Chapter 5

COUNTABLE-STATE MARKOV CHAINS

5.1 Introduction and classification of states

Markov chains with a countably-infinite state space (more briefly, *countable-state Markov chains*) exhibit some types of behavior not possible for chains with a finite state space. With the exception of the first example to follow and the section on branching processes, we label the states by the nonnegative integers. This is appropriate when modeling things such as the number of customers in a queue, and causes no loss of generality in other cases.

The following two examples give some insight into the new issues posed by countable state spaces.

Example 5.1.1. Consider the familiar Bernoulli process $\{S_n = X_1 + \cdots X_n;\, n \geq 1\}$ where $\{X_n;\, n \geq 1\}$ is an IID binary sequence with $\mathsf{p}_X(1) = p$ and $\mathsf{p}_X(-1) = (1-p) = q$. The sequence $\{S_n;\, n \geq 1\}$ is a sequence of integer random variables (rv's) where $S_n = S_{n-1} + 1$ with probability p and $S_n = S_{n-1} - 1$ with probability q. This sequence can be modeled by the Markov chain in Figure 5.1.

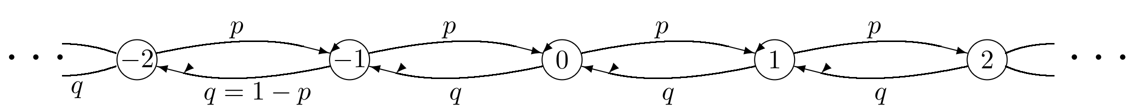

Figure 5.1: A Markov chain with a countable state space modeling a Bernoulli process. If $p > 1/2$, then as time n increases, the state X_n becomes large with high probability, *i.e.*, $\lim_{n\to\infty} \Pr\{X_n \geq j\} = 1$ for each integer j. Similarly, for $p < 1/2$, the state becomes highly negative.

Using the notation of Markov chains, P^n_{0j} is the probability of being in state j at the end of the nth transition, conditional on starting in state 0. The final state j is the number of positive transitions k less the number of negative transitions $n-k$, *i.e.*, $j = 2k - n$. Thus, using the binomial formula,

$$P_{0j}^n = \binom{n}{k} p^k q^{n-k} \qquad \text{where } k = \frac{j+n}{2}; \quad j+n \text{ even.} \tag{5.1}$$

All states in this Markov chain communicate with all other states, and are thus in the same class. The formula makes it clear that this class, *i.e.*, the entire set of states in the Markov chain, is periodic with period 2. For n even, the state is even and for n odd, the state is odd.

What is more important than the periodicity, however, is what happens to the state probabilities for large n. As we saw in (1.88) while proving the central limit theorem for the binomial case,

$$P_{0j}^n \sim \frac{1}{\sqrt{2\pi npq}} \exp\left[\frac{-(k-np)^2}{2pqn}\right]\} \qquad \text{where } k = \frac{j+n}{2}; \quad j+n \text{ even.} \tag{5.2}$$

In other words, P_{0j}^n, as a function of j, looks like a quantized form of the Gaussian density for large n. The significant terms of that distribution are close to $k = np$, *i.e.*, to $j = n(2p-1)$. For $p > 1/2$, the state increases with increasing n. Its distribution is centered at $n(2p-1)$, but the distribution is also spreading out as $\sqrt{n}$. For $p < 1/2$, the state similarly decreases and spreads out. The most interesting case is $p = 1/2$, where the distribution remains centered at 0, but due to the spreading, the PMF approaches 0 as $1/\sqrt{n}$ for all j.

For this example, then, the probability of each state approaches zero as $n \to \infty$, and this holds for all choices of p, $0 < p < 1$. If we attempt to define a steady-state probability as 0 for each state, then these probabilities do not sum to 1, so they cannot be viewed as a steady-state distribution. Thus, for countable-state Markov chains, the notions of recurrence and steady-state probabilities will have to be modified from that with finite-state Markov chains. The same type of situation occurs whenever $\{S_n; n \geq 1\}$ is a sequence of sums of arbitrary IID integer-valued rv's.

Most countable-state Markov chains that are useful in applications are quite different from Example 5.1.1, and instead are quite similar to finite-state Markov chains. The following example bears a close resemblance to Example 5.1.1, but at the same time is a countable-state Markov chain that will keep reappearing in a large number of contexts. It is a special case of a birth-death process, which we study in Section 5.2.

Example 5.1.2. Figure 5.2 is similar to Figure 5.1 except that the negative states have been eliminated. A sequence of IID binary rv's $\{X_n; n \geq 1\}$, with $\mathsf{p}_X(1) = p$ and $\mathsf{p}_X(-1) = q = 1 - p$, controls the state transitions. Now, however, $S_n = \max(0, S_{n-1} + X_n$, so that S_n is a nonnegative rv. All states again communicate, and because of the self transition at state 0, the chain is aperiodic.

For $p > 1/2$, transitions to the right occur with higher frequency than transitions to the left. Thus, reasoning heuristically, we expect the state S_n at time n to drift to the right with increasing n. Given $S_0 = 0$, the probability P_{0j}^n of being in state j at time n, should then tend to zero for any fixed j with increasing n. As in Example 5.1.1, we see that a

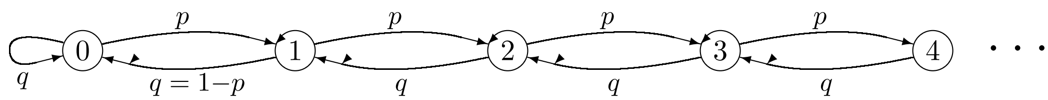

Figure 5.2: A Markov chain with a countable state space. If $p > 1/2$, then as time n increases, the state X_n becomes large with high probability, *i.e.*, $\lim_{n\to\infty} \Pr\{X_n \geq j\} = 1$ for each integer j.

steady state does not exist. In more poetic terms, the state wanders off into the wild blue yonder.

One way to understand this chain better is to look at what happens if the chain is truncated The truncation of Figure 5.2 to k states is analyzed in Exercise 3.9. The solution there defines $\rho = p/q$ and shows that if $\rho \neq 1$, then $\pi_i = (1-\rho)\rho^i/(1-\rho^k)$ for each i, $0 \leq i < k$. For $\rho = 1$, $\pi_i = 1/k$ for each i. For $\rho < 1$, the limiting behavior as $k \to \infty$ is $\pi_i = (1-\rho)\rho^i$. Thus for $\rho < 1$ ($p < 1/2$), the steady state probabilities for the truncated Markov chain approaches a limit which we later interpret as the steady state probabilities for the untruncated chain. For $\rho > 1$ ($p > 1/2$), on the other hand, the steady-state probabilities for the truncated case are geometrically decreasing from the right, and the states with significant probability keep moving to the right as k increases. Although the probability of each fixed state j approaches 0 as k increases, the truncated chain never resembles the untruncated chain.

Perhaps the most interesting case is that where $p = 1/2$. The nth order transition probabilities, P^n_{0j} can be calculated exactly for this case (see Exercise 5.3) and are very similar to those of Example 5.1.1. In particular,

$$P^n_{0j} = \begin{cases} \binom{n}{(j+n)/2} 2^{-n} & \text{for } j \geq 0, (j+n) \text{ even} \\ \} & \\ \binom{n}{(j+n+1)/2} 2^{-n} & \text{for } j \geq 0, (j+n) \text{ odd} \end{cases} \tag{5.3}$$

$$\sim \sqrt{\frac{\}2}{\pi n}} \exp\left[\frac{-j^2}{2n}\right] \qquad \text{for } j \geq 0. \tag{5.4}$$

We see that P^n_{0j} for large n is approximated by the positive side of a quantized Gaussian distribution. It looks like the positive side of the PMF of (5.1) except that it is no longer periodic. For large n, P^n_{0j} is concentrated in a region of width $\sqrt{n}$ around $j = 0$, and the PMF goes to 0 as $1/\sqrt{n}$ for each j as $n \to \infty$.

Fortunately, the strange behavior of Figure 5.2 when $p \geq q$ is not typical of the Markov chains of interest for most applications. For typical countable-state Markov chains, a steady state does exist, and the steady-state probabilities of all but a finite number of states (the number depending on the chain and the application) can almost be ignored for numerical calculations.

5.1.1 Using renewal theory to classify and analyze Markov chains

The matrix approach used to analyze finite-state Markov chains does not generalize easily to the countable-state case. Fortunately, renewal theory is ideally suited for this purpose, especially for analyzing the long term behavior of countable-state Markov chains. We must first revise the definition of recurrent states. The definition for finite-state Markov chains does not apply here, and we will see that, under the new definition, the Markov chain in Figure 5.2 is recurrent for $p \leq 1/2$ and transient for $p > 1/2$. For $p = 1/2$, the chain is called null-recurrent, as explained later.

In general, we will find that for a recurrent state j, the sequence of subsequent entries to state j, conditional on starting in j, forms a renewal process. The renewal theorems then specify the time-average relative-frequency of state j, the limiting probability of j with increasing time and a number of other relations.

We also want to understand the sequence of epochs at which one state, say j, is entered, conditional on starting the chain at some other state, say i. We will see that, subject to the classification of states i and j, this gives rise to a delayed renewal process. In preparing to study these renewal processes and delayed renewal process, we need to understand the inter-renewal intervals. The probability mass functions (PMF's) of these intervals are called *first-passage-time probabilities* in the notation of Markov chains.

Definition 5.1.1. *The first-passage-time probability, $f_{ij}(n)$, of a Markov chain is the probability, conditional on $X_0 = i$, that the first subsequent entry to state j occurs at discrete epoch n. That is, $f_{ij}(1) = P_{ij}$ and for $n \geq 2$,*

$$f_{ij}(n) = \Pr\{X_n{=}j, X_{n-1}{\neq}j, X_{n-2}{\neq}j, \dots, X_1{\neq}j | X_0{=}i\}. \tag{5.5}$$

The distinction between $f_{ij}(n)$ and $P_{ij}^n = \Pr\{X_n = j | X_0 = i\}$ is that $f_{ij}(n)$ is the probability that the *first* entry to j (after time 0) occurs at time n, whereas P_{ij}^n is the probability that *any* entry to j occurs at time n, both conditional on starting in state i at time 0. The definition in (5.5) also applies for $j = i$; $f_{ii}(n)$ is thus the probability, given $X_0 = i$, that the first occurrence of state i after time 0 occurs at time n. Since the transition probabilities are independent of time, $f_{kj}(n-1)$ is also the probability, given $X_1 = k$, that the first subsequent occurrence of state j occurs at time n. Thus we can calculate $f_{ij}(n)$ from the iterative relations

$$f_{ij}(n) = \sum_{k \neq j} P_{ik} f_{kj}(n-1); \quad n > 1; \qquad f_{ij}(1) = P_{ij}. \tag{5.6}$$

Note that the sum excludes $k = j$, since $P_{ij} f_{jj}(n-1)$ is the probability that state j occurs first at epoch 1 and next at epoch n. Note also from the Chapman-Kolmogorov equation that $P_{ij}^n = \sum_k P_{ik} P_{kj}^{n-1}$. In other words, the only difference between the iterative expressions to calculate $f_{ij}(n)$ and P_{ij}^n is the exclusion of $k = j$ in the expression for $f_{ij}(n)$.

With this iterative approach, the first-passage-time probabilities $f_{ij}(n)$ for a given n must be calculated for all i before proceeding to calculate them for the next larger value of n. This also gives us $f_{jj}(n)$, although $f_{jj}(n)$ is not used in the iteration.

Let $\mathsf{F}_{ij}(n)$, for $n \geq 1$, be the probability, given $X_0 = i$, that state j occurs at some time between 1 and n inclusive. Thus,

$$\mathsf{F}_{ij}(n) = \sum_{m=1}^{n}\} f_{ij}(m). \tag{5.7}$$

For each i, j, $\mathsf{F}_{ij}(n)$ is non-decreasing in n and (since it is a probability) is upper bounded by 1. Thus $\mathsf{F}_{ij}(\infty)$ (*i.e.*, $\lim_{n\to\infty} \mathsf{F}_{ij}(n)$) must exist, and is the probability, given $X_0 = i$, that state j will ever occur. If $\mathsf{F}_{ij}(\infty) = 1$, then, given $X_0 = i$, it is certain (with probability 1) that the chain will eventually enter state j. In this case, we can define a random variable (rv) T_{ij}, conditional on $X_0 = i$, as the *first-passage time* from i to j. Then $f_{ij}(n)$ is the PMF of T_{ij} and $\mathsf{F}_{ij}(n)$ is the distribution function of T_{ij}. If $\mathsf{F}_{ij}(\infty) < 1$, then T_{ij} is a defective rv, since, with some non-zero probability, there is no first-passage to j. Defective rv's are not considered to be rv's (in the theorems here or elsewhere), but they do have many of the properties of rv's.

The first-passage time T_{jj} from a state j back to itself is of particular importance. It has the PMF $f_{jj}(n)$ and the distribution function $\mathsf{F}_{jj}(n)$. It is a rv (as opposed to a defective rv) if $\mathsf{F}_{jj}(\infty) = 1$, *i.e.*, if the state eventually returns to state j with probability 1 given that it starts in state j. This leads to the definition of recurrence.

Definition 5.1.2. *A state j in a countable-state Markov chain is* recurrent *if* $\mathsf{F}_{jj}(\infty) = 1$. *It is* transient *if* $\mathsf{F}_{jj}(\infty) < 1$.

Thus each state j in a countable-state Markov chain is either recurrent or transient, and is recurrent if and only if an eventual return to j (conditional on $X_0 = j$) occurs with probability 1. Equivalently, j is recurrent if and only if T_{jj}, the time of first return to j, is a rv. Note that for the special case of finite-state Markov chains, this definition is consistent with the one in Chapter 3. For a countably-infinite state space, however, the earlier definition is not adequate. An example is provided by the case $p > 1/2$ in Figure 5.2. Here i and j communicate for all states i and j, but it is intuitively obvious (and shown in Exercise 5.2, and further explained in Section 5.2) that each state is transient.

