C. Kraaikamp

PROBSTAT-IIii*

SPIN Springer's internal project number, if known

– Monograph –

31st May 2005

Springer

Berlin Heidelberg New York
Hong Kong London
Milan Paris Tokyo

Contents

1 Conditional expectations 1
1.1 Conditional expectations: the discrete case 1
1.2 Conditional expectations: the continuous case 4
1.3 It's a handy tool! 6
1.4 Solutions to the quick exercises 8
1.5 Exercises 10

2 Generating Functions 15
2.1 Generating Functions 15
2.2 Probability Generating Functions 16
2.3 Moment Generating Functions 21
2.4 Convergence in distribution 24
2.5 The central limit theorem 25
2.6 Solutions to the quick exercises 27
2.7 Exercises 29

3 Discrete-time Markov chains 33
3.1 An example 34
3.2 Definition of a discrete Markov chain 36
3.3 Classification of the states 37
The simple random walk 42
3.4 Solutions to the quick exercises 44
3.5 Exercises 45

4 Limit behavior of discrete Markov chains 51
4.1 Branching processes 51
4.2 Asymptotic behavior 55
Some examples 55
Periodicity 59
The main theorem 60
4.3 Solutions to the quick exercises 63

4.4 Exercises ... 64

5 Continuous-time Markov chains ... 69
5.1 The Markov property ... 69
5.2 The generator matrix ... 71
5.3 Kolmogorov's backward and forward equations ... 74
5.4 The generator matrix revisited ... 75
5.5 Asymptotic behavior ... 79
5.6 Birth-death processes ... 83
5.7 Solutions to the quick exercises ... 86
5.8 Exercises ... 88

A Answers to exercises ... 91

B Solutions to selected exercises ... 93

References ... 99

Index ... 101

1

Conditional expectations

Knowing whether a certain event has occurred (or not) often makes it easier to determine the (desired) probability of another event. The concept of conditional probability is very useful and powerful, see e.g. the examples given in [5], Chapter 3. The law of total probability and Bayes' rule are handy tools.

Conditioning is also a powerful concept when applied to random variables, and occurs in a natural manner. Suppose for example that the average life span of women in Holland is 80 years, and that your mother is 50. It would be a mistake to think that she can expect to live for another 30 years. What you would like to know in this situation is the expected life span of women in Holland, *given* that they are at least 50 years of age.

In order to define the conditional expectation of a random variable X, given that some other random variable Y has attained a certain value y, we need to know the distribution of X, *given* that $Y = y$. We treat the discrete and continuous cases separately.

1.1 Conditional expectations: the discrete case

For two events E and F, with $\mathrm{P}(F) > 0$, one defines (see e.g. Section 3.4 in [5]) the conditional probability of E, *given* F, as

$$\mathrm{P}(E \mid F) = \frac{\mathrm{P}(E \cap F)}{\mathrm{P}(F)}. \tag{1.1.1}$$

Now let X and Y be two discrete random variables, and let y be a number, such that $\mathrm{P}(Y = y) > 0$. In view of (1.1.1), setting $E = \{X = x\}$ and $F = \{Y = y\}$, we define the *conditional probability* that $X = x$, *given* $Y = y$, as

$$\mathrm{P}(X = x \mid Y = y) = \frac{\mathrm{P}(X = x, Y = y)}{\mathrm{P}(Y = y)}.$$

Quick exercise 1.1 In Chapters 4 and 9 of [5] we considered two random variables S and M, the sum and the maximum of two independent throws of an unbiased die. In Table 1.1 (see also Table 9.1 in [5]), the joint probability mass function $p(a,b) = \mathrm{P}(S = a, M = b)$ is given. Use this table to determine

Table 1.1. Joint probability mass function $p(a,b) = \mathrm{P}(S = a, M = b)$.

	b					
a	1	2	3	4	5	6
2	1/36	0	0	0	0	0
3	0	2/36	0	0	0	0
4	0	1/36	2/36	0	0	0
5	0	0	2/36	2/36	0	0
6	0	0	1/36	2/36	2/36	0
7	0	0	0	2/36	2/36	2/36
8	0	0	0	1/36	2/36	2/36
9	0	0	0	0	2/36	2/36
10	0	0	0	0	1/36	2/36
11	0	0	0	0	0	2/36
12	0	0	0	0	0	1/36

the distribution of S, given that $M = 4$. What is the sum of the various probabilities? (See also Exercise 1.2.) □

If $p(x,y)$ denotes the joint probability mass function of X and Y, and p_Y the (marginal) probability mass function of Y, we see that *given* $Y = y$, the conditional probability mass function $p_{X|Y}$ of X is

$$p_{X|Y}(x|y) = \mathrm{P}(X = x \,|\, Y = y) = \frac{\mathrm{P}(X = x, Y = y)}{\mathrm{P}(Y = y)} = \frac{p(x,y)}{p_Y(y)},$$

where y is a number for which $\mathrm{P}(Y = y) > 0$.

The *conditional expectation* $\mathrm{E}(X \,|\, Y = y)$ of X, *given* $Y = y$ is—following the classical definition for discrete random variables—defined by

$$\mathrm{E}(X \,|\, Y = y) = \sum_x x p_{X|Y}(x|y).$$

Quick exercise 1.2 For the discrete random variables S and M from Quick exercise 1.1, determine the conditional expectations $\mathrm{E}(S \,|\, M = b)$ for $b = 1, 4$, and $b = 6$. □

In this quick exercise we have seen that $\mathrm{E}(S \mid M = b)$ is actually a function of the value b that M attains! To be complete, all the values of $\mathrm{E}(S \mid M = b)$ are given in Table 1.2.

Table 1.2. Probability mass function of $\mathrm{E}(S \mid M)$.

b	1	2	3	4	5	6
$\mathrm{E}(S \mid M = b)$	2	$\frac{10}{3}$	$\frac{24}{5}$	$\frac{44}{7}$	$\frac{70}{9}$	$\frac{102}{11}$

If we denote by $\mathrm{E}(X \mid Y)$ that function $k(Y)$ of Y which attains at $Y = y$ the value $\mathrm{E}(X \mid Y = y)$, then $k(Y)$ is a random variable, and thus has an expected value! Using the change of variable formula (see [5], Section 7.3), we find that

$$\mathrm{E}[\mathrm{E}(X \mid Y)] = \sum_y \mathrm{E}(X \mid Y = y)\mathrm{P}(Y = y)\,.$$

By definition, $\mathrm{E}(X \mid Y = y) = \sum_x x\mathrm{P}(X = x \mid Y = y)$, so it follows that

$$\begin{aligned}
\mathrm{E}[\mathrm{E}(X \mid Y)] &= \sum_y \Big(\sum_x x\mathrm{P}(X = x \mid Y = y)\Big)\mathrm{P}(Y = y) \\
&= \sum_y \sum_x x\mathrm{P}(X = x, Y = y) \\
&= \sum_x x\Big(\sum_y \mathrm{P}(X = x, Y = y)\Big) \\
&= \sum_x x\mathrm{P}(X = x) \\
&= \mathrm{E}[X]\,.
\end{aligned}$$

We have derived the following theorem.

Theorem 1.1 Let X and Y be two discrete random variables. Then

$$\mathrm{E}[\mathrm{E}(X \mid Y)] = \mathrm{E}[X]\,. \tag{1.1.2}$$

Quick exercise 1.3 For the discrete random variables S and M from Quick exercise 1.1, use (1.1.2) to determine the expectation value $\mathrm{E}[S]$ of S. Check your answer by calculating $\mathrm{E}[S]$ directly from the (marginal) probability mass function p_S of S. □

Apart from being a bit strange at first sight ('the expectation of the conditional expectation of X given Y, is the expectation of X'), this result is very handy and powerful.

Suppose for example, that we repeatedly throw a biased coin, with P('heads') = p, and let X be the number of throws needed to throw 'heads' for the first time. Then X has a *Geo*(p) distribution, and we have seen in Section 7.2 of [5] (see also Remark 11.1 in [5]), that $\mathrm{E}[X] = 1/p$. One can easily obtain this result with conditional expectations! Define the random variable Y as

$$Y = \begin{cases} 1, & \text{if the first throw is `heads'} \\ 0, & \text{if the first throw is `tails'}. \end{cases}$$

According to our theorem we then have that

$$\begin{aligned} \mathrm{E}[X] = \mathrm{E}[\mathrm{E}(X \,|\, Y)] &= \sum_{y=0}^{1} \mathrm{E}(X \,|\, Y = y) \mathrm{P}(Y = y) \\ &= \mathrm{E}(X \,|\, Y = 0)(1 - p) + \mathrm{E}(X \,|\, Y = 1)p. \end{aligned}$$

By definition of Y, $Y = 1$ means that we have 'heads' on the first throw, so $\mathrm{E}(X \,|\, Y = 1) = 1$; we're done right away! In case $Y = 0$, the first throw is 'tails', so we need to throw again, and we find that $\mathrm{E}(X \,|\, Y = 0) = 1 + \mathrm{E}[X]$ (see also Exercise 1.7 where these conditional expectation are calculated explicitly, and not—like here—heuristically derived). But then we have, that

$$\mathrm{E}[X] = (1 + \mathrm{E}[X])(1 - p) + 1 \cdot p = (1 - p)\mathrm{E}[X] + 1,$$

i.e., $p\mathrm{E}[X] = 1$, implying that $\mathrm{E}[X] = 1/p$. □

Using the change of variable formula, one can get a result which implies our theorem.

Quick exercise 1.4 Let X and Y be two discrete random variables, and let $k(Y)$ denote the conditional expectation of X given Y, i.e., $k(Y) = \mathrm{E}(X \,|\, Y)$. Furthermore, let g be a function, for which the expectations $\mathrm{E}[k(Y)g(Y)]$ and $\mathrm{E}[Yg(Y)]$ exist. Then

$$\mathrm{E}[\mathrm{E}(X \,|\, Y)g(Y)] = \mathrm{E}[Xg(Y)]\,.$$

1.2 Conditional expectations: the continuous case

As usual, the situation is more complicated for continuous random variables. Let X and Y be two continuous random variables, with joint probability density function f, and (marginal) densities f_X respectively f_Y. In this case $\mathrm{P}(Y = y) = 0$ for all y. We need to find a sensible definition for the conditional density function $F_{X|Y}$ of X *given* that $Y = y$. In order to find such a definition, consider

$$\mathrm{P}(X \le x \,|\, y \le Y \le y + h),$$

where $h > 0$, and y is such that $f_Y(y) > 0$ and $\mathrm{P}(y \leq Y \leq y+h) > 0$. From (1.1.1) we find that

$$\mathrm{P}(X \leq x \,|\, y \leq Y \leq y+h) = \frac{F(x, y+h) - F(x,y)}{F_Y(y+h) - F_Y(y)}, \tag{1.1.3}$$

where F is the joint distribution function of X and Y, and F_Y is the (marginal) distribution function of Y. Taking $h \to 0$, the left-hand side of (1.1.3) is the conditional distribution function $F_{X|Y}(x|y)$ of X *given* that $Y = y$. The right-hand side of (1.1.3) can be written as

$$\frac{F(x, y+h) - F(x,y)}{h} \cdot \frac{h}{F_Y(y+h) - F_Y(y)}.$$

Letting $h \to 0$ yields that

$$F_{X|Y}(x|y) = \frac{\partial(F(x,y))/\partial y}{f_Y(y)}.$$

Since the derivative of $F_{X|Y}(x|y)$ to x should yield the conditional probability density function $f_{X|Y}(x|y)$ of X *given* $Y = y$, we have the following definition.

> **Definition.** Let X and Y be two continuous random variables with joint probability density function f, and let f_Y the (marginal) density function of Y. Then the conditional probability density function $f_{X|Y}$ of X *given* Y is defined by
>
> $$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)},$$
>
> for all y such that $f_Y(y) > 0$.

Following the definition of the expectation of a continuous random variable, we define the *conditional expectation* of X *given* $Y = y$ as

$$\mathrm{E}(X \,|\, Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\,\mathrm{d}x.$$

As in the discrete case, $\mathrm{E}(X \,|\, Y = y)$ is a function of y, say $k(y) = \mathrm{E}(X \,|\, Y = y)$. Again we denote the random variable $k(Y)$ as $\mathrm{E}(X \,|\, Y)$. As in the discrete case, we have the following important property

$$\mathrm{E}[\mathrm{E}(X \,|\, Y)] = \mathrm{E}[X]\,. \tag{1.1.4}$$

Quick exercise 1.5 Use the change of variable formula to prove (1.1.4). □

1.3 It's a handy tool!

In the previous two sections we showed that

$$\mathrm{E}[\mathrm{E}(X \mid Y)] = \mathrm{E}[X]$$

holds in case both X and Y are discrete, or both are continuous. In fact this formula holds for *all* random variables X and Y, and this can be used for example to calculate all kinds of probabilities involving X and Y. To see this, we have the following lemma.

Lemma. Let E be some event, and let Y be some random variable. Then

$$\mathrm{P}(E) = \begin{cases} \sum_y \mathrm{P}(E \mid Y = y)\mathrm{P}(Y = y)\,, & \text{if } Y \text{ is discrete} \\ \int_{-\infty}^{\infty} \mathrm{P}(E \mid Y = y) f_Y(y)\,\mathrm{d}y, & \text{if } Y \text{ is continuous.} \end{cases}$$

Proof. Define the Bernoulli variable X by

$$X = \begin{cases} 1, & \text{if } E \text{ occurs} \\ 0, & \text{if } E^c \text{ occurs,} \end{cases}$$

i.e., $X = 1_E$, the indicator function of E. But then

$$\mathrm{E}[X] = \sum_{x=0}^{1} x\mathrm{P}(X = x) = 0 \cdot \mathrm{P}(X = 0) + 1 \cdot \mathrm{P}(X = 1) = \mathrm{P}(E)\,.$$

Furthermore, $\mathrm{E}(X \mid Y = y) = \mathrm{P}(E \mid Y = y)$, so in case Y is discrete we find that

$$\begin{aligned} \mathrm{P}(E) &= \sum_y \mathrm{E}(X \mid Y = y)\mathrm{P}(Y = y) \\ &= \sum_y \mathrm{P}(E \mid Y = y)\mathrm{P}(Y = y)\,. \end{aligned}$$

In case Y is continuous we find

$$\begin{aligned} \mathrm{P}(E) &= \int_{-\infty}^{\infty} \mathrm{E}(X \mid Y = y) f_Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} \mathrm{P}(E \mid Y = y) f_Y(y)\,\mathrm{d}y. \end{aligned}$$

□

Recall that for two independent continuous random variables X and Y the joint probability density function of the sum $X+Y$ of two continuous random variables X and Y is given by

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_X(z-y) f_X(y)\,\mathrm{d}y, \qquad \text{for } -\infty < z < \infty,$$

see e.g. [5], Section 11.2.

Setting $E = \{X + Y \leq z\}$, we first will use the above lemma to find the distribution function of $X + Y$. Differentiation will then yield f_{X+Y}. For z between $-\infty$ and ∞,

$$\begin{aligned} F_{X+Y}(z) = \mathrm{P}(E) &= \int_{-\infty}^{\infty} \mathrm{P}(E \mid Y = y) f_Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} \mathrm{P}(X + Y \leq z \mid Y = y) f_Y(y)\,\mathrm{d}y \\ &= \int_{-\infty}^{\infty} \mathrm{P}(X + y \leq z \mid Y = y) f_Y(y)\,\mathrm{d}y. \end{aligned}$$

Since X and Y are independent, $\mathrm{P}(X \leq z - y \mid Y = y) = \mathrm{P}(X \leq z - y)$, and we find that

$$F_{X+Y}(z) = \int_{-\infty}^{\infty} F_X(z-y) f_Y(y)\,\mathrm{d}y.$$

Differentiating $F_{X+Y}(z)$ to z yields the desired result. □

The formula $\mathrm{E}[\mathrm{E}(X \mid Y)] = \mathrm{E}[X]$ is also handy in situations where X is the sum of an unknown number of random variables, all independent and having the same distribution. For example, in a hospital N clients will enter, each needing a certain amount of time T for care. So we assume that $N, T_1, T_2, \dots$ are all independent, and that the T_i's have the same distribution. A natural question now is: 'what is the expected total time spend on all the patients', i.e., what is

$$\mathrm{E}[T_1 + T_2 + \cdots + T_N]?$$

Intuitively, one would say that the correct answer should be $\mathrm{E}[N]\,\mathrm{E}[T]$, and this is correct! To see this, note that due to the independence of N and the T_i's,

$$\mathrm{E}(\sum_{i=1}^{N} T_i \mid N = n) = \mathrm{E}(\sum_{i=1}^{n} T_i \mid N = n) = \mathrm{E}\left[\sum_{i=1}^{n} T_i\right] = n\mathrm{E}[T]\,,$$

so from $\mathrm{E}[\mathrm{E}(X \mid Y)] = \mathrm{E}[X]$ it follows that

$$\begin{aligned} \mathrm{E}\left[\sum_{i=1}^{N} T_i\right] &= \sum_n \mathrm{E}(\sum_{i=1}^{N} T_i \mid N = n)\mathrm{P}(N = n) \\ &= \sum_n \mathrm{E}(\sum_{i=1}^{n} T_i \mid N = n)\mathrm{P}(N = n) \\ &= \sum_n n\mathrm{E}[T]\,\mathrm{P}(N = n) = \mathrm{E}[T] \cdot \sum_n n\mathrm{P}(N = n) \\ &= \mathrm{E}[N]\,\mathrm{E}[T]\,. \end{aligned}$$

The following rule holds generally.

Expectation and variance of a random sum of random variables. Let $N, T_1, T_2, \ldots$ be independent random variables, where the T_i's all have the same distribution, and N is a discrete random variable attaining only non-negative integer values. Then

$$\mathrm{E}\left[\sum_{i=1}^{N} T_i\right] = \mathrm{E}[N]\,\mathrm{E}[T] \tag{1.1.5}$$

and

$$\mathrm{Var}\left(\sum_{i=1}^{N} T_i\right) = \mathrm{E}[N]\,\mathrm{Var}(T) + (\mathrm{E}[X])^2\mathrm{Var}(N)\,.$$

(See also [12], Section 3.4).

1.4 Solutions to the quick exercises

1.1 We need to determine the probabilities of S, *given* that $M = 4$. From Table 9.1 (in [5]) one sees that $\mathrm{P}(S = a \mid M = 4) = 0$ for $a = 2, 3, 4, 9, 10, 11$, and 12. Furthermore, $\mathrm{P}(S = 5 \mid M = 4) = \frac{2/36}{7/36} = \frac{2}{7}$, $\mathrm{P}(S = 6 \mid M = 4) = \frac{2/36}{7/36} = \frac{2}{7}$, $\mathrm{P}(S = 7 \mid M = 4) = \frac{2/36}{7/36} = \frac{2}{7}$, and $\mathrm{P}(S = 8 \mid M = 4) = \frac{1/36}{7/36} = \frac{1}{7}$. We see that, leaving out the terms in the sum which are zero,

$$\sum_{a=2}^{12} \mathrm{P}(S = a, M = b) = \frac{2}{7} + \frac{2}{7} + \frac{2}{7} + \frac{1}{7} = 1.$$

1.2 By definition,

$$\mathrm{E}(S \mid M = 1) = \sum_{a=2}^{12} s\mathrm{P}(S = a \mid M = 1) = 2 \cdot \frac{1/36}{1/36} = 2,$$

$$\begin{aligned}\mathrm{E}(S \mid M = 4) &= \sum_{a=2}^{12} s\mathrm{P}(S = a \mid M = 4) \\ &= 5 \cdot \frac{2/36}{11/36} + 6 \cdot \frac{2/36}{11/36} + 7 \cdot \frac{2/36}{11/36} + 8 \cdot \frac{1/36}{11/36} = \frac{44}{7},\end{aligned}$$

and

$$\begin{aligned}\mathrm{E}(S \mid M = 6) &= \sum_{a=2}^{12} s\mathrm{P}(S = a \mid M = 6) \\ &= 7 \cdot \frac{2/36}{11/36} + 8 \cdot \frac{2/36}{11/36} + 9 \cdot \frac{2/36}{11/36} + 10 \cdot \frac{2/36}{11/36} \\ &\quad +11 \cdot \frac{2/36}{11/36} + 12 \cdot \frac{1/36}{11/36} = \frac{102}{11}.\end{aligned}$$

1.3 From (1.1.2) it follows that, using Table 1.2,

$$\begin{aligned}
\mathrm{E}[S] &= \sum_{b=1}^{6} \mathrm{E}(S \mid M=b) \\
&= 2\cdot\frac{1}{36} + \frac{10}{3}\cdot\frac{3}{36} + \frac{24}{5}\cdot\frac{5}{36} + \frac{44}{7}\cdot\frac{7}{36} + \frac{70}{9}\cdot\frac{9}{36} + \frac{102}{11}\cdot\frac{9}{36} \\
&= \frac{2+10+24+44+70+102}{36} = 7.
\end{aligned}$$

In the margin of Table 9.2 in [5] the marginal (!) probability mass function p_S is given. We find that

$$\begin{aligned}
\mathrm{E}[S] = \sum_{a=2}^{12} a\mathrm{P}(S=a) &= 2\cdot\frac{1}{36} + 3\cdot\frac{2}{36} + 4\cdot\frac{3}{36} + 5\cdot\frac{4}{36} + 6\cdot\frac{5}{36} + 7\cdot\frac{6}{36} \\
&\quad + 8\cdot\frac{5}{36} + 9\cdot\frac{4}{36} + 10\cdot\frac{3}{36} + 11\cdot\frac{2}{36} + 12\cdot\frac{1}{36} = 7.
\end{aligned}$$

1.4 Clearly this result implies our theorem; choose $g(y)=1$ for all y. Now due to the change of variable formula we have

$$\begin{aligned}
\mathrm{E}[\mathrm{E}(X \mid Y)g(Y)] &= \sum_y \mathrm{E}(X \mid Y=y)g(y)\mathrm{P}(Y=y) \\
&= \sum_y \Big(\sum_x xg(y)\mathrm{P}(X=x \mid Y=y)\Big)\mathrm{P}(Y=y) \\
&= \sum_{x,y} xg(y)\mathrm{P}(X=x, Y=y) \\
&= \mathrm{E}[Xg(Y)]\,.
\end{aligned}$$

1.5 According to the change of variable formula, $\mathrm{E}[k(Y)] = \int_{-\infty}^{\infty} k(y) f_Y(y)\,\mathrm{d}y$. Setting $k(y) = \mathrm{E}(X \mid Y=y)$, we find

$$\begin{aligned}
\mathrm{E}[\mathrm{E}(X \mid Y)] &= \int_{-\infty}^{\infty} \mathrm{E}(X \mid Y=y) f_Y(y)\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \Big(\int_{-\infty}^{\infty} x f_{X|Y}(x|y)\ \mathrm{d}x\Big) f_Y(y)\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} x\Big(\int_{-\infty}^{\infty} f(x,y)\ \mathrm{d}y\Big)\mathrm{d}x \\
&= \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x \\
&= \mathrm{E}[X]\,.
\end{aligned}$$

We see that the Theorem from Section 1.1.1 also holds for continuous random variables.

1.5 Exercises

1.1 Suppose your neighbors have two children.[1]

a. What is the conditional probability that these two children are girls, *given* that at least one of them is a girl?

b. During a walk in the park you met the father of these children, who desperately tried to keep up with one of his children—a girl—driving a small bicycle. What is the conditional probability that both the neighbors' children are girls, *given* that you saw that at least one of them is a girl?

Did you get the same answer as in **a.**? Why (not)?

1.2 Let X and Y be two discrete random variables. Explain why

$$\sum_x p_{X|Y}(x|y) = 1,$$

for all y for which $\mathrm{P}(Y = y) > 0$.

1.3 In [5], Exercise 9.1, the joint probabilities $\mathrm{P}(X = a, Y = b)$ of two discrete random variables X and Y are given, based on the magical square in Albrecht's Dürer's engraving *Melencolia I* (see also Figure 1.1 on page 11). We have the following table:

	a			
b	1	2	3	4
1	16/136	3/136	2/136	13/136
2	5/136	10/136	11/136	8/136
3	9/136	6/136	7/136	12/136
4	4/136	15/136	14/136	1/136

a. Determine $\mathrm{E}(X \mid Y = y)$ for $y = 1, 2, 3, 4$.

b. Use your answer from **a.** to calculate $\mathrm{E}[X]$, and check this with the answer you would get if you would use the (marginal) probability mass function p_X of X.

1.4 Let X, Y, and Z be three discrete random variables, with joint probability density function $p(a, b, c) = \mathrm{P}(X = a, Y = b, Z = c)$, given by

[1] Although technically not really an exercise which 'belongs' in this chapter, it is a good thing to think about the meaning of conditional statements. In fact this is an old exercise which has stirred a lot of discussion and controversy, see e.g. Chapter 10 in [2], or more recently [9], and several papers in subsequent issues of the same journal commenting on [9].

Fig. 1.1. Albrecht Dürer's *Melencolia I.*

$$p(0,0,0) = \tfrac{1}{8},\ \ p(0,0,1) = \tfrac{3}{16},\ p(0,1,0) = 0,\ p(0,1,1) = \tfrac{3}{16}$$

$$p(1,0,0) = \tfrac{1}{16},\ p(1,0,1) = 0,\ \ p(1,1,0) = \tfrac{3}{8},\ p(1,1,1) = \tfrac{1}{16}$$

a. Determine $\mathrm{E}(X \mid Y = 0, Z = 0)$.
b. Determine $\mathrm{E}(X \mid Y = 0)$.

1.5 In a nightmare you are locked-up in a dungeon with three doors. After an hour you finally try the doors; all three are unlocked! Behind the first door is (unknown to you) a tunnel which returns you to your cell after 2 hellish hours, behind the second door there is a trapdoor which will return you to the dungeon right-away, while the horrors behind the third door will wake you up immediately.

a. Suppose that you always choose—independent of possibly earlier choices—any of the three doors completely randomly. What is the expected length of your nightmare?

b. Even in a nightmare it doesn't make sense to choose the same door twice, so let us assume that you always will choose a door completely randomly from the doors you haven't picked before. What is now the expected length of your nightmare?

1.6 Every Friday-night, the police of Nergenshuizen stops cars on the local main street, to test the drivers for driving under the influence of illegal substances. Let p be the probability that a driver is intoxicated, and let the number of drivers who are stopped be a $Pois(\lambda)$ distributed random variable.

a. What is the probability that not a single driver stopped by the police will be intoxicated?
b. What is the probability that of the drivers stopped by the police, k will be intoxicated?

1.7 A biased coin is thrown until 'heads' shows for the first time, where $p = \mathrm{P}(\text{'}heads')$ is a number between 0 and 1. Let X be the number of throws needed, and let Y be a random variable with a $Ber(p)$ distribution, where $Y = 1$ if the first throw resulted in 'heads', and $Y = 0$ otherwise.

a. Give a formula for $\mathrm{P}(X = k)$, for $k = 1, 2, \ldots$.
b. Calculate $\mathrm{P}(X = k \mid Y = 0)$, for $k = 1, 2, \ldots$.
c. Show that $\mathrm{E}(X \mid Y = 0) = 1 + \mathrm{E}[X]$.
d. Calculate $\mathrm{E}(X \mid Y = 1)$.
e. Give a function h which satisfies $h(Y) = \mathrm{E}(X \mid Y)$.
f. Use the formula $\mathrm{E}[X] = \mathrm{E}[\mathrm{E}(X \mid Y)]$ to find that $\mathrm{E}[X] = 1/p$.
g. Use the above method to determine $\mathrm{E}\left[X^2\right]$.

