Recommended books:


Contents

1 Introduction 3
2 Measure-theoretic probability 3
3 Important inequalities 8
4 Independence 9
  4.1 Independent events and random variables 9
  4.2 Existence of independent random variables with given distributions, product measure 10
  4.3 Products of independent random variables 12
5 Convergence of random variables and distributions 13
6 Laws of large numbers 15
7 Conditional expectation 15
8 Stochastic processes and martingales 20
9 Stopping times 23
10 Maximal inequalities 25
11 Martingale convergence theorems 26
  11.1 Doob’s downcrossing inequality 26
  11.2 Doob’s convergence theorem 29
  11.3 Uniformly integrable martingales 30
12 Backward martingales, sub- and supermartingales 35
13 Applications of the martingale convergence theorems 36
14 Construction of stochastic processes with prescribed properties 38

*Notes of a course given at the University of Luxembourg during the winter semesters 2013/2014, 2014/2015 and 2015/2016.*
15 Markov kernels, semigroups and processes

16 Brownian motion and Kolmogorov’s continuity theorem
  16.1 Brownian motion ......................................................... 48
  16.2 Modifications and indistinguishable processes .................. 51
  16.3 Kolmogorov’s continuity theorem ..................................... 52
  16.4 Brownian motion is a martingale .................................... 56

17 Gaussian processes ....................................................... 57

18 Path properties of Brownian motion .................................... 64
1 Introduction

The word “martingale” has many different meanings, see e.g. http://en.wikipedia.org/wiki/Martingale. In mathematics, and more precisely probability theory, it stands for a mathematical model of a “fair game”. What does that mean?

Example 1.1 (Coin tossing). We repeatedly toss a fair coin. We win 1 EUR if the result is heads, and lose 1 EUR if it is tails. Mathematically, let $\xi_i = 1$ if the result of the $i$-th toss is heads, and $\xi_i = -1$ if it is tails. Then the random variables $\xi_i$ are independent and identically distributed with $P[\xi_i = 1] = P[\xi_i = -1] = 1/2$. We are interested in the total gain after the $n$-th toss, $X_n := \sum_{i=1}^{n} \xi_i$.

Intuitively, the sequence $(X_n)_{n \in \mathbb{N}}$ is an example of a “fair game”. What does that mean?

First observation: $E(X_n) = \sum_{i=1}^{n} E(\xi_i) = 0$, so that the expectation of $X_n$ is constant. This property however is not enough to call a game “fair”. If for example we define $Y_n := n\xi_1$, we also have $E(Y_n) = nE(\xi_1) = 0$, but once we know $Y_1$ we would not consider the game to be fair any more.

Preliminary definition: We say that a sequence $(X_n)_{n \in \mathbb{N}}$ of random variables is a fair game if for all $n \in \mathbb{N}$ and all $a_1, \ldots, a_n \in \mathbb{R}$ such that $P[X_1 = a_1, \ldots, X_n = a_n] > 0$ we have $E[X_n+1 | X_1 = a_1, \ldots, X_n = a_n] = a_n$.

Let us check whether our two sequences $(X_n)_{n \in \mathbb{N}}$ and $(Y_n)_{n \in \mathbb{N}}$ are fair games in the sense of this definition: we first observe that $E[Y_2 | Y_1 = -1] = -2 < -1$, so that the sequence $(Y_n)_{n \in \mathbb{N}}$ is not fair. (Once we observe that $Y_1 = -1$, we would like to stop the game.) On the other hand,

$$E[X_{n+1} | X_1 = a_1, \ldots, X_n = a_n] = E[X_n + \xi_{n+1} | X_1 = a_1, \ldots, X_n = a_n] = a_n + E[\xi_{n+1} | X_1 = a_1, \ldots, X_n = a_n] = a_n + E[\xi_{n+1}] = a_n,$$

so that the sequence $(Y_n)_{n \in \mathbb{N}}$ is fair.

Problem: In many cases (for instance if the random variables $X_n$ have continuous distributions), the probability $P[X_1 = a_1, \ldots, X_n = a_n]$ is 0 for all $a_1, \ldots, a_n \in \mathbb{R}$. Here we arrive at the limits of elementary probability, and therefore we introduce measure-theoretic probability.

2 Measure-theoretic probability

In elementary probability theory a probability space is a pair $(\Omega, p)$ where $\Omega$ is a countable set and $p : \Omega \to \mathbb{R}_+$ a function with $\sum_{\omega \in \Omega} p(\omega) = 1$; $p(\omega)$ is the probability of the elementary event $\omega \in \Omega$. The probability of a general event $A \subset \Omega$ is then $P(A) = \sum_{\omega \in A} p(\omega)$. The function $P : \mathcal{P}(\Omega) \to [0, 1]$ has the following properties:
1. For all singletons, \( P(\{\omega\}) = p(\omega) \).
2. \( P(\emptyset) = 0 \).
3. Whenever \((A_i)_{i \in I}\) is a collection of pairwise disjoint subsets of \( \Omega \), we have
\[
P\left( \bigcup_{i \in I} A_i \right) = \sum_{i \in I} P(A_i). \tag{1}\]
4. \( P(\Omega) = 1 \).

**Example 2.1.** We choose a point “uniformly” from the unit cube \([0, 1]^3\), in the sense that
\[
P(A) = \operatorname{vol}(A)
\]
for all \( A \subseteq [0, 1]^3 \). Two problems arise:
1. On the one hand we have \( P(\{\omega\}) = 0 \) for all \( \omega \in [0, 1]^3 \), and consequently
\[
\sum_{\omega \in [0,1]^3} P(\{\omega\}) = 0;
\]
on the other hand
\[
P\left( \bigcup_{\omega \in [0,1]^3} \{\omega\} \right) = P([0,1]^3) = 1 \neq 0;
\]
consequently (1) is not satisfied.
2. The *Banach-Tarski paradox*\(^1\) shows that it is not possible to define the volume of all subsets of \([0, 1]^3\) in a sensible way.

The solution to this sort of problems is to use *measure theory*.

**Definition 2.2.** A *probability space* is a measure space \((\Omega, \mathcal{A}, P)\) with \( P(\Omega) = 1 \). This means:

1. \((\Omega, \mathcal{A})\) is a *measurable space*, i.e. \( \Omega \) is a set (not necessarily countable) and \( \mathcal{A} \) a \( \sigma \)-field on \( \Omega \), i.e. a collection of subsets of \( \Omega \) satisfying
   (a) \( \emptyset \in \mathcal{A} \),
   (b) for every countable collection \((A_i)_{i \in I}\) of elements of \( \mathcal{A} \) we have \( \bigcup_{i \in I} A_i \in \mathcal{A} \),
   (c) for every \( A \in \mathcal{A} \) we have \( A^c \in \mathcal{A} \).
2. \( P \) is a *measure* on \((\Omega, \mathcal{A})\), i.e. a function \( P : \mathcal{A} \to [0,\infty] \) satisfying
   (a) \( P(\emptyset) = 0 \),
   (b) whenever \((A_i)_{i \in I}\) is a countable collection of pairwise disjoint subsets of \( \Omega \), we have \( P\left( \bigcup_{i \in I} A_i \right) = \sum_{i \in I} P(A_i) \).
3. \( P(\Omega) = 1 \).

\(^1\)S. Banach, A. Tarski, Sur la décomposition des ensembles de points en parties respectivement congruentes. Fund. Math 6 (1924), 244–277. (Stefan Banach (1892–1945), Polish mathematician; Alfred Tarski (1901–1983), Polish mathematician and philosopher)
Definition 2.3. Let $(\Omega, \mathcal{A}, P)$ be a probability space and $(E, \mathcal{B})$ a measurable space.

1. An $(E, \mathcal{B})$-valued random variable on $(\Omega, \mathcal{A}, P)$ is a map $X : \Omega \to E$ which is measurable with respect to $\mathcal{A}$ and $\mathcal{B}$.

2. The distribution of $X$ is the image measure of $P$ under $X$, i.e. the probability measure $P_X$ on $(E, \mathcal{B})$ defined by $P_X(B) := P\{X \in B\}$.

Remark 2.4.

1. If $E$ is a topological space, we always use its Borel $\sigma$-field $\mathcal{B} = \mathcal{B}(E)$ (i.e. the $\sigma$-field generated by the topology of $E$) unless otherwise specified.

2. If $E$ is not specified, the word “random variable” always means “real-valued random variable”, i.e. $E = \mathbb{R}$.

Definition 2.5. Let $X$ be a random variable defined on a probability space $(\Omega, \mathcal{A}, P)$. The expectation of $X$ is its Lebesgue integral:

$$E(X) := \int_\Omega X dP,$$

provided that this integral exists. (If $\int_\Omega X_+ dP$ and $\int_\Omega X_- dP$ are both infinite, the Lebesgue integral and hence the expectation of $X$ is not defined.)

Remark 2.6. The Lebesgue integral of a random variable $X$ is defined in three steps:

1. If $X$ is an elementary or simple random variable, i.e. of the form $X = \sum_{i=1}^k \alpha_i 1_{A_i}$ with $\alpha_i \in \mathbb{R}$ and $A_i \in \mathcal{A}$, one sets

$$\int_\Omega X dP := \sum_{i=1}^k \alpha_i P(A_i).$$

One has to check that this is well-defined, i.e. independent of the particular representation of $X$.

2. If $X$ is nonnegative, there exists a non-decreasing sequence $(X_n)_{n \in \mathbb{N}}$ of nonnegative elementary random variables converging pointwise to $X$, i.e. $0 \leq X_1(\omega) \leq X_2(\omega) \leq \ldots$ and $X_n(\omega) \to X(\omega)$ for all $\omega \in \Omega$. One then sets

$$\int_\Omega X dP := \lim_{n \to \infty} \int_\Omega X_n dP$$

and must again show that this is well-defined, i.e. independent of the choice of the sequence $(X_n)_{n \in \mathbb{N}}$.

3. In the general case one sets

$$\int_\Omega X dP := \int_\Omega X_+ dP - \int_\Omega X_- dP,$$

provided that at least one of these two integrals is finite. If they are both infinite, the Lebesgue integral of $X$ is not defined.

---

$^2$Émile Borel (1871–1956), French mathematician

Definition 2.7. Let $X$ be a random variable defined on a probability space $(\Omega, \mathcal{A}, P)$.

1. We say that $X$ is quasi-integrable if $E(X)$ exists, i.e. if $E(X_+)$ or $E(X_-)$ is finite.

2. We say that $X$ is integrable if $E(X)$ exists and is finite. Clearly, this holds if and only if $E(X_+)$ and $E(X_-)$ are both finite or, equivalently, if $E(|X|) < \infty$.

3. Let $p \geq 1$. We say that $X$ belongs to $L^p$ if $E(|X|^p) < \infty$. If $X$ belongs to $L^2$ we say that $X$ is square-integrable.

Remark 2.8. An often used sufficient criterion for the existence of $E(X)$ is that $X$ is non-negative or integrable. Such random variables are sometimes called admissible.

Remark 2.9. In general the pointwise convergence of a sequence $(X_n)_{n\in\mathbb{N}}$ to another random variable $X$ does NOT imply the convergence of $E(X_n)$ to $E(X)$.

Example 2.10. Let $\Omega := [0, 1]$, $\mathcal{A}$ its Borel $\sigma$-field, $P$ the Lebesgue measure and $X_n := n^2 \cdot 1_{(0,1/n)}$. Then $X_n(\omega) \to 0$ for all $\omega \in \Omega$, but $E(X_n) = n \to \infty$.

Proposition 2.11 (Beppo Levi’s\footnote{Beppo Levi (1875–1961), Italian mathematician} monotone convergence theorem). Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables and $X$ another random variable, all defined on the same probability space $(\Omega, \mathcal{A}, P)$. If $0 \leq X_n \nearrow X$, then $E(X_n) \to E(X)$.

Proposition 2.12 (Lebesgue’s dominated convergence theorem). Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables and $X$ another random variable, all defined on the same probability space $(\Omega, \mathcal{A}, P)$. If the sequence $(X_n)_{n\in\mathbb{N}}$ converges pointwise to $X$ and if moreover there exists an integrable random variable $X'$ on $(\Omega, \mathcal{A}, P)$ such that $|X_n| \leq X'$ for all $n \in \mathbb{N}$, then $E(X_n) \to E(X)$.

Proposition 2.13 (Fatou’s lemma\footnote{P. Fatou, Séries trigonométriques et séries de Taylor. Acta Math. 30 (1906), 335–400. (Pierre Fatou (1878–1929), French mathematician and astronomer)}). Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of nonnegative random variables. Then

$$E(\liminf_{n\to\infty} X_n) \leq \liminf_{n\to\infty} E(X_n).$$

Definition 2.14. The variance of $X$ is

$$\text{Var}(X) := \begin{cases} E[(X - E(X))^2] & \text{if } E(|X|) < \infty \\ +\infty & \text{if } E(|X|) = \infty. \end{cases}$$

Remark 2.15. If $E(|X|) < \infty$, one clearly has $\text{Var}(X) = E(X^2) - E(X)^2$.

Definition 2.16. The square root of the variance is called the standard deviation of $X$:

$$\sigma(X) := \sqrt{\text{Var}(X)}.$$

Definition 2.17. Let $X$ and $Y$ be two integrable random variables such that their product $XY$ is also integrable. Then their covariance is

$$\text{Cov}(X, Y) := E(XY) - E(X)E(Y).$$

If $\text{Cov}(X, Y) = 0$, $X$ and $Y$ are called uncorrelated.
Lemma 2.18 (Transformation lemma). Let \((\Omega, \mathcal{A}, P)\) be a probability space, \((E, \mathcal{B})\) a measurable space, \(X\) an \((E, \mathcal{B})\)-valued random variable on \((\Omega, \mathcal{A}, P)\) and \(f : E \to \mathbb{R}\) a measurable function. Then the expectation of the random variable \(f(X)\) exists if and only if the integral of \(f\) with respect to the measure \(P_X\) exists, and in this case one has

\[ E(f(X)) = \int_E f \, dP_X. \]

In particular, for \(E = \mathbb{R}\) and \(f(x) = x\) or \(f(x) = x^2\) we obtain

\[ E(X) = \int_{\mathbb{R}} x \, dP_X \]

and

\[ E(X^2) = \int_{\mathbb{R}} x^2 \, dP_X. \]

Consequently, the expectation and the variance of \(X\) can be computed without knowing the probability space \((\Omega, \mathcal{A}, P)\), provided one knows the distribution of \(X\).

Proof. We first assume that \(f\) is elementary, i.e. of the form \(f = \sum_{i=1}^{k} \alpha_i 1_{B_i}\) with \(\alpha_i \in \mathbb{R}\) and \(B_i \in \mathcal{B}\). Then \(f(X) = \sum_{i=1}^{k} \alpha_i 1_{\{X \in B_i\}}\) and consequently

\[ E(f(X)) = \sum_{i=1}^{k} \alpha_i P\{X \in B_i\} \]

\[ = \sum_{i=1}^{k} \alpha_i P_X(B_i) \]

\[ = \int_E f \, dP_X. \]

Now let \(f \geq 0\). Then we know that there exists a sequence \((f_n)_{n \in \mathbb{N}}\) of elementary functions such that \(0 \leq f_n \uparrow f\). It follows that \(0 \leq f_n(X) \uparrow f(X)\), and consequently

\[ E(f(X)) = \lim_{n \to \infty} E(f_n(X)) \]

\[ = \lim_{n \to \infty} \int_E f_n \, dP_X \]

\[ = \int_E f \, dP_X. \]

Finally let \(f\) be arbitrary. Then we have

\[ E(f_+(X)) = \int_E f_+ \, dP_X \]

and

\[ E(f_-(X)) = \int_E f_- \, dP_X, \]

and the claim follows. \(\square\)

Remark 2.19. The proof used a typical technique called measure theoretic induction.
3 Important inequalities

Proposition 3.1 (Chebyshev’s inequality\(^6\)). For all random variables \( X \) and all \( \alpha, p > 0 \) one has

\[
P\{|X| \geq \alpha\} \leq \frac{1}{\alpha^p} E(|X|^p)
\]

In particular, replacing \( X \) with \( X - E(X) \) and choosing \( p = 2 \), one obtains

\[
P\{|X - E(X)| \geq \alpha\} \leq \frac{1}{\alpha^2} \text{Var}(X).
\]

Proof.  

\[
E(|X|^p) = \int_{\Omega} |X|^p dP
\geq \int_{\{|X| \geq \alpha\}} |X|^p dP
\geq \alpha^p P\{|X| \geq \alpha\}. \quad \Box
\]

Proposition 3.2 (Hölder’s inequality\(^7\)). Let \( X \) and \( Y \) be two random variables defined on the same probability space \((\Omega, A, P)\), and let \( p, q > 1 \) such that \( 1/p + 1/q = 1 \). Then

\[
E(|XY|) \leq E(|X|^p)^{1/p} E(|Y|^q)^{1/q}.
\]

Proof. See any book on measure and integration theory. \(\Box\)

Corollary 3.3. Let \( X \) be a random variable and \( q \geq p > 0 \). Then

\[
E(|X|^p) \leq E(|X|^q)^{p/q}.
\]

In particular, if \( X \) belongs to \( L^q \), then it also belongs to \( L^p \).

Proof. We may assume that \( q > p \), so that \( r := q/p > 1 \). Let \( s > 1 \) be such that \( 1/r + 1/s = 1 \). Then Hölder’s inequality implies that

\[
E(|X|^p) = E(|X|^p \cdot 1)
\leq E(|X|^{p/r})^{1/r} E(1^s)^{1/s}
\leq E(|X|^q)^{p/q}. \quad \Box
\]

Proposition 3.4 (Jensen’s inequality\(^8\)). Let \( X \) be an integrable random variable and \( q : \mathbb{R} \to \mathbb{R} \) a convex function. Then \( E(q(X)) \) exists and

\[
E(q(X)) \geq q(E(X)).
\]

---


Proof. Since $q$ is convex, we have for all $x, y \in \mathbb{R}$

$$q(y) \geq q(x) + q'_+(x)(y - x),$$  \hspace{1cm} (2)

where $q'_+(x)$ is the right-hand derivative of $q$ at $x$ (which exists since $q$ is convex). (The same statement is true if we replace $q'_+(x)$ with $q'_-(x)$, the left-hand derivative of $q$ at $x$.)

By choosing $x = E(X)$ and $y = X$ in (2) it follows that

$$q(X) \geq q(E(X)) + q'_+(E(X))(X - E(X)).$$  \hspace{1cm} (3)

By assumption, the right-hand side of this inequality is integrable, so that in particular the integral of its negative part is finite. Therefore $E(q(X)_-)$ is finite as well, so that $E(q(X))$ exists. Taking expectations on both sides of (3) yields the claim. \qed

4 Independence

"Probability theory = measure theory + independence."

4.1 Independent events and random variables

Definition 4.1. Let $(\Omega, \mathcal{A}, P)$ be a probability space and $I$ an arbitrary index set. A collection $(A_i)_{i \in I}$ of events (i.e. $A_i \in \mathcal{A}$ for all $i \in I$) is said to be independent if for every finite subset $J$ of $I$ one has

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i).$$ \hspace{1cm} (4)

Remark 4.2. One cannot ask (4) to hold for all subsets $J$ of $I$ because for general $J$ the set $\bigcap_{i \in J} A_i$ need not belong to $\mathcal{A}$. One may however replace the word “finite” with “countable”.

Definition 4.3.

1. A collection $(\mathcal{E}_i)_{i \in I}$ of sets of events (i.e. $\mathcal{E}_i \subseteq \mathcal{A}$ for all $i \in I$) is said to be independent if for every finite subset $J$ of $I$ and every choice of events $A_i \in \mathcal{E}_i$ ($i \in J$) holds

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i).$$

2. A collection $(X_i)_{i \in I}$ of random variables on $(\Omega, \mathcal{A}, P)$ ($X_i$ taking values in a measurable space $(E_i, \mathcal{B}_i)$) is independent if the family $(\sigma(X_i))_{i \in I}$ is independent, where $\sigma(X_i) \subseteq \mathcal{A}$ is the $\sigma$-field generated by $X_i$.

Remark 4.4. Since $\sigma(X_i) = \{\{X_i \in B_i\} \mid B_i \in \mathcal{B}_i\}$, the family $(X_i)_{i \in I}$ is independent if and only if for all choices $B_i \in \mathcal{B}_i$ ($i \in I$) the family of events $(\{X_i \in B_i\})_{i \in I}$ is independent, i.e. if for every finite subset $J$ of $I$ and every choice of sets $B_i \in \mathcal{B}_i$ ($i \in J$),

$$P\left(\bigcap_{i \in J} \{X_i \in B_i\}\right) = \prod_{i \in J} P\{X_i \in B_i\}. \hspace{1cm} (5)$$
4.2 Existence of independent random variables with given distributions, product measure

Let \((E_i, B_i)\) be a family of measurable spaces, and for each \(i \in I\) let \(\mu_i\) be a probability measure on \((E_i, B_i)\).

Question: Does there exist a probability space \((\Omega, A, P)\) and random variables \((X_i)_{i \in I}\) on it \((X_i\) taking values in \((E_i, B_i))\) such that

1. the family \((X_i)_{i \in I}\) is independent, and
2. the law of \(X_i\) equals \(\mu_i\) for all \(i \in I\)?

If \(I\) is finite, the answer is yes. Namely one can take:

- \(\Omega := \prod_{i \in I} E_i\) (the Cartesian product),
- \(X_i = \pi_i\), where \(\pi_i : \Omega \to E_i\) is the \(i\)th projection, defined by \(\pi_i((\omega_j)_{j \in I}) := \omega_i\).
- \(A := \bigotimes_{i \in I} B_i\) the product \(\sigma\)-field, i.e. the smallest \(\sigma\)-field on \(\Omega\) such that all projections \(\pi_i : \Omega \to E_i\) are measurable,
- \(P := \bigotimes_{i \in I} \mu_i\) the product measure, i.e. the unique measure \(P\) on \(\Omega\) such that for every choice of sets \((B_i \in B_i)_{i \in I}\)

\[
P \left( \prod_{i \in I} B_i \right) = \prod_{i \in I} \mu_i(B_i).
\]

In the general case one has to be more careful because then \(\prod_{i \in I} B_i\) need not belong to \(A\). Therefore we define:

**Definition 4.5.** A probability measure \(P\) on \((\Omega, A)\) is called a **product measure** of the probability measures \((\mu_i)_{i \in I}\) if for every non-empty finite subset \(J\) of \(I\) and every choice of sets \((B_i \in B_i)_{i \in J}\), one has

\[
P \left( \prod_{i \in J} B_i \times \prod_{i \in I \setminus J} E_i \right) = \prod_{i \in J} \mu_i(B_i).
\]

**Theorem 4.6.** There is always a unique product measure.

**Proof.** The basic idea is the same as for the case of finite \(I\), namely to apply Carathéodory’s\(^{10}\) extension theorem. To do so one has to show that

1. the set \(S\) of sets of the form \(\prod_{i \in J} B_i \times \prod_{i \in I \setminus J} E_i\) is a semi-ring, i.e.
   - (a) \(\emptyset \in S\),
   - (b) for all \(A, B \in S\), we have \(A \cap B \in S\), and
   - (c) for all \(A, B \in S\), there exist finitely many pairwise disjoint sets \(K_1, \ldots, K_n \in S\) such that \(A \setminus B = \bigcup_{i=1}^n K_i\), and
2. the function \(P : S \to \mathbb{R}_+\) is a pre-measure, i.e.

---

\(^9\)René Descartes (1596–1650), French philosopher and mathematician

\(^{10}\)Constantin Carathéodory (1873–1950), Greek mathematician
(a) \( P(\emptyset) = 0, \)
(b) for every countable collection \( (A_i)_{i \in I} \) of pairwise disjoint elements of \( S \) whose union \( \bigcup_{i \in I} A_i \) also belongs to \( S \), one has

\[
P \left( \bigcup_{i \in I} A_i \right) = \sum_{i \in I} P(A_i).
\]

Once one has shown this, the result follows immediately from Carathéodory’s extension theorem. Since the details are very technical, they are omitted. They can be found e.g. in the book *Wahrscheinlichkeitstheorie* by Bauer or the book *Probability with Martingales* by Williams.

**Remark 4.7.** A family \( (X_i)_{i \in I} \) of random variables (each \( X_i \) taking values in a measurable space \( (E_i, B_i) \)) can be regarded as one random variable \( X \) with values in the product space \( \prod_{i \in I} E_i, \bigotimes_{i \in I} B_i \). (Check that \( X \) is indeed measurable with respect to \( A \) and \( \bigotimes_{i \in I} B_i \)!) The joint distribution of the family \( (X_i)_{i \in I} \) is then the image measure of \( P \) under \( X \) (hence a probability measure on \( (\prod_{i \in I} E_i, \bigotimes_{i \in I} B_i) \)).

**Proposition 4.8.** A family \( (X_i)_{i \in I} \) of random variables (each \( X_i \) taking values in a measurable space \( (E_i, B_i) \)) is independent if and only if their joint distribution is equal to the product of their distributions.

**Proof.** Let us first assume that the variables \( (X_i)_{i \in I} \) are independent. To show that \( P_{(X_i)_{i \in I}} = \bigotimes_{i \in I} P_{X_i} \), it suffices to show that these two measures agree on every set of the form \( \prod_{i \in J} B_i \times \prod_{i \in I \setminus J} E_i \), where \( J \subseteq I \) is finite and \( B_i \in B_i \) for every \( i \in J \). Using the independence of the variables \( X_i \) (in the form of (5)) and the definition of the product measure, one obtains

\[
P_{(X_i)_{i \in I}} \left( \prod_{i \in J} B_i \times \prod_{i \in I \setminus J} E_i \right) = P \{ X_i \in B_i \quad \forall i \in J \}
\]

\[
= \prod_{i \in J} P \{ X_i \in B_i \}
\]

\[
= \prod_{i \in J} P_{X_i}(B_i)
\]

\[
= \left( \bigotimes_{i \in I} P_{X_i} \right) \left( \prod_{i \in J} B_i \times \prod_{i \in I \setminus J} E_i \right).
\]

Conversely, let us now assume that \( P_{(X_i)_{i \in I}} = \bigotimes_{i \in I} P_{X_i} \). Then we have to show that

\[
P \left( \bigcap_{i \in J} \{ X_i \in B_i \} \right) = \prod_{i \in J} P \{ X_i \in B_i \}
\]
for every finite subset \( J \) of \( I \) and every choice of sets \((B_i \in B_i)_{i \in J}\):

\[
P\left(\bigcap_{i \in J} \{X_i \in B_i\}\right) = P_{(X_i)_{i \in I}}\left(\prod_{i \in J} B_i \times \prod_{i \in I \setminus J} E_i\right)
= \bigotimes_{i \in I} P_{X_i}\left(\prod_{i \in J} B_i \times \prod_{i \in I \setminus J} E_i\right)
= \prod_{i \in J} P_{X_i}(B_i)
= \prod_{i \in J} P\{X_i \in B_i\}. \qed
\]

4.3 Products of independent random variables

**Proposition 4.9.** Let \( X_1, \ldots, X_n \) be independent random variables on the same probability space \((\Omega, \mathcal{A}, P)\) which are all nonnegative or all integrable. Then

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

Moreover, in the second case, the product \(\prod_{i=1}^{n} X_i\) is integrable as well.

