Stochastic Processes

Lecture Notes

Zakhar Kabluchko

WWU Münster

Vorwort

Dies ist ein unvollständiges Skript zur Vorlesung "Stochastik II (Stochastische Prozesse)", die an der Universität Ulm im Wintersemester 2013/2014 gehalten wurde. Ich bedanke mich bei Judith Schmidt für die Erstellung der ersten LaTeX-Version dieses Skriptes. Danach wurde das Skript von mir überarbeitet, korrigiert und ergänzt.

Bei Fragen, Wünschen und Verbesserungsvorschlägen können Sie gerne eine E-Mail an

`zakhar DOT kabluchko AT uni-muenster DOT de`

schreiben.

November 17, 2017

Zakhar Kabluchko

References

1. R. Durrett. *Probability: Theory and Examples*, Cambridge University Press.
2. R. Durrett. *Essentials of Stochastic Processes*, Springer Texts in Statistics.
3. D. Williams. *Probability with Martingales*, Cambridge University Press.
4. G. Lawler. *Introduction to Stochastic Processes*. Chapman and Hall.
5. G. Grimmett, D. Stirzaker. *Probability and Random Processes*. Oxford University Press.
6. P. Mörters, Y. Peres. *Brownian Motion*. Cambridge University Press.
7. J. Lamperti. *Probability: A Survey of Mathematical Theory*, Wiley.
8. A. Klenke. *Probability Theory: A Comprehensive Course*. Deutsche Version: A. Klenke. *Wahrscheinlichkeitstheorie*.
9. R. Schilling, L. Partzsch. *Brownian motion: An Introduction to Stochastic Processes*, de Gruyter. Mit Beiträgen von B. Böttcher.
10. S. Resnick. *Adventures in Stochastic Processes*. Birkhäuser.
11. G. Lawler, L. Coyle. *Lectures on Contemporary Probability*, American Mathematical Society.
12. J. F. C. Kingman. *Poisson Processes*, Oxford University Press.
13. G. Kersting, A. Wakolbinger. *Zufallsvariable und Stochastische Prozesse*, Birkhäuser.
14. R. Bass. *Stochastic Processes*, Cambridge University Press.

Contents

CHAPTER 1

General theory of stochastic processes

1.1. Definition of stochastic process

First let us recall the definition of a random variable. A random variable is a random number appearing as a result of a random experiment. If the random experiment is modeled by a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then a random variable is defined as a function $\xi : \Omega \to \mathbb{R}$ which is measurable. Measurability means that for every Borel set $B \subset \mathbb{R}$ it holds that $\xi^{-1}(B) \in \mathcal{F}$. Performing the random experiment means choosing the outcome $\omega \in \Omega$ at random according to the probability measure $\mathbb{P}$. Then, $\xi(\omega)$ is the value of the random variable which corresponds to the outcome ω.

A *stochastic process* is a *random function* appearing as a result of a random experiment.

Definition 1.1.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let T be an arbitrary set (called the index set). Any collection of random variables $X = \{X_t : t \in T\}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *stochastic process* with index set T.

So, to every $t \in T$ corresponds some random variable $X_t : \Omega \to \mathbb{R}$, $\omega \mapsto X_t(\omega)$. Note that in the above definition we require that all random variables X_t are defined on the *same* probability space. Performing the random experiment means choosing an outcome $\omega \in \Omega$ at random according to the probability measure $\mathbb{P}$.

Definition 1.1.2. The function (defined on the index set T and taking values in $\mathbb{R}$)

$$t \mapsto X_t(\omega)$$

is called the *sample path* (or the *realization*, or the *trajectory*) of the stochastic process X corresponding to the outcome ω.

So, to every outcome $\omega \in \Omega$ corresponds a trajectory of the process which is a function defined on the index set T and taking values in $\mathbb{R}$.

Stochastic processes are also often called *random processes*, *random functions* or simply *processes*.

Depending on the choice of the index set T we distinguish between the following types of stochastic processes:

1. If T consists of just one element (called, say, 1), then a stochastic process reduces to just one random variable $X_1 : \Omega \to \mathbb{R}$. So, the concept of a stochastic process includes the concept of a random variable as a special case.

2. If $T = \{1, \ldots, n\}$ is a finite set with n elements, then a stochastic process reduces to a collection of n random variables $X_1, \ldots, X_n$ defined on a common probability space. Such

a collection is called a random vector. So, the concept of a stochastic process includes the concept of a random vector as a special case.

3. Stochastic processes with index sets $T = \mathbb{N}$, $T = \mathbb{Z}$, $T = \mathbb{N}^d$, $T = \mathbb{Z}^d$ (or any other countable set) are called stochastic processes with *discrete time*.

4. Stochastic processes with index sets $T = \mathbb{R}$, $T = \mathbb{R}^d$, $T = [a, b]$ (or other similar uncountable sets) are called stochastic processes with *continuous time*.

5. Stochastic processes with index sets $T = \mathbb{R}^d$, $T = \mathbb{N}^d$ or $T = \mathbb{Z}^d$, where $d \geq 2$, are sometimes called *random fields*.

The parameter t is sometimes interpreted as "time". For example, X_t can be the price of a financial asset at time t. Sometimes we interpret the parameter t as "space". For example, X_t can be the air temperature measured at location with coordinates $t = (u, v) \in \mathbb{R}^2$. Sometimes we interpret t as "space-time". For example, X_t can be the air temperature measured at location with coordinates $(u, v) \in \mathbb{R}^2$ at time $s \in \mathbb{R}$, so that $t = (u, v, s) \in \mathbb{R}^3$.

1.2. Examples of stochastic processes

1. *I.i.d. Noise.* Let $\{X_n : n \in \mathbb{Z}\}$ be independent and identically distributed (i.i.d.) random variables. This stochastic process is sometimes called the i.i.d. noise. A realization of this process is shown in Figure 1, left.

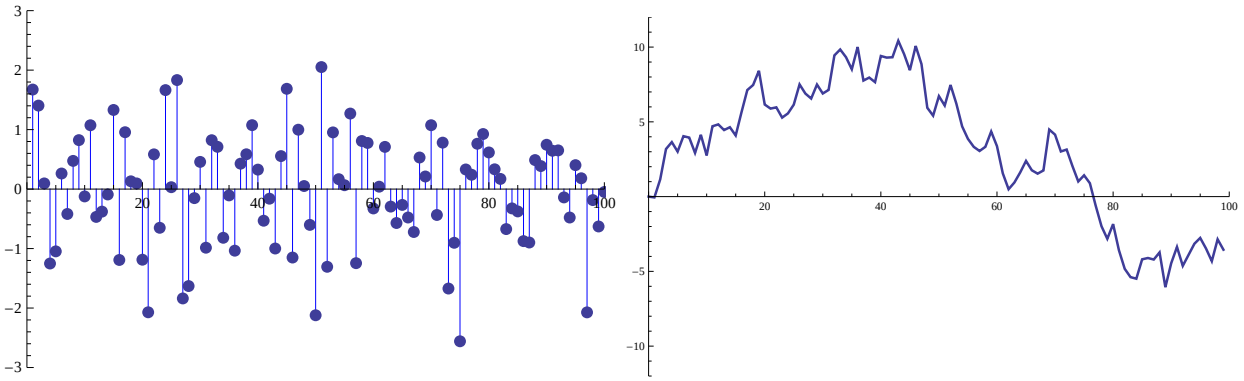


FIGURE 1. Left: A sample path of the i.i.d. noise. Right: A sample path of the random walk. In both cases, the variables X_n are standard normal

2. *Random walk.* Let $\{X_n : n \in \mathbb{N}\}$ be independent and identically distributed random variables. Define

$$S_n := X_1 + \ldots + X_n, \quad n \in \mathbb{N}, \quad S_0 = 0.$$

The process $\{S_n : n \in \mathbb{N}_0\}$ is called the random walk. A sample path of the random walk is shown in Figure 1, right.

3. *Geometric random walk.* Let $\{X_n : n \in \mathbb{N}\}$ be independent and identically distributed random variables such that $X_n > 0$ almost surely. Define

$$G_n := X_1 \cdot \ldots \cdot X_n, \quad n \in \mathbb{N}, \quad G_n = 1.$$

The process $\{G_n : n \in \mathbb{N}_0\}$ is called the geometric random walk. Note that $\{\log S_n : n \in \mathbb{N}_0\}$ is a (usual) random walk.

4. *Random lines and polynomials.* Let $\xi_0, \xi_1 : \Omega \to \mathbb{R}$ be two random variables defined on the same probability space. Define

$$X_t = \xi_0 + \xi_1 t, \quad t \in \mathbb{R}.$$

The process $\{X_t : t \in \mathbb{R}\}$ might be called "a random line" because the sample paths $t \mapsto X_t(\omega)$ are linear functions.
More generally, one can consider random polynomials. Fix some $d \in \mathbb{N}$ (the degree of the polynomial) and let $\xi_0, \ldots, \xi_d$ be random variables defined on a common probability space. Then, the stochastic process

$$X_t = \xi_0 + \xi_1 t + \xi_2 t^2 + \ldots + \xi_d t^d, \quad t \in \mathbb{R},$$

might be called a "random polynomial" because its sample paths are polynomial functions.

5. *Renewal process.* Consider a device which starts to work at time 0 and works T_1 units of time. At time T_1 this device is replaced by another device which works for T_2 units of time. At time $T_1 + T_2$ this device is replaced by a new one, and so on. Let us denote the working time of the i-th device by T_i. Let us assume that $T_1, T_2, \ldots$ are independent and identically distributed random variables with $\mathbb{P}[T_i > 0] = 1$. The times

$$S_n = T_1 + \ldots + T_n, \quad n \in \mathbb{N},$$

are called *renewal times* because at time S_n some device is replaced by a new one. Note that $0 < S_1 < S_2 < \ldots$. The number of renewal times in the time interval $[0, t]$ is

$$N_t = \sum_{n=1}^{\infty} \mathbb{1}_{S_n \le t} = \#\{n \in \mathbb{N} : S_n \le t\}, \quad t \ge 0.$$

The process $\{N_t : t \ge 0\}$ is called a *renewal process*.

Many further examples of stochastic processes will be considered later (Markov chains, Brownian Motion, Lévy processes, martingales, and so on).

1.3. Finite-dimensional distributions

A random variable is usually described by its distribution. Recall that the distribution of a random variable ξ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability measure P^ξ on the real line $\mathbb{R}$ defined by

$$P^\xi(A) = \mathbb{P}[\xi \in A] = \mathbb{P}[\{\omega \in \Omega : \xi(\omega) \in A\}], \quad A \subset \mathbb{R} \text{ Borel}.$$

Similarly, the distribution of a random vector $\xi = (\xi_1, \ldots, \xi_n)$ (with values in $\mathbb{R}^n$) is a probability measure P^ξ on $\mathbb{R}^n$ defined by

$$P^\xi(A) = \mathbb{P}[\xi \in A] = \mathbb{P}[\{\omega \in \Omega : (\xi_1(\omega), \ldots, \xi_n(\omega)) \in A\}], \quad A \subset \mathbb{R}^n \text{ Borel}.$$

Now, let us define similar concepts for stochastic processes. Let $\{X_t : t \in T\}$ be a stochastic process with index set T. Take some $t_1, \ldots, t_n \in T$. For Borel sets $B_1, \ldots, B_n \subset \mathbb{R}$ define

$$P_{t_1,\ldots,t_n}(B_1 \times \ldots \times B_n) = \mathbb{P}[X_{t_1} \in B_1, \ldots, X_{t_n} \in B_n].$$

More generally, define $P_{t_1,\ldots,t_n}$ (a probability measure on $\mathbb{R}^n$) by

$$P_{t_1,\ldots,t_n}(B) = \mathbb{P}[(X_{t_1},\ldots,X_{t_n}) \in B], \quad B \subset \mathbb{R}^n \text{ Borel}.$$

Note that $P_{t_1,\ldots,t_n}$ is the distribution of the random vector $(X_{t_1},\ldots,X_{t_n})$. It is called *a finite-dimensional distribution* of X. We can also consider the collection of *all* finite dimensional distributions of X:

$$\mathcal{P} := \{P_{t_1,\ldots,t_n} : n \in \mathbb{N}, t_1,\ldots,t_n \in T\}.$$

It is an exercise to check that the collection of all finite-dimensional distributions if a stochastic process X has the following two properties.

1. *Permutation invariance.* Let $\pi : \{1,\ldots,n\} \to \{1,\ldots,n\}$ be a permutation. Then, for all $n \in \mathbb{N}$, for all $t_1,\ldots,t_n \in T$, and for all $B_1,\ldots,B_n \in \mathcal{B}(\mathbb{R})$,

$$P_{t_1,\ldots,t_n}(B_1 \times \ldots \times B_n) = P_{t_{\pi(1)},\ldots,t_{\pi(n)}}(B_{\pi(1)} \times \ldots \times B_{\pi(n)}).$$

2. *Projection invariance.* For all $n \in \mathbb{N}$, all $t_1,\ldots,t_n,t_{n+1} \in T$, and all $B_1,\ldots,B_n \in \mathcal{B}(\mathbb{R})$ it holds that

$$P_{t_1,\ldots,t_n,t_{n+1}}(B_1 \times \ldots \times B_n \times \mathbb{R}) = P_{t_1,\ldots,t_n}(B_1 \times \ldots \times B_n).$$

To a given stochastic process we can associate the collection of its finite-dimensional distributions. This collection has the properties of permutation invariance and projection invariance. One may ask a converse question. Suppose that we are given an index set T and suppose that for every $n \in \mathbb{N}$ and every $t_1,\ldots,t_n \in T$ some probability measure $P_{t_1,\ldots,t_n}$ on $\mathbb{R}^n$ is given. [A priori, this probability measure need not be related to any stochastic process. No stochastic process is given at this stage.] We can now ask whether we can construct a stochastic process whose finite-dimensional distributions are given by the probability measures $P_{t_1,\ldots,t_n}$. Necessary conditions for the existence of such stochastic process are the permutation invariance and the projection invariance. The following theorem of Kolmogorov says that these conditions are also sufficient.

THEOREM 1.3.1 (Kolmogorov's existence theorem). *Fix any non-empty set T. Let*

$$\mathcal{P} = \{P_{t_1,\ldots,t_n} : n \in \mathbb{N}, t_1,\ldots,t_n \in T\}$$

be a collection of probability measures (so that $P_{t_1,\ldots,t_n}$ is a probability measure on $\mathbb{R}^n$) which has the properties of permutation invariance and projection invariance stated above. Then, there exist a probability space $(\Omega,\mathcal{F},\mathbb{P})$ and a stochastic process $\{X_t : t \in T\}$ on $(\Omega,\mathcal{F},\mathbb{P})$ whose finite-dimensional distributions are given by the collection $\mathcal{P}$. This means that for every $n \in \mathbb{N}$ and every $t_1,\ldots,t_n \in \mathbb{N}$ the distribution of the random vector $(X_{t_1},\ldots,X_{t_n})$ coincides with $P_{t_1,\ldots,t_n}$.

IDEA OF PROOF. We have to construct a suitable probability space $(\Omega,\mathcal{F},\mathbb{P})$ and an appropriate stochastic process $\{X_t : t \in T\}$ defined on this probability space.

STEP 1. Let us construct Ω first. Usually, Ω is the set of all possible outcomes of some random experiment. In our case, we would like the outcomes of our experiment to be functions (the realizations of our stochastic process). Hence, let us define Ω to be the set of all functions defined on T and taking values in $\mathbb{R}$:

$$\Omega = \mathbb{R}^T = \{f : T \to \mathbb{R}\}.$$

STEP 2. Let us construct the functions $X_t : \Omega \to \mathbb{R}$. We want the sample path $t \mapsto X_t(f)$ of our stochastic process corresponding to an outcome $f \in \Omega$ to coincide with the function f. In order to fulfill this requirement, we need to define

$$X_t(f) = f(t), \quad f \in \mathbb{R}^T.$$

The functions X_t are called the canonical *coordinate mappings*. For example, if $T = \{1, \dots, n\}$ is a finite set of n elements, then $\mathbb{R}^T$ can be identified with $\mathbb{R}^n = \{f = (f_1, \dots, f_n) : f_i \in \mathbb{R}\}$. Then, the mappings defined above are just the maps $X_1, \dots, X_n : \mathbb{R}^n \to \mathbb{R}$ which map a vector to its coordinates:

$$X_1(f) = f_1, \ \ \dots, \ \ X_n(f) = f_n, \ \ f = (f_1, \dots, f_n) \in \mathbb{R}^n.$$

STEP 3. Let us construct the σ-algebra $\mathcal{F}$. We have to define what subsets of $\Omega = \mathbb{R}^T$ should be considered as measurable. We want the coordinate mappings $X_t : \Omega \to \mathbb{R}$ to be measurable. This means that for every $t \in T$ and every Borel set $B \in \mathcal{B}(\mathbb{R})$ the preimage

$$X_t^{-1}(B) = \{f : T \to \mathbb{R} : f(t) \in B\} \subset \Omega$$

should be measurable. By taking finite intersections of these preimages we obtain the so-called cylinder sets, that is sets of the form

$$A_{t_1,\dots,t_n}^{B_1,\dots,B_n} := \{f \in \Omega : f(t_1) \in B_1, \dots, f(t_n) \in B_n\},$$

where $t_1, \dots, t_n \in T$ and $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$. If we want the coordinate mappings X_t to be measurable, then we must declare the cylinder sets to be measurable. Cylinder sets do not form a σ-algebra (just a semi-ring).
This is why we define $\mathcal{F}$ as the σ-algebra generated by the collection of cylinder sets:

$$\mathcal{F} = \sigma\left\{A_{t_1,\dots,t_n}^{B_1,\dots,B_n} : n \in \mathbb{N}, t_1, \dots, t_n \in T, B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})\right\}.$$

We will call $\mathcal{F}$ the *cylinder σ-algebra*. Equivalently, one could define $\mathcal{F}$ as the smallest σ-algebra on Ω which makes the coordinate mappings $X_t : \Omega \to \mathbb{R}$ measurable.
Sometimes cylinder sets are defined as sets of the form

$$A_{t_1,\dots,t_n}^{B} := \{f \in \Omega : (f(t_1), \dots, f(t_n)) \in B\},$$

where $t_1, \dots, t_n \in T$ and $B \in \mathcal{B}(\mathbb{R}^n)$. One can show that the σ-algebra generated by these sets coincides with $\mathcal{F}$.

STEP 4. We define a probability measure $\mathbb{P}$ on $(\Omega, \mathcal{F})$. We want the distribution of the random vector $(X_{t_1}, \dots, X_{t_n})$ to coincide with the given probability measure $P_{t_1,\dots,t_n}$, for all $t_1, \dots, t_n \in T$. Equivalently, we want the probability of the event $\{X_{t_1} \in B_1, \dots, X_{t_n} \in B_n\}$ to be equal to $P_{t_1,\dots,t_n}(B_1 \times \dots \times B_n)$, for every $t_1, \dots, t_n \in T$ and $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$. However, with our definition of X_t as coordinate mappings, we have

$$\begin{aligned}\{X_{t_1} \in B_1, \dots, X_{t_n} \in B_n\} &= \{f \in \Omega : X_{t_1}(f) \in B_1, \dots, X_{t_n}(f) \in B_n\} \\ &= \{f \in \Omega : f(t_1) \in B_1, \dots, f(t_n) \in B_n\} \\ &= A_{t_1,\dots,t_n}^{B_1,\dots,B_n}.\end{aligned}$$

Hence, we must define the probability of a cylinder set $A^{B_1,\ldots,B_n}_{t_1,\ldots,t_n}$ as follows:

$$\mathbb{P}[A^{B_1,\ldots,B_n}_{t_1,\ldots,t_n}] = P_{t_1,\ldots,t_n}(B_1 \times \ldots \times B_n).$$

It can be shown that $\mathbb{P}$ can be extended to a well-defined probability measure on $(\Omega, \mathcal{F})$. This part of the proof is non-trivial but similar to the extension of the Lebesgue measure from the semi-ring of all rectangles to the Borel σ-algebra. We will omit this argument here. The properties of permutation invariance and projection invariance are used to show that $\mathbb{P}$ is well-defined. □

EXAMPLE 1.3.2 (Independent random variables). Let T be an index set. For all $t \in T$ let a probability measure P_t on $\mathbb{R}$ be given. Can we construct a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a collection of *independent* random variables $\{X_t : t \in T\}$ on this probability space such that X_t has distribution P_t for all $t \in T$? We will show that the answer is yes. Consider the family of probability distributions $\mathcal{P} = \{P_{t_1,\ldots,t_n} : n \in \mathbb{N}, t_1, \ldots, t_n \in T\}$ defined by

$$P_{t_1,\ldots,t_n}(B_1 \times \ldots \times B_n) = P_{t_1}(B_1) \cdot \ldots \cdot P_{t_n}(B_n), \tag{1.3.1}$$

where $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R})$. It is an exercise to check that permutation invariance and projection invariance hold for this family. By Kolmogorov's theorem, there is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a collection of random variables $\{X_t : t \in T\}$ on this probability space such that the distribution of $(X_{t_1}, \ldots, X_{t_n})$ is $P_{t_1,\ldots,t_n}$. In particular, the one-dimensional distribution of X_t is P_t. Also, it follows from (1.3.1) that the random variables $X_{t_1}, \ldots, X_{t_n}$ are independent. Hence, the random variables $\{X_t : t \in T\}$ are independent.

1.4. The law of stochastic process

Random variables, random vectors, stochastic processes (=random functions) are special cases of the concept of *random element*.

DEFINITION 1.4.1. Let $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ be two measurable spaces. That is, Ω and Ω' are any sets and $\mathcal{F} \subset 2^\Omega$ and $\mathcal{F}' \subset 2^{\Omega'}$ are σ-algebras of subsets of Ω, respectively Ω'. A function $\xi : \Omega \to \Omega'$ is called *$\mathcal{F}$-$\mathcal{F}'$-measurable* if for all $A' \in \mathcal{F}'$ it holds that $\xi^{-1}(A') \in \mathcal{F}$.

DEFINITION 1.4.2. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\Omega', \mathcal{F}')$ a measurable space. A *random element* with values in Ω' is a function $\xi : \Omega \to \Omega'$ which is $\mathcal{F}$-$\mathcal{F}'$-measurable.

DEFINITION 1.4.3. The *probability distribution* (or the *probability law*) of a random element $\xi : \Omega \to \Omega'$ is the probability measure P^ξ on $(\Omega', \mathcal{F}')$ given by

$$P^\xi(A') = \mathbb{P}[\xi \in A'] = \mathbb{P}[\{\omega \in \Omega : \xi(\omega) \in A'\}], \quad A' \in \mathcal{F}'.$$

Special cases:

1. If $\Omega' = \mathbb{R}$ and $\mathcal{F}' = \mathcal{B}(\mathbb{R})$, then we recover the notion of random variable.

