{n=1}^{\infty} \emptyset\big) = \sum_{n=1}^{\infty} \nu(\emptyset)$$

which can only be true if $\nu(\emptyset) = 0$.

Considering the second result, let $F_{N+1} = F_{N+2} = \cdots = \emptyset$ and apply the σ–additivity again such that

$$\nu\Big(\bigcup_{n=1}^{N} F_n\Big) = \nu\Big(\bigcup_{n=1}^{N} F_n \cup \bigcup_{n=N+1}^{\infty} \emptyset\Big) = \sum_{n=1}^{N} \nu(F_n) + \sum_{n=N+1}^{\infty} 0 = \sum_{n=1}^{N} \nu(F_n)\,.$$

Finally we demonstrate the third result in the case where $F_n \uparrow F$. Define $G_1 = F_1$, $G_2 = F_2 \setminus F_1$, $G_3 = F_3 \setminus F_2, \ldots$. Then $G_1, G_2, \ldots$ are disjoint with

$$\bigcup_{n=1}^{N} G_n = F_N\,, \qquad \bigcup_{n=1}^{\infty} G_n = F\,,$$

so

$$\nu(F_N) = \nu\big(\bigcup_{n=1}^{N} G_n\big) = \sum_{n=1}^{N} \nu(G_n) \to \sum_{n=1}^{\infty} \nu(G_n) = \nu(F) \quad \text{as } N \to \infty.$$

□

From the definition of a bounded, signed measure and Theorem 4.1.2 we almost immediately see that bounded, signed measures with non–negative values are in fact bounded, positive measures.

Theorem 4.1.3. *Assume that ν is a bounded, signed measure on $(\Omega, \mathcal{F})$. If ν only has values in $[0, \infty)$, then ν is a bounded, positive measure.*

Proof. That ν is a measure in the classical sense follows since it satisfies the σ–additivity condition, and we furthermore have $\nu(\emptyset) = 0$ according to (1) in Theorem 4.1.2. That $\nu(\Omega) < \infty$ is obviously a consequence of (1) in Definition 4.1.1. □

Example 4.1.4. Let $\Omega = \{1, 2, 3, 4\}$ and assume that ν is a bounded, signed measure on Ω given by

$$\nu(\{1\}) = 2 \quad \nu(\{2\}) = -1 \quad \nu(\{3\}) = 4 \quad \nu(\{4\}) = -2\,.$$

Then e.g.

$$\nu(\{1,2\}) = 1 \quad \nu(\{3,4\}) = 2$$

and

$$\nu(\Omega) = 3\,.$$

so we see that although $\{3\} \subsetneq \Omega$, it is possible that $\nu(\{3\}) > \nu(\Omega)$. Hence, the condition 1 in the definition is indeed meaningful: Only demanding that $\nu(\Omega) < \infty$ as for positive measures would not ensure that ν is bounded on all sets, and in particular $\nu(\Omega)$ is not necessarily an upper bound for ν. ◦

Recall, that if μ is a bounded, positive measure then, for $F_1, F_2 \in \mathcal{F}$ with $F_1 \subseteq F_2$, it holds that $\mu(F_1) \leq \mu(F_2)$. Hence, condition (1) in Definition 4.1.1 will (for μ) be equivalent to $\mu(\Omega) < \infty$.

If ν is a bounded, signed measure and $F_1, F_2 \in \mathcal{F}$ with $F_1 \subseteq F_2$ then it need not hold that $\nu(F_1) \leq \nu(F_2)$:

In general we have using the finite additivity that

$$\nu(F_2) = \nu(F_1) + \nu(F_2 \backslash F_1),$$

but $\nu(F_2 \backslash F_1) \geq 0$ and $\nu(F_2 \backslash F_1) < 0$ are both possible.

Recall from classical measure theory that new positive measures can be constructed by integrating non–negative functions with respect to other positive measures. Similarly, we can construct a bounded, signed measure by integrating an integrable function with respect to a bounded, positive measure.

Theorem 4.1.5. *Let μ be a bounded, positive measure on $(\Omega, \mathcal{F})$ and let $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ be a μ-integrable function, i.e.*

$$\int |f| \, \mathrm{d}\mu < \infty.$$

Then

$$\nu(F) = \int_F f \, \mathrm{d}\mu \qquad (F \in \mathcal{F}) \tag{4.1}$$

defines a bounded, signed measure on $(\Omega, \mathcal{F})$. Furthermore it holds that ν is a bounded, positive measure if and only if $f \geq 0$ μ-a.e. (almost everywhere).

Proof. For all $F \in \mathcal{F}$ we have

$$|\nu(F)| \leq \int_F |f| \, \mathrm{d}\mu \leq \int |f| \, \mathrm{d}\mu < \infty$$

which gives (1). To obtain that (2) is satisfied, let $F_1, F_2, \ldots \in \mathcal{F}$ be disjoint and define $F = \cup F_n$. Observe that

$$|1_{\cup_{n=1}^N F_n} f| \leq |f|$$

for all $N \in \mathbb{N}$. Then dominated convergence yields

$$\nu\Big(\bigcup_{n=1}^{\infty} F_n\Big) = \nu(F) = \int_F f \, \mathrm{d}\mu = \int \lim_{N\to\infty} 1_{\bigcup_{n=1}^N F_n} f \, \mathrm{d}\mu = \lim_{N\to\infty} \int 1_{\bigcup_{n=1}^N F_n} f \, \mathrm{d}\mu$$
$$= \lim_{N\to\infty} \sum_{n=1}^{N} \int_{F_n} f \, \mathrm{d}\mu = \lim_{N\to\infty} \sum_{n=1}^{N} \nu(F_n) = \sum_{n=1}^{\infty} \nu(F_n) \,.$$

The last statement follows from Theorem 4.1.3, since $f \geq 0$ μ–a.e. implies that $\nu(F) \geq 0$ for all $F \in \mathcal{F}$. □

In the following definition we introduce two possible relations between a signed measure and a positive measure. A main result in this chapter will be that the two definitions are equivalent.

Definition 4.1.6. *Assume that μ is a bounded, positive measure, and that ν is a bounded, signed measure on $(\Omega, \mathcal{F})$.*

(1) ν is absolutely continuous *with respect to μ, (we write $\nu \ll \mu$) if $\mu(F) = 0$ implies $\nu(F) = 0$.*

(2) ν has density *with respect to μ if there exists a μ-integrable function f (the density), such that (4.1) holds. If ν has density with respect to μ we write $\nu = f \cdot \mu$ and $f = \frac{d\nu}{d\mu}$. f is called the* Radon-Nikodym derivative *of ν with respect to μ.*

Lemma 4.1.7. *Assume that μ is a bounded, positive measure on $(\Omega, \mathcal{F})$ and that ν is a bounded signed measure on $(\Omega, \mathcal{F})$. If $\nu = f \cdot \mu$, then $\nu \ll \mu$.*

Proof. Choose $F \in \mathcal{F}$ with $\mu(F) = 0$. Then $1_F f = 0$ μ–a.e. so

$$\nu(F) = \int_F f \, \mathrm{d}\mu = \int 1_F f \, \mathrm{d}\mu = \int 0 \, \mathrm{d}\mu = 0 \,,$$

and the proof is complete. □

The following definition will be convenient as well:

Definition 4.1.8. *Assume that ν, ν_1, and ν_2 are bounded, signed measures on $(\Omega, \mathcal{F})$.*

(1) ν is concentrated *on $F \in \mathcal{F}$ if $\nu(G) = 0$ for all $G \in \mathcal{F}$ with $G \subseteq F^c$.*

(2) ν_1 *and* ν_2 *are* singular *(we write* $\nu_1 \perp \nu_2$*), if there exist disjoint sets* $F_1, F_2 \in \mathcal{F}$ *such that* ν_1 *is concentrated on* F_1 *and* ν_2 *is concentrated on* F_2*.*

Example 4.1.9. Let μ be a bounded, positive measure on $(\Omega, \mathcal{F})$, and assume that f is μ-integrable. Define the bounded, signed measure ν by $\nu = f \cdot \mu$. Then ν is concentrated on $(f \neq 0)$: Take $G \subseteq (f \neq 0)^c$, or equivalently $G \subseteq (f = 0)$. Then $1_G f = 0$, so $\nu(G) = \int 1_G f \,\mathrm{d}\mu = \int 0 \,\mathrm{d}\mu = 0$ and we have the result.

Now assume that both f_1 and f_2 are μ-integrable and define $\nu_1 = f_1 \cdot \mu$ and $\nu_2 = f_2 \cdot \mu$. Then it holds that $\nu_1 \perp \nu_2$ if $(f_1 \neq 0) \cap (f_2 \neq 0) = \emptyset$.

In fact, the result would even be true if we only have $\mu((f_1 \neq 0) \cap (f_2 \neq 0)) = 0$ (why?). ◦

Lemma 4.1.10. *Let* ν *be a bounded, signed measure on* $(\Omega, \mathcal{F})$*. If* ν *is concentrated on* $F \in \mathcal{F}$ *then* ν *is also concentrated on any* $G \in \mathcal{F}$ *with* $G \supseteq F$*.*

A bounded, positive measure μ *on* $(\Omega, \mathcal{F})$ *is concentrated on* $F \in \mathcal{F}$ *if and only if* $\mu(F^c) = 0$*.*

Proof. To show the first statement, assume that ν is concentrated on $F \in \mathcal{F}$, and let $G \in \mathcal{F}$ satisfy $G \supseteq F$. For any set $G' \subseteq G^c$ we have that $G' \subseteq F^c$, so by the definition we have $\nu(G') = 0$.

For the second result, we only need to show that if $\mu(F^c) = 0$, then μ is concentrated on F. So assume that $G \subseteq F^c$. Then, since μ is assumed to be a positive measure, we have

$$0 \leq \mu(G) \leq \mu(F^c) = 0$$

and we have that $\mu(G) = 0$ as desired. □

The following theorem is a deep result from classical measure theory, stating that any bounded, signed measure can be constructed as the difference between two bounded, positive measures.

Theorem 4.1.11. *(The Jordan-Hahn decomposition). A bounded, signed measure* ν *can be decomposed in exactly one way,*

$$\nu = \nu^+ - \nu^-,$$

where ν^+, ν^- *are positive, bounded measures and* $\nu^+ \perp \nu^-$*.*

Proof. The existence: Define $\lambda = \inf\{\nu(F) : F \in \mathcal{F}\}$. Then $-\infty < \lambda \leq 0$ and for all $n \in \mathbb{N}$ there exists $F_n \in \mathcal{F}$ with

$$\nu(F_n) \leq \lambda + \frac{1}{2^n}.$$

We first show that with $G = (F_n \text{ evt.}) = \cup_{n=1}^{\infty} \cap_{k=n}^{\infty} F_k$ it holds that

$$\nu(G) = \lambda\,. \tag{4.2}$$

Note that

$$\bigcap_{k=n}^{\infty} F_k \uparrow \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} F_k = G \quad \text{as} \quad n \to \infty$$

so since ν is continuous, we have

$$\nu(G) = \lim_{n\to\infty} \nu\Big(\bigcap_{k=n}^{\infty} F_k\Big)\,.$$

Similarly we have $\cap_{k=n}^{N} F_k \downarrow \cap_{k=n}^{\infty} F_k$ (as $N \to \infty$) so

$$\nu(G) = \lim_{n\to\infty} \lim_{N\to\infty} \nu\Big(\bigcap_{k=n}^{N} F_k\Big)\,.$$

Let n be fixed and suppose it is shown that for all $N \geq n$

$$\nu\Big(\bigcap_{k=n}^{N} F_k\Big) \leq \lambda + \sum_{k=n}^{N} \frac{1}{2^k}\,. \tag{4.3}$$

Since we have that $\sum_{k=1}^{\infty} \frac{1}{2^k} < \infty$, we must have that $\lim_{n\to\infty} \lim_{N\to\infty} \sum_{k=n}^{N} \frac{1}{2^k} = 0$. Hence

$$\lambda \leq \nu(G) \leq \lambda + \lim_{n\to\infty} \lim_{N\to\infty} \sum_{k=n}^{N} \frac{1}{2^k} = \lambda + 0 = \lambda\,.$$

So we have that $\nu(G) = \lambda$, if we can show (4.3). This is shown by induction for all $N \geq n$. If $N = n$ the result is trivial from the choice of F_n:

$$\nu\Big(\bigcap_{k=n}^{N} F_k\Big) = \nu(F_n) \leq \lambda + \frac{1}{2^n} = \lambda + \sum_{k=n}^{N} \frac{1}{2^k}\,.$$

If (4.3) is true for $N-1$ we obtain

$$\begin{aligned}\nu\Big(\bigcap_{k=n}^{N} F_k\Big) &= \nu\Big(\bigcap_{k=n}^{N-1} F_k \cap F_N\Big) = \nu\Big(\bigcap_{k=n}^{N-1} F_k\Big) + \nu(F_N) - \nu\Big(\bigcap_{k=n}^{N-1} F_k \cup F_N\Big) \\ &\leq \Big(\lambda + \sum_{k=n}^{N-1} \frac{1}{2^k}\Big) + \Big(\lambda + \frac{1}{2^N}\Big) - \lambda = \lambda + \sum_{k=n}^{N} \frac{1}{2^k}\,.\end{aligned}$$

In the inequality we have used (4.3) for $N-1$, the definition of F_N, and that $\nu(F) \geq \lambda$ for all $F \in \mathcal{F}$.

We have thus shown $\nu(G) = \lambda$ and may now define, for $F \in \mathcal{F}$,

$$\nu^-(F) = -\nu(F \cap G), \quad \nu^+(F) = \nu(F \cap G^c).$$

Obviously e.g.

$$\sup\{|\nu^-(F)| \ : \ F \in \mathcal{F}\} \leq \sup\{|\nu(F)| \ : \ F \in \mathcal{F}\} < \infty$$

and

$$\begin{aligned} \nu^-\Big(\bigcup_{n=1}^{\infty} F_n\Big) &= -\nu\Big((\bigcup_{n=1}^{\infty} F_n) \cap G\Big) = -\nu\Big(\bigcup_{n=1}^{\infty} (F_n \cap G)\Big) \\ &= -\sum_{n=1}^{\infty} \nu(F_n \cap G) = \sum_{n=1}^{\infty} \nu^-(F_n) \end{aligned}$$

for $F_1, F_2, \ldots \in \mathcal{F}$ disjoint sets, so ν^+ and ν^- are bounded, signed measures. It is easily seen (since G and G^c are disjoint) that $\nu = \nu^+ - \nu^-$. We furthermore have that ν^- is concentrated on G, since for $F \subseteq G^c$

$$\nu^-(F) = -\nu(F \cap G) = -\nu(\emptyset) = 0\,.$$

Similarly ν^+ is concentrated on G^c, so we must have $\nu^- \perp \nu^+$.

The existence part of the proof can now be completed by showing that $\nu^+ \geq 0, \nu^- \geq 0$. For $F \in \mathcal{F}$ we have $F \cap G = G \setminus (F^c \cap G)$ so

$$\begin{aligned} \nu^-(F) &= -\nu(F \cap G) \\ &= -(\nu(G) - \nu(F^c \cap G)) \\ &= -\lambda + \nu(F^c \cap G) \geq 0 \end{aligned}$$

and

$$\begin{aligned} \nu^+(F) &= \nu(F) + \nu^-(F) = \nu(F) - \lambda + \nu(F^c \cap G) \\ &= -\lambda + \nu(F \cup (F^c \cap G)) \geq 0\,, \end{aligned}$$

and the argument is complete.

The uniqueness: In order to show uniqueness of the decomposition, let $\nu = \tilde{\nu}^+ - \tilde{\nu}^-$ be another decomposition satisfying $\tilde{\nu}^+ \geq 0$, $\tilde{\nu}^- \geq 0$ and $\tilde{\nu}^+ \perp \tilde{\nu}^-$. Choose $\tilde{G} \in \mathcal{F}$ to be a set, such that $\tilde{\nu}^-$ is concentrated on $\tilde{G}$ and $\tilde{\nu}^+$ is concentrated on $\tilde{G}^c$. Then for $F \in \mathcal{F}$

$$\nu(F \cap G \cap \tilde{G}^c) = \begin{cases} -\nu^-(F \cap \tilde{G}^c) \leq 0\,, & \text{since it is a subset of } G \\ \tilde{\nu}^+(F \cap G) \geq 0\,, & \text{since it is a subset of } \tilde{G}^c \end{cases},$$

and hence $\nu(F \cap G \cap \tilde{G}^c) = 0$. Similarly we observe that $\nu(F \cap G^c \cap \tilde{G}) = 0$, so

$$\begin{aligned}\nu^-(F) &= -\nu(F \cap G) = -\nu(F \cap G \cap \tilde{G}) - \nu(F \cap G \cap \tilde{G}^c) \\ &= -\nu(F \cap G \cap \tilde{G}) = -\nu(F \cap G \cap \tilde{G}) - \nu(F \cap G^c \cap \tilde{G}) \\ &= -\nu(F \cap \tilde{G}) = \tilde{\nu}^-(F)\end{aligned}$$

for all $F \in \mathcal{F}$. □

Example 4.1.12. If $\nu = f \cdot \mu$, where μ is a bounded, positive measure and f is μ-integrable, we see that the decomposition is given by

$$\nu^+ = f^+ \cdot \mu, \quad \nu^- = f^- \cdot \mu,$$

where $f^+ = f \vee 0$ and $f^- = -(f \wedge 0)$ denote the positive and the negative part of f, respectively. The argument is by inspection: It is clear that $\nu = \nu^+ - \nu^-$, $\nu^+ \geq 0, \nu^- \geq 0$ and moreover $\nu^+ \perp \nu^-$ since ν^+ is concentrated on $(f \geq 0)$ and ν^- is concentrated on $(f < 0)$. ◦

Theorem 4.1.13 (The Lebesgue decomposition)**.** *If ν is a bounded, signed measure on $(\Omega, \mathcal{F})$ and μ is a bounded, positive measure on $\Omega, \mathcal{F})$ then there exists a $\mathcal{F}$–measurable, μ–integrable function f and a bounded, signed measure ν_s with $\nu_s \perp \mu$ such that*

$$\nu = f \cdot \mu + \nu_s\,.$$

The decomposition is unique in the sense that if $\nu = \widetilde{f} \cdot \mu + \widetilde{\nu}_s$ is another decomposition, then

$$\widetilde{f} = f \;\; \mu - a.e., \quad \widetilde{\nu}_s = \nu_s.$$

If $\nu \geq 0$, then $f \geq 0$ $\mu - a.e.$ and $\nu_s \geq 0$.

Proof. We begin with the uniqueness part of the theorem: Assume that

$$\nu = f \cdot \mu + \nu_s = \widetilde{f} \cdot \mu + \widetilde{\nu}_s,$$

where $\mu \perp \nu_s$ and $\mu \perp \widetilde{\nu}_s$. Choose $F_0, \widetilde{F}_0 \in \mathcal{F}$ such that ν_s is concentrated on F_0, μ is concentrated on F_0^c, $\widetilde{\nu}_s$ is concentrated on $\widetilde{F}_0$ and μ is concentrated on $\widetilde{F}_0^c$. Define $G_0 = F_0 \cup \widetilde{F}_0$. According to Lemma 4.1.10 we only need to show that $\mu(G_0) = 0$ in order to conclude that μ is concentrated on G_0^c, since $\mu \geq 0$. This is true since

$$0 \leq \mu(G_0) = \mu(F_0 \cup \tilde{F}_0) \leq \mu(F_0) + \mu(\tilde{F}_0) = 0 + 0 = 0\,.$$

Furthermore we have that ν_s is concentrated on G_0 since $F_0 \subseteq G_0$. Similarly $\tilde{\nu}_s$ is concentrated on G_0. Then for $F \in \mathcal{F}$

$$\begin{aligned}\tilde{\nu}_s(F) - \nu_s(F) &= \tilde{\nu}_s(F \cap G_0) - \nu_s(F \cap G_0) \\ &= \big(\nu(F \cap G_0) - (\tilde{f}\mu)(F \cap G_0)\big) - \big(\nu(F \cap G_0) - (f\mu)(F \cap G_0)\big) \\ &= \big(\nu(F \cap G_0) - 0\big) - \big(\nu(F \cap G_0) - 0\big) = 0\,,\end{aligned}$$

where we have used $\mu(F \cap G_0) = 0$ such that

$$\int_{F \cap G_0} f \,\mathrm{d}\mu = 0 \quad \text{and} \quad \int_{F \cap G_0} \tilde{f} \,\mathrm{d}\mu = 0\,.$$

Then $\nu_s = \tilde{\nu}_s$. The equation $f \cdot \mu + \nu_s = \tilde{f} \cdot \mu + \tilde{\nu}_s$ gives $f \cdot \mu = \tilde{f} \cdot \mu$, which leads to $f = \tilde{f}$ μ–a.e.

To prove existence, it suffices to consider the case $\nu \geq 0$. For a general ν we can find the Jordan–Hahn decomposition, $\nu = \nu^+ - \nu^-$, and then apply the Lebesgue decomposition to ν^+ and ν^- separately:

$$\nu^+ = f\mu + \nu_s \quad \text{and} \quad \nu^- = g\mu + \kappa_s$$

where there exist F_0 and $\tilde{F}_0$ such that ν_s is concentrated on F_0, μ is concentrated on F_0^c, κ_s is concentrated on $\tilde{F}_0$, and μ is concentrated on $\tilde{F}_0^c$. Defining $G_0 = F_0 \cup \tilde{F}_0$ we can obtain, similarly to the argument above, that

$$\nu_s, \kappa_s \text{ both are concentrated on } G_0 \quad \text{and} \quad \mu \text{ is concentrated on } G_0^c\,.$$

Obviously, the bounded, signed measure $\nu_s - \kappa_s$ is then concentrated on G_0 as well, leading to $\nu_s - \kappa_s \perp \mu$. Writing

$$\nu = (f - g)\mu + (\nu_s - \kappa_s)$$

gives the desired decomposition.

So assume that $\nu \geq 0$. Let $\mathcal{L}(\mu)^+$ denote the set of non–negative, μ–integrable functions and define

$$\mathbb{H} = \Big\{ g \in \mathcal{L}(\mu)^+ \;\Big|\; \nu(F) \geq \int_F g \,\mathrm{d}\mu \text{ for all } F \in \mathcal{F} \Big\}$$

Recall that $\nu \geq 0$ such that e.g. $0 \in \mathbb{H}$. Define furthermore

$$\alpha = \sup\Big\{ \int g \;\mathrm{d}\mu \;\Big|\; g \in \mathbb{H} \Big\}\,.$$

Since $\int_\Omega g \,\mathrm{d}\mu \leq \nu(\Omega)$ for all $g \in \mathbb{H}$, we must have

$$0 \leq \alpha \leq \nu(\Omega) < \infty\,.$$

We will show that there exists $f \in \mathbb{H}$ with $\int f \, d\mu = \alpha$.

Note that if $h_1, h_2 \in \mathbb{H}$ then $h_1 \vee h_2 \in \mathbb{H}$: For $F \in \mathcal{F}$ we have

$$\begin{aligned}\int_F h_1 \vee h_2 \, d\mu &= \int_{F \cap (h_1 \geq h_2)} h_1 \, d\mu + \int_{F \cap (h_1 < h_2)} h_2 \, d\mu \\ &\leq \nu(F \cap (h_1 \geq h_2)) + \nu(F \cap (h_1 < h_2)) = \nu(F)\,.\end{aligned}$$

At the inequality it is used that both h_1 and h_2 are in $\mathbb{H}$.

Now, for each $n \in \mathbb{N}$, choose $g_n \in \mathbb{H}$ so that

$$\int g_n \, d\mu \geq \alpha - \frac{1}{n}$$

and define $f_n = g_1 \vee \cdots \vee g_n$ for each $n \in \mathbb{N}$. According to the result shown above, we have $f_n \in \mathbb{H}$ and furthermore it is seen that the sequence (f_n) is increasing. Then the pointwise limit $f = \lim_{n\to\infty} f_n$ exists, and by monotone convergence we obtain for $F \in \mathcal{F}$ that

$$\int_F f \, d\mu = \lim_{n\to\infty} \int_F f_n \, d\mu \leq \nu(F)\,.$$

Hence $f \in \mathbb{H}$. Furthermore we have for all $n \in \mathbb{N}$ that $f \geq g_n$, so

$$\int f \, d\mu \geq \int g_n \, d\mu \geq \alpha - \frac{1}{n}$$

leading to the conclusion that $\int f \, d\mu = \alpha$.

Now we can define the bounded measure ν_s by

$$\nu_s = \nu - f \cdot \mu$$

Then $\nu_s \geq 0$ since $f \in \mathbb{H}$ such that

$$\nu_s(F) = \nu(F) - \int_F f \, d\mu \geq 0$$

for all $F \in \mathcal{F}$.

What remains in the proof is showing that $\nu_s \perp \mu$. For all $n \in \mathbb{N}$ define the bounded, signed measure (see e.g. Exercise 4.1) λ_n by

$$\lambda_n = \nu_s - \frac{1}{n}\mu$$

Let $\lambda_n = \lambda_n^+ - \lambda_n^-$ be the Jordan–Hahn decomposition of λ_n. Then we can find $F_n \in \mathcal{F}$ such that

$$\lambda_n^- \text{ is concentrated on } F_n \quad \text{and} \quad \lambda_n^+ \text{ is concentrated on } F_n^c$$

For $F \in \mathcal{F}$ and $F \subseteq F_n^c$ we obtain

$$\begin{aligned}\nu(F) = \nu_s(F) + \int_F f \,\mathrm{d}\mu &= \lambda_n(F) + \frac{1}{n}\mu(F) + \int_F f \,\mathrm{d}\mu \\ &= \lambda_n^+(F) + \frac{1}{n}\mu(F) + \int_F f \,\mathrm{d}\mu \geq \int_F f + \frac{1}{n}\,\mathrm{d}\mu\,.\end{aligned}$$

If we define

$$\tilde{f}_n = \begin{cases} f & \text{on } F_n \\ f + \frac{1}{n} & \text{on } F_n^c \end{cases},$$

then for $F \in \mathcal{F}$

$$\begin{aligned}\int_F \tilde{f}_n \,\mathrm{d}\mu &= \int_{F\cap F_n} f \,\mathrm{d}\mu + \int_{F\cap F_n^c} f + \frac{1}{n}\,\mathrm{d}\mu \\ &\leq \nu(F\cap F_n) + \nu(F \cap F_n^c) = \nu(F)\end{aligned}$$

so $\tilde{f}_n \in \mathbb{H}$. Hence

$$\alpha \geq \int \tilde{f}_n \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu + \frac{1}{n}\mu(F_n^c) = \alpha + \frac{1}{n}\mu(F_n^c)\,.$$

This implies that $\mu(F_n^c) = 0$ leading to

$$\mu\Big(\bigcup_{n=1}^{\infty} F_n^c\Big) = 0\,.$$

Thus μ is concentrated on $F_0 = (\cup_1^\infty F_n^c)^c = \cap_1^\infty F_n$. Finally, we have for all $n \in \mathbb{N}$ (recall that λ_n^+ is concentrated on F_n^c) that

$$\begin{aligned}0 \leq \nu_s(F_0) \leq \nu_s(F_n) &= \frac{1}{n}\mu(F_n) + \lambda_n(F_n) \\ &= \frac{1}{n}\mu(F_n) - \lambda_n^-(F_n) \leq \frac{1}{n}\mu(\Omega)\end{aligned}$$

which for $n \to \infty$ implies that $\nu_s(F_0) = 0$. Hence (since $\nu_s \geq 0$) ν_s is concentrated on F_0^c. □

Theorem 4.1.14 (Radon-Nikodym)**.** *Let μ be a positive, bounded measure and ν a bounded, signed measure on $(\Omega, \mathcal{F})$. Then $\nu \ll \mu$ if and only if there exists a $\mathcal{F}$–measurable, μ–integrable f such that $\nu = f \cdot \mu$.*

If $\nu \ll \mu$ then the density f is uniquely determined μ-a.e. If in addition $\nu \geq 0$ then $f \geq 0$ μ-a.e.

Proof. That f is uniquely determined follows from the uniqueness in the Lebesgue decomposition. Also $\nu \geq 0$ implies $f \geq 0$ μ–a.e.

In the "if and only if" part it only remains to show, that $\nu \ll \mu$ implies the existence of a $\mathcal{F}$–measurable and μ–integrable function f with $\nu = f\mu$. So assume that $\nu \ll \mu$ and consider the Lebesgue decomposition of ν

$$\nu = f \cdot \mu + \nu_s\,.$$

Choose F_0 such that ν_s is concentrated on F_0 and μ is concentrated on F_0^c. For $F \in \mathcal{F}$ we then obtain that

$$\nu_s(F) = \nu_s(F \cap F_0) = \nu(F \cap F_0) - (f\mu)(F \cap F_0) = 0$$

since $\mu(F \cap F_0) = 0$ and since $\nu \ll \mu$ implies that $\nu(F \cap F_0) = 0$. Hence $\nu_s = 0$ and the claim $\nu = f \cdot \mu$ follows. □

4.2 Conditional Expectations given a σ-algebra

In this section we will return to considering a probability space $(\Omega, \mathcal{F}, P)$ and real random variables defined on this space. We shall see how the existence of conditional expectations can be shown using a Radon-Nikodym derivative. In the course MI the existence is shown from $\mathcal{L}^2$-theory using projetions on the subspace $\mathcal{L}^2(\Omega, \mathcal{D}, P)$ of $\mathcal{L}^2(\Omega, \mathcal{F}, P)$, when $\mathcal{D} \subseteq \mathcal{F}$ is a sub σ-algebra.

Let X be a real random variable defined on $(\Omega, \mathcal{F}, P)$ with $E|X| < \infty$. A conditional expectation of X (given something) can be interpreted as a guess on the value of $X(\omega)$ based on varying amounts of information about which $\omega \in \Omega$ has been drawn. If we know nothing about ω, then it is not possible to say very much about the value of $X(\omega)$. Perhaps the best guess we can come up with is suggesting the value $E(X) = \int X \,\mathrm{d}P$!

Now let $D_1, \ldots, D_n$ be a system of disjoint sets in $\mathcal{F}$ with $\cup_{i=1}^n D_i = \Omega$, and assume that for a given $\omega \in \Omega$, we know whether $\omega \in D_i$ for each $i = 1, \ldots, n$. Then we actually have *some* information about the ω that has been drawn, and an educated guess on the value of $X(\omega)$ may not be as simple as $E(X)$ any more. Instead our guessing strategy will be

$$\text{guess on } X(\omega) \;=\; \frac{1}{P(D_i)} \int_{D_i} X \,\mathrm{d}P \qquad \text{if } \omega \in D_i\,. \tag{4.4}$$

We are still using an integral of X, but we only integrate over the set D_i, where we know that ω is an element. It may not be entirely clear, why this is a good strategy for our guess (that will probably depend on the definition of a good guess), but at least it seems reasonable that we give the same guess on $X(\omega)$ for all $\omega \in D_i$.

Example 4.2.1. Suppose $\Omega = \{a, b, c, d\}$ and that the probability measure P is given by

$$P(\{a\}) = P(\{b\}) = P(\{c\}) = P(\{d\}) = \frac{1}{4}$$

and furthermore that $X : \Omega \to \mathbb{R}$ is defined by

$$X(a) = 5 \quad X(b) = 4 \quad X(c) = 3 \quad X(d) = 2\,.$$

If we know nothing about ω then the guess is

$$E(X) = \frac{5+4+3+2}{4} = 3.5\,.$$

Now let $D = \{a, b\}$ and assume that we want to guess $X(\omega)$ in a situation where we know whether $\omega \in D$ or $\omega \in D^c$. The strategy described above gives that if $\omega \in D$ then

$$\text{guess on } X(\omega) = \frac{\int_D X \, \mathrm{d}P}{P(D)} = \frac{\frac{1}{4}(5+4)}{\frac{1}{2}} = 4.5\,.$$

Similarly, if we know $\omega \in D^c = \{c, d\}$, then the best guess would be 2.5. Given the knowledge of whether $\omega \in D$ or $\omega \in D^c$ we can write the guess as a function of ω, namely

$$\text{guess}(\omega) = 1_D(\omega) \cdot 4.5 + 1_{D^c}(\omega) \cdot 2.5\,.$$

$\circ$

Note that the collection $\{D_1, \ldots, D_n\}$ is a sub σ–algebra of $\mathcal{F}$. The stability requirements for σ–algebras fit very well into the knowledge of whether $\omega \in D_i$ for all i: If we know whether $\omega \in D_i$, then we also know whether $\omega \in D_i^c$. And we know whether $\omega \in \cup A_i$, if we know whether $\omega \in A_i$ for all i.

The concept of conditional expectations takes the guessing strategy to a general level, where the conditioning σ-algebra $\mathcal{D}$ is general sub σ–algebra of $\mathcal{F}$. The result will be a random variable (as in Example 4.2.1) which we will denote $E(X|\mathcal{D})$ and call the *conditional expectation of X given $\mathcal{D}$*. We will show in Example 4.2.5 that when $\mathcal{D}$ has the form $\{D_1, \ldots, D_n\}$ as above, then $E(X|\mathcal{D})$ is given by (4.4).

Definition 4.2.2. *let X be a real random variable defined on $(\Omega, \mathcal{F}, P)$ with $E|X| < \infty$. A conditional expectation X given $\mathcal{D}$ is a $\mathcal{D}$-measurable real random variable, denoted $E(X|\mathcal{D})$ which satisfies*

$$E|E(X|\mathcal{D})| < \infty, \tag{1}$$

$$\int_D E(X|\mathcal{D}) \, \mathrm{d}P = \int_D X \, \mathrm{d}P \qquad \textit{for all } D \in \mathcal{D}\,. \tag{2}$$

Note that one cannot in general use $E(X|\mathcal{D}) = X$ (even though it satisfies (1) and (2)): X is assumed to be $\mathcal{F}$-measurable but need not be $\mathcal{D}$-measurable.

Given this definition, conditional expectations are almost surely unique:

Theorem 4.2.3. *(1) If U and $\widetilde{U}$ are both conditional expectations of X given $\mathcal{D}$, then $U = \widetilde{U}$ a.s.*

(2) If U is a conditional expectation of X given $\mathcal{D}$ and $\widetilde{U}$ is $\mathcal{D}$-measurable with $\widetilde{U} = U$ a.s. then $\widetilde{U}$ is also a conditional expectation of X given $\mathcal{D}$.

Proof. For the first result consider, e.g., $D = (\widetilde{U} > U)$. Then

$$\int_D (\widetilde{U} - U)\,\mathrm{d}P = \int_D \widetilde{U} dP - \int_D U\,\mathrm{d}P = \int_D X\,\mathrm{d}P - \int_D X\,\mathrm{d}P = 0$$

according to (2) in Definition 4.2.2 . But $\widetilde{U} > U$ on D, so therefore $P(D) = P(\widetilde{U} > U) = 0$. Similarly, $P(\widetilde{U} < U) = 0$.

The second statement is trivial: Simply use that

$$E|\tilde{U}| = E|U| \quad \text{and} \quad \int_D \tilde{U}\,\mathrm{d}P = \int_D U\,\mathrm{d}P$$

so $\tilde{U}$ satisfies (1) and (2). □

Theorem 4.2.4. *If X is a real random variable with $E|X| < \infty$, then there exists a conditional expectation of X given $\mathcal{D}$.*

Proof. Define for $D \in \mathcal{D}$

$$\nu(D) = \int_D X\,\mathrm{d}P\,.$$

Then ν is a bounded, signed measure on $(\Omega, \mathcal{D})$. Let P^0 denote the restriction of P to $\mathcal{D}$: P^0 is the probability measure on $(\Omega, \mathcal{D})$ given by

$$P^0(D) = P(D)$$

for all $D \in \mathcal{D}$. Now we obviously have for all $D \in \mathcal{D}$

$$P^0(D) = 0 \quad \Rightarrow \quad P(D) = 0 \quad \Rightarrow \quad \nu(D) = 0\,,$$

so $\nu \ll P^0$. According to the Radon–Nikodym Theorem we can find the Radon–Nikodym derivative $U = d\nu/dP^0$ satisfying

$$\nu(D) = \int_D U \,\mathrm{d}P^0 \,.$$

By construction in the Radon–Nikodym Theorem, U is automatically $\mathcal{D}$–measurable and P^0–integrable. For all $D \in \mathcal{D}$ we now have that

$$\int_D X \,\mathrm{d}P = \nu(D) = \int_D U \,\mathrm{d}P^0 = \int_D U \,\mathrm{d}P \tag{4.5}$$

so it is shown that U is a conditional expectation of X given $\mathcal{D}$.

That the last equation in (4.5) is true is just basic measure theory: The integral of a function with respect to some measure does not change if the measure is extended to a larger σ–algebra; the function is also a measurable function on the larger measurable space.

A direct argument could be first looking at indicator functions. Let $D \in \mathcal{D}$ and note that 1_D is $\mathcal{D}$–measurable. Then

$$\int 1_D \,\mathrm{d}P^0 = P^0(D) = P(D) = \int 1_D \,\mathrm{d}P \,.$$

Then it follows that

$$\int Y \,\mathrm{d}P^0 = \int Y \,\mathrm{d}P$$

if Y is a linear combination of indicator functions, and finally the result is shown to be true for general $\mathcal{D}$–measurable functions Y by a standard approximation argument. □

Example 4.2.5. consider a probability space $(\Omega, \mathcal{F}, P)$ and a real random variable X defined on Ω with $E|X| < \infty$. Assume that $D_1, \dots, D_n \in \mathcal{F}$ form a partition of Ω: $D_i \cap D_j = \emptyset$ for $i \neq j$ and $\cup_{i=1}^n D_i = \Omega$. Also assume (for convenience) that $P(D_i) > 0$ for all $i = 1, \dots, n$. Let $\mathcal{D}$ be the σ–algebra generated by the D_j–sets. Then $D \in \mathcal{D}$ if and only if D is a union of some D_j's.

We will show that

$$U = \sum_{i=1}^n \frac{1}{P(D_i)} \Big(\int_{D_i} X \;\, \mathrm{d}P \Big) 1_{D_i} \,.$$

is a conditional expectation of X given $\mathcal{D}$. First note that U is $\mathcal{D}$–measurable, since the indicator functions 1_{D_i} are $\mathcal{D}$–measurable. Furthermore

$$
\begin{aligned}
E|U| &\le \sum_{i=1}^n \int \Big| \frac{1}{P(D_i)} \Big(\int_{D_i} X \;\mathrm{d}P \Big) 1_{D_i} \Big| \,\mathrm{d}P \\
&\le \sum_{i=1}^n \frac{1}{P(D_i)} \Big(\int_{D_i} |X| \;\mathrm{d}P \Big) \int 1_{D_i} \,\mathrm{d}P \\
&= \sum_{i=1}^n \int_{D_i} |X| \;\mathrm{d}P = E|X| < \infty \,.
\end{aligned}
$$

Finally let $D \in \mathcal{D}$. Then $D = \cup_{k=1}^m D_{i_k}$ for some $1 \le i_1 < \cdots < i_m \le n$. We therefore obtain

$$
\int_D X \,\mathrm{d}P = \sum_{k=1}^m \int_{D_{i_k}} X \,\mathrm{d}P
$$

so

$$
\begin{aligned}
\int_D U \,\mathrm{d}P &= \sum_{k=1}^m \int_{D_{i_k}} \sum_{i=1}^n \frac{1}{P(D_i)} \Big(\int_{D_i} X \,\mathrm{d}P \Big) 1_{D_i} \,\mathrm{d}P \\
&= \sum_{k=1}^m \int_{D_{i_k}} \frac{1}{P(D_{i_k})} \Big(\int_{D_{i_k}} X \,\mathrm{d}P \Big) 1_{D_{i_k}} \,\mathrm{d}P \\
&= \sum_{k=1}^m \frac{1}{P(D_{i_k})} \Big(\int_{D_{i_k}} X \,\mathrm{d}P \Big) \int_{D_{i_k}} 1_{D_{i_k}} \,\mathrm{d}P \\
&= \sum_{k=1}^m \int_{D_{i_k}} X \,\mathrm{d}P = \int_D X \,\mathrm{d}P \,.
\end{aligned}
$$

Hence U satisfies the conditions in Definition 4.2.2 and is therefore a conditional expectation of X given $\mathcal{D}$. ◦

We shall now show a series of results concerning conditional expectations. The results and the proofs are well-known from the course MI. In Theorem 4.2.6, X, Y and X_n are real random variables, all of which are integrable.

Theorem 4.2.6. *(1) If $X = c$ a.s., where $c \in \mathbb{R}$ is a constant, then $E(X|\mathcal{D}) = c$ a.s.*

(2) For $\alpha, \beta \in \mathbb{R}$ it holds that

$$
E(\alpha X + \beta Y|\mathcal{D}) = \alpha E(X|\mathcal{D}) + \beta E(Y|\mathcal{D}) \text{ a.s.}
$$

(3) If $X \ge 0$ a.s. then $E(X|\mathcal{D}) \ge 0$ a.s. If $Y \ge X$ a.s. then $E(Y|\mathcal{D}) \ge E(X|\mathcal{D})$ a.s.

(4) If $\mathcal{D} \subseteq \mathcal{E}$ are sub σ-algebras of $\mathcal{F}$ then

$$E(X|\mathcal{D}) = E[E(X|\mathcal{E})|\mathcal{D}] = E[E(X|\mathcal{D})|\mathcal{E}] \ a.s.$$

(5) If $\sigma(X)$ and $\mathcal{D}$ are independent then

$$E(X|\mathcal{D}) = EX \ \ a.s.$$

(6) If X is $\mathcal{D}$-measurable then

$$E(X|\mathcal{D}) = X \ \ a.s.$$

(7) If it holds for all $n \in \mathbb{N}$ that $X_n \geq 0$ a.s. and $X_{n+1} \geq X_n$ a.s. with $\lim X_n = X$ a.s., then

$$\lim_{n\to\infty} E(X_n|\mathcal{D}) = E(X|\mathcal{D}) \ \ a.s.$$

(8) If X is $\mathcal{D}$-measurable and $E|XY| < \infty$, then

$$E(XY|\mathcal{D}) = X\, E(Y|\mathcal{D}) \ \ a.s.$$

(9) If $f : \mathbb{R} \to \mathbb{R}$ is a measurable function that is convex on an interval I, such that $P(X \in I) = 1$ and $E|f(X)| < \infty$, then it holds that

$$f\big(E(X|\mathcal{D})\big) \leq E\big(f(X)|\mathcal{D}\big) \ \ a.s.$$

Proof. (1) We show that the constant variable U given by $U(\omega) = c$ meets the conditions from Definition 4.2.2. Firstly, it is $\mathcal{D}$–measurable, since for $B \in \mathcal{B}$ we have

$$U^{-1}(B) = \begin{cases} \Omega & c \in B \\ \emptyset & c \notin B \end{cases}$$

which is $\mathcal{D}$ measurable in either case. Furthermore $E|U| = |c| < \infty$ and obviously

$$\int_D U \,\mathrm{d}P = \int_D c \,\mathrm{d}P = \int_D X \,\mathrm{d}P\,.$$

(2) $\alpha E(X|\mathcal{D}) + \beta E(Y|\mathcal{D})$ is $\mathcal{D}$-measurable and integrable, so all we need to show is (see part (2) of Definition 4.2.2), that

$$\int_D (\alpha E(X|\mathcal{D}) + \beta E(Y|\mathcal{D}))\,\mathrm{d}P = \int_D (\alpha X + \beta Y)\,\mathrm{d}P$$

for all $D \in \mathcal{D}$. But here, the left hand side is

$$\alpha \int_D E(X|\mathcal{D})\,\mathrm{d}P + \beta \int_D E(Y|\mathcal{D})\,\mathrm{d}P,$$

which is seen to equal the right hand side when we use Definition 4.2.2 on both terms.

(3) For the first claim define $D = (E(X|\mathcal{D}) < 0) = E(X|\mathcal{D})^{-1}((-\infty, 0))$ and note that $D \in \mathcal{D}$ since $E(X|\mathcal{D})$ is $\mathcal{D}$–measurable. Then

$$\int_D E(X|\mathcal{D})\,\mathrm{d}P = \int_D X\,\mathrm{d}P \geq 0\,,$$

since $X \geq 0$ a.s. But the fact that $E(X|\mathcal{D}) < 0$ on D, makes $\int_D E(X|\mathcal{D})\,\mathrm{d}P < 0$, if $P(D) > 0$. Hence $P(D) = 0$ and $E(X|\mathcal{D}) \geq 0$ a.s.

For the second claim, just use the first result on $Y - X$ and apply (2).

(4) Firstly, we show that $E(X|\mathcal{D}) = E[E(X|\mathcal{E})|\mathcal{D}]$ a.s. By definition we have that $E[E(X|\mathcal{E})|\mathcal{D}]$ is $\mathcal{D}$–measurable with finite expectation, and for $D \in \mathcal{D}$

$$\int_D E[E(X|\mathcal{E})|\mathcal{D}]\,\mathrm{d}P = \int_D E(X|\mathcal{E})\,\mathrm{d}P = \int_D X\,\mathrm{d}P\,.$$

Hence we have the result from Definition 4.2.2. In the first equality, it is used that $E[E(X|\mathcal{E})|\mathcal{D}]$ is a conditional expectation of $E(X|\mathcal{E})$ given $\mathcal{D}$. In the second equality Definition 4.2.2 is applied to $E(X|\mathcal{E})$, using that $D \in \mathcal{D} \subseteq \mathcal{E}$.

Secondly, we prove that $E(X|\mathcal{D}) = E[E(X|\mathcal{D})|\mathcal{E}]$ a.s. by showing that $E(X|\mathcal{D})$ is a conditional expectation of $E(X|\mathcal{D})$ given $\mathcal{E}$. But that follows directly from 6, since $E(X|\mathcal{D})$ is $\mathcal{E}$–measurable.

(5) As in (1), the constant map $\omega \mapsto EX$ is $\mathcal{D}$–measurable and has finite expectation, so it remains to show that for $D \in \mathcal{D}$

$$\int_D EX\,\mathrm{d}P = \int_D X\,\mathrm{d}P\,.$$

The left hand side is $EX \cdot P(X)$. For the right hand side we obtain the following, using that 1_D and X are independent

$$\int_D X\,\mathrm{d}P = \int 1_D \cdot X\,\mathrm{d}P = \int 1_D\,\mathrm{d}P \cdot \int X\,\mathrm{d}P = P(D) \cdot EX\,,$$

so the stated equality is true.

(6) Trivial.

(7) According to (3) we have for all $n \in \mathbb{N}$ that $E(X_{n+1}|\mathcal{D}) \geq E(X_n|\mathcal{D})$ a.s., so with

$$F_n = (E(X_{n+1}|\mathcal{D}) \geq E(X_n|\mathcal{D})) \in \mathcal{D}$$

we have $P(F_n) = 1$. Let $F_0 = (E(X_1|\mathcal{D}) \geq 0)$ such that $P(F_0) = 1$. With the definition $F = \cap_{n=0}^{\infty} F_n$ we have $F \in \mathcal{D}$ and $P(F) = 1$. For $\omega \in F$ it holds that the sequence $\big(E(X_n|\mathcal{D})(\omega)\big)_{n\in\mathbb{N}}$ is increasing and $E(X_1|\mathcal{D})(\omega) \geq 0$. Hence for $\omega \in F$ the number $Y(\omega) = \lim_{n\to\infty} E(X_n|\mathcal{D})(\omega)$

is well–defined in $[0, \infty]$. Defining e.g. $Y(\omega) = 0$ for $\omega \in F^c$ makes Y a $\mathcal{D}$–measurable random variable (since F is $\mathcal{D}$–measurable, and Y is the point–wise limit of $1_F E(X_n|\mathcal{D})$ that are all $\mathcal{D}$–measurable variables) with values in $[0, \infty]$. Thus the integrals $\int_G Y \,\mathrm{d}P$ of Y makes sense for all $G \in \mathcal{F}$.