**Proof.** Using the transformation lemma, the independence of the random variables \( X_i \) and Tonelli’s theorem\(^{11}\) we obtain

\[
E\left(\prod_{i=1}^{n} |X_i|\right) = \int_{\mathbb{R}^n} \prod_{i=1}^{n} |x_i| \, dP_{(X_1, \ldots, X_n)}
= \int_{\mathbb{R}^n} \prod_{i=1}^{n} |x_i| \, d\left(\bigotimes_{i=1}^{n} P_{X_i}\right)
= \prod_{i=1}^{n} \int_{\mathbb{R}} |x_i| \, dP_{X_i}
= \prod_{i=1}^{n} E(|X_i|),
\]

which shows the first statement for nonnegative variables as well as the second statement. Consequently, if the \( X_i \) are all integrable, we may drop the absolute value signs, and the third step in the above computation is now justified by Fubini’s\(^{12}\) instead of Tonelli’s theorem.

**Corollary 4.10.** Independent integrable random variables are pairwise uncorrelated.


\(^{12}\)G. Fubini, Sugli integrali multipli. *Rend. Acad. Lincei* 16 (1907), 608–614. (Guido Fubini (1879–1943), Italian mathematician)
5 Convergence of random variables and distributions

**Definition 5.1.** Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of $\mathbb{R}^d$-valued random variables and $X$ another $\mathbb{R}^d$-valued random variable, all defined on the same probability space $(\Omega, \mathcal{A}, P)$. We say that the sequence $(X_n)_{n \in \mathbb{N}}$ converges to $X$

1. **almost surely** if $P\{X_n \to X\} = 1,$
2. **in probability** if for all $\varepsilon > 0$ we have $\lim_{n \to \infty} P\{|X_n - X| > \varepsilon\} = 0,$
3. **in $L^p$ ($p \geq 1$)** if $E(|X_n - X|^p) \to 0.$

**Proposition 5.2.** Let $p \leq q$. Then convergence in $L^q$ implies convergence in $L^p$.

**Proof.** By Corollary 3.3 we have
\[
E(|X_n - X|^p) \leq E(|X_n - X|^q)^{p/q} \to 0.
\]

**Proposition 5.3.** Convergence in $L^p$ for some $p \geq 1$ implies convergence in probability.

**Proof.** Using Chebyshev’s inequality we obtain
\[
P\{|X_n - X| > \varepsilon\} \leq \frac{1}{\varepsilon^p} E(|X_n - X|^p) \to 0.
\]

**Lemma 5.4.** The sequence $(X_n)_{n \in \mathbb{N}}$ converges almost surely to $X$ if and only if for all $\varepsilon > 0$
\[
\lim_{n \to \infty} P\left\{\sup_{m \geq n} |X_m - X| > \varepsilon\right\} = 0.
\]

**Proof.**
\[
X_n \to X \text{ a.s. } \iff P\{X_n \to X\} = 1
\]
\[
\iff P\left[\forall k \in \mathbb{N} : \exists n \in \mathbb{N} : \forall m \geq n : |X_m - X| \leq \frac{1}{k}\right] = 1
\]
\[
\iff P\left[\bigcap_{k \in \mathbb{N}} \bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} \left\{|X_m - X| \leq \frac{1}{k}\right\}\right] = 1
\]
\[
\iff \forall k \in \mathbb{N} : P\left[\bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} \left\{|X_m - X| \leq \frac{1}{k}\right\}\right] = 1
\]
\[
\iff \forall \varepsilon > 0 : P\left[\bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} \{|X_m - X| \leq \varepsilon\}\right] = 1
\]
\[
\iff \forall \varepsilon > 0 : P\left[\bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} \{|X_m - X| > \varepsilon\}\right] = 0
\]
\[
\iff \forall \varepsilon > 0 : \lim_{n \to \infty} P\left[\bigcup_{m \geq n} \{|X_m - X| > \varepsilon\}\right] = 0
\]
\[
\iff \forall \varepsilon > 0 : \lim_{n \to \infty} P\left[\exists m \geq n : \{|X_m - X| > \varepsilon\}\right] = 0
\]
\[
\iff \forall \varepsilon > 0 : \lim_{n \to \infty} P\left[\sup_{m \geq n} |X_m - X| > \varepsilon\right] = 0. \quad \square
\]
Corollary 5.5. Almost sure convergence implies convergence in probability.

Proof. Use the preceding lemma and note that

$$|X_n - X| \leq \sup_{m \geq n} |X_m - X|.$$  \hfill \Box

Warning. Almost sure convergence does not convergence in $L^p$, nor does convergence in $L^p$ imply almost sure convergence.

Example 5.6. Let $\Omega := [0, 1]$, $\mathcal{A}$ its Borel $\sigma$-field, $P$ = Lebesgue measure and $X_n := n^2 \cdot 1_{(0, 1/n)}$. Then $X_n \to 0$ a.s., but $E(|X_n|^p) = n^{2p-1} \to \infty$, so that $X_n$ does not converge to 0 in $L^p$.

Example 5.7. Convergence in $L^p$ does not imply almost sure convergence.

Definition 5.8. Let $E$ be a topological space equipped with its Borel $\sigma$-field $\mathcal{B}(E)$.

1. A sequence $(\mu_n)_{n \in \mathbb{N}}$ of probability measures on $E$ converges weakly to another probability measure $\mu$ if

$$\int_E f d\mu_n \to \int_E f d\mu$$

for all bounded continuous functions $f : E \to \mathbb{R}$.

2. A sequence $(X_n)_{n \in \mathbb{N}}$ of $E$-valued random variables (not necessarily defined on the same probability space) converges in distribution to another $E$-valued random variable $X$ if the sequence of their distributions converges weakly to the distribution of $X$. By the transformation lemma this holds if and only if

$$E(f(X_n)) \to E(f(X))$$

for all bounded continuous functions $f : E \to \mathbb{R}$.

Proposition 5.9. Let $E = \mathbb{R}^d$. Then convergence in probability implies convergence in distribution.

Proof. We have to show that

$$E(f(X_n)) \to E(f(X))$$

for all $f \in C_b(\mathbb{R}^d)$. Choose $f \in C_b(\mathbb{R}^d)$ and $\varepsilon > 0$. We will show that

$$|E(f(X_n)) - E(f(X))| \leq (4\|f\|_\infty + 1) \varepsilon$$

for all sufficiently large $n \in \mathbb{N}$;

Since $B_R(0) \not\supset \mathbb{R}^d$, we have $P\{X \in B_R(0)\} \to 1$ as $R \to \infty$. Consequently, there exists $R > 0$ such that

$$P(|X| > R) \leq \varepsilon.$$

Since $f$ is uniformly continuous on the compact set $B_{2R}(0)$, there exists $\delta \in (0, R]$ such that

$$|f(x) - f(y)| \leq \varepsilon$$

for all $x, y \in B_{2R}(0)$ such that $|x - y| \leq \delta$. Moreover, the convergence in probability of $X_n$ to $X$ implies the existence of $n_0 \in \mathbb{N}$ such that

$$P(|X_n - X| > \delta) \leq \varepsilon$$

14
for all $n \geq n_0$. It follows that
\[
|E(f(X_n)) - E(f(X))| \\
\leq E(|f(X_n) - f(X)|) \\
= \int_{\Omega} |f(X_n) - f(X)| dP \\
= \int_{\{|X| > R\}} |f(X_n) - f(X)| dP + \int_{\{|X| \leq R, |X_n - X| \leq \delta\}} |f(X_n) - f(X)| dP \\
+ \int_{\{|X| \leq R, |X_n - X| > \delta\}} |f(X_n) - f(X)| dP \\
\leq 2\|f\|_\infty P\{|X| > R\} + \varepsilon + 2\|f\|_\infty P\{|X_n - X| > \delta\} \\
\leq (4\|f\|_\infty + 1)\varepsilon
\]
for all $n \geq n_0$. \qed

6 Laws of large numbers

Definition 6.1 (Law of large numbers). Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of integrable random variables with the same expectation, $E(X_n) = \mu$ for all $n \in \mathbb{N}$. We say that the sequence satisfies the \{strong, weak, $L^p$ ($p \geq 1$)\} law of large numbers if $\frac{1}{n} \sum_{i=1}^{n} X_i$ converges to $\mu$ almost surely, in probability, in $L^p$.

Remark 6.2. The results of the preceding section imply that both the strong and the $L^p$ law of large numbers imply the weak law.

Example 6.3. Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of independent and identically distributed square-integrable random variables. Then it is very easy to show that it satisfies the $L^2$ and consequently also the weak law of large numbers (exercise!). To prove that it satisfies the strong law is much harder (see e.g. Theorem 12.1 in Bauer’s book). We will later see a proof based on martingale theory.

7 Conditional expectation

Let us come back to the fair game example of the beginning of this course. There we gave the preliminary definition that a sequence $(X_n)_{n \in \mathbb{N}}$ of random variables is a fair game if for all $n \in \mathbb{N}$ and all $a_1, \ldots, a_n \in \mathbb{R}$ such that $P[X_1 = a_1, \ldots, X_n = a_n] > 0$ we have
\[
E[X_{n+1} | X_1 = a_1, \ldots, X_n = a_n] = a_n.
\]
Here the elementary conditional expectation of a random variable $X$ given an event $B$ of positive probability is defined as
\[
E(X|B) := \frac{E(X \cdot 1_B)}{P(B)}.
\]
Clearly this definition only makes sense if indeed $P(B) > 0$. Consequently, the above fair game definition works well if the variables $(X_n)_{n \in \mathbb{N}}$ are all discrete, but not in general because it can happen that $P[X_1 = a_1, \ldots, X_n = a_n] = 0$ for all $a_1, \ldots, a_n \in \mathbb{R}$. We therefore need a more general concept of conditional expectation.
Let \( X \) be a quasi-integrable random variable on a probability space \((Ω, A, P)\), and let \((B_i)_{i ∈ I}\) be a countable collection of pairwise disjoint events with \(P(B_i) > 0\) for all \(i ∈ I\) and such that \(∪_{i ∈ I} B_i = Ω\). Then the conditional expectation of \(X\) given \(B_i\) is

\[
E(X|B_i) = \frac{1}{P(B_i)} E(X \cdot 1_{B_i}).
\]

Let now

\[
X_0 := \sum_{i ∈ I} E(X|B_i)1_{B_i}.
\]

Then \(X_0\) is a random variable on \((Ω, A, P)\) satisfying

\[
X_0(ω) = E(X|B_i) \quad \text{if} \quad ω ∈ B_i.
\]

In particular, \(X_0\) is constant on each \(B_i\) and therefore measurable with respect to \(C := σ((B_i)_{i ∈ I}) = \{∪_{i ∈ J} B_i | J ⊂ I\}\). Moreover, for each \(i ∈ I\),

\[
\int_{B_i} X_0 dP = E(X|B_i)P(B_i) = E(X \cdot 1_{B_i}) = \int_{B_i} X dP,
\]

and consequently

\[
\int_C X_0 dP = \int_C X dP
\]

for all \(C ∈ C\).

**Definition 7.1.** Let \( X \) be a quasi-integrable random variable on a probability space \((Ω, A, P)\), and let \(C\) be a sub-\(σ\)-field of \(A\). A quasi-integrable random variable \(X_0\) on \((Ω, A, P)\) is called a *conditional expectation* of \(X\) given \(C\) if

1. \(X_0\) is measurable with respect to \(C\), and
2. \(\int_C X_0 dP = \int_C X dP\) for all \(C ∈ C\).

To prove existence and uniqueness of the conditional expectation we need the Radon-Nikodym theorem:\(^{13}\)

**Theorem 7.2** (Radon-Nikodym). Let \((Ω, A)\) be a measurable space, \(μ\) a \(σ\)-finite measure on it and \(ν\) a signed measure which is absolutely continuous with respect to \(μ\), i.e. \(ν(A) = 0\) whenever \(μ(A) = 0\). Then \(ν\) has a density with respect to \(μ\), i.e. there exists a quasi-integrable function \(f : Ω → \mathbb{R}\) such that

\[
ν(A) = \int_A f dμ
\]

for all \(A ∈ A\). Moreover, \(f\) is unique \(μ\)-almost everywhere.

**Theorem 7.3.** A conditional expectation always exists. Moreover, it is almost surely unique (i.e. if \(X_0\) and \(X'_0\) are two conditional expectations, then \(X_0 = X'_0\) almost surely).

Proof. Let $Q$ be the signed measure with density $X$ with respect to $P$, i.e.

$$Q(A) := \int_A X \, dP.$$ 

Then a $\mathcal{C}$-measurable quasi-integrable random variable $X_0$ is a conditional expectation of $X$ given $\mathcal{C}$ if and only if

$$\int_{\mathcal{C}} X_0 \, dP = Q(\mathcal{C})$$

for all $C \in \mathcal{C}$, i.e. if and only if $X_0$ is a density of $Q|_{\mathcal{C}}$ with respect to $P|_{\mathcal{C}}$. Since $Q|_{\mathcal{C}}$ is absolutely continuous with respect to $P|_{\mathcal{C}}$ (because $Q$ is absolutely continuous with respect to $P$), we can apply the Radon-Nikodym theorem which tells us that such a density exists and is almost surely unique. \hfill \Box

Notation. The conditional expectation of $X$ with respect to $\mathcal{C}$ is denoted by $E^\mathcal{C}(X)$ or $E(X|\mathcal{C})$.

Warning. Contrary to the usual expectation and the elementary conditional expectation, the conditional expectation of a random variable with respect to a sub-$\sigma$-field is NOT a real number, but again a random variable. Moreover, it is only almost surely uniquely determined.

Proposition 7.4. The conditional expectation has the following properties:

1. If $X$ is integrable, then so is $E(X|\mathcal{C})$ (in particular it is almost surely finite).
2. If $\mathcal{C} = \{\emptyset, \Omega\}$, then $E(X|\mathcal{C}) = E(X)$.
3. If $\mathcal{C} = A$, then $E(X|\mathcal{C}) = X$.
4. Let $(B_i)_{i \in I}$ be a countable collection of pairwise disjoint events with $P(B_i) > 0$ for all $i \in I$ and such that $\bigcup_{i \in I} B_i = \Omega$, and let $\mathcal{C} := \sigma((B_i)_{i \in I}) = \{\bigcup_{i \in J} B_i | J \subset I\}$. Then

$$E(X|\mathcal{C}) = \sum_{i \in I} E(X|B_i)1_{B_i}.$$ 

5. $E(E(X|\mathcal{C})) = E(X)$ (often useful!)
6. If $X$ is $\mathcal{C}$-measurable, then $E(X|\mathcal{C}) = X$.
7. If $X$ and $Y$ are integrable and $\alpha, \beta \in \mathbb{R}$, then $E(\alpha X + \beta Y|\mathcal{C}) = \alpha E(X|\mathcal{C}) + \beta E(Y|\mathcal{C})$.
8. If $X \leq Y$, then $E(X|\mathcal{C}) \leq E(Y|\mathcal{C})$.
9. $|E(X|\mathcal{C})| \leq E(|X| |\mathcal{C})$.
10. Whenever $0 \leq X_n \nearrow X$, one has $E(X_n|\mathcal{C}) \nearrow E(X|\mathcal{C})$ (conditional Beppo Levi).
11. If $X_n \to X$ a.e. and there exists an integrable random variable $Y$ with $|X_n| \leq Y$ for all $n \in \mathbb{N}$, then $E(X_n|\mathcal{C}) \to E(X|\mathcal{C})$ (conditional Lebesgue).
12. If $X$ is integrable and $q : \mathbb{R} \to \mathbb{R}$ is convex, then $q(X)$ is quasi-integrable (as we already know), and

$$E(q(X)|\mathcal{C}) \geq q(E(X|\mathcal{C}))$$
(conditional Jensen). Since both sides of this inequality are quasi-integrable, it follows that

\[ E(q(E(X|C))) \leq E(q(X)). \]

In particular, for \( p \geq 1 \),

\[ E(|E(X|C)|^p) \leq E(|X|^p). \]

In particular, if \( X \) belongs to \( L^p \), then so does \( E(X|C) \).

13. Let \( C_1 \subseteq C_2 \subseteq A \). Then

\[ E(E(X|C_2)|C_1) = E(X|C_1). \]

**Proposition 7.5.** Let \( X \) and \( Y \) be both nonegative or such that \( X \) and \( XY \) are all integrable. If \( X \) is \( \mathcal{C} \)-measurable, then

\[ E(XY|\mathcal{C}) =XE(Y|\mathcal{C}). \]

“We measurable factors can be taken out of the conditional expectation.”

**Proof.** Let \( Y_0 := E(Y|\mathcal{C}) \). By the definition of conditional expectation we have to show that

\[ \int_C XYdP = \int_C XY_0dP \]

for all \( C \in \mathcal{C} \). We proceed in several steps:

1. \( X \) is the indicator function of some \( \mathcal{C} \)-measurable set \( \tilde{C} \), \( X = 1_{\tilde{C}} \). Then

\[ \int_C 1_{\tilde{C}} YdP = \int_{C\cap \tilde{C}} YdP = \int_{C\cap \tilde{C}} Y_0dP = \int_C 1_{\tilde{C}} Y_0dP. \]

2. \( X \) is elementary: This follows from 1. by linearity.

3. \( X \) and \( Y \) are both nonnegative: Let \( (X_n)_{n \in \mathbb{N}} \) be a sequence of nonnegative elementary \( \mathcal{C} \)-measurable random variables such that \( X_n \nearrow X \). Then \( X_nY \nearrow XY \) and \( X_nY_0 \nearrow XY_0 \) so that by 2.

\[
\int_C XYdP = \int_C \lim_{n \to \infty} X_nYdP \\
= \lim_{n \to \infty} \int_C X_nYdP \\
= \lim_{n \to \infty} \int_C X_nY_0dP \\
= \int_C XY_0dP.
\]

4. General case: Write \( XY = X_+Y_+ + X_-Y_- - X_+Y_- - X_-Y_+ \). Then

\[
\int_C XYdP = \int_C X_+Y_+dP + \int_C X_-Y_-dP - \int_C X_+Y_-dP - \int_C X_-Y_+dP \\
= \int_C X_+Y_0dP + \int_C X_-Y_0dP - \int_C X_+Y_0dP - \int_C X_-Y_0dP \\
= \int_C XY_0dP. \quad \square
\]
Corollary 7.6. Let \( X \) and \( Y \) be both nonegative or such that \( X, Y \) and \( XY \) are all integrable. If \( X \) is \( \mathcal{C} \)-measurable, then
\[
E(XY) = E(XE(Y|\mathcal{C})).
\]

Proof. Take expectation on both sides of the previous proposition. \( \Box \)

Proposition 7.7. If \( X \) is independent of \( \mathcal{C} \), then
\[
E(X|\mathcal{C}) = E(X).
\]

Proof. Let \( C \in \mathcal{C} \). Then
\[
\int_C XdP = E(X \cdot 1_C) = E(X)P(C) = \int_C E(X)dP.
\]

Lemma 7.8. Let \( X, Y \in L^2 \), and let \( Y \) be \( \mathcal{C} \)-measurable. Write \( X_0 := E(X|\mathcal{C}) \). Then
\[
E((X - Y)^2) = E((X - X_0)^2) + E((X_0 - Y)^2).
\]

Remark 7.9. In the case \( \mathcal{C} = \{\emptyset, \Omega\} \), this simplifies to the well-known formula
\[
E((X - a)^2) = \text{Var}(X) + (E(X) - a)^2
\]
for all \( a \in \mathbb{R} \). “Mean squared error = variance + squared bias.”

Proof of Lemma 7.8. By Corollary 7.6 we have
\[
E(XY) = E(X_0Y) \quad \text{and} \quad E(XX_0) = E(X_0^2).
\]
Consequently
\[
E((X - X_0)^2) + E((X_0 - Y)^2) = E(X^2) - 2E(XX_0) + 2E(X_0^2) - 2E(X_0Y) + E(Y^2)
= E(X^2) - 2E(XY) + E(Y^2)
= E((X - Y)^2). \quad \Box
\]

Corollary 7.10. If \( X \in L^2(\Omega, \mathcal{A}, P) \), \( E(X|\mathcal{C}) \) is the unique (up to almost sure equality) best approximation of \( X \) in the subspace \( L^2(\Omega, \mathcal{C}, P) \). In this sense \( E(X|\mathcal{C}) \) is the best possible prediction of \( X \) given the information contained in \( \mathcal{C} \).

Remark 7.11. This observation leads to an alternative proof of existence for the conditional expectation using the projection theorem of Hilbert space\(^{14}\) theory.

Theorem 7.12 (Projection theorem). Let \( H \) be a Hilbert space (in our case \( H = L^2(\Omega, \mathcal{A}, P) \)), \( x \in H \) (in our case \( x = X \)) and \( S \) a non-empty convex closed subset of \( H \) (in our case \( S = L^2(\Omega, \mathcal{C}, P) \); being itself a Hilbert space it is complete and therefore closed).

1. There is a unique best approximation \( x_0 \) to \( x \) in \( S \), i.e. an element \( x_0 \) of \( S \) satisfying
\[
\|x - x_0\| \leq \|x - y\|
\]
for all \( y \in S \).

2. If $S$ is a linear subspace of $H$, $x_0$ is characterized by the property that $x - x_0$ is orthogonal to $S$, i.e. $(x - x_0, y) = 0$ for all $y \in S$. Picture!!!

Alternative proof of Theorem 7.3.

1. Assume first that $X \in L^2(\Omega, A, P)$, and let $X_0$ be the best approximation to $X$ in $L^2(\Omega, C, P)$. It satisfies $E(XY) = E(X_0Y)$ for all $Y \in L^2(\Omega, C, P)$. By choosing $Y = 1_C$ we obtain

$$\int_C X dP = \int_C X_0 dP$$

for all $C \in \mathcal{C}$ so that $X_0$ is a conditional expectation of $X$ given $\mathcal{C}$.

2. If $X$ is nonnegative, choose a sequence $(X_n)_{n \in \mathbb{N}}$ in $L^2(\Omega, A, P)$ such that $0 \leq X_n \nearrow X$ (use the fact that all elementary random variables belong to $L^2$). Then the sequence $(X_{n0})_{n \in \mathbb{N}}$ is non-decreasing, so that $X_0 := \lim_{n \to \infty} X_{n0}$ exists, and

$$\int_C X dP = \lim_{n \to \infty} \int_C X_n dP = \lim_{n \to \infty} \int_C X_{n0} dP = \int_C X_0 dP$$

for all $C \in \mathcal{C}$.

3. If $X$ is quasi-integrable, choose $X_0 := X^{+} - X^{-}$.

8 Stochastic processes and martingales

**Definition 8.1.** Let $(\Omega, A, P)$ be a probability space, $(E, B)$ a measurable space and $I$ a set. A **stochastic process** on $(\Omega, A, P)$ with state space $(E, B)$ and index set $I$ is a family $(X_t)_{t \in I}$ of $(E, B)$-valued random variables on $(\Omega, A, P)$.

**Remark 8.2.** Usually $I \subseteq \mathbb{R}$, and then $t \in I$ is interpreted as “time”. The two most important cases are

1. $I = \mathbb{N}_0$, “discrete time”, and
2. $I = \mathbb{R}_+$, “continuous time”.

**Definition 8.3.** Let $I$ be an **ordered set**, i.e. a set equipped with a binary relation $\leq$ which is

1. reflexive, i.e. $t \leq t$ for all $t \in I$,  
2. transitive, i.e. $(s \leq t$ and $t \leq u) \Rightarrow s \leq u$, and
3. antisymmetric, i.e. $(s \leq t$ and $t \leq s) \Rightarrow s = t$.

(As above, usually $I \subseteq \mathbb{R}$, in particular $I = \mathbb{N}_0$ or $I = \mathbb{R}_+$.) In this context we define:

1. A **filtration** on $(\Omega, A, P)$ indexed by $I$ is a family $(\mathcal{F}_t)_{t \in I}$ of sub-$\sigma$-fields of $A$ satisfying $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s \leq t$.

2. A **filtered probability space** is a quadruple $(\Omega, A, P, (\mathcal{F}_t)_{t \in I})$ where $(\Omega, A, P)$ is a probability space and $(\mathcal{F}_t)_{t \in I}$ a filtration on it.

3. A stochastic process $(X_t)_{t \in I}$ on $(\Omega, A, P)$ is **adapted** to the filtration $(\mathcal{F}_t)_{t \in I}$ if $X_t$ is $\mathcal{F}_t$-measurable for each $t \in I$. 

20
4. The canonical filtration of a stochastic process $(X_t)_{t \in I}$ is given by

$$F^X_t := \sigma(X_s, s \leq t).$$

It is the smallest filtration with respect to which the process $(X_t)_{t \in I}$ is adapted.

**Remark 8.4.** $F_t$ is often interpreted as “the information available at time $t$”. Adaptedness of a process $(X_t)_{t \in I}$ means that the information available at time $t$ contains at least the information on the values of $X_s$ for $s \leq t$. The canonical filtration of a process $(X_t)_{t \in I}$ contains exactly the information on the values of $X_s$ for $s \leq t$. If a stochastic process, but no filtration is specified, we work with the canonical filtration.

**Definition 8.5.** Let $(\Omega, \mathcal{A}, P, (F_t)_{t \in I})$ be a filtered probability space. A real-valued stochastic process $(M_t)_{t \in I}$ on $(\Omega, \mathcal{A}, P)$ is called a {martingale, submartingale, supermartingale} with respect to the filtration $(F_t)_{t \in I}$ if

1. it is adapted to $(F_t)_{t \in I}$,
2. $E(|M_t|) < \infty$ for all $t \in I$, and
3. $E(M_t|F_s) \{=, \geq, \leq\} M_s$ for all $s \leq t$.