2. If $\Omega' = \mathbb{R}^d$ and $\mathcal{F}' = \mathcal{B}(\mathbb{R}^d)$, we recover the notion of random vector.

3. If $\Omega' = \mathbb{R}^T$ and $\mathcal{F}' = \sigma_{cyl}$ is the cylinder σ-algebra, then we recover the notion of stochastic process. Indeed, a stochastic process $\{X_t : t \in T\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ leads to the mapping $\xi : \Omega \to \mathbb{R}^T$ which maps an outcome $\omega \in \Omega$ to the corresponding

trajectory of the process $\{t \mapsto X_t(\omega)\} \in \mathbb{R}^T$. This mapping is $\mathcal{F}$-σ_{cyl}-measurable because the preimage of any cylinder set

$$A^{B_1,\dots,B_n}_{t_1,\dots,t_n} = \{f \in \mathbb{R}^T : f(t_1) \in B_1, \dots, f(t_n) \in B_n\}$$

is given by

$$\xi^{-1}(A^{B_1,\dots,B_n}_{t_1,\dots,t_n}) = \{\omega \in \Omega : X_{t_1}(\omega) \in B_1, \dots, X_{t_n}(\omega) \in B_n\} = X^{-1}_{t_1}(B_1) \cap \dots \cap X^{-1}_{t_n}(B_n).$$

This set belongs to the σ-algebra $\mathcal{F}$ because $X^{-1}_{t_i}(B_i) \in \mathcal{F}$ by the measurability of the function $X_{t_i} : \Omega \to \mathbb{R}$. Hence, the mapping ξ is $\mathcal{F}$-σ_{cyl}-measurable.

To summarize, we can consider a stochastic process with index set T as a random element defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in $\mathbb{R}^T$.

In particular, the probability distribution (or the probability law) of a stochastic process $\{X_t, t \in T\}$ is a probability measure P^X on $(\mathbb{R}^T, \sigma_{cyl})$ whose values on cylinder sets are given by

$$P^X(A^{B_1,\dots,B_n}_{t_1,\dots,t_n}) = \mathbb{P}[X_{t_1} \in B_1, \dots, X_{t_n} \in B_n].$$

1.5. Equality of stochastic processes

There are several (non-equivalent) notions of equality of stochastic processes.

Definition 1.5.1. Two stochastic processes $X = \{X_t : t \in T\}$ and $Y = \{Y_t : t \in T\}$ with the same index set T have the *same finite-dimensional distributions* if for all $t_1, \dots, t_n \in T$ and all $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R})$:

$$\mathbb{P}[X_{t_1} \in B_1, \dots, X_{t_n} \in B_n] = \mathbb{P}[Y_{t_1} \in B_1, \dots, Y_{t_n} \in B_n].$$

Definition 1.5.2. Let $\{X_t : t \in T\}$ and $\{Y_t : t \in T\}$ be two stochastic processes defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and having the same index set T. We say that X is a *modification* of Y if

$$\forall t \in T : \quad \mathbb{P}[X_t = Y_t] = 1.$$

With other words: For the random events $A_t = \{\omega \in \Omega : X_t(\omega) = Y_t(\omega)\}$ it holds that

$$\forall t \in T : \quad \mathbb{P}[A_t] = 1.$$

Note that in this definition the random event A_t may depend on t.

The next definition looks very similar to Definition 1.5.2. First we formulate a preliminary version of the definition and will argue later why this preliminary version has to be modified.

Definition 1.5.3. Let $\{X_t : t \in T\}$ and $\{Y_t : t \in T\}$ be two stochastic processes defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and having the same index set T. We say that the processes X and Y are *indistinguishable* if

$$\mathbb{P}[\forall t \in T : X_t = Y_t] = 1.$$

With other words, it should hold that

$$\mathbb{P}[\{\omega \in \Omega : X_t(\omega) = Y_t(\omega) \text{ for all } t \in T\}] = 1.$$

Another reformulation: the set of outcomes $\omega \in \Omega$ for which the sample paths $t \mapsto X_t(\omega)$ and $t \mapsto Y_t(\omega)$ are equal (as functions on T), has probability 1. This can also be written as

$$\mathbb{P}[\cap_{t\in T} A_t] = 1.$$

Unfortunately, the set $\cap_{t\in T} A_t$ may be non-measurable if T is not countable, for example if $T = \mathbb{R}$. That's why we have to reformulate the definition as follows.

DEFINITION 1.5.4. Let $\{X_t : t \in T\}$ and $\{Y_t : t \in T\}$ be two stochastic processes defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and having the same index set T. The processes X and Y are called *indistinguishable* if there exists a measurable set $A \in \mathcal{F}$ so that $\mathbb{P}[A] = 1$ and for every $\omega \in A$, $t \in T$ it holds that $X_t(\omega) = Y_t(\omega)$.

If the processes X and Y are indistinguishable, then they are modifications of each other. The next example shows that the converse is not true, in general.

EXAMPLE 1.5.5. Let U be a random variable which is uniformly distributed on the interval $[0, 1]$. The probability space on which U is defined is denoted by $(\Omega, \mathcal{F}, \mathbb{P})$. Define two stochastic processes $\{X_t : t \in [0, 1]\}$ and $\{Y_t : t \in [0, 1]\}$ by

1. $X_t(\omega) = 0$ for all $t \in [0, 1]$ and $\omega \in \Omega$.
2. For all $t \in [0, 1]$ and $\omega \in \Omega$,

$$Y_t(\omega) = \begin{cases} 1, & \text{if } t = U(\omega), \\ 0, & \text{otherwise.} \end{cases}$$

Then,

(a) X is a modification of Y because for all $t \in [0, 1]$ it holds that

$$\mathbb{P}[X_t = Y_t] = \mathbb{P}[Y_t = 0] = \mathbb{P}[U \neq t] = 1.$$

(b) X and Y are *not* indistinguishable because for every $\omega \in \Omega$ the sample paths $t \mapsto X_t(\omega)$ and $t \mapsto Y_t(\omega)$ are not equal as functions on T. Namely, $Y_{U(\omega)}(\omega) = 1$ while $X_{U(\omega)}(\omega) = 0$.

PROPOSITION 1.5.6. *Let $\{X_t : t \in T\}$ and $\{Y_t : t \in T\}$ be two stochastic processes defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and having the same index set T. Consider the following statements:*

1. *X and Y are indistinguishable.*
2. *X and Y are modifications of each other.*
3. *X and Y have the same finite-dimensional distributions.*

Then, $1 \Rightarrow 2 \Rightarrow 3$ and none of the implications can be inverted, in general.

PROOF. Exercise. □

EXERCISE 1.5.7. Let $\{X_t : t \in T\}$ and $\{Y_t : t \in T\}$ be two stochastic processes defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and having the same *countable* index set T. Show that X and Y are indistinguishable if and only if they are modifications of each other.

1.6. Measurability of subsets of $\mathbb{R}^T$

Let $\{X_t : t \in T\}$ be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. To every outcome $\omega \in \Omega$ we can associate a trajectory of the process which is the function $t \mapsto X_t(\omega)$. Suppose we would like to compute the probability that the trajectory is everywhere equal to zero. That is, we would like to determine the probability of the set

$$Z := \{\omega \in \Omega : X_t(\omega) = 0 \text{ for all } t \in T\} = \cap_{t\in T}\{\omega \in \Omega : X_t(\omega) = 0\} = \cap_{t\in T} X_t^{-1}(0).$$

But first we need to figure out whether Z is a measurable set, that is whether $Z \in \mathcal{F}$. If T is countable, then Z is measurable since any of the sets $X_t^{-1}(0)$ is measurable (because X_t is a measurable function) and a countable intersection of measurable sets is measurable. However, if the index set T is not countable (for example $T = \mathbb{R}$), then the set Z may be non-measurable, as the next example shows.

EXAMPLE 1.6.1. We will construct a stochastic process $\{X_t : t \in \mathbb{R}\}$ for which the set Z is not measurable. As in the proof of Kolmogorov's theorem, our stochastic process will be defined on the "canonical" probability space $\Omega = \mathbb{R}^{\mathbb{R}} = \{f : \mathbb{R} \to \mathbb{R}\}$, with $\mathcal{F} = \sigma_{cyl}$ being the cylinder σ-algebra. Let $X_t : \mathbb{R}^{\mathbb{R}} \to \mathbb{R}$ be defined as the canonical coordinate mappings: $X_t(f) = f(t)$, $f \in \mathbb{R}^{\mathbb{R}}$. Then, the set Z consists of just one element, the function which is identically 0.
We show that Z does not belong to the cylinder σ-algebra. Let us call a set $A \subset \mathbb{R}^{\mathbb{R}}$ countably generated if one can find $t_1, t_2, \ldots \in \mathbb{R}$ and a set $B \subset \mathbb{R}^{\mathbb{N}}$ such that

$$f \in A \quad \Leftrightarrow \quad \{i \mapsto f(t_i)\} \in \mathbb{R}^{\mathbb{N}}. \tag{1.6.1}$$

With other words, a set A is countably generated if we can determine whether a given function $f : \mathbb{R} \to \mathbb{R}$ belongs to this set just by looking at the values of f at a countable number of points $t_1, t_2, \ldots$ and checking whether these values have some property represented by the set B.
One can easily check that the countably generated sets form a σ-algebra (called σ_{cg}) and that the cylinder sets belong to this σ-algebra. Since the cylinder σ-algebra is the *minimal* σ-algebra containing all cylinder sets, we have $\sigma_{cyl} \subset \sigma_{cg}$.
Let us now take some (nonempty) set $A \in \sigma_{cyl}$. Then, $A \in \sigma_{cg}$. Let us show that A is infinite. Indeed, since A is non-empty, it contains at least one element $f \in A$. We will show that it is possible to construct infinitely many modifications of f (called f_a, $a \in \mathbb{R}$) which are still contained in A. Since A is countably generated we can find $t_1, t_2, \ldots \in \mathbb{R}$ and a set $B \subset \mathbb{R}^{\mathbb{N}}$ such that (1.6.1) holds. Since the sequence $t_1, t_2, \ldots$ is countable while $\mathbb{R}$ is not, we can find $t_0 \in \mathbb{R}$ such that t_0 is not a member of the sequence $t_1, t_2, \ldots$. For every $a \in \mathbb{R}$ let $f_a : \mathbb{R} \to \mathbb{R}$ be the function given by

$$f_a(t) = \begin{cases} a, & \text{if } t = t_0, \\ f(t), & \text{if } t \neq t_0. \end{cases}$$

The function f_a belongs to A because f belongs to A and the functions $i \mapsto f(t_i)$, $i \in \mathbb{N}$, and $i \mapsto f_a(t_i)$, $i \in \mathbb{N}$, coincide; see (1.6.1). Hence, the set A contains infinitely many elements, namely f_a, $a \in \mathbb{R}$. In particular, the set A cannot contain exactly one element. It follows that the set Z (which contains exactly one element) does not belong to the cylinder σ-algebra.

EXERCISE 1.6.2. Show that the following subsets of $\mathbb{R}^{\mathbb{R}}$ do *not* belong to the cylinder σ-algebra:

(1) $C = \{f : \mathbb{R} \to \mathbb{R} : f \text{ is continuous}\}$.
(2) $B = \{f : \mathbb{R} \to \mathbb{R} : f \text{ is bounded}\}$.
(3) $M = \{f : \mathbb{R} \to \mathbb{R} : f \text{ is monotone increasing}\}$.

1.7. Continuity of stochastic processes

There are several non-equivalent notions of continuity for stochastic processes. Let $\{X_t \colon t \in \mathbb{R}\}$ be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For concreteness we take the index set to be $T = \mathbb{R}$, but everything can be generalized to the case when $T = \mathbb{R}^d$ or T is any metric space.

DEFINITION 1.7.1. We say that the process X has *continuous sample paths* if for all $\omega \in \Omega$ the function $t \mapsto X_t(\omega)$ is continuous in t.

So, the process X has continuous sample paths if every sample path of this process is a continuous function.

DEFINITION 1.7.2. We say that the process X has *almost surely continuous sample paths* if there exists a set $A \in \mathcal{F}$ such that $\mathbb{P}[A] = 1$ and for all $\omega \in A$ the function $t \mapsto X_t(\omega)$ is continuous in t.

Note that we *do not* state this definition in the form

$$\mathbb{P}[\omega \in \Omega \colon \text{the function } t \mapsto X_t(\omega) \text{ is continuous in } t] = 1$$

because the corresponding set need not be measurable; see Section 1.6.

DEFINITION 1.7.3. We say that the process X is *stochastically continuous* or *continuous in probability* if for all $t \in \mathbb{R}$ it holds that

$$X_s \xrightarrow{P} X_t \text{ as } s \to t.$$

That is,

$$\forall t \in \mathbb{R}\ \forall \varepsilon > 0 : \lim_{s \to t} \mathbb{P}[|X_t - X_s| > \varepsilon] = 0.$$

DEFINITION 1.7.4. We say that the process X is *continuous in L^p*, where $p \geq 1$, if for all $t \in \mathbb{R}$ it holds that

$$X_s \xrightarrow{L^p} X_t \text{ as } s \to t.$$

That is,

$$\forall t \in \mathbb{R} : \lim_{s \to t} \mathbb{E}|X_t - X_s|^p = 0.$$

EXAMPLE 1.7.5. Let U be a random variable which has continuous distribution function F. For concreteness, one can take the uniform distribution on $[0, 1]$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space on which U is defined. Consider a stochastic process $\{X_t : t \in \mathbb{R}\}$ defined as follows: For all $t \in \mathbb{R}$ and $\omega \in \Omega$ let

$$X_t(\omega) = \begin{cases} 1, & \text{if } t > U(\omega), \\ 0, & \text{if } t \leq U(\omega). \end{cases}$$

1. For every outcome $\omega \in \Omega$ the trajectory $t \mapsto X_t(\omega)$ is discontinuous because it has a jump at $t = U(\omega)$. Thus, the process X does not have continuous sample paths.

2. However, we will show that the process X is continuous in probability. Take some $\varepsilon \in (0,1)$. Then, for any $t, s \in [0,1]$,

$$\mathbb{P}[|X_t - X_s| > \varepsilon] = \mathbb{P}[|X_t - X_s| = 1] = \mathbb{P}[U \text{ is between } t \text{ and } s] = |F(t) - F(s)|,$$

which converges to 0 as $s \to t$ because the distribution function F was supposed to be continuous. Hence, the process X is continuous in probability.

3. We show that X is continuous in L^p, for every $p \geq 1$. Since the random variable $|X_t - X_s|$ takes only values 0 and 1 and since the probability of the value 1 is $|F(t) - F(s)|$, we have

$$\mathbb{E}|X_t - X_s|^p = |F(t) - F(s)|,$$

which goes to 0 as $s \to t$.

EXERCISE 1.7.6. Show that if a process $\{X(t) \colon t \in \mathbb{R}\}$ has continuous sample paths, the it is stochastically continuous. (The converse is not true by Example 1.7.5).

We have seen in Section 1.6 that for general stochastic processes some very natural events (for example, the event that the trajectory is everywhere equal to 0) may be non-measurable. This nasty problem disappears if we are dealing with processes having continuous sample paths.

EXAMPLE 1.7.7. Let $\{X_t, t \in \mathbb{R}\}$ be a process with continuous sample paths. We show that the set

$$A := \{\omega \in \Omega \colon X_t(\omega) = 0 \text{ for all } t \in \mathbb{R}\}$$

is measurable. A continuous function is equal to 0 for all $t \in \mathbb{R}$ if and only if it is equal to 0 for all $t \in \mathbb{Q}$. Hence, we can write

$$A = \{\omega \in \Omega \colon X_t(\omega) = 0 \text{ for all } t \in \mathbb{Q}\} = \cap_{t \in \mathbb{Q}} \{\omega \in \Omega \colon X_t(\omega) = 0\} = \cap_{t \in \mathbb{Q}} X_t^{-1}(0)$$

which is a measurable set because $X_t^{-1}(0) \in \mathcal{F}$ for every t (since $X_t : \Omega \to \mathbb{R}$ is a measurable function) and because the intersection over $t \in \mathbb{Q}$ is *countable*.

EXERCISE 1.7.8. Let $\{X \colon t \in \mathbb{R}\}$ be a stochastic process with continuous sample paths. The probability space on which X is defined is denoted by $(\Omega, \mathcal{F}, \mathbb{P})$. Show that the following subsets of Ω belong to the σ-algebra $\mathcal{F}$:

(1) $B = \{\omega \in \Omega : \text{the function } t \mapsto X_t(\omega) \text{ is bounded}\}$.
(2) $M = \{\omega \in \Omega : \text{the function } t \mapsto X_t(\omega) \text{ is monotone increasing}\}$
(3) $I = \{\omega \in \Omega : \lim_{t \to +\infty} X_t(\omega) = +\infty\}$.

CHAPTER 2

Markov chains

2.1. Examples

EXAMPLE 2.1.1 (Markov chain with two states). Consider a phone which can be in two states: "free"= 0 and "busy"= 1. The set of the states of the phone is

$$E = \{0, 1\}.$$

We assume that the phone can randomly change its state in time (which is assumed to be discrete) according to the following rules.

1. If at some time n the phone is free, then at time $n + 1$ it becomes busy with probability p or it stays free with probability $1 - p$.

2. If at some time n the phone is busy, then at time $n + 1$ it becomes free with probability q or it stays busy with probability $1 - q$.

Denote by X_n the state of the phone at time $n = 0, 1, \ldots$. Thus, $X_n : \Omega \to \{0, 1\}$ is a random variable and our assumptions can be written as follows:

$$p_{00} := \mathbb{P}[X_{n+1} = 0 | X_n = 0] = 1 - p, \qquad p_{01} := \mathbb{P}[X_{n+1} = 1 | X_n = 0] = p,$$
$$p_{10} := \mathbb{P}[X_{n+1} = 0 | X_n = 1] = q, \qquad p_{11} := \mathbb{P}[X_{n+1} = 1 | X_n = 1] = 1 - q.$$

We can write these probabilities in form of a *transition matrix*

$$P = \begin{pmatrix} 1 - p & p \\ q & 1 - q \end{pmatrix}.$$

Additionally, we will make the following assumption which is called the *Markov property*: Given that at some time n the phone is in state $i \in \{0, 1\}$, the behavior of the phone after time n does not depend on the way the phone reached state i in the past.

PROBLEM 2.1.2. Suppose that at time 0 the phone was free. What is the probability that the phone will be free at times $1, 2$ and then becomes busy at time 3?

SOLUTION. This probability can be computed as follows:

$$\mathbb{P}[X_1 = X_2 = 0, X_3 = 1] = p_{00} \cdot p_{00} \cdot p_{01} = (1 - p)^2 p.$$

PROBLEM 2.1.3. Suppose that the phone was free at time 0. What is the probability that it will be busy at time 3?

SOLUTION. We have to compute $\mathbb{P}[X_3 = 1]$. We know the values $X_0 = 0$ and $X_3 = 1$, but the values of X_1 and X_2 may be arbitrary. We have the following possibilities:

(1) $X_0 = 0, X_1 = 0, X_2 = 0, X_3 = 1$. Probability: $(1 - p) \cdot (1 - p) \cdot p$.
(2) $X_0 = 0, X_1 = 0, X_2 = 1, X_3 = 1$. Probability: $(1 - p) \cdot p \cdot (1 - q)$.

(3) $X_0 = 0, X_1 = 1, X_2 = 0, X_3 = 1$. Probability: $p \cdot q \cdot p$.
(4) $X_0 = 0, X_1 = 1, X_2 = 1, X_3 = 1$. Probability: $p \cdot (1-q) \cdot (1-q)$.

The probability we look for is the sum of these 4 probabilities:

$$\mathbb{P}[X_3 = 1] = (1-p)^2 p + (1-p)(1-q)p + p^2 q + p(1-q)^2.$$

EXAMPLE 2.1.4 (Gambler's ruin). At each unit of time a gambler plays a game in which he can either win 1€ (which happens with probability p) or he can loose 1€ (which happens with probability $1-p$). Let X_n be the capital of the gambler at time n. Let us agree that if at some time n the gambler has no money (meaning that $X_n = 0$), then he stops to play (meaning that $X_n = X_{n+1} = \ldots = 0$). We can view this process as a Markov chain on the state space $E = \{0, 1, 2, \ldots\}$ with transition matrix

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \ldots \\ 1-p & 0 & p & 0 & 0 & \ldots \\ 0 & 1-p & 0 & p & 0 & \ldots \\ 0 & 0 & 1-p & 0 & p & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{pmatrix}.$$

2.2. Definition of Markov chains

Let us consider some system. Assume that the system can be in some states and that the system can change its state in time. The set of all states of the system will be denoted by E and called the *state space* of the Markov chain. We always assume that the state space E is a finite or countable set. Usually, we will denote the states so that $E = \{1, \ldots, N\}$, $E = \mathbb{N}$, or $E = \mathbb{Z}$.

Assume that if at some time the system is in state $i \in E$, then in the next moment of time it can switch to state $j \in E$ with probability p_{ij}. We will call p_{ij} the *transition probability* from state i to state j. Clearly, the transition probabilities should be such that

(1) $p_{ij} \geq 0$ for all $i, j \in E$.
(2) $\sum_{j \in E} p_{ij} = 1$ for all $i \in E$.

We will write the transition probabilities in form of a *transition matrix*

$$P = (p_{ij})_{i,j \in E}.$$

The rows and the columns of this matrix are indexed by the set E. The element in the i-th row and j-th column is the transition probability p_{ij}. The elements of the matrix P are non-negative and the sum of elements in any row is equal to 1. Such matrices are called *stochastic*.

DEFINITION 2.2.1. A *Markov chain* with state space E and transition matrix P is a stochastic process $\{X_n \colon n \in \mathbb{N}_0\}$ taking values in E such that for every $n \in \mathbb{N}_0$ and every states $i_0, i_1, \ldots, i_{n-1}, i, j$ we have

$$\begin{aligned} \mathbb{P}[X_{n+1} = j | X_n = i] &= \mathbb{P}[X_{n+1} = j | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i] \\ &= p_{ij}, \end{aligned} \tag{2.2.1}$$

provided that $\mathbb{P}[X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i] \neq 0$ (which ensures that the conditional probabilities are well-defined).

Condition (2.2.1) is called the *Markov property*.

In the above definition it is not specified at which state the Markov chain starts at time 0. In fact, the initial state can be in general arbitrary and we call the probabilities

$$\alpha_i := \mathbb{P}[X_0 = i], \quad i \in E, \tag{2.2.2}$$

the *initial probabilities*. We will write the initial probabilities in form of a row vector $\alpha = (\alpha_i)_{i\in E}$. This vector should be such that $\alpha_i \geq 0$ for all $i \in E$ and $\sum_{i\in E} \alpha_i = 1$.

THEOREM 2.2.2. *For all $n \in \mathbb{N}_0$ and for all $i_0, \ldots, i_n \in E$ it holds that*

$$\mathbb{P}[X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n] = \alpha_0 p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n}. \tag{2.2.3}$$

PROOF. We use the induction over n. The induction basis is the case $n = 0$. We have $\mathbb{P}[X_0 = i_0] = \alpha_{i_0}$ by the definition of initial probabilities, see (2.2.2). Hence, Equation (2.2.3) holds for $n = 0$.

Induction assumption: Assume that (2.2.3) holds for some n. We prove that (2.2.3) holds with n replaced by $n + 1$. Consider the event $A = \{X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\}$. By the induction assumption,

$$\mathbb{P}[A] = \alpha_{i_0} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n}.$$

By the Markov property,

$$\mathbb{P}[X_{n+1} = i_{n+1} | A] = p_{i_n i_{n+1}}.$$

It follows that

$$\begin{aligned} \mathbb{P}[X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n, X_{n+1} = i_{n+1}] &= \mathbb{P}[X_{n+1} = i_{n+1} | A] \cdot \mathbb{P}[A] \\ &= p_{i_n i_{n+1}} \cdot \alpha_{i_0} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n} \\ &= \alpha_{i_0} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n} p_{i_n i_{n+1}}. \end{aligned}$$

This completes the induction. □

REMARK 2.2.3. If $\mathbb{P}[A] = 0$, then in the above proof we cannot use the Markov property. However, in the case $\mathbb{P}[A] = 0$ both sides of (2.2.3) are equal to 0 and (2.2.3) is trivially satisfied.

