Stochastic Processes

Peter Pfaffelhuber

Version: November 30, 2024

Prelude

Stochastic processes play a central role in modern stochastics. They are used in various application fields, including financial mathematics, as well as in biology and physics. Stochastic processes are always used when a variable - for example a stock price, the frequency of an allele in a population or the position of a small particle – changes randomly over time.

The aim here is to provide important tools for dealing with stochastic processes. We will deal with important examples, such as the Poisson process or Brownian motion. The latter also plays a decisive role in the construction of stochastic integrals.

The following books have guided me as references for the purpose of this manuscript.

- Durrett, Rick. Probability: Theory and Examples, Cambridge Series in Statistical and Probabilistic Mathematics, 2019
- Kallenberg, Olaf. Foundations of Modern Probability Theory. Springer, third edition, 2021
- Klenke, Achim. Probability theory. A comprehensive course. Springer, 2014

This manuscript is based on the courses in Measure Theory and Probability Theory, which cover Sections 1–3, and 4–12, respectively.

The present english version of this manuscript was written based on the German version with the help of DeepL.

Contents

13 Introduction

Stochastic processes are nothing more than families of random variables. It is important to realize that this family is indexed by with time. In the course of time, more and more random variables are realized.

In the following, let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, (E, r) a complete and separable metric space with with Borel's σ-algebra $\mathcal{B}(E)$ and I an ordered subset of $\overline{\mathbb{R}}$, which we also call index set. We will always consider the two cases $I \subseteq \overline{\mathbb{Z}}$ and $I \subseteq \overline{\mathbb{R}}$. Note already here that an uncountable index set, such as $I = \mathbb{R}$, raises new questions, as probability measures are only known to be able to deal with countably number of events.

13.1 Definition and existence

First of all, we take care of the elementary question of what a stochastic process is and how it can be defined in an unambiguous way.

Definition 13.1 (Stochastic process). *1. Let $\mathcal{X} = (X_t)_{t \in I}$ such that $X_t : \Omega \to E$ is $\mathcal{F}/\mathcal{B}(E)$-measurable. Then, $\mathcal{X}$ is called an E-valued* (stochastic) process. *For $\omega \in \Omega$, the mapping given by $X(\omega) : t \mapsto X_t(\omega)$ is called a* path *of X.*

2. *If in 1., the probability space $\Omega = E^I$ and $X_t = \pi_t$ is the projection, then $\mathcal{X}$ is called* canonical process.

3. *Let $0 < p < \infty$. A real-valued process $\mathcal{X} = (X_t)_{t \in I}$ is called p-fold integrable if $\mathbf{E}[|X_t|^p] < \infty$ for all $t \in I$. It is called L^p-bounded, if $\sup_{t \in I} \mathbf{E}[|X_t|^p] < \infty$.*

In the Sections 13.2 and 13.3, we will become familiar with two examples of stochastic processes. In particular, the Poisson process (see Section 13.2) is the first process with an uncountable index set $I = [0, \infty)$.

Example 13.2 (Sums of independent random variables and Markov chains). *From the lecture* Elementary Probability Theory*, some stochastic processes are already known, even if they were not called stochastic processes.*

1. *Let $(X_t)_{t \in I}$ be independent. Then $\mathcal{X} = (X_t)_{t \in I}$ is a (very simple) stochastic process.*

2. *Let $X_1, X_2, ...$ be real-valued, independent, identically distributed random variables. Then, $\mathcal{S} = (S_t)_{t=0,1,2,...}$ with $S_0 = 0$ and*

$$S_t = \sum_{i=1}^{t} X_i$$

for $t = 1, 2, ...$ is a real-valued, stochastic process with index set $I = \{0, 1, 2, ...\}$. In particular, if $\mathbf{P}(X_i = \pm 1) = 1/2$, then $\mathcal{S}$ is called a one-dimensional, simple random walk; see Figure 1.

3. *Let $\kappa(.,.)$ be a stochastic kernel (see Definition 5.9) from $(E, \mathcal{B}(E))$ to $(E, \mathcal{B}(E))$. Further, let X_0 be an E-valued random variable and given X_t, X_{t+1} has the distribution $\kappa(X_t, .)$, $t = 0, 1, 2, ...$ Then $(X_t)_{t=0,1,...}$ is called an E-valued* Markov chain.

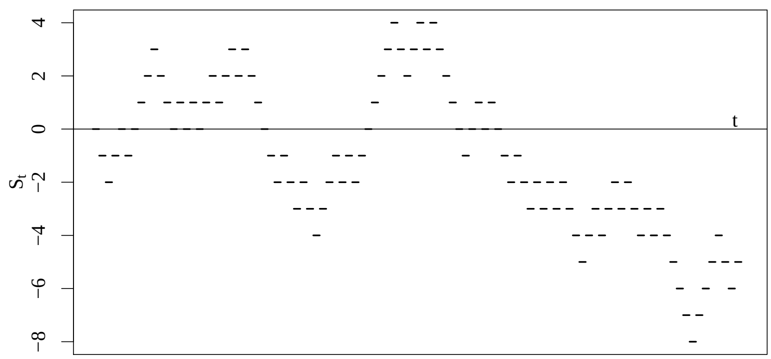

Figure 1: A path of a one-dimensional random walk.

Remark 13.3 (Repetition: Existence of stochastic Processes)**.**

1. *Recall from Section 5: the product σ-algebra on the space E^I is defined as the smallest σ-algebra with respect to which all projections $\pi_t, t \in I$ are measurable. In particular, for an E-valued stochastic process $\mathcal{X} = (X_t)_{t\in I}$, the mapping $\omega \mapsto X(\omega)$ is $\mathcal{F}/\mathcal{B}(E)^I$-measurable. Furthermore, a projective family on $\mathcal{F}$ is a family of distributions $(\mathbf{P}_J)_{J\subseteq_f I}$ with $\mathbf{P}_H = (\pi_H^J)_*\mathbf{P}_J$ for $H \subseteq J$, where π_H^J is the projection of E^J onto E^H.*

2. *Often, the finite-dimensional distributions of a stochastic process $\mathcal{X} = (X_t)_{t\in I}$, i.e. the joint distribution of $(X_{t_1}, ..., X_{t_n})$ for any $t_1, ..., t_n \in I$, are given. For example, in Sections 13.2 and 13.3, the Poisson process and Brownian motion are given by specifying the joint distribution of $(X_{t_1}, X_{t_2} - X_{t_1}, ..., X_{t_n} - X_{t_{n-1}})$. This also uniquely defines the finite-dimensional distributions. In order to ensure that there is a stochastic process for these finite-dimensional distributions, we need Kolmogorov's extension theorem; see Theorem 5.24. It should be noted that finite dimensional distributions of stochastic processes are always projective; see also Example 5.22.2.*

Definition 13.4 (Equality of stochastic processes)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ and $\mathcal{Y} = (Y_t)_{t\in I}$ be two E-valued stochastic processes.*

1. *If $\mathcal{X} \stackrel{d}{=} \mathcal{Y}$, then $\mathcal{Y}$ is a* version *of $\mathcal{X}$ (and vice versa).*

2. *If $\mathcal{X}$ and $\mathcal{Y}$ are defined on the same probability space and $\mathbf{P}(X_t = Y_t) = 1$ for all $t \in I$, then $\mathcal{X}$ is called a* modification *of $\mathcal{Y}$ (and vice versa).*

3. *If $\mathcal{X}$ and $\mathcal{Y}$ are defined on the same probability space and $\mathbf{P}(X_t = Y_t \text{ for all } t \in I) = 1$, then $\mathcal{X}$ and $\mathcal{Y}$ are called* indistinguishable.

The paths $t \mapsto X_t(\omega)$ of a stochastic process can have have certain properties. For example, they can be continuous functions $I \to E$. In addition to processes with continuous paths, we will need processes with right-continuous paths and left-limits.

Definition 13.5 (Right-continuous functions, left limits)**.** *A function $f : I \to E$ is called* *right-continuous in $t \in I$ with left-sided limit value*[1] *if*

$$f(t) = \lim_{s \downarrow t} f(s) \text{ and } \lim_{s \uparrow t} f(s) \text{ exists.}$$

It is called right-continuous with left limit values if this property holds for all $t \in I$. The set of right-continuous functions with left limits is denoted by $\mathcal{D}_E(I)$.

Proposition 13.6 (Versions, modifications, indistinguishable processes)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ and $\mathcal{Y} = (Y_t)_{t\in I}$ be stochastic processes with values in E.*

1. *The process $\mathcal{Y}$ is a version of $\mathcal{X}$ (and vice versa) if both processes have the same finite-dimensional distributions, i.e. $(X_{t_1}, ..., X_{t_n}) \stackrel{d}{=} (Y_{t_1}, ..., Y_{t_n})$ for any choice of $n \in \mathbb{N}$ and $t_1, ..., t_n \in I$.*
2. *If $\mathcal{X}$ and $\mathcal{Y}$ are indistinguishable, then $\mathcal{X}$ is a modification of $\mathcal{Y}$ (or vice versa). If $\mathcal{X}$ is a modification of $\mathcal{Y}$, then $\mathcal{X}$ is a version of $\mathcal{Y}$.*
3. *If I is at most countable and $\mathcal{X}$ is a modification of $\mathcal{Y}$ (or vice versa), then $\mathcal{X}$ and $\mathcal{Y}$ are indistinguishable.*
4. *If $I = [0, \infty)$ and $\mathcal{X}$ and $\mathcal{Y}$ have almost surely right-continuous paths and $\mathcal{X}$ is a modification of $\mathcal{Y}$, then $\mathcal{X}$ and $\mathcal{Y}$ are indistinguishable.*

Proof. 1. '$\Rightarrow$': clear. '$\Leftarrow$': Let $(\Omega, \mathcal{F}, \mathbf{P})$ and $(\Omega', \mathcal{F}', \mathbf{P}')$ the probability spaces on which probability spaces on which $\mathcal{X}$ and $\mathcal{Y}$ are defined. We consider the $\cap$-stable generator

$$\mathcal{C} := \{\pi_J^{-1}(A) : A \in \mathcal{B}(E)^{|J|}; J \subseteq_f I\} \subseteq \mathcal{B}(E)^I$$

of $\mathcal{B}(E)^I$. Further, for $J \subseteq_f I$, $A \in \mathcal{B}(E)^{|J|}$,

$$\mathbf{P}((X_t)_{t\in J} \in A) = \mathbf{P}'_I((Y_t)_{t\in J} \in A),$$

i.e. $\mathcal{X}_*\mathbf{P}$ and $\mathcal{Y}_*\mathbf{P}'$ coincide on $\mathcal{C}$. According to Theorem 2.11, this means that $\mathcal{X}_*\mathbf{P} = \mathcal{Y}_*\mathbf{P}'$. So $\mathcal{Y}$ is a version of $\mathcal{X}$.
2. Let $t \in I$. If $\mathcal{X}$ and $\mathcal{Y}$ are indistinguishable, then $\mathbf{P}(X_t \neq Y_t) \leq \mathbf{P}(X_s \neq Y_s \text{ for a } s \in I) = 0$. If $\mathcal{X}$ and $\mathcal{Y}$ are modifications and $t_1, ..., t_n \in I$, then

$$\mathbf{P}(X_{t_1} = Y_{t_1}, ..., X_{t_n} = Y_{t_n}) = 1$$

since finite unions of null-sets are null-sets. In particular, $\mathcal{X}$ and $\mathcal{Y}$ have the same finite-dimensional distributions. According to 1. $\mathcal{Y}$ is therefore is a version of $\mathcal{X}$.
3. The statement is clear because of the σ-subadditivity of probability measures,

$$\mathbf{P}(X_t \neq Y_t \text{ for a } t \in I) \leq \sum_{t\in I} \mathbf{P}(X_t \neq Y_t) = 0.$$

4. Let R be a set with $\mathbf{P}(R) = 1$ such that $\mathcal{X}$ and $\mathcal{Y}$ have right-continuous paths on R and $N_t := \{X_t \neq Y_t\}$. Further, let $I' = I \cap \mathbb{Q}$. Then, $\mathbf{P}(\bigcup_{t\in I'} N_t) = 0$ and

$$\mathbf{P}\Big(\bigcup_{t\in I} N_t\Big) \leq \mathbf{P}\Big(R \cap \bigcup_{t\in I} \bigcup_{r\geq t, r\in I'} N_r\Big) = \mathbf{P}\Big(R \cap \bigcup_{r\in I'} N_r\Big) = 0.$$

□

[1] Such functions are also called rcll (right-continuous with left limits) or càdlàg (continue à droite, limite à gauche)

Remark 13.7 (Versions with different path properties)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be an E-valued stochastic process and $I = [0,\infty)$. Each path $t \mapsto X_t(\omega)$ is therefore a mapping $I \to E$. A distinction is made between stochastic processes according to their path properties. For example, if $t \mapsto X_t(\omega)$ is a continuous function for almost all ω, we say that $\mathcal{X}$ has (almost certainly) continuous paths. It is important to realize that the property of the process to have continuous paths cannot be read from its distribution:*

Let $\mathcal{Y} = (Y_t)_{t\in I}$ with $Y_t = 0$, and $T \sim \exp(1)$ and $\mathcal{X} = (X_t)_{t\in I}$ given by

$$X_t = \begin{cases} 1, & t = T, \\ 0, & \textit{otherwise}. \end{cases}$$

Then $\mathbf{P}(X_t = Y_t) = \mathbf{P}(T \neq t) = 1$ for each $t \in I$. So $\mathcal{X}$ is a modification of $\mathcal{Y}$. In particular, according to the last proposition, the distributions of $\mathcal{X}$ and $\mathcal{Y}$ coincide. However, only $\mathcal{Y}$ has continuous paths, but every path of $\mathcal{X}$ is discontinuous (at T). In particular, $\mathcal{X}$ and $\mathcal{Y}$ are not indistinguishable.

Theorem 13.8 (Continuous modifications; Kolmogorov, Chentsov)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be an E-valued stochastic process with $I = \mathbb{R}$ or $I = [0,\infty)$. For every $\tau > 0$ there are numbers $\alpha, \beta, C > 0$ with*

$$\mathbf{E}[r(X_s, X_t)^\alpha] \leq C|t-s|^{1+\beta}$$

for all $0 \leq s, t \leq \tau$. Then there is a modification $\widetilde{\mathcal{X}} = (\widetilde{X}_t)_{t\in I}$ of $\mathcal{X}$ with continuous paths. The paths are even almost surely local Hölder-continuous of any order $\gamma \in (0, \beta/\alpha)$.[2]

Proof. It is sufficient to show the statement for $I = [0,1]$. The general case follows by dividing I into countably many intervals of length 1. We consider the set of time points

$$D_n := \{0, 1, ..., 2^n\} \cdot 2^{-n}$$

for $n = 0, 1, ...$, $D = \bigcup_{n=0}^\infty D_n$ and the random variable

$$\xi_n := \max\{r(X_s, X_t) : s, t \in D_n, |t-s| = 2^{-n}\}.$$

Let $0 < \gamma < \beta/\alpha$. Then for some $C > 0$,

$$\begin{aligned} \mathbf{E}\Big[\sum_{n=0}^\infty (2^{\gamma n}\xi_n)^\alpha\Big] &= \sum_{n=0}^\infty 2^{\alpha\gamma n}\mathbf{E}[\xi_n^\alpha] \leq \sum_{n=0}^\infty 2^{\alpha\gamma n} \sum_{s,t\in D_n, |t-s|=2^{-n}} \mathbf{E}[r(X_s, X_t)^\alpha] \\ &\leq C\sum_{n=0}^\infty 2^{\alpha\gamma n}2^n 2^{-n(1+\beta)} = C\sum_{n=0}^\infty 2^{(\alpha\gamma-\beta)n} < \infty. \end{aligned} \tag{13.1}$$

Therefore, there is a random variable C' with $\xi_n \leq C'2^{-\gamma n}$ for all $n = 0, 1, ...$ Now let $m \in \{0, 1, ...\}$ and $r \in [2^{-m-1}, 2^{-m}] \cap D$. Then,

$$\begin{aligned} \sup\big\{r(X_s, X_t) : s, t \in D, |s-t| \leq r\big\} &= \sup_{n\geq m}\big\{r(X_s, X_t) : s, t \in D_n, |s-t| \leq r\big\} \\ &\leq 2\sum_{n\geq m}\xi_n \leq 2C'\sum_{n\geq m} 2^{-\gamma n} \leq C''2^{-\gamma(m-1)} \leq C''r^\gamma. \end{aligned} \tag{13.2}$$

[2]As a reminder: a function $f : I \to E$ is locally Hölder-continuous of order γ, if for every $\tau > 0$ there is a C with $r(f(s), f(t)) \leq C|t-s|^\gamma$ for all $0 \leq s, t \leq \tau$.

for a random variable C''. It follows that almost every path on D is Hölder-continuous to the parameter γ. This means that $\mathcal{X}$ can be extended Hölder-continuously to I. We call this continuous extension $\mathcal{Y} = (Y_t)_{t\in I}$. To show that $\mathcal{Y}$ is a modification of $\mathcal{X}$, we consider a $t \in I$ and a sequence $t_1, t_2, ... \in D$ with $t_n \to t$ with $n \to \infty$. Because of the condition, $\mathbf{P}(r(X_{t_n}, X_t) > \varepsilon) \leq \mathbf{E}[r(X_{t_n}, X_t)^\alpha]/\varepsilon^\alpha \xrightarrow{n\to\infty} 0$ for each $\varepsilon > 0$, i.e. $X_{t_n} \xrightarrow{n\to\infty}_p X_t$. Furthermore, due to the continuity of $\mathcal{Y}$, we find $Y_{t_n} \xrightarrow{n\to\infty}_{fs} Y_t$. In particular, $\mathbf{P}(X_t = Y_t) = 1$. This completes the proof. □

13.2 Example 1: The Poisson process

For the first time, we consider a concrete stochastic process with process with index set $I = [0, \infty)$. A path of the Poisson process is shown in Figure 2.

Remark 13.9 (Modeling by a Poisson process)**.** *We want to model clicks of a Geiger counter, calls to a call-center, mutation events along ancestral lines, or something else which has events randomly occurring in time. We want to analyze such counting processes with the help of a stochastic process $\mathcal{X} = (X_t)_{t\in I}$ with $I = [0, \infty)$. Let X_t be the number of clicks/calls/mutations up to time t. For such a process it makes sense to make a few assumptions:*

1. Independent increments: *If $0 = t_0 < t_1 < ... < t_n$, then $(X_{t_i} - X_{t_{i-1}} : i = 1, ..., n)$ is an independent family.*
2. Identically distributed increments: *If $0 < t_1 < t_2$, then $X_{t_2} - X_{t_1} \stackrel{d}{=} X_{t_2 - t_1} - X_0$.*
3. No double-points: *It is $\limsup_{\varepsilon\to 0} \frac{1}{\varepsilon}\mathbf{P}(X_\varepsilon - X_0 > 1) = 0$.*

Definition 13.10 (Poisson process)**.** *A real-valued stochastic process $\mathcal{X} = (X_t)_{t\in[0,\infty)}$ with $X_0 = 0$ is called a Poisson process with intensity λ if the following applies:*

1. *For $0 = t_0 < ... < t_n$, the family $(X_{t_i} - X_{t_{i-1}} : i = 1, ..., n)$ is independent.*
2. *For $0 \leq t_1 < t_2$ is $X_{t_2} - X_{t_1} \sim Poi(\lambda(t_2 - t_1))$.*

Proposition 13.11 (Existence of Poisson processes)**.** *Let $\lambda \geq 0$. Then there is exactly one distribution $\mathbf{P}_I$ on $(\mathcal{B}(\mathbb{R}))^I$ such that the canonical process with respect to $\mathbf{P}_I$ is a Poisson process with intensity λ.*

Proof. As for uniqueness: The finite-dimensional distributions of $\mathbf{P}_I$ as given by 1. and 2. from Definition 13.10 are uniquely defined. Therefore the uniqueness follows from Proposition 13.6.1.

For existence, we define the Poisson process as a projective limit. For $J = \{t_1, ..., t_n\} \subseteq_f I$ with $0 = t_0 < t_1 < ... < t_n$ we set for $x_0 = 0$

$$S^n : (x_1 - x_0, ..., x_n - x_{n-1}) \mapsto (x_1, ..., x_n).$$

Further,

$$\mathbf{P}_J := S^n_* \bigotimes_{i=1}^n \mathrm{Poi}(\lambda(t_i - t_{i-1})). \tag{13.3}$$

In other words: If $Y_{t_i - t_{i-1}}$ for $i = 1, ..., n$ are independently Poisson distributed with parameter $\lambda(t_i - t_{i-1})$, then $S^n(Y_{(t_1 - t_0)}, ..., Y_{t_n - t_{n-1}})) \sim \mathbf{P}_J$.

We now show that the family $(\mathbf{P}_J : J \subseteq_f I)$ is projective: let $J = \{t_1, ..., t_n\}$ as above and $H = J \setminus \{t_i\}$ for one i. Then,

$$\text{Poi}(\lambda(t_{i+1} - t_i)) * \text{Poi}(\lambda(t_i - t_{i-1})) = \text{Poi}(\lambda(t_{i+1} - t_{i-1}))$$

and therefore

$$(\pi_H^J)_* \mathbf{P}_J = (\pi_H^J \circ S^n)_* \bigotimes_{j=1}^n \text{Poi}(\lambda(t_j - t_{j-1})) = \mathbf{P}_H.$$

According to Theorem 5.24, there is the projective limit $\mathbf{P}_I$. Let us consider the canonical process $\mathcal{X} = (X_t)_{t \in I}$ with respect to $\mathbf{P}_I$. It has the finite-dimensional distributions $(\mathbf{P}_J : J \subseteq_f I)$. In particular, because of (13.3) increments are independent and Poisson distributed. Thus $\mathcal{X}$ fulfills the conditions 1. and 2. from Definition 13.10. $\square$

Proposition 13.12 (Characterization of Poisson processes)**.** *A non-decreasing stochastic process $\mathcal{X} = (X_t)_{t \in I}$ with $X_0 = 0$ and values in $\mathbb{Z}_+$ is a Poisson process with intensity λ iff $\lambda = \mathbf{E}[X_1 - X_0] < \infty$ and 1.-3. from Remark 13.9 are fulfilled.*

Proof. '$\Rightarrow$': 1. and 2. from remark 13.9 are clearly fulfilled. For 3. we calculate directly

$$\tfrac{1}{\varepsilon}\mathbf{P}(X_\varepsilon > 1) = \frac{1 - e^{-\lambda\varepsilon}(1 + \lambda\varepsilon)}{\varepsilon} \leq \frac{1 - (1 - \lambda\varepsilon)(1 + \lambda\varepsilon)}{\varepsilon} \xrightarrow{\varepsilon \to 0} 0.$$

'$\Leftarrow$': 1. from Definition 13.10 is fulfilled. It remains to show that $X_t \sim \text{Poi}(\lambda t)$. Let for $n \in \mathbb{N}, k = 1, ..., n$,

$$Z_k^n := (X_{tk/n} - X_{t(k-1)/n}) \wedge 1, \qquad X_t^n = \sum_{k=1}^n Z_k^n.$$

This means that Z_k^n indicates whether in the interval $(t(k-1)/n; tk/n]$ at least one jump has taken place. Then, since X_t^n is monotonic in n,

$$\begin{aligned}
\mathbf{P}(\lim_{n \to \infty} X_t^n \neq X_t) &= \lim_{n \to \infty} \mathbf{P}(X_t^n \neq X_t) \\
&= \lim_{n \to \infty} \mathbf{P}(X_{tk/n} - X_{t(k-1)/n} > 1 \text{ for a } k) \\
&\leq \lim_{n \to \infty} \sum_{k=1}^n \mathbf{P}(X_{tk/n} - X_{t(k-1)/n} > 1) \\
&= \lim_{n \to \infty} n\mathbf{P}(X_{t/n} > 1) \xrightarrow{n \to \infty} 0
\end{aligned}$$

from 3. Further, X_t^n is binomially distributed with n and probability of success $p_n := \mathbf{P}(X_{t/n} > 0)$. Because of the linearity of the mapping $t \mapsto \mathbf{E}[X_t]$ and, since $X_t^n \uparrow X_t$, it follows from the theorem on monotone convergence,

$$\lambda t = \mathbf{E}[X_t] = \lim_{n \to \infty} \mathbf{E}[X_t^n] = \lim_{n \to \infty} np_n.$$

By a Poisson approximation (see Example 10.1),

$$\begin{aligned}\mathbf{P}(X_t = k) &= \lim_{n\to\infty} \mathbf{P}(X_t^n = k) - \mathbf{P}(X_t^n = k; X_t \neq X_t^n) + \mathbf{P}(X_t = k; X_t \neq X_t^n) \\ &= \lim_{n\to\infty} \mathbf{P}(X_t^n = k) = \text{Poi}(\lambda t)(k),\end{aligned}$$

i.e. $X_t \sim \text{Poi}(\lambda t)$ and the assertion follows. □

Proposition 13.13 (Construction by exponential distributions)**.** *Let $S_1, S_2, ...$ be independent, exponentially distributed with parameter λ. Further, let $\mathcal{X} = (X_t)_{t\in I}$ be given by*

$$X_t := \max\{i : S_1 + ... + S_i < t\}$$

with $\max \emptyset = 0$. Then $\mathcal{X}$ is a Poisson process with intensity λ.

Proof. We must show that for $0 = t_0 < ... < t_n$, $k_1, ..., k_n \in \mathbb{N}$

$$\mathbf{P}(X_{t_1} - X_{t_0} = k_1, ..., X_{t_n} - X_{t_{n-1}} = k_n) = \prod_{j=1}^n \text{Poi}(\lambda(t_j - t_{j-1}))(k_j).$$

This will only be calculated for the case $n = 2$, the general case follows analogously. In the following calculation, let $0 \leq s < t$ and $U_1, U_2, ...$ uniformly distributed random variables on $[0, t]$. We calculate

$$\begin{aligned}&\mathbf{P}(X_s - X_0 = k, X_t - X_s = \ell) \\ &\quad= \int_0^s \int_{s_1}^s \cdots \int_{s_{k-1}}^s \int_s^t \int_{s_{k+1}}^t \cdots \int_{s_{k+\ell-1}}^t \int_t^\infty \lambda^{k+\ell+1} e^{-\lambda s_1} e^{-\lambda(s_2 - s_1)} \cdots \\ &\qquad\qquad \cdots e^{-\lambda(s_{k+\ell+1} - s_{k+\ell})} ds_{k+\ell+1} ... ds_1 \\ &\quad= \lambda^{k+\ell} \int_0^s \int_{s_1}^s \cdots \int_{s_{k-1}}^s \int_s^t \int_{s_{k+1}}^t \cdots \int_{s_{k+\ell-1}}^t \Big(\int_t^\infty \lambda e^{-\lambda s_{k+\ell+1}} ds_{k+\ell+1} \Big) ds_{k+\ell} \cdots ds_1 \\ &\quad= e^{-\lambda t} \lambda^{k+\ell} t^{k+\ell} \mathbf{P}[U_1 < ... < U_k < s < U_{k+1} < ... < U_{k+\ell}] \\ &\quad= e^{-\lambda t} \lambda^\ell t^\ell \binom{k+\ell}{k} \Big(\frac{s}{t}\Big)^k \Big(\frac{t-s}{t}\Big)^\ell \frac{1}{(k+\ell)!} \\ &\quad= e^{-\lambda s} \frac{(\lambda s)^k}{k!} \cdot e^{-\lambda(t-s)} \frac{(\lambda(t-s))^\ell}{\ell!},\end{aligned}$$

and the assertion follows. □

Example 13.14 (Left- and right-continuous Poisson process)**.** *Let, similar to Proposition 13.13, the stochastic process process $\mathcal{Y} = (Y_t)_{t\in I}$ given by*

$$Y_t := \max\{i : S_1 + ... + S_i \leq t\}.$$

Paths of the processes $\mathcal{X}$ from proposition 13.13 and $\mathcal{Y}$ can be seen in Figure 2. The two processes differ in that $\mathcal{X}$ is right-continuous and $\mathcal{Y}$ is left-continuous. However, both processes are Poisson processes with intensity λ, as you can easily see. This is because $\mathbf{P}(X_t = Y_t) = 1$ applies for all $t \in [0, \infty)$ and thus $\mathcal{Y}$ is a version of $\mathcal{X}$ according to Proposition 13.6. As you can see from this example, two processes with the same distribution can have completely different paths.

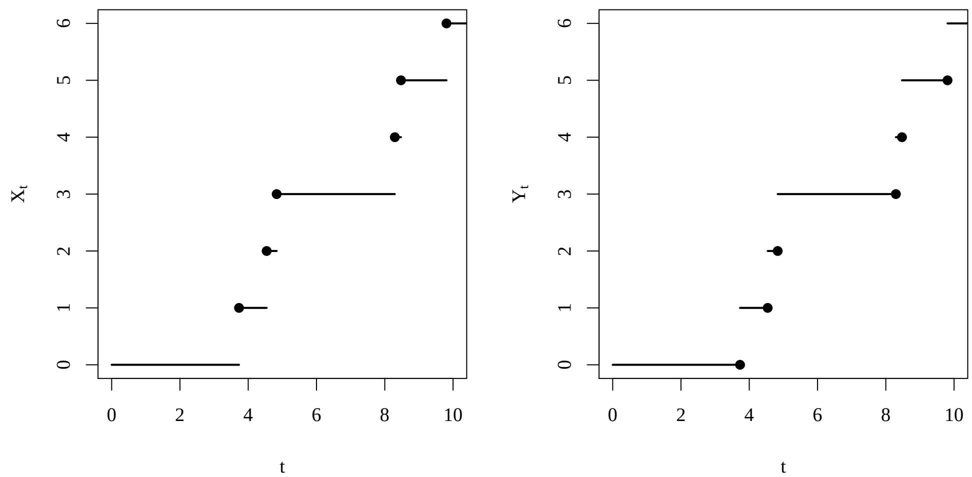

Figure 2: The right-continuous Poisson process $\mathcal{X}$ and the left-continuous Poisson process $\mathcal{Y}$.

13.3 Example 2: Brownian motion

Brownian motion is named after the botanist Robert Brown who observed in a microscope how pollen appears to move under thermal fluctuations seem to move erratically. We will give a mathematical definition for this process, that will be particularly important in stochastic analysis. Moreover, the normal distribution will play an important role in this process. A path of a one-dimensional Brownian motion can be found in Figure 3.

This section only serves to introduce Brownian motion. We will learn more about properties of Brownian motion later.

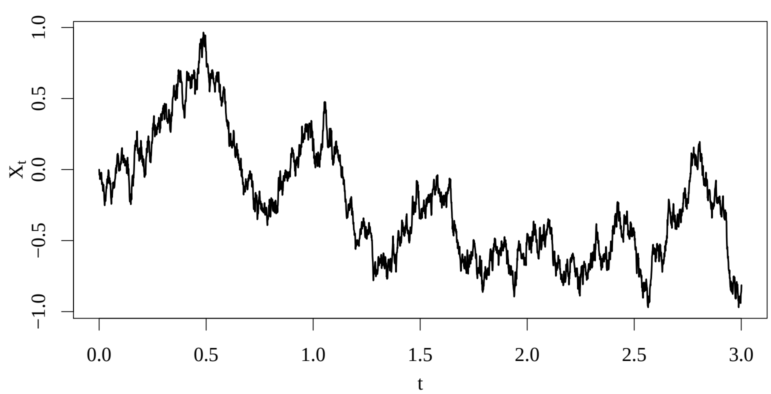

Figure 3: A path of a Brownian motion.

Definition 13.15 (Brownian motion and Gaussian processes)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be a stochastic process with values in $\mathbb{R}$.*

1. *The process $\mathcal{X}$ is called* Gaussian *if $c_1 X_{t_1} + \cdots + c_n X_{t_n}$ for each choice of $c_1, ..., c_n \in \mathbb{R}$ and $t_1, ..., t_n \in I$ is normally distributed. For a Gaussian process, $t \mapsto \mathbb{E}[X_t]$ denotes its*

expectation and $(s,t) \mapsto \mathbf{COV}(X_s, X_t)$ *its covariance structure.*

2. *If* $I = [0, \infty)$, *then* $\mathcal{X}$ *is called a* Brownian motion *with start in* x, *if the process has continuous paths and if for each choice of* $0 = t_0 < t_1 < \cdots < t_n$ *it holds that* $X_{t_0} = x$ *and* $X_{t_i} - X_{t_{i-1}}$ *are independently distributed according to* $N(0, t_i - t_{i-1})$, $i = 1, ..., n$. *If* $x = 0$, *then* $\mathcal{X}$ *is also called* standardized *or also* Wiener process.

3. *Let* $\mathcal{X}^1 = (X_t^1)_{t\in[0,\infty)}, ..., \mathcal{X}^d = (X_t^d)_{t\in[0,\infty)}$ *Brownian motion. Then the* $\mathbb{R}^d$*-valued process* $\mathcal{X} = (X_t)_{t\in[0,\infty)}$ *with* $X_t = (X_t^1, ..., X_t^d)$ *is called a*d-dimensional Brownian motion*.*

Remark 13.16 (Continuity of Brownian motion)**.** *According to Theorem 5.24 it is clear that there is a process whose increments are normally distributed as specified in Definition 13.15.2. The specified distributions* $(X_{t_1}, ..., X_{t_n})_{n\in\mathbb{N}, t_1,...,t_n \in I}$ *are in fact a projective family. For example, if* $X_{t_i} - X_{t_{i-1}} \sim N(0, t_i - t_{i-1})$ *and* $X_{t_{i-1}} - X_{t_{i-2}} \sim N(0, t_{i-1} - t_{i-2})$, *then* $X_{t_i} - X_{t_{i-2}} = X_{t_i} - X_{t_{i-1}} + X_{t_{i-1}} - X_{t_{i-2}} \sim N(0, t_i - t_{i-2})$ *because of example 5.20. However, it is less clear whether there is there is also a process with such increments that has continuous paths has. To check this, we use the criterion from Theorem 13.8.*

Proposition 13.17 (Existence of Brownian motion)**.** *Let* $\mathcal{X} = (X_t)_{t\in[0,\infty)}$ *be a real-valued stochastic process such that for any choice of* $0 = t_0 < t_1 < ... < t_n$ *it holds that* $X_{t_0} = x$ *and* $X_{t_i} - X_{t_{i-1}}$ *are independent and are distributed according to* $N(0, t_i - t_{i-1})$, $i = 1, ..., n$. *Then there exists a modification* $\mathcal{Y}$ *of* $\mathcal{X}$ *with continuous paths. In other words,* $\mathcal{Y}$ *is a Brownian motion. The process* $\mathcal{Y}$ *is even locally Hölder continuous for every parameter* $\gamma < 1/2$. *Furthermore, the covariance structure of Brownian motion* $\mathcal{Y}$ *is given by* $\mathbf{COV}(X_s, X_t) = s \wedge t$.

Proof. Wlog let $x = 0$. The existence and uniqueness of a process with independent normally distributed increments follows as in the proof of Proposition 13.11. Since $X_s \sim N(0, s)$, $X_s \stackrel{d}{=} s^{1/2} X_1$, as can be seen, for example, from Example 6.13.3. For $a > 2$,

$$\mathbf{E}[|X_t - X_s|^a] = \mathbf{E}[|X_{t-s}|^a] = \mathbf{E}[((t-s)^{1/2}|X_1|)^a] = (t-s)^{a/2}\mathbf{E}[|X_1|^a].$$

According to Theorem 13.8, there is therefore a modification of $\mathcal{X}$ with continuous paths. According to the above calculation, these are Hölder-continuous for each parameter $\gamma \in (0, ((a/2) - 1)/a) = (0, (a-2)/(2a))$. Since a was arbitrary, the Hölder continuity follows for each $\gamma \in (0, 1/2)$.

To determine the covariance structure of $\mathcal{X}$, we calculate for $0 \leq s \leq t$

$$\mathbf{COV}(X_s, X_t) = \mathbf{COV}(X_s, X_s) + \mathbf{COV}(X_s, X_t - X_s) = \mathbf{V}[X_s] = s.$$

An analogous calculation for $t < s$ provides the result $\mathbf{COV}(X_s, X_t) = s \wedge t$. □

Lemma 13.18 (Characterization of Gaussian processes)**.**
Let $\mathcal{X} = (X_t)_{t\in[0,\infty)}$ *and* $\mathcal{Y} = (Y_t)_{t\in[0,\infty)}$ *be Gaussian processes with* $\mathbf{E}[X_t] = \mathbf{E}[Y_t]$ *and* $\mathbf{COV}(X_s, X_t) = \mathbf{COV}(Y_s, Y_t)$. *Then* $\mathcal{Y}$ *is a version of* $\mathcal{X}$ *(and vice versa).*

Proof. Let $n \in \mathbb{N}$ and $c_1, ..., c_n \in \mathbb{R}$ be arbitrary. Then, for each choice of $t_1, ..., t_n \in I$ both $Z_X := c_1 X_{t_1} + \cdots + c_n X_{t_n}$ as well as $Z_Y := c_1 Y_{t_1} + \cdots c_n Y_{t_n}$ are normally distributed. Furthermore, according to the assumption,

$$\mathbf{E}[Z_X] = c_1\mathbf{E}[X_{t_1}] + \cdots + c_n\mathbf{E}[X_{t_n}] = c_1\mathbf{E}[Y_{t_1}] + \cdots + c_n\mathbf{E}[Y_{t_n}] = \mathbf{E}[Z_Y]$$

and

$$\mathbf{V}(Z_X) = \sum_{i,j=1}^n c_i c_j \mathbf{COV}(X_{t_i}, X_{t_j}) = \sum_{i,j=1}^n c_i c_j \mathbf{COV}(Y_{t_i}, Y_{t_j}) = \mathbf{V}(Z_Y).$$

This means that $Z_X \stackrel{d}{=} Z_Y$. Since $c_1, ..., c_n$ were arbitrary, it follows from Proposition 10.17 that $(X_{t_1}, ..., X_{t_n}) \stackrel{d}{=} (Y_{t_1}, ..., Y_{t_n})$. With Proposition 13.6.1 the statement follows. $\square$

Theorem 13.19 (Brownian scaling)**.** *Let $\mathcal{X} = (X_t)_{t\in[0,\infty)}$ be a Brownian motion. Then, the processes $(X_{c^2t}/c)_{t\in[0,\infty)}$ are for each $c > 0$ and $(tX_{1/t})_{t\in[0,\infty)}$ also Brownian motions.*

Proof. It is clear that both $(X_{c^2t}/c)_{t\in[0,\infty)}$ and $(tX_{1/t})_{t\in[0,\infty)}$ are Gaussian processes. Furthermore,

$$\begin{aligned} \mathbf{E}[X_{c^2t}/c] &= 0, \\ \mathbf{E}[tX_{1/t}] &= 0, \end{aligned}$$

and for $s, t \geq 0$

$$\begin{aligned} \mathbf{COV}[X_{c^2s}/c, X_{c^2t}/c] &= \frac{1}{c^2}(c^2 s \wedge c^2 t) = s \wedge t, \\ \mathbf{COV}[sX_{1/s}, tX_{1/t}] &= st\Big(\frac{1}{s} \wedge \frac{1}{t}\Big) = s \wedge t. \end{aligned}$$

Now the assertion follows with Lemma 13.18. $\square$

13.4 Filtrations and stopping times

In a stochastic process, more and more of the underlying random variables of the underlying random variables are realized as time goes by. This means that more and more information about the path of the process becomes process becomes visible. Now information is synonymous with the measurability with respect to a σ-algebra, as can be seen from the lecture *Probability Theory*. Since the information grows over over time, a stochastic process involves an increasing family of σ-algebras, which we will call a filtration in the in the following.