**Remark 8.6.**

1. Martingales describe fair games, supermartingales describe “real” games (i.e. games you can play in a casino). Submartingales describe games from the point of view of the casino.
2. The third condition in the previous definition can also be formulated as

$$\int_A M_t dP \{=, \geq, \leq\} \int_A M_s dP$$

for all $s \leq t$ and all $A \in F_s$ (hence without explicitly using the notion of conditional expectation).
3. Let $(\tilde{F}_t)_{t \in I}$ be a smaller filtration (i.e. $\tilde{F}_t \subseteq F_t$ for all $t \in I$). If $(M_t)_{t \in I}$ is a martingale (submartingale, supermartingale) with respect to $(F_t)_{t \in I}$ and adapted to $(\tilde{F}_t)_{t \in I}$, then it is also a martingale (submartingale, supermartingale) with respect to $(\tilde{F}_t)_{t \in I}$. (Use (6).)

In the sequel we will be mainly interested in the case of discrete time, i.e. $I = \mathbb{N}_0$. (The study of martingales in the case of continuous time, i.e. $I = \mathbb{R}_+$, is the topic of stochastic analysis, see next semester.)

**Proposition 8.7.** Let $(M_n)_{n \in \mathbb{N}_0}$ be a martingale (submartingale, supermartingale) with respect to a filtration $(F_n)_{n \in \mathbb{N}_0}$, and let $(X_n)_{n \in \mathbb{N}_0}$ be a bounded adapted process which we assume to be non-negative if $(M_n)_{n \in \mathbb{N}_0}$ is only a sub- or supermartingale. Then the (sub-/super-)martingale transform of $M$ by $X$ defined by

$$(X \cdot M)_n := \sum_{k=1}^n X_{k-1}(M_k - M_{k-1})$$

is also a martingale (submartingale, supermartingale).
Proof.

\[
E ((X \cdot M)_{n+1} - (X \cdot M)_n \mid \mathcal{F}_n) = E (X_n(M_{n+1} - M_n) \mid \mathcal{F}_n) = X_n E ((M_{n+1} - M_n) \mid \mathcal{F}_n)
\]

\{=, \geq, \leq\} 0. \quad \square

Example 8.8. Let \( M_n \) be the price of a financial asset at time \( n \), and let \( X_n \) be the number of shares you hold from time \( n \) to time \( n + 1 \) (since \( X_n \) has to be known at time \( n \), the process \((X_n)_{n \in \mathbb{N}_0}\) should be adapted). Then \((X \cdot M)_n\) is the amount of money you have won (or lost) between time 0 and time \( n \), and the proposition says that if the original asset is a martingale (submartingale, supermartingale), then so is your own value process. (If the original asset is only a sub- or supermartingale, one has to assume that \( X_n \geq 0 \) for all \( n \in \mathbb{N}_0 \), i.e. that no short-selling (Leerverkauf, vente à découvert) occurs.

Definition 8.9. Let \((\Omega, \mathcal{A}, P, (\mathcal{F}_n)_{n \in \mathbb{N}_0})\) be a filtered probability space. A stochastic process \((X_n)_{n \in \mathbb{N}}\) is predictable if

1. \( X_0 \) is constant, and
2. \( X_n \) is \( \mathcal{F}_{n-1} \)-measurable for all \( n \geq 1 \).

“The value of \( X_n \) can be predicted given the information available at time \( n - 1 \).”

Lemma 8.10. Let \((M_n)_{n \in \mathbb{N}_0}\) be a predictable martingale. Then it is constant, i.e. there exists \( c \in \mathbb{R} \) such that \( M_n = c \) for all \( n \in \mathbb{N}_0 \).

Proof. By induction: Since \((M_n)_{n \in \mathbb{N}_0}\) is predictable, there exists a constant \( c \in \mathbb{R} \) such that \( M_0 = c \). Suppose we have proved that \( M_n = c \). Then predictability and the martingale property imply that

\[
M_{n+1} = E(M_{n+1} \mid \mathcal{F}_n) = M_n = c.
\]

\( \square \)

Proposition 8.11 (Doob-Meyer decomposition). Let \((\Omega, \mathcal{A}, P, (\mathcal{F}_n)_{n \in \mathbb{N}_0})\) be a filtered probability space and \((X_n)_{n \in \mathbb{N}_0}\) an adapted integrable process. Then there exists a martingale \((M_n)_{n \in \mathbb{N}_0}\) and a predictable process \((A_n)_{n \in \mathbb{N}_0}\) such that

\[
X_n = M_n + A_n
\]

for all \( n \in \mathbb{N}_0 \). Moreover,

1. this decomposition is unique up to an additive constant (if \( X_n = \tilde{M}_n + \tilde{A}_n \) is another decomposition of this type, there exists a constant \( c \in \mathbb{R} \) such that for all \( n \in \mathbb{N}_0 \) we have \( \tilde{M}_n = M_n + c \) and \( \tilde{A}_n = A_n - c \)).

2. If the process \((X_n)_{n \in \mathbb{N}_0}\) is a submartingale (supermartingale), the process \((A_n)_{n \in \mathbb{N}_0}\) is increasing (decreasing), i.e. \( A_{n+1} \geq A_n \) \( (A_{n+1} \leq A_n) \) for all \( n \in \mathbb{N}_0 \).

Proof. Uniqueness up to an additive constant is clear from the preceding lemma (the process \( M_n - M_n = \tilde{A}_n - A_n \) is a predictable martingale and hence constant). For existence let

\[
A_n := \sum_{k=1}^{n} E(X_k - X_{k-1} \mid \mathcal{F}_{k-1})
\]

and

\[
M_n := X_n - A_n.
\]
Clearly, $A_0 = 0$ (in particular it is constant), and $A_n$ is $\mathcal{F}_{n-1}$-measurable for all $n \in \mathbb{N}$, so that the process $(A_n)_{n \in \mathbb{N}_0}$ is predictable. Moreover, the process $(M_n)_{n \in \mathbb{N}_0}$ is clearly adapted and integrable, and for $n \geq 1$

\[ E(M_n - M_{n-1}|\mathcal{F}_{n-1}) = E(X_n - X_{n-1} - (A_n - A_{n-1})|\mathcal{F}_{n-1}) = E(X_n - X_{n-1} - E(X_n - X_{n-1}|\mathcal{F}_{n-1})|\mathcal{F}_{n-1}) = 0, \]

so that the process $(M_n)_{n \in \mathbb{N}_0}$ is indeed a martingale. Finally, if the process $(X_n)_{n \in \mathbb{N}_0}$ is a submartingale (supermartingale), we have

\[ A_n - A_{n-1} = E(X_n - X_{n-1}|\mathcal{F}_{n-1}) \geq 0 \ (\leq 0) \]

for all $n \geq 1$.

### 9 Stopping times

**Definition 9.1.** Let $(\Omega, \mathcal{A}, P, (\mathcal{F}_n)_{n \in \mathbb{N}_0})$ be a filtered probability space. A random variable $T$ on $(\Omega, \mathcal{A}, P, (\mathcal{F}_n)_{n \in \mathbb{N}_0})$ with values in $\mathbb{N}_0 \cup \{\infty\}$ is called a **stopping time** if for all $n \in \mathbb{N}_0$ we have

\[ \{T \leq n\} \in \mathcal{F}_n. \]

**Interpretation.** At any time $n$ we know whether $T$ has occurred.

**Remark 9.2.** An equivalent condition is that

\[ \{T = n\} \in \mathcal{F}_n \]

for all $n \in \mathbb{N}_0$.

**Proof.** Use $\{T = n\} = \{T \leq n\} \setminus \{T \leq n - 1\}$ and $\{T \leq n\} = \bigcup_{k=1}^{n} \{T = k\}$. $\square$

**Example 9.3.** Let $(E, B)$ be a measurable space, $(X_n)_{n \in \mathbb{N}_0}$ an adapted $(E, B)$-valued stochastic process, $B \in B$, and

\[ T_B := \begin{cases} \inf \{n \in \mathbb{N}_0 | X_n \in B\} & \text{if such an } n \text{ exists} \\ \infty & \text{otherwise} \end{cases} \]

the **first hitting time** of $B$. $T_B$ is a stopping time because

\[ \{T \leq n\} = \{\exists k \leq n : X_k \in B\} = \bigcup_{k=0}^{n} \{X_k \in B\} \in \mathcal{F}_n \]

On the other hand, the last hitting time of $B$ is not in general a stopping time.

**Notation.** Let $T$ be a stopping time on a filtered probability space $(\Omega, \mathcal{A}, P, (\mathcal{F}_n)_{n \in \mathbb{N}_0})$ and $(X_n)_{n \in \mathbb{N}_0}$ an adapted $(E, B)$-valued stochastic process. We write

\[ \mathcal{F}_T := \{A \in \mathcal{A} | A \cap \{T \leq n\} \in \mathcal{F}_n \text{ for all } n \in \mathbb{N}_0\} \]

and

\[ X_T(\omega) := X_{T(\omega)}(\omega) \quad \text{if } T(\omega) < \infty \]

**Exercise.** Show that $\mathcal{F}_T$ is a $\sigma$-field and that $T$ and $X_T : \{T < \infty\} \to E$ are $\mathcal{F}_T$-measurable.
Proposition 9.4 (Optional stopping). Let \((M_n)_{n \in \mathbb{N}_0}\) be a martingale (submartingale, supermartingale), and let \(T\) be a stopping time. Then the stopped process \((M_{T \wedge n})_{n \in \mathbb{N}}\) is a martingale (submartingale, supermartingale) as well.

Proof. Let \(X_n := 1_{\{T \geq n+1\}}\). The process \((X_n)_{n \in \mathbb{N}_0}\) is adapted because \(\{T \geq n+1\} = \{T \leq n\}^c\). Moreover, the martingale transform of \(M\) by \(X\) equals \((X \cdot M)_n = \sum_{k=1}^{n} X_{k-1}(M_k - M_{k-1}) = \sum_{k=1}^{n} 1_{\{T \geq k\}}(M_k - M_{k-1}) = \sum_{k=1}^{T \wedge n} (M_k - M_{k-1}) = M_{T \wedge n} - M_0\), so that \(M_{T \wedge n} - M_0\) equals the martingale transform of \(M\) by \(X\) and is therefore a martingale (submartingale, supermartingale) as well (note that the process \((X_n)_{n \in \mathbb{N}_0}\) is non-negative).

Theorem 9.5 (Optional sampling theorem). Let \((X_n)_{n \in \mathbb{N}_0}\) be a martingale (submartingale, supermartingale), and let \(S \leq T\) be two bounded stopping times. Then

\[\mathbb{E}(X_T | \mathcal{F}_S) \{=, \geq, \leq\} X_S.\]

Proof. (for supermartingales only) Let \(k\) be an upper bound for \(T\). We have to show that

\[\int_A X_T dP \leq \int_A X_S dP\]

for all \(A \in \mathcal{F}_S\).

Let us first assume that \(T \leq S + 1\). Then

\[
\begin{align*}
\int_A X_T dP &= \int_{A \cap \{S = T\}} X_T dP + \int_{A \cap \{S < T\}} X_T dP \\
&= \int_{A \cap \{S = T\}} X_S dP + \sum_{i=0}^{k} \int_{A \cap \{S = i\} \cap \{T > i\}} X_{i+1} dP \\
&\leq \int_{A \cap \{S = T\}} X_S dP + \sum_{i=0}^{k} \int_{A \cap \{S = i\} \cap \{T > i\}} X_i dP \\
&= \int_A X_S dP
\end{align*}
\]

For the general case, let \(S_i := (S+i) \wedge T\). Then \(S = S_0 \leq \ldots \leq S_k = T\) and \(S_{i+1} \leq S_i + 1\) for all \(i \in \{0, \ldots, k-1\}\). Consequently,

\[\int_A X_T dP = \int_A X_{S_k} dP \leq \ldots \leq \int_A X_{S_0} dP = \int_A X_S dP.\]

Corollary 9.6. Let \((X_n)_{n \in \mathbb{N}_0}\) be a martingale (submartingale, supermartingale), and let \(T_1 \leq \ldots \leq T_k\) be \(k\) bounded stopping times. Then the sequence \((X_{T_i})_{i=1}^k\) is a martingale (submartingale, supermartingale) with respect to the filtration \((\mathcal{F}_{T_i})_{i=1}^k\).
Remark 9.7. If $T$ is almost surely finite, but not bounded, the result is wrong.

Example 9.8. Let $(\xi_i)_{i \in \mathbb{N}}$ be a sequence of independent random variables with $P(\xi_i = 1) = P(\xi_i = -1) = 1/2$, and let $X_n := \sum_{i=1}^n \xi_i$. Then the process $(X_n)_{n \in \mathbb{N}}$ is a martingale (exercise!). Moreover, let $S := 0$ and $T := \inf\{n \in \mathbb{N}_0 : X_n = 1\}$. $T$ is almost surely finite because the process $(X_n)_{n \in \mathbb{N}}$ is a one-dimensional simple random walk which is known to be recurrent. We have $X_T = 1$ and consequently $E(X_T | \mathcal{F}_0) = 1 \neq X_0 = 0$.

10 Maximal inequalities

Proposition 10.1. Let $(X_n)_{n \in \mathbb{N}}$ be a submartingale on a filtered probability space $(\Omega, \mathcal{A}, P, (\mathcal{F}_n)_{n \in \mathbb{N}_0})$. Then for all $n \in \mathbb{N}_0$ and all $c > 0$

$$P\left( \max_{0 \leq k \leq n} X_k \geq c \right) \leq \frac{1}{c} E\left( X_n \cdot 1_{\{\max_{0 \leq k \leq n} X_k \geq c\}} \right) \leq \frac{1}{c} E(X_n^+) .$$

Proof. Fix $n \in \mathbb{N}_0$, and let $X_n^* := \max\{X_k : k \leq n\}$ and

$$T := n \wedge \inf\{k \in \mathbb{N} : X_k \geq c\} .$$

$T$ is a stopping time and $T \leq n$. Therefore the optional sampling theorem implies that

$$E(X_n | \mathcal{F}_T) \geq X_T$$

so that

$$\int_A X_n dP \geq \int_A X_T dP$$

for all $A \in \mathcal{F}_T$. We want to apply this to the event $A := \{X_n^* \geq c\}$, so we have to check that $A \in \mathcal{F}_T$, i.e. $A \cap \{T \leq k\} \in \mathcal{F}_k$ for all $k \in \mathbb{N}_0$. For $k \leq n$ we have

$$A \cap \{T \leq k\} = \{X_n^* \geq c\} \cap \{X_k^* \geq c\} = \{X_k^* \geq c\} \in \mathcal{F}_k \subseteq \mathcal{F}_n ,$$

and for $k > n$ we have

$$A \cap \{T \leq k\} = \{X_n^* \geq c\} \in \mathcal{F}_n .$$

Consequently $A \in \mathcal{F}_T$, so that

$$cP\{X_n^* \geq c\} \leq \int_{\{X_n^* \geq c\}} X_T dP \leq \int_{\{X_n^* \geq c\}} X_n dP \leq E(X_n^+) . \quad \square$$

Proposition 10.2. Let $(X_n)_{n \in \mathbb{N}}$ be a martingale on a filtered probability space $(\Omega, \mathcal{A}, P, (\mathcal{F}_n)_{n \in \mathbb{N}_0})$. Then for all $p > 1$, all $n \in \mathbb{N}_0$ and all $c > 0$

$$E\left( \max_{0 \leq k \leq n} |X_k|^p \right) \leq \left( \frac{p}{p - 1} \right)^p E(|X_n|^p) .$$

Proof. Let $Y := \max_{0 \leq k \leq n} |X_k|$ and $q := p/(p - 1)$ (such that $1/p + 1/q = 1$). Since for $n \geq m$ we have

$$E(|X_n| | \mathcal{F}_m) \geq |E(X_n | \mathcal{F}_m)| = |X_m| ,$$

the process $(|X_n| | \mathcal{F}_m)_{n \in \mathbb{N}_0}$ is a submartingale. Consequently the previous proposition implies that for all $c > 0$

$$P(|Y| \geq c) \leq \frac{1}{c} E(|X_n| \cdot 1_{\{|Y| \geq c\}}) .$$

25
Consequently

\[ E(Y^p) = E \left( \int_{0}^{Y} pt^{p-1} dt \right) \]
\[ = E \left( \int_{0}^{\infty} pt^{p-1} 1_{\{Y \geq t\}} dt \right) \]
\[ = \int_{0}^{\infty} E \left( pt^{p-1} 1_{\{Y \geq t\}} \right) dt \]
\[ = \int_{0}^{\infty} pt^{p-1} P\{Y \geq t\} dt \]
\[ \leq \int_{0}^{\infty} pt^{p-2} E(|X_n| 1_{\{|Y| \geq t\}}) dt \]
\[ = E \left( \int_{0}^{\infty} pt^{p-2} |X_n| 1_{\{|Y| \geq t\}} dt \right) \]
\[ = E \left( |X_n| \int_{0}^{Y} pt^{p-2} dt \right) \]
\[ = \frac{p}{p-1} E( |X_n| Y^{p-1}) \]
\[ \leq \frac{p}{p-1} E( |X_n|^p)^{1/p} E(Y^{(p-1)q})^{1/q} \]
\[ = \frac{p}{p-1} E( |X_n|^p)^{1/p} E(Y^p)^{(p-1)/p}, \]

so that

\[ E(Y^p)^p \leq \left( \frac{p}{p-1} \right)^p E(|X_n|^p) E(Y^p)^{p-1} \]

and therefore

\[ E(Y^p) \leq \left( \frac{p}{p-1} \right)^p E(|X_n|^p). \]

11 Martingale convergence theorems

Key observation: “A sequence of real numbers converges in \( \bar{\mathbb{R}} \) if and only if it does not cross any nontrivial interval infinitely often.”

11.1 Doob’s downcrossing inequality

**Definition 11.1.** Let \((x_n)_{n \in \mathbb{N}_0}\) be a sequence of real numbers, and let \(a < b\). The number of downcrossings of the interval \([a, b]\) by the sequence \((x_n)_{n \in \mathbb{N}}\) is

\[ \gamma_{a,b}((x_n)_{n \in \mathbb{N}_0}) := \sup\{ n \in \mathbb{N}_0 : \exists k_1 < \ldots < k_{2n} \in \mathbb{N}_0 : x_{k_1} > b, x_{k_2} < a, \ldots, x_{k_{2n-1}} > b, x_{k_{2n}} < a \}. \]

**Lemma 11.2.** Let \((x_n)_{n \in \mathbb{N}_0}\) be a sequence of real numbers. Then the following are equivalent:

1. The sequence converges in \( \bar{\mathbb{R}} \).
2. \( \gamma_{a,b}((x_n)_{n \in \mathbb{N}_0}) \) is finite for all \( a < b \in \mathbb{R} \).
3. \( \gamma_{a,b}((x_n)_{n \in \mathbb{N}_0}) \) is finite for all \( a < b \in \mathbb{Q} \).
Proof. 1. ⇒ 2. Suppose that $\gamma_{a,b}((x_n)_{n\in\mathbb{N}_0}) = \infty$ for some $a < b$. Then $\liminf_{n \to \infty} x_n \leq a$ and $\limsup_{n \to \infty} x_n \geq b$, so that the sequence $(x_n)_{n\in\mathbb{N}_0}$ does not converge in $\bar{\mathbb{R}}$.

2. ⇒ 3. This is trivial.

3. ⇒ 1. Suppose that the sequence $(x_n)_{n\in\mathbb{N}_0}$ does not converge in $\bar{\mathbb{R}}$. Then $\liminf_{n \to \infty} x_n < \limsup_{n \to \infty} x_n$, so that there exist $a, b \in \mathbb{Q}$ such that $\liminf_{n \to \infty} x_n < a < b < \limsup_{n \to \infty} x_n$. Consequently $\gamma_{a,b}((x_n)_{n\in\mathbb{N}_0}) = \infty$. □

Lemma 11.3. Let $X$ be a random variable with values in $\mathbb{N}_0$. Then

$$E(X) = \sum_{k=1}^{\infty} P(X \geq k).$$

Proof.

$$\sum_{k=1}^{\infty} P(X \geq k) = \sum_{k=1}^{\infty} \sum_{l=k}^{\infty} P(X = l) = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} 1\{l \geq k\} P(X = l) = \sum_{l=1}^{\infty} \sum_{k=1}^{l} P(X = l) = \sum_{l=1}^{\infty} l P(X = l) = E(X).$$

Theorem 11.4 (Doob’s downcrossing inequality). Let $(X_n)_{n\in\mathbb{N}_0}$ be a submartingale on a filtered probability space $(\Omega, \mathcal{A}, P, (\mathcal{F}_n)_{n\in\mathbb{N}_0})$, and let $a < b$. Then

$$E(\gamma_{a,b}((X_n)_{n\in\mathbb{N}_0})) \leq \frac{1}{b-a} \sup_{n \in \mathbb{N}} E((X_n - b)_+).$$

Proof. For $k \in \mathbb{N}$ let

$$\gamma_{a,b}^k((X_n)_{n\in\mathbb{N}_0}) := \sup\{n \in \mathbb{N}_0 : \exists k_1 < \ldots < k_{2n} \in \{0, \ldots, k\} : X_{k_1} > b, X_{k_2} < a, \ldots, X_{k_{2n-1}} > b, X_{k_{2n}} < a\}$$

be the number of downcrossings of the interval $[a, b]$ by the first $k+1$ terms of the sequence $(X_n)_{n\in\mathbb{N}}$. Clearly

$$0 \leq \gamma_{a,b}^k((X_n)_{n\in\mathbb{N}_0}) \leq \gamma_{a,b}((X_n)_{n\in\mathbb{N}_0}),$$

so that by Beppo Levi’s theorem

$$E(\gamma_{a,b}((X_n)_{n\in\mathbb{N}_0})) = \lim_{k \to \infty} E(\gamma_{a,b}^k((X_n)_{n\in\mathbb{N}_0})).$$

---

15 Joseph L. Doob (1910–2004), American mathematician
Therefore it suffices to show that for each $k \in \mathbb{N}$ we have
\[
E(\gamma_{a,b}^k((X_n)_{n \in \mathbb{N}_0})) \leq \frac{1}{b-a} \sup_{n \in \mathbb{N}} E((X_n - b)_+).
\]

We now fix $k \in \mathbb{N}$ and let
\[
T_1 := (k + 1) \wedge \inf\{n \in \mathbb{N}_0 \mid X_n > b\},
T_2 := (k + 1) \wedge \inf\{n > T_1 \mid X_n < a\},
T_3 := (k + 1) \wedge \inf\{n > T_2 \mid X_n > b\}, \text{ etc.}
\]

Each $T_i$ is a bounded stopping time (exercise!). Now let $A_i := \{T_i \leq k\} \in \mathcal{F}_{T_i}$ (because $T_i$ is $\mathcal{F}_{T_i}$-measurable). Then we have:

1. $A_1 \supseteq A_2 \supseteq \ldots$,
2. $T_{2l} \leq k \Leftrightarrow \gamma_{a,b}^k((X_n)_{n \in \mathbb{N}_0}) \geq l$, so that $P\left(\gamma_{a,b}^k((X_n)_{n \in \mathbb{N}_0}) \geq l\right) = P(A_{2l})$,
3. $X_{T_{2l-1}} > b$ on $A_{2l-1} = \{T_{2l-1} \leq k\}$,
4. and $X_{T_{2l}} < a$ on $A_{2l} = \{T_{2l} \leq k\}$.

By the optional sampling theorem we have
\[
E(X_{T_{2l}} - b \mid \mathcal{F}_{T_{2l-1}}) \geq X_{T_{2l-1}} - b
\]
and therefore
\[
\int_{A_{2l-1}} (X_{T_{2l-1}} - b) dP \leq \int_{A_{2l-1}} (X_{T_{2l}} - b) dP;
\]
and
\[
E(X_{k+1} - b \mid \mathcal{F}_{T_{2l}}) \geq X_{T_{2l}} - b
\]
and therefore
\[
\int_{A_{2l-1} \setminus A_{2l}} (X_{T_{2l}} - b) dP \leq \int_{A_{2l-1} \setminus A_{2l}} (X_{k+1} - b) dP.
\]

It follows that
\[
0 \leq \int_{A_{2l-1}} (X_{T_{2l-1}} - b) dP \leq \int_{A_{2l-1}} (X_{T_{2l}} - b) dP
\]
\[= \int_{A_{2l-1} \setminus A_{2l}} (X_{T_{2l}} - b) dP + \int_{A_{2l}} (X_{T_{2l}} - b) dP
\]
\[\leq \int_{A_{2l-1} \setminus A_{2l}} (X_{k+1} - b) dP + \int_{A_{2l}} (a - b) dP
\]
\[= \int_{A_{2l-1} \setminus A_{2l}} (X_{k+1} - b) dP - (b - a) P(A_{2l}),
\]
so that
\[
(b - a) P(A_{2l}) \leq \int_{A_{2l-1} \setminus A_{2l}} (X_{k+1} - b) dP.
\]
Consequently
\[
E(\gamma_{a,b}((X_n)_{n \in \mathbb{N}_0})) = \sum_{l=1}^{\infty} P \left( \gamma_{a,b}((X_n)_{n \in \mathbb{N}_0}) \geq l \right)
= \sum_{l=1}^{\infty} P(T_{2l} \leq k)
= \sum_{l=1}^{\infty} P(A_{2l})
\leq \frac{1}{b-a} \sum_{l=1}^{\infty} \int_{A_{2l-1} \setminus A_{2l}}^{} (X_{k+1} - b)_+ dP.
\]

Since \(A_1 \supseteq A_2 \supseteq \ldots\), the sets \((A_{2l-1} \setminus A_{2l})_{l \in \mathbb{N}}\) are pairwise disjoint, so that
\[
\sum_{l=1}^{\infty} \int_{A_{2l-1} \setminus A_{2l}}^{} (X_{k+1} - b)_+ dP = \int_{\bigcup_{l \in \mathbb{N}} (A_{2l-1} \setminus A_{2l})}^{} (X_{k+1} - b)_+ dP
\leq E((X_{k+1} - b)_+)
\leq \sup_{n \in \mathbb{N}} E((X_n - b)_+). \quad \square
\]

11.2 Doob’s convergence theorem

Lemma 11.5. For a submartingale \((X_n)_{n \in \mathbb{N}_0}\) the following conditions are equivalent:

1. \(\sup_{n \in \mathbb{N}_0} E((X_n)_+) < \infty\),
2. \(\sup_{n \in \mathbb{N}_0} E(|X_n|) < \infty\).

