"Stochastic Processes"
Course notes

Anton Wakolbinger

— Summer Semester 2001 —

Contents

Chapter 1

Discrete Markov chains

1.1 Random paths; stochastic and Markovian dynamics

Recall: An *S-valued random variable* X (where S is a space of possible outcomes) models the random choice of an element in S. We will focus on the case where S consists of *paths* $\underline{x} = (x_t)$ in some *state space* S_0. Here, x_t is the *state* at time t, and time may be modelled either as discrete or continuouos.

A *stochastic process* is thus a random variable taking its values in a *path space* S.

In later chapters, we will turn to continuous time and look e.g. at continuous real-valued paths. In the present chapter we will concentrate on *discrete time* and *discrete state space* – many important concepts will become clear already in this case. Thus, in this chapter our path space will be of the form $S = S_0^{\mathbb{N}_0}$, with S_0 some finite or countable set.

In order to specify the *distribution* of our random path $X = (X_0, X_1, \ldots)$ within some probability model, we will have to agree, for a reasonably rich class of subsets B of S, on the probability of the event $\{X \in B\}$ (read "X falls in B"). A reasonable and intuitive procedure for this is to tell *how to start* and *how to proceed.* In other words, we are going to specify an *initial distribution* and a *stochastic dynamics.* Now then!

Let μ be a probability measure on S_0 (given by the nonnegative probability weights $\mu(x_0), x_0 \in S_0$), and let, for each $n \in \mathbb{N}_0$ and $x_0, \ldots, x_{n-1} \in S_0$, $P_n((x_0, \ldots, x_{n-1}), .)$ be probability weights on S_0.

Imagining that $P_n((x_0, \ldots, x_{n-1}), x_n)$ should be the conditional probability of the event $\{X_n = x_n\}$, given $\{(X_0, \ldots, X_{n-1}) = (x_0, \ldots, x_{n-1})\}$, it makes perfect sense in view of the multiplication rule to define

$$\mathbf{P}[(X_0, \ldots, X_n) = (x_0, \ldots, x_n)] := \mu(x_0) P_1(x_0, x_1) \ldots P_n((x_0, \ldots, x_{n-1}), x_n) \tag{1.1}$$

The l.h.s. of (1.1) can be written as $\mathbf{P}[X \in B_{x_0, \ldots, x_n}]$, where

$$B_{x_0, \ldots, x_n} = \{(x_0, \ldots x_n)\} \times S_0^{\{n+1, n+2, \ldots\}} \subseteq S \tag{1.2}$$

A theorem due to Ionescu-Tulcea tells us that (1.1) uniquely extends to a probability measure $B \mapsto \mathbf{P}[X \in B]$, $B \in \mathcal{S}$, where $\mathcal{S}$ is the σ-algebra generated by all sets of the form (1.2).

Little can be said on interesting properties of X on this level of generality. There is, however, a rich theory for *time homogeneous, Markovian* stochastic dynamics, where

$$P_n((x_0,\dots,x_{n-1}),x_n) = P(x_{n-1},x_n) \tag{1.3}$$

for a *stochastic matrix* $P = P(y,z)$, $y,z \in S_0$, i.e. a matrix with nonnegative entries and each of whose rows sums to one.

In the following, we will consider a time-homogenous Markovian dynamics (given by a stochastic matrix P on S_0) as fixed, and vary the initial distribution μ, which we wrte as a subscript to $\mathbf{P}$:

$$\mathbf{P}_\mu[(X_0,\dots,X_n) = (x_0,\dots,x_n)] := \mu(x_0)P(x_0,x_1)\dots P(x_{n-1},x_n). \tag{1.4}$$

We call a random path X with distribution given by (1.4) a *Markov chain with transition matrix P and initial distribution μ* (or *Markov-(μ,P)*- for short).

For a deterministic start in $z \in S_0$, i.e. $\mu = \delta_z$, we write $\mathbf{P}_{\delta_z} =: \mathbf{P}_z$ for short.

It is elementary to verify the so called *Markov property*: For all $z_0,\dots,z_n \in S_0$ with $\mathbf{P}_\mu[(X_0,\dots,X_n) = (z_0,\dots,z_n)] > 0$

$$\begin{aligned}\mathbf{P}_\mu[(X_n,\dots,X_{n+h}) = (x_0,\dots x_h) \,|\, (X_0,\dots X_n) = (z_0,\dots z_n)]&\\ = \mathbf{P}_{z_n}[(X_0,\dots,X_h) = (x_0,\dots x_h)].&\end{aligned} \tag{1.5}$$

This extends to

$$\mathbf{P}_\mu[(X_n,X_{n+1},\dots) \in B \,|\, (X_0,\dots X_n) = (z_0,\dots z_n)] = \mathbf{P}_{z_n}[(X_0,X_1,\dots) \in B]$$

for all $B \in \mathcal{S}$. In other words, conditional on $\{X_n = z\}$, $(X_{n+k})_{k\geq 0}$ is *Markov-(δ_z,P)* and independent of $(X_0,\dots X_n)$.

1.2 Excursions from a state; recurrence and transience

The path of a Markov chain "starts a new life" (independent of its past) given its present state not only at a fixed time n, but also at certain random times read off from the path, e.g. at the so called *return times* to a fixed state z.

For $z \in S_0$, we put $R_z := R_z^1 := \inf\{n > 0 : X_n = z\}$, and call it the *first passage time* of X to z. For paths starting in z, we speak also of the *(first) return time* to z. Using the convention that the infimum of the empty set is ∞, we observe that the event $\{R_z = \infty\}$ equals the event that X never returns to z. Write $X^z := (X_k)_{k<R_z}$; in case $X_0 = z$, we call it *(the first) excursion* of X from z. We say that X^z *escapes from z* if $R_z = \infty$.

Now let $X^{z,1}, X^{z,2},\dots$ be independent, identically distributed ("i.i.d") copies of X^z under $\mathbf{P}_z$, and let L be the smallest integer for which $X^{z,L}$ escapes from z (we put $L = \infty$ if all of the $X^{z,k}$ return to z). By "piecing together" the $X^{z,1}, X^{z,2},\dots,X^{z,L}$ in case L is finite, and all the $X^{z,1}, X^{z,2},\dots$ in case L is infinite, we arrive at a random path Y whose distribution is the same as that of X under $\mathbf{P}_z$.

Obviously, L can be viewed as the waiting time to the first success in a coin tossing experiment with success probability $\mathbf{P}_z[R_z = \infty]$. Such a waiting time is finite iff the success probability is strictly positive. In this case, i.e. if $\mathbf{P}_z[R_z = \infty] > 0$, we call the state z *transient*. Otherwise, i.e. if $\mathbf{P}_z[R_z < \infty] = 1$, we call the state z *recurrent*.

In a coin tossing experiment with success probability p, the expected waiting time to the first success is $1/p$ (check!), hence it is finite iff $p > 0$. This, in turn, is the case iff the (random) waiting time to the first success is finite a.s.

Coming back to our picture of excursions from z, and writing

$$V(z) := \#\{n \geq 0 : X_n = z\} = \sum_{n=0}^{\infty} I_{\{X_n = z\}}$$

for the *number of visits in the state* z of the path X, we see that we have proved

Proposition 1.2.1 *: Let z be a state in S_0.*
(i) z is recurrent $\Leftrightarrow \mathbf{P}_z[V(z) = \infty] = 1 \Leftrightarrow \mathbf{E}_z[V(z)] = \infty$.
(ii)z is transient $\Leftrightarrow \mathbf{P}_z[V(z) < \infty] = 1 \Leftrightarrow \mathbf{E}_z[V(z)] < \infty$.

Can it happen that some state is transient, whereas another one is recurrent? Yes, as the following simple example shows:

$$S_0 = \{0, 1\},\ P(0, 0) = P(1, 0) = 1.$$

Here, state 1 ist transient and state 0 is recurrent. Note however, that state 1 "cannot be reached" from state 1. Let's make precise what we mean by this.

Definition 1.2.1 *For two states $y, z \in S_0$ we say that z can be reached from y if, for some $n \geq 0$, $P^n(y, z) > 0$.*

Here, P^n denotes the n-th power of the stochastic matrix P, defined inductively by $P^0(y, z) := \delta_{y,z}$, $P^1 := P$, $P^n(y, z) := \sum_x P^{n-1}(y, x)\, P(x, z)$.

Remark 1.2.1 *For $y \neq z$, z can be reached from y iff $\mathbf{P}_y[R_z < \infty] > 0$. Indeed, for $n > 1$,*

$$P^n(y, z) = \mathbf{P}_y[X_n = z] \leq \mathbf{P}_y[R_z \leq n] \leq \mathbf{P}_y[R_z < \infty],$$

and conversely,

$$\mathbf{P}_y[R_z < \infty] = \mathbf{P}_y[X_n = z \textit{ for some } n > 0] \leq \sum_{n>0} P^n(y, z).$$

The next lemma states that no transient state can be reached from a recurrent one, and that reachability is in fact an equivalence relation on the recurrent states.

Lemma 1.2.1 *Assume $y \in S_0$ is recurrent, and $z \in S_0$ can be reached from y. Then*
a) $\mathbf{P}_y[V(z) = \infty] = 1$
b) y can be reached from z
c) z is recurrent.

Proof. a) The random path X visits z in every excursion from y with positive probability, and, since y is recurrent, there are infinitely many trials.

b) Assume y cannot be reached from z. Then, starting from y, X would reach z with positive probability and afterwards never return to y, which contradicts the recurrence of y.

c) Starting from z, X hits y with positive probability, and from there it has, in each of its infitely many independent excursions, the same positive probability to visit z. Hence $\mathbf{P}_z[V(z) = \infty] > 0$, and z is recurrent. □

The previous lemma implies that S_0 can be partitioned into its subset of transient states and an at most countable number of so-called *irreducible recurrent components*, all of which consist of mutually reachable recurrent states.

Definition 1.2.2 *We call S_0 (or, more exactly, P),* irreducible recurrent *if any two states can be reached from each other, and one (and hence any) state in S_0 is recurrent. (In other words, S_0 is irreducible recurrent if it consists of exactly one irreducible recurrent component.)*

1.3 Renewal chains

The following class of examples will be basic for what follows.

Take $\mathbb{N}$ as the set of states, and consider a transition matrix $p = (p_{i,j})_{i,j\in\mathbb{N}}$ on $\mathbb{N}$ with the property $p_{i,i-1} = 1\,(i > 1)$. The dynamics of the corresponding Markov chain Y can be described as follows:

Y moves down to 1 "at unit speed"; after having reached 1, it jumps to k with probability $p_{1,k}$. Think of Y as the residual lifetime of a certain device. When it expires, the device is replaced by a new one; all the lifetimes are independent copies of the return time R_1 under $\mathbf{P}_1$ (note also that $\mathbf{P}_1[R_1 = k] = p_{1,k}, k \geq 1$). Having this picture in mind, we call Y a *renewal chain* (with *lifetime distribution* $(p_{1,k})$).

Obviously, the state 1 is always recurrent. What about all the other states? If there are arbitrarily large $k \in \mathbb{N}$ such that $p_{1,k} > 0$, then p is irreducible recurrent. If, on the other hand, $K := \sup\{k\ :\ p_{1,k} > 0\} < \infty\}$, then all $k > K$ are transient. However, in this case $(p_{i,j})_{1\leq i,j\leq K}$ is an irreducible recurrent stochastic matrix.

1.4 Equilibrium distributions

Definition 1.4.1 *A probability measure π on S_0 is called an equilibrium distribution for P if*

$$\pi P = \pi.$$

This is equivalent to the distribution of X under $\mathbf{P}_\pi$ being time-stationary (or invariant under time shift):

$$\mathbf{P}_\pi[(X_0, X_1, \dots) \in \cdot] = \mathbf{P}_\pi[(X_n, X_n, \dots) \in \cdot], \quad , n \in \mathbb{N}.$$

Does there exist an equilibrium distribution, ν, say, for the renewal chain with transition matrix p as described in Subsection 1.3? This amounts to require

$$\nu_i = \nu_{i+1} + \nu_1 p_{1,i}, \quad i = 1, 2, \dots \tag{1.6}$$

Summing this from $i = 1$ to $i = k - 1$ we get

$$-\nu_k + \nu_1 = \nu_1 \sum_{i=1}^{k-1} p_{1,i}$$

or equivalently

$$\nu_k = \nu_1 \sum_{i=k}^{\infty} p_{1,i} = \nu_1 \mathbf{P}[R \geq k], \tag{1.7}$$

where R is a random variable with $\mathbf{P}[R = j] := p_{1,j}, \quad j = 1, 2 \dots$

Summing this from $k = 1$ to ∞, we arrive at

$$\begin{aligned} 1 = \nu_1 \sum_{k=1}^{\infty} \mathbf{P}[R \geq k] &= \nu_1 \sum_{k=1}^{\infty} \sum_{j\geq k} \mathbf{P}[R = j] = \nu_1 \sum_{j=1}^{\infty} \sum_{k\leq j} \mathbf{P}[R = j] \\ &= \nu_1 \sum_{j=1}^{\infty} j\mathbf{P}[R = j] = \nu_1 \mathbf{E}R \end{aligned}$$

This can be satisfied iff $\mathbf{E}R < \infty$, i.e. if the *expected return time* to 0 is finite when starting from 0. In this case we have the fundamental identity

$$\nu_1 = \frac{1}{\mathbf{E}R} \tag{1.8}$$

which has a very intuitive interpretation:

In a renewal chain, the equilibrium weight of state 1 is the inverse of the expected duration of an excursion from 1.

We will soon see how (1.8) extends to general discrete state spaces: all we have to require is, apart from irreducibility, that the *expected return times* are finite.

To prepare this, let us state a simple

Proposition 1.4.1 *a) If z is transient, then $\mathbf{E}_y[V(z)] < \infty$ for all $y \in S_0$.*

b) If π is an equilibrium distribution and $\pi(z) > 0$, then z is recurrent.

Proof. a) We may assume $y \neq z$. Then

$$\mathbf{E}_y[V(z)] = \mathbf{P}_y[R_z < \infty] \cdot \mathbf{E}_z[V(z)] < \infty.$$

b) Assume z were transient. Then, because of a),

$$\sum_{n=0}^{\infty} P^n(y,z) < \infty \qquad \text{for all } y \in S_0.$$

A fortiori, $P^n(y,z) \to 0$ for all $y \in S_0$. Because of dominated convergence (see Lemma 1.4.1 below),

$$\sum_{y \in S_0} \pi(y) P^n(y,z) \to 0. \tag{1.9}$$

But the l.h.s. of (1.9) is for all $n \in \mathbb{N}$ equal to $\pi(z)^{>}0$, which is a contradiction. □

We have to append

Lemma 1.4.1 *(Dominated convergence, discrete case) Let m be a measure on S_0 (given by the nonnegative weights $m(y)$, $y \in S_0$), and let $g : S_0 \to \mathbb{R}_+$ be m-integrable, i.e. $\sum_{y \in S_0} g(y)m(y) < \infty$. In addition, let f_n be a sequence of real-valued functions on S_0 which is dominated by g (in the sense that $|f_n| \leq g$ for all n) and which converges pointwise to some f. Then*

$$\sum_{y \in S_0} |f_n(y) - f(y)| m(y) \to 0$$

and a fortiori

$$\sum_{y \in S_0} f_n(y)m(y) \to \sum_{y \in S_0} f(y)m(y).$$

Proof. For given ε choose a finite $K \subseteq S_0$ such that $\sum_{y \notin K} g(y) < \varepsilon$. Then

$$\limsup \sum_{y \in S_0} |f_n(y) - f(y)| m(y) \leq \limsup \sum_{y \in K} |f_n(y) - f(y)| m(y) + 2\varepsilon = 2\varepsilon.$$

□

Remark 1.4.1 *Let π be an equilibrium distribution on S_0, and let $C_1, C_2, \dots$ be the irrudicible recurrent components of S_0. If, for some i, $\pi(C_i) > 0$, then also $\pi(.|C_i)$ is an equilibrium distribution (check!), and π has the so-called* ergodic decomposition

$$\pi = \sum_{i:\pi(C_i)>0} \pi(C_i)\pi(.|C_i)$$

We say that π is *ergodic* if it is concentrated on a single irreducible recurrent component.

Theorem 1.4.1 *For an ergodic equilibrium distribution π and any state z such that $\pi(z) > 0$,*

$$\pi(z) = \frac{1}{\mathbf{E}_z(R_z)}.$$

Proof. First note that by the assumed ergodicity of π and because of Lemma 1.2.1.a), $\mathbf{P}_\pi$-almost all paths X visit z infinitely many often. For such paths, define $T_{z,k}$ as the time of the k-th visit to z, and put $Y_n := \inf\{T_{z,k} - n \ : \ T_{z,k} > n, \ k \geq 1\}$. Then under $\mathbf{P}_\pi$, $Y = (Y_n)$ is a time stationary renewal chain whose lifetime distribution equals the distribution of R_z under $\mathbf{P}_z$. Since $\pi(z) = \mathbf{P}_\pi[X = z] = \mathbf{P}_\pi[X_0 = z] = \mathbf{P}_\pi[Y_0 = 0]$, the assertion follows from (1.8). □

Corollary 1.4.1 *An irreducible recurrent Markov chain has at most one equilibrium distribution. If it has one, then all states z are* positive recurrent*, that is $\mathbf{E}_z[R_z] < \infty$.*

What about *existence* of an equilibrium distribution in the positive recurrent case? The situation is as nice as it can be: *for fixed z, the expected number of visits in y in an excursion from z, divided by the expected duration of the excursion, is an equilibrium distribution!* (Because of Corollary 1.4.1, this is then *the* equilibrium distribution.)

Well then! For fixed $z \in S_0$, we put

$$m_z(y) := \mathbf{E}_z[\sum_{n=1}^{R_z} I_{\{X_n = y\}}]. \tag{1.10}$$

This defines a measure on S_0 whose total mass $m_z(S_0) = \mathbf{E}_z[R_z]$ is finite iff z is positive recurrent.

Theorem 1.4.2 *Assume $z \in S_0$ is recurrent. Then m_z is a P-*invariant measure*, that is, for all $y \in S_0$*

$$\sum_{x \in S_0} m_z(x) P(x, y) = m_z(y). \tag{1.11}$$

In particular, if z is positive recurrent, then $\frac{1}{m_z(S_0)} m_z =: \pi$ is the unique equilibrium distribution, and for all π-integrable $f : S_0 \to \mathbb{R}$, that is, for all f with $\sum |f(y)|\pi(y) < \infty$ we have

$$\mathbf{E}_z\left[\sum_{n=1}^{R_z} f(X_n)\right] = \mathbf{E}_z[R_z] \sum_{y \in S_0} f(y)\pi(y). \tag{1.12}$$

Proof. Since z is recurrent, under $\mathbf{P}_z$ we have $R_z < \infty$ and $X_0 = X_{R_z} = z$ with probability one. Therefore, it makes no difference if we count the visit to z at the end or at the beginning of the excursion, and so we have for all $x \in S_0$

$$m_z(x) = \mathbf{E}_z[\sum_{n=1}^{R_z} I_{\{X_{n-1} = x\}}] = \sum_{n=1}^{\infty} \mathbf{P}_z[X_{n-1} = x, n \leq R_z].$$

Noting that the event $\{R_z \geq n\}$ depends only on $X_0, \dots, X_{n-1}$, we observe, using the Markov property at time $n-1$,

$$\begin{aligned}
\sum_x m_z(x)P(x,y) &= \sum_x \sum_{n=1}^{\infty} \mathbf{P}_z[X_{n-1}=x, n\le R_z]P(x,y)\\
&= \sum_x \sum_{n=1}^{\infty} \mathbf{P}_z[X_{n-1}=x, X_n=y, n\le R_z]\\
&= \sum_{n=1}^{\infty} \mathbf{P}_z[X_n=y, n\le R_z]\\
&= \mathbf{E}_z\left[\sum_{n=1}^{R_z} I_{\{X_n=y\}}\right] = m_z(y).
\end{aligned}$$

Thus, m_z defined by (1.10) obeys (1.11). If z is positive recurrent, then $m_z(S_0) = \mathbf{E}_z[R_z] < \infty$. Hence, because of Corollary 1.4.1, $\frac{1}{m_z(S_0)}m_z =: \pi$ is *the* equilibrium distribution. For $f := 1_{\{x\}}$, $x \in S_0$, (1.12) is clear; for general π-integrable f, (1.12) follows by linearity. $\square$

1.5 The ergodic theorem for Markov chains

The next result is intimately connected with the law of large numbers which we recall first.

Theorem 1.5.1 *(Strong law of large numbers) Let $Z_1, Z_2, \ldots$ be i.i.d. real-valued random variables with finite expectation μ. Then*

$$\frac{1}{k}\sum_{j=1}^{k} Z_j \to \mu \qquad \textit{almost surely as } k\to\infty.$$

In words: The "empirical mean" (i.e. the arithmetic mean of $Z_1, \ldots Z_k$) converges (as $k\to\infty$) with probability 1 to the "theoretical mean" (i.e. the expectation μ).

Theorem 1.5.2 *Assume P is irreducible and positive recurrent, and denote by π its equilibrium distribution. Let $f : S_0 \to \mathbb{R}$ be π-integrable, that is $\sum_y |f(y)|\pi(y) < \infty$. Then for all $z \in S_0$,*

$$\frac{1}{n}\sum_{i=0}^{n-1} f(X_i) \overset{n\to\infty}{\longrightarrow} \sum_y f(y)\pi(y) \qquad \mathbf{P}_z \textit{ almost surely.}$$

In words: For any (reasonable) real-valued function f defined on the state space, the "time average" (i.e. the arithmetical mean of $f(X_0), \ldots, f(X_{n-1})$) along the path converges a.s. to the "space average", i.e. the expectation of $f(y)$ with respect to the equilibrium distribution concentrated on the respective component – provided this equilibrium distribution exists.

Proof: It suffices to consider the case $f \ge 0$ (write f as the difference of its positive part $f_+ := \sup(f,0)$ and its negative part $f_- := -\inf(f,0)$). We first consider the random sequence $T_1, T_2, \ldots$ of the first, second, $\ldots$ return time to z and put $T_0 := 0$. Then $\sum_{i=0}^{T_k-1} f(X_i)$ is the sum of k i.i.d. random variables, each with expectation $\mathbf{E}_z[R_z]\sum_y f(y)\pi(y)$ (by theorem 1.4.2), Thus we obtain by the law of large numbers

$$\frac{1}{k}\sum_{i=0}^{T_k-1} f(X_i) \longrightarrow \mathbf{E}_z[R_z]\sum_y f(y)\pi(y) \quad \mathbf{P}_z \text{ a.s. as } k\to\infty.$$

On the other hand, also T_k is, under $\mathbf{P}_z$, the sum of k i.i.d. random variables, each with expectation $\mathbf{E}_z[R_z]$. Hence, again by the strong law of large numbers,

$$\frac{T_k}{k} \longrightarrow \mathbf{E}[R_z] \quad \mathbf{P}_z \text{ a.s. as } k\to\infty, \tag{1.13}$$

and therefore

$$\frac{1}{T_k}\sum_{i=0}^{T_k-1} f(X_i) \longrightarrow \sum_y f(y)\pi(y) \quad \mathbf{P}_z \text{ a.s. as } k\to\infty. \tag{1.14}$$

Put $K(n) := \max\{k : T_k \le n\}$, that is, $K(n)$ is the number of returns up to (and including) time n (note also that $K(n) = \sum_{i=1}^{n} I_{\{X_i=z\}}$). Obviously we have $T_{K(n)} \le n < T_{K(n)+1}$ and

$$K(n)\to\infty \quad \mathbf{P}_z \text{ a.s. as } n\to\infty \tag{1.15}$$

because of the assumed recurrence.

We then have

$$\frac{1}{T_{K(n)+1}}\sum_{i=0}^{T_{K(n)}-1} f(X_i) \le \frac{1}{n}\sum_{i=0}^{n-1} f(X_i) \le \frac{1}{T_{K(n)}}\sum_{i=0}^{T_{K(n)+1}-1} f(X_i).$$

Noting that $T_{K(n)}/T_{K(n)+1} \longrightarrow 1$ $\mathbf{P}_z$ a.s. as $n\to\infty$ because of (1.15) and (1.13), we infer from (1.14) that both the the l.h.s. and the r.h.s. converge to $\sum_y f(y)\pi(dy)$ $\mathbf{P}_z$ a.s. Hence the claim follows. □

1.6 Convergence to equilibrium

If π is an equilibrium distribution, does $\mathbf{P}_z[X_n = x] = P^n(z,x)$ converge to $\pi(x)$ as $n\to\infty$ for arbitrary initial states z ?

