An Introduction to Stochastic Processes in Continuous Time

Harry van Zanten

November 8, 2004 (this version)

always under construction

Preface

Contents

1

Stochastic processes

1.1 Stochastic processes

Loosely speaking, a stochastic process is a phenomenon that can be thought of as evolving in time in a random manner. Common examples are the location of a particle in a physical system, the price of a stock in a financial market, interest rates, etc.

A basic example is the erratic movement of pollen grains suspended in water, the so-called Brownian motion. This motion was named after the English botanist R. Brown, who first observed it in 1827. The movement of the pollen grain is thought to be due to the impacts of the water molecules that surround it. These hits occur a large number of times in each small time interval, they are independent of each other, and the impact of one single hit is very small compared to the total effect. This suggest that the motion of the grain can be viewed as a random process with the following properties:

(i) The displacement in any time interval $[s,t]$ is independent of what happened before time s.

(ii) Such a displacement has a Gaussian distribution, which only depends on the length of the time interval $[s,t]$.

(iii) The motion is continuous.

The mathematical model of the Brownian motion will be the main object of investigation in this course. Figure 1.1 shows a particular realization of this stochastic process. The picture suggest that the BM has some remarkable properties, and we will see that this is indeed the case.

Mathematically, a stochastic process is simply an indexed collection of random variables. The formal definition is as follows.

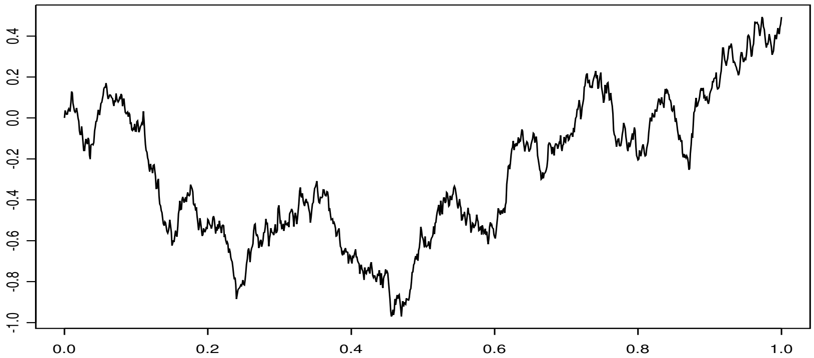

Figure 1.1: A realization of the Brownian motion

Definition 1.1.1. Let T be a set and $(E, \mathcal{E})$ a measurable space. A *stochastic process* indexed by T, taking values in $(E, \mathcal{E})$, is a collection $X = (X_t)_{t \in T}$ of measurable maps X_t from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $(E, \mathcal{E})$. The space $(E, \mathcal{E})$ is called the *state space* of the process.

We think of the index t as a time parameter, and view the index set T as the set of all possible time points. In these notes we will usually have $T = \mathbb{Z}_+ = \{0, 1, \ldots\}$ or $T = \mathbb{R}_+ = [0, \infty)$. In the former case we say that time is *discrete*, in the latter we say time is *continuous*. Note that a discrete-time process can always be viewed as a continuous-time process which is constant on the intervals $[n-1, n)$ for all $n \in \mathbb{N}$. The state space $(E, \mathcal{E})$ will most often be a Euclidean space $\mathbb{R}^d$, endowed with its Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$. If E is the state space of a process, we call the process *E-valued*.

For every fixed $t \in T$ the stochastic process X gives us an E-valued random element X_t on $(\Omega, \mathcal{F}, \mathbb{P})$. We can also fix $\omega \in \Omega$ and consider the map $t \mapsto X_t(\omega)$ on T. These maps are called the *trajectories*, or *sample paths* of the process. The sample paths are functions from T to E, i.e. elements of E^T. Hence, we can view the process X as a random element of the function space E^T. (Quite often, the sample paths are in fact elements of some nice subset of this space.)

The mathematical model of the physical Brownian motion is a stochastic process that is defined as follows.

Definition 1.1.2. The stochastic process $W = (W_t)_{t \geq 0}$ is called a (standard) *Brownian motion*, or *Wiener process*, if

(i) $W_0 = 0$ a.s.,

(ii) $W_t - W_s$ is independent of $(W_u : u \leq s)$ for all $s \leq t$,

(iii) $W_t - W_s$ has a $N(0, t-s)$-distribution for all $s \leq t$,

(iv) almost all sample paths of W are continuous.

We abbreviate 'Brownian motion' to BM in these notes. Property (i) says that a standard BM *starts in* 0. A process with property (ii) is called a process with *independent increments*. Property (iii) implies that that the distribution of the increment $W_t - W_s$ only depends on $t-s$. This is called the *stationarity of the increments*. A stochastic process which has property (iv) is called a *continuous process*. Similarly, we call a stochastic process *right-continuous* if almost all of its sample paths are right-continuous functions. We will often use the acronym *cadlag* (continu à droite, limites à gauche) for processes with sample paths that are right-continuous have finite left-hand limits at every time point.

It is not clear from the definition that the BM actually exists! We will have to prove that there exists a stochastic process which has all the properties required in Definition 1.1.2.

1.2 Finite-dimensional distributions

In this section we recall Kolmogorov's theorem on the existence of stochastic processes with prescribed finite-dimensional distributions. We use it to prove the existence of a process which has properties (i), (ii) and (iii) of Definition 1.1.2.

Definition 1.2.1. Let $X = (X_t)_{t\in T}$ be a stochastic process. The distributions of the finite-dimensional vectors of the form $(X_{t_1}, \ldots, X_{t_n})$ are called the *finite-dimensional distributions* (fdd's) of the process.

It is easily verified that the fdd's of a stochastic process form a consistent system of measures, in the sense of the following definition.

Definition 1.2.2. Let T be a set and $(E, \mathcal{E})$ a measurable space. For all $t_1, \ldots, t_n \in T$, let $\mu_{t_1,\ldots,t_n}$ be a probability measure on $(E^n, \mathcal{E}^n)$. This collection of measures is called *consistent* if it has the following properties:

(i) For all $t_1, \ldots, t_n \in T$, every permutation π of $\{1, \ldots, n\}$ and all $A_1, \ldots, A_n \in \mathcal{E}$

$$\mu_{t_1,\ldots,t_n}(A_1 \times \cdots \times A_n) = \mu_{t_{\pi(1)},\ldots,t_{\pi(n)}}(A_{\pi(1)} \times \cdots \times A_{\pi(n)}).$$

(ii) For all $t_1, \ldots, t_{n+1} \in T$ and $A_1, \ldots, A_n \in \mathcal{E}$

$$\mu_{t_1,\ldots,t_{n+1}}(A_1 \times \cdots \times A_n \times E) = \mu_{t_1,\ldots,t_n}(A_1 \times \cdots \times A_n).$$

The Kolmogorov consistency theorem states that conversely, under mild regularity conditions, every consistent family of measures is in fact the family of fdd's of some stochastic process. Some assumptions are needed on the state space $(E, \mathcal{E})$. We will assume that E is a Polish space (a complete, separable metric space) and $\mathcal{E}$ is its Borel σ-algebra, i.e. the σ-algebra generated by the open sets. Clearly, the Euclidean spaces $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ fit into this framework.

Theorem 1.2.3 (Kolmogorov's consistency theorem). *Suppose that E is a Polish space and $\mathcal{E}$ is its Borel σ-algebra. Let T be a set and for all $t_1, \dots, t_n \in T$, let $\mu_{t_1,\dots,t_n}$ be a measure on $(E^n, \mathcal{E}^n)$. If the measures $\mu_{t_1,\dots,t_n}$ form a consistent system, then on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ there exists a stochastic process $X = (X_t)_{t\in T}$ which has the measures $\mu_{t_1,\dots,t_n}$ as its fdd's.*

Proof. See for instance Billingsley (1995). □

The following lemma is the first step in the proof of the existence of the BM.

Corollary 1.2.4. *There exists a stochastic process $W = (W_t)_{t\geq 0}$ with properties (i), (ii) and (iii) of Definition 1.1.2.*

Proof. Let us first note that a process W has properties (i), (ii) and (iii) of Definition 1.1.2 if and only if for all $t_1, \dots, t_n \geq 0$ the vector $(W_{t_1}, \dots, W_{t_n})$ has an n-dimensional Gaussian distribution with mean vector 0 and covariance matrix $(t_i \wedge t_j)_{i,j=1..n}$ (see Exercise 1). So we have to prove that there exist a stochastic process which has the latter distributions as its fdd's. In particular, we have to show that the matrix $(t_i \wedge t_j)_{i,j=1..n}$ is a valid covariance matrix, i.e. that it is nonnegative definite. This is indeed the case since for all $a_1, \dots, a_n$ it holds that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j (t_i \wedge t_j) = \int_0^\infty \left(\sum_{i=1}^{n} a_i 1_{[0,t_i]}(x)\right)^2 dx \geq 0.$$

This implies that for all $t_1, \dots, t_n \geq 0$ there exists a random vector $(X_{t_1}, \dots, X_{t_n})$ which has the n-dimensional Gaussian distribution $\mu_{t_1,\dots,t_n}$ with mean 0 and covariance matrix $(t_i \wedge t_j)_{i,j=1..n}$. It easily follows that the measures $\mu_{t_1,\dots,t_n}$ form a consistent system. Hence, by Kolmogorov's consistency theorem, there exists a process W which has the distributions $\mu_{t_1,\dots,t_n}$ as its fdd's. □

To prove the existence of the BM, it remains to consider the continuity property (iv) in the definition of the BM. This is the subject of the next section.

1.3 Kolmogorov's continuity criterion

According to Corollary 1.3.4 there exists a process W which has properties (i)–(iii) of Definition 1.1.2. We would like this process to have the continuity property (iv) of the definition as well. However, we run into the problem that there is no particular reason why the set

$$\{\omega : t \mapsto W_t(\omega) \text{ is continuous}\} \subseteq \Omega$$

should be measurable. Hence, the probability that the process W has continuous sample paths is not well defined in general.

One way around this problem is to ask whether we can modify the given process W in such a way that the resulting process, say $\tilde{W}$, has continuous sample paths and still satisfies (i)–(iii), i.e. has the same fdd's as W. To make this idea precise, we need the following notions.

Definition 1.3.1. Let X and Y be two processes indexed by the same set T and with the same state space $(E, \mathcal{E})$, defined on probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\Omega', \mathcal{F}', \mathbb{P}')$ respectively. The processes are called *versions* of each other if they have the same fdd's, i.e. if for all $t_1, \dots, t_n \in T$ and $B_1, \dots, B_n \in \mathcal{E}$

$$\mathbb{P}(X_{t_1} \in B_1, \dots, X_{t_n} \in B_n) = \mathbb{P}'(Y_{t_1} \in B_1, \dots, Y_{t_n} \in B_n).$$

Definition 1.3.2. Let X and Y be two processes indexed by the same set T and with the same state space $(E, \mathcal{E})$, defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The processes are called *modifications* of each other if for every $t \in T$

$$X_t = Y_t \quad \text{a.s.}$$

The second notion is clearly stronger than the first one: if processes are modifications of each other, then they are certainly versions of each other. The converse is not true in general (see Exercise 2).

The following theorem gives a sufficient condition under which a given process has a continuous modification. The condition (1.1) is known as *Kolmogorov's continuity condition*.

Theorem 1.3.3 (Kolmogorov's continuity criterion). *Let $X = (X_t)_{t\in[0,T]}$ be an $\mathbb{R}^d$-valued process. Suppose that there exist constants $\alpha, \beta, K > 0$ such that*

$$\mathbb{E}\|X_t - X_s\|^\alpha \leq K|t-s|^{1+\beta} \tag{1.1}$$

for all $s, t \in [0, T]$. Then there exists a continuous modification of X.

Proof. For simplicity, we assume that $T = 1$ in the proof. First observe that by Chebychev's inequality, condition (1.1) implies that the process X is *continuous in probability*. This means that if $t_n \to t$, then $X_{t_n} \to X_t$ in probability. Now for $n \in \mathbb{N}$, define $D_n = \{k/2^n : k = 0, 1, \dots, 2^n\}$ and let $D = \bigcup_{n=1}^\infty D_n$. Then D is a countable set, and D is dense in $[0, 1]$. Our next aim is to show that with probability 1, the process X is uniformly continuous on D.

Fix an arbitrary $\gamma \in (0, \beta/\alpha)$. Using Chebychev's inequality again, we see that

$$\mathbb{P}(\|X_{k/2^n} - X_{(k-1)/2^n}\| \geq 2^{-\gamma n}) \lesssim 2^{-n(1+\beta-\alpha\gamma)}.$$ [1]

[1] The notation '$\lesssim$' means that the left-hand side is less than a positive constant times the right-hand side.

It follows that

$$\begin{aligned}\mathbb{P}\Big(\max_{1\le k\le 2^n}\|X_{k/2^n}-X_{(k-1)/2^n}\|\ge 2^{-\gamma n}\Big)\\ \le \sum_{k=1}^{2^n}\mathbb{P}(\|X_{k/2^n}-X_{(k-1)/2^n}\|\ge 2^{-\gamma n})\lesssim 2^{-n(\beta-\alpha\gamma)}.\end{aligned}$$

Hence, by the Borel-Cantelli lemma, there almost surely exists an $N\in\mathbb{N}$ such that

$$\max_{1\le k\le 2^n}\|X_{k/2^n}-X_{(k-1)/2^n}\|<2^{-\gamma n} \tag{1.2}$$

for all $n\ge N$.

Next, consider an arbitrary pair $s,t\in D$ such that $0<t-s<2^{-N}$. Our aim in this paragraph is to show that

$$\|X_t-X_s\|\lesssim |t-s|^{\gamma}. \tag{1.3}$$

Observe that there exists an $n\ge N$ such that $2^{-(n+1)}\le t-s<2^{-n}$. We claim that if $s,t\in D_m$ for $m\ge n+1$, then

$$\|X_t-X_s\|\le 2\sum_{k=n+1}^{m}2^{-\gamma k}. \tag{1.4}$$

The proof of this claim proceeds by induction. Suppose first that $s,t\in D_{n+1}$. Then necessarily, $t=k/2^{n+1}$ and $s=(k-1)/2^{n+1}$ for some $k\in\{1,\dots,2^{n+1}\}$. By (1.2), it follows that

$$\|X_t-X_s\|<2^{-\gamma(n+1)},$$

which proves the claim for $m=n+1$. Now suppose that it is true for $m=n+1,\dots,l$ and assume that $s,t\in D_{l+1}$. Define the numbers $s',t'\in D_l$ by

$$s'=\min\{u\in D_l: u\ge s\},\quad t'=\max\{u\in D_l: u\le t\}.$$

Then by construction, $s\le s'\le t'\le t$ and $s'-s\le 2^{-(l+1)}$, $t'-t\le 2^{-(l+1)}$. Hence, by the triangle inequality, (1.2) and the induction hypothesis,

$$\begin{aligned}\|X_t-X_s\|&\le\|X_{s'}-X_s\|+\|X_{t'}-X_t\|+\|X_{t'}-X_{s'}\|\\ &\le 2^{-\gamma(l+1)}+2^{-\gamma(l+1)}+2\sum_{k=n+1}^{l}2^{-\gamma k}=2\sum_{k=n+1}^{l+1}2^{-\gamma k},\end{aligned}$$

so the claim is true for $m=l+1$ as well. The proof of (1.3) is now straightforward. Indeed, since $t,s\in D_m$ for some large enough m, relation (1.4) implies that

$$\|X_t-X_s\|\le 2\sum_{k=n+1}^{\infty}2^{-\gamma k}=\frac{2}{1-2^{-\gamma}}\,2^{-\gamma(n+1)}\le\frac{2}{1-2^{-\gamma}}\,|t-s|^{\gamma}.$$

Observe that (1.3) implies in particular that almost surely, the process X is uniformly continuous on D. In other words, we have an event $\Omega^*\subseteq\Omega$ with $\mathbb{P}(\Omega^*)=1$ such that for all $\omega\in\Omega^*$, the sample path $t\mapsto X_t(\omega)$ is uniformly continuous on the countable, dense set D. Now we define a new stochastic process $Y = (Y_t)_{t\in[0,1]}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ as follows: for $\omega \notin \Omega^*$, we put $Y_t = 0$ for all $t \in [0,1]$, for $\omega \in \Omega^*$ we define

$$Y_t = \begin{cases} X_t & \text{if } t \in D, \\ \lim\limits_{\substack{t_n \to t \\ t_n \in D}} X_{t_n}(\omega) & \text{if } t \notin D. \end{cases}$$

The uniform continuity of X implies that Y is a well-defined, continuous stochastic process. Since X is continuous in probability (see the first part of the proof), Y is a modification of X (see Exercise 3). □

Corollary 1.3.4. *Brownian motion exists.*

Proof. By Corollary 1.3.4, there exists a process $W = (W_t)_{t\geq 0}$ which has properties (i)–(iii) of Definition 1.1.2. By property (iii) the increment $W_t - W_s$ has a $N(0, t-s)$-distribution for all $s \leq t$. It follows that $\mathbb{E}(W_t - W_s)^4 = (t-s)^2\mathbb{E}Z^4$, where Z is a standard Gaussian random variable. This means that Kolmogorov's continuity criterion (1.1) is satisfied with $\alpha = 4$ and $\beta = 1$. So for every $T \geq 0$, there exist a continuous modification $W^T = (W^T_t)_{t\in[0,T]}$ of the process $(W_t)_{t\in[0,T]}$. Now define the process $X = (X_t)_{t\geq 0}$ by

$$X_t = \sum_{n=1}^{\infty} W^n_t 1_{[n-1,n)}(t).$$

The process X is a Brownian motion (see Exercise 5). □

1.4 Gaussian processes

Brownian motion is an example of a Gaussian process. The general definition is as follows.

Definition 1.4.1. A real-valued stochastic process is called *Gaussian* if all its fdd's are Gaussian.

If X is a Gaussian process indexed by the set T, the *mean function* of the process is the function m on T defined by $m(t) = \mathbb{E}X_t$. The *covariance function* of the process is the function r on $T \times T$ defined by $r(s,t) = \mathbb{C}\mathrm{ov}(X_s, X_t)$. The functions m and r determine the fdd's of the process X.

Lemma 1.4.2. *Two Gaussian processes with the same mean function and covariance function are versions of each other.*

Proof. See Exercise 6. □

The mean function m and covariance function r of the BM are given by $m(t) = 0$ and $r(s,t) = s \wedge t$ (see Exercise 1). Conversely, the preceding lemma implies that every Gaussian process with the same mean and covariance function has the same fdd's as the BM. It follows that such a process has properties (i)–(iii) of Definition 1.1.2. Hence, we have the following result.

Lemma 1.4.3. *A continuous Gaussian process $X = (X_t)_{t\geq 0}$ is a BM if and only if it has mean function $\mathbb{E}X_t = 0$ and covariance function $\mathbb{E}X_s X_t = s \wedge t$.*

Using this lemma we can prove the following symmetries and scaling properties of BM.

Theorem 1.4.4. *Let W be a BM. Then the following are BM's as well:*

(i) (time-homogeneity) *for every $s \geq 0$, the process $(W_{t+s} - W_s)_{t\geq 0}$,*

(ii) (symmetry) *the process $-W$,*

(iii) (scaling) *for every $a > 0$, the process W^a defined by $W^a_t = a^{-1/2} W_{at}$,*

(iv) (time inversion) *the process X_t defined by $X_0 = 0$ and $X_t = tW_{1/t}$ for $t > 0$.*

Proof. Parts (i), (ii) and (iii) follow easily from he preceding lemma, see Exercise 7. To prove part (iv), note first that the process X has the same mean function and covariance function as the BM. Hence, by the preceding lemma, it only remains to prove that X is continuous. Since $(X_t)_{t>0}$ is continuous, it suffices to show that if $t_n \downarrow 0$, then

$$\mathbb{P}(X_{t_n} \to 0 \text{ as } n \to \infty) = 1. \tag{1.5}$$

But this probability is determined by the fdd's of the process $(X_t)_{t>0}$ (see Exercise 8). Since these are the same as the fdd's of $(W_t)_{t>0}$, we have

$$\mathbb{P}(X_{t_n} \to 0 \text{ as } n \to \infty) = \mathbb{P}(W_{t_n} \to 0 \text{ as } n \to \infty) = 1.$$

This completes the proof. □

Using the scaling and the symmetry properties we can prove that that the sample paths of the BM oscillate between $+\infty$ and $-\infty$.

Corollary 1.4.5. *Let W be a BM. Then*

$$\mathbb{P}(\sup_{t\geq 0} W_t = \infty) = \mathbb{P}(\inf_{t\geq 0} W_t = -\infty) = 1.$$

Proof. By the scaling property we have for all $a > 0$

$$\sup_t W_t =_d \sup_t \frac{1}{\sqrt{a}} W_{at} = \frac{1}{\sqrt{a}} \sup_t W_t.$$

It follows that for $n \in \mathbb{N}$,

$$\mathbb{P}(\sup_t W_t \leq n) = \mathbb{P}(n^2 \sup_t W_t \leq n) = \mathbb{P}(\sup_t W_t \leq 1/n).$$

By letting n tend to infinity we see that

$$\mathbb{P}(\sup_t W_t < \infty) = \mathbb{P}(\sup_t W_t = 0),$$

so the distribution of $\sup_t W_t$ is concentrated on $\{0, \infty\}$. Hence, to prove that $\sup_t W_t = \infty$ a.s., it suffices to show that $\mathbb{P}(\sup_t W_t = 0) = 0$. Now we have

$$\begin{aligned}
\mathbb{P}(\sup_t W_t = 0) &\leq \mathbb{P}(W_1 \leq 0\ ,\ \sup_{t \geq 1} W_t \leq 0) \\
&\leq \mathbb{P}(W_1 \leq 0\ ,\ \sup_{t \geq 1} W_t - W_1 < \infty) \\
&= \mathbb{P}(W_1 \leq 0) \mathbb{P}(\sup_{t \geq 1} W_t - W_1 < \infty),
\end{aligned}$$

by the independence of the Brownian increments. By the time-homogeneity of the BM, the last probability is the probability that the supremum of a BM is finite. We just showed that this must be equal to $\mathbb{P}(\sup_t W_t = 0)$. Hence, we find that

$$\mathbb{P}(\sup_t W_t = 0) \leq \tfrac{1}{2} \mathbb{P}(\sup_t W_t = 0),$$

which shows that $\mathbb{P}(\sup_t W_t = 0) = 0$, and we obtain the first statement of the corollary. By the symmetry property,

$$\sup_{t \geq 0} W_t =_d \sup_{t \geq 0} -W_t = -\inf_{t \geq 0} W_t.$$

Together with the first statement, this proves the second one. □

Since the BM is continuous, the preceding result implies that almost every sample path visits every point of $\mathbb{R}$ infinitely often. A real-valued process with this property is called *recurrent.*

Corollary 1.4.6. *The BM is recurrent.*

An interesting consequence of the time inversion is the following strong law of large numbers for the BM.

Corollary 1.4.7. *Let W be a BM. Then*

$$\frac{W_t}{t} \xrightarrow{\text{as}} 0$$

as $t \to \infty$.

Proof. Let X be as in part (iv) of Theorem 1.4.4. Then

$$\mathbb{P}(W_t/t \to 0 \text{ as } t \to \infty) = \mathbb{P}(X_{1/t} \to 0 \text{ as } t \to \infty) = 1,$$

since X is continuous and $X_0 = 0$. □

1.5 Non-differentiability of the Brownian sample paths

We have already seen that the sample paths of the BM are continuous functions that oscillate between $+\infty$ and $-\infty$. Figure 1.1 suggests that the sample paths are 'very rough'. The following theorem states that this is indeed the case.

Theorem 1.5.1. *Let W be a Brownian motion. For all ω outside a set of probability zero, the sample path $t \mapsto W_t(\omega)$ is nowhere differentiable.*

Proof. Let W be a BM. Consider the upper and lower right-hand derivatives

$$D^W(t,\omega) = \limsup_{h \downarrow 0} \frac{W_{t+h}(\omega) - W_t(\omega)}{h}$$

and

$$D_W(t,\omega) = \liminf_{h \downarrow 0} \frac{W_{t+h}(\omega) - W_t(\omega)}{h}.$$

Consider the set

$$A = \{\omega : \text{there exists a } t \geq 0 \text{ such that } D^W(t,\omega) \text{ and } D_W(t,\omega) \text{ are finite}\}.$$

Note that A is not necessarily measurable! We will prove that A is contained in a measurable set B with $\mathbb{P}(B) = 0$, i.e. that A has *outer probability* 0.

To define the event B, consider first for $k, n \in \mathbb{N}$ the random variable

$$X_{n,k} = \max\left\{\left|W_{\frac{k+1}{2^n}} - W_{\frac{k}{2^n}}\right|, \left|W_{\frac{k+2}{2^n}} - W_{\frac{k+1}{2^n}}\right|, \left|W_{\frac{k+3}{2^n}} - W_{\frac{k+2}{2^n}}\right|\right\}$$

and for $n \in \mathbb{N}$, define

$$Y_n = \min_{k \leq n2^n} X_{n,k}.$$

The event B is then defined by

$$B = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \left\{Y_k \leq \frac{k}{2^k}\right\}.$$

We claim that $A \subseteq B$ and $\mathbb{P}(B) = 0$.

To prove the inclusion, take $\omega \in A$. Then there exist $K, \delta > 0$ such that

$$|W_s(\omega) - W_t(\omega)| \leq K|s-t| \quad \text{for all } s \in [t, t+\delta]. \tag{1.6}$$

Now take $n \in \mathbb{N}$ so large that

$$\frac{4}{2^n} < \delta, \quad 8K < n, \quad t < n. \tag{1.7}$$

Given this n, determine $k \in \mathbb{N}$ such that

$$\frac{k-1}{2^n} \leq t < \frac{k}{2^n}. \tag{1.8}$$

Then by the first relation in (1.7) we have $k/2^n, \dots, (k+3)/2^n \in [t, t+\delta]$. By (1.6) and the second relation in (1.7) it follows that $X_{n,k}(\omega) \leq n/2^n$. By (1.8) and the third relation in (1.7) it holds that $k - 1 \leq t2^n < n2^n$. Hence, we have $k \leq n2^n$ and therefore $Y_n(\omega) \leq X_{n,k}(\omega) \leq n/2^n$. We have shown that if $\omega \in A$, then $Y_n(\omega) \leq n/2^n$ for all n large enough. This precisely means that $A \subseteq B$.

To complete the proof, we have to show that $\mathbb{P}(B) = 0$. For $\varepsilon > 0$, the basic properties of the BM imply that

$$\mathbb{P}(X_{n,k} \leq \varepsilon) = \mathbb{P}\left(|W_1| \leq 2^{n/2}\varepsilon\right)^3 \lesssim \left(\int_0^{2^{n/2}\varepsilon} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,dx\right)^3 \leq 2^{3n/2}\varepsilon^3.$$

It follows that

$$\mathbb{P}(Y_n \leq \varepsilon) \leq n2^n \mathbb{P}(X_{n,k} \leq \varepsilon) \lesssim n2^{5n/2}\varepsilon^3.$$

In particular we see that $\mathbb{P}(Y_n \leq n/2^n) \to 0$, which implies that $\mathbb{P}(B) = 0$. □

1.6 Filtrations and stopping times

If W is a BM, the increment $W_{t+h} - W_t$ is independent of 'what happened up to time t'. In this section we introduce the concept of a *filtration* to formalize this notion of 'information that we have up to time t'. The probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is fixed again and we suppose that T is a subinterval of $\mathbb{Z}_+$ or $\mathbb{R}_+$.

Definition 1.6.1. A collection $(\mathcal{F}_t)_{t\in T}$ of sub-σ-algebras of $\mathcal{F}$ is called a *filtration* if $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s \leq t$. A stochastic process X defined on $(\Omega, \mathcal{F}, \mathbb{P})$ and indexed by T is called *adapted* to the filtration if for every $t \in T$, the random variable X_t is $\mathcal{F}_t$-measurable.

We should think of a filtration as a flow of information. The σ-algebra $\mathcal{F}_t$ contains the events that can happen 'up to time t'. An adapted process is a process that 'does not look into the future'. If X is a stochastic process, we can consider the filtration $(\mathcal{F}^X_t)_{t\in T}$ defined by

$$\mathcal{F}^X_t = \sigma(X_s : s \leq t).$$

We call this filtration the *filtration generated by* X, or the *natural filtration of* X. Intuitively, the natural filtration of a process keeps track of the 'history' of the process. A stochastic process is always adapted to its natural filtration.

If $(\mathcal{F}_t)$ is a filtration, then for $t \in T$ we define the σ-algebra

$$\mathcal{F}_{t+} = \bigcap_{n=1}^{\infty} \mathcal{F}_{t+1/n}.$$

This is the σ-algebra $\mathcal{F}_t$, augmented with the events that 'happen immediately after time t'. The collection $(\mathcal{F}_{t+})_{t \in T}$ is again a filtration (see Exercise 14). Cases in which it coincides with the original filtration are of special interest.

Definition 1.6.2. We call a filtration $(\mathcal{F}_t)$ *right-continuous* if $\mathcal{F}_{t+} = \mathcal{F}_t$ for every t.

Intuitively, the right-continuity of a filtration means that 'nothing can happen in an infinitesimally small time interval'. Note that for every filtration $(\mathcal{F}_t)$, the corresponding filtration $(\mathcal{F}_{t+})$ is right-continuous.

In addition to right-continuity it is often assumed that $\mathcal{F}_0$ contains all events in $\mathcal{F}_\infty$ that have probability 0, where

$$\mathcal{F}_\infty = \sigma\left(\mathcal{F}_t : t \geq 0\right).$$

As a consequence, every $\mathcal{F}_t$ then contains all those events.

Definition 1.6.3. A filtration $(\mathcal{F}_t)$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to satisfy the *usual conditions* if it is right-continuous and $\mathcal{F}_0$ contains all the $\mathbb{P}$-negligible events in $\mathcal{F}_\infty$.

We now introduce a very important class of 'random times' that can be associated with a filtration.

Definition 1.6.4. A $[0, \infty]$-valued random variable τ is called a *stopping time* with respect to the filtration $(\mathcal{F}_t)$ if for every $t \in T$ it holds that $\{\tau \leq t\} \in \mathcal{F}_t$. If $\tau < \infty$ almost surely, we call the stopping time *finite*.

Loosely speaking, τ is a stopping time if for every $t \in T$ we can determine whether τ has occurred before time t on the basis of the information that we have up to time t. With a stopping time τ we associate the σ-algebra

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t \text{ for all } t \in T\}$$

(see Exercise 15). This should be viewed as the collection of all events that happen prior to the stopping time τ. Note that the notation causes no confusion since a deterministic time $t \in T$ is clearly a stopping time and its associated σ-algebra is simply the σ-algebra $\mathcal{F}_t$.

If the filtration $(\mathcal{F}_t)$ is right-continuous, then τ is a stopping time if and only if $\{\tau < t\} \in \mathcal{F}_t$ for every $t \in T$ (see Exercise 21). For general filtrations, we introduce the following class of random times.

Definition 1.6.5. A $[0, \infty]$-valued random variable τ is called an *optional time* with respect to the filtration $\mathbb{F}$ if for every $t \in T$ it holds that $\{\tau < t\} \in \mathcal{F}_t$. If $\tau < \infty$ almost surely, we call the optional time *finite*.

Lemma 1.6.6. *τ is an optional time with respect to $(\mathcal{F}_t)$ if and only if it is a stopping time with respect to $(\mathcal{F}_{t+})$. Every stopping time is an optional time.*

Proof. See Exercise 22. □

The so-called *hitting times* form an important class of stopping times and optional times. The hitting time of a set B is the first time that a process visits that set.

Lemma 1.6.7. *Let (E, d) be a metric space. Suppose that $X = (X_t)_{t \geq 0}$ is a continuous, E-valued process and that B is a closed set in E. Then the random variable*

$$\sigma_B = \inf\{t \geq 0 : X_t \in B\}$$

is an $(\mathcal{F}_t^X)$-stopping time. [2]

Proof. Denote the distance of a point $x \in E$ to the set B by $d(x, B)$, so

$$d(x, B) = \inf\{d(x, y) : y \in B\}.$$

Since X is continuous, the real-valued process $Y_t = d(X_t, B)$ is continuous as well. Moreover, since B is closed, it holds that $X_t \in B$ if and only if $Y_t = 0$. Using the continuity of Y, it follows that $\sigma_B > t$ if and only if $Y_s > 0$ for all $s \leq t$ (check!). But Y is continuous and $[0, t]$ is compact, so we have

$$\begin{aligned}\{\sigma_B > t\} &= \{Y_s \text{ is bounded away from } 0 \text{ for all } s \in \mathbb{Q} \cap [0, t]\} \\ &= \{X_s \text{ is bounded away from } B \text{ for all } s \in \mathbb{Q} \cap [0, t]\}.\end{aligned}$$

The event on the right-hand side clearly belongs to $\mathcal{F}_t^X$. □

Lemma 1.6.8. *Let (E, d) be a metric space. Suppose that $X = (X_t)_{t \geq 0}$ is a right-continuous, E-valued process and that B is an open set in E. Then the random variable*

$$\tau_B = \inf\{t > 0 : X_t \in B\}$$

is an $(\mathcal{F}_t^X)$-optional time.

---

[2] As usual, we define $\inf \varnothing = \infty$.

Proof. Since B is open and X is right-continuous, it holds that $X_t \in B$ if and only if there exists an $\varepsilon > 0$ such that $X_s \in B$ for all $s \in [t, t+\delta]$. Since this interval always contains a rational number, it follows that

$$\{\tau_B < t\} = \bigcup_{\substack{s<t \\ s\in\mathbb{Q}}} \{X_s \in B\}.$$

The event on the right-hand side is clearly an element of $\mathcal{F}_t^X$. $\square$

Example 1.6.9. Let W be a BM and for $x > 0$, consider the random variable

$$\tau_x = \inf\{t > 0 : W_t = x\}.$$

Since $x > 0$ and W is continuous, τ_x can be written as

$$\tau_x = \inf\{t \geq 0 : W_t = x\}.$$

By Lemma 1.6.7 this is an $(\mathcal{F}_t^W)$-stopping time. Moreover, by the recurrence of the BM (see Corollary 1.4.6), τ_x is a finite stopping time. $\blacksquare$

We often want to consider a stochastic process X, evaluated at a finite stopping τ. However, it is not a-priori clear that the map $\omega \mapsto X_{\tau(\omega)}(\omega)$ is measurable, i.e. that X_τ is in fact a random variable. This motivates the following definition.

Definition 1.6.10. An $(E, \mathcal{E})$-valued stochastic process X is called *progressively measurable* with respect to the filtration $(\mathcal{F}_t)$ if for every $t \in T$, the map $(s, \omega) \mapsto X_s(\omega)$ is measurable as a map from $([0,t] \times \Omega, \mathcal{B}([0,t]) \times \mathcal{F}_t)$ to $(E, \mathcal{E})$.

Lemma 1.6.11. *Every adapted, right-continuous, $\mathbb{R}^d$-valued process X is progressively measurable.*

Proof. Let $t \geq 0$ be fixed. For $n \in \mathbb{N}$, define the process

$$X_s^n = X_0 1_{\{0\}}(s) + \sum_{k=1}^n X_{kt/n} 1_{((k-1)t/n, kt/n]}(s), \quad s \in [0,t].$$

Clearly, the map $(s, \omega) \mapsto X_s^n(\omega)$ on $[0,t] \times \Omega$ is $\mathcal{B}([0,t]) \times \mathcal{F}_t$-measurable. Now observe that since X is right-continuous, the map $(s, \omega) \mapsto X_s^n(\omega)$ converges pointwise to the map $(s, \omega) \mapsto X_s(\omega)$ as $n \to \infty$. It follows that the latter map is $\mathcal{B}([0,t]) \times \mathcal{F}_t$-measurable as well. $\square$

Lemma 1.6.12. *Suppose that X is a progressively measurable process and let τ be a finite stopping time. Then X_τ is an $\mathcal{F}_\tau$-measurable random variable.*

Proof. To prove that X_τ is $\mathcal{F}_\tau$-measurable, we have to show that for every $B \in \mathcal{E}$ and every $t \geq 0$, it holds that $\{X_\tau \in B\} \cap \{\tau \leq t\} \in \mathcal{F}_t$. Hence, it suffices to show that the map $\omega \mapsto X_{\tau(\omega)\wedge t}(\omega)$ is $\mathcal{F}_t$-measurable. This map is the composition of the maps $\omega \mapsto (\tau(\omega) \wedge t, \omega)$ from Ω to $[0,t] \times \Omega$ and $(s,\omega) \mapsto X_s(\omega)$ from $[0,t] \times \Omega$ to E. The first map is measurable as a map from $(\Omega, \mathcal{F}_t)$ to $([0,t] \times \Omega, \mathcal{B}([0,t]) \times \mathcal{F}_t)$ (see Exercise 23). Since X is progressively measurable, the second map is measurable as a map from $([0,t]\times\Omega, \mathcal{B}([0,t])\times\mathcal{F}_t)$ to $(E, \mathcal{E})$. This completes the proof, since the composition of measurable maps is again measurable. □

It is often needed to consider a stochastic process X up to a given stopping time τ. For this purpose we define the *stopped process* X^τ by

$$X_t^\tau = X_{\tau\wedge t} = \begin{cases} X_t & \text{if } t < \tau, \\ X_\tau & \text{if } t \geq \tau. \end{cases}$$

By Lemma 1.6.12 and Exercises 16 and 18, we have the following result.

Lemma 1.6.13. *If X is progressively measurable with respect to $(\mathcal{F}_t)$ and τ an $(\mathcal{F}_t)$-stopping time, then the stopped process X^τ is adapted to the filtrations $(\mathcal{F}_{\tau\wedge t})$ and $(\mathcal{F}_t)$.*

In the subsequent chapters we repeatedly need the following technical lemma. It states that every stopping time is the decreasing limit of a sequence of stopping times that take on only finitely many values.

Lemma 1.6.14. *Let τ be a stopping time. Then there exist stopping times τ_n that only take finitely many values and such that $\tau_n \downarrow \tau$ almost surely.*

Proof. Define

$$\tau_n = \sum_{k=1}^{n2^n-1} \frac{k}{2^n} 1_{\{\tau\in[(k-1)/2^n, k/2^n)\}} + \infty\, 1_{\{\tau\geq n\}}.$$

Then τ_n is a stopping time and $\tau_n \downarrow \tau$ almost surely (see Exercise 24). □

Using the notion of filtrations, we can extend the definition of the BM as follows.

Definition 1.6.15. Suppose that on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we have a filtration $(\mathcal{F}_t)_{t\geq 0}$ and an adapted stochastic process $W = (W_t)_{t\geq 0}$. Then W is called a (standard) *Brownian motion*, (or *Wiener process) with respect to the filtration* $(\mathcal{F}_t)$ if

(i) $W_0 = 0$,

(ii) $W_t - W_s$ is independent of $\mathcal{F}_s$ for all $s \leq t$,

(iii) $W_t - W_s$ has a $N(0, t-s)$-distribution for all $s \leq t$,

(iv) almost all sample paths of W are continuous.

Clearly, a process W that is a BM in the sense of the 'old' Definition 1.1.2 is a BM with respect to its natural filtration. If in the sequel we do not mention the filtration of a BM explicitly, we mean the natural filtration. However, we will see that it is sometimes necessary to consider Brownian motions with respect to larger filtrations as well.

1.7 Exercises

1. Prove that a process W has properties (i), (ii) and (iii) of Definition 1.1.2 if and only if for all $t_1, \ldots, t_n \geq 0$ the vector $(W_{t_1}, \ldots, W_{t_n})$ has an n-dimensional Gaussian distribution with mean vector 0 and covariance matrix $(t_i \wedge t_j)_{i,j=1..n}$.

2. Give an example of two processes that are versions of each other, but not modifications.

3. Prove that the process Y defined in the proof of Theorem 1.3.3 is indeed a modification of the process X.

4. Let $\alpha > 0$ be given. Give an example of a right-continuous process X that is not continuous and which statisfies

$$\mathbb{E}|X_t - X_s|^\alpha \leq K|t - s|$$

for some $K > 0$ and all $s, t \geq 0$. (Hint: consider a process of the form $X_t = 1_{\{Y \leq t\}}$ for a suitable chosen random variable Y).

5. Prove that the process X in the proof of Corollary 1.3.4 is a BM.

6. Prove Lemma 1.4.2.

7. Prove parts (i), (ii) and (iii) of Theorem 1.4.4.

8. Consider the proof of the time-inversion property (iv) of Theorem 1.4.4. Prove that the probability in (1.5) is determined by the fdd's of the process X.

9. Let W be a BM and define $X_t = W_{1-t} - W_1$ for $t \in [0, 1]$. Show that $(X_t)_{t \in [0,1]}$ is a BM as well.

10. Let W be a BM and fix $t > 0$. Define the process B by

$$B_s = W_{s \wedge t} - (W_s - W_{s \wedge t}) = \begin{cases} W_s, & s \leq t \\ 2W_t - W_s, & s > t. \end{cases}$$

Draw a picture of the processes W and B and show that B is again a BM. We will see another version of this so-called *reflection principle* in Chapter 3.

11. (i) Let W be a BM and define the process $X_t = W_t - tW_1$, $t \in [0, 1]$. Determine the mean and covariance function of X.

    (ii) The process X of part (i) is called the (standard) *Brownian bridge* on $[0, 1]$, and so is every other continuous, Gaussian process indexed by the interval $[0, 1]$ that has the same mean and covariance function. Show that the processes Y and Z defined by $Y_t = (1 - t)W_{t/(1-t)}$, $t \in [0, 1)$, $Y_1 = 0$ and $Z_0 = 0$, $Z_t = tW_{(1/t)-1}$, $t \in (0, 1]$ are standard Brownian bridges.

