8

Stochastic Processes

8.1 INTRODUCTION

Many real-world applications of probability theory have the particular feature that data are collected sequentially in time. A few examples are weather data, stock market indices, air-pollution data, demographic data, and political tracking polls. These also have in common that successive observations are typically not independent. We refer to any such collection of observations as a *stochastic process*. Formally, a stochastic process is a collection of random variables that take values in a set S, the *state space*. The collection is indexed by another set T, the *index set*. The two most common index sets are the natural numbers $T = \{0, 1, 2, ...\}$, and the nonnegative real numbers $T = [0, \infty)$, which usually represent discrete time and continuous time, respectively. The first index set thus gives a sequence of random variables $\{X_0, X_1, X_2, ...\}$ and the second, a collection of random variables $\{X(t), t \geq 0\}$, one random variable for each time t. In general, the index set does not have to describe time but is also commonly used to describe spatial location. The state space can be finite, countably infinite, or uncountable, depending on the application.

In order to be able to analyze a stochastic process, we need to make assumptions on the dependence between the random variables. In this chapter, we will focus on the most common dependence structure, the so called *Markov property*, and in the next section we give a definition and several examples.

8.2 DISCRETE-TIME MARKOV CHAINS

You are playing roulette, in each round betting $\$1$ on odd. You start with $\$10$ and after each round record your new fortune. Suppose that the first five rounds gives the sequence loss, loss, win, win, win, which gives the sequence of fortunes

$$9,\ 8,\ 9,\ 10,\ 11$$

and that you wish to find the distribution of your fortune after the next round, given this information. Your fortune will be 12 if you win, which has probability $\frac{18}{38}$, and 10 if you lose, which has probability $\frac{20}{38}$. One thing we realize is that this depends only on the fact that the current fortune is $\$11$ and not the values prior to that. Generally, if your fortunes in the first n rounds are the random variables $X_1, ..., X_n$, then the conditional distribution of X_{n+1} given $X_1, ..., X_n$ depends only on X_n. This is a fundamental property, and we state the following general definition.

***Definition* 8.2.1.** Let $X_0, X_1, X_2, ...$ be a sequence of discrete random variables, taking values in some set S and that are such that

$$P(X_{n+1} = j | X_0 = i_0, ..., X_{n-1} = i_{n-1}, X_n = i) = P(X_{n+1} = j | X_n = i)$$

for all $i, j, i_0, ..., i_{n-1}$ in S and all n. The sequence $\{X_n\}$ is then called a *Markov chain*.

We often think of the index n as discrete time and say that X_n is the *state* of the chain at time n, where the state space S may be finite or countably infinite. The defining property is called the *Markov property*, which can be stated in words as "conditioned on the present, the future is independent of the past."

In general, the probability $P(X_{n+1} = j | X_n = i)$ depends on i, j, and n. It is, however, often the case (as in our roulette example) that there is no dependence on n. We call such chains *time-homogeneous* and restrict our attention to these chains. Since the conditional probability in the definition thus depends only on i and j, we use the notation

$$p_{ij} = P(X_{n+1} = j | X_n = i), \ \ i, j \in S$$

and call these the *transition probabilities* of the Markov chain. Thus, if the chain is in state i, the probabilities p_{ij} describe how the chain chooses which state to jump to next. Obviously the transition probabilities have to satisfy the following two criteria:

$$\text{(a) } p_{ij} \geq 0, \quad \text{for all } i, j \in S, \qquad \text{(b) } \sum_{j \in S} p_{ij} = 1, \quad \text{for all } i \in S$$

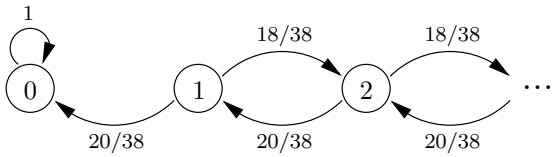
***Example* 8.2.1.** In the roulette example above, the state space is

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

and if the chain is in state $i \geq 1$, it can jump to either $i-1$ or $i+1$ according to the transition probabilities

$$p_{i,i-1} = \frac{20}{38} \quad \text{and} \quad p_{i,i+1} = \frac{18}{38}$$

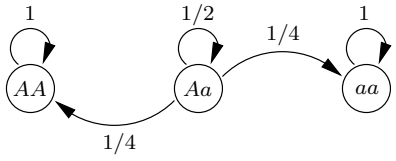
When $i = 0$, this means that you are ruined and cannot play anymore. Thus, you can jump to 0 but not from it. It is customary to describe this by letting $p_{00} = 1$, thus imagining that the chain performs an eternal sequence of jumps from 0 to itself. The diagram below shows a way to describe a Markov chain as a graph, which we refer to as the *transition graph*. The arrows show the possible transitions and their corresponding probabilities. Note that the sum of the numbers on the arrows going *out* from each state is 1. This is criterion (b) above.



□

***Example* 8.2.2.** A certain gene in a plant has two alleles, A and a (see Section 1.6.2). Thus, its genotype with respect to this gene can be AA, Aa, or aa. Now suppose that a plant is crossed with itself and one offspring selected that is crossed with itself and so on and so forth. Describe the sequence of genotypes as a Markov chain.

The state space is $S = \{AA, Aa, aa\}$, which also shows that states do not have to be numbers.[1] The Markov property is clear, since the offspring's genotype depends only on the parent plant, not the grandparent. Clearly, genotypes AA and aa can have only themselves as offspring and for the type Aa, we recall the Punnett square from Section 1.6.2 to get the following transition graph.



□

[1]The picky probabilist then refers to the X_k as *random objects* rather than random variables. If we wish, we can rename the states $\{1, 2, 3\}$ instead, where the numbers have no role other than serving as labels.

It is convenient to summarize the transition probabilities in the *transition matrix* P, which has p_{ij} as its (i,j)th entry. Depending on the state space, the transition matrix may be finite or infinite. Thus, in the genetics example we have

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 0 & 1 \end{pmatrix}$$

and in the roulette example the infinite matrix

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \dots \\ 20/38 & 0 & 18/38 & 0 & 0 & \dots \\ 0 & 20/38 & 0 & 18/38 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

8.2.1 Time Dynamics of a Markov Chain

The most fundamental aspect of a Markov chain in which we are interested is how it develops over time. The transition matrix provides us with a description of the stepwise behavior, but suppose that we want to compute the distribution of the chain two steps ahead. Let

$$p_{ij}^{(2)} = P(X_2 = j | X_0 = i)$$

and condition on the intermediate step X_1. The law of total probability gives

$$\begin{aligned} p_{ij}^{(2)} &= \sum_{k \in S} P(X_2 = j | X_0 = i, X_1 = k) P(X_1 = k | X_0 = i) \\ &= \sum_{k \in S} P(X_2 = j | X_1 = k) P(X_1 = k | X_0 = i) \; = \; \sum_{k \in S} p_{ik} p_{kj} \end{aligned}$$

where we used the Markov property for the second-to-last equality. We switched the order between the factors in the sum to get the intuitively appealing last expression; in order to go from i to j in two steps, we need to visit *some* intermediate step k and jump from there to j. Now recall how matrix multiplication works to help us realize from the expression above that $p_{ij}^{(2)}$ is the (i,j)th entry in the matrix P^2. Thus, in order to get the two-step transition probabilities, we square the transition matrix. Generally, define the *n-step transition probabilities* as

$$p_{ij}^{(n)} = P(X_n = j | X_0 = i)$$

and let $P^{(n)}$ be the n-step transition matrix. Repeating the argument above gives $P^{(n)} = P^n$, the nth power of the one-step transition matrix. In particular, this gives the relation

$$P^{(n+m)} = P^{(n)} P^{(m)}$$

for all m, n, commonly referred to as thehapman–Kolmogorov equations *Chapman–Kolmogorov equations.* Spelled out coordinatewise, they become

$$p_{ij}^{(n+m)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(m)}$$

for all m, n and all $i, j \in S$. In words, to go from i to j in $n + m$ steps, we need to visit some intermediate step k after n steps. We let $P^{(0)} = I$, the identity matrix.

***Example* 8.2.3.** Find the n-step transition matrix in the genetics example (Example 8.2.2).

The state space is $S = \{AA, Aa, aa\}$, and let us start with $n = 2$. We get

$$\begin{aligned} P^{(2)} &= \begin{pmatrix} 1 & 0 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 3/8 & 1/4 & 3/8 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

We now realize that 0s and 1s will remain in all powers of P, that the middle entry in $P^{(n)}$ is $p_{22}^{(n)} = (\frac{1}{2})^n$, and that by symmetry $p_{21}^{(n)} = p_{23}^{(n)}$. This gives the n-step transition matrix

$$P^{(n)} = \begin{pmatrix} 1 & 0 & 0 \\ (1 - (1/2)^n)/2 & (1/2)^n & (1 - (1/2)^n)/2 \\ 0 & 0 & 1 \end{pmatrix}$$

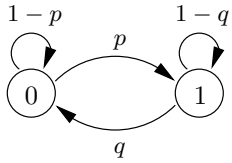
It is obvious without computations that the 0s and 1s remain unchanged; the types AA and aa can have offspring only of their own type. Also note how the probability to find the type Aa declines rapidly with n, indicating that eventually this genotype will disappear. We will return to this aspect of the transition matrix. □

It should be pointed out that computation of $P^{(n)}$ is seldom this simple and may be more or less impossible if the state space is large. Even for a small state space, the computation is not trivial, as the next example shows.

***Example* 8.2.4. (*ON/OFF System*).** Consider a system that alternates between the two states 0 (OFF) and 1 (ON) and that is checked at discrete timepoints. If the system is OFF at one timepoint, the probability that it has switched to ON at the next timepoint is p, and if it is ON, the probability that it switches to OFF is q. **(a)** Describe the system as a Markov chain. **(b)** Find the n-step transition matrix. **(c)** Suppose that $p = \frac{3}{4}$ and

$q = \frac{1}{2}$. If the system starts being OFF, what is the probability that it is ON at time $n = 3$?

For (a), the transition graph is



and the transition matrix

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

and computation of the powers of P needed for (b) is facilitated by diagonalization techniques from linear algebra. The eigenvalues of P are $\lambda_1 = 1$ and $\lambda_2 = 1 - p - q$ and it can be shown that

$$P^n = \frac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix} + \frac{\lambda_2^n}{p+q}\begin{pmatrix} p & -p \\ -q & q \end{pmatrix}$$

and you may verify that this satisfies the relation $P^{n+1} = P^n P$. To find the answer to (c), we need $p_{01}^{(3)}$, which is the $(0,1)$ entry in $P^{(3)}$. Thus, the probability that the system is ON at time $n = 3$, given that it starts being OFF, is

$$p_{01}^{(3)} = \frac{p}{p+q} + \frac{-p\lambda_2^3}{p+q} = \frac{3/4}{5/4} + \frac{-3/4 \times (-1/4)^3}{5/4} \approx 0.61$$ □

One interesting aspect of a Markov chain is its long term behavior. As it turns out, there are simple and elegant asymptotic results for Markov chains that makes this easy to deal with. Before we get to those results, let us consider asymptotics in some of our examples.

***Example* 8.2.5.** Recall the genetics example (Example 8.2.2). Find the limits of the transition probabilities as $n \to \infty$.

The n-step transition matrix is

$$P^{(n)} = \begin{pmatrix} 1 & 0 & 0 \\ (1-(1/2)^n)/2 & (1/2)^n & (1-(1/2)^n)/2 \\ 0 & 0 & 1 \end{pmatrix}$$

and letting $n \to \infty$ gives the matrix

$$\lim_{n\to\infty} P^{(n)} = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 0 & 1 \end{pmatrix}$$

Thus, if we start in state AA or aa, we stay there, and if we start in state Aa, we eventually end up in either AA or aa with equal probabilities. □

***Example* 8.2.6.** Recall the ON/OFF system in Example 8.2.4. Find the limits of the transition probabilities as $n \to \infty$.

The n-step transition matrix is

$$P^{(n)} = \frac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix} + \frac{\lambda_2^n}{p+q}\begin{pmatrix} p & -p \\ -q & q \end{pmatrix}$$

and, since $\lambda_2 = 1-p-q$ and thus $|\lambda_2| < 1$ (unless $p = q = 0$ or $p = q = 1$), letting $n \to \infty$ gives the matrix

$$\lim_{n\to\infty} P^{(n)} = \frac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix}$$

Note that the rows of this matrix are identical. Thus, at a late timepoint, the probabilities that the system is OFF and ON are approximately $q/(p+q)$ and $p/(p+q)$ respectively, regardless of the initial state. Note that if $q > p$, the probability to be OFF is larger, which makes sense. □

In the last example, the asymptotic probabilities do not depend on how the chain was started, and we call the distribution $(q/(p+q), p/(p+q))$ on the state space $\{0, 1\}$ a *limit distribution*. Compare with the genetics example where no limit distribution exists, since the asymptotic probabilities depend on the initial state. A question of general interest is when a Markov chain has a limit distribution. To be able to answer this, we need to introduce some criteria that enables us to classify Markov chains.

8.2.2 Classification of States

The graphic representation of a Markov chain illustrates in which ways states can be reached from each other. In the roulette example, state 1 can, for example, reach state 2 in one step and state 3, in two steps. It can also reach state 3 in four steps, through the sequence $2, 1, 2, 3$, and so on. One important property of state 1 is that it can reach any other state. Compare this to state 0 which cannot reach any other state. Whether or not states can reach each other in this way is of fundamental importance in the study of Markov chains, and we state the following definition.

***Definition* 8.2.2.** If $p_{ij}^{(n)} > 0$ for some n, we say that state j is *accessible* from state i, written $i \to j$. If $i \to j$ and $j \to i$, we say that i and j *communicate* and write this $i \leftrightarrow j$.

If j is accessible from i, this means that it is *possible* to reach j from i but not that this necessarily happens. In the roulette example, $1 \to 2$ since $p_{12} > 0$, but if the chain starts in 1, it may jump directly to 0, and thus it will never be able to visit state 2. In this example, all nonzero states communicate with each other and 0 communicates only with itself.

In general, if we fix a state i in the state space of a Markov chain, we can find all states that communicate with i and form the *communicating class* containing i. It is easy to realize that not only does i communicate with all states in this class but they all communicate with each other. By convention, every state communicates with itself (it can "reach itself in 0 steps") so every state belongs to a class. If you wish to be more mathematical, the relation "$\leftrightarrow$" is an equivalence relation and thus divides the state space into equivalence classes that are precisely the communicating classes. In the roulette example, there are two classes

$$C_0 = \{0\}, \;\; C_1 = \{1, 2, ...\}$$

and in the genetics example, each state forms its own class and we thus have

$$C_1 = \{AA\}, \;\; C_2 = \{Aa\}, \;\; C_3 = \{aa\}$$

In the ON/OFF system, there is only one class, the entire state space S. In this chain, all states communicate with each other, and it turns out that this is a desirable property.

***Definition* 8.2.3.** If all states in S communicate with each other, the Markov chain is said to be *irreducible*.

Another important property of Markov chains has to do with returns to a state. For example, in the roulette example, if the chain starts in state 1, it may happen that it never returns. Compare this with the ON/OFF system where the chain eventually returns to where it started (assuming that $p > 0$ and $q > 0$). We next classify states according to whether return is certain. We introduce the notation P_i for the probability distribution of the chain when the initial state X_0 is i.

***Definition* 8.2.4.** Consider a state $i \in S$ and let τ_i be the number of steps it takes for the chain to first visit i. Thus

$$\tau_i = \min\{n \geq 1 : X_n = i\}$$

where $\tau_i = \infty$ if i is never visited. If $P_i(\tau_i < \infty) = 1$, state i is said to be *recurrent* and if $P_i(\tau_i < \infty) < 1$, it is said to be *transient*.

A recurrent state thus has the property that if the chain starts in it, the time until it returns is finite. For a transient state, there is a positive probability that the time until return is infinite, meaning that the state is never revisited. This means that a recurrent state is visited over and over but a transient state is eventually never revisited.

Now consider a transient state i and another state j such that $i \leftrightarrow j$. We will argue that j must also be transient. By the Markov property, every visit to j starts a fresh Markov chain and since $i \leftrightarrow j$, there is a positive probability to visit i before coming back to j. We may think of this as repeated trials to reach i every time the chain is in j, and since the success probability is positive, eventually there will be a success. If j were recurrent, the chain would return to j infinitely many times and the trial would also succeed infinitely many times. But this means that there would be infinitely many visits to i, which is impossible since i is transient. Hence j must also be transient.

We have argued that transience (and hence also recurrence) is a *class property*, a property that is shared by all states in a communicating class. In particular, the following holds.

Corollary 8.2.1. In an irreducible Markov chain, either all states are transient or all states are recurrent.

This is convenient since we can classify the entire Markov chain as transient or recurrent by checking only one state. In the case of a finite state space, there is an easy way to classify the transient and recurrent states.

Corollary 8.2.2. Suppose that S is finite. A state i is transient if and only if there is another state j such that $i \to j$ but $j \not\to i$.

We omit the proof, referring instead to an intuitive argument. Every time the chain visits a transient state, there is a chance that it will never return again. In a finite state space, the only way in which this can happen is if there is some other state that can be reached but from where there is no path back. In an infinite state space, however, there is enough room for states to be transient even if they all communicate with each other. We also realize that if the state space is finite, there is not enough room for all states to be transient.

Corollary 8.2.3. If a Markov chain has finite state space, there is at least one recurrent state.

***Example* 8.2.7.** Classify the states as recurrent/transient in the ON/OFF system in Example 8.2.4.

To avoid trivialities, we assume that both p and q are strictly positive. Since the state space is finite, we can use Corollary 8.2.2 and note that since i and j communicate, they must both be recurrent. □

***Example* 8.2.8.** Classify the states as recurrent/transient in the roulette example (Example 8.2.1).

Here we must use the general definition. Let us start with state 0, which is trivially recurrent since if we start there, we are stuck there forever, that is, $\tau_0 \equiv 1$. As for state 1, if we start there and the first jump is to 0, we never return to 1, and thus $\tau_1 = \infty$ in this case. Hence $P(\tau_1 < \infty) < 1$ and state 1 is transient. Since 1 communicates with the states $2, 3, ...$, they are all transient.

The recurrent state 0 has the additional property that once the chain is there, it can never leave.[2] Such a state is called *absorbing*. □

A transient state i is revisited a number of times, which has a geometric distribution with success probability $P_i(\tau_i = \infty)$ (where "success" means that the state is never revisited). This means that the expected number of returns is $1/P_i(\tau_i = \infty)$, which is finite since $P_i(\tau_i = \infty) > 0$. On the other hand, for a recurrent state the expected

[2] A quote from *Hotel California* is sometimes given at this point in the presentation, but we resist the temptation.

number of returns is infinite (since already the *actual* number is infinite). Now let

$$I_n = \begin{cases} 1 & \text{if } X_n = i \\ 0 & \text{otherwise} \end{cases}$$

and let

$$S = \sum_{n=1}^{\infty} I_n$$

the total number of returns to state i. We get

$$E_i[S] = \sum_{n=1}^{\infty} E_i[I_n] = \sum_{n=1}^{\infty} P_i(X_n = i) = \sum_{n=1}^{\infty} p_{ii}^{(n)}$$

which gives the following nice characterization of transience/recurrence.[3]

Proposition 8.2.4. State i is

$$\text{transient if } \sum_{n=1}^{\infty} p_{ii}^{(n)} < \infty$$

$$\text{recurrent if } \sum_{n=1}^{\infty} p_{ii}^{(n)} = \infty$$

Since we have noted that it is often difficult to compute the $p_{ii}^{(n)}$, one may wonder how useful the last result is. However, we do not need to compute the exact value of the infinite sum, only determine whether it is convergent, and for this, it will suffice if we have some idea of how $p_{ii}^{(n)}$ behaves asymptotically in n. We will later see examples of this.