There are simple counterexamples. Perhaps the simplest is

$$S_0 = \{0,1\}, \quad P(0,1) = P(1,0) = 1.$$

Then $\pi(0) = \pi(1) = \frac{1}{2}$, $P^{2n}(0,0) = 1$, $P^{2n+1}(0,0) = 0$. But in fact this kind of periodicity is all what can cause troubles.

A state z is called *aperiodic* for P if $P^n(z,z) > 0$ for some $n\in\mathbb{N}$ and if the greatest common divisor of $\{n\in\mathbb{N} : P^n(z,z) > 0\}$ is 1. A transition matrix P is called aperiodic if all states have this property. Let us state the following elementary lemma from the realm of "discrete mathematics". We leave its proof as an exercise (see, e.g. Appendix 1 of P. Bremaud, Markov Chains, Springer 1999)

Lemma 1.6.1 *Let K be a nonempty set of positive integers which is closed under addition. Then the greatest common divisor of K is 1 iff $\mathbb{N}\setminus K$ is finite.*

Corollary 1.6.1 *: a) z is aperiodic iff $P^n(z,z) > 0$ for all sufficiently large n.*

b) If S_0 is irreducible and some $z\in S_0$ is aperiodic, then all states in S_0 are aperiodic.

Proof: a) Put $K := \{n \in \mathbb{N} \mid P^n(z,z) > 0\}$ and note that K is closed under addition.

b) For some k, l, and all sufficiently large n we have

$$P^{k+l+n}(y,y) \geq P^k(y,z) \cdot P^n(z,z) \cdot P^l(z,y) > 0. \quad \square$$

Theorem 1.6.1 *Let P be irreducible and aperiodic, and suppose that P has an equilibrium distribution π. Let ρ be any distribution. Then $\mathbf{P}_\rho[X_n = x] \longrightarrow \pi(x)$ as $n \to \infty$ for all x. In particular,*

$$P^n(z,x) \longrightarrow \pi(x) \quad \textit{for all } z, x.$$

Proof: We follow the book of J.R. Norris (Markov Chains, Cambridge University Press 1997). The proof uses the idea of coupling, which goes back to Wolfgang Doeblin, who died as a soldier in World War II in his twenties. Intuitively, the trick is as follows: Take a chain Y which starts in equilibrium π and is independent of X. First follow X, wait until X and Y meet, and from this time proceed with Y instead of X. Then the distance from the distribution of X_n to π can be estimated by the probability that X and Y haven't met by time n. To make things still simpler, let us wait till X and Y meet at some predescribed state b, i.e. we put

$$T := \inf\{n \geq 1 : X_n = Y_n = b\}$$

We claim that $\mathbf{P}[T < \infty] = 1$. Indeed, the process $W_n = (X_n, Y_n)$ is a Markov chain on $S_0 \times S_0$ with transition matrix

$$\tilde{P}((x,y),(u,v)) := P(x,u)P(y,v)$$

and initial distribution

$$\tilde{\mu}(x,y) := \mu(x)\pi(y).$$

Since P is aperiodic, for all states x, y, u, v we have

$$\tilde{P}^n((x,y),(u,v)) = P^n(x,u)P^n(y,v) > 0$$

for all sufficiently large n, so $\tilde{P}$ is irreducible. Also, $\tilde{P}$ has an invariant distribution given by

$$\tilde{\pi}(x,y) := \pi(x)\pi(y)$$

(check!) so by Corollary 1.4.1, $\tilde{P}$ is positive recurrent, and the claim follows.

Let us now put

$$Z_n = \begin{cases} X_n & \text{if } n < T \\ Y_n & \text{if } n \geq T \end{cases}$$

It is rather clear that Z is (ρ, P)-Markov. Here is a formal argument. For $j > k$,

$$\mathbf{P}[(Z_0, \ldots, Z_k) = (x_0, \ldots, x_k); T = j] = \mathbf{P}[(X_0, \ldots, X_k) = (x_0, \ldots, x_k); T = j]$$

and for $j \leq k$,

$$\begin{aligned} & \mathbf{P}[(Z_0, \ldots, Z_k) = (x_0, \ldots, x_k); T = j] \\ = \; & \mathbf{P}[(X_0, \ldots, X_j) = (x_0, \ldots, x_j); (Y_j, \ldots, Y_k) = (x_j, \ldots, x_k); T = j] \\ = \; & \mathbf{P}[(X_0, \ldots, X_j) = (x_0, \ldots, x_j); T = j] P(x_j, x_{j+1}) \ldots P(x_{k-1}, x_k) \\ = \; & \mathbf{P}[(X_0, \ldots, X_k) = (x_0, \ldots, x_k); T = j]. \end{aligned}$$

Now sum over j to see that $(X_0, \ldots X_k)$ and $(Z_0, \ldots Z_k)$ have the same distribution.

We therefore have for all $B \subseteq S_0$

$$\begin{aligned} & \mid \mathbf{P}[X_n \in B] - \pi(B) \mid \;=\; \mid \mathbf{P}[Z_n \in B] - \mathbf{P}[Y_n \in B] \mid \\ = & \mid \mathbf{P}[X_n \in B; n < T] + \mathbf{P}[Y_n \in B, n \geq T] - \mathbf{P}[Y_n \in B] \mid \\ = & \mid \mathbf{P}[X_n \in B; n < T] - \mathbf{P}[Y_n \in B; n < T] \mid \\ \leq & \; \mathbf{P}[n < T] \longrightarrow 0 \quad \text{as } n \to \infty, \end{aligned}$$

since $T < \infty$ $\mathbf{P}$ a.s

□

1.7 Optimal Stopping

Let f be a nonnegative function on S_0, and think of $f(x), x \in S_0$ as the state-dependent payoff you receive when stopping in x. You are allowed to specify your stopping rule, in terms of some finite stopping time T. Recall that this is a $\mathbb{Z}_+$-valued random variable $T = T(X)$ such that $\{T = n\}$ depends only on $(X_0, X_1, \dots, X_n)$ for all n.
The expected payoff when using T and starting in x is $\mathbf{E}_x[f(X_T)]$. The task is to maximize this. Question: What is the *value*

$$v(x) := \sup_T \mathbf{E}_x[f(X_T)] \tag{1.16}$$

where $\sup_T$ extends over all finite stopping times T. And what about the best stopping rule ?

Clearly, $v \geq f$, since $T \equiv 0$ is a stopping rule. Intuitively, we would expect that a best stopping rule, whenever it exists, should be of the form

$$T = T_C,$$

where C is some subset of S_0, and T_C is the *first hitting time* to C, i.e.

$$T_C := \inf\{n \geq 0 : X_n \in C\}.$$

Assume that $v(x) = \mathbf{E}_x[f(X_{T_C})]$, for some $C \subseteq S_0$. Then a "first-step decomposition" shows that, for all $x \notin C$,

$$v(x) = \sum_{y \in S_0} P(x,y)v(y) =: Pv(x).$$

Moreover, we claim that

$$v \geq Pv \quad \text{on } S_0 \tag{1.17}$$

(we say that v is *superharmonic on* S_0).

Indeed, let us consider the stopping rule "first go one step according to P and only afterwards stop when reaching C". This cannot be better than T_C, hence

$$v(x) \geq \mathbf{E}_x[f(X_{1+T_C((X_1,X_2,\dots))})] = \mathbf{E}_x[\mathbf{E}_{X_1}[f(X_{T_C})]] = \mathbf{E}_x[v(X_1)] = Pv(x).$$

Recalling that $v \geq f$, $v \geq Pv$ and

$$\begin{aligned} v(x) &= f(x) \quad \text{for } x \in C \ \ (\text{in other words, } C \subseteq \{y : v(y) = f(y)\}) \\ v(x) &= Pv(x) \quad \text{for } x \notin C, \end{aligned}$$

we observe that

$$v = \max(Pv, f),$$

In particular,

$$v = Pv \quad \text{on } \{v > f\} \tag{1.18}$$

(we say that v is *harmonic on* $\{v > f\}$).

Because of $C \subseteq \{v = f\}$, the path enters $\{v = f\}$ not later than it enters C, and as soon as we are in $\{v = f\}$, we can't do better than stopping immediately. So why not try $T_{\{v=f\}}$, the first hitting time of $\{y : v(y) = f(y)\}$, as a stopping rule ? All we have to show for this to work is

$$v(x) = \mathbf{E}_x[v(X_{T_{\{v=f\}}})], \quad x \in S_0. \tag{1.19}$$

What will help us is (1.18), saying that v is P-harmonic outside of $\{v = f\}$.
Put $Y_n := v(X_n)$. Because of the Markov property we have

$$\mathbf{E}_x[Y_{n+1} \mid (X_0, X_1, \dots, X_n) = (x_0, .., x_n)] = \sum_y P(x_n, y)v(y).$$

Hence, using (1.17) we have

$$\mathbf{E}_x[Y_{n+1} \mid (X_0, \dots, X_n) = (x_0, \dots, x_n)] \le v(x_n)$$

This we write briefly as

$$\mathbf{E}_x[Y_{n+1} \mid (X_0, \dots, X_n)] \le v(X_n) = Y_n$$

We say that $Y = (Y_n)$ is an X-*supermartingale*: Y is "adapted to the past of X" (here in fact even to the present), and the conditional expectation of Y_{n+1}, given the past of X up to time n, is less or equal Y_n. Now let us "stop" Y at $T_{\{v=f\}}$, putting

$$M_n := Y_{n \wedge T_{\{v=f\}}}.$$

By considering separately the events $\{T_{\{v=f\}} \le n\}$ and $\{T_{\{v=f\}} > n\}$ and using (1.18), it is easy to verify that

$$\mathbf{E}_x[M_{n+1} \mid (X_0, \dots X_n)] = M_n$$

We say that $M = (M_n)$ is a *martingale*. Later we will treat martingales more systematically, and we will prove the important "stopping theorem":
Let τ be an a.s. finite stopping time. Then

a) For any non-negative supermartingale $\tilde{Y}$,

$$\mathbf{E}[\tilde{Y}_\tau] \le \mathbf{E}[\tilde{Y}_0]$$

b) For any bounded martingale $\tilde{M}$,

$$\mathbf{E}[\tilde{M}_\tau] = \mathbf{E}[\tilde{M}_0].$$

Putting $\tau := T_{\{v=f\}}$ and $\tilde{M}_n := M_n$, b) translates into (1.19).

All we did henceforth was starting from the hypothesis of a best stopping rule of the form $T = T_C$. Let us now, without this hypothesis, deduce the following result on v given by (1.16)

Theorem 1.7.1 *:*

a) v is the smallest superharmonic majorant of f

b) $v = \lim v_n$, where $v_0 := f, v_{n+1} := \max(v_n, Pv_n)$

c) $v = \max(Pv, f)$ (and in particular, v is harmonic on $\{v > f\}$.

Proof:

a) (i) Any superharmonic $g \geq f$ is even $\geq v$. Indeed, for any stopping time T, by the stopping theorem:

$$g(x) \geq \mathbf{E}_x[g(X_T)] \geq \mathbf{E}_x[f(X_T)]$$

Now take $\sup\limits_T$ on the r.h.s

(ii) v is superharmonic:
Let T^n be a sequence of stopping times such that for all x

$$\mathbf{E}_x[f(X_{T^n})] \uparrow v(x) \text{ as } n \to \infty$$

Consider the stopping time $\tilde{T}^n := 1 + T^n((X_1, X_2, \dots))$ Then

$$\begin{aligned} v(x) \geq \mathbf{E}_x[f(X_{\tilde{T}^n})] &= \mathbf{E}_x[f(X_{1+T^n((X_1,X_2,\dots))})] \\ &= \mathbf{E}_x[\mathbf{E}_{X_1}[f(X_{T^n})]] \quad \uparrow \quad \mathbf{E}_x[v(X_1)] \end{aligned}$$

by monotone convergence. The r.h.s., however, equals $Pv(x)$.

b) Let us show that $\tilde{v} := \lim v_n$ is the smallest superharmonic majorant of f:

i) $\tilde{v}$ exists since $v_n \uparrow$

ii) $\tilde{v}$ is superharmonic, since by monotone convergence

$$P\tilde{v} = \lim_{n\to\infty} Pv_n \leq \lim_{n\to\infty} v_{n+1} = \tilde{v}.$$

iii) Let $f \leq g$ and g be superharmonic. We show $v_n \leq g$ by induction. This holds for $n = 0$, since $v_0 = f$. The induction hypothesis $v_n \leq g$ implies $Pv_n \leq Pg \leq g$, since g is superharmonic. Hence $v_{n+1} = \max(v_n, Pv_n) \leq g$. Letting n tend to ∞, we obtain $\tilde{v} \leq g$.

c) By induction we show

$$v_n = \max(f, Pv_{n-1}) :$$

$n = 1 :\ \ v_1 = \max(f, Pv_0) = \max(v_0, Pv_0)$
$n \to n+1 : v_{n+1} = \max(v_n, Pv_n) = \max(f, Pv_{n-1}, Pv_n) = \max(f, Pv_n)$
Letting $n \to \infty$, we obtain from b): $v = \tilde{v} = \max(f, P\tilde{v}) = \max(f, Pv)$.

□

Let us now assume that

$$T_{\{v=f\}} < \infty \quad \mathbf{P}_x \text{ a.s.} \tag{1.20}$$

and that f is bounded.
Then also v is bounded, and the stopping theorem for martingales implies

$$v(x) = \mathbf{E}_x[v(X_{T_{\{v=f\}}})] = \mathbf{E}_x[f(X_{T_{\{v=f\}}})],$$

hence $T_{\{v=f\}}$ is an optimal stopping rule. Moreover, if there is any optimal stopping rule T , then (1.20) is automatic. Indeed, in this case we have, since v is a superharmonic majorant of f:

$$v(x) \geq \mathbf{E}_x[v(X_T)] \geq \mathbf{E}_x[f(X_T)] = v(x).$$

Hence $\mathbf{E}_x[v(X_T)] = \mathbf{E}_x[f(X_T)]$. Together with $v(X_T) \geq f(X_T)$ this implies

$$v(X_T) = f(X_T) \quad \mathbf{P}_x \text{ a.s.}$$

Hence $X_T \in \{v = f\}$ $\mathbf{P}_x$ a.s., or in other words

$$T_{\{v=f\}} \leq T \qquad \mathbf{P}_x \text{ a.s.}$$

Thus, $T_{\{v=f\}}$ is the *smallest* optimal stopping time.

1.8 Renewal chains revisited

Let us recall our picture of renewal chains. We fix a probability distribution ϱ on $\mathbb{N}$ and imagine a path moving down the nonnegative integers one by one. Whenever the path hits state 0, God throws a die whose outcome R has distribution ϱ, and resets the state in the next step as R. In this way we obtain an $\mathbb{N}$-valued Markov chain with transition matrix $(p_{i,j})$ given by

$$\begin{array}{lll} p_{i,i-1} & = 1 & i > 1 \\ p_{1,j} & = \varrho_j & j = 1, 2, \ldots \end{array}$$

Let us now keep track not only of the current state y but also of the length of the excursion we are curently in. What is its equilibrium (resp. limiting) distribution? The stochastic dynamics which describes the joint evolution of the current total excursion length ℓ and the residual lifetime j is given by

$$\left\{ \begin{array}{lll} (\ell, y) \longrightarrow (\ell, y-1) & \text{with prob. } 1 & (1 < y \leq \ell) \\ (k, 1) \longrightarrow (\ell, \ell) & \text{with prob. } p_{1,\ell} & (k, \ell \geq 1) \end{array} \right. \tag{1.21}$$

With a similar proviso as in Subsection 1.3, this dynamics is irreducible recurrent. The conditions on the weights $\nu(\ell, y)$ of an invariant measure ν are

$$\left\{ \begin{array}{llll} \nu(\ell, \ell) & = \sum\limits_{k=1}^{\infty} \nu(k, 1) p_{1,\ell} & , & \ell \geq 1 \\ \nu(\ell, y) & = \nu(\ell, y+1) & , & 1 \leq y < \ell \end{array} \right. \tag{1.22}$$

The condition that ν has total mass 1 is

$$1 = \sum_{\ell \geq 1} \sum_{1 \leq y \leq \ell} \nu(\ell, y) = \sum_{\ell \geq 1} \ell \nu(\ell, \ell) \tag{1.23}$$

$$= \sum_{\ell \geq 1} \ell p_{1,\ell} \sum_{k=0}^{\infty} \nu(k, 1) \tag{1.24}$$

Recalling that $p_{1,\ell} = \varrho_\ell$, $\ell = 1, 2, \ldots$, and writing $\mathbf{E}R := \sum\limits_{j \geq 1} j \varrho_j$ for the expected value of ϱ, we arrive at

$$\nu_1 := \sum_{k=1}^{\infty} \nu(k, 1) = \frac{1}{\mathbf{E}R}$$

which is consistent with Subsection 1.4. For all $\ell \geq 1$, $1 \leq y \leq \ell$ we have the following chain of equalities:

$$\nu(\ell, y) = \nu(\ell, \ell) = \nu_1 p_{1,\ell} = \frac{1}{\mathbf{E}R}\varrho_\ell = \frac{1}{\ell}\frac{\ell}{\mathbf{E}R}\varrho_\ell\,. \tag{1.25}$$

Let us define the size-biased distribution $\hat{\varrho}$ obtained from ϱ by putting

$$\hat{\varrho}_k := \frac{k}{\mathbf{E}R}\varrho_k\,. \tag{1.26}$$

With this, (1.25) can be written as

$$\nu(\ell, y) = \frac{1}{\ell}\hat{\varrho}_\ell, \quad \ell \geq 1,\ 1 \leq y \leq \ell.$$

We have proved:

Proposition 1.8.1 *Assume* $\mathbf{E}R < \infty$.

Then the equilibrium distribution for the stochastic dynamics (1.21) is the distribution of the pair

$$(\hat{R}, U[1, \ldots, \hat{R}])$$

where $\hat{R}$ has the size-biased distribution $\hat{\rho}$ given by (1.26), and given $\hat{R}$, $U[1, \ldots, \hat{R}]$ is uniform on $\{1, \ldots, \hat{R}\}$.

Remark 1.8.1 *(cf. (1.7) and (1.8)) From (1.25) we obtain*

$$\nu_y := \sum_{k \geq y} \nu(k, y) = \sum_{k \geq y} \nu(k, k) = \frac{1}{\mathbf{E}R}\sum_{k \geq y} \varrho_k = \frac{1}{\mathbf{E}R}\mathbf{P}[R \geq y].$$

Remark 1.8.2 *If we assume in addition that the greatest common divisor of* $\{k : \varrho_k > 0\}$ *is* 1*, then the stochastic dynamics (1.21) is aperiodic, and we have convergence to equilibrium in the sense of Theorem 1.6.1.*

Chapter 2

Renewal processes

2.1 The renewal points and the residual lifetime process

We are going to parallel the picture of subsections 1.3 and 1.8, but now in continuous time. For the whole chapter we fix the following framework: Let ϱ be a distribution on $\mathbb{R}_+ = [0, \infty)$, and $R, R_1, R_2, \ldots$ be i.i.d with distribution ϱ. Assume $\mu := \mathbf{E}R \in (0, \infty)$. Let $Y = (Y_t)$ be an $\mathbb{R}_+$-valued stochastic process constructed as follows: Starting from some Y_0 in $(0, \infty), Y$ moves down the positive half axis at unit speed. When reaching 0, i.e. at time $\tau := Y_0$, Y is set equal to R_1 (so that $Y_{\tau-} = 0$, $Y_\tau = R_1$) and from there continues to move down at unit speed. At time $Y_0 + R_1, Y$ is set equal to R_2 and so on. In this way we get the *renewal points*

$$(T_1, T_2, \ldots) := (Y_0, Y_0 + R_1, Y_0 + R_1 + R_2, \ldots).$$

For $t \geq 0$ we define $N(t)$ to be the number of renewals up to and including time t. In other words, putting $T_0 := 0$, we have

$$N(t) = \max\{n : T_n \leq t\}. \tag{2.1}$$

The strong law of large numbers gives

$$\frac{T_n}{n} \longrightarrow \mu \quad \text{a.s.}, \tag{2.2}$$

hence with probability one only finitely many renewals happen before t. Let us also observe that, because of (2.2),

$$N(t) \longrightarrow \infty \text{ a.s. as } t \longrightarrow \infty. \tag{2.3}$$

Note that

$$T_{N(t)} \leq t < T_{N(t)+1}$$

and

$$Y_t = T_{N(t)+1} - t.$$

Proposition 2.1.1 $\lim\limits_{t \to \infty} \frac{N(t)}{t} = \frac{1}{\mu}$ *a.s.*

Proof: Compare to previous and next renewal times:

$$\frac{T_{N(t)}}{N(t)} \leq \frac{t}{N(t)} \leq \frac{T_{N(t)+1}}{N(t)+1} \frac{N(t)+1}{N(t)}.$$

Now use (2.2) and (2.3). $\square$

Almost sure convergence does not imply convergence of expectations. However, in our case we have

Proposition 2.1.2 *("Elementary Renewal Theorem")*

$$\lim_{t\to\infty} \frac{\mathbf{E}[N(t)]}{t} = \frac{1}{\mu}$$

provided that $\mathbf{E}Y_0 < \infty$.

Those who are interested in a proof can find it e.g. in the course notes "Applied Stochastic Processes" by Russel Lyons which are based on Sheldon Ross' book "Stochastic processes", 2nd ed., Wiley 1996, and are downloadable from Russel Lyons' homepage: http://php.ucs.indiana.edu/˜ rdlyons/home.html. The proof of Proposition 2.1.2 given there uses Wald's identity, which is of interest in its own right - we'll come back to this later.

2.2 Stationary renewal processes

Proposition 2.2.1 *If* Y_0 *has distribution density*

$$g(y) := \frac{1}{\mu}\mathbf{P}[R \geq y] = \frac{1}{\mu}\varrho((y,\infty)) \tag{2.4}$$

then for all $t > 0$ *, also* Y_t *has distribution density* g.

Before giving a clean (and still beautiful) proof of this proposition, we present a quick, a bit dirty and still nice "differential" argument that the equilibrium density of Y must be given by (2.4): The differential analogue of (1.6) in Subsection 1.4 is

$$g(r) = g(r + dr) + g(0)\varrho(dr) \tag{2.5}$$

Integrating this over $r \in [0, y)$ gives

$$-g(y) + g(0) = g(0)\,\varrho([0, y)),$$

or equivalently

$$g(y) = g(0)\int 1_{\{r\geq y\}}\varrho(dr).$$

Integrating this from $y = 0$ to $y = \infty$ and using the interchangeability of integrals with nonnegative integrands, we arrive at

$$1 = g(0)\int_0^\infty r\,\rho(dr).$$

Although this argument, which parallels the derivation of (1.7) and (1.8) in Subsection (1.4), is appealing and easy to remember, a direct rigorous derivation of the differential equation (2.5) seems tricky. Let's therefore have another go and prepare our rigorous proof of Proposition 2.2.1 with the following

Lemma 2.2.1 *Assume* Y_0 *has distribution density* g *given by (2.4).*

a) $\mathbf{P}[T_1 \in dt] = (1 - \mathbf{P}[R < t])dt/\mu$

b) For all $n \geq 2$,

$$\mathbf{P}[T_n \in dt] = (\mathbf{P}[R_1 + \ldots + R_{n-1} < t] - \mathbf{P}[R + R_1 + \ldots + R_{n-1} < t])\,dt/\mu.$$

c) $\sum_{n=1}^{\infty} \mathbf{P}[T_n \in dt] = dt/\mu$

Proof: a) $\mathbf{P}[T_1 \in dt] = \mathbf{P}[Y_0 \in dt] = g(t)dt = (1 - \mathbf{P}[R < t])dt/\mu$.

$$\begin{aligned} \text{b) } \mathbf{P}[T_n \in dt] &= \int 1_{[0,t)}(s)\mathbf{P}[R_1 + \ldots + R_{n-1} \in ds]\,\mathbf{P}[Y_0 + s \in dt] \\ &= \int 1_{[0,t)}(s)\mathbf{P}[R_1 + \ldots + R_{n-1} \in ds]\frac{1}{\mu}(1 - \mathbf{P}[R < t - s])dt \\ &= (\mathbf{P}[R_1 + \ldots + R_{n-1} < t] - \mathbf{P}[R_1 + \ldots + R_{n-1} + R < t])\,dt/\mu. \end{aligned}$$

c) follows by telescope summation. □

In words, Lemma 2.2.1 says that the expected number of renewal points in a set $B \subseteq \mathbb{R}_+$, when starting with Y_0 in density g, is $\frac{1}{\mu}\cdot$ Lebesgue measure of B. In other words, the expected number of renewal points per time unit (the so called *renewal density*) then is a constant (namely $\frac{1}{\mu}$), which clearly should be crucial for stationarity.