Definition 13.20 (Filtrations, stopping times)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be an E-valued stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$.*

1. *A family $(\mathcal{F}_t)_{t\in I}$ of σ-algebras with $\mathcal{F}_t \subseteq \mathcal{F}, t \in I$, is called* filtration *if $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s \leq t$.*

2. *The filtration $\mathcal{F} = (\mathcal{F}_t)_{t\in I}$ with $\mathcal{F}_t = \sigma(X_s : s \leq t)$ is called the filtration generated by $\mathcal{X}$.*

3. *The stochastic process $(X_t)_{t\in I}$ is called* adapted *to the filtration $(\mathcal{F}_t)_{t\in I}$ if X_t is a $\mathcal{F}_t/\mathcal{B}(E)$-measurable random variable for all $t \in I$.*

Now let $\mathcal{F} = (\mathcal{F}_t)_{t\in I}$ be a filtration.

5. *A random time is a random variable with values in $\bar{I}$ (the completion of I). A random time T is called* $((\mathcal{F}_t)_{t\in I}$-)stopping time *if $\{T \leq t\} \in \mathcal{F}_t$ for all $t \in I$. If $I = [0, \infty)$, then a random time T is called* $((\mathcal{F}_t)_{t\in I}$-)option time *if $\{T < t\} \in \mathcal{F}_t$ for all $t \in I$. (In the case $I = \{0, 1, 2, ...\}$ we do not need this term.)*

6. *Each stopping time T defines the σ-algebra*

$$\mathcal{F}_T := \{A \in \mathcal{A} : A \cap \{T \leq t\} \in \mathcal{F}_t, t \in I\}$$

of the T-past.

7. *For a random time T, X_T is defined by $\omega \mapsto X_{T(\omega)}(\omega)$. Further, $\mathcal{X}^T := (X_{T\wedge t})_{t\in I}$ is the process stopped at T.*

Remark 13.21 (Interpretation of the definition of stopping times)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be a stochastic process and $(\mathcal{F}_t)_{t\in I}$ the canonical filtration. One can view $\mathcal{F}_t$ as the information that is available at time t through knowledge of $(X_s)_{0\leq s\leq t}$. If T is a stopping time, then $\{T \leq t\} \in \mathcal{F}_t$. So the occurrence of the event $\{T \leq t\}$ can be predicted by knowledge of $(X_s)_{s\leq t}$. In other words by knowing the stochastic process up to time t it can be decided whether the stopping time T has occurred by now at the latest. If T is an option time, then by knowing the stochastic process up to time t, it can be decided whether the stopping time T has already occurred in the past of t.*

Example 13.22 (Hitting times in the Poisson process)**.** *Let $\mathcal{X} = (X_t)_{t\in[0,\infty)}$ and $\mathcal{Y} = (Y_t)_{t\in[0,\infty)}$ the right and left continuous Poisson process from example 13.14, respectively, and $(\mathcal{F}_t^{\mathcal{X}})_{t\in[0,\infty)}$ and $(\mathcal{F}_t^{\mathcal{Y}})_{t\in 0,\infty)}$ the corresponding filtrations. Further, let*

$$T_1 := \inf\{t \geq 0 : X_t = 1\} = \inf\{t \geq 0 : Y_t = 1\}$$

be the hitting time *of* 1*. (The last equality holds because the processes $\mathcal{X}$ and $\mathcal{Y}$ jump from 0 to 1 at the same time). Then:*

- *T_1 is both $(\mathcal{F}_t^{\mathcal{X}})_{t\in[0,\infty)}$-stopping time, as well as a $(\mathcal{F}_t^{\mathcal{X}})_{t\in[0,\infty)}$ option time.*

Indeed: If $T_1 = t$ is the jump time from 0 to 1, then $X_t = 1$, i.e. $\{T_1 \leq t\} = \{X_t \geq 1\} \in \sigma((X_s)_{s\leq t}) = \mathcal{F}_t^{\mathcal{X}}$ and $\{T_1 < t\} = \{X_{t-} \geq 1\} \in \sigma((X_s)_{s<t}) \subseteq \mathcal{F}_t^{\mathcal{X}}$.

- *T_1 is indeed an $(\mathcal{F}_t^{\mathcal{Y}})_{t\in[0,\infty)}$ option time, but not a $(\mathcal{F}_t^{\mathcal{Y}})_{t\in[0,\infty)}$ stopping time.*

Indeed: If $T_1 = t$ is the jump time from 0 to 1, then $Y_t = 0$, but $Y_{t+} = 1$, i.e. $\{T_1 \leq t\} = \{X_{t+} \geq 1\} \in \sigma((Y_s)_{s\leq t+h})$ for every $h > 0$, but not $\{T_1 \leq t\} \in \mathcal{F}_t^{\mathcal{Y}}$. However, still $\{T_1 < t\} = \{Y_t \geq 1\} \in \sigma((Y_s)_{s\leq t}) \subseteq \mathcal{F}_t^{\mathcal{Y}}$.

Lemma 13.23 (Simple properties of stopping times)**.** *Let $(\mathcal{F}_t)_{t\in I}$ be a filtration.*

1. *Each time $T = s \in I$ is a stopping time*

2. *For stopping times S, T, the times $S \wedge T$ and $S \vee T$ are also stopping times.*

3. *For stopping times $S, T \geq 0$, $S + T$ is a stopping time.*

4. *Each stopping time T is $\mathcal{F}_T$ measurable.*

5. *For stopping times S, T with $S \le T$ is $\mathcal{F}_S \subseteq \mathcal{F}_T$.*

Proof. 1. for $t \in I$ is $\{s \le t\} \in \{\emptyset, \Omega\} \subseteq \mathcal{F}_t$, i.e. $T = s$ is a stopping time.
2. for $t \in I$ is $\{S \wedge T \le t\} = \{S \le t\} \cup \{T \le t\} \in \mathcal{F}_t$ and $\{S \vee T \le t\} = \{S \le t\} \cap \{T \le t\} \in \mathcal{F}_t$.
3. Let $t \in I$. There are $S \wedge t$ and $T \wedge t$ stopping times, i.e. for $s \le t$ is $\{S \wedge t \le s\} \in \mathcal{F}_s \subseteq \mathcal{F}_t$. For $s > t$, $\{S \wedge t \le s\} = \Omega \in \mathcal{F}_t$, i.e. $S \wedge t$ is $\mathcal{F}_t$-measurable. Analogously, it follows that $T \wedge t$ is $\mathcal{F}_t$-measurable. Furthermore, $1_{\{S>t\}}, 1_{\{T>t\}}$ $\mathcal{F}_t$-measurable. If we set $S' = S \wedge t + 1_{\{S>t\}}, T' = T \wedge t + 1_{\{T>t\}}$, then $S' + T'$ is $\mathcal{F}_t$-measurable and $\{S + T \le t\} = \{S' + T' \le t\} \in \mathcal{F}_t$.
4. Since T is a stopping time, $\{T \le t\} \in \mathcal{F}_t$. According to the definition of $\mathcal{F}_T$ this means $\{T \le t\} \in \mathcal{F}_T$. Since $\mathcal{H} := \{(-\infty; t] : t \in \mathbb{R}\}$ is a generator of $\mathcal{B}(\mathbb{R})$, so the assertion follows.
5. Let $A \in \mathcal{F}_S$ and $t \in I$. Since $B := A \cap \{S \le t\} \in \mathcal{F}_t$,

$$A \cap \{T \le t\} = B \cap \{T \le t\} \in \mathcal{F}_t,$$

i.e. $A \in \mathcal{F}_T$. $\square$

Definition 13.24 (Continuous and complete filtration)**.** 1. *Let $(\mathcal{F}_t)_{t\in[0,\infty)}$ be a filtration. We define $(\mathcal{F}_t^+)_{t\in[0,\infty)}$ by $\mathcal{F}_t^+ := \bigcap_{s>t} \mathcal{F}_s$. Further, $(\mathcal{F}_t)_{t\in[0,\infty)}$ is* continuous *if $\mathcal{F}_t^+ = \mathcal{F}_t$.*

2. *Let $\mathcal{N} = \{A :$ there exists a $N \supseteq A$ with $N \in \mathcal{F}$ and $\mathbf{P}(N) = 0\}$. Then, the filtration $(\mathcal{F}_t)_{t\in I}$ is called* complete *if $\mathcal{N} \subseteq \mathcal{F}_t$ for each $t \in I$.*

Lemma 13.25 (Usual completion of a filtration)**.** *Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, $(\mathcal{F}_t)_{t\in[0,\infty)}$ a filtration and $\mathcal{N}$ as in Definition 13.24. Then there is a smallest continuous and complete filtration $(\mathcal{G}_t)_{t\in[0,\infty)}$ with $\mathcal{F}_t \subseteq \mathcal{G}_t, t \in [0,\infty)$. This is given by*

$$\mathcal{G}_t = \sigma(\mathcal{F}_t^+, \mathcal{N}).$$

Furthermore, $\sigma(\mathcal{F}_t^+, \mathcal{N}) = \sigma(\mathcal{F}_t, \mathcal{N})^+$.

Proof. First we show the last equation. It is clear that

$$\sigma(\mathcal{F}_t^+, \mathcal{N}) \subseteq \sigma(\sigma(\mathcal{F}_t, \mathcal{N})^+, \mathcal{N}) = \sigma(\mathcal{F}_t, \mathcal{N})^+.$$

Conversely, let $A \in \sigma(\mathcal{F}_t, \mathcal{N})^+$. Then, $A \in \sigma(\mathcal{F}_{t+h}, \mathcal{N})$ for all $h > 0$. So there is an $A_h \in \mathcal{F}_{t+h}$ with $\mathbf{P}((A \setminus A_h) \cup (A_h \setminus A)) = 0$. Now choose $h_1, h_2, ...$ with $h_n \downarrow 0$ and

$$A' = \{A_{h_n} \text{ infinitely often}\}.$$

Then, obviously, $A' \in \mathcal{F}_t^+$ and $\mathbf{P}((A \setminus A') \cup (A' \setminus A)) = 0$, i.e. $A \in \sigma(\mathcal{F}_t^+, \mathcal{N})$. From this follows $\sigma(\mathcal{F}_t, \mathcal{N})^+ \subseteq \sigma(\mathcal{F}_t^+, \mathcal{N})$.

To prove the minimality of $(\mathcal{G}_t)_{t\in[0,\infty)}$ let $(\mathcal{H}_t)_{t\in[0,\infty)}$ be another right-continuous complete extension of $(\mathcal{F}_t)_{t\in[0,\infty)}$. Then,

$$\mathcal{G}_t = \sigma(\mathcal{F}_t^+, \mathcal{N}) \subseteq \sigma(\mathcal{H}_t, \mathcal{N}) = \mathcal{H}_t$$

for all $t \in [0, \infty)$. $\square$

Lemma 13.26 (Option and stopping times)**.** *Let $(\mathcal{F}_t)_{t\in[0,\infty)}$ be a filtration. A random time T is an $(\mathcal{F}_t)_{t\in[0,\infty)}$ option time iff T is a $(\mathcal{F}_t^+)_{t\in[0,\infty)}$-stopping time. In this case,*

$$\mathcal{F}_T^+ = \{A \in \mathcal{F} : A \cap \{T < t\} \in \mathcal{F}_t, t > 0\}.$$

In particular, if $(\mathcal{F}_t)_{t\in[0,\infty)}$ is continuous, then every random time is a $(\mathcal{F}_t)_{t\in[0,\infty)}$-stopping time if it is a $(\mathcal{F}_t)_{t\in[0,\infty)}$ option time.

Proof. First,

$$\{T \le t\} = \bigcap_{\mathbb{Q} \ni s > t} \{T < s\}, \qquad \{T < t\} = \bigcup_{\mathbb{Q} \ni s < t} \{T \le s\}.$$

If T is a $(\mathcal{F}_t^+)_{t \in [0,\infty)}$-stopping time and $A \cap \{T \le t\} \in \mathcal{F}_t^+$. Then,

$$A \cap \{T < t\} = \bigcup_{\mathbb{Q} \ni s < t} (A \cap \{T \le s\}) \in \mathcal{F}_t.$$

If, on the other hand, $A \cap \{T < t\} \in \mathcal{F}_t$, then

$$A \cap \{T \le t\} = \bigcap_{h > 0} \bigcap_{t < s < t+h} (A \cap \{T < s\}) \in \bigcap_{h>0} \mathcal{F}_{t+h} = \mathcal{F}_t^+.$$

If you set $A = \Omega$ in the last two equations, the first assertion follows. For general A the second also follows. □

Lemma 13.27 (Suprema and infima of stopping times)**.** *Let $T_1, T_2, ...$ be random times and $(\mathcal{F}_t)_{t \in I}$ a filtration. Then the following applies:*

1. *If $T_1, T_2, ...$ are stopping times, then $T := \sup_n T_n$ is also a stopping time.*
2. *If $I = \{0, 1, 2, ...\}$ and $T_1, T_2, ...$ are stopping times, then $T := \inf_n T_n$ is also a stopping time.*
3. *If $I = [0, \infty)$ and $T_1, T_2, ...$ are option times, then $T := \inf_n T_n$ is also an option time. In addition, $\mathcal{F}_T^+ = \bigcap_n \mathcal{F}_{T_n}^+$.*

Proof. 1. We have $\{T \le t\} = \bigcap_n \{T_n \le t\} \in \mathcal{F}_t$ and the assertion follows.
2. It holds $\{T \le t\} = \bigcup_n \{T_n \le t\} \in \mathcal{F}_t$, from which the assertion follows.
3. Here, $\{T < t\} = \bigcup_n \{T_n < t\} \in \mathcal{F}_t$. Since $T \le T_n$, $\mathcal{F}_T^+ \subseteq \bigcap_n \mathcal{F}_{T_n}^+$ according to Lemma 13.23.5. If, on the other hand, $A \in \bigcap_n \mathcal{F}_{T_n}^+$, then

$$A \cap \{T < t\} = A \cap \bigcup_n \{T_n < t\} = \bigcup_n (A \cap \{T_n < t\}) \in \mathcal{F}_t.$$

Thus $A \in \mathcal{F}_T^+$. □

Proposition 13.28 (Approximation by countable stopping times)**.** *If $I = [0, \infty)$, each option time T can be replaced by a sequence of stopping times $T_1, T_2, ...$, such that T_n only assumes values in a discrete (in particular countable) quantity and $T_n \downarrow T$.*

Proof. We define $T_n = 2^{-n}[2^n T + 1]$. Then $T_1, T_2, ...$ is a sequence decreasing towards T, where T_n only contains the values $\{1, 2, ...\} \cdot 2^{-n}$, $n = 1, 2, ...$ Further, $\{T_n \le k2^{-n}\} = \{T < k2^{-n}\} \in \mathcal{F}_{k2^{-n}}$, so T_n is a stopping time, $n = 1, 2, ...$ □

Definition 13.29 (Hitting time)**.** *Let $B \in \mathcal{B}(E)$. Then the* hitting time of B *is given by*

$$T_B := \inf\{t : X_t \in B\}.$$

To find out whether the hitting time T_B is a stopping (or option) time, the following result is important.

Proposition 13.30 (Hitting times as option and stopping times)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be an E-valued process that is adapted with respect to a filtration $(\mathcal{F}_t)_{t\in I}$. Then, for $B \in \mathcal{B}(E)$:*

1. *If $I = \{0, 1, 2, ...\}$, then the time T_B is a stopping time.*
2. *If $I = [0, \infty)$, B is open and $\mathcal{X}$ has right-continuous paths, then T_B is an option time.*
3. *If $I = [0, \infty)$, B is closed and $\mathcal{X}$ has continuous paths, then T_B is a stopping time.*

Proof. 1. Here,

$$\{T_B \le t\} = \bigcup_{s\le t} \{X_s \in B\} \in \mathcal{F}_t.$$

For 2. we write

$$\{T_B < t\} = \bigcup_{\mathbb{Q}\ni s<t} \{X_s \in B\} \in \mathcal{F}_t.$$

For 3. with $B_n := \{x \in E : r(x, B) < 1/n\}$

$$\{T_B \le t\} = \bigcap_n \{T_{B_n} \le t\} = \bigcap_n (\{T_{B_n} < t\} \cup \{X_t \in \overline{B_n}\}) \in \mathcal{F}_t.$$

This shows all assertions. □

13.5 Progressive measurability

By definition, for a stochastic process $\mathcal{X} = (X_t)_{t\in I}$, each of the variables X_t is measurable, $t \in I$. However, it is (still) unclear when exactly for a random time T the quantity $X_T : \omega \mapsto X_{T(\omega)}(\omega)$ is measurable and therefore a random variable. For this we need a stronger measurability concept for the process $\mathcal{X}$.

Definition 13.31 (Progressive measurability)**.** *Let $(\mathcal{F}_t)_{t\in I}$ be a filtration and $\mathcal{X} = (X_t)_{t\in I}$ a stochastic process adapted to it. Then $\mathcal{X}$ is called* progressively measurable with respect to $(\mathcal{F}_t)_{t\in I}$*, if for all $t \in I$ the mapping*

$$\begin{cases} I \cap [0, t] \times \Omega & \to E \\ (s, \omega) & \mapsto X_s(\omega) \end{cases}$$

is measurable with respect to $I \cap \mathcal{B}([0, t]) \otimes \mathcal{F}_t/\mathcal{B}(E)$.

Lemma 13.32 (Right-continuous paths and progressive measurability)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be a stochastic process adapted to the filtration $(\mathcal{F}_t)_{t\in I}$. If either I is countable, or $\mathcal{X}$ has right-continuous paths, then $\mathcal{X}$ is progressively measurable with respect to $(\mathcal{F}_t)_{t\in I}$.*

Proof. Let $t \in I$. We consider the mapping

$$Y : \begin{cases} I \cap [0, t] \times \Omega & \to E \\ (s, \omega) & \mapsto X_s(\omega). \end{cases}$$

First, let I be countable and $B \in \mathcal{B}(E)$. Then,

$$Y^{-1}(B) = \bigcup_{s\in I, s\le t} \{s\} \times X_s^{-1}(B) \in \mathcal{B}(I \cap [0, t]) \otimes \mathcal{F}_t.$$

Next, let I be uncountable and let $\mathcal{X}$ have right-continuous paths. Consider the processes $\mathcal{X}^n = (X_s^n)_{t\in I\cap[0,t]}, n = 1, 2, ...$ with $X_s^n := X_{(2^{-n}\lceil 2^n s\rceil)\wedge t}$ and the corresponding mappings Y_n. Due to the right continuity of the paths, $Y_n \xrightarrow{n\to\infty}_{as} Y$. Furthermore

$$\begin{aligned} Y_n^{-1}(B) = &\bigcup_{k:(k+1)2^{-n}\leq t} [k2^{-n}, (k+1)2^{-n}) \times X_{(k+1)2^{-n}}^{-1}(B) \cup [2^{-n}\lfloor 2^n t\rfloor, t] \times X_t^{-1}(B) \\ &\in \mathcal{B}([0,t]) \otimes \mathcal{F}_t. \end{aligned}$$

□

Proposition 13.33 (Measurability of X_T)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be adapted to the filtration $(\mathcal{F}_t)_{t\in I}$, progressively measurable, and T a $(\mathcal{F}_t)_{t\in I}$ stopping time. Then*

$$X_T : \begin{cases} \{T < \infty\} & \to E \\ \omega & \mapsto X_{T(\omega)}(\omega) \end{cases}$$

is measurable with respect to $\{T < \infty\} \cap \mathcal{F}_T/\mathcal{B}(E)$.

Proof. We have to show that $\{X_T \in B, T \leq t\} \in \mathcal{F}_t$ for $B \in \mathcal{B}(E)$ holds, $t \in I$. By definition of $\mathcal{F}_T$, it then holds that $\{X_T \in B\} \in \mathcal{F}_T$, from which the assertion follows. However, since $\{X_T \in B, T \leq t\} = \{X_{T\wedge t} \in B, T \leq t\}$, it suffices to show that $X_{T\wedge t}$ is measurable with respect to $\mathcal{F}_t$, $t \in I$. We can therefore wlog assume that $T \leq t$ applies. We write $X_T = Y_t \circ \psi$, where $\psi(\omega) := (T(\omega), \omega)$ is measurable with respect to $\mathcal{F}_t/(I\cap\mathcal{B}([0,t])\otimes\mathcal{F}_t)$ and $Y_t(s,\omega) = X_s(\omega)$ according to condition $I \cap \mathcal{B}([0,t]) \otimes \mathcal{F}_t/\mathcal{B}(E)$-measurable. The assertion now follows with Lemma 3.6.2. □

14 Martingales

We now begin to deal with a particular class of stochastic processes, martingales. They are often referred to as *fair games*. Simply put, a martingale is a real-valued stochastic process whose increments vanish on average.

14.1 Introduction

Throughout the section, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, (E, r) a complete and separable metric space and $I \subseteq \mathbb{R}$ an ordered index set (usually $I = \{0, 1, 2, ...\}$ or $I = \mathbb{R}_+$). In addition, let a filtration $(\mathcal{F}_t)_{t\in I}$ be given. Adaptedness of a stochastic process is always with respect to $(\mathcal{F}_t)_{t\in I}$.

Example 14.1 (A simple martingale)**.** *For a $\mathcal{F}$-measurable random variable X one can define a stochastic process, namely $\mathcal{X} = (X_t)_{t\in I}$ with*

$$X_t := \mathbf{E}[X|\mathcal{F}_t]. \tag{$*$}$$

Of course, because of Theorem 11.2.7,

$$\mathbf{E}[X_t|\mathcal{F}_s] = \mathbf{E}[\mathbf{E}[X|\mathcal{F}_t]|\mathcal{F}_s] = \mathbf{E}[X|\mathcal{F}_s] = X_s.$$

Stochastic processes $\mathcal{X}$ with this property will be called martingales. In Section 14.4, we will then (among other things) deal with when a martingale $(X_t)_{t\in I}$ is associated with a random variable X so that $(*)$ applies.

Definition 14.2 ((Sub-, Super-)martingale). *Let $\mathcal{X} = (X_t)_{t\in I}$ be an adapted, real-valued stochastic process with $\mathbf{E}[|X_t|] < \infty$, $t \in I$. Then $\mathcal{X}$ is called*

$$\begin{aligned}
&\textit{martingale, if } \mathbf{E}[X_t|\mathcal{F}_s] = X_s \textit{ for } s,t \in I,\ s < t,\\
&\textit{sub-martingale, if } \mathbf{E}[X_t|\mathcal{F}_s] \geq X_s \textit{ for } s,t \in I,\ s < t,\\
&\textit{super-martingale, if } \mathbf{E}[X_t|\mathcal{F}_s] \leq X_s \textit{ for } s,t \in I,\ s < t.
\end{aligned}$$

More precisely, we say that $\mathcal{X}$ is a $(\mathcal{F}_t)_{t\in I}$-(sub, super)-martingale.

Remark 14.3 (Martingale property with discrete index set). *If I is discrete, for example $I = \{0, 1, 2, ...\}$, then a real-valued stochastic process $\mathcal{X} = (X_t)_{t\in I}$ is a martingale iff $\mathbf{E}[|X_t|] < \infty$, $t \in I$ and $\mathbf{E}[X_t|\mathcal{F}_{t-1}] = X_{t-1}$ for all $t = 1, 2, ...$ Then, for $s, t \in I, s \leq t$,*

$$\mathbf{E}[X_t|\mathcal{F}_s] = \mathbf{E}[\cdots \mathbf{E}[\mathbf{E}[X_t|\mathcal{F}_{t-1}]|\mathcal{F}_{t-2}]\cdots \mathcal{F}_s] = X_s$$

according to theorem 11.2.7 The same holds to sub- and super martingales.

Example 14.4 (Sums and products of integrable random variables).

1. *Let $X_1, X_2, ...$ be a sequence of independent, integrable random variables with $\mathbf{E}[X_i] = 0, i = 1, 2, ...$ and $\mathcal{F}_t := \sigma(X_1, ..., X_t)$. Further, let $S_0 := 0$ and for $t = 1, 2, ...$*

$$S_t := \sum_{i=1}^{t} X_i.$$

   Then,

$$\mathbf{E}[S_t|\mathcal{F}_{t-1}] = \mathbf{E}[S_{t-1} + X_t|\mathcal{F}_{t-1}] = S_{t-1} + \mathbf{E}[X_t|\mathcal{F}_{t-1}] = S_{t-1} + \mathbf{E}[X_t] = S_{t-1},$$

   i.e. $(S_t)_{t=0,1,2,...}$ is a martingale.

   If $\mathbf{E}[X_i] \geq 0$ for all $i = 1, 2, ...,$ then $(S_t)_{t\geq 0}$ is a sub-martingale.

2. *Let $I = \{-1, -2, ...\}$ and $X_1, X_2, ...$ be integrable, independent, identically distributed random variables. Further, we set for $t \in I$*

$$S_t := \frac{1}{|t|} \sum_{i=1}^{|t|} X_i$$

   and $\mathcal{F}_t := \sigma(..., S_{t-1}, S_t)$. Then for $t \in I$,

$$\begin{aligned}
\mathbf{E}[S_t|\mathcal{F}_{t-1}] &= \mathbf{E}\Big[\frac{1}{|t|}\sum_{i=1}^{|t|} X_i \Big| S_{t-1}, S_{t-2}, ...\Big]\\
&= \frac{1}{|t|}\sum_{i=1}^{|t|} \mathbf{E}\Big[X_i\Big| \sum_{i=1}^{|t|+1} X_i\Big]\\
&= \mathbf{E}\Big[X_1\Big| \sum_{i=1}^{|t|+1} X_i\Big]\\
&= \frac{1}{|t-1|}\sum_{i=1}^{|t-1|} X_i\\
&= S_{t-1}
\end{aligned}$$

according to Example 11.9. Specifically,

$$\begin{aligned}
\mathbf{E}[X_1|\mathcal{F}_t] &= \mathbf{E}\Big[X_1\Big|\sum_{i=1}^{|t|} X_i\Big] \\
&= \frac{1}{|t|}\sum_{i=1}^{|t|} X_i \\
&= S_t.
\end{aligned}$$

3. *Let $I = \{0, 1, 2, ...\}$ and $X_1, X_2, ...$ be a sequence of independent, integrable random variables with $\mathbf{E}[X_i] = 1, i = 1, 2, ...$ and $\mathcal{F}_t := \sigma(X_1, ..., X_t)$. Further, $S_0 := 1$ and for $t = 1, 2, ...$*

$$S_t := \prod_{i=1}^{t} X_i.$$

Then, $S_1, S_2, ...$ are integrable and

$$\mathbf{E}[S_t|\mathcal{F}_{t-1}] = \mathbf{E}[S_{t-1}X_t|\mathcal{F}_{t-1}] = S_{t-1} \cdot \mathbf{E}[X_t|\mathcal{F}_{t-1}] = S_{t-1} \cdot \mathbf{E}[X_t] = S_{t-1},$$

i.e. $(S_t)_{t\in I}$ is a martingale.

If $\mathbf{E}[X_i] \geq 1$ for all $i = 1, 2, ...$, then $(S_t)_{t\in I}$ is a sub-martingale.

Example 14.5 (Branching processes in discrete time)**.** *We consider a simple model for a randomly evolving population. evolving population. Let $X_i^{(t)}$) be independent, $\{0, 1, 2, ...\}$-valued random variable and $\mu = \mathbf{E}[X_i^{(t)}]$. Here, $X_i^{(t)}$ stands for the number of offspring of the ith individual of generation t with $i, t = 0, 1, ..., t = 1, 2, ...$ Starting with $Z_0 = k$ we set*

$$Z_{t+1} = \sum_{i=1}^{Z_t} X_i^{(t)},$$

so $\mathcal{Z} = (Z_t)_{t=0,1,2,...}$ is the stochastic process of the total process of the total number of individuals. The distribution of $X_i^{(t)}$ is also called the offspring distribution.

The process $\mathcal{Z}$ a (non-negative) martingale (with respect to the filtration generated by $\mathcal{Z}$), iff $\mathbf{E}[X_i^{(t)}] = 1$, i.e. each individual has on average has one offspring. Then, for $t = 1, 2, ...$

$$\mathbf{E}[Z_{t+1} - Z_t|\mathcal{F}_t] = \mathbf{E}\Big[\sum_{i=1}^{Z_t} X_i^{(t)} - Z_t|\mathcal{F}_t\Big] = (\mu - 1)Z_t.$$

If $\mu > 1$, $\mathcal{Z}$ is a sub-martingale, and if $\mu < 1$, $\mathcal{Z}$ is a super-martingale. Also, we call[3].

$$\begin{aligned}
&\mathcal{Z} \text{ a critical branching process if } \mu = 1, \\
&\mathcal{Z} \text{ a super-critical branching process if } \mu > 1, \\
&\mathcal{Z} \text{ a sub-critical branching process if } \mu < 1.
\end{aligned}$$

[3] It may seem irritating that a *supercritical branching process* is a sub-martingale and a *subcritical branching process* is a super martingale

In general, $(Z_t/\mu^t)_{t=0,1,2,...}$ is a (non-negative) martingale, because just like in the last calculation,

$$\mathbf{E}[Z_{t+1} - \mu Z_t | \mathcal{F}_t] = \mu Z_t - \mu Z_t = 0.$$

It is also worth noting that $\mathbf{E}[Z_{t+1}|\mathcal{F}_t] = \mu Z_t$, from which one can recursively conclude that

$$\mathbf{E}[Z_t] = \mu^t.$$

□

Example 14.6 (Martingales derived from Markov chains)**.** *With (discrete-time) Markov chains, we have already described a fairly simple dependency structure between random variables. Let $I = \{0, 1, 2, ...\}$, E at most countable and $P = (p_{xy})_{x,y \in E}$ a stochastic matrix, i.e.*

$$p_{xy} \geq 0, \qquad \sum_{z \in E} p_{xz} = 1$$

for all $x, y \in E$. We find for $f : E \to \mathbb{R}$ bounded and all $s = 0, 1, 2, ...,$

$$\mathbf{E}[f(X_{s+1}) - f(X_s)|\mathcal{F}_s] = \mathbf{E}[f(X_{s+1}) - f(X_s)|X_s] = \sum_{x \in E} p_{X_s,y}(f(y) - f(X_s)).$$

Therefore, setting $\mathcal{M} = (M_t)_{t=0,1,2,...}$ with

$$M_t = f(X_t) - \sum_{s=1}^{t-1} \mathbf{E}[f(X_{s+1}) - f(X_s)|X_s],$$

we have

$$\mathbf{E}[M_t - M_{t-1}|\mathcal{F}_{t-1}] = \mathbf{E}[f(X_t) - f(X_{t-1})|\mathcal{F}_{t-1}] - \mathbf{E}[f(X_t) - f(X_{t-1})|X_{t-1}] = 0.$$

In other words, $\mathcal{M}$ is a martingale.

We conclude this section with a simple statement on how to obtain further sub-martingales from known (sub)-martingales.

Proposition 14.7 (Convex functions of martingales are sub-martingales)**.** *Let $\mathcal{X} = (X_t)_{t \in I}$ be a stochastic process and $\varphi : \mathbb{R} \to \mathbb{R}$ convex. If $\varphi(X) = (\varphi(X_t))_{t \in I}$ is integrable and one of the two conditions*

1. *$\mathcal{X}$ is a martingale*
2. *$\mathcal{X}$ is a sub-martingale and varphi is non-decreasing*

is satisfied, then $\varphi(\mathcal{X}) = (\varphi(X_t))_{t \in I}$ is a sub-martingale.

Proof. If $\mathcal{X}$ is a martingale, then $\varphi(X_s) = \varphi(\mathbf{E}[X_t|\mathcal{F}_s])$. If $\mathcal{X}$ is a sub-martingale and φ is non-decreasing, $\varphi(X_s) \leq \varphi(\mathbf{E}[X_t|\mathcal{F}_s])$. In both cases, for $s \leq t$ because of Jensen's inequality for conditional expectations, Proposition 11.4,

$$\varphi(X_s) \leq \varphi(\mathbf{E}[X_t|\mathcal{F}_s]) \leq \mathbf{E}[\varphi(X_t)|\mathcal{F}_s],$$

i.e. $\varphi(\mathcal{X})$ is a sub-martingale. □

14.2 The stochastic integral as a martingale

In this section, let always $I = \{0, 1, 2, ...\}$ (whereby all results can be transferred to a discrete index set $I = \{t_0, t_1, ...\}$ with $t_0 < t_1 < ...$). All concepts introduced here have an analogue for processes in continuous time. However, the statements are much more complex to formulate and prove in that case. Some of these analogous statements are first formulated in the lecture *Stochastic analysis.*

Definition 14.8 (Previsible process). *A stochastic process $\mathcal{X}$ is called $(\mathcal{F}_t)_{t\in I}$-*previsible *if $X_0 = 0$ and X_t is $\mathcal{F}_{t-1}$-measurable, $t = 1, 2, ...$*

Proposition 14.9 (Doob decomposition). *Let $I = \{0, 1, 2, ...\}$. Each adapted process $\mathcal{X} = (X_t)_{t\in I}$ has an almost surely unique decomposition $\mathcal{X} = \mathcal{M} + \mathcal{A}$, where $\mathcal{M}$ is a martingale and $\mathcal{A}$ is previsible. In particular, $\mathcal{X}$ is a sub-martingale iff $\mathcal{A}$ almost surely non-decreasing.*

Proof. Define the previsible process $\mathcal{A} = (A_t)_{t\in I}$ by

$$A_t = \sum_{s=1}^{t} \mathbf{E}[X_s - X_{s-1}|\mathcal{F}_{s-1}]. \tag{14.1}$$

Then $\mathcal{M} = \mathcal{X} - \mathcal{A}$ is a martingale, because

$$\mathbf{E}[M_t - M_{t-1}|\mathcal{F}_{t-1}] = \mathbf{E}[X_t - X_{t-1}|\mathcal{F}_{t-1}] - (A_t - A_{t-1}) = 0.$$

Now we come to the uniqueness of the representation. If $\mathcal{X} = \mathcal{M} + \mathcal{A}$ for a martingale $\mathcal{M}$ and a previsible process $\mathcal{A}$, then $A_t - A_{t-1} = \mathbf{E}[X_t - X_{t-1}|\mathcal{F}_{t-1}]$ for all $t = 1, 2, ...$, i.e. (14.1) is almost surely true. □

Definition 14.10 (Quadratic variation, increasing process). *Let $I = \{0, 1, 2, ...\}$ and $\mathcal{X} = (X_t)_{t\in I}$ be a square integrable martingale. The almost surely uniquely determined, previsible process $(\langle\mathcal{X}\rangle_t)_{t\in I}$, for which $(X_t^2 - \langle\mathcal{X}\rangle_t)_{t\in I}$ is a is a martingale, is the* quadratic variation process *(or also the* increasing process*) of $\mathcal{X}$.*

Proposition 14.11 (Increasing process and variance). *Let $I = \{0, 1, 2, ...\}$, $\mathcal{X} = (X_t)_{t\in I}$ be a martingale with quadratic variation process $\langle\mathcal{X}\rangle = (\langle\mathcal{X}\rangle_t)_{t\in I}$. Then*

$$\langle\mathcal{X}\rangle_t = \sum_{s=1}^{t} \mathbf{E}[X_s^2 - X_{s-1}^2|\mathcal{F}_{s-1}] = \sum_{s=1}^{t} \mathbf{E}[(X_s - X_{s-1})^2|\mathcal{F}_{s-1}]$$

and

$$\mathbf{E}[\langle X\rangle_t] = \mathbf{V}[X_t - X_0].$$

Proof. As in the proof of Proposition 14.9, the process $\langle\mathcal{X}\rangle$ using (14.1) can be written. This immediately results in the first equals sign. The second follows, since $\mathbf{E}[X_s X_{s-1}|\mathcal{F}_{s-1}] = X_{s-1}^2$. Further is

$$\mathbf{E}[\langle\mathcal{X}\rangle_t] = \sum_{s=1}^{t} \mathbf{E}[X_s^2 - X_{s-1}^2] = \mathbf{E}[X_t^2 - X_0^2] = \mathbf{E}[(X_t - X_0)^2] = \mathbf{V}[X_t - X_0].$$

□

Example 14.12 (Increasing processes). *1. Let $\mathcal{S} = (S_t)_{\in I}$ with $S_t = \sum_{s=1}^t X_s$ as in example 14.4.1 with square integrable random variables $X_1, X_2, ...$ Then, with Proposition 14.11*

$$\langle \mathcal{S} \rangle_t = \sum_{s=1}^t \mathbf{E}[X_s^2].$$

In particular, the quadratic variation process of $\mathcal{S}$ is deterministic.

2. *Let $\mathcal{S} = (S_t)_{\in I}$ with $S_t = \prod_{s=1}^t X_s$ as in Example 14.4.3 with square integrable random variables $X_1, X_2, ...$ Then*

$$\langle \mathcal{S} \rangle_t = \sum_{s=1}^t \mathbf{E}[(S_s - S_{s-1})^2 | \mathcal{F}_{s-1}] = \sum_{s=1}^t S_{s-1}^2 \mathbf{E}[(X_s - 1)^2 | \mathcal{F}_{s-1}] = \sum_{s=1}^t S_{s-1}^2 \mathbf{V}[X_s].$$

In particular, in this example the process $\langle \mathcal{S} \rangle$ is truly stochastic.

3. *Let $I = [0, \infty)$ and $(X_t)_{t \in I}$ be a Brownian motion. Even in continuous time, the increasing process $((\langle \mathcal{X} \rangle_t)_{t \in I}$ is defined such that $(X_t^2 - \langle \mathcal{X} \rangle_t)_{t \in I}$ is a martingale. According to Example 14.47, $\langle \mathcal{X} \rangle_t = t$ is a candidate for the increasing process of the Brownian motion. However, in continuous time it is more difficult to say what the equivalent of a* previsible *process should be.*

Definition 14.13 (Discrete stochastic integral). *Let $I = \{0, 1, 2, ...\}$ and $\mathcal{H} = (H_t)_{t \in I}, \mathcal{X} = (X_t)_{t \in I}$ be a stochastic processes with values in $\mathbb{R}$. If $\mathcal{X}$ is adapted and $\mathcal{H}$ is previsible, then we define the stochastic integral $\mathcal{H} \cdot \mathcal{X} = ((\mathcal{H} \cdot \mathcal{X})_t)_{t \in I}$ by*

$$(\mathcal{H} \cdot \mathcal{X})_t = \sum_{s=1}^t H_s (X_s - X_{s-1})$$

for all $t \in I$. If $\mathcal{X}$ is a martingale, then we call $\mathcal{H} \cdot \mathcal{X}$ a martingale transform of $\mathcal{X}$.

