Probability theory II

Dr. Peter Gracar
Universität zu Köln

Version of January 30, 2020
Based on lecture notes of Prof. Dr. Alexander Drewitz

If you spot any typos or mistakes, please drop an email to pgracar@math.uni-koeln.de.

Contents

Chapter 1

Conditional probabilities and expectations revisited

We have seen in [Dre19, Definition 3.2.1] how to define the conditional probability

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

for $A, B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$. In particular, for any such B, the function

$$\mathcal{F} \ni A \mapsto \mathbb{P}(A \,|\, B)$$

defines a probability measure on $\mathcal{F}$. We can therefore also define the conditional expectation of a real random variable X (if it exists) as

$$\mathbb{E}[X \,|\, B] := \frac{\mathbb{E}[X \mathbb{1}_B]}{\mathbb{P}(B)} = \frac{\int_B X \, \mathrm{d}\mathbb{P}}{\mathbb{P}(B)} = \int_\Omega X \, \mathrm{d}\mathbb{P}(\cdot \,|\, B), \tag{1.0.1}$$

where the latter equality follows using algebraic induction.

Example 1.0.1. *Consider the example of the random variable X modeling a fair dice roll. The fairness assumption means*

$$\mathbb{P}(X = i) = \frac{1}{6} \quad \forall i \in \{1, 2, \ldots, 6\}.$$

Thus, the expected value of the dice roll is $\mathbb{E}[X] = \frac{7}{2}$.
Now assume that for whatever reason you do get hold of the insider information that the event $\{X \geq 4\}$ will occur – what is the expected value of the dice roll conditionally on this event? According to the above definition we have with $B := \{X \geq 4\}$ that

$$\mathbb{E}[X \,|\, B] = \frac{\int_B X \, \mathrm{d}\mathbb{P}}{\mathbb{P}(B)} = \frac{\sum_{i=4}^{6} i \frac{1}{6}}{\frac{1}{2}} = 5,$$

which coincides with our intuition.

Similarly to the interpretation of the common expectation, the conditional expectation can be seen as a weighted average of X restricted to B. It turns out, however, that we would like to have a definition for the conditional expectation and probability not only in the case $\mathbb{P}(B) > 0$, but also for events $B \in \mathcal{F}$ for which we have $\mathbb{P}(B) = 0$.

Example 1.0.2. *Consider the situation of two independent Gaussian random variables X and Y distributed according to $\mathcal{N}(0, 1)$, and set $Z := X + Y$. In this case, if for some reason we know the value of the realisation of X but not that of Y, we would intuitively like to have that*

$$\mathbb{P}(Z \in \cdot \,|\, \{X = x\}) \tag{1.0.2}$$

makes sense, and that it would denote the law $\mathcal{N}(x,1)$ *– however, the event* $\{X = x\}$ *has probability* 0 *under* $\mathbb{P}$*, so our common definition of conditional probability does not apply. In what follows we will introduce the machinery to fix this shortcoming. We will, however, start with introducing the conditional expectation in this generalised setting first and then consider the special case of indicator functions in order to obtain the notion of conditional probability.*

Definition 1.0.3. *Let* $\mathcal{G}$ *be a sub-*σ*-algebra of* $\mathcal{F}$*, and let* $X \in \mathcal{L}^1$ *be a real random variable on* $(\Omega, \mathcal{F}, \mathbb{P})$*. A random variable* Y *defined on* $(\Omega, \mathcal{F}, \mathbb{P})$ *is called the* conditional expectation of X with respect to $\mathcal{G}$*, if*

(a) Y *is* $\mathcal{G} - \mathcal{B}(\mathbb{R})$*-measurable, and*

(b) for any $A \in \mathcal{G}$*,*

$$\mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[Y \mathbb{1}_A].$$

In this case we also write $\mathbb{E}[X \,|\, \mathcal{G}] := Y$*. Furthermore, for* $F \in \mathcal{F}$ *we write*

$$\mathbb{P}(F \,|\, \mathcal{G}) := \mathbb{E}[\mathbb{1}_F \,|\, \mathcal{G}]$$

for the conditional probability of F given $\mathcal{G}$*.*

Remark 1.0.4. *It is possible to define the conditional expectation also for a non-negative random variable that is not necessarily integrable. We will usually restrict ourselves to the case where the random variable* X *is integrable (as in Definition 1.0.3); we will however prove the existence of conditional expectation also for the non-integrable case below in Theorem 1.0.5.*

We will see in Exercise 1.0.8 below that the conditional expectation introduced in Definition 1.0.3 can actually be interpreted as a generalisation of the one introduced in (1.0.1). On the other hand, in Corollary 2.2.9 later we will learn how in a multitude of situations one obtains that the general conditional expectation of Definition 1.0.3 can be approximated by the basic conditional expectation introduced in (1.0.1).
It turns out that conditional probabilities and expectations exhibit properties to those known from common probabilities and expectations. Before making this more precise, however, we establish their existence and almost sure uniqueness.

Theorem 1.0.5. *[Existence and uniqueness of conditional expectations] For* $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ *or* $X \in \mathcal{L}^0(\Omega, \mathcal{F}, \mathbb{P})$ *and non-negative, and any sub-*σ*-algebra* $\mathcal{G} \subset \mathcal{F}$*, the conditional expectation* $\mathbb{E}[X \,|\, \mathcal{G}]$ *exists and is unique* $\mathbb{P}$*-a.s.*

Proof. We begin by proving the statement for X integrable. The two maps

$$\mu^+ : \mathcal{G} \ni G \mapsto \mathbb{E}[X^+ \mathbb{1}_G] \qquad \mu^- : \mathcal{G} \ni G \mapsto \mathbb{E}[X^- \mathbb{1}_G]$$

define two non-negative finite measures on $(\Omega, \mathcal{G})$, which are both absolutely continuous with respect to $\mathbb{P}$. Therefore, the Radon-Nikodym theorem [Dre19, Corollary 2.2.17] provides us with $\mathcal{G} - \mathcal{B}(\mathbb{R})$-measurable densities $Y^+, Y^- \in \mathcal{L}^1(\Omega, \mathcal{G}, \mathbb{P})$ such that

$$\mu^\pm(G) = \int_G Y^\pm \, \mathrm{d}\mathbb{P}.$$

As a consequence, setting $Y := Y^+ - Y^-$ has the properties of the conditional expectation of X given $\mathcal{G}$.
Uniqueness:

Assume yet another conditional expectation Z of X given $\mathcal{G}$ to be given. Then $\{Y > Z\}, \{Y < Z\} \in \mathcal{G}$, so by definition of the conditional expectation,

$$0 = \mathbb{E}[X\mathbb{1}_{\{Y>Z\}}] - \mathbb{E}[X\mathbb{1}_{\{Y>Z\}}] = \mathbb{E}[Y\mathbb{1}_{\{Y>Z\}}] - \mathbb{E}[Z\mathbb{1}_{\{Y>Z\}}] = \mathbb{E}[(Y-Z)\mathbb{1}_{\{Y>Z\}}].$$

Using [Dre19, Lemma 2.2.3], this implies that $\mathbb{P}((Y-Z)\mathbb{1}_{\{Y>Z\}} > 0) = 0$ and therefore $\mathbb{P}(Y > Z) = 0$; similarly we obtain $\mathbb{P}(Y < Z) = 0$, so $\mathbb{P}(Y = Z) = 1$, which finishes the proof for X integrable.
For X non-negative, but not necessarily integrable, define first $X_n = \min\{n, X\}$. As X_n is bounded, the conditional expectation $Y_n := \mathbb{E}[X_n \,|\, \mathcal{G}]$ exists by the first part of the proof. Furthermore, notice that $Y_0 = X_0 = 0$ and that for any n, the event $G = \{Y_{n+1} < Y_n\} \in \mathcal{G}$. It follows that

$$\mathbb{E}[\mathbb{1}_G(Y_n - Y_{n+1})] = \mathbb{E}[\mathbb{1}_G(X_n - X_{n+1})] \le 0$$

and from that $Y_{n+1} \ge Y_n$ a.s.
Setting $Y = \sup_n Y_n$ the monotone convergence theorem gives

$$\mathbb{E}[\mathbb{1}_G Y] = \lim_{n\to\infty} \mathbb{E}[\mathbb{1}_G Y_n] = \lim_{n\to\infty} \mathbb{E}[\mathbb{1}_G X_n] = \mathbb{E}[1_G X]$$

as required.
Finally, suppose that Z is also a non-negative $\mathcal{G}$-measurable random variable satisfying the definition of conditional expectation. Then, for any $x \in \mathbb{R}$, setting $G = \{Y > Z, x > Y\}$ makes $\mathbb{1}_G Y$ and $\mathbb{1}_G Z$ bounded and

$$\mathbb{E}[\mathbb{1}_G(Y - Z)] = \mathbb{E}[\mathbb{1}_G X] - \mathbb{E}[\mathbb{1}_G X] = 0,$$

so that $\mathbb{P}(G) = 0$. Letting x go to infinity, we obtain that almost surely $Z \ge Y$. Reversing the roles of Y and Z gives that $Y = Z$ a.s. □

According to the above result, conditional expectations are only uniquely defined $\mathbb{P}$-almost surely, and any random variable fulfilling the properties of the conditional expectation will be called a *version* of the conditional expectation. Therefore, all (in-)equalities involving conditional expectations will always be understood in a $\mathbb{P}$-a.s. sense in what follows.
Note also that in the case the random variable X is non-negative, but possibly not integrable, we must allow $\mathbb{E}[X \,|\, \mathcal{G}]$ in the statement of the theorem to take the values in $\mathbb{R}_+ \cup \{\infty\}$. As we will see shortly, we can omit the second possibility precisely when X is integrable.
It should be noted that one can prove the existence of conditional expectation also without relying on the Radon-Nikodym theorem. Such a construction relies on knowledge of function-analytical tools; a short proof can be found in the appendix (see Theorem A.1.1 and Corollary A.1.2).

Definition 1.0.6. *For random variables X, Y defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and such that $X \in \mathcal{L}^1$, whereas Y can map to an arbitrary measurable space $(E, \mathcal{E})$, we define*

$$\mathbb{E}[X \,|\, Y] := \mathbb{E}[X \,|\, \sigma(Y)].$$

Remark 1.0.7. *Note that there exists a measurable mapping $\varphi : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that*

$$\mathbb{E}[X \,|\, Y] = \varphi \circ Y.$$

In fact, this map φ not only helps to define the conditional $\mathbb{E}[X \,|\, Y]$ as a random variable from Ω to $\mathbb{R}$, but it will help us to define the so-called regular conditional probability ('reguläre bedingte Wahrscheinlichkeit')*, where in particular we want to be able to interpret expressions such as that in (1.0.2) as measurable functions in x.*

Exercise 1.0.8. *(a) Show that in the setting of Example 1.0.2 we have*

$$\mathbb{E}[Z \,|\, X] = X,$$

so φ from the previous remark equals the identity mapping from $\mathbb{R}$ to $\mathbb{R}$.

(b) For $B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$ we can recover the conditional expectation $\mathbb{E}[X \,|\, B]$ defined in (1.0.1) from our new definition by considering $\mathbb{E}[X \,|\, \sigma(B)]$. Then the latter is $\mathbb{P}$-a.s. constant and equal to $\mathbb{E}[X \,|\, B]$ on B.

More generally, for a finite partition $\mathcal{P}$ of Ω with $\mathcal{P} \subset \mathcal{F}$, and $X \in \mathcal{L}^p$ some $p \in [1, \infty)$, we have

$$\mathbb{E}[X \,|\, \sigma(\mathcal{P})] = \sum_{\substack{A \in \mathcal{P} \\ \mathbb{P}(A) > 0}} \mathbb{E}[X \,|\, A] \mathbb{1}_A.$$

1.0.1 Basic properties of conditional expectations

Theorem 1.0.9 (Properties of conditional expectations)**.** *Let $X, Y \in \mathcal{L}^1$ and $\mathcal{A} \subset \mathcal{G}$ be sub-σ-algebras of $\mathcal{F}$. Furthermore, let $c \in \mathbb{R}$. Then $\mathbb{P}$-a.s.:*

(a) For $c \in \mathbb{R}$,

$$\mathbb{E}[cX + Y \,|\, \mathcal{G}] = c\mathbb{E}[X \,|\, \mathcal{G}] + \mathbb{E}[Y \,|\, \mathcal{G}]. \quad \textit{(linearity)} \tag{1.0.3}$$

(b) If $X \le Y$, then

$$\mathbb{E}[X \,|\, \mathcal{G}] \le \mathbb{E}[Y \,|\, \mathcal{G}]. \quad \textit{(monotonicity)}$$

(c)

$$|\mathbb{E}[X \,|\, \mathcal{G}]| \le \mathbb{E}[|X| \,|\, \mathcal{G}]. \quad \textit{(triangle inequality)}$$

(d) For $X_n \in \mathcal{L}^1$, $n \in \mathbb{N}$, with $X_n \uparrow X$ $\mathbb{P}$-a.s. as $n \to \infty$,

$$\mathbb{E}[X_n \,|\, \mathcal{G}] \uparrow \mathbb{E}[X \,|\, \mathcal{G}] \quad \mathbb{P}-a.s. \quad \textit{(Monotone Convergence theorem for conditional expectations)}$$

(e) If $XY \in \mathcal{L}^1$ and X is $\mathcal{G} - \mathcal{B}(\mathbb{R})$-measurable, then

$$\mathbb{E}[XY \,|\, \mathcal{G}] = X\mathbb{E}[Y \,|\, \mathcal{G}]. \tag{1.0.4}$$

(f)

$$\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}] \,|\, \mathcal{A}] = \mathbb{E}[\mathbb{E}[X \,|\, \mathcal{A}] \,|\, \mathcal{G}] = \mathbb{E}[X \,|\, \mathcal{A}]. \quad \textit{(tower property)}$$

(g) If $\sigma(X)$ and $\mathcal{G}$ are independent, then

$$\mathbb{E}[X \,|\, \mathcal{G}] = \mathbb{E}[X]. \tag{1.0.5}$$

(h)

$$\mathbb{E}[X \,|\, \{\Omega, \emptyset\}] = \mathbb{E}[X].$$

Proof.

(a) The right-hand side of (1.0.3) is $\mathcal{G} - \mathcal{B}$-measurable, and furthermore, due to the linearity of the expectation and the characterising properties of the conditional expecation, we have for any $G \in \mathcal{G}$ that

$$\begin{aligned} \mathbb{E}\big[\big(c\mathbb{E}[X \,|\, \mathcal{G}] + \mathbb{E}[Y \,|\, \mathcal{G}]\big)\mathbb{1}_G\big] &= c\mathbb{E}\big[\mathbb{E}[X \,|\, \mathcal{G}]\mathbb{1}_G\big] + \mathbb{E}\big[\mathbb{E}[Y \,|\, \mathcal{G}]\mathbb{1}_G\big] \\ &= c\mathbb{E}[X\mathbb{1}_G] + \mathbb{E}[Y\mathbb{1}_G] = \mathbb{E}[(cX + Y)\mathbb{1}_G], \end{aligned}$$

which shows that the desired equality.

(b) Define the set $G := \{\mathbb{E}[X \,|\, \mathcal{G}] > \mathbb{E}[Y \,|\, \mathcal{G}]\} \in \mathcal{G}$. Then due to the characterising properties of the conditional expectation as well as its linearity,

$$\mathbb{E}\big[\underbrace{\big(\mathbb{E}[X \,|\, \mathcal{G}] - \mathbb{E}[Y \,|\, \mathcal{G}]\big)\mathbb{1}_G}_{=:Z}\big] = \mathbb{E}\big[\big(\mathbb{E}[X - Y \,|\, \mathcal{G}]\big)\mathbb{1}_G\big] = \mathbb{E}\big[(X - Y)\mathbb{1}_G\big] \le 0,$$

where the last inequality follows from the assumption that $X \le Y$. Since $Z \ge 0$ by definition, the above implies $Z = 0$ $\mathbb{P}$-a.s., and hence $\mathbb{E}[X \,|\, \mathcal{G}] \le \mathbb{E}[Y \,|\, \mathcal{G}]$.

(c) This follows from the linearity and monotonicity of the conditional expectation in combination with $X = X^+ - X^- \le X^+ + X^- = |X|$.

(d) Using the monotonicity of conditional expecations we infer that the monotone limit

$$Z := \lim_{n\to\infty} \mathbb{E}[X_n \,|\, \mathcal{G}]$$

exists $\mathbb{P}$-almost surely, and Z is $\mathcal{G}-\mathcal{B}$-measurable with $Z \le \mathbb{E}[X \,|\, \mathcal{G}]$. Furthermore, for any $G \in \mathcal{G}$ we have by applying the standard Monotone Convergence theorem twice that

$$\int_\Omega Z\mathbb{1}_G \,\mathrm{d}\mathbb{P} = \lim_{n\to\infty} \int_\Omega \mathbb{E}[X_n \,|\, \mathcal{G}]\mathbb{1}_G \,\mathrm{d}\mathbb{P} = \lim_{n\to\infty} \int_\Omega X_n\mathbb{1}_G \,\mathrm{d}\mathbb{P} = \int_\Omega X\mathbb{1}_G \,\mathrm{d}\mathbb{P}.$$

Therefore, Z satisfies the properties of a conditional expectation of X with respect to $\mathcal{G}$, and using the uniqueness result Theorem 1.0.5, we infer $Z = \mathbb{E}[X \,|\, \mathcal{G}]$.

(e) Using the linearity of conditional expectation, w.l.o.g. we can restrict ourselves to the case that $X, Y \ge 0$. By [Dre19, Lemma 2.0.10] there exists a monotone sequence (X_n) of simple non-negative function (X_n) such that $X_n \to X$. In combination with the Monotone Convergence Theorem for conditional expectations we infer the second convergence in

$$X_n\mathbb{E}[Y \,|\, \mathcal{G}] \uparrow X\mathbb{E}[Y \,|\, \mathcal{G}], \quad \mathbb{E}[X_nY \,|\, \mathcal{G}] \uparrow \mathbb{E}[XY \,|\, \mathcal{G}],$$

and thus it will be sufficient to prove the result for the case that X is a simple function. In this case, we can furthermore reduce it to the case $X = \mathbb{1}_G$, $G \in \mathcal{G}$ due to the linearity of conditional expectations. In this case we obtain for any $A \in \mathcal{G}$ that

$$\int_A \mathbb{1}_G\mathbb{E}[Y \,|\, \mathcal{G}] \,\mathrm{d}\mathbb{P} = \int_{A\cap G} \mathbb{E}[Y \,|\, \mathcal{G}] \,\mathrm{d}\mathbb{P} = \int_{A\cap G} Y \,\mathrm{d}\mathbb{P} = \int_A \mathbb{1}_G Y \,\mathrm{d}\mathbb{P},$$

where we used $A \cap G \in \mathcal{G}$ and the defining properties of the conditional expectation in the second equality. Using the uniqueness result Theorem 1.0.5, we infer the desired equation.

(f) The second equality is a consequence of the assumption $\mathcal{A} \subset \mathcal{G}$ and (1.0.4).

Now for arbitrary $A \in \mathcal{A}$ we have

$$\int_\Omega \mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}] \,|\, \mathcal{A}]\mathbb{1}_A \,\mathrm{d}\mathbb{P} = \int_\Omega \mathbb{E}[X \,|\, \mathcal{G}]\mathbb{1}_A \,\mathrm{d}\mathbb{P} = \int_\Omega X\mathbb{1}_A \,\mathrm{d}\mathbb{P},$$

where in the first and second equality we used the definition of the conditional expectation (and in the second equality also the fact that $\mathcal{A} \subset \mathcal{G}$). Using the uniqueness result in Theorem 1.0.5, we infer

$$\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}] \,|\, \mathcal{A}] = \mathbb{E}[X \,|\, \mathcal{A}]$$

as desired.

(g) Exercise.

(h) Exercise.

$\square$

Note that the proof of property (d) works also when the random variables are not integrable but instead non-negative. Observe furthermore that by the tower property (f), for X non-negative but not necessarily integrable we have with $\mathcal{A} = \Omega$

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X],$$

which implies that X is integrable precisely when $\mathbb{E}[X \mid \mathcal{G}]$ is integrable.
We will now turn to deriving further properties that are known for common probabilities and expectations, and extend them to conditional probabilities and expectations.

Theorem 1.0.10 (Dominated convergence theorem)**.** *Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is the underlying probability space. Furthermore, let (X_n) be a sequence in $\mathcal{L}^1$ and let $Y \in \mathcal{L}^1$ such that $|X_n| \leq |Y|$ for all $n \in \mathbb{N}$. Then, if $X_n \to X$ $\mathbb{P}$-a.s. for some random variable X, we have*

$$\lim_{n\to\infty} \mathbb{E}[X_n \,|\mathcal{G}] = \mathbb{E}[X \mid \mathcal{G}] \quad \mathbb{P} - a.s.$$

Proof. Set

$$U_n := \inf_{m\geq n} X_m \leq X_n \leq \sup_{m\geq n} X_m =: V_n.$$

Then, since $-|Y| \leq U_n \uparrow X \leq |Y|$, the MCT for conditional expectations supplies us with

$$\mathbb{E}[U_n \mid \mathcal{G}] \uparrow \mathbb{E}[X \mid \mathcal{G}].$$

Similarly, since $-|Y| \leq V_n \downarrow X \leq |Y|$, the MCT for conditional expectations supplies us with

$$\mathbb{E}[V_n \mid \mathcal{G}] \downarrow \mathbb{E}[X \mid \mathcal{G}].$$

Combining this with $\mathbb{E}[U_n \mid \mathcal{G}] \leq \mathbb{E}[X_n \mid \mathcal{G}] \leq \mathbb{E}[V_n \mid \mathcal{G}]$ (which follows from the monotonicity property of conditional expectations) supplies us with

$$\mathbb{E}[X_n \mid \mathcal{G}] \to \mathbb{E}[X \mid \mathcal{G}],$$

which finishes the proof. $\square$

Theorem 1.0.11 (Conditional Jensen's inequality)**.** *Let $I \subset \mathbb{R}$ be an interval and $\varphi : I \mapsto \mathbb{R}$ a convex function. Then if $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ such that $X(\Omega) \subset I$, if $\varphi(X) \in \mathcal{L}^1$, and if $\mathcal{G}$ is a sub-σ-algebra of $\mathcal{F}$,*

$$\varphi(\mathbb{E}[X \mid \mathcal{G}]) \leq \mathbb{E}[\varphi(X) \mid \mathcal{G}] \quad \mathbb{P} - a.s.$$

Proof.
If φ is an affine function (i.e., $\varphi(x) = ax + b$, for constants $a, b \in \mathbb{R}$), then due to $X \in \mathcal{L}^1$, the linearity of the conditional expectation (Theorem 1.0.9 (a)) supplies us with

$$\varphi(\mathbb{E}[X \mid \mathcal{G}]) = \mathbb{E}[\varphi(X) \mid \mathcal{G}].$$

If φ is not affine, we consider the countable space of affine functions

$$V := \big\{ f : \mathbb{R} \ni x \mapsto ax + b \,:\, a, b \in \mathbb{Q},\, \varphi(x) \geq f(x)\, \forall x \in \mathbb{R} \big\}.$$

Since φ is convex we have the identity

$$\varphi(x) = \sup \big\{ f(x) \,:\, f \in V \big\}, \tag{1.0.6}$$

and we note that as a countable supremum of measurable functions, the left-hand side is measurable again.
Furthermore, Theorem 1.0.9 Parts (b) and (a) imply that for $f \in V$ we have

$$\mathbb{E}[\varphi(X) \,|\, \mathcal{G}] \geq f(\mathbb{E}[X \,|\, \mathcal{G}]) \quad \mathbb{P}-a.s.$$

Taking the (countable) supremum over $f \in V$ in this inequality yields

$$\mathbb{E}[\varphi(X) \,|\, \mathcal{G}] \geq \varphi(\mathbb{E}[X \,|\, \mathcal{G}]) \quad \mathbb{P}-a.s.$$

□

Remark 1.0.12. *With Theorem 1.0.11 at hand, the triangle inequality for conditional expectations would also follow immediately by setting $I = \mathbb{R}$ and $\varphi = |\cdot|$.*

The following result admits a very nice and intuitive interpretation of the conditional expectation for a large class of random variables. In geometric terms, it states that $\mathbb{E}[X \,|\, \mathcal{G}]$ is the orthogonal projection of $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ onto the closed subspace $L^2(\Omega, \mathcal{G}, \mathbb{P})$ of $L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Theorem 1.0.13. *Let $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra of $\mathcal{F}$. Then the functional*

$$L^2(\Omega, \mathcal{G}, \mathbb{P}) \ni Y \mapsto \mathbb{E}[(Y - X)^2] \in [0, \infty)$$

is minimised if and only if $Y = \mathbb{E}[X \,|\, \mathcal{G}]$.

Proof.
We start by noting that $\mathbb{E}[X \,|\, \mathcal{G}] \in L^2$ due to

$$\mathbb{E}\big[\big(\mathbb{E}[X \,|\, \mathcal{G}]\big)^2\big] \leq \mathbb{E}\big[\mathbb{E}[X^2 \,|\, \mathcal{G}]\big] = \mathbb{E}[X^2] < \infty,$$

where we used the conditional Jensen inequality in the first inequality, the tower property in the equality, and the assumption $X \in L^2$ in the last inequality.
Noting that

$$\mathbb{E}\big[(X - \mathbb{E}[X \,|\, \mathcal{G}]) \cdot (\underbrace{\mathbb{E}[X \,|\, \mathcal{G}] - Y}_{:=Z})\big] = 0$$

by Theorem 1.0.9 (f) (since Z is $\mathcal{G} - \mathcal{B}$-measurable) we rewrite

$$\mathbb{E}[(X - Y)^2] = \mathbb{E}\big[(X - \mathbb{E}[X \,|\, \mathcal{G}])^2 + Z^2\big] = \mathbb{E}\big[(X - \mathbb{E}[X \,|\, \mathcal{G}])^2\big] + \mathbb{E}\big[Z^2\big]. \tag{1.0.7}$$

This shows that the left-hand side is minimised if and only if $Y = \mathbb{E}[X \,|\, \mathcal{G}]$. □

As a generalisation of the characterisation of events $A, B \in \mathcal{F}$ with $\mathbb{P}(B) > 0$ as independent via the equation

$$\mathbb{P}(A \,|\, B) = \mathbb{P}(A),$$

and complementing the property of conditional expectations given in (1.0.5), we have a corresponding property for our generalised version of conditional expectations.

Proposition 1.0.14. *Let $\mathcal{A}, \mathcal{G}$ be sub-σ-algebras of $\mathcal{F}$. Then $\mathcal{A}$ and $\mathcal{G}$ are independent if and only if*

$$\mathbb{P}(A \,|\, \mathcal{G}) = P(A) \quad \forall A \in \mathcal{A}. \tag{1.0.8}$$

Proof. It follows from (1.0.5) that if $\mathcal{A}$ and $\mathcal{G}$ are independent, then (1.0.8) holds.
Conversely, if (1.0.8) holds true, then for all $A \in \mathcal{A}$, $G \in \mathcal{G}$,

$$\mathbb{P}(A \cap G) = \mathbb{E}[\mathbb{1}_A \mathbb{1}_G] = \mathbb{E}\big[\mathbb{P}(A \,|\, \mathcal{G})\mathbb{1}_G\big] \overset{(1.0.8)}{=} \mathbb{P}(A)\mathbb{P}(G),$$

so $\mathcal{A}$ and $\mathcal{G}$ are independent. □

Chapter 2

Martingales

We recall Definition 4.1.1 from [Dre19] of a stochastic process.

Definition 2.0.1. *A* stochastic process *is a family* (X_t), $t \in T$, *of random variables mapping from some probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *into a measurable space* $(E, \mathcal{E})$. *Here,* T *is an arbitrary non-empty set.*

An important class of processes is formed by those processes for which the index set T is totally ordered (e.g., $T = \mathbb{R}$).
In this course we will mostly be interested in the case where $T = \mathbb{N}_0$, $T = \mathbb{N}$, or $T = \mathbb{Z}$, and $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$. This suggests that T will in fact often play the role of 'time'. Therefore, we will assume $T = \mathbb{N}_0$ as well as $(E, \mathcal{E}) = (\mathbb{R}, \mathcal{B})$ in the following unless mentioned otherwise.

Definition 2.0.2. *A family* $(\mathcal{F}_t)$, $t \in T$ *is called a* filtration ('Filtration', 'Filtrierung') *if* $(\mathcal{F}_t)$, $t \in T$, *is an increasing family of* σ*-algebras on* Ω *with* $\mathcal{F}_t \subset \mathcal{F}$ *for all* $t \in T$.

Example 2.0.3. *To a stochastic process* (X_t), $t \in T$, *we can associate the filtration* $(\mathcal{F}_t)$, $t \in T$, *where*

$$\mathcal{F}_t := \sigma(X_s \, : \, s \in T, \, s \leq t).$$

It is not hard to check that this actually defines a filtration, and it is usually called the filtration generated by the process (X_t).

If we interpret T as denoting time, then one interpretation is that $\mathcal{F}_t$ contains 'all the information that is available at time t'. In particular, this makes sense if $(\mathcal{F}_t)$ is the canonical filtration since in this case Remark 1.0.7 tells us that any $\mathcal{F}_t - \mathcal{B}$-measurable function can be written as $\varphi(X_1, \ldots, X_t)$, some φ measurable from $(\mathbb{R}^t, \mathcal{B}(\mathbb{R}^t))$ to $(\mathbb{R}, \mathcal{B})$.

Definition 2.0.4. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and* $(\mathcal{F}_t)$, $t \in T$, *be a filtration.*
The quadruple $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in T}))$ *is called a* filtered probability space.
A stochastic process (X_t), $t \in T$, *is called* adapted to $(\mathcal{F}_t)$ *if for each* $t \in T$, *the random variable* X_t *is* $\mathcal{F}_t - \mathcal{B}$*-measurable.*
Similarly, in the special case $T = \mathbb{N}_0$, *a stochastic process* (X_t), $t \in \mathbb{N}_0$, *is called* predictable (or previsible) ('vorhersagbar', 'previsibel') with respect to $(\mathcal{F}_t)$ *if for each* $n \in \mathbb{N}$, *the random variable* X_n *is* $\mathcal{F}_{n-1} - \mathcal{B}$*-measurable.*
In this case it is also sometimes helpful to define

$$\mathcal{F}_\infty := \sigma\Big(\bigcup_{n \in \mathbb{N}_0} \mathcal{F}_n \Big). \tag{2.0.1}$$

Example 2.0.5. *(a) Any process* (X_t) *is adapted to the filtration* $\mathcal{F}_t$ *it generates. Furthermore, the filtration generated by a process is minimal in the sense that for any other filtration* $\mathcal{G}_t$ *on* $(\Omega, \mathcal{F}, \mathbb{P})$ *to which it is adapted, one has* $\mathcal{F}_t \subset \mathcal{G}_t$ *for all* $t \in T$.

(b) If $T = \mathbb{N}_0$, any predictable process is in particular adapted.

Definition 2.0.6. *Let (X_t), $t \in T$, be a stochastic process which is adapted to a filtration $(\mathcal{F}_t)$, $t \in T$. Then (X_t) is called a* martingale ('Martingal') (with respect to $(\mathcal{F}_t)$)[1] *if*

(a)

$$X_t \in \mathcal{L}^1 \quad \forall t \in T; \tag{2.0.2}$$

(b)

$$\mathbb{E}[X_t \,|\, \mathcal{F}_s] = X_s \quad \forall s, t \in T \text{ with } s \le t; \tag{2.0.3}$$

If '=' in (2.0.3) is replaced by '$\ge$' (or '$\le$'), then (X_n) is called a 'submartingale' (or 'supermartingale').

Remark 2.0.7. *In the case $T = \mathbb{N}_0$ (and for $T = \mathbb{N}$, $T = \mathbb{Z}$), on which we will focus for most of this class, instead of checking (2.0.3) it is sufficient to show that*

$$\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] = X_n \quad \forall n \in T. \tag{2.0.4}$$

Indeed, if (2.0.4) holds true, using the tower property we obtain

$$\mathbb{E}[X_{n+2} \,|\, \mathcal{F}_n] = \mathbb{E}[\underbrace{\mathbb{E}[X_{n+2} \,|\, \mathcal{F}_{n+1}]}_{=X_{n+1} \text{ by assumption}} \,|\, \mathcal{F}_n] = X_n \quad \forall n \in T,$$

and inductively one infers

$$\mathbb{E}[X_{n+k} \,|\, \mathcal{F}_n] = X_n \quad \forall k, n \in T,$$

i.e., the validity of (2.0.3).

The most interesting property in the definition of a martingale in Definition 2.0.6 is arguably (2.0.3). Recalling Theorem 1.0.13, which stated that for a random variable with finite second moment the conditional expectation corresponds to the orthogonal projection on the closed subspace $L^2(\Omega, \mathcal{G}, P)$ of $L^2(\Omega, \mathcal{F}, P)$, it means that for $s, t \in T$ with $s < t$, the best approximation to X_t which is $\mathcal{F}_s - \mathcal{B}$ measurable, is X_s itself. This is the reason why martingales are also referred to as *fair games*: Using the present information, i.e., all $\mathcal{F}_s - \mathcal{B}$-measurable random variables, it is not possible to predict anything better about the future than X_s (such as e.g. that the value at a future time would have a tendency to be higher or lower than at present).
The following can also be interpreted as a game where at each time step one is winning or losing one unit with equal probability, and independently of everything else. Hence, a lot of the problems related to it contain the term 'gambling', such as e.g. *the gambler's ruin problem.* We do currently not have the time to consider many of the interesting details concerning these problems, but hope to be able to return to it elsewhere.

Example 2.0.8. *(a) Consider (X_n), $n \in \mathbb{N}$, a sequence of independent random variables such that $\mathbb{E}[X_n] = 0$ for each $n \in \mathbb{N}$ (in particular, this implies $X_n \in \mathcal{L}^1$), as well as the canonical filtration characterised by $\mathcal{F}_n := \sigma(X_1, \dots, X_n)$. Then with $S_n := \sum_{j=1}^n X_j$, the sequence (S_n) is a martingale with respect to the filtration $\mathcal{F}_n$.*

In particular, we have $S_n \in \mathcal{L}^1$ since $\mathbb{E}[X_n] = 0$ by assumption. Furthermore, since the function $\mathbb{R}^n \ni (x_1, \dots, x_n) \mapsto \sum_{j=1}^n x_j$ is continuous, we infer in combination with [Dre19, Corollary 2.3.9] that S_n is $\mathcal{F}_n - \mathcal{B}$-measurable. Last but not least, observe that

$$\mathbb{E}[S_{n+1} \,|\, \mathcal{F}_n] \overset{\text{Thm. 1.0.9 (e) and (a)}}{=} S_n + \mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n].$$

[1]For the origins of the word 'martingale', see http://www.jehps.net/juin2009/Mansuy.pdf.

But now $\sigma(X_{n+1})$ and $\mathcal{F}_n$ are independent due to [Dre19, Proposition 3.2.7], so Theorem 1.0.9 (g) implies that $\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] = \mathbb{E}[X_{n+1}] = 0$, which finishes the proof that (S_n) is a martingale. Therefore, in particular, we consider (S_n) to describe a fair game.

If (S_n) is a martingale, then the process defined via $(S_{n+1} - S_n)$ is usually referred to as a martingale difference sequence*. The above already suggests that a martingale difference sequence can be interpreted as a generalisation of an i.i.d. sequence of random variables.*

(b) Consider any random variable $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, and any filtration $(\mathcal{F}_n)$, $n \in \mathbb{N}$. Then the sequence (X_n) defined via

$$\mathbb{E}[X \,|\, \mathcal{F}_n]$$

is a martingale with respect to $(\mathcal{F}_n)$. We will see below that under certain circumstances we may follow a converse procedure, and given a martingale (X_n) we will be able to obtain a random variable X such that $\mathbb{E}[X \,|\, \mathcal{F}_n] = X_n$.

(c) Let (S_n) be simple random walk and show that the process

$$\cosh(\sigma)^{-n} \exp\left\{\sigma S_n\right\}$$

is a martingale with respect to the canonical filtration.

(d) Describe all stochastic processes on $\{-1, 0, 1\}$ that are martingales. We note first that if (X_n) is a martingale, then it must satisfy

$$\mathbb{P}(X_{n+k} = X_n \,\forall k \geq 0 \,|\, X_n = 1) = 1.$$

If not, we have on $\{X_n = 1\}$ that $\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] = \mathbb{P}(X_{n+1} = X_n \,|\, X_n = 1) - \mathbb{P}(X_{n+1} = -1 \,|\, X_n = 1) < 1 = X_n$, so that X_n can't be a martingale. Similarly we must have that

$$\mathbb{P}(X_{n+k} = X_n \,\forall k \geq 0 \,|\, X_n = -1) = 1.$$

Define now $T = \min\{n \in \mathbb{N}_0 \cup \infty : X_n \in \{-1, 1\}\}$. Then a.s. $X_n = 0, \forall n < T$ and $X_n = X_T, \forall n \geq T$. The first case follows directly from the definition of T. Let now $\mathbb{P}(X_n \neq X_T) > 0$ for some $n > T$. Then, we have that either $\mathbb{P}(X_{n+k} = X_n \,\forall k \geq 0 \,|\, X_n = 1) < 1$ or $\mathbb{P}(X_{n+k} = X_n \,\forall k \geq 0 \,|\, X_n = -1) < 1$, contradicting our earlier observation.

Therefore, we have that the vector (X_T, T) completely describes all possible martingales on $\{-1, 0, 1\}$. The sole constriction of this vector on $\mathbb{N}_0 \cup \{\infty\}\} \times \{-1, 1\}$ is that

$$\mathbb{P}(X_T = 1 \,|\, T = n) = \mathbb{P}(X_T = -1 \,|\, T = n) = \frac{1}{2} \,\forall n \in \mathbb{N}.$$

To see why, note that on $\{X_n = 0\}$ it must hold, that $\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] = 0$. At the same time, we must have that

$$\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] = \mathbb{P}(T = n+1, X_{n+1} = 1) - \mathbb{P}(T = n+1, X_{n+1} = -1)$$

and therefore $\mathbb{P}(X_T = 1, T = n) = \mathbb{P}(X_T = -1, T = n) = \frac{1}{2}\mathbb{P}(T = n)$.

Finally, we must show that every such pair represents a martingale. Let therefore (T, X_T) be a vector satisfying the above restriction. The corresponding stochastic process is then integrable (since it is bounded from above and below by 1) and adapted (since all quantities needed to consider X_n depend only on X_m for $m \leq n$. Finally, $\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] = X_n$, since on $\{T \leq n\}$ we have that $X_{n+1} = X_n = X_T$ almost surely by definition, on $\{T = n+1\}$ the restriction of (T, X_T) gives that $\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] = 0 = X_n$ and on $\{T > n+1\}$ we have that $X_{n+1} = X_n = 0$ almost surely.

Remark 2.0.9. *(a) If we do not explicitly mention the filtration of a martingale (X_n), we will usually assume that $\mathcal{F}_n := \sigma(X_1, \dots, X_n)$.*

(b) It might look puzzling at a first glance that we call the process a super*martingale if in fact we have that $\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n]$ is* smaller *than X_n. If, for example, we choose the i.i.d. variables X_n in Example 2.0.8 in such a way that $\mathbb{P}(X_1 = 1) = p$ and $\mathbb{P}(X_1 = -1) = 1 - p$ for some $p \in (0,1)\backslash\{\frac{1}{2}\}$, then we call S_n a* random walk with drift*; we will choose $p \in (0, \frac{1}{2})$ for the rest of this remark for the sake of definiteness. Everything goes through as before, but we obtain $\mathbb{E}[S_{n+1} \,|\, \mathcal{F}_n] = S_n + (2p-1) < S_n$. Hence, in this case we see that (S_n) is a supermartingale; if for instance one interprets (S_n) as the evolution of a sequence of biased coin tosses such that for heads (say $X_n = 1$) one wins one unit in the n-th step, whereas if it shows tails (say $X_n = -1$) one loses one unit, then on average things are going down which is far from being super.*

Indeed, the reason (S_n) is called a supermartingale in this case is the following. In analysis, a function $f \in C^2(\mathbb{R}^d, \mathbb{R})$ is called harmonic *if*

$$\Delta f(x) = 0 \quad \forall x \in \mathbb{R}^d. \tag{2.0.5}$$

[2] *In this case, it can be shown that for all $r > 0$ and $x \in \mathbb{R}^d$, f satisfies the* mean value property

$$f(x) = \frac{1}{|B(x,r)|} \int_{B(x,r)} f(y) \,\mathrm{d}y = \frac{1}{\sigma(\partial B(x,r))} \int_{\partial B(x,r)} f(y) \,\sigma(\mathrm{d}y) \tag{2.0.6}$$

where $B(x,r) := \{y \in \mathbb{R}^d : \|x-y\|_2 \le r\}$. Furthermore, f is called superharmonic *(*subharmonic*, respectively), if the equality in (2.0.5) is replaced by $\le$ ($\ge$, respectively). In this case, the equality in (2.0.6) can be replaced by $\ge$ ($\le$, respectively).*

Exercise 2.0.10. *Let $\varphi \in C^2(\mathbb{R}^d, \mathbb{R})$ be subharmonic, and let (X_n) be an i.i.d. sequence of random variables which are uniformly distributed on either $\partial B(0,1)$ or otherwise $B(0,1)$. Show that with $S_n := \sum_{i=1}^n X_i$ the sequence $\varphi(S_n)$ is a submartingale with respect to the canonical filtration $\mathcal{F}_n := \sigma(X_1, \dots, X_n)$.*

Theorem 2.0.11. *(a) Let (X_t) be a martingale with respect to a filtration $(\mathcal{F}_t)$ and $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function. If*

$$\mathbb{E}[\varphi(X_t)^+] < \infty \quad \forall t \in T, \tag{2.0.7}$$

then $\varphi(X_t)$ is a submartingale with respect to $(\mathcal{F}_t)$.

(b) Let (X_t) be a submartingale with respect to a filtration $(\mathcal{F}_t)$ and $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex and non-decreasing function. If (2.0.7) continues to hold, then $\varphi(X_t)$ is a submartingale with respect to $(\mathcal{F}_t)$.

Proof. (a) We start with noting that since (X_t) is a process adapted to $(\mathcal{F}_t)$, so is the process $(\varphi(X_t))$, since as a convex function φ is $\mathcal{B} - \mathcal{B}$-measurable (as follows e.g. from the representation (1.0.6)).

Furthermore, since φ is convex, we can find $a, b \in \mathbb{R}$ such $\varphi(x) \ge ax + b$ for all $x \in \mathbb{R}$, and so $\varphi(x)^- \le (ax+b)^-$. Thus, we deduce

$$\mathbb{E}[\varphi(X_t)^-] \le \mathbb{E}[(aX_t + b)^-] \le |b| + |a|\mathbb{E}[|X_t|] < \infty.$$

[2] This is in fact consistent with the terminology for (discrete) harmonic functions from [Dre18], where for the transition matrix P of a Markov chain we called a function f harmonic if $(P - I)f = 0$. The operator $P - I$ (or the one given by $\lim_{t \downarrow 0}(P^t - \mathrm{Id})/t$ in the context of continuous time) will be called the *generator* (see also Def. 3.3.1 below) in the context of Markov processes, and if time admits we will see that $\frac{1}{2}\Delta$ is the generator of Brownian motion.

In combination with (2.0.7), this implies

$$\mathbb{E}[|\varphi(X_t)|] < \infty, \quad i.e., \quad \varphi(X_t) \in \mathcal{L}^1$$

for all $t \in T$ (cf. (2.0.2)).

Therefore, the conditional Jensen inequality of Theorem 1.0.11 supplies us with

$$\mathbb{E}[\varphi(X_t) \,|\, \mathcal{F}_s] \geq \varphi\big(\mathbb{E}[X_t \,|\, \mathcal{F}_s]\big) = \varphi(X_s), \tag{2.0.8}$$

which finishes the proof.

(b) The proof proceeds in exactly the same way as that of the first part, with the equality in (2.0.8) replaced by '$\geq$' since φ is non-decreasing and $\mathbb{E}[X_t \,|\, \mathcal{F}_s] \geq X_s$.

□

Exercise 2.0.12. *Find an example which shows that part (b) of Theorem 2.0.11 does not generally hold if one does without the assumption of φ being non-decreasing.*

Proposition 2.0.13. *(a) (X_n) is a submartingale if and only if $(-X_n)$ is a supermartingale.*

(b) If (X_n) and (Y_n) are submartingales, then so is $(X_n \vee Y_n)$.

Proof. (a) (X_n) has the desired integrability and measurability properties if and only if $(-X_n)$ does. As regards to the conditional expectations, we have

$$\mathbb{E}[X_t \,|\, \mathcal{F}_s] \geq X_s \quad \text{iff} \quad \mathbb{E}[-X_t \,|\, \mathcal{F}_s] \leq -X_s$$

due to the linearity of conditional expectations.

(b) As a maximum of integrable and $\mathcal{F}_n - \mathcal{B}$-measurable random variables we have that $X_n \vee Y_n$ is integrable and $\mathcal{F}_n - \mathcal{B}$-measurable. Furthermore, we have, using the monotonicity of conditional expectation as well as the definition of a submartingale, for $s, t \in T$ with $s \leq t$, that

$$\mathbb{E}[X_t \vee Y_t \,|\, \mathcal{F}_s] \geq \mathbb{E}[X_t \,|\, \mathcal{F}_s] \vee \mathbb{E}[Y_t \,|\, \mathcal{F}_s] \geq X_s \vee Y_s.$$

□

2.0.1 Discrete stochastic integral

We will now get back to the gambling interpretation of martingales in more detail. Indeed, since times at which one can trade at a stock exchange are discrete anyway, the notion of a discrete stochastic integral suggests itself as a suitable candidate for modeling this situation. At each time $t \in \mathbb{N}$ the investor has to decide how much to invest in a certain stock from time t to $t+1$. The process describing this investment should be adapted to the canonical filtration of the stock price, since the investor should not be able to take advantage of any information other than the past up to the present of the stock prices – in particular, she should not be able to consider any information on future stock prices in that decision.

Definition 2.0.14. *Denote by (X_t), (Y_t), $t \in \mathbb{N}_0$, two processes adapted to a filtration $(\mathcal{F}_t)$. Then the* stochastic integral *$((Y \bullet X)_n)$, $n \in \mathbb{N}_0$, is defined via $(Y \bullet X)_0 := 0$, and*

$$(Y \bullet X)_n := \sum_{k=0}^{n-1} Y_k \cdot (X_{k+1} - X_k), \quad n \in \mathbb{N}.$$

If (X_t) is a martingale, then the process $((Y \bullet X)_n)$ is also referred to as a martingale transform.

It is easy to check that the stochastic integral defined in Definition 2.0.14 is adapted to the filtration $(\mathcal{F}_t)$.