1.8 (See also Quick exercise 1.4) Let X and Y be two continuous random variables, and let $k(Y)$ denote the conditional expectation of X given Y. Furthermore, let g be a function, for which the expectations $\mathrm{E}[k(Y)g(Y)]$ and $\mathrm{E}[Y g(Y)]$ exist. Then

$$\mathrm{E}[\mathrm{E}(X \mid Y)g(Y)] = \mathrm{E}[Xg(Y)] \,.$$

1.9 On many provincial highways in Holland the situation is quite dangerous, mainly due to reckless driving and speeding. Suppose that the expected number of accidents per month on a particularly notorious provincial highway is 4, and that per accident the expected number of deceased people is 2. The number of accidents per month, and the number of deceased people per accident are assumed to be independent random variables. What is the expected number of deceased people per month?

1.10 Consider the following (highly unrealistic) model for age-span distribution of women. Suppose that the age-span X of a woman is uniformly distributed over the years $0, 1, 2, \ldots, 160$. So the probability that your mother will become i years of age is in this model given by

$$P(X = i) = \frac{1}{161}, \qquad \text{for } i = 0, 1, 2, \ldots, 160.$$

In this model $E[X] = 80$ (years). Now suppose that your mother has turned 50, can she expect to live another 30 years? More? Less? In order to answer this question, set

$$Y = \begin{cases} 0 & \text{if } X < 50 \\ 1 & \text{if } X \geq 50. \end{cases}$$

a. Determine the joint probability mass function of X and Y.
b. Calculate the desired expectation $E(X \mid X \geq 50)$.

1.11 Let X and Y be two continuous random variables, with joint probability density function given by

$$f(x, y) = x + y, \qquad \text{for } 0 \leq x \leq 1,\ 0 \leq y \leq 1,$$

and $f(x, y) = 0$ for all other values of x and y.

a. Determine for y between 0 and 1 the conditional probability density $f_{X|Y}(x|y)$ of X, given that $Y = y$.
b. Determine the conditional expectation of X, given that $Y = y$ (of course, with $0 \leq y \leq 1$).
c. Use **b.** to calculate $E[X]$. Check your answer by calculating $E[X]$ directly from the (marginal) probability density of X.

1.12 Let X and Y be two continuous random variables, with joint probability density function f, and (marginal) probability density functions f_X respectively f_Y. Furthermore, suppose that

$$f_Y(y) = 5y^4 1_{(0,1)}(y)$$

and

$$f_{X|Y}(x|y) = \frac{K(y)x}{y^2} 1_{(0,y)}(x).$$

a. Determine $K(y)$; note that $K(y)$ does not depend on y!
b. Calculate $P(\frac{1}{4} < X < \frac{1}{2} \mid Y = \frac{5}{8})$.
c. Determine $f(x, y)$, the joint probability density function of X and Y.
d. Calculate $P\big(\frac{1}{4} < X < \frac{1}{2}\big)$.

1.13 The casino 'La Bella Fortuna' has come up with a new game. Throw with a fair die. If the outcome is less than 6, the casino will cash out the outcome in euros. If the outcome is 6, one is allowed to throw the die another time. If the outcome of this second throw is less than 6, the casino will cash out this second outcome + the outcome of the first throw (which is 6) in euros. In case the second throw was also a 6, one is allowed to throw for a third time. Etc. The direction has not yet figured out what a fair price should be, for someone to participate in this game. You—as a probabilist—are asked to determine the fair price.

2

Generating Functions

In this chapter the concept of generating function will be introduced, and then applied to probabilities and expectations. The resulting probability and moment generating functions are powerful tools, and in several examples applications will be given. One such application is to see—in an easy way—what the distribution is of the sum of two random variables.

2.1 Generating Functions

An easy—and at first view a seemingly not very helpful—way to 'code' an infinite sequence of numbers $a_0, a_1, a_2, a_3, \ldots$ by a single object is the formal power series

$$A(s) = \sum_{k=0}^{\infty} a_k s^k.$$

In what follows we will treat $A(s)$ as a formal power series, i.e., we will not bother ourselves to find for which values s the power series $A(s)$ converges. There exists a number $R \geq 0$, the *radius of convergence*, such that the sum $A(s)$ converges absolutely for $s < R$, and diverges if $S > R$. The sum $A(s)$ is absolutely convergent on $\{s : |s| < R'\}$, for any $R' < R$. As a consequence, $A(s)$ may be differentiated or integrated as often as you like when $|s| < R$.

Several interesting sequences have a 'nice' generating function. For example, if $a_k = \binom{n}{k}$ for $k = 0, 1, \ldots, n$, and $a_k = 0$ for $k \geq n+1$, one has that

$$A(s) = \sum_{k=0}^{n} \binom{n}{k} s^k = (1+s)^n.$$

An even easier example is the sequence, consisting of only 1's: $a_k = 1$ for $k = 0, 1, 2, \ldots$. In this case the generating function $A(s)$ of the sequence $(a_k)_{k \geq 0}$ is given by

$$A(s) = \sum_{k=0}^{\infty} s^k = \frac{1}{1-s}.$$

Quick exercise 2.1 What are the radii of convergence in these two examples?

Now let $A(s)$ be the generating function of the sequence $(a_k)_{k\geq 0}$, and let $B(s)$ be the generating function of the sequence $(b_k)_{k\geq 0}$. Concider

$$C(s) = A(s)B(s) = (a_0 + a_1 s + a_2 s^2 + \cdots)(b_0 + b_1 s + b_2 s^2 + \cdots).$$

The coefficient of s^n in this product is

$$c_n = \sum_{k=0}^{n} a_k b_{n-k}, \quad n \geq 0,$$

when $C(s) = \sum_{k=0}^{\infty} c_k$. Although simple, this fact makes generating functions well suited for applications in probability theory. The sequence $(c_k)_{k\geq 0}$ is called the *convolution* of $(a_k)_{k\geq 0}$ and $(b_k)_{k\geq 0}$. Recall that we have seen the convolution of the probability mass functions or probability density functions of two random variables X and Y, when dealing with the sum $X+Y$. We will see that the convolution is directly related to the distribution of the sum $X+Y$ of the random variables X and Y.

Quick exercise 2.2 Show, that for every non-negative integer n,

$$\binom{n}{0}^2 + \binom{n}{1}^2 + \cdots + \binom{n}{n}^2 = \binom{2n}{n}.$$

Note that there is a one-one relation between an infinite sequence $(a_n)_{n\geq 0}$ and its generating sequence $A(s)$. If $(b_n)_{n\geq 0}$ has as generating function $B(s)$, and $A(s) = B(s)$ for $|s| < R'$, where $R' > 0$ is smaller than the radii of convergence of both $A(s)$ and $B(s)$, then we have that $a_n = b_n$ for $n \geq 0$. Conversely, if $a_n = b_n$ for $n \geq 0$, then obviously we have that $A(s) = B(s)$. We see that the generating function $A(s)$ is uniquely determined by the sequence $(a_n)_{n\geq 0}$, and conversely.

2.2 Probability Generating Functions

Generating functions are very handy tools when applied to random variables. In this section we will consider discrete random variables having values in the non-negative integers $\mathbb{N}$. In the next section we will widen our view to general random variables.

Probability Generating Functions. Let X be a discrete random variables, with values in the non-negative integers $\mathbb{N}$. Then the *probability generating function* of X is defined as

$$G_X(s) = \sum_{k=0}^{\infty} \mathrm{P}(X = k)\, s^k.$$

Quick exercise 2.3 Show that that the radius of convergence of $G_X(s)$ is at least 1. □

Examples 2.1

(i) Let X have a *Ber*(p) distribution, i.e., $\mathrm{P}(X = 1) = p$, and $\mathrm{P}(X = 0) = 1-p$. Then

$$G_X(s) = \mathrm{P}(X = 0) + \mathrm{P}(X = 1)\, s + 0 \cdot s^2 + 0 \cdot s^3 + \cdots = 1 - p + ps$$

is the probability generating function of X.

(ii) If X has a *Bin*(n, p) distribution, then it follows from Newton's Binonium that the probability generating function of X is given by:

$$\begin{aligned} G_X(s) &= \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} s^k \\ &= \sum_{k=0}^{n} \binom{n}{k} (ps)^k (1-p)^{n-k} \\ &= (1 - p + ps)^n. \end{aligned}$$

(iii) Let X be a discrete random variable with a *Pois*(μ) distribution. Then its probability generating function is given by

$$\begin{aligned} G_X(s) &= \sum_{k=0}^{\infty} \frac{\mu^k}{k!} \mathrm{e}^{-\mu} s^k = \mathrm{e}^{-\mu} \sum_{k=0}^{\infty} \frac{(\mu s)^k}{k!} \\ &= \mathrm{e}^{-\mu} \mathrm{e}^{\mu s} = \mathrm{e}^{\mu(s-1)}. \end{aligned}$$

Quick exercise 2.4 Let X be a discrete random variable, with values in $\mathbb{N}$. Then

$$G_X(s) = \mathrm{E}\big[s^X\big]\,.$$

We could have used the expression in Quick exercise 2.4 as a definition of the probability generating function of any random variable X. We decided to stick to our definition to be able to derive some nice properties holding only in the particular case that X has its values in $\mathbb{N}$. For general random variables we will slightly alter the definition, and consider the so-called *moment generating function*, see also the next section, Section 2.3.

Now we will derive some properties of probability generating functions. All these properties follow from the fact that a probability generating function has a radius of convergence which is at least 1.

Quick exercise 2.5 Let X be a discrete random variable with values in $\mathbb{N}$, and let $G_X(s)$ be its probability generating function. Show that

$$\mathrm{P}(X=0) = G_X(0), \qquad \mathrm{P}(X=1) = G'_X(0),$$

and more generally,

$$\mathrm{P}(X=k) = \frac{1}{k!} G_X^{(k)}(0), \qquad \text{for } k = 0, 1, 2, \ldots.$$

Since the radius of convergence of a probability generating function is at least 1, the following theorem is a handy tool, see also [11], p. 120, or Abel's original paper [1] (the theorem had already been stated and used by Gauss, but Gauss' proof contained a gap involving the interchange of two limit processes).

Theorem 2.1 (Abel) If $A(s)$ converges for a certain positive value s_0 of s, then it also converges also for every real number s for which $|s| < s_0$, and $A(s-\alpha)$ converges to $A(s)$ if $|s| \le s_0$ and α decreases to zero, i.e.,

$$\lim_{s \uparrow s_0} A(s) = \sum_{k=0}^{\infty} a_k s_0^k.$$

As direct consequences of Abel's theorem one has the following properties.

Theorem 2.2 Let X be a discrete random variable with values in $\mathbb{N}$, and suppose that $\mathrm{E}\left[X^2\right] < \infty$. If $G_X(s)$ is the probability generating function of X, then

(a) $\mathrm{E}[X] = G'_X(1)$,
(b) $\mathrm{Var}(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2$.

Proof. As we have seen in Quick exercise 2.3, the series $\sum_{k=0}^{\infty} \mathrm{P}(X=k)\, s^k$ has radius of convergence at least 1. For s with $|s| < 1$ the probability generating function is differentiable, and

$$G'_X(s) = \sum_{k=0}^{\infty} k \mathrm{P}(X=k)\, s^{k-1}, \qquad G''_X(s) = \sum_{k=0}^{\infty} k(k-1) \mathrm{P}(X=k)\, s^{k-2}.$$

Due to Abel's theorem (with $s_0 = 1$), we find that

$$G'_X(1) = \sum_{k=0}^{\infty} k \mathrm{P}(X=k) = \mathrm{E}[X],$$

and

$$\begin{aligned} G''_X(1) &= \sum_{k=0}^{\infty} k(k-1)\mathrm{P}(X=k) = \sum_{k=0}^{\infty} k^2 \mathrm{P}(X=k) - \sum_{k=0}^{\infty} k \mathrm{P}(X=k) \\ &= \mathrm{E}\left[X^2\right] - \mathrm{E}[X]. \end{aligned}$$

But then we find that $\mathrm{Var}(X) = G''_X(1) + G'_X(1) - (G'_X(1))^2$. $\square$

Quick exercise 2.6 Use Theorem 2.2 to find the expectation and variance of a random variable X with a *Bin*(n, p) distribution. □

In [5], Chapter 9, we considered the sum of two independent random variables X and Y. In case these random variables are discrete, it was derived that the probability mass function of $X + Y$ is given as the convolution of the probability mass functions of X and Y:

$$p_{X+Y}(c) = \sum_b p_X(c-b)p_Y(b),$$

i.e.,

$$\mathrm{P}(X+Y=c) = \sum_b \mathrm{P}(X=c-b)\,\mathrm{P}(Y=b)\,.$$

It is no coincidence that the probability mass function of $X + Y$ is called the convolution of the probability mass functions of X and Y, as the following theorem shows.

Theorem 2.3 Let X and Y be two independent discrete random variables, both having values in $\mathbb{N}$. Furthermore, let $Z = X + Y$. Then

$$G_Z(s) = G_X(s)G_Y(s).$$

Proof. Since X and Y are independent, we have—due to the propagation of independence property from Section 9.5 in [5]—that s^X and S^Y are independent for all s with $|s| < 1$. But then we have, that

$$G_Z(s) = \mathrm{E}\big[s^{X+Y}\big] = \mathrm{E}\big[s^X s^Y\big] = \mathrm{E}\big[s^X\big]\,\mathrm{E}\big[s^Y\big] = G_X(s)G_Y(s).$$

□

This theorem is trivially generalized to the sum of n independent random variables. We have the following corollary.

Corollary 2.1 Let $X_1, X_2, \ldots, X_n$ be n independent discrete random variables, all having values in $\mathbb{N}$. Furthermore, let $G_{X_i}(s)$ be the probability generating function if X_i, and let $G_{X_1+X_2+\cdots+X_n}(s)$ the probability generating function of $X_1 + X_2 + \cdots + X_n$. Then

$$G_{X_1+X_2+\cdots+X_n}(s) = G_{X_1}(s)G_{X_2}(s)\cdots G_{X_n}(s).$$

Examples 2.2

(a) Let X and Y be two independent random variables, where X has a *Pois*(λ) distribution, and Y a *Pois*(μ) distribution. Then

$$G_{X+Y}(s) = G_X(s)G_Y(s) = \mathrm{e}^{\lambda(s-1)}\mathrm{e}^{\mu(s-1)} = \mathrm{e}^{(\lambda+\mu)(s-1)}.$$

From Example 2.1(c) we know that $\mathrm{e}^{(\lambda+\mu)(s-1)}$ is the probability generating function of a *Pois*$(\lambda + \mu)$ distributed random variable. Due to the uniqueness property, mentioned at the end of Section 2.1, it follows that $X+Y$ has a *Pois*$(\lambda+\mu)$ distribution.

(b) Let X be a *Bin*(n,p) distributed random variable. In Examples 2.1(b) we obtained the probability generating function of X, essentially by applying Newton's Binonium Formula. However, due to our Corollary we can find this probability generating function almost without any calculation whatsoever. Recall that X can be represented as the sum $X_1+X_2+\cdots+X_n$ of n independent, *Ber*(p) distributed random variables. In Example 2.1(a) we have seen that $G_{X_i}(s)=1-p+ps$, for each i. So due to the Corollary we find that $G_X(s)=(1-p+ps)^n$. □

Note that n in Corollary 2.1 is a fixed number, not the realization of some non-constant random variable N. In this situation—i.e., when we have a random sum of random numbers—the situation is more complicated, but also more interesting! We have the following theorem.

Theorem 2.4 Let $X_1, X_2, X_3, \ldots$ be a sequence of independent random variables, all having the same distribution, and let N be a $\mathbb{N}$-valued random variable, which is independent of the X_i's. Let G_X be the probability generating function of X_i (for $i \geq 1$), and let G_N be the probability generating function of N. Then the probability generating function of $S=X_1+X_2+\cdots+X_N$ is given by

$$G_S(s)=G_N\left(G_{X_1}(s)\right).$$

In Chapter 1 we have seen, that for random variables satisfying the conditions of Theorem 2.4,

$$\mathrm{E}[S]=\mathrm{E}[N]\,\mathrm{E}[X_1],$$

and this result was obtained using conditional expectations. This is also the instrument of choice to prove Theorem 2.4 too.

Proof. Since

$$\mathrm{E}\big[s^S\big]=\mathrm{E}\big[\mathrm{E}(s^S\mid N)\big]=\sum_{n=0}^{\infty}\mathrm{E}(s^S\mid N=n)\mathrm{P}(N=n),$$

and

$$\begin{aligned}\mathrm{E}(s^S\mid N=n)&=\mathrm{E}(s^{X_1+X_2+\cdots+X_N}\mid N=n)\\&=\mathrm{E}(s^{X_1+X_2+\cdots+X_n}\mid N=n)\\&=\mathrm{E}\big[s^{X_1+X_2+\cdots+X_n}\big],\end{aligned}$$

(where the last equality is due to the independence of N and the X_i's), it follows that

$$G_S(s)=\sum_{n=0}^{\infty}\mathrm{E}\big[s^{X_1+X_2+\cdots+X_n}\big]\,\mathrm{P}(N=n).$$

Due to the propagation of independence rule (see [5], Section 9.5), we have that $s^{X_1}, s^{X_2}, \ldots$ are independent random variables, and therefore

$$\mathrm{E}\left[s^{X_1+X_2+\cdots+X_n}\right] = \mathrm{E}\left[s^{X_1}\right]\mathrm{E}\left[s^{X_2}\right]\cdots\mathrm{E}\left[s^{X_n}\right] = (G_{X_1}(s))^n.$$

This yields the desired result,

$$G_S(s) = \sum_{n=0}^{\infty}(G_{X_1}(s))^n \mathrm{P}(N=n) = G_N\left(G_{X_1}(s)\right).$$

□

2.3 Moment Generating Functions

In the previous section we have seen that probability generating functions are a handy tool when dealing with random variables having values in $\mathbb{N}$. In this section we will introduce a similar tool, which can be applied to *any* random variable X, the moment generating function of X. Here is its definition.

Moment Generating Functions. Let X be a random variables. Then the *moment generating function* of X is given by

$$M_X(t) = \mathrm{E}\left[\mathrm{e}^{tX}\right].$$

One of the problems with the moment generating function is, that we do not know a priori for which value of t the function $M_X(t)$ exists. In the next Quick exercise we even will see a random variable for which $M_X(t)$ exists only for $t=0$. Note that we always have that $M_X(0)=1$.

Quick exercise 2.7 Let X be a discrete random variable, with probability distribution given by

$$\mathrm{P}(X=n) = \mathrm{P}(X=-n) = \frac{1}{2n(n-1)}, \qquad \text{for } n = 2, 3, \ldots.$$

Then $M_X(t)$ exists only for $t=0$.

Examples 2.3

(a) If X is a random variable, having its values in $\mathbb{N}$, then there is a simple relation between the probability and moment generating function of X; we only need to replace s in $G_X(s)$ by e^t:

$$M_X(t) = G_X(\mathrm{e}^t).$$

For example, if X has a $Bin(n,p)$ distribution, then

$$M_X(t) = (1-p+p\mathrm{e}^t)^n.$$

(b) Let X have a *Exp*(λ) distribution, i.e., the probability density function of X is zero for $x < 0$, and $f_X(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. Then for all t with $t < \lambda$, the moment generating function of X is given by

$$\begin{aligned} M_X(t) &= \int_0^\infty e^{tx} \lambda e^{-\lambda x}\, dx = \lambda \int_0^\infty e^{(t-\lambda)}\, dx \\ &= \frac{\lambda}{t-\lambda} \lim_{A\to\infty} \left[e^{(t-\lambda)x} \right]_{x=0}^{x=A} = \frac{\lambda}{\lambda - t}. \end{aligned}$$

(c) Let X have a $U(\alpha, \beta)$ distribution. Then for all $t \neq 0$

$$M_X(t) = \int_\alpha^\beta \frac{e^{tx}}{\beta - \alpha} = \frac{e^{\beta t} - e^{\alpha t}}{t(\beta - \alpha)}.$$

Quick exercise 2.8 Let X have a $N(\mu, \sigma^2)$ distribution. Then for all t

$$M_X(t) = e^{\frac{\sigma^2 t^2}{2} + \mu t}.$$

The probability generating function of a discrete random variable X having values in $\mathbb{N}$, is a generating function, which has as coefficients exactly all possible probabilities of X. Something similar applies to the moment generating function of X. Recalling that $e^{tk} = \sum_{n=0}^\infty \frac{(tk)^n}{n!}$, we find that

$$\begin{aligned} M_X(t) &= \sum_{k=0}^\infty e^{tk} P(X = k) = \sum_{k=0}^\infty \sum_{n=0}^\infty \frac{(tk)^n}{n!} P(X = k) \\ &= \sum_{n=0}^\infty \frac{t^n}{n!} \left(\sum_{k=0}^\infty k^n P(X = k) \right) = \sum_{n=0}^\infty \frac{t^n}{n!} E[X^n]. \end{aligned}$$

So $M_X(t)$ is the generating function of the sequence $E[X^0], E[X^1], E[X^2]/2!, \ldots, E[X^n]/n!, \ldots$ of X. This motivates its name.

For any random variable X, for which $M_X(t)$ exists for some $t \neq 0$, one has that

$$M_X'(t) = \frac{d}{dt} E[e^{tX}] = E\left[\frac{d}{dt} e^{tX}\right] = E[Xe^{tX}],$$

implying[1] that $E[X] = M_X'(0)$. We have the following theorem.

Theorem 2.5 Let X be a random variable for which $M_X(t)$ exists for some $t \neq 0$. Then

$$E[X^k] = M_X^{(k)}(0), \qquad \text{for } k = 0, 1, 2, \ldots.$$

[1] Note that here we exchanged E and d/dt without giving the non-trivial reasons why this can be done!

Example 2.4

Let X be random variable, with a $Exp(\lambda)$ distribution. In Example 2.3(b) we have seen that $M_X(t) = \lambda/(\lambda - t)$, for $t < \lambda$. Note that we can write $M_X(t)$ as

$$M_X(t) = \frac{\lambda}{\lambda - t} = \frac{1}{1 - t/\lambda} = \sum_{k=0}^{\infty} \frac{t^k}{\lambda^k}.$$

We find that $\mathrm{E}\big[X^k\big] = M_X^{(k)}(0) = \frac{k!}{\lambda^k}$. In particular we have that

$$\mathrm{E}[X] = \frac{1}{\lambda}, \qquad \text{and } \mathrm{E}\big[X^2\big] = \frac{2}{\lambda^2}.$$

In Theorem 2.3 we have seen that the probability generating function of the sum of two random variables X and Y is equal to the product of the probability generating functions of X and Y. The essential ingredient of the proof of Theorem 2.3 is the propagation of independence rule, from [5], Section 9.5. The same rule also lies beneath the following theorem, which we mention without (the obvious) proof.

Theorem 2.6 Let X and Y be two independent random variables, with moment generating functions $M_X(t)$ respectively $M_Y(t)$. Then

$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

Uniqueness? In the previous section we have seen that the probability generating function of a random variable X, having values in $\mathbb{N}$, uniquely determines the distribution of X. Seen from this perspective, it is possible to recognize the type of distribution of X from its probability generating function.

In [12], Section 2.6, it is written on page 63 that "*the moment generating function uniquely determines the distribution.* That is, there is a one-to-one correspondence between the moment generating function and the distribution function of a random variable." In other words, once you know all the moments of a random variable X, you know its distribution. In [6], page 166, it is remarked that this is not entirely correct. In one of their exercises (Problem (5.12.43)) Grimmett & Stirzaker show there exists a family of distributions based on the so-called *log-normal* distribution, all having the same moments, and all with a different distribution (i.e., with a different distribution function and probability density function). It is then remarked that no attractive known condition exists which is both necessary and sufficient to guarantee that a distribution is uniquely determined by its moments. The easiest known sufficient condition is, that the moment generating function $M_X(t)$ of X is finite in some open neighborhood of $t = 0$.

The above discussion might at first sight appear as somewhat academic. However, to know whether or not a distribution is determined by its moments (and thus by its moment generating function), comes in very handy. For example, in [5], Section 11.2, it was shown after tedious calculations that the sum of two (standard) normal distributed random variables again has a normal distribution. In Quick exercise 2.8 we have seen that the moment generating function of a $N(\mu, \sigma^2)$ distributed random variable X is given by

$$M_X(t) = \mathrm{e}^{\frac{\sigma^2 t^2}{2} + \mu t}, \qquad -\infty < t < \infty.$$

Clearly $M_X(t)$ is finite on some open neighborhood of $t = 0$, so by the discussion above we know that any random variable Y with $M_X(t)$ as moment generating function has a $N(\mu, \sigma^2)$ distribution. Now let X and Y be two independent, normally distributed random variables, say X with a $N(\mu, \sigma^2)$ distribution, and Y with a $N(\rho, \tau^2)$ distribution. By Theorem 2.5 we have, that

$$M_{X+Y}(t) = M_X(t) M_Y(t) = \mathrm{e}^{\frac{(\sigma^2+\tau^2)t^2}{2} + (\mu+\rho)t}, \qquad -\infty < t < \infty.$$

We find that $X + Y$ has a $N(\mu + \rho, \sigma^2 + \tau^2)$ distribution.

Quick exercise 2.9 Let X be a random variable, with moment generating function $M_X(t)$, and let $a, b \in \mathbb{R}$. Show that

$$M_{aX+b}(t) = \mathrm{e}^{bt} M_X(at).$$

2.4 Convergence in distribution

Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed random variables, say with finite expectation μ and finite variance σ^2. Setting

$$Z_n = \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma}, \qquad \text{for } n \geq 1,$$

then the *central limit theorem* tells us that as n tends to infinity, the distribution function of Z_n converges to the distribution function Φ of the standard normal distribution. This is an example of random variables converging in distribution. Here is the definition.

Convergence in distribution. Let $(X_n)_{n \geq 1}$ be a sequence of random variables. Then this sequence *converges in distribution* to a random variable X if

$$F_{X_n}(x) = \mathrm{P}(X_n \leq x) \to \mathrm{P}(X \leq x) = F_X(x), \qquad \text{as } n \to \infty,$$

for all points x where F_X is continuous.

Examples 2.5

(a) For $n \geq 1$, let the distribution of the random variables X_n be given by

$$\mathrm{P}\left(X_n = \frac{k}{n}\right) = \frac{1}{n}, \qquad \text{for } k = 0, 1, \ldots, n-1.$$

Furthermore, let X be a $U(0,1)$ distributed random variable. Then X_n converges to X in distribution.