If the initial state X_0 of a Markov chain is a recurrent state j, then T_{jj} is the integer time of the first recurrence of state j. At that recurrence, the Markov chain is in the same state j as it started in, and the discrete interval from T_{jj} to the next occurrence of state j, say $T_{jj,2}$ has the same distribution as T_{jj} and is clearly independent of T_{jj}. Similarly, the sequence of successive recurrence intervals, $T_{jj}, T_{jj,2}, T_{jj,3}, \dots$ is a sequence of IID rv's. This sequence of recurrence intervals[1] is then the sequence of inter-renewal intervals of a renewal process, where each renewal interval has the distribution of T_{jj}. These inter-renewal intervals have the PMF $f_{jj}(n)$ and the distribution function $\mathsf{F}_{jj}(n)$.

Since results about Markov chains depend very heavily on whether states are recurrent or transient, we will look carefully at the probabilities $\mathsf{F}_{ij}(n)$. Substituting (5.6) into (5.7), we

[1]Note that in Chapter 4 the inter-renewal intervals were denoted $X_1, X_2, \dots$, whereas here $X_0, X_1, \dots$, is the sequence of states in the Markov chain and $T_{jj}, T_{jj,2}, \dots$, is the sequence of inter-renewal intervals.

obtain

$$\mathsf{F}_{ij}(n) = P_{ij} + \sum_{k\neq j} \} P_{ik}\mathsf{F}_{kj}(n-1); \quad n > 1; \qquad \mathsf{F}_{ij}(1) = P_{ij}. \tag{5.8}$$

To understand the expression $P_{ij} + \sum_{k\neq j} P_{ik}\mathsf{F}_{kj}(n-1)$, note that the first term, P_{ij}, is $f_{ij}(1)$ and the second term, $\sum_{k\neq j} P_{ik}\mathsf{F}_{kj}(n-1)$, is equal to $\sum_{\ell=2}^{n} f_{ij}(\ell)$.

We have seen that $\mathsf{F}_{ij}(n)$ is non-decreasing in n and upper bounded by 1, so the limit $\mathsf{F}_{ij}(\infty)$ must exist. Similarly, $\sum_{k\neq j} P_{ik}\mathsf{F}_{kj}(n-1)$ is non-decreasing in n and upper bounded by 1, so it also has a limit, equal to $\sum_{k\neq j} P_{ik}\mathsf{F}_{kj}(\infty)$. Thus

$$\mathsf{F}_{ij}(\infty) = P_{ij} + \sum_{k\neq j} \} P_{ik}\mathsf{F}_{kj}(\infty). \tag{5.9}$$

For any given j, (5.9) can be viewed as a set of linear equations in the variables $\mathsf{F}_{ij}(\infty)$ for each state i. There is not always a unique solution to this set of equations. In fact, the set of equations

$$x_{ij} = P_{ij} + \sum_{k\neq j} P_{ik}x_{kj}; \qquad \text{all states } i \tag{5.10}$$

always has a solution in which $x_{ij} = 1$ for all i. If state j is transient, however, there is another solution in which x_{ij} is the true value of $\mathsf{F}_{ij}(\infty)$ and $\mathsf{F}_{jj}(\infty) < 1$. Exercise 5.1 shows that if (5.10) is satisfied by a set of nonnegative numbers $\{x_{ij}; 1 \leq i \leq J\}$, then $\mathsf{F}_{ij}(\infty) \leq x_{ij}$ for each i.

We have defined a state j to be recurrent if $\mathsf{F}_{jj}(\infty) = 1$ and have seen that if j is recurrent, then the returns to state j, given $X_0 = j$ form a renewal process. All of the results of renewal theory can then be applied to the random sequence of integer times at which j is entered. The main results from renewal theory that we need are stated in the following lemma.

Lemma 5.1.1. *Let $\{N_{jj}(t); t \geq 0\}$ be the counting process for occurrences of state j up to time t in a Markov chain with $X_0 = j$. The following conditions are then equivalent.*

1. *state j is recurrent.*
2. $\lim_{t\to\infty} N_{jj}(t) = \infty$ *with probability* 1.
3. $\lim_{t\to\infty} \mathsf{E}[N_{jj}(t)] = \infty$.
4. $\lim_{t\to\infty} \sum_{1\leq n\leq t} P_{jj}^{n} = \infty$.

Proof: First assume that j is recurrent, *i.e.*, that $\mathsf{F}_{jj}(\infty) = 1$. This implies that the inter-renewal times between occurrences of j are IID rv's, and consequently $\{N_{jj}(t);\, t \geq 1\}$ is a renewal counting process. Recall from Lemma 4.3.1 of Chapter 4 that, whether or not the expected inter-renewal time $\mathsf{E}[T_{jj}]$ is finite, $\lim_{t\to\infty} N_{jj}(t) = \infty$ with probability 1 and $\lim_{t\to\infty} \mathsf{E}[N_{jj}(t)] = \infty$.

Next assume that state j is transient. In this case, the inter-renewal time T_{jj} is not a rv, so $\{N_{jj}(t); t \geq 0\}$ is not a renewal process. An eventual return to state j occurs only with probability $\mathsf{F}_{jj}(\infty) < 1$, and, since subsequent returns are independent, the total number of returns to state j is a geometric rv with mean $\mathsf{F}_{jj}(\infty)/[1 - \mathsf{F}_{jj}(\infty)]$. Thus the total number of returns is finite with probability 1 and the expected total number of returns is finite. This establishes the first three equivalences.

Finally, note that P_{jj}^n , the probability of a transition to state j at integer time n, is equal to the expectation of a transition to j at integer time n (*i.e.*, a single transition occurs with probability P_{jj}^n and 0 occurs otherwise). Since $N_{jj}(t)$ is the sum of the number of transitions to j over times 1 to t, we have

$$\mathsf{E}\left[N_{jj}(t)\right] = \sum_{1 \leq n \leq t}\} P_{jj}^n,$$

which establishes the final equivalence. □

Our next objective is to show that all states in the same class as a recurrent state are also recurrent. Recall that two states are in the same class if they communicate, *i.e.*, each has a path to the other. For finite-state Markov chains, the fact that either all states in the same class are recurrent or all transient was relatively obvious, but for countable-state Markov chains, the definition of recurrence has been changed and the above fact is no longer obvious.

Lemma 5.1.2. *If state j is recurrent and states i and j are in the same class, i.e., i and j communicate, then state i is also recurrent.*

Proof: From Lemma 5.1.1, state j satisfies $\lim_{t\to\infty} \sum_{1\leq n \leq t} P_{jj}^n = \infty$. Since j and i communicate, there are integers m and k such that $P_{ij}^m > 0$ and $P_{ji}^k > 0$. For every walk from state j to j in n steps, there is a corresponding walk from i to i in $m + n + k$ steps, going from i to j in m steps, j to j in n steps, and j back to i in k steps. Thus

$$\begin{aligned} P_{ii}^{m+n+k} &\geq P_{ij}^m \, P_{jj}^n \, P_{ji}^k \\ \sum_{n=1}^{\infty}\} P_{ii}^n &\geq \sum_{n=1}^{\infty}\} P_{ii}^{m+n+k} \geq P_{ij}^m \, P_{ji}^k \sum_{n=1}^{\infty}\} P_{jj}^n = \infty. \end{aligned}$$

Thus, from Lemma 5.1.1, i is recurrent, completing the proof. □

Since each state in a Markov chain is either recurrent or transient, and since, if one state in a class is recurrent, all states in that class are recurrent, we see that if one state in a class is transient, they all are. Thus we can refer to each class as being recurrent or transient. This result shows that Theorem 3.2.1 also applies to countable-state Markov chains. We state this theorem separately here to be specific.

Theorem 5.1.1. *For a countable-state Markov chain, either all states in a class are transient or all are recurrent.*

We next look at the delayed counting process $\{N_{ij}(n); n \geq 1\}$. If this is a delayed *renewal* counting process, then we can use delayed renewal processes to study whether the effect of the initial state eventually dies out. If state j is recurrent, we know that $\{N_{jj}(n);\, n \geq 1\}$ is a renewal counting process. In addition, in order for $\{N_{ij}(n);\, n \geq 1\}$ to be a delayed renewal counting process, it is necessary for the first-passage time to be a rv, *i.e.*, for $\mathsf{F}_{ij}(\infty)$ to be 1.

Lemma 5.1.3. *Let states i and j be in the same recurrent class. Then $\mathsf{F}_{ij}(\infty) = 1$.*

Proof: Since i is recurrent, the number of visits to i by time t, given $X_0 = i$, is a renewal counting process $N_{ii}(t)$. There is a path from i to j, say of probability $\alpha > 0$. Thus the probability that the first return to i occurs before visiting j is at most $1-\alpha$. The probability that the second return occurs before visiting j is thus at most $(1-\alpha)^2$ and the probability that the nth occurs without visiting j is at most $(1-\alpha)^n$. Since i is visited infinitely often with probability 1 as $n \to \infty$, the probability that j is never visited is 0. Thus $\mathsf{F}_{ij}(\infty) = 1$. □

Lemma 5.1.4. *Let $\{N_{ij}(t); t \geq 0\}$ be the counting process for transitions into state j up to time t for a Markov chain given $X_0 = i \neq j$. Then if i and j are in the same recurrent class, $\{N_{ij}(t); t \geq 0\}$ is a delayed renewal process.*

Proof: From Lemma 5.1.3, T_{ij}, the time until the first transition into j, is a rv. Also T_{jj} is a rv by definition of recurrence, and subsequent intervals between occurrences of state j are IID, completing the proof. □

If $\mathsf{F}_{ij}(\infty) = 1$, we have seen that the first-passage time from i to j is a rv, *i.e.*, is finite with probability 1. In this case, the mean time $\overline{T}_{ij}$ to first enter state j starting from state i is of interest. Since T_{ij} is a nonnegative random variable, its expectation is the integral of its complementary distribution function,

$$\overline{T}_{ij} = 1 + \sum_{n=1}^{\infty}(1 - \mathsf{F}_{ij}(n)). \tag{5.11}$$

It is possible to have $\mathsf{F}_{ij}(\infty) = 1$ but $\overline{T}_{ij} = \infty$. As will be shown in Section 5.2, the chain in Figure 5.2 satisfies $\mathsf{F}_{ij}(\infty) = 1$ and $\overline{T}_{ij} < \infty$ for $p < 1/2$ and $\mathsf{F}_{ij}(\infty) = 1$ and $\overline{T}_{ij} = \infty$ for $p = 1/2$. As discussed before, $\mathsf{F}_{ij}(\infty) < 1$ for $p > 1/2$. This leads us to the following definition.

Definition 5.1.3. *A state j in a countable-state Markov chain is* positive-recurrent *if $\mathsf{F}_{jj}(\infty) = 1$ and $\overline{T}_{jj} < \infty$. It is* null-recurrent *if $\mathsf{F}_{jj}(\infty) = 1$ and $\overline{T}_{jj} = \infty$.*

Each state of a Markov chain is thus classified as one of the following three types — positive-recurrent, null-recurrent, or transient. For the example of Figure 5.2, null-recurrence lies on a boundary between positive-recurrence and transience, and this is often a good way to look at null-recurrence. Part f) of Exercise 6.3 illustrates another type of situation in which null-recurrence can occur.

Assume that state j is recurrent and consider the renewal process $\{N_{jj}(t); t \geq 0\}$. The limiting theorems for renewal processes can be applied directly. From the strong law for renewal processes, Theorem 4.3.1,

$$\lim_{t\to\infty} N_{jj}(t)/t = 1/\overline{T}_{jj} \qquad \text{with probability 1.} \tag{5.12}$$

From the elementary renewal theorem, Theorem 4.6.1,

$$\lim_{t\to\infty} \mathsf{E}\left[N_{jj}(t)/t\right] = 1/\overline{T}_{jj}. \tag{5.13}$$

Equations (5.12) and (5.13) are valid whether j is positive-recurrent or null-recurrent.

Next we apply Blackwell's theorem to $\{N_{jj}(t); t \geq 0\}$. Recall that the period of a given state j in a Markov chain (whether the chain has a countable or finite number of states) is the greatest common divisor of the set of integers $n > 0$ such that $P_{jj}^n > 0$. If this period is d, then $\{N_{jj}(t); t \geq 0\}$ is arithmetic with span λ (*i.e.*, renewals occur only at times that are multiples of λ). From Blackwell's theorem in the arithmetic form of (4.61),

$$\lim_{n\to\infty} \Pr\{X_{n\lambda} = j \mid X_0 = j\} = \lambda/\overline{T}_{jj}. \tag{5.14}$$

If state j is aperiodic (*i.e.*, $\lambda = 1$), this says that $\lim_{n\to\infty} \Pr\{X_n = j \mid X_0 = j\} = 1/\overline{T}_{jj}$. Equations (5.12) and (5.13) suggest that $1/\overline{T}_{jj}$ has some of the properties associated with a steady-state probability of state j, and (5.14) strengthens this if j is aperiodic. For a Markov chain consisting of a single class of states, all positive-recurrent, we will strengthen this association further in Theorem 5.1.4 by showing that there is a unique *steady-state distribution*, $\{\pi_j, j \geq 0\}$ such that $\pi_j = 1/\overline{T}_{jj}$ for all j and such that $\pi_j = \sum_i \pi_i P_{ij}$ for all $j \geq 0$ and $\sum_j \pi_j = 1$. The following theorem starts this development by showing that (5.12-5.14) are independent of the starting state.