Theorem 2.0.15. *(a) The process (X_t), $t \in \mathbb{N}_0$, from Definition 2.0.14 is a martingale if and only if for any adapted process (Y_t), $t \in \mathbb{N}_0$, such that each Y_t is a bounded random variable, the stochastic integral $((Y \bullet X)_t)$, $t \in \mathbb{N}_0$ is a martingale.*

(b) The process (X_t), $t \in \mathbb{N}_0$, from Definition 2.0.14 is a sub-/ supermartingale if and only if for any adapted process (Y_t), $t \in \mathbb{N}_0$, with $Y_t \geq 0$ and Y_t bounded for each $t \in \mathbb{N}_0$, the stochastic integral $((Y \bullet X)_t)$, $t \in \mathbb{N}_0$ is a sub-/ supermartingale, respectively.

Proof. (a) Let (X_t) be a martingale. Then we have

$$\begin{aligned}\mathbb{E}\big[(Y \bullet X)_{n+1} \,|\, \mathcal{F}_n\big] &= \mathbb{E}\big[(Y \bullet X)_n + Y_n(X_{n+1} - X_n) \,|\, \mathcal{F}_n\big] \\ &\overset{(1.0.4)}{=} (Y \bullet X)_n + Y_n \underbrace{\mathbb{E}\big[(X_{n+1} - X_n) \,|\, \mathcal{F}_n\big]}_{=0,\text{ since } (X_n) \text{ martingale}} = (Y \bullet X)_n,\end{aligned}$$

so $((Y \bullet X)_n)$ is a martingale.

Conversely, assume that for any (Y_n) as in the assumptions, $((Y \bullet X)_n)$, $n \in \mathbb{N}_0$, is a martingale. Then for any $n \in \mathbb{N}$, setting $Y_n := 1$ as well as $Y_m = 0$ for $m \neq n$, we have that

$$0 = (Y \bullet X)_{n-1} = \mathbb{E}[(Y \bullet X)_n \,|\, \mathcal{F}_{n-1}] = \mathbb{E}[X_n - X_{n-1} \,|\, \mathcal{F}_{n-1}] = \mathbb{E}[X_n \,|\, \mathcal{F}_{n-1}] - X_{n-1},$$

which shows that (X_n) is a martingale (with respect to $(\mathcal{F}_n)$).

(b) The proof is similar to that of the martingale case and is left as an exercise.

$\square$

Example 2.0.16. *Let (X_n) be an i.i.d. sequence of Rademacher distributed random variables and set $S_n := \sum_{i=1}^n X_i$, with $S_0 := 0$. Then (S_n) defines a martingale (simple random walk) with respect to the canonical filtration $\mathcal{F}_n := \sigma(S_1, \dots, S_n) = \sigma(X_1, \dots, X_n)$, and we can define an adapted process (H_n) via $H_0 := 1$ and recursively*

$$H_n := (H \bullet S)_n \quad \text{for } n \geq 1.$$

In particular, note that the definition of $(H \bullet S)_n$ only requires us to know the values of (H_k), $k \in \{0, \dots, n-1\}$, so H_n is well-defined. This stochastic integral describes the value of a portfolio for which at each time step from n to $n+1$ we are entirely invested, and the portfolio either doubles in value (if $X_{n+1} = 1$) or becomes worthless (if $X_{n+1} = -1$), until we're bankrupt.[3] The previous result then tells us that the value of the portfolio is a martingale.

The interpretation of this Theorem 2.0.15 is as follows: If you invest in a process constituting a martingale, and if you do so by just using the information that is available at present, then you can't beat the system in the sense that the resulting portfolio will be a martingale again.
Similarly, if you invest (non-negative capital) in a process constituting a sub- / supermartingale, then the resulting portfolio process will be a sub- / supermartingale again.
Given this possibly disappointing result, you might still ask if you can leave the investment game at a certain point in time (e.g. at the first time your capital exceeds a given amount) in order to 'beat the system'. We will investigate this question below in the section of stopping times; before, however, we introduce a representation of martingales which will be useful in what is about to come.

[3] I'm not a financial mathematician, but I guess it seems unrealistic to assume that the value of a stock becomes negative; so instead you might think of the value of a so-called future contract.

2.0.2 Doob decomposition

The following is the first theorem in our series of results attributed to Joseph Doob (see https://en.wikipedia.org/wiki/Joseph_L._Doob).

Theorem 2.0.17 (Doob decomposition). *Let (X_t), $t \in \mathbb{N}_0$, be an adapted process such that $X_t \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ for all $t \in \mathbb{N}_0$.*

(a) Then there exists a martingale (M_t), $t \in \mathbb{N}_0$, and a previsible process (A_t), $t \in \mathbb{N}_0$, with $A_0 = 0$ such that

$$X_n = M_n + A_n, \quad \forall n \in \mathbb{N}_0. \tag{2.0.9}$$

Furthermore, (M_n) and (A_n) are $\mathbb{P}$-a.s. uniquely determined via (2.0.9), i.e., if $(\widetilde{M}_t)$, $t \in \mathbb{N}_0$, is a martingale and $(\widetilde{A}_t)$, $t \in \mathbb{N}_0$, a previsible process with $\widetilde{A}_0 = 0$ such that

$$X_n = \widetilde{M}_n + \widetilde{A}_n, \quad \forall n \in \mathbb{N}_0.$$

then $\mathbb{P}(A_n = \widetilde{A}_n, M_n = \widetilde{M}_n \, \forall n \in \mathbb{N}_0) = 1$.

(b) Furthermore, (X_n) is a submartingale if and only if the process (A_n) is increasing, i.e., $\mathbb{P}(A_n \le A_{n+1}) = 1$ for all $n \in \mathbb{N}_0$.

Proof. (a) The canonical way to extract a martingale contribution from (X_n) is to define

$$M_n := X_0 + \sum_{i=1}^{n} (X_i - \mathbb{E}[X_i \,|\, \mathcal{F}_{i-1}]), \quad n \in \mathbb{N}_0, \tag{2.0.10}$$

which by definition is a martingale with respect to $(\mathcal{F}_n)$. To make up for the fact that M_n is generally not yet equal to X_n, we set

$$A_n := X_n - M_n, \quad n \in \mathbb{N},$$

as well as $A_0 := 0$, which in combination with (2.0.10) yields at once that (A_n) is previsible, and also the identity (2.0.9).

In order to show the $\mathbb{P}$-a.s. uniqueness, we observe that for processes $(\widetilde{M}_n)$ and $(\widetilde{A}_n)$ as in the assumptions, the process defined via

$$M_n - \widetilde{M}_n = A_n - \widetilde{A}_n$$

is a previsible martingale and, using $A_0 = \widetilde{A}_0 = 0$, equal to 0 (exercise), which yields the desired uniquness.

(b) Using the first part we have

$$\mathbb{E}\big[X_{n+1} - X_n \,|\, \mathcal{F}_n\big] = \mathbb{E}\big[M_{n+1} + A_{n+1} - M_n - A_n \,|\, \mathcal{F}_n\big] = A_{n+1} - A_n,$$

which shows that (X_n) is a submartingale if and only if (A_n) is increasing.

□

2.0.3 Quadratic variation

Definition 2.0.18. *Let (M_n) be a square integrable martingale, i.e., a martingale with $M_n \in \mathcal{L}^2$ for all $n \in T$. The previsible process (A_n) for which we have that $A_0 = 0$ and that*

$$(M_n^2 - A_n)$$

is a martingale, is called the quadratic variation process ('quadratischer Variationsprozess') *or* angle bracket process *of (M_n), and it is also denoted by $(\langle M \rangle_n)$.*

We observe that due to Theorem 2.0.17, the process (A_n) exists and is $\mathbb{P}$-a.s. unique. Furthermore, due to Theorem 2.0.11, in the context of Definition 2.0.18, the process (M_n^2) is a submartingale and hence, using Theorem 2.0.17 again, the process $(\langle M\rangle_n)$ is non-decreasing.
Also, note that with (M_n) a square integrable martingale, for $m,n \in \mathbb{N}$ with $m < n$ Cauchy-Schwarz supplies us with $(M_{n+1}-M_n)(M_{m+1}-M_m) \in \mathcal{L}^1$, so in combination with $\mathbb{E}[M_{n+1}-M_n] = \mathbb{E}[M_{m+1}-M_m] = 0$ Theorem 1.0.9 (e) immediately implies

$$\mathbb{E}[(M_{n+1}-M_n)(M_{m+1}-M_m)\,|\,\mathcal{F}_n] = (M_{m+1}-M_m)\mathbb{E}[(M_{n+1}-M_n)\,|\,\mathcal{F}_n] = 0,$$

and in particular, taking expectations in the above, we deduce that $M_{n+1}-M_n$ and $M_{m+1}-M_m$ are uncorrelated.
Furthermore, from the fact that $(M_n^2 - \langle M\rangle_n)$ is a martingale we infer that

$$\langle M\rangle_n = \sum_{i=1}^n \mathbb{E}[M_i^2\,|\,\mathcal{F}_{i-1}] - M_{i-1}^2 = \sum_{i=1}^n \mathbb{E}[(M_i - M_{i-1})^2\,|\,\mathcal{F}_{i-1}]. \tag{2.0.11}$$

This justifies that $(\langle M\rangle_n)$ can be interpreted as a pathwise measurement of the conditional variances of the trajectories of (M_n), and in this interpretation $\langle M\rangle_\infty := \lim_{n\to\infty}\langle M\rangle_n$ is the total conditional variance along a path.
Furthermore, from (2.0.11) we obtain by taking expectations

$$\mathbb{E}[\langle M\rangle_n] = \sum_{i=1}^n \mathbb{E}[(M_i - M_{i-1})^2] = \mathrm{Var}(M_n - M_0). \tag{2.0.12}$$

Example 2.0.19. *Let $X \in \mathcal{L}^2$, and for a sub-σ-algebra $\mathcal{G}$ of $\mathcal{F}$ define the* conditional variance of X given G *as*

$$\mathrm{Var}(X\,|\,\mathcal{G}) := \mathbb{E}\big[X^2 - \mathbb{E}[X\,|\,\mathcal{G}]^2\,|\,\mathcal{G}\big].$$

For a filtration $(\mathcal{F}_n)$ we set $M_n := \mathbb{E}[X\,|\,\mathcal{F}_n]$. Then the claim is that

$$\langle M\rangle_n = \underbrace{\mathrm{Var}(X) - \mathrm{Var}(X\,|\,\mathcal{F}_{n-1})}_{=:A_n}. \tag{2.0.13}$$

W.l.o.g. assume that $\mathbb{E}[X] = 0$. We start with observing that (A_n) defines a previsible process with $A_0 = 0$ a.s. Furthermore, using the (conditional) orthogonality of the increments in the first equality,

$$\begin{aligned}
&\mathbb{E}[M_{n+1}^2 - A_{n+1}\,|\,\mathcal{F}_n] \\
&= \mathbb{E}[(M_{n+1}-M_n)^2 + M_n^2 - (\mathrm{Var}(X) - \mathrm{Var}(X\,|\,\mathcal{F}_n))\,|\,\mathcal{F}_n] \\
&= \mathbb{E}[(\mathbb{E}[X\,|\,\mathcal{F}_{n+1}] - \mathbb{E}[X\,|\,\mathcal{F}_n])^2 - (\mathrm{Var}(X\,|\,\mathcal{F}_{n-1}) - \mathrm{Var}(X\,|\,\mathcal{F}_n)\,|\,\mathcal{F}_n] \\
&\quad + M_n^2 - \underbrace{(\mathrm{Var}(X) - \mathrm{Var}(X\,|\,\mathcal{F}_{n-1}))}_{=A_n} \\
&= M_n^2 - A_n.
\end{aligned}$$

2.1 Stopping times

It will turn out useful to consider stochastic processes not only at fixed or deterministic times, but also at random times. As a motivation one might consider the investing interpretation of the discrete stochastic integral again. As an example, you can consider the strategy that sells the stock once it exceeds a certain value for the first time. We will see later on (see the optional stopping theorem, Theorem 2.1.14 below) that, under suitable assumptions, if the underlying process is a martingale, one cannot 'beat' the market using such strategies.
In order to have a somehow compatible notion of random times, we introduce the following definition.

Definition 2.1.1. *A random variable* $\tau : (\Omega, \mathcal{F}, \mathbb{P}) \to T \cup \{\infty\}$ *is called a* stopping time (with respect to a filtration $(\mathcal{F}_t)$, $t \in T$), *if for each* $t \in T$,

$$\{\tau \le t\} \in \mathcal{F}_t. \tag{2.1.1}$$

Intuitively, this means that in order to decide whether or not the process is 'stopped' by time t, you are only allowed to take advantage of all the information that is available at time t; in particular, you're not allowed to peek into the future!
The condition given in (2.1.1) is to be interpreted as follows: If we consider T as time, then (2.1.1) says that in order to decide whether or not $\tau \le t$ occurs, we only need to know the 'information' provided by $\mathcal{F}_t$. In the important case of $\mathcal{F}_t = \sigma(\{X_s : s \le t\})$ this means that whether or not $\tau \le t$ occurs can be judged just on the basis of the X_s, $s \le t$.

Example 2.1.2. *(a) For any process* (X_t) *adapted to some filtration* $(\mathcal{F}_t)$*, for any* $t \in T$*, the constant random variable* $\tau := t$ *is a stopping time.*

(b) Consider random walk (S_n) *as in Example 2.0.8, but with the assumption* $\mathbb{E}[X_n] = 0$ *replaced by* $\mathbb{E}[X_n] \in \mathbb{R}\backslash\{0\}$ *and* $\operatorname{Var}(X_n) > 0$*. For* $A \in \mathcal{B}$ *we can define the so-called* first entrance time to the set A *via*

$$\tau_A := \inf\{n \in T : S_n \in A\}.$$

Then τ_A *is a stopping time. Indeed, we have for any* $n \in \mathbb{N}_0$ *that*

$$\{\tau_A \le n\} = \bigcup_{k=0}^{n} \underbrace{\{S_k \in A\}}_{\in \mathcal{F}_k \subset \mathcal{F}_n},$$

which proves that τ_A *is a stopping time.*

Now consider the case $A = (1, \infty)$*. Show that*

(i) if $\mathbb{E}[X_n] > 0$*, then* $\mathbb{P}(\tau_A < \infty) = 1$*;*

(ii) if $\mathbb{E}[X_n] < 0$*, then* $\mathbb{P}(\tau_A = \infty) > 0$*.*

A popular interpretation is that of e^{S_t} *modeling the price of a stock over time. In particular, note that under suitable integrability assumptions (such as assuming* $\mathbb{P}(X_n = 1) = \mathbb{P}(X_n = -1) = \frac{1}{2}$ *for the underlying random variables for the sake of simplicity), due to Theorem 2.0.11 the stock price* (e^{S_t})*,* $t \in T$*, is a submartingale. This corresponds to the fact that in the long run stocks are expected to outperform investments in the money market since they are more risky, and investors have to have a benefit for taking risks.*

Lemma 2.1.3. *A random variable* $\tau : (\Omega, \mathcal{F}, \mathbb{P}) \to T \cup \{\infty\}$ *is a stopping time if and only if for all* $n \in T$ *(recall that* T *is assumed to be discrete) we have*

$$\{\tau = n\} \in \mathcal{F}_n \quad \forall n \in T.$$

Proof. Exercise □

Exercise 2.1.4. *Show by giving a counterexample that the equivalence given in Lemma 2.1.3 does not apply if* T *is chosen more generally as before (such that* $T = [0, \infty)$*).*
Consider the following setting: $\Omega = [0, 1]$*,* $\mathcal{F} = \mathcal{B}([0, 1])$*, and let* $T \subset [0, 1]$ *be such that* $T \notin \mathcal{F}$*. Furthermore, set* $X_t = \mathrm{id} : \Omega \to \Omega$ *for all* $t \in T$ *and consider the filtration given by* $\mathcal{F}_t = \mathcal{F}$ *for all* $t \in T$*. Define the random variable* $\tau := \inf\{t \in T : X_t = t\}$*. Then:*

- *The process* (X_t)*,* $t \in T$*, is adapted;*

- *The random variable τ is a* not *a stopping time: Indeed, we have*

$$\{\tau \leq 1\} = T \notin \mathcal{F}_1.$$

- $$\{\tau = t\} = \{t\} \in \mathcal{F}_t \quad \forall t \in T.$$

Lemma 2.1.5. *Let τ and σ be two stopping times.*

(a) The minimum $\sigma \wedge \tau$ and the maximum $\sigma \vee \tau$ are stopping times.

(b) If e.g. $T = \mathbb{Z}$, or $T = \mathbb{N}_0$, then $\sigma + \tau$ is a stopping time.

Proof. (a) We have

$$\{\sigma \wedge \tau \leq n\} = \underbrace{\{\sigma \leq n\}}_{\in \mathcal{F}_n} \cup \underbrace{\{\tau \leq n\}}_{\in \mathcal{F}_n},$$

and hence the right-hand side is in $\mathcal{F}_n$.

(b) We have

$$\{\sigma + \tau \leq n\} = \bigcup_{j \in T, j \leq n} \bigcup_{k \in T, k \leq j} \{\sigma = j - k\} \cap \{\tau = k\},$$

and all the sets on the right-hand side are contained in $\mathcal{F}_n$ due to Lemma 2.1.3 which proves the claim.

□

Exercise 2.1.6. *Give an example to show that in the setting of the above lemma, $\tau - \sigma$ is generally* not *a stopping time.*

Definition 2.1.7. *Let τ be a stopping time with respect to a filtration $(\mathcal{F}_t)_{t \in T}$. We can define the σ-algebra*

$$\mathcal{F}_\tau := \big\{F \in \mathcal{F} \, : \, F \cap \{\tau \leq n\} \in \mathcal{F}_n\big\},$$

which is commonly referred to as the σ-algebra of the τ-past ('σ-Algebra der τ-Vergangenheit')*.*

Exercise 2.1.8. *Show that $\mathcal{F}_\tau$ as defined above actually defines a σ-algebra.*

Remark 2.1.9. *Note that, using Lemma 2.1.3 we immediately infer that for $F \in \mathcal{F}_\tau$ we also have that $F \cap \{\tau = n\} \in \mathcal{F}_n$ for all $n \in T$.*

One of the principal reasons for introducing it is the formulation of the optional sampling theorem below. We will take advantage of the following simple properties of $\mathcal{F}_\tau$.

Lemma 2.1.10. *Let σ and τ be stopping times.*

(a) If $\sigma \leq \tau$, then

$$\mathcal{F}_\sigma \subset \mathcal{F}_\tau.$$

(b)

$$\mathcal{F}_{\sigma \wedge \tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau.$$

Proof. (a) Let $F \in \mathcal{F}_\sigma$. Then, since $\sigma \leq \tau$, we have

$$F \cap \{\tau \leq t\} = \underbrace{(F \cap \{\sigma \leq t\})}_{\in \mathcal{F}_t} \cap \underbrace{\{\tau \leq t\}}_{\in \mathcal{F}_t} \in \mathcal{F}_t,$$

which proves the desired inclusion.

(b) Exercise

□

One of the main reasons for introducing the σ-algebra of the τ-past, which looks slightly unnatural and technical at a first glance, is given by the following lemma.

Lemma 2.1.11. *Let (X_t) be an adapted process and let $\tau < \infty$ be a stopping time. Then $\omega \mapsto X_{\tau(\omega)}(\omega)$ defines a $\mathcal{F}_\tau - \mathcal{B}$-measurable random variable.*

Proof. Indeed, for $B \in \mathcal{B}$ we can write

$$\{X_\tau \in B\} \cap \{\tau \le t\} = \bigcup_{s \le t} \{X_s \in B\} \cap \{\tau = s\}.$$

Since all the sets occurring on the right-hand side are contained in $\mathcal{F}_t$, and since the union on the right-hand side is countable, we infer that the left-hand side is contained in $\mathcal{F}_t$ as well. □

In particular, this lemma implies that for any adapted process (X_t), any $B \in \mathcal{B}$, and any stopping time τ, one has $\{X_\tau \in B\} \in \mathcal{F}_\tau$.
One of the fundamental results in the theory of martingales is that the martingale property is preserved under the evaluation at stopping times (under certain assumptions).

Theorem 2.1.12 (Optional sampling theorem)**.** *Let (X_t), $t \in T$, with T countable be a (sub-, super-) martingale with respect to a filtration $(\mathcal{F}_t)$, and assume that σ and τ are stopping times with $\sigma \le \tau \le N$, some $N \in \mathbb{N}$. Then (X_σ, X_τ) is an integrable process adapted to $(\mathcal{F}_\sigma, \mathcal{F}_\tau)$, and*

$$\mathbb{E}[X_\tau \,|\, \mathcal{F}_\sigma] = X_\sigma \tag{2.1.2}$$

(with '$=$' replaced by '$\ge$' and '$\le$', respectively, in the case of a sub- and supermartingale).

Proof. We give the proof in the case of (X_t) being a submartingale; the case of a supermartingale follows by considering $(-X_t)$, and the case of a martingale follows from combining the corresponding results for the sub- and supermartingale.
Adaptedness follows from Lemma 2.1.11, and the integrability follows from $|X_\sigma| \vee |X_\tau| \le \max_{s \in T,\, s \le N} |X_s|$ and the fact that each X_s is integrable.
It remains to show (2.1.2), and for this purpose we start with showing that if (M_t) is a martingale, then

$$\mathbb{E}[M_N \,|\, \mathcal{F}_\tau] = M_\tau. \tag{2.1.3}$$

Indeed, the right-hand side of (2.1.3) is an $\mathcal{F}_\tau - \mathcal{B}$-measurable random variable due to Lemma 2.1.11, and furthermore, for any $F \in \mathcal{F}_\tau$ we have

$$\begin{aligned}
\mathbb{E}[M_\tau \mathbb{1}_F] &= \sum_{s \in T,\, s \le N} \mathbb{E}[M_s \mathbb{1}_{\tau = s} \mathbb{1}_F] = \sum_{s \in T,\, s \le N} \mathbb{E}\big[\mathbb{E}[M_N \,|\, \mathcal{F}_s] \mathbb{1}_{\{\tau = s\} \cap F}\big] \\
&= \sum_{s \in T,\, s \le N} \mathbb{E}\big[M_N \mathbb{1}_{\{\tau = s\} \cap F}\big] = \mathbb{E}\big[M_N \mathbb{1}_F\big],
\end{aligned}$$

where in the second equality we used the martingale property, in the third equality we took advantage of the fact that $\{\tau = s\} \cap F \in \mathcal{F}_s$, cf. Remark 2.1.9, and the tower property again. In the last equality we used $\tau \le N$. This establishes (2.1.3).
Now for (X_t) a submartingale the Doob decomposition (Theorem 2.0.17) yields

$$X_t = M_t + A_t,$$

with (M_t) a martingale and (A_t) a previsible non-decreasing process with $A_0 = 0$. Hence, we can deduce

$$\mathbb{E}[X_\tau \,|\, \mathcal{F}_\sigma] = \mathbb{E}[M_\tau + A_\tau \,|\, \mathcal{F}_\sigma] = \mathbb{E}[\mathbb{E}[M_N \,|\, \mathcal{F}_\tau] \,|\, \mathcal{F}_\sigma] + \mathbb{E}[A_\tau \,|\, \mathcal{F}_\sigma] \ge M_\sigma + A_\sigma = X_\sigma,$$

where we took advantage of (2.1.3) in the second equality and the inequality, and in the inequality we also used the tower property and the fact that (A_t) is non-decreasing. □

Example 2.1.13.
In order to see that the previous result is not valid if one drops the boundedness assumptions on σ and τ, one can e.g. consider the following: Let $X_1, X_2, \dots$ be a sequence of independent random variables such that

$$\mathbb{P}(X_n = 2^n) = \mathbb{P}(X_n = -2^n) = \frac{1}{2}.$$

Then $S_n := \sum_{i=1}^n X_i$ defines a martingale with respect to the canonical filtration $(\mathcal{F}_n)$. Define the random time $\tau_+ := \inf\{n \in \mathbb{N} : S_n > 0\}$. Then τ_+ is a stopping time with respect to the canonical filtration $(\mathcal{F}_n)$, and

$$\mathbb{P}(\tau_+ < \infty) = 1,$$

and

$$S_{\tau_+} = 2.$$

In particular, (S_1, S_{τ_+}) is not a martingale with respect to $(\mathcal{F}_1, \mathcal{F}_{\tau_+})$.
Thus, the implication of the optional sampling theorem is violated, and the reason it cannot be applied is that the stopping time τ_+ is not bounded.

Theorem 2.1.14 (Optional stopping theorem)**.** *Let (X_t), $t \in T$, be a (sub-, super-) martingale with respect to a filtration $\mathcal{F}_t$, and let τ be a stopping time. Then $(X_{t\wedge\tau})$, $t \in T$, is a (sub-, super-) martingale with respect to $(\mathcal{F}_t)$ as well as with respect to $(\mathcal{F}_{\tau\wedge t})$.*

Proof. As in the proof of Theorem 2.1.12, we will give the proof for the case of (X_t) being a submartingale only.
We first observe that $X_{t\wedge\tau} \in \mathcal{L}^1$, and due to Lemma 2.1.11 we deduce that $X_{t\wedge\tau}$ is $\mathcal{F}_{t\wedge\tau} - \mathcal{B}$-measurable (and hence also $\mathcal{F}_t - \mathcal{B}$-measurable).
Thus, it remains to show the corresponding inequalities for the conditional expectations.

$$\mathbb{E}[X_{(t+1)\wedge\tau} \,|\, \mathcal{F}_{t\wedge\tau}] \geq X_{t\wedge\tau}, \quad \forall t \in T,$$

follows from Theorem 2.1.12, and hence $(X_{t\wedge\tau})$ is a submartingale with respect to $(\mathcal{F}_{t\wedge\tau})$. For showing that it is a submartingale with respect to $(\mathcal{F}_t)$ also, we
first observe that on $\{\tau \leq t\} \in \mathcal{F}_t$ we have

$$X_{(t+1)\wedge\tau} = X_{t\wedge\tau}.$$

As a consequence,

$$\begin{aligned}\mathbb{E}[(X_{(t+1)\wedge\tau} - X_{t\wedge\tau}) \,|\, \mathcal{F}_t] &= \mathbb{E}[(X_{(t+1)\wedge\tau} - X_{t\wedge\tau})\mathbb{1}_{\tau>t} \,|\, \mathcal{F}_t] = \mathbb{E}[(X_{t+1} - X_t)\mathbb{1}_{\tau>t} \,|\, \mathcal{F}_t] \\ &= \mathbb{1}_{\tau>t}\mathbb{E}[(X_{t+1} - X_t) \,|\, \mathcal{F}_t] \geq 0,\end{aligned}$$

where the third equality takes advantage of $\{\tau > t\} \in \mathcal{F}_t$, and the inequality follows since (X_t) is a submartingale with respect to $(\mathcal{F}_t)$. □

(Sub-, super-) Martingales and the optional stopping theorem provide very powerful tools, and part of the difficulties in / art of applying them is to discover the actual martingale to which to apply the above machinery. This is still fairly easy in the following example, which is a gambler's ruin problem.

Example 2.1.15. *Consider simple random walk (S_n) on the integers endowed with the canonical filtration $(\mathcal{F}_n)$ so that it becomes a martingale.*

(a) *For $a < 0 < b$ with $a, b \in \mathbb{Z}$ we are interested in the probability that (S_n) hits a before it hits b. For this purpose, define the random time $\tau := \tau_{a,b} := \inf\big\{n \in \mathbb{N}_0 \,:\, S_n \in \{a, b\}\big\}$. and check that it is a stopping time with respect to $(\mathcal{F}_n)$.*

As a consequence, the optional stopping theorem supplies us with the fact that the process $(S_{n\wedge\tau})$, $n \in \mathbb{N}_0$, is a martingale. Thus, in particular we have

$$0 = \mathbb{E}[S_0] = \mathbb{E}[S_{n\wedge\tau}].$$

Applying the dominated convergence theorem to the right-hand side we get

$$0 = \mathbb{E}[S_\tau],$$

but we also have that

$$\mathbb{E}[S_\tau] = a\mathbb{P}(S_\tau = a) + b\mathbb{P}(S_\tau = b) = a\mathbb{P}(S_\tau = a) + b(1 - \mathbb{P}(S_\tau = a)),$$

which all in all supplies us with

$$\mathbb{P}(S_\tau = a) = \frac{b}{b-a}.$$

(b) *Although we know that (S_n) is a martingale and hence a fair game, we could hope to outfox fate / the casino by applying the following strategy: We start at time 0 with 0 capital (corresponding to $S_0 = 0$) and stop playing at all once we have reached capital 1. I.e., setting*

$$\tau_1 := \inf\{n \in \mathbb{N}_0 \,:\, S_n = 1\},$$

we consider the sequence of random variables $(S_{n\wedge\tau_1})$. Now it is not too hard to show that SRW is recurrent, i.e.,

$$\mathbb{P}\big(S_n = 0 \text{ for infinitely many } n \in \mathbb{N}\big) = 1,$$

and from this one can also infer that $\mathbb{P}(\tau_1 < \infty) = 1$. This means, that we will almost surely reach the point at which we have capital 1 in finite time. However, the optional stopping theorem tells us that $(S_{n\wedge\tau_1})$, $n \in \mathbb{N}$, still is a martingale with respect to $(\mathcal{F}_n)$ – so in particular we have $\mathbb{E}[S_{n\wedge\tau_1}] = 0$ for each $n \in \mathbb{N}$ (this is yet another example in the flavour of Example 2.1.13). Think about what this means in terms of the gambling strategy!

In the following example it is already slightly harder to identify the 'right' martingale, and for the sake of completeness we give another version of the optional stopping theorem that will be proved as a homework exercise and employed in the following example.

Theorem 2.1.16. *Let (M_n) be a martingale with bounded increments, i.e., there exists $K \in (0, \infty)$ such that*

$$|M_{n+1} - M_n| \leq K \quad \forall n \in \mathbb{N},$$

and let τ be a stopping time which is integrable. Then

$$\mathbb{E}[M_\tau] = \mathbb{E}[M_0].$$

Proof. Problem sheet 3 (Exercise 3.1) $\square$

Example 2.1.17. (Patterns in coin tossing) *We want to compute the expected time it takes to first see a certain pattern in a sequence of fair coin tosses. For the sake of conciseness let us consider the pattern 'heads-heads-tail', abbreviated HHT. The first time that this pattern occurs in a sequence of fair coin tosses modeled by the sequence (X_n) is a stopping time with respect to the canonical filtration $(\mathcal{F}_n)$ defined via $\mathcal{F}_n := \sigma(X_1, \ldots, X_n)$ (exercise). More formally, as in the Example 4.2.7 of [Dre19] we consider an i.i.d. sequence (X_n) such that X_n is Rademacher-distributed; we identify $\{X_n = 1\}$ with the n-th coin toss resulting in heads, and $\{X_n = -1\}$ with the n-th coin toss resulting in tails. Hence,*

$$\tau_{HHT} := \inf\{n \in \mathbb{N} : X_n = X_{n+1} = 1, X_{n+2} = -1\} + 2.$$

Before each time $n \in \mathbb{N}$ we consider a new player entering the scene, borrowing one unit from the bank, and betting it unit on $\{X_n = 1\}$. If she wins, she receives 2 units from the casino, if she loses, she loses the one unit she borrowed. In the latter case she stops playing, whereas in the former she continues and bets all her 2 units on $\{X_{n+1} = 1\}$. Again, if she wins this bet, her stakes are doubled and she receives 4 units from the casino, whereas if she loses, she stops playing being broke and still having the loan with the bank. In the former case, she bets 4 units on $\{X_{n+2} = -1\}$; if she wins, she gets 8 units, whereas if she loses, she's broke (still with a loan). In either case she stops playing, but notice that she wins the last bet if and only if $\tau_{HHT} = n + 2$.
Now writing A^i_n to denote the capital of the i-th player at time n, we can write, using the notation $S_n := \sum_{i=1}^n X_i$, that

$$A^i_n = (Y^i \bullet S)_n,$$

where

$$Y^i_n = \begin{cases} 1, & \text{if } n = i, \\ 2, & \text{if } n = i+1 \text{ and } X_n = 1, \\ 4, & \text{if } n = i+2 \text{ and } X_{n+1} = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, $A^i_n - 1$, $n \in \mathbb{N}$ (recall the loan of 1 unit from the bank, which implies the -1) is a martingale due to Theorem 2.0.15, and similarly we have for the accumulated wealth process $V_n := \sum_{i=1}^n A^i_n - n$ that it is a martingale with $V_0 = 0$. Furthermore, we also have that $\mathbb{E}[\tau_{HHT}] < \infty$, since τ_{HHT} can be dominated by three times a geometric random variable. Therefore, and since there exists a constant $C \in (0, \infty)$ such that

$$|V_{n+1} - V_n| \le C \quad \forall n \in \mathbb{N},$$

we can apply the version of the optional stopping theorem given above as Theorem 2.1.16 to obtain that $\mathbb{E}[V_{\tau_{HHT}}] = \mathbb{E}[V_0] = 0$. Furthermore, conditional on $\{\tau_{HHT} = n\}$, $n \ge 2$, we have that the

$$A^i_{\tau_{HHT}} = \begin{cases} 0, & \text{if } i < n, \\ 8, & \text{if } i = n, \end{cases}$$

Thus, $V_{\tau_{HHT}} = 8 - \tau_{HHT}$, which in combination with the above yields

$$8 - \mathbb{E}[\tau_{HHT}] = 0,$$

and hence

$$\mathbb{E}[\tau_{HHT}] = 8.$$

Note that e.g. $\mathbb{E}[\tau_{HHH}]$ is strictly bigger than $\mathbb{E}[\tau_{HHT}]$ (or the expected time of any other triplet of results except for TTT for that matter) – think about why this is the case!

Example 2.1.18 (The "abracadabra" problem)**.** *Let $(U_i)_{i\in\mathbb{N}}$ denote random letters drawn independently and uniformly from the english alphabet. We are interested in the expected amount of draws needed to observe the the pattern "abracadabra", i.e. the expectation of*

$$\tau = \min\{n \in \mathbb{N}, n \geq 11 : U_i = x_i \forall i \in \{1, \dots, 11\}\},$$

where $x_1 =$ 'a', $x_2 =$ 'b', $x_3 =$ 'r' and so on. To show this, just like in Example 2.1.17, we must find the right martingale, show that it has bounded increments and show that the stopping time τ is integrable.
Finding the right process and proving that it is a martingale is using the same strategy as in Example 2.1.17 straightforward, so we only describe the strategy of the i-th player and leave proving that it is a martingale with bounded increments as an exercise. Using the notation of Example 2.1.17, we have

$$Y_n^i = \begin{cases} 1 & \text{if } n < i \\ Y_{n-1}^i \cdot 26\mathbb{1}_{\{U_n = x_{(n-i)+1}\}} & \text{if } i \leq n \leq i+10 \\ Y_{i+10}^i & \text{if } n > i+10 \end{cases}.$$

With this strategy, we get that the capital of player i $A_n^i = (Y^i \bullet S)_n$ is a martingale with expectation 1. Assuming that τ is integrable, we get like in Example 2.1.17 that

$$0 = \mathbb{E}[V_0] = \mathbb{E}[V_\tau] = 26^{11} + 26^4 + 26 - \mathbb{E}[\tau],$$

which gives us the desired answer. Before we proceed to show that τ is integrable, let us comment on why $V_\tau = 26^{11} + 26^4 + 26 - \tau$. Player $\tau - 10$ has entered the game precisely at the beginning of the sequence of letters that spell "abracadabra" and has successfully won at every step, thus his capital is $26^{11} - 1$. All the players that came before him had to have lost at some step of their play (since "abracadabra" has not appeared before) and thus their capital is -1. The same is true for most players that have entered the game after player $\tau - 10$. For example, player $\tau - 7$ has entered the game when the second "a" of the word "abracadabra" is drawn. Therefore, he initially doubled his capital, but then lost it all in the following step (betting on "b" when the letter "c" was drawn). This is not the case only for two other players. Player $\tau - 3$ has entered the game precisely when the second-last "a" of "abracadabra" has been drawn, and has in his first 4 steps correctly bet on the letters "a","b","r" and "a", resulting in his capital being $26^4 - 1$. The last player to enter has similarly made a correct bet on "a". Note that these two players have not made a wrong bet as of time τ, so their capital hasn't dropped to -1 yet. Combining all these players together, we observe the value -1 appearing τ times, which together with the terms 26^{11}, 26^4 and 26 gives the full value of V_τ.
It remains to argue that $E[\tau] < \infty$. To show that, we will use Lemma A.2.1. By assumption, the variables U_i are independent and therefore the probability of observing the sequence "abracadabra" is p^{11}, where $p = 1/26$. Since the event A_n that we observe the sequence "abracadabra" during the steps $n+1$ to $n+11$ is contained in the event $\tau \leq n+11$, we have that

$$\mathbb{P}(\tau < n + 11 \,|\, \tau > n) \geq \mathbb{P}(A_n \,|\, \tau > n)$$

for every n. However, the events A_n and $\{\tau > n\}$ are independent so that

$$\mathbb{P}(\tau < n + 11 \,|\, \tau > n) \geq \mathbb{P}(A_n \,|\, \tau > n) = P(A_n) = p^{11}.$$

By Lemma A.2.1 with $N = 11$ and $\epsilon = p^{11}$ we get that $\mathbb{E}[\tau] < \infty$.

Example 2.1.19. St. Petersburg paradox *(due to Daniel Bernoulli, see `https://en.wikipedia.org/wiki/Daniel_Bernoulli`, who also was a professor in St. Petersburg)*
Recall the setting of Example 2.0.16.

2.1.1 Inequalities for (sub-)martingales

In what comes below, the following definition comes in handy.

Definition 2.1.20. *For $(X_t)_{t\in T}$ (with $T \subset \mathbb{N}_0$) a real-valued stochastic process we define its* running maximum

$$X_t^* := \max_{s\in T,\, s\le t} X_s. \tag{2.1.4}$$

In the following we will get to know some inequalities which are usually associated to J. Doob. There is, however, some ambiguity about the exact naming of them, and the names I quote below are the result of a short and non-representative survey of the nomenclature in some standard sources.
The following can be interpreted as an extension of Markov's inequality.

Lemma 2.1.21 (Doob's (submartingale) inequality)**.** *Let $(X_t)_{t\in T}$ be a submartingale. Then for any $\lambda \in (0,\infty)$,*

$$\lambda \cdot \mathbb{P}(X_t^* \ge \lambda) \le \mathbb{E}[X_t \cdot \mathbb{1}_{X_t^* \ge \lambda}]$$

Proof. We define the stopping time

$$\tau_\lambda := \min\big\{ s \in T \,:\, X_s \ge \lambda \big\}.$$

Since $\tau_\lambda \wedge t$ is bounded, the optional sampling theorem (Theorem 2.1.12) supplies us with

$$\mathbb{E}[X_t] \ge \mathbb{E}[X_{\tau_\lambda \wedge t}] = \mathbb{E}[X_t \cdot \mathbb{1}_{X_t^* < \lambda}] + \mathbb{E}[X_{\tau_\lambda} \cdot \mathbb{1}_{X_t^* \ge \lambda}] \ge \mathbb{E}[X_t \cdot \mathbb{1}_{X_t^* < \lambda}] + \lambda \mathbb{P}(X_t^* \ge \lambda),$$

so

$$\lambda \mathbb{P}(X_t^* \ge \lambda) \le \mathbb{E}[X_t \cdot \mathbb{1}_{X_t^* \ge \lambda}].$$

□

Remark 2.1.22. *It should be noted that although Markov's inequality is sharp e.g. in the case of the constant random variable $X = \lambda$, the above allows us to replace the left-hand side of Markov's inequality by a sum of probabilities of corresponding events.*

We will apply Lemma 2.1.21 in the proof of the result below.
In the following, similarly to (2.1.4), we define

$$|X|_t^* := \max_{s \le t} |X_s|.$$

Theorem 2.1.23 (Doob's $\mathcal{L}^p$-inequality or also Doob's maximum inequality)**.**
Let $(X_t)_{t\in T}$ be a martingale or a positive submartingale with countable index set T. Then for any $p \in (1,\infty)$ and $t \in T$,

$$\mathbb{E}\big[|X_t|^p\big] \le \mathbb{E}\big[(|X|_t^*)^p\big] \le \Big(\frac{p}{p-1}\Big)^p \mathbb{E}\big[|X_t|^p\big]. \tag{2.1.5}$$

The statement of Theorem 2.1.23 fails to hold in the case $p = 1$ in the sense that there exists no finite constant $C \in (0,\infty)$ such that for any (X_t) as in the assumptions of Theorem 2.1.23 one has

$$\mathbb{E}\big[(|X|_t^*)^p\big] \le C\mathbb{E}\big[|X_t|^p\big]. \tag{2.1.6}$$

Indeed, consider simple random walk starting in 1, i.e., $S_n = 1 + \sum_{j=1}^n X_j$ and the (X_j) i.i.d. Rademacher distributed. Furthermore, let $\tau_0 := \inf\{n \in \mathbb{N} \,:\, S_n = 0\}$. Then Theorem 2.1.14 implies that $(S_{n\wedge\tau_0})$ is a martingale and hence $\mathbb{E}[|S_{n\wedge\tau_0}|] = \mathbb{E}[S_0] = 1$. On the other hand, one can show that $\mathbb{E}[|S_{n\wedge\tau_0}|^*] \to \infty$ as $n \to \infty$, which implies that (2.1.6) cannot hold.
One does, however, have a version of the result with $1 + \mathbb{E}[|X_t| \ln(|X_t| \vee 1)]$ on the right-hand side.

Proof of Theorem 2.1.23.
The first inequality is clear.
We start with noticing that $(|X|_t)$ is a submartingale (either by assumption or taking advantage of Theorem 2.0.11) and thus Lemma 2.1.21 yields

$$\lambda\mathbb{P}(|X|_t^* \geq \lambda) \leq \mathbb{E}[|X_t| \cdot \mathbb{1}_{|X|_t^* \geq \lambda}]. \tag{2.1.7}$$

We will truncate $|X|_t^*$ and consider $|X|_t^* \wedge M$ for $M > 0$ instead.
Then, using that for $r > 0$ we have $r^p = \int_0^r p\lambda^{p-1}\,\mathrm{d}\lambda$, we get

$$\mathbb{E}\big[(|X|_t^* \wedge M)^p\big] = \mathbb{E}\Big[\int_0^{|X|_t^* \wedge M} p\lambda^{p-1}\,\mathrm{d}\lambda\Big] = \mathbb{E}\Big[\int_0^M p\lambda^{p-1} \cdot \mathbb{1}_{|X|_t^* \geq \lambda}\,\mathrm{d}\lambda\Big],$$

which, using Tonelli's theorem, equals

$$\begin{aligned}\int_0^M p\lambda^{p-1} \cdot \mathbb{P}(|X|_t^* \geq \lambda)\,\mathrm{d}\lambda \leq \int_0^M p\lambda^{p-2} \cdot \mathbb{E}[|X|_t \cdot \mathbb{1}_{|X|_t^* \geq \lambda}]\,\mathrm{d}\lambda = p\mathbb{E}\Big[|X_t| \int_0^{|X|_t^* \wedge M} \lambda^{p-2}\,\mathrm{d}\lambda\Big] \\ = \frac{p}{p-1}\mathbb{E}\Big[|X_t|(|X|_t^* \wedge M)^{p-1}\Big],\end{aligned} \tag{2.1.8}$$

and where we used (2.1.7) in the inequality.
Now since $\frac{1}{p} + \frac{1}{p/(p-1)} = 1$, we can apply Hölder's inequality with exponents p and $p/(p-1)$ to obtain

$$\mathbb{E}\Big[|X_t|(|X|_t^* \wedge M)^{p-1}\Big] \leq \mathbb{E}[|X_t|^p]^{\frac{1}{p}}\mathbb{E}[(|X|_t^* \wedge M)^p]^{\frac{p-1}{p}}.$$

Hence, we can continue (2.1.8) to get

$$\mathbb{E}\big[(|X|_t^* \wedge M)^p\big] \leq \frac{p}{p-1}\mathbb{E}[|X_t|^p]^{\frac{1}{p}}\mathbb{E}[(|X|_t^* \wedge M)^p]^{\frac{p-1}{p}}.$$

Raising both sides of the inequality to the p-th power and dividing by $\mathbb{E}[(|X|_t^* \wedge M)^p]^{p-1}$ (which is finite due to the truncation with M), we obtain

$$\mathbb{E}\big[(|X|_t^* \wedge M)^p\big] \leq \Big(\frac{p}{p-1}\Big)^p\mathbb{E}[|X_t|^p].$$

Taking $M \to \infty$ yields (2.1.5) and hence finishes the proof. $\square$

2.2 A first martingale convergence theorem

We will now see some basic results on martingale convergence.
Our start will be a fundamental inequality involving the so-called upcrossings of an interval $[a, b]$ by a submartingale (X_n). For this purpose, we keep in mind the following trading strategy: We 'buy' a unit of the submartingale once it falls below a, and we sell once it exceeds b.
More formally: Define the stopping times $\sigma_0 := 0$,

$$\tau_k := \inf\{n \geq \sigma_{k-1} \,:\, X_n \leq a\}, \quad k \geq 1,$$

$$\sigma_k := \inf\{n \geq \tau_k \,:\, X_n \geq b\}, \quad k \geq 1.$$

We can then define the *number of upcrossings of* (X_n) *over* $[a, b]$ *up to time* $m \in \mathbb{N}$ via

$$U_m[a, b] := \sup\{k \,:\, \sigma_k \leq m\}.$$

Lemma 2.2.1 (Upcrossing inequality). *Let (X_n) be a submartingale. Then*

$$\mathbb{E}\big[U_n[a,b]\big] \le \frac{\mathbb{E}[(X_n-a)^+]-\mathbb{E}[(X_0-a)^+]}{b-a}.$$

Proof. Recall from Theorem 2.0.15 that the discrete stochastic integral $(H \bullet X)_n$ is a submartingale if (X_n) is a submartingale and (H_n) is an adapted, bounded, non-negative process. Putting the above trading strategy in mathematical terms we can define $H_n := 1$ for $n \in \{\tau_k, \dots, \sigma_k - 1\}$ and all $k \in \mathbb{N}$, and $H_n := 0$ otherwise. Then (H_n) is certainly adapted, bounded, non-negative (check!). If we define $Y_n := a + (X_n - a)^+$ this yields a submartingale (e.g. due to part (b) of Theorem 2.0.11), and as a consequence of Theorem 2.0.15, $(H \bullet Y)_n$ defines a submartingale. Furthermore, we have for $k \in \mathbb{N}$ on the event $\{\sigma_k < \infty\}$ that

$$(H \bullet Y)_{\sigma_k} = \sum_{j=1}^{k} (\underbrace{Y_{\sigma_j}}_{\ge b} - \underbrace{Y_{\tau_j}}_{\le a}) \ge (b-a)k.$$

By our definition of (Y_n), for $n \in \{\sigma_k, \dots, \sigma_{k+1}-1\}$, we have $(H \bullet Y)_n - (H \bullet Y)_{\sigma_k} \ge 0$, so

$$(H \bullet Y)_n \ge (b-a)U_n[a,b].$$

As a consequence, taking expectations we arrive at

$$\mathbb{E}[(H \bullet Y)_n] \ge (b-a)\mathbb{E}[U_n[a,b]].$$

Setting $K_n := 1 - H_n \ge 0$ defines a bounded and adapted process, and hence we deduce in an analogous way that $(K \bullet Y)_n$ is a submartingale, which implies that

$$\mathbb{E}[(K \bullet Y)_n] \ge \mathbb{E}[(K \bullet Y)_0] = 0. \tag{2.2.1}$$

We have

$$(H \bullet Y)_n + (K \bullet Y)_n = Y_n - Y_0,$$

so

$$\mathbb{E}[(H \bullet Y)_n] + \mathbb{E}[(K \bullet Y)_n] = \mathbb{E}[Y_n] - \mathbb{E}[Y_0],$$

which in combination with (2.2.1) yields

$$\mathbb{E}[(H \bullet Y)_n] \le \mathbb{E}[Y_n] - \mathbb{E}[Y_0],$$

so

$$\mathbb{E}[U_n[a,b]] \le \frac{\mathbb{E}[(X_n-a)^+]-\mathbb{E}[(X_0-a)^+]}{b-a},$$

which proves the result.