Proof. Since \((X_n)_+ \leq |X_n|\), it is clear that the second condition implies the first one. Conversely, since
\[
|X_n| = X_n^+ + X_n^- = 2X_n^+ - X_n,
\]
we have
\[
E(|X_n|) \leq 2E(X_n^+) - E(X_0),
\]
so that the second condition implies the first one. \quad \square

Theorem 11.6 (Doob’s convergence theorem). Let \((X_n)_{n \in \mathbb{N}_0}\) be a submartingale satisfying \(\sup_{n \in \mathbb{N}_0} E((X_n)_+) < \infty\) (or, equivalently, \(\sup_{n \in \mathbb{N}_0} E(|X_n|) < \infty\)). Then the sequence \((X_n)_{n \in \mathbb{N}_0}\) converges almost surely to an integrable random variable \(X_\infty\).

Proof. Since \((X_n - b)_+ \leq (X_n)_+ + b_-\), Doob’s downcrossing inequality implies that \(E(\gamma_{a,b}((X_n)_{n \in \mathbb{N}_0}))\) is finite for all \(a < b \in \mathbb{R}\). Consequently, \(\gamma_{a,b}((X_n)_{n \in \mathbb{N}_0})\) is finite almost surely for all \(a < b \in \mathbb{R}\). More precisely, letting \(\Omega_{a,b} := \{\gamma_{a,b}((X_n)_{n \in \mathbb{N}_0}) < \infty\}\), we have \(P(\Omega_{a,b}) = 1\) for all \(a < b \in \mathbb{R}\). Now let
\[
\Omega^* := \bigcap_{a < b \in \mathbb{Q}} \Omega_{a,b}.
\]
Then \(P(\Omega^*) = 1\), and on \(\Omega^*\) the sequence \((X_n)_{n \in \mathbb{N}}\) converges to an \(\bar{\mathbb{R}}\)-valued random variable \(X_\infty\). By Fatou’s lemma we have
\[
E(|X_\infty|) \leq \liminf_{n \to \infty} E(|X_n|) < \infty. \quad \square
\]
Now the following two questions naturally arise:

1. Does the convergence $X_n \to X_\infty$ also take place in $L^1$?

2. Letting $\mathcal{F}_\infty := \sigma(\bigcup_{n \in \mathbb{N}_0} \mathcal{F}_n)$, is the process $(X_n)_{n \in \mathbb{N}_0 \cup \{\infty\}}$ a submartingale with respect to the filtration $(\mathcal{F}_n)_{n \in \mathbb{N}_0 \cup \{\infty\}}$?

**Example 11.7.** Let $(\xi_i)_{i \in \mathbb{N}}$ be a sequence of independent random variables with $P(\xi_i = 1) = P(\xi_i = -1) = 1/2$, and let $X_n := \sum_{i=1}^{n} \xi_i$. Then the process $(X_n)_{n \in \mathbb{N}}$ is a martingale (exercise!). Now fix $k > 0$, let $T_k := \inf\{n \in \mathbb{N}_0 : X_n = k\}$ and $Y_n := X_n \wedge T_k$. By the optional stopping theorem the process $(Y_n)_{n \in \mathbb{N}_0}$ is a martingale as well. Moreover, $Y_n \leq k$ for all $n \in \mathbb{N}_0$, so that we can apply Doob's convergence theorem:

$$\exists Y_\infty : Y_n \to Y_\infty \text{ a.s } (n \to \infty).$$

Obviously, $Y_\infty = k$. Since $E(Y_n) = 0$ for all $n \in \mathbb{N}$, but $E(Y_\infty) = k > 0$, we neither have $Y_n \to Y_\infty$ in $L^1$, nor is the process $(Y_n)_{n \in \mathbb{N}_0 \cup \{\infty\}}$ a martingale with respect to the filtration $(\mathcal{F}_n)_{n \in \mathbb{N}_0 \cup \{\infty\}}$.

However, the answer to the above questions is YES for *uniformly integrable martingales*.

### 11.3 Uniformly integrable martingales

**Lemma 11.8.** A random variable $X$ is integrable if and only if

$$\lim_{M \to \infty} \int_{\{|X| > M\}} |X|dP = 0. \tag{7}$$

**Proof.** Assume first that (7) holds. Then $\int_{\{|X| > M\}} |X|dP$ is finite for some $M$, and consequently

$$E(|X|) = \int_{\{|X| \leq M\}} |X|dP + \int_{\{|X| > M\}} |X|dP \leq M + \int_{\{|X| > M\}} |X|dP < \infty.$$  

Conversely, assume that $X$ is integrable. We have $0 \leq |X| \cdot 1_{\{|X| \leq M\}} \nearrow X$, so that by Beppo Levi’s theorem

$$\int_{\{|X| \leq M\}} |X|dP \nearrow E(|X|).$$

Consequently, since by assumption $E(|X|)$ is finite,

$$\int_{\{|X| > M\}} |X|dP = E(|X|) - \int_{\{|X| \leq M\}} |X|dP \to 0$$
as $M \to \infty$. \qed 

**Definition 11.9.** A family $(X_i)_{i \in I}$ of random variables is said to be *uniformly integrable* if it satisfies (7) uniformly in $i$, i.e. if

$$\lim_{M \to \infty} \sup_{i \in I} \int_{\{|X_i| > M\}} |X_i|dP = 0. \tag{8}$$
Remark 11.10. Any finite family of integrable random variables is clearly uniformly integrable.

Proposition 11.11. Let $\mathcal{A}(\varepsilon) := \{ A \in \mathcal{A} \mid P(A) \leq \varepsilon \}$. A family $(X_i)$ of random variables is uniformly integrable if and only if

1. it is bounded in $L^1$, i.e. $\sup_{i \in I} E(|X_i|) < \infty$, and
2. \[
\lim_{\varepsilon \to 0} \sup_{A \in \mathcal{A}(\varepsilon), i \in I} \int_A |X_i| dP = 0. \tag{9}
\]

Proof of Proposition 11.11. Assume first that the family $(X_i)_{i \in I}$ is uniformly integrable. Then there exists $M > 0$ such that $\int_{\{ |X_i| > M \}} |X_i| dP \leq 1$ for all $i \in I$. Consequently,

$$E(|X_i|) = \int_{\{ |X_i| \leq M \}} |X_i| dP + \int_{\{ |X_i| > M \}} |X_i| dP \leq M + 1,$$

so that the family $(X_i)_{i \in I}$ is indeed bounded in $L^1$. To prove (9), we have to show that for any given $\delta > 0$ there exists $\varepsilon > 0$ such that

$$\sup_{A \in \mathcal{A}(\varepsilon), i \in I} \int_A |X_i| dP \leq \delta.$$

To do so, we first choose $M > 0$ in such a way that

$$\sup_{i \in I} \int_{\{ |X_i| > M \}} |X_i| dP \leq \delta/2,$$

and then choose $\varepsilon := \delta/(2M)$. Then for all $A \in \mathcal{A}(\varepsilon)$ and all $i \in I$ we have

$$\int_A |X_i| dP = \int_{A \cap \{ |X_i| \leq M \}} |X_i| dP + \int_{A \cap \{ |X_i| > M \}} |X_i| dP \leq MP(A) + \frac{\delta}{2} \leq \delta,$$

which finishes the proof of the first implication.

Now assume that the family $(X_i)_{i \in I}$ is bounded in $L^1$ (say $E(|X_i|) \leq C$ for all $i \in I$) and satisfies (9). Then by Chebyshev’s inequality for each $i \in I$

$$P\{ |X_i| > M \} \leq \frac{1}{M} E(|X_i|) \leq \frac{C}{M},$$

i.e. $\{ |X_i| > M \} \in \mathcal{A}(C/M)$. Consequently

$$\sup_{i \in I} \int_{\{ |X_i| > M \}} |X_i| dP \leq \sup_{A \in \mathcal{A}(C/M), i \in I} \int_A |X_i| dP \to 0 \quad (M \to \infty). \quad \square$$

Example 11.12. Let $X$ be an integrable random variable on a probability space $(\Omega, \mathcal{A}, P)$ and $\mathcal{A}^*$ the set of all sub-$\sigma$-fields of $\mathcal{A}$. Then the family $(E(X|\mathcal{F}))_{\mathcal{F} \in \mathcal{A}^*}$ is uniformly integrable.
Proof. We have to show that
\[
\sup_{F \in \mathcal{A}} \int_{\{|E(X|F)| > M\}} |E(X|F)|dP \to 0
\]
as \(M \to \infty\). Chebyshev’s inequality implies that for each sub-\(\sigma\)-field \(F\) of \(\mathcal{A}\) we have
\[
P\{|E(X|F)| > M\} \leq \frac{1}{M} E(|X|),
\]
so that
\[
\{|E(X|F)| > M\} \in \mathcal{A}(E(|X|)/M).
\]
Consequently
\[
\sup_{F \in \mathcal{A}^*} \int_{\{|E(X|F)| > M\}} |E(X|F)|dP \leq \sup_{F \in \mathcal{A}^*} \int_{\{|E(X|F)| > M\}} E(|X||F)dP
\]
\[
= \sup_{F \in \mathcal{A}^*} \int_{\{|E(X|F)| > M\}} |X|dP
\]
\[
\leq \sup_{A \in \mathcal{A}(E(|X|)/M)} \int_A |X|dP
\]
\[
\to 0 \quad (M \to \infty).
\]

Proposition 11.13 (Sufficient criteria for uniform integrability). Let \((X_i)\) be a family of random variables. Each of the following two conditions implies that it is uniformly integrable:

1. It is dominated in \(L^1\), i.e. there exists an integrable random variable \(X\) such that \(|X_i| \leq X\) for all \(i \in I\).

2. It is bounded in \(L^p\) for some \(p > 1\), i.e. \(\sup_{i \in I} E(|X_i|^p) < \infty\)

Proof. If the first condition is satisfied, the family \((X_i)_{i \in I}\) is clearly bounded in \(L^1\) (by \(E(|X|))\). Moreover we have
\[
\sup_{A \in \mathcal{A}(\varepsilon), i \in I} \int_A |X_i|dP \leq \sup_{A \in \mathcal{A}(\varepsilon)} \int_A |X|dP \to 0
\]
because every single integrable random variable is uniformly integrable.

Now assume that the second condition holds. Then
\[
\sup_{i \in I} \int_{\{|X_i| > M\}} |X_i|dP \leq \sup_{i \in I} \int_{\{|X_i| > M\}} \frac{|X_i|^p}{M^{p-1}}dP \leq \frac{1}{M^{p-1}} \sup_{i \in I} E(|X_i|^p) \to 0 \quad (M \to \infty).
\]

Theorem 11.14 (Vitali’s convergence theorem). Let \((X_n)_{n \in \mathbb{N}}\) be a sequence of integrable random variables, and let \(X\) be another integrable random variable. Then the following are equivalent:

1. \(X_n \to X\) in \(L^1\).

2. The sequence \((X_n)_{n \in \mathbb{N}}\) is uniformly integrable, and \(X_n \to X\) in probability.

\(^{16}\)Giuseppe Vitali (1875–1932), Italian mathematician
Remark 11.15. Compare with Lebesgue’s dominated convergence theorem.

Proof of Theorem 11.14. 1. ⇒ 2. We already know that convergence in $L^1$ implies convergence in probability. To show uniform integrability of the sequence $(X_n)_{n \in \mathbb{N}}$, we first show uniform integrability of the sequence $(X_n - X)_{n \in \mathbb{N}}$: For given $\varepsilon > 0$ let $n_0(\varepsilon) \in \mathbb{N}$ be such that $E(|X_n - X|) \leq \varepsilon$ for $n > n_0(\varepsilon)$. Now write

$$
\sup_{n \in \mathbb{N}} \int_{\{|X_n - X| > M\}} |X_n - X|dP = \max \left( \sup_{n \leq n_0(\varepsilon)} \int_{\{|X_n - X| > M\}} |X_n - X|dP, \sup_{n > n_0(\varepsilon)} \int_{\{|X_n - X| > M\}} |X_n - X|dP \right).
$$

Since any finite family of integrable random variables is uniformly integrable, the first supremum is $\leq \varepsilon$ for $M$ large enough. Moreover, the second supremum is bounded by $\sup_{n > n_0(\varepsilon)} E(|X_n - X|) \leq \varepsilon$.

Uniform integrability of the sequence $(X_n - X)_{n \in \mathbb{N}}$ implies that

$$
\lim_{\varepsilon \to 0} \sup_{A \in \mathcal{A}(\varepsilon)} \sup_{n \in \mathbb{N}} \int_A |X_n - X|dP = 0,
$$

and consequently

$$
\sup_{A \in \mathcal{A}(\varepsilon)} \sup_{n \in \mathbb{N}} \int_A |X_n|dP \leq \sup_{A \in \mathcal{A}(\varepsilon)} \sup_{n \in \mathbb{N}} \int_A |X_n - X|dP + \sup_{A \in \mathcal{A}(\varepsilon)} \int_A |X|dP \to 0 \quad (\varepsilon \to 0),
$$

so that the sequence $(X_n)_{n \in \mathbb{N}}$ is uniformly integrable as well.

2. ⇒ 1. We first show that the sequence $(X_n - X)_{n \in \mathbb{N}}$ is uniformly integrable as well:

$$
\sup_{A \in \mathcal{A}(\varepsilon)} \sup_{n \in \mathbb{N}} \int_A |X_n - X|dP \leq \sup_{A \in \mathcal{A}(\varepsilon)} \sup_{n \in \mathbb{N}} \int_A |X_n|dP + \sup_{A \in \mathcal{A}(\varepsilon)} \int_A |X|dP \to 0 \quad (\varepsilon \to 0).
$$

To prove that $X_n \to X$ in $L^1$ we have to show that for each $\varepsilon > 0$ we have

$$
E(|X_n - X|) \leq \varepsilon
$$

for all large enough $n \in \mathbb{N}$. For each $\delta > 0$ choose $n_0(\delta) \in \mathbb{N}$ in such a way that

$$
P(|X_n - X| > \delta) \leq \delta.
$$

for all $n \geq n_0(\delta)$ (this is possible because of the convergence in probability of $X_n \to X$).

Then we have $\{|X_n - X| > \delta\} \in \mathcal{A}(\delta)$ for all $n \geq n_0(\delta)$, so that

$$
\sup_{n \geq n_0(\delta)} \int_{\{|X_n - X| > \delta\}} |X_n - X|dP \leq \sup_{A \in \mathcal{A}(\delta)} \sup_{n \in \mathbb{N}} \int_A |X_n - X|dP \to 0 \quad (\delta \to 0).
$$

Now choose $\delta > 0$ in such a way that $\delta \leq \varepsilon/2$ and

$$
\sup_{n \geq n_0(\delta)} \int_{\{|X_n - X| > \delta\}} |X_n - X|dP \leq \frac{\varepsilon}{2}.
$$

Then we have for all $n \geq n_0(\delta)$:

$$
E(|X_n - X|) \leq \int_{\{|X_n - X| \leq \delta\}} |X_n - X|dP + \int_{\{|X_n - X| > \delta\}} |X_n - X|dP \leq \delta + \frac{\varepsilon}{2} \leq \varepsilon. \quad \square
$$
Theorem 11.16. Let \((X_n)_{n \in \mathbb{N}_0}\) be a submartingale on a filtered probability space \((\Omega, \mathcal{A}, \mathbb{P}, (\mathcal{F}_n)_{n \in \mathbb{N}_0})\), and let \(\mathcal{F}_\infty := \sigma(\mathcal{F}_n, n \in \mathbb{N}_0)\).

1. It converges in \(L^1\) if and only if it is uniformly integrable.

2. In this case the process \((X_n)_{n \in \mathbb{N}_0 \cup \{\infty\}}\) (where \(X_\infty := \lim_{n \to \infty} X_n\)) is a submartingale with respect to the filtration \((\mathcal{F}_n)_{n \in \mathbb{N}_0 \cup \{\infty\}}\).

Proof. Convergence in \(L^1\) implies uniform integrability by Vitali's convergence theorem. Conversely, assume that the submartingale is uniformly integrable. Then it is bounded in \(L^1\), so that by Doob's convergence theorem it converges almost surely to an integrable random variable \(X_\infty\). Together with uniform integrability this implies convergence in \(L^1\) by Vitali's convergence theorem.

Finally, to show that the process \((X_n)_{n \in \mathbb{N}_0 \cup \{\infty\}}\) is a submartingale, we have to show that
\[
\int_A X_\infty d\mathbb{P} \geq \int_A X_n d\mathbb{P}
\]
for all \(n \in \mathbb{N}_0\) and all \(A \in \mathcal{F}_n\). Since \(X_n \to X_\infty\) in \(L^1\) we have
\[
\int_A X_\infty d\mathbb{P} = \lim_{m \to \infty} \int_A X_m d\mathbb{P} \geq \int_A X_n d\mathbb{P}. \quad \Box
\]

Theorem 11.17. Let \((X_n)_{n \in \mathbb{N}_0}\) be a martingale on a filtered probability space \((\Omega, \mathcal{A}, \mathbb{P}, (\mathcal{F}_n)_{n \in \mathbb{N}_0})\), and let \(\mathcal{F}_\infty := \sigma(\mathcal{F}_n, n \in \mathbb{N}_0)\). Then the following conditions are equivalent:

1. The martingale \((X_n)_{n \in \mathbb{N}_0}\) is uniformly integrable.

2. It converges in \(L^1\) to some limit \(X^*\).

3. It is closable, i.e. there exists an \(\mathcal{F}_\infty\)-measurable random variable \(X_\infty\) such that the process \((X_n)_{n \in \mathbb{N}_0 \cup \{\infty\}}\) is a martingale with respect to the filtration \((\mathcal{F}_n)_{n \in \mathbb{N}_0 \cup \{\infty\}}\).

Moreover, in this case \(X_\infty = X^*\).

Proof. 1. \(\iff\) 2. \(\implies\) 3. These implications follow from the previous theorem.

3. \(\implies\) 1. Since the family \((E(X_\infty|\mathcal{F}))_F\) sub-\(\sigma\)-field of \(\mathcal{A}\) is uniformly integrable, the sequence \((X_n)_{n \in \mathbb{N}}\) is uniformly integrable as well.

To prove that \(X_\infty = X^*\), note that for all \(A \in \bigcup_{m \in \mathbb{N}_0} \mathcal{F}_m\) we have
\[
\int_A X^* d\mathbb{P} = \lim_{n \to \infty} \int_A X_n d\mathbb{P} = \int_A X_\infty d\mathbb{P}. \quad (10)
\]

Since \(\{ A \in \mathcal{A} | \int_A X^* d\mathbb{P} = \int_A X_\infty d\mathbb{P} \}\) is a \(\sigma\)-field, it follows that (10) holds for all \(A \in \mathcal{F}_\infty\). Since \(X^*\) and \(X_\infty\) are both \(\mathcal{F}_\infty\)-measurable \((X^*\) being a limit of \(\mathcal{F}_\infty\)-measurable random variables and \(X_\infty\) by assumption), it follows that \(X_\infty = X^*\). \(\Box\)

Theorem 11.18 \((L^p\)-convergence theorem\). Let \((X_n)_{n \in \mathbb{N}_0}\) be a martingale satisfying \(\sup_{n \in \mathbb{N}_0} E(|X_n|^p) < \infty\) for some \(p > 1\). Then it converges almost surely and in \(L^p\).
Proof. Since \((X_n)_{n \in \mathbb{N}_0}\) is bounded in \(L^p\), it is uniformly integrable and therefore converges almost surely to some limit \(X_\infty\). To prove convergence in \(L^p\), let \(Y := \sup_{n \in \mathbb{N}_0} |X_n|\). Clearly,

\[|X_n - X_\infty|^p \leq (2Y)^p.\]

By the second maximal inequality,

\[E(Y^p) = E\left(\sup_{n \in \mathbb{N}_0} |X_n|^p\right) \leq \left(\frac{p}{p-1}\right)^p \sup_{n \in \mathbb{N}_0} E(|X_n|^p) < \infty,\]

so that \(Y^p \in L^1\). Consequently, Lebesgue’s dominated convergence theorem implies that

\[\lim_{n \to \infty} E(|X_n - X_\infty|^p) = 0.\]

Remark 11.19. This is wrong for \(p = 1\).

Example 11.20. Let \((\xi_i)_{i \in \mathbb{N}}\) be a sequence of independent random variables such that \(P(\xi_i = 1) = P(\xi_i = -1) = 1/2\) for all \(i \in \mathbb{N}\). Let \(X_n := 1 + \sum_{i=1}^n \xi_i\) and \(Y_n := X_{T(0)^\wedge n}\). Then the process \((Y_n)_{n \in \mathbb{N}_0}\) is a non-negative martingale so that

\[E(|Y_n|) = E(Y_n) = E(Y_0) = 1.\]

Moreover, it converges almost surely to 0, but not in \(L^1\) (because \(E(Y_n) = 1 \neq 0\) for all \(N \in \mathbb{N}_0\)).

12 Backward martingales, sub- and supermartingales

So far we have mainly studied (sub-/super-)martingales with index set \(\mathbb{N}_0 = \{0, 1, 2, \ldots\}\). In this section we study “backward (sub-/super-)martingales”, i.e. (sub-/super-)martingales with index set \(-\mathbb{N}_0 = \{\ldots, -2, -1, 0\}\). In contrast to “usual” (sub)martingales, backward (sub)martingales always have the following nice properties:

1. Because of the convexity and monotonicity of the function \(x \mapsto x^+\) we have \(E(X_0^+|\mathcal{F}_n) \geq E(X_0|\mathcal{F}_n)^+ \geq X_n^+\) and therefore \(E(X_n^+)^- \leq E(X_0^+)\) for all \(n \in -\mathbb{N}_0\), so that every backward submartingale satisfies

\[\sup_{n \in -\mathbb{N}_0} E(X_n^+) = E(X_0^+) < \infty.\]

2. Every backward martingale is uniformly integrable (because the family \((X_n)_{n \in -\mathbb{N}_0}\) is contained in the uniformly integrable family \((E(X_0)\mathcal{C})_{\mathcal{C} \subseteq \mathcal{A}^*}\)).

Theorem 12.1 (Downcrossing inequality for backward submartingales). Let \((X_n)_{n \in -\mathbb{N}_0}\) be a backward submartingale on a filtered probability space \((\Omega, \mathcal{A}, P, (\mathcal{F}_n)_{n \in -\mathbb{N}_0})\), and let \(a < b\). Then

\[E(\gamma_{a,b}(X_n)_{n \in -\mathbb{N}_0}) \leq \frac{1}{b-a} \sup_{n \in -\mathbb{N}_0} E((X_n - b)^+) = \frac{1}{b-a} E((X_0 - b)^+).\]

Proof. The first inequality can be obtained in the same way as the usual downcrossing inequality. For the second inequality note that convexity and monotonicity of the function \(x \mapsto x^+\) imply that

\[E((X_0 - b)^+|\mathcal{F}_n) \geq E(X_0 - b|\mathcal{F}_n)^+ \geq (X_n - b)^+.\]
Corollary 12.2 (Convergence theorem for backward submartingales). Any backward submartingale \((X_n)_{n \in \mathbb{N}_0}\) converges almost surely as \(n \to -\infty\).

Warning. The limit need not be finite.

Example 12.3. Let \(X_n := n\). Then \((X_n)_{n \in \mathbb{N}_0}\) is a submartingale which converges to \(-\infty\) as \(n \to -\infty\).

Theorem 12.4 (Convergence theorem for backward martingales). Any backward martingale \((X_n)_{n \in \mathbb{N}_0}\) converges almost surely and in \(L^1\) to \(E(X_0|\mathcal{F}_-\infty)\) as \(n \to -\infty\), where \(\mathcal{F}_-\infty := \bigcap_{n \in \mathbb{N}} \mathcal{F}_n\).

Proof. By the convergence theorem for backward submartingales the sequence \((X_n)_{n \in \mathbb{N}_0}\) converges almost surely to a (not necessarily finite) random variable \(X_\infty\) as \(n \to -\infty\).

Since the process \(|X_n|\)_{n \in \mathbb{N}_0} is a submartingale, Fatou’s lemma implies that

\[
E(|X_\infty|) = \liminf_{n \to -\infty} E(|X_n|) \leq E(|X_0|),
\]

so that \(X_\infty\) is integrable. Since the sequence \((X_n)_{n \in \mathbb{N}_0}\) is uniformly integrable (because \(X_n = E(X_0|\mathcal{F}_n)\) the convergence \(X_n \to X_\infty\) takes also place in \(L^1\) (by Vitali’s theorem).

To show that \(X_\infty = E(X_0|\mathcal{F}_-\infty)\), note that \(X_\infty\) is \(\mathcal{F}_-\infty\)-measurable, so that it suffices to check that

\[
\int_A X_\infty dP = \int_A E(X_0|\mathcal{F}_-\infty) dP
\]

for all \(A \in \mathcal{F}_-\infty\): Thanks to \(L^1\)-convergence we have

\[
\int_A X_\infty dP = \lim_{n \to -\infty} \int_A X_n dP = \lim_{n \to -\infty} \int_A E(X_n|\mathcal{F}_-\infty) dP = \int_A E(X_0|\mathcal{F}_-\infty) dP. \quad \Box
\]

13 Applications of the martingale convergence theorems

Definition 13.1. Let \((\Omega, \mathcal{A}, P)\) be a probability space and \((\mathcal{A}_n)_{n \in \mathbb{N}_0}\) a sequence of sub-\(\sigma\)-fields of \(\mathcal{A}\). Then the **tail \(\sigma\)-field** of the sequence \((\mathcal{A}_n)_{n \in \mathbb{N}_0}\) is defined as

\[
\bigcap_{n \in \mathbb{N}_0} \sigma(\mathcal{A}_m, m \geq n).
\]

Theorem 13.2 (Kolmogorov’s\(^{17}\) zero-one law). Let \((\Omega, \mathcal{A}, P)\) be a probability space and \((\mathcal{A}_n)_{n \in \mathbb{N}_0}\) an independent sequence of sub-\(\sigma\)-fields of \(\mathcal{A}\). Then every event \(A\) belonging to the tail \(\sigma\)-field of this sequence satisfies \(P(A) \in \{0, 1\}\).