THEOREM 2.2.4. *For every $n \in \mathbb{N}$ and every state $i_n \in E$ we have*

$$\mathbb{P}[X_n = i_n] = \sum_{i_0, \ldots, i_{n-1} \in E} \alpha_{i_0} p_{i_0 i_1} \ldots p_{i_{n-1} i_n}.$$

PROOF. We have

$$\begin{aligned} \mathbb{P}[X_n = i_n] &= \sum_{i_0, \ldots, i_{n-1} \in E} \mathbb{P}[X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n] \\ &= \sum_{i_0, \ldots, i_{n-1} \in E} \alpha_{i_0} p_{i_0 i_1} \ldots p_{i_{n-1} i_n}, \end{aligned}$$

where the last step is by Theorem 2.2.2. □

2.3. n-step transition probabilities

NOTATION 2.3.1. If we want to indicate that the Markov chain starts at state $i \in E$ at time 0, we will write $\mathbb{P}_i$ instead of $\mathbb{P}$.

DEFINITION 2.3.2. The n-step transition probabilities of a Markov chain are defined as

$$p_{ij}^{(n)} := \mathbb{P}_i[X_n = j].$$

We will write these probabilities in form of the *n-step transition matrix* $P^{(n)} = (p_{ij}^{(n)})_{i,j\in E}$.

By Theorem 2.2.4 we have the formula

$$p_{ij}^{(n)} = \sum_{i_1,\ldots,i_{n-1}\in E} p_{ii_1}p_{i_1i_2}\ldots p_{i_{n-1}j}.$$

The next theorem is crucial. It states that the n-step transition matrix $P^{(n)}$ can be computed as the n-th power of the transition matrix P.

THEOREM 2.3.3. *We have $P^{(n)} = P^n = P \cdot \ldots \cdot P$.*

PROOF. We use induction over n. For $n = 1$ we have $p_{ij}^{(1)} = p_{ij}$ and hence, $P^{(1)} = P$. Thus, the statement of the theorem is true for $n = 1$.

Let us now assume that we already proved that $P^{(n)} = P^n$ for some $n \in \mathbb{N}$. We compute $P^{(n+1)}$. By the formula of total probability, we have

$$p_{ij}^{(n+1)} = \mathbb{P}_i[X_{n+1} = j] = \sum_{k\in E} \mathbb{P}_i[X_n = k]\mathbb{P}[X_{n+1} = j|X_n = k] = \sum_{k\in E} p_{ik}^{(n)}p_{kj}.$$

On the right hand-side we have the scalar product of the i-th row of the matrix $P^{(n)}$ and the j-th column of the matrix P. By definition of the matrix multiplication, this scalar product is exactly the entry of the matrix product $P^{(n)}P$ which is located in the i-th row and j-th column. We thus have the equality of matrices

$$P^{(n+1)} = P^{(n)}P.$$

But now we can apply the induction assumption $P^{(n)} = P^n$ to obtain

$$P^{(n+1)} = P^{(n)}P = P^n \cdot P = P^{n+1}.$$

This completes the induction. □

In the next theorem we consider a Markov chain with initial distribution $\alpha = (\alpha_i)_{i\in E}$ and transition matrix P. Let $\alpha^{(n)} = (\alpha_j^{(n)})_{j\in E}$ be the distribution of the position of this chain at time n, that is

$$\alpha_j^{(n)} = \mathbb{P}[X_n = j].$$

We write both $\alpha^{(n)}$ and α as *row* vectors. The next theorem states that we can compute $\alpha^{(n)}$ by taking α and multiplying it by the n-step transition matrix $P^{(n)} = P^n$ from the right.

THEOREM 2.3.4. *We have*

$$\alpha^{(n)} = \alpha P^n.$$

PROOF. By the formula of the total probability

$$\alpha_j^{(n)} = \mathbb{P}[X_n = j] = \sum_{i \in E} \alpha_i \mathbb{P}_i[X_n = j] = \sum_{i \in E} \alpha_i p_{ij}^{(n)}.$$

On the right-hand side we have the scalar product of the row α with the j-th column of $P^{(n)} = P^n$. By definition of matrix multiplication, this means that $\alpha^{(n)} = \alpha P^n$. □

2.4. Invariant measures

Consider a Markov chain on state space E with transition matrix P. Let $\lambda : E \to \mathbb{R}$ be a function. To every state $i \in E$ the function assigns some value which will be denoted by $\lambda_i := \lambda(i)$. Also, it will be convenient to write the function λ as a row vector $\lambda = (\lambda_i)_{i \in E}$.

DEFINITION 2.4.1. A function $\lambda : E \to \mathbb{R}$ is called a *measure* on E if $\lambda_i \geq 0$ for all $i \in E$.

DEFINITION 2.4.2. A function $\lambda : E \to \mathbb{R}$ is called a *probability measure* on E if $\lambda_i \geq 0$ for all $i \in E$ and

$$\sum_{i \in E} \lambda_i = 1.$$

DEFINITION 2.4.3. A measure λ is called *invariant* if $\lambda P = \lambda$. That is, for every state $j \in E$ it should hold that

$$\lambda_j = \sum_{i \in E} \lambda_i p_{ij}.$$

REMARK 2.4.4. If the initial distribution α of a Markov chain is invariant, that is $\alpha P = \alpha$, then for every $n \in \mathbb{N}$ we have $\alpha P^n = \alpha$ which means that at every time n the position of the Markov chain has the same distribution as at time 0:

$$X_0 \stackrel{d}{=} X_1 \stackrel{d}{=} X_2 \stackrel{d}{=} \dots.$$

EXAMPLE 2.4.5. Let us compute the invariant distribution for the Markov chain from Example 2.1.1. The transition matrix is

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}.$$

The equation $\lambda P = \lambda$ for the invariant probability measure takes the following form:

$$(\lambda_0, \lambda_1) \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} = (\lambda_0, \lambda_1).$$

Multiplying the matrices we obtain the following two equations:

$$\lambda_0(1-p) + \lambda_1 q = \lambda_0,$$
$$\lambda_0 p + \lambda_1(1-q) = \lambda_1.$$

From the first equation we obtain that $\lambda_1 q = \lambda_0 p$. Solving the second equation we obtain the same relation which means that the second equation does not contain any information not contained in the first equation. However, since we are looking for invariant *probability* measures, we have an additional equation

$$\lambda_0 + \lambda_1 = 1.$$

Solving this equation together with $\lambda_1 q = \lambda_0 p$ we obtain the following result:

$$\lambda_0 = \frac{q}{p+q}, \quad \lambda_1 = \frac{p}{p+q}.$$

PROBLEM 2.4.6. Consider the phone from Example 2.1.1. Let the phone be free at time 0. What is (approximately) the probability that it is free at time $n = 1000$?

SOLUTION. The number $n = 1000$ is large. For this reason it seems plausible that the probability that the phone is free (busy) at time $n = 1000$ should be approximately the same as the probability that it is free (busy) at time $n + 1 = 1001$. Denoting the initial distribution by $\alpha = (1, 0)$ and the distribution of the position of the chain at time n by $\alpha^{(n)} = \alpha P^n$ we thus must have

$$\alpha^{(n)} \approx \alpha^{(n+1)} = \alpha P^{n+1} = \alpha P^n \cdot P = \alpha^{(n)} P.$$

Recall that the equation for the invariant probability measure has the same form $\lambda = \lambda P$. It follows that $\alpha^{(n)}$ must be approximately the invariant probability measure:

$$\alpha^{(n)} \approx \lambda.$$

For the probability that the phone is free (busy) at time $n = 1000$ we therefore obtain the approximations

$$p_{00}^{(n)} \approx \lambda_0 = \frac{q}{p+q}, \quad p_{01}^{(n)} \approx \lambda_1 = \frac{p}{p+q}.$$

Similar considerations apply to the case when the phone is busy at time 0 leading to the approximations

$$p_{10}^{(n)} \approx \lambda_0 = \frac{q}{p+q}, \quad p_{11}^{(n)} \approx \lambda_1 = \frac{p}{p+q}.$$

Note that $p_{00}^{(n)} \approx p_{10}^{(n)}$ and $p_{01}^{(n)} \approx p_{11}^{(n)}$ which can be interpreted by saying that the Markov chain almost forgets its initial state after many steps. For the n-step transition matrix we therefore may conjecture that

$$\lim_{n\to\infty} P^n = \lim_{n\to\infty} \begin{pmatrix} p_{00}^{(n)} & p_{01}^{(n)} \\ p_{10}^{(n)} & p_{11}^{(n)} \end{pmatrix} = \begin{pmatrix} \lambda_0 & \lambda_1 \\ \lambda_0 & \lambda_1 \end{pmatrix}.$$

The above considerations are not rigorous. We will show below that if a general Markov chain satisfies appropriate conditions, then

(1) The invariant probability measure λ exists and is unique.
(2) For every states $i, j \in E$ we have $\lim_{n\to\infty} p_{ij}^{(n)} = \lambda_j$.

EXAMPLE 2.4.7 (Ehrenfest model). We consider a box which is divided into 2 parts. Consider N balls (molecules) which are located in this box and can move from one part to the other according to the following rules. Assume that at any moment of time one of the N balls is chosen at random (all balls having the same probability $1/N$ to be chosen). This ball moves to the other part. Then, the procedure is repeated. Let X_n be the number of balls at time n in Part 1. Then, X_n takes values in $E = \{0, 1, \ldots, N\}$ which is our state space. The transition probabilities are given by

$$p_{0,1} = 1, \quad p_{N,N-1} = 1, \quad p_{i,i+1} = \frac{N-i}{N}, \quad p_{i,i-1} = \frac{i}{N}, \quad i = 1, \ldots, N-1.$$

For the invariant probability measure we obtain the following system of equations

$$\lambda_0 = \frac{\lambda_1}{N}, \quad \lambda_N = \frac{\lambda_{N-1}}{N}, \quad \lambda_j = \frac{N-j+1}{N}\,\lambda_{j-1} + \frac{j+1}{N}\,\lambda_{j+1}, \quad j = 1, \dots, N-1.$$

Additionally, we have the equation $\lambda_0 + \ldots + \lambda_N = 1$. This system of equations can be solved directly, but one can also guess the solution without doing computations. Namely, it seems plausible that after a large number of steps every ball will be with probability $1/2$ in Part 1 and with probability $1/2$ in Part 2. Hence, one can guess that the invariant probability measure is the binomial distribution with parameter $1/2$:

$$\lambda_j = \frac{1}{2^N}\binom{N}{j}.$$

One can check that this is indeed the unique invariant probability measure for this Markov chain.

EXAMPLE 2.4.8. Let $X_0, X_1, \dots$ be independent and identically distributed random variables with values $1, \dots, N$ and corresponding probabilities

$$\mathbb{P}[X_n = i] = p_i, \quad p_1, \dots, p_N \geq 0, \quad \sum_{i=1}^{N} p_i = 1.$$

Then, $X_0, X_1, \dots$ is a Markov chain and the transition matrix is

$$P = \begin{pmatrix} p_1 & \dots & p_N \\ \dots & \dots & \dots \\ p_1 & \dots & p_N \end{pmatrix}.$$

The invariant probability measure is given by $\lambda_1 = p_1, \dots, \lambda_N = p_N$.

2.5. Class structure and irreducibility

Consider a Markov chain on a state space E with transition matrix P.

DEFINITION 2.5.1. We say that state $i \in E$ *leads* to state $j \in E$ if there exists $n \in \mathbb{N}_0$ such that $p_{ij}^{(n)} \neq 0$. We use the notation $i \rightsquigarrow j$.

REMARK 2.5.2. By convention, $p_{ii}^{(0)} = 1$ and hence, every state leads to itself: $i \rightsquigarrow i$.

THEOREM 2.5.3. *For two states $i, j \in E$ with $i \neq j$, the following statements are equivalent:*

(1) $i \rightsquigarrow j$.
(2) $\mathbb{P}_i[\exists n \in \mathbb{N}\colon X_n = j] \neq 0$.
(3) *There exist $n \in \mathbb{N}$ and states $i_1, \dots, i_{n-1} \in E$ such that $p_{ii_1} \dots p_{i_{n-1}j} > 0$.*

PROOF. We prove that Statements 1 and 2 are equivalent. We have the inequality

$$p_{ij}^{(n)} \leq \mathbb{P}_i[\exists n \in \mathbb{N} : X_n = j] \leq \sum_{n=1}^{\infty} \mathbb{P}_i[X_n = j] = \sum_{n=1}^{\infty} p_{ij}^{(n)}. \tag{2.5.1}$$

If Statement 1 holds, then for some $n \in \mathbb{N}$ we have $p_{ij}^{(n)} > 0$. Hence, by (2.5.1), we have $\mathbb{P}_i[\exists n \in \mathbb{N} : X_n = j] > 0$ and Statement 2 holds. If, conversely, Statement 2 holds, then

$\mathbb{P}_i[\exists n \in \mathbb{N} : X_n = j] > 0$. Hence, by (2.5.1), $\sum_{n=1}^{\infty} p_{ij}^{(n)} > 0$, which implies that at least one summand $p_{ij}^{(n)}$ must be strictly positive. This proves Statement 1.

We prove the equivalence of Statements 1 and 3. We have the formula

$$p_{ij}^{(n)} = \sum_{i_1,\ldots,i_{n-1} \in E} p_{ii_1} \ldots p_{i_{n-1}j}. \tag{2.5.2}$$

If Statement 1 holds, then for some $n \in \mathbb{N}$ we have $p_{ij}^{(n)} > 0$ which implies that at least one summand on the right-hand side of (2.5.2) must be strictly positive. This implies Statement 3. If, conversely, Statement 3 holds, then the sum on the right-hand side of (2.5.2) is positive which implies that $p_{ij}^{(n)} > 0$. Hence, Statement 1 holds. □

Definition 2.5.4. States $i, j \in E$ *communicate* if $i \rightsquigarrow j$ and $j \rightsquigarrow i$. Notation: $i \leftrightsquigarrow j$.

Theorem 2.5.5. *$i \leftrightsquigarrow j$ is an equivalence relation, namely*

(1) $i \leftrightsquigarrow i$.
(2) $i \leftrightsquigarrow j \Longleftrightarrow j \leftrightsquigarrow i$.
(3) $i \leftrightsquigarrow j, j \leftrightsquigarrow k \Rightarrow i \leftrightsquigarrow k$.

Proof. Statements 1 and 2 follow from the definition. We prove Statement 3. If $i \leftrightsquigarrow j$ and $j \leftrightsquigarrow k$, then, in particular, $i \rightsquigarrow j$ and $j \rightsquigarrow k$. By Theorem 2.5.3, Statement 3, we can find $r \in \mathbb{N}$, $s \in \mathbb{N}$ and states $u_1, \ldots, u_{r-1} \in E$ and $v_1, \ldots, v_{s-1} \in E$ such that $p_{iu_1} p_{u_1u_2} \ldots p_{u_{r-1}j} > 0$ and $p_{jv_1} p_{v_1v_2} \ldots p_{v_{s-1}k} > 0$. Multiplying both inequalities, we get

$$p_{iu_1} p_{u_1u_2} \ldots p_{u_{r-1}j} p_{jv_1} p_{v_1v_2} \ldots p_{v_{s-1}k} > 0.$$

By Theorem 2.5.3, Statement 3, we have $i \rightsquigarrow k$. In a similar way one shows that $k \rightsquigarrow i$. □

Definition 2.5.6. The *communication class* of state $i \in E$ is the set $\{j \in E : i \leftrightsquigarrow j\}$. This set consists of all states j which communicate to i.

Since communication of states is an equivalence relation, the state space E can be decomposed into a disjoint union of communication classes. Any two communication classes either coincide completely or are disjoint sets.

Definition 2.5.7. A Markov chain is *irreducible* if every two states communicate. Hence, an irreducible Markov chain consists of just one communication class.

Definition 2.5.8. A communication class C is *open* if there exist a state $i \in C$ and a state $k \notin C$ such that $i \rightsquigarrow k$. Otherwise, a communication class is called *closed*.

If a Markov chain once arrived in a closed communication class, it will stay in this class forever.

Exercise 2.5.9. Show that a communication class C is open if and only if there exist a state $i \in C$ and a state $k \notin C$ such that $p_{ik} > 0$.

Theorem 2.5.10. *If the state space E is a finite set, then there exists at least one closed communication class.*

PROOF. We use a proof by contradiction. Assume that there is no closed communication class. Hence, all communication classes are open. Take some state and let C_1 be the communication class of this state. Since C_1 is open, there is a path from C_1 to some other communication class $C_2 \neq C_1$. Since C_2 is open, we can go from C_2 to some other communication class $C_3 \neq C_3$, and so on. Note that in the sequence $C_1, C_2, C_3, \ldots$ all classes are different. Indeed, if for some $l < m$ we would have $C_l = C_m$ (a "cycle"), this would mean that there is a path starting from C_l, going to C_{l+1} and then to $C_m = C_l$. But this is a contradiction since then C_l and C_{l+1} should be a single communication class, and not two different classes, as in the construction. So, the classes $C_1, C_2, \ldots$ are different (in fact, disjoint) and each class contains at least one element. But this is a contradiction since E is a finite set. □

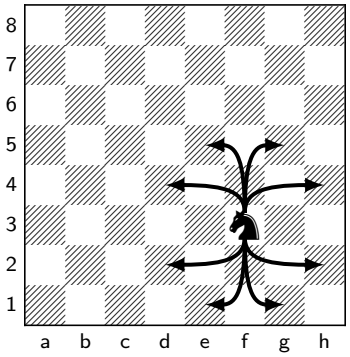
2.6. Aperiodicity

DEFINITION 2.6.1. The *period* of a state $i \in E$ is defined as

$$\gcd\{n \in \mathbb{N} : p_{ii}^{(n)} > 0\}.$$

Here, gcd states for the greatest common divisor. A state $i \in E$ is called *aperiodic* if its period is equal to 1. Otherwise, the state i is called *periodic*.

EXAMPLE 2.6.2. Consider a knight on a chessboard moving according to the usual chess rules in a random way. For concreteness, assume that at each moment of time all moves of the knight allowed by the chess rules are counted and then one of these moves is chosen, all moves being equiprobable.



This is a Markov chain on a state space consisting of 64 squares. Assume that at time 0 the knight is in square i. Since the knight changes the color of its square after every move, it cannot return to the original square in an odd number of steps. On the other hand, it can return to i in an even number of steps with non-zero probability (for example by going to some other square and then back, many times). So,

$$p_{ii}^{(2n+1)} = 0, \quad p_{ii}^{(2n)} > 0.$$

Hence, the period of any state in this Markov chain is 2.

EXAMPLE 2.6.3. Consider a Markov chain on a state space of two elements with transition matrix

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

We have

$$p_{ii}^{(2n+1)} = 0, \quad p_{ii}^{(2n)} = 1.$$

Hence, the period of any state in this Markov chain is 2.

EXERCISE 2.6.4. Show that in the Ehrenfest Markov chain (Example 2.4.7) every state is periodic with period 2.

LEMMA 2.6.5. *Let $i \in E$ be any state. The following conditions are equivalent:*

(1) *State i is aperiodic.*
(2) *There is $N \in \mathbb{N}$ such that for every natural number $n > N$ we have $p_{ii}^{(n)} > 0$.*

PROOF. If Statement 2 holds, then for some sufficiently large n we have $p_{ii}^{(n)} > 0$ and $p_{ii}^{(n+1)} > 0$. Since $\gcd(n, n+1) = 1$, the state i has period 1. Hence, Statement 1 holds.

Suppose, conversely, that Statement 1 holds. Then, we can find $n_1, \ldots, n_r \in \mathbb{N}$ such that $\gcd\{n_1, \ldots, n_r\} = 1$ and $p_{ii}^{(n_1)} > 0, \ldots, p_{ii}^{(n_r)} > 0$. By a result from number theory, the condition $\gcd\{n_1, \ldots, n_r\} = 1$ implies that there is $N \in \mathbb{N}$ such that we can represent any natural number $n > N$ in the form $n = l_1 n_1 + \ldots + l_r n_r$ for suitable $l_1, \ldots, l_r \in \mathbb{N}$. We obtain that

$$p_{ii}^{(l_1 n_1 + \ldots + l_r n_r)} \geq (p_{ii}^{(n_1)})^{l_1} \cdot \ldots \cdot (p_{ii}^{(n_r)})^{l_r} > 0.$$

This proves Statement 2.

LEMMA 2.6.6. *If state $i \in E$ is aperiodic and $i \leftrightsquigarrow j$, then j is also aperiodic.*

REMARK 2.6.7. We can express this by saying that aperiodicity is a *class property*: If some state in a communication class is aperiodic, then all states in this communication class are aperiodic. Similarly, if some state in a communication class is periodic, then all states in this communication class must be periodic. We can thus divide all communication classes into two categories: the aperiodic communication classes (consisting of only aperiodic states) and the periodic communication classes (consisting only of periodic states).

DEFINITION 2.6.8. An irreducible Markov chain is called aperiodic if some (and hence, all) states in this chain are aperiodic.

PROOF OF LEMMA 2.6.6. From $i \leftrightsquigarrow j$ it follows that $i \rightsquigarrow j$ and $j \rightsquigarrow i$. Hence, we can find $r, s \in \mathbb{N}_0$ such that $p_{ji}^{(r)} > 0$ and $p_{ij}^{(s)} > 0$. Since the state i is aperiodic, by Lemma 2.6.5 we can find $N \in \mathbb{N}$ such that for all $n > N$, we have $p_{ii}^{(n)} > 0$ and hence,

$$p_{jj}^{(n+r+s)} \geq p_{ji}^{(r)} \cdot p_{ii}^{(n)} \cdot p_{ij}^{(s)} > 0.$$

It follows that $p_{jj}^{(k)} > 0$ for all $k := n + r + s > N + r + s$. By Lemma 2.6.5, this implies that j is aperiodic. □

2.7. Recurrence and transience

Consider a Markov chain $\{X_n : n \in \mathbb{N}_0\}$ on state space E with transition matrix P.

DEFINITION 2.7.1. A state $i \in E$ is called *recurrent* if

$$\mathbb{P}_i[X_n = i \text{ for infinitely many } n] = 1.$$

DEFINITION 2.7.2. A state $i \in E$ is called *transient* if

$$\mathbb{P}_i[X_n = i \text{ for infinitely many } n] = 0.$$

A recurrent state has the property that a Markov chain starting at this state returns to this state infinitely often, with probability 1. A transient state has the property that a Markov chain starting at this state returns to this state only finitely often, with probability 1.

The next theorem is a characterization of recurrent/transient states.

THEOREM 2.7.3. *Let $i \in E$ be a state. Denote by f_i the probability that a Markov chain which starts at i returns to i at least once, that is*

$$f_i = \mathbb{P}_i[\exists n \in \mathbb{N} : X_n = i].$$

Then,

(1) *The state i is recurrent if and only if $f_i = 1$.*
(2) *The state i is transient if and only if $f_i < 1$.*

COROLLARY 2.7.4. *Every state is either recurrent or transient.*

PROOF. For $k \in \mathbb{N}$ consider the random event

$$B_k = \{X_n = i \text{ for at least } k \text{ different values of } n \in \mathbb{N}\}.$$

Then, $\mathbb{P}_i[B_k] = f_i^k$. Also, $B_1 \supset B_2 \supset \ldots$. It follows that

$$\mathbb{P}_i[X_n = i \text{ for infinitely many } n] = \mathbb{P}_i[\cap_{k=1}^{\infty} B_k] = \lim_{k\to\infty} \mathbb{P}_i[B_k] = \lim_{k\to\infty} f_i^k = \begin{cases} 1, & \text{if } f_i = 1, \\ 0, & \text{if } f_i < 1. \end{cases}$$

It follows that state i is recurrent if $f_i = 1$ and transient if $f_i < 1$. □

Here is one more characterization of recurrence and transience.