(b) In example (a) the distribution function F of the random variable X is continuous on $\mathbb{R}$. In order to see why we demanded in the definition of convergence in distribution that the convergence holds 'for all points x where F_X is continuous', we have the following example. For $n \geq 1$, let the distribution of random variables X_n and Y_n be given by

$$\mathrm{P}\left(X_n = \frac{1}{n}\right) = 1, \quad \mathrm{P}\left(Y_n = -\frac{1}{n}\right) = 1, \qquad \text{for } n \geq 1.$$

Let U be the random variable with $\mathrm{P}(U = 0) = 1$. Intuitively both X_n and Y_n converge to U in distribution. Indeed, for all $u \in \mathbb{R}$ one has that

$$F_{Y_n}(u) = \mathrm{P}(Y_n \leq u) \to \mathrm{P}(U \leq u) = F_U(u),$$

so certainly $F_{Y_n}(u) \to F_U(u)$ for all the points u where F_U is continuous. However, for all n we have that $F_{X_n}(0) = 0$, and therefore

$$0 = F_{X_n}(0) \not\to F_U(0) = 1.$$

According to the definition of convergence in distribution we can still say that X_n converges to U in distribution; $u = 0$ is excluded since F_U is discontinuous in $u = 0$. □

Whether or not a sequence of random variables X_n converges in distribution to a random variable X can be 'seen' from the behavior of their moment generating functions. We have the following result, which we mention without proof.

Theorem 2.7 Let $(X_n)_{n \geq 1}$ be a sequence of random variables. Then this sequence converges in distribution to a random variable X if and only if

$$\lim_{n \to \infty} M_{X_n}(t) = M_X(t),$$

for all $t \in [-t_0, t_0]$ for some $t_0 > 0$.

2.5 The central limit theorem

In Chapter 14 of [5] a basic version is given of the *central limit theorem*, and some applications of this classical result are worked out. Here we will present an easy proof of a slightly different version of this theorem, using moment generating functions. Before giving the proof, we state the theorem.

The central limit theorem. Let $X_1, X_2, \ldots$ be any sequence of independent identically distributed random variables with finite positive variance. Let $M_{X_i}(t)$ be the moment generating function of X_i, and suppose that $M_{X_i}(t)$ exists for some $t \neq 0$. Let μ be the expected value and σ^2 the variance of each of the X_i's. For $n \geq 1$, let Z_n be defined by

$$Z_n = \sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma},$$

then for any number a

$$\lim_{n\to\infty} F_{Z_n}(a) = \Phi(a),$$

where Φ is the distribution function of the $N(0,1)$ distribution. In words: the distribution function of Z_n converges to the distribution function Φ of the standard normal distribution.

Proof. Since we can replace the random variable $X_n - \mu$ by Y_n, for $n \geq 1$, we might assume without loss of generality that $\mu = \mathrm{E}[X_n] = 0$. Since the X_i's all have the same distribution, it follows that they have the same moment generating function: $M_{X_i}(t) = M(t)$ for all $i \geq 1$. Since the X_i's are independent, it follows from Theorem 2.5 that $(M(t))^n$ is the moment generating function of $X_1 + X_2 + \cdots + X_n$. Due to Quick exercise 2.9 it follows that the function $M_n(t)$, defined by

$$M_n(t) = \left(M\left(\frac{t}{\sigma\sqrt{n}}\right)\right)^n,$$

is the moment generating function of

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n}{\sigma\sqrt{n}}.$$

Theorem 2.4 now yields that for $t \to 0$,

$$\begin{aligned} M(t) &= M(0) + M'(0)t + \frac{1}{2}M''(0)t^2 + \mathcal{O}(t^3) \\ &= 1 + \frac{1}{2}\sigma^2 t^2 + \mathcal{O}(t^3), \end{aligned}$$

since $\mu = M'(0) = 0$, and $M''(0) = \sigma^2$. But then we find that

$$M_n(t) = \left(1 + \frac{1}{2}\frac{t^2}{n} + \mathcal{O}\left((t/\sqrt{n})^3\right)\right)^n.$$

Recall the following exercise from your calculus class.

A calculus exercise. Let $(a_n)_{n\geq 1}$ be a sequence of real number converging to $a \in \mathbb{R}$. Then

$$\lim_{n\to\infty}\left(1 + \frac{a_n}{n}\right)^n = \mathrm{e}^a.$$

An immediate consequence of this 'exercise' is, that

$$\lim_{n\to\infty} M_n(t) = \mathrm{e}^{\frac{1}{2}t^2}.$$

From Quick exercise 2.8 we know that $\mathrm{e}^{\frac{1}{2}t^2}$ is the moment generating function of a standard normally distributed random variable, so due to Theorem 2.6 we find that Z_n converges to a $N(0,1)$ distributed random variable as n tends to infinity. □

2.6 Solutions to the quick exercises

2.1 In the first example the series converges for all $s \in \mathbb{R}$, so $R = \infty$. In the second example we have convergence if $|s| < 1$, and divergence if $|s| > 1$, so $R = 1$. Note that in this last example the series need not converge for s with $|s| = 1$.

2.2 Let n be a fixed non-negative integer, and set for $k = 0, 1, \ldots, n$, $a_k = \binom{n}{k}$, and $a_k = 0$ for $k \geq n+1$. Let $A(s)$ be the generating function of the sequence $(a_k)_{k\geq 0}$, i.e., $A(s) = (1+s)^n$. Now recall that for $k = 0, 1, \ldots, n$,

$$a_k = \binom{n}{k} = \binom{n}{n-k} = a_{n-k},$$

i.e.,

$$\binom{n}{k}^2 = \binom{n}{k}\binom{n}{n-k} = a_k a_{n-k}.$$

Let $(c_n)_{n\geq 0}$ be the convolution of $(a_k)_{n\geq 0}$ with itself, then

$$C(s) = (A(s))^2 = ((1+s)^n)^2 = (1+s)^{2n}.$$

Now c_n is the coefficient of s^n in $(1+s)^{2n}$, so

$$c_n = \sum_{k=0}^{n} \binom{n}{k}^2 = \sum_{k=0}^{n} a_k a_{n-k}.$$

We find that $c_n = \binom{2n}{n}$.

2.3 Since $\sum_{k=0}^{\infty} \mathrm{P}(X = k) = 1$, we find that the radius of convergence of $G_X(s)$ is at least 1.

2.4 This is a simple consequence of the change of variable formula from [5], Section 7.3. Setting $g(x) = s^x$, we find that

$$\mathrm{E}[g(X)] = \sum_{k=0}^{\infty} g\left(s^k\right) \mathrm{P}(X = k)\,.$$

2.5 Since the radius of convergence if $G_X(s)$ is at least 1, we find that $G_X^{(k)}(0)$ exist, and can be obtained by differentiating k times every term in $G_X(s)$, substitute s by 0, and take the sum of the term. This yields

$$P(X = k) = \frac{1}{k!} G_X^{(k)}(0), \qquad \text{for } k = 0, 1, 2, \ldots.$$

Taking $k = 0$ then gives $P(X = 0) = G_X(0)$, while $k = 1$ yields $P(X = 1) = G'_X(0)$.

2.6 In Example 2.1(b) we have seen that $G_X(s) = (1 - p + ps)^n$. But then

$$G'_X(s) = np(1 - p + ps)^{n-1} \quad \text{and} \quad G''_X(s) = n(n-1)p^2(1 - p + ps)^{n-2},$$

and it follows from Theorem 2.2 that

$$E[X] = G'_X(1) = np, \quad \text{and} \quad \mathrm{Var}(X) = n(n-1)p^2 + np - (np)^2 = np(1-p).$$

2.7 Let $t > 0$ (the case $t < 0$ runs along similar lines). Then there exists a positive integer n_0 such that $e^{nt} \geq 2(n-1)$ for all $n \geq n_0$. But then

$$E\left[e^{tX}\right] \geq \sum_{n=n_0}^{\infty} e^{nt} P(X = n) \geq \sum_{n=n_0}^{\infty} \frac{1}{n} = \infty.$$

We see that $M_X(t)$ does not converge to a finite limit for $t \neq 0$.

2.8 By definition,

$$\begin{aligned} M_X(t) = E\left[e^{tX}\right] &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x^2 - 2\mu x + \mu^2 - 2\sigma^2 t x)/2\sigma^2} dx. \end{aligned}$$

Now applying some high-school math yields that

$$\begin{aligned} x^2 - 2\mu x + \mu^2 - 2\sigma^2 t x &= x^2 - 2(\mu + \sigma^2 t)x + \mu^2 \\ &= \big(x - (\mu + \sigma^2 t)\big)^2 - \sigma^4 t^2 + \mu^2, \end{aligned}$$

implying that

$$\begin{aligned} M_X(t) &= \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{\sigma^4 t^2 + 2\mu\sigma^2 t}{2\sigma^2}} \int_{-\infty}^{\infty} e^{-\left(x - (\mu + \sigma^2 t)\right)^2/2\sigma^2} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{\sigma^2 t^2}{2} + \mu t} \int_{-\infty}^{\infty} e^{-\left(x - (\mu + \sigma^2 t)\right)^2/2\sigma^2} dx. \end{aligned}$$

Since

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\left(x - (\mu + \sigma^2 t)\right)^2/2\sigma^2}$$

is the probability density function of a $N(\mu + \sigma^2 t, \sigma^2)$ distribution, we find that

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-\left(x-(\mu+\sigma^2 t)\right)^2/2\sigma^2} \mathrm{d}x = 1,$$

implying that

$$M_X(t) = \mathrm{e}^{\frac{\sigma^2 t^2}{2} + \mu t},$$

as desired.

2.9 By definition,

$$M_{aX+b}(t) = \mathrm{E}\left[\mathrm{e}^{t(aX+b)}\right] = \mathrm{E}\left[\mathrm{e}^{atX}\mathrm{e}^{bt}\right] = \mathrm{e}^{bt}\mathrm{E}\left[\mathrm{e}^{atX}\right] = \mathrm{e}^{bt} M_X(at).$$

2.7 Exercises

2.1 Let X be a discrete random variable, with probability mass function given by

$$\mathrm{P}(X = i) = \frac{1}{4}, \qquad \text{for } i = 0, 1, 2, 3.$$

a. Determine the probability generating function of X.
b. Use **a.** to find the expected value and the variance of X.

2.2 Let X be an $\mathbb{N}$-valued random variable, with probability generating function G_X, given by

$$G_X(s) = A(1 + s + s^2).$$

Here $A \in \mathbb{R}$ is a constant.

a. Determine A.
b. Determine the probability generating function of $U = X + 1$.
c. Determine the probability generating function of $V = 3X$.
d. Determine the probability generating function of $W = X^2$.
e. Is $\left(G_X(s)\right)^2$ also a probability generating function?

2.3 Let $G(s)$ be a probability generating function. Does there exist a constant K, such that $KG'(s)$ is also a probability generating function?

2.4 Let X be a *Geo*(p) distributed random variable. Determine the probability generating function of X.

2.5 Let $X_1, X_2, \ldots$ be a sequence of independent Bernoulli variables with parameter p, and let N be a random variable with a *Pois*(μ) distribution, which is independent of the X_i's. Furthermore, let $S = X_1 + X_2 + \cdots + X_N$.

a. Determine the probability generating function of X.
b. What type of distribution does S have?

2.6 Let X and Y be two independent random variables, where X has a $Bin(n,p)$ distribution, and Y has a $Bin(m,p)$ distribution. Use probability generating functions to determine the distribution of $X+Y$.

2.7 Let the probability generating function of the random variable X be given by

$$G(s) = \frac{1+s-s^6-s^7}{12-12s},\ s \neq 1, \qquad \text{and } G_X(1) = 1.$$

a. Determine $\mathrm{P}(X=0)$ and $\mathrm{P}(X=1)$.
b. Determine $\mathrm{P}(X=666)$. *Hint:* $\sum_{k=0}^{\infty} s^k = 1/(1-s)$.

2.8 Let n be a fixed positive integer. Someone throws a fair dice until 6 shows for the nth time. What is the expected number of times this person will throw?

2.9 Let X be a discrete random variable, with probability mass function given by

$$\mathrm{P}(X=i) = \frac{1}{4}, \qquad \text{for } i = 0,1,2,3.$$

a. Determine the moment generating function of X.
b. Use **a.** to find the expected value and the variance of X.

2.10 Let X and Y be two independent random variables, where X has a $U(0,1)$ distribution, Y a discrete uniform distribution over $k = 0,1,\ldots,9$, i.e., $\mathrm{P}(Y=k) = 1/10$ for $k = 0,1,\ldots,9$.

a. Heuristically, what do you think is the distribution of $X+Y$?
b. Determine the moment generation function of $X+Y$. What does this imply about the type of distribution of $X+Y$?

2.11 Let X have a gamma distribution, with parameters n and λ, i.e., let the probability density function of X be given by

$$f(x) = \frac{1}{\Gamma(s)}\lambda^s x^{s-1}\mathrm{e}^{-\lambda x}, \qquad \text{for } x \geq 0,$$

and $f(x) = 0$ for $x < 0$ (see e.g. [5], Section 11.2 for the definition of the gamma distribution).

a. Show that the moment generating function M_X of X satisfies

$$M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^n, \qquad \text{for } t < \lambda.$$

b. Use M_X to find the expectation and variance of X.

c. Determine the moment generating function of $Y = \lambda X$. What type of distribution does Y have?

2.12 Let X and Y be two independent, gamma distributed random variables, with parameters s_x and λ_X, respectively s_Y and λ_Y. Determine the moment generating function of $X + Y$. What type of distribution does $X + Y$ have?

2.13 In [5], Section 11.2, it was shown—after tedious calculations—that the sum of n independent $\mathit{Exp}(\lambda)$ distributed random variables has a $\mathit{Gam}(n, \lambda)$ distribution. Show this using moment generating functions.

2.14 Let Y_n have a $\mathit{Geo}(\lambda/n)$ distribution, i.e.,

$$\mathrm{P}(Y_n = k) = \frac{\lambda}{n}\left(1 - \frac{\lambda}{n}\right)^{k-1}, \qquad \text{for } k = 1, 2, \ldots.$$

Show that $\frac{1}{n}Y_n$ converges in distribution to an exponential distribution.

3

Discrete-time Markov chains

Up to now we primarily encountered (finite) samples; (finite) sequences of independent random variables $X_1, X_2, \dots, X_n$, each having the same distribution. These samples were used to obtain information about some parameter of interest. Apart from using sequences of random variables in statistics, one can also use them to 'describe the state of a system'. For instance, X_n may denote the number of customers in a shop at 'time n', where the time step n is measured in minutes, or seconds, or in any convenient discrete time step. Of course one could also continuously report the state of the system; we consider random variables X_t, with $t \in I$, where I is some interval. For example, $I = [0, \infty)$ or $I = [0, 1]$. In both the discrete and the continuous case we call $\{X_n\}_{n \in \mathbb{N}}$ respectively $\{X_t\}_{t \in I}$ a *stochastic process*. In contrast to a sample we do not demand that the random variables in a stochastic process are either independent, or identically distributed, only that the random variables X_n (resp. X_t) are all defined on the *same* sample space S. In fact, describing the state of a system, assumptions such as independence seem highly unlikely and questionable.

In this chapter we will introduce a stochastic process where *given* the state of the system at the present time n, the future is independent of the past. Although this seems only a small step away from the models we studied up to this point, where all the random variables are independent, it turns out that this stochastic process, which is called a *Markov chain* after A.A. Markov[1], is widely applicable, and one of the most important stochastic processes studied. In this chapter discrete Markov chains will be introduced, and some of its elementary properties studied. In the next chapter we will study the long-term behavior of these stochastic processes. In Chapter 5 continuous time Markov chains will be introduced and studied.

[1] In fact, Markov was not the first to study the stochastic processes we now call Markov processes.

3.1 An example

As an example of the kind of stochastic process we have in mind, suppose that we have a gas of N molecules, moving freely between two connected chambers. How can we describe this in a simple, more-or-less realistic manner? Let the chambers be denoted by A and B, and let X_n be the number of molecules in chamber A at time n. In the 1930ies, Paul Ehrenfest modeled the situation along the following lines. Suppose the molecules are numbered from 1 to N, and at time n we select one of these numbers/molecules at random (so the ith molecule is selected with probability $1/N$), and move it to the other chamber. Note that knowing the realization of X_n at time n determines the possible realizations (and the probabilities of these realizations) of X_{n+1} at time $n+1$. For instance, if all the molecules are in chamber A at time n, then one molecule will be in chamber B at time $n+1$ with probability one, so

$$\mathrm{P}(X_{n+1} = N-1 \mid X_n = N) = 1.$$

Similarly, $\mathrm{P}(X_{n+1} = 1 \mid X_n = 0) = 1$, and more generally

$$\mathrm{P}(X_{n+1} = i-1 \mid X_n = i) = \frac{i}{N}, \qquad \text{for } i = 1, 2, \ldots, N,$$

and

$$\mathrm{P}(X_{n+1} = i+1 \mid X_n = i) = \frac{N-i}{N}, \qquad \text{for } i = 0, 1, \ldots, N-1.$$

Due to the way we modeled the movement of the gas we moreover have that

$$\mathrm{P}(X_{n+1} = j \mid X_n = i) = 0 \qquad \text{for } j = 0, 1, \ldots, N,\ j \neq i-1, i+1.$$

Note that X_{n+1} is determined *only* by X_n. In fact we will see that for processes such as in this example one has that *given* X_n, the random variable X_{n+1} is independent of X_k, for $k \leq n-1$. I.e., for all $i, x_{n-1}, \ldots, x_0 \in \{0, 1, \ldots, N\}$ with

$$\mathrm{P}(X_n = i, X_{n-1} = x_{i-1}, \ldots, X_0 = x_0) > 0,$$

we have that

$$\mathrm{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = x_{i-1}, \ldots, X_0 = x_0) = \mathrm{P}(X_{n+1} = j \mid X_n = i).$$

This is called the *Markov property*.

Since the conditional probabilities in our example do not depend on n, we define for all $n \geq 0$

$$p_{i,j} = \mathrm{P}(X_{n+1} = j \mid X_n = i) \qquad \text{for } i, j \in S = \{0, 1, \ldots, N\}\ .$$

We say that the stochastic process $\{X_n\}_{n \in \mathbb{N}}$ is a *time homogeneous* Markov chain, on the *state space* $S = \{0, 1, \ldots, N\}$.

Quick exercise 3.1 In Ehrenfest's model of a gas moving between chambers, determine the matrix $P = (p_{i,j})$ of transition probabilities for $N = 5$. Also, show that the rows of this matrix add up to 1, i.e., show that for $i = 0, 1, \dots, 5$,

$$\sum_{j=0}^{5} p_{i,j} = 1.$$

In order to know the long-term behavior of the system, i.e., to know the distribution of X_n for large n, it seems to be important to know how the system started out. That is, it seems to be important how the molecules were distributed over the two chambers A and B at time 0. After all, due to the total law of probability (see e.g. [5], Section 3.3), we have that

$$\mathrm{P}(X_n = j) = \sum_{i=0}^{N} \mathrm{P}(X_n = j \mid X_0 = i)\mathrm{P}(X_0 = i)\,, \qquad \text{for } i, j \in S.$$

Let the vector $\mu = (\mu_0 \;\; \mu_1 \;\; \cdots \;\; \mu_N)$ describe the *initial distribution*, i.e.,

$$\mathrm{P}(X_0 = i) = \mu_i, \qquad \text{for } i \in S,$$

where—of course—we have that $\mu_i \geq 0$ and $\mu_0 + \mu_1 + \cdots + \mu_N = 1$.
For example, if at time 0 all the molecules are in chamber A, then $\mu_N = 1$ (and $\mu_i = 0$ for all other i's). But then

$$\mathrm{P}(X_2 = j) = \begin{cases} \frac{N-1}{N} & \text{if } j = N-2 \\ \frac{1}{N} & \text{if } j = N \\ 0 & \text{for all other values of } j \in S. \end{cases}$$

Quick exercise 3.2 In our molecules example, suppose that $\mu_i = \binom{N}{i}2^{-N}$ for $i \in S$. Then $\mathrm{P}(X_n = j) = \binom{N}{j}2^{-N}$ for each $j \in S$ and each $n \geq 1$. Show this holds for $N = 5$, $n = 1$, and for a general $n \geq 2$. Why is this initial distribution not so far fetched as it may seem to be at first view?

So we showed in this Quick exercise, that if the initial distribution μ of the number of molecules in chamber A is $Bin(N, \frac{1}{2})$ at time $n = 0$, it will remain so for all n. In fact we will see in the next chapter for this particular example, whatever the initial distribution, the distribution of the X_n's will be approximately $Bin(N, \frac{1}{2})$ for n large, i.e., that

$$\lim_{n\to\infty} \mathrm{P}(X_n = j) = \binom{N}{j}2^{-N}, \qquad \text{for all } j \in S.$$

In this example the $Bin(N, \frac{1}{2})$ distribution is the so-called *stationary distribution*, and we will see in Theorem 4.? that the distribution of the X_n's converges to this stationary distribution.

3.2 Definition of a discrete Markov chain

A (discrete) Markov chain is a discrete stochastic process $(X_n)_{n\geq 0}$, taking its values in a finite or countably infinite set S, which is called the *state space*. The elements of S are called *states*. The process starts at time $n = 0$ in one of the states (in which state it starts is determined by the initial distribution μ), and then 'moves' at integer times from one state to the other. The conditional probability p_{ij} that[2] the process 'moves' to state j, *given* that it is currently in state i, only depends on the current state i, and does **not** depend on the way it got to this state i. The process can also remain in state i; this happens with probability p_{ii}. So we must have that

$$p_{ij} \geq 0, \qquad \sum_{j\in S} p_{ij} = 1,$$

and

$$\mathrm{P}(X_{n+1} = j \,|\, X_n = i, X_{n-1} = x_{i-1}, \ldots, X_0 = x_0) = \mathrm{P}(X_{n+1} = j \,|\, X_n = i),$$

which is m — as already mentioned in the previous section — the so-called Markov property. The *transition matrix* P is the matrix consisting of the various transition probabilities: $P = (p_{i,j})$. So the matrix P is a *Markov matrix*, i.e., it is a square matrix with non-negative entries, whose rows add up to 1. Such a matrix is also called a *stochastic matrix*. For example, if $S = \{1, 2, \ldots, N\}$, then

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & \cdots & p_{2N} \\ \vdots & \vdots & & & \vdots \\ p_{N1} & p_{N2} & \cdots & \cdots & p_{NN} \end{pmatrix}.$$

The Markov property can be generalized as follows. Let $K_n \subset S^n$ be a subset of the set S^n of vectors of length n with entries from S, and let V be the event, given by

$$V = \{(X_0, X_1, \ldots, X_{n-1}) \in K_n\}.$$

Then the Markov property states that

$$\mathrm{P}(X_{n+1} = j \,|\, X_n = i, V) = \mathrm{P}(X_{n+1} = j \,|\, X_n = i) = p_{ij}.$$

Let $m \geq 1$, and let the event Z for $L_m \subset S^m$ be given by

$$Z = \{(X_{n+1}, X_{n+2}, \ldots, X_{n+m}) \in L_m\},$$

then we find that the Markov property can also be given by

[2] Note that in the previous section we denoted these transition probabilities as $p_{i,j}$.

$$P(Z \mid X_n = i, V) = P(Z \mid X_n = i). \tag{3.3.1}$$

We have the following theorem, which states that in a Markov chain, *given* that the chain is now in state i, the 'future' (which is the event Z) is independent of the 'past' (i.e., the event V). Conversely, if the 'future' is independent of the 'past', *given* the present state i, then the stochastic process is a Markov chain.

Theorem 3.1 Let V and Z be defined as above. Then the Markov property (3.3.1) is equivalent with

$$P(V \cap Z \mid X_n = i) = P(V \mid X_n = i)P(Z \mid X_n = i) \tag{3.3.2}$$

Proof. Suppose that the Markov property (3.3.1) holds. Then

$$\begin{aligned} P(V \cap Z \mid X_n = i) &= \frac{P(Z, X_n = i, V)}{P(X_n = i)} = \frac{P(Z, X_n = i, V)}{1} \frac{1}{P(X_n = i)} \\ &= \frac{P(Z, X_n = i, V)}{P(X_n = i, V)} \frac{P(V, X_n = i)}{P(X_n = i)} \\ &= P(Z \mid X_n = i, V)P(V \mid X_n = i) \quad \text{(use (3.3.1))} \\ &= P(Z \mid X_n = i)P(V \mid X_n = i). \end{aligned}$$

Conversely, if (3.3.2) holds, then

$$\begin{aligned} P(V \mid X_n = i)P(Z \mid X_n = i) &= P(V \cap Z \mid X_n = i) \\ &= P(Z \mid X_n = i, V)P(V \mid X_n = i), \end{aligned}$$

which yields (3.3.1). □

3.3 Classification of the states

In the molecules example in the previous section, it is intuitively clear that every state j can be 'reached' from every state i. By this we mean, that for every $i, j \in S$ there exists a non-negative integer n, such that $P(X_n = j \mid X_0 = i) > 0$. We denote this last probability by $p_{i,j}^{[n]}$, i.e., the n-step transition probability is given by

$$p_{i,j}^{[n]} = P(X_n = j \mid X_0 = i), \qquad \text{for } i, j \in S \text{ and } n \geq 0,$$

and we write[3] $i \to j$. In particular we have for $n = 0$ that

$$p_{ij}^{[0]} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j. \end{cases}$$

[3] Some authors, see e.g. [6] and [7], say that i and j *communicate*, but to me this is a rather one-way form of communication.

Note, that since our Markov chains are time homogeneous, we have for every $k \leq 0$ that $p_{i,j}^{[n]} = \mathrm{P}(X_{k+n} = j \mid X_k = i)$, for $i, j \in S$ and $n \geq 0$. Obviously, $p_{ij}^{[1]} = p_{ij}$ for all $i, j \in S$.

Theorem 3.2 (Chapman-Kolmogorov) Let $(X_n)_{n\geq 0}$ be a discrete time-homogeneous Markov chain on the state space S. Furthermore, let $p_{ij}^{[n]}$ be the n-step transition probabilities of this Markov chain. Then for $i, j \in S$ and $m, n \geq 0$ we have

$$p_{ij}^{[m+n]} = \sum_{k \in S} p_{ik}^{[m]} \cdot p_{kj}^{[n]}.$$

Proof. The proof rests on the total law of probability ([5], Section 3.3), and the Markov property (3.3.1). We have that

$$\begin{aligned}
p_{ij}^{[m+n]} &= \mathrm{P}(X_{m+n} = j \mid X_0 = i) \\
&= \sum_{k \in S} \mathrm{P}(X_{m+n} = j, X_m = k \mid X_0 = i) \\
&= \sum_{k \in S} \frac{\mathrm{P}(X_{m+n} = j, X_m = k, X_0 = i)}{\mathrm{P}(X_m = k, X_0 = i)} \frac{\mathrm{P}(X_m = k, X_m = k, X_0 = i)}{\mathrm{P}(X_0 = i)} \\
&= \sum_{k \in S} \mathrm{P}(X_{m+n} = j \mid X_m = k, X_0 = i)\mathrm{P}(X_m = k \mid X_0 = i) \\
&= \sum_{k \in S} \mathrm{P}(X_{m+n} = j \mid X_m = k)\mathrm{P}(X_m = k \mid X_0 = i) \\
&= \sum_{k \in S} p_{ik}^{[m]} \cdot p_{kj}^{[n]}.
\end{aligned}$$

□

This theorem has a nice corollaries; In the first corollary the n-step transition probabilities $p_{ij}^{[n]}$ are linked to powers of the transition matrix P.

Corollary 3.1 Let $(X_n)_{n\geq 0}$ be a discrete time-homogeneous Markov chain on the state space S. Furthermore, let $p_{ij}^{[n]}$ be the n-step transition probabilities of this Markov chain, and let $P^{[n]} = \left(p_{ij}^{[n]}\right)$ be the matrix of the n-step transition probabilities. Then for $m, n \geq 0$ we have

$$P^{[m+n]} = P^{[m]} P^{[n]}.$$

In particular, for $n \geq 0$ we have that $P^{[n]} = P^n$.