Proof. For each \(n \in \mathbb{N}_0 \cup \{\infty\}\) let \(\mathcal{F}_n := \sigma(\mathcal{A}_0, \ldots, \mathcal{A}_n)\) and \(X_n := E(1_A|\mathcal{F}_n)\). Then the sequence \((X_n)_{n \in \mathbb{N}_0 \cup \{\infty\}}\) is a closed martingale, so that

\[X_n \to X_\infty\]

almost surely and in \(L^1\). Since \(A\) is independent of \(\mathcal{F}_n\) for each \(n \in \mathbb{N}_0\), we have

\[X_n = E(1_A|\mathcal{F}_n) = E(1_A) = P(A)\]

for each \(n \in \mathbb{N}_0\). On the other hand, since \(A\) is measurable with respect to \(\mathcal{F}_\infty\), we have

\[X_\infty = E(1_A|\mathcal{F}_\infty) = 1_A.\]

Consequently, \(1_A = P(A)\), so that \(P(A) \in \{0, 1\}\). \(\Box\)

\(^{17}\)Andrey Nikolaevich Kolmogorov (1903–1987), Russian mathematician
Corollary 13.3. Let \((X_n)_{n \in \mathbb{N}}\) be a sequence of independent random variables. Then \(\limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i\) and \(\liminf_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i\) are almost surely constant.

**Proof.** Let \(\mathcal{T}\) be the tail \(\sigma\)-field of the sequence \((\sigma(X_n))_{n \in \mathbb{N}}\). Since for every \(m \in \mathbb{N}\)
\[
\limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \limsup_{n \to \infty} \frac{1}{n} \sum_{i=m}^{n} X_i,
\]
this lim sup is measurable with respect to \(\mathcal{T}\). The claim now follows from Kolmogorov’s zero-one law. \qed

**Theorem 13.4** (Kolmogorov’s strong law of large numbers). Let \((X_n)_{n \in \mathbb{N}}\) be a sequence of independent and identically distributed random variables, and let \(S_n := \sum_{i=1}^{n} X_i\). Then \[
\frac{S_n}{n} \to E(X_1) \quad \text{almost surely and in } L^1.
\]

For the proof we need the following lemma:

**Lemma 13.5.** For each \(n \in \mathbb{N}\) we have
\[
E(X_1 | \sigma(S_k, k \geq n)) = \frac{S_n}{n}.
\]

**Proof.** For reasons of symmetry
\[
E(X_1 | \sigma(S_k, k \geq n)) = E(X_1 | \sigma(S_k, k \geq n))
\]
for all \(i \in \{1, \ldots, n\}\). Therefore
\[
E(X_1 | \sigma(S_k, k \geq n)) = \frac{1}{n} \sum_{i=1}^{n} E(X_1 | \sigma(S_k, k \geq n)) = E \left( \frac{1}{n} \sum_{i=1}^{n} X_i | \sigma(S_k, k \geq n) \right) = E \left( \frac{S_n}{n} | \sigma(S_k, k \geq n) \right) = \frac{S_n}{n}. \quad \square
\]

**Proof of Theorem 13.4.** For each \(n \in \mathbb{N}_0\) let \(\mathcal{F}_n := \sigma(S_k, k \geq n)\) and \(Y_n := \frac{S_n}{n} \). Then \((\mathcal{F}_m)_{m \in \mathbb{N}}\) is a filtration, and the process \((Y_m)_{m \in \mathbb{N}}\) is adapted to it. Moreover by the lemma
\[
Y_n = \frac{S_n}{n} = E(X_1 | \mathcal{F}_n),
\]
so that \((Y_m)_{m \in \mathbb{N}}\) is a martingale with respect to \((\mathcal{F}_m)_{m \in \mathbb{N}}\). By the convergence theorem for backward martingales it converges almost surely and in \(L^1\) to \(E(X_1 | \mathcal{F}_{-\infty})\) as \(m \to -\infty\). By the previous corollary
\[
E(X_1 | \mathcal{F}_{-\infty}) = \lim_{n \to \infty} \frac{S_n}{n}
\]
is almost surely constant and hence equal to \(E(X_1)\). \qed
14 Construction of stochastic processes with prescribed properties

Recall the definition of a stochastic process:

**Definition 14.1.** Let \((\Omega, \mathcal{A}, P)\) be a probability space, \((E, \mathcal{B})\) a measurable space and \(I\) a set. A **stochastic process** on \((\Omega, \mathcal{A}, P)\) with **state space** \((E, \mathcal{B})\) and index set \(I\) is a family \((X_t)_{t \in I}\) of \((E, \mathcal{B})\)-valued random variables on \((\Omega, \mathcal{A}, P)\).

One is often interested in the following question: Given a set \(I\) and a measurable space \((E, \mathcal{B})\), do there exist a probability space \((\Omega, \mathcal{A}, P)\) and an \((E, \mathcal{B})\)-valued stochastic process \((X_t)_{t \in I}\) on \((\Omega, \mathcal{A}, P)\) having certain prescribed properties? Often these properties can be formulated in terms of the **finite-dimensional distributions** of the process.

**Definition 14.2.** Let \((E, \mathcal{B})\) be a measurable space and \((X_t)_{t \in I}\) an \((E, \mathcal{B})\)-valued stochastic process on a probability space \((\Omega, \mathcal{A}, P)\). Let \(I^*\) be the set of non-empty finite subsets of \(I\). For each \(J \in I^*\) we denote by \(X_J\) the \(E^J\)-valued random variable \(X_J := (X_t)_{t \in J}\) and by \(P_J(\cdot)\) its distribution, i.e. the joint distribution of the random variables \((X_t)_{t \in J}\). We then define the **family of finite-dimensional distributions** of the stochastic process \((X_t)_{t \in I}\) to be the family \(\{P_J(\cdot)\}_{J \in I^*}\).

**Remark 14.3.**

1. If \(J \subseteq J'\), then \(P_{J'}(\cdot)\) determines \(P_J(\cdot)\).
2. If \(J = \{t_0, \ldots, t_n\}\), \(P_J(\cdot)\) can be specified by specifying \(E(f(X_{t_0}, \ldots, X_{t_n}))\) for all bounded measurable functions \(f: \mathbb{R}^{n+1} \to \mathbb{R}\).

**Example 14.4** (Brownian motion\(^{18}\)). A one-dimensional standard **Brownian motion** (or **Wiener**\(^{19}\) **process**) is a real-valued stochastic process \((X_t)_{t \in \mathbb{R}_+}\) with the following properties:

1. \(X_0 = 0\) almost surely.
2. It has independent increments, i.e. for all \(n \in \mathbb{N}\) and all \(0 < t_1 < \ldots < t_n\) the random variables \(X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}\) are independent.
3. For all \(0 \leq t_1 < t_2\) the increment \(X_{t_2} - X_{t_1}\) is normally distributed with mean 0 and variance \(t_2 - t_1\).
4. For almost all \(\omega \in \Omega\) the map \(t \to X_t(\omega)\) is continuous.

We will now see that the first three properties of Brownian motion can be formulated in terms of the finite-dimensional distributions:

**Lemma 14.5.** A real-valued stochastic process \((X_t)_{t \in \mathbb{R}_+}\) satisfies the first three properties of Brownian motion if and only if for all \(0 = t_0 < t_1 < \ldots < t_n\) and every bounded measurable function \(f: \mathbb{R}^{n+1} \to \mathbb{R}\),

\[
E(f(X_{t_0}, \ldots, X_{t_n})) = \int_{\mathbb{R}^n} f(0, x_1, \ldots, x_n) \prod_{i=1}^n \nu_{t_i-t_{i-1}}(x_i - x_{i-1}) dx_1 \ldots dx_n. \tag{11}
\]

\(^{18}\)R. Brown, A brief account of microscopical observations made in the months of June, July and August, 1827, on the particles contained in the pollen of plants; and on the general existence of active molecules in organic and inorganic bodies. Phil. Mag. 4 (1828), 161–173. (Robert Brown (1773–1858), Scottish botanist)

\(^{19}\)Norbert Wiener (1894–1964), American mathematician and philosopher
Here $x_0 := 0$, and $\nu_t(x)$ denotes the density of the normal distribution with mean 0 and variance $t$, i.e.
\[
\nu_t(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}.
\]

**Remark 14.6.** Note that (11) specifies $P_J(X)$ for $J = \{0, t_1, \ldots, t_n\}$.

**Proof of Lemma 14.5.** Suppose first that $(X_t)_{t \in \mathbb{R}^+}$ satisfies the first three properties of Brownian motion. Let
\[
\tilde{f}(y_1, \ldots, y_n) := f(0, y_1, y_2, \ldots, y_1 + \ldots + y_n),
\]
so that
\[
f(0, x_1, \ldots, x_n) = \tilde{f}(x_1, x_2 - x_1, \ldots, x_n - x_{n-1}).
\]
Then it follows that
\[
E\left(f(X_{t_1}, \ldots, X_{t_n})\right) = E\left(f(0, X_{t_1}, \ldots, X_{t_n})\right)
= E\left(\tilde{f}(X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}})\right)
= \int_{\mathbb{R}^n} \tilde{f}(y_1, \ldots, y_n) P_{X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}}(dy_1, \ldots, dy_n)
= \int_{\mathbb{R}^n} \tilde{f}(y_1, \ldots, y_n) \prod_{i=1}^n P_{X_{t_i} - X_{t_{i-1}}}(dy_1, \ldots, dy_n)
= \int_{\mathbb{R}^n} \tilde{f}(y_1, \ldots, y_n) \prod_{i=1}^n \nu_{t_i - t_{i-1}}(y_i) dy_1, \ldots, dy_n
= \int_{\mathbb{R}^n} \tilde{f}(x_1, x_2 - x_1, \ldots, x_n - x_{n-1}) \prod_{i=1}^n \nu_{t_i - t_{i-1}}(x_i - x_{i-1}) dx_1, \ldots, dx_n
= \int_{\mathbb{R}^n} f(0, x_1, \ldots, x_n) \prod_{i=1}^n \nu_{t_i - t_{i-1}}(x_i - x_{i-1}) dx_1, \ldots, dx_n.
\]

Now assume that (11) holds. Clearly, (11) implies that $E(f(X_0)) = f(0)$, so that $X_0 = 0$ almost surely.

To prove that $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent and that $X_{t_i} - X_{t_{i-1}}$ is normally distributed with mean 0 and variance $t_i - t_{i-1}$, we have to show that
\[
E\left(\prod_{i=1}^n g(X_{t_i} - X_{t_{i-1}})\right) = \prod_{i=1}^n \int_{\mathbb{R}} g(x) \nu_{t_i - t_{i-1}}(x) dx
\]
for all bounded measurable functions $g_1, \ldots, g_n$. Let
\[
f(x_1, \ldots, x_n) := \prod_{i=1}^n g(x_i - x_{i-1})
\]
Definition 14.7. Let \((E, B)\) be a measurable space and \(I\) a set. For each non-empty finite subset \(J\) of \(I\) let \(P_J\) be a probability measure on \((E^J, B^J)\), where \(B^J := \otimes_{i \in J} B\).

1. The family \((P_J)_{J \in I^*}\) is consistent if the image measure of \(P_{J'}\) under \(\pi^J_{J'}\) equals \(P_J\) whenever \(J\) and \(J'\) are two non-empty finite subsets of \(I\) such that \(J \subseteq J'\).

2. A probability measure \(P_I\) on \((E^I, B^I)\) is called a projective limit of the family \((P_J)_{J \in I^*}\) if for every finite subset \(J\) of \(I\) the image of \(P_I\) under the projection \(\pi_J := \pi^J_J\) equals \(P_J\).

Remark 14.8. Clearly a projective limit can only exist if the family \((P_J)_{J \in I^*}\) is consistent.

Proposition 14.9. A consistent family \((P_J)_{J \in I^*}\) of probability measures is the family of finite-dimensional distributions of a stochastic process if and only if it has a projective limit.

Proof. Suppose first that \((X_t)_{t \in I}\) is a stochastic process such that \(P_J(X) = P_J\) for all \(J \in I^*\). Then its distribution is a projective limit of the family \((P_J)_{J \in I^*}\).

Conversely, let \(P_I\) be a projective limit of the family \((P_J)_{J \in I^*}\). Then choose

- \(\Omega := E^I\),
- \(A := B^I\),
- \(P := P_I\),
- \(X_t = \pi_t\), where \(\pi_t : \Omega \to E\) is the \(t\)-th projection, defined by \(\pi_t((\omega_i)_{i \in I}) := \omega_t\).  \(\square\)
It turns out that in many cases consistency of a family \((P_J)_{J \in I^*}\) is sufficient for the existence of a projective limit and hence of a stochastic process \((X_t)_{t \in I}\) with finite-dimensional distributions \((P_J)_{J \in I^*}\).

**Definition 14.10.** A topological space \(E\) is **Polish** if

1. it is completely metrizable, i.e. if there exists a metric \(d\) on \(E\) such that
   - (a) \(d\) induces the topology of \(E\) and
   - (b) the metric space \((E, d)\) is complete, and
2. it is separable, i.e. it contains a countable dense set.

**Example 14.11.** \(\mathbb{R}^n\), separable Banach spaces.

**Remark 14.12.** If \(E\) and \(F\) are both Polish, then so is \(E \times F\) (choose \(d((e, f), (e', f')) := d_1(e, e') + d_2(f, f')\)).

**Theorem 14.13.** Every finite measure \(\mu\) on a Polish space \(E\) is regular, i.e. for every \(B \in \mathcal{B}(E)\),

\[
\mu(B) = \sup \{ \mu(K) \mid K \subseteq B, \ K \text{ compact} \} \tag{12}
\]

\[
= \inf \{ \mu(U) \mid U \supseteq B, \ U \text{ open} \}. \tag{13}
\]

**Proof.** See books on measure theory, e.g. *Maß- und Integrationstheorie* by Bauer or *Maß- und Integrationstheorie* by Elstrodt.

**Theorem 14.14** (Kolmogorov’s consistency theorem, measure-theoretic version). Let \(E\) be a Polish space (equipped with its Borel \(\sigma\)-field \(\mathcal{B}\)), \(I\) a set and \((P_J)_{J \in I^*}\) a consistent family of probability measures. Then it has a unique projective limit.

**Corollary 14.15** (Probabilistic version of Kolmogorov’s consistency theorem). Let \(E\) be a Polish space (equipped with its Borel \(\sigma\)-field \(\mathcal{B}\)), \(I\) a set and \((P_J)_{J \in I^*}\) a consistent family of probability measures. Then there exists a stochastic process with finite-dimensional distributions \((P_J)_{J \in I^*}\).

**Proof of Theorem 14.14.** For each \(J \in I^*\) let

\[
\mathcal{B}_J := \left\{ B \times E^{I \setminus J} \mid B \in \mathcal{B}^J \right\}.
\]

Clearly, \(\mathcal{B}_J\) is a \(\sigma\)-field on \(E^J\), and \(\mathcal{B}_J \subseteq \mathcal{B}_{J'}\) whenever \(J \subseteq J'\). Consequently, \(\mathcal{B}_I := \bigcup_{J \in I^*} \mathcal{B}_J\) is an algebra, i.e.

1. \(\emptyset \in \mathcal{B}_I\),
2. for all \(A, B \in \mathcal{B}_I\), we have \(A \cup B \in \mathcal{B}_I\), and
3. for all \(A \in \mathcal{B}_I\), we have \(A^c \in \mathcal{B}_I\).

Moreover, \(\mathcal{B}_I^\ast\) is the \(\sigma\)-field generated by \(\mathcal{B}_I\). We now define \(P_I\) on \(\mathcal{B}_I\) in the following way: for any \(A \in \mathcal{B}_I\) there exist \(J \in I^*\) and \(B \in \mathcal{B}^J\) such that \(A = B \times E^{I \setminus J}\). We then set

\[
P_I(A) := P_J(B).
\]
Consistency of the family \((P_J)_{J \in I}\) precisely means that this definition does not depend on the choice of \(J\). Since each \(P_J\) is a probability measure, it is clear that \(P_I\) is a finitely additive set function on \(\mathcal{B}_I\) satisfying \(P_I(\emptyset) = 1\).

In order to apply Carathéodory’s extension theorem, we have to show that \(P_I\) is \(\sigma\)-additive. From measure theory we know that this is the case if and only if for every sequence \((A_n)_{n \in \mathbb{N}}\) of elements of \(\mathcal{B}_I\) satisfying \(A_n \searrow \emptyset\) (i.e. \(A_n \subseteq A_{n+1}\) for \(n \geq m\) and \(\bigcap_{n \in \mathbb{N}} A_n = \emptyset\)) we have

\[
\lim_{n \to \infty} P_I(A_n) = 0.
\]

Suppose that this is not true. Then there exist \(\varepsilon > 0\) and a sequence \((A_n)_{n \in \mathbb{N}}\) of elements of \(\mathcal{B}_I\) satisfying \(A_n \searrow \emptyset\) and \(P_I(A_n) \geq \varepsilon\) for all \(n \in \mathbb{N}\). Let \(A_n = B_n \times E^I \setminus J_n\) with \(J_n \in I^*\) and \(B_n \in \mathcal{B}^{J_n}\). Without loss of generality we may assume that \(J_n \subseteq J_{n+1}\) for all \(n \in \mathbb{N}\). By Remark 14.12 \(E^{J_n}\) is Polish, so that by Proposition 14.13 \(P_{J_n}\) is regular. Therefore there exists a compact subset \(K_n\) of \(B_n\) such that

\[
P_{J_n}(B_n \setminus K_n) \leq 2^{-n} \varepsilon.
\]

Let \(\tilde{K}_n := K_n \times E^I \setminus J_n\) and \(L_n := \bigcap_{m=1}^n \tilde{K}_m\). Then the sequence \((L_n)_{n \in \mathbb{N}}\) is decreasing. Moreover, we have \(L_n \subseteq \tilde{K}_n \subseteq A_n\) and therefore

\[
L_n \searrow \emptyset.
\]

Moreover,

\[
P_I(A_n \setminus \tilde{K}_n) = P_{J_n}(B_n \setminus K_n) \leq 2^{-n} \varepsilon
\]

and consequently

\[
P_I(A_n \setminus L_n) = P_I\left( A_n \setminus \bigcap_{m=1}^n \tilde{K}_m \right)
= P_I\left( \bigcup_{m=1}^n \left( A_n \setminus \tilde{K}_m \right) \right)
\leq P_I\left( \bigcup_{m=1}^n \left( A_m \setminus \tilde{K}_m \right) \right)
\leq \sum_{m=1}^n P_I\left( A_m \setminus \tilde{K}_m \right)
\leq \sum_{m=1}^n 2^{-m} \varepsilon < \varepsilon.
\]

Since \(P_I(A_n) \geq \varepsilon\), this implies that \(P_I(L_n) > 0\), in particular \(L_n\) is non-empty. For each \(n \in \mathbb{N}\) let \(\omega_n\) be an arbitrary element of \(L_n\). Since the sequence \((L_n)_{n \in \mathbb{N}}\) is decreasing, we have \(\omega_m \in L_n \subseteq \tilde{K}_n\) for all \(m \geq n\), and consequently \(\pi_{J_n}(\omega_m) \in K_n\). This implies that for all \(t \in \bigcup_{n \in \mathbb{N}} J_n\) all but a finite number of terms of the sequence \((\pi_t(\omega_m))_{m \in \mathbb{N}}\) belong to a compact set. Namely, if \(t \in J_n\) we have

\[
\pi_t(\omega_m) \in \pi_t(K_n)
\]

for all \(m \geq n\). Let now \((t_1, t_2, \ldots)\) be an enumeration of the countable set \(\bigcup_{n \in \mathbb{N}} J_n\). Because of (15) there exists a subsequence \((\omega_m^1)_{m \in \mathbb{N}}\) of the sequence \((\omega_m^2)_{m \in \mathbb{N}}\) such that \(\pi_{t_1}(\omega_m^1)\) converges as \(m \to \infty\), a further subsequence \((\omega_m^2)_{m \in \mathbb{N}}\) such that also \(\pi_{t_2}(\omega_m^2)\)
converges, etc. Then the diagonal sequence \((ω^m_m)_{m∈N}\) is such that \(π_t(ω^m_m)\) converges for all \(t ∈ \bigcup_{n∈N} J_n\). Let

\[ z_t := \lim_{m→∞} π_t(ω^m_m) \]

for \(t ∈ \bigcup_{n∈N} J_n\). Since \(π_{J_n}(ω^m_m) ∈ K_n\) for all but finitely many \(m\) and since \(K_n\) is compact (and therefore closed), we also have

\[ (z_t)_{t∈J_n} = \lim_{m→∞} π_{J_n}(ω^m_m) ∈ K_n. \]

Let now \(z\) an arbitrary element of \(E\), and define \(ω ∈ E^I\) by

\[ π_t(ω) := \begin{cases} z_t & \text{if } t ∈ \bigcup_{n∈N} J_n, \\ z & \text{otherwise}. \end{cases} \]

Then \(π_{J_n}(ω) ∈ K_n\) for all \(n ∈ N\), hence \(ω ∈ K_n \subseteq A_n\) for all \(n ∈ N\). Hence \(\bigcap_{n∈N} A_n\) is not empty, in contradiction to (14).

Hence we have shown that indeed

\[ \lim_{n→∞} P_I(A_n) = 0 \]

for every sequence \((A_n)_{n∈N}\) of elements of \(B_I\) satisfying \(A_n \searrow \emptyset\).

\[ \square \]

15 Markov kernels, semigroups and processes

**Definition 15.1.** Let \((E, B)\) be a measurable space. A map \(K : E × B → [0,1]\) is called a Markov kernel on \((E, B)\) if

1. For every \(B ∈ B\) the map \(x ↦ K(x, B)\) is measurable with respect to \(B\).
2. For every \(x ∈ E\) the map \(B ↦ K(x, B)\) is a probability measure on \((E, B)\).

**Remark 15.2.** The notion of a Markov kernel is a generalization of the notion of a stochastic matrix. Namely, let \(E\) be a countable set, and let \(K : E × P(E) → [0,1]\) be a Markov kernel on \((E, P(E))\). Then the matrix \(P ∈ \mathbb{R}^E×E\) defined by

\[ P(x, y) := K(x, \{y\}) \]

is stochastic, i.e.

1. \(P(x, y) ≥ 0\) for all \(x, y ∈ E\),
2. \(∑_{y∈E} P(x, y) = 1\) for all \(x ∈ E\).

Conversely, let \(P ∈ \mathbb{R}^E×E\) be a stochastic matrix on \(E\). Then the map \(K : E × P(E) → [0,1]\) defined by

\[ K(x, B) := ∑_{y∈B} P(x, y) \]

is a Markov kernel on \((E, P(E))\).

**Definition 15.3.** Let \((E, B)\) be a measurable space.

\[^{20}\text{Andrey Andreyevich Markov (1856–1922), Russian mathematician}\]
1. The identity kernel on \((E, B)\) is defined by
   \[ I(x, B) = 1_{\{x \in B\}}. \]

2. Let \(\mu\) be a probability measure and \(K\) a Markov kernel on \((E, B)\). Then the probability measure \(\mu K\) on \((E, B)\) is defined by
   \[ (\mu K)(B) := \int_E K(x, B)\mu(dx) \]

3. Let \(f : E \to \mathbb{R}\) be bounded and measurable and \(K\) a Markov kernel on \((E, B)\). Then the function \(Kf : E \to \mathbb{R}\) is defined by
   \[ (Kf)(x) := \int_E f(y)K(x, dy). \]

4. Let \(K_1\) and \(K_2\) be two Markov kernels on \((E, B)\). Their composition is defined by
   \[ (K_1 K_2)(x, B) := \int_E K_2(y, B)K_1(x, dy). \]

**Remark 15.4.** If \(E\) is countable, the identity kernel corresponds to the identity matrix, and the operations 2.–4. are given by matrix multiplications. To highlight this analogy, one sometimes writes
   \[ (\mu K)(B) = \int_E \mu(dx)K(x, B) \]

instead of \( (\mu K)(B) = \int_E K(x, B)\mu(dx), \) etc.

**Definition 15.5.** Let \((E, B)\) be a measurable space, \((P_t)_{t \in I}\) a Markov semigroup on \((E, B)\) and \((F_t)_{t \in I}\) a filtration on a probability space \((\Omega, \mathcal{A}, P)\). We say that an \((E, B)\)-valued stochastic process \((X_t)_{t \in I}\) on \((\Omega, \mathcal{A}, P)\) is a Markov process with transition semigroup \((P_t)_{t \in I}\) with respect to the filtration \((F_t)_{t \in I}\) if

1. \(P_0 = I,\) and
2. \(P_{s+t} = P_s P_t\) for all \(s, t \in I,\) i.e.
   \[ P_{s+t}(x, B) = \int_E P_t(y, B)P_s(x, dy) \]

   for all \(s, t \in I,\) all \(x \in E\) and all \(B \in B\) (Chapman-Kolmogorov\(^{21}\) equations).

**Interpretation.** We interprete \(P_t(x, B)\) as the probability that a particle initially at \(x\) will be in \(B\) at time \(t.\)

**Definition 15.6.** Let \((E, B)\) be a measurable space, \((P_t)_{t \in I}\) a Markov semigroup on \((E, B)\) and \((F_t)_{t \in I}\) a filtration on a probability space \((\Omega, \mathcal{A}, P)\). We say that an \((E, B)\)-valued stochastic process \((X_t)_{t \in I}\) on \((\Omega, \mathcal{A}, P)\) is a Markov process with transition semigroup \((P_t)_{t \in I}\) with respect to the filtration \((F_t)_{t \in I}\) if

1. it is adapted to the filtration \((F_t)_{t \in I},\) and

---

\(^{21}\)Sydney Chapman (1888–1970), English mathematician and physicist; Andrey Nikolaevich Kolmogorov (1903–1987), Russian mathematician
2. \[ P(X_t \in B \mid \mathcal{F}_s) = P_{t-s}(X_s, B) \] for all \( s \leq t \in I \) and all \( B \in \mathcal{B} \). Here the \textit{conditional probability} of an event \( A \) given a \( \sigma \)-field \( \mathcal{C} \) is defined as
\[ P(A \mid \mathcal{C}) := E(1_A \mid \mathcal{C}). \]
(Note that \( P(A \mid \mathcal{C}) \) is NOT a number, but a random variable, and that it is only determined almost surely.)

\textbf{Remark 15.7.} The second condition of the definition is equivalent to
\[ E(f(X_t) \mid \mathcal{F}_s) = (P_{t-s}f)(X_s) \] for all \( s \leq t \in I \) and all bounded measurable functions \( f : E \to \mathbb{R} \).