In particular we obtain the following for $D \in \mathcal{D}$ using monotone convergence in the third and the sixth equality

$$\begin{aligned}\int_D Y \,\mathrm{d}P &= \int_{D\cap F} Y \,\mathrm{d}P = \int_{D\cap F} \lim_{n\to\infty} E(X_n|\mathcal{D}) \,\mathrm{d}P = \lim_{n\to\infty} \int_{D\cap F} E(X_n|\mathcal{D}) \,\mathrm{d}P \\ &= \lim_{n\to\infty} \int_D E(X_n|\mathcal{D}) \,\mathrm{d}P = \lim_{n\to\infty} \int_D X_n \,\mathrm{d}P = \int_D \lim_{n\to\infty} X_n \,\mathrm{d}P = \int_D X \,\mathrm{d}P\,.\end{aligned}$$

Letting $D = \Omega$ shows that $E|Y| = EY = EX < \infty$, so we can conclude $Y = E(X|\mathcal{D})$ a.s.. Thereby we have shown (7).

(8) Since $XE(Y|\mathcal{D})$ is obviously $\mathcal{D}$–measurable, it only remains to show that $E|XE(Y|\mathcal{D})| < \infty$ and that

$$\int_D XE(Y|\mathcal{D}) \,\mathrm{d}P = \int_D XY \,\mathrm{d}P \tag{4.6}$$

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

We now prove the result for all $X \geq 0$ and $Y \geq 0$ by showing the equation in the following steps:

(i) when $X = 1_{D_0}$ for $D_0 \in \mathcal{D}$

(ii) when $X = \sum_{k=1}^n a_k 1_{D_k}$ for $D_k \in \mathcal{D}$ and $a_k \geq 0$

(iii) when $X \geq 0$ is a general $\mathcal{D}$–measurable variable

So firstly assume that $X = 1_{D_0}$ with $D_0 \in \mathcal{D}$. Then

$$E|1_{D_0}E(Y|\mathcal{D})| \leq E|E(Y|\mathcal{D})| < \infty$$

and since $D \cap D_0 \in \mathcal{D}$ we obtain

$$\begin{aligned}\int_D 1_{D_0}E(Y|\mathcal{D})\,\mathrm{d}P &= \int_{D\cap D_0} E(Y|\mathcal{D})\,\mathrm{d}P \\ &= \int_{D\cap D_0} Y\,\mathrm{d}P \\ &= \int_D 1_{D_0}Y\,\mathrm{d}P\,.\end{aligned}$$

Hence formula (4.6) is shown in case (i). If

$$X = \sum_{k=1}^{n} a_k 1_{D_k}$$

with $D_k \in \mathcal{D}$ and $a_k \geq 0$, we easily obtain (4.6) from linearity

$$\int_D XE(Y|\mathcal{D})\,\mathrm{d}P = \sum_{k=1}^{n} a_k \int_D 1_{D_k}E(Y|\mathcal{D})\,\mathrm{d}P = \sum_{k=1}^{n} a_k \int_D 1_{D_k}Y\,\mathrm{d}P = \int_D XE(Y|\mathcal{D})\,\mathrm{d}P\,.$$

For a general $\mathcal{D}$–measurable X we can obtain X through the approximation

$$X = \lim_{n\to\infty} X_n\,,$$

where

$$X_n = \sum_{k=1}^{n2^n} \frac{k-1}{2^n} 1_{(\frac{k-1}{2^n} \leq X < \frac{k}{2^n})}\,.$$

Note that all sets $(\frac{k-1}{2^n} \leq X < \frac{k}{2^n})$ are $\mathcal{D}$–measurable, so each X_n have the form from step (ii). Hence

$$\int_D X_n E(Y|\mathcal{D})\,\mathrm{d}P = \int_D X_n Y\,\mathrm{d}P \tag{4.7}$$

for all $n \in \mathbb{N}$. Furthermore the construction of $(X_n)_{n\in\mathbb{N}}$ makes it non–negative and increasing.

Since we have assumed that $Y \geq 0$ we must have $XE(Y|\mathcal{D}) \geq 0$ a.s. Thereby the integral $\int_D XE(Y|\mathcal{D})\,\mathrm{d}P$ is defined for all $D \in \mathcal{D}$ (but it may be $+\infty$). Since the sequence

$(X_n E(Y|\mathcal{D}))_{n\in\mathbb{N}}$ is almost surely increasing (increasing for all ω with $E(Y|\mathcal{D})(\omega) \geq 0$) we obtain

$$\begin{aligned}\int_D XE(Y|\mathcal{D})\,\mathrm{d}P &= \int_D \lim_{n\to\infty} X_n E(Y|\mathcal{D})\,\mathrm{d}P = \lim_{n\to\infty}\int_D X_n E(Y|\mathcal{D})\,\mathrm{d}P \\ &= \lim_{n\to\infty}\int_D X_n Y\,\mathrm{d}P = \int_D XY\,\mathrm{d}P\,,\end{aligned}$$

where the second and the fourth equality follow from monotone convergence, and the third equality is a result of (4.7). From this (since $E|XY| < \infty$) we in particular see, that

$$E|XE(Y|\mathcal{D})| = E(XE(Y|\mathcal{D})) = \int_\Omega XE(Y|\mathcal{D})\,\mathrm{d}P = \int_\Omega XY\,\mathrm{d}P = E(XY) < \infty\,.$$

Hence (8) is shown in the case, where $X \geq 0$ and $Y \geq 0$. That (8) holds in general then easily follows by splitting X and Y up into their positive and negative parts, $X = X^+ - X^-$ and $Y = Y^+ - Y^-$, and then applying the version of (8) that deals with positive X and Y on each of the terms we obtain the desired result by multiplying out the brackets.

(9) The full proof is given in the lecture notes of the course MI. That

$$E(X|\mathcal{D})^+ \leq E(X^+|\mathcal{D}) \;\; a.s.$$

is shown in Exercise 4.11, and that

$$|E(X|\mathcal{D})| \leq E(|X||\mathcal{D}) \;\; a.s.$$

is shown in Exercise 4.12. □

4.3 Conditional expectations given a random variable

In this section we will consider the special case, where the conditioning σ–algebra $\mathcal{D}$ is generated by a random variable Y. So assume that $Y : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ is a random variable with values in the space E that is not necessarily $\mathbb{R}$. If $\mathcal{D} = \sigma(Y)$, i.e. the σ-algebra generated by Y, we write

$$E(X|Y)$$

rather than $E(X|\mathcal{D})$ and the resulting random variable is referred to as the conditional expectation of X *given* Y. Recall that $D \in \sigma(Y)$ is always of the form $D = (Y \in A)$ for some $A \in \mathcal{E}$. Then we immediately have the following characterization of $E(X|Y)$:

Theorem 4.3.1. *Let X be a real random variable with $E|X| < \infty$, and assume that Y is a random variable with values in $(E, \mathcal{E})$. Then the conditional expectation $E(X|Y)$ of X given Y is characterised by being $\sigma(Y)$-measurable and satisfying $E|E(X|Y)| < \infty$ and*

$$\int_{(Y \in A)} E(X|Y)dP = \int_{(Y \in A)} X \, \mathrm{d}P \quad \text{for all } A \in \mathcal{E}.$$

Note that if $\sigma(Y) = \sigma(\widetilde{Y})$, then $E(X|Y) = E(X|\widetilde{Y})$ a.s. If e.g. Y takes values in the real numbers and $\psi : E \to E$ is a bijective and bimeasurable map (ψ and ψ^{-1} are both measurable), then

$$E(X|Y) = E(X|\psi(Y)) \text{ a.s.}$$

The following lemma will be extremely useful in the comprehension of conditional expectations given random variables.

Lemma 4.3.2. *A real random variable Z is $\sigma(Y)$-measurable if and only if there exists a measurable map $\phi : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B})$ such that*

$$Z = \phi \circ Y.$$

Proof. First the easy implication:

Assume that $Z = \phi \circ Y$, where ϕ is $\mathcal{E}$–$\mathcal{B}$–measurable, and obviously Y is $\sigma(Y)$–$\mathcal{E}$–measurable. Then it is well–known that Z is $\sigma(Y)$–$\mathcal{B}$–measurable.

Now assume that Z is $\sigma(Y)$–measurable. We can write Z as

$$Z = \lim_{n \to \infty} \sum_{k=-n2^n}^{n2^n} \frac{k-1}{2^n} 1_{\left(\frac{k-1}{2^n} \le Z < \frac{k}{2^n}\right)}, \tag{4.8}$$

where each set $\left(\frac{k-1}{2^n} \le Z < \frac{k}{2^n}\right) \in \sigma(Y)$, since Z is $\sigma(Y)$–measurable. Define the class $\mathbb{H}$ of real random variables by

$$\mathbb{H} = \{Z' \; : \; \text{there exists a } \mathcal{E} - \mathcal{B}\text{–measurable function } \psi \text{ with } Z' = \psi \circ Y\}$$

Because of the approximation (4.8) the argument will be complete, if we can show that $\mathbb{H}$ has the following properties

(i) $1_D \in \mathbb{H}$ for all $D \in \sigma(Y)$

(ii) $a_1Z_1 + \cdots + a_nZ_n \in \mathbb{H}$ if $Z_1, \ldots, Z_n \in \mathbb{H}$ and $a_1, \ldots, a_n \in \mathbb{R}$

(iii) $Z \in \mathbb{H}$, where $Z = \lim_{n\to\infty} Z_n$ and all $Z_n \in \mathbb{H}$

because in that case we will have shown that $Z \in \mathbb{H}$.

(i): Assume that $D \in \sigma(Y)$. Then there exists a set $A \in \mathcal{E}$ such that $D = (Y \in A)$ (simply from the definition of $\sigma(Y)$). But then $1_D = 1_A \circ Y$, since

$$\begin{aligned} 1_D(\omega) = 1 \quad &\Leftrightarrow \quad \omega \in D = (Y \in A) \\ &\Leftrightarrow \quad Y(\omega) \in A \\ &\Leftrightarrow \quad (1_A \circ Y)(\omega) = 1\,. \end{aligned}$$

We have that 1_A is $\mathcal{E} - \mathcal{B}$–measurable, so (i) is shown.

(ii): Assume that $Z_k = \phi_k \circ Y$, where ϕ_k is $\mathcal{E} - \mathcal{B}$–measurable, for $k = 1, \ldots, n$. Then we get

$$\sum_{k=1}^n a_k Z_k(\omega) = \sum_{k=1}^n a_k(\phi_k(Y(\omega))) = \left(\Big(\sum_{k=1}^n a_k\phi_k\Big) \circ Y\right)(\omega)\,.$$

So

$$\sum_{k=1}^n a_k Z_k = (\sum_{k=1}^n a_k\phi_k) \circ Y\,,$$

where $\sum_{k=1}^n a_k\phi_k$ is measurable. Hence we have shown (ii).

(iii): Assume that $Z_n = \phi_n \circ Y$ for all $n \in \mathbb{N}$, where ϕ_n is $\mathcal{E} - \mathcal{B}$–measurable. Then

$$Z(\omega) = \lim_{n\to\infty} Z_n(\omega) = \lim_{n\to\infty} (\phi_n \circ Y)(\omega) = \lim_{n\to\infty} \phi_n(Y(\omega))$$

for all $\omega \in \Omega$. In particular the limit $\lim_{n\to\infty} \phi_n(y)$ exists for all $y \in Y(\Omega) = \{Y(\omega) \ : \ \omega \in \Omega\}$. Define $\phi : E \to \mathbb{R}$ by

$$\phi(y) = \begin{cases} \lim_{n\to\infty} \phi_n(y) & \text{if the limit exists} \\ 0 & \text{otherwise} \end{cases}\,,$$

then ϕ is $\mathcal{E} - \mathcal{B}$–measurable, since $F = (\lim_n \phi_n \text{ exists}) \in \mathcal{E}$ and $\phi = \lim_n(1_F\phi_n)$ with each $1_F\phi_n$ being $\mathcal{E} - \mathcal{B}$–measurable. Furthermore note that $Z(\omega) = \phi(Y(\omega))$, so $Z = \phi \circ Y$. Hence (iii) is shown. □

Now we return to the discussion of the $\sigma(Y)$-measurable random variable $E(X|Y)$. By the lemma, we have that

$$E(X|Y) = \phi \circ Y,$$

for some measurable $\phi : E \to \mathbb{R}$. We call $\phi(y)$ a *conditional expectation of X given $Y = y$* and write

$$\phi(y) = E(X|Y = y).$$

This type of conditional expectations is characterized in Theorem 4.3.3 below.

Theorem 4.3.3. *A measurable map $\phi : E \to \mathbb{R}$ defines a conditional expectation of X given $Y = y$ for all y if and only if ϕ is integrable with respect to the distribution $Y(P)$ of Y and*

$$\int_B \phi(y)\,\mathrm{d}Y(P)(y) = \int_{(Y\in B)} X\,\mathrm{d}P \quad (B \in \mathcal{E}).$$

Proof. Firstly, assume that ϕ defines a conditional expectation of X given $Y = y$ for all y. Then we have $E(X|Y) = \phi \circ Y$ so

$$\int_E |\phi(y)|\,\mathrm{d}Y(P)(y) = \int_\Omega |\phi \circ Y|\,\mathrm{d}P = \int_\Omega |E(X|Y)|\,\mathrm{d}P = E|E(X|Y)| < \infty\,,$$

and we have shown that ϕ is $Y(P)$–integrable. Above, the first equality is a result of the Change–of–variable Formula. Similarly we obtain for all $B \in \mathcal{E}$

$$\int_B \phi(y)\,\mathrm{d}Y(P)(y) = \int_{(Y\in B)} E(X|Y)\,\mathrm{d}P = \int_{(Y\in B)} X\,\mathrm{d}P\,.$$

Thereby the "only if" claim is obtained.

Conversely, assume that ϕ is integrable with respect to $Y(P)$ and that

$$\int_B \phi(y)\,\mathrm{d}Y(P)(y) = \int_{(Y\in B)} X\,\mathrm{d}P$$

for all $B \in \mathcal{E}$.

Firstly, we note that $\phi\circ Y$ is $\sigma(Y)$–measurable (as a result of the trivial implication in Lemma 4.3.2). Furthermore we have

$$\int_\Omega |\phi \circ Y|\,\mathrm{d}P = \int_E |\phi(y)|\,\mathrm{d}Y(P)(y) < \infty$$

using the change–of–variable formula again (simply the argument from above backwards). Finally for $D \in \sigma(Y)$ we have $B \in \mathcal{E}$ with $D = (Y \in B)$ so

$$\int_D \phi \circ Y\,\mathrm{d}P = \int_B \phi(y)\,\mathrm{d}Y(P)(y) = \int_D X\,\mathrm{d}P\,,$$

where we have used the assumption. This shows that $\phi \circ Y$ is a conditional expectation of X given Y, so $\phi \circ Y = E(X|Y)$. From that we have by definition, that $\phi(y)$ is a conditional expectation of X given $Y = y$. □

4.4 Exercises

Exercise 4.1. Assume that ν_1 and ν_2 are bounded, signed measures. Show that $\alpha\nu_1 + \beta\nu_2$ is a bounded, signed measure as well, when $\alpha, \beta \in \mathbb{R}$ are real–valued constants, using the (obvious) definition

$$(\alpha\nu_1 + \beta\nu_2)(A) = \alpha\nu_1(A) + \beta\nu_2(A)\,.$$

$\circ$

Note that the definition $\nu \ll \mu$ also makes sense if μ is a positive measure (not necessarily bounded).

Exercise 4.2. Let τ be the counting measure on $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ (equipped with the σ–algebra $\mathbb{P}(\mathbb{N}_0)$ that contains all subsets). Let μ be the Poisson distribution with parameter λ:

$$\mu(\{n\}) = \frac{\lambda^n}{n!} e^{-\lambda} \qquad \text{for } n \in \mathbb{N}_0\,.$$

Show that $\mu \ll \tau$. Does μ have a density f with respect to τ? In that case find f.

Now let ν be the binomial distribution with parameters (N, p). Decide whether $\mu \ll \nu$ and/or $\nu \ll \mu$. $\circ$

Exercise 4.3. Assume that μ is a bounded, positive measure and that $\nu_1, \nu_2 \ll \mu$ are bounded, signed measures. Show

$$\frac{d(\nu_1 + \nu_2)}{d\mu} = \frac{d\nu_1}{d\mu} + \frac{d\nu_2}{d\mu} \qquad \mu\text{–a.e.}$$

$\circ$

Exercise 4.4. Assume that π, μ are bounded, positive measures and that ν is a bounded, signed measure, such that $\nu \ll \pi \ll \mu$. Show that

$$\frac{d\nu}{d\mu} = \frac{d\nu}{d\pi}\frac{d\pi}{d\mu} \qquad \mu\text{–a.e.}$$

$\circ$

Exercise 4.5. Assume that μ, ν are bounded, positive measures such that $\nu \ll \mu$ and $\mu \ll \nu$. Show

$$\frac{d\nu}{d\mu} = \left(\frac{d\mu}{d\nu}\right)^{-1}.$$

both ν–a.e. and μ–a.e. ◦

Exercise 4.6. Assume that ν is a bounded, signed measure and that μ is a σ–finite measure with $\nu \ll \mu$. Show that there exists $f \in \mathcal{L}(\mu)$ such that $\nu = f \cdot \mu$ (meaning $\nu(F) = \int_F f \, \mathrm{d}\mu$). ◦

In the following exercises we assume that all random variables are defined on a probability space $(\Omega, \mathcal{F}, P)$.

Exercise 4.7. Let X and Y be random variables with $E|X| < \infty$ and $E|Y| < \infty$ that are both measurable with respect to some sub σ–algebra $\mathcal{D}$. Assume furthermore that

$$\int_D X \;\mathrm{d}P = \int_D Y \;\mathrm{d}P \quad \text{for all } D \in \mathcal{D}\,.$$

Show that $X = Y$ a.s. ◦

Exercise 4.8. Assume that X_1 and X_2 are independent random variables satisfying $X_1 \sim \exp(\beta)$ and $X_2 \sim \mathcal{N}(0,1)$. Define $Y = X_1 + X_2$ and the sub σ–algebra $\mathcal{D}$ by $\mathcal{D} = \sigma(X_1)$. Show that $E(Y|\mathcal{D}) = X_1$ a.s. ◦

Exercise 4.9. Assume that X is a real random variable with $EX^2 < \infty$ and that $\mathcal{D}$ is some sub σ–algebra. Let $Y = E(X|\mathcal{D})$. Show that

$$X(P) = Y(P) \quad \Leftrightarrow \quad X = Y \text{ a.s.}$$

◦

Exercise 4.10. Let X and Y be random variables with $EX^2 < \infty$ and $EY^2 < \infty$. The conditional variance of X given the sub σ–algebra $\mathcal{D}$ is defined by

$$V(X|\mathcal{D}) = E(X^2|\mathcal{D}) - (E(X|\mathcal{D}))^2$$

and the conditional covariance between X and Y given $\mathcal{D}$ is

$$\mathrm{Cov}(X,Y|\mathcal{D}) = E(XY|\mathcal{D}) - E(X|\mathcal{D})E(Y|\mathcal{D})$$

Show that

$$\begin{aligned} V(X) &= E(V(X|\mathcal{D})) + V(E(X|\mathcal{D})) \\ \mathrm{Cov}(X,Y) &= E(\mathrm{Cov}(X,Y|\mathcal{D})) + \mathrm{Cov}(E(X|\mathcal{D}), E(Y|\mathcal{D})) \end{aligned}$$

◦

Exercise 4.11. Let X be a real random variable with $E|X| < \infty$. Let $\mathcal{D}$ be a sub σ–algebra. Show without referring to (9) in Theorem 4.2.6 that

$$(E(X|\mathcal{D}))^+ \leq E(X^+|\mathcal{D}) \text{ a.s.}$$

∘

Exercise 4.12. Let X be a real random variable with $E|X| < \infty$. Let $\mathcal{D}$ be a sub σ–algebra. Show without referring to (9) in Theorem 4.2.6 that

$$|E(X|\mathcal{D})| \leq E(|X|\,|\mathcal{D}) \text{ a.s.}$$

∘

Exercise 4.13. Let $(\Omega, \mathcal{F}, P) = ((0,1), \mathcal{B}, \lambda)$ (where λ is the Lebesgue measure on $(0,1)$). Define the real random variable X by

$$X(\omega) = \omega$$

and

$$\mathcal{D} = \{D \subseteq (0,1) \mid D \text{ or } D^c \text{ is countable}\}\,.$$

Then $\mathcal{D}$ is a sub σ–algebra of $\mathcal{B}$ (you can show this if you want...). Find a version of $E(X|\mathcal{D})$.
∘

Exercise 4.14. Let $(\Omega, \mathbb{F}, P) = ([0,1], \mathcal{B}, \lambda)$, where λ is the Lebesgue measure on $[0,1]$. Consider the two real valued random variables

$$X_1(\omega) = 1 - \omega \qquad X_2(\omega) = \omega^2$$

Show that for any given real random variable Y it holds that $E(Y|X_1) = E(Y|X_2)$.

Show by giving an example that $E(Y|X_1 = x)$ and $E(Y|X_2 = x)$ may be different on a set of x's with positive Lebesgue measure. ∘

Exercise 4.15. Assume that X is a real random variable with $E|X| < \infty$ and that $\mathcal{D}$ is a sub σ–algebra of $\mathbb{F}$. Assume that Y is a $\mathcal{D}$–measurable real random variable with $E|Y| < \infty$ that satisfies

$$E(X) = E(Y)$$

and

$$\int_D Y \, \mathrm{d}P = \int_D X \, \mathrm{d}P$$

for all $D \in \mathcal{G}$, where $\mathcal{G}$ is a $\cap$–stable set of subsets of Ω with $\sigma(\mathcal{G}) = \mathcal{D}$.

Show that Y is a conditional expectation of X given $\mathcal{D}$. ∘

Exercise 4.16. Let $\mathbb{X} = (X_1, X_2, \ldots)$ be a stochastic process, and assume that Y and Z are real random variables, such that (Z, Y) is independent of $\mathbb{X}$. Assume that Y has finite expectation.

(1) Show that
$$\int_{(Z\in B,\mathbb{X}\in C)} E(Y|Z)\,\mathrm{d}P = \int_{(Z\in B,\mathbb{X}\in C)} Y\,\mathrm{d}P$$
for all $B \in \mathcal{B}$ and $C \in \mathcal{B}_\infty$.

(2) Show that
$$E(Y|Z) = E(Y|Z, \mathbb{X})$$

∘

Exercise 4.17. Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with $E|X_1| < \infty$. Define $S_n = X_1 + \cdots + X_n$.

(1) Show that $E(X_1|S_n) = E(X_1|S_n, S_{n+1}, S_{n+2}, \ldots)$ a.s.

(2) Show that $\frac{1}{n}S_n = E(X_1|S_n, S_{n+1}, S_{n+2}, \ldots)$ a.s.

∘

Exercise 4.18. Assume that (X, Y) follows the two–dimensional Normal distribution with mean vector (μ_1, μ_2) and covariance matrix
$$\begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$
where $\sigma_{12} = \sigma_{21}$. Then $X \sim \mathcal{N}(\mu_1, \sigma_{11})$, $Y \sim \mathcal{N}(\mu_2, \sigma_{22})$ and $\mathrm{Cov}(X, Y) = \sigma_{12}$.

Show that
$$E(X|Y) = \mu_1 + \beta(Y - \mu_2)\,,$$
where $\beta = \sigma_{12}/\sigma_{22}$. ∘

Chapter 5

Martingales

In this chapter we will present the classical theory of martingales. Martingales are sequences of real random variables, where the index set $\mathbb{N}$ (or $[0, \infty)$) is regarded as a time line, and where – conditionally on the present level – the level of the sequence at a future time point is expected to be as the current level. So it is sequences that evolves over time without a drift in any direction. Similarly, submartingales are expected to have the same or a higher level at future time points, conditioned on the present level.

In Section 5.1 we will give an introduction to the theory based on a motivating example from gambling theory. The basic definitions will be presented in Section 5.2 together with results on the behaviour of martingales observed at random time points. The following Section 5.3 will mainly address the very important martingale theorem, giving conditions under which martingales and submartingales converge. In Section 5.4 we shall introduce the concept of uniform integrability and see how this interplays with martingales. Finally, in Section 5.5 we will prove a central limit theorem for martingales. That is a result that relaxes the independence assumption from Section 3.5.

5.1 Introduction to martingale theory

Let $Y_1, Y_2, \ldots$ be mutually independent identically distributed random variables with

$$P(Y_n = 1) = 1 - P(Y_n = -1) = p$$

where $0 < p < 1$. We will think of Y_n as the result of a game where the probability of winning is p, and where if you bet 1 dollar. and win you receive 1 dollar. and if you lose, you lose the 1 dollar you bet.

If you bet 1 dollar, then, the expected winnings in each game is

$$EY_n = p - (1-p) = 2p - 1,$$

and the game is called favourable if $p > \frac{1}{2}$, fair if $p = \frac{1}{2}$ and unfavourable if $p < \frac{1}{2}$ corresponding to whether EY_n is > 0, $=$ or < 0, respectively.

If the player in each game makes a bet of 1, his (signed) winnings after n games will be $S_n = Y_1 + \cdots + Y_n$. According to the strong law of large numbers,

$$\frac{1}{n} S_n \overset{\text{a.s.}}{\to} 2p - 1,$$

so it follows that if the game is favourable, the player is certain to win in the long run ($S_n > 0$ evt. almost surely) and if the game is unfavourable, the player is certain to lose in the long run.

Undoubtedly, in practice, it is only possible to participate in unfavourable games (unless you happen to be, e.g, a casino or a state lottery). Nevertheless, it may perhaps be possible for a player to turn an unfavourable game into a favourable one, by choosing his bets in a clever fashion Assume that the player has a starting capital of $X_0 > 0$, where X_0 is a constant. Let $\mathcal{F}_n = \sigma(Y_1, \ldots, Y_n)$ for $n \geq 1$. A *strategy* is a sequence $(\phi_n)_{n \geq 1}$ of functions

$$\phi_n : (0, \infty) \times \{-1, 1\}^{n-1} \to [0, \infty),$$

such that the value of

$$\phi_n(X_0, y_1, \ldots, y_{n-1})$$

is the amount the player will bet in the n'th game, when the starting capital is X_0 and $y_1, \ldots, y_{n-1}$ are the results of the $n-1$ first games (the observed values of $Y_1, \ldots, Y_{n-1}$).

A strategy thus allows for the player to take the preceding outcomes into account when he makes his n'th bet.

Note that ϕ_1 is given by X_0 alone, making it constant. Further note that it is possible to let $\phi_n = 0$, corresponding to the player not making a bet, for instance because he or she has been winning up to this point and therefore wishes to stop.

Given the strategy (ϕ_n) the (signed) winnings in the n'th game become

$$Z_n = Y_n \phi_n(X_0, Y_1, \ldots, Y_{n-1})$$

and the capital after the n'th game is

$$X_n = X_0 + \sum_{k=1}^{n} Z_k.$$

It easily follows that for all n, X_n is $\mathcal{F}_n$-measurable, integrable and

$$\begin{aligned} E(X_{n+1}|\mathcal{F}_n) &= E(X_n + Y_{n+1}\phi_{n+1}(X_0, Y_1, \ldots, Y_n)|\mathcal{F}_n) \\ &= X_n + \phi_{n+1}(X_0, Y_1, \ldots, Y_n)E(Y_{n+1}|\mathcal{F}_n) \\ &= X_n + \phi_{n+1}(X_0, Y_1, \ldots, Y_n)(2p-1). \end{aligned}$$

The most important message here is that

$$E(X_{n+1}|\mathcal{F}_n) \begin{smallmatrix}\geq \\ = \\ \leq\end{smallmatrix} X_n \quad \text{for} \quad p \begin{smallmatrix}\geq \\ = \\ <\end{smallmatrix} \frac{1}{2}, \tag{5.1}$$

meaning that $(X_n, \mathcal{F}_n)$ is a submartingale ($\geq$), martingale ($=$) or a supermartingale ($\leq$) (see Definition 5.2.3 below).

For instance, if $p < \frac{1}{2}$, the conditional expected value of the capital after the $n+1$'st game is at most X_n, so the game with strategy (ϕ_n) is at best fair. But what if one simply chooses to focus on the development of the capital at points in time that are advantageous for the player, and where he or she can just decide to quit?

With $0 < p < 1$ infinitely many wins, $Y_n = 1$ and, of course, infinitely many losses, $Y_n = -1$, will occur with probability 1. Let τ_k be the time of the k'th win, i.e.

$$\tau_1 = \inf\{n : Y_n = 1\}$$

and for $k \geq 1$,

$$\tau_{k+1} = \inf\{n > \tau_k : Y_n = 1\}.$$

Each τ_k can be shown to be a stopping time (see Definition 5.2.6 below) and Theorem 5.2.12 provides conditions for when '(X_n) is a supermartingale (fair or unfavourable)' implies that

'(X_{τ_k}) is a supermartingale'. The conditions of Theorem 5.2.12 are for instance met if (X_n) is a supermartingale and we require, not unrealistically, that $X_n \geq a$ always, where a is some given constant. (The player has limited credit and any bet made must, even if the player loses, leave a capital of at least a). It is this result we phrase by stating that it is not possible to turn an unfavourable game into a favourable one.

Even worse, if $p \leq \frac{1}{2}$ and we require that $X_n \geq a$, it can be shown that if there is a minimum amount that one must bet (if one chooses to play) and the player keeps playing, he or she will eventually be ruined! (If $p > \frac{1}{2}$ there will still be a strictly positive probability of ruin, but it is also possible that the capital will beyond all any number).

The result just stated 'only' holds under the assumption of, for instance, all $X_n \geq a$. As we shall see, it is in fact easy to specify strategies such that $X_{\tau_k} \uparrow \infty$ for $k \uparrow \infty$. The problem is that such strategies may well prove costly in the short run.

A classic strategy is to double the amount you bet every game until you win and then start all over with a bet of, e.g., 1, i.e.

$$\phi_1(X_0) = 1,$$

$$\phi_n(X_0, y_1, \ldots, y_{n-1}) = \begin{cases} 2\phi_{n-1}(X_0, y_1, \ldots, y_{n-2}) & \text{if } \ y_{n-1} = -1, \\ 1 & \text{if } \ y_{n-1} = 1. \end{cases}$$

If, say, $\tau_1 = n$, the player loses $\sum_{k=1}^{n-1} 2^{k-1}$ in the $n-1$ first games and wins 2^{n-1} in the n'th game, resulting in the total winnings of

$$-\sum_{1}^{n-1} 2^{k-1} + 2^{n-1} = 1.$$

Thus, at the random time τ_k the total amount won is k and the capital is

$$X_{\tau_k} = X_0 + k.$$

But if p is small, one may experience long strings of consecutive losses and X_n can become very negative.

In the next sections we shall - without referring to gambling - discuss sequences (X_n) of random variables for which the inequalities (1.1) hold. A main result is the martingale convergence theorem (Theorem 5.3.2).

The proof presented here is due to the American probabilist J.L. Doob.

5.2 Martingales and stopping times

Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $(\mathcal{F}_n)_{n\geq 1}$ be a sequence of sub σ-algebras of $\mathcal{F}$, which is increasing $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ for all n. We say that $(\Omega, \mathcal{F}, \mathcal{F}_n, P)$ is a *filtered probability space* with *filtration* $(\mathcal{F}_n)_{n\geq 1}$. The interpretation of a filtration is that we think of n as a point in time and $\mathcal{F}_n$ as consisting of the events that are decided by what happens up to and including time n.

Now let $(X_n)_{n\geq 1}$ be a sequence of random variables defined on $(\Omega, \mathcal{F})$.

Definition 5.2.1. *The sequence (X_n) is* adapted *to $(\mathcal{F}_n)$ if X_n is $\mathcal{F}_n$-measurable for all n.*

Instead of writing that (X_n) is adapted to $(\mathcal{F}_n)$ we often write that $(X_n, \mathcal{F}_n)$ is adapted.

Example 5.2.2. Assume that (X_n) is a sequence of random variables defined on $(\Omega, \mathcal{F}, P)$, and define for each $n \in \mathbb{N}$ the σ–algebra $\mathcal{F}_n$ by

$$\mathcal{F}_n = \sigma(X_1, \dots, X_n)\,.$$

Then $(\mathcal{F}_n)$ is a filtration on $(\Omega, \mathcal{F}, P)$ and $(X_n, \mathcal{F}_n)$ is adapted. ◦

Definition 5.2.3. *An adapted sequence $(X_n, \mathcal{F}_n)$ of real random variables is called a* martingale *if for all $n \in \mathbb{N}$ it holds that $E|X_n| < \infty$ and*

$$E(X_{n+1}|\mathcal{F}_n) = X_n \quad a.s.\,,$$

and a submartingale *if for all $n \in \mathbb{N}$ it holds that $E|X_n| < \infty$ and*

$$E(X_{n+1}|\mathcal{F}_n) \geq X_n \;\; a.s.\,,$$

and a supermartingale *if $(-X_n, \mathcal{F}_n)$ is a submartingale.*

Note that if $(X_n, \mathcal{F}_n)$ is a submartingale (martingale), then for all $m < n \in \mathbb{N}$

$$\begin{aligned} E(X_n|\mathcal{F}_m) &= E(E(X_n|\mathcal{F}_{n-1})|\mathcal{F}_m) \underset{(=)}{\geq} E(X_{n-1}|\mathcal{F}_m) \\ &= E(E(X_{n-1}|\mathcal{F}_{n-2})|\mathcal{F}_m) \underset{(=)}{\geq} \dots \geq X_m \end{aligned}$$

and in particular $EX_n \underset{(=)}{\geq} EX_m$.

The following lemma will be very useful and has a corollary that gives an equivalent formulation of the submartingale (martingale) property.

Lemma 5.2.4. *Suppose that X and Y are real random variables with $E|X| < \infty$ and $E|Y| < \infty$, and let $\mathcal{D}$ be a sub σ–algebra of $\mathcal{F}$ such that X is $\mathcal{D}$–measurable. Then*

$$E(Y|\mathcal{D}) \underset{(=)}{\geq} X \qquad a.s.$$

if and only if

$$\int_D Y \, \mathrm{d}P \underset{(=)}{\geq} \int_D X \, \mathrm{d}P \qquad \text{for all } D \in \mathcal{D}\,.$$

Proof. Since both $E(Y|\mathcal{D})$ and X are $\mathcal{D}$–measurable, we have that

$$E(Y|\mathcal{D}) \underset{(=)}{\geq} X \tag{5.2}$$

if and only if

$$\int_D E(Y|\mathcal{D}) \, \mathrm{d}P \underset{(=)}{\geq} \int_D X \, \mathrm{d}P$$

for all $D \in \mathcal{D}$ (The "if" implication should be obvious. For "only if", consider the integral $\int_D (E(Y|\mathcal{D}) - X) \, \mathrm{d}P$, where $D = (E(Y|\mathcal{D}) < X)$). The left integral above equals $\int_D Y \, \mathrm{d}P$ because of the definition of conditional expectations, so (5.2) holds if and only if

$$\int_D Y \, \mathrm{d}P \underset{(=)}{\geq} \int_D X \, \mathrm{d}P$$

for all $D \in \mathcal{D}$. □

Corollary 5.2.5. *Assume that $(X_n \mathcal{F}_n)$ is adapted with $E|X_n| < \infty$ for all $n \in \mathbb{N}$. Then $(X_n, \mathcal{F}_n)$ is a submartingale (martingale) if and only if for all $n \in \mathbb{N}$*

$$\int_F X_{n+1} \, \mathrm{d}P \underset{(=)}{\geq} \int_F X_n \, \mathrm{d}P$$

for all $F \in \mathcal{F}_n$.

When handling martingales and submartingales it is often fruitful to study how they behave at random time points of a special type called *stopping times*.

Definition 5.2.6. *A* stopping time *is a random variable $\tau : \Omega \to \overline{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ such that*

$$(\tau = n) \in \mathcal{F}_n$$

for all $n \in \mathbb{N}$. If $\tau < \infty$ we say that τ is a finite stopping time.

Example 5.2.7. Let (X_n) be a sequence of real random variables, and define the filtration $(\mathcal{F}_n)$ by $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Assume that τ is a stopping time with respect to this filtration, and consider the set $(\tau = n)$ that belongs to $\mathcal{F}_n$. Since $\mathcal{F}_n$ is generated by the vector $(X_1, \dots, X_n)$ there exists a set $B \in \mathcal{B}_n$ such that

$$(\tau = n) = \big((X_1, \dots, X_n) \in B_n\big)\,.$$

The implication of this is, that we are able to read off from the values of $X_1, \dots, X_n$ whether $\tau = n$ or not. So by observing the sequence (X_n) for some time, we know if τ has occurred or not. ◦

We have an equivalent definition of a stopping time:

Lemma 5.2.8. *A random variable $\tau : \Omega \to \overline{\mathbb{N}}$ is a stopping time if and only if*

$$(\tau \leq n) \in \mathcal{F}_n$$

for all $n \in \mathbb{N}$.

Proof. Firstly, assume that τ is a stopping time. We can write

$$(\tau \leq n) = \cup_{k=1}^{n} (\tau = k)$$

which belongs to $\mathcal{F}_n$, since each $(\tau = k) \in \mathcal{F}_k$ and the filtration $(\mathcal{F}_n)$ is increasing, so $\mathcal{F}_k \subseteq \mathcal{F}_n$ for $k \leq n$.

Assume conversely, that $(\tau \leq n) \in \mathcal{F}_n$ for all n. Then the stopping time property follows from

$$(\tau = n) = (\tau \leq n) \backslash (\tau \leq n-1)\,,$$

since $(\tau \leq n) \in \mathcal{F}_n$ and $(\tau \leq n-1) \in \mathcal{F}_{n-1} \subseteq \mathcal{F}_n$. □

Example 5.2.9. If $n_0 \in \mathbb{N}$ then the constant function $\tau = n_0$ is a stopping time:

$$(\tau = n) \in \{\emptyset, \Omega\} \subseteq \mathcal{F}_n \qquad \text{for all } n \in \mathbb{N}\,.$$

If σ and τ are stopping times then $\sigma \wedge \tau, \sigma \vee \tau$ and $\sigma + \tau$ are also stopping times. E.g. for $\sigma \wedge \tau$ write

$$(\sigma \wedge \tau \leq n) = (\sigma \leq n) \cup (\tau \leq n)$$

and note that $(\sigma \leq n), (\tau \leq n) \in \mathcal{F}_n$. ◦

We now define a σ-algebra $\mathcal{F}_\tau$, which consists of all the events that are decided by what happens up to and including the random time τ.

Consider for τ a stopping time the collection of sets

$$\mathcal{F}_\tau = \{F \in \mathcal{F} : F \cap (\tau = n) \in \mathcal{F}_n \quad \text{for all } n \in \mathbb{N}\}.$$

Then we have

Theorem 5.2.10. *(1) $\mathcal{F}_\tau$ is a σ-algebra.*

(2) If σ, τ are stopping times and $\sigma \leq \tau$, then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.

Proof. (1) We have

$$\Omega \cap (\tau = n) = (\tau = n) \in \mathcal{F}_n \text{ for all } n \in \mathbb{N}\,,$$

since τ is a stopping time. Hence $\Omega \in \mathcal{F}_\tau$. Now assume that $F \in \mathcal{F}_\tau$. Then by definition $F \cap (\tau = n) \in \mathcal{F}_n$ for all $n \in \mathbb{N}$, so

$$F^c \cap (\tau = n) = (\tau = n) \setminus \big(F \cap (\tau = n)\big) \in \mathcal{F}_n$$

for all $n \in \mathbb{N}$. This shows that $F^c \in \mathcal{F}_n$. Finally assume that $F_1, F_2, \ldots \in \mathcal{F}_\tau$. Then

$$\big(\bigcap_{k=1}^{\infty} F_k\big) \cap (\tau = n) = \bigcap_{k=1}^{\infty} \big(F_k \cap (\tau = n)\big) \in \mathcal{F}_n$$

for all $n \in \mathbb{N}$.

Altogether it is shown that $\mathcal{F}_\tau$ is a σ–algebra.

(2) Assume that $F \in \mathcal{F}_\sigma$. Since $\sigma \leq \tau$ we can write

$$\begin{aligned} F \cap (\tau = n) &= \bigcup_{k=1}^{n} F \cap (\sigma = k, \tau = n) \\ &= \bigcup_{k=1}^{n} \big((F \cap (\sigma = k)) \cap (\tau = n)\big) \in \mathcal{F}_n \end{aligned}$$

using that for $k \leq n$ we have $\mathcal{F}_k \subseteq \mathcal{F}_n$ and $F \cap (\sigma = k) \in \mathcal{F}_k$. Hence by definition we have $F \in \mathcal{F}_\tau$. □

With τ a finite stopping time, we consider the process (X_n) at the random time τ and define

$$X_\tau(\omega) = X_{\tau(\omega)}(\omega).$$

From now on we only consider real-valued X_n's.

Although definition of $\mathcal{F}_\tau$ may not seem very obvious, Theorem 5.2.11 below shows that both X_τ and τ are $\mathcal{F}_\tau$–measurable. Hence certain events at time τ are $\mathcal{F}_\tau$–measurable, and the intuitive interpretation of $\mathcal{F}_\tau$ as consisting of all events up to time τ is still reasonable.

Theorem 5.2.11. *If $(X_n, \mathcal{F}_n)$ is adapted and τ is a finite stopping time, then both τ and X_τ are $\mathcal{F}_\tau$-measurable.*

Proof. The proof is straightforward manipulations. If we consider the event $(\tau = k)$ then we have

$$(\tau = k) \cap (\tau = n) = \begin{cases} (\tau = n) & \text{if } k = n \\ \emptyset & \text{if } k \neq n \end{cases},$$

so in both cases we get that $(\tau = k) \cap (\tau = n) \in \mathcal{F}_n$, and hence $(\tau = k) \in \mathcal{F}_\tau$. This shows the measurability of τ.

For the second statement, let $B \in \mathcal{B}$ and realize that for all n,

$$(X_\tau \in B) \cap (\tau = n) = (X_n \in B) \cap (\tau = n) \in \mathcal{F}_n,$$

implying that $(X_\tau \in B) \in \mathcal{F}_\tau$ as desired. □

Theorem 5.2.12 (Optional sampling, first version)**.** *Let $(X_n, \mathcal{F}_n)$ be a submartingale (martingale) and assume that σ and τ are finite stopping times with $\sigma \leq \tau$. If $E|X_\tau| < \infty$, $E|X_\sigma| < \infty$ and*

$$\liminf_{N\to\infty} \int_{(\tau>N)} X_N^+ dP = 0$$
$$(\liminf_{N\to\infty} \int_{(\tau>N)} |X_N| dP = 0),$$

then

$$E(X_\tau|\mathcal{F}_\sigma) \underset{(=)}{\geq} X_\sigma\,.$$

Proof. According to Theorem 5.2.4 we only need to show that

$$\int_A X_\tau \,\mathrm{d}P \underset{(=)}{\geq} \int_A X_\sigma \,\mathrm{d}P \tag{5.3}$$

for all $A \in \mathcal{F}_\sigma$. So we fix A in the following and define $D_j = A \cap (\sigma = j)$. If we can show

$$\int_{D_j} X_\tau \, \mathrm{d}P \underset{(=)}{\geq} \int_{D_j} X_\sigma \, \mathrm{d}P \tag{5.4}$$

for all $j = 1, 2, \ldots$, then (5.3) will follow since then

$$\int_A X_\tau \, \mathrm{d}P = \sum_{j=1}^{\infty} \int_{D_j} X_\tau \, \mathrm{d}P \underset{(=)}{\geq} \sum_{j=1}^{\infty} \int_{D_j} X_\sigma \, \mathrm{d}P = \int_A X_\sigma \, \mathrm{d}P$$

In the two equalities we have used dominated convergence: E.g. for the first equality we have the integrable upper bound $|X_\tau|$, so

$$\begin{aligned}\int_A X_\tau \, \mathrm{d}P &= \int \sum_{j=1}^{\infty} 1_{D_j} X_\tau \, \mathrm{d}P = \int \lim_{M\to\infty} \sum_{j=1}^{M} 1_{D_j} X_\tau \, \mathrm{d}P \\ &= \lim_{M\to\infty} \int \sum_{j=1}^{M} 1_{D_j} X_\tau \, \mathrm{d}P = \lim_{M\to\infty} \sum_{j=1}^{M} \int 1_{D_j} X_\tau \, \mathrm{d}P = \sum_{j=1}^{\infty} \int_{D_j} X_\tau \, \mathrm{d}P\end{aligned}$$

Hence the argument will be complete if we can show (5.4). For this, first define for $N \geq j$

$$I_N = \int_{D_j \cap (\tau \leq N)} X_\tau \, \mathrm{d}P + \int_{D_j \cap (\tau > N)} X_N \, \mathrm{d}P \, .$$

We claim that

$$I_j \leq I_{j+1} \underset{(=)}{\leq} I_{j+2} \underset{(=)}{\leq} \cdots$$

For $N > j$ we get

$$\begin{aligned}I_N &= \int_{D_j \cap (\tau \leq N)} X_\tau \, \mathrm{d}P + \int_{D_j \cap (\tau > N)} X_N \, \mathrm{d}P \\ &= \int_{D_j \cap (\tau < N)} X_\tau \, \mathrm{d}P + \int_{D_j \cap (\tau = N)} X_\tau \, \mathrm{d}P + \int_{D_j \cap (\tau > N)} X_N \, \mathrm{d}P \\ &= \int_{D_j \cap (\tau < N)} X_\tau \, \mathrm{d}P + \int_{D_j \cap (\tau = N)} X_N \, \mathrm{d}P + \int_{D_j \cap (\tau > N)} X_N \, \mathrm{d}P \\ &= \int_{D_j \cap (\tau < N)} X_\tau \, \mathrm{d}P + \int_{D_j \cap (\tau \geq N)} X_N \, \mathrm{d}P\end{aligned}$$

Note that $(\tau \geq N) = (\tau \leq N-1)^c \in \mathcal{F}_{N-1}$. Also note that $D_j = A \cap (\sigma = j) \in \mathcal{F}_j$ from the definition of $\mathcal{F}_\sigma$ (since it is assumed that $A \in \mathcal{F}_\sigma$). Then (recall that $j < N$) $D_j \cap (\tau \geq N) \in \mathcal{F}_{N-1}$, so Theorem 5.2.4 yields

$$\int_{D_j \cap (\tau \geq N)} X_N \, \mathrm{d}P \underset{(=)}{\geq} \int_{D_j \cap (\tau \geq N)} X_{N-1} \, \mathrm{d}P$$

So we have

$$\begin{aligned} I_N &= \int_{D_j\cap(\tau<N)} X_\tau \,\mathrm{d}P + \int_{D_j\cap(\tau\geq N)} X_N \,\mathrm{d}P \\ &\underset{(=)}{\geq} \int_{D_j\cap(\tau\leq N-1)} X_\tau \,\mathrm{d}P + \int_{D_j\cap(\tau>N-1)} X_{N-1} \,\mathrm{d}P = I_{N-1} \end{aligned}$$

and thereby the sequence $(I_N)_{N\geq j}$ is shown to be increasing. For the left hand side in (5.4) this implies that

$$\begin{aligned} \int_{D_j} X_\tau \,\mathrm{d}P &= \int_{D_j\cap(\tau\leq N)} X_\tau \,\mathrm{d}P + \int_{D_j\cap(\tau>N)} X_\tau \,\mathrm{d}P \\ &\quad + \int_{D_j\cap(\tau>N)} X_N \,\mathrm{d}P - \int_{D_j\cap(\tau>N)} X_N \,\mathrm{d}P \\ &= I_N + \int_{D_j\cap(\tau>N)} X_\tau \,\mathrm{d}P - \int_{D_j\cap(\tau>N)} X_N \,\mathrm{d}P \\ &\underset{(=)}{\geq} I_j + \int_{D_j\cap(\tau>N)} X_\tau \,\mathrm{d}P - \int_{D_j\cap(\tau>N)} X_N \,\mathrm{d}P \\ &\geq I_j + \int_{D_j\cap(\tau>N)} X_\tau \,\mathrm{d}P - \int_{D_j\cap(\tau>N)} X_N^+ \,\mathrm{d}P \qquad (5.5) \\ &\Big(I_j + \int_{D_j\cap(\tau>N)} X_\tau \,\mathrm{d}P - \int_{D_j\cap(\tau>N)} X_N \,\mathrm{d}P\Big) \end{aligned}$$

Recall the assumption $\sigma\leq\tau$. Then

$$D_j\cap(\tau\leq j) = A\cap(\sigma=j)\cap(\tau\leq j) = D_j\cap(\tau=j)\,,$$

so

$$\begin{aligned} I_j &= \int_{D_j\cap(\tau\leq j)} X_\tau \,\mathrm{d}P + \int_{D_j\cap(\tau>j)} X_j \,\mathrm{d}P \\ &= \int_{D_j\cap(\tau=j)} X_j \,\mathrm{d}P + \int_{D_j\cap(\tau>j)} X_j \,\mathrm{d}P \\ &= \int_{D_j} X_j \,\mathrm{d}P = \int_{A\cap(\sigma=j)} X_j \,\mathrm{d}P = \int_{D_j} X_\sigma \,\mathrm{d}P \qquad (5.6) \end{aligned}$$

Hence we have shown (5.4) if we can show that the two last terms in (5.5) can be ignored. Since $(\tau>N)\downarrow\emptyset$ for $N\to\infty$ and X_τ is integrable, we have from dominated convergence that

$$\lim_{N\to\infty}\int_{D_j\cap(\tau>N)} X_\tau \,\mathrm{d}P = \lim_{N\to\infty}\int 1_{D_j\cap(\tau>N)}X_\tau \,\mathrm{d}P = \int \lim_{N\to\infty} 1_{D_j\cap(\tau>N)}X_\tau \,\mathrm{d}P = 0 \quad (5.7)$$

And because of the assumption from the theorem, we must have a subsequence of natural numbers $N_1, N_2, \ldots$ such that

$$\lim_{\ell \to \infty} \int_{(\tau > N_\ell)} X_{N_\ell}^+ \,\mathrm{d}P = 0$$
$$\Big(\lim_{\ell \to \infty} \int_{(\tau > N_\ell)} |X_{N_\ell}| \,\mathrm{d}P = 0\Big)$$

Hence from using (5.7) we have

$$\lim_{\ell \to \infty} \Big(\int_{D_j \cap (\tau > N_\ell)} X_\tau \,\mathrm{d}P - \int_{D_j \cap (\tau > N_\ell)} X_{N_\ell}^+ \,\mathrm{d}P \Big) = 0$$

and combining this with (5.5) and (5.6) yields

$$\int_{D_j} X_\tau \,\mathrm{d}P \underset{(=)}{\geq} I_j = \int_{D_j} X_\sigma \,\mathrm{d}P$$

which is (5.4). □

Corollary 5.2.13. *Let $(X_n, \mathcal{F}_n)$ be a submartingale (martingale), and let $\sigma \leq \tau$ be bounded stopping times. Then $E|X_\tau| < \infty$, $E|X_\sigma| < \infty$ and*

$$E(X_\tau | \mathcal{F}_\sigma) \underset{(=)}{\geq} X_\sigma$$

Proof. We show that the conditions from Theorem 5.2.12 are fulfilled. There exists $K < \infty$ such that $\sup_{\omega \in \Omega} \tau(\omega) \leq K$. Then

$$E|X_\tau| = \int \Big|\sum_{k=1}^K 1_{(\tau=k)} X_k\Big| \,\mathrm{d}P \leq \sum_{k=1}^K \int 1_{(\tau=k)} |X_k| \,\mathrm{d}P \leq \sum_{k=1}^K \int |X_k| \,\mathrm{d}P = \sum_{k=1}^K E|X_k| < \infty\,.$$

That $E|X_\sigma| < \infty$ follows similarly. Furthermore it must hold that $(\tau > N) = \emptyset$ for all $N \geq K$, so obviously

$$\int_{(\tau > N)} X_N^+ \,\mathrm{d}P = 0$$

for all $N \geq K$. □

We can also translate Theorem 5.2.12 into a result concerning the process considered at a sequence of stopping times. Firstly, we need to specify the sequence of stopping times.