The idea behind the proof of the second corollary is the same idea behind the solution of Quick exercise 3.2; the law of total probability, i.e.,

$$\mathrm{P}(X_n = j) = \sum_{i \in S} \mathrm{P}(X_n = j \mid X_0 = i)\mathrm{P}(X_0 = i)\,.$$

We have the following corollary.

Corollary 3.2 Let $(X_n)_{n\geq 0}$ be a discrete time-homogeneous Markov chain on the state space S. Furthermore, let $p_{ij}^{[n]}$ be the n-step transition probabilities of this Markov chain, and let μ be the vector of initial probabilities, i.e., $\mu_i = \mathrm{P}(X_0 = i)$ for $i \in S$. Then for $n \geq 0$ and $j \in S$ we have

$$\mathrm{P}(X_n = j) = (\mu P^n)_i = \sum_{i\in S} \mu_i p_{ij}^{[n]}.$$

Now let $i, j \in S$ be two states, such that i can be reached from j, and conversely, j can be reached from i. I.e., there exist $m, n \geq 0$ such that $p_{ij}^{[m]} > 0$ and $p_{ji}^{[n]} > 0$. In this case we[4] say that i and j *communicate*, and we write $i \leftrightarrow j$. In the molecules-example each state i communicates with each other state j, but in general this needs not to be the case. In fact, $\leftrightarrow$ is an equivalence relation on S, and consequently S can be written as the disjoint union of subsets of elements of S (the so-called equivalence classes) which only communicate with one-another. To show that $\leftrightarrow$ is indeed an equivalence relation on S is not very hard. Obviously $\leftrightarrow$ is reflexive (since by definition $p_{ii}^{[0]} = 1$ for every $i \in S$), and symmetric. Finally, the relation $\leftrightarrow$ is also transitive; suppose that $i \leftrightarrow j$ and $j \leftrightarrow k$, then there exist $n, m \geq 0$ such that $p_{ij}^{[n]} > 0$ and $p_{jk}^{[m]} > 0$. But then, due to the Theorem of Chapman-Kolmogorov, we have that

$$p_{ik}^{[n+m]} = \sum_{\ell\in S} p_{i\ell}^{[n]} \cdot p_{\ell k}^{[m]} \geq p_{ij}^{[n]} \cdot p_{jk}^{[m]} > 0,$$

i.e., $i \to k$. In the same way one finds that i can be reached from k. If there is only one equivalence class (so when all the states communicate with one-another), we say that the Markov chain is *irreducible*. In this case also the matrix of transition probabilities P is called irreducibe.

Examples 3.1

Let $(X_n)_{n\geq 0}$ be a Markov chain on the state space $S = \{1, 2, 3, 4\}$, and let P be the matrix of transition probabilities. We will consider three different P's, and see that the behavior of the chain is qualitatively very different for each of these three cases.

(a) Let

$$P = \begin{pmatrix} \frac{2}{3} & 0 & \frac{1}{3} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & \frac{1}{9} & 0 & \frac{8}{9} \end{pmatrix}.$$

It is also very instructive to put in one figure the various transitions between the states, see Figure 3.1. With the aid of Figure 3.1 one convinces oneself

[4] Some authors, see e.g. [6] and [7], say that i and j *intercommunicate*, see also the footnote on the previous page.

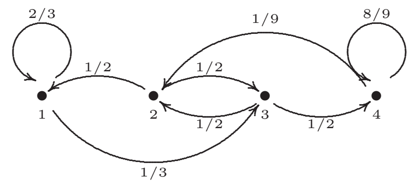

Fig. 3.1. The various transitions in Example 3.1(a)

quickly that there is only one communicating class; the Markov chain is irreducible.

(b) Now suppose that P is given as follows:

$$P = \begin{pmatrix} \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{9} & 0 & 0 & \frac{8}{9} \end{pmatrix}.$$

From Figure 3.2 we see that there are two equivalence classes: $\{1,4\}$ and $\{2,3\}$. The Markov chain is reducible; each of these classes acts as a "subworld".

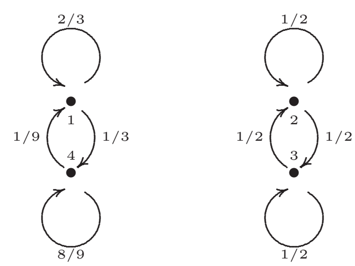

Fig. 3.2. The transitions in Example 3.1(b)

(c) Finally, let P be given as follows:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{9} & 0 & 0 & \frac{8}{9} \end{pmatrix}.$$

From Figure 3.3 it is now clear that there are three classes: $\{1\}$, $\{2,3\}$, and $\{4\}$. The last two classes are special; they are *transient*. Starting in $\{2,3\}$ or in $\{4\}$, one will eventually leave this class, never to return. Note that the time it takes to leave the class $\{4\}$ is geometrically distributed, with parameter $p = 8/9$. The class $\{1\}$ is *recurrent*; it will occur infinitely often.

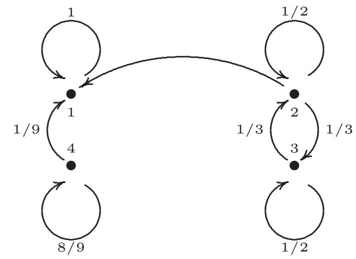

Fig. 3.3. The transitions in Example 3.1(c)

Note that for a reducible Markov chain the recurrent states are like 'sub-worlds', behaving as irreducible Markov chains (once you get in such a state, you'll never get out again). It is for this reason that we often only will consider irreducible Markov chains; reducible ones can be studied piecemeal. Not only classes can be recurrent, also states can be so. It is intuitively clear, that if one state in a class is recurrent (transient), all the other states in that class are also recurrent (transient) (since all the states communicate with each other). We will give a proof of this at the end of this section. We start with a formal definition of a recurrent (transient) state.

Recurrent and transient states. Let $(X_n)_{n\geq 0}$ be a discrete time-homogeneous Markov chain on the state space S. A state $i \in S$ is called *recurrent* (or *persistent*) if

$$f_i = \mathrm{P}(X_n = i \text{ for some } n \geq 1 \,|\, X_0 = i) = 1.$$

A state i which is not recurrent is called *transient*.

Setting

$$f_{ii}^{[n]} = \begin{cases} \mathrm{P}(X_1 = i \,|\, X_0 = i) & n = 1 \\ \mathrm{P}(X_n = i, X_m \neq i, 1 \leq m < n \,|\, X_0 = i) & n > 1, \end{cases}$$

we have that

$$f_i = \sum_{n=1}^{\infty} f_{ii}^{[n]}.$$

Clearly, $f_{ii}^{[1]} = p_{ii}$ for all states i, and for all $n \geq 2$,

$$f_{ii}^{[n]} \leq p_{ii}^{[n]}.$$

Now let N_i be the number of visits to the state i:

$$N_i = \sum_{n=0}^{\infty} 1_{\{X_n = i\}}.$$

Clearly, $\mathrm{P}(N_i = 1 \mid X_0 = i) = 1 - f_i$. In Exercise 3.11 you are invited to show that for $k \geq 1$,

$$\mathrm{P}(N_i = k+1 \mid X_0 = i) = f_i \mathrm{P}(N_i = k \mid X_0 = i). \tag{3.3.3}$$

Repeating this another $k-1$ times we find that

$$\mathrm{P}(N_i = k+1 \mid X_0 = i) = (1-f_i) f_i^{k-1}, \qquad \text{for } k = 1, 2, \ldots.$$

For a recurrent state i this yields that

$$\mathrm{P}(N_i = \infty \mid X_0 = i) = 1,$$

while for a transient state i we find that

$$\mathrm{E}(N_i \mid X_0 = i) = \frac{1}{1-f_i}.$$

Since

$$\mathrm{E}(N_i \mid X_0 = i) = \sum_{n=0}^{\infty} \mathrm{E}(1_{\{X_n = i\}} \mid X_0 = i) = \sum_{n=0}^{\infty} p_{ii}^{[n]},$$

we find the following characterization of the states.

Theorem 3.3 The state i is recurrent if and only if $\displaystyle\sum_{n=0}^{\infty} p_{ii}^{[n]} = \infty$.

Equivalently, the state i is transient if and only if $\displaystyle\sum_{n=0}^{\infty} p_{ii}^{[n]} < \infty$.

An immediate consequence of this theorem is the following corollary.

Corollary 3.3 If i is a transient state, then $p_{ii}^{[n]} \to 0$ as $n \to \infty$.

In Exercises 3.12 you will show that that two communicating states are either both recurrent or both transient. Furthermore, in Exercise 3.13 you will show that any state j which can be reached from a recurrent state i communicates with i, and is therefore also recurrent. In view of this is seems natural to call an equivalence class recurrent if one (and thus all) of its states is recurrent, and to call it transient otherwise. The following—classical—example shows that somehow the concept of a recurrent state is not completely satisfactory.

The simple random walk

Up to now we have seen only examples of Markov chains with a finite state space S. One of the oldest and easiest examples of a Markov chain on an infinite state space S is the so-called *simple random walk*. Let $S = \mathbb{Z}$, the collection of the real integers, and let the transition probabilities be given by

$$p_{i,i+1} = p, \quad p_{i,i-1} = q = 1-p, \qquad \text{for some } p \in (0,1),$$

Fig. 3.4. The various transitions in Example 3.1(a)

and $p_{ij} = 0$ for $j \notin \{i-1, i+1\}$; see also Figure 3.3.
Since $p \neq 0, 1$ it is clear that the Markov chain is irreducible; all states i and j communicate with one-another. Due to Exercise 3.12 we know that every state is either recurrent, or transient. In view of Theorem 3.3 we need to find $\sum_{n=0}^{\infty} p_{ii}^{[n]}$.

Quick exercise 3.3 Show that for every $n \geq 1$ one has, that

$$p_{ii}^{[n]} = \begin{cases} 0 & \text{when } n \text{ is odd} \\ \binom{2k}{k} p^k q^k & \text{when } n = 2k. \end{cases}$$

From this quick exercise we see that

$$\sum_{n=0}^{\infty} p_{ii}^{[n]} = \sum_{k=0}^{\infty} \binom{2k}{k} p^k q^k.$$

Using Stirling's formula, which states that

$$k! \sim \left(\frac{k}{\mathrm{e}}\right)^k \sqrt{2\pi k}, \qquad k \to \infty,$$

one finds that

$$\binom{2k}{k} = \frac{(2k)!}{k!} \sim \frac{(2k/\mathrm{e})^{2k}\sqrt{4\pi k}}{\Big((k/\mathrm{e})^k \sqrt{2\pi k}\Big)^2} \equiv \frac{2^{2k}}{\sqrt{\pi k}}.$$

But the we see that

$$\binom{2k}{k} p^k q^k \sim \frac{\big(4p(1-p)\big)^k}{\sqrt{\pi k}}.$$

Now $4p(1-p) < 1$ if and only if $p \neq \frac{1}{2}$. So if $p \neq \frac{1}{2}$ we find that

$$\sum_{n=0}^{\infty} p_{ii}^{[n]} < \infty,$$

while

$$\sum_{n=0}^{\infty} p_{ii}^{[n]} = \infty, \qquad \text{when } p = \frac{1}{2}.$$

We conclude that the Markov chain is transient if $p \neq \frac{1}{2}$, and is recurrent if $p = \frac{1}{2}$. □

Starting in any state i, how long does it take to return to i? In order to answer this—rather natural—question, we define the *first return time to state* i, by

$$T_i = \min\{n \geq 1;\, X_n = i\},$$

with the convention that $T_i = \infty$ if $\{n \geq 1;\, X_n = i\} = \emptyset$, i.e., if this this visit never happens. We have the following definition.

Definition 3.1 The *mean recurrence time* m_i of a state i is defined as

$$m_i = \mathrm{E}(T_i \,|\, X_0 = i) = \begin{cases} \sum_{n=1}^{\infty} n f_{ii}^{[n]} & \text{when } i \text{ is recurrent} \\ \infty & \text{when } i \text{ is transient.} \end{cases}$$

It is important to note that m_i may be infinite, even if i is recurrent. In fact the simple random walk with $p = \frac{1}{2}$ is an example of this (but this is not so easy to show; in fact it is a consequence of the main theorem in Chapter 4; see Exercise 4.11). We have the following definition.

Definition 3.2 The *recurrent state* i is called *null-recurrent* if $m_i = \infty$, and *positive recurrent* (or *non-null recurrent*) if $m_i < \infty$.

In a finite irreducible Markov chain all states are positive recurrent, see also Exercise 3.14.

3.4 Solutions to the quick exercises

3.1 In this example $S = \{0, 1, 2, 3, 4, 5\}$, and the matrix of transition probabilities P is given by

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{1}{5} & 0 & \frac{4}{5} & 0 & 0 & 0 \\ 0 & \frac{2}{5} & 0 & \frac{3}{5} & 0 & 0 \\ 0 & 0 & \frac{3}{5} & 0 & \frac{2}{5} & 0 \\ 0 & 0 & 0 & \frac{4}{5} & 0 & \frac{1}{5} \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Clearly, all the rows of this matrix sum up to 1.

3.2 Due to the law of total probability we have

$$\mathrm{P}(X_1 = j) = \sum_{i=0}^{5} \mathrm{P}(X_1 = j \,|\, X_0 = i)\mathrm{P}(X_0 = i)\,,$$

Now plugging-in the transition probabilities we have found in Quick exercise 3.1 the desired result follows.

We prove the general statement by induction. We have just shown that the statement holds for $n = 1$. Next suppose that for some $n \geq 2$ one has that $\mathrm{P}(X_{n-1} = j) = \binom{N}{j} 2^{-N}$ for $j \in S$. We need to show that $\mathrm{P}(X_n = j) = \binom{N}{j} 2^{-N}$ for $j \in S$. We have,

$$\mathrm{P}(X_n = j) = \sum_{i=0}^{N} \mathrm{P}(X_n = j \mid X_{n-1} = i)\mathrm{P}(X_{n-1} = i)\,.$$

Because the Markov chain is time homogeneous, this is essentially checking the same thing as in the case $n = 1$, i.e., the result holds for all $n \geq 1$.

The initial distribution "tells" us how we started, so if $\mu_0 = 1$ (and $\mu_i = 0$ for $i = 1, 2, \ldots, N$), then initially all molecules were put in A). If we randomly put each molecule either in A, or in B, then μ_i "automatically" is equal to $\binom{N}{i} 2^{-N}$ for $i \in S$.

3.3 Starting in state i, one cannot return to i in an odd number of steps, i.e., $p_{ii}^{[n]} = 0$ whenever n is odd. Furthermore, one can only return to i is the number of 'steps to the left' is equal to the number of 'steps to the right'. So if $n = 2k$, k steps must be to the left, and k to the right. There are $\binom{2k}{k}$ ways to do this, each having probability $p^k q^k$.

3.5 Exercises

3.1 Let A, B, and C be three events. Show that the statements

$$\mathrm{P}(A \mid B \cap C) = \mathrm{P}(A \mid B)$$

and

$$\mathrm{P}(A \cap C \mid B)\mathrm{P}(A \mid B) \cdot \mathrm{P}(C \mid B)$$

are equivalent.

3.2 Suppose that the weather of tomorrow only depends on the weather conditions of today. If it rains today, it will rain tomorrow with probability 0.7. If there was no rain today, there will be no rain tomorrow with probability 0.7. Define for $n \geq 0$ the stochastic process (X_n) by

$$X_n = \begin{cases} 1 & \text{no rain on the } n\text{th day} \\ 0 & \text{rain on the } n\text{th day.} \end{cases}$$

a. What is the state space S, and what are the transition probabilities p_{ij}?
b. If today it rained, what is the probability it will rain the day after tomorrow? After three days?

3.3 Clearly the weather model in Exercise 3.2 is not very realistic. Suppose the weather of today depends on the weather conditions of the previous two days. Suppose that it will rain today with probability 0.7, if it is *given* that it rained the previous two days. If it rained only yesterday, but not the day before yesterday, it will rain today with probability 0.5. If it rained two days ago, but not yesterday, then the probability that it will rain today is 0.4. Finally, if the past two days were without rain, it will rain today with probability 0.2.

a. 'Translate' the statement '*it will rain today with probability* 0.7*, if it rained the previous two days*' in a statement involving probabilities of the process $(X_n)_{n\geq 0}$.
b. Why is the process $(X_n)_{n\geq 0}$ not a Markov chain?
c. Define

$$Y_n = \begin{cases} 0 & \text{if } X_{n-1} = 1 \text{ and } X_n = 1 \\ 1 & \text{if } X_{n-1} = 0 \text{ and } X_n = 1 \\ 2 & \text{if } X_{n-1} = 1 \text{ and } X_n = 0 \\ 3 & \text{if } X_{n-1} = 0 \text{ and } X_n = 0. \end{cases}$$

Show that $(Y_n)_{n\geq 0}$ is a Markov chain. What is the state space, and what are the transition probabilities?

3.4 You repeatedly throw a fair die. Let X_n be the outcome of the nth throw, and let M_n be the maximum of the first n throws:

$$M_n = \max\{X_1, \dots , X_n\},$$

(so $M_n = X_{(n)}$). You may assume that the random variables X_n are independent, and discrete uniformly distributed on $S = \{1, 2, 3, 4, 5, 6\}$.

a. Show that the stochastic process $(M_n)_{n\geq 1}$ is a Markov chain with state space S.
b. Find the matix P of transition probabilities of the Markov chain $(M_n)_{n\geq 1}$, and classify the states.
c. Let T be the first time a 6 has appeared:

$$T = \begin{cases} \min\{n;\, M_n = 6\} & \text{in case } \{n : M_n = 6\} \neq \emptyset \\ \infty & \text{in case } \{n : M_n = 6\} = \emptyset \end{cases}$$

Determine the probability distribution of T.

3.5 Supply a proof of Corollaries 3.1 and 3.2.

3.6 Consider the Markov chain $(X_n)_{n\geq 0}$, with state space $S = \{0, 1\}$, and with matrix of transition probabilities P, given by

$$P = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

a. Classify the states. Is the chain irreducible?
b. Determine $f_{00}^{[n]}$ and f_0. Does this answer surprise you? (or not?). Why?
c. Let the initial distribution μ be given by

$$\mu = \begin{pmatrix} 3/7 & 4/7 \end{pmatrix}.$$

Determine $\mathrm{P}(X_n = 0)$ for all $n \geq 1$. What do you notice?

3.7 In the previous exercise, Exercise 3.6, we saw that $\mathrm{P}(X_n) = \mu_i$ for $i \in S$, if $\mu = \begin{pmatrix} 3/7 & 4/7 \end{pmatrix}$. It is quite natural to wonder what happens if μ is some arbitrary initial distribution. I.e., what happens if

$$\mu = \begin{pmatrix} \mu_0 & \mu_1 \end{pmatrix},$$

with $\mu_0 \geq 0$, $\mu_1 \geq 0$, and $\mu_0 + \mu_1 = 1$?

In the next chapter we will investigate this, but here we will use our simple set-up to 'get a feel' of what's going on. In view of Corollary 3.2 we are interested in powers P^n of the matrix of transition probabilities P.

a. Find the determinant, the eigenvalues, and eigenvectors of P.
b. Let T be the matrix, whose columns are the eigenvectors of P, and let D be a diagonal matrix, with the corresponding eigenvalues in the main diagonal. Argue that $P = TDT^{-1}$. Use this, to find

$$P^{\infty} = \lim_{n \to \infty} P^n.$$

c. Let μ be an initial distribution. Use Corollary 3.2 and your results from **b.** to show that

$$\lim_{n \to \infty} \mathrm{P}(X_n = 0) = \frac{3}{7}.$$

3.8 Let $(Y_n)_{n \geq 0}$ be a sequence of independent random variables, all with the same distribution, given by

$$\mathrm{P}(Y_n = 0) = \frac{2}{3}, \qquad \mathrm{P}(Y_n = 1) = \frac{1}{6}, \qquad \mathrm{P}(Y_n = 2) = \frac{1}{6}$$

Consider the stochastic process $(X_n)_{n \geq 0}$, defined by $X_0 = 0$, and

$$X_{n+1} = \begin{cases} X_n - Y_n & \text{if } X_n = 4 \\ X_n - 1 + Y_n & \text{if } 1 \leq X_n \leq 3 \\ Y_n & \text{if } X_n = 0, \end{cases}$$

for $n \geq 0$.

a. Determine the matrix of transition probabilities P.
b. Determine $p_{34}^{[2]}$, the probability to go from state 3 to state 4 in two steps.

3.9 Consider the Markov chain $(X_n)_{n\geq 0}$, with state space $S = \{1,2,3,4\}$, and with matrix of transition probabilities P. Classify the states, determine f_i for $i \in S$, and find out whether the chain is irreducible in the following two cases:

a. The matrix P of transition probabilities is first given by:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix}.$$

b. Next, P is given by:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}.$$

3.10 Show that for every $n \geq 1$ one has, that

$$p_{ii}^{[n]} = \sum_{k=1}^{n} p_{ii}^{[n-k]} f_{ii}^{[k]}.$$

3.11 In this exercise we show that for $k \geq 1$,

$$\mathrm{P}(N_i = k+1 \,|\, X_0 = i) = f_i \mathrm{P}(N_i = k \,|\, X_0 = i),$$

see also (3.3.3).

a. First show that $\mathrm{P}(N_i = k+1 \,|\, X_0 = i)$ is equal to

$$\sum_{m=1}^{\infty} \mathrm{P}(X_\ell \neq i,\, \ell = 1, \ldots, m-1, X_m = i, \sum_{n=\ell}^{\infty} 1_{\{X_\ell = i\}} = k \,|\, X_0 = i).$$

b. Setting $A_m = \{X_j \neq i \text{ for } j = 1, 2, \ldots, m-1, X_0 = i\}$, and

$$B_m = \left\{ \sum_{n=0}^{\infty} 1_{\{X_{n+m} = i\}} = k \right\},$$

show that the expression in **a.** is equal to $\mathrm{P}(X_m = i, A_m \cap B_m \,|\, X_0 = i)$.

c. Use (3.3.2) to show that $\mathrm{P}(X_m = i, A_m \cap B_m \,|\, X_0 = i)$ is equal to

$$\mathrm{P}(A_m \,|\, X_m = i)\mathrm{P}(B_m \,|\, X_M = i)\frac{\mathrm{P}(X_m = i)}{\mathrm{P}(X_0 = i)}.$$

d. Next show that

$$\mathrm{P}(A_m \,|\, X_m = i)\frac{\mathrm{P}(X_m = i)}{\mathrm{P}(X_0 = i)} = f_{ii}^{[m]},$$

and derive the desired result.

3.12 Let i and j be two communicating states, i.e., $i \leftrightarrow j$. Then i is recurrent if and only if j is recurrent.

3.13 Let i be a recurrent state, and suppose moreover that $i \to j$. Then $i \leftrightarrow j$. (Due to Exercise 3.12 it follows that j is also recurrent.)

3.14 Let (X_n) be an irreducible Markov chain on a finite state space S. Then $m_i < \infty$.

4

Limit behavior of discrete Markov chains

In this chapter the long-term behavior of a discrete Markov chain will be investigated. Already in Exercise 3.7 we have seen that if the Markov chain is "nice," something remarkable is happening; the matrix P^n of n-step transition probabilities converges to a matrix in which the values per column are identical. In fact each row of this limiting matrix is equal to the stationary distribution. In Section 4.2 this will be further investigated. However, we will start with a process where the Markov chain does not have the "nice" properties of the chain in Exercise 3.7, and where we still can say a lot about the long term behavior of the chain.

4.1 Branching processes

It is well know that the study of Probability Theory has its roots in gambling. Another major reason to study Probability Theory is to understand the evolution of systems. This is in general quite a difficult problem, and we will address it with a rather simple model; Branching processes. These branching processes were first introduced by Galton in 1889 to understand how family names become extinct, but they can also be used to model the growth (or decline) of a group of bacteria, or neutrons hitting atoms and in this way releasing more neutrons, thus (possibly) starting a chain reaction. In this model we have the following assumption: for $n \geq 0$, the nth generation has X_n "individuals," and at the end of the nth generation each "individual" has Z offsprings (and dies itself), independently from what happened in previous generations, and independently from the number of offsprings of the other members of the nth generation. Here Z is a discrete random variable, with probability mass function given by

$$p_j = \mathrm{P}(Z = j)\,, \qquad j = 0, 1, 2, \ldots.$$

We will assume that for every j one has that $p_j < 1$ (otherwise the whole process becomes "deterministic;" with probability one every individual will

have exactly j offsprings), and also that $p_0 > 0$ (otherwise it is trivial what will happen in the long-run; the number of offsprings will eventually pass any given limit). Finally, we will assume that $X_0 = 1$; this is not really necessary, but makes the analysis more transparent.
Setting $Z_i^{(n-1)}$ as the number of offspring of the ith member of the $n-1$th generation, we thus have that

$$\begin{cases} X_0 = 1 \\ X_n = Z_1^{(n-1)} + Z_2^{(n-1)} + \cdots + Z_{X_{n-1}}^{(n-1)}, \end{cases}$$

where for each n the random variables $Z_1^{(n-1)}, Z_2^{(n-1)}, \ldots$ are independent and identically distributed (with the same distribution as Z), and for different n these sequences of random variables are independent.
Clearly every X_n has its values in $S = \mathbb{N}$, and that the number of individuals X_{n+1} in the $n+1$st generation *only* depends on the number of individuals X_n in the nth generation. I.e., the stochastic process $(X_n)_{n\geq 0}$ is a Markov chain, with transition probabilities p_{ij} given by

$$p_{ij} = \mathrm{P}(X_{n+1} = j \,|\, X_n = i) = \mathrm{P}\Big(Z_1^{(n)} + Z_2^{(n)} + \cdots + Z_i^{(n)} = j\Big)\,. \qquad (4.4.1)$$

Quick exercise 4.1 Show that the assumption that $p_0 > 0$ implies that every state k different from 0 is transient (*Hint:* determine p_{k0}).

From Quick exercise 4.1 we conclude that for every $k \geq 1$ the set $\{1, 2, \ldots, k\}$ will be visited only of finite number of times. So if $p_0 > 0$ we either have (with probability one) that the population will become extinct or that eventually the population size will grow beyond any given bound. In order to see which of these two possible cases will happen we will first determine the expected number of individual in the nth generation. Setting $\mu = \mathrm{E}[Z]$, i.e.,

$$\mu = \sum_{j=0}^{\infty} j\mathrm{P}(Z = j) \left(= \sum_{j=0}^{\infty} jp_j \right),$$

we see that the expected number of individuals $\mathrm{E}[X_1]$ in the first generation is μ. In general we have the following result.