Proof of Proposition 2.2.1:

$$\begin{aligned} \mathbf{P}[Y_t \geq b] &= \sum_{n=0}^{\infty} \mathbf{P}[Y_t \geq b; N_t = n] \\ &= \mathbf{P}[Y_0 \geq t + b] + \sum_{n=1}^{\infty} \int_0^t \mathbf{P}[T_n \in ds; R_n \geq t - s + b] \\ &= \mathbf{P}[Y_0 \geq t + b] + \sum_{n=1}^{\infty} \int_0^t \mathbf{P}[T_n \in ds]\,\mathbf{P}[R \geq t - s + b] \\ &= \int_{t+b}^{\infty} \frac{1}{\mu}\mathbf{P}[R \geq s]ds + \int_0^t \frac{1}{\mu}ds\,\mathbf{P}[R \geq t - s + b] \\ &= \int_b^{\infty} \frac{1}{\mu}\mathbf{P}[R \geq s]ds = \mathbf{P}[Y_0 \geq b]. \qquad \square \end{aligned}$$

Like in subsection 1.8 we define the size-biased lifetime distribution $\hat{\varrho}$ by

$$\hat{\varrho}(dr) := \frac{1}{\mu} r \varrho(dr)$$

and denote by $\hat{R}$ a random variable with distribution $\hat{\varrho}$. Further, let U be a random variable which is uniformly distributed on $[0, 1]$ and independent of $\hat{R}$.

Lemma 2.2.2 *(compare to Remark 1.8.1 and Proposition 1.8.1)*

$$g(r) = \frac{1}{\mathbf{E}} R\mathbf{P}[R > r]$$

is the distribution density of $U\hat{R}$, where $\hat{R}$ has distribution $\hat{\varrho}$ and U is uniform on $[0, 1]$ and independent of $\hat{R}$.

Proof: Let $h : \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ be measurable

$$\begin{aligned} \mathbf{E}[h(U\hat{R})] &= \int \mathbf{E}[h(Ur)]\hat{\varrho}(dr) \\ &= \frac{1}{\mu}\int r\mathbf{E}[h(Ur)]\varrho(dr) \\ &= \frac{1}{\mu}\int r\frac{1}{r}\int_0^r h(t)dt\,\varrho(dr) \\ &= \frac{1}{\mu}\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} h(t)1_{\{t \leq r\}}dt\,\varrho(dr) \\ &= \frac{1}{\mu}\int_{\mathbb{R}_+} \varrho([t, \infty))h(t)dt = \int g(t)h(t)\,dt. \qquad \square \end{aligned}$$

Paralleling subsection 1.8, we describe the dynamics which keeps track not only of the residual lifetime Y_t, but also of the current total lifetime L_t. The first component of (L_t, Y_t) remains constant while the second component decreases with unit speed till it hits 0. Then both components jump to (R, R), where R has distribution ϱ and is independent of what happened before. In view of the previous results and Proposition 2.2.1, the following proposition should not be mysterious any more.

Proposition 2.2.2 *If (L_0, Y_0) equals in distribution $(\hat{R}, U\hat{R})$, then (L_t, Y_t) has the same distribution for all $t \geq 0$.*

For a proof, see e.g. S. Asmussen, Applied Probability and Queues, Wiley 1987, p.116/117.

This proposition gives us a neat way to construct a time-stationary process of renewal points on the real line:

Let $R_1, R_2, \ldots, R_{-1}, R_{-2}, \ldots$ be i.i.d. copies of R, and independent of $\hat{R}$ and U. Put

$$T_n := U\hat{R} + R_1 + \ldots + R_{n-1}, \quad T_{-n} := -(1-U)\hat{R} - R_{-1} - \ldots - R_{-(n-1)},$$

and

$$\Phi := \sum_{i \in \mathbb{Z} \setminus \{0\}} \delta_{T_i}.$$

Φ is a counting measure on $\mathbb{R}$: for $B \subseteq \mathbb{R}$,

$$\Phi(B) := \#\{i : T_i \in B\}$$

counts the number of renewal points falling int B. Because of Proposition 2.2.2, the distribution of Φ is invariant w.r.to time shift, i.e. $\Phi = \sum_{i \in \mathbb{Z} \setminus \{0\}} \delta_{T_i}$ and $\theta_t \Phi := \sum_{i \in \mathbb{Z} \setminus \{0\}} \delta_{T_i + t}$ have the same distribution. Thus, it makes sense to call Φ a *stationary renewal point process* with lifetime distribution ϱ.

2.3 Convergence to equilibrium

In view of subsection 1.6, it is not too astonishing that, in order to guarantee convergence to equilibrium, we need something like an aperiodicity condition.

Definition 2.3.1 *We say that ϱ is* non-lattice *if there does not exist any $d > 0$ s.th.*

$$\varrho(\{0, d, 2d, 3d, \ldots\}) = 1.$$

The next theorem, which we won't prove, is in the spirit of the convergence theorem Thm. 1.6.1.

Theorem 2.3.1 *(Key Renewal Theorem) Assume that ϱ is non-lattice. Then, irrespective of the distribution of $Y_0 = T_1$,*
$(T_{N(t)+1} - T_{N(t)}, T_{N(t)+1} - t)$ converges in distribution to $(\hat{R}, U\hat{R})$ as $t \longrightarrow \infty$.

The next theorem, which we won't prove either, states that in the non-lattice case the expected number of renewals in a late time interval of lenght a is approximately a/μ:

Theorem 2.3.2 *(Blackwell's Renewal Theorem) Assume that ϱ is non-lattice and $\mathbf{E}[T_1] < \infty$. Then for all $a > 0$,*

$$\mathbf{E}[N(t+a) - N(t)] \longrightarrow a/\mu \quad \text{as } t \to \infty.$$

2.4 Homogeneous Poisson processes on the line

We specialize to the important case

$$\varrho := \text{ exponential distribution (with parameter } \alpha)$$

i.e. $\varrho(dr) = \alpha e^{-\alpha r} dr$. Note that $\mathbf{P}[R > r] = e^{-\alpha r}$, which immediately shows that R has *no memory* in the following sense

$$\mathbf{P}[R > t + h | R > t] = \mathbf{P}[R > h].$$

This explains why ϱ coincides with the equilibrium distribution of the residual lifetime. Indeed, since $\mu = \mathbf{E}R = \frac{1}{\alpha}$,

$$\frac{1}{\mu}\mathbf{P}[R > r] = \alpha e^{-\alpha r}$$

Moreover, $\hat{R}$ has density $\alpha^2 r e^{-\alpha r}$. We claim that this is the density of the sum of two independent, exponential (α) distributed random variables X_1, X_2. Indeed,

$$\mathbf{P}[X_1 + X_2 \in [r, r + dr]] = dr \int_0^r \alpha e^{-\alpha s} \alpha e^{-\alpha(r-s)} ds = dr\, \alpha^2 r e^{-\alpha r}.$$

Thus the following definition makes sense.

Definition 2.4.1 *Let $R_0, R_0^-, R_1, R_2 \dots, R_{-1}, R_{-2} \dots$ be independent, exponential(α) distributed random variables, put*

$$T_n := R_0 + \sum_{i=1}^{n-1} R_i, \quad T_{-n} := -R_0^- - \sum_{i=1}^{n-1} R_{-i}, \quad n = 1, 2, \dots$$

Then $\Phi = \sum_{i \in \mathbb{Z}\setminus\{0\}} \delta_{T_i}$ is called a stationary (or homogeneous) Poisson point process *with intensity α.*

The name "Poisson process" is explained by the following

Proposition 2.4.1 *In the context of Definition 2.4.1, $N(1) = \#\{i \mid 0 \le T_i \le 1\}$ is Poisson (α)-distributed, and given $N(1) = n$, $(T_1, \dots T_n)$ is distributed like the* order statistics *(i.e. the increasing reordering) $(U_{(1)}, \dots U_{(n)})$ of independent uniform (on $[0, 1]$) random variables $U_1, \dots, U_n$.*

Proof: Let $B \subseteq \{(t_1, \dots t_n) : 0 \le t_1 \le \dots \le t_n \le 1\}$ and put

$$\tilde{B} := \{(r_0, \dots, r_{n-1}) \mid r_i \ge 0, (r_0, r_0 + r_1, \dots r_1 + \dots + r_{n-1}) \in B\}.$$

We then have

$$\begin{aligned} & \mathbf{P}[N(1) = n, (T_1, \dots, T_n) \in B] \\ = \; & \mathbf{P}[R_0 + \dots + R_n > 1, (R_0, \dots, R_{n-1}) \in \tilde{B}] \\ = \; & \int\limits_{r_0 + \dots + r_n > 1, (r_0, \dots, r_{n-1}) \in \tilde{B}} \alpha^{n+1} e^{-\alpha r_0} \dots e^{-\alpha r_n} dr_0 \dots .dr_n \end{aligned}$$

Using the 1-1 transformation $t_i = r_0 + \dots + r_{i-1}$, $i = 1, \dots n + 1$, we can write this as

$$\begin{aligned} & \int\limits_{(t_1, \dots, t_n) \in B,\, t_{n+1} > 1} \alpha^{n+1} e^{-\alpha t_{n+1}} dt_1 \dots dt_{n+1} \\ = \; & e^{-\alpha} \alpha^n \lambda^n(B) = e^{-\alpha} \frac{\alpha^n}{n!} n! \lambda^n(B) \\ = \; & Pois_\alpha(n) \cdot \mathbf{P}[(U_{(1)}, \dots ., U_{(n)}) \in B]. \end{aligned}$$

□

Since the random counting measures $\sum_{i=1}^{n} \delta_{U_{(i)}}$ and $\sum_{i=1}^{n} \delta_{U_i}$ obviously have the same distribution, we obtain from the previous proposition the following

Corollary 2.4.1 : *Let Z be Poisson(α)-distributed, and $U_1, U_2, \ldots$ be independent, uniform $[0,1]$ (and independent of Z).*

Let $0 < T_1 < T_2 < \ldots$ be the random time points of a stationary Poisson (α) process. Then $\sum_{i=1}^{N(1)} \delta_{T_i}$ and $\sum_{i=1}^{Z} \delta_{U_i}$ have the same distribution.

Chapter 3

Poisson processes

3.1 Heuristics

In the previous chapter, we have made acquaintance with homogeneous Poisson processes on the real line. Recall the intuition that in each small time interval of length dr, the probability of a point landing there is $\alpha \cdot dr$, independently of everything else.

This latter intuition carries beyond the line. Think of an arbitrary measurable space $(E, \mathcal{E})$, and let m be a σ-finite measure on $\mathcal{E}$ (where σ-finiteness means that $m(B_n) < \infty$ for some $B_1 \subseteq B_2 \subseteq \ldots$ with $\cup B_n = E$). For the moment let us also assume that $m(\{z\}) = 0$ for all $z \in E$.

Imagine throwing a configuration of points randomly into E, assuming that for each small volume element dy

$$\mathbf{P}[\text{a point lands in } dy] = m(dy),$$

and these events are independent for disjoint $dy_1, dy_2, \ldots$.
With some good sense of humor we can write:

$$\begin{aligned} \mathbf{P}[\text{no point lands in } B] &= \prod_{y \in B} (1 - m(dy)) \\ &= \prod_{y \in B} e^{-m(dy)} = e^{-\int_B m(dy)} = e^{-m(B)} \end{aligned}$$

This "taboo probability" is perfectly compatible with what we saw in the previous chapter. What is the distribution of the total number of points landing in B ?

We recall the Poisson limit law: The total number of successes for many independent trials whose (small) success probabilities sum up to α is asymptotically Pois(α)-distributed.
Thus, we guess that

$$\begin{aligned} \mathbf{P}[k \text{ points land in } B] &= Pois_{m(B)}(k) \\ &= e^{-m(B)} \frac{m(B)^k}{k!} \end{aligned}$$

3.2 Characterization

Let's now leave the realm of heuristics. Let E be a non-empty set, $\mathcal{E}$ be a σ-algebra on E, $n \in \cup\{\infty\}$, and $(x_i)_{i=1,\ldots,n}$ be a (finite or infinite) sequence in E. The measure

$$\varphi = \sum_{i=1}^{n} \delta_{x_i} \quad (*)$$

counts how many points of the sequence (x_i) fall into the various subsets of E. That is:

$$\varphi(B) := \#\{i \mid x_i \in B\},\ B \subseteq E. \tag{3.1}$$

Note that φ forgets about the ordering (but not about possible multiplicities) of (x_i). Literally the only thing that counts is the number of points of (x_i) falling into the set B.

Measures of the form (3.1) are called *point measures* or *counting measures*. The simplest of that kind are the *Dirac measures* δ_x, $x \in E$, that is, $n = 1$ in $(*)$. The definition $(**)$ then turns into

$$\delta_x(B) = 1 \text{ if } x \in B, \text{ and } = 0 \text{ if } x \notin B.$$

Integration of functions with respect to point measures is particularly simple. For $\varphi = \sum_{i=1}^{n} \delta_{x_i}$ and $f : E \to \mathbb{R}_+$,

$$\int f(x)\, \varphi(dx) = \sum_{i=1}^{n} f(x_i).$$

For the special case $n = 1$ this is just the definition (3.1).

Lemma 3.2.1 *a) The distribution of a random point configuration Φ is uniquely determined by the distribution of*

$$((\Phi(B_1), \dots, \Phi(B_n)), \quad n \in \mathbb{N}, B_i \in \mathcal{E}.$$

b) The distribution of an $\mathbb{R}_+^n$-valued random variable $Z = (Z_1, \dots, Z_n)$ is uniquely determined by all the expectations $\mathbf{E}e^{-<\beta,Z>}$, $\beta = (\beta_1, \dots, \beta_n) \in G$, where G is some non-empty open subset of $\mathbb{R}_+^n$.

Proof: a) see O. Kallenberg, Foundations of Modern Probability, Springer 1997, Thm 4.3.
b) see loc.cit. Prop.2.2, and O. Kallenberg, Random measures, 4th ed., Akademie-Verlag and Academic Press 1986, p. 167 □
Note that (3.2.1), viewed as a function of β, is called the *Laplace transform* of (the distribution of) Z.

Definition 3.2.1 *For a random point configuration Φ, the measure $B \longmapsto \mathbf{E}\Phi(B)$, $B \in \mathcal{E}$, is called the* intensity measure *of Φ.*

Proposition 3.2.1 *The distribution of a random point configuration Φ is uniquely determined by the expectations*

$$\mathbf{E}\exp(-\int f(z)\Phi(dz)),\ \ f : E \to \mathbb{R}_+ \textit{ measurable} \tag{3.2}$$

Proof: Fix $B_1, \dots, B_n \in \mathcal{E}$ and consider the $\mathbb{R}_+^n$-valued random variable

$$Z = (Z_1, \dots, Z_n) = (\Phi(B_1), \dots, \Phi(B_n)).$$

For $\beta = (\beta_1, \dots, \beta_n) \in \mathbb{R}_+^n$, we have, putting $f := \sum_{i=1}^{n} \beta_i 1_{B_i}$,

$$\mathbf{E}e^{-<\beta,Z>} = \mathbf{E}\exp(-\int f(z)\Phi(dz)) \tag{3.3}$$

Thus, by Lemma 3.1 c), the expectations (3.2) determine the distribution of the random variables Z. These, in turn, determine the distribution of Φ by Lemma 3.1a). $\square$

Definition 3.2.2 *A random point configuration Φ is called a Poisson point process (PPP) on E if, for all disjoint $B_1, B_2, \ldots \in \mathcal{E}$,*

$$\Phi(B_1), \ldots, \Phi(B_n)$$

are independent and Poisson-distributed.

We do not exclude the case that $\Phi(B_i) \equiv \infty$ a.s. for some B_i, in this case we say $\Phi(B_i)$ is Poisson(∞)-distributed. Next, we identify the "Laplace transform" of a PPP.

Proposition 3.2.2 *For a PPP Φ with intensity measure m, and $f : E \to \mathbb{R}_+$ measurable,*

$$\mathbf{E}\exp(-\int f(z)\Phi(dz)) = \exp(-\int (1 - e^{-f(z)})m(dz)) \tag{3.4}$$

Proof: a) Let N be Pois$_\alpha$-distributed, $\beta > 0$. Then

$$\mathbf{E}\exp(-\beta N) = \exp\left(-(1 - e^{-\beta})\alpha\right)$$

(check!)

b) For disjoint $B_1, \ldots, B_n \in \mathcal{E}$, and $\beta_1, \ldots, \beta_n > 0$,

$$\begin{aligned} \mathbf{E}\exp(-\sum_{i=1}^n \beta_i \Phi(B_i)) &= \prod_{i=1}^n \mathbf{E}\exp(-\beta_i \Phi(B_i)) \\ &= \prod_{i=1}^n \exp\left(-(1 - e^{\beta_i})m(B_i)\right) \\ &= \exp\left(-\sum_{i=1}^n (1 - e^{-\beta_i})m(B_i)\right) \end{aligned}$$

For $f := \sum_{i=1}^n \beta_i 1_{B_i}$, this translates into (3.4).

c) Since any nonnegative measurable f is the pointwise limit of functions in b) , the assertion follows by dominated convergence (the continuous analogue of Lemma 1.4.1, cf. Kallenberg Thm 1.21).

Immediate from Propositions 3.1 and 3.2 is

Corollary 3.2.1 *a) A random point configuration Φ is a PPP with intensity measure m if (3.4) holds.*
b) Two PPP with the same intensity measure have the same distribution.

3.3 Construction

What about *existence* of a PPP for a given intensity measure m ? We will give a simple and useful construction, first in the special case of *finite* m.

Proposition 3.3.1 *Let m be a finite measure on E. Let N be a Poisson$(m(E))$-distributed r.v., and $U_1, U_2, \ldots$. be independent with distribution $m/m(E)$.*
Then $\Phi := \sum_{i=1}^N \delta_{U_i}$ is a PPP with intensity measure m.

Proof: Let $B_1, \dots, B_n \in \mathcal{E}$ be disjoint, and put

$$B_{n+1} := E \backslash (B_1 \cup \ldots \cup B_n).$$

For $k_1, \dots, k_{n+1} \in \mathbb{N}_0, \quad k := k_1 + \ldots + k_{n+1}$, we have

$$\begin{aligned}
&\mathbf{P}[\Phi(B_1) = k_1, \dots, \Phi(B_{n+1}) = k_{n+1}] \\
&= \mathbf{P}[N = k] \binom{k}{k_1, \dots, k_{n+1}} \left(\frac{m(B_1)}{m(E)}\right)^{k_1} \cdots \left(\frac{m(B_{k+1})}{m(E)}\right)^{k_{n+1}} \\
&= \frac{e^{-m(E)}}{k!} (m(E))^k \frac{k!}{k_1! \dots k_{n+1}!} \frac{m(B_1)^{k_1} \dots m(B_{k+1})^{k_{n+1}}}{m(E)^k} \\
&= \prod_{i=1}^{n+1} \frac{e^{-m(B_i)}}{k_i!} (m(B_i))^{k_i}
\end{aligned}$$

Hence we see that the $\Phi(B_i)$ are independent and Poisson $(m(B_i))$ distributed. □

The next result states that the independent superposition of PPP's is again a PPP, and the intensity measures add up.

Lemma 3.3.1 *Let $\Phi_1, \Phi_2, \dots$ be independent PPP's with intensity measures $m_1, m_2, \dots$. Then $\Phi := \sum_{i=1}^{\infty} \Phi_i$ is a PPP with intensity measure $m := \sum_{i=1}^{\infty} m_i$.*

Proof: We use Corollary 3.2.1:

$$\begin{aligned}
&\mathbf{E} \exp\left(-\int f(z) (\sum \Phi_i)(dz)\right) \\
&\qquad = \prod_i \mathbf{E} \exp\left(-\int f(z) \Phi_i(dz)\right) \\
&\qquad = \prod_i \exp\left(-\int (1 - e^{-f(z)}) m_i(dz)\right) \\
&\qquad = \exp\left(-\int (1 - e^{-f(z)}) (\sum m_i)(dz)\right). \quad \square
\end{aligned}$$

Corollary 3.3.1 *Let m be a σ-finite measure on E, that is, there exist $B_1 \subseteq B_2 \subseteq \dots$ such that $\bigcup B_n = E$ and $m(B_n) < \infty$ for all n. Then there exist finite measures m_i (even concentrated on disjoint sets) such that $m = \sum m_i$. Now construct a PPP Φ_i with intensity measure m_i as in Proposition 3.4. Then $\Phi := \sum_i \Phi_i$ is a PPP with intensity measure m.*

3.4 Independent labelling and thinning

Let $\Phi = \sum \delta_{Y_i}$ be a PPP. Attach to every point Y_i a label L_i whose distribution may depend on Y_i but is independent of all the other Y_j and all the other labels. We claim that $\Psi := \sum \delta_{(Y_i, L_i)}$ then is a PPP on the product space of positions and labels. To formalize this, let us specify a space $(\mathbb{L}, \mathcal{L})$ of labels, and a transition probability $P(y, d\ell)$ from E to $\mathbb{L}$ (that is, for all $y \in E, P(y, .)$ is a probability measure on $\mathbb{L}$, and for alll $G \in \mathcal{L}, y \mapsto P(y, G)$ is measurable.)

Proposition 3.4.1 *Let Φ be a PPP with σ-finite intensity measure m. Given $\Phi = \sum_i \delta_{y_i}$, let (L_i) be independent, and L_i have distribution $P(y_i, .)$. Then $\Psi := \sum_i \delta_{(Y_i, L_i)}$ is a PPP on $E \times \mathbb{L}$ with intensity measure $(m \otimes P)(dy, d\ell) := m(dy) P(y, d\ell)$.*

Proof: Proceeding like in Lemma 3.1 and Corollary 3.2, it suffices to consider the case $m(E) < \infty$. Use the construction of Φ given in Proposition 3.3. If U has distribution $m/m(E)$, and given $U = u, L$ has distribution $P(u,.)$, then (U, L) has distribution $\frac{m}{m(E)} \otimes P$. Thus, Ψ arises as $\sum_{i=1}^{N} \delta_{(U_i,L_i)}$, where N is Poisson distributed with parameter $m(E) = (m \otimes P)(E \times \mathbb{L})$, and the (U_i, L_i) are independent with distribution $(m \otimes P)/(m \otimes P)(E \times \mathbb{L})$. This identifies Φ as a PPP with intensity measure $m \otimes P$. □

Corollary 3.4.1 *Let Φ be a PPP on E with σ-finite intensity measure m , and $p : E \to [0,1]$ be measurable. Given $\Phi = \sum \delta_{y_i}$, for each i throw an independent coin with success probability $p(y_i)$, thus arriving at the labelled configuration$\sum \delta_{(y_i,L_i)}$, where $L_i \in \{0,1\}$ and $\mathbf{P}[L_i = 1] = p(y_i)$.*
Then $\chi := \sum_{i:L_i=1} \delta_{Y_i}$ is a PPP with intensity measure

$$m_p(B) := \int_B p(y)m(dy).$$

(Indeed, because of Proposition 3.3.1 $\Psi := \sum \delta_{(Y_i,L_i)}$ is a PPP on $E \times \{0,1\}$ with intensity measure $m(dy)(p(y)\delta_1 + (1 - p(y))\delta_0)$ and therefore $\sum_{i:L_i=1} \delta_{(Y_i,1)}$ is a PPP on $E \times \{1\}$ with intensity measure $m(dy)p(y)\delta_1)$. We call χ a *p-thinning* of Φ.

Example (Minimum of independent exponentially distributed random variables)
Let $\alpha_1, \alpha_2, \ldots > 0$ with $\sum \alpha_\ell =: \alpha < \infty$. Let W_ℓ, $\ell = 1, 2, \ldots$, be independent and Exp(α_ℓ)-distributed. We claim that $H := \min W_j$ is Exp(α)-distributed, and $\mathbf{P}[H = W_\ell] = \frac{\alpha_\ell}{\alpha}$.
Indeed, consider a homogeneous Poisson(α) process $(T_1, T_2, \ldots)$. Do an independent labelling

$$\mathbf{P}[L_i = \ell] = \frac{\alpha_\ell}{\alpha}.$$

The resulting point processes $(T_1^{(\ell)}, T_2^{(\ell)}, \ldots)$ are Poisson(α_ℓ) and independent.
Obviously,

$$\begin{aligned}
T_1 &= \min_j T_1^{(j)} \\
\mathcal{L}(T_1) &= \text{Exp}(\alpha) \\
\mathcal{L}(T_1^{(l)}) &= \text{Exp}(\alpha_l) \\
\mathbf{P}[T_1 = T_1^{(\ell)}] &= \alpha_\ell/\alpha, \quad \ell = 1, 2, \ldots
\end{aligned}$$

□

3.5 Poisson integrals, subordinators and Lévy processes

Let Φ be a PPP with σ-finite intensity measure m, and $f : E \longrightarrow \mathbb{R}$ be measurable.