□

Theorem 2.2.2 (Martingale convergence theorem). *Let (X_n) be a submartingale with*

$$\sup_{n \in T} \mathbb{E}[X_n^+] < \infty. \tag{2.2.2}$$

Then there exists a random variable $X_\infty \in \mathcal{L}^1(\Omega, \mathcal{F}_\infty, \mathbb{P})$ such that

$$X_n \to X_\infty \quad \mathbb{P}-a.s.$$

as $n \to \infty$.

Proof of Theorem 2.2.2. For any $a, b \in \mathbb{R}$ with $a < b$, the monotone limit $U[a,b] := \lim_{n\to\infty} U_n[a,b]$ exists as a random variable mapping from $(\Omega, \mathcal{F}, \mathbb{P})$ to $([0,\infty], \mathcal{B}([0,\infty]))$. Using the monotone convergence theorem and Lemma 2.2.1, we thus infer

$$\mathbb{E}[U[a,b]] = \lim_{n\to\infty} \mathbb{E}[U_n[a,b]] \le \limsup_{n\to\infty} \frac{\mathbb{E}[(X_n - a)^+] - \mathbb{E}[(X_0 - a)^+]}{b-a} \le \limsup_{n\to\infty} \frac{\mathbb{E}[X_n^+] + |a|}{b-a} < \infty,$$

where the latter inequality follows from (2.2.2). In particular,

$$\mathbb{P}(U[a,b] < \infty) = 1. \tag{2.2.3}$$

But we can write

$$\begin{aligned}\{X_n \text{ does not converge in } [-\infty,\infty]\} &\subset \{\liminf_{n\to\infty} X_n < \limsup_{n\to\infty} X_n\} \\ &\subset \bigcup_{a,b\in\mathbb{Q}} \{\liminf_{n\to\infty} X_n < a < b < \limsup_{n\to\infty} X_n\} \subset \bigcup_{a,b\in\mathbb{Q}} \{U[a,b] = \infty\}.\end{aligned}$$

Thus, in combination with (2.2.3) we infer that

$$\mathbb{P}(\{X_n \text{ converges in } [-\infty,\infty]\}) \ge 1 - \sum_{a,b\in\mathbb{Q}} \mathbb{P}(U[a,b] = \infty) = 1,$$

and we denote the $\mathbb{P}$-a.s. limit of (X_n) by X_∞ which, as the limit of $\mathcal{F}_\infty - \mathcal{B}(\overline{\mathbb{R}})$-measurable random variables is $\mathcal{F}_\infty - \mathcal{B}(\overline{\mathbb{R}})$-measurable itself. To show we do in fact even have $X_\infty \in \mathcal{L}^1$, observe that Fatou's lemma supplies us with

$$\mathbb{E}[X_\infty^+] = \mathbb{E}[(\lim_{n\to\infty} X_n)^+] = \mathbb{E}[\liminf_{n\to\infty} (X_n^+)] \le \liminf_{n\to\infty} \mathbb{E}[X_n^+],$$

and the right-hand side is finite due to (2.2.2). On the other hand, since (X_n) is a submartingale, we obtain

$$\mathbb{E}[X_n^-] = \mathbb{E}[X_n^+] - \mathbb{E}[X_n] \le \mathbb{E}[X_n^+] - \mathbb{E}[X_0].$$

Applying Fatou's lemma again supplies us with

$$\mathbb{E}[X_\infty^-] = \mathbb{E}[\liminf_{n\to\infty} X_n^-] \le \liminf_{n\to\infty} \mathbb{E}[X_n^-] \le \liminf_{n\to\infty} \mathbb{E}[X_n^+] - \mathbb{E}[X_0],$$

and again the right-hand side is finite due to (2.2.2). This shows $X_\infty \in \mathcal{L}^1$ and finishes the proof. □

Exercise 2.2.3. *Prove that in the context of Theorem 2.2.2, the condition in (2.2.2) is equivalent to the sequence (X_n) being L^1-bounded, i.e.,*

$$\sup_{n\in\mathbb{N}} \mathbb{E}[|X_n|] < \infty.$$

The above will enable us to give a generalisation of the Borel-Cantelli lemma: The independence assumption on the events (A_n) can be dropped in this version of the lemma. Before, however, we have to give a small auxiliary lemma which is of interest on its own.

Lemma 2.2.4. *Let (M_n) be a martingale and assume there exists a constant $K \in (0,\infty)$ such that the increments of (M_n) are bounded by K, i.e.,*

$$|M_n - M_{n+1}| \le K \quad \forall n \in \mathbb{N}.$$

Then for

$$C_M := \big\{(M_n) \text{ converges in } \mathbb{R} \text{ as } n \to \infty\big\}$$

and

$$F_M := \big\{\limsup_{n\to\infty} M_n = -\liminf_{n\to\infty} M_n = \infty\big\},$$

we have $\mathbb{P}(C_M \cup F_M) = 1$.

Proof. First observe that $\widetilde{M}_n := M_n - M_0$ defines a martingale with $\widetilde{M}_0 = 0$. For $L < 0$ we define the stopping time

$$\tau_L := \inf\{n \in \mathbb{N} : \widetilde{M}_n \leq L\}.$$

The optional stopping theorem (Theorem 2.1.14) implies that

$$(\widetilde{M}_{n\wedge\tau_L})$$

is a martingale, and furthermore we have $\widetilde{M}_{n\wedge\tau_L} \geq L-K$ for all $n \in \mathbb{N}$. Hence, $(\widetilde{M}_{n\wedge\tau_L} - (L-K))$ defines a non-negative martingale, which due to Theorem 2.2.2 converges $\mathbb{P}$-a.s. to a random variable in $\mathcal{L}^1$. Hence, on $\{\tau_L = \infty\}$, the martingale $(\widetilde{M}_n - (L-K))$ and hence also the martingale (M_n) converge $\mathbb{P}$-a.s. to a finite limit. Taking $L \to -\infty$ we deduce that (M_n) converges in $\mathbb{R}$ $\mathbb{P}$-a.s. on

$$\bigcup_{L\in-\mathbb{N}} \{\tau_L = \infty\} = \big\{\liminf_{n\to\infty} M_n > -\infty\big\}.$$

Repeating the above procedure for the martingale $(-M_n)$, we obtain that $(-M_n)$ and hence also (M_n) converges $\mathbb{P}$-a.s. on

$$\big\{\limsup_{n\to\infty} M_n < \infty\big\}.$$

But (M_n) converges if and only $(-M_n)$ does, so

$$\big\{\liminf_{n\to\infty} M_n > -\infty\big\} = \big\{\limsup_{n\to\infty} M_n < \infty\big\},$$

and hence

$$\big\{\liminf_{n\to\infty} M_n = -\infty\big\} = \big\{\limsup_{n\to\infty} M_n = \infty\big\},$$

which proves the lemma. □

Corollary 2.2.5 (Generalised Borel-Cantelli lemma)**.** *Let $(\mathcal{F}_n)$ be a filtration with $\mathcal{F}_0 = \{\emptyset, \Omega\}$, and let (A_n) be a sequence of events with $A_n \in \mathcal{F}_n$ for each $n \in \mathbb{N}$.*
Define

$$Y_m := \sum_{n=1}^{m} \mathbb{P}(A_n \,|\, \mathcal{F}_{n-1})$$

and

$$Z_m := \sum_{n=1}^{m} \mathbb{1}_{A_n}$$

as well as their limits

$$Y_\infty := \lim_{m\to\infty} Y_m,$$

and

$$Z_\infty := \lim_{m\to\infty} Z_m.$$

Then $\mathbb{P}$-a.s. the following implications hold true:

(a)

$$Y_\infty < \infty \Longrightarrow Z_\infty < \infty;$$

(b)

$$Y_\infty = \infty \Longrightarrow \frac{Z_m}{Y_m} \to 1 \qquad \text{as } m \to \infty.$$

Proof. We will only prove the weaker version

$$\mathbb{P}\big(\{Z_\infty = \infty\}\Delta\{Y_\infty = \infty\}\big) = 0. \tag{2.2.4}$$

A proof of the fully-fledged result given in the theorem requires a finer understanding of convergence of square integrable martingales and can be found as the proof of Theorem 12.15 in [Wil91].
To prove (2.2.4), we observe that for the martingale (check!) $M_m := Z_m - Y_m$ we have $|M_{m+1} - M_m| \le 1$ for all $m \in \mathbb{N}$. Hence, we can apply Lemma 2.2.4 which (in the notation used in there) yields that on F_M one has that

$$Z_\infty = \infty \quad \text{and} \quad Y_\infty = \infty,$$

whereas on C_M one has that

$$Z_\infty = \infty \quad \text{if and only if} \quad Y_\infty = \infty.$$

Since Lemma 2.2.4 implies $\mathbb{P}(C_M \cup F_M) = 1$, this proves the result.

□

Corollary 2.2.6. *If (X_n) is a non-negative supermartingale, then $X_\infty := \lim_{n\to\infty} X_n$ exists $\mathbb{P}$-a.s. with $\mathbb{E}[X_\infty] \le \mathbb{E}[X_0]$.*

Proof. We know that $(-X_n)$ is a submartingale with $\sup_{n\in T} \mathbb{E}[(-X_n)^+] \le 0 < \infty$, so the previous theorem implies that $-X_\infty := \lim_{n\to\infty} -X_n$ exists in $\mathcal{L}^1$ $\mathbb{P}$-a.s. In addition, Fatou's lemma yields

$$\mathbb{E}[X_\infty] = \mathbb{E}[\liminf_{n\to\infty} X_n] \le \liminf_{n\to\infty} \mathbb{E}[X_n] \le \mathbb{E}[X_0],$$

where the last inequality follows from (X_n) being a supermartingale. □

Example 2.2.7. *(a) Consider simple random walk starting in 1, i.e., $S_n := 1 + \sum_{k=1}^n X_k$, where the X_k are i.i.d. Rademacher distributed, and consider the first hitting time of 0 defined via*

$$\tau_0 := \inf\{n \in \mathbb{N} \,:\, S_n = 0\}.$$

Keeping in mind that $(S_{n\wedge\tau_0})$ is a non-negative martingale due to Theorem 2.1.14, we can use Corollary 2.2.6 in order to infer that $S_{n\wedge\tau_0}$ converges $\mathbb{P}$-a.s. to 0 as $n \to \infty$. Obviously, this implies $\mathbb{P}(\tau_0 < \infty) = 1$.

Also, note that $\mathbb{E}[S_{n\wedge\tau_0}] = \mathbb{E}[S_0] = 1$, since $(S_{n\wedge\tau_0})$ is a martingale. In particular, this shows that $(S_{n\wedge\tau_0})$ cannot converge in L^1. This is different from so-called uniformly integrable *martingales that we will get to know below, see Section 2.3.*

(b) Still considering S_n to be simple random walk (can start it in 0 though, so $S_n := \sum_{k=1}^n X_k$), since the martingale (S_n) has increments bounded by 1, we infer using Lemma 2.2.4 (again taking advantage of the fact that S_n cannot converge since increments are of size 1 at each time, and hence (S_n) cannot converge in $\mathbb{R}$) that we must have

$$\mathbb{P}(\limsup_{n\to\infty} S_n = -\liminf_{n\to\infty} S_n = \infty) = 1.$$

This implies the recurrence *of simple random walk, namely that $\mathbb{P}(S_n = 0$ for infinitely many $n \in \mathbb{N}) = 1$.*

Corollary 2.2.8. *Let $p \in [1, \infty)$, $X \in \mathcal{L}^p$, and let $(\mathcal{F}_n)$ be a filtration. Then*

$$\mathbb{E}[X \,|\, \mathcal{F}_n] \to \mathbb{E}[X \,|\, \mathcal{F}_\infty] \quad \mathbb{P} - a.s.$$

Proof. According to Example 2.0.8 (b), setting $X_n := (\mathbb{E}[X \,|\, \mathcal{F}_n])$ constitutes a martingale (X_n) with

$$\sup_{n\in T} \mathbb{E}[X_n^+] \le \sup_{n\in T} \mathbb{E}[|X_n|] = \sup_{n\in T} \mathbb{E}[\big|\mathbb{E}[X \,|\, \mathcal{F}_n]\big|] \le \sup_{n\in T} \mathbb{E}[\mathbb{E}[|X| \,|\, \mathcal{F}_n]] = \mathbb{E}[|X|] < \infty,$$

where the penultimate inequality follows from the conditional Jensen inequality, the last equality from the tower property for conditional expectations, and the last equality from the fact that $X \in \mathcal{L}^p \subset \mathcal{L}^1$. Hence, the previous theorem implies the $\mathbb{P}$-a.s. convergence of X_n to X_∞ which is $\mathcal{F}_\infty - \mathcal{B}$-measurable.
It remains to argue that

$$X_\infty = \mathbb{E}[X \,|\, \mathcal{F}_\infty], \tag{2.2.5}$$

and this will be done in Example 2.3.12 as a demonstration how powerful the machinery of uniform integrability is. □

The following is a useful corollary to Corollary 2.2.8. It states that if the σ-algebra we condition on is countably generated, then we can approximate the general conditional expectation introduced in Definition 1.0.3 by the basic one we had introduced before in (1.0.1). To make this more precise, recall (see Exercise 1.0.8) that for a finite partition $\mathcal{P}$ of Ω with $\mathcal{P} \subset \mathcal{F}$, and $X \in \mathcal{L}^p$ some $p \in [1, \infty)$ we have

$$\mathbb{E}[X \,|\, \sigma(\mathcal{P})] = \sum_{\substack{A\in\mathcal{P} \\ \mathbb{P}(A)>0}} \mathbb{E}[X \,|\, A] \mathbb{1}_A.$$

Corollary 2.2.9. *Let $(\mathcal{P}_n)$ be a sequence of finite partitions of Ω with $\sigma(\mathcal{P}_n) \subset \sigma(\mathcal{P}_{n+1}) \subset \mathcal{F}$ and set $X_n := \mathbb{E}[X \,|\, \sigma(\mathcal{P}_n)]$ as well as*

$$\mathcal{G} := \sigma\Big(\bigcup_{n\in\mathbb{N}} \mathcal{P}_n\Big).$$

Then

$$X_n \to \mathbb{E}[X \,|\, \mathcal{G}] \quad \mathbb{P} - a.s. \text{ and in } L^p.$$

Proof. Exercise. □

This last corollary can also be used to transfer other results which we know for expectations (and hence also for conditional expectations as defined in (1.0.1)) to general conditional expectations – an example for this is Jensen's inequality.
It turns out that under stronger assumptions on the martingales in question, our results of this section can be significantly improved (as one example, the convergence in Corollary 2.2.8 can be extended to convergence in L^p). For this purpose, we will introduce the concept of uniform integrability.

2.3 Uniform integrability

It will turn out that the concept of uniform integrability will prove very helpful in the context of (convergence of) martingales. In order to be able to take full advantage of it, we have to establish some basic properties first, which are interesting in their own right. We will afterwards apply this machinery to our setting of martingales.
To start with, observe that for a single random variable $X \in \mathcal{L}^1$ one has that the mapping

$$\mathcal{F} \ni F \mapsto \mathbb{E}[X \mathbb{1}_F] = \int_F X \, \mathrm{d}\mathbb{P}$$

defines a finite (signed) measure on $\mathcal{F}$. By definition, this measure is then absolutely continuous with respect to $\mathbb{P}$, and since $X \in \mathcal{L}^1$ we obtain that for each $\varepsilon > 0$ there exists $M \in (0, \infty)$ such that $\int_{|X| \geq M} |X| \, \mathrm{d}\mathbb{P} \leq \varepsilon$. The notion of uniform integrability introduced below is a generalisation of this property to families of random variables for which the respective M can be chosen uniformly in all members of the family.

Definition 2.3.1. *A family $\mathcal{S} \subset \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ is called* uniformly integrable ('gleichgradig integrierbar') *if for every $\varepsilon > 0$ there exists $M > 0$ such that*

$$\sup_{X \in \mathcal{S}} \int_{|X| \geq M} |X| \, \mathrm{d}\mathbb{P} \leq \varepsilon.$$

The notion of uniform integrability can be extended to infinite measure spaces $(\Omega, \mathcal{F}, \mu)$ with $\mu(\Omega) = \infty$. Since this makes some of the reasonings slightly more technical, and since we will primarily be interested in probability spaces, we stick to the setting of probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$ here.

Example 2.3.2. *(a) If (f_λ), $\lambda \in \Lambda$, with $|\Lambda| < \infty$ is a finite family of random variables with $f_\lambda \in \mathcal{L}^1$ for all $\lambda \in \Lambda$, then the family (f_λ) is uniformly integrable.*

(b) If $\mathcal{S} \subset \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ is a uniformly integrable family, then any subset $\mathcal{R} \subset \mathcal{S}$ is also a uniformly integrable family.

(c) If $\mathcal{R}, \mathcal{S} \subset \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ are uniformly integrable families, then $\mathcal{R} \cup \mathcal{S}$ is also a uniformly integrable family.

(d) Let $\mathcal{S} \subset \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ such that there exists $g \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ with

$$|f| \leq g \quad \forall f \in \mathcal{S}. \tag{2.3.1}$$

Then $\mathcal{S}$ is uniformly integrable. Indeed, since $g \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, for each $\varepsilon > 0$ we find $M \in (0, \infty)$ such that

$$\int_{|g| \geq M} g \, \mathrm{d}\mathbb{P} \leq \varepsilon.$$

Furthermore, from (2.3.1) we deduce

$$\sup_{f \in \mathcal{S}} \int_{|f| \geq M} |f| \, \mathrm{d}\mathbb{P} \leq \int_{|g| \geq M} g \, \mathrm{d}\mathbb{P},$$

which in combination with the previous gives us the desired uniform integrability.

(e) If for some $p \in (1, \infty)$ the family $(|f_\lambda|^p)$, $\lambda \in \Lambda$, is uniformly integrable, then, for any $q \in (1, p)$, so is the family $(|f_\lambda|^q)$, $\lambda \in \Lambda$.

Indeed, if $(|f_\lambda|^p)$, $\lambda \in \Lambda$, is uniformly integrable, then for any $\varepsilon > 0$ we have $M \in (1, \infty)$ such that the last inequality of the following holds true,

$$\sup_{\lambda \in \Lambda} \int_{|f_\lambda|^q \geq M^{\frac{q}{p}}} |f_\lambda|^q \, \mathrm{d}\mathbb{P} \leq \sup_{\lambda \in \Lambda} \int_{|f_\lambda|^p \geq M} (1 \vee |f_\lambda|^p) \, \mathrm{d}\mathbb{P} = \sup_{\lambda \in \Lambda} \int_{|f_\lambda|^p \geq M} |f_\lambda|^p \, \mathrm{d}\mathbb{P} < \infty,$$

where the equality follows from $M > 1$, and this proves the claim.

(f) Let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. Then the family

$$\big\{ \mathbb{E}[X \,|\, \mathcal{G}] \, : \, \mathcal{G} \text{ is a } \sigma\text{-algebra with } \mathcal{G} \subset \mathcal{F} \big\}$$

is uniformly integrable (easier to prove with the proof of Theorem 2.3.6 below).

Definition 2.3.3. *For $p \in [1, \infty]$, we call a family $\mathcal{S} \subset \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ (L^p)*-bounded*, if*

$$\sup_{X \in \mathcal{S}} \|X\|_p < \infty.$$

The following lemma can be seen as a generalisation to families of uniformly integrable random variables of the following absolute continuity property for integrable random variables: If $X \in \mathcal{L}^1$, then for each $\varepsilon > 0$ there exists $\delta > 0$ such that for all $F \in \mathcal{F}$ with $\mathbb{P}(F) \leq \delta$, one has $\mathbb{E}[|X| \mathbb{1}_F] \leq \varepsilon$.

Lemma 2.3.4. *A family (X_n), $n \in T$, of random variables is uniformly integrable if and only if*

(a) it is L^1-bounded, and

(b) for every $\varepsilon > 0$ there is $\delta > 0$ such that for all $A \in \mathcal{F}$ with $\mathbb{P}(A) < \delta$, one has

$$\mathbb{E}[|X_n| \cdot \mathbb{1}_A] < \varepsilon \quad \forall n \in T.$$

Proof. If the family (X_n) is uniformly integrable, then for any $\varepsilon > 0$ we have $M_\varepsilon \in (0, \infty)$ such that $\mathbb{E}[|X_n| \mathbb{1}_{|X_n| \geq M_\varepsilon}] \leq \varepsilon$, so for each $n \in T$,

$$\int |X_n| \, d\mathbb{P} \leq \int_{\{|X_n| \geq M_\varepsilon\}} |X_n| \, d\mathbb{P} + M_\varepsilon \leq M_\varepsilon + \varepsilon,$$

which means that the family is L^1-bounded. In addition, for $\varepsilon > 0$ arbitrary, we have with $\delta := \frac{\varepsilon}{M_\varepsilon}$ that for $F \in \mathcal{F}$ with $\mathbb{P}(F) \leq \delta$,

$$\mathbb{E}[|X_n| \cdot \mathbb{1}_F] = \mathbb{E}[|X_n| \cdot \mathbb{1}_F \mathbb{1}_{|X_n| \geq M_\varepsilon}] + \mathbb{E}[|X_n| \cdot \mathbb{1}_F \mathbb{1}_{|X_n| < M_\varepsilon}] \leq \varepsilon + M_\varepsilon \cdot \delta = 2\varepsilon,$$

which shows part (b).
Conversely, assume that (a) and (b) are fulfilled. Fix $\varepsilon > 0$ arbitrarily and choose $\delta > 0$ accordingly as in (b). Then (a) implies $C := \sup_{n \in T} \mathbb{E}[|X_n|] < \infty$, and we can choose $M_\delta \in (0, \infty)$ such that the last inequality of

$$\mathbb{P}(|X_n| \geq M_\delta) \leq \frac{\mathbb{E}[|X_n|]}{M_\delta} \leq \frac{C}{M_\delta} \overset{!}{\leq} \delta \quad \forall n \in T$$

holds true. Thus, by part (b) we have

$$\mathbb{E}[|X_n| \cdot \mathbb{1}_{|X_n| \geq M_\delta}] \leq \varepsilon \quad \forall n \in T$$

which shows the uniform integrability. $\square$

Exercise 2.3.5. *(a) The sequence $X_n := n \mathbb{1}_{[0, \frac{1}{n}]}$ of random variables defined on $([0,1], \mathcal{B}([0,1]), \lambda_{|[0,1]})$ is L^1-bounded, but (b) of Lemma 2.3.4 is not fulfilled.*

(b) Find an example of a family of random variables which fulfills part (b) of Lemma 2.3.4 but which does not fulfill part (a), and hence is not uniformly integrable.

A main tool for showing the uniform integrability of a family of random variables is given by the following characterisation.

Theorem 2.3.6. *Let $\mathcal{S} \subset \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. Then $\mathcal{S}$ is uniformly integrable if and only if there exists a function $\varphi : [0, \infty) \to [0, \infty)$ such that*

(a)

$$\frac{\varphi(x)}{x} \to \infty \qquad \text{as } x \to \infty, \tag{2.3.2}$$

(b) and

$$\sup_{X\in\mathcal{S}}\int_\Omega \varphi(|X|)\,\mathrm{d}\mathbb{P}<\infty. \tag{2.3.3}$$

Furthermore, φ can be chosen to be non-decreasing and convex.

Proof. If $\mathcal{S}$ is uniformly integrable, then for any $n\in\mathbb{N}$ there exists $M_n\in(0,\infty)$ such that

$$\sup_{X\in\mathcal{S}}\int_{\{|X|\geq M_n\}}|X|\,\mathrm{d}\mathbb{P}\leq 2^{-n},$$

and without loss of generality we can and do choose the M_n in such a way that $\lim_{n\to\infty}M_n=\infty$. A fortiori

$$\sup_{X\in\mathcal{S}}\int(|X|-M_n)^+\,\mathrm{d}\mathbb{P}\leq 2^{-n}. \tag{2.3.4}$$

Setting

$$\varphi:\,[0,\infty)\ni x\mapsto\sum_{n\in\mathbb{N}}(x-M_n)^+\in[0,\infty)$$

defines an non-decreasing function, which is also convex, since it is the sum of convex functions. In addition,

$$\frac{\varphi(x)}{x}=\sum_{n\in\mathbb{N}}(1-M_n/x)^+\to\infty\qquad\text{as } x\to\infty,$$

so (2.3.2) holds true. In addition, we deduce using (2.3.4) in combination with the monotone convergence theorem that

$$\sup_{X\in\mathcal{S}}\int\varphi(|X|)\,\mathrm{d}\mathbb{P}\leq\sum_{n\in\mathbb{N}}\sup_{X\in\mathcal{S}}\int(|X|-M_n)^+\,\mathrm{d}\mathbb{P}\leq 1,$$

which establishes (2.3.3).
To prove the converse implication, assume that (2.3.3) and (2.3.2) hold true. These properties imply that for M large enough, for any $\varepsilon>0$ there exists $M_\varepsilon\in(0,\infty)$ such that we have that

$$\varphi(x)\geq x\frac{\sup_{X\in\mathcal{S}}\int_\Omega\varphi(|X|)\,\mathrm{d}\mathbb{P}}{\varepsilon}\qquad\text{for all } x\geq M_\varepsilon.$$

Hence, uniformly in $X\in\mathcal{S}$,

$$\int_{|X|\geq M_\varepsilon}|X|\,\mathrm{d}\mathbb{P}\leq\frac{\varepsilon}{\sup_{X\in\mathcal{S}}\int_\Omega\varphi(|X|)\,\mathrm{d}\mathbb{P}}\int_{|X|\geq M_\varepsilon}\varphi(|X|)\,\mathrm{d}\mathbb{P}\leq\varepsilon,$$

which shows the uniform integrability of $\mathcal{S}$. □

Corollary 2.3.7. *Let $p>1$ and let $\mathcal{S}\subset\mathcal{L}^p$ be an L^p-bounded family. Then for any $q\in[1,p)$, the family*

$$\{|X|^q\,:\,X\in\mathcal{S}\}$$

is uniformly integrable.

Proof. We define $\varphi:[0,\infty)\to[0,\infty)$ via

$$\varphi(x):=x^{\frac{p}{q}}.$$

Then $\varphi(x)/x\to\infty$ as $x\to\infty$, and

$$\sup_{X\in\mathcal{S}}\int_\Omega\varphi(|X|^q)\,\mathrm{d}\mathbb{P}=\sup_{X\in\mathcal{S}}\int_\Omega|X|^p\,\mathrm{d}\mathbb{P}<\infty$$

due to the assumption that $\mathcal{S}$ is a L^p-bounded family. Via Theorem 2.3.6 this implies the result. □

Exercise 2.3.8. *Show that under the assumptions of Corollary 2.3.7, the family*

$$\{|X|^p \,:\, X \in \mathcal{S}\}$$

is not necessarily uniformly integrable.

One of the reasons that the notion of uniform integrability is so powerful is the following result.

Theorem 2.3.9. *Let (X_n), $n \in \mathbb{N}$, and X all be elements of $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. Assume furthermore that X_n converges to X in probability. Then the following are equivalent:*

(a) The family (X_n), $n \in \mathbb{N}$, is uniformly integrable.

(b) X_n converges to X in L^1.

(c)
$$\mathbb{E}[|X_n|] \to \mathbb{E}[|X|] < \infty.$$

Remark 2.3.10. *Note that the above describes a very special situation since we have seen in 'probability theory I' that convergence in probability is generally weaker than convergence in L^1.*

Proof of Theorem 2.3.9.
'(a) $\implies$ (b)':
For notational convenience write $X_0 := X$. Then the family (X_n), $n \in \mathbb{N}_0$, is uniformly integrable. Therefore, for $\varepsilon > 0$ arbitrary, Lemma 2.3.4 implies that we find $\delta > 0$ such that for all $F \in \mathcal{F}$ with $\mathbb{P}(F) \le \delta$ we have

$$\mathbb{E}[|X_n| \cdot \mathbb{1}_F] \le \varepsilon \quad \forall n \in \mathbb{N}_0. \tag{2.3.5}$$

Since X_n converges to X in probability as $n \to \infty$, we can find N_δ such that for all $n \ge N_\delta$ we have $\mathbb{P}(|X_n - X| \ge \varepsilon) \le \delta$. As a consequence, for such n we deduce

$$\begin{aligned}\mathbb{E}[|X_n - X|] &\le \mathbb{E}[|X_n - X| \cdot \mathbb{1}_{|X_n - X| \le \varepsilon}] + \mathbb{E}[|X_n - X| \cdot \mathbb{1}_{|X_n - X| \ge \varepsilon}] \\ &\le \varepsilon + \mathbb{E}[|X_n| \cdot \mathbb{1}_{|X_n - X| \ge \varepsilon}] + \mathbb{E}[|X| \cdot \mathbb{1}_{|X_n - X| \ge \varepsilon}],\end{aligned}$$

and the last two summands are each upper bounded by ε due to (2.3.5), which establishes (b).
'(b) $\implies$ (c)': Jensen's inequality implies the first and the reverse triangle the second inequality of

$$\Big|\mathbb{E}[|X_n|] - \mathbb{E}[|X|]\Big| \le \mathbb{E}\Big[\big||X_n| - |X|\big|\Big] \le \mathbb{E}[|X_n - X|],$$

and the right-hand side converges to 0 as $n \to \infty$ by assumption.
'(c) $\implies$ (a)':
We set $X_0 := X$ and start with noting that the family (X_n), $n \in \mathbb{N}_0$, is L^1-bounded due to (c). Now let $\varepsilon > 0$ be arbitrary. Then since $X \in \mathcal{L}^1$, there exists $\delta \in (0, \varepsilon)$ such that for all $\mathbb{P}(A) < \delta$, we have

$$\mathbb{E}[|X| \cdot \mathbb{1}_A] < \varepsilon. \tag{2.3.6}$$

Furthermore, due to $X_n \xrightarrow{\mathbb{P}} X$ and (c), there exists $N_\delta \in \mathbb{N}$ such that, setting $G_{\delta,n} := \{|X_n - X| < \delta\}$, for all $n \ge N_\delta$ we have

$$\mathbb{P}(G_{\delta,n}^c) < \delta \quad \text{and} \quad |\mathbb{E}[|X_n|] - \mathbb{E}[|X|]| < \varepsilon. \tag{2.3.7}$$

Hence, due to our choice of δ,

$$\Big|\mathbb{E}[|X_n| \cdot \mathbb{1}_{G_{\delta,n}^c}] - \mathbb{E}[|X| \cdot \mathbb{1}_{G_{\delta,n}^c}]\Big| \le \big|\mathbb{E}[|X_n|] - \mathbb{E}[|X|]\big| + \underbrace{\big|\mathbb{E}\big[(|X_n| - |X|) \cdot \mathbb{1}_{G_{\delta,n}}\big]\big|}_{\le \delta} < 2\varepsilon, \quad n \ge N_\delta, \tag{2.3.8}$$

so in combination with (2.3.7) and (2.3.6),

$$\mathbb{E}[|X_n| \cdot \mathbb{1}_{G^c_{\delta,n}}] \leq 3\varepsilon. \tag{2.3.9}$$

Now for arbitrary $A \in \mathcal{F}$ with $\mathbb{P}(A) \leq \delta$ we get that for $n \geq N_\delta$,

$$\mathbb{E}[|X_n| \cdot \mathbb{1}_A] \leq \mathbb{E}[|X_n| \cdot \mathbb{1}_{G^c_{\delta,n}}] + \mathbb{E}[|X_n| \cdot \mathbb{1}_{G_{\delta,n} \cap A}], \tag{2.3.10}$$

and by definition of $G_{\delta,n}$, the second summand on the right-hand side is upper bounded by

$$\mathbb{E}[|X_n| \cdot \mathbb{1}_{G_{\delta,n} \cap A}] \leq \mathbb{E}[|X| \cdot \mathbb{1}_A] + \delta \leq 2\varepsilon, \tag{2.3.11}$$

again due to (2.3.6). Summarising (2.3.11), (2.3.10), and (2.3.9), we infer that for $n \geq N_\delta$,

$$\mathbb{E}[|X_n| \cdot \mathbb{1}_A] \leq 5\varepsilon.$$

Since $\varepsilon > 0$ was chosen arbitrarily, and since as observed above, the family (X_n) is L^1-bounded, in combination with Lemma 2.3.4, this shows the desired uniform integrability.

□

Proposition 2.3.11. *Let (X_n) be a (sub-, super-) martingale with $X_n \to X_\infty$ in L^1 as $n \to \infty$. Then $X_n = \mathbb{E}[X_\infty \,|\, \mathcal{F}_n]$ (with '$=$' replaced by '$\leq$' and '$\geq$' in the case of a sub- and a supermartingale, respectively).*

Proof. We give the proof in the case of (X_n) being a martingale; the modifications in the case of a sub- or supermartingale are minor and left to the Reader.
Note that on the one hand, for $F \in \mathcal{F}_n$ we have

$$\int_F X_n \,\mathrm{d}\mathbb{P} = \int_F X_k \,\mathrm{d}\mathbb{P} \tag{2.3.12}$$

for all $k \geq n$ due to the martingale property $\mathbb{E}[X_k \,|\, \mathcal{F}_n] = X_n$. On the other hand, due to the L^1-convergence in the assumptions, we get

$$\lim_{k\to\infty} \int_F X_k \,\mathrm{d}\mathbb{P} = \int_F X_\infty \,\mathrm{d}\mathbb{P}.$$

Combining this with (2.3.12) and the fact that $F \in \mathcal{F}_n$ had been chosen arbitrarily, we infer that

$$X_n = \mathbb{E}[X_\infty \,|\, \mathcal{F}_n].$$

□

Example 2.3.12. *(a)*
Recall that we still had to prove (2.2.5); with Theorem 2.3.9 at our disposal, this boils down to a piece of cake: Indeed, Example 2.3.2 (f) tells us that the family of random variables $(\mathbb{E}[X \,|\, \mathcal{F}_n])$ is uniformly integrable. But in the partial proof of Corollary 2.2.8 we had that $(\mathbb{E}[X \,|\, \mathcal{F}_n])$ converges $\mathbb{P}$-a.s. to some limit X_∞, and hence Theorem 2.3.9 implies that it converges to X_∞ in L^1 as well. Now the characterisation of the limit follows from Proposition 2.3.11.

(b) Now that we have completed the proof of Corollary 2.2.8 we show how it provides us with an easy proof of Kolmogorov's $0-1$-law (cf. Theorem 3.2.16 in [Dre19]). Indeed, let (X_n) be an independent sequence of random variables, $\mathcal{G}_n := \sigma(X_n)$, and let $\mathcal{F}_n := \sigma(\cup_{i=1}^n \mathcal{G}_n)$. Then for F in the tail-σ-algebra, i.e.,

$$F \in \mathcal{T} := \bigcap_{n=1}^\infty \sigma\Big(\bigcup_{m=n}^\infty \mathcal{G}_m\Big).$$

we have $\mathbb{P}(F) \in \{0, 1\}$.

Indeed, since F is independent of $\mathcal{F}_n$ for any $n \in \mathbb{N}$, we have that $\mathbb{P}(F \mid \mathcal{F}_n) = \mathbb{P}(F)$. On the other hand, Corollary 2.2.8 implies that $\mathbb{P}$-a.s.,

$$\mathbb{P}(F \mid \mathcal{F}_n) \to \mathbb{P}(F \mid \mathcal{F}_\infty), \quad n \to \infty,$$

and the right-hand side equals $\mathbb{1}_F$ since $\mathcal{T} \subset \mathcal{F}_\infty$, so we have $\mathbb{P}(\mathbb{P}(F) \neq \mathbb{1}_F) = 0$, which means that $\mathbb{P}(F) \in \{0, 1\}$.

Exercise 2.3.13. *Derive the following version of the dominated convergence theorem using Theorem 2.3.9:*

Theorem. *Let X, X_n, $n \in \mathbb{N}$, be real random variables such that X_n converges to X in probability, and assume there is $Y \in \mathcal{L}^1$ such that for all $n \in \mathbb{N}$ we have $|X_n| \leq Y$ $\mathbb{P}$-a.s. Then $X \in \mathcal{L}^1$, and X_n converges to X in L^1, so in particular $\int X_n \,\mathrm{d}\mathbb{P} \to \int X \,\mathrm{d}\mathbb{P}$ as $n \to \infty$.*

In fact, if one introduces the notion of uniform integrability for general measure spaces and develops the analogous machinery for uniformly integrable families in this setting, one can deduce the general version of the dominated convergence theorem from the result corresponding to Theorem 2.3.9.

2.3.1 Convergence theorems for uniformly integrable martingales

As we have seen before, if a sequence of random variables is uniformly integrable, this might add to its convergence properties. In particular, this is true for martingales as we will see below. As a first observation, if (M_n) is a uniformly integrable martingale, then Lemma 2.3.4 implies that the family (M_n) is bounded in L^1 and hence (M_n) converges $\mathbb{P}$-a.s. as $n \to \infty$ due to Theorem 2.2.2. We can then use Theorem 2.3.9 to deduce further properties, and this is summarised in the following result.

Theorem 2.3.14. *Let (M_n), $n \in \mathbb{N}_0$, be a submartingale. Then the following are equivalent:*

(a) the family (M_n), $n \in \mathbb{N}_0$, is uniformly integrable;

(b) the sequence (M_n), $n \in \mathbb{N}_0$ converges $\mathbb{P}$-a.s. and in L^1 to $M_\infty \in \mathcal{L}^1$;

(c) the sequence (M_n), $n \in \mathbb{N}_0$ converges in L^1 to $M_\infty \in \mathcal{L}^1$.

Proof. '(a) $\Longrightarrow$ (b)': Since (M_n) is a uniformly integrable submartingale, Lemma 2.3.4 implies that the family (M_n) is bounded in L^1 and hence (M_n) converges $\mathbb{P}$-a.s. as $n \to \infty$ to some $M_\infty \in \mathcal{L}^1$ due to Theorem 2.2.2. Theorem 2.3.9 (c) implies that this convergence also is in L^1.
'(b) $\Longrightarrow$ (c)': This is evident.
'(c) $\Longrightarrow$ (a)': This is a direct consequence of Theorem 2.3.9. □

By the usual reasoning the previous theorem also applies to martingales and supermartingales.

Theorem 2.3.15. *If one of the equivalent conditions in Theorem 2.3.14 is fulfilled, then there exists $M_\infty \in \mathcal{L}^1$ such that*

$$M_n = \mathbb{E}[M_\infty \mid \mathcal{F}_n] \quad \forall n \in \mathbb{N}$$

(with '$=$' replaced by '$\leq$' and '$\geq$' in the case of (M_n) being a submartingale and supermartingale, respectively).

Proof. This follows from Proposition 2.3.11. □

Compare the previous theorem to Example 2.0.8, where we had shown that for any random variable X and a filtration $(\mathcal{F}_n)$, the sequence $(\mathbb{E}[X \mid \mathcal{F}_n])$ supplies us with a martingale. The last item of the above is a converse to this, and a martingale with this property is also called *right-closable* (which, due to the above, is the same as the martingale being uniformly integrable). In the case where we do have that a martingale even is L^p-bounded, $p \in (1, \infty)$, we can strengthen the convergence given above.

Theorem 2.3.16. *Let $p \in (1, \infty)$, and let (M_n) be a martingale which is L^p-bounded. Then there exists $M_\infty \in \mathcal{L}^p$ such that $M_n \to M_\infty$ $\mathbb{P}$-a.s. and in L^p.*
In particular, the sequence $(|M_n|^p)$ is uniformly integrable.

Proof. Since

$$\sup_{n \in T} \mathbb{E}[|M_n|] \le \sup_{n \in T} \mathbb{E}[|M_n|^p]^{\frac{1}{p}} < \infty$$

since (M_n) is L^p-bounded, our plain vanilla martingale convergence theorem, Theorem 2.2.2, implies that (M_n) converges $\mathbb{P}$-a.s. to some $M_\infty \in \mathcal{L}^1$.
Doob's L^p-inequality (Theorem 2.1.23) stated that

$$\mathbb{E}[(|M|_n^*)^p] \le \Big(\frac{p}{p-1}\Big)^p \mathbb{E}[|M_n|^p],$$

and since the right-hand side is bounded in n by assumption, we deduce taking $n \to \infty$ on the left-hand side that $M_\infty \in \mathcal{L}^p$. Furthermore, since we have

$$|M_n - M_\infty| \le 2 \lim_{n \to \infty} |M|_n^*,$$

and since the right-hand side is in $\mathcal{L}^p$ due to the previous, we deduce by the DCT and $M_n \to M_\infty$ $\mathbb{P}$-a.s., that $M_n \to M_\infty$ in L^p also.
The stated uniform integrability immediately follows from

$$|M_n| \le \lim_{n \to \infty} |M|_n^* \in \mathcal{L}^p.$$

□

Example 2.3.17. *If we investigate our favourite martingale simple random walk (S_n) under the light of convergence, then, since $|S_{n+1} - S_n| = 1$, we know that it does not converges $\mathbb{P}$-a.s., and this is consistent with the above since (S_n) is not uniformly integrable.*

2.3.2 Square integrable martingales

As we have seen before in the context of conditional expectations, having square integrability can nicely bring in geometry. This is also one of the reasons why square integrable martingales deserve some special attention. Indeed, in the words of Williams [Wil91]: 'When it works, one of the easiest ways of proving that a martingale M is bounded in L^1 is to prove that it is bounded in L^2...'
Another reason is that square integrable naturally have finite second moments which is necessary if you might hope for a central limit theorem to hold at all. As it will turn out, we can actually prove such a central limit theorem, which is yet another manifestation that martingales can be considered as a generalisation of sums of i.i.d. variables.[4]

[4] Indeed, in [Dre19] we had only studied a central limit theorem for i.i.d. random variables. There are, however, pretty general versions of central limit theorems for sequences of dependent random variables also (which are not totally easy to prove), see e.g. Section 7.7 of [Dur10]. We will focus here on the case of martingales since it fits our studies and still comes relatively simple, but already with some important ideas.

Theorem 2.3.18. *If (M_n) is a square integrable martingale with $M_0 = 0$ then*

$$\mathbb{E}[\sup_{n\in\mathbb{N}_0} M_n^2] \le 4\mathbb{E}[\langle M\rangle_\infty].$$

Proof. For $p = 2$, Doob's L^p-inequality (Theorem 2.1.23) yields

$$\mathbb{E}[\sup_{k\in\{0,\dots,n\}} M_k^2] \le 4\mathbb{E}[M_n^2] = 4\mathbb{E}[\langle M\rangle_n],$$

and taking $n \to \infty$ on both sides as well as applying the monotone convergence theorem, we obtain the desired result. □

Corollary 2.3.19. *In the setting of Theorem 2.3.18, on the set $\{\langle M\rangle_\infty < \infty\}$, $\mathbb{P}$-a.s. the limit $\lim_{n\to\infty} M_n$ exists and is finite.*

Proof. Again denoting by $(\langle M\rangle_n)$ the square variation process of (M_n), we get for $K \in (0,\infty)$ that $\tau_K := \inf\{n \in \mathbb{N} : \langle M\rangle_{n+1} > K\}$ is a stopping time (since $(\langle M\rangle_n)$ is a predictable process). Hence, the optional sampling theorem yields that $M_{\tau_K\wedge n}$ is a (in fact square integrable) martingale (with square variation $\langle M_{\tau_K\wedge\cdot}\rangle_n = \langle M\rangle_{\tau_K\wedge n} \le K$ for all $n \in \mathbb{N}$) and thus Theorem 2.3.18 implies that

$$\mathbb{E}[\sup_{k\in\{0,\dots,n\}} M_{\tau_K\wedge n}^2] \le 4K.$$

Thus, Theorem 2.3.16 implies that $\lim_{n\to\infty} M_{\tau_K\wedge n}$ exists and is finite $\mathbb{P}$-a.s. Taking $K \to \infty$ the result follows. □

Corollary 2.3.20. *Let (M_n) be a square integrable martingale with square variation process $(\langle M\rangle_n)$. The following are equivalent:*

(a)

$$\lim_{n\to\infty} \mathbb{E}[\langle M\rangle_n] < \infty;$$

(b)

$$\sum_{i=1}^{\infty} \mathbb{E}[(M_{i+1} - M_i)^2] < \infty;$$

(c)

$$\sup_{n\in\mathbb{N}} \mathbb{E}[M_n^2] < \infty;$$

(d) (M_n) converges in L^2;

(e) (M_n) converges $\mathbb{P}$-a.s. and in L^2.

Proof. The equivalence of (a), (b), and (c) follows from (2.0.12).
The implications (e) implies (d) implies (c) are obvious.
The implication (c) implies (e) follows from Theorem 2.3.16. □

2.3.3 A central limit theorem for martingales

Here, as in other contexts, it turns out conceptually useful to consider the *increments* of a martingale (also called *martingale differences*) instead of the martingale itself. This is captured in the following definition, and it also underlines the interpretation of a martingale of sums of independent random variables.

Definition 2.3.21. *If (X_n), $n \in \mathbb{N}$, is a stochastic process with $X_n \in \mathcal{L}^1$ and adapted to $(\mathcal{F}_n)$, then it is called a* martingale difference sequence *if $\mathbb{E}[X_{n+1} \,|\, \mathcal{F}_n] = 0$ for all $n \in \mathbb{N}$.*
If for each $n \in \mathbb{N}$ we have a (finite) sequence of random variables $(X_{n,k})$, $1 \le k \le m_n$, some $m_n \in \mathbb{N}$, with $X_{n,k} \in \mathcal{L}^1$, and a corresponding filtration $(\mathcal{F}_{n,k})$, $1 \le k \le m_n$, such that for all $n \in \mathbb{N}$ we have that for all $k \in \{1, \dots, m_n\}$, the equation

$$\mathbb{E}[X_{n,k} \,|\, \mathcal{F}_{n,k-1}] = 0$$

holds true (with $\mathcal{F}_{n,0} := \{\Omega, \emptyset\}$), then we call the pair $(X_{n,k})$, $(\mathcal{F}_{n,k})$, $1 \le k \le m_n$, a martingale difference array.

In particular, if (M_n) is a martingale with respect to $(\mathcal{F}_n)$, then $X_n := M_n - M_{n+1}$ is a martingale difference sequence. Hence, in our heuristic correspondence of martingales to generalisations of sums of independent random variables, martingale differences correspond to the independent random variables themselves.