8.2.3 Stationary Distributions

Consider the ON/OFF system and suppose that we choose the initial state according to the probabilities $\nu_0 = P(X_0 = 0), \nu_1 = P(X_0 = 1) = 1 - \nu_0$. The distribution $\boldsymbol{\nu} = (\nu_0, \nu_1)$ is called an *initial distribution* and the probability distribution of the first state X_1 is computed by conditioning on X_0, which gives

$$P(X_1 = j) = p_{0j}\nu_0 + p_{1j}\nu_1, \quad j = 0, 1$$

[3]In the calculation, we interchanged summation and expectation, which is not always allowed when the sum is infinite. However, it can be shown that it is justified if the summands are nonnegative as in this case.

or in matrix notation

$$(P(X_1 = 0), P(X_1 = 1)) = \boldsymbol{\nu} P$$

Suppose in particular that we take $\nu_0 = q/(p+q), \nu_1 = p/(p+q)$, the limit distribution, as initial distribution. We then get

$$P(X_1 = 0) = (1-p)\frac{q}{p+q} + q\frac{p}{p+q} = \frac{q}{p+q} = \nu_0$$

and $P(X_1 = 1) = \nu_1$. In matrix notation, $\boldsymbol{\nu} = \boldsymbol{\nu} P$, which means that the distribution does not change over time. This is an important observation, and we state the following general definition.

***Definition* 8.2.5.** Let P be the transition matrix of a Markov chain with state space S. A probability distribution $\boldsymbol{\pi} = (\pi_1, \pi_2, ...)$ on S satisfying

$$\boldsymbol{\pi} P = \boldsymbol{\pi}$$

is called a *stationary distribution* of the chain.

The entries of $\boldsymbol{\pi}$ thus satisfy

$$\pi_j = \sum_{i \in S} p_{ij} \pi_i, \quad \text{for all } j \in S$$

and together with the condition

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

this determines the stationary distribution. The intuition behind the probability π_j is that it describes what proportion of time that is spent in state j in the long run. Other terms are *invariant distribution* and *equilibrium distribution*.

There are, however, some caveats: (1) a stationary distribution may not always exist and (2) there may be more than one. The uniqueness problem goes away if we make our usual assumption of irreducibility, an observation that we state without proof.

Proposition 8.2.5. Consider an irreducible Markov chain. If a stationary distribution exists, it is unique.

This is helpful, but we can still not guarantee that a stationary distribution exists. Things simplify if the state space is finite.

Proposition 8.2.6. If S is finite and the Markov chain is irreducible, a unique stationary distribution $\boldsymbol{\pi}$ exists.

Rather than giving the proof, we examine our examples to illustrate how to compute the stationary distribution and what can go wrong if the chain is not irreducible.

***Example* 8.2.9.** Find the stationary distribution for the ON/OFF system in Example 8.2.4.

Since the chain is finite and irreducible, the stationary distribution exists and is unique. The equation $\boldsymbol{\pi}P = \boldsymbol{\pi}$ becomes

$$(\pi_0\ \pi_1)\begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} = (\pi_0\ \pi_1)$$

from which we take the first equation

$$(1-p)\pi_0 + q\pi_1 = \pi_0$$

which gives

$$\pi_1 = \frac{p}{q}\pi_0$$

The second equation is

$$p\pi_0 + (1-q)\pi_1 = \pi_1$$

which also gives $\pi_1 = (p/q)\pi_0$. To get a solution, we note that $\pi_0 + \pi_1 = 1$, which gives

$$\pi_0\left(1 + \frac{p}{q}\right) = 1$$

which gives stationary distribution

$$\boldsymbol{\pi} = \left(\frac{q}{p+q},\ \frac{p}{p+q}\right)$$

Note how $\pi_0 > \pi_1$ if $q > p$. This makes sense since if the chain is more likely to jump from 1 to 0 than the other way, in the long run it spends more time in 0. □

Note how the two equations for π_0 and π_1 in the example turned out to be the same. In general, if the state space has r states, the equation $\boldsymbol{\pi}P = \boldsymbol{\pi}$ gives at most $r-1$ linearly independent equations, and in addition to these, we also have the equation $\sum_{j\in S}\pi_j = 1$. Recalling results from linear algebra, this means that there always exists a solution to this system of equations, but unless the chain is irreducible, there may be more than one solution.

***Example* 8.2.10.** Find the stationary distribution in the genetics example (Example 8.2.2).

The chain is not irreducible, but let us still attempt to find the stationary distribution. The states are AA, Aa, and aa, and the equation $\boldsymbol{\pi}P = \boldsymbol{\pi}$ becomes

$$(\pi_{AA}\ \pi_{Aa}\ \pi_{aa})\begin{pmatrix} 1 & 0 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 0 & 1 \end{pmatrix} = (\pi_{AA}\ \pi_{Aa}\ \pi_{aa})$$

from which we get the first equation

$$\pi_{AA} + \frac{1}{4}\pi_{Aa} = \pi_{AA}$$

which gives $\pi_{Aa} = 0$. Knowing this, the second equation gives only $0 = 0$, and the third gives $\pi_{aa} = \pi_{aa}$. Thus, any distribution of the form

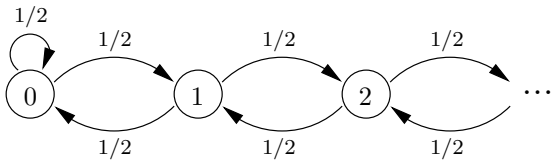
$$\boldsymbol{\pi} = (\alpha,\ 0,\ 1-\alpha)$$

where $0 \le \alpha \le 1$ qualifies as a stationary distribution. The evolution of the chain is simple; we choose either AA or aa, according to $\boldsymbol{\pi}$, and whatever state we choose stays forever. Thus, we get the sequence $AA, AA, ...$ with probability α and $aa, aa, ...$ with probability $1-\alpha$. Perhaps not very exciting, but a good illustration of what can happen without irreducibility. □

Things are a little more complicated if the state space is infinite. Consider, for example, the following variant of Example 1.6.17.

***Example* 8.2.11.** Recall the gambler's ruin problem in Example 1.6.17, where Ann starts with one dollar and Bob is infinitely wealthy. Also suppose that Ann has an infinitely wealthy and benevolent uncle who gives her a dollar to bet every time she goes broke. Describe the Markov chain and find the stationary distribution.

The state space of Ann's possible fortune is $S = \{0, 1, 2, ...\}$. If $i \ge 1$, the possible transitions are to states $i-1$ and $i+1$ and in state 0, Ann either wins and pays back her uncle's dollar or loses her uncle's dollar, stays in 0, and borrows another dollar. The transition graph is



and the transition matrix is

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & \dots \\ 1/2 & 0 & 1/2 & 0 & 0 & \dots \\ 0 & 1/2 & 0 & 1/2 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

The equation $\boldsymbol{\pi} P = \boldsymbol{\pi}$ now gives the first equation

$$\frac{1}{2}\pi_0 + \frac{1}{2}\pi_1 = \pi_0$$

which gives $\pi_1 = \pi_0$. The second equation is

$$\frac{1}{2}\pi_0 + \frac{1}{2}\pi_2 = \pi_1$$

which gives $\pi_2 = 2\pi_1 - \pi_0 = \pi_0$. The remaining equations all look the same:

$$\frac{1}{2}\pi_{n-2} + \frac{1}{2}\pi_n = \pi_{n-1}, \;\; n \geq 3$$

which gives $\pi_n = 2\pi_{n-1} - \pi_{n-2} = \pi_0$. Thus, a stationary distribution must be of the form

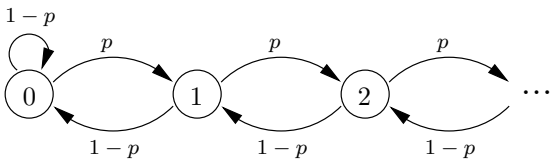
$$\boldsymbol{\pi} = (\pi_0, \; \pi_0, \; \pi_0, ...)$$

which is obviously a problem since we cannot sum these probabilities to 1. If $\pi_0 = 0$, the sum is 0, and if $\pi_0 > 0$, the sum is infinite. We conclude that *no stationary distribution exists.* □

The technique used in this example to express all the π_n in terms of π_0 is the standard way to try to find a stationary distribution. In this case it turned out to be impossible since the π_n did not sum to one.

***Example* 8.2.12.** Reconsider the previous problem under the assumption that Bob has an edge in the game, so that in each round Ann wins with probability $p < \frac{1}{2}$.

The transition graph is now



and the transition matrix

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

The equation $\boldsymbol{\pi} P = \boldsymbol{\pi}$ gives

$$(1-p)\pi_0 + (1-p)\pi_1 = \pi_0$$

which gives

$$\pi_1 = \frac{p}{1-p}\pi_0$$

The next equation is

$$p\pi_0 + (1-p)\pi_2 = \pi_1$$

which gives

$$\pi_2 = \frac{1}{1-p}(\pi_1 - p\pi_0) = \left(\frac{p}{1-p}\right)^2 \pi_0$$

The remaining equations are

$$p\pi_{n-2} + (1-p)\pi_n = \pi_{n-1}$$

and it is easily verified that

$$\pi_n = \left(\frac{p}{1-p}\right)^n \pi_0, \;\; n = 0, 1, 2, ...$$

satisfy these equations. To find π_0, we use the condition $\sum_{j \in S} \pi_j = 1$ and get

$$1 = \pi_0 \sum_{n=0}^{\infty} \left(\frac{p}{1-p}\right)^n = \pi_0 \frac{1-p}{1-2p}$$

which gives stationary distribution

$$\pi_n = \frac{1-2p}{1-p}\left(\frac{p}{1-p}\right)^n, \;\; n = 0, 1, 2, ...$$

□

The Markov chains are irreducible and recurrent in both examples; the only difference is that the probability $\frac{1}{2}$ was replaced by $p < \frac{1}{2}$ in the second. Why, then, does a stationary distribution exist in the second but not the first chain? We need to introduce yet another classifying property for Markov chains.

***Definition* 8.2.6.** Let i be a recurrent state. If $E_i[\tau_i] < \infty$, then i is said to be *positive recurrent*. If $E_i[\tau_i] = \infty$, i is said to be *null recurrent*.

This is a more subtle distinction than between recurrence and transience. Any recurrent state is revisited infinitely many times, but only a positive recurrent state is revisited in such a way that the expected time between visits is finite (recall from Example 2.4.9 that a random variable can be finite and yet have an infinite expectation). Thus, recurrence/transience is distinguished by the τ_i themselves and positive/null recurrence by the expected values $E_i[\tau_i]$. It can be shown that positive recurrence is also a class property, and hence we have the following corollary.

Corollary 8.2.7. For an irreducible Markov chain, there are three possibilities: (a) all states are positive recurrent, (b) all states are null recurrent, (c) all states are transient.

Now consider a finite state space. There cannot be any null recurrent states, a fact that we will not prove, but the intuition is that there simply is not enough room for very long paths of return. Also recall that there must be at least one recurrent state and hence this state must be positive recurrent. Thus, if a finite chain is irreducible, it must also be positive recurrent. If the state space is infinite, this is not true because of (b) in Corollary 8.2.7. The following result covers both finite and infinite state spaces.

Proposition 8.2.8. Consider an irreducible Markov chain $\{X_n\}$. Then

$$\text{A stationary distribution } \boldsymbol{\pi} \text{ exists } \Leftrightarrow \{X_n\} \text{ is positive recurrent}$$

If this is the case, $\boldsymbol{\pi}$ is unique and has $\pi_j > 0$ for all $j \in S$.

The intuition behind the result is not obvious, but we can look at the last two examples for a comparison. By Proposition 8.2.8, all states are positive recurrent when $p < \frac{1}{2}$,

something which can also be shown directly. We can think of this as there being a "pull" toward 0 and the chain settles in toward a stationary distribution. Compare this with the case $p = \frac{1}{2}$ where there is no such pull and the chain wanders around aimlessly forever. This is reflected in how the stationary distribution "tries to be uniform," but this is not possible on an infinite state space. In general, think of a positive recurrent Markov chain as one that makes a deliberate effort to revisit states, in contrast to a null recurrent chain, which just happens to revisit states without really trying.

Note that Proposition 8.2.8 goes in both directions. Thus, if we can find a stationary distribution of an irreducible chain, we know two things: (1) the stationary distribution is unique and (2) the chain is positive recurrent.

8.2.4 Convergence to the Stationary Distribution

In this section we will state the main limit result for a Markov chain. Although the proof is not beyond the scope of the text, we will not give it; instead we will focus on its interpretation and applications. For a proof, the interested reader may consult, for example, Grimmett and Stirzaker, *Probability and Random Processes* [7]. In Example 8.2.6, we found the limit distribution of the Markov chain, and let us now formally define this concept.

***Definition* 8.2.7.** Let $p_{ij}^{(n)}$ be the n-step transition probabilities of a Markov chain. If there exists a probability distribution $\mathbf{q}$ on S such that

$$p_{ij}^{(n)} \to q_j \quad \text{as} \quad n \to \infty \quad \text{for all } i, j \in S$$

we call $\mathbf{q}$ the *limit distribution* of the Markov chain.

Note that the limit distribution is the same for every initial state $i \in S$. Another way to express this is that the n-step transition matrix $P^{(n)}$ converges to a limit matrix in which all rows are equal. The intuition behind the limit distribution is that q_j describes the probability that the chain is in state j at some late timepoint and that at this time, the chain has "forgotten how it started." We have seen in the examples that a limit distribution does not always exist. If it does, however, it also qualifies as a stationary distribution (see Problem 13).

The more interesting question is whether the converse is true: is the stationary distribution also the limit distribution? This would give a nice computational recipe: in order to find the limit distribution we solve $\boldsymbol{\pi} P = \boldsymbol{\pi}$, which is typically much easier than computing powers of the transition matrix. The following example shows that there may be a problem.

***Example* 8.2.13.** In the ON/OFF system in Example 8.2.4, suppose that $p = q = 1$, that is, that the system always changes. Find the stationary distribution and limit distribution.

The stationary distribution satisfies

$$(\pi_0\ \pi_1)\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = (\pi_0\ \pi_1)$$

which gives $\pi_0 = \pi_1$, so the stationary distribution is $\boldsymbol{\pi} = (\frac{1}{2}, \frac{1}{2})$. In this case it is easy to find the n-step transition probabilities explicitly. For example

$$p_{00}^{(n)} = \begin{cases} 0 & \text{if } n \text{ is even} \\ 1 & \text{if } n \text{ is odd} \end{cases}$$

and similarly for the other three n-step transition probabilities. But this means that the n-step transition probabilities do not converge, and thus there *is no limit distribution*. Recall how a limit distribution forgets where the chain started; in this case, if we start in state 0, we know that the system will be in state 0 at every even timepoint and in state 1 at every odd timepoint, no matter how late. □

Thus, stationary distributions and limit distributions are not necessarily the same. What is the intuition behind these concepts? Suppose that we look at a Markov chain at some late time n. The stationary distribution then gives the long-term proportions of time spent in the different states *up to* time n. The limit distribution, on the other hand, gives the proportions of time spent in the states *at* time n (so we have to think of the Markov chain being run up to time n multiple times). In the previous example, let $n = 1000$. The stationary distribution $(\frac{1}{2}, \frac{1}{2})$ tells us that equal amounts of time have been spent in both states up to time $n = 1000$, regardless of the initial state. However, if we look only at precisely time $n = 1000$, the chain must be in the same state that it started in, and if we run the chain up to time $n = 1000$ from the same initial state, the proportion of time in the other state is 0. For a theoretical result that motivates the interpretation of the stationary distribution, see Problem 21.

The existence of a limit distribution is a desirable property of a Markov chain, since it means that we can get an idea of the distribution over the state space at some late, arbitrary timepoint. It turns out that the problem in the last example is that the chain is *periodic*, in the sense that returns from a state to itself can only occur in an even number of steps.

***Definition* 8.2.8.** The *period* of state i is defined as

$$d(i) = \gcd\{n \geq 1 : p_{ii}^{(n)} > 0\}$$

the greatest common divisor of lengths of cycles through which it is possible to return to i. If $d(i) = 1$, state i is said to be *aperiodic*; otherwise it is called *periodic*.

The concept of a period may not be immediately clear. Let us look at two examples.

***Example* 8.2.14.** Find the periods of the states in the ON/OFF system with $p = q = 1$.

Since $p_{00}^{(n)} > 0$ whenever n is even and 0 otherwise, the set of n such that $p_{00}^{(n)} > 0$ is $\{2, 4, 6, ...\}$ which has greatest common divisor 2. Thus, the period of state 0 is 2, which means that the only possible return paths to state 0 have lengths that are multiples of 2. The period of state 1 is also 2. □

***Example* 8.2.15.** Find the period of the state 1 in the gambler's ruin example (Example 8.2.11).

We have $p_{11}^{(1)} = 0$, $p_{11}^{(2)} > 0$, and $p_{11}^{(3)} > 0$, and since the greatest common divisor of 2 and 3 is 1, we do not need to go any further. State 1 is aperiodic, and we note that this does *not* mean that it can reach itself in one step. See also Problem 14. □

If all states are aperiodic, we call the whole Markov chain aperiodic. It can be shown that periodicity is a class property in the sense that communicating states have the same period. Thus, if we can show that one state is aperiodic in an irreducible chain, the whole chain must be aperiodic. Aperidodicity is the last property we need to be able to say that the stationary distribution and the limit distribution coincide. The following is the main convergence theorem for Markov chains.

Theorem 8.2.9. Consider an irreducible, positive recurrent, and aperiodic Markov chain with stationary distribution $\boldsymbol{\pi}$ and n-step transition probabilities $p_{ij}^{(n)}$. Then

$$p_{ij}^{(n)} \to \pi_j \quad \text{as} \quad n \to \infty$$

for all $i, j \in S$.

An irreducible, positive recurrent, and aperiodic Markov chain is called *ergodic*. We have seen examples of what can go wrong when any of the three conditions are removed. Take away irreducibility, and there may be more than one stationary distribution, take away positive recurrence, and there may be none at all; take away aperiodicity, and there may be a unique stationary distribution that is not the limit distribution.

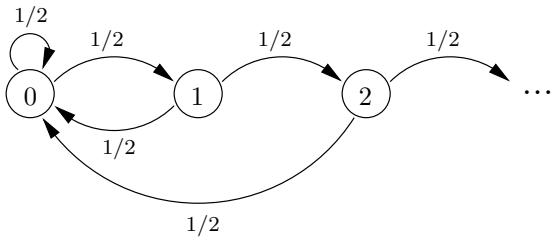
We should also point out that positive recurrence is listed as an assumption. Positive recurrence is an important characteristic of a Markov chain, describing its long-term behavior, but it is typically not checked since it is easier to find the stationary distribution. The practical way to use the theorem is thus to check irreducibility and aperiodicity and then go about solving $\boldsymbol{\pi}P = \boldsymbol{\pi}$. If this can be done, $\boldsymbol{\pi}$ is the limit distribution, and we get positive recurrence for free.

***Example* 8.2.16. (*Success Runs*).** A fair coin is flipped repeatedly, and at time n, we let X_n be the length of the current run of heads. For example, if we get the sequence $HTHHHT$, we have (let $X_0 \equiv 0$)

$$X_1 = 1,\ X_2 = 0,\ X_3 = 1,\ X_4 = 2,\ X_5 = 3,\ X_6 = 0$$

Describe this sequence of *success runs* as a Markov chain and find its limit distribution.