Lemma 3.5.1 *a) For $f \geq 0$ or $\int | f | \, dm < \infty$,*

$$\mathbf{E} \int f(z)\Phi(dz) = \int f(z)m(dz) \tag{3.5}$$

b) For $\int | f | dm < \infty$,

$$\text{Var} \int f(z)\Phi(dz) = \int f^2(z)m(dz). \tag{3.6}$$

Proof: a) is clear from monotone convergence.
b) Again, we can restrict to finite m. For functions $f = \sum_{i=1}^{n} \beta_i 1_{B_i}, B_i$ pairwise disjoint $\beta_i \geq 0$, (3.6) is clear from independence and the fact that Var $Z = \alpha$ for a Poisson(α)-distributed Z (check!) Hence, for all such f, the *second moment* of $\int f(z)\Phi(dz)$ is

$$\mathbf{E}[(\int f(z)\phi(dz))^2] = \int f^2 dm + \int f dm. \tag{3.7}$$

Monotone convergence gives (3.7) for all measurable $f \geq 0$, and hence also (3.6) provided that $\int f dm < \infty$. Finally, split E into $\{f \geq 0\}$ and $\{f < 0\}$, and use independence. □

As a preparation for the Poisson representation of Lévy processes, we show two nice little lemmata.

Lemma 3.5.2 *Let* Φ *be a PPP with intensity measure* m*, and* $f : E \to \mathbb{R}_+$ *be measurable.*
a) If $f \geq 1$*, then*

$$\int f(z)\Phi(dz) < \infty \quad a.s. \Leftrightarrow m(E) < \infty$$

b) If $f \leq 1$*, then*

$$\int f(z)\Phi(dz) < \infty \quad a.s. \Leftrightarrow \int f(z)m(dz) < \infty$$

Proof: a) This is clear since both is equivalent to $\Phi(E) < \infty$ a.s.
b) "$\Rightarrow$": From $\mathbf{E}\exp(-\int f(dz)\Phi(dz)) > 0$ and Proposition 3.2.2 we have

$$\int (1 - e^{-f(z)})m(dz) < \infty.$$

Since $f \leq 1$, there exists a $c > 0$ such that $cf \leq 1 - e^{-f}$, hence $\int f(z)m(dz) < \infty$.
"$\Leftarrow$": By dominated convergence, we have

$$\lim_{c\to 0} \mathbf{E}\exp(-c\int f(z)\Phi(dz)) = \mathbf{P}[\int f(z)\Phi(dz) < \infty].$$

On the other hand, since by assumption

$$\infty > \int f(z)m(dz) \geq \int (1 - e^{-f(z)})m(dz),$$

we get again by dominated convergence

$$\lim_{c\to 0} \exp(-\int (1 - e^{-cf(z)})m(dz)) = \exp(0) = 1$$

Hence, together with Proposition 3.2 the assertion follows. □

Example (Poisson representation of the Gamma distribution)
Fix $k > 0$, and let Φ be a PPP on $\mathbb{R}_+$ with intensity measure m given by

$$\nu(dh) := k\frac{1}{h}e^{-h}dh \tag{3.8}$$

Consider the Poisson integral (or "'Poissonian superposition")

$$Z := \int h\,\Phi(dh) \tag{3.9}$$

We claim that

$$\mathbf{E}e^{-\beta Z} = (1+\beta)^{-k},\ \beta > 0.$$

Indeed, because of Proposition 3.2.2 and the well known fact that $\int\limits_{\mathbb{R}_+} h^n e^{-h}\,dh = n!$, $n \in \mathbb{N}$, we have for all $\beta \in [0,1]$

$$\begin{aligned}
\mathbf{E}e^{-\beta Z} &= \exp\left(-\int\limits_{\mathbb{R}_+}(1-e^{-\beta h})k\frac{1}{h}e^{-h}\,dh\right)\\
&= \exp\left(-k\int\limits_{\mathbb{R}_+}(-\sum_{j=1}^{\infty}\frac{(-\beta h)^j}{j!})\frac{1}{h}e^{-h}\,dh\right)\\
&= \exp\left(k\sum_{j=1}^{\infty}(-\beta)^j\frac{1}{j!}\int\limits_{\mathbb{R}_+}h^{j-1}e^{-h}\,dh\right)\\
&= \exp\left(-k\sum_{j=1}^{\infty}(-1)^{j-1}\frac{\beta^j}{j!}(j-1)!\right)\\
&= \exp\left(-k\sum_{j=1}^{\infty}(-1)^{j-1}\beta^j\frac{1}{j}\right)\\
&= \exp\left(-k\ln(1+\beta)\right) = (1+\beta)^{-k}.
\end{aligned}$$

Recall that for $k \in \mathbb{R}_+$, the *Gamma-distribution* with form parameter k (and scale parameter 1) (or *Gamma*(k)-distribution for short) has density

$$g_k(y) := \frac{1}{\Gamma(k)}y^{k-1}e^{-y}, \quad y > 0,$$

where

$$\Gamma(k) := \int\limits_0^{\infty} y^{k-1}e^{-y}\,dy, \quad k \in \mathbb{R}_+,$$

denotes the Γ-function. For a Gamma(k)-distributed Y we have

$$\mathbf{E}e^{-\beta Y} = \frac{1}{(1+\beta)^k}, \quad \beta > 0$$

(check!) So, because of Lemma 3.2.1b) we conclude that the "'Poissonian superposition" (3.9) represents a Gamma(k)-distributed random variable. □

Example Let ν be a measure on $\mathbb{R}_+$, with

$$\int y\,\nu(dy) < \infty. \tag{3.10}$$

Let $\Phi = \sum_i \delta_{(S_i,Y_i)}$ be a PPP on $\mathbb{R}_+ \times (\mathbb{R}\setminus\{0\})$ with intensity measure $m(d(s,y)) = ds \cdot \nu(dy)$, and put

$$X_t := \int_0^t \int_{\mathbb{R}_+} y\,\Phi(dy) = \sum_{S_i \le t} Y_i. \tag{3.11}$$

Then X has *homogeneous independent nonnegative increments*, that is:
(i) $\mathcal{L}(X_t - X_r)$ depends only on $t - r$,
(ii) $X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent if $t_0 < t_1 < \ldots < t_n$,
(iii) $X_t - X_r \geq 0$ if $t \geq r$.

A process X with the properties (i)-(iii) is called a *subordinator*. Perhaps the most prominent subordinator is the *Gamma process*. It is of the form (3.11) where ν is given by (3.8).

Notably, any subordinator is of the form

$$Z_t = c + bt + X_t,$$

where $b \geq 0$ and X is of the form (3.11) for *some* ν meeting the requirement (3.10) ((see O. Kallenberg, Foundations of Modern Probability, p. 290 (in the 2nd ed)).

Let us now come back to Poisson integrals. If we look, instead of the random point measure Φ, on the compensated (signed) random measure $\Phi - m$, we might hope that this integrates a larger class of functions f. Indeed, we will show that we can make sense out of $\int f(z)(\Phi - m)(dz)$ through a suitable limit procedure, provided that $\mid f \mid \leq 1$ and $\int f^2 dm < \infty$.

Lemma 3.5.3 *Let* $|f| \leq 1$ *and*

$$\int f^2(z) m(dz) < \infty \tag{3.12}$$

a) *Let* $B_1 \subseteq B_2 \subseteq \ldots, \bigcup B_n = E$, *and assume* $\mu(B_n) < \infty \quad \forall n$. *Then*

$$I_n := I_n(f) := \int f(z) 1_{B_n}(z)(\Phi - m)(dz)$$

converges, as $n \to \infty$, *in* L^2 *(or mean-square) to a random variable*

$$I(f) := \int f(z)(\Phi - m)(dz).$$

b) *The special choice of* (B_n) *doesn't matter so much:*
Assume $C_1 \subseteq C_2 \subseteq \ldots, \bigcup C_n = E$, $\mu(C_n) < \infty \quad \forall n$, *and assume further that each* C_n *is contained in some* B_k, $k \geq n$. *Then*

$$J_n := \int f(z) 1_{C_n}(z)(\Phi - m)(dz) \to I(f) \text{ in } L^2.$$

Proof: a) For $n \geq k$, because of Lemma 3.5.1 b) and dominated convergence,

$$\mathbf{E}[(I_n - I_k)^2] = \int_{B_n \setminus B_k} f^2(z) m(dz) \leq \int_{E \setminus B_k} f^2(z) m(dz) \longrightarrow 0 \text{ as } k \to \infty \text{ as } k \to \infty.$$

Thus, (I_n) is a Cauchy sequence in the sense of mean square (or L^2-) convergence and hence converges in L^2 towards some random variable J (cf. Kallenberg, Lemma 1.31).

b) For all $n \in \mathbb{N}$ let $k = k(n) \geq n$ be such that $C_n \subseteq B_k$. Then

$$\mathbf{E}[(J_n - I_k)^2] = \int_{B_k \setminus C_n} f^2(z)\, m(dz) \leq \int_{E \setminus C_n} f^2(z)\, m(dz) \to 0 \text{ as } n \to \infty,$$

hence also (J_n) converges in L^2 to $I(f)$. □

Example Let ν be a measure on $\mathbb{R} \setminus \{0\}$, with

$$\int (y^2 \wedge 1)\, \nu(dy) < \infty. \tag{3.13}$$

Let Φ be a PPP on $\mathbb{R} \setminus \{0\}$ with intensity measure $m(d(s,y)) = ds \cdot \nu(dy)$, and put

$$X_t := \int_0^t \int_{|y| \le 1} y\, (\Phi - m)(d(s,y)) + \int_0^t \int_{|y|>1} y\, \Phi(d(s,y)), \tag{3.14}$$

where the first integral is defined according to Lemma 3.5.3, with $B_n := [0,t] \times ([-1,1] \setminus [-1/n,\, 1/n])$. Then X has homogeneous independent increments, that is:
(i) $\mathcal{L}(X_t - X_r)$ depends only on $t - r$,
(ii) $X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent if $t_0 < t_1 < \ldots < t_n$.

A process X with the properties (i) and (ii) is called a *Lévy process*. Notably, any Lévy process is of the form

$$Z_t = c + bt + \sigma W_t + X_t,$$

where W is a standard Wiener process (we'll come back to this later), $\sigma \ge 0$, and X is of the form (3.14) for *some* ν meeting the requirement (3.13) (see O. Kallenberg, Foundations of Modern Probability, p. 290 (in the 2nd ed)).

Chapter 4

Markov chains in continuous time

4.1 Jump rates

Like in the first lesson, we start by considering a finite or countable set S_0, and a stochastic matrix Π on S_0. Other than in chapter 1, however, time is now thought to be continuous. Moreover, we introduce state-dependent rates $\alpha_x, x \in S_0$ (having in mind that in different states, time may pass in different speed). We think of α_x as the parameter of an exponential distribution: when starting in x, our process keeps waiting there for an Exp(α_x)-distributed time, then moves to y with probability $\Pi(x, y)$, then keeps waiting there for an independent Exp(α_y)-distributed waiting time, and so on.
Question: What is the distribution of the time at which our process, when starting in x, jumps for the first time *away* from x (the so-called *holding time* in x)? The "jumping away" happens already after the first Exp(α_x)-distributed waiting time with probability $p := 1 - \Pi(x, x)$. With probability $\Pi(x, x)$, however, there follows another independent Exp(α_x)-distributed waiting time, and so on. Overall, we are faced with a p-thinning of a Poisson(α_x)-process. Thus, the holding time in x has an exponential distribution with parameter

$$q_x := \alpha_x p = \alpha_x(1 - \Pi(x, x)).$$

At this time, the process jumps to $y\,(\neq x)$ with probability

$$J(x, y) := \frac{\Pi(x, y)}{1 - \Pi(x, x)}.$$

(Here, we assume that $\Pi(x, x) < 1$; otherwise, q_x would be zero and the process would remain in x forever.)
Let us forget about α and Π, and fix q and J as our basic ingredients.
$q_x \geq 0, \quad x \in S_0$ are called the *jump rates*,
$J(x, y) \geq 0, \quad x \neq y \in S_0$, with the property

$$\sum_{y \in S_0 \setminus \{x\}} J(x, y) = 1,$$

is called the *jump matrix*.
We now define a random path $(X_t)_{t \leq 0}$ starting in $x \in S_0$. After an Exp(q_x)-distributed time H_0, jump to y with probability $J(x, y)$.
Then after an Exp (q_y)-distributed time H_1 (independent of H_0), jump to z with

probabiltiy $J(y, z)$, and so on.
The process (X_t) starting in x can be defined in the following way (check !):
Let $(Y_0, Y_1 \dots)$ be a discrete time Markov chain starting in x with transition matrix J. Given $(Y_0, Y_1, \dots) = (y_0, y_1, \dots)$, let H_i be independent and Exp (q_{y_i})-distributed. Put

$$\begin{aligned} X_t &:= x \quad \text{for } x \leq t < H_0 \\ X_t &:= y_i \quad \text{for } H_0 + \dots + H_{i-1} \leq t < H_0 + \dots + H_i, i \geq 1 \end{aligned}$$

4.2 The minimal process and its transition semigroup

Does for given q and J the construction in section 4.1 define the random path (X_t) for all times $t \geq 0$? Yes, if the time axis $[0, \infty)$ is exhausted by the sum of the holding times.
Note: The construction defines X_t only for $t < H_0 + H_1 + \dots =; \zeta$
If $\zeta < \infty$, we say that X *explodes*, and call ζ the *explosion time*.

Here is an example for explosion:

$$S_0 = \mathbb{N}, \quad q_k = k^2, \quad J(k, k+1) = 1.$$

Starting in $x = 1$ we have

$$\mathbf{E}[\zeta] = \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty.$$

Let us now extend the construction (4.1) extend beyond ther time ζ.

Definition 4.2.1 : *Let Δ be an element not belonging to S_0 and put $S_\Delta := S_0 \cup \{\Delta\}$.*

The *minimal process* X following the dynamics (q, J) is constructed as above for $t < \zeta$, and set equal to Δ for $t \geq \zeta$. Thus, $X = (X_t)_{t \geq 0}$ is a random variable taking its values in the right-continuous S_Δ-valued paths which never return from Δ.
By construction, our X obeys the *Markov property*:

$$\mathbf{P}_x[X_{s+t} = z \mid X_{s_1} = y_1, \dots, X_s = y] = \mathbf{P}_y[X_t = z] \tag{4.1}$$
$$(s_1 \leq \dots \leq s, \, t > 0, \, x, y_1, \dots, y, \, z \in S_\Delta) \tag{4.2}$$

We put

$$P_t^\Delta(x, y) := \mathbf{P}_x[X_t = y], \quad x, y \in S_\Delta \tag{4.3}$$

and

$$P_t(x, y) := \mathbf{P}_x[X_t = y], \quad x, y \in S_0. \tag{4.4}$$

Note that $P_t(x, S_0) = 1$ is guaranteed only if $\mathbf{P}_x[\zeta > t] = 1$. In general, we have

$$P_t(x, S_0) \leq 1 \tag{4.5}$$

and consequently call P_t a *substochastic* matrix.
The law of total probability and the non-returning from Δ gives

$$P_{s+t}(x, y) = \sum_{z \in S_0} P_s(x, z) P_t(z, y) =: (P_s P_t)(x, y), \quad x, y \in S_0 \tag{4.6}$$

We say that (P_t) satisfies the *Chapman-Kolmogorov equations* (or, that it is a *semigroup* of substochastic matrices). For short, we call (P_t) the *transition semigroup* of (X_t). There is a formal analogy between the relation

$$P_{s+t} = P_s P_t, \quad s, t \geq 0; \quad P_0 = I := \text{ identity matrix} \tag{4.7}$$

and the relation

$$f(s+t) = f(s) \cdot f(t), \quad s, t \geq 0, \quad f(0) = 1. \tag{4.8}$$

Equations(4.8) are satisfied by $f(t) := e^{\alpha t}, \alpha \in \mathbb{R}$, which obeys the differential equation

$$\frac{d}{dt} f(t) = \alpha f(t) \tag{4.9}$$

and

$$\frac{d}{dt} f(t) \mid_{t=0} = \alpha. \tag{4.10}$$

We'll explore the counterpart of (4.10) for our semigroup (P_t). To this end, let's analyze (P_t) near $t = 0$, and start with some heuristics. Neglecting the effect of multiple jumps in small time intervals (which in fact can be justified) we have

$$P_h(x,x) = e^{-q_x h} + o(h) = 1 - q_x \cdot h + o(h), P_h(x,y) = h q_x J(x,y) + o(h), \quad x \neq y \tag{4.11}$$

which can be written compactly as

$$P_h = I + hQ + o(h), \tag{4.12}$$

where I denotes the identity matrix on S_0 , and

$$Q(x,y) := \begin{cases} -q_x & x = y \\ q_x J(x,y) & x \neq y \end{cases} \tag{4.13}$$

is the so-called *Q-matrix associated with* q and J. Since

$$P_0(x,y) = \mathbf{P}_x[X_0 = y] = \delta_{xy} = I(x,y),$$

(4.12) translates into

$$\frac{d}{dt} P_t \mid_{t=0} = Q.$$

4.3 Backward and forward equations

We have seen in the previous section that the semigroup property of (P_t) is intimately connected with the law of total probability. We can now apply the "total probability decomposition" near time 0 or near time t. This will give us two systems of differential equations for P_t, called the backward and the forward euqations, respectively. In a nutshell, the argument is as follows:

$$P_{h+t} - P_t = P_h P_t - P_t = (P_h - I) P_t, \tag{4.14}$$

thus (assuming that limits and summations can be interchanged)

$$\frac{d}{dt} P_t = Q P_t \tag{4.15}$$

On the other hand

$$P_{t+h} - P_t = P_t P_h - P_t = P_t(P_h - I), \tag{4.16}$$

thus (again assuming that limits and summations can be interchanged)

$$\frac{d}{dt} P_t = P_t Q. \tag{4.17}$$

(4.15) and (4.17) are called the *backward* and *forward* equations. We'll agree to understand them component-wise. Written more explicitly, (4.15) reads as

$$\frac{d}{dt} P_t(x, y) = \sum_{z \in S_0} Q(x, z) P_t(z, y) \quad (x, y \in S_0) \tag{4.18}$$

and (4.17) reads as

$$\frac{d}{dt} P_t(x, y) = \sum_{z \in S_0} P_t(x, z) Q(z, y) \qquad x, y \in S_0. \tag{4.19}$$

Thus, for fixed $y \in S_0$, the backward equations (4.18) are a system of differential equations for the $P_t(x, y)$, $x \in S_0$, and for fixed $x \in S_0$, the forward equations (4.19) are a system of differential equations for the $P_t(x, y)$, $y \in S_0$.
Let us illustrate this point still more. Take a real valued function $f = f(y)$, and a probability meausre $\mu = \mu(x)$. How do the expectations $\mathbf{E}_x[f(X_t)] =: u(t, x)$ and the probabilties $\mathbf{P}_\mu[X_t = y] =: \mu_t(y)$ evolve in time ? We have

$$\begin{aligned} u(t, x) &= \sum_y P_t(x, y) f(y) =: P_t f(x) \\ \mu_t(y) &= \sum_x \mu(x) P_t(x, y) =: \mu P_t(y) \end{aligned}$$

Again assuming that limits and summations interchange, we get that u satisfies the backward equations

$$\begin{aligned} \frac{\partial}{\partial t} u(t, x) &= \sum_y \sum_z Q(x, z) P_t(z, y) f(y) \\ &= \sum_z Q(x, z) u(t, z) =: (Q u(t, .))(x) \end{aligned}$$

and μ satisfies the forward equation

$$\begin{aligned} \frac{\partial}{\partial t} \mu_t(y) &= \sum_x \mu(x) \sum_z P_t(x, z) Q(z, y) \\ &= \sum_z \mu_t(z) Q(z, y) =: (\mu_t Q)(y) \end{aligned}$$

We are now going to prove that P_t defined by (4.4) satisfies the backward equation (4.15). This will be achieved by establishing an integral equation equivalent to (4.15) through a "first jump decomposition".

Proposition 4.3.1 *Consider jump rates q_x and a jump matrix J, and define the matrix Q as in (4.13). Let (X_t) be the minimal process constructed in section 4.2, and P_t its transition semigroup defined by (4.4). Then (P_t) satisfies the backward equations (4.18).*

Proof: Following the strategy of a "first jump decomposition", we obtain

$$\begin{aligned}
P_t(x,y) &= \mathbf{P}_x[H_0 > t, X_t = y] + \sum_{z\neq x} \mathbf{P}_x[H_0 \le t, X_{H_0} = z, X_t = y] \\
&= e^{-q_x t}\delta_{xy} + \sum_{z\neq x} \int_0^t q_x e^{-q_x \cdot s} ds\, J(x,z) P_{t-s}(z,y) \\
&= e^{-q_x t}\delta_{xy} + \sum_{z\neq y} \int_0^t q_x e^{-q_x (t-u)} du\, J(x,z) P_u(z,y)
\end{aligned}$$

Multiplying by $e^{q_x t}$ we arrive at

$$e^{q_x t} P_t(x,y) = \delta_{xy} + \int_0^t \sum_{z\neq x} q_x e^{q_x u} du J(x,z) P_u(z,y).$$

Hence, taking the derivative with respect to t,

$$e^{q_x t}[q_x P_t(x,y) + \frac{d}{dt} P_t(x,y)] = e^{q_x t} \sum_{z\neq x} q_x J(x,z) P_t(z,y)$$

or in other words

$$\frac{d}{dt} P_t(x,y) = (QP_t)(x,y).$$

□

The proof of the next proposition is similar in spirit but slightly more involved than the previous one: Here one decomposes according to the last jump before t, and uses a time reversal argument. We won't give the details, but refer to J.R. Norris, Markov Chains, CUP, 1997, p 100-103.

Proposition 4.3.2 *Let Q and P_t be as in Proposition 4.3.1. Then (P_t) satisfies also the forward equations (4.17).*

The previous two propositions describe how to construct, starting from a given Q as in (4.13), a "probabilistic" solution to Kolmogorov's equations (4.18) and (4.19) in terms of the minimal process. The next proposition states that this solution is in fact the *minimal* one.

Proposition 4.3.3 *The transition probabilities $\mathbf{P}_x[X_t = y]$ are the minimal non-negative solutions both of the backward equations (4.18) and the forward equations (4.19), always with initial condition $P_0(x,y) = \delta_{xy}$.*

Proof: see Norris, loc.cit., p.98 and p.100.

We conclude the subsection with a statement on the equivalence of the differential and the integral form of the backward and forward equation, respectively. For this, let $Q = Q(x,y)$ be a matrix with non-negative entries off the diagonal, and

$$\sum_{y\neq x} Q(x,y) \le -Q(x,x) =: q_x < \infty, \quad x \in S_0.$$

(So Q may be of a slightly more general form than in (4.13).)

Proposition 4.3.4 *a) $P_t(x,y)$, x, $y \in S_0$, satisfies the backward differential equation (4.18) iff it satisfies the backward integral equation*

$$P_t(x,y) = \delta_{xy} e^{-q_x t} + \int_0^t e^{-q_x s} \sum_{z \neq x} Q(x,z) P_{t-s}(z,y) ds, \quad t \geq 0; \quad x,\, y \in S_0. \tag{4.20}$$

b) $P_t(x,y)$, x, $y \in S_0$, satisfies the forward differential equation (4.19) iff it satisfies the forward integral equation

$$P_t(x,y) = \delta_{xy} e^{-q_x t} + \int_0^t e^{-q_y s} \sum_{z \neq y} P_{t-s}(x,z) Q(z,y) ds, \quad t \geq 0; \quad x,\, y \in S_0. \tag{4.21}$$

Proof: see e.g. W.J. Anderson, Continuous Time Markov Chains, Springer 1981, Propositions 2.1.1 and 2.1.2. □

4.4 Revival after explosion

We saw in the previous sections how to construct, for a given Q-matrix as in (4.13), an S_Δ-valued Markov chain (X_t) (the minimal process) whose transition semigroup (P_t) obeyed

$$\frac{d}{dt} P_t \,|_{t=0} = Q \tag{4.22}$$

If starting from x, an explosion in finite time happens with positive probability then we have for some $t > 0$

$$P_t(x, S_0) < 1$$

(and we say that (P_t) is *non-conservative*). Can we modify (X_t) such that

$$\mathbf{P}_x[X_t \in S_0] = P_t(x, S_0) = 1 \text{ for all } t \text{ and } x,$$

and still (4.22) is valid ? Indeed, we can, and even in many ways:
Let π be an arbitrary probability distribution on S_0, and let the process, instead of remaining in Δ for $t \geq \zeta$, jump back at time ζ into S_0, arriving at z with probability $\pi(z)$. After the next explosion, apply the same procedure independently, and so on.