\textit{Proof.} (16) follows from (17) by choosing \( f = 1_B \). Conversely, (17) follows from (16) by measure-theoretic induction. \( \square \)

\textbf{Remark 15.8.} In this case we have
\[ E(f(X_t) \mid \mathcal{F}_s) = E(f(X_t) \mid \sigma(X_s)) \] for all \( s \leq t \in I \) and all bounded measurable functions \( f : E \to \mathbb{R} \) and in particular
\[ P(X_t \in B \mid \mathcal{F}_s) = P(X_t \in B \mid \sigma(X_s)) \] for all \( B \in \mathcal{B} \). (Markov property: The future \( (X_t) \) depends on the past \( (\mathcal{F}_s) \) only via the present \( (X_s) \)).

\textit{Proof.} We will show that \((P_{t-s}f)(X_s)\) has the defining properties of the conditional expectation \( E(f(X_t) \mid \sigma(X_s))\). It is clearly measurable with respect to \( \sigma(X_s) \), and for all \( A \in \sigma(X_s) \) we have
\[ \int_A (P_{t-s}f)(X_s) dP = \int_A E(f(X_t) \mid \mathcal{F}_s) dP = \int_A f(X_t) dP. \] \( \square \)

\textbf{Proposition 15.9.} Let \((E, \mathcal{B})\) be a measurable space, \((P_t)_{t \in I}\) a Markov semigroup, \( \mu \) a probability measure on \((E, \mathcal{B})\) and \((X_t)_{t \in I}\) an \((E, \mathcal{B})\)-valued Markov process with transition semigroup \((P_t)_{t \in I}\) (with respect to any filtration) and initial distribution \( \mu \). Then its finite-dimensional distributions are given as follows: For all \( n \in \mathbb{N} \), all \( t_1 < \cdots < t_n \in I \) and all \( B \in \mathcal{B}^n \) we have
\[ P(X_{t_1} \in B_1, \ldots, X_{t_n} \in B_n) = \int_{E} \cdots \int_{B_n} P_{t_n-t_{n-1}}(x_{n-1}, dx_n) \cdots P_{t_1}(x_0, dx_1) \mu(dx_0). \] Conversely, if the finite-dimensional distributions of a stochastic process \((X_t)_{t \in I}\) are given by (18), it is a Markov process with transition semigroup \((P_t)_{t \in I}\) with respect to its canonical filtration and with initial distribution \( \mu \).
Proof. First note that an equivalent formulation of (18) is that for all bounded measurable functions \( f_1, \ldots, f_n : E \to \mathbb{R} \) we have

\[
E \left( \prod_{i=1}^{n} f_i(X_{t_i}) \right) = \int_E \cdots \int_{E} \prod_{i=1}^{n} f_i(x_i) P_{t_n-\ldots-t_1}(x_{n-1}, dx_n) \ldots P_{t_1}(x_0, dx_1) \mu(dx_0). \tag{19}
\]

((18) follows from (19) by choosing \( f_i = 1_{B_i} \). Conversely, (19) follows from (18) by measure-theoretic induction.)

Suppose now that \((X_t)_{t \in I}\) an \((E, B)\)-valued Markov process with transition semigroup \((P_t)_{t \in I}\) and initial distribution \(\mu\). To prove (19), we proceed by induction, starting with the trivial case \(n = 0\). Assuming that (19) is true for the value \(n\), we obtain

\[
E \left( \prod_{i=1}^{n+1} f_i(X_{t_i}) \right) = E \left( E \left( \prod_{i=1}^{n+1} f_i(X_{t_i}) \mid \mathcal{F}_{t_N} \right) \right)
= E \left( \prod_{i=1}^{n} f_i(X_{t_i}) E \left( f_{n+1}(X_{t_{n+1}}) \mid \mathcal{F}_{t_n} \right) \right)
= E \left( \prod_{i=1}^{n} f_i(X_{t_i}) (P_{t_{n+1}-t_n} f_{n+1})(X_{t_n}) \right)
= \int_E \cdots \int_{E} \prod_{i=1}^{n} f_i(x_i) (P_{t_{n+1}-t_n} f_{n+1})(x_n)
\]

\[
= \int_E \cdots \int_{E} \prod_{i=1}^{n} f_i(x_i) \int_E f(x_{n+1}) P_{t_{n+1}-t_n}(x_n, dx_{n+1})
\]

\[
= \int_E \cdots \int_{E} \prod_{i=1}^{n+1} f_i(x_i)
\]

\[
P_{t_{n+1}-t_n}(x_n, dx_{n+1}) P_{t_{n}-t_{n-1}}(x_{n-1}, dx_n) \ldots P_{t_1}(x_0, dx_1) \mu(dx_0),
\]

i.e. (19) with \(n\) replaced with \(n+1\).

Now suppose that the finite-dimensional distributions of \((X_t)_{t \in I}\) are given by (19). Fix \(n \in \mathbb{N}, 0 \leq s_1 < \ldots < s_n \leq t\) and a bounded measurable function \(f : E \to \mathbb{R}\). To show that

\[
E(f(X_t) \mid \mathcal{F}_s) = (P_{t-s}f)(X_s),
\]

we have to show that

\[
\int_A f(X_t) dP = \int_A (P_{t-s}f)(X_s) dP \tag{20}
\]

for all \(A \in \mathcal{F}_s\). Suppose first that \(A\) is of the form

\[
A = \{X_{s_1} \in B_1, \ldots, X_{s_n} \in B_n\}
\]

46
with \(0 \leq s_1 < \ldots < s_n = s\). Letting \(s_{n+1} := t\), \(f_i := 1_{B_i}\) for \(i \in \{1, \ldots, n\}\) and \(f_{n+1} := f\), we have on the one hand

\[
\int_A f(X_t) dP = E \left( \prod_{i=1}^n 1_{B_i}(X_{s_i}) f(X_t) \right)
\]

\[
= \int_E \int_E \ldots \int_E \prod_{i=1}^n 1_{B_i}(x_i) f(x_{n+1}) P_{s_{n+1} - s_n}(x_n, dx_{n+1}) \ldots P_{s_1}(x_0, dx_1) \mu(dx_0),
\]

and on the other hand

\[
\int_A (P_t - s)f(X_s) dP
\]

\[
= E \left( \prod_{i=1}^n 1_{B_i}(X_{s_i})(P_t - s f)(X_s) \right)
\]

\[
= \int_E \int_E \ldots \int_E \prod_{i=1}^n 1_{B_i}(x_i)(P_{s_{n+1} - s_n} f)(x_n) P_{s_{n-1} - s_n - 1}(x_{n-1}, dx_n) \ldots P_{s_1}(x_0, dx_1) \mu(dx_0)
\]

\[
= \int_E \int_E \ldots \int_E \prod_{i=1}^n 1_{B_i}(x_i) \int_E f(y) P_{s_{n+1} - s_n}(x_n, dy) P_{s_{n-1} - s_n - 1}(x_{n-1}, dx_n) \ldots P_{s_1}(x_0, dx_1) \mu(dx_0),
\]

and this coincides with (21).

We have hence shown that (20) holds for all \(A\) belonging to

\[
S := \bigcup_{n \in \mathbb{N}} \bigcup_{0 \leq s_1 < \ldots < s_n = s} \sigma(X_{s_1}, \ldots, X_{s_n}).
\]

Since in particular \(\Omega \in S\) it is clear that set \(D\) of events for which (20) holds is a Dynkin system, i.e.

1. \(\emptyset \in D\),
2. \(A \in D \Rightarrow A^c \in D\),
3. \(\bigcup_{i \in I} A_i \in D\) for every countable union of pairwise disjoint elements of elements of \(D\).

Since moreover \(S\) is stable under intersection, we have

\[
D = \sigma(S) = \mathcal{F}_s,
\]

i.e. (20) holds for all \(A \in \mathcal{F}_s\).

\(\square\)

**Theorem 15.10.** Let \((E, B)\) be a measurable space, and let \(\mu\) be a probability measure and \((P_t)_{t \in I}\) a Markov semigroup on it \((I = \mathbb{N}_0\ or\ I = \mathbb{R}_+\). For any non-empty finite subset \(J = \{t_1 < \ldots < t_n\}\) of \(I\) and any \(B \in B^J\) we set

\[
P_J(B) := \int_E \int_E \ldots \int_E 1_{\{(x_1, \ldots, x_n) \in B\}} P_{t_n - t_{n-1}}(x_{n-1}, dx_n) \ldots P_{t_1}(x_0, dx_1) \mu(dx_0). \quad (22)
\]

Then
1. each \( P_J \) is a probability measure, and
2. the family \( (P_J)_{J \in I^*} \) is consistent.

**Corollary 15.11.** If \( E \) is Polish, then for any Markov semigroup \((P_t)_{t \in I}\) and any probability measure \( \mu \) on \( E \) there exists an \( E \)-valued Markov process \((X_t)_{t \in I}\) with transition semigroup \((P_t)_{t \in I}\) and initial distribution \( \mu \).

**Proof of Theorem 15.10.** The first statement is clear. To prove that the family \( (P_J)_{J \in I^*} \) is consistent, one has to show that for all \( J, J' \in I^* \) with \( J \subseteq J' \) the image measure of \( P_{J'} \) under the projection \( \pi_{J'}^J \) equals \( P_J \). Clearly it suffices to do so in the case when \( I \setminus J \) contains exactly one element. In order to keep notation simple, we restrict ourselves to the case when \( J' \) contains three elements, \( J' = \{t_1, t_2, t_3\} \) with \( t_1 < t_2 < t_3 \), and \( J = \{t_1, t_3\} \). Then we have to show that for all \( B_1, B_3 \in \mathcal{B} \) we have

\[
P_{J'}(B_1 \times E \times B_3) = P_J(B_1 \times B_3): \]

\[
P_{J'}(B_1 \times E \times B_3) = \int_E \int_{B_1} \int_{B_3} P_{t_3-t_2}(x_2, dx_3) P_{t_2-t_1}(x_1, dx_2) P_{t_1}(x_0, dx_1) \mu(dx_0) \]

The semigroup property implies that the underbraced integral equals \( P_{t_3-t_1}(x_1, B_3) \), so that we get

\[
P_J(B_1 \times E \times B_3) = \int_E \int_{B_1} P_{t_3-t_1}(x_1, B) P_{t_1}(x_0, dx_1) \mu(dx_0) \]

\[
= \int_E \int_{B_1} \int_{B_3} P_{t_3-t_1}(x_1, dx_3) P_{t_1}(x_0, dx_1) \mu(dx_0) \]

\[
= P_J(B_1 \times B_3). \quad \square \]

16 Brownian motion and Kolmogorov’s continuity theorem

16.1 Brownian motion

**Definition 16.1 (Heat kernel).** For \( t > 0 \) and \( x, y \in \mathbb{R}^d \) let

\[
p_t(x, y) := (2\pi t)^{-d/2} e^{-\frac{|x-y|^2}{2t}} \]

be the \( d \)-dimensional heat kernel.

**Remark 16.2.** The heat kernel is of fundamental importance in both PDE and probability theory:

1. It is the fundamental solution of the heat equation

\[
\frac{\partial p}{\partial t} = \frac{1}{2} \Delta p, \]

i.e. for each \( x \in \mathbb{R}^d \)

(a) the function \((t, y) \mapsto p_t(x, y)\) satisfies the heat equation in the domain \( \mathbb{R}_+^* \times \mathbb{R}^d \), and
(b) as \( t \to 0 \), the measure with density \( y \mapsto p_t(x, y) \) with respect to Lebesgue measure converges weakly to the Dirac measure at \( x \).

2. For each \( x \in \mathbb{R}^d \) the function \( (t, y) \mapsto p_t(x, y) \) is the density of the \( d \)-dimensional normal distribution with mean \( x \) and covariance matrix \( tI \), i.e. of the joint distribution of \( d \) independent normally distributed random variables with means \( x_i (i = 1, \ldots, d) \) and variance \( t \). To see this, note that

\[
p_t(x, y) = \prod_{i=1}^{d} \left( \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x_i - y_i)^2}{2t}} \right).
\]

**Definition 16.3.** The Brownian semigroup \((P_t)_{t \in \mathbb{R}^+}\) on \( \mathbb{R}^d \) is defined as follows: \( P_0 = I \), and

\[
P_t(x, B) := \int_B p_t(x, y)dy
\]

for \( t > 0 \) and \( B \in \mathcal{B}(\mathbb{R}^d) \).

**Proposition 16.4.** The family \((P_t)_{t \in \mathbb{R}^+}\) defined above is indeed a Markov semigroup on \( \mathbb{R}^d \).

**Proof.** We first show that

\[
\frac{|x - y|^2}{2s} + \frac{|y - z|^2}{2t} = \frac{|x - z|^2}{2(s + t)} + \frac{s + t}{2st} \left| x - y - \frac{s(x - z)}{s + t} \right|^2
\]

for \( x, y, z \in \mathbb{R}^d \) and \( s, t > 0 \):

\[
\begin{align*}
\frac{|x - z|^2}{2(s + t)} + \frac{s + t}{2st} \left| x - y - \frac{s(x - z)}{s + t} \right|^2 &= \frac{|x - z|^2}{2(s + t)} + \frac{s + t}{2st} \left( \frac{|x - y|^2}{s + t} - \frac{2s(x - y) \cdot (x - z)}{s + t} + \frac{s^2|x - z|^2}{(s + t)^2} \right) \\
&= \frac{|x - z|^2}{2(s + t)} + \frac{s + t}{2st} \frac{(s + t)|x - y|^2}{t} - \frac{(x - y) \cdot (x - z)}{t} + \frac{s|x - z|^2}{2t(s + t)} \\
&= \frac{|x - z|^2}{2t} + \frac{(s + t)|x - y|^2 - 2s(x - y) \cdot (x - z)}{2st} \\
&= \frac{s|x - z|^2 + (s + t)|x - y|^2 - 2s(x - y) \cdot (x - z)}{2st} \\
&= \frac{s(|x - z|^2 - 2(x - y) \cdot (x - z) + |x - y|^2) + t|x - y|^2}{2st} \\
&= \frac{s|y - z|^2 + t|x - y|^2}{2st} \\
&= \frac{|x - y|^2}{2s} + \frac{|y - z|^2}{2t}.
\end{align*}
\]

49
We then obtain (for \( s, t > 0 \))
\[
(P_s P_t)(x, B) = \int_{\mathbb{R}} P_t(y, B) P_s(x, dy)
\]
\[
= \int_{\mathbb{R}} \left( (2\pi t)^{-d/2} \int_{\mathbb{R}^d} e^{-\frac{|y-z|^2}{2t}} \, dz \right) (2\pi s)^{-d/2} e^{-\frac{|x-y|^2}{2s}} \, dy
\]
\[
= (4\pi^2 s t)^{-d/2} \int_{\mathbb{R}^d} \exp \left( -\frac{|y-z|^2}{2t} - \frac{|x-y|^2}{2s} \right) \, dy \, dz
\]
\[
= (4\pi^2 s t)^{-d/2} \int_{\mathbb{R}^d} \exp \left( -\frac{|x-z|^2}{2(s+t)} \right) \int_{\mathbb{R}^d} \exp \left( -\frac{s+t}{2st} |y|^2 \right) \, dy \, dz.
\]
Since moreover
\[
\int_{\mathbb{R}^d} \exp \left( -\frac{s+t}{2st} |y|^2 \right) \, dy = \left( \int_{\mathbb{R}} \exp \left( -\frac{s+t}{2st} y^2 \right) \, dy \right)^d
\]
\[
= \frac{2\pi st}{s+t}
\]
it follows that
\[
(P_s P_t)(x, B) = (2\pi (s+t))^{-d/2} \int_{\mathbb{R}^d} \exp \left( -\frac{(x-z)^2}{2(s+t)} \right) \, dz
\]
\[
= P_{s+t}(x, B).
\]

**Proposition 16.5.** Let \((P_t)_{t \in \mathbb{R}_+}\) be the Brownian semigroup on \(\mathbb{R}^d\). For an \(\mathbb{R}^d\)-valued stochastic process \((X_t)_{t \in \mathbb{R}_+}\) the following properties are equivalent:

1. It has independent increments (i.e. for all \( n \in \mathbb{N} \) and all \( 0 < t_1 < \ldots < t_n \) the random variables \( X_0, X_{t_1} - X_0, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}} \) are independent), and for all \( 0 \leq s < t \) the increment \( X_t - X_s \) is normally distributed with mean 0 and covariance matrix \((t-s)I_d\), i.e. the distribution of \( X_t - X_s \) has density
\[
p_{t-s}(0, x) = (2\pi(t-s))^{-d/2} e^{-\frac{|x|^2}{t-s}}
\]
with respect to Lebesgue measure.

2. It is a Markov process with transition semigroup \((P_t)_{t \in \mathbb{R}_+}\) (with respect to its canonical filtration).

**Proof.** Let \( \mu_0 \) be the distribution of \( X_0 \). Then both properties are equivalent to the following statement on the finite-dimensional distributions of the process: For all \( 0 \leq t_1 < \ldots < t_n \) and all \( B_1, \ldots, B_n \in B(\mathbb{R}^d) \) we have
\[
P(X_{t_1} \in B_1, \ldots, X_{t_n} \in B_n) = \int_{E} \int_{B_1} \ldots \int_{B_n} P_{t_n-t_{n-1}}(x_{n-1}, dx_{n}) \ldots P_{t_1}(x_0, dx_1) \mu(dx_0).
\]

**Definition 16.6.** An \(\mathbb{R}^d\)-valued stochastic process \((X_t)_{t \in \mathbb{R}_+}\) on a probability space \((\Omega, A, \mathbb{P})\) is called a \(d\)-dimensional Brownian motion if
1. it has the equivalent properties of the last proposition, and
2. almost all its paths are continuous, i.e. for almost all \( \omega \in \Omega \) the map \( t \mapsto X_t(\omega) \) is continuous (from \( \mathbb{R}_+ \) to \( \mathbb{R}^d \)).

If moreover \( X_0 = 0 \) a.s., then it is called a standard Brownian motion.

**Remark 16.7.** The existence theorem for Markov processes assures the existence of a stochastic process with the first property. The existence of a process which also has the second property will follow from Kolmogorov’s continuity theorem.

### 16.2 Modifications and indistinguishable processes

**Definition 16.8.** Let \((E, B)\) be a measurable space and \(I\) a set.

1. Two \((E, B)\)-valued stochastic processes \((X_t)_{t \in I}\) and \((X'_t)_{t \in I}\) defined on the same probability space \((\Omega, A, P)\) are said to be **modifications** of each other if for each \(t \in I\)
   \[ X_t = X'_t \quad \text{a.s.} \]

2. They are called **indistinguishable** if almost surely
   \[ X_t = X'_t \quad \text{for all } t \in I. \]

**Remark 16.9.**

1. If two stochastic processes are indistinguishable, they are modifications of each other.
2. If two stochastic processes are modifications of each other, they have the same finite-dimensional distributions.

“Indistinguishable \(\Rightarrow\) Modifications of each other \(\Rightarrow\) Same finite-dimensional distributions”

The converse implications are not true in general.

**Example 16.10.** Let \((\Omega, A, P) = [0, 1]\) equipped with its Borel \(\sigma\)-field and Lebesgue measure, and let
\[ X_t(\omega) := \begin{cases} 1 & \text{if } t = \omega \\ 0 & \text{otherwise.} \end{cases} \]

Then for each \(t \in [0, 1]\) we have \(X_t = 0\) a.s., so that the process \((X_t)_{t \in [0,1]}\) is a modification of the process which is identically 0. On the other hand, for there is not a single \(\omega \in \Omega\) for which the path \(t \mapsto X_t(\omega)\) is identically 0. Consequently, the process \((X_t)_{t \in [0,1]}\) is not indistinguishable from the process which is identically 0.

**Proposition 16.11.** Let \(E\) be a Hausdorff\(^{22}\) topological space, and let \((X_t)_{t \in \mathbb{R}_+}\) and \((X'_t)_{t \in \mathbb{R}_+}\) be two continuous \(E\)-valued stochastic processes on the same probability space \((\Omega, A, P)\) (i.e. almost all paths of both processes are continuous). These two processes are indistinguishable if and only if they are modifications of each other.

**Proof.** We already know that indistinguishable processes are modifications of each other, so let us assume that \((X_t)_{t \in \mathbb{R}_+}\) and \((X'_t)_{t \in \mathbb{R}_+}\) are modifications of each other. Let \(\Omega^*\) be an event of probability 1 such that on \(\Omega^*\) all paths of both processes are continuous, and for each \(t \in \mathbb{R}_+\) let \(\Omega_t := \{X_t = X'_t\}\). Then \(\Omega^* \cap \bigcap_{t \in \mathbb{Q}_+} \Omega_t\) has probability 1. Moreover,

\(^{22}\)Felix Hausdorff (1868–1942), German mathematician
choosing for each \( t \in \mathbb{R}_+ \) a sequence \( (t_n)_{n \in \mathbb{N}} \) in \( \mathbb{Q}_+ \) converging to \( t \) we obtain on that event:

\[
X'_t = \lim_{n \to \infty} X'_{t_n} = \lim_{n \to \infty} X_{t_n} = X_t,
\]

where in the second equality we used the fact that in Hausdorff spaces limits are unique. \( \square \)

16.3 Kolmogorov’s continuity theorem

**Lemma 16.12** (Borel-Cantelli lemma\(^{23}\)). Let \( (A_n)_{n \in \mathbb{N}} \) be a sequence of events such that the series \( \sum_{n=1}^{\infty} P(A_n) \) converges. Then

\[
P \left( \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} A_m \right) = 0.
\]

**Proof.** Convergence of the series \( \sum_{n=1}^{\infty} P(A_n) \) is equivalent to the property that \( \lim_{n \to \infty} \sum_{m \geq n} P(A_m) = 0 \). Consequently

\[
P \left( \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} A_m \right) = \lim_{n \to \infty} P \left( \bigcup_{m \geq n} A_m \right) \leq \liminf_{n \to \infty} \sum_{m \geq n} P(A_m) = 0. \quad \square
\]

**Remark 16.13.** We have \( \omega \in \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} A_m \) if and only if \( \omega \in A_m \) for infinitely many \( m \). Consequently, the Borel-Cantelli lemma can be reformulates as follows: If the series \( \sum_{n=1}^{\infty} P(A_n) \) converges, the probability that infinitely many of the events \( A_m \) occur is 0.

**Theorem 16.14** (Kolmogorov’s continuity theorem). Let \( (E,d) \) be a complete metric space (e.g. \( E = \mathbb{R}^d \)), and let \( (X_t)_{t \in \mathbb{R}_+} \) be an \( E \)-valued stochastic process with the following property: There exist constants \( \alpha, \beta, C > 0 \) such that

\[
E \left( d(X_s,X_t)^\alpha \right) \leq C |t - s|^{1 + \beta}
\]

for all \( s,t \in \mathbb{R}_+ \). Then this process has a continuous modification \( (X'_t)_{t \in \mathbb{R}_+} \). Moreover, this (and consequently any) continuous modification has the following property: Almost every path is locally Hölder continuous of any exponent \( \gamma \in (0, \beta/\alpha) \), i.e. there exists an event \( \Omega' \) of probability 1 such that for each \( \omega \in \Omega \) and each \( k \in \mathbb{N} \) there exists a constant \( C(k,\gamma,\omega) \) such that

\[
d(X'_s(\omega),X'_t(\omega)) \leq C(k,\gamma,\omega)|t - s|^{\gamma}.
\]

for all \( s,t \in [0,k] \).

**Proof.** Fix \( \gamma \in (0, \beta/\alpha) \). For \( m,k \in \mathbb{N} \) let \( S_{m,k} := \{ i \cdot 2^{-m} | i \in \{ 0, \ldots, k \cdot 2^m \} \} \subset [0,k] \),

\( D_k := \bigcup_{m \in \mathbb{N}} S_{m,k} \) and \( D := \bigcup_{k \in \mathbb{N}} D_k \). \( D_k \) is the set of dyadic numbers belonging to \([0,k]\), and \( D \) is the set of all dyadic numbers. The crucial step of the proof is the following lemma:

Lemma 16.15. Let
\[ K(\gamma) := \frac{2}{1 - 2^{-\gamma}} + 2^\gamma. \]
Then there exist an event \( \Omega^*(\gamma) \) of probability 1 and for each \( k \in \mathbb{N} \) and each \( \omega \in \Omega^*(\gamma) \) a number \( m(k, \gamma, \omega) \in \mathbb{N} \) such that
\[ d(X_s(\omega), X_t(\omega)) \leq K(\gamma)|t - s|\gamma \]
for all \( \omega \in \Omega^*(\gamma) \) and all \( s, t \in D \cap [0, k] \) satisfying \( |t - s| \leq 2^{-m(k, \gamma, \omega)} \).