Proposition 14.14 (Stability of stochastic integrals). *Let $I = \{0, 1, 2, ...\}$ and $\mathcal{X} = (X_t)_{t \in I}$ be an adapted, real-valued process with $\mathbf{E}[|X_0|] < \infty$.*

1. *$\mathcal{X}$ is a martingale if and only if for each previsible process $\mathcal{H} = (H_t)_{t \in I}$, the stochastic integral $\mathcal{H} \cdot \mathcal{X}$ is a martingale.*

2. *$\mathcal{X}$ is a sub-martingale (super-martingale) if and only if for every previsible, non-negative process $\mathcal{H} = (H_t)_{t \in I}$ the stochastic integral $\mathcal{H} \cdot \mathcal{X}$ is a sub-martingale (super-martingale).*

Proof. 1. '$\Rightarrow$': We immediately write

$$\begin{aligned} \mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_{t+1} - (\mathcal{H} \cdot \mathcal{X})_t | \mathcal{F}_t] &= \mathbf{E}[H_{t+1}(X_{t+1} - X_t) | \mathcal{F}_t] \\ &= H_{t+1} \mathbf{E}[X_{t+1} - X_t | \mathcal{F}_t] \\ &= 0. \end{aligned}$$

'$\Leftarrow$': Let $t \in I$ and $H_s := 1_{\{s=t\}}$. Then $\mathcal{H} = (H_s)_{s \in I}$ is deterministic, in particular previsible. Since $(\mathcal{H} \cdot \mathcal{X})_{t-1} = 0$, it follows that

$$0 = \mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_t | \mathcal{F}_{t-1}] = \mathbf{E}[X_t - X_{t-1} | \mathcal{F}_{t-1}] = \mathbf{E}[X_t | \mathcal{F}_{t-1}] - X_{t-1}$$

From this, the assertion follows.
2. follows analogously. □

Example 14.15 (Quadratic variation for stochastic integrals)**.** *Let $I = \{0, 1, 2, ...\}$, $\mathcal{X} = (X_t)_{t\in I}$ a martingale and $\mathcal{H} = (H_t)_{t\in I}$ previsible. Then, because of Proposition 14.11,*

$$\begin{aligned}\langle \mathcal{H}\cdot\mathcal{X}\rangle_t &= \sum_{s=1}^t \mathbf{E}[((\mathcal{H}\cdot\mathcal{X})_s - (\mathcal{H}\cdot\mathcal{X})_{s-1})^2|\mathcal{F}_{s-1}] = \sum_{s=1}^t \mathbf{E}[H_s^2(X_s - X_{s-1})^2|\mathcal{F}_{s-1}]\\ &= \sum_{s=1}^t H_s^2\cdot\mathbf{E}[(X_s - X_{s-1})^2|\mathcal{F}_{s-1}].\end{aligned}$$

In particular,

$$\mathbf{V}[(\mathcal{H}\cdot\mathcal{X})_t] = \sum_{s=1}^t \mathbf{E}[H_s^2\cdot(X_s - X_{s-1})^2].$$

Example 14.16 (Payout for games)**.** *Martingale transforms can also be interpreted as payoffs of games. Given a random variable evolves according to the adapted process $\mathcal{X} = (X_t)_{t=0,1,2,...}$. If you bet before time t with a stake H_t (based on the experience gained from $X_0, ..., X_{t-1}$) on the change in the random variable $X_t - X_{t-1}$, then $(\mathcal{H}\cdot\mathcal{X})_t$ is the profit realized up to time time t. Given the underlying process $\mathcal{X}$ is a martingale, Proposition 14.14 shows that the profit realized $\mathcal{H}\cdot\mathcal{X}$ for each strategy $\mathcal{H}$ is a martingale. In particular, the expected profit is 0.*

As an example, consider the Petersburg paradox: a fair coin is tossed infinitely often. In each round, a player places a stake of any amount. If `heads` *comes up, he loses it, if* `tails` *comes up, the stake is doubled and paid out again. The paradox consists of the following strategy: starting with a stake of 1 on the first coin toss, the player doubles his stake with every failure. If the first success comes on the t-th toss, his previous stake is $\sum_{i=1}^t 2^{i-1} = 2^t - 1$. Since the last bet was 2^{t-1}, the player gets 2^t back, so he has certainly made a profit of 1 even though the game was fair.*

To analyze this game using martingales, let $X_1, X_2, ...$ be an independent, identically distributed sequence with $\mathbf{P}(X_1 = -1) = \mathbf{P}(X_1 = 1) = \frac{1}{2}$, and $S_0 = 0, S_t = \sum_{i=1}^t X_i$. Then $\mathcal{S} = (S_t)_{t=0,1,2,...}$ is a martingale. Further, let H_t be the stake in the tth game. Therefore,

$$(\mathcal{H}\cdot\mathcal{S})_t = \sum_{i=1}^t H_i(S_i - S_{i-1}) = \sum_{i=1}^t H_iX_i$$

is the profit after the tth game. Since with $\mathcal{S}$, $\mathcal{H}\cdot\mathcal{S}$ is also a martingale, we find

$$\lim_{t\to\infty}\mathbf{E}[(\mathcal{H}\cdot\mathcal{S})_t] = \mathbf{E}[(\mathcal{H}\cdot\mathcal{S})_1] = \mathbf{E}[X_1] = 0,$$

i.e. the mean profit after a long time is 0, independent of the strategy $\mathcal{H}$. Above we have the bet

$$H_t := 2^{t-1}1_{\{S_{t-1}=-(t-1)\}} \tag{14.2}$$

and show that for the gain $(\mathcal{H}\cdot\mathcal{S})_t \xrightarrow{t\to\infty}_{fs} 1$ holds.

How do we now evaluate the strategy (14.2)? Let T be the random time of the win, i.e. T is geometrically distributed with parameter $\frac{1}{2}$. In particular, T is almost surely finite. Then

$$\mathbf{E}\Big[\sum_{t=1}^\infty H_t\Big] = \sum_{k=1}^\infty \frac{1}{2^k}(2^k - 1) = \infty,$$

i.e. for the above strategy you may need a lot of capital.

14.3 Stopped martingales

Let $\mathcal{X} = (X_t)_{t\in I}$ be a stochastic process. A stopped stochastic process is given by $\mathcal{X}^T := (X_{T\wedge t})_{t\in I}$, where T is an I-valued random variable. The process $\mathcal{X}^T$ therefore stops when T is reached. Special random variables T are called stopping times, whose occurrence at time t can be decided by means of the σ-algebra $\mathcal{F}_t$. (Consider, for example, a player who plays a fair game and stops at a random time, e.g. when he has won or lost enough. In this section we will learn about the Optional Stopping Theorem, which states that stopped (at a stopping time) martingales are martingales again; see Theorem 14.19. The Optional Sampling Theorem specifies conditions for which the martingale property applies not only to fixed, but also at random stopping times; see Theorem 14.22.

We start by recalling some facts on random times; see also Definition 13.20.

Remark 14.17 (Stopping time).
1. *A random time is a random variable with values in $\bar{I}$ (the end of I). A random time T is called $((\mathcal{F}_t)_{t\in I}$-)*stopping time *if $\{T \le t\} \in \mathcal{F}_t$ for all $t \in I$.*

2. *Each stopping time T defines the σ-algebra*
$$\mathcal{F}_T := \{A \in \mathcal{A} : A \cap \{T \le t\} \in \mathcal{F}_t, t \in I\}$$
of the T-past.

3. *Let $B \in \mathcal{B}(E)$. Then the* hitting time of B *is defined as*
$$T_B := \inf\{t : X_t \in B\}.$$

4. *For a random time T, X_T is defined by $\omega \mapsto X_{T(\omega)}(\omega)$. Further, $\mathcal{X}^T := (X_{T\wedge t})_{t\in I}$ is the process stopped at T.*

Remark 14.18 (Interpretation and hitting times).
1. *Let $\mathcal{X} = (X_t)_{t\in I}$ be a stochastic process and $(\mathcal{F}_t)_{t\in I}$ the canonical filtration. $\mathcal{F}_t$ can be understood as the information that is available at time t through knowledge of $(X_s)_{0\le s\le t}$. If T is a stopping time, then $\{T \le t\} \in \mathcal{F}_t$. Therefore, the occurrence of the event $\{T \le t\}$ can be decided by knowing $(X_s)_{s\le t}$. In other words, by knowing the stochastic process up to time t, it is possible to decide whether the stopping time T has already occurred.*

2. *If I is at most countable and $B \in \mathcal{B}(E)$, then T_B is a stopping time. Indeed, we write*
$$\{T_B \le t\} = \bigcup_{s\le t} \underbrace{\{X_s \in B\}}_{\in \mathcal{F}_s \subseteq \mathcal{F}_t} \in \mathcal{F}_t.$$

Proposition 14.19 (Optional Stopping). *Let $I = \{0, 1, 2, ...\}$ and $\mathcal{X} = (X_t)_{t\in I}$ be a (sub-, super-) martingale and T a stopping time. Then $\mathcal{X}^T = (X_{T\wedge t})_{t\in I}$ is a (sub-, super-) martingale.*

Proof. We show the assertion only for the case that $\mathcal{X}$ is a sub-martingale. The other statements follows analogously. For a sub-martingale $\mathcal{X}$ and $\{T > t-1\} \in \mathcal{F}_t$,
$$\begin{aligned}\mathbf{E}[X_{T\wedge t} - X_{T\wedge(t-1)}|\mathcal{F}_{t-1}] &= \mathbf{E}[(X_t - X_{t-1})1_{\{T>t-1\}}|\mathcal{F}_{t-1}] \\ &= 1_{\{T>t-1\}}\mathbf{E}[X_t - X_{t-1}|\mathcal{F}_{t-1}] \ge 0,\end{aligned}$$
i.e. $\mathcal{X}^T$ is a sub-martingale. $\square$

Lemma 14.20 (Conditions on $\mathcal{F}_T$)**.** *Let $I = \{0, 1, 2, ...\}$, $\mathcal{X} = (X_t)_{t \in I}$ be a martingale and T a stopping time bounded by t. Then $X_T = \mathbf{E}[X_t|\mathcal{F}_T]$.*

Proof. According to the definition of the conditional expectation and since X_T is $\mathcal{F}_T$ measurable (see Proposition 13.33), we must show that $\mathbf{E}[X_t; A] = \mathbf{E}[X_T; A]$ for $A \in \mathcal{F}_T$. It is $\{T = s\} \cap A \in \mathcal{F}_s$ for $s \in I$, i.e.

$$\begin{aligned}
\mathbf{E}[X_T; A] &= \sum_{s=1}^{t} \mathbf{E}[X_s; \{T = s\} \cap A] \\
&= \sum_{s=1}^{t} \mathbf{E}[\mathbf{E}[X_t|\mathcal{F}_s]; \{T = s\} \cap A] \\
&= \sum_{s=1}^{t} \mathbf{E}[X_t; \{T = s\} \cap A] \\
&= \mathbf{E}[X_t; A].
\end{aligned}$$

□

Lemma 14.21 (Uniform integrability and stopping times)**.**
Let $I = \{0, 1, 2, ...\}$. A martingale $\mathcal{X} = (X_t)_{t \in I}$ is uniformly integrable if the family $\{X_T : T$ almost surely finite stopping time$\}$ is uniformly integrable.

Proof. '$\Leftarrow$': clear.
'$\Rightarrow$': According to Lemma 7.9 there is a convex function $f : \mathbb{R}_+ \to \mathbb{R}_+$ with $\frac{f(x)}{x} \xrightarrow{x \to \infty} \infty$ and $\sup_{t \in I} \mathbf{E}[f(|X_t|)] =: L < \infty$. If T is almost surely a finite stopping time, then according to Lemma 14.20 (applied to the almost surely finite stopping time $T \wedge t$) $\mathbf{E}[X_t|\mathcal{F}_{T \wedge t}] = X_{T \wedge t}$. Since $\{T \leq t\} \in \mathcal{F}_{T \wedge t}$, we find with Jensen's inequality

$$\begin{aligned}
\mathbf{E}[f(|X_T|), \{T \leq t\}] &= \mathbf{E}[f(|X_{T \wedge t}|), \{T \leq t\}] \\
&= \mathbf{E}[f(|\mathbf{E}[X_t|\mathcal{F}_{T \wedge t}]|), \{T \leq t\}] \\
&\leq \mathbf{E}[\mathbf{E}[f(|X_t|)|\mathcal{F}_{T \wedge t}], \{T \leq t\}] \\
&= \mathbf{E}[f(|X_t|), \{T \leq t\}] \leq L.
\end{aligned}$$

Thus $\mathbf{E}[f(|X_T|)] \leq L$, i.e. the assertion follows with lemma 7.9. □

In example 14.16, $\mathcal{H} \cdot \mathcal{S}$ was a martingale, T a stopping time and $\mathbf{E}[(\mathcal{H} \cdot \mathcal{S})_t] = 0 \neq 1 = (\mathcal{H} \cdot \mathcal{S})_T$. If T had been bounded, this inequality would not have been possible, as we now show.

Theorem 14.22 (Optional Sampling Theorem)**.** *Let $I = \{0, 1, 2, ...\}$, $S \leq T$ almost certainly finite stopping times and $\mathcal{X} = (X_t)_{t \in I}$ a sub-martingale. If either T is bounded or $\mathcal{X}$ is uniformly integrable, then X_T is is integrable and $X_S \leq \mathbf{E}[X_T|\mathcal{F}_S]$.*

Proof. We first carry out the proof in the case of a bounded stopping time T. Let $T \leq t$ be for a $t \in I$. We use the Doob decomposition $\mathcal{X} = \mathcal{M} + \mathcal{A}$ of $\mathcal{X}$ into the martingale $\mathcal{M}$ and the monotonically non-decreasing process $\mathcal{A}$. Then with Lemma 14.20 and $\mathcal{F}_S \subseteq \mathcal{F}_T$ according to theorem 11.2.7

$$\begin{aligned} X_S = M_S + A_S &= \mathbf{E}[M_t + A_S|\mathcal{F}_S] \\ &\leq \mathbf{E}[M_t + A_T|\mathcal{F}_S] \\ &= \mathbf{E}[\mathbf{E}[M_t|\mathcal{F}_T] + A_T|\mathcal{F}_S] \\ &= \mathbf{E}[M_T + A_T|\mathcal{F}_S] \\ &= \mathbf{E}[X_T|\mathcal{F}_S]. \end{aligned}$$

Now let T be unbounded and $\mathcal{X}$ be uniformly integrable. Let $\mathcal{X} = \mathcal{M} + \mathcal{A}$ be the Doob decomposition of $\mathcal{X}$ into the martingale $\mathcal{M}$ and the non-falling previsible process $\mathcal{A} \geq 0$ with $A_0 = 0$. Since

$$\mathbf{E}[|A_t|] = \mathbf{E}[A_t] = \mathbf{E}[X_t - X_0] \leq \mathbf{E}[|X_0|] + \sup_{s \in I} \mathbf{E}[|X_s|]$$

we find $A_t \uparrow A_\infty$ for an $A_\infty \geq 0$ with $\mathbf{E}[A_\infty] < \infty$. With Lemma 7.9 one can conclude that $\mathcal{M}$ is also uniformly integrable. We now apply the Optional Sampling Theorem to the bounded stopping times $S \wedge t$, $T \wedge t$ and $\mathcal{M}$. For $A \in \mathcal{F}_S$ is $\{S \leq t\} \cap A \in \mathcal{F}_{S \wedge t}$, therefore

$$\mathbf{E}[M_{T\wedge t}, \{S \leq t\} \cap A] = \mathbf{E}[\mathbf{E}[M_{T\wedge t}|\mathcal{F}_{S\wedge t}], \{S \leq t\} \cap A] = \mathbf{E}[M_{S\wedge t}, \{S \leq t\} \cap A].$$

Since according to Lemma 14.21 the set $\{M_{S\wedge t}, M_{T\wedge t} : t \in I\}$ is uniformly integrable, then by Theorem 7.11

$$\mathbf{E}[M_T, A] = \lim_{t\to\infty} \mathbf{E}[M_{T\wedge t}, \{S \leq t\} \cap A] = \lim_{t\to\infty} \mathbf{E}[M_{S\wedge t}, \{S \leq t\} \cap A] = \mathbf{E}[M_S, A],$$

i.e. $\mathbf{E}[M_T|\mathcal{F}_S] = M_S$. Furthermore,

$$\mathbf{E}[X_T|\mathcal{F}_S] = \mathbf{E}[M_T|\mathcal{F}_S] + A_S + \mathbf{E}[A_T - A_S|\mathcal{F}_S] \geq M_S + A_S = X_S.$$

□

The Optional Sampling Theorem offers a simple way of characterizing martingales.

Lemma 14.23 (Characterization of martingales)**.** *Let $I = \{0, 1, 2, ...\}$, and $\mathcal{X} = (X_t)_{t\in I}$ be an adapted stochastic process. Then $\mathcal{X}$ is a martingale iff $\mathbf{E}[X_S] = \mathbf{E}[X_T]$ for stopping times S, T that only take two values.*

Proof. '⇒': Clear according to the Optional Sampling Theorem.
'⇐': Let $s \leq t$, $A \in \mathcal{F}_s$ and $T = s1_A + t1_{A^c}$. Then T is a stopping time and

$$0 = \mathbf{E}[X_t - X_T] = \mathbf{E}[X_t] - \mathbf{E}[X_s, A] - \mathbf{E}[X_t, A^c] = \mathbf{E}[X_t - X_s, A].$$

Since A was arbitrary, it follows that $\mathbf{E}[X_t|\mathcal{F}_s] = X_s$, so $\mathcal{X}$ is a martingale. □

Example 14.24 (Wald's identities, ruin problem)**.** *1. Let $X_1, X_2, ... \in \mathcal{L}^1$ be independent with $\mu := \mathbf{E}[X_1] = \mathbf{E}[X_2] = ...$, and $S_t := \sum_{s=1}^{t} X_s$. Furthermore, let T be an almost certainly limited stopping time. Then the first Wald identity is*

$$\mathbf{E}[S_T] = \mathbf{E}[T]\mu.$$

Indeed: the process $\mathcal{M} = (M_t)_{t=0,1,2,...}$ *with* $M_0 = 0$, $M_t = S_t - t\mu$ *for* $t = 1, 2, ...$ *is a martingale, and according to the Optional Sampling Theorem*

$$0 = \mathbf{E}[M_T] = \mathbf{E}[S_T] - \mathbf{E}[T]\mu.$$

Furthermore, if $X_1, X_2, ... \in L^2$ *with* $\sigma^2 = \mathbf{V}[X_1] = \mathbf{V}[X_2] = ...$ *and* T *is independent of* $X_1, X_2, ...$, *then the second Wald identity is*

$$\mathbf{V}[S_T] = \mathbf{E}[T]\sigma^2 + \mathbf{V}[T]\mu^2.$$

Indeed: $(M_t^2 - \langle M\rangle_t)_{t=0,1,2,...}$ *is a martingale, and* $\langle M\rangle_t = t\sigma^2$ *according to Example 14.12, thus*

$$0 = \mathbf{E}[M_T^2 - \langle M\rangle_T] = \mathbf{E}[M_T^2] - \mathbf{E}[T]\sigma^2.$$

Furthermore, due to the independence of T *and* $X_1, X_2, ...$,

$$\mathbf{COV}[S_T, T] = \mathbf{E}[\mathbf{E}[X_1 + \cdots + X_T|T]T] - \mu\mathbf{E}[T]^2 = \mu\mathbf{V}[T],$$

as well as

$$\mathbf{E}[M_T^2] = \mathbf{V}[S_T - T\mu] = \mathbf{V}[S_T] + \mu^2\mathbf{V}[T] - 2\mu\mathbf{COV}[S_T, T] = \mathbf{V}[S_T] - \mu^2\mathbf{V}[T].$$

In both Wald identities, the condition that T *is bounded can be weakened.*

2. *Let* $k \in \mathbb{N}$ *and* $X_1, X_2, ...$ *be independent and identically distributed random variables with* $\mathbf{P}(X_1 = 1) = 1 - \mathbf{P}(X_1 = -1) = p := 1 - q$. *For* $N \in \mathbb{N}$ *with* $0 < k < N$ *let* $S_0 = k$ *and* $S_t = S_0 + \sum_{i=1}^t X_i$. *Further, let* $T := \inf\{t : S_t \in \{0, N\}\}$ *and* $p_k := \mathbf{P}(S_T = 0)$. *This means that you play a game, starting with k (money) units, until you are either ruined or have N units. In each step you win with probability p one unit and loses with probability* $q = 1 - p$ *one unit. Then the probability of being ruined (having 0 units) is given by* p_k.

   In the case $p = \frac{1}{2}$, $(S_t)_{t=0,1,2,...}$ *is a martingale, and thus according to the Optional Sampling Theorem*

   $$k = \mathbf{E}[S_T] = N(1 - \mathbf{P}(S_T = 0)),$$

   thus

   $$\mathbf{P}(S_T = 0) = \frac{N-k}{N}.$$

   A similar calculation allows the determination of p_k *for the case* $p \neq \frac{1}{2}$.

   We now calculate further using the optional sampling theorem for $p \neq \frac{1}{2}$

   $$p_k := \mathbf{P}(S_T = 0) = \frac{\left(\frac{q}{p}\right)^k - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N}. \tag{14.3}$$

   Indeed: the following applies

   $$\mathbf{E}\Big[\Big(\frac{q}{p}\Big)^{X_1}\Big] = \frac{q}{p}p + \frac{p}{q}q = 1$$

and thus $\mathcal{Y} = (Y_t)_{t=0,1,2,\dots}$*, is defined by* $Y_t := \left(\frac{q}{p}\right)^{S_t}$ *according to Example 14.4.3 a martingale. Since* T *is almost surely finite,* $Y_{T\wedge t}$ *is a martingale due to Proposition 14.19, which is bounded by* 1 *and* $\left(\frac{q}{p}\right)^N$*. Because of Theorem 14.22,*

$$\left(\frac{q}{p}\right)^k = \mathbf{E}[Y_0] = \mathbf{E}[Y_T] = p_k + (1-p_k)\left(\frac{q}{p}\right)^N,$$

from which (14.3) *follows.*

3. *Let's consider a fair coin toss. How long does it take until the pattern* `ZKZK` *occurs for the first time? (*`K` *and* `Z` *stand for heads and tails).*

   To calculate this, let's consider the following game: before the first coin toss, a player bets one euro on `Z`*. If she loses, she stops, if she wins, she bets two euros on* `K` *before the next toss. If she loses in the second throw, she stops; if she wins, she bets four euros on euros on* `Z`*. If she loses on the third throw, she stops, if she wins, she bets eight euros on* `K`*. So if she wins on the fourth throw, she has won a total of 15 euros. In all other cases, she loses one euro.*

   Let us now assume that before each coin toss a new player plays according to the above strategy. The game ends when the first player first time a player wins 15 euros.

   Let X_t *be the total winnings of all players up to time* t *and* T *is the time at which the game is stopped because for the first time the pattern* `ZKZK` *has occurred. Certainly,*

$$|X_t| \le 15 \cdot t, \qquad \mathbf{P}[T > 4t] \le \frac{15}{16}^t.$$

   This means that $(X_{t\wedge T} : t = 1,2,\dots)$ *has a cominating integrable random variable, so according to Example 7.8.2 it is uniformly integrable. This allows us to apply the optional stopping theorem, i.e.* $(X_{T\wedge t})_{t=1,2,\dots}$ *is a martingale.*

   It is certain that

$$X_T = 15 - 1 + 3 - 1 - (T-4)$$

   since the first $T-4$ *players, as well as players* $T-3$ *and* $T-1$ *had to accept a loss of one euro. Player* $T-2$ *currently has at time* T *a profit of three euros and player* $T-4$ *has won 15 euros. So,*

$$0 = \mathbf{E}[X_T] = \mathbf{E}[15 - 1 + 3 - 1 - (T-4)] = -\mathbf{E}[T] - 20,$$

   therefore $\mathbf{E}[T] = 20$*. It is interesting to note that it can be expected that, for example, the pattern* `ZZKK` *can already occur after* 16 *coin tosses using a similar calculation.*

14.4 Martingale convergence results

Again, $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space, I countable (here it is also allowed that I is dense in $[0,\infty)$) and $(\mathcal{F}_t)_{t\in I}$ is a filtration. We are familiar with convergence theorems, such as the strong law of of large numbers. Martingales converge under relatively weak conditions.

We start in Proposition 14.26 with Doob's inequalities. These make statements about the distribution of $\sup_{s\le t} X_s$ if $\mathcal{X} = (X_t)_{t\in I}$ is a (sub, super)-martingale.

Lemma 14.25 (maximum inequality)**.** *If I is at most countable and $\mathcal{X} = (X_t)_{t\in I}$ is a sub-martingale, then for $\lambda > 0$*

$$\lambda \mathbf{P}[\sup_{s\leq t} X_s \geq \lambda] \leq \mathbf{E}[X_t, \sup_{s\leq t} X_s \geq \lambda] \leq \mathbf{E}[|X_t|, \sup_{s\leq t} X_s \geq \lambda].$$

Proof. The second inequality is trivial. For the first one, we note that due to monotonic convergence (by choosing finer and finer index sets in index sets in $[0, t]$) it is sufficient to consider the discrete case, e.g. $I = \{0, 1, 2, ...\}$. We recall the definition of T_B from Definition 14.17, which is given after Remark 14.18.2 is a stopping time and set

$$T = t \wedge T_{[\lambda;\infty)}.$$

According to the Optional Sampling Theorem 14.22 is

$$\begin{aligned}\mathbf{E}[X_t] \geq \mathbf{E}[X_T] &= \mathbf{E}[X_T; \sup_{s\leq t} X_s \geq \lambda] + \mathbf{E}[X_T; \sup_{s\leq t} X_s < \lambda] \\ &\geq \lambda \mathbf{P}[\sup_{s\leq t} X_s \geq \lambda] + \mathbf{E}[X_t; \sup_{s\leq t} X_s < \lambda].\end{aligned}$$

Subtracting the last term gives the inequality. □

Proposition 14.26 (Doob's L^p inequality)**.** *Let I be at most countable and $\mathcal{X} = (X_t)_{t\in I}$ be a martingale or a positive sub-martingale.*

1. *For $p \geq 1$ and $\lambda > 0$ is*

$$\lambda^p \mathbf{P}[\sup_{s\leq t} |X_s| \geq \lambda] \leq \mathbf{E}[|X_t|^p].$$

2. *For $p > 1$ is*

$$\mathbf{E}[|X_t|^p] \leq \mathbf{E}[\sup_{s\leq t} |X_s|^p] \leq \Big(\frac{p}{p-1}\Big)^p \mathbf{E}[|X_t|^p].$$

Proof. Again, it suffices – due to monotonic convergence – to consider the case $I = \{0, 1, 2, ...\}$ to consider.

1 According to proposition 14.7, $(|X_t|^p)_{t\in I}$ is a sub-martingale and the assertion follows from Lemma 14.25.

2 The first inequality is clear. For the second inequality, note that according to Lemma 14.25 it holds that

$$\lambda \mathbf{P}\{\sup_{s\leq t} |X_s| \geq \lambda\} \leq \mathbf{E}[|X_s|; \sup_{s\leq t} |X_s| \geq \lambda].$$

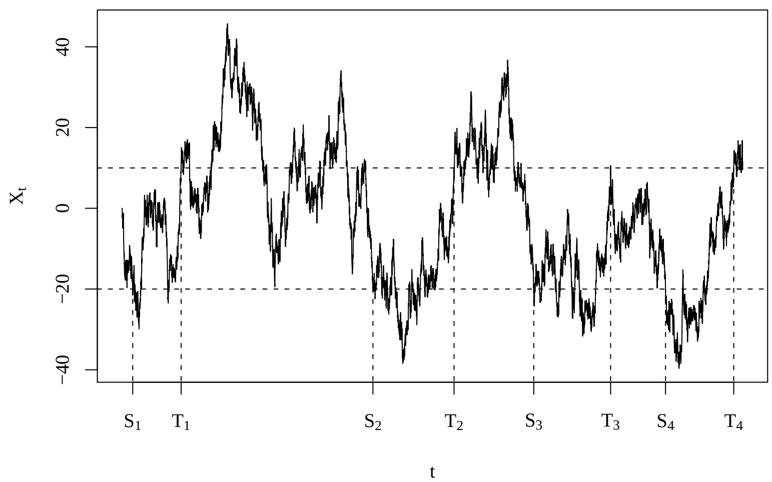

Figure 4:
An illustration of the stopping times $S_1, T_1, S_2, T_2, ...$ from definition 14.27

Therefore, for $K > 0$

$$\begin{aligned}
\mathbf{E}[\sup_{s\le t}(|X_s|\wedge K)^p] &= \mathbf{E}\Big[\int_0^{\sup_{s\le t}|X_s|\wedge K} p\lambda^{p-1}d\lambda\Big] \\
&= \mathbf{E}\Big[\int_0^K p\lambda^{p-1}1_{\{\lambda<\sup_{s\le t}|X_s|\}}d\lambda\Big] \\
&= \int_0^K p\lambda^{p-1}\mathbf{P}(\sup_{s\le t}|X_s|\ge\lambda)d\lambda \\
&\le \int_0^K p\lambda^{p-2}\mathbf{E}[|X_t|, \sup_{s\le t}|X_s|\ge\lambda]d\lambda \\
&= p\mathbf{E}\Big[|X_t|\int_0^{\sup_{s\le t}|X_s|\wedge K}\lambda^{p-2}d\lambda\Big] \\
&= \frac{p}{p-1}\mathbf{E}[|X_t|(\sup_{s\le t}|X_s|\wedge K)^{p-1}] \\
&\le \frac{p}{p-1}\mathbf{E}[\sup_{s\le t}(|X_s|\wedge K)^p]^{(p-1)/p}\cdot\mathbf{E}[|X_t|^p]^{1/p},
\end{aligned}$$

where we used the Hölder inequality in the last step. If you exponentiate both sides by p and then divide by $\mathbf{E}[\sup_{s\le t}(|X_s|\wedge K)^p]^{p-1}$, the it follows

$$\mathbf{E}[\sup_{s\le t}(|X_s|)^p] = \lim_{K\to\infty}\mathbf{E}[\sup_{s\le t}(|X_s|\wedge K)^p] \le \Big(\frac{p}{p-1}\Big)^p\mathbf{E}[|X_t|^p].$$

□

For the martingale convergence theorems, the upcrissong lemma 14.28 is central. Figure 4 illustrates the definition of an upcrossing.

Definition 14.27. *Let I be at most countable and $\mathcal{X} = (X_t)_{t\in I}$ a real-valued stochastic process. For $a < b$ an upcrossing is a piece of path $(X_r)_{s\leq r\leq s'}$ with $X_s \leq a$ and $X_{s'} \geq b$. To count the number of such upcrossings, we carry out stopping times $0 =: T_0 < S_1 < T_1 < S_2 < T_2 < ...$*

$$S_k := \inf\{t \geq T_{k-1} : X_t \leq a\},$$
$$T_k := \inf\{t \geq S_k : X_t \geq b\}$$

with $\inf \emptyset = \infty$. The k-th intersection between a and b is here between S_k and T_k. Further is

$$U^t_{a,b} := \sup\{k : T_k \leq t\}$$

is the number of crossings between a and b up to time t.

Lemma 14.28 (Upcrossing lemma)**.** *Let I be at most countable and $\mathcal{X} = (X_t)_{t\in I}$ a sub-martingale. Then*

$$\mathbf{E}[U^t_{a,b}] \leq \frac{\mathbf{E}[(X_t - a)^+]}{b-a}.$$

Proof. Again, we can assume – due to monotonic convergence – that $I = \{0, 1, 2, ...\}$. Since according to proposition 14.7 with $\mathcal{X}$ $((X_t - a)^+)_{t\in I}$ is also a sub-martingale and the upcrossings between a and b of $\mathcal{X}$ are the same as the upcrossings of $((X_t - a)^+)_{t\in I}$ between 0 and $b - a$, we can wlog assume that $\mathcal{X} \geq 0$ and $a = 0$. We define the process $\mathcal{H} = (H_t)_{t\in I}$ by

$$H_t := \sum_{k\geq 1} 1_{\{S_k<t\leq T_k\}},$$

i.e. $H_t = 1$ exactly when t lies in an upcrossing. Since

$$\{H_t = 1\} = \bigcup_{k\geq 1}\{S_k \leq t-1\} \cap \{T_k > t-1\},$$

H is previsible.

Given $T_k < \infty$ is obviously $X_{T_k} - X_{S_k} \geq b$. Further, in this case

$$(\mathcal{H} \cdot \mathcal{X})_{T_k} = \sum_{i=1}^{k} \sum_{s=S_i+1}^{T_i} (X_s - X_{s-1}) = \sum_{i=1}^{k}(X_{T_i} - X_{S_i}) \geq kb.$$

For $t \in \{T_k, ..., S_{k+1}\}$ is $(\mathcal{H} \cdot \mathcal{X})_t = (\mathcal{H} \cdot \mathcal{X})_{T_k}$ and for $t \in \{S_k + 1, ..., T_k\}$ is $(\mathcal{H} \cdot \mathcal{X})_t \geq (\mathcal{H} \cdot \mathcal{X})_{S_k} = (\mathcal{H} \cdot \mathcal{X})_{T_{k-1}}$. Therefore, $(\mathcal{H} \cdot \mathcal{X})_t \geq bU^t_{0,b}$. From Proposition 14.14 it follows that $((1 - \mathcal{H}) \cdot \mathcal{X})$ is a sub-martingale, in particular $\mathbf{E}[((1 - \mathcal{H}) \cdot \mathcal{X})_t] \geq 0$. With $X_t - X_0 = (1 \cdot \mathcal{X})_t = (\mathcal{H} \cdot \mathcal{X})_t + ((1 - \mathcal{H}) \cdot \mathcal{X})_t$ applies

$$\mathbf{E}[X_t] \geq \mathbf{E}[X_t - X_0] \geq \mathbf{E}[(\mathcal{H} \cdot \mathcal{X})_t] \geq b\mathbf{E}[U^t_{0,b}].$$

□

Theorem 14.29 (martingale convergence theorem for sub-martingales)**.** *Let $I \subseteq [0, \infty)$ be countable, $\sup I = u \in (0, \infty]$, $\mathcal{F}_u = \sigma(\bigcup_{t\in I} \mathcal{F}_t)$ and $\mathcal{X} = (X_t)_{t\in I}$ a sub-martingale with $\sup_{t\in I} \mathbf{E}[X^+_t] < \infty$. Then there is a null set N such that $\mathcal{X}$ converges outside of N along every ascending or descending sequence in I.*

In particular, if $I = \{0, 1, 2, ...\}$, $\mathcal{X}$ is a sub-martingale with $\sup_{t\in I} \mathbf{E}[X^+_t] < \infty$, then there exists a $\mathcal{F}_\infty$-measurable, integrable random variable X_∞ and $X_t \xrightarrow{t\to\infty}_{fs} X_\infty$.

Proof. Because of Lemma 14.28, $\mathbf{P}(U^t_{a,b} < \infty) = 1$ for all a, b, t. Therefore

$$N := \bigcup_{\substack{a<b \\ a,b\in\mathbb{Q}}} \{\sup_{t\in I} U^t_{a,b} = \infty\}$$

is a null set. Assuming that there is an ascending or descending sequence $t_1, t_2, ... \in I$ exists such that $\mathbf{P}(\liminf_{n\to\infty} X_{t_n} < \limsup_{n\to\infty} X_{t_n}) > 0$. For $a, b \in \mathbb{Q}$ let

$$B(a, b) := \{\liminf_{n\to\infty} X_{t_n} < a < b < \limsup_{n\to\infty} X_{t_n}\}.$$

Since $\{\liminf_{n\to\infty} X_{t_n} < \limsup_{n\to\infty} X_{t_n}\} = \bigcup_{a,b\in\mathbb{Q}} B(a, b)$, there exist $a, b \in \mathbb{Q}$ with $\mathbf{P}(B(a, b)) > 0$. However, $\sup_t U^t_{a,b} = \infty$ applies to $B(a, b)$ in contradiction to the fact that N is a null set. Thus follows the almost sure convergence along every ascending or descending sequence.

Now let $I = \{0, 1, 2, ...\}$. Since all X_t are $\mathcal{F}_\infty$-measurable, X_∞ is also $\mathcal{F}_\infty$-measurable. It remains to show that X_∞ is integrable. According to Fatou's Lemma,

$$\mathbf{E}[X^+_\infty] \le \sup_{t\in I} \mathbf{E}[X^+_t] < \infty.$$

Moreover, since $\mathcal{X}$ is a sub-martingale, again using Fatou's lemma,

$$\mathbf{E}[X^-_\infty] \le \liminf_{t\to\infty} \mathbf{E}[X^-_t] = \liminf_{t\to\infty} \big(\mathbf{E}[X^+_t] - \mathbf{E}[X_t]\big) \le \sup_{t\in I} \mathbf{E}[X^+_t] - \mathbf{E}[X_0] < \infty.$$

□

Corollary 14.30 (martingale convergence theorem for positive super martingales)**.** *Let $I \subseteq [0, \infty)$ be at most countable, $\sup I = u \in (0, \infty]$, $\mathcal{F}_u = \sigma(\bigcup_{t\in I} \mathcal{F}_t)$ and $\mathcal{X} = (X_t)_{t\in I}$ a non-negative super martingale. Then there exists a $\mathcal{F}_u$-measurable, integrable random variable X_u with $\mathbf{E}[X_u] \le \mathbf{E}[X_0]$ and $X_t \xrightarrow{t\to u}_{fs} X_u$.*

Proof. Theorem 14.29, applied to the sub-martingale $-\mathcal{X}$ provides the almost sure limit. With the Lemma of Fatou also

$$\mathbf{E}[X_u] \le \liminf_{t\to u} \mathbf{E}[X_t] \le \mathbf{E}[X_0].$$

□

Example 14.31 (Convergence of branching processes)**.** *Let us consider a critical or sub-critical branching process $\mathcal{Z} = (Z_t)_{t=0,1,2,...}$ from Example 14.5 (where the offspring distribution is not degenerate, i.e. $X^{(t)}_i = 1$ is not almost certain). These are non-negative super-martingales, so they must converge according to Corollary 14.30 almost surely against a random variable Z_∞. In this case, $\mathbf{P}(Z_\infty > 0) = 0$ must apply, otherwise the almost sure convergence is violated. (A population with a positive number of individuals has a positive probability of changing its size in one generation.) Therefore,*

$$Z_t \xrightarrow{t\to\infty} Z_\infty := 0$$

is almost certain.

In the case of the critical branching process, it is important to realize that $(Z_t)_{t=0,1,2,...,\infty}$ is not a martingale, because $\mathbf{E}[Z_\infty|\mathcal{F}_t] = \mathbf{E}[0|\mathcal{F}_t] \ne Z_t$ applies with positive probability.

If $\mathcal{Z}$ is supercritical, then $(Z_t/\mu^t)_{t=0,1,2,...}$ is a non-negative martingale that also converges almost surely according to the above corollary.