Definition 5.2.14. *A sequence $(\tau_n)_{n\geq 1}$ of positive random variables is a sequence of* sampling times *if it is increasing and each τ_n is a finite stopping time.*

With (τ_n) a sequence of sampling times it holds, according to 1. in Theorem 5.2.10, that $\mathcal{F}_{\tau_n} \subseteq \mathcal{F}_{\tau_{n+1}}$ for all n. If $(X_n, \mathcal{F}_n)$ is adapted then, according to Theorem 5.2.11, X_{τ_n} is $\mathcal{F}_{\tau_n}$-measurable. Hence, the sampled sequence $(X_{\tau_n}, \mathcal{F}_{\tau_n})$ is adapted.

Theorem 5.2.15. *Let $(X_n, \mathcal{F}_n)$ be a submartingale (martingale) and let (τ_k) be a sequence of sampling times. If*

$$E|X_{\tau_k}| < \infty \quad \textit{for all } k, \tag{a}$$

$$\liminf_{N\to\infty} \int_{(\tau_k > N)} X_N^+ dP = 0 \quad \textit{for all } k, \tag{b}$$

$$\left(\liminf_{N\to\infty} \int_{(\tau_k > N)} |X_N| dP = 0 \quad \textit{for all } k\right),$$

then $(X_{\tau_k}, \mathcal{F}_{\tau_k})$ is a submartingale (martingale).

Proof. Use Theorem 5.2.12 for each k separately. □

5.3 The martingale convergence theorem

We shall need the following result on transformations of martingales and submartingales.

Lemma 5.3.1. *Assume that $(X_n, \mathcal{F}_n)$ is an adapted sequence.*

(1) If $(X_n, \mathcal{F}_n)$ is a martingale, then both $(|X_n|, \mathcal{F}_n)$ and $(X_n^+, \mathcal{F}_n)$ are submartingales.

(2) If $(X_n, \mathcal{F}_n)$ is a submartingale, then $(X_n^+, \mathcal{F}_n)$ is a submartingale.

Proof. (1) Assume that $(X_n, \mathcal{F}_n)$ is a martingale. Then X_n is $\mathcal{F}_n$–measurable, so also $|X_n|$ and X_n^+ are $\mathcal{F}_n$–measurable. Furthermore we have $E(X_{n+1}|\mathcal{F}_n) = X_n$ a.s., so that $E(X_{n+1}|\mathcal{F}_n)^+ = X_n^+$ a.s. and $|E(X_{n+1}|\mathcal{F}_n)| = |X_n|$ a.s. We also have $EX_n^+ < \infty$ (recall that $0 \leq X_n^+ \leq |X_n|$) and obviously $E|X_n| < \infty$.

Then

$$E(X_{n+1}^+|\mathcal{F}_n) \geq E(X_{n+1}|\mathcal{F}_n)^+ = X_n^+ \text{ a.s.},$$

where the inequality follows from (9) in Theorem 4.2.6, since the function $x \mapsto x^+$ is convex. Similarly since $x \mapsto |x|$ is convex, (9) in Theorem 4.2.6 gives that

$$E(|X_{n+1}||\mathcal{F}_n) \geq |E(X_{n+1}|\mathcal{F}_n)| = |X_n| \text{ a.s.}$$

which proves (1).

(2) If $(X_n, \mathcal{F}_n)$ is a submartingale, we similarly have that X_n^+ is $\mathcal{F}_n$–measurable with $EX_n^+ < \infty$. Furthermore it holds

$$E(X_{n+1}|\mathcal{F}_n) \geq X_n \text{ a.s.},$$

and since $x \mapsto x^+$ is increasing, we also have

$$E(X_{n+1}|\mathcal{F}_n)^+ \geq X_n^+ \text{ a.s.}.$$

We obtain

$$E(X_{n+1}^+|\mathcal{F}_n) \geq E(X_{n+1}|\mathcal{F}_n)^+ \geq X_n^+ \text{ a.s.}$$

□

We shall now prove the main theorem of classic martingale theory.

Theorem 5.3.2 (The martingale convergence theorem)**.** *If $(X_n, \mathcal{F}_n)$ is a submartingale such that $\sup_n EX_n^+ < \infty$, then $X = \lim_{n\to\infty} X_n$ exists almost surely and $E|X| < \infty$.*

The proof is given below. Note that, cf. Lemma 5.3.1 the sequence EX_n^+ is increasing, so the assumption $\sup_n EX_n^+ < \infty$ is equivalent to assuming

$$\lim_{n\to\infty} EX_n^+ < \infty.$$

The proof is based on a criterion for convergence of a sequence of real numbers, which we shall now discuss.

Let $(x_n)_{n\geq 1}$ be a sequence of real numbers. For $a < b$ consider

$$n_1 = \inf\{n \geq 1 \mid x_n \geq b\}, \quad m_1 = \inf\{n > n_1 \mid x_n \leq a\}$$

and recursively define

$$n_k = \inf\{n > m_{k-1} \mid x_n \geq b\}, \quad m_k = \inf\{n > n_k \mid x_n \leq a\},$$

always using the convention $\inf \emptyset = \infty$. We now define the number of *down-crossings from b to a* for the sequence (x_n) as $+\infty$ if all $m_k < \infty$ and as k if $m_k < \infty$ and $m_{k+1} = \infty$ (in particular, 0 down-crossings in the case $m_1 = \infty$). Note that $n_1 \leq m_1 \leq n_2 \leq m_2 \ldots$ with equality only possible, if the common value is ∞.

Lemma 5.3.3. *The limit* $x = \lim_{n\to\infty} x_n$ *exists (as a limit in* $\overline{\mathbb{R}} = [-\infty, \infty]$*) if and only if for all rational* $a < b$ *it holds that the number of down-crossings from* b *to* a *for* (x_n) *is finite.*

Proof. Firstly, consider the "only if" implication. So assume that the limit $x = \lim_{n\to\infty}$ exists in $\overline{\mathbb{R}}$ and let $a < b$ be given. We must have that either $x > a$ or $x < b$ (if not both of them are true). If for instance $x > a$ then we must have some n_0 such that

$$x_n > a \qquad \text{for all } n \geq n_0\,,$$

since e.g there exists n_0 with $|x - x_n| < x - a$ for all $n \geq n_0$. But then we must have that $m_k = \infty$ for k large enough, which makes the number of down–crossings finite. The case $x < b$ is treated analogously.

Now we consider the "if" part of the result, so assume that the limit $\lim_{n\to\infty} x_n$ does not exist. Then $\liminf_{n\to\infty} x_n < \limsup_{n\to\infty} x_n$, so in particular we can find rational numbers $a < b$ such that

$$\liminf_{n\to\infty} x_n < a < b < \limsup_{n\to\infty} x_n\,.$$

This implies that $x_n \leq a$ for infinitely many n and $x_n \geq b$ for infinitely many n. Especially we must have that the number of down–crossings from b to a is infinite. □

In the following proofs we shall apply the result to a sequence (X_n) of real random variables. For this we will need some notation similar to the definition of n_k and m_k above. Define the random times

$$\sigma_1 = \inf\{n \mid X_n \geq b\},\quad \tau_1 = \inf\{n > \sigma_1 \mid X_n \leq a\},$$

and recursively

$$\sigma_k = \inf\{n > \tau_{k-1} \mid X_n \geq b\},\quad \tau_k = \inf\{n > \sigma_k \mid X_n \leq a\}.$$

We have that $\sigma_1 \leq \tau_1 \leq \sigma_2 \leq \tau_2 \leq \ldots$, where equality is only possible if the common value is ∞.

Furthermore, all τ_k and σ_k are stopping times: We can write

$$(\sigma_1 = n) = \Big(\bigcup_{i=1}^{n-1}(X_i < b)\Big) \cap (X_n \geq b)$$

which is in $\mathcal{F}_n$, since all $X_1, \ldots, X_n$ are $\mathcal{F}_n$–measurable, implying that $(X_i > a) \in \mathcal{F}_n$ (when $i < n$) and $(X_n \leq a) \in \mathcal{F}_n$. Furthermore if e.g. σ_k is a stopping time, then

$$(\tau_k = n) = \bigcap_{j=1}^{n-1}(\sigma_k = j) \cap (X_{j+1} > a, \ldots, X_{n-1} > a, X_n \leq a)$$

which belongs to $\mathcal{F}_n$, since $(\sigma_k = j) \in \mathcal{F}_n$ and all the X–variables involved are $\mathcal{F}_n$–measurable. Hence τ_k is a stopping time as well.

Let $\beta_{ab}(\omega)$ be the number of down–crossings from b to a for the sequence $(X_n(\omega))$ for each $\omega \in \Omega$. Then we have (we suppress the ω)

$$\beta_{ab} = \sum_{k=1}^{\infty} 1_{(\tau_k < \infty)}$$

so we see that β_{ab} is an integer–valued random variable. With this notation, we can also define the number of down–crossings β_{ab}^N from b to a in the time interval $\{1, \dots, N\}$ as

$$\beta_{ab}^N = \sum_{k=1}^{\infty} 1_{(\tau_k \leq N)} = \sum_{k=1}^{N} 1_{(\tau_k \leq N)} .$$

The last equality follows since we necessarily must have $\tau_{N+1} \geq N+1$ (either $\tau_{N+1} = \infty$ or the strict inequality holds, making $\sigma_1 < \tau_1 < \cdots < \sigma_{N+1} < \tau_{N+1}$).

Finally note, that $\beta_{ab}^N \uparrow \beta_{ab}$ as $N \to \infty$.

Proof of Theorem 5.3.2. In order to show that $X = \lim X_n$ exists as a limit in $\overline{\mathbb{R}}$ it is, according to Lemma 5.3.3, sufficient to show that

$$P(\beta_{ab} \text{ is finite for all rational } a < b) = 1 ,$$

But it follows directly from the down–crossing lemma (Lemma 5.3.4), that for all rational pairs $a < b$ we have $P(\beta_{ab} < \infty) = 1$. Hence also

$$1 = P\Big(\bigcap_{a<b \text{ rational}} (\beta_{ab} < \infty) \Big) = P(\beta_{ab} \text{ is finite for all rational } a < b) .$$

We still need to show that $E|X| < \infty$. In the affirmative, we will also know that X is finite almost surely. Otherwise we would have

$$P(|X| = \infty) = \epsilon > 0$$

such that

$$E|X| \geq E(1_{(|X|=\infty)} \cdot \infty) = \infty \cdot \epsilon = \infty$$

which is a contradiction.

So we prove that $E|X| < \infty$. Fatou's Lemma yields

$$E|X| = E(\lim_{n\to\infty} |X_n|) = E(\liminf_{n\to\infty} |X_n|) \leq \liminf_{n\to\infty} E|X_n|$$

Note that since $(X_n, \mathcal{F}_n)$ is a submartingale, then

$$EX_1 \le EX_n = EX_n^+ - EX_n^- .$$

which implies that $EX_n^- \le EX_n^+ - EX_1$ such that

$$E|X_n| = EX_n^+ + EX_n^- \le 2EX_n^+ - EX_1 .$$

Then we obtain that

$$E|X| \le \liminf_{n\to\infty} E|X_n| \le 2 \liminf_{n\to\infty} EX_n^+ - EX_1 = 2 \lim_{n\to\infty} EX_n^+ - EX_1 < \infty ,$$

due to the assumption in Theorem 5.3.2. □

The most significant and most difficult part of the proof of Theorem 5.3.2 is contained in the next result.

Lemma 5.3.4 (The down–crossing lemma)**.** *If $(X_n, \mathcal{F}_n)$ is a submartingale then it holds, for all $N \in \mathbb{N}, a < b$, that*

$$E\beta_{ab}^N \le \frac{E(X_N - b)^+}{b-a}, \tag{5.8}$$

$$E\beta_{ab} \le \frac{1}{b-a} \lim_{n\to\infty} E(X_n - b)^+. \tag{5.9}$$

In particular, $\beta_{ab} < \infty$ a.s. if $\lim_{n\to\infty} EX_n^+ < \infty$.

Proof. The last claim follows directly from (5.9) and the inequality $(X_n - b)^+ \le X_n^+ + |b|$, since it is assumed in the theorem that $\sup EX_n^+ = \lim EX_n^+ < \infty$.

Note that for all $N, k \in \mathbb{N}$ it holds that $1_{(\sigma_k \le N)} = 1_{(\tau_k \le N)} + 1_{(\sigma_k \le N < \tau_k)}$. Then we can write

$$\begin{aligned}
\sum_{k=1}^N (X_{\tau_k \wedge N} - X_{\sigma_k \wedge N}) 1_{(\sigma_k \le N)} &= \sum_{k=1}^N (X_{\tau_k \wedge N} - X_{\sigma_k \wedge N})(1_{(\tau_k \le N)} + 1_{(\sigma_k \le N < \tau_k)}) \\
&= \sum_{k=1}^N (X_{\tau_k} - X_{\sigma_k}) 1_{(\tau_k \le N)} + \sum_{k=1}^N (X_N - X_{\sigma_k}) 1_{(\sigma_k \le N < \tau_k)} \\
&\le (a-b) \sum_{k=1}^N 1_{(\tau_k \le N)} + (X_N - b) \sum_{k=1}^N 1_{(\sigma_k \le N < \tau_k)} \\
&\le (a-b)\beta_{ab}^N + (X_N - b)^+ .
\end{aligned}$$

In the first inequality we have used, that if $\tau_k < \infty$, then $X_{\sigma_k} \geq b$ and $X_{\tau_k} \leq a$, and also if only $\sigma_k < \infty$ it holds that $X_{\sigma_k} \geq b$. By rearranging the terms above, we obtain the inequality

$$\beta_{ab}^N \leq \frac{(X_N - b)^+}{b-a} - \frac{1}{b-a} \sum_{k=1}^{N} (X_{\tau_k \wedge N} - X_{\sigma_k \wedge N}) 1_{(\sigma_k \leq N)} \,. \tag{5.10}$$

Note that for each k we have that $\sigma_k \wedge N \leq \tau_k \wedge N$ are bounded stopping times. Hence Corollary 5.2.13 yields that $E|X_{\sigma_k \wedge N}| < \infty$, $E|X_{\tau_k \wedge N}| < \infty$ and $E(X_{\tau_k \wedge N} | \mathcal{F}_{\sigma_k \wedge N}) \geq X_{\sigma_k \wedge N}$. Then according to Lemma 5.2.4 it holds that

$$\int_F X_{\tau_k \wedge N} \, \mathrm{d}P \geq \int_F X_{\sigma_k \wedge N} \, \mathrm{d}P$$

for all $F \in \mathcal{F}_{\sigma_k \wedge N}$. Since

$$(\sigma_k \leq N) \cap (\sigma_k \wedge N = n) = \begin{cases} (\sigma_k = n) & n \leq N \\ \emptyset & n > N \end{cases} \in \mathcal{F}_n$$

for all $n \in \mathbb{N}$, we must have that $(\sigma_k \leq N) \in \mathcal{F}_{\sigma_k \wedge N}$ which implies that

$$E\Big((X_{\tau_k \wedge N} - X_{\sigma_k \wedge N}) 1_{(\sigma_k \leq N)} \Big) = \int_{(\sigma_k \leq N)} X_{\tau_k \wedge N} \, \mathrm{d}P - \int_{(\sigma_k \leq N)} X_{\sigma_k \wedge N} \, \mathrm{d}P \geq 0 \,.$$

Then the sum in (5.10) has positive expectation, so in particular we have

$$E\beta_{ab}^N \leq \frac{E(X_N - b)^+}{b-a}$$

Finally note that $(X_n - b, \mathcal{F}_n)$ is a submartingale, since $(X_n, \mathcal{F}_n)$ is a submartingale. Hence the sequence $((X_n - b)^+, \mathcal{F}_n)$ will be a submartingale as well according to 2. in Lemma 5.3.1, such that $E(X_N - b)^+$ is increasing and thereby convergent. So applying monotone convergence to the left hand side (recall $\beta_{ab}^N \uparrow \beta_{ab}$) leads to

$$E\beta_{ab} \leq \frac{1}{b-a} \lim_{n \to \infty} E(X_n - b)^+ \,.$$

□

It is useful to highlight the following immediate consequences of Theorem 5.3.2.

Corollary 5.3.5. *1. If $(X_n, \mathcal{F}_n)$ is a submartingale and there exists $c \in \mathbb{R}$ such that $X_n \leq c$ a.s. for all n, then $X = \lim X_n$ exists almost surely and $E|X| < \infty$.*

2. If $(X_n, \mathcal{F}_n)$ is a martingale and there exists $c \in \mathbb{R}$ such that either $X_n \leq c$ a.s. for all n or $X_n \geq c$ a.s. for all n, then $X = \lim X_n$ exists almost surely and $E|X| < \infty$. □

5.4 Martingales and uniform integrability

If $(X_n, \mathcal{F}_n)$ is a martingale then in particular $EX_n = EX_1$ for all n, but even though $X_n \overset{\text{a.s.}}{\to} X$ and $E|X| < \infty$, it is not necessarily true that $EX = EX_1$. We shall later obtain some results where not only $EX = EX_1$ but where in addition, the martingale $(X_n, \mathcal{F}_n)$ has a number of other attractive properties. Let $(X_n)_{n\geq 1}, X$ be real random variables with $E|X_n| < \infty, \forall n$ and $E|X| < \infty$. Recall that by $X_n \xrightarrow{\mathcal{L}^1} X$ we mean that

$$\lim_{n\to\infty} E|X_n - X| = 0,$$

and if this is the case then

$$\lim_{n\to\infty} |\int_F X_n \, dP - \int_F X \, dP| = 0$$

will hold for all $F \in \mathcal{F}$, (because $|\int_F X_n \, dP - \int_F X \, dP| \leq E|X_n - X|$). In particular the $\mathcal{L}^1$–convergence gives that $EX_1 = EX$. We are looking for a property that implies this $\mathcal{L}^1$–convergence.

Definition 5.4.1. *A family $(X_i)_{i\in I}$ of real random variables is* uniformly integrable *if $E|X_i| < \infty$ for all $i \in I$ and*

$$\lim_{x\to\infty} \sup_I \int_{(|X_i|>x)} |X_i| dP = 0.$$

Example 5.4.2. (1) The family $\{X\}$ consisting of only one real variable is uniformly integrable if $E|X| < \infty$:

$$\lim_{x\to\infty} \int_{(|X|>x)} |X| \, \mathrm{d}P = \int \lim_{x\to\infty} 1_{(|X|>x)} |X| \, \mathrm{d}P = 0$$

by dominated convergence (since $1_{(|X|>x)}|X| \leq |X|$, and $|X|$ is integrable).

(2) Now consider a finite family $(X_i)_{i=1,\dots,n}$ of real random variables. This family is uniformly integrable, if each $E|X_i| < \infty$:

$$\lim_{x\to\infty} \sup_{i\in I} \int_{(|X_i|>x)} |X_i| \, \mathrm{d}P = \lim_{x\to\infty} \max_{i=1,\dots,n} \int_{(|X_i|>x)} |X_i| \, \mathrm{d}P = 0 \,,$$

since each integral has limit 0 because of (1). ◦

Example 5.4.3. Let $(X_i)_{i\in I}$ be a family of real random variables. If $\sup_{i\in I} E|X_i|^r < \infty$ for some $r > 1$, then (X_i) is uniformly integrable:

$$\int_{(|X_i|>x)} |X_i| \,\mathrm{d}P \le \int_{(|X_i|>x)} \Big|\frac{X_i}{x}\Big|^{r-1} |X_i| \,\mathrm{d}P = \frac{1}{x^{r-1}} \int_{(|X_i|>x)} |X_i|^r \,\mathrm{d}P$$
$$\le \frac{1}{x^{r-1}} E|X_i|^r \le \frac{1}{x^{r-1}} \sup_{j\in I} E|X_j|^r$$

so we obtain

$$\sup_{i\in I} \int_{(|X_i|>x)} |X_i| \,\mathrm{d}P \le \frac{1}{x^{r-1}} \sup_{j\in I} E|X_j|^r$$

which has limit 0 as $x \to \infty$. ◦

We have the following equivalent definition of uniform integrability:

Theorem 5.4.4. *The family $(X_i)_{i\in I}$ is uniformly integrable if and only if,*

(1) $\sup_I E|X_i| < \infty$,

(2) there for all $\epsilon > 0$ exists a $\delta > 0$ such that

$$\sup_{i\in I} \int_A |X_i| dP < \epsilon$$

for all $A \in \mathcal{F}$ with $P(A) < \delta$.

Proof. First we demonstrate the "only if" statement. So assume that (X_i) is uniformly integrable. For all $x > 0$ we have for all $i \in I$ that

$$E|X_i| = \int_\Omega |X_i| \,\mathrm{d}P = \int_{(|X_i|\le x)} |X_i| \,\mathrm{d}P + \int_{(|X_i|>x)} |X_i| \,\mathrm{d}P$$
$$\le xP(|X_i| \le x) + \int_{(|X_i|>x)} |X_i| \,\mathrm{d}P \le x + \int_{(|X_i|>x)} |X_i| \,\mathrm{d}P$$

so also

$$\sup_{i\in I} E|X_i| \le x + \sup_{i\in I} \int_{(|X_i|>x)} |X_i| \,\mathrm{d}P\,.$$

The last term is assumed to $\to 0$ as $x \to \infty$. In particular it is finite for x sufficiently large, so we conclude that $\sup_{i\in I} E|X_i|$ is finite, which is (1).

To show (2) let $\epsilon > 0$ be given. Then for all $A \in \mathcal{F}$ we have (with a similar argument to the one above)

$$\begin{aligned}\int_A |X_i| \, \mathrm{d}P &= \int_{A\cap(|X_i|\leq x)} |X_i| \, \mathrm{d}P + \int_{A\cap(|X_i|>x)} |X_i| \, \mathrm{d}P \\ &\leq xP(A \cap (|X_i| \leq x)) + \int_{A\cap(|X_i|>x)} |X_i| \, \mathrm{d}P \\ &\leq xP(A) + \int_{(|X_i|>x)} |X_i| \, \mathrm{d}P\end{aligned}$$

so

$$\sup_{i\in I} \int_A |X_i| \, \mathrm{d}P \leq xP(A) + \sup_{i\in I} \int_{(|X_i|>x)} |X_i| \, \mathrm{d}P \, .$$

Now choose $x = x_0 > 0$ according to the assumption of uniform integrability such that

$$\sup_{i\in I} \int_{(|X_i|>x_0)} |X_i| \, \mathrm{d}P < \frac{\epsilon}{2} \, .$$

Then for $A \in \mathcal{F}$ with $P(A) < \frac{\epsilon}{2x_0}$ we must have

$$\sup_{i\in I} \int_A |X_i| \, \mathrm{d}P < x_0 \frac{\epsilon}{2x_0} + \frac{\epsilon}{2} = \epsilon$$

so if we choose $\delta = \frac{\epsilon}{2x_0}$ we have shown (2).

For the "if" part of the theorem, assume that both (1) and (2) hold. Assume furthermore that it is shown that

$$\lim_{x\to\infty} \sup_{i\in I} P(|X_i| > x) = 0 \, . \tag{5.11}$$

In order to obtain that the definition of uniform integrability is fulfilled, let $\epsilon > 0$ be given. We want to find $x_0 > 0$ such that

$$\sup_{i\in I} \int_{(|X_i|>x)} |X_i| \, \mathrm{d}P \leq \epsilon$$

for $x > x_0$. Find the $\delta > 0$ corresponding to ϵ according to (2) and then let x_0 satisfy that

$$\sup_{i\in I} P(|X_i| > x) < \delta$$

for $x > x_0$. Now for all $x > x_0$ and $i \in I$ we have $P(|X_i| > x) < \delta$ such that (because of (2))

$$\int_{(|X_i|>x)} |X_i| \, \mathrm{d}P \leq \sup_{j\in I} \int_{(|X_i|>x)} |X_j| \, \mathrm{d}P \leq \epsilon$$

so also

$$\sup_{i\in I} \int_{(|X_i|>x)} |X_i| \, \mathrm{d}P \leq \epsilon \, .$$

Hence the proof is complete, if we can show (5.11). But for $x > 0$ it is obtained from Markov's inequality that

$$\sup_{i\in I} P(|X_i| > x) \le \frac{1}{x} \sup_{i\in I} E|X_i|$$

and the last term $\to 0$ as $x \to \infty$, since $\sup_{i\in I} E|X_i|$ is finite by the assumption (1). □

The next result demonstrates the importance of uniform integrability if one aims to show $\mathcal{L}^1$ convergence.

Theorem 5.4.5. *Let $(X_n)_{n\ge1}, X$ be real random variables with $E|X_n| < \infty$ for all n. Then $E|X| < \infty$ and $X_n \xrightarrow{\mathcal{L}^1} X$ if and only if $(X_n)_{n\ge1}$ is uniformly integrable and $X_n \xrightarrow{P} X$.*

Proof. Assume that $E|X| < \infty$ and $X_n \xrightarrow{\mathcal{L}^1} X$. Then in particular $X_n \xrightarrow{P} X$.

We will show that (X_n) is uniformly integrable by showing (1) and (2) from Theorem 5.4.4 are satisfied.

That (1) is satisfied follows from the fact that

$$E|X_n| = E|X_n - X + X| \le E|X_n - X| + E|X|$$

so

$$\sup_{n\ge1} E|X_n| \le E|X| + \sup_{n\ge1} E|X_n - X|\,.$$

Since each $E|X_n - X| \le E|X_n| + E|X| < \infty$ and the sequence $(E|X_n - X|)_{n\in\mathbb{N}}$ converges to 0 (according to the $\mathcal{L}^1$–convergence), it must be bounded, so $\sup_{n\ge1} E|X_n - X| < \infty$.

To show (2) first note that for $A \in \mathcal{F}$

$$\int_A |X_n|\,\mathrm{d}P = \int_A |X_n - X + X|\,\mathrm{d}P \le \int_A |X_n - X|\,\mathrm{d}P + \int_A |X|\,\mathrm{d}P \le E|X_n - X| + \int_A |X|\,\mathrm{d}P\,.$$

Now let $\epsilon > 0$ be given and choose n_0 so that $E|X_n - X| < \frac{\epsilon}{2}$ for $n > n_0$. Furthermore (since the one–member family $\{X\}$ is uniformly integrable) we can choose $\delta_1 > 0$ such that

$$\int_A |X|\,\mathrm{d}P < \frac{\epsilon}{2} \quad \text{if } P(A) < \delta_1\,.$$

Then we have

$$\int_A |X_n|\,\mathrm{d}P \le E|X_n - X| + \int_A |X|\,\mathrm{d}P < \epsilon$$

whenever $n > n_0$ and $P(A) < \delta_1$.

Now choose $\delta_2 > 0$ (since the finite family $(X_n)_{1\le n\le n_0}$ is uniformly integrable) such that

$$\max_{1\le n\le n_0} \int_A |X_n| \, \mathrm{d}P < \epsilon$$

if $P(A) < \delta_2$. We have altogether with $\delta = \delta_1 \wedge \delta_2$ that for all $n \in \mathbb{N}$ it holds that $\int_A |X_n| \, \mathrm{d}P < \epsilon$ if $P(A) < \delta$, and this shows (2) since then

$$\sup_{n\ge 1} \int_A |X_n| \, \mathrm{d}P \le \epsilon \quad \text{ if } P(A) < \delta \, .$$

For the converse implication, assume that $(X_n)_{n\ge 1}$ is uniformly integrable and $X_n \xrightarrow{P} X$. Then we can choose (according to Theorem 1.2.13) a subsequence $(n_k)_{k\ge 1}$ such that $X_{n_k} \xrightarrow{\text{a.s.}} X$. From Fatou's Lemma and the fact that (1) is satisfied we obtain

$$E|X| = E \lim_{k\to\infty} |X_{n_k}| = E \liminf_{k\to\infty} |X_{n_k}| \le \liminf_{k\to\infty} E|X_{n_k}| \le \sup_{k\ge 1} E|X_{n_k}| \le \sup_{n\ge 1} E|X_n| < \infty \, .$$

In order to show that $E|X_n - X| \to 0$ let $\epsilon > 0$ be given. We obtain for all $n \in \mathbb{N}$

$$\begin{aligned} E|X_n - X| &= \int_{(|X_n - X| \le \frac{\epsilon}{2})} |X_n - X| \, \mathrm{d}P + \int_{(|X_n - X| > \frac{\epsilon}{2})} |X_n - X| \, \mathrm{d}P \\ &\le \frac{\epsilon}{2} + \int_{(|X_n - X| > \frac{\epsilon}{2})} |X_n - X| \, \mathrm{d}P \\ &\le \frac{\epsilon}{2} + \int_{(|X_n - X| > \frac{\epsilon}{2})} |X_n| \, \mathrm{d}P + \int_{(|X_n - X| > \frac{\epsilon}{2})} |X| \, \mathrm{d}P \end{aligned}$$

In accordance with (2) applied to the two families (X_n) and (X) choose $\delta > 0$ such that

$$\sup_{m\in\mathbb{N}} \int_A |X_m| \, \mathrm{d}P < \frac{\epsilon}{4} \qquad \text{and} \qquad \int_A |X| \, \mathrm{d}P < \frac{\epsilon}{4}$$

for all $A \in \mathcal{F}$ with $P(A) < \delta$. Since $X_n \xrightarrow{P} X$ we can find $n_0 \in \mathbb{N}$ such that $P(|X_n - X| > \frac{\epsilon}{2}) < \delta$ for $n \ge n_0$. Then for $n \ge n_0$ we have

$$\begin{aligned} &\int_{(|X_n - X| > \frac{\epsilon}{2})} |X_n| \, \mathrm{d}P \le \sup_{m\ge 1} \int_{(|X_n - X| > \frac{\epsilon}{2})} |X_m| \, \mathrm{d}P < \frac{\epsilon}{4} \\ &\int_{(|X_n - X| > \frac{\epsilon}{2})} |X| \, \mathrm{d}P < \frac{\epsilon}{4} \end{aligned}$$

so we have shown that for $n \ge n_0$ it holds

$$E|X_n - X| < \frac{\epsilon}{2} + \frac{\epsilon}{4} + \frac{\epsilon}{4} = \epsilon$$

which completes the proof. □

After this digression into the standard theory of integration we now return to adapted sequences, martingales and submartingales.

Definition 5.4.6. *If* $(X_n, \mathcal{F}_n)$ *is a submartingale (martingale) and* Y *is a real random variable with* $E|Y| < \infty$, *we say that* Y closes *the submartingale (martingale) if*

$$E(Y|\mathcal{F}_n) \geq X_n, \quad (E(Y|\mathcal{F}_n) = X_n) \quad a.s.$$

for all $n \geq 1$.

Theorem 5.4.7. *(1) Let* $(X_n, \mathcal{F}_n)$ *be a submartingale. If* $(X_n^+)_{n\geq 1}$ *is uniformly integrable, then* $X = \lim_{n\to\infty} X_n$ *exists almost surely,* X *closes the the submartingale and* $X_n^+ \xrightarrow{\mathcal{L}^1} X^+$.

A sufficient condition for uniform integrability of (X_n^+) *is the existence of a random variable* Y *that closes the submartingale.*

(2) Let $(X_n, \mathcal{F}_n)$ *be a martingale. If* (X_n) *is uniformly integrable, then* $X = \lim_{n\to\infty} X_n$ *exists almost surely,* X *closes the martingale and* $X_n \xrightarrow{\mathcal{L}^1} X$.

A sufficient condition for uniform integrability of (X_n) *is the existence of a random variable* Y *that closes the martingale.*

Proof. (1) We start with the final claim, so assume that there exists Y such that

$$E(Y|\mathcal{F}_n) \geq X_n \quad \text{a.s.}$$

for all $n \in \mathbb{N}$. Then from (9) in Theorem 4.2.6 we obtain

$$X_n^+ \leq (E(Y|\mathcal{F}_n))^+ \leq E(Y^+|\mathcal{F}_n) \quad \text{a.s.} \tag{5.12}$$

and taking expectations on both sides yields $EX_n^+ \leq EY^+$ so also $\sup_n EX_n^+ \leq EY^+ < \infty$. Then according to the martingale convergence theorem (Theorem 5.3.2) we have that $X = \lim_{n\to\infty} X_n$ exists almost surely. Since $(X_n(\omega))_{n\to\infty}$ is convergent for almost all $\omega \in \Omega$ it is especially bounded (not by the same constant for each ω), and in particular $\sup_n X_n^+(\omega) < \infty$ for almost all ω. We obtain that for all $x > 0$ and $n \in \mathbb{N}$

$$\int_{(X_n^+ > x)} X_n^+ \, \mathrm{d}P \leq \int_{(X_n^+ > x)} E(Y^+|\mathcal{F}_n) \, \mathrm{d}P = \int_{(X_n^+ > x)} Y^+ \, \mathrm{d}P \leq \int_{(\sup_k X_k^+ > x)} Y^+ \, \mathrm{d}P \,,$$

where the first inequality is due to (5.12) and the equality follows from the definition of conditional expectations, since X_n is $\mathcal{F}_n$–measurable such that $(X_n^+ > x) \in \mathcal{F}_n$. Since this is true for all $n \in \mathbb{N}$ we have

$$\sup_{n\in\mathbb{N}} \int_{(X_n^+>x)} X_n^+ \,\mathrm{d}P \leq \int_{(\sup_n X_n^+>x)} Y^+ \,\mathrm{d}P$$

Furthermore we have that Y^+ is integrable with $1_{(\sup_n X_n^+>x)} Y^+ \leq Y^+$ for all x, and obviously since $\sup_n X_n^+ < \infty$ almost surely, we have $1_{(\sup X_n^+>x)} Y^+ \to 0$ a.s. when $x \to \infty$. Then from Dominated convergence the right hand integral above will $\to 0$ as $x \to \infty$. Hence we have shown that (X_n^+) is uniformly integrable.

Now we return to the first statement, so assume that (X_n^+) is uniformly integrable. In particular (according to 1 in Theorem 5.4.4) we have $\sup_n EX_n^+ < \infty$. Then The Martingale convergence Theorem yields that $X = \lim_{n\to\infty} X_n$ exists almost surely with $E|X| < \infty$. Obviously we must also have $X_n^+ \xrightarrow{\text{a.s.}} X^+$, and then Theorem 5.4.5 implies that $X_n^+ \xrightarrow{\mathcal{L}^1} X^+$.

In order to show that

$$E(X|\mathcal{F}_n) \geq X_n \quad \text{a.s.}$$

for all $n \in \mathbb{N}$ it is equivalent (according to Lemma 5.2.4) to show that for all $n \in \mathbb{N}$ it holds

$$\int_F X_n \,\mathrm{d}P \leq \int_F X \,\mathrm{d}P \tag{5.13}$$

for all $F \in \mathcal{F}_n$. For $F \in \mathcal{F}_n$ and $n \leq N$ we have (according to Corollary 5.2.5, since $(X_n, \mathcal{F}_n)$ is a submartingale)

$$\int_F X_n \,\mathrm{d}P \leq \int_F X_N \,\mathrm{d}P = \int_F X_N^+ \,\mathrm{d}P - \int_F X_N^- \,\mathrm{d}P\,.$$

Since it is shown that $X_N^+ \xrightarrow{\mathcal{L}^1} X^+$ we have from the remark just before Definition 5.4.1 that

$$\lim_{N\to\infty} \int_F X_N^+ \,\mathrm{d}P = \int_F X^+ \,\mathrm{d}P$$

Furthermore Fatou's lemma yields (when using that $X_N \xrightarrow{\text{a.s.}} X$ so $X_N^- \xrightarrow{\text{a.s.}} X^-$ and thereby $\liminf_{N\to\infty} X_N^- = X^-$ a.s.)

$$\limsup_{N\to\infty} \Big(-\int_F X_N^- \,\mathrm{d}P \Big) = -\liminf_{N\to\infty} \int_F X_N^- \,\mathrm{d}P \leq -\int_F \liminf_{N\to\infty} X_N^- \,\mathrm{d}P = -\int_F X^- \,\mathrm{d}P\,.$$

When combining the obtained inequalities we have for all $n \in \mathbb{N}$ and $F \in \mathcal{F}_n$ that

$$\begin{aligned}\int_F X_n \,\mathrm{d}P &\leq \limsup_{N\to\infty}\Big(\int_F X_N^+ \,\mathrm{d}P - \int_F X_N^- \,\mathrm{d}P\Big)\\ &= \lim_{N\to\infty}\int_F X_N^+ \,\mathrm{d}P + \limsup_{N\to\infty}\Big(-\int_F X_N^- \,\mathrm{d}P\Big)\\ &\leq \int_F X^+ \,\mathrm{d}P - \int_F X^- \,\mathrm{d}P = \int_F X \,\mathrm{d}P\end{aligned}$$

which is the inequality (5.13) we were supposed to show.

(2) Once again we start proving the last claim, so assume that Y closes the martingale, so $E(Y|\mathcal{F}_n) = X_n$ a.s. for all $n \in \mathbb{N}$. From (10) in Theorem 4.2.6 we have

$$|X_n| = |E(Y|\mathcal{F}_n)| \leq E(|Y||\mathcal{F}_n) \quad \text{a.s.}\,.$$

Similar to the argument in 1 we then have $\sup_{n\geq 1} E|X_n| \leq E|Y| < \infty$ so in particular $\sup_{n\geq 1} EX_n^+ < \infty$. Hence $X = \lim_{n\to\infty} X_n$ exists almost surely leading to the fact that $\sup_{n\geq 1}|X_n| < \infty$ almost surely. Then for all $n \in \mathbb{N}$ and $x > 0$

$$\int_{(|X_n|>x)} |X_n| \,\mathrm{d}P \leq \int_{(|X_n|>x)} E(|Y||\mathcal{F}_n)\,\mathrm{d}P = \int_{(|X_n|>x)} |Y| \,\mathrm{d}P \leq \int_{(\sup_k |X_k|>x)} |Y| \,\mathrm{d}P\,,$$

where the last integral $\to 0$ as $x \to \infty$ as a result of dominated convergence. Hence

$$0 \leq \lim_{x\to\infty}\sup_{n\in\mathbb{N}}\int_{(|X_n|>x)} |X_n| \,\mathrm{d}P \leq \lim_{x\to\infty}\int_{(\sup_k |X_k|>x)} |Y| \,\mathrm{d}P = 0$$

so (X_n) is uniformly integrable.

Finally assume that (X_n) is uniformly integrable. Then $\sup_n E|X_n| < \infty$ (from Theorem 5.4.4) and in particular $\sup_n EX_n^+ < \infty$. According to the martingale convergence theorem we have $X_n \xrightarrow{\text{a.s.}} X$ with $E|X| < \infty$. From Theorem 5.4.5 we have $X_n \xrightarrow{\mathcal{L}^1} X$ which leads to (see the remark just before Definition 5.4.1)

$$\lim_{N\to\infty}\int_F X_N \,\mathrm{d}P = \int_F X \,\mathrm{d}P\,.$$

for all $F \in \mathcal{F}$. We also have from the martingale property of $(X_n, \mathcal{F}_n)$ that for all $n \leq N$ and $F \in \mathcal{F}_n$

$$\int_F X_n \,\mathrm{d}P = \int_F X_N \,\mathrm{d}P$$

so we must have

$$\int_F X_n \,\mathrm{d}P = \int_F X \,\mathrm{d}P$$

for all $n \in \mathbb{N}$ and $F \in \mathcal{F}_n$. This shows that $E(X|\mathcal{F}_n) = X_n$ a.s., so X closes the martingale.
□

An important warning is the following: Let $(X_n, \mathcal{F}_n)$ be a submartingale and assume that $(X_n^+)_{n\geq 1}$ is uniformly integrable. As we have seen, we then have $X_n \xrightarrow{\text{a.s.}} X$ and $X_n^+ \xrightarrow{\mathcal{L}^1} X^+$, but we do not in general have $X_n \xrightarrow{\mathcal{L}^1} X$. If, e.g., $(X_n^-)_{n\geq 1}$ is also uniformly integrable, then it is true that $X_n \xrightarrow{\mathcal{L}^1} X$ since $X_n^- \xrightarrow{\text{a.s.}} X^-$ implies that $X_n^- \xrightarrow{\mathcal{L}^1} X^-$ by Theorem 5.4.5 and then $E|X_n - X| \leq E|X_n^+ - X^+| + E|X_n^- - X^-| \to 0$.

Theorem 5.4.8. *Let $(\mathcal{F}_n)_{n\in\mathbb{N}}$ be a filtration on $(\Omega, \mathcal{F}, P)$ and let Y be a random variable with $E|Y| < \infty$. Define for all $n \in \mathbb{N}$*

$$X_n = E(Y|\mathcal{F}_n)\,.$$

Then $(X_n, \mathcal{F}_n)$ is a martingale, and $X_n \to X$ a.s. and in $\mathcal{L}^1$, where

$$X = E(Y|\mathcal{F}_\infty)$$

with $\mathcal{F}_\infty \subseteq \mathcal{F}$ the smallest σ–algebra containing $\mathcal{F}_n$ for all $n \in \mathbb{N}$.

Proof. Clearly, $(X_n, \mathcal{F}_n)$ is adapted and $E|X_n| < \infty$ and since

$$E(X_{n+1}|\mathcal{F}_n) = E(E(Y|\mathcal{F}_{n+1})|\mathcal{F}_n) = E(Y|\mathcal{F}_n) = X_n,$$

we see that $(X_n, \mathcal{F}_n)$ is a martingale. It follows directly from the definition of X_n that Y closes the martingale, so by 2 in Theorem 5.4.7 it holds that $X = \lim X_n$ exists a.s. and that $X_n \xrightarrow{\mathcal{L}^1} X$.

The remaining part of the proof is to show that

$$X = E(Y|\mathcal{F}_\infty)\,. \tag{5.14}$$

For this, first note that $X = \lim_{n\to\infty} X_n$ can be chosen $\mathcal{F}_\infty$–measurable, since

$$F = (\lim_{n\to\infty} X_n \text{ exists}) = \bigcap_{N=1}^{\infty} \bigcup_{k=1}^{\infty} \bigcap_{m,n=k}^{\infty} \Big(|X_n - X_m| \leq \frac{1}{N}\Big) \in \mathcal{F}_\infty\,,$$

and X can be defined as

$$X = \lim_{n\to\infty} X_n 1_F\,,$$

where each $X_n 1_F$ is $\mathcal{F}_\infty$–measurable, making X measurable with respect to $\mathcal{F}_\infty$ as well. Hence, in order to show (5.14), we only need to show that

$$\int_F X \,\mathrm{d}P = \int_F Y \,\mathrm{d}P$$

for all $F \in \mathcal{F}_\infty$. Note that $\mathcal{F}_\infty = \sigma\big(\cup_{n=1}^\infty \mathcal{F}_n\big)$ where $\cup_{n=1}^\infty \mathcal{F}_n$ is $\cap$–stable. Then according to Exercise 4.15 it is enough to show the equality above for all $F \in \mathcal{F}_n$ for any given $n \in \mathbb{N}$. So let $n \in \mathbb{N}$ be given and assume that $F \in \mathcal{F}_n$. Then we have for all $N \geq n$ that

$$\int_F X_N \, \mathrm{d}P = \int_F Y \, \mathrm{d}P$$

since $X_N = E(Y|\mathcal{F}_N)$ and $F \in \mathcal{F}_n \subseteq \mathcal{F}_N$. Furthermore we have $X_N \xrightarrow{\mathcal{L}^1} X$, so

$$\int_F X \, \mathrm{d}P = \lim_{N\to\infty} \int_F X_N \, \mathrm{d}P$$

which leads to the conclusion that

$$\int_F X \, \mathrm{d}P = \int_F Y \, \mathrm{d}P$$

□

In conclusion of this chapter, we shall discuss an extension of the optional sampling theorem (Theorem 5.2.12).

Let $(X_n, \mathcal{F}_n)_{n\geq 1}$ be a (usual) submartingale or martingale, and let τ be an arbitrary stopping time. If $\lim X_n = X$ exists a.s., we can define X_τ even if τ is not finite as is assumed in Theorem 5.2.12:

$$X_\tau(\omega) = \begin{cases} X_{\tau(\omega)}(\omega) & \text{if } \tau(\omega) < \infty \\ X(\omega) & \text{if } \tau(\omega) = \infty. \end{cases}$$

With this, X_τ is a $\mathcal{F}_\tau$-measurable random variable which is only defined if (X_n) converges a.s.