Theorem 5.1.2. *Let j be a recurrent state in a Markov chain and let i be any state in the same class as j. Given $X_0 = i$, let $N_{ij}(t)$ be the number of transitions into state j by time t and let $\overline{T}_{jj}$ be the expected recurrence time of state j (either finite or infinite). Then*

$$\lim_{t\to\infty} N_{ij}(t)/t = 1/\overline{T}_{jj} \qquad \textit{with probability } 1 \tag{5.15}$$

$$\lim_{t\to\infty} \mathsf{E}\left[N_{ij}(t)/t\right] = 1/\overline{T}_{jj}. \tag{5.16}$$

If j is also aperiodic, then

$$\lim_{n\to\infty} \Pr\{X_n = j \mid X_0 = i\} = 1/\overline{T}_{jj}. \tag{5.17}$$

Proof: Since i and j are recurrent and in the same class, Lemma 5.1.4 asserts that $\{N_{ij}(t); t \geq 0\}$ is a delayed renewal process for $j \neq i$. Thus (5.15) and (5.16) follow from Theorems 4.8.1 and 4.8.2 of Chapter 4. If j is aperiodic, then $\{N_{ij}(t); t \geq 0\}$ is a delayed renewal process for which the inter-renewal intervals T_{jj} have span 1 and T_{ij} has an integer span. Thus, (5.17) follows from Blackwell's theorem for delayed renewal processes, Theorem 4.8.3. For $i = j$, (5.15-5.17) follow from (5.12-5.14), completing the proof. □

Theorem 5.1.3. *All states in the same class of a Markov chain are of the same type — either all positive-recurrent, all null-recurrent, or all transient.*

Proof: Let j be a recurrent state. From Theorem 5.1.1, all states in a class are recurrent or all are transient. Next suppose that j is positive-recurrent, so that $1/\overline{T}_{jj} > 0$. Let i be in the same class as j, and consider the renewal-reward process on $\{N_{jj}(t); t \geq 0\}$ for which $R(t) = 1$ whenever the process is in state i (*i.e.*, if $X_n = i$, then $R(t) = 1$ for $n \leq t < n+1$). The reward is 0 whenever the process is in some state other than i. Let $\mathsf{E}[\mathrm{R}_n]$ be the expected reward in an inter-renewal interval; this must be positive since i is accessible from j. From the strong law for renewal-reward processes, Theorem 4.4.1,

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t R(\tau)d\tau = \frac{\mathsf{E}[\mathrm{R}_n]}{\overline{T}_{jj}} \qquad \text{with probability 1.}$$

The term on the left is the time-average number of transitions into state i, given $X_0 = j$, and this is $1/\overline{T}_{ii}$ from (5.15). Since $\mathsf{E}[\mathrm{R}_n] > 0$ and $\overline{T}_{jj} < \infty$, we have $1/\overline{T}_{ii} > 0$, so i is positive-recurrent. Thus if one state is positive-recurrent, the entire class is, completing the proof. □

If all of the states in a Markov chain are in a null-recurrent class, then $1/\overline{T}_{jj} = 0$ for each state, and one might think of $1/\overline{T}_{jj} = 0$ as a "steady-state" probability for j in the sense that 0 is both the time-average rate of occurrence of j and the limiting probability of j. However, these "probabilities" do not add up to 1, so a steady-state probability *distribution* does not exist. This appears rather paradoxical at first, but the example of Figure 5.2, with $p = 1/2$ will help to clarify the situation. As time n increases (starting in state i, say), the random variable X_n spreads out over more and more states around i, and thus is less likely to be in each individual state. For each j, $\lim_{n\to\infty} P_{ij}(n) = 0$. Thus, $\sum_j \{\lim_{n\to\infty} P_{ij}^n\} = 0$. On the other hand, for every n, $\sum_j P_{ij}^n = 1$. This is one of those unusual examples where a limit and a sum cannot be interchanged.

In Chapter 3, we defined the steady-state distribution of a finite-state Markov chain as a probability vector $\boldsymbol{\pi}$ that satisfies $\boldsymbol{\pi} = \boldsymbol{\pi}[P]$. Here we define $\{\pi_i; i \geq 0\}$ in the same way, as a set of numbers that satisfy

$$\pi_j = \sum_i \pi_i P_{ij} \quad \text{for all } j; \qquad \pi_j \geq 0 \quad \text{for all } j; \qquad \sum_j \pi_j = 1. \tag{5.18}$$

Suppose that a set of numbers $\{\pi_i; i \geq 0\}$ satisfying (5.18) is chosen as the initial probability distribution for a Markov chain, *i.e.*, if $\Pr\{X_0 = i\} = \pi_i$ for all i. Then $\Pr\{X_1 = j\} = \sum_i \pi_i P_{ij} = \pi_j$ for all j, and, by induction, $\Pr\{X_n = j\} = \pi_j$ for all j and all $n \geq 0$. The fact that $\Pr\{X_n = j\} = \pi_j$ for all j motivates the definition of steady-state distribution above. Theorem 5.1.2 showed that $1/\overline{T}_{jj}$ is a 'steady-state' probability for state j, both in a time-average and a limiting ensemble-average sense. The following theorem brings these ideas together. An irreducible Markov chain is a Markov chain in which all pairs of states communicate. For finite-state chains, irreducibility implied a single class of recurrent states, whereas for countably infinite chains, an irreducible chain is a single class that can be transient, null-recurrent, or positive-recurrent.

Theorem 5.1.4. *Assume an irreducible Markov chain with transition probabilities $\{P_{ij}\}$. If (5.18) has a solution, then the solution is unique, $\pi_i = 1/\overline{T}_{ii} > 0$ for all $i \geq 0$, and the states are positive-recurrent. Also, if the states are positive-recurrent then (5.18) has a solution.*

Proof*: Let $\{\pi_j; j \geq 0\}$ satisfy (5.18) and be the initial distribution of the Markov chain, *i.e.*, $\Pr\{X_0{=}j\} = \pi_j$, $j \geq 0$. Then, as shown above, $\Pr\{X_n{=}j\} = \pi_j$ for all $n \geq 0$, $j \geq 0$. Let $\tilde{N}_j(t)$ be the number of occurrences of any given state j from time 1 to t. Equating $\Pr\{X_n{=}j\}$ to the expectation of an occurrence of j at time n, we have,

$$(1/t)\mathsf{E}\left[\tilde{N}_j(t)\right] = (1/t) \sum_{1 \leq n \leq t} \Pr\{X_n{=}j\} = \pi_j \qquad \text{for all integers } t \geq 1.$$

Conditioning this on the possible starting states i, and using the counting processes $\{N_{ij}(t); t \geq 0\}$ defined earlier,

$$\pi_j = (1/t)\mathsf{E}\left[\tilde{N}_j(t)\right] = \sum_i \pi_i \mathsf{E}\left[N_{ij}(t)/t\right] \quad \text{for all integer } t \geq 1. \tag{5.19}$$

For any given state i, let T_{ij} be the time of the first occurrence of state j given $X_0 = i$. Then if $T_{ij} < \infty$, we have $N_{ij}(t) \leq N_{ij}(T_{ij} + t)$. Thus, for all $t \geq 1$,

$$\mathsf{E}\left[N_{ij}(t)\right] \leq \mathsf{E}\left[N_{ij}(T_{ij} + t)\right] = 1 + \mathsf{E}\left[N_{jj}(t)\right]. \tag{5.20}$$

The last step follows since the process is in state j at time T_{ij}, and the expected number of occurrences of state j in the next t steps is $\mathsf{E}\left[N_{jj}(t)\right]$.

Substituting (5.20) in (5.19) for each i, $\pi_j \leq 1/t + \mathsf{E}\left[N_{jj}(t)/t\right]$. Taking the limit as $t \to \infty$ and using (5.16), $\pi_j \leq \lim_{t\to\infty} \mathsf{E}\left[N_{jj}(t)/t\right]$. Since $\sum_i \pi_i = 1$, there is at least one value of j for which $\pi_j > 0$, and for this j, $\lim_{t\to\infty} \mathsf{E}\left[N_{jj}(t)/t\right] > 0$, and consequently $\lim_{t\to\infty} \mathsf{E}\left[N_{jj}(t)\right] = \infty$. Thus, from Lemma 5.1.1, state j is recurrent, and from Theorem 5.1.2, j is positive-recurrent. From Theorem 5.1.3, all states are then positive-recurrent. For any j and any integer M, (5.19) implies that

$$\pi_j \geq \sum_{i \leq M} \pi_i \mathsf{E}\left[N_{ij}(t)/t\right] \quad \text{for all } t. \tag{5.21}$$

From Theorem 5.1.2, $\lim_{t\to\infty} \mathsf{E}\left[N_{ij}(t)/t\right] = 1/\overline{T}_{jj}$ for all i. Substituting this into (5.21), we get $\pi_j \geq 1/\overline{T}_{jj} \sum_{i \leq M} \pi_i$. Since M is arbitrary, $\pi_j \geq 1/\overline{T}_{jj}$. Since we already showed that $\pi_j \leq \lim_{t\to\infty} \mathsf{E}\left[N_{jj}(t)/t\right] = 1/\overline{T}_{jj}$, we have $\pi_j = 1/\overline{T}_{jj}$ for all j. This shows both that $\pi_j > 0$ for all j and that the solution to (5.18) is unique. Exercise 5.5 completes the proof by showing that if the states are positive-recurrent, then choosing $\pi_j = 1/\overline{T}_{jj}$ for all j satisfies (5.18). □

In practice, it is usually easy to see whether a chain is irreducible. We shall also see by a number of examples that the steady-state distribution can often be calculated from (5.18). Theorem 5.1.4 then says that the calculated distribution is unique and that its existence guarantees that the chain is positive recurrent.

Example 5.1.3. Age of a renewal process: Consider a renewal process $\{N(t); t > 0\}$ in which the inter-renewal random variables $\{W_n; n \geq 1\}$ are arithmetic with span 1. We will use a Markov chain to model the age of this process (see Figure 5.3). The probability that a renewal occurs at a particular integer time depends on the past only through the integer time back to the last renewal. The state of the Markov chain during a unit interval

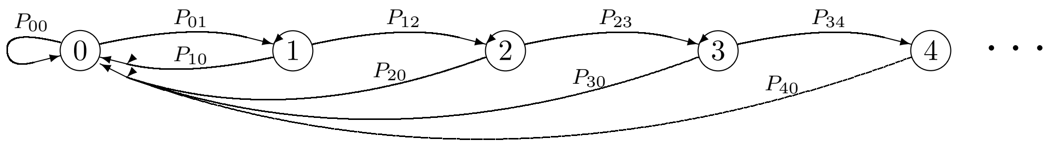

Figure 5.3: A Markov chain model of the age of a renewal process.

will be taken as the age of the renewal process at the beginning of the interval. Thus, each unit of time, the age either increases by one or a renewal occurs and the age decreases to 0 (*i.e.*, if a renewal occurs at time t, the age at time t is 0).

$\Pr\{W > n\}$ is the probability that an inter-renewal interval lasts for more than n time units. We assume that $\Pr\{W > 0\} = 1$, so that each renewal interval lasts at least one time unit. The probability $P_{n,0}$ in the Markov chain is the probability that a renewal interval has duration $n+1$, given that the interval exceeds n. Thus, for example, P_{00} is the probability that the renewal interval is equal to 1. $P_{n,n+1}$ is $1 - P_{n,0}$, which is $\Pr\{W > n+1\} / \Pr\{W > n\}$. We can then solve for the steady state probabilities in the chain: for $n > 0$,

$$\pi_n = \pi_{n-1} P_{n-1,n} = \pi_{n-2} P_{n-2,n-1} P_{n-1,n} = \pi_0 P_{0,1} P_{1,2} \dots P_{n-1,n}.$$

The first equality above results from the fact that state n, for $n > 0$ can be entered only from state $n-1$. The subsequent equalities come from substituting in the same expression for π_{n-1}, then p_{n-2}, and so forth.

$$\pi_n = \pi_0 \frac{\Pr\{W > 1\} \Pr\{W > 2\}}{\Pr\{W > 0\} \Pr\{W > 1\}} \cdots \frac{\Pr\{W > n\}}{\Pr\{W > n-1\}} = \pi_0 \Pr\{W > n\}. \tag{5.22}$$

We have cancelled out all the cross terms above and used the fact that $\Pr\{W > 0\} = 1$. Another way to see that $\pi_n = \pi_0 \Pr\{W > n\}$ is to observe that state 0 occurs exactly once in each inter-renewal interval; state n occurs exactly once in those inter-renewal intervals of duration n or more.

Since the steady-state probabilities must sum to 1, (5.22) can be solved for π_0 as

$$\pi_0 = \frac{1}{\sum_{n=0}^{\infty} \Pr\{W > n\}} = \frac{1}{\mathsf{E}[W]}. \tag{5.23}$$

The second equality follows by expressing $\mathsf{E}[W]$ as the integral of the complementary distribution function of W. Combining this with (5.22), the steady-state probabilities for $n \geq 0$ are

$$\pi_n = \frac{\Pr\{W > n\}}{\mathsf{E}[W]}. \tag{5.24}$$

In terms of the renewal process, π_n is the probability that, at some large integer time, the age of the process will be n. Note that if the age of the process at an integer time is n, then the age increases toward $n+1$ at the next integer time, at which point it either drops to 0 or continues to rise. Thus π_n can be interpreted as the fraction of time that the age of the process is between n and $n+1$. Recall from (4.28) (and the fact that residual life and age are equally distributed) that the distribution function of the time-average age is given by $\mathsf{F}_Z(n) = \int_0^n \Pr\{W > w\}\, dw / \mathsf{E}[W]$. Thus, the probability that the age is between n and $n+1$ is $\mathsf{F}_Z(n+1) - \mathsf{F}_Z(n)$. Since W is an integer random variable, this is $\Pr\{W > n\} / \mathsf{E}[W]$ in agreement with our result here.

The analysis here gives a new, and intuitively satisfying, explanation of why the age of a renewal process is so different from the inter-renewal time. The Markov chain shows the ever increasing loops that give rise to large expected age when the inter-renewal time is heavy tailed (*i.e.*, has a distribution function that goes to 0 slowly with increasing time). These loops can be associated with the isosceles triangles of Figure 4.7. The advantage here is that we can associate the states with steady-state probabilities if the chain is recurrent. Even when the Markov chain is null-recurrent (*i.e.*, the associated renewal process has infinite expected age), it seems easier to visualize the phenomenon of infinite expected age.