Theorem 14.32 (Convergence theorem for uniformly integrable martingales)**.** *Let I be countable with $\sup I = u \in (0, \infty]$, $\mathcal{F}_u = \sigma(\bigcup_{t\in I} \mathcal{F}_t)$ and $\mathcal{X} = (X_t)_{t\in I}$ a (super, sub)-martingale. Then the following statements are equivalent:*

1. *$\mathcal{X}$ is uniformly integrable.*
2. *There exists a $\mathcal{F}_u$-measurable random variable X_u such that $(X_t)_{t\in I\cup u}$ is a (super, sub)martingale.*
3. *There exists a $\mathcal{F}_u$-measurable random variable X_u with $X_t \xrightarrow{t\to u}_{fs,L^1} X_u$.*

Proof. 2. $\Rightarrow$ 1 follows directly from Lemma 11.5.
1. $\Rightarrow$ 3. By Lemma 7.9, $\sup_{t\in I} \mathbf{E}[|X_t|] < \infty$. The almost certain convergence follows from theorem 14.29 and the L^1-convergence thus from theorem 7.11.
3. $\Rightarrow$ 2.: As for the proof that $(X_t)_{t\in I\cup\{u\}}$ is a (super, sub)-martingale, we only give the argument for sub-martingales, i.e. $\mathbf{E}[\mathbf{E}[X_u|\mathcal{F}_s]; A] \geq \mathbf{E}[X_s; A]$ for $A \in \mathcal{F}_s$ and $s \in I$. Because of the L^1 convergence according to Theorem 11.2.3, $\mathbf{E}[|\mathbf{E}[X_t|\mathcal{F}_s] - \mathbf{E}[X_u|\mathcal{F}_s]|] \xrightarrow{t\to u} 0$ and thus for $A \in \mathcal{F}_s$, so

$$\mathbf{E}[\mathbf{E}[X_u|\mathcal{F}_s]; A] = \lim_{t\to\infty} \mathbf{E}[\mathbf{E}[X_t|\mathcal{F}_s]; A] \geq \mathbf{E}[X_s; A],$$

i.e. $\mathbf{E}[X_u|\mathcal{F}_s] \geq X_s$ almost surely. □

Theorem 14.33 (Martingale convergence theorem for L^p-bounded martingales)**.** *Let I be countable with $\sup I = u \in [0, \infty)$, $\mathcal{F}_u = \sigma(\bigcup_{t\in I} \mathcal{F}_t)$, $p > 1$ and $\mathcal{X} = (X_t)_{t\in I}$ an L^p-bounded martingale. Then there is a $\mathcal{F}_u$-measurable random variable X_u with $\mathbf{E}[|X_u|^p] < \infty$, $X_t \xrightarrow{t\uparrow u}_{fs,L^p} X_u$. Furthermore, $(|X_t|^p)_{t\in I}$ is uniformly integrable.*

Proof. Because of Lemma 7.9, $\mathcal{X}$ is uniformly integrable. According to Theorem 14.32 there is thus the limit X_u with $X_t \xrightarrow{t\uparrow u}_{fs,L^1} X_u$. According to Doob's inequality from Proposition 14.26, for $t \in I$

$$\mathbf{E}[\sup_{t\in I} |X_t|^p] = \lim_{t\uparrow u} \mathbf{E}[\sup_{s\leq t} |X_s|^p] \leq \lim_{t\uparrow u} \Big(\frac{p}{p-1}\Big)^p \mathbf{E}[|X_t|^p] < \infty.$$

Thus $(|X_t|^p)_{t\in I}$ is uniformly integrable according to Example 7.8.3 According to Fatou's Lemma and Lemma 7.9, $\mathbf{E}[|X_u|^p] \leq \sup_{t\in I} \mathbf{E}[|X_t|^p] < \infty$ and Theorem 7.11 provides the convergence in L^p. □

Example 14.34 (Branching process)**.** *Let $\mathcal{Z}$ be a branching process as in Example 14.5 and Example 14.31 with $Z_0 = k$. The quadratic variation of $\mathcal{Y} = (Y_t)_{t=0,1,2,\dots}$, given $Y_t = Z_t/\mu^t$ is according to Proposition 14.11 is given as*

$$\begin{aligned}
\langle \mathcal{Y} \rangle_t &= \sum_{s=1}^{t} \frac{1}{\mu^{2s}} \mathbf{E}\Big[\Big(\sum_{i=1}^{Z_{s-1}} X_i^{(s-1)} - \mu Z_{s-1}\Big)^2 |\mathcal{F}_{s-1}\Big] \\
&= \sum_{s=1}^{t} \frac{1}{\mu^{2s}} \mathbf{V}\Big[\sum_{i=1}^{Z_{s-1}} X_i^{(s-1)} | Z_{s-1}\Big] \\
&= \sum_{s=1}^{t} \frac{1}{\mu^{2s}} Z_{s-1} \cdot \mathbf{V}[X_1^{(1)}].
\end{aligned}$$

In particular, the offspring distribution has a second moment, therefore $\mathbf{V}[X_1^{(1)}] =: \sigma^2 < \infty$, *so*

$$\mathbf{V}[Y_t] = \sum_{s=1}^{t} \frac{1}{\mu^{2s}} \mathbf{E}[Z_s] \cdot \sigma^2 = k\sigma^2 \sum_{s=1}^{t} \frac{1}{\mu^s}.$$

If $\mu \leq 1$, *then* $\mathcal{Y}$ *is not* L^2*-bounded, but for* $\mu > 1$ $\sup_{t=0,1,2,...} \mathbf{V}[Y_t] < \infty$. *This means that there is a* $\mathcal{F}_\infty$*-measurable, square-integrable random variable* Y_∞, *so that* $(Y_t)_{t=0,1,2,...,\infty}$ *is a martingale.*

Example 14.35 (product of random variables)**.** *Let* $I = \{1, 2, ...\}$, $X_1, X_2, ...$ *be non-negative, independent, integrable random variable with* $\mathbf{E}[X_t] = 1, t \in I$ *and* $S_t := \prod_{s=1}^t X_s$ *according to Example 14.4.2 a martingale. According to the corollary 14.30 there is thus a* S_∞, *so that* $S_t \xrightarrow{t\to\infty}_{fs} S_\infty$. *Define*

$$a_t := \mathbf{E}[\sqrt{X_t}].$$

We now show:

$$\{S_t : t \in I\} \text{ uniformly integrable} \qquad \iff \qquad \prod_{t=1}^{\infty} a_t > 0.$$

In particular, then also $S_t \xrightarrow{t\to\infty}_{L^1} S_\infty$. *In the proof we set for* $t = 1, 2, ..$

$$W_t := \prod_{s=1}^{t} \frac{\sqrt{X_s}}{a_s}.$$

This means that $(W_t)_{t=1,2,...}$ *is a martingale. Here, too, it follows that there is a* W_∞ *with* $W_t \xrightarrow{t\to\infty}_{fs} W_\infty$.

'$\Leftarrow$*': Because of Jensen's inequality* $a_t^2 = (\mathbf{E}[\sqrt{X_t}])^2 \leq \mathbf{E}[X_t] = 1$, *thus* $a_t \leq 1$. *The following applies*

$$\sup_{t\in I} \mathbf{E}[W_t^2] = \sup_{t\in I} \mathbf{E}\Big[\prod_{s=1}^{t} \frac{X_s}{a_s^2}\Big] = \sup_{t\in I} \prod_{s=1}^{t} \frac{\mathbf{E}[X_s]}{a_s^2} \leq \frac{1}{\Big(\prod_{s=1}^\infty a_s\Big)^2} < \infty.$$

Thus $(W_t)_{t\in I}$ *is an* L^2*-bounded martingale, according to Theorem 14.33,* $\{W_t^2 : t \in I\}$ *is uniformly integrable. From this also the uniform integrability of* $\{S_t : t \in I\}$ *follows.*
'$\Rightarrow$*': Let us assume that* $\prod_{s=1}^\infty a_s = 0$. *Since* W_t *has an almost certain finite limit,* $S_t = \prod_{s=1}^t X_s \xrightarrow{t\to\infty}_{fs} 0$ *must hold. If* $\{S_t : t \in I\}$ *were uniformly integrable,* $0 = \mathbf{E}[S_\infty] = \lim_{t\to\infty} \mathbf{E}[S_t] = 1$, *i.e. a contradiction.*

Theorem 14.36 (Convergence of conditional expected values)**.**

1. *Let* $I \subseteq [0, \infty)$ *be countable with* $\sup I = u \in (0, \infty]$, $(\mathcal{F}_t)_{t\in I}$ *a filtration and* $\mathcal{F}_u = \sigma(\bigcup_{t\in I} \mathcal{F}_t)$. *Then the following applies for* $X \in \mathcal{L}^1$ *that*

$$\mathbf{E}[X|\mathcal{F}_t] \xrightarrow{t\uparrow u}_{fs,L^1} \mathbf{E}[X|\mathcal{F}_u].$$

2. *Let* $I \subseteq (-\infty, \infty)$ *be countable with* $\inf I = u \in [-\infty, \infty)$, $(\mathcal{F}_t)_{t\in I}$ *a filtration and* $\mathcal{F}_u = \bigcap_{t\in I} \mathcal{F}_t$. *Then the following applies for* $X \in \mathcal{L}^1$ *that*

$$\mathbf{E}[X|\mathcal{F}_t] \xrightarrow{t\downarrow u}_{fs,L^1} \mathbf{E}[X|\mathcal{F}_u].$$

Proof. We only show 1. since the proof of 2. proceeds analogously. With $\mathbf{E}[|\mathbf{E}[X|\mathcal{F}_t]|] \leq \mathbf{E}[|X|] < \infty$ converges according to Theorem 14.29 the martingale $(\mathbf{E}[X|\mathcal{F}_t])_{t\in I}$ converges almost surely. The L^1-convergence follows with Theorem 14.32 and Lemma 11.5. The limit value X_u can be chosen $\mathcal{F}_u$-measurable can be chosen. We will now show that $X_u = \mathbf{E}[X|\mathcal{F}_u]$, from which the assertion follows.

It is clear that $\mathbf{E}[\mathbf{E}[X|\mathcal{F}_t], A] = \mathbf{E}[X, A]$ applies to all $A \in \mathcal{F}_s$ and $s \leq t$. With $t \uparrow u$ is therefore $\mathbf{E}[X_u, A] = \mathbf{E}[X, A]$ for all $A \in \mathcal{F}_s$ and with $s \uparrow u$ also $\mathbf{E}[X_u, A] = \mathbf{E}[X, A]$ for all $A \in \mathcal{F}_u$. Since X_u is measurable with respect to $\mathcal{F}_u$-measurable, this means that $X_u = \mathbf{E}[X|\mathcal{F}_u]$. □

We now come to backward martingales, which are martingales with an index set downward unlimited index set $I \subseteq (-\infty, 0]$. These converge under very weak conditions.

Theorem 14.37 (Martingale convergence theorem for backward martingales)**.** *Let $I \subseteq (\infty, 0]$ be discrete, $\inf I = u \in (-\infty, 0]$, $\mathcal{F}_u = \bigcap_{t\in I} \mathcal{F}_t$ and $\mathcal{X} = (X_t)_{t\in I}$ a sub-martingale. Then are equivalent*

1. *There is a $\mathcal{F}_u$-measurable, integrable random variable X_u with $X_t \xrightarrow{t\downarrow u}_{fs,\ L^1} X_u$*
2. $\inf_{t\in I} \mathbf{E}[X_t] > -\infty$.

Then $(X_t)_{t\in I\cup\{u\}}$ is also a sub-martingale. In particular, every backward martingale converges almost surely and in L^1.

Proof. Wlog, let $I = \{..., -2, -1, 0\}$ and $u = -\infty$.
'1. $\Rightarrow$ 2.': From the convergence in the mean follows

$$\inf_{t\in I} \mathbf{E}[X_t] = \lim_{t\to-\infty} \mathbf{E}[X_t] = \mathbf{E}[X_{-\infty}] > -\infty.$$

'2. $\Rightarrow$ 1.': The almost sure convergence follows as in the proof of Theorem 14.29, where the condition $\sup_{t\in I} \mathbf{E}[X_t^+] < \infty$ because of $I \subseteq (-\infty, 0]$ must be replaced by $\inf_{t\in I} \mathbf{E}[X_t^-] < \infty$. We further define for $t = ..., -2, -1, 0$

$$Y_t := \mathbf{E}[X_t - X_{t-1}|\mathcal{F}_{t-1}] \geq 0.$$

Then,

$$\mathbf{E}\Big[\sum_{t=0}^{-\infty} Y_t\Big] = \mathbf{E}[X_0] - \inf_{t\in I} \mathbf{E}[X_t] < \infty.$$

Thus, $\sum_{t=0}^{-\infty} Y_t < \infty$ almost surely, and we define

$$A_t = \sum_{s\leq t} Y_s, \qquad M_t = X_t - A_t$$

Now $(A_t)_{t\in I}$ is uniformly integrable because $\mathbf{E}[A_0] < \infty$, and $(M_t)_{t\in I}$ is integrable because it is uniformly integrable by Lemma 11.5. Thus, $\mathcal{X}$ is uniformly integrable, and the L^1-convergence follows. The proof that $(X_t)_{t\in I\cup\{-\infty\}}$ is a sub-martingale proceeds analogous to the proof in 14.32. □

Example 14.38 (The strong law of large numbers)**.** *Let $X_1, X_2, ... \in L^1$ be independently identically distributed. For $t \in \{..., -2, -1\}$ we set as in the Example 14.4.2,*

$$S_t := \frac{1}{|t|}\sum_{s=1}^{|t|} X_s$$

and $\mathcal{F}_t = \sigma(..., S_{t-1}, S_t) = \sigma(S_t, X_{t+1}, X_{t+2}, ...)$. Then $(S_t)_{t\in I}$ is a backward martingale with $S_t = \mathbf{E}[X_1|\mathcal{F}_t]$. According to Theorem 14.37, S_t converges almost surely and in L^1 against a random variable $S_{-\infty}$. This is measurable with respect to $\mathcal{F}_{-\infty}$, but also with respect to $\mathcal{T}(X_1, X_2, ...)$, the terminal σ-algebra of the family $\{X_1, X_2, ...\}$. Since this σ-algebra is trivial according to Kolmogoroff's 0-1-law, $S_{-\infty}$ is almost certainly constant. Since $(S_t)_{t\in I\cup\{-\infty\}}$ is a martingale, it follows that

$$\frac{1}{|t|}\sum_{s=1}^{|t|} X_s = S_t \xrightarrow{t\to-\infty}_{fs, L^1} S_{-\infty} = \mathbf{E}[S_{-\infty}] = \mathbf{E}[S_{-1}] = \mathbf{E}[X_1].$$

However, the almost sure convergence is exactly the statement of the law of large numbers.

We now come to an application of the martingale convergence theorems, an improvement of the Borel-Cantelli lemma, Theorem 8.8. For this we need a lemma.

Lemma 14.39 (Convergence and increasing process)**.** *Let $\mathcal{M} = (M_t)_{t=0,1,2,...}$ be an L^2-integrable martingale, where $|M_t - M_{t-1}| \leq K$ for some K and all $t = 1, 2, ...$ holds. Then there is a nullset N such that*

$$\begin{aligned}\{\langle\mathcal{M}\rangle_\infty < \infty\} &\subseteq \{\lim_{t\to\infty} M_t \text{ exists}\} \cup N,\\ \{\langle\mathcal{M}\rangle_\infty = \infty\} &\subseteq \{\lim_{t\to\infty} M_t/\langle\mathcal{M}\rangle_t = 0\} \cup N.\end{aligned}$$

Proof. We start with the first statement. First, for each $k = 1, 2, 3, ...$ the random time

$$T_k := \inf\{t : \langle\mathcal{M}\rangle_t > k\}$$

is a stopping time. From this already follows

$$\{\langle\mathcal{M}\rangle_\infty < \infty\} = \bigcup_{k=1}^\infty \{T_k = \infty\}. \tag{14.4}$$

Furthermore, the stopped process $(\langle\mathcal{M}\rangle_{t\wedge T_k})_{t=0,1,2,...}$ is previsible, because for $A \in \mathcal{B}(\mathbb{R})$,

$$\{\langle\mathcal{M}\rangle_{t\wedge T_k} \in A\} = \big(\{T_k > t-1\} \cap \{\langle\mathcal{M}\rangle_t \in B\}\big) \cup \bigcup_{s=0}^{t-1}\{T_k = s, \langle\mathcal{M}\rangle_s \in A\} \in \mathcal{F}_{t-1}.$$

Let us now consider the martingale $(\mathcal{M}^{T_k})^2 - \langle\mathcal{M}\rangle^{T_k} = (\mathcal{M}^2 - \langle\mathcal{M}\rangle)^{T_k}$ for $k = 1, 2, ...$ It is $\langle\mathcal{M}^{T_k}\rangle = \langle\mathcal{M}\rangle^{T_k}$ and $\langle\mathcal{M}\rangle^{T_k}$ is bounded by $k + K^2$. Thus $\mathcal{M}^{T_k}$ is bounded in L^2 and thus converges almost surely. However, on the set $\{T_k = \infty\}$, the process $\mathcal{M}^{T_k}$ converges if and only if $\mathcal{M}$ converges. Together with (14.4) the statement follows.

For the second statement, we consider the martingale $\mathcal{X} := (1 + \langle\mathcal{M}\rangle)^{-1} \cdot \mathcal{M}$. Since $(1 + \langle\mathcal{M}\rangle)^{-1}$ is bounded and $\mathcal{M}$ is an L^2-integrable martingale, $\mathcal{X}$ is an L^2-integrable martingale. Furthermore, according to Example 14.15,

$$\begin{aligned}
\langle\mathcal{X}\rangle_t &= \Big(\frac{1}{(1+\langle\mathcal{M}\rangle)^2} \cdot \langle\mathcal{M}\rangle\Big)_t = \sum_{s=1}^{t} \frac{1}{(1+\langle\mathcal{M}\rangle_s)^2}(\langle\mathcal{M}\rangle_s - \langle\mathcal{M}\rangle_{s-1}) \\
&\leq \sum_{s=1}^{t} \frac{1}{(1+\langle\mathcal{M}\rangle_s)(1+\langle\mathcal{M}\rangle_{s-1})}(\langle\mathcal{M}\rangle_s - \langle\mathcal{M}\rangle_{s-1}) = \sum_{s=1}^{t} \frac{1}{1+\langle\mathcal{M}\rangle_{s-1}} - \frac{1}{1+\langle\mathcal{M}\rangle_s} \\
&= 1 - \frac{1}{1+\langle\mathcal{M}\rangle_t}.
\end{aligned}$$

This means that the martingale $\mathcal{X}$ converges after 1. i.e. in particular

$$\sum_{s=1}^{\infty} \frac{M_s - M_{s-1}}{1+\langle\mathcal{M}\rangle_s} < \infty.$$

Now the Kronecker lemma 8.24 provides that

$$\frac{\sum_{s=1}^{t} M_s - M_{s-1}}{\langle\mathcal{M}\rangle_t} \xrightarrow{t\to\infty} 0$$

on $\{\langle\mathcal{M}\rangle_\infty = \infty\}$. □

Theorem 14.40 (Extension of the Borel-Cantelli lemma)**.** *Let $A_t \in \mathcal{F}_t$, $t = 0, 1, 2, ...$ and*

$$X_s := \mathbf{P}(A_s|\mathcal{F}_{s-1}).$$

1. *On $\sum_{t=1}^{\infty} X_t < \infty$ only a finite number of the A_t occur, i.e.*

$$\Big\{\sum_{t=1}^{\infty} X_t < \infty\Big\} \subseteq \Big\{\sum_{t=1}^{\infty} 1_{A_t} < \infty\Big\}.$$

2. *On $\sum_{t=1}^{\infty} X_t = \infty$ applies $\sum_{t=1}^{\infty} 1_{A_t} / \sum_{t=1}^{\infty} X_t = 1$, thus*

$$\Big\{\sum_{t=1}^{\infty} X_t = \infty\Big\} \subseteq \Big\{\sum_{t=1}^{\infty} 1_{A_t} / \sum_{t=1}^{\infty} X_t = 1\Big\} \subseteq \Big\{\sum_{t=1}^{\infty} 1_{A_t} = \infty\Big\}.$$

Remark 14.41 (Extension)**.** *The Borel-Cantelli Lemma from theorem 8.8 can now be easily be derived. Namely, if*

$$\mathbf{E}\Big[\sum_{t=1}^{\infty} X_t\Big] = \sum_{t=1}^{\infty} \mathbf{P}(A_t) < \infty,$$

then $\sum_{t=1}^{\infty} X_t < \infty$ almost certainly applies. The statement now gives that at most a finite number of the A_n occur. If further $A_1, A_2, ...$ are independent, then we set $\mathcal{F}_t = \sigma(A_1, ..., A_t)$ and thus $X_s = \mathbf{E}[1_{A_s}|\mathcal{F}_{s-1}] = \mathbf{P}(A_s)$. Now, $\sum_{t=1}^{\infty} \mathbf{P}(A_t) = \infty$, infinitely many of the A_n's occur.

Proof. We consider the martingale $\mathcal{M}$ with

$$M_t = \sum_{s=1}^{t} 1_{A_s} - X_s.$$

Then,

$$\langle \mathcal{M} \rangle_t = \sum_{s=1}^{t} \mathbf{E}[1_{A_s}^2 X_s^2 | \mathcal{F}_{s-1}] = \sum_{s=1}^{t} X_s(1 - X_s) \leq \sum_{s=1}^{t} X_s.$$

If now $\sum_{t=1}^{\infty} X_t < \infty$, then $\mathcal{M}$ converges according to Lemma 14.39.1. therefore also $\sum_{t=1}^{\infty} 1_{A_t} < \infty$.
If now $\sum_{t=1}^{\infty} X_t = \infty$ and $\langle \mathcal{M} \rangle_\infty < \infty$, then $\mathcal{M}$ converges and the assertion is clear.
If now $\sum_{t=1}^{\infty} X_t = \infty$ and $\langle \mathcal{M} \rangle_\infty = \infty$, then $M_t / \langle \mathcal{M} \rangle_t \xrightarrow{t\to\infty} 0$ according to Lemma 14.39.2
From this,

$$\Big| \frac{\sum_{s=1}^{t} 1_{A_s}}{\sum_{s=1}^{t} X_s} - 1 \Big| = \Big| \frac{M_t}{\sum_{s=1}^{t} X_s} \Big| \leq \Big| \frac{M_t}{\langle \mathcal{M} \rangle_t} \Big| \xrightarrow{t\to\infty} 0.$$

$\square$

14.5 The Central Limit Theorem for martingales

The Central Limit Theorem from Section 10.2 states the convergence of a sum of *independent* random variables – suitably transformed – to a normally distributed random variable. Now we treat the case of a sequence of martingales $\mathcal{M}^1 = (M_t^1)_{t=0,1,2,...}, \mathcal{M}^2 = (M_t^2)_{t=0,1,2,...}, ...$, each started in 0, which are given by $X_t^n := M_t^n - M_{t-1}^n, t = 1, 2, ...$ as a sum through $M_t^n = X_1^n + \cdots + X_t^n$ now applies. Now note that the family $X_1^n, X_2^n, ...$ do not have to be independent. Nevertheless, we can – under suitable conditions – still prove convergence in distribution against a normally distributed random variable.

Theorem 14.42 (Central limit theorem for martingales)**.** *Let $I^n = \{0, 1, 2, ..., t_n\}$ and $\mathcal{M}^n = (M_t^n)_{t \in I^n}$ a martingale with $M_0^n = 0$ with respect to a filtration $\mathcal{F}^n = (\mathcal{F}_t^n)_{t \in \mathcal{I}^n}$, $n = 1, 2, ...$ For $X_t^n := M_t^n - M_t^{n-1}$ (with $t = 1, ..., t_n$) the following applies*

$$\mathbf{E}[\max_{1 \leq s \leq t_n} |X_s^n|] \xrightarrow{n\to\infty} 0, \tag{14.5}$$

$$\sum_{s=1}^{t_n} (X_s^n)^2 \xrightarrow{n\to\infty}_p \sigma^2 > 0. \tag{14.6}$$

Then $M_{t_n}^n \xrightarrow{n\to\infty} X$ with $X \sim N(0, \sigma^2)$.

We need two lemmas in the proof of the theorem.

Lemma 14.43 (Convergence of products of random variables)**.** *Let $U_1, U_2, ..., T_1, T_2, ...$ be random variables that satisfy the following conditions:*

1. $U_n \xrightarrow{n\to\infty}_p u$,
2. $(T_n)_{n=1,2,...}$ *and* $(T_n U_n)_{n=1,2,...}$ *are uniformly integrable,*
3. $\mathbf{E}[T_n] \xrightarrow{n\to\infty} 1$.

Then $\mathbf{E}[T_nU_n] \xrightarrow{n\to\infty} u$.

Proof. Because of 3. it suffices to show that $\mathbf{E}[T_n(U_n-u)] \xrightarrow{n\to\infty} 0$. To do this, we simply show $T_n(U_n-u) \xrightarrow{n\to\infty}_p 0$, which implies the L^1-convergence due to 2. In particular, then $\mathbf{E}[T_n(U_n-u)] \xrightarrow{n\to\infty} 0$. Let $\varepsilon > 0$ and K be large enough so that $\sup_n \mathbf{P}(|T_n| >> K) \leq \varepsilon$. (Such a K exists because of the uniform integrability of $T_1, T_2, ...$) Now we write (note that for $x, y \geq 0$ and $\delta, \varepsilon > 0$ it always holds that $xy > \delta\varepsilon \to x > \delta$ or $y > \varepsilon$)

$$\limsup_{n\to\infty} \mathbf{P}(|T_n(U_n-u)| > \varepsilon) \leq \limsup_{n\to\infty} \mathbf{P}(|U_n-u| > \varepsilon/K) + \mathbf{P}(|T_n| > K) \leq \varepsilon.$$

The assertion follows from this. □

Lemma 14.44 (Estimation of the exponential function). *1. There is a $C > 0$ and a function r with $|r(x)| \leq C|x^3|$ such that*

$$\exp(ix) = (1+ix)\exp(-x^2/2 + r(x))$$

for all $x \in \mathbb{R}$ is valid.

2. $|1+ix| \leq e^{x^2/2}$ applies to all $x \in \mathbb{R}$.

Proof. 1. It is sufficient to show the assertion for small $|x|$, since it is trivial for large $|x|$. With the help of lemma 10.12, we write

$$\begin{aligned}
&\Big|\exp(ix) - (1+ix)\exp(-x^2/2)\Big| \\
&\qquad = \Big|\exp(ix) - 1 - ix + x^2/2 - (1+ix)(\exp(-x^2/2) - 1 + x^2/2) + ix^3/3\Big| \\
&\qquad \leq \Big|\exp(ix) - 1 - ix + x^2/2\Big| + |1+ix| \cdot \Big|\exp(-x^2/2) - 1 + x^2/2\Big| + \Big|x^3/3\Big| \\
&\qquad \leq \tfrac{|x^3|}{6} + |1+ix| \cdot \Big(\tfrac{|x^2|}{2} \wedge \tfrac{|x^4|}{8}\Big) + \tfrac{|x^3|}{3} \leq |x^3|
\end{aligned}$$

for all x. From this follows the assertion for small $|x|$, and thus 1. is proven. For 2. it is sufficient to use $|1+ix|^2 = 1 + x^2 \leq e^{x^2}$ and take the root. □

Proof of Theorem 14.42. First we define

$$Z_s^n := X_s^n 1_{\sum_{r=1}^{s-1}(X_r^n)^2 \leq 2\sigma^2}$$

and $N_t^n := \sum_{s=1}^t Z_s^n$. Then $(N_t^n)_{t=1,2,...}$ is a $(\mathcal{F}_t^n)_{t\in I^n}$ martingale, because

$$\mathbf{E}[N_t^n - N_{t-1}^n|\mathcal{F}_{t-1}^n] = \mathbf{E}[Z_t^n|\mathcal{F}_{t-1}^n] = 1_{\sum_{r=1}^{s-1}(X_r^n)^2 \leq 2\sigma^2} \cdot \mathbf{E}[X_t^n|\mathcal{F}_{t-1}^n] = 0,$$

since $M_t^n = X_1^n + \cdots + X_t^n$. Now,

$$\begin{aligned}
\mathbf{P}(\max_{t=1,\ldots,t_n} |M_t^n - N_t^n| > 0) &= \mathbf{P}(M_t^n \neq N_t^n \text{ for one } t \in I^n) \\
&= \mathbf{P}(X_t^n \neq Z_t^n \text{ for a } t \in I^n) \\
&= \mathbf{P}\Big(\sum_{s=1}^{t_n}(X_s^n)^2 > 2\sigma^2\Big) \xrightarrow{n\to\infty} 0,
\end{aligned} \tag{14.7}$$

where the convergence follows from (14.6). Now the following applies $M^n_{t_n} - N^n_{t_n} \xrightarrow{n\to\infty}_p 0$, so it suffices according to Slutzky's theorem, Corollary 9.9, $N^n_{t_n} \xRightarrow{n\to\infty} X \sim N(0,\sigma^2)$ to show. For this we will for any $\lambda \in \mathbb{R}$

$$\mathbf{E}[e^{i\lambda N^n_{t_n}}] \xrightarrow{n\to\infty} e^{-i\lambda^2\sigma^2/2}$$

show. With the function r from Lemma 14.44 now applies

$$\mathbf{E}[e^{i\lambda N^n_{t_n}}] = \prod_{s=1}^{t_n}(1 + i\lambda Z^n_s) \cdot \exp\Big(- \tfrac{\lambda^2}{2}\sum_{s=1}^{t_n}(Z^n_s)^2 + \sum_{s=1}^{t_n} r(\lambda Z^n_s)\Big).$$

We now set

$$T_n := \prod_{s=1}^{t_n}(1 + i\lambda Z^n_s), \qquad\qquad U_n := \exp\Big(- \tfrac{\lambda^2}{2}\sum_{s=1}^{t_n}(Z^n_s)^2 + \sum_{s=1}^{t_n} r(\lambda Z^n_s)\Big)$$

and show that for these random variables the conditions of Lemma 14.43 apply to these random variables (with $u = e^{-\lambda^2\sigma^2/2}$). For 1. first because of (14.7)

$$\lim_{n\to\infty}\sum_{s=1}^{t_n}(Z^n_s)^2 = \lim_{n\to\infty}\sum_{s=1}^{t_n}(X^n_s)^2 = \sigma^2.$$

Further, with C from Lemma 14.44

$$\begin{aligned}\Big|\sum_{s=1}^{t_n} r(\lambda Z^n_s)\Big| &\le C\cdot|\lambda^3|\cdot\sum_{s=1}^{t_n}|Z^n_s|^3 \le C\cdot|\lambda^3|\cdot\sum_{s=1}^{t_n}|X^n_s|^3 \\ &\le C\cdot|\lambda^3|\cdot\max_{1\le s\le t_n}|X^n_s|\cdot\sum_{s=1}^{t_n}|X^n_s|^2 \xrightarrow{n\to\infty} 0,\end{aligned}$$

where the convergence follows from (14.5) and (14.6).
For 2. $|T_nU_n| = |e^{i\lambda N^n_{t_n}}| = 1$, from which the uniform integrability of $(T_nU_n)_{n\in I^n}$ already follows. For the uniform integrability of $(T_n)_{n\in I^n}$ we define

$$J_n := \inf\Big\{ s \le t_n : \sum_{r=1}^{s}(X^n_r)^2 > 2\sigma^2\Big\} \wedge t_n$$

and write

$$\begin{aligned}|T_n| &= \prod_{s=1}^{J_n-1}|1 + i\lambda Z^n_s|\cdot|1 + i\lambda Z^n_{J_n}| \le \exp\Big(\tfrac{\lambda^2}{2}\sum_{s=1}^{J_n-1}(X^n_s)^2\Big)(1+\lambda|X^n_{J_n}|) \\ &\le \exp(\lambda^2\sigma^2)\cdot(1 + |\lambda|\cdot\max_{1\le s\le t_n}|X^n_s|).\end{aligned}$$

Since $\max_{1\le s\le t_n}|X^n_s| \xrightarrow{n\to\infty}_{L^1} 0$, in particular the family $(\max_{1\le s\le t_n}|X^n_s|)_{n=1,2,\dots}$ is uniformly integrable, from which the uniform integrability of $(T_n)_{n=1,2,\dots}$ follows.

We now come to 3. by showing $\mathbf{E}[T_n] = 1$. Since $\mathbf{E}[Z^n_s|\mathcal{F}^n_{s-1}] = 0$ for all $s = 1,\dots,t_n$,

$$\begin{aligned}\mathbf{E}[T_n] &= \mathbf{E}\Big[\prod_{s=1}^{t_n}(1 + i\lambda Z^n_s)\Big] \\ &= \mathbf{E}\Big[(1 + i\lambda Z^n_1)\cdot\mathbf{E}[(1 + i\lambda Z^n_2)\cdots\mathbf{E}[1 + \lambda Z^n_{t_n}|\mathcal{F}^n_{t_n-1}]\cdots|\mathcal{F}^n_1]\Big] = 1.\end{aligned}$$

Now the assertion follows directly with Lemma 14.44. $\square$

Example 14.45. *1. Let $X_1, X_2, \ldots$ be independent, identically distributed, real-valued random variable with $\mathbf{E}[X_1] = 0$ and finite variance $\mathbf{V}[X_1] = \sigma^2$. It is then known that $\mathcal{M}^n = (M_t^n)_{t=0,1,2,\ldots}$ with*

$$M_t^n = \tfrac{1}{\sqrt{n}} \sum_{s=1}^t X_t$$

is a martingale and

$$M_n^n \xRightarrow{n\to\infty} X \sim N(0, \sigma^2).$$

This can also be realized by means of Theorem 14.42: first we establish that $\int_0^\infty t\mathbf{P}(|X_1| > t)dt < \infty$ because of the finite second moment. This means that $\mathbf{P}(|X_1| > t) = o(1/t^2)$ for $t \to \infty$, can therefore be written as $\mathbf{P}(|X_1| > t) = a(t)/t^2$ with $a(t) \xrightarrow{t\to\infty} 0$. From this,

$$\begin{aligned}\mathbf{E}[\max_{1\le s\le n} |X_s|/\sqrt{n}] &= \int_0^\infty \mathbf{P}(\max_{1\le s\le n} |X_s| > t\sqrt{n})dt = \int_0^\infty 1 - (1 - \mathbf{P}(|X_1| > t\sqrt{n}))^n dt \\ &= \int_0^\infty 1 - \Big(1 - \frac{a(t\sqrt{n})}{t^2 n}\Big)^n dt \xrightarrow{n\to\infty} 0\end{aligned}$$

due to dominated convergence. Furthermore, with the law of large numbers,

$$\frac{1}{n}\sum_{s=1}^n X_s^2 \xrightarrow{n\to\infty}_{fs} \sigma^2.$$

So, the conditions of theorem 14.42 are fulfilled.

2. We bring another example of a sequence of martingales that lead to sums of dependent random variables. For this, we recall the stochastic integral from Definition 14.13. Let $Y_1, Y_2, \ldots$ be independent, identically distributed, restricted random variables with $\mathbf{E}[Y_1] = 0$ and $\mathbf{V}[Y_1] = 1$ and $\mathcal{H} = (H_t)_{t=0,1,2,\ldots}$ and $\mathcal{M}^n = (M_t^n)_{t=0,1,2,\ldots}$ given as

$$H_s = \frac{1}{s-1}(Y_1^2 + \cdots + Y_{s-1}^2), \qquad M_t^n = \frac{1}{\sqrt{n}}\sum_{s=1}^t Y_s.$$

Then,

$$(\mathcal{H}\cdot\mathcal{M}^n)_t = \frac{1}{\sqrt{n}}\sum_{s=1}^t Y_s \frac{1}{s-1}\sum_{r=1}^{s-1} Y_r^2$$

is a martingale with

$$X_t^n := (\mathcal{H}\cdot\mathcal{M}^n)_t - (\mathcal{H}\cdot\mathcal{M}^n)_{t-1} = \frac{1}{\sqrt{n}} Y_t \frac{1}{t-1}\sum_{r=1}^{t-1} Y_r^2.$$

(Note that $(X_1^n, X_2^n, \ldots)$ is not an independent family). Now, (14.5) applies by the boundedness of $Y_1, Y_2, \ldots$ We further calculate

$$\sum_{s=1}^n (X_s^n)^2 = \frac{1}{n}\sum_{s=1}^n Y_s^2\Big(\frac{1}{s-1}\sum_{r=1}^{s-1} Y_r^2\Big)^2 \xrightarrow{n\to\infty} 1,$$

from which now $(\mathcal{H}\cdot\mathcal{M}^n)_n \xRightarrow{n\to\infty} X \sim N(0,1)$ follows.

14.6 Properties of martingales in continuous time

Example 14.46 (Martingales derived from the Poisson process)**.** *Let $I = [0, \infty)$, $\mathcal{X} = (X_t)_{t \in I}$ be a Poisson process with intensity λ and $\mathcal{F}_t = \sigma(X_s : s \leq t)$. Then,*

$$(X_t - \lambda t)_{t \in I} \qquad \textit{and} \qquad \Big(X_t^2 - \lambda \int_0^t (2X_r + 1)dr\Big)_{t \in I}$$

is a martingale. The following applies for $0 \leq s \leq t$

$$\begin{aligned} &\mathbf{E}[X_t - \lambda t | \mathcal{F}_s] = \mathbf{E}[X_s + X_t - X_s - \lambda t | \mathcal{F}_s] = X_s + \lambda(t-s) - \lambda t = X_s - \lambda s, \\ &\mathbf{E}\Big[X_t^2 - X_s^2 - \lambda \int_s^t (2X_r + 1)dr | \mathcal{F}_s\Big] \\ &\qquad = \mathbf{E}\Big[(X_t - X_s)^2 + 2(X_t - X_s)X_s - \lambda((2X_s + 1)(t-s) + 2\int_s^t (X_r - X_s)dr)|\mathcal{F}_s] \\ &\qquad = \lambda(t-s) + \lambda^2(t-s)^2 + 2\lambda(t-s)X_s - \lambda((2X_s+1)(t-s) - \lambda^2(t-s)^2 = 0. \end{aligned}$$

Example 14.47 (Martingales derived from Brownian motion)**.** *Let $I = [0, \infty)$, $\mathcal{X} = (X_t)_{t \in I}$ be a Brownian motion, $\mathcal{F}_t = \sigma(X_s : s \leq t)$ and $\alpha \in \mathbb{R}$.*

1. *The processes*

$$(\alpha X_t)_{t \in I}, \qquad (\alpha X_t^2 - \alpha t)_{t \in I} \qquad \textit{and} \qquad \big(\exp(\alpha X_t - \alpha^2 t/2)\big)_{t \in I} \tag{14.8}$$

are martingales. The following applies for $0 \leq s \leq t$

$$\begin{aligned} \mathbf{E}[\alpha X_t | \mathcal{F}_s] &= \mathbf{E}[\alpha X_s + \alpha(X_t - X_s)|\mathcal{F}_s] = \alpha X_s, \\ \mathbf{E}\Big[\alpha X_t^2 - \alpha t | \mathcal{F}_s] &= \alpha \mathbf{E}[(X_t - X_s)^2 + 2(X_t - X_s)X_s + X_s^2 - t|\mathcal{F}_s] \\ &= \alpha(t-s) + \alpha X_s^2 - \alpha t = \alpha X_s^2 - \alpha s, \\ \mathbf{E}\Big[\exp(\alpha X_t - \alpha^2 t/2)|\mathcal{F}_s] &= \exp(\alpha X_s - \alpha^2 t/2) \cdot \mathbf{E}[\exp(\alpha(X_t - X_s))] \\ &= \exp(\alpha X_s - \alpha^2 t/2 + \alpha^2(t-s)/2) = \exp(\alpha X_s - \alpha^2 s/2) \end{aligned}$$

according to Example 6.13.3.
Since the process $\big(\exp(\alpha X_t - \alpha^2 t/2)\big)_{t \in I}$ is a non-negative martingale with $\mathbf{E}[\exp(\alpha X_t - \alpha t/2)] = 1$, it represents a density. Therefore, for $\tau > 0$,

$$\mathbf{Q}_\tau : \begin{cases} \mathcal{B}(\mathbb{R})^{[0,\tau]} & \to [0,1] \\ A & \mapsto \mathbf{E}[\exp(\alpha X_\tau - \alpha^2 \tau/2), A] \end{cases}$$

is another probability measure on $\mathcal{B}(\mathbb{R})^{[0,\tau]}$, which leads to a probability measure $\mathbf{Q}$ on $\mathcal{B}(\mathbb{R})^I$ since

$$\mathbf{Q}|_{\mathcal{F}_\tau} = \mathbf{Q}_\tau \tag{14.9}$$

can be continued.

2. *For $\mu \in \mathbb{R}$ the process $(X_t + \mu t)_{t \in [0,\infty)}$ is called* Brownian motion with drift μ. *This is a martingale if and only if $\mu = 0$. For $\mu > 0$ it is a sub-martingale and for $\mu < 0$ it is a super-martingale.*

There is a close connection between the Brownian motion with drift and the martingale $\big(\exp(\mu X_t - \mu^2 t/2)\big)_{t\in I}$ from (14.8).

Proposition 14.48 (Brownian motion with drift and change of measure)**.** *Let $I = [0,\infty)$ and $\mathcal{X} = (X_t)_{t\in I}$ be a Brownian motion defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Further, let $\mathcal{Y} = (Y_t)_{t\in I}$ with $Y_t = X_t + \mu t$ for a $\mu \in \mathbb{R}$ and $\mathbf{Q}$ from* (14.9)*. Then,*

$$\mathcal{X}_*\mathbf{Q} = \mathcal{Y}_*\mathbf{P} \qquad \textit{and} \qquad \mathcal{Y}_*\mathbf{Q} = \mathcal{X}_*\mathbf{P},$$

i.e. the distribution of $\mathcal{Y}$ under the measure $\mathbf{Q}$ is that of a Brownian motion with drift μ under $\mathbf{P}$. In particular, is a martingale under $\mathbf{Q}$.