THEOREM 2.7.5. *Let $i \in E$ be a state. Recall that $p_{ii}^{(n)} = \mathbb{P}_i[X_n = i]$ denotes the probability that a Markov chain which started at state i visits state i at time n. Then,*

(1) *The state i is recurrent if and only if $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$.*
(2) *The state i is transient if and only if $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$.*

PROOF. Let the Markov chain start at state i. Consider the random variable

$$V_i := \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n = i\}}$$

which counts the number of returns of the Markov chain to state i. Note that the random variable V_i can take the value $+\infty$. Then,

$$\mathbb{P}_i[V_i \geq k] = \mathbb{P}[B_k] = f_i^k, \quad k \in \mathbb{N}.$$

Thus, the expectation of V_i can be computed as follows:

$$\mathbb{E}_i[V_i] = \sum_{k=1}^{\infty} \mathbb{P}_i[V_i \geq k] = \sum_{k=1}^{\infty} f_i^k. \tag{2.7.1}$$

On the other hand,

$$\mathbb{E}_i[V_i] = \mathbb{E}_i \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n=i\}} = \sum_{n=1}^{\infty} \mathbb{E}_i \mathbb{1}_{\{X_n=i\}} = \sum_{n=1}^{\infty} p_{ii}^{(n)}. \tag{2.7.2}$$

Case 1. Assume that state i is recurrent. Then, $f_i = 1$ by Theorem 2.7.3. It follows that $\mathbb{E}_i[V_i] = \infty$ by (2.7.1). (In fact, $\mathbb{P}_i[V_i = +\infty] = 1$ since $\mathbb{P}[V_i \geq k] = 1$ for every $k \in \mathbb{N}$). Hence, $\sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$ by (2.7.2)

Case 2. Assume that state i is transient. Then, $f_i < 1$ by Theorem 2.7.3. Thus, $\mathbb{E}_i V_i < \infty$ by (2.7.1) and hence, $\sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$ by (2.7.2). □

The next theorem shows that recurrence and transience are class properties: If some state in a communicating class is recurrent (resp. transient), then all states in this class are recurrent (resp. transient).

Theorem 2.7.6.

1. *If $i \in E$ be a recurrent state and $j \leftrightsquigarrow i$, then j is also recurrent.*
2. *If $i \in E$ be a transient state and $j \leftrightsquigarrow i$, then j is also transient.*

Proof. It suffices to prove Part 2. Let i be a transient state and let $j \leftrightsquigarrow i$. It follows that there exist $s, r \in \mathbb{N}_0$ with $p_{ij}^{(s)} > 0$ and $p_{ji}^{(r)} > 0$. For all $n \in \mathbb{N}$ it holds that

$$p_{ii}^{(n+r+s)} \geq p_{ij}^{(s)} p_{jj}^{(n)} p_{ji}^{(r)}.$$

Therefore,

$$\sum_{n=1}^{\infty} p_{jj}^{(n)} \leq \frac{1}{p_{ij}^{(s)} p_{ji}^{(r)}} \sum_{n=1}^{\infty} p_{ii}^{(n+r+s)} \leq \frac{1}{p_{ij}^{(s)} p_{ji}^{(r)}} \sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty,$$

where the last step holds because i is transient. It follows that state j is also transient. □

Theorem 2.7.6 allows us to introduce the following definitions.

Definition 2.7.7. A communicating class is called recurrent if at least one (equivalently, every) state in this class is recurrent. A communicating class is transient if at least one (equivalently, every) state in this class is transient.

Definition 2.7.8. An irreducible Markov chain is called recurrent if at least one (equivalently, every) state in this chain is recurrent. An irreducible Markov chain is called transient if at least one (equivalently, every) state in this chain is transient.

The next theorem states that it is impossible to leave a recurrent class.

Theorem 2.7.9. *Every recurrent communicating class is closed.*

Proof. Let C be a non-closed class. We need to show that it is not recurrent. Since C is not closed, there exist states i, j so that $i \in C$, $j \notin C$ and $i \rightsquigarrow j$. This means that there exists $m \in \mathbb{N}$ so that $p_{ij}^{(m)} = \mathbb{P}_i[X_m = j] > 0$. If the event $\{X_m = j\}$ occurs, then after

time m the chain cannot return to state i because otherwise i and j would be in the same communicating class. It follows that

$$\mathbb{P}_i[\{X_m = j\} \cap \{X_n = i \text{ for infinitely many } n\}] = 0.$$

This implies that

$$\mathbb{P}_i[X_n = i \text{ for infinitely many } n] < 1.$$

Therefore, state i is not recurrent. □

If some communicating class contains only finitely states and the chain cannot leave this class, then it looks very plausible that the chain which started in some state of this class will return to this state infinitely often (and, in fact, will visit any state of this class infinitely often), with probability 1. This is stated in the next theorem.

THEOREM 2.7.10. *Every finite closed communicating class is recurrent.*

PROOF. Let C be a closed communicating class with finitely many elements. Take some state $i \in C$. A chain starting in i stays in C forever and since C is finite, there must be at least one state $j \in C$ which is visited infinitely often with positive probability:

$$\mathbb{P}_i[X_n = j \text{ for infinitely many } n \in \mathbb{N}] > 0.$$

At the moment it is not clear whether we can take $i = j$. But since i and j are in the same communicating class, there exists $m \in \mathbb{N}_0$ so that $p_{ji}^{(m)} > 0$. From the inequality

$$\mathbb{P}_j[X_n = j \text{ for infinitely many } n] > p_{ji}^{(m)} \cdot \mathbb{P}_i[X_n = j \text{ for infinitely many } n] > 0$$

it follows that state j is recurrent. The class C is then recurrent because it contains at leats one recurrent state, namely j. □

So, in a Markov chain with finitely many states we have the following equivalencies

(1) A communicating class is recurrent if and only if it is closed.
(2) A communicating class is transient if and only if it is not closed.

LEMMA 2.7.11. *Consider an irreducible, recurrent Markov chain with an arbitrary initial distribution α. Then, for every state $j \in E$ the number of visits of the chain to j is infinite with probability* 1.

PROOF. Exercise. □

2.8. Recurrence and transience of random walks

EXAMPLE 2.8.1. A *simple random walk on* $\mathbb{Z}$ is a Markov chain with state space $E = \mathbb{Z}$ and transition probabilities

$$p_{i,i+1} = p, \quad p_{i,i-1} = 1 - p, \quad i \in \mathbb{Z}.$$

So, from every state the random walk goes one step to the right with probability p, or one step to the left with probability $1 - p$; see Figure 3. Here, $p \in [0, 1]$ is a parameter.

THEOREM 2.8.2. *If $p = \frac{1}{2}$, then any state of the simple random walk is recurrent. If $p \neq \frac{1}{2}$, then any state is transient.*

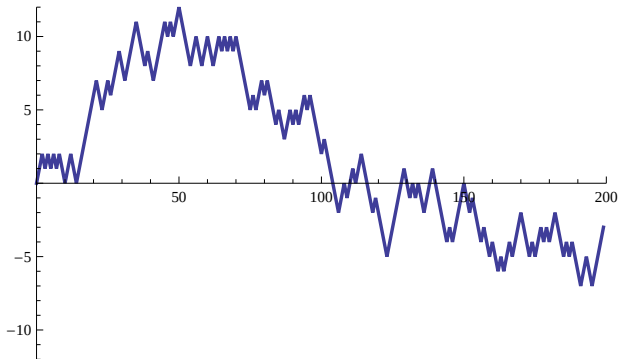


FIGURE 1. Sample path of a simple random walk on $\mathbb{Z}$ with $p = \frac{1}{2}$. The figure shows 200 steps of the walk.

PROOF. By translation invariance, we can restrict our attention to state 0. We can represent our Markov chain as $X_n = \xi_1 + \ldots + \xi_n$, where $\xi_1, \xi_2, \ldots$ are independent and identically distributed random variables with Bernoulli distribution:

$$\mathbb{P}[\xi_k = 1] = p, \quad \mathbb{P}[\xi_k = -1] = 1 - p.$$

CASE 1. Let $p \neq \frac{1}{2}$. Then, $\mathbb{E}\xi_k = p - (1-p) = 2p - 1 \neq 0$. By the strong law of large numbers,

$$\lim_{n\to\infty} \frac{1}{n} X_n = \lim_{n\to\infty} \frac{\xi_1 + \ldots + \xi_n}{n} = \mathbb{E}\xi_1 \neq 0 \quad \text{a.s.}$$

In the case $p > \frac{1}{2}$ we have $\mathbb{E}\xi_1 > 0$ and hence, $\lim_{n\to\infty} X_n = +\infty$ a.s. In the case $p < \frac{1}{2}$ we have $\mathbb{E}\xi_1 < 0$ and hence, $\lim_{n\to\infty} X_n = -\infty$ a.s. In both cases it follows that

$$\mathbb{P}[X_n = 0 \text{ for infinitely many } n] = 0.$$

Hence, state 0 is transient.

CASE 2. Let $p = \frac{1}{2}$. In this case, $\mathbb{E}\xi_k = 0$ and the argument of Case 1 does not work. We will use Theorem 2.7.5. The n-step transition probability from 0 to 0 is given by

$$p_{00}^{(n)} = \begin{cases} 0, & \text{if } n = 2k+1 \text{ odd}, \\ \frac{1}{2^{2k}}\binom{2k}{k}, & \text{if } n = 2k \text{ even}. \end{cases}$$

The Stirling formula $n! \sim \sqrt{2\pi n}(\frac{n}{e})^n$, as $n \to \infty$, yields that

$$p_{00}^{(2k)} \sim \frac{1}{\sqrt{\pi k}}, \quad \text{as } k \to \infty.$$

Since the series $\sum_{k=1}^{\infty} \frac{1}{\sqrt{k}}$ diverges, it follows that $\sum_{n=1}^{\infty} p_{00}^{(n)} = \sum_{k=1}^{\infty} p_{00}^{(2k)} = \infty$. By Theorem 2.7.5, this implies that 0 is a recurrent state. □

EXAMPLE 2.8.3. The *simple, symmetric random walk on* $\mathbb{Z}^d$ is a Markov chain defined as follows. The state space is the d-dimensional lattice

$$\mathbb{Z}^d = \{(n_1, \ldots, n_d) : n_1, \ldots, n_d \in \mathbb{Z}\}.$$

Let $e_1, \ldots, e_d$ be the standard basis of $\mathbb{R}^d$, that is

$$e_1 = (1, 0, 0 \ldots, 0), \quad e_2 = (0, 1, 0, \ldots, 0), \quad e_3 = (0, 0, 1, \ldots, 0), \quad \ldots, \quad e_d = (0, 0, 0, \ldots, 1).$$

Let $\xi_1, \xi_2, \ldots$ be independent and identically distributed d-dimensional random vectors such that

$$\mathbb{P}[\xi_i = e_k] = \mathbb{P}[\xi_i = -e_k] = \frac{1}{2d}, \quad k = 1, \ldots, d, \quad i \in \mathbb{N}.$$

Define $S_n = \xi_1 + \ldots + \xi_n$, $n \in \mathbb{N}$, and $S_0 = 0$. The sequence $S_0, S_1, S_2, \ldots$ is called the simple symmetric random walk on $\mathbb{Z}^d$. It is a Markov chain with transition probabilities

$$p_{i,i+e_1} = p_{i,i-e_1} = \ldots = p_{i,i+e_d} = p_{i,i-e_d} = \frac{1}{2d}, \quad i \in \mathbb{Z}^d.$$

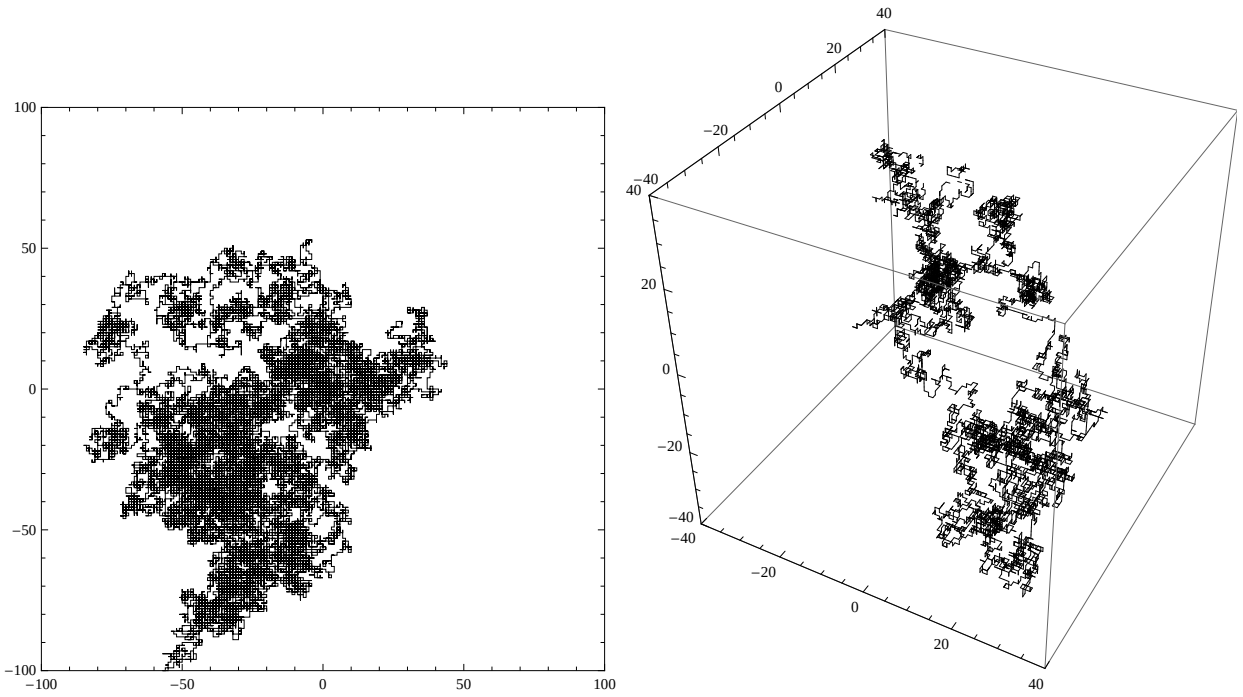


FIGURE 2. *Left:* Sample path of a simple symmetric random walk on $\mathbb{Z}^2$. *Right:* Sample path of a simple symmetric random walk on $\mathbb{Z}^3$. In both cases the random walk makes 50000 steps.

THEOREM 2.8.4 (Pólya, 1921). *The simple symmetric random walk on $\mathbb{Z}^d$ is recurrent if and only if $d = 1, 2$ and transient if and only if $d \geq 3$.*

PROOF. For $d = 1$ we already proved the statement in Theorem 2.8.2.

Consider the case $d = 2$. We compute the n-step transition probability $p_{00}^{(n)}$. For an odd n this probability is 0. For an even $n = 2k$ we have

$$p_{00}^{(2k)} = \frac{1}{4^{2k}} \sum_{i=0}^{k} \binom{2k}{i, i, k-i, k-i} = \frac{1}{4^{2k}} \binom{2k}{k} \sum_{i=0}^{k} \binom{k}{i} \binom{k}{k-i} = \left(\frac{1}{2^{2k}} \binom{2k}{k} \right)^2 \sim \frac{1}{\pi k},$$

as $k \to \infty$, where the last step is by the Stirling formula. The harmonic series $\sum_{k=1}^{\infty} \frac{1}{k}$ diverges. Therefore, $\sum_{n=1}^{\infty} p_{00}^{(n)} = \infty$ and the random walk is recurrent in $d = 2$ dimensions.

Generalizing the cases $d = 1, 2$ one can show that for an arbitrary dimension $d \in \mathbb{N}$ we have, as $k \to \infty$,

$$p_{00}^{(2k)} \sim \frac{1}{(\pi k)^{d/2}}.$$

Since the series $\sum_{k=1}^{\infty} k^{-d/2}$ is convergent for $d \geq 3$ it holds that $\sum_{n=1}^{\infty} p_{00}^{(n)} < \infty$ and the random walk is transient in $d = 3$ dimensions. □

2.9. Existence and uniqueness of the invariant measure

The next two theorems state that any irreducible and recurrent Markov chain has a *unique invariant measure* λ, up to a multiplication by a constant. This measure may be finite (that is, $\sum_{i \in E} \lambda_i < +\infty$) or infinite (that is, $\sum_{i \in E} \lambda_i = +\infty$).

First we provide an explicit construction of an invariant measure for an irreducible and recurrent Markov chain. Consider a Markov chain starting at state $k \in E$. Denote the time of the first return to k by

$$T_k = \min\{n \in \mathbb{N} : X_n = k\} \in \mathbb{N} \cup \{+\infty\}.$$

The minimum of an empty set is by convention $+\infty$. For a state $i \in E$ denote the expected number of visits to i before the first return to k by

$$\gamma_i = \gamma_i^{(k)} = \mathbb{E}_k \sum_{n=0}^{T_k - 1} \mathbb{1}_{\{X_n = i\}} \in [0, +\infty].$$

THEOREM 2.9.1. *For an irreducible and recurrent Markov chain starting at state $k \in E$ we have*

(1) $\gamma_k = 1$.
(2) *For all $i \in E$ it holds that $0 < \gamma_i < \infty$.*
(3) $\gamma = (\gamma_i)_{i \in E}$ *is an invariant measure.*

PROOF.

STEP 1. We show that $\gamma_k = 1$. By definition of T_k, we have $\sum_{n=0}^{T_k - 1} \mathbb{1}_{\{X_n = k\}} = 1$, if the chain starts at k. It follows that $\gamma_k = \mathbb{E}_k 1 = 1$.

STEP 2. We show that for every state $j \in E$,

$$\gamma_j = \sum_{i \in E} p_{ij} \gamma_i. \tag{2.9.1}$$

(At this moment, both sides of (2.9.1) are allowed to be infinite, but in Step 3 we will show that both sides are actually finite). The Markov chain is recurrent, thus $T_k < \infty$ almost surely. By definition, $X_{T_k} = k = X_0$. We have

$$\gamma_j = \mathbb{E}_k \sum_{n=1}^{T_k} \mathbb{1}_{\{X_n = j\}} = \mathbb{E}_k \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n = j, n \leq T_k\}} = \sum_{n=1}^{\infty} \mathbb{P}_k[X_n = j, T_k \geq n].$$

Before visiting state j at time n the chain must have been in some state i at time $n-1$, where $i \in E$ can be, in general, arbitrary. We obtain that

$$\gamma_j = \sum_{i \in E} \sum_{n=1}^{\infty} \mathbb{P}_k[X_n = j, X_{n-1} = i, T_k \geq n] = \sum_{i \in E} \sum_{n=1}^{\infty} p_{ij} \mathbb{P}_k[X_{n-1} = i, T_k \geq n].$$

Introducing the new summation variable $m = n - 1$, we obtain that

$$\gamma_j = \sum_{i \in E} p_{ij} \sum_{m=0}^{\infty} \mathbb{E}_k \mathbb{1}_{\{X_m = i, T_k \geq m+1\}} = \sum_{i \in E} p_{ij} \, \mathbb{E}_k \sum_{m=0}^{T_k - 1} \mathbb{1}_{\{X_m = i\}} = \sum_{i \in E} p_{ij} \gamma_i.$$

This proves that (2.9.1) holds.

Step 3. Let $i \in E$ be an arbitrary state. We show that $0 < \gamma_i < \infty$. Since the chain is irreducible, there exist $n, m \in \mathbb{N}_0$ such that $p_{ik}^{(m)} > 0$ and $p_{ki}^{(n)} > 0$. From (2.9.1) it follows that

$$\gamma_i = \sum_{l \in E} p_{li}^{(n)} \gamma_l \geq p_{ki}^{(n)} \gamma_k = p_{ki}^{(n)} > 0.$$

On the other hand, again using (2.9.1), we obtain that

$$1 = \gamma_k = \sum_{l \in E} p_{lk}^{(m)} \gamma_l \geq p_{ik}^{(m)} \gamma_i.$$

This implies that $\gamma_i \leq 1/p_{ik}^{(m)} < \infty$. □

The next theorem states the uniqueness of the invariant measure, up to multiplication by a constant.

Theorem 2.9.2. *Consider an irreducible and recurrent Markov chain and fix some state $k \in E$. Then, every invariant measure λ can be represented in the form*

$$\lambda_j = c \gamma_j^{(k)} \quad \text{for all } j \in E,$$

where c is a constant (not depending on j). In fact, $c = \lambda_k$.

Remark 2.9.3. Hence, the invariant measure is unique up to a multiplication by a constant. In particular, the invariant measures $(\gamma_i^{(k_1)})_{i \in E}$ and $(\gamma_i^{(k_2)})_{i \in E}$, for different states $k_1, k_2 \in E$, differ by a multiplicative constant.

Proof. Let λ be an invariant measure.

Step 1. We show that $\lambda_j \geq \lambda_k \gamma_j^{(k)}$ for all $j \in E$. We will *not* use the irreducibility and the recurrence of the chain in this step. The invariance of the measure λ implies that

$$\lambda_j = \sum_{i_0 \in E} \lambda_{i_0} p_{i_0 j} = \sum_{i_0 \neq k} \lambda_{i_0} p_{i_0 j} + \lambda_k p_{kj}.$$

Applying the same procedure to λ_{i_0}, we obtain

$$\begin{aligned}\lambda_j &= \sum_{i_0 \neq k} \left(\sum_{i_1 \neq k} \lambda_{i_1} p_{i_1 i_0} + \lambda_k p_{k i_0} \right) p_{i_0 j} + \lambda_k p_{kj} \\ &= \sum_{i_0 \neq k} \sum_{i_1 \neq k} \lambda_{i_1} p_{i_1 i_0} p_{i_0 j} + \left(\lambda_k p_{kj} + \lambda_k \sum_{i_0 \neq k} p_{k i_0} p_{i_0 j} \right).\end{aligned}$$

Applying the procedure to λ_{i_1} and repeating it over and over again we obtain that for every $n \in \mathbb{N}$,

$$\lambda_j = \sum_{i_0, i_1, \ldots, i_n \neq k} \lambda_{i_n} p_{i_n i_{n-1}} \cdots p_{i_1 i_0} p_{i_0 j} + \lambda_k \left(p_{kj} + \sum_{i_0 \neq k} p_{k i_0} p_{i_0 j} + \ldots + \sum_{i_0, \ldots, i_{n-1} \neq k} p_{k i_0} p_{i_0 i_1} \cdots p_{i_{n-1} j} \right).$$

Noting that the first term is non-negative, we obtain that

$$\lambda_j \geq 0 + \lambda_k \mathbb{P}_k[X_1 = j, T_k \geq 1] + \lambda_k \mathbb{P}_k[X_2 = j, T_k \geq 2] + \ldots + \lambda_k \mathbb{P}_k[X_n = j, T_k \geq n].$$

Since this holds for every $n \in \mathbb{N}$, we can pass to the limit as $n \to \infty$:

$$\lambda_j \geq \lambda_k \sum_{n=1}^{\infty} \mathbb{P}_k[X_n = j, T_k \geq n] = \lambda_k \gamma_j^{(k)}.$$

It follows that $\lambda_j \geq \lambda_k \gamma_j^{(k)}$.