5.2 Birth-death Markov chains

A *birth-death Markov chain* is a Markov chain in which the state space is the set of nonnegative integers; for all $i \geq 0$, the transition probabilities satisfy $P_{i,i+1} > 0$ and $P_{i+1,i} > 0$, and for all $|i-j| > 1$, $P_{ij} = 0$ (see Figure 5.4). A transition from state i to $i+1$ is regarded as a birth and one from $i+1$ to i as a death. Thus the restriction on the transition probabilities means that only one birth or death can occur in one unit of time. Many applications of birth-death processes arise in queueing theory, where the state is the number of customers, births are customer arrivals, and deaths are customer departures. The restriction to only one arrival or departure at a time seems rather peculiar, but usually such a chain is a finely sampled approximation to a continuous-time process, and the time increments are then small enough that multiple arrivals or departures in a time increment are unlikely and can be ignored in the limit.

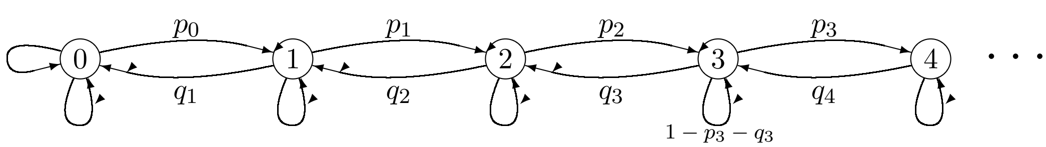

Figure 5.4: Birth-death Markov chain.

We denote $P_{i,i+1}$ by p_i and $P_{i,i-1}$ by q_i. Thus $P_{ii} = 1 - p_i - q_i$. There is an easy way to find the steady-state probabilities of these birth-death chains. In any sample function of the process, note that the number of transitions from state i to $i+1$ differs by at most 1 from the number of transitions from $i+1$ to i. If the process starts to the left of i and ends to the right, then one more $i \to i+1$ transition occurs than $i+1 \to i$, etc. Thus if we visualize a renewal-reward process with renewals on occurrences of state i and unit reward on transitions from state i to $i+1$, the limiting time-average number of transitions per unit time is $\pi_i\, p_i$. Similarly, the limiting time-average number of transitions per unit time from

$i+1$ to i is $\pi_{i+1}q_{i+1}$. Since these two must be equal in the limit,

$$\pi_i p_i = \pi_{i+1}q_{i+1} \qquad \text{for } i \geq 0. \tag{5.25}$$

The intuition in (5.25) is simply that the rate at which downward transitions occur from $i+1$ to i must equal the rate of upward transitions. Since this result is very important, both here and in our later study of continuous-time birth-death processes, we show that (5.25) also results from using the steady-state equations in (5.18):

$$\pi_i = p_{i-1}\pi_{i-1} + (1-p_i-q_i)\pi_i + q_{i+1}\pi_{i+1}; \quad i > 0 \tag{5.26}$$

$$\pi_0 = (1-p_0)\pi_0 + q_1\pi_1. \tag{5.27}$$

From (5.27), $p_0\pi_0 = q_1\pi_1$. To see that (5.25) is satisfied for $i > 0$, we use induction on i, with $i = 0$ as the base. Thus assume, for a given i, that $p_{i-1}\pi_{i-1} = q_i\pi_i$. Substituting this in (5.26), we get $p_i\pi_i = q_{i+1}\pi_{i+1}$, thus completing the inductive proof.

It is convenient to define ρ_i as p_i/q_{i+1}. Then we have $\pi_{i+1} = \rho_i\pi_i$, and iterating this,

$$\pi_i = \pi_0 \prod_{j=0}^{i-1} \rho_j; \quad \pi_0 = \frac{1}{1+\sum_{i=1}^{\infty}\ \prod_{j=0}^{i-1}\ \rho_j}. \tag{5.28}$$

If $\sum_{i\geq 1}\prod_{0\leq j<i}\rho_j < \infty$, then π_0 is positive and all the states are positive-recurrent. If this sum of products is infinite, then no state is positive-recurrent. If ρ_j is bounded below 1, say $\rho_j \leq 1-\epsilon$ for some fixed $e > 0$ and all sufficiently large j, then this sum of products will converge and the states will be positive-recurrent.

For the simple birth-death process of Figure 5.2, if we define $\rho = q/p$, then $\rho_j = \rho$ for all j. For $\rho < 1$, (5.28) simplifies to $\pi_i = \pi_o\rho^i$ for all $i \geq 0$, $\pi_0 = 1-\rho$, and thus $\pi_i = (1-\rho)\rho^i$ for $i \geq 0$. Exercise 5.2 shows how to find $\mathsf{F}_{ij}(\infty)$ for all i, j in the case where $\rho \geq 1$. We have seen that the simple birth-death chain of Figure 5.2 is transient if $\rho > 1$. This is not necessarily so in the case where self-transitions exist, but the chain is still either transient or null-recurrent. An example of this will arise in Exercise 6.3.

5.3 Reversible Markov chains

Many important Markov chains have the property that, in steady state, the sequence of states looked at backwards in time, *i.e.*,$\dots X_{n+1}, X_n, X_{n-1}, \dots$, has the same probabilistic structure as the sequence of states running forward in time. This equivalence between the forward chain and backward chain leads to a number of results that are intuitively quite surprising and that are quite difficult to derive without using this equivalence. We shall study these results here and then extend them in Chapter 6 to Markov processes with a discrete state space. This set of ideas, and its use in queueing and queueing networks, has been an active area of queueing research over many years . It leads to many simple results for systems that initially look very complex. We only scratch the surface here and refer the interested reader to [13] for a more comprehensive treatment. Before going into reversibility, we describe the backward chain for an arbitrary Markov chain.

The defining characteristic of a Markov chain $\{X_n; n \geq 0\}$ is that for all $n \geq 0$,

$$\Pr\{X_{n+1} \mid X_n, X_{n-1}, \dots, X_0\} = \Pr\{X_{n+1} \mid X_n\}. \tag{5.29}$$

For homogeneous chains, which we have been assuming throughout, $\Pr\{X_{n+1} = j \mid X_n = i\} = P_{ij}$, independent of n. For any $k > 1$, we can extend (5.29) to get

$$\begin{aligned}
&\Pr\{X_{n+k}, X_{n+k-1}, \dots, X_{n+1} \mid X_n, X_{n-1}, \dots, X_0\} \\
&\quad = \Pr\{X_{n+k} \mid X_{n+k-1}\} \Pr\{X_{n+k-1} \mid X_{n+k-2}\} \dots \Pr\{X_{n+1} \mid X_n\} \\
&\quad = \Pr\{X_{n+k}, X_{n+k-1}, \dots, X_{n+1} \mid X_n\}.
\end{aligned} \tag{5.30}$$

By letting A^+ be any event defined on the states X_{n+1} to X_{n+k} and letting A^- be any event defined on X_0 to X_{n-1}, this can be written more succinctly as

$$\Pr\big\{A^+ \mid X_n, A^-\big\} = \Pr\big\{A^+ \mid X_n\big\}. \tag{5.31}$$

This says that, given state X_n, any future event A^+ is statistically independent of any past event A^-. This result, namely that past and future are independent given the present state, is equivalent to (5.29) for defining a Markov chain, but it has the advantage of showing the symmetry between past and future. This symmetry is best brought out by multiplying both sides of (5.31) by $\Pr\{A^- \mid X_n\}$, obtaining[2]

$$\Pr\big\{A^+, A^- \mid X_n\big\} = \Pr\big\{A^+ \mid X_n\big\} \Pr\big\{A^- \mid X_n\big\}. \tag{5.32}$$

This symmetric form says that, conditional on the current state, the past and future states are statistically independent. Dividing both sides by $\Pr\{A^+ \mid X_n\}$ then yields

$$\Pr\big\{A^- \mid X_n, A^+\big\} = \Pr\big\{A^- \mid X_n\big\}. \tag{5.33}$$

By letting A^- be X_{n-1} and A^+ be $X_{n+1}, X_{n+2}, \dots, X_{n+k}$, this becomes

$$\Pr\{X_{n-1} \mid X_n, X_{n+1}, \dots, X_{n+k}\} = \Pr\{X_{n-1} \mid X_n\}.$$

This is the equivalent form to (5.29) for the backward chain, and says that the backward chain is also a Markov chain. By Bayes' law, $\Pr\{X_{n-1} \mid X_n\}$ can be evaluated as

$$\Pr\{X_{n-1} \mid X_n\} = \frac{\Pr\{X_n \mid X_{n-1}\} \Pr\{X_{n-1}\}}{\Pr\{X_n\}}. \tag{5.34}$$

Since the distribution of X_n can vary with n, $\Pr\{X_{n-1} \mid X_n\}$ can also depend on n. *Thus the backward Markov chain is not necessarily homogeneous.* This should not be surprising, since the forward chain was defined with some arbitrary distribution for the initial state at time 0. This initial distribution was not relevant for equations (5.29) to (5.31), but as soon as $\Pr\{A^- \mid X_n\}$ was introduced, the initial state implicitly became a part of each equation and destroyed the symmetry between past and future. For a chain in steady state, however, $\Pr\{X_n = j\} = \Pr\{X_{n-1} = j\} = \pi_j$ for all j, and we have

$$\Pr\{X_{n-1} = j \mid X_n = i\} = P_{ji}\pi_j/\pi_i. \tag{5.35}$$

[2]Much more broadly, any 3 events, say A^-, X_0, A^+ are said to be Markov if $\Pr\big\{A^+ \mid X_0 A^-\big\} = \Pr\big\{A^+ \mid X_0\big\}$, and this implies the more symmetric form $\Pr\big\{A^- A^+ \mid X_0)\big\} = \Pr\big\{A^- \mid X_0\big\} \Pr\big\{A^+ \mid X_0\big\}$.

Thus the backward chain is homogeneous if the forward chain is in steady state. For a chain with steady-state probabilities $\{\pi_i;\ i \geq 0\}$, we define the backward transition probabilities P^*_{ij} as

$$\pi_i P^*_{ij} = \pi_j P_{ji}. \tag{5.36}$$

From (5.34), the backward transition probability P^*_{ij}, for a Markov chain in steady state, is then equal to $\Pr\{X_{n-1} = j \mid X_n = i\}$, the probability that the previous state is j given that the current state is i.

Now consider a new Markov chain with transition probabilities $\{P^*_{ij}\}$. Over some segment of time for which both this new chain and the old chain are in steady state, the set of states generated by the new chain is statistically indistinguishable from the backward running sequence of states from the original chain. It is somewhat simpler, in talking about forward and backward running chains, however, to visualize Markov chains running in steady state from $t = -\infty$ to $t = +\infty$. If one is uncomfortable with this, one can also visualize starting the Markov chain at some very negative time with the initial distribution equal to the steady-state distribution.

Definition 5.3.1. *A Markov chain that has steady-state probabilities* $\{\pi_i;\ i \geq 0\}$ *is reversible if* $P_{ij} = \pi_j P_{ji}/\pi_i$ *for all* i, j, i.e., *if* $P^*_{ij} = P_{ij}$ *for all* i, j.

Thus the chain is reversible if, in steady state, the backward running sequence of states is statistically indistinguishable from the forward running sequence. Comparing (5.36) with the steady-state equations (5.25) that we derived for birth-death chains, we have the important theorem:

Theorem 5.3.1. *Every birth-death chain with a steady-state probability distribution is reversible.*

We saw that for birth-death chains, the equation $\pi_i P_{ij} = \pi_j P_{ji}$ (which only had to be considered for $|i-j| \leq 1$) provided a very simple way of calculating the steady-state probabilities. Unfortunately, it appears that we must first calculate the steady-state probabilities in order to show that a chain is reversible. The following simple theorem gives us a convenient escape from this dilemma.

Theorem 5.3.2. *Assume that an irreducible Markov chain has transition probabilities* $\{P_{ij}\}$. *Suppose* $\{\pi_i\}$ *is a set of positive numbers summing to* 1 *and satisfying*

$$\pi_i P_{ij} = \pi_j P_{ji}; \quad \textit{all } i, j. \tag{5.37}$$

then, first, $\{\pi_i; i \geq 0\}$ *is the steady-state distribution for the chain, and, second, the chain is reversible.*

Proof: Given a solution to (5.37) for all i and j, we can sum this equation over i for each j.

$$\sum_i \}\pi_i P_{ij} = \pi_j \sum_i \}P_{ji} = \pi_j. \tag{5.38}$$

Thus the solution to (5.37), along with the constraints $\pi_i > 0$, $\sum_i \pi_i = 1$, satisfies the steady-state equations, (5.18), and, from Theorem 5.1.4, this is the unique steady-state distribution. Since (5.37) is satisfied, the chain is also reversible.

It is often possible, sometimes by using an educated guess, to find a solution to (5.37). If this is successful, then we are assured both that the chain is reversible and that the actual steady-state probabilities have been found.

Note that the theorem applies to periodic chains as well as to aperiodic chains. If the chain is periodic, then the steady-state probabilities have to be interpreted as average values over the period, but from Theorem 5.1.4 shows that (5.38) still has a unique solution (assuming an irreducible chain). On the other hand, for a chain with period $d > 1$, there are d subclasses of states and the sequence $\{X_n\}$ must rotate between these classes in a fixed order. For this same order to be followed in the backward chain, the only possibility is $d = 2$. Thus periodic chains with periods other than 2 cannot be reversible.

There are several simple tests that can be used to show that some given irreducible chain is not reversible. First, the steady-state probabilities must satisfy $\pi_i > 0$ for all i, and thus, if $P_{ij} > 0$ but $P_{ji} = 0$ for some i, j, then (5.37) cannot be satisfied and the chain is not reversible. Second, consider any set of three states, i, j, k. If $P_{ji}P_{ik}P_{kj}$ is unequal to $P_{jk}P_{ki}P_{ij}$ then the chain cannot be reversible. To see this, note that (5.37) requires that

$$\pi_i = \pi_j P_{ji}/P_{ij} = \pi_k P_{ki}/P_{ik}.$$

Thus, $\pi_j P_{ji} P_{ik} = \pi_k P_{ki} P_{ij}$. Equation (5.37) also requires that $\pi_j P_{jk} = \pi_k P_{kj}$. Taking the ratio of these equations, we see that $P_{ji}P_{ik}P_{kj} = P_{jk}P_{ki}P_{ij}$. Thus if this equation is not satisfied, the chain cannot be reversible. In retrospect, this result is not surprising. What it says is that for any cycle of three states, the probability of three transitions going around the cycle in one direction must be the same as the probability of going around the cycle in the opposite (and therefore backwards) direction.