Proof. First, let f be continuous and bounded, and $0 \le s \le t$. Then,

$$\begin{aligned}
\mathbf{E_Q}[f(X_t)|\mathcal{F}_s] &= \mathbf{E_P}[f(X_t)e^{\mu X_t - \mu^2 t/2}|\mathcal{F}_s] \\
&= \frac{1}{\sqrt{2\pi(t-s)}} e^{\mu X_s - \mu^2 t/2} \int f(X_s + y) e^{\mu y} e^{-y^2/(2(t-s))} dy \\
&= \frac{1}{\sqrt{2\pi(t-s)}} e^{\mu X_s - \mu^2 t/2 + \mu^2 (t-s)/2} \int f(X_s + y) e^{-(y-\mu(t-s))^2/(2(t-s))} dy \\
&= \frac{1}{\sqrt{2\pi(t-s)}} e^{\mu X_s - \mu^2 s/2} \int f(X_s + y + \mu(t-s)) e^{-y^2/(2(t-s))} dy \\
&= \mathbf{E_P}[f(X_t + \mu(t-s))|\mathcal{F}_s] \cdot e^{\mu X_s - \mu^2 s/2}.
\end{aligned}$$

Now let $0 \le t_1 \le \cdots \le t_n$ and $f_1, ..., f_n$ be continuous and bounded. Then,

$$\begin{aligned}
\mathbf{E_Q}[f_1(X_{t_1}) \cdots f_n(X_{t_n})] &= \mathbf{E_P}[f_1(X_{t_1}) \cdots f_{n-1}(X_{t_{n-1}}) \mathbf{E_P}[f_n(X_{t_n}) e^{\mu X_{t_n} - \mu^2 t_n/2} | \mathcal{F}_{t_{n-1}}]] \\
&= \mathbf{E_P}[f_1(X_{t_1}) \cdots f_{n-2}(X_{t_{n-2}}) \cdot \\
&\qquad \mathbf{E_P}[f_{n-1}(X_{t_{n-1}}) \mathbf{E_P}[f_n(X_{t_n} + \mu(t_n - t_{n-1}) | \mathcal{F}_{t_{n-1}}] e^{\mu X_{t_{n-1}} - \mu^2 t_{n-1}/2}] | \mathcal{F}_{t_{n-2}}]] \\
&= \cdots = \mathbf{E_P}[f_1(X_{t_1} + \mu t_1) \cdots f_n(X_{t_n} + \mu t_n)] = \mathbf{E_P}[f_1(Y_{t_1}) \cdots f_n(Y_{t_n})].
\end{aligned}$$

Since $f_1, ..., f_n$ were arbitrary, the finite-dimensional distributions of $\mathcal{X}_*\mathbf{Q}$ and $\mathcal{Y}_*\mathbf{P}$ are identical. The statement now follows from Proposition 13.6.1. $\square$

We will now apply the results of martingales with a countable index set to the case of an uncountable index set, $I = [0,\infty)$. Central to this is Theorem 14.49, in which we will see that there is a right-continuous modification for very many sub-martingales.

Theorem 14.49 (Regularization of martingales in continuous time)**.** *Let $I = [0,\infty)$ and $\mathcal{X} = (X_t)_{t\in I}$ be a sub-martingale. Further, $\mathcal{Y} = (Y_t)_{t\in I\cap\mathbb{Q}}$ with $Y_t = X_t$ for $t \in I \cap \mathbb{Q}$. Then, with $(\mathcal{G}_t)_{t\in I}$ from Lemma 13.25, the following holds:*

1. *There is a null set N such that $Y_t^+ := \lim_{s\downarrow t} Y_t$ for all $t \in I$ outside N exists. The process $\mathcal{Z} = (Z_t)_{t\in I}$ with $Z_t = 1_{N^c} Y_t^+$ is a $(\mathcal{G}_t)_{t\in I}$ sub-martingale.*

2. *If $(\mathcal{F}_t)_{t\in I}$ is right-continuous, then $\mathcal{X}$ has a modification with paths in $\mathcal{D}_{\mathbb{R}}([0,\infty))$ if $t \mapsto \mathbf{E}[X_t]$ is right-continuous.*

Proof. Since $(|\mathcal{Y}_t|)_{t\in I\cap\mathbb{Q}}$ is a sub-martingale, $\sup_{t\leq\tau}\mathbf{E}[|Y_t|]<\infty$ for $\tau<\infty$. Thus, according to Theorem 14.29, for each $t\in I$ the limits $Y_{t\pm}, t\in I$ outside a nullset N. This means that $(Z_t)_{t\in I}$ with $Z_t = 1_{N^c}Y_t^+$ is right-continuous with left-sided limits. Furthermore, Z_t is measurable with respect to $\sigma(\mathcal{F}_t,\mathcal{N})^+, t\in I$.

We now show that $(Z_t)_{t\in I}$ is a sub-martingale. Let $s<t$ and $s_n\downarrow s$, as well as $t_n\downarrow t$ (and $s_n\leq t$, $n=1,2,...$). Then obviously $Y_{s_m}\leq\mathbf{E}[Y_{t_n}|\mathcal{F}_{s_m}]$ for all m,n. This means that $Z_s\leq\mathbf{E}[Y_{t_n}|\mathcal{F}_{s+}]$ according to Theorem 14.36. Since $\sup_n\mathbf{E}[Y_{t_n}]<\infty$, the sub-martingale $(Y_{t_n})_{n=1,2,...}$ is according to Theorem 14.37 uniformly integrable with $Y_{t_n}\xrightarrow{n\to\infty}_{fs,L^1} Z_t$, and thus $\mathbf{E}[Y_{t_n}|\mathcal{F}_{s+}]\xrightarrow{n\to\infty}_{fs,L^1}\mathbf{E}[Z_t|\mathcal{F}_{s+}]$. From this, $Z_s\leq\mathbf{E}[Z_t|\mathcal{F}_{s+}]=\mathbf{E}[Z_t|\mathcal{G}_s]$.

2. With the same notation, for $t\in I$ and $t_n\downarrow t$ with $t_1,t_2,...\in\mathbb{Q}$,

$$\mathbf{E}[X_{t_n}]=\mathbf{E}[Y_{t_n}],\qquad X_t\leq\mathbf{E}[Y_{t_n}|\mathcal{F}_t].$$

Because of $t_n\downarrow t$, $\lim_{s\downarrow t}\mathbf{E}[X_s]=\mathbf{E}[Z_t]$. Furthermore, due to the right-continuity of $(\mathcal{F}_t)_{t\in I}$ and Theorem 14.37 $X_t\leq\mathbf{E}[Z_t|\mathcal{F}_t]=Z_t$. If $\mathcal{X}$ has a right-continuous modification, then $Z_t=X_t$ is almost certain, and thus $\lim_{s\downarrow t}\mathbf{E}[X_s]=\mathbf{E}[X_t]$, therefore $t\mapsto\mathbf{E}[X_t]$ right-handed. On the other hand, if $t\mapsto\mathbf{E}[X_t]$ is right-handed, then $\mathbf{E}[|Z_t-X_t|]=0$, and thus $Z_t=X_t$ almost surely. Thus $(Z_t)_{t\in I}$ is a right-continuous modification of $\mathcal{X}$. □

Remark 14.50 (Usual conditions)**.** *Let $I=[0,\infty)$. In the following, we will always assume that the filtration $(\mathcal{F}_t)_{t\geq0}$ is right-continuous and complete. Furthermore, Theorem 14.49 shows that, under these assumptions, for each sub-martingale $\mathcal{X}$ there is a modification with paths in $\mathcal{D}_{\mathbb{R}}([0,\infty))$ if $t\mapsto\mathbf{E}[X_t]$ is right-continuous. We also want to assume this modification of each sub-martingale has paths in $\mathcal{D}_{\mathbb{R}}([0,\infty))$. All this we will summarize and say that the* usual conditions *hold.*

Theorem 14.51 (Martingale convergence theorems for continuous I)**.** *Let $I\subseteq[0,\infty)$ be an interval. Under the usual conditions, the statements of Lemma 14.25, Proposition 14.26, Lemma 14.28, Theorem 14.29, Corollary 14.30, Theorem 14.32, Theorem 14.33, Theorem 14.36 and Theorem 14.37 apply accordingly.*

Proof. Note that all statements already apply in the case of countable index set, e.g. $I\cap\mathbb{Q}$, have already been shown. All statements follow in the continuous case, because under the usual conditions, the process $\mathcal{X}=(X_t)_{t\in I}$, as well as all its limit values, can be uniquely constructed from $(X_t)_{t\in I\cap\mathbb{Q}}$ and its limits can be constructed. □

All martingale convergence theorems are now also shown for the case of continuous index set. The following are the statements of the Optional Sampling (Theorem 14.22) and Optional Stopping Theorem (Proposition 14.19) in the continuous case.

Theorem 14.52 (Optional Sampling Theorem in the continuous case)**.** *Let $I\subseteq[0,\infty)$ be an interval, $S\leq T$ almost surely finite stopping times and $\mathcal{X}=(X_t)_{t\in I}$ a sub-martingale. If either T is bounded or $\mathcal{X}$ is uniformly integrable, then X_T is integrable and $X_S\geq\mathbf{E}[X_T|\mathcal{F}_S]$. Furthermore, Lemma 14.23 is also valid for $I=[0,\infty)$.*

Proof. Without restriction, $I=[0,\infty)$. Let $S_n:=2^{-n}[2^nS+1]$ and $T_n:=2^{-n}[2^nT+1]$ such that $S_n\downarrow S$ and $T_n\downarrow T$ as in Proposition 13.28. With Theorem 14.22 follows $X_{S_m}\leq\mathbf{E}[X_{T_n}|\mathcal{F}_{S_m}]$ for all $m\geq n$. With $m\to\infty$ and Theorem 14.36.2,

$$X_S\leq\mathbf{E}[X_{T_n}|\mathcal{F}_S]. \tag{14.10}$$

If T is almost surely bounded, then $..., X_{T_2}, X_{T_1}$ is a sub-martingale with $\inf_n \mathbf{E}[X_{T_n}] > -\infty$. Therefore, according to Theorem 14.37 it is a uniformly integrable, almost surely and in L^1 convergent sub-martingale with limit X_T. Now follows the statement from (14.10) with $m \to \infty$.

If $\mathcal{X}$ is uniformly integrable, then it converges according to Theorem 14.32 (or Theorem 14.51) as $X_t \xrightarrow{t\to\infty}_{fs,L^1} X_\infty$ with integrable limit X_∞. and $X_s \leq \mathbf{E}[X_\infty|\mathcal{F}_s]$ applies.

As above, first $X_S \leq \mathbf{E}[X_{T_n}|\mathcal{F}_S]$, and the sub-martingale $..., X_{T_2}, X_{T_1}$ converges almost surely and in L^1 against X_T. So the statement applies again because of (14.10).

The proof of Lemma 14.23 applies unchanged. □

Corollary 14.53 (Optional stopping in the continuous case). *Let $I \subseteq [0,\infty)$ be an interval and $\mathcal{X} = (X_t)_{t\in I}$ a (sub, super)-martingale and T an almost surely finite stopping time. Then $\mathcal{X}^T = (X_{T\wedge t})_{t\in I}$ is a (sub, super) martingale.*

Proof. The corollary follows with the Optional Sampling Theorem, since $T\wedge s \leq T \wedge t$, thus $X_{T\wedge s} \leq \mathbf{E}[X_{T\wedge t}|\mathcal{F}_{T\wedge s}] \leq \mathbf{E}[X_{T\wedge t}|\mathcal{F}_s]$. □

15 Markov processes

The simplest stochastic processes $\mathcal{X} = (X_t)_{t\in I}$ are those in which $\mathcal{X}$ is an independent family. We now come to the second simplest dependency structure that occurs in stochastic processes. By a Markov process $\mathcal{X}$ we understand a process in which at time t the future $(X_u)_{u>t}$ depends only on X_t, but not on $(X_s)_{s<t}$. In other words: $(X_s)_{s>t}$ and $(X_s)_{s<t}$ are given independently X_t.

Many of the stochastic processes already introduced are Markov processes and will serve as examples in this section. Throughout this section, let (E, r) be a complete and separable metric space.

15.1 Definition and examples

In this section, we will introduce the notion of conditional independence from Section 11.4 will be needed. Finally, Markov processes are those in which the future – given the present – does not depend on the past. After the introduction of Markov processes and some examples, we will determine in Theorem 15.5, when Gaussian processes are Markov. A central notion will be Markov kernels $\mu^{\mathcal{X}}_{s,t}$, which represent just the transition probabilities between two points in time s and t. Formally equivalent, we introduce operators $T^{\mathcal{X}}_{s,t}$, which indicate how expected values of functions $f(X_t)$ change over time.

Definition 15.1 (Markov process). *Let $(\mathcal{F}_t)_{t\in I}$ be a filtration and $\mathcal{X} = (X_t)_{t\in I}$ an adapted stochastic process.*

1. *The process $\mathcal{X}$ is called* Markov process *if $\mathcal{F}_s$ is independent of X_t given X_s, $s \leq t$. This means that for $A \in \mathcal{B}(E)$ (see Proposition 11.18)*
$$\mathbf{P}(X_t \in A|\mathcal{F}_s) = \mathbf{P}(X_t \in A|X_s) \tag{15.1}$$
or equivalently
$$\mathbf{E}(f(X_t)|\mathcal{F}_s) = \mathbf{E}(f(X_t)|X_s)$$
for all measurable and bounded $f : E \to \mathbb{R}$.

2. *The Markov kernels (or transition kernels) $\mu^{\mathcal{X}}_{s,t}$ (from E to E) of $\mathcal{X}$ are given by*

$$\mu^{\mathcal{X}}_{s,t}(X_s, B) = \mathbf{P}(X_t \in B|X_s) = \mathbf{P}(X_t \in B|\mathcal{F}_s).$$

3. *Let $\mathcal{B}(E)$ be (not only the Borel's σ-algebra on E, but also) the set of bounded, measurable functions $f : E \to \mathbb{R}$. Then we define for $s \le t$ the transition operator*

$$T^{\mathcal{X}}_{s,t} : \begin{cases} \mathcal{B}(E) & \to \mathcal{B}(E) \\ f & \mapsto x \mapsto \mathbf{E}[f(X_t)|X_s = x] = \int \mu^{\mathcal{X}}_{s,t}(x, dy) f(y). \end{cases}$$

4. *For Markov kernels μ, ν from E to E we define a Markov kernel from E to E^2 by*

$$(\mu \otimes \nu)(x, A \times B) = \int \mu(x, dy)\nu(y, dz) 1_{y \in A, z \in B}$$

and a Markov kernel from E to E by

$$(\mu\nu)(x, A) = (\mu \otimes \nu)(x, E \times A).$$

Remark 15.2 (Interpretations)**.** 1. *Just as with martingales, the Markov property is formulated with respect to a filtration $(\mathcal{F}_t)_{t \in I}$. In the following, however, we will always use $\mathcal{F}_t = \sigma((X_s)_{s \le t})$, $t \in I$.*

2. *We want the transition kernels $(\mu^{\mathcal{X}}_{s,t})_{s \le t}$ as regular versions of the conditional expectation of X_t given X_s. This is possible because E is Polish and according to Theorem 11.23, then the regular version of the conditional distribution exists.*

3. *The transition operator $T^{\mathcal{X}}_{s,t}$ is best interpreted as follows: Given a function f and X_s), then $(T^{\mathcal{X}}_{s,t} f)(X_s)$ is the expectation of $f(X_t)$ at the start in X_s. This naturally depends on the value X_s so $T^{\mathcal{X}}_{s,t} f$ is a function of X_s.*

4. *To interpret the Markov kernels $\mu^{\mathcal{X}}_{s,t} \otimes \mu^{\mathcal{X}}_{t,u}$ and $\mu^{\mathcal{X}}_{s,t}\mu^{\mathcal{X}}_{t,u}$ for $s \le t \le u$ note the following: It is $\mu^{\mathcal{X}}_{s,t} \otimes \mu^{\mathcal{X}}_{t,u}(x, A \times B)$ is the probability, given $X_s = x$, that is both $X_t \in A$ and $X_u \in B$. In addition, under $\mu^{\mathcal{X}}_{s,t}\mu^{\mathcal{X}}_{t,u}$ the state at time t is integrated out, i.e. $\mu^{\mathcal{X}}_{s,t}\mu^{\mathcal{X}}_{t,u}(x, B)$ is the probability, given $X_s = x$, that $X_u \in B$. (Of course, in the case of a of a Markov process must be equal to $\mu^{\mathcal{X}}_{s,u}(x, B)$; see also the Chapman-Kolmogorov equations in Corollary 15.16.)*

Example 15.3 (Markov chains)**.** *(See also example 5.10.) Markov processes $\mathcal{X} = (X_t)_{t \in I}$ with at most countable state space E are called* Markov chains*. Furthermore, if $I = \{0, 1, 2, ...\}$, then the transition kernel $\mu^{\mathcal{X}}_{t,t+1}$ is represented by a matrix $P_{t,t+1} = (p_{t,t+1}(x, y))_{x,y \in E}$ so that*

$$p_{t,t+1}(x, y) = \mathbf{P}(X_{t+1} = y|X_t = x)$$

and

$$\mu^{\mathcal{X}}_{t,t+1}(x, A) = \sum_{y \in A} p_{t,t+1}(x, y).$$

Further here is

$$(\mu^{\mathcal{X}}_{t,t+1} \otimes \mu^{\mathcal{X}}_{t+1,t+2})(x, A \times B) = \sum_{y \in A, z \in B} p_{t,t+1}(x,y) p_{t+1,t+2}(y,z)$$

and

$$(\mu^{\mathcal{X}}_{t,t+1} \mu^{\mathcal{X}}_{t+1,t+2})(x, A) = \sum_{y \in E, z \in A} p_{t,t+1}(x,y) p_{t+1,t+2}(y,z).$$

For the transition operator $(T^{\mathcal{X}}_{s,t})_{s \leq t}$ can be written $f : E \to \mathbb{R}$ can be written as a restricted vector, namely as $f = (f(x))_{x \in E}$ and thus

$$(T^{\mathcal{X}}_{t,t+1} f)(x) = \sum_{y \in E} \mu^{\mathcal{X}}_{t,t+1}(x, dy) f(y) = \sum_{y \in E} p_{t,t+1}(x,y) f(y),$$

so the application of $T^{\mathcal{X}}_{t,t+1}$ to f corresponds to a multiplication of the matrix $p_{t,t+1}$ with the vector f.

Example 15.4 (sums and products of independent random variables etc.)**.**

1. *Let $X_1, X_2, ...$ be real-valued, almost certainly finite and independent. Then $\mathcal{S} = (S_t)_{t=0,1,2,...}$ with $S_t = \sum_{s=1}^t X_s$ and also $\mathcal{S} = (S_t)_{t=0,1,2,...}$ with $S_t = \prod_{s=1}^t X_s$ Markov processes. The following applies for example for $A \in \mathcal{B}(\mathbb{R})$*

$$\begin{aligned} \mathbf{P}(S_{t+1} \in A | \mathcal{F}_t) &= \int \mathbf{P}(S_t \in A - x, X_{t+1} \in dx | \mathcal{F}_t) \\ &= \int 1_{S_t \in A - x} \mathbf{P}(X_{t+1} \in dx) = \mathbf{P}(S_{t+1} \in A | S_t). \end{aligned}$$

   In this case

$$\mu^{\mathcal{S}}_{t,t+1}(x, A) = \mathbf{P}(X_{t+1} \in A - x)$$

   and

$$(T^{\mathcal{S}}_{t,t+1} f)(x) = \mathbf{E}[f(x + X_{t+1})].$$

2. *Let $\mathcal{X} = (X_t)_{t \geq 0}$ be a Poisson process with intensity λ. Then $(X_t)_{t \geq 0}$ and $(X_{f(t)})_{t \geq 0}$ for each growing function f Markov processes, just like $(X_t - \lambda t)_{t \geq 0}$. However, $(X_t^2 - \lambda \int_0^t (2X_r + 1) dr)_{t \geq 0}$ is not a Markov process; see also Example 14.46. (Note for the last process: assuming $X_t^2 - \lambda \int_0^t (2X_r + 1) dr = x$, the process decreases linearly with slope $\lambda(2X_t + 1)$. However this slope is not a function of x).*

   Let's look at the Poisson process $\mathcal{X}$. Here the Markov kernels for $x \in \{0, 1, 2, ...\}$ are given as

$$\mu^{\mathcal{X}}_{s,t}(x, A) = \sum_{k \in A \cap \{x, x+1, ...\}} e^{-\lambda(t-s)} \frac{(\lambda(t-s))^{k-x}}{(k-x)!},$$

   and the transition operator for $f : \{0, 1, 2, ...\} \to \mathbb{R}$ is bounded

$$(T^{\mathcal{X}}_{s,t} f)(x) = \sum_{k=0}^{\infty} e^{-\lambda(t-s)} \frac{(\lambda(t-s))^k}{k!} f(x+k) = \mathbf{E}[f(x+P)],$$

   where P is a Poisson distributed random variable with parameters $\lambda(t-s)$.

3. *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion. Then both $(\mu X_t)_{t\geq 0}$ and $(\mu X_t^2 - \mu t)_{t\geq 0}$ as well as $(\exp(\mu X_t - \mu^2 t/2))_{t\geq 0}$ for $\mu \in \mathbb{R}$ Markov processes (as well as martingales according to example 14.47). For example*

$$\begin{aligned}\mathbf{P}[X_u^2 - u \leq x|\mathcal{F}_t] &= \mathbf{P}[(X_u - X_t)^2 + 2(X_u - X_t)X_t + X_t^2 \leq u + x|\mathcal{F}_t]\\ &= \mathbf{P}[(X_u - X_t)^2 + 2(X_u - X_t)X_t + X_t^2 \leq u + x|X_t] = \mathbf{P}[X_u^2 - u \leq x|X_t].\end{aligned}$$

Let us consider the Brownian motion $\mathcal{X}$. Its Markov kernels is given by

$$\mu^{\mathcal{X}}_{s,t}(x, A) = \frac{1}{\sqrt{2\pi(t-s)}} \int_A \exp\Big(-\frac{(y-x)^2}{2(t-s)}\Big)dy$$

and the transition operator for $f \in \mathcal{B}(\mathbb{R})$

$$(T^{\mathcal{X}}_{s,t} f)(x) = \frac{1}{\sqrt{2\pi(t-s)}} \int \exp\Big(-\frac{y^2}{2(t-s)}\Big) f(x+y)dy = \mathbf{E}[f(x + \sqrt{t-s}Z)],$$

where Z is a $N(0,1)$-distributed random variable.

Theorem 15.5 (Gaussian Markov processes)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Gaussian process. Then, $\mathcal{X}$ is Markov iff*

$$\mathbf{COV}(X_s, X_u) \cdot \mathbf{V}(X_t) = \mathbf{COV}(X_s, X_t) \cdot \mathbf{COV}(X_t, X_u) \tag{15.2}$$

for all $s \leq t \leq u$.

Proof. By subtracting the expected values, we can assume wlog that $\mathbf{E}[X_t] = 0$ holds for all $t \geq 0$ is valid. We note that (if $\mathbf{V}(X_t) > 0$) with

$$X_u' = X_u - \frac{\mathbf{COV}(X_t, X_u)}{\mathbf{V}(X_t)} X_t$$

holds that $\mathbf{COV}(X_u', X_t) = 0$. Therefore, X_u' and X_t are independent (and the joint distribution is a normal distribution). In the case $\mathbf{V}(X_t) = 0$ we set $X_u' = X_u$ from which the same follows.

First, let $\mathcal{X}$ be Markov and $s \leq t \leq u$. Then X_s is independent of X_u given X_t, so X_s is also independent of X_u' given X_t. Since X_t and X_u' are independent, we find

$$\begin{aligned}\mathbf{P}(X_s \in A, X_u' \in B) &= \mathbf{E}[\mathbf{P}(X_s \in A|X_t) \cdot \mathbf{P}(X_u' \in B|X_t)]\\ &= \mathbf{E}[\mathbf{P}(X_s \in A|X_t) \cdot \mathbf{P}(X_u' \in B)] = \mathbf{P}(X_s \in A) \cdot \mathbf{P}(X_u' \in B)\end{aligned}$$

and therefore X_s and X_u' are independent. This means that

$$0 = \mathbf{COV}(X_s, X_u') = \mathbf{COV}(X_s, X_u) - \frac{\mathbf{COV}(X_t, X_u)}{\mathbf{V}(X_t)} \mathbf{COV}(X_s, X_t)$$

and (15.2) follows.

Conversely, let $\mathcal{X}$ fulfill (15.2). Then (with the same calculation as above), X_s is independent of X_u' for all $s \leq t$. This means that X_u' is independent of $\mathcal{F}_t = \sigma((X_s)_{s\leq t})$ and

$$\begin{aligned}\mathbf{P}(X_u \in A|\mathcal{F}_t) &= \int \mathbf{P}\Big(X_u' \in dx, \tfrac{\mathbf{COV}(X_t,X_u)}{\mathbf{V}(X_t)} X_t \in A - x|\mathcal{F}_t\Big)\\ &= \int \mathbf{P}\Big(X_u' \in dx, \tfrac{\mathbf{COV}(X_t,X_u)}{\mathbf{V}(X_t)} X_t \in A - x|X_t\Big)\\ &= \mathbf{P}(X_u \in A|X_t).\end{aligned}$$

$\square$

Example 15.6 (Examples of Gaussian Markov processes). 1. *We have already shown that a Brownian motion $\mathcal{X}$ is a Markov process. To be on the safe side, also note that in this case for $s \leq t \leq u$*

$$\mathbf{COV}(X_s, X_u) \cdot \mathbf{V}(X_t) = s \cdot t = \mathbf{COV}(X_s, X_t) \cdot \mathbf{COV}(X_t, X_u).$$

2. *A fractional Brownian motion with Hurst parameter h is a Gaussian process $\mathcal{X} = (X_t)_{t\geq 0}$ with $\mathbf{E}[X_t] = 0$, $t \geq 0$ and*

$$\mathbf{COV}(X_s, X_t) = \frac{1}{2}(t^{2h} + s^{2h} - (t-s)^{2h}).$$

As you can easily calculate, this is only for $h = \frac{1}{2}$ a Markov process. Then $\mathcal{X}$ is the Brownian motion.

3. *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion and $\mathcal{Y} = (Y_t)_{t\in[0,1]}$ given as $Y_t = X_t - tX_1$. Then $\mathcal{Y}$ is called Brown's bridge; see also Figure 5. It is $\mathbf{E}[Y_t] = 0, t \geq 0$ and $s \leq t$*

$$\mathbf{COV}(Y_s, Y_t) = \mathbf{COV}(X_s - sX_1, X_t - tX_1) = s - 2st + st = s(1-t).$$

This means that for $s \leq t \leq u$

$$\mathbf{COV}(Y_s, Y_u) \cdot \mathbf{V}(Y_t) = s(1-u)t(1-t) = \mathbf{COV}(Y_s, Y_t) \cdot \mathbf{COV}(Y_t, Y_u),$$

so the Brownian bridge is a Markov process.

(A) (B)

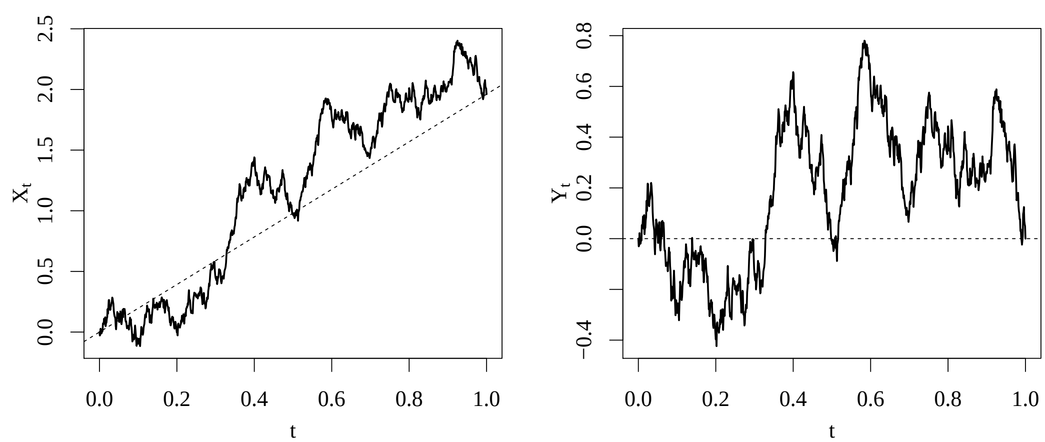

Figure 5: (A) The path of a Brownian motion $\mathcal{X} = (X_t)_{t\in[0,1]}$. (B) The corresponding path of the Brownian bridge $\mathcal{Y} = (Y_t)_{t\in[0,1]}$ with $Y_t = X_t - tX_1$.

The verbal description of Markov processes states that the future of the process is independent of the past, given the present. However, in Definition (15.1) it is only required that individual time points in the future are independent of the past, given the present. The fact that this in fact corresponds to with the verbal description is now shown.

Lemma 15.7 (Extended Markov property). *Let $\mathcal{X} = (X_t)_{t\in I}$ be a Markov process. Then $(X_u)_{u\geq t}$ is independent of $\mathcal{F}_t$ given X_t*

Proof. Let $t = t_0 < t_1 < ... < t_n \in I$ and $A_0, ..., A_n \in E$. Then the following applies

$$\begin{aligned}\mathbf{P}(X_{t_0} \in A_0,..., X_{t_n} \in A_n|\mathcal{F}_t) &= \mathbf{E}[1_{X_{t_0}\in A_0}, ..., 1_{X_{t_{n-1}}\in A_{n-1}} \cdot \mathbf{E}[1_{X_{t_n}\in A_n}|\mathcal{F}_{t_{n-1}}]|\mathcal{F}_t] \\ &= \mathbf{E}[1_{X_{t_0}\in A_0}, ..., 1_{X_{t_{n-1}}\in A_{n-1}} \cdot \mathbf{E}[1_{X_{t_n}\in A_n}|X_{t_{n-1}}]|\mathcal{F}_t] \\ &= \mathbf{E}[1_{X_{t_0}\in A_0}, ..., 1_{X_{t_{n-2}}\in A_{n-2}} \cdot \underbrace{\mathbf{E}[1_{X_{t_{n-1}}\in A_{n-1}}\mathbf{E}[1_{X_{t_n}\in A_n}|X_{t_{n-1}}]|\mathcal{F}_{t_{n-2}}]}_{\substack{=\mathbf{E}[1_{X_{t_{n-1}}\in A_{n-1}}\mathbf{E}[1_{X_{t_n}\in A_n}|X_{t_{n-1}},X_{t_{n-2}}]|X_{t_{n-2}}] \\ =\mathbf{E}[1_{X_{t_{n-1}}\in A_{n-1}}\cdot 1_{X_{t_n}\in A_n}|X_{t_{n-2}}]}} \\ mathcal F_t] \\ &= \cdots = \mathbf{E}[1_{X_{t_0}\in A_0}\mathbf{E}[1_{X_1\in A_1}, ..., 1_{X_{t_n}\in A_n}|X_{t_0}]|\mathcal{F}_t] \\ &= \mathbf{E}[1_{X_{t_0}\in A_0}, ..., 1_{X_{t_n}\in A_n}|X_t] = \mathbf{P}[X_{t_0} \in A_0, ..., X_{t_n} \in A_n|X_t].\end{aligned}$$

where we have used Proposition 11.18. This shows that $(X_{t_0}, ..., X_{t_n})$ is independent of $\mathcal{F}_t$ given X_t, i.e. the independence on cylinder sets $\{X_{t_0} \in A_0, ..., X_{t_n} \in A_n\}$. This is extended by means ofan argument with a Dynkin system to all sets in $\sigma((X_u)_{u\geq t})$. $\square$

A special case is that of a Markov process that is spatially homogeneous. This always behaves in the same way, regardless of its current value is. We have already become familiar with such processes via the Brownian motion and the Poisson process.Equivalent to this is that the process has independent increments, as Lemma 15.9 shows.

Definition 15.8 (Spatially homogeneous Markov process).
Let E be an Abelian group.

1. *A Markov kernel from E to E is called* homogeneous *if $\mu(x, B) = \mu(0, B - x)$ for all $x \in E$ and $B \in \mathcal{B}(E)$ is valid. (Here $B - x = \{y - x : y \in B\}$.)*
2. *A Markov process $\mathcal{X}$ is called* spatially homogeneous, *if the Markov kernels $\mu^{\mathcal{X}}_{s,t}$ are homogeneous, $s \leq t$.*
3. *A Markov process $\mathcal{X} = (X_t)_{t\geq 0}$ has independent increments if $X_t - X_s$ is independent of $\mathcal{F}_s$, $s \leq t$.*

Lemma 15.9 (homogeneity and independent increments). *Let $\mathcal{X} = (X_t)_{t\in I}$ be a Markov process with state space E, where E is an Abelian group. The process $\mathcal{X}$ has independent increments if and only if $\mathcal{X}$ is spatially homogeneous. In this case, the completion of the filtration $(\mathcal{F}_t)_{t\geq 0}$ with $\mathcal{F}_t = \sigma((X_s)_{s\leq t})$ is right-continuous.*

Proof. First, let $\mathcal{X}$ be a spatially homogeneous Markov process, i.e. $\mu^{\mathcal{X}}_{s,t}(x, B) = \mu^{\mathcal{X}}_{s,t}(0, B-x)$ for all $x \in E$ and $B \in \mathcal{B}(E)$. Then,

$$\mathbf{P}(X_t - X_s \in B|\mathcal{F}_s) = \mathbf{P}(X_t \in X_s + B|\mathcal{F}_s) = \mu_{s,t}(X_s, X_s + B) = \mu^{\mathcal{X}}_{s,t}(0, B).$$

Thus $X_t - X_s$ is according to Lemma 11.13 independent of $\mathcal{F}_s$, so $\mathcal{X}$ has independent increments.

Conversely, $\mathcal{X}$ has independent increments. Then $(X_t - X_s)_{t \geq s}$ is also a Markov process with the same Markov kernels and

$$\mu^{\mathcal{X}}_{s,t}(X_s, B) = \mathbf{P}(X_t \in B | \mathcal{F}_s) = \mathbf{P}(X_t - X_s \in B - X_s | \mathcal{F}_s) = \mu^{\mathcal{X}}_{s,t}(0, B - X_s).$$

We now come to the second part of the statement, the right continuity of the filtration generated by $\mathcal{X}$. Let $t \in I$ and $u_1, u_2, ... \in I$ with $u_n \downarrow t$. Wlog we assume that $\mathcal{F}_t$ is complete. We must show that $\mathcal{F}_t^+ = \bigcap_n \mathcal{F}_{u_n} = \mathcal{F}_t$. First of all, $(\mathcal{F}_t, \mathcal{G}_1, \mathcal{G}_2, ...)$ is with $\mathcal{G}_n = \sigma(X_{u_{n-1}} - X_{u_n})$ an independent family. It is $\mathcal{F}_t^+$ independent of $(\mathcal{G}_1, ..., \mathcal{G}_n)$ for each n. Let $A \in \mathcal{F}_t^+$ be. Then, according to Proposition 11.18,

$$\mathbf{P}(A | \mathcal{F}_t) = \mathbf{P}(A | \mathcal{F}_t, \mathcal{G}_1, ..., \mathcal{G}_n) \xrightarrow{n \to \infty} 1_A$$

almost surely by Theorem 14.36 and because 1_A is measurable with respect to $\sigma(\mathcal{F}_t, \mathcal{G}_1, .\mathcal{G}_2, ...)$. In particular, since $\mathcal{F}_t$ is complete, $\mathcal{F}_t^+ \subseteq \mathcal{F}_t \subseteq \mathcal{F}_t^+$. $\square$

15.2 Strong Markov processes

With martingales, we have become familiar with the procedure that a property that applies for fixed times (e.g. $X_s = \mathbf{E}[X_t | \mathcal{F}_s]$) is transferred to stopping times. (This led to the Optional Sampling Theorem, i.e. $X_S = \mathbf{E}[X_T | \mathcal{F}_S]$ for almost surely bounded stopping times $S \leq T$).

The Markov property is initially again a property for fixed points in time, which can be written, for example, as

$$\mathbf{P}(X_{s+t} \in A | \mathcal{F}_s) = \mu^{\mathcal{X}}_{s,s+t}(X_s, A).$$

Replacing the fixed time s in the last equation with a stopping time S leads to strong Markov processes. Most of the processes discussed here belong to this class, however Example 15.14 is an exception.

Definition 15.10 (Strong Markov process). *Let $\mathcal{X} = (X_t)_{t \in I}$ be a Markov process with generated filtration $(\mathcal{F}_t)_{t \in I}$ and progressively measurable. Further let S be a $(\mathcal{F}_t)_{t \in I}$ stopping time. Then $\mathcal{X}$ has the* strong Markov property at S *if*

$$\mathbf{P}(X_{S+t} \in A | \mathcal{F}_S) = \mu^{\mathcal{X}}_{S,S+t}(X_S, A)$$

for $A \in \mathcal{B}(E)$ or equivalent to this

$$\mathbf{E}[f(X_{S+t}) | \mathcal{F}_S] = (T^{\mathcal{X}}_{S,S+t} f)(X_S)$$

applies to $f \in \mathcal{B}(E)$. Further, $\mathcal{X}$ is a strong Markov process *if $\mathcal{X}$ has the strong Markov property at all almost surely finite stopping times.*

Proposition 15.11 (Strong Markov at discrete stopping times). *Let $\mathcal{X} = (X_t)_{t \in I}$ be a Markov process with generated filtration $(\mathcal{F}_t)_{t \in I}$ and progressively measurable. Further let S be an almost surely finite $(\mathcal{F}_t)_{t \in I}$-stopping time, which only assumes discrete (i.e. in particular only countably many) values. Then $\mathcal{X}$ has the strong Markov property for S.*

If I in particular is discrete, then every Markov process $\mathcal{X}$ also has the strong Markov property.

Proof. Let $\{s_1, s_2, ...\}$ be the range of values of S and $f \in \mathcal{B}(E)$ and $A \in \mathcal{F}_S$. Then (since the range of values of S is discrete) $A \cap \{S = s_i\} \in \mathcal{F}_{s_i}$ and

$$\begin{aligned}
\mathbf{E}[f(X_{S+t}), A] &= \sum_i \mathbf{E}[f(X_{S+t}), A \cap \{S = s_i\}] \\
&= \sum_i \mathbf{E}[f(X_{s_i+t}), A \cap \{S = s_i\}] \\
&= \sum_i \mathbf{E}\big[\mathbf{E}[f(X_{s_i+t})|X_{s_i}], A \cap \{S = s_i\}\big] \\
&= \sum_i \mathbf{E}[(T_{s_i,s_i+t}f)(X_{s_i}), A \cap \{S = s_i\}] \\
&= \sum_i \mathbf{E}[(T_{S,S+t}f)(X_S), A \cap \{S = s_i\}] \\
&= \mathbf{E}[(T_{S,S+t}f)(X_S), A\big].
\end{aligned}$$

Since $(T_{S,S+t}f)(X_S)$ is measurable according to $\mathcal{F}_S$, the assertion follows. □

Theorem 15.12 (Strong Markov with continuous transition operator)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be a Markov process with generated filtration $(\mathcal{F}_t)_{t\in I}$ with right-continuous paths. If $T^{\mathcal{X}}_{s,t}f$ is continuous for $f \in \mathcal{C}_b(E)$ and $s \mapsto T^{\mathcal{X}}_{s,s+t}f$ continuous for all $f \in \mathcal{C}_b(E)$ (with respect to the supremum norm on $\mathcal{C}_b(E)$), then $\mathcal{X}$ is a strong Markov process.*

Proof. First, according to Lemma 13.32, the process $\mathcal{X}$ is progressively measurable. Let S be an almost surely finite stopping time, which, according to Proposition 13.28, we replace by stopping times $S_1, S_2, ...$ with $S_n \downarrow S$ so that S_n only takes on assumes discrete values, $n = 1, 2, ...$ Then, because of the right continuity of the paths of $\mathcal{X}$ that $X_{S_n} \xrightarrow{n\to\infty} X_S$ is almost certain and for $f \in \mathcal{C}_b(E)$ is

$$\begin{aligned}
\mathbf{E}[f(X_{S+t})|\mathcal{F}_S] &= \lim_{n\to\infty} \mathbf{E}[\mathbf{E}[f(X_{S_n+t})|\mathcal{F}_{S_n}]|\mathcal{F}_S] \\
&= \lim_{n\to\infty} \mathbf{E}[(T^{\mathcal{X}}_{S_n,S_n+t}f)(X_{S_n})|\mathcal{F}_S] \\
&= \mathbf{E}[(T^{\mathcal{X}}_{S,S+t}f)(X_S)|\mathcal{F}_S] = (T^{\mathcal{X}}_{S,S+t}f)(X_S),
\end{aligned}$$

where the continuity conditions in the third equality are included. □

Example 15.13 (Poisson process and Brownian motion are strong Markov)**.**

1. *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Poisson process with intensity $\lambda \geq 0$. Then $\mathcal{X}$ is strongly Markov, because:*
   According to Example 15.4.2, $(T^{\mathcal{X}}_{s,t}f)(x) = \mathbf{E}[f(x + P)]$, where $P \sim Poi(\lambda(t - s))$. Thus $s \mapsto T^{\mathcal{X}}_{s,s+t}f$ is constant. Further, $x \mapsto (T^{\mathcal{X}}_{s,s+t}f)(x)$ is measurable and due to the discrete topology on $\{0, 1, 2, ...\}$ also continuous. The strong Markov property thus follows from Theorem 15.12.