Theorem 2.3.22. *Let $(X_{n,k})$, $1 \le k \le m_n$, some $m_n \in \mathbb{N}$, and $(\mathcal{F}_{n,k})$, $1 \le k \le m_n$, as above be a martingale difference array, and set*

$$S_{n,k} := \sum_{i=1}^{k} X_{n,i}, \quad n \in \mathbb{N},\, k \in \{0, 1, \dots, m_n\}.$$

If

$$\mathbb{E}[\max_{i \in \{1,\dots,m_n\}} |X_{n,i}|] \to 0 \quad \text{as } n \to \infty, \tag{2.3.13}$$

and if in addition for some $\sigma^2 \in (0, \infty)$ we have that

$$\sum_{i=1}^{m_n} X_{n,i}^2 \to \sigma^2 \quad \text{in probability as} \quad n \to \infty, \tag{2.3.14}$$

then

$$S_{n,m_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2) \quad \text{as } n \to \infty.$$

Remark 2.3.23. *In order to put this result into context, we compare it to the standard CLT we know from 'Probability Theory I', see Theorem 3.8.1 of [Dre19] (with $d = 1$ for the sake of simplicity).*
Indeed, for (X_n) a sequence of i.i.d. random variables with $\sigma^2 := \mathbb{E}[X_1^2] \in (0, \infty)$, [Dre19, Theorem 3.8.1] states that

$$\frac{\sum_{i=1}^{n} (X_i - \mathbb{E}[X_i])}{\sqrt{n}}$$

converges in distribution to a $\mathcal{N}(0, \sigma^2)$-distributed random variable.
This can be recovered from the previous result by setting $m_n := n$, $X_{n,i} := (X_i - \mathbb{E}[X_i])/\sqrt{n}$, $1 \le i \le n$, and $\mathcal{F}_{n,k} := \sigma(X_1, \dots, X_k)$ for any $1 \le k \le n$. It is easy to check that this actually defines a martingale difference array, and furthermore we have

$$\mathbb{E}\big[\max_{i \in \{1,\dots,n\}} |X_{n,i}|\big] = \int_0^\infty \mathbb{P}(\max_{i \in \{1,\dots,n\}} X_i^2 \ge n s^2) \,\mathrm{d}s = \int_0^\infty 1 - \mathbb{P}(X_1^2 < n s^2)^n \,\mathrm{d}s.$$

But since $X_1 \in \mathcal{L}^2$, we have

$$\mathbb{E}[X_1^2] = \int_0^\infty \mathbb{P}(X_1^2 \ge s) \,\mathrm{d}s < \infty,$$

whence the integral test of convergence ('Integralkriterium') yields $\mathbb{P}(X_1^2 \geq s) \in o((s \ln s)^{-1})$*, so continuing the above we get*

$$\mathbb{E}\big[\max_{i\in\{1,\dots,n\}} |X_{n,i}|\big] \leq \int_0^\infty 1-\Big(1-\frac{1}{1 \vee n(s^2 \ln(ns^2))}\Big)^n \,\mathrm{d}s \leq \int_0^\infty \frac{1}{1 \vee (s^2 \ln(ns^2))}\,\mathrm{d}s \to 0 \qquad \textit{as } n \to \infty.$$

Furthermore, the (weak) law of large numbers yields

$$\sum_{i=1}^n X_{n,i}^2 = \frac{1}{n}\sum_{i=1}^n X_i^2 \to \mathbb{E}[X_1^2] = \sigma^2 \qquad \textit{in probability as} \qquad n \to \infty,$$

which establishes (2.3.14).

Before we can actually prove this result we need the following lemma.

Lemma 2.3.24. *Let* (X_n) *and* (Y_n) *be two sequences of random variables such that the following three conditions hold true:*

(a) There exists a constant $c \in \mathbb{R}$ *such that*

$$X_n \xrightarrow{\mathbb{P}} c;$$

(b) the sequences (Y_n) *and* $(X_n Y_n)$ *are both uniformly integrable;*

(c)

$$\mathbb{E}[Y_n] \to 1 \qquad \textit{as } n \to \infty.$$

Then

$$\mathbb{E}[X_n Y_n] \to c \qquad \textit{as } n \to \infty.$$

Proof of Theorem 2.3.22. Dividing all elements of the entire array by $\sqrt{\sigma^2}$ we can assume $\sigma^2 = 1$ without loss of generality.
In order to obtain a good quality of the Taylor expansion below, in a standard cutoff procedure we replace the original array by the random variables

$$\widetilde{X}_{n,1} := X_{n,1}, \quad \widetilde{X}_{n,k} := X_{n,k} \mathbb{1}_{\sum_{i=1}^{k-1} X_{n,i}^2 \leq 2}, \qquad k \in \{2, \dots, m_n\}.$$

Then $(\widetilde{X}_{n,k})$ still constitutes a martingale difference array since we have that

$$\Big\{\sum_{i=1}^{k-1} X_{n,i}^2 \leq 2\Big\} \in \mathcal{F}_{n,k-1}.$$

Introducing the stopping times

$$\tau_n := \inf\Big\{k \in \mathbb{N} : \sum_{i=1}^k X_{n,i}^2 > 2\Big\} \wedge m_n$$

we infer that

$$\mathbb{P}\big(X_{n,i} \neq \widetilde{X}_{n,i} \text{ some } i \in \{1, \dots, m_n\}\big) = \mathbb{P}(\tau_n < m_n) \leq \mathbb{P}\Big(\sum_{i=1}^{m_n} X_{n,i}^2 > 2\Big) \to 0 \qquad \text{as } n \to \infty,$$

where the convergence follows in combination with (2.3.14). Hence it is sufficient to prove that

$$\widetilde{S}_n := \sum_{i=1}^{m_n} \widetilde{X}_{n,i}$$

converges in distribution to $\mathcal{N}(0,1)$. Now recall that due to Theorem 3.8.10 of [Dre19], for this purpose it is sufficient to show that

$$\mathbb{E}[e^{\mathrm{i}t\widetilde{S}_n}] \to e^{-t^2/2} \qquad \text{as } n \to \infty.$$

It turns out that a suitable way to write the Taylor expansion of the exponential function around 0 in this context is given by

$$e^{\mathrm{i}x} = (1+\mathrm{i}x)e^{-\frac{x^2}{2}+\epsilon(x)},$$

where ϵ describes the error and satisfies

$$|\epsilon(x)| \le C|x|^3 \tag{2.3.15}$$

for some constant $C \in (0,\infty)$ and all $|x|$ small enough. As a consequence,

$$e^{\mathrm{i}t\widetilde{S}_n} = \underbrace{\Big(\prod_{i=1}^{m_n}(1+\mathrm{i}t\widetilde{X}_{n,i})\Big)}_{=:Y_n}\underbrace{\exp\Big\{-\frac{t^2}{2}\sum_{i=1}^{m_n}\widetilde{X}_{n,i}^2+\sum_{i=1}^{m_n}\epsilon(t\widetilde{X}_{n,i})\Big\}}_{=:X_n}. \tag{2.3.16}$$

We now show that the assumptions of Lemma 2.3.24 hold true. Indeed, we have

$$\Big|\sum_{i=1}^{m_n}\epsilon(t\widetilde{X}_{n,i})\Big| \le C|t|^3\sum_{i=1}^{m_n}|\widetilde{X}_{n,i}|^3 \le C|t|^3\sum_{i=1}^{m_n}|X_{n,i}|^3 \le C|t|^3\max_{i\in\{1,\dots,m_n\}}|X_{n,i}|\sum_{i=1}^{m_n}|X_{n,i}|^2 \xrightarrow{\mathbb{P}} 0,$$

where in the first inequality we used (2.3.15), and in the convergence we took advantage of (2.3.14) and (2.3.13), where the latter implied that for $\varepsilon > 0$ we have

$$\mathbb{P}\big(\max_{i\in\{1,\dots,m_n\}}|X_{n,i}| \ge \varepsilon\big) \le \varepsilon^{-1}\mathbb{E}\big[\max_{i\in\{1,\dots,m_n\}}|X_{n,i}|\big] \to 0 \qquad \text{as } n\to\infty.$$

Using (2.3.14) again, this implies

$$X_n \xrightarrow{\mathbb{P}} e^{-\frac{t^2}{2}},$$

and hence assumption (a) of Lemma 2.3.24 is fulfilled with $c = e^{-\frac{t^2}{2}}$.
Now by definition we have $|\widetilde{X}_{n,k}| \le 2$ for all $n \in \mathbb{N}$, $k \in \{1,\dots,m_n\}$, and so we deduce in combination with the tower property, the fact that we're dealing with a martingale difference sequence, and Theorem 1.0.9 (e) that

$$\begin{aligned}
\mathbb{E}[Y_n] &= \mathbb{E}\Big[\mathbb{E}\Big[(1+\mathrm{i}t\widetilde{X}_{n,m_n})\prod_{i=1}^{m_{n-1}}(1+\mathrm{i}t\widetilde{X}_{n,i})\,\big|\,\mathcal{F}_{n,m_n-1}\Big]\Big]\\
&= \mathbb{E}\Big[\prod_{i=1}^{m_{n-1}}(1+\mathrm{i}t\widetilde{X}_{n,i})\underbrace{\mathbb{E}\Big[(1+\mathrm{i}t\widetilde{X}_{n,m_n})\,\big|\,\mathcal{F}_{n,m_n-1}\Big]}_{=1}\Big]\\
&= \mathbb{E}\Big[\prod_{i=1}^{m_{n-1}}(1+\mathrm{i}t\widetilde{X}_{n,i})\Big] = \ldots = 1.
\end{aligned}$$

In particular, this means that (c) of Lemma 2.3.24 holds true.
Also, note that the sequence (X_nY_n) is uniformly integrable since we have $|X_nY_n| = 1$ for all $n \in \mathbb{N}$ due to (2.3.16). Therefore, the only thing to check in order to apply Lemma 2.3.24 is the uniform integrability of the sequence (Y_n). Note that due to

$$|1+\mathrm{i}x|^2 = 1+x^2 \le e^{x^2}$$

we infer, taking square roots on both sides, that

$$|Y_n| = \prod_{i=1}^{\tau_n} |1 + \mathrm{i}tX_{n,i}| \le \exp\Big\{\frac{t^2 \sum_{i=1}^{\tau_n - 1} X_{n,i}^2}{2}\Big\}(1 + |tX_{n,\tau_n}|) \le e^{t^2}(1 + |t| \max_{i\in\{1,\dots,m_n\}} |X_{n,i}|).$$

But now note that the sequence $(\max_{j\in\{1,\dots,m_n\}} |X_{n,j}|)$ is uniformly integrable due to Theorem 2.3.9 since it converges to 0 in L^1, and hence the same applies to the right-hand side and thus the left-hand side of the previous display.

□

Remark 2.3.25. *In exactly the same way as Example 3.8.12 of [Dre19], the above Theorem 2.3.22 implies that $S_{n,m_n}/\sqrt{m_n}$ converges in probability to 0, which can be interpreted as a weak law of large numbers type result.*

Since it makes a better fit and will prop up quite naturally, we will see applications of Theorem 2.3.22 in Chapter 4 below.

2.4 Backwards martingales

The martingale convergence results we have investigated so far corresponded to more and more information becoming available. In particular, on a heuristic level this means that the case of a martingale converging to a constant is the exception rather than the rule. However, if you recall fundamental results such as the law of large numbers, it becomes evident that convergence to constants also plays a significant role in probability theory, and this is in some sense better captured by the backwards martingales defined below.
The general definition of a martingale given in Definition 2.0.6 did not presume the index set to be the natural numbers. In fact, one of the reasons for keeping things general over there is becoming apparent in this section.

Definition 2.4.1. *If (M_n), $n \in -\mathbb{N}_0$, (or $n \in -\mathbb{N}$) is a martingale with respect to a filtration $(\mathcal{F}_n)$, $n \in -\mathbb{N}_0$, (or $n \in -\mathbb{N}$) then (M_n) is also called a* backwards martingale *(also:* reversed martingale*) ('Rückwärtsmartingal'). Similarly, if it is a sub- or supermartingale, it is called a* backwards sub- *or* supermartingale.

Note that the 'right-closedness' property we had established for (uniformly integrable) martingales holds by definition (i.e., due to the martingale property) for backwards martingales: $M_n = \mathbb{E}[M_0 \,|\, \mathcal{F}_n]$ for all $n \in -\mathbb{N}_0$. Equivalently, cf. Example 2.3.2 (f), we observe that backwards martingales are always uniformly integrable and hence in general as well-behaved as only the nicest martingales can be. This will also manifest itself in the convergence theorem below as well as in its applications.
For a filtration $(\mathcal{F}_n)$, $n \in -\mathbb{N}$, we define, analogously to $\mathcal{F}_\infty$, the σ-algebra

$$\mathcal{F}_{-\infty} := \bigcap_{n\in-\mathbb{N}_0} \mathcal{F}_n = \lim_{n\to-\infty} \mathcal{F}_n.$$

We now directly dive into our main convergence result for backwards martingales, the proof of which also takes advantage of our basic upcrossing lemma, Lemma 2.2.1.

Theorem 2.4.2. *Let (M_n), $n \in -\mathbb{N}_0$, be a backwards martingale. Then*

$$M_n \to M_{-\infty} = \mathbb{E}[M_0 \,|\, \mathcal{F}_{-\infty}] \quad \text{in } L^1 \text{ and } \mathbb{P}-a.s. \quad \text{as } n \to -\infty.$$

Similarly, if (M_n), $n \in -\mathbb{N}_0$, is a backwards submartingale with $\inf_{n\in-\mathbb{N}_0} \mathbb{E}[M_n] > -\infty$, then (M_n), $n \in -\mathbb{N}_0$, is a uniformly integrable family and

$$M_{-\infty} := \lim_{n\to-\infty} M_n \quad \text{exists in } L^1 \text{ and } \mathbb{P}-a.s. \quad \text{as } n \to -\infty,$$

and

$$M_\infty \le \mathbb{E}[M_n \,|\, \mathcal{F}_{-\infty}] \quad \forall n \in -\mathbb{N}_0.$$

Proof. We give the proof in the case of a martingale and leave the case of a submartingale to the Reader. The proof proceeds quite similarly to the proof of the standard martingale convergence theorem.
Denoting for $n \in -\mathbb{N}_0$ and $a, b \in \mathbb{R}$ with $a < b$ by $U_n[a, b]$ the number of upcrossings of $M_n, M_{n+1}, \dots, M_0$ of the interval $[a, b]$ similarly to Section 2.2, Lemma 2.2.1 provides us with

$$\mathbb{E}\big[U_n[a, b]\big] \le \frac{\mathbb{E}[(M_0 - a)^+]}{b - a},$$

Defining the monotone non-decreasing limit $U_\infty[a, b] := \lim_{n\to-\infty} U_n[a, b]$ we infer from the previous display in combination with the monotone convergence theorem that

$$\mathbb{E}\big[U_\infty[a, b]\big] \le \frac{\mathbb{E}[(M_0 - a)^+]}{b - a}.$$

As in the proof of Theorem 2.2.2 we therefore conclude that the limit $M_{-\infty} := \lim_{n\to-\infty} M_n$ exists $\mathbb{P}$-a.s. in $[-\infty, \infty]$. Since the family (M_n), $n \in -\mathbb{N}_0$, is uniformly integrable due to Example 2.3.2 (f), and since $M_{-\infty} \in \mathcal{L}^1$ (check!), we deduce that $M_n \to M_{-\infty}$ in L^1 as well due to Theorem 2.3.9.
In order to show that

$$M_{-\infty} = \mathbb{E}[M_0 \,|\, \mathcal{F}_{-\infty}], \tag{2.4.1}$$

as before for $F \in \mathcal{F}_{-\infty}$ we have due to $\mathcal{F}_{-\infty} \subset \mathcal{F}_n$ that

$$\mathbb{E}[M_0 \mathbb{1}_F] = \mathbb{E}[M_n \mathbb{1}_F] \quad \forall n \in -\mathbb{N}_0,$$

and taking $n \to -\infty$ this becomes $\mathbb{E}[M_0 \mathbb{1}_F] = \mathbb{E}[M_{-\infty} \mathbb{1}_F]$ due to uniform integrability and Theorem 2.3.9, which by definition of the conditional expectation shows (2.4.1). □

Exercise 2.4.3. *Show that if in the setting of the previous theorem, $M_0 \in \mathcal{L}^p$ some $p \in (1, \infty)$, then $M_n \to M_{-\infty}$ even in L^p as $n \to \infty$.*

The above convergence result is sort of the converse to what we had seen before in martingale convergence: Indeed, taking $n \to -\infty$ corresponds to having less and less information available. We will give another example of an application of convergence of backwards martingales here.

Example 2.4.4 (Ballot theorem). *Consider (X_n), $1 \le n \le N \in \mathbb{N}$ a finite sequence of $\mathbb{N}_0$-valued i.i.d. random variables and set $S_n := \sum_{i=1}^n X_i$. The claim is that*

$$\mathbb{P}(\underbrace{S_n < n \,\forall n \in \{1, \dots, N\}}_{=:G} \,|\, S_N) = \Big(1 - \frac{S_N}{N}\Big)^+. \tag{2.4.2}$$

To motivate the name of this problem, consider the case where N voters each consecutively have to cast a vote for one of two candidates A and B, and where the event $\{X_n = 2\}$ corresponds to the n-th voter casting her vote for candidate A, and the event $\{X_n = 0\}$ corresponds to the n-th voter casting her vote for candidate B. Then candidate B leads at time n if and only if $S_n < n$, so the event in the probability of (2.4.2) corresponds to B having led throughout the entire voting process from the first to the N-th vote.

On the event $\{S_N \geq N\}$ the statement of (2.4.2) trivially holds true, so assume $S_N < N$ from now on. The crucial observation here is that defining for $n \in \{1, \ldots, N\}$ the random variables $M_{-n} := S_n/n$, they can be interpreted as a backwards martingale with respect to the filtration characterised by $\mathcal{F}_{-n} := \sigma(S_n, S_{n+1}, \ldots, S_N)$. Indeed, we start with noting that $\mathcal{F}_{-n} = \sigma(S_n, X_{n+1}, \ldots, X_N)$, which in combination with the fact that the (X_i) are i.i.d. implies that for $1 \leq i \leq n$ we have $\mathbb{E}[X_i \,|\, \mathcal{F}_{-n}] = \mathbb{E}[X_1 \,|\, \sigma(S_n)] = \frac{S_n}{n}$. As a consequence,

$$\begin{aligned}\mathbb{E}[M_{-n+1} \,|\, \mathcal{F}_{-n}] &= \mathbb{E}\Big[\frac{S_{n-1}}{n-1} \,|\, \mathcal{F}_{-n}\Big] = \mathbb{E}\Big[\frac{S_n}{n-1} - \frac{X_n}{n-1} \,|\, \mathcal{F}_{-n}\Big] \\ &= \frac{S_n}{n-1} - \frac{S_n}{n(n-1)} = \frac{S_n}{n} = M_{-n},\end{aligned}$$

which establishes the martingale property.
Furthermore, we claim that setting

$$\tau := \begin{cases} \max\big\{n \in \{1, \ldots, N\} \,:\, S_n \geq n\big\}, & \text{if the set on the RHS is non-empty,} \\ 1, & \text{otherwise,} \end{cases}$$

we have that $-\tau$ is a stopping time with respect to $(\mathcal{F}_{-n})$. Indeed, we have

$$\{-\tau = -n\} = \{\tau = n\} = \big\{S_n \geq n \text{ and } S_k < k \,\forall k \in \{n+1, \ldots, N\}\big\} \in \mathcal{F}_{-n}$$

for $n \in \{2, \ldots, N\}$, and the case $n = 1$ can be proven separately by a distinction of cases regarding the definition of τ; this shows the stopping time property for $-\tau$.
Therefore, the optional sampling theorem yields

$$\mathbb{E}[M_{-\tau} \,|\, \mathcal{F}_{-N}] = M_{-N} = S_N/N. \tag{2.4.3}$$

We now claim that $M_{-\tau} = 1$ on G^c. To check this, note that if $S_{j+1} < j+1$, then the fact that the X_i are non-negative integers implies $S_j \leq S_{j+1} \leq j$. Furthermore, since $G \subset \{-\tau = -1\}$ and $S_1 < 1$ implies $S_1 = 0$, we have that $M_{-\tau} = 0$ on G.
Combining the two we get

$$M_{-\tau} = \mathbb{1}_{G^c}.$$

Hence, in combination with $\mathcal{F}_{-N} = \sigma(S_N)$, display (2.4.3) implies

$$\mathbb{P}(G \,|\, S_N) = 1 - \mathbb{P}(G^c \,|\, S_N) = 1 - S_N/N,$$

which proves the result.

Exercise 2.4.5. *Using parts of the results of the previous Example 2.4.4 show that an i.i.d. sequence of random variables $(X_n)_{n \in \mathbb{N}}$ with $X_n \in \mathcal{L}^1$ satisfies a strong law of large numbers, and show that this convergence holds in L^1 as well.*
In particular, note that this result is as strong as we can get in terms of as weak as possible integrability assumptions on (X_n), and in particular it is stronger than Theorem 3.6.9 in [Dre19], where we had to assume $X_n \in \mathcal{L}^4$.

2.5 A discrete Black-Scholes formula

The Black-Scholes formula for pricing options is a central result of probabilistic financial mathematics. In this section, we will with the help of martingale theory discuss a discrete version of this formula.
In our (simple) economy, we will consider two types of securities (documents holding a certain worth): *bonds* that deliver a fixed rate r of return, and *shares/stocks*, whose value fluctuates. We will write S_n for the value of a share during the time interval $(n, n+1)$ and $B_n = (1+r)^n B_0$

for the corresponding value of a bond. A *portfolio* (A_n, V_n) at time n consists of A_n shares and V_n bonds. With that, the initial capital of the investor is

$$x = X_0 := A_0 S_0 + V_0 B_0$$

and through trading during the time interval $(0, 1)$, the investor can obtain any portfolio (A_1, V_1), which satisfies $x = A_1 S_0 + V_1 B_0$. The value of this portfolio in the interval $(1, 2)$ is then

$$X_1 := A_1 S_1 + V_1 B_1$$

and so on. The income in the n-th step is therefore

$$X_n - X_{n-1} = A_n(S_n - S_{n-1}) + V_n(B_n - B_{n-1}).$$

Since we can write $B_n - B_{n-1} = rB_{n-1}$ and $S_n - S_{n-1} = R_n S_{n-1}$ for a fluctuating rate R_n, we obtain

$$X_n - X_{n-1} = rX_{n-1} + A_n S_{n-1}(R_n - r).$$

It is often easier to consider the *discounted* value of a portfolio, defined as

$$Y_n := (1 + r)^{-n} X_n,$$

which leads to

$$Y_n - Y_{n-1} = (1 + r)^{-n+1} A_n S_{n-1}(R_n - r).$$

In our simplified model, we will only allow R_n to take two possible values, specifically $a \in (-1, r)$ and $b \in (r, \infty)$.
Problem: A European option allows the holder of the option to buy a share at time $n = N$ for the price K. What is a fair price for such an option at time $n = 0$?
A *hedging strategy* with starting value x for the option is a predictable process $(A_n, V_n) : n \in \mathbb{N})$ as described above, such that $X_n \geq 0$ for all $n \in \{0, \ldots N\}$ and such that

$$X_N = (S_N - K)^+.$$

Note that the left hand side is precisely the value of portfolio at time N, while the right hand side is the value of the option at time N. Black and Scholes postulated that x is a fair price for the option precisely when there exists a portfolio with starting value x, whose value at time N matches the value of the option, independently of the evolution of the *underlying* stock price. Note that this statement does not make any assumption on the choice of the underlying model.

Theorem 2.5.1. *A hedging strategy with starting value x exists if and only if*

$$x = \mathbb{E}[(1 + r)^{-N}(S_N - k)^+],$$

where the expectation is with respect to i.i.d. R_n, that take value b with probability $p = \frac{r-a}{b-a}$ and value a with probability $1 - p = \frac{b-r}{b-a}$.

Proof. Assume that a trading strategy with starting value x exists. We consider a sequence of i.i.d. random variables $\epsilon_1, \epsilon_2, \ldots$ with $\mathbb{P}(\epsilon_i = 1) = p = 1 - \mathbb{P}(\epsilon_i = -1)$ and set

$$Z_n = \sum_{k=1}^{n} (\epsilon_k - 2p + 1).$$

With this choice (Z_n) is a martingale and we can express R_n from our model as

$$R_n = r + \frac{1}{2}(b - a)(Z_n - Z_{n-1}).$$

Using this we can consider Y_n as a stochastic integral, namely

$$Y_n = Y_0 + (F \bullet Z)_n,$$

where $F_n = (1+r)^{-n+1} A_n S_{n-1}$. Importantly, such F_n is predictable and bounded. Using Theorem 2.0.15 we obtain that Y_n is therefore also a martingale. For $Y_0 = x$ it follows

$$x = \mathbb{E}[Y_N] = \mathbb{E}[(1+r)^{-N} X_N] = \mathbb{E}[(1+r)^{-N}(S_N - K)^+],$$

as claimed. Note that non-negativity of X_n was not required.
We now show the converse - that for our choice of x a hedging strategy exists. We define therefore

$$Y_n := \mathbb{E}[(1+r)^{-N}(S_N - K)^+ \,|\, \mathcal{F}_n].$$

With this definition (Y_n) is a martingale and it remains to show that a predictable process (A_n) exists, for which

$$Y_n - Y_{n-1} = (1+r)^{-n+1} A_n S_{n-1}(R_n - r).$$

Assuming this existence, we can then set $X_n = (1+r)^n Y_n$ and $V_n = (X_n - A_n S_n)/B_n$. This ensures that the equations that connect X_n with (A_n, V_n) hold and furthermore that $((A_n, V_n))$ is predictable. Since $X_0 = Y_0 = x$ and $X_N = \mathbb{E}[(S_N - K)^+ \,|\, \mathcal{F}_N] = (S_N - K)^+$ we have found the desired hedging strategy. It remains to show the existence of A_n, which is implied by the following lemma. □

Lemma 2.5.2. *Let $(M_n : 0 \leq n \leq N)$ be a martingale with respect to the filtration $\mathcal{F}_n = \sigma(\epsilon_1, \ldots, \epsilon_n)$ and let $Z_n = \sum_{k=1}^n (\epsilon_k - 2p + 1)$. Then there exists a predictable process (H_n) such that*

$$M_n = M_0 + \sum_{k=1}^n H_k(Z_k - Z_{k-1}).$$

Proof. Since M_n is measurable with respect to $\mathcal{F}_n$, there exists a function

$$f_n : \{-1, 1\}^n \to \mathbb{R}$$

such that $M_n = f_n(\epsilon_1, \ldots, \epsilon_n)$. The martingale property gives

$$\begin{aligned} 0 &= \mathbb{E}[M_n - M_{n-1} \,|\, \mathcal{F}_{n-1}] \\ &= p f_n(\epsilon_1, \ldots, \epsilon_{n-1}, 1) + (1-p) f_n(\epsilon_1, \ldots, \epsilon_{n-1}, -1) - f_{n-1}(\epsilon_1, \ldots, \epsilon_{n-1}). \end{aligned}$$

Setting

$$\begin{aligned} H_n &= \frac{f_n(\epsilon_1, \ldots, \epsilon_{n-1}, 1) - f_{n-1}(\epsilon_1, \ldots, \epsilon_{n-1})}{2(1-p)} \\ &= \frac{f_{n-1}(\epsilon_1, \ldots, \epsilon_{n-1}) - f_n(\epsilon_1, \ldots, \epsilon_{n-1}, -1)}{2p}, \end{aligned}$$

the result follows by simple algebra. □

2.6 A discrete Itô-formula

Itô's formula (also known as Itô's-Doeblin theorem in recognition of posthumously published work of Wolfgang Doeblin) is a useful tool, in particular in stochastic analysis and applications to mathematical finance. Since this will not be the subject of this course, we will restrict ourselves to giving a discrete version of it, which already provides some insight and intuition. Indeed, when (M_n) is a martingale and we investigate the process $f(M_n)$ obtained by applying a nice

function to the original process (M_n), then Itô's lemma supplies us with the Doob decomposition of this process into a martingale (expressed as a stochastic integral) and a previsible process – this usually helps in obtaining a better understanding of the process.
From a different point of view, Itô's lemma can also be seen as a generalisation of the Fundamental Theorem of Calculus which says that if $f : \mathbb{R} \to \mathbb{R}$ is continuously differentiable, then

$$f(x) = \int_0^x f'(s)\, \mathrm{d}s + f(0).$$

In particular, this tells us that we can write the function f as the integral of an appropriate function, namely its derivative. In view of the discrete stochastic integral introduced in Definition 2.0.14 one might hope that we can also write

$$f(X_n) = (f' \bullet X)_n + C \tag{2.6.1}$$

for some constant C and a martingale (X_n). As it turns out, this is not entirely true but we will be able to recover a corresponding result.
In the following it suggests itself to either consider continuous time processes, i.e., processes (X_t), $t \in [0, \infty)$, or otherwise $\mathbb{Z}$-valued processes in discrete time, and the latter is what we will actually stick to for the remaining part of this section. In order to formulate a result corresponding to the Fundamental Theorem of Calculus we have to introduce the discrete derivative of a function $f : \mathbb{Z} \to \mathbb{R}$ in a standard way as

$$f'(x) = \frac{f(x+1) - f(x-1)}{2}, \quad x \in \mathbb{Z},$$

as well as its second derivative (or Laplace operator)

$$f''(x) = f(x+1) - 2f(x) + f(x-1), \quad x \in \mathbb{Z}.$$

Now let (X_n) be a $\mathbb{Z}$-valued process with $X_{n+1} - X_n \in \{-1, 1\}$ for all $n \in \mathbb{N}_0$. Decomposing according to the possible increments we obtain that

$$\begin{aligned} f(X_{n+1}) - f(X_n) &= \frac{1}{2}\Big((f(X_n+1) - f(X_n-1))(X_{n+1} - X_n) + f(X_n+1) + f(X_n-1)\Big) - f(X_n) \\ &= f'(X_n)(X_{n+1} - X_n) + \frac{1}{2} f''(X_n), \end{aligned}$$

and hence summation yields the discrete Itô formula

$$f(X_n) = \sum_{i=0}^{n-1} f'(X_i)(X_{i+1} - X_i) + \frac{1}{2}\sum_{i=0}^{n-1} f''(X_i) + f(X_0) = (f'(X_\cdot) \bullet X)_n + \frac{1}{2}\sum_{i=0}^{n-1} f''(X_i) + f(X_0). \tag{2.6.2}$$

The first term on the right-hand side is the stochastic integral, but in addition to the terms expected in (2.6.1), we see that there is an additional term $\frac{1}{2}\sum_{i=0}^{n-1} f''(X_i)$ where the sum can also be interpreted as a (discrete) integral.
We observe that if (X_n) is a martingale with $X_0 = c$ some constant $c \in \mathbb{Z}$, then since the process $(f'(X_n))$ is adapted, due to Theorem 2.0.15 the process $(f'(X_\cdot) \bullet X)_n$ is a martingale, too (observe here that $f'(X_n)$ is bounded due to $X_0 = c$ and $|X_{n+1} - X_n| = 1$). In addition, the process $(\frac{1}{2}\sum_{i=0}^{n-1} f''(X_i))$ is previsible. Hence, as a consequence of Theorem 2.0.17 we deduce that the Doob decomposition of the process $(f(X_n))$ is given by

$$f(X_n) = \Big(\sum_{i=0}^{n-1} f'(X_i)(X_{i+1} - X_i) + f(X_0)\Big) + \frac{1}{2}\sum_{i=0}^{n-1} f''(X_i),$$

with the term in brackets on the right-hand side being a martingale and the last summand being a previsible process. In particular, if f is convex (in a discrete sense), then the second derivative of f is non-negative and the above yields that $f(X_n)$ is a submartingale due to Theorem 2.0.17 (which otherwise we had already concluded using Jensen's inequality in Theorem 2.0.11).

Chapter 3

Markov processes

See Section 1.17 of [Dre18] for a short introduction to Markov chains in discrete time. We will mostly focus on time continuous Markov chains in this section. Although strictly not required for understanding the results of this chapter, it is useful and guiding for intuition to have a look at Section 1.17 of [Dre18] to obtain some basic knowledge in discrete time Markov chains. For those of you who do know Markov chains in discrete time, continuous time Markov chains can essentially be thought of as discrete time Markov chains which are run through at random jump times (instead of the deterministic jump times of $\mathbb{N}$ in the case of discrete time Markov chains). It will turn out that in order to enjoy the crucial property that the transition probabilities of a Markov chain depend on the entire past solely through the state of the present, the times in between jumps will necessarily have to be distributed according to a sequence of i.i.d. exponentially distributed random variables.
We start with a generalisation of the Markov property for discrete time stochastic processes that had been treated Section 1.17 of [Dre18] as follows.

Definition 3.0.1. *Let (X_t), $t \in T$, be a stochastic process taking values in a Polish space $(S, \mathcal{O})$ (endowed with the Borel σ-algebra) and let $(\mathcal{F}_t)$, $t \in T$, denote the filtration generated by the process (X_t). Then (X_t) has the* Markov property *if for any $0 < s < t$, and $B \in \mathcal{B}(S)$, with $s, t \in T$,*

$$\mathbb{P}(X_t \in B \,|\, \mathcal{F}_s) = \mathbb{P}(X_t \in B \,|\, X_s). \tag{3.0.1}$$

The Reader is invited to check that in the case of countable S and $T = \mathbb{N}_0$, a stochastic process exhibits the Markov property with respect to its canonical filtration if and only if it is a Markov chain as introduced in [Dre18].
It can be shown that one of the properties equivalent to the Markov property is that 'the past and the future are conditionally independent given the present.'

Exercise 3.0.2. *(a) Random walk as introduced in Example 2.0.8 (a) has the Markov property (check!).*

(b) Show that if $f : S \to \mathbb{R}$ is a measurable function and the stochastic process (X_t) has the Markov property then $f(X_t)$ does not necessarily have the Markov property.

(c) Recall the setting of Section 1.17 of [Dre18], where S was a finite or countable set. A sequence of random variables (X_n) on $(\Omega, \mathcal{F}, \mathbb{P})$ was called a homogeneous Markov chain with state space S and transition matrix $P = (P(x,y))_{x,y \in S}$ *if for all $n \in \mathbb{N}_0$ and $x_0, x_1, \ldots, x_{n+1} \in S$, one has*

$$\begin{aligned} P(x_n, x_{n+1}) &= \mathbb{P}(X_{n+1} = x_{n+1} \,|\, X_n = x_n) \\ &= \mathbb{P}(X_{n+1} = x_{n+1} \,|\, X_0 = x_0,\, X_1 = x_1, \ldots, X_n = x_n), \end{aligned} \tag{3.0.2}$$

whenever $\mathbb{P}(X_0 = x_0,\, X_1 = x_1, \ldots, X_n = x_n) > 0$.

Show that (X_t) has the Markov property.

Exercise 3.0.3. *Let $T = \mathbb{N}_0$ and $S = \mathbb{R}$. In this context, is any martingale a Markov process, and vice versa? Find counterexamples or proofs!*

Remark 3.0.4. *For $T = \mathbb{N}_0$, every stochastic process can be turned into a Markov chain (here: a stochastic process exhibiting the Markov property) by considering the state space $\cup_{n\in\mathbb{N}} S^n$ instead, and by considering the process $(\widetilde{X}_n)$, $n \in \mathbb{N}_0$, defined via*

$$\widetilde{X}_n := (X_0, \ldots, X_n).$$

Intuitively, the above display (3.0.1) tells us that that knowing all the past of the process (X_t) up to time s does not provide us with any advantage over only knowing X_s. Also, equation (3.0.1) naturally leads to the concept of regular conditional probabilities. Indeed, the factorisation lemma implies that for each $B \in \mathcal{B}(S)$, there exists a $\mathcal{B}(S) - \mathcal{B}$-measurable function $\varphi_B : S \to \mathbb{R}$ such that

$$\mathbb{P}(X_t \in B \,|\, \mathcal{F}_s) = \varphi_B(X_s),$$

where, as always, this equality is to be understood in an $\mathbb{P}$-a.s. sense. It would be nice, however, to have that this identity is 'well-behaved' in B also, in the sense that the right-hand side, for X_s fixed, defines a probability measure in B also. This is the reason for us to insert the following section here.

3.1 Interlude: Regular conditional distributions

Recall from Chapter 1 that we defined the conditional probability of $B \in \mathcal{F}$ given a sub-σ-algebra $\mathcal{G}$ as $\mathbb{P}(B \,|\, \mathcal{G})(\omega) = \mathbb{E}[\mathbb{1}_B \,|\, \mathcal{G}](\omega)$ for almost all $\omega \in \Omega$. Therefore, for almost all ω, $\mathbb{P}(B \,|\, \mathcal{G})(\omega)$ satisfies countable additivity due to Theorem 1.0.9 (a) and (b). However, the set of $\omega \in \Omega$ with measure 0 where countable additivity fails might depend on the sequence of sets in question, and there might be uncountably many such sequences, so that their union covers all of Ω. To avoid this problem, we must choose the conditional probabilities carefully, which we do in this section.

Definition 3.1.1. *Let $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ be measurable spaces. A mapping*

$$\mu : \Omega_1 \times \mathcal{F}_2 \to [0, 1]$$

is called transition kernel *or* stochastic kernel *(from $(\Omega_1, \mathcal{F}_1)$ to $(\Omega_2, \mathcal{F}_2)$) if the following properties hold true:*

(a) For each $F \in \mathcal{F}_2$ the mapping

$$\Omega_1 \ni \omega \mapsto \mu(\omega, F)$$

is $\mathcal{F}_1 - \mathcal{B}$-measurable;

(b) For each $\omega \in \Omega_1$, the mapping

$$\mathcal{F}_2 \ni F \mapsto \mu(\omega, F)$$

defines a probability measure on $\mathcal{F}_2$.

Example 3.1.2. *If $(S, 2^S)$ is the countable state space (endowed with its power set) of a (time homogeneous) Markov chain with transition matrix P, then we can define the transition kernel*

$$p : S \times 2^S \ni (x, B) \mapsto \sum_{y\in B} P(x, y)$$

from $(S, 2^S)$ to itself. In this sense, transition kernels can be considered generalisations of transition matrices for Markov chains. As we will see in Theorem 3.1.19 below, however, transition *kernels can be used to construct more general stochastic processes than just Markov processes, by not only taking elements from S as an argument (corresponding to the state of a Markov chain at time $n-1$), but by instead considering the n-fold product space $S^{\{0,\dots,n-1\}}$ (corresponding to the entire history of a Markov chain up to time $n-1$). This will then be a kernel from $(S^{\{0,\dots,n-1\}}, 2^{S^{\{0,\dots,n-1\}}})$ to $(S, 2^S)$.*

Definition 3.1.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $(E, \mathcal{E})$ a measurable space, and $X : (\Omega, \mathcal{F}) \to (E, \mathcal{E})$ a random variable, as well as $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. A function $\mu : \Omega \times \mathcal{E} \to [0,1]$ is called a* regular conditional distribution for X given $\mathcal{G}$ *if*

(a) for each $F \in \mathcal{E}$, the function $\Omega \ni \omega \mapsto \mu(\omega, F)$ is a version[1] *of the conditional probability $\mathbb{P}(X \in F \,|\, \mathcal{G})$ (in particular, $\mu(\cdot, F)$ is a $\mathcal{G} - \mathcal{B}(\mathbb{R})$-measurable function);*

(b) for $\mathbb{P}$-almost all $\omega \in \Omega$,

$$\mathcal{E} \ni F \mapsto \mu(\omega, F)$$

is a probability measure on $\mathcal{E}$.

In other words, a regular conditional distribution for X given $\mathcal{G}$ is a stochastic kernel μ such that for $\mathbb{P}$-almost all $\omega \in \Omega$ and all $F \in \mathcal{E}$,

$$\mu(\omega, F) = \mathbb{P}(X \in F \,|\, \mathcal{G})(\omega) \tag{3.1.1}$$

In the case $(E, \mathcal{E}) = (\Omega, \mathcal{F})$ and X the identity map, μ is also called a regular conditional probability*.*

Indeed, note that for any fixed $F \in \mathcal{E}$, we can just define the stochastic kernel μ via (3.1.1). However, it is far from clear that there exists a set $\widetilde{\Omega} \in \mathcal{F}$ of probability 1 such that for all $\omega \in \widetilde{\Omega}$ *and* all $F \in \mathcal{G}$, display (3.1.1) holds true. Indeed, there are settings where a regular conditional probability distribution does not exist – see [Fad85] for conditions of on the existence of regular conditional probability distributions.

Theorem 3.1.4. *Let $(S, \mathcal{O})$ be a Polish space and let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (S, \mathcal{O})$ be a random variable. Furthermore, let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra.*
Then a conditional distribution μ of X given $\mathcal{G}$ exists. Furthermore, it is unique in the sense that if μ' is another conditional distribution of X given $\mathcal{G}$, then for $\mathbb{P}$-almost all ω, the measures $\mu(\omega, \cdot)$ and $\mu'(\omega, \cdot)$ coincide.

As a consequence, we call any such regular conditional distribution *the* regular conditional distribution.
In order to prove this result, we introduce the following notion: For $\mathcal{A} \subset \mathcal{C}$ two algebras of subsets of some Polish space $(S, \mathcal{O})$ and a finite, additive, non-negative function μ defined on $\mathcal{C}$, we say that *μ is regular on $\mathcal{A}$ for $\mathcal{C}$*, if for each $A \in \mathcal{A}$ we have

$$\mu(A) = \sup \big\{ \mu(C) \, : \, C \in \mathcal{C} \text{ with } C \subset A \text{ and } C \text{ compact} \big\}.$$

We will also take advantage of the following auxiliary result.

Lemma 3.1.5. *If in the above context μ is regular on $\mathcal{A}$ for $\mathcal{C}$, then μ is σ-additive on $\mathcal{A}$.*

Proof. Assume that μ is not σ-additive. Then by Proposition 1.2.15 of [Dre19] we infer that μ cannot be continuous in $\emptyset$ either, so there exists $\delta > 0$ and a sequence of sets (A_n) with $A_n \in \mathcal{A}$

[1] I.e, we have $\mathbb{P}(\{\omega \in \Omega \, : \, \mu(\omega, F) \neq \mathbb{P}(X \in F \,|\, \mathcal{G})(\omega)\}) = 0$. In particular, recall that the random variable $\mathbb{P}(X \in F \,|\, \mathcal{G})$ is only uniquely defined as an element in L^1, not in $\mathcal{L}^1$. So the previous can be formulated as $\mu(\cdot, F) = \mathbb{P}(X \in F \,|\, \mathcal{G})$ in L^1.

and $A_n \downarrow \emptyset$, as well as $\mu(A_n) > \delta$ for all $n \in \mathbb{N}$. Again, since μ is regular on $\mathcal{A}$ for $\mathcal{C}$, we have that for each $n \in \mathbb{N}$ there is $K_n \subset A_n$ compact with $K_n \in \mathcal{C}$ and $\mu(A_n \backslash K_n) \leq \delta 2^{-n-1}$. This implies

$$\mu\Big(\bigcap_{i=1}^{n} K_i\Big) \geq \mu(A_n) - \sum_{i=1}^{n} \delta 2^{-i-1} \geq \delta/2.$$

In particular, all the $\cap_{i=1}^n K_i$ are non-empty, and

$$\bigcap_{i=1}^{\infty} K_i \neq \emptyset \tag{3.1.2}$$

as well. Indeed, if it wasn't, then the $(\cap_{i=1}^n K_i)^c$, $n \in \mathbb{N}$, would form an open covering of K_1, so since K_1 is compact, there would exist a finite subcover, which would imply that one of the $\cap_{i=1}^n K_i$, $n \in \mathbb{N}$ would have to be disjoint from K_1, a contradiction, so (3.1.2) holds true. This, however, is a contradiction to $A_n \downarrow \emptyset$. Hence, μ must be σ-additive. □

Proof of Theorem 3.1.4. We start with observing that since $(S, \mathcal{O})$ is Polish, $\mathcal{B}(S)$ is countably generated, e.g. by the set $\mathcal{U} := \{B(x, \frac{1}{n}) : x \in \widetilde{S}, n \in \mathbb{N}\}$, where $\widetilde{S} \subset S$ is any countable dense subset of S and $B(x, \frac{1}{n})$ is the open ball around x of radius $\frac{1}{n}$ with respect to any (but a fixed one) of the metrics which induce the topology $\mathcal{O}$. Therefore, the algebra $\mathcal{V}$ generated by $\mathcal{U}$ can be written as a countable union of non-decreasing, finite algebras $\mathcal{V}_n$. By the regularity Lemma 4.2.3 from [Dre19] (Ulam's theorem), for each $B \in \mathcal{V}$ there exists a non-decreasing sequence (B_j^B) with $B_j^B \subset B$, B_j^B compact, such that

$$\mathbb{P}_X(B_j^B) \uparrow \mathbb{P}_X(B) \quad \text{ as } j \to \infty.$$

Since finite unions of compact sets remain compact, we can choose a specific sequence (B_j^B) for B such that $B_1^B \subset B_2^B \subset \cdots$ and hence by the conditional monotone convergence theorem (Theorem 1.0.9 (d)) we deduce that

$$\mathbb{P}_X(B_j^B \,|\, \mathcal{G}) \uparrow \mathbb{P}_X(B \,|\, \mathcal{G}) \quad \mathbb{P} - a.s. \tag{3.1.3}$$

But now the union of $\mathcal{V}$ with all the possible B_j^B, $B \in \mathcal{V}$, $j \in \mathbb{N}$, is countable again, and we denote the countable algebra that is generated by this union by $\mathcal{A}$. For a given choice of $\mathbb{P}_X(A) \,|\, \mathcal{G})(\omega)$ for all $A \in \mathcal{A}$ and all $\omega \in \Omega$, which is $\mathcal{G}$-$\mathcal{B}(S)$-measurable we have

(a) For each $A \in \mathcal{A}$,
$$\mathbb{P}_X(A \,|\, \mathcal{G})(\omega) \geq 0 \text{ almost surely.}$$

(b)
$$\mathbb{P}_X(S \,|\, \mathcal{G})(\omega) = 1 \text{ a.s.,} \quad \text{ and } \quad \mathbb{P}_X(\emptyset \,|\, \mathcal{G})(\omega) = 0 \text{ a.s.}$$

(c) For any finite sequence of pairwise disjoint sets $A_1, \ldots, A_n$ with $A_j \in \mathcal{A}$ for all $j \in \{1, \ldots, n\}$, and almost all ω we have
$$\mathbb{P}_X\Big(\bigcup_{j=1}^{n} A_j \,|\, \mathcal{G}\Big)(\omega) = \sum_{j=1}^{n} \mathbb{P}_X(A_j \,|\, \mathcal{G})(\omega);$$

(d) For almost all ω and all sequences (B_j^B), $j \in \mathbb{N}$, $B \in \mathcal{V}$, the convergence in (3.1.3) holds true.