12. Let $H \in (0,1)$ be given. A continuous, zero-mean Gaussian process X with covariance function $2\mathbb{E}X_sX_t = (t^{2H} + s^{2H} - |t-s|^{2H})$ is called a *fractional Brownian motion (fBm)* with *Hurst index* H. Show that the fBm with Hurst index $1/2$ is simply the BM. Show that if X is a fBm with Hurst index H, then for all $a > 0$ the process $a^{-H}X_{at}$ is a fBm with Hurst index H as well.

13. Let W be a Brownian motion and fix $t > 0$. For $n \in \mathbb{N}$, let π_n be a partition of $[0,t]$ given by $0 = t_0^n < t_1^n < \cdots < t_{k_n}^n = t$ and suppose that the mesh $\|\pi_n\| = \max_k |t_k^n - t_{k-1}^n|$ tends to zero as $n \to \infty$. Show that
$$\sum_k (W_{t_k^n} - W_{t_{k-1}^n})^2 \xrightarrow{L^2} t$$
as $n \to \infty$. (Hint: show that the expectation of the sum tends to t, and the variance to zero.)

14. Show that if $(\mathcal{F}_t)$ is a filtration, $(\mathcal{F}_{t+})$ is a filtration as well.

15. Prove that the collection $\mathcal{F}_\tau$ associated with a stopping time τ is a σ-algebra.

16. Show that if σ and τ are stopping times such that $\sigma \le \tau$, then $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$.

17. Let σ and τ be two $(\mathcal{F}_t)$-stopping times. Show that $\{\sigma \le \tau\} \in \mathcal{F}_\sigma \cap \mathcal{F}_\tau$.

18. If σ and τ are stopping times w.r.t. the filtration $(\mathcal{F}_t)$, show that $\sigma \wedge \tau$ and $\sigma \vee \tau$ are also stopping times and determine the associated σ-algebras.

19. Show that if σ and τ are stopping times w.r.t. the filtration $(\mathcal{F}_t)$, then $\sigma + \tau$ is a stopping time as well. (Hint: for $t > 0$, write
$$\begin{aligned}\{\sigma + \tau > t\} = \{\tau = 0, \sigma > t\} &\cup \{0 < \tau < t, \sigma + \tau > t\} \\ &\cup \{\tau > t, \sigma = 0\} \cup \{\tau \ge t, \sigma > 0\}.\end{aligned}$$
Only for the second event on the right-hand side it is non-trivial to prove that it belongs to $\mathcal{F}_t$. Now observe that if $\tau > 0$, then $\sigma + \tau > t$ if and only if there exists a positive $q \in \mathbb{Q}$ such that $q < \tau$ and $\sigma + q > t$.)

20. Show that if σ and τ are stopping times with respect to the filtration $(\mathcal{F}_t)$ and X is an integrable random variable, then a.s. $1_{\{\tau=\sigma\}}\mathbb{E}(X \mid \mathcal{F}_\tau) = 1_{\{\tau=\sigma\}}\mathbb{E}(X \mid \mathcal{F}_\sigma)$. (Hint: show that $\mathbb{E}(1_{\{\tau=\sigma\}}X \mid \mathcal{F}_\tau) = \mathbb{E}(1_{\{\tau=\sigma\}}X \mid \mathcal{F}_\tau \cap \mathcal{F}_\sigma)$.)

21. Show that if the filtration $(\mathcal{F}_t)$ is right-continuous, then τ is an $(\mathcal{F}_t)$-stopping time if and only if $\{\tau < t\} \in \mathcal{F}_t$ for all $t \in T$.

22. Prove Lemma 1.6.6.

23. Show that the map $\omega \mapsto (\tau(\omega) \wedge t, \omega)$ in the proof of Lemma 1.6.12 is measurable as a map from $(\Omega, \mathcal{F}_t)$ to $([0,t] \times \Omega, \mathcal{B}([0,t]) \times \mathcal{F}_t)$.

24. Show that the τ_n in the proof of Lemma 1.6.14 are indeed stopping times and that they converge to τ almost surely.

25. Translate the definitions of Section 1.6 to the special case that time is discrete, i.e. $T = \mathbb{Z}_+$.

26. Let W be a BM and let $Z = \{t \geq 0 : W_t = 0\}$ be its zero set. Show that with probability one, the set Z has Lebesgue measure zero, is closed and unbounded.

2

Martingales

2.1 Definitions and examples

In this chapter we introduce and study a very important class of stochastic processes: the so-called martingales. Martingales arise naturally in many branches of the theory of stochastic processes. In particular, they are very helpful tools in the study of the BM. In this section, the index set T is an arbitrary interval of $\mathbb{Z}_+$ or $\mathbb{R}_+$.

Definition 2.1.1. An $(\mathcal{F}_t)$-adapted, real-valued process M is called a *martingale* (with respect to the filtration $(\mathcal{F}_t)$) if

(i) $\mathbb{E}|M_t| < \infty$ for all $t \in T$,

(ii) $\mathbb{E}(M_t \,|\, \mathcal{F}_s) \overset{\text{as}}{=} M_s$ for all $s \leq t$.

If property (ii) holds with '$\geq$' (resp. '$\leq$') instead of '$=$', then M is called a *submartingale* (resp. *supermartingale*).

Intuitively, a martingale is a process that is 'constant on average'. Given all information up to time s, the best guess for the value of the process at time $t \geq s$ is simply the current value M_s. In particular, property (ii) implies that $\mathbb{E}M_t = \mathbb{E}M_0$ for all $t \in T$. Likewise, a submartingale is a process that increases on average, and a supermartingale decreases on average. Clearly, M is a submartingale if and only if $-M$ is a supermartingale and M is a martingale if it is both a submartingale and a supermartingale.

The basic properties of conditional expectations (see Appendix A) give us the following examples.

Example 2.1.2. Suppose that X is an integrable random variable and $(\mathcal{F}_t)_{t\in T}$ a filtration. For $t \in T$, define $M_t = \mathbb{E}(X \,|\, \mathcal{F}_t)$, or, more precisely, let M_t be a version of $\mathbb{E}(X \,|\, \mathcal{F}_t)$. Then $M = (M_t)_{t\in T}$ is an $(\mathcal{F}_t)$-martingale and M is uniformly integrable (see Exercise 1). ■

Example 2.1.3. Suppose that M is a martingale and that φ is a convex function such that $\mathbb{E}|\varphi(M_t)| < \infty$ for all $t \in T$. Then the process $\varphi(M)$ is a submartingale. The same is true if M is a submartingale and φ is an increasing, convex function (see Exercise 2). ■

The BM generates many examples of martingales. The most important ones are presented in the following example.

Example 2.1.4. Let W be a BM. Then the following processes are martingales with respect to the same filtration:

(i) W itself,

(ii) $W_t^2 - t$,

(iii) for every $a \in \mathbb{R}$, the process $\exp(aW_t - a^2t/2)$.

You are asked to prove this in Exercise 3. ■

In the next section we first develop the theory for discrete-time martingales. The generalization to continuous time is carried out in Section 2.3. In Section 2.4 we return to our study of the BM.

2.2 Discrete-time martingales

In this section we restrict ourselves to martingales and (filtrations) that are indexed by (a subinterval of) $\mathbb{Z}_+$. Note that as a consequence, it only makes sense to consider $\overline{\mathbb{Z}}_+$-valued stopping times. In discrete time, τ a stopping time with respect to the filtration $(\mathcal{F}_n)_{n\in\mathbb{Z}_+}$ if $\{\tau \le n\} \in \mathcal{F}_n$ for all $n \in \mathbb{N}$.

2.2.1 Martingale transforms

If the value of a process at time n is already known at time $n-1$, we call a process predictable. The precise definition is as follows.

Definition 2.2.1. We call a discrete-time process X *predictable* with respect to the filtration $(\mathcal{F}_n)$ if X_n is $\mathcal{F}_{n-1}$-measurable for every n.

In the following definition we introduce discrete-time 'integrals'. This is a useful tool in martingale theory.

Definition 2.2.2. Let M and X be two discrete-time processes. We define the process $X\cdot M$ by $(X\cdot M)_0 = 0$ and

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

We call $X\cdot M$ the *discrete integral of X with respect to M*. If M is a (sub-, super-)martingale, it is often called the *martingale transform of M by X*.

The following lemma explains why these 'integrals' are so useful: the integral of a predictable process with respect to a martingale is again a martingale.

Lemma 2.2.3. *If M is a submartingale (resp. supermartingale) and X is a bounded, nonnegative predictable process, then $X\cdot M$ is a submartingale (resp. supermartingale) as well. If M is a martingale and X is a bounded, predictable process, then $X\cdot M$ is a martingale.*

Proof. Put $Y = X\cdot M$. Clearly, the process Y is adapted. Since X is bounded, say $|X_n| \le K$ for all n, we have $\mathbb{E}|Y_n| \le 2K\sum_{k\le n}\mathbb{E}|M_k| < \infty$. Now suppose first that M is a submartingale and X is nonnegative. Then a.s.

$$\begin{aligned}\mathbb{E}(Y_n \mid \mathcal{F}_{n-1}) &= \mathbb{E}\left(Y_{n-1} + X_n(M_n - M_{n-1}) \mid \mathcal{F}_{n-1}\right)\\ &= Y_{n-1} + X_n\mathbb{E}(M_n - M_{n-1} \mid \mathcal{F}_{n-1}) \ge Y_{n-1},.\end{aligned}$$

hence Y is a submartingale. If M is a martingale the last inequality is an equality, irrespective of the sign of X_n, which implies that Y is a martingale in that case. □

Using this lemma it is easy to see that a stopped (sub-, super-)martingale is again a (sub-, super-)martingale.

Theorem 2.2.4. *Let M be a (sub-, super-)martingale and τ a stopping time. Then the stopped process M^τ is a (sub-, super-)martingale as well.*

Proof. Define the process X by $X_n = 1_{\{\tau\ge n\}}$ and verify that $M^\tau = M_0 + X\cdot M$. Since τ is a stopping time, we have $\{\tau \ge n\} = \{\tau \le n-1\}^c \in \mathcal{F}_{n-1}$, which shows that the process X is predictable. It is also a bounded process, so the statement follows from the preceding lemma. □

The following result can be viewed as a first version of the so-called *optional stopping theorem*. The general version will be treated in Section 2.2.5.

Theorem 2.2.5. *Let M be a submartingale and σ, τ two stopping times such that $\sigma \le \tau \le K$ for some constant $K > 0$. Then*

$$\mathbb{E}(M_\tau \,|\, \mathcal{F}_\sigma) \ge M_\sigma \quad \text{a.s.}$$

An adapted process M is a martingale if and only if

$$\mathbb{E}M_\tau = \mathbb{E}M_\sigma$$

for any such pair of stopping times.

Proof. Suppose for simplicity that M is a martingale and define the predictable process $X_n = 1_{\{\tau \ge n\}} - 1_{\{\sigma \ge n\}}$, so that $X \cdot M = M^\tau - M^\sigma$. By Lemma 2.2.3 the process $X \cdot M$ is a martingale, hence $\mathbb{E}(M_n^\tau - M_n^\sigma) = \mathbb{E}(X \cdot M)_n = 0$ for all n. For $\sigma \le \tau \le K$, it follows that

$$\mathbb{E}M_\tau = \mathbb{E}M_K^\tau = \mathbb{E}M_K^\sigma = \mathbb{E}M_\sigma.$$

Now take $A \in \mathcal{F}_\sigma$ and define the 'truncated random times'

$$\sigma^A = \sigma 1_A + K 1_{A^c}, \quad \tau^A = \tau 1_A + K 1_{A^c}.$$

By definition of $\mathcal{F}_\sigma$ it holds for every n that

$$\{\sigma^A \le n\} = (A \cap \{\sigma \le n\}) \cup (A^c \cap \{K \le n\}) \in \mathcal{F}_n,$$

so σ^A is a stopping time. Similarly, τ^A is a stopping time, and we clearly have $\sigma^A \le \tau^A \le K$. By the first part of the proof, it follows that $\mathbb{E}M_{\sigma^A} = \mathbb{E}M_{\tau^A}$, which implies that

$$\int_A M_\sigma \, d\mathbb{P} = \int_A M_\tau \, d\mathbb{P}.$$

Since $A \in \mathcal{F}_\sigma$ is arbitrary, this shows that $\mathbb{E}(M_\tau \,|\, \mathcal{F}_\sigma) = M_\sigma$ almost surely (recall that M_σ is $\mathcal{F}_\sigma$-measurable, cf. Lemma 1.6.12).

If in the preceding paragraph σ is replaced by $n-1$ and τ is replaced by n, the argument shows that if M is an adapted process for which $\mathbb{E}M_\tau = \mathbb{E}M_\sigma$ for every bounded pair $\sigma \le \tau$ of stopping times, then $\mathbb{E}(M_n \,|\, \mathcal{F}_{n-1}) = M_{n-1}$ a.s., i.e. M is a martingale.

If M is a submartingale, the same reasoning applies, but with inequalities instead of equalities. □

2.2.2 Inequalities

Markov's inequality implies that if M is a discrete-time process, then $\lambda \mathbb{P}(M_n \ge \lambda) \le \mathbb{E}|M_n|$ for all $n \in \mathbb{N}$ and $\lambda > 0$. Doob's classical submartingale inequality states that for submartingales, we have a much stronger result.

Theorem 2.2.6 (Doob's submartingale inequality). *Let M be a submartingale. For all $\lambda > 0$ and $n \in \mathbb{N}$,*

$$\lambda \mathbb{P}\left(\max_{k \le n} M_k \ge \lambda\right) \le \mathbb{E} M_n 1_{\{\max_{k \le n} M_k \ge \lambda\}} \le \mathbb{E}|M_n|.$$

Proof. Define $\tau = n \wedge \inf\{k : M_k \ge \lambda\}$. This is a stopping time and $\tau \le n$ (see Lemma 1.6.7), so by Theorem 2.2.5, we have $\mathbb{E}M_n \ge \mathbb{E}M_\tau$. It follows that

$$\begin{aligned}\mathbb{E}M_n &\ge \mathbb{E}M_\tau 1_{\{\max_{k \le n} M_k \ge \lambda\}} + \mathbb{E}M_\tau 1_{\{\max_{k \le n} M_k < \lambda\}} \\ &\ge \lambda \mathbb{P}\left(\max_{k \le n} M_k \ge \lambda\right) + \mathbb{E}M_n 1_{\{\max_{k \le n} M_k < \lambda\}}.\end{aligned}$$

This yields the first inequality, the second one is obvious. □

Theorem 2.2.7 (Doob's L^p-inequality). *If M is a martingale or a nonnegative submartingale and $p > 1$, then for all $n \in \mathbb{N}$*

$$\mathbb{E}\left(\max_{k \le n} |M_n|^p\right) \le \left(\frac{p}{p-1}\right)^p \mathbb{E}\,|M_n|^p\,.$$

Proof. We define $M^* = \max_{k \le n} |M_k|$. By Fubini's theorem, we have for every $m \in \mathbb{N}$

$$\begin{aligned}\mathbb{E}|M^* \wedge m|^p &= \mathbb{E}\int_0^{M^* \wedge m} px^{p-1}\,dx \\ &= \mathbb{E}\int_0^m px^{p-1} 1_{\{M^* \ge x\}}\,dx \\ &= \int_0^m px^{p-1}\mathbb{P}(M^* \ge x)\,dx.\end{aligned}$$

By the conditional version of Jensen's inequality, the process $|M|$ is a submartingale (see Example 2.1.3). Hence, the preceding theorem implies that

$$\begin{aligned}\mathbb{E}|M^* \wedge m|^p &\le \int_0^m px^{p-2}\mathbb{E}|M_n| 1_{\{M^* \ge x\}}\,dx \\ &= p\mathbb{E}\left(|M_n| \int_0^{M^* \wedge m} x^{p-2}\,dx\right) \\ &= \frac{p}{p-1}\mathbb{E}\Big(|M_n||M^* \wedge m|^{p-1}\Big).\end{aligned}$$

By Hölder's inequality, it follows that with $1/p + 1/q = 1$,

$$\mathbb{E}|M^* \wedge m|^p \le \frac{p}{p-1}\Big(\mathbb{E}|M_n|^p\Big)^{1/p}\Big(\mathbb{E}|M^* \wedge m|^{(p-1)q}\Big)^{1/q}.$$

Since $p > 1$ we have $q = p/(p-1)$, so

$$\mathbb{E}|M^* \wedge m|^p \leq \frac{p}{p-1}\Big(\mathbb{E}|M_n|^p\Big)^{1/p}\Big(\mathbb{E}|M^* \wedge m|^p\Big)^{(p-1)/p}.$$

Now raise both sides to the pth power and cancel common factors to see that

$$\mathbb{E}|M^* \wedge m|^p \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}|M_n|^p.$$

The proof is completed by letting m tend to infinity. □

2.2.3 Doob decomposition

An adapted, integrable process X can always be written as a sum of a martingale and a predictable process. This is called the *Doob decomposition* of the process X.

Theorem 2.2.8. *Let X be an adapted, integrable process. There exist a martingale M and a predictable process A such that $A_0 = M_0 = 0$ and $X = X_0 + M + A$. The processes M and A are a.s. unique. The process X is a submartingale if and only if A is increasing.*

Proof. Suppose first that there exist a martingale M and a predictable process A such that $A_0 = M_0 = 0$ and $X = X_0 + M + A$. Then the martingale property of M and the predictability of A show that

$$\mathbb{E}(X_n - X_{n-1} \,|\, \mathcal{F}_{n-1}) = A_n - A_{n-1}. \tag{2.1}$$

Since $A_0 = 0$ it follows that

$$A_n = \sum_{k=1}^{n} \mathbb{E}(X_k - X_{k-1} \,|\, \mathcal{F}_{k-1}) \tag{2.2}$$

for $n \geq 1$, and hence $M_n = X_n - X_0 - A_n$. This shows that M and A are a.s. unique.

Conversely, given a process X, (2.2) defines a predictable process A, and it is easily seen that the process M defined by $M = X - X_0 - A$ is a martingale. This proves existence of the decomposition.

Equation (2.1) show that X is a submartingale if and only if the process A is increasing. □

Using the Doob decomposition in combination with the submartingale inequality, we obtain the following result.

Theorem 2.2.9. *Let X be a submartingale or a supermartingale. For all $\lambda > 0$ and $n \in \mathbb{N}$,*

$$\lambda \mathbb{P}\left(\max_{k\leq n} |X_k| \geq 3\lambda\right) \leq 4\mathbb{E}|X_0| + 3\mathbb{E}|X_n|.$$

Proof. Suppose that X is a submartingale. By the preceding theorem there exist a martingale M and an increasing, predictable process A such that $M_0 = A_0 = 0$ and $X = X_0 + M + A$. By the triangle inequality and the fact that A is increasing,

$$\mathbb{P}\left(\max_{k\leq n} |X_k| \geq 3\lambda\right) \leq \mathbb{P}\left(|X_0| \geq \lambda\right) + \mathbb{P}\left(\max_{k\leq n} |M_k| \geq \lambda\right) + \mathbb{P}\left(A_n \geq \lambda\right).$$

Hence, by Markov's inequality and the submartingale inequality,

$$\lambda \mathbb{P}\left(\max_{k\leq n} |X_k| \geq 3\lambda\right) \leq \mathbb{E}|X_0| + \mathbb{E}|M_n| + \mathbb{E}A_n.$$

Since $M_n = X_n - X_0 - A_n$, the right-hand side is bounded by $2\mathbb{E}|X_0| + \mathbb{E}|X_n| + 2\mathbb{E}A_n$. We know that A_n is given by (2.2). Taking expectations in the latter expression shows that $\mathbb{E}A_n = \mathbb{E}X_n - \mathbb{E}X_0 \leq \mathbb{E}|X_0| + \mathbb{E}|X_n|$. This completes the proof. □

2.2.4 Convergence theorems

Let M be a supermartingale and consider a compact interval $[a, b] \subseteq \mathbb{R}$. The number of *upcrossings* of $[a, b]$ that the process M makes up to time n is the number of times that the process 'moves' from a level below a to a level above b. The precise definition is as follows.

Definition 2.2.10. The number $U_n[a, b]$ is the largest $k \in \mathbb{Z}_+$ such that there exist $0 \leq s_1 < t_1 < s_2 < t_2 < \cdots < s_k < t_k \leq n$ with $M_{s_i} < a$ and $M_{t_i} > b$.

Lemma 2.2.11 (Doob's upcrossings lemma). *Let M be a supermartingale. Then for all $a < b$, the number of upcrossings $U_n[a, b]$ of the interval $[a, b]$ by M up to time n satisfies*

$$(b - a)\mathbb{E}U_n[a, b] \leq \mathbb{E}(M_n - a)^-.$$

Proof. Consider the bounded, predictable process X given by $X_0 = 1_{\{M_0<a\}}$ and

$$X_n = 1_{\{X_{n-1}=1\}} 1_{\{M_{n-1}\leq b\}} + 1_{\{X_{n-1}=0\}} 1_{\{M_{n-1}<a\}}$$

for $n \in \mathbb{N}$, and define $Y = X \cdot M$. The process X is 0 until M drops below the level a, then is 1 until M gets above b etc. So every completed upcrossing of

$[a, b]$ increases the value of Y by at least $(b - a)$. If the last upcrossing has not been completed at time n, this can cause Y to decrease by at most $(M_n - a)^-$. Hence, we have the inequality

$$Y_n \geq (b - a)U_n[a, b] - (M_n - a)^-. \tag{2.3}$$

By Lemma 2.2.3, the process $Y = X \cdot M$ is a supermartingale. In particular, it holds that $\mathbb{E}Y_n \leq \mathbb{E}Y_0 = 0$. Hence, the proof is completed by taking expectations on both sides of (2.3). □

Observe that the upcrossings lemma implies that if M is a supermartingale that is bounded in L^1, i.e. for which $\sup_n \mathbb{E}|M_n| < \infty$, then $\mathbb{E}U_\infty[a, b] < \infty$ for all $a \leq b$. In particular, the total number $U_\infty[a, b]$ of upcrossings of the interval $[a, b]$ is finite almost surely. The proof of the classical martingale convergence theorem is now straightforward.

Theorem 2.2.12 (Doob's martingale convergence theorem). *If M is a supermartingale that is bounded in L^1, then M_n converges almost surely to a limit M_∞ as $n \to \infty$, and $\mathbb{E}|M_\infty| < \infty$.*

Proof. Suppose that M_n does not converge to a limit in $[-\infty, \infty]$. Then there exist two rational numbers $a < b$ such that $\liminf M_n < a < b < \limsup M_n$. In particular, we must have $U_\infty[a, b] = \infty$. This contradicts the fact that by the upcrossings lemma, the number $U_\infty[a, b]$ is finite with probability 1. We conclude that almost surely, M_n converges to a limit M_∞ in $[-\infty, \infty]$. By Fatou's lemma,

$$\mathbb{E}|M_\infty| = \mathbb{E}(\liminf |M_n|) \leq \liminf \mathbb{E}(|M_n|) \leq \sup \mathbb{E}|M_n| < \infty.$$

This completes the proof. □

If the supermartingale M is not only bounded in L^1 but also uniformly integrable, then in addition to almost sure convergence we have convergence in L^1. Moreover, in that case the whole sequence $M_0, M_1, \ldots, M_\infty$ is a supermartingale.

Theorem 2.2.13. *Let M be a supermartingale that is bounded in L^1. Then $M_n \to M_\infty$ in L^1 if and only if $\{M_n : n \in \mathbb{Z}_+\}$ is uniformly integrable. In that case*

$$\mathbb{E}(M_\infty \,|\, \mathcal{F}_n) \leq M_n \quad \text{a.s.},$$

with equality if M is a martingale.

Proof. By the preceding theorem the convergence $M_n \to M_\infty$ holds almost surely, so the first statement follows from Theorem A.3.5 in the appendix. To prove the second statement, suppose that $M_n \to M_\infty$ in L^1. Since M is a supermartingale we have

$$\int_A M_m \, d\mathbb{P} \leq \int_A M_n \, d\mathbb{P} \tag{2.4}$$

for all $A \in \mathcal{F}_n$ and $m \geq n$. For the integral on the left-hand side it holds that

$$\left|\int_A M_m \, d\mathbb{P} - \int_A M_\infty \, d\mathbb{P}\right| \leq \mathbb{E}|M_m - M_\infty| \to 0$$

as $m \to \infty$. Hence, by letting m tend to infinity in (2.4) we find that

$$\int_A M_\infty \, d\mathbb{P} \leq \int_A M_n \, d\mathbb{P}$$

for every $A \in \mathcal{F}_n$. This completes the proof. □

If X is an integrable random variable and $(\mathcal{F}_n)$ is a filtration, then $\mathbb{E}(X \mid \mathcal{F}_n)$ is a uniformly integrable martingale (cf. Example 2.1.2). For martingales of this type we can identify the limit explicitly in terms of the 'limit σ-algebra' $\mathcal{F}_\infty$ defined by

$$\mathcal{F}_\infty = \sigma\left(\bigcup_n \mathcal{F}_n\right).$$

Theorem 2.2.14 (Lévy's upward theorem). *Let X be an integrable random variable and $(\mathcal{F}_n)$ a filtration. Then as $n \to \infty$,*

$$\mathbb{E}(X \mid \mathcal{F}_n) \to \mathbb{E}(X \mid \mathcal{F}_\infty)$$

almost surely and in L^1.

Proof. The process $M_n = \mathbb{E}(X \mid \mathcal{F}_n)$ is a uniformly integrable martingale (see Example 2.1.2 and Lemma A.3.4). Hence, by Theorem 2.2.13, $M_n \to M_\infty$ almost surely and in L^1. It remains to show that $M_\infty = \mathbb{E}(X \mid \mathcal{F}_\infty)$ a.s. Suppose that $A \in \mathcal{F}_n$. Then by Theorem 2.2.13 and the definition of M_n,

$$\int_A M_\infty \, d\mathbb{P} = \int_A M_n \, d\mathbb{P} = \int_A X \, d\mathbb{P}.$$

By a standard monotone class argument, it follows that this holds for all $A \in \mathcal{F}_\infty$. □

We also need the corresponding result for decreasing families of σ-algebras. If we have a filtration of the form $(\mathcal{F}_n : n \in -\mathbb{N})$, i.e. a collection of σ-algebras such that $\mathcal{F}_{-(n+1)} \subseteq \mathcal{F}_{-n}$ for all n, then we define

$$\mathcal{F}_{-\infty} = \bigcap_n \mathcal{F}_{-n}.$$

Theorem 2.2.15 (Lévy-Doob downward theorem). *Let $(\mathcal{F}_n : n \in -\mathbb{N})$ be a collection of σ-algebras such that $\mathcal{F}_{-(n+1)} \subseteq \mathcal{F}_{-n}$ for every $n \in \mathbb{N}$ and let $M = (M_n : n \in -\mathbb{N})$ be supermartingale, i.e.*

$$\mathbb{E}(M_n \mid \mathcal{F}_m) \leq M_m \quad \text{a.s.}$$

for all $m \leq n \leq -1$. *Then if* $\sup \mathbb{E}M_n < \infty$, *the process* M *is uniformly integrable, the limit*

$$M_{-\infty} = \lim_{n\to-\infty} M_n$$

exists a.s. and in L^1, *and*

$$\mathbb{E}(M_n \,|\, \mathcal{F}_{-\infty}) \leq M_{-\infty} \quad \text{a.s.},$$

with equality if M *is a martingale.*

Proof. For every $n \in \mathbb{N}$ the upcrossings inequality applies to the supermartingale $(M_k : k = -n, \ldots, -1)$. So by reasoning as in the proof of Theorem 2.2.12 we see that the limit $M_{-\infty} = \lim_{n\to-\infty} M_n$ exists and is finite almost surely. Now for all $K > 0$ and $n \in -\mathbb{N}$ we have

$$\int_{|M_n|>K} |M_n| \, d\mathbb{P} = -\int_{M_n<-K} M_n \, d\mathbb{P} - \int_{M_n \leq K} M_n \, d\mathbb{P} + \mathbb{E}M_n.$$

As n tends to $-\infty$, the expectation $\mathbb{E}M_n$ increases to a finite limit. So for an arbitrary $\varepsilon > 0$, there exists an $m \in -\mathbb{N}$ such that $\mathbb{E}M_n - \mathbb{E}M_m \leq \varepsilon$ for all $n \leq m$. Together with the supermartingale property this implies that for all $n \leq m$,

$$\begin{aligned}\int_{|M_n|>K} |M_n| \, d\mathbb{P} &\leq -\int_{M_n<-K} M_m \, d\mathbb{P} - \int_{M_n \leq K} M_m \, d\mathbb{P} + \mathbb{E}M_m + \varepsilon \\ &= \int_{|M_n|>K} |M_m| \, d\mathbb{P} + \varepsilon.\end{aligned}$$

Hence, to prove the uniform integrability of M it suffices (in view of Lemma A.3.1) to show that by choosing K large enough, we can make $\mathbb{P}(M_n > K)$ arbitrarily small for all n simultaneously. Now consider the process $M^- = \max\{-M, 0\}$. It is an increasing, convex function of the submartingale $-M$, whence it is a submartingale itself (see Example 2.1.3). In particular, $\mathbb{E}M_n^- \leq \mathbb{E}M_{-1}^-$ for all $n \in -\mathbb{N}$. It follows that

$$\mathbb{E}|M_n| = \mathbb{E}M_n + 2\mathbb{E}M_n^- \leq \sup \mathbb{E}M_n + 2\mathbb{E}|M_{-1}|$$

and, consequently,

$$\mathbb{P}(|M_n| > K) \leq \frac{1}{K}\left(\sup \mathbb{E}M_n + 2\mathbb{E}|M_{-1}|\right).$$

So indeed, M is uniformly integrable. The limit $M_{-\infty}$ therefore exists in L^1 as well and the proof can be completed by reasoning as in the proof of the upward theorem. □

Note that the downward theorem includes the 'downward version' of Theorem 2.2.14 as a special case. Indeed, if X is an integrable random variable and $\mathcal{F}_1 \supseteq \mathcal{F}_2 \supseteq \cdots \supseteq \bigcap_n \mathcal{F}_n = \mathcal{F}_\infty$ is a decreasing sequence of σ-algebras, then as $n \to \infty$,

$$\mathbb{E}(X \,|\, \mathcal{F}_n) \to \mathbb{E}(X \,|\, \mathcal{F}_\infty)$$

almost surely and in L^1. This is generalized in the following corollary of Theorems 2.2.14 and 2.2.15, which will be useful in the sequel.

Corollary 2.2.16. *Suppose that $X_n \to X$ a.s. and that $|X_n| \leq Y$ for all n, where Y is an integrable random variable. Moreover, suppose that $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots$ (resp. $\mathcal{F}_1 \supseteq \mathcal{F}_2 \supseteq \cdots$) is an increasing (resp. decreasing) sequence of σ-algebras. Then $\mathbb{E}(X_n \,|\, \mathcal{F}_n) \to \mathbb{E}(X \,|\, \mathcal{F}_\infty)$ a.s. as $n \to \infty$, where $\mathcal{F}_\infty = \sigma\left(\bigcup_n \mathcal{F}_n\right)$ (resp. $\mathcal{F}_\infty = \bigcap_n \mathcal{F}_n$).*

Proof. For $m \in \mathbb{N}$, put $U_m = \inf_{n\geq m} X_n$ and $V_m = \sup_{n\geq m} X_n$. Then since $X_n \to X$ a.s., we a.s. have $V_m - U_m \to 0$ as $m \to \infty$. It holds that $|V_m - U_m| \leq 2|Y|$, so dominated convergence implies that $\mathbb{E}(V_m - U_m) \to 0$ as $m \to \infty$. Now fix an arbitrary $\varepsilon > 0$ and choose m so large that $\mathbb{E}(V_m - U_m) \leq \varepsilon$. For $n \geq m$ we have

$$U_m \leq X_n \leq V_m, \tag{2.5}$$

hence $\mathbb{E}(U_m \,|\, \mathcal{F}_n) \leq \mathbb{E}(X_n \,|\, \mathcal{F}_n) \leq \mathbb{E}(V_m \,|\, \mathcal{F}_n)$ a.s. The processes on the left-hand side and the right-hand side are martingales which satisfy the conditions of the upward (resp. downward) theorem, so letting n tend to infinity we obtain, a.s.,

$$\mathbb{E}(U_m \,|\, \mathcal{F}_\infty) \leq \liminf \mathbb{E}(X_n \,|\, \mathcal{F}_n) \leq \limsup \mathbb{E}(X_n \,|\, \mathcal{F}_n) \leq \mathbb{E}(V_m \,|\, \mathcal{F}_\infty). \tag{2.6}$$

It follows that

$$\begin{aligned} \mathbb{E}\big(\limsup & \mathbb{E}(X_n \,|\, \mathcal{F}_n) - \liminf \mathbb{E}(X_n \,|\, \mathcal{F}_n)\big) \\ &\leq \mathbb{E}\big(\mathbb{E}(V_m \,|\, \mathcal{F}_\infty) - \mathbb{E}(U_m \,|\, \mathcal{F}_\infty)\big) \\ &= \mathbb{E}(V_m - U_m) \leq \varepsilon. \end{aligned}$$

By letting $\varepsilon \downarrow 0$ we see that $\limsup \mathbb{E}(X_n \,|\, \mathcal{F}_n) = \liminf \mathbb{E}(X_n \,|\, \mathcal{F}_n)$ a.s., so $\mathbb{E}(X_n \,|\, \mathcal{F}_n)$ converges a.s.

To identify the limit we let $n \to \infty$ in (2.5), finding that $U_m \leq X \leq V_m$, hence

$$\mathbb{E}(U_m \,|\, \mathcal{F}_\infty) \leq \mathbb{E}(X \,|\, \mathcal{F}_\infty) \leq \mathbb{E}(V_m \,|\, \mathcal{F}_\infty) \tag{2.7}$$

a.s. From (2.6) and (2.7) we see that both $\lim \mathbb{E}(X_n \,|\, \mathcal{F}_n)$ and $\mathbb{E}(X \,|\, \mathcal{F}_\infty)$ are a.s. between the random variables $\mathbb{E}(U_m \,|\, \mathcal{F}_\infty)$ and $\mathbb{E}(V_m \,|\, \mathcal{F}_\infty)$. It follows that

$$\mathbb{E}\big|\lim \mathbb{E}(X_n \,|\, \mathcal{F}_n) - \mathbb{E}(X \,|\, \mathcal{F}_\infty)\big| \leq \mathbb{E}(V_m - U_m) \leq \varepsilon.$$

By letting $\varepsilon \downarrow 0$ we see that $\lim \mathbb{E}(X_n \,|\, \mathcal{F}_n) = \mathbb{E}(X \,|\, \mathcal{F}_\infty)$ a.s. □

2.2.5 Optional stopping theorems

Theorem 2.2.5 implies that if M is a martingale and $\sigma \leq \tau$ are two *bounded* stopping times, then $\mathbb{E}(M_\tau \,|\, \mathcal{F}_\sigma) = M_\sigma$ a.s. The following theorem extends this result.

Theorem 2.2.17 (Optional stopping theorem). *Let M be a uniformly integrable martingale. Then the family of random variables $\{M_\tau :$ τ is a finite stopping time$\}$ is uniformly integrable and for all stopping times $\sigma \leq \tau$ we have*

$$\mathbb{E}(M_\tau \,|\, \mathcal{F}_\sigma) = M_\sigma$$

almost surely.

Proof. By Theorem 2.2.13, $M_\infty = \lim M_n$ exist almost surely and in L^1 and $\mathbb{E}(M_\infty \,|\, \mathcal{F}_n) = M_n$ a.s. Now let τ be an arbitrary stopping time and $n \in \mathbb{N}$. Since $\tau \wedge n \leq n$ we have $\mathcal{F}_{\tau\wedge n} \subseteq \mathcal{F}_n$. By the tower property of conditional expectations (see Appendix A), it follows that for every $n \in \mathbb{N}$

$$\mathbb{E}(M_\infty \,|\, \mathcal{F}_{\tau\wedge n}) = \mathbb{E}\left(\mathbb{E}(M_\infty \,|\, \mathcal{F}_n) \,|\, \mathcal{F}_{\tau\wedge n}\right) = \mathbb{E}(M_n \,|\, \mathcal{F}_{\tau\wedge n}),$$

a.s. Hence, by Theorem 2.2.5, we almost surely have

$$\mathbb{E}(M_\infty \,|\, \mathcal{F}_{\tau\wedge n}) = M_{\tau\wedge n}.$$

Now let n tend to infinity. Then the right-hand side converges almost surely to M_τ, and by the upward convergence theorem, the left-hand side converges a.s. (and in L^1) to $\mathbb{E}(M_\infty \,|\, \mathcal{G})$, where

$$\mathcal{G} = \sigma\left(\bigcup_n \mathcal{F}_{\tau\wedge n}\right),$$

so

$$\mathbb{E}(M_\infty \,|\, \mathcal{G}) = M_\tau \tag{2.8}$$

almost surely. Now take $A \in \mathcal{F}_\tau$. Then

$$\int_A M_\infty \, d\mathbb{P} = \int_{A\cap\{\tau<\infty\}} M_\infty \, d\mathbb{P} + \int_{A\cap\{\tau=\infty\}} M_\infty \, d\mathbb{P}.$$

Since $A \cap \{\tau < \infty\} \in \mathcal{G}$ (see Exercise 6), relation (2.8) implies that

$$\int_{A\cap\{\tau<\infty\}} M_\infty \, d\mathbb{P} = \int_{A\cap\{\tau<\infty\}} M_\tau \, d\mathbb{P}.$$

Trivially, we also have

$$\int_{A\cap\{\tau=\infty\}} M_\infty \, d\mathbb{P} = \int_{A\cap\{\tau=\infty\}} M_\tau \, d\mathbb{P}.$$

Hence, we find that

$$\int_A M_\infty \, d\mathbb{P} = \int_A M_\tau \, d\mathbb{P}$$

for every $A \in \mathcal{F}_\tau$. We conclude that

$$\mathbb{E}(M_\infty \,|\, \mathcal{F}_\tau) = M_\tau$$

almost surely. The first statement of the theorem now follows from Lemma A.3.4 in Appendix A. The second statement follows from the tower property of conditional expectations and the fact that $\mathcal{F}_\sigma \subseteq \mathcal{F}_\tau$ if $\sigma \leq \tau$. □

For the equality $\mathbb{E}(M_\tau \mid \mathcal{F}_\sigma) = M_\sigma$ in the preceding theorem to hold it is necessary that M is uniformly integrable. There exist (positive) martingales that are bounded in L^1 but not uniformly integrable, for which the equality fails in general (see Exercise 12)! For positive supermartingales without additional integrability properties we only have an inequality.

Theorem 2.2.18. *Let M be a nonnegative supermartingale and let $\sigma \le \tau$ be stopping times. Then*

$$\mathbb{E}(M_\tau 1_{\{\tau<\infty\}} \mid \mathcal{F}_\sigma) \le M_\sigma 1_{\{\sigma<\infty\}}$$

almost surely.

Proof. Fix $n \in \mathbb{N}$. The stopped supermartingale $M^{\tau\wedge n}$ is a supermartingale again (cf. Theorem 2.2.4) and is uniformly integrable (check!). By reasoning exactly as in the proof of the preceding theorem we find that

$$\mathbb{E}(M_{\tau\wedge n} \mid \mathcal{F}_\sigma) = \mathbb{E}(M_\infty^{\tau\wedge n} \mid \mathcal{F}_\sigma) \le M_\sigma^{\tau\wedge n} = M_{\sigma\wedge n},$$

a.s. Hence, by the conditional version of Fatou's lemma (see Appendix A),

$$\begin{aligned}\mathbb{E}(M_\tau 1_{\{\tau<\infty\}} \mid \mathcal{F}_\sigma) &\le \mathbb{E}(\liminf M_{\tau\wedge n} \mid \mathcal{F}_\sigma)\\ &\le \liminf \mathbb{E}(M_{\tau\wedge n} \mid \mathcal{F}_\sigma)\\ &\le \liminf M_{\sigma\wedge n}\end{aligned}$$

almost surely. On the event $\{\sigma < \infty\}$ the lim inf in the last line equals M_σ. □

2.3 Continuous-time martingales

In this section we consider general martingales, indexed by a subset T of $\mathbb{R}_+$. If the martingale $M = (M_t)_{t\ge0}$ has 'nice' sample paths, for instance right-continuous, then M can be approximated 'accurately' by a discrete-time martingale. Simply choose a countable, dense subset $\{t_n\}$ of the index set T and compare the continuous-time martingale M with the discrete-time martingale $(M_{t_n})_n$. This simple idea allows us to transfer many of the discrete-time results to the continuous-time setting.

2.3.1 Upcrossings in continuous time

For a continuous-time process X we define the number of upcrossings of the interval $[a, b]$ in the set of time points $T \subseteq \mathbb{R}_+$ as follows. For a finite set

$F = \{t_1, \ldots, t_n\} \subseteq T$ we define $U_F[a,b]$ as the number of upcrossings of $[a,b]$ of the discrete-time process $(X_{t_i})_{i=1..n}$ (see Definition 2.2.10). We put

$$U_T[a,b] = \sup\{U_F[a,b] : F \subseteq T, F \text{ finite}\}.$$

Doob's upcrossings lemma has the following extension.