It is also true (see [16] for a proof), that a necessary and sufficient condition for a chain to be reversible is that the product of transition probabilities around any cycle of arbitrary length must be the same as the product of transition probabilities going around the cycle in the opposite direction. This doesn't seem to be a widely useful way to demonstrate reversibility.

There is another result, generalizing Theorem 5.3.2, for finding the steady-state probabilities of an arbitrary Markov chain and simultaneously finding the transition probabilities of the backward chain.

Theorem 5.3.3. *Assume that an irreducible Markov chain has transition probabilities* $\{P_{ij}\}$*. Suppose* $\{\pi_i\}$ *is a set of positive numbers summing to* 1 *and that* $\{P^*_{ij}\}$ *is a set of transition probabilities satisfying*

$$\pi_i P_{ij} = \pi_j P^*_{ji}; \quad \text{all } i, j. \tag{5.39}$$

Then $\{\pi_i\}$ *is the steady-state distribution and* $\{P^*_{ij}\}$ *is the set of transition probabilities for the backward chain.*

Proof: Summing (5.39) over i, we get the steady-state equations for the Markov chain, so the fact that the given $\{\pi_i\}$ satisfy these equations asserts that they are the steady-state probabilities. Equation (5.39) then asserts that $\{P^*_{ij}\}$ is the set of transition probabilities for the backward chain.

The following two sections illustrate some important applications of reversibility.

5.4 The M/M/1 sample-time Markov chain

The M/M/1 Markov chain is a sampled-time model of the M/M/1 queueing system. Recall that the M/M/1 queue has Poisson arrivals at some rate λ and IID exponentially distributed service times at some rate μ. We assume throughout this section that $\lambda < \mu$ (this is required to make the states positive-recurrent). For some given small increment of time δ, we visualize observing the state of the system at the sample times $n\delta$. As indicated in Figure 5.5, the probability of an arrival in the interval from $(n-1)\delta$ to $n\delta$ is modeled as $\lambda\delta$, independent of the state of the chain at time $(n-1)\delta$ and thus independent of all prior arrivals and departures. Thus the arrival process, viewed as arrivals in subsequent intervals of duration δ, is Bernoulli, thus approximating the Poisson arrivals. This is a sampled-time approximation to the Poisson arrival process of rate λ for a continuous-time M/M/1 queue.

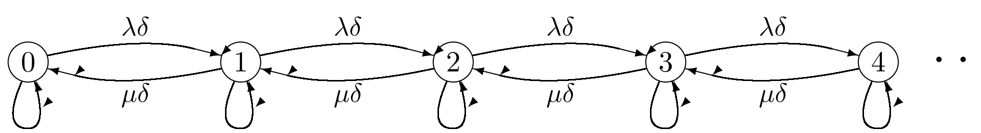

Figure 5.5: Sampled-time approximation to M/M/1 queue for time increment δ.

When the system is non-empty (*i.e.*, the state of the chain is one or more), the probability of a departure in the interval $(n-1)\delta$ to $n\delta$ is $\mu\delta$, thus modelling the exponential service times. When the system is empty, of course, departures cannot occur.

Note that in our sampled-time model, there can be at most one arrival or departure in an interval $(n-1)\delta$ to $n\delta$. As in the Poisson process, the probability of more than one arrival, more than one departure, or both an arrival and a departure in an increment δ is of order δ^2 for the actual continuous-time M/M/1 system being modeled. Thus, for δ very small, we expect the sampled-time model to be relatively good. At any rate, we can now analyze the model with no further approximations.

Since this chain is a birth-death chain, we can use (5.28) to determine the steady-state probabilities; they are

$$\pi_i = \pi_0 \rho^i \;;\; \rho = \lambda/\mu < 1.$$

Setting the sum of the π_i to 1, we find that $\pi_0 = 1 - \rho$, so

$$\pi_i = (1-\rho)\rho^i \;; \qquad \text{all } i \geq 0. \tag{5.40}$$

Thus the steady-state probabilities exist and the chain is a birth-death chain, so from Theorem 5.3.1, it is reversible. We now exploit the consequences of reversibility to find some rather surprising results about the M/M/1 chain in steady state. Figure 5.6 illustrates a sample path of arrivals and departures for the chain. To avoid the confusion associated with the backward chain evolving backward in time, we refer to the original chain as the chain moving to the right and to the backward chain as the chain moving to the left.

There are two types of correspondence between the right-moving and the left-moving chain:

1. The left-moving chain has the same Markov chain description as the right-moving chain, and thus can be viewed as an M/M/1 chain in its own right. We still label the sampled-time intervals from left to right, however, so that the left-moving chain makes transitions from X_{n+1} to X_n to X_{n-1}. Thus, for example, if $X_n = i$ and $X_{n-1} = i+1$, the left-moving chain has an arrival in the interval from $n\delta$ to $(n-1)\delta$.

2. Each sample function $\dots x_{n-1}, x_n, x_{n+1} \dots$ of the right-moving chain corresponds to the same sample function $\dots x_{n+1}, x_n, x_{n-1} \dots$ of the left-moving chain, where $X_{n-1} = x_{n-1}$ is to the left of $X_n = x_n$ for both chains. With this correspondence, an arrival to the right-moving chain in the interval $(n-1)\delta$ to $n\delta$ is a departure from the left-moving chain in the interval $n\delta$ to $(n-1)\delta$, and a departure from the right-moving chain is an arrival to the left-moving chain. Using this correspondence, each event in the left-moving chain corresponds to some event in the right-moving chain.

In each of the properties of the M/M/1 chain to be derived below, a property of the left-moving chain is developed through correspondence 1 above, and then that property is translated into a property of the right-moving chain by correspondence 2.

Property 1: Since the arrival process of the right-moving chain is Bernoulli, the arrival process of the left-moving chain is also Bernoulli (by correspondence 1). Looking at a sample function x_{n+1}, x_n, x_{n-1} of the left-moving chain (*i.e.*, using correspondence 2), an arrival in the interval $n\delta$ to $(n-1)\delta$ of the left-moving chain is a departure in the interval $(n-1)\delta$ to $n\delta$ of the right-moving chain. Since the arrivals in successive increments of the left-moving chain are independent and have probability $\lambda\delta$ in each increment δ, we conclude that departures in the right-moving chain are similarly Bernoulli.

The fact that the departure process is Bernoulli with departure probability $\lambda\delta$ in each increment is surprising. Note that the probability of a departure in the interval $(n\delta - \delta, n\delta]$ is $\mu\delta$ conditional on $X_{n-1} \geq 1$ and is 0 conditional on $X_{n-1} = 0$. Since $\Pr\{X_{n-1} \geq 1\} = 1 - \Pr\{X_{n-1} = 0\} = \rho$, we see that the unconditional probability of a departure in the interval $(n\delta - \delta, n\delta]$ is $\rho\mu\delta = \lambda\delta$ as asserted above. The fact that successive departures are independent is much harder to derive without using reversibility (see exercise 5.13).

Property 2: In the original (right-moving) chain, arrivals in the time increments after $n\delta$ are independent of X_n. Thus, for the left-moving chain, arrivals in time increments to the left of $n\delta$ are independent of the state of the chain at $n\delta$. From the correspondence between sample paths, however, a left chain arrival is a right chain departure, so that for the right-moving chain, departures in the time increments prior to $n\delta$ are independent of X_n, which is equivalent to saying that the state X_n is independent of the prior departures.

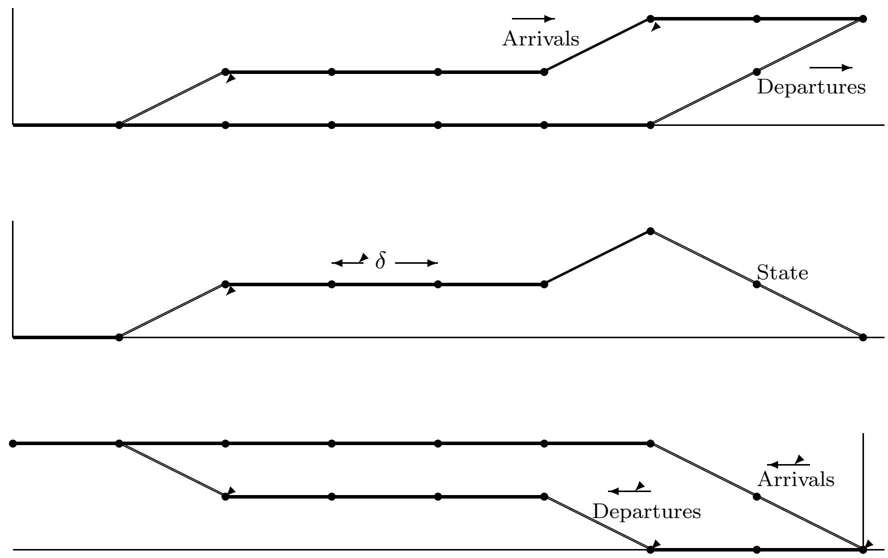

Figure 5.6: Sample function of M/M/1 chain over a busy period and corresponding arrivals and departures for right and left-moving chains. Arrivals and departures are viewed as occurring between the sample times, and an arrival in the left-moving chain between time $n\delta$ and $(n+1)\delta$ corresponds to a departure in the right-moving chain between $(n+1)\delta$ and $n\delta$.

This means that if one observes the departures prior to time $n\delta$, one obtains no information about the state of the chain at $n\delta$. This is again a surprising result. To make it seem more plausible, note that an unusually large number of departures in an interval from $(n-m)\delta$ to $n\delta$ indicates that a large number of customers were probably in the system at time $(n-m)\delta$, but it doesn't appear to say much (and in fact it says exactly nothing) about the number remaining at $n\delta$.

The following theorem summarizes these results.

Theorem 5.4.1 (Burke's theorem for sampled-time). *Given an M/M/1 Markov chain in steady state with $\lambda < \mu$,*

a) the departure process is Bernoulli,

b) the state X_n at any time $n\delta$ is independent of departures prior to $n\delta$.

The proof of Burke's theorem above did not use the fact that the departure probability is the same for all states except state 0. Thus these results remain valid for any birth-death chain with Bernoulli arrivals that are independent of the current state (*i.e.*, for which $P_{i,i+1} = \lambda\delta$ for all $i \geq 0$). One important example of such a chain is the sampled time approximation to an M/M/m queue. Here there are m servers, and the probability of departure from state i in an increment δ is $\mu i\delta$ for $i \leq m$ and $\mu m\delta$ for $i > m$. For the states to be recurrent, and thus for a steady state to exist, λ must be less than μm. Subject to this restriction, properties a) and b) above are valid for sampled-time M/M/m queues.

5.5 Branching processes

Branching processes provide a simple model for studying the population of various types of individuals from one generation to the next. The individuals could be photons in a photomultiplier, particles in a cloud chamber, micro-organisms, insects, or branches in a data structure.

Let X_n be the number of individuals in generation n of some population. Each of these X_n individuals, independently of each other, produces a random number of offspring, and these offspring collectively make up generation $n+1$. More precisely, a *branching process* is a Markov chain in which the state X_n at time n models the number of individuals in generation n. Denote the individuals of generation n as $\{1, 2, ..., X_n\}$ and let $Y_{k,n}$ be the number of offspring of individual k. The random variables $Y_{k,n}$ are defined to be IID over k and n, with a PMF $p_j = \Pr\{Y_{k,n} = j\}$. The state at time $n+1$, namely the number of individuals in generation $n+1$, is

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{k,n}. \tag{5.41}$$

Assume a given distribution (perhaps deterministic) for the initial state X_0. The transition probability, $P_{ij} = \Pr\{X_{n+1} = j \mid X_n = i\}$, is just the probability that $Y_{1,n}+Y_{2,n}+\cdots+Y_{i,n} = j$. The zero state (i e., the state in which there are *no* individuals) is a trapping state (*i.e.*, $P_{00} = 1$) since no future offspring can arise in this case.

One of the most important issues about a branching process is the probability that the population dies out eventually. Naturally, if p_0 (the probability that an individual has no offspring) is zero, then each generation must be at least as large as the generation before, and the population cannot die out unless $X_0 = 0$. We assume in what follows that $p_0 > 0$ and $X_0 > 0$. Recall that $\mathsf{F}_{ij}(n)$ was defined as the probability, given $X_0 = i$, that state j is entered between times 1 and n. From (5.8), this satisfies the iterative relation

$$\mathsf{F}_{ij}(n) = P_{ij} + \sum_{k \neq j} P_{ik}\mathsf{F}_{kj}(n-1), n > 1; \quad \mathsf{F}_{ij}(1) = P_{ij}. \tag{5.42}$$

The probability that the process dies out by time n or before, given $X_0 = i$, is thus $\mathsf{F}_{i0}(n)$. For the n^{th} generation to die out, starting with an initial population of i individuals, the descendants of each of those i individuals must die out. Since each individual generates descendants independently, we have $\mathsf{F}_{i0}(n) = [\mathsf{F}_{10}(n)]^i$ for all i and n. Because of this relationship, it is sufficient to find $\mathsf{F}_{10}(n)$, which can be determined from (5.42). Observe that P_{1k} is just p_k, the probability that an individual will have k offspring. Thus, (5.42) becomes

$$\mathsf{F}_{10}(n) = p_0 + \sum_{k=1}^{\infty} p_k[\mathsf{F}_{10}(n-1)]^k = \sum_{k=0}^{\infty} p_k[\mathsf{F}_{10}(n-1)]^k. \tag{5.43}$$

Let $h(z) = \sum_k p_k z^k$ be the z transform of the number of an individual's offspring. Then (5.43) can be written as

$$\mathsf{F}_{10}(n) = h(\mathsf{F}_{10}(n-1)). \tag{5.44}$$

This iteration starts with $\mathsf{F}_{10}(1) = p_0$. Figure 5.7 shows a graphical construction for evaluating $\mathsf{F}_{10}(n)$. Having found $\mathsf{F}_{10}(n)$ as an ordinate on the graph for a given value of n, we find the same value as an abscissa by drawing a horizontal line over to the straight line of slope 1; we then draw a vertical line back to the curve $h(z)$ to find $h(\mathsf{F}_{10}(n)) = \mathsf{F}_{10}(n+1)$.