Theorem 4.1 For $n \geq 0$ we have, that

$$\mathrm{E}[X_n] = \mu^n.$$

Proof. Because $X_0 = 1$, the statement in the theorem is correct for $n = 0$, and we just have seen that it is also correct for $n = 1$. We may therefore assume that $n \geq 2$. Using (1.1.2), we find that

$$\mathrm{E}[X_n] = \mathrm{E}[\mathrm{E}(X_n \,|\, X_{n-1})] = \sum_{i=0}^{\infty} \mathrm{E}(X_n \,|\, X_{n-1} = i)\mathrm{P}(X_{n-1} = i)\,.$$

Since $\mathrm{E}(X_n \mid X_{n-1} = i) = \mathrm{E}\Big[Z_1^{(n-1)} + Z_2^{(n-1)} \cdots + Z_i^{(n-1)}\Big] = i\mu$, we find that

$$\begin{aligned}
\mathrm{E}[X_n] &= \sum_{i=0}^{\infty} \mu i \mathrm{P}(X_{n-1} = i) \\
&= \mu \sum_{i=0}^{\infty} i \mathrm{P}(X_{n-1} = i) \qquad \text{(by definition of } \mathrm{E}[X_{n-1}]) \\
&= \mu \mathrm{E}[X_{n-1}]\,.
\end{aligned}$$

Since $\mathrm{E}[X_0] = 1$, we find by iteration the desired result: $\mathrm{E}[X_n] = \mu^n$. □

Trivially, the $n+1$st generation does not have any individual, when the nth generation does not have any individuals, i.e., $\{X_n = 0\} \subset \{X_{n+1} = 0\}$, implying that for $n \geq 1$, $\mathrm{P}(X_{n+1} = 0) \geq \mathrm{P}(X_n = 0)$. I.e., the sequence $(c_n)_{n\geq 0}$ where $c_n = \mathrm{P}(X_n = 0)$, is a monotonically non-decreasing sequence of probabilities, and is therefore bounded by 1. But then the limit exists as n tends to infinity; say this limit is π_0, then

$$\pi_0 = \lim_{n\to\infty} \mathrm{P}(X_n = 0)\,.$$

So π_0 is the probability of ultimate extinction. We have the following two cases: $\mu \leq 1$, and $\mu > 1$.

If $\mu < 1$, we have that $\pi_0 = 1$. To see this, note that

$$\begin{aligned}
\mu^n = \mathrm{E}[X_n] &= \sum_{j=0}^{\infty} j \mathrm{P}(X_n = j) \\
&\geq \sum_{j=1}^{\infty} \mathrm{P}(X_n = j) = \mathrm{P}(X_n \geq 1) = 1 - \mathrm{P}(X_n = 0)\,,
\end{aligned}$$

yielding that

$$1 - \mu^n \leq \mathrm{P}(X_n = 0) \leq 1,$$

from which it follows that $\pi_0 = 1$. Also in case $\mu = 1$ one can show that $\pi_0 = 1$. However, in case $\mu > 1$ we have that $\pi_0 < 1$.

Theorem 4.2 The probability of ultimate extinction π_0 is the smallest positive solution of the equation

$$x = \sum_{j=0}^{\infty} p_j x^j. \tag{4.4.2}$$

Proof. Note that the right-hand side of (4.4.2) is the probability generating function of Z;

$$G_Z(s) = \sum_{j=0}^{\infty} p_j s^j.$$

The consequences of this are two-fold. First, $x = 1$ is a solution of (4.4.2). Furthermore, $\mu = \mathrm{E}[Z] = G'_Z(1)$, so μ is the slope of the tangent of $G_Z(s)$ in $s = 1$. Because it is assumed that $p_0 > 0$, it follows in case $\mu > 1$ from Figure 4.1 that (4.4.2) must have another solution x between 0 and 1. Also note that

$$0 < G'_Z(x) < G'_Z(1), \qquad \text{for } x \text{ between 0 and 1},$$

so $x = 1$ is the smallest solution of (4.4.2) in case $\mu \leq 1$; the curve $y = G_Z(x)$ does not cross $y = x$ for x between 0 and 1.

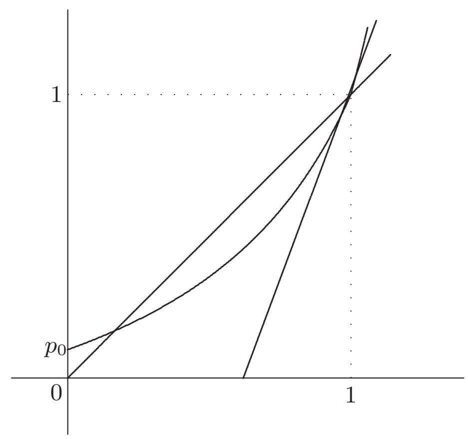

Fig. 4.1. The tangent of $y = G_Z(x)$ in $x = 1$, in case $\mu > 1$.

By definition of π_0,

$$\begin{aligned}\pi_0 &= \lim_{n\to\infty} \mathrm{P}(X_n = 0)\\ &= \lim_{n\to\infty} \sum_{j=0}^{\infty} \mathrm{P}(X_n = 0, X_1 = j)\\ &= \lim_{n\to\infty} \sum_{j=0}^{\infty} \mathrm{P}(X_n = 0 \,|\, X_1 = j)\mathrm{P}(X_1 = j)\\ &= \lim_{n\to\infty} \sum_{j=0}^{\infty} p_j \mathrm{P}(X_n = 0 \,|\, X_1 = j).\end{aligned}$$

What can one say about $\mathrm{P}(X_n = 0 \,|\, X_1 = j)$? Note this is the probability of extinction at time n, given we started j independent trees at time 1. Each of these trees is extinct at time n with probability $\mathrm{P}(X_{n-1} = 0)$, so—due to the independence of these j trees—we find

$$P(X_n = 0 \mid X_1 = j) = P(X_{n-1} = 0)^j .$$

But then it follows that

$$\pi_0 = \sum_{j=0}^{\infty} p_j \pi^j ,$$

and we find that π_0 is indeed a solution of (4.4.2).

Next, let ξ be a non-negative solution of (4.4.2). We are done if we show that $\xi \geq \pi_0$. Since $G_Z(0) = p_0 > 0$, it follows from $X_0 = 1$ that

$$\xi > 0 = P(X_0 = 0) .$$

Now suppose that $\xi \geq P(X_{n-1} = 0)$ for some $n \geq 1$. We want to show that $\xi \geq P(X_n = 0)$, so that by induction for all $n \geq 1$ we have that $\xi \geq P(X_n = 0)$, implying that $\xi \geq \lim_{n\to\infty} P(X_n = 0) = \pi_0$. Recycling what we derived earlier in this proof, we find

$$\begin{aligned} P(X_n = 0) &= \sum_{j=0}^{\infty} P(X_n = 0, X - 1 = j) \\ &= \sum_{j=0}^{\infty} p_j P(X_n = 0 \mid X_1 = j) \\ &= \sum_{j=0}^{\infty} p_j \left(P(X_{n-1} = 0)\right)^j \qquad \text{(by induction hypothesis)} \\ &\leq \sum_{j=0}^{\infty} p_j \xi^j = \xi. \end{aligned}$$

Thus we find that π_0 is the smallest positive solution of (4.4.2). □

4.2 Asymptotic behavior

Some examples

In Section 3.3 we introduced some notions that describe the long-time behavior of a Markov chain. If $(X_n)_{n\geq 0}$ is a irreducible Markov chain on a finite state space S, then every state must be recurrent; see Exercise 3.14. However, if the state space if infinite, we need to "sharpen" the definition of recurrence; as in case of the symmetric random walk, each state is recurrent, but the expected return time might be infinite. One could wonder whether there are other properties important for the long-term behavior of the Markov chain? And what happens to P^n when the chain is reducible? Consider the following examples.

Example 4.1

Let $(X_n)_{n\geq 0}$ be a Markov chain on the finite state space $S = \{1,2,3,4\}$, and let the matrix of transition probabilities be given by

$$P = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix};$$

see also Figure 4.2.

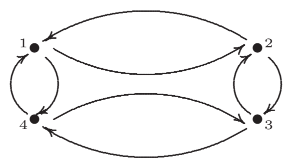

Fig. 4.2. The transitions in Example 4.1 (all transitions are with probability 1/2)

From Figure 4.2 we see that the Markov chain is irreducible, something that also follows from the following Quick exercise. But for every $i \in S$ we have that $f_{ii}^{[n]} = 0$ if n is odd, while $f_{ii}^{[n]} = \frac{1}{2}$ if n is even; the states are *periodic*.

Quick exercise 4.2 Show that for the Markov chain in Example 4.1 we have that

$$P^n = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix} \quad \text{if } n \text{ is odd,}$$

and

$$P^n = \begin{pmatrix} 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix} \quad \text{if } n \geq 2 \text{ is even.}$$

In the following example we will see, that a small change in the values of P has dramatic consequences for the values of P^n, with $n \geq 2$.

Example 4.2

Again, let $(X_n)_{n\geq 0}$ be a Markov chain on the finite state space $S = \{1,2,3,4\}$, but let the matrix of transition probabilities P now be given by

$$P = \begin{pmatrix} 1/3 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix};$$

see also Figure 4.3.

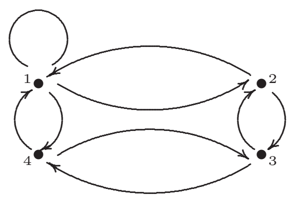

Fig. 4.3. The transitions in Example 4.2

From Figure 4.3 we see that the Markov chain is irreducible. However, using MAPLE (or any other software-package) we see that P^n converges to a matrix with constant values per column (the values in the matrices are rounded off);

$$P^2 = \begin{pmatrix} 4/9 & 1/9 & 1/3 & 1/9 \\ 1/6 & 5/12 & 0 & 5/12 \\ 1/2 & 0 & 1/2 & 0 \\ 1/6 & 5/12 & 0 & 5/12 \end{pmatrix}, \quad P^5 = \begin{pmatrix} 67/243 & 283/972 & 23/162 & 283/972 \\ 283/648 & 8/81 & 79/216 & 8/81 \\ 23/108 & 79/216 & 1/18 & 79/216 \\ 283/648 & 8/81 & 79/216 & 8/81 \end{pmatrix},$$

$$P^{11} = \begin{pmatrix} .3099029337 & .2501914018 & .1897142627 & .2501914018 \\ .3752871027 & .1721414630 & .2804299713 & .1721414630 \\ .2845713941 & .2804299713 & .1545686633 & .2804299713 \\ .3752871027 & .1721414630 & .2804299713 & .1721414630 \end{pmatrix},$$

and

$$P^{100} = \begin{pmatrix} .3333333694 & .2222221792 & .2222222723 & .2222221792 \\ .3333332687 & .2222222993 & .2222221326 & .2222222993 \\ .3333334084 & .2222221326 & .2222223264 & .2222221326 \\ .3333332687 & .2222222993 & .2222221326 & .2222222993 \end{pmatrix}.$$

Of course, things get really interesting when n becomes really big. For instance, if $n = 1000$, we find

$$P^{1000} = \begin{pmatrix} .3333333333 & .2222222222 & .2222222222 & .2222222222 \\ .3333333333 & .2222222222 & .2222222222 & .2222222222 \\ .3333333333 & .2222222222 & .2222222222 & .2222222222 \\ .3333333333 & .2222222222 & .2222222222 & .2222222222 \end{pmatrix}.$$

This suggests (but certainly does not prove!) that

$$\lim_{n\to\infty} P^n = \begin{pmatrix} 1/3 & 2/9 & 2/9 & 2/9 \\ 1/3 & 2/9 & 2/9 & 2/9 \\ 1/3 & 2/9 & 2/9 & 2/9 \\ 1/3 & 2/9 & 2/9 & 2/9 \end{pmatrix}. \tag{4.4.3}$$

In fact, following Exercise 3.7 it is in fact easy to show that (4.4.3) holds; see also Exercise 4.5.

Quick exercise 4.3 Let μ be the vector, given by

$$\mu = \begin{pmatrix} 1/3 & 2/9 & 2/9 & 2/9 \end{pmatrix},$$

i.e., μ is any of the rows of the matrix in (4.4.3). Show that $\mu P = \mu$. □

In Examples 3.1 (b) and (c) the Markov chain was reducible. In Example 3.1 (b) there are clearly two "subworlds"; these are the equivalence-classes $\{1,4\}$ and $\{2,3\}$. Again using MAPLE one sees that

$$P^n \approx \begin{pmatrix} 1/4 & 0 & 0 & 3/4 \\ 0 & 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 & 0 \\ 1/4 & 0 & 0 & 3/4 \end{pmatrix} \qquad \text{as } n \to \infty.$$

So on the equivalence-class $\{1,4\}$ the Markov chain $(X_n)_{n\geq 0}$ behaves as a chain with only two states 1 and 4, with matrix of transition probabilities

$$P_{\{1,4\}} = \begin{pmatrix} 2/3 & 1/3 \\ 1/9 & 8/9 \end{pmatrix} \quad \to \quad \begin{pmatrix} 1/4 & 3/4 \\ 1/4 & 3/4 \end{pmatrix} \quad \text{as } n \to \infty,$$

while the chain behaves on $\{2,3\}$ as a chain with matrix of transition probabilities

$$P_{\{2,3\}} = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}.$$

Note that $P^n_{\{2,3\}} = P_{\{2,3\}}$ for every $n \geq 1$.

In Example 3.1 (c) we have seen that state 1 behaves as a "sink"; everything is eventually sucked into it. Using MAPLE this also becomes apparent. For example, for $n = 100$ one finds that

$$P^n \approx \begin{pmatrix} 1 & 0 & 0 & 0 \\ .9999999903 & .482986938910^{-8} & .482986938910^{-8} & 0 \\ .9999999855 & .724480408310^{-8} & .724480408310^{e-8} & 0 \\ .9999923308 & 0 & 0 & .766915923710^{e-5} \end{pmatrix},$$

while for $n = 1000$,

$$P^n \approx \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & .263520219610^{-79} & .263520219610^{-79} & 0 \\ 1 & .395280329410^{-79} & .395280329410^{-79} & 0 \\ 1 & 0 & 0 & .703845847310^{-51} \end{pmatrix}.$$

Periodicity

In view of Example 4.1 we have the following definition.

Definition 4.1 Let i be a recurrent state. The *period* $d(i)$ of state i is the greatest common divisor of the set

$$T_i = \{n \geq 1;\, p_{ii}^{[n]} > 0\}.$$

A recurrent state with period 1 is called *aperiodic.*

At first this definition might be a little bit cryptic; in words: $d(i)$ is the greatest positive integer, that divides those positive integers n for which $p_{ii}^{[n]} > 0$. So in Example 4.1 we have that $T_1 = \{2, 4, 6, 8, \dots\}$, and therefore $d(1) = 2$. Note that in fact $d(i) = 2$ for all $i \in S$. This is no coincidence, as the following theorem shows that two recurrent states in the same equivalence-class have the same period.

Theorem 4.2 Let i and j be two recurrent states with periods $d(i)$ respectively $d(j)$. Furthermore, let $i \leftrightarrow j$. Then $d(i) = d(j)$.

Proof. Because $i \leftrightarrow j$, there exist positive integers ℓ and m, such that

$$p_{ij}^{[\ell]} > 0 \quad \text{and} \quad p_{ji}^{[m]} > 0.$$

Because i is recurrent, there exists a positive integer n such that $p_{ii}^{[n]} > 0$. Due to Theorem 3.2 ('Chapman-Kolmogorov') we have, that

$$p_{ii}^{[2n]} \geq p_{ii}^{[n]} p_{ii}^{[n]} > 0,$$

and

$$p_{jj}^{[\ell+m+n]} \geq p_{ji}^{[m]} p_{ii}^{[n]} p_{ij}^{[\ell]} > 0 \quad \text{and} \quad p_{jj}^{[\ell+m+2n]} \geq p_{ji}^{[m]} p_{ii}^{[2n]} p_{ij}^{[\ell]} > 0.$$

By definition, $d(j)$ divides both $\ell + m + n$ and $\ell + m + 2n$. But then we have that $d(j)$ also divides the difference of these two positive integers; $d(j)$ divides n. Since $d(i)$ is the *greatest* common divisor of all positive integers n for which $p_{ii}^{[n]} > 0$, we must have that $d(j) \leq d(i)$. Exchanging the role of i and j in the preceding discussion yields that $d(i) \leq d(j)$, and we find that $d(i) = d(j)$. □

In view of this theorem we can therefore speak of the period of an equivalence class (of communicating states), or even of the period of a Markov chain, if this chain is irreducible.

Theorem 4.3 Let P be an irreducible stochastic matrix with period d, then there exist non-negative integers m, n_0, such that for all $n \geq n_0$,

$$p_{ij}^{[m+nd]} > 0.$$

Remark. If $d = 1$ and P is a finite matrix (i.e., an $s \times s$-matrix for some $s \in \mathbb{N}$), then it follows from Theorem 4.3 that there exists a non-negative integer n_0 such that all the entries in P^n for $n \geq n_0$, i.e., $p_{ij}^{[n]} > 0$ for $n \geq n_0$.

Proof. It follows from $i \leftrightarrow j$, that there exist positive integers $k, \ell \leq s$ such that $p_{ij}^{[k]} > 0$ and $p_{ji}^{[\ell]} > 0$. Note that it now suffices—due to Chapman-Kolmogorov—to show that there exists a non-negative integer n_0 such that $p_{jj}^{[nd]} >$ for all $n \geq n_0$. Since the period is d, we know that the gcd of the set $T_j = \{n \geq 1;\, p_{jj}^{[n]} > 0\}$ is d. Again due to Chapman-Kolmogorov the set T_j is closed under addition; if $n_1, n_2 \in T_j$, then $n_1 + n_2 \in T_j$:

$$p_{jj}^{[n_1+n_2]} = \sum_{s \in S} p_{js}^{[n_1]} p_{sj}^{[n_2]} \geq p_{jj}^{[n_1]} p_{jj}^{[n_2]} > 0.$$

The desired result now follows from the following lemma from Number Theory; see [3], section 2.4 and appendix A.

> **Lemma** Let d be the gcd of the set of positive integers $A = \{a_n;\, n \in \mathbb{N}\}$, and suppose that A is closed under addition (i.e., $a_n + a_m \in A$ for all $n, m \in \mathbb{N}$). Then there exists a positive integer n_0 such that $nd \in A$ for all $n \geq n_0$.

□

The main theorem

In the Ehrenfest gas-of-molecules-example in Section 3.1 we saw that the vector $\mu = (\, \mu_0 \ \mu_1 \ \ldots \ \mu_N \,)$, with

$$\mu_i = \binom{N}{i} 2^{-N}, \qquad i = 0, 1, \ldots, N,$$

is a *stationary distribution*; if $\mathrm{P}(X_0 = i) = \mu_i$, then $\mathrm{P}(X_n = j) = \mu_j$, for $i, j = 0, 1, \ldots, N$.
Here is the general definition.

> **Stationary distribution** Let $(X_n)_{n\geq 0}$ be a Markov chain on a finite or countable state space S, and let P be the matrix of transition probabilities of $(X_n)_{n\geq 0}$. Then the vector $\mu = (\, \mu_i \,)$, with
>
> $$\mu_i \geq 0 \ \text{ for } i \in S, \qquad \text{and} \quad \sum_{i \in S} \mu_i = 1,$$
>
> is a *stationary distribution* of the Markov chain if for all $i, j \in S$, and all $n \in \mathbb{N}$,
>
> $$\mathrm{P}(X_0 = i) = \mu_i \text{ implies that } \mathrm{P}(X_n = j) = \mu_j,$$
>
> In case there is only one stationary distribution, it is usually denoted by π.

It follows from Corollary 3.2 that μ is a stationary distribution if and only if $\mu = \mu P$.

In the following Quick exercise you are invited to show that there might be more than one stationary distribution.

Quick exercise 4.4 Let N be some positive integer, and let $(X_n)_{n\geq 0}$ be a Markov chain on the state space $S = \{1, 2, \dots, N\}$, satisfying $p_{ii} = 1$ for each state $i \in S$. Show that *any* distribution on the state space S is a stationary distribution.

In Examples 4.1 and 4.2 we have seen that the long term behavior of a Markov chain depends on the (ir)reducibility and (a)periodicity the chain. Clearly, if the chain is reducible, it will fall apart in "subworlds," which can be studied separately, and if the chain is periodic the distribution will not settle down for lange n. On the other hand, we have seen in various example that if the chain is "nice" (say irreducible and aperiodic), that the distribution of X_n for n large "stabilizes."

We have the following theorem, which we mention here without proof. Various proofs exist in the literature; see for example the excellent books [3], [6], [7], and [8].

> **Main theorem** Let $(X_n)_{n\geq 0}$ be an irreducible Markov chain on a finite or countable state space S. Furthermore, let $P = (p_{ij})$ be its matrix of transition probabilities. Then the chain has a stationary distribution π if and only if all the states are non-null recurrent. In case all the states are non-null recurrent, π is the unique stationary distribution, and satisfies
>
> $$\pi_i = \frac{1}{m_i}, \qquad \text{for each } i \in S,$$
>
> where m_i is the mean recurrence time of state i (c.f. Definition 3.1). Finally, if the chain is aperiodic (and non-null recurrent), then
>
> $$p_{ij}^{[n]} \to \pi_j \qquad \text{as } n \to \infty, \text{ for all } i, j \in S.$$

Remarks. (i) If the state space S is finite and the chain is irreducible, automatically all states are non-null recurrent; see Exercise 3.14.

(ii) In an irreducible aperiodic Markov chain in which all states are non-null recurrent, the limit $\lim_{n\to\infty} p_{ij}^{[n]}$ does not depend on the starting point $X_0 = i$; the "chain forgets its origin." By Corollary 3.2,

$$\mathrm{P}(X_n = j) \to \pi_j \qquad \text{as } n \to \infty,$$

irrespective of the initial distribution.

In fact a more general result holds, which is known as the *ergodic theorem*.

Ergodic theorem Let $(X_n)_{n\geq 0}$ be an irreducible Markov chain on a finite state space S. Let $P = (p_{ij})$ be its transition matrix, and π its unique stationary distribution. Furthermore, let $f : S \to \mathbb{R}$ be a function on S. Then with probability 1 we have that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \sum_{i\in S} f(i)\pi_i.$$

In fact, this theorem can be formulated—with some constraints on f—for states spaces S which are countable, but not finite; see e.g. [8]. The ergodic theorem—in fact a generalization of the law of large numbers—is often interpreted as: "time mean = space mean." The ergodic theorem has its roots in theoretical physics, but is nowadays widely used in many fields in mathematics, such as number theory; see e.g. [4].

(iii) In fact,when the state space S is finite, a proof of the main theorem can be obtained using linear algebra, essentially following the lines of Exercises 3.7 and 4.5. In these exercises the key point is, that the largest eigenvalue of P is 1, and that the other eigenvalues are smaller than 1 (in absolute value). Then P can be written as $P = TDT^{-1}$, where T is a matrix whose columns are the eigenvectors of P, and D is a diagonal matrix with the eigenvalues of P in its diagonal. Consequently, $P^n = TD^nT^{-1}$, and we find the desired result, since D^n is a diagonal matrix with the nth powers of the eigenvalues of P in its diagonal. In order to see that this approach works in general for a transition matrix P of a finite irreducible Markov chain of period d, the famous theorem of Perron-Frobenius comes to our aid; see section 6.6 in [6] where this theorem is stated, and a more detailed proof of this approach is given.

(iv) The main theorem implicitly contains an algorithm how to determine the stationary distribution π; one can find π from

$$\pi P = \pi,$$

and the fact that

$$\sum_{i\in S} \pi_i = 1.$$

As an example, consider the Markov chain from Quick exercise 3.1. In this example (Ehrenfest's model for $N = 5$), we have that

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ \frac{1}{5} & 0 & \frac{4}{5} & 0 & 0 & 0 \\ 0 & \frac{2}{5} & 0 & \frac{3}{5} & 0 & 0 \\ 0 & 0 & \frac{3}{5} & 0 & \frac{2}{5} & 0 \\ 0 & 0 & 0 & \frac{4}{5} & 0 & \frac{1}{5} \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

Since $\pi = \pi P$, we find that

$$\pi_0 = \tfrac{1}{5}\pi_1,\ \pi_1 = \pi_0 + \tfrac{2}{5}\pi_2,\ \pi_2 = \tfrac{4}{5}\pi_1 + \tfrac{3}{5}\pi_3,$$
$$\pi_3 = \tfrac{3}{5}\pi_2 + \tfrac{4}{5}\pi_4,\ \pi_4 = \tfrac{2}{5}\pi_3 + \pi_5,\ \pi_5 = \tfrac{1}{5}\pi_4,$$

yielding that

$$\pi_1 = 5\pi_0,\ \pi_2 = 10\pi_0,\ \pi_3 = 10\pi_0,\ \pi_4 = 5\pi_0,\ \pi_5 = \pi_0,$$

and therefore,

$$\pi_0 + 5\pi_0 + 10\pi_0 + 10\pi_0 + 5\pi_0 + \pi_0 = 1 \quad \Leftrightarrow \quad \pi_0 = \tfrac{1}{32}.$$

We find that

$$\pi = \left(\ \tfrac{1}{32}\ \tfrac{5}{32}\ \tfrac{10}{32},\ \tfrac{10}{32}\ \tfrac{5}{32}\ \tfrac{1}{32}\ \right)$$

is the unique stationary distribution of the Markov chain in the Ehrenfest model (in case $N = 5$). Note that in Quick exercise 3.2 you have showed that the stationary distribution is $Bin(N, \frac{1}{2})$, where N is the number of molecules in the gas. It is a simple exercise to show that this is indeed the case here.

4.3 Solutions to the quick exercises

4.1 From (4.4.1) it follows that

$$\begin{aligned} p_{k0} &= \mathrm{P}\Big(Z_1^{(n)} + \cdots + Z_k^{(n)} = 0\Big) \\ &= \mathrm{P}\Big(Z_1^{(n)} = 0, Z_2^{(n)} = 0, \ldots, Z_k^{(n)} = 0\Big) \\ &= p_0^k > 0. \end{aligned}$$

4.2 The result immediately follows from the fact that

$$P^2 = \begin{pmatrix} 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{pmatrix}, \quad \text{and} \quad P^3 = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix} = P.$$

4.3

$$\begin{aligned} \mu P &= \left(\ 1/3\ \ 2/9\ \ 2/9\ \ 2/9\ \right) \begin{pmatrix} 1/2 & 1/3 & 0 & 1/3 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix} \\ &= \left(\ 1/3\ \ 2/9\ \ 2/9\ \ 2/9\ \right) \\ &= \mu. \end{aligned}$$

4.4 Let $\mu = (\ \mu_1\ \mu_2\ \ldots\ \mu_N\ )$ be any distribution on S. Because $p_{ii} = 1$ for each state $i \in S$ we have that the matrix of transition probabilities P is the identity matrix, sone trivially has that $\mu P = \mu$.

4.4 Exercises

4.1 Consider a branching process, where the distribution of the number of offsprings Z is given by

$$p_j = \mathrm{P}(Z = j)\,, \qquad j = 0, 1, 2, \dots\,.$$

Determine the probability of ultimate extinction in each of the following three cases:

a. $p_0 = \frac{1}{4}$; $p_2 = \frac{3}{4}$.
b. $p_0 = \frac{1}{4}$; $p_1 = \frac{1}{2}$; $p_2 = \frac{1}{4}$.
c. $p_0 = \frac{1}{6}$; $p_1 = \frac{1}{2}$; $p_2 = \frac{1}{3}$.

4.2 Suppose a branching process is given, where the distribution of the number of offsprings Z is given by

$$\mathrm{P}(Z = j) = bc^{j-1}, \qquad j = 1, 2, \dots\,,$$

with $0 < b \le 1 - c$.

a. Determine $\mathrm{P}(Z = 0)$.
b. Find the probability generating function of Z, and use this generating function to determine the expectation $\mu = \mathrm{E}[Z]$.
c. Determine for various values of μ the probability of ultimate extinction π_0.