Lemma 2.3.1. *Let M be a supermartingale and let $T \subseteq \mathbb{R}_+$ be countable. Then for all $a < b$, the number of upcrossings $U_T[a,b]$ of the interval $[a,b]$ by M satisfies*

$$(b-a)\mathbb{E}U_T[a,b] \leq \sup_{t \in T} \mathbb{E}(M_t - a)^-.$$

Proof. Let T_n be a nested sequence of finite sets such that $U_T[a,b] = \lim U_{T_n}[a,b]$. For every n, the discrete-time upcrossings inequality states that

$$(b-a)\mathbb{E}U_{T_n}[a,b] \leq \mathbb{E}(M_{t_n} - a)^-,$$

where t_n is the largest element of T_n. By the conditional version of Jensen's inequality the process $(M-a)^-$ is a submartingale (see Example 2.1.3). In particular the function $t \mapsto \mathbb{E}(M_t - a)^-$ is increasing, so

$$\mathbb{E}(M_{t_n} - a)^- = \sup_{t \in T_n} \mathbb{E}(M_t - a)^-.$$

Hence, for every n we have the inequality

$$(b-a)\mathbb{E}U_{T_n}[a,b] \leq \sup_{t \in T_n} \mathbb{E}(M_t - a)^-.$$

The proof is completed by letting n tend to infinity. □

2.3.2 Regularization

Theorem 2.3.2. *Let M be a supermartingale. Then for almost all ω, the limits*

$$\lim_{\substack{q \uparrow t \\ q \in \mathbb{Q}}} M_q(\omega) \quad \textit{and} \quad \lim_{\substack{q \downarrow t \\ q \in \mathbb{Q}}} M_q(\omega)$$

exist and are finite for every $t \geq 0$.

Proof. Fix $n \in \mathbb{N}$. By Lemma 2.3.1 there exists an event Ω_n of probability 1 on which the upcrossing numbers of the process M satisfy

$$U_{[0,n] \cap \mathbb{Q}}[a,b] < \infty$$

for every pair of rational numbers $a < b$. Now suppose that for $t \leq n$, the limit

$$\lim_{\substack{q\uparrow t\\ q\in\mathbb{Q}}} M_q$$

does not exist. Then there exist two rational numbers $a < b$ such that

$$\liminf_{\substack{q\uparrow t\\ q\in\mathbb{Q}}} M_q < a < b < \limsup_{\substack{q\uparrow t\\ q\in\mathbb{Q}}} M_q.$$

But this implies that $U_{[0,n]\cap\mathbb{Q}}[a, b] = \infty$. We conclude that on the event Ω_n the limit

$$\lim_{\substack{q\uparrow t\\ r\in\mathbb{Q}}} M_q$$

exists for every $t \leq n$. Similarly, the other limit exists for every $t \leq n$. It follows that on the event $\Omega' = \bigcap \Omega_n$, both limits exist in $[-\infty, \infty]$ for every $t \geq 0$.

To prove that the left limit is in fact finite, let Q be the set of all rational numbers less than t and let Q_n be nested finite sets of rational numbers increasing to Q. For fixed n, the process $(M_q)_{q\in\{0\}\cup Q_n\cup\{t\}}$ is a discrete-time supermartingale. Hence, by Theorem 2.2.9,

$$\lambda\mathbb{P}(\max_{q\in Q_n} |M_q| \geq 3\lambda) \leq 4\mathbb{E}|M_0| + 3\mathbb{E}|M_t|.$$

Letting $n \to \infty$ and then $\lambda \to \infty$, we see that $\sup_{q\in Q} |M_q| < \infty$ almost surely, and hence the left limit is finite as well. The finiteness of the right limit is proved similarly. □

Corollary 2.3.3. *Almost every sample path of a right-continuous supermartingale is cadlag.*

Proof. See Exercise 7. □

Given a supermartingale M we now define for every $t \geq 0$

$$M_{t+} = \lim_{\substack{q\downarrow t\\ q\in\mathbb{Q}}} M_q.$$

These random variables are well defined by Theorem 2.3.2 and we have the following result regarding the process $(M_{t+})_{t\geq 0}$.

Lemma 2.3.4. *Let M be a supermartingale. Then $\mathbb{E}|M_{t+}| < \infty$ for every t and*

$$\mathbb{E}(M_{t+} \mid \mathcal{F}_t) \leq M_t \quad \text{a.s.}$$

If $t \mapsto \mathbb{E}M_t$ is continuous, this inequality is an equality. Moreover, the process $(M_{t+})_{t\geq 0}$ is a supermartingale with respect to the filtration $(\mathcal{F}_{t+})_{t\geq 0}$ and it is a martingale if M is a martingale.

Proof. Let t_n be a sequence of rational numbers decreasing to t. Then the process M_{t_n} is a 'backward supermartingale' like we considered in Theorem 2.2.15. Hence, M_{t+} is integrable and M_{t_n} converges to M_{t+} in L^1. It follows that in the inequality

$$\mathbb{E}(M_{t_n} \mid \mathcal{F}_t) \leq M_t \quad \text{a.s.},$$

we may let n tend to infinity to find that

$$\mathbb{E}(M_{t_+} \mid \mathcal{F}_t) \leq M_t \quad \text{a.s.}.$$

The L^1 convergence also implies that $\mathbb{E}M_{t_n} \to \mathbb{E}M_{t+}$. So if $t \mapsto \mathbb{E}M_t$ is continuous, we have $\mathbb{E}M_{t+} = \lim \mathbb{E}M_{t_n} = \mathbb{E}M_t$, hence

$$\mathbb{E}(M_{t_+} \mid \mathcal{F}_t) = M_t \quad \text{a.s.}$$

(see Exercise 8).

To prove the final statement, take $s < t$ and let s_n be a sequence of rational numbers smaller than t and decreasing to s. Then

$$\mathbb{E}(M_{t+} \mid \mathcal{F}_{s_n}) = \mathbb{E}(\mathbb{E}(M_{t+} \mid \mathcal{F}_t) \mid \mathcal{F}_{s_n}) \leq \mathbb{E}(M_t \mid \mathcal{F}_{s_n}) \leq M_{s_n}$$

almost surely, with equality if M is a martingale. The right-hand side of this display converges to M_{s+} as $n \to \infty$. The process $(\mathbb{E}(M_{t+} \mid \mathcal{F}_{s_n}))_n$ is a 'backward supermartingale', so by Theorem 2.2.15, the left-hand side converges to $\mathbb{E}(M_{t+} \mid \mathcal{F}_{s+})$ almost surely. □

We can now prove the main regularization theorem for supermartingales with respect to filtrations satisfying the usual conditions (see Definition 1.6.3).

Theorem 2.3.5. *Let M be a supermartingale with respect to a filtration $(\mathcal{F}_t)$ that satisfies the usual conditions. Then if $t \mapsto \mathbb{E}M_t$ is right-continuous, the process M has a cadlag modification.*

Proof. By Theorem 2.3.2, there exists an event Ω' of probability 1 on which the limits

$$M_{t-} = \lim_{\substack{q \uparrow t \\ q \in \mathbb{Q}}} M_q \quad \text{and} \quad M_{t+} = \lim_{\substack{q \downarrow t \\ q \in \mathbb{Q}}} M_q$$

exist for every t. We define the process $\tilde{M}$ by $\tilde{M}_t(\omega) = M_{t+}(\omega)1_{\Omega'}(\omega)$ for $t \geq 0$. Then $\tilde{M}_t = M_{t+}$ a.s. and since M_{t+} is $\mathcal{F}_{t+}$-measurable, we have $\tilde{M}_t = \mathbb{E}(M_{t+} \mid \mathcal{F}_{t+})$ a.s. Hence, by the right-continuity of the filtration and the preceding lemma, $\tilde{M}_t = \mathbb{E}(M_{t+} \mid \mathcal{F}_t) \leq M_t$ a.s. Since $t \mapsto \mathbb{E}M_t$ is right-continuous, it holds that

$$\mathbb{E}\tilde{M}_t = \lim_{s \downarrow t} \mathbb{E}M_s = \mathbb{E}M_t.$$

Together with the fact that $\tilde{M}_t \leq M_t$ a.s. this implies that $\tilde{M}_t = M_t$ a.s., in other words, $\tilde{M}$ is a modification of M (see Exercise 8). The usual conditions on the filtration imply that $\tilde{M}$ is adapted, and the process is cadlag by construction (see Exercise 9). By the preceding lemma and the right-continuity of the filtration it is a supermartingale. □

Corollary 2.3.6. *A martingale with respect to a filtration that satisfies the usual conditions has a cadlag modification.*

2.3.3 Convergence theorems

In view of the results of the preceding section, we will only consider right-continuous martingales from this point on. Under this assumption, many of the discrete-time theorems can be generalized to continuous time.

Theorem 2.3.7 (Martingale convergence theorem). *Let M be a right-continuous supermartingale that is bounded in L^1. Then M_t converges almost surely to a limit M_∞ as $t \to \infty$, and $\mathbb{E}|M_\infty| < \infty$.*

Proof. First we show that since M is right-continuous, it holds that $M_t \to M_\infty$ as $t \to \infty$ if and only if

$$\lim_{\substack{t\to\infty \\ t\in\mathbb{Q}_+}} M_t = M_\infty. \tag{2.9}$$

To prove the non-trivial implication in this assertion, assume that (2.9) holds and fix an $\varepsilon > 0$. Then there exists a number $a > 0$ such that for all $t \in [a, \infty) \cap \mathbb{Q}$, it holds that $|M_t - M_\infty| < \varepsilon$. Now let $t > a$ be arbitrary. Since M is right-continuous, there exists an $s > t > a$ such that $s \in \mathbb{Q}$ and $|M_s - M_t| < \varepsilon$. By the triangle inequality, it follows that $|M_t - M_\infty| \le |M_t - M_s| + |M_s - M_\infty| \le 2\varepsilon$, which proves the first assertion.

So to prove the convergence, we may assume that M is indexed by the countable set $\mathbb{Q}_+$. The proof can now be finished by arguing as in the proof of Theorem 2.2.12, replacing Doob's discrete-time upcrossings inequality by Lemma 2.3.1. □

Corollary 2.3.8. *A nonnegative, right-continuous supermartingale M converges almost surely as $t \to \infty$.*

Proof. Simple exercise. □

The following continuous-time extension of Theorem 2.2.13 can be derived by reasoning exactly as in the discrete-time case.

Theorem 2.3.9. *Let M be a right-continuous supermartingale that is bounded in L^1. Then $M_t \to M_\infty$ in L^1 as $t \to \infty$ if and only if M is uniformly integrable. In that case*

$$\mathbb{E}(M_\infty \,|\, \mathcal{F}_t) \le M_t \quad \text{a.s.},$$

with equality if M is a martingale.

2.3.4 Inequalities

Doob's submartingale inequality and L^p-inequality are easily extended to the setting of general right-continuous martingales.

Theorem 2.3.10 (Doob's submartingale inequality). *Let M be a right-continuous submartingale. Then for all $\lambda > 0$ and $t \geq 0$,*

$$\mathbb{P}\left(\sup_{s\leq t} M_s \geq \lambda\right) \leq \frac{1}{\lambda}\mathbb{E}|M_t|.$$

Proof. Let T be a countable, dense subset of $[0, t]$ and choose a sequence of finite subsets $T_n \subseteq T$ such that $t \in T_n$ for every n and $T_n \uparrow T$ as $n \to \infty$. Then by the discrete-time version of the submartingale inequality

$$\mathbb{P}\left(\max_{s\in T_n} M_s \geq \lambda\right) \leq \frac{1}{\lambda}\mathbb{E}|M_t|.$$

Now let n tend to infinity. Then the probability on the left-hand side increases to

$$\mathbb{P}\left(\sup_{s\in T} M_s \geq \lambda\right).$$

Since M is right-continuous, the supremum over T equals the supremum over $[0, t]$. $\square$

By exactly the same reasoning we can generalize the L^p-inequality to continuous time.

Theorem 2.3.11 (Doob's L^p-inequality). *Let M be a right-continuous martingale or a right-continuous, nonnegative submartingale. Then for all $p > 1$ and $t \geq 0$*

$$\mathbb{E}\left(\sup_{s\leq t} |M_s|^p\right) \leq \left(\frac{p}{p-1}\right)^p \mathbb{E}\,|M_t|^p.$$

2.3.5 Optional stopping

We now come to the continuous-time version of the optional stopping theorem.

Theorem 2.3.12 (Optional stopping theorem). *Let M be a right-continuous, uniformly integrable martingale. Then for all stopping times $\sigma \leq \tau$ we have*

$$\mathbb{E}(M_\tau \mid \mathcal{F}_\sigma) = M_\sigma$$

almost surely.

Proof. By Lemma 1.6.14 there exist stopping times σ_n and τ_n that only take finitely many values and such that $\tau_n \downarrow \tau$ and $\sigma_n \downarrow \sigma$ almost surely. Moreover, it is easily seen that for the stopping times σ_n and τ_n constructed as in the proof of Lemma 1.6.14, it holds that $\sigma_n \leq \tau_n$ for every n. Hence, by the discrete-time optional sampling theorem,

$$\mathbb{E}(M_{\tau_n} \,|\, \mathcal{F}_{\sigma_n}) = M_{\sigma_n}$$

almost surely. Since $\sigma \leq \sigma_n$ it holds that $\mathcal{F}_\sigma \subseteq \mathcal{F}_{\sigma_n}$. It follows that a.s.

$$\mathbb{E}(M_{\tau_n} \,|\, \mathcal{F}_\sigma) = \mathbb{E}(\mathbb{E}(M_{\tau_n} \,|\, \mathcal{F}_{\sigma_n}) \,|\, \mathcal{F}_\sigma) = \mathbb{E}(M_{\sigma_n} \,|\, \mathcal{F}_\sigma). \tag{2.10}$$

The discrete-time optional sampling theorem also implies that

$$\mathbb{E}(M_{\tau_n} \,|\, \mathcal{F}_{\tau_{n+1}}) = M_{\tau_{n+1}}$$

almost surely. Hence, M_{τ_n} is a 'backward' martingale in the sense of Theorem 2.2.15. Since $\sup \mathbb{E}M_{\tau_n} = \mathbb{E}M_0 < \infty$, Theorem 2.2.15 implies that M_{τ_n} is uniformly integrable. Together with the fact that $M_{\tau_n} \to M_\tau$ a.s. by the right-continuity of M_τ, this implies that $M_{\tau_n} \to M_\tau$ in L^1. Similarly, we have that $M_{\sigma_n} \to M_\sigma$ in L^1. Now take $A \in \mathcal{F}_\sigma$. Then by equation (2.10) it holds that

$$\int_A M_{\tau_n}\, d\mathbb{P} = \int_A M_{\sigma_n}\, d\mathbb{P}.$$

By the L^1-convergence we just proved, this yields the relation

$$\int_A M_\tau\, d\mathbb{P} = \int_A M_\sigma\, d\mathbb{P}$$

if we let n tend to infinity. This completes the proof. $\square$

Corollary 2.3.13. *A right-continuous, adapted process M is a martingale if and only if for every bounded stopping time τ, the random variable M_τ is integrable and $\mathbb{E}M_\tau = \mathbb{E}M_0$ almost surely.*

Proof. Suppose first that M is a martingale. Since τ is bounded, there exists a constant $K > 0$ such that $\tau \leq K$. The process $(M_t)_{t \leq K}$ is uniformly integrable. Hence, by the optional sampling theorem, $\mathbb{E}M_\tau = \mathbb{E}M_0$. The converse statement is proved by arguing as in the proof of Theorem 2.2.5. $\square$

Corollary 2.3.14. *If M is a right-continuous martingale and τ is a stopping time, then the stopped process M^τ is a martingale as well.*

Proof. Let σ be a bounded stopping time. Then by applying the preceding corollary to the martingale M we find

$$\mathbb{E}M^\tau_\sigma = \mathbb{E}M_{\tau\wedge\sigma} = \mathbb{E}M_0 = \mathbb{E}M^\tau_0.$$

Since M^τ is right-continuous and adapted (see Lemmas 1.6.11 and 1.6.13), another application of the preceding corollary completes the proof. $\square$

Just as in discrete time the assumption of uniform integrability is crucial for the optional stopping theorem. If this condition is dropped we only have an inequality in general. Theorem 2.2.18 carries over to continuous time by using the same arguments as in the proof of Theorem 2.3.12.

Theorem 2.3.15. *Let M be a nonnegative, right-continuous supermartingale and let $\sigma \leq \tau$ be stopping times. Then*

$$\mathbb{E}(M_\tau 1_{\{\tau<\infty\}} \,|\, \mathcal{F}_\sigma) \leq M_\sigma 1_{\{\sigma<\infty\}}$$

almost surely.

A consequence of this result is that nonnegative, right-continuous supermartingales stay at zero once they have hit it.

Corollary 2.3.16. *Let M be a nonnegative, right-continuous supermartingale and define $\tau = \inf\{t : M_t = 0 \text{ or } M_{t-} = 0\}$. Then almost surely, the process M vanishes on $[\tau, \infty)$.*

Proof. Define $\tau_n = \inf\{t : M_t \leq 1/n\}$. Then for a rational number $q > 0$ we have $\tau_n \leq \tau + q$, so that by the preceding theorem,

$$\mathbb{E} M_{\tau+q} 1_{\{\tau<\infty\}} \leq \mathbb{E} M_{\tau_n} 1_{\{\tau_n<\infty\}} \leq \frac{1}{n}.$$

It follows that almost surely on the event $\{\tau < \infty\}$, it holds that $M_{\tau+q} = 0$ for every rational $q > 0$. Since M is right-continuous, it holds for all $q \geq 0$. □

2.4 Applications to Brownian motion

In this section we apply the developed martingale theory to the study of the Brownian motion.

2.4.1 Quadratic variation

The following theorem extends the result of Exercise 13 of Chapter 1.

Theorem 2.4.1. *Let W be a Brownian motion and fix $t > 0$. For $n \in \mathbb{N}$, let π_n be a partition of $[0, t]$ given by $0 = t_0^n \leq t_1^n \leq \cdots \leq t_{k_n}^n = t$ and suppose that the mesh $\|\pi_n\| = \max_k |t_k^n - t_{k-1}^n|$ tends to zero as $n \to \infty$. Then*

$$\sum_k (W_{t_k^n} - W_{t_{k-1}^n})^2 \xrightarrow{L^2} t$$

as $n \to \infty$. If the partitions are nested we have

$$\sum_k (W_{t_k^n} - W_{t_{k-1}^n})^2 \xrightarrow{\text{as}} t$$

as $n \to \infty$.

Proof. For the first statement, see Exercise 13 of Chapter 1. To prove the second one, denote the sum by X_n and put $\mathcal{F}_n = \sigma(X_n, X_{n+1}, \ldots)$. Then $\mathcal{F}_{n+1} \subseteq \mathcal{F}_n$ for every $n \in \mathbb{N}$. Now suppose that we can show that $\mathbb{E}(X_n \,|\, \mathcal{F}_{n+1}) = X_{n+1}$ a.s. Then, since $\sup \mathbb{E}X_n = t < \infty$, Theorem 2.2.15 implies that X_n converges almost surely to a finite limit X_∞. By the first statement of the theorem the X_n converge in probability to t. Hence, we must have $X_\infty = t$ a.s.

So it remains to prove that $\mathbb{E}(X_n \,|\, \mathcal{F}_{n+1}) = X_{n+1}$ a.s. Without loss of generality, we assume that the number elements of the partition π_n equals n. In that case, there exists a sequence t_n such that the partition π_n has the numbers $t_1, \ldots, t_n$ as its division points: the point t_n is added to π_{n-1} to form the next partition π_n. Now fix n and consider the process W' defined by

$$W'_s = W_{s \wedge t_{n+1}} - (W_s - W_{s \wedge t_{n+1}}).$$

By Exercise 10 of Chapter 1, W' is again a BM. For W', denote the analogues of the sums X_k by X'_k. Then it is easily seen that for $k \geq n+1$ it holds that $X'_k = X_k$. Moreover, it holds that $X'_n - X'_{n+1} = X_{n+1} - X_n$ (check!). Since both W and W' are BM's, the sequences $(X_1, X_2, \ldots)$ and $(X'_1, X'_2, \ldots)$ have the same distribution. It follows that a.s.,

$$\begin{aligned}
\mathbb{E}(X_n - X_{n+1} \,|\, \mathcal{F}_{n+1}) &= \mathbb{E}(X'_n - X'_{n+1} \,|\, X'_{n+1}, X'_{n+2}, \ldots) \\
&= \mathbb{E}(X'_n - X'_{n+1} \,|\, X_{n+1}, X_{n+2}, \ldots) \\
&= \mathbb{E}(X_{n+1} - X_n \,|\, X_{n+1}, X_{n+2}, \ldots) \\
&= -\mathbb{E}(X_n - X_{n+1} \,|\, \mathcal{F}_{n+1}).
\end{aligned}$$

This implies that $\mathbb{E}(X_n - X_{n+1} \,|\, \mathcal{F}_{n+1}) = 0$ almost surely. □

A real-valued function f is said to be of *finite variation* on an interval $[a, b]$ if there exists a finite number $K > 0$ such that for every finite partition $a = t_0 < \cdots < t_n = b$ of $[a, b]$ it holds that

$$\sum_k |f(t_k) - f(t_{k-1})| < K.$$

Roughly speaking, this means that the graph of the function f on $[a, b]$ has finite length. Theorem 2.4.1 shows that the sample paths of the BM have positive and finite *quadratic* variation. This has the following consequence.

Corollary 2.4.2. *Almost every sample path of the BM has unbounded variation on every interval.*

Proof. Fix $t > 0$. Let π_n be nested partitions of $[0,t]$ given by $0 = t_0^n \leq t_1^n \leq \cdots \leq t_{k_n}^n = t$ and suppose that the mesh $\|\pi_n\| = \max_k |t_k^n - t_{k-1}^n|$ tends to zero as $n \to \infty$. Then

$$\sum_k (W_{t_k^n} - W_{t_{k-1}^n})^2 \leq \sup_{|s-t| \leq \|\pi_n\|} |W_s - W_t| \sum_k |W_{t_k^n} - W_{t_{k-1}^n}|.$$

By the continuity of the Brownian sample paths the first factor on the right-hand side converges to zero a.s. as $n \to \infty$. Hence, if the BM would have finite variation on $[0,t]$ with positive probability, then

$$\sum_k (W_{t_k^n} - W_{t_{k-1}^n})^2$$

would converge to zero with positive probability, which contradicts Theorem 2.4.1. □

2.4.2 Exponential inequality

Let W be a Brownian motion. We have the following exponential inequality for the tail probabilities of the running maximum of the Brownian motion.

Theorem 2.4.3. *For every $t \geq 0$ and $\lambda > 0$,*

$$\mathbb{P}\left(\sup_{s \leq t} W_s \geq \lambda\right) \leq e^{-\frac{1}{2}\frac{\lambda^2}{t}}$$

and

$$\mathbb{P}\left(\sup_{s \leq t} |W_s| \geq \lambda\right) \leq 2e^{-\frac{1}{2}\frac{\lambda^2}{t}}.$$

Proof. For $a > 0$ consider the exponential martingale M defined by $M_t = \exp(aW_t - a^2 t/2)$ (see Example 2.1.4). Observe that

$$\mathbb{P}\left(\sup_{s \leq t} W_s \geq \lambda\right) \leq \mathbb{P}\left(\sup_{s \leq t} M_s \geq e^{a\lambda - a^2 t/2}\right).$$

By the submartingale inequality, the probability on the right-hand side is bounded by

$$e^{a^2 t/2 - a\lambda}\mathbb{E}M_t = e^{a^2 t/2 - a\lambda}\mathbb{E}M_0 = e^{a^2 t/2 - a\lambda}.$$

The proof of the first inequality is completed by minimizing the latter expression in $a > 0$. To prove the second one, note that

$$\begin{aligned}\mathbb{P}\left(\sup_{s \leq t} |W_s| \geq \lambda\right) &\leq \mathbb{P}\left(\sup_{s \leq t} W_s \geq \lambda\right) + \mathbb{P}\left(\inf_{s \leq t} W_s \leq -\lambda\right) \\ &= \mathbb{P}\left(\sup_{s \leq t} W_s \geq \lambda\right) + \mathbb{P}\left(\sup_{s \leq t} -W_s \geq \lambda\right).\end{aligned}$$

The proof is completed by applying the first inequality to the BM's W and $-W$. □

The exponential inequality also follows from the fact that $\sup_{s \leq t} W_s =_d |W_t|$ for every fixed t. We will prove this equality in distribution in the next chapter.

2.4.3 The law of the iterated logarithm

The law of the iterated logarithm describes how the BM oscillates near zero and infinity. In the proof we need the following simple lemma.

Lemma 2.4.4. *For every $a > 0$,*

$$\int_a^\infty e^{-\frac{1}{2}x^2}\,dx \geq \frac{a}{1+a^2} e^{-\frac{1}{2}a^2}.$$

Proof. The proof starts from the inequality

$$\int_a^\infty \frac{1}{x^2} e^{-\frac{1}{2}x^2}\,dx \leq \frac{1}{a^2} \int_a^\infty e^{-\frac{1}{2}x^2}\,dx.$$

Integration by parts shows that the left-hand side equals

$$\begin{aligned} -\int_a^\infty e^{-\frac{1}{2}x^2}\,d\left(\frac{1}{x}\right) &= \frac{1}{a} e^{-\frac{1}{2}a^2} + \int_a^\infty \frac{1}{x}\,d\left(e^{-\frac{1}{2}x^2}\right) \\ &= \frac{1}{a} e^{-\frac{1}{2}a^2} - \int_a^\infty e^{-\frac{1}{2}x^2}\,dx. \end{aligned}$$

Hence, we find that

$$\left(1 + \frac{1}{a^2}\right) \int_a^\infty e^{-\frac{1}{2}x^2}\,dx \geq \frac{1}{a} e^{-\frac{1}{2}a^2}$$

and the proof is complete. □

Theorem 2.4.5 (Law of the iterated logarithm). *It almost surely holds that*

$$\limsup_{t \downarrow 0} \frac{W_t}{\sqrt{2t \log\log 1/t}} = 1, \qquad \liminf_{t \downarrow 0} \frac{W_t}{\sqrt{2t \log\log 1/t}} = -1,$$

$$\limsup_{t \to \infty} \frac{W_t}{\sqrt{2t \log\log t}} = 1, \qquad \liminf_{t \to \infty} \frac{W_t}{\sqrt{2t \log\log t}} = -1.$$

Proof. It suffices to prove the first statement, since the second statement follows by applying the first one to the BM $-W$. The third and fourth follow by applying the first two to the BM $tW_{1/t}$ (see Theorem 1.4.4).

Put $h(t) = (2t \log\log 1/t)^{1/2}$ and choose two numbers $\theta, \delta \in (0,1)$. We put

$$\alpha_n = (1+\delta)\theta^{-n} h(\theta^n), \quad \beta_n = h(\theta^n)/2.$$

Using the submartingale inequality as in the proof of Theorem 2.4.3 we find that

$$\mathbb{P}\left(\sup_{s\le 1}(W_s - \alpha_n s/2) \ge \beta_n\right) \le e^{-\alpha_n \beta_n} \le K n^{-1-\delta}$$

for some constant $K > 0$ that does not depend on n. Hence, by the Borel-Cantelli lemma,

$$\sup_{s\le 1}(W_s - \alpha_n s/2) \le \beta_n$$

for all n large enough. In particular, it holds for all n large enough and $s \in [0, \theta^{n-1}]$ that

$$W_s \le \frac{\alpha_n s}{2} + \beta_n \le \frac{\alpha_n \theta^{n-1}}{2} + \beta_n = \left(\frac{1+\delta}{2\theta} + \frac{1}{2}\right) h(\theta^n).$$

Since h is increasing in a neighbourhood of 0, it follows that for all n large enough,

$$\sup_{\theta^n \le s \le \theta^{n-1}} \frac{W_s}{h(s)} \le \left(\frac{1+\delta}{2\theta} + \frac{1}{2}\right),$$

which implies that

$$\limsup_{t\downarrow 0} \frac{W_t}{h(t)} \le \frac{1+\delta}{2\theta} + \frac{1}{2}.$$

Now let $\theta \uparrow 1$ and $\delta \downarrow 0$ to find that

$$\limsup_{t\downarrow 0} \frac{W_t}{h(t)} \le 1.$$

To prove the reverse inequality, choose $\theta \in (0,1)$ and consider the events

$$A_n = \left\{W_{\theta^n} - W_{\theta^{n+1}} \ge \left(1 - \sqrt{\theta}\right) h(\theta^n)\right\}.$$

By the independence of the increments of the BM the events A_n are independent. Moreover, it holds that

$$\mathbb{P}(A_n) = \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-\frac{1}{2}x^2}\, dx.$$

with $a = (1-\sqrt{\theta})(2(1-\theta)^{-1} \log\log \theta^{-n})^{1/2}$. By Lemma (2.4.4), it follows that

$$\sqrt{2\pi}\mathbb{P}(A_n) \ge \frac{a}{1+a^2} e^{-\frac{1}{2}a^2}.$$

It is easily seen that the right-hand side is of order

$$n^{-\frac{(1-\sqrt{\theta})^2}{1-\theta}} = n^{-\alpha},$$

with $\alpha < 1$. It follows that $\sum \mathbb{P}(A_n) = \infty$, so by the Borel-Cantelli lemma,

$$W_{\theta^n} \ge \left(1 - \sqrt{\theta}\right) h(\theta^n) + W_{\theta^{n+1}}$$

for infinitely many n. Since $-W$ is also a BM, the first part of the proof implies that

$$-W_{\theta^{n+1}} \leq 2h(\theta^{n+1})$$

for all n large enough. Combined with the preceding inequality and the fact that $h(\theta^{n+1}) \leq 2\sqrt{\theta}h(\theta^n)$, this implies that

$$W_{\theta^n} \geq \left(1 - \sqrt{\theta}\right) h(\theta^n) - 2h(\theta^{n+1}) \geq h(\theta^n)(1 - 5\sqrt{\theta})$$

for infinitely many n. Hence,

$$\limsup_{t\downarrow 0} \frac{W_t}{h(t)} \geq 1 - 5\sqrt{\theta},$$

and the proof is completed by letting θ tend to zero. □

As a corollary we have the following result regarding the zero set of the BM that was considered in Exercise 26 of Chapter 1.

Corollary 2.4.6. *The point* 0 *is an accumulation point of the zero set of the BM, i.e. for every* $\varepsilon > 0$*, the BM visits* 0 *infinitely often in the time interval* $[0, \varepsilon)$.

Proof. By the law of the iterated logarithm, the exist sequences t_n and s_n converging to 0 such that

$$\frac{W_{t_n}}{\sqrt{2t_n \log\log 1/t_n}} = 1$$

and

$$\frac{W_{s_n}}{\sqrt{2s_n \log\log 1/s_n}} = -1$$

for every n. The corollary follows by the continuity of the Brownian sample paths. □

2.4.4 Distribution of hitting times

Let W be a standard BM and for $a > 0$, let τ_a be the (a.s. finite) hitting time of the level a, cf. Example 1.6.9.

Theorem 2.4.7. *For* $a > 0$*, the Laplace transform of the hitting time* τ_a *is given by*

$$\mathbb{E}e^{-\lambda\tau_a} = e^{-a\sqrt{2\lambda}}, \quad \lambda \geq 0.$$

Proof. For $b \geq 0$, consider the exponential martingale $M_t = \exp(bW_t - b^2t/2)$ (see Example 2.1.4). The stopped process M^{τ_a} is again a martingale (see Corollary 2.3.14) and is bounded by $\exp(ab)$. A bounded martingale is uniformly integrable, hence, by the optional stopping theorem,

$$\mathbb{E}M_{\tau_a} = \mathbb{E}M_\infty^{\tau_a} = \mathbb{E}M_0^{\tau_a} = 1.$$

Since $W_{\tau_a} = a$, it follows that

$$\mathbb{E}e^{ab - b^2\tau_a/2} = 1.$$

The expression for the Laplace transform follows by substituting $b^2 = 2\lambda$. □

We will see later that τ_a has the density

$$x \mapsto \frac{ae^{-\frac{a^2}{2x}}}{\sqrt{2\pi x^3}} 1_{\{x \geq 0\}}.$$

This can be shown for instance by inverting the Laplace transform of τ_a. We will however use an alternative method in the next chapter. At this point we only prove that although the hitting times τ_a are finite almost surely, we have $\mathbb{E}\tau_a = \infty$ for every $a > 0$. A process with this property is called *null recurrent*.

Corollary 2.4.8. *For every $a > 0$ it holds that $\mathbb{E}\tau_a = \infty$.*

Proof. Denote the distribution function of τ_a by F. By integration by parts we have for every $\lambda > 0$

$$\mathbb{E}e^{-\lambda\tau_a} = \int_0^\infty e^{-\lambda x}\, dF(x) = -\int_0^\infty F(x)\, de^{-\lambda x}$$

Combined with the fact that

$$-1 = \int_0^\infty de^{-\lambda x}$$

it follows that

$$\frac{1 - \mathbb{E}e^{-\lambda\tau_a}}{\lambda} = -\frac{1}{\lambda}\int_0^\infty (1 - F(x))\, de^{-\lambda x} = \int_0^\infty (1 - F(x))e^{-\lambda x}\, dx.$$

Now suppose that $\mathbb{E}\tau_a < \infty$. Then by the dominated convergence theorem, the right-hand side converges to $\mathbb{E}\tau_a$ as $\lambda \to 0$. In particular,

$$\lim_{\lambda \downarrow 0} \frac{1 - \mathbb{E}e^{-\lambda\tau_a}}{\lambda}$$

is finite. However, the preceding theorem shows that this is not the case. □

2.5 Exercises

1. Prove the assertion in Example 2.1.2.

2. Prove the assertion in Example 2.1.3.

3. Show that the processes defined in Example 2.1.4 are indeed martingales.

4. Let $X_1, X_2, \ldots$ be i.i.d. random variables and consider the tail σ-algebra defined by
$$\mathcal{T} = \bigcap_{n=1}^{\infty} \sigma(X_n, X_{n+1}, \ldots).$$
   (a) Show that for every n, $\mathcal{T}$ is independent of the σ-algebra $\sigma(X_1, \ldots, X_n)$ and conclude that for every $A \in \mathcal{T}$
$$\mathbb{P}(A) = \mathbb{E}(1_A \,|\, X_1, X_2, \ldots, X_n)$$
   almost surely.

   (b) Give a "martingale proof" of Kolmogorov's 0–1 law : for every $A \in \mathcal{T}$, $\mathbb{P}(A) = 0$ or $\mathbb{P}(A) = 1$.

5. In this exercise we present a "martingale proof" of the law of large numbers. Let $X_1, X_2, \ldots$ be i.i.d. random variables with $\mathbb{E}|X_1| < \infty$. Define $S_n = X_1 + \cdots + X_n$ and $\mathcal{F}_n = \sigma(S_n, S_{n+1}, \ldots)$.

   (a) Note that for $i = 1, \ldots n$, the distribution of the pair (X_i, S_n) is independent of i. From this fact, deduce that $\mathbb{E}(X_n \,|\, \mathcal{F}_n) = S_n/n$ and, consequently,
$$\mathbb{E}\left(\frac{1}{n}S_n \,\middle|\, \mathcal{F}_{n+1}\right) = \frac{1}{n+1}S_{n+1}$$
   almost surely.

   (b) Show that S_n/n converges almost surely to a finite limit.

   (c) Derive from Kolmogorov's 0–1 law that the limit must be a constant and determine its value.

6. Consider the proof of Theorem 2.2.17. Prove that for the stopping time τ and the event $A \in \mathcal{F}_\tau$ it holds that $A \cap \{\tau < \infty\} \in \mathcal{G}$.

7. Prove Corollary 2.3.3.

8. Let X and Y be two integrable random variables, $\mathcal{F}$ a σ-algebra and suppose that Y is $\mathcal{F}$-measurable. Show that if $\mathbb{E}(X \,|\, \mathcal{F}) \leq Y$ a.s. and $\mathbb{E}X = \mathbb{E}Y$, then $\mathbb{E}(X \,|\, \mathcal{F}) = Y$ a.s.

9. Show that the process $\tilde{M}$ constructed in the proof of Theorem 2.3.5 is cadlag.

10. Let M be a supermartingale. Prove that if M has a right-continuous modification, then $t \mapsto \mathbb{E}M_t$ is right-continuous.

11. Show that for every $a \neq 0$, the exponential martingale of Example 2.1.4 converges to 0 a.s. as $t \to \infty$. (Hint: use for instance the recurrence of the Brownian motion). Conclude that these martingales are not uniformly integrable.

12. Given an example of two stopping times $\sigma \leq \tau$ and a martingale M that is bounded in L^1 but not uniformly integrable, for which the equality $\mathbb{E}(M_\tau \mid \mathcal{F}_\sigma) = M_\sigma$ fails. (Hint: see Exercise 11.)

13. Let M be a positive, continuous martingale that converges a.s. to zero as t tends to infinity.

    (a) Prove that for every $x > 0$,
    $$\mathbb{P}\Big(\sup_{t \geq 0} M_t > x \mid \mathcal{F}_0\Big) = 1 \wedge \frac{M_0}{x}$$
    almost surely. (Hint: stop the martingale when it gets above the level x.)

    (b) Let W be a standard BM. Using the exponential martingales of Example 2.1.4, show that for every $a > 0$, the random variable
    $$\sup_{t \geq 0} (W_t - \tfrac{1}{2}at)$$
    has an exponential distribution with parameter a.

14. Let W be a BM and for $a \in \mathbb{R}$, let τ_a be the first time that W hits a. Suppose that $a > 0 > b$. By considering the stopped martingale $W^{\tau_a \wedge \tau_b}$, show that
$$\mathbb{P}(\tau_a < \tau_b) = \frac{-b}{a - b}.$$

15. Consider the setup of the preceding exercise. By stopping the martingale $W_t^2 - t$ at an appropriate stopping time, show that $\mathbb{E}(\tau_a \wedge \tau_b) = -ab$. Deduce that $\mathbb{E}\tau_a = \infty$.

16. Let W be a BM and for $a > 0$, let $\tilde{\tau}_a$ be the first time that the process $|W|$ hits the level a.

    (a) Show that for every $b > 0$, the process $M_t = \cosh(b|W_t|)\exp(-b^2 t/2)$ is martingale.

    (b) Find the Laplace transform of the stopping time $\tilde{\tau}_a$.

    (c) Calculate $\mathbb{E}\tilde{\tau}_a$.

3

Markov processes

3.1 Basic definitions

We begin with the following central definition.

Definition 3.1.1. Let $(E, \mathcal{E})$ be a measurable space. A *transition kernel* on E is a map $P : E \times \mathcal{E} \to [0, 1]$ such that

(i) for every $x \in E$, the map $B \mapsto P(x, B)$ is a probability measure on $(E, \mathcal{E})$,

(ii) for every $A \in \mathcal{E}$, the map $x \mapsto P(x, B)$ is measurable.

A transition kernel provides a mechanism for a random motion in the space E which can be described as follows. Say at time zero, one starts at a point $x_0 \in E$. Then at time 1, the next position x_1 is chosen according to the probability measure $P(x_0, dx)$. At time 2 the point x_2 is chosen according to $P(x_1, dx)$, etc. This motion is clearly *Markovian*, in the sense that at every point in time, the next position does not depend on the entire past, but only on the current position. Roughly speaking, a Markov process is a continuous-time process with this property.

Suppose that we have such a process X for which, for every $t \geq 0$, there exists a transition kernel P_t such that

$$\mathbb{P}(X_{t+s} \in B \,|\, \mathcal{F}_s^X) = P_t(X_s, B) \quad \text{a.s.}$$

for all $s \geq 0$. Then it is necessary that the transition kernels $(P_t)_{t \geq 0}$ satisfy a certain consistency relation. Indeed, by a standard approximation argument, we have

$$\mathbb{E}(f(X_{t+s}) \,|\, \mathcal{F}_s^X) = \int f(x) P_t(X_s, dx) \quad \text{a.s.}$$

for every nonnegative, measurable function f. It follows that for $s, t \geq 0$,

$$\begin{aligned} P_{t+s}(X_0, B) &= \mathbb{P}(X_{t+s} \in B \,|\, \mathcal{F}_0^X) \\ &= \mathbb{E}(\mathbb{P}(X_{t+s} \in B \,|\, \mathcal{F}_s^X) \,|\, \mathcal{F}_0^X) \\ &= \mathbb{E}(P_t(X_s, B) \,|\, \mathcal{F}_0^X) \\ &= \int P_t(y, B) P_s(X_0, dy) \end{aligned}$$

almost surely. This motivates the following definition.

Definition 3.1.2. Let $(E, \mathcal{E})$ be a measurable space. A collection of transition kernels $(P_t)_{t \geq 0}$ is called a *(homogenous) transition function* if for all $s, t \geq 0$, $x \in E$ and $B \in \mathcal{E}$

$$P_{t+s}(x, B) = \int P_s(x, dy) P_t(y, B).$$

This relation is known as the *Chapman-Kolmogorov equation.*

Integrals of the form $\int f \, d\nu$ are often written in operator notation as νf. It is useful to introduce a similar notation for transition kernels. If $P(x, dy)$ is a transition kernel on a measurable space $(E, \mathcal{E})$ and f is a nonnegative, measurable function on E, we define the function Pf by

$$Pf(x) = \int f(y) P(x, dy).$$

For $B \in \mathcal{E}$ we have $P1_B(x) = P(x, B)$, so $P1_B$ is a measurable function. By a standard approximation argument it is easily seen that for every nonnegative, measurable function f, the function Pf is measurable as well. Moreover, we see that if f is bounded, then Pf is also a bounded function. Hence, P can be viewed a linear operator on the space of bounded, measurable functions on E.