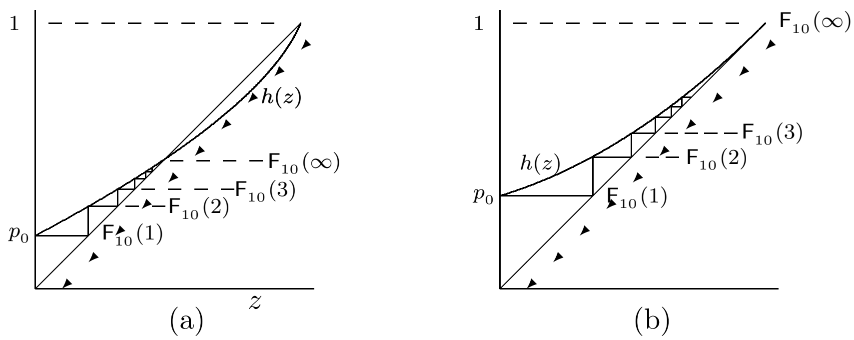

Figure 5.7: Graphical construction to find the probability that a population dies out. Here $\mathsf{F}_{10}(n)$ is the probability that a population starting with one member at generation 0 dies out by generation n or before. Thus $\mathsf{F}_{10}(\infty)$ is the probability that the population ever dies out.

For the two subfigures shown, it can be seen that $\mathsf{F}_{10}(\infty)$ is equal to the smallest root of the equation $h(z) - z = 0$. We next show that these two figures are representative of all possibilities. Since $h(z)$ is a z transform, we know that $h(1) = 1$, so that $z = 1$ is one root of $h(z) - z = 0$. Also, $h'(1) = \overline{Y}$, where $\overline{Y} = \sum_k kp_k$ is the expected number of an individual's offspring. If $\overline{Y} > 1$, as in Figure 5.7a, then $h(z) - z$ is negative for z slightly smaller than 1. Also, for $z = 0$, $h(z) - z = h(0) = p_0 > 0$. Since $h''(z) \geq 0$, there is exactly one root of $h(z) - z = 0$ for $0 < z < 1$, and that root is equal to $\mathsf{F}_{10}(\infty)$. By the same type of analysis, it can be seen that if $\overline{Y} \leq 1$, as in Figure 5.7b, then there is no root of $h(z) - z = 0$ for $z < 1$, and $\mathsf{F}_{10}(\infty) = 1$.

As we saw earlier, $\mathsf{F}_{i0}(\infty) = [\mathsf{F}_{10}(\infty)]^i$, so that for any initial population size, there is a probability strictly between 0 and 1 that successive generations eventually die out for $\overline{Y} > 1$, and probability 1 that successive generations eventually die out for $\overline{Y} \leq 1$. Since state 0 is accessible from all i, but $\mathsf{F}_{0i}(\infty) = 0$, it follows from Lemma 5.1.3 that all states other than state 0 are transient.

We next evaluate the expected number of individuals in a given generation. Conditional on $X_{n-1} = i$, (5.41) shows that the expected value of X_n is $i\overline{Y}$. Taking the expectation over X_{n-1}, we have

$$\mathsf{E}[X_n] = \overline{Y}\mathsf{E}[X_{n-1}]. \tag{5.45}$$

Iterating this equation, we get

$$\mathsf{E}[X_n] = \overline{Y}^n\mathsf{E}[X_0]. \tag{5.46}$$

Thus, if $\overline{Y} > 1$, the expected number of individuals in a generation increases exponentially with n, and $\overline{Y}$ gives the rate of growth. Physical processes do not grow exponentially forever, so branching processes are appropriate models of such physical processes only over some finite range of population. Even more important, the model here assumes that the number of offspring of a single member is independent of the total population, which is highly questionable in many areas of population growth. The advantage of an oversimplified model such as this is that it explains what would happen under these idealized conditions, thus providing insight into how the model should be changed for more realistic scenarios.

It is important to realize that, for branching processes, the mean number of individuals is not a good measure of the actual number of individuals. For $\overline{Y} = 1$ and $X_0 = 1$, the expected number of individuals in each generation is 1, but the probability that $X_n = 0$ approaches 1 with increasing n; this means that as n gets large, the n^{th} generation contains a large number of individuals with a very small probability and contains no individuals with a very large probability. For $\overline{Y} > 1$, we have just seen that there is a positive probability that the population dies out, but the expected number is growing exponentially.

A surprising result, which is derived from the theory of martingales in Chapter 7, is that if $X_0 = 1$ and $\overline{Y} > 1$, then the sequence of random variables $X_n/\overline{Y}^n$ has a limit with probability 1. This limit is a random variable; it has the value 0 with probability $\mathsf{F}_{10}(\infty)$, and has larger values with some given distribution. Intuitively, for large n, X_n is either 0 or very large. If it is very large, it tends to grow in an orderly way, increasing by a multiple of $\overline{Y}$ in each subsequent generation.

5.6 Round-robin and processor sharing

Typical queueing systems have one or more servers who each serve customers in FCFS order, serving one customer completely while other customers wait. These typical systems have larger average delay than necessary. For example, if two customers with service requirements of 10 and 1 units respectively are waiting when a single server becomes empty, then serving the first before the second results in departures at times 10 and 11, for an average delay of 10.5. Serving the customers in the opposite order results in departures at times 1 and 11, for an average delay of 6. Supermarkets have recognized this for years and have special express checkout lines for customers with small service requirements.

Giving priority to customers with small service requirements, however, has some disadvantages; first, customers with high service requirements can feel discriminated against, and second, it is not always possible to determine the service requirements of customers before they are served. The following alternative to priorities is popular both in the computer and data network industries. When a processor in a computer system has many jobs to accomplish, it often serves these jobs on a time-shared basis, spending a small increment of time on one, then the next, and so forth. In data networks, particularly high-speed networks, messages are broken into small fixed-length packets, and then the packets from different messages can be transmitted on an alternating basis between messages.

A *round-robin* service system is a system in which, if there are m customers in the system, say $c_1, c_2, \dots, c_m$, then c_1 is served for an incremental interval δ, followed by c_2 being served for an interval δ, and so forth up to c_m. After c_m is served for an interval δ, the server returns and starts serving c_1 for an interval δ again. Thus the customers are served in a cyclic, or "round-robin" order, each getting a small increment of service on each visit from the server. When a customer's service is completed, the customer leaves the system, m is reduced, and the server continues rotating through the now reduced cycle as before. When a new customer arrives, m is increased and the new customer must be inserted into the cycle of existing customers in a way to be discussed later.

Processor sharing is the limit of round-robin service as the increment δ goes to zero. Thus, with processor sharing, if m customers are in the system, all are being served simultaneously, but each is being served at $1/m$ times the basic server rate. For the example of two customers with service requirement 1 and 10, each customer is initially served at rate 1/2, so one customer departs at time 2. At that time, the remaining customer is served at rate 1 and departs at time 11. For round-robin service with an increment of 1, the customer with unit service requirement departs at either time 1 or 2, depending on the initial order of service. With other increments of service, the results are slightly different.

We first analyze round-robin service and then go to the processor-sharing limit as $\delta \to 0$. As the above example suggests, the results are somewhat cleaner in the limiting case, but more realistic in the round-robin case. Round robin provides a good example of the use of backward transition probabilities to find the steady-state distribution of a Markov chain. The techniques used here are quite similar to those used in the next chapter to analyze queueing networks.

Assume a Bernoulli arrival process in which the probability of an arrival in an interval δ is $\lambda\delta$. Assume that the i^{th} arriving customer has a service requirement W_i. The random variables W_i, $i \geq 1$, are IID and independent of the arrival epochs. Thus, in terms of the arrival process and the service requirements, this is the same as an M/G/1 queue (see Section 4.5.5), but with M/G/1 queues, the server serves each customer completely before going on to the next customer. We shall find that the round-robin service here avoids the "slow truck effect" identified with the M/G/1 queue.

For simplicity, assume that W_i is arithmetic with span δ, taking on only values that are positive integer multiples of δ. Let $f(j) = \Pr\{W_i = j\delta\}, j \geq 1$ and let $\overline{F}(j) = \Pr\{W_i > j\delta\}$. Note that if a customer has already received j increments of service, then the probability that that customer will depart after 1 more increment is $f(j+1)/\overline{F}(j)$. This probability of departure on the next service increment after the j^{th} is denoted by

$$g(j) = f(j+1)/\overline{F}(j); j \geq 1. \tag{5.47}$$

The state $\boldsymbol{s}$ of a round-robin system can be expressed as the number, m, of customers in the system, along with an ordered listing of how many service increments each of those m customers have received, *i.e.*,

$$\boldsymbol{s} = (m, z_1, z_2, \dots, z_m), \tag{5.48}$$

where $z_1\delta$ is the amount of service already received by the customer at the front of the queue, $z_2\delta$ is the service already received by the next customer in order, etc. In the special case of an idle queue, $\boldsymbol{s} = (0)$, which we denote as ϕ.

Given that the state X_n at time $n\delta$ is $\boldsymbol{s} \neq \phi$, the state X_{n+1} at time $n\delta + \delta$ evolves as follows:

- A new arrival enters with probability $\lambda\delta$ and is placed at the front of the queue;
- The customer at the front of the queue receives an increment δ of service;
- The customer departs if service is complete.
- Otherwise, the customer goes to the back of the queue

It can be seen that the state transition depends, first, on whether a new arrival occurs (an event of probability $\lambda\delta$), and, second, on whether a departure occurs. If no arrival and no departure occurs, then the queue simply rotates. The new state is $\boldsymbol{s}' = r(\boldsymbol{s})$, where the rotation operator $r(\boldsymbol{s})$ is defined by $r(\boldsymbol{s}) = (m, z_2, \dots, z_m, z_1 + 1)$. If a departure but no arrival occurs, then the customer at the front of the queue receives its last unit of service and departs. The new state is $\boldsymbol{s}' = \delta(\boldsymbol{s})$, where the departure operator $\delta(\boldsymbol{s})$ is defined by $\delta(\boldsymbol{s}) = (m - 1, z_2, \dots, z_m)$.

If an arrival occurs, the new customer receives one unit of service and goes to the back of the queue if more than one unit of service is required. In this case, the new state is $\boldsymbol{s}' = a(\boldsymbol{s})$ where the arrival operator $a(\boldsymbol{s})$ is defined by $a(\boldsymbol{s}) = (m+1, z_1, z_2, \dots, z_m, 1)$. If only one unit of service is required by a new arrival, the arrival departs and $\boldsymbol{s}' = \boldsymbol{s}$. In the special case of an empty queue, $\boldsymbol{s} = \phi$, the state is unchanged if either no arrival occurs or an arrival requiring one increment of service arrives. Otherwise, the new state is $\boldsymbol{s} = (1, 1)$, *i.e.*, the one customer in the system has received one increment of service.

We next find the probability of each transition for $\boldsymbol{s} \neq \phi$. The probability of no arrival is $1 - \lambda\delta$. Given no arrival, and given a non-empty system, $\boldsymbol{s} \neq \phi$, the probability of a departure is $g(z_1) = f(z_1 + 1)/\overline{F}(z_1)$, i e., the probability that one more increment of service allows the customer at the front of the queue to depart. Thus the probability of a departure is $(1 - \lambda\delta)g(z_1)$ and the probability of a rotation is $(1 - \lambda\delta)[1 - g(z_1)]$. Finally, the probability of an arrival is $\lambda\delta$, and given an arrival, the new arrival will leave the system after one unit of service with probability $g(0) = f(1)$. Thus the probability of an arrival and no departure is $\lambda\delta[1 - f(1)]$ and the probability of an unchanged system is $\lambda\delta f(1)$. To summarize, for $\boldsymbol{s} \neq \phi$,

$$
\begin{aligned}
P_{\boldsymbol{s},r(\boldsymbol{s})} &= (1 - \lambda\delta)[1 - g(z_1)]; && r(\boldsymbol{s}) = (m, z_2, \dots, z_m, z_1 + 1) \\
P_{\boldsymbol{s},d(\boldsymbol{s})} &= (1 - \lambda\delta)g(z_1); && d(\boldsymbol{s}) = (m - 1, z_2, \dots, z_m) \\
P_{\boldsymbol{s},a(\boldsymbol{s})} &= \lambda\delta[1 - f(1)]; && a(\boldsymbol{s}) = (m + 1, z_1, z_2, \dots, z_m, 1) \\
P_{\boldsymbol{s},s} &= \lambda\delta f(1). && \qquad (5.49)
\end{aligned}
$$

For the special case of the idle state, $P_{\phi,\phi} = (1 - \lambda\delta) + \lambda\delta f(1)$ and $P_{\phi,(1,1)} = \lambda\delta(1 - f(1))$.

We now find the steady-state distribution for this Markov chain by looking at the backward Markov chain. We will hypothesize backward transition probabilities, and then use Theorem 5.3.3 to verify that the hypothesis is correct. Consider the backward transitions corresponding to each of the forward transitions in (5.49). A rotation in forward time causes the elements $z_1, \dots, z_m$ in the state $\boldsymbol{s} = (m, z_1, \dots, z_m)$ to rotate left, and the left most element (corresponding to the front of the queue) is incremented while rotating to the right end. The backward transition from $r(\boldsymbol{s})$ to $\boldsymbol{s}$ corresponds to the elements $z_2, \dots, z_m, z_1 + 1$

rotating to the right, with the right most element being decremented while rotating to the left end. If we view the transitions in backward time as a kind of round-robin system, we see that the rotation is in the opposite direction from the forward time system.