Proof. Fix \( k \in \mathbb{N} \) and let \( \delta := \beta - \alpha \gamma > 0 \). For each \( m \in \mathbb{N} \) let
\[ A_m := \bigcup_{t \in S_{m,k} \setminus \{0\}} \{ d(X_t, X_{t-2^{-m}}) \geq 2^{-\gamma m} \}. \]
For all \( t \in S_{m,k} \setminus \{0\} \) we have using Chebyshev’s inequality and the fact that \( \alpha \gamma < \beta \)
\[ P( d(X_t, X_{t-2^{-m}}) \geq 2^{-\gamma m} ) \leq 2^{\alpha \gamma m} E(d(X_t, X_{t-2^{-m}})\gamma) \leq 2^{\alpha \gamma m} C \cdot 2^{-m(1+\beta)} \leq C \cdot 2^{-(\delta+1)m}, \]
so that
\[ P(A_m) \leq k \cdot 2^m C \cdot 2^{-(\delta+1)m} = kC \cdot 2^{-\delta m}. \]
Since \( \delta > 0 \) we have \( 2^{-\delta} < 1 \), so that the series \( \sum_{m=1}^{\infty} P(A_m) \) converges. Let
\[ \Omega_k := \left( \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} A_m \right)^c. \]
By the Borel-Cantelli lemma \( P(\Omega_k) = 1 \). Moreover, for each \( \omega \in \Omega_k \) there exists \( m(k, \gamma, \omega) \) such that
\[ d(X_t(\omega), X_{t-2^{-m}}(\omega)) \leq 2^{-\gamma m} \]
for all \( m \geq m(k, \gamma, \omega) \) and all \( t \in S_{m,k} \setminus \{0\} \).
Fix now \( \omega \in \Omega_k \), and let \( s, t \in D_k \) with \( 0 < |s - t| \leq 2^{-m(k, \gamma, \omega)} \). Choose \( m \in \mathbb{N} \) in such a way that
\[ 2^{-(m+1)} < |s - t| \leq 2^{-m} \quad (23) \]
(clearly this implies that \( m \geq m(k, \gamma, \omega) \)). Write \( s \) and \( t \) in the form
\[ s = A \cdot 2^{-m} + \sum_{i=m+1}^{\infty} \sigma_i 2^{-i}, \quad t = B \cdot 2^{-m} + \sum_{j=m+1}^{\infty} \tau_j 2^{-j} \]
with \( A, B \in \{0, \ldots, k \cdot 2^m\} \) and \( \sigma_i, \tau_j \in \{0, 1\} \) and such that only finitely many of the numbers \( \sigma_i \) and \( \tau_j \) are non-zero. Moreover, let
\[ s_n := A \cdot 2^{-m} + \sum_{i=m+1}^{m+n} \sigma_i 2^{-i}, \quad t_n := B \cdot 2^{-m} + \sum_{j=m+1}^{m+n} \tau_j 2^{-j} \]
(so that \( s_n = s \) and \( t_n = t \) for all sufficiently large \( n \)). Then (23) implies that either \( s_0 = t_0 \) or \( |s_0 - t_0| = 2^{-m} \), and consequently
\[ d(X_{s_0}(\omega), X_{t_0}(\omega)) \leq 2^{-m}. \]
Moreover,
\[
d(X_{s_0}(\omega), X_s(\omega)) \leq \sum_{n=1}^{\infty} d(X_{s_n}(\omega), X_{s_{n-1}}(\omega))
\]
\[
\leq \sum_{n=1}^{\infty} 2^{-\gamma(m+n)}
\]
\[
= 2^{-\gamma m} \sum_{n=1}^{\infty} 2^{-\gamma n}
\]
\[
= 2^{-\gamma m} \frac{2^{-\gamma}}{1 - 2^{-\gamma}}
\]
\[
= \frac{2^{-\gamma(m+1)}}{1 - 2^{-\gamma}}
\]

and, by the same argument also
\[
d(X_{t_0}(\omega), X_t(\omega)) \leq \frac{2^{-\gamma(m+1)}}{1 - 2^{-\gamma}}.
\]

Combining these three estimates we obtain
\[
d(X_s(\omega), X_t(\omega)) \leq d(X_{s_0}(\omega), X_{s_0}(\omega)) + d(X_{t_0}(\omega), X_{t_0}(\omega)) + d(X_{t_0}(\omega), X_{t_0}(\omega))
\]
\[
\leq \frac{2^{-\gamma(m+1)}}{1 - 2^{-\gamma}} + 2^{-\gamma m}
\]
\[
= K(\gamma)(2^{-(m+1)}\gamma)
\]
\[
\leq K(\gamma)|s - t|^\gamma.
\]

We conclude the proof of the lemma by choosing \(\Omega^*(\gamma) := \bigcap_{k \in \mathbb{N}} \Omega_k\).

**Lemma 16.16.** Let \(t \in \mathbb{R}_+\) and \((t_n)_{n \in \mathbb{N}}\) a sequence in \(D\) converging to \(t\). Then for all \(\omega \in \Omega^*(\gamma)\) the limit
\[
X'_t(\omega) := \lim_{n \to \infty} X_{t_n}(\omega)
\]
exists and does not depend on the choice of the sequence \((t_n)_{n \in \mathbb{N}}\).

**Proof.** Fix \(\omega \in \Omega^*(\gamma)\). We first show that the sequence \((X_{t_n}(\omega))_{n \in \mathbb{N}}\) is a Cauchy sequence: To do so choose \(k \in \mathbb{N}\) and \(n_0 \in \mathbb{N}\) such that \(t_n \leq k\) and \(|t_n - t_{n'}| \leq 2^{-m(k,\gamma,\omega)}\) for all \(n, n' \geq n_0\). For all such \(n, n'\) we then obtain
\[
d(X_{t_n}(\omega), X_{t_{n'}}(\omega)) \leq K|t_n - t_{n'}|^\gamma \to 0
\]
as \(n, n' \to \infty\). Since the metric space \(E\) is complete (by assumption), this implies the existence of the limit \(\lim_{n \to \infty} X_{t_n}(\omega)\).

To show that the limit does not depend on the choice of the sequence \((t_n)_{n \in \mathbb{N}}\), let \((s_n)_{n \in \mathbb{N}}\) be another sequence in \(D\) also converging to \(t\). By the same argument as above we obtain for all sufficiently large \(n\)
\[
d(X_{s_n}(\omega), X_{t_n}(\omega)) \leq K|s_n - t_n|^\gamma \to 0
\]
as \(n \to \infty\) and therefore
\[
\lim_{n \to \infty} X_{s_n}(\omega) = \lim_{n \to \infty} X_{t_n}(\omega).
\]
Let now \[ \Omega^*: = \bigcap_{\gamma \in Q \cap (0, \beta/\alpha)} \Omega^*(\gamma). \]

To conclude the proof of the theorem we have to show that

1. For each \( \omega \in \Omega^* \) the map \( t \mapsto X'_t(\omega) \) is locally Hölder continuous of any order \( \gamma \in (0, \beta/\alpha) \), i.e. for each \( \omega \in \Omega^* \), each \( \gamma \in (0, \beta/\alpha) \) and each \( k \in \mathbb{N} \) there exists a constant \( C(k, \gamma, \omega) \) such that
\[
d(X'_s(\omega), X'_t(\omega)) \leq C(k, \gamma, \omega)|t - s|^\gamma
\]
for all \( s, t \in [0, k] \).

2. For each \( t \in \mathbb{R}_+ \) we have
\[
X'_t = X_t \quad \text{a.s.}
\]

We start with the first property. Fix \( k \in \mathbb{N} \) and \( \gamma \in (0, \beta/\alpha) \). We know that on \( \Omega^* \) we have
\[
d(X'_s(\omega), X'_t(\omega)) \leq K(\gamma)|t - s|^\gamma
\]
for all \( \gamma \in Q \cap (0, \beta/\alpha) \) and all \( s, t \in D \cap [0, k] \) satisfying \( |t - s| \leq 2^{-m(k, \gamma, \omega)} \). By continuity this extends to all \( \gamma \in (0, \beta/\alpha) \) and all \( s, t \in [0, k] \) satisfying \( |t - s| \leq 2^{-m(k, \gamma, \omega)} \). Consequently, for all \( s, t \in [0, k] \) (not necessarily satisfying \( |t - s| \leq 2^{-m(k, \gamma, \omega)} \)) we obtain
\[
d(X'_s(\omega), X'_t(\omega)) \leq 2^{m(k, \gamma, \omega)}K(\gamma)|t - s|^\gamma.
\]

To prove the second property let \( (t_n)_{n \in \mathbb{N}} \) be a sequence in \( D \) converging to \( t \). Then we have
\[
X_{t_n} \to X'_t
\]
almost surely and hence in probability, and on the other hand, by Chebyshev’s inequality
\[
X_{t_n} \to X_t
\]
in probability. Consequently \( X'_t = X_t \) almost surely.

We now want to apply this result to show existence of Brownian motion.

**Lemma 16.17.** Let \( (X_t)_{t \in \mathbb{R}_+} \) be an \( \mathbb{R}^d \)-valued stochastic process whose increments are normally distributed with mean 0 and covariance matrix \( (t - s)I_d \). Then for each \( \alpha > 0 \) there is a constant \( C_\alpha > 0 \) such that
\[
E(|X_t - X_s|^\alpha) = C_\alpha(t - s)^{\alpha/2}
\]

**Proof.** Let
\[
Z_{s,t} := \frac{X_t - X_s}{\sqrt{t - s}}.
\]
\( Z_{s,t} \) is normally distributed with mean 0 and covariance matrix \( I_d \), so that \( C_\alpha := E(|Z_{s,t}|^\alpha) \) does not depend on \( s \) and \( t \). It follows that
\[
E(|X_t - X_s|^\alpha) = (t - s)^{\alpha/2}E(|Z_{s,t}|^\alpha) = C_\alpha(t - s)^{\alpha/2}.
\]

**Theorem 16.18.** For any probability measure \( \mu \) on \( \mathbb{R}^d \) there exists a Brownian motion with initial distribution \( \mu \). Moreover almost all its paths are locally Hölder continuous of any order \( \gamma < 1/2 \).

**Proof.** For any \( \alpha > 0 \) let \( \beta := \alpha/2 - 1 \). Then \( \beta \) is positive as soon as \( \alpha > 2 \). Moreover \( \beta/\alpha \to 1/2 \) as \( \alpha \to \infty \). The result therefore follows from Kolmogorov’s continuity theorem.
16.4 Brownian motion is a martingale

Lemma 16.19. A real-valued stochastic process \((X_t)_{t \in \mathbb{R}_+}\) has independent increments if and only if for all \(s \leq t \in \mathbb{R}_+\) the increment \(X_t - X_s\) is independent of \(\mathcal{F}_s^X\). (Recall that \((\mathcal{F}_t^X)_{t \geq 0}\) denotes the canonical filtration of the process \((X_t)_{t \in \mathbb{R}_+}\).)

Proof. Suppose first that the process has independent increments, and let \(s \leq t\). Let \(\mathcal{D}\) be the set of events which are independent of \(X_t - X_s\). On the one hand, \(\mathcal{D}\) is clearly a Dynkin system. On the other hand, independence of increments implies that for all \(0 \leq s_0 < \ldots < s_n \leq s\) the increment \(X_t - X_s\) is independent of \(\sigma(X_{s_0}, X_{s_1} - X_{s_0}, \ldots, X_{s_n} - X_{s_{n-1}})\). Since this \(\sigma\)-field coincides with \(\sigma(X_{s_0}, \ldots, X_{s_n})\), it follows that \(X_t - X_s\) is independent of \(S := \bigcup_{0 \leq s_0 < \ldots < s_n \leq s} \sigma(B_{s_0}, \ldots, B_{s_n})\). Since \(S\) is stable under intersection and \(\mathcal{D}\) is a Dynkin system, it follows that \(\mathcal{D} \supseteq \sigma(S) = \mathcal{F}_s\), i.e. that \(X_t - X_s\) is independent of \(\mathcal{F}_s\).

Suppose now that for all \(s \leq t \in \mathbb{R}_+\) the increment \(X_t - X_s\) is independent of \(\mathcal{F}_s\). We will show by induction that for all \(n \in \mathbb{N}_0\) and all \(0 = t_0 < t_1 < \ldots < t_n\) the increments \(X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}\) are independent. For \(n = 0\) there is nothing to prove. So let \(0 = t_0 < t_1 < \ldots < t_{n+1}\) and suppose that \(X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}\) are independent. Since by assumption \(X_{t_{n+1}} - X_{t_n}\) is independent of \(\mathcal{F}_s\), it follows that he family \(X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}, X_{t_{n+1}} - X_{t_n}\) is independent as well. \(\square\)

Corollary 16.20. An \(\mathbb{R}^d\)-valued stochastic process \((B_t)_{t \in \mathbb{R}_+}\) is a Brownian motion if and only if

1. for all \(s \leq t \in \mathbb{R}_+\) the increment \(B_t - B_s\) is
   (a) normally distributed with mean 0 and covariance matrix \(tI\), and
   (b) independent of \(\mathcal{F}_s^B\), and

2. almost all its paths are continuous, i.e. for almost all \(\omega \in \Omega\) the map \(t \mapsto X_t(\omega)\) is continuous.

This result motivates the following definition:

Definition 16.21. Let \((\mathcal{F}_t)_{t \in \mathbb{R}_+}\) be a filtration. An \(\mathbb{R}^d\)-valued stochastic process \((B_t)_{t \in \mathbb{R}_+}\) is called a Brownian motion with respect to the filtration \((\mathcal{F}_t)_{t \in \mathbb{R}_+}\) if and only if it is adapted to \((\mathcal{F}_t)_{t \in \mathbb{R}_+}\), and has the properties of the above corollary, but with the given filtration \((\mathcal{F}_t)_{t \in \mathbb{R}_+}\) instead of the canonical filtration \((\mathcal{F}_t^B)_{t \in \mathbb{R}_+}\).

Remark 16.22. Let \((B_t)_{t \in \mathbb{R}_+}\) be a one-dimensional Brownian motion with respect to a filtration \((\mathcal{F}_t)_{t \in \mathbb{R}_+}\). If \(B_0\) is integrable, then \((B_t)_{t \in \mathbb{R}_+}\) is a martingale with respect to \((\mathcal{F}_t)_{t \in \mathbb{R}_+}\).

Proof. Since the increment \(B_t - B_0\) is normally distributed, integrability of \(B_0\) implies integrability of \(B_t\). Moreover, since \(B_t - B_s\) is independent of \(\mathcal{F}_s\) and normally distributed with mean 0, it follows that

\[
E(B_t - B_s | \mathcal{F}_s) = E(B_t - B_s) = 0.
\]

\(\square\)

Remark 16.23. In a similar way one can also show that the processes \((B_t^2 - t)_{t \in \mathbb{R}_+}\) and \((e^{\alpha B_t - a^2 t/2})_{t \in \mathbb{R}_+}\) are martingales, provided their initial values are integrable.
17 Gaussian processes

“An $\mathbb{R}^d$-valued stochastic process is Gaussian if all its finite-dimensional distributions are Gaussian measures.” Example: Any Brownian motion with Gaussian initial distribution.

**Definition 17.1** (One-dimensional Gaussian measure). Let $m \in \mathbb{R}$ and $\sigma \geq 0$. The Gaussian measure (or normal distribution) with mean $m$ and variance $\sigma^2$, denoted $\mathcal{N}_{m,\sigma^2}$, is

a) the Dirac measure at $m$ if $\sigma = 0$, and

b) the measure with density

$$
\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}}
$$

with respect to Lebesgue measure if $\sigma > 0$.

**Remark 17.2.** Let $a \in \mathbb{R}$. Then the image measure of $\mathcal{N}_{m,\sigma^2}$ under the map $x \mapsto ax$ equals $\mathcal{N}_{am,a^2\sigma^2}$. (In probabilistic terms this means that if $X$ is normally distributed with expectation $m$ and variance $\sigma^2$, then $aX$ is normally distributed with expectation $am$ and variance $a^2\sigma^2$.)

This observation motivates the following definition:

**Definition 17.3** (Multidimensional Gaussian measure).

1. A probability measure $\mu$ on $\mathbb{R}^n$ is a Gaussian measure if for every vector $\lambda \in \mathbb{R}^n$ the image measure of $\mu$ under the map $x \mapsto \lambda \cdot x$ is a one-dimensional Gaussian measure.

2. An $\mathbb{R}^n$-valued random variable $X$ is normally distributed if its distribution is a Gaussian measure.

**Remark 17.4.** From the definition it follows immediately that an $\mathbb{R}^n$-valued random variable $X$ is Gaussian if and only if for every vector $\lambda \in \mathbb{R}^n$ the real-valued random variable $\lambda \cdot X$ is Gaussian, i.e. if and only if every linear combination of the components of $X$ is Gaussian.

**Remark 17.5.** For every matrix $A \in \mathbb{R}^{m \times n}$, every $\mathbb{R}^n$-valued Gaussian random variable $X$ and every vector $b \in \mathbb{R}^m$ the $\mathbb{R}^m$-valued random variable $AX + b$ is Gaussian as well.

**Proof.** Let $\lambda \in \mathbb{R}^m$. Then $\lambda \cdot (AX + b) = (A^T\lambda) \cdot X + \lambda \cdot b$ is Gaussian.

**Remark 17.6.** Let $X$ and $Y$ be independent Gaussian random variables, where $X$ is $\mathbb{R}^m$-valued and $Y$ is $\mathbb{R}^n$-valued. Then the $\mathbb{R}^{m+n}$-valued random variable $(X,Y)$ is Gaussian as well.

**Proof.** Let $\lambda = (\lambda_1,\lambda_2) \in \mathbb{R}^{m+n}$. Then $\lambda \cdot (X,Y) = \lambda_1 \cdot X + \lambda_2 \cdot Y$ is the sum of two independent one-dimensional Gaussian random variables and hence itself Gaussian.

**Remark 17.7.** Let $\mu$ be a Gaussian measure on $\mathbb{R}^n$. Then its mean $m \in \mathbb{R}^n$ defined by

$$
m_i := \int_{\mathbb{R}^n} x_i \mu(dx)
$$

and its covariance matrix $C \in \mathbb{R}^{N \times N}$ defined by

$$
C_{ij} := \int_{\mathbb{R}^n} (x_i - m_i)(x_j - m_j) \mu(dx)
$$

both exist.

---

$^{24}$Carl Friedrich Gauss (1777–1855), German mathematician
Proof. Let \( X \) be a random variable with distribution \( \mu \). Then its components \( X_i, i = 1, \ldots, N \), are real-valued normally distributed random variables and therefore square-integrable. Consequently, \( m_i = E(X_i) \) and \( C_{ij} = \text{Cov}(X_i, X_j) \) exist. \( \square \)

Remark 17.8. The covariance matrix \( C \) is symmetric (i.e. \( C_{ij} = C_{ji} \)) and positive semi-definite: For all \( \lambda \in \mathbb{R}^n \) we have

\[
\sum_{i,j=1}^{n} C_{ij} \lambda_i \lambda_j = \sum_{i,j=1}^{n} \lambda_i \lambda_j \text{Cov}(X_i, X_j) = \text{Var} \left( \sum_{i=1}^{n} \lambda_i X_i \right) \geq 0.
\]

Theorem 17.9. For every vector \( m \in \mathbb{R}^n \) and every symmetric and positive semi-definite matrix \( C \in \mathbb{R}^{n \times n} \) there exists exactly one Gaussian measure on \( \mathbb{R}^n \) with mean \( m \) and covariance matrix \( C \).

Proof. To prove existence, let \( R \in \mathbb{R}^{n \times n} \) be such that \( RR^T = C \) (such a matrix \( R \) exists because \( C \) is symmetric and positive semi-definite), \( X = (X_1, \ldots, X_n) \) a vector of independent standard normally distributed random variables and \( Y := RX + m \). Then we know that \( Y \) is Gaussian. Moreover, clearly \( E(Y_i) = m_i \), and

\[
\text{Cov}(Y_i, Y_j) = \text{Cov} \left( \sum_{\alpha=1}^{N} R_{i\alpha} X_\alpha, \sum_{\beta=1}^{N} R_{j\beta} X_\beta \right)
= \sum_{\alpha,\beta=1}^{N} R_{i\alpha} R_{j\beta} \text{Cov}(X_\alpha, X_\beta)
= (RR^T)_{ij}
= C_{ij}.
\]

To prove uniqueness we show that the Fourier transform\(^{25} \) of a Gaussian measure depends only on its mean and its covariance matrix. \( \square \)

Remark 17.10. The Fourier transform of a finite measure \( \mu \) on \( \mathbb{R}^d \) is the function \( \hat{\mu} : \mathbb{R}^d \to \mathbb{C} \) defined by

\[
\hat{\mu}(\xi) := \int_{\mathbb{R}^d} e^{i\xi \cdot y} \mu(dy).
\]

By the uniqueness theorem, two finite measures on \( \mathbb{R}^d \) coincide if and only if their Fourier transforms coincide.

Lemma 17.11. Let \( \mu \) be a Gaussian measure on \( \mathbb{R}^N \) with mean \( m \) and covariance matrix \( C \). Then its Fourier transform is given by

\[
\hat{\mu}(\xi) = \exp \left( im \cdot \xi - \frac{1}{2} \xi^T C \xi \right).
\]

Proof. Let \( X \) be a random variable with distribution \( \mu \), so that \( \hat{\mu}(x) = E(e^{i\xi \cdot X}) \). The random variable \( \xi \cdot X \) is normally distributed with expectation \( \xi \cdot m \) and variance

\[
\text{Var}(\xi \cdot X) = \text{Var} \left( \sum_{i=1}^{N} \xi_i X_i \right) = \sum_{i,j=1}^{N} \xi_i \xi_j \text{Cov}(X_i, X_j) = \xi^T C \xi.
\]

Consequently
\[ \hat{\mu}(\xi) = E(e^{i\xi \cdot X}) = \int_{\mathbb{R}} e^{ix} \frac{1}{\sqrt{2\pi \xi^T C \xi}} e^{-\frac{(x - \xi \cdot m)^2}{2\xi^T C \xi}} dx = \frac{1}{\sqrt{2\pi \xi^T C \xi}} e^{i\xi \cdot m} \int_{\mathbb{R}} e^{ix} e^{-\frac{x^2}{2\xi^T C \xi}} dx. \]

Note that this integral cannot be computed by completing the square and applying the translation invariance of Lebesgue measure, because here we would need translation by a complex vector, while the translation invariance of Lebesgue measure holds only for translations by real vectors. Instead, the integral can be computed as follows: Let
\[ \Phi(t) := \int_{\mathbb{R}} e^{itx} e^{-\frac{x^2}{2\xi^T C \xi}} dx, \]
so that \( \Phi(1) \) is the integral we are interested in, and \( \Phi(0) = \int_{\mathbb{R}} e^{-\frac{x^2}{2\xi^T C \xi}} dx = \sqrt{2\pi \xi^T C \xi} \).

Moreover,
\[ \Phi'(t) = i \int_{\mathbb{R}} xe^{itx} e^{-\frac{x^2}{2\xi^T C \xi}} dx, \]
and
\[ t\Phi(t) = -i \int_{\mathbb{R}} ite^{itx} e^{-\frac{x^2}{2\xi^T C \xi}} dx = -i \left[ e^{itx} e^{-\frac{x^2}{2\xi^T C \xi}} \right]_{-\infty}^{\infty} = -i \frac{\int_{\mathbb{R}} xe^{itx} e^{-\frac{x^2}{2\xi^T C \xi}} dx}{\xi^T C \xi} \]
so that
\[ \Phi'(t) = -\xi^T C \xi t \Phi(t). \]

Since the function \( \Psi(t) := e^{-\frac{t^2 \xi^T C \xi}{2}} \) solves the same linear ODE, \( \Psi'(t) = -\xi^T C \xi t \Psi(t) \), it follows that
\[ \Phi(t) = \frac{\Phi(0)}{\Psi(0)} \Psi(t) = \sqrt{2\pi \xi^T C \xi} e^{-\frac{t^2 \xi^T C \xi}{2}}, \]
and therefore
\[ \hat{\mu}(\xi) = \frac{1}{\sqrt{2\pi \xi^T C \xi}} e^{i\xi \cdot m} \Phi(1) = \exp \left( i\xi m - \frac{1}{2} \xi^T C \xi \right). \]

**Definition 17.12.** An \( \mathbb{R}^d \)-valued stochastic process \((X_t)_{t \in I}\) is called Gaussian if all its finite-dimensional distributions are Gaussian measures.

To keep notation simple we consider only real-valued Gaussian processes in the sequel. However, all definitions and results can be easily extended to \( \mathbb{R}^d \)-valued processes.

**Definition 17.13.** Let \((X_t)_{t \in I}\) be a real-valued Gaussian process. Its expectation function \( m : I \to \mathbb{R} \) and its covariance function \( \Gamma : I \times I \to \mathbb{R} \) are defined by
\[ m(t) := E(X_t) \]
and
\[ \Gamma(s, t) := \text{Cov}(X_s, X_t). \]
Remark 17.14. The covariance function is symmetric and positive semi-definite, i.e. for all \( t_1, \ldots, t_n \in I \) and all \( \lambda_1, \ldots, \lambda_n \in \mathbb{R} \) we have

\[
\sum_{i,j=1}^{n} \Gamma(t_i, t_j) \lambda_i \lambda_j \geq 0.
\]

Proof. \[
\sum_{i,j=1}^{n} \Gamma(t_i, t_j) \lambda_i \lambda_j = \sum_{i,j=1}^{n} \text{Cov}(X_{t_i}, X_{t_j}) \lambda_i \lambda_j = \text{Var} \left( \sum_{i=1}^{n} \lambda_i X_{t_i} \right) \geq 0.
\]

Remark 17.15. Since every Gaussian measure is determined by its mean and its covariance matrix, the finite-dimensional distributions of a Gaussian process are determined by its expectation and covariance functions (i.e. two Gaussian processes with the same expectation and covariance functions have the same finite-dimensional distributions).

Theorem 17.16. Let \( I \) be any set. For every function \( m : I \to \mathbb{R} \) and every symmetric and positive semi-definite function \( \Gamma : I \times I \to \mathbb{R} \) there exists a real-valued Gauss process with expectation function \( m \) and covariance function \( \Gamma \).

Proof. For each non-empty finite subset \( J \) of \( I \) let \( P_J \) be the Gauss measure with mean \( m_{|J} \) and covariance function \( C_{|J \times J} \). We have to check that the family \( (P_J)_{J \subseteq I^*} \) is projective, i.e. that for all \( J \subseteq J' \subseteq I^* \) the image measure of \( P_{J'} \) under \( \pi_{J'} \) equals \( P_J \). Let \( Y = (Y_t)_{t \in J} \) be a random variable with distribution \( P_J \). Then the distribution of the variable \( (Y_t)_{t \in J} \) is the image measure of \( P_{J'} \) under \( \pi_{J'} \). Moreover, the variable \( (Y_t)_{t \in J} \) is clearly Gaussian with mean \( m_{|J} \) and covariance matrix \( C_{|J \times J} \).

Example 17.17 (Standard Brownian motion). A real-valued stochastic process \( (B_t)_{t \in \mathbb{R}_+} \) with continuous paths is a standard Brownian motion if and only if it is a Gaussian process with expectation function \( m \equiv 0 \) and covariance function \( C \) given by \( C(s, t) = s \wedge t \).

Proof. Let \( (B_t)_{t \in \mathbb{R}_+} \) be a standard Brownian motion, and let \( 0 \leq t_1 < \ldots < t_n \). To check that the random variable \( (B_t)_{t \in\mathbb{R}_+} \) is Gaussian, let \( Y_t := \sum_{i=1}^{n} Y_{t_i} \) for \( i = 1, \ldots, n \), where \( t_0 := 0 \). By definition of Brownian motion, the increments \( Y_1, \ldots, Y_n \) are independent and normally distributed, so that the vector \( (Y_1, \ldots, Y_n) \) is Gaussian. Since \( B_{t_i} = \sum_{j=1}^{i} Y_j \), it follows that the vector \( (B_{t_1}, \ldots, B_{t_n}) \) is Gaussian as well. Moreover, \( E(B_t) = 0 \), and for \( s \leq t \) we obtain

\[
\text{Cov}(B_s, B_t) = E(B_s B_t) = E(B_s^2) + E(B_s(B_t - B_s)) = s + E(B_s E(B_t - B_s)) = s + E(B_s B_t - B_s^2) = s.
\]

The converse implication follows from the fact that the finite-dimensional distributions of a Gaussian process are determined by its expectation and covariance functions.