4.3 Let $(X_n)_{n\ge 0}$ be a branching process, with $X_0 = 1$, and let the number of offspring of each individual be given by the random variable Z, with probability mass function given by

$$\mathrm{P}(Z = 0) = \mathrm{P}(Z = 1) = \frac{1}{4} \qquad \text{and} \;\; \mathrm{P}(Z = 2) = \frac{1}{2}.$$

a. Determine the conditional expectation of X_2, if it is known that $X_1 = 2$. Next, what s the conditional expectation of X_{10}, given that $X_1 = 2$?
b. Determine the probability of ultimate extinction if it is given that $X_1 = 2$.

4.4 Suppose a branching process is given, where the distribution of the number of offsprings Z is given by

$$\mathrm{P}(Z = 0) = 1 - 2p, \qquad \text{and} \;\; \mathrm{P}(Z = 1) = \mathrm{P}(Z = 2) = p,$$

where p is a parameter satisfying $0 < p < \frac{1}{2}$. Determine the probability of ultimate extinction π_0 for each p satisfying $0 < p < \frac{1}{2}$.

4.5 Find the determinant, the eigenvalues, and eigenvectors of the matrix P from Example 4.2. Use these to show that (4.4.3) holds.

4.6 Find the stationary distribution of the Markov chain $(X_n)_{n\geq 0}$ of the weather model in Exercise 3.2. What percentage of the time will it rain?

4.7 Find the stationary distribution of the Markov chain $(Y_n)_{n\geq 0}$ of the more elaborate weather model in Exercise 3.3. What is the percentage of the time it will rain?

4.8 In Exercise 3.6 we consider the Markov chain $(X_n)_{n\geq 0}$, with state space $S = \{0, 1\}$, and with matrix of transition probabilities P, given by

$$P = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Determine the stationary distribution π of this Markov chain. Do you find the same answer as in Exercise 3.7? Why or why not?

4.9 Suppose we have a vase with N ball. At each time $n \in \mathbb{N}$ some of these balls will be white and the others black (although it is possible that there are times n where all the balls are black, or all white. At every time n we throw a fair coin: if "heads" show we select completely at random a ball from the urn, and replace it by a *white* ball. In case we throw "tails", we also select completely randomly a ball from the urn, but now replace it by a *black* ball. Let X_n be the number of white balls in the urn at time n.

a. Explain in words why $(X_n)_{n\geq 0}$ is Markov chain. What is the state space S?

b. Determine the transition matrix P, i.e., determine the transition probabilities $p_{ij} = \mathrm{P}(X_{n+1} = j \mid X_n = i)$ for $i = 0, 1, \ldots, N$, $j = 0, 1, \ldots, N$.

c. What are the equivalence classes? Is the chain irreducible? What is the period $d(i)$ for $i = 0, 1, \ldots, N$?

d. Suppose that $N = 2$. Determine the stationary distribution π. What is m_i for $i = 0, 1, 2$?

4.10 A transition matrix P is called *doubly stochastic* if the sum over the entries of each column is equal to 1, i.e., if

$$\sum_i p_{ij} = 1, \qquad \text{for each } j.$$

Let $(X_n)_{n\geq 0}$ be an aperiodic irreducible Markov chain on the finite state space S, say $S = \{0, 1, \ldots M\}$. Show that the stationary distribution π is equal to the discrete uniform distribution on S, i.e., that

$$\pi_i = \frac{1}{M+1}, \qquad \text{for each } i \in S.$$

4.11 In Section 3.3 we considered as an example of a Markov chain on an infinite state space the simple random walk on $\mathbb{Z}$; cf. page 42. Here the transition probabilities for all $i \in \mathbb{Z}$ are given by

$$p_{i,i+1} = p, \quad p_{i,i-1} = q = 1 - p, \qquad \text{for some } p \in (0,1),$$

and we saw that the chain is irreducible, but that only in case $p = \frac{1}{2}$ the states are recurrent.

a. Show that the transition matrix P is doubly stochastic.
b. Use the main theorem to show that in case $p = \frac{1}{2}$ each state $i \in \mathbb{Z}$ is null-recurrent.

4.12 Five hippies are sitting in a circle, smoking a pipe. After every puff the smoker gives the pipe to the person to her/his right with probability p, and to the person to the left with probability $q = 1 - p$. For sake of convenience, the five people are numbered 1 to 5. Let X_n be the person holding the pipe after n puffs. Clearly, $(X_n)_{n\geq 0}$ is a Markov chain on the state space $S = \{1, \ldots, 5\}$.

a. Find the transition matrix P and classify the states.
b. What is the fraction of time the person who lit the pipe is actually smoking it?
c. Discuss the case when there are six smokers.

4.13 Suppose that we repeatedly throw an unbiased die, and that X_n is the outcome of the nth throw. We assume that the X_n's are independent. Define for for $n = 1, 2, \ldots$ the random variables S_n by

$$S_n = X_1 + X_2 + \cdots + X_n.$$

Show that

$$\lim_{n\to\infty} \mathrm{P}(S_n \text{ is a multiple of } 5) = \frac{1}{5}.$$

Hint: find a suitable Markov chain.

4.14 From the examples in Section 4.2 it is clear that aperiodicity is an important ingredient of the main theorem. In fact it is a theorem, that if $(X_n)_{n\geq 0}$ is an irreducible Markov chain on a finite state space S with transition matrix P, we always have that

$$\pi_i = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{[k]}, \qquad \text{for all } i, j \in S, \tag{4.4.4}$$

where π is the unique probability vector satisfying $\pi = \pi P$. We know from the main theorem that

$$\pi_i = \lim_{n\to\infty} p_{ij}^{[n]}, \qquad \text{for all } i, j \in S. \tag{4.4.5}$$

This suggests that (4.4.5) is a "stronger property" than (4.4.4) (i.e., that (4.4.5) implies (4.4.4)). This is indeed so, as the following will show.

a. Investigate whether the limits

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}a_k \qquad \text{and} \qquad \lim_{n\to\infty}a_n$$

exist, in the following two cases:
(i) $a_n=(-1)^n$, for $n\geq 1$.
(ii) $a_n=\frac{1}{n}$, for $n\geq 1$.

b. Let a be a real number, and let $(a_n)_{n\geq 1}$ be a sequence of real numbers converging to a, i.e.,

$$\lim_{n\to\infty}a_n=a.$$

Show that the limit

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n}a_k$$

exists, and is equal to a.

Hints: First show that it is enough to show that the statement holds in cases $a=0$. Next, set for $n\geq 1$,

$$S_n=\sum_{k=1}^{n}a_k,$$

and show that for every $\varepsilon>0$ there exists a positive integer N, such that

$$S_N-(n-N-1)<S_n<S_N+(n-N-1)\,,$$

for all $n\geq N$. Use this to show that $S_n/n\to 0$ as $n\to\infty$.

4.15 Let $(X_n)_{n\geq 0}$ be a Markov chain on the state space $S=\{1,2,3\}$, with transition matrix P given by

$$P=\begin{pmatrix}\frac{1}{3}&\frac{1}{2}&\frac{1}{6}\\ \frac{1}{6}&\frac{1}{3}&\frac{1}{2}\\ \frac{1}{2}&\frac{1}{6}&\frac{1}{3}\end{pmatrix}.$$

Furthermore, the initial distribution μ is given by

$$\mu=\begin{pmatrix}\frac{1}{3}&\frac{1}{3}&\frac{1}{3}\end{pmatrix}.$$

Determine:

a. $\displaystyle\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}X_k$. (*Hint*: use the ergodic theorem.)

b. $\mathrm{P}(X_n=i\,|\,X_{n+1}=j)$, for $i,j\in S$.

4.16 (Continuation of Exercise 4.15). Suppose that we are also given that for $n \geq 1$,

$$Y_n = X_n + X_{n-1}.$$

a. Calculate

$$\mathrm{P}(Y_3 = 5 \,|\, Y_2 = 3, Y_1 = 2) \qquad \text{and} \quad \mathrm{P}(Y_3 = 5 \,|\, Y_2 = 3, Y_1 = 3).$$

Is the stochastic process $(Y_n)_{n \geq 1}$ a Markov chain (on $S_Y = \{2,3,4,5,6\}$)?

b. Determine

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} Y_k.$$

4.17 Let us return once more to Ehrenfest's model of molecules in a gas. Suppose we made a film of the transitions, and we started this movie somewhere in the middle, without telling you whether it moved "forward" or "backward" in time. You wouldn't be able to tell the difference! In other words, the "forward transition probabilities" $p_{ij} = \mathrm{P}(X_{n+1} = j \,|\, X_n = i)$ are identical to the "backward transition probabilities" $q_{ij} = \mathrm{P}(X_n = j \,|\, X_{n+1} = i)$, for each $i, j \in S$. In fact, for any Markov chain one can "reverse the order of time," and get a new Markov chain, with transition matrix $Q = (q_{ij})$. Here, in the Ehrenfest example, we moreover have that $P = Q$. In general, when $P = Q$, we say that the Markov chain is *time-reversible*.

a. Show that the Markov chain $(X_n)_{n \geq 0}$ is time-reversible if and only if

$$\pi_i p_{ij} = \pi_j p_{ji}, \qquad \text{for all } i, j \in S.$$

b. Let $(X_n)_{n \geq 0}$ be an irreducible Markov chain, and let π be a probability vector, satisfying

$$\pi_i \geq 0, \quad \sum_{i \in S} \pi_1 = 1, \quad \text{and} \quad \pi_i p_{ij} = \pi_j p_{ji}, \quad \text{for all } i, j \in S.$$

Show that π is the unique stationary distribution of the chain.

4.18 In the remarks following the main theorem in Section 4.2 a method was given, how to find the stationary distribution π of an irreducible Markov chain. However, when this chain is time-reversible, the results of Exercise 4.17 yield an easier way to find π.

Find the stationary distribution π of the Ehrenfest example, by using Exercise 4.17.

5

Continuous-time Markov chains

In this chapter we will study the continuous-time analogue of the discrete time Markov chains from Chapters 3 and 4. This continuous-time Markov chain is a stochastic process $(X_t)_{t\in I}$, where the random variables X_t are indexed by some (time-)interval I (usually the interval $[0,\infty)$), and where the Markov property

"given the present state, the future is independent of the past"

applies. Continuous-time Markov chains have many important applications, and some of these—in queuing theory—will be outlined.

The analysis of continuous-time Markov chains is harder than that of its "brother," the discrete-time Markov chain, so often the proofs of its properties will only be mentioned, or heuristically motivated. Having said this, we will see that there are important similarities between these two stochastic processes. We will see for example, that the main theorem from Chapter 4 plays an important role in understanding the long-term behavior of continuous-time Markov chains. In the next section, a formal definition will be given, and we will briefly consider an important example of such stochastic processes, which is a process that you already have met in chapter 12 of [5]: the Poisson process. We will see in later sections that one of the fundamental properties of the Poisson process also allies to general continuous-time Markov chains; the interarrival times are exponentially distributed.

5.1 The Markov property

As in the discrete case we will assume that the random variables in the stochastic process $(X(t))_{t\geq 0}$ will[1] attain their values in some finite or countably infinite set S, the state space. The stochastic process $(X(t))_{t\geq 0}$ is a continuous time Markov chain if it satisfies the *Markov property*:

[1] We changed the notation from X_t to $X(t)$ for typographical reasons.

$$\mathrm{P}(X(t_{n+1}) = j \mid X(t_k) = i_k, 1 \le k \le n) = \mathrm{P}(X(t_{n+1}) = j \mid X(t_n) = i_n),$$

for any sequence $0 < t_1 < \cdots < t_n < t_{n+1}$ and for all $i_1, i_2, \dots, i_n, j \in S$. Furthermore, we will always assume that

$$\mathrm{P}(X(t_k) = i_k, 1 \le k \le n) > 0.$$

In case $\mathrm{P}(X(s) = i) > 0$, we define for $t \ge s$ the transition probability $p_{ij}(s,t)$ by

$$p_{ij}(s,t) = \mathrm{P}(X(t) = j \mid X(s) = i),$$

and we denote by $P(s,t)$ the stochastic matrix with entries $p_{ij}(s,t)$.

Example 6.1 (The Poisson process)

As an example of the kind of stochastic process we have in mind, recall the Poisson process $(N(t))_{t\ge 0}$ with intensity λ from chapter 12 in [5]. Such a process has the following properties:

(i) $N(0) = 0$, and $N(t) \ge N(s)$ whenever $t > s \ge 0$;
(ii) the process has *independent increments*, i.e., for all $t_0, t_1, \dots, t_n$ with

$$0 \le t_0 < t_1 < \cdots < t_n,$$

the random variables

$$N(t_0), N(t_1) - N(t_0), N(t_2) - N(t_1), \dots, N(t_n) - N(t_{n-1})$$

are independent;
(iii) for all $t, s \ge 0$,

$$\mathrm{P}(N(t+s) - N(t) = k) = \frac{(\lambda s)^k}{k!}\mathrm{e}^{-\lambda s}. \tag{5.5.1}$$

In order to show that the Poisson process $(N(t))_{t\ge 0}$ is a continuous time Markov chain, set

$$\mathcal{V} = \{X(t_k) = i_k, 1 \le k \le n\}$$

for $0 < t_1 < \cdots < t_n < s < t$ and $i_1, i_2, \dots, i_n, j \in S$. Then

$$\begin{aligned}
\mathrm{P}(N(t) = j \mid N(s) = i, \mathcal{V}) &= \frac{\mathrm{P}(N(t) = j, N(s) = i, \mathcal{V})}{\mathrm{P}(N(s) = i, \mathcal{V})} \\
&= \frac{\mathrm{P}(N(t) - N(s) = j - i, N(s) = i, \mathcal{V})}{\mathrm{P}(N(s) = i, \mathcal{V})} \qquad \text{(ii)} \\
&= \frac{\mathrm{P}(N(t) - N(s) = j - i)\,\mathrm{P}(N(s) = i, \mathcal{V})}{\mathrm{P}(N(s) = i, \mathcal{V})} \\
&= \mathrm{P}(N(t) - N(s) = j - i)\,.
\end{aligned}$$

A similar calculation yields that

$$\mathrm{P}(N(t) = j \mid N(s) = i) = \mathrm{P}(N(t) - N(s) = j - i)\,,$$

so we find that

$$\mathrm{P}(N(t) = j \mid N(s) = i, \mathcal{V}) = \mathrm{P}(N(t) = j \mid N(s) = i),$$

and it follows that the Poisson process $(N(t))_{t\geq 0}$ is indeed a continuous time Markov chain. It is time-homogeneous, because it follows from (iii) that $\mathrm{P}(N(t+s) = j \mid N(s) = i)$ is independent s, but only depends on t.

Quick exercise 5.1 Let $(N(t))_{t\geq 0}$ be a Poisson process with intensity λ. Determine $p_{ij}(s,t)$, for all $i, j \in \mathbb{N}$ and all $t \geq s \geq 0$. □

Another assumption we will make throughout is that we will only consider chains which are *time homogeneous*; the transition probability $p_{ij}(s,t)$ only depends on the time difference $t - s$, not on the values of s and t, i.e.,

$$p_{ij}(s,t) = p_{ij}(0, t-s), \qquad \text{for all } i, j \in S \text{ and } t \geq s.$$

In view of this it suffices to write $p_{ij}(t)$ in stead of $p_{ij}(0,t)$, and P_t in stead of $P(0,t)$. So we have that

$$p_{ij}(t) \geq 0, \qquad \text{and} \quad \sum_{j\in S} p_{ij}(t) = 1.$$

As in the discrete case we have the Chapman-Kolmogorov equations:

$$p_{ij}(s+t) = \sum_{k\in S} p_{ik}(s)p_{kj}(t), \qquad \text{for } s, t \geq 0. \tag{5.5.2}$$

Quick exercise 5.2 Give a proof of (5.5.2). □

5.2 The generator matrix

One of the handy things in dealing with discrete-time Markov chains is the transition matrix P. Clearly, for continuous-time Markov chains such a single matrix P does not exist, but rather a family $\{P_t\}_{t\geq 0}$ of transition matrices. This family has a few nice properties, which makes it into a *semigroup*; it satisfies

(a) $P_0 = I$, where I is the identity matrix;
(b) for each $t \geq 0$ we have that P_t is a stochastic matrix (i.e., the rows of P_t add up to 1);
(c) for all $t, s \geq 0$ we have that $P_{t+s} = P_t P_s$ ('Chapman-Kolmogorov').

Henceforth we will assume that the semigroup is continuous at the origin,

$$\lim_{h \downarrow 0} P_h = P_0,$$

where the convergence is pointwise for each entry, i.e.,

$$\lim_{h \downarrow 0} p_{ij}(h) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j. \end{cases} \tag{5.5.3}$$

This makes the semigroup $\{P_t\}_{t \geq 0}$ a *standard semigroup* (or: *continuous semigroup*). We have the following proposition.

Proposition 6.1 Let $\{P_t\}_{t \geq 0}$ be a standard semigroup on S. Then for all $i, j \in S$ and all $t \geq 0$ we have that

$$\lim_{h \to 0} p_{ij}(t+h) = p_{ij}(t),$$

that is, for all $i, j \in S$ and all $t \geq 0$ the function $t \mapsto p_{ij}(t)$ is continuous at t.

Proof. Since the halfgroup is continuous only from the right in $t = 0$, we need to address the limits from the left and from the right in t separately. From 'Chapman-Kolmogorov' and (5.5.3) it follows that

$$\lim_{h \downarrow 0} p_{ij}(t+h) = \lim_{h \downarrow 0} \sum_{k \in S} p_{ik}(t) p_{kj}(h) = p_{ij}(t).$$

The limit from the left is slightly more "tricky." Let $t > 0$, then for $0 \leq h \leq t$ we have that, again using 'Chapman-Kolmogorov',

$$\begin{aligned} p_{ij}(t) - p_{ij}(t-h) &= \sum_{k \in S} p_{ik}(t-h) p_{kj}(h) \ - p_{ij}(t-h) \\ &= \sum_{k \neq j} p_{ik}(t-h) p_{kj}(h) \ - p_{ij}(t-h) \left(1 - p_{jj}(h)\right). \end{aligned}$$

But then (5.5.3) implies that

$$\lim_{h \downarrow 0} \left(p_{ij}(t) - p_{ij}(t-h)\right) = 0.$$

□

The standard semigroup $\{P_t\}_{t \geq 0}$ on S is not a very "handy" object to work with. However, if the functions $t \mapsto p_{ij}(t)$ are continuous at $t = 0$ one can show that they are also differentiable in $t = 0$; for a proof, see e.g. [3], chapter 8, theorem 2.1. Now things become more easier. We can define for $i, j \in S$ numbers q_{ij} by

$$q_{ij} = \lim_{h \downarrow 0} \frac{p_{ij}(h) - p_{ij}(0)}{h} = p'_{ij}(0). \tag{5.5.4}$$

Quick exercise 5.3 Show that $q_{ii} \leq 0$, and that $q_{ij} \geq 0$ whenever $i \neq j$.

From (5.5.4) it follows that for $h > 0$ small,

$$p_{ij}(h) \approx q_{ij} \text{ if } i \neq j, \qquad \text{and} \quad p_{ii}(h) \approx 1 + q_{ii}(h).$$

Since $\sum_{j \in S} p_{ij}(h) = 1$, we find that

$$1 = \sum_{j \in S} p_{ij}(h) \approx 1 + h \sum_{j \in S} q_{ij},$$

yielding that

$$\sum_{j \in S} q_{ij} = 0, \qquad \text{for all } i \in S.$$

If S is finite, this can also be seen as follows:

$$\sum_{j \in S} q_{ij} = \sum_{j \in S} p'_{ij}(0) = \frac{\mathrm{d}}{\mathrm{d}t}\Big(\sum_{j \in S} p_{ij}(t)\Big)|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t}(1) = 0.$$

Note that in case S in countably infinite the exchange of sum and derivative is not always allowed. In fact, when this interchange is not allowed, the preceding (heuristical) derivation that $\sum_{j \in S} q_{ij} = 0$ is incorrect!

The matrix $Q = (q_{ij})$ is called the *generator* (or: *intensity matrix*) of the standard semigroup $\{P_t\}_{t \geq 0}$ on S. It takes over the role of the transition matrix P for discrete-time Markov chains. In a compact notation,

$$Q = \lim_{h \downarrow 0} (P_h - I)\,.$$

Setting $q(i) = -q_{ii}$, we have found that

$$\mathrm{P}(X(t+h) = j \mid X(t) = i) = p_{ij}(t) = \begin{cases} q_{ij}h + o(h),\ h \downarrow 0, & \text{if } i \neq j \\ 1 - q(i)h + o(h),\ h \downarrow 0, & \text{if } i = j. \end{cases}$$

We will assume for the rest of this chapter that for every state i

$$q(i) < \infty, \qquad \text{(the semigroup } \{P_t\}_{t \geq 0} \text{ is } \textit{stable}\text{)},$$

and

$$\sum_{j \in S} q_{ij} = 0 \qquad \text{(the semigroup } \{P_t\}_{t \geq 0} \text{ is } \textit{conservative}\text{)}.$$

Quick exercise 5.4 Let $(N(t))_{t \geq 0}$ be a Poisson process with intensity λ. Show that $q_{ii} = -\lambda$, $q_{i,i+1} = \lambda$, and that $q_{ii} = 0$ for all other values of $i, j \in \mathbb{N}$. $\square$

5.3 Kolmogorov's backward and forward equations

It follows from the property (c) of the semigroup $\{P_t\}_{t\geq 0}$, that for all $t \geq 0$ and all $h \geq 0$ one has that

$$\frac{P_{t+h} - P_0}{h} = P_t \frac{P_h - I}{h} = \frac{P_h - I}{h} P_t. \tag{5.5.5}$$

So in case S is a finite state space, we find that

$$\frac{\mathrm{d}}{\mathrm{d}t} P_t = P_t Q = Q P_t,$$

where Q is the generator of the semigroup $\{P_t\}_{t\geq 0}$. The equation

$$\frac{\mathrm{d}}{\mathrm{d}t} P_t = Q P_t, \tag{5.5.6}$$

is know as Kolmogorov's *backward equation*, while the equation

$$\frac{\mathrm{d}}{\mathrm{d}t} P_t = P_t Q, \tag{5.5.7}$$

is Kolmogorov's *forward equation*. Note, that these equations are differential equations, and that we have as initial condition that $P_0 = I$. In case S is finite, the unique solution of these equations is given by

$$P_t = \mathrm{e}^{tQ}.$$

$\big($Recall from your linear algebra classes that if A is a finite dimensional matrix,

$$\mathrm{e}^C = \sum_{n=0}^{\infty} \frac{C^n}{n!},$$

where the sum is taken per entry.$\big)$

In case S in infinite, problems may arise when taking the limit $h \downarrow 0$ in (5.5.5). We have the following theorem.

Theorem 6.1 (Kolmogorov's backward equation) If the standard semigroup $\{P_t\}_{t\geq 0}$ is stable and conservative, Kolmogorov's backward equation (5.5.6) is satisfied.

For the forward equation the result is far less general, mainly due to the lack of regularity assumptions on the trajectories of the Markov chain.

Theorem 6.2 (Kolmogorov's forward equation) If the standard semigroup $\{P_t\}_{t\geq 0}$ is stable and conservative, and moreover, if for all states i and all $t \geq 0$,

$$\sum_{j\in S} p_{ij}(t) q_j < \infty,$$

then Kolmogorov's forward equation (5.5.7) is satisfied.

For a proof of these results, see [3] (chapter 8, theorems 3.1 and 3.2), [6] (section 6.10), or [10] (theorem 2.1.1).

5.4 The generator matrix revisited

Up to now it is not clear why the generator Q is such a handy thing, comparable to the transition matrix P in the discrete case. To get an idea why this is so, let us first return to the Poisson process. Let $(N(t))_{t\geq 0}$ be a Poisson process with intensity λ. Define the random variable S_1 by

$$S_1 = \inf\{t \geq 0;\, N(t) = 1\},$$

(recall that by definition $N(0) = 0$), so S_1 is the first time t for which $N(t) = 1$. In general, for $k \geq 2$ define S_k by

$$S_k = \inf\{t \geq 0;\, N(t) = 1\}.$$

If we view the Poisson process as a bookkeeping in time of certain occurrences ("$N(t)$ is the number of customers who have entered a shop up to and including time t"), then S_k is the time when the kth occurrence took place. Setting $S_0 = 0$, and

$$T_k = S_k - S_{k-1} \qquad \text{for } k = 1, 2, \ldots,$$

one can show that the random variables T_k—the so-called *interarrival times*—are independent, $Exp(\lambda)$ distributed random variables (in fact, in [5] the Poisson process was defined in this way, and the properties (i), (ii), and (iii) are equivalent to this). In Quick exercise 5.4 we saw that $q(i) = \lambda = q_{i,i+1}$ for every $i \in \mathbb{N}$. This is no coincidence, as we now show.

Theorem 6.3 Let $(X(t))_{t\geq 0}$ be a continuous-time Markov chain on the state space S, with standard semigroup $\{P_t\}_{t\geq 0}$ and initial distribution μ. Moreover, suppose that, for $t, u \geq 0$,

$$\mathrm{P}(X(s) = i, t \leq s \leq t+u) = \lim_{n\to\infty} \mathrm{P}\Big(X\big(t + k\frac{u}{2^n}\big) = i, k = 0, 1, \ldots, 2^n\Big).$$

Then we have for $t, u \geq 0$,

$$\mathrm{P}(X(s) = i, t \leq s \leq t+u \,|\, X(t) = i) = \mathrm{e}^{-q(i)u}.$$

Proof. Note that repeatedly use of the Markov properties yields that

$$\mathrm{P}\Big(X\big(t + k\frac{u}{2^n}\big) = i, k = 0, 1, \ldots, 2^n\Big) = \mathrm{P}(X(t) = i)\,\Big(p_{ii}\big(\frac{u}{2^n}\big)\Big)^{2^n}.$$

By definition of $q(i)$ it follows that (by taking $h = u/2^n$),

$$p_{ii}\big(\frac{u}{2^n}\big) = 1 - \frac{q(i)u}{2^n} + o\big(\frac{1}{2^n}\big), \qquad \text{as } n \to \infty.$$

But then we have, due to the "calculus exercise" from page 26, that

$$\lim_{n\to\infty} \Big(p_{ii}\big(\frac{u}{2^n}\big)\Big)^{2^n} = \mathrm{e}^{-q(i)u},$$

and we find that

$$\mathrm{P}(X(s) = i, t \leq s \leq t+u) = \mathrm{P}(X(t) = i)\,\mathrm{e}^{-q(i)u}.$$

The desired follows if we divide the left- and righthand side of this last equation by $\mathrm{P}(X(s) = i)$. □

Now suppose that the chain is in state i at time t. Given this event, we define the random variable T_i as the remaining time (the "holding time") that the chain will be in i;

$$T_i = \inf\{s \geq 0;\ X(t+s) \neq i\}.$$

It is no coincidence that this random variable bears the same name as the random variable T_i we defined for the Poisson process at the beginning of this Section! Theorem 6.3 states that *given* that the chain is at time t in state i, T_i has an $Exp(q(i))$ distribution. So at time $t + T_i$ the chain is for the first time after time t in a state different from i. To which state j (different from i) does the chain jump? Note that

$$\begin{aligned}
&\mathrm{P}(X(t+T_i) = j, T_i > u, X(t) = i) = \\
&= \lim_{n\to\infty} \sum_{N=1}^{\infty} \mathrm{P}\Big(X(t+k\frac{u}{2^n}) = i, k = 0, \ldots, 2^n + N - 1, X(t + (2^n + N)\frac{u}{2^n}) = j\Big) \\
&= \lim_{n\to\infty} \mathrm{P}(X(t) = i)\Big(p_{ii}(\frac{u}{2^n})\Big)^{2^n} \sum_{N=1}^{\infty} \Big(p_{ii}(\frac{u}{2^n})\Big)^{N-1} p_{ij}(\frac{u}{2^n}) \\
&= \lim_{n\to\infty} \mathrm{P}(X(t) = i)\Big(p_{ii}(\frac{u}{2^n})\Big)^{2^n} \frac{\left(\frac{u}{2^n}\right)}{1 - \left(\frac{u}{2^n}\right)} \\
&= \mathrm{P}(X(t) = i)\,\mathrm{e}^{-q(i)u}\,\frac{q_{ij}}{q(i)}.
\end{aligned}$$

We find that

$$\mathrm{P}(X(t+T_i) = j, T_i > u \,|\, X(t) = i) = \mathrm{e}^{-q(i)u}\,\frac{q_{ij}}{q(i)},$$

or in words: given that $X(t) = i$, the chain stays an exponentially distributed time T_i in state i, and then jumps (independently of T_i) to state j with probability $\frac{q_{ij}}{q(i)}$.