The state space is $S = \{0, 1, 2, ...\}$, and from a state i, transitions are possible to either $i + 1$ or 0, with equal probabilities, giving the following transition graph:



The chain is clearly irreducible, so let us look for the stationary distribution. The equation $\boldsymbol{\pi}P = \boldsymbol{\pi}$ becomes

$$(\pi_0, \pi_1, ...) \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & ... \\ 1/2 & 0 & 1/2 & 0 & 0 & ... \\ 1/2 & 0 & 0 & 1/2 & 0 & ... \\ 1/2 & 0 & 0 & 0 & 1/2 & ... \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} = (\pi_0, \pi_1, ...)$$

and it is easily checked that the solution is

$$\pi_k = \frac{1}{2^{k+1}},\ \ k = 0, 1, 2, ...$$

which we recognize from Section 2.5.3 as a geometric distribution including 0, with success probability $p = \frac{1}{2}$. The last question is whether this also qualifies as the limit distribution and we need to check for aperiodicity. Consider state 0 and note that $p_{00} > 0$, which means that state 0 is aperiodic. Thus, by irreducibility, the entire chain is aperiodic, Theorem 8.2.9 applies, and $\boldsymbol{\pi}$ is the limit distribution. □

In a Markov chain with stationary distribution $\boldsymbol{\pi}$, π_i is the long-term frequency spent in state i; thus, state i is revisited on average every $1/\pi_i$ steps. Now consider another state j. In any sequence of N steps, it is visited on average $N\pi_j$ times, and we get the following nice result, which we state without formal proof. We use the notation E_i for expected value when the initial state is i.

Proposition 8.2.10. Consider an ergodic Markov chain with stationary distribution $\boldsymbol{\pi}$ and choose two states i and j. Let τ_i be the return time to state i, and let N_j be the number of visits to j between consecutive visits to i. Then

$$E_i[\tau_i] = \frac{1}{\pi_i} \quad \text{and} \quad E_i[N_j] = \frac{\pi_j}{\pi_i}$$

Note that by positive recurrence, all the $E_i[\tau_i]$ are finite and hence all the π_i are strictly positive. The $E_i[\tau_i]$ are called the *mean recurrence times*.

***Example* 8.2.17.** Consider Example 8.2.12. Suppose that Ann wins with probability $p = \frac{1}{3}$. **(a)** If Ann just went broke, what is the expected number of rounds until she is broke again? **(b)** If Ann reaches a fortune of $\$5$, how many times can she expect to go broke before reaching that fortune again?

We are asking for $E_0[\tau_0]$ and $E_5[N_0]$, and by Proposition 8.2.10, these are

$$E_0[\tau_0] = \frac{1}{\pi_0} = \frac{1}{1/2} = 2$$

and

$$E_5[N_0] = \frac{\pi_0}{\pi_5} = \frac{1/2}{1/64} = 32$$

□

8.3 RANDOM WALKS AND BRANCHING PROCESSES

In this section we will look at two special cases of Markov chains: *random walks* and *branching processes*. Although they are examples of Markov chains, their properties

are such that the methods we have explored do not reach very far and we instead analyze them by methods that are suited to their particular nature.

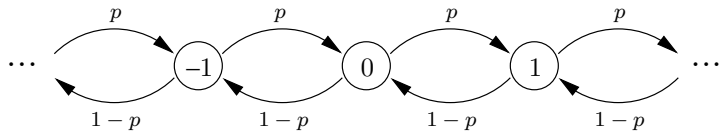
8.3.1 The Simple Random Walk

Many of the examples we looked at in the previous section are similar in nature. For example, the roulette example and the various versions of gambler's ruin have in common that the states are integers and the only possible transitions are one step up or one step down. We now take a more systematic look at such Markov chains, called *simple random walks*. A simple random walk can be described as a Markov chain $\{S_n\}$ that is such that

$$S_n = \sum_{k=1}^{n} X_k$$

where the X_k are i.i.d. such that $P(X_k = 1) = p, P(X_k = -1) = 1 - p$.

The term "simple" refers to the fact that only unit steps are possible; more generally we could let the X_k have any distribution on the integers. The initial state S_0 is usually fixed but could also be chosen according to some probability distribution. Unless otherwise mentioned, we will always have $S_0 \equiv 0$. If $p = \frac{1}{2}$, the walk is said to be *symmetric*. It is clear from the construction that the random walk is a Markov chain with state space $S = \{..., -2, -1, 0, 1, 2, ...\}$ and transition graph



Note how the transition probabilities $p_{i,i+1}$ and $p_{i,i-1}$ do not depend on i, a property called *spatial homogeneity*. We can also illustrate the random walk as a function of time, as was done in Example 1.6.16. Note that this illustrates one particular outcome of the sequence $S_0, S_1, S_2, ...$, called a *sample path*, or a *realization*, of the random walk.

It is clear that the random walk is irreducible, so it has to be either transient, null recurrent, or positive recurrent, and which one it is may depend on p. The random walk is a Markov chain where we can compute the n-step transition probabilities explicitly and apply Proposition 8.2.4. Consider any state i, and note first that

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

since we cannot make it back to a state in an odd number of steps. To make it back in $2n$ steps, we must take n steps up and n steps down, which has probability

$$p_{ii}^{(2n)} = \binom{2n}{n} p^n (1-p)^n = \frac{(2n)!}{n!n!} (p(1-p))^n \tag{8.3.1}$$

and since convergence of the sum of the $p_{ii}^{(2n)}$ is not affected by the values of any finite number of terms in the beginning, we can use an asymptotic approximation of

$n!$, *Stirling's formula*, which says that

$$n! \sim n^n \sqrt{n} e^{-n} \sqrt{2\pi}$$

where "$\sim$" means that the ratio of the two sides goes to 1 as $n \to \infty$. The practical use is that we can substitute one for the other for large n in the sum in Proposition 8.2.4. Thus

$$p_{ii}^{(2n)} \sim \frac{(4p(1-p))^n}{\sqrt{\pi n}}$$

and if $p = \frac{1}{2}$, this equals $1/\sqrt{\pi n}$ and the sum over n is infinite. If instead $p \neq \frac{1}{2}$, $p_{ii}^{(2n)}$ is of the form $x^n/\sqrt{\pi n}$ where $|x| < 1$ and the sum over n converges. This shows that the simple random walk is recurrent if $p = \frac{1}{2}$ and transient if $p \neq \frac{1}{2}$.

The next question is whether the case $p = \frac{1}{2}$ is positive recurrent or null recurrent. Repeating the argument from Example 3.7.9 shows that regardless of whether the first step is up or down, the expected time until return to 0 is infinite. We summarize as follows.

Proposition 8.3.1. The simple random walk is null recurrent if $p = \frac{1}{2}$ and transient if $p \neq \frac{1}{2}$.

In particular, this means that there is never a stationary distribution, so the theory for Markov chains does not give us anything further, but there are still interesting questions regarding the behavior of the random walk. Recall τ_1, the time of the first visit to state 1. From Examples 1.6.17 and 3.7.9, we know that $P_0(\tau_1 < \infty) = 1$ and $E_0[\tau_1] = \infty$, if $p = \frac{1}{2}$. What about other values of p? If $p > \frac{1}{2}$, we must still have $P_0(\tau_1 < \infty) = 1$ but what if $p < \frac{1}{2}$? Let us use recursion and again condition on the first step. With $r = P_0(\tau_1 < \infty)$, we get

$$r = p + (1-p)r^2$$

which has solutions $r = 1$ and $r = p/(1-p)$. We can exclude $r = 1$, since if the probability to reach 1 in finite time equals 1, the same must be true for -1 (more likely to go down than up), but then by symmetry, the probability is also 1 that the walk gets back to 0 again in finite time, which contradicts transience. We get the following result.

Corollary 8.3.2. The probability that the walk ever visits 1 is

$$P_0(\tau_1 < \infty) = \begin{cases} 1 & \text{if } p \geq 1/2 \\ \dfrac{p}{1-p} & \text{if } p < 1/2 \end{cases}$$

Thus, if $p < \frac{1}{2}$, the probability is $p/(1-p)$ that the walk *never* visits the positive axis. If it does visit the positive axis, how far does it get? According to the following result, not very far.

Corollary 8.3.3. The simple random walk with $p < \frac{1}{2}$ visits only finitely many states on the positive axis.

Proof. Let A be the event that the walk visits all positive states and A_r the event that the walk eventually visits state r. Then

$$A = \bigcap_{r=1}^{\infty} A_r$$

where the events A_r are decreasing (must have visited r in order to visit $r+1$). Moreover, in order to visit r, the walk must first visit 1, then 2, and so on, and by Corollary 8.3.2 and Proposition 1.3.5, we obtain

$$P(A) = \lim_{r\to\infty} P(A_r) = \lim_{r\to\infty} \left(\frac{p}{1-p}\right)^r = 0$$

since $p < \frac{1}{2}$. ■

We can continue the argument in the proof. Since the walk makes it up only to some maximum integer, it must eventually leave the positive axis for good. But then it is at -1, and the same argument says that it must eventually hit -2, never to return to -1 again. Thus, for any number, the random walk will eventually stay below it forever, and we have argued that the simple random walk with $p < \frac{1}{2}$ drifts toward $-\infty$. In obvious analogy, if $p > \frac{1}{2}$, the walk drifts toward ∞. In Problem 27, you are asked to compute the probability that the transient random walk ever returns to 0 (right now all we know is that this probability is < 1). In the remaining case, $p = \frac{1}{2}$, the walk must visit all states infinitely many times (why?) and must, like Ahasverus, wander aimlessly forever.

We next turn to expected values. If the walk starts in 0, what is the expected time until its first visit to 1? If $p < \frac{1}{2}$, it must be infinite, since the random variable τ_1 itself may be infinite. If $p = \frac{1}{2}$, can $E_0[\tau_1]$ be finite? If it were, then the expected time to visit -1 would by symmetry also be finite. But then the expected time back to 0 again would also be finite, which contradicts null recurrence. Thus, if $p \leq \frac{1}{2}$, we have $E_0[\tau_1] = \infty$. It remains to investigate what happens when $p > \frac{1}{2}$. We will use

a recursive approach, in the spirit of Example 3.7.9. Let $\mu = E_0[\tau_1]$ and condition on the first step to get the equation

$$\mu = p \times 1 + (1-p)(1+2\mu)$$

which we solve for μ to get $\mu = 1/(2p-1)$. Note, however, that also $\mu = \infty$ is a solution, so we need to argue that this can be ruled out. First note that $\tau_1 = 2n+1$ if the walk is back at 0 at time $2n$, without having visited 1, then goes up to 1 in the following step. The probability to be back at 0 at time $2n$ without any restrictions is $p_{00}^{(2n)}$, and hence

$$P_0(\tau_1 = 2n+1) \leq p \times p_{00}^{(2n)} \sim p\frac{(4p(1-p))^n}{\sqrt{\pi n}}$$

and since $4p(1-p) < 1$, $E_0[\tau_1] = \sum_n (2n+1)P_0(\tau_1 = 2n+1)$ must be finite. We summarize as follows.

Corollary 8.3.4. The expected time until the first visit to 1 is

$$E_0[\tau_1] = \begin{cases} \dfrac{1}{2p-1} & \text{if } p > 1/2 \\ \\ \infty & \text{if } p \leq 1/2 \end{cases}$$

8.3.2 Multidimensional Random Walks

Let us now consider a two-dimensional simple random walk and investigate it with regard to transience/recurrence. First, how do we define it? In one dimension, if the walk is at i, it chooses one of the neighboring points $i-1$ and $i+1$. In two dimensions, there are different ways to define neighboring points. One way is to consider the four neighbors parallel with the axes; another, to consider the four neighboring corner points.

Regardless of definition, we will assume that the walk is symmetric so that each neighbor is chosen with probability $\frac{1}{4}$; otherwise the walk is transient (why?). Also, since all we are interested in is transience/recurrence, it does not matter which definition we choose, since in each case there are four equally likely neighbors (one case is just a 45° rotation of the other). We will choose the second version, choosing the corners with equal probabilities. The reason for this is that we can then view the two-dimensional random walk as two independent one-dimensional random walks, one on each axis. Thus, let

$$\mathbf{S}_n = (S_n^{(1)}, S_n^{(2)}), \quad n = 0, 1, 2, ...$$

where we let $\mathbf{S}_0 = (0,0)$. Then $\mathbf{S}_n = (0,0)$ if and only if $S_n^{(1)} = S_n^{(2)} = 0$ and by independence and Equation (8.3.1),

$$P(\mathbf{S}_n = (0,0)) = P(S_n^{(1)} = 0)P(S_n^{(2)} = 0) = \left(\binom{2n}{n}\frac{1}{4^n}\right)^2$$

and by Stirling's formula

$$P(\mathbf{S}_n = (0,0)) \sim \frac{1}{\pi n}$$

and since $\sum_n \frac{1}{n} = \infty$, we conclude that the simple random walk in two dimensions is recurrent (and must be null recurrent). But look now what happens in three dimensions. By defining

$$P(\mathbf{S}_n = (S_n^{(1)}, S_n^{(2)}, S_n^{(3)}), \;\; n = 0,1,2,...$$

where $S_n^{(1)}, S_n^{(2)}$, and $S_n^{(3)}$ are independent symmetric random walks and letting $\mathbf{S}_0 = (0,0,0)$, we get

$$P(\mathbf{S}_n = (0,0,0)) \sim \frac{1}{(\pi n)^{3/2}}$$

and since $\sum_n \frac{1}{n^{3/2}} < \infty$, we conclude that the walk is now transient. This might be a bit surprising, and there is no immediate intuition for why the walk always returns to the origin in one and two dimensions but not in three. We know that each of the three individual walks returns to 0 infinitely many times, but only finitely many times will they do so simultaneously. We can define the random walk in any number of dimensions and conclude the following.

Corollary 8.3.5. The symmetric simple random walk in n dimensions is recurrent for $n = 1$ and $n = 2$, and transient for $n \geq 3$.

One thing needs to be pointed out. In two dimensions, we argued that the definition of a neighboring point does not matter, since there are four neighbors with either definition. This is not true in dimensions $n \geq 3$. For example, in three dimensions, there is a difference between choosing between the eight corner points and the six points along the axes. We chose the first definition, and the chance of return to the origin ought to be higher with the second, since there are fewer choices in each step. However, it can be shown that the probability of return is still less than one and also this variant of the three-dimensional random walk is transient (see Problem 29).

8.3.3 Branching Processes

To steal a line of British humo(u)r from Grimmett and Stirzaker, [7]: "Besides gambling, many probabilists have been interested in reproduction." In this section we

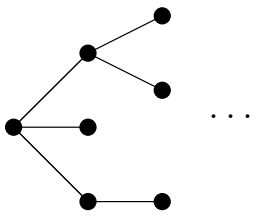


Fig. 8.1 A branching process with $Z_0 = 1, Z_1 = 3$, and $Z_2 = 3$

analyze a simple model for populations that are composed of individuals who reproduce independently of each other. Suppose that we start from one individual, the *ancestor*, who gets a number of children, X, with range $\{0, 1, 2, ...\}$ and pmf p_X. We refer to this pmf as the *offspring distribution*. Each child then reproduces independently according to the offspring distribution and their children reproduce in the same way and so on. The resulting evolving population is an example of a *branching process* (see Figure 8.1).[4] To describe it mathematically, we let Z_n be the number of individuals in the nth generation, and let $Z_0 \equiv 1$. The generation sizes relate to each other as

$$Z_n = \sum_{k=1}^{Z_{n-1}} X_k, \quad n = 1, 2, ... \tag{8.3.2}$$

where the X_k are i.i.d. with pmf p_X. The formula states that in order to get the number of individuals in any generation, we go through the individuals in the preceding generation and sum their numbers of children. Note that in each generation we get a new set of X_k, which if needed can be indicated by a superscript, $X_k^{(n-1)}$, the number of children of the kth individual in the $(n-1)$st generation.

We are interested in the behavior of Z_n and will focus on two issues: extinction and population growth. Note that the process $\{Z_n\}$ is a Markov chain but the transition probabilities are complicated and it is also clear that state 0 is absorbing, so the Markov chain methods that we know will not be of much help. Let us from now on exclude the two uninteresting cases $p_X(0) = 1$ and $p_X(1) = 1$. (Why are these uninteresting?) From Equation (8.3.2) it is clear that

$$Z_n = 0 \;\; \Rightarrow \;\; Z_{n+1} = 0 \tag{8.3.3}$$

in which case the population has gone extinct. Now let E be the event that the population goes extinct *eventually*, that is, the event that some generation size is 0.

[4] In this simple form, it is usually called a *Galton–Watson* process, after the previously mentioned Sir Francis Galton, who worried about extinction of the English nobility, and Henry W. Watson, mathematician, clergyman, and mountaineer, who advocated the use of generating functions to solve the problems. In *les pays francophones*, the name of *Bienaymé* is usually also added, which is only fair since the work of I. J. Bienaymé precedes that of Galton and Watson.

This event can be described as

$$E = \bigcup_{n=1}^{\infty} \{Z_n = 0\}$$

and to find the probability of E, we will use probability generating functions. Thus, let X have pgf G, and let Z_n have pgf G_n. Then $G_1 = G$ (why?) and by Equation (8.3.2) and Proposition 3.11.5, we get the relation

$$G_n(s) = G_{n-1}(G(s)), \;\; n = 2, 3, ... \tag{8.3.4}$$

By Equation (8.3.3), we realize that the events $\{Z_n = 0\}$ are increasing in n, and Corollary 3.11.1 and Proposition 1.3.5 together give

$$P(E) = \lim_{n\to\infty} P(Z_n = 0) = \lim_{n\to\infty} G_n(0)$$

***Example* 8.3.1.** Consider a population of cells that may either die or reproduce, according to the following random variable:

$$X = \begin{cases} 0 & \text{with probability } 1/4 \\ 2 & \text{with probability } 3/4 \end{cases}$$

Find the probability of extinction.

We have

$$G(s) = \frac{1}{4} + \frac{3}{4}s^2$$

which gives $G(0) = \frac{1}{4}$, the probability that extinction occurs already in the first generation. For G_2, we use Equation (8.3.4) and obtain

$$G_2(s) = G(G(s)) = \frac{1}{4} + \frac{3}{4}\left(\frac{1}{4} + \frac{3}{4}s^2\right)^2$$

which gives $G_2(0) = \frac{19}{64}$. Already here we realize that it will be hard to find $P(E)$ in this way; remember that we need to find the limit of $G_n(0)$ as $n \to \infty$ and a pattern for the sequence does not readily emerge. We give up. □

The branching process in this example is about as simple as they come, and yet it is virtually impossible to find the extinction probability as the limit of $G_n(0)$. Luckily, there is a much quicker way.

Proposition 8.3.6. Consider a branching process where the offspring distribution has pgf G, and let E be the event of extinction. Then $q = P(E)$ is the smallest solution in $[0, 1]$ to the equation $s = G(s)$.

Proof. Let us first show that q is a solution and then that it must be smaller than any other solution. Condition on the number of children, X, of the ancestor and note that if $X = k$, then there are k independent branching processes that must go extinct, which has probability q^k (true also for $k = 0$). This gives the following observation

$$q = \sum_{k=0}^{\infty} P(E|X = k)P(X = k) = \sum_{k=0}^{\infty} q^k P(X = k) = G(q)$$

and we see that the extinction probability solves the equation $s = G(s)$.

Suppose that there is another solution $r \in [0, 1]$. By Problem 133 in Chapter 3, $G(s)$ is increasing, and since $r \geq 0$, we get

$$r = G(r) \geq G(0)$$

Applying G again gives $r = G(r) \geq G(G(0)) = G_2(0)$ and repeating the argument gives

$$r \geq G_n(0) \quad \text{for all } n$$

But since $G_n(0) \to q$, we also get $r \geq q$, and hence q is the smallest solution to the equation $s = G(s)$. ■

***Example* 8.3.2.** In Example 8.3.1 we get the equation

$$s = \frac{1}{4} + \frac{3}{4}s^2$$

which has solutions $\frac{1}{3}$ and 1. Thus, the probability of extinction is $\frac{1}{3}$. □

Note that any pgf has the property $G(1) = 1$, and hence $s = 1$ is always a solution to the equation $s = G(s)$.[5] Let us now turn to the question of population growth. More specifically, we will find the mean and variance of the nth-generation size Z_n. Equation (8.3.2) is central, and we can apply Corollary 3.11.6 to obtain the following result.