Note that in this operator notation, the Chapman-Kolmogorov equation states that for a transition function $(P_t)_{t \geq 0}$ it holds that

$$P_{t+s} f = P_t(P_s f) = P_s(P_t f)$$

for every bounded, measurable function f and $s, t \geq 0$. In other words, the operators $(P_t)_{t \geq 0}$ form a *semigroup* of operators on the space of bounded, measurable functions on E. In the sequel we will not distinguish between this semigroup and the corresponding (homogenous) transition function on $(E, \mathcal{E})$, since there is a one-to-one relation between the two concepts.

We can now give the definition of a Markov process.

Definition 3.1.3. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ be a filtered probability space, $(E, \mathcal{E})$ a measurable space and (P_t) a transition function on E. An adapted process X is called a *(homogenous) Markov process* with respect to the filtration $(\mathcal{F}_t)$ if for all $s, t \geq 0$ and every nonnegative, measurable function f on E,

$$\mathbb{E}(f(X_{t+s}) \,|\, \mathcal{F}_s) = P_t f(X_s) \quad \mathbb{P}\text{-a.s.}$$

The distribution of X_0 under $\mathbb{P}$ is called the *initial distribution* of the process, E is called the *state space.*

The following lemma will be useful in the next section. It shows in particular that the fdd's of a Markov process are determined by the transition function and the initial distribution.

Lemma 3.1.4. *A process X is Markov with respect to its natural filtration with transition function (P_t) and initial distribution ν if and only if for all $0 = t_0 < t_1 < \cdots < t_n$ and nonnegative, measurable functions $f_0, f_1, \ldots, f_n$,*

$$\mathbb{E}\prod_{i=0}^{n} f_i(X_{t_i}) = \nu f_0 P_{t_1-t_0} f_1 \cdots P_{t_n-t_{n-1}} f_n.$$

Proof. Exercise 1. □

The Brownian motion is a Markov process. In the following example we calculate its transition function.

Example 3.1.5. Let W be a standard BM and let $(\mathcal{F}_t)$ be its natural filtration. For $s, t \geq 0$ and a nonnegative, measurable function f, consider the conditional expectation $\mathbb{E}(f(W_{t+s}) \,|\, \mathcal{F}_s)$. Given $\mathcal{F}_s$, the random variable $W_{t+s} = (W_{t+s} - W_s) + W_s$ has a normal distribution with mean W_s and variance t . It follows (check!) that a.s.

$$\mathbb{E}(f(W_{t+s}) \,|\, \mathcal{F}_s) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi t}} f(y) e^{-\frac{1}{2}\frac{(W_s-y)^2}{t}}\, dy = P_t f(W_s),$$

where the transition kernel P_t on $\mathbb{R}$ is defined by

$$P_t f(x) = \int_{\mathbb{R}} f(y) p(t,x,y) dy,$$

with

$$p(t,x,y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{1}{2}\frac{(x-y)^2}{t}}.$$

Hence, the BM is a Markov process with respect to its natural filtration. ■

In general it is not true that a function of a Markov process is again Markovian. The following lemma gives a sufficient condition under which this is the case.

Lemma 3.1.6. *Let X be a Markov process with state space $(E, \mathcal{E})$ and transition function (P_t). Suppose that $(E', \mathcal{E}')$ is a measurable space and let $\varphi : E \to E'$ be measurable and onto. If (Q_t) is a collection of transition kernels such that*

$$P_t(f \circ \varphi) = (Q_t f) \circ \varphi$$

for all bounded, measurable functions f on E', then $Y = \varphi(X)$ is a Markov process with respect to its natural filtration, with state space $(E', \mathcal{E}')$ and transition function (Q_t).

Proof. Let f be a bounded, measurable function on E'. Then by the assumption and the semigroup property of (P_t),

$$(Q_tQ_sf)\circ\varphi = P_t((Q_sf)\circ\varphi) = P_tP_s(f\circ\varphi) = P_{t+s}(f\circ\varphi) = (Q_{t+s}f)\circ\varphi.$$

Since φ is onto, this implies that (Q_t) is a semigroup. Using the preceding lemma and the assumption it is easily verified that Y has the Markov property (see Exercise 2). □

3.2 Existence of a canonical version

In this section we show that for a given transition function (P_t) and probability measure ν on a measurable space $(E,\mathcal{E})$, we can construct a so-called canonical Markov process X which has ν as initial distribution and (P_t) as its transition function.

An E-valued process can be viewed as random element of the space $E^{\mathbb{R}_+}$ of E-valued functions on $\mathbb{R}_+$. Recall that a subset $A\subseteq E^{\mathbb{R}_+}$ is called a *cylinder* if it is of the form

$$A = \{f\in E^{\mathbb{R}_+} : f(t_1)\in E_1,\dots,f(t_n)\in E_n\}$$

for certain $t_1,\dots,t_n\geq 0$ and $E_1,\dots,E_n\in\mathcal{E}$. The product-$\sigma$-algebra $\mathcal{E}^{\mathbb{R}_+}$ on $E^{\mathbb{R}_+}$ is defined as the smallest σ-algebra which contains all cylinders. Equivalently, it is the smallest σ-algebra which makes all projections $E^{\mathbb{R}_+}\to\mathbb{R}$ given by $f\mapsto f(t)$ measurable. Now we define $\Omega = E^{\mathbb{R}_+}$ and $\mathcal{F} = \mathcal{E}^{\mathbb{R}_+}$ and on $(\Omega,\mathcal{F})$ we consider the process $X = (X_t)_{t\geq 0}$ defined by

$$X_t(\omega) = \omega(t), \quad \omega\in\Omega.$$

Viewed as a map from $\Omega\to E^{\mathbb{R}_+}$, X is simply the identity map. In particular, X is a measurable map $(\Omega,\mathcal{F})\to(E^{\mathbb{R}_+},\mathcal{E}^{\mathbb{R}_+})$ and it follows that for every $t\geq 0$, X_t is a measurable map $(\Omega,\mathcal{F})\to(E,\mathcal{E})$. Hence, X is an E-valued stochastic process on $(\Omega,\mathcal{F})$ in the sense of Definition 1.1.1. Note however that we have not yet defined a probability measure on $(\Omega,\mathcal{F})$.

Definition 3.2.1. The process X is called the *canonical process* on $(E^{\mathbb{R}_+},\mathcal{E}^{\mathbb{R}_+})$.

In Chapter 1 we presented (without proof) Kolmogorov's consistency theorem (see Theorem 1.2.3). It states that for every consistent collection of probability measures (cf. Definition 1.2.2) there exists a certain probability space and a process on this space which has these probability measures as its fdd's. At this point we need the following refinement of this result.

Theorem 3.2.2 (Kolmogorov's consistency theorem). *Suppose that E is a Polish space and $\mathcal{E}$ is its Borel σ-algebra. For all $t_1, \ldots, t_n \geq 0$, let $\mu_{t_1,\ldots,t_n}$ be a probability measure on $(E^n, \mathcal{E}^n)$. If the measures $\mu_{t_1,\ldots,t_n}$ form a consistent system, then there exists a probability measure $\mathbb{P}$ on the measurable space $(E^{\mathbb{R}_+}, \mathcal{E}^{\mathbb{R}_+})$ such that under $\mathbb{P}$, the canonical process X on $(E^{\mathbb{R}_+}, \mathcal{E}^{\mathbb{R}_+})$ has the measures $\mu_{t_1,\ldots,t_n}$ as its fdd's.*

Proof. See for instance Billingsley (1995). □

From this point on we assume that $(E, \mathcal{E})$ is a Polish space, endowed with its Borel σ-algebra. We have the following existence result for Markov processes with a given transition function and initial distribution.

Corollary 3.2.3. *Let (P_t) be a transition function on $(E, \mathcal{E})$ and let ν be a probability measure on $(E, \mathcal{E})$. Then there exists a unique probability measure $\mathbb{P}_\nu$ on $(\Omega, \mathcal{F}) = (E^{\mathbb{R}_+}, \mathcal{E}^{\mathbb{R}_+})$ such that under $\mathbb{P}_\nu$, the canonical process X is a Markov process with respect to its natural filtration $(\mathcal{F}_t^X)$, with initial distribution ν.*

Proof. For all $0 = t_0 \leq t_1 \leq \cdots \leq t_n$ we define the probability measure on $(E^{n+1}, \mathcal{E}^{n+1})$ by

$$\mu_{t_0,t_1,\ldots,t_n}(A_0 \times A_1 \times \cdots \times A_n) = \nu 1_{A_0} P_{t_1-t_0} 1_{A_1} \cdots P_{t_n-t_{n-1}} 1_{A_n}.$$

For arbitrary $t_1, \ldots, t_n \geq 0$, let π be a permutation of $\{1, \ldots, n\}$ such that $t_{\pi(1)} \leq \cdots \leq t_{\pi(n)}$ and define

$$\mu_{t_1,\ldots,t_n}(A_1 \times \cdots \times A_n) = \mu_{t_{\pi(1)},\ldots,t_{\pi(n)}}(A_{\pi(1)} \times \cdots \times A_{\pi(n)}).$$

Then by construction, the probability measures $\mu_{t_1,\ldots,t_n}$ satisfy condition (i) of Definition (1.2.2). By the Chapman-Kolmogorov equation we have $P_s 1_E P_t = P_{s+t}$ for all $s, t \geq 0$. From this fact it follows that condition (ii) is also satisfied, so the measures $\mu_{t_1,\ldots,t_n}$ form a consistent system (see Exercise 5). Hence, by Kolmogorov's consistency theorem there exists a probability measure $\mathbb{P}_\nu$ on $(\Omega, \mathcal{F}) = (E^{\mathbb{R}_+}, \mathcal{E}^{\mathbb{R}_+})$ such that under $\mathbb{P}_\nu$, the canonical process X has the measures $\mu_{t_1,\ldots,t_n}$ as its fdd's. In particular, we have for $0 = t_0 < t_1 < \cdots < t_n$ and $A_0, A_1, \ldots, A_n \in \mathcal{E}$

$$\begin{aligned}
\mathbb{P}_\nu(X_{t_0} \in A_0, \ldots, X_{t_n} \in A_n) &= \mu_{t_0,t_1,\ldots,t_n}(A_0 \times A_1 \times \cdots \times A_n) \\
&= \nu 1_{A_0} P_{t_1-t_0} 1_{A_1} \cdots P_{t_n-t_{n-1}} 1_{A_n}.
\end{aligned}$$

By Lemma 3.1.4 this implies that X is Markov w.r.t. its natural filtration. □

Consider a given transition function (P_t) on $(E, \mathcal{E})$ again. For $x \in E$, let δ_x be the Dirac measure concentrated at the point x. By Corollary 1.3.4 there exists a probability measure $\mathbb{P}_{\delta_x}$ such that under this measure, the canonical process X has δ_x as initial distribution. In the remainder of these notes, we simply write $\mathbb{P}_x$ instead of $\mathbb{P}_{\delta_x}$ and the corresponding expectation is denoted by

$\mathbb{E}_x$. Since $\mathbb{P}_x(X_0 = x) = 1$ we say that under $\mathbb{P}_x$, the process X *starts in* x. Note also that for all $x \in E$ and $A \in \mathcal{E}$,

$$\mathbb{P}_x(X_t \in A) = \int \delta_x(y) P_t(y, A) = P_t(x, A).$$

In particular, the map $x \mapsto \mathbb{P}_x(X_t \in A)$ is measurable for every $A \in \mathcal{E}$. This is generalized in the following lemma.

Lemma 3.2.4. *Let Z be an $\mathcal{F}^X_\infty$-measurable random variable, nonnegative or bounded. Then the map $x \mapsto \mathbb{E}_x Z$ is measurable and*

$$\mathbb{E}_\nu Z = \int \nu(dx) \mathbb{E}_x Z,$$

for every initial distribution ν.

Proof. It is easily seen that the class

$$\mathcal{S} = \left\{ \Gamma \in \mathcal{F}^X_\infty : x \mapsto \mathbb{E}_x 1_\Gamma \text{ is measurable and } \mathbb{E}_\nu 1_\Gamma = \int \nu(dx) \mathbb{E}_x 1_\Gamma \right\}$$

is a monotone class of subsets of $\mathcal{F}^X_\infty$. Let $\mathcal{G} \subseteq \mathcal{F}^X_\infty$ be the class of rectangles of the form $\{X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\}$. Using Lemma 3.1.4, it is easy to see that $\mathcal{G} \subseteq \mathcal{S}$. Since $\mathcal{G}$ is closed under finite intersections, it follows that

$$\mathcal{F}^X_\infty = \sigma(\mathcal{G}) \subseteq \mathcal{S},$$

by the monotone class theorem (see the appendix). So for every $\Gamma \in \mathcal{F}^X_\infty$, the statement of the lemma is true for the random variable $Z = 1_\Gamma$. By a standard approximation argument is follows that the statement is true for every nonnegative or bounded $\mathcal{F}^X_\infty$-measurable random variable Z. (See also Exercise 6) $\square$

For $t \geq 0$ we define the *translation operator* $\theta_t : E^{\mathbb{R}_+} \to E^{\mathbb{R}_+}$ by $\theta_t f(s) = f(t+s)$. So θ_t just cuts of the part of the path of f before time t and shifts the remaining part to the origin. Clearly, $\theta_t \circ \theta_s = \theta_{t+s}$ and every θ_t is $\mathcal{E}^{\mathbb{R}_+}$-measurable. Using the translation operators, we can formulate the Markov property as follows.

Theorem 3.2.5. *Let Z be an $\mathcal{F}^X_\infty$-measurable random variable, nonnegative or bounded. Then for every $t > 0$ and initial distribution ν,*

$$\mathbb{E}_\nu(Z \circ \theta_t \,|\, \mathcal{F}^X_t) = \mathbb{E}_{X_t} Z \qquad \mathbb{P}_\nu\text{-a.s.}$$

Before we turn to the proof, let us remark that the right-hand side of the display should be read as the evaluation at the (random) point X_t of the function $x \mapsto \mathbb{E}_x Z$. So by Lemma 3.2.4 it is a measurable function of X_t.

Proof. We have to show that for $A \in \mathcal{F}_t^X$

$$\int_A Z \circ \theta_t \, d\mathbb{P}_\nu = \int_A \mathbb{E}_{X_t} Z \, d\mathbb{P}_\nu.$$

By the usual approximation arguments it is enough to show this equality for A of the form $A = \{X_{t_0} \in A_0, \dots, X_{t_n} \in A_n\}$ with $0 = t_0 < \dots < t_n = t$ and Z of the form

$$Z = \prod_{i=1}^m f_i(X_{s_i}),$$

for $s_1 \le \dots \le s_m$ and certain nonnegative, measurable functions f_i. In this case the left-hand side equals

$$\mathbb{E}_\nu \prod_{j=0}^n 1_{A_j}(X_{t_j}) \prod_{i=1}^m f_i(X_{t+s_i}).$$

By Lemma 3.1.4, the right-hand side equals

$$\mathbb{E}_\nu \prod_{j=0}^n 1_{A_j}(X_{t_j}) \left(P_{s_1} f_1 P_{s_2 - s_1} f_2 \cdots P_{s_m - s_{m-1}} f_m\right)(X_t).$$

Using Lemma 3.1.4 again we see that the two expressions are equal. □

3.3 Feller processes

3.3.1 Feller transition functions and resolvents

In Section 3.1 we saw that a homogenous transition function $(P_t)_{t \ge 0}$ on a measurable space $(E, \mathcal{E})$ can be viewed as a semigroup of operators on the space of bounded, measurable functions on E. In this section we consider semigroups with additional properties. For simplicity, the state space E is assumed to be a subset of $\mathbb{R}^d$, and $\mathcal{E}$ is its Borel-σ-algebra. By $C_0 = C_0(E)$ we denote the space of real-valued, continuous functions on E which vanish at infinity. In other words, a function $f \in C_0$ is continuous on E and has the property that $f(x) \to 0$ as $\|x\| \to \infty$. Functions in C_0 are bounded, so we can endow the space with the sup-norm, which is defined by

$$\|f\|_\infty = \sup_{x \in E} |f(x)|$$

for $f \in C_0$. Note that C_0 is a subset of the space of bounded, measurable functions on E, so we can consider the restriction of the transition operators P_t to C_0.

Definition 3.3.1. The transition function $(P_t)_{t\geq 0}$ is called a *Feller transition function* if

(i) $P_t C_0 \subseteq C_0$ for all $t \geq 0$,

(ii) for every $f \in C_0$ and $x \in E$, $P_t f(x) \to f(x)$ as $t \downarrow 0$.

A Markov process with a Feller transition function is a called a *Feller process*.

Observe that the operators P_t are *contractions* on C_0, i.e. for every $f \in C_0$, we have

$$\|P_t f\|_\infty = \sup_{x\in E} |P_t f(x)| = \sup_{x\in E} \left| \int_E f(y) P_t(x, dy) \right| \leq \|f\|_\infty.$$

So for all $t \geq 0$ we have $\|P_t\| \leq 1$, where $\|P_t\|$ is the norm of P_t as a linear operator on the normed linear space C_0, endowed with the supremum norm (see Appendix B for the precise definitions of these notions).

If $f \in C_0$, then $P_t f$ is also in C_0 by part (i) of Definition 3.3.1. By the semigroup property and part (ii) it follows that

$$P_{t+h} f(x) = P_h P_t f(x) \to P_t f(x)$$

as $h \downarrow 0$. In other words, the map $t \mapsto P_t f(x)$ is right-continuous for all $f \in C_0$ and $x \in E$. In particular we see that this map is measurable, so for all $\lambda > 0$ we may define

$$R_\lambda f(x) = \int_0^\infty e^{-\lambda t} P_t f(x)\, dt.$$

If we view $t \mapsto P_t$ as an operator-valued map, then R_λ is simply its Laplace transform, calculated at the 'frequency' λ. The next theorem collects the most important properties of the operators R_λ. In particular, it states that for all $\lambda > 0$, R_λ is in fact an operator which maps C_0 into itself. It is called the *resolvent of order* λ.

Theorem 3.3.2. *Let R_λ, $\lambda > 0$ be the resolvents of a Feller transition function. Then $R_\lambda C_0 \subseteq C_0$ for every $\lambda > 0$ and the* resolvent equation

$$R_\mu - R_\lambda + (\mu - \lambda) R_\mu R_\lambda = 0$$

holds for all $\lambda, \mu > 0$. Moreover, the image of R_λ does not depend on λ and is dense in C_0.

Proof. Denote the transition function by P_t. For $f \in C_0$, we have $P_t f \in C_0$ for every $t \geq 0$. Hence, if $x_n \to x$ in E, the function $t \mapsto P_t f(x_n)$ converges pointwise to $t \mapsto P_t f(x)$. By dominated convergence it follows that

$$R_\lambda f(x_n) = \int_0^\infty e^{-\lambda t} P_t f(x_n)\, dt \to \int_0^\infty e^{-\lambda t} P_t f(x)\, dt = R_\lambda f(x),$$

hence R_λ maps C_0 into the space of continuous functions on E. Using the same reasoning, we see that if $\|x_n\| \to \infty$, then $R_\lambda f(x_n) \to 0$. So indeed, $R_\lambda C_0 \subseteq C_0$.

To prove the resolvent equation, note that

$$e^{-\mu t} - e^{-\lambda t} = (\lambda - \mu)e^{-\lambda t} \int_0^t e^{(\lambda-\mu)s}\, ds.$$

Hence,

$$\begin{aligned}
R_\mu f(x) - R_\lambda f(x) &= \int_0^\infty (e^{-\mu t} - e^{-\lambda t}) P_t f(x)\, dt \\
&= (\lambda - \mu) \int_0^\infty e^{-\lambda t} \left(\int_0^t e^{(\lambda-\mu)s} P_t f(x)\, ds \right) dt \\
&= (\lambda - \mu) \int_0^\infty e^{-\mu s} \left(\int_s^\infty e^{-\lambda(t-s)} P_t f(x)\, dt \right) ds,
\end{aligned}$$

by Fubini's theorem. A change of variables, the semigroup property of the transition function and another application of Fubini show that the inner integral equals

$$\begin{aligned}
\int_0^\infty e^{-\lambda u} P_{s+u} f(x)\, du &= \int_0^\infty e^{-\lambda u} P_s P_u f(x)\, du \\
&= \int_0^\infty e^{-\lambda u} \left(\int_E P_u f(y)\, P_s(x, dy) \right) du \\
&= \int_E \left(\int_0^\infty e^{-\lambda u} P_u f(y)\, du \right) P_s(x, dy) \\
&= P_s R_\lambda f(x).
\end{aligned} \tag{3.1}$$

Combined with the preceding display, this yields the desired equation.

The resolvent equation implies that $R_\lambda = R_\mu(I + (\mu - \lambda)R_\lambda)$, which shows that the image of the map R_λ is contained in the image of R_μ. Since this also holds with the roles of λ and μ reversed, the image D of R_λ is independent of λ.

To prove that D is dense in C_0, consider an arbitrary, bounded linear functional A on C_0 that vanishes on D. By the Riesz representation theorem (Theorem B.2.1), there exist finite Borel measures ν and ν' on E such that

$$A(f) = \int_E f\, d\nu - \int_E f\, d\nu' = \int_E f\, d(\nu - \nu')$$

for every $f \in C_0$. Now fix $f \in C_0$. Then by part (ii) of Definition 3.3.1 and dominated convergence, it holds for every $x \in E$ that

$$\lambda R_\lambda f(x) = \int_0^\infty \lambda e^{-\lambda t} P_t f(x)\, dt = \int_0^\infty e^{-s} P_{s/\lambda} f(x)\, ds \to f(x) \tag{3.2}$$

as $\lambda \to \infty$. Noting also that $\|\lambda R_\lambda f\|_\infty \le \|f\|_\infty$, dominated convergence implies that

$$0 = A(\lambda R_\lambda f) = \int_E \lambda R_\lambda f(x)\, (\nu - \nu')(dx) \to \int_E f(x)\, (\nu - \nu')(dx) = A(f).$$

We conclude that the functional A vanishes on the entire space C_0. By Corollary B.1.2 to the Hahn-Banach theorem, this shows that D is dense in C_0. $\square$

Observe that for every $f \in C_0$, the resolvent R_λ satisfies

$$\|R_\lambda f\|_\infty \leq \int_0^\infty e^{-\lambda t} \|P_t f\|_\infty \, dt \leq \|f\|_\infty \int_0^\infty e^{-\lambda t} \, dt = \frac{\|f\|_\infty}{\lambda}.$$

Hence, for the linear transformation $R_\lambda : C_0 \to C_0$ we have $\|R_\lambda\| \leq 1/\lambda$. Let us also note that in the proof of Theorem 3.3.2 we saw that

$$P_t R_\lambda f(x) = \int_0^\infty e^{-\lambda u} P_{t+u} f(x) \, du. \tag{3.3}$$

By the semigroup property of the transition function, the right-hand side of this equation equals $R_\lambda P_t f(x)$. In other words, the operators P_t and R_λ commute.

The following corollary to the theorem states that for Feller transition functions, the pointwise convergence of part (ii) of Definition 3.3.1 is actually equivalent to the seemingly stronger uniform convergence. This is called the *strong continuity* of a Feller semigroup. Similarly, the pointwise convergence in (3.2) is strengthened to uniform convergence.

Corollary 3.3.3. *For a Feller transition function P_t and its resolvents R_λ it holds for every $f \in C_0$ that $\|P_t f - f\|_\infty \to 0$ as $t \downarrow 0$, and $\|\lambda R_\lambda f - f\|_\infty \to 0$ as $\lambda \to \infty$.*

Proof. Relation (3.1) in the proof of Theorem 3.3.2 shows that for $f \in C_0$,

$$P_t R_\lambda f(x) = e^{\lambda t} \int_t^\infty e^{-\lambda s} P_s f(x) \, ds.$$

It follows that

$$P_t R_\lambda f(x) - R_\lambda f(x) = (e^{\lambda t} - 1) \int_t^\infty e^{-\lambda s} P_s f(x) \, ds - \int_0^t e^{-\lambda s} P_s f(x) \, ds,$$

hence

$$\|P_t R_\lambda f - R_\lambda f\|_\infty \leq (e^{\lambda t} - 1) \|R_\lambda f\|_\infty + t \|f\|_\infty.$$

Since the right-hand side vanishes as $t \downarrow 0$, this proves the first statement of the corollary for functions in the joint image D of the resolvents. Now let $f \in C_0$ be arbitrary. Then for every $g \in D$, it holds that

$$\begin{aligned} \|P_t f - f\|_\infty &\leq \|P_t f - P_t g\|_\infty + \|P_t g - g\|_\infty + \|g - f\|_\infty \\ &\leq \|P_t g - g\|_\infty + 2\|g - f\|_\infty. \end{aligned}$$

Since D is dense in C_0, the second term can be made arbitrarily small by choosing an appropriate function g. For that choice of $g \in D$, the first term vanishes as $t \downarrow 0$, by the first part of the proof. This completes the proof of the first statement of the lemma. The second statement follows easily from the first one by dominated convergence. $\square$

Example 3.3.4. The BM is a Feller process. Its resolvents are given by

$$R_\lambda f(x) = \int_{\mathbb{R}} f(y) r_\lambda(x, y)\, dy,$$

where $r_\lambda(x, y) = \exp\left(-\sqrt{2\lambda}|x - y|\right)/\sqrt{(2\lambda)}$ (see Exercise 7). ■

Lemma 3.1.6 gives conditions under which a function of a Markov process is again Markovian. The corresponding result for Feller processes is as follows.

Lemma 3.3.5. *Let X be a Feller process with state space E and transition function (P_t). Suppose that $\varphi : E \to E'$ is continuous and onto, and that $\|\varphi(x_n)\| \to \infty$ in E' if and only if $\|x_n\| \to \infty$ in E. Then if (Q_t) is a collection of transition kernels such that*

$$P_t(f \circ \varphi) = (Q_t f) \circ \varphi$$

for all $f \in C_0(E')$, the process $Y = \varphi(X)$ is Feller with respect to its natural filtration, with state space E' and transition function (Q_t).

Proof. By Lemma 3.1.6, we only have to show that the semigroup (Q_t) is Feller. The assumptions on φ imply that if $f \in C_0(E')$, then $f \circ \varphi \in C_0(E)$. The Feller property of (Q_t) therefore follows from that of (P_t) (Exercise 8). □

3.3.2 Existence of a cadlag version

In this section we consider a Feller transition function P_t on $(E, \mathcal{E})$, where $E \subseteq \mathbb{R}^d$ and $\mathcal{E}$ is the Borel-σ-algebra of E. By Corollary 3.2.3 there exists for every probability measure ν on $(E, \mathcal{E})$ a probability measure $\mathbb{P}_\nu$ on the canonical space $(\Omega, \mathcal{F}) = (E^{\mathbb{R}_+}, \mathcal{E}^{\mathbb{R}_+})$ such that under $\mathbb{P}_\nu$, the canonical process X is a Markov process with respect to the natural filtration $(\mathcal{F}_t^X)$, with transition function (P_t) and initial distribution ν. As in the preceding section we write C_0 instead of $C_0(E)$ and we denote the resolvent of order $\lambda > 0$ by R_λ.

Our first aim in this section is to prove that a Feller process always admits a cadlag modification. We need the following lemma, which will allow us to use the regularization results for supermartingales of the preceding chapter.

Lemma 3.3.6. *For every $\lambda > 0$ and every nonnegative function $f \in C_0$, the process*

$$e^{-\lambda t} R_\lambda f(X_t)$$

is a $\mathbb{P}_\nu$-supermartingale with respect to the filtration $(\mathcal{F}_t^X)$, for every initial distribution ν.

Proof. By the Markov property we have

$$\mathbb{E}_\nu(e^{-\lambda t}R_\lambda f(X_t)\,|\,\mathcal{F}_s^X) = e^{-\lambda t}P_{t-s}R_\lambda f(X_s)$$

$\mathbb{P}_\nu$-a.s. (see Theorem 3.2.5). Hence, to prove the statement of the lemma it suffices to show that

$$e^{-\lambda t}P_{t-s}R_\lambda f(x) \le e^{-\lambda s}R_\lambda f(x)$$

for every $x \in E$, which follows from a straight-forward calculation (see Exercise 9). □

We can now prove that a Feller process admits a cadlag modification. The proof relies on some topological facts that have been collected in Exercises 10–12.

Theorem 3.3.7. *The canonical Feller process X admits a cadlag modification. More precisely, there exists a cadlag process Y on the canonical space $(\Omega, \mathcal{F})$ such that for all $t \ge 0$ and every initial distribution ν on $(E, \mathcal{E})$, we have $Y_t = X_t$, $\mathbb{P}_\nu$-a.s.*

Proof. Fix an arbitrary initial distribution ν on $(E, \mathcal{E})$. Let $\mathcal{H}$ be a countable, dense subset of the space $C_0^+(E)$ of nonnegative functions in $C_0(E)$. Then $\mathcal{H}$ separates the points of the one-point compactification E^* of E (see Exercise 10) and by the second statement of Corollary 3.3.3, the class

$$\mathcal{H}' = \{nR_n h : h \in \mathcal{H}, n \in \mathbb{N}\}$$

has the same property. Lemma 3.3.6 and Theorem 2.3.2 imply that $\mathbb{P}_\nu$-a.s., the limit

$$\lim_{\substack{r \downarrow t \\ r \in \mathbb{Q}}} h(X_r)$$

exists for all $h \in \mathcal{H}'$ and $t \ge 0$. In view of Exercise 11, it follows that $\mathbb{P}_\nu$-a.s., the limit

$$\lim_{\substack{r \downarrow t \\ r \in \mathbb{Q}}} X_r$$

exists in E^*, for all $t \ge 0$. So if $\Omega' \subseteq \Omega$ is the event on which these limits exist, it holds that $\mathbb{P}_\nu(\Omega') = 1$ for every initial distribution ν.

Now fix an arbitrary point $x_0 \in E$ and define a new process Y as follows. For $\omega \notin \Omega'$, we put $Y_t(\omega) = x_0$ for all $t \ge 0$. For $\omega \in \Omega'$ and $t \ge 0$, define

$$Y_t(\omega) = \lim_{\substack{r \downarrow t \\ r \in \mathbb{Q}}} X_r(\omega).$$

We claim that for every initial distribution ν and $t \ge 0$, we have $Y_t = X_t$, $\mathbb{P}_\nu$-a.s. To prove this, let f and g be two bounded, continuous functions on E^*. Then by dominated convergence and the Markov property,

$$\begin{aligned}
\mathbb{E}_\nu f(X_t)g(Y_t) &= \lim_{\substack{r \downarrow t \\ r \in \mathbb{Q}}} \mathbb{E}_\nu f(X_t)g(X_r) \\
&= \lim_{\substack{r \downarrow t \\ r \in \mathbb{Q}}} \mathbb{E}_\nu \mathbb{E}_\nu(f(X_t)g(X_r)\,|\,\mathcal{F}_t^X) \\
&= \lim_{\substack{r \downarrow t \\ r \in \mathbb{Q}}} \mathbb{E}_\nu f(X_t)P_{r-t}g(X_t).
\end{aligned}$$

By Corollary 3.3.3, we $\mathbb{P}_\nu$-a.s. have $P_{r-t}g(X_t) \to g(X_t)$ as $r \downarrow t$. By dominated convergence, it follows that $\mathbb{E}_\nu f(X_t)g(Y_t) = \mathbb{E}_\nu f(X_t)g(X_t)$. Hence, by Exercise 12 we indeed have $Y_t = X_t$, $\mathbb{P}_\nu$-a.s.

The process Y is right-continuous by construction, and we have shown that Y is a modification of X. It remains to prove that for every initial distribution ν, Y has left limits with $\mathbb{P}_\nu$-probability 1. To this end, note that for all $h \in \mathcal{H}'$, the process $h(Y)$ is a right-continuous martingale. By Corollary 2.3.3 this implies that $h(Y)$ has left limits with $\mathbb{P}_\nu$-probability one. In view of Exercise 11, it follows that Y has left limits with $\mathbb{P}_\nu$-probability one. □

3.3.3 Existence of a good filtration

Let X be the canonical, cadlag version of a Feller process with state space $E \subseteq \mathbb{R}^d$ and transition function P_t (see Theorem 3.3.7). So far we have only been working with the natural filtration $(\mathcal{F}_t^X)$. In general, this filtration is neither complete nor right-continuous. We would like to replace it by a larger filtration that satisfies the usual conditions (see Definition 1.6.3), and with respect to which the process X is still a Markov process.

We first construct a new filtration for every fixed initial distribution ν. We let $\mathcal{F}_\infty^\nu$ be the completion of $\mathcal{F}_\infty^X$ with respect to $\mathbb{P}_\nu$. This means that $\mathcal{F}_\infty^\nu$ consists of the sets B for which there exist $B_1, B_2 \in \mathcal{F}_\infty^X$ such that $B_1 \subseteq B \subseteq B_2$ and $\mathbb{P}_\nu(B_2 \backslash B_1) = 0$. The $\mathbb{P}_\nu$-probability of such a set B is defined as $\mathbb{P}_\nu(B) = \mathbb{P}_\nu(B_1) = \mathbb{P}_\nu(B_2)$, thus extending the definition of $\mathbb{P}_\nu$ to $\mathcal{F}_\infty^\nu$. By N^ν we denote the $\mathbb{P}_\nu$-negligible sets in $\mathcal{F}_\infty^\nu$, i.e. the sets $B \in \mathcal{F}_\infty^\nu$ for which $\mathbb{P}_\nu(B) = 0$. The filtration $(\mathcal{F}_t^\nu)$ is then defined by

$$\mathcal{F}_t^\nu = \sigma\left(\mathcal{F}_t^X \cup N^\nu\right).$$

We define the filtration $(\mathcal{F}_t)$ by putting

$$\mathcal{F}_t = \bigcap_\nu \mathcal{F}_t^\nu,$$

where the intersection is taken over all probability measures on the state space $(E, \mathcal{E})$. We call $(\mathcal{F}_t)$ the *usual augmentation* of the natural filtration $(\mathcal{F}_t^X)$.

It turns out that by just adding null sets, we have made the filtration right-continuous.

Theorem 3.3.8. *The filtrations $(\mathcal{F}_t^\nu)$ and $(\mathcal{F}_t)$ are right-continuous.*

Proof. It is easily seen that the right-continuity of $(\mathcal{F}_t^\nu)$ implies the right-continuity of $(\mathcal{F}_t)$, so it suffices to prove the former. Let us first show that for every nonnegative, $\mathcal{F}_\infty^X$-measurable random variable Z,

$$\mathbb{E}_\nu(Z \,|\, \mathcal{F}_t^\nu) = \mathbb{E}_\nu(Z \,|\, \mathcal{F}_{t+}^\nu) \quad \mathbb{P}_\nu\text{-a.s.} \tag{3.4}$$

By a monotone class argument, it is enough to prove this equality for Z of the form
$$Z = \prod_{i=1}^{n} f_i(X_{t_i}),$$
with $t_1 < \cdots < t_n$ and $f_i \in C_0$ for $i = 1, \ldots, n$. Now suppose that $t_{k-1} \leq t < t_k$. Since $\mathcal{F}^\nu_{t+h}$ differs only from $\mathcal{F}^X_{t+h}$ by $\mathbb{P}_\nu$-negligible sets we have $\mathbb{E}_\nu(Z \,|\, \mathcal{F}^\nu_{t+h}) = \mathbb{E}_\nu(Z \,|\, \mathcal{F}^X_{t+h})$, $\mathbb{P}_\nu$-a.s. Hence, the Markov property implies that for h small enough,

$$\begin{aligned}
\mathbb{E}_\nu(Z \,|\, \mathcal{F}^\nu_{t+h}) &= \prod_{i=1}^{k-1} f_i(X_{t_i}) \mathbb{E}_\nu\big(\prod_{i=k}^{n} f_i(X_{t_i}) \,|\, \mathcal{F}^X_{t+h}\big) \\
&= g_h(X_{t+h}) \prod_{i=1}^{k-1} f_i(X_{t_i})
\end{aligned}$$

$\mathbb{P}_\nu$-a.s., where

$$g_h(x) = P_{t_k-(t+h)} f_k P_{t_{k+1}-t_k} f_{k+1} \cdots P_{t_n - t_{n-1}} f_n(x).$$

By the strong continuity of the semigroup P_t (Corollary 3.3.3) we have $\|g_h - g\|_\infty \to 0$ as $h \downarrow 0$, where

$$g(x) = P_{t_k-t} f_k P_{t_{k+1}-t_k} f_{k+1} \cdots P_{t_n - t_{n-1}} f_n(x).$$

Moreover, the right-continuity of X implies that $X_{t+h} \to X_t$, $\mathbb{P}_\nu$-a.s., so as $h \downarrow 0$, we $\mathbb{P}_\nu$-a.s. have

$$\mathbb{E}_\nu(Z \,|\, \mathcal{F}^\nu_{t+h}) \to g(X_t) \prod_{i=1}^{k-1} f_i(X_{t_i}) = \mathbb{E}_\nu(Z \,|\, \mathcal{F}^\nu_t).$$

On the other hand, by Theorem 2.2.15, the conditional expectation on the left-hand side converges $\mathbb{P}_\nu$-a.s. to $\mathbb{E}_\nu(Z \,|\, \mathcal{F}^\nu_{t+})$ as $h \downarrow 0$, which completes the proof of (3.4).

Now suppose that $B \in \mathcal{F}^\nu_{t+}$. Then 1_B is $\mathcal{F}^\nu_\infty$-measurable and since $\mathcal{F}^\nu_\infty$ is the $\mathbb{P}_\nu$-completion of $\mathcal{F}^X_\infty$, there exists an $\mathcal{F}^X_\infty$- measurable random variable Z such that $\mathbb{P}_\nu$-a.s., $1_B = Z$. By (3.4), it follows that we $\mathbb{P}_\nu$-a.s. have

$$1_B = \mathbb{E}_\nu(1_B \,|\, \mathcal{F}^\nu_{t+}) = \mathbb{E}_\nu(Z \,|\, \mathcal{F}^\nu_{t+}) = \mathbb{E}_\nu(Z \,|\, \mathcal{F}^\nu_t).$$

This shows that the random variable 1_B is $\mathcal{F}^\nu_t$-measurable, so $B \in \mathcal{F}^\nu_t$. □

Our next aim is to prove that the statement of Theorem 3.2.5 remains true if we replace the natural filtration $(\mathcal{F}^X_t)$ by its usual augmentation $(\mathcal{F}_t)$. In particular, we want to show that X is still a Markov process with respect to $(\mathcal{F}_t)$.

But first we have to address some measurability issues. We begin by considering the completion of the Borel-σ-algebra $\mathcal{E} = \mathcal{B}(E)$ of the state space E. If μ is a probability measure on $(E, \mathcal{E})$, we denote by $\mathcal{E}^\mu$ the completion of $\mathcal{E}$ w.r.t. μ. Next, we define
$$\mathcal{E}^* = \bigcap_\mu \mathcal{E}^\mu,$$

where the intersection is taken over all probability measures on $(E, \mathcal{E})$. The σ-algebra $\mathcal{E}^*$ is called the σ-algebra of *universally measurable sets.*

Lemma 3.3.9. *If Z is a bounded, $\mathcal{F}_\infty$-measurable random variable, then the map $x \mapsto \mathbb{E}_x Z$ is $\mathcal{E}^*$-measurable and*

$$\mathbb{E}_\nu Z = \int \nu(dx)\mathbb{E}_x Z,$$

for every initial distribution ν.

Proof. Fix the initial distribution ν and note that $\mathcal{F}_\infty \subseteq \mathcal{F}_\infty^\nu$. By definition of $\mathcal{F}_\infty^\nu$ there exist two $\mathcal{F}_\infty^X$ random variables Z_1, Z_2 such that $Z_1 \leq Z \leq Z_2$ and $\mathbb{E}_\nu(Z_2 - Z_1) = 0$. It follows that for all $x \in E$, we have $\mathbb{E}_x Z_1 \leq \mathbb{E}_x Z \leq \mathbb{E}_x Z_2$. Moreover, the maps $x \mapsto \mathbb{E}_x Z_i$ are $\mathcal{E}$-measurable by Lemma 3.2.4 and

$$\int (\mathbb{E}_x Z_2 - \mathbb{E}_x Z_1)\, \nu(dx) = \mathbb{E}_\nu(Z_2 - Z_1) = 0.$$

By definition of $\mathcal{E}^\nu$ this shows that $x \mapsto \mathbb{E}_x Z$ is $\mathcal{E}^\nu$-measurable and that

$$\mathbb{E}_\nu Z = \mathbb{E}_\nu Z_1 = \int \nu(dx)\mathbb{E}_x Z_1 = \int \nu(dx)\mathbb{E}_x Z.$$

Since ν is arbitrary, it follows that $x \mapsto \mathbb{E}_x Z$ is in fact $\mathcal{E}^*$-measurable. $\square$

Lemma 3.3.10. *For all $t \geq 0$, the random variable X_t is measurable as a map from $(\Omega, \mathcal{F}_t)$ to $(E, \mathcal{E}^*)$.*

Proof. Take $A \in \mathcal{E}^*$ and fix an initial distribution ν on $(E, \mathcal{E})$. Denote the distribution of X_t on $(E, \mathcal{E})$ under $\mathbb{P}_\nu$ by μ. Since $\mathcal{E}^* \subseteq \mathcal{E}^\mu$, there exist $A_1, A_2 \in \mathcal{E}$ such that $A_1 \subseteq A \subseteq A_2$ and $\mu(A_2 \backslash A_1) = 0$. Consequently, $X_t^{-1}(A_1) \subseteq X_t^{-1}(A) \subseteq X_t^{-1}(A_2)$. Since $X_t^{-1}(A_1), X_t^{-1}(A_2) \in \mathcal{F}_t^X$ and

$$\mathbb{P}_\nu(X_t^{-1}(A_2) \backslash X_t^{-1}(A_1)) = \mathbb{P}_\nu(X_t^{-1}(A_2 \backslash A_1)) = \mu(A_2 \backslash A_1) = 0,$$

this shows that $X_t^{-1}(A)$ is contained in the $\mathbb{P}_\nu$-completion of $\mathcal{F}_t^X$. But ν is arbitrary, so the proof is complete. $\square$

We can now prove that formulation of the Markov property in terms of the shift operators θ_t is still valid for the usual augmentation $(\mathcal{F}_t)$ of the natural filtration of our Feller process.