2. *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion. Then $\mathcal{X}$ is strongly Markov, because:*
   According to Example 15.4.3 is $(T^{\mathcal{X}}_{s,t}f)(x) = \mathbf{E}[f(x + \sqrt{t - s}Z)]$, where $Z \sim N(0, 1)$. This means that $s \mapsto T^{\mathcal{X}}_{s,s+t}f$ is constant and $x \mapsto (T^{\mathcal{X}}_{s,s+t}f)(x)$ is constant. Again, the strong Markov property follows from Theorem 15.12.

It is not so easy to specify non-strong Markov processes. However, here is an example.

Example 15.14 (A non-strong Markov process)**.** *Let $T \sim \exp(1)$ be distributed. We further define the stochastic process $\mathcal{X} = (X_t)_{t\geq 0}$ with*

$$X_t = (t - T)^+$$

and completion of the canonical filtration $(\mathcal{F}_t)_{t\geq 0}$. Then for $f \in \mathcal{B}(\mathbb{R})$

$$\mathbf{E}[f(X_{s+t})|\mathcal{F}_s] = \begin{cases} \mathbf{E}[f((t-T)^+)], & \text{if } X_s = 0, \\ f(x+t), & \text{if } X_s > 0. \end{cases}$$

In particular, the right-hand side only depends on X_s and therefore $\mathcal{X}$ is a Markov process with transition operator

$$(T^{\mathcal{X}}_{s,s+t})f(x) = 1_{x=0}\mathbf{E}[f((t-T)^+)] + 1_{x>0}f(x+t).$$

Now consider the random time $S = \inf\{t : X_t > 0\}$ (i.e. $S = T$). According to Proposition 13.30.2, T is an option time and thus, since $\{T = t\}$ is a zero set and $\mathcal{F}_t$ is complete, $\{T \leq t\} = \{T < t\} \cup \{T = t\} \in \mathcal{F}_t$. Thus T is $(\mathcal{F}_t)_{t\geq 0}$ a stopping time. Now,

$$\mathbf{E}[f(X_{S+t})|\mathcal{F}_S] = f(t),$$

da S is measurable according to $\mathcal{F}_S$ and $X_{S+t} = t$ almost surely is valid. On the other hand, $X_S = 0$ and therefore

$$(T^{\mathcal{X}}_{S,S+t}f)(X_S) = (T^{\mathcal{X}}_{S,S+t}f)(0) = \mathbf{E}[f((t-T)^+)].$$

Since the right-hand sides of the last two equations for many $f \in \mathcal{B}(E)$ do not match, $\mathcal{X}$ is not a strong Markov process.

15.3 The distribution of a Markov process

For a Markov process $\mathcal{X}$, the Markov kernels $\mu^{\mathcal{X}}_{s,t}$ and the transition operators $T^{\mathcal{X}}_{s,t}$ are important tools. We will discuss in Theorem 15.17 that a consistency condition (the Chapman-Kolmogorov equations, see Corollary 15.16) is not only necessary but also sufficient for a family of Markov kernels to be Markov kernels for a Markov process.

Lemma 15.15 (Finite-dimensional distributions)**.** *Let $\mathcal{X} = (X_t)_{t\in I}$ be a Markov process with $X_t \sim \nu^{\mathcal{X}}_t$ for distributions $\nu^{\mathcal{X}}_t$ on E and Markov kernels $(\mu^{\mathcal{X}}_{s,t})_{s\leq t}$. Then, for $t_0 < \cdots < t_n$*

$$(X_{t_0}, ...X_{t_n}) \sim \nu^{\mathcal{X}}_{t_0} \otimes \mu^{\mathcal{X}}_{t_0,t_1} \otimes \cdots \otimes \mu^{\mathcal{X}}_{t_{n-1},t_n}$$

and

$$\mathbf{P}((X_{t_1}, ..., X_{t_n}) \in \cdot|\mathcal{F}_{t_0}) = (\mu^{\mathcal{X}}_{t_0,t_1} \otimes \cdots \otimes \mu^{\mathcal{X}}_{t_{n-1},t_n})(X_{t_0}, \cdot)$$

Proof. The proof of the first formula is done by induction. For $n = 0$ the statement is clear. If it applies for n, then the following applies for $f \in \mathcal{C}_b(E^{n+2})$

$$\begin{aligned}\mathbf{E}[f(X_{t_0},..., X_{t_{n+1}})] &= \mathbf{E}[\mathbf{E}[f(X_{t_0}, ..., X_{t_{n+1}})|\mathcal{F}_{t_n}]] \\ &= \mathbf{E}\Big[\int f(X_{t_0}, ..., X_{t_n}, x_{n+1})\mu^{\mathcal{X}}_{t_n,t_{n+1}}(X_{t_n}, dx_{n+1})\Big] \\ &= \int \nu^{\mathcal{X}}_{t_0} \otimes \mu^{\mathcal{X}}_{t_0,t_1} \otimes \cdots \otimes \mu^{\mathcal{X}}_{t_n,t_{n+1}}(dx_0, ..., dx_{n+1})f(x_0, ..., x_{n+1})\end{aligned}$$

so the first formula applies to $n+1$. For the second formula, we note that the right-hand side X_{t_0} is measurable. Furthermore, with Lemma 15.7

$$\mathbf{P}((X_{t_1},...,X_{t_n}) \in \cdot|\mathcal{F}_{t_0}) = \mathbf{P}((X_{t_1},...,X_{t_n}) \in \cdot|X_{t_0})$$

and for $A \in \mathcal{B}(E)$ and $B \in \mathcal{B}(E^n)$ with the first formula

$$\begin{aligned}
&\mathbf{E}[1_{(X_{t_1},...,X_{t_n})\in B}, X_{t_0} \in A] = \mathbf{P}((X_{t_0},...,X_{t_n}) \in A \times B) \\
&\quad = \int_A \nu_{t_0}^{\mathcal{X}}(dx)(\mu_{t_0,t_1}^{\mathcal{X}} \otimes \cdots \otimes \mu_{t_n,t_{n+1}}^{\mathcal{X}}(x,B) = \mathbf{E}[(\mu_{t_0,t_1}^{\mathcal{X}} \otimes \cdots \otimes \mu_{t_n,t_{n+1}}^{\mathcal{X}})(X_{t_0},B), X_{t_0} \in A],
\end{aligned}$$

from which the assertion follows. □

Corollary 15.16 (Chapman-Kolmogorov equations)**.** *Let $\mathcal{X}$ be a Markov process with $X_t \sim \nu_t^{\mathcal{X}}$ for distributions $\nu_t^{\mathcal{X}}$ on E, Markov kernels $(\mu_{s,t}^{\mathcal{X}})_{s\le t}$ and transition operators $(T_{s,t}^{\mathcal{X}})_{s\le t}$. Then, for $s \le t \le u$*

$$\mu_{s,t}^{\mathcal{X}}\mu_{t,u}^{\mathcal{X}} = \mu_{s,u}^{\mathcal{X}}, \tag{15.3}$$

and for $f \in \mathcal{B}(E)$

$$(T_{s,t}^{\mathcal{X}}(T_{t,u}^{\mathcal{X}}f))(X_s) = (T_{s,u}^{\mathcal{X}}f)(X_s) \tag{15.4}$$

$\nu_s^{\mathcal{X}}$-almost certain.

Proof. According to Proposition 15.15, for $\nu_s^{\mathcal{X}}$-almost all X_s for $A \in \mathcal{B}(E)$

$$\begin{aligned}
\mu_{s,u}^{\mathcal{X}}(X_s, A) &= \mathbf{P}(X_u \in A|\mathcal{F}_s) = \mathbf{P}((X_t, X_u) \in E \times A|\mathcal{F}_s) \\
&= (\mu_{s,t}^{\mathcal{X}} \otimes \mu_{t,u}^{\mathcal{X}})(X_s, E \times A) = (\mu_{s,t}^{\mathcal{X}}\mu_{t,u}^{\mathcal{X}})(X_s, A)
\end{aligned}$$

and for $f \in \mathcal{B}(E)$.

$$\begin{aligned}
(T_{s,u}^{\mathcal{X}}f)(X_s) &= \mathbf{E}[f(X_u)|\mathcal{F}_s] = \mathbf{E}\big[\mathbf{E}[f(X_u)|\mathcal{F}_t]|\mathcal{F}_s\big] \\
&= \mathbf{E}[(T_{t,u}^{\mathcal{X}}f)(X_t)|\mathcal{F}_s] = (T_{s,t}^{\mathcal{X}}(T_{t,u}^{\mathcal{X}}f))(X_s).
\end{aligned}$$

□

It is clear that for each Markov process there are the Markov kernels $(\mu_{s,t}^{\mathcal{X}})_{s\le t}$ exist. Conversely, we now show that for every family of Markov kernels $(\mu_{s,t})_{s\le t}$, which satisfies the Chapman-Kolmogorov equations, there is a Markov process.

Theorem 15.17 (Existence of Markov processes)**.**
Let I be an index set with $\min I = 0$, ν_0 a probability measure on E. Then the following applies:

1. *If $(\mu_{s,t})_{s\le t}$ is a family of Markov kernels with $\mu_{s,t}\mu_{t,u} = \mu_{s,u}$ for all $s \le t \le u$. Then there is a Markov process with starting distribution ν_0 and transition kernels $(\mu_{s,t})_{s\le t}$.*

2. *If $(T_{s,t})_{s\le t}$ is a family of transition operators with $T_{s,t}T_{t,u} = T_{s,u}$ for all $s \le t \le u$. Then there is a Markov process with starting distribution ν_0 and transition operators $(T_{s,t})_{s\le t}$.*

Proof. Given $(\mu_{s,t})_{s\leq t}$, it is easy to calculate that

$$(T_{s,t}f)(x) := \int \mu_{s,t}(x,dy)f(y)$$

with $f \in \mathcal{B}(E)$ a family of transition operators $(T_{s,t})_{s\leq t}$, which exactly then (15.4)is fulfilled if $(\mu_{s,t})_{s\leq t}$ fulfills the conditions (15.3) are fulfilled. If the other way around $(T_{s,t})_{s\leq t}$ is given, then defines

$$\mu_{s,t}(x,A) = (T_{s,t}1_A)(x)$$

a family of Markov kernels that (15.3) is fulfilled iff $(T_{s,t})_{s\leq t}$ fulfills the condition (15.4). It is therefore sufficient to show 1.For this we first define the measures for $t_1 < ... < t_n$ with $\{t_1, ..., t_n\} \subseteq_f I$

$$\nu_{t_1,...,t_n} = \nu_0\mu_{0,t_1} \otimes \mu_{t_1,t_2} \otimes \cdots \otimes \mu_{t_{n-1},t_n}.$$

To show that $(\nu_{t_1,...,t_n})_{\{t_1,...,t_n\}\subseteq_f I}$ a projective family is $J = \{t_1, ..., t_n\}$ and $H = \{t_1, ..., t_{k-1}, t_{k+1}, ..., t_n\}$. Then for $B = B_1 \times \cdots \times B_{k-1} \times B_{k+1} \times \cdots \times B_n \in \mathcal{B}(E^H)$

$$\begin{aligned}(\pi^J_H)_*\nu_J(B) &= \nu_J((\pi^J_H)^{-1}(B))\\ &= (\nu_0\mu_{0,t_1} \otimes \mu_{t_1,t_2} \otimes \cdots \mu_{t_{n-1},t_n})(B_1 \times \cdots \times B_{k-1} \times E \times B_{k+1} \times \cdots \times B_n)\\ &= (\nu_0\mu_{0,t_1} \otimes \mu_{t_1,t_2} \otimes \cdots \mu_{t_{k-1},t_k}\mu_{t_k,t_{k+1}} \otimes \mu_{t_{k+1},t_{k+2}} \otimes \cdots \mu_{t_{n-1},t_n})(B)\\ &= (\nu_0\mu_{0,t_1} \otimes \mu_{t_1,t_2} \otimes \cdots \mu_{t_{k-1},t_{k+1}} \otimes \mu_{t_{k+1},t_{k+2}} \otimes \cdots \mu_{t_{n-1},t_n})(B)\\ &= \nu_H(B).\end{aligned}$$

According to Theorem 5.24 there is a process $\mathcal{X} = (X_t)_{t\in I}$ with the finite-dimensional distributions $(\nu_J)_{J\subseteq_f I}$ and starting distribution ν_0. It remains to show that $\mathcal{X}$ is a Markov process. For this, let $A \in \mathcal{B}(E^J)$ for a $J \subset I$ and $\max J = s \leq t$ and $B \in \mathcal{B}(E)$. Then,

$$\mathbf{P}((X_r)_{r\in J} \in A, X_t \in B) = \nu_{J\cup\{t\}}(A \times B) = \mathbf{E}[\mu_{s,t}(X_s, B), (X_r)_{r\in J} \in A].$$

If $(\mathcal{F}_t)_{t\in I}$ is the filtration generated by $\mathcal{X}$, the filtration, then the following applies to $A \in \mathcal{F}_s$

$$\mathbf{P}(X_t \in B, A) = \mathbf{E}[\mu_{s,t}(X_s, B), A].$$

From the definition of the conditional expectation, we can read that $\mathbf{P}(X_s \in B|\mathcal{F}_s) = \mu_{s,t}(X_s, B) = \mathbf{P}(X_s \in B|X_s)$. From this the assertion follows. □

Corollary 15.18 (Distribution of Markov processes)**.** *Let ν and $(\mu_{s,t})_{s\leq t}$ be as in Theorem 15.17. Then there is a probability distribution $\mathbf{P}_\nu$ on $\mathcal{B}(E)^I$, such that $\mathbf{P}_\nu$ is the distribution of the Markov process with transition kernels $(\mu_{s,t})_{s\leq t}$ and initial distribution ν. Furthermore, $x \mapsto \mathbf{P}_x := \mathbf{P}_{\delta_x}$ defines a transition kernel from E to $\mathcal{B}(E)^I$ and*

$$\mathbf{P}_\nu = \int \nu(dx)\mathbf{P}_x.$$

Proof. It is easy to calculate that $\mathbf{P}_\nu(A) = \int \nu(dx)\mathbf{P}_x(A)$ applies to cylinder sets A. As usual, one extends this statement to all $A \in \mathcal{B}(E)^I$. □

15.4 Semigroups and generators

Temporally homogeneous Markov processes play a special role. With these, $\mu^{\mathcal{X}}_{s,t}$ depends only on the time difference $t - s$.

Definition 15.19 (Temporally homogeneous Markov process and its semigroups)**.** *Let I be closed under addition. A Markov process $\mathcal{X}$ is called* temporally homogeneous *if there is a family of Markov kernels $(\mu_t)_{t\in I}$ with $\mu^{\mathcal{X}}_{s,t} = \mu_{t-s}$. Then we also write $\mu^{\mathcal{X}}_t = \mu_t$ and denote $(\mu^{\mathcal{X}}_t)_{t\in I}$ as* transitional semigroup[4].

This is (of course) exactly the case if there is a family of transition operators $(T_t)_{t\in I}$ with $T^{\mathcal{X}}_{s,t} = T_{t-s}$. In this case, we write $T^{\mathcal{X}}_t = T_t$ and denote $(T^{\mathcal{X}}_t)_{t\in I}$ as operator semigroup.

Remark 15.20 (Transfer to temporally homogeneous Markov processes)**.** *Let $\mathcal{X}$ be a temporally homogeneous Markov process with transition and operator semigroup $(\mu^{\mathcal{X}}_t)_{t\in I}$ and $(T^{\mathcal{X}}_t)_{t\in I}$. Then, according to the results from Section 15.3,*

$$(X_{t_0}, ...X_{t_n}) \sim \nu^{\mathcal{X}}_{t_0} \otimes \mu^{\mathcal{X}}_{t_1-t_0} \otimes \cdots \otimes \mu^{\mathcal{X}}_{t_n-t_{n-1}}$$

and

$$\mathbf{P}((X_{t_1}, ..., X_{t_n}) \in \cdot|\mathcal{F}_{t_0}) = (\mu^{\mathcal{X}}_{t_1-t_0} \otimes \cdots \otimes \mu^{\mathcal{X}}_{t_n-t_{n-1}})(X_{t_0}, \cdot).$$

In addition, the Chapman-Kolmogorov equations become

$$\begin{aligned} \mu^{\mathcal{X}}_s \mu^{\mathcal{X}}_t &= \mu^{\mathcal{X}}_{s+t}, \\ T^{\mathcal{X}}_s T^{\mathcal{X}}_t &= T^{\mathcal{X}}_{s+t} \end{aligned}$$

for all $s, t \in I$. The strong Markov property is in this case

$$\begin{aligned} \mathbf{P}[X_{S+t} \in A|\mathcal{F}_S] &= \mu_t(X_S, A), \\ \mathbf{E}[f(X_{S+t})|\mathcal{F}_S] &= (T_t f)(X_S) \end{aligned}$$

for all almost surely finite stopping times S, $A \in \mathcal{B}(E)$ or $f \in \mathcal{B}(E)$.

Remark 15.21 (Semigroup property)**.** *Let $(\mu^{\mathcal{X}}_t)_{t\in I}$ be the transition semigroup and $(T^{\mathcal{X}}_t)_{t\in I}$ the operator semigroup of a temporally homogeneous Markov process $\mathcal{X}$. Then, by the Chapman-Kolmogorov equations*

$$\begin{aligned} \mu^{\mathcal{X}}_s \mu^{\mathcal{X}}_t &= \mu^{\mathcal{X}}_{s+t}, \\ T^{\mathcal{X}}_s T^{\mathcal{X}}_t &= T^{\mathcal{X}}_{s+t} \end{aligned}$$

for all $s, t \in I$. For this reason, one speaks of (commutative) transition and operator semigroups.

Certain properties of operator semigroups often facilitate proofs. This leads to the concept of the Feller semigroup. To save us save paperwork, we use the distributions $\mathbf{P}_x$ from Corollary 15.18 and denote the expected value with respect to this distribution with $\mathbf{E}_x$.

Definition 15.22 (Feller semigroup, Feller process)**.** *Let $I = \mathbb{R}_+$.*

[4] A semigroup is a pair $(I, *)$, where $*$ is an associative map $I \times I \to I$

1. *Let $(T_t)_{t\in I}$ be a family of operators with $T_t : \mathbf{B}(E) \to \mathcal{B}(E)$. This is called an* operator semigroup *if $T_t(T_s f) = T_{t+s} f$ for all $f \in \mathcal{B}(E)$. Such a semigroup is called*

   (a) positive *if $T_t f \geq 0$ if $f \geq 0$ for all $t \in I$,*

   (b) contraction *if $0 \leq T_t f \leq 1$ for $0 \leq f \leq 1$ for a*

   (c) conservative *if $T_t 1 = 1$ for all $t \in I$,*

   (d) strongly continuous *if $||T_t f - f||_\infty \xrightarrow{t\to 0} 0$ for all $f \in \mathcal{C}_b(E)$.*

   (e) Feller semigroup *if $T_t f(x) \xrightarrow{t\to 0} f(x)$ for $x \in E$ and $f \in \mathcal{C}_b(E)$ and $T_t f \in \mathcal{C}_b(E)$ for all $f \in \mathcal{C}_b(E)$ and $t \in I$.*

2. *A temporally homogeneous Markov process $\mathcal{X} = (X_t)_{t\in\mathcal{I}}$ is called* Feller process *if its operator semigroup $(T_t^{\mathcal{X}})_{t\in I}$ is a Feller semigroup.*

Remark 15.23 (Probabilistic properties of Feller processes)**.** *Let $I = \mathbb{R}_+$ and $(T_t^{\mathcal{X}})_{t\in I}$ be the operator semigroup of a Markov process $\mathcal{X} = (X_t)_{t\in I}$.*

1. *The semigroup $(T_t^{\mathcal{X}})_{t\in I}$ is conservative and a positive contraction.*
   Indeed: Of course, $T_t^{\mathcal{X}} 1(x) = \mathbf{E}_x[1] = 1$, which shows the conservativeness of $(T_t^{\mathcal{X}})_{t\in I}$. Similarly, one writes for $f \in \mathcal{B}(E)$ with $0 \leq f \leq 1$

   $$T_t^{\mathcal{X}} f(x) = \mathbf{E}_x[f(x)] \leq \mathbf{E}_x[1] = 1$$

   and thus $(T_t^{\mathcal{X}})_{t\in I}$ is a contraction.

2. *Let $X_0 = x$. Then $T_t^{\mathcal{X}} f(x) \xrightarrow{t\to 0} f(x)$ for all $f \in \mathcal{C}_b(E)$ if and only if $X_t \xrightarrow{t\to 0}_p x$.*

   Indeed: '$\to$': It follows with $g(y) := r(x, y) \wedge 1$ that $\mathbf{E}_x[r(x, Y_t) \wedge 1] = T_t^{\mathcal{X}} g(x) \xrightarrow{t\to 0} g(x) = 0$, which shows the claimed convergence. '$\Leftarrow$': $X_t \xRightarrow{t\to 0} x$ applies and thus according to the definition of weak convergence for $f \in \mathcal{C}_b(E)$ in particular $T_t^{\mathcal{X}} f(x) = \mathbf{E}_x[f(X_t)] \xrightarrow{t\to\infty} \mathbf{E}_x[f(x)] = f(x)$.

Lemma 15.24 (Poisson process and Brownian motion are Feller)**.** *Both the Poisson process (with rate $\lambda \geq 0$) and the Brownian motion are Feller processes.*

Proof. Let $\mathcal{X}^x = (X_t^x)_{t\geq 0}$ be a Poisson process and $\mathcal{Y}^y = (Y_t^y)_{t\geq 0}$ a Brownian motion, each started in $x \in \mathbb{R}$ and $y \in \mathbb{R}$. The following applies $\mathcal{X}^x \stackrel{d}{=} x + \mathcal{X}^0$ and $\mathcal{Y}^y \stackrel{d}{=} y + \mathcal{Y}^0$. Then $X_t^x \sim N(x, t)$ and $Y_t^y \sim y + \text{Poi}(t\lambda)$. In particular, obviously $X_t \xrightarrow{t\to 0}_p x, Y_t \xrightarrow{t\to 0}_p$. Therefore, $T_t^{\mathcal{X}} f(x) \xrightarrow{t\to 0} f(x)$ and $T_t^{\mathcal{Y}} f(y) \xrightarrow{t\to 0} f(y)$ for $f \in \mathcal{C}_b(\mathbb{R})$ according to remark 15.23.2 Further,

$$T_t^{\mathcal{X}} f(x) = \mathbf{E}_x[f(X_t)] = \mathbf{E}_0[f(x + X_t)] \xrightarrow{x\to x'} \mathbf{E}_0[f(x' + X_t)] = T_t^{\mathcal{X}} f(x')$$

and analogously for the process $\mathcal{Y}$. From this follow all assertions. □

For concrete Markov processes, semigroups are usually difficult to specify. (However, see the exceptions of the Poisson process and the Brownian motion from example 15.4). It is easier to define what happens in an infinitesimally short time. This is described by the generator of the operator semigroup.

Definition 15.25 (generator)**.** *Let $I = [0,\infty)$, $\mathcal{X} = (X_t)_{t\in I}$ be a temporally homogeneous Markov process with operator semigroup $(T_t^{\mathcal{X}})_{t\in I}$. Then the* generator *of $\mathcal{X}$ (or of its operator semigroup) is defined as*

$$(G^{\mathcal{X}} f)(x) = \lim_{t\to 0} \frac{\mathbf{E}_x[f(X_t) - f(x)]}{t} = \lim_{t\to 0} \frac{1}{t}((T_t^{\mathcal{X}} f)(x) - f(x)),$$

for all f for which the limit value exists. The set of functions f for which $(G^{\mathcal{X}} f)(x)$ exists for all $x \in E$ exists is the domain of $G^{\mathcal{X}}$ *and is denoted by $\mathcal{D}(G^{\mathcal{X}})$.*

Example 15.26 (Generator for Poisson process and Brownian motion)**.**

1. *Let $\mathcal{X} = (X_t)_{t\in I}$ be a Poisson process with parameter λ and $G^{\mathcal{X}}$ its generator. Then,*

   $$(G^{\mathcal{X}} f)(x) = \lambda(f(x+1) - f(x)$$

   for $x \in \mathbb{N}$ and $f \in \mathcal{B}(\mathbb{N})$.

   Because we calculate, if P_t is a Poisson distributed random variable with parameter λt

   $$\begin{aligned}(G^{\mathcal{X}} f)(x) &= \lim_{t\to 0} \frac{1}{t}(\mathbf{E}_x[f(x+P_t) - f(x)]) = \lim_{t\to 0} \frac{1}{t} \sum_{k=1}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!}(f(x+k) - f(x)) \\ &= \lim_{t\to 0} \lambda \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{(k+1)!}(f(x+1+k) - f(x)) \\ &= \lambda(f(x+1) - f(x)\end{aligned}$$

   due to dominated convergence.

2. *Let $\mathcal{X} = (X_t)_{t\in I}$ be a Brownian motion and $G^{\mathcal{X}}$ its generator. Then*

   $$(G^{\mathcal{X}} f)(x) = \tfrac{1}{2} f''(x)$$

   for $x \in \mathbb{R}$ and $f \in \mathcal{C}_b^2(\mathbb{R})$, the set of bounded, twice continuously differentiable functions with with bounded derivatives.

   Because we calculate, if Z is a $N(0,1)$-distributed random variable with the Taylor approximation and a random variable Y with $|Y| \leq |Z|$

   $$\begin{aligned}(G^{\mathcal{X}} f)(x) &= \lim_{t\to 0} \frac{1}{t}(\mathbf{E}_x[f(x+\sqrt{t}Z) - f(x)]) \\ &= \lim_{t\to 0} \frac{1}{t}(\mathbf{E}_x[f'(x)\sqrt{t}Z + \tfrac{1}{2} f''(x) t Z^2 + \tfrac{1}{2}(f''(x+\sqrt{t}Y) - f''(x)) t Z^2]) \qquad (15.5) \\ &= \tfrac{1}{2} f''(x) + \lim_{t\to\infty} \mathbf{E}[\tfrac{1}{2}(f''(x+\sqrt{t}Y) - f''(x)) Z^2] = \tfrac{1}{2} f''(x)\end{aligned}$$

   by dominated convergence.

   We calculate analogously: If $\mathcal{X} = (X_t)_{t\in I}$ with $X_t = (X_t^1, ..., X_t^d)$ is a d-dimensional Brownian motion. Then,

   $$(G^{\mathcal{X}} f)(x) = \frac{1}{2} \sum_{i=1}^{d} \frac{\partial^2 f}{\partial x_i^2}(x)$$

   for $x \in \mathbb{R}^d$ and $f \in \mathcal{C}_b^2(\mathbb{R}^2)$.

Remark 15.27 (Feller semigroups and strong continuity)**.** *If E is at least locally compact, one can – if one replaces $\mathcal{C}_b(E)$ by $\mathcal{C}_0(E)$, the continuous functions vanishing at infinity – at least show that every Feller semigroup is strongly continuous. This makes it easier in some proofs to verifying the uniform convergence for strong continuity. In particular, according to Lemma 15.24 the (Feller) semigroups of the Poisson process and Brownian motion are strongly continuous.*

Lemma 15.28 (Relationship between operator semigroup and generator)**.** *Let $\mathcal{X}$ be a Feller process with operator semigroup $(T_t^{\mathcal{X}})_{t\in I}$. Further, let $G^{\mathcal{X}}$ be the generator of $\mathcal{X}$ and $\mathcal{D} \subseteq \mathcal{D}(G^{\mathcal{X}})$ with $G^{\mathcal{X}}(\mathcal{D}) \subseteq \mathcal{C}_b(E)$. For $f \in \mathcal{C}_b(E)$ is then $\int_0^t (T_s^{\mathcal{X}} f)ds \in \mathcal{D}(G^{\mathcal{X}})$ with*

$$(T_t^{\mathcal{X}} f)(x) - f(x) = \Big(G^{\mathcal{X}}\Big(\int_0^t (T_s^{\mathcal{X}} f)ds\Big)\Big)(x) \tag{15.6}$$

and for $f \in \mathcal{D}$ and $t \geq 0$ is also $T_t^{\mathcal{X}} f \in \mathcal{D}(G^{\mathcal{X}})$ and the following applies

$$\begin{aligned} G^{\mathcal{X}}(T_t^{\mathcal{X}} f) &= T_t^{\mathcal{X}}(G^{\mathcal{X}} f), \\ (T_t^{\mathcal{X}} f)(x) - f(x) &= \int_0^t (T_s^{\mathcal{X}}(G^{\mathcal{X}} f))(x)ds, \end{aligned} \tag{15.7}$$

thus

$$\mathbf{E}_x[f(X_t)] = f(x) + \int_0^t \mathbf{E}[(G^{\mathcal{X}} f)(X_s)]ds.$$

Proof. For $x \in E$ and $f \in \mathcal{C}_b(E)$, $t \mapsto (T_t^{\mathcal{X}} f)(x)$ is continuous. Because of the Feller property of $(T_t^{\mathcal{X}})_{t\in I}$,

$$(T_{t+h}^{\mathcal{X}} f)(x) = (T_t^{\mathcal{X}}(T_h^{\mathcal{X}} f))(x) = (T_t^{\mathcal{X}} f)(x).$$

For the first equation,

$$\begin{aligned} \frac{1}{h}\mathbf{E}_x\Big[\int_0^t (T_s^{\mathcal{X}} f)(X_h) - (T_s^{\mathcal{X}} f)(x)ds\Big] &= \frac{1}{h}\Big(\int_0^t (T_{s+h}^{\mathcal{X}} f)(x) - (T_s^{\mathcal{X}} f)(x)ds\Big) \\ &= \frac{1}{h}\Big(\int_h^{t+h} (T_s^{\mathcal{X}} f)(x)ds - \int_0^t (T_s^{\mathcal{X}} f)(x)ds\Big) \\ &= \frac{1}{h}\int_t^{t+h} (T_s^{\mathcal{X}} f)(x)ds - \frac{1}{h}\int_0^h (T_s^{\mathcal{X}} f)(x)ds \\ &\xrightarrow{h\to 0} (T_t^{\mathcal{X}} f)(x) - f(x). \end{aligned}$$

For the other statements, first of all

$$\begin{aligned} \frac{d}{dt}\mathbf{E}_x[f(X_t)] &= \lim_{h\to 0} \tfrac{1}{h}\mathbf{E}_x[f(X_{t+h}) - f(X_t)] \\ &= (T_t^{\mathcal{X}} \lim_{h\to 0} \tfrac{1}{h}\mathbf{E}_x[f(X_h) - f(x)] = (T_t^{\mathcal{X}}(G^{\mathcal{X}} f))(x), \end{aligned}$$

but also

$$\begin{aligned} \frac{d}{dt}\mathbf{E}_x[f(X_t)] &= \lim_{h\to 0} \tfrac{1}{h}\mathbf{E}_x[f(X_{t+h}) - f(X_t)] \\ &= \lim_{h\to 0} \tfrac{1}{h}\mathbf{E}_x[(T_t^{\mathcal{X}} f)(X_h) - (T_t^{\mathcal{X}} f)(x)] = (G^{\mathcal{X}}(T_t^{\mathcal{X}} f))(x), \end{aligned}$$

which shows the first equation. For the second equation, we note that $t \mapsto (T_t^{\mathcal{X}}(G^{\mathcal{X}} f))(x)$ is continuous according to the condition, so

$$(T_t^{\mathcal{X}} f)(x) - f(x) = \int_0^t \frac{d}{ds} \mathbf{E}_x[f(X_s)]ds = \int_0^t (T_s^{\mathcal{X}}(G^{\mathcal{X}} f))(x)ds.$$

$\square$

Corollary 15.29 (Domain is dense). *Let $\mathcal{X}$, $(T_t^{\mathcal{X}})_{t\in I}$ and $G^{\mathcal{X}}$ as in Lemma 15.28 and the conditions in Lemma 15.28 apply with $\mathcal{D} = \mathcal{C}_b(E)$. Furthermore, let $(T_t^{\mathcal{X}})_{t\in I}$ be strongly continuous. Then $\mathcal{D}(G^{\mathcal{X}})$ is dense in $\mathcal{C}_b(E)$ with respect to the supremum norm, i.e. each $f \in \mathcal{C}_b(E)$ can be approximated by functions from $\mathcal{D}(G^{\mathcal{X}})$.*

Proof. For each $f \in \mathcal{C}_b(E)$ the following applies according to the condition

$$\tfrac{1}{t} \int_0^t (T_s^{\mathcal{X}} f)ds \xrightarrow{t\to 0} f$$

with respect to the supremum norm. Since the function on the left-hand side after (15.6) lie in $\mathcal{D}(G^{\mathcal{X}})$, the assertion is shown. $\square$

Theorem 15.30 (Martingales derived from Markov processes). *Let $\mathcal{X} = (X_t)_{t\in I}$ be a Feller process with generator $G^{\mathcal{X}}$ and domain $\mathcal{D}(G^{\mathcal{X}})$. Further let $\mathcal{D} \subseteq \mathcal{D}(G^{\mathcal{X}})$ be such that $G^{\mathcal{X}}(\mathcal{D}) \subseteq \mathcal{C}_b(E)$. Then, for $f \in \mathcal{D}$ both*

$$\Big(f(X_t) - \int_0^t (G^{\mathcal{X}} f)(X_s)ds\Big)_{t\in I}$$

as well as, in the case of $(G^{\mathcal{X}} f)/f \in L$

$$\Big(f(X_t) \exp\Big(-\int_0^t \frac{(G^{\mathcal{X}} f)(X_s)}{f(X_s)}ds\Big)\Big)_{t\in I}$$

are martingales.

Proof. Let $t \geq s$. For the first process, we note

$$\begin{aligned}
\mathbf{E}\Big[f(X_t) - f(X_s) &- \int_s^t (G^{\mathcal{X}} f)(X_r)dr\Big|\mathcal{F}_s\Big] \\
&= \mathbf{E}\Big[f(X_t) - f(X_s) - \int_s^t (G^{\mathcal{X}} f)(X_r)dr\Big|X_s\Big] \\
&= (T_{t-s} f)(X_s) - f(X_s) - \int_s^t (T_{r-s}(G^{\mathcal{X}} f))(X_s)dr = 0
\end{aligned}$$

according to Lemma 15.28. Furthermore,

$$\begin{aligned}
&\mathbf{E}_x\Big[f(X_t) \exp\Big(-\int_0^t \frac{(G^{\mathcal{X}} f)(X_r)}{f(X_r)}dr\Big) - f(X_s) \exp\Big(-\int_0^s \frac{(G^{\mathcal{X}} f)(X_r)}{f(X_r)}dr\Big)\Big|\mathcal{F}_s\Big] \\
&\quad = \mathbf{E}_x\Big[f(X_t) \exp\Big(-\int_s^t \frac{(G^{\mathcal{X}} f)(X_r)}{f(X_r)}dr\Big) - f(X_s)\Big|X_s\Big] \cdot \exp\Big(-\int_0^s \frac{(G^{\mathcal{X}} f)(X_r)}{f(X_r)}dr\Big)
\end{aligned}$$

and

$$\frac{d}{dt}\mathbf{E}_{X_s}\Big[f(X_t)\exp\Big(-\int_0^t \frac{(G^{\mathcal{X}}f)(X_r)}{f(X_r)}dr\Big)\Big]$$
$$= \mathbf{E}_{X_s}\Big[(G^{\mathcal{X}}f)(X_t)\exp\Big(-\int_0^t \frac{(G^{\mathcal{X}}f)(X_r)}{f(X_r)}dr\Big)$$
$$- f(X_t)\exp\Big(-\int_0^t \frac{(G^{\mathcal{X}}f)(X_r)}{f(X_r)}dr\Big)\frac{(G^{\mathcal{X}}f)(X_t)}{f(X_t)}\Big] = 0.$$

Again, integration from s to t provides the assertion. □

Example 15.31 (Ordinary differential equation)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ with values in $\mathbb{R}^d$ which is a solution of the ordinary differential equation*

$$\frac{d}{dt}X_t = g(X_t)$$

where $g = (g_i)_{i=1,\dots,d} : \mathbb{R}^d \to \mathbb{R}^d$ is a Lipshitz function. Then $\mathcal{X}$ is deterministic, but can also be regarded as homogeneous in time (because g does not additionally depend on t) Markov process. The generator of $\mathcal{X}$ is calculated for $f \in \mathcal{C}_b^1(\mathbb{R}^d)$ and $X_0 = x$ as

$$(G^{\mathcal{X}}f)(x) = \lim_{t\to 0}\tfrac{1}{t}(f(X_t) - f(x)) = \frac{d}{dt}(f(X_t))\Big|_{t=0} = \sum_{i=1}^d \frac{\partial f}{\partial x_i}(g(x))\cdot g_i(x) = (\nabla f)(g(x))\cdot g(x).$$

Example 15.32 (Poisson process and Brownian motion)**.** *In the following, let $f_n(x) = xe^{-x/n}$, i.e. $f_n \in \mathcal{C}_b(\mathbb{R}_+)$ and $g_n(x) = x^2e^{-x/n}$ such that $f_n(x) \xrightarrow{n\to\infty} f(x)$ and $g_n(x) \xrightarrow{n\to\infty} g(x)$ with $f(x) = x$ and $g(x) = x^2$.*

1. *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Poisson process with rate λ. Thus, according to theorem 15.30 and Example 15.26*
$$\big(X_t \wedge n - \int_0^t \lambda 1_{X_s\leq n-1}ds\big)_{t\geq 0}$$
is a martingale. Since X_t is integrable, it follows from dominated convergence that
$$\big(X_t - \lambda t\big)_{t\geq 0}$$
is a martingale. Analogously, one concludes (from the integrability of X_t^2 that
$$\big(X_t^2 - \lambda\int_0^t (X_s+1)^2 - X_s^2 ds\big)_{t\geq 0}$$
is a martingale. See also example 14.46.

2. *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion. From the integrability of X_t, X_t^2 and $e^{\mu X_t}$, one concludes from Theorem 15.30 that because of $G^{\mathcal{X}}h(x) = \frac{1}{2}h''(x)$*
$$\Big(X_t - \tfrac{1}{2}\int_0^t id''(X_s)ds\Big)_{t\geq 0} = (X_t)_{t\geq 0},$$
$$\Big(X_t^2 - \tfrac{1}{2}\int_0^t (id^2)''(X_s)ds\Big)_{t\geq 0} = (X_t^2 - t)_{t\geq 0},$$
$$\Big(\exp\Big(\mu X_t - \tfrac{1}{2}\int_0^t \frac{(e^{\mu\cdot})''(X_s)}{e^{\mu X_s}}ds\Big)\Big)_{t\geq 0} = \Big(\exp\Big(\mu X_t - \tfrac{1}{2}\mu^2 t\Big)\Big)_{t\geq 0}$$
are all martingales. See also example 14.47.