If $\sigma \leq \tau$ are two stopping times, we are interested in investigating when

$$E(X_\tau|\mathcal{F}_\sigma) \geq X_\sigma \,, \tag{5.15}$$

if $(X_n, \mathcal{F}_n)$ is a submartingale, and when

$$E(X_\tau|\mathcal{F}_\sigma) = X_\sigma \,, \tag{5.16}$$

if $(X_n, \mathcal{F}_n)$ is a martingale.

By choosing $\tau = \infty$, $\sigma = n$ for a given, arbitrary, $n \in \mathbb{N}$, the two conditions amount to the demand that X closes the submartingale, respectively the martingale. So if (5.15), (5.16) is to hold for *all* pairs $\sigma \leq \tau$ of stopping times, we must consider submartingales or martingales, which satisfy the conditions from Theorem 5.4.7. For pairs with $\tau < \infty$ we can use Theorem 5.2.12 to obtain weaker conditions.

Theorem 5.4.9 (Optional sampling, full version)**.** *Let $(X_n, \mathcal{F}_n)$ be a submartingale (martingale) and let $\sigma \leq \tau$ be stopping times. If one of the three following conditions is satisfied, then $E|X_\tau| < \infty$ and*

$$E(X_\tau|\mathcal{F}_\sigma) \underset{(=)}{\geq} X_\sigma.$$

(1) τ is bounded.
(2) $\tau < \infty, E|X_\tau| < \infty$ and

$$\liminf_{N\to\infty} \int_{(\tau>N)} X_N^+ dP = 0$$

$$(\liminf_{N\to\infty} \int_{(\tau>N)} |X_N| dP = 0).$$

(3) $(X_n^+)_{n\geq 1}$ is uniformly integrable ($(X_n)_{n\geq 1}$ is uniformly integrable).

If 3 holds, then $\lim X_n = X$ exists a.s. with $E|X| < \infty$ and for an arbitrary stopping time τ it holds that

$$E|X_\tau| \leq 2EX^+ - EX_1. \tag{5.17}$$

Proof. That the conclusion is true under assumption (1) is simply Corollary 5.2.13. Comparing condition (2) with the conditions in Theorem 5.2.12 shows, that if (2) implies that $E|X_\sigma| < \infty$, then the argument concerning condition (2) is complete. For this consider the increasing sequence of bounded stopping times given by $(\sigma \wedge n)_{n\geq 1}$. For each n the pair of stopping times $\sigma \wedge n \leq \sigma \wedge (n+1)$ fulfils the conditions of Corollary 5.2.13, so $E|X_{\sigma\wedge n}| < \infty$, $E|X_{\sigma\wedge(n+1)}| < \infty$, and

$$E(X_{\sigma\wedge(n+1)}|\mathcal{F}_{\sigma\wedge n}) \underset{(=)}{\geq} X_{\sigma\wedge n} \quad \text{a.s.}\,,$$

which shows that the adapted sequence $(X_{\sigma\wedge n}, \mathcal{F}_{\sigma\wedge n})$ is a submartingale (martingale). We have similarly that the pair of stopping times $\sigma \wedge n \leq \tau$ fulfils the conditions from Theorem 5.2.12 (because of the assumption (2) and that $E|X_{\sigma\wedge n}| < \infty$ as argued above). Hence the theorem yields that for each $n \in \mathbb{N}$ we have

$$E(X_\tau|\mathcal{F}_{\sigma\wedge n}) \underset{(=)}{\geq} X_{\sigma\wedge n} \quad \text{a.s.}\,,$$

which shows that X_τ closes the submartingale (martingale) $(X_{\sigma\wedge n}, \mathcal{F}_{\sigma\wedge n})$. Hence according to Theorem 5.4.7 it converges almost surely, where the limit is integrable. Since obviously, $X_{\sigma\wedge n} \xrightarrow{\text{a.s.}} X_\sigma$ we conclude that $E|X_\sigma| < \infty$ as desired.

We finally show that (3) implies (5.15) if $(X_n, \mathcal{F}_n)$ is a submartingale. That (3) implies (5.16) if $(X_n, \mathcal{F}_n)$ is a martingale, is then seen as follows: From fact that $(|X_n|)$ is uniformly integrable we have that both (X_n^+) and (X_n^-) are uniformly integrable. Since both $(X_n, \mathcal{F}_n)$ and $(-X_n, \mathcal{F}_n)$ are submartingales (with (X_n^+) and $((-X_n)^+)$ uniformly integrable) the result for submartingales can be applied to obtain

$$E(X_\tau|\mathcal{F}_\sigma) \geq X_\sigma \qquad \text{and} \qquad E(X_\tau|\mathcal{F}_\sigma) \leq X_\sigma$$

from which the desired result can be derived.

So, assume that $(X_n, \mathcal{F}_n)$ is a submartingale and that (X_n^+) is uniformly integrable. From (1) in Theorem 5.4.7 we have that $X = \lim_{n\to\infty} X_n$ exists almost surely and that X closes the submartingale $(X_n, \mathcal{F}_n)$. Since $(X_n^+, \mathcal{F}_n)$ is a submartingale as well we can apply Theorem 5.4.7 again and obtain that X^+ closes the submartingale $(X_n^+, \mathcal{F}_n)$. Using the notation $X_\infty^+ = X^+$ we get

$$\begin{aligned} EX_\tau^+ &= \int \sum_{1\leq n\leq\infty} 1_{(\tau=n)} X_n^+ \,\mathrm{d}P = \sum_{1\leq n\leq\infty} \int_{(\tau=n)} X_n^+ \,\mathrm{d}P \\ &\leq \sum_{1\leq n\leq\infty} \int_{(\tau=n)} X^+ \,\mathrm{d}P = \int \sum_{1\leq n\leq\infty} 1_{(\tau=n)} X^+ \,\mathrm{d}P = \int X^+ \,\mathrm{d}P = EX^+ \end{aligned}$$

at the inequality we have used Lemma 5.2.4 and $(\tau = n) \in \mathcal{F}_n$ since $E(X^+|\mathcal{F}_n) \geq X_n^+$. Let $N \in \mathbb{N}$. Then $\tau \wedge N$ is a bounded stopping time with $\tau \wedge N \uparrow \tau$ as $N \to \infty$. Applying the inequality above to $\tau \wedge N$ yields

$$EX_{\tau\wedge N}^+ \leq EX^+ . \tag{5.18}$$

Furthermore we have according to part (1) in the theorem (since $1 \leq \tau \wedge N$ are bounded stopping times), that

$$EX_1 \leq EX_{\tau\wedge N} . \tag{5.19}$$

Then combining (5.18) and (5.19) gives

$$EX_{\tau\wedge N}^- = EX_{\tau\wedge N}^+ - EX_{\tau\wedge N} \leq EX^+ - EX_1$$

such that by Fatou's lemma it holds that

$$EX_\tau^- = E \liminf_{N\to\infty} X_{\tau\wedge N}^- \leq \liminf_{N\to\infty} EX_{\tau\wedge N}^- \leq EX^+ - EX_1$$

Thereby we have

$$E|X_\tau| = EX_\tau^+ + EX_\tau^- \leq 2EX^+ - EX_1$$

which in particular is finite.

The proof will be complete, if we can show

$$E(X_\tau|\mathcal{F}_\sigma) \geq X_\sigma$$

which is equivalent to showing

$$\int_F X_\sigma \,\mathrm{d}P \le \int_F X_\tau \,\mathrm{d}P$$

for all $F \in \mathcal{F}_\sigma$. For showing this inequality it will suffice to show

$$\int_{F\cap(\sigma=k)} X_k \,\mathrm{d}P \le \int_{F\cap(\sigma=k)} X_\tau \,\mathrm{d}P \tag{5.20}$$

for all $F \in \mathcal{F}_\sigma$ and $k \in \mathbb{N}$ since in that case we can obtain (using dominated convergence, since $E|X_\tau| < \infty$ and $E|X_\sigma| < \infty$)

$$\int_F X_\sigma \,\mathrm{d}P = \sum_{1\le k\le\infty} \int_{F\cap(\sigma=k)} X_k \,\mathrm{d}P \le \sum_{1\le k\le\infty} \int_{F\cap(\sigma=k)} X_\tau \,\mathrm{d}P = \int_F X_\tau \,\mathrm{d}P$$

and we obviously have

$$\int_{F\cap(\sigma=\infty)} X_\infty \,\mathrm{d}P = \int_{F\cap(\sigma=\infty)} X_\tau \,\mathrm{d}P\,.$$

So we will show the inequality (5.20). The Theorem in the 1–case yields (since $\sigma\wedge N \le \tau\wedge N$ are bounded stopping times) that $E(X_{\tau\wedge N}|\mathcal{F}_{\sigma\wedge N}) \ge X_{\sigma\wedge N}$ a.s. Now let $F_k = F\cap(\sigma=k)$ and assume that $N \ge k$. Then $F_k \in \mathcal{F}_{\sigma\wedge N}$:

$$F_k \cap (\sigma\wedge N = n) = \begin{cases} F\cap(\sigma=n) \in \mathcal{F}_n & n=k \\ \emptyset \in \mathcal{F}_n & n \ne k \end{cases}$$

From this we obtain

$$\begin{aligned} \int_{F_k} X_k \,\mathrm{d}P = \int_{F_k} X_{\sigma\wedge N} \,\mathrm{d}P &\le \int_{F_k} X_{\tau\wedge N} \,\mathrm{d}P \\ &= \int_{F_k\cap(\tau\le N)} X_\tau \,\mathrm{d}P + \int_{F_k\cap(\tau>N)} X_N \,\mathrm{d}P\,. \end{aligned}$$

We will spend the rest of the proof on finding an upper limit of the two terms, as $N\to\infty$. Considering the first term, we have from dominated convergence (since $|1_{F_k\cap(\tau\le N)}X_\tau| \le |X_\tau|$, X_τ is integrable, and $(\tau\le N)\uparrow(\tau<\infty)$) that

$$\lim_{N\to\infty} \int_{F_k\cap(\tau\le N)} X_\tau \,\mathrm{d}P = \int_{F_k\cap(\tau<\infty)} X_\tau \,\mathrm{d}P$$

For the second term we will use that $F_k \in \mathcal{F}_{\sigma\wedge N} \subseteq \mathcal{F}_N$ and $(\tau>N) = (\tau\le N)^c \in \mathcal{F}_N$, such that $F_k\cap(\tau>N) \in \mathcal{F}_N$. Then, since X closes the submartingale $(X_n,\mathcal{F}_n)$, we obtain

$$\int_{F_k\cap(\tau>N)} X_N \,\mathrm{d}P \le \int_{F_k\cap(\tau>N)} X \,\mathrm{d}P$$

where the right hand side converges using dominated convergence:

$$\lim_{N\to\infty} \int_{F_k\cap(\tau>N)} X \, \mathrm{d}P = \int_{F_k\cap(\tau=\infty)} X \, \mathrm{d}P = \int_{F_k\cap(\tau=\infty)} X_\tau \, \mathrm{d}P$$

Altogether we have shown

$$\begin{aligned}
\int_{F_k} X_k \, \mathrm{d}P &\le \limsup_{N\to\infty} \Big(\int_{F_k\cap(\tau\le N)} X_\tau \, \mathrm{d}P + \int_{F_k\cap(\tau>N)} X_N \, \mathrm{d}P \Big) \\
&\le \int_{F_k\cap(\tau<\infty)} X_\tau \, \mathrm{d}P + \int_{F_k\cap(\tau=\infty)} X_\tau \, \mathrm{d}P \\
&= \int_{F_k} X_\tau \, \mathrm{d}P
\end{aligned}$$

which was the desired inequality (5.20). □

5.5 The martingale central limit theorem

The principle, that sums of independent random variables almost follow a normal distribution, is sound. But it underestimates the power of central limit theorems: Sums of *dependent* variables are very often approximately normal as well. Many common dependence structures are weak in the sense the terms may be strongly dependent on a few other variables, but almost independent of the major part of the variables. Hence the sum of such variables will have a probabilistic structure resembling the sum of independent variables.

An important theme of the probability theory in the 20th century has been the refinement of this loose reasoning. How should "weak dependence" be understood, and how is it possible to inspect the difference between the sum of interest and the corresponding sum of independent variables? Typically, smaller and rather specific classes of models have been studied, but the general drive has been missing. Huge amounts of papers exists focusing on

1) U-statistics (a type of sums that are very symmetric).
2) Stationary processes.
3) Markov processes.

The big turning point was reached around 1970 when a group of mathematicians, more or less independent of each other, succeeded in stating and proving central limit theorems in the frame of martingales. Almost all later work in the area have been based on the martingale results.

In the following we will consider a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n), P)$, and for notational reasons we will often need a "time 0" σ–algebra $\mathcal{F}_0$. In the lack of any other suggestions, we will use the trivial σ–algebra

$$\mathcal{F}_0 = \{\emptyset, \Omega\}.$$

Definition 5.5.1. *A real stochastic process $(X_n)_{n\geq 1}$ is a martingale difference, relative to the filtration $(\mathcal{F}_n)_{n\geq 1}$, if*

$$\begin{aligned}
&(1) \quad (X_n)_{n\geq 1} \textit{ is adapted to } (\mathcal{F}_n)_{n\geq 1}\,,\\
&(2) \quad E\,|X_n| < \infty \quad \textit{for all } n \geq 1\,,\\
&(3) \quad E(X_n \mid \mathcal{F}_{n-1}) = 0 \quad \textit{a.s. for all} \;\; n \geq 1\,.
\end{aligned}$$

Note that a martingale difference $(X_n)_{n\geq 1}$ satisfies that

$$E(X_n) = E(E(X_n \mid \mathcal{F}_{n-1})) = E(0) = 0 \quad \text{for all } n \geq 1\,.$$

If $(X_n)_{n\geq 1}$ is a martingale difference, then

$$S_n = \sum_{i=1}^{n} X_i \quad \text{for} \;\; n \geq 1$$

is a martingale, relative to the same filtration, and all the variables in this martingale will have mean 0. Conversely, if $(S_n)_{n\geq 1}$ is a martingale with all variables having mean 0, then

$$X_1 = S_1\,, \quad X_n = S_n - S_{n-1} \quad \text{for} \;\; n = 2, 3, \ldots$$

is a martingale difference. Hence a martingale difference somehow represents the same stochastic property as a martingale, just with a point of view that is shifted a little bit.

Example 5.5.2. If the variables $(X_n)_{n\geq 1}$ are independent and all have mean 0, then the sequence form a martingale difference with respect to the natural filtration

$$\mathcal{F}_n = \sigma(X_1, \ldots, X_n)\,,$$

since

$$E(X_n \mid \mathcal{F}_{n-1}) = E(X_n \mid X_1, \ldots, X_{n-1}) \overset{\text{a.s.}}{=} E(X_n) = 0 \quad \text{for all} \;\; n \geq 1\,.$$

The martingale corresponding to this martingale difference is what we normally interpret as a *random walk*. ∘ We will typically be interested in *square-integrable* martingale differences, that is martingale differences $(X_n)_{n\geq 1}$ such that

$$E\,X_n^2 < \infty \quad \text{for all } n \in \mathbb{N}\,.$$

This leads to the introduction of conditional variances defined by

$$V_n = V(X_n \mid \mathcal{F}_{n-1}) = E(X_n^2 \mid \mathcal{F}_{n-1}) \quad \text{a.s. for all} \quad n \geq 1\,.$$

It may also be useful to define the variables

$$W_n = \sum_{m=1}^{n} V_m\,.$$

In the terminology of martingales the process $(W_n)_{n\geq 1}$ is often denoted *the compensator* for the martingale $(S_n)_{n\geq 1}$. It is easily shown that

$$S_n^2 - W_n$$

is a martingale. It should be noted that in the case of a random walk, where the X–variables are independent, then the compensator is non–random, more precisely

$$W_n = \sum_{m=1}^{n} E\,X_m^2\,.$$

We shall study the so–called *martingale difference arrays*, abbreviated MDA's. These are triangular arrays (X_{nm}) of real random variables,

$$\begin{array}{llll} X_{11} & & & \\ X_{21} & X_{22} & & \\ X_{31} & X_{32} & X_{33} & \\ \vdots & \vdots & \vdots & \ddots \end{array}$$

such that each row in the array forms a martingale difference.

To avoid heavy notation we assume that the same fixed filtration $(\mathcal{F}_n)_{n\geq 1}$ is used in all rows. In principle, it had been possible to use an entire triangular array of σ-algebras $(\mathcal{F}_{nm})$, since we will not need anywhere in the arguments that the σ-algebras in different rows a related, but in practice the higher generality will not be useful at all.

Under these notation–dictated conditions, the assumptions for being a MDA will be

1) X_{nm} is $\mathcal{F}_m$–measurable for all $n \in \mathbb{N},\ m = 1, \ldots, n\,,$

2) $E\,|X_{nm}| < \infty$ for all $n \geq 1,\ m = 1, \ldots, n\,,$

3) $E(X_{nm} \mid \mathcal{F}_{m-1}) = 0$ a.s. for all $n \geq 1,\ m = 1, \ldots, n\,.$

Usually we will assume that all the variables in the array have second order moments. Similarly to the notation in Section 3.5 we introduce the cumulated sums within rows, defined by

$$S_{nm} = \sum_{k=1}^{m} X_{nk} \quad \text{for } n \geq 1,\ m = 1, \ldots, n\,.$$

A central limit theorem in this framework will be a result concerning the full row sums S_{nn} converging in distribution towards a normal distribution as $n \to \infty$.

In Section 3.5 we saw that under a condition of independence within rows a central limit theorem is constructed by demanding that the variance of the row sums converges towards a fixed constant and that the terms in the sums are sufficiently small (Lyapounov's conditions or Lindeberg's condition).

When generalizing to martingale difference arrays, it is still important to ensure that the terms are small. But the condition concerning convergence of the variance of the row sums is changed substantially. The new condition will be that the compensators of the rows

$$\sum_{m=1}^{n} E(X_{nm}^2 \mid \mathcal{F}_{m-1})\,, \tag{5.21}$$

(that are random variables) converge in probability towards a non–zero constant. This constant will serve as the variance in the limit normal distribution. Without loss of generality, we shall assume that this constant is 1.

In order to ease notation, we introduce the conditional variances of the variables in the array

$$V_{nm} = E(X_{nm}^2 \mid \mathcal{F}_{m-1}) \quad \text{for} \;\; n \geq 1,\ m = 1, \ldots, n\,,$$

and the corresponding cumulated sums

$$W_{nm} = \sum_{k=1}^{m} V_{nk} \quad \text{for} \;\; n \geq 1,\ m = 1, \ldots, n\,,$$

representing the compensators within rows. Note that the V_{nm}'s are all non–negative (almost surely), and that W_{nm} thereby grows as m increases. Furthermore note, that W_{nm} is $\mathcal{F}_{m-1}$–measurable.

At first, we shall additionally assume that $|W_{nn}| \leq 2$, or equivalently that

$$W_{nm} \leq 2 \qquad \text{a.s. for all} \;\; n \geq 1,\ m = 1, \ldots, n\,.$$

We will use repeatedly that for a bounded sequence of random variables, that converges in probability to constant, the integrals will also converge:

Lemma 5.5.3. *Let (X_n) be a sequence of real random variables such that $|X_n| \leq C$ for all $n \geq 1$ and some constant C. If $X_n \xrightarrow{P} x$, then $EX_n \to x$ as well.*

Proof. Note that $(X_n)_{n\geq 1}$ is uniformly integrable, and thereby it converges in $\mathcal{L}^1$. Hence the convergence of integrals follows. □

Lemma 5.5.4. *Let $(X_{n\,m})$ be a triangular array consisting of real random variables with third order moment. Assume that there exists a filtration $(\mathcal{F}_n)_{n\geq 1}$, making each row in the array a martingale difference. Assume furthermore that*

$$\sum_{m=1}^{n} E(X_{nm}^2 \mid \mathcal{F}_{m-1}) \xrightarrow{P} 1 \quad \textit{as } n \to \infty \ .$$

If

$$\sum_{m=1}^{n} E(X_{nm}^2 \mid \mathcal{F}_{m-1}) \leq 2 \qquad \textit{a.s. for all} \ \ n \geq 1, \tag{5.22}$$

and if the array fulfils Lyapounov's condition

$$\sum_{m=1}^{n} E\,|X_{nm}|^3 \to 0 \quad \textit{as} \ \ n \to \infty\,, \tag{5.23}$$

then the row sums $S_{nn} = \sum_{m=1}^{n} X_{nm}$ will satisfy

$$S_{nn} \xrightarrow{wk} \mathcal{N}(0,1) \quad \textit{as } n \to \infty\,.$$

Note that it is not important which upper bound is used in (5.22) - the number 2 could be replaced by any constant $c > 1$ without changing the proof and without changing the utility of the lemma.

Proof. It will be enough to show the following convergence

$$\int e^{i\,S_{nn}\,t + W_{nn}\,t^2/2}\,\mathrm{d}P \to 1 \quad \text{as} \quad n \to \infty \tag{5.24}$$

for each $t \in \mathbb{R}$. That is seen from the following: Let $\phi_n(t)$ be the characteristic function of S_{nn}, then we have

$$\phi_n(t)\,e^{t^2/2} = \int e^{i\,S_{nn}\,t + t^2/2}\,\mathrm{d}P = \int e^{i\,S_{nn}\,t + W_{nn}\,t^2/2}\,\mathrm{d}P + \int e^{i\,S_{nn} t}(e^{t^2/2} - e^{W_{nn}\,t^2/2})\,\mathrm{d}P\,.$$

Since we have assumed that $W_{nn} \xrightarrow{P} 1$ and the function $x \mapsto \exp(t^2/2) - \exp(xt^2/2)$ is continuous in 1 (with the value 0 in 1), we must have that

$$e^{t^2/2} - e^{W_{nn}t^2/2} \xrightarrow{P} 0\,.$$

Then

$$P\big(\big|e^{iS_{nn}t}\big|\,\big|e^{t^2/2} - e^{W_{nn}t^2/2}\big| > \epsilon\big) = P\big(\big|e^{t^2/2} - e^{W_{nn}t^2/2}\big| > \epsilon\big) \to 0$$

so the integrand in the last integral above converges to 0 in probability. Furthermore recall that W_{nn} is bounded by 2, so the integrand must be bounded by e^{t^2}. Then the integral converges to 0 as $n \to \infty$ because of Lemma 5.5.3. So if (5.24) is shown, we will obtain that $\phi_n(t)e^{t^2/2} \to 1$ as $n \to \infty$ which is equivalent to having obtained

$$\phi_n(t) \to e^{-t^2/2} \qquad \text{as } n \to \infty\,.$$

and according to the Theorem 3.4.20 we have thereby shown that $S_{nn} \xrightarrow{wk} \mathcal{N}(0,1)$.

In order to show (5.24), we define the variables

$$Q_{nm} = e^{iS_{nm}t + W_{nm}t^2/2} \qquad \tilde{Q}_{nm} = e^{iS_{n\,(m-1)}t + W_{nm}t^2/2}$$

and we will be done, if we can show

$$E(Q_{nn} - 1) = \int Q_{nn}\,\mathrm{d}P - 1 \to 0$$

Firstly, we can rewrite (when defining $S_{n0} = 0$ and $Q_{n0} = 1$)

$$Q_{nn} - 1 = \sum_{m=1}^{n} Q_{nm} - Q_{n\,(m-1)}$$

and we observe

$$Q_{nm} = e^{iX_{n\,m}t}\tilde{Q}_{n\,m} \qquad Q_{n\,(m-1)} = e^{-V_{n\,m}t^2/2}\tilde{Q}_{nm}$$

such that

$$Q_{nn} - 1 = \sum_{m=1}^{n} \big(e^{iX_{nm}t} - e^{-V_{nm}t^2/2}\big)\tilde{Q}_{nm}\,.$$

Recall from the definitions, that both $S_{n\,(m-1)}$ and W_{nm} are $\mathcal{F}_{m-1}$–measurable, such that $\tilde{Q}_{nm}$ is $\mathcal{F}_{m-1}$–measurable as well. Then we can write

$$\begin{aligned}
E(Q_{nn} - 1) &= \sum_{m=1}^{n} E\big((e^{iX_{nm}t} - e^{-V_{nm}t^2/2})\tilde{Q}_{nm}\big) \\
&= \sum_{m=1}^{n} E\big(E\big((e^{iX_{nm}t} - e^{-V_{n\,m}t^2/2})\tilde{Q}_{nm}|\mathcal{F}_{m-1}\big) \\
&= \sum_{m=1}^{n} E\big(E\big((e^{iX_{nm}t} - e^{-V_{n\,m}t^2/2})|\mathcal{F}_{m-1}\big)\tilde{Q}_{nm}\big)
\end{aligned}$$

Furthermore recall that $|W_{nm}| \leq |W_{nn}| \leq 2$ a.s., such that

$$|\tilde{Q}_{nm}| = |e^{iS_{n\,(m-1)}t}||e^{W_{nm}t^2/2}| \leq e^{t^2} \qquad \text{a.s.}$$

such that we together with using the triangle inequality obtain

$$\begin{aligned}
|E(Q_{nn}-1)| &= \Big|\sum_{m=1}^{n} E\Big(E\big((e^{iX_{nm}t} - e^{-V_{nm}t^2/2})|\mathcal{F}_{m-1}\big)\tilde{Q}_{nm}\Big)\Big| \\
&\leq \sum_{m=1}^{n} E\Big(\big|E\big((e^{iX_{nm}t} - e^{-V_{nm}t^2/2}|\mathcal{F}_{m-1}\big)\big||\tilde{Q}_{nm}|\Big) \\
&\leq \sum_{m=1}^{n} E\big|E\big(e^{iX_{nm}t} - e^{-V_{nm}t^2/2}|\mathcal{F}_{m-1}\big)\big|e^{t^2}
\end{aligned}$$

We have from Lemma 3.5.2 that

$$e^{iy} = 1 + iy - \frac{y^2}{2} + r_1(y)\,, \qquad |r_1(y)| \leq \frac{|y|^3}{3}\,.$$

for $y \in \mathbb{R}$ and that

$$e^{-y/2} = 1 - \frac{y}{2} + r_2(y)\,, \qquad |r_2(y)| \leq \frac{y^2}{8}$$

for $y \geq 0$. Then

$$\begin{aligned}
&E\big(e^{iX_{nm}t} - e^{-V_{nm}t^2/2}|\mathcal{F}_{m-1}\big) \\
&\quad = E\Big(\big(1 + iX_{nm}t - \frac{X_{nm}^2t^2}{2} + r_1(X_{nm}t)\big) - \big(1 - \frac{V_{nm}t^2}{2} + r_2(V_{nm}t^2)\big)|\mathcal{F}_{m-1}\Big) \\
&\quad = itE(X_{nm}|\mathcal{F}_{m-1}) - \frac{1}{2}t^2E(X_{nm}^2|\mathcal{F}_{m-1}) + E(r_1(X_{nm}t)|\mathcal{F}_{m-1}) + \frac{1}{2}V_{nm}t^2 - r_2(V_{nm}t^2) \\
&\quad = E(r_1(X_{nm}t)|\mathcal{F}_{m-1}) - r_2(V_{nm}t^2)\,.
\end{aligned}$$

And if we apply the upper bounds for the remainder terms, we obtain

$$\begin{aligned}
\big|E\big(e^{iX_{nm}t} - e^{-V_{nm}t^2/2}|\mathcal{F}_{m-1}\big)\big| &= |E(r_1(X_{nm}t)|\mathcal{F}_{m-1}) - r_2(V_{nm}t^2)| \\
&\leq E(|r_1(X_{nm}t)|\,|\mathbb{F}_{m-1}) + |r_2(V_{nm}t^2)| \\
&\leq E(|X_{nm}|^3|\mathcal{F}_{m-1})\frac{|t|^3}{3} + \frac{V_{nm}^2t^4}{8}
\end{aligned}$$

Collecting all the obtained inequalities yields that

$$|E(Q_{nn}-1)| \leq e^{t^2}\Big(\frac{|t|^3}{3}\sum_{m=1}^{n} E|X_{nm}|^3 + \frac{t^4}{8}\sum_{m=1}^{n} EV_{nm}^2\Big)$$

That the first sum above converges to 0, is simply the Lyapounov condition that is assumed to be true in the lemma. Hence the proof will be complete, if we can show that

$$\sum_{m=1}^{n} EV_{nm}^2 \to 0$$

as $n \to \infty$. This is obviously the same as showing that

$$E\sum_{m=1}^{n} V_{nm}^2 \to 0 \quad \text{as } n \to \infty\,. \tag{5.25}$$

The integrand above must have the following upper bound

$$\sum_{m=1}^{n} V_{nm}^2 \le \sum_{m=1}^{n} V_{nm}\big(\max_{k=1,\dots,n} V_{nk}\big) = \big(\max_{k=1,\dots,n} V_{nk}\big)\sum_{m=1}^{n} V_{nm} \tag{5.26}$$

and this must be bounded by 4, since $\sum_{m=1}^{n} V_{nm} \le 2$ such that especially also all $V_{nk} \le 2$. Hence the integrand in (5.25) is bounded, so Lemma 5.5.3 gives that the proof is complete, if we can show

$$\sum_{m=1}^{n} V_{nm}^2 \xrightarrow{P} 0\,,$$

which (because of the inequality (5.26)) will be the case, if we can show

$$\big(\max_{k=1,\dots,n} V_{nk}\big)\sum_{m=1}^{n} V_{nm} \xrightarrow{P} 0\,.$$

Since we have assumed that $\sum_{m=1}^{n} V_{nm} \xrightarrow{P} 1$, it will according to Theorem 3.3.3 be enough to show that

$$\max_{k=1,\dots,n} V_{nk} \xrightarrow{P} 0\,. \tag{5.27}$$

In order to show (5.27) we will utilize the fact that for each $c > 0$ exists a $d > 0$ such that

$$x^2 \le c + d|x|^3 \qquad \text{for all } x \in \mathbb{R}\,.$$

So let $c > 0$ be some arbitrary number and find the corresponding $d > 0$. Then

$$\begin{aligned} V_{nm} &= E(X_{nm}^2|\mathcal{F}_{m-1}) \le E(c + d|X_{nm}|^3|\mathcal{F}_{m-1}) \\ &= c + dE(|X_{nm}|^3|\mathcal{F}_{m-1}) \le c + d\sum_{m=1}^{n} E(|X_{nm}|^3|\mathcal{F}_{m-1})\,, \end{aligned}$$

and since this upper bound does not depend on m, we have the inequality

$$\max_{m=1,\dots,n} V_{nm} \le c + d\sum_{m=1}^{n} E(|X_{nm}|^3|\mathcal{F}_{m-1})$$

Then from integration

$$E\max_{m=1,\dots,n} V_{nm} \le c + d\sum_{m=1}^{n} E\big(E(|X_{nm}|^3|\mathcal{F}_{m-1})\big) = c + d\sum_{m=1}^{n} E|X_{nm}|^3$$

and letting $n \to \infty$ yields

$$\limsup_{n\to\infty} E \max_{m=1,\dots,n} V_{nm} \le \limsup_{n\to\infty} \left(c + d\sum_{m=1}^{n} E|X_{nm}|^3\right) = c + d \lim_{n\to\infty} \sum_{m=1}^{n} E|X_{nm}|^3 = c\,.$$

Since $c > 0$ was chosen arbitrarily, and the left hand side is independent of c, we must have (recall that it is non–negative)

$$\lim_{n\to\infty} E \max_{m=1,\dots,n} V_{nm} = 0$$

And since $E\max_{m=1,\dots,n} V_{nm} = E|\max_{m=1,\dots,n} V_{nm} - 0|$, we actually have that

$$\max_{m=1,\dots,n} V_{nm} \xrightarrow{\mathcal{L}^1} 0$$

which in particular implies (5.27). □

Theorem 5.5.5 (Brown)**.** *Let (X_{nm}) be a triangular array of real random variables with third order moment. Assume that there exists a filtration $(\mathcal{F}_n)_{n\ge1}$ that makes each row in the array a martingale difference. Assume furthermore that*

$$\sum_{m=1}^{n} E(X_{nm}^2 \mid \mathcal{F}_{m-1}) \xrightarrow{P} 1 \quad \text{as } n \to \infty\ . \tag{5.28}$$

If the array fulfils the conditional Lyapounov condition

$$\sum_{m=1}^{n} E\left(|X_{nm}|^3 \mid \mathcal{F}_{m-1}\right) \xrightarrow{P} 0 \quad \text{as} \;\; n \to \infty\,, \tag{5.29}$$

then the row sums $S_{nn} = \sum_{m=1}^{n} X_{nm}$ satisfies that

$$S_{nn} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \quad \text{as } n \to \infty$$

Proof. Most of the work is already done in lemma 5.5.4 - we only need to use a little bit of martingale technology in order to reduce the general setting to the situation in the lemma.

Analogous to the W_{nm}-variables from before we define the cumulated third order moments within each row

$$Z_{nm} = \sum_{k=1}^{m} E\left(|X_{nk}|^3 \mid \mathcal{F}_{k-1}\right).$$

Furthermore define the variable

$$X_{nm}^* = X_{nm}\, 1_{(W_{nm} \le 2,\, Z_{nm} \le 1)}\,.$$

It is not important exactly which upper limit is chosen above for the Z–variables – any strictly positive upper limit would give the same results as 1. The trick will be to see that the triangular array (X_{nm}^*) consisting of the *star* variables fulfils the conditions from Lemma 5.5.4.

Note that since both W_{nm} and Z_{nm} are $\mathcal{F}_{m-1}$–measurable, then the indicator function will be $\mathcal{F}_{m-1}$–measurable. Hence each X_{nm}^* must be $\mathcal{F}_m$–measurable. Furthermore (using that (X_{nm}) is a martingale difference array)

$$E|X_{nm}^*| = E|X_{nm} 1_{(W_{nm} \le 2, Z_{nm} \le 1)}| \le E|X_{nm}| < \infty$$

and also

$$E(X_{nm}^* | \mathcal{F}_{m-1}) = E(X_{nm}|\mathcal{F}_{m-1}) 1_{(W_{nm} \le 2, Z_{nm} \le 1)} = 0\,.$$

Altogether this shows that (X_{nm}^*) is a martingale difference array. We will define the variables V_{nm}^* and W_{nm}^* for (X_{nm}^*) similar to the variables V_{nm} and W_{nm} for (X_{nm}):

$$V_{nm}^* = E((X_{nm}^*)^2 | \mathcal{F}_{m-1})\,, \qquad W_{nm}^* = \sum_{k=1}^{m} V_{nk}^*\,.$$

Then (as before)

$$V_{nm}^* = V_{nm} 1_{(W_{nm} \le 2, Z_{nm} \le 1)}$$

so

$$W_{nm}^* = \sum_{k=1}^{n} V_{nk} 1_{(W_{nk} \le 2, Z_{nk} \le 1)}\,.$$

From this we obtain that $W_{nn}^* \le 2$ (we only add V's to the sum as long as the W's are below 2). We also have that

$$W_{nm}^* = W_{nm} \quad \text{for } m = 1, \ldots, n \text{ on } (W_{nn} \le 2, Z_{nn} \le 1) \tag{5.30}$$

because, since both W_{nk} and Z_{nk} are increasing in k, all the indicator functions $1_{(W_{nk} \le 2, Z_{nk} \le 1)}$ are 1 on the set $(W_{nn} \le 2, Z_{nn} \le 1)$. Since $W_{nn} \xrightarrow{P} 1$ and $Z_{nn} \xrightarrow{P} 0$ it holds that $P(W_{nn} \le 2) = P(|W_{nn} - 1| \le 1) \to 1$ and $P(Z_{nn} \le 1) = P(|Z_{nn} - 0| \le 1) \to 1$. Hence also $P(W_{nn} \le 2, Z_{nn} \le 1) \to 1$. Combining this with (5.30) yields

$$1 \ge P(|W_{nn}^* - W_{nn} - 0| \le \epsilon) \ge P(W_{nn}^* = W_{nn}) \ge P(W_{nn} \le 2, Z_{nn} \le 1) \to 1\,,$$

which shows that $W_{nn}^* - W_{nn} \xrightarrow{P} 0$. Then

$$W_{nn}^* = (W_{nn}^* - W_{nn}) + W_{nn} \xrightarrow{P} 0 + 1 = 1\,.$$

To be able to apply Lemma 5.5.4 to the triangular array we still need to show that the array satisfies the unconditional Lyapounov condition (5.23). For this define

$$Z_{nn}^* = \sum_{k=1}^n E(|X_{nk}^*|^3|\mathcal{F}_{k-1}) = \sum_{k=1}^n E(|X_{nk}|^3|\mathcal{F}_{k-1})1_{(W_{nk}\leq 2, Z_{nk}\leq 1)}$$

It is obvious that $Z_{nn}^* \leq 1$ and from using that all terms in Z_{nm} are non–negative, such that Z_{nm} increases for $m = 1, \dots, n$ we also see (like above) that $Z_{nn}^* \leq Z_{nn}$. The assumption $Z_{nn} \xrightarrow{P} 0$ then implies

$$P(|Z_{nn}^* - 0| > \epsilon) = P(Z_{nn}^* > \epsilon) \leq P(Z_{nn} > \epsilon) \to 0$$

so $Z_{nn}^* \xrightarrow{P} 0$. The fact that all $0, \leq Z_{nn}^* \leq 1$, makes (Z_{nn}^*) uniformly integrable. For $x > 1$:

$$\sup_{n\in\mathbb{N}} \int_{(Z_{nn}^* \geq x)} Z_{nn}^* \,\mathrm{d}P = \int_{\emptyset} Z_{nn}^* \,\mathrm{d}P = 0$$

So Theorem 5.4.5 gives, that $Z_{nn}^* \xrightarrow{\mathcal{L}^1} 0$. Hence $E(Z_{nn}^*) = E|Z_{nn}^* - 0| \to 0$, such that

$$\sum_{k=1}^n E|X_{nk}^*|^3 = \sum_{k=1}^n E\big(E(|X_{nk}^*|^3|\mathcal{F}_{k-1})\big) = E\Big(\sum_{k=1}^n E(|X_{nk}^*|^3|\mathcal{F}_{k-1})\Big) = E(Z_{nn}^*) \to 0$$

Summarising the results so far, we have shown that all conditions from Lemma 5.5.4 are satisfied for the martingale difference array (X_{nm}^*). Then the lemma gives that

$$\sum_{m=1}^n X_{nm}^* \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)\,.$$

We have already argued, that on the set $(W_{nn} \leq 2, Z_{nn} \leq 1)$ all the indicator functions $1_{(W_{nk}\leq 2, Z_{nk}\leq 1)}$ are 1. Hence

$$X_{nm}^* = X_{nm} \quad \text{for all } m = 1, \dots, n \quad \text{on} \quad (W_{nn} \leq 2, Z_{nn} \leq 1)$$

so also

$$\sum_{m=1}^n X_{nm}^* = \sum_{m=1}^n X_{nm} \qquad \text{on} \quad (W_{nn} \leq 2, Z_{nn} \leq 1)\,.$$

Then (using an argument similar to the previous) we obtain

$$1 \geq P\Big(\big|\sum_{m=1}^n X_{nm}^* - \sum_{m=1}^n X_{nm} - 0\big| \leq \epsilon\Big) \geq P(W_{nn} \leq 2, Z_{nn} \leq 1) \to 1$$

so

$$\sum_{m=1}^n X_{nm}^* - \sum_{m=1}^n X_{nm} \xrightarrow{P} 0\,.$$

Referring to Slutsky's lemma completes the proof since then

$$\sum_{m=1}^{n} X_{nm} = \sum_{m=1}^{n} X_{nm}^{*} + \Big(\sum_{m=1}^{n} X_{nm} - \sum_{m=1}^{n} X_{nm}^{*} \Big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

□

By some work it is possible to replace the third order conditions by some Lindeberg–inspired conditions. It is sufficient that all the X–variables have second order moment, satisfy (5.28), and fulfil

$$\sum_{m=1}^{n} E\left(X_{n\,m}^2 1_{(|X_{n\,m}|>c)} \mid \mathcal{F}_{m-1}\right) \xrightarrow{P} 0 \qquad \text{as} \;\; n \to \infty\,, \tag{5.31}$$

for all $c > 0$ in order for the conclusion in Brown's Theorem to be maintained.

5.6 Exercises

All random variables in the following exercises are assumed to be real valued. **Exercise 5.1.** Characterise the *mean function* $n \to E(X_n)$ if $(X_n, \mathcal{F}_n)$ is

(1) a martingale.

(2) a submartingale.

(3) a supermartingale.

Show that a submartingale is a martingale, if and only if the mean function is constant. ○

Exercise 5.2. Let $(\mathcal{F}_n)$ be a filtration on $(\Omega, \mathcal{F})$ and assume that τ and σ are stopping times. Show that $\tau \vee \sigma$ and $\tau + \sigma$ are stopping times. ○

Exercise 5.3. Let $X_1, X_2, \ldots$ be independent and identically distributed real random variables such that $EX_1 = 0$ and $VX_1 = \sigma^2$. Let $(\mathcal{F}_n)$ be the filtration on $(\Omega, \mathcal{F})$ defined by $\mathcal{F}_n = \mathcal{F}(X_1, \ldots, X_n)$. Define

$$Y_n = \left(\sum_{k=1}^{n} X_k \right)^2$$
$$Z_n = Y_n - n\sigma^2$$

Show that $(Y_n, \mathcal{F}_n)$ is a submartingale and that $(Z_n, \mathcal{F}_n)$ is a martingale. ○

Exercise 5.4. Assume that $(X_n, \mathcal{F}_n)$ is an adapted sequence, where each X_n is a real valued random variable. Let $A \in \mathbb{B}$. Define $\tau : \Omega \to \mathbb{N} \cup \{\infty\}$ by

$$\tau(\omega) = \inf\{n \in \mathbb{N} \ : \ X_n(\omega) \in A\} \, .$$

Show that τ is a stopping time. ○

Exercise 5.5. Let $(\mathcal{F}_n)$ be a filtration on $(\Omega, \mathcal{F})$. Let $\tau : \Omega \to \mathbb{N} \cup \{\infty\}$ be a random variable. Show that τ is a stopping time if and only if there exists a sequence of sets (F_n), such that $F_n \in \mathcal{F}_n$ for all $n \in \mathbb{N}$ and

$$\tau(\omega) = \inf\{n \in \mathbb{N} \ : \ \omega \in F_n\}$$

○

Exercise 5.6. Let $(\mathcal{F}_n)$ be a filtration on $(\Omega, \mathcal{F})$ and consider a sequence of sets (F_n) where for each n $F_n \in \mathcal{F}_n$. Let σ be a stopping time and define

$$\tau(\omega) = \inf\{n > \sigma(\omega) \ : \ \omega \in F_n\}$$

Show that τ is a stopping time. ○

Exercise 5.7.

(1) Assume that $(X_n, \mathcal{F}_n)$ is an adapted sequence. Show that $(X_n, \mathcal{F}_n)$ is a martingale if and only if

$$E(X_k | \mathcal{F}_\tau) = X_\tau \text{ a.s.}$$

for all $k \in \mathbb{N}$ and all stopping times τ, where $\tau \leq k$.

(2) Show that if $(X_n, \mathcal{F}_n)$ is a martingale and $\tau \leq m \in \mathbb{N}$, then

$$E(X_\tau) = E(X_1)$$

(3) Show that if $(X_n, \mathcal{F}_n)$ is a submartingale and $\tau \leq m \in \mathbb{N}$, then

$$E(X_1) \leq E(X_\tau) \leq E(X_m)$$

○

Exercise 5.8. Let $(X_1, X_2, \ldots)$ be a sequence of independent and identically distributed random variables, such that $X_1 \sim Pois(\lambda)$. Define for $n \in \mathbb{N}$

$$S_n = \sum_{k=1}^{n} X_k \quad \text{and} \quad \mathcal{F}_n = \mathcal{F}(X_1, \ldots, X_n)\,.$$

(1) Show that $(S_n - n\lambda, \mathcal{F}_n)$ is a martingale.

(2) Define

$$\tau = \inf\{n \in \mathbb{N} \;:\; S_n \geq 1\}$$

Show that τ is a stopping time.

(3) Show that $E(S_{\tau \wedge n}) = \lambda E(\tau \wedge n)$ for all $n \in \mathbb{N}$.

(4) Argue that $P(\tau < \infty) = 1$.

(5) Show that $E(S_\tau) = \lambda E(\tau)$.

∘

Exercise 5.9. Let $X_1, X_2, \ldots$ be a sequence of independent random variables with $EX_n = 0$ for all $n \in \mathbb{N}$. Assume that $\sum_{n=1}^{\infty} EX_n^2 < \infty$ and define $S_n = X_1 + \cdots + X_n$ for all $n \in \mathbb{N}$. Show that $\lim_{n\to\infty} S_n$ exists almost surely. ∘

Exercise 5.10. Assume that $X_1, X_2, \ldots$ are independent and identically distributed random variables with

$$P(X_n = 1) = p \qquad P(X_n = -1) = 1 - p\,,$$

where $0 < p < 1$ with $p \neq \frac{1}{2}$. Define

$$S_n = X_1 + \cdots + X_n$$

and $\mathcal{F}_n = \mathcal{F}(X_1, \ldots, X_n)$ for all $n \geq 1$.

(1) Let $r = \frac{1-p}{p}$ and show that $E(M_n) = 1$ for all $n \geq 1$, where

$$M_n = r^{S_n}\,.$$

and show that $(M_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is a martingale.

(2) Show that $M_\infty = \lim_{n\to\infty} M_n$ exists a.s.

(3) Show that $EM_\infty \leq 1$.

(4) Show that
$$\lim_{n\to\infty} \frac{1}{n} S_n = 2p - 1 \quad \text{a.s.}$$
and conclude that $S_n \to +\infty$ a.s. if $p > \frac{1}{2}$ and $S_n \to -\infty$ a.s. if $p < \frac{1}{2}$.

(5) Let $a, b \in \mathbb{Z}$ with $a < 0 < b$ and define
$$\tau = \inf\{n \in \mathbb{N} \mid S_n = a \text{ or } S_n = b\}$$
Show that τ is a stopping time.

(6) Show that $P(\tau < \infty) = 1$, and realise that $P(S_\tau \in \{a, b\}) = 1$.

(7) Show that $EM_{\tau\wedge n} = 1$ for all $n \geq 1$.

(8) Show that for all $n \in \mathbb{N}$
$$|S_{\tau\wedge n}| \leq |a| \vee b\,.$$
and conclude that the sequence $(M_{\tau\wedge n})_{n\in\mathbb{N}}$ is bounded.

(9) Show that $EM_\tau = 1$.

(10) Show that
$$P(S_\tau = b) = \frac{1 - r^a}{r^b - r^a}$$
$$P(S_\tau = a) = \frac{r^b - 1}{r^b - r^a}$$

∘

Exercise 5.11. The purpose of this exercise is to show that for a random variable X with $EX^2 < \infty$ and a sub σ–algebra $\mathcal{D}$ of $\mathcal{F}$ we have the following version of Jensen's inequality for conditional expectations
$$E(X^2|\mathcal{D}) \geq E(X|\mathcal{D})^2 \quad \text{a.s.}$$

(1) Show that $x^2 - y^2 \geq 2y(x - y)$ for all $x, y \in \mathbb{R}$ and show that for all $n \in \mathbb{N}$ it holds that
$$1_{D_n}\big(X^2 - E(X|\mathcal{D})^2\big) \geq 1_{D_n} 2E(X|\mathcal{D})\big(X - E(X|\mathcal{D})\big)\,,$$
where $D_n = (|E(X|\mathcal{D})| \leq n)$. Show that both the left hand side and the right hand side are integrable.

(2) Show that

$$E\Big(1_{D_n} 2E(X|\mathcal{D})\big(X - E(X|\mathcal{D})\big)|\mathcal{D}\Big) = 0 \text{ a.s.}$$

(3) Show that

$$1_{D_n} E(X^2|\mathcal{D}) \geq 1_{D_n} E(X|\mathcal{D})^2 \quad \text{a.s. for all } n \in \mathbb{N}.$$

and conclude

$$E(X^2|\mathcal{D}) \geq E(X|\mathcal{D})^2 \quad \text{a.s.}$$

∘

Exercise 5.12.(The Chebychev–Kolmogorov inequality) Let $(X_n, \mathcal{F}_n)$ be a martingale where $EX_n^2 < \infty$ for all $n \in \mathbb{N}$.