Theorem 3.3.11 (Markov property). *Let Z be an $\mathcal{F}_\infty$-measurable random variable, nonnegative or bounded. Then for every $t > 0$ and initial distribution ν,*

$$\mathbb{E}_\nu(Z \circ \theta_t \,|\, \mathcal{F}_t) = \mathbb{E}_{X_t} Z \qquad \mathbb{P}_\nu\text{-a.s.}$$

In particular, X is still a Markov process with respect to $(\mathcal{F}_t)$.

Proof. By combining the two preceding lemmas we see that the random variable $\mathbb{E}_{X_t} Z$ is $\mathcal{F}_t$-measurable, so we only have to prove that for any $A \in \mathcal{F}_t$,

$$\int_A Z \circ \theta_t \, d\mathbb{P}_\nu = \int_A \mathbb{E}_{X_t} Z \, d\mathbb{P}_\nu. \tag{3.5}$$

Assume that Z is bounded and denote the law of X_t under $\mathbb{P}_\nu$ by μ. By definition of $\mathcal{F}_\infty$, there exists an $\mathcal{F}^X_\infty$- measurable random variable Z' such that $\{Z \neq Z'\} \subseteq \Gamma$, with $\Gamma \in \mathcal{F}^X_\infty$ and $\mathbb{P}_\mu(\Gamma) = 0$. We have

$$\{Z \circ \theta_t \neq Z' \circ \theta_t\} = \theta_t^{-1}(\{Z \neq Z'\}) \subseteq \theta_t^{-1}(\Gamma)$$

and by Theorem 3.2.5,

$$\mathbb{P}_\nu(\theta_t^{-1}(\Gamma)) = \mathbb{E}_\nu(1_\Gamma \circ \theta_t) = \mathbb{E}_\nu \mathbb{E}_\nu(1_\Gamma \circ \theta_t \,|\, \mathcal{F}^X_t) = \mathbb{E}_\nu \mathbb{E}_{X_t} 1_\Gamma = \mathbb{E}_\nu \varphi(X_t),$$

where $\varphi(x) = \mathbb{E}_x 1_\Gamma = \mathbb{P}_x(\Gamma)$. Since the distribution of X_t under $\mathbb{P}_\nu$ is given by μ, we have

$$\mathbb{E}_\nu \varphi(X_t) = \int \mu(dx)\varphi(x) = \int \mu(dx)\mathbb{P}_x(\Gamma) = \mathbb{P}_\mu(\Gamma) = 0,$$

so $\mathbb{P}_\nu(\theta_t^{-1}(\Gamma)) = 0$. This shows that on the left-hand side of (3.5), we may replace Z by Z'.

The last two displays show that the two probability measures $B \mapsto \mathbb{E}_\nu \mathbb{E}_{X_t} 1_B$ and $\mathbb{P}_\mu$ coincide. Hence, since $\mathbb{P}_\mu(Z \neq Z') \leq \mathbb{P}_\mu(\Gamma) = 0$,

$$\mathbb{E}_\nu \left| \mathbb{E}_{X_t} Z - \mathbb{E}_{X_t} Z' \right| \leq \mathbb{E}_\nu \mathbb{E}_{X_t} |Z - Z'| = \mathbb{E}_\mu |Z - Z'| = 0.$$

It follows that $\mathbb{E}_{X_t} Z = \mathbb{E}_{X_t} Z'$, $\mathbb{P}_\nu$-a.s., so on the right-hand side of (3.5) we may also replace Z be Z'. Since Z' is $\mathcal{F}^X_\infty$-measurable, the statement now follows from Theorem 3.2.5. □

3.4 Strong Markov property

3.4.1 Strong Markov property of a Feller process

Let X again be a Feller process with state space $E \subseteq \mathbb{R}^d$. In view of the preceding sections, we consider the canonical, cadlag version of X, which is a Markov process with respect of the usual augmentation $(\mathcal{F}_t)$ of the natural filtration of the canonical process X. As before, we denote the shift operators by θ_t.

In this section we will prove that for Feller processes, the Markov property of Theorem 3.3.11 does not only hold for deterministic times t, but also for $(\mathcal{F}_t)$-stopping times. This is called the *strong Markov property.* Recall that for deterministic $t \geq 0$, the shift operator θ_t on the canonical space Ω maps a path

$s \mapsto \omega(s)$ to the path $s \mapsto \omega(t+s)$. Likewise, we now define θ_τ for a random time τ as the operator that maps the path $s \mapsto \omega(s)$ to the path $s \mapsto \omega(\tau(\omega)+s)$. This definition does not cause confusion since if τ equals the deterministic time t, then $\tau(\omega) = t$ for all $\omega \in \Omega$, so θ_τ is equal to the old operator θ_t. Observe that since the canonical process X is just the identity on the space Ω, we have for instance $(X_t \circ \theta_\tau)(\omega) = X_t(\theta_\tau(\omega)) = \theta_\tau(\omega)(t) = \omega(\tau(\omega)+t) = X_{\tau(\omega)+t}(\omega)$, i.e. $X_t \circ \theta_\tau = X_{\tau+t}$. So the operators θ_τ can still be viewed as time shifts, but now also certain random times are allowed.

Theorem 3.4.1 (Strong Markov property). *Let Z be an $\mathcal{F}_\infty$-measurable random variable, nonnegative or bounded. Then for every $(\mathcal{F}_t)$-stopping time τ and initial distribution ν, we have $\mathbb{P}_\nu$-a.s.*

$$\mathbb{E}_\nu(Z \circ \theta_\tau \,|\, \mathcal{F}_\tau) = \mathbb{E}_{X_\tau} Z \qquad \text{on } \{\tau < \infty\}.$$

Proof. Suppose first that τ is an a.s. finite stopping time that takes values in a countable set D. Then since θ_τ equals θ_d on the event $\{\tau = d\}$, we have (see Exercise 20 of Chapter 1)

$$\begin{aligned}
\mathbb{E}_\nu(Z \circ \theta_\tau \,|\, \mathcal{F}_\tau) &= \sum_{d\in D} 1_{\{\tau=d\}} \mathbb{E}_\nu(Z \circ \theta_\tau \,|\, \mathcal{F}_\tau) \\
&= \sum_{d\in D} 1_{\{\tau=d\}} \mathbb{E}_\nu(Z \circ \theta_d \,|\, \mathcal{F}_d) \\
&= \sum_{d\in D} 1_{\{\tau=d\}} \mathbb{E}_{X_d} Z \\
&= \mathbb{E}_{X_\tau} Z
\end{aligned}$$

$\mathbb{P}_\nu$-a.s., by the Markov property.

Let us now consider a general finite stopping time τ and assume that Z is of the form

$$Z = \prod_i f_i(X_{t_i})$$

for certain $t_1 < \cdots < t_k$ and $f_1, \ldots, f_k \in C_0$. By Lemma 1.6.14 there exist stopping times τ_n taking values in a finite set and decreasing to τ. By the preceding paragraph we $\mathbb{P}_\nu$-a.s. have

$$\mathbb{E}_\nu\Big(\prod_i f_i(X_{t_i}) \circ \theta_{\tau_n} \,|\, \mathcal{F}_{\tau_n}\Big) = \mathbb{E}_{X_{\tau_n}} \prod_i f_i(X_{t_i}) = g(X_{\tau_n}),$$

where

$$g(x) = P_{t_1} f_1 P_{t_2-t_1} f_2 \cdots P_{t_k-t_{k-1}} f_k(x).$$

By the right-continuity of the paths the right-hand side converges a.s. to $g(X_\tau)$ as $n \to \infty$. The right-continuity of the filtration and Corollary 2.2.16 imply that the left-hand side converges a.s. to

$$\mathbb{E}_\nu\Big(\prod_i f_i(X_{t_i}) \circ \theta_\tau \,|\, \mathcal{F}_\tau\Big).$$

Hence, the statement of the theorem holds for $Z = \prod_i f_i(X_{t_i})$. By a monotone class type argument we see that it holds for all $\mathcal{F}_\infty^X$-measurable random variables Z.

Next, let Z be a general $\mathcal{F}_\infty$-measurable random variable. Denote the distribution of X_τ under $\mathbb{P}_\nu$ by μ. By construction, $\mathcal{F}_\infty$ is contained in the $\mathbb{P}_\mu$-completion of $\mathcal{F}_\infty^X$ so there exist two $\mathcal{F}_\infty^X$-measurable random variables Z' and Z'' such that $Z' \leq Z \leq Z''$ and $\mathbb{E}_\mu(Z'' - Z') = 0$. It follows that $Z' \circ \theta_\tau \leq Z \circ \theta_\tau \leq Z'' \circ \theta_\tau$ and by the preceding paragraph

$$\begin{aligned} \mathbb{E}_\nu(Z'' \circ \theta_\tau - Z' \circ \theta_\tau) &= \mathbb{E}_\nu \mathbb{E}_\nu(Z'' \circ \theta_\tau - Z' \circ \theta_\tau \,|\, \mathcal{F}_\tau) \\ &= \mathbb{E}_\nu \mathbb{E}_{X_\tau}(Z'' - Z') \\ &= \int \mathbb{E}_x(Z'' - Z')\,\mu(dx) \\ &= \mathbb{E}_\mu(Z'' - Z') = 0. \end{aligned}$$

So $Z \circ \theta_\tau$ is measurable with respect to the $\mathbb{P}_\nu$-completion of $\mathcal{F}_\infty^X$, and since ν is arbitrary we conclude that $Z \circ \theta_\tau$ is $\mathcal{F}_\infty$-measurable. Now observe that we have

$$\mathbb{E}_\nu(Z' \circ \theta_\tau \,|\, \mathcal{F}_\tau) \leq \mathbb{E}_\nu(Z \circ \theta_\tau \,|\, \mathcal{F}_\tau) \leq \mathbb{E}_\nu(Z'' \circ \theta_\tau \,|\, \mathcal{F}_\tau)$$

$\mathbb{P}_\nu$-a.s. By the preceding paragraph the outer terms terms $\mathbb{P}_\nu$-a.s. equal $\mathbb{E}_{X_\tau} Z'$ and $\mathbb{E}_{X_\tau} Z''$, respectively. The preceding calculation shows that these random variables are $\mathbb{P}_\nu$-a.s. equal. Since $Z' \leq Z \leq Z''$ it follows that they are both a.s. equal to $\mathbb{E}_{X_\tau} Z$.

We have now shown that the theorem holds for finite stopping times. If τ is not finite, apply the result to the finite stopping time $\sigma = \tau 1_{\{\tau<\infty\}}$ and verify that $\mathbb{P}_\nu$-a.s., $\mathbb{E}_\nu(V 1_{\{\tau<\infty\}} \,|\, \mathcal{F}_\sigma) = \mathbb{E}_\nu(V 1_{\{\tau<\infty\}} \,|\, \mathcal{F}_\tau)$ for every bounded or nonnegative random variable V. The result then easily follows. □

We can say more for Feller processes with stationary and independent increments, i.e. for which the increment $X_t - X_s$ is independent of $\mathcal{F}_s$ for all $s \leq t$ and its $\mathbb{P}_\nu$- distribution depends only on $t - s$ for all initial distributions ν. For such processes, which are called *Lévy processes*, we have the following corollary.

Corollary 3.4.2. *If X has stationary, independent increments and τ is a finite stopping time, the process $(X_{\tau+t} - X_\tau)_{t\geq 0}$ is independent of $\mathcal{F}_\tau$ and under each $\mathbb{P}_\nu$, it has the same law as X under $\mathbb{P}_0$, provided that $0 \in E$.*

Proof. Put $Y_t = X_{\tau+t} - X_\tau$ for $t \geq 0$. For $t_1 < \cdots < t_n$ and measurable, nonnegative functions $f_1, \ldots, f_n$ we have

$$\begin{aligned} \mathbb{E}_\nu\Big(\prod_k f_k(Y_{t_k}) \,|\, \mathcal{F}_\tau\Big) &= \mathbb{E}_\nu\Big(\prod_k f_k(X_{t_k} - X_0) \circ \theta_\tau \,|\, \mathcal{F}_\tau\Big) \\ &= \mathbb{E}_{X_\tau} \prod_k f_k(X_{t_k} - X_0), \end{aligned}$$

by the strong Markov property. Hence, the proof is complete once we have shown that for arbitrary $x \in E$

$$\mathbb{E}_x \prod_{k=1}^n f_k(X_{t_i} - X_0) = P_{t_1} f_1 \cdots P_{t_n - t_{n-1}} f_n(0)$$

(see Lemma 3.1.4). We prove this by induction on n. Suppose first that $n = 1$. Then by the stationarity of the increments, the distribution of $X_{t_1} - X_0$ under $\mathbb{P}_x$ is independent of x. In particular we can take $x = 0$, obtaining

$$\mathbb{E}_x f_1(X_{t_1} - X_0) = \mathbb{E}_0 f_1(X_{t_1}) = P_{t_1} f_1(0).$$

Now suppose that the statement is true for $n-1$ and all nonnegative, measurable functions $f_1, \ldots, f_{n-1}$. We have

$$\begin{aligned}
\mathbb{E}_x \prod_{i=1}^n f_i(X_{t_i} - X_0) &= \mathbb{E}_x \mathbb{E}_x \Big(\prod_{i=1}^n f_i(X_{t_i} - X_0) \,|\, \mathcal{F}_{t_{n-1}} \Big) \\
&= \mathbb{E}_x \Big(\prod_{i=1}^{n-1} f_i(X_{t_i} - X_0) \mathbb{E}_x (f_n(X_{t_n} - X_0) \,|\, \mathcal{F}_{t_{n-1}}) \Big).
\end{aligned}$$

By the independence of the increments

$$\begin{aligned}
\mathbb{E}_x (f_n(X_{t_n} - X_0) \,|\, \mathcal{F}_{t_{n-1}}) &= \mathbb{E}_x (f_n(X_{t_n} - X_{t_{n-1}} + X_{t_{n-1}} - X_0) \,|\, \mathcal{F}_{t_{n-1}}) \\
&= g_x(X_{t_{n-1}} - X_0),
\end{aligned}$$

where

$$g_x(y) = \mathbb{E}_x f_n(X_{t_n} - X_{t_{n-1}} + y).$$

The $\mathbb{P}_x$-distribution of $X_{t_n} - X_{t_{n-1}}$ is the same as the distribution of $X_{t_n - t_{n-1}} - X_0$ and is independent of x. It follows that

$$g_x(y) = \mathbb{E}_x f_n(X_{t_n - t_{n-1}} - X_0 + y) = \mathbb{E}_y f_n(X_{t_n - t_{n-1}}) = P_{t_n - t_{n-1}} f_n(y),$$

so that

$$\begin{aligned}
\mathbb{E}_x \prod_{i=1}^n f_i(X_{t_i} - X_0) &= \mathbb{E}_x \Big(\prod_{i=1}^{n-1} f_i(X_{t_i} - X_0) P_{t_n - t_{n-1}} f_n(X_{t_{n-1}} - X_0) \Big) \\
&= \mathbb{E}_x \Big(\prod_{i=1}^{n-2} f_i(X_{t_i} - X_0) \Big) (f_{n-1} P_{t_n - t_{n-1}} f_n)(X_{t_{n-1}} - X_0).
\end{aligned}$$

By the induction hypothesis this equals $P_{t_1} f_1 \cdots P_{t_n - t_{n-1}} f_n(0)$ and the proof is complete. □

The following lemma is often useful in connection with the strong Markov property.

Lemma 3.4.3. *If σ and τ are finite $(\mathcal{F}_t)$-stopping times, then $\sigma + \tau \circ \theta_\sigma$ is also a finite $(\mathcal{F}_t)$-stopping time.*

Proof. Since $(\mathcal{F}_t)$ is right-continuous, it suffices to prove that $\{\sigma + \tau \circ \theta_\sigma < t\} \in \mathcal{F}_t$ for every $t > 0$. Observe that

$$\{\sigma + \tau \circ \theta_\sigma < t\} = \bigcup_{\substack{q \geq 0 \\ q \in \mathbb{Q}}} \{\tau \circ \theta_\sigma < q\} \cap \{\sigma \leq t - q\}.$$

The indicator of $\{\tau \circ \theta_\sigma < q\}$ can be written as $1_{\{\tau < q\}} \circ \theta_\sigma$. By Exercise 14, it follows that $\{\tau \circ \theta_\sigma < q\} \in \mathcal{F}_{\sigma+q}$. Hence, by definition of the latter σ-algebra, it holds that

$$\{\tau \circ \theta_\sigma < q\} \cap \{\sigma \leq t - q\} = \{\tau \circ \theta_\sigma < q\} \cap \{\sigma + q \leq t\} \in \mathcal{F}_t.$$

This completes the proof. □

3.4.2 Applications to Brownian motion

In this section W is the canonical version of the Brownian motion, and $(\mathcal{F}_t)$ is the usual augmentation of its natural filtration. Since the BM has stationary, independent increments, Corollary 3.4.2 implies that for every $(\mathcal{F}_t)$-stopping time τ, the process $(W_{\tau+t} - W_\tau)_{t \geq 0}$ is a standard BM. The first application that we present is the so-called reflection principle (compare with Exercise 10 of Chapter 1).

Recall that we denote the hitting time of $x \in \mathbb{R}$ by τ_x. This is a finite stopping time with respect to the natural filtration of the BM (see Example 1.6.9) so it is certainly an $(\mathcal{F}_t)$-stopping time.

Theorem 3.4.4 (Reflection principle). *Let W be a Brownian motion and for $x \in \mathbb{R}$, let τ_x be the first hitting time of x. Define the process W' by*

$$W'_t = \begin{cases} W_t, & \text{if } t \leq \tau_x \\ 2x - W_t, & \text{if } t > \tau_x. \end{cases}$$

Then W' is a standard Brownian motion.

Proof. Define the processes Y and Z by $Y = W^{\tau_x}$ and $Z_t = W_{\tau_x+t} - W_{\tau_x} = W_{\tau_x+t} - x$ for $t \geq 0$. By Corollary 3.4.2 the processes Y and Z are independent and Z is a standard BM. By the symmetry of the BM it follows that $-Z$ is also a BM that is independent of Y, so the two pairs (Y, Z) and $(Y, -Z)$ have the same distribution. Now observe that for $t \geq 0$,

$$W_t = Y_t + Z_{t-\tau_x} 1_{\{t>\tau_x\}}, \quad W'_t = Y_t - Z_{t-\tau_x} 1_{\{t>\tau_x\}}.$$

In other words we have $W = \varphi(Y, Z)$ and $W' = \varphi(Y, -Z)$, where $\varphi : C[0,\infty) \times C_0[0,\infty) \to C[0,\infty)$ is given by

$$\varphi(y, z)(t) = y(t) + z(t - \psi(y)) 1_{\{t>\psi(y)\}},$$

and $\psi : C[0,\infty) \to [0,\infty]$ is defined by $\psi(y) = \inf\{t > 0 : y(t) = x\}$. Now it is easily verified that ψ is a Borel measurable map, so φ is also measurable, being a composition of measurable maps (see Exercise 17). Since $(Y, Z) =_d (Y, -Z)$, it follows that $W = \varphi(Y, Z) =_d \varphi(Y, -Z) = W'$. □

The refection principle allows us to calculate the distributions of certain functionals related to the hitting times of the BM. We first consider the joint distribution of W_t and the running maximum

$$S_t = \sup_{s \le t} W_s.$$

Corollary 3.4.5. *Let W be a standard BM and S its running maximum. Then for $x \le y$,*

$$\mathbb{P}(W_t \le x, S_t \ge y) = \mathbb{P}(W_t \le x - 2y).$$

The pair (W_t, S_t) has joint density

$$(x, y) \mapsto \frac{(2y - x)e^{-\frac{(2y-x)^2}{2t}}}{\sqrt{\pi t^3/2}} 1_{\{x \le y\}}$$

with respect to Lebesgue measure.

Proof. Let W' be the process obtained by reflecting W at the hitting time τ_y. Observe that $S_t \ge y$ if and only if $t \ge \tau_y$, so the probability of interest equals $\mathbb{P}(W_t \le x, t \ge \tau_y)$. On the event $\{t \ge \tau_y\}$ we have $W_t = 2y - W'_t$, so it reduces further to $\mathbb{P}(W'_t \ge 2y - x, t \ge \tau_y)$. But since $x \le y$ we have $2y - x \ge y$, hence $\{W'_t \ge 2y - x\} \subseteq \{W'_t \ge y\} \subseteq \{t \ge \tau_y\}$. It follows that $\mathbb{P}(W'_t \ge 2y - x, t \ge \tau_y) = \mathbb{P}(W'_t \ge 2y - x)$. By the reflection principle and the symmetry of the BM this proves the first statement. For the second statement, see Exercise 18. □

It follows from the preceding corollary that for all $x > 0$ and $t \ge 0$,

$$\mathbb{P}(S_t \ge x) = \mathbb{P}(\tau_x \le t) = 2\mathbb{P}(W_t \ge x) = \mathbb{P}(|W_t| \ge x)$$

(see Exercise 19). This shows in particular that $S_t =_d |W_t|$ for every $t \ge 0$, and we can now derive an explicit expression for the density of the hitting time τ_x. It is easily seen from this expression that $\mathbb{E}\tau_x = \infty$, as was proved by martingale methods in Exercise 15 of Chapter 2.

Corollary 3.4.6. *The first time τ_x that the standard BM hits the level $x > 0$ has density*

$$t \mapsto \frac{xe^{-\frac{x^2}{2t}}}{\sqrt{2\pi t^3}} 1_{\{t \ge 0\}}$$

with respect to Lebesgue measure.

Proof. See Exercise 20. □

We have seen in the first two chapters that the zero set of the standard BM is a.s. closed, unbounded, has Lebesgue measure zero and that 0 is an accumulation point of the set, i.e. 0 is not an isolated point. Using the strong Markov property we can prove that in fact, the zero set contains no isolated points at all.

Corollary 3.4.7. *The zero set $Z = \{t \geq 0 : W_t = 0\}$ of the standard BM is a.s. closed, unbounded, contains no isolated points and has Lebesgue measure zero.*

Proof. In view of Exercise 26 of Chapter 1 we only have to prove that Z contains no isolated points. For rational $q \geq 0$, define $\sigma_q = q + \tau_0 \circ \theta_q$ and observe that σ_q is the first time after time q that the BM W visits 0. By Lemma 3.4.3 the random time σ_q is a stopping time, hence the strong Markov property implies that the process $W_{\sigma_q+t} = W_{\sigma_q+t} - W_{\sigma_q}$ is a standard BM. By Corollary 2.4.6 it follows that a.s., σ_q is an accumulation point of Z, hence with probability 1 it holds that for every rational $q \geq 0$, σ_q is an accumulation point of Z. Now take an arbitrary point $t \in Z$ and choose rational points q_n such that $q_n \uparrow t$. Then since $q_n \leq \sigma_{q_n} \leq t$, we have $\sigma_{q_n} \to t$. The limit of accumulation points is also an accumulation point, so the proof is complete. □

3.5 Generators

3.5.1 Generator of a Feller process

Throughout this section we consider the 'good version' of a Feller process X with state space $E \subseteq \mathbb{R}^d$, transition semigroup (P_t) and resolvents R_λ.

The transition function and the resolvents give little intuition about the way in which the Markov process X moves from point to point as time evolves. In the following definition we introduce a new operator which captures the behaviour of the process in infinitesimally small time intervals.

Definition 3.5.1. A function $f \in C_0$ is said to belong to the domain $\mathcal{D}(A)$ of the infinitesimal generator of X if the limit

$$Af = \lim_{t \downarrow 0} \frac{P_t f - f}{t}$$

exists in C_0. The linear transformation $A : \mathcal{D}(A) \to C_0$ is called the *infinitesimal generator* of the process.

Immediately from the definition we see that for $f \in \mathcal{D}(A)$, it holds $\mathbb{P}_\nu$-a.s. holds that

$$\mathbb{E}_\nu(f(X_{t+h}) - f(X_t) \,|\, \mathcal{F}_t) = hAf(X_t) + o(h)$$

as $h \downarrow 0$. So in this sense, the generator indeed describes the movement of the process in an infinitesimally small amount of time.

Lemma 3.3.5 gives conditions under which a function φ of the Feller process X is again Feller. Immediately from the definition we see that the generators of the two processes are related as follows.

Lemma 3.5.2. *Let $\varphi : E \to E'$ be continuous and onto, and assume that $\|\varphi(x_n)\| \to \infty$ in E' if and only if $\|x_n\| \to \infty$ in E. Suppose that (Q_t) is a collection of transition kernels such that $P_t(f \circ \varphi) = (Q_t f) \circ \varphi$ for all $f \in C_0(E')$, so that $Y = \varphi(X)$ is a Feller process with state space E' and transition function (Q_t). Then the generator B of Y satisfies $\mathcal{D}(B) = \{f \in C_0(E') : f \circ \varphi \in \mathcal{D}(A)\}$ and $A(f \circ \varphi) = (Bf) \circ \varphi$ for $f \in \mathcal{D}(B)$.*

Proof. See Exercise 21. □

The following lemma gives some of the basic properties of the generator. The differential equations in part (ii) are *Kolmogorov's forward and backward equations.*

Lemma 3.5.3. *Suppose that $f \in \mathcal{D}(A)$.*

(i) For all $t \geq 0$, $P_t\mathcal{D}(A) \subseteq \mathcal{D}(A)$.

(ii) The function $t \mapsto P_t f$ is right-differentiable in C_0 and

$$\frac{d}{dt} P_t f = AP_t f = P_t Af.$$

More precisely, this means that

$$\lim_{h \downarrow 0} \left\| \frac{P_{t+h}f - P_t f}{h} - P_t Af \right\|_\infty = \lim_{h \downarrow 0} \left\| \frac{P_{t+h}f - P_t f}{h} - AP_t f \right\|_\infty = 0.$$

(iii) We have

$$P_t f - f = \int_0^t P_s Af \, ds = \int_0^t AP_s f \, ds$$

for every $t \geq 0$.

Proof. All properties follow from the semigroup property and the strong continuity of (P_t), see Exercise 22. To prove (iii) it is useful to note that since the Feller process X with semigroup (P_t) has a right-continuous and quasi left-continuous modification, the map $t \mapsto P_t f(x)$ is continuous for all $f \in C_0$ and $x \in E$. □

The next theorem gives a full description of the generator in terms of the resolvents R_λ. In the proof we need the following lemma.

Lemma 3.5.4. *For $h, s > 0$ define the linear operators*

$$A_h f = \frac{1}{h}(P_h f - f), \quad B_s f = \frac{1}{s}\int_0^s P_t f\,dt.$$

Then $B_h f \in \mathcal{D}(A)$ for all $h > 0$ and $f \in C_0$, and $AB_h = A_h$.

Proof. It is easily verified that $A_h B_s = A_s B_h$ for all $s, h > 0$. Also note that for $f \in C_0$,

$$\|B_h f - f\|_\infty \le \frac{1}{h}\int_0^h \|P_t f - f\|_\infty\,dt \to 0$$

as $h \downarrow 0$, by the strong continuity of the Feller semigroup. It follows that for $s > 0$ and $f \in C_0$,

$$A_h B_s f = A_s B_h f \to A_s f$$

in C_0 as $h \downarrow 0$, and the proof is complete. □

Theorem 3.5.5. *The domain $\mathcal{D}(A)$ equals the joint image of the resolvents, whence $\mathcal{D}(A)$ is dense in C_0. For every $\lambda > 0$ the transformation $\lambda I - A : \mathcal{D}(A) \to C_0$ is invertible and its inverse is R_λ.*

Proof. For $f \in \mathcal{D}(A)$ we have

$$\begin{aligned} R_\lambda(\lambda f - Af) &= \int_0^\infty e^{-\lambda t} P_t(\lambda f - Af)\,dt \\ &= \lambda \int_0^\infty e^{-\lambda t} P_t f\,dt - \int_0^\infty e^{-\lambda t}\left(\frac{d}{dt}P_t f\right)\,dt, \end{aligned}$$

by Lemma 3.5.3. Integration by parts of the last integral shows that $R_\lambda(\lambda f - Af) = f$. We conclude that $\mathcal{D}(A) \subseteq \mathrm{Im} R_\lambda$ and $R_\lambda(\lambda I - A) = I$ on $\mathcal{D}(A)$.

To prove the converse, we use the notation of the preceding lemma. By the lemma and the fact that P_h and R_λ commute, we have for $f \in C_0$ and $h > 0$,

$$A_h R_\lambda f = R_\lambda A_h f = R_\lambda A B_h f = \int_0^\infty e^{-\lambda t} P_t A B_h f\,dt.$$

Using Lemma 3.5.3 and integration by parts we see that the right-hand side equals

$$\int_0^\infty e^{-\lambda t}\left(\frac{d}{dt}P_t B_h f\right)\,dt = \lambda R_\lambda B_h f - B_h f,$$

so we have $A_h R_\lambda f = \lambda R_\lambda B_h f - B_h f$. In the proof of the preceding lemma we saw that as $h \to 0$, $B_h f \to f$, uniformly. This shows that $R_\lambda f \in \mathcal{D}(A)$ and $AR_\lambda f = \lambda R_\lambda f - f$. It follows that $\mathrm{Im} R_\lambda \subseteq \mathcal{D}(A)$ and $(\lambda I - A)R_\lambda = I$ on C_0. □

The theorem states that for $\lambda > 0$, it holds that $R_\lambda = (\lambda I - A)^{-1}$, so the generator A determines the resolvents. By uniqueness of Laplace transforms the resolvents determine the semigroup (P_t). This shows that the generator determines the semigroup. In other words, two Feller processes with the same generator have the same semigroup.

Corollary 3.5.6. *The generator determines the semigroup.*

The preceding theorem also shows that for all $\lambda > 0$, the generator is given by $A = \lambda I - R_\lambda^{-1}$. This gives us a method for explicit computations.

Example 3.5.7. For the standard BM we have that $\mathcal{D}(A) = C_0^2$, the space of functions on $\mathbb{R}$ with two continuous derivatives that vanish at infinity. For $f \in \mathcal{D}(A)$ it holds that that $Af = f''/2$ (see Exercise 23). ■

From the definition of the generator it is easy to see that if $f \in \mathcal{D}(A)$ and $x \in E$ is such that $f(y) \leq f(x)$ for all $y \in E$, then $Af(x) \leq 0$. When an operator has this property, we say it satisfies the *maximum principle*.

The following result will be useful below.

Theorem 3.5.8. *Suppose the generator A extends to $A' : D' \to C_0$, where $D' \subseteq C_0$ is a linear space, and A' satisfies the maximum principle. Then $D' = \mathcal{D}(A)$.*

Proof. In the first part of the proof we show that for a linear space $D \subseteq C_0$, an operator $B : D \to C_0$ that satisfies the maximum principle is dissipative, i.e.

$$\lambda \|f\|_\infty \leq \|(\lambda I - B)f\|_\infty$$

for all $f \in D$ and $\lambda > 0$. To prove this, let $x \in E$ be such that $|f(x)| = \|f\|_\infty$ and define $g(y) = f(y)\mathrm{sgn}f(x)$. Then $g \in D$ (D is a linear space) and $g(y) \leq g(x)$ for all y. Hence, by the maximum principle, $Bg(x) \leq 0$. It follows that

$$\lambda \|f\|_\infty = \lambda g(x) \leq \lambda g(x) - Bg(x) \leq \|(\lambda I - B)g\|_\infty = \|(\lambda I - B)f\|_\infty,$$

as claimed.

For the proof of the theorem, take $f \in D'$ and define $g = (I - A')f$. By the first part of the proof A' is dissipative, hence

$$\|f - R_1 g\|_\infty \leq \|(I - A')(f - R_1 g)\|_\infty.$$

By Theorem 3.5.5 we have $(I - A)R_1 = I$ on C_0, so $(I - A')(f - R_1 g) = g - (I - A)R_1 g = 0$. It follows that $f = R_1 g$, whence $f \in \mathcal{D}(A)$. □

Generators provide an important link between Feller processes and martingales.

Theorem 3.5.9. *For every $f \in \mathcal{D}(A)$ and initial measure ν, the process*

$$M_t^f = f(X_t) - f(X_0) - \int_0^t Af(X_s)\,ds$$

is a $\mathbb{P}_\nu$-martingale.

Proof. Since f and Af are in C_0 and therefore bounded, M_t^f is integrable for every $t \geq 0$. Now for $s \leq t$,

$$\mathbb{E}_\nu(M_t^f \,|\, \mathcal{F}_s) = M_s^f + \mathbb{E}_\nu\left(f(X_t) - f(X_s) - \int_s^t Af(X_u)\,du \,\middle|\, \mathcal{F}_s\right).$$

By the Markov property, the conditional expectation on the right-hand side equals

$$\mathbb{E}_{X_s}\left(f(X_{t-s}) - f(X_0) - \int_0^{t-s} Af(X_u)\,du\right).$$

But for every $x \in E$,

$$\begin{aligned}
&\mathbb{E}_x\left(f(X_{t-s}) - f(X_0) - \int_0^{t-s} Af(X_u)\,du\right)\\
&\qquad = P_{t-s}f(x) - f(x) - \int_0^{t-s} P_uAf(x)\,du = 0,
\end{aligned}$$

by Fubini's theorem and part (iii) of Lemma 3.5.3. □

3.5.2 Characteristic operator

The definition of the generator and its expression in terms of resolvents given by Theorem 3.5.5 are analytical in nature. In the present section we give a probabilistic description of the generator. The key is the following corollary of Theorem 3.5.9.

Corollary 3.5.10 (Dynkin's formula). *For every $f \in \mathcal{D}(A)$ and every stopping time τ such that $\mathbb{E}_x\tau < \infty$, we have*

$$\mathbb{E}_x f(X_\tau) = f(x) + \mathbb{E}_x \int_0^\tau Af(X_s)\,ds$$

for every $x \in E$.

Proof. By Theorem 3.5.9 and the optional stopping theorem, we have

$$\mathbb{E}_x f(X_{\tau\wedge n}) = f(x) + \mathbb{E}_x \int_0^{\tau\wedge n} Af(X_s)\,ds$$

for every $n \in \mathbb{N}$. By the quasi-left continuity of X (see Exercise 16), the left-hand side converges to $\mathbb{E}_x f(X_\tau)$ as $n \to \infty$. Since $Af \in C_0$ we have $\|Af\|_\infty < \infty$ and

$$\left| \int_0^{\tau \wedge n} Af(X_s)\, ds \right| \leq \|Af\|_\infty \tau.$$

Hence, by the fact that $\mathbb{E}_x \tau < \infty$, the dominated convergence theorem implies that the integral on the right-hand side of the first display converges to

$$\mathbb{E}_x \int_0^\tau Af(X_s)\, ds.$$

This completes the proof. □

We call a point $x \in E$ *absorbing* if for all $t \geq 0$ it holds that $P_t(x, \{x\}) = 1$. This means that if the process is started in x, it never leaves x (see also Exercise 13). For $r > 0$, define the stopping time

$$\eta_r = \inf\{t \geq 0 : \|X_t - X_0\| > r\}. \tag{3.6}$$

If x is absorbing, then clearly we $\mathbb{P}_x$-a.s. have $\eta_r = \infty$ for all $r > 0$. For non-absorbing points however, the *escape time* η_r is a.s. finite and has a finite mean if r is small enough.

Lemma 3.5.11. *If $x \in E$ is not absorbing, then $\mathbb{E}_x \eta_r < \infty$ for $r > 0$ sufficiently small.*

Proof. Let $B(x, \varepsilon) = \{y : \|y - x\| \leq \varepsilon\}$ be the closed ball of radius ε around the point x. If x is not absorbing, it holds that $P_t(x, B(x, \varepsilon)) < p < 1$ for some $t, \varepsilon > 0$. By the Feller property of the semigroup (P_t) we have the weak convergence $P_t(y, \cdot) \Rightarrow P_t(x, \cdot)$ as $y \to x$. Hence, by the portmanteau theorem, the fact that $B(x, \varepsilon)$ is closed implies that

$$\limsup_{y \to x} P_t(y, B(x, \varepsilon)) \leq P_t(x, B(x, \varepsilon)).$$

It follows that for all y close enough to x, say $y \in B(x, r)$ for $r \in (0, \varepsilon)$, we have $P_t(y, B(x, r)) \leq P_t(x, B(x, \varepsilon)) < p$. Using the Markov property it can easily be shown that consequently, $\mathbb{P}_x(\eta_r \geq nt) \leq p^n$ for $n = 0, 1, \ldots$ (see Exercise 24). Hence,

$$\mathbb{E}_x \eta_r = \int_0^\infty \mathbb{P}_x(\eta_r \geq s)\, ds \leq t \sum_{n=0}^\infty \mathbb{P}_x(\eta_r \geq nt) \leq \frac{t}{1-p} < \infty.$$

This completes the proof of the lemma. □

We can now prove the following alternative description of the generator.

Theorem 3.5.12. *For $f \in \mathcal{D}(A)$ we have $Af(x) = 0$ if x is absorbing, and*

$$Af(x) = \lim_{r \downarrow 0} \frac{\mathbb{E}_x f(X_{\eta_r}) - f(x)}{\mathbb{E}_x \eta_r} \tag{3.7}$$

otherwise.

Proof. If x is absorbing we have $P_t f(x) = f(x)$ for all $t \geq 0$, which shows that $Af(x) = 0$. For non-absorbing $x \in E$ the stopping time η_r has finite mean for r small enough (see the lemma), so by Dynkin's formula (Corollary 3.5.10),

$$\mathbb{E}_x f(X_{\eta_r}) = f(x) + \mathbb{E}_x \int_0^{\eta_r} Af(X_s)\, ds.$$

It follows that

$$\begin{aligned}\Big|\frac{\mathbb{E}_x f(X_{\eta_r}) - f(x)}{\mathbb{E}_x \eta_r} - Af(x)\Big| &\leq \frac{\mathbb{E}_x \int_0^{\eta_r} |Af(X_s) - Af(x)|\, ds}{\mathbb{E}_x \eta_r} \\ &\leq \sup_{\|y-x\| \leq r} |Af(y) - Af(x)|.\end{aligned}$$

This completes the proof, since Af is continuous. □

The operator defined by the right-hand side of (3.7) is called the *characteristic operator* of the Markov process X, its domain is simply the collection of all functions $f \in C_0$ for which the limit in (3.7) exists. The theorem states that for Feller processes, the characteristic operator extends the infinitesimal generator. It is easily seen that the characteristic operator satisfies the maximum principle. Hence, by Theorem 3.5.8, the characteristic operator and the generator coincide.

3.6 Killed Feller processes

3.6.1 Sub-Markovian processes

Up to this point we have always assumed that the transition function (P_t) satisfies $P_t(x, E) = 1$, i.e. that the $P_t(x, \cdot)$ are probability measures. It is sometimes useful to consider semigroups for which $P_t(x, E) < 1$ for some t, x. We call the semigroup *sub-Markovian* in that case. Intuitively, a sub-Markovian semigroup describes the motion of a particle that can disappear from the state space E, or *die*, in a finite time. A sub-Markovian transition function can be turned into a true Markovian one by adjoining a new point Δ to the state space E, called the *cemetery*. We then put $E_\Delta = E \cup \{\Delta\}$, $\mathcal{E}_\Delta = \sigma(\mathcal{E}, \{\Delta\})$ and define a new semigroup $(\tilde{P}_t)$ by setting $\tilde{P}_t(x, A) = P_t(x, A)$ if $A \subseteq E$, $\tilde{P}_t(x, \{\Delta\}) = 1 - P_t(x, E)$ and $\tilde{P}_t(\Delta, \{\Delta\}) = 1$ (see Exercise 25). Note that by construction, the point Δ is absorbing for the new process. By convention, all functions on E are extended to E_Δ by putting $f(\Delta) = 0$. In the sequel we will not distinguish between P_t and $\tilde{P}_t$ in our notation.

In the case of sub-Markovian Feller semigroups, we extend the topology of E to E_Δ in such a way that E_Δ is the one-point compactification of E if E is not compact, and Δ is an isolated point otherwise. Then the result of Theorem 3.3.7 is still true for the process on E_Δ, provided that we call an E_Δ-valued function on $\mathbb{R}_+$ cadlag if it is right-continuous and has left limits with respect to this topology on E_Δ. Note that by construction, a function $f \in C_0(E_\Delta)$ satisfies $f(\Delta) = 0$. As before, we always consider the cadlag version of a given Feller process.

We already noted that Δ is absorbing by construction. So once a process has arrived in the cemetery state, it stays there for the remaining time. We define the *killing time* of the Feller process X by

$$\zeta = \inf\{t \geq 0 : X_{t-} = \Delta \text{ or } X_t = \Delta\}.$$

Clearly $\zeta < \infty$ with positive probability if the process X is sub-Markovian. For a cadlag Feller process X we have the following strong version of the absorption property of the cemetery.

Lemma 3.6.1. *It $\mathbb{P}_\nu$-almost surely holds that $X_t = \Delta$ for $t \geq \zeta$, for every initial distribution ν.*

Proof. Let $f \in C_0(E_\Delta)$ be strictly positive on E. Then the right-continuous $\mathbb{P}_\nu$-supermartingale $M_t = \exp(-t)R_1 f(X_t)$ (see Lemma 3.3.6) is nonnegative, and clearly $M_t = 0$ if and only if $X_t = \Delta$ and $M_{t-} = 0$ if and only if $X_{t-} = \Delta$. Hence, the result follows from Corollary 2.3.16. □

We close this subsection with an example of a sub-Markovian Feller semigroup, given by a Brownian motion which is killed when it first hits zero.