[5]Watson, whose work was published in 1875, found the solution $s = 1$ but overlooked the fact that there could be more solutions and erroneously concluded that extinction is always inevitable. It took yet another half-century and Danish ingenuity to completely solve the problem, done by J. F. Steffensen in 1930.

Proposition 8.3.7. Consider a branching process where the offspring distribution has mean μ and variance σ^2. Then

$$E[Z_n] = \mu^n$$

$$\text{Var}[Z_n] = \begin{cases} n\sigma^2 & \text{if } \mu = 1 \\ \dfrac{\sigma^2(\mu^n - 1)\mu^{n-1}}{\mu - 1} & \text{if } \mu \neq 1 \end{cases}$$

Proof. For the mean, repeatedly apply Corollary 3.11.6 to Equation (8.3.2) and get

$$E[Z_n] = E[Z_{n-1}]\mu = E[Z_{n-2}]\mu^2 = \cdots = \mu^n$$

since $E[Z_0] = 1$. We leave it as an exercise to verify that the expression for the variance satisfies Corollary 3.11.6. ■

The proposition thus tells us that

$$E[Z_n] \begin{cases} \to 0 & \text{if } \mu < 1 \\ \equiv 1 & \text{if } \mu = 1 \\ \to \infty & \text{if } \mu > 1 \end{cases}$$

and that

$$\text{Var}[Z_n] \begin{cases} \to 0 & \text{if } \mu < 1 \\ \to \infty & \text{if } \mu = 1 \\ \to \infty & \text{if } \mu > 1 \end{cases}$$

This suggests that the population always goes extinct if $\mu < 1$ since the mean and variance both go to 0 and Z_n itself can take on only integer values. It is less clear what happens in the other two cases, but the following result gives the answer.

Proposition 8.3.8. Consider a branching process with mean number of children μ. Then

$$\mu \leq 1 \Rightarrow P(E) = 1$$
$$\mu > 1 \Rightarrow P(E) < 1$$

Proof. If $p_X(0) = 0$, then $\mu > 1$ and $P(E) = 0$. Now suppose that $p_X(0) > 0$, and recall from Section 3.11 that $\mu = G'(1)$, the slope at the point $s = 1$. By Problem 133 in Chapter 3, $G(s)$ is convex and increasing, and since it has to increase from $G(0) = p_X(0) > 0$ to $G(1) = 1$, it must intersect the line $y = s$ if $\mu > 1$, and the intersection is the extinction probability, which is < 1. If instead $\mu \leq 1$, there can be no such intersection, and the extinction probability is 1. See Figure 8.2 for the two possible cases. Note that this also shows that there can never be more than two solutions in $[0, 1]$ to the equation $s = G(s)$. ■

This result is quite remarkable. It says that whether extinction occurs for certain depends only on the mean number of children, and if this is less than or equal to 1, there will be extinction sooner or later. If $\mu > 1$, extinction may be avoided and the probability of this is found by solving the equation $s = G(s)$. The cases $\mu > 1$, $\mu = 1$, and $\mu < 1$ are called the *supercritical*, *critical*, and *subcritical*, respectively. Although extinction is certain in the last two, they exhibit some differences in behavior that motivates the distinction (see Problem 30).

***Example* 8.3.3.** An individual with a contagious disease enters a large city. Suppose that he passes the disease on to a number of people, who in turn pass it on to others and so on. Suppose that each individual remains contagious for one day and in this day interacts with a number of people that has a Poisson distribution with mean 10 and that for each person, the probability of infection is p. **(a)** For which values of p does the disease eventually die out? **(b)** If $p = 0.2$, what is the probability that the disease still exists in the population on day 2? **(c)** For $p = 0.2$, what is the probability that the disease eventually disappears?

The number of infected "children" of an individual has a Poisson distribution with mean $10p$, so the disease eventually dies out if $p \leq 0.1$. For part (b), we need to

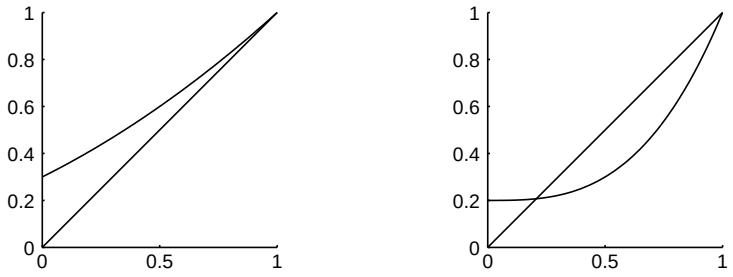


Fig. 8.2 Plots of the pgf $G(s)$ of the offspring distribution and the line $y = s$, in the cases $\mu = G'(1) \leq 1$ (left) and $\mu = G'(1) > 1$ (right).

compute $P(Z_2 = 0) = G_2(0)$, and since

$$G(s) = e^{2(s-1)}$$

we get

$$P(Z_2 = 0) = e^{2(e^{-2}-1)} \approx 0.18$$

so the probability that the disease still exists is 0.82. For (c), we need numerically to solve the equation

$$s = e^{2(s-1)}$$

which gives $q \approx 0.2$. The spread of a disease may be adequately modeled by a branching process in its early stages but as time goes on, such a model becomes less realistic. There are several reasons for this, which we leave for the reader to ponder.

□

The final question is what happens if there is no extinction. As it turns out, the only other possibility is for Z_n to go to infinity. We will not give a strict proof but refer to an intuitive argument. Suppose that Z_n does not go to infinity. Then there is some integer K such that Z_n drops below K infinitely many times. But each time it does, there is a chance that it goes extinct before the next time it drops below K. The probability of this is at least $p_X(0)^K$, since the "worst case" is when the population goes extinct already in the next generation. But this means that the population must become extinct sooner or later (think geometric trials here), and we have argued that if Z_n does not go to infinity, the population must become extinct. In terms of probabilities

$$P(Z_n \not\to \infty) \leq P(E)$$

and since the reversed inequality obviously also holds (if there is extinction, the populations size cannot go to infinity), the probabilities are equal. We have argued for the following result.

Proposition 8.3.9. Consider a branching process with extinction probability q. Then

$$Z_n \to \begin{cases} 0 & \text{with probability } q \\ \infty & \text{with probability } 1-q \end{cases}$$

as $n \to \infty$.

Note that "$Z_n \to 0$" is just another way of saying "extinction," since Z_n takes on only integer values. We can describe this result by saying that the sequence of random variables $Z_1, Z_2, ...$ converges to a random variable Z as $n \to \infty$, where Z is either 0 or ∞ with probabilities q and $1-q$, respectively. To be perfectly strict, we need to

say that this convergence takes place *with probability* 1 since we can describe other types of sequences. For example, in a population of dividing cells let us say that $Z_1 = 2$. It could then happen that one divides and the other dies so that $Z_2 = 2$ also. If the same thing happens again, $Z_3 = 2$, and we can continue, to describe a sequence $Z_1, Z_2, ...$ where $Z_n = 2$ for all n. However, it can be shown that this sequence together with all other sequences that do not converge to 0 or ∞ belong to an event that has probability 0. This is an example of convergence almost surely, mentioned in footnote 1 in Section 4.2.

8.4 CONTINUOUS-TIME MARKOV CHAINS

We have studied Markov chains in discrete time and will now turn to their counterpart in continuous time. This means that the chains stays in each state a random time that is a continuous random variable with a distribution that may depend on the state. The state of the chain at time t is denoted $X(t)$, where t ranges over the nonnegative real numbers. In addition to having the Markov property for the jumps, we also want the jumps to be independent of how long a time that is spent in a specific state, and in order to achieve this, we recall that there is only one continuous distribution that would ensure this property: the exponential distribution.[6] We state the following definition.

***Definition* 8.4.1.** Let $\{X(t),\ t \geq 0\}$ be a collection of discrete random variables taking values in some set S and that evolves in time as follows:

(a) If the current state is i, the time until the state is changed has an exponential distribution with parameter $\lambda(i)$.

(b) When state i is left, a new state $j \neq i$ is chosen according to the transition probabilities of a discrete-time Markov chain.

Then $\{X(t)\}$ is called a *continuous-time Markov chain*.

Thus, a continuous-time Markov chain $\{X(t)\}$ is composed of a discrete-time Markov chain $\{X_n\}$, the *jump chain*, for the transitions and exponential random variables for the *holding times*. Recall that the holding times in a discrete Markov chain are geometric, the discrete counterpart of the exponential (see Section 2.6), so this is a natural assumption. The $\lambda(i)$ are called the *holding-time parameters*. Note that

[6] We are leaving out some subtle technicalities here that the interested reader may find, for example, in Grimmett and Stirzaker, *Probability and Random Processes* [7].

the state space is still finite or countably infinite; the discrete/continuous distinction refers to how time is measured. Let us also mention that sometimes the term *Markov process* is used in continuous time. We will not formally state the Markov property but intuitively it says that conditioned on the current state and time, where and when the chain jumps next is independent of the complete history of the chain.

Our construction also ensures that the process is *time-homogeneous*, that is, the probability $P(X(s+t) = j \,|X(s) = i)$ depends only on time through the difference $(s + t) - s = t$, and we can define the transition probabilities as

$$p_{ij}(t) = P(X(t) = j \,|X(0) = i)$$

the probability that the chain is in state j, t time units after having been in state i. For each t, we then get a transition matrix $P(t)$ with entries $p_{ij}(t)$, $i, j \in S$, which has the following properties.

Proposition 8.4.1. Let $P(t)$ be the transition matrix for a continuous-time Markov chain with state space S. Then

(a) $P(0) = I$, the identity matrix

(b) $\sum_{j\in S} p_{ij}(t) = 1$, for all $i \in S$ and $t \geq 0$

(c) $P(s + t) = P(s)P(t)$(Chapman–Kolmogorov equations)

When you talk to mathematicians, make sure to refer to the set $\{P(t),\ t \geq 0\}$ as a *stochastic semigroup*.

Proof. Parts (a) and (b) are obvious. For (c), consider an element $p_{ij}(s + t)$ of $P(s + t)$ and condition on an intermediate state k at time s to obtain

$$\begin{aligned} p_{ij}(s+t) &= \sum_{k\in S} P_i(X(s+t) = j \,|X(s) = k)P_i(X(s) = k|X_0 = i) \\ &= \sum_{k\in S} p_{ik}(s)p_{kj}(t) \end{aligned}$$

which is the (i, j)th entry in the matrix $P(s)P(t)$. ■

One problem is that $P(t)$ is usually difficult or impossible to compute, in the same way that $P^{(n)}$ may be in the discrete case. In the discrete case, however, we know that $P^{(n)} = P^n$, so all the information we need is contained in the one-step transition

matrix P. In the continuous case there is no analog of "one step," so we need to proceed differently in search of a more compact description.

Let the jump chain have transition probabilities p_{ij} for $i \neq j$ and consider the chain in a state i. The holding time is $\exp(\lambda(i))$ and when it leaves, the chain jumps to state j with probability p_{ij}. Now, if we consider the chain only when it is in state i and disregard everything else, we can view the jumps from i as a Poisson process with rate $\lambda(i)$. For any other state j, the jumps from i to j is then a thinned Poisson process with rate $\lambda(i)p_{ij}$. Thus, for any pair of states i and j, we can define the *transition rate* between i and j as

$$\gamma_{ij} = \lambda(i)p_{ij}$$

In addition to these, we also let

$$\gamma_{ii} = -\sum_{j \neq i} \gamma_{ij}$$

and define the *generator* as the matrix G whose (i, j)th entry is γ_{ij}. Note that once the γ_{ij} have been inserted, the diagonal elements γ_{ii} are chosen such that G has row sums equal to 0. The generator completely describes the Markov chain, since if we are given G, we can retrieve the holding-time parameters as

$$\lambda(i) = -\gamma_{ii}, \quad i \in S$$

and the jump probabilities as

$$p_{ij} = -\frac{\gamma_{ij}}{\gamma_{ii}}, \quad j \neq i$$

Note that $p_{ii} = 0$ for all $i \in S$ since the p_{ij} give the probability distribution when the chain leaves a state and there can be no jumps from a state to itself (see also Problem 36). Let us look at a few examples.

***Example* 8.4.1.** An ON/OFF system stays OFF for a time that is $\exp(\lambda)$ and ON for a time $\exp(\mu)$ (μ does not denote the mean here). Describe the system as a continuous-time Markov chain.

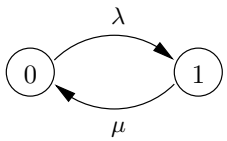
The holding-time parameters are λ and μ, and the only possible jumps are from 0 (OFF) to 1 (ON) and vice versa. Thus we have

$$\gamma_{01} = \lambda, \;\; \gamma_{10} = \mu$$

and after filling in the diagonal elements, we get generator

$$G = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$$

We can also describe the system in a graph as follows:



This is similar to how we described discrete-time Markov chains but the numbers on the arrows are now rates, not probabilities. The jump chain has transition matrix

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

□

***Example* 8.4.2.** A continuous-time Markov chain on state space $\{1, 2, 3\}$ has generator

$$G = \begin{pmatrix} -6 & 2 & 4 \\ 1 & -2 & 1 \\ 3 & 1 & -4 \end{pmatrix}$$

Suppose that the chain is in state 1. What is the expected time until it leaves, and what is the probability that it next jumps to state 2?

The holding-time parameter in state 1 is $\lambda(1) = -\gamma_{11} = 6$, so the expected holding time is $\frac{1}{6}$. The probability to jump to state 2 is

$$p_{12} = -\frac{\gamma_{12}}{\gamma_{11}} = -\frac{2}{-6} = \frac{1}{3}$$

□

The generator now plays the role that the transition matrix did in the discrete case, and a logical question is how G relates to $P(t)$. The following proposition gives the answer, where $P'(t)$ denotes the matrix of the derivatives $p'_{ij}(t)$.

Proposition 8.4.2. The transition matrix $P(t)$ and generator G satisfy the *backward equations*

$$P'(t) = GP(t), \quad t \geq 0$$

and *forward equations*

$$P'(t) = P(t)G, \quad t \geq 0$$

If we spell out the backward equations elementwise, we get

$$p'_{ij}(t) = \sum_{k \in S} \gamma_{ik} p_{kj}(t), \quad i, j \in S, t \geq 0$$

and the forward equations

$$p'_{ij}(t) = \sum_{k \in S} p_{ik}(t)\gamma_{kj}, \quad i, j \in S, t \geq 0$$

Proof. We refer to an intuitive argument and leave out the technical details. Consider the probability $p_{ij}(t+h) = P_i(X(t+h) = j)$. The Chapman–Kolmogorov equations give

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

and if h is small, we have, with T_j denoting the holding time in state j

$$\begin{aligned} p_{jj}(h) &= P(X(h) = j | X(0) = j) \approx P(T_j > h) = e^{-\lambda(j)h} \\ &\approx 1 - \lambda(j)h = 1 + \gamma_{jj}h \end{aligned}$$

The intuition behind this is that if h is small and the chain is in state j at times 0 and h, most likely nothing happened in the interval $(0, h)$. With a similar argument, if there is a jump in $(0, h)$, we neglect the possibility of more than one jump and obtain

$$p_{kj}(h) \approx \gamma_{kj}h, \quad k \neq j$$

This gives

$$\begin{aligned} p_{ij}(t+h) &\approx p_{ij}(t)(1 + \gamma_{jj}h) + \sum_{k \neq j} p_{ik}(t)\gamma_{kj}h \\ &= p_{ij}(t) + h \sum_{k \in S} p_{ik}(t)\gamma_{kj} \end{aligned}$$

which gives

$$\frac{p_{ij}(t+h) - p_{ij}(t)}{h} = \sum_{k \in S} p_{ik}(t)\gamma_{kj}$$

and letting $h \downarrow 0$ gives the forward equations, with a similar type of argument for the backward equations. ■

It turns out that the forward equations are usually easier to solve but do not always exist (a fact that is not revealed by our intuitive argument above). In all examples and applications we consider, they do, however, exist. It is usually difficult to solve the backward and forward equations and only in simple cases can we easily find the explicit form of $P(t)$. In Problem 34, you are asked to find the solution for the simple ON/OFF system.

Since $P(0) = I$, the backward and forward equations also suggest a way to obtain the generator from $P(t)$ according to

$$G = P'(0) \tag{8.4.1}$$

8.4.1 Stationary Distributions and Limit Distributions

Just as in the discrete case, we are interested in asymptotic behavior, which is described by the limit of $P(t)$ as $t \to \infty$ and also as in the discrete case, we would like to do this via stationary distributions instead of direct calculations. A limit distribution in the continuous case is the obvious analog of the discrete case: a distribution $\mathbf{q}$ such that

$$p_{ij}(t) \to q_j, \quad \text{as } t \to \infty \quad \text{for all } i, j \in S$$

How should we define a stationary distribution? In the discrete case, it is defined through the relation $\boldsymbol{\pi} = \boldsymbol{\pi}P$, but in the continuous case there is no P. However, since a stationary distribution "stays forever," we also have $\boldsymbol{\pi} = \boldsymbol{\pi}P^{(n)}$ for all n, and we can imitate this in the continuous case.

***Definition* 8.4.2.** Consider a continuous-time Markov chain with transition matrix $P(t)$. A probability distribution $\boldsymbol{\pi}$ which is such that

$$\boldsymbol{\pi}P(t) = \boldsymbol{\pi} \quad \text{for all } t \geq 0$$

is called a *stationary distribution* of the chain.

The intuition is the same as in the discrete case; the probability π_j is the proportion of time spent in state j in the long run. Since we have pointed out how difficult it typically is to find $P(t)$, the definition does not give a computational recipe. Instead, first differentiate with respect to t on both sides in the definition to obtain

$$\frac{d}{dt}(\boldsymbol{\pi}P(t)) = \boldsymbol{\pi}P'(t) = \frac{d}{dt}(\boldsymbol{\pi}) = \mathbf{0}, \;\; t \geq 0$$

since $\boldsymbol{\pi}$ does not depend on t. In particular, with $t = 0$, Equation (8.4.1) gives $P'(0) = G$ and we have shown the following.

Corollary 8.4.3. The stationary distribution satisfies the equation

$$\boldsymbol{\pi}G = \mathbf{0}$$

where $\mathbf{0}$ is a vector of zeros.

Elementwise, the equations are

$$\sum_{i \in S} \gamma_{ij}\pi_i = 0, \;\; j \in S$$

and as in the discrete case, we have the additional condition that the entries in $\boldsymbol{\pi}$ sum to 1.

***Example* 8.4.3.** Consider the ON/OFF system in Example 8.4.1. Find the stationary distribution.

The equation $\boldsymbol{\pi}G = \mathbf{0}$ is

$$(\pi_0,\ \pi_1)\begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} = (0,\ 0)$$

which gives the first equation

$$-\pi_0\lambda + \pi_1\mu = 0$$

which in turn gives

$$\pi_1 = \frac{\lambda}{\mu}\pi_0$$

and the condition $\pi_0 + \pi_1 = 1$ gives

$$1 = \pi_0\left(1 + \frac{\lambda}{\mu}\right)$$

and we get stationary distribution

$$(\pi_0,\ \pi_1) = \left(\frac{\mu}{\mu+\lambda}, \frac{\lambda}{\mu+\lambda}\right)$$

Note how $\pi_0 > \pi_1$ if $\mu > \lambda$, that is, when the transition rate is higher from 1 to 0 than vice versa. Also note that

$$(\pi_0, \pi_1) = \left(\frac{1/\lambda}{1/\lambda + 1/\mu},\ \frac{1/\mu}{1/\lambda + 1/\mu}\right)$$

which is intuitively appealing since the expected times spent in states 0 and 1 are $1/\lambda$ and $1/\mu$, respectively. Thus, in the long run, the proportions of time spent in states 0 and 1 are π_0 and π_1.