Define the "jump-matrix" $R = (r_{ij})$ by

$$r_{ij} = \begin{cases} \frac{q_{ij}}{q(i)} & \text{if } i \neq j \\ 0 & \text{if } i = j. \end{cases}$$

Quick exercise 5.5 Show that R is a Markov matrix (i.e., a stochastic matrix) on S.

So our previous characterization of a continuous-time Markov chain is as follows: given that $X(t) = i$, the chain stays an exponentially distributed time T_i in state i, and then jumps (independently of T_i) to state j according to the stochastic matrix R.

This characterization makes it possible to simulate continuous-time Markov chains. We also see that in essence discrete-time and continuous-time "behave" in the same way, the difference being that the "time" between two steps is discrete in the former, and exponentially distributed in the latter case.

Example 6.1

Consider a factory with two machines and one repairman in case one or both machines break down. Operating machines break down after an exponentially distributed time with parameter λ. Repair times are exponentially distributed with parameter μ.

Let $X(t)$ be the number of operational machines at time t, then $X(t)$ can attain as values 0, 1, and 2, so we have $S = \{0, 1, 2\}$ as state space. Suppose that at time t both machines work, so $X(t) = 2$, and let X_1 and X_2 be the time until failure of the first respectively the second machine. Let $T_2 = \min\{X_1, X_2\}$, then T_2 is—as the minimum of two independent exponentially distributed random variables, both with parameter λ—exponentially distributed with parameter 2λ. So $X(s) = 2$ for $s \in [t, t + T_2)$ and $X(s) = 1$ for[2] $s = t + T_2$. Suppose that at this time $s = t + T_2$ the first machine breaks down. Due to the memoryless property the residual lifetime of the second machine starts all over again! We now have one operating machine, and one machine the repairman is trying to fix. Let Y be the time needed to fix the broken machine, and X_{res} the time the operating machine still runs. Then $T_1 = \min\{X_{\text{res}}, Y\}$, and we have that $X(s) = 1$ if $s \in [t + T_2, t + T_2 + T_1)$, and that at time $s = t + T_2 + T_1$ the chain either jumps with probability r_{10} to state 0 (the second machine also breaks down, while the repairman is still working on the broken machine), or to state 2 with probability r_{12} (the repairman finished repairing the broken machine before the other machine broke down). Etcetera. Note that T_1 is exponentially distributed with parameter $\lambda + \mu$ (because $T_1 = \min\{X_{\text{res}}, Y\}$), and T_0 is exponentially distributed with parameter μ. The schematic representation in Figure 5.1 helps us to find the generator matrix Q.

[2] Note that with probability zero both machines will break down at exactly the same moment.

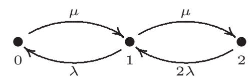

Fig. 5.1. The intensities of the transitions in Example 6.1

We find that

$$Q = \begin{pmatrix} -\mu & \mu & 0 \\ \lambda & -(\lambda+\mu) & \mu \\ 0 & 2\lambda & -2\lambda \end{pmatrix}.$$

Quick exercise 5.6 Determine the jump matrix R.

Example 6.2

Suppose that in the previous example we have two repairmen, each having an exponentially distributed repair time with parameter μ, and where these repair times are independent. Furthermore, we assume that each repairmen works alone. In this case the transition intensities are given in Figure 5.2.

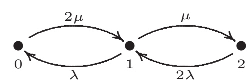

Fig. 5.2. The transition intensities in Example 6.2

For this example we find that

$$Q = \begin{pmatrix} -2\mu & 2\mu & 0 \\ \lambda & -(\lambda+\mu) & \mu \\ 0 & 2\lambda & -2\lambda \end{pmatrix}.$$

Example 6.3

In the previous example the assumption that both repairmen can do the job "in the same time" seems to be somewhat unrealistic. Suppose that the first repairman has an exponential repair time with parameter μ_1, and that the second repairman has an exponential repair time with parameter μ_2, and say $\mu_1 < \mu_2$ (i.e., in the mean the second repairman works faster than the first one). In view of this, the board of directors of the factory have decided that the second repairman always should start to repair a broken machine after a period in which both machines were operational. Once a repairman starts to work on a machine, he/she will also finish the job (so the first repairman is not taken from his job if the second repairman finished repairing her machine before the first repairman was finished). Clearly, we cannot describe this anymore as a continuous Markov chain with state space $S = \{0, 1, 2\}$ (as we did in the previous two examples). We need to "split" state 1. We set $S = \{0, 1_1, 1_2, 2\}$, and $X(t) = 1_1$ now means that at time t one machine is operational, and that repairman 1 is working on the other—broken—machine. In the same way, $X(t) = 1_2$ now means that at time t one machine is operational, and that repairman 2 is working on the other machine. The transition intensities are given in Figure 5.3.

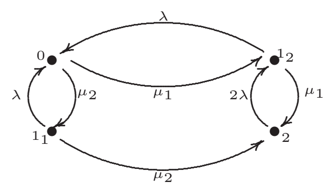

Fig. 5.3. The transition intensities in Example 6.2

Now the intensities matrix Q is given by

$$Q = \begin{pmatrix} -(\mu_1 + \mu_2) & \mu_2 & \mu_1 & 0 \\ \lambda & -(\lambda + \mu_1) & 0 & \mu_1 \\ \lambda & 0 & -(\lambda + \mu_2) & \mu_2 \\ 0 & 0 & 2\lambda & -2\lambda \end{pmatrix}.$$

5.5 Asymptotic behavior

In this section we will study the asymptotic behavior of a continuous-time Markov chain on a finite or countably infinite state space S. A lot of notation will be recycled from the discrete case. For example, the chain is called *irreducible* if for any pair $i, j \in S$ we have that $p_{ij}(t) > 0$ for some t.

Lemma 6.1 For any pair of states $i, j \in S$ we either have either that $p_{ij}(t) = 0$ for all $t > 0$, or that $p_{ij}(t) > 0$ for all $t > 0$.

Proof. Since $\lim_{h \downarrow 0} p_{ii}(h) = 1$, we find that $p_{ii}(t) > 0$ for small values of t. We can use the inequality

$$p_{ii}(t + h) \geq p_{ii}(t) p_{ii}(h),$$

to show that $p_{ii}(t) > 0$ for all $t \geq 0$.

Now suppose that $i \neq j$, and there exists a value $t > 0$ for which $p_{ij}(t) > 0$, so $i \to j$. But then we can find states $i_0 = i, i_1, i_2, \ldots, i_{n-1}, i_n = j$ for which

$$q_{i_0 i_1} q_{i_1 i_2} q_{i_2 i_3} \cdots q_{i_{n-1} i_n} > 0.$$

Since $i_{k-1} \neq i_k$ for $k = 1, 2 \ldots, n$, we see that

$$p_{i_{j-1} i_j}(h) = q_{i_{j-1} i_j} h + o(h), \qquad h \downarrow 0$$

implies that $p_{i_{j-1} i_j}(h) > 0$ for h sufficiently small. But then we find that

$$p_{ij}(t) \geq p_{i_0 i_1}(t/n) p_{i_1 i_2}(t/n) \cdots p_{i_{n-1} i_n}(t/n) > 0.$$

□

Another definition which is recycled from the discrete case, is that of a *stationary distribution*. The vector $\pi = (\pi_i)_{i \in S}$ is a stationary distribution of the chain if π is a probability vector (i.e., $\pi_i \geq 0$ for each $i \in S$, and $\sum_{i \in S} \pi_i = 1$), and $\pi = \pi P_t$ for all $t \geq 0$. Note that if the probability vector $\mu^{(0)}$ is the initial distribution, so

$$\mathrm{P}(X(0) = i) = \mu_i \qquad \text{for } i \in S,$$

then the distribution $\mu^{(t)}$ at time t is given by

$$\mu^{(t)} = \mu^{(0)} P_t.$$

So if $\mu^{(0)} = \pi$, then $\mu^{(t)} = \pi$ for all t.

Recall, that for discrete-time Markov chains we find π by solving $\pi = \pi P$. For continuous-time Markov chains we need to solve $\pi = \pi P_t$ for all t. This might seem to be a hard task, but the intensities matrix Q comes to our aid here. As in the discrete case the stationary distribution need not exist, or—if it exists—need not to be unique. We have the following theorem, which we state for *finite* state spaces S; from the proof it will be clear what extra conditions are needed in case S is countably infinite.

Theorem 6.4 Let $(X(t))_{t \geq 0}$ be an irreducible continuous-time Markov chain with finite state space S, standard semigroup $\{P_t\}_{t \geq 0}$, and generator matrix Q. Then for every $i, j \in S$ the limit

$$\lim_{t \to \infty} p_{ij}(t) = \pi_j, \tag{5.5.8}$$

exists, where $\pi = (\pi_i)_{i \in S}$ is the unique probability vector, satisfying

$$\pi Q = 0.$$

Proof. First we show that the limit exists. Choose $h > 0$, then $P(h)$ is a stochastic matrix, whose entries are all positive due to Lemma 6.1. But then we see that $P(h)$ is equal to the transition matrix P of an irreducible and aperiodic discrete-time Markov chain $(Y_n)_{n \in \mathbb{N}}$ on S (and because S is finite this discrete chain is non-null recurrent); this discrete-time chain $(Y_n)_{n \in \mathbb{N}}$ is called the *skeleton* of the continuous-time chain $(X(t))_{t \geq 0}$. So the transition probabilities p_{ij} of the skeleton are given by $p_{ij} = p_{ij}(h)$, and therefore the n-step transition probabilities of the skeleton satisfy $p_{ij}^{[n]} = p_{ij}(nh)$, for $n \in \mathbb{N}$. Due to the main theorem from Chapter 4 we know that there exists for the skeleton a unique probability vector $\pi = (\pi_i)_{i \in S}$, satisfying

$$\lim_{n\to\infty} p_{ij}(nh) = \pi_j \qquad \text{for all } i, j \in S. \tag{5.5.9}$$

We now show that (5.5.8) holds. For all $t \geq 0$ and $0 \leq s \leq h$ one has on the one hand that

$$\begin{aligned}
p_{ij}(t+s) - p_{ij}(t) &= \sum_{k\in S} p_{ik}(s)p_{kj}(t) - p_{ij}(t) \\
&= \sum_{k\neq i} p_{ik}(s)p_{kj}(t) - (1 - p_{ii}(s))\, p_{ij}(t) \\
&\leq \sum_{k\neq i} p_{ik}(s) - (1 - p_{ii}(s))\, p_{ij}(t) \\
&= 1 - p_{ii}(s) - (1 - p_{ii}(s))\, p_{ij}(t) \\
&= (1 - p_{ii}(s))\,(1 - p_{ij}(t)) \qquad \text{since } 1 - p_{ij}(t) \leq 1 \\
&\leq 1 - p_{ii}(s),
\end{aligned}$$

while on the other hand we have that

$$\begin{aligned}
p_{ij}(t+s) - p_{ij}(t) &= \sum_{k\in S} p_{ik}(s)p_{kj}(t) - p_{ij}(t) \\
&= \sum_{k\neq i} p_{ik}(s)p_{kj}(t) - (1 - p_{ii}(s))\, p_{ij}(t) \\
&\geq -\,(1 - p_{ii}(s))\, p_{ij}(t) \\
&\geq -\,(1 - p_{ii}(s))\,.
\end{aligned}$$

This yields that

$$-\,(1 - p_{ii}(s)) \leq p_{ij}(t+s) - p_{ij}(t) \leq 1 - p_{ii}(s),$$

i.e.,

$$|p_{ij}(t+s) - p_{ij}(t)| \leq 1 - p_{ii}(s).$$

Note that,

$$\begin{aligned}
p_{ii}(s) &= \mathrm{P}(X(s) = i \,|\, X(0) = i) \geq \mathrm{P}(T_i > s \,|\, X(0) = i) \\
&= 1 - \mathrm{P}(T_i \leq s \,|\, X(0) = i),
\end{aligned}$$

and because T_i has an $Exp(q(i))$ distribution this yields that

$$1 - p_{ii}(s) \leq \mathrm{P}(T_i \leq s \,|\, X(0) = i) = \int_0^s q(i)\mathrm{e}^{-q(i)x}\mathrm{d}x = 1 - \mathrm{e}^{-q(i)s}.$$

So for all s with $0 \leq s \leq h$ we have:

$$|p_{ij}(t+s) - p_{ij}(t)| \leq 1 - \mathrm{e}^{-q(i)s}.$$

Now let $\varepsilon > 0$ and suppose we originally choose h is such a way, that

$$1 - \mathrm{e}^{-q(i)h} \leq \varepsilon/2.$$

$\big($Note this can always be done; for $h > 0$ fixed, let $m \in \mathbb{N}$ be such, that $\tilde{h} = h/m$ satisfies $1 - \mathrm{e}^{-q(i)\tilde{h}} \leq \varepsilon/2$. But then we have that, due to Chapman-Kolmogorov,

$$P(h) = \underbrace{P(\tilde{h})P(\tilde{h}) \cdots P(\tilde{h})}_{m \text{ times}}.$$

Let $\tilde{\pi}$ be the unique stationary distribution of $P(\tilde{h})$ (we know that $\tilde{\pi}$ exists and is unique for exactly the same reason we know that π exists and is unique), then

$$\tilde{\pi}P(h) = \tilde{\pi}\left(P(\tilde{h})\right)^m = \tilde{\pi},$$

and we find that $\tilde{\pi} = \pi$.$\big)$
From (5.5.9) it follows that the exists an $N \in \mathbb{N}$ such that

$$|p_{ij}(nh) - \pi_j| < \varepsilon/2 \qquad \text{for all } n \geq N.$$

For each $t \geq Nh$ there exists a unique $n \geq N$ such that $nh \leq t < (n+1)h$, and we find that

$$|p_{ij}(t) - \pi_j| \leq |p_{ij}(t) - p_{ij}(nh)| + |p_{ij}(nh) - \pi_j| < \varepsilon.$$

Since $\varepsilon > 0$ was chosen arbitrarily, we see that (5.5.8) follows. Note that this limit implies that π is unique. It now follows from Kolmogorov's backwards equation (5.5.6) that

$$\frac{\mathrm{d}}{\mathrm{d}h}\pi P_h = \pi Q P_h \qquad \text{for all } h > 0.$$

Because $\pi P_h = \pi$ we also see that

$$\frac{\mathrm{d}}{\mathrm{d}h}\pi P_h = \frac{\mathrm{d}}{\mathrm{d}h}(\pi) = 0,$$

so together we find that

$$\pi Q P_h = 0 \qquad \text{for all } h > 0.$$

Taking the limit $h \downarrow 0$ yields that

$$\lim_{h \downarrow 0} \pi Q P_h = \pi Q,$$

and we find that $\pi Q = 0$. $\square$

Remark. As we mentioned before, in case S is countable, but infinite, it is possible that π does not exist, even if he chain $(X(t))_{t\geq 0}$ is irreducible; it is still possible that the skeleton $(Y_n)_{n\in\mathbb{N}}$ is null-recurrent! In that case one has that

$$\lim_{t\to\infty} p_{ij}(t) = 0 \qquad \text{for all states } i \text{ and } j. \tag{5.5.10}$$

For a sketch of a proof of this, see [6]. So we see that for an irreducible continuous-time chain either the stationary distribution π exists, in which case (5.5.8) hold, or that π does not exist, in which case we have (5.5.10).

5.6 Birth-death processes

Continuous-time Markov chains which have wide applications in queuing theory are the so-called *birth-death processes*. Here we usually have that $S = \mathbb{N}$, and that $X(t) = i$ (for $i \in \mathbb{N}$) means that there are i persons "in the system" at time t (think for instance of the number of people waiting to be served in the bakery on the corner). New people arrive after an exponentially distributed time with parameter λ_i, and leave after an exponentially distributed time with parameter μ_i (so the arrival intensities λ_i and the departure intensities μ_i may depend on the state i). We call the parameters λ_i the *birth rates*, and the parameters μ_i the *death rates*. It is handy to set $\mu_0 = 0$.

Quick exercise 5.7 Why is the Poisson process a birth-death process? What are the λ_i's and μ_i's?

In Figure 5.4 the transition intensities are given.

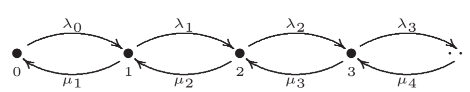

Fig. 5.4. The transition intensities in a birth-death process

From Figure 5.4 we see that the intensity matrix Q is given by:

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 \cdots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & 0 \cdots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \ddots \end{pmatrix}.$$

Quick exercise 5.8 Determine the jump matrix R for a birth-death process with birth rates λ_i and death rates μ_i.

In order to find the stationary distribution we must solve $\pi = \pi Q$. Writing this out, we find for birth-death processes the so-called *rate out = rate in* principle:

State i	Rate at which leave state i = rate at which state i is entered
0	$\lambda_0\pi_0 = \mu_1\pi_1$
1	$(\lambda_1+\mu_1)\pi_1 = \lambda_0\pi_0 + \mu_2\pi_2$
2	$(\lambda_2+\mu_2)\pi_2 = \lambda_1\pi_1 + \mu_3\pi_3$
$\vdots$	$\vdots$
$n\,(n \geq 1)$	$(\lambda_n+\mu_n)\pi_n = \lambda_{n-1}\pi_{n-1} + \mu_{n+1}\pi_{n+1}$

If we now add to each of these (infinitely many) equations the equation directly above it we find that

$$\begin{aligned}
\lambda_0\pi_0 &= \mu_1\pi_1\\
\lambda_1\pi_1 &= \mu_2\pi_2\\
\lambda_2\pi_2 &= \mu_3\pi_3\\
&\vdots\\
\lambda_n\pi_n &= \mu_{n+1}\pi_{n+1}, \quad n \geq 0,
\end{aligned}$$

i.e.,

$$\begin{aligned}
\pi_1 &= \tfrac{\lambda_0}{\mu_1}\pi_0\\
\pi_2 &= \tfrac{\lambda_1}{\mu_2}\pi_1\\
\pi_3 &= \tfrac{\lambda_2}{\mu_3}\pi_2\\
&\vdots\\
\pi_{n+1} &= \tfrac{\lambda_n}{\mu_{n+1}}\pi_n, \quad n \geq 0,
\end{aligned}$$

so recursively we find that

$$\pi_n = \frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}\pi_0.$$

Since $\pi_0 + \pi_1 + \cdots = 1$, we obtain

$$\pi_0 = \frac{1}{1+\sum_{n=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}},$$

and therefore, for $n \in \mathbb{N}$:

$$\pi_n = \frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}\frac{1}{1+\sum_{n=1}^{\infty}\frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n}}, \tag{5.5.11}$$

We see from (5.5.11) that the birth-death process with birth rates λ_i and death rates μ_i has a stationary distribution π if and only if

$$\sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} < \infty.$$

Examples

The $M/M/1$ queue. Suppose that $\lambda_i = \lambda$ for $i \geq 0$ and $\mu_i = \mu$ for $i \geq 1$. This models a shop with one server with an exponentially distributed service time with parameter μ, where customers arrive according to a Poisson process with intensity λ. So both the service time and the arrival times are exponentially distributed, and this "explains" the name of this model, where the M stands for "Markov." We find, setting $\rho = \lambda/\mu$, that

$$\pi_1 = \rho\pi_0, \quad \pi_2 = \rho^2\pi_0, \quad \pi_3 = \rho^3\pi_0,$$

and in general, $\pi_n = \rho^n \pi_0$, for $n \in \mathbb{N}$. We have now two possible situations:
(a) The "traffic intensity" ρ is smaller than 1, i.e., $0 < \rho < 1$. So $\lambda < \mu$, and we see that the expected time between two consecutive customers is greater than the expected time to serve a customer; in view of this we do not expect that the shop will fill up with ever growing number of customers. That this—heuristic—point of view is correct follows from the fact that

$$\sum_{i=0}^{\infty} \rho^i = \frac{1}{1-\rho} < \infty,$$

so π exists, and $\pi_i = \rho^i(1-\rho)$ for $i \in \mathbb{N}$. The expected number of customers in the shop will be (when the chain has attained its stationary distribution):

$$\mathrm{E}[X(\infty)] = \sum_{i=0}^{\infty} i\pi_i = \sum_{i=0}^{\infty} i\rho^i(1-\rho) = \frac{\rho}{1-\rho}.$$

(b) In case $\rho \geq 1$ the equation $\pi Q = 0$ implies that $\pi_i = 0$ for all i, so no stationary distribution exist. In time the number of customers in the shop will grow beyond any bound.

The $M/M/2$ queue. Suppose that $\lambda_i = \lambda$ for $i \geq 0$ and

$$\mu_i = \begin{cases} 0 & i = 0 \\ \mu & i = 1 \\ 2\mu & i \geq 2, \end{cases}$$

see also Figure 5.5
This models a shop with two servers, both with exponentially distributed service times with parameter μ, where customers arrive according to a Poisson process with intensity λ. In the name $M/M/2$ the M again stands for "Markov," while the "2" tells you that there are two servers. The "rate in = rate out" principle now yields that

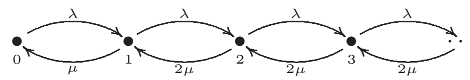

Fig. 5.5. The transition intensities in an $M/M/2$ queue

5.7 Solutions to the quick exercises

5.1 Obviously $p_{ij}(s,t) = 0$ if $i > j$, and for $i \leq j$ it follows from (5.1) that for $t \geq s$ the discrete random variable $N(t) - N(s)$ has a $Pois(\lambda(t-s))$ distribution. But then we have, that

$$\begin{aligned}
p_{ij}(s,t) &= \mathrm{P}(N(t) = j \mid N(s) = i) \\
&= \frac{\mathrm{P}(N(t+s) - N(t) = j - i, N(s) = i)}{\mathrm{P}(N(s) = i)} \qquad \text{(use (ii))} \\
&= \frac{\mathrm{P}(N(t+s) - N(t) = j - i)\,\mathrm{P}(N(s) = i)}{\mathrm{P}(N(s) = i)} \\
&= \mathrm{P}(N(t+s) - N(t) = j - i) \\
&= \frac{(\lambda(t-s))^{(j-i)}}{(j-i)!}\mathrm{e}^{-\lambda(t-s)}.
\end{aligned}$$

5.2 The proof of (5.5.2) is essentially the same as that of Theorem 3.2;

$$\begin{aligned}
p_{ij}(s+t) &= \mathrm{P}(X(s+t) = j \mid X(0) = i) \\
&= \sum_{k \in S} \mathrm{P}(X(s+t) = j, X(s) = k \mid X(0) = i) \\
&= \sum_{k \in S} \mathrm{P}(X(s+t) = j \mid X(s) = k)\mathrm{P}(X(s) = k \mid X(0) = i) \\
&= \sum_{k \in S} p_{ik}(s) p_{kj}(t).
\end{aligned}$$

5.3 Since $p_{ii} = 1$ and $0 \leq p_{ii}(h) \leq 1$ (after all, $p_{ii}(h)$ is a probability!), we see that $p_{ii}(h) - p_{ii}(0) \leq 0$ for all $h \geq 0$, implying that $q_{ii} \leq 0$. In case $i \neq j$ we have that $p_{ij}(0) = 0$, and thus we find—since $p_{ij}(h)$ is a probability—that $p_{ij}(h) - p_{ij}(0) \geq 0$, yielding that $q_{ij} \geq 0$.

5.4 In solving Quick exercise 5.1 we have seen that $p_{ij}(t) = 0$ whenever $i > j$, and that

$$p_{ij}(t) = \frac{(\lambda t)^{(j-i)}}{(j-i)!}\mathrm{e}^{-\lambda t}, \qquad \text{if } i \leq j.$$

But then we have that

$$p'_{ij}(t) = \begin{cases} 0 & \text{if } i > j \\ -\lambda \mathrm{e}^{-\lambda t} & \text{if } i = j \\ \lambda \mathrm{e}^{-\lambda t} - \lambda^2 t \mathrm{e}^{-\lambda t} & \text{if } j = i+1 \\ \lambda \left(\frac{(\lambda t)^{(j-i-1)}}{(j-i-1)!} \mathrm{e}^{-\lambda t} - \frac{(\lambda t)^{(j-i)}}{(j-i)!} \mathrm{e}^{-\lambda t} \right) & \text{if } j > i+1, \end{cases}$$

yielding that

$$q_{ij} = p'_{ij}(0) = \begin{cases} -\lambda & \text{if } i = j \\ \lambda & \text{if } j = i+1 \\ 0 & \text{in all other cases.} \end{cases}$$

5.5 Since $\sum_{j \in S} q_{ij} = 0$ and $q(i) = -q_{ii}$ for each $j \in S$, it follows that

$$q(i) = \sum_{j \in S, j \neq i} q_{ij}.$$

But the we see for $j \in S$, that

$$\sum_{j \in S} r_{ij} = \frac{1}{q(i)} \sum_{j \in S, j \neq i} q_{ij} = \frac{q(i)}{q(i)} = 1.$$

5.6 Since $q(0) = \mu$, $q(1) = \lambda + \mu$, and $q(2) = 2\lambda$, we find that

$$R = \begin{pmatrix} 0 & 1 & 0 \\ \frac{\lambda}{\lambda+\mu} & 0 & \frac{\mu}{\lambda+\mu} \\ 0 & 1 & 0 \end{pmatrix}.$$

5.7 In the Poisson process there are only births; $\lambda_i = \lambda$ for all $i \in \mathbb{N}$, and $\mu_i = 0$ for all $i \in \mathbb{N}$.

5.8 Note that $q(0) = \lambda_0$, and $q(i) = \lambda_i + \mu_i$ for $i \geq 1$. From Q we see that $r_{0i} = 1$ and

$$r_{i,i+1} = \frac{\lambda_i}{\lambda_i + \mu_i}, \quad r_{i,i-1} = \frac{\mu_i}{\lambda_i + \mu_i}, \qquad \text{for } i \geq 1.$$

For all other values of i and j we have that $r_{ij} = 0$.