The transition semigroup (P_t) of our new Markov chain (X_t) then has the following properties:

$$P_t \text{ is a \textit{stochastic} matrix on } S_0, \quad t \geq 0, \tag{4.23}$$

$$P_0 = I, \tag{4.24}$$

$$P_{s+t} = P_s P_t, \quad s,\, t \geq 0, \tag{4.25}$$

$$\lim_{t \downarrow 0} P_t(x,x) = 1. \tag{4.26}$$

4.5 Standard transition semigroups and their Q-matrices

Let's turn the tables and start from a family (P_t) satisfying (4.23) to (4.26). Such a family is sometimes called a *standard transition semigroup.*

Proposition 4.5.1 *:*

$$\lim_{h\to 0}\frac{1}{h}(P_h - I)(x,y) =: Q(x,y) \tag{4.27}$$

exists in $\mathbb{R}\cup\{-\infty\}$.
(i) For $x = y$, $\quad Q(x,x) \in [-\infty, 0]$
(ii) For $x \neq y$, $\quad Q(x,y) \in [0,\infty)$
(iii) For all x, $\quad \sum\limits_{y\neq x} Q(x,y) \leq -Q(x,x)$.

Proof: see S. Karlin, H.M. Taylor: A second course in stochastic processes, Academic press 1981, p.139-142.

Definition 4.5.1 *a) A matrix* Q *with the above stated properties (i), (ii), (iii) is called a* Q*-matrix.*
b) Q *as defined in (4.27) is called the* Q*-matrix of the semigroup* (P_t).

Definition 4.5.2 *Let* Q *be the* Q*-matrix of a standard semigroup* (P_t)*. A state* x *is called*

- instantaneous *if* $Q(x,x) = -\infty$
- stable *if* $Q(x,x) > -\infty$
- conservative *if it is stable and* $\sum\limits_{y} Q(x,y) = 0$.

What is the probabilistic meaning of an instantaneous state x ? The process should jump away immediately from x, but because of (4.26) should for small times be in x with probability close to 1. Is such a thing possible ?
And what is the probabilistic meaning of a stable non-conservative state x ? Because of

$$\sum_{y\neq x}\frac{1}{q_x}Q(x,y) < 1,$$

the process should get lost from S_0 for a moment with positive probability (at the random time when it jumps away from x), but because of (4.23) should return immediately to S_0.

These two effects are illustrated by two nice examples due to Kolmogorov, now known as K1 and K2 (cf W.J. Anderson, loc.cit, p.28-32, and K.L. Chung, Markov chains with stationary transition probabilities, 2nd ed., Springer 1967, p.275 ff)
We will outline them briefly, starting with K2.

Example K2
Consider $S_0 := \mathbb{N}_0$, and the Q-matrix

$$Q = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & 4 & -4 & 0 & 0 & \dots \\ 0 & 0 & 9 & -9 & 0 & \dots \\ 0 & 0 & 0 & 16 & -16 & \dots \\ . & . & . & . & . & \dots \end{pmatrix}$$

Construct (X_t) as follows: Starting from $x = k \geq 1$, things are simple: with $\mathrm{Exp}(n^2)-$ distributed holding times, the process jumps down to state 1 where it remains forever. Starting from the stable but non-conservative state $x = 0$, the

process jumps " to ∞" after an Exp(1)-distributed holding time, and from there performs an immediate "implosion" until it comes to eternal rest in state 1.
For this, let $W_n, n = 1, 2, \ldots$ be independent and Exp(n^2)-distributed, and put

$$X_t := \begin{cases} 0 & \text{if} \quad 0 \le t < W_1 \\ n & \text{if} \quad W_1 + \sum\limits_{k=n+1}^{\infty} W_k \le t < W_1 + \sum\limits_{k=n}^{\infty} W_k, \quad n > 1 \\ 1 & \text{if} \quad \sum\limits_{k=1}^{\infty} W_k \le t \end{cases}$$

(Note that $\sum\limits_{k=1}^{\infty} W_k < \infty$ a.s., since $\mathbf{E}[\sum\limits_{k=1}^{\infty} W_k] = \sum\limits_{k=1}^{\infty} \frac{1}{k^2} < \infty$.)

Example K1

Consider $S_0 := \mathbb{N}_0$, and the Q-matrix

$$Q = \begin{pmatrix} -\infty & 1 & 1 & 1 & \ldots \\ 1 & -1 & 0 & 0 & \ldots \\ 4 & 0 & -4 & 0 & \ldots \\ 9 & 0 & 0 & -9 & \ldots \\ \cdot & \cdot & \cdot & \cdot & \ldots \end{pmatrix}$$

The intuition about Q-matrices, which we developed in the previous sections, seems to leave us in the lurch. How should it be possible to jump away from the instantaneous state $x = 0$ immediately and uniformly to $1, 2, \ldots$, and still be back to state 0 after a short time with high probabiltiy ?
Things clear up if one first considers, for $M \in \mathbb{N}$, the Q-matrix

$$Q_{M^*} = \begin{pmatrix} -M & 1 & 1 & \ldots & 1 \\ 1 & -1 & 0 & \ldots & 0 \\ 4 & 0 & -4 & \ldots & 0 \\ \vdots & & & & \\ M^2 & 0 & 0 & \ldots & -M^2 \end{pmatrix}$$

on $S_M := \{0, 1, \ldots, M\}$.

A path starting in 0 remains there for an Exp(M)-distributed time, then chooses uniformly a $k \in \{1, \ldots, M\}$ where it stays for an Exp(k^2)-distributed time, then jumps back to state 0, where it remains for another (independent) Exp(M)-distributed waiting time and so on.
A crucial idea is now to sum up all the holding times in 0 along a time axis which counts only the time spent in 0 (the so called "local time" in 0).
We can now construct a random path $X^{(M)}$ following the Q_M-dynamics:
Let $\Phi^{(M)} = \sum\limits_i \delta_{(L_i, K_i)}$ be a Poissson process on $\mathbb{R}_+ \times \{1, 2, \ldots, M\}$, homogeneous with unit intensity on all $\mathbb{R}_+ \times \{k\}, K = 1, \ldots M$. Every point in Φ stands for an excursion from state 0; a point (L_i, K_i) means that at local time L_i the process jumps to state K_i.
Given the points (L_i, K_i), attach independent Exp(K_i^2)-distributed time spans W_i (the duration of excursion no. i). Excursion no. i starts at real time

$$T_i := L_i + \sum_{j: L_j < L_i} W_j \tag{4.28}$$

and ends at real time $T_i + W_i$; during the time interval $[T_i, T_i + W_i)$ the process is in state K_i.

Thus,

$$X_t^{(M)} := \sum_i K_i 1_{[T_i, T_i+W_i)}(t) \tag{4.29}$$

defines an $\{0, \dots, M\}$-valued random walk following the Q_M-dynamics.
The same construction can be carried out for the "full picture":
Let $\Phi := \sum_i \delta_{(L_i,K_i)}$ be a Poisson process on $\mathbb{R}_+ \times \mathbb{N}$, homogeneous with unit intensity on all the $\mathbb{R}_+ \times \{k\}$, $k \in \mathbb{N}$. Again, attach independent, $\mathrm{Exp}(K_i^2)$ distributed labels W_i.
The crucial observation is that, although infinitely many excursions from 0 happen up to a positive local time $l > 0$, the total time A_l spent outside of state 0 up to local time l remains finite. Indeed, we compute its expectations as

$$\begin{aligned}
\mathbf{E}[A_l] &= \mathbf{E}[\sum_{j:L_j<l} W_j] \\
&= \sum_{k=1}^{\infty} \mathbf{E}[\sum_{\substack{j:L_j<l,\\ K_j=k}} W_j] \\
&= \sum_{k=1}^{\infty} \frac{1}{k^2} \mathbf{E}\#\{j : L_j < l, K_j = k] \\
&= l \cdot \sum_{k=1}^{\infty} \frac{1}{k^2} < \infty.
\end{aligned}$$

We define T_i as in (4.28), but now in terms of Φ (instead of $\Phi^{(M)}$). Noting that $T_i < \infty$ a.s. because of the previous estimate, we can define X_t as in (4.29).

Is it indeed true that (4.26) is met, i.e. that $\mathbf{P}_0[X_t = 0] \to 1$ as $t \to 0$? Yes! we won't give a formal proof, but content ourselves with an observation which hits the core of the mattter.
Claim: For small l, the process stays with high probability only for a short fraction of time outside of 0.
(Intuition: there were so much more very short excursions than long ones.)
Claim reformulated: For all $\varepsilon > 0$

$$\mathbf{P}[A_l \le \varepsilon l] \ge 1 - \varepsilon \quad \text{for } l \text{ sufficiently small.}$$

Proof: Write $A_l = \sum_{k=1}^{\infty} A_{l,k}$, where

$$A_{l,k} := \sum_{j:L_j<l;K_j=k} W_j$$

We have:

$$\mathbf{E}[A_{l,k}] = l \frac{1}{k^2}$$

Choose M so big that $\sum_{k=M+1}^{\infty} \frac{1}{k^2} < \frac{\varepsilon^2}{2}$. Let l be so small that

$$\mathbf{P}[\Phi \text{ has a point in } [0,l] \times \{1, \dots, M\}] < \frac{\varepsilon}{2}.$$

Then, by Markov's inequality ($\mathbf{P}[Z > c] \leq \frac{1}{c}\mathbf{E}[Z]$),

$$\begin{aligned}\mathbf{P}[A_l > \varepsilon l] \leq \mathbf{P}[\sum_{k=1}^{M} A_{k,l} > 0 \text{ or } &\sum_{k=M+1}^{\infty} A_{k,l} > \varepsilon \cdot l] \\ &\leq \frac{\varepsilon}{2} + \frac{1}{\varepsilon l} \cdot l \frac{\varepsilon^2}{2} \leq \varepsilon.\end{aligned}$$

□

Chapter 5

Conditional Expectation

Let Z be an $\bar{\mathbb{R}}$-valued random variable which is non-negative or obeys $\mathbf{E}[|Z|] < \infty$ (in the latter case Z is called *integrable*). For an event A with $\mathbf{P}[A] > 0$ we call

$$\mathbf{E}[Z|A] := \frac{\mathbf{E}[ZI_A]}{\mathbf{P}[A]} \tag{5.1}$$

the *conditional expectation* of Z , given A.

Now assume we are interested in events $A = \{X = x\}$, $x \in S$, where S is some discrete space and X is an S valued random variable.
Writing

$$\varphi(x) := \mathbf{E}[Z \mid \{X = x\}] \tag{5.2}$$

for all $x \in S$ with $\mathbf{P}[X = x] > 0$, we have found a random variable $Y := \varphi(X)$ which for all $B \subseteq S$ obeys

$$\mathbf{E}[ZI_{\{X \in B\}}] = \mathbf{E}[\varphi(X)I_{\{X \in B\}}.] \tag{5.3}$$

Indeed,

$$\begin{aligned} \mathbf{E}[ZI_{\{X\in B\}}] &= \sum_{x\in B:\mathbf{P}[X=x]>0} \mathbf{E}[ZI_{\{X=x\}}] \\ &= \sum_{x\in B:\mathbf{P}[X=x]>0} \varphi(x)\mathbf{P}[X=x] = \mathbf{E}[\varphi(X)I_{\{X\in B\}}]. \end{aligned}$$

In view of (5.3), it makes sense to call the random variable $Y := \varphi(X)$ the conditional expectation of Z given X.
Now let us turn to the case of uncountable S. Then typically $\mathbf{P}[X = x] = 0$, and we are in trouble with (5.2). However, it still makes sense to require (5.3).

Definition 5.0.3 *Let Z be an $\bar{\mathbb{R}}$-valued random variable. Assume $Z \geq 0$ or $\mathbf{E}|Z| < \infty$. In addition, let X be an S-valued random variable, where $(S, \mathcal{S})$ is some measurable space. We call a random variable $\varphi(X)$* conditional expectation of Z given X *if*

$$\mathbf{E}[ZI_{\{X\in B\}}] = \mathbf{E}[\varphi(X)I_{\{X\in B\}}] \tag{5.4}$$

for all $B \in \mathcal{S}$.

The conditional expectation of Z given X is a.s. unique. This is a corollary of the following

Lemma 5.0.1 *Assume that φ_1 and φ_2 obey*

$$\mathbf{E}[\varphi_1(X)I_{\{X\in B\}}] \le \mathbf{E}[\varphi_2(X)I_{\{X\in B\}}] \quad \text{for all } B \in \mathcal{S} \tag{5.5}$$

Then

$$\varphi_1(X) \le \varphi_2(X) \ \ a.s. \tag{5.6}$$

Proof: Put $B := \{x \mid \varphi_1(x) > \varphi_2(x)\}$. Then $0 \ge \mathbf{E}[(\varphi_1(X) - \varphi_2(X))I_{\{\varphi_1(X)>\varphi_2(X)\}}]$. On the other hand, $Y := (\varphi_1(X) - \varphi_2(X))I\{\varphi_1(X) > \varphi_2(X)\} \ge 0$. Together, this implies that $Y = 0$ a.s., which enforces (5.6). □

Notation: If $\varphi(X)$ meets (5.4), we write $\varphi(X) = \mathbf{E}[Z|X]$ a.s. What about existence of conditional expectations ?
There is a beautiful geometrical picture which gives this existence (almost) for free. To begin with, let Z be *square integrable*, i.e.

$$\mathbf{E}Z^2 < \infty.$$

Look for a random variable $\varphi(X)$ which, among all those of the form $\psi(X)$, minimizes the *mean square distance* $\mathbf{E}[(\psi(X) - Z)^2]$. We claim that $\varphi(X) = \mathbf{E}[Z \mid X]$ a.s.
(This is not too astonishing if one remembers that $\mathbf{E}[Z]$ is *that* constant, which among all constants c, minimizes $\mathbf{E}[(c - Z)^2]$.)
It remains to make sure that
a) the problem "minimze $\mathbf{E}[(\psi(X) - Z)^2]$" indeed has a solution
b) the solution obeys (5.4).

We won't go into every detail, but just state that the space $\mathcal{L}^2$ of square integrable random variables (more precisely, the space of L^2 of equivalence classes of square integrable random variables being almost surely equal) carries a scalar product given by

$$< Y_1, Y_2 >:= \mathbf{E}[Y_1 Y_2],$$

generating the norm $|| \ Y \ ||:= \mathbf{E}[Y^2]^{1/2}$. This norm is complete (i.e. every Cauchy sequence has an a.s. limit in $\mathcal{L}^2$), and the subspace $\mathcal{L}^2(X)$ of all square integrable random variables of the form $\psi(X)$ is complete as well. Let $\varphi(X)$ be the orthogonal projection of Z on $\mathcal{L}^2(X)$. Then $Z - \varphi(X)$ is orthogonal to all $Y \in \mathcal{L}^2(X)$, in particular also to $1_B(X) = I_{\{X\in B\}}$. That is

$$< Z, I_{\{X\in B\}} > = < \varphi(X), I_{\{X\in B\}} > \ \text{ for all } B \in \mathcal{S},$$

which is nothing but (5.4). This guarantees already the existence of $\mathbf{E}[Z|X]$ for $Z \in \mathcal{L}^2$.
For an arbitrary random variable $Z \ge 0$, put $Z_n := \min(Z, n)$ and, noting that $Z_n \in \mathcal{L}^2$, put $\varphi_n(X) := \mathbf{E}[Z_n|X]$. Because of Lemma 5.2 we have $\varphi_n(X) \uparrow$ a.s. Writing $\varphi(X)$ for the a.s. limit of $\varphi_n(X)$, we obtain from monotone convergence for all $B \in \mathcal{S}$

$$\begin{aligned}
\mathbf{E}[ZI_{\{X\in B\}}] &= \lim_n \mathbf{E}[Z_n I_{\{X\in B\}}] \\
&= \lim_n \mathbf{E}[\varphi_n(X) I_{\{X\in B\}}] = \mathbf{E}[\varphi(X) I_{\{X\in B\}}],
\end{aligned}$$

which is (5.4).
This guarantees existence of $\mathbf{E}[Z|X]$ for any non-negative random variable Z. Finally, for a real-valued random variable Z with $\mathbf{E}[|Z|] < \infty$, decompose Z in its positive and negative part ($Z = Z^+ - Z^-$) and put

$$\mathbf{E}[Z|X] := \mathbf{E}[Z^+|X] - \mathbf{E}[Z^-|X].$$

Overall, we have proved:

Theorem 5.0.1 *Let Z and X be as in Definition 5.0.3. Then $\mathbf{E}[Z|X]$ exists and is a.s. unique.*

We now show that (5.4) extends from the indicator functions to all bounded measurable $f : S \to \mathbb{R}$.

Proposition 5.0.2 *Let Z and X be as in Definition 5.0.3 and $\varphi(X) := \mathbf{E}[Z|X]$. Then*

$$\mathbf{E}[Zf(X)] = \mathbf{E}[\varphi(X)f(X)] \tag{5.7}$$

for all bounded measurable $f : S \to \mathbb{R}$.

Proof: For non-negative f , approximate f from below by functions of the form $\sum c_k 1_{B_k}$ and obtain (5.7), using (5.4) , linearity of the expectation and monotone convergence. For general f, write $f = f^+ - f^-$ and again use linearity □

Let us now collect some important properties of conditional expectations.
Fact 5.1: (Law of total probability)

$$\mathbf{E}[\mathbf{E}[Z|X]] = \mathbf{E}[Z]$$

(put $B := S$ in (5.4))
Fact 5.2: (Respect what you completely depend on)
If Z depends completely on X, i.e. $Z = g(X)$ for some $g : \mathcal{S} \to \overline{\mathbb{R}}$,then

$$\mathbf{E}[g(X)|X] = g(X) \text{ a.s.}$$

(since (5.4) is clearly satisfied with $Z = g(X) = \varphi(X)$)
Fact 5.3: (Ignore what you are independent of)
If Z and X are independent, then

$$\mathbf{E}[Z|X] = \mathbf{E}[Z] \text{ a.s.}$$

(since $\mathbf{E}[ZI_{\{X\in B\}}] = \mathbf{E}[Z]\mathbf{P}[X \in B] = \mathbf{E}[\mathbf{E}[Z]I_{\{X\in B\}}]$)
Fact 5.4: (Linearity of conditional expectation)

$$\mathbf{E}[\alpha Z_1 + \beta Z_2 \mid X] = \alpha\mathbf{E}[Z_1 \mid X] + \beta\mathbf{E}[Z_2 \mid X] \quad \text{a.s.}$$

(check!)
Fact 5.5: (Monotonicity of conditional expectation)

$$Z_1 \le Z_2 \text{ a.s} \implies \mathbf{E}[Z_1|X] \le \mathbf{E}[Z_2|X] \quad \text{a.s.}$$

Lemma 5.0.2 *(Monotone convergence of conditional expectations)*
If $0 \le Z_n \uparrow Z$ a.s., then

$$\mathbf{E}[Z_n|X] \uparrow \mathbf{E}[Z|X] \quad \textit{a.s.}$$

Proof: Let $\varphi_n(X) = \mathbf{E}[Z_n|X]$ a.s.
Then by Fact 5.5, $\varphi_n(X) \uparrow \varphi(X)$ -a.s. Put $\varphi := \limsup_n \varphi_n$. Then by monotone convergence

$$\mathbf{E}[ZI_{\{X\in B\}}] = \lim_n \mathbf{E}[Z_nI_{\{X\in B\}}] = \lim_n \mathbf{E}[\varphi_n(X)I_{\{X\in B\}}] = \mathbf{E}[\varphi(X)I_{\{X\in B\}}].$$

□

Lemma 5.0.3 *(Projection property of conditional expectations)*
Let Y be X-measurable, i.e. $Y = g(X)$ for some y. Then

$$\mathbf{E}[\mathbf{E}[Z|X]|Y] = \mathbf{E}[Z|Y]$$

Proof: Since both sides are Y-measurable, it suffices to show (cf. Lemma 5.0.1)

$$\mathbf{E}[\mathbf{E}[Z|X] \cdot I_{\{Y \in B\}}] = \mathbf{E}[Z I_{\{Y \in B\}}]$$

This, however, is true since $\{Y \in B\} = \{X \in g^{-1}(B)\}$. □

Lemma 5.0.4 *(Taking out what is known) Assume $g(X)$ bounded, $\mathbf{E}[|Z|] < \infty$. Then $\mathbf{E}[g(X)Z|X] = g(X)\mathbf{E}[Z|X]$ a.s.*

Proof: Because of linearity and monotone convergence, it suffices to assume Z and g as non-negative and bounded. Since the r.h.s. is X-measurable, it suffices to show

$$\mathbf{E}[Zg(X)I_{\{X \in B\}}] = \mathbf{E}[\mathbf{E}[Z|X]g(X)I_{\{X \in B\}}].$$

This, however , is valid because of Proposition 5.0.2, □

Lemma 5.0.5 *("Integrating out independent stuff") Let X be an S-valued random variable, and Y be an S'-valued random variable independent of X. Also, let $h : S \times S' \to \mathbb{R}$ with $\mathbf{E}|h(X,Y)| < \infty$. Then*

$$\mathbf{E}[h(X,Y)|X] = \int h(X,y)\mu_Y(dy),$$

where μ_Y is the distribution of Y.

Proof: For $B \in \mathcal{S}$,

$$\begin{aligned}
\mathbf{E}[h(X,Y)1_B(X)] &= \int h(x,y)1_B(x)\mu_X \otimes \mu_Y(d(x,y)) \\
&= \int (\int h(x,y)\mu_Y(dy))1_B(x)\mu_X(dx) \\
&= \mathbf{E}[\int h(X,y)\mu_Y(dy)1_B(X)] \qquad \square
\end{aligned}$$

Recall: $g : \mathbb{R} \to \mathbb{R}$ is *convex*: $\iff$

$$\sum_i g(z_i)\mu(z_i) \geq g\left(\sum_i z_i\mu(z_i)\right)$$

$\forall z_1, \ldots, z_n \in \mathbb{R}$ and probability weights $\mu(z_1) \ldots, \mu(z_n)$.

Lemma 5.0.6 *If g is convex and $\mathbf{E}|Z| < \infty$, then*

$$\mathbf{E}[g(Z)] \geq g(\mathbf{E}[Z])$$

and, more generally,

$$\mathbf{E}[g(Z)|X] \geq g(\mathbf{E}[Z|X]) \qquad a.s.$$

Proof: We use the well-known fact that g is the countable supremum of straight lines (see e.g D. Williams, Probability of Martingales, Cambridge University Press 1991, p.61): There exist sequences (a_n) and (b_n) in $\mathbb{R}$ such that

$$g(z) = \sup_n (a_n z + b_n), \ z \in \mathbb{R}.$$

For all n, we have because of monotonicity and linearity of the conditional expectation:

$$\mathbf{E}[g(Z)|X] \geq \mathbf{E}[a_n Z + b_n|X] = a_n \mathbf{E}[Z|X] + b_n \quad \text{a.s}$$

and hence

$$\mathbf{E}[g(Z)|X] \geq \sup_n (a_n \mathbf{E}[Z|X] + b_n] = g(\mathbf{E}[Z|X]) \quad \text{a.s}$$

□

Remark 5.0.1 *Let $\mathcal{A}(X) := \{\{X \in B\}, B \in \mathcal{S}\}$ be the σ-field of events generated by X. Then one also writes $\mathbf{E}[Z|\mathcal{A}(X)]$ instead of $\mathbf{E}[Z|X]$.*

By the way, we could also consider, as our given information, a "σ-field of events" $\mathcal{F}$ instead of a random variable X.

Definition 5.0.4 *$\mathcal{F}$ is called a σ-field of events : $\Longleftrightarrow$*

*(i) $\mathcal{F}$ contains two events $\bigwedge$ and $\bigvee$ called **impossible** and **certain**.*

(ii) For all events $A_1, A_2, \ldots$ in $\mathcal{F}$, also the event

$$\bigcup A_n := \text{“}A_1 \text{ or } A_2 \text{ or } \ldots\text{” belongs to } \mathcal{F}$$

(iii) For each event A in $\mathcal{F}$, also the event $A^c :=$ " not A" belongs to $\mathcal{F}$.