Recall that the jump chain has stationary distribution $(\frac{1}{2}, \frac{1}{2})$, which tells us that the jump chain on average visits 0 and 1 equally many times but does not take holding times into account. □

The existence of a stationary distribution is again closely related to the concepts of irreducibility and positive recurrence. Irreducibility is only a property of how the

states communicate and has nothing to do with holding times, so we call a continuous-time Markov chain irreducible if its jump chain is irreducible. As for recurrence and transience, they are defined in the analogous way, letting

$$S_i = \inf\{t : X(t) = i\}$$

where $S_i = \infty$ if i is never visited. The only difference from the discrete case is that S_i is now a continuous random variable and the following definition is a direct analog.

***Definition* 8.4.3.** If $P_i(S_i < \infty) = 1$, state i is called *recurrent* and if $P_i(S_i < \infty) < 1$, state i is called *transient*. If i is recurrent and $E_i[S_i] < \infty$, i is called *positive recurrent*; otherwise it is called *null recurrent*.

We use the notation S for "sum," since S_i is the sum of the holding times in all states visited before reaching i. We keep the notation τ_i for the return times in the jump chain. Thus, if the holding time in state k is T_k, then S_i and τ_i relate as

$$S_i = \sum_{n=0}^{\tau_i - 1} T_{X_n} \tag{8.4.2}$$

Now suppose that state i is recurrent in the jump chain $\{X_n\}$. This means that τ_i presented above is finite, and since also the T_k are finite, S_i must be finite and i is recurrent also in $\{X(t)\}$. Thus, if the jump chain is recurrent, so is the continuous-time chain $\{X(t)\}$. When it comes to positive recurrence, things are more complicated, as the following example shows.

***Example* 8.4.4.** Consider the following continuous-time version of the success run chain from Example 8.2.16. The holding time parameters are $\lambda(k) = 1/2^k$ for $k = 0, 1, 2...$, and the success run chain functions as the jump chain, with the exception that $p_{01} = 1$. Show that the jump chain $\{X_n\}$ is positive recurrent but $\{X(t)\}$ is not.

The state space is $S = \{0, 1, 2, ...\}$ and the transition matrix of the jump chain differs from that of the success run chain only in that it has $p_{01} = 1$ instead of $p_{00} = p_{01} = \frac{1}{2}$. It is easy to find the stationary distribution (see Problem 16) and this shows that $\{X_n\}$ is positive recurrent.

Now consider state 0. Equation (8.4.2) is extra simple since the jump chain increases by unit steps until it drops back to 0, which gives $X_n = n$ for $n < \tau_0$, and we get

$$S_0 = \sum_{n=0}^{\tau_0 - 1} T_n$$

Since the state 0 is recurrent in $\{X_n\}$, it is recurrent also in $\{X(t)\}$. However, we will show that state 0 is null recurrent in $\{X(t)\}$ and thus proceed to compute the expected value of S_0. To do so, we condition on τ_0 and note that the range of τ_0 is $\{2, 3, ...\}$. If $\tau_0 = k$, S_0 is the sum of $T_1, ..., T_{k-1}$ and hence

$$E_0[S_0|\tau_0 = k] = \sum_{n=0}^{k-1} E[T_n] = \sum_{n=0}^{k-1} 2^n = 2^k - 1$$

and the distribution of τ_0 is

$$P_0(\tau_0 = k) = \frac{1}{2^{k-1}}, \;\; k = 2, 3, ...$$

since $\tau_0 = k$ precisely when there are $k - 1$ successes followed by a failure, where k must be at least 2. We get

$$\begin{aligned} E_0[S_0] &= \sum_{k=2}^{\infty} E_0[S_0|\tau_0 = k]P_0(\tau_0 = k) \\ &= \sum_{k=2}^{\infty} (2^k - 1)\frac{1}{2^{k-1}} \;=\; \sum_{k=2}^{\infty} \left(2 - \frac{1}{2^{k-1}}\right) = \infty \end{aligned}$$

which means that state 0 is null recurrent in the continuous-time chain $X(t)$. By irreducibility, the entire continuous-time chain is null recurrent. The problem is that even though the jump chain is positive recurrent, the holding times get so long that the continuous chain becomes null recurrent. □

If the holding times are instead very short, it is possible that $\{X_n\}$ is null recurrent but $\{X(t)\}$ is positive recurrent (see Problem 38). It is also possible to construct examples where the jump chain is transient but the continuous-time chain has a stationary distribution in the sense of a solution to $\boldsymbol{\pi} G = \mathbf{0}$. However, in such examples we get the unpleasant property of infinitely many jumps in finite time, and to rule out such anomalies, we always assume that the jump chain is recurrent. The following is the continuous-time analog of Proposition 8.2.8.

Proposition 8.4.4. Consider an irreducible continuous-time Markov chain with a recurrent jump chain. Then

A stationary distribution $\boldsymbol{\pi}$ exists $\Leftrightarrow$ $\{X(t)\}$ is positive recurrent

The stationary distribution is unique and has $\pi_j > 0$ for all $j \in S$.

As in the discrete case, positive recurrence is an important concept for describing the behavior of the chain, but it is typically not checked directly. Instead, we look for a solution to $\boldsymbol{\pi}G = \mathbf{0}$.

We next turn to the question of convergence to the stationary distribution. In the discrete case this was complicated by possible periodicity, but in the continuous case we have no unit step size, and thus the concept of period does not exist. We can state the convergence theorem already.

Theorem 8.4.5. Consider an irreducible, continuous-time Markov chain with a recurrent jump chain, stationary distribution $\boldsymbol{\pi}$, and transition probabilities $p_{ij}(t)$. Then

$$p_{ij}(t) \to \pi_j \quad \text{as} \quad t \to \infty$$

for all $i, j \in S$.

As in the discrete case, a continuous-time Markov chain that satisfies the assumptions in the theorem is called *ergodic*. The obvious analog of Proposition 8.2.10, regarding mean recurrence times and mean number of visits to intermediate states, holds in the continuous case as well. Finally, we refer to Problem 38 to see how the stationary distributions for $\{X(t)\}$ and its jump chain relate to each other.

***Example* 8.4.5.** Consider the ON/OFF system from Example 8.4.3, where the jump chain has stationary distribution $(\frac{1}{2}, \frac{1}{2})$ and the continuous-time chain has stationary distribution $(\mu/(\lambda+\mu), \lambda/(\lambda+\mu))$. These describe the proportion of jumps and the proportion of time, respectively, spent in each state in the long run. However, only the continuous-time chain also has a limit distribution. If the system starts in state 0, it forgets how it started if we consider it in real time, but not if we count the jumps, as we saw in Example 8.2.13. □

8.4.2 Birth–Death Processes

In this section we will examine continuous-time analogs of random walks. Thus we will consider integer-valued, continuous-time Markov chains that can only step up or down, so the only generator entries that can be positive (but do not have to be) are $\gamma_{i,i-1}$ and $\gamma_{i,i+1}$. We also restrict the state space to $S = \{0, 1, 2, \ldots\}$, the nonnegative integers. Such Markov chains are called *birth–death processes*. Let us explain why in an example.

***Example* 8.4.6.** Consider a population of cells. Each cell lives for a time that is $\exp(\alpha)$ and then either splits into two new cells with probability p or dies with probability $1-p$, independently of all other cells. Let $X(t)$ be the number of cells at time

t and describe this as a continuous-time Markov chain.

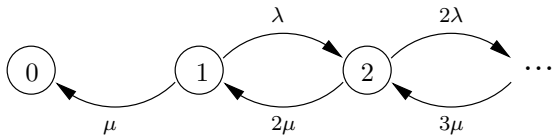
The state space is $S = \{0, 1, 2, ...\}$. If there are i cells, the next change comes after a time that is the minimum of i lifetimes that are independent and $\exp(\alpha)$. By Example 3.10.3, we thus have holding-time parameters

$$\lambda(i) = i\alpha, \;\; i = 0, 1, 2, ...$$

where $\lambda(0) = 0$ means that state 0 is absorbing. This gives the transition rates

$$\gamma_{i,i+1} = i\alpha p \;\; \text{and} \;\; \gamma_{i,i-1} = i\alpha(1-p), \;\; i = 1, 2, ...$$

and it is common to define the *birth rate* $\lambda_i = \gamma_{i,i+1}$ and *death rate* $\mu_i = \gamma_{i,i-1}$ [do not confuse λ_i and $\lambda(i)$]. Since the birth and death rates are linear in i, this is called a *linear* birth–death process. It is also customary to denote $\lambda = \alpha p$ and $\mu = \alpha(1-p)$, the *individual* birth and death rates. The transition graph is then



and the generator

$$G = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \dots \\ \mu & -(\lambda+\mu) & \lambda & 0 & 0 & \dots \\ 0 & 2\mu & -2(\lambda+\mu) & 2\lambda & 0 & \dots \\ 0 & 0 & 3\mu & -3(\lambda+\mu) & 3\lambda & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

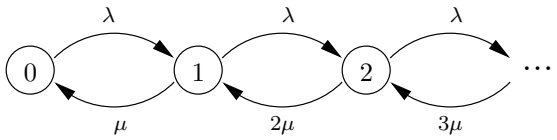
The jump chain is the simple random walk, which we know is transient if $p > \frac{1}{2}$, the only case in which absorption in 0 can be avoided. □

***Example* 8.4.7.** Consider a population where the individual death rate is μ and there are no births. Instead, there is constant immigration into the population according to a Poisson process with rate λ. Describe the process, determine when a limit distribution exists, and find what it is.

The birth and death rates are

$$\lambda_i = \lambda, \;\; \mu_i = i\mu, \;\; i = 0, 1, ...$$

and we have the following transition graph:



The chain is clearly irreducible, so we need only to look for a stationary distribution, which, if it exists, is also the limit distribution. The equation $\boldsymbol{\pi} G = \mathbf{0}$ becomes

$$(\pi_0, \pi_1, ...) \begin{pmatrix} -\lambda & \lambda & 0 & 0 & 0 & \dots \\ \mu & -(\lambda+\mu) & \lambda & 0 & 0 & \dots \\ 0 & 2\mu & -(\lambda+2\mu) & \lambda & 0 & \dots \\ 0 & 0 & 3\mu & -(\lambda+3\mu) & \lambda & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} = (0, 0, ...)$$

which gives the first equation

$$-\lambda\pi_0 + \mu\pi_1 = 0$$

which gives

$$\pi_1 = \frac{\lambda}{\mu}\pi_0$$

The next equation

$$\lambda\pi_0 - (\lambda+\mu)\pi_1 + 2\mu\pi_2 = 0$$

gives, after some algebra

$$\pi_2 = \frac{\lambda^2}{2\mu^2}\pi_0$$

The remaining equations all look the same:

$$\lambda\pi_{n-1} - (\lambda + n\mu)\pi_n + (n+1)\mu\,\pi_{n+1}, \;\; n = 2, 3, ...$$

and it is easy to check that the general solution is

$$\pi_n = \frac{\rho^n}{n!}\pi_0, \; n = 0, 1, 2, ...$$

where $\rho = \lambda/\mu$. The condition $\sum_n \pi_n = 1$ gives

$$1 = \pi_0 \sum_{n=0}^{\infty} \frac{\rho^n}{n!} = \pi_0 e^{\rho}$$

which gives stationary distribution

$$\pi_n = e^{-\rho}\frac{\rho^n}{n!}, \; n = 0, 1, 2, ...$$

which we recognize as a Poisson distribution with mean ρ. Note that the stationary distribution always exists. The intuitive reason for this is that the larger the population

becomes, the more likely that the next event is a death rather than an immigration. Even if the immigration rate λ is enormous compared to the individual death rate μ, sooner or later there will be so many individuals that deaths start to compensate for immigration. The jump chain has transition probabilities

$$p_{i,i+1} = -\frac{\gamma_{i,i+1}}{\gamma_{ii}} = \frac{\lambda}{\lambda + i\mu}$$

and $p_{i,i-1} = i\mu/(\lambda + i\mu), i = 0, 1, ...$ and also note that the larger the population becomes, the more frequent the events, since the expected holding time in state i is $1/(\lambda + i\mu)$. □

Since the structure of the generator is so simple in a birth–death process, it is possible to find a general formula for the stationary distribution. The general form of the generator is

$$G = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \dots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & 0 & \dots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 & \dots \\ 0 & 0 & \mu_3 & -(\lambda_3 + \mu_3) & \lambda_3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

from which it is easily seen that the equation $\boldsymbol{\pi} G = \mathbf{0}$ gives the first equation

$$-\lambda_0 \pi_0 + \mu_1 \pi_1 = 0$$

which gives

$$\pi_1 = \frac{\lambda_0}{\mu_1} \pi_0$$

and the remaining equations

$$\lambda_{n-2}\pi_{n-2} - (\lambda_{n-1} + \mu_{n-1})\pi_{n-1} + \mu_n \pi_n = 0, \;\; n = 2, 3, ...$$

and it is easy to check that these are satisfied by

$$\pi_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} \pi_0, \;\; n = 1, 2, ... \tag{8.4.3}$$

Summing over n now yields that a stationary distribution exists if and only if

$$1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} < \infty$$

and the stationary distribution is then given by

$$\pi_0 = \left(1 + \sum_{n=1}^{\infty} \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n} \right)^{-1}$$

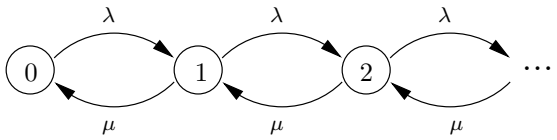
and the remaining π_n by Equation (8.4.3). Also see Problem 42 for a nice interpretation of the equation $\boldsymbol{\pi} G = \mathbf{0}$.

8.4.3 Queueing Theory

A particular class of birth–death processes arise in models for certain service systems. Let us introduce the subject of *queueing theory* with an example.

***Example* 8.4.8. (*The* $M/M/1$ *queue*).** Customers arrive according to a Poisson process with rate λ to a service station with one server. Service times are i.i.d. exponential with rate μ and independent of the arrivals (note that μ does not denote the mean here; the mean service time is $1/\mu$). If the server is busy, incoming customers wait in line and as soon as a service is completed, the next begins. Describe the system as a birth–death process, determining when it has a stationary distribution and what it is.

We let the state $X(t)$ be the number of customers in the system (under service and in line) at time t. Transition rates are given already in the problem; $\lambda_i = \lambda$ for $i \geq 0$ and $\mu_i = \mu$ for $i \geq 1$. The transition graph is



and to find the stationary distribution, note that

$$\frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n} = \frac{\lambda^n}{\mu^n}$$

so with $\rho = \lambda/\mu$ we see that a stationary distribution exists if and only if

$$1 + \sum_{n=1}^{\infty} \rho^n = \sum_{n=0}^{\infty} \rho^n < \infty$$

which is to say that $\rho < 1$. Using the formula for the geometric series, it is easy to see that the stationary distribution in this case is

$$\pi_n = (1-\rho)\rho^n, \;\; n = 0, 1, 2...$$

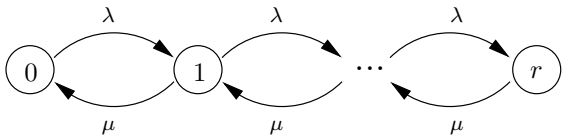
a geometric distribution including 0. The constant ρ is called the *traffic intensity* and in order for a stationary distribution to exist, this must be strictly less than one, meaning that service rates are higher than arrival rates. Only in this way can the server be efficient to regularly clear out the system.

The jump chain is the simple random walk where the probability to step up is $p = \lambda/(\lambda+\mu) = \rho/(1+\rho)$, which means that $\rho < 1$ if and only if $p < \frac{1}{2}$. □

The system in this example is called an $M/M/1$ *queue*, where "M" stands for "Markov," meaning that both interarrival times (first M) and service times (second M) are exponential, which is the only way in which this system satisfies the Markov property. The "1" indicates that there is one server. Other, non-Markovian, queueing systems are for example $M/D/1$, where service times are deterministic and $G/G/1$, where both interarrival times and service times have some general distribution, not necessarily exponential. The analysis of such systems require other methods, and we will stick with the M/M queues and analyze them as continuous time Markov chains. There are many variants of the $M/M/1$ queue. In the following three examples, we examine some of these, leaving others for the Problems section.

***Example* 8.4.9. (*Finite waiting room*).** Consider the $M/M/1$ queue, but suppose that there is only room for r customers in the system (the maximum queue length is $r-1$), denoted $M/M/1/r$.

The transition graph is



Again let $\rho = \lambda/\mu$. We have

$$\sum_{n=0}^{r} \rho^n = \begin{cases} r+1 & \text{if } \rho = 1 \\ \dfrac{1-\rho^{r+1}}{1-\rho} & \text{if } \rho \neq 1 \end{cases}$$

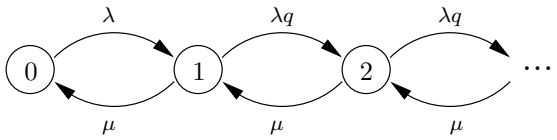
which gives stationary distribution

$$\pi_n = \begin{cases} \dfrac{1}{r+1}, \ \ n = 0, 1, ..., r & \text{if } \rho = 1 \\ \dfrac{(1-\rho)\rho^n}{1-\rho^{r+1}}, \ \ n = 0, 1, ..., r & \text{if } \rho \neq 1 \end{cases}$$

Note how the stationary distribution is uniform if $\rho = 1$ and how the probability π_r approaches 1 as $\rho \to \infty$. □

***Example* 8.4.10. (*Balking*).** Consider the $M/M/1$ queue and suppose that an arriving customer who finds the system not empty joins with probability q and leaves otherwise.

The only difference from the $M/M/1$ queue is that transitions from i to $i+1$ where $i \geq 1$ now occur according to a thinned Poisson process with rate λq. Thus we have the transition graph



and we get

$$\frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n} = \frac{\lambda^n q^{n-1}}{\mu^n}$$

where we let $\rho = \lambda q/\mu$ to obtain

$$1 + \sum_{n=1}^{\infty} \frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n} = 1 + \frac{1}{q}\sum_{n=1}^{\infty}\rho^n$$

which is finite if and only if $\rho < 1$, which makes intuitive sense. Since

$$\sum_{n=1}^{\infty}\rho^n = \frac{\rho}{1-\rho}$$

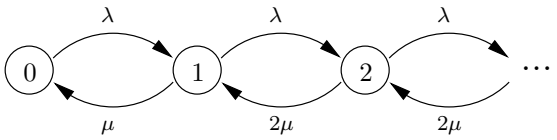
the stationary distribution is given by

$$\begin{cases} \pi_0 &= \left(1 + \dfrac{\rho}{q(1-\rho)}\right)^{-1} \\ \pi_n &= \dfrac{\rho^n\pi_0}{q}, \quad n = 1, 2, ... \end{cases}$$

Note how we redefined the traffic intensity ρ. We did this since we want $\rho < 1$ to be the criterion for when the system is efficient, that is, when the server manages to deal with the arrivals. □

***Example* 8.4.11. (*More than one server*).** Consider the $M/M/2$ queue, which is just like the $M/M/1$ queue except that there are two servers instead of one. An arriving customer can thus get immediate service if the system is empty or if only one server is busy.

The difference this time is what happens when both servers are busy. The time for a service to be completed is now the minimum of two exponentials with rate μ and is thus $\exp(2\mu)$ (recall Example 3.10.3). The transition graph is



This time we get

$$\frac{\lambda_0\lambda_1\cdots\lambda_{n-1}}{\mu_1\mu_2\cdots\mu_n} = \frac{\lambda^n}{\mu(2\mu)^{n-1}} = 2\left(\frac{\lambda}{2\mu}\right)^n$$

which suggests that we define the traffic intensity as $\rho = \lambda/2\mu$. We then get

$$\pi_n = 2\rho^n\pi_0, \;\; n \geq 1$$

and the stationary distribution exists if and only if $\rho < 1$. Since

$$\left(1 + 2\sum_{n=1}^{\infty}\rho^n\right) = \frac{1+\rho}{1-\rho}$$

the stationary distribution is given by

$$\begin{cases} \pi_0 = & \dfrac{1-\rho}{1+\rho} \\ \\ \pi_n = & \dfrac{2\rho^n(1-\rho)}{1+\rho}, \;\; n \geq 1 \end{cases}$$

□

8.4.4 Further Properties of Queueing Systems

When a queueing system has settled in to have the stationary distribution, we say that it is *in equilibrium.* Several different measures can be used to assess the efficiency of the system, often called *performance measures*. Let us examine some of them for the $M/M/1$ queue.