Items (a) to (d) consist of countably many equations or inequalities in terms of $\mathcal{G}$-$\mathcal{B}(S)$-measurable functions, including limits of sequences of $\mathcal{G}$-$\mathcal{B}(S)$-measurable functions, each holding almost surely. Therefore, there exists some $\widetilde{\Omega} \in \mathcal{G}$ for which $\mathbb{P}(\widetilde{\Omega}) = 0$ and for which we can replace the a.s. statements in (a) to (d) with "for all $\omega \in \widetilde{\Omega}$".
Now by definition of $\widetilde{\Omega}$, for $\omega \in \widetilde{\Omega}$, the function $\mathbb{P}_X(\cdot \,|\, \mathcal{G})(\omega)$ is regular on $\mathcal{V}$ for $\mathcal{A}$, and hence is σ-additive on the algebra $\mathcal{V}$ by Lemma 3.1.5. By Corollary 1.3.9 from [Dre19], we can extend this function uniquely to a probability measure $\mu(\omega, \cdot)$ on $\mathcal{B}(S)$.
From this we construct the regular conditional distribution as follows. Denote by $\mathcal{D}$ the set of all $B \in \mathcal{B}(S)$ such that

$$\omega \mapsto \mu(\omega, B)$$

defines a version of the conditional probability $\mathbb{P}_X(B \,|\, \mathcal{G})$. Then due to the monotone convergence theorem for conditional expectations, $\mathcal{D}$ is a Dynkin system, and furthermore it contains $\mathcal{V}$. Hence, due to Dynkin's $\pi - \lambda$-Theorem (see Theorem 1.1.30 in [Dre19]), $\mathcal{D} = \mathcal{O}$, and therefore μ as above, extended to being equal to 0 on $\Omega \backslash \widetilde{\Omega}$, defines a regular conditional distribution.
As regards to the uniqueness property, observe that if μ' is another regular conditional distribution of X given $\mathcal{G}$, then $\mathbb{P}$-a.s. the two functions coincide on the countable algebra $\mathcal{A}$. But since $\mathcal{A}$ is a π-system generating $\mathcal{B}(S)$, Dynkin's $\pi - \lambda$-theorem again yields that $\mathbb{P}$-a.s.,

$$\mu(\cdot, B) = \mu'(\cdot, B) \quad \forall B \in \mathcal{B}(S),$$

which proves the desired uniqueness. □

Remark 3.1.6. *In fact, the previous result not only holds for Polish spaces $(S, \mathcal{O})$, but we can even replace $(S, \mathcal{O})$ by a standard Borel space. A measurable space $(E, \mathcal{E})$ is called a* (standard) Borel space *if there exists a bijection $\varphi : E \to B$, some $B \in \mathcal{B}(\mathbb{R})$, such that φ and φ^{-1} are both measurable. See e.g. Theorem 4.1.6 in [Dur10] or Theorem 8.37 in [Kle14]. Indeed, we can also deduce this without much effort from our general result for Polish spaces above, since in particular $\mathbb{R}$ is a Polish space.*

Example 3.1.7. *(a) Let X, Y be real random variables such that the random vector (X, Y) has probability density f with respect to the two-dimensional Lebesgue measure λ^2 on $\mathcal{B}^2$.*

Then the marginal density of X *(with respect to the one-dimensional Lebesgue measure λ) is given by*

$$f_X(x) := \int_{\mathbb{R}} f(x, y)\, \mathrm{d}y,$$

and for each $x \in \mathbb{R}$ with $f_X(x) \in (0, \infty)$ and $y \in \mathbb{R}$ we define the conditional density of Y given $X = x$ *via*

$$f_{Y \,|\, X}(y \,|\, x) := \frac{f(x, y)}{f_X(x)}.$$

Defining $\mu(x, \cdot)$ as the probability measure on $\mathcal{B}$ with density $f_{Y \,|\, X}(\cdot \,|\, x)$ if $f_X(x) \in (0, \infty)$, and as an arbitrary but fixed probability measure on $\mathcal{B}$ otherwise, we obtain a regular conditional distribution of Y given X.

(b) We can now also make precise our motivating Example 1.0.2 from the very beginning. There we considered the situation of two independent Gaussian random variables X and Y distributed according to $\mathcal{N}(0, 1)$, and setting $Z := X + Y$ we wanted to make sense of

$$\mathbb{P}(Z \in \cdot \,|\{X = x\}). \tag{3.1.4}$$

Take $\Omega := \mathbb{R}^2$, $\mathcal{F} := \mathcal{B}^2$, as well as $X(x, y) := x$, $Y(x, y) := y$, and

$$\mathbb{P} = \frac{1}{2\pi} e^{\frac{-x^2 - y^2}{2}} \cdot \lambda^2.$$

In this context, we interpret (3.1.4) as the regular conditional distribution μ of Z given X evaluated at (x, y), $y \in \mathbb{R}$ arbitrary (i.e., $\mu((x, y), \cdot)$), and we claim that it is given by

$$\mu((x, y), B) := \frac{1}{\sqrt{2\pi}} \int_B e^{\frac{-(z-x)^2}{2}} \, \mathrm{d}z, \quad B \in \mathcal{B}, \, (x, y) \in \mathbb{R}^2.$$

Indeed, for each $(x, y) \in \mathbb{R}^2$ we have that $\mu((x, y), \cdot)$ defines a probability measure on $\mathcal{B}$, so Part (b) of Definition 3.1.3 holds true.
On the other hand, for each $B \in \mathcal{B}^2$ we have that $(x, y) \mapsto \mu((x, y), B)$ is $\sigma(X) - \mathcal{B}$-measurable (recall that $\sigma(X) = \mathcal{B} \otimes \{\mathbb{R}, \emptyset\}$, and by Tonelli's theorem, 'integrating out one of the variables of a measurable function keeps measurability of the resulting integral in the other variables'), and furthermore, for $A \in \sigma(X)$ we have $A = A_1 \times \mathbb{R}$, some $A_1 \in \mathcal{B}$, so

$$\begin{aligned} \int_A \mu((x, y), B) \, \mathbb{P}(\mathrm{d}x, \mathrm{d}y) &= \frac{1}{2\pi} \int_A \frac{1}{\sqrt{2\pi}} \Big(\int_B e^{\frac{-(z-x)^2}{2}} \, \mathrm{d}z \Big) \, e^{\frac{-x^2-y^2}{2}} \, \lambda^2(\mathrm{d}x, \mathrm{d}y) \\ &= \frac{1}{2\pi} \int_{A_1 \times B} e^{\frac{-(z-x)^2 - x^2}{2}} \, \lambda^2(\mathrm{d}x, \mathrm{d}z). \end{aligned} \tag{3.1.5}$$

On the other hand,

$$\begin{aligned} \mathbb{P}[\{Z \in B\} \cap A] &= \frac{1}{2\pi} \int \mathbb{1}_{x+y \in B} \mathbb{1}_{x \in A_1} e^{\frac{-x^2-y^2}{2}} \, \lambda(\mathrm{d}y)\lambda(\mathrm{d}x) \\ &\overset{y \mapsto y-x}{=} \frac{1}{2\pi} \int_{A_1 \times B} e^{\frac{-x^2-(y-x)^2}{2}} \, \lambda^2(\mathrm{d}x, \mathrm{d}y), \end{aligned}$$

so it equals (3.1.5), and hence Part (a) of Definition 3.1.3 holds true as well.

Theorem 3.1.8 (Disintegration formula)**.** *Let $(E, \mathcal{E})$ be a measurable space, and let $X : (\Omega, \mathcal{F}, \mathbb{P}) \to (E, \mathcal{E})$ be a random variable. Furthermore, let $f : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B})$ be measurable with $f \circ X \in \mathcal{L}^1$, and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Then f is integrable with respect to $\mu(\omega, \cdot)$ for $\mathbb{P}$-a.a. $\omega \in \Omega$, and*

$$\mathbb{E}[f \circ X \,|\, \mathcal{G}](\omega) = \int_E f(x) \, \mu(\omega, \mathrm{d}x) \quad \mathbb{P} - \textit{almost all } \omega \in \Omega, \tag{3.1.6}$$

where μ denotes the regular conditional distribution of X given $\mathcal{G}$.

Proof. Due to the linearity of (3.1.6) in f, and since $f = f^+ - f^-$, we can assume w.l.o.g. that $f \geq 0$.
If $f = \mathbb{1}_B$ some $B \in \mathcal{E}$, then by the definitions, for $\mathbb{P}$-almost all $\omega \in \Omega$ we have

$$\mathbb{E}[f \circ X \,|\, \mathcal{G}](\omega) = \mathbb{P}(\{X \in B\} \,|\, \mathcal{G})(\omega)$$

and

$$\int f(x) \, \mu(\omega, \mathrm{d}x) = \mu(\omega, B) = \mathbb{P}(\{X \in B\} \,|\, \mathcal{G})(\omega),$$

so the claim holds true, and due to linearity it also extends to simple functions. Choose a sequence (f_n) of simple functions with $f_n \geq 0$, and $f_n \uparrow f$. Hence, the monotone convergence theorem for conditional expectations (Theorem 1.0.9(b)) implies that

$$\mathbb{E}[f_n \circ X \,|\, \mathcal{G}] \uparrow \mathbb{E}[f \circ X \,|\, \mathcal{G}]$$

$\mathbb{P}$-almost surely. On the other hand, for any ω fixed we have that

$$\int f_n(x) \, \mu(\omega, \mathrm{d}x) \uparrow \int f(x) \, \mu(\omega, \mathrm{d}x)$$

due to the standard monotone convergence theorem, and the result follows. $\square$

An application of the previous result is given by the following exercise.

Exercise 3.1.9. *Regular conditional probability distributions can be used to derive a conditional Hölder inequality: For X, Y real-valued random variables and $\mathcal{G} \subset \mathcal{F}$, and $p, q \in (1, \infty)$ with $p^{-1} + q^{-1} = 1$, we have*

$$\mathbb{E}[|XY| \,|\, \mathcal{G}] \le \mathbb{E}[|X|^p \,|\, \mathcal{G}]^{\frac{1}{p}} \mathbb{E}[|Y|^q \,|\, \mathcal{G}]^{\frac{1}{q}}.$$

3.1.1 Properties of transition kernels

While we have seen that Kolmogorov's existence and uniqueness result is quite helpful in constructing stochastic processes once we have a consistent family of probability measures, it turns out that oftentimes (and in particular in the setting of Markov processes) a more hands-on approach is more useful: Instead of starting with a consistent family of probability measures, it might be easier to start from a sequence of transition kernels.
We start with the following basic lemma.

Lemma 3.1.10. *Let μ be a transition kernel from $(E_1, \mathcal{E}_1)$ to $(E_2, \mathcal{E}_2)$, and let $f : (E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2) \to ([0, \infty], \mathcal{B}([0, \infty]))$ be a measurable function. Then the mapping*

$$\varphi_f : E_1 \ni x_1 \mapsto \int f(x_1, x_2)\, \mu(x_1, \mathrm{d}x_2) \in [0, \infty]$$

is well-defined and $\mathcal{E}_1 - \mathcal{B}([0, \infty])$-measurable.

Proof. Obviously the mapping is well-defined (since $x_2 \mapsto f(x_1, x_2)$ is measurable for any $x_1 \in E_1$ due to a pre-result to Fubini's / Tonelli's theorem).
As almost always, the first step is to show the statement for $f = \mathbb{1}_F$ with $F \in \mathcal{E}_1 \otimes \mathcal{E}_2$. For this purpose, consider $F = F_1 \times F_2$ with $F_i \in \mathcal{E}_i$, $i \in \{1, 2\}$. Then

$$\int \mathbb{1}_{F_1 \times F_2}(x_1, x_2)\, \mu(x_1, \mathrm{d}x_2) = \mathbb{1}_{F_1}(x_1)\mu(x_1, F_2),$$

which evidently is a $\mathcal{E}_1 - \mathcal{B}([0, \infty])$-measurable function. We now claim that the set of $F \in \mathcal{E}_1 \otimes \mathcal{E}_2$ for which $\varphi_{\mathbb{1}_F}$ is $\mathcal{E}_1 - \mathcal{B}([0, \infty])$-measurable forms a Dynkin system $\mathcal{D}$. To start off with, we certainly have $E_1 \times E_2 \in \mathcal{D}$ since $\varphi_{\mathbb{1}_\Omega}(x_1) = \mu(x_1, E_2) \in [0, 1]$ which is measurable by assumption. Furthermore, for $D \in \mathcal{D}$ we have

$$\varphi_{\mathbb{1}_{D^c}}(x_1) = \varphi_{\mathbb{1}_\Omega}(x_1) - \varphi_{\mathbb{1}_D}(x_1),$$

which in combination with the previous entails $D^c \in \mathcal{D}$. It remains to show the stability of $\mathcal{D}$ under countable disjoint unions. For this purpose let (D_n), $n \in \mathbb{N}$ be a sequence with $D_n \in \mathcal{D}$ for all $n \in \mathbb{N}$. We then infer

$$\varphi_{\mathbb{1}_{\bigcup_{n \in \mathbb{N}} D_n}} = \int \mathbb{1}_{\bigcup_{n \in \mathbb{N}} D_n}(x_1, x_2)\, \mu(x_1, \mathrm{d}x_2) = \sum_{n \in \mathbb{N}} \int \mathbb{1}_{D_n}(x_1, x_2)\, \mu(x_1, \mathrm{d}x_2),$$

where the last equality follows from the monotone convergence theorem. Since each summand on the right-hand side is $\mathcal{E}_1 - \mathcal{B}([0, \infty])$-measurable, we obtain that the sum is $\mathcal{E}_1 - \mathcal{B}([0, \infty])$-measurable again as a limit of $\mathcal{E}_1 - \mathcal{B}([0, \infty])$-measurable functions.
All in all, the above shows that $\mathcal{D}$ is a Dynkin system containing the π-system of rectangles in $\mathcal{E}_1 \otimes \mathcal{E}_2$, which generates $\mathcal{E}_1 \otimes \mathcal{E}_2$. Hence, Dynkin's $\pi - \lambda$-Theorem implies $\mathcal{D} = \mathcal{E}_1 \otimes \mathcal{E}_2$.
This shows that for any $F \in \mathcal{E}_1 \otimes \mathcal{E}_2$., the function $\varphi_{\mathbb{1}_F}$ defines a $\mathcal{E}_1 - \mathcal{B}([0, \infty])$-measurable function, and by linearity this extends to simple functions. Now any arbitrary measurable function $f : (E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2) \to ([0, \infty], \mathcal{B}([0, \infty]))$ can be written as the monotone pointwise limit $f_n \uparrow f$ of simple functions f_n. Then, again by monotone convergence,

$$\varphi_f(x_1) = \lim_{n \to \infty} \varphi_{f_n}(x_1);$$

Each of the functions on the right-hand side is measurable in x_1 and hence so is the limit on the left-hand side which finishes the proof. $\square$

Theorem 3.1.11.
Let $(E_i, \mathcal{E}_i)$, $i \in \{0,1,2\}$, be measurable spaces and let μ_1 be a transition kernel from $(E_0, \mathcal{E}_0)$ to $(E_1, \mathcal{E}_1)$, and μ_2 be a transition kernel from $(E_0 \times E_1, \mathcal{E}_0 \otimes \mathcal{E}_1)$ to $(E_2, \mathcal{E}_2)$. We define the mapping

$$\mu_1 \otimes \mu_2 : E_0 \times (\mathcal{E}_1 \otimes \mathcal{E}_2) \to [0,1]$$

via

$$(x_0, F) \mapsto \int_{E_1} \Big(\int_{E_2} \mathbb{1}_F(x_1, x_2)\, \mu_2((x_0, x_1), \mathrm{d}x_2) \Big) \mu_1(x_0, \mathrm{d}x_1). \tag{3.1.7}$$

Then $\mu_1 \otimes \mu_2$ is well-defined and a transition kernel from $(E_0, \mathcal{E}_0)$ to $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$. It is denoted by $\mu_1 \otimes \mu_2$ and also referred to as the product of μ_1 and μ_2*.*

We get the following immediate corollary. It is tailor-made for the setting where we have an initial distribution of the stochastic process to construct.

Corollary 3.1.12. *Let $(E_i, \mathcal{E}_i)$, $i \in \{1,2\}$, be measurable spaces and let $\mathbb{P}_1$ be a probability measure on $(E_1, \mathcal{E}_1)$, and let μ_2 be a transition kernel from $(E_1, \mathcal{E}_1)$ to $(E_2, \mathcal{E}_2)$. We define the mapping*

$$\mathbb{P}_1 \otimes \mu_2 : \mathcal{E}_1 \otimes \mathcal{E}_2 \to [0,1]$$

via

$$F \mapsto \int_{E_1} \Big(\int_{E_2} \mathbb{1}_F(x_1, x_2)\, \mu_2(x_1, \mathrm{d}x_2) \Big) \mathbb{P}_1(\mathrm{d}x_1). \tag{3.1.8}$$

Then $\mathbb{P}_1 \otimes \mu_2$ is well-defined and a probability measure on $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$. It is denoted by $\mathbb{P}_1 \otimes \mu_2$.

Proof. This follows directly by interpreting $\mathbb{P}_1$ as a transition kernel from $(E_0, \mathcal{E}_0)$ to $(E_1, \mathcal{E}_1)$ which is constant in E_0. $\square$

Also, inductively, Theorem 3.1.11 in combination with the proof of the previous corollary supplies us with the following result.

Corollary 3.1.13. *Let $(E_i, \mathcal{E}_i)$, $i \in \{0,1,\dots,n\}$, be measurable spaces and for $i \in \{1,\dots,n\}$ let μ_i be a transition kernel from $(\times_{j=0}^{i-1} E_j, \otimes_{j=0}^{i-1} \mathcal{E}_j)$ to $(E_i, \mathcal{E}_i)$.*
We can define the mapping

$$\bigotimes_{i=1}^{n} \mu_i : E_0 \times (\otimes_{i=1}^{n} \mathcal{E}_i) \to [0,1]$$

as

$$\Big(\bigotimes_{i=1}^{n-1} \mu_i \Big) \otimes \mu_n. \tag{3.1.9}$$

Then $\otimes_{i=1}^{n} \mu_i$ is well-defined and a transition kernel from $(E_0, \mathcal{E}_0)$ to $(\times_{i=1}^{n} E_i, \otimes_{i=1}^{n} \mathcal{E}_i)$.
Also, as above, we can replace the kernel μ_1 by a probability measure $\mathbb{P}_1$ on $(E_1, \mathcal{E}_1)$ to obtain that $\mathbb{P}_1 \otimes \bigotimes_{i=2}^{n} \mu_i$ is a probability measure on $(\times_{i=1}^{n} E_i, \otimes_{i=1}^{n} \mathcal{E}_i)$.

Remark 3.1.14. *In particular in the context of transition kernels it turns out that it is sometimes more convenient and easier to read if we write the integrating kernel directly after the integral sign (which likely originates in physics). This usually provides a better overview of which integration is performed at which point. For example, the iterated integral on the right-hand side of (3.1.7) would be written as*

$$\int_{E_1} \mu_1(x_0, \mathrm{d}x_1) \Big(\int_{E_2} \mu_2((x_0, x_1), \mathrm{d}x_2)\, \mathbb{1}_F(x_1, x_2) \Big).$$

Proof of Theorem 3.1.11. We start with observing that

$$(E_0 \times E_1, \mathcal{E}_0 \otimes \mathcal{E}_1) \ni (x_0, x_1) \mapsto \int_{E_2} \mathbb{1}_F(x_1, x_2)\, \mu_2((x_0, x_1), \mathrm{d}x_2) \in ([0, \infty], \mathcal{B}([0, \infty]))$$

defines a measurable mapping due to Lemma 3.1.10, and applying that lemma once again we infer that $(\mu_1 \otimes \mu_2)(x_0, F)$ is well-defined and $\mathcal{E}_1 - \mathcal{B}$-measurable as a function in x_0; this shows Part (a) of Definition 3.1.1. Furthermore, by monotone convergence we deduce that for any $x_0 \in E_0$ we have that

$$\mathcal{E}_1 \otimes \mathcal{E}_2 \ni F \mapsto \int_{E_1} \Big(\int_{E_2} \mathbb{1}_F(x_1, x_2)\, \mu_2((x_0, x_1), \mathrm{d}x_2) \Big) \mu_1(x_0, \mathrm{d}x_1)$$

is σ-additive. Since it is furthermore non-negative and evaluates to 1 for $E_1 \times E_2$, we infer that Part (b) of Definition 3.1.1 if fulfilled. All in all, we conclude that $\mu_1 \otimes \mu_2$ defines a transition kernel. □

The preceding theorem can be generalised to n-fold products of correspondingly defined kernels in an obvious way.

Remark 3.1.15. *Note that in the special case of μ_1 in the above theorem being constant in x_0 (so μ_1 is just a probability measure on $\mathcal{E}_1$), we infer that $\mu_1 \otimes \mu_2$ is a probability measure on $\mathcal{E}_1 \otimes \mathcal{E}_2$; in particular, we can therefore define the product of a probability measure and a transition kernel, which yields a probability measure again.*

Furthermore, oftentimes the composition of kernels is of interest as well.

Definition 3.1.16. *For measurable spaces $(E_i, \mathcal{E}_i)$, $i \in \{0, 1, 2\}$, and transition kernels μ_i from $(E_{i-1}, \mathcal{E}_{i-1})$ to $(E_i, \mathcal{E}_i)$, $i \in \{1, 2\}$, we define the* composition ('Verkettung') $\mu_1\mu_2$ of μ_1 and μ_2 *via*

$$\mu_1\mu_2 : E_0 \times \mathcal{E}_2 \ni (x_0, F_2) \mapsto \int_{E_1} \mu_1(x_0, \mathrm{d}x_1)\, \mu_2(x_1, F_2).$$

The relation between the composition and the product of kernels is given by the following.

Theorem 3.1.17. *For measurable spaces $(E_i, \mathcal{E}_i)$, $i \in \{0, 1, 2\}$, and transition kernels μ_i from $(E_{i-1}, \mathcal{E}_{i-1})$ to $(E_i, \mathcal{E}_i)$, $i \in \{1, 2\}$, we have*

$$\mu_1\mu_2(x_0, F_2) = (\mu_1 \otimes \mu_2)(x_0, \pi_2^{-1}(F_2)) \quad \forall x_0 \in E_0, \quad F_2 \in \mathcal{E}_2, \tag{3.1.10}$$

where $\pi_2 : E_1 \times E_2 \to E_2$ denotes the projection $\pi_2 : (x_1, x_2) \mapsto x_2$, and $\mu_1\mu_2$ defines a transition kernel from $(E_0, \mathcal{E}_0)$ to $(E_2, \mathcal{E}_2)$. I.e., for each $x_0 \in E_0$, we have that

$$\mu_1\mu_2(x_0, \cdot) = (\mu_1 \otimes \mu_2)(x_0, \cdot) \circ \pi_2^{-1}, \quad \forall x_0 \in E_0, \tag{3.1.11}$$

so $\mu_1\mu_2(x_0, \cdot)$ is the image measure of $\mu_1 \otimes \mu_2$ under π_2.

Remark 3.1.18. *Here, as is commonly done, we interpret the transition kernel μ_2 as a transition kernel from $(E_0 \times E_1, \mathcal{E}_0 \otimes \mathcal{E}_1)$ to $(E_2, \mathcal{E}_2)$ in order to construct $\mu_1 \otimes \mu_2$.*

Proof. For all $x_0 \in E_0$ and $F_2 \in \mathcal{E}_2$, we have

$$\begin{aligned}(\mu_1 \otimes \mu_2)(x_0, \pi_2^{-1}(F_2)) &= \int_{E_1} \mu_1(x_0, \mathrm{d}x_1) \int_{E_2} \mathbb{1}_{E_1 \times F_2}(x_1, x_2)\, \mu_2(x_1, \mathrm{d}x_2) \\ &= \int_{E_1} \mu_1(x_0, \mathrm{d}x_1)\mu_2(x_1, F_2) = \mu_1\mu_2(x_0, F_2),\end{aligned}$$

which establishes (3.1.10).
Since $\pi_2^{-1}(F_2) \in \mathcal{E}_1 \otimes \mathcal{E}_2$, Theorem 3.1.11 provides us with the desired measurability of $\mu_1\mu_2(\cdot, F_2)$ for $F_2 \in \mathcal{E}_2$. Furthermore, from (3.1.11) we infer that for each $x_0 \in E_0$, we have that $\mu_1\mu_2(x_0, \cdot)$ is a probability measure indeed. □

Arguably the most intuitive and most accessible version of Ionescu-Tulcea's theorem is formulated in the product space setting. For this purpose, assume the following: Let $(E_n, \mathcal{E}_n)$, $n \in \mathbb{N}_0$ be sequence of measurable spaces and let $\mathbb{P}_0$ be a probability measure on $(E_0, \mathcal{E}_0)$. For $n \in \mathbb{N}$, set $\Omega_n := \times_{i=0}^n E_i$ and $\Omega := \times_{i=0}^\infty E_i$, and define the σ-algebras $\mathcal{F}_n := \otimes_{i=0}^n \mathcal{E}_i$ as well as $\mathcal{F} := \otimes_{i=0}^\infty \mathcal{E}_i$ on these spaces. Furthermore, let (μ_n), $n \in \mathbb{N}$, be a sequence of transition kernels such that μ_n is a transition kernel from $(\Omega_{n-1}, \mathcal{F}_{n-1})$ to $(E_n, \mathcal{E}_n)$. Using Corollary 3.1.13 we observe that

$$\mathbb{P}_n := \mathbb{P}_0 \otimes \bigotimes_{i=1}^n \mu_i$$

defines a probability measure on $\mathcal{F}_n$.
Furthermore, it is not hard to see inductively that the sequence $(\mathbb{P}_n)$, $n \in \mathbb{N}$, fulfills the following consistency condition: For $m, n \in \mathbb{N}$ with $m < n$ we have

$$\mathbb{P}_m(F_m) = \mathbb{P}_n\Big(F_m \times \bigtimes_{i=m+1}^n E_i\Big) \quad \forall F_m \in \mathcal{F}_m. \tag{3.1.12}$$

Therefore, in a way similar to the context of Kolmogorov's existence and uniqueness theorem in [Dre19] (where we also had an underlying consistency condition), we can and do now enquire about the existence (and uniqueness) of a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ such that

$$\mathbb{P}_m(F) = \mathbb{P}\Big(F \times \bigtimes_{k=m+1}^\infty E_i\Big) \quad \forall m \in \mathbb{N},\, F \in \mathcal{F}_m. \tag{3.1.13}$$

Theorem 3.1.19 (Ionescu-Tulcea)**.** *In the setting described above, there exists a unique probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$ such that (3.1.13) holds true.*

Proof. If the $(E_i, \mathcal{E}_i)$ were Polish spaces, taking advantage of the consistency condition (3.1.12) we could deduce that the assumptions of Kolmogorov's existence and uniqueness theorem in order to deduce the existence of the probability measures that we are looking for here. Indeed, to check the assumptions of Theorem 4.2.1 in [Dre19] in this setting, set $\Lambda := \mathbb{N}_0$, and E_λ as in the assumptions with their Borel σ-algebra $\mathcal{B}_\lambda$. For $J \subset \Lambda$ finite, we can write $J = \{\lambda_1, \dots, \lambda_n\}$ with $\lambda_1 < \dots < \lambda_n$ and define

$$\mathbb{P}_J(B_J) := \mathbb{P}_{\lambda_n}\Big(x \in \bigtimes_{i=0}^{\lambda_n} E_i \,:\, (x_{\lambda_1}, \dots, x_{\lambda_n}) \in B_J\Big), \quad \forall B_J \in \bigotimes_{\lambda \in J} \mathcal{B}_\lambda,$$

i.e.,

$$\mathbb{P}_J = \mathbb{P}_{\lambda_n} \circ \pi^{-1}_{\lambda_1, \dots, \lambda_n},$$

where $\pi_{\lambda_1, \dots, \lambda_n}$ is the coordinate projection from $\times_{i=0}^{\lambda_n} E_i$ to $\times_{\lambda \in J} E_\lambda$, and so $\mathbb{P}_J$ defines a probability measure on $(\times_{\lambda \in J} E_\lambda, \otimes_{\lambda \in J} \mathcal{B}_\lambda)$. Then one can check that the family $\mathbb{P}_J$, $J \subset \Lambda$, is consistent and we can apply Theorem 4.2.1 of [Dre19] in order to deduce the existence of a probability measure with the desired properties. For the general case of measurable spaces $(E_i, \mathcal{E}_i)$ we refer to e.g. [Kle14] for a proof. □

3.2 Back to Markov processes again

In the previous section the product of transition kernels had been defined in a pretty general way in Theorem 3.1.11. We will mostly be interested in the simpler case where $(E_i, \mathcal{E}_i)$ does not depend on i, (so we can write $(E, \mathcal{E})$ instead) and the transition kernel μ_n depends on $\times_{i=0}^{n-1} E$ only through its last coordinate – this will be crucial for obtaining Markov processes, which we will do for possibly uncountable index sets T at once.
In particular, having the above basic theory of transition kernels and induced distributions at our disposal, we can introduce the following definition.

Definition 3.2.1. *As before, let $T \subset \mathbb{R}$ be a semigroup and assume a measurable space $(E, \mathcal{E})$ to be given, as well as a family (μ_t), $t \in T$, of transition kernels from $(E, \mathcal{E})$ to $(E, \mathcal{E})$.*[2] *Then the family (μ_t), $t \in T$, is called a* Markov semigroup (of transition kernels) *if the equation*

$$\mu_s \mu_t = \mu_{s+t} \tag{3.2.1}$$

holds true for all $s, t \in T$.
Equation (3.2.1) is also called the Chapman-Kolmogorov *equation (cf. Proposition 1.17.32 in [Dre18]).*

The name *semigroup* is motivated by the fact that the kernels (μ_n), $n \in T$, form a semigroup in the algebraic sense with respect to composition.
In the case $T = \mathbb{N}$, and countable state space $(E, 2^E)$ the Markov transition kernels μ_n give exactly the probability that a Markov process (or Markov chain in this setting) jumps from state $x \in E$ to state $y \in E$ in n time steps: $\mu_n(x, \{y\})$, and as is further detailed in [Dre18], the law of the entire Markov chain is then uniquely characterised (modulo the initial distribution) by the *transition matrix* $P(x, y)$, $x, y \in E$, which give the probability that the Markov chain jumps from x to y in one time step, so

$$P(x, y) = \mu_1(x, \{y\}). \tag{3.2.2}$$

Display (3.2.1) then corresponds to the fact that for $m, n \in N$ we have $P^m P^n = P^{m+n}$.

Example 3.2.2. *In the case $(E, \mathcal{E}) = (\mathbb{R}^d, \mathcal{B}^{\otimes d})$, a class of Markov semigroups is given by those which have a density with respect to Lebesgue measure.*

(a) *A case of a particularly important semigroup is the* Brownian semigroup *of transition kernels, which is named after botanist Robert Brown and will give rise to Brownian motion in a later chapter. We choose $T = [0, \infty)$ and define the transition kernels*

$$\mu_t(x, \cdot) := \frac{1}{(2\pi t)^{\frac{d}{2}}} \exp\Big\{ - \frac{(y - x) \cdot (y - x)}{2t} \Big\} \cdot \lambda^d(\mathrm{d}y), \quad t > 0,$$

and $\mu_0(x, \cdot) = \delta_x$. Hence, we have that $\mu_t(x, \cdot)$ is the law of a $\mathcal{N}(x, t\mathrm{Id})$-distributed random variable, with the common convention that for $t = 0$ this equals δ_x.

We first have to check that this defines a semigroup of transition kernels and for this purpose, observe that for $s, t > 0$ (the case $s = 0$ can be checked separately) as well as $x_0 \in \mathbb{R}^d$ and $B \in \mathcal{B}^{\otimes d}$ we have

$$\begin{aligned}(\mu_s \mu_t)(x_0, B) &= \int_{\mathbb{R}^d} \mu_s(x_0, \mathrm{d}x_1) \mu_t(x_1, B) = \int_{\mathbb{R}^d} \mu_s(x_0, \mathrm{d}x_1) \mu_t(0, B - x_1) \\ &= (\mu_s(x_0, \cdot) * \mu_t(0, \cdot))(B),\end{aligned}$$

where we took advantage of the translation invariance *of the semigroup (μ_t) in the sense that for any $x, y \in \mathbb{R}^d$ and $B \in \mathcal{B}^{\otimes d}$ we have $\mu_t(x + y, B) = \mu_t(x, B - y)$.*

Now due to Corollary 3.8.8 from [Dre19] we infer that the convolution on the right-hand side $\mathcal{N}(x_0, (s+t)\mathrm{Id})$ distribution, so it equals $\mu_{s+t}(x_0, \cdot)$, which shows the validity of (3.2.1).

In particular, we obtain the existence of random walk (X_n), $n \in \mathbb{N}_0$, with Gaussian increments via Theorem 3.1.19. However, this is not big news since we knew before (e.g. through Kolmogorov's existence and uniqueness theorem)

[2] In this case we will also say that μ_n is a transition kernel *on* $(E, \mathcal{E})$.

(b) *For $T = [0,\infty)$ define the Markov semigroup (μ_t), $t \in T$, of transition kernels on $(\mathbb{N}_0, 2^{\mathbb{N}_0})$ via $\mu_0 = \delta_0$, and for $\varrho \in (0,\infty)$ fixed, denoting by ν_t the law of a* $\mathrm{Poi}(\varrho t)$ *distributed random variable, we define*

$$\mu_t(x,B) := \nu_t(B-x), \quad \forall x \in \mathbb{N}_0,\, B \in 2^{\mathbb{N}_0}.$$

Then (μ_t), $t \in T$, defines a Markov semigroup on $(\mathbb{N}_0, 2^{\mathbb{N}_0})$: Indeed, for $t \in T$, we have that μ_t is a mapping from $\mathbb{N}_0 \times 2^{\mathbb{N}_0}$ to $[0,1]$, and for each $F \in 2^{\mathbb{N}_0}$, the mapping

$$\mathbb{N}_0 \ni x \mapsto \mu_t(x,F)$$

is certainly $2^{\mathbb{N}_0} - \mathcal{B}$-measurable. Furthermore, for each $x \in \mathbb{N}_0$, the mapping

$$2^{\mathbb{N}_0} \ni F \mapsto \mu_t(x,F)$$

describes a probability measure by definition. It remains to show the Chapman-Kolmogorov equation (3.2.1). *For this purpose, recall that by Corollary 3.8.8 (b) of [Dre19] we have that the convolution of a* $\mathrm{Poi}(\varrho_1 t)$ *with a* $\mathrm{Poi}(\varrho_2 t)$ *distribution results in a* $\mathrm{Poi}((\varrho_1+\varrho_2)t)$ *distribution. Hence, we infer for $s,t \in T$, $x \in \mathbb{N}_0$, and $B \in 2^{\mathbb{N}_0}$ that*

$$\begin{aligned}\mu_s\mu_t(x,B) &= \int_{\mathbb{N}_0} \mu_s(x,\mathrm{d}y)\mu_t(y,B) = \int_{\mathbb{N}_0} \mu_s(x,\mathrm{d}y)\mu_t(0,B-y) = (\nu_s(\cdot - x) * \nu_t)(B)\\ &= (\nu_s * \nu_t)(B-x) = \nu_{s+t}(B-x) = \mu_{s+t}(x,B),\end{aligned}$$

and we're done. This semigroup is also referred to as the Poisson semigroup (with parameter ϱ).

In both of the above examples we have seen that the fact that we were able to write the product of Markov kernels as a convolution of probability measures was crucial. In fact, there is a general theory of (translation invariant) convolution semigroups making this more precise, but we will not go into further detail here.

While Markov semigroups are the most important ones to us, our fundamental existence result below can be formulated for more general families of transition kernels.

Definition 3.2.3. *Let $(E,\mathcal{E})$ be a measurable space and let a two parameter family of transition kernels $\mu_{s,t}$, $s,t \in T$, where $T \subset \mathbb{R}$, on $(E,\mathcal{E})$ be given. The family is called* consistent *if $\mu_{s,t}\mu_{t,u} = \mu_{s,u}$ for all $s,t,u \in T$ with $s < t < u$.*

Example 3.2.4. *For any Markov semigroup μ_t, $t \in [0,\infty)$, the family defined via*

$$\widetilde{\mu}_{s,t} := \mu_{t-s}, \quad s,t \in [0,\infty),\, s < t,$$

is a consistent family of transition kernels.

The following can be interpreted as a generalisation of Ionescu-Tulcea's theorem to some uncountable index sets $T \subset \mathbb{R}$. However, the price we pay is that we also lose some generality when comparing to Ioncescu-Tulcea since we only deal with consistent families of transition kernels as well as Markov kernels in what comes below. The stochastic processes that naturally arise from Markov kernels (i.e., Markov processes, see Theorem 3.2.8 below) generally provide a good approximation to real world processes on the one hand, and on the other hand they are quite amenable to mathematical investigations and numerical simulations.

Theorem 3.2.5. *Let $\mu_{s,t}$, $s,t \in T \subset [0,\infty)$, with $s < t$ be a consistent family of transition kernels on a Polish space $(S,\mathcal{O})$ with $0 \in T$.*
Then there exists a (unique) transition kernel μ from $(S,\mathcal{B}(S))$ to $(S^T, \mathcal{B}(S)^{\otimes T})$ such that for all $x \in S$ and all finite subsets $J = \{t_0, t_1, \ldots, t_n\} \subset T$ with $0 = t_0 < t_1 < \ldots < t_n$, we have

$$\mu(x,\cdot) \circ \pi^{-1}_{(t_0,\ldots,t_n)}(B) = \Big(\delta_x \otimes \bigotimes_{i=1}^n \mu_{t_{i-1},t_i}\Big)(B) \quad \forall B \in \mathcal{B}(S)^{\otimes J}. \tag{3.2.3}$$

Proof. Similarly to the proof we gave for Ionescu-Tulcea's theorem before, the proof can be performed using Kolmogorov's existence and uniqueness theorem. We leave the details to the Reader and refer to [Kle14, Thm. 14.42] for a proof. □

We can generalise the previous result to the case of Markov semigroups with an arbitrary given initial distribution.

Corollary 3.2.6. *Let (μ_t), $t \in T \subset [0, \infty)$, be a Markov semigroup of transition kernels on a Polish space $(S, \mathcal{O})$ with $0 \in T$.*
If ν is a probability measure on $(S, \mathcal{B}(S))$, then there exists a unique probability measure $\mathbb{P}_\nu$ on $(S^T, \mathcal{B}(S)^{\otimes T})$ such that for all $0 = t_0 < t_1 < \ldots < t_n$ with $t_i \in T$,

$$\mathbb{P}_\nu \circ (\pi_{t_0}, \ldots, \pi_{t_n})^{-1} = \nu \otimes \bigotimes_{i=0}^{n-1} \mu_{t_{i+1}-t_i}.$$

In the case $\nu = \delta_x$ some $x \in E$, we also write $\mathbb{P}_x$ for $\mathbb{P}_\nu$.

Proof. This follows from Theorem 3.2.5. □

The previous result tells us that given a suitable family of transition kernels, we can find a probability measure on the T-fold product space which projects down on the finite dimensional spaces and fulfills the corresponding consistency condition. As we have seen before in the context of Kolmogorov's existence and uniqueness theorem, this gives rise to a stochastic process (X_t), $t \in T$, with the property that its law is given by $\mathbb{P}$.
Since we will not delve in the depths of Markov processes here but only give some basic results, we will restrict ourselves to giving the definition of so-called time homogeneous processes – i.e, the transition probabilities from x at time s to B at time t depend on s and t only through the difference $t - s$.
However, every so often in mathematics, when concepts become more advanced and more technical, there is oftentimes an entire zoo of just slightly different definitions around. This is the case for Markov processes as well.

Definition 3.2.7. *Let a Polish space $(S, \mathcal{O})$ be given, as well as a family $(\mathbb{P}_x)$, $x \in S$, of probability measures on $(\Omega, \mathcal{F})$, and measurable mappings (X_t), $t \in T$, from $(\Omega, \mathcal{F})$ to $(S, \mathcal{B}(S))$ i.e., for each $x \in S$, the family (X_t), $t \in T$, is a stochastic process on $(\Omega, \mathcal{F}, \mathbb{P}_x)$. If*

(a)
$$\mathbb{P}_x(X_0 = x) = 1 \quad \forall x \in S;$$

(b) for every $t \in T$, the mapping
$$S \times \mathcal{B}(S) \ni (x, F) \mapsto \mathbb{P}_x(X_t \in F)$$
defines a transition kernel on $(S, \mathcal{B}(S))$;

(c) the Markov property is fulfilled in the sense that for all $x \in S$, $s, t \in T$ with $s < t$, and for all $B \in \mathcal{B}(S)$, we have
$$\mathbb{P}_x(X_{t+s} \in B \,|\, \mathcal{F}_s) = \mathbb{P}_{X_s}(X_t \in B) \quad \mathbb{P}_x - a.s., \tag{3.2.4}$$
with $(\mathcal{F}_t)$ the canonical filtration.

then we call the process (X_t), $t \in T$, *a* Markov process.[3]

We will also get back to the general Definition 3.2.7 in Chapter 5 when introducing Brownian motion. For the time being, however, before giving examples in Section 3.2.1, we will show how Markov processes generally arise from the semigroups introduced above.

Theorem 3.2.8.
Let (μ_t), $t \in T \subset [0, \infty)$, *be a Markov semigroup of transition kernels on a Polish space* $(S, \mathcal{O})$ *with* $0 \in T$.
Then there exists a family of probability measures $(\mathbb{P}_x)$, $x \in S$, *on a measurable space* $(\Omega, \mathcal{F})$ *as well as a stochastic process* (X_t), $t \in T$, *defined* $(\Omega, \mathcal{F}, \mathbb{P}_x)$, *such that the process* (X_t) *is a Markov process on* $(\Omega, \mathcal{F}, \mathbb{P}_x)$, $x \in S$, *with respect to the canonical filtration* $(\mathcal{F}_t)$, *and with transition probabilities given by*

$$\mathbb{P}_x(X_t \in B) = \mu_t(x, B), \quad \forall t \in T,\, x \in S,\, B \in \mathcal{B}(S). \tag{3.2.5}$$

Vice versa, if there is a family of probability measures $(\mathbb{P}_x)$, $x \in S$, *and a probability space* $(\Omega, \mathcal{F})$ *such that* (X_t), $t \in T$, *is a Markov process on* $(\Omega, \mathcal{F}, \mathbb{P}_x)$, $x \in S$, *then defining* μ_t *via the equation in* (3.2.5) *we obtain a Markov semigroup.*

Proof. Let a Markov semigroup as in the assumptions be given. Corollary 3.2.6 implies that for the given Markov semigroup, for each $x \in S$ there exists a unique probability measure $\mathbb{P}_x$ on $(S^T, \mathcal{B}(S)^{\otimes T})$ such that for $0 = t_0 < t_1 < \ldots < t_n$ with $t_i \in T$,

$$\mathbb{P}_x \circ \pi^{-1}_{t_0,\ldots,t_n} = \delta_x \otimes \bigotimes_{i=0}^{n-1} \mu_{t_{i+1}-t_i}, \tag{3.2.6}$$

Define on $(S^T, \mathcal{B}(S)^{\otimes T}, \mathbb{P}_x)$ the coordinate process

$$X_t : S^T \ni \omega \mapsto \omega_t \in S, \quad t \in T.$$

Then the general setup of Definition 3.2.7 is fulfilled, and also Part (a) holds true. Furthermore, Part (b) is fulfilled since (3.2.6) implies

$$\mathbb{P}_x(X_t \in F) = \mu_t(x, F), \quad \forall x \in S,\, F \in \mathcal{B}(S), \tag{3.2.7}$$

and μ_t is a transition kernel by assumption. It remains to show

$$\mathbb{P}_x(X_{t_n} \in F_n \,|\, \mathcal{F}_{t_{n-1}}) = \mathbb{P}_{X_{t_{n-1}}}(X_{t_n - t_{n-1}} \in F_n) \quad \mathbb{P}_x - a.s., \tag{3.2.8}$$

and we do so by showing that the right-hand side has the required measurability and integrability properties of (a version of) $\mathbb{P}_x(X_{t_n} \in F_n \,|\, \mathcal{F}_{t_{n-1}})$.
To start with, note that the $\mathcal{F}_{t_{n-1}} - \mathcal{B}$-measurability of the random variable on the right-hand side of (3.2.8) follows from (3.2.7) and the $\mathcal{F}_{t_{n-1}} - \mathcal{B}$-measurability of $X_{t_{n-1}}$. Furthermore, from (3.2.6) we infer that for $0 = s_0 < s_1 < \ldots < s_m := t_{n-1}$, $m \in \mathbb{N}$, as well as $F_i \in \mathcal{B}(S)$, $i \in \{0, \ldots, m\}$, we have

$$\begin{aligned}
&\mathbb{P}_x\big(X_{t_n} \in F_n,\, X_{s_i} \in F_i,\, \forall i \in \{0, 1, \ldots, m\}\big) \\
&= \int_{F_{n-1}} \mathbb{P}_x\big(X_{s_i} \in F_i,\, \forall i \in \{0, 1, \ldots, m-1\},\, X_{t_{n-1}} \in \mathrm{d}x_{n-1}\big) \cdot \mu_{t_n - t_{n-1}}(x_{n-1}, F_n) \\
&= \mathbb{E}_x\Big[\prod_{i=0}^{m} \mathbb{1}_{F_i}(X_{s_i}) \cdot \mu_{t_n - t_{n-1}}(X_{t_{n-1}}, F_n)\Big] \\
&= \mathbb{E}_x\Big[\prod_{i=0}^{m} \mathbb{1}_{F_i}(X_{s_i}) \cdot \mathbb{P}_{X_{t_{n-1}}}(X_{t_n - t_{n-1}} \in F_n)\Big].
\end{aligned} \tag{3.2.9}$$

[3]Similar concepts also go by 'universal Markov process' ([Bau92]) or 'Markov family' ([KS91], [Str11]) in the literature.

Since the π-system $\{X_{s_i} \in F_i,\ 0 = s_0 < s_1 < \ldots < s_m = t_{n-1},\ F_i \in \mathcal{B}(S),\ m \in \mathbb{N}\}$ generates $\mathcal{F}_{t_{n-1}}$, in combination with Corollary 1.2.18 of [Dre19] this implies (3.2.8), and hence (X_t) as defined above is a Markov process indeed.
To establish the converse direction, let a Markov process be given. So we have a family of probability measures $(\mathbb{P}_x)$, $x \in S$, with the required properties, and we define a transition kernel via

$$\mu_t(x, F) := \mathbb{P}_x(X_t \in F), \quad t \in T,\ x \in S,\ F \in \mathcal{B}(S). \tag{3.2.10}$$

Then μ_t defines a transition kernel on $(S, \mathcal{B}(S))$ due to Definition 3.2.7 (b).
Using the Markov property we find that

$$\begin{aligned}(\mu_s \mu_t)(x, F) &= \int \mu_s(x, \mathrm{d}x_1)\mu_t(x_1, F) = \int \mathbb{P}_x(X_s \in \mathrm{d}x_1)\mathbb{P}_{x_1}(X_t \in F) \\ &= \mathbb{E}_x\big[\mathbb{P}_{X_s}(X_t \in F)\big] = \mathbb{P}_x(X_{s+t} \in F) = \mu_{s+t}(x, F),\end{aligned}$$

which all in all establishes that (μ_t), $t \in T$, forms a Markov semigroup.