Example 3.6.2. Let W be a BM started at some positive point and let τ_0 be the first time it hits zero. Define a new process X by putting $X_t = W_t$ for $t < \tau_0$, and $X_t = \Delta$ for $t \geq \tau_0$. We can view the process X as a sub-Markovian Feller process with state space $(0, \infty)$. Moreover, we can compute its transition semigroup (P_t) explicitly. For $f \in C_0(0, \infty)$ and $x > 0$ we have

$$P_t f(x) = \mathbb{E}_x f(X_t) = \mathbb{E}_x f(W_t) 1_{\{\tau_0 > t\}}.$$

Now observe that under $\mathbb{P}_x$, the pair (W, τ_0) has the same distribution as the pair $(x + W, \tau_{-x})$ under $\mathbb{P}_0$. By symmetry of the BM, the latter pair has the same distribution as $(x - W, \tau_x)$ under $\mathbb{P}_0$. Noting also that $\{\tau_0 > t\} = \{S_t < x\}$, where $S_t = \sup_{s \leq t} W_s$, we find that

$$P_t f(x) = \mathbb{E}_0 f(x - W_t) 1_{\{S_t < x\}}.$$

The joint distribution of (W_t, S_t) under $\mathbb{P}_0$ is given by Corollary 3.4.5. We can use this to show that

$$P_t f(x) = \int_0^\infty f(y) p(t, x, y)\, dy,$$

where

$$p(t,x,y) = \frac{1}{\sqrt{2\pi t}}\Big(\exp\Big(-\frac{1}{2t}(y-x)^2\Big) - \exp\Big(-\frac{1}{2t}(y+x)^2\Big)\Big)$$

(see Exercise 26). ■

3.6.2 Feynman-Kac formula

Let X be a Feller process and τ an exponential random variable with parameter $q > 0$, independent of X. Then we can define a new process Y by

$$Y_t = \begin{cases} X_t, & t < \tau, \\ \Delta, & t \geq \tau. \end{cases}$$

So Y is constructed by killing X at an independent exponential time. Note that since $f \in C_0$ satisfies $f(\Delta) = 0$, we have

$$Q_t f(x) = \mathbb{E}_x f(Y_t) = \mathbb{E}_x f(X_t) 1_{\{t<\tau\}} = e^{-qt}\mathbb{E}_x f(X_t).$$

It follows that $(Q_t)_{t\geq 0}$ is a sub-Markovian Feller semigroup. Denote the semigroup of the original process X by (P_t) and its generator by A. Then for $f \in \mathcal{D}(A)$ we have

$$\frac{Q_t f - f}{t} = e^{-qt}\frac{P_t f - f}{t} + f\frac{e^{-qt}-1}{t} \to Af - qf.$$

So the generator A^q of the killed process Y is related to the original generator A by $A^q = A - qI$.

The preceding example extends to Feller processes which are killed at nonconstant rates. More precisely, one can prove that if q is a continuous, nonnegative function on the state space of the Feller process X, then

$$Q_t f(x) = \mathbb{E}_x f(X_t) e^{-\int_0^t q(X_u)\,du}$$

defines a (sub-Markovian) Feller semigroup. We think of this as the semigroup of a process obtained by killing X at the rate q. It is then not hard to show that as in the example above, the generator A^q of the killed process is given by $A^q = A - qI$. The λ-resolvent of the killed process, denoted by R^q_λ, is given by

$$R^q_\lambda f(x) = \mathbb{E}_x \int_0^\infty f(X_t) e^{-\lambda t - \int_0^t q(X_u)\,du}\,dt$$

for $f \in C_0$. Since $\lambda I - A^q$ is the inverse of R^q_λ (cf. Theorem 3.5.5), it follows that the function $u = R^q_\lambda f$ satisfies

$$(\lambda + q)u - Au = f.$$

If we apply the resolvent R_λ of the original process X to this equation we get the equivalent equation

$$u + R_\lambda q u = R_\lambda f.$$

Each of these formulas is called the *Feynman-Kac formula.*

In fact, the formula holds under less restrictive conditions. We can drop the continuity requirements on f and q.

Theorem 3.6.3 (Feynman-Kac formula). *Let X be a Feller process with λ-resolvent R_λ. Let q be a nonnegative, bounded measurable function on the state space and let f be bounded and measurable. Then the function*

$$u(x) = \mathbb{E}_x \int_0^\infty f(X_t) e^{-\lambda t - \int_0^t q(X_u)\,du}\,dt$$

satisfies

$$u + R_\lambda q u = R_\lambda f. \tag{3.8}$$

Proof. Define $A_t = \int_0^t q(X_u)\,du$. Then since $dA_t/dt = q(X_t)\exp(A_t)$, we have

$$\begin{aligned}
R_\lambda f(x) - u(x) &= \mathbb{E}_x \int_0^\infty f(X_t) e^{-\lambda t - A_t}\Big(e^{A_t} - 1\Big)\,dt \\
&= \mathbb{E}_x \int_0^\infty f(X_t) e^{-\lambda t - A_t}\Big(\int_0^t q(X_s) e^{A_s}\,ds\Big)\,dt \\
&= \mathbb{E}_x \int_0^\infty q(X_s) e^{A_s}\Big(\int_s^\infty f(X_t) e^{-\lambda t - A_t}\,dt\Big)\,ds.
\end{aligned}$$

Observe that A is an additive functional, i.e. $A_{t+s} - A_t = A_s \circ \theta_t$. Hence, a substitution $v = t - s$ yields

$$\begin{aligned}
R_\lambda f(x) - u(x) &= \mathbb{E}_x \int_0^\infty q(X_0 \circ \theta_s) e^{-\lambda s}\Big(\int_0^\infty f(X_v \circ \theta_s) e^{-\lambda v - A_v \circ \theta_s}\,dv\Big)\,ds \\
&= \mathbb{E}_x \int_0^\infty e^{-\lambda s}(Z \circ \theta_s)\,ds,
\end{aligned}$$

where

$$Z = q(X_0)\int_0^\infty f(X_t) e^{-\lambda t - \int_0^t q(X_u)\,du}\,dt.$$

By the Markov property, $\mathbb{E}_x(Z \circ \theta_s) = \mathbb{E}_x \mathbb{E}_{X_s} Z = \mathbb{E}_x q(X_s) u(X_s)$. Hence, we get

$$R_\lambda f(x) - u(x) = \mathbb{E}_x \int_0^\infty e^{-\lambda s}\mathbb{E}_x q(X_s) u(X_s)\,ds = R_\lambda q u(x).$$

This completes the proof. □

3.6.3 Feynman-Kac formula and arc-sine law for the BM

If we specialize the Feyman-Kac formula to the Brownian motion case, we get the following result.

Theorem 3.6.4 (Feynman-Kac formula for BM). *Let W be a standard BM, f a bounded, measurable function, q a nonnegative, bounded, measurable function and $\lambda > 0$. Then the function u defined by*

$$u(x) = \mathbb{E}_x \int_0^\infty f(W_t) e^{-\lambda t - \int_0^t q(X_u)\, du}\, dt$$

is twice differentiable, and satisfies the differential equation

$$(\lambda + q)u - \tfrac{1}{2}u'' = f.$$

Proof. By Example 3.3.4, the resolvent R_λ of the standard BM is given by

$$R_\lambda f(x) = \int_{\mathbb{R}} f(y) r_\lambda(x, y) dy,$$

where $r_\lambda(x, y) = \exp(-\sqrt{2\lambda}|x - y|)/\sqrt{(2\lambda)}$. This shows that if g is bounded and measurable, then $R_\lambda g$ is continuous. Moreover, a simple calculation shows that $R_\lambda g$ is also differentiable and

$$\frac{d}{dx} R_\lambda g(x) = \int_{y>x} g(y) e^{-\sqrt{2\lambda}(y-x)}\, dy - \int_{y<x} g(y) e^{-\sqrt{2\lambda}(x-y)}\, dy.$$

In particular, we see that $R_\lambda g$ is continuously differentiable and in fact twice differentiable. Differentiating once more we find that

$$\frac{d^2}{dx^2} R_\lambda g(x) = -2g(x) + 2\lambda R_\lambda g(x)$$

(see also Exercise 23). Under the conditions of the theorem the functions qu and f are clearly bounded. Hence, (3.8) implies that u is continuously differentiable and twice differentiable. The equation for u is obtained by differentiating (3.8) twice and using the identity in the last display. □

We now apply the Feynman-Kac formula to prove the arc-sine law for the amount of time that the BM spends above (or below) zero in a finite time interval. The name of this law is explained by the fact that the density in the statement of the theorem has distribution function

$$F(x) = \frac{2}{\pi} \arcsin \sqrt{\frac{x}{t}}, \quad x \in [0, t].$$

Theorem 3.6.5 (Arc-sine law). *Let W be a Brownian motion. For $t > 0$, the random variable*

$$\int_0^t 1_{(0,\infty)}(W_s)\, ds$$

has an absolutely continuous distribution with density

$$s \mapsto \frac{1}{\pi\sqrt{s(t-s)}}$$

on $(0, t)$.

Proof. For $\lambda, \omega > 0$, consider the double Laplace transform

$$u(x) = \int_0^\infty e^{-\lambda t} \mathbb{E}_x e^{-\omega \int_0^t 1_{(0,\infty)}(W_s)\, ds}\, dt.$$

By the Feynman-Kac formula (applied with $f = 1$, $q = \omega 1_{(0,\infty)}$) the function u is continuously differentiable and

$$\begin{aligned} (\lambda + \omega) u(x) - \tfrac{1}{2} u''(x) &= 1, \quad x > 0, \\ \lambda u(x) - \tfrac{1}{2} u''(x) &= 1, \quad x < 0. \end{aligned}$$

For $a > 0$ the general solution of the equation $ag - g''/2 = 1$ on an interval is given by

$$C_1 e^{x\sqrt{2a}} + C_2 e^{-x\sqrt{2a}} + \frac{1}{a}.$$

Since u is bounded, it follows that for certain constants C_1, C_2,

$$u(x) = \begin{cases} C_1 e^{-x\sqrt{2(\lambda+\omega)}} + \dfrac{1}{\lambda + \omega}, & x > 0, \\ C_2 e^{x\sqrt{2\lambda}} + \dfrac{1}{\lambda}, & x < 0. \end{cases}$$

Continuity of u and u' at $x = 0$ gives the equations $C_1 - 1/(\lambda + \omega) = C_2 - 1/\lambda$ and $-C_1\sqrt{2(\lambda + \omega)} = C_2\sqrt{2\lambda}$. Solving these yields

$$C_1 = \frac{\sqrt{\lambda + \omega} - \sqrt{\lambda}}{(\lambda + \omega)\sqrt{\lambda}}$$

and hence

$$\int_0^\infty e^{-\lambda t} \mathbb{E}_x e^{-\omega \int_0^t 1_{(0,\infty)}(W_s)\, ds}\, dt = u(0) = \frac{1}{\sqrt{\lambda(\lambda + \omega)}}.$$

On the other hand, Fubini's theorem and a change of variables show that for the same double Laplace transform of the density given in the statement of the theorem, it holds that

$$\begin{aligned} \int_0^\infty & e^{-\lambda t} \Big(\int_0^t e^{-\omega s} \frac{1}{\pi\sqrt{s(t-s)}}\, ds \Big)\, dt \\ &= \int_0^\infty \frac{1}{\pi\sqrt{s}} e^{-\omega s} \Big(\int_s^\infty \frac{1}{\sqrt{t-s}} e^{-\lambda t}\, dt \Big)\, ds \\ &= \frac{1}{\pi} \int_0^\infty \frac{1}{\sqrt{s}} e^{-(\lambda+\omega)s}\, ds \int_0^\infty \frac{1}{\sqrt{u}} e^{-\lambda u}\, du. \end{aligned}$$

A substitution $x = \sqrt{t}$ shows that for $\alpha > 0$,

$$\int_0^\infty \frac{1}{\sqrt{t}} e^{-\alpha t}\, dt = \sqrt{\frac{\pi}{\alpha}}.$$

Hence, we get

$$\int_0^\infty e^{-\lambda t} \Big(\int_0^t e^{-\omega s} \frac{1}{\pi\sqrt{s(t-s)}}\, ds \Big)\, dt = \frac{1}{\sqrt{\lambda(\lambda + \omega)}}.$$

By the uniqueness of Laplace transforms, this completes the proof. $\square$

3.7 Exercises

1. Prove lemma 3.1.4.

2. Complete the proof of Lemma 3.1.6.

3. Let W be a BM. Show that the *reflected Brownian motion* defined by $X = |W|$ is a Markov process (with respect to its natural filtration) and compute its transition function. (Hint: calculate the conditional probability $\mathbb{P}(X_t \in B \,|\, \mathcal{F}_s^X)$ by conditioning further on $\mathcal{F}_s^W$).

4. Let X be a Markov process with state space E and transition function (P_t). Show that for every bounded, measurable function f on E and for all $t \geq 0$, the process $(P_{t-s}f(X_s))_{s\in[0,t]}$ is a martingale.

5. Prove that the probability measure $\mu_{t_1,\ldots,t_n}$ defined in the proof of Corollary 3.2.3 form a consistent system.

6. Work out the details of the proof of Lemma 3.2.4.

7. Prove the claims made in Example 3.3.4. (Hint: to derive the explicit expression for the resolvent kernel it is needed to calculate integrals of the form
$$\int_0^\infty \frac{e^{-a^2t-b^2t^{-1}}}{\sqrt{t}}\,dt.$$
To this end, first perform the substitution $t = (b/a)s^2$. Next, make a change of variables $u = s - 1/s$ and observe that $u(s) = s - 1/s$ is a smooth bijection from $(0,\infty)$ to $\mathbb{R}$, whose inverse $u^{-1} : \mathbb{R} \to (0,\infty)$ satisfies $u^{-1}(t) - u^{-1}(-t) = t$, whence $(u^{-1})'(t) + (u^{-1})'(-t) = 1$.)

8. Finish the proof of Lemma 3.3.5.

9. Prove the inequality in the proof of Lemma 3.3.6.

10. Suppose that $E \subseteq \mathbb{R}^d$. Show that every countable, dense subset $\mathcal{H}$ of the space $C_0^+(E)$ of nonnegative functions in $C_0(E)$ separates the points of the one-point compactification of E. This means that for all $x \neq y$ in E, there exists a function $h \in \mathcal{H}$ such that $h(x) \neq h(y)$ and for all $x \in E$, there exists a function $h \in \mathcal{H}$ such that $h(x) \neq 0$.

11. Let (X, d) be a compact metric space and let $\mathcal{H}$ be a class of nonnegative, continuous functions on X that *separates the points of* X, i.e. for all $x \neq y$ in X, there exists a function $h \in \mathcal{H}$ such that $h(x) \neq h(y)$. Prove that $d(x_n, x) \to 0$ if and only if for all $h \in \mathcal{H}$, $h(x_n) \to h(x)$. (Hint: suppose that $\mathcal{H} = \{h_1, h_2, \ldots\}$, endow $\mathbb{R}^\infty$ with the product topology and consider the map $A : X \to \mathbb{R}^\infty$ defined by $A(x) = (h_1(x), h_2(x), \ldots)$.)

12. Let X and Y be two random variables defined on the same probability space, taking values in a Polish space E. Show that $X = Y$ a.s. if and only if $\mathbb{E}f(X)g(Y) = \mathbb{E}f(X)g(X)$ for all bounded and continuous functions f and g on E. (Hint: use the monotone class theorem.)

13. Let X be a (canonical) Feller process with state space E and for $x \in E$, consider the stopping time $\sigma_x = \inf\{t > 0 : X_t \neq x\}$.

    (i) Using the Markov property, show that for every $x \in E$

    $$\mathbb{P}_x(\sigma_x > t + s) = \mathbb{P}_x(\sigma_x > t)\mathbb{P}_x(\sigma_x > s)$$

    for all $s, t \geq 0$.

    (ii) Conclude that there exists an $a \in [0, \infty]$, possibly depending on x, such that

    $$\mathbb{P}_x(\sigma_x > t) = e^{-at}.$$

    (Remark: this leads to a classification of the points in the state space of a Feller process. A point for which $a = 0$ is called an *absorption point* or a *trap*. If $a \in (0, \infty)$ the point is called a *holding point*. Points for which $a = \infty$ are called *regular*.)

14. Let $(\mathcal{F}_t)$ be the usual augmentation of the natural filtration of a canonical Feller process. Show that for every nonnegative, $\mathcal{F}_t$-measurable random variable Z and every finite stopping time τ, the random variable $Z \circ \theta_\tau$ is $\mathcal{F}_{\tau+t}$-measurable. (Hint: First prove it for $Z = 1_A$ for $A \in \mathcal{F}_t^X$. Next, prove it for $Z = 1_A$ for $A \in \mathcal{F}_t$ and use the fact that $A \in \mathcal{F}_t^\nu$ if and only if there exist $B \in \mathcal{F}_t^X$ and $C, D \in N^\nu$ such that $B \backslash C \subseteq A \subseteq B \cup D$ (cf. Revuz and Yor (1991), pp. 3–4). Finally, prove it for arbitrary Z.)

15. Consider the situation of Exercise 13. Suppose that $x \in E$ is a holding point, i.e. a point for which $a \in (0, \infty)$.

    (i) Observe that $\sigma_x < \infty$, $\mathbb{P}_x$-a.s. and that $\{X_{\sigma_x} = x\} \subseteq \{\sigma_x \circ \theta_{\sigma_x} = 0\}$.

    (ii) Using the strong Markov property, show that

    $$\mathbb{P}_x(X_{\sigma_x} = x) = \mathbb{P}_x(X_{\sigma_x} = x)\mathbb{P}_x(\sigma_x = 0).$$

    (iii) Conclude that $\mathbb{P}_x(X_{\sigma_x} = x) = 0$, i.e. a Feller process can only leave a holding point by a jump.

16. Show that if X is a Feller process and we have $(\mathcal{F}_t)$-stopping times $\tau_n \uparrow \tau$ a.s., then $\lim_n X_{\tau_n} = X_\tau$ a.s. on $\{\tau < \infty\}$. This is called the *quasi-left continuity* of Feller processes. (Hint: First observe that is enough to prove the result for bounded τ. Next, put $Y = \lim_n X_{\tau_n}$ (why does the limit exist?) and use the strong Markov property to show that for $f, g \in C_0$,

    $$\mathbb{E}_x f(Y)g(X_\tau) = \lim_{t \downarrow 0} \lim_n \mathbb{E}_x f(X_{\tau_n})g(X_{\tau_n + t}) = \mathbb{E}_x f(Y)g(Y).$$

    The claim then follows from Exercise 12.)

17. Show that the maps ψ and φ in the proof of Theorem 3.4.4 are Borel measurable.

18. Derive the expression for the joint density of the BM and its running maximum given in Corollary 3.4.5.

19. Let W be a standard BM and S its running maximum. Show that for all $x > 0$ and $t \geq 0$,
$$\mathbb{P}(S_t \geq x) = \mathbb{P}(\tau_x \leq t) = 2\mathbb{P}(W_t \geq x) = \mathbb{P}(|W_t| \geq x).$$
20. Prove Corollary 3.4.6.
21. Prove Lemma 3.5.2.
22. Prove Lemma 3.5.3.
23. Prove the claim of Example 3.5.7. (Hint: use the explicit formula for the resolvent (Example 3.3.4) to show that for $f \in C_0$, it holds that $R_\lambda f \in C_0^2$ and $\lambda R_\lambda f - f = (R_\lambda f)''/2$. The fact that $C_0^2 \subseteq \mathcal{D}(A)$ follows from Theorem 4.1.1.)
24. In the proof of Lemma 3.5.11, show that $\mathbb{P}_x(\eta_r \geq nt) \leq p^n$ for $n = 0, 1, \ldots$.
25. Let (P_t) be a sub-Markovian semigroup on $(E, \mathcal{E})$ and let Δ be a point which does not belong to E. Put $E_\Delta = E \cup \{\Delta\}$, $\mathcal{E}_\Delta = \sigma(\mathcal{E} \cup \{\Delta\})$ and define $\tilde{P}_t$ by setting $\tilde{P}_t(x, A) = P_t(x, A)$ if $A \subseteq E$, $\tilde{P}_t(x, \{\Delta\}) = 1 - P_t(x, E)$ and $\tilde{P}_t(\Delta, \{\Delta\}) = 1$. Show that $(\tilde{P}_t)$ is a semigroup on $(E_\Delta, \mathcal{E}_\Delta)$.
26. Verify the expression for the transition density of the killed BM given in example 3.6.2.

4

Special Feller processes

4.1 Brownian motion in $\mathbb{R}^d$

In this section we consider the Brownian motion in $\mathbb{R}^d$ for $d \in \mathbb{N}$. We will see that the recurrence and transience properties of this process depend on the dimension d of the state space.

A standard BM in $\mathbb{R}^d$ (notation: BM^d) is simply an $\mathbb{R}^d$-valued process $W = (W^1, \dots, W^d)^T$, such that the components $W^1, \dots, W^d$ are independent, standard BM's in $\mathbb{R}$. The BM^d is a process with stationary and independent increments. For $s \le t$, the increment $W_t - W_s$ has a d-dimensional Gaussian distribution with mean vector zero and covariance matrix $(t-s)I$. As in the one-dimensional case it follows that the BM^d is a Markov process with transition function P_t given by

$$P_t f(x) = \frac{1}{(2\pi t)^{d/2}} \int_{\mathbb{R}^d} f(y) e^{-\frac{\|y-x\|^2}{2t}} \, dy \tag{4.1}$$

(see Example 3.1.5). From this explicit expression it easily follows that the process is Feller.

Now let W be the canonical version of the BM^d and consider the measures $\mathbb{P}_x$ for $x \in \mathbb{R}$. By construction W is a standard BM^d under $\mathbb{P}_0$. The semigroup (4.1) has the property that if $f \in C_0$ and $x \in \mathbb{R}$, then for $g(y) = f(y-x)$ it holds that

$$P_t g(x) = P_t f(0).$$

This translation invariance implies that the law of W under $\mathbb{P}_x$ is equal to the law of $x + W$ under $\mathbb{P}_0$ (see Exercise 1). In other words, under $\mathbb{P}_x$ the process W is just a standard BM^d, translated over the vector x.

The following theorem describes the infinitesimal generator of the BM^d. It generalizes Example 3.5.7 to higher dimensions.

Theorem 4.1.1. *The domain of the generator of the* BM^d *contains the space* C_0^2 *of functions on* $\mathbb{R}^d$ *with two continuous derivatives that vanish at infinity. On* C_0^2 *the generator equals* $\Delta/2$*, where* $\Delta = \sum_{i=1}^d \partial^2/\partial y_i^2$ *is the Laplacian on* $\mathbb{R}^d$*.*

Proof. Let $x \in \mathbb{R}^d$ be fixed. Define $p_t(y) = (2\pi t)^{-d/2} \exp(-\|y\|^2/2t)$, so that for $f \in C_0$,

$$P_t f(x) = \int_{\mathbb{R}^d} f\,(x+y)\, p_t(y) dy.$$

The functions $p_t(y)$ satisfy the partial differential equation (the heat equation)

$$\frac{\partial}{\partial t} p_t(y) = \frac{1}{2}\Delta p_t(y).$$

It follows that for $0 < \varepsilon \le t$,

$$\begin{aligned} P_t f(x) - P_\varepsilon f(x) &= \int_{\mathbb{R}^d} f(x+y) \left(\int_\varepsilon^t \frac{\partial}{\partial s} p_s(y)\, ds \right) dy \\ &= \int_{\mathbb{R}^d} f(x+y) \left(\int_\varepsilon^t \frac{1}{2}\Delta p_s(y)\, ds \right) dy \\ &= \int_\varepsilon^t \left(\int_{\mathbb{R}^d} f(x+y) \frac{1}{2}\Delta p_s(y)\, dy \right) ds. \end{aligned}$$

The interchanging of the integrals is justified by the fact that $f \in C_0$. Now suppose that $f \in C_0^2$. Then integrating by parts twice yields

$$\int_{\mathbb{R}^d} f(x+y) \frac{1}{2}\Delta p_s(y)\, dy = \int_{\mathbb{R}^d} p_s(y) \frac{1}{2}\Delta f(x+y)\, dy = P_s \frac{1}{2}\Delta f(x).$$

If we insert this in the preceding display and let $\varepsilon \to 0$, we find that

$$P_t f(x) - f(x) = \int_0^t P_s \frac{1}{2}\Delta f(x)\, ds.$$

Divide by t and let $t \to 0$ to complete the proof. □

In combination with Dynkin's formula, the explicit expression $\Delta/2$ for the generator of the BM^d allows us to investigate the recurrence properties of the higher dimensional BM. For $a \ge 0$, consider the stopping time

$$\tau_a = \inf\{t > 0 : \|W_t\| = a\},$$

which is the first hitting time of the ball of radius a around the origin.

The follow theorem extends the result of Exercise 14 of Chapter 2 to higher dimensions.

Theorem 4.1.2. *For* $0 < a < \|x\| < b$,

$$\mathbb{P}_x(\tau_a < \tau_b) = \begin{cases} \dfrac{\log b - \log \|x\|}{\log b - \log a} & d = 2, \\[2ex] \dfrac{\|x\|^{2-d} - b^{2-d}}{a^{2-d} - b^{2-d}} & d \ge 3. \end{cases}$$

Proof. Let f be a C^∞-function with compact support on $\mathbb{R}^d$ such that for all $x \in \mathbb{R}^d$ with $a \leq \|x\| \leq b$,

$$f(x) = \begin{cases} \log \|x\| & d = 2 \\ \|x\|^{2-d} & d \geq 3. \end{cases}$$

Observe that $\tau_a \wedge \tau_b \leq \inf\{t : |W_t^1| = b\}$ under $\mathbb{P}_x$, so by Exercise 15 of Chapter 2, $\mathbb{E}_x \tau_a \wedge \tau_b < \infty$. Hence, by Dynkin's formula and the fact that $f \in C_0^2$ and $\Delta f(x) = 0$ for $a \leq \|x\| \leq b$, we have $\mathbb{E}_x f(W_{\tau_a \wedge \tau_b}) = f(x)$. For $d = 2$ this yields

$$\begin{aligned} \log \|x\| = f(x) &= \mathbb{E}_x f(W_{\tau_a}) 1_{\{\tau_a < \tau_b\}} + \mathbb{E}_x f(W_{\tau_b}) 1_{\{\tau_a > \tau_b\}} \\ &= \mathbb{P}_x(\tau_a < \tau_b) \log a + \mathbb{P}_x(\tau_a > \tau_b) \log b, \end{aligned}$$

which proves the first statement of the theorem. The second statement follows from Dynkin's formula by the same reasoning, see Exercise 2. □

The following corollary follows from the preceding theorem by letting b tend to infinity.

Corollary 4.1.3. *For all* $x \in \mathbb{R}^d$ *such that* $0 < a < \|x\|$,

$$\mathbb{P}_x(\tau_a < \infty) = \begin{cases} 1 & d = 2, \\ \|x\|^{2-d}/a^{2-d} & d \geq 3. \end{cases}$$

Proof. Note that by the continuity of sample paths,

$$\mathbb{P}_x\left(\tau_a < \infty\right) = \mathbb{P}_x\left(\bigcup_{b > \|x\|} \{\tau_a < \tau_b\}\right).$$

It follows that

$$\mathbb{P}_x\left(\tau_a < \infty\right) = \lim_{b \to \infty} \mathbb{P}_x\left(\tau_a < \tau_b\right),$$

whence the statement follows from Theorem 4.1.2. □

By the translation invariance of the BM, this corollary shows that with probability 1, the BM^2 hits every disc with positive radius, and hence every non-empty open set. It also implies that for $d \geq 3$, the BM^d ultimately leaves every bounded set forever (see Exercise 3). We say that the BM^2 is *neighbourhood-recurrent* and for $d \geq 3$, the BM^d is *transient*.

The following corollary shows that the BM^2 is not 'point-recurrent' like the BM^1. A given point is hit with probability zero.

Corollary 4.1.4. *For the BM*^{2} *we have* $\mathbb{P}_x(\tau_0 < \infty) = 0$ *for all* $x \neq 0$.

Proof. Note that

$$\{\tau_0 < \infty\} = \bigcup_n \{\tau_0 < \tau_n\} = \bigcup_n \bigcap_{m > \|x\|^{-1}} \{\tau_{1/m} < \tau_n\}.$$

Hence, the corollary follows from the fact that

$$\mathbb{P}_x\left(\bigcap_{m>\|x\|^{-1}}\{\tau_{1/m} < \tau_n\}\right) = \lim_{m\to\infty} \mathbb{P}_x(\tau_{1/m} < \tau_n) = 0,$$

by Theorem 4.1.2. □

If we combine all the recurrence-transience results, we arrive at the following theorem.

Theorem 4.1.5. *The BM^1 is recurrent, the BM^2 is neighbourhood-recurrent and for $d \geq 3$, the BM^d is transient.*

4.2 Feller diffusions

Markov processes with continuous sample paths are often called diffusions. In the present section we consider Feller diffusions. Their precise definition is given below. In addition to continuity we require that the domain of the generator of the process contains the collection of functions that are infinitely often differentiable, and whose support is compact and contained in the interior of the state space E. This collection of functions is denoted by $C_c^\infty(\mathrm{int}(E))$.

Definition 4.2.1. A Feller process with state space $E \subseteq \mathbb{R}^d$ is called a *Feller diffusion* if it has continuous sample paths and the domain of its generator contains the function space $C_c^\infty(\mathrm{int}(E))$.

Theorem 4.1.1 shows that the BM^d is a Feller diffusion. More generally we have the following example.

Example 4.2.2. Let W be a BM^d, σ an invertible $d \times d$ matrix, $b \in \mathbb{R}^d$ and define the process X by $X_t = bt + \sigma W_t$. Then X is a Feller diffusion. Its generator A satisfies

$$Af(x) = \sum_i b_i \frac{\partial f}{\partial x_i}(x) + \tfrac{1}{2}\sum_i\sum_j a_{ij}\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$$

for $f \in C_0^2(\mathbb{R}^d)$, where $a = \sigma\sigma^T$. If $b = 0$ this can be seen by relating the semigroup of X to that of W (see Exercise 4). The general case is a consequence of Exercise 5. ■

The process X in the preceding example is called an n-dimensional Brownian motion with *drift* b and *diffusion* matrix a. Below we will prove that a general Feller diffusion locally behaves like such a process. In the general case the drift vector b and diffusion matrix a depend on the space variable x.

We need the following lemma, which collects some properties of the generator of a Feller diffusion.

Lemma 4.2.3. *Let X be a Feller diffusion on E with generator A.*

(i) If $f \in \mathcal{D}(A)$ and $f(y) \leq f(x)$ for all $y \in E$, then $Af(x) \leq 0$.

(ii) If $f, g \in \mathcal{D}(A)$ coincide in a neighborhood of $x \in \text{int}(E)$, then $Af(x) = Ag(x)$.

(iii) If $f \in \mathcal{D}(A)$ has a local maximum in $x \in \text{int}(E)$, then $Af(x) \leq 0$.

(iv) If $f \in \mathcal{D}(A)$ is constant in a neighborhood of $x \in \text{int}(E)$, then $Af(x) = 0$.

Proof. Part (i) is valid for general Feller processes, see the remarks preceding Theorem 3.5.8. Part (ii) follows from Theorem 3.5.12 and the continuity of the sample paths. Combined, (i) and (ii) yield (iii). The last part follows from Theorem 3.5.12 again, by noting that if f is constant near x, the numerator in (3.7) vanishes for small r. □

We can now prove that the generator of a Feller diffusion is a second order differential operator.

Theorem 4.2.4. *Let X be a Feller diffusion on $E \subseteq \mathbb{R}^d$. Then there exist continuous functions a_{ij} and b_j on $\text{int}(E)$ such that for $f \in C_c^\infty(\text{int}(E))$ and $x \in \text{int}(E)$,*

$$Af(x) = \sum_i b_i(x) \frac{\partial f}{\partial x_i}(x) + \tfrac{1}{2} \sum_i \sum_j a_{ij}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x).$$

Moreover, the matrix $(a_{ij}(x))$ is symmetric and nonnegative definite for every $x \in \text{int}(E)$.

Proof. Fix $x \in \text{int}(E)$ and pick $\varphi_i \in C_c^\infty(\text{int}(E))$ such that $\varphi_i(y) = y_i - x_i$ for y in a neighborhood of x. Let $\varphi \in C_c^\infty(\text{int}(E))$ be a function which is identically 1 in a neighborhood of x. Define $b_i(x) = A\varphi_i(x)$ and $a_{ij}(x) = A\varphi_i\varphi_j(x)$. Note that a and b are well defined by part (ii) of the lemma, and since $A : \mathcal{D}(A) \to C_0$, they are continuous. Also observe that for all $\lambda_1, \ldots, \lambda_d \in \mathbb{R}$, the function

$$y \mapsto -\Big(\sum_i \lambda_i \varphi_i(y)\Big)^2,$$

defined in a neighborhood of x, has a local maximum at x. By part (iii) of the lemma this implies that

$$\sum_i \sum_j \lambda_i \lambda_j a_{ij}(x) \geq 0,$$

i.e. $(a_{ij}(x))$ is nonnegative definite.

By Taylor's formula it holds that $f(y) = g(y) + o(\|x - y\|^2)$ for y near x, where

$$g(y) = f(x)\varphi(y) + \sum_i \varphi_i(y)\frac{\partial f}{\partial x_i}(x) + \tfrac{1}{2}\sum_i\sum_j \varphi_i(y)\varphi_j(y)\frac{\partial^2 f}{\partial x_i \partial x_j}(x).$$

Part (iv) of the lemma and the definitions of a and b imply that

$$Ag(x) = \sum_i b_i(x)\frac{\partial f}{\partial x_i}(x) + \tfrac{1}{2}\sum_i\sum_j a_{ij}(x)\frac{\partial^2 f}{\partial x_i \partial x_j}(x),$$

so it remains to show that $Af(x) = Ag(x)$. To prove this, fix $\varepsilon > 0$ and consider the function

$$y \mapsto f(y) - g(y) - \varepsilon\|y - x\|^2,$$

defined in a neighborhood of x. Near x this function is $\|y - x\|^2(o(1) - \varepsilon)$, and hence it has a local maximum at x. By part (iii) of the lemma we get

$$Af(x) - Ag(x) - \varepsilon\sum_i a_{ii}^2(x) \le 0,$$

and by letting $\varepsilon \to 0$ we obtain $Af(x) \le Ag(x)$. Reversing the roles of f and g yields $Af(x) \ge Ag(x)$. This completes the proof. □

In view of Example 4.2.2, we call the functions b and a exhibited in Theorem 4.2.4 the *drift* and *diffusion* of the process X, respectively. The proof of the theorem shows that b and a determine the infinitesimal movements of the diffusion process $X = (X^1, \ldots, X^d)$, in the sense that for every $x \in \text{int}(E)$ and small h, it holds that $\mathbb{E}_x X_h^i - x_i \approx P_h\varphi_i(x) \approx hA\varphi_i(x) = b_i(x)h$. In other words,

$$\mathbb{E}(X_{t+h} \,|\, X_t = x) \approx x + hb(x).$$

Similarly, we have

$$\mathbb{E}((X_{t+h}^i - x_i)(X_{t+h}^j - x_j) \,|\, X_t = x) \approx ha_{ij}(x).$$

Next we investigate when a function of a Feller diffusion is again a Feller diffusion, and how the drift and diffusion functions of the new process are related to the old one.

Theorem 4.2.5. *Let X be a Feller diffusion on $E \subseteq \mathbb{R}^d$ and let $\varphi : E \to \tilde{E} \subseteq \mathbb{R}^n$ be continuous and onto, and assume that $\|\varphi(x_n)\| \to \infty$ in $\tilde{E}$ if and only if $\|x_n\| \to \infty$ in E. Suppose that $(\tilde{P}_t)$ is a collection of transition kernels such that $P_t(f \circ \varphi) = (\tilde{P}_t f) \circ \varphi$ for all $f \in C_0(\tilde{E})$, so that $\tilde{X} = \varphi(X)$ is a Feller process with state space $\tilde{E}$ and transition function $(\tilde{P}_t)$.*

(i) If φ is infinitely often differentiable on int(E) and for every compact $K \subseteq$ int$(\tilde{E})$, $\varphi^{-1}(K)$ is contained in a compact subset of int(E), then $\tilde{X}$ is a Feller diffusion on $\tilde{E}$.

(ii) In that case the drift and diffusion $\tilde{b}$ and $\tilde{a}$ of $\tilde{X}$ satisfy

$$\tilde{b}_k(\varphi(x)) = \sum_i b_i(x)\frac{\partial\varphi_k}{\partial x_i}(x) + \tfrac{1}{2}\sum_i\sum_j a_{ij}(x)\frac{\partial^2\varphi_k}{\partial x_i \partial x_j}(x),$$

$$\tilde{a}_{kl}(\varphi(x)) = \sum_i\sum_j a_{ij}(x)\frac{\partial\varphi_k}{\partial x_i}(x)\frac{\partial\varphi_l}{\partial x_j}(x)$$

for $\varphi(x) \in int(\tilde{E})$ and $k, l = 1, \dots, n$.

Proof. (i). Since φ is continuous the process $\tilde{X}$ is continuous, so we only have to show that $C_c^\infty(\text{int}(\tilde{E}))$ is contained in the domain of the generator $\tilde{A}$ of $\tilde{X}$. So take $f \in C_c^\infty(\text{int}(\tilde{E}))$. The assumptions on φ imply that $f \circ \varphi \in C_c^\infty(\text{int}(E)) \subseteq \mathcal{D}(A)$. Hence, by Lemma 3.5.2, $f \in \mathcal{D}(\tilde{A})$.

(ii). To simplify the notation we write ∂_i instead of $\partial/\partial x_i$ for the differential w.r.t. the ith variable. By the chain rule we have

$$\partial_i(f \circ \varphi)(x) = \sum_k \partial_k f(\varphi(x))\partial_i\varphi_k(x)$$

and (check!)

$$\partial_i\partial_j(f \circ \varphi)(x) = \sum_k\sum_l \partial_k\partial_l f(\varphi(x))\partial_i\varphi_k(x)\partial_j\varphi_k(x) + \sum_k \partial_i\partial_j\varphi_k(x)\partial_k f(\varphi(x)).$$

Now compare $A(f \circ \varphi)(x)$ with $\tilde{A}f(\varphi(x))$ and apply Lemma 3.5.2 to complete the proof (Exercise 6). □

Example 4.2.6 (Bessel processes). Let X be a BM^d with $d \geq 2$ and define $\tilde{X} = \|X\|$. This process is called the *Bessel process* of order d and is denoted by BES^d. The transition function P_t of the BM^d (see (4.1)) has the property that if $f \in C_0(\mathbb{R}^d)$ and $f(x)$ depends only on $\|x\|$, then $P_t f(x)$ also depends only on $\|x\|$ (check!). So if $f \in C_0(\mathbb{R}_+)$, x is a point on the unit sphere in $\mathbb{R}^d$ and $r \geq 0$ then $\tilde{P}_t f(r) = P_t(f \circ \|\cdot\|)(rx)$ is independent of x, and this defines a transition kernel $\tilde{P}_t$ on $\mathbb{R}_+$. The function $\varphi = \|\cdot\|$ and the semigroup $(\tilde{P}_t)$ satisfy the conditions of Theorem 4.2.5, so the Bessel process $\tilde{X}$ is a Feller diffusion on $\mathbb{R}_+$. For $\varphi(x) = \|x\| > 0$, we have

$$\frac{\partial\varphi}{\partial x_i}(x) = \frac{x_i}{\|x\|}, \quad \frac{\partial^2\varphi}{\partial x_i \partial x_j}(x) = \frac{\delta_{ij}}{\|x\|} - \frac{x_i x_j}{\|x\|^3}.$$

By part (ii) of the preceding theorem, it follows that the drift $\tilde{b}$ and diffusion $\tilde{a}$ of the BES^d are given by

$$\tilde{b}(x) = \frac{d-1}{2x}, \quad \tilde{a}(x) = 1. \tag{4.2}$$

(See Exercise 7).

Observe that for x close to 0 the drift upward becomes very large. As a consequence the process always gets “pulled up” when it approaches 0, so that it stays nonnegative all the time. ■

If we combine Theorem 4.2.4 and Example 4.2.2 we see that a Feller diffusion process X can be viewed as a BM with space dependent drift and diffusion. Loosely speaking, it holds that around time t, X_t moves like a BM with drift $b(X_t)$ and diffusion $a(X_t)$. If we consider the one-dimensional case for simplicity, this means that for s close to t, we have something like

$$X_s = b(X_t)s + \sigma(X_t)W_s,$$

where W is a BM (this is of course purely heuristic). If we consider this equation for $s = t + h$ and $s = t$ and subtract, we see that for small h, we must have

$$X_{t+h} - X_t \approx b(X_t)h + \sigma(X_t)(W_{t+h} - W_t).$$

This suggests that a one-dimensional Feller diffusion with generator drift b and diffusion σ^2 satisfies a *stochastic differential equation (SDE)* of the form

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t, \tag{4.3}$$

with W a Brownian motion.

Of course, it is not at all exactly clear what we mean by this. In particular, the term dW_t is somewhat disturbing. Since we know that the Brownian sample paths are of unbounded variation, the path $W(\omega)$ does not define a (signed) measure $dW(\omega)$. This means that in general, we can not make sense of the SDE (4.3) by viewing it as a pathwise equation for the process X. It turns out however that there is a way around this, using the fact that the Brownian motion has finite *quadratic* variation. The SDE (4.3) can be given a precise meaning and it can be shown that indeed, under certain assumptions on b and σ, it has a unique solution X which is a Markov process with generator $Af(x) = b(x)f'(x) + \frac{1}{2}\sigma^2(x)f''(x)$. The theory of SDEs and the associated *stochastic calculus* are very useful tools in the study of diffusions.