***Example* 8.4.12.** Consider the $M/M/1$ queue with $\rho < 1$ in equilibrium. **(a)** What is the expected number of customers in the system? **(b)** What is the expected queue length? **(c)** When a customer arrives, what is the probability that she does not have to wait in line? **(d)** When a customer arrives, what is her expected waiting time until service? **(e)** When a customer arrives, what is her expected total time in the system?

Let us introduce some random variables. Thus, let

$$\begin{array}{rcl} N & = & \text{the number of customers in the system} \\ Q & = & \text{the queue length} \\ W & = & \text{the waiting time until service} \\ T & = & \text{the total time spent in the system} \end{array}$$

For (a), we know that N has distribution

$$\pi_k = (1-\rho)\rho^k, \;\; k = 0, 1, ...$$

the geometric distribution including 0 with success probability $1-\rho$, and from Section 2.5.3 we know that

$$E[N] = \frac{\rho}{1-\rho}$$

which answers (a). For (b), note that

$$Q = \begin{cases} 0 & \text{if } N = 0 \text{ or } N = 1 \\ N-1 & \text{if } N > 1 \end{cases}$$

and hence

$$\begin{aligned} P(Q=0) &= \pi_0 + \pi_1 \\ P(Q=k) &= \pi_{k+1}, \;\; k \geq 1 \end{aligned}$$

which gives

$$\begin{aligned} E[Q] &= \sum_{k=0}^{\infty} kP(Q=k) \; = \; \sum_{k=1}^{\infty} k\pi_{k+1} \\ &= \rho \sum_{k=1}^{\infty} k(1-\rho)\rho^k \; = \; \rho E[N] \; = \; \frac{\rho^2}{1-\rho} \end{aligned}$$

The answer to (c) is simply $\pi_0 = 1-\rho$, and for (d), note that $W = 0$ if the system is empty and the sum of N i.i.d. exponentials with mean $1/\mu$ if there are N customers in the system (keep in mind that μ does not denote the mean but the service rate). By Corollary 3.11.6 we get

$$E[W] = E[N]\frac{1}{\mu} = \frac{\rho}{\mu(1-\rho)}$$

Finally, for (e), let S be a service time and note that $T = W + S$ to obtain

$$E[T] = \frac{\rho}{\mu(1-\rho)} + \frac{1}{\mu} = \frac{1}{\mu(1-\rho)}$$

We summarize as follows. In the $M/M/1$ system in equilibrium, we have

$$\begin{aligned} E[N] &= \frac{\rho}{1-\rho}, \qquad E[Q] \; = \; \frac{\rho^2}{1-\rho} \\ E[W] &= \frac{\rho}{\mu(1-\rho)}, \qquad E[T] \; = \; \frac{1}{\mu(1-\rho)} \end{aligned}$$

□

There are many interesting relations between these expected values. For example,

$$\frac{E[Q]}{E[N]} = \frac{E[W]}{E[T]}$$

which is also equal to the traffic intensity ρ. Also, since $\mu = \rho\lambda$, we obtain

$$E[N] = \lambda E[T] \tag{8.4.4}$$

formulas that are intuitively reasonable. These relations hold for many queueing systems, not only the $M/M/1$. Equation (8.4.4) is known as *Little's formula*. Also note how all the expectations go to ∞ as ρ approaches 1.

If there is finite waiting room, an obvious performance measure is how likely it is that the system is full, in which case arriving customers are lost. Thus, the probability that the system is full is the long-term proportion of arriving customers that are lost. Let us consider one example.

***Example* 8.4.13.** Consider the $M/M/1/r$ queue from Example 8.4.9. What proportion of customers are lost?

This is the probability that the system is full, π_r, which is

$$\pi_r = \begin{cases} \dfrac{1}{r+1} & \text{if } \rho = 1 \\ \dfrac{(1-\rho)\rho^r}{1-\rho^{r+1}} & \text{if } \rho \neq 1 \end{cases}$$

where we can note that $\pi_r \to 1$ as $\rho \to \infty$. □

8.5 MARTINGALES

In this section, we will introduce a class of stochastic processes called *martingales*[7] that is particularly useful in a wide variety of situations where asymptotic properties are of interest.

Let us go directly to the definition.

[7]The term originated in France in the 17th century as a class of betting strategies to increase the chances in various games of gambling.

***Definition* 8.5.1.** A sequence of random variables $Y_1, Y_2, \ldots$ is called a martingale in discrete time with respect to the sequence $X_1, X_2, \ldots$ if $E[|Y_n|] < \infty$ and

$$E[Y_{n+1}|X_1, \ldots, X_n] = Y_n \tag{8.5.1}$$

for all $n = 1, 2, \ldots$.

The conditional expectation used in Definition 8.5.1 is a generalization of Definition 3.7.4, where we interpreted conditional expectation as a random variable. In this context, (8.5.1) is a random variable assuming the value $E[Y_{n+1}|X_1 = x_1, \ldots, X_n = x_n]$ whenever the event $\{X_1 = x_1, \ldots, X_n = x_n\}$ occurs.

In many applications, we can actually choose $X_k = Y_k$ for all $k = 1, 2, \ldots$ so that (8.5.1) takes the form

$$E[Y_{n+1}|Y_1, \ldots, Y_n] = Y_n$$

which better illustrates the fundamental property of a martingale. Basically, it says that if we have observed the process for n steps, we at least know that the process in the next step on average will not deviate from the last value Y_n. The more general definition above is useful in situations where we can write Y_n as a function of the underlying variables $X_1, \ldots, X_n$.

***Example* 8.5.1.** One process that we have considered earlier that fits nicely into this theory is the symmetric random walk from Section 8.3.1 where $X_1, X_2, \ldots$ are i.i.d. random variables where $P(X_k = 1) = P(X_k = -1) = \frac{1}{2}$ and

$$S_n = \sum_{k=1}^{n} X_k$$

The first property in Definition 8.5.1 is clearly satisfied since $E[|S_n|] \leq n < \infty$ and

$$E[S_{n+1}|X_1, \ldots, X_n] = E[X_{n+1} + S_n|X_1, \ldots, X_n] = E[X_{n+1}] + S_n = S_n$$

shows that $S_1, S_2, \ldots$ is a martingale with respect to $X_1, X_2, \ldots$. □

Example* 8.5.2. *(Branching Processes). Another example where martingales are useful is the branching process introduced in Section 8.3.3. Here, the variables $X_1, X_2, \ldots$ represent the number of offspring of individuals in a generation and

$$Z_n = \sum_{k=1}^{Z_{n-1}} X_k, \quad n = 1, 2, \ldots$$

where $Z_0 \equiv 1$, denotes the number of individuals in generation n. If $\mu = E[X_k] < \infty$, Proposition 8.3.7 implies that $E[|Z_n|] = E[Z_n] = \mu^n < \infty$ and

$$E[Z_{n+1}|Z_1, \ldots, Z_n] = E\left[\sum_{k=1}^{Z_n} X_k \middle| Z_1, \ldots, Z_n\right] = Z_n\mu$$

showing that $Z_1, Z_2, \ldots$ is a martingale with respect to itself, but only if $\mu = 1$. However, it is actually quite easy to construct a martingale for arbitrary μ by rescaling Z_n with respect to the mean as

$$Y_n = \frac{Z_n}{\mu^n}$$

Clearly, $E[|Y_n|] = E[Y_n] = 1$ for all $n = 1, 2, \ldots$ and

$$\begin{aligned} E[Y_{n+1}|Z_1, \ldots, Z_n] &= E\left[\frac{1}{\mu^{n+1}}\sum_{k=1}^{Z_n} X_k \middle| Z_1, \ldots, Z_n\right] \\ &= \frac{1}{\mu^{n+1}} Z_n\mu = Y_n \end{aligned}$$

showing that $Y_1, Y_2, \ldots$ is a martingale with respect to $Z_1, Z_2, \ldots$. □

8.5.1 Martingale Convergence

The main result of this section is the following.

Proposition 8.5.1 (Martingale Convergence Theorem)**.** If $Y_1, Y_2, \ldots$ is a martingale with respect to some sequence $X_1, X_2, \ldots$ and $E[Y_n^2] < c < \infty$ for all $n = 1, 2, \ldots$, then there exists a random variable Y such that

$$Y_n \to Y \quad \text{as} \quad n \to \infty$$

with probability one.

Note that the convergence is almost surely, the stronger mode of convergence mentioned in the discussion of the law of large numbers in Section 4.2. This implies that Y_n also converges in probability and in distribution. Proposition 8.5.1 says that there exists a random variable Y in the limit, but it does not say anything about its properties. If we want to find out its distribution, say, we have to use other methods, e.g. limits of moment generating functions as described in the proof of Theorem 4.3.1.

Example* 8.5.3. *(Strong Law of Large Numbers). As a first example, we are going to demonstrate how to use Proposition 8.5.1 to strengthen Theorem 4.2.1[8] which says that the sample mean $\bar{X}$ converges in probability to the mean μ as $n \to \infty$.

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables and let

$$Y_n = \sum_{k=1}^{n} \frac{X_k - \mu}{k}$$

Clearly, $E[Y_n] = 0$ and

$$E[Y_n^2] = \text{Var}[Y_n] = \sum_{k=1}^{n} \frac{\sigma^2}{k^2} < \sigma^2 \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\sigma^2 \pi^2}{6}$$

so as long as $\sigma^2 < \infty$, the condition in Proposition 8.5.1 is satisfied. To conclude that Y_n is a martingale, we have to show that $E[|Y_n|] < \infty$ for all n, which follows from

$$\begin{aligned} E[|Y_n|] &= E[|Y_n| \mid |Y_n| \leq 1]P(|Y_n| \leq 1) + E[|Y_n| \mid |Y_n| > 1]P(|Y_n| > 1) \\ &< 1 \times P(|Y_n| \leq 1) + E[Y_n^2] < 1 + \frac{\sigma^2 \pi^2}{6} \end{aligned}$$

Proposition 8.5.1 now says that there exists some Y such that $Y_n \to Y$ with probability one. To obtain the strong law of large numbers we need a mathematical result called Kronecker's lemma, which says that if we have a sequence of arbitrary real numbers $x_1, x_2, \ldots$ and positive constants $b_1, b_2, \ldots$ that increase strictly to infinity then

$$\sum_{k=1}^{n} \frac{x_k}{b_k} \to z \quad \text{implies that} \quad \frac{1}{b_n} \sum_{k=1}^{n} x_k \to 0$$

as $n \to \infty$, where $|z| < \infty$. We have to be a bit cautious since we are dealing with random variables, but without going into details it turns out that

$$\sum_{k=1}^{n} \frac{X_k - \mu}{k} \to Y \quad \text{implies that} \quad \frac{1}{n} \sum_{k=1}^{n} (X_k - \mu) \to 0$$

as $n \to \infty$ with probability one. Finally, we note that since $\bar{X} - \mu \to 0$ it follows that $\bar{X} \to \mu$ with probability one. □

[8]Sometimes called the *Weak Law of Large Numbers* to distinguish it from this case.

Example* 8.5.4. *(Branching Processes). Recall Example 8.5.2 where we showed that $Y_n = Z_n/\mu^n$ for $n = 1, 2, \ldots$ is a martingale with respect to $Z_1, Z_2, \ldots$. Proposition 8.3.7 implies that

$$\begin{aligned} E[Y_n^2] &= (E[Y_n])^2 + \text{Var}[Y_n] = 1 + \frac{1}{\mu^{2n}} \times \frac{\sigma^2(\mu^n - 1)\mu^{n-1}}{\mu - 1} \\ &= 1 + \frac{\sigma^2(1 - \mu^{-n})}{\mu(\mu - 1)} < 1 + \frac{\sigma^2}{\mu(\mu - 1)} \end{aligned}$$

if $\mu > 1$. Hence, Proposition 8.5.1 says that there only exists a limit Y for the martingale Y_n in the supercritical case. Luckily, this is the most interesting case because we know from Proposition 8.3.8 that subcritical and critical branching processes always goes extinct eventually and it is possible, using other methods, to conclude that $Y_n \to 0$ for those cases.

The properties of Y are not that easy to obtain and we are not going to go much further here. It is possible, though, to show that Y is continuous except for a point mass at 0 equal to the extinction probability. This means that the event $\{Y > 0\}$ is equivalent to non-extinction and if we condition on this event, we can say that $Z_n \approx Y\mu^n$ for large n. Essentially, the branching process grows with a more or less deterministic rate but from a random level. □

8.5.2 Stopping Times

Let us assume that we can interpret a martingale as the accumulated fortune of a gambler playing a fair game. The game is fair in the sense that the expected amount after each play is equal to the gambler's fortune before the play. However, if we are lucky we may win some money or if we are unlucky we may lose some money in each individual play.

Let us, for simplicity, consider the symmetric random walk as a simple model for gambling where we bet \$1 and win \$2 if we get heads and lose our bet if we get tails. Throughout history, innumerable attempts have been made to beat the odds, to come up with the perfect foolproof strategy to win money no matter how the dice fall or the roulette wheel spins. One of the most (in)famous is the so called doubling strategy, where you always double your bet if you lose one play.

If it takes n plays to get heads, we have lost $1 + 2 + 4 + \ldots + 2^{n-2} = 2^{n-1} - 1$ dollars along the way but since the winning play brings in 2^{n-1} dollars, we will gain \$1 altogether. Seems solid, doesn't it, so, what's the problem? If we let N denote the number of plays it takes to get heads, we know that $p_n = P(N = n) = \left(\frac{1}{2}\right)^n$. The expected amount that we will have to bet to gain \$1 is then

$$\sum_{n=1}^{\infty}(2^{n-1} - 1)p_n = \sum_{n=1}^{\infty}\left(\frac{1}{2} - \frac{1}{2^n}\right) = \infty$$

Hence, no matter how big your initial fortune is, you will most likely be ruined before you stand to win any significant amount of money.

This is an example of a strategy using a *stopping time*, i.e. a predetermined rule when to stop playing. The rule has to be formulated so that we know for sure, after each play, if we will stop or not and that we are not allowed to use future, unobserved observations. In mathematical terms, it can be defined as follows.

***Definition* 8.5.2.** A random variable T that takes values in $\{1, 2, \ldots, \infty\}$ is called a stopping time for the sequence $X_1, X_2, \ldots$ if

$$P(T = n | X_1 = x_1, X_2 = x_2, \ldots) =$$

$$P(T = n | X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = 0 \text{ or } 1$$

This means that if we know exactly what values the stochastic process $X_1, X_2, \ldots$ has assumed up to and including step n, we will know for sure if the event $\{T = n\}$ has occurred or not. Also, note that we allow T to assume an infinite value, which corresponds to a situation where our rule is never met and we therefore never stop.

A poorly chosen stopping time usually means, as illustrated above, that there is a clear risk that we may have to wait for a very long time until we stop and that things may go awry before that. Exactly what conditions a useful stopping time should satisfy are covered in the following result.

Proposition 8.5.2 (Optional Stopping Theorem)**.** Let $Y_1, Y_2, \ldots$ be a martingale and T a stopping time with respect to some sequence $X_1, X_2, \ldots$. If

i) $P(T < \infty) = 1$

ii) $E[|Y_T|] < \infty$

iii) $E[Y_n | T > n] P(T > n) \to 0$ as $n \to \infty$

then $E[Y_T] = E[Y_1]$.

There are several versions of the optional stopping theorem giving slightly different conditions, but those described in Proposition 8.5.2 are usually easy to verify. This is bad news for all gamblers since it shows that there is no (reasonable) strategy that will increase your expected fortune in a fair game.[9] The doubling strategy described

[9]It is actually even worse because casinos and other gambling venues rarely offer any fair games. The odds are usually stacked in their favour, bringing in a small but steady profit. Processes where

above satisfies the first two conditions in Proposition 8.5.2 but not the third since

$$E[Y_n|T > n]P(T > n) = -(2^n - 1) \times \left(\frac{1}{2}\right)^n \to -1$$

as $n \to \infty$.

Example* 8.5.5. *(Gambler's Ruin). As a first example of the optional stopping theorem, let us revisit Example 1.6.16 where Ann and Bob flipped a coin and Ann paid Bob one dollar if it turned up heads and Bob paid Ann one dollar if it turned up tails. Ann started with a dollars and Bob with b dollars and the question was which one would be ruined first.

If we let X_k be 1 if we get heads in the kth flip and -1 otherwise, we can write Bob's total gain after n coin flips as

$$S_n = \sum_{k=1}^{n} X_K$$

This is clearly a martingale with respect to $X_1, X_2, \ldots$ since

$$E[S_{n+1}|X_1, \ldots, X_n] = (S_n - 1) \times \frac{1}{2} + (S_n + 1) \times \frac{1}{2} = S_n$$

The game stops either when $S_n = a$ (Ann is ruined) or $S_n = -b$ (Bob is ruined), so

$$T = \min\{n : S_n = a \text{ or } S_n = -b\}$$

is clearly a stopping time.

We know from Section 8.3.1 that the symmetric random walk is recurrent, which means that it will hit either a or $-b$ eventually. This implies both that $P(T < \infty) = 1$, verifying the first condition of Proposition 8.5.2, and that $P(T > n) \to 0$ as $n \to \infty$. We also know that the random variable S_T only assumes the values a and $-b$ so that $E[|S_T|] \leq \max(a, b) < \infty$. Finally, we realize that if the event $\{T > n\}$ occurs, the martingale has not stopped at time n so that $-b < S_n < a$ and the third condition follows.

Let p be the probability that Bob wins all the money (and Ann gets ruined) and note that $p = P(S_T = a)$. Proposition 8.5.2 now implies that

$$E[S_T] = ap - b(1 - p) = E[S_1] = 0$$

which gives us that

$$p = \frac{b}{a + b}$$

$E[Y_{n+1}|X_1, X_2, \ldots, X_n] \leq Y_n$ are called *supermartingales* (replace $\leq$ with $\geq$ and you get *submartingales*) and have similar properties.

We can also calculate the expected value of T by considering the sequence $Y_n = S_n^2 - n$, which also is a martingale with respect to $X_1, X_2, \ldots$ since

$$E[Y_{n+1}|X_1, \ldots, X_n] = E[(S_n + X_{n+1})^2 - (n+1)|S_n]$$

$$= \quad S_n^2 + 2S_n E[X_{n+1}] + E[X_{n+1}^2] - (n+1) = S_n^2 - n = Y_n$$

It is not quite as easy to verify that all conditions of Proposition 8.5.2 are satisfied, but we will skip those details. Hence, we know that

$$E[Y_T] = E[S_T^2 - T] = E[S_T^2] - E[T] = E[Y_1] = E[S_1^2 - 1] = 0$$

and, since we know that S_T is either equal to a with probability p or equal to $-b$ with probability $1 - p$, we get that

$$E[T] = E[S_T^2] = a^2 p + b^2(1-p) = \frac{a^2 b}{a+b} + \frac{b^2 a}{a+b} = ab$$

which would be very difficult to calculate in any other way. □

Example* 8.5.6.** ***(Ballot Theorem). There has been a ballot between Ann and Bob where Ann won with a votes whereas Bob only got b votes. If the votes where counted one by one in random order, what is the probability that Ann was ahead of Bob the whole time?

Let X_k be equal to 1 if vote number k was a vote for Ann and -1 if the vote was for Bob. Then $S_n = \sum_{k=1}^{n} X_k$ denotes the difference between the number of votes for Ann and Bob after n votes have been counted. The probability we are looking for can be expressed as $P(S_n > 0; 1 \leq n \leq N)$, where $N = a + b$ is the total number of votes.

There are several ways to solve this classic probabilistic problem[10] but we will use the optional stopping theorem to do it here. It turns out that we can simplify the problem by going backwards in time, creating a so called *backwards martingale*.