Remark 17.18. It is not always easy to prove that a given symmetric function \( \Gamma : I \times I \to \mathbb{R} \) is positive semidefinite.

Exercise. Give an elementary proof that the covariance function of standard Brownian motion, \( \Gamma(s, t) = s \wedge t \), is positive semidefinite. (We already know that it is positive semidefinite because we know that it is the covariance function of a stochastic process.)
Solution. Note that
\[ s \wedge t = \int_0^\infty 1_{[0,s]}(r)1_{[0,t]}(r)dr. \]

Consequently
\[
\sum_{i,j=1}^n t_i \wedge t_j \lambda_i \lambda_j = \sum_{i,j=1}^n \lambda_i \lambda_j \int_0^\infty 1_{[0,t_i]}(r)1_{[0,t_j]}(r)dr \\
= \int_0^\infty \sum_{i,j=1}^n \lambda_i \lambda_j 1_{[0,t_i]}(r)1_{[0,t_j]}(r)dr \\
= \int_0^\infty \left( \sum_{i=1}^n \lambda_i 1_{[0,t_i]}(r) \right)^2 dr \\
\geq 0.
\]

Example 17.19 (Fractional Brownian motion). Let \( H > 0 \). A real-valued stochastic process \((B_t)_{t \in \mathbb{R}_+}\) is called a fractional Brownian motion of Hurst index \( H \) if

1. it is Gaussian with expectation function \( m \equiv 0 \) and covariance function
   \[ \Gamma(s,t) = \frac{1}{2} \left( s^{2H} + t^{2H} - |s-t|^{2H} \right), \]
   and

2. almost all its paths are continuous.

Remark 17.20. If \( H = 1/2 \), we get \( \Gamma(s,t) = s \wedge t \). Hence, a fractional Brownian motion of Hurst index 1/2 is the same as a standard Brownian motion.

Proposition 17.21. The function \( \Gamma(s,t) = \frac{1}{2} \left( s^{2H} + t^{2H} - |s-t|^{2H} \right) \) is positive semi-definite if and only if \( H \leq 1 \). Consequently, a stochastic process with the first property of fractional Brownian motion with Hurst index \( H \) exists if and only if \( H \leq 1 \).

Proof. Assume first that \( H > 1 \). Choosing \( n = 2, t_1 = 1, t_2 = 2, \lambda_1 = -2 \) and \( \lambda_2 = 1 \) we obtain
\[
\sum_{i,j=1}^n \Gamma(t_i,t_j) \lambda_i \lambda_j = 1 \cdot (-2) \cdot (-2) + 2 \cdot \frac{1}{2} \cdot 2^{2H} \cdot (-2) \cdot 1 + 2^{2H} \cdot 1 \cdot 1 \\
= -4 - 2 \cdot 2^{2H} + 2^{2H} \\
= -4 \\
< 0
\]
because \( H > 1 \). Hence \( \Gamma \) is not positive semidefinite.

The case \( H = 1 \) is easy. In this case
\[ \Gamma(s,t) = \frac{1}{2} (s^2 + t^2 - (s-t)^2) = st \]
and consequently
\[
\sum_{i,j=1}^n \Gamma(t_i,t_j) \lambda_i \lambda_j = \sum_{i,j=1}^n t_i t_j \lambda_i \lambda_j = \left( \sum_{i=1}^n \lambda_i t_i \right)^2 \geq 0.
\]

\[ ^{26}\text{H. E. Hurst, Long-term storage capacity in reservoirs. Trans. Amer. Soc. Civil Eng. 116 (1951), 400–410. (Harold Edwin Hurst (1880–1978), English hydrologist)} \]
Assume now that $H \in (0,1)$ and consider for $t > 0$ the integral

$$\int_{0}^{\infty} \frac{1 - e^{-t^2 x^2}}{x^{1+2H}} \, dx.$$ 

To check that this integral is finite we have to study the integrand as $x \to \infty$ and $x \to 0$. As $x \to \infty$ the integrand behaves as $x^{-1-2H}$ and, since $H > 0$, it is therefore unproblematic. As $x \to 0$ the series expansion of the exponential function tells us that the integrand behaves as $t^2 x^{1-2H}$ which, since $H < 1$, is unproblematic as well. Consequently, the integral is finite. Moreover, the substitution $x = y/t$ implies that

$$\int_{0}^{\infty} \frac{1 - e^{-t^2 x^2}}{x^{1+2H}} \, dx = \frac{1}{t} \int_{0}^{\infty} \frac{1 - e^{-y^2}}{(y/t)^{1+2H}} \, dy = t^{2H} \int_{0}^{\infty} \frac{1 - e^{-y^2}}{y^{1+2H}} \, dy. $$

Consequently,

$$t^{2H} = \frac{1}{C} \int_{0}^{\infty} \frac{1 - e^{-t^2 x^2}}{x^{1+2H}} \, dx.$$ 

It follows that

$$C \left( s^{2H} + t^{2H} - |s-t|^{2H} \right)$$

$$= \int_{0}^{\infty} \frac{1 - e^{-s^2 x^2}}{x^{1+2H}} \, dx + \int_{0}^{\infty} \frac{1 - e^{-t^2 x^2}}{x^{1+2H}} \, dx - \int_{0}^{\infty} \frac{1 - e^{-(s-t)^2 x^2}}{x^{1+2H}} \, dx$$

$$= \int_{0}^{\infty} \frac{1 + e^{-(s+t)^2 x^2} - e^{-s^2 x^2} - e^{-t^2 x^2}}{x^{1+2H}} \, dx$$

$$= \int_{0}^{\infty} \frac{(1 - e^{-s^2 x^2})(1 - e^{-t^2 x^2})}{x^{1+2H}} \, dx + \int_{0}^{\infty} \frac{e^{-s^2 x^2} e^{2 s t x} e^{-t^2 x^2} - e^{-s^2 x^2} e^{-t^2 x^2}}{x^{1+2H}} \, dx$$

$$= \int_{0}^{\infty} \frac{(1 - e^{-s^2 x^2})(1 - e^{-t^2 x^2})}{x^{1+2H}} \, dx + \int_{0}^{\infty} \frac{e^{-s^2 x^2} e^{-t^2 x^2} (e^{2 st x} - 1)}{x^{1+2H}} \, dx$$

$$= \int_{0}^{\infty} \frac{(1 - e^{-s^2 x^2})(1 - e^{-t^2 x^2})}{x^{1+2H}} \, dx + \int_{0}^{\infty} \frac{e^{-s^2 x^2} e^{-t^2 x^2} \sum_{n=1}^{\infty} \frac{(2 s t x)^n}{n!}}{x^{1+2H}} \, dx$$

$$= \int_{0}^{\infty} \frac{(1 - e^{-s^2 x^2})(1 - e^{-t^2 x^2})}{x^{1+2H}} \, dx + \sum_{n=1}^{\infty} \frac{2^n}{n!} \int_{0}^{\infty} \frac{e^{-s^2 x^2} e^{-t^2 x^2} (s t x)^n}{x^{1+2H}} \, dx$$

We have hence succeeded in representing $\Gamma(s,t)$ in the form

$$\Gamma(s,t) = \int f(s,\xi) f(t,\xi) \mu(d\xi)$$

62
(a sort of “diagonalization” with a nonnegative diagonal matrix). It follows that

\[
\sum_{i,j=1}^{n} \Gamma(t_i, t_j) \lambda_i \lambda_j = \sum_{i,j=1}^{n} \lambda_i \lambda_j \int f(t_i, \xi) f(t_j, \xi) \mu(d\xi) \\
= \int \sum_{i,j=1}^{n} \lambda_i \lambda_j f(t_i, \xi) f(t_j, \xi) \mu(d\xi) \\
= \int \left( \sum_{i=1}^{n} \lambda_i f(t_i, \xi) \right)^2 \mu(d\xi) \\
\geq 0.
\]

**Proposition 17.22.** Fractional Brownian motion of Hurst index \(H\) exists if and only if \(H \leq 1\). In this case, almost all its paths are locally Hölder continuous of any exponent \(\gamma < H\).

**Proof.** We have already seen that a process with the first property of fractional Brownian motion of Hurst index \(H\) exists if and only if \(H \leq 1\). It remains to show that such a process \((X_t)_{t \in \mathbb{R}_+}\) admits a continuous modification. To do so observe that for all \(s,t \in \mathbb{R}_+\) the variable \(X_t - X_s\) is Gaussian with mean 0 and variance

\[
\text{Var}(X_t - X_s) = \text{Var}(X_t) + \text{Var}(X_s) - 2 \text{Cov}(X_s, X_t) \\
= \frac{1}{2} \left( 2t^{2H} + 2s^{2H} - 2(s^{2H} + t^{2H} - |s-t|^{2H}) \right) \\
= |s-t|^{2H}
\]

(note the analogy with usual Brownian motion). Let now

\[
Z_{s,t} := \frac{X_t - X_s}{|t-s|^{H}}.
\]

\(Z_{s,t}\) is normally distributed with mean 0 and variance 1, so that \(C_{\alpha, H} := E(|Z_{s,t}|^\alpha)\) does not depend on \(s\) and \(t\). It follows that

\[
E(|X_t - X_s|^\alpha) = |t-s|^\alpha E(|Z_{s,t}|^\alpha) = C_{\alpha, H} |t-s|^\alpha H.
\]

For any \(\alpha > 0\) let \(\beta := \alpha H - 1\). Then \(\beta\) is positive as soon as \(\alpha > 1/H\). Moreover \(\beta/\alpha \to H\) as \(\alpha \to \infty\). The result therefore follows from Kolmogorov’s continuity theorem.

The characterization of standard Brownian motion as a Gaussian process with expectation function \(m \equiv 0\) and covariance function \(\Gamma(s, t) = s \wedge t\) allows to prove the following remarkable fact:

**Proposition 17.23** (Time inversion of Brownian motion). Let \((B_t)_{t \in \mathbb{R}_+}\) be a one-dimensional standard Brownian motion. Then the process \((Y_t)_{t \in \mathbb{R}_+}\) defined by

\[
Y_t := \begin{cases} 
  tB_{1/t} & \text{if } t > 0 \\
  0 & \text{if } t = 0
\end{cases}
\]

is a standard Brownian motion as well.
Proof. The process \((Y_t)_{t \in \mathbb{R}^+}\) is clearly a Gauss process with \(E(Y_t) = tE(B_{1/t}) = 0\) and
\[
\text{Cov}(Y_s, Y_t) = st \text{Cov}(B_{1/s}, B_{1/t}) = st \frac{1}{t} = s.
\]
Hence it has the finite-dimensional distributions of Brownian motion and consequently satisfies
\[
E(|Y_t - Y_s|^\alpha) = C_\alpha (t - s)^{\alpha/2}.
\]
Therefore, by Kolmogorov’s continuity theorem it has a continuous modification \((\tilde{Y}_t)_{t \in \mathbb{R}^+}\).

Hence for each \(t \in \mathbb{R}^+\) we have \(Y_t = \tilde{Y}_t\) almost surely, there exists an event \(\Omega_1\) with \(P(\Omega_1) = 1\) such that on \(\Omega_1\) we have \(Y_t = \tilde{Y}_t\) for all \(t \in \mathbb{Q}^+\). Moreover there exists an event \(\Omega_2\) with \(P(\Omega_2) = 1\) such that on \(\Omega_2\) the process \((Y_t)_{t \in \mathbb{R}^+}\) is continuous everywhere except possibly at 0. It follows that on \(\Omega_1 \cap \Omega_2\) we have \(Y_t = \tilde{Y}_t\) for all \(t \in \mathbb{R}^+\), hence almost all paths of the process \((Y_t)_{t \in \mathbb{R}^+}\) are continuous even at 0.

18 Path properties of Brownian motion

We already know that almost every path of a Brownian motion is locally Hölder continuous of any exponent \(\gamma < 1/2\). A natural question is whether Brownian paths have even more regularity, in particular if they are differentiable. This question is answered by the law of the iterated logarithm. For its proof we will need the second part of the Borel-Cantelli lemma:

**Proposition 18.1.** Let \((A_n)_{n \in \mathbb{N}}\) be a sequence of independent events such that \(\sum_{n=1}^{\infty} P(A_n) = \infty\). Then
\[
P\left( \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} A_m \right) = 1,
\]
i.e. with probability 1 infinitely many of the events \(A_n\) occur.

Proof. It suffices to prove that for each \(n \in \mathbb{N}\) we have \(P\left( \bigcup_{m \geq n} A_m \right) = 1\) or, in other words, \(P\left( \bigcap_{m \geq n} A_m^c \right) = 0\). Using the independence of the events \((A_n)_{n \in \mathbb{N}}\) we obtain
\[
P\left( \bigcap_{m \geq n} A_m^c \right) = \lim_{N \to \infty} P\left( \bigcap_{m=n}^{N} A_m^c \right)
\]
\[
= \lim_{N \to \infty} \prod_{m=n}^{N} P(A_m^c)
\]
\[
= \lim_{N \to \infty} \prod_{m=n}^{N} (1 - P(A_m))
\]
\[
\leq \liminf_{N \to \infty} \prod_{m=n}^{N} e^{-P(A_m)}
\]
\[
= \liminf_{N \to \infty} \exp\left( - \sum_{m=n}^{N} P(A_m) \right)
\]
\[
= 0. \qed
\]

Moreover, we need the following lemma:
Lemma 18.2.

\[ P \left( \sup_{s \in [0,t]} B_s \geq \alpha t \right) \leq e^{-\alpha^2 t/2}. \]

Proof. Clearly

\[
\left\{ \sup_{s \in [0,t]} B_s \geq \alpha t \right\} = \left\{ \sup_{s \in [0,t]} (B_s - \alpha t/2) \geq \alpha t/2 \right\} \subseteq \left\{ \sup_{s \in [0,t]} (B_s - \alpha s/2) \geq \alpha t/2 \right\} = \left\{ \sup_{s \in [0,t]} e^{\alpha B_s - \alpha^2 s/2} \geq e^{\alpha^2 t/2} \right\}.
\]

We now apply the first maximal inequality to the martingale \((e^{\alpha B_t - \alpha^2 t/2})_{t \in \mathbb{R}_+}\) and obtain

\[
P \left( \sup_{s \in [0,t]} e^{\alpha B_s - \alpha^2 s/2} \geq e^{\alpha^2 t/2} \right) \leq e^{-\alpha^2 t/2} E \left[ e^{\alpha B_t - \alpha^2 t/2} \right] = e^{-\alpha^2 t/2}.
\]

Theorem 18.3 (Law of the iterated logarithm for the short-time behavior of Brownian motion). Let \((B_t)_{t \in \mathbb{R}_+}\) be a one-dimensional standard Brownian motion. Then almost surely

\[
\limsup_{t \to 0} \frac{B_t}{\sqrt{2t \log \log 1/t}} = 1,
\]

\[
\liminf_{t \to 0} \frac{B_t}{\sqrt{2t \log \log 1/t}} = -1.
\]

Remark 18.4.

1. For \(t \in (0, 1/e)\) we have \(1/t > e\), hence \(\log 1/t > 1\), and consequently \(\log \log 1/t > 0\). Consequently, the functions

\[ \psi(t) := t \log \log 1/t \]

and

\[ \varphi(t) := \sqrt{2t \log \log 1/t} \]

are defined for all \(t \in (0, 1/e)\). Moreover,

\[ \psi'(t) = \log \log 1/t + t \frac{1}{\log 1/t} \left( -\frac{1}{t^2} \right) = \log \log 1/t - \frac{1}{\log 1/t}, \]

so that for \(t \leq e^{-e}\)

\[ \psi'(t) \geq 1 - 1/e > 0. \]

Hence \(\varphi\) is strictly increasing on the interval \((0, e^{-e})\). In the sequel we will only work on this interval.

2. As \(t \to 0\), \(\log \log 1/t \to \infty\), but only extremely slowly. For example, \(\log \log 10^3 \approx 1.9\), \(\log \log 10^6 \approx 2.6\), \(\log \log 10^{100} \approx 5.4\). In particular, \(t \log \log 1/t \to 0\) as \(t \to 0\).
3. The result implies that almost every path of Brownian motion is not even pointwise Hölder continuous of exponent 1/2 and in particular not differentiable.

4. Since the process $(-B_t)_{t \in \mathbb{R}^+}$ is a standard Brownian motion as well, it suffices to prove the first statement.

5. There is also a law of the iterated logarithm for the long-time behavior of Brownian motion which can be obtained from the short-time law by time inversion.

Proof of Theorem 18.5. We will show that for every $\varepsilon > 0$,

$$P \left\{ \limsup_{t \to 0} \frac{B_t}{\varphi(t)} > 1 + \varepsilon \right\} = 0,$$

$$P \left\{ \limsup_{t \to 0} \frac{B_t}{\varphi(t)} < 1 - \varepsilon \right\} = 0.$$

To prove the first of these statements let $t_n := (1 + \varepsilon)^{-n}$. Since $t_n \searrow 0$, there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ we have $t_n < e^{-\varepsilon}$. For $n \geq n_0$ let

$$C_n := \left\{ \sup_{t \in [t_{n+1}, t_n]} \frac{B_t}{\varphi(t)} > 1 + \varepsilon \right\}$$

Then each $\omega \in \left\{ \limsup_{t \to 0} \frac{B_t}{\varphi(t)} > 1 + \varepsilon \right\}$ is contained in infinitely many of the events $C_n$, i.e.

$$\left\{ \limsup_{t \to 0} \frac{B_t}{\varphi(t)} > 1 + \varepsilon \right\} \subseteq \bigcap_{n \geq n_0} \bigcup_{m \geq n} C_m.$$

In order to apply the Borel-Cantelli lemma, we have to show that

$$\sum_{n=n_0}^{\infty} P(C_n) < \infty.$$

To do so, note that since $\varphi$ is increasing on the interval $(0, e^{-\varepsilon})$,

$$C_n \subseteq \left\{ \sup_{t \in [0, t_n]} B_t \varphi(t_{n+1}) > 1 + \varepsilon \right\} = \left\{ \sup_{t \in [0, t_n]} B_t > (1 + \varepsilon) \varphi(t_{n+1}) \right\},$$

so that

$$P(C_n) \leq P \left\{ \sup_{t \in [0, t_n]} B_t > (1 + \varepsilon) \varphi(t_{n+1}) \right\}$$

$$= P \left\{ \sup_{t \in [0, t_n]} B_t > \frac{(1 + \varepsilon) \varphi(t_{n+1})}{t_n} \right\}$$

$$\leq \exp \left( -\frac{(1 + \varepsilon)^2 \varphi(t_{n+1})^2}{2t_n} \right)$$

$$= \exp \left( -\frac{(1 + \varepsilon)^2 2(1 + \varepsilon)^{-(n+1)} \log \log((1 + \varepsilon)^{n+1})}{2(1 + \varepsilon)^{-n}} \right)$$

$$= \exp \left( -(1 + \varepsilon) \log \log((1 + \varepsilon)^{n+1}) \right)$$

$$= \exp \left( -(1 + \varepsilon) \log((n + 1) \log(1 + \varepsilon)) \right)$$

$$= \left( (n + 1) \log(1 + \varepsilon) \right)^{-(1+\varepsilon)}$$

$$= \log(1 + \varepsilon)^{-(1+\varepsilon)} (n + 1)^{-(1+\varepsilon)},$$

66
and consequently
\[ \sum_{n=n_0}^{\infty} P(C_n) < \infty, \]
so that
\[ P \left( \limsup_{t \to 0} \frac{B_t}{\varphi(t)} > 1 + \varepsilon \right) \leq P \left( \bigcap_{n \geq n_0} \bigcup_{m \geq n} C_m \right) = 0. \]

To prove that
\[ P \left\{ \limsup_{t \to 0} \frac{B_t}{\varphi(t)} < 1 - \varepsilon \right\} = 0, \]
let \( \delta := \varepsilon/2 \) and \( q := \varepsilon^2/64 \), so that
\[ \delta + 2(1 + \delta) \sqrt{q} = \varepsilon^2 + (2 + \varepsilon) \frac{\varepsilon}{8} < \varepsilon \]
if \( \varepsilon \leq 1 \), and
\[ \alpha := 2(1 - \delta)^2 = \frac{2(1 - \varepsilon/2)^2}{1 - \varepsilon^2/64} = \frac{2(1 - \varepsilon/2)^2}{(1 - \varepsilon/8)(1 + \varepsilon/8)} < \frac{2(1 - \varepsilon/2)}{(1 - \varepsilon/8)} < 2. \]

Since \( q < 1 \), the sequence \((t_n)_{n \in \mathbb{N}}\) defined by \( t_n := q^n \) is strictly decreasing and converges to 0. The random variables \( Z_n := B_{t_n} - B_{t_{n+1}} \) are independent. Since \( Z_n \) is normally distributed with mean 0 and variance \( t_n - t_{n+1} \), \( Z_n/\sqrt{t_n - t_{n+1}} \) is standard normally distributed, so that for each \( x \geq 1 \) we obtain
\[
P \left( Z_n > x \sqrt{t_n - t_{n+1}} \right) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-y^2/2} dy > \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} \frac{1 + y^2}{1 + x^2} e^{-y^2/2} dy
\]
\[
= \frac{1}{\sqrt{2\pi}} \left( \int_{x}^{\infty} e^{-y^2/2} dy + \int_{x}^{\infty} \frac{1}{y^2} e^{-y^2/2} dy \right)
= \frac{1}{\sqrt{2\pi}} \left( \int_{x}^{\infty} e^{-y^2/2} dy + \left[ -\frac{1}{y} e^{-y^2/2} \right]_{y=x}^{y=\infty} - \int_{x}^{\infty} e^{-y^2/2} dy \right)
= \frac{1}{\sqrt{2\pi}} \frac{1}{2x} e^{-x^2/2}
\]
We apply this inequality to
\[
x_n := (1 - \delta) \frac{\varphi(t_n)}{\sqrt{t_n - t_{n+1}}}
= \frac{(1 - \delta) \sqrt{2q^n} \log q^{-n}}{\sqrt{q^n - q^{n+1}}}
= \frac{(1 - \delta) \sqrt{2}}{\sqrt{1 - q}} \sqrt{\log(n \log 1/q)}
= \sqrt{\alpha} \sqrt{\log(n \log 1/q)}.
\]
Clearly \( x_n \to \infty \) as \( n \to \infty \); in particular \( x_n > 1 \) for all sufficiently large \( n \). It follows that for all \( n > \log 1/q \) we have

\[
P(Z_n > (1 - \delta) \varphi(t_n)) = P(Z_n > x_n \sqrt{t_n - t_{n+1}}) > \frac{1}{\sqrt{2\pi}} \frac{1}{2x_n} e^{-x_n^2 \frac{\alpha}{2}}
\]

\[
= \frac{1}{2\sqrt{2\pi}} \frac{\exp \left(-\frac{\alpha}{2} \log(n \log 1/q)\right)}{\sqrt{\alpha \log(n^2)}}
\]

\[
> \frac{1}{2\sqrt{2\pi}} \frac{\exp \left(-\frac{\alpha}{2} \log n + \log 1/q\right)}{\sqrt{\alpha \log(n^2)}}
\]

\[
= \frac{\exp \left(-\frac{\alpha}{2} \log 1/q\right) \exp \left(-\frac{\alpha}{2} \log n\right)}{2\sqrt{2\pi} \sqrt{2\alpha} \log n}
\]

\[
= C n^{-\alpha/2} \frac{1}{\sqrt{\log n}}.
\]

Since \( \alpha < 2 \), the series \( \sum_{n=2}^{\infty} n^{-\alpha/2} \frac{1}{\sqrt{\log n}} \) diverges, so that the second part of the Borel-Cantelli lemma implies that

\[
P \left( \bigcap_{n \in \mathbb{N}} \bigcup_{m \geq n} \{ Z_n > (1 - \delta) \varphi(t_n) \} \right) = 1,
\]

i.e. with probability 1 we have

\[
Z_n > (1 - \delta) \varphi(t_n)
\]

for infinitely many \( n \).

Consequently,

\[
\frac{B(t_n)}{\varphi(t_n)} = \frac{B(t_{n+1} + Z_n)}{\varphi(t_n)} > 1 - \delta + \frac{B(t_{n+1})}{\varphi(t_n)}
\]

for infinitely many \( n \).

Since by the upper estimate with probability 1

\[
\limsup_{t \to 0} -\frac{B(t)}{\varphi(t)} < 1 + \delta,
\]

we have

\[
B(t_{n+1}) > -(1 + \delta) \varphi(t_{n+1})
\]

for infinitely many \( n \) and therefore

\[
\frac{B(t_n)}{\varphi(t_n)} > 1 - \delta - (1 + \delta) \frac{\varphi(t_{n+1})}{\varphi(t_n)}
\]

for infinitely many \( n \).

Since

\[
\frac{\varphi(t_{n+1})}{\varphi(t_n)} = \frac{\sqrt{2q^{n+1} \log \log 1/q^{n+1}}}{\sqrt{2q^n \log \log 1/q^n}} = \sqrt{q} \sqrt{\frac{\log \log 1/q^{n+1}}{\log \log 1/q^n}} \to \sqrt{q},
\]

it follows that

\[
\frac{B(t_n)}{\varphi(t_n)} > 1 - \delta - 2(1 + \delta) \sqrt{q} > 1 - \varepsilon
\]

for infinitely many \( n \) (because \( \delta \) and \( q \) were chosen in such a way that \( \delta + 2(1 + \delta) \sqrt{q} < \varepsilon \)).
**Corollary 18.5** (Law of the iterated logarithm for the long-time behavior of Brownian motion). Let \((B_t)_{t \in \mathbb{R}_+}\) be a one-dimensional standard Brownian motion. Then almost surely

\[
\limsup_{t \to \infty} \frac{B_t}{\sqrt{2t \log \log t}} = 1,
\]

\[
\liminf_{t \to \infty} \frac{B_t}{\sqrt{2t \log \log t}} = -1.
\]

**Proof.** We know that the process \((Y_t)_{t \in \mathbb{R}_+}\) defined by

\[
Y_t := \begin{cases} 
  tB_{1/t} & \text{if } t > 0 \\
  0 & \text{if } t = 0.
\end{cases}
\]

is a standard Brownian motion as well. Consequently,

\[
1 = \limsup_{t \to 0} \frac{Y_t}{\sqrt{2t \log \log 1/t}}
= \limsup_{t \to 0} \frac{tB_{1/t}}{\sqrt{2t \log \log 1/t}}
= \limsup_{t \to 0} \frac{B_{1/t}}{\sqrt{2/t \log \log 1/t}}
= \limsup_{t \to \infty} \frac{B_t}{\sqrt{2t \log \log t}}.
\]