Step 2. We prove the converse inequality. Consider $\mu_j := \lambda_j - \lambda_k \gamma_j^{(k)}$, $j \in E$. By the above, $\mu_j \geq 0$ for all $j \geq 0$ so that $\mu = (\mu_j)_{j \in E}$ is a measure. Moreover, this measure is invariant because it is a linear combination of two invariant measures. Finally, note that by definition, $\mu_k = 0$. We will prove that this implies that $\mu_j = 0$ for all $j \in E$. By the irreducibility of our Markov chain, for every $j \in E$ we can find $n \in \mathbb{N}_0$ such that $p_{jk}^{(n)} > 0$. By the invariance property of μ,

$$0 = \mu_k = \sum_{i \in E} \mu_i p_{ik}^{(n)} \geq \mu_j p_{jk}^{(n)}.$$

It follows that $\mu_j p_{jk}^{(n)} = 0$ but since $p_{jk}^{(n)} > 0$, we must have $\mu_j = 0$. By the definition of μ_j this implies that $\lambda_j = \lambda_k \gamma_j^{(k)}$. □

We can now summarize Theorems 2.9.1 and 2.9.2 as follows:

Theorem 2.9.4. *A recurrent, irreducible Markov chain has unique (up to a constant multiple) invariant measure.*

This invariant measure may be finite or infinite. However, if the Markov chain has only finitely many states, then the measure must be finite and we can even normalize it to be a *probability* measure.

Corollary 2.9.5. *A finite and irreducible Markov chain has a unique invariant probability measure.*

PROOF. A finite and irreducible Markov chain is recurrent by Theorem 2.7.10. By Theorem 2.9.1, there exists an invariant measure $\lambda = (\lambda_i)_{i\in E}$. Since the number of states in E is finite by assumption and $\lambda_i < \infty$ by Theorem 2.9.1, we have $M := \sum_{i\in E} \lambda_i < \infty$ and hence, the measure λ is finite. To obtain an invariant *probability* measure, consider the measure $\lambda'_i = \lambda_i/M$.

To show that the invariant probability measure is unique, assume that we have two invariant probability measures $\nu' = (\nu'_i)_{i\in E}$ and $\nu'' = (\nu''_i)_{i\in E}$. Take an arbitrary state $k \in E$. By Theorem 2.9.2, there are constants c' and c'' such that $\nu'_i = c'\gamma_i^{(k)}$ and $\nu''_i = c''\gamma_i^{(k)}$, for all $i \in E$. But since both ν' and ν'' are *probability* measures, we have

$$1 = \sum_{i\in E} \nu'_i = c' \sum_{i\in E} \gamma_i^{(k)}, \quad 1 = \sum_{i\in E} \nu''_i = c'' \sum_{i\in E} \gamma_i^{(k)}.$$

This implies that $c' = c''$ and hence, the measures ν' and ν'' are equal. □

Above, we considered only irreducible, recurrent chains. What happens if the chain is irreducible and transient? It turns out that in this case everything is possible:

(1) It is possible that there is no invariant measure at all (except the zero measure).
(2) It is possible that there is a unique (up to multiplication by a constant) invariant measure.
(3) It is possible that there are at least two invariant measures which are not constant multiples of each other.

EXERCISE 2.9.6. Consider a Markov chain on $\mathbb{N}$ with transition probabilities $p_{i,i+1} = 1$, for all $i \in \mathbb{N}$. Show that the only invariant measure is $\lambda_i = 0$, $i \in \mathbb{N}$.

EXERCISE 2.9.7. Consider a Markov chain on $\mathbb{Z}$ with transition probabilities $p_{i,i+1} = 1$, for all $i \in \mathbb{Z}$. Show that the invariant measures have the form $\lambda_i = c$, $i \in \mathbb{Z}$, where $c \geq 0$ is constant.

EXERCISE 2.9.8. Consider a simple random walk on $\mathbb{Z}$ with $p \neq \frac{1}{2}$. Show that any invariant measure has the form

$$\lambda_i = c_1 + c_2 \left(\frac{p}{1-p}\right)^i, \quad i \in \mathbb{Z},$$

for some constants $c_1 \geq 0$, $c_2 \geq 0$.

2.10. Positive recurrence and null recurrence

The set of recurrent states of a Markov chain can be further subdivided into the set of positive recurrent states and the set of negative recurrent states. Let us define the notions of positive recurrence and null recurrence.

Consider a Markov chain on state space E. Take some state $i \in E$, assume that the Markov chain starts at state i and denote by T_i the time of the first return of the chain to state i:

$$T_i = \min\{n \in \mathbb{N} : X_n = i\} \in \mathbb{N} \cup \{+\infty\}.$$

Denote by m_i the expected return time of the chain to state i, that is

$$m_i = \mathbb{E}_i T_i \in (0, \infty]$$

Note that for a transient state i we always have $m_i = +\infty$ because the random variable T_i takes the value $+\infty$ with strictly positive probability $1 - f_i > 0$, see Theorem 2.7.3. However, for a recurrent state i the value of m_i may be both finite and infinite, as we shall see later.

DEFINITION 2.10.1. A state $i \in E$ as called *positive recurrent* if $m_i < \infty$.

DEFINITION 2.10.2. A state $i \in E$ is called *null recurrent* if it is recurrent and $m_i = +\infty$.

REMARK 2.10.3. Both null recurrent states and positive recurrent states are recurrent. For null recurrent states this is required by definition. For a positive recurrent state we have $m_i < \infty$ which means that T_i cannot attain the value $+\infty$ with strictly positive probability and hence, state i is recurrent.

THEOREM 2.10.4. *Consider an irreducible Markov chain. Then the following statements are equivalent:*

(1) *Some state is positive recurrent.*
(2) *All states are positive recurrent.*
(3) *The chain has invariant* probability *measure* $\lambda = (\lambda_i)_{i \in E}$.

Also, if these statements hold, then $m_i = \frac{1}{\lambda_i}$ *for all* $i \in E$.

PROOF. The implication $2 \Rightarrow 1$ is evident.

PROOF OF $1 \Rightarrow 3$. Let $k \in E$ be a positive recurrent state. Then, k is recurrent and all states of the chain are recurrent by irreducibility. By Theorem 2.9.1, $(\gamma_i^{(k)})_{i \in E}$ is an invariant measure. However, we need an invariant *probability* measure. To construct it, note that

$$\sum_{j \in E} \gamma_j^{(k)} = m_k < \infty$$

(since k is positive recurrent). We can therefore define $\lambda_i = \gamma_i^{(k)}/m_k$, $i \in E$. Then, $\sum_{i \in E} \lambda_i = 1$, and $(\lambda_i)_{i \in E}$ is an invariant probability measure.

PROOF OF $3 \Rightarrow 2$. Let $(\lambda_i)_{i \in E}$ be an invariant probability measure. First we show that $\lambda_k > 0$ for every state $k \in E$. Since λ is a probability measure, we have $\lambda_l > 0$ for at least one $l \in E$. By irreducibility, we have $p_{lk}^{(n)} > 0$ for some $n \in \mathbb{N}_0$ and by invariance of λ, we have

$$\lambda_k = \sum_{i \in E} p_{ik}^{(n)} \lambda_i \geq p_{lk}^{(n)} \lambda_l > 0.$$

This proves that $\lambda_k > 0$ for every $k \in E$.

By Step 1 from the proof of Theorem 2.9.2 (note that this step does not use recurrence), we have for all $j \in E$,

$$\lambda_i \geq \lambda_k \gamma_i^{(k)}.$$

Hence,

$$m_k = \sum_{i \in E} \gamma_i^{(k)} \leq \sum_{i \in E} \frac{\lambda_i}{\lambda_k} = \frac{1}{\lambda_k} < \infty.$$

It follows that k is positive recurrent, thus establishing statement 2.

PROOF THAT $m_k = \frac{1}{\lambda_k}$. Assume that statements 1,2,3 hold. In particular, the chain is recurrent and by Theorem 2.9.2, we must have $\lambda_i = \lambda_k \gamma_i^{(k)}$ for all $i \in E$. It follows that

$$m_k = \sum_{i \in E} \gamma_i^{(k)} = \sum_{i \in E} \frac{\lambda_i}{\lambda_k} = \frac{1}{\lambda_k},$$

thus proving the required formula. □

EXAMPLE 2.10.5. Any state in a *finite* irreducible Markov chain is positive recurrent. Indeed, such a chain has an invariant probability measure by Corollary 2.9.5.

EXAMPLE 2.10.6. Consider a simple symmetric random walk on $\mathbb{Z}$ or on $\mathbb{Z}^2$. This chain is irreducible. Any state is recurrent by Pólya's Theorem 2.8.4. We show that in fact, any state is *null* recurrent. To see this, note that the measure assigning the value 1 to every state $i \in E$ is invariant by the definition of the chain. By Theorem 2.9.2, any other invariant measure must be of the form $\lambda_i = c$, $i \in E$, for some constant $c \geq 0$. However, no measure of this form is a probability measure. So, there is no invariant probability measure and by Theorem 2.10.4, all states must be null recurrent.

2.11. Convergence to the invariant probability measure

We are going to state and prove a "strong law of large numbers" for Markov chains. First recall that the usual strong law of large numbers states that if $\xi_1, \xi_2, \ldots$ are i.i.d. random variables with $\mathbb{E}|\xi_1| < \infty$, then

$$\frac{\xi_1 + \ldots + \xi_n}{n} \xrightarrow[n\to\infty]{a.s.} \mathbb{E}\xi_1. \tag{2.11.1}$$

The statement is not applicable if $\mathbb{E}|\xi_1| = \infty$. However, it is an exercise to show that if $\xi_1, \xi_2, \ldots$ are i.i.d. random variables which are a.s. *nonnegative* with $\mathbb{E}\xi_1 = +\infty$, then

$$\frac{\xi_1 + \ldots + \xi_n}{n} \xrightarrow[n\to\infty]{a.s.} +\infty. \tag{2.11.2}$$

Consider a Markov chain $\{X_n : n \in \mathbb{N}_0\}$ with initial distribution $\alpha = (\alpha_i)_{i \in E}$. Given a state $i \in E$, denote the number of visits to state i in the first n steps by

$$V_i(n) = \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k = i\}}.$$

THEOREM 2.11.1. *Consider an irreducible Markov chain $\{X_n : n \in \mathbb{N}_0\}$ with an arbitrary initial distribution $\alpha = (\alpha_i)_{i \in E}$.*

1. *If the Markov chain is transient or null recurrent, then for all $i \in E$ it holds that*

$$\frac{V_i(n)}{n} \xrightarrow[n\to\infty]{} 0 \quad a.s. \tag{2.11.3}$$

2. *If the Markov chain is positive recurrent with invariant probability measure λ, then for all $i \in E$ it holds that*

$$\frac{V_i(n)}{n} \xrightarrow[n\to\infty]{} \lambda_i \quad a.s. \tag{2.11.4}$$

PROOF. If the chain is transient, then $V_i(n)$ stays bounded as a function of n, with probability 1. This implies (2.11.3). In the sequel, let the chain be recurrent.

For simplicity, we will assume in this proof that the chain starts in state i. Denote the time of the k-th visit of the chain to i by S_k, that is

$$\begin{aligned} S_1 &= \min\{n \in \mathbb{N} : X_n = i\}, \\ S_2 &= \min\{n > S_1 : X_n = i\}, \\ S_3 &= \min\{n > S_2 : X_n = i\}, \end{aligned}$$

and so on. Note that $S_1, S_2, S_3, \ldots$ are a.s. finite by the recurrence of the chain. Let also $\xi_1, \xi_2, \xi_3, \ldots$ be the excursion times between the returns to i, that is

$$\xi_1 = S_1, \;\; \xi_2 = S_2 - S_1, \;\; \xi_3 = S_3 - S_2, \;\; \ldots.$$

Then, $\xi_1, \xi_2, \xi_3, \ldots$ are i.i.d. random variables by the Markov property.

By definition of $V_i(n)$ we have

$$\xi_1 + \xi_2 + \ldots + \xi_{V_i(n)-1} \le n \le \xi_1 + \xi_2 + \ldots + \xi_{V_i(n)}.$$

Dividing this by $V_i(n)$ we get

$$\frac{\xi_1 + \xi_2 + \ldots + \xi_{V_i(n)-1}}{V_i(n)} \le \frac{n}{V_i(n)} \le \frac{\xi_1 + \xi_2 + \ldots + \xi_{V_i(n)}}{V_i(n)}. \tag{2.11.5}$$

Note that by recurrence, $V_i(n) \underset{n\to\infty}{\longrightarrow} \infty$ a.s.

CASE 1. Let the chain be *null* recurrent. It follows that $\mathbb{E}\xi_1 = \infty$. By using (2.11.2) and (2.11.5), we obtain that

$$\frac{n}{V_i(n)} \xrightarrow[n\to\infty]{a.s.} \infty.$$

This proves (2.11.3).

CASE 2. Let the chain be *positive* recurrent. Then, by Theorem 2.10.4, $\mathbb{E}\xi_1 = m_i = \frac{1}{\lambda_i} < \infty$. Using (2.11.1) and (2.11.5) we obtain that

$$\frac{n}{V_i(n)} \xrightarrow[n\to\infty]{a.s.} \frac{1}{\lambda_i}.$$

This proves (2.11.4). □

In the next theorem we prove that the n-step transition probabilities converge, as $n \to \infty$, to the invariant probability measure.

THEOREM 2.11.2. *Consider an irreducible, aperiodic, positive recurrent Markov chain* $\{X_n : n \in \mathbb{N}_0\}$ *with transition matrix* P *and invariant probability measure* $\lambda = (\lambda_i)_{i\in E}$. *The initial distribution* $\alpha = (\alpha_i)_{i\in E}$ *may be arbitrary. Then, for all* $j \in E$ *it holds that*

$$\lim_{n\to\infty} \mathbb{P}[X_n = j] = \lambda_j.$$

In particular, $\lim_{n\to\infty} p_{ij}^{(n)} = \lambda_j$ *for all* $i, j \in E$.

REMARK 2.11.3. In particular, the theorem applies to any irreducible and aperiodic Markov chain with finite state space.

For the proof we need the following lemma.

LEMMA 2.11.4. *Consider an irreducible and aperiodic Markov chain. Then, for every states $i, j \in E$ we can find $N = N(i,j) \in \mathbb{N}$ such that for all $n > N$ we have $p_{ij}^{(n)} > 0$.*

PROOF. The chain is irreducible, hence we can find $r \in \mathbb{N}_0$ such that $p_{ij}^{(r)} > 0$. Also, the chain is aperiodic, hence we can find $N_0 \in \mathbb{N}$ such that for all $k > N_0$ we have $p_{ii}^{(k)} > 0$. It follows that for all $k > N_0$,

$$p_{ij}^{(k+r)} > p_{ii}^{(k)} p_{ij}^{(r)} > 0.$$

It follows that for every $n := k + r$ such that $n > N_0 + r$, we have $p_{ij}^{(n)} > 0$. □

PROOF OF THEOREM 2.11.2. We use the "coupling method".

STEP 1. Consider *two* Markov chains called $\{X_n : n \in \mathbb{N}_0\}$ and $\{Y_n : n \in \mathbb{N}_0\}$ such that

(1) X_n is a Markov chain with initial distribution α and transition matrix P.
(2) Y_n is a Markov chain with initial distribution λ (the invariant probability measure) and the same transition matrix P.
(3) The process $\{X_n : n \in \mathbb{N}_0\}$ is independent of the process $\{Y_n : n \in \mathbb{N}_0\}$.

Note that both Markov chains have the same transition matrix but different initial distributions. Fix an arbitrary state $b \in E$. Denote by T be the time at which the chains meet at state b:

$$T = \min\{n \in \mathbb{N} : X_n = Y_n = b\} \in \mathbb{N} \cup \{+\infty\}.$$

If the chains do not meet at b, we set $T = +\infty$.

STEP 2. We show that $\mathbb{P}[T < \infty] = 1$. Consider the stochastic process $W_n = (X_n, Y_n)$ taking values in $E \times E$. It is a Markov chain on $E \times E$ with transition probabilities given by

$$\tilde{p}_{(i,k),(j,l)} = p_{ij} p_{kl}, \quad (i,k) \in E \times E, \quad (j,l) \in E \times E.$$

The initial distribution of W_0 is given by

$$\mu_{(i,k)} = \alpha_i \lambda_k, \quad (i,k) \in E \times E.$$

Since the chains X_n and Y_n are aperiodic and irreducible by assumption of the theorem, we can apply Lemma 2.11.4 to obtain for every $i, j, k, l \in E$ a number $N = N(i,j,k,l) \in \mathbb{N}$ such that for all $n > N$ we have

$$\tilde{p}_{(i,k),(j,e)}^{(n)} = p_{ij}^{(n)} p_{ke}^{(n)} > 0.$$

Thus, the chain W_n is irreducible. Also, it is an exercise to check that the probability measure $\tilde{\lambda}_{(i,k)} := \lambda_i \lambda_k$ is invariant for W_n. Thus, by Theorem 2.10.4, the Markov chain W_n is positive recurrent and thereby recurrent. Therefore, $T < \infty$ a.s. by Lemma 2.7.11.

STEP 3. Define the stochastic process $\{Z_n : n \in \mathbb{N}_0\}$ by

$$Z_n = \begin{cases} X_n, & \text{if } n \le T, \\ Y_n, & \text{if } n \ge T. \end{cases}$$

Then, Z_n is a Markov chain with initial distribution α and the same transition matrix P as X_n and Y_n. (The Markov chain Z_n is called the coupling of X_n and Y_n). The chain Y_n starts with the invariant probability measure λ and hence, at every time n, Y_n is distributed according to λ. Also, the chain Z_n has the same initial distribution α and the same transition

matrix P as the chain X_n, so that in particular, the random elements X_n and Z_n have the same distribution at every time n. Using these facts, we obtain that

$$|\mathbb{P}[X_n = j] - \lambda_j| = |\mathbb{P}[X_n = j] - \mathbb{P}[Y_n = j]| = |\mathbb{P}[Z_n = j] - \mathbb{P}[Y_n = j]|.$$

By definition of Z_n, we can rewrite this as

$$\begin{aligned} |\mathbb{P}[X_n = j] - \lambda_j| &= |\mathbb{P}[X_n = j, n < T] + \mathbb{P}[Y_n = j, n \geq T] - \mathbb{P}[Y_n = j]| \\ &= |\mathbb{P}[X_n = j, n < T] - \mathbb{P}[Y_n = j, n < T]| \\ &\leq \mathbb{P}[T > n]. \end{aligned}$$

But we have shown in Step 2 that $\mathbb{P}[T = \infty] = 0$, hence $\lim_{n\to\infty} \mathbb{P}[T > n] = 0$. It follows that

$$\lim_{n\to\infty} \mathbb{P}[X_n = j] = \lambda_j,$$

thus establishing the theorem. □

CHAPTER 3

Renewal processes and Poisson process

3.1. Definition of renewal processes and limit theorems

Let $\xi_1, \xi_2, \dots$ be independent and identically distributed random variables with $\mathbb{P}[\xi_k > 0] = 1$. Define their partial sums

$$S_n = \xi_1 + \ldots + \xi_n, \quad n \in \mathbb{N}, \quad S_0 = 0.$$

Note that the sequence $S_1, S_2, \dots$ is increasing. We call $S_1, S_2, \dots$ the *renewal times* (or simply *renewals*) and $\xi_1, \xi_2, \dots$ the *interrenewal times*.

DEFINITION 3.1.1. The process $\{N_t : t \geq 0\}$ given by

$$N_t = \sum_{n=1}^{\infty} \mathbb{1}_{\{S_n \leq t\}}$$

is called the *renewal process*.

THEOREM 3.1.2 (Law of large numbers for renewal processes). *Let $m := \mathbb{E}\xi_1 \in (0, \infty)$, then*

$$\frac{N_t}{t} \overset{a.s.}{\to} \frac{1}{m}, \quad \text{as } t \to \infty.$$

IDEA OF PROOF. By the definition of N_t we have the inequality

$$S_{N_t} \leq t \leq S_{N_t+1}.$$

Dividing this by N_t we obtain

$$\frac{S_{N_t}}{N_t} \leq \frac{t}{N_t} \leq \frac{S_{N_t+1}}{N_t + 1} \cdot \frac{N_t + 1}{N_t}. \tag{3.1.1}$$

We have $N_t \to \infty$ as $t \to \infty$ since there are infinitely many renewals and thus, the function N_t (which is non-decreasing by definition) cannot stay bounded. By the law of large numbers, both sides of (3.1.1) a.s. converge to m as $t \to \infty$. By the sandwich lemma, we have

$$\frac{t}{N_t} \overset{a.s.}{\to} m, \quad \text{as } t \to \infty.$$

This proves the claim. □

THEOREM 3.1.3 (Central limit theorem for renewal processes). *Let $m := \mathbb{E}\xi_1 \in (0, \infty)$ and $\sigma^2 := \operatorname{Var} \xi_1 \in (0, \infty)$. Then,*

$$\frac{N_t - \frac{t}{m}}{\frac{\sigma}{m^{3/2}}\sqrt{t}} \overset{d}{\to} \mathrm{N}(0, 1), \quad \text{as } t \to \infty.$$

IDEA OF PROOF. The usual central limit theorem for $S_n = \xi_1 + \ldots + \xi_n$ states that

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{d} \mathrm{N}(0,1).$$

Denoting by N a standard normal random variable we can write this as follows: For large n, we have an approximate equality of distributions

$$S_n \approx nm + \sigma\sqrt{n}N.$$

This means that the interval $[0, nm + \sigma\sqrt{n}N]$ contains approximately n renewals. By the law of large numbers for renewal processes, see Theorem 3.1.2, it seems plausible that the interval $[nm, nm + \sigma\sqrt{n}N]$ contains approximately $\sigma\sqrt{n}N/m$ renewals. It follows that the interval $[0, nm]$ contains approximately $n - \sigma\sqrt{n}N/m$ renewals. Let us now introduce the variable $t = nm$. Then, $n \to \infty$ is equivalent to $t \to \infty$. Consequently, for large t in the interval $[0, t]$ we have approximately

$$\frac{t}{m} - \frac{\sigma\sqrt{t}}{m^{3/2}}N$$

renewals. By definition, this number of renewals is N_t. This means that

$$\frac{N_t - \frac{t}{m}}{\frac{\sigma}{m^{3/2}}\sqrt{t}} \approx N,$$

for large t. □

DEFINITION 3.1.4. The *renewal function* $H(t)$ is the expected number of renewals in the interval $[0, t]$:

$$H(t) = \mathbb{E}N_t, \quad t \geq 0.$$

REMARK 3.1.5. Denoting by $F^{*k}(t) = \mathbb{P}[S_k \leq t]$ the distribution function of S_k, we have the formula

$$H(t) = \mathbb{E}N_t = \mathbb{E}\sum_{k=1}^{\infty} \mathbb{1}_{S_k \leq t} = \sum_{k=1}^{\infty} \mathbb{E}\mathbb{1}_{S_k \leq t} = \sum_{k=1}^{\infty} \mathbb{P}[S_k \leq t] = \sum_{k=1}^{\infty} F^{*k}(t).$$

THEOREM 3.1.6 (Weak renewal theorem). *Let $m := \mathbb{E}\xi_1 \in (0, \infty)$. It holds that*

$$\lim_{t\to\infty} \frac{H(t)}{t} = \frac{1}{m}.$$

IDEA OF PROOF. By Theorem 3.1.2, $\frac{N_t}{t} \overset{a.s.}{\to} \frac{1}{m}$ as $t \to \infty$. In order to obtain Theorem 3.1.6, we have to take expectation of both sides and interchange the limit and the expectation. The rigorous justification will be omitted. □

DEFINITION 3.1.7. The random variables ξ_k are called *lattice* if there are $a > 0, b \in \mathbb{R}$ so that ξ_k with probability 1 takes values in the set $a\mathbb{Z} + b$, that is

$$\mathbb{P}[\xi_k \in \{an + b : n \in \mathbb{Z}\}] = 1.$$

THEOREM 3.1.8 (Blackwell renewal theorem). *Assume that ξ_1 is non-lattice and let $m := \mathbb{E}\xi_1 \in (0, \infty)$. Then, for all $s > 0$,*

$$\lim_{t\to\infty} (H(t+s) - H(t)) = \frac{s}{m}.$$

PROOF. Omitted □

3.2. Stationary processes and processes with stationary increments

Consider a stochastic process $\{X_t, t \geq 0\}$. For concreteness, we have chosen the index set T to be $[0, \infty)$, but similar definitions apply to stochastic processes with index sets $T = \mathbb{R}, \mathbb{N}, \mathbb{N}_0, \mathbb{Z}$.