First, we note that after n votes have been counted, Ann have $(n + S_n)/2$ votes and Bob have $(n - S_n)/2$ votes. This implies that

$$P(S_n = S_{n+1} + 1|S_{n+1}) \quad = \quad \frac{(n+1-S_{n+1})/2}{n+1}$$

$$P(S_n = S_{n+1} - 1|S_{n+1}) \quad = \quad \frac{(n+1+S_{n+1})/2}{n+1}$$

[10]Introduced by the French mathematician Joseph Bertrand (1822–1900) in 1887. Martingale theory had not been developed at the time, so he solved it using combinatorical methods.

These conditional probabilities consider the events that we take away a vote for Bob and we take away a vote for Ann, respectively, from $n+1$ counted votes.

Now, we define the sequence $Y_1, Y_2, \ldots, Y_N$ as

$$Y_n = \frac{S_{N-n+1}}{N-n+1} \quad n = 1, 2, \ldots, N$$

which is a martingale with respect to itself since

$$\begin{aligned} E[Y_{n+1}|Y_1, \ldots, Y_n] &= E\left[\frac{S_{N-n}}{N-n}\middle| S_N, \ldots, S_{N-n+1}\right] \\ &= \frac{1}{N-n}\{(S_{N-n+1}+1)P(S_{N-n} = S_{N-n+1}+1|S_{N-n+1}) \\ &\quad +(S_{N-n+1}-1)P(S_{N-n} = S_{N-n+1}-1|S_{N-n+1})\} \\ &= \frac{1}{N-n}\left(S_{N-n+1} + \frac{(N-n+1-S_{N-n+1})/2}{N-n+1}\right. \\ &\quad \left. -\frac{(N-n+1+S_{N-n+1})/2}{N-n+1}\right) \\ &= \frac{S_{N-n+1}}{N-n}\left(1 - \frac{1}{N-n+1}\right) = \frac{S_{N-n+1}}{N-n+1} = Y_n \end{aligned}$$

Next, we define the stopping time T as

$$T = \min\{\min\{n : Y_n = 0\}, N\}$$

Since we are going backwards in time, T denotes the last time the candidates were even and if this never happens, we let $T = N$, which corresponds to the first vote counted. This means that if $T < N$, the candidates were even at T and $Y_T \equiv 0$. On the other hand, if $T = N$, Ann was always ahead and $Y_T \equiv 1$. Since both the martingale and the stopping time are bounded, it is obvious that all conditions in Proposition 8.5.2 are satisfied, so we can conclude that

$$E[Y_T] = E[Y_1] = E\left[\frac{S_N}{N}\right] = \frac{a-b}{a+b}$$

Finally, since the random variable Y_T only assumes the values 0 and 1, we get that

$$\begin{aligned} E[Y_T] &= 0 \times P(Y_T = 0) + 1 \times P(Y_T = 1) \\ &= 0 \times P(T < N) + 1 \times P(T = N) = P(T = N) \end{aligned}$$

which yields the probability that Ann was always in the lead as

$$P(\text{Ann always in the lead}) = P(T = N) = \frac{a-b}{a+b}$$

□

8.6 RENEWAL PROCESSES

In Section 3.12, we introduced the Poisson process as a point process with i.i.d. exponentially distributed inter-arrival times. In this section, we will outline some theory of the most obvious generalization of this, namely to allow the inter-arrival times to have any non-negative distribution.

Let us start with the formal definition.

***Definition* 8.6.1.** Let $T_1, T_2, \ldots$ be i.i.d. non-negative random variables such that $P(T_k = 0) < 1$ and $S_n = \sum_{k=1}^{n} T_k$, then

$$N(t) = \max\{n : S_n \leq t\}$$

is a *renewal process* for $t \geq 0$.

The time points $S_1, S_2, \ldots$ are called the *renewals* since it is like starting all over from the beginning every time an inter-arrival time ends. Since $N(t) = n$ is equivalent to $S_n \leq t < S_{n+1}$, we see that $N(t)$ counts the number of renewals in the interval $[0, t]$. The cdf of T_k and S_n will be denoted $F(t) = P(T_k \leq t)$ and $F_n(t) = P(S_n \leq t)$, respectively, and we let μ and σ^2 denote the mean and the variance of T_k.

The exact distribution of $N(t)$ is usually very difficult to derive, except for a few simple cases, but it can at least theoretically be expressed in the form

$$\begin{aligned} P(N(t) = n) &= P(S_n \leq t < S_{n+1}) = P(S_n \leq t) - P(S_{n+1} \leq t) \\ &= F_n(t) - F_{n+1}(t) \end{aligned}$$

for $n = 0, 1, 2, \ldots$, where $F_0(t) = 1$. We can also write the mean $m(t) = E[N(t)]$, using Proposition 2.4.1, as

$$m(t) = \sum_{n=1}^{\infty} P(N(t) \geq n) = \sum_{n=1}^{\infty} P(S_n \leq t) = \sum_{n=1}^{\infty} F_n(t)$$

The mean $m(t)$ is also called the *renewal function* and it turns out that it uniquely determines the renewal process, i.e. if we are given an expression for $m(t)$, we can in principle calculate the distribution $F(t)$ of the inter-arrival times.

***Example* 8.6.1.** Let us have a look at the Poisson process expressed as a renewal process. Since the inter-arrival times are exponentially distributed, we know that

$$F(t) = 1 - e^{-\lambda t}$$

and, by Proposition 3.10.6 and Section 2.8.2, that

$$F_n(t) = 1 - e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!}$$

The distribution of $N(t)$ can now be calculated as

$$P(N(t)=n) = \left(1 - e^{-\lambda t}\sum_{k=0}^{n-1}\frac{(\lambda t)^k}{k!}\right) - \left(1 - e^{-\lambda t}\sum_{k=0}^{n}\frac{(\lambda t)^k}{k!}\right) = e^{\lambda t}\frac{(\lambda t)^n}{n!}$$

which is the Poisson distribution with mean λt. The renewal function becomes

$$\begin{aligned} m(t) &= \sum_{n=1}^{\infty}\left(1 - e^{-\lambda t}\sum_{k=0}^{n-1}\frac{(\lambda t)^k}{k!}\right) = \sum_{n=1}^{\infty}\left(1 - e^{-\lambda t}\left(e^{\lambda t} - \sum_{k=n}^{\infty}\frac{(\lambda t)^k}{k!}\right)\right) \\ &= e^{-\lambda t}\sum_{n=1}^{\infty}\sum_{k=n}^{\infty}\frac{(\lambda t)^k}{k!} = e^{-\lambda t}\sum_{k=1}^{\infty}\sum_{n=1}^{k}\frac{(\lambda t)^k}{k!} = e^{-\lambda t}\sum_{k=1}^{\infty}\frac{(\lambda t)^k}{(k-1)!} \\ &= e^{-\lambda t}\lambda t\sum_{k=1}^{\infty}\frac{(\lambda t)^{k-1}}{(k-1)!} = \lambda t \end{aligned}$$

where we used the Taylor expansion of $e^{\lambda t}$. This is a very longwinded way to derive the mean of the Poisson process but at least it shows that the theory works. What we have shown is that the Poisson process has a linear renewal function and, since the renewal function determines the renewal process, it is worth noting that the Poisson process is the *only* renewal process with a linear renewal function. □

For continuous inter-arrival times, we can use the property that the renewal process starts over at every renewal S_n to obtain the following result.

Proposition 8.6.1 (The Renewal Equation). If $T_1, T_2, \ldots$ are continuous inter-arrival times with cdf $F(t)$ and pdf $f(t)$, the renewal function $m(t)$ satisfies

$$m(t) = F(t) + \int_0^t m(t-u)f(u)\,du$$

Proof. If we condition on the time of the first renewal, Proposition 3.7.1 says that we can write the renewal function

$$m(t) = E[N(t)] = \int_0^{\infty} E[N(t)|T_1 = u]f(u)\,du \tag{8.6.1}$$

Now, if the first renewal occurs after t, i.e. that $u > t$ holds, clearly $E[N(t)|T_1 = u] = 0$. On the other hand, if $u \leq t$, we count the first renewal and start the process from the beginning at time u. This means that

$$E[N(t)|T_1 = u] = 1 + E[N(t-u)] = 1 + m(t-u) \quad u \leq t$$

which means that we can write (8.6.1) as

$$m(t) = \int_0^t (1 + m(t-u))f(u)\,du = F(t) + \int_0^t m(t-u)f(u)\,du$$ ■

The main advantage of Proposition 8.6.1 is that we only need to consider the distribution of the first renewal to get the renewal function, but instead we need to solve an integral equation, which is not that easy in most cases. Still, it gives us a tool to verify whether a proposed $m(t)$ actually is a renewal function for a given renewal process.

8.6.1 Asymptotic Properties

Since the exact distribution of the renewal process is difficult to obtain in most cases, it would at least be interesting to see how $N(t)$ behaves as $t \to \infty$. The first result is essentially a version of the Law of Large Numbers.

Proposition 8.6.2. For $\mu > 0$, it holds that

$$\frac{N(t)}{t} \xrightarrow{P} \frac{1}{\mu} \quad \text{as } t \to \infty$$

Proof. Since $S_{N(t)}$ is the time of the last renewal before t, we note that

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

which yields that

$$\frac{S_{N(t)}}{N(t)} \leq \frac{t}{N(t)} < \frac{S_{N(t)+1}}{N(t)} \tag{8.6.2}$$

for $N(t) > 0$. Now, it holds that $N(t) \to \infty$ as $t \to \infty$ (Why is that?), so the Law of Large Numbers implies that

$$\frac{S_{N(t)}}{N(t)} \xrightarrow{P} \mu \quad \text{as } t \to \infty$$

Furthermore, using a similar argument yields that

$$\frac{S_{N(t)+1}}{N(t)} = \frac{S_{N(t)+1}}{N(t)+1} \times \frac{N(t)+1}{N(t)} \xrightarrow{P} \mu \times 1 \quad \text{as } t \to \infty$$

We have shown that both the upper and lower bound in (8.6.2) converge in probability to μ and since $t/N(t)$ is between these, it has to converge in probability to μ as well. Taking the reciprocals completes the proof. ■

The ratio $1/\mu$ is called the *renewal rate* because it denotes the average number of renewals per time unit in the long run. Note that the result holds also for infinite μ, in which case we interpret the renewal rate as 0.

In light of Proposition 8.6.2, the following result seems quite natural.

Proposition 8.6.3 (Elementary Renewal Theorem). For $\mu > 0$, it holds that

$$\frac{m(t)}{t} \to \frac{1}{\mu} \quad \text{as } t \to \infty$$

It would seem that this is a simple consequence of Proposition 8.6.2, but that is actually not the case. Proving the elementary renewal theorem turns out to be rather difficult and requires some asymptotic theory that we have not introduced.

The following result is slightly more general but holds only for continuous inter-arrival times.

Proposition 8.6.4 (Renewal Theorem). For continuous $F(t)$ and $\mu > 0$, it holds that

$$m(t+s) - m(t) \to \frac{s}{\mu} \quad \text{as } t \to \infty$$

This result says essentially that if we slide a ruler of length s on the time axis, the expected number of renewals covered by the ruler will be, at least approximately, proportional to the length of the ruler. For a fixed and large t, Propositions 8.6.3 and 8.6.4 can be summarized as

$$\begin{aligned} m(t) &\approx \frac{t}{\mu} \\ m(t+s) - m(t) &\approx \frac{s}{\mu} \end{aligned}$$

which illustrates how they are connected. Proposition 8.6.3 can be regarded as a global average of renewals while Proposition 8.6.4 gives a similar local property.

There is also a central limit theorem for renewal processes.

Proposition 8.6.5. For $\mu < \infty$ and $\sigma^2 < \infty$, it holds that

$$P\left(\frac{N(t) - t/\mu}{\sigma\sqrt{t/\mu^3}} \leq x\right) \to \Phi(x)$$

as $t \to \infty$, where $\Phi(x)$ is the cdf of the standard normal distribution.

Proof. For a fixed x, let

$$t = n\mu - x\sigma\sqrt{n} \tag{8.6.3}$$

and consider the probability

$$P(N(t) < n) = P\left(\frac{N(t) - t/\mu}{\sigma\sqrt{t/\mu^3}} < \frac{n\mu - t}{\sigma\sqrt{t/\mu}}\right) = P\left(\frac{N(t) - t/\mu}{\sigma\sqrt{t/\mu^3}} < x\sqrt{\frac{n\mu}{t}}\right)$$

Now, if we let $n \to \infty$ we see from (8.6.3) that $t \to \infty$ and vice versa. Also, (8.6.3) implies that $\sqrt{n\mu/t} \to 1$, which means that

$$\lim_{n\to\infty} P(N(t) < n) = \lim_{t\to\infty} P\left(\frac{N(t) - t/\mu}{\sigma\sqrt{t/\mu^3}} < x\right)$$

Finally, since the events $\{N(t) < n\}$ and $\{S_n > t\}$ are equivalent, we get that

$$\begin{aligned} P(N(t) < n) &= P(S_n > t) = P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} > \frac{t - n\mu}{\sigma\sqrt{n}}\right) \\ &= P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} > -x\right) \to 1 - \Phi(-x) = \Phi(x) \end{aligned}$$

as $n \to \infty$ by the central limit theorem. ■

***Example* 8.6.2.** A Geiger-Müller counter (or Geiger counter for short) is an electronic device that detects radioactive particles, usually from beta and gamma radiation. One problem is that every time a particle is registered, the counter has to be reset before it can detect new particles. These periods are called *dead periods* because any particles that arrive while the counter is reset are lost. Therefore, some appropriate adjustment needs to be done in order to avoid underestimation of the radioactive intensity.

Let us assume that the radioactive particles arrive according to a Poisson process with rate λ and that the lengths of the dead periods $Y_1, Y_2, \ldots$ are i.i.d. random variables. Let $X_1, X_2, \ldots$ denote the times until a reset counter registers a particle and,

from the properties of the Poisson process, these are independent and exponentially distributed with parameter λ. Furthermore, we assume that the lengths of the dead periods are independent of the Poisson process.

If we let the inter-arrival times be $T_k = X_k + Y_k$ for $k = 1, 2, \ldots$, we can define the renewals as

$$S_n = \sum_{k=1}^{n} T_k = \sum_{k=1}^{n} (X_k + Y_k)$$

In this case, we get the mean and variance of the inter-arrival times as

$$\begin{aligned} \mu &= \frac{1}{\lambda} + \mu_Y \\ \sigma^2 &= \frac{1}{\lambda^2} + \sigma_Y^2 \end{aligned} \tag{8.6.4}$$

Let us say that we have run the counter for a long time t and registered $N(t)$ particle emissions.[11] Proposition 8.6.2 implies that

$$N(t) \approx \frac{t}{\mu} = \frac{t}{1/\lambda + \mu_Y}$$

which gives us the appropriate estimator of λ as

$$\widehat{\lambda} = \frac{1}{\dfrac{t}{N(t)} - \mu_Y}$$

Having obtained an estimate for λ, we can then use Proposition 8.6.5 to calculate a confidence interval. First, we get an interval for $1/\mu$ with approximate confidence level q as

$$\frac{1}{\mu} \approx \frac{N(t)}{t} \pm z\,\frac{\sigma}{\sqrt{t\mu^3}} \approx \frac{N(t)}{t} \pm z\widehat{\sigma}\,\frac{\sqrt{N(t)^3}}{t^2}$$

where $\Phi(z) = (1+q)/2$. In the second approximation above, we used the fact that $\mu \approx t/N(t)$ and

$$\widehat{\sigma}^2 = \frac{1}{\widehat{\lambda}^2} + \sigma_Y^2$$

Then, (8.6.4) can be used to transform the interval for λ.

For a practical example, let us assume that we have detected $N(t) = 5630$ particles during one second and the dead periods are uniformly distributed between 0 and 200 μs. Let us use 1 ms as a convenient time unit so that λ denotes the average number of particle emissions per ms. It is clear that $\mu_Y = 0.1$, which yields the point estimate

$$\widehat{\lambda} = \frac{1}{\dfrac{1000}{5630} - 0.1} = 12.9$$

[11] If the time point t ends up in a dead period, the number of detected particles is actually $N(t) + 1$ since the last particle is registered before t although the corresponding inter-arrival time ends after t. However, if t is large, this is negligible.

Now, we get

$$\widehat{\sigma}^2 = \frac{1}{12.9^2} + \frac{0.2^2}{12} = 0.0094$$

which yields the 95 % interval

$$\frac{1}{\mu} \approx \frac{5630}{1000} \pm 1.96\sqrt{0.0094}\frac{\sqrt{5630^3}}{1000^2} = 5.63 \pm 0.08$$

Finally, this yields the limits for λ as

$$\frac{1}{\frac{1}{5.58} - 0.1} \leq \lambda \leq \frac{1}{\frac{1}{5.71} - 0.1} \quad (\approx 95\%)$$

or

$$12.5 \leq \lambda \leq 13.3 \quad (\approx 95\%)$$

□

Example* 8.6.3.** ***(Delayed Renewal Process). There is one variation of the regular renewal process that is of particular interest, where the distribution of the initial inter-arrival time T_1 may be different from the subsequent. Such a situation may emerge if we start the process in between two adjacent renewals rather than at a specific renewal time, hence the term delayed renewal process. Since only the first inter-arrival time differs, all previous limiting results can be shown to hold also for the delayed renewal process. In fact, it has some interesting theoretical properties that actually can be used to prove some of these results for the regular renewal process.

Let $F(t)$ denote the cdf of $T_2, T_3, \ldots$ as before and denote the cdf of T_1 by $\widetilde{F}(t)$. Furthermore, let $\widetilde{S}_n$ denote the renewals, $\widetilde{N}(t)$ the delayed renewal process and $\widetilde{m}(t)$ the renewal function. Now, we can use the same method as in the proof of Proposition 8.6.1 to obtain

$$\begin{aligned} \widetilde{m}(t) &= \int_0^t E[\widetilde{N}(t)|T_1 = u]\widetilde{f}(u)\,du = \int_0^t (1 + m(t-u))\widetilde{f}(u)\,du \\ &= \widetilde{F}(t) + \int_0^t m(t-u)\widetilde{f}(u)\,du \end{aligned} \tag{8.6.5}$$

Note that we get the mean $m(t-u)$ in the integral since we get an ordinary renewal process after conditioning on $\{T_1 = u\}$. If we apply Proposition 8.6.1 to $m(t-u)$, we can write (8.6.5) as

$$\widetilde{F}(t) + \int_0^t \left(F(t-u) + \int_0^{t-u} m(t-u-v)f(v)\,dv \right) \widetilde{f}(u)\,du$$

$$= \widetilde{F}(t) + \int_0^t F(t-u)\widetilde{f}(u)\,du + \int_0^t \left(\int_0^{t-v} m(t-v-u)\widetilde{f}(u)\,du \right) f(v)\,dv$$

The first integral can be transformed, using integration by parts, into

$$\begin{aligned}\int_0^t F(t-u)\widetilde{f}(u)\,du &= \Big[F(t-u)\widetilde{F}(u)\Big]_0^t + \int_0^t f(t-u)\widetilde{F}(u)\,du \\ &= \int_0^t \widetilde{F}(t-u)f(u)\,du\end{aligned}$$

and we can use (8.6.5) again to rewrite the inner integral as

$$\int_0^{t-v} m(t-v-u)\widetilde{f}(u)\,du = \widetilde{m}(t-v) - \widetilde{F}(t-v)$$

Together, these results imply that

$$\begin{aligned}\widetilde{m}(t) &= \widetilde{F}(t) + \int_0^t \widetilde{F}(t-u)f(u)\,du + \int_0^t (\widetilde{m}(t-v) - \widetilde{F}(t-v))f(v)\,dv \\ &= \widetilde{F}(t) + \int_0^t \widetilde{m}(t-v)f(v)\,dv \qquad (8.6.6)\end{aligned}$$

which is the renewal equation for the delayed renewal process.