□

While for the case $T = \mathbb{N}_0$, this is essentially all we want to have, in the case $T = [0, \infty)$ very interesting questions arise concerning the so-called *path properties* of a process (X_t), $t \in [0, \infty)$. Indeed, for $(\Omega, \mathcal{F}, \mathbb{P}) := ([0,1], \mathcal{B}([0,1]), \lambda_{|[0,1]})$ we can define stochastic processes (X_t), $t \in [0,1]$, via

$$X_t(\omega) = \mathbb{1}_\omega(t) \quad t \in [0,1],\ \omega \in \Omega.$$

as well as (Y_t), $t \in [0,1]$, via

$$Y_t(\omega) = 0 \quad t \in [0,1],\ \omega \in \Omega.$$

Then (X_t) is a process for which its *sample paths*

$$[0,1] \ni t \mapsto X_t(\omega), \quad \omega \in \Omega,$$

are not continuous, whereas the paths

$$[0,1] \ni t \mapsto Y_t(\omega), \quad \omega \in \Omega,$$

are continuous for all $\omega \in \Omega$.
However, all finite dimensional distributions of (X_t) and (Y_t) coincide, and hence so does the projective limit, i.e., the consistent probability measure on $(\mathbb{R}^{[0,1]}, \mathcal{B}^{\otimes [0,1]})$. In particular, this implies that the path properties are therefore not uniquely characterised by the projective limit. Therefore, in addition to comparing processes by comparing their finite-dimensional distributions, it turns out to be useful to consider the following stronger properties.

Definition 3.2.9. *Let (X_t) and (Y_t) be two stochastic processes. We say that (X_t) is a* modification *of (Y_t) if for each $t \in T$ we have*

$$\mathbb{P}(X_t = Y_t) = 1.$$

Definition 3.2.10. *Let (X_t) and (Y_t) be two stochastic processes. We say that (X_t) and (Y_t) are* indistinguishable *if there exists a set A with $\mathbb{P}(A) = 1$ such that*

$$A \subseteq \{X_t = Y_t\, \forall t \in T\}$$

It is easy to check that if (X_t) and (Y_t) are indistinguishable, then they are also modifications of each other, and if they are modifications of each other, then they have the same finite-dimensional distributions.

Exercise 3.2.11. *Show that the converses of the previous implications are not generally true.*

In particular, Example 3.2.2 (a) that Theorem 3.2.5 applied to the Brownian semigroup would almost already give us Brownian motion – however, using this construction we can so far not guarantee continuous sample paths (or let alone that the paths are measurable functions from $[0, \infty)$ to $\mathbb{R}$).
There is, however, a powerful abstract result which gives a sufficient criterion for the existence of a modification of the original process such that this modification has continuous sample paths. Before stating the result we recall the notion of Hölder continuity.

Definition 3.2.12. *Let (X_1, d_1) and (X_2, d_2) be metric spaces. For $\alpha \in (0, 1]$, a function $f : X_1 \to X_2$ is called* α-Hölder continuous at $x \in X_1$ *if for each $\varepsilon > 0$,*

$$\sup_{y \in X_1, \mathrm{d}_1(x,y) \le \varepsilon} \frac{\mathrm{d}_2(f(x), f(y))}{\mathrm{d}_1(x, y)^\alpha} \text{ is finite.} \tag{3.2.11}$$

f is called locally α-Hölder continuous if for each $x \in X_1$ there exists $\varepsilon > 0$ such that

$$\sup_{\substack{y,z \in X_1, \\ \mathrm{d}_1(x,y) \le \varepsilon \\ \mathrm{d}_1(x,z) \le \varepsilon}} \frac{\mathrm{d}_2(f(y), f(z))}{\mathrm{d}_1(y, z)^\alpha} \text{ is finite.}$$

It is called α-Hölder continuous *if (3.2.11) is bounded from above as a function in $x \in X_1$.*

Theorem 3.2.13 (Kolmogorov-Chentsov). *Let (X_t), $t \in [0, \infty)$, be a stochastic process taking values in a complete metric space (X, d). If for every $T > 0$ there exist constants $\alpha, \beta, C \in (0, \infty)$ such that*

$$\mathbb{E}[\mathrm{d}(X_s, X_t)^\alpha] \le C|s - t|^{1+\beta}, \quad \forall s, t \in [0, T],$$

then there exists a modification $(\widetilde{X}_t)$, $t \in [0, \infty)$, of (X_t), $t \in [0, \infty)$, such that for each $\gamma \in (0, \frac{\beta}{\alpha})$, the sample paths $[0, \infty) \ni t \mapsto \widetilde{X}_t(\omega)$ are locally γ-Hölder continuous for $\mathbb{P}$-almost all $\omega \in \Omega$.

Proof. A complete proof of this result would divert us too far from the main topics of this class. We therefore refer to [Kal02, Thm. 3.23] for further details. □

Exercise 3.2.14. *(a) Consider Example 3.2.2 (a). Start with showing the following result.*

> **Lemma 3.2.15.** *Let $Y \sim \mathcal{N}(0, t \cdot \mathrm{Id}_d)$ be a d-dimensional normally distributed random variable with covariance matrix $t \cdot \mathrm{Id}_d$. Then for $n \in \mathbb{N}$ there exists a constant $C_n \in (0, \infty)$ such that for all $t \in (0, \infty)$,*
>
> $$\mathbb{E}[\|Y\|_2^{2n}] = C_n t^n.$$

(b) Show that the Markov process arising from the Poisson semigroup from Example 3.2.2 (b) does not satisfy the assumptions of Theorem 3.2.13 for any admissible choice of parameters (in distribution this is the 'Poisson process', but we will require some sample paths regularity and hence construct it explicitly below).

Taking advantage of this powerful result, with a yet another small definition we are now even able to prove the existence of Brownian motion.

Definition 3.2.16. *An $(\mathbb{R}^d, \mathcal{B}^d)$-valued stochastic process (X_t), $t \in [0, \infty)$, has* independent increments *if*

$$\forall\, s, t \in [0, \infty) \text{ with } s < t, \text{ we have that } X_t - X_s \text{ is independent of } \mathcal{F}_s, \tag{3.2.12}$$

with $(\mathcal{F}_t)$, $t \in [0, \infty)$, denoting the canonical filtration of (X_t), $t \in [0, \infty)$.
Furthermore, we say that the process (X_t) has stationary increments, *if for any $0 < s < t$, we have that all random variables $X_{t+r} - X_{s+r}$, $r \ge 0$, have the same distribution.*

Remark 3.2.17. *Condition (3.2.12) is equivalent to the following: For all $t_1, \dots, t_n \in [0, \infty)$ with $t_1 < \dots < t_n$, the random variables $X_{t_{i+1}} - X_{t_i}$, $i \in \{1, \dots, n-1\}$, form an independent family.*

Definition 3.2.18. *Let $x \in \mathbb{R}^d$, and let (B_t), $t \in [0, \infty)$ be an $(\mathbb{R}^d, \mathcal{B}^d)$-valued stochastic process defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that:*

(a)

$$B_0 = x;$$

(b) the increments of the process (B_t) are independent;

(c) for any $0 \le s < t$, the random variable $B_t - B_s$ is $\mathcal{N}(0, (t-s)\mathrm{Id}_d)$-distributed;

(d) $\mathbb{P}$-almost surely, the mapping

$$[0, \infty) \ni t \mapsto B_t$$

is continuous.

Then (B_t) is called Brownian motion starting in x*. If $x = 0$, then this process is also referred to as* standard Brownian motion.

Remark 3.2.19. *As in Corollary 3.2.6, if we can construct Brownian starting in $x \in \mathbb{R}^d$, then, given a probability distribution ν on $(\mathbb{R}^d, \mathcal{B}^d)$, we can also construct Brownian motion with initial distribution ν.*

Theorem 3.2.20. *For any $x \in \mathbb{R}^d$, Brownian motion starting in x exists, and $\mathbb{P}$-a.s. the sample paths are even locally α-Hölder continuous for each $\alpha \in (0, \frac{1}{2})$.*

Proof. Recall the Brownian semigroup defined in Example 3.2.2 (a). Hence, due to Theorem 3.2.8 and (3.2.6), we can infer the existence of a Markov process with independent and appropriately normally distributed increments.
The only property which has not been addressed so far is that of local α-Hölder continuous sample paths. Due to Exercise 3.2.14 (a), however, for $\alpha \in (0, \frac{n-1}{2n})$ we can invoke Theorem 3.2.13 to provide us with the existence of a modification of the original process such that this modification has the desired local α-Hölder continuity of the sample paths.
We now observe that since if two processes are modifications of each other, we have that $\mathbb{P}$-a.s., they coincide at all times $t \in \mathbb{Q} \cap [0, \infty)$. But since $\mathbb{P}$-a.s., the sample paths of continuous processes (indexed by $[0, \infty)$) are already characterised by their evaluations at rational times, we infer that α-Hölder continuous version from above must already be $\mathbb{P}$-a.s. locally α-Hölder continuous for *all* $\alpha \in (0, \frac{1}{2})$. □

Before we move on from Brownian motion to another process, it is natural to ask at this stage whether more than one Brownian motion that starts at x exists, that is, does the choice of the probability space lead to "different" processes. We answer this question in the following proposition.

Proposition 3.2.21. *Let (X_t), $\mathbb{P}_x, x \in S$ be a Markov process and let (Y_t) be a stochastic process on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ such that*

$$\tilde{\mathbb{P}}(Y_t \in B \,|\, \mathcal{F}_s) = \mathbb{P}_{Y_s}(X_{t-s} \in B), \quad s < t \in T, B \in \mathcal{B}(S). \tag{3.2.13}$$

Then the distribution of (Y_t) under $\tilde{\mathbb{P}}$ is the same as the distribution of (X_t) under $\mathbb{P}_\nu$, where $\mathbb{P}_\nu := \int \mathbb{P}_x \nu(dx)$ and $Y_0 \sim \nu$.

Proof. Note first that (3.2.13) and algebraic induction imply that

$$\tilde{\mathbb{E}}[f(Y_t)\,|\,\tilde{\mathcal{F}}_s] = \mathbb{E}_{Y_s}[f(X_{t-s})]$$

for any measurable function f. To show that the distributions of (X_t) and (Y_t) under their respective laws are the same, it suffices to show that for all $n \in \mathbb{N}_0$ and any choice of measurable functions $f_1, \ldots, f_n$ it holds that

$$\tilde{\mathbb{E}}[f_1(Y_{t_1}) \cdots f_n(Y_{t_n})] = \mathbb{E}_\nu[f_1(X_{t_1}), \cdots, f_n(X_{t_n})].$$

We prove this by induction, starting with $n = 1$. Here it holds that

$$\tilde{\mathbb{E}}[f_1(Y_{t_1})] = \tilde{\mathbb{E}}[\tilde{\mathbb{E}}[f_1(Y_{t_1})\,|\,\tilde{\mathcal{F}}_0]] = \tilde{\mathbb{E}}[\mathbb{E}_{Y_0}[f_1(X_{t_1})]] = \mathbb{E}_\nu[f_1(X_{t_1})],$$

where the second equality follows by our initial observation and the last equality follows by the definition of $\mathbb{P}_\nu$.
We now argue the inductive step $n \to n+1$. Writing $g_n(x) = f_n(x)\mathbb{E}_x[f_{n+1}(X_{t_{n+1}-t_n})]$ and noting that this defines a measurable function, we have

$$\begin{aligned}
\mathbb{E}_\nu[f_1(X_{t_1}) \cdots f_{n+1}(X_{t_{n+1}})] &= \mathbb{E}_\nu[f_1(X_{t_1}) \cdots f_n(X_{t_n})\mathbb{E}_{X_{t_n}}[f_{n+1}(X_{t_{n+1}-t_n})]] \\
&= \mathbb{E}_\nu[f_1(X_{t_1}) \cdots g_n(X_{t_n})] = \tilde{\mathbb{E}}[f_1(Y_{t_1}) \cdots g_n(Y_{t_n})] \\
&= \tilde{\mathbb{E}}[f_1(Y_{t_1}) \cdots f_n(Y_{t_n})\mathbb{E}_{Y_{t_n}}[f_{n+1}(X_{t_{n+1}-t_n})]] \\
&= \tilde{\mathbb{E}}[f_1(Y_{t_1}) \cdots f_n(Y_{t_n})\tilde{\mathbb{E}}[f_{n+1}(Y_{t_{n+1}})\,|\,\tilde{\mathcal{F}}_{t_n}]] \\
&= \tilde{\mathbb{E}}[f_1(Y_{t_1}) \cdots f_{n+1}(Y_{t_{n+1}})],
\end{aligned}$$

which concludes the proof. □

In the following section we construct in detail (and without referring to results we have not yet proven) another process that will be important to us. In addition to the distributional requirements, we will also impose additional properties on the sample paths.

3.2.1 Poisson process

In order to construct the Poisson process, we could proceed using the machinery introduced above (see Remark 3.2.26); we will, however, follow a more hands-on approach here which will provide us with the fact that $\mathbb{P}$-a.s., the sample paths of the process will be càdlàg, i.e., 'continue à droite, limite à gauche' (English: RCLL, 'right-continuous with left-limits').

Definition 3.2.22. *For $\kappa > 0$ let (T_n), $n \in \mathbb{N}$, be a sequence of i.i.d.* Exp(κ)*-distributed random variables. For $t \in [0, \infty)$ define*

$$N_t := \sup\Big\{n \in \mathbb{N}_0 \,:\, \sum_{i=1}^n T_i \le t\Big\}. \tag{3.2.14}$$

The stochastic process (N_t), $t \in [0,\infty)$, is called a Poisson process with rate *κ.*

Lemma 3.2.23. *Let N_t be a Poisson process with rate $\kappa \in (0, \infty)$. Then for $\mathbb{P}$-a.a. $\omega \in \Omega$, the sample paths*

$$[0, \infty) \ni t \mapsto N_t(\omega)$$

are right-continuous with left limits.

Proof. From (3.2.14) it follows that for any $t \in (0, \infty)$, we have that $\mathbb{P}$-a.s., $N_t \in \mathbb{N}_0$. In particular, $\mathbb{P}$-a.s., (N_t) has only finitely many jumps on the interval $[0, t]$, and using (3.2.14) again, we infer that $\mathbb{P}$-a.s., (N_t) has left limits and is right-continuous on $[0, t]$. Summing over $t \in \mathbb{N}$ establishes the claim.

□

The reason that the process introduced above is called *Poisson process* is given by the following lemma.

Proposition 3.2.24. *Let (N_t), $t \in [0,\infty)$, be a Poisson process with rate $\kappa > 0$. Then (N_t) has stationary and independent increments; furthermore, for any $s,t \in [0,\infty)$ with $s < t$,*

$$N_t - N_s \sim \mathrm{Poi}(\kappa(t-s));$$

We furthermore take advantage of the following claim.

Claim 3.2.25. *Let $X_1,\dots,X_n$ be independent real random variables whose distributions have densities $\varphi_1,\dots,\varphi_n$ with respect to the Lebesgue measure. Then the random vector $(X_1,\dots,X_n)$ has density $(x_1,\dots,x_n) \mapsto \varphi_1(x_1)\cdot\ldots\cdot\varphi_n(x_n)$ with respect to the n-dimensional Lebesgue measure.*

Proof. For $B_1,\dots,B_n \in \mathcal{B}(\mathbb{R})$ we have

$$\begin{aligned}\mathbb{P}((X_1,\dots,X_n) \in B_1\times\ldots\times B_n) &= \prod_{i=1}^n \mathbb{P}(X_i \in B_i) = \prod_{i=1}^n \int_{B_i} \varphi_i(x_i)\,\mathrm{d}x_i \\ &= \int_{B_1\times\ldots\times B_n} \prod_{i=1}^n \varphi_i(x_i)\,\mathrm{d}(x_1,\dots,x_n),\end{aligned}$$

Hence, the two measures on $\mathcal{B}(\mathbb{R}^n)$ defined via

$$\mathbb{P}((X_1,\dots,X_n)\in\cdot)$$

and

$$\int_{\cdot} \prod_{i=1}^n \varphi_i(x_i)\,\mathrm{d}(x_1,\dots,x_n)$$

coincide on the generating π-system of $\mathcal{B}(\mathbb{R}^d)$ given by the hypercubes, and thus Theorem 1.2.16 of [Dre19] yields the result.

□

Proof sketch of Proposition 3.2.24. We show that for $k,n \in \mathbb{N}_0$ as well $0 < s < t$ we have

$$\mathbb{P}(N_s = k, N_t - N_s = n) = e^{-\kappa s}\frac{(\kappa s)^k}{k!}e^{-\kappa(t-s)}\frac{(\kappa(t-s))^n}{n!}. \tag{3.2.15}$$

While this only proves that the two increments N_s and $N_t - N_s$ are independent and have the desired distribution, the general case follows in a similar way but is notationally strictly more involved, hence we leave it to the reader.
Using that the $T_1, T_2, \dots$ form an i.i.d. sequence of $\mathrm{Exp}(\kappa)$ variables and taking advantage of Claim 3.2.25, we obtain integrating iteratively from 'the inside out' that

$$\begin{aligned}\mathbb{P}(N_s = k, N_t = n+k) = \kappa^{n+k+1} &\int_{[0,\infty)} \mathrm{d}x_1 \dots \int_{[0,\infty)} \mathrm{d}x_{n+k+1} \\ &e^{-\kappa\sum_{i=1}^{n+k+1} x_i} \mathbb{1}_{\{\sum_{i=1}^k x_i \le s,\, x_{k+1} > s - \sum_{i=1}^k x_i,\, \sum_{i=1}^{n+k} x_i \le t,\, x_{n+k+1} > t - \sum_{i=1}^{n+k} x_i\}}.\end{aligned}$$

The crucial observation now is that for $x_1,\dots,x_{k+n}$ fixed with $\sum_{i=1}^{n+k} x_i \le t$,

$$\kappa\int_{[0,\infty)} \mathrm{d}x_{n+k+1}\, e^{-\kappa\sum_{i=1}^{n+k+1} x_i}\mathbb{1}_{\{x_{n+k+1} > t - \sum_{i=1}^{n+k} x_i\}} = e^{-\kappa t},$$

so after this integration we're left with only indicators in the previous iterated integral. Furthermore, for $x_1, \dots, x_k$ such that $\sum_{i=1}^k x_i \le s$, by induction on n we obtain

$$\int_{[0,\infty)} \mathrm{d}x_{k+1} \dots \int_{[0,\infty)} \mathrm{d}x_{n+k}\, \mathbb{1}_{\{x_{k+1} > s - \sum_{i=1}^k x_i,\, \sum_{i=k+1}^{n+k} x_i \le t - \sum_{i=1}^k x_i\}} = \frac{(t-s)^n}{n!}.$$

Hence, the first display of the proof can be continued to obtain

$$\mathbb{P}(N_s = k, N_t = n+k) = \frac{(\kappa s)^k}{k!} \frac{(\kappa(t-s))^n}{n!} e^{-\kappa t},$$

which establishes (3.2.15) and hence proves what we wanted. □

It is often assumed that the interarrival times between customers to e.g. a restaurant or a store, or the times between calls to a customers service department are i.i.d. exponentially distributed.[4] As pointed out in the discussion subsequent to Theorem 1.9.9 of [Dre19] already, it seems reasonable to assume the number of customers that arrive in a certain time interval to be Poisson distributed. Therefore, the above gives yet another manifestation of it (hinging on the same mathematical assumptions, though).

Remark 3.2.26. *Alternatively, we can also introduce the Poisson process (modulo continuous sample paths) as the Markov process arising from the Poisson semigroup via Example 3.2.2 (b) and Theorem 3.2.8, where* (3.2.6) *ensures the independent and appropriately Poisson distributed increments. In particular, the conclusions of Proposition 3.2.24 would essentially have followed from the construction.*
This, however, would not yet have given us the connection to the sum of independent exponentially distributed random variables as derived above. More importantly, as we have seen before, the construction of the process by specifying a Markov semigroup / the finite dimensional distributions, does not yet give us information on the regularity of the sample paths $t \mapsto X_t(\omega)$. Instead, for the path regularity we could try to invoke the Kolmogorov-Chentsov theorem; you can check, cf. Example 3.2.14 (b), that the assumptions of this result are not fulfilled in the context of the Poisson process, so we could not deduce any path regularity from it. This makes sense in the light of the hands-on construction that we followed above, where we obtain that $\mathbb{P}$-a.s., the sample paths are càdlàg, but not continuous.

It can be shown that the above can be generalised to Poisson processes $N(t)$ with non-constant rates, i.e., when we have a function $\kappa : [0,\infty) \to [0,\infty)$, then $N(t) - N(s)$ is Poisson distributed with parameter $\int_s^t \kappa(r)\mathrm{d}r$. Like the Poisson process, the generalised Poisson process turns out very useful in investigating mathematical problems that find their motivations in applications such as statistical physics or finance, see Example 4.1.12 below.
One reason the Poisson process is so important in probability theory is that the time between two jumps is exponentially distributed and hence has no memory. Indeed, we (might) have seen already, that the only distributions without memory are the exponential distributions (and accommodating the discrete nature of the geometric distribution, it can also be considered memoryless). This meant that if $X \sim \mathrm{Exp}(\kappa)$, then

$$\mathbb{P}(X \ge s + t \,|\, X \ge t) = \mathbb{P}(X \ge s) \quad \forall s, t \in [0,\infty).$$

If we now want to construct a continuous time Markov process which is essentially made up of finitely many jumps on compact intervals (think of simple random walk, but now with time indexed by $[0,\infty)$ instead of $\mathbb{N}_0$; Brownian motion, however, does not fall into this regime), then the waiting times between two consecutive jumps *have to be* exponentially distributed in order

[4] Of course, this is only an approximation since there are usually more people eating out for lunch or dinner than at 4 pm in the afternoon. During peak times, however, this is considered a reasonable approximation.

to preserve the Markov property. Indeed, if the time between two jumps was not memoryless, we could gain some knowledge about the time of a future jump by knowing for how long the process had been sitting at the current site. This, however, translates to $\mathcal{F}_t$ having strictly more information about the future that just X_t, a contradiction to (3.0.1).

Example 3.2.27 (Continuous time random walk)**.** *Let (N_t) be a Poisson process with parameter $\kappa > 0$, and let (X_n), $n \in \mathbb{N}$, be an i.i.d. sequence of real random variables. Then*

$$S_t := \sum_{i=1}^{N_t} X_i$$

defines a continuous time random wxalk (with increment distribution $\mathbb{P} \circ X_1^{-1}$).

3.3 Martingales and the Markov property

There is a fundamental connection between Markov processes and martingales which goes back (at least) to [SV79]. While the resulting theory is particularly useful in the case of continuous time when investigating stochastic differential equations and diffusion processes, we will touch upon it in the setting of discrete time and countable state space in order to keep technicalities light and understanding deep. In particular, we will denote by P the corresponding transition matrix.
To introduce the principal ideas assume the following setting: Assume given a stochastic process (X_n), $n \in \mathbb{N}_0$, with each X_n taking values in a countable state space S, and suppose that we want to characterise the distribution of the entire process as a probability measure on $(S^{\mathbb{N}_0}, (2^S)^{\otimes \mathbb{N}_0})$. In order to stick to the notation of Theorem 3.1.19, for $n \in \mathbb{N}$, set $\Omega_n := \times_{i=0}^n S_i$ and $\Omega := \times_{i=0}^\infty S_i$, and define the σ-algebras $\mathcal{F}_n := \otimes_{i=0}^n 2^S$, as well as $\mathcal{F} := \otimes_{i=0}^\infty 2^S$ on these spaces. Then the sequence of regular conditional probability distributions of X_n given $\mathcal{F}_{n-1}$ in combination with the distribution of X_0 already completely determines the distribution of the process (X_n). Indeed, since the cylinder sets form a π-system generating the product σ-algebra, the distribution of the process (X_n) is completely determined by its projections down to $(\Omega_n, \mathcal{F}_n)$, and for $F_0 \times \ldots \times F_n \in \mathcal{F}_n$ we have using the the tower property for conditional expectations in combination with the disintegration formula, that

$$\mathbb{P}((X_0, \ldots, X_n) \in F_0 \times \ldots \times F_n) = \mathbb{E}[\ldots \mathbb{E}[\underbrace{\mathbb{E}[\mathbb{1}_{X_n \in F_n} \,|\, \mathcal{F}_{n-1}]}_{\overset{\text{Theorem 3.1.8}}{=} \mu_n(\cdot, \{X_n \in F_n\})} \mathbb{1}_{X_{n-1} \in F_{n-1}} \,|\, \mathcal{F}_{n-2}] \ldots],$$

with μ_n the regular conditional probability of X_n given $\mathcal{F}_{n-1}$.[5] Specifying these (μ_n) again is equivalent to specifying the conditional expectations

$$\mathbb{E}[f(X_n) \,|\, \mathcal{F}_{n-1}]$$

for a sufficiently rich class of bounded functions $f : S \to \mathbb{R}$. More specifically, for such f fixed we can introduce the notation

$$h_{n-1}(X_0, \ldots, X_{n-1}) := \mathbb{E}[f(X_n) \,|\, \mathcal{F}_{n-1}] - f(X_{n-1}), \tag{3.3.1}$$

which in particular implies

$$\mathbb{E}\big[f(X_n) - f(X_{n-1}) - h_{n-1}(X_0, \ldots, X_{n-1}) \,|\, \mathcal{F}_{n-1}\big] = 0,$$

[5] Essentially, this follows from Ionescu-Tulcea's theorem 3.1.19 since the μ_n we have here can be 'extended' to transition kernels $\widetilde{\mu}_n$ from $(\Omega_{n-1}, \mathcal{F}_{n-1})$ to $(\Omega_n, \mathcal{F}_n)$ by defining

$$\widetilde{\mu}_n(\cdot, (X_1, \ldots, X_n) \in F_1 \times \ldots \times F_n) := \mathbb{1}_{(X_1, \ldots, X_{n-1}) \in F_1 \times \ldots \times F_{n-1}} \mu_n(\cdot, X_n \in F_n).$$

and so summing the previous display over time yields that the process

$$f(X_n) - f(X_0) - \sum_{i=0}^{n-1} h_i(X_0, \dots, X_i), \quad n \in \mathbb{N}_0,$$

defines a martingale.
Now if (X_n) is not only a stochastic process but even satisfies the Markov property (3.0.1), then we immediately infer that h_{n-1} can be written as a function of X_{n-1} only instead of $X_0, \dots, X_{n-1}$, and furthermore h_{n-1} does not depend on $n-1$. Hence, in a slight abuse of notation we can rewrite h through (3.3.1) as

$$h(x) := \sum_{y \in S} P(x, y) f(y) - f(x) = ((P - I)f)(x),$$

i.e.,

$$h = (P - \mathrm{Id})f \tag{3.3.2}$$

as functions on S. Thus we infer that if (X_n) is a Markov chain and $f : S \to \mathbb{R}$ bounded, then

$$f(X_n) - f(X_0) - \sum_{i=0}^{n-1} h(X_i), \quad n \in \mathbb{N}_0, \tag{3.3.3}$$

is a martingale again.
The power of this approach stems from the fact that the converse of the above is true as well; i.e., the law of a Markov chain with transition matrix P and some initial state x_0 is the unique probability measure on $(S^{\mathbb{N}_0}, (2^S)^{\mathbb{N}_0})$ such that (3.3.3) describes a martingale for all f and corresponding h.
We introduce the following definition in order to formalise (3.3.2).

Definition 3.3.1. *For a Markov chain on a countable state space S and with transition matrix P, we call the matrix*

$$L := P - \mathrm{Id} \tag{3.3.4}$$

the generator *of the Markov chain.*

Remark 3.3.2. *The definition corresponding to (3.3.4) in continuous time would be the operator L defined via*

$$(Lf)(x) := \lim_{t \downarrow 0} \frac{1}{t} \big(\mathbb{E}_x[f(X_t)] - f(x) \big);$$

of course, existence of Lf is non-trivial here and one would have to investigate for which f that expression actually makes sense. By investigating the discrete time case we avoid these technicalities.

Using this notation we can rewrite (3.3.3) as

$$M_n := f(X_n) - f(X_0) - \sum_{i=0}^{n-1} (Lf)(X_i), \quad n \in \mathbb{N}_0. \tag{3.3.5}$$

Also observe that this describes the Doob decomposition of the process $f(X_n)$, $n \in \mathbb{N}_0$, since (M_n) is a martingale and the sum on the right-hand side defines a previsible process. Now coming back to the above, we can formulate the following result which provides us with a link between Markov chains and those matrices L which make the process in (3.3.5) a martingale.

Theorem 3.3.3 (Martingale problem). *Let (X_n), $n \in \mathbb{N}_0$, be a stochastic process taking values in the countable state space S and adapted to a filtration $(\mathcal{F}_n)$. Furthermore, let L be an operator on $\mathbb{R}^S$.*[6]
Then (X_n) is a Markov chain if and only if for each bounded functions f, the process (M_n) defined in (3.3.5) is a martingale.
In this case, the relation between P and L is given by

$$P = \mathrm{Id} + L. \tag{3.3.6}$$

Proof. We have seen above that if (X_n) is a Markov chain, then (3.3.5) describes a martingale indeed.
To prove the converse direction, in order to show that (X_n) is a Markov process indeed, using the notation of (3.3.5) we compute

$$\begin{aligned}
\mathbb{E}[f(X_{n+m}) \,|\, \mathcal{F}_n] &= \mathbb{E}[M_{n+m} \,|\, \mathcal{F}_n] + f(X_0) + \sum_{i=0}^{n+m-1} \mathbb{E}[(Lf)(X_i) \,|\, \mathcal{F}_n] \\
&= M_n + f(X_0) + \sum_{i=0}^{n-1} (Lf)(X_i) + \sum_{i=n}^{n+m-1} \mathbb{E}[(Lf)(X_i) \,|\, \mathcal{F}_n] \\
&= f(X_n) + \sum_{i=n}^{n+m-1} \mathbb{E}[(Lf)(X_i) \,|\, \mathcal{F}_n].
\end{aligned}$$

Plugging in $m = 1$ we infer that

$$\mathbb{E}[f(X_{n+1}) \,|\, \mathcal{F}_n] = f(X_n) + \mathbb{E}[(Lf)(X_n) \,|\, \mathcal{F}_n] = f(X_n) + (Lf)(X_n),$$

from which by choosing indicator functions $f = \mathbb{1}_B$ it follows that

$$\mathbb{P}(X_{n+1} \in B \,|\, \mathcal{F}_n) = \mathbb{1}_B(X_n) + (L\mathbb{1}_B)(X_n) = ((L + \mathrm{Id})(\mathbb{1}_B))(X_n).$$

But the right-hand side is $\sigma(X_n)$-measurable, so (3.0.1) follows by iteration. This shows that (X_n) defines a Markov chain with transition matrix as in (3.3.6) and hence finishes the proof. □

Remark 3.3.4. *Of course we know that generators of Markov chains with countable state space S are of the form*

$$L = P - \mathrm{Id},$$

where P is any stochastic matrix in $[0, 1]^{S \times S}$. As alluded to before, however, in the continuous setting it is oftentimes not so clear if an operator generates a Markov chain, and it is in that setting where the theory from above unfolds its full power.

Remark 3.3.5. *We can generalise the notion of (super- / sub)harmonic functions introduced in Remark 2.0.9. In fact, in the continuous context the $\frac{1}{2}\Delta$ defined on suitable functions from $\mathbb{R}^d$ to $\mathbb{R}$ is the generator of Brownian motion (with its discrete counterpart*

$$\begin{aligned}
\Delta : \mathbb{R}^{\mathbb{Z}^d} &\to \mathbb{R}^{\mathbb{Z}^d}, \\
f &\mapsto \Big(\frac{1}{2d} \sum_{\substack{y \in \mathbb{Z}^d, \\ \|x-y\|_1 = 1}} \big(f(y) - f(x)\big) \Big)_{x \in \mathbb{Z}^d}
\end{aligned}$$

being the generator of discrete time simple random walk). Therefore, in the same way that we were able to define harmonic functions for general Markov chains in [Dre18], we can generalise

[6] I.e., L can be interpreted as a matrix in $\mathbb{R}^{S \times S}$ taking a function $f : S \to \mathbb{R}$ to $(Lf)(x) := \sum_{y \in S} L(x, y) f(y)$.

the notion of sub- / superharmonic functions. Indeed, a function $f : S \to \mathbb{R}$ is called harmonic, subharmonic, or superharmonic *(on $U \subset S$) with respect to a generator L, if*

$$Lf = 0, \quad Lf \geq 0, \quad \text{or} \quad Lf \leq 0,$$

respectively, where all (in-)equalities are to be understood as functions on S (or as functions on U only).
Theorem 3.3.3 again supplies us with the fact that for a Markov chain (X_n) with generator L we have that, independently of the initial condition, $(f(X_n))$ is a martingale, submartingale, or supermartingale, respectively, if (and essentially only if) f is harmonic, subharmonic, or superharmonic with respect to L, respectively.

We now turn to some application of the above.

Exercise 3.3.6. *The details of the following are left to the reader.*

(a) *Show that for $d \geq 3$, simple random walk is transient, i.e., that*

$$\mathbb{P}_x(\tau_0 < \infty) < 1 \tag{3.3.7}$$

for all $x \in \mathbb{Z}^d$. For this purpose, we first observe that it is sufficient to show that (3.3.7) holds true for all x with $\|x\|$ large enough.

(b) *Show that simple random walk in $\mathbb{Z}^2$ is recurrent, i.e.,*

$$\mathbb{P}_x(\tau_0 < \infty) = 1.$$

3.4 Doob's h-transform

An important tool for investigating Markov processes is the so-called h-transform which we will introduce below. The motivation mainly is to investigate Markov processes conditioned on first hitting some set in a certain subset – what can we say about the resulting process?
For the sake of simplicity we stick to the setting of Markov processes with countable state space S and discrete time $\mathbb{N}_0$.
Now choose some $x \in S$ as well as some $A, B \subset S$ with $A \subset B$. What can we say about the process (X_n) under the measure $\mathbb{P}_x^h := \mathbb{P}_x(\cdot \,|\, X_{H_B} \in A)$? While this conditioned process will still be a process in the sense that the sequences (X_n) are well-defined under this probability measure, it is a priori not clear what the law of this conditioned process is, nor if it will still enjoy the Markov property.
Also, in order to facilitate things notationally, we will consider the modification $(\widetilde{X}_n)$ of (X_n) for which for each $y \in B$, the transition probabilities are changed to

$$\widetilde{P}(y, z) := \mathbb{1}_{y=z},$$

i.e., B plays the role of a so-called *cemetery*, and all other transition probabilities are left unchanged; all corresponding quantities for this modified process will carry a $\widetilde{\ }$ on them.
Introducing the function

$$h : S \ni x \mapsto \mathbb{P}_x(\widetilde{\tau}_B = \widetilde{\tau}_A),$$

the probability measure $\mathbb{P}_x^h$ has a density with respect to the standard probability measure $\mathbb{P}_x$ which can be written as

$$\frac{\mathrm{d}\mathbb{P}_x^h}{\mathrm{d}\mathbb{P}_x} = \frac{\mathbb{1}_{\widetilde{\tau}_B = \widetilde{\tau}_A}}{h(x)}.$$

Note that for $y \in S$ we have

$$h(y) = \mathbb{P}_y(\widetilde{\tau}_A = \widetilde{\tau}_B) = \mathbb{E}_y\big[\mathbb{E}_y[\mathbb{1}_{\widetilde{\tau}_A = \widetilde{\tau}_B} \,|\, \widetilde{X}_1]\big] = \mathbb{E}_y[h(\widetilde{X}_1)] = (\widetilde{P}h)(y),$$

so the function h is harmonic with respect to $\widetilde{L} = \widetilde{P} - \mathrm{Id}$ (even on S due to our modification, and otherwise the last equality in (3.4.1) below would generally not be true anymore). Therefore, the density is well-behaved under conditioning in the sense that

$$\mathbb{E}_x\Big[\frac{\mathrm{d}\mathbb{P}_x^h}{\mathrm{d}\mathbb{P}_x} \,\Big|\, \widetilde{\mathcal{F}}_n\Big] = \frac{\mathbb{E}_x[\mathbb{1}_{\widetilde{\tau}_B = \widetilde{\tau}_A} \,|\, \widetilde{\mathcal{F}}_n]}{h(x)} = \frac{\mathbb{P}_x(\widetilde{\tau}_B = \widetilde{\tau}_A \,|\, \widetilde{X}_n)}{h(x)} = \frac{h(\widetilde{X}_n)}{h(x)}, \tag{3.4.1}$$

and where we have taken advantage of the harmonicity of h with respect to $\widetilde{L}$, as well as (of a consequence) of the Markov property of $(\widetilde{X}_n)$.
This again implies that for any $f : S \to \mathbb{R}$ bounded we obtain, using that $(\widetilde{X}_n)$ has the Markov property with respect to $\mathbb{P}_x$, that

$$\mathbb{E}_x^h[f(\widetilde{X}_{n+1}) \,|\, \widetilde{\mathcal{F}}_n] = \mathbb{E}_x\Big[f(\widetilde{X}_{n+1})\frac{h(\widetilde{X}_{n+1})}{h(x)} \,|\, \widetilde{\mathcal{F}}_n\Big] = \mathbb{E}_x\Big[f(\widetilde{X}_{n+1})\frac{h(\widetilde{X}_{n+1})}{h(x)} \,|\, \widetilde{X}_n\Big],$$

and so we infer that $(\widetilde{X}_n)$ has the Markov property under $\mathbb{P}_x^h$ also. Plugging in indicator functions, we infer that it has transition probabilities given by

$$\mathbb{P}_x^h(\widetilde{X}_{n+1} = z \,|\, \widetilde{X}_n = y) = \frac{h(z)}{h(y)}\widetilde{P}(y,z), \tag{3.4.2}$$

for all $y, z \in S$ such that $\mathbb{P}_x^h(\widetilde{X}_n = y) > 0$. In particular, for $y \in S \backslash B$, the transition probability to $z \in S$ is given by

$$\frac{h(z)}{h(y)}P(y,z).$$

Also note that the above construction did not depend on the explicit nature of the function h, but only on the fact that h was harmonic with respect to $\widetilde{P}$. This leads us to the following definition.

Definition 3.4.1. *Let P be the transition matrix of a Markov chain and let $h : S \to [0,\infty)$ be a function which is harmonic with respect to P. Then* Doob's h-transform *is the Markov chain with transition probabilities P^h given by*

$$P^h(y,z) = \frac{h(z)}{h(y)}\widetilde{P}(y,z), \quad y,z \in S.$$

Remark 3.4.2. *It is possible to define P^h also directly for a harmonic h and transition matrix P, without resorting to $\widetilde{P}$. In this case, we set*

$$P^h(y,z) = \begin{cases} \frac{h(z)}{h(y)}P(y,z), & \text{if } h(y) > 0 \\ \mathbb{1}_{y=z}, & \text{otherwise} \end{cases}; \qquad y,z \in S.$$

Note that P^h defined in this way agrees with the one from Definition 3.4.1.

Example 3.4.3 (Gambler's ruin)**.** *Consider simple random walk on $\{0,1,\dots,M\}$, some $M \in \mathbb{N}$, starting in $x \in \{1,\dots,M-1\}$. In Section 2.0.2 of [Dre18] (using harmonic functions) as well as in Example 2.1.15 (using martingales) we found that the probability of reaching M before 0 when starting in i was given by i/M. We now want to obtain a more profound understanding of* the process itself *(instead of just the hitting probabilities) by investigating the law of this simple random walk conditioned on $H_{\{0,M\}} = H_M$.*

The above derivation tells us that this process again has the Markov property, so it has a transition matrix. The adequate function is given by

$$h(x) := \mathbb{P}_x(\tau_{\{0,M\}} = \tau_M), \quad x \in \{0, 1, \dots, M\},$$

which is harmonic and positive on $\{1, \dots, M\}$*. Using* $h(0) = 0$ *and* $h(M) = 1$ *it's not hard to see that in fact*

$$h(i) = \frac{i}{M}, \quad i \in \{0, \dots, M\}.$$

Hence, from (3.4.2) we infer that the transition probabilities of this h*-transform are given by*

$$P^h(x, x \pm 1) = \frac{1}{2}\frac{h(x \pm 1)}{h(x)} = \frac{1}{2} \pm \frac{1}{2x}, \quad x \in \{1, \dots, M-1\}.$$

Lemma 3.4.4. *Let* (X_n)*,* $n \in \mathbb{N}_0$*, be a Markov process with countable state space* S *and transition matrix* P*. Let* $h : S \to [0, \infty)$ *be a function which is harmonic on* $\{x \in S : h(x) > 0\}$ *with respect to* $P - \mathrm{Id}$*, and denote by*

$$F := \big\{x \in S : \exists y \in S \text{ such that } h(y) = 0 \text{ and } P(x, y) > 0\big\}$$

the set of points in S *from which the process can jump to an element of the zero set of* h*. Then, for* $x \in S$ *with* $h(x) > 0$*, under* $\mathbb{P}_x^h$ *the process*

$$\frac{1}{h(X_{n \wedge \tau_F})}$$

is a martingale with respect to the canonical filtration $(\mathcal{F}_n)$ *of the process* $(X_{n\wedge\tau_F})$*,* $n \in [0, \infty)$*.*

Proof. We have for $n > 0$ that

$$\begin{aligned}
\mathbb{E}_x^h\Big[\frac{1}{h(X_{(n+1)\wedge\tau_F})}\,\Big|\,\mathcal{F}_n\Big] &= \sum_{y \in S\,:\,h(y)>0} \frac{P^h(X_{n\wedge\tau_F}, y)}{h(y)} = \sum_{y \in S\,:\,h(y)>0} \frac{P(X_{n\wedge\tau_F}, y)h(y)}{h(y)h(X_{n\wedge\tau_F})} \\
&= \sum_{y \in S\,:\,h(y)>0} \frac{P(X_{n\wedge\tau_F}, y)}{h(X_{n\wedge\tau_F})} = \frac{\mathbb{P}(h(X_{(n+1)\wedge\tau_F}) > 0)}{h(X_{n\wedge\tau_F})} = \frac{1}{h(X_{n\wedge\tau_F})},
\end{aligned}$$

which completes the proof. □

Example 3.4.5. *As in Example 3.4.3 let* (X_n) *denote one-dimensional simple random walk, and denote by* $\mathbb{P}_x^h$*,* $x \in \mathbb{Z}\backslash\{0\}$*, the Doob transform under which* $\mathbb{P}_x^h(X_0 = x) = 1$*, and under which* (X_n) *is simple random walk conditioned on not hitting* 0 *(corresponding to taking* $M \to \infty$*, which is OK since we've seen that the transition probabilities actually do not depend on* M*, so we can use an existence theorem of our choice (Ionescu-Tulcea, Kolmogorov) for this process). Then, for* $x, y, z \in \mathbb{N}$ *with* $1 < x < y < z$ *the following hold true:*

(a)
$$\mathbb{P}_y^h(\tau_x < \tau_z) = \frac{x(z-y)}{y(z-x)};$$

(b)
$$\mathbb{P}_y^h(\tau_x < \infty) = \frac{x}{y};$$

(c)
$$\mathbb{P}_x^h(\nexists n \in \mathbb{N} : X_n = x) = \frac{1}{2x}.$$

Proof. Left to the reader. □

Chapter 4

Stationarity & ergodicity

We have seen that the law of large numbers is one of the most important limit theorems we have gotten to know so far (along with the central limit theorem). The goal of this section is to extend this result significantly – in particular, we will be able to cover situations where implications similar to that of the strong law of large numbers are obtained in regimes where we do have certain dependence between the respective random variables. See Remark 4.1.2 below also for a embedding ergodicity in its historic context.

Definition 4.0.1. *Let (X_t), $t \in T$, with $T = \mathbb{N}, \mathbb{N}_0, \mathbb{Z}$, or $T = \mathbb{R}$,*[1] *be a stochastic process taking values in a Polish space $(S, \mathcal{O})$ with Borel σ-algebra $\mathcal{B}(S)$. Then if for all $s \in T$ we have that for all $t_1 < \ldots < t_n$,*

$$\mathbb{P}((X_{t_1}, \ldots, X_{t_n}) \in \cdot) = \mathbb{P}((X_{t_1+s}, \ldots, X_{t_n+s}) \in \cdot)$$

as measures on $(S^n, \mathcal{B}(S^n))$, the stochastic process (X_t), $t \in T$, is called time homogeneous *or* stationary.

Remark 4.0.2. *In the setting of Definition 4.0.1, a stochastic process is stationary if and only if for all $s \in T$, the processes (X_t), $t \in T$, and (X_{t+s}), $t \in T$, have the same law on $(S^T, \mathcal{B}(S)^{\otimes T})$.*

Exercise 4.0.3. *In the setting of Definition 4.0.1, find an example which shows that in order for the process (X_t) to be stationary it is not sufficient to have that the one-dimensional distributions of time shifts coincide, i.e., that $\mathbb{P} \circ X_t^{-1} = \mathbb{P} \circ X_{t+s}^{-1}$ for all $s \in T$.*

Example 4.0.4. *(a)*

You might have seen in Proposition 1.17.11 of [Dre18] that every Markov chain on a finite state space S has a stationary distribution, and in Proposition 1.17.15 of [Dre18] that if the chain is irreducible (see [Dre18]), then the stationary distribution is unique. Let π be such a distribution and denote the transition matrix of the chain by P. Then if we start the chain in the initial distribution π, then for the resulting stochastic process (X_n) we have that (4.0.1) holds true.

(b) Recall that the Ornstein-Uhlenbeck process as defined in the exercise classes. As we did before in the construction of Markov processes, instead of starting the process in a deterministic point, we could also start it from an initial distribution $\mathcal{N}(0, \frac{\sigma^2}{2\theta})$. One can show that with this initial distribution, the resulting process is a stationary process and fulfills (4.0.1).

(c) Let (X_n), $n \in \mathbb{N}_0$, be an i.i.d. sequence of real random variables and for $m \in \mathbb{N}$ fixed as well as $c_0, \ldots, c_m \in \mathbb{R}$ fixed, define the sequence $Y_n := \sum_{i=0}^m c_i X_{n+i}$. Show that (Y_n) is a stationary process ('moving average process').

[1] More generally, any monoid $T \subset \mathbb{R}$ with respect to addition would work here, and this is what we assume for the rest of this chapter.

(d) *Show that Brownian motion is not stationary, no matter which initial distribution you choose.*

The concept of stationary processes can be generalised to more general settings as follows.

Definition 4.0.5. *Let $(\Omega, \mathcal{F}, \mu)$ be σ-finite measure space. An $\mathcal{F}-\mathcal{F}$-measurable mapping $\theta : \Omega \to \Omega$ is called a* measure-preserving transformation ('maßerhaltende Transformation') *if $\mu = \mu \circ \theta^{-1}$.*
In this setting, μ is also called an invariant measure *for θ.*

Example 4.0.6. (a) *Given the distribution $\mathbb{P}$ on $(S^{\mathbb{N}_0}, \mathcal{B}(S)^{\otimes \mathbb{N}_0})$ of a stationary process, the coordinate shift $\theta : (x_n)_{n\in\mathbb{N}_0} \mapsto (x_{n+1})_{n\in\mathbb{N}_0}$ is a measure-preserving transformation on $(S^{\mathbb{N}_0}, \mathcal{B}(S)^{\otimes \mathbb{N}_0})$.*

Certainly θ is a map from $S^{\mathbb{N}_0}$ onto itself, and it's measurable since for an arbitrary cylinder set $A \in \mathcal{B}(S)^{\otimes \mathbb{N}_0}$ we have $\theta^{-1}(A) = S \times A$ which is a cylinder set again, and hence it's in $\mathcal{B}(S)^{\otimes \mathbb{N}_0}$. But the cylinder sets generate $\mathcal{B}(S)^{\otimes \mathbb{N}_0}$ and hence the desired measurability of θ follows from Theorem 1.4.8 of [Dre19].