Example 15.33 (Jump processes)**.** *The simplest Markov processes are piecewise constant processes. We now describe the following process: Given $X_s = x$, the process jumps after an exponentially distributed time with rate $\lambda(x)$. The process jumps according to the Markov kernel $\mu(X_s, .)$, i.e. it jumps with probability $\mu(X_s, dy)$ to y.*

Let $\lambda \in \mathcal{B}(E)$ be given with $0 \leq \lambda \leq \lambda^*$ and the Markov kernel μ from E to E. Further, let $(Y_k)_{k=0,1,2,...}$ be a Markov chain in discrete time with $\mathbf{P}(Y_{k+1} \in A|Y_k) = \mu(Y_k, A)$ for all $A \in \mathcal{B}(E)$. Furthermore, let $T_1, T_2, ...$ be independent and $\exp(1)$*-distributed. (We note that this means that T_k/λ according to* $\exp(\lambda)$ *is distributed). We define the jump process $(X_t)_{t\geq 0}$ by*

$$X_t = \begin{cases} Y_0, & t < \dfrac{T_0}{\lambda(Y_0)}, \\ Y_k, & \displaystyle\sum_{j=0}^{k-1} \frac{T_j}{\lambda(Y_j)} \leq t < \sum_{j=0}^{k} \frac{T_j}{\lambda(Y_j)}. \end{cases} \tag{15.8}$$

This is a Markov process since it is memoryless by the exponential distribution. To calculate the generator of $\mathcal{X}$, we note that the probability that in time t more than than 2 jumps take place is at most $1 - e^{-\lambda^* t}(1 + \frac{1}{2}\lambda^* t) = \mathcal{O}(t^2)$. So the following applies for $f \in \mathcal{C}_b(E)$

$$\begin{aligned} (G^{\mathcal{X}} f)(x) &= \lim_{t\to 0} \frac{\mathbf{E}_x[f(X_t) - f(x)]}{t} \\ &= \lim_{t\to 0} \tfrac{1}{t}\Big((e^{-\lambda(x)t} - 1)f(x) + \lambda(x)te^{-\lambda(x)t} \int \mu(x, dy)f(y)\Big) \\ &= \lambda(x) \int \mu(x, dy)\big(f(y) - f(x)\big)dy. \end{aligned} \tag{15.9}$$

We now give some more examples of Markov jump processes on countable state spaces.

Example 15.34 (Master equation)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a jump process on a countable state space E, given as in the last example by the functions λ and the Markov kernel $\mu(,.)$. We now set $\lambda(x, y) := \lambda(x)\mu(x, y)$ and denote this quantity as the jump rate from x to y, i.e.*

$$Gf(x) = \sum_{y\in E} \lambda(x, y)(f(y) - f(x))$$

is the generator of $\mathcal{X}$. If you insert the following into this equation function $f(y) = 1_{y=x}$ (for a fixed x) into this equation, you get

$$\begin{aligned} \frac{d}{dt}\mathbf{P}(X_t = x) &= \frac{d}{dt}\mathbf{E}[f(X_t)] = \mathbf{E}[(Gf)(X_t)] \\ &= \mathbf{E}\Big[\sum_{y\in E} \lambda(X_t, y)(1_{y=x} - 1_{X_t=x})\Big] \\ &= \sum_{z\in E} \mathbf{P}(X_t = z) \sum_{y\in E} \lambda(z, y)(1_{x=y} - 1_{x=z}) \\ &= \sum_{z\in E} \lambda(z, x)\mathbf{P}(X_t = z) - \lambda(x, z)\mathbf{P}(X_t = x). \end{aligned} \tag{15.10}$$

This equation is therefore a differential equation for $(\mathbf{P}(X_t = x))_{x\in E}$. The solution of this equation thus provides the exact distribution of X_t. This equation is also known in physics as the master equation*.*

We will now also replace the generator equation with

$$\mathbf{E}_x[f(X_h)] = f(x) + hGf(x) + o(h).$$

write.

Example 15.35 (Branching processes in continuous time)**.** *In a continuous-time branching process (with state space $\mathbb{Z}_+$), each individual dies at rate 1 and is replaced by a random number of random number of offspring (with distribution μ). Here the generator results in*

$$Gf(x) = x\sum_{n=0}^{\infty}\mu(n)(f(x-1+n) - f(x)).$$

For example, for $f_r(x) = r^x$,

$$Gf_r(x) = xr^{x-1}\sum_{n=0}^{\infty}\mu(n)(r^n - r) = xr^{x-1}(g_\mu(r) - r) = (g_\mu(r) - r)\frac{d}{dr}f_r(x).$$

From this you calculate

$$\mathbf{E}_x[r^{X_t}] = r^x + (g_\mu(r) - r)\int_0^t \frac{d}{dr}\mathbf{E}_x[r^{X_s}]ds,$$

so the function $u : (t,r) \mapsto \mathbf{E}_x[r^{X_t}]$ solves the equation

$$\frac{d}{dt}u(t,r) = (g_\mu(r) - r)\frac{d}{dr}u(t,r) \tag{15.11}$$

with the boundary conditions $u(0,r) = r^x, u(t,1) = 1$.

Example 15.36 (Yule process)**.** *The simplest branching process is the Yule process, in which each individual is replaced by two offspring. In this case $\mu = \delta_2$ and thus $g_\mu(r) = r^2$, so here in* (15.11)

$$\frac{d}{dt}u(t,r) = -r(1-r)\frac{d}{dr}u(t,r)$$

apply. We now claim that this equation in the case $x = 1$ is given by

$$u(t,r) = \frac{e^{-t}r}{1 - r(1-e^{-t})}.$$

Indeed,

$$(1 - r(1-e^{-t}))^2\frac{d}{dt}u(t,r) = -(1 - r(1-e^{-t}))re^{-t} + e^{-2t}r^2 = -r(1-r)e^{-t}$$
$$(1 - r(1-e^{-t}))^2\frac{d}{dr}u(t,r) = (1 - r(1-e^{-t}))e^{-t} + e^{-t}r(1-e^{-t}) = e^{-t}.$$

Since the generating function of the geometric distribution is just

$$g_{geo(p)}(r) = \sum_{n=1}^{\infty}(1-p)^{n-1}pr^n = \frac{pr}{1-r(1-p)}$$

we have shown that in this case $X_t \sim geo(e^{-t})$. This can also be shown using the master equation

$$\frac{d}{dt}\mathbf{P}(X_t = x) = (x-1)\mathbf{P}(X_t = x-1) - x\mathbf{P}(X_t = x).$$

This is because for $\mathbf{P}(X_t = x) = (1-e^{-t})^{x-1}e^{-t}$

$$\begin{aligned}\frac{d}{dt}(1-e^{-t})^{x-1}e^{-t} &= (x-1)(1-e^{-t})^{x-2}e^{-2t} - (1-e^{-t})^{x-1}e^{-t} \\ &= (1-e^{-t})^{x-2}e^{-t}((x-1)e^{-t} - (1-e^{-t})) = (1-e^{-t})^{x-2}e^{-t}(xe^{-t}-1)\end{aligned}$$

and

$$\begin{aligned}(x-1)\mathbf{P}(X_t = x-1) - x\mathbf{P}(X_t = x) &= (1-e^{-t})^{x-2}e^{-t}(x-1-x(1-e^{-t})) \\ &= (1-e^{-t})^{x-2}e^{-t}(xe^{-t}-1).\end{aligned}$$

Example 15.37 (Probability of extinction of a branching process)**.** *Let $T = T_0$ be the extinction time of a branching process. Then obviously $\mathbf{P}_x(T<\infty) = \mathbf{P}_1(T<\infty)^x$ and*

$$\mathbf{P}_1(T<\infty) = (1-h)\mathbf{P}_1(T<\infty) + h\sum_{n=0}^{\infty}\mu(n)\mathbf{P}_1(T<\infty)^n + o(h)$$

therefore, for $r := \mathbf{P}_1(T<\infty)$ just

$$r = g_\mu(r) \tag{15.12}$$

apply. This equation trivially has the solution $r = 1$. In the case $\sum_n n\mu(n) \le 1$ this is the only solution, which shows that the extinction probability in this case is 1. (This we have already seen through martingale theory). In the case $\mu = p\delta_0 + q\delta_2$ with $q > p$ (i.e. $\sum_n n\mu(n) > 1$) you calculate that (15.12) *applies exactly when $0 = qr^2 - r + p$, i.e. for*

$$r = \frac{1 \pm \sqrt{1-4pq}}{2q} = \frac{1 \pm 2q - 1}{2q}.$$

Since the extinction probability must be less than 1 is therefore just $(p/q) \wedge 1$.

Example 15.38 (Hitting times)**.** *Let $E' \subseteq E$ and $T := T_{E'}$ be the hitting time of E'. We want to calculate the mapping $u : x \mapsto \mathbf{E}_x[T]$. Obviously, $u(x) = \mathbf{E}_x[T] = 0$ for $x \in E'$, so with $\lambda(x) = \sum_y \lambda(x,y)$*

$$\begin{aligned}\mathbf{E}_x[T] &= (1-h\lambda(x))\mathbf{E}_x[T+h] + \sum_y \mathbf{E}_x[T|X_h = y]\cdot\mathbf{P}(X_h = y) \\ &= \mathbf{E}_x[T] + h(1-\lambda(x)\mathbf{E}_x[T] + \sum_y \lambda(x,y)\mathbf{E}_y[T] + O(h^2) \\ &= \mathbf{E}_x[T] + h(1 + G\mathbf{E}_\bullet[T]) + O(h^2).\end{aligned}$$

Therefore, the function u must fulfill the equation

$$\begin{aligned}Gu(x) &= -1, \qquad x \notin E', \\ u(x) &= 0, \qquad x \in E'.\end{aligned}$$

Example 15.39 (Birth-death processes)**.** *Markov processes with $E = \mathbb{Z}_+$ and transition rate $\lambda(x, y) = 0$ for $x - y| > 1$ are called birth-death processes. Typically one denotes*

$$\lambda(n, n+1) =: \lambda_n, \qquad \lambda(n, n-1) =: \mu_n,$$

and thus the generator is given by

$$Gf(n) = \lambda_n(f(n+1) - f(n)) + \mu_n(f(n-1) - f(n)).$$

For the expected hitting times of 0, i.e. $u(n) := \mathbf{E}_n[T_0]$, we now show that

$$\begin{aligned} u(0) &= 0, \\ u(n) &= \sum_{k=1}^{n} \frac{1}{\mu_k \pi_k} \sum_{j=k}^{\infty} \pi_j \end{aligned}$$

with $\pi_1 = 1$ and

$$\pi_i = \prod_{j=2}^{i} \frac{\lambda_{j-1}}{\mu_j}.$$

Then,

$$\begin{aligned} Gu(n) &= \lambda_n \frac{1}{\mu_{n+1}\pi_{n+1}} \sum_{j=n+1}^{\infty} \pi_j - \mu_n \frac{1}{\mu_n \pi_n} \sum_{j=n}^{\infty} \pi_j \\ &= \frac{1}{\pi_n} \sum_{j=n+1}^{\infty} \pi_j - \frac{1}{\pi_n} \sum_{j=n}^{\infty} \pi_j = -1. \end{aligned}$$

16 Properties of Brownian motion

Although we have already introduced Brownian motion in Chapter 13.3, there is still a lot of properties we have not covered yet. We already know that Brownian motion is a martingale, a Gaussian process and a strong Markov process with independent and identically distributed increments and has continuous paths. From this, we can deduce new properties, for example, Blumenthal's 0-1 law, which is an addition to the Kolmogorov's 0-1 law.

Theorem 16.1 (Blumenthal's 0-1 law)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion, defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, started in $x \in \mathbb{R}$, and $\mathcal{F}_{0+} := \bigcap_{t>0} \sigma(X_s : s \leq t)$. Then $\mathcal{F}_{0+}$ $\mathbf{P}$-trivial, i.e. $\mathbf{P}(A) \in \{0, 1\}$ for $A \in \mathcal{F}_{0+}$.*

Let further $\mathcal{T} := \bigcap_{s\geq 0} \sigma(X_t : t \geq s)$ be the terminal σ-algebra of $\mathcal{X}$. Then $\mathcal{T}$ is $\mathbf{P}$-trivial.

Proof. According to Lemma 15.9, the filtration $(\mathcal{F}_t)_{t\geq 0}$ with $\mathcal{F}_t = \sigma(X_s : s \leq t)$ is right-continuous. From the right continuity in 0 follows $\mathcal{F}_{0+} = \sigma(X_0)$. Since $X_0 = x$ is constant, $\mathcal{F}_{0+}$ must therefore be a $\mathbf{P}$-trivial σ-algebra.

Furthermore, with $\mathcal{X}$ according to theorem 13.19 also $\mathcal{X}' = (X'_t)_{t\geq 0}$ with $X'_t = tX_{1/t}$ a Brownian motion started in 0. With what has just been shown, $\bigcap_{t\geq 0} \sigma(X'_s : s \leq t)$ is $\mathbf{P}$-trivial. It follows, however, that

$$\bigcap_{s\geq 0} \sigma(X_t : t \geq s) = \bigcap_{s\geq 0} \sigma(tX_{1/t} : t \leq s) = \bigcap_{s\geq 0} \sigma(X'_t : t \leq s)$$

is $\mathbf{P}$-trivial, i.e. the assertion. $\square$

Remark 16.2. *Although Blumenthal's 0-1 law looks simple, it may nevertheless be surprising. As we will show later, the Brownian motion – in a suitable sense – can be thought of as the limit of random walks. If we start a random walk in 0, then this random walk either jumps upwards first or downwards first. In particular, for small times they spend either more time in the positive or in the negative.*

Let us define analogously for Brownian motion

$$A_t := \Big\{ \int_0^t 1_{X_s>0} ds \geq \int_0^t 1_{X_s<0} ds \Big\}$$

the set of Brownian paths that have spent more time in the positive by time t and and $A := \bigcap_{t>0} \bigcap_{0<s\leq t} A_s$, which is the set of paths that have spent more time in the positive up to some small time t. Then $A \in \rangle sF_{0+}$, so for reasons of symmetry $\mathbf{P}(A) = 0$ must apply. So there is almost certainly no Brownian path that has spent more time in the positive, for very small times.

However, this law is only the prelude to a series of further properties. Here we examine the quadratic variation in Section 16.1, the reflection principle based on the strong Markov property in Section 16.2, the law of the iterated logarithm in Section 16.3, the convergence of random walks against Brownian motion in Section 16.4 and a further connection with random walks in Section 16.5.

16.1 Quadratic variation

The paths of Brownian motion in Figures 3 and 5 look – albeit steady – very *rough.* This property should now be specified.

Definition 16.3 (Variation and quadratic variation)**.** *Let $f \in \mathcal{D}_{\mathbb{R}}([0,\infty))$, $t \geq 0$ and for $n = 1, 2, ...$ let $0 = t_{n,0} < t_{n,1} < \cdots < t_{n,k_n} = t$ be given. We denote $\zeta_n := \{t_{n,0}, ..., t_{n,k_n}\}$ as n-th partition (of $[0,t]$). Assuming $\max_k(t_{n,k} - t_{n,k-1}) \xrightarrow{n\to\infty} 0$, i.e. the partitions exploit the interval $[0,t]$ for $n \to \infty$ better and better. Then we define the ℓ-variation of f with respect to $\zeta = (\zeta_n)_{n=1,2,...}$ as*

$$\nu_{\ell,t,\zeta}(f) := \lim_{n\to\infty} \nu^n_{\ell,t,\zeta}(f)$$

with

$$\nu^n_{\ell,t,\zeta}(f) = \sum_{k=1}^{k_n} |f(t_{n,k}) - f(t_{n,k-1})|^\ell.$$

If the limit value is independent of ζ, we call this the ℓ-variation and denote it by $\nu_{\ell,t}(f)$. The 1-variation is also called variation and the 2-variation is also called quadratic variation.

In addition, ζ is called ascending *if $\zeta_n \subseteq \zeta_{n+1}$ holds for all $n = 1, 2, ...$*

Lemma 16.4 (Elementary properties of the (quadratic) variation)**.** *Let f be continuous and $t \geq 0$. Then the following applies to ζ as in Definition 16.3*

$$\nu_{\ell,t,\zeta}(f) < \infty \to \nu_{\ell+1,t,\zeta}(f) = 0,$$
$$\nu_{\ell+1,t,\zeta}(f) > 0 \Rightarrow \nu_{\ell,t,\zeta}(f) = \infty.$$

Proof. It is sufficient to show the first property. We write

$$
\begin{aligned}
0 &\le \lim_{n\to\infty} \sum_{k=1}^{k_n} |f(t_{n,k}) - f(t_{n,k-1})|^{\ell+1} \\
&\le \lim_{n\to\infty} \sup_k |f(t_{n,k}) - f(t_{n,k-1})| \cdot \lim_{n\to\infty} \sum_{k=1}^{k_n} |f(t_{n,k}) - f(t_{n,k-1})|^{\ell} = 0
\end{aligned}
$$

since f is uniformly continuous on $[0,t]$. □

Proposition 16.5 (Quadratic variation of Brownian motion)**.** *Let $\mathcal{X} = (X_t)_{t\ge 0}$ be a Brownian motion. Then for ζ as in Definition 16.3,*

$$\nu^n_{2,t,\zeta}(\mathcal{X}) \xrightarrow{n\to\infty}_{L^2} t.$$

If ζ is ascending, then also

$$\nu^n_{2,t,\zeta}(\mathcal{X}) \xrightarrow{n\to\infty}_{fs} t.$$

In particular, the variation of $\mathcal{X}$ is almost certainly infinite.

Proof. We write $\nu^n_{2,t,\zeta} := \nu^n_{2,t,\zeta}(\mathcal{X})$. First to the L^2-convergence. It is known that $X_t - X_s \sim \sqrt{t-s}X_1$ is valid for $s \le t$. Therefore

$$\mathbf{E}[\nu^n_{2,\zeta}] = \sum_{k=1}^{k_n} \mathbf{E}[(X_{t_{n,k}} - X_{t_{n,k-1}})^2] = \sum_{k=1}^{k_n} (t_{n,k} - t_{n,k-1})\mathbf{E}[X_1^2] = \sum_{k=1}^{k_n} (t_{n,k} - t_{n,k-1}) = t$$

as well as

$$\mathbf{E}[(\nu^n_{2,\zeta} - t)^2] = \mathbf{V}[\nu^n_{2,\zeta}] = \sum_{k=1}^{k_n} \mathbf{V}[(X_{n,k} - X_{n,k-1})^2] = \sum_{k=1}^{k_n} (t_{n,k} - t_{n,k-1})^2 \mathbf{E}[X_1^4] \xrightarrow{n\to\infty} 0.$$

For the almost sure convergence, we first assume wlog that there is $0 \le t_1, t_2, ... \le t$, so that $\zeta_n = \{t_1, ..., t_n\}$. We will further show that $(\nu^{-n}_{2,\zeta})_{n=...,-2,-1}$ is a (backward) martingale, so that

$$\mathbf{E}[\nu^{n-1}_{2,\zeta} - \nu^n_{2,\zeta} | \nu^n_{2,\zeta}, \nu^{n+1}_{2,\zeta}, ...] = 0$$

applies to all n. If t_n' and t_n'' are the points in time directly before and after t_n in ζ_n,

$$
\begin{aligned}
\nu^{n-1}_{2,\zeta} - \nu^n_{2,\zeta} &= (X_{t_n''} - X_{t_n'})^2 - (X_{t_n''} - X_{t_n})^2 - (X_{t_n} - X_{t_n'})^2 \\
&= 2(X_{t_n''} - X_{t_n})(X_{t_n} - X_{t_n'}).
\end{aligned}
$$

We define a second Brownian motion $(\widetilde{X}_t)_{t\ge 0}$ by an independent random variable Y with $\mathbf{P}(Y=1) = \mathbf{P}(Y=-1) = \frac{1}{2}$ and

$$\widetilde{X}_s = X_{s\wedge t_n} + Y(X_s - X_{s\wedge t_n}).$$

This means that $(\widetilde{X}_s)_{0\le s\le t}$ after t_n at X_{t_n} is mirrored. In particular, $(X_{t_n} - X_{t_n'}) = (\widetilde{X}_{t_n} - \widetilde{X}_{t_n'})$ and $(X_{t_n''} - X_{t_n}) = -(\widetilde{X}_{t_n''} - \widetilde{X}_{t_n})$. It is $\nu^k_{2,t,\zeta}(\mathcal{X}) = \nu^k_{2,t,\zeta}(\widetilde{\mathcal{X}})$ for $k = n, n+1, ...$ and thus

$$\mathbf{E}[\nu^{n-1}_{2,t,\zeta}(\mathcal{X}) - \nu^n_{2,t,\zeta}(\mathcal{X}) | \nu^n_{2,\zeta}, \nu^{n+1}_{2,\zeta}, ...] = \mathbf{E}[\nu^{n-1}_{2,t,\zeta}(\widetilde{\mathcal{X}}) - \nu^n_{2,t,\zeta}(\widetilde{\mathcal{X}}) | \nu^n_{2,\zeta}, \nu^{n+1}_{2,\zeta}, ...],$$

thus

$$\begin{aligned}\mathbf{E}[\nu_{2,t,\zeta}^{n-1}(\mathcal{X}) - \nu_{2,t,\zeta}^{n}(\mathcal{X})|\nu_{2,\zeta}^{n}, \nu_{2,\zeta}^{n+1}, ...]\\ &= \tfrac{1}{2}\big(\mathbf{E}[\nu_{2,t,\zeta}^{n-1}(\mathcal{X}) - \nu_{2,t,\zeta}^{n}(\mathcal{X})|\nu_{2,\zeta}^{n}, \nu_{2,\zeta}^{n+1}, ...] + \mathbf{E}[\nu_{2,t,\zeta}^{n-1}(\widetilde{\mathcal{X}}) - \nu_{2,t,\zeta}^{n}(\widetilde{\mathcal{X}})|\nu_{2,\zeta}^{n}, \nu_{2,\zeta}^{n+1}, ...]\big)\\ &= \mathbf{E}[(X_{t_n''} - X_{t_n})(X_{t_n} - X_{t_n'}) + (\widetilde{X}_{t_n''} - \widetilde{X}_{t_n})(\widetilde{X}_{t_n} - \widetilde{X}_{t_n'})|\nu_{2,\zeta}^{n}, \nu_{2,\zeta}^{n+1}, ...] = 0,\end{aligned}$$

which shows the desired martingale property. According to Theorem 14.37, $(\nu_{2,t,\zeta}^{n})_{n=1,2,...}$ converges almost surely towards t. □

Corollary 16.6 (Brownian motion has nowhere differentiable paths)**.** *A Brownian motion $\mathcal{X} = (X_t)_{t\geq 0}$ almost certainly has nowhere differentiable paths. This means that*

$$\mathbf{P}\Big(\lim_{h\to 0}\frac{X_{t+h}-X_t}{h} \text{ exists for some } t>0\Big) = 0.$$

Proof. It is sufficient to consider the set of paths of Brownian motion whose quadratic variation in time $[0,t]$ is exactly t. (The set of these paths has probability 1, as Proposition 16.5 shows). Each path in this set has positive quadratic variation in every small time interval, i.e. according to Lemma 16.4 infinite variation. Since differentiability requires at least a finite variation in a small time interval the assertion follows. □

16.2 Strong Markov property and reflection principle

In Example 15.13 we saw that Brownian motion is a strong Markov process. This has some useful consequences, as we will now see. The reflection principle is illustrated in Figure 6.

Lemma 16.7 (Reflection principle)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion and T is a stopping time. Then the reflected process is $\mathcal{X}' = (X'_t)_{t\geq 0}$ with*

$$X'_t := X_{t\wedge T} - (X_t - X_{t\wedge T}) = \begin{cases} X_t, & t\leq T,\\ 2X_T - X_t, & t > T\end{cases}$$

is also a Brownian motion.

Proof. First of all, it is clear from the construction that $\mathcal{X}'$ has continuous paths. Wlog, we assume that $T<\infty$ holds. We define $\mathcal{Y} = (Y_t)_{t\geq 0}$ by $Y_t := X_{t\wedge T}$ and $\mathcal{Z} = (Z_t)_{t\geq 0}$ by $Z_t := X_{T+t} - X_T$. Then $\mathcal{Z}$ is a Brownian motion, since by the strong Markov property, $(T,\mathcal{Y})$ is independent. This means that $(T,\mathcal{Y},\mathcal{Z}) \stackrel{d}{=} (T,\mathcal{Y},-\mathcal{Z})$, since $\mathcal{Z} \stackrel{d}{=} -\mathcal{Z}$. It also follows that $(\mathcal{Y},\mathcal{Z}^T) \stackrel{d}{=} (\mathcal{Y},-\mathcal{Z}^T)$ with $\mathcal{Z}^T := (Z_t^T)_{t\geq 0}$, $Z_t^T := Z_{(t-T)^+}$. From this,

$$\mathcal{X} = \mathcal{Y} + \mathcal{Z}^T \stackrel{d}{=} \mathcal{Y} - \mathcal{Z}^T = \mathcal{X}'.$$

This shows the assertion. □

As an application of the reflection principle, we now calculate the distribution of the maximum of a Brownian motion up to a time t. First, however, we note that from Doob's L^p inequality, Proposition 14.26, estimates about the distribution of the maximum. Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion and $\mathcal{M} = (M_t)_{t\geq 0}$ with $M_t = \sup_{s\leq t} X_s$ the maximum process. Then

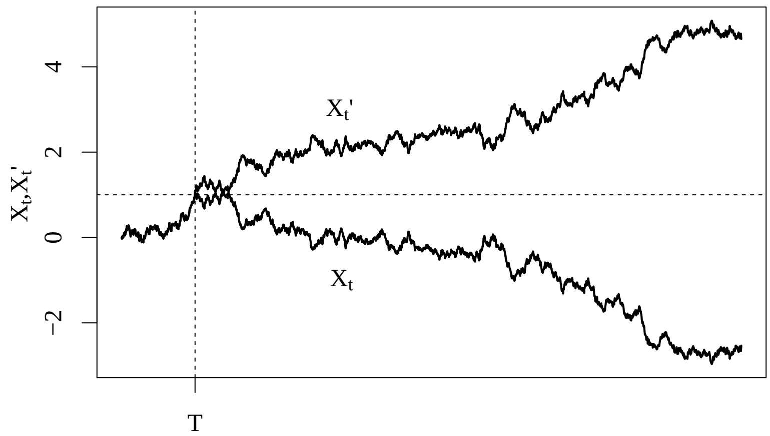

Figure 6:
The reflection principle of Brownian motion states that for a Brownian motion $(X_t)_{t\geq 0}$ the process reflected to T at $x = X_T$ $(X_t')_{t\geq 0}$ is also a Brownian motion.

it follows from Proposition 14.26 (or the extension to continuous-time processes from Theorem 14.51) with $p = 2$

$$\mathbf{P}(M_t \geq x) \leq \tfrac{1}{x^2}\mathbf{E}[X_t^2] = \tfrac{t}{x^2}.$$

However, especially for large x, this probability is in fact much smaller, as the next result shows.

Theorem 16.8 (Maximum of Brownian motion)**.** *Let $\mathcal{X} = (X_t)_{t\geq 0}$ be a Brownian motion started in $X_0 = 0$. We define the maximum process $\mathcal{M} = (M_t)_{t\geq 0}$ by $M_t = \sup_{0\leq s\leq t} X_s$. Then,*

$$M_t \stackrel{d}{=} M_t - X_t \stackrel{d}{=} |X_t|.$$

All three random variables have the density

$$x \mapsto \sqrt{\frac{2}{\pi t}} \exp\Big(- \frac{x^2}{2t}\Big) 1_{x\geq 0}.$$

Proof. Let $\varphi_t(x) = \frac{1}{\sqrt{2\pi t}} \exp\big(- \frac{x^2}{2t}\big)$ the density of Brownian motion at time t. Then the density of $|X_t|$ is given by $2\varphi_t(x)1_{x\geq 0}$. So it remains to show that both M_t and $M_t - X_t$ have exactly this density. For this we set $T := T_x = \inf\{s \geq 0 : X_s = x\}$. For $0, y \leq x$, because of Lemma 16.7, if $(X_t')_{t\geq 0}$ is the process mirrored at T,

$$\mathbf{P}(M_t \geq x, X_t \leq y) = \mathbf{P}(X_t' \geq 2x - y) = \int_{2x-y}^{\infty} \varphi_t(z)dz$$

and thus for $x \geq 0$

$$\begin{aligned}\mathbf{P}(M_t \geq x) &= \mathbf{P}(M_t \geq x, X_t \leq x) + \mathbf{P}(X_t \geq x)\\ &= 2\int_x^{\infty} \varphi_t(z)dz\end{aligned}$$

from which it follows that $M_t \stackrel{d}{=} |X_t|$. We further calculate

$$\begin{aligned}
\mathbf{P}(M_t - X_t \geq x) &= \lim_{\varepsilon \to 0} \tfrac{1}{\varepsilon} \int_0^\infty \mathbf{P}(z \leq M_t \leq z + \varepsilon, X_t \leq z - x) dz \\
&= \lim_{\varepsilon \to 0} \tfrac{1}{\varepsilon} \int_0^\infty \mathbf{P}(M_t \geq z, X_t \leq z - x) - \mathbf{P}(M_t \geq z + \varepsilon, X_t \leq z - x) dz \\
&= \lim_{\varepsilon \to 0} \tfrac{1}{\varepsilon} \int_0^\infty 2\varphi_t(z + x) dz = \int_x^\infty 2\varphi(z) dz.
\end{aligned}$$

Again, $M_t - X_t \stackrel{d}{=} |X_t|$ applies. □

Remark 16.9 (The path-valued reflection principle)**.** *The reflection principle only shows the equality of the distributions of $|X_t|, M_t, M_t - X_t$ at a fixed time t. It now remains open whether $(|X_t|)_{t\geq 0} \sim (M_t - X_t)_{t \geq 0}$ is also valid. Even if we do not show this here, this assertion turns out to be correct. (By the way: Surely $(M_t)_{t\geq 0}$ is distributed differently than $(|X_t|)_{t\geq 0}$ or $(M_t - X_t)_{t\geq 0}$, since the last two processes can also decrease, but $(M_t)_{t \geq 0}$ not).*

16.3 The Law of the Iterated Logarithm

We want to determine how a Brownian motion $\mathcal{X} = (X_t)_{t \geq 0}$ maximally *grows*. This means that we have a function $t \mapsto h_t$ so that

$$0 < \limsup_{t \to \infty} \frac{X_t}{h_t} < \infty. \tag{16.1}$$

We already know from the law of large numbers that $\frac{X_t}{t} \xrightarrow{t \to \infty} 0$. The following also applies

$$\limsup_{t \to \infty} \frac{X_t}{\sqrt{t}} = \infty. \tag{16.2}$$

Indeed: Certainly $\limsup_{t\to\infty} \frac{X_t}{\sqrt{t}}$ is measurable with respect to the terminal σ-algebra of $\mathcal{X}$, i.e. according to Theorem 16.1 almost certainly constant. Suppose, $\limsup_{t\to\infty} \frac{X_t}{\sqrt{t}} \xrightarrow{t\to\infty} \gamma$ for a $0 < \gamma < \infty$. Then it would apply in particular that $\mathbf{P}(\frac{X_t}{\sqrt{t}} > 2\gamma) \xrightarrow{t \to \infty} 0$, in contradiction to the central limit theorem.

The task now is to find a function $t \mapsto h_t$ with $\sqrt{t} \leq h_t \leq t$ so that (16.1) applies. This is determined by the *iterated logarithm* as follows:

Theorem 16.10 (Iterated logarithm for Brownian motion)**.** *Let $\mathcal{X} = (X_t)_{t \geq 0}$ be a Brownian motion. Then*

$$\limsup_{t \to \infty} \frac{X_t}{\sqrt{2t \log \log t}} = \limsup_{t \to 0} \frac{X_t}{\sqrt{2t \log \log 1/t}} = 1, \tag{16.3}$$

almost surely.

Remark 16.11. *For reasons of symmetry, i.e. because $-\mathcal{X}$ is also a Brownian motion,*

$$\liminf_{t \to \infty} \frac{X_t}{\sqrt{2t \log \log t}} = \liminf_{t \to 0} \frac{X_t}{\sqrt{2t \log \log 1/t}} = -1$$

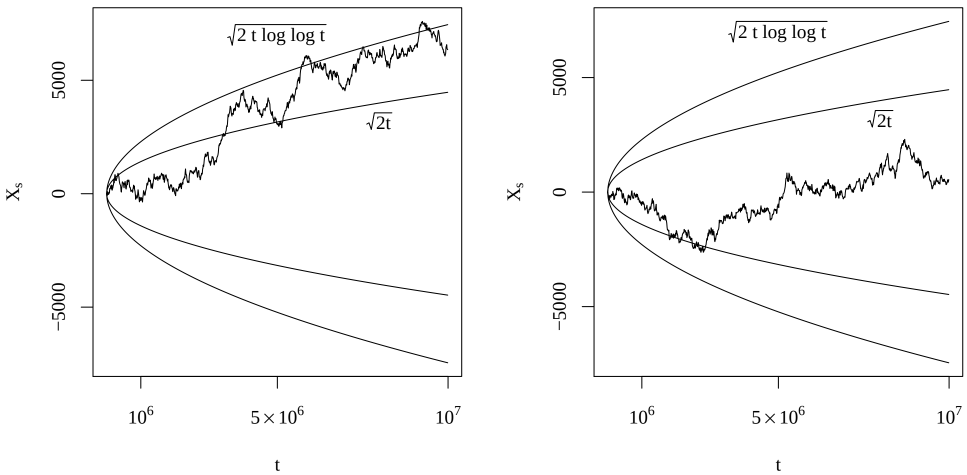

Figure 7:
Here are two paths of a Brownian movement are given. As you can see, the two paths leave the curves $t \mapsto \pm\sqrt{2t}$ much more frequently than the curve $t \mapsto \pm h_t$.

almost surely. For illustration see Figure 7. The fact that $h_t := \sqrt{2t \log\log t}$ is the correct function means that almost every path of the Brownian motion is only finitely often outside the two curves $t \mapsto \pm h_t$ but infinitely often outside the two curves $t \mapsto \pm(1-\varepsilon)h_t$, where $0 < \varepsilon < 1$ is arbitrary.

Proof. First of all, we note that with Theorem 13.19 also $(tX_{1/t})_{t\geq 0}$ is also a Brownian motion. If we apply the statement for the $t \to \infty$ limit, it follows that

$$\limsup_{t\to 0} \frac{X_t}{\sqrt{2t \log\log 1/t}} = \limsup_{t\to\infty} \frac{X_{1/t}}{\sqrt{2\frac{1}{t}\log\log t}} = \limsup_{t\to\infty} \frac{tX_{1/t}}{\sqrt{2t\log\log t}} = 1$$

almost surely. In addition, we write $h_t := h(t) := \sqrt{2t\log\log t}$. The proof for $t \to \infty$ requires a few estimations. We divide the proof into three steps.

Step 1: Estimation of the normal distribution: Let $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ the density of X_1. Then

$$\mathbf{P}(X_1 > x) \leq \tfrac{1}{x}\varphi(x), \tag{16.4}$$

$$\mathbf{P}(X_1 > x) \geq \tfrac{x}{1+x^2}\varphi(x), \tag{16.5}$$

Indeed: $\varphi'(y) = -y\varphi(y)$ and therefore

$$\varphi(x) = \int_x^\infty y\varphi(y)dy \geq x\int_x^\infty \varphi(y)dy = x \cdot \mathbf{P}(X > x),$$

which shows (16.4). For (16.5) we write, quite similarly, $\left(\frac{\varphi(y)}{y}\right)' = -\frac{1+y^2}{y^2}\varphi(y)$, and thus

$$\frac{\varphi(x)}{x} = \int_x^\infty \frac{1+y^2}{y^2}\varphi(y)dy \leq \frac{1+x^2}{x^2}\int_x^\infty \varphi(y)dy = \frac{1+x^2}{x^2}\cdot \mathbf{P}(X > x).$$

In the following we write $a(x) \overset{x\to\infty}{\approx} b(x)$, if $\frac{a(x)}{b(x)} \xrightarrow{x\to\infty} 1$ applies. So, for example, according to what has just been shown

$$\mathbf{P}(X_t > x\sqrt{t}) \overset{x\to\infty}{\approx} \tfrac{1}{x}\varphi(x).$$

2nd step: upper estimate: According to Theorem 16.8 is for $x > 0$

$$\mathbf{P}(\sup_{0\le s\le t} X_s > x\sqrt{t}) = 2\cdot\mathbf{P}(X_t > x\sqrt{t}) \overset{x\to\infty}{\approx} \tfrac{2}{x}\varphi(x).$$

Now let $r > 1$. We first notice

$$h(r^{n-1}) = \sqrt{\frac{2(\log(n-1)+\log\log r)}{r}}\sqrt{r^n} \overset{n\to\infty}{\approx} \sqrt{\frac{2\log n}{r}}\sqrt{r^n}$$

Now for $c > 0$ with the last two estimates

$$\begin{aligned}\mathbf{P}\big(\sup_{0\le t\le r^n} X_t > ch(r^{n-1})\big) &\overset{n\to\infty}{\approx} 2\cdot\mathbf{P}\Big(X_{r^n} > c\sqrt{\frac{2\log n}{r}}\sqrt{r^n}\Big)\\ &\overset{n\to\infty}{\approx} \frac{1}{c}\sqrt{\frac{2r}{\log n}}\varphi(c\sqrt{2\log n^{1/r}}) \\ &\overset{n\to\infty}{\approx} \frac{1}{c}\sqrt{\frac{r}{\pi\log n}}\frac{1}{n^{c^2/r}}.\end{aligned} \tag{16.6}$$

Therefore, for $c > 1$ and $1 < r < c^2$, the right-hand side of the last equation is summable, so the following follows with the Borel-Cantelli lemma

$$\mathbf{P}\Big(\limsup_{t\to\infty}\frac{X_t}{h_t} \ge c\Big) \le \mathbf{P}\big(\sup_{0\le t\le r^n} X_t > ch_{r^{n-1}} \text{ for infinitely many } n\big) = 0.$$

Thus '$\le$' follows in (16.3).

3rd step: lower estimate: Let $r > 1$ (typically large) and $c > 0$ (typically close to 1). Define the events

$$A_n := \{X_{r^n} - X_{r^{n-1}} > ch(r^n - r^{n-1})\}.$$

Since $X_{r^n} - X_{r^{n-1}} \sim N(0, r^n - r^{n-1})$, the following applies according to Step 1

$$\begin{aligned}\mathbf{P}(A_n) &= \mathbf{P}\Big(\frac{X_{r^n} - X_{r^{n-1}}}{\sqrt{r^n - r^{n-1}}} > c\frac{h(r^n - r^{n-1})}{\sqrt{r^n - r^{n-1}}}\Big)\\ &= \mathbf{P}\big(X_1 > c\sqrt{2\log\log(r^n - r^{n-1})}\big)\\ &\overset{n\to\infty}{\approx} \frac{1}{c}\frac{1}{\sqrt{4\pi\log\log(r^n - r^{n-1})}}\exp\big(-c^2\log\log(r^n - r^{n-1})\big)\\ &\overset{n\to\infty}{\approx} \frac{1}{c}\frac{1}{\sqrt{4\pi\log n}}\frac{1}{n^{c^2}}\end{aligned}$$

If $c < 1$, these probabilities cannot be summed up in n. Since the events $A_1, A_2, \ldots$ are independent, according to the Borel-Cantelli lemma, an infinite number of A_n occur. Thus, for an infinite number of n, if $c < 1$

$$X_{r^n} > ch(r^n - r^{n-1}) + X_{r^{n-1}}.$$

According to the '$\le$' direction, $X_{r^{n-1}} > -2h(r^{n-1})$ for almost all n, i.e. $\liminf_{n\to\infty} \frac{X_{r^{n-1}}}{h(r^n)} \le -\liminf_{n\to\infty} \frac{h(r^{n-1})}{h(r^n)} = -\frac{1}{\sqrt{r}}$ is almost certain. Further, $h(r^n - r^{n-1})/h(r^n) \xrightarrow{n\to\infty} 1$ and thus

$$\limsup_{t\to\infty} \frac{X_t}{h_t} \ge \limsup_{n\to\infty} \frac{X_{r^n}}{h(r^n)} \ge \limsup_{n\to\infty} \frac{X_{r^n} - X_{r^{n-1}}}{h(r^n - r^{n-1})} - \frac{1}{\sqrt{r}} \ge c - \frac{1}{\sqrt{r}}.$$

Since $0 < c < 1$ and $r > 0$ were arbitrary, '$\ge$' follows in (16.3). $\square$

16.4 Donsker's Theorem

Brownian motion $\mathcal{X} = (X_t)_{t\ge0}$ is a stochastic process with continuous paths. Paul Lévy considered approximated Brownian motion as the path of a random walk, where

$$X_{t+dt} - X_t = \pm\sqrt{dt}, \text{ each with probability } \tfrac{1}{2}.$$

(Of course, this can only be a formal representation, after all it is unclear what $\sqrt{dt}$ is supposed to be). Donsker's Theorem presented here makes the connection between random walks and Brownian motion. It asserts the convergence of random walks against Brownian motion in distribution.