Definition 5.0.5 *A_n S-valued random variable Y is $\mathcal{F}$-adapted $:\Longleftrightarrow$ all the events $\{Y \in C\}, C \in \mathcal{S}$, belong to $\mathcal{F}$.*

Definition 5.0.6 *An $\mathcal{F}$-adapted $\mathbb{R}$-valued random variable Y is called conditional expectation of Z given $\mathcal{F}$ $:\Longleftrightarrow$ for all events A in $\mathcal{F}$,*

$$\mathbf{E}[YI_A] = \mathbf{E}[ZI_A].$$

In this case, we write

$$Y = E[Z|\mathcal{F}] \quad \text{a.s .}$$

All results which we proved for $\mathbf{E}[Z|X]$ (existence, uniqueness, linearity, monotonicity, ...) can easily be carried over to this framework. The "way back" from Definition 5.0.6 to Definition 5.0.3 in case $\mathcal{F} = A(X)$ is provided by the following

Lemma 5.0.7 *An $\bar{\mathbb{R}}$-valued random variable Y is $\mathcal{A}(X)$-adapted iff there exists a measurable $\varphi : S \to \bar{\mathbb{R}}$ with $Y = \varphi(X)$*

Proof:
a) We first prove the assertion in case Y takes only finitely many values $y_1, \ldots y_k$. By assumption, there exist $B_1, \ldots B_k \in \mathcal{S}$ such that

$$\{Y = y_i\} = \{X \in B_i\} \quad , i = 1, \ldots, k$$

Since for $i \neq j$, the event

$$\{X \in B_i \cap B_j\} = \{X \in B_i\} \cap \{X \in B_j\} = \{Y = y_i\} \cap \{Y = y_j\}$$

is impossible, we can redefine the B_n such that they are pairwise disjoint.
Now put

$$\varphi(x) := \begin{cases} y_i & \text{for } x \in B_i \qquad , i = 1, \ldots, k \\ 0 & \text{otherwise} \end{cases}$$

We then have for all $i = 1, \dots, k$

$$\varphi(X) I_{\{Y=y_i\}} = \varphi(X) I_{\{X \in B_i\}} = y_i I_{\{X \in B_i\}} = y_i I_{\{Y=y_i\}}.$$

Summing over i, we arrrive at $\varphi(X) = Y$.
b) Now we turn to the general case. Without loss of generality we can assume $Y \geq 0$.
Let Y_n be random variables each taking finitely many values, such that $Y_n \uparrow Y$.
Let φ_n be such that $\varphi_n(X) = Y_n$, and put

$$C := \{x \in S : \lim_{n \to \infty} \varphi_n(x) \text{exists}\}$$

Since the event $\{X \in C\}$ is certain,

$$\varphi := \lim_{n \to \infty} \varphi_n 1_C$$

fulfills $\varphi(X) = Y$. □

Chapter 6

Martingales

6.1 Basic concepts

A martingale is a real-valued stochastic process whose conditional expectation at a future time point, given the overall information at present time, equals its present value.
We have to specify what we mean by the overall information at present time.

Definition 6.1.1 *a) A family $\mathbb{F} := (\mathcal{F}_n)_{n=0,1,\ldots}$ of σ-fields of events is called a **filtration** if it is increasing, i.e.*

$$\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \qquad , n = 0, 1, \ldots$$

b) A stochastic process $Z = (Z_n)_{n=0,1,\ldots}$ is called $\mathbb{F}$-adapted *if each Z_n is $\mathcal{F}_n$-adapted, $n = 0, 1, 2, \ldots$.*
(cf. Definition 5.0.5)

Remark 6.1.1 *Think of a stochastic process $X = (X_0, X_1, \ldots)$, where $X_{0\ldots n} := (X_0, \ldots, X_n)$ describes the states of the world (or at least all what you observe about them) up to time n. Then $\mathcal{F}_n := \mathcal{A}(X_{0\ldots n})$ defines a filtration $\mathbb{F}$ and (see Lemma 5.0.7) an $\mathbb{R}$-valued process $Z = (Z_n)$ is $\mathbb{F}$-adapted iff*

$$Z_n = g_n(X_{0\ldots n}) \quad , n = 0, 1, \ldots$$

for some measurable g_n.

For the rest of the chapter, let $(\mathcal{F}_n) = \mathbb{F}$ be a filtration.

Definition 6.1.2 *An $\mathbb{F}$-adapted sequence $Z = (Z_n)$ of integrable random variables is called an* $\mathbb{F}$-martingale *if*

$$\mathbf{E}[Z_{n+1}|\mathcal{F}_n] = Z_n \qquad a.s.$$

Z is called $\mathbb{F}$-supermartingale *if*

$$\mathbf{E}[Z_{n+1}|\mathcal{F}_n] \leq Z_n \qquad a.s. \;,$$

and submartingale if $(-Z_n)$ is a supermartingale.

Remark 6.1.2 *If Z is an $\mathbb{F}$-martingale, then*

$$\mathbf{E}[Z_{n+1}|(Z_0, \ldots, Z_n)] = Z_n \qquad a.s.$$

Indeed, Z_n is $\mathcal{A}(Z_0, \dots, Z_n)$-adapted, and each event $A \in \mathcal{A}(Z_0, \dots, Z_n)$ also belongs to $\mathcal{F}_n$. Hence, for all, $A \in \mathcal{A}(Z_0, \dots, Z_n)$,

$$\mathbf{E}[Z_{n+1} I_A] = \mathbf{E}[Z_n I_A].$$

Definition 6.1.3 *A sequence $(\xi_n)_{n \geq 1}$ of random variables is called $\mathbb{F}$-**previsible** : $\iff \xi_n$ is $\mathcal{F}_{n-1}$-adapted for all $n \geq 1$.*

Lemma 6.1.1 *let ξ be a real-valued, $\mathbb{F}$-previsible process, and*

$$G_n := \sum_{k=1}^{n} \xi_k (Z_k - Z_{k-1})$$

be integrable $(n = 1, 2, \dots)$.

a) If (Z_n) is a martingale, then also (G_n) is one.
b) If $(Z_n)$is a supermartingale and ξ_n is non-negative, $n \geq 1$, then also (G_n) is a supermartingale.

Proof:

$$\begin{aligned}
\mathbf{E}[G_{n+1}|\mathcal{F}_n] - G_n = \mathbf{E}[G_{n+1} - G_n|\mathcal{F}_n] &= \mathbf{E}[\xi_{n+1}(Z_{n+1} - Z_n)|\mathcal{F}_n] \\
&= \xi_{n+1}\mathbf{E}[Z_{n+1} - Z_n|\mathcal{F}] \begin{cases} = 0 & \text{a.s. in } a) \\ \geq 0 & \text{a.s. in } b) \end{cases}
\end{aligned}$$

(check which of the facts on conditional expectation we have used !)

6.2 The supermartingale convergence theorem

How often does a supermartingale (Z_n) transverse an interval $[a, b]$ from below to above? In any case, the tendency of (Z_n) is not to go upwards. The proof of the following estimate, which is due to Doob, relies on a simple idea: bet on the upcrossings, and estimate the gain from below in terms of the number of upcrossings. Since the gain process is a supermartingale (whose expectation is ≤ 0 since it starts in 0), this gives - under a mild additional assumption - a uniform upper bound for the expected number of upcrossings.
Let (Z_n) be a supermartingale, and fix $a < b \in \mathbb{R}$. Think of (Z_n) as the price of some asset. Trade one unit of the asset (by betting on increasing Z) as soon as Z has fallen below a, and do this as long as Z has risen above b:

$$\begin{aligned}
\xi_1 &:= I_{\{Z_0 < a\}} \\
\xi_n &:= I_{\{Z_{n-1} < a\} \cup \{Z_{n-1} \leq b, \xi_{n-1} = 1\}} \\
G_n &:= \sum_{k=1}^{n} \xi_k (Z_k - Z_{k-1})
\end{aligned}$$

Let U_n denote the number of upcrossings of $[a, b]$ till time n. Obviously, with $x^- := -\min(x, 0)$,

$$\begin{aligned}
G_n &\geq (b - a)U_n - (Z_n - a)^- \\
&\geq (b - a)U_n - (Z_n - |a|)^- \\
&\geq (b - a)U_n - (Z_n^- + |a|),
\end{aligned}$$

that is,

$$(b - a)U_n \leq G_n + |a| + Z_n^-.$$

Since $\mathbf{E}G_n \leq 0$ (see Lemma 6.1.1) we have

Lemma 6.2.1 *(Doob's upcrossing inequality)*

$$(b-a)\mathbf{E}U_n \leq |a| + \mathbf{E}Z_n^- \leq |a| + \sup_k \mathbf{E}Z_k^-$$

If, moreover, $\sup_k \mathbf{E}Z_k^- < \infty$*, then monotone convergence implies*

$$\mathbf{E}U_\infty < \infty, \ \textit{where } U_\infty := \lim_n U_n$$

Theorem 6.2.1 *(Supermartingale convergence theorem)*
Let (Z_n) *be a supermartingale with* $\sup_n \mathbf{E}Z_n^- < \infty$*. Then* (Z_n) *converges a.s. to an integrable random variable* Z_∞*.*

Proof: For all $a < b \in \mathbb{R}$, Lemma 6.2.1 yields

$$\mathbf{P}[\liminf Z_n < a, \limsup Z_n > b] \leq \mathbf{P}[U_\infty = \infty] = 0.$$

Hence

$$\begin{aligned}
\mathbf{P}[\liminf Z_n < \limsup Z_n] &= \mathbf{P}[\bigcup_{\substack{a<b \\ a,b,\in\mathbb{Q}}} \{\liminf Z_n < a, \limsup Z_n > b\}] \\
&\leq \sum_{\substack{a<b \\ a,b\in\mathbb{Q}}} \mathbf{P}[\liminf Z_n < a, \limsup Z_n > b] = 0.
\end{aligned}$$

This implies

$$Z_n \to Z_\infty := \limsup X_n \qquad \text{a.s.}$$

Finally, Fatou's lemma (see Lemma 6.2.2 below) yields

$$\begin{aligned}
\mathbf{E}|Z_\infty| &= \mathbf{E}[\liminf |Z_n|] \leq \liminf \mathbf{E}[|Z_n|] \\
&= \liminf_n \mathbf{E}[Z_n + 2Z_n^-] \\
&\leq \mathbf{E}Z_0 + 2\sup_n \mathbf{E}Z_n^- < \infty. \qquad \Box
\end{aligned}$$

We have to append

Lemma 6.2.2 *(Fatou's lemma)*
Let Y_n *be non-negative random variables. Then*

$$\mathbf{E}[\liminf_n Y_n] \leq \liminf_n \mathbf{E}[Y_n]$$

Proof: Since $\liminf_n Y_n = \lim_n \inf_{m\geq n} Y_n$, we have by monotone convergence

$$\begin{aligned}
\mathbf{E}[\liminf_n Y_n] &= \lim_n \mathbf{E}[\inf_{m\geq n} Y_m] \\
&= \liminf_n \mathbf{E}[\inf_{m\geq n} Y_m] \leq \liminf_n \mathbf{E}[Y_n]
\end{aligned}$$

$\Box$

6.3 Doob's submartingale inequalities

Let (Z_n) be a non-negative submartingale. (As a prominent example, think of $Z_n := |M_n|$, for a martingale (M_n). Indeed, by Jensen's inequality

$$\mathbf{E}[|M_{n+1}| \quad |\mathcal{F}_n] \ \geq \ |\mathbf{E}[M_{n+1}|\mathcal{F}_n]| = |M_n| \qquad \text{a.s)}$$

Put

$$Z_n^* := \max_{0\le k\le n} Z_k$$

(the "'current maximum" of the path up to time n).
Since Z_n has an upward tendency, there is some hope for a "stochastic estimate" of Z_n^* by Z_n.
Because of

$$cI_{\{Z_n^*\ge c\}} \le Z_n^* I_{\{Z_n^*\ge c\}},$$

we have

$$c\mathbf{P}[Z_n^* \ \ge \ c] \ \le \ \mathbf{E}[Z_n^* I_{\{Z_n^*\ge c\}}] \tag{6.1}$$

It turns out that in the r.h.s. one can replace Z_n^* by Z_n. This is

Proposition 6.3.1 *(Doob's first submartingale inequality)*
For $c > 0$,

$$c\mathbf{P}[Z_n^* \ge c] \le \mathbf{E}[Z_n I_{\{Z_n^*\ge c\}}]. \tag{6.2}$$

Proof: Put $F_k := \{Z$ exceeds the level c for the first time at time $k\}$
In other words,

$$\begin{aligned} F_0 &= \{Z_0 \ge c\} \\ F_k &= \{Z_{k-1}^* < c, Z_k \ge c\} \quad , k = 1, \dots, n \end{aligned}$$

Because of the submartingale property we have

$$\mathbf{E}[Z_n I_{F_k}] \ge \mathbf{E}[Z_k I_{F_k}] \ge c\mathbf{P}[F_k].$$

Since $I_{\{Z_n^*\ge c\}} = \sum\limits_{k=0}^{n} I_{F_k}$, the claim follows by summation. □

The assertion (6.2) can be rephrased as follows:

$$\text{For all } c \ge 0, \quad \mathbf{E}[Z_n | \{Z_n^* \ge c\}] \ge c.$$

It turns out that this provides an estimate of the 2nd moment of Z_n in terms of that of Z_n^*.

Lemma 6.3.1 *Let X and Y be non-negative random variables with*

$$c\mathbf{P}[X \ge c] \le \mathbf{E}[Y I_{\{X\ge c\}}], c > 0. \tag{6.3}$$

Then

$$E[X^2] \le 4\mathbf{E}[Y^2]. \tag{6.4}$$

Proof: Without loss of generality, $0 < \mathbf{E}X$, and $\mathbf{E}Y < \infty$. First we observe:

$$\mathbf{P}[X = \infty] = \lim_{c\to\infty} \mathbf{P}[X \ge c] \le \lim_{c\to\infty} \frac{1}{c}\mathbf{E}[Y] = 0.$$

Next we state a useful formula for the 2nd moment:

$$\mathbf{E}X^2 = \int_0^\infty 2c\mathbf{P}[X \ge c]dc \tag{6.5}$$

Indeed, writing μ_X for the distribution of X and using Fubini's lemma) we have

$$\begin{aligned}\int_0^\infty 2c\mu_X([c,\infty))dc &= \int_0^\infty\int_0^\infty 2c1_{\{x\geq c\}}\mu_X(dx)dc \\ &= \int_0^\infty\int_0^\infty 2c1_{\{x\geq c\}}dc\,\mu_X(dx) \\ &= \int_0^\infty x^2\mu_X(dx) = \mathbf{E}X^2 \quad .\end{aligned}$$

Writing $\mu_{(X,Y)}$ for the joint distribution of X and Y , and using successively (6.5), (6.3), once again Fubini, and the Cauchy-Schwarz inequality, we arrive at

$$\begin{aligned}\mathbf{E}X^2 &= \int_0^\infty 2c\mathbf{P}[X\geq c] \leq \int_0^\infty 2\mathbf{E}[YI_{\{X\geq c\}}]\; dc \\ &= \int_0^\infty\int_{\mathbb{R}_+^2} 2\; 1_{\{X\geq c\}}y \quad \mu_{(X,Y)}(d(x,y))\; dc \\ &= \int_{\mathbb{R}_+^2} 2xy\,\mu_{(X,Y)}(d(x,y)) = 2\mathbf{E}[XY] \leq 2\sqrt{\mathbf{E}X^2}\sqrt{\mathbf{E}Y^2}\end{aligned}$$

Dividing by $\sqrt{\mathbf{E}X^2}$ and squaring yields this assertion. □
It is now easy to prove

Theorem 6.3.1 *(Doob's L^2-inequality):*
If (Z_n) is a non-negative, L^2-bounded submartingale, then

$$\mathbf{E}[(\sup_k Z_k)^2] \leq 4\sup_k \mathbf{E}Z_k^2 \tag{6.6}$$

Moreover, (Z_k) converges not only a.s. but also in L^2.

Proof: Doob's first submartingale inequality together with Lemma 6.4 implies (with $Z_n^* := \sup_{0\leq k\leq n} Z_n$)

$$\mathbf{E}(Z_n^*)^2 \leq 4\mathbf{E}Z_n^2 \leq 4\sup_{k>0}\mathbf{E}Z_k^2$$

Since $Z_n^* \uparrow Z^* := \sup_k Z_k$ $(= \sup_k Z_k^*)$, (6.6) follows by monotne convergence. Because of $\mathbf{E}|Z_n| \leq \mathbf{E}Z_n^2$, $(-Z_n)$ is an L^1-bounded supermartingale. Hence the martingale convergence theorem tells us that Z_n converges a.s. to a randam variable Z_∞. Because of

$$|Z_n - Z_\infty| \leq 2Z^* \quad \text{a.s.} \ ,$$

and because $\mathbf{E}(Z^*)^2 = \sup \mathbf{E}(Z_n^*)^2 < \infty$, the L^2-convergence of Z_n to Z_∞ follows by dominated convergence. □
We have to append

Lemma 6.3.2 *(Lebesgue's dominated convergence theorem)*
If $X_n \to X$ in probability, and $|X_n| \leq Y$ for some integrable Y, then $\mathbf{E}|X_n - X| \to 0$ (and a fortiori $\mathbf{E}X_n \to \mathbf{E}X$).

This is a consequence of the observation

$$\begin{aligned}\lim_{c\to\infty}\sup_n \mathbf{E}[|X_n| I_{\{|X_n|\geq c\}}] &\leq \lim_{c\to\infty}\mathbf{E}[Y I_{\{Y\geq c\}}] \\ &= \lim_{c\to\infty}\mathbf{E}[Y] - \lim_{c\to\infty}\mathbf{E}[Y I_{\{Y<c\}}] = 0\end{aligned}$$

(monotone convergence!) and the stronger

Lemma 6.3.3 *If $X_n \to X$ in probability, and*

$$\lim_{c\to\infty}\sup_n \mathbf{E}[|X_n| \quad I_{\{|X_n|>c\}}] = 0, \tag{6.7}$$

then X is integrable, and $\mathbf{E}|X_n - X| \to 0$.

Proof: Let us write $\mathbf{E}[Z; A] := \mathbf{E}[Z I_A]$.

1) First we claim that (6.7) implies

$$\sup_n \mathbf{E}|X_n| < \infty.$$

Indeed, choose c so large that

$$\sup_n \mathbf{E}[|X_n|; \{|X_n| > c] \leq 1.$$

Then

$$\mathbf{E}|X_n| = \mathbf{E}[|X_n|; \{|X_n| \leq c\}] + \mathbf{E}[|X_n|; \{|X_n| > c\}] \leq c + 1.$$

2) Convergence in probability implies convergence of a suitable subsequence X_{n_k}. Hence by Fatou

$$\mathbf{E}|X| = \mathbf{E}\liminf |X_{n_k}| \leq \liminf \mathbf{E}|X_{n_k}| < \infty.$$

3) For given ε let c be so large that

$$\mathbf{E}[|X_n|; \{|X_n| > c\}] < \varepsilon, \quad n \in \mathbb{N}.$$

and

$$\mathbf{E}[|X|; \{|X| < c\}] > \varepsilon.$$

Then

$$\begin{aligned}\mathbf{E}|X_n - X| &\leq \mathbf{E}[|X_n - X|; |X_n - X| \leq \varepsilon] \\ &+ \mathbf{E}[|X_n|; |X_n - X| \geq \varepsilon; |X_n| > c] \\ &+ \mathbf{E}[|X|; |X_n - X| \geq \varepsilon; |X| > c] \\ &+ \mathbf{E}[|X_n|; |X_n - X| \geq \varepsilon; |X_n| \leq c] \\ &+ \mathbf{E}[|X_n|; |X_n - X| \geq \varepsilon; |X| \leq c] \\ &\leq 3\varepsilon + 2c\mathbf{P}[|X_n - X| \geq \varepsilon] \\ &\to 3\varepsilon \text{ as } n \to \infty.\end{aligned}$$

Since ε was arbitrary, $\mathbf{E}|X_n - X| \to 0$. □

Property (6.7) is important enough to be given a name.

Definition 6.3.1 *A family $(X_i)_{i\in I}$ of $\bar{\mathbb{R}}$-valued random variables is called* **uniformly integrable** *if*

$$\lim_{c\to\infty}\sup_{i\in I} \mathbf{E}[|X_i \mid I_{\{|X_i|>c\}}] = 0.$$

Remark 6.3.1 *The first step in the proof of Lemma 6.3.3 shows that uniform integrability implies boundedness in* L^1.

Lemma 6.3.4 rephrased: *Convergence in probability and uniform integrability imply* L^1*- convergence.*

(In fact, also the converse is true, see D. Williams, loc.cit, Theorem 13.7)

6.4 Stopping times

Let $\mathbb{F} = (\mathcal{F}_n)$ be a filtration, and $\mathcal{F}_\infty$ be the smallest σ-field of events containing all the $\mathcal{F}_n$.

Definition 6.4.1 *An* $\mathbb{N}_0 \cup \{\infty\}$*-valued random variable* T *is called an* $\mathbb{F}$*-stopping time:* $\Longleftrightarrow$

$$\{T \leq n\} \in \mathcal{F}_n \quad , n \in \mathbb{N}_0. \tag{6.8}$$

Remark 6.4.1 *a) (6.8)* $\Longleftrightarrow \{T = n\} \in \mathcal{F}_n \quad \forall n \in \mathbb{N}_0$, *since*

$$\{T \leq n\} = \bigcup_{k=0}^{n} \{T = k\}, \text{ and } \{T = n\} = \{T \leq n\} \cap \{T \leq n-1\}^c.$$

b) Every constant in $\mathbb{N}_0 \cup \{\infty\}$ *is a stopping time.*
c) Together with T, T' *, also* $\max\{T, T'\}$ *and* $\min\{T, T'\}$ *are stopping times (check!)*

In the sequel let T be an $\mathbb{F}$-stopping time.

Definition 6.4.2

$$\mathcal{F}_T := \{A \in \mathcal{F}_\infty : A \cap \{T \leq n\} \in \mathcal{F}_n \quad \forall n\}$$

is called the **(σ-field of) T-past**.

Definition 6.4.3 *For* $\mathbb{F}$*-adapted* (X_n) *and* $\mathcal{F}_\infty$*-adapted* X_∞ *we define*

$$X_T := \sum_{k \in \mathbb{N}_0 \cup \{\infty\}} X_k I_{\{T=k\}}.$$

Remark 6.4.2 X_T *is* $\mathcal{F}_T$*-adapted, since*

$$\begin{aligned} \{X_T \in B\} \cap \{T \leq n\} &= \bigcup_{k=0}^{n} (\{X_T \in B\} \cap \{T = k\}) \\ &= \bigcup_{k=0}^{n} (\underbrace{\{X_k \in B\} \cap \{T = k\}}_{\in \mathcal{F}_k}) \in \mathcal{F}_k. \end{aligned}$$

(In particular putting $X_n := n$*, we see that* T *is* $\mathcal{F}_T$*-adapted.)*

6.5 Stopped supermartingales

Let $\mathcal{F}$ be a filtration, (X_n) be an $\mathbb{F}$-supermartingale, and T be an $\mathbb{F}$-stopping time.

Proposition 6.5.1 $(X_{T \wedge n})$ *is an* $\mathbb{F}$*-supermartingale as well.*

Proof:

$$\begin{aligned} X_{T\wedge n} &= X_0 + \sum_{j=1}^{T\wedge n}(X_j - X_{j-1}) \\ &= X_0 + \sum_{j=1}^{n} I_{\{T\ge j\}}(X_j - X_{j-1}) \end{aligned}$$

Since $\{T \ge n\} = \{T \le n-1\}^c \in \mathcal{F}_{n-1}$, $\xi_n := I_{\{T\ge n\}}$ is a previsible process and the claim follows from Lemma 6.1.1 b). □

Proposition 6.5.2 *(Stopping theorem, baby version)*
Let S, T be $\mathbb{F}$-stopping times with $S \le T \le n$ for some $n \in \mathbb{N}$. Then

$$X_S \ge \mathbf{E}[X_T|\mathcal{F}_S] \quad a.s.$$

(and in particular $\mathbf{E}X_S \ge \mathbf{E}X_T$).

Proof: X_S is $\mathcal{F}_S$-adapted. Hence it suffices to show

$$\mathbf{E}[X_T; G] \le \mathbf{E}[X_S; G],\ G \in \mathcal{F}_S. \tag{6.9}$$

Put $G_k := G \cap \{S = k\}, k = 0, \ldots, n$. Recalling that $G_k \in \mathcal{F}_k$ and using Proposition 6.5.1, we infer

$$\begin{aligned} \mathbf{E}[X_T; G_k] &= [X_{T\wedge n}; G_k] \le \mathbf{E}[X_{T\wedge k}; G_k] \\ &= \mathbf{E}[X_{T\wedge S}; G_k] = \mathbf{E}[X_S; G_k]. \end{aligned}$$

Summation over k yields (6.9). □

Theorem 6.5.1 *(Stopping theorem, adult version) Let (X_n) be a uniformly integrable supermartingale, S and T be stopping times with $S \le T$. Then*

$$\mathbf{E}[X_S] \ge \mathbf{E}[X_T] \tag{6.10}$$

(where we put $X_\infty := \limsup X_n$).