Similarly, for a cylinder set $A \in \mathcal{B}(S)^{\otimes \mathbb{N}_0}$ we have $\mathbb{P}(A) = \mathbb{P} \circ \theta^{-1}(A)$ (since the process is stationary), hence the two measures coincide on a $\cap$-stable generator of $\mathcal{B}(S)^{\otimes \mathbb{N}_0}$ and therefore coincide due to Dynkin's π-λ Theorem. Thus, θ is measure-preserving.

(b) *Consider the space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda^d)$. Show that for any $y \in \mathbb{R}^d$, the shift*

$$\begin{aligned} \theta_y : \mathbb{R}^d &\to \mathbb{R}^d, \\ x &\mapsto x + y, \end{aligned}$$

is a measure-preserving transformation.

(c) *Let (x_n), $n \in \mathbb{N}_0$, be an S-valued sequence with period $p \in \mathbb{N}$, i.e., $x_n = x_{n+p}$ for all $n \in \mathbb{N}_0$. Define the discrete probability measure*

$$\mathbb{P} := \frac{1}{p} \sum_{k=0}^{p-1} \delta_{(x_{n+k})_{n\in\mathbb{N}_0}}.$$

Then $\mathbb{P}$ is invariant under the shift $\theta : S^{\mathbb{N}_0} \to S^{\mathbb{N}_0}$ which maps $(x_n)_{n\in\mathbb{N}_0}$ to $(x_{n+1})_{n\in\mathbb{N}_0}$.

(d) *Consider the unit circle $\mathbb{S}^1 = \{x \in \mathbb{C} : |x| = 1\}$, which can be reparametrised as $\mathbb{S}^1 = \{e^{\mathrm{i}\alpha} : \alpha \in [0, 2\pi)\}$. For an angle $\beta \in \mathbb{R}$ we can define the rotation by the angle β via*

$$\theta_\beta : e^{\mathrm{i}\alpha} \mapsto e^{\mathrm{i}(\alpha+\beta)}.$$

Show that this defines a measure-preserving transformation on the space $(\mathbb{S}^1, \mathcal{B}(S^1), \mu)$, with μ denoting the uniform measure on $(\mathbb{S}^1, \mathcal{B}(S^1))$ normalised to $\mu(\mathbb{S}^1) = 1$.

(e) *Consider the following map θ from $([0,1], \mathcal{B}([0,1]), \lambda_{|[0,1]})$ onto itself:*

$$\theta(x) := \begin{cases} 2x, & \text{if } x \in [0, \frac{1}{2}), \\ 2 - 2x, & \text{if } x \in [\frac{1}{2}, 1] \end{cases}$$

The map is called 'baker's map' since if you visualise how the map is acting on $[0,1]$ this is slightly reminiscent of kneading dough.

Show that the map is measure-preserving.

The following definition will most probably not sound very useful until we get to know Theorem 4.1.1 below.

Definition 4.0.7. *Let θ be a measure-preserving transformation on $(\Omega, \mathcal{F}, \mu)$. The σ-algebra*

$$\mathcal{I} := \{F \in \mathcal{F} : \theta^{-1}(F) = F\}$$

is called the σ-algebra of θ-invariant sets*.*
A measure-preserving transformation on $(\Omega, \mathcal{F}, \mu)$ is called ergodic ('ergodisch')*, if for all $F \in \mathcal{I}$, either $\mu(F) = 0$ or $\mu(\Omega \backslash F) = 0$.*

Remark 4.0.8. *It is immediate that if μ is a probability measure, then the transformation is ergodic if and only if $\mu(F) \in \{0, 1\}$ for all $F \in \mathcal{I}$.*

Exercise 4.0.9. *(a) Check that $\mathcal{I}$ of Definition 4.0.7 defines a σ-algebra indeed.*

(b) Let $n \in \mathbb{N}$ with $n \geq 2$. Show that the ergodicity of a transformation θ does not imply the ergodicity of θ^n, $n \geq 2$. Does the ergodicity of θ follow from the ergodicity of θ^n? Find proofs or counterexamples!

Since we're studying probability theory, the following notion is sufficiently important to deserve its own definition.

Definition 4.0.10. *In the setting of Example 4.0.6 (a), if the underlying stochastic process (X_t) is stationary (i.e., the shift operator θ is measure-preserving on the given probability space), then the process is called* ergodic *if and only if the shift operator θ is ergodic.*

Remark 4.0.11. *According to the previous definition, a stochastic process in particular has to be stationary in order for it to possibly be ergodic. As mentioned in Example 4.0.4, if you start a Markov chain in equilibrium this in particular supplies us with a stationary chain, i.e., the time shift is measure-preserving. Especially in the context of Markov chains, however, you will encounter chains (X_n) called 'ergodic', but which do not even have the same distributions at different times $\mathbb{P} \circ X_n^{-1} = \mathbb{P} \circ X_{n+m}^{-1}$ for all $m, n \in \mathbb{N}_0$.*
Nevertheless, these chains are sometimes called ergodic if the implication of the (say, individual) ergodic theorem still hold true: For any initial distribution of the chain $\mathbb{P}$-a.s.,

$$\frac{1}{n} \sum_{i=0}^{n-1} f(X_n) \to \sum_{x \in S} \pi(x) f(x),$$

for any (bounded) function $f : S \to \mathbb{R}$, and with π denoting the unique stationary distribution of the chain. (see e.g. Theorem 4.7 in [LPW09]). This means that even though the chain is neither stationary nor ergodic, we still have the mantra that spatial averages equal asymptotic time averages, and this holds true e.g. for

Exercise 4.0.12. *Determine which of the measure-preserving shifts defined in Example 4.0.6 (b) are ergodic (for some of them that is not completely easy and you might want to take advantage of Lemma 4.0.15 below to show it).*

In order to understand the above notions a bit better, we give the following auxiliary result.

Lemma 4.0.13.
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let θ be a transformation on Ω which is measure-preserving.

(a) A real $\mathcal{F} - \mathcal{B}$-measurable function $f : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ is $\mathcal{I} - \mathcal{B}$-measurable if and only if $f \circ \theta = f$.

(b) The transformation θ is ergodic if and only if any $\mathcal{I} - \mathcal{B}$-measurable function $f : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ is $\mathbb{P}$-a.s. constant.

Proof. (a) Left as an exercise.

(b) Let θ be ergodic, and f a $\mathcal{I} - \mathcal{B}$-measurable function. Then for each $s \in \mathbb{R}$ we have that

$$\mathbb{P}(f^{-1}((-\infty, s))) \in \{0, 1\},$$

and since the left-hand side is monotone increasing in s with $\lim_{s\downarrow -\infty} \mathbb{P}(f^{-1}((-\infty, s))) = 0$ and $\lim_{s\uparrow\infty} \mathbb{P}(f^{-1}((-\infty, s))) = 1$ we get the $\mathbb{P}$-a.s. equality

$$f = \sup\{s \in \mathbb{R} : \mathbb{P}(f^{-1}((-\infty, s))) = 0\} \in \mathbb{R}.$$

Conversely, if any $\mathcal{I} - \mathcal{B}$-measurable function is $\mathbb{P}$-a.s. constant, then in particular, for any $F \in \mathcal{I}$ we have $\mathbb{1}_F$ is $\mathbb{P}$-a.s. constant, which implies $\mathbb{P}(F) \in \{0, 1\}$. Since $F \in \mathcal{I}$ was chosen arbitrarily, this completes the proof.

□

We can introduce another pretty general and sufficient criterion for a shift to be ergodic. While there are in fact different notions of 'mixing' around, we will just introduce the following.

Definition 4.0.14. *Let θ be a measure-preserving transformation on $(\Omega, \mathcal{F}, \mathbb{P})$. Then θ is called* mixing *if for all $F, G \in \mathcal{F}$ we have*

$$\mathbb{P}(F \cap \theta^{-n}(G)) \to \mathbb{P}(F)\mathbb{P}(G) \qquad \text{as } n \to \infty. \tag{4.0.1}$$

Lemma 4.0.15. *Let θ be a measure-preserving transformation on $(\Omega, \mathcal{F}, \mathbb{P})$. Then if θ is mixing, it is also ergodic.*

Proof. Let $F \in \mathcal{I}$, with $\mathcal{I}$ the σ-algebra of θ-invariant elements of $\mathcal{F}$. We have

$$\mathbb{P}(F) = \mathbb{P}(F \cap \theta^{-n}(F)) \to \mathbb{P}(F)^2,$$

so $\mathbb{P}(F) = \mathbb{P}(F)^2$ which implies $\mathbb{P}(F) \in \{0, 1\}$, so θ is ergodic with respect to $\mathbb{P}$. □

Example 4.0.16. *(a) In the same way that we transferred the property of ergodicity from transformations to processes (cf. Example 4.0.6 (a)), we can do so for the property of mixing. Show that the moving average process from Example 4.0.4 (c) is mixing (hint: show that (4.0.1) holds true for F, G cylinder events, and then reason that this is sufficient to deduce the mixing property).*

4.1 Ergodic theorems

See [Var01, Dud02, Str11] for more analytically flavoured approaches to the following ergodic theorems.

Theorem 4.1.1 (Individual ergodic theorem, Birkhoff's ergodic theorem)**.**
Let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$, and let θ be a measure-preserving transformation on $(\Omega, \mathcal{F}, \mathbb{P})$. Then $\mathbb{P}$-a.s.,

$$\frac{1}{n}\sum_{k=0}^{n-1} X \circ \theta^k \to \mathbb{E}[X \,|\, \mathcal{I}]. \tag{4.1.1}$$

Remark 4.1.2. *(a) The theorem is oftentimes applied to random variables $f \circ X$, where X is a random variable and f a real valued function such that $f \circ X$ satisfies the assumptions of the theorem (consider $X \in \mathcal{L}^2$ and $f(x) := x^2$ as an example). In this case (4.1.1) becomes*

$$\frac{1}{n}\sum_{k=0}^{n-1} f(X \circ \theta^k) \to \mathbb{E}[f(X) \,|\, \mathcal{I}].$$

(b) *There's a version of Theorem 4.1.1 for σ-finite measure spaces also (see e.g. [Dud02] or [Str11]), but we'll stick to the setting of probability spaces for simplicity.*

(c) *While a general intuition for the law of large numbers had been around for some time even before a rigorous proof of the result had been found, ergodicity is a bit harder to grasp. Indeed, it arguably first appeared in the incipient works on statistical mechanics by Boltzmann, notably [Bol68]. The guiding interpretation has been that the 'probability' of finding a particle at a certain part of the state space*[2] *should be given the normalized long time local time – i.e., the relative amount of time a particle has spent in that area in a long-term observation. Hence, ergodicity is oftentimes also phrased as 'time average equals ensemble average (i.e., expectation)'.*

Mathematically, ergodic theory only developed from the beginning of the 20th century onwards

(d) *Theorem 4.1.1 can be used to give yet another proof of recurrence of (simple) random walk (in $\mathrm{d} = 1$) – since we have seen a variety of different proofs of this fact already, we refer to [Kle14, Section 20.4] for further details.*

(e) *It is immediate that combining Example 4.0.6 (a) with Theorem 4.1.1 we obtain our third proof of the law of large numbers after Theorem 3.6.9 in [Dre19] and Exercise 2.4.5. The version we can derive here, like Exercise 2.4.5, only needs first moments of X_n.*

Most proofs of this result proceed via the so-called *maximal ergodic lemma*, which is the following result.

Lemma 4.1.3 (Hopf's maximal ergodic lemma)**.**
Let $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, and let θ be a measure-preserving transformation on $(\Omega, \mathcal{F}, \mathbb{P})$. Writing

$$S_n := \sum_{i=0}^{n-1} X \circ \theta^i, \quad n \geq 1,$$

$S_0 := 0$,

$$M_n := \max_{i=0}^{n} S_i$$

and introducing the set

$$P_n := \{M_n > 0\},$$

we have

$$\mathbb{E}[X \mathbb{1}_{P_n}] \geq 0.$$

Proof. We start with observing that

$$S_{j+1} = X + S_j \circ \theta, \quad j \geq 1,$$

hence for $j \in \{1, \ldots, n\}$ we have

$$S_j = X + S_{j-1} \circ \theta \leq X + M_n \circ \theta. \tag{4.1.2}$$

Since furthermore on P_n we have

$$M_n = \max_{j=1}^{n} S_j,$$

we infer, taking the maximum over $j \in \{1, \ldots, n\}$ in (4.1.2) that

$$M_n \leq X + M_n \circ \theta \quad \text{on } P_n.$$

[2]Indeed, as we will also be investigating in the construction of Brownian motion below, there is a priori nothing intrinsically random in the motion of a particle, say, molecule – however, a computation of its trajectory on the basis of the determining parameters is oftentimes too complex.

This implies the first inequality in the following line of (in-)equalities:

$$\int_{P_n} X \, d\mathbb{P} \geq \int_{P_n} (M_n - M_n \circ \theta) \, d\mathbb{P} = \int_\Omega M_n \, d\mathbb{P} - \int_{P_n} M_n \circ \theta \, d\mathbb{P} \geq \int_\Omega M_n \, d\mathbb{P} - \int_\Omega M_n \circ \theta \, d\mathbb{P} = 0,$$

where the last inequality follows since $M_n \circ \theta \geq 0$, and the last equality follows since θ is measure-preserving. □

We can now prove the ergodic theorem, and we follow the standard proof these days.

Proof of Theorem 4.1.1. The first step is to observe that we can replace X by $X - \mathbb{E}[X \mid \mathcal{I}]$, and hence assume w.l.o.g. that

$$\mathbb{E}[X \mid \mathcal{I}] = 0; \tag{4.1.3}$$

so it boils down to showing

$$\frac{1}{n} \sum_{k=0}^{n-1} X \circ \theta^k \to 0, \tag{4.1.4}$$

and for brevity we introduce $X_n := X \circ \theta^n$ and $S_n := \sum_{k=0}^{n-1} X_k$. We want to show that for each $\varepsilon > 0$,

$$\mathbb{P}\Big(\underbrace{\limsup_{n\to\infty} \frac{1}{n} S_n > \varepsilon}_{=:L_\varepsilon} \Big) = 0, \tag{4.1.5}$$

and replacing X_k by $-X_k$ we would obtain (4.1.4).
Introduce the random variables

$$X_k^\varepsilon := (X_k - \varepsilon) \mathbb{1}_{L_\varepsilon}, \quad k \in \mathbb{N}_0$$

as well as

$$S_n^\varepsilon := \sum_{k=0}^{n-1} X_k^\varepsilon.$$

Observing that $L_\varepsilon = \theta^{-1}(L_\varepsilon)$, we infer that $L_\varepsilon \in \mathcal{I}$, and as a consequence, (X_k^ε) is a stationary process.
Using the notation $M_n^\varepsilon := \max_{i=0}^n S_i^\varepsilon$, with $S_0^\varepsilon := 0$ we apply the maximal ergodic lemma (Lemma 4.1.3) to obtain

$$0 \leq \mathbb{E}[X_0^\varepsilon \mathbb{1}_{M_n^\varepsilon > 0}],$$

for each $n \in \mathbb{N}$. Using the dominated convergence theorem, the right-hand side then converges to $\mathbb{E}[X_0^\varepsilon \mathbb{1}_{\sup_{n\in\mathbb{N}} S_n^\varepsilon > 0}]$.
But now

$$\{\sup_{n\in\mathbb{N}} S_n^\varepsilon > 0\} = \{\sup_{n\in\mathbb{N}} S_n > \varepsilon\} \cap L_\varepsilon = L_\varepsilon,$$

and hence we infer with the above that

$$0 \leq \mathbb{E}[(X_0 - \varepsilon) \cdot \mathbb{1}_{L_\varepsilon}] = \mathbb{E}[\mathbb{1}_{L_\varepsilon} \underbrace{\mathbb{E}[X \mid \mathcal{I}]}_{\overset{(4.1.3)}{=} 0}] - \varepsilon \mathbb{P}(L_\varepsilon) = -\varepsilon \mathbb{P}(L_\varepsilon).$$

This entails $\mathbb{P}(L_\varepsilon) = 0$, and therefore (4.1.5) and (4.1.4) follow. □

Theorem 4.1.4 (Mean ergodic theorem). *Let $p \in [1, \infty)$, $X \in \mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$, and let θ be a measure-preserving shift on $(\Omega, \mathcal{F}, \mathbb{P})$. Then*

$$\frac{1}{n} \sum_{k=0}^{n-1} X \circ \theta^k \to \mathbb{E}[X \mid \mathcal{I}] \quad \text{in } L^p(\Omega, \mathcal{F}, \mathbb{P}). \tag{4.1.6}$$

Birkhoff's individual ergodic theorem supplies us with the $\mathbb{P}$-a.s. convergence. Therefore, recalling our results for uniformly integrable families of random variables, we infer that it's sufficient to show that the family of random variables occurring on the left-hand side of (4.1.4), raised to their p-th powers, is uniformly integrable.

Lemma 4.1.5. *Let $p \in [1,\infty)$ as well as (X_n) a sequence of identically distributed but not necessarily independent real random variables with $X_n \in L^p$. Then setting*

$$Y_n := \Big|\frac{1}{n}\sum_{k=0}^{n-1} X_k\Big|^p,$$

the sequence (Y_n), $n \in \mathbb{N}_0$, is uniformly integrable.

Proof. By Theorem 2.3.6 there exists $\varphi : [0,\infty) \to [0,\infty)$ non-decreasing and convex with $\varphi(x)/x \to \infty$ as $x \to \infty$, and such that

$$C_0 := \mathbb{E}[\varphi(|X_0|^p)] < \infty. \tag{4.1.7}$$

We then obtain by applying Jensen's inequality twice that

$$\varphi(Y_n) \le \varphi\Big(\frac{1}{n}\sum_{k=0}^{n-1} |X_k|^p\Big) \le \frac{1}{n}\sum_{k=0}^{n-1} \varphi(|X_k|^p);$$

indeed, for both inequalities we took the average $\frac{1}{n}\sum_{k=0}^{n-1}\dots$ as expectation, and for the first inequality we applied Jensen's inequality with the convex function $x \mapsto |x|^p$, whereas for the second we used the convexity of φ.
Since the X_k are identically distributed, in combination with (4.1.7) this implies that for each $n \in \mathbb{N}_0$ we have

$$\mathbb{E}[\varphi(Y_n)] \le C_0,$$

which again due to Theorem 2.3.6 yields the desired uniform integrability. □

Proof of Theorem 4.1.4. We take advantage of Lemma 4.1.5 in order to deduce that, since θ is measure-preserving and thus all the $X \circ \theta^k$, $k \in \mathbb{N}_0$, are identically distributed, the family of random variables defined via

$$Z_n := \Big|\frac{1}{n}\sum_{k=0}^{n-1} X \circ \theta^k - \mathbb{E}[X \,|\, \mathcal{I}]\Big|^p$$

is uniformly integrable. Since Birkhoff's ergodic theorem implies that $Z_n \to 0$ holds $\mathbb{P}$-a.s. as $n \to \infty$, Theorem 2.3.9 implies that $Z_n \to 0$ in L^1 as well, which entails the desired L^p-convergence. □

Remark 4.1.6. *In the important case of a transformation θ being ergodic, Lemma 4.0.13 (b) implies that $\mathbb{E}[X] = \mathbb{E}[X \,|\, \mathcal{I}]$. Therefore, in the settings of Theorems 4.1.1 and 4.1.4 we have convergence of*

$$\frac{1}{n}\sum_{i=0}^{n-1} X \circ \theta^i$$

to $\mathbb{E}[X]$ $\mathbb{P}$-a.s., and in L^p, respectively.

The following example gives a nice connection between probability theory, number theory, and applications.

Example 4.1.7. *Benford's law has even been used in e.g. in the context of detection tax evasion, see* `http://search.proquest.com/docview/211023799?pq-origsite=gscholar`. *We refer to [Kle14, p. 447] for a short version and to [Ben38] as well as to [BH+11] for longer bedtime stories.*

4.1.1 Subadditive ergodic theorem

A substantial enhancement of Birkhoff's ergodic theorem given in the previous section is the subadditive ergodic theorem. Indeed, in many models inspired by real-world applications one does not have stationarity of the corresponding sequence of random variables, but some weaker properties instead. Nevertheless, under some suitable assumptions one can still derive a result in the spirit of the ergodic theorems we investigated before.

Definition 4.1.8. *A sequence* $(x_n)_{n\in\mathbb{N}}$ *of real numbers is called* subadditive *if*

$$x_{n+m} \le x_n + x_m \quad \forall n, m \in \mathbb{N}$$

(and it's called superadditive *if* $(-x_n)$, $n \in \mathbb{N}_0$, *is subadditive).*

Lemma 4.1.9 (Fekete's lemma). *If* $(x_n)_{n\in\mathbb{N}}$ *is a subadditive sequence, then* $\lim_{n\to\infty} \frac{x_n}{n}$ *exists in* $\mathbb{R} \cup \{-\infty\}$ *and*

$$\lim_{n\to\infty} \frac{x_n}{n} = \inf_{n\in\mathbb{N}} \frac{x_n}{n} \tag{4.1.8}$$

Proof. Once we have proven (4.1.8), the fact that the limit is an element of $\mathbb{R} \cup \{-\infty\}$ is immediate.
It is clear that

$$\liminf_{n\to\infty} \frac{x_n}{n} \ge \inf_{n\in\mathbb{N}} \frac{x_n}{n}$$

holds, so in order to show the converse inequality, let $\varepsilon > 0$ be arbitrary and let $n_\varepsilon \in \mathbb{N}$ such that

$$\frac{x_{n_\varepsilon}}{n_\varepsilon} \le \inf_{n\in\mathbb{N}} \frac{x_n}{n} + \varepsilon. \tag{4.1.9}$$

For $n > n_\varepsilon$ write $n = kn_\varepsilon + r$ with $r \in \{0, \dots, n_\varepsilon\}$ and $k \in \mathbb{N}$. Then using the subadditivity we infer

$$\frac{x_n}{n} \le \frac{kx_{n_\varepsilon} + x_r}{n} = \frac{kx_{n_\varepsilon}}{n} + \frac{x_r}{n},$$

and taking the lim sup we infer

$$\limsup_{n\to\infty} \frac{x_n}{n} \le \frac{x_{n_\varepsilon}}{n_\varepsilon} \overset{(4.1.9)}{\le} \inf_{n\in\mathbb{N}} \frac{x_n}{n} + \varepsilon.$$

Since $\varepsilon > 0$ was chosen arbitrarily, this proves the result. □

The following result is a stochastic generalisation of Fekete's lemma (Lemma 4.1.9), and it is originally due to Kingman [Kin73] and has been improved to its current form by Liggett, see [Lig85]. In fact, the latter article is a very good read and entirely accessible at your level of probability theory.

Theorem 4.1.10 (Subadditive ergodic theorem). *Let* $(X_{m,n})$ $n \in \mathbb{N}_0$, $0 \le m \le n$, *be an array of real random variables satisfying the following properties:*

(a)

$$X_{0,n} \in \mathcal{L}^1 \quad \forall n \in \mathbb{N},$$

$$\inf_{n\in\mathbb{N}} \frac{\mathbb{E}[X_{0,n}]}{n} > -\infty;$$

(b)

$$X_{0,n} \le X_{0,m} + X_{m,n} \quad \forall n \in \mathbb{N},\ m \in \{1, \dots, n-1\}; \tag{4.1.10}$$

(c) for each $k \in \mathbb{N}$, the sequence $(X_{kn,k(n+1)})$, $n \in \mathbb{N}$, forms a stationary process;

(d) for each $k \in \mathbb{N}_0$, the distribution of the sequence $(X_{k,k+n})$, $n \in \mathbb{N}$, in $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\otimes \mathbb{N}})$ is the same as that of the sequence $(X_{k+1,k+1+n})$, $n \in \mathbb{N}$, in $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\otimes \mathbb{N}})$.

Then:

$$X_\infty := \lim_{n\to\infty} \frac{X_{0,n}}{n} \quad \mathbb{P}\text{-a.s. and in } L^1,$$

with

$$\mathbb{E}[X_\infty] = \lim_{n\to\infty} \frac{1}{n}\mathbb{E}[X_{0,n}] = \inf_{n\to\infty} \frac{1}{n}\mathbb{E}[X_{0,n}].$$

If in addition the stationary processes in assumption (c) are ergodic (i.e., the canonical shift on the product space is ergodic), then X_∞ is $\mathbb{P}$-a.s. constant and equal to $\mathbb{E}[X_\infty]$.

Remark 4.1.11. *(a) A first version of the above result had earlier been proven by Kingman [Kin73], and this is why the above theorem is often referred to as (an improved version / Liggett's version of) Kingman's subadditive ergodic theorem.*

A crucial difference to the above result is that Kingman [Kin73] assumed

$$X_{l,n} \le X_{l,m} + X_{m,n} \quad \forall 0 \le l \le m \le n,$$

instead of (4.1.10). This assumption is strictly stronger than (4.1.10).

(b) In the case of the corresponding transformation θ being ergodic, Theorem 4.1.10 is indeed an improvement over the ergodic theorems we have investigated in the previous section.

Assuming that the conditions of the (individual, or mean for $p = 1$) ergodic theorem are fulfilled, for $m, n \in \mathbb{N}_0$ with $m < n$ set

$$X_{m,n} := \sum_{i=m+1}^{n} X \circ \theta^i.$$

Then

$$X_{0,n} = X_{0,m} + X_{m,n}.$$

Therefore, the subadditive ergodic theorem supplies us with

$$X_\infty = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} X \circ \theta^i \quad \mathbb{P} - a.s. \quad \text{and in } L^1,$$

with $X_\infty = \mathbb{E}[X]$, which is the statement the previous ergodic theorems would supply us with, too.

Proof. The proof can be found in [Lig85] and, although not very short, is not overly complicated either. It takes advantage of Birkhoff's ergodic theorem itself so we omit proving it here. □

We do give an example for its application here which is essentially motivated by current research, it is a simplified version of some results of [DGRS12].

Example 4.1.12 (Existence of quenched Lyapunov exponents in the parabolic Anderson model)**.** *Let $\xi := (\xi(x))_{x\in\mathbb{Z}^d}$ be a real-valued random field which is ergodic with respect to the canonical shifts $\theta_z \xi := (\xi(x+z))_{x\in\mathbb{Z}^d}$, for all $z \in \mathbb{Z}^d$.*[3] *For simplicity assume further that*

[3] Again, in our lingo this means that with our canonical probability space $(\mathbb{R}^{\mathbb{Z}^d}, \mathcal{B}^{\otimes \mathbb{Z}^d}, \mathbb{P})$, with $\mathbb{P}$ the law of $(\xi(x))_{x\in\mathbb{Z}^d}$, the canonical shifts $\theta_z : \mathbb{Z}^d \ni x \mapsto x + z \in \mathbb{Z}^d$ are measure-preserving and ergodic.

$|\xi(0)| \leq A$ $\mathbb{P}$-a.s. some $A > 0$. *The solution of the parabolic Anderson model (PAM) is the solution of the following (random) parabolic difference equation with random potential* ξ:[4]

$$\begin{aligned} \frac{\partial}{\partial t} u(t,x) &= \Delta u(t,x) - \gamma \xi(x)\, u(t,x), \qquad\qquad x \in \mathbb{Z}^d,\ t \geq 0, \\ u(0,x) &= 1, \end{aligned} \tag{4.1.11}$$

where ξ *are as before,* $\Delta f(x) = \frac{1}{2d}\sum_{\|y-x\|=1}(f(y)-f(x))$ *is the discrete Laplacian on* $\mathbb{Z}^d$ *that we have bumped into before and* $\gamma > 0$ *is a constant.*
By the so-called Feynman-Kac formula, *the solution* u *admits the probabilistic representation*

$$u(t,0) = \mathbb{E}_0^X \left[\exp\left\{ -\gamma \int_0^t \xi(X_s)\, \mathrm{d}s \right\} \right],$$

where $(X_s)_{s\in[0,\infty)}$ *is a continuous time simple random walk with jump rate* 1, *which is independent from the potential* $(\xi(z))_{z\in\mathbb{Z}^d}$.[5] *For* $x \in \mathbb{Z}^d$ *let* $\mathbb{P}_x^X$ *and* $\mathbb{E}_x^X$ *denote respectively probability and expectation for a jump rate* 1 *simple symmetric random walk* X, *starting from* x. *For each* $0 \leq t$ *and* $x, y \in \mathbb{Z}^d$, *define*

$$\begin{aligned} e(t,x,y,\xi) &:= \ \mathbb{E}_x^X \left[\exp\left\{ -\gamma \int_0^t \xi(X_u)\, \mathrm{d}u \right\} 1_{\{X_t = y\}} \right], \\ a(t,x,y,\xi) &:= \ -\log e(t,x,y,\xi). \end{aligned} \tag{4.1.12}$$

$a(t,x,y,\xi)$ *is also referred to as the* point to point *passage function from* x *to* y. *We will prove parts of the following so-called* shape theorem *for* $a(t,0,y,\xi)$.

Theorem 4.1.13. **[Shape theorem]** *There exists a deterministic convex function* $\alpha : \mathbb{R}^d \to \mathbb{R}$, *which we call the* shape function, *such that* $\mathbb{P}$*-a.s., for any compact* $K \subset \mathbb{R}^d$,

$$\lim_{t\to\infty} \sup_{y \in tK \cap \mathbb{Z}^d} |t^{-1} a(t,0,y,\xi) - \alpha(y/t)| = 0. \tag{4.1.13}$$

Remark 4.1.14. *The term* shape theorem *is motivated by the fact that* $\alpha + \operatorname{ess\,inf} \xi(0)$ *defines a norm on* $\mathbb{R}^d$ *(check!) and the theorem gives information about the shape of the balls* $\{x \in \mathbb{R}^d : \alpha(x) + \operatorname{ess\,inf} \xi(0) \leq r\}$, $r \in (0,\infty)$, *in this norm.*

More precisely, we will show the following weaker result.

Proposition 4.1.15. *There exists a deterministic function* $\alpha : \mathbb{Q}^d \to [-A, \infty)$ *such that* $\mathbb{P}$*-a.s., for every* $y \in \mathbb{Q}^d$,

$$\lim_{\substack{t\to\infty \\ ty \in \mathbb{Z}^d}} t^{-1} a(t,0,ty,\xi) = \alpha(y). \tag{4.1.14}$$

Proof. Since we assume $y \in \mathbb{Q}^d$ and $ty \in \mathbb{Z}^d$, without loss of generality, it suffices to consider $y \in \mathbb{Z}^d$ and $t \in \mathbb{N}$. Note that by the definition of the passage function a in (4.1.12), $\mathbb{P}$-a.s.,

$$a(t_1 + t_2, x_1, x_3) \leq a(t_1, x_1, x_2) + a(t_2, x_2, x_3) \quad \forall\, t_1 < t_2,\ x_1, x_2, x_3 \in \mathbb{Z}^d.$$

Together with our assumption on the ergodicity of ξ, this implies that the two-parameter family $a(t-s, sy, ty)$, $0 \leq s \leq t$ with $s, t \in \mathbb{N}_0$, satisfies the conditions of the subadditive ergodic theorem (Theorem 4.1.10). Therefore, there exists a deterministic constant $\alpha(y)$ such that (4.1.14) holds. ∎

[4] See [Kï6] for a survey of this model.

[5] So note that we have two kinds of randomness here, one for the random potential ξ and one for the random walk, and these two are supposed to be independent; so one way to actually set this up is to construct both processes on different probability spaces, and then take the product of these two probability spaces as our new probability space, which is big enough to accommodate for all the randomness needed.
Oftentimes, we just want to take expectations with respect to one of the two processes. Formally, this corresponds to taking the conditional expectation, so taking the expectation with respect to the random walk starting in 0 technically boils down to taking the conditional expectation $\mathbb{E}[\dots \mid (\xi(z))_{z\in\mathbb{Z}^d}]$.

To extend the definition of $\alpha(y)$ in Theorem 4.1.13 to $y \notin \mathbb{Q}^d$ and to prove the uniform convergence in (4.1.13), we need to establish the equicontinuity of $t^{-1}a(t, 0, ty)$ in y, as $t \to \infty$. For more details, see [DGRS12].

Chapter 5

Brownian motion and random walks

For a profound study of Brownian motion we refer to the monograph [MP10], which is one of the standard sources on the topic, including relatively recent and advanced results.
An important fact in this section, which will be used without further mentioning, is that the distribution of a d-dimensional normally distributed random variable is characterised by its expectation and its covariance matrix already (cf. Examples 2.2.8 and 1.5.8 of [Dre19]).

5.1 Another constructions of Brownian motion

We recall the definition of Brownian motion in Definition 3.2.18.
While we have already seen in Theorem 3.2.20 that Brownian motion exists, we want to give some further possibilities of construction, in particular since in Theorem 3.2.20 we took advantage of the result by Kolmogorov and Chentsov (which we did not prove) in order to deduce the desired $\mathbb{P}$-a.s. continuity of the sample paths.
As a warm-up, we start with a simple exercise.

Exercise 5.1.1. *Let (B_t), $t \in [0, \infty)$ be standard Brownian motion and let $r \in [0, \infty)$. Then the process $(B_{t+r} - B_r)$, $t \in [0, \infty)$, is also standard Brownian motion.*

5.1.1 Lévy's construction of Brownian motion

This construction is due to P. Lévy [L9́2], the first edition of which had been published in 1948. We will follow the version of [MP10] here. There are other constructions of Brownian motion as a uniform limiting process (in the supremum norm for continuous functions on compact intervals) of continuous processes (e.g. using Haar functions, see Section 21.5 in [Kle14] or Section 2.5 in [KS91]), but the one we give is very intuitive and furthermore has nice consistency properties.

Theorem 5.1.2 (N. Wiener (1923))**.** *Standard Brownian motion exists.*

For the proof we will need a couple of lemmas that are interesting in their own rights.

Lemma 5.1.3. *For $X \sim \mathcal{N}(0, 1)$ we have for $x > 0$ that*

$$\Big(\frac{1}{x} - \frac{1}{x^3}\Big)\frac{1}{\sqrt{2\pi}}e^{-x^2/2} \le \mathbb{P}(X \ge x) \le \frac{1}{x}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

Proof.
We have

$$\sqrt{2\pi}\mathbb{P}(X \ge x) = \int_x^\infty e^{-\frac{s^2}{2}}\,\mathrm{d}s \le \int_x^\infty \frac{s}{x}e^{-\frac{s^2}{2}}\,\mathrm{d}s = x^{-1}e^{-\frac{x^2}{2}},$$

which is the upper bound.

The lower is slightly more involved, and we compute using the transformation $s \mapsto x + \frac{s}{x}$ as well as the fact that $1 - z < e^{-z}$ for $z > 0$,

$$\begin{aligned}\sqrt{2\pi}\mathbb{P}(X \geq x) &= \int_x^\infty e^{-\frac{s^2}{2}}\,\mathrm{d}s = x^{-1}\int_0^\infty e^{-\frac{x^2+2s+s^2/x^2}{2}}\,\mathrm{d}s \\ &= x^{-1}e^{-x^2/2}\int_0^\infty e^{-\frac{2s+s^2/x^2}{2}}\,\mathrm{d}s \geq x^{-1}e^{-x^2/2}\int_0^\infty e^{-s}(1-\frac{s^2}{2x^2})\,\mathrm{d}s,\end{aligned}$$

and the latter integral equals $1 - x^{-2}$, which finishes the proof. □

Lemma 5.1.4. *Let X be a d-dimensional Gaussian random variable, i.e., distributed according to $\mathcal{N}(\mu, \Sigma)$ with mean $\mu \in \mathbb{R}^d$ and covariance matrix $\Sigma \in \mathbb{R}^{d\times d}$. Then the coordinates (X_i), $i \in \{1, \dots, d\}$, are independent random variables if and only if*

$$\operatorname{Cov}(X_i, X_j) = 0 \quad \forall\, i, j \in \{1, \dots, d\} \text{ with } i \neq j. \tag{5.1.1}$$

Proof. We have gotten to know d-dimensional normally distributed random variables in Examples 2.2.8 and 1.5.8 of [Dre19] for instance, which in particular showed that the distribution of a d-dimensional normally distributed random variable is completely determined by its mean and its covariance matrix.
Since independence of the (X_i) implies that they are uncorrelated also, the interesting direction is the reverse; so assume (5.1.1) to hold. Then, if $(Y_1, \dots, Y_d)$ denotes a random vector such that its entries are i.i.d. $\mathcal{N}(0,1)$-distributed, we infer that the vector $(\mu_1 + \sqrt{\operatorname{Var}(X_1)}Y_1, \dots, \mu_d + \sqrt{\operatorname{Var}(X_d)}Y_d)$ has the same distribution as the vector $(X_1, \dots, X_d)$, which implies the independence of the (X_i), $i \in \{1, \dots, d\}$. □

Exercise 5.1.5. *Convince yourself that it is essential in the previous lemma to not only have that the X_i are all Gaussian but that indeed the vector X is a Gaussian random vector. Indeed, consider the case $d = 2$, $X_1 \sim \mathcal{N}(0,1)$, and ε an independent Rademacher-distributed random variable. Set $X_2 := \varepsilon \cdot X_1$. Check that X_1 and X_2 are both $\mathcal{N}(0,1)$-distributed random variables, that they are uncorrelated, but not independent.*

Lemma 5.1.6. *Let X and Y be real random variables which are independent and $\mathcal{N}(0,\sigma^2)$-distributed ($\sigma^2 \in (0,\infty)$). Then the random variables $X+Y$ and $X-Y$ are independent and $\mathcal{N}(0, 2\sigma^2)$-distributed.*

Proof.
It is obvious from Example 3.3.11 (a) in [Dre19] that $X+Y$ and $X-Y$ are $\mathcal{N}(0,2\sigma^2)$-distributed, and that $(X+Y, X-Y)$ is a Gaussian random vector since it equals

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} X \\ Y \end{pmatrix};$$

you may convince yourself that this means that the vector $(X+Y, X-Y)$ is Gaussian distributed with mean 0 and covariance matrix given by

$$\sigma^2 \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^t.$$

But we have

$$\mathbb{E}[(X+Y)(X-Y)] = \mathbb{E}[X^2] - \mathbb{E}[Y^2] = 0,$$

which means $\operatorname{Cov}((X+Y)(X-Y)) = 0$, and so Lemma 5.1.4 implies the desired independence. □

Proof of Theorem 5.1.2. We first note that it is sufficient to standard construct Brownian motion (B_t) on the interval $[0,1]$ (i.e., a stochastic process satisfying all the properties of Brownian motion but with $[0,\infty)$ replaced by $[0,1]$), since we can then extend it to $[0,\infty)$ by 'patching together' a countable number of independent Brownian motions. Indeed, denoting by $(B_t^{(n)})$, $t \in [0,1]$, $n \in \mathbb{N}$, an independent family of such processes, we can define

$$B_t := \sum_{i=1}^{\lfloor t \rfloor} B_1^{(i)} + B_{t-\lfloor t \rfloor}^{(\lfloor t \rfloor + 1)}, \quad t \in [0,\infty),$$

and check that this satisfies the properties of Brownian motion.
Now in order to define Brownian motion (B_t), $t \in [0,1]$, on $[0,1]$, the idea is to first construct it on the dyadic numbers. For this purpose, set

$$D_n := \{k2^{-n} : k \in \{0,\dots,2^n\}\}, \quad n \in \mathbb{N}.$$

as well as $D := \cup_{n\in\mathbb{N}} D_n$.
In order to do this, let a family of i.i.d. random variables (X_t), $t \in D$, be given such that $X_t \sim \mathcal{N}(0,1)$.
To start with, we want to construct a process (B_t), $t \in D$, such that the following properties are fulfilled for each $n \in \mathbb{N}$:

(a) For $r < s < t \in D_n$ the random variables $B_t - B_s$ and $B_s - B_r$ are independent, and the former is $\mathcal{N}(0, t-s)$ distributed whereas the latter is $\mathcal{N}(0, s-r)$ distributed;

(b) the family of random variables B_t, $t \in D_n$, is independent of the family X_t, $t \in D \backslash D_n$.

Define

$$B_0 := 0 \quad \text{and} \quad B_1 := X_1, \tag{5.1.2}$$

and inductively, given B_t for all $t \in D_{n-1}$, $n \in \mathbb{N}$, we define for $t \in D_n \backslash D_{n-1}$

$$B_t := \frac{B_{t-2^{-n}} + B_{t+2^{-n}}}{2} + 2^{-\frac{n+1}{2}} \cdot X_t = \frac{B_{t+2^{-n}} - B_{t-2^{-n}}}{2} + B_{t-2^{-n}} + 2^{-\frac{n+1}{2}} \cdot X_t, \tag{5.1.3}$$

and where we note that $\frac{B_{t+2^{-n}} - B_{t-2^{-n}}}{2}$ as well as $2^{-\frac{n+1}{2}} \cdot X_t$ are both $\mathcal{N}(0, 2^{-(n+1)})$-distributed. While, as alluded to before, the first summand is a linear interpolation between the values of the process at the neighbouring dyadic sites of the next largest scale, the second is an independent normal perturbation of exactly the right scale.
We now prove inductively that the process (B_t), $t \in D$, constructed above fulfills the desired properties. Indeed, for $n = 0$ it is obvious from (5.1.2) that (b) and (a) hold true. Now assume that (b) and (a) hold for all numbers in $\{0,1,\dots,n\}$, some $n \in \mathbb{N}_0$. Then we infer in combination with (b) and the independence assumption on the (X_t), $t \in D$, that the families

$$\begin{aligned} (B_t), &\quad t \in D_n, \\ (X_t), &\quad t \in D_{n+1} \backslash D_n, \\ (X_t), &\quad t \in D \backslash D_{n+1}, \end{aligned}$$

are all independent, and therefore from the definition of B_t for $t \in D_{n+1}$ we infer that (b) holds for $n+1$ as well.
In order to establish (a), observe that due to the recursive definition of (5.1.3), it is sufficient to show that the family of all increments $(B_{t+2^{-(n+1)}} - B_t)$, $t \in D_{n+1} \backslash \{1\}$, is an independent family, since they're all $\mathcal{N}(0, 2^{-(n+1)})$-distributed. Since the vector formed by these increments is a Gaussian random vector (check!), it is in fact sufficient to show that they are pairwise uncorrelated due to Lemma 5.1.4.
Now for this purpose, we distinguish two cases:

(a) If $t - s = 2^{-(n+1)}$, then from (5.1.3) we infer that

$$B_{t+2^{-(n+1)}} - B_t = \frac{B_{t+2^{-n}} - B_{t-2^{-n}}}{2} - 2^{-\frac{n+1}{2}} \cdot X_t,$$

as well as

$$B_{s+2^{-(n+1)}} - B_s = \frac{-B_{t+2^{-n}} + B_{t-2^{-n}}}{2} - 2^{-\frac{n+1}{2}} \cdot X_t.$$

Now note that in both displays, the minuend on the right-hand side as well as the subtrahend (in the outer difference) are $\mathcal{N}(0, 2^{-(n+1)})$-distributed. With Lemma 5.1.6 we therefore infer that $B_{t+2^{-(n+1)}} - B_t$ and $B_{s+2^{-(n+1)}} - B_s$ are independent.

(b) If $t - s > 2^{-(n+1)}$, let $m \in \mathbb{N}$ be minimal such that there exists $d \in D_m$ such that we have the interval inclusion

$$[s, s + 2^{-(n+1)}] \subset [d - 2^{-m}, d] \quad \text{and} \quad [t, t + 2^{-(n+1)}] \subset [d, d + 2^{-m}]$$

(convince yourself that such d exists). Note that

$$m < n + 1. \tag{5.1.4}$$

Since the increments over intervals of length $2^{-(n+1)}$ which are contained in $[d - 2^{-m}, d]$, $d \in D_m$, are constructed using $B_{d+2^{-m}} - B_d$ as well as X_r, $r \in [d - 2^{-m}, d]$, and similarly for $[d, d + 2^{-m}]$, we infer the desired independence.