(1) Show that $(X_n^2, \mathcal{F}_n)$ is a submartingale.

(2) Define for some $\epsilon > 0$

$$\tau = \inf\{n \in \mathbb{N} \ : \ |X_n| \geq \epsilon\}$$

Show that τ is a stopping time.

(3) Show that for some $n \in \mathbb{N}$ it holds that

$$EX_{\tau\wedge n}^2 \leq EX_n^2$$

(4) Show that

$$EX_{\tau\wedge n}^2 \geq \epsilon^2 P(\max_{k=1,\dots,n} |X_k| \geq \epsilon)$$

(5) Conclude the Chebychev–Kolmogorov Inequality:

$$P(\max_{k=1,\dots,n} |X_k| \geq \epsilon) \leq \frac{EX_n^2}{\epsilon^2}$$

∘

Exercise 5.13.(Doob's Inequality) Assume that $(Y_n, \mathcal{F}_n)_{n\in\mathbb{N}}$ is a submartingale. Let $t > 0$ be a given constant.

(1) Define $\tau : \Omega \to \mathbb{N} \cup \{\infty\}$ by

$$\tau = \inf\{k \in \mathbb{N} \ : \ Y_k \geq t\}$$

Show that τ is a stopping time.

(2) Let $n \in \mathbb{N}$ and define

$$A_n = (\max_{k=1,\dots,n} Y_k \geq t)$$

Use the definition of A_n and τ to show

$$tP(A_n) \leq \int_{A_n} Y_{\tau \wedge n} \,\mathrm{d}P$$

(3) Show "Doob's Inequality":

$$tP(A_n) \leq \int_{A_n} Y_n \,\mathrm{d}P$$

∘

Exercise 5.14.

(1) Assume that $X_1, X_2, \dots$ are random variables with each $X_n \geq 0$ and $E|X_n| < \infty$, such that $X_n \to 0$ a.s. and $EX_n \to 0$. Show that (X_n) is uniformly integrable.

(2) Find a sequence $X_1, X_2, \dots$ of random variables on $(\Omega, \mathcal{F}, P) = ([0,1], \mathcal{B}, \lambda)$ such that

$$X_n \to 0 \quad \text{a.s.} \qquad \text{and} \qquad EX_n \to 0$$

but where (X_n) is not uniformly integrable.

∘

Exercise 5.15. Let X be a random variable with $E|X| < \infty$. Let $\mathbb{G}$ be collection of sub σ–algebras of $\mathcal{F}$. In this exercise we shall show that the following family of random variables

$$(E(X|\mathcal{D}))_{\mathcal{D} \in \mathbb{G}}$$

is uniformly integrable.

(1) Let $\mathcal{D} \in \mathbb{G}$. Show that for all $x > 0$ it holds that

$$\int_{(|E(X|\mathcal{D})|>x)} |E(X|\mathcal{D})| \,\mathrm{d}P \leq \int_{(E(|X||\mathcal{D})>x)} |X| \,\mathrm{d}P$$

(2) Show that for all $K \in \mathbb{N}$ and $x > 0$

$$\int_{(E(|X||\mathcal{D})>x)} |X| \,\mathrm{d}P \leq \int_{(|X|>K)} |X| \,\mathrm{d}P + K\frac{E|X|}{x}$$

(3) Show that $(E(X|\mathcal{D}))_{\mathbb{D}\in\mathbb{G}}$ is uniformly integrable.

$\circ$

Exercise 5.16. Assume that $(\Omega, \mathcal{F}, (\mathcal{F}_n), P)$ is a filtered probability space. Let τ be a stopping time with $E\tau < \infty$. Assume that $(X_n, \mathcal{F}_n)$ is a martingale.

(1) Argue that X_τ is almost surely well–defined and that $X_{\tau\wedge n} \to X_\tau$ a.s.

(2) Assume that $(X_{\tau\wedge n})_{n\in\mathbb{N}}$ is uniformly integrable. Show that $E|X_\tau| < \infty$ and $X_{\tau\wedge n} \xrightarrow{L^1} X_\tau$.

(3) Assume that $(X_{\tau\wedge n})_{n\in\mathbb{N}}$ is uniformly integrable. Show that $EX_\tau = EX_1$.

(4) Assume that a random variable Y exists such that $E|Y| < \infty$ and $|X_{\tau\wedge n}| \leq |Y|$ a.s. for all $n \in \mathbb{N}$. Show that $(X_{\tau\wedge n})$ is uniformly integrable.

In the rest of the exercise you can use without proof that

$$|X_{\tau\wedge n}| \leq |X_1| + \sum_{m=1}^{\infty} 1_{(\tau>m)}|X_{m+1} - X_m| \tag{5.32}$$

for all $n \in \mathbb{N}$.

(5) Show that

$$E\tau = \sum_{n=0}^{\infty} P(\tau > n)$$

(6) Assume that there exists a constant $B > 0$ such that

$$E(|X_{n+1} - X_n|\big|\mathcal{F}_n) \leq B \quad \text{a.s.}$$

for all $n \in \mathbb{N}$. Show that $EX_\tau = EX_1$

Let $Y_1, Y_2, \ldots$ be a sequence of independent and identically distributed random variables satisfying $E|Y_1| < \infty$. Define $\mathcal{G}_n = \mathcal{F}(Y_1, \ldots, Y_n)$ and

$$S_n = \sum_{k=1}^{n} Y_k \qquad Z_n = S_n - n\xi$$

where $\xi = EY_1$.

(7) Show that $(Z_n, \mathcal{F}_n)$ is a martingale.

Let σ be a stopping time (with respect to the filtration $(\mathcal{G}_n)$) such that $E\sigma < \infty$.

(8) Show that $E(|Z_{n+1} - Z_n| \big| \mathcal{G}_n) = E(|Y_1 - \xi|)$ a.s for all $n \in \mathbb{N}$.

(9) Show that $ES_\sigma = E\sigma EY_1$.

◦

Exercise 5.17. Assume that $X_1, X_2, \ldots$ are independent random variables such that for each n it holds $X_n \geq 0$ and $EX_n = 1$. Define $\mathcal{F}_n = \mathcal{F}(X_1, \ldots, X_n)$ and

$$Y_n = \prod_{k=1}^{n} X_k$$

(1) Show that $(Y_n, \mathcal{F}_n)$ is a martingale.

(2) Show that $Y = \lim_{n\to\infty} Y_n$ exists almost surely with $E|Y| < \infty$.

(3) Show that $0 \leq EY \leq 1$.

Assume furthermore that all X_n's are identically distributed satisfying

$$P(X_n = \frac{1}{2}) = P(X_n = \frac{3}{2}) = \frac{1}{2}\,.$$

(4) Show that $Y = 0$ a.s.

(5) Conclude that there does not exist a random variable Z such that $Y_n \xrightarrow{\mathcal{L}^1} Z$.

◦

Exercise 5.18. Let $X_1, X_2, \ldots$ be a sequence of real random variables with $E|X_n| < \infty$ for all $n \in \mathbb{N}$. Assume that X is another random variable with $E|X| < \infty$. The goal of this exercise is to show

$$X_n \xrightarrow{\mathcal{L}^1} X \quad \text{if and only if} \quad E|X_n| \to E|X| \text{ and } X_n \xrightarrow{P} X$$

(1) Assume that $X_n \xrightarrow{\mathcal{L}^1} X$. Show that $E|X_n| \to E|X|$ and $X_n \xrightarrow{P} X$.

Let $U_1, U_2, \ldots$ and $V, V_1, V_2, \ldots$ be two sequences of random variables such that $E|V| < \infty$ and for all $n \in \mathbb{N}$

$$E|V_n| < \infty$$
$$|U_n| \leq V_n$$
$$V_n \xrightarrow{\text{a.s.}} V \text{ as } n \to \infty$$
$$EV_n \to EV \text{ as } n \to \infty$$

(2) Apply Fatou's lemma on the sequence $(V_n - |U_n|)$ to show that

$$\limsup_{n\to\infty} E|U_n| \leq E \limsup_{n\to\infty} |U_n|$$

Hint: You can use that if (a_n) is a real sequence, then

$$\liminf_{n\to\infty}(-a_n) = -\limsup_{n\to\infty} a_n$$

and if (b_n) is another real sequence with $b_n \to b$, then

$$\liminf_{n\to\infty} b_n = b$$

and

$$\liminf_{n\to\infty}(a_n + b_n) = (\liminf_{n\to\infty} a_n) + b$$
$$\limsup_{n\to\infty}(a_n + b_n) = (\limsup_{n\to\infty} a_n) + b$$

(3) Use (2) to show that if $E|X_n| \to E|X|$ and $X_n \cap X$ then $X_n \xrightarrow{\mathcal{L}^1} X$.

(4) Assume that $E|X_n| \to E|X|$ and $X_n \xrightarrow{P} X$. Show that $X_n \xrightarrow{\mathcal{L}^1} X$.

Now let $(Y_n, \mathcal{F}_n)$ be a martingale. Assume that a random variable Y exists with $E|Y| < \infty$, such that $Y_n \xrightarrow{P} Y$.

(5) Assume that $E|Y_n| \to E|Y|$. Show that $Y_n \xrightarrow{\text{a.s.}} Y$.

(6) Show that Y closes the martingale if and only if $E|Y_n| \to E|Y|$.

◦

Exercise 5.19. Consider the gambling strategy discussed in Section 5.1: Let $Y_1, Y_2, \dots$ be independent and identically distributed random variables with

$$P(Y_1 = 1) = p \qquad P(Y_1 = -1) = 1 - p\,,$$

where $0 < p < \frac{1}{2}$. We think of Y_n as the the result of a game, where the probability of winning is p, and where if you bet 1 dollar, you will receive 1 dollar if you win, and lose the 1 dollar, if you lose the game. We consider the sequence of strategies where the bet is doubled for each lost game, and when a game finally is won, the bet is reset to 1. That is defining the sequence of strategies (ϕ_n) such that

$$\phi_1 = 1$$

and furthermore recursively for $n \geq 2$

$$\phi_n(y_1, \dots, y_{n-1}) = \begin{cases} 2\phi_{n-1}(y_1, \dots, y_{n-2}) & \text{if } y_{n-1} = -1 \\ 1 & \text{if } y_{n-1} = 1 \end{cases}$$

Then the winnings in the n'th game is

$$Y_n \phi_n(Y_1, \dots, Y_{n-1})$$

and the total winnings in game $1, \dots, n$ is

$$X_n = \sum_{k=1}^{n} Y_k \phi_k(Y_1, \dots, Y_{k-1})$$

If e.g. we lose the first three games and win the fourth, then

$$X_1 = -1, \quad X_2 = -1 - 2, \quad X_3 = -1 - 2 - 2^2, \quad X_4 = -1 - 2 - 2^2 + 2^3 = 1$$

Define for each $n \in \mathbb{N}$ the σ–algebra $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$.

(1) Show that $(X_n, \mathcal{F}_n)$ is a *true* supermartingale (meaning a supermartingale that is not a martingale).

Define the sequence (τ_k) by

$$\tau_1 = \inf\{n \in \mathbb{N} \mid Y_n = 1\}$$
$$\tau_{k+1} = \inf\{n > \tau_k \mid Y_n = 1\}$$

(2) Show that (τ_k) is a sequence of sampling times.

(3) Realise that $X_{\tau_k} = k$ for all $k \in \mathbb{N}$ and conclude that $(X_{\tau_k}, \mathcal{F}_{\tau_k})$ is a true submartingale.

Hence we have stopped a true supermartingale and obtained a true submartingale!! In the next questions we shall compare that result to Theorem 5.4.9.

(4) See that on the set $(\tau_1 > n)$ we must have $X_n = -\sum_{k=1}^n 2^{k-1} = 1 - 2^n$ and show that

$$\int_{(\tau_1 > n)} X_n^- \, \mathrm{d}P = q^n - (2q)^n \to -\infty \text{ as } n \to \infty ,$$

where $q = 1 - p$.

(5) Compare the result from 4 with assumption 2 in Theorem 5.4.9.

Now assume that we change the strategy sequence (ϕ_n) in such a way, that we limit our betting in order to avoid $X_n < -7$. Hence we always have $X_n \geq -7$. Since all bettings are non–negative we still have, that $(X_n, \mathbb{F}_n)$ is a supermartingale.

(6) Let (σ_k) be an increasing sequence of stopping times. Show that $(X_{\sigma_k}, \mathbb{F}_{\sigma_k})$ is a supermartingale.

∘

Exercise 5.20. The purpose of this exercise is to show the following theorem:

Let (X_n) be a martingale and assume that for some $p > 1$ it holds that

$$\sup_{n \geq 1} E|X_n|^p < \infty .$$

Then a random variable X exists with $E|X|^p < \infty$ such that

$$X_n \xrightarrow{\text{a.s.}} X , \qquad X_n \xrightarrow{L^p} X$$

(1) Assume that $\sup_{n \geq 1} E|X_n|^p < \infty$. Show that there exists X such that $X_n \xrightarrow{\text{a.s.}} X$ and $E|X| < \infty$.

(2) Assume that both $\sup_{n\geq 1} E|X_n|^p < \infty$ and $E\big[\sup_n |X_n|^p\big] < \infty$. Show that $E|X|^p < \infty$ (with X from 1)) and $X_n \xrightarrow{L^p} X$.

In the rest of the exercise we shall show that $E\big[\sup_n |X_n|^p\big] < \infty$ under the assumption that $\sup_{n\geq 1} E|X_n|^p < \infty$.

(3) Assume that $Z \geq 0$ and let $r > 0$. Show

$$EZ^r = \int_0^\infty r t^{r-1} P(Z \geq t)\,\mathrm{d}t$$

Define $M_n = \max_{1\leq k\leq n} |X_k|$.

(4) Show that

$$EM_n^p \leq \int_0^\infty p t^{p-2} \int_{(M_n \geq t)} |X_n|\,\mathrm{d}P\,\mathrm{d}t$$

(5) Show that

$$EM_n^p \leq \frac{p}{p-1} E\big(M_n^{p-1}|X_n|\big)$$

Recall that Hölder's Inequality gives: If $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$ and furthermore Y, Z are random variables with $E|Y|^p < \infty$ and $E|Z|^q < \infty$, then

$$E|YZ| \leq \big(E|Y|^p\big)^{1/p} \big(E|Z|^q\big)^{1/q}.$$

(6) Show that

$$EM_n^p \leq \Big(\frac{p}{p-1}\Big)^p E(|X_n|^p)$$

(7) Conclude that $E\big[\sup_n |X_n|^p\big] < \infty$ under the assumption $\sup_{n\in\mathbb{N}} E|X_n|^p < \infty$.

∘

Exercise 5.21. (Continuation of Exercise 5.10)
Assume that $X_1, X_2, \dots$ are independent and identically distributed random variables with

$$P(X_n = 1) = P(X_n = -1) = \frac{1}{2}.$$

Define

$$S_n = X_1 + \cdots + X_n$$

and $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ for all $n \in \mathbb{N}$.

(1) Show that $(S_n, \mathcal{F}_n)$ and $(S_n^2 - n, \mathcal{F}_n)$ are martingales.

Let $a, b \in \mathbb{Z}$ with $a < 0 < b$ and define

$$\tau_{a,b} = \inf\{n \in \mathbb{N} \ : \ S_n = a \text{ or } S_n = b\}$$

It was seen in Exercise 5.10, that $\tau_{a,b}$ is a stopping time.

(2) Show for all $n, m \in \mathbb{N}$ that

$$P(\tau_{a,b} > nm) \leq \prod_{k=1}^{n} P(|S_{km} - S_{(k-1)m}| < b - a) = P(|S_m| < b - a)^n$$

Defining $S_0 = 0$.

(3) Show that $P(\tau_{a,b} < \infty) = 1$ and conclude that $P(S_{\tau_{a,b}} \in \{a, b\}) = 1$.

(4) Show that $ES_{\tau_{a,b} \wedge n} = 0$ for all $n \in \mathbb{N}$ and conclude that $ES_{\tau_{a,b}} = 0$.

(5) Show that

$$P(S_{\tau_{a,b}} = a) = \frac{b}{b-a} \qquad P(S_{\tau_{a,b}} = b) = \frac{-a}{b-a}$$

(6) Show that $ES^2_{\tau_{a,b}} = E\tau_{a,b}$ and conclude that $E\tau_{a,b} = -ab$.

Define the stopping time

$$\tau_b = \inf\{n \in \mathbb{N} \ : \ S_n = b\}$$

(7) Show that $P(F) = 1$, where $F = \bigcap_{n=1}^{\infty}(\tau_{-n,b} < \infty)$.

(8) Show $P((\tau_{-n,b} = \tau_b) \cap F) \to 1$ as $n \to \infty$. Conclude that $P(G) = 1$, where

$$G = \left(\bigcup_{n=1}^{\infty}(\tau_{-n,b} = \tau_b)\right) \cap F$$

(9) Show that $P(\tau_b < \infty) = 1$.

(10) Show that $E\tau_b = \infty$.

(11) Argue that

$$\liminf_{N\to\infty} \int_{(\tau_b > N)} |S_N| \, \mathrm{d}P \neq 0$$

(12) Show that

$$P(\sup_{n\geq 1} S_n = \infty) = 1$$

From symmetry it is seen that also

$$P(\inf_{n\geq 1} S_n = -\infty) = 1$$

∘

Exercise 5.22. Let $(\Omega, \mathcal{F}, \mathcal{F}_n, P)$ be a filtered probability space, and let $(Y_n)_{n\geq 1}$ and $(Z_n)_{n\geq 1}$ be two adapted sequences of real random variables. Define furthermore $Z_0 \equiv 1$. Assume that $Y_1, Y_2, \ldots$ are independent and identically distributed with $E|Y_1|^3 < \infty$ and $EY_1 = 0$. Assume furthermore that for all $n \geq 2$ it holds that Y_n is independent of $\mathcal{F}_{n-1}$. Finally, assume that $E|Z_n|^3 < \infty$ for all $n \in \mathbb{N}$. Define for all $n \in \mathbb{N}$

$$M_n = \sum_{m=1}^{n} Z_{m-1} Y_m$$

(1) Show that $(M_n, \mathcal{F}_n)$ is a martingale.

(2) Assume that

$$\frac{1}{n}\sum_{m=0}^{n-1} Z_m^2 \xrightarrow{P} \alpha^2 > 0$$

and

$$\frac{1}{n^{3/2}}\sum_{m=0}^{n-1} |Z_m|^3 \xrightarrow{\text{a.s.}} 0\,.$$

Show that

$$\frac{1}{\sqrt{n}} M_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \alpha^2\sigma^2)\,,$$

where $\sigma^2 = EY_1^2$.

Define $N_1 \equiv Y_1$ and for $n \geq 2$

$$N_n = Y_1 + \sum_{m=2}^{n} \frac{1}{m} Y_{m-1} Y_m$$

(3) Argue that $(N_n, \mathcal{F}_n)$ is a martingale.

(4) Show that for all $n \geq 2$

$$EN_n^2 = EN_{n-1}^2 + \frac{1}{n^2}(\sigma^2)^2$$

(5) Show that the sequence (N_n) is uniformly integrable.

(6) Show that $N_\infty = \lim_{n\to\infty} N_n$ exists almost surely and in $\mathcal{L}^1$. Find EN_∞.

(7) Show that for $1 \leq i < j$ it holds that

$$EY_{i-1}Y_iY_{j-1}Y_j = 0$$

and use this to conclude that for $k, n \in \mathbb{N}$

$$E(N_{n+k} - N_n)^2 = (\sigma^2)\sum_{m=n+1}^{n+k}\frac{1}{m^2}$$

(8) Show that $N_n \to N_\infty$ in L^2.

Define for all $n \in \mathbb{N}$

$$M_n^* = \sum_{m=2}^{n} Y_{m-1}Y_m$$

with the definition $Y_0 \equiv 1$. In the following questions you can use Kronecker's Lemma that is a mathematical result: If (x_n) is a sequence of real numbers such that $\lim_{n\to\infty}\sum_{k=1}^n x_k = s$ exists, and if $0 < b_1 \leq b_2 \leq \cdots$ with $b_n \to \infty$, then

$$\lim_{n\to\infty}\frac{1}{b_n}\sum_{k=1}^{n} b_k x_k = 0\,.$$

(9) Show that

$$\lim_{n\to\infty}\frac{1}{n}M_n^* = 0 \quad \text{a.s.}$$

(10) Use the strong law of large numbers to show

$$\frac{1}{n}\sum_{k=1}^{n} Y_k^2 \xrightarrow{\text{a.s.}} \sigma^2 \quad \text{a.s.}$$

and

$$\frac{1}{n^{3/2}}\sum_{k=1}^{n} |Y_k|^3 \xrightarrow{\text{a.s.}} 0 \quad \text{a.s.}$$

(11) Show that

$$\frac{1}{\sqrt{n}} M_n^* \xrightarrow{\mathcal{D}} \mathcal{N}(0, (\sigma^2)^2)$$

○

Chapter 6

The Brownian motion

The first attempt to define the stochastic process which is now known as the Brownian motion was made by the Frenchman Bachelier, who at the end of the 19th century tried to give a statistical description of the random price fluctuations on the stock exchange in Paris. Some years later, a variation of the Brownian motion is mentioned in Einstein's theory of relativity, but the first precise mathematical definition is due to Norbert Wiener (1923) (which explains the name one occasionally sees: the Wiener process). The Frenchman Paul Lévy explored and discovered some of the fundamental properties of Brownian motion and since that time thousands of research papers have been written concerning what is unquestionably the most important of all stochastic processes.

Brown himself has only contributed his name to the theory of the process: he was a botanist and in 1828 observed the seemingly random motion of flower pollen suspended in water, where the pollen grains constantly changed direction, a phenomenon he explained as being caused by the collision of the microscopic pollen grains with water molecules.

So far the largest collection of random variables under study have been sequences indexed by $\mathbb{N}$. In this chapter we study stochastic processes indexed by $[0, \infty)$. In Section 6.1 we discuss how to define such processes indexed by $[0, \infty)$, and we furthermore define and show the existence of the important Brownian motion. In following sections we study the behaviour of the so–called sample paths of the Brownian motion. In Section 6.2 we prove that there exists a continuous version, and in the remaining sections we study how well–behaved the sample paths are – apart from being continuous.

6.1 Definition and existence

We begin with a brief presentation of some definitions and results from the general theory of stochastic processes.

Definition 6.1.1. *A stochastic process in continuous time is a family $X = (X_t)_{t\geq 0}$ of real random variables, defined on a probability space $(\Omega, \mathcal{F}, P)$.*

In Section 2.3 we regarded a sequence $(X_n)_{n\geq 1}$ of real random variables as a random variable with values in $\mathbb{R}^\infty$ equipped with the σ–algebra $\mathcal{B}_\infty$. Similarly we will regard a stochastic process X in continuous time as having values in the space $\mathbb{R}^{[0,\infty)}$ consisting of all functions $x : [0,\infty) \to \mathbb{R}$. The next step is to equip $\mathbb{R}^{[0,\infty)}$ with a σ–algebra. For this, define the coordinate projections $\hat{X}_t$ by

$$\hat{X}_t(x) = x_t \quad \text{for } x \in \mathbb{R}^{[0,\infty)}$$

for all $t \geq 0$. Then we define

Definition 6.1.2. *Let $\mathcal{B}_{[0,\infty)}$ denote the smallest σ–algebra that makes $\hat{X}_t$ ($\mathcal{B}_{[0,\infty)} - \mathcal{B}$) measurable for all $t \geq 0$.*

Then we have

Lemma 6.1.3. *Let $X : \Omega \to \mathbb{R}^{[0,\infty)}$. Then X is is $\mathcal{F} - \mathcal{B}_{[0,\infty)}$ measurable, if and only if $\hat{X}_t \circ X$ is $\mathcal{F} - \mathcal{B}$ measurable for all $t \geq 0$.*

Proof. The proof will be identical to the proof of Lemma 2.3.3: If X is $\mathcal{F} - \mathcal{B}_{[0,\infty)}$ measurable we can use that $\hat{X}_t$ by definition is $\mathcal{B}_{[0,\infty)} - \mathcal{B}$ measurable, so the composition is $\mathcal{F} - \mathcal{B}$ measurable. Conversely, assume that $\hat{X}_t \circ X$ is $\mathcal{F} - \mathcal{B}$ measurable for all $t \geq 0$. To show that X is $\mathcal{F} - \mathcal{B}_{[0,\infty)}$ measurable, it suffices to show that $X^{-1}(A) \in \mathcal{F}$ for all A in the generating system $\mathbb{H} = \{\hat{X}_t^{-1}(B) \mid t \geq 0, B \in \mathcal{B}\}$ for $\mathcal{B}_{[0,\infty)}$. But for any $t \geq 0$ and $B \in \mathcal{B}$ we have $X^{-1}(\hat{X}_t^{-1}(B)) = (\hat{X}_t \circ X)^{-1}(B) \in \mathcal{F}$ by our assumptions. □

Lemma 6.1.4. *Let $X = (X_t)_{t\geq 0}$ be a stochastic process. Then X is $\mathcal{F} - \mathcal{B}_{[0,\infty)}$ measurable.*

Proof. Note that $\hat{X}_t \circ X = X_t$ and X_t is $\mathcal{F} - \mathcal{B}$ measurable by assumption. The result follows from Lemma 6.1.3. □

If $X = (X_t)_{t\geq 0}$ is a stochastic process, we can consider the distribution $X(P)$ on $(\mathbb{R}^{[0,\infty)}, \mathcal{B}_{[0,\infty)}$. For determining such a distribution, the following lemma will be useful.

Lemma 6.1.5. *Define* $\mathbb{H}$ *as the family of sets on the form*

$$\{x \in \mathbb{R}^{[0,\infty)} \mid (x_{t_1}, \ldots, x_{t_n}) \in B_n\} = \big((\hat{X}_{t_1}, \ldots, \hat{X}_{t_n}) \in B_n\big),$$

where $n \in \mathbb{N}$*,* $0 \leq t_1 < \cdots < t_n$ *and* $B_n \in \mathcal{B}_n$*. Then* $\mathbb{H}$ *is a generating family for* $\mathcal{B}_{[0,\infty)}$ *which is stable under finite intersections.*

Proof. It is immediate (but notationally heavy) to see that $\mathbb{H}$ is stable under finite intersections. Let $F = \big((\hat{X}_{t_1}, \ldots, \hat{X}_{t_n}) \in B_n\big) \in \mathbb{H}$ and note that the vector $(\hat{X}_{t_1}, \ldots, \hat{X}_{t_n})$ is $\mathcal{B}_{[0,\infty)} - \mathcal{B}_n$ measurable, so $F \in \mathcal{B}_{[0,\infty)}$. Therefore $\mathbb{H} \subseteq \mathcal{B}_{[0,\infty)}$, so also $\sigma(\mathbb{H}) \subseteq \mathcal{B}_{[0,\infty)}$. For the converse inclusion, note that for all $t \geq 0$ and $B \in \mathcal{B}$ it holds that $\hat{X}_t^{-1}(B) = (\hat{X}_t \in B) \in \mathbb{H}$, so each coordinate projection must be $\sigma(\mathbb{H}) - \mathcal{B}$ measurable. As $\mathcal{B}_{[0,\infty)}$ is the smallest $\sigma-$algebra with this property, we conclude that $\mathcal{B}_{[0,\infty)} \subseteq \sigma(\mathbb{H})$. All together we have the desired result $\mathcal{B}_{[0,\infty)} = \sigma(\mathbb{H})$. □

If $\hat{P}$ is a probability on $(\mathbb{R}^{[0,\infty)}, \mathcal{B}_{[0,\infty)})$ then

$$P^{(n)}_{t_1 \ldots t_n}(B_n) = \hat{P}((\hat{X}_{t_1}, \ldots, \hat{X}_{t_n}) \in B_n) \tag{6.1}$$

defines a probability on $(\mathbb{R}^n, \mathcal{B}_n)$ for all $n \in \mathbb{N}$, $t_1 < \cdots < t_n$. The class of all $P^{(n)}_{t_1 \ldots t_n}$ is the class of *finite-dimensional distributions* for $\hat{P}$.

If X is a real stochastic process with distribution $\hat{P}$ then $P^{(n)}_{t_1 \ldots t_n}$ given by (6.1) is the distribution of $(X_{t_1}, \ldots, X_{t_n})$ and the class $(P^{(n)}_{t_1 \ldots t_n})$ is called the class or family of *finite-dimensional distributions for* X.

From Lemma 6.1.5 and Theorem A.2.4, it follows that a probability $\hat{P}$ on $(\mathbb{R}^{[0,\infty)}, \mathcal{B}_{[0,\infty)})$ is uniquely determined by the finite-dimensional distributions. The main result concerning the construction of stochastic processes, Kolmogorov's consistency theorem, gives a simple condition for when a given class of finite-dimensional distributions is the class of finite-dimensional distributions for one (and necessarily only one) probability on $(\mathbb{R}^{[0,\infty)}, \mathcal{B}^{[0,\infty)})$.

With $\hat{P}$ a probability on $(\mathbb{R}^{[0,\infty)}, \mathcal{B}_{[0,\infty)})$, it is clear that the finite-dimensional distributions for $\hat{P}$ fit together in the following sense: the $\hat{P}$-distribution of $(\hat{X}_{t_1}, \ldots, \hat{X}_{t_n})$ can be obtained as the marginal distribution of $(\hat{X}_{u_1}, \ldots, \hat{X}_{u_m})$ for any choice of m and $0 \leq u_1 < \ldots < u_m$ such that $\{t_1, \ldots, t_n\} \subseteq \{u_1, \ldots, u_m\}$. In particular, a class of finite-dimensional distributions must always fulfil the following consistency condition: for all $n \in \mathbb{N}$, $0 \le t_1 < \ldots < t_{n+1}$ and all k ,$1 \le k \le n+1$, we have

$$P^{(n)}_{t_1 \ldots t_{k-1} t_{k+1} \ldots t_{n+1}} = \pi_k(P^{(n+1)}_{t_1 \ldots t_{n+1}}), \tag{6.2}$$

where $\pi_k : \mathbb{R}^{n+1} \to \mathbb{R}^n$ is given by

$$\pi_k(y_1, \ldots, y_{n+1}) = (y_1, \ldots, y_{k-1}, y_{k+1}, \ldots, y_{n+1}).$$

If $X = (X_t)_{t \ge 0}$ has distribution $\hat{P}$ with finite-dimensional distributions $(P^{(n)}_{t_1 \ldots t_n})$, then (6.2) merely states that the distribution of $(X_{t_1}, \ldots, X_{t_{k-1}}, X_{t_{k+1}}, \ldots, X_{t_{n+1}})$, is the marginal distribution in the distribution of $(X_{t_1}, \ldots, X_{t_{n+1}})$ which is obtained by excluding X_{t_k}.

We will without proof use

Theorem 6.1.6 (Kolmogorov's consistency theorem)**.** *If $\mathcal{P} = (P^{(n)}_{t_1 \ldots t_n})$ is a family of finite-dimensional distributions, defined for $n \in \mathbb{N}$, $0 \le t_1 < \cdots < t_n$, which satisfies the consistency condition (6.2), then there exists exactly one probability $\hat{P}$ on $(\mathbb{R}^{[0,\infty)}, \mathcal{B}_{[0,\infty)})$ which has $\mathcal{P}$ as its family of finite-dimensional distributions.*

We shall use the consistency theorem to prove the existence of a Brownian motion, that is defined by

Definition 6.1.7. *A real stochastic process $X = (X_t)_{t \ge 0}$ defined on a probability space $(\Omega, \mathbb{F}, P)$ is a Brownian motion with drift $\xi \in \mathbb{R}$ and variance $\sigma^2 > 0$, if the following three conditions are satisfied*

(1) $P(X_0 = 0) = 1$.

(2) For all $0 \le s < t$ the increments $X_t - X_s$ are normally distributed $\mathcal{N}((t-s)\xi, (t-s)\sigma^2))$.

(3) The increments $X_{t_1} = X_{t_1} - X_0, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are for all $n \in \mathbb{N}$ and $0 \le t_1 < \cdots < t_n$ mutually independent.

Definition 6.1.8. *A normalised Brownian motion is a Brownian motion with drift $\xi = 0$ and variance $\sigma^2 = 1$.*

Theorem 6.1.9. *For any $\xi \in \mathbb{R}$ and $\sigma^2 > 0$ there exists a Brownian motion with drift ξ and variance σ^2.*

Proof. We shall use Kolmogorov's consistency theorem. The finite dimensional distributions for the Brownian motion are determined by (1)–(3):
Let $0 \leq t_1 < \cdots < t_{n+1}$. Then we know that

$$X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_{n+1}} - X_{t_n}$$

are independent and normally distributed. Then the vector

$$(X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_{n+1}} - X_{t_n}) \tag{6.3}$$

is $n+1$–dimensional normally distributed. Since

$$(X_{t_1}, X_{t_2}, \ldots, X_{t_{n+1}}) \tag{6.4}$$

is a linear transformation of (6.3), then (6.4) is $n+1$-dimensional normally distributed as well. The distribution of such a normal vector is determined by the mean vector and covariance matrix. We have

$$E(X_t) = E(X_t - X_0) = (t-0)\xi = t\xi$$

and for $s \leq t$

$$\begin{aligned}
\mathrm{Cov}(X_s, X_t) &= \mathrm{Cov}(X_s, X_s + (X_t - X_s)) \\
&= \mathrm{V}(X_s) + \mathrm{Cov}(X_s, X_t - X_s) \\
&= \mathrm{V}(X_s) + 0 = \mathrm{V}(X_s - X_0) = (s-0)\sigma^2 = s\sigma^2\,.
\end{aligned}$$

We have shown that the finite-dimensional distributions of a Brownian motion with drift ξ and variance σ^2 are given by

$$P^{(n+1)}_{t_1,\ldots,t_{n+1}} = \mathcal{N}\left(\begin{pmatrix} t_1\xi \\ t_2\xi \\ t_3\xi \\ \vdots \\ t_{n+1}\xi \end{pmatrix}, \begin{pmatrix} t_1\sigma^2 & t_1\sigma^2 & t_1\sigma^2 & \cdots & t_1\sigma^2 \\ t_1\sigma^2 & t_2\sigma^2 & t_2\sigma^2 & \cdots & t_2\sigma^2 \\ t_1\sigma^2 & t_2\sigma^2 & t_3\sigma^2 & \cdots & t_3\sigma^2 \\ \vdots & \vdots & \vdots & & \vdots \\ t_1\sigma^2 & t_2\sigma^2 & t_3\sigma^2 & \cdots & t_{n+1}\sigma^2 \end{pmatrix}\right)$$

Finding $\pi_k(P^{(n+1)}_{t_1\ldots t_{n+1}})$ (cf. (6.2)) is now simple: the result is an n-dimensional normal distribution, where the mean vector is obtained by deleting the k'th entry in the mean vector for $P^{(n+1)}_{t_1\ldots t_{n+1}}$, and the covariance matrix is obtained by deleting the k'th row and the k'th column of the covariance matrix for $P^{(n+1)}_{t_1\ldots t_{n+1}}$. It is immediately seen that we thus obtain $P^{(n)}_{t_1\ldots t_{k-1}t_{k+1}\ldots t_{n+1}}$, so by the consistency theorem there is exactly one probability $\hat{P}$ on $(\mathbb{R}^{[0,\infty)}, \mathcal{B}^{[0,\infty)})$ with finite-dimensional distributions given by the normal distribution above. With this probability measure $\hat{P}$, the process consisting of all the coordinate projections $\hat{X} = (\hat{X}_t)_{t\geq 0}$ becomes a Brownian motion with drift ξ and variance σ^2. □

The following lemma will be useful in Section 6.2:

Lemma 6.1.10. *Assume that $X = (X_t)$ is a Brownian Motion with drift ξ and variance σ^2. Let $u \geq 0$. Then*

$$(X_s)_{s\geq 0} \overset{\mathcal{D}}{=} (X_{u+s} - X_u)_{s\geq 0}$$

Proof. We will show that the two processes have the same finite–dimensional distributions. So let $0 \leq t_1 < \cdots < t_n$. Then we show

$$(X_{t_1}, \ldots, X_{t_n}) \overset{\mathcal{D}}{=} (X_{t_1+u} - X_u, X_{t_2+u} - X_u, \ldots, X_{t_n+u} - X_u)\,. \tag{6.5}$$

In the proof of Theorem 6.1.9 we obtained that

$$(X_{t_1}, \ldots, X_{t_n})$$

is n–dimensional normally distributed, since it is a linear transformation of

$$(X_{t_1} - X_0, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}})$$

where the coordinates are independent and normally distributed. In the exact same way we can see that

$$(X_{u+t_1} - X_u, \ldots, X_{u+t_n} - X_u)$$

is n–dimensional normally distributed, since it is a linear transformation of

$$(X_{u+t_1} - X_u, X_{u+t_2} - X_{u+t_1}, \ldots, X_{u+t_n} - X_{u+t_{n-1}})$$

that have independent and normally distributed coordinates. So both of the vectors in (6.5) are normally distributed. To see that the two vectors have the same mean vector and covariance matrix, it suffices to show that for $0 \leq s$

$$EX_s = E(X_{u+s} - X_u)$$

and for $0 \leq s_1 < s_2$

$$\mathrm{Cov}(X_{s_1}, X_{s_2}) = \mathrm{Cov}(X_{u+s_1} - X_u, X_{u+s_2} - X_u)\,.$$

We obtain

$$E(X_{u+s} - X_u) = EX_{u+s} - EX_u = \xi(u+s) - \xi u = \xi s = EX_s$$

and

$$\begin{aligned}
&\mathrm{Cov}(X_{u+s_1} - X_u, X_{u+s_2} - X_u) \\
=&\mathrm{Cov}(X_{u+s_1} - X_u, X_{u+s_1} - X_u + X_{u+s_2} - X_{u+s_1}) \\
=&\mathrm{Cov}(X_{u+s_1} - X_u, X_{u+s_1} - X_u) + \mathrm{Cov}(X_{u+s_1} - X_u, X_{u+s_2} - X_{u+s_1}) \\
=&V(X_{u+s_1} - X_u) = \sigma^2 s_1 = \mathrm{Cov}(X_{s_1}, X_{s_2})
\end{aligned}$$

□

6.2 Continuity of the Brownian motion

In the previous section we saw how it is possible using Kolmogorov's consistency theorem to construct probabilities on the function space $(\mathbb{R}^{[0,\infty)}, \mathcal{B}_{[0,\infty)})$. Thereby we also obtained a construction of stochastic processes $X = (X_t)_{t\geq 0}$ with given finite-dimensional distributions. However, if one aims to construct processes (X_t), which are well-behaved when viewed as functions of t, the function space $(\mathbb{R}^{[0,\infty)}, \mathcal{B}_{[0,\infty)})$ is much too large, as we shall presently see. Let $X = (X_t)_{t\geq 0}$ be a real process, defined on $(\Omega, \mathcal{F}, P)$. The *sample paths* of the process are those elements

$$t \to X_t(\omega)$$

in $\mathbb{R}^{[0,\infty)}$ which are obtained by letting ω vary in Ω. One might then be interested in determining whether (almost all) the sample paths are continuous, i.e., whether

$$P(X \in C_{[0,\infty)}) = 1\,,$$

where $C_{[0,\infty)} \subseteq \mathbb{R}^{[0,\infty)}$ is the set of continuous $x : [0,\infty) \to \mathbb{R}$. The problem is, that $C_{[0,\infty)}$ is *not* in $\mathcal{B}_{[0,\infty)}$!

We will show this by finding two $\mathcal{B}_{[0,\infty)}$–measurable processes X and Y defined on the same $(\Omega, \mathcal{F}, P)$ and with the same finite dimensional distributions, but such that all sample paths for X are continuous, and all sample paths for Y are discontinuous in all $t \geq 0$. The processes X and Y are constructed in Example 6.2.1. The existence of such processes X and Y gives that

$$(X \in C_{[0,\infty)}) = \Omega \qquad\qquad (Y \in C_{[0,\infty)}) = \emptyset$$

and if $C_{[0,\infty)}$ was measurable the identical distributions would lead to

$$P(X \in C_{[0,\infty)}) = P(Y \in C_{[0,\infty)})\,,$$

which is a contradiction!

Example 6.2.1. Let U be defined on $(\Omega, \mathcal{F}, P)$ and assume that U has the uniform distribution on $[0,1]$.

Define

$$X_t(\omega) = 0 \qquad \text{for all } \omega \in \Omega, t \geq 0$$

and

$$Y_t(\omega) = \begin{cases} 0, & \text{if } U(\omega) - t \text{ is irrational} \\ 1, & \text{if } U(\omega) - t \text{ is rational} \end{cases}\,.$$

The finite dimensional distributions of X are degenerated

$$P(X_{t_1} = \cdots = X_{t_n} = 0) = 1$$

for all $n \in \mathbb{N}$ and $0 \leq t_1 < \cdots < t_n$. For Y we have

$$P(Y_t = 1) = P(U - t \in \mathbb{Q}) = 0$$

so $P(Y_t = 0) = 1$ and thereby also

$$P(Y_{t_1} = \cdots = Y_{t_n} = 0) = 1$$

This shows that X and Y have the same finite dimensional distributions. ◦

Thus constructing a continuous process will take more than distributional arguments. In the following we discuss a concrete approach that leads to the construction of a continuous Brownian motion.

Definition 6.2.2. *If the processes $X = (X_t)_{t \geq 0}$ and $Y = (Y_t)_{t \geq 0}$ are both defined on $(\Omega, \mathcal{F}, P)$, then we say that Y is a version of X if*

$$P(Y_t = X_t) = 1$$

for all $t \geq 0$.

We see that Definition 6.2.2 is symmetric: If Y is a version of X, then X is also a version of Y.

Example 6.2.3. With X and Y as in Example 6.2.1 from above, we have

$$(Y_t = X_t) = (Y_t = 0)$$

and since we have seen that $P(Y_t = 0) = 1$, then Y is a version of X. ◦

Theorem 6.2.4. *If Y is a version of X, then Y has the same distribution as X.*

Proof. The idea is to show that Y and X have the same finite–dimensional distributions: With $t_1 < \cdots < t_n$ we have $P(Y_{t_k} = X_{t_k}) = 1$ for $k = 1, \ldots, n$. Then also

$$P((Y_{t_1}, \ldots, Y_{t_n}) = (X_{t_1}, \ldots, X_{t_n})) = P\Big(\bigcap_{k=1}^{n} (Y_{t_k} = X_{t_k})\Big) = 1$$

□

The aim is to show that there exists a continuous version of the Brownian motion. Define for $n \in \mathbb{N}$

$$C_n^\circ = \{x \in \mathbb{R}^{[0,\infty)} \; : \; x \text{ is uniformly continuous on } [0,n] \cap \mathbb{Q}\}$$

and

$$C_\infty^\circ = \bigcap_{n=1}^{\infty} C_n^\circ$$

Lemma 6.2.5. *If $x \in C_\infty^\circ$ then there exists a uniquely determined continuous function $y \in \mathbb{R}^{[0,\infty)}$ such that $y_q = x_q$ for all $q \in [0,\infty) \cap \mathbb{Q}$.*

Proof. Let $x \in C_\infty^\circ$ and $t \geq 0$. Then choose n such that $n > t$. We have that $x \in C_n^\circ$, so x is uniformly continuous on $[0,n] \cap \mathbb{Q}$. That means

$$\forall \epsilon > 0 \; \exists \delta > 0 \; \forall q_1, q_2 \in [0,n] \cap \mathbb{Q} \; : \; |q_1 - q_2| < \delta \Rightarrow |x_{q_1} - x_{q_2}| < \epsilon$$

Choose a sequence $(q_k) \subseteq [0,n] \cap \mathbb{Q}$ with $q_k \to t$. Then in particular (q_k) is a Cauchy sequence. The uniform continuity of x gives that x_{q_k} is a Cauchy sequence as well: Let $\epsilon > 0$ and find the corresponding $\delta > 0$. We can find $K \in \mathbb{N}$ such that for all $m, n \geq K$ it holds

$$|q_m - q_n| < \delta$$

But then we must have that

$$|x_{q_m} - x_{q_n}| < \epsilon$$

if only $m, n \geq K$. This shows, that (x_{q_k}) is Cauchy, and therefore the limit $y_t = \lim_{k\to\infty} x_{q_k}$ exists in $\mathbb{R}$. We furthermore have, that the limit y_t does not depend on the choice of (q_k): Let $(\tilde{q}_k) \subseteq [0,n] \cap \mathbb{Q}$ be another sequence with $\tilde{q}_k \to t$. Then

$$|\tilde{q}_k - q_k| \to 0 \qquad \text{as } k \to \infty$$

and this yields (using the uniform continuity again) that

$$|x_{\tilde{q}_k} - x_{q_k}| \to 0 \qquad \text{as } k \to \infty$$

so $\lim x_{\tilde{q}_k} = \lim x_{q_k}$.

For all $t \in \mathbb{Q}$ we see that $y_t = x_t$, since the continuity of x in t gives $\lim_{k\to\infty} x_{q_k} = x_t$.

Finally we have, that y is continuous in all $t \geq 0$: Let $t \geq 0$ and $\epsilon > 0$ be given, and find $\delta > 0$ according to the uniform continuity. Now choose t' with $|t' - t| < \delta/2$. Assume that $q_k \to t$ and $q_k' \to t'$. We can find $K \in \mathbb{N}$ such that $|q_k' - q_k| < \delta$ for $k \geq K$. Then

$$|x_{q_k'} - x_{q_k}| < \epsilon$$

for all $k \geq K$, and thereby we obtain that

$$|y_{t'} - y_t| \leq \epsilon\,.$$

This shows the desired continuity of y in t. □

It is a critical assumption, that the continuity is uniform. Consider x given by

$$x_t = 1_{[\sqrt{2},\infty)}$$

Then x is continuous on $[0,n] \cap \mathbb{Q}$, but an y does not exist with the required properties.

We obtain, that $C^\circ_\infty \in \mathbb{B}^{[0,\infty)}$ since

$$C_n^\circ = \bigcap_{M=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{q_1,q_2 \in [0,n]\cap\mathbb{Q}, |q_1-q_2| \leq \frac{1}{N}} \Big\{ x \in \mathbb{R}^{[0,\infty)} \;:\; |x_{q_2} - x_{q_1}| < \frac{1}{M} \Big\}\,,$$

which is a $\mathcal{B}_{[0,\infty)}$–measurable set, since

$$\Big\{ x \in \mathbb{R}^{[0,\infty)} \;:\; |x_{q_2} - x_{q_1}| < \frac{1}{M} \Big\} = \Big(|\hat{X}_{q_2} - \hat{X}_{q_1}| < \frac{1}{M} \Big)\,,$$

where $\hat{X}_{q_1}, \hat{X}_{q_2} : \mathbb{R}^{[0,\infty)} \to \mathbb{R}$ are both $\mathcal{B}_{[0,\infty)} - \mathcal{B}$–measurable.

Definition 6.2.6. *A real process $X = (X_t)_{t\geq 0}$ is continuous in probability if for all $t \geq 0$ and all sequences $(t_k)_{k\in\mathbb{N}}$ with $t_k \geq 0$ and $t_k \to t$ it holds that $X_{t_k} \xrightarrow{P} X_t$*

Theorem 6.2.7. *Let $X = (X_t)_{t\geq 0}$ be a real process which is continuous in probability. If*

$$P(X \in C^\circ_\infty) = 1\,,$$

then there exists a version Y of X which is continuous.

Proof. Let $F = (X \in C^\circ_\infty)$. Assume that $\omega \in F$. According to Lemma 6.2.5 there exists a uniquely determined continuous function $t \mapsto Y_t(\omega)$ such that

$$Y_q(\omega) = X_q(\omega) \qquad \text{for all } q \in [0,\infty) \cap \mathbb{Q}\,. \tag{6.6}$$

Furthermore we must have for each $t \geq 0$ that a rational sequence (q_k) can be chosen with $q_k \to t$. Then using the continuity of $t \to Y_t(\omega)$ and the property in (6.6) yields that for all $\omega \in F$

$$Y_t(\omega) = \lim_{k\to\infty} Y_{q_k}(\omega) = \lim_{k\to\infty} X_{q_k}(\omega)\,.$$

If we furthermore define $Y_t(\omega) = 0$ for $\omega \in F^c$, then we have

$$Y_t = \lim_{k\to\infty} 1_F X_{q_k}\,.$$

Since all $1_F X_{q_k}$ are random variables (measurable), then Y_t is a random variable as well. And since $t \geq 0$ was chosen arbitrarily, then $Y = (Y_t)_{t\geq 0}$ is a continuous real process (for $\omega \in F^c$ we chose $(Y_t(\omega))$ to be constantly 0 – which is a continuous function) that satisfies.

$$P(Y_q = X_q) = 1 \qquad \text{for all } q \in [0,\infty)\cap\mathbb{Q}\,,$$

since $P(F) = 1$ and $Y_q(\omega) = X_q(\omega)$ when $\omega \in F$.