Proof:
a) $(X_{T\wedge n})$ is a supermartingale fulfilling the requirement

$$\sup_n \mathbf{E}X^-_{n\wedge T} < \infty \tag{6.11}$$

in the supermartingale convergence theorem 6.2.1. Indeed, since $\varphi(x) := -x^-$ is concave and increasing, $(-X^-_{n\wedge T})_{n\ge 0}$ is again a supermartingale (check !)
Hence, because of the baby version of the stopping theorem,

$$\mathbf{E}[-X^-_{n\wedge T}] \ge \mathbf{E}[-X^-_n] \ge -\mathbf{E}|X_n|.$$

Together with Remark 6.3.1 this implies (6.11)
b) Because of a) and the supermartingale convergence theorem 6.2.1, $\lim_{n\to\infty} X_{n\wedge T}$ exists a.s. and is integrable. On $\{T < \infty\}$,

$$\lim_{n\to\infty} X_{n\wedge T} = X_T,$$

and on $\{T < \infty\}$,

$$\lim_{n\to\infty} X_{n\wedge T} = \limsup_n = X_\infty = X_T \text{ a.s.}$$

Hence $\lim_{n\to\infty} X_{n\wedge T} = X_T$ a.s., and X_T is integrable.
c) Replacing T by S in a) and b) we see that $\lim_{n\to\infty} X_{n\wedge S} = X_S$ a.s., and X_T is integrable.
d) We know from Proposition 6.5.1 and the baby version of the stopping theorem that

$$\mathbf{E}[X_{S\wedge n}] \geq \mathbf{E}[X_{T\wedge n}],\ n = 0, 1, \ldots \tag{6.12}$$

It remains to check that $(X_{S\wedge n})$ and $(X_{T\wedge n})$ are uniformly integrable (the assertion (6.10) then follows from (6.12) together with Lemma 6.6). Indeed,

$$\begin{aligned}
& \mathbf{E}[|X_{n\wedge T}|; |X_{n\wedge T}| > c] \\
= \ & \mathbf{E}[|X_T|; |X_T| > c; T \leq n] + \mathbf{E}[|X_n|; |X_n| > c; T > n] \\
\leq \ & \mathbf{E}[|X_T|; |X_T| > c] + \mathbf{E}[|X_n|; |X_n| > c] \longrightarrow 0 \text{ as } c \to 0,
\end{aligned}$$

since X_T is integrable by part a), and (X_n) is uniformly integrable by assumption. □

Example: How long does it take till in a fair coin tossing game the pattern $\mathcal{THTH}$ occurs for the first time ?
Consider the following fair game: Before the first toss, a gambler enters the casino and bets 1 Euro on tail. If she loses, she goes home, with a loss of 1 Euro. If she wins, she bets two Euro on head. If she loses in the second toss, she goes home, with a total loss of 1 Euro. If she wins in the second toss, she bets 4 Euro on tail. If she then loses in the third toss, she goes home with a total loss of 1 Euro. If she wins in the 3rd toss, she bets 8 Euro on head. If she loses in the 4th toss, she goes home with a total loss of one Euro. If she wins in the 4th toss, the game is stopped, and she goes home gaining 15 Euro.
Now imagine that before any new loss, a new gambler enters the casino, following exactly the same strategy (ie. starting to bet one Euro on tail, and playing at most 4 rounds). The game is stopped when the pattern $\mathcal{THTH}$ occurs for the first time, i.e. at the first time when one of the gamblers wins 15 Euro.
Denote by X_n the total gain of all the gamblers (having entered so far) at time n. Obviously (X_n) is a martingale, and $|X_n| \leq$ const $\cdot\, n$. Hence

$$|X_{n\wedge T}| \leq \text{ const } \cdot (n \wedge T) \leq \text{ const } \cdot T.$$

Since $\mathbf{P}[T \geq m] \leq K^{-cm}$ for some K, C, we have $\mathbf{E}T < \infty$, and consequently $(X_{n\wedge T})$ is uniformly integrable. Since by Proposition 6.5.1 $X_{n\wedge T}$ is a martingale, we obtain from the stopping theorem

$$\mathbf{E}X_T = \mathbf{E}X_0 = 0$$

However,

$$X_T = 15 - 1 + 3 - 1 - (T - 4) = 20 - T.$$

Hence

$$\mathbf{E}T = 20.$$

Finally, let us consider the pattern $\mathcal{TTHH}$ (instead of $\mathcal{THTH}$). Then

$$X_t = 15 - 1 - 1 - 1 - (T - 4) = 16 - T,$$

hence

$$\mathbf{E}T = 16$$

Thus, although for each of the two patterns and each fixed time point n, the probability that the pattern starts at n is 2^{-4}, the expected waiting time for the pattern $\mathcal{THTH}$ is larger than that for $\mathcal{TTHH}$. An intuitive explanation for this ist that the pattern $\mathcal{THTH}$ tends to come in clumps like $(\mathcal{THTHTH})$, thus, by poetic justice, the expected waiting times between clumps should be longer.

Chapter 7

The Wiener Process

7.1 Heuristics and basics

How to scale an ordinary random walk to get a "diffusion limit" ? Consider the increments of an ordinary random walk:

$$Y_k = \begin{cases} +1 & \text{with prob } \frac{1}{2} \\ -1 & \text{with prob } \frac{1}{2} \end{cases}$$

Now consider n steps, and take each increment of size $\frac{1}{\sqrt{n}}$. The central limit theorem tells us that

$$\mathbf{P}[\sum_{k=1}^{n} \frac{1}{\sqrt{n}} Y_k \in [a,b]] \longrightarrow \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

Next, define

$$S_t^{(n)} := \sum_{k=1}^{[nt]} \frac{1}{\sqrt{n}} Y_k = \sqrt{t} \underbrace{\sum_{k=1}^{[nt]} \frac{1}{\sqrt{nt}} Y_k}_{\longrightarrow \mathcal{N}(0,1) \text{in distribution}}$$

Hence

$$S_t^{(n)} \to \mathcal{N}(0,t) \text{ in distribution}$$

A candidate for a limit in distribution of S^n on the *space of paths* would be a $C(\mathbb{R}_+, \mathbb{R})$-valued random variable W with the properties

(i) $W(t+h) - W(t)$ $\mathcal{N}(0,h)$-distributed

(ii) $W(t_1 - W(t_0), \ldots, W(t_k) - W(t_{k-1})$ independent for $t_0 \le t_1 \le \ldots \le t_k$

(iii) $W(0) = 0$ a.s.

A continuous random path W with these properties is called a **standard Wiener** process.
Since joint normal (or Gaussian) distributions on $\mathbb{R}^d$ are determined by their mean vector and covariance matrix, the following is equivalent to (i) - (iii):
$(W(t_1), \ldots, (W(t_n))$ has a joint normal distribution with mean zero and covariance

$$\mathbf{E} W(t_i) W(t_j) = \min(t_i, t_j)$$

7.2 Lévy's construction of W

Basic observation: If W is a standard Wiener process, then
$Y := Y_{t_1,t_2} := W(\frac{t_1+t_2}{2}) - \frac{1}{2}(W(t_1) + W(t_2))$ is $\mathcal{N}(0, \frac{1}{4}(t_2 - t_2))$ distributed and independent of $W(t_1)$ and $W(t_2)$. Indeed,

$$\begin{aligned}
\mathbf{E}[Y \cdot W(t_1)] &= t_1 - \frac{1}{2}t_1 - \frac{1}{2}t_1 = 0, \\
\mathbf{E}[Y \cdot W(t_2)] &= \frac{t_1+t_2}{2} - \frac{1}{2}t_1 - \frac{1}{2}t_2 = 0, \\
\mathbf{E}[Y^2] &= \frac{1}{4}t_1 + \frac{1}{4}t_2 - t_1 + \frac{1}{2}t_1 \\
&= \frac{1}{4}(t_2 - t_1).
\end{aligned}$$

Successive construction of $W(\frac{k}{2^n}), 0 < k < 2^n$, k odd (inductive over n): Let $W(1)$ and $Z_{2^n,k}$ be independent standard normal random variables. Put

$$W(\frac{1}{2}) := \frac{1}{2}W(1) + \frac{1}{2}Z_{2^0,1}$$

This defines $W(\frac{0}{2^1}), W(\frac{0}{2^1}), W(\frac{0}{2^1})$. Proceed inductively by

$$W(\frac{k}{2^n}) := \frac{1}{2}(W(\frac{k-1}{2^n}) + W(\frac{k+1}{2^n})) + \frac{1}{2^{\frac{n+1}{2}}}Z_{2^n,k}.$$

Put $W_n :=$ linear interpolation of the $W(\frac{k}{2^n}), 0 < k < 2^n$, k odd. This defines a sequence of $C([0,1],\mathbb{R})$-valued random variables.
Let us now estimate the distance between W_{n-1} and W_n:

$$\sup_{0 \le t \le 1} \mid W_n(t) - W_{n-1}(t) \mid \le 2^{-\frac{n+1}{2}} \max\{|Z_{2^n,k}| : 0 < k < 2^n\}$$

Since $\mathbf{P}[|Z_{2^n,k}| > n] \le 2\frac{1}{\sqrt{2\pi}} \int\limits_n^\infty xe^{-\frac{x^2}{2}}dx = \sqrt{\frac{2}{\pi}}e^{-\frac{n^2}{2}}$, we conclude that

$$\sum_{n=1}^{\infty}\sum_{k=1}^{2^n} \mathbf{P}[|Z_{2^n,k}| \ge n] \le \sum_{n=1}^{\infty} 2^n e^{-\frac{n^2}{2}} < \infty,$$

and hence $\sum\limits_{n=1}^{\infty} \mathbf{P}[\exists k : |Z_{2^n,k}| \ge n] < \infty$.
Using Borel-Cantelli, we get

$$\mathbf{P}[\exists n_0 \quad \forall n \ge n_0 : \sup_{0 \le t \le 1} \mid W_n(t) - W_{n-1}(t) \mid < 2^{-\frac{n+1}{2}}n\} = 1.$$

Therefore: a.s. , (W_n) is a Cauchy sequence w.r.to uniform convergence.
Put $W :=$ a.s. limit of W (w.r. to uniform convergence in $C[0,1],\mathbb{R})$.
Claim: For all $k \in \mathbb{N}$ and $t_1 < \ldots < t_k$, $(W(t_1), \ldots, W(t_k))$ is jointly normal with expectation 0 and covariances $\mathbf{E}[W(t_i)W(t_j)] = \min(t_i, t_j)$.
Indeed: approximate t_i by dyadic rationals $t_{n,i}$. Then, because of the a.s. uniform convergence and the continuity of W_n,

$$(W_n(t_{n,1}), \ldots .W_n(t_{n,k})) \longrightarrow (W(t_1), \ldots, W(t_k)) \text{ a.s.}$$

The claim about the joint distribution of $(W(t_1), \ldots W(t_k))$ then follows e.g. by using characteristic functions (O.Kallenberg, Foundations of modern probability, Springer 97, Thm.4.4).

7.3 Quadratic variation of Wiener paths

Proposition 7.3.1 *Let W be a standard Wiener process. Then*

$$Q_n := \sum_{\frac{k}{n}\le t} \left(W(\frac{k}{n}) - W(\frac{k-1}{n}) \right)^2 \longrightarrow t \ \textit{ in probability} .$$

Proof:

$$\mathbf{E}Q_n = \sum \left(\frac{k}{n} - \frac{k-1}{n} \right) = \max \left\{ \frac{k}{n} : \frac{k}{n} \le t \right\} \longrightarrow t.$$

Since the $W(\frac{k}{n}) - W(\frac{k-1}{n})$, $k = 1, 2, \ldots$ are independent and distributed as $\frac{1}{\sqrt{n}}Z$, where Z is a standard normal random variable, we have

$$\begin{aligned} \operatorname{Var} Q_n &= [nt] \cdot \operatorname{Var} \left(W(\frac{1}{n}) \right)^2 \\ &= [nt] \cdot \operatorname{Var}(\frac{1}{\sqrt{n}} Z)^2 \\ &= [nt] \, \frac{1}{n^2} \, \operatorname{Var} Z^2 \overset{\mathrm{n}\to\infty}{\longrightarrow} 0. \end{aligned}$$

The assertion now follows by Tschebyshev's inequality. □

Corollary 7.3.1 *W has a.s. "infinite total variation" on $[0,t]$, since*

$$\sum_{\frac{k}{n}\le t} |W(\frac{k}{n}) - W(\frac{k-1}{n}) \,| \ge \frac{1}{\sup\limits_k |W(\frac{k}{n}) - W(\frac{k-1}{n})|} \quad \sum \left(W(\frac{k}{n}) - W(\frac{k-1}{n}) \right)^2 .$$

The first factor on the r.h.s. converges to ∞ by (uniform) continuity of W on $[0,t]$, the second factor converges to t along a subsequence (n').

The corollary indicates that it won't be possible to define an integral $\int_0^t \xi_s dW_s$ naively à la Riemann-Stieltjes. Indeed, the following classical example of Itō (1942) illustrates this with $\int W_s dW_s$.

Example 7.3.1

$$\begin{aligned} I_1 &:= \sum_{\frac{k}{n}\le t} W(\frac{k-1}{n}) \left(W(\frac{k}{n}) - W(\frac{k-1}{n}) \right) \\ I_2 &:= \sum_{\frac{k}{n}\le t} W(\frac{k}{n}) \left(W(\frac{k}{n}) - W(\frac{k-1}{n}) \right) \end{aligned}$$

We then have

$$\begin{aligned} I_2 + I_1 &= W^2_{\frac{[nt]}{n}} - W_0^2 \ \longrightarrow W_t^2 \\ I_2 - I_1 &= \sum_{\frac{k}{n}\le t} \left(W(\frac{k}{n}) - W(\frac{k-1}{n}) \right)^2 \longrightarrow t \quad \text{in probability} \end{aligned}$$

and consequently

$$\begin{aligned} I_1 &\overset{\mathrm{n}\to\infty}{\longrightarrow} \frac{1}{2}W_t^2 - \frac{t}{2} \quad \text{in probability,} \\ I_2 &\overset{\mathrm{n}\to\infty}{\longrightarrow} \frac{1}{2}W_t^2 + \frac{t}{2} \quad \text{in probability.} \end{aligned}$$

7.4 Intermezzo: Filtrations and stopping in continuous time

Definition 7.4.1 *An increasing family $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ of σ-fields is called a* **filtration** *(in continuous time).*

We put $\mathcal{F}_{t+} := \bigcap_{s>t} \mathcal{F}_t, \quad \mathbb{F}_+ := (\mathcal{F}_{t+})_{t\geq 0}$,
$\mathcal{F}_\infty :=$ the smallest σ-field containing all the $\mathcal{F}_t$.
An $\bar{\mathbb{R}}$-valued random variable is called an $\mathbb{F}$-**stopping time**: $\Longleftrightarrow$

$$\Longleftrightarrow \{\tau < t\} \in \mathcal{F}_t, \quad t > 0.$$

Lemma 7.4.1 *τ is an $\mathbb{F}$-stopping time $\Longleftrightarrow \{\tau \leq t\} \in \mathcal{F}_{t+} \quad \forall t$.*

Proof: "$\Longrightarrow$" for all $s > t$,

$$\{\tau \leq t\} = \bigcap_{n: t+\frac{1}{n}<s} \{\tau < t + \frac{1}{n}\} \in \mathcal{F}_s$$

Hence $\{\tau \leq t\} \in \mathcal{F}_{t+}$.
"$\Longleftarrow$" $\{\tau < t\} = \bigcup_n \{\tau \leq t - \frac{1}{n}\} \in \mathcal{F}_t$. □

Definition 7.4.2 *For a stopping time τ, we put*

$$\mathcal{F}_{\tau+} := \{A \in \mathcal{F}_\infty : A \cap \{\tau < t\} \in \mathcal{F}_t, \quad t > 0\}$$

and call it the σ-field of pre-τ events.

Typical example: Let X be a process with continuous paths in $\mathbb{R}^d$, $\mathcal{F}_t$ be the σ-field generated by $(X_s)_{0\leq s\leq t}$, B, C some (open or closed) sets in $\mathbb{R}^d$. Consider the stopping time

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

(the first hitting time of B).
Then the event $A := \{X \text{ hits } C \text{ before it hits } B\}$ belongs to $\mathcal{F}_{\tau+}$.

Remark 7.4.1 *For $\tau =$ a constant time s, the definition of $\mathcal{F}_{\tau+}$ is consistent with that of $\mathcal{F}_{s+}$ (check!)*

Definition 7.4.3 *a) A process X is callled $\mathbb{F}$-adapted if X_t is $\mathcal{F}_t-$ adapted $\forall t$. We say that X is continuous if it has continuous paths.*
*b) An $\mathbb{F}$-adapted process X with $\mathbf{E}|X_t| < \infty$ is called an $\mathbb{F}$-**martingale** if*

$$\mathbf{E}[X_{t+h}|\mathcal{F}_t] = X_t \ a.s. \ , \quad t, h \geq 0.$$

c) A Wiener process W is called an $\mathbb{F}$-Wiener process if it is $\mathbb{F}$-adapted and the increments $W_{t+h} - W_t$ are independent of $\mathcal{F}_t$ for all $t \geq 0$. (In particular, W then is an $\mathbb{F}$-martingale.)

Proposition 7.4.1 *(Strong Markov property of the Wiener process, Ka. 11.11)*[1] *Let W be an $\mathbb{F}$-Wiener process and τ be an a.s. finite $\mathbb{F}$-stopping time. Then $(W_{\tau+t} - W_\tau)_{t\geq 0}$ is again a standard Wiener process, independent of $\mathcal{F}_{\tau+}$.*

Corollary 7.4.1 *An $\mathbb{F}$-Wiener process is also an $\mathbb{F}_+$-Wiener process.*

Henceforth, we will always assume that our filtration $\mathbb{F}$ is right continuous (i.e. obeys $\mathbb{F} = \mathbb{F}_+$).

[1]Here and below, the citation Ka XX.YY will refer to the book O. Kallenberg, Foundations of Modern Probability, 2nd ed, Springer 2002.

7.5 The Itō-integral for simple integrands

Definition 7.5.1 *A random path $H = (H_s)$ is called a* **simple integrand** *: $\Longleftrightarrow$*

$$H_s = \sum_{k=0}^{m-1} \xi_k 1_{(t_k, t_{k+1}]}(s)$$

for some $0 =: t_0 < t_1 < \ldots < t_m$ and $\mathcal{F}_{t_k}$-adapted ξ_k.

Definition 7.5.2

$$\int_0^t H_s dW_s := \sum_k \xi_k (W_{t \wedge t_{k+1}} - W_{t \wedge t_k}).$$

Observation: $G_t := \int_0^t H_s dW_s =: (H \bullet W)_t$ is a continuous $\mathbb{F}$-martingale, and

$$\mathbf{E}[G_t^2] = \mathbf{E}[\sum_k \xi_k^2 (t \wedge t_{k+1} - t \wedge t_k)] = \mathbf{E}[\int_0^t H_s^2 ds].$$

This so-called Itō-isometry allows to define the "stochastic integral" $\int_0^t H_s dW_s$ for a much larger class of integrands H.
Idea: Let $H = (H_s)$ be such that

(i) H is $\mathbb{F}$-adapted

(ii) $\mathbf{E}[\int_0^t H_s^2 ds] < \infty, \quad t \geq 0.$

Approximate H by simple integrands $H^{(n)}$ such that

$$\mathbf{E}[\int_0^t (H_s - H_s^{(n)})^2 ds] \longrightarrow 0.$$

Then $\int_0^t H_s^{(n)} dW_s$ converges in L^2. The limits - denoted by $\int_0^t H_s dW_s$ - constitute a continuous $\mathbb{F}$-martingale. (For proving the continuity of paths, one uses Doob's submartingale inequalities.)

One can even go beyond integrands obeying $\mathbf{E}[\int_0^t H_s^2 ds] < \infty$.
Assume

$$\left.\begin{array}{l} H \text{ is } \mathbb{F}\text{-adapted and} \\ \mathbf{P}[\int_0^t H_s^2 ds < \infty] = 1 \end{array}\right\} (*)$$

By introducing stopping times $\tau_n := \inf\{t : \int_0^t H_s^2 ds \geq n\}$ we can define, with the above recipe, the martingales

$$\int_0^{t \wedge \tau_n} H_s dW_s, \quad t \geq 0$$

and put $\int\limits_0^t H_s dW_s := \lim\limits_n \int\limits_0^{t\wedge\tau_n} H_s dW_s$.
This is not necessarily a martingale, but in any case a so called local martingale.

Definition 7.5.3 *$(M_t)_{t\geq 0}$ is called a* **local $\mathbb{F}$-martingale:** $\Longleftrightarrow$ *there exists a sequence of $\mathbb{F}$-stopping times τ_n with $\tau_n \to \infty$ a.s. and, for alll $n = 0, 1, \dots$, $(M_{t\wedge\tau_n} - M_0)_{t\geq 0}$ is an $\mathbb{F}$-martingale.*

It turns out that - beside the classical integrands of bounded variation - the local martingales are the right class of "'stochastic integrands". In this context, a fundamental role is played by the quadratic variation and covariation process of local martingales.

7.6 Integrators of locally finite variation

Definition 7.6.1 *A function $a : \mathbb{R} \longrightarrow \mathbb{R}$ is called of* **locally finite variation***: $\Longleftrightarrow$ for all $t > 0$,*

$$v_t(a) := \sup_{(t_k) \text{ partition of } [0,t]} \sum |a(t_{k+1}) - a(t_k)| < \infty$$

Note that $t \mapsto v_t(a)$ is increasing (and continuous if a is continuous). Every such a can be uniquely written in the form

$$a = a^{(+)} - a^{(-)}, \text{ with nondecreasing } a^{(+)} \text{ and } a^{(-)},$$

and

$$v_t(a) = a_t^{(+)} + a_t^{(-)}.$$

Therefore, functions of locally finite variation are just differences of incresasing functions. But we know how to treat increasing functions as integrands, viewing them as weight functions of measures. For measurable $h : \mathbb{R}_+ \longrightarrow \mathbb{R}$ we define

$$\int\limits_0^t h(s)da(s) := \int\limits_0^t h(s)da^{(+)}(s) - \int\limits_0^t h(s)da^{(-)}(s)$$

provided both terms on the r.h.s are finite.
If $\int\limits_0^t |h(s)|dv_t(s) < \infty \quad \forall t$, we say that h is **locally a-integrable**.

Example ("Absolutely continuous functions") Consider

$$a(t) := \int\limits_0^t b(s)ds$$

Then

$$v_t(a) = \int\limits_0^t |b(s)|ds, \text{ and } \int\limits_0^t h(s)da(s) = \int\limits_0^t h(s)b(s)ds.$$

It turns out (and is not even difficult to prove) that there are no non-trivial continuous local martingales of locally finite variation.

Proposition 7.6.1 *(Ka 17.2)*
If M is a continuous local martingale of locally finite variation, then $M = M_0$ a.s.

7.7 Continuous local martingales as integrators

The previous proposition suggests that in a calculus dealing with a local martingale M, not only the increment dM but also the squared increment $(dM)^2$ might play a role.
An intuitive key to the understanding of $(dM)^2$ is the follow easy

Remark 7.7.1 *For a square integrable martingale M,*

$$\begin{aligned}\mathbf{E}[(M_{t+h}-M_t)^2|\mathcal{F}_t] &= \mathbf{E}[M_{t+h}^2 - 2M_{t+h}M_t + M_t^2|\mathcal{F}_t] \\ &= \mathbf{E}[M_{t+h}^2 - M_t^2|\mathcal{F}_t]\end{aligned}$$

This says that the predicted squared increment of M is the predicted increment of the square of M. Thus, it seems reasonable to relate $(dM)^2$ to the "predictor" of the submartingale M^2 (which would be an increasing process $[M]$ making $M^2-[M]$ a martingale). Indeed, this is no vain hope.

Theorem 7.7.1 *(quadratic variation and covariation of continuous local martingales Ka 17.5)*

a) For any continuous local martingale M there exists an a.s. unique continuous increasing process $[M]$ with $[M]_0 = 0$ such that

$$M^2 - [M] \text{ is a local martingale.}$$

$[M]$ is called the quadratic vatiation (process) of M.

b) For any continuous local martingales M and N,

$$[M,N] := \frac{1}{4}([M+N]-[M-N])$$

is the a.s. unique continuous process of locally finite variation and with $[M,N]_0 = 0$ such that

$$MN - [M,N] \text{ is a local martingale .}$$

$[M,N]$ is called the covariation (process) of M and N.