DEFINITION 3.2.1. The process $\{X_t : t \geq 0\}$ is called *stationary* if for all $n \in \mathbb{N}$, $0 \leq t_1 \leq \ldots \leq t_n$ and all $h \geq 0$,

$$(X_{t_1}, \ldots, X_{t_n}) \stackrel{d}{=} (X_{t_1+h}, \ldots, X_{t_n+h}).$$

EXAMPLE 3.2.2. Let $\{X_t : t \in \mathbb{N}_0\}$ be independent and identically distributed random variables. We claim that the process X is stationary. Let μ be the probability distribution of X_t, that is $\mu(A) = \mathbb{P}[X_t \in A]$, for all Borel sets $A \subset \mathbb{R}$. Then, for all Borel sets $A_1, \ldots, A_n \subset \mathbb{R}$,

$$\mathbb{P}[X_{t_1+h} \in A_1, \ldots, X_{t_n+h} \in A_n] = \mu(A_1) \cdot \ldots \cdot \mu(A_n) = \mathbb{P}[X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n].$$

This proves that X is stationary.

EXAMPLE 3.2.3. Let $\{X_t : t \in \mathbb{N}_0\}$ be a Markov chain starting with an *invariant* probability distribution λ. Then, X_t is stationary.

PROOF. Let us first compute the joint distribution of $(X_h, X_{h+1}, \ldots, X_{h+m})$. For any states $i_0, \ldots, i_m \in E$ we have

$$\mathbb{P}[X_h = i_0, X_{h+1} = i_1, \ldots, X_{h+m} = i_m] = \mathbb{P}[X_h = i_0] \cdot p_{i_0 i_1} \cdot \ldots \cdot p_{i_{m-1} i_m}.$$

Since the initial measure λ of the Markov chain is invariant, we have $\mathbb{P}[X_h = i_0] = \lambda_{i_0}$. We therefore obtain that

$$\mathbb{P}[X_h = i_0, X_{h+1} = i_1, \ldots, X_{h+m} = i_m] = \lambda_{i_0} p_{i_0 i_1} \cdot \ldots \cdot p_{i_{m-1} i_m}.$$

This expression does not depend on h thus showing that

$$(X_h, X_{h+1}, \ldots, X_{h+m}) \stackrel{d}{=} (X_0, X_1, \ldots, X_m).$$

If we drop some components in the first vector and the corresponding components in the second vector, the vectors formed by the remaining components still have the same distribution. In this way we can prove that $(X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_n+h})$ has the same distribution as $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$. □

DEFINITION 3.2.4. The process $\{X_t : t \geq 0\}$ has *stationary increments* if for all $n \in \mathbb{N}$, $h \geq 0$ and $0 \leq t_0 \leq t_1 \leq \ldots \leq t_m$, we have the following equality in distribution:

$$(X_{t_1+h} - X_{t_0+h}, X_{t_2+h} - X_{t_1+h}, \ldots, X_{t_n+h} - X_{t_{n-1}+h}) \stackrel{d}{=} (X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}).$$

DEFINITION 3.2.5. The process $\{X_t : t \geq 0\}$ has *independent increments* if for all $n \in \mathbb{N}$ and $0 \leq t_0 \leq t_1 \leq \ldots \leq t_n$, the random variables

$$X_{t_0}, X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$$

are independent.

Later we will consider two examples of processes which have both stationary and independent increments: the Poisson Process and the Brownian Motion.

3.3. Poisson process

The Poisson process is a special case of renewal process in which the interrenewal times are exponentially distributed. Namely, let $\xi_1, \xi_2, \ldots$ be independent identically distributed random variables having exponential distribution with parameter $\lambda > 0$, that is

$$\mathbb{P}[\xi_k \leq x] = 1 - e^{-\lambda x}, \quad x \geq 0.$$

Define the renewal times S_n by

$$S_n = \xi_1 + \ldots + \xi_n, \quad n \in \mathbb{N}, \quad S_0 = 0.$$

It's an exercise to show (for example, by induction) that the density of S_n is given by

$$f_{S_n}(x) = \frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x}, \quad x \geq 0.$$

The distribution of S_n is called the Erlang distribution with parameters n and λ. It is a particular case of the Gamma distribution.

Definition 3.3.1. The Poisson process with intensity $\lambda > 0$ is a process $\{N_t : t \geq 0\}$ defined by

$$N_t = \sum_{k=1}^{\infty} \mathbb{1}_{\{S_k \leq t\}}.$$

Note that N_t counts the number of renewals in the interval $[0, t]$. The next theorem explains why the Poisson process was named after Poisson.

Theorem 3.3.2. *For all $t \geq 0$ it holds that $N_t \sim \mathrm{Poi}(\lambda t)$.*

Proof. We need to prove that for all $n \in \mathbb{N}_0$,

$$\mathbb{P}[N_t = n] = \frac{(\lambda t)^n}{n!} e^{-\lambda t}.$$

Step 1. Let first $n = 0$. Then,

$$\mathbb{P}[N_t = 0] = \mathbb{P}[\xi_1 > t] = e^{-\lambda t},$$

thus establishing the required formula for $n = 0$.

Step 2. Let $n \in \mathbb{N}$. We compute the probability $\mathbb{P}[N_t = n]$. By definition of N_t we have

$$\mathbb{P}[N_t = n] = \mathbb{P}[N_t \geq n] - \mathbb{P}[N_t \geq n+1] = \mathbb{P}[S_n \leq t] - \mathbb{P}[S_{n+1} \leq t].$$

Using the formula for the density of S_n we obtain that

$$\mathbb{P}[N_t = n] = \int_0^t f_{S_n}(x)dx - \int_0^t f_{S_{n+1}}(x)dx = \int_0^t \left(\frac{\lambda^n x^{n-1}}{(n-1)!} e^{-\lambda x} - \frac{\lambda^{n+1} x^n}{n!} e^{-\lambda x} \right) dx.$$

The expression under the sign of the integral is equal to

$$\frac{d}{dx} \left(\frac{(\lambda x)^n}{n!} e^{-\lambda x} \right).$$

Thus, we can compute the integral as follows:

$$\mathbb{P}[N_t = n] = \left(\frac{(\lambda x)^n}{n!} e^{-\lambda x} \right) \Big|_{x=0}^{x=t} = \frac{(\lambda t)^n}{n!} e^{-\lambda t},$$

where the last step holds since we assumed that $n \neq 0$. $\square$

REMARK 3.3.3. From the above theorem it follows that the renewal function of the Poisson process is given by $H(t) = \mathbb{E}N_t = \lambda t$.

For the next theorem let $U_1, \ldots, U_n$ be independent random variables which are uniformly distributed on the interval $[0, t]$. Denote by $U_{(1)} \leq \ldots \leq U_{(n)}$ the order statistics of $U_1, \ldots, U_n$.

THEOREM 3.3.4. *The conditional distribution of the random vector $(S_1, \ldots, S_n)$ given that $\{N_t = n\}$ coincides with the distribution of $(U_{(1)}, \ldots, U_{(n)})$:*

$$(S_1, \ldots, S_n)|\{N_t = n\} \stackrel{d}{=} (U_{(1)}, \ldots, U_{(n)}).$$

PROOF. We will compute the densities of both vectors and show these densities are equal.

STEP 1. The joint density of the random variables $(\xi_1, \ldots, \xi_{n+1})$ has (by independence) the product form

$$f_{\xi_1,\ldots,\xi_{n+1}}(u_1, \ldots, u_{n+1}) = \prod_{k=1}^{n+1} \lambda e^{-\lambda u_k}, \quad u_1, \ldots, u_{n+1} > 0.$$

STEP 2. We compute the joint density of $(S_1, \ldots, S_{n+1})$. Consider a linear transformation A defined by

$$A(u_1, u_2, \ldots, u_{n+1}) = (u_1, u_1 + u_2, \ldots, u_1 + \ldots + u_{n+1}).$$

The random variables $(S_1, \ldots, S_{n+1})$ can be obtained by applying the linear transformation A to the variables $(\xi_1, \ldots, \xi_{n+1})$:

$$(S_1, \ldots, S_{n+1}) = A(\xi_1, \ldots, \xi_{n+1}).$$

The determinant of the transformation A is 1 since the matrix of this transformation is triangular with 1's on the diagonal. By the density transformation theorem, the density of $(S_1, \ldots, S_{n+1})$ is given by

$$f_{S_1,\ldots,S_{n+1}}(t_1, \ldots, t_{n+1}) = \prod_{k=1}^{n+1} \lambda e^{-\lambda(t_k - t_{k-1})} = \lambda^{n+1} e^{-\lambda t_{n+1}},$$

where $0 = t_0 < t_1 < \ldots < t_{n+1}$. Otherwise, the density vanishes. Note that the formula for the density depends only on t_{n+1} and does not depend on $t_1, \ldots, t_n$.

STEP 3. We compute the conditional density of $(S_1, \ldots, S_n)$ given that $N_t = n$. Let $0 < t_1 < \ldots < t_n < t$. Intuitively, the conditional density of $(S_1, \ldots, S_n)$ given that $N_t = n$ is given by

$$\begin{aligned} f_{S_1,\ldots,S_n}(t_1, \ldots, t_n|N_t = n) &= \lim_{\varepsilon \downarrow 0} \frac{\mathbb{P}[t_1 < S_1 < t_1 + \varepsilon, \ldots, t_n < S_1 < t_n + \varepsilon | N_t = n]}{\varepsilon^n} \\ &= \lim_{\varepsilon \downarrow 0} \frac{\mathbb{P}[t_1 < S_1 < t_1 + \varepsilon, \ldots, t_n < S_n < t_n + \varepsilon, N_t = n]}{\varepsilon^n \mathbb{P}[N_t = n]} \\ &= \lim_{\varepsilon \downarrow 0} \frac{\mathbb{P}[t_1 < S_1 < t_1 + \varepsilon, \ldots, t_n < S_n < t_n + \varepsilon, S_{n+1} > t]}{\varepsilon^n \mathbb{P}[N_t = n]}. \end{aligned}$$

Using the formula for the joint density of $(S_1, \ldots, S_{n+1})$ and noting that this density does not depend on $t_1, \ldots, t_n$, we obtain that

$$\frac{\mathbb{P}[t_1 < S_1 < t_1 + \varepsilon, \ldots, t_n < S_n < t_n + \varepsilon, S_{n+1} > t]}{\varepsilon^n \mathbb{P}[N_t = n]} = \frac{\int_t^\infty \lambda^{n+1} e^{-\lambda t_{n+1}} dt_{n+1}}{\mathbb{P}[N_t = n]} = \frac{n!}{t^n},$$

where in the last step we used that N_t has Poisson distribution with parameter λt. So, we have

$$f_{S_1,\ldots,S_n}(t_1, \ldots, t_n | N_t = n) = \begin{cases} \frac{n!}{t^n}, & \text{for } 0 < t_1 < \ldots < t_n < t, \\ 0, & \text{otherwise.} \end{cases}$$

Step 4. The joint density of the order statistics $(U_{(1)}, \ldots, U_{(n)})$ is known (Stochastik I) to be given by

$$f_{U_{(1)},\ldots,U_{(n)}}(t_1, \ldots, t_n) = \begin{cases} \frac{n!}{t^n}, & \text{for } 0 < t_1 < \ldots < t_n < t, \\ 0, & \text{otherwise.} \end{cases}$$

This coincides with the conditional density of $(S_1, \ldots, S_n)$ given that $N_t = n$, thus proving the theorem. □

Theorem 3.3.5. *The Poisson process $\{N_t : t \geq 0\}$ has independent increments and these increments have Poisson distribution, namely for all $t, s \geq 0$ we have*

$$N_{t+s} - N_t \sim \text{Poi}(\lambda s).$$

Proof. Take some points $0 = t_0 \leq t_1 \leq \ldots \leq t_n$. We determine the distribution of the random vector

$$(N_{t_1}, N_{t_2} - N_{t_1}, \ldots, N_{t_n} - N_{t_{n-1}}).$$

Take some $x_1, \ldots, x_n \in \mathbb{N}_0$. We compute the probability

$$P := \mathbb{P}[N_{t_1} = x_1, N_{t_2} - N_{t_1} = x_2, \ldots, N_{t_n} - N_{t_{n-1}} = x_n].$$

Let $x = x_1 + \ldots + x_n$. By definition of conditional probability,

$$P = \mathbb{P}[N_{t_1} = x_1, N_{t_2} - N_{t_1} = x_2, \ldots, N_{t_n} - N_{t_{n-1}} = x_n | N_{t_n} = x] \cdot \mathbb{P}[N_{t_n} = x].$$

Given that $N_{t_n} = x$, the Poisson process has x renewals in the interval $[0, t_n]$ and by Theorem 3.3.4 these renewals have the same distribution as x independent random variables which have uniform distribution on the interval $[0, t_n]$, after arranging them in an increasing order. Hence, in order to compute the conditional probability we can use the multinomial distribution:

$$P = \left(\frac{x!}{x_1! \ldots x_n!} \prod_{k=1}^n \frac{(t_k - t_{k-1})^{x_k}}{t_n^{x_k}} \right) \cdot \frac{(\lambda t_n)^x}{x!} e^{-\lambda t_n}.$$

After making transformations we arrive at

$$P = \prod_{k=1}^n \left(\frac{(\lambda(t_k - t_{k-1}))^{x_k}}{x_k!} e^{-\lambda(t_k - t_{k-1})} \right).$$

From this formula we see that the random variables N_{t_1}, $N_{t_2} - N_{t_1}, \ldots$, $N_{t_n} - N_{t_{n-1}}$ are independent and that they are Poisson distributed, namely

$$N_{t_k} - N_{t_{k-1}} \sim \text{Poi}(\lambda(t_k - t_{k-1})).$$

This proves the theorem. □

THEOREM 3.3.6. *The Poisson process has stationary increments.*

PROOF. Take some $h \geq 0$, and some $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n$. We have to show that the distribution of the random vector

$$(N_{t_1+h} - N_{t_0+h}, N_{t_2+h} - N_{t_1+h}, \ldots, N_{t_n+h} - N_{t_{n-1}+h})$$

does not depend on h. However, we know from Theorem 3.3.5 that the components of this vector are independent and that

$$N_{t_k+h} - N_{t_{k-1}+h} \sim \text{Poi}(\lambda(t_k - t_{k-1})),$$

which does not depend on h. □

3.4. Lattice renewal processes

In this section we show how the theory of Markov chains can be used to obtain some properties of renewal processes whose interrenewal times are integer. Let $\xi_1, \xi_2, \ldots$ be independent and identically distributed random variables with values in $\mathbb{N} = \{1, 2, \ldots\}$. Let us write

$$r_n := \mathbb{P}[\xi_1 = n], \quad n \in \mathbb{N}.$$

We will make the aperiodicity assumption:

$$\gcd\{n \in \mathbb{N} : r_n \neq 0\} = 1. \tag{3.4.1}$$

For example, this condition excludes renewal processes for which the ξ_k's take only even values. Define the renewal times $S_n = \xi_1 + \ldots + \xi_n$, $n \in \mathbb{N}$.

THEOREM 3.4.1. *Let $m := \mathbb{E}\xi_1$ be finite. Then,*

$$\lim_{n\to\infty} \mathbb{P}[\exists k \in \mathbb{N} : S_k = n] = \frac{1}{m}.$$

So, the probability that there is a renewal at time n converges, as $n \to \infty$, to $\frac{1}{m}$.

PROOF. STEP 1. Consider a Markov chain defined as follows: Let

$$X_n = \inf\{t \geq n : t \text{ is renewal time}\} - n.$$

The random variable X_n (which is called the forward renewal time) represents the length of the time interval between n and the first renewal following n. (Please think why X_n has the Markov property). Note that at renewal times we have $X_n = 0$.
The state space of this chain is

$$E = \{0, 1, \ldots, M-1\}, \text{if } M < \infty,$$
$$E = \{0, 1, 2, \ldots\}, \text{if } M = \infty,$$

where M is the maximal value which the ξ_k's can attain:

$$M = \sup\{i \in \mathbb{N} : r_i > 0\} \in \mathbb{N} \cup \{\infty\}.$$

The transition probabilities of this Markov chain are given by

$$p_{i,i-1} = 1 \text{ for } i = 1, 2, \ldots, M-1,$$
$$p_{0,i} = r_{i+1} \text{ for } i = 1, \ldots, M-1.$$

STEP 2. We prove that the chain is irreducible. Starting at any state $i \in E$ we can reach state 0 by following the path

$$i \to i-1 \to i-2 \to \ldots \to 0.$$

So, every state leads to state 0. Let us prove that conversely, state 0 leads to every state. Let first M be finite. Starting in state 0 we can reach any state $i \in E$ with positive probability by following the path

$$0 \to M-1 \to M-2 \to \ldots \to i.$$

If M is infinite, then for every $i \in E$ we can find some $K > i$ such that $r_K > 0$. Starting at state 0 we can reach state i by following the path

$$0 \to K-1 \to K-2 \to \ldots \to i.$$

We have shown that every state leads to 0 and 0 leads to every state, so the chain is irreducible.

STEP 3. We prove that the chain is aperiodic. By irreducibility, we need to show that state 0 is aperiodic. For every i such that $r_i \neq 0$ we can go from 0 to 0 in i steps by following the path

$$0 \to i-1 \to i-2 \to \ldots \to 0.$$

By (3.4.1) the greatest common divisor of all such i's is 1, so the period of state 0 is 1 and it is aperiodic.

STEP 4. We claim that the unique invariant probability measure of this Markov chain is given by

$$\lambda_i = \frac{r_{i+1} + r_{i+2} + \ldots}{m}, \quad i \in E.$$

Indeed, the equations for the invariant probability measure look as follows:

$$\lambda_j = \sum_{i=0}^{M-1} p_{ij}\lambda_i = p_{0,j}\lambda_0 + p_{j+1,j}\lambda_{j+1} = r_{j+1}\lambda_0 + \lambda_{j+1}.$$

It follows that

$$\lambda_j - \lambda_{j+1} = r_{j+1}\lambda_0.$$

We obtain the following equations:

$$\begin{aligned}
\lambda_0 - \lambda_1 &= r_1\lambda_0, \\
\lambda_1 - \lambda_2 &= r_2\lambda_0, \\
\lambda_2 - \lambda_3 &= r_3\lambda_0, \\
&\ldots
\end{aligned}$$

By adding all these equations starting with the $(j+1)$-st one, we obtain that

$$\lambda_j = (r_{j+1} + r_{j+2} + \ldots)\lambda_0.$$

It remains to compute λ_0. By adding the equations for all $j = 0, 1, \ldots, M-1$ we obtain that

$$1 = \lambda_0 + \lambda_1 + \ldots = (r_1 + 2r_2 + 3r_3 + \ldots)\lambda_0 = m\lambda_0.$$

It follows that

$$\lambda_0 = \frac{1}{m}.$$

This proves the formula for the invariant probability distribution.

Step 5. Our chain is thus irreducible, aperiodic, and positive recurrent. By the theorem on the convergence to the invariant probability distribution we have

$$\lim_{n\to\infty} \mathbb{P}[X_n = 0] = \lambda_0 = \frac{1}{m}.$$

Recalling that we have $X_n = 0$ if and only if n is a renewal time, we obtain that

$$\lim_{n\to\infty} \mathbb{P}[\exists k \in \mathbb{N} : S_n = k] = \lim_{n\to\infty} \mathbb{P}[X_n = 0] = \frac{1}{m},$$

thus proving the claim of the theorem. □

CHAPTER 4

Brownian motion

Brownian motion is one of the most important and interesting stochastic processes. The history of the Brownian motion began in 1827 when the botanist Robert Brown looked through a microscope at small particles (pollen grains) suspended in water. He noted that the particles were moving chaotically. The mechanism causing this chaotic motion can be explained as follows. The particle collides with water molecules. Any collisions results in a displacement of the particle in some direction. The number of collisions is large, but the impact of any collision is small. To compute the total displacement of the particle caused by all collisions we have to add a very large number of very small random variables (impacts of individual collisions), like in the central limit theorem.

A similar situation appears when we try to model a price of an asset. The price, considered as a function of time, is subject to random changes due to the influence of some random events. If we assume that any random event has a very small impact on the price and that the number of events is very large, we are in the same situation when modelling the Brownian particle. This is why the Brownian motion is one of the main building blocks for stochastic processes used in financial mathematics.

In this chapter we will define a stochastic process $\{B(t)\colon t \geq 0\}$ (called the *Brownian motion* or the *Wiener process*) which is a mathematical model for the experiment described above.

4.1. Discrete approximation to the Brownian motion

Let us now try to model the motion of a small pollen grain particle in a fluid mathematically. First of all, we will model the motion of the particle in *one* dimension (that is, on the real line), because to model the motion in three dimensions we can model the three coordinates of the particle separately. So, we want to model a particle which moves on the real line due to random impacts which can shift the particle to the left or to the right. Assume without restriction of generality that at time 0 the particle starts at position 0. Denote by N the parameter describing the number of collisions of the particle with water molecules per unit time. This parameter should be very large. Assume that any collision causes a displacement of the particle by a distance $\delta > 0$ (which should be very small) either to the left or to the right, both possibilities having the same probability 1/2. A sample path of such particle (the coordinate of the particle as a function of time) is shown on Figure 1, left. Note that in this model we ignore the inertia of the particle. That is, the impacts are assumed to change the position of the particle, but we don't try to model the speed of the particle. This approach is justified if the fluid has large viscosity.