5.8 Exercises

5.1 Consider the factory with two machines and one repairman, as described in Examples 6.1 and 6.2; see pages 77 and 78.

a. Determine the stationary distribution π for both examples.
b. If $\lambda = 1$ and $\mu = 2$, what is the proportion of the time in each of these examples that both machines are out-of-order?

5.2 Answer the same questions posed in Exercise 5.1, but now for the factory described in Example 6.3 on page 78.

5.3 Suppose that each bacteria in a group of bacteria either splits into two new bacteria after an exponentially distributed time with parameter λ, or dies after an exponentially distributed time with parameter μ.

a. Describe this as a birth-death process. What are the birth rates λ_i, and the death rates μ_i.
b. Can you explain heuristically why it is not possible to find a stationary distribution π?

5.4 In a birth-death process with birth rates λ_i for $i \geq 0$, and death rates μ_i for $i \geq 1$, determine the expected time to go from state 0 to state 3.

5.5 Show that Kolmogorov's backward equation yields for birth-death processes that

$$p'_{0j}(t) = \lambda_0 \left(p_{1j}(t) - p_{0j}(t)\right),$$
$$p'_{ij}(t) = \lambda_i p_{i+1,j}(t) + \mu_i p_{i-1,j}(t) - (\lambda_i + \mu_i) p_{ij}(t), \qquad i \geq 1.$$

5.6 A machine works for an exponentially distributed amount of time before breaking down. After breaking down, it takes an exponentially distributed amount of time to repair it. Let it be known that the expected time it runs is 1, and also that the expected repair time is 1. If the machine is working at time $t = 0$, what is the probability that it is also running at time $t = 100$? *Hint:* model this as a continuous Markov chain, with state space $S = \{0, 1\}$, where $X(t) = 0$ means that the machine is running at time t. What is Q? Q^n for $n \geq 0$? Finally, what is e^{tQ}?

5.7 Suppose we have almost the same situation as in Exercise 5.6, the difference being that now the expected time the machine runs is $1/\lambda$, and that the expected repair time is $1/\mu$. In this case it is more tedious to determine e^{tQ} (although you could use MAPLE to get an idea). In order to find $p_{00}(100)$ we use Kolmogorov's backward equation.

a. Setting $\lambda_0 = \lambda$, $\lambda_1 = 0$, $\mu_0 = 0$, and $\mu_1 = \mu$, show that

$$p'_{00}(t) = \lambda\left(p_{10}(t) - p_{00}(t)\right),$$
$$p'_{10}(t) = \mu\left(p_{00}(t) - p_{10}(t)\right).$$

b. Use your result from **a** to obtain that $\mu p'_{00}(t) + \lambda p'_{10}(t) = 0$. Show this implies that $\mu p_{00}(t) + \lambda p_{10}(t) = c$, for some constant c.

c. Argue that $c = \mu$, and derive that $p'_{00}(t) = \mu - (\lambda + \mu)p_{00}(t)$.

d. Setting

$$h(t) = p_{00}(t) - \frac{\mu}{\lambda + \mu},$$

derive that

$$h'(t) = -(\lambda + \mu)h(t). \tag{5.5.12}$$

e. Show that

$$h(t) = K\mathrm{e}^{-(\lambda+\mu)t}$$

is the solution of the differential equation (5.5.12).

f. Conclude from your answer in **e** that

$$p_{00}(t) = K\mathrm{e}^{-(\lambda+\mu)t} + \frac{\mu}{\lambda + \mu}.$$

What is K? Plug in $t = 100$.

g. Find P_t.

5.8 The probability and statistics group has two printers: a new and fast one, and one which is old and ragged. The old machine is used as a back-up for the new one (so effectively the group only has one printer). As soon as the new printer breaks down (after an exponentially distributed time, with parameter μ_{new}), it is replaced by the old one, and the systems manager will repair the new machine. Of course, the old machine can also break down (after an exponentially distributed time, with parameter μ_{old}). If the new machine will break down while the old one is being repaired, the systems manager will immediately stop working on the old one, and start working on the new printer. The repair time of the system manager is (for both machines) exponentially distributed, with parameter λ. What is the proportion of the time the probability and statistics group will have no printer available?

5.9 Customers arrive according to a Poisson process with intensity λ at *Jaap's Fish & Chips Joint*; a dinghy little place with one 'chef' where one can eat the best fried fish in Delft. The serving time of this 'chef' is exponentially distributed, with parameter μ. This would be the set-up of a standard $M/M/1$ queue, except that customers joint the queue in *Jaap's* restaurant with probability $1/(n+1)$ (and go with probability $n/(n+1)$ to another restaurant) if there are already n customers in *Jaap's* culinary Kingdom; here $n \geq 0$. Show that the stationary distribution $\pi = (\pi_0\ \pi_1\ \pi_2\ \dots)$ of this birth-death process has a Poisson distribution, with parameter λ/μ.

A

Answers to exercises

1.1a 1/3.

1.1b 1/2.

1.2 XXX

1.3a XXX

1.3b XXX

1.4a 1/3.

1.4b 4/9.

1.5a XXX

1.5b XXX

1.6a $e^{-\lambda p}$.

1.6b $e^{-\lambda p} \frac{(\lambda p)^k}{k!}$.

1.7a $P(X = k) = (1-p)^k p$, for $k = 1, 2, \ldots$.

1.7b $P(X = 1 \mid Y = 0) = 0$. For $k \geq 2$, $P(X = k \mid Y = 0) = (1-p)^{k-2} p$.

1.7c Use the answer of **b.**

1.7d $E(X \mid Y = 1) = 1$.

1.7e $h(Y) = 1 + Y E[X]$.

1.7f Calculate $E[h(Y)]$, with $h(Y)$ as found in **e.**

1.7g XXX

1.9 8.

1.10a $P(X = i, Y = j) = \frac{1}{161} \begin{cases} \text{if } 0 \leq i < 50 \text{ and } j = 0 \\ \text{or } 50 \leq i \leq 160 \text{ and } j = 1. \end{cases}$
$P(X = i, Y = j) = 0$ for all other choices of i and j.

1.10b 105 years.

1.11a $f_{X|Y}(x|y) = \frac{x+y}{y+\frac{1}{2}} 1_{[0,1]}(x)$.

1.11b $\frac{3y+2}{6y+3}$.

1.11c 7/12.

1.12a XXXXX

1.12b XXX

1.12c XXX

1.12d XXX

1.13 € 4.20

2.1a $G_X(s) = \frac{1}{4}\left(1 + s + s^2 + s^3\right)$.

2.1b $\mathrm{E}[X] = \frac{3}{2}$ and $\mathrm{Var}(X) = \frac{5}{4}$.

2.2a $A = \frac{1}{3}$.

2.2b XXX

2.2c XXX

2.2d XXX

2.2e XXX

2.3 XXX

2.4 $\frac{ps}{1-(1-p)s}$.

2.5a $G_S(s) = G_N\left(G_X(s)\right) = \mathrm{e}^{\mu p(s-1)}$.

2.5b $Pois(\mu p)$.

2.7a XXXXX

2.7b XXX

2.8 n/p.

2.9a $M_X(t) = \frac{1}{4}\left(1 + \mathrm{e}^t + \mathrm{e}^{2t} + \mathrm{e}^{3t}\right)$

2.9b $\mathrm{E}[X] = \frac{3}{2}$ and $\mathrm{Var}(X) = \frac{5}{4}$.

2.10a XXXXX

2.10b XXXXX

2.11a XXXXX

2.11b XXX

2.11c XXX

2.12 XXXXX

2.14 *Hint:* Use moment generating functions.

5.1a XXXXX

B

Solutions to selected exercises

1.1a Assuming that each child is either a boy or a girl with probability 1/2, independent from other brothers and/or sisters, we see that in the present situation the sample space is given by $\Omega = \{(G,G),\ (G,B),\ (B,G),\ (B,B)\}$, where for example (G,B) should be understood as: 'the first-born child is a girl, the younger child is a boy'. But then

$$\mathrm{P}((G,G)\,|\,(G,G),(G,B),(B,G)) = \frac{\mathrm{P}((G,G))}{\mathrm{P}((G,G),(G,B),(B,G))} = \frac{1/4}{3/4} = \frac{1}{3},$$

which is the desired probability.

1.1b In the present situation the sample space Ω will be the same as in **a.**, but the interpretation of (G,B) should be adapted to your observation. To distinguish between youngest and oldest child is now irrelevant; one should distinguish between the child you have seen, and the other child. So (G,B) should now be understood as: 'the child you have seen is a girl, the other child is a boy'. But then the *given* part is represented by the event $\{(G,G),(G,B)\}$, not by $\{(G,G),(G,B),(B,G)\}$ (as in **a.**). So the desired probability is now

$$\mathrm{P}((G,G)\,|\,(G,G),(G,B)) = \frac{\mathrm{P}((G,G))}{\mathrm{P}((G,G),(G,B))} = \frac{1/4}{2/4} = \frac{1}{2}.$$

1.4a One has that

$$\begin{aligned}\mathrm{P}(X=0\,|\,Y=0,Z=0) &= \frac{\mathrm{P}(X=0,Y=0,Z=0)}{\mathrm{P}(Y=0,Z=0)}\\ &= \frac{\mathrm{P}(X=0,Y=0,Z=0)}{\mathrm{P}(X=0,Y=0,Z=0)+\mathrm{P}(X=1,Y=0,Z=0)}\\ &= \frac{1/8}{1/8+1/16} = \frac{2}{3}.\end{aligned}$$

and $\mathrm{P}(X=1\,|\,Y=0,Z=0) = 1-\mathrm{P}(X=0\,|\,Y=0,Z=0) = \frac{1}{3}$. But then

$$\mathrm{E}(X\,|\,Y=0,Z=0) = \sum_{a=0}^{1} a\mathrm{P}(X=a\,|\,Y=0,Z=0) = \frac{1}{3}.$$

1.4b One has that

$$\begin{aligned} P(X=0 \,|\, Y=0) &= \frac{P(X=0, Y=0)}{P(Y=0)} \\ &= \frac{P(X=0,Y=0,Z=0)+P(X=0,Y=0,Z=1)}{P(X=0,Y=0,Z=0)+P(X=1,Y=0,Z=0)+P(X=0,Y=0,Z=1)+P(X=1,Y=0,Z} \\ &= \frac{1/8+3/16}{1/8+1/16+3/16+0} = \frac{4}{9}. \end{aligned}$$

and $P(X=1 \,|\, Y=0) = 1 - P(X=0 \,|\, Y=0) = \frac{5}{9}$. But then

$$E(X \,|\, Y=0, Z=0) = \sum_{a=0}^{1} a P(X=a \,|\, Y=0) = \frac{4}{9}.$$

1.6a Let X be the number of intoxicated people, stopped by the police, and let N be the total number of cars, stopped by the police. Then

$$\begin{aligned} P(X=0) &= \sum_{n=0}^{\infty} P(X=0 \,|\, N=n) P(N=n) \\ &= \sum_{n=0}^{\infty} P(X=0 \,|\, N=n) \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \sum_{n=0}^{\infty} (1-p)^n \frac{\lambda^n}{n!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda(1-p))^n}{n!} \\ &= e^{-\lambda} e^{\lambda(1-p)} \\ &= e^{-\lambda p}. \end{aligned}$$

1.6b Let X be the number of intoxicated people, stopped by the police, and let N be the total number of cars, stopped by the police. Then

$$\begin{aligned} P(X=0) &= \sum_{n=0}^{\infty} P(X=0 \,|\, N=n) P(N=n) \\ &= \sum_{n=0}^{\infty} P(X=0 \,|\, N=n) \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \sum_{n=k}^{\infty} \binom{n}{k} p^k (1-p)^{n-k} \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \dots \\ &= \frac{(\lambda p)^k}{k!} e^{-\lambda p}. \end{aligned}$$

So we see that X has a *Pois*(λp) distribution.

1.7a X has a *Geo*(p) distribution, so $P(X=k) = (1-p)^k p$, for $k = 1, 2, \dots$.

1.7b If $k = 1$, the event $\{X = 1, Y = 0\}$ is empty, so $\mathrm{P}(X = 1 \,|\, Y = 0) = 0$. For $k \geq 2$ we have that $\{X = k, Y = 0\} = \{X = k\}$, and consequently

$$\mathrm{P}(X = k \,|\, Y = 0) = \frac{\mathrm{P}(X = k, Y = 0)}{\mathrm{P}(Y = 0)} = \frac{\mathrm{P}(X = k)}{\mathrm{P}(Y = 0)} = \frac{(1-p)^{k-1}p}{1-p} = (1-p)^{k-2}p.$$

1.7c By definition,

$$\begin{aligned}
\mathrm{E}(X \,|\, Y = 0) &= \sum_{k=1}^{\infty} k\mathrm{P}(X = k \,|\, Y = y) = \sum_{k=2}^{\infty} k(1-p)^{k-2}p \\
&= \sum_{\ell=1}^{\infty} (\ell+1)(1-p)^{\ell-1}p \\
&= \sum_{\ell=1}^{\infty} (1-p)^{\ell-1}p + \sum_{\ell=1}^{\infty} \ell(1-p)^{\ell-1}p \\
&= \sum_{\ell=1}^{\infty} \mathrm{P}(X = \ell) + \sum_{\ell=1}^{\infty} \ell\mathrm{P}(X = \ell) \\
&= 1 + \mathrm{E}[X] \,.
\end{aligned}$$

1.7d Note that $\mathrm{P}(X = 1 \,|\, Y = 1) = \mathrm{P}(X = 1, Y = 1) \,/\mathrm{P}(Y = 1) = p/p = 1$. Furthermore, for $k \geq 2$ one has that $\mathrm{P}(X = k, Y = 1) = \mathrm{P}(\emptyset) = 0$. But then we find that

$$\mathrm{E}(X \,|\, Y = 1) = \sum_{k=1}^{\infty} k\mathrm{P}(X = k \,|\, Y = 1) = 1.$$

1.7e From **c.** and **d.** it follows that $\mathrm{E}(X \,|\, Y) = 1 + (1-Y)\mathrm{E}[X]$. So $h(Y) = 1 + Y\mathrm{E}[X]$.

1.7f In **e.** we found that $h(Y) = 1 + (1-Y)\mathrm{E}[X]$. Since $\mathrm{E}[X] = \mathrm{E}[\mathrm{E}(X \,|\, Y)]$, we find that $\mathrm{E}[X] = \mathrm{E}[h(Y)] = \mathrm{E}[1 + (1-Y)\mathrm{E}[X]] = 1 + \mathrm{E}[X]\,(1 - \mathrm{E}[Y]) = 1 + (1-p)\mathrm{E}[X]$. Solving for $\mathrm{E}[X]$ yields the desired result.

1.8 Clearly this result implies our theorem; choose $g(y) = 1$ for all y. Now due to the change of variable formula we have

$$\begin{aligned}
\mathrm{E}[\mathrm{E}(X \,|\, Y)g(Y)] &= \int_{-\infty}^{\infty} \mathrm{E}(X \,|\, Y = y)g(y)f_Y(y)\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \Big(\int_{-\infty}^{\infty} xg(y)f_{X|Y}(x|y)\,\mathrm{d}x \Big) f_Y(y)\,\mathrm{d}y \\
&= \iint_{-\infty}^{\infty} xg(y)f(x,y)\,\mathrm{d}x\mathrm{d}y \\
&= \mathrm{E}[Xg(Y)] \,.
\end{aligned}$$

1.11a Since $f_Y(y) = \int_0^1 (x+y)\,\mathrm{d}x = y + \frac{1}{2}$, for $0 \leq y \leq 1$, we find for x between 0 and 1 that

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{x+y}{y+\frac{1}{2}}.$$

Furthermore, $f_{X|Y}(x|y) = 0$ since $f(x,y) = 0$ for values of x outside the interval $[0,1]$.

1.11b By definition, for y between 0 and 1,

$$\mathrm{E}(X \mid Y = y) = \int_0^1 \frac{x(x+y)}{y+\frac{1}{2}}\,\mathrm{d}x = \left[\frac{\frac{1}{3}x^3 + \frac{1}{2}x^2 y}{y+\frac{1}{2}}\right]_{x=0}^{x=1} = \frac{3y+2}{6y+3}.$$

1.11c Using (1.1.4) we find

$$\mathrm{E}[X] = \int_{-infty}^{\infty} x f_{X|Y}\,\mathrm{d}x = \int_0^1 \frac{3y+2}{6y+3}\left(y+\frac{1}{2}\right)\mathrm{d}y = \frac{1}{6}\int_0^1 (3y+2)\,\mathrm{d}y = \frac{7}{12}.$$

By symmetry, $f_X(x) = x + \frac{1}{2}$ for x between 0 and 1 (and $f_X(x) = 0$ for x outside the interval $[0,1]$). But then,

$$\mathrm{E}[X] = \int_0^1 x(x+\frac{1}{2})\,\mathrm{d}x = \left[\frac{1}{3}x^3 + \frac{1}{4}x^2\right]_0^1 = \frac{7}{12}.$$

1.13 Let M be the money the casino has to pay, and let N be the number of throws needed for the first appearance of an outcome different from 6. Clearly N has a *Geo*(5/6) distribution, so $\mathrm{E}[N] = 6/5$. Now let X be a random variable with a discrete uniform distribution on $1,\ldots,5$, i.e., $\mathrm{P}(X = i) = 1/5$ for $i = 1,\ldots,5$. Heuristically, X represents the last throw, the throw in which for the first time the outcome was different from 6. Then one has that $\mathrm{E}(M \mid N = 1) = \mathrm{E}[X] = 3$, $\mathrm{E}(M \mid N = 2) = 6 + \mathrm{E}[X] = 9$, $\mathrm{E}(M \mid N = 3) = 2 \cdot 6 + \mathrm{E}[X] = 15$, and in general, for $n \geq 1$:

$$\mathrm{E}(M \mid N = n) = \mathrm{E}[X] = 6(n-1) + 3 = 6n - 3. \tag{B.B.1}$$

Due to (1.1) we find that

$$\begin{aligned}\mathrm{E}[M] = \mathrm{E}[\mathrm{E}(M \mid N)] &= \sum_{k=1}^{\infty} \mathrm{E}(M \mid N = k)\mathrm{P}(N = k) \\ &= \sum_{k=1}^{\infty} (6n-3)\mathrm{P}(N = k) = 6\mathrm{E}[N] - 3 = 4.20.\end{aligned}$$

Note that we can write (B.B.1) also as $\mathrm{E}(M \mid N) = 6N - 3$, so

$$\mathrm{E}[M] = \mathrm{E}[\mathrm{E}(M \mid N)] = \mathrm{E}[6N - 3] = 6\mathrm{E}[N] - 3 = 4.20.$$

2.1a By definition,

$$G_X(s) = \sum_{k=0}^{3} \mathrm{P}(X = k)\, s^k = \frac{1}{4}\left(1 + s + s^2 + s^3\right).$$

2.1b Note that

$$G'_X(s) = \frac{1}{4}\left(1 + 2s + 3s^2\right), \qquad \text{and } G''_X(s) = \frac{1}{4}(2 + 6s),$$

so by Theorem 2.2 we have

$$\mathrm{E}[X] = G'_X(1) = \frac{3}{2}, \qquad \text{and } \mathrm{Var}(X) = G''_X(1) + G'_X(1) - \left(G'_X(1)\right)^2 = \frac{5}{4}.$$

2.2a By definition,

$$G_X(s) = \sum_{k=0}^{\infty} \mathrm{P}(X = k)\, s^k,$$

implying that $G_X(1) = \sum_{k=0}^{\infty} \mathrm{P}(X = k) = 1$. So

$$1 = A(1 + 1 + 1^2) = 3A,$$

i.e., $A = \frac{1}{3}$.

2.3 YYYY

2.4 By definition,

$$\begin{aligned} G_X(s) &= \sum_{k=1}^{\infty} (1-p)^{k-1} p s^k = ps \sum_{k=1}^{\infty} \left((1-p)s\right)^{k-1} \\ &= ps \sum_{\ell=0}^{\infty} \left((1-p)s\right)^{\ell} = \frac{ps}{1-(1-p)s}. \end{aligned}$$

2.5a From Examples 2.1 we know that $G_X(s) = 1 - p + ps$ is the probability generating function of each of the X_i's, while $G_N(s) = \mathrm{e}^{\mu(s-1)}$ is the probability generating function of N. But then Theorem 2.4 yields that

$$G_S(s) = G_N\left(G_X(s)\right) = \mathrm{e}^{\mu(1-p+ps-1)} = \mathrm{e}^{\mu p(s-1)}.$$

2.5b From Example 2.1(c) and the remark following Quick exercise 2.4 it follows that G_S is the probability generating function of a *Pois*(μp) distributed random variable.

2.6 In [5], Section 10.1, we have seen that $X + Y$ has a *Bin*($n + m, p$) distribution. This can also be seen using probability generating functions. In Example 2.2 we have seen that the probability generating functions of X respectively Y are given by

$$G_X(s) = (1 - p + ps)^n, \quad \text{respectively } G_Y(s) = (1 - p + ps)^m.$$

Due to Theorem 2.3 we have that

$$G_{X+Y}(s) = (1 - p + ps)^{n+m},$$

and since the probability generating function of a random variable uniquely determines its distribution, it follows that $X + Y$ has a *Bin*($n + m, p$) distribution.

2.8 There are several ways to ind the answer. Let N_n be the number of throws needed, then $N_n = X_1 + X_2 + \cdots + X_n$, where $X_1, X_2, \ldots, X_n$ are independent *Geo*(1/6) distributed random variable. But then we have, that

$$\mathrm{E}[N_n] = \mathrm{E}[X_1 + X_2 + \cdots + X_n] = n\mathrm{E}[X_1] = \frac{1}{p}.$$

Another way is to determine the probability generating function of N_n. Due to Corollary 2.1 and Exercise 2.4 we have

$$G_{N_n}(s) = g_{X_1}(s) G_{X_2}(s) \cdots G_{X_n}(s) = \left(G_{X_1}(s)\right)^n = \left(\frac{ps}{1-(1-p)s}\right)^n.$$

From Theorem 2.2(a) we find that

$$\mathrm{E}[N_n] = G'_{N_n}(1) = n\left(\frac{p\cdot 1}{1-(1-p)\cdot 1}\right)^{n-1}\frac{p}{(1-(1-p)\cdot 1)^2} = \frac{n}{p}.$$

2.9a $M_X(t) = \sum_{k=0}^{3}\mathrm{e}^{tk}\mathrm{P}(X=k) = \frac{1}{4}\left(1+\mathrm{e}^t+\mathrm{e}^{2t}+\mathrm{e}^{3t}\right)$.

2.9b Note that

$$M'_X(t) = \frac{1}{4}\left(\mathrm{e}^t+2\mathrm{e}^{2t}+3\mathrm{e}^{3t}\right), \qquad \text{and } M''_X(t) = \frac{1}{4}\left(\mathrm{e}^t+4\mathrm{e}^{2t}+9\mathrm{e}^{3t}\right).$$

So by Theorem 2.5 we have

$$\mathrm{E}[X] = M'_X(0) = \frac{3}{2}, \qquad \text{and } \mathrm{E}\left[X^2\right] = M''_X(0) = \frac{7}{2},$$

and we find that $\mathrm{Var}(X) = \frac{5}{4}$. See also Exercise 2.1.

2.13 XXXX

2.14

3.6b Clearly $f_{00}^{[1]} = \frac{1}{3}$, and $f_{00}^{[2]} = \frac{2}{3}\cdot\frac{1}{2} = \frac{1}{3}$. For $n \geq 2$: $f_{00}^{[n]} = \frac{2}{3}\cdot\left(\frac{1}{2}\right)^{n-2}\frac{1}{2} = \frac{2}{3}\cdot\left(\frac{1}{2}\right)^{n-1}$. So

$$\begin{aligned} f_0 &= \frac{1}{3} + \frac{2}{3}\frac{1}{2} + \frac{2}{3}\left(\frac{1}{2}\right)^2 + \frac{2}{3}\left(\frac{1}{2}\right)^3 + \cdots \\ &= \frac{1}{3} + \frac{2}{3}\left(\sum_{k=0}^{\infty}\left(\frac{1}{2}\right)^k - 1\right) \\ &= \frac{1}{3} + \frac{2}{3}\left(\frac{1}{1-\frac{1}{2}} - 1\right) = 1. \end{aligned}$$

3.6c By Corollary 3.2 we have, that for $j \in S$,

$$\mathrm{P}(X_n = j) = (\mu P^n)_i = \sum_{i\in S}\mu_i p_{ij}^{[n]}.$$

Note that for $n = 1$

$$\begin{aligned} \mu P &= \begin{pmatrix} 3/7 & 4/7 \end{pmatrix}\begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \\ &= \begin{pmatrix} 3/7 & 4/7 \end{pmatrix}, \end{aligned}$$

so $\mu P^2 = \mu P = \mu$, and in general (using induction), $\mu P^n = \mu$. We find that $\mathrm{P}(X_n = 0) = \frac{3}{7}$ for all $n \geq 1$, so μ is *stationary.*

3.7a $\det(P) = \frac{1}{3}\cdot\frac{1}{2} - \frac{2}{3}\cdot\frac{1}{2} = -\frac{1}{6}$

3.7b

3.7c

References

1. N.H. Abel. "untersuchungen über die reihe $1 + \frac{m}{1}x + \frac{m(m-1)}{1\cdot 2}x^2 + \frac{m(m-1)(m-2)}{1\cdot 2\cdot 3}x^3 + \cdots$ usw". *Jour. für Reine und Angew. Math*, 1:311–339, 1826.
2. Deborah J. Bennett. *Randomness*. Harvard University Press, Cambridge, Massachusetts, third edition, 1998.
3. Pierre Brémaud. *Markov chains*, volume 31 of *Texts in Applied Mathematics*. Springer-Verlag, New York, 1999. Gibbs fields, Monte Carlo simulation, and queues.
4. Karma Dajani and Cor Kraaikamp. *Ergodic theory of numbers*, volume 29 of *Carus Mathematical Monographs*. Mathematical Association of America, Washington, DC, 2002.
5. F.M. Dekking, C. Kraaikamp, H.P. Lopuhaä, and L.E. Meester. *A modern whatever*. Springer, London, first edition, 2005.
6. G. R. Grimmett and D. R. Stirzaker. *Probability and random processes*. The Clarendon Press Oxford University Press, New York, second edition, 1992.
7. Olle Häggström. *Finite Markov chains and algorithmic applications*, volume 52 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2002.
8. Marius Iosifescu. *Finite Markov processes and their applications*. John Wiley & Sons Ltd., Chichester, 1980. Wiley Series in Probability and Mathematical Statistics.
9. R.W.J. Meester. Verwarring door onvolledige informatie: conditionele kansen. *Nieuwe Wiskrant*, 17(3):20–21, 1998.
10. J. R. Norris. *Markov chains*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998. Reprint of 1997 original.
11. Reinhold Remmert. *Theory of Complex Functions*. Springer, New York, first edition, 1991.
12. S.M. Ross. *A first course in probability*. Prentice-Hall, Inc., New Jersey, sixth edition, 1984.

Index

central limit theorem
  for averages, 25
convergence in distribution, 24
convolution, 16

doubly stochastic, 65

joint
  discrete distribution
    of sum and maximum, 2
  probability mass function
    of sum and maximum, 2