In the backward time system, we view the numbers $z_1, \dots, z_m$ in the state as the remaining service required before the corresponding customers can depart. Thus, these numbers decrease in the backward moving system. Also, since the customer rotation in the backward moving system is opposite to that in the forward moving system, z_m is the remaining service of the customer at the front of the queue, and z_1 is the remaining service of the customer at the back of the queue. We also view departures in forward time as arrivals in backward time. Thus the backward transition from $d(\boldsymbol{s}) = (m-1, z_2, \dots, z_m)$ to $\boldsymbol{s} = (m, z_1, \dots, z_m)$ corresponds to an arrival requiring $z_1 + 1$ units of service; the arrival goes to the front of the queue, receives one increment of service, and then goes to the back of the queue with z_1 increments of remaining service.

The nicest thing we could now hope for is that the arrivals in backward time are Bernoulli. This is a reasonable hypothesis to make, partly because it is plausible, and partly because it is easy to check via Theorem 5.3.3. Fortunately, we shall find that it is valid. According to this hypothesis, the backward transition probability $P^*_{r(\boldsymbol{s}),\boldsymbol{s}}$ is given by $1-\lambda\delta$; that is, given that X_{n+1} is $r(\boldsymbol{s}) = (m, z_2, \dots, z_m, z_1{+}1)$, and given that there is no arrival in the backward system at time $(n+1)\delta$, then the only possible state at time n is $\boldsymbol{s} = (m, z_1, \dots, z_n)$. Next consider a backward transition from $d(\boldsymbol{s}) = (m-1, z_2, \dots, z_n)$ to $\boldsymbol{s} = (m, z_1, z_2, \dots, z_m)$. This corresponds to an arrival in the backward moving system; the arrival requires $z_1 + 1$ increments of service, one of which is provided immediately, leaving the arrival at the back of the queue with z_1 required increments of service remaining. The probability of this transition is $P^*_{d(\boldsymbol{s}),\boldsymbol{s}} = \lambda\delta f(z_1+1)$. Calculating the other backward transitions in the same way, the hypothesized backward transition probabilities are given by

$$\begin{aligned} P^*_{r(\boldsymbol{s}),\boldsymbol{s}} &= 1-\lambda\delta \qquad P^*_{d(\boldsymbol{s}),\boldsymbol{s}} = \lambda\delta f(z_1+1) \\ P^*_{a(\boldsymbol{s}),\boldsymbol{s}} &= 1-\lambda\delta \qquad P^*_{\boldsymbol{s},\boldsymbol{s}} = \lambda\delta f(1). \end{aligned} \tag{5.50}$$

One should view (5.50) as an hypothesis for the backward transition probabilities. The arguments leading up to (5.50) are simply motivation for this hypothesis. If the hypothesis is correct, we can combine (5.49) and (5.50) to express the steady-state equations of Theorem 5.3.3 (for $\boldsymbol{s} \neq f$) as

$$\pi_{\boldsymbol{s}} P_{\boldsymbol{s},r(\boldsymbol{s})} = \pi_{r(\boldsymbol{s})} P^*_{r(\boldsymbol{s}),\boldsymbol{s}}; \qquad (1-\lambda\delta)[1-g(z_1)]\pi_{\boldsymbol{s}} = (1-\lambda\delta)\pi_{r(\boldsymbol{s})} \tag{5.51}$$

$$\pi_{\boldsymbol{s}} P_{\boldsymbol{s},d(\boldsymbol{s})} = \pi_{d(\boldsymbol{s})} P^*_{d(\boldsymbol{s}),\boldsymbol{s}}; \qquad (1-\lambda\delta)g(z_1)\pi_{\boldsymbol{s}} = \lambda\delta f(z_1+1)\pi_{d(\boldsymbol{s})} \tag{5.52}$$

$$\pi_{\boldsymbol{s}} P_{\boldsymbol{s},a(\boldsymbol{s})} = \pi_{a(\boldsymbol{s})} P^*_{a(\boldsymbol{s}),\boldsymbol{s}}; \qquad \lambda\delta[1-f(1)]\pi_{\boldsymbol{s}} = (1-\lambda\delta)\pi_{a(\boldsymbol{s})} \tag{5.53}$$

$$\pi_{\boldsymbol{s}} P_{\boldsymbol{s},\boldsymbol{s}} = \pi_{\boldsymbol{s}} P^*_{\boldsymbol{s},\boldsymbol{s}}; \qquad \lambda\delta f(1)\pi_{\boldsymbol{s}} = \lambda\delta f(1)\pi_{\boldsymbol{s}}. \tag{5.54}$$

We next show that (5.52), applied repeatedly, will allow us to solve for $\pi_{\boldsymbol{s}}$ (if λ is small enough for the states to be positive recurrent). Verifying that the solution also satisfies (5.51) and (5.53), will then verify the hypothesis. Since $f(z_1+1)/g(z_1)$ is $\overline{F}(z_1)$ from (5.47), we have

$$\pi_{\boldsymbol{s}} = \frac{\lambda\delta}{1-\lambda\delta}\,\overline{F}(z_1)\pi_{d(\boldsymbol{s})}. \tag{5.55}$$

For $m > 1$, $d(\boldsymbol{s}) = (m-1, z_2, \dots, z_m)$, so we can apply (5.55) to $\pi_{d(\boldsymbol{s})}$, and substitute the result back into (5.55), yielding

$$\pi_{\boldsymbol{s}} = \left(\frac{\lambda\delta}{1-\lambda\delta}\right)^2 \overline{F}(z_1)\overline{F}(z_2)\pi_{d(d(\boldsymbol{s}))}, \tag{5.56}$$

where $d(d(\boldsymbol{s})) = (m-2, z_3, \dots, z_m)$. Applying (5.55) repeatedly to $\pi_{d(d(\boldsymbol{s}))}, \pi_{d(d(d(\boldsymbol{s})))}$, and so forth, we eventually get

$$\pi_{\boldsymbol{s}} = \left(\frac{\lambda\delta}{1-\lambda\delta}\right)^m \left(\prod_{j=1}^{m} \overline{F}(z_j)\right)\pi_{\phi}. \tag{5.57}$$

Before this can be accepted as a steady-state distribution, we must verify that it satisfies (5.51) and (5.53). The left hand side of (5.51) is $(1-\lambda\delta)[1-g(z_1)]\pi_{\boldsymbol{s}}$, and, from (5.47), $1-g(z_1) = [\overline{F}(z_1) - f(z_1+1)]/\overline{F}(z_1) = \overline{F}(z_1+1)/(z_1)$. Thus using (5.57), the left side of (5.51) is

$$(1-\lambda\delta)\,\frac{\overline{F}(z_1{+}1)}{\overline{F}(z_1)}\left(\frac{\lambda\delta}{1{-}\lambda\delta}\right)^m \left(\prod_{j=1}^{m}\overline{F}(z_j)\right)\pi_{\phi} = (1{-}\lambda\delta)\left(\frac{\lambda\delta}{1{-}\lambda\delta}\right)^m \left(\prod_{j=2}^{m}\}\overline{F}(z_j)\right)\overline{F}(z_1{+}1)\pi_{\phi}.$$

This is equal to $(1-\lambda\delta)\pi_{r(\boldsymbol{s})}$, verifying (5.51). Equation (5.53) is verified in the same way. We now have to find whether there is a solution for pf such that these probabilities sum to 1. First define $P_m = \sum z_1, \dots, z_m \pi(m, z_1, \dots, z_m)$. This is the probability of m customers in the system. Whenever a new customer enters the system, it receives one increment of service immediately, so each $z_i \geq 1$. Using the hypothesized solution in (5.57),

$$P_m = \left(\frac{\lambda\delta}{1-\lambda\delta}\right)^m \left(\prod_{j=1}^{m}\sum_{i=1}^{\infty}\overline{F}(i)\right)\pi_{\phi}. \tag{5.58}$$

Since $\overline{F}(i) = \Pr\{W > i\delta\}$, since W is arithmetic with span δ, and since the mean of a nonnegative random variable is the integral of its complementary distribution function, we have

$$\begin{aligned} \delta\sum_{i=1}^{\infty}\}\overline{F}(i) &= \mathsf{E}[W] - \delta & (5.59)\\ P_m &= \left(\frac{\lambda}{1-\lambda\delta}\right)^m \left(\mathsf{E}[W]-\delta\right)^m \pi_{\phi}. & (5.60) \end{aligned}$$

Defining $\rho = [\lambda/(1-\lambda\delta)]\{\mathsf{E}[W]-\delta\}$, we see $P_m = \rho^m\pi_{\phi}$. If $\rho < 1$, then $\pi_{\phi} = 1-\rho$, and

$$P_m = (1-\rho)\rho^m; \quad m \geq 0. \tag{5.61}$$

The condition $r < 1$ is required for the states to be positive-recurrent. The expected number of customers in the system for a round-robin queue is $\sum_m mP_m = \rho/(1-\rho)$, and using Little's theorem, Theorem 4.5.3, the expected delay is $\rho/[\lambda(1-\rho)]$. In using Little's theorem here, however, we are viewing the time a customer spends in the system as starting when the number m in the state increases; that is, if a customer arrives at time $n\delta$, it goes to the front of the queue and receives one increment of service, and then, assuming it needs more than one increment, the number m in the state increases at time $(n+1)\delta$. Thus the actual expected delay, including the original d when the customer is being served but not counted in the state, is $\delta + \rho/[\lambda(1-\rho)]$.

The relation between ρ and $\lambda \mathsf{E}[W]$ is shown in Figure 5.8, and it is seen that $\rho < 1$ for $\lambda \mathsf{E}[W] < 1$. The extreme case where $\lambda\delta = \lambda \mathsf{E}[W]$ is the case for which each customer requires exactly one unit of service. Since at most one customer can arrive per time increment, the state always remains at $\boldsymbol{s} = \phi$, and the delay is δ, *i.e.*, the original increment of service received when a customer arrives.

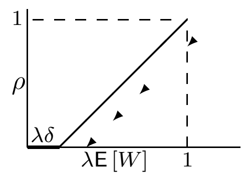

Figure 5.8: ρ as a function of $\lambda \mathsf{E}[W]$ for given $\lambda\delta$.

Note that (5.61) is the same as the distribution of customers in the system for the M/M/1 Markov chain in (5.40), except for the anomaly in the definition of ρ here. We then have the surprising result that if round-robin queueing is used rather than FCFS, then the distribution of the number of customers in the system is approximately the same as that for an M/M/1 queue. In other words, the slow truck effect associated with the M/G/1 queue has been eliminated.

Another remarkable feature of round-robin systems is that one can also calculate the expected delay for a customer conditional on the required service of that customer. This is done in Exercise 5.16, and it is found that the expected delay is linear in the required service.

Next we look at processor sharing by going to the limit as $\delta \to 0$. We first eliminate the assumption that the service requirement distribution is arithmetic with span δ. Assume that the server always spends an increment of time δ on the customer at the front of the queue, and if service is finished before the interval of length δ ends, the server is idle until the next sample time. The analysis of the steady-state distribution above is still valid if we define $\overline{F}(j) = \Pr\{W > j\delta\}$, and $f(j) = \overline{F}(j) - \overline{F}(j+1)$. In this case $\delta \sum_{i=1}^{\infty} \overline{F}(i)$ lies between $\mathsf{E}[W] - \delta$ and $\mathsf{E}[W]$. As $\delta \to 0$, $\rho = \lambda \mathsf{E}[W]$, and distribution of time in the system becomes identical to that of the M/M/1 system.

5.7 Summary

This chapter extended the finite-state Markov chain results of Chapter 3 to the case of countably-infinite state spaces. It also provided an excellent example of how renewal processes can be used for understanding other kinds of processes. In Section 5.1, the first-passage-time random variables were used to construct renewal processes with renewals on successive transitions to a given state. These renewal processes were used to rederive the basic properties of Markov chains using renewal theory as opposed to the algebraic Perron-Frobenius approach of Chapter 3. The central result of this was Theorem 5.1.4, which showed that, for an irreducible chain, the states are positive-recurrent if and only if the steady-state equations, (5.18), have a solution. Also if (5.18) has a solution, it is positive and unique. We also showed that these steady-state probabilities are, with probability 1, time-averages for sample paths, and that, for an ergodic chain, they are limiting probabilities independent of the starting state.

We found that the major complications that result from countable state spaces are, first, different kinds of transient behavior, and second, the possibility of null-recurrent states. For finite-state Markov chains, a state is transient only if it can reach some other state from which it can't return. For countably infinite chains, there is also the case, as in Figure 5.2 for $p > 1/2$, where the state just wanders away, never to return. Null recurrence is a limiting situation where the state wanders away and returns with probability 1, but with an infinite expected time. There is not much engineering significance to null recurrence; it is highly sensitive to modeling details over the entire infinite set of states. One usually uses countably infinite chains to simplify models; for example, if a buffer is very large and we don't expect it to overflow, we assume it is infinite. Finding out, then, that the chain is transient or null-recurrent simply means that the modeling assumption is not very good.

We next studied birth-death Markov chains and reversibility. Birth-death chains are widely used in queueing theory as sample time approximations for systems with Poisson arrivals and various generalizations of exponentially distributed service times. Equation (5.28) gives their steady-state probabilities if positive-recurrent, and shows the conditions under which they are positive-recurrent. We showed that these chains are reversible if they are positive-recurrent.

Theorems 5.3.2 and 5.3.3 provides a simple way to find the steady-state distribution of reversible chains and also of chains where the backward chain behavior could be hypothesized or deduced. We used reversibility to show that M/M/1 and M/M/m Markov chains satisfy Burke's theorem for sampled-time — namely that the departure process is Bernoulli, and that the state at any time is independent of departures before that time.