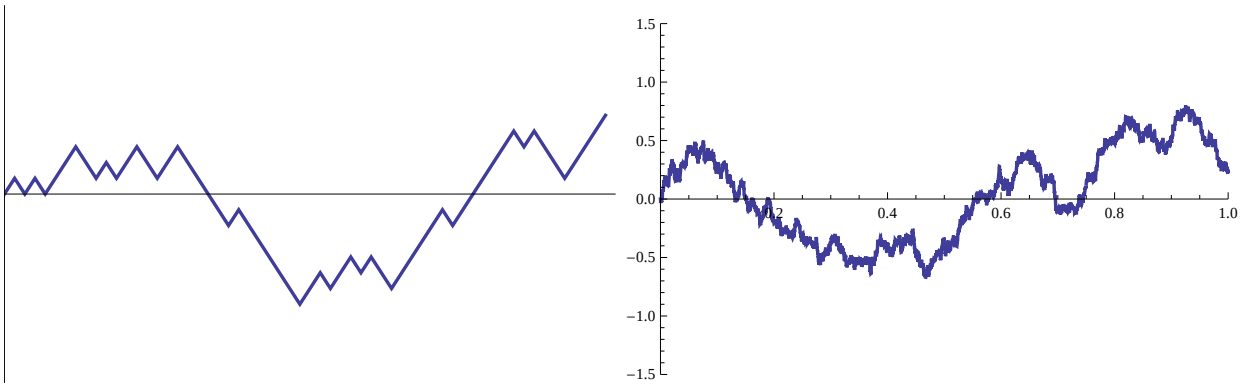


FIGURE 1. Left: A sample path of the process $B_{N,\delta}$. Right: A sample path of the Brownian motion

A more precise description of the model is as follows. Let $\xi_1, \xi_2, \ldots$ be independent and identically distributed random variables with

$$\mathbb{P}[\xi_i = +1] = \mathbb{P}[\xi_i = -1] = \frac{1}{2}.$$

Define a stochastic process $\{B_{N,\delta}(t) \colon t \geq 0\}$ describing the position of the particle at time t as follows. The position of the particle at time $t = \frac{k}{N}$, where $k \in \mathbb{N}_0$, is given by the sum of the first k impacts:

$$B_{N,\delta}\left(\frac{k}{N}\right) = \delta \cdot (\xi_1 + \ldots + \xi_k).$$

For $t \in (\frac{k}{N}, \frac{k+1}{N})$ we can define $B_{N,\delta}(t)$ by linear interpolation, as in Figure 1.

It is clear from the definition that the process $\{B_{N,\delta}(t) \colon t \geq 0\}$ has the following two properties:

(1) $B_{n,\delta}(0) = 0$.
(2) For every integer numbers $0 \leq k_1 \leq k_2 \leq \ldots \leq k_n$, the increments

$$B_{N,\delta}\left(\frac{k_1}{N}\right), B_{N,\delta}\left(\frac{k_2}{N}\right) - B_{N,\delta}\left(\frac{k_1}{N}\right), \ldots, B_{N,\delta}\left(\frac{k_n}{N}\right) - B_{N,\delta}\left(\frac{k_{n-1}}{N}\right)$$

are independent.

Let us now determine the approximate distribution of these increments. First of all, let us look at the position of the particle at time 1:

$$B_{N,\delta}(1) = \delta \cdot (\xi_1 + \ldots + \xi_N).$$

This position is a random variable and its expectation and variance are given by

$$\mathbb{E}B_{N,\delta}(1) = 0, \quad \operatorname{Var} B_{N,\delta}(1) = \delta^2 N.$$

Now, we want to see what happens in the scaling limit as $N \to \infty$ (meaning that the number of collisions of particle with water molecules is very large) and, at the same time, $\delta \to 0$ (meaning that the displacement caused by any collision is very small); see Figure 1, right. It is natural to require that $\operatorname{Var} B_{N,\delta}(1)$ should stay constant (independent of N and δ) because

otherwise we will not obtain any meaningful limit. We will choose this constant to be equal to 1 which leads to the requirement

$$\delta = \frac{1}{\sqrt{N}}.$$

If this relation holds, then by the central limit theorem we obtain that

$$B_{N,\delta}(1) = \frac{\xi_1 + \ldots + \xi_N}{\sqrt{N}} \xrightarrow[N\to\infty]{d} \mathrm{N}(0,1).$$

Similarly, for more general increments one obtains the following property:

$$B_{N,\delta}(t+h) - B_{N,\delta}(t) \xrightarrow[N\to\infty]{d} \mathrm{N}(0,h).$$

So, in the limit, the increments of our process should have the normal distribution.

4.2. Definition of the Brownian motion

The considerations of the preceding section make the following definition natural.

DEFINITION 4.2.1. A stochastic process $B = \{B(t)\colon t \geq 0\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called *Brownian motion* or *Wiener process* if

(1) $B(0) = 0$.
(2) B has *independent increments*, that is for all $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n$ the random variables
$$B(t_1), B(t_2) - B(t_1), \ldots, B(t_n) - B(t_{n-1})$$
are independent.
(3) B has *normal increments*, that is for all $t \geq 0$ and $h > 0$,
$$B(t+h) - B(t) \sim \mathrm{N}(0,h).$$
(4) B has *continuous sample paths*, that is for all $\omega \in \Omega$, the function $t \mapsto B(t;\omega)$ is continuous in t.

First of all, one has to ask whether a process satisfying these four requirements exists. This question is non-trivial and will be positively answered in Section 4.5 below. Here we sketch an idea of a possible approach to proving existence. The first three properties in the definition of the Brownian motion deal with the finite dimensional distributions of the process B only. It can be shown using Kolmogorov's existence theorem that a process with finite-dimensional distributions satisfying coonditions 1, 2, 3 exists. To be able to apply Kolmogorov's existence theorem one has to verify that the family of finite-dimensional distributions given by conditions 1, 2, 3 is consistent, that is that these conditions do not contradict each other. Essentially, this verification boils down to the following argument. If we know that for some $0 \leq t_1 \leq t_2 \leq t_3$ the increments

$$B(t_2) - B(t_1) \sim \mathrm{N}(0, t_2 - t_1) \text{ and } B(t_3) - B(t_2) \sim \mathrm{N}(0, t_3 - t_2)$$

are independent, then by the convolution property of the normal distribution, we must have

$$B(t_3) - B(t_1) = (B(t_3) - B(t_2)) + (B(t_2) - B(t_1)) \sim \mathrm{N}(0, (t_3 - t_2) + (t_2 - t_1)) = \mathrm{N}(0, t_3 - t_1).$$

Since this is in agreement with condition 3, there seems to be no contradiction between the conditions 1, 2, 3. Thus, we can apply Kolmogorov's existence theorem to construct a process satisfying conditions 1, 2, 3. However, Kolmogorov's theorem does not guarantee that the

resulting process satisfies condition 4, so that an additional modification of the construction is needed to make condition 4 satisfied. This is why we choose a different approach to prove the existence of a process satisfying conditions 1, 2, 3, 4; see Section 4.5.

The following example shows that it is not possible to drop condition 4 from the definition of the Brownian motion.

EXAMPLE 4.2.2. Assume that we have a process $\{B(t)\colon t \geq 0\}$ satisfying conditions 1, 2, 3, 4. We will show how, by modifying B, we can construct a process $\tilde{B}$ which satisfies properties 1, 2, 3, but violates property 4. This proves that property 4 is not a corollary of properties 1, 2, 3. Take a random variable $U \sim \mathrm{U}[0,1]$ independent of the process B. Define a new process $\{\tilde{B}(t)\colon t \geq 0\}$ by

$$\tilde{B}(t) = \begin{cases} B(t), & \text{if } t \neq U, \\ 0, & \text{if } t = U. \end{cases}$$

This process has the same finite-dimensional distributions as B. Indeed, the vectors

$$(B(t_1), \ldots, B(t_n)) \text{ and } (\tilde{B}(t_1), \ldots, \tilde{B}(t_n))$$

are equal unless $U \in \{t_1, \ldots, t_n\}$, but this event has probability 0. So, both random vectors are a.s. equal and hence, have the same distribution. This implies that the process $\{\tilde{B}(t)\colon t \geq 0\}$ also satisfies conditions 1, 2, 3. However, it does not satisfy condition 4 because the probability that its sample path is continuous is 0. Namely, we have

$$\lim_{t \to U, t \neq U} \tilde{B}(t) = \lim_{t \to U, t \neq U} B(t) = B(U).$$

This limit is a.s. different from $\tilde{B}(U) = 0$ because

$$\mathbb{P}[B(U) = 0] = \int_0^1 \mathbb{P}[B(u) = 0] du = \int_0^1 0 du = 0.$$

Thus, the probability that the sample path of $\tilde{B}$ has a discontinuity at U is 1.

4.3. Multivariate Gaussian distributions and Gaussian processes

It follows from the definition of the Brownian motion that its one-dimensional distributions are Gaussian, namely

$$B(t) \sim \mathrm{N}(0, t).$$

What about the multidimensional distributions of the Brownian motion? It turns out that these distributions are so-called multivariate Gaussian distributions. The aim of this section is to define the multivariate Gaussian distributions.

By definition, a random variable X has a (univariate) Gaussian distribution with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ (notation: $X \sim \mathrm{N}(\mu, \sigma^2)$) if the density of X has the form

$$f_X(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}, \quad t \in \mathbb{R}.$$

It is convenient to extend this definition to the case $\mu \in \mathbb{R}$, $\sigma^2 = 0$ by declaring $X \sim \mathrm{N}(\mu, 0)$ if $X = \mu$ almost surely. The characteristic function of a Gaussian random variable $X \sim$

$\mathrm{N}(\mu, \sigma^2)$ has the form

$$\varphi_X(s) = e^{is\mu - \frac{1}{2}\sigma^2 s^2}, \quad s \in \mathbb{R}.$$

The random variable X is called standard Gaussian if it is Gaussian with $\mu = 0$ and $\sigma^2 = 1$, that is if the density of X is given by

$$f_X(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}, \quad t \in \mathbb{R}.$$

We will now extend the definition of the Gaussian distribution from random variables to random vectors. Let us start with the definition of a standard Gaussian random vector.

DEFINITION 4.3.1. Fix dimension $d \in \mathbb{N}$. A random vector $X = (X_1, \ldots, X_d)^T$ is called *d-dimensional standard Gaussian* if

(1) $X_1, \ldots, X_d \sim \mathrm{N}(0, 1)$ are standard Gaussian random variables and
(2) $X_1, \ldots, X_d$ are independent random variables.

By independence, the joint density of a d-dimensional standard Gaussian vector X is given by

$$f_{X_1,\ldots,X_d}(t_1, \ldots, t_d) = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}(t_1^2 + \ldots + t_d^2)} = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}\langle t, t \rangle},$$

where $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$; see Figure 2.

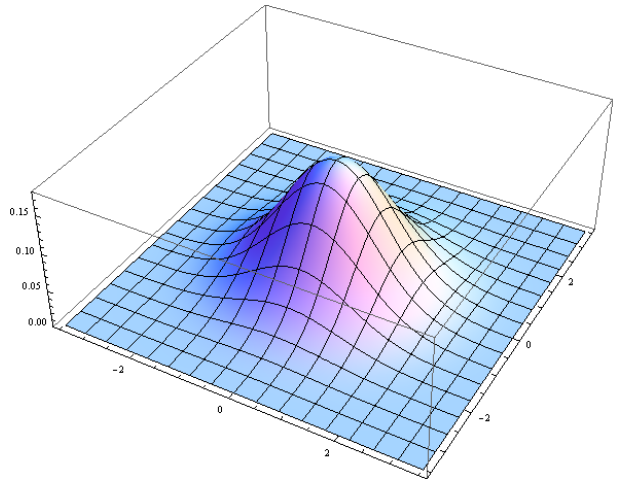


FIGURE 2. The 2-dimensional standard Gaussian density.

The expectation vector of X is equal to zero (because all components X_i have zero mean by definition). The covariance matrix of X is the $d \times d$-identity matrix (because the variance of any component X_i is 1 and different components are independent and hence uncorrelated):

$$\mathbb{E}X = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathrm{Cov}\, X = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}.$$

The next lemma states that the standard Gaussian distribution remains unchanged under rotations of the space around the origin.

LEMMA 4.3.2. *If X is d-dimensional standard Gaussian random vector and A an orthogonal $d \times d$-matrix, then the random vector AX is also standard Gaussian.*

PROOF. Recall that the orthogonality of the matrix A means that $AA^T = A^TA = \mathrm{Id}$. It follows that $\det A = \pm 1$ and in particular, A is invertible. By the transformation formula, the density of the random vector AX is

$$f_{AX}(t) = f_X(A^{-1}t)|\det(A^{-1})| = f_X(A^{-1}t) = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}\langle A^{-1}t, A^{-1}t\rangle} = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}\langle t,t\rangle} = f_X(t),$$

where we used that $\langle A^{-1}t, A^{-1}t\rangle = \langle (A^{-1})^T A^{-1}t, t\rangle = \langle (AA^T)^{-1}t, t\rangle = \langle t, t\rangle$. □

Then next lemma will be used in the construction of the Brownian motion in Section 4.5.

LEMMA 4.3.3. *Let X_1 and X_2 be independent Gaussian random variables with mean 0 and $\operatorname{Var} X_1 = \operatorname{Var} X_2 = \sigma^2$. Then, the random variables*

$$Y_1 = \frac{X_1 + X_2}{\sqrt{2}} \text{ and } Y_2 = \frac{X_1 - X_2}{\sqrt{2}}$$

are also independent and Gaussian with mean zero and variance σ^2.

PROOF. By definition, the random vector $(X_1/\sigma, X_2/\sigma)^T$ is 2-dimensional standard Gaussian. By Lemma 4.3.2, we obtain that the random vector

$$\begin{pmatrix} \frac{Y_1}{\sigma} \\ \frac{Y_2}{\sigma} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \frac{X_1}{\sigma} \\ \frac{X_2}{\sigma} \end{pmatrix}$$

is also two-dimensional standard Gaussian, because the matrix in the above equality is orthogonal. It follows that the random vector $(Y_1/\sigma, Y_2/\sigma)^T$ is also 2-dimensional standard Gaussian. Hence, the random variables Y_1/σ and Y_2/σ are independent and standard Gaussian. □

Now we are going to define the general (non-standard) multivariate Gaussian distribution. Essentially, we declare a random vector to be multivariate Gaussian if this random vector can be represented as an affine transform of some standard Gaussian random vector.

DEFINITION 4.3.4. A random vector $Y = (Y_1, \ldots, Y_d)^T$ is called *d-dimensional Gaussian* if there is some $m \in \mathbb{N}$, some m-dimensional *standard* Gaussian vector $X = (X_1, \ldots, X_m)^T$, some $d \times m$-matrix A and some $\mu \in \mathbb{R}^d$ so that

$$Y \stackrel{d}{=} AX + \mu.$$

EXERCISE 4.3.5. Show that the expectation and the covariance matrix of Y are given by

$$\mathbb{E}Y = \mu, \quad \operatorname{Cov} Y = AA^T.$$

NOTATION 4.3.6. We usually denote the covariance matrix by $\Sigma := \operatorname{Cov} Y = AA^T$ (not by Σ^2), and write $Y \sim \mathrm{N}_d(\mu, \Sigma)$. Note that the parameter μ takes values in $\mathbb{R}^d$, whereas the covariance matrix Σ can be any symmetric, positive semidefinite matrix.

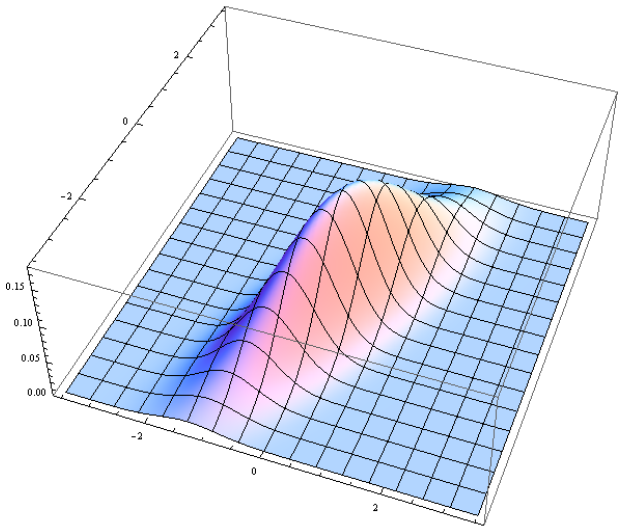


FIGURE 3. A two-dimensional (non-standard) Gaussian density

Any affine transformation of a Gaussian vector is again a Gaussian vector:

LEMMA 4.3.7. *If $Y \sim \mathrm{N}_d(\mu, \Sigma)$ is a d-dimensional Gaussian vector, A' is a $d' \times d$-matrix and $\mu' \in \mathbb{R}^{d'}$, then*

$$A'Y + \mu' \sim \mathrm{N}_{d'}(A'\mu + \mu', A'\Sigma A'^T)$$

PROOF. By definition, we can represent Y in the form $Y = AX + \mu$, where $AA^T = \Sigma$ and the vector X is m-dimensional standard Gaussian. The d'-dimensional random vector

$$A'Y + \mu' = A'(AX + \mu) + \mu' = (A'A)X + (A'\mu + \mu')$$

is also an affine transform of X and hence, multivariate Gaussian. The parameters of $A'Y + \mu'$ are given by

$$\mathbb{E}[A'Y + \mu'] = A'\mu + \mu', \quad \mathrm{Cov}(A'Y + \mu') = (A'A)(A'A)^T = A'AA^TA'^T = A'\Sigma A'^T.$$

□

REMARK 4.3.8. In particular, any component Y_i of a Gaussian random vector $(Y_1, \ldots, Y_d)^T$ is a Gaussian random variable. The converse is not true: If $Y_1, \ldots, Y_d$ are Gaussian random variables, then it's in general not true that $(Y_1, \ldots, Y_d)^T$ is a Gaussian random vector. However, if we additionally require that $Y_1, \ldots, Y_d$ should be independent, the statement becomes true.

LEMMA 4.3.9. *Let $Y_1, \ldots, Y_d$ be independent Gaussian random variables. Then, $(Y_1, \ldots, Y_d)^T$ is a Gaussian random vector.*

PROOF. Let $Y_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$. Then, we can write $Y_i = \sigma_i X_i + \mu_i$, where X_i are standard normal and independent. So, the random vector $(Y_1, \ldots, Y_d)^T$ is an affine transformation of some standard Gaussian random vector $(X_1, \ldots, X_d)^T$ and hence, itself d-dimensional Gaussian. □

LEMMA 4.3.10. *The characteristic function of a d-dimensional Gaussian random vector* $Y \sim \mathrm{N}_d(\mu, \Sigma)$ *is given by*

$$\varphi_Y(t) := \mathbb{E}e^{i\langle t, Y\rangle} = e^{i\langle \mu, t\rangle - \frac{1}{2}\langle t, \Sigma t\rangle}, \quad t \in \mathbb{R}^d.$$

PROOF. Fix $t = (t_1, \ldots, t_d) \in \mathbb{R}^d$. The mapping $y \mapsto \langle t, y\rangle$ is a linear map from $\mathbb{R}^d$ to $\mathbb{R}$ whose matrix is given by $(t_1, \ldots, t_d)$. By Lemma 4.3.7, the random variable $Z := \langle t, Y\rangle$ is Gaussian with expectation $\langle \mu, t\rangle$ and variance $\langle t, \Sigma t\rangle$. We have

$$\varphi_Y(t) = \mathbb{E}e^{i\langle t, Y\rangle} = \mathbb{E}e^{iZ} = \varphi_Z(1) = e^{i\langle \mu, t\rangle - \frac{1}{2}\langle t, \Sigma t\rangle}.$$

where in the last step we used the known formula for the characteristic function of the Gaussian random variable Z. □

EXERCISE 4.3.11. Let $X_1, X_2, \ldots$ be a sequence of d-dimensional Gaussian vectors whose expectations μ_n converge to μ and covariance matrices Σ_n converge to Σ. Show that X_n converges in distribution to $\mathrm{N}_d(\mu, \Sigma)$.

What is the density of a multivariate Gaussian distribution $\mathrm{N}_d(\mu, \Sigma)$? First of all, this density does not always exist, as the following example shows.

EXAMPLE 4.3.12. Let us construct an example of a two-dimensional Gaussian random vector which has no density. Let X be a standard normal random variable. The two-dimensional vector $(X, X)^T$ is Gaussian because it can be represented as a linear transformation AX, where

$$A : x \mapsto \begin{pmatrix} x \\ x \end{pmatrix}.$$

However, the random vector $(X, X)^T$ has no density (with respect to the two-dimensional Lebesgue measure) because X takes values in the line $\{(x, x) : x \in \mathbb{R}\}$ which has Lebesgue measure 0. Note that the covariance matrix of $(X, X)^T$ is equal to

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

This matrix is degenerate, meaning that its determinant is 0.

The next lemma gives a formula for the density of the multivariate Gaussian distribution in the case when Σ a non-degenerate matrix.

LEMMA 4.3.13. *The density of a d-dimensional Gaussian random vector* $Y \sim \mathrm{N}_d(\mu, \Sigma)$, *where* Σ *is a non-degenerate matrix, is given by*

$$f_Y(t) = \frac{1}{(\sqrt{2\pi})^d \sqrt{\det \Sigma}} e^{-\frac{1}{2}\langle t-\mu, \Sigma^{-1}(t-\mu)\rangle}.$$

If the matrix Σ *is degenerate, then* Y *has no density with respect to the Lebesgue measure on* $\mathbb{R}^d$.

PROOF. Since the matrix Σ is positive semidefinite, we can write $\Sigma = \Sigma^{1/2} \cdot \Sigma^{1/2}$ for some symmetric matrix $\Sigma^{1/2}$. We have the representation

$$Y \stackrel{d}{=} \Sigma^{1/2} X + \mu,$$

where X is a standard Gaussian vector on $\mathbb{R}^d$. Consider the transformation

$$T : \mathbb{R}^d \to \mathbb{R}^d, \quad x \mapsto \Sigma^{1/2}x + \mu.$$

Then, $T(X) \stackrel{d}{=} Y$.

1. If Σ is degenerate, then the image of T is a subspace of $\mathbb{R}^d$ having dimension strictly smaller than d. It follows that the image of T has Lebesgue measure 0. So, Y takes values in a subset of $\mathbb{R}^d$ which has Lebesgue measure 0. It follows that Y has no density.

2. If we assume that $\det \Sigma \neq 0$, we have the inverse transformation

$$T^{-1}(y) = \Sigma^{-1/2}(y - \mu).$$

The density of X is

$$f_X(x) = \frac{1}{(\sqrt{2\pi})^d} e^{-\frac{1}{2}\langle x,x\rangle}, \quad x \in \mathbb{R}^d.$$

Now we can compute the density of Y by using the transformation of density theorem:

$$f_Y(y) = f_X(T^{-1}(y))|\det T^{-1}| = \frac{1}{(\sqrt{2\pi})^d\sqrt{\det \Sigma}} e^{-\frac{1}{2}\langle \Sigma^{-1/2}(y-\mu),\Sigma^{-1/2}(y-\mu)\rangle}, \quad y \in \mathbb{R}^d.$$

Using the symmetry of the matrix $\Sigma^{1/2}$, we obtain

$$f_Y(y) = \frac{1}{(\sqrt{2\pi})^d\sqrt{\det \Sigma}} e^{-\frac{1}{2}\langle (y-\mu),\Sigma^{-1}(y-\mu)\rangle}, \quad y \in \mathbb{R}^d.$$

which is the required formula. □

For general random vectors it is known that the independence of components implies their uncorrelatedness, but the converse is, generally speaking, not true. It is an important property of the multivariate Gaussian distribution that for this distribution, the independence and the uncorrelatedness of the components are *equivalent*.