Remark 16.12 (Random walks and Brownian motion)**.** *In this section, $Y_1, Y_2, \dots$ are independent and identically distributed random variables with $\mathbf{E}[Y_1] = 0$ and $\mathbf{V}[Y_1] = \sigma^2$ and $\widetilde{X}_{n,t} := \frac{Y_1+\cdots+Y_{\lfloor nt\rfloor}}{\sqrt{n\sigma^2}}$ for $t \ge 0$ and $\widetilde{\mathcal{X}}_n = (\widetilde{X}_{n,t})_{t\ge0}$. We know from the central limit theorem that for $t > 0$*

$$\widetilde{X}_{n,t} \xRightarrow{n\to\infty} X_t,$$

where $X_t \sim N(0,t)$ is distributed. Analogously, for $0 < t_1 < \cdots < t_k < \infty$

$$\big(\widetilde{X}_{n,t_1}, \widetilde{X}_{n,t_2} - \widetilde{X}_{n,t_1}, \dots, \widetilde{X}_{n,t_k} - \widetilde{X}_{n,t_{k-1}}\big) \xRightarrow{n\to\infty} (X_{t_1}, X_{t_2} - X_{t_1}, \dots, X_{t_k} - X_{t_{k-1}}),$$

if $(X_{t_1}, \dots, X_{t_k})$ is Brownian motion $\mathcal{X}$ at the points in time $t_1, .., t_k$. Does this now mean already the convergence of the random walks against the Brownian motion, therefore $\mathcal{X}_n \xRightarrow{n\to\infty} \mathcal{X}$? No! For this convergence, we must use both $\mathcal{X}_n$ and $\mathcal{X}$ as random variables with values in a topological space – let's call it $\mathcal{C}$ – where the convergence in distribution is based on the convergence of expected values with respect to continuous, bounded functions $f : \mathcal{C} \to \mathbb{R}$. However, for the uncountable product space, the σ-algebra $\mathcal{B}(\mathbb{R})^{\otimes[0,\infty)}$ is not the Borel σ-algebra on the product space, and we have developed the theory of weak convergence only for the case of probability measures on a Borel's σ-algebra.

In order to formulate the convergence in distribution against the Brownian motion we first need a suitable state space. This is defined as $\mathcal{C} := \mathcal{C}_{\mathbb{R}}([0,\infty))$, provided with the topology of compact convergence (see Definition 16.13). In order to convergence in this space, we define the linear interpolation of the processes $\widetilde{\mathcal{X}}_n$ so that their paths are also continuous. For this we set

$$X_{n,t} := \widetilde{X}_{n,t} + (nt - \lfloor nt\rfloor)\frac{Y_{\lfloor nt\rfloor+1}}{\sqrt{n\sigma^2}}. \tag{16.7}$$

and $\mathcal{X}_n = (X_{n,t})_{t\ge0}$. Now it makes sense to ask whether

$$\mathcal{X}_n \xRightarrow{n\to\infty} \mathcal{X}$$

applies, whereby here the weak convergence with respect to the distributions on $\mathcal{B}(\mathcal{C}_{\mathbb{R}}([0,\infty)))$ *is meant here.*

is

Definition 16.13 (Uniform convergence on compacta)**.** *Let* (E, r) *be a metric space. For* $f, f_1, f_2, ... \in \mathcal{C}_E([0,\infty))$ *let* $f_n \xrightarrow{n\to\infty} f$ uniform on compacta *if and only if* $\sup_{0\le s\le t} r(f_n(s), f(s)) \xrightarrow{n\to\infty} 0$ *for all* $t > 0$*.*

Lemma 16.14 ($\mathcal{C}_E([0,\infty))$ is Polish)**.** *Let* E *be Polish with complete metric* r*. Then the topology of uniform convergence on compacta on* $\mathcal{C}_E([0,\infty))$ *is separable. Moreover, defined*

$$r_{\mathcal{C}}(f,g) := \int_0^\infty e^{-t} \cdot (1 \wedge \sup_{0\le s\le t} |r(f(s), g(s))|)dt$$

a complete metric on $\mathcal{C}_E([0,\infty))$*, which induces this topology induces. In particular,* $\mathcal{C}_E([0,\infty))$ *is Polish.*

Proof. To show separability, it is sufficient to name a countable class of functions that every function in $\mathcal{C}_E([0,\infty))$ can be locally approximated by such functions on compacta. For this purpose, let $D \subseteq E$ be dense and countable. For every finite sequence $x_1, ..., x_n \in D$ and $t_1, ..., t_n$ let $f = f_{x_1,...,x_n,t_1,...,t_n}$ be a continuous function with $f(t_i) = x_i$. Then $\bigcup_n \{f_{x_1,...,x_n,t_1,...,t_n} : x_1, ..., x_n \in D, t_1, ..., t_n \ge 0\}$ is countable and dense in $\mathcal{C}_E([\infty))$.

Now to the metric. Since $t \mapsto \sup_{0\le s\le t} r(f(s), g(s)) \wedge 1$ is monotonically increasing, $r_{\mathcal{C}}(f_n, f) \xrightarrow{t\to\infty} 0$ holds if and only if $\sup_{0\le s\le t} r(f_n(s), f(s)) \xrightarrow{n\to\infty} 0$ for all t is valid. But this is exactly the compact convergence. Let further $f_1, f_2, ...$ be a Cauchy sequence with respect to $r_{\mathcal{C}}$. Then for every $t > 0$ the sequence $f_1, f_2, ...,$ restricted to $[0, t]$ is a Cauchy sequence with respect to the supremum norm on $[0, t]$, i.e. uniformly convergent on $[0, t]$. The assertion now follows by means of of a diagonal sequence argument. ☐

First, we define two types of convergence of stochastic processes that we have just learned about.

Definition 16.15 (Convergence of stochastic processes)**.** *Let* $\mathcal{X} = (X_t)_{t\ge0}, \mathcal{X}^1 = (X_t^1)_{t\ge0}, \mathcal{X}^2 = (X_t^2)_{t\ge0}, ...$ *stochastic processes with state space* E*.*

1. *For each choice of* $t_1, ..., t_k, k = 1, 2, ...,$ *it holds that*

$$(X_{t_1}^n, ..., X_{t_k}^n) \xRightarrow{n\to\infty} (X_{t_1}, ..., X_{t_k}),$$

   we say that the finite-dimensional distributions of $\mathcal{X}^1, \mathcal{X}^2, ...$ *converge to those of* $\mathcal{X}$ *converge and write*

$$\mathcal{X}^n \underset{fdd}{\xRightarrow{n\to\infty}} \mathcal{X}.$$

   (Here fdd *stands for* finite dimensional distributions*).*

2. *If the processes* $\mathcal{X}, \mathcal{X}^1, \mathcal{X}^2, ...$ *have paths in* $\mathcal{C}_E([0,\infty))$ *and*

$$\mathcal{X}^n \xRightarrow{n\to\infty} \mathcal{X},$$

   where we use $\mathcal{X}, \mathcal{X}^1, \mathcal{X}^2, ...$ *as the random variable in* $\mathcal{C}_E([0,\infty))$*, we say that* $\mathcal{X}^1, \mathcal{X}^2, ...$ *converges in distribution against* $\mathcal{X}$*.*

The fdd convergence is weaker than the weak convergence of processes. However, if the processes are tight (see Definition 9.14), both terms coincide.

Proposition 16.16 (Weak and fdd convergence)**.** *Let $\mathcal{X}, \mathcal{X}^1, \mathcal{X}^2, ...$ be random variables with values in $\mathcal{C}_E([0,\infty))$. Then are equivalent*

1. $\mathcal{X}^n \xRightarrow{n\to\infty} \mathcal{X}$.
2. $\mathcal{X}^n \xRightarrow[fdd]{n\to\infty} \mathcal{X}$ *and* $\{\mathcal{X}^n : n = 1, 2, ...\}$ *is tight in* $\mathcal{C}_E([0,\infty))$.

Proof. '1.→2.': First, from the weak convergence, according to Corollary 9.18 the tightness of $\{\mathcal{X}^n : n = 1, 2, ...\}$ follows. Furthermore, the mappings $f \mapsto (f(t_1), ..., f(t_k))$ are continuous for $t_1, ..., t_k \in [0,\infty)$, so the fdd convergence follows according to Theorem 9.10.

'2.⇒1.': We define the function class

$$\mathcal{M} := \{f \mapsto \varphi(f(t_1), ..., f(t_k)) : t_1, ..., t_k \in [0,\infty), \varphi \in \mathcal{C}_b(E^k)\} \subseteq \mathcal{C}_b(\mathcal{C}_E([0,\infty)))).$$

It is clear that the fdd convergence $\mathcal{X}^n \xRightarrow[fdd]{n\to\infty} \mathcal{X}$ is equivalent to $\mathbf{E}[\varphi(\mathcal{X}^n)] \xrightarrow{n\to\infty} \mathbf{E}[\varphi(\mathcal{X})]$ for all $\varphi \in \mathcal{M}$. Furthermore $\mathcal{M}$ is an algebra and separates points, according to Theorem 9.24 is therefore separating. Now follows the weak convergence follows from Proposition 9.27. □

To show the convergence of processes, after Proposition 16.16 both the convergence of the finite-dimensional distributions as well as the tightness must be shown. In applications, the verification of tightness is usually non-trivial. In particular, one needs to understand how (relatively) compact subsets of $\mathcal{C}_E([0,\infty))$ can be characterized. This is done using the theorem of Arzela-Ascoli's theorem, which is based on the modulus of continuity.

Definition 16.17 (Modulus of continuity)**.** *For $f \in \mathcal{C}_E([0,\infty))$ we define the* modulus of continuity

$$w(f, \tau, h) := \sup\{r(f(s), f(t)) : s, t \le \tau, |t - s| \le h\}.$$

Theorem 16.18 (Arzela-Ascoli)**.** *A set $A \subseteq \mathcal{C}_E([0,\infty))$ is relatively compact if and only if $\{f(t) : f \in A\}$ for all $t \in \mathbb{Q}_+ := [0,\infty) \cap \mathbb{Q}$ is relatively compact in E and for all $\tau > 0$*

$$\lim_{h\to 0} \sup_{f\in A} w(f, \tau, h) = 0. \tag{16.8}$$

Proof. First, let A be relatively compact. Then $\{f(t) : t \in A\}$ must be relatively compact for all $t \ge 0$, otherwise it would be easy to construct a divergent sequence. Furthermore, A is according to Proposition A.9 totally bounded. Further, let $\tau > 0$, $\varepsilon > 0$ and $f_1, ..., f_N$, so that $A \subseteq \bigcup_{i=1}^N B_{\varepsilon/3}(f_i)$. Since $f_1, ..., f_N$ is based on $[0, \tau]$ are uniformly continuous, there is an $h > 0$ with

$$0 \le s, t \le \tau, |t - s| < h \qquad \Longrightarrow \qquad r(f_i(t), f_i(s)) \le \varepsilon/3, \qquad i = 1, ..., N.$$

So, for every $f \in A$ and $s, t \le \tau, |t - s| \le h$, that

$$r(f(s), f(t)) \le \min_{i=1,...,N} r(f(s), f_i(s)) + r(f_i(s), f_i(t)) + r(f_i(t), f(t)) \le \varepsilon$$

and thus

$$w(f,\tau,h) = \sup\{r(f(t),f(s)) : s,t \le \tau, |t-s| \le h\} \le \varepsilon,$$

independent of f. From this follows (16.8).

Conversely, (16.8) applies. It suffices to show that every sequence in A has a subsequence that is Cauchy. By the relative compactness of $\{f(t) : f \in A\}$ for $t \in \mathbb{Q}_+$ it is clear that for each sequence there is a subsequence $f_1, f_2, ...$ such that $f_1(t_i), f_2(t_i), ...$ for all $t_i \in \mathbb{Q}_+$ is a Cauchy sequence (i.e. convergent). Now let $\varepsilon > 0$. According to the condition there is an $h > 0$, so that from $|t-s| \le h$ and $f \in A$ it follows that $r(f(s),f(t)) \le \varepsilon/3$ holds. Further, let $M = \lceil \tau/h \rceil$ and $0 = t_1, ..., t_M \in \mathbb{Q}_+$, so that $|t_{i+1} - t_i| \le h$, $i = 1, ..., M-1$ and $t_M \ge \tau$. Further there is an N such that from $n, m > N$ it follows that $\sup_{t=t_1,...,t_M} r(f_n(t), f_m(t)) \le \varepsilon/3$. Thus, for $0 \le s \le t$

$$r(f_n(s), f_m(s)) \le r(f_n(s), f_n(t_{\lceil s/h \rceil})) + r(f_n(t_{\lceil s/h \rceil}), f_m(t_{\lceil s/h \rceil})) + r(f_m(t_{\lceil s/h \rceil}), f_m(s)) \le \varepsilon.$$

It follows that $f_1, f_2, ...$ is a Cauchy sequence with respect to compact convergence on $[0,t]$, i.e. it converges on this range converges uniformly. A diagonal sequence argument extends this statement to compact convergence. □

Theorem 16.19 (Tightness in $\mathcal{C}_{\mathbb{E}}([0,\infty))$)**.** *Let $\mathcal{X}, \mathcal{X}^1, \mathcal{X}^2, ...$ be random variables with values in $\mathcal{C}_E([0,\infty))$. Then $\mathcal{X}^n \xRightarrow{n\to\infty} \mathcal{X}$ iff $\mathcal{X}^n \xRightarrow[fdd]{n\to\infty} \mathcal{X}$ and*

$$\lim_{h\to 0} \limsup_{n\to\infty} \mathbf{E}[w(\mathcal{X}^n, \tau, h) \wedge 1] = 0 \tag{16.9}$$

for all $\tau > 0$.

Proof. According to Proposition 16.16 it suffices to show that (16.9) is equivalent to the tightness of the family $(\mathcal{X}^n)_{n=1,2,...}$.

First, let $(\mathcal{X}^n)_{n=1,2,...}$ be tight and $\varepsilon > 0$. Then there is a compact set $K \subseteq \mathcal{C}_E([0,\infty))$ such that $\limsup_{n\to\infty} \mathbf{P}(\mathcal{X}^n \notin K) \le \varepsilon$. For $\tau > 0$ you can use the Arzela-Ascoli Theorem h can be chosen small enough so that $w(f,\tau,h) \le \varepsilon$ applies to $f \in K$. This means that

$$\limsup_{n\to\infty} \mathbf{E}[w(\mathcal{X}^n,\tau,h) \wedge 1] \le \varepsilon + \sup_{n=1,2,...} \mathbf{P}[w(\mathcal{X}^n,\tau,h) > \varepsilon] \le 2\varepsilon,$$

from which (16.9) follows.

Conversely, (16.9) and $\mathcal{X}^n \xRightarrow[fdd]{n\to\infty} \mathcal{X}$. The mapping w is increasing in h and $w(\mathcal{X}^n,\tau,h) \xrightarrow{h\to 0} 0$ almost certainly for $n = 1, 2, ...$ So $\lim_{h\to 0} \sup_{n=1,2,...} \mathbf{E}[w(\mathcal{X}^n,\tau,h) \wedge 1] = \lim_{h\to 0} \sup_{n=k,k+1,...} \mathbf{E}[w(\mathcal{X}^n,\tau,h) \wedge 1]$ for all k, i.e. also $\lim_{h\to 0} \sup_{n=1,2,...} \mathbf{E}[w(\mathcal{X}^n,\tau,h) \wedge 1] = \lim_{h\to 0} \limsup_{n\to\infty} \mathbf{E}[w(\mathcal{X}^n,\tau,h) \wedge 1]$. So (16.9) is equivalent to

$$\lim_{h\to 0} \sup_{n=1,2,...} \mathbf{P}[w(\mathcal{X}^n,\tau,h) > \varepsilon] = 0$$

for all $\varepsilon > 0$ and $\tau > 0$. Let $\tau_k = k$ and $\varepsilon > 0$. Then there exist $h_1, h_2, ... > 0$ such that

$$\sup_{n=1,2,...} \mathbf{P}(w(\mathcal{X}^n, \tau_k, h_k) > 2^{-k}) \le 2^{-(k+1)}\varepsilon.$$

Further, let $t_1, t_2, ...$ be a count of $\mathbb{Q}_+$ and $C_1, C_2, ... \subseteq \mathbb{R}$ compact such that

$$\sup_{n=1,2,...} \mathbf{P}(X^n(t_k) \notin C_k) \leq 2^{-(k+1)}\varepsilon.$$

Now we define

$$B := \bigcap_{k=1}^{\infty} \{f \in \mathcal{C}_E([0,\infty)) : f(t_k) \in C_k, w(f, \tau_k, h_k) \leq 2^{-k}\}.$$

According to Arzela-Ascoli's theorem, $B \subseteq \mathcal{C}_E([0,\infty))$ is relatively compact. Furthermore,

$$\begin{aligned}\sup_{n=1,2,...} \mathbf{P}(\mathcal{X}^n \notin B) &\leq \sup_{n=1,2,...} \sum_{k=1}^{\infty} \mathbf{P}(X^n(t_k) \notin C_k) + \mathbf{P}(w(\mathcal{X}^n, \tau_k, h_k) > 2^{-k}) \\ &\leq \sum_{k=1}^{\infty} 2^{-(k+1)}\varepsilon + 2^{-(k+1)}\varepsilon = \varepsilon.\end{aligned}$$

It follows that $(\mathcal{X}^n)_{n=1,2,...}$ is tight. $\square$

We want to apply the last result to prove the convergence of the random walk against Brownian motion. For this we need one more lemma.

Lemma 16.20. *Let $Y_1, Y_2, ...$ be independent and identically distributed random variables with $\mathbf{E}[Y_1] = 0$ and $\mathbf{V}[Y_1] = \sigma^2 > 0$ and $S_n := Y_1 + \cdots + Y_n$. Then the following applies for $r > 1$*

$$\mathbf{P}(\max_{1 \leq k \leq n} S_k > 2r\sqrt{n}) \leq \frac{\mathbf{P}(|S_n| > r\sqrt{n})}{1 - \sigma^2 r^{-2}}.$$

Proof. We define $T := \inf\{k : |S_k| > 2r\sqrt{n}\}$. Then, since $(S_n)_{n=1,2,...}$ is strongly Markov,

$$\begin{aligned}\mathbf{P}(|S_n| > r\sqrt{n}) &\geq \mathbf{P}(|S_n| > r\sqrt{n}, \max_{1 \leq k \leq n} S_k > 2r\sqrt{n}) \\ &\geq \mathbf{P}(T \leq n, |S_n - S_T| \leq r\sqrt{n}) \\ &\geq \mathbf{P}(\max_{1 \leq k \leq n} S_k > 2r\sqrt{n}) \cdot \min_{1 \leq k \leq n} \mathbf{P}(|S_k| \leq r\sqrt{n}).\end{aligned}$$

From Chebychev's inequality,

$$\min_{1 \leq k \leq n} \mathbf{P}(|S_k| \leq r\sqrt{n}) \geq \min_{1 \leq k \leq n} 1 - \frac{\sigma^2 k}{r^2 n} = 1 - \frac{\sigma^2}{r^2}.$$

$\square$

Theorem 16.21 (Donsker's theorem)**.** *Let $Y_1, Y_2, ...$ be independent, identically distributed random variables with $\mathbf{E}[Y_1] = 0$ and $\mathbf{V}[Y_1] = \sigma^2 > 0$, and $\mathcal{X}_n = (X_{n,t})_{t \geq 0}$ given by*

$$X_{n,t} := \frac{1}{\sqrt{n\sigma^2}}\big(Y_1 + \cdots + Y_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor)Y_{\lfloor nt \rfloor + 1}\big)$$

and $\mathcal{X} = (X_t)_{t \geq 0}$ a Brownian motion. Then,

$$\mathcal{X}_n \xRightarrow{n \to \infty} \mathcal{X}.$$

Proof. Let wlog $\sigma^2 = 1$. As stated in Remark 16.12 it holds that $\mathcal{X}_n \xrightarrow[fdd]{n\to\infty} \mathcal{X}$. Therefore, according to Proposition 16.16, the tightness of the family $\{\mathcal{X}_n : n \in \mathbb{N}\}$, so (16.9) from Theorem 16.19, must be proven. We write $S_n := Y_1 + \cdots + Y_n$ in the following. With Lemma 16.20,

$$\begin{aligned}
&\lim_{h\to 0} \frac{1}{h} \limsup_{n\to\infty} \mathbf{P}\Big(\sup_{0\le s\le h} |X_{n,t+s} - X_{n,t}| > \varepsilon \Big) \\
&\quad \le \lim_{h\to 0} \frac{1}{h} \limsup_{n\to\infty} \mathbf{P}\Big(\sup_{k=1,\dots,\lceil nh\rceil} |S_k| > \frac{\varepsilon}{\sqrt{h}} \sqrt{nh} \Big) \\
&\quad \le \lim_{h\to 0} \frac{1}{h} \limsup_{n\to\infty} \mathbf{P}\Big(\frac{|S_{\lceil nh\rceil}|}{\sqrt{nh}} > \frac{\varepsilon}{2\sqrt{h}} \Big) \\
&\quad \le \lim_{h\to 0} \frac{2}{h} \int_{\varepsilon/(2\sqrt{h})}^{\infty} \varphi(x)dx \\
&\quad = \lim_{h\to 0} \frac{2}{h} \frac{2\sqrt{h}}{\varepsilon} \varphi(\varepsilon/(2\sqrt{h})) = 0
\end{aligned}$$

by (16.5), where φ is the density of the $N(0,1)$ distribution. Now let $\delta > 0$ and h be small enough for

$$\limsup_{n\to\infty} \mathbf{P}\Big(\sup_{0\le s\le h} |X_{n,t+s} - X_{n,t}| > \varepsilon \Big) \le \delta h.$$

With this we can write

$$\begin{aligned}
\limsup_{n\to\infty} \mathbf{P}(w(\mathcal{X}_n, \tau, h) > 2\varepsilon) &= \limsup_{n\to\infty} \mathbf{P}(\sup_{0\le t\le \tau-h, 0\le s\le h} |X_{n,t+s} - X_{n,t}| > 2\varepsilon) \\
&\le \limsup_{n\to\infty} \mathbf{P}(\sup\{|X_{n,kh+s} - X_{n,kh}| : k = 0, 1, \dots, [\tau/h], 0 \le s \le h\} > \varepsilon) \\
&\le \sum_{k=0}^{[\tau/h]} \limsup_{n\to\infty} \mathbf{P}(\sup\{|X_{n,kh+s} - X_{n,kh}| : 0 \le s \le h\} > \varepsilon) \\
&\le [\tau/h]\delta h \xrightarrow{h\to 0} \tau\delta.
\end{aligned}$$

Since $\delta > 0$ was arbitrary, the result follows (16.9). $\square$

We end this section with a tightness criterion that is often is applicable. It builds on theorem 13.8.

Theorem 16.22 (Kolmogorov-Chentsov criterion for tightness)**.** *Let $\mathcal{X}_1 = (X_1(t))_{t\ge 0}, \mathcal{X}_2 = (X_2(t))_{t\ge 0}, \dots$ are stochastic processes with continuous paths. Assuming $\{X_n(0) : n \in \mathbb{N}\}$ is tight and for every each $\tau > 0$ there are numbers $\alpha, \beta, C > 0$ with*

$$\sup_n \mathbf{E}[r(X_n(s), X_n(t))^\alpha] \le C|t-s|^{1+\beta}$$

for all $0 \le s, t \le \tau$. Then $\{\mathcal{X}_n : n \in \mathbb{N}\}$ is tight in $\mathcal{C}_E([\infty))$.

Proof. Let $0 < \gamma < \beta/\alpha$ be arbitrary. We use the notation from the proof of theorem 13.8, e.g. $\xi_{nk} := \max\{r(X_n(s), X_n(t)) : s, t \in D_k, |t-s| = 2^{-k}\}$. Elog let $\tau = 1$. Just as in (13.1) we calculate

$$\sum_{k=0}^{\infty} 2^{\alpha\gamma k} \mathbf{E}[\xi_{nk}^\alpha] \le C \sum_{k=0}^{\infty} 2^{(\alpha\gamma-\beta)k}.$$

Since the right-hand side does not depend on n, there is a C' with $\sup_n \mathbf{E}[\xi_{nk}^\alpha] \le C' e^{-\alpha\gamma k}$. It is important to realize that $w(\mathcal{X}_n, 1, 2^{-m}) \le \sum_{k=m}^\infty \xi_{nk}$. From this,

$$\begin{aligned}\sup_n \mathbf{E}[w(\mathcal{X}_n, 1, 2^{-m})^\alpha \wedge 1] &\le \sup_n \mathbf{E}\Big[\Big(\sum_{k=m}^\infty \xi_{nk}\Big)^\alpha\Big] \le \sup_n \Big(\sum_{k=m}^\infty \mathbf{E}[\xi_{nk}^\alpha]^{1/\alpha}\Big)^\alpha \\ &\le C'\Big(\sum_{k=m}^\infty e^{-\gamma k}\Big)^{1/\alpha} \xrightarrow{m\to\infty} 0,\end{aligned}$$

from which the assertion follows. $\square$

16.5 The Skorohod embedding Theorem

The name Skorohod was already mentioned in the connection between weak and almost sure convergence, see Theorem 9.11. Simply spoken, a sequence of random variables converges weakly iff it converges almost surely in a suitable probability space. If we look again at Donsker's theorem, we can ask ourselves the question as to what the probability space should look like, on which the random walks converge almost surely against a Brownian motion. In other words: how must one define the random walks and the Brownian motion so that both *always* are close together. This is answered by Skorohod's embedding theorem, Theorem 16.26. It allows further conclusions to be drawn about the error, such as the law of the iterated logarithm, Theorem 16.29. The following lemma is fundamental:

Lemma 16.23 (Randomization). *For $w < 0 < z$ let $Y_{w,z}$ be a random variable with state space $\{w, z\}$ with*

$$\mathbf{P}(Y_{w,z} = w) = \frac{z}{z + |w|}$$

and $Y_{w,z} = 0$ for $w, z = 0$. Further, let Y be a real-valued random variable with $\mathbf{E}[Y] = 0$. Then there is a pair of random variables (W, Z) with $W \le 0, Z \ge 0$, so that Y has the distribution $Y_{W,Z}$.

Proof. We set $c = \mathbf{E}[Y^+] = \mathbf{E}[Y^-]$. Further, let $f : \mathbb{R} \to \mathbb{R}_+$ is measurable with $f(0) = 0$. Then, if $Y \sim \mu$,

$$\begin{aligned}c \cdot \mathbf{E}[f(Y)] &= \mathbf{E}[Y^+] \cdot \mathbf{E}[f(-Y^-)] + \mathbf{E}[Y^-] \cdot \mathbf{E}[f(Y^+)] \\ &= \int\int (zf(w) + |w|f(z)) 1_{z\ge 0} 1_{w \le 0} \mu(dw)\mu(dz) \\ &= \int\int (z + |w|)\mathbf{E}[f(Y_{w,z})] 1_{z\ge 0} 1_{w\le 0} \mu(dw)\mu(dz).\end{aligned}$$

This means that we define (W, Z) as a random variable with a joint distribution

$$\mu_{W,Z}(dw, dz) = \mu(0)\delta_{0,0}(dw, dz) + \tfrac{1}{c}(z + |w|) 1_{w\le 0} 1_{z\ge 0} \mu(dw)\mu(dz)$$

can be selected. (It is easy to check that the total mass of this measure is 1). Then,

$$c\mathbf{E}[f(Y_{W,Z})] = c\mathbf{E}[\mathbf{E}[f(W_{W,Z})|(W, Z)]] = \int\int (z + |w|)\mathbf{E}[f(Y_{w,z})] 1_{w\le 0} 1_{z\ge 0} \mu(dw)\mu(dz)$$

and the assertion is shown, since f was arbitrary. $\square$

Remark 16.24 (strong embedding)**.** *The lemma initially only asserts equality in distribution, $Y \sim Y_{W,Z}$. Furthermore, it is also possible to define the probability space on which Y is defined by adding random variables (W, Z) and $Y_{W,Z}$, so that $Y = Y_{W,Z}$ is almost certain.*

Lemma 16.25 (Embedding of a random variable in a Brownian motion)**.** *Let Y be a real-valued random variable with $\mathbf{E}[Y] = 0$. Further, let (W, Z) be distributed as in Lemma 16.23, and $\mathcal{X} = (X_t)_{t\geq 0}$ is an independent Brownian motion. Then*

$$T_{W,Z} = \inf\{t \geq 0 : X_t \in \{W, Z\}\}$$

is a stopping time with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$ with $\mathcal{F}_t = \sigma(W, Z; X_s : s \leq t)$. In addition,

$$X_{T_{W,Z}} \sim Y, \qquad \mathbf{E}[T_{W,Z}] = \mathbf{E}[Y^2].$$

Proof. The Brownian motion $\mathcal{X}$ is adapted to $(\mathcal{F}_t)_{t\geq 0}$ is adapted. Therefore, $T_{W,Z}$ according to Proposition 13.30 is a stopping time. Clearly, for $w < 0 \leq z$ the random variable $X_{T_{w,z}}$ only takes the values w and z. According to Proposition 14.19, $(X_{T_{w,z}\wedge t})_{t\geq 0}$ is is a martingale which, according to Theorem 14.22 converges in L^1 against $X_{T_{w,z}}$. Therefore

$$0 = \mathbf{E}[X_{T_{w,z}}] = w\mathbf{P}(X_{T_{w,z}} = w) + z(1 - \mathbf{P}(X_{T_{w,z}} = w)),$$

also

$$\mathbf{P}(X_{T_{w,z}} = w) = \frac{z}{z + |w|}.$$

So $X_{T_{w,z}}$ has the same distribution as $Y_{w,z}$ from Lemma 16.23 and is independent of X. According to the lemma it follows that $X_{T_{W,Z}} \sim Y_{W,Z} \sim Y$. Further, $(X_t^2 - t)_{t\geq 0}$ is a martingale and for $y < 0 \leq z$, the process $(X^2_{T_{w,z}\wedge t} - T_{w,z} \wedge t)_{t\geq 0}$ is a martingale. This means that with monotone and dominated convergence,

$$\begin{aligned}\mathbf{E}[T_{W,Z}] &= \mathbf{E}[\mathbf{E}[T_{W,Z}|W, Z]] = \mathbf{E}[\lim_{t\to\infty} \mathbf{E}[T_{W,Z} \wedge t]|W, Z] \\ &= \mathbf{E}[\mathbf{E}[X^2_{T_{W,Z}}|W, Z]] = \mathbf{E}[X^2_{T_{W,Z}}] = \mathbf{E}[Y^2].\end{aligned}$$

□

Theorem 16.26 (Skorohod's embedding theorem)**.** *Let $Y_1, Y_2, ...$ be independent and identically distributed with $\mathbf{E}[Y_1] = 0$, and $S_n = Y_1 + \cdots + Y_n$. Then there is a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with filtration $(\mathcal{F}_t)_{t\geq 0}$, as well as a Brownian motion $\mathcal{X} = (X_t)_{t\geq 0}$ on this probability space, which is a $(\mathcal{F}_t)_{t\geq 0}$ martingale and stopping times $T_1, T_2, ...$, so that:*

1. *$(X_{T_1}, X_{T_2}, ...) \sim S_1, S_2, ...$ and*
2. *$(T_{n+1} - T_n)_{n=0,1,2,...}$ are independent with $\mathbf{E}[T_{n+1} - T_n] = \mathbf{V}[Y_1]$ for $n = 1, 2, ...$*

Remark 16.27 (Strong embedding)**.**

1. *As in remark 16.24, it is possible to define the probability space on which $Y_1, Y_2, ...$ are defined so that $(X_{T_1}, X_{T_2}, ...) = S_1, S_2, ...$ almost certainly holds.*
2. *Without the restriction of the integrability of $T_{n+1} - T_n$ the statement of the theorem would be trivial. Then you could simply recursively $0 = T_0 \leq T_1, ...$ by means of*

$$T_n = \inf\{t \geq T_{n-1} : X_t = S_n\}.$$

However, these waiting times cannot be integrated.

Proof of theorem 16.26. Let the pairs $(W_1, Z_1), (W_2, Z_2), ...$ be distributed exactly as in Lemma 16.23. We extend the probability space by an independent Brownian motion $\mathcal{X} = (X_t)_{t\geq 0}$. We recursively define $0 = T_0 \leq T_1 \leq T_2...$ by

$$T_n := \inf\{t \geq T_{n-1} : X_t - X_{T_{n-1}} \in \{W_n, Z_n\}\}.$$

Thus $T_1, T_2, ...$ are stopping times with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$ with $\mathcal{F}_t = \sigma(W_1, Z_1, W_2, Z_2, ...; X_s : s \leq t)$ and $\mathcal{X}$ is a martingale with respect to $(\mathcal{F}_t)_{t\geq 0}$. Furthermore, the pairs $(T_{n+1} - T_n, X_{T_{n+1}} - X_{T_n})_{n=0,1,2,...}$ because of the strong Markov property of the Brownian motion are independent of each other. Therefore, it follows from Lemma 16.25 that

$$(X_{T_1}, X_{T_2} - X_{T_1}, ...) \sim (Y_1, Y_2, ...),$$

also

$$(X_{T_1}, X_{T_2}, ...) \sim (S_1, S_2, ...),$$

and $\mathbf{E}[T_{n+1} - T_n] = \mathbf{E}[Y_n]$. □

Since, thanks to the last theorem, the relationship between the random walks and Brownian motion is shown, it is obvious to formulate another extension of Donsker's theorem, Theorem 16.21.

Corollary 16.28 (Stochastic convergence of the random walks)**.** *Let $Y_1, Y_2, ...$ be real-valued, independent, identically distributed random variables with $\mathbf{E}[Y_1] = 0$, $\mathbf{V}[Y_1] = 1$ and $S_n = Y_1 + \cdots + Y_n$. Then you can expand the probability space so that there is a Brownian motion $\mathcal{X} = (X_t)_{t\geq 0}$ with*

$$\sup_{0\leq s\leq t} \Big| \frac{1}{\sqrt{n}} S_{[sn]} - \frac{1}{\sqrt{n}} X_{sn} \Big| \xrightarrow{n\to\infty}_p 0 \tag{16.10}$$

for all $t > 0$.

Proof. We use the construction from Theorem 16.26 and Remark 16.27. Since $T_{n+1} - T_n$ are independent and identically distributed with $\mathbf{E}[T_{n+1} - T_n] = 1$ and $T_n/n \xrightarrow{n\to\infty} 1$ according to the law of large numbers. This means that $\frac{1}{n}\sup_{s\leq t} |T_{[sn]} - sn| \xrightarrow{n\to\infty}_{fs} 0$. (To see this, we consider the set $\{\frac{1}{n}\sup_{s\leq t} |T_{[sn]} - sn| > \varepsilon\}$ for a $\varepsilon > 0$. On this set there are $s_1, s_2, ... \leq t$ with $|T_{[s_n n]} - s_n n| > \varepsilon n$. However, this contradicts $\lim_{n\to\infty} T_{[s_n n]}/[s_n n] = \lim_{n\to\infty} T_n/n = 1$.)

We recall the definition of the continuity modulus w from Definition 16.17. With the scaling property of the Brownian motion from Theorem 13.19, it follows that $S_{[sn]} = X_{T_{[sn]}}$,

$$\begin{aligned}
\limsup_{n\to\infty} \mathbf{P}\Big(\frac{1}{\sqrt{n}} \sup_{0\leq s\leq t} |S_{[sn]} - X_{sn}| > \varepsilon\Big) \\
&\leq \inf_h \limsup_{n\to\infty} \mathbf{P}(w(\mathcal{X}, (t+h)n, nh) > \varepsilon\sqrt{n}) + \mathbf{P}(\sup_{s\leq t} |T_{[sn]} - sn| > nh) \\
&= \inf_h \mathbf{P}(w(\mathcal{X}, t+h, h) > \varepsilon) = 0.
\end{aligned}$$

□

Now that the random walks and Brownian motion are directly related to each other, it makes sense to transfer the properties of Brownian motion to the random walks. We do this for the Law of the iterated logarithm.

Theorem 16.29 (Law of the iterated logarithm for random walks)**.** *Let $Y_1, Y_2, ...$ be real-valued, independent, identically distributed random variable with $\mathbf{E}[Y_1] = 0$, $\mathbf{V}[Y_1] = 1$ and $S_n = Y_1 + \cdots + Y_n$. Then,*

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n\log\log n}} = 1$$

almost certainly.

Proof. We only show that the probability space can be extended in such a way such that there is a Brownian motion $\mathcal{X} = (X_t)_{t\geq 0}$ exists with

$$\frac{S_{[t]} - X_t}{\sqrt{2t\log\log t}} \xrightarrow{t\to\infty}_{fs} 0. \tag{16.11}$$

Then the statement follows from the law of the iterated logarithm for Brownian motion, theorem 16.10.

According to Theorem 16.26 there is an extension of the probability space and stopping times $0 = T_0, T_1, ...,$ so that $X_{T_n} = S_n$. Again, $T_n/n \xrightarrow{n\to\infty} 1$ applies according to the law of large numbers, which is also $T_{[t]}/t \xrightarrow{t\to\infty} 1$ implies. Now let $r > 1$, $c^2 > r - 1$ and $h(t) = \sqrt{2t\log\log t}$. Then, with a similar calculation as in (16.6)

$$\begin{aligned}\mathbf{P}\Big(\sup_{r^{n-1}\leq t\leq r^n} |X_t - X_{r^{n-1}}| > ch(r^{n-1})\Big) &= \mathbf{P}\Big(\sup_{0\leq t\leq r^n - r^{n-1}} |X_t| > ch(r^{n-1})\Big)\\ &= 2\mathbf{P}(X_{r^n - r^{n-1}} > ch(r^{n-1})) = 2\mathbf{P}(X_1 > ch(r^{n-1})/\sqrt{r^n - r^{n-1}})\\ &\overset{n\to\infty}{\approx} \frac{1}{c}\sqrt{\frac{(r-1)}{\pi\log n}} n^{-c^2/(r-1)},\end{aligned}$$

since $h(r^{n-1})/\sqrt{r^n - r^{n-1}} \overset{n\to\infty}{\approx} \sqrt{(2\log n)/(r-1)}$. The right-hand side is summable, so with the Borel-Cantelli lemma and $X_{T_{[t]}} = S_{[t]}$

$$\begin{aligned}\mathbf{P}\Big(\limsup_{t\to\infty} \frac{|S_{[t]} - X_t|}{h(t)} = 0\Big) &\geq \mathbf{P}\Big(\lim_{r\downarrow 1}\limsup_{t\to\infty}\sup_{t\leq u\leq rt} \frac{|X_u - X_t|}{h(t)} = 0\Big)\\ &\geq \mathbf{P}\Big(\lim_{r\downarrow 1}\limsup_{n\to\infty}\sup_{r^{n-1}\leq t\leq r^n} \frac{|X_t - X_{r^{n-1}}|}{h(r^{n-1})} = 0\Big)\\ &= \inf_{c>0}\mathbf{P}\Big(\lim_{r\downarrow 1, r<c^2+1}\limsup_{n\to\infty}\sup_{r^{n-1}\leq t\leq r^n} \frac{|X_t - X_{r^{n-1}}|}{h(r^{n-1})} \leq c\Big) = 1.\end{aligned}$$

Therefore follows (16.11). □