As for the ordinary renewal process, this equation uniquely determines the renewal function. If we look at the asymptotic properties as $t \to \infty$, we realize that they depend largely on $F(t)$ and that the initial inter-arrival time eventually becomes irrelevant. This means that we may choose whatever $\widetilde{F}(t)$ we like without affecting any limiting results. Proposition 8.6.3 says that the scaled renewal function $m(t)/t$ for a regular renewal process converges towards $1/\mu$ and, as mentioned above, the same can be shown to hold for $\widetilde{m}(t)/t$. Let us see if we can choose $\widetilde{F}(t)$ so that

$$\widetilde{m}(t) = \frac{t}{\mu}$$

for *all* $t \geq 0$. In that case, (8.6.6) implies that

$$\begin{aligned}\widetilde{F}(t) &= \widetilde{m}(t) - \int_0^t \widetilde{m}(t-v)f(v)\,dv = \frac{t}{\mu} - \int_0^t \frac{t-v}{\mu}f(v)\,dv \\ &= \frac{t}{\mu} - \left[\frac{t-v}{\mu}F(v)\right]_0^t - \int_0^t \frac{1}{\mu}F(v)\,dv = \frac{1}{\mu}\int_0^t (1-F(v))\,dv\end{aligned}$$

as long as $\mu < \infty$. □

8.7 BROWNIAN MOTION

The simple random walk presented in Section 8.3.1 describes a discrete process that at unit time points jumps one unit up or one unit down with fixed probabilities p and

$1-p$, respectively, independent of all previous jumps. Let us see if we can construct a continuous version of this that makes infinitesimally small jumps infinitesimally often. To achieve this, we can consider random walks with jump sizes Δx and time intervals of length Δt and then gradually decrease Δx and Δt towards zero. In the following, we will only consider the symmetric random walk where $p = 1 - p = \frac{1}{2}$.

Let $X_1, X_2, \ldots$ be independent random variables such that

$$X_k = \begin{cases} +1 & \text{with probability } 1/2. \\ \\ -1 & \text{with probability } 1/2. \end{cases} \tag{8.7.1}$$

Then we can define the process

$$S_n(t) = \Delta x\, X_1 + \Delta x\, X_2 + \ldots + \Delta x\, X_{[t/\Delta t]} = \Delta x \sum_{i=1}^{n} X_i \tag{8.7.2}$$

where $n = [t/\Delta t]$ denotes the largest integer less than or equal to the real number $t/\Delta t$.

Since $E[X_k] = 0$ and $\text{Var}[X_k] = 1$, we can calculate the mean and variance of (8.7.2) as

$$\begin{aligned} E[S_n(t)] &= \Delta x \sum_{i=1}^{n} E[X_k] = 0 \\ \text{Var}[S_n(t)] &= (\Delta x)^2 \sum_{i=1}^{n} \text{Var}[X_k] = (\Delta x)^2 n = (\Delta x)^2 \left[\frac{t}{\Delta t}\right] \end{aligned} \tag{8.7.3}$$

If we let $\Delta t \to 0$, then $n \to \infty$ and we can use the Central Limit Theorem to conclude that the sum in (8.7.2), properly standardized, converges to the standard normal distribution.

To obtain a similar result for $S_n(t)$, we also have to let $\Delta x \to 0$ in some orderly fashion. It is clear from (8.7.3) that Δx would have to decrease at a slower rate than Δt, of the order $\sqrt{\Delta t}$, if we are to obtain a positive and finite variance in the limit. For simplicity, let $\Delta x = \sigma\sqrt{\Delta t}$ to obtain the variance limit $\text{Var}[S_n(t)] \to \sigma^2 t$ and let $\Delta t \to 0$. Then, the Central Limit Theorem implies that

$$\frac{S_n(t) - E[S_n(t)]}{\sqrt{\text{Var}[S_n(t)]}} \xrightarrow{d} N(0,1) \tag{8.7.4}$$

as $n \to \infty$.

Since both $E[S_n(t)]$ and $\text{Var}[S_n(t)]$ converge to finite and, for the variance, positive limits, we can reformulate (8.7.4) as

$$S_n(t) \xrightarrow{d} B(t) \sim N(0, \sigma^2 t) \tag{8.7.5}$$

Another important property of the limiting process $B(t)$ is that it has independent increments, i.e. that the changes $B(t_3) - B(t_2)$ and $B(t_2) - B(t_1)$ are independent

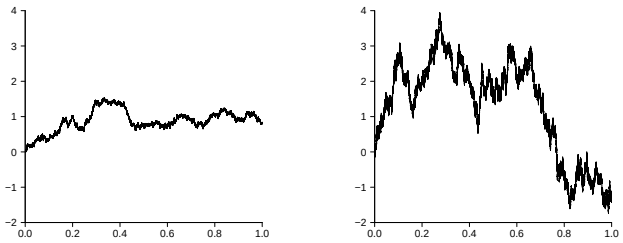


Fig. 8.3 Two realizations of Brownian motions in the unit interval for (a) $\sigma = 1$ and (b) $\sigma = 4$.

random variables for $t_1 < t_2 < t_3$, which should be intuitively clear from the random walk definition above. In fact, these are the two fundamental defining properties of the process.

***Definition* 8.7.1.** The *Brownian Motion*[12] $B(t)$ is a real valued stochastic process in real time $t \geq 0$ that satisfies

i) $B(0) = 0$

ii) If $t_1 < t_2 < \ldots < t_n$, then $B(t_1), B(t_2) - B(t_1), \ldots, B(t_n) - B(t_{n-1})$ are independent.

iii) $B(t+s) - B(s) \sim N(0, \sigma^2 t)$ for $t, s \geq 0$

Note that part iii) of Definition 8.7.1 is slightly more general than (8.7.5) since it says that all increments are also normally distributed. It does behave very erratically, which is illustrated in Figure 8.3 where two simulated Brownian motions are shown for $\sigma = 1$ and $\sigma = 4$.

This is one of the most studied stochastic processes in the mathematical literature partly because it has a lot of fascinating theoretical properties but also because it has been found to be very useful in quite different areas like physics (quantum mechanics), electronics (filtering theory) and economics (option pricing). We will not have time to

[12]Named after the Scottish botanist Robert Brown (1773–1858) who studied pollen grains submerged in liquid. To his amazement he observed that particles ejected from the pollen grains moved in a very erratic random fashion. He was the first to observe this phenomenon but he did not manage to explain it. Later, the American mathematician Norbert Wiener (1894–1964) developed the mathematical theory behind it and therefore the process $B(t)$ is sometimes also called the *Wiener process*

explore the Brownian motion in detail here except for some of the most fundamental and useful properties.

One interesting feature of $B(t)$ is that it is scale invariant in both time and space. It is quite easy to see, by checking the conditions in Definition 8.7.1, that if $B(t)$ is a Brownian motion with variance σ^2, then $B(t)/\sigma$ is a Brownian motion with variance 1. This is called the *standard Brownian motion* and since any Brownian motion can be standardized in this way, it is common to assume that $\sigma^2 = 1$. Similarily, it can be shown that if $B(t)$ is a standard Brownian motion, then $B(\sigma^2 t)$ is a Brownian motion with variance σ^2. These scalings can also be combined.

Proposition 8.7.1 (Self-similarity)**.** Let $B(t)$ be a standard Brownian motion. Then the process

$$\widetilde{B}(t) = \frac{B(a^2 t)}{a}$$

is also a standard Brownian motions for any $a \neq 0$.

Essentially, what this says is that if we would "zoom in" on the Brownian motion we would see a process that would look quite similar to what we started with, i.e. the Brownian motion can be charaterized as a *random fractal*. However, in order for this to work, we see that the rescaling of the horizontal time axis has to be the square of the rescaling of the vertical space axis. Also note that the rescaling constants does not have to be positive. A negative a means that we reverse the vertical axis and, by symmetry of the normal distribution, we still get a Brownian motion.

8.7.1 Hitting Times

One quantity of interest is the time until the Brownian motion attains some predetermined level. As a practical example, let us say that we buy a share of stock whose value can be described by a Brownian motion and decides to sell this after its value has increased by a certain amount. Then we would like to know how long time it takes until our investment pays off. We start with the following definition.

***Definition* 8.7.2.** The time until the Brownian motion $B(t)$ hits a is called the *hitting time* and is defined

$$T_a = \min\{t : B(t) = a\}$$

Clearly, T_a is a non-negative continuous random variable whose distribution is given below.

Proposition 8.7.2. The hitting time T_a of a standard Brownian motion has pdf

$$f_{T_a}(t) = \frac{|a|}{\sqrt{2\pi t^3}} e^{-a^2/2t} \quad t > 0$$

for any $a \neq 0$.

Proof. Let us first assume that $a > 0$ and consider the event that the Brownian motion exceeds a at time t. The law of total probability gives us

$$\begin{aligned} P(B(t) \geq a) &= P(B(t) \geq a | T_a > t) P(T_a > t) \\ &\quad + P(B(t) \geq a | T_a \leq t) P(T_a \leq a) \end{aligned} \tag{8.7.6}$$

The first conditional probability above is clearly 0 since the event $\{T_a > t\}$ means that the Brownian motion has not hit a at time t and since $B(t)$ is continuous, it cannot be above a. If we turn to the second conditional probability, the condition says that we have hit a before t. Then we can split up $B(t)$ in the two parts

$$B(t) = B(T_a) + (B(t) - B(T_a)) = a + (B(t) - a)$$

Now, the increment $B(t) - a$ is normally distributed with mean 0, which means that

$$P(B(t) \geq a | T_a > t) = P(B(t) - a \geq 0) = \frac{1}{2}$$

by symmetry. This means that (8.7.6) can be written

$$P(T_a \leq t) = 2P(B(t) \geq a) = 2\left(1 - \Phi\left(\frac{a-0}{\sqrt{t}}\right)\right)$$

For $a < 0$ we consider the event $\{B(t) \leq a\}$ and, again by symmetry, we get that

$$P(T_a \leq t) = 2P(B(t) \leq a) = 2\Phi\left(\frac{a-0}{\sqrt{t}}\right) = 2\left(1 - \Phi\left(-\frac{a}{\sqrt{t}}\right)\right)$$

Hence, the cdf can be written

$$F_{T_a}(t) = 2\left(1 - \Phi\left(\frac{|a|}{\sqrt{t}}\right)\right) \tag{8.7.7}$$

and taking the derivative with respect to t yields the pdf

$$-2\phi\left(\frac{|a|}{\sqrt{t}}\right) \times \left(-\frac{|a|}{2t^{3/2}}\right) = \frac{|a|}{\sqrt{2\pi t^3}} e^{-a^2/2t}$$

■

The properties of this distribution are not that easy to investigate. Even if T_a is finite with probability 1, implying that we will hit a eventually, it turns out that the mean is infinite, in analogy with Corollary 8.3.4, suggesting that it might take awhile. The median, however, turns out to be just $1.67a^2$, so for small a there is a fair chance that we will reach a fairly soon.

***Example* 8.7.1.** Since the Brownian motion can be characterized as the scaled limit of a symmetric random walk, we can use its properties to say something, at least approximately, about random walks. Therefore, let

$$S_n = \sum_{k=1}^{n} X_k$$

where $X_1, X_2, \ldots$ are defined in (8.7.1), be a symmetric random walk. As an example, let us consider the probability that the random walk will reach $S_n = 100$ within 1000 steps.

If we let $\Delta t = 0.001$ in and $\Delta x = \sqrt{\Delta t} = \sqrt{0.001}$, by the proper scaling, we see that the event that S_n reaches 100 within 1000 steps is equivalent to $S_n(t)$ defined by (8.7.2) reaching $100 * \Delta x = \sqrt{10}$ in the interval $[0, 1]$. Since Δt is comparatively small, we can use (8.7.5) to conclude that $S_n(t) \overset{d}{\approx} B(t)$, which means that we can express the event approximately as $\{T_{\sqrt{10}} \leq 1\}$ and (8.7.7) yields that

$$P(\max_{0 \leq n \leq 1000} S_n \geq 100) \approx P(T_{\sqrt{10}} \leq 1) = 2(1 - \Phi(3.16)) = 0.0016$$

This is quite unlikely, so let us consider 10,000 steps instead. We could do this by another rescaling where $\Delta t = 10^{-4}$, but it suffices to use the scaling we have and instead look at the time $t = n\Delta t = 10$. Then we get that

$$P(\max_{0 \leq n \leq 10,000} S_n \geq 100) \approx P(T_{\sqrt{10}} \leq 10) = 2(1 - \Phi(1)) = 0.32$$

□

Another related quantity of interest is the maximum of a Brownian motion in a fixed interval. Therefore, we define the random variable

$$M_t = \max_{0 \leq s \leq t} B(t)$$

for the standard Brownian motion $B(t)$. This variable has the following distribution.

Proposition 8.7.3. The maximum M_t of a standard Brownian motion in the interval $[0, t]$ has pdf

$$f_{M_t}(x) = \frac{2}{\sqrt{2\pi t}} e^{-x^2/2t} \quad x > 0$$

Proof. The cdf of M_t can be written

$$F_{M_t}(x) = P(M_t \leq x) = 1 - P(M_t > x) = 1 - P(T_x < t)$$

since we have to hit the level x in order to exceed it before the time point t. Now, equation (8.7.7) yields that

$$F_{M_t}(x) = 1 - F_{T_x}(t) = 1 - 2\left(1 - \Phi\left(\frac{|x|}{\sqrt{t}}\right)\right) = 2\Phi\left(\frac{x}{\sqrt{t}}\right) + 1$$

since $x > 0$, and taking the derivative with respect to x gives us the pdf

$$f_{M_t}(x) = 2\phi\left(\frac{x}{\sqrt{t}}\right) \times \frac{1}{\sqrt{t}} = \frac{2}{\sqrt{2\pi t}} e^{-x^2/2t}$$

■

Since t is fixed and x is variable, this distribution is easier to understand. Essentially, it consists of the positive part of a normal distribution with mean 0 and variance t, which among other things means that the mean and variance can be calculated analytically (see Problem 66).

8.7.2 Variations of the Brownian Motion

The standard Brownian motion $B(t)$ is quite interesting in itself, but it is also used as a component in other, more realistic processes applied in various fields. In this section we are going to look at some of the most common variations of $B(t)$.

Example* 8.7.2. *(Brownian Motion with Drift). Sometimes it is not enough to look at a Brownian motion with zero mean but rather a process that have a tendency to drift in a particular direction. We say that a Brownian motion with drift parameter μ and variance σ^2 is defined as

$$X(t) = \mu t + \sigma B(t)$$

For a fixed t, we get that $X(t)$ is normally distributed with mean μt and variance $\sigma^2 t$. This means that the mean is increasing (or decreasing for $\mu < 0$) towards infinity, but

if the variance also increase towards infinity, can we really be sure that the process will have a tendency to drift upwards? This is indeed the case, which can be seen by considering the limit of

$$P(X(t) > a) = 1 - \Phi\left(\frac{a - \mu t}{\sigma\sqrt{t}}\right) = 1 - \Phi\left(\frac{a}{\sigma}\frac{1}{\sqrt{t}} - \frac{\mu}{\sigma}\sqrt{t}\right)$$

as $t \to \infty$. If $\mu > 0$, we see that the argument of Φ above tends to $-\infty$, which means that $P(X(t) > a) \to 1$ for any a showing that $X(t)$ will eventually exceed any finite level. □

Example* 8.7.3. *(Geometric Brownian Motion). Let $X(t)$ be Brownian motion with drift as defined in Example 8.7.2. Geometric Brownian motion is then defined as

$$Y(t) = e^{X(t)} = e^{\mu t + \sigma B(t)}$$

Since ordinary Brownian motion can be characterized as a sum of independent normal increments, this gives us a process that can handle products of independent increments. Furthermore, since $X(t)$ is normally distributed for any fixed t, we get that $Y(t)$ follows a lognormal distribution (see Section 2.8.1). This means that we can get the mean and variance as

$$\begin{aligned} E[Y(t)] &= e^{\mu t + \sigma^2 t/2} = e^{(\mu + \sigma^2/2)t} \\ \text{Var}[Y(t) &= e^{2\mu t + \sigma^2 t}\left(e^{\sigma^2 t} - 1\right) \end{aligned} \tag{8.7.8}$$

It is interesting to note that we can have an increasing mean even if μ is negative as long as $\mu > -\sigma^2/2$. It also holds that we can have an increasing variance if $\mu > -\sigma^2$. This yields an interesting interval $-\sigma^2 < \mu < -\sigma^2/2$ where we get a process whose mean decreases exponentially to zero but whose variance increases exponentially to infinity. Basically, it means that the drift downwards is not strong enough to prevent brief excursions upwards and the exponential functions amplifies the effect rather strongly.

One of the most common applications of the geometric Brownian motion is to use it as a model for stock prices and even whole stock exchange indices. Since price fluctuations are relative to the stock, we get a natural situation where increments are multiplicative rather than additive. In this context, μ is usually called the *expected logarithmic return* and σ the *volatility* of the stock. The *risk* is usually interpreted as the probability that the stock will decrease in value over a fixed time period, i.e.

$$P(Y(t) < Y(0) = 1) = P(X(t) < 0) = \Phi\left(\frac{0 - \mu t}{\sigma\sqrt{t}}\right) = \Phi\left(-\frac{\mu}{\sigma}\sqrt{t}\right)$$

This shows that in order to minimize the risk, we should choose a stock with high expected return but low volatility. This is unfortunately not always possible in reality,

stock with high expected return often suffer from high volatility and vice versa. On the other hand, a high volatility also means a higher expected value, as shown in (8.7.8), so it might still be worth to accept the higher risk. The geometric Brownian motion has also proved to be a valuable tool in the field of *option pricing*, but we will not go further into that area here. □

Example* 8.7.4.** ***(Brownian Bridge). In some situations, we may be interested in the properties of a Brownian motion that returns to its starting point. Without loss of generality, we will consider a standard Brownian motion $B(t)$ that returns to 0 at time $t = 1$. Brownian bridges with arbitrary variance σ^2 and arbitrary time points t can be obtain by appropriate time and space scaling. One way to do that is to condition on the event $\{B(1) = 0\}$, but an equivalent and more convenient construction is

$$B^\circ(t) = B(t) - tB(1)$$

The Brownian bridge $B^\circ(t)$ is a linear combination of two normally distributed quantities and, hence, also normally distributed. The mean is clearly 0 and to calculate the variance, we first note that $B(t)$ and $B(1) - B(t)$ are independent and normally distributed increments, which yields that

$$\begin{aligned} \text{Var}[B^\circ(t)] &= \text{Var}[B(t) - tB(1)] = \text{Var}[(1-t)B(t) - t(B(1) - B(t))] \\ &= (1-t)^2\text{Var}[B(t)] + t^2\text{Var}[B(1) - B(t)] \\ &= (1-t)^2 t + t^2(1-t) = t(1-t) \end{aligned}$$

One of the most important applications of the Brownian bridge is as a large sample approximation of the empirical distribution function

$$\widehat{F}_n(x) = \frac{1}{n}\sum_{k=1}^{n} I_{\{X_k \le x\}}$$

introduced in Section 6.9.1. We note that, for a fixed x, the total number Y of events $\{X_k \le x\}$ that occur is binomially distributed with parameters n and $p = F(x)$. Then, the Central Limit Theorem implies that

$$\frac{Y - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$$

as $n \to \infty$ and, since $\widehat{F}_n(x) = Y/n$, we get that

$$\sqrt{n}(\widehat{F}_n(x) - F(x)) \xrightarrow{d} N(0, F(x)(1-F(x))) \stackrel{d}{=} B^\circ(F(x))$$

Now, this result holds pointwise for every fixed x, but can actually be generalized to hold uniformly over the whole interval and is then usually expressed as

$$\sqrt{n}(\widehat{F}(\cdot) - F(\cdot)) \xrightarrow{d} B^\circ(F(\cdot)) \tag{8.7.9}$$