The mapping $(M,N) \longrightarrow [M,N]$ is a.s. symmetric and bilinear, with

$$[M,N] = [M-M_0, N-N_0] \quad a.s.$$

Furthermore, for every stopping time τ,

$$[M^\tau, N] = [M^\tau, N^\tau] = [M,N]^\tau \quad a.s.$$

(where $M_t^\tau := M_{t\wedge\tau}$).

The following fact sheds light on the meaning of $[M]$ (and plays a role in the proof of the previous theorem).

Proposition 7.7.1 *(Ka 17.18) Let M be a continuous local martingale, fix any t and consider a sequence of partitions (t_k^n) of $[0,t]$ with mesh size $\longrightarrow 0$. Then*

$$\sum_k (M_{t_{k+1}^n} - M_{t_k^n})^2 \longrightarrow [M]_t \quad \text{in probability, and} \tag{7.1}$$

$$\sum_k (M_{t_{k+1}^n} - M_{t_k^n})(N_{t_{k+1}^n} - N_{t_k^n}) \longrightarrow [M,N]_t \quad \text{in probability.}$$

Corollary 7.7.1

$$|[X;Y]|_t \leq \sqrt{[X]_t}\sqrt{[Y]_t} \qquad a.s$$

Remark 7.7.2 *For a simple integrand H and a continuous local martingale M, we define $\int_0^\bullet H_s dM_s$ analogous to Definition 7.5.2. This is again a local martingale, and its quadratic variation is*

$$[\int_0^\bullet H_s dM_s] = \int_0^\bullet H_s^2 d[M]_s.$$

Its covariation with a local martingale N is

$$[\int_0^\bullet H_s dM_s, N] = \int_0^\bullet H_s d[M,N]_s].$$

It is probably easier to recall the following differential mnemonics:

$$(H_t\ dM_t)^2 = H_t^2(dM_t)^2$$
$$H_t\ dM_t\ dN_t = H_t\ dM_t\ dN_t$$

Definition 7.7.1 *For a continuous local martingale M, we put*
$L(M) := \{H : H\ \mathbb{F}\text{-adpated and} \int\limits_0^t H_s^2 d[M]_s < \infty \text{ for all } t\}$
(calling these H the **locally M-integrable processes**)

An elegant geometric approach to the stochastic integral is to characterize it in terms of its covariation with all the other continuous local martingales:

Theorem 7.7.2 *(stochastic integral, Itō, Kunita and Watanabe)(Ka 15.12) For every continuous local martingale M and every process $H \in L(M)$ there exists an a.s. unique continuous local martingale $H \bullet M$ with $(H \bullet M)_0 = 0$ such that*

$$[H \bullet M, N] = \int_0^\bullet H_s d[M,N]_s =: H \bullet [M,N] \qquad a.s.$$

for every continuous local martingale N.

Theorem 7.7.3 *(Ka 15.23) The integral $H \bullet M$ is the a.s. unique linear extension of the elementary stochastic integral such that for every $t > 0$ the convergence*

$$(H_n^2 \bullet [M])_t \longrightarrow 0 \text{ in probability}$$

implies

$$\sup_{0 \leq s \leq t} (H_n \bullet M)_s \longrightarrow 0 \text{ in probability}$$

In particular we have in the setting of Proposition 7.7.1

$$\sum_k M_{t_k^n}(M_{t_{k+1}^n} - M_{t_k^n}) \to \int_0^t M_s dM_s.$$

Together with (7.1) and the same reasoning as in Example 7.3.1 this shows:

Remark 7.7.3 *For any continuous local martingale M,*

$$M_t^2 = M_0^2 + 2\int\limits_0^t M_s dM_s + [M]_t.$$

7.8 Stochastic calculus for continuous local semimartingales

Definition 7.8.1 *A continuous $\mathbb{F}$-semimartingale X is the sum of a continuous local $\mathbb{F}$-martingale M and a continuous $\mathbb{F}$-adapted process A of locally finite variation.*

By Proposition 7.6.1, the decomposition

$$X = M + A$$

is a.s. unique.

We put

$$L(X) := \ H : H \text{ is locally } M\text{-integrable and locally } A\text{-integrable}\}$$

For $H \in L(X)$, we define the stochastic integral $H \bullet X$ as

$$H \bullet X := H \bullet M + H \bullet A.$$

For two semimartingales $X = M + A$, $Y = N + B$ we put

$$[X, Y] = [M, N].$$

In particular

$$[X] = [M].$$

Proposition 7.8.1 *(chain rule, Ka 17.15) Let X be a contiunous semimartingale, and $H \in L(X)$. Then $J \in L(H \bullet X)$ iff $J \bullet H \in L(X)$, in which case $J \bullet (H \bullet X) = (JH) \bullet X$.*

Proposition 7.8.2 *(stopping) For any continuous semimartingale X, any process $H \in L(X)$ and stopping time τ we have a.s.*

$$(H \bullet X)^\tau = H \bullet X^\tau = (H 1_{[0,\tau]}) \bullet X.$$

The following gives one more interpretation of the covariation term $[X, Y]$ as the "remainder term" in the integration by parts formula (compare with Remark 7.7.3)

Theorem 7.8.1 *(Integration by parts, Ka ch. 17) For any continuous semimartingale X, Y we have a.s.*

$$XY = X_0 Y_0 + X \bullet Y + Y \bullet X + [X, Y]$$

Theorem 7.8.2 *(substitution rule, Itō's formula, Ka 17.18)*
a) Let X be a continuous semimartingale, $f \in C^2(\mathbb{R})$,. Then

$$f(X_t) = f(X_0) + \int_0^t f'(X_s) dX_s + \frac{1}{2} \int_0^t f''(X_s) d[X]_s.$$

b) Let $X^1, \dots, X^d$ be continuous semimartingales, $f \in C^2(\mathbb{R}^d)$, $X = (X^1, \dots, X^d)$. Then

$$f(X) = f(X_0) + \sum_{i=1}^d \frac{\partial}{\partial x_i} f(X_s) dX_s^i + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} f(X_s) d[X^i, X^j]_s \qquad a.s.$$

A most prominent example is $X_t = (W_t, t)$:

$$df(W_t,t) = \frac{\partial}{\partial x_1} f(W_t,t)dW_t + \frac{\partial}{\partial x_2} f(W_t,t)dt + \frac{1}{2}\frac{\partial^2}{\partial x_1^2} f(W_t,t)dt,$$

since $Y_t = t$ has locally finite variation, hence $[Y] \equiv 0$ and (see Corollary 7.7.1) $[Y,W] \equiv 0$.

It is suggestive to think of Itō's formula as a second order Taylor expansion:

$$df(X) = \sum_i f_i'(X)dX_i + \frac{1}{2}\sum_{i,j} f_{ij}''(X)dX_i dX_j$$

(where f_i denotes $\frac{\partial}{\partial x_i} f$.)

Example 7.8.1 *For a standard Wiener process W, constant σ and $X_0 > 0$, consider the strictly positive process*

$$S_t := S_0 \exp(\sigma W_t - \frac{1}{2}\sigma^2 t). \tag{7.2}$$

Writing $f(z) := e^z$, $Z_t := \sigma W_t - \frac{1}{2}\sigma^2 t$, we obtain from Itō's formula

$$\begin{aligned} dS_t &= X_t dZ_t + \frac{1}{2} X_t d[Z]_t \qquad (7.3)\\ &= X_t(\sigma dW_t - \frac{1}{2}\sigma^2 dt + \frac{1}{2}\sigma^2 dt)\\ &= X_t \sigma dt. \end{aligned}$$

Hence X_t is a local martingale. Indeed, an elementary calculation shows that

$$\mathbf{E}e^{\sigma W_t} = \frac{1}{\sqrt{2\pi t}}\int e^{-\frac{x^2}{2t}} e^{\sigma x}dx = e^{\frac{1}{2}\sigma^2 t}\frac{1}{\sqrt{2\pi t}}\int e^{-(\frac{x}{\sqrt{2t}} - \frac{\sigma\sqrt{t}}{\sqrt{2}})^2}dx = e^{\frac{1}{2}\sigma^2 t}.$$

Together with the independence of increments of W, this shows that X is even a martingale (provided X_0 is integrable). X is called **geometric Brownian motion with volatility** σ *(and initial value X_0.)*

By the way, the solution to the "stochastic differential equation" (7.3) with initial condition $X_0 = x_0$ is a.s. unique. Indeed, consider some continuous local martingale Y with $Y_0 = x_0$ and

$$dY = YdW.$$

Then $Z_t := \frac{Y_t}{X_t}$ obeys, by Itō's formula, $dZ_t = 0$ (check!), hence $Y = X$ a.s.

7.9 Lévy's characterisation of W

We saw in Subsection 7.3. and 7.4 that a standard $\mathbb{F}$-Wiener process W is an $\mathbb{F}$-martingale with $W_0 = 0$ and quadratic variation $[W]_t = t$ a.s. We will now prove that each continuous local martingale M with $M_0 = 0$ and $[M]_t = t$ is in fact a Wiener process. This will be achieved by analysing the process $\exp(i\alpha M_t + \frac{\alpha^2}{2}t)$, which will turn out to be a complex-valued martingale, and will help to identify the conditional distribution of the increment $M_t - M_s$, given $\mathcal{F}_s$.

As a preparation, we state Itō's formula for complex-valued continuous semimartingales.

By a complex-valued continuous semimartingale we mean a process of the form $Z = X + iY$, where X and Y are real continuous semimartingales. We put

$$\begin{aligned}[Z] := [Z, Z] &:= [X + iY, X + iY] = \\ &= [X] + i[X, Y] - [Y].\end{aligned}$$

For $Z = X + iY, K = H + iJ$, $H, J \in L(X) \cap L(Y)$, we put

$$K \bullet Z := H \bullet X - J \bullet Y + i(H \bullet Y + J \bullet X),$$

An easy consequence of Theorem 7.5 is

Corollary 7.9.1 *(Ka 17.20) Let $f : \mathbb{C} \to \mathbb{C}$ be differentiable, and Z be a complex-valued continuous semimartingale. Then*

$$f(Z) = f(Z_0) + f'(Z) \bullet Z + \frac{1}{2} f''(Z) \bullet [Z] \qquad a.s\ .$$

Example 7.9.1 *Let W be a standard Wiener Process. Put $X_t := e^{i\alpha W_t + \frac{\alpha^2}{2}t} =: e^{Z_t}$. Then $[Z]_t = -\alpha^2 t$, and by Itō's formula (with $f(x) = e^x$)*

$$\begin{aligned} dX &= XdZ + \frac{1}{2} X d[Z] \\ &= X(i\alpha dW + \frac{\alpha^2}{2}dt - \frac{1}{2}\alpha^2 dt) \\ &= Xi\alpha dW. \end{aligned}$$

Hence X is a local martingale, and since

$$\sup_{0 \le s \le t} |X_s| = \exp(\frac{\alpha^2}{2}t) < \infty,$$

X is even a martingale (use dominated convergence!). Thus: $\mathbf{E}e^{i\alpha W_t} = e^{\frac{\alpha^2}{2}t}$.

Corollary 7.9.2 *(Characteristic function of the normal distribution) For the normal distribution v with mean 0 and variance t,*

$$\int_{\mathbb{R}} e^{i\alpha y} v(dy) = e^{\frac{\alpha^2}{2}t}, \alpha \in \mathbb{R}.$$

Theorem 7.9.1 *(Lévy's characterization of the Wiener process) Let M by a continuous local $\mathbb{F}$-martingale with quadratic variation $[M]_t = t$, $\quad t \ge 0$, and $M_0 = 0$. Then M is a standard $\mathbb{F}$-Wiener process.*

Proof: Since a distribution on $\mathbb{R}$ is uniquely determined by its characteristic function (Ka 4.3), it suffices to show

$$\mathbf{E}[e^{i\alpha(M_t - M_s)} | \mathcal{F}] = e^{-\frac{1}{2}\alpha^2(t-s)} \qquad \text{a.s.}$$

Or, in other words

$$\mathbf{E}[e^{i\alpha M_t + \frac{1}{2}\alpha^2 t} | \mathcal{F}_s] = e^{i\alpha M_s + \frac{1}{2}\alpha^2 s} \qquad \text{a.s.} \tag{7.4}$$

Putting $X_t. = e^{\alpha M_t + \frac{\alpha^2}{2}t}$, we infer exactly as in Example 7.1. that X is martingale, which yields property (7.4). □

7.10 Reweighting the probability = changing the drift

Up to now we always considered one single probability measure P on our σ-field of events. We will now consider another probability measure Q which s "locally absolutely continuous" with respect to P in the sense of (7.5) below. If M is a continuous local martingale under P, it is not necessarily a continuous local martingale under Q. Indeed, if the "density process" of Q with respect to P has a positive covariance with M, passing from P to Q will generate a positive drift. However, such a "change of measure" does not affect the quadratic variation process $[M]$.

Let's now make things precise. Let P and Q be two probability measures on $\mathcal{F}_\infty$. Assume that for any t there exists a Z_t such that

$$Q(F) = \mathbf{E}[I_F Z_t] \quad , \quad F \in \mathcal{F}_t. \tag{7.5}$$

We then say that Q has *density* Z_t with respect to P on $\mathcal{F}_t$ (writing $Q = Z_t \cdot P$ on $\mathcal{F}_t$), and call (Z_t) *density process* of Q w.r.t. to P.
Since $\mathcal{F}_s \subseteq \mathcal{F}_t$ we have for all $F \in \mathcal{F}_t$

$$\mathbf{E}[Z_s I_F] = Q(F) = \mathbf{E}[Z_t I_F].$$

Hence (Z_t) is a martingale.

Lemma 7.10.1 *(Ka 18.25, 18.26, 18.17) Let $Q = Z_t \cdot P$ on $\mathcal{F}_t$ for all $t \geq 0$. (That is, assume that (7.5) holds.) Then Z is a P-martingale. Moreover, if Z is P-a.s. continuous, then*

a) an adapted continuous process X is a local Q-martingale iff $X \cdot Z$ is a local P-martingale.

b) for all $t > 0$, $\quad \inf_{s \leq t} Z_s > 0 \quad Q$ a.s

Theorem 7.10.1 *(transformation of drift, Ka 18.19) Let $Q = Z_t P$ on $\mathcal{F}_t$ for all $t \geq 0$, and assume that Z is a.s continuous. Then for any continuous local P-martingale M, the process*

$$\tilde{M} = M - \frac{1}{Z} \bullet [M, Z]$$

is a local Q-martingale.

Sketch of proof:

a) If Z^{-1} is bounded, then $\tilde{M}$ is a continuous P-semimartingale, and we get

$$\begin{aligned} d(\tilde{M}Z) &= \tilde{M}dZ + Z \cdot d\tilde{M} + d[\tilde{M}, Z] \\ &= \tilde{M}dZ + Z \cdot dM - d[M, Z] + d[\tilde{M}, Z] \end{aligned}$$

However, since $M$and $\tilde{M}$ differ only by a process of locally finite variation, the last two terms cancel. Hence $\tilde{M}Z$ is a continuous local P-martingale,and, by Lemma 7.2, $\tilde{M}$ is a local Q-martingale.

b) In general, consider $\tau_n := \inf\{t \geq 0 : Z_t < \frac{1}{n}\}$ and argue as in a) that $\tilde{M}^{\tau_n} Z$ is a continuous local P-martingale. Hence every $\tilde{M}^{\tau_n}$, and therefore also $\tilde{M}$, is a local Q-martingale.

$\square$

Example 7.10.1 *Let M be a continuous local martingale, and $B \in L(M)$.
Then $Z = \exp(B \cdot M - \frac{1}{2}B^2 \cdot [M])$ is a continuous local martingale. Indeed, by Itō's formula (check!)*

$$dZ = ZB \quad dM.$$

Assume that, for a fixed $t > 0$,
$(Z_s)_{0 \le s \le t}$ is even a martingale
(equivalent to this is

$$EZ_t \equiv 1 \qquad (\text{“Girsanov's condition”})$$

and sufficient for this is

$$\mathbf{E}[\exp(\frac{1}{2}\int_0^t B_s^2 d[M]_s)] < \infty \qquad (\text{"Novikov's condition"}))$$

Then, under $Q := Z_t \bullet P$,

$$\tilde{M} := M - B \bullet [M]$$

is a local martingale.
Indeed, by Theorem 7.10.1 it suffices to check that

$$\frac{1}{Z} \bullet [M, Z] = B \bullet [M]$$

However,

$$\begin{aligned} [M, Z] &= [M, ZB \bullet M] \\ &= ZB \bullet [M] \end{aligned}$$

(Again, the differential abbreviation is more suggestive:

$$\frac{1}{Z} dM dZ = \frac{1}{Z} ZB(dM)^2 = B(dM)^2.)$$

Special case: Let W be a standard Wiener process under P. Put

$$Z_s := \exp\left(\int_0^s B_u dW_u - \frac{1}{2}\int_0^s B_u^2 ds\right)$$

Assume $\mathbf{E}Z_t = 1$. Then, under $Q := Z_t \cdot P$,
$\tilde{W}_s := W_s - \int_0^s B_u du$,

$$0 \le s \le t,$$

is a local martingale with quadratic variation process $[\tilde{W}] = [W]$, hence $\tilde{W}$ is a standard Wiener process by Lévy's characterization!
This is the classical Girsanov theorem.

Remark 7.10.1 *In the situation of Example 7.1, the density Z_t of Q with respect to P on $\mathcal{F}_t$ is strictly positive, and hence an event $F \in \mathcal{F}_t$ has P-probability zero iff it has Q-probability 0. We say in this case that P and Q are* **equivalent** *probability measures on $\mathcal{F}_t$.*

7.11 A strategy for (almost) all cases

Example 7.11.1

Consider a geometric Brownian motion X with volatility $\sigma > 0$ and (possibly random) initial value X_0 (see Example 7.1). Let $\mathbb{F}$ be the filtration generated by X, and $T > 0$ be a fixed time.
Task: For given (integrable) $f(X_T)$, look for adapted processes H and G such that

$$dG_t = H_t dX_t$$

and

$$G_T = f(X_T)$$

or in other words,

$$f(X_T) = G_0 + \int_0^T H_t \, dX_t \quad \text{a.s.}$$

Thus we look for an initial value G_0 and a "strategy" H that yields the final value $f(X_T)$ for almost all paths X.

Observe that (because of the projection property of conditional expectations)

$$G_t := \mathbf{E}[f(X_T)|\mathcal{F}_t], \quad t \geq 0$$

is a martingale.
On the other hand, G_t is of the form $g(t, X_t)$. Indeed, since $X_T/X_t = \exp(\sigma(W_T - W_t) - \frac{1}{2}\sigma^2(T-t))$ and W has independent increments,

$$G_t = \mathbf{E}[f(X_T)|\mathcal{F}_t] = \mathbf{E}[f(X_t \frac{X_T}{X_t}|\mathcal{F}_t] = g(t, X_t) \quad \text{a.s.}$$

where

$$g(t,x) := \mathbf{E}[f(x \exp(\sigma W_{T-t} - \frac{\sigma^2}{2}(T-t))]$$

By Itō's formula,

$$g(t, X_t) - g(0, X_0) = \int_0^t \frac{\partial}{\partial x} g(s, X_s) dX_s + A_t, \tag{7.6}$$

where A is of locally finite variation. Since all the other terms in (7.6) are continuous martingales, so is A. Hence A vanishes by Proposition 7.2.
Since

$$g(T, X_T) = G_T = \mathbf{E}[f(X_T)|\mathcal{F}_T] = f(X_T) \quad \text{a.s.}$$

and

$$g(0, X_0) = G_0 = \mathbf{E}[f(X_T)|X_0] \quad \text{a.s.}$$

we get from (7.6) (recalling that $A \equiv 0$) the representation

$$f(X_T) = \mathbf{E}[f(X_T)|X_0] + \int_0^t \frac{\partial}{\partial x} g(s, X_s) dX_s \quad \text{a.s.}$$

This gives a formula both for the required *initial value*

$$G_0 = \mathbf{E}[f(X_T)|X_0] = \mathbf{E}[f(X_0 \exp(\sigma W_T - \frac{\sigma^2}{2} T))|X_0]$$

and the *hedging strategy*

$$H_s = \frac{\partial}{\partial x} g(s, X_s).$$

Example 7.11.2

Let W be a standard Wiener process, and S be a geometric Brownian motion with

$$dS_t = S_t \, \sigma \, dW_t.$$

Let $\mathbb{F}$ be the filtration generated by S, and fix a constant $r > 0$ (think of S as a "stock price" and r as an "interest rate".) Let $T > 0$ be a fixed time.
Task: for given (integrable) $h(S_T)$, look for adapted processes H and V such that

$$dV_t = H_t dS_t + [V_t - H_t S_t] \, r \, dt \tag{7.7}$$

and

$$V_T = h(S_T). \tag{7.8}$$

Think of (V_t) as the value process of a portfolio, and (H_t) as trading strategy: at time t, one holds H_t units of the stock and puts an amount of $V_t - H_t S_t$ on the savings account. This leads to an increment dV_t given by (7.7). Like in Example 7.4, we look for an initial value V_0 and a strategy H that yields the final value $h(S_T)$ for almost all paths S.
Consider the "discounted processes"

$$X_t := e^{-rt} S_t \tag{7.9}$$

and

$$G_t := e^{-rt} V_t \tag{7.10}$$

By Itō's formula we have

$$\begin{aligned} dX_t &= e^{-rt}[dS_t - S_t r \, dt] = e^{-rt} S_t [\sigma dW_t - r \, dt] \\ &= X_t [\sigma dW_t - r dt] \end{aligned} \tag{7.11}$$

and

$$\begin{aligned} dG_t &= e^{-rt}[dV_t - V_t r \, dt] \\ &= e^{-rt}[H_t dS_t - H_t S_t r \, dt] \\ &= H_t e^{-rt} S_t [\sigma dW_t - r \, dt] \\ &= H_t dX_t \end{aligned} \tag{7.12}$$

Let Q be a probability measure equivalent to P on $\mathcal{F}_T$ and such that

$$d\widetilde{W} := dW - \frac{r}{\sigma} dt$$

defines a standard Wiener process under Q. Then (7.11) and 7.12) translate into

$$dX_t = \sigma d\widetilde{W}_t. \tag{7.13}$$

Moreover, the final condition for G is

$$G_T = e^{-rT} V_T = e^{-rT} h(S_T) = e^{-rT} h(e^{rT} X_T). \tag{7.14}$$

Putting

$$f(x) := e^{-rT} h(e^{rT} x),$$

(7.14) writes as

$$G_T = f(X_T). \tag{7.15}$$

Looking at example 7.11.1, we see that a solution to (7.13), (7.12), (7.15) is given by

$$G_t = \mathbf{E}_Q[f(X_T)|\mathcal{F}_t] = g(t, X_t),$$

where

$$g(t,x) = \mathbf{E}_Q[f(x\exp(\sigma\widetilde{W}_{T-t} - \frac{\sigma^2}{2}(T-t))],$$

and

$$H_t = \frac{\partial}{\partial x} g(t, X_t).$$

In particular,

$$\begin{aligned} V_0 &= G_0 = g(0, X_0) \qquad (7.16) \\ &= \mathbf{E}_Q[f(X_0\exp(\sigma\widetilde{W}_T - \frac{\sigma^2}{2}T))|X_0] \\ &= \mathbf{E}_Q[e^{-rt}h(e^{rT}S_0\exp(\sigma\widetilde{W}_T - \frac{\sigma^2}{2}T))|S_0] \quad \text{a.s.} \end{aligned}$$

Example 7.11.3

Let the process S obey

$$dS_t = S_t\ dY_t$$

where Y is a continuous semimartingale with quadratic variation $\sigma^2 t$:

$$dS_t = \sigma dW_t + dA_t$$

Task: for given $h(S_T)$, find adapted processes H and V such that (7.7) and (7.8) are valid.
For X and G defined by (7.9) and (7.10), we have (compare (7.11) and (7.12))

$$dX_t = X_t[dY_t - r\ dt] = X_t[\sigma dW_t + dA_t - r\ dt]$$

and

$$dG_t = H_t dX_t.$$

Let the probability measure Q (equivalent to P on $\mathcal{F}_T$) be such that

$$d\widetilde{W} := dW + \frac{1}{\sigma}(dA_t - r\ dt)$$

defines a standard Wiener process under Q. Then (7.13) is valid again, and the formula (7.16) for H and V also provide a solution to (7.7) and (7.8). This is a variant of the celebrated Black-Scholes formula, giving the fair price V_0 and the "replication"

$$h(S_T) = V_0 + \int_0^T H_t dS_t + \int_0^T (V_t - H_t S_t) r\, dt$$

of a claim $h(S_T)$ via a self-finacing trading strategy.