THEOREM 4.3.14. *Let $Y = (Y_1, \ldots, Y_d)^T$ be a random vector with multivariate Gaussian distribution. Then, the following properties are equivalent:*

(1) *The random variables $Y_1, \ldots, Y_d$ are independent.*
(2) $\mathrm{Cov}(Y_i, Y_j) = 0$ *for all $i \neq j$.*

PROOF. It is known that (1) implies (2) even without the multivariate Gaussian assumption. We prove that (2) implies (1). Assume that $\mathrm{Cov}(Y_i, Y_j) = 0$ for all $i \neq j$. The components Y_k are Gaussian, say $Y_k \sim \mathrm{N}(\mu_k, \sigma_k^2)$. By the uncorrelatedness, the covariance matrix of Y is a diagonal matrix, whereas the expectation vector of Y may be, in general, arbitrary:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \ldots & 0 \\ 0 & \sigma_2^2 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & \sigma_d^2 \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{pmatrix}.$$

The characteristic function of Y is given by

$$\varphi_{Y_1,\ldots,Y_d}(t_1, \ldots, t_d) = e^{i\langle \mu,t\rangle - \frac{1}{2}\langle t,\Sigma t\rangle} = e^{i\sum_{k=1}^d \mu_k t_k - \frac{1}{2}\sum_{k=1}^d \sigma_k^2 t_k^2} = \prod_{k=1}^d e^{i\mu_k t_k - \frac{1}{2}\sigma_k^2 t_k^2} = \prod_{k=1}^d \varphi_{Y_k}(t_k).$$

This implies that $Y_1, \ldots, Y_d$ are independent. $\square$

Recall that two random vectors $X = (X_1, \ldots, X_n)^T$ and $Y = (Y_1, \ldots, Y_m)^T$ defined on a common probability space are called independent if for every Borel sets $A \subset \mathbb{R}^n$ and $B \subset \mathbb{R}^m$ we have

$$\mathbb{P}[X \in A, Y \in B] = \mathbb{P}[X \in A] \cdot \mathbb{P}[Y \in B].$$

EXERCISE 4.3.15. Let $(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ be a Gaussian random vector. Show that the random vectors $(X_1, \ldots, X_n)$ and $(Y_1, \ldots, Y_m)$ are independent if and only if

$$\mathrm{Cov}(X_i, Y_j) = 0$$

for all $i = 1, \ldots, n$ and $j = 1, \ldots, m$.

4.4. Brownian motion as a Gaussian process

A stochastic process is called Gaussian if its finite-dimensional distributions are multivariate Gaussian. More precisely:

DEFINITION 4.4.1. A stochastic process $\{X(t) \colon t \in T\}$ is called Gaussian if for every $n \in \mathbb{N}$ and every $t_1, \ldots, t_n \in T$, the random vector $(X(t_1), \ldots, X(t_n))^T$ is n-dimensional Gaussian.

EXAMPLE 4.4.2. Let us show that the Brownian motion is a Gaussian process. Take some $0 \le t_1 \le t_2 \le \ldots \le t_n$. We show that the vector $(B(t_1), \ldots, B(t_n))$ is Gaussian. Consider the random variables

$$\Delta_i = B(t_i) - B(t_{i-1}).$$

By the definition of the Brownian motion, these random variables are independent and each has Gaussian distribution. It follows from Lemma 4.3.9 that the random vector $(\Delta_1, \ldots, \Delta_n)$ is n-dimensional Gaussian. We can represent $(B(t_1), \ldots, B(t_n))$ as a linear transform of $(\Delta_1, \ldots, \Delta_n)$:

$$B(t_i) = \Delta_1 + \ldots + \Delta_i.$$

It follows from Lemma 4.3.7 that the vector $(B(t_1), \ldots, B(t_n))$ is also n-dimensional Gaussian.

REMARK 4.4.3. The finite dimensional distributions of a Gaussian process are uniquely determined by the expectation function $\mu(t) = \mathbb{E}X(t)$ and the covariance function

$$\Gamma(t_1, t_2) = \mathrm{Cov}(X(t_1), X(t_2)).$$

EXAMPLE 4.4.4. If B is a Brownian motion, then

$$\mathbb{E}B(t) = 0, \quad \Gamma(t_1, t_2) = \min(t_1, t_2).$$

Conversely, we have the following characterization of the Brownian motion.

THEOREM 4.4.5. *A stochastic process $\{B(t) \colon t \ge 0\}$ is a Brownian motion if and only if*

(1) *B is Gaussian;*
(2) *$\mathbb{E}B(t) = 0$ for all $t \ge 0$;*
(3) *$\mathrm{Cov}(B(t_1), B(t_2)) = \min(t_1, t_2)$ for all $t_1, t_2 \ge 0$;*
(4) *B has continuous sample paths.*

PROOF. It is an exercise to show that the above four conditions are equivalent to the conditions from the definition of the Brownian motion. □

The next theorem is called the *weak Markov property* of the Brownian motion.

THEOREM 4.4.6. *Let $\{B(t)\colon t \geq 0\}$ be a Brownian motion. Fix some $u \geq 0$. Then:*

(1) *The process $B_u(s) = B(u+s) - B(u), s \geq 0$, is also a Brownian motion.*
(2) *The processes $\{B(t)\colon 0 \leq t \leq u\}$ and $\{B_u(s)\colon s \geq 0\}$ are independent.*

PROOF. We will verify conditions of Theorem 4.4.5. The process B_u is Gaussian. Indeed, for every $s_1, \ldots, s_n$, the random vector $(B_u(s_1), \ldots, B_u(s_n))$ can be written as a linear transformation of the Gaussian random vector $(B(u+t_1), \ldots, B(u+t_n), B(u))$. Also, the process B_u has continuous sample paths because B does so by definition of the Brownian motion. In order to show that B_u is a Brownian motion, we compute the expectation and the covariance function of B_u. The expectation is given by

$$\mathbb{E}B_u(s) = \mathbb{E}(B(u+s) - B(u)) = 0.$$

The covariance function is given by

$$\begin{aligned}\operatorname{Cov}(B_u(s_1), B_u(s_2)) &= \operatorname{Cov}(B(u+s_1) - B(u), B(u+s_2) - B(u)) \\ &= \min(u+s_1, u+s_2) - u - u + u \\ &= \min(s_1, s_2).\end{aligned}$$

So, B_u is a Brownian motion.

We prove that the processes $\{B(t)\colon 0 \leq t \leq u\}$ and $\{B_u(s)\colon s \geq 0\}$ are independent. First of all, we need to define what does this mean.

DEFINITION 4.4.7. Two stochastic process $\{X(t)\colon t \in T\}$ and $\{Y(s)\colon s \in S\}$ defined on the same probability space are called independent if for all $t_1, \ldots, t_n \in T$ and $s_1, \ldots, s_m \in S$ the vector $(X(t_1), \ldots, X(t_n))$ is independent of $(Y(s_1), \ldots, Y(s_m))$.

To show that the processes $\{B(t)\colon 0 \leq t \leq u\}$ and $\{B_u(s)\colon s \geq 0\}$ are independent, it suffices to show that there is no correlation between these two processes. Take some $0 \leq t \leq u$ and $s \geq 0$. Then,

$$\operatorname{Cov}(B(t), B_u(s)) = \operatorname{Cov}(B(t), B(u+s) - B(u)) = t_i - t_i = 0.$$

This proves the independence. □

The next theorem states the *self-similarity property* of the Brownian motion.

THEOREM 4.4.8. *Let $\{B(t)\colon t \geq 0\}$ be a Brownian motion and let $a > 0$. Then, the process*

$$\left\{\frac{B(at)}{\sqrt{a}}\colon t \geq 0\right\}$$

is again a Brownian motion.

PROOF. Exercise.

4.5. Lévy's construction of the Brownian motion

THEOREM 4.5.1. *The Brownian motion exists. Concretely: It is possible to construct a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a stochastic process $\{B(t)\colon t \geq 0\}$ on this probability space such that*

(1) $B(0) = 0$.
(2) *B has independent increments.*
(3) $B(t+h) - B(t) \sim \mathrm{N}(0, h)$ *for all* $t, h \geq 0$.
(4) *For every $\omega \in \Omega$ the function $t \mapsto B(t;\omega)$ is continuous in t.*

PROOF. First we will show how to construct the Brownian motion for $t \in [0,1]$.

STEP 1: Construction on the set of dyadic rationals. Consider the sets

$$D_n = \left\{ \frac{k}{2^n} \colon k = 0, 1, \ldots, 2^n \right\}, \quad n \in \mathbb{N}_0.$$

The first few sets are given by

$$D_0 = \{0, 1\}, \quad D_1 = \left\{0, \frac{1}{2}, 1\right\}, \quad D_2 = \left\{0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1\right\}, \quad \ldots$$

Note also that $D_0 \subset D_1 \subset \ldots$. Let D be the set of dyadic rationals in $[0,1]$:

$$D = \bigcup_{n=0}^{\infty} D_n.$$

By Kolmogorov's existence theorem, we can construct a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ carrying a collection $\{Z_t \colon t \in D\}$ of independent standard normal random variables indexed by D.

For every $n \in \mathbb{N}_0$ we will construct a family of random variables $\{B(d) \colon d \in D_n\}$ such that

(1) For all $r < s < t$ in D_n the random variables $B(t) - B(s) \sim \mathrm{N}(0, t-s)$ and $B(s) - B(r) \sim \mathrm{N}(0, s-r)$ are independent.
(2) The processes $\{B(d) \colon d \in D_n\}$ and $\{Z_t \colon t \in D \backslash D_n\}$ are independent.

We use induction over n.

Case $n = 0$: For $n = 0$ we define $B(0) = 0, B(1) = Z_1 \sim \mathrm{N}(0,1)$.

Passing from $n-1$ to n: Assume we have constructed $\{B(d) \colon d \in D_{n-1}\}$ for which Properties (1) and (2) hold. We construct $\{B(d) \colon d \in D_n\}$ as follows. For $d \in D_n \backslash D_{n-1}$ define

$$B(d) = \frac{B(d - 2^{-n}) + B(d + 2^{-n})}{2} + \frac{Z_d}{2^{\frac{n+1}{2}}}.$$

Property (2) holds because for $d \in D_n$, the random variable $B(d)$ is defined in terms of the random variables $\{Z_t \colon t \in D_n\}$ only.

We prove Property (1). Define random variables:

$$X_1 := \frac{B(d + 2^{-n}) - B(d - 2^{-n})}{2} \sim \mathrm{N}(0, 2^{-n-1}),$$
$$X_2 := \frac{Z_d}{2^{\frac{n+1}{2}}} \sim \mathrm{N}(0, 2^{-n-1}).$$

The random variables X_1 and X_2 are independent by Property (2) (which, by the induction assumption, holds for $n-1$ instead of n). By Lemma 4.3.3, the random variables

$$X_1 + X_2 = B(d) - B(d - 2^{-n}) \sim \mathrm{N}(0, 2^{-n}),$$
$$X_1 - X_2 = B(d - 2^{-n}) - B(d) \sim \mathrm{N}(0, 2^{-n})$$

are independent.

The above shows that any two "neighboring" increments of the form $B(d) - B(d - 2^{-n})$, $B(d-2^{-n})-B(d)$, where $d \in D_n \backslash D_{n-1}$, are independent. In fact, we show that all increments

$$B(d) - B(d - 2^{-n}), \quad d \in D_n \backslash \{0\},$$

are independent. This implies Property (1). The vector formed by these increments is Gaussian since it is a linear transform of the standard Gaussian vector $\{Z_t \colon t \in D_n\}$. Consider two intervals of the form

$$I_1 = [d_1 - 2^{-n}, d_1], \quad I_2 = [d_2 - 2^{-n}, d_2], \quad d_1, d_2 \in D_n \backslash \{0\}, \quad d_1 < d_2.$$

They are separated by some $d \in D_j$, where we choose j to be minimal with this property. We prove that the increments of B over these intervals are independent. We have considered the case when $j = n$ above. Therefore, let $j < n$. The intervals I_1 and I_2 are contained in $K_1 = [d - 2^{-j}, d]$ and $K_2 = [d + 2^{-j}, d]$ since otherwise, we could replace d by $d \pm 2^{-j}$ which has smaller j. By the induction assumption, the increments of B over the intervals K_1 and K_2 are independent. The increments over the intervals I_1 and I_2 are defined using the increments over K_1 and K_2 and some disjoint subsets of the family $\{Z_t \colon t \in D_n\}$. Hence, the increments over I_1 and I_2 are independent.

This completes the construction of $\{B(t) \colon t \in D\}$.

Step 2: Extending the construction to $[0, 1]$. Define a sequence $F_0, F_1, \ldots$ of random functions on the interval $[0, 1]$ as follows. Let $F_0(t) = Z_1 t$, for $t \in [0, 1]$. Further, define

$$F_n(t) = \begin{cases} 0, & t \in D_{n-1}, \\ 2^{-\frac{n+1}{2}} Z_t, & t \in D_n \backslash D_{n-1}, \end{cases}$$

and let $F_n(t)$ be defined by linear interpolation between the points from D_n.

For $d \in D_n$ we defined in Step 1

$$B(d) = \sum_{i=0}^{n} F_i(d) = \sum_{i=0}^{\infty} F_i(d).$$

We prove that there is a measurable set $\Omega_1 \subset \Omega$ with $\mathbb{P}[\Omega_1] = 1$ such that for all $\omega \in \Omega_1$ there exists $N = N(\omega) \in \mathbb{N}$ such that for all $n > N$,

$$\sup_{t \in [0,1]} |F_n(t)| \le 3\sqrt{n} 2^{-n/2}. \tag{4.5.1}$$

Let us prove (4.5.1). Let $c > 1$. Then, for large enough n,

$$\mathbb{P}[|Z_d| > c\sqrt{n}] = 2\mathbb{P}[Z_d > c\sqrt{n}] \le 2e^{-c^2 n/2}. \tag{4.5.2}$$

Here, we used the asymptotics

$$\mathbb{P}[Z_d > x] \sim \frac{1}{\sqrt{2\pi}x} e^{-x^2/2}, \quad x \to \infty,$$

which can be proven using the L'Hôspital rule. We have, using (4.5.2),

$$\sum_{n=0}^{\infty} \mathbb{P}[\exists d \in D_n \colon |Z_d| \geq c\sqrt{n}] \leq \sum_{n=0}^{\infty} \sum_{d \in D_n} \mathbb{P}[|Z_d| \geq c\sqrt{n}] \leq C + \sum_{n=0}^{\infty} (2^n + 1) \cdot 2e^{-c^2 n/2} < \infty,$$

where the last step holds if $c > \sqrt{2 \log 2}$, for example, if $c = 3$. By the Borel-Cantelli lemma, we obtain that (4.5.1) holds.

It follows from (4.5.1) that for all $\omega \in \Omega_1$ the series $\sum_{n=0}^{\infty} F_n(t;\omega)$ converges uniformly over $t \in [0,1]$. The sum of the series is denoted by $B(t;\omega)$. Since the sum of a uniformly convergent series of continuous functions is continuous, we have that for all $\omega \in \Omega_1$ the function $t \mapsto B(t;\omega)$ is continuous.

Step 3: We show that the process $\{B(t) \colon t \in [0,1]\}$ constructed in Step 2 has independent and normal increments. Take some $0 \leq t_1 \leq \ldots \leq t_n \leq 1$. Since the set D is dense in $[0,1]$ we can find for every $k \in \mathbb{N}$ dyadic rationals $0 \leq t_{1,k} \leq \ldots \leq t_{n,k} \leq 1$ so that $\lim_{k\to\infty} t_{i,k} = t_i$ for all $i = 1, \ldots, n$. By the continuity of B we have

$$\Delta_i := B(t_i) - B(t_{i-1}) = \lim_{k\to\infty} (B(t_{i,k}) - B(t_{i-1,k})) = \lim_{k\to\infty} \Delta_{i,k},$$

where $\Delta_{i,k} := B(t_{i,k}) - B(t_{i-1,k})$ and we put $t_0 = t_{0,k} = 0$. The vector $(\Delta_{1,k}, \ldots, \Delta_{n,k})$ is Gaussian by the construction from Step 1, with mean 0. Again, by the construction of Step 1, we have

$$\operatorname{Cov}(\Delta_{i,k}, \Delta_{j,k}) = (t_{i,k} - t_{i-1,k}) \mathbb{1}_{i=j} \to (t_i - t_{i-1}) \mathbb{1}_{i=j}, \quad \text{as } k \to \infty.$$

It follows (see Exercise 4.3.11), the random vector $(\Delta_1, \ldots, \Delta_n)$ is also Gaussian, with mean 0 and with covariance matrix

$$\operatorname{Cov}(\Delta_i, \Delta_j) = (t_i - t_{i-1}) \mathbb{1}_{i=j}.$$

In particular, the components of this vector are independent and the variance of Δ_i is $t_i - t_{i-1}$. This proves that $\{B(t) \colon t \in [0,1]\}$ has independent increments and that $B(t+h) - B(t) \sim \mathrm{N}(0,h)$.

Step 4: We extend the construction to all $t \geq 0$. Take independent copies $B_0 = B, B_1, B_2, \ldots$ of the process $\{B(t) \colon t \in [0,1]\}$ constructed in Steps 1–3 and glue them together. Concretely, for $t \in [n, n+1]$ define

$$B(t) = B_n(t-n) + \sum_{i=0}^{n-1} B_i(1).$$

The process $\{B(t) \colon t \geq 0\}$ defined in this way is Gaussian and has continuous sample paths. It is not difficult to check that its covariance function coincides with the covariance function of the Brownian motion. So, the process $\{B(t) \colon t \geq 0\}$ is the Brownian motion. $\square$

4.6. Non-differentiability of Brownian motions paths

THEOREM 4.6.1 (Paley, Wiener, Zygmund). *Let $\{B(t)\colon t \geq 0\}$ be a Brownian motion defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, with probability 1, the function $t \to B(t)$ is nowhere differentiable. Concretely: There is a measurable set $\Omega' \subset \Omega$ with $\mathbb{P}[\Omega'] = 1$ such that for all $\omega \in \Omega'$ and for all $t_0 \geq 0$ the function $t \to B(t;\omega)$ has no derivate at t_0.*

REMARK 4.6.2. We will prove even more. For a function $f : \mathbb{R} \to \mathbb{R}$ define

$$D^+f(t) = \limsup_{h \downarrow 0} \frac{f(t+h) - f(t)}{h} \text{ (upper right derivative)},$$

$$D^-f(t) = \liminf_{h \downarrow 0} \frac{f(t+h) - f(t)}{h} \text{ (lower right derivative)}.$$

If $D^+f(t) = D^-f(t)$ is finite, then we say that f is differentiable from the right. In a similar way one can define the upper left derivative and the lower left derivative. Consider the set

$$A := \{\omega \in \Omega : \exists t_0 \in [0,1] \text{ such that } -\infty < D^-B(t_0;\omega) \leq D^+B(t_0;\omega) < +\infty\}.$$

We would like to show that $\mathbb{P}[A] = 0$, that is for almost every sample path of the Brownian motion and for every $t_0 \geq 0$ we have $D^+B(t_0) = +\infty$, or $D^-B(t_0) = -\infty$, or both. However, it is not immediately clear whether the set A is measurable. Therefore, we will prove a somewhat weaker statement: There is a measurable set A' with $\mathbb{P}[A'] = 0$ such that $A \subset A'$.

PROOF. We have $A \subset \cup_{M \in \mathbb{N}} A_M$, where

$$A_M = \left\{ \omega \in \Omega : \exists t_0 \in [0,1] \text{ such that } \sup_{h \in [0,1]} \left| \frac{B(t_0+h) - B(t_0)}{h} \right| \leq M \right\}.$$

Fix some $M \in \mathbb{N}$. We show that $\mathbb{P}[A_M] = 0$. Take some $n \in \mathbb{N}, n \geq 3$. Any $t_0 \in [0,1]$ must be in some interval $t_0 \in [\frac{k-1}{2^n}, \frac{k}{2^n}], k = 1, \ldots, 2^n$. If the event A_M occurs and $t_0 \in [\frac{k-1}{2^n}, \frac{k}{2^n}]$, then the following three events also occur:

(1) $F^{(1)}_{n,k} : |B(\frac{k+1}{2^n}) - B(\frac{k}{2^n})| \leq |B(\frac{k+1}{2^n}) - B(t_0)| + |B(t_0) - B(\frac{k}{2^n})| \leq \frac{3M}{2^n}$.
(2) $F^{(2)}_{n,k} : |B(\frac{k+2}{2^n}) - B(\frac{k+1}{2^n})| \leq |B(\frac{k+2}{2^n}) - B(t_0)| + |B(t_0) - B(\frac{k+1}{2^n})| \leq \frac{5M}{2^n}$.
(3) $F^{(3)}_{n,k} : |B(\frac{k+3}{2^n}) - B(\frac{k+2}{2^n})| \leq |B(\frac{k+3}{2^n}) - B(t_0)| + |B(t_0) - B(\frac{k+2}{2^n})| \leq \frac{7M}{2^n}$.

Consider the event $F_{n,k} = F^{(1)}_{n,k} \cap F^{(2)}_{n,k} \cap F^{(3)}_{n,k}$. Then, for every $n \geq 3$ we have

$$A_M \subset \bigcup_{k=1}^{2^n} F_{n,k}.$$

We will estimate the probabilities $\mathbb{P}[F^{(1)}_{n,k}]$, $\mathbb{P}[F^{(2)}_{n,k}]$, $\mathbb{P}[F^{(3)}_{n,k}]$. For example, for $\mathbb{P}[F^{(3)}_{n,k}]$ we have

$$\mathbb{P}[F^{(3)}_{n,k}] = \mathbb{P}\left[\left|B\left(\frac{k+3}{2^n}\right) - B\left(\frac{k+2}{2^n}\right)\right| \leq \frac{7M}{2^n}\right] = \mathbb{P}\left[\frac{|N|}{\sqrt{2^n}} < \frac{7M}{2^n}\right] = \mathbb{P}\left[|N| \leq \frac{7M}{\sqrt{2^n}}\right],$$

where N is a standard normal random variable. Denoting by $f_N(t)$ its density (which is smaller than $1/\sqrt{2\pi} < 1/2$), we have

$$\mathbb{P}[F^{(3)}_{n,k}] = \int_{-\frac{7M}{2^{n/2}}}^{\frac{7M}{2^{n/2}}} f_N(t)dt \leq \frac{7M}{2^{n/2}}.$$

Similarily, one shows that

$$\mathbb{P}[F_{n,k}^{(1)}] \leq \frac{7M}{2^n}, \quad \mathbb{P}[F_{n,k}^{(2)}] \leq \frac{7M}{2^{n/2}}.$$

Since the events $F_{n,k}^{(1)}$, $F_{n,k}^{(2)}$, $F_{n,k}^{(3)}$ are independent (by the independence of increments of the Brownian motion), we have

$$\mathbb{P}[F_{n,k}] = \mathbb{P}[F_{n,k}^{(1)}] \cdot \mathbb{P}[F_{n,k}^{(2)}] \cdot \mathbb{P}[F_{n,k}^{(3)}] \leq \frac{(7M)^3}{2^{3n/2}}.$$

It follows that

$$\mathbb{P}[A_M] \leq \mathbb{P}[\cup_{k=1}^{2^n} F_{n,k}] \leq 2^n \frac{(7M)^3}{2^{3n/2}} = \frac{(7M)^3}{2^{n/2}}.$$

Since this holds for every $n \geq 3$, we have $\mathbb{P}[A_M] = 0$ and hence, the set $A' := \cup_{M \in \mathbb{N}} A_M$ has probability 0. We can now take $\Omega' = \Omega \backslash A'$. $\square$