It is interesting to see how Theorem 4.2.5 translates in the language of SDEs. The theorem states that if X is a one-dimensional Feller diffusion with drift b and diffusion σ^2 and φ is a smooth function, then, under certain conditions, $Y = \varphi(X)$ is a Feller diffusion with drift $\tilde{b}$ and diffusion $\tilde{\sigma}^2$ given by

$$\tilde{b}(\varphi(x)) = b(x)\varphi'(x) + \tfrac{1}{2}\sigma^2(x)\varphi''(x),$$
$$\tilde{\sigma}^2(\varphi(x)) = \sigma^2(x)(\varphi'(x))^2.$$

In the language of SDE's, this show that if X satisfies (4.3), then $\varphi(X)$ satisfies

$$d\varphi(X_t) = \Big(b(X_t)\varphi'(X_t) + \tfrac{1}{2}\sigma^2(X_t)\varphi''(X_t)\Big)\,dt + \sigma(X_t)\varphi'(X_t)\,dW_t. \tag{4.4}$$

This is a special case of a central result in stochastic calculus, known as *Itô's formula*. It shows in particular that there is a difference between stochastic calculus and the ordinary calculus. Indeed, if the rules of ordinary calculus were valid in this case, we would have

$$d\varphi(X_t) = \varphi'(X_t)\,dX_t = b(X_t)\varphi'(X_t)\,dt + \sigma(X_t)\varphi'(X_t)\,dW_t.$$

The presence of the additional term

$$\tfrac{1}{2}\sigma^2(X_t)\varphi''(X_t)\,dt$$

in (4.4) is typical for stochastic calculus, and is caused by the quadratic variation of the Brownian sample paths.

See the references for texts on stochastic calculus and stochastic differential equations.

4.3 Feller processes on discrete spaces

Consider a Markov process X on a discrete state space E, meaning that E is countable and that it is endowed with the discrete topology. Then E is a Polish space and its Borel σ-algebra $\mathcal{E}$ consists of all subsets of E.

As usual we denote the semigroup of X by (P_t) and we consider the canonical version with underlying measures $\mathbb{P}_x$, $x \in E$. It is useful to introduce the numbers

$$p_t(x, y) = P_t(x, \{y\}) = \mathbb{P}_x(X_t = y)$$

for $t \geq 0$ and $x, y \in E$. These transition probabilities of course satisfy

$$\sum_y p_t(x, y) = 1 \tag{4.5}$$

for $x \in E$ and $t \geq 0$, and the Chapman-Kolmogorov equations reduce to

$$p_{s+t}(x, y) = \sum_z p_s(x, z) p_t(z, y) \tag{4.6}$$

for $s, t \geq 0$ and $x, y \in E$.

Conversely, every collection of nonnegative numbers $\{p_t(x, y) : t \geq 0, x, y \in E\}$ satisfying (4.5) and (4.6) defines a transition function (P_t) on $(E, \mathcal{E})$, given by

$$P_t(x, B) = \sum_{y \in B} p_t(x, y),$$

and hence for every $x \in E$ there exist a measure $\mathbb{P}_x$ on the canonical space such that under $\mathbb{P}_x$, the canonical process X is a Markov process which starts in x and which has transition probabilities $p_t(x, y)$.

To verify if a Markov process on a discrete space is Feller we have to understand what the function space C_0 looks like in this case. Since the space E is discrete, every function on E is automatically continuous. For a sequence x_n in E, the convergence $\|x_n\| \to \infty$ means that x_n is outside every given finite subset of E for n large enough. Hence, $f \in C_0$ means that for every $\varepsilon > 0$, there exists a finite set $F \subseteq E$ such that that $|f(x)| < \varepsilon$ for all x outside F. Note in particular that if the state space E itself is finite, every function on E belongs to C_0.

Example 4.3.1 (Finite state space). If the state space E is finite, then every function belongs to C_0, so $P_t C_0 \subseteq C_0$ is always true. It follows that a Markov process on a finite state space E is Feller if and only if its transition probabilities satisfy, for all $x, y \in E$,

$$p_t(x, y) \to \delta_{x,y}$$

as $t \to 0$. ■

Example 4.3.2 (Poisson process). Let $\lambda > 0$ be a fixed parameter. For $t > 0$ and $x, y \in \mathbb{Z}_+$, define

$$p_t(x,y) = \begin{cases} \frac{(\lambda t)^k}{k!} e^{-\lambda t}, & \text{if } y = x + k \text{ for } k \in \mathbb{Z}_+, \\ 0, & \text{otherwise,} \end{cases} \tag{4.7}$$

and put $p_0(x,y) = \delta_{x,y}$. Upon recognizing Poisson probabilities we see that (4.5) holds. As for (4.6), observe that

$$\begin{aligned} \sum_z p_s(x,z)p_t(z,y) &= \sum_{l=0}^{k} p_s(x,x+l)p_t(x+l,x+k) \\ &= \frac{e^{\lambda(t+s)}}{k!} \sum_{l=0}^{k} \binom{k}{l} (\lambda s)^l (\lambda t)^{k-l}. \end{aligned}$$

By Newton's binomial theorem the sum on the right-hand side equals $(\lambda(t+s))^k$, so the Chapman-Kolmogorov equalities are indeed satisfied. It follows that the transition probabilities (4.7) define a Markov process on $\mathbb{Z}_+$. This process is called the *Poisson process* with *intensity* λ. Observe that if N is a Poisson process with intensity λ, then for $k \in \mathbb{Z}_+$

$$\mathbb{P}_0(N_t = k) = p_t(0,k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

So under $\mathbb{P}_0$, the random variable N_t has a Poisson distribution with parameter λt.

To verify that this process is Feller, take a function f on $\mathbb{Z}_+$ such that $f(x) \to 0$ as $x \to \infty$. Then

$$P_t f(x) = \sum_y f(y)p_t(x,y) = e^{-\lambda t} \sum_{k \geq 0} \frac{(\lambda t)^k}{k!} f(x+k).$$

The summand on the right-hand side vanishes as $x \to \infty$ and is bounded by $\|f\|_\infty (\lambda t)^k/k!$. Hence, by dominated convergence, $P_t f(x) \to 0$ as $x \to \infty$. This shows that $P_t C_0 \subseteq C_0$. Note that we can also write

$$P_t f(x) = e^{-\lambda t} f(x) + e^{-\lambda t} \sum_{k \geq 1} \frac{(\lambda t)^k}{k!} f(x+k).$$

This shows that $P_t f(x) \to f(x)$ as $t \to 0$. Hence, the Poisson process is Feller.

Using the explicit expression for P_t above it is rather straightforward to compute the generator A of the Poisson process with intensity λ. It holds that $\mathcal{D}(A) = C_0$ and

$$Af(x) = \lambda(f(x+1) - f(x))$$

(see Exercise 8). We will see below that the form of the generator reveals much about the structure of the Poisson process. ■

The motion of a Feller process X on a discrete space E is of a simple nature. The process stays in a certain state for a random amount of time, then jumps to the next state, etc. For $x \in E$, consider the stopping time $\sigma_x = \inf\{t > 0 : X_t \neq x\}$. From Exercise 13 of Chapter 3 we know that under $\mathbb{P}_x$, the random time σ_x has an exponential distribution, say with parameter $\lambda(x) \in [0, \infty)$ ($\lambda(x) = 0$ meaning that $\sigma_x = \infty$, $\mathbb{P}_x$-a.s., i.e. that x is an absorption point). This defines a function $\lambda : E \to [0, \infty)$, where $\lambda(x)$ is interpreted as the rate at which the state x is left by the process X. Now start the process in a point $x \in E$ with $\lambda(x) > 0$. Then at time σ_x the process jumps from state x to X_{σ_x}. So if for $x \neq y$ in E we define

$$q(x, y) = \mathbb{P}_x(X_{\sigma_x} = y),$$

we can interpret $q(x, y)$ as the probability that the process jumps to y when it leaves state x. Clearly, we have that $q(x, y) = Q(x, \{y\})$, where Q is the transition kernel on E defined by

$$Q(x, B) = \mathbb{P}_x(X_{\sigma_x} \in B).$$

Together, the rate function λ and the transition kernel Q (or, equivalently, the transition probabilities $q(x, y)$) determine the entire motion of the process. When started in x the process stays there for an exponential($\lambda(x)$)-distributed amount of time. Then it chooses a new state y according to the probability distribution $Q(x, \cdot)$ and jumps there. By the strong Markov property this procedure then repeats itself. The process stays in y for an exponential($\lambda(y)$)-distributed amount of time and then jumps to a new state z chosen according to the probability distribution $Q(z, \cdot)$, etc. We call the constructed function λ the *rate function* of the process, and Q the *transition kernel*.

The rate function and transition kernel of a Feller process on a discrete space can easily be read of from the generator.

Theorem 4.3.3. *Let X be a Feller process on a discrete space E, with rate function λ and transition kernel Q. Then the domain $\mathcal{D}(A)$ of its generator A consists of all functions $f \in C_0$ such that $\lambda(Qf - f) \in C_0$, and for $f \in \mathcal{D}(A)$,*

$$Af(x) = \lambda(x)(Qf(x) - f(x)) = \lambda(x)\Big(\sum_{y \neq x} f(y)q(x, y) - f(x)\Big).$$

Proof. This follows from Theorem 3.5.12, see Exercise 9. □

Example 4.3.4 (Poisson process, continued). Using Theorem 4.3.3 we can now read off the rate function and transition kernel of the Poisson process from the expression $Af(x) = \lambda(f(x + 1) - f(x))$ that we have for its generator. We conclude that the rate function is identically equal to λ, and that $q(x, x + 1) = Q(x, \{x + 1\}) = 1$, and all other transition probabilities are 0. So the jumps of the Poisson process are all equal to 1 (a process with this property is called a *counting process*), and the time between two consecutive jumps is exponentially distributed with parameter λ. ■

The preceding example shows that the Poisson process can be constructed by starting with the deterministic process which has jumps of size 1 at times $1, 2, 3, \ldots$ and then replacing the deterministic times between the jumps by independent exponential times. More generally, any discrete-time Markov chain can be transformed into a continuous-time Markov process by replacing the fixed times between the jumps by independent exponential times.

To make this precise, let Q be a transition kernel on E and $\lambda > 0$. For $t \geq 0$ and f a bounded function on E, define

$$P_t f(x) = e^{-\lambda t} \sum_{n=0}^{\infty} Q^n f(x) \frac{(\lambda t)^n}{n!}, \tag{4.8}$$

where $Q^n = Q \circ \cdots \circ Q$ (n times). Using Newton's binomial theorem it is easily verified that (P_t) is a transition semigroup on E (see Exercise 10), so we can consider the associated Markov process on E.

The intuition behind the construction of this process is as follows. Let $Y = (Y_n)$ be a discrete-time Markov chain with transition kernel Q, meaning that given $Y_n = y$, the random variable Y_{n+1} has distribution $Q(y, \cdot)$. Set $\tau_0 = 0$ and let $\tau_1, \tau_2, \ldots$ be i.i.d., all exponentially distributed with parameter λ. Now define the process X by

$$X_t = \sum_{n=0}^{\infty} Y_n 1_{[\tau_n, \tau_{n+1})}(t).$$

From the lack of memory of the exponential distribution and the Markov property of Y it is intuitively clear that X has the Markov property as well. Its transition function (P_t) is given by

$$\begin{aligned} P_t f(x) = \mathbb{E}_x f(X_t) &= \sum_{n=0}^{\infty} \mathbb{E}_x f(Y_n) \mathbb{P}_x(\tau_n \leq t < \tau_n) \\ &= e^{-\lambda t} \sum_{n=0}^{\infty} Q^n f(x) \frac{(\lambda t)^n}{n!}. \end{aligned}$$

Here we have used the Markov property of Y to see that $\mathbb{E}_x f(Y_n) = \mathbb{E}_x \mathbb{E}_x(f(Y_n) \,|\, Y_{n-1}, \ldots, Y_0) = \mathbb{E}_x Q f(Y_{n-1})$. Iterating this yields $\mathbb{E}_x f(Y_n) = Q^n f(x)$. Moreover, with N_t a Poisson process that starts in 0, we have that $\mathbb{P}(\tau_n \leq t < \tau_n) = \mathbb{P}(N_t = n) = \exp(-\lambda t)(\lambda t)^n/n!$.

Theorem 4.3.5. *If Q is a transition kernel on E such that $QC_0 \subseteq C_0$ and $\lambda > 0$, then (4.8) defines a Feller process on E whose generator A satisfies $\mathcal{D}(A) = C_0$ and $Af = \lambda(Qf - f)$.*

Proof. The properties $P_t C_0 \subseteq C_0$ and $P_t f(x) \to f(x)$ as $t \to 0$ follow from (4.8) by dominated convergence (see Exercise 11). We also see from (4.8) that

$$\frac{P_t f(x) - f(x)}{t} = \lambda Q f(x) e^{-\lambda t} + \frac{e^{-\lambda t} - 1}{t} f(x) + O(t).$$

Letting $t \to 0$ the right-hand side converges to $\lambda(Qf(x) - f(x))$. □

If in the present setting we define $q(x, y) = Q(x, \{y\})$ for $x, y \in E$, the generator A can be expressed as

$$Af(x) = \lambda(1 - q(x,x))\Big(\sum_{y \neq x} \frac{q(x,y)}{1 - q(x,x)} f(y) - f(x)\Big).$$

Hence, by Theorem 4.3.3, the construction (4.8) yields a Feller process on E with rate function $\lambda(x) = \lambda(1 - q(x,x))$ and transition probabilities

$$\frac{q(x,y)}{1 - q(x,x)}$$

for $x \neq y$.

Example 4.3.6 (Continuous-time random walk). Ordinary random walk in $\mathbb{Z}^d$ is a Markov chain with transition kernel Q given by

$$Q(x, \{y\}) = \begin{cases} \dfrac{1}{2d}, & \text{if } y \in N_x, \\ 0, & \text{otherwise,} \end{cases}$$

where N_x is the set consisting of the $2d$ neighbors of x. In other words, at each time step the chain moves to one of the neighbors of the current state, all neighbors being equally likely. Clearly $QC_0 \subseteq C_0$ in this case. The associated Feller process constructed above is called the *continuous-time random walk on* $\mathbb{Z}^d$. If we take $\lambda = 1$ its generator is given by

$$Af(x) = \sum_{y \in N_x} \frac{1}{2d} f(y) - f(x),$$

as is easily checked. ■

4.4 Lévy processes

We already encountered processes with independent and stationary increments in Corollary 3.4.2. In this section we study this class of processes in some more detail. We restrict ourselves to real-valued processes, although most results can be generalized to higher dimensions.

4.4.1 Definition, examples and first properties

We begin with the basic definition.

Definition 4.4.1. An $(\mathcal{F}_t)$-adapted process X is called a *Lévy process* (with respect to $(\mathcal{F}_t)$) if

(i) $X_t - X_s$ is independent of $\mathcal{F}_s$ for all $s \le t$,

(ii) $X_t - X_s =_d X_{t-s}$ for all $s \le t$,

(iii) the process is continuous in probability, i.e. if $t_n \to t$ then $X_{t_n} \to X_t$ in probability.

Observe that the first two items of the definition imply that for $u \in \mathbb{R}$, the function $f(t) = \mathbb{E}\exp(iuX_t)$ satisfies

$$f(s+t) = \mathbb{E}e^{iu((X_{s+t}-X_t)+X_t)} = \mathbb{E}e^{iuX_s}\mathbb{E}e^{iuX_t} = f(s)f(t).$$

By item (iii) the function f is continuous, and it follows that the characteristic function of X_t is of the form

$$\mathbb{E}e^{iuX_t} = f(t) = e^{-t\psi(u)},$$

where $\psi : \mathbb{R} \to \mathbb{C}$ is a continuous function with $\psi(0) = 0$. The function ψ is called the *characteristic exponent* of the Lévy process. Note in particular that $X_0 = 0$ almost surely. The independence and stationarity of the increments implies that the characteristic exponent determines the whole distribution of a Lévy process (see Exercise 13).

Recall that if Y is an integrable random variable with characteristic function φ, then $\varphi'(0) = i\mathbb{E}Y$. So if X is a Lévy process and X_t is integrable, then the characteristic exponent ψ is differentiable and $\mathbb{E}X_t = it\psi'(0)$. If X_t has a finite second moment as well, we have $\mathbb{V}\mathrm{ar}X_t = \sigma^2 t$ for some $\sigma^2 \ge 0$ (see Exercise 14).

The independence and stationarity of the increments imply that for a Lévy process X and a bounded measurable function f we have, for $s \le t$,

$$\mathbb{E}(f(X_t) \,|\, \mathcal{F}_s) = \mathbb{E}(f((X_t - X_s) + X_s) \,|\, \mathcal{F}_s) = P_{t-s}f(X_s),$$

where $P_t f(x) = \mathbb{E}f(x + X_t)$. Hence, a Lévy process is a Markov process with transition kernel $P_t f(x) = \mathbb{E}f(x + X_t)$. It is not difficult to see that the semigroup (P_t) is Feller (Exercise 15), so a Lévy process is always a Feller process. From this point on, we will always consider the canonical, cadlag version of X and the usual augmentation of its natural filtration. As usual, we denote by $\mathbb{P}_x$ the law under which X is a Markov process with transition function (P_t) and initial distribution δ_x. Under $\mathbb{P}_0$ the process is Lévy, and since $\mathbb{E}_x f(X_t) = P_t f(x) = \mathbb{E}_0 f(x + X_t)$, the law of X under $\mathbb{P}_x$ is equal to the law of $x + X$ under $\mathbb{P}_0$. Below we will write $\mathbb{P}$ for $\mathbb{P}_0$ and $\mathbb{E}$ for $\mathbb{E}_0$.

Example 4.4.2 (Brownian motion). If W is a standard Brownian motion and $b \in \mathbb{R}$ and $\sigma \geq 0$, then the Brownian motion with drift defined by $X_t = bt + \sigma W_t$ is clearly a Lévy process. Since $\exp(iuZ) = \exp(-u^2/2)$ if Z has a standard normal distribution, the characteristic exponent of the process X is given by

$$\psi(u) = -ibu + \tfrac{1}{2}\sigma^2 u^2.$$

We will see below that the Brownian motion with linear drift is in fact the only Lévy process with continuous sample paths (see Theorem 4.4.13). ■

Example 4.4.3 (Poisson process). Let N be a Poisson process with intensity $\lambda > 0$. Then by the Markov property $\mathbb{E}(f(N_t - N_s) \,|\, \mathcal{F}_s) = \mathbb{E}_{N_s} f(N_{t-s} - N_0)$ for every bounded, measurable function f and $s \leq t$. For $x \in \mathbb{Z}_+$ we have

$$\mathbb{E}_x f(N_{t-s} - N_0) = P_{t-s} g(x),$$

where $g(y) = f(y - x)$. Using the explicit expression for the semigroup (P_t) (cf. Example 4.3.2) we see that $P_{t-s}g(x) = P_{t-s}f(0)$. Hence,

$$\mathbb{E}(f(N_t - N_s) \,|\, \mathcal{F}_s) = P_{t-s}f(0) = \mathbb{E}f(N_{t-s}).$$

This shows that the Poisson process, when started in 0, is a Lévy process. Recall that if Z has a Poisson distribution with parameter $\alpha > 0$, then $\mathbb{E}\exp(iuZ) = \exp(-\alpha(1 - e^{iu}))$. It follows that the characteristic exponent of the Poisson process with intensity λ is given by

$$\psi(u) = \lambda(1 - e^{iu}).$$

In fact, the Poisson process is the only non-trivial Lévy process that is also counting process. Indeed, let N be a process of the latter type. Then N is a Feller process on the discrete space $\mathbb{Z}_+$, of the type studied in Section 4.3. Since it is a counting process, its transition kernel is given by $Q(x, \{x+1\}) = 1$ for $x \in \mathbb{Z}_+$. As for the rate function, let σ_x be the first time the process N leaves x. Under $\mathbb{P}_x$, N has the same law as $x + N$ under $\mathbb{P}_0$. Hence, the law of σ_x under $\mathbb{P}_x$ is equal to the law of σ_0 under $\mathbb{P}_0$. This shows that the rate function is equal to a constant $\lambda \geq 0$ (if $\lambda = 0$, the process is trivial). Theorem 4.3.3 now shows that N has generator $Af(x) = \lambda(f(x+1) - f(x))$. By Example 4.3.2 and the fact that the generator determines the semigroup, it follows that N is a Poisson process. ■

Example 4.4.4 (Compound Poisson process). Let N be a Poisson process with intensity $\lambda > 0$, started in 0, and $Y_1, Y_s, \ldots$ i.i.d. random variables with common distribution function F, independent of N. Then the process X defined by

$$X_t = \sum_{i=1}^{N_t} Y_i$$

is called a *compound Poisson process*. It is easily seen to be a Lévy process by conditioning on N and using the fact that N is a Lévy process (see Exercise 16). It can also be shown by conditioning that the characteristic exponent of the compound Poisson process is given by

$$\psi(u) = \lambda \int (1 - e^{iux})\, dF(x)$$

(see the same exercise). ■

Example 4.4.5 (First passage times). Let W be a standard Brownian motion and for $a \geq 0$, define

$$\sigma_a = \inf\{t : W_t > a\},$$

the *first passage time* of the level a. By construction it holds that $W_{\sigma_a + t} > a$ for t small enough, which implies that the process $(\sigma_a)_{a\geq 0}$ is right-continuous. Moreover, it is clear that for $a_n \uparrow a$ it holds that $\sigma_{a_n} \uparrow \tau_a$, where $\tau_a = \inf\{t : W_t = a\}$ is the first hitting time of the level a. It follows in particular that the process $(\sigma_a)_{a\geq 0}$ is cadlag. For $b > a$ we have

$$\sigma_b - \sigma_a = \inf\{t : W_{\sigma_a + t} - W_{\sigma_a} > b - a\}.$$

Hence, by the strong Markov property (see Corollary 3.4.2), $\sigma_b - \sigma_a$ is independent of $\mathcal{F}_{\sigma_a}$ and has the same distribution as σ_{b-a}. This shows that the process of first passage times $(\sigma_a)_{a\geq 0}$ is a Lévy process. It is in fact the cadlag modification of the process $(\tau_a)_{a\geq 0}$ of first hitting times. To see this, let $a_n \uparrow a$. As we just noted, $\sigma_{a_n} \to \tau_a$ a.s.. On the other hand, the quasi-left continuity of the first passage time process implies that a.s. $\sigma_{a_n} \to \sigma_a$ (see Exercise 16 of Chapter 3). It follows that a.s., $\sigma_a = \tau_a$. ■

Since a Lévy process is cadlag, the number of jumps ΔX_s before some time t such that $|\Delta X_s| \geq \varepsilon$ has to be finite for all $\varepsilon > 0$. Hence, if B is Borel set bounded away from 0, then for $t \geq 0$

$$N_t^B = \#\{s \in [0,t] : \Delta X_s \in B\}$$

is well defined and a.s. finite. The process N^B is clearly a counting process, and we call it the *counting process of* B. It inherits the Lévy property from X and hence, by Example 4.4.3, N^B is a Poisson process with a certain intensity, say $\nu^X(B) < \infty$. If B is a disjoint union of Borel sets B_i, then $N^B = \sum_i N^{B_i}$, hence $\nu^X(B) = \mathbb{E}N_1^B = \sum \mathbb{E}N_1^{B_i} = \sum \nu(B_i)$. We conclude that ν^X is a Borel measure on $\mathbb{R}\backslash\{0\}$. It holds that $\nu^X(\mathbb{R}\backslash(-\varepsilon,\varepsilon)) < \infty$ for all $\varepsilon > 0$, in particular ν^X is σ-finite. The measure ν^X is called the *Lévy measure* of the process X.

It is easily seen that for the Brownian motion and the Poisson process with intensity λ, we have $\nu^X = 0$ and $\nu^X = \lambda\delta_1$, respectively. As for the compound Poisson process $X_t = \sum_{i\leq N_t} Y_i$ of Example 4.4.4, observe that the counting process of the Borel set B is given by

$$N_t^B = \sum_{i=1}^{N_t} 1_{\{Y_i \in B\}}.$$

Hence, by conditioning on N we see that the Lévy measure is given by $\nu^X(B) = \mathbb{E}N_1^B = \lambda \int_B dF$, i.e. $\nu^X(dx) = \lambda\, dF(x)$.

4.4.2 Jumps of a Lévy process

In this subsection we analyze the structure of the jumps of a Lévy process. The first theorem provides useful information about Lévy process with bounded jumps.

Theorem 4.4.6. *A Lévy process with bounded jumps has finite moments of all orders.*

Proof. Let X be a Lévy process and let the number $C > 0$ be such that $|\Delta X_t| \leq C$ for all $t \geq 0$. Define the stopping times $\tau_0, \tau_1, \tau_2, \ldots$ by $\tau_0 = 0$,

$$\tau_1 = \inf\{t : |X_t - X_0| \geq C\}, \ldots, \tau_{n+1} = \inf\{t > \tau_n : |X_t - X_{\tau_n}| \geq C\}, \ldots.$$

The fact that X is right-continuous implies that τ_1 is a.s. strictly positive. Using the uniqueness of Laplace transforms, it follows that for some $\lambda > 0$, we have $a = \mathbb{E}_0 \exp(-\lambda\tau_1) < 1$. By the strong Markov property we have, on the event $\{\tau_{n-1} < \infty\}$,

$$\begin{aligned}
&\mathbb{E}\big(e^{-\lambda(\tau_n - \tau_{n-1})} \,|\, \mathcal{F}_{\tau_{n-1}}\big) \\
&\quad = \mathbb{E}\big(e^{-\lambda\tau_1} \circ \theta_{\tau_{n-1}} \,|\, \mathcal{F}_{\tau_{n-1}}\big) = \mathbb{E}_{X_{\tau_{n-1}}} e^{-\lambda\tau_1} = \mathbb{E}e^{-\lambda\tau_1},
\end{aligned}$$

where the last equality in the display follows from the fact that the law of $X - X_0$ under $\mathbb{P}_x$ is the same as the law of X under $\mathbb{P}_0$. By repeated conditioning, it follows that

$$\begin{aligned}
\mathbb{E}e^{-\lambda\tau_n} &= \mathbb{E}e^{-\lambda\tau_n} 1_{\{\tau_n < \infty\}} = \mathbb{E}\Big(e^{-\lambda\sum_{k=1}^{n}(\tau_k - \tau_{k-1})} 1_{\{\tau_n < \infty\}}\Big) \\
&= \mathbb{E}\Big(e^{-\lambda\sum_{k=1}^{n-1}(\tau_k - \tau_{k-1})} 1_{\{\tau_{n-1} < \infty\}} \mathbb{E}\Big(e^{-\lambda(\tau_n - \tau_{n-1})} \,|\, \mathcal{F}_{\tau_{n-1}}\Big)\Big) \\
&= a\mathbb{E}\Big(e^{-\lambda\sum_{k=1}^{n-1}(\tau_k - \tau_{k-1})} 1_{\{\tau_{n-1} < \infty\}}\Big) \\
&= \ldots = a^{n-1}\mathbb{E}e^{-\lambda\tau_1} 1_{\{\tau_1 < \infty\}} = a^n.
\end{aligned}$$

Now observe that by construction $\sup_{t \leq \tau_n} |X_t| \leq 2nC$, so

$$\mathbb{P}(|X_t| > 2nC) \leq \mathbb{P}(\tau_n < t) \leq \mathbb{P}(\exp(-\lambda\tau_n) > \exp(-\lambda t)) \leq e^{\lambda t} a^n.$$

This implies that X_t has finite moments of all orders. $\square$

The next theorem deals with two Lévy processes X and Y which never jump at the same time, meaning that $\Delta X_t \Delta Y_t = 0$ for all $t \geq 0$. One of the processes is assumed to have sample paths with finite variation on finite intervals. Recall that a function f on $\mathbb{R}_+$ is said to have this property if for every $t \geq 0$

$$\sup \sum |f(t_k) - f(t_{k-1})| < \infty,$$

where the supremum is taken over all finite partitions $0 = t_0 < \cdots < t_n = t$ of $[0, t]$. It is clear that for instance monotone functions and continuously differentiable functions have finite variation on finite intervals.

In the proof of the theorem we use that fact that if X is a Lévy process, the process M defined by

$$M_t = \frac{e^{iuX_t}}{\mathbb{E}e^{iuX_t}}$$

is a martingale (see Exercise 17).

Theorem 4.4.7. *Let X and Y be two Lévy processes with respect to the same filtration, and let one of them have sample paths with finite variation on finite intervals. If X and Y never jump at the same time, they are independent.*

Proof. In this proof we repeatedly use the simple fact that if Z is a martingale satisfying $\sup_{s \leq t} |Z_s| \leq C$ for some constant $C > 0$, and for $n \in \mathbb{N}$, π_n is a partition of $[0, t]$ given by $0 = t_0^n < \cdots < t_n^n = t$, then

$$\begin{aligned}\mathbb{E}\Big(\sum_k (Z_{t_k^n} - Z_{t_{k-1}^n})^2\Big)^2 &\lesssim \mathbb{E}\sum_k (Z_{t_k^n} - Z_{t_{k-1}^n})^4 \\ &\lesssim C^2 \mathbb{E}\sum_k (Z_{t_k^n} - Z_{t_{k-1}^n})^2 = C^2(\mathbb{E}Z_t^2 - \mathbb{E}Z_0^2).\end{aligned}$$

In particular, the sequence $\sum_k (Z_{t_k^n} - Z_{t_{k-1}^n})^2$ is bounded in L^2.

Now say that Y has finite variation on finite intervals and consider the martingales

$$M_t = \frac{e^{iuX_t}}{\mathbb{E}e^{iuX_t}} - 1, \quad N_t = \frac{e^{ivY_t}}{\mathbb{E}e^{ivY_t}} - 1.$$

Since N has finite variation on $[0, t]$ the sum $\sum_{s \leq t} |\Delta N_s|$ converges a.s. so we can write $N_t = \sum_{s \leq t} \Delta N_s + N_t^c$, where N^c is continuous. For $n \in \mathbb{N}$, let π_n be a partition of $[0, t]$ given by $0 = t_0^n < \cdots < t_n^n = t$ and suppose that $\|\pi_n\| = \sup_k |t_k^n - t_{k-1}^n| \to 0$ as $n \to \infty$. We have

$$\begin{aligned}&\sum_k (M_{t_k^n} - M_{t_{k-1}^n})(N_{t_k^n} - N_{t_{k-1}^n}) \\ &\quad = \sum_k (M_{t_k^n} - M_{t_{k-1}^n})(N_{t_k^n} - N_{t_{k-1}^n}) - \sum_{s \leq t} \Delta M_s \Delta N_s \\ &\quad = \sum_{s \leq t} (M_s^n - \Delta M_s)\Delta N_s + \sum_k (M_{t_k^n} - M_{t_{k-1}^n})(N^c_{t_k^n} - N^c_{t_{k-1}^n}),\end{aligned} \tag{4.9}$$

where $M_s^n = \sum_k (M_{t_k^n} - M_{t_{k-1}^n}) 1_{(t_{k-1}^n, t_k^n]}(s)$. By Cauchy-Schwarz the square of the second term on the right-hand side is bounded by

$$\sup_{\substack{|u-v| \leq \|\pi_n\| \\ u,v \leq t}} |N_u^c - N_v^c| \sum_k |N^c_{t_k^n} - N^c_{t_{k-1}^n}| \sum_k (M_{t_k^n} - M_{t_{k-1}^n})^2.$$

Since N^c is uniformly continuous and has finite variation on $[0,t]$, the product of the first two factors converges a.s. to 0. The last factor is bounded in L^2 by the argument in the first paragraph, whence the second term on the right-hand side of (4.9) tends to 0 in probability. As for the first term, observe that the sum is in fact countable and a.s. $M^n_s \to \Delta M_s$ for every $s \geq 0$. Since M is bounded and N has finite variation on $[0,t]$, it follows by dominated convergence that the first term on the right-hand side of (4.9) a.s. converges to 0.

All together we see that

$$\sum_k (M_{t^n_k} - M_{t^n_{k-1}})(N_{t^n_k} - N_{t^n_{k-1}})$$

converges to 0 in probability. Since by Cauchy-Schwarz and the first paragraph the sequence is bounded in L^2 and hence uniformly integrable (see Lemma A.3.4) we conclude that the convergence takes place in L^1 as well (Theorem A.3.5), so

$$\begin{aligned}\mathbb{E}M_tN_t &= \mathbb{E}\sum_k (M_{t^n_k} - M_{t^n_{k-1}})\sum_k (N_{t^n_k} - N_{t^n_{k-1}}) \\ &= \mathbb{E}\sum_k (M_{t^n_k} - M_{t^n_{k-1}})(N_{t^n_k} - N_{t^n_{k-1}}) \to 0.\end{aligned}$$

In view of the definitions of M and N this implies that

$$\mathbb{E}e^{iuX_t+ivY_t} = \mathbb{E}e^{iuX_t}\mathbb{E}e^{ivY_t},$$

so X_t and Y_t are independent. Since X and Y are Lévy processes, it follows that the processes are independent (see Exercise 18). □

The full jump structure of a Lévy process X is described by the random measure associated with its jumps. Formally, a *random measure* on a measurable space $(E, \mathcal{E})$ is a map $\mu : \Omega \times \mathcal{E} \to [0,\infty]$ such that for fixed $B \in \mathcal{E}$, $\mu(\cdot, B)$ is a (measurable) random variable and for fixed ω, $\mu(\omega, \cdot)$ is a measure on $(E, \mathcal{E})$. *The random measure μ^X associated with the jumps of the Lévy process* X is the random measure on $\mathbb{R}_+ \times \mathbb{R}\backslash\{0\}$ defined by

$$\mu^X(\omega, \cdot) = \sum_{t\geq 0} \delta_{(t, \Delta X_t(\omega))}.$$

We usually suppress the ω in the notation and simply write $\mu^X = \sum_{t\geq 0} \delta_{(t,\Delta X_t)}$. Observe that the sum is in fact countable and with N^B the counting process of the Borel set B, we have $\mu^X([0,t] \times B) = N^B_t$. This shows that $\mu^X([0,t] \times B)$ is indeed a random variable. By a standard approximation argument, μ^X is then seen to be a well-defined random measure on $(\mathbb{R}_+ \times \mathbb{R}\backslash\{0\}, \mathcal{B}(\mathbb{R}_+ \times \mathbb{R}\backslash\{0\}))$.

We will prove below that μ^X is a so-called Poisson random measure. The general definition is as follows.

Definition 4.4.8. Let μ be a random measure on $(E, \mathcal{E})$ and ν an ordinary measure. Then μ is called a *Poisson random measure* with *intensity measure* ν if

(i) for any disjoint sets $B_1, \dots, B_n \in \mathcal{E}$, the random variables $\mu(B_1), \dots, \mu(B_n)$ are independent,

(ii) if $\nu(B) < \infty$ then $\mu(B)$ has a Poisson distribution with parameter $\nu(B)$.

We have in fact already encountered Poisson random measures on the positive half-line.

Example 4.4.9. If N is a Poisson process with intensity $\lambda > 0$, we can define a random measure μ on $\mathbb{R}_+$ by setting $\mu(s,t] = N_t - N_s$ for $s \leq t$. The properties of the Poisson process show that μ is a Poisson random measure on $\mathbb{R}_+$ with intensity measure $\lambda\,\mathrm{Leb}$, where Leb denotes the Lebesgue measure. ■

Theorem 4.4.10. *The random measure μ^X associated with the jumps of a Lévy process X with Lévy measure ν^X is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}\backslash\{0\}$ with intensity measure Leb$\otimes\nu^X$.*

Proof. We provide the main ingredients of the proof, leaving the details as an exercise (see Exercise 21).

Suppose that $(s_1, t_1] \times B_1$ and $(s_2, t_2] \times B_2$ are disjoint Borel subsets of $\mathbb{R}_+ \times \mathbb{R}\backslash\{0\}$, and the sets B_i are bounded away from 0. Observe that $\mu^X((s_i, t_i] \times B_i) = N^{B_i}_{t_i} - N^{B_i}_{s_i}$. Since the sets are disjoint, either $(s_1, t_1]$ and $(s_2, t_2]$ are disjoint, or B_1 and B_2 are disjoint. In the first case, the independence of the random variables $\mu^X((s_i, t_i] \times B_i)$ follows from the independence of the increments of the processes N^{B_i}. In the second case the processes N^{B_i} never jump at the same time, so they are independent by Theorem 4.4.7. It is clear that this line of reasoning can be extended to deal with n disjoint sets $(s_i, t_i] \times B_i$ instead of just two. Appropriate approximation arguments complete the proof of the fact that μ^X satisfies the first requirement of Definition 4.4.8.

For a Borel subset of $\mathbb{R}_+ \times \mathbb{R}\backslash\{0\}$ of the form $(s,t] \times A$ we have that $\mu^X((s,t] \times A) = N^A_t - N^A_s$ is Poisson distributed with parameter $(t-s)\nu^X(A) = (\mathrm{Leb} \otimes \nu^X)((s,t] \times A)$. Using the fact that the sum of independent Poisson random variables is again Poisson and appropriate approximation arguments, it is straightforward to conclude that the second requirement of the definition is fulfilled. □

At several places below certain integrals with respect to the jump measure μ^X will appear. A measurable function f is called *μ^X-integrable* if for every $t \geq 0$,

$$\int_{(0,t]} \int_{\mathbb{R}\backslash\{0\}} |f(x)|\, \mu^X(ds, dx) < \infty$$

almost surely. For a μ^X-integrable function f we define the stochastic process $f * \mu^X$ by

$$f * \mu^X_t = \int_{(0,t]} \int_{\mathbb{R}\backslash\{0\}} f(x)\, \mu^X(ds, dx) = \sum_{s \leq t} f(\Delta X_s) 1_{\{\Delta X_s \neq 0\}}.$$

For a specific function f we often write $f(x) * \mu^X$ instead of $f * \mu^X$, for instance

$$x 1_{\{|x|>1\}} * \mu_t^X = \int_{(0,t]} \int_{|x|>1} x\, \mu^X(ds, dx) = \sum_{s \le t} \Delta X_s 1_{\{|\Delta X_s|>1\}}.$$

It is clear that $f * \mu^X$ is an adapted process. Observe that $1_B * \mu^X = N^B$, the counting process of the set B.

The following lemma collects some useful properties of integrals relative to the jump measure.

Lemma 4.4.11. *Let μ^X be the Poisson random measure associated with the jumps of the Lévy process X with Lévy measure ν^X, and let f be a measurable function.*

(i) The function f is μ^X-integrable if and only if $\int (|f| \wedge 1)\, d\nu^X < \infty$, and in that case

$$\mathbb{E} e^{iuf * \mu_t^X} = e^{-t \int (1 - e^{iuf})\, d\nu^X}.$$

(ii) If $f \in L^1(\nu^X)$ then $\mathbb{E}\, f * \mu_t^X = t \int f\, d\nu^X$.

(iii) If $f \in L^1(\nu^X) \cap L^2(\nu^X)$ then $\mathbb{V}\mathrm{ar}\, f * \mu_t^X = t \int f^2\, d\nu^X$.

(iv) If f is μ^X-integrable, the processes $f * \mu^X$ and $X - f * \mu^X$ are Lévy processes.

Proof. Recall that if the random variable Y has a Poisson distribution with parameter $\alpha > 0$, its Laplace transform is given by $\mathbb{E}\exp(-\lambda Y) = \exp(-\alpha(1 - \exp(-\lambda)))$. Let f be a simple, nonnegative function of the form $f = \sum_{i=1}^n c_i 1_{B_i}$, for certain disjoint Borel sets B_i and numbers $c_i \ge 0$. Then $f * \mu^X = \sum c_i N^{B_i}$ and hence, since the processes N^{B_i} are independent by Theorem 4.4.7,

$$\mathbb{E} e^{-\lambda (f * \mu^X)_t} = \prod \mathbb{E} e^{-c_i \lambda N_t^{B_i}} = \prod e^{-t\nu^X(B_i)(1 - e^{-c_i \lambda}))} = e^{-t \int (1 - e^{-\lambda f})\, d\nu^X}.$$

For an arbitrary nonnegative, measurable function f we can choose a sequence f_n of simple, nonnegative functions increasing to f. By monotone and dominated convergence we can then conclude that again,

$$\mathbb{E} e^{-\lambda (f * \mu^X)_t} = e^{-t \int (1 - e^{-\lambda f})\, d\nu^X}.$$

It follows that a measurable function f is μ^X-integrable if and only if

$$\lim_{\lambda \downarrow 0} \int (1 - e^{-\lambda |f|})\, d\nu^X = 0.$$

For $x \ge 0$ it holds that $1 - \exp(-x) \le x \wedge 1$, whence, for λ small enough, $|1 - \exp(-\lambda |f|)| \le \lambda |f| \wedge 1 \le |f| \wedge 1$. In view of the dominated convergence theorem, we see that $\int (|f| \wedge 1)\, d\nu < \infty$ is a sufficient condition for the μ^X-integrability of f. The fact that it is also necessary follows for instance from the inequality

$$\frac{1}{2}(x \wedge 1) \le (\log 2x) \wedge \frac{1}{2} \le 1 - e^{-x}$$

for $x \ge 0$, which follows from the concavity of $x \mapsto 1 - \exp(-x)$.

A similar computation as above yields the expression for the characteristic function, see Exercise 22.