This shows that (b) and (a) hold true for all $n \in \mathbb{N}$.
We now define our limiting process as follows. Start with introducing the functions

$$F_0(t) := \begin{cases} X_1, & \text{if } t = 1, \\ 0, & \text{if } t = 0, \\ \text{linear interpolation between } X_0 \text{ and } X_1, & \text{otherwise,} \end{cases}$$

as well as

$$F_n(t) := \begin{cases} 2^{-\frac{n+1}{2}} X_t, & \text{if } t \in D_n \backslash D_{n-1}, \\ 0, & \text{if } t \in D_{n-1}, \\ \text{linear interpolation of } X \text{ between consecutive points of } D_n, & \text{otherwise.} \end{cases}$$

for $n \geq 1$, and note in passing that they are continuous on $[0, 1]$.
Then for $t \in D_n$ it is not hard to see that we have

$$B_t = \sum_{i=0}^{n} F_i(t) = \sum_{i=0}^{\infty} F_i(t). \tag{5.1.5}$$

Our goal now is to show that the infinite sum converges uniformly on $[0, 1]$ and hence the limit is a continuous function again.
For this purpose we notice that as a consequence of Lemma 5.1.3, for any $C \in (0, \infty)$ fixed, we have

$$\mathbb{P}(|X_t| \geq C\sqrt{n}) \leq e^{-\frac{C^2 n}{2}} \tag{5.1.6}$$

for all $n \geq N$, some $N \in \mathbb{N}$.
Hence, we infer

$$\sum_{n \in \mathbb{N}_0} \mathbb{P}\big(\exists t \in D_n \text{ such that } |X_t| \geq C\sqrt{n}\big) \leq \sum_{n \leq N} (2^n + 1) + \sum_{n > N} (2^n + 1) e^{-\frac{C^2 n}{2}}.$$

Now for $C > \sqrt{2 \ln 2}$ the series on the right-hand side converges and therefore the Borel-Cantelli lemma implies that there exists a random but finite N such that for all $n \geq N$,

$$\forall t \in D_n \quad |X_t| \leq C\sqrt{n},$$

so

$$\|F_n\|_\infty \leq C\sqrt{n} \cdot 2^{-\frac{n}{2}}.$$

This entails that $\mathbb{P}$-a.s., the infinite sequence on the right-hand side of (5.1.5) converges uniformly on $[0,1]$, and the continuous limit on $[0,1]$ is what we define (B_t), $t \in [0,1]$, to be.
It remains to check that the increments of (B_t), $t \in [0,1]$, have the required finite-dimensional distributions. For this purpose, let $0 \leq t_1 < \ldots < t_n \leq 1$, and let for each $i \in \{1, \ldots, n\}$, let $(t_k^{(i)})$, $k \in \mathbb{N}$, be a D-valued sequence such that $(t_k^{(i)}) \to t_i$ as $k \to \infty$. Then, for all k large enough, $(B_{t_k^{(i+1)}} - B_{t_k^{(i)}}) \sim \mathcal{N}(0, t_k^{(i+1)} - t_k^{(i)})$, and they form an independent family in $i \in \{1, \ldots, n-1\}$. Furthermore, $\mathbb{P}$-a.s., as $k \to \infty$,

$$(B_{t_k^{(i+1)}} - B_{t_k^{(i)}}) \to (B_{t_{i+1}} - B_{t_{(i)}}).$$

Using e.g. characteristic functions and Theorem 2.6.3 of [Dre19], it is not hard to show that $(B_{t_{i+1}} - B_{t_{(i)}}) \sim \mathcal{N}(0, t_{i+1} - t_i)$, and that the $(B_{t_{i+1}} - B_{t_{(i)}})$, $i \in \{1, \ldots, n-1\}$, form an independent family, which finishes the proof. □

5.2 Some properties of Brownian motion

Proposition 5.2.1 (Scaling invariance)**.** *If $c > 0$ and (B_t) is a Brownian motion starting in $x \in \mathbb{R}^d$, then the process $(\frac{1}{c} B_{c^2 t})$ is a Brownian motion starting in $\frac{x}{c}$.*

Proof. The only property of Definition 3.2.18 that is not clear is Part (c). But by assumption we have that $B_{c^2 t} - B_{c^2 s} \sim \mathcal{N}(0, c^2(t-s)\mathrm{Id})$, and hence we have that

$$\frac{1}{c} B_{c^2 t} - \frac{1}{c} B_{c^2 s} \sim \mathcal{N}(0, (t-s)\mathrm{Id}),$$

which finishes the proof. □

Remark 5.2.2. *The above result gives rise to the term* diffusive scaling, *which means 'speeding up' time by a factor n^2 and 'shrinking' via division by n, and then taking $n \to \infty$. This is the same scaling that you have seen in the central limit theorem (where the only notational difference was that time was of order n and space was rescaled by $n^{-1/2}$, which is the same in the sense that time is rescaled by the inverse square of the rescaling of space).*

Proposition 5.2.3 (Time inversion)**.** *If (B_t), $t \in [0, \infty)$, is Brownian motion, then so is $(tB_{\frac{1}{t}})$, $t \in [0, \infty)$, with $0 B_{\frac{1}{0}} := 0$.*

Proof. Recall that the finite-dimensional distributions $(B_{t_1}, \ldots, B_{t_n})$ of Brownian motion are Gaussian random vectors and are therefore characterised by $\mathbb{E}[B_{t_i}] = 0$ and $\mathrm{Cov}(B_{t_i}, B_{t_j}) = t_i$ for $0 \leq t_i \leq t_j$. Define now $X_t := tB_{1/t}$ with $X_0 = 0$. Trivially, this process is also Gaussian and the corresponding Gaussian vectors have expectation zero. It remains to check what the covariances are. For $t > 0, h \geq 0$, they are given by

$$\mathrm{Cov}(X_{t+h}, X_t) = (t+h)t \,\mathrm{Cov}(B_{1/(t+h)}, B_{1/t}) = t(t+h)\frac{1}{t+h} = t.$$

This confirms that the law of all finite-dimensional distributions of X_t are the same as for Brownian motion. The continuity of paths $t \to X_t$ is clear for $t > 0$. For $t = 0$ we use the following two facts. First, since $\mathbb{Q}$ is countable, the distribution of $\{X_t : t \geq 0, t \in \mathbb{Q}\}$ is the same as for a Brownian motion and hence

$$\lim_{\substack{t \to 0 \\ t \in \mathbb{Q}}} X_t = 0 \text{ almost surely.}$$

Secondly, $\mathbb{Q} \cap (0, \infty)$ is dense in $(0, \infty)$ and $\{X_t : t \geq 0\}$ is almost surely continuous on $(0, \infty)$ so that

$$0 = \lim_{\substack{t \to 0 \\ t \in \mathbb{Q}}} X_t = \lim_{t \to \infty} X_t \text{ almost surely.}$$

This proves the continuity of paths and yields the stated claim. $\square$

While so far we've mainly been collecting regularity properties of Brownian motion, we now turn to a property which explains why I would fail miserably at sketching Brownian motion at the blackboard.

Proposition 5.2.4. *Let (B_t) be Brownian motion, and let $[a, b] \subset [0, \infty)$, $a < b$, be an interval. Then $\mathbb{P}$-a.s., (B_t) is not monotone on the interval $[a, b]$.*

Proof. We give the proof for 'increasing', and the case of 'decreasing' follows from considering $(-B_t)$, $t \in [0, \infty)$, using that this also constitutes a Brownian motion.
For arbitrary $n \in \mathbb{N}$ we partition $[a, b]$ into the subintervals

$$[a_{n,k}, b_{n,k}), \quad k \in \{0, \ldots, n-1\},$$

where $a_{n,k} := a + \frac{k}{n}(b-a)$, and $b_{n,k} := a_{n,k} + \frac{1}{n}(b-a)$. Denoting by M_I the event that Brownian motion is increasing on the interval I, we have for each $n \in \mathbb{N}$ that

$$M_{[a,b]} = \bigcap_{k=0}^{n-1} M_{[a_{n,k}, b_{n,k})}.$$

Furthermore, since increments of Brownian motion are independent, we have that the event in the intersection on the right-hand side are independent, and we also have

$$\mathbb{P}(M_{[a_{n,k}, b_{n,k})}) \leq \mathbb{P}(B_{b_{n,k}} - B_{a_{n,k}} \geq 0) = \frac{1}{2},$$

since on any such interval the increments are centered Gaussian. All in all, this supplies us with

$$\mathbb{P}(M_{[a,b]}) = \prod_{k=0}^{n-1} \mathbb{P}(M_{[a_{n,k}, b_{n,k})}) \leq 2^{-n} \quad \forall n \in \mathbb{N},$$

so $\mathbb{P}(M_{[a,b]}) = 0$ which is what we wanted to prove. $\square$

5.2.1 Law of the iterated logarithm

The law of the iterated logarithm is in some sense a regime that is in between that of the law of large numbers and the central limit theorem. In fact, while the (strong version of) the former implies $S_n/n \to 0$ almost surely under suitable circumstances, the latter states the convergence in distribution of $S_n/\sqrt{n}$ to a normal random variable.
In fact, for arbitrary $C > 0$ we have in combination with Corollary 2.1.6 of [Dre19] that

$$\mathbb{P}\big(\limsup_{n \to \infty} B_n > C\sqrt{n}\big) \geq \limsup_{n \to \infty} \mathbb{P}(B_n > C\sqrt{n}) = \mathbb{P}(B_1 > C) \in (0, 1),$$

so Kolmogorov's $0-1$-law implies that almost surely,

$$\limsup_{n\to\infty} \frac{B_n}{\sqrt{n}} = \infty.$$

On the other hand, we had already seen in Remark 3.8.2 that the sequence cannot converge in probability either.
The law of the iterated logarithm gives the right scaling (including constant) for it to converge.

Theorem 5.2.5 (Law of the iterated logarithm). *Let (B_t) be one-dimensional Brownian motion starting in $x \in \mathbb{R}$. Then $\mathbb{P}$-a.s.,*

$$\limsup_{t\to\infty} \frac{B_t}{\sqrt{2t\ln\ln t}} = 1. \tag{5.2.1}$$

Proof. Since Brownian motion starting in x can be obtained by shifting standard Brownian motion by x, and since the influence of the starting point vanishes in (5.2.1), it is sufficient to prove the result for standard Brownian motion.
We start with showing that $\mathbb{P}$-a.s.,

$$\limsup_{t\to\infty} \frac{B_t}{\sqrt{2t\ln\ln t}} \leq 1 \tag{5.2.2}$$

by chopping time into intervals which are growing exponentially. More precisely, for $\alpha > 1$ we set $t_n := \alpha^n$ as well as $f(t) := 2\alpha^2 \ln\ln t$. Since $t \mapsto tf(t)$ is increasing, we observe

$$\mathbb{P}\Big(\max_{t_n\leq s\leq t_{n+1}} \frac{B_s}{\sqrt{2s\ln\ln s}} > \alpha\Big) \leq \mathbb{P}\Big(\max_{t_n\leq s\leq t_{n+1}} B_s > \sqrt{t_n f(t_n)}\Big). \tag{5.2.3}$$

Then, using scale invariance of Brownian motion in the first equality, the reflection principle Theorem 5.2.12 and Lemma 5.1.3 to obtain the second inequality, and the identity $\frac{f(t_n)}{2\alpha} = \alpha(\ln(n\ln\alpha)) = \alpha\ln n + \alpha\ln\ln\alpha$ in the penultimate step, we get

$$\begin{aligned}
&\mathbb{P}\Big(\max_{t_n\leq s\leq t_{n+1}} B_s > \sqrt{t_n f(t_n)}\Big) \leq \mathbb{P}\Big(\frac{1}{\sqrt{t_{n+1}}}\max_{0\leq s\leq t_{n+1}} B_s > \sqrt{\frac{f(t_n)}{\alpha}}\Big) = \mathbb{P}\Big(\max_{0\leq s\leq 1} B_s > \sqrt{\frac{f(t_n)}{\alpha}}\Big)\\
&\leq \frac{2}{\sqrt{2\pi}}\sqrt{\frac{\alpha}{f(t_n)}}e^{-\frac{f(t_n)}{2\alpha}} \leq \sqrt{\frac{\alpha}{f(t_n)}}(\ln\alpha)^{-\alpha}n^{-\alpha} \leq n^{-\alpha},
\end{aligned}$$

where the latter inequality holds true for all n large enough.
Now since $\alpha > 1$, the right-hand side is summable in n, and we obtain using the Borel-Cantelli Lemma 1.13.6 that almost surely,

$$\limsup_{t\to\infty} \frac{B_t}{\sqrt{tf(t)}} < \alpha.$$

Taking $\alpha \downarrow 1$ we recover (5.2.2).
It remains to show the lower bound

$$\limsup_{t\to\infty} \frac{B_t}{\sqrt{2t\ln\ln t}} \geq 1 \tag{5.2.4}$$

For this purpose, we will take $\alpha \to \infty$ in the above scales. Define, with α and t_n as above the quantities $\widetilde{\alpha} := \frac{\alpha-1}{\alpha} < 1$ and $g(t) := 2\widetilde{\alpha}^2\ln\ln t$. Using that $t_{n+1} - t_n = \alpha^n(\alpha-1) = t_{n+1}\widetilde{\alpha}$, recalling Lemma 5.1.3 (and taking advantage of the fact that $\frac{1}{x} - \frac{1}{x^3} \geq \frac{1}{2x}$ for $x \geq 1$ in that bound) we obtain

$$\begin{aligned}
&\mathbb{P}(B_{t_{n+1}} - B_{t_n} > \sqrt{t_{n+1}g(t_{n+1})}) = \mathbb{P}\Big(\frac{B_{t_{n+1}} - B_{t_n}}{\sqrt{t_{n+1}-t_n}} > \sqrt{\frac{t_{n+1}g(t_{n+1})}{t_{n+1}-t_n}}\Big) = \mathbb{P}(B_1 > \sqrt{g(t_{n+1})/\widetilde{\alpha}})\\
&\geq \frac{1}{\sqrt{2\pi}}\frac{1}{2}\sqrt{\frac{\widetilde{\alpha}}{g(t_{n+1})}}e^{-\frac{g(t_{n+1})}{2\widetilde{\alpha}}} = \frac{1}{\sqrt{2\pi}}\frac{1}{2}\sqrt{\frac{\widetilde{\alpha}}{g(t_{n+1})}}(\ln\alpha)^{-\widetilde{\alpha}}(n+1)^{-\widetilde{\alpha}}.
\end{aligned}$$

Since $\widetilde{\alpha} < 1$ we infer that the right-hand side summed over n yields ∞, so, using the independence of the respective events, the Borel-Cantelli Lemma supplies us with

$$\mathbb{P}\big(\underbrace{B_{t_{n+1}} - B_{t_n} > \sqrt{t_{n+1} g(t_{n+1})}}_{=:G_n} \quad \text{for infinitely many } n\big) = 1. \tag{5.2.5}$$

Now using the upper bound (5.2.2) in combination with the symmetry of Brownian motion and the fact that

$$\lim_{n\to\infty} \frac{t_{n+1}\ln\ln t_{n+1}}{t_n \ln\ln t_n} = \alpha,$$

we infer that for $\varepsilon > 0$, $\mathbb{P}$-a.s.,

$$B_{t_n} > -\frac{1+\varepsilon}{\sqrt{\alpha}}\sqrt{2t_{n+1}\ln\ln t_{n+1}}$$

for all n large enough. Using this in combination with (5.2.5) we infer that $\mathbb{P}$-a.s., for all n large enough, we have that on G_n,

$$B_{t_{n+1}} - B_{t_n} + B_{t_n} \geq \sqrt{t_{n+1}g(t_{n+1})} - \frac{1+\varepsilon}{\sqrt{\alpha}}\sqrt{2t_{n+1}\ln\ln t_{n+1}}$$

and hence

$$\limsup_{n\to\infty} \frac{B_{t_n}}{\sqrt{2t_n \ln\ln t_n}} \geq \widetilde{\alpha} - \frac{1+\varepsilon}{\sqrt{\alpha}} = \frac{\alpha - 1}{\alpha} - \frac{1+\varepsilon}{\sqrt{\alpha}}.$$

Taking $\alpha \to \infty$ we recover (5.2.4), which finishes the proof.

□

5.2.2 Non-differentiability of Brownian motion

Corollary 5.2.6 (non-differentiability of Brownian motion). *For $t \in [0,\infty)$ fixed, $\mathbb{P}$-a.s., Brownian motion is not differentiable at t. More precisely, we even have that $\mathbb{P}$-a.s.,*

$$\limsup_{h\downarrow 0} \frac{B_{t+h} - B_t}{h} = -\liminf_{h\downarrow 0} \frac{B_{t+h} - B_t}{h} = \infty.$$

Remark 5.2.7. *The fact that $\mathbb{P}$-a.s., sample paths of Brownian motion are continuous (even α-Hölder continuous for $\alpha \in (0, \frac{1}{2})$, cf. Theorem 3.2.20) but nowhere differentiable, is quite interesting. Indeed, if you want to construct examples of such functions explicitly (see e.g. the 'Weierstraß function'), this is non-trivial.*

Proof of Corollary 5.2.6. In combination with Proposition 5.2.3 we infer

$$\limsup_{h\downarrow 0} \frac{B_h - B_0}{h} = \limsup_{h\downarrow 0} \frac{hB_{h^{-1}}}{h} = \limsup_{s\to\infty} B_s = \infty,$$

where the last equality is due to Theorem 5.2.5, and similarly for the limes inferior, which entails the statement for $t = 0$. For general t we observe that the process $(B_{s+t} - B_t)$, $s \in [0,\infty)$, is a standard Brownian motion again (cf. Exercise 5.1.1), and this finishes the proof.

□

Theorem 5.2.8 (Blumenthal's 0-1 law). *Let (B_t) be Brownian motion. Write*

$$\mathcal{F}_0^+ := \bigcap_{t>0} \mathcal{F}_t,$$

where $(\mathcal{F}_t)$ denotes the canonical σ-algebra of (B_t), and $\mathcal{F}_0^+$ is oftentimes referred to as the germ σ-algebra.
Then $\mathcal{F}_0^+$ is $\mathbb{P}$-trivial.

In order to prove Theorem 5.2.8 let us first generalise slightly the germ σ-algebra and set

$$\mathcal{F}_s^+ = \bigcap_{t>s} \mathcal{F}_t.$$

Clearly, due to the growing nature of filtrations, we have that $\mathcal{F}_s \subset \mathcal{F}_s^+$. Due to the fact that $\mathcal{F}_s^+$ might be bigger, it is not immediately clear anymore that the process $(B_{t+s} - B_s)_{t\geq 0}$ is a Brownian motion independent of the σ-algebra $\mathcal{F}_s^+$ (due to the independence of increments, this statement is trivially true for the σ-algebra $\mathcal{F}_s$). It turns out however, that we retain independence also for the slightly bigger σ-algebra $\mathcal{F}_s^+$.

Theorem 5.2.9. *For every $s \geq 0$, the process $(B_{t+s} - B_s)_{t\geq 0}$ is a standard Brownian motion independent of $\mathcal{F}_s^+$.*

Proof. Suppose $s_n > s$ such that $s_n \downarrow s$. By the Markov property of Brownian motion, the process $(B_{t+s_n} - B_{s_n})_{t\geq 0}$ is a standard Brownian motion independent of $\mathcal{F}_{s_n}$. By continuity of Brownian motion, for all $t_j > t_{j-1} > \cdots > t_1 > 0$, we have

$$(B_{t_j+s} - B_{t_{j-1}+s}, \dots B_{t_2+s} - B_{t_1+s}) = \lim_{n\to\infty} (B_{t_j+s_n} - B_{t_{j-1}+s_n}, \dots B_{t_2+s_n} - B_{t_1+s_n}),$$

so we infer that $(B_{t+s} - B_s)_{t\geq 0}$ is independent of $\bigcap_n \mathcal{F}_{s_n} = \mathcal{F}_s^+$. □

Proof of Theorem 5.2.8. For $s = 0$, we have that any $A \in \sigma(B_t : t \geq 0)$ is independent of $\mathcal{F}_0^+$ by the preceding theorem. This applies in particular to $A \in \mathcal{F}_0^+$, which makes it independent of itself and must therefore have probability 0 or 1. □

Remark 5.2.10. *We might ask ourselves whether the σ-algebras $\mathcal{F}_s^+$ contribute anything to the understanding of Brownian motion beyond their applications to $\mathbb{P}$-trivial events. An affirmative answer to this question follows immediately if we consider for example the random variable $\tau_A = \inf\{t \geq 0 : B_t \in A\}$ for an* open *set $A \subset \mathbb{R}$ (for simplicity, let $B_0 \notin A$). Then it can easily be verified that τ_A is a stopping time with respect to the filtration $(\mathcal{F}_s^+)_{s\geq 0}$, but not with respect to $(\mathcal{F}_s)_{s\geq 0}$. The intuitive reason for this is due to the fact that we can only learn whether B has entered an open set by looking infinitesimally into the future, where this foresight is not needed for closed sets (which we already know since for closed sets τ_A is a stopping time also with respect to $(\mathcal{F}_s)_{s\geq 0}$).*

5.2.3 The strong Markov property of Brownian motion

We already know that Brownian motion satisfies the Markov property, which can be interpreted as Brownian motion "restarting" itself at each deterministic time. Crucially, for Brownian motion this is true also for stopping times.

Theorem 5.2.11 (Strong Markov property)**.** *For every almost surely finite stopping time τ, the process $(B_{\tau+t} - B_\tau)_{t\geq 0}$ is a standard Brownian motion independent of $\mathcal{F}_\tau^+$.*

Proof. We first show the statement for discrete approximations τ_n of τ, i.e. for $\tau_n = (m+1)2^{-n}$ if $m2^{-n} \leq \tau < (m+1)2^{-n}$. Write $B^k = (B_t^k)_{t\geq 0}$ for the Brownian motion defined by $B_t^k = B_{t+k/2^n} - B_{k/n^2}$ and $B^* = (B_t^*)_{t\geq 0}$ for the process defined by $B_t^* = B_{t+\tau_n} - B_{\tau_n}$. Let $E \in \mathcal{F}_{\tau_n}^+$. It then holds for every event $\{B^* \in A\}$

$$\begin{aligned}\mathbb{P}(\{B^* \in A\} \cap E) &= \sum_{k=0}^{\infty} \mathbb{P}(\{B^k \in A\} \cap E \cap \{\tau_n = k2^{-n}\}) \\ &= \sum_{k=0}^{\infty} \mathbb{P}(B^k \in A)\mathbb{P}(E \cap \{\tau_n = k2^{-n}\}),\end{aligned}$$

where we used the independence of $\{B^k \in A\}$ from $E \cap \{\tau_n = k2^{-n}\} \in \mathcal{F}^+_{k2^{-n}}$, which we have due to the preceding theorem. Using the (normal) Markov property, we have that $\mathbb{P}(B^k \in A) = \mathbb{P}(B \in A)$ and therefore does not depend on k. This gives

$$\begin{aligned}\sum_{k=0}^{\infty} \mathbb{P}(B^k \in A)\mathbb{P}(E \cap \{\tau_n = k2^{-n}\}) &= \mathbb{P}(B \in A) \sum_{k=0}^{\infty} \mathbb{P}(E \cap \{\tau_n = k2^{-n}\}) \\ &= \mathbb{P}(B \in A)\mathbb{P}(E),\end{aligned}$$

which shows that B^* is a Brownian motion and independent of E, hence of $\mathcal{F}^+_{\tau_n}$.
We now generalise this statement to a general stopping time τ. As $\tau_n \downarrow \tau$ we have that

$$(B_{t+\tau_n} - B_{\tau_n})_{s\geq 0}$$

is a Brownian motion independent of $\mathcal{F}^+_{\tau_n} \supset \mathcal{F}^+_{\tau}$. Therefore, the increments

$$B_{s+t+\tau} - B_{t+\tau} = \lim_{n\to\infty} B_{s+t+\tau_n} - B_{t+\tau_n}$$

of the process $(B_{r+\tau} - B_\tau)_{r\geq 0}$ are independent and normally distributed with mean zero and variance s. Since the process is trivially almost surely continuous, it is a Brownian motion. Furthermore, all increments $B_{s+t+\tau} - B_{t+\tau} = \lim B_{s+t+\tau_n} - B_{t+\tau_n}$ are independent of $\mathcal{F}^+_\tau$, as is therefore the process itself.

□

Theorem 5.2.12 (Reflection principle for Brownian motion (essentially D. André, 1887))**.** *Let (B_t) be standard Brownian motion. Then for $t \in (0, \infty)$ and $M > 0$ we have*

$$\mathbb{P}(\sup_{s\in[0,t]} B_s \geq M) = 2\mathbb{P}(B_t \geq M).$$

In order to prove Theorem 5.2.12, we will first show that taking a Brownian motion and reflecting it upon hitting some value results in a process that is itself a Brownian motion. More precisely, we have the following result.

Theorem 5.2.13. *If τ is a stopping time and $(B_t)_{t\geq 0}$ is a standard Brownian motion, then the process $(B^*_t)_{t\geq 0}$ called the Brownian motion reflected at τ and defined by*

$$B^*_t = B_t \mathbb{1}_{t\leq\tau} + (2B_\tau - B_t)\mathbb{1}_{t>\tau}$$

is also a standard Brownian motion.

Proof. On the event that τ is finite, by the strong Markov property both paths

$$(B_{t+\tau} - B_\tau)_{t\geq 0} \text{ and } -(B_{t+\tau} - B_\tau)_{t\geq 0}$$

are Brownian motions and independent of the beginning $(B_t)_{0\leq t\leq\tau}$. The concatenation mapping, which takes a continuous path $(g_t)_{t\geq 0}$ and glues it to the end point of a finite continuous path $(f_t)_{0\leq t\leq\tau}$ to form a new continuous path, is measurable. Therefore, the process arising from glueing the first path of the pair above to $(B_t)_{0\leq t\leq\tau}$ and the process arising from gluing the second path of the pair to $(B_t)_{0\leq t\leq\tau}$ have the same distribution. Clearly the first case results in the process $(B_t)_{t\geq 0}$ and the second to $(B^*_t)_{t\geq 0}$. On the event that τ is infinite, the two processes agree by definition. □

Proof of Theorem 5.2.12. Let $\tau = \inf\{t \geq 0 : B_t = M\}$ and let B^* be the Brownian motion reflected at stopping time τ. Then

$$\{\sup_{s\in[0,t]} B_s \geq M\} = \{B_t \geq M\} \cup \{\sup_{s\in[0,t]} B_s \geq M, B_t < M\}.$$

This is by choice a disjoint union and the second event coincides with the event $\{B^*_t > M\}$. By the preceding theorem we obtain the desired result. □

5.2.4 The zero set of Brownian motion

What we will show in this section is the perhaps surprising property of standard Brownian motion, that the set of points $\{t \geq 0 : B_t = 0\}$ is closed and contains no isolated points (such sets are sometimes called *perfect* sets). This is especially surprising, as the Brownian motion almost surely has zeroes that are isolated from the left (e.g. the first zero after time 1/2) or from the right (e.g. last zero before 1/2). As a direct consequence we also get that the set of such zeroes is uncountable, which one would not expect for a continuous function that is $\mathbb{P}$-a.s. not monotone on any interval (as per Proposition 5.2.4).

Theorem 5.2.14. *Let $(B_t)_{t\geq 0}$ be a one dimensional standard Brownian motion and*

$$A = \{t \geq 0 : B_t = 0\}$$

its zero set. Then A is almost surely a closed set with no isolated points.

Proof. Since Brownian motion is almost surely continuous, we have that with probability one, A is closed. To prove the second property, take for each $q \in \mathbb{Q} \cap [0, \infty)$ the first zero after q, i.e. $\tau_q = \inf\{t \geq q : B_t = 0\}$ and note that τ_q is an almost surely finite stopping time. Since A is almost surely closed, the infimum is achieved and is therefore almost surely a minimum. By the strong Markov property applied to τ_q, we have that for each q almost surely τ_q is not an isolated zero from the right. Therefore, almost surely for all of the countably many rationals q, τ_q is not isolated from the right. For $0 < t \in A$ that are different from τ_q for all $q \in \mathbb{Q}$ we now show they are not isolated from the left. To see this, take $\mathbb{Q} \ni q_n \uparrow t$ and consider $t_n := \tau_{q_n}$. It holds that $q_n \leq t_n < t$ and therefore $t_n \uparrow t$, so t is not isolated from the left. □

Corollary 5.2.15. *Let A be the zero set of standard Brownian motion. Then A is uncountable.*

Proof. The uncountability of A follows from the known result that perfect sets are uncountable, which we now prove. We will do this by constructing a subset of A with cardinality $\{1, 2\}^{\mathbb{N}}$ which is known to be uncountable. Start by by choosing a point $x_1 \in A$. As this point is not isolated there exists a second point $x_2 \in A$. Take now two disjoint balls B_1 and B_2 around these two points. As x_1 is not isolated, we can find two points in $B_1 \cap A$ around which we can place two disjoint balls contained in $B_1$with no more than half its radius; we do similarly for $B_2 \cap A$ and so on. There now exists a bijection between $\{1, 2\}^{\mathbb{N}}$ and the decreasing sequence of balls. As A is complete the intersection of the balls in each such sequence contains at least one point of A and two points belonging to two different sequences are different. This proves our claim. □

5.3 Local central limit theorem

The presentation of this section is strongly guided by Section 2 of [LL10].
The local central limit theorem is important in many contexts in probability and arises quite naturally from a finer investigation of what's happening at microscopic scales in the standard central limit theorem (see Theorem [Dre19, Theorem 3.8.1]) applied to certain random walks.
Let (X_n) be an i.i.d. sequence of $\mathbb{Z}^d$-valued random variables and denote by $S_n := \sum_{k=1}^{n} X_k$ the corresponding random walk. For simplicity denote by p_n the distribution $\mathbb{P}(S_n \in \cdot)$ of S_n on $(\mathbb{Z}^d, \mathcal{B}(\mathbb{Z}^d))$.
Furthermore, denote by

$$\overline{p}_n(x) := \frac{1}{(2\pi n)^{\frac{d}{2}} \sqrt{\det \Sigma}} e^{-\frac{(x, \Sigma^{-1} x)}{2n}}$$

the probability density of a $\mathcal{N}(0, n\Sigma)$-distributed $\mathbb{R}^d$-valued normal random variable. As you know from the basic probability theory lectures, this is in fact the distribution of the sum of n independent $\mathcal{N}(0, \Sigma)$-distributed random variables. Also, recall that the gist of the central limit theorem was that asymptotically, for partial sums S_n of suitable random variables (keeping notation one-dimensional for simplicity),

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}(S_n)}}$$

is well-approximated by a $\mathcal{N}(0,1)$-variable. While this is in some sense a 'macroscopic' result (since space is rescaled by $\sqrt{\mathrm{Var}(S_n)}$ which is of order $\sqrt{n}$), the local central limit theorem gives insight into the 'microscopic ' behaviour in the sense that $S_n - \mathbb{E}[S_n]$ is well approximated by a $\mathcal{N}(0, \mathrm{Var}(S_n))$-distributed random variables (heuristically, this is a multiplication of the scale in the central limit theorem by $\sqrt{\mathrm{Var}(S_n)}$).

Theorem 5.3.1 (Local central limit theorem (LCLT), Theorem 2.3.5 in [LL10])**.** *Let $S_n := \sum_{k=1}^n X_k$ be a random walk such that $\mathbb{E}[X_1] = 0 \in \mathbb{R}^d$, and $\mathrm{Var}(\pi_i(X_1)) \in (0,\infty)$ for all $i \in \{1,\dots,d\}$ (where π_i denotes the projection on the i-th coordinate), and such that*

$$\mathbb{E}[|X_1|^3] < \infty. \tag{5.3.1}$$

Furthermore, assume that (S_n) is an aperiodic and irreducible random walk on $\mathbb{Z}^d$, i.e., for each $x \in \mathbb{Z}^d$ there exists $N_x \in \mathbb{N}$ such that $p_n(x) > 0$ for all $n \geq N_x$.
Then, with the above notation, there exists a constant $C \in (0,\infty)$ such that for all $n \in \mathbb{N}$ and $x \in \mathbb{Z}^d$,

$$|p_n(x) - \overline{p}_n(x)| \leq \frac{C}{n^{(d+1)/2}}.$$

As in the case of the standard central limit theorem (as well as the one for martingales), the proof will be based on computations of characteristic functions. Indeed, we have the following auxiliary result.

Lemma 5.3.2 (Lemma 2.3.3 of [LL10])**.** *Let $X := X_1$ with X_1 as in Theorem 5.3.1. Using the Taylor expansion of the exponential function around 0 in combination with (5.3.1), we write the characteristic function as*

$$\varphi_X(t) = 1 - \frac{(t, \Sigma t)}{2} + g(t), \quad t \in \mathbb{R}^d, \tag{5.3.2}$$

for some function $g : \mathbb{R}^d \to \mathbb{C}$ with

$$g(t) \in O(|t|^3) \quad \text{as } t \to 0. \tag{5.3.3}$$

Then there exist $\varepsilon, C \in (0,\infty)$ such that for all $n \in \mathbb{N}$, $|t| \leq \varepsilon\sqrt{n}$, we have

$$\Big(\varphi_X\Big(\frac{t}{\sqrt{n}}\Big)\Big)^n = \exp\Big\{-\frac{(t,\Sigma t)}{2} + h(t,n)\Big\},$$

where

$$|h(t,n)| \leq \min\Big\{\frac{(t,\Sigma t)}{4}, n\Big|g\Big(\frac{t}{\sqrt{n}}\Big)\Big| + \frac{C|t|^4}{n}\Big\}. \tag{5.3.4}$$

Proof. From (5.3.2) we infer by taking the logarithm and using its second order Taylor expansion around 1 that

$$\ln(\varphi_X(t)) = -\frac{(t,\Sigma t)}{2} + g(t) - \frac{(t,\Sigma t)^2}{8} + O(|g(t)|\,|t|^2) + O(|t|^6).$$

Hence, multiplying this display by n and at the same time replacing t by $t/\sqrt{n}$, we infer that

$$n\ln\Big(\varphi_X\Big(\frac{t}{\sqrt{n}}\Big)\Big) = -\frac{(t,\Sigma t)}{2} + \underbrace{ng\Big(\frac{t}{\sqrt{n}}\Big) - \frac{(t,\Sigma t)^2}{8n} + O\Big(\Big|g\Big(\frac{t}{\sqrt{n}}\Big)\Big|\cdot|t|^2\Big) + n^{-2}O(|t|^6)}_{=:h(t,n)}.$$

From this definition of h, in combination with (5.3.3) it immediately follows that we can find $\varepsilon > 0$ such that for all t with $|t| \le \varepsilon\sqrt{n}$, we have that $|h(t,n)|$ is bounded from above by the second term in the minimum of the right-hand side of (5.3.4). Choosing ε even smaller, we also get that $|h(t,n)|$ is smaller than the right-hand side of (5.3.4) by using that $g(t) \in o(|t|^2)$.

$\square$

Lemma 5.3.3 (Lemma 2.3.4 of [LL10])**.** *Assume the setting and notation from Lemma 5.3.2. Then there exist $C', \gamma \in (0,\infty)$ such that for all $r \in [0, \varepsilon\sqrt{n}]$, defining $\delta_n(x,r)$ via*

$$p_n(x) = \overline{p}_n(x) + \delta_n(x,r) + \frac{1}{(2\pi)^d n^{d/2}} \int_{|t|\le r} \exp\Big\{-\frac{\mathrm{i}x\cdot t}{\sqrt{n}} - \frac{(t,\Sigma t)}{2}\Big\}(e^{h(t,n)} - 1)\lambda^d(\mathrm{d}t),$$

then

$$|\delta_n(x,r)| \le C' n^{-\frac{d}{2}} e^{-\gamma r^2}.$$

For the proof of Lemma 5.3.3 we recall the Fourier inversion formula Corollary 3.8.6 of [Dre19]. While a standing assumption in that result was that the measure to which it was applied had a density with respect to λ^d, the result can be extended to cover the case of discrete random variables, and we can therefore infer, using that the characteristic function 'turns sums of independent random variables into products', that we have

$$p_n(x) = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi)^d} \varphi_{X_1}^n(t) e^{-\mathrm{i}t\cdot x}\, \lambda^d(\mathrm{d}t) \tag{5.3.5}$$

(see e.g. display (2.12) in [LL10]).
Also, we remark that for each $\varepsilon > 0$,

$$\sup\Big\{|\varphi(t)| \,:\, t \in [-\pi,\pi]^d,\, |t| \ge \varepsilon\Big\} < 1, \tag{5.3.6}$$

as well as the fact that

$$\frac{1}{(2\pi)^d n^{\frac{d}{2}}} \int_{\mathbb{R}^d} e^{-\mathrm{i}t\cdot x/\sqrt{n}} e^{-\frac{(t,\Sigma t)}{2}}\, \lambda^d(\mathrm{d}t) = \overline{p}_n(x), \tag{5.3.7}$$

which follows from explicit calculation (see e.g. display (2.2) in [LL10]).
Scrutinising (5.3.7) we can upper bound part of the integral via

$$\Big|\frac{1}{(2\pi)^d n^{\frac{d}{2}}} \int_{|t|\ge\varepsilon\sqrt{n}} e^{-\mathrm{i}t\cdot x/\sqrt{n}} e^{-\frac{(t,\Sigma t)}{2}}\, \lambda^d(\mathrm{d}t)\Big| \le \frac{1}{(2\pi)^d n^{\frac{d}{2}}} \int_{|t|\ge\varepsilon\sqrt{n}} e^{-\frac{(t,\Sigma t)}{2}}\, \lambda^d(\mathrm{d}t) \le O(e^{-\widetilde{\varepsilon}n}), \tag{5.3.8}$$

which turns out to be negligible as $n \to \infty$.

Proof of Lemma 5.3.3. In order to employ the asymptotic behaviour established in (5.3.2), we use (5.3.5) and the transformation formula to infer that

$$p_n(x) = \frac{1}{(2\pi)^d} \int_{[-\pi,\pi)^d} \varphi_X^n(t) e^{-\mathrm{i}t\cdot x}\, \lambda^d(\mathrm{d}t) = \frac{1}{(2\pi)^d n^{\frac{d}{2}}} \int_{[-\sqrt{n}\pi,\sqrt{n}\pi)^d} \varphi_X^n(t/\sqrt{n}) e^{-\mathrm{i}t\cdot x/\sqrt{n}}\, \lambda^d(\mathrm{d}t).$$

In combination with (5.3.6) we find that for $\varepsilon > 0$ there exists $c > 0$ such that $|\varphi_X(t)| \le e^{-c}$ for all $t \in [-\pi, \pi)^d$ with $|t| \ge \varepsilon$. Hence, the previous display can be continued as

$$= \frac{1}{(2\pi)^d n^{\frac{d}{2}}} \int\limits_{|t| \le \varepsilon\sqrt{n}} \varphi_X^n(t/\sqrt{n}) e^{-\mathrm{i}t\cdot x/\sqrt{n}} \, \lambda^d(\mathrm{d}t) + O((e^{-c})^n). \tag{5.3.9}$$

Using Lemma 5.3.2 we have

$$\varphi_X^n(t/\sqrt{n}) = e^{-\frac{(t,\Sigma t)}{2}} + e^{-\frac{(t,\Sigma t)}{2}}(e^{h(n,x)} - 1).$$

As a consequence, plugging this into (5.3.9) and combining it with (5.3.8) and (5.3.7), we infer that

$$p_n(x) = \overline{p}_n(x) + O(e^{-\widetilde{\varepsilon}n}) + \frac{1}{(2\pi)^d n^{d/2}} \int_{|t| \le \widetilde{\varepsilon}\sqrt{n}} \exp\Big\{ - \frac{\mathrm{i}x \cdot t}{\sqrt{n}} - \frac{(t, \Sigma t)}{2} \Big\} (e^{h(t,n)} - 1) \, \lambda^d(\mathrm{d}t),$$

which proves the result for $r = \widetilde{\varepsilon}\sqrt{n}$. Otherwise, take advantage of

$$|e^{h(t,n)} - 1| \le e^{\frac{(t,\Sigma t)}{4}} + 1, \tag{5.3.10}$$

which follows from (5.3.4). From this we then get

$$\Big| \int\limits_{r \le |t| \le \varepsilon\sqrt{n}} \exp\Big\{ - \frac{\mathrm{i}x \cdot t}{\sqrt{n}} - \frac{(t, \Sigma t)}{2} \Big\} (e^{h(t,n)} - 1) \, \lambda^d(\mathrm{d}t) \Big| \le 2 \int_{|t| \ge r} e^{-\frac{(t,\Sigma t)}{4}} \, \lambda^d(\mathrm{d}t) = O(e^{-\delta r^2}),$$

which finishes the proof. □

Proof of Theorem 5.3.1. Choosing $r = n^{\frac{1}{8}}$ in Lemma 5.3.3 we deduce

$$p_n(x) = \overline{p}_n(x) + O(e^{-\alpha n^{\frac{1}{4}}}) + \int\limits_{|t| \le n^{\frac{1}{8}}} \exp\Big\{ - \frac{\mathrm{i}x \cdot t}{\sqrt{n}} - \frac{(t, \Sigma t)}{2} \Big\} (e^{h(t,n)} - 1) \, \lambda^d(\mathrm{d}t)$$

where for $|t| \le n^{\frac{1}{8}}$ we have with (5.3.3) and (5.3.4) that

$$|h(t,n)| \le C|t|^3 n^{-\frac{1}{2}},$$

and thus $|e^{h(t,n)} - 1| \le C|t|^3/\sqrt{n}$, which again implies that

$$\int\limits_{|t| \le n^{\frac{1}{8}}} \exp\Big\{ - \frac{\mathrm{i}x \cdot t}{\sqrt{n}} - \frac{(t, \Sigma t)}{2} \Big\} (e^{h(t,n)} - 1) \, \lambda^d(\mathrm{d}t) \le \frac{C}{\sqrt{n}} \int\limits_{\mathbb{R}^d} |t|^3 e^{-\frac{(t,\Sigma t)}{2}} \, \lambda^d(\mathrm{d}t) \le C/\sqrt{n},$$

and finishes the proof. □

A more profound treatment of this important coupling can be found in [LL10, Chapter 7].

Acknowledgement: I would like to thank students of this class, as well as Alexei Prévorst and Lars Schmitz, for several useful comments and for pointing out a variety of typos. Finally, many thanks to Prof. Alexander Drewitz for kindly letting me base this lecture on his lecture notes.

Bibliography

[Bau92] Heinz Bauer. *Maß- und Integrationstheorie*. de Gruyter Lehrbuch. [de Gruyter Textbook]. Walter de Gruyter & Co., Berlin, second edition, 1992.

[Ben38] Frank Benford. The law of anomalous numbers. *Proceedings of the American Philosophical Society*, 78(4):551–572, 1938.

[BH$^+$11] Arno Berger, Theodore P Hill, et al. A basic theory of Benford's Law. *Probability Surveys*, 8(1):1–126, 2011.

[Bol68] L Boltzmann. Lösung eines mechanischen problems. *Wiener Berichte*, 58:1035–1044, 1868.

[DGRS12] Alexander Drewitz, Jürgen Gärtner, Alejandro F. Ram´irez, and Rongfeng Sun. Survival probability of a random walk among a Poisson system of moving traps. In *Probability in complex physical systems*, volume 11 of *Springer Proc. Math.*, pages 119–158. Springer, Heidelberg, 2012.

[Dre18] Alexander Drewitz. Introduction to probability and statistics (lecture notes), 2018.

[Dre19] Alexander Drewitz. Probability theory I (lecture notes), 2019.

[Dud02] R. M. Dudley. *Real analysis and probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. Revised reprint of the 1989 original.

[Dur10] Rick Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.

[Fad85] Arnold M. Faden. The existence of regular conditional probabilities: necessary and sufficient conditions. *Ann. Probab.*, 13(1):288–298, 1985.

[KÏ6] Wolfgang König. *The parabolic Anderson model*. Pathways in Mathematics. Birkhäuser/Springer, [Cham], 2016. Random walk in random potential.

[Kal02] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[Kin73] J. F. C. Kingman. Subadditive ergodic theory. *Ann. Probab.*, 1:883–909, 1973. With discussion by D. L. Burkholder, Daryl Daley, H. Kesten, P. Ney, Frank Spitzer and J. M. Hammersley, and a reply by the author.

[Kle14] Achim Klenke. *Probability theory*. Universitext. Springer, London, second edition, 2014. A comprehensive course.

[KS91] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.

[L92] Paul Lévy. *Processus stochastiques et mouvement brownien.* Les Grands Classiques Gauthier-Villars. [Gauthier-Villars Great Classics]. Éditions Jacques Gabay, Sceaux, 1992. Followed by a note by M. Loève, Reprint of the second (1965) edition.

[Lig85] Thomas M. Liggett. An improved subadditive ergodic theorem. *Ann. Probab.*, 13(4):1279–1285, 1985.

[LL10] Gregory F. Lawler and Vlada Limic. *Random walk: a modern introduction*, volume 123 of *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge, 2010.

[LPW09] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times.* American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.

[MP10] Peter Mörters and Yuval Peres. *Brownian motion.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010. With an appendix by Oded Schramm and Wendelin Werner.

[Str11] Daniel W. Stroock. *Probability theory.* Cambridge University Press, Cambridge, second edition, 2011. An analytic view.

[SV79] Daniel W. Stroock and S. R. Srinivasa Varadhan. *Multidimensional diffusion processes*, volume 233 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, 1979.

[Var01] S. R. S. Varadhan. *Probability theory*, volume 7 of *Courant Lecture Notes in Mathematics.* New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2001.

[Wil91] David Williams. *Probability with martingales.* Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.

Appendix A

Alternative constructions and proofs

A.1 Conditional expectation: a different approach

Theorem A.1.1. *Let $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. Then, for any sub-$\sigma$-algebra $\mathcal{G} \subset \mathcal{F}$ there exists a $\mathcal{G}$-$\mathcal{B}(\mathbb{R})$-measurable Y with $\mathbb{E}[Y^2] < \infty$ satisfying $\mathbb{E}[\mathbb{1}_G Y] = \mathbb{E}[\mathbb{1}_G X]$ for all $G \in \mathcal{G}$.*

Proof. Consider the norm $\|Y\|_2 := \mathbb{E}[Y^2]^{1/2}$ on the vector space $V = \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ of real valued random variables Y with finite second moment, with inner product given by

$$\langle Y_1, Y_2 \rangle = \mathbb{E}[Y_1 Y_2].$$

As $\mathcal{L}^p$ spaces are complete, this makes V into a *Hilbert space*. Then, $U = \mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ is a complete and therefore closed subspace of V.
By the existence of orthogonal projections onto closed subspaces of Hilbert spaces, there exists an orthogonal projection $\pi : V \to U$. In particular, $\langle \pi Y - Y, Z \rangle = 0$ for all $Y \in V$ and $Z \in U$. Setting $Y = \pi X$ gives

$$\mathbb{E}[\mathbb{1}_G Y] - \mathbb{E}[\mathbb{1}_G X] = \langle \mathbb{1}_G, \pi X - X \rangle = 0$$

as required. □

Corollary A.1.2. *Let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. Then, for any sub-$\sigma$-algebra $\mathcal{G} \subset \mathcal{F}$ there exists a $\mathcal{G}$-$\mathcal{B}(\mathbb{R})$-measurable Y with $\mathbb{E}[Y] < \infty$ satisfying $\mathbb{E}[\mathbb{1}_G Y] = \mathbb{E}[\mathbb{1}_G X]$ for all $G \in \mathcal{G}$.*

Proof. Let $X = X^+ - X^-$, where $X^\pm$ are non-negative, square integrable random variables. Define $X_n^\pm = \min\{X^\pm, n\}$. Since $X_n^\pm$ is bounded, we have by Theorem A.1.1 that $Y_n^\pm = \mathbb{E}[X^\pm \,|\, \mathcal{G}]$ exists and is unique a.s. Using the same steps as in the proof of Theorem 1.0.5 for non-integrable X, we obtain the existence of $Y^\pm$ and therefore also $Y = Y^+ - Y^-$. □

A.2 Martingales

Lemma A.2.1. *A positive random variable τ is integrable if $\exists N \in \mathbb{N}, \epsilon > 0$ so that*

$$\mathbb{P}(\tau \le n + N \,|\, \tau > n) \ge \epsilon \quad \forall n \in \mathbb{N}_0$$

Proof. We calculate

$$\begin{aligned}\mathbb{E}[\tau] &= \int_0^\infty \mathbb{P}(\tau > t) dt = \sum_{\ell \in \mathbb{N}_0} \int_{\ell N}^{(\ell+1)N} \mathbb{P}(\tau > t) dt \le \sum_{\ell \in \mathbb{N}_0} \mathbb{P}(\tau > \ell N) \int_{\ell N}^{(\ell+1)N} dt \\ &\le \sum_{\ell \in \mathbb{N}_0} (1 - \epsilon)^\ell N = \frac{N}{1 - (1 - \epsilon)} = \frac{N}{\epsilon} < \infty,\end{aligned}$$

which gives the desired result, once we argue the validity of the inequality

$$\mathbb{P}(\tau > \ell N) \leq (1-\epsilon)^{\ell}.$$

We prove this inequality by induction. For $\ell = 0$ the inequality trivially holds. For general ℓ, note that $\{\tau > (\ell+1)N\} \subseteq \{\tau > \ell N\}$, so we can write

$$\begin{aligned}\mathbb{P}(\tau > (\ell+1)N) &= \mathbb{P}(\tau > (\ell+1)N, \tau > \ell N) \\ &= \mathbb{P}(\tau > \ell N)\mathbb{P}(\tau > (\ell+1)N \mid \tau > \ell N).\end{aligned}$$

The assumption of the lemma gives $\mathbb{P}(\tau > n + N \mid \tau > n) \leq (1-\epsilon)$. Choosing $n = \ell N$ yields $\mathbb{P}(\tau > (\ell+1)N \mid \tau > \ell N) \leq (1-\epsilon)$, which combined with the induction assumption $\mathbb{P}(\tau > \ell N) \leq (1-\epsilon)^{\ell}$ yields the desired inequality. $\square$