Branching processes were introduced in Section 5.5 as a model to study the growth of various kinds of elements that reproduce. In general, for these models (assuming $p_0 > 0$), there is one trapping state and all other states are transient. Figure 5.7 showed how to find the probability that the trapping state is entered by the nth generation, and also the probability that it is entered eventually. If the expected number of offspring of an element is at most 1, then the population dies out with probability 1, and otherwise, the population dies out with some given probability q, and grows without bound with probability $1 - q$.

Round-robin queueing was then used as a more complex example of how to use the backward process to deduce the steady-state distribution of a rather complicated Markov chain; this also gave us added insight into the behavior of queueing systems and allowed us to show that, in the processor-sharing limit, the distribution of number of customers is the same as that in an M/M/1 queue.

For further reading on Markov chains with countably-infinite state spaces, see [8], [16], or [22]. Feller [8] is particularly complete, but Ross [16] and Wolff [22] are somewhat more accessible. Harris, [12] is the standard reference on branching processes and Kelly, [13] is the standard reference on reversibility. The material on round-robin systems is from [24] and is generalized there.

5.8 Exercises

Exercise 5.1. Let $\{P_{ij}; i, j \geq 0\}$ be the set of transition probabilities for an countable-state Markov chain. For each i, j, let $\mathsf{F}_{ij}(n)$ be the probability that state j occurs sometime between time 1 and n inclusive, given $X_0 = i$. For some given j, assume that $\{x_k; k \geq 0\}$ is a set of nonnegative numbers satisfying $x_i = P_{ij} + \sum_{k \neq j} P_{ik} x_k$. Show that $x_i \geq \mathsf{F}_{ij}(n)$ for all n and i, and hence that $x_i \geq \mathsf{F}_{ij}(\infty)$ for all i. Hint: use induction.

Exercise 5.2. a) For the Markov chain in Figure 5.2, show that, for $p \geq 1/2$, $\mathsf{F}_{00}(\infty) = 2(1-p)$ and show that $\mathsf{F}_{i0}(\infty) = [(1-p)/p]^i$ for $i \geq 1$. Hint: first show that this solution satisfies (5.9) and then show that (5.9) has no smaller solution (see Exercise 5.1). Note that you have shown that the chain is transient for $p > 1/2$ and that it is recurrent for $p = 1/2$.

b) Under the same conditions as part a), show that $\mathsf{F}_{ij}(\infty)$ equals $2(1-p)$ for $j = i$, equals $[(1-p)/p]^{i-j}$ for $i > j$, and equals 1 for $i < j$.

Exercise 5.3. a): Show that the nth order transition probabilities, starting in state 0, for the Markov chain in Figure 5.2 satisfy

$$P_{0j}^n = pP_{0,i-1}^{n-1} + qP_{0,i+1}^{n-1} \quad j \neq 0; \qquad P_{00}^n = qP_{00}^{n-1} + qP_{01}^{n-1}.$$

Hint: Use the Chapman-Kolmogorov equality, (3.8).

b) For $p = 1/2$, use this equation to calculate P_{0j}^n iteratively for $n = 1, 2, 3, 4$. Verify (5.3) for these values and then use induction to verify (5.3) in general. Note: this becomes an absolute mess for $p \neq 1/2$, so don't attempt this in general.

c) As a more interesting approach, which brings out the relationship of Figures 5.2 and 5.1, note that (5.3), with $j + n$ even, is the probability that $S_n = j$ for the chain in 5.1 and (5.3) with $j + n$ odd is the probability that $S_n = -j - 1$ for the chain in 5.1. By viewing each transition over the self loop at state 0 as a sign reversal for the chain in 5.1, explain why this surprising result is true. (Again, this doesn't work for $p \neq 1/2$, since the sign reversals also reverse the +1, -1 transitions.)

Exercise 5.4. Let j be a transient state in a Markov chain and let j be accessible from i. Show that i is transient also. Interpret this as a form of Murphy's law (if something bad can happen, it will, where the bad thing is the lack of an eventual return). Note: give a direct demonstration rather than using Lemma 5.1.3.

Exercise 5.5. Consider an irreducible positive-recurrent Markov chain. Consider the renewal process $\{N_{jj}(t); t \geq 0\}$ where, given $X_0 = j$, $N_{jj}(t)$ is the number of times that state j is visited from time 1 to t. For each $i \geq 0$, consider a renewal-reward function $R_i(t)$ equal to 1 whenever the chain is in state i and equal to 0 otherwise. Let π_i be the time-average reward.

a) Show that $\pi_i = 1/\overline{T}_{ii}$ for each i with probability 1.

b) Show that $\sum_i \pi_i = 1$. Hint: consider $\sum_{i \leq M} \pi_i$ for any integer M.

c) Consider a renewal-reward function $R_{ij}(t)$ that is 1 whenever the chain is in state i and the next state is state j. $R_{ij}(t) = 0$ otherwise. Show that the time-average reward is equal to $\pi_i P_{ij}$ with probability 1. Show that $p_k = \sum_i \pi_i P_{ik}$ for all k.

Exercise 5.6. Let $\{X_n; n \geq 0\}$ be a branching process with $X_0 = 1$. Let $\overline{Y}$, σ^2 be the mean and variance of the number of offspring of an individual.

a) Argue that $\lim_{n\to\infty} X_n$ exists with probability 1 and either has the value 0 (with probability $\mathsf{F}_{10}(\infty)$) or the value ∞ (with probability $1 - \mathsf{F}_{10}(\infty)$).

b) Show that $\mathsf{VAR}[X_n] = \sigma^2 \overline{Y}^{n-1}(\overline{Y}^n - 1)/(\overline{Y} - 1)$ for $\overline{Y} \neq 1$ and $\mathsf{VAR}[X_n] = n\sigma^2$ for $\overline{Y} = 1$.

Exercise 5.7. There are n states and for each pair of states i and j, a positive number $d_{ij} = d_{ji}$ is given. A particle moves from state to state in the following manner: Given that the particle is in any state i, it will next move to any $j \neq i$ with probability P_{ij} given by

$$P_{ij} = \frac{d_{ij}}{\sum_{j \neq i} d_{ij}}.$$

Assume that $P_{ii} = 0$ for all i. Show that the sequence of positions is a reversible Markov chain and find the limiting probabilities.

Exercise 5.8. Consider a reversible Markov chain with transition probabilities P_{ij} and limiting probabilities π_i. Also consider the same chain truncated to the states $0, 1, \ldots, M$. That is, the transition probabilities $\{P'_{ij}\}$ of the truncated chain are

$$P'_{ij} = \begin{cases} \frac{P_{ij}}{\sum_{k=0}^{m} P_{ik}} & ; \quad 0 \leq i, j \leq M \\ \} \qquad 0 & ; \quad \text{elsewhere.} \end{cases}$$

Show that the truncated chain is also reversible and has limiting probabilities given by

$$\overline{\pi}_i = \frac{\pi_i \sum_{j=0}^{M} P_{ij}}{\sum_{k=0}^{M} \pi_i \sum_{m=0}^{M} P_{km}}.$$

Exercise 5.9. A Markov chain (with states $\{0, 1, 2, \ldots, J-1\}$ where J is either finite or infinite) has transition probabilities $\{P_{ij}; i, j \geq 0\}$. Assume that $P_{0j} > 0$ for all $j > 0$ and $P_{j0} > 0$ for all $j > 0$. Also assume that for all i, j, k, we have $P_{ij}P_{jk}P_{ki} = P_{ik}P_{kj}P_{ji}$.

a) Assuming also that all states are positive recurrent, show that the chain is reversible and find the steady-state probabilities $\{\pi_i\}$ in simplest form.

b) Find a condition on $\{P_{0j}; j \geq 0\}$ and $\{P_{j0}; j \geq 0\}$ that is sufficient to ensure that all states are positive recurrent.

Exercise 5.10. **a)** Use the birth and death model described in figure 5.4 to find the steady-state probability mass function for the number of customers in the system (queue plus service facility) for the following queues:

i) M/M/1 with arrival probability $\lambda\delta$, service completion probability $\mu\delta$.

ii) M/M/m with arrival probability $\lambda\delta$, service completion probability $i\mu\delta$ for i servers busy, $1 \leq i \leq m$.

iii) M/M/∞ with arrival probability $\lambda\delta$, service probability $i\mu\delta$ for i servers. Assume δ so small that $i\mu\delta < 1$ for all i of interest.

Assume the system is positive recurrent.

b) For each of the queues above give necessary conditions (if any) for the states in the chain to be i) transient, ii) null recurrent, iii) positive recurrent.

c) For each of the queues find:

L = (steady-state) mean number of customers in the system.

L^q = (steady-state) mean number of customers in the queue.

W = (steady-state) mean waiting time in the system.

W^q = (steady-state) mean waiting time in the queue.

Exercise 5.11. **a)** Given that an arrival occurs in the interval $(n\delta, (n+1)\delta)$ for the sampled-time M/M/1 model in figure 5, find the conditional PMF of the state of the system at time $n\delta$ (assume n arbitrarily large and assume positive recurrence).

b) For the same model, again in steady state but not conditioned on an arrival in $(n\delta, (n+1)\delta)$, find the probability $Q(i, j)(i \geq j > 0)$ that the system is in state i at $n\delta$ and that $i-j$ departures occur before the next arrival.

c) Find the expected number of customers seen in the system by the first arrival after time $n\delta$. (Note: the purpose of this exercise is to make you cautious about the meaning of "the state seen by a random arrival").

Exercise 5.12. Find the backward transition probabilities for the Markov chain model of age in figure 2. Draw the graph for the backward Markov chain, and interpret it as a model for residual life.

Exercise 5.13. Consider the sample time approximation to the M/M/1 queue in figure 5.5.

a) Give the steady-state probabilities for this chain (no explanations or calculations required–just the answer).

In parts b) to g) do not use reversibility and do not use Burke's theorem. Let X_n be the state of the system at time $n\delta$ and let D_n be a random variable taking on the value 1 if a departure occurs between $n\delta$ and $(n+1)\delta$, and the value 0 if no departure occurs. Assume that the system is in steady state at time $n\delta$.

b) Find $\Pr\{X_n = i, D_n = j\}$ for $i \geq 0$, $j = 0, 1$

c) Find $\Pr\{D_n = 1\}$

d) Find $\Pr\{X_n = i \mid D_n = 1\}$ for $i \geq 0$

e) Find $\Pr\{X_{n+1} = i \mid D_n = 1\}$ and show that X_{n+1} is statistically independent of D_n. Hint: Use part d); also show that $\Pr\{X_{n+1} = i\} = \Pr\{X_{n+1} = i \mid D_n = 1\}$ for all $i \geq 0$ is sufficient to show independence.

f) Find $\Pr\{X_{n+1} = i, D_{n+1} = j \mid D_n\}$ and show that the pair of variables (X_{n+1}, D_{n+1}) is statistically independent of D_n.

g) For each $k > 1$, find $\Pr\{X_{n+k} = i, D_{n+k} = j \mid D_{n+k-1}, D_{n+k-2}, \dots, D_n\}$ and show that the pair (X_{n+k}, D_{n+k}) is statistically independent of $(D_{n+k-1}, D_{n+k-2}, \dots, D_n)$. Hint: use induction on k; as a substep, find $\Pr\{X_{n+k} = i \mid D_{n+k-1} = 1, D_{n+k-2}, \dots, D_n\}$ and show that X_{n+k} is independent of $D_{n+k-1}, D_{n+k-2}, \dots, D_n$.

h) What do your results mean relative to Burke's theorem.

Exercise 5.14. Let $\{X_n, n \geq 1\}$ denote a irreducible recurrent Markov chain having a countable state state space. Now consider a new stochastic process $\{Y_n, n \geq 0\}$ that only accepts values of the Markov chain that are between 0 and some integer m. For instance, if $m = 3$ and $X_1 = 1$, $X_2 = 3$, $X_3 = 5$, $X_4 = 6$, $X_5 = 2$, then $Y_1 = 1$, $Y_2 = 3$, $Y_3 = 2$.

a) Is $\{Y_n, n \geq 0\}$ a Markov chain? Explain briefly.

b) Let p_j denote the proportion of time that $\{X_n, n \geq 1\}$ is in state j. If $p_j > 0$ for all j, what proportion of time is $\{Y_n, n \geq 0\}$ in each of the states $0, 1, \dots, m$?

c) Suppose $\{X_n\}$ is null-recurrent and let $p_i(m), i = 0, 1, \dots, m$ denote the long-run proportions for $\{Y_n, n \geq 0\}$. Show that $p_j(m) = p_i(m)E[\text{time the } X \text{ process spends in } j \text{ between returns to } i]$, $j \neq i$.}

Exercise 5.15. Verify that (5.53) is satisfied by the hypothesized solution to p in (5.57). Also show that the equations involving the idle state f are satisfied.

Exercise 5.16. Replace the state $\mathbf{m} = (m, z_1, \dots, z_m)$ in Section 5.6 with an expanded state $\mathbf{m} = (m, z_1, w_1, z_2, w_2, \dots, z_m, w_m)$ where m and $\{z_i; 1 \leq i \leq m\}$ are as before and $w_1, w_2, \dots, w_m$ are the original service requirements of the m customers.

a) Hypothesizing the same backward round-robin system as hypothesized in Section 5.6, find the backward transition probabilities and give the corresponding equations to (5.51-5.54) for the expanded state description.

b) Solve the resulting equations to show that

$$\pi_{\mathbf{m}} = \pi + \phi\Big(\frac{\lambda\delta}{1-\lambda\delta}\Big)^m \prod_{j=1}^{m} f(w_j).$$

c) Show that the probability that there are m customers in the system, and that those customers have original service requirements given by $w_1, \dots, w_m$, is

$$\Pr\{m, w_1, \dots, w_m\} = \pi_\phi \quad \frac{\lambda\delta}{1-\lambda\delta}\Bigg)^m \prod_{j=1}^{m} (w_j - 1) f(w_j).$$

d) Given that a customer has original service requirement w, find the expected time that customer spends in the system.