We still need to show, that Y is a version of X. So let $t \geq 0$ and find a rational sequence (q_k) with $q_k \to t$. Since X is assumed to be continuous in probability we must have

$$X_{q_k} \xrightarrow{P} X_t$$

and since we have $Y_{q_k} \overset{\text{a.s.}}{=} X_{q_k}$ it holds

$$Y_{q_k} \xrightarrow{P} X_t\,.$$

From the (true) continuity we have the convergence (for all $\omega \in \Omega$)

$$Y_{q_k} \to Y_t$$

Then

$$P(Y_t = X_t) = 1\,.$$

as desired. □

Theorem 6.2.8. *Let $X = (X_t)_{t\geq 0}$ be a Brownian motion with drift ξ and variance $\sigma^2 > 0$. Then X has a continuous version.*

Proof. It is sufficient to consider the normalized case, where $\xi = 0$ and $\sigma^2 = 1$. For a general choice of ξ and σ^2 we have that

$$\tilde{X}_t = \frac{X_t - \xi t}{\sigma}$$

is a normalized Brownian motion. And obviously, $(X_t)_{t\geq 0}$ is continuous if and only if $(\tilde{X}_t)_{t\geq 0}$ is continuous.

So let $X = (X_t)_{t\geq 0}$ be a normalized Brownian motion. Firstly, we show that X is continuous in probability. For all $0 \leq s < t$ we have

$$X_t - X_s \sim \mathcal{N}(0, t-s)$$

such that

$$\frac{1}{\sqrt{t-s}}(X_t - X_s) \sim \mathcal{N}(0,1)$$

Then for $\epsilon > 0$ we have

$$\begin{aligned} P(|X_t - X_s| > \epsilon) &= P\big(\frac{1}{\sqrt{t-s}}|X_t - X_s| > \frac{\epsilon}{\sqrt{t-s}}\big) \\ &= \int_{-\infty}^{-\frac{\epsilon}{\sqrt{t-s}}} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\,\mathrm{d}u + \int_{\frac{\epsilon}{\sqrt{t-s}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\,\mathrm{d}u \\ &= 2\int_{\frac{\epsilon}{\sqrt{t-s}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\,\mathrm{d}u \end{aligned}$$

For general $s, t \geq 0$ with $s \neq t$ we clearly have

$$P(|X_t - X_s| > \epsilon) = 2\int_{\frac{\epsilon}{\sqrt{|t-s|}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\,\mathrm{d}u$$

and this decreases to 0 as $|t-s| \to 0$. Hence in particular, we have for $t_k \to t$ that

$$P(|X_t - X_{t_k}| > \epsilon) \to 0 \qquad \text{as } k \to \infty$$

which demonstrates the continuity in probability.

The following, that is actually a stronger version of the continuity in probability, will be useful. It holds for all $\epsilon > 0$ that

$$\lim_{h \downarrow 0} \frac{1}{h} P(|X_h| > \epsilon) = 0\,. \tag{6.7}$$

This follows from Markov's inequality, since

$$\frac{1}{h}P(|X_h| > \epsilon) = \frac{1}{h}P(X_h^4 > \epsilon^4) \leq \frac{1}{h\epsilon^4} E X_h^4 = \frac{h}{\epsilon^4} E(\frac{1}{\sqrt{h}} X_h)^4 = \frac{3h}{\epsilon^4}$$

which has limit 0, as $h \to 0$. In the last equality we used that $\frac{1}{\sqrt{h}}X_h$ is $\mathcal{N}(0,1)$–distributed and that the $\mathcal{N}(0,1)$–distribution has fourth moment $= 3$.

It is left to show that

$$P(X \in C_\infty^\circ) = 1$$

and for this it suffices to show that

$$P(X \in C_n^\circ) = 1$$

for all $n \in \mathbb{N}$, recalling that a countable intersection of sets with probability 1 has probability 1. We show this for $n = 1$ (higher values of n would not change the argument, only make the notation more involved). Define

$$V_N = \sup\left\{ |X_{q'} - X_q| \ : \ q, q' \in \mathbb{Q} \cap [0,1], |q' - q| \leq \frac{1}{2^N} \right\}.$$

Then V_N decreases as $N \to \infty$, so

$$\begin{aligned}(X \in C_1^\circ) &= \bigcap_{M=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{q_1,q_2 \in [0,1]\cap\mathbb{Q}, |q_2-q_1| \leq \frac{1}{N}} \left(|X_{q_2} - X_{q_1}| \leq \frac{1}{M} \right) \\ &= \bigcap_{M=1}^{\infty} \bigcup_{N=1}^{\infty} \left(V_N \leq \frac{1}{M} \right) = (\lim_{N\to\infty} V_N = 0)\end{aligned}$$

Hence we need to show that $P(\lim_{N\to\infty} V_N = 0) = 1$. Since we already know, that V_N is decreasing, it will be enough to show that $V_N \xrightarrow{P} 0$. So we need to show that for any $\epsilon > 0$, $P(V_N > \epsilon) \to 0$ as $N \to \infty$.

For this define for $N \in \mathbb{N}$ and $k = 1, \ldots, 2^N$

$$Y_{k,N} = \sup\{|X_q - X_{\frac{k-1}{2^N}}| \mid q \in J_{k,N}\},$$

where

$$J_{k,N} = \left[\frac{k-1}{2^N}, \frac{k}{2^N}\right] \cap \mathbb{Q}.$$

If we can show that

(1) $V_N \leq 3 \max\{Y_{k,N} \mid 1 \leq k \leq 2^N\}$

(2) $P(Y_{k,N} > y) = P(Y_{1,N} > y) \leq 2P(|X_{\frac{1}{2^N}}| > y)$

then we obtain

$$\begin{aligned}P(V_N > \epsilon) &\leq P\Big(\max_{k=1,\ldots,2^N} Y_{k,N} > \frac{\epsilon}{3}\Big) = P\left(\bigcup_{k=1}^{2^N} \Big(Y_{k,N} > \frac{\epsilon}{3}\Big)\right) \\ &\leq \sum_{k=1}^{2^N} P\Big(Y_{k,N} > \frac{\epsilon}{3}\Big) = 2^N P\Big(Y_{1,N} > \frac{\epsilon}{3}\Big) \leq 2^{N+1} P\Big(|X_{\frac{1}{2^N}}| > \frac{\epsilon}{3}\Big)\end{aligned}$$

which has limit 0 as $N \to \infty$ because of (6.7). The first inequality is according to (1), the second inequality follows from Boole's inequality, and the last inequality is due to (2). Hence, the proof is complete if we can show (1) and (2).

For (1): Consider for some fixed $N \in \mathbb{N}$ the q, q' that are used in the definition of V_N. Hence $q < q' \in \mathbb{Q} \cap [0,1]$ where $|q' - q| \leq \frac{1}{2^N}$. We have two possibilities:

Either q, q' belong to the same $J_{k,N}$ such that

$$\begin{aligned}
|X_{q'} - X_q| &= |X_{q'} - X_{\frac{k-1}{2^N}} + X_{\frac{k-1}{2^N}} - X_q| \\
&\leq |X_{q'} - X_{\frac{k-1}{2^N}}| + |X_q - X_{\frac{k-1}{2^N}}| \\
&\leq 2Y_{k,N} \\
&\leq 2\max\{Y_{k,N} \mid 1 \leq k \leq 2^N\}\,,
\end{aligned}$$

or $q \in J_{k-1,N}$ and $q' \in J_{k,N}$. Then

$$\begin{aligned}
|X_{q'} - X_q| &= |X_{q'} - X_{\frac{k-1}{2^N}} + X_{\frac{k-1}{2^N}} - X_{\frac{k-2}{2^N}} + X_{\frac{k-2}{2^N}} - X_q| \\
&\leq |X_{q'} - X_{\frac{k-1}{2^N}}| + |X_{\frac{k-1}{2^N}} - X_{\frac{k-2}{2^N}}| + |X_q - X_{\frac{k-2}{2^N}}| \\
&\leq Y_{k,N} + 2Y_{k-1,N} \\
&\leq 3\max\{Y_{k,N} \mid 1 \leq k \leq 2^N\}\,.
\end{aligned}$$

In any case we have

$$|X_{q'} - X_q| \leq 3\max\{Y_{k,N} \mid 1 \leq k \leq 2^N\}\,,$$

where the right hand side does not depend on q, q'. Property (1) follows from taking the supremum.

For (2): Note that for all $k = 2, \ldots, 2^N$, the variable $Y_{k,N}$ is calculated from the process

$$(X_{\frac{k-1}{2^N}+s} - X_{\frac{k-1}{2^N}})_{s \geq 0}$$

in the exact same way as $Y_{1,N}$ is calculated from the process $(X_s)_{s \geq 0}$.

Also note that because of the Lemma 6.1.10 the two processes

$$(X_{\frac{k-1}{2^N}+s} - X_{\frac{k-1}{2^N}})_{s \geq 0} \quad \text{and} \quad (X_s)_{s \geq 0}$$

have the same distributions. Then also $Y_{k,N} \stackrel{\mathcal{D}}{=} Y_{1,N}$ for all $k = 2, \ldots, 2^N$, such that in particular

$$P(Y_{k,N} > y) = P(Y_{1,N} > y) \leq 2P(|X_{\frac{1}{2^N}}| > y)$$

for all $y > 0$. The inequality comes from Lemma 6.2.9 below, since $J_{1,N}$ is countable with $J_{1,N} \subseteq [0, \frac{1}{2^N}]$. □

Lemma 6.2.9. *Let $X = (X_t)_{t\geq 0}$ be a normalized Brownian Motion and let $D \subseteq [0, t_0]$ be an at most countable set. Then it holds for $x > 0$ that*

$$P(\sup_{t\in D} X_t > x) \leq 2P(X_{t_0} > x)$$
$$P(\sup_{t\in D} |X_t| > x) \leq 2P(|X_{t_0}| > x)$$

Proof. First assume that D is finite such that $D = \{t_1, \dots, t_n\}$, where $0 \leq t_1 < \cdots < t_n \leq t_0$. Define

$$\tau = \min\{k \in \{1, \dots, n\} \mid X_{t_k} > x\}$$

and let $\tau = n$ if $X_{t_k} \leq x$ for all $k = 1, \dots, n$. Then

$$P(\sup_{t\in D} X_t > x) = \sum_{k=1}^{n-1} P(\tau = k) + P(\tau = n, X_{t_n} > x)$$

Let $k \leq n - 1$. Then

$$(\tau = k) = \bigcap_{j=1}^{k-1} (X_{t_j} \leq x) \cap (X_{t_k} > x)$$

and note that $(X_{t_1}, \dots, X_{t_k}) \perp\!\!\!\perp (X_{t_n} - X_{t_k})$, so in particular

$$(\tau = k) \quad \perp\!\!\!\perp \quad (X_{t_n} - X_{t_k}) .$$

Furthermore $(X_{t_n} - X_{t_k}) \sim \mathcal{N}(0, t_n - t_k)$ so $P(X_{t_n} - X_{t_k} > 0) = \frac{1}{2}$. Hence

$$\begin{aligned} P(\tau = k) &= 2P(\tau = k)P(X_{t_n} - X_{t_k} > 0) = 2P(\tau = k, X_{t_n} - X_{t_k} > 0) \\ &= 2P(\tau = k, X_{t_n} > X_{t_k}) \leq 2P(\tau = k, X_{t_n} > x) , \end{aligned}$$

where it is used that $X_{t_k} > x$ on $(\tau = k)$. Then

$$\begin{aligned} P(\sup_{t\in D} X_t > x) &= \sum_{k=1}^{n-1} P(\tau = k) + P(\tau = n, X_{t_n} > x) \\ &\leq 2\sum_{k=1}^{n-1} P(\tau = k, X_{t_n} > x) + 2P(\tau = n, X_{t_n} > x) \\ &= 2P(X_{t_n} > x) \leq 2P(X_{t_0} > x) . \end{aligned}$$

In the last inequality it is used that $t_n \leq t_0$ such that X_{t_0} have a larger variance than X_{t_n} (both variables have mean 0).

Thereby we have shown the first result in the case, where D is finite. To obtain the second result for a finite D, consider the process $-X = (-X_t)_{t\geq 0}$, which is again a normalised Brownian motion. Hence for $x > 0$ we have

$$P(\inf_{t\in D} X_t < -x) = P(\sup_{t\in D}(-X_t) > x) \leq 2P(-X_{t_0} > x) = 2P(X_{t_0} < -x)$$

so we can obtain

$$\begin{aligned} P(\sup_{t\in D}|X_t| > x) &= P\big((\sup_{t\in D} X_t > x) \cup (\inf_{t\in D} X_t < -x)\big) \\ &\leq P(\sup_{t\in D} X_t > x) + P(\inf_{t\in D} X_t < -x) \\ &\leq 2P(X_{t_0} > x) + 2P(X_{t_0} < -x) \\ &= 2P(|X_{t_0}| > x) \end{aligned}$$

Then we have also shown the second result, when D is finite.

For a general D find a sequence (D_n) of finite subsets of D where $D_n \uparrow D$. Then the two inequalities holds for each D_n. Since furthermore

$$\begin{aligned} (\sup_{t\in D_n} X_t > x) &\uparrow (\sup_{t\in D} X_t > x) \\ (\sup_{t\in D_n} |X_t| > x) &\uparrow (\sup_{t\in D} |X_t| > x) \end{aligned}$$

the continuity of the probability measure P yields that

$$\begin{aligned} P(\sup_{t\in D} X_t > x) &= \lim_{n\to\infty} P(\sup_{t\in D_n} X_t > x) \leq 2P(X_{t_0} > x) \\ P(\sup_{t\in D} |X_t| > x) &= \lim_{n\to\infty} P(\sup_{t\in D_n} |X_t| > x) \leq 2P(|X_{t_0}| > x) \end{aligned}$$

which completes the proof of the lemma. □

6.3 Variation and quadratic variation

In this and the subsequent section we study the sample paths of a continuous Brownian motion. In this framework it will be useful to consider the space $C_{[0,\infty)}$ consisting of all functions $x \in \mathbb{R}^{[0,\infty)}$ that are continuous. Like the projections on $\mathbb{R}^{[0,\infty)}$ we let $\tilde{X}_t$ denote the coordinate projections on $C_{[0,\infty)}$, that is $\tilde{X}_t(x) = x_t$ for all $x \in C_{[0,\infty)}$. Let $\mathcal{C}_{[0,\infty)}$ denote the smallest σ–algebra, that makes all $\tilde{X}_t$ $\mathcal{C}_{[0,\infty)} - \mathcal{B}$–measurable

$$\mathcal{C}_{[0,\infty)} = \sigma(\tilde{X}_t \mid t \geq 0) .$$

Similarly to what we have seen previously, $\mathcal{C}_{[0,\infty)}$ is generated by all the finite–dimensional cylinder sets

$$\mathcal{C}_{[0,\infty)} = \sigma\Big(\big((\tilde{X}_{t_1}, \dots, \tilde{X}_{t_n}) \in B_n\big) \mid n \in \mathbb{N}, 0 < t_1 < \cdots < t_n, B_n \in \mathcal{B}_n\Big) .$$

We demonstrated in Section 6.2 that there exists a process X defined on $(\Omega, \mathcal{F}, P)$ with values in $(\mathbb{R}^{[0,\infty)}, \mathcal{B}_{[0,\infty)})$ such that X is a Brownian motion $X = (X_t)$ and the sample paths $t \mapsto X_t(\omega)$ are continuous for *all* $\omega \in \Omega$. Equivalently, we have $X(\omega) \in C_{[0,\infty)}$ for all $\omega \in \Omega$, so we can regard the process X as having values in $C_{[0,\infty)}$. That X is measurable with values in $(\mathbb{R}^{[0,\infty)}, \mathcal{B}_{[0,\infty)})$ means that $\hat{X}_t(X)$ is $\mathcal{F} - \mathcal{B}$ measurable for all $t \geq 0$. But $\tilde{X}_t(X) = \hat{X}_t(X)$ since X is continuous, so $\tilde{X}_t(X)$ is also $\mathcal{F} - \mathcal{B}$ measurable for all $t \geq 0$. Then X is measurable, when regarded as a variable with values in $(C_{[0,\infty)}, \mathcal{C}_{[0,\infty)})$. The distribution $X(P)$ of X will be a distribution on $(C_{[0,\infty)}, \mathcal{C}_{[0,\infty)})$, and this is uniquely determined by the behaviour on the finite–dimensional cylinder sets on the form $\big((\tilde{X}_{t_1}, \dots, \tilde{X}_{t_n}) \in B_n\big)$.

The space $(C_{[0,\infty)}, \mathcal{C}_{[0,\infty)})$ is significantly easier to deal with than $(\mathbb{R}^{\infty}, \mathcal{B}_{[0,\infty)})$, and a number of interesting functionals become measurable on $C_{[0,\infty)}$, while they are not measurable on $\mathbb{R}^{[0,\infty)}$. For instance, for $t > 0$ we have that

$$\tilde{M} = \sup_{s \in [0,t]} \tilde{X}_s$$

is a measurable function (a random variable) on $(C_{[0,\infty)}, \mathcal{C}_{[0,\infty)})$, which can be seen by

$$(\tilde{M} \leq y) = \bigcap_{s \in [0,t]} (\tilde{X}_s \leq y) = \bigcap_{q \in [0,t] \cap \mathbb{Q}} (\tilde{X}_q \leq y)$$

where the last intersection is countable – hence measurable. For the last equality, the inclusion '$\subseteq$' is trivial. For the converse inclusion, assume that

$$x \in \bigcap_{q \in [0,t] \cap \mathbb{Q}} (\tilde{X}_q \leq y) .$$

Then $x_q \leq y$ for all $q \in [0,t] \cap \mathbb{Q}$. Let $s \in [0,t]$ and find a rational sequence $q_n \to s$. Then $x_s = \lim_{n \to \infty} x_{q_n} \leq y$ and since s was arbitrarily chosen, it holds that

$$x \in \bigcap_{s \in [0,t]} (\tilde{X}_s \leq y) .$$

We will define various concepts that can be used to describe the behaviour of the sample paths of a process.

Definition 6.3.1. *Let $x \in C_{[0,\infty)}$. We say that x is nowhere monotone if it for all $0 \leq s < t$ holds that x is neither increasing nor decreasing on $[s,t]$. Let $S \subseteq C_{[0,\infty)}$ denote the set of nowhere monotone functions.*

Related to the set S we define M_{st} to be the set of functions, which are either increasing or decreasing on the interval $[s,t]$:

$$M_{st} = \bigcap_{N=1}^{\infty} \{x \in C_{[0,\infty)} \mid x_{t_{kN}} \geq x_{t_{k-1,N}},\, 1 \leq k \leq 2^N\} \\ \cup \bigcap_{N=1}^{\infty} \{x \in C_{[0,\infty)} \mid x_{t_{kN}} \leq x_{t_{k-1,N}},\, 1 \leq k \leq 2^N\},$$

where $t_{kN} = s + \frac{k}{2^N}(t-s)$ for $0 \leq k \leq 2^N$. We note that $M_{st} \in \mathcal{C}_{[0,\infty)}$, since e.g.

$$\{x \in C_{[0,\infty)} \mid x_{t_{kN}} \leq x_{t_{k-1,N}},\, 1 \leq k \leq 2^N\} = \left(\tilde{X}_{t_{kN}} \leq \tilde{X}_{t_{k-1,N}},\, 1 \leq k \leq 2^N\right).$$

Since $x \in S^c$ if and only if there exists intervals with rational endpoints where x is monotone, then we can write

$$S^c = \bigcup_{\substack{0 \leq q_1 < q_2 \\ q_1, q_2 \in \mathbb{Q}}} M_{q_1 q_2},$$

which shows that $S \in \mathcal{C}_{[0,\infty)}$. We shall see later that $P(X \in S) = 1$ for a continuous Brownian motion X.

Definition 6.3.2. *Let $x \in C_{[0,\infty)}$ and $0 \leq s < t$. The variation of x on $[s,t]$ is defined as*

$$V_{st}(x) = \sup \sum_{k=1}^{n} |x_{t_k} - x_{t_{k-1}}|,$$

where sup *is taken over all finite partitions $s \leq t_0 < \cdots < t_n \leq t$ of $[s,t]$.*

The variation has some simple properties:

Lemma 6.3.3. *Let $x, y \in C_{[0,\infty)}$, $c \in \mathbb{R}$, $0 \leq s < t$ and $[s,t] \subseteq [s',t']$. Then it holds that*

(1) $V_{st}(x) \leq V_{s't'}(x)$.

(2) $V_{st}(cx) = |c| V_{st}(x)$.

(3) $V_{st}(x+y) \leq V_{st}(x) + V_{st}(y)$.

Proof. The first statement is because the sup in $V_{s't'}(x)$ is over more partitions than the sup in $V_{st}(x)$. For the second result we have

$$V_{st}(cx) = \sup \sum_{k=1}^{n} |cx_{t_k} - cx_{t_{k-1}}| = |c| \sup \sum_{k=1}^{n} |x_{t_k} - x_{t_{k-1}}| = |c| V_{st}(x)\,,$$

and the third property follows from

$$\begin{aligned} V_{st}(x+y) &= \sup \sum_{k=1}^{n} |x_{t_k} + y_{t_k} - x_{t_{k-1}} - y_{t_{k-1}}| \\ &\leq \sup \sum_{k=1}^{n} |x_{t_k} - x_{t_{k-1}}| + |y_{t_k} - y_{t_{k-1}}| \\ &\leq \sup \sum_{k=1}^{n} |x_{t_k} - x_{t_{k-1}}| + \sup \sum_{k=1}^{n} |y_{t_k} - y_{t_{k-1}}| \\ &= V_{st}(x) + V_{st}(y) \end{aligned}$$

□

Furthermore we have situations, where the variation is particularly simple

Lemma 6.3.4. *If $x \in C_{[0,\infty)}$ is monotone on $[s,t]$ then*

$$V_{st} = |x_t - x_s|$$

Proof. If x is monotone on $[s,t]$ then

$$\sum_{k=1}^{n} |x_{t_k} - x_{t_{k-1}}| = |x_{t_n} - x_{t_0}|$$

for any partition $s \leq t_0 < \cdots < t_n \leq t$ of $[s,t]$. □

Let $x \in C_{[0,\infty)}$ and assume that $s \leq t_{k-1} < t_k \leq t$ are given. If $(q_n), (r_n) \subseteq [s,t]$ are rational sequences with $q_n \to t_{k-1}$ and $r_n \to t_k$, it holds due to the continuity of x that

$$\lim_{n\to\infty} |x_{r_n} - x_{q_n}| = |x_{t_k} - x_{t_{k-1}}|\,.$$

This shows that all partitions can be approximated arbitrarily well by rational partitions, so the sup in the definition of V_{st} needs only to be over all rational partitions. Hence V_{st} is $\mathcal{C}_{[0,\infty)} - \mathcal{B}$ measurable.

Definition 6.3.5. *(1) Let $x \in C_{[0,\infty)}$ and $0 \le s < t$. Then x is of bounded variation on $[s,t]$ if $V_{st}(x) < \infty$. The set of functions of bounded variation on $[s,t]$ is denoted*

$$F_{st} = \{x \in C_{[0,\infty)} \mid V_{st}(x) < \infty\}$$

(2) Let $x \in C_{[0,\infty)}$. Then x is everywhere of unbounded variation, if $x \in F_{st}^c$ for all $0 \le s < t$. Let $G = \cap_{0 \le s < t} F_{st}^c$ denote the set of continuous functions, which are everywhere of unbounded variation.

Since V_{st} is $\mathcal{C}_{[0,\infty)} - \mathcal{B}$ measurable we observe that $F_{st} \in \mathcal{C}_{[0,\infty)}$. Furthermore we can rewrite G as

$$G = \bigcap_{\substack{0 \le q_1 < q_2 \\ q_1, q_2 \in \mathbb{Q}}} F_{q_1,q_2}^c \,,$$

which shows that $G \in \mathcal{C}_{[0,\infty)}$. The equality above is a direct consequence of (1) in Lemma 6.3.3.

The following lemmas shows which type of continuous functions have bounded variation.

Lemma 6.3.6. *Let $x \in C_{[0,\infty)}$. Then $x \in F_{st}$ if and only if x on $[s,t]$ has the form*

$$x = y - \tilde{y} \,,$$

where both y and $\tilde{y}$ are increasing.

Proof. If x has the form $x = y - \tilde{y}$ on $[s,t]$, where both y and $\tilde{y}$ are increasing, then using Lemma 6.3.3 yields

$$V_{st}(x) = V_{st}(y - \tilde{y}) \le V_{st}(y) + V_{st}(-\tilde{y}) = V_{st}(y) + V_{st}(\tilde{y}) = |y_t - y_s| + |\tilde{y}_t - \tilde{y}_s|$$

which is finite. Conversely, assume that $V_{st}(x) < \infty$ and define for $u \in [s,t]$

$$y_u = \frac{1}{2}(x_u + V_{su}(x)) \quad \text{and} \quad \tilde{y}_u = \frac{1}{2}(-x_u + V_{su}(x)) \,.$$

Then $x = y - \tilde{y}$ and furthermore we have, that e.g. $u \to y_u$ is increasing: If $x_{u+h} \ge x_u$, then $y_{u+h} \ge y_u$ since always $V_{s,u+h}(x) \ge V_{s,u}(x)$. If $x_{u+h} < x_u$, then

$$\begin{aligned} V_{s,u}(x) + |x_{u+h} - x_u| &= \tilde{\sup} \sum_{j=1}^{n'} |x_{s_j} - x_{s_{j-1}}| + |x_{u+h} - x_u| \\ &\le \sup \sum_{k=1}^{n} |x_{t_k} - x_{t_{k-1}}| = V_{s,u+h}(x) \,, \end{aligned}$$

where $\tilde{\sup}$ is over all partitions $s \le s_0 < \cdots < s_{n'} \le u$ and sup is over all partitions $s \le t_0 < \cdots < t_n \le u + h$. Hence we have seen that

$$y_u = \frac{1}{2}\Big(x_{u+h} + |x_{u+h} - x_u| + V_{su}(x)\Big) \le \frac{1}{2}\Big(x_{u+h} + V_{s,u+h}(x)\Big) = y_{u+h}\,.$$

□

Corollary 6.3.7. *Let $x \in C_{[0,\infty)}$. If $x \in G$ then $x \in S$.*

Proof. Assume that $x \in S^c$. Then there exists $s < t$ such that x is monotone on $[s,t]$. But then x has the form $x - 0$ on $[s,t]$, where both x and 0 are increasing. Thus $x \in F_{st}$, so $x \in G^c$. □

Lemma 6.3.8. *If $x \in C_{[0,\infty)}$ is continuously differentiable on $[s,t]$, then $x \in F_{st}$.*

Proof. The derivative x' of x is continuous on $[s,t]$, so x' must be bounded on $[s,t]$. Let $K = \sup_{u\in[s,t]} |x'(u)|$ and consider an arbitrary partition $s \le t_0 < \cdots < t_n \le t$. Then

$$\sum_{k=1}^{n} |x_{t_k} - x_{t_{k-1}}| = \sum_{k=1}^{n} (t_k - t_{k-1})|x'_{u_k}| \le (t-s)K < \infty\,,$$

where each $u_k \in [s,t]$ is chosen according to the mean value theorem. □

Definition 6.3.9. *Let $x \in C_{[0,\infty)}$. The quadratic variation of $x \in C_{[0,\infty)}$ on $[s,t]$ is defined as*

$$Q_{st}(x) = \limsup_{N\to\infty} \sum_{\substack{k\in\mathbb{N} \\ s \le \frac{k-1}{2^N} < \frac{k}{2^N} \le t}} (x_{\frac{k}{2^N}} - x_{\frac{k-1}{2^N}})^2\,.$$

We observe that Q_{st} is $\mathcal{C}_{[0,\infty)} - \mathcal{B}$ measurable and that

$$Q_{st}(x) \le Q_{s't'}(x)\,, \tag{6.8}$$

if $[s,t] \subseteq [s',t']$.

Lemma 6.3.10. *For $x \in C_{[0,\infty)}$ we have*

$$V_{st}(x) < \infty \quad \Rightarrow \quad Q_{st}(x) = 0\,,$$

or equivalently

$$Q_{st}(x) > 0 \quad \Rightarrow \quad V_{st}(x) = \infty\,.$$

Proof. For $N \in \mathbb{N}$ define

$$K_N = \max\{|x_{\frac{k}{2^N}} - x_{\frac{k-1}{2^N}}| \mid k \in \mathbb{N},\, s \leq \frac{k-1}{2^N} < \frac{k}{2^N} \leq t\}\,.$$

Since x is uniformly continuous on the compact interval $[s,t]$, we will have $K_N \to 0$ as $N \to \infty$. Furthermore

$$\begin{aligned} Q_{st}(x) &= \limsup_{N\to\infty} \sum_{\substack{k\in\mathbb{N} \\ s\leq\frac{k-1}{2^N}<\frac{k}{2^N}\leq t}} (x_{\frac{k}{2^N}} - x_{\frac{k-1}{2^N}})^2 \\ &\leq \limsup_{N\to\infty} K_N \sum_{\substack{k\in\mathbb{N} \\ s\leq\frac{k-1}{2^N}<\frac{k}{2^N}\leq t}} |x_{\frac{k}{2^N}} - x_{\frac{k-1}{2^N}}| \\ &\leq \limsup_{N\to\infty} V_{st}(x) K_N\,, \end{aligned}$$

from which the lemma follows. □

The main result of the section is the following theorem which describes exactly how "wild" the sample paths of the Brownian motion behaves.

Theorem 6.3.11. *If $X = (X_t)_{t\geq 0}$ is a continuous Brownian motion with drift ξ and variance σ^2, then*

$$P \bigcap_{0\leq s<t} (Q_{st}(X) = (t-s)\sigma^2) = 1\,.$$

Before turning to the proof, we observe:

Corollary 6.3.12. *If $X = (X_t)_{t\geq 0}$ is a continuous Brownian motion with drift ξ and variance σ^2, then X is everywhere of unbounded variation,*

$$P(X \in G) = 1\,.$$

Proof. Follows by combining Theorem 6.3.11 and Lemma 6.3.10. □

Corollary 6.3.13. *If $X = (X_t)_{t\geq 0}$ is a continuous Brownian motion with drift ξ and variance σ^2, then X is nowhere monotone,*

$$P(X \in S) = 1\,.$$

Proof. Is a result of Corollary 6.3.12 and Corollary 6.3.7. □

Proof of Theorem 6.3.11. Firstly we can note that

$$\bigcap_{0\le s<t}\{x\in C_{[0,\infty)}\ :\ Q_{st}(x)=(t-s)\sigma^2\}=\bigcap_{0\le q_1<q_2,\,q_1,q_2\in\mathbb{Q}}\{x\in C_{[0,\infty)}\ :\ Q_{q_1,q_2}(x)=(q_2-q_1)\sigma^2\}$$

The inclusion $\subseteq$ is trivial. The converse inclusion $\supseteq$ is argued as follows: Assume that x is an element of the right hand side and let $0\le s<t$ be given. Then we are supposed to show that $Q_{st}(x)=(t-s)\sigma^2$. Let $(q_n^1),(q_n^2),(r_n^1),(r_n^2)$ be rational sequences such that $q_n^1\uparrow s$, $q_n^2\downarrow t$, $r_n^1\downarrow s$, and $r_n^2\uparrow t$. Then for all $n\in\mathbb{N}$ we must have $[r_n^1,r_n^2]\subseteq[s,t]\subseteq[q_n^1,q_n^2]$ such that because of (6.8) it holds that

$$Q_{r_n^1,r_n^2}(x)\le Q_{st}(x)\le Q_{q_n^1,q_n^2}(x)$$

for all $n\in\mathbb{N}$. By assumption

$$Q_{r_n^1,r_n^2}(x)=(r_n^2-r_n^1)\sigma^2\qquad\text{and}\qquad Q_{q_n^1,q_n^2}(x)=(q_n^2-q_n^1)\sigma^2$$

for all $n\in\mathbb{N}$, leading to

$$\lim_{n\to\infty}Q_{r_n^1,r_n^2}(x)=\lim_{n\to\infty}(r_n^2-r_n^1)\sigma^2=(t-s)\sigma^2$$
$$\lim_{n\to\infty}Q_{q_n^1,q_n^2}(x)=\lim_{n\to\infty}(q_n^2-q_n^1)\sigma^2=(t-s)\sigma^2$$

which combined with the inequality above gives the desired result that $Q_{st}(x)=(t-s)\sigma^2$.

Since the intersection above is countable, we only need to conclude that each of the sets in the intersection has probability 1 in order to conclude the result. Hence it will suffice to show that

$$P(Q_{st}(X)=(t-s)\sigma^2)$$

for given $0\le s<t$ (we only need to show it for $s,t\in\mathbb{Q}$, but that makes no difference in the rest of the proof). Furthermore, we show the result for $s=0$. The general result can be seen in the exact same way, using even more notation.

Define for each $n\in\mathbb{N}$

$$U_{k,n}=\sqrt{\frac{2^n}{\sigma^2}}\Big(X_{\frac{k}{2^n}}-X_{\frac{k-1}{2^n}}-\frac{1}{2^n}\xi\Big)\,.$$

For the increment in $U_{k,n}$ we have

$$X_{\frac{k}{2^n}}-X_{\frac{k-1}{2^n}}\sim\mathcal{N}\Big(\frac{1}{2^n}\xi,\frac{1}{2^n}\sigma^2\Big)$$

so each $U_{k,n} \sim \mathcal{N}(0,1)$. Furthermore for fixed n and varying k, the increments are independent, such that also $U_{1,n}, U_{2,n}, \ldots$ are independent. We can write

$$\begin{aligned}
Q_{0t}(X) &= \limsup_{n\to\infty} \sum_{k=1}^{[2^n t]} (X_{\frac{k}{2^n}} - X_{\frac{k-1}{2^n}})^2 \\
&= \limsup_{n\to\infty} \sum_{k=1}^{[2^n t]} \Big(\sqrt{\frac{\sigma^2}{2^n}} U_{k,n} + \frac{1}{2^n}\xi\Big)^2 \\
&= \limsup_{n\to\infty} \left(\sum_{k=1}^{[2^n t]} \Big(\frac{\sigma^2}{2^n} U_{k,n}^2 + 2\sqrt{\frac{\sigma^2}{2^n}} \frac{1}{2^n} U_{k,n}\xi + \frac{1}{4^n}\xi^2 \Big) \right) \\
&= \limsup_{n\to\infty} \left(\sum_{k=1}^{[2^n t]} \frac{\sigma^2}{2^n} \big(U_{k,n}^2 + \frac{2}{\sqrt{\sigma^2 2^n}} \xi U_{k,n} \big) + \frac{[2^n t]}{4^n}\xi^2 \right) .
\end{aligned}$$

Which gives

$$\begin{aligned}
&Q_{0t}(X) - t\sigma^2 \\
&= \limsup_{n\to\infty} \left(\sum_{k=1}^{[2^n t]} \frac{\sigma^2}{2^n} \big(U_{k,n}^2 + \frac{2}{\sqrt{\sigma^2 2^n}} \xi U_{k,n} \big) - t\sigma^2 + \frac{[2^n t]}{4^n}\xi^2 \right) \\
&= \limsup_{n\to\infty} \left(\sigma^2 \sum_{k=1}^{[2^n t]} \frac{1}{2^n} \big(U_{k,n}^2 + \frac{2}{\sqrt{\sigma^2 2^n}} \xi U_{k,n} - 1 \big) + \sigma^2 \Big(\frac{[2^n t]}{2^n} - t \Big) + \frac{[2^n t]}{4^n}\xi^2 \right)
\end{aligned}$$

We have that

$$t - \frac{1}{2^n} = \frac{2^n t - 1}{2^n} < \frac{[2^n t]}{2^n} \leq \frac{2^n t}{2^n} = t$$

which shows that

$$\frac{[2^n t]}{2^n} \to t \quad \text{and} \quad \frac{[2^n t]}{4^n} = \frac{1}{2^n}\frac{[2^n t]}{2^n} \to 0$$

as $n \to \infty$ So for deterministic part $Q_{0t}(X) - t\sigma^2$ it holds

$$\sigma^2 (\frac{[2^n t]}{2^n} - t) + \frac{[2^n t]}{4^n}\xi^2 \to 0$$

as $n \to \infty$. Then the proof will be complete, if we can show that $S_n \xrightarrow{\text{a.s.}} 0$ where

$$S_n = \sum_{k=1}^{[2^n t]} \frac{1}{2^n} \big(U_{k,n}^2 + \frac{2}{\sqrt{\sigma^2 2^n}} \xi U_{k,n} - 1 \big) .$$

Note that

$$ES_n = \sum_{k=1}^{[2^n t]} \frac{1}{2^n} \big(EU_{k,n}^2 + \frac{2}{\sqrt{\sigma^2 2^n}} \xi EU_{k,n} - 1 \big) = 0$$

since $EU_{k,n}^2 = 1$. According to Lemma 1.2.12 the convergence of S_n is obtained, if we can show

$$\sum_{n=1}^{\infty} P(|S_n| > \epsilon) < \infty \tag{6.9}$$

for all $\epsilon > 0$. Each term in this sum satisfies (using Chebychev's Inequality)

$$P(|S_n| > \epsilon) = P(|S_n - ES_n| > \epsilon) \leq \frac{1}{\epsilon^2} V(S_n)$$

such that the sum in (6.9) converges, if we can show

$$\sum_{n=1}^{\infty} V(S_n) < \infty \tag{6.10}$$

Using that $U_{1,n}, U_{2,n}$ are independent and identically distributed gives

$$\begin{aligned}
V(S_n) =& \sum_{k=1}^{[2^n t]} \frac{1}{4^n} V\big(U_{k,n}^2 + \frac{2}{\sqrt{\sigma^2 2^n}} \xi U_{k,n} - 1\big) \\
=& [2^n t] \frac{1}{4^n} V\big(U_{1,n}^2 + \frac{2}{\sqrt{\sigma^2 2^n}} \xi U_{1,n} - 1\big) \\
=& \frac{[2^n t]}{4^n} E\big(U_{1,n}^2 + \frac{2}{\sqrt{\sigma^2 2^n}} \xi U_{1,n} - 1\big)^2 \\
=& \frac{[2^n t]}{4^n} \Bigg(EU_{1,n}^4 + \frac{2}{\sqrt{\sigma^2 2^n}} \xi EU_{1,n}^3 - EU_{1,n}^2 + \frac{2}{\sqrt{\sigma^2 2^n}} \xi EU_{1,n}^3 + \frac{4}{\sigma^2 2^n} \xi^2 EU_{1,n}^2 \\
&+ \frac{2}{\sqrt{\sigma^2 2^n}} \xi EU_{1,n} - EU_{1,n}^2 - \frac{2}{\sqrt{\sigma^2 2^n}} \xi EU_{1,n} + 1 \Bigg)
\end{aligned}$$

and since $EU_{1,n} = 0$, $EU_{1,n}^2 = 1$, $EU_{1,n}^3 = 0$, and $EU_{1,n}^4 = 3$ we have

$$\begin{aligned}
V(S_n) &= \frac{[2^n t]}{4^n} \big(3 - 1 + \frac{4\xi^2}{\sigma^2 2^n} - 1 + 1\big) = \frac{[2^n t]}{4^n} \big(2 + \frac{4\xi^2}{\sigma^2 2^n}\big) \\
&\leq \frac{2^n t}{4^n} \big(2 + \frac{4\xi^2}{\sigma^2 2}\big) = \frac{1}{2^n} t \big(2 + \frac{4\xi^2}{\sigma^2 2}\big)
\end{aligned}$$

from which it is seen that the sum in (6.10) is finite. □

6.4 The law of the iterated logarithm

In this section we shall show a classical result concerning Brownian motion which in more detail describes the behaviour of the sample paths immediately after the start of the process.

If $X = (X_t)_{t\geq 0}$ is a continuous Brownian motion we in particular have that $\lim_{t\to 0} X_t = X_0 = 0$ a.s. For a more precise description of the behaviour near 0, we seek a function h, increasing and continuous on an interval of the form $[0, t_0)$ with $h(0) = 0$, such that

$$\limsup_{t\to 0} \frac{1}{h(t)} X_t, \quad \liminf_{t\to 0} \frac{1}{h(t)} X_t \tag{6.11}$$

are both something interesting, i.e., finite and different from 0. A good guess for a sensible h can be obtained by considering a Brownian motion without drift ($\xi = 0$), and using that then

$$\frac{1}{\sqrt{t}} X_t$$

has the same distribution for all $t > 0$ which could be taken as an indication that $\frac{1}{\sqrt{t}} X_t$ behaves sensibly for $t \to 0$. But alas, $h(t) = \sqrt{t}$ is too small, although not much of an adjustment is needed before (6.11) yields something interesting.

In the sequel we denote by ϕ the function

$$\phi(t) = \sqrt{2t \, \log\log \frac{1}{t}}, \tag{6.12}$$

which is defined and finite for $0 < t < \frac{1}{e}$. Since

$$\lim_{t\to 0} t \log\log \frac{1}{t} = \lim_{t\to\infty} \frac{\log\log t}{t} = 0$$

we have $\lim_{t\to 0} \phi(t) = 0$ so it makes sense to define $\phi(0) = 0$. Then ϕ is defined, non-negative and continuous on $[0, \frac{1}{e})$. We shall also need the useful limit

$$\lim_{x\to\infty} \frac{\int_x^\infty e^{-u^2/2} du}{\frac{1}{x} e^{-x^2/2}} = 1 \tag{6.13}$$

which follows from the following inequalities, that all hold for $x > 0$: since $\frac{u}{x} \geq 1$ for $u \geq x$ we have

$$\int_x^\infty e^{-u^2/2} \, \mathrm{d}u \leq \int_x^\infty \frac{u}{x} e^{-u^2/2} \, \mathrm{d}u = \frac{1}{x} [-e^{-u^2/2}]_x^\infty = \frac{1}{x} e^{-x^2/2}$$

and since $\frac{u}{x+1} \leq 1$ for $x \leq u \leq x+1$ we have

$$\begin{aligned}
\int_x^\infty e^{-u^2/2} du &\geq \int_x^{x+1} \frac{u}{x+1} e^{-u^2/2} du \\
&= \frac{1}{x+1} (e^{-x^2/2} - e^{-(x+1)^2/2}) \\
&= \frac{1}{x} e^{-x^2/2} \frac{x}{x+1} (1 - e^{-x-1/2}),
\end{aligned}$$

and then (6.13) follows once we note that

$$\lim_{x\to\infty} \frac{x}{x+1}(1-e^{-x-1/2}) = 1.$$

Theorem 6.4.1 (The law of the iterated logarithm)**.** *For a continuous Brownian motion $X = (X_t)_{t\geq 0}$ with drift ξ and variance $\sigma^2 > 0$, it holds that*

$$P(\limsup_{t\to 0} \frac{X_t}{\sqrt{\sigma^2}\phi(t)} = 1) = P(\liminf_{t\to 0} \frac{X_t}{\sqrt{\sigma^2}\phi(t)} = -1) = 1\,,$$

where ϕ is given by (6.12).

Proof. We show the theorem for X a continuous, normalised Brownian motion. Since

$$\lim_{t\to 0} \frac{\xi t}{\phi(t)} = 0,$$

it then holds that with X normalised,

$$\limsup_{t\to 0} \frac{\sigma X_t + \xi t}{\sqrt{\sigma^2}\phi(t)} = \limsup_{t\to 0} \frac{X_t}{\phi(t)} = 1 \text{ a.s.}$$

and similarly for lim inf. Since $(\sigma X_t + \xi t)_{t\geq 0}$ has drift ξ and variance σ^2, the theorem follows for an arbitrary Brownian motion.

In the following it is therefore assumed that X is a continuous, normalised Brownian motion. We show the theorem by showing the two following claims:

$$\limsup_{t\to 0} \frac{X_t}{\phi(t)} \leq 1+\epsilon \text{ a.s.} \qquad \text{for all } \epsilon > 0\,, \tag{6.14}$$

$$\limsup_{t\to 0} \frac{X_t}{\phi(t)} \geq 1-\epsilon \text{ a.s.} \qquad \text{for all } \epsilon > 0\,. \tag{6.15}$$

From (6.14) and (6.15) it immediately follows that

$$P\Big(\limsup_{t\to 0} \frac{X_t}{\phi(t)} = 1\Big) = 1$$

and, applying this result to the normalised Brownian motion $-X$,

$$P\Big(\liminf_{t\to 0} \frac{X_t}{\phi(t)} = -1\Big) = 1.$$

To show (6.14), let $0 < u < 1$, put $t_n = u^n$ and

$$C_{n,\epsilon,u} = \bigcup_{t\in[t_{n+1},t_n]} (X_t > (1+\epsilon)\phi(t)).$$

Since X is continuous, the union can be replaced by a countable union so $C_{n,\epsilon,u}$ is measurable. For a given $\epsilon > 0$ and $0 < u < 1$ it is seen that

$$
\begin{aligned}
(C_{n,\epsilon,u} \text{ i.o.}) &= \big(\forall n_0 \geq 1\, \exists n \geq n_0\, \exists t \in [t_{n+1}, t_n] \,:\, X_t > (1+\epsilon)\phi(t)\big) \\
&= \big(\forall n \geq 1\, \exists t \leq t_n \,:\, X_t > (1+\epsilon)\phi(t)\big) = \left(\limsup_{t\to 0} \frac{X_t}{\phi(t)} > 1+\epsilon\right)
\end{aligned}
$$

so it is thus clear that (6.14) follows if there for all $\epsilon > 0$ exists a u, $0 < u < 1$, such that

$$P(C_{n,\epsilon,u} \text{ i.o.}) = 0$$

and to deduce this, it is by the Borel-Cantelli Lemma (Lemma 1.2.11) sufficient that

$$\sum_n P(C_{n,\epsilon,u}) < \infty. \tag{6.16}$$

(Note that $C_{n,\epsilon,u}$ is only defined for n so large that $t_n = u^n \in [0, \frac{1}{e})$, the interval where ϕ is defined. In all computations we of course only consider such n)

Since the function ϕ is continous on $[0, \frac{1}{e})$ with $\phi(0) = 0$ and $\phi(t) > 0$ for $t > 0$, there exists $0 < \delta_0 < \frac{1}{e}$ such that ϕ is increasing on the interval $[0, \delta_0]$. Therefore it holds for n large (so large that $t_n \leq \delta_0$) that

$$
\begin{aligned}
P(C_{n,\epsilon,u}) &\leq P\Big(\bigcup_{t:t_{n+1}\leq t\leq t_n} (X_t > (1+\epsilon)\phi(t_{n+1}))\Big) = P\big(\sup_{t:t_{n+1}\leq t\leq t_n} X_t > (1+\epsilon)\phi(t_{n+1})\big) \\
&\leq P\big(\sup_{t:t\leq t_n} X_t > (1+\epsilon)\phi(t_{n+1})\big) \leq 2P(X_{t_n} > (1+\epsilon)\phi(t_{n+1}))\,,
\end{aligned}
$$

where we in the last inequality have used Lemma 6.2.9 and the continuity of X (which implies that $\sup_{t\leq t_n} X_t = \sup_{q\in\mathbb{Q}\cap[0,t_n]} X_q$). Since $\frac{1}{\sqrt{t_n}} X_{t_n}$ is $N(0,1)$-distributed it follows that

$$P(C_{n,\epsilon,u}) \leq \sqrt{\frac{2}{\pi}} \int_{x_n}^{\infty} e^{-s^2/2}\, ds,$$

where

$$x_n = (1+\epsilon)\frac{1}{\sqrt{t_n}}\phi(t_{n+1}) = (1+\epsilon)\sqrt{2u\ \log((n+1)\log\frac{1}{u})}.$$

We see that $x_n \to \infty$ for $n \to \infty$ and hence it holds by (6.13) that

$$\frac{\int_{x_n}^{\infty} e^{-s^2/2} ds}{\frac{1}{x_n} e^{-x_n^2/2}} \to 1\ .$$

In particular we have

$$\frac{\int_{x_n}^{\infty} e^{-s^2/2} ds}{e^{-x_n^2/2}} \to 0$$

For n large, we thus have

$$P(C_{n,\epsilon,u}) \leq \sqrt{\frac{2}{\pi}} e^{-x_n^2/2} = K\,(n+1)^{-(1+\epsilon)^2 u},$$

where $K = \sqrt{\frac{2}{\pi}}(\log \frac{1}{u})^{-(1+\epsilon)^2 u}$. If we for a given $\epsilon > 0$ choose $u < 1$ sufficiently close to 1, we obtain that $(1+\epsilon)^2 u > 1$, making

$$\sum_{n=1}^{\infty} K(n+1)^{-(1+\epsilon)^2 u} < \infty$$

Hence (6.16) and thereby also (6.14) follow.

To show (6.15), let $t_n = v^n$, where $0 < v < 1$, put $Z_n = X_{t_n} - X_{t_{n+1}}$ and define

$$D_{n,\epsilon,v} = \Big(Z_n > (1 - \frac{\epsilon}{2})\phi(t_n) \Big).$$

Note that the events $D_{n,\epsilon,v}$ for fixed ϵ and v and varying n are mutually independent.

We shall show that, given $\epsilon > 0$, there exists a v, $0 < v < 1$, such that

$$P(D_{n,\epsilon,v} \text{ i.o.}) = 1 \tag{6.17}$$

and we claim that this implies (6.15): if we apply (6.14) to $-X$, we get

$$P\Big(\liminf_{t \to 0} \frac{X_t}{\phi(t)} \geq -1 \Big) = 1,$$

so with (6.17) satisfied, we have for almost all $\omega \in \Omega$ that

$$\liminf_{t \to 0} \frac{X_t(\omega)}{\phi(t)} \geq -1$$

and that there exists a subsequence (n'), $n' \to \infty$ of natural numbers (depending on ω) such that for all n'

$$\omega \in D_{n',\epsilon,v}.$$

But then, for all n',

$$\frac{X_{t_{n'}}(\omega)}{\phi(t_{n'})} = \frac{Z_{n'}(\omega)}{\phi(t_{n'})} + \frac{X_{t_{n'+1}}(\omega)}{\phi(t_{n'})} > 1 - \frac{\epsilon}{2} + \frac{X_{t_{n'+1}}(\omega)}{\phi(t_{n'+1})} \frac{\phi(t_{n'+1})}{\phi(t_{n'})}$$

and since

$$\frac{\phi(t_{n'+1})}{\phi(t_{n'})} = \frac{\sqrt{2v^{n+1}\log\big((n+1)\log\frac{1}{v}\big)}}{\sqrt{2v^n \log\big(n \log \frac{1}{v}\big)}} = \sqrt{v}\sqrt{\frac{\log\big((n+1)\log\frac{1}{v}\big)}{\log\big(n\log\frac{1}{v}\big)}} \to \sqrt{v}$$

and furthermore

$$\liminf_{n'\to\infty}\frac{X_{t_{n'+1}}(\omega)}{\phi(t_{n'+1})}\geq\liminf_{t\to 0}\frac{X_t(\omega)}{\phi(t)},$$

we see that

$$\begin{aligned}\limsup_{t\to 0}\frac{X_t(\omega)}{\phi(t)}\geq\limsup_{n'\to\infty}\frac{X_{t_{n'}}(\omega)}{\phi(t_{n'})}&\geq 1-\frac{\epsilon}{2}+\liminf_{n'\to\infty}\frac{X_{t_{n'+1}}(\omega)}{\phi(t_{n'+1})}\sqrt{v}\\&\geq 1-\frac{\epsilon}{2}-\sqrt{v}\geq 1-\epsilon,\end{aligned}$$

if v for given $\epsilon>0$ is chosen so small that $\sqrt{v}<\frac{\epsilon}{2}$. Hence we have shown that (6.17) implies (6.15).

We still need to show (6.17). Since the $D_{n,\epsilon,v}$'s for varying n are independent, (6.17) follows from second version of the Borel-Cantelli lemma (Lemma 1.3.12) by showing that

$$\sum_n P(D_{n,\epsilon,v})=\infty. \tag{6.18}$$

We conclude the proof by showing that this may be achieved by choosing $v>0$ sufficiently small for any given $\epsilon>0$ (this was already needed to conclude that (6.17) implies (6.15)). But since Z_n is $\mathcal{N}(0,t_n-t_{n+1})$–distributed, we obtain

$$P(D_{n,\epsilon,v})=\frac{1}{\sqrt{2\pi}}\int_{y_n}^{\infty}e^{-s^2/2}ds,$$

where

$$y_n=\Big(1-\frac{\epsilon}{2}\Big)\frac{\phi(t_n)}{\sqrt{t_n-t_{n+1}}}=(1-\frac{\epsilon}{2})\sqrt{\frac{2}{1-v}\log\Big(n\log\frac{1}{v}\Big)}\,.$$

Since $y_n\to\infty$, (6.13) implies that

$$\frac{\int_{y_n}^{\infty}e^{-s^2/2}ds}{\frac{1}{y_n}e^{-y_n^2/2}}\to 1,$$

and since

$$\frac{y_n}{\sqrt{\log n}}=\text{const.}\cdot\frac{\sqrt{\log\left(n\log\frac{1}{v}\right)}}{\sqrt{\log n}}=\text{const.}\cdot\sqrt{\frac{\log n+\log\log\frac{1}{v}}{\log n}}\to\text{const.}>0\,,$$

the proof is now finished by realising that for given $\epsilon>0$ we have

$$\sum\frac{1}{\sqrt{\log n}}e^{-y_n^2/2}=\infty \tag{6.19}$$

if only v is sufficiently close to 0. But

$$e^{-y_n^2/2} = \exp\left(-\frac{\left(1-\frac{\epsilon}{2}\right)}{1-v}\log\left(\log\frac{1}{v}\right)\right) = K\,n^{-\alpha},$$

where $\alpha = \frac{(1-\epsilon/2)^2}{1-v}$ and $K = (\log\frac{1}{v})^{-\alpha}$, so $\alpha < 1$ if v is sufficiently small, and it is then a simple matter to obtain (6.19): if, e.g., $\beta > 0$ is so small that $\alpha + \beta < 1$ then the n'th term of (6.19) becomes

$$\frac{K}{\sqrt{\log n}}\frac{1}{n^{\alpha}} = K\frac{n^{\beta}}{\sqrt{\log n}}\frac{1}{n^{\alpha+\beta}} > \frac{K}{n^{\alpha+\beta}}$$

for n sufficiently large, and since $\sum n^{-(\alpha+\beta)} = \infty$ the desired conclusion follows. □

As an immediate consequence of Theorem 6.4.1 we obtain the following result concerning the number of points where the Brownian motion is zero.

Corollary 6.4.2. *If X is a continuous Brownian motion, it holds for almost all ω that for all $\epsilon > 0$, $X_t(\omega) = 0$ for infinitely many values of $t \in [0, \epsilon]$.*

Note that it trivially holds that

$$P\bigcap_{q\in\mathbb{Q}\cap(0,\infty)}(X_q \neq 0) = 1,$$

so for almost all ω there exists, for each rational $q > 0$, an open interval around q where $t \to X_t(\omega)$ does not take the value 0. In some sense therefore, $X_t(\omega)$ is only rarely 0 but it still happens an infinite number of times close to 0. □

Proof. Theorem 6.4.1 implies that for almost every ω there exist sequences $0 < s_n \downarrow 0$ and $0 < t_n \downarrow 0$ such that

$$X_{s_n}(\omega) > \frac{1}{2}\phi(s_n)\sqrt{\sigma^2} > 0, \quad X_{t_n}(\omega) < -\frac{1}{2}\phi(t_n)\sqrt{\sigma^2} < 0$$

for all n. Since X is continuous, the corollary follows. □

Our final result is an analogue to Theorem 6.4.1, with $t \to \infty$ instead of $t \to 0$. However, this result only holds for Brownian motions with drift $\xi = 0$.

Theorem 6.4.3. *If X is a continuous standard Brownian motion, then*

$$P\Big(\limsup_{t\to\infty}\frac{X_t}{\sqrt{2t\log\log t}} = 1\Big) = 1,$$

$$P\Big(\liminf_{t\to\infty}\frac{X_t}{\sqrt{2t\log\log t}} = -1\Big) = 1\,.$$

Proof. Define a new process Y by

$$Y_t = \begin{cases} t\,X_{\frac{1}{t}} & (t > 0) \\ 0 & (t = 0). \end{cases}$$

Then $t \mapsto Y_t$ is continuous on the open interval $(0,\infty)$ and for arbitrary n and $0 < t_1 < \cdots < t_n$ it is clear that $(Y_{t_1},\ldots,Y_{t_n})$ follows an n-dimensional normal distribution. Since $Y_0 = X_0 = 0$ and we for $0 < s < t$ have $EY_t = 0$ while (recall the finite–dimensional distributions of the Brownian motion)

$$\text{Cov }(Y_s, Y_t) = st\,\text{Cov }(X_{\frac{1}{s}}, X_{\frac{1}{t}}) = s,$$

it follows that Y and X have the same finite-dimensional distributions. In particular we therefore have

$$P\Big(\lim_{q\to 0, q\in\mathbb{Q}} Y_q = 0\Big) = P\Big(\lim_{q\to 0, q\in\mathbb{Q}} X_q = 0\Big) = 1\,,$$

and with Y continuous on $(0,\infty)$ we see that Y becomes continuous on $[0,\infty)$. But then the continuous process Y has the same distribution as X, and thus, Y is a continuous, normalized Brownian motion. Theorem 6.4.1 applied to Y then shows us that for instance

$$\limsup_{s\to 0}\frac{sX_{\frac{1}{s}}}{\sqrt{2s\log\log\frac{1}{s}}} = 1 \quad \text{a.s.}$$

If the s here is replaced by $\frac{1}{t}$ we obtain

$$\limsup_{t\to\infty}\frac{X_t}{\sqrt{2t\log\log t}} = 1 \quad \text{a.s.}$$

as we wanted. □

From Theorem 6.4.3 it easily follows, by an argument similar to the one we used in the proof of Corollary 6.4.2, that for $t \to \infty$ a standard Brownian motion will cross any given level $x \in \mathbb{R}$ infinitely many times.

Corollary 6.4.4. *If X is a continuous, normalized Brownian motion it holds for almost every ω that for all $T > 0$ and all $x \in \mathbb{R}$, $X_t(\omega) = x$ for infinitely many values of $t \in [T,\infty)$.*

6.5 Exercises

Exercise 6.1. Let $X = (X_t)_{t\geq 0}$ be a Brownian motion with drift $\xi \in \mathbb{R}$ and variance $\sigma^2 > 0$. Define for each $t \geq 0$

$$\tilde{X}_t = \frac{X_t - \xi t}{\sigma}$$

Show that $\tilde{X} = (\tilde{X}_t)_{t\geq 0}$ is a normalised Brownian motion. ◦

Exercise 6.2. Assume that $(\Omega, \mathcal{F}, P)$ is a probability space, and assume that $\mathcal{D}_1, \mathcal{D}_2 \subseteq \mathcal{F}$ both are ∩–stable collections of sets. Assume that $\mathcal{D}_1$ and $\mathcal{D}_2$ are independent, that is

$$P(D_1 \cap D_2) = P(D_1)P(D_2) \qquad \text{for all } D_1 \in \mathcal{D}_1, D_2 \in \mathcal{D}_2$$

Show that $\sigma(\mathcal{D}_1)$ and $\sigma(\mathcal{D}_2)$ are independent:

$$P(D_1 \cap D_2) = P(D_1)P(D_2) \qquad \text{for all } D_1 \in \sigma(\mathcal{D}_1), D_2 \in \sigma(\mathcal{D}_2) \tag{6.20}$$

◦

Exercise 6.3. Let $\mathbb{X} = (X_t)_{t\geq 0}$ be a Brownian Motion with drift ξ and variance $\sigma^2 > 0$. Define for each $t > 0$ the σ–algebra

$$\mathcal{F}_t = \mathcal{F}(X_s \;:\; 0 \leq s \leq t)$$

Show that $\mathcal{F}_t$ is independent of $\sigma(X_u - X_t)$, where $u > t$.

You can use without argument that (similarly to the arguments in the beginning of Section 6.1) $(X_s)_{0\leq s\leq t}$ has values in $(\mathbb{R}^{[0,t]}, \mathcal{B}_{[0,t]})$ where the σ–algebra $\mathcal{B}_{[0,t]}$ is generated by

$$\mathbb{G} = \{\big((\hat{X}_{t_1}, \dots, \hat{X}_{t_n}) \in B_n\big) \;:\; n \in \mathbb{N}, 0 \leq t_1 < \cdots < t_n \leq t, B_n \in \mathcal{B}_n\}$$

Then $\mathcal{F}_t$ must be generated by the pre images of these sets

$$\begin{aligned}\mathbb{D} &= \{((X_s)_{0\leq s\leq t})^{-1}(G) \;:\; G \in \mathbb{G}\}\\ &= \{\big((X_{t_1}, \dots, X_{t_n}) \in B_n\big) \;:\; n \in \mathbb{N}, 0 \leq t_1 < \cdots < t_n \leq t, B_n \in \mathbb{B}^n\}\end{aligned}$$

◦

Exercise 6.4. Assume that $X = (X_t)_{t\geq 0}$ is a Brownian Motion with drift $\xi \in \mathbb{R}$ and variance $\sigma^2 > 0$. Define $\mathcal{F}_t$ as in Exercise 6.3. Show that X has the following Markov property for $t \geq s$

$$E(X_t|\mathcal{F}_s) = E(X_t|X_s) \quad \text{a.s.}$$

◦

Exercise 6.5. Assume that $X = (X_t)_{t\geq 0}$ is a normalised Brownian motion and define $\mathcal{F}_t$ as in Exercise 6.3 for each $t \geq 0$. Show that

(1) X_t is $\mathcal{F}_t$ measurable for all $t \geq 0$.

(2) $E|X_t| < \infty$ for all $t \geq 0$

(3) $E(X_t|\mathcal{F}_s) = X_s$ a.s. for all $0 \leq s < t$.

We say that $(X_t, \mathcal{F}_t)_{t\geq 0}$ is a *martingale* in continuous time. ◦

Exercise 6.6. Let $X = (X_t)_{t\geq 0}$ be a normalised Brownian motion. Let $T > 0$ be fixed and define the process $B^T = (B_t^T)_{0\leq t\leq T}$ by

$$B_t^T = X_t - \frac{t}{T} X_T$$

The process B^T is called a *Brownian bridge* on $[0, T]$.

1) Show that for all $0 < t_1 < \cdots < t_n < T$ then

$$(B_{t_1}^T, \ldots, B_{t_n}^T)$$

is n–dimensional normally distributed. And find for $0 < s < t < T$

$$EB_s^T \qquad \text{and} \qquad \mathrm{Cov}(B_s^T, B_t^T)$$

2) Show that for all $T > 0$ then

$$(B_{Tt}^T)_{0\leq t\leq 1} \stackrel{\mathcal{D}}{=} (\sqrt{T} B_t^1)_{0\leq t\leq 1}$$

◦

Exercise 6.7. A stochastic process $X = (X_t)_{t\geq 0}$ is self–similar if there exists $H > 0$ such that

$$(X_{\gamma t})_{t\geq 0} \stackrel{\mathcal{D}}{=} (\gamma^H X_t)_{t\geq 0} \qquad \text{for all } \gamma > 0\,.$$

intuitively, this means that if we "zoom in" on the process, then it looks like a scaled version of the original process.

(1) Assume that $X = (X_t)_{t\geq 0}$ is a Brownian motion with drift 0 and variance σ^2. Show that X is self–similar and find the parameter H.

Now assume that $X = (X_t)_{t\geq 0}$ is a stochastic process (defined on $(\Omega, \mathcal{F}, P)$) that is self–similar with parameter $0 < H < 1$. Assume that X has stationary increments:

$$X_t - X_s \overset{\mathcal{D}}{=} X_{t-s} \qquad \text{for all } 0 \leq s \leq t\,.$$

Assume furthermore that $P(X_0 = 0) = 1$ and $P(X_1 = 0) = 0$.

(2) Show that for all $0 \leq s < t$

$$\frac{X_t - X_s}{t - s} \overset{\mathcal{D}}{=} (t-s)^{H-1} X_1\,.$$

(3) Show that $P(X_t = 0) = 0$ for all $t > 0$.

(4) Show that X is continuous in probability.

$\circ$

Exercise 6.8. Let Y be a normalised Brownian motion with continuous sample paths. Hence Y can be considered as a random variable with values in $(C_{[0,\infty)}, \mathcal{C}_{[0,\infty)})$. Define for all $n, M \in \mathbb{N}$ the set

$$C_{n,M} = \left\{ x \in C_{[0,\infty)} \;\Big|\; \sup_{t\in[n,n+1]} \frac{|x_t|}{t} > \frac{1}{M} \right\}$$

(1) Show that $C_{n,M} \in \mathcal{C}_{[0,\infty)}$.

(2) Show that

$$P(Y \in C_{n,M}) \leq 2P\big(Y_{n+1} > \frac{n}{M}\big)$$

(3) Show that

$$P\big(Y_{n+1} > \frac{n}{M}\big) \leq \frac{3(n+1)^2 M^4}{n^4}$$

and conclude that

$$\sum_{n=1}^{\infty} P(Y \in C_{n,M}) < \infty\,.$$

(4) Show that for all $M \in \mathbb{N}$

$$P\left(\Big(\sup_{t\in[n,n+1]} \frac{|Y_t|}{t} > \frac{1}{M}\Big) \text{ i.o. }\right) = 0\,.$$

(5) Show that

$$P\bigcap_{M=1}^{\infty}\left(\Big(\sup_{t\in[n,n+1]} \frac{|Y_t|}{t} \leq \frac{1}{M}\Big) \text{ evt. }\right) = 1\,.$$

(6) Show that

$$\frac{Y_t}{t} \xrightarrow{\text{a.s.}} 0 \quad \text{as } t \to \infty\,.$$

∘

Chapter 7

Further reading

In this final chapter, we reflect on the theory presented in the previous chapters and give recommendations for further reading.

The material covered in Chapter 1 can be found scattered in many textbooks on probability theory, such as Breiman (1968), Loève (1977a), Kallenberg (2002) and Rogers & Williams (2000a).

Abstract results in ergodic theory as presented in Chapter 2 can be found in Breiman (1968) and Loève (1977b). A major application of ergodic theory is to the theory of the class of stochastic processes known as Markov processes. In Markov process theory, stationary processes are frequently encountered, and thus the theory presents an opportunity for utilizing the ergodic theorem for stationary processes. Basic introductions to Markov processes in both discrete and continuous time can be found in Norris (1999) and Brémaud (1999). In Meyn & Tweedie (2009), a more general theory is presented, which includes a series of results on ergodic Markov processes.

In Chapter 3, we dealt with weak convergence. In its most general form, weak convergence of probability measures can be cast in the context of probability measures on complete, separable metric spaces, where the metric space considered is endowed with the Borel-σ-algebra generated by the open sets. A classical exposition of this theory is found in Billingsley (1999), with Parthasarathy (1967) also being a useful resource.

Being one of the cornerstones of modern probability, the discrete-time martingale theory of Chaper 5 can be found in many textbooks. A classical source is Rogers & Williams (2000a).

The results on Brownian motion in Chapter 6 represents an introduction to the problems and results of the theory of continuous-time stochastic processes. This very large subject encompasses many branches, prominent among them are continuous-time martingale theory, stochastic integration theory and continuous-time Markov process theory, to name a few. A good introduction to several of the major themes can be found in Rogers & Williams (2000a) and Rogers & Williams (2000b). Karatzas & Shreve (1988) focuses on martingales with continuous paths and the theory of stochastic integration. A solid introduction to continuous-time Markov processes is Ethier & Kurtz (1986).

Appendix A

Supplementary material

In this chapter, we outline results which are either assumed to be well-known, or which are of such auxiliary nature as to merit separation from the main text.

A.1 Limes superior and limes inferior

In this section, we recall some basic results on the supremum and infimum of a set in the extended real numbers, as well as the limes superior and limes inferior of a sequence in $\mathbb{R}$. By $\mathbb{R}^*$, we denote the set $\mathbb{R} \cup \{-\infty, \infty\}$, and endow $\mathbb{R}^*$ with its natural ordering, in the sense that $-\infty < x < \infty$ for all $x \in \mathbb{R}$. We refer to $\mathbb{R}^*$ as the extended real numbers. In general, working with $\mathbb{R}^*$ instead of merely $\mathbb{R}$ is useful, although somewhat technically inconvenient from a formal point of view.

Definition A.1.1. *Let $A \subseteq \mathbb{R}^*$. We say that $y \in \mathbb{R}^*$ is an upper bound for A if it holds for all $x \in A$ that $x \leq y$. Likewise, we say that $y \in \mathbb{R}^*$ is a lower bound for A if it holds for all $x \in A$ that $y \leq x$.*

Theorem A.1.2. *Let $A \subseteq \mathbb{R}^*$. There exists a unique element $\sup A \in \mathbb{R}^*$ characterized by that $\sup A$ is an upper bound for A, and for any upper bound y for A, $\sup A \leq y$. Likewise, there exists a unique element $\inf A \in \mathbb{R}^*$ characterized by that $\inf A$ is a lower bound for A, and for any lower bound y for A, $y \leq \inf A$.*

Proof. See Theorem C.3 of Hansen (2009). □

The elements $\sup A$ and $\inf A$ whose existence and uniqueness are stated in Theorem A.1.2 are known as the supremum and infimum of A, respectively, or as the least upper bound and greatest lower bound of A, respectively.

In general, the formalities regarding the distinction between $\mathbb{R}$ and $\mathbb{R}^*$ are necessary to keep in mind when concerned with formal proofs, however, in practice, the supremum and infimum of a set in $\mathbb{R}^*$ is what one expects it to be: For example, the supremum of $A \subseteq \mathbb{R}^*$ is infinity precisely if A contains "arbitrarily large elements", otherwise it is the "upper endpoint" of the set, and similarly for the infimum.

The following yields useful characterisations of the supremum and infimum of a set when the supremum and infimum is finite.

Lemma A.1.3. *Let $A \subseteq \mathbb{R}^*$ and let $y \in \mathbb{R}$. Then y is the supremum of A if and only if the following two properties hold:*

(1). y is an upper bound for A.

(2). For each $\varepsilon > 0$, there exists $x \in A$ such that $y - \varepsilon < x$.

Likewise, y is the infimum of A if and only if the following two properties hold:

(1). y is a lower bound for A.

(2). For each $\varepsilon > 0$, there exists $x \in A$ such that $x < y + \varepsilon$.

Proof. We just prove the result on the supremum. Assume that y is the supremum of A. By definition, y is then an upper bound for A. Let $\varepsilon > 0$. If $y - \varepsilon$ were an upper bound for A, we would have $y \leq y - \varepsilon$, a contradiction. Therefore, $y - \varepsilon$ is not an upper bound for A, and so there exists $x \in A$ such that $y - \varepsilon < x$. This proves that the two properties are necessary for y to be the supremum of A.

To prove the converse, assume that the two properties hold, we wish to show that y is the supremum of A. By our assumptions, y is an upper bound for A, so it suffices to show that for any upper bound $z \in \mathbb{R}^*$, we have $y \leq z$. To obtain this, note that by the second of our assumptions, A is nonempty. Therefore, $-\infty$ is not an upper bound for A. Thus, it suffices to consider an upper bound $z \in \mathbb{R}$ and prove that $y \leq z$. Letting z be such an upper bound, assume that $z < y$ and put $\varepsilon = y - z$. There then exists $x \in A$ such that $z = y - \varepsilon < x$. This shows that z is not an upper bound for A, a contradiction. We conclude that for any upper bound z of A, it must hold that $y \leq z$. Therefore, y is the supremum of A, as desired. □

We also have the following useful results.

Lemma A.1.4. *Let* $A, B \subseteq \mathbb{R}^*$. *If* $A \subseteq B$, $\sup A \leq \sup B$ *and* $\inf B \leq \inf A$.

Proof. See Lemma C.4 of Hansen (2009). □

Lemma A.1.5. *Let* $A \subseteq \mathbb{R}^*$ *and assume that* A *is nonempty. Then* $\inf A \leq \sup A$.

Proof. See Lemma C.5 of Hansen (2009). □

Lemma A.1.6. *Let* $A \subseteq \mathbb{R}^*$. *Put* $-A = \inf\{-x \mid x \in A\}$. *Then* $-\sup A = \inf(-A)$ *and* $-\inf A = \sup(-A)$.

Proof. See p. 4 of Carothers (2000). □

A particular result which we will be of occasional use to us is the following.

Lemma A.1.7. *Let* $A \subseteq \mathbb{R}^*$, *and let* $y \in \mathbb{R}$. *Then* $\sup A > y$ *if and only if there exists* $x \in A$ *with* $x > y$. *Analogously,* $\inf A < y$ *if and only if there exists* $x \in A$ *with* $x < y$.

Proof. We prove the result on the supremum. Assume that $\sup A > y$. If $\sup A$ is infinite, A is not bounded from above, and so there exists arbitrarily large elements in A, in particular there exists $x \in A$ with $x > y$. If $\sup A$ is finite, Lemma A.1.3 shows that with $\varepsilon = \sup A - y$, there exists $x \in A$ such that $y = \sup A - \varepsilon < x$. This proves that if $\sup A > y$, there exists $x \in A$ with $x > y$. Conversely, if there is $x \in A$ with $x > y$, we also obtain $y < x \leq \sup A$, since $\sup A$ is an upper bound for A. This proves the other implication. □

Note that the result of Lemma A.1.7 is false if the strict inequalities are exchanged with inequalities. For example, $\sup[0, 1) \geq 1$, but there is no $x \in [0, 1)$ with $x \geq 1$. Next, we turn our attention to sequences.

Definition A.1.8. *Let* (x_n) *be a sequence in* $\mathbb{R}$. *We define*

$$\limsup_{n\to\infty} x_n = \inf_{n\geq 1} \sup_{k\geq n} x_k$$
$$\liminf_{n\to\infty} x_n = \sup_{n\geq 1} \inf_{k\geq n} x_k,$$

and refer to $\limsup_{n\to\infty} x_n$ *and* $\liminf_{n\to\infty} x_n$ *as the limes superior and limes inferior of* (x_n), *respectively.*

The limes superior and limes inferior are useful tools for working with sequences and in particular for proving convergence.

Lemma A.1.9. *Let* (x_n) *be a sequence in* $\mathbb{R}$. *Then* $\liminf_{n\to\infty} x_n \leq \limsup_{n\to\infty} x_n$.

Proof. See Lemma C.11 of Hansen (2009). □

Theorem A.1.10. *Let* (x_n) *be a sequence in* $\mathbb{R}$, *and let* $c \in \mathbb{R}^*$. x_n *converges to* c *if and only if* $\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n = c$. *In particular,* (x_n) *is convergent to a finite limit if and only if the limes inferior and limes superior are finite and equal, and in the affirmative, the limit is equal to the common value of the limes inferior and the limes superior.*

Proof. See Theorem C.15 and Theorem C.16 of Hansen (2009). □

Corollary A.1.11. *Let* (x_n) *be a sequence of nonnegative numbers. Then* x_n *converges to zero if and only if* $\limsup_{n\to\infty} x_n = 0$.

Proof. By Theorem A.1.10, it holds that $\limsup_{n\to\infty} x_n = 0$ if x_n converges to zero. Conversely, assume that $\limsup_{n\to\infty} x_n = 0$. As zero is a lower bound for (x_n), we find $0 \leq \liminf_{n\to\infty} x_n \leq \limsup_{n\to\infty} x_n = 0$, so Theorem A.1.10 shows that x_n converges to zero. □

We will often use Corollary A.1.11 to show various kinds of convergence results. We also have the following useful results for the practical manipulation of expressions involving the limes superior and limes inferior.

Lemma A.1.12. *Let* (x_n) *and* (y_n) *be sequences in* $\mathbb{R}$. *Given that all the sums are well-defined, the following holds.*

$$\liminf_{n\to\infty} x_n + \liminf_{n\to\infty} y_n \leq \liminf_{n\to\infty}(x_n + y_n). \tag{A.1}$$

$$\limsup_{n\to\infty}(x_n + y_n) \leq \limsup_{n\to\infty} x_n + \limsup_{n\to\infty} y_n. \tag{A.2}$$

Furthermore, if (y_n) *is convergent with limit in* $\mathbb{R}$, *it holds that*

$$\liminf_{n\to\infty}(x_n + y_n) = \liminf_{n\to\infty} x_n + \lim_{n\to\infty} y_n. \tag{A.3}$$

$$\limsup_{n\to\infty}(x_n + y_n) = \limsup_{n\to\infty} x_n + \lim_{n\to\infty} y_n. \tag{A.4}$$

Also, we always have

$$-\limsup_{n\to\infty} x_n = \liminf_{n\to\infty}(-x_n). \tag{A.5}$$

$$-\liminf_{n\to\infty} x_n = \limsup_{n\to\infty}(-x_n). \tag{A.6}$$

If $x_n \leq y_n$, *it holds that*

$$\liminf_{n\to\infty} x_n \leq \liminf_{n\to\infty} y_n. \tag{A.7}$$

$$\limsup_{n\to\infty} x_n \leq \limsup_{n\to\infty} y_n. \tag{A.8}$$

Proof. The relationships in (A.1) and (A.2) are proved in Lemma C.14 of Hansen (2009). Considering (A.3), let y be the limit of (y_n) and let $m \geq 1$ be so large that $y - \varepsilon \leq y_n \leq y + \varepsilon$ for $n \geq m$. For such n, we then have

$$(\inf_{k\geq n} x_n) + y - \varepsilon = \inf_{k\geq n}(x_n + y - \varepsilon) \leq \inf_{k\geq n}(x_n + y_n) \leq \inf_{k\geq n}(x_n + y + \varepsilon) = (\inf_{k\geq n} x_n) + y + \varepsilon,$$

yielding $\liminf_{n\to\infty} x_n + y - \varepsilon \leq \liminf_{n\to\infty}(x_n + y_n) \leq \liminf_{n\to\infty} x_n + y + \varepsilon$. As $\varepsilon > 0$ was arbitrary, this yields (A.3). By a similar argument, we obtain (A.4). Furthermore, (A.5) and (A.6) follow from Lemma A.1.6. The relationships (A.7) and (A.8) are proved in Lemma C.12 of Hansen (2009). □

A.2 Measure theory and real analysis

In this section, we recall some of the main results from measure theory and real analysis which will be needed in the following. We first recall some results from basic measure theory, see Hansen (2009) for a general exposition.

Definition A.2.1. *Let E be a set. Let $\mathcal{E}$ be a collection of subsets of E. We say that $\mathcal{E}$ is a σ-algebra on E if it holds that $E \in \mathcal{E}$, that if $A \in \mathcal{E}$, then $A^c \in \mathcal{E}$, and that if $(A_n)_{n\geq 1} \subseteq \mathcal{E}$, then $\cup_{n=1}^{\infty} A_n \in \mathcal{E}$.*

We say that a pair $(E, \mathcal{E})$, where E is some set and $\mathcal{E}$ is a σ-algebra on E, is a measurable space. Also, if $\mathbb{H}$ is some set of subsets of E, we define $\sigma(\mathbb{H})$ to be the smallest σ-algebra containing $\mathbb{H}$, meaning that $\mathbb{H}$ is the intersection of all σ-algebras on E containing $\mathbb{H}$. For a σ-algebra $\mathcal{E}$ on E and a family $\mathbb{H}$ of subsets of E, we say that $\mathbb{H}$ is a generating family for $\mathcal{E}$ if $\mathcal{E} = \sigma(\mathbb{H})$. One particular example of this is the Borel σ-algebra $\mathcal{B}_A$ on $A \subseteq \mathbb{R}^n$, which is the smallest σ-algebra on A containing all open sets in A. In particular, we denote by $\mathcal{B}_n$ the Borel σ-algebra on $\mathbb{R}^n$.

If it holds for all $A, B \in \mathbb{H}$ that $A \cap B \in \mathbb{H}$, we say that $\mathbb{H}$ is stable under finite intersections. Also, if $\mathbb{D}$ is a family of subsets of E, we say that $\mathbb{D}$ is a Dynkin class if it satisfies the following requirements: $E \in \mathbb{D}$, if $A, B \in \mathbb{D}$ with $A \subseteq B$ then $B \setminus A \in \mathbb{D}$, and if $(A_n) \subseteq \mathbb{D}$ with $A_n \subseteq A_{n+1}$ for all $n \geq 1$, then $\cup_{n=1}^{\infty} A_n \in \mathbb{D}$. We have the following useful result.

Lemma A.2.2 (Dynkin's lemma)**.** *Let $\mathbb{D}$ be a Dynkin class on E, and let $\mathbb{H}$ be a set of subsets of E which is stable under finite intersections. If $\mathbb{H} \subseteq \mathbb{D}$, then $\sigma(\mathbb{H}) \subseteq \mathbb{D}$.*

Proof. See Theorem 3.6 of Hansen (2009), or Theorem 4.1.2 of Ash (1972). □

Definition A.2.3. *Let $(E, \mathcal{E})$ be a measurable space. We say that a function $\mu : \mathcal{E} \to [0, \infty)$ is a measure, if it holds that $\mu(\emptyset) = 0$ and that whenever $(A_n) \subseteq \mathcal{E}$ is a sequence of pairwise disjoint sets, $\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$.*

We say that a triple $(E, \mathcal{E}, \mu)$ is a measure space. Also, if there exists an increasing sequence of sets $(E_n) \subseteq \mathcal{E}$ with $E = \cup_{n=1}^{\infty} E_n$ and such that $\mu(E_n)$ is finite, we say that μ is σ-finite and refer to $(E, \mathcal{E}, \mu)$ as a σ-finite measure space. If $\mu(E)$ is finite, we say that μ is finite, and if $\mu(E) = 1$, we say that μ is a probability measure. In the latter case, we refer to $(E, \mathcal{E}, \mu)$ as a probability space. An important application of Lemma A.2.2 is the following.

Theorem A.2.4 (Uniqueness theorem for probability measures)**.** *Let P and Q be two probability measures on $(E, \mathcal{E})$. Let $\mathbb{H}$ be a generating family for $\mathcal{E}$ which is stable under finite intersections. If $P(A) = Q(A)$ for all $A \in \mathbb{H}$, then $P(A) = Q(A)$ for all $A \in \mathcal{E}$.*

Proof. See Theorem 3.7 in Hansen (2009). □

Next, we consider measurable mappings.

Definition A.2.5. *Let $(E, \mathcal{E})$ and $(F, \mathcal{F})$ be two measurable spaces. Let $f : E \to F$ be some mapping. We say that f is $\mathcal{E}$-$\mathcal{F}$ measurable if $f^{-1}(A) \in \mathcal{E}$ whenever $A \in \mathcal{F}$.*

For a family of mappings $(f_i)_{i \in I}$ from E to F_i, where $(F_i, \mathcal{F}_i)$ is some measurable space, we may introduce $\sigma((f_i)_{i \in I})$ as the smallest σ-algebra $\mathcal{E}$ on E such that all the f_i are $\mathcal{E}$-$\mathcal{F}_i$ measurable. Formally, $\mathcal{E}$ is the σ-algebra generated by $\{(f_i \in A) \mid i \in I, A \in \mathcal{F}_i\}$. For measurability with respect to such σ-algebras, we have the following very useful lemma.

Lemma A.2.6. *Let E be a set, let $(f_i)_{i \in I}$ be a family of mappings from E to F_i, where $(F_i, \mathcal{F}_i)$ is some measurable space, and let $\mathcal{E} = \sigma((f_i)_{i \in I})$. Let $(H, \mathcal{H})$ be some other measurable space, and let $g : H \to E$. Then g is $\mathcal{H}$-$\mathcal{E}$ measurable if and only if $f_i \circ g$ is $\mathcal{H}$-$\mathcal{F}_i$ measurable for all $i \in I$.*

Proof. See Lemma 4.14 of Hansen (2009) for a proof in the case of a single variable. □

If $f : E \to \mathbb{R}$ is $\mathcal{E}$-$\mathcal{B}$ measurable, we say that f is Borel measurable. In the context of probability spaces, we refer to Borel measurable mappings as random variables. For any measure space $(E, \mathcal{E}, \mu)$ and any Borel measurable mapping $f : E \to [0, \infty]$, the integral $\int f \, \mathrm{d}\mu$ is well-defined as the supremum of the explicitly constructed integral of an appropriate class of simpler mappings. If instead we consider some $f : E \to \mathbb{R}$, the integral $\int f \, \mathrm{d}\mu$ is well-defined as the difference between the integrals of the positive and negative parts of f whenever $\int |f| \, \mathrm{d}\mu$ is finite. The integral has the following important properties.

Theorem A.2.7 (The monotone convergence theorem)**.** *Let $(E, \mathcal{E}, \mu)$ be a measure space, and let (f_n) be a sequence of measurable mappings $f_n : E \to [0, \infty]$. Assume that the sequence (f_n) is increasing μ-almost everywhere. Then*

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

Proof. See Theorem 6.12 in Hansen (2009). □

Lemma A.2.8 (Fatou's lemma)**.** *Let $(E, \mathcal{E}, \mu)$ be a measure space, and let (f_n) be a sequence of measurable mappings $f_n : E \to [0, \infty]$. It holds that*

$$\int \liminf_{n \to \infty} f_n \, \mathrm{d}\mu \le \liminf_{n \to \infty} \int f_n \, \mathrm{d}\mu.$$

Proof. See Lemma 6.25 in Hansen (2009). □

Theorem A.2.9 (The dominated convergence theorem). *Let $(E, \mathcal{E}, \mu)$ be a measure space, and let (f_n) be a sequence of measurable mappings from E to $\mathbb{R}$. Assume that the sequence (f_n) converges μ-almost everywhere to some mapping f. Assume that there exists a measurable, integrable mapping $g : E \to [0, \infty)$ such that $|f_n| \leq g$ μ-almost everywhere for all n. Then f_n is integrable for all $n \geq 1$, f is measurable and integrable, and*

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

Proof. See Theorem 7.6 in Hansen (2009). □

For the next result, recall that for two σ-finite measure spaces $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$, $\mathcal{E} \otimes \mathcal{F}$ denotes the σ-algebra on $E \times F$ generated by $\{A \times B \mid A \in \mathcal{E}, B \in \mathcal{F}\}$, and $\mu \otimes \nu$ denotes the unique σ-finite measure such that $(\mu \otimes \nu)(A \times B) = \mu(A)\nu(B)$ for $A \in \mathcal{E}$ and $B \in \mathcal{F}$, see Chapter 9 of Hansen (2009).

Theorem A.2.10 (Tonelli's theorem). *Let $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$ be two σ-finite measure spaces, and assume that f is nonnegative and $\mathcal{E} \otimes \mathcal{F}$ measurable. Then*

$$\int f(x, y) \, \mathrm{d}(\mu \otimes \nu)(x, y) = \int \int f(x, y) \, \mathrm{d}\nu(y) \, \mathrm{d}\mu(x).$$

Proof. See Theorem 9.4 of Hansen (2009). □

Theorem A.2.11 (Fubini's theorem). *Let $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$ be two σ-finite measure spaces, and assume that f is $\mathcal{E} \otimes \mathcal{F}$ measurable and $\mu \otimes \nu$ integrable. Then $y \mapsto f(x, y)$ is integrable with respect to ν for μ-almost all x, the set where this is the case is measurable, and it holds that*

$$\int f(x, y) \, \mathrm{d}(\mu \otimes \nu)(x, y) = \int \int f(x, y) \, \mathrm{d}\nu(y) \, \mathrm{d}\mu(x).$$

Proof. See Theorem 9.10 of Hansen (2009). □

Theorem A.2.12 (Jensen's inequality). *Let $(E, \mathcal{E}, \mu)$ be a probability space. Let $X : E \to \mathbb{R}$ be a Borel mapping. Let $f : \mathbb{R} \to \mathbb{R}$ be another Borel mapping. Assume that X and $f(X)$ are integrable and that f is convex. Then $f(\int X \, \mathrm{d}\mu) \leq \int f(X) \, \mathrm{d}\mu$.*

Proof. See Theorem 16.31 in Hansen (2009). □

Theorem A.2.7, Lemma A.2.8 and Theorem A.2.9 are the three main tools for working with integrals. Theorem A.2.12 is frequently useful as well, and can in purely probabilistic terms be stated as the result that $f(EX) \leq Ef(X)$ when f is convex.

Also, for measurable spaces $(E, \mathcal{E})$ and $(F, \mathcal{F})$, μ a measure on $(E, \mathcal{E})$ and $t : E \to F$ an $\mathcal{E}$-$\mathcal{F}$ measurable mapping, we define the image measure $t(\mu)$ as the measure on $(F, \mathcal{F})$ given by putting $t(\mu)(A) = \mu(t^{-1}(A))$ for $A \in \mathcal{F}$. We then have the following theorem on successive transformations.

Theorem A.2.13. *Let $(E, \mathcal{E})$, $(F, \mathcal{F})$ and $(G, \mathcal{G})$ be measurable spaces. Let μ be a measure on $(E, \mathcal{E})$. Let $t : E \to F$ and $s : F \to G$ be measurable. Then $s(t(\mu)) = (s \circ t)(\mu)$.*

Proof. See Theorem 10.2 of Hansen (2009). □

The following abstract change-of-variable formula also holds.

Theorem A.2.14. *Let $(E, \mathcal{E}, \mu)$ be a measurable space and let $(F, \mathcal{F})$ be some measure space. Let $t : E \to F$ be measurable, and let $f : F \to \mathbb{R}$ be Borel measurable. Then f is $t(\mu)$-integrable if and only if $f \circ t$ is μ-integrable, and in the affirmative, it holds that $\int f \,\mathrm{d}t(\mu) = \int f \circ t \,\mathrm{d}\mu$.*

Proof. See Corollary 10.9 of Hansen (2009). □

Next, we recall some results on $\mathcal{L}^p$ spaces.

Definition A.2.15. *Let $(E, \mathcal{E}, \mu)$ be a measurable space, and let $p \geq 1$. By $\mathcal{L}^p(E, \mathcal{E}, \mu)$, we denote the set of measurable mappings $f : E \to \mathbb{R}$ such that $\int |f|^p \,\mathrm{d}\mu$ is finite.*

We endow $\mathcal{L}^p(E, \mathcal{E}, \mu)$ with the norm $\|\cdot\|_p$ given by $\|f\|_p = (\int |f|^p \,\mathrm{d}\mu)^{1/p}$. That $\mathcal{L}^p(E, \mathcal{E}, \mu)$ is a vector space and that $\|\cdot\|_p$ is a seminorm on this space is a consequence of the Minkowski inequality, see Theorem 2.4.7 of Ash (1972). We refer to $\mathcal{L}^p(E, \mathcal{E}, \mu)$ as an $\mathcal{L}^p$-space. For $\mathcal{L}^p$-spaces, the following two main results hold.

Theorem A.2.16 (Hölder's inequality)**.** *Let $p > 1$ and let q be the dual exponent to p, meaning that $q > 1$ is uniquely determined as the solution to the equation $\frac{1}{p} + \frac{1}{q} = 1$. If $f \in \mathcal{L}^p(E, \mathcal{E}, \mu)$ and $g \in \mathcal{L}^q(E, \mathcal{E}, \mu)$, it holds that $fg \in \mathcal{L}^1(E, \mathcal{E}, \mu)$, and $\|fg\|_1 \leq \|f\|_p \|g\|_q$.*

Proof. See Theorem 2.4.5 of Ash (1972). □

Theorem A.2.17 (The Riesz-Fischer completeness theorem)**.** *Let $p \geq 1$. The seminormed vector space $\mathcal{L}^p(E, \mathcal{E}, \mu)$ is complete.*

Proof. See Theorem 2.4.11 of Ash (1972). □

Following these results, we recall a simple lemma which we make use of in the proof of the law of large numbers.

Lemma A.2.18. *Let (x_n) be some sequence in $\mathbb{R}$, and let x be some element of $\mathbb{R}$. If $\lim_{n\to\infty} x_n = x$, then $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} x_k = x$ as well.*

Proof. See Lemma 15.5 of Carothers (2000). □

Also, we recall some properties of the integer part function. For any $x \in \mathbb{R}$, we define $[x] = \sup\{n \in \mathbb{Z} \mid n \leq x\}$.

Lemma A.2.19. *It holds that $[x]$ is the unique integer such that $[x] \leq x < [x] + 1$, or equivalently, the unique integer such that $x - 1 < [x] \leq x$.*

Proof. We first show that $[x]$ satisfies the bounds $[x] \leq x < [x] + 1$. As x is an upper bound for the set $\{n \in \mathbb{Z} \mid n \leq x\}$, and $[x]$ is the least upper bound, we obtain $[x] \leq x$. On the other hand, as $[x]$ is an upper bound for $\{n \in \mathbb{Z} \mid n \leq x\}$, $[x] + 1$ cannot be an element of this set, yielding $x < [x] + 1$.

This shows that $[x]$ satisfies the bounds given. Now assume that m is an integer satisfying $m \leq x < m + 1$, we claim that $m = [x]$. As $m \leq x$, we obtain $m \leq [x]$. And as $x < m + 1$, $m + 1$ is not in $\{n \in \mathbb{Z} \mid n \leq x\}$. In particular, for all $n \leq x$, $n < m + 1$. As $[x] \leq x$, this yields $[x] < m + 1$, and so $[x] \leq m$. We conclude that $m = [x]$, as desired. □

Lemma A.2.20. *Let $x \in \mathbb{R}$ and let $n \in \mathbb{Z}$. Then $[x + n] = [x] + n$.*

Proof. As $[x] \leq x < [x] + 1$ is equivalent to $[x] + n \leq x + n < [x] + n + 1$, the characterization of Lemma A.2.19 yields the result. □

Finally, we state Taylor's theorem with Lagrange form of the remainder.

Theorem A.2.21. *Let $n \geq 1$ and assume that f is n times differentiable, and let $x, y \in \mathbb{R}^p$. It then holds that*

$$f(y) = \sum_{k=0}^{n-1} \frac{f^{(k)}(x)}{n!}(y-x)^k + \frac{f^{(n)}(\xi(x,y))}{n!}(y-x)^n,$$

where $f^{(k)}$ denotes the k'th derivative of f, with the convention that $f^{(0)} = f$, and $\xi(x,y)$ is some element on the line segment between x and y.

Proof. See Apostol (1964) Theorem 7.6. □

A.3 Existence of sequences of random variables

In this section, we state a result which yield the existence of particular types of sequences of random variables.

Theorem A.3.1 (Kolmogorov's consistency theorem)**.** *Let $(Q_n)_{n\geq 1}$ be a sequence of probability measures such that Q_n is a probability measure on $(\mathbb{R}^n, \mathcal{B}_n)$. For each $n \geq 2$, let $\pi_n : \mathbb{R}^n \to \mathbb{R}^{n-1}$ denote the projection onto the first $n-1$ coordinates. Assume that $\pi_n(Q_n) = Q_{n-1}$ for all $n \geq 2$. There exists a probability space $(\Omega, \mathcal{F}, P)$ and a sequence of random variables $(X_n)_{n\geq 1}$ on $(\Omega, \mathcal{F}, P)$ such that for all $n \geq 1$, $(X_1, \ldots, X_n)$ have distribution Q_n.*

Proof. This follows from Theorem II.30.1 of Rogers & Williams (2000a). □

Corollary A.3.2. *Let $(Q_n)_{n\geq 1}$ be a sequence of probability measures on $(\mathbb{R}, \mathcal{B})$. There exists a probability space $(\Omega, \mathcal{F}, P)$ and a sequence of random variables $(X_n)_{n\geq 1}$ on $(\Omega, \mathcal{F}, P)$ such that for all $n \geq 1$, $(X_1, \ldots, X_n)$ are independent, and X_n has distribution Q_n.*

Proof. This follows from applying Theorem A.3.1 with the sequence of probability measures $(Q_1 \otimes \cdots \otimes Q_n)_{n\geq 1}$. □

From Corollary A.3.2, it follows for example that there exists a probability space $(\Omega, \mathcal{F}, P)$ and a sequence of independent random variables $(X_n)_{n\geq 1}$ on $(\Omega, \mathcal{F}, P)$ such that X_n is distributed on $\{0,1\}$ with $P(X_n = 1) = p_n$, where (p_n) is some sequence in $[0,1]$. Such sequences are occasionally used as examples or counterexamples regarding certain propositions.

A.4 Exercises

Exercise A.1. Let $A = \{1 - \frac{1}{n} | n \geq 1\}$. Find $\sup A$ and $\inf A$. ◦

Exercise A.2. Let $y \in \mathbb{R}$. Define $A = \{x \in \mathbb{Q} \mid x < y\}$. Find $\sup A$ and $\inf A$. ◦

Exercise A.3. Let $(E, \mathcal{E}, \mu)$ be a measure space, and let (f_n) be a sequence of measurable mappings $f_n : E \to [0, \infty)$. Assume that there is $g : E \to [0, \infty)$ such that $f_n \leq g$ for all $n \geq 1$, where g is integrable with respect to μ. Show that $\limsup_{n\to\infty} \int f_n \,\mathrm{d}\mu \leq \int \limsup_{n\to\infty} f_n \,\mathrm{d}\mu$.
◦

Appendix B

Hints for exercises

B.1 Hints for Chapter 1

Hints for exercise 1.2. Consider the probability space $(\Omega, \mathcal{F}, P) = ([0,1], \mathcal{B}_{[0,1]}, \lambda)$, where λ denotes the Lebesgue measure on $[0,1]$. As a counterexample, consider variables defined as $X_n(\omega) = n\lambda(A_n)^{-1}1_{A_n}$ for an appropriate sequence of intervals (A_n). ○

Hints for exercise 1.4. Show that the sequence $(EX_n)_{n\geq 1}$ diverges and use this to obtain the result. ○

Hints for exercise 1.5. Show that for any ω with $P(\{\omega\}) > 0$, $X_n(\omega)$ converges to $X(\omega)$. Obtain the desired result by noting that $\{\omega \mid P(\{\omega\}) > 0\}$ is an almost sure set. ○

Hints for exercise 1.6. Show that $P(|X_n - X| \geq \varepsilon) \leq P(|X_n 1_{F_k} - X 1_{F_k}| \geq \varepsilon) + P(F_k^c)$, and use this to obtain the result. ○

Hints for exercise 1.7. To prove that $\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon_k) = 0$ for all $k \geq 1$ implies $X_n \xrightarrow{P} X$, take $\varepsilon > 0$ and pick k such that $0 \leq \varepsilon_k \leq \varepsilon$. ○

Hints for exercise 1.8. Using that the sequence $(\sup_{k\geq n} |X_k - X| > \varepsilon)_{n\geq 1}$ is decreasing, show that $\lim_{n\to\infty} P(\sup_{k\geq n} |X_k - X| > \varepsilon) = P(\cap_{n=1}^\infty \cup_{k=n}^\infty |X_k - X| > \varepsilon)$. Use this to prove the result. ○

Hints for exercise 1.9. To obtain that d is a pseudometric, use that $x \mapsto x(1+x)^{-1}$ is increasing on $[0,\infty)$. To show that $X_n \xrightarrow{P} X$ implies convergence in d, prove that for all $\varepsilon > 0$, it holds that $d(X_n, X) \leq P(|X_n - X| > \varepsilon) + \frac{\varepsilon}{1+\varepsilon}$. In order to obtain the converse, apply Lemma 1.2.7. ◦

Hints for exercise 1.10. Choose c_n as a positive number such that $P(|X_n| \geq \frac{1}{nc_n}) \leq \frac{1}{2^n}$. Use Lemma 1.2.11 to show that this choice yields the desired result. ◦

Hints for exercise 1.11. To prove the first claim, apply Lemma 1.2.7 with $p = 4$. To prove the second claim, apply Lemma 1.2.12. ◦

Hints for exercise 1.12. Use Lemma 1.2.13 and Fatou's lemma to show that $E|X|^p$ is finite. Apply Hölder's inequality to obtain convergence in $\mathcal{L}^q$ for $1 \leq q < p$. ◦

Hints for exercise 1.13. Define $S(n,m) = \cap_{k=n}^{\infty}(|X_k - X_n| \leq \frac{1}{m})$ and argue that it suffices for each $\varepsilon > 0$ to show that there exists $F \in \mathcal{F}$ with $P(F^c) \leq \varepsilon$ such that for all $m \geq 1$, there is $n \geq 1$ with $F \subseteq S(n,m)$. To obtain such a set F, consider a sequence $(\varepsilon_m)_{m\geq 1}$ of positive numbers with $\sum_{m=1}^{\infty} \varepsilon_m \leq \varepsilon$ and choose for each m an n_m with $P(S(n_m, n)) \geq 1 - \varepsilon_m$. ◦

Hints for exercise 1.14. Apply Lemma 1.2.12 and Lemma 1.2.7. ◦

Hints for exercise 1.15. Use Lemma 1.2.13, also recalling that all sequences in $\mathbb{R}$ which are monotone and bounded are convergent. ◦

Hints for exercise 1.16. First argue that $(|X_{n+1} - X_n| \leq \varepsilon_n$ evt.) is an almost sure set. Use this to show that almost surely, for $n > m$ large enough, $|X_n - X_m| \leq \sum_{k=m}^{\infty} \varepsilon_k$. Using that $\sum_{k=m}^{\infty} \varepsilon_k$ tends to zero as m tends to infinity, conclude that (X_n) is almost surely Cauchy. ◦

Hints for exercise 1.17. To prove almost sure convergence, calculate an explicit expression for $P(|X_n - 1| \geq \varepsilon)$ and apply Lemma 1.2.12. To prove convergence in $\mathcal{L}^p$, apply the dominated convergence theorem. ◦

Hints for exercise 1.18. Apply Lemma 1.2.11. ◦

Hints for exercise 1.19. Use Lemma 1.3.12 to prove the contrapositive of the desired implication. ◦

Hints for exercise 1.20. To calculate $P(X_n/\log n > c \text{ i.o.})$, use the properties of the exponential distribution to obtain an explicit expression for $P(X_n/\log n > c)$, then apply Lemma 1.3.12. To prove $\limsup_{n\to\infty} X_n/\log n = 1$ almost surely, note that for all $c > 0$, it holds that $\limsup_{n\to\infty} X_n/\log n \le c$ when $X_n/\log n \le c$ eventually and $\limsup_{n\to\infty} X_n/\log n \ge c$ when $X_n/\log n > c$ infinitely often. ◦

Hints for exercise 1.21. Use that sequence $(\cup_{k=n}(X_k \in B))_{n\ge 1}$ is decreasing to obtain that $(X_n \in B \text{ i.o.})$ is in $\mathcal{J}$. Take complements to obtain the result on $(X_n \in B \text{ evt.})$. ◦

Hints for exercise 1.22. Show that

$$\left(\lim_{n\to\infty}\sum_{k=1}^{n} a_{n-k+1}X_k \in B\right) = \left(\lim_{n\to\infty}\sum_{k=m}^{n} a_{n-k+1}X_k \in B\right),$$

and use this to obtain the result. ◦

Hints for exercise 1.23. For the result on convergence in probability, work directly from the definition of convergence in probability and consider $0 < \varepsilon < 1$ in this definition. For the result on almost sure convergence, note that X_n converges to zero if and only if X_n is zero eventually, and apply Lemma 1.3.12. ◦

Hints for exercise 1.24. Use the monotone convergence theorem. ◦

Hints for exercise 1.25. Use Theorem 1.3.10 to show that $\sum_{k=1}^{n} a_k X_k$ either is almost surely divergent or almost surely convergent. To obtain the sufficient criterion for convergence, apply Theorem 1.4.2. ◦

Hints for exercise 1.26. Let (X_n) be an sequence of independent random variables concentrated on $\{0, n\}$ with $P(X_n = n) = p_n$. Use Lemma 1.3.12 to choose (p_n) so as to obtain the result. ◦

Hints for exercise 1.27. Apply Theorem 1.4.3. ◦

Hints for exercise 1.28. Use Lemma 1.3.12 to conclude that $P(|X_n| > n \text{ i.o.}) = 0$ if and only if $\sum_{n=1}^{\infty} P(|X_1| > n)$ is finite. Apply the monotone convergence theorem and Tonelli's theorem to conclude that the latter is the case if and only if $E|X_1|$ is finite. ◦

Hints for exercise 1.29. Apply Exercise 1.28 to show that $E|X_1|$ is finite. Apply Theorem 1.5.3 to show that $EX_1 = c$. ◦

B.2 Hints for Chapter 2

Hints for exercise 2.1. To show that T is measure preserving, find simple explicit expressions for $T(x)$ for $0 \leq x < \frac{1}{2}$ and $\frac{1}{2} \leq x < 1$, respectively, and use this to show the relationship $P(T^{-1}([0,\alpha))) = P([0,\alpha))$ for $0 \leq \alpha \leq 1$. Apply Lemma 2.2.1 to obtain that T is P-measure preserving. To show that S is measure preserving, first show that it suffices to consider the case where $0 \leq \lambda < 1$. Fix $0 \leq \alpha \leq 1$. Prove that for $\alpha \geq \mu$, $S^{-1}([0,\alpha)) = [0,\alpha-\mu)\cup[1-\mu,1)$, and for $\alpha < \mu$, $S^{-1}([0,\alpha)) = [1-\mu, 1-\mu+\alpha)$. Use this and Lemma 2.2.1 to obtain the result. ◦

Hints for exercise 2.2. Apply Lemma 2.2.1. To do so, find a simple explicit expression for $T(x)$ when $\frac{1}{n+1} < x \leq \frac{1}{n}$, and use this to calculate, for $0 \leq \alpha < 1$, $T^{-1}([0,\alpha))$ and subsequently $P(T^{-1}([0,\alpha)))$, ◦

Hints for exercise 2.3. Assume, expecting a contradiction, that P is a probability measure such that T is measure preserving for P. Show that this implies $P(\{0\}) = 0$ and that $P((\frac{1}{2^n}, \frac{2}{2^n}]) = 0$ for all $n \geq 1$. Use this to obtain the desired contradiction. ◦

Hints for exercise 2.4. Let $\lambda = n/m$ for $n \in \mathbb{Z}$ and $m \in \mathbb{N}$. Show that $T^m(x) = x$ in this case. Fix $0 \leq \alpha \leq 1$ and put $F_\alpha = \cup_{k=0}^{m-1} T^{-k}([0,\alpha])$ and show that for α small and positive, F_α is a set in the T-invariant σ-algebra which has a measure not equal to zero or one. ◦

Hints for exercise 2.5. Prove that $\int X - X \circ T \, \mathrm{d}P = 0$ and use this to obtain the result. ◦

Hints for exercise 2.6. Show that $\mathcal{I}_T \subseteq \mathcal{I}_{T^2}$ and use this to prove the result. ◦

Hints for exercise 2.7. Consider a space Ω containing only two points. ◦

Hints for exercise 2.8. For part two, note that $\cup_{k=n}^\infty T^{-k}(F) \subseteq \cup_{k=0}^\infty T^{-k}(F)$ and use that T is measure preserving. For part three, use that $F \subseteq \cup_{k=0}^\infty T^{-k}(F)$. For part four, use that $F = (F \cap (T^k \in F^c \text{ evt.})) \cup (F \cap (T^k \in F^c \text{ evt.})^c)$ ◦

Hints for exercise 2.9. To show that the criterion is sufficient for T to be ergodic, use Theorem 2.2.3. For the converse implication, assume that T is ergodic and use Theorem 2.2.3 to argue that the result holds when X and Y are indicators for sets in $\mathcal{F}$. Consider $X = 1_G$ and Y nonnegative and bounded and use linearity and approximation with simple functions to obtain that the criterion also holds in this case. Use a similar argument to obtain the criterion for general X and Y such that X is nonnegative and integrable and Y is nonnegative and bounded. Use linearity to obtain the final extension to X integrable and Y bounded. ∘

Hints for exercise 2.10. First use Lemma 2.2.6 to argue that it suffices to show that for $\alpha, \beta \in [0,1)$, $\lim_{n\to\infty} P([0,\beta) \cap T^{-n}([0,\alpha))) = P([0,\beta))P([0,\alpha))$. To do so, first show that $T^n(x) = 2^n x - [2^n x]$ and use this to obtain a simple explicit expression for $T^{-n}([0,\alpha))$. Use this to prove the desired result. ∘

Hints for exercise 2.11. For part one, use that the family $\{F_1 \times F_2 \mid F_1 \in \mathcal{F}_1, F_2 \in \mathcal{F}_2\}$ is a generating family for $\mathcal{F}_1 \otimes \mathcal{F}_2$ which is stable under finite intersections and apply Lemma 2.2.1. For part two, show that whenever F_1 is T_1-invariant and F_2 is T_2-invariant, $F_1 \times F_2$ is T-invariant, and use this to obtain the desired result. For part three, use that for $F_1 \in \mathcal{F}_1$ and $F_2 \in \mathcal{F}_2$, it holds that $P_1(F_1) = P(F_1 \times \Omega_2)$ and $P_2(F_2) = P(\Omega_1 \times F_2)$. For part four, use Lemma 2.2.6. ∘

Hints for exercise 2.12. Let $A = (\hat{X}_n \in B \text{ i.o.})$ and note that $(X_n \in B \text{ i.o.}) = X^{-1}(A)$. Show that A is θ-invariant to obtain the result. ∘

Hints for exercise 2.13. For $B \in \mathcal{B}_\infty$, express $Z(P)(B)$ in terms of $X(P)(B)$, $Y(P)(B)$ and p. Use this to obtain that θ is measure preserving for $Z(P)$. ∘

Hints for exercise 2.14. Assume that (X_n) is stationary. Using that θ is $X(P)$-measure preserving, argue that all X_n have the same distribution and conclude that $EX_n = EX_k$ for all $n, k \geq 1$. Using a similar argument, argue that for all $1 \leq n \leq k$, (X_n, X_k) has the same distribution as $(X_1, X_{k-(n-1)})$ and conclude that $\text{Cov}(X_n, X_k) = \text{Cov}(X_1, X_{k-(n-1)})$. Use this to conclude that (X_n) is weakly stationary. ∘

Hints for exercise 2.15. Use Exercise 2.14 to argue that if (X_n) is stationary, it is also weakly stationary. To obtain the converse implication, assume that (X_n) is stationary and argue that for all $n \geq 1$, $(X_2, \ldots, X_{n+1})$ has the same distribution as $(X_1, \ldots, X_n)$. Combine this with the assumption that (X_n) has Gaussian finite-dimensional distributions in order to obtain stationarity. ∘

B.3 Hints for Chapter 3

Hints for exercise 3.1. First assume that (θ_n) converges with limit θ. In the case where $\theta > 0$, apply Lemma 3.1.9 to obtain weak convergence. In the case where $\theta = 0$, prove weak convergence directly by proving convergence of $\int f \, \mathrm{d}\mu_n$ for $f \in C_b(\mathbb{R})$.

Next, assume that (μ_n) is weakly convergent. Use Lemma 3.1.6 to argue that (θ_n) is bounded. Assume that (θ_n) is not convergent, and argue that there must exist two subsequences (θ_{n_k}) and (θ_{m_k}) with different limits θ and θ^*. Use what was already shown and Lemma 3.1.5 to obtain a contradiction. ◦

Hints for exercise 3.2. To obtain weak convergence when the probabilities converge, apply Lemma 3.1.9. To obtain the converse implication, use Lemma 3.1.3 to construct for each k a mapping in $C_b(\mathbb{R})$ which takes the value 1 on k and takes the value zero for, say $(k-1, k+1)^c$. Use this mapping to obtain convergence of the probabilities. ◦

Hints for exercise 3.3. Applying Stirlings's formula and the fact that $\lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n = e^x$ for all $x \in \mathbb{R}$ to prove that the densities f_n converges pointwise to the density of the normal disribution. Invoke Lemma 3.1.9 to obtain the desired result. ◦

Hints for exercise 3.4. Using Stirling's formula as well as the result that if (x_n) is a sequence converging to x, then $\lim_{n\to\infty} \left(1 + \frac{x_n}{n}\right)^n = e^x$, prove that the probability functions converge pointwise. Apply Lemma 3.1.9 to obtain the result. ◦

Hints for exercise 3.5. Apply Stirling's formula and Lemma 3.1.9. ◦

Hints for exercise 3.6. Let F_n be the cumulative distribution function for μ_n. Using the properties of cumulative distribution functions, show that for $x \in \mathbb{R}$ satisfying the inequalities $q(k/(n+1)) < x < q((k+1)/(n+1))$, with $k \leq n$, it holds that $|F_n(x) - F(x)| \leq 2/(n+1)$. Also show that for $x < q(1/(n+1))$ and $x > q(n/(n+1))$, $F_n(x) - F(x)| \leq 1/(n+1)$. Then apply Theorem 3.2.3 to obtain the result. ◦

Hints for exercise 3.7. First assume that (ξ_n) and (σ_n) converge to limits ξ and σ. In the case where $\sigma > 0$, apply Lemma 3.1.9 to obtain weak convergence of μ_n. In the case where $\sigma = 0$, use Theorem 3.2.3 to obtain weak convergence.

Next, assume that μ_n converges weakly. Use Lemma 3.1.6 to show that both (ξ_n) and (σ_n)

are bounded. Then apply a proof by contradiction to show that ξ_n and σ_n both must be convergent. ◦

Hints for exercise 3.8. Let $\varepsilon > 0$, and take n so large that $|x_n - x| \leq \varepsilon$. Use Lemma 3.2.1 and the monotonicity properties of cumulative distribution functions to obtain the set of inequalities $F(x-\varepsilon) \leq \liminf_{n\to\infty} F_n(x_n) \leq \limsup_{n\to\infty} F_n(x_n) \leq F(x+\varepsilon)$. Use this to prove the desired result. ◦

Hints for exercise 3.9. Argue that with F_n denoting the cumulative distribution function for μ_n, it holds that $F_n(x) = 1 - (1-1/n)^{[nx]}$, where $[nx]$ denotes the integer part of nx. Use l'Hôpital's rule to prove pointwise convergence of $F_n(x)$ as n tends to infinity, and invoke Theorem 3.2.3 to conclude that the desired result holds. ◦

Hints for exercise 3.10. Use the binomial theorem. ◦

Hints for exercise 3.11. Use the Taylor expansion of the exponential function. ◦

Hints for exercise 3.12. Use independence of X and Y to express the characteristic function of XY as an integral with respect to $\mu \otimes \nu$. Apply Fubini's theorem to obtain the result. ◦

Hints for exercise 3.13. Argue that XY and $-ZW$ are independent and follow the same distribution. Use Lemma 3.4.15 to 3.13 to express the characteristic function of $XY - ZW$ in terms of the characteristic function of XY. Apply Exercise 3.12 and Example 3.4.10 to obtain a closed expression for this characteristic function. Recognizing this expression as the characteristic function for the Laplace distribution, use Theorem 3.4.19 to obtain the desired distributional result. ◦

Hints for exercise 3.14. Define a triangular array by putting $X_{nk} = (X_n - EX_n)/\sqrt{\sum_{k=1}^n VX_k}$. Apply Theorem 3.5.6 to show that $\sum_{k=1}^n X_{nk}$ converges to the standard normal distribution, and conclude the desired result from this. ◦

Hints for exercise 3.15. Fix $\varepsilon > 0$. Use independence and Lemma 1.3.12 to conclude that $\sum_{n=1}^\infty P(|X_n| > \varepsilon)$ converges. To argue that $\sum_{n=1}^\infty VX_n 1_{(|X_n| \leq \varepsilon)}$, assume that the series is divergent. Put $Y_n = X_n 1_{(|X_n|\leq\varepsilon)}$ and $S_n = \sum_{k=1}^n Y_n$. Use Exercise 3.15 to argue that S_n converges almost surely, while $(S_n - ES_n)/\sqrt{VS_n}$ converges in distribution to the standard normal distribution. Use Lemma 3.3.2 to conclude that $ES_n/\sqrt{VS_n}$ converges in distribution to the standard normal distribution. Obtain a contradiction from this. For the convergence of the final series, apply Theorem 1.4.2. ◦

Hints for exercise 3.16. Use Theorem 3.5.3 to obtain that using the probability space $(\Omega, \mathcal{F}, P_\lambda)$, $\overline{X}_n$ is asymptotically normal. Fix a differentiable mapping $f : \mathbb{R} \to \mathbb{R}$, and use Theorem 3.6.3 to show that $f(\overline{X}_n)$ is asymptotically normal as well. Use the form of the asymptotic parameters to obtain a requirement on f' for the result of the exercise to hold. Identify a function f satisfying the requirements from this. ∘

Hints for exercise 3.17. In the case $\alpha > 1/2$, use Lemma 1.2.7 to obtain the desired convergence in probability. In the cas $\alpha \leq 1/2$, note by Theorem 3.5.3, that $(S_n - n\xi)/n^{1/2}$ converges in distribution to the standard normal distribution. Fix $\varepsilon > 0$ and use Lemma 3.1.3 to obtain a mapping $g \in C_b(\mathbb{R})$ such that $1_{(\xi-2\varepsilon,\xi+2\varepsilon)^c}(x) \leq g(x) \leq 1_{[\xi-\varepsilon,\xi+\varepsilon]^c}(x)$. Use this to prove that $\liminf_{n\to\infty} P(|S_n - n\xi|/n^\alpha \geq \varepsilon)$ is positive, and conclude that $(S_n - n\xi)/n^\alpha$ does not converge in probability to zero. ∘

Hints for exercise 3.18. First calculate EX_n^2 and EX_n^4. Use Theorem 3.5.3, to obtain that $\overline{X}_n$ is asymptotically normal. Apply Theorem 3.6.3 to obtain that $\hat{\theta}_n$ is asymptotically normal as well. ∘

Hints for exercise 3.19. Use Theorem 3.5.3 and Theorem 3.6.3 to obtain the results on $\overline{X}_n$ and $\overline{X}_n^{-1}$. In order to show $Y_n \xrightarrow{P} 1/\mu$, calculate EY_n and VY_n and use Lemma 1.2.7 to prove the convergence. ∘

Hints for exercise 3.20. Use Theorem 3.5.3 to argue that $\overline{X}_n$ is asymptotically normal. To obtain the result on $(Y_n - \theta)/(\sqrt{4\theta^2/9n})$, define a triangular array $(X_{nk})_{n\geq k\geq 1}$ by putting $X_{nk} = \frac{\sqrt{36}}{\theta n^{3/2}}(kU_k - \frac{k\theta}{2})$ and apply Theorem 3.5.7. ∘

B.4 Hints for Chapter 4

Hints for exercise 4.1. Use that ν_1 and ν_2 are signed measures to show that $\alpha\nu_1 + \beta\nu_2$ satisfies (i) and (ii) in the definition. ∘

Hints for exercise 4.2. For $\mu \ll \tau$: Use that $\tau(A) = 0$ if and only if $A = \emptyset$.

For the density f: Recall that *any* function $g : \mathbb{N}_0 \to \mathbb{R}$ is measurable and

$$\int_A g \,\mathrm{d}\tau = \sum_{a\in A} g(a) \qquad \text{for } A \in \mathbb{P}(\mathbb{N}_0)$$

and write $\mu(A) = \sum_{a\in A} f(a)$ for a suitable function f.

Finally show that $\nu \ll \mu$ and find a counter example to $\mu \ll \nu$. ◦

Hints for exercise 4.3. Define $\nu = \nu_1 + \nu_2$ and argue that $\nu \ll \mu$. Argue that there exists measurable and μ–integrable functions f, f_1, f_2 with $\nu(F) = \int_F f \, \mathrm{d}\mu$, $\nu_1(F) = \int_F f_1 \, \mathrm{d}\mu$, and $\nu_2(F) = \int_F f_2 \, \mathrm{d}\mu$ for all $F \in \mathbb{F}$. Show that

$$\int_F f \, \mathrm{d}\mu = \int_F f_1 \, \mathrm{d}\mu + \int_F f_2 \, \mathrm{d}\mu \quad \text{for all } F \in \mathbb{F}$$

and conclude the desired result. ◦

Hints for exercise 4.4. Argue that $\nu \ll \mu$ and let $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$, $h = \frac{\mathrm{d}\nu}{\mathrm{d}\pi}$, and $g = \frac{\mathrm{d}\pi}{\mathrm{d}\mu}$. Note that π and g are non–negative measure and density as known from Sand1. Show that $\nu(F) = \int_F hg \, \mathrm{d}\mu$ and $\nu(F) \int_F f \, \mathrm{d}\mu$. ◦

Hints for exercise 4.5. Let $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ and $g = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}$. Show that $\nu(F) = \int_F fg \, \mathrm{d}\nu$ and $\nu(F) = \int_F 1 \, \mathrm{d}\nu$. Conclude the desired result ν–a.e. Obtain the result μ–a.e. by symmetry. ◦

Hints for exercise 4.6. Find a disjoint sequence of sets (F_n) such that $\mu(F_n) < \infty$ and $\bigcup F_n = \Omega$. Define the measures $\mu_n(F) = \mu(F \cap F_n)$ and $\nu_n(F) = \nu(F \cap F_n)$ and show that $\nu_n \ll \mu_n$ for all n. Let $f_n = \frac{\mathrm{d}\nu_n}{\mathrm{d}\mu_n}$ and $f_n = 0$ on F_n^c (why is that OK?). Define $f = \sum_{n=1}^{\infty} f_n$. Show that

$$\int |f| \, \mathrm{d}\mu = \cdots = \sum_{n=1}^{\infty} \Big(\int_{(f_n > 0)} f_n \, \mathrm{d}\mu - \int_{(f_n < 0)} f_n \, \mathrm{d}\mu \Big) \leq \cdots \leq 2 \sup_{F \in \mathbb{F}} |\nu(F)| < \infty$$

and that $\nu(F) = \int_F f \, \mathrm{d}\mu$. ◦

Hints for exercise 4.7. Show that X is a conditional expectation of Y given $\mathcal{D}$, and that Y is a conditional expectation of X given $\mathcal{D}$. ◦

Hints for exercise 4.8. Straightforward application of Theorem 4.2.6 *(2)*, *(5)* and *(7)*. ◦

Hints for exercise 4.9. "$\Leftarrow$" is trivial. For "$\Rightarrow$" show that $EX^2 = EY^2$ and use that $X = Y$ a.s. if and only if $E[(X - Y)^2] = 0$. Apply Theorem 4.2.6. ◦

Hints for exercise 4.10. Straightforward calculations! ◦

Hints for exercise 4.11. Recall that $x^+ = \max\{x, 0\}$. Show and use $P(0 \leq E(X^+|\mathcal{D})) = 1$ and $P(E(X|\mathcal{D}) \leq E(X^+|\mathcal{D})) = 1$. ◦

Hints for exercise 4.12. Use that $|x| = x^+ + x^-$ and $(-x)^+ = x^-$ and Exercise 4.11. ◦

Hints for exercise 4.13. Show that $E(X|\mathcal{D}) = \frac{1}{2}$. Show and use that if D is countable, then $1_D = 0$ a.s. and $1_{D^c} = 1$ a.s. ∘

Hints for exercise 4.14. Compare $\sigma(X_1)$ and $\sigma(X_2)$. ∘

Hints for exercise 4.15. Define

$$\mathbb{H} = \{F \in \mathbb{D} \mid \int_D Y \,\mathrm{d}P = \int_D X \,\mathrm{d}P\}$$

and use Dynkin's lemma (Lemma A.2.2) to show that $\sigma(\mathcal{G}) \subseteq \mathbb{H}$ (it is assumed that $\mathcal{G} \subseteq \mathbb{H}$). ∘

Hints for exercise 4.16.

(1) Write $E(Y|Z) = \phi(Z)$ (!) so e.g. the left hand side equals $E(\phi(Z)1_{(Z\in B)}1_{(\mathbb{X}\in C)})$. Use that $Z \perp\!\!\!\perp \mathbb{X}$ and $(Y,Z) \perp\!\!\!\perp \mathbb{X}$.

(2) Use Exercise 4.15 and 1 to show that $E(Y|Z)$ is a conditional expectation of Y given $\sigma(Z,\mathbb{X})$.

∘

Hints for exercise 4.17.

(1) Show that $\sigma(S_n, S_{n+1}, S_{n+2}, \ldots) = \sigma(S_n, X_{n+1}, X_{n+2}, \ldots)$. Use Exercise 4.16 and that the X_n–variables are independent.

(2) First show that $\frac{1}{n}S_n = E(X_1|S_n)$ by checking the conditions for being a conditional expectation of X_1 given S_n. For the proof of

$$\int_{(S_n\in B)} \frac{1}{n} S_n \,\mathrm{d}P = \int_{(S_n\in B)} X_1 \,\mathrm{d}P \quad \text{for all } B \in \mathbb{B}\,,$$

use (and show) that $1_{(S_n\in B)}X_k \overset{\mathcal{D}}{=} 1_{(S_n\in B)}X_1$ (the distributions are equal) for all $k = 1,\ldots,n$.

∘

Hints for exercise 4.18. Write $X = Z + \mu_1 + cY$, where $Z = (X - \mu_1) - cY$ and c is chosen such that $\text{Cov}(Z, Y) = 0$ (Recall that in that case Z and Y will be independent!) ◦

B.5 Hints for Chapter 5

Hints for exercise 5.5. For ⇒, let $F_n = (\tau = n)$. For ⇐, show and use that

$$(\tau = m) = \cap_{n=1}^{m} F_n^c \cap F_m \, .$$

◦

Hints for exercise 5.6. Use the partition $(\tau = m) = \cup_{k=1}^{m-1} (\tau = m, \sigma = k)$. ◦

Hints for exercise 5.7. (1): For ⇐, let $\tau = n$ and $k = n + 1$. For ⇒, use Corollary 5.2.13. ◦

Hints for exercise 5.8. For (2): Use Exercise 5.4. For 3: Show $E(S_1) = 0$ and use (2) in Exercise 5.7. For (4): Use The Strong of Large Numbers to show that $S_n \xrightarrow{\text{a.s.}} +\infty$. For 5: Use monotone convergence of both sides in (3). ◦

Hints for exercise 5.9. First show that $(S_n, \mathcal{F}_n)$ is a martingale with a suitable choice of $(\mathcal{F}_n)$. Then use the independence to show that $ES_n^2 \leq \sum_{n=1}^{\infty} EX_n^2 < \infty$ for all $n \in \mathbb{N}$. Finally use The martingale convergence theorem (for the argument of $\sup_n ES_n^+ < \infty$, recall that $x^+ \leq |x| \leq x^2 + 1$). ◦

Hints for exercise 5.10. For (2): See that $M_n \geq 0$ and use Theorem 5.3.2. For (3): Use Fatou's lemma. For (5): Use Exercise 5.4. For (7): Use Corollary 5.2.13 and the fact that $\tau \wedge n$ and n are bounded stopping times. For (9): Use (7)+(8)+ dominated convergence. For (10): Let $q = P(S_\tau = b)$ and write $EM_\tau = qr^b + (1-q)r^a$. ◦

Hints for exercise 5.11. For (1): Exploit the inequality $(x-y)^2 \geq 0$. For the integrability, use that $1_{D_n} E(X|\mathbb{D})$ is bounded by n. For (2): use that both 1_{D_n} and $E(X|\mathcal{D})$ are $\mathcal{D}$–measurable. For (3): Use (1) and (2) to obtain that

$$E\Big(1_{D_n}\big(X^2 - E(X|\mathbb{D})^2\big)|\mathbb{D}\Big) \geq 0 \text{ a.s.}$$

◦

Hints for exercise 5.12.

(1) Use Exercise 5.11 and that $E(X_{n+1}|\mathcal{F}_n)^2 = X_n^2$ a.s. by the martingale property.

(2) Use Exercise 5.4.

(3) Use Corollary 5.2.13, since $\tau \wedge n$ and n are bounded stopping times.

(4) Write

$$EX_{\tau\wedge n}^2 = \int_{(\max_{k=1,\dots,n}|X_k|\geq\epsilon)} X_{\tau\wedge n}^2 \,\mathrm{d}P + \int_{(\max_{k=1,\dots,n}|X_k|<\epsilon)} X_{\tau\wedge n}^2 \,\mathrm{d}P$$

and use (and prove) that $|X_{\tau\wedge n}| \geq \epsilon$ on the set $(\max_{k=1,\dots,n}|X_k| \geq \epsilon)$.

∘

Hints for exercise 5.13. For (3): Show that $A_n \in \mathcal{F}_{\tau\wedge n}$ (recall the definition of $\mathcal{F}_{\tau\wedge n}$) and use this to show

$$\int_{A_n} Y_{\tau\wedge n}\,\mathrm{d}P \leq \int_{A_n} E(Y_n|\mathcal{F}_{\tau\wedge n})\,\mathrm{d}P = \int_{A_n} Y_n\,\mathrm{d}P\,.$$

∘

Hints for exercise 5.14. (1): Use Theorem 5.4.5 and the fact that $E|X_n - 0| = EX_n$.
(2): According to (1), the variables should have both positive and negative values. Use linear combinations of indicator functions like $1_{[0,\frac{1}{n})}$ and $1_{[\frac{1}{n},\frac{2}{n})}$. ∘

Hints for exercise 5.15. For (1): Use 10 in Theorem 4.2.6 and the definition of conditional expectations. For (2): First divide into the two situations $|X| \leq K$ and $|X| > K$, and secondly use Markov's inequality. For (3): Obtain that for all $K \in \mathbb{N}$

$$\lim_{x\to\infty}\sup_{\mathcal{D}\in\mathbb{G}}\int_{|(E(X|\mathcal{D})|>x)} |E(X|\mathcal{D})|\,\mathrm{d}P \leq \int_{(|X|>K)} |X|\,\mathrm{d}P\,.$$

Let $K \to \infty$ and use dominated convergence. ∘

Hints for exercise 5.16.

(1) This IS very easy.

(2) Use Theorem 5.4.5 and that $X_{\tau\wedge n} \xrightarrow{P} X_\tau$.

(3) Show that $EX_{\tau\wedge n} \to EX_\tau$ (use e.g. the remark before Definition 5.4.1) and that $EX_{\tau\wedge n} = EX_1$.

(4) Show and use that

$$\int_{(|X_{\tau\wedge n}|>x)} |X_{\tau\wedge n}| \,\mathrm{d}P \leq \int_{(|Y|>x)} |Y| \, df P$$

for all $x > 0$ and $n \in \mathbb{N}$. Use dominated convergence to see that the right hand side $\to 0$ as $x \to \infty$.

(5) Write

$$\sum_{n=0}^{\infty} P(\tau > n) = \sum_{n=0}^{\infty} 1_{(k\geq n)} P(\tau = k)$$

and interchange the sums.

(6) Define Y as the right hand side of (5.32) and expand $E|Y|$:

$$E|Y| = \ldots = E|X_1| + \sum_{m=1}^{\infty} \int_{(\tau>m)} E(|X_{m+1} - X_m| \big| \mathcal{F}_m) \,\mathrm{d}P$$

Use $E(|X_{n+1} - X_n| \big| \mathcal{F}_n) \leq B$ a.s. and (5) to obtain that $E|Y| < \infty$. Use (1)-(4).

(9) Use (6)-(8) to show that $EZ_\sigma = 0$. Furthermore realise that $Z_\sigma = S_\sigma - \sigma\xi$.

◦

Hints for exercise 5.17. For (2): Use that all $Y_n \geq 0$. For (3): Write $Y = \liminf_{n\to\infty} Y_n$ and use Fatou's Lemma. For 4: Write $Y_n = \exp\left(\sum_{k=1}^{n} \log(X_k)\right)$ and use The Strong Law of Large Numbers to show $\frac{1}{n}\sum_{k=1}^{n} \log(X_k) \to \xi < 0$ a.s.. For 5: Use that if $Y_n \xrightarrow{\mathcal{L}^1} Z$ then $Y_n \xrightarrow{P} Z$ and conclude that $Z = 0$ a.s. Realise that $Y_n \overset{L^1}{\nrightarrow} 0$ (You might need that if $Y_n \xrightarrow{P} Y$ and $Y_n \xrightarrow{P} Z$, then $Y = Z$ a.s.). ◦

Hints for exercise 5.18.

(1) Use the triangle inequality to obtain $|E|X_n| - E|X|| \leq E|X_n - X|$. For the second claim use Theorem 1.2.8.

(3) Define $U_n = X_n - X$ and $V_n = |X_n| + |X|$. Showing $X_n \xrightarrow{\mathcal{L}^1} X$ will be equivalent (?) to show $\limsup_{n\to\infty} E|X_n - X| = 0$. Argue and use that $\limsup_{n\to\infty} |X_n - X| = 0$ a.s.

(4) Let (n_ℓ) be a subsequence such that

$$E|X_{n_\ell} - X| \xrightarrow{\ell\to\infty} \limsup_{n\to\infty} E|X_n - X|$$

(such a subsequence will always exist). Find a subsequence (n_{ℓ_k}) of (n_ℓ) such that

$$X_{n_{\ell_k}} \xrightarrow{\text{a.s.}} X\,.$$

Conclude from (3) that $E|X_{n_{\ell_k}} - X| \to 0$ and use the uniqueness of this limit to derive $\limsup_{n\to\infty} E|X_n - X| = 0$.

(5) Use (4) and Theorem 5.4.5 to conclude that (Y_n) is uniformly integrable. Use (2) in Thm 5.4.7 (You might need that if $Y_n \xrightarrow{P} Y$ and $Y_n \xrightarrow{P} Z$, then $Y = Z$ a.s.).

(6) For $\Leftarrow$, do as in 5 and use furthermore Thm 5.4.7 to conclude that Y closes.
For $\Rightarrow$, use 5.4.7, 5.4.5, and (1).

∘

Hints for exercise 5.19.

(3) Use that if $\tau_1 = n$ then we have lost the first $n-1$ games and won the n'th games, such that (like the example in the exercise)

$$X_1 = -1, X_2 = -1-2, \ldots, X_{n-1} = \sum_{k=1}^{n-1} 2^{k-1}, \quad X_n = \sum_{k=1}^{n-1} 2^{k-1} + 2^{n-1} = 1$$

(5) It may be useful to recall that $(-X_n, \mathcal{F}_n)$ is a submartingale, and $(-X_{\tau_k}, \mathcal{F}_{\tau_k})$ is a supermatingale.

6: Show and use that $((-X_n)^+)$ is uniformly integrable.

∘

Hints for exercise 5.20.

(1) Use $x^+ \leq 1 + |x|^p$ to show $\sup EX_n^+ < \infty$.

(2) For $E|X|^p < \infty$, apply Fatou's lemma to $E|X|^p = E \liminf_n |X_n|^p$. For $X_n \xrightarrow{L^p} X$, show that $|X_n - X|^p \leq 2^p \sup_n |X_n|^p$ and use dominated convergence on $E\big(|X_n - X|^p\big)$.

(3) Hint: Write $P(Z \geq t) = \int_0^\infty 1_{[t,\infty)}(x)\,\mathrm{d}Z(P)(x)$ and apply Tonelli's Theorem to the right hand side. Use that $1_{[t,\infty)}(x) = 1_{[0,x]}(t)$.

(4) Combine (3) and Doob's inequality from Exercise 5.13.

(5) Apply Toenlli's theorem to the double integral in (4).

(6) Apply Hölder to $E(M_n^{p-1}|X_n|)$ (use that $\frac{1}{p/(p-1)}+\frac{1}{p}=1$). Combine with the inequality from (5).

(7) Write $E\big[\sup_n |X_n|^p\big] = E\big(\lim_n M_n^p\big)$ and use monotone convergence together with the inequality from (6).

○

Hints for exercise 5.21.

(2) Realize that if $\tau_{a,b} > nm$, then especially

$$S_m, S_{2m}, S_{3m}, \dots, S_{nm} \in (a,b)$$

Which leads to the conclusion that

$$|S_m - S_0| < b-a, \quad |S_{2m} - S_m| < b-a, \dots, |S_{nm} - S_{(n-1)m}| < b-a\,.$$

Use that all these differences are independent and identically distributed.

(3) Use (2). Choose a fixed m such that $P(|S_m| < b-a) < 1$ and let $n \to \infty$. The second statement is trivial.

(4) For the first result, use optional sampling for the martingale $(S_n, \mathcal{F}_n)$ and e.g. the two bounded stopping times 1 and $\tau_{a,b} \wedge n$. For the second result, let $n \to \infty$ using dominated (since $S_{\tau_{a,b}\wedge n}$ is bounded) and monotone convergence.

(5) Write $ES_{\tau_{a,b}} = aP(S_{\tau_{a,b}} = a) + bP(S_{\tau_{a,b}} = b)$. Use (3).

(6) Apply the arguments from (4) to the martingale $(S_n^2 - n, \mathcal{F}_n)$. For the second statement, use that the distribution of $S_{\tau_{a,b}}$ is well–known from (5).

(7) Use (3).

(8) Use that on F we have $\tau_{-n,b} = \tau_b$ if and only if $S_{\tau_{-n,b}} = b$.

(9) Use and show that if $\omega \in G$, then $\tau_b(\omega) < \infty$.

(10) Use that $\tau_{-n,b} \uparrow \tau_b$ as $n \to \infty$.

(11) See that $ES_{\tau_b} \neq ES_1$ and compare with Theorem 5.4.9.

(12) Write

$$(\sup_{n\in\mathbb{N}} S_n = \infty) = \bigcap_{n=1}^{\infty} (\tau_n < \infty)$$

∘

Hints for exercise 5.22.

(2): Define the triangular array (X_{nm}) by

$$X_{nm} = \frac{1}{\sqrt{n\alpha^2\sigma^2}} Z_{m-1} Y_m$$

and use Brown's Theorem to show that

$$\frac{1}{\sqrt{n\alpha^2\sigma^2}} M_n \xrightarrow{wk} \mathcal{N}(0,1)\,,$$

(4) Recall that $EY_n = 0$.

(5) Show that $\sup_{n\in\mathbb{N}} EN_n^2 < \infty$ and use the

(6) Use appropriate theorems from Chapter 5.

(9) Use the almost sure convergence from (6) and Kronecker's lemma.

(11) Use the result from (2) with $Z_m = Y_m$ for all $m \geq 0$. Use (10) to see that the assumptions are fulfilled.

∘

B.6 Hints for Chapter 6

Hints for exercise 6.1. Find the distribution of $(\tilde{X}_{t_1}, \ldots, \tilde{X}_{t_n})$. ∘

Hints for exercise 6.2. Show the result in two steps:

(1) Show that $P(D_1 \cap D_2) = P(D_1)P(D_2)$ for all $D_1 \in \mathbb{D}_1, D_2 \in \sigma(\mathbb{D}_2)$.

(2) Show (6.20).

For (1): Let $D_1 \in \mathbb{D}_1$ and define

$$\mathbb{E}_{D_1} = \{F \in \sigma(\mathcal{D}_2) \ : \ P(D_1 \cap F) = P(D_1)P(F)\}$$

and then show that $\mathbb{E}_{D_1} = \sigma(\mathcal{D}_2)$ using Lemma A.2.2. Note that you already have $\mathcal{D}_2 \subseteq \mathbb{E}_{D_1} \subseteq \sigma(\mathcal{D}_2)$.

For (2): Let $A \in \sigma(\mathcal{D}_2)$ and define

$$\mathbb{E}_A = \{F \in \mathbb{F}(\mathcal{D}_1) \ : \ P(F \cap A) = P(F)P(A)\}\,.$$

Show that $\mathbb{E}_A = \sigma(\mathcal{D}_2)$ (as for (1)). ○

Hints for exercise 6.3. Show that $\mathbb{D} \perp\!\!\!\perp \sigma(X_u - X_t)$ and use Exercise 6.2. ○

Hints for exercise 6.4. Write $X_t = X_t - X_s + X_s$ and use Exercise 6.3. ○

Hints for exercise 6.6. Show that the finite–dimensional distributions are the same. ○

Hints for exercise 6.7.

(1) Find $H > 0$ such that for all $\gamma > 0$ and all $0 \leq t_1 < \cdots < t_n$

$$(X_{\gamma t_1}, \ldots, X_{\gamma t_n}) \stackrel{\mathcal{D}}{=} (\gamma^H X_{t_1}, \ldots, \gamma^H X_{t_n})\,.$$

(2) Use both of the assumptions: stationary increments and self–similarity.

(3) Use (2) and the assumption $P(X_1 = 0) = 0$.

(4) You need to show that for some $t \geq 0$ and $t_n \to t$, then for all $\epsilon > 0$

$$\lim_{n\to\infty} P(|X_{t_n} - X_t| > \epsilon) = 0\,.$$

Use (2).

○

Hints for exercise 6.8.

(1) Write $C_{n,M}$ on the form

$$C_{n,M} = \Big\{x \in C_{[0,\infty)} \mid \sup_{q \in [n,n+1]\cap\mathbb{Q}} \frac{x_q}{q} > \frac{1}{M}\Big\} = \bigcup_{q\in[n,n+1]\cap\mathbb{Q}} (|\tilde{X}_q| \ldots)$$

(2) Use that

$$(Y \in C_{n,M}) = \Big(\sup_{q\in[n,n+1]\cap\mathbb{Q}} \frac{|Y_q|}{q} > \frac{1}{M}\Big)$$

and Lemma 6.2.9.

(3) For the first result use Markov's inequality and that $EU^4 = 3$ if $U \sim \mathcal{N}(0,1)$.

(4) Use Borel–Cantelli.

(5) Start showing that

$$P\left(\Big(\sup_{t\in[n,n+1]} \frac{|Y_t|}{t} \leq \frac{1}{M}\Big) \text{ evt.}\right) = 1\,.$$

for all $M \in \mathbb{N}$.

(6) Use (5).

∘

B.7 Hints for Appendix A

Hints for exercise A.1. To obtain the supremum, use that weak inqequalities are preserved by limits. ∘

Hints for exercise A.2. To obtain the supremum, use Lemma A.1.3 and the fact that $\mathbb{Q}$ is dense in $\mathbb{R}$. ∘

Hints for exercise A.3. Apply Fatou's lemma to the sequence $(g - f_n)_{n\geq 1}$. ∘

Index

Bibliography

R. B. Ash: Real analysis and Probability, Academic Press, 1972.

T. M. Apostol: Calculus, Volume 1, Blaisdell Publishing Company, 1964.

L. Breiman: Probability, Addison-Wesley, 1968.

M. Loève: Probability Theory I, Springer-Verlag, 1977.

M. Loève: Probability Theory II, Springer-Verlag, 1977.

P. Brémaud: Markov Chains: Gibbs fiends, Monte Carlo simulation and queues, Springer-Verlag, 1999.

S. N. Ethier & T. G. Kurtz: Markov Processes: Characterization and Convergence, Wiley, 1986.

J. R. Norris: Markov Chains, Cambridge University Press, 1999.

S. Meyn & R. L. Tweedie: Markov chains and stochastic stability, Cambridge University Press, 2009.

P. Billingsley: Convergence of Probability Measures, 2nd edition, 1999.

K. R. Parthasarathy: Probability measures on Metric Spaces, Academic Press, 1967.

I. Karatzas, S. E. Shreve: Brownian Motion and Stochastic Calculus, Springer-Verlag, 1988.

N. L. Carothers: Real Analysis, Cambridge University Press, 2000.

E. Hansen: Measure theory, Department of Mathematical Sciences, University of Copenhagen, 2004.

O. Kallenberg: Foundations of Modern Probability, Springer-Verlag, 2002.

L. C. G. Rogers & D. Williams: Diffusions, Markov Processes and Martingales, Volume 1: Foundations, 2nd edition, Cambridge University Press, 2000.

L. C. G. Rogers & D. Williams: Diffusions, Markov Processes and Martingales, Volume 1: Itô calculus, 2nd edition, Cambridge University Press, 2000.