Items (ii) and (iii) can be proved by first considering simple functions and using an approximation argument. The proofs of (ii), (iii) and (iv) are left to the reader, see Exercise 23. □

4.4.3 Lévy-Itô decomposition

Let X be a Lévy process with Lévy measure ν^X and $f \in L^1(\nu^X)$. Then by Lemma 4.4.11 the process $f * \mu_t^X - t \int f \, d\nu^X$ is a process with independent increments and constant expectation, whence it is a martingale. For convenience we denote this martingale by $f * (\mu^X - \nu^X)$.

For every $\varepsilon > 0$ we can write

$$X = (X - x1_{\{\varepsilon<|x|\le 1\}} * (\mu^X - \nu^X)) + x1_{\{\varepsilon<|x|\le 1\}} * (\mu^X - \nu^X). \tag{4.10}$$

Both processes on the right-hand side are Lévy processes by Lemma 4.4.11. The first one is obtained by subtracting from X its jumps with absolute value in $(\varepsilon, 1]$ and then subtracting $t \int_{\varepsilon<|x|\le 1} x \, \nu^X(dx)$ to ensure that we get a martingale. The jumps of the second process on the right-hand side are precisely those of X with absolute value outside $(\varepsilon, 1]$. Hence, by Theorem 4.4.7, the two processes are independent. It turns out as $\varepsilon \to 0$ they tend to well-defined, independent Lévy processes. The jumps of the first one equal the "small jumps" of X, while the jumps of the second one coincide with the "large jumps" of X.

Theorem 4.4.12. *Let X be a Lévy process with Lévy measure ν^X. Then $\int (x^2 \wedge 1) \, \nu^X(dx) < \infty$ and for $t \ge 0$ the limit*

$$M_t = \lim_{\varepsilon \downarrow 0} x1_{\{\varepsilon<|x|\le 1\}} * (\mu^X - \nu^X)_t$$

exists in L^2. The processes $X - M$ and M are independent Lévy processes. It holds that $\Delta M_t = \Delta X_t 1_{\{|\Delta X_t| \le 1\}}$ and $\Delta(X - M)_t = \Delta X_t 1_{\{|\Delta X_t|>1\}}$.

Proof. Define $Y = X - x1_{\{|x|>1\}} * \mu^X$. By part (iv) of Lemma 4.4.11, Y is a Lévy process. Observe that $\Delta Y_t = \Delta X_t 1_{\{|\Delta X_t| \le 1\}}$. Hence by Theorem 4.4.6 the process Y has finite moments of all orders, and it is easily seen that the Lévy measure ν^Y of Y is given by $\nu^Y(B) = \nu^X(B \cap [-1, 1])$ (see Exercise 24). For $\varepsilon > 0$, define $Y^\varepsilon = x1_{\{\varepsilon<|x|\le 1\}} * \mu^Y = x1_{\{\varepsilon<|x|\le 1\}} * \mu^X$. By Lemma 4.4.11 and monotone convergence we have

$$\mathbb{V}\mathrm{ar}\, Y_1^\varepsilon = \int_{\varepsilon<|x|\le 1} x^2 \, \nu^X(dx) \uparrow \int_{|x|\le 1} x^2 \, \nu^X(dx)$$

as $\varepsilon \to 0$. Lemma 4.4.11 also implies that $Y - Y^\varepsilon$ is independent of Y^ε, so that $\mathbb{V}\mathrm{ar}\, Y_1^\varepsilon \le \mathbb{V}\mathrm{ar}\, Y_1 < \infty$. It follows that

$$\int_{|x|\le 1} x^2 \, \nu^X(dx) < \infty,$$

which proves the first statement of the theorem.

Now put $M^\varepsilon = x1_{\{\varepsilon<|x|\le 1\}} * (\mu^X - \nu^X)$. Then for $t \ge 0$ and $\varepsilon' < \varepsilon < 1$ we have

$$\mathbb{E}(M_t^\varepsilon - M_t^{\varepsilon'})^2 = \int_{\varepsilon'<|x|\le\varepsilon} x^2\,\nu^X(dx).$$

Hence, in view of the first paragraph of the proof, the random variables M_t^ε are Cauchy in L^2. So as $\varepsilon \to 0$, M_t^ε converges to some random variable M_t in L^2. The process M inherits the Lévy property from M^ε (see Exercise 25) and $(\mathbb{E}M_t)^2 \le \mathbb{E}(M_t^\varepsilon - M_t)^2 \to 0$, so M is centered. In particular, M is a martingale. By Theorem 4.4.7, $X - x1_{\{\varepsilon<|x|\le 1\}} * \mu^X$ and $x1_{\{\varepsilon<|x|\le 1\}} * \mu^X$ are independent, hence $X - M^\varepsilon$ and M^ε are independent as well. By letting $\varepsilon \to 0$ we see that $X - M$ and M are independent.

To complete the proof, note that by Doob's L^2-inequality (Theorem 2.3.11) $\mathbb{E}(\sup_{s\le t}|M_s^\varepsilon - M_s|)^2 \le 4\mathbb{E}(M_t^\varepsilon - M_t)^2 \to 0$, so there exists a sequence $\varepsilon_n \to 0$ such that a.s. $\sup_{s\le t}|(Y - M^{\varepsilon_n})_s - (Y-M)_s| \to 0$. By construction, $|\Delta(Y - M^\varepsilon)_u| \le \varepsilon$ for all $u \ge 0$. It follows that the process $Y - M$ is continuous, hence $X - M$ is the sum of a continuous process and $x1_{\{|x|>1\}} * \mu^X$. In particular, the jumps of $X - M$ are precisely the jumps of X which are larger than 1 in absolute value. □

The limit process M of the preceding theorem is from now on denoted by

$$M = x1_{\{|x|\le 1\}} * (\mu^X - \nu^X).$$

According to the theorem it is a Lévy process and $\Delta M_t = \Delta X_t 1_{\{|\Delta X_t|\le 1\}}$, so its Lévy measure is given by $B \mapsto \nu^X(B \cap [-1,1])$.

The following theorem proves the claim made in Example 4.4.2.

Theorem 4.4.13. *A continuous Lévy process X is a Brownian motion with linear drift, i.e. $X_t = bt + \sigma W_t$ for certain numbers $\mu \in \mathbb{R}$, $\sigma \ge 0$ and a standard Brownian motion W.*

Proof. By Theorem 4.4.6 the process X has moments of all orders, so there exist $b \in \mathbb{R}$ and $\sigma \ge 0$ such that $\mathbb{E}X_t = bt$ and $\mathbb{V}\mathrm{ar}X_t = \sigma^2 t$ (see Subsection 4.4.1). If we define $Y_t = X_t - bt$, it follows easily that under $\mathbb{P}_x$, the processes $Y_t - x$ and $(Y_t - x)^2 - \sigma^2 t$ are centered martingales. For $a \in \mathbb{R}$, let τ_a be the first time that the process Y hits the level a. By arguing exactly as we did for the BM in Exercises 14 and 15 of Chapter 2 we find that for $a < x < b$,

$$\mathbb{P}_x(\tau_a < \tau_b) = \frac{b-x}{b-a}, \quad \mathbb{E}_x\tau_a \wedge \tau_b = -\frac{(a-x)(b-x)}{\sigma^2}.$$

Now consider the escape time $\eta_r = \inf\{t : |Y_t - Y_0| > r\}$. Under $\mathbb{P}_x$ we have that Y_{η_r} is $x \pm r$ and

$$\mathbb{P}_x(Y_{\eta_x} = x - r) = \mathbb{P}_x(\tau_{x-r} < \tau_{x+r}) = \frac{r}{2r} = \frac{1}{2}.$$

Also note that $\mathbb{E}_x\eta_r = \mathbb{E}_x\tau_{x-r} \wedge \tau_{x+r} = r^2/\sigma^2$. Hence, using Taylor's formula we see that for $f \in C_0^2$ we have

$$\begin{aligned}\frac{\mathbb{E}_x f(Y_{\eta_r}) - f(x)}{\mathbb{E}_x\eta_r} &= \sigma^2 \frac{\frac{1}{2}f(x+r) + \frac{1}{2}f(x-r) - f(x)}{r^2} \\ &= \frac{1}{4}\sigma^2(f''(y_1) + f''(y_2)),\end{aligned}$$

where $y_1, y_2 \in [x-r, x+r]$. By Theorem 3.5.12 it follows that for the generator A of the process Y and $f \in C_0^2$ it holds that $Af = \sigma^2 f''/2$. This is precisely the generator of σW, where W is a standard Brownian motion (see Example 3.5.7). □

We have now collected all ingredients for the proof of the main result of this subsection.

Theorem 4.4.14 (Lévy-Itô decomposition). *Let X be a Lévy process with Lévy measure ν^X and jump measure μ^X. Then $\int (x^2 \wedge 1)\, \nu^X(dx) < \infty$ and*

$$X_t = bt + \sigma W_t + x1_{\{|x|\le 1\}} * (\mu^X - \nu^X)_t + x1_{\{|x|>1\}} * \mu^X_t,$$

where $\mu \in \mathbb{R}$, $\sigma \ge 0$ and W is a standard Brownian motion. The three processes on the right-hand side are independent. If $\int (|x| \wedge 1)\, \nu^X(dx) < \infty$ the decomposition simplifies to

$$X_t = ct + \sigma W_t + x * \mu^X_t,$$

with $c = b + \int_{|x|\le 1} x\, \nu^X(dx)$.

Proof. By Lemma 4.4.11 the processes $Y = X - x1_{\{|x|>1\}} * \mu^X$ and $x1_{\{|x|>1\}} * \mu^X$ are Lévy processes. Since the latter has finite variation on finite intervals and they jump at different times, the processes are independent by Theorem 4.4.7. Note that the jump measure and Lévy measure of Y are given by $\mu^Y = \sum_{s\le t} \delta_{(t,\Delta X_t)} 1_{\{|\Delta X_t|\le 1\}}$ and $\nu^Y(B) = \nu^X(B \cap [-1,1])$, respectively. By Theorem 4.4.12 and the process Y can be written the sum of

$$x1_{\{|x|\le 1\}} * (\mu^Y - \nu^Y) = x1_{\{|x|\le 1\}} * (\mu^X - \nu^X)$$

and an independent Lévy process without jumps. According to Theorem 4.4.13 the latter is a Brownian motion with drift. This completes the proof of the first statement of the theorem. The proof of the second statement is left as Exercise 26. □

The triplet (b, σ^2, ν^X) of appearing in the decomposition theorem is called the *characteristic triplet* of the Lévy process X. It is clear from the theorem that it is uniquely determined by X. The first process in the decomposition is a Brownian motion with drift, and the third one is a compound Poisson process (see Exercise 27). The second one can be viewed as a mixture of independent, compensated Poisson processes. Observe that the triplets of the BM with drift, the Poisson process and the compound Poisson process of Examples 4.4.2 - 4.4.4 are given by $(b, \sigma^2, 0)$, $(0, 0, \lambda\delta_1)$ and $(0, 0, \lambda\, dF)$, respectively. The Lévy measure of the first passage time process of Example 4.4.5 is computed in Example 4.4.20 below.

The Lévy-Itô decomposition can be used to characterize certain properties of the process X in terms of its triplet. Here is an example (see also Exercise 28 for another example).

Corollary 4.4.15. *A Lévy process with characteristic triplet (b, σ^2, ν) has finite variation on finite intervals if and only if $\sigma^2 = 0$ and $\int (|x| \wedge 1)\, \nu(dx) < \infty$.*

Proof. Sufficiency follows immediately from Theorem 4.4.14.

Suppose now that X has finite variation on finite intervals. Then in particular $\sum_{s \le t} |\Delta X_s| < \infty$ a.s. for every $t \ge 0$. Hence, by Lemma 4.4.11, $\int (|x| \wedge 1)\, \nu(dx) < \infty$. Theorem 4.4.14 then implies that

$$X_t = ct + \sigma W_t + \sum_{s \le t} \Delta X_s,$$

where W is a Brownian motion. It follows that the process σW has finite variation on finite intervals. By Corollary 2.4.2 this can only happen if $\sigma^2 = 0$. □

A Lévy process X is called a *subordinator* if it has increasing sample paths. In particular the sample paths of a subordinator have finite variation on finite intervals, so by the preceding corollary the Lévy measure satisfies $\int (|x| \wedge 1)\, \nu(dx) < \infty$ and we can write

$$X_t = ct + \sum_{s \le t} \Delta X_s.$$

The number c is called the *drift coefficient* of the process. All jumps of a subordinator must be positive, hence its Lévy measure ν is concentrated on $(0, \infty)$. To ensure that the process is increasing between jumps, we must have $c \ge 0$. We say that (c, ν) are the *characteristics* of the subordinator X. Clearly, they are uniquely determined by the process X. Since a subordinator is nonnegative, it is convenient to work with Laplace transforms instead of characteristic functions. The Lévy property implies that for $\lambda, t \ge 0$ it holds that $\mathbb{E} \exp(-\lambda X_t) = \exp(-t\varphi(\lambda))$ for some $\varphi : \mathbb{R}_+ \to \mathbb{R}_+$. The function φ is called the *Laplace exponent* of the subordinator X.

4.4.4 Lévy-Khintchine formula

Below we derive the Lévy-Khintchine formula, which gives an expression for the characteristic exponent in terms of the triplet. We need the following lemma to deal with the second process in the Lévy-Itô decomposition.

Lemma 4.4.16. *Let X be a Lévy process with jump measure μ^X and Lévy measure ν^X. Then*

$$\mathbb{E}e^{iu\,x1_{\{|x|\le 1\}}*(\mu^X-\nu^X)_t} = \exp\Big(-t\int_{|x|\le 1}\Big(1-iux-e^{iux}\Big)\,\nu^X(dx)\Big)$$

Proof. The random variable $M_t = x1_{\{|x|\le 1\}} * (\mu^X - \nu^X)_t$ is the L^2-limit of

$$M_t^\varepsilon = x1_{\{\varepsilon<|x|\le 1\}} * \mu_t^X - t\int_{\varepsilon<|x|\le 1} x\,\nu^X(dx)$$

for $\varepsilon \downarrow 0$, see Theorem 4.4.12. By Lemma 4.4.11 the characteristic function of M_t^ε is given by

$$\mathbb{E}e^{iuM_t^\varepsilon} = \exp\Big(-t\int_{\varepsilon<|x|\le 1}\Big(1-iux-e^{iux}\Big)\,\nu^X(dx)\Big).$$

From the power series expansion of the exponential we see that $|1-z-e^z| \le |z|^2e^{|z|}$ and hence $|1-iux-e^{iux}| \le u^2e^ux^2$ for $|x|\le 1$. Since $x\mapsto x^21_{\{|x|\le 1\}}$ is ν^X-integrable we can now apply the dominated convergence theorem to complete the proof. □

Theorem 4.4.17 (Lévy-Khintchine formula). *Let X be a Lévy process with characteristic triplet (b,σ^2,ν). Then its characteristic exponent ψ is given by*

$$\psi(u) = -ibu + \tfrac{1}{2}\sigma^2u^2 + \int_{\mathbb{R}}\Big(1-e^{iu}+iu1_{\{|x|\le 1\}}\Big)\,\nu(dx).$$

If X is a subordinator with characteristics (c,ν), its Laplace exponent is given by

$$\varphi(\lambda) = c\lambda + \int_0^\infty (1-e^{-\lambda x})\,\nu(dx).$$

Proof. The first part of the theorem is a straightforward consequence of the Lévy-Itô decomposition and the expressions that we have for the characteristic exponents of the three independent processes appearing in the decomposition (see Examples 4.4.2, 4.4.3 and Lemma 4.4.16).

For the second statement of the Theorem write $X_t = ct + x*\mu_t^X$ and use the fact that for a Borel function $f\ge 0$,

$$\mathbb{E}e^{-\lambda(f*\mu^X)_t} = e^{-t\int(1-e^{-\lambda f})\,d\nu^X},$$

see the proof of Lemma 4.4.11. □

As we observed above, the characteristic exponent of a Lévy process determines its distribution. Hence, the Lévy-Khintchine formula shows that the distribution of a Lévy process (or subordinator) is determined by its characteristics.

4.4.5 Stable processes

A Lévy process X is called a *stable process* with *index* $\alpha > 0$ if for all $r > 0$, the process $(r^{-1/\alpha} X_{rt})_{t \geq 0}$ has the same distribution as X. Processes with this scaling property are also called *self-similar* (with index $1/\alpha$). The term "stable" is usually reserved for self-similar Lévy processes. Observe that in terms of the characteristic exponent ψ of X, α-stability means that $\psi(ur) = r^\alpha \psi(u)$ for all $r > 0$ and $u \in \mathbb{R}$. By the scaling property, the Brownian motion is a stable process with index 2. The deterministic process $X_t = bt$ is stable with index 1.

The scaling property of the BM also implies that the process of first passage times of Example 4.4.5 is stable.

Example 4.4.18. Consider the first passage times process $(\sigma_a)_{a \geq 0}$ of the BM. Then

$$r^{-2}\sigma_{ra} = r^{-2} \inf\{t : W_t > ra\} = \inf\{t : r^{-1} W_{r^2 t} > a\}.$$

Hence, by the scaling property of the BM, the processes $(r^{-2}\sigma_{ra})_{a \geq 0}$ and $(\sigma_a)_{a \geq 0}$ have the same distribution, so the first passage times process is stable with index $1/2$. ■

The following theorem characterizes stable processes in terms of their triplets. We exclude the degenerate, deterministic processes with triplets of the form $(b, 0, 0)$. Trivially, they are 1-stable if $b \neq 0$ and α-stable for any α if $b = 0$. We call a subordinator with characteristics (c, ν) degenerate if $\nu = 0$.

Theorem 4.4.19. *Let X be a non-degenerate Lévy process with characteristic triplet (b, σ^2, ν).*

(i) If X is stable with index α, then $\alpha \in (0, 2]$.

(ii) The process X is stable with index 2 if and only if $b = 0$ and $\nu = 0$.

(iii) The process X is stable with index 1 if and only if $\sigma^2 = 0$ and $\nu(dx) = C_\pm |x|^{-2}\, dx$ on $\mathbb{R}_\pm$, for certain constants $C_\pm \geq 0$.

(iv) The process X is stable with index $\alpha \in (0, 1) \cup (1, 2)$ if and only if $b = 0$, $\sigma^2 = 0$ and $\nu(dx) = C_\pm |x|^{-1-\alpha}\, dx$ on $\mathbb{R}_\pm$, for certain constants $C_\pm \geq 0$.

A non-degenerate subordinator with characteristics (c, ν) is α-stable if and only $\alpha \in (0, 1)$, $c = 0$ and $\nu(dx) = Cx^{-1-\alpha}\, dx$ for some $C \geq 0$.

Proof. The process X is α-stable if and only if for all $r > 0$, the processes $(X_{r^\alpha t})_{t \geq 0}$ and $(rX_t)_{t \geq 0}$ have the same triplets. These triplets are given by $(r^\alpha b, r^\alpha \sigma^2, r^\alpha \nu)$ and $(rb, r^2\sigma^2, \nu(\frac{1}{r}\,\cdot))$, respectively (check!). Hence, X is stable with index α if and only if for all $r > 0$: $r^\alpha b = rb$, $r^\alpha \sigma^2 = r^2 \sigma^2$, and $r^\alpha \nu(\cdot) = \nu(\frac{1}{r}\,\cdot)$. The conditions on the triplets given in the statement of the theorem are now easily seen to be sufficient. It remains to prove that they are also necessary.

If the Lévy measures $r^\alpha \nu(\cdot)$ and $\nu(\frac{1}{r}\,\cdot)$ are equal then for $x > 0$ we have $r^\alpha \nu[x, \infty) = \nu[x/r, \infty)$ for all $r > 0$. Taking $r = x$ we see that

$$\nu[x, \infty) = \frac{1}{x^\alpha}\nu[1, \infty) = C_+ \int_x^\infty \frac{1}{y^{1+\alpha}}\, dy,$$

with $C_+ = \nu[1, \infty)/\alpha$, hence $\nu(dx) = C_+ x^{-1-\alpha}\, dx$ on $\mathbb{R}_+$. If we replace $[1, \infty)$ by $(-\infty, -1]$ and repeat the argument we see that if X is α-stable, then $\nu(dx) = C_-|x|^{-1-\alpha}\, dx$ on $\mathbb{R}_-$, for some constant $C_- \geq 0$.

Now if X is stable with index 2, then clearly we must have $b = 0$ and the preceding paragraph shows that $\nu(dx) = C_\pm |x|^{-3}\, dx$ on $\mathbb{R}_\pm$, for certain constants $C_\pm \geq 0$. Since it holds that $\int_{-1}^1 x^2\, \nu(dx) < \infty$ (see Theorem 4.4.12), we must have $\nu = 0$, proving (ii). For the proof of (i), (iii) and (iv), suppose X is stable with an index $\alpha \neq 2$. Then we must have $\sigma^2 = 0$ and $\nu(dx) = C_\pm |x|^{-1-\alpha}\, dx$ on $\mathbb{R}_\pm$, for certain constants $C_\pm \geq 0$. Since we assumed that X is non-degenerate it holds that $\nu \neq 0$. The fact that we have $\int_{-1}^1 x^2\, \nu(dx) < \infty$ then forces $\alpha \in (0, 2)$. If $\alpha \neq 1$ then clearly we must have $b = 0$ as well.

You are asked to provide the proof for the subordinator case in Exercise 29. □

We can now compute the characteristic of the process of first passage times of the Brownian motion.

Example 4.4.20. The process $(\sigma_a)_{a \geq 0}$ of first passage times of a standard BM W is a subordinator with certain characteristics (c, ν). We saw in Example 4.4.18 that the process is stable with index 1/2. Hence, by Theorem 4.4.19 we have $c = 0$ and $\nu(dx) = Cx^{-3/2}\, dx$ for some $C \geq 0$. It remains to determine the constant C. By Theorem 4.4.17 we have

$$\mathbb{E}e^{-\sigma_1} = \exp\Big(-C\int_0^\infty \frac{1 - e^{-x}}{x^{3/2}}\, dx\Big) = \exp\Big(-2C\int_0^\infty \frac{e^{-x}}{\sqrt{x}}\, dx\Big) = e^{-2\sqrt{\pi}C}.$$

Here the second and third equalities follow from integration by parts and a substitution $x = \frac{1}{2}y^2$, respectively. On the other hand, we a.s. have that $\sigma_a = \tau_a = \inf(t : W_t = a\}$ (see Example 4.4.5) and hence, by Theorem 2.4.7,

$$\mathbb{E}e^{-\sigma_1} = \mathbb{E}e^{-\tau_1} = e^{-\sqrt{2}}.$$

It follows that $C = (2\pi)^{-1/2}$, so $\nu(dx) = (2\pi)^{-1/2}x^{-3/2}\, dx$. The Laplace exponent is given by $\varphi(\lambda) = \sqrt{2\lambda}$. ■

4.5 Exercises

1. Let W be the canonical version of the BM^d. Show that the law of W under $\mathbb{P}_x$ is equal to the law of $x + W$ under $\mathbb{P}_0$.

2. Prove the second statement of Theorem 4.1.2.

3. Derive from Corollary 4.1.3 that for $d \geq 3$, it holds with probability one that the BM^d ultimately leaves every bounded set forever. (Hint: for $0 < a < b$, consider the stopping times

$$\begin{aligned}
\tau_a^{(1)} &= \tau_a \\
\tau_b^{(1)} &= \inf\{t > \tau_a^{(1)} : \|W_t\| = b\} \\
\tau_a^{(2)} &= \inf\{t > \tau_b^{(1)} : \|W_t\| = a\} \\
&\vdots
\end{aligned}$$

Using the strong Markov property, show that if $\|x\| = b$, then

$$\mathbb{P}_x\left(\tau_a^{(n)} < \infty\right) = \left(\frac{a}{\|x\|}\right)^{(d-2)n}.$$

Conclude that $\mathbb{P}_x(\tau_a^{(n)} < \infty \text{ for all } n) = 0$.)

4. Let W be a BM^d, σ an invertible $d \times d$ matrix and define $X_t = \sigma W_t$. Show that X is a Feller process and compute its generator.

5. Let X be a Feller process on $\mathbb{R}^d$ with generator A. For $b \in \mathbb{R}^d$, define $Y_t = bt + X_t$ and let B be the generator of Y. Show that for $f \in \mathcal{D}(A) \cap C_0^2(\mathbb{R}^d)$, it holds that

$$Bf(x) = Af(x) + \sum_i b_i \frac{\partial f}{\partial x_i}(x).$$

6. Complete the proof of Theorem 4.2.5.

7. Verify the assertions in Example 4.2.6.

8. Prove that the generator A of a Poisson process with intensity λ is given by $Af(x) = \lambda(f(x+1) - f(x))$ for $f \in C_0$.

9. Prove Theorem 4.3.3.

10. Let Q be a transition kernel on E and $\lambda > 0$. Show that (4.8) defines a transition semigroup.

11. Show that the semigroup (P_t) in the proof of Theorem 4.3.5 is Feller.

12. Give a "continuous-time proof" of the fact that the ordinary random walk on $\mathbb{Z}$ is recurrent. (Hint: consider the continuous-time random walk X on $\mathbb{Z}$. Fix $a \leq x \leq b$ in $\mathbb{Z}$ and consider a function $f \in C_0$ such that $f(y) = y$

for $a \leq y \leq b$. As in the proof of Theorem 4.1.2, use Dynkin's formula with this f to show that

$$\mathbb{P}_x(\tau_a < \tau_b) = \frac{b-x}{b-a}.$$

Complete the proof by arguing as in the proof of 4.1.3.)

13. Show that the characteristic exponent of a Lévy process determines all finite dimensional distributions.

14. Show that if X is a Lévy process with finite second moments, then $\mathbb{V}\mathrm{ar} X_t = \sigma^2 t$ for some $\sigma^2 \geq 0$. (Hint: for a proof without using the characteristic function, show first that the claim holds for $t \in \mathbb{Q}$. Then for arbitrary $t \geq 0$, consider a sequence of rationals $q_n \downarrow t$ and prove that X_{q_n} is a Cauchy sequence in L^2. Finally, use the continuity in probability.)

15. Show that any Lévy process is a Feller process.

16. Show that the compound Poisson process is a Lévy process and compute its characteristic exponent.

17. If X is a Lévy process, show that $M_t = \exp(iuX_t)/\mathbb{E}\exp(iuX_t)$ is a martingale.

18. Let X and Y be two Lévy processes with respect to the same filtration. Show that if X_t and Y_t are independent for each $t \geq 0$, then the processes X and Y are independent.

19. By constructing a counter example, show that in Theorem 4.4.7, the assumption that one of the processes has finite variation on finite intervals can not be dropped.

20. By constructing a counter example, show that in Theorem 4.4.7, the assumption that X and Y are Lévy processes with respect with the same filtration can not be dropped. (Hint: keeping the strong Markov property in mind, construct two suitable Poisson processes w.r.t. different filtrations.)

21. Work out the details of the proof of Theorem 4.4.10.

22. Derive the expression for the characteristic function given in Lemma 4.4.11. (Hint: use the bound $|1 - \exp(ix)|^2 = 2(1 - \cos x) \leq 2(x^2 \wedge 1)$.)

23. Prove items (ii), (iii) and (iv) of Lemma 4.4.11.

24. Let X be a Lévy process with Lévy measure ν^X and f a μ^X-integrable function. Determine the Lévy measures of $f * \mu^X$ and $X - f * \mu^X$.

25. Let X^n be a sequence of Lévy processes relative to the same filtration and for every $t \geq 0$, $X^n_t \to X_t$ in L^2. Then X is a Lévy process as well.

26. Proof the second statement of Theorem 4.4.14.

27. Let X be a Lévy process. Show that $\sum_{s \leq t} \Delta X_s 1_{\{|\Delta X_s| > 1\}}$ is a compound Poisson process.

28. Let X be a Lévy process with Lévy measure ν. Show that X has finite first moments if and only $\int (|x| \wedge x^2)\, \nu(dx) < \infty$ and X has finite second moments if and only $\int x^2\, \nu(dx) < \infty$.

29. Complete the proof of Theorem 4.4.19.

A

Elements of measure theory

A.1 Definition of conditional expectation

The following theorem justifies the definition of the conditional expectation.

Theorem A.1.1. *Let X be an integrable random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra. Then there exists an integrable random variable Y such that*

(i) Y is $\mathcal{G}$-measurable,

(ii) for all $A \in \mathcal{G}$

$$\int_A X \, d\mathbb{P} = \int_A Y \, d\mathbb{P}.$$

If Y' is another random variable with these properties, then $Y = Y'$ almost surely.

Proof. To prove existence, suppose first that $X \geq 0$. Define the measure $\mathbb{Q}$ on $(\Omega, \mathcal{G})$ by

$$\mathbb{Q}(A) = \int_A X \, d\mathbb{P}, \quad A \in \mathcal{G}.$$

Since X is integrable, the measure $\mathbb{Q}$ is finite. Note that for all $A \in \mathcal{G}$, we have that $\mathbb{P}(A) = 0$ implies that $\mathbb{Q}(A) = 0$, in other words, $\mathbb{Q}$ is absolutely continuous with respect to the restriction of $\mathbb{P}$ to $\mathcal{G}$. By the Radon-Nikodym theorem, it follows that there exists an integrable, measurable function Y on $(\Omega, \mathcal{G})$ such that

$$\mathbb{Q}(A) = \int_A Y \, d\mathbb{P}, \quad A \in \mathcal{G}.$$

Clearly, Y has the desired properties. The general case follows by linearity.

Now suppose that Y and Y' are two integrable random variables that both satisfy (i) and (ii). For $n \in \mathbb{N}$, define the event $A_n = \{Y - Y' \geq 1/n\}$. Then $A_n \in \mathcal{G}$, so

$$0 = \int_{A_n} (Y - Y')\, d\mathbb{P} \geq \frac{1}{n}\mathbb{P}(A_n),$$

hence $\mathbb{P}(A_n) = 0$. It follows that

$$\mathbb{P}(Y > Y') \leq \sum_{n=1}^{\infty} \mathbb{P}(A_n) = 0,$$

so $Y \leq Y'$, almost surely. By switching the roles of Y and Y' we see that it also holds that $Y' \leq Y$, almost surely. □

Definition A.1.2. An integrable random variable Y that satisfies conditions (i) and (ii) of Theorem A.1.1 is called *(a version of) the conditional expectation of X given $\mathcal{G}$*. We write $Y = \mathbb{E}(X \mid \mathcal{G})$, almost surely.

If the σ-algebra $\mathcal{G}$ is generated by another random variable, say $\mathcal{G} = \sigma(Z)$, then we usually write $\mathbb{E}(X \mid Z)$ instead of $\mathbb{E}(X \mid \sigma(Z))$. Similarly, if $Z_1, Z_2, \ldots$ is a sequence of random variable, then we write $\mathbb{E}(X \mid Z_1, Z_2, \ldots)$ instead of $\mathbb{E}(X \mid \sigma(Z_1, Z_2, \ldots))$.

Note that if X is square integrable, then $\mathbb{E}(X \mid \mathcal{G})$ is simply the orthogonal projection in $L^2(\Omega, \mathcal{F}, \mathbb{P})$ of X on the closed subspace $L^2(\Omega, \mathcal{G}, \mathbb{P})$. In other words, it is the $\mathcal{G}$-measurable random variable that is closest to X in mean square sense. This shows that we should think of the conditional expectation $\mathbb{E}(X \mid \mathcal{G})$ as the 'best guess of X that we can make on the basis of the information in $\mathcal{G}$'.

A.2 Basic properties of conditional expectation

The following properties follow either immediately from the definition, or from the corresponding properties of ordinary expectations.

Theorem A.2.1.

(i) *If X is $\mathcal{G}$-measurable, then $\mathbb{E}(X \mid \mathcal{G}) = X$.*

(ii) $\mathbb{E}(X \mid \{\varnothing, \Omega\}) = \mathbb{E}X$ *a.s.*

(iii) *Linearity:* $\mathbb{E}(aX + bY \mid \mathcal{G}) \stackrel{\text{as}}{=} a\mathbb{E}(X \mid \mathcal{G}) + b\mathbb{E}(Y \mid \mathcal{G})$.

(iv) *Positivity: if $X \geq 0$ a.s., then $\mathbb{E}(X \mid \mathcal{G}) \geq 0$ a.s.*

(v) *Monotone convergence: if $X_n \uparrow X$ a.s., then $\mathbb{E}(X_n \mid \mathcal{G}) \uparrow \mathbb{E}(X \mid \mathcal{G})$ a.s.*

(vi) *Fatou: it $X_n \geq 0$ a.s., then $\mathbb{E}(\liminf X_n \,|\, \mathcal{G}) \leq \liminf \mathbb{E}(X_n \,|\, \mathcal{G})$ a.s.*

(vii) *Dominated convergence: suppose that $|X_n| \leq Y$ a.s. and $\mathbb{E}Y < \infty$. Then if $X_n \to X$ a.s., it follows that $\mathbb{E}(X_n \,|\, \mathcal{G}) \to \mathbb{E}(X \,|\, \mathcal{G})$ a.s.*

(viii) *Jensen: if φ is a convex function such that $\mathbb{E}|\varphi(X)| < \infty$, then $\mathbb{E}(\varphi(X) \,|\, \mathcal{G}) \geq \varphi(\mathbb{E}(X \,|\, \mathcal{G}))$ a.s.*

(ix) *Tower property: if $\mathcal{G} \subseteq \mathcal{H}$, then $\mathbb{E}(X \,|\, \mathcal{G}) = \mathbb{E}(\mathbb{E}(X \,|\, \mathcal{H}) \,|\, \mathcal{G})$ a.s.*

(x) *Taking out what is known: if Y is $\mathcal{G}$-measurable and bounded, then $\mathbb{E}(YX \,|\, \mathcal{G}) = Y\mathbb{E}(X \,|\, \mathcal{G})$ a.s.*

(xi) *Role of independence: if $\mathcal{H}$ is independent of $\sigma(\sigma(X), \mathcal{G})$, then $\mathbb{E}(X \,|\, \sigma(\mathcal{H}, \mathcal{G})) = \mathbb{E}(X \,|\, \mathcal{G})$ a.s.*

Proof. Exercise! The proofs of parts (iv), (viii), (x) and (xi) are the most challenging. See Williams (1991). □

A.3 Uniform integrability

For an integrable random variable X, the map $A \mapsto \int_A |X| \, d\mathbb{P}$ is "continuous" in the following sense.

Lemma A.3.1. *Let X be an integrable random variable. Then for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $A \in \mathcal{F}$, $\mathbb{P}(A) \leq \delta$ implies that*

$$\int_A |X| \, d\mathbb{P} < \varepsilon.$$

Proof. See for instance Williams (1991), Lemma 13.1. □

Roughly speaking, a class of random variables is called uniformly integrable if this continuity property holds uniformly over the entire class. The precise definition is as follows.

Definition A.3.2. Let $\mathcal{C}$ be an arbitrary collection of random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We call the collection *uniformly integrable* if for every $\varepsilon > 0$ there exists a $K \geq 0$ such that

$$\int_{|X|>K} |X| \, d\mathbb{P} \leq \varepsilon, \quad \text{for all } X \in \mathcal{C}.$$

The following lemma gives important examples of uniform integrable classes.

Lemma A.3.3. *Let $\mathcal{C}$ be a collection of random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.*

(i) If the class $\mathcal{C}$ is bounded in $L^p(\mathbb{P})$ for some $p > 1$, then $\mathcal{C}$ is uniformly integrable.

(ii) If $\mathcal{C}$ is uniformly integrable, then $\mathcal{C}$ is bounded in $L^1(\mathbb{P})$.

Proof. Exercise, see Williams (1991). □

Conditional expectations give us the following important example of a uniformly integrable class.

Lemma A.3.4. *Let X be an integrable random variable. Then the class*

$$\{\mathbb{E}(X \,|\, \mathcal{G}) : \mathcal{G} \text{ is a sub-}\sigma\text{-algebra of } \mathcal{F}\}$$

is uniformly integrable.

Proof. Exercise. See Williams (1991), Section 13.4. □

Uniform integrability is what is needed to strengthen convergence in probability to convergence in L^1.

Theorem A.3.5. *Let $(X_n)_{n\in\mathbb{N}}$ and X be integrable random variables. Then $X_n \to X$ in L^1 if and only if*

(i) $X_n \to X$ in probability, and

(ii) the sequence (X_n) is uniformly integrable.

Proof. See Williams (1991), Section 13.7. □

A.4 Monotone class theorem

There exist many formulations of the monotone class theorem. In Chapter 3 we use a version for monotone classes of sets.

Definition A.4.1. A collection $\mathcal{S}$ of subsets of Ω is called a *monotone class* if

(i) $\Omega \in \mathcal{S}$,

(ii) if $A, B \in \mathcal{S}$ and $A \subseteq B$, then $B \backslash A \in \mathcal{S}$,

(iii) if A_n is an increasing sequence of sets in $\mathcal{S}$, then $\bigcup A_n \in \mathcal{S}$.

Theorem A.4.2. *Let $\mathcal{S}$ be a monotone class of subsets of Ω and let $\mathcal{F}$ be a class of subsets that is closed under finite intersections. Then $\mathcal{F} \subseteq \mathcal{S}$ implies $\sigma(\mathcal{F}) \subseteq \mathcal{S}$.*

Proof. See Appendix A1 of Williams (1991). □

B

Elements of functional analysis

B.1 Hahn-Banach theorem

Recall that a (real) vector space V is called a *normed linear space* if there exists a *norm* on V, i.e. a map $\|\cdot\| : V \to [0, \infty)$ such that

(i) $\|v + w\| \leq \|v\| + \|w\|$ for all $v, w \in V$,

(ii) $\|av\| = |a|\,\|v\|$ for all $v \in V$ and $a \in \mathbb{R}$,

(iii) $\|v\| = 0$ if and only if $v = 0$.

A normed linear space may be regarded as a metric space, the distance between to vectors $v, w \in V$ being given by $\|v - w\|$.

If V, W are two normed linear spaces and $A : V \to W$ is a linear map, we define the *norm* $\|A\|$ of the operator A by

$$\|A\| = \sup\{\|Av\| : v \in V \text{ and } \|v\| = 1\}.$$

If $\|A\| < \infty$, we call A a *bounded linear transformation* from V to W. Observe that by construction, it holds that $\|Av\| \leq \|A\|\|v\|$ for all $v \in V$. A bounded linear transformation from V to $\mathbb{R}$ is called a *bounded linear functional on* V.

We can now state the Hahn-Banach theorem.

Theorem B.1.1 (Hahn-Banach theorem). *Let W be a linear subspace of a normed linear space V and let A be a bounded linear functional on W. Then A can be extended to a bounded linear functional on V, without increasing its norm.*

Proof. See for instance Rudin (1987), pp. 104–107. □

In Chapter 3 we use the following corollary of the Hahn-Banach theorem.

Corollary B.1.2. *Let W be a linear subspace of a normed linear space V. If every bounded linear functional on V that vanishes on W, vanishes on the whole space V, then W is dense in V.*

Proof. Suppose that W is not dense in V. Then there exists a $v \in V$ and an $\varepsilon > 0$ such that for $w \in W$, $\|v - w\| > \varepsilon$. Now let W' be the subspace generated by W and v and define a linear functional A on W' by putting $A(w + \lambda v) = \lambda$ for all $w \in W$ and $\lambda \in \mathbb{R}$. Note that for $\lambda \neq 0$, $\|w + \lambda v\| = |\lambda| \, \|v - (-\lambda^{-1} w)\| \geq |\lambda|\varepsilon$, hence $|A(w + \lambda v)| = |\lambda| \leq \|w + \lambda v\|/\varepsilon$. It follows that $\|A\| \leq 1/\varepsilon$, so A is a bounded linear functional on W'. By the Hahn-Banach theorem, it can be extended to a bounded linear functional on the whole space V. Since A vanishes on W and $A(v) = 1$, the proof is complete. □

B.2 Riesz representation theorem

Let $E \subseteq \mathbb{R}^d$ be an arbitrary set and consider the class $C_0(E)$ of continuous functions on E that vanish at infinity. We endow $C_0(E)$ with the supremum norm

$$\|f\|_\infty = \sup_{x \in E} |f(x)|,$$

turning $C_0(E)$ into a normed linear space (and even a Banach space). The Riesz representation theorem (the version that we consider here) describes the bounded linear functionals on the space $C_0(E)$.

If μ is a finite Borel measure on E, then clearly, the map

$$f \mapsto \int_E f \, d\mu$$

is a linear functional on $C_0(E)$ and its norm is equal to $\mu(E)$. The Riesz representation theorem states that every bounded linear functional on $C_0(E)$ can be represented as the difference of two functionals of this type.

Theorem B.2.1. *Let A be a bounded linear functional on $C_0(E)$. Then there exist two finite Borel measures μ and ν on E such that*

$$A(f) = \int_E f \, d\mu - \int_E f \, d\nu$$

for every $f \in C_0(E)$.

Proof. See Rudin (1987), pp. 130–132. □

References

Billingsley, P. (1995). *Probability and measure*. John Wiley & Sons Inc., New York, third edition.

Kallenberg, O. (1997). *Foundations of modern probability*. Springer-Verlag, New York.

Karatzas, I. and Shreve, S. E. (1991). *Brownian motion and stochastic calculus*. Springer-Verlag, New York.

Revuz, D. and Yor, M. (1991). *Continuous martingales and Brownian motion*. Springer-Verlag, Berlin.

Rogers, L. C. G. and Williams, D. (1987). *Diffusions, Markov processes, and martingales. Vol. 1*. John Wiley & Sons Inc., New York.

Rudin, W. (1987). *Real and complex analysis*. McGraw-Hill Book Co., New York, third edition.

Williams, D. (1991). *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge.