Chapter 7

Markov Chains

7.1 INTRODUCTION

In the previous chapter, we presented the Generalized Semi-Markov Process (GSMP) framework as a means of modeling stochastic DES. By allowing event clocks to tick at varying speeds, we also provided an extension to the basic GSMP. In addition, we introduced the Poisson process as a basic building block for a class of stochastic DES which possess the Markov (memoryless) property. Thus, we obtained the class of stochastic processes known as *Markov chains*, which we will study in some detail in this chapter. It should be pointed out that the analysis of Markov chains provides a rich framework for studying many DES of practical interest, ranging from gambling and the stock market to the design of "high-tech" computer systems and communication networks.

The main characteristic of Markov chains is that their stochastic behavior is described by transition probabilities of the form $P[X(t_{k+1}) = x' \mid X(t_k) = x]$ for all state values x, x' and $t_k \leq t_{k+1}$. Given these transition probabilities and a distribution for the initial state, it is possible to determine the probability of being at any state at any time instant. Describing precisely how to accomplish this and appreciating the difficulties involved in the process are the main objectives of this chapter.

We will first study *discrete-time* Markov chains. This will allow us to present the main ideas, techniques, and analytical results before proceeding to *continuous-time* Markov chains. In principle, we can always "solve" Markov chains, that is, determine the probability of being at any state at any time instant. However, in practice this task can be a formidable one, as it often involves the solution of complicated differential equations. Even though we would like to obtain general *transient* solutions, in most cases we have to settle for *steady-state* or *stationary* solutions, which describe the probability of being at any state in the long run only (after the system has been in operation for a "sufficiently long" period of time). Even then we often need to resort to numerical techniques. Fortunately, however, explicit closed-form expressions are available for several cases of practical interest. In particular, we will focus attention on the class of *birth–death Markov chains*, of which, as we shall see, the Poisson process turns out to be a special case. Finally, we will present a methodology for converting continuous-time Markov chains into equivalent discrete-time Markov chains. This is known as *uniformization*, and it relies on the fundamental memoryless property of these models.

7.2 DISCRETE-TIME MARKOV CHAINS

Recall that in a discrete-time Markov chain events (and hence state transitions) are constrained to occur at time instants $0, 1, 2, \ldots, k, \ldots$ Thus, we form a stochastic sequence $\{X_1, X_2, \ldots\}$ which is characterized by the Markov (memoryless) property:

$$\begin{aligned} &P[X_{k+1} = x_{k+1} \mid X_k = x_k, X_{k-1} = x_{k-1}, \ldots, X_0 = x_0] \\ &\quad = P[X_{k+1} = x_{k+1} \mid X_k = x_k] \end{aligned} \tag{7.1}$$

Given the current state x_k, the value of the next state depends only on x_k and not on any past state history (no state memory). Moreover, the amount of time spent in the current state is irrelevant in determining the next state (no age memory).

In the rest of this section, we first discuss what it means to "obtain a model" of a stochastic DES that qualifies as a discrete-time Markov chain. We then proceed with analyzing such models. One of our main objectives is the determination of probabilities for the chain being at various states at different times.

7.2.1 Model Specification

Thus far, we have been modeling stochastic DES by means of the stochastic timed automaton formalism based on the six-tuple

$$(\mathcal{E}, \mathcal{X}, \Gamma, p, p_0, G)$$

In the framework of the previous chapter, state transitions are driven by events belonging to the set $\mathcal{E}$. Thus, the transition probabilities are expressed as $p(x'; x, e')$ where $e' \in \Gamma(x)$ is the triggering event, and $\Gamma(x)$ is the feasible event set at state x. In Markov chains, however, we will only be concerned with the total probability $p(x', x)$ of making a transition from x to x', regardless of which event actually causes the transition. Thus, similar to (6.68), we apply the rule of total probability to get

$$p(x', x) = \sum_{i \in \Gamma(x)} p(x'; x, i) \cdot p(i, x)$$

where $p(i, x)$ is the probability that event i occurs at state x. This transition probability is an aggregate over all events $i \in \Gamma(x)$ which may cause the transition from x to x'. In general, we will allow transition probabilities to depend on the time instant at which the transition occurs, so we set

$$p_k(x', x) = P[X_{k+1} = x' \mid X_k = x]$$

With these observations in mind, we will henceforth identify events with state transitions in our specification of Markov chain models. Although it is always useful to keep in mind what the underlying event set is, we will omit the specification of the sets $\mathcal{E}$ and $\Gamma(x)$. As we shall see, the clock structure G is implicitly defined by the Markov property in (7.1). Therefore, to specify a Markov chain model we only need to identify:

1. A state space $\mathcal{X}$.
2. An initial state probability $p_0(x) = P[X_0 = x]$, for all $x \in \mathcal{X}$.
3. Transition probabilities $p(x', x)$ where x is the current state and x' is the next state.

Since the state space $\mathcal{X}$ is a countable set, we will subsequently map it onto the set of non-negative integers (or any subset thereof):

$$\mathcal{X} = \{0, 1, 2, \ldots\}$$

which will allow us to keep notation simple.

In the next few sections, we provide the basic definitions and notation associated with discrete-time Markov chains.

7.2.2 Transition Probabilities and the Chapman-Kolmogorov Equations

Since we have decided to use the nonnegative integers as our state space, we will use the symbols i, j to denote a typical current and next state respectively. We will also modify our notation and define *transition probabilities* as follows:

$$p_{ij}(k) \equiv P[X_{k+1} = j \mid X_k = i] \tag{7.2}$$

where $i, j \in \mathcal{X}$ and $k = 0, 1, \ldots$ As already pointed out, transition probabilities are allowed to be time-dependent, hence the need to express $p_{ij}(k)$ as a function of the time instant k.

Clearly, $0 \leq p_{ij}(k) \leq 1$. In addition, observe that for any state i and time instant k:

$$\sum_{\text{all } j} p_{ij}(k) = 1 \tag{7.3}$$

since we are summing over all possible mutually exclusive events causing a transition from i to some new state.

The transition probabilities $p_{ij}(k)$ defined in (7.2) refer to state transitions that occur in one step. A natural extension is to consider state transitions that occur over n steps, $n = 1, 2, \ldots$ Thus, we define the *n-step transition probabilities*:

$$p_{ij}(k, k+n) \equiv P[X_{k+n} = j \mid X_k = i] \tag{7.4}$$

Let us now condition the event $[X_{k+n} = j \mid X_k = i]$ above on $[X_u = r]$ for some u such that $k < u \leq k+n$. Using the rule of total probability, (7.4) becomes

$$p_{ij}(k, k+n) = \sum_{\text{all } r} P[X_{k+n} = j \mid X_u = r, X_k = i] \cdot P[X_u = r \mid X_k = i] \tag{7.5}$$

By the memoryless property (7.1),

$$P[X_{k+n} = j \mid X_u = r, X_k = i] = P[X_{k+n} = j \mid X_u = r] = p_{rj}(u, k+n)$$

Moreover, the second term in the sum in (7.5) is simply $p_{ir}(k,u)$. Therefore,

$$p_{ij}(k,k+n) = \sum_{\text{all } r} p_{ir}(k,u)p_{rj}(u,k+n), \qquad k < u \leq k+n \tag{7.6}$$

This is known as the *Chapman-Kolmogorov equation.* It is one of the most general relationships we can derive about discrete-time Markov chains.

It is often convenient to express the Chapman-Kolmogorov equation (7.6) in matrix form. For this purpose, we define the matrix

$$\mathbf{H}(k,k+n) \equiv [p_{ij}(k,k+n)], \qquad i,j = 0,1,2,\ldots \tag{7.7}$$

Then, (7.6) can be rewritten as a matrix product:

$$\mathbf{H}(k,k+n) = \mathbf{H}(k,u)\mathbf{H}(u,k+n) \tag{7.8}$$

Choosing $u = k+n-1$ we get

$$\mathbf{H}(k,k+n) = \mathbf{H}(k,k+n-1)\mathbf{H}(k+n-1,k+n) \tag{7.9}$$

This relationship is known as the *forward Chapman-Kolmogorov equation.* The *backward Chapman-Kolmogorov equation* is obtained by choosing $u = k+1$, which gives

$$\mathbf{H}(k,k+n) = \mathbf{H}(k,k+1)\mathbf{H}(k+1,k+n) \tag{7.10}$$

7.2.3 Homogeneous Markov Chains

Whenever the transition probability $p_{ij}(k)$ is independent of k for all $i,j \in \mathcal{X}$, we obtain a *homogeneous* Markov chain. In this case, (7.2) is written as

$$p_{ij} = P[X_{k+1} = j \mid X_k = i] \tag{7.11}$$

where p_{ij} is independent of k. In simple terms, a state transition from i to j always occurs with the same probability, regardless of the point in time when it is observed. Note that homogeneity is a form of stationarity which applies to transition probabilities only, but not necessarily to the Markov chain itself. In particular, $P[X_{k+1} = j \mid X_k = i]$ may be independent of k, but the joint probability $P[X_{k+1} = j, X_k = i]$ need not be independent of k. To see this more clearly, observe that

$$\begin{aligned} P[X_{k+1} = j, X_k = i] &= P[X_{k+1} = j \mid X_k = i] \cdot P[X_k = i] \\ &= p_{ij} \cdot P[X_k = i] \end{aligned}$$

where $P[X_k = i]$ need not be independent of k.

Example 7.1

Consider a machine which can be in one of two states, UP or DOWN. We choose a state space $\mathcal{X} = \{0,1\}$ where 1 denotes UP and 0 denotes DOWN. The state of the machine is checked every hour, and we index hours by $k = 0,1,\ldots$ Thus, we form the stochastic sequence $\{X_k\}$, where X_k is the state of the machine at the kth hour. Let us further assume that if the machine is UP, it has a probability α of failing during the next hour. If the machine is in the DOWN state, it has a probability β of being repaired during the next hour.

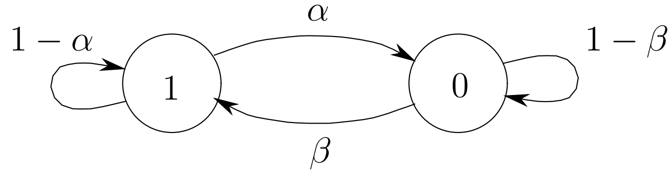


Figure 7.1: State transition diagram for Example 7.1.

We therefore obtain a homogeneous discrete-time Markov chain with transition probabilities:

$$p_{10} = \alpha, \qquad p_{11} = 1-\alpha, \qquad p_{01} = \beta, \qquad p_{00} = 1-\beta$$

where $0 \le \alpha \le 1$ and $0 \le \beta \le 1$. We can also draw a state transition diagram (as we did for automata), but the arcs connecting states now refer to *probabilities* rather than events (see Fig. 7.1).

Alternatively, it is reasonable to expect that as the machine wears out over time it becomes more failure-prone. Suppose we modify the transition probabilities above so that

$$p_{10}(k) = 1-\gamma^k, \qquad p_{11}(k) = \gamma^k$$

for some $0 < \gamma < 1$ and $k = 0, 1, \ldots$ In this case, the probability of the machine failing increases with k and approaches 1 as $k \to \infty$. The new chain is no longer homogeneous, since $p_{10}(k)$ is time-dependent.

In a homogeneous Markov chain, the n-step transition probability $p_{ij}(k, k+n)$ is also independent of k. In this case, we denote it by p_{ij}^n, and (7.4) is written as

$$p_{ij}^n = P[X_{k+n} = j \mid X_k = i], \qquad n = 1, 2, \ldots \tag{7.12}$$

Then, by setting $u = k+m$ in the Chapman-Kolmogorov equation (7.6), we get

$$p_{ij}^n = \sum_{\text{all } r} p_{ir}^m p_{rj}^{n-m}$$

and by choosing $m = n-1$:

$$p_{ij}^n = \sum_{\text{all } r} p_{ir}^{n-1} p_{rj} \tag{7.13}$$

or, in the matrix notation of (7.7), where $\mathbf{H}(n) \equiv [p_{ij}^n]$,

$$\mathbf{H}(n) = \mathbf{H}(n-1)\mathbf{H}(1) \tag{7.14}$$

In simple terms, (7.13) breaks up the process of moving from state i to state j in n steps into two parts. First, we move from i to some intermediate state r in $(n-1)$ steps, and then we take one more step from r to j. By aggregating over all possible intermediate states r, we obtain all possible paths from i to j in n steps.

From this point on, we will limit ourselves to *homogeneous* Markov chains (unless explicitly stated otherwise).

7.2.4 The Transition Probability Matrix

The transition probability information for a discrete-time Markov chain is conveniently summarized in matrix form. We define the *transition probability matrix* $\mathbf{P}$ to consist of all p_{ij}:

$$\mathbf{P} \equiv [p_{ij}], \qquad i, j = 0, 1, 2, \ldots \tag{7.15}$$

By this definition, in conjunction with (7.3), all elements of the ith row in this matrix, $i = 1, 2, \ldots$, must always sum up to 1. Note that $\mathbf{P} = \mathbf{H}(1)$ in the matrix equation (7.14).

Example 7.2 (A simplified telephone call process)
We describe a simple telephone call process in discrete time. Let the time line consist of small intervals indexed by $k = 0, 1, 2, \ldots$, which are sometimes called *time slots*. The process operates as follows:

- At most one telephone call can occur in a single time slot, and there is a probability α that a telephone call occurs in any one slot.
- If the phone is busy, the call is lost (no state transition); otherwise, the call is processed.
- There is a probability β that a call in process completes in any one time slot.
- If both a call arrival and a call completion occur in the same time slot, the new call will be processed.

We also assume that call arrivals and call completions in any time slot occur independently of each other and of the state of the process.

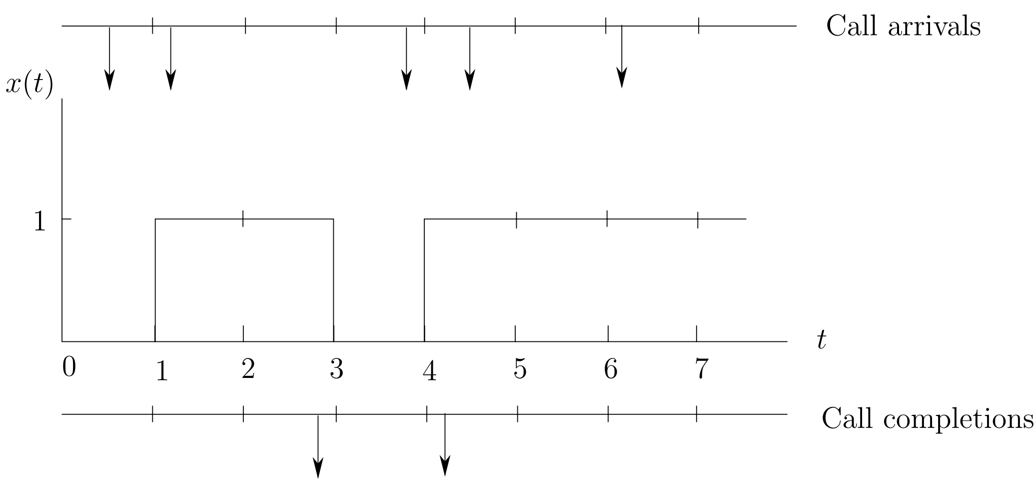


Figure 7.2: Typical sample path for Example 7.2.

Let X_k denote the state of the system at the kth slot, which is either 0 (phone is idle) or 1 (phone is busy). A typical sample path is shown in Fig. 7.2. In this example, $x_0 = 0$ and a call arrives in slot 0, hence $x_1 = 1$. Note that in the slot that starts at $t = 4$ both a new call arrival and a call completion occur; the new call is not lost in this case, and we have $x_5 = 1$.

The transition probabilities are as follows:

$p_{00} = 1 - \alpha$	(phone remains idle if no new call occurs in current slot)
$p_{01} = \alpha$	(phone becomes busy if new call occurs in current slot)
$p_{10} = \beta \cdot (1 - \alpha)$	(phone becomes idle if call completes in current slot *and* no new call occurs)
$p_{11} = (1 - \beta) + \alpha\beta$	(phone remains busy if call does not complete in current slot *or* call does complete but a new call also occurs)

A state transition diagram is shown in Fig. 7.3. The probability transition matrix for this Markov chain is

$$\mathbf{P} = \begin{bmatrix} 1-\alpha & \alpha \\ \beta(1-\alpha) & (1-\beta)+\alpha\beta) \end{bmatrix}$$

where we can easily check that (7.3) is satisfied for both rows. Clearly, once p_{10} is determined, we can simply write $p_{11} = 1 - p_{10}$. Independently evaluating p_{11} serves as a means of double-checking that we have correctly evaluated transition probabilities.

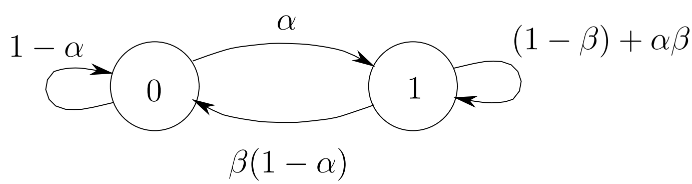


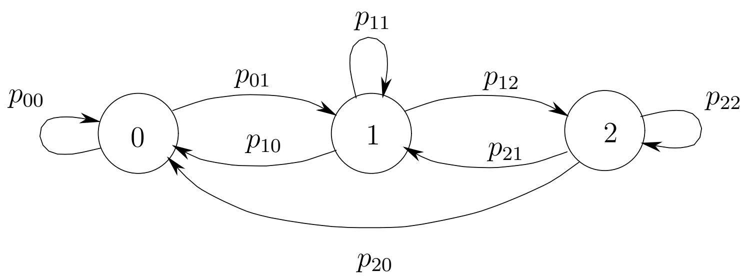
Figure 7.3: State transition diagram for Example 7.2.

Remark. As mentioned earlier, in Markov chain modeling we concentrate on state transition probabilities regardless of which event causes the transition. As a result, the stochastic behavior of a DES is compactly captured in the transition probability matrix $\mathbf{P}$. But this comes at some cost: The *structural information* contained in the event-driven description is lost. This can be seen in Example 7.2 when one considers $p_{11} = (1 - \alpha) + \alpha\beta$. This is just a number telling us the probability of a self-loop transition at state 1, but it does not reveal how this transition might come about (because of no call completion or because of both a call completion and call arrival). Nor does it provide any information on possible call arrivals that were lost during such self-loop transitions. One should be aware of this basic tradeoff: compact representation versus structural information.

Example 7.3 (A two-processor computer system)
Let us consider a computer system consisting of two identical processors working in parallel. Here, we have two "servers" (the processors), as opposed to only one (the phone) in Example 7.2. Once again, let the time line consist of slots indexed by $k = 1, 2, \ldots$ The operation of this system is only slightly more complicated than that of the telephone call process:

- At most one job can be submitted to the system in a single time slot, and such an event occurs with probability α.
- When a job is submitted to the system, it is served by whichever processor is available.
- If both processors are available, the job is given to processor 1.
- If both processors are busy, the job is lost.
- When a processor is busy, there is a probability β that it completes processing in any one time slot.
- If a job is submitted in a slot where both processors are busy and either one of the processors completes in that slot, then the job will be processed.

We also assume that job submissions and completions at either processor in any time slot all occur independently of each other and of the state of the system.



$$
\begin{array}{lll}
p_{00} = (1-\alpha) & p_{01} = \alpha & p_{02} = 0 \\
p_{10} = \beta \cdot (1-\alpha) & p_{11} = (1-\beta)\cdot(1-\alpha) + \beta\cdot\alpha & p_{12} = (1-\beta)\cdot\alpha \\
p_{20} = \beta^2 \cdot (1-\alpha) & \multicolumn{2}{l}{p_{21} = \beta\cdot(1-\beta)\cdot(1-\alpha) + \beta\cdot(1-\beta)\cdot(1-\alpha) + \beta^2\cdot\alpha}
\end{array}
$$

$$p_{22} = (1-\beta)^2 + \beta\cdot(1-\beta)\cdot\alpha + \beta\cdot(1-\beta)\cdot\alpha$$

Figure 7.4: State transition diagram for Example 7.3.

Let X_k denote the number of jobs being processed by the system at slot k. Thus, the state space is the set $\{0, 1, 2\}$. The transition probability matrix $\mathbf{P}$ is obtained from the state transition diagram shown in Fig. 7.4, where we summarize all $p_{ij}, i, j = 0, 1, 2$. The derivation of each p_{ij} is straightforward. As an example, we explain the derivation of p_{21}:

$$
\begin{aligned}
p_{21} &= \beta\cdot(1-\beta)\cdot(1-\alpha) && \text{(proc. 1 completes, proc. 2 remains busy, no new job submitted)} \\
&\quad +\beta\cdot(1-\beta)\cdot(1-\alpha) && \text{(proc. 2 completes, proc. 1 remains busy, no new job submitted)} \\
&\quad +\beta^2\cdot\alpha && \text{(both processors complete, new job submitted)}
\end{aligned}
$$

Let us write $\mathbf{P}$ for a specific numerical example when $\alpha = 0.5$ and $\beta = 0.7$. Evaluating all p_{ij} shown in Fig. 7.4 for these values, we obtain

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.35 & 0.5 & 0.15 \\ 0.245 & 0.455 & 0.3 \end{bmatrix}$$

7.2.5 State Holding Times

It might be argued that the specification of a discrete-time Markov chain model should require the specification of probability distributions characterizing the amount of time spent at some state i whenever this state is visited. However, due to the memoryless property, these distributions are automatically determined. We will use $V(i)$ to denote the number of time steps spent at state i when it is visited. This is a discrete random variable referred to as the *state holding time*.

It is a fundamental property of discrete-time Markov chains that *the distribution of the state holding time* $V(i), P[V(i) = n], n = 1, 2, \ldots,$ *is geometric with parameter* p_{ii}. This is a direct consequence of the memoryless property, as shown below.

Suppose state i is entered at the kth time step. Note that the event $[V(i) = n]$ is identical to the event $[X_{k+1} = i, X_{k+2} = i, \ldots, X_{k+n-1} = i, X_{k+n} \neq i \mid X_k = i]$, since the chain remains at state i for precisely n time steps. Therefore,

$$\begin{aligned} P[V(i) = n] &= P[X_{k+1} = i, \ldots, X_{k+n-1} = i, X_{k+n} \neq i \mid X_k = i] \\ &= P[X_{k+n} \neq i \mid X_{k+n-1} = i, \ldots, X_k = i] \\ &\quad \cdot P[X_{k+n-1} = i, \ldots, X_{k+1} = i \mid X_k = i] \\ &= [1 - P[X_{k+n} = i \mid X_{k+n-1} = i, \ldots, X_k = i]] \\ &\cdot P[X_{k+n-1} = i, \ldots, X_{k+1} = i \mid X_k = i] \end{aligned}$$

By the memoryless property (7.1), we have

$$\begin{aligned} &P[X_{k+n} = i \mid X_{k+n-1} = i, \ldots, X_k = i] \\ &\quad = P[X_{k+n} = i \mid X_{k+n-1} = i] = p_{ii} \end{aligned}$$

and, therefore,

$$P[V(i) = n] = (1 - p_{ii}) \cdot P[X_{k+n-1} = i, \ldots, X_{k+1} = i \mid X_k = i] \tag{7.16}$$

Similarly,

$$\begin{aligned} &P[X_{k+n-1} = i, \ldots, X_{k+1} = i \mid X_k = i] \\ &\quad = P[X_{k+n-1} = i \mid X_{k+n-2} = i, \ldots, X_k = i] \\ &\qquad \cdot P[X_{k+n-2} = i, \ldots, X_{k+1} = i \mid X_k = i] \\ &\quad = P[X_{k+n-1} = i \mid X_{k+n-2} = i] \cdot P[X_{k+n-2} = i, \ldots, X_{k+1} = i \mid X_k = i] \\ &\quad = p_{ii} \cdot P[X_{k+n-2} = i, \ldots, X_{k+1} = i \mid X_k = i] \end{aligned}$$

and (7.16) becomes

$$P[V(i) = n] = (1 - p_{ii}) \cdot p_{ii} \cdot P[X_{k+n-2} = i, \ldots, X_{k+1} = i \mid X_k = i]$$

Repeating this process finally yields

$$P[V(i) = n] = (1 - p_{ii}) \cdot (p_{ii})^{n-1} \tag{7.17}$$

This is a geometric distribution with parameter p_{ii}. It is independent of the amount of time already spent in state i, which is a consequence of the memoryless property: At each time step the chain determines its future based only on the fact that the current state is i. Hence, $P[V(i) = n]$ depends only on p_{ii}. The geometric distribution is the discrete-time counterpart of the exponential distribution which we saw in the previous chapter, and which we shall have the opportunity to see once again later in this chapter.

7.2.6 State Probabilities

One of the main objectives of Markov chain analysis is the determination of probabilities of finding the chain at various states at specific time instants. We define *state probabilities* as follows:

$$\pi_j(k) \equiv P[X_k = j] \tag{7.18}$$

Accordingly, we define the *state probability vector*

$$\boldsymbol{\pi}(k) = [\pi_0(k), \pi_1(k), \ldots] \tag{7.19}$$

This is a row vector whose dimension is specified by the dimension of the state space of the chain. Clearly, it is possible for $\boldsymbol{\pi}(k)$ to be infinite-dimensional.

A discrete-time Markov chain model is completely specified if, in addition to the state space $\mathcal{X}$ and the transition probability matrix $\mathbf{P}$, we also specify an initial state probability vector

$$\boldsymbol{\pi}(0) = [\pi_0(0), \pi_1(0), \ldots]$$

which provides the probability distribution of the initial state, X_0, of the chain.

7.2.7 Transient Analysis

Once a model is specified through $\mathcal{X}$, $\mathbf{P}$, and $\boldsymbol{\pi}(0)$, we can start addressing questions such as: What is the probability of moving from state i to state j in n steps? Or: What is the probability of finding the chain at state i at time k? In answering such questions, we limit ourselves to given finite numbers of steps over which the chain is analyzed. This is what we refer to as *transient* analysis, in contrast to the *steady-state* analysis we will discuss in later sections.

Our main tool for transient analysis is provided by the recursive equation (7.13). A similar recursive relationship can also be derived, making explicit use of the transition probability matrix $\mathbf{P}$. Using the definitions of $\boldsymbol{\pi}(k)$ and $\mathbf{P}$ we get

$$\boldsymbol{\pi}(k+1) = \boldsymbol{\pi}(k)\mathbf{P}, \qquad k = 0, 1, \ldots \tag{7.20}$$

This relationship is easily established as follows. Let $\pi_j(k+1)$ be a typical element of $\boldsymbol{\pi}(k+1)$. Then, by conditioning the event $[X_{k+1} = j]$ on $[X_k = i]$ for all possible i, we get

$$\begin{aligned}\pi_j(k+1) = P[X_{k+1} = j] &= \sum_{\text{all } i} P[X_{k+1} = j \mid X_k = i] \cdot P[X_k = i] \\ &= \sum_{\text{all } i} p_{ij} \cdot \pi_i(k)\end{aligned}$$

which, written in matrix form, is precisely (7.20).

Moreover, by using (7.20) we can obtain an expression for $\boldsymbol{\pi}(k)$ in terms of a given initial state probability vector $\boldsymbol{\pi}(0)$ and the transition probability matrix $\mathbf{P}$. Specifically, for $k = 0$ (7.20) becomes

$$\boldsymbol{\pi}(1) = \boldsymbol{\pi}(0)\mathbf{P}$$

Then, for $k = 1$, we get

$$\boldsymbol{\pi}(2) = \boldsymbol{\pi}(1)\mathbf{P} = \boldsymbol{\pi}(0)\mathbf{P}^2$$

and continuing in the same fashion (or using a formal induction argument) we obtain

$$\boldsymbol{\pi}(k) = \boldsymbol{\pi}(0)\mathbf{P}^k, k = 1, 2, \ldots \tag{7.21}$$

Using (7.13), (7.20), and (7.21) we can completely study the transient behavior of homogeneous discrete-time Markov chains.

Example 7.4 (A simple gambling model)

Many games of chance are stochastic DES and can be modeled through discrete-time Markov chains. Suppose a gambler starts out with a capital of 3 dollars and plays a game of "spin-the-wheel" where he bets 1 dollar at a time. The wheel has 12 numbers, and the gambler bets on only one of them. Thus, his chance of winning in a spin is 1/12. If his number comes up, he receives 3 dollars, hence he has a net gain of 2 dollars. If his number does not come up, his bet is lost.

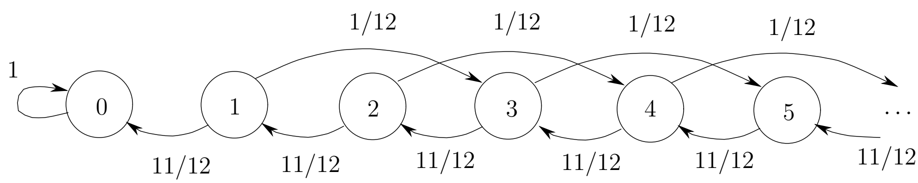


Figure 7.5: State transition diagram for Example 7.4.
In the "spin-the-wheel" game a gambler either wins two dollars per bet with probability 1/12*, or loses one dollar with probability* 11/12*. The game ends if the gambler's capital becomes* 0*.*

This is a Markov chain with state space $\{0, 1, \ldots\}$, where the state X_k represents the gambler's capital after the kth bet, and the initial state is $X_0 = 3$. The state transition diagram is shown in Fig. 7.5. From any state $i > 0$, there are two possible transitions: to state $(i-1)$ when the gambler loses, and to state $(i+2)$ when the gambler wins. At state 0 the gambler has no more capital to bet with; the game is effectively over, which we model by a self-loop transition with probability 1. The corresponding transition probability matrix is

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & \cdots \\ 11/12 & 0 & 0 & 1/12 & 0 & \cdots \\ 0 & 11/12 & 0 & 0 & 1/12 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

We now use (7.13) to compute the probability that the gambler doubles his capital after three bets. Note that since $X_0 = 3$,

$$\begin{aligned} &P[\text{gambler doubles his capital after three bets}] \\ &= P[X_3 = 6 \mid X_0 = 3] = p_{36}^3 \end{aligned}$$

It follows from (7.13) that

$$p_{36}^3 = \sum_{r=0}^{\infty} p_{3r}^2 \cdot p_{r6}$$

Looking at $\mathbf{P}$ (or the state transition diagram), observe that $p_{r6} = 0$ for all $r \neq 4, 7$. Therefore, we can rewrite the equation above as

$$p_{36}^3 = p_{34}^2 \cdot p_{46} + p_{37}^2 \cdot p_{76} = p_{34}^2 \cdot \frac{1}{12} + p_{37}^2 \cdot \frac{11}{12} \tag{7.22}$$

Using (7.13) again, we get

$$p_{34}^2 = \sum_{r=0}^{\infty} p_{3r} \cdot p_{r4}$$

where $p_{3r} = 0$ for all $r \neq 2, 5$, therefore,

$$p_{34}^2 = p_{32} \cdot p_{24} + p_{35} \cdot p_{54} = \frac{11}{12} \cdot \frac{1}{12} + \frac{1}{12} \cdot \frac{11}{12} = \frac{22}{(12)^2} \tag{7.23}$$

Similarly,

$$p_{37}^2 = p_{32} \cdot p_{27} + p_{35} \cdot p_{57} = p_{32} \cdot 0 + \frac{1}{12} \cdot \frac{1}{12} = \frac{1}{(12)^2} \tag{7.24}$$

Combining (7.22) through (7.24) we get

$$p_{36}^3 = \frac{33}{(12)^3} \approx 0.019$$

In this case, the same result can be obtained by inspection of the state transition diagram in Fig. 7.5. One can see that there are only three possible ways to get from state 3 to state 6 in exactly three steps:

$$\begin{aligned} 3 \to 2 \to 4 \to 6 &\text{ with probability } \frac{11}{12} \cdot \frac{1}{12} \cdot \frac{1}{12} \\ 3 \to 5 \to 4 \to 6 &\text{ with probability } \frac{1}{12} \cdot \frac{11}{12} \cdot \frac{1}{12} \\ 3 \to 5 \to 7 \to 6 &\text{ with probability } \frac{1}{12} \cdot \frac{1}{12} \cdot \frac{11}{12} \end{aligned}$$

which gives the same answer as before.

Remark. The type of Markov chain described in Example 7.4 is often referred to as a gambler's ruin chain: If the gambler continues to play, he is condemned to eventually enter state 0 and hence be ruined. The reason is that there is always a positive probability to reach that state, no matter how large the initial capital is. This situation is typical of most games of chance. In Example 7.4, note that the expected gain for the gambler after every bet is $2(1/12) + (-1)(11/12) = -9/12 < 0$, so the game hardly seems "fair" to play. In a "fair" game, on the other hand, the expected gain after every bet is 0. The chain describing such a game is called a *martingale*. In particular, a martingale is a sequence of random variables $\{X_k\}$ with the property:

$$E[X_{k+1} \mid X_k = i, X_{k-1} = x_{k-1}, \dots, X_0 = x_0] = i$$

Thus, if in a game of chance X_k represents the gambler's capital at time k, this property requires that the expected capital after the kth bet, given all past capital values, is the same as the present value $X_k = i$. A martingale is not necessarily a Markov chain. If, however, $\{X_k\}$ is a Markov chain, then the property above is equivalent to the requirement that the transition probability p_{ij} satisfy

$$\sum_{\text{all } j} j \cdot p_{ij} = i$$

for all states i.

Example 7.5

Let us return to the two-processor system in Example 7.3. With $\alpha = 0.5$ and $\beta = 0.7$, we obtained the transition probability matrix

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.35 & 0.5 & 0.15 \\ 0.245 & 0.455 & 0.3 \end{bmatrix}$$

We will now pose three questions regarding this system, and answer them using (7.21). Suppose the system starts out empty, that is,

$$\boldsymbol{\pi}(0) = [1, 0, 0]$$

Then:

1. What is the probability that the system is empty at $k = 3$?
2. What is the probability that no job completion occurs in the third slot?
3. What is the probability that the system remains empty through slots 1 and 2?

To answer the first question, we need to calculate $\pi_0(3)$. Let us determine the entire vector $\boldsymbol{\pi}(3)$ using (7.21). Thus, we first calculate $\mathbf{P}^3$:

$$\begin{aligned} \mathbf{P}^3 &= \begin{bmatrix} 0.425 & 0.5 & 0.075 \\ 0.38675 & 0.49325 & 0.12 \\ 0.35525 & 0.4865 & 0.15825 \end{bmatrix} \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.35 & 0.5 & 0.15 \\ 0.245 & 0.455 & 0.3 \end{bmatrix} \\ &= \begin{bmatrix} 0.405875 & 0.496625 & 0.0975 \\ 0.3954125 & 0.4946 & 0.1099875 \\ 0.3866712 & 0.4928788 & 0.12045 \end{bmatrix} \end{aligned}$$

and since $\boldsymbol{\pi}(0) = [1, 0, 0]$, (7.21) gives

$$\boldsymbol{\pi}(3) = [0.405875, 0.496625, 0.0975] \tag{7.25}$$

Therefore, $\boldsymbol{\pi}_0(3) = 0.405875$. Note that the same result could also be obtained by calculating the three-step transition probability p_{00}^3 from (7.13).

To answer the second question, note that

$$\begin{aligned} &P[\text{no job completes at } k = 3] \\ &\quad = \sum_{j=0}^{2} P[\text{no job completes at } k = 3 \mid X_3 = j] \cdot \pi_j(3) \end{aligned}$$

We then see that $P[\text{no job completes at } k = 3 \mid X_3 = 0] = 1$, $P[\text{no job completes at } k = 3 \mid X_3 = 1] = (1 - \beta)$, and $P[\text{no job completes at } k = 3 \mid X_3 = 2] = (1 - \beta)^2$. Moreover, the values of $\pi_j(3), j = 0, 1, 2$, were determined in (7.25). We then get

$$P[\text{no job completes at } k = 3] = 0.563612$$

To answer the third question, we need to calculate the probability of the event $[X_1 = 0, X_2 = 0]$. We have

$$\begin{aligned} P[X_1 = 0, X_2 = 0] &= P[X_2 = 0 \mid X_1 = 0] \cdot P[X_1 = 0] \\ &= p_{00} \cdot \pi_0(1) \end{aligned} \tag{7.26}$$

From the transition probability matrix we have $p_{00} = 0.5$, and from (7.21):

$$\boldsymbol{\pi}(1) = [1, 0, 0]\mathbf{P} = [0.5, 0.5, 0]$$

Therefore,

$$P[X_1 = 0, X_2 = 0] = 0.25$$

As (7.26) clearly indicates (and one should expect), $P[X_1 = 0, X_2 = 0] \neq P[X_1 = 0] \cdot P[X_2 = 0] = \pi_0(1) \cdot \pi_0(2)$.

In general, solving (7.20) or (7.21) to obtain $\boldsymbol{\pi}(k)$ for any $k = 1, 2, \dots$ is not a simple task. As in classical system theory, where we resort to z-transform or Laplace transform techniques in order to obtain transient responses, a similar approach may be used here as well. We will not provide details, but refer the reader to Appendix I. The main result is that $\mathbf{P}^k$, which we need to calculate $\boldsymbol{\pi}(k)$ through (7.21), is the inverse z-transform of the matrix $[\mathbf{I} - z\mathbf{P}]^{-1}$, where $\mathbf{I}$ is the identity matrix.

7.2.8 Classification of States

We will now provide several definitions in order to classify the states of a Markov chain in ways that turn out to be particularly convenient for our purposes, specifically the study of steady-state behavior. We begin with the notion of state reachability, which we have encountered earlier in our study of untimed DES models.

Definition. A state j is said to be *reachable* from a state i if $p_{ij}^n > 0$ for some $n = 1, 2, \dots$ ♦

If we consider a state transition diagram, then finding a path from i to j is tantamount to reachability of j from i.

Next, let S be a subset of the state space $\mathcal{X}$. If there is no feasible transition from any state in S to any state outside S, then S forms a *closed* set:

Definition. A subset S of the state space $\mathcal{X}$ is said to be *closed* if $p_{ij} = 0$ for any $i \in S$, $j \notin S$. ♦

A particularly interesting case of a closed set is one consisting of a single state:

Definition. A state i is said to be *absorbing* if it forms a single-element closed set. ♦

Clearly, by (7.3), if i is an absorbing state we have $p_{ii} = 1$. Another interesting case of a closed set is one consisting of mutually reachable states:

Definition. A closed set of states S is said to be *irreducible* if state j is reachable from state i for any $i, j \in S$. ♦

Definition. A Markov chain is said to be *irreducible* if its state space $\mathcal{X}$ is irreducible. ♦

If a Markov chain is not irreducible, it is called *reducible*. In this case, there must exist at least one closed set of states which prohibits irreducibility.

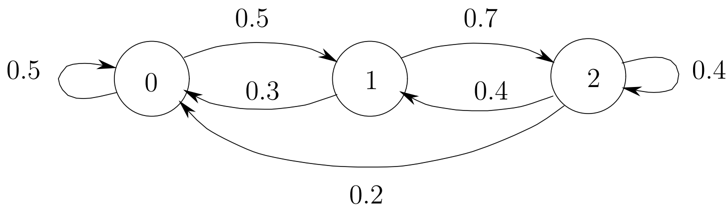


Figure 7.6: State transition diagram for Example 7.6.

Example 7.6
Consider the discrete-time Markov chain with state transition diagram shown in Fig. 7.6. This is a simple case of a three-state irreducible chain, since every state is reachable from any other state. Its transition probability matrix is

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.3 & 0 & 0.7 \\ 0.2 & 0.4 & 0.4 \end{bmatrix}$$

Example 7.7
Consider the five-state Markov chain shown in Fig. 7.7, with transition probability matrix

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 \\ 0.3 & 0 & 0.3 & 0 & 0.4 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

This is a reducible Markov chain, since there are states which are not reachable from other states; for instance, state 0 cannot be reached from state 4 or from state 3. There are two closed sets in this chain (both proper subsets of the state space), which are easy to detect by inspection in Fig. 7.7:

$$S_1 = \{4\} \quad \text{and} \quad S_2 = \{2, 3\}$$

State 4 is absorbing. The closed set S_2 is irreducible, since states 2 and 3 are mutually reachable. These observations can also be made by inspection of $\mathbf{P}$. Note that $p_{ii} = 1$ indicates that state i is absorbing, as in the case $i = 4$. Also, in the case of $i = 2, 3$, we can see that $p_{ij} = 0$ for all $j \neq 2, 3$, indicating that the chain is trapped in these two states forming a closed set.

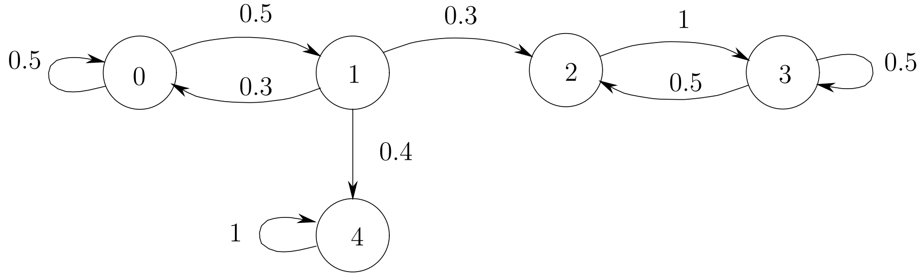


Figure 7.7: State transition diagram for Example 7.7.
This is a reducible chain. State 4 is absorbing. States 2 and 3 form a closed irreducible set.

Transient and Recurrent States

Suppose a chain is in state i. It is reasonable to ask the question: Will the chain ever return to state i? If the answer is "definitely yes," state i is *recurrent*, otherwise it is *transient*. To formalize this distinction, we introduce the notion of the *hitting time*, T_{ij}, defined as follows:

$$T_{ij} \equiv \min\{k > 0 : X_0 = i, X_k = j\} \tag{7.27}$$

The hitting time represents the *first* time the chain enters state j given that it starts out at state i. If we let $j = i$, then T_{ii} is the first time that the chain returns to state i given that it is currently in state i. We refer to the random variable T_{ii} as the *recurrence time* of state i.

In discrete time, T_{ii} can take on values $1, 2, \ldots$ (including ∞). We define ρ_i^k to be the probability that the recurrence time of state i is k:

$$\rho_i^k \equiv P[T_{ii} = k] \tag{7.28}$$

Then, let ρ_i be the probability of the event [ever return to i | current state is i], which is given by

$$\rho_i = \sum_{k=1}^{\infty} \rho_i^k \tag{7.29}$$

Observe that the event [ever return to i | current state is i] is identical to the event $[T_{ii} < \infty]$. Therefore, we also have

$$\rho_i = P[T_{ii} < \infty] \tag{7.30}$$

Definition. A state i is said to be *recurrent* if $\rho_i = 1$. If $\rho_i < 1$, state i is said to be *transient*.

♦

Thus, recurrence implies that a state is definitely visited again. On the other hand, a transient state may be visited again, but, with some positive probability $(1 - \rho_i)$, it will not.

Example 7.8
In the Markov chain shown in Fig. 7.8, state 2 is absorbing. Clearly, this is a recurrent state, since the chain will forever be returning to 2, (i.e., $\rho_2 = 1$). On the other hand, if the current state is 0 or 1, there is certainly some positive probability of returning to this state. However, this probability is less than 1, since the chain will eventually be trapped in state 2. For example, if the current state is 0, then with probability $p_{02} = 0.5$ the chain will never return to 0; and if the current state is 1, then with probability $p_{10}p_{02} = 0.25$ the chain will never return to state 1.

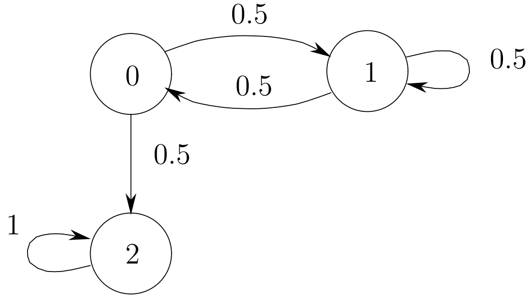


Figure 7.8: State transition diagram for Example 7.8.
In this chain, state 2 is recurrent. States 0 and 1 are transient.

Example 7.9
Returning to the gambler's ruin chain of Example 7.4 (see Fig. 7.5), observe that state 0 is absorbing and hence recurrent (the gambler has lost his capital and can no longer play). All remaining states in the countably infinite set $1, 2, \ldots$ are transient. Thus, the gambler wanders around the chain, possibly for a long period of time, but he eventually gets trapped at state 0.

There are several simple facts regarding transient and recurrent states which can be formally proved. We state some of the most important ones without providing proofs (the proofs are left as exercises; the reader is also referred to Chap. 1 of Hoel et al., 1972).

Theorem 7.1 If a Markov chain has a finite state space, then at least one state is recurrent. ♦

Theorem 7.2 If i is a recurrent state and j is reachable from i, then state j is recurrent. ♦

Theorem 7.3 If S is a finite closed irreducible set of states, then every state in S is recurrent. ♦

We can use Fig. 7.7 to illustrate these results. The finite Markov chain in Fig. 7.7 has five states of which three (states 2, 3, 4) are recurrent (Theorem 7.1). Also, state 2 is recurrent and 3 is reachable from 2; as expected, state 3 is also recurrent (Theorem 7.2). Finally, states 2 and 3 form a finite closed irreducible set which consists of recurrent states only (Theorem 7.3).

Null and Positive Recurrent States

Let i be a recurrent state. We denote by M_i the *mean recurrence time* of state i, given by

$$M_i \equiv E[T_{ii}] = \sum_{k=1}^{\infty} k \rho_i^k \tag{7.31}$$

Depending on whether M_i is finite or not we classify state i as *positive recurrent* or *null recurrent*.

Definition. A recurrent state i is said to be *positive* (or *non-null*) *recurrent* if $M_i < \infty$. If $M_i = \infty$, state i is said to be *null recurrent.* ♦

A null recurrent state is not a transient state, because the probability of recurrence is 1; however, the expected recurrence time is infinite. We can view transient states and positive recurrent states as two extremes: Transient states may never be revisited, whereas positive recurrent states are definitely revisited with finite expected recurrence time. Null recurrent states may be viewed as "weakly recurrent" states: They are definitely revisited, but the expected recurrence time is infinite.

A result similar to Theorem 7.2 is the following (see also Chap. 2 of Hoel et al., 1972):

Theorem 7.4 If i is a positive recurrent state and j is reachable from i, then state j is positive recurrent. ♦

By combining Theorems 7.2 and 7.4, we obtain a very useful fact pertaining to irreducible closed sets, and hence also irreducible Markov chains:

Theorem 7.5 If S is a closed irreducible set of states, then every state in S is positive recurrent or every state in S is null recurrent or every state in S is transient. ♦

We can also obtain a stronger version of Theorem 7.3:

Theorem 7.6 If S is a finite closed irreducible set of states, then every state in S is positive recurrent. ♦

Example 7.10 (Discrete-time birth–death chain)
To illustrate the distinctions between transient, positive recurrent and null recurrent states, let us take a close look at the Markov chain of Fig. 7.9. In this model, the state increases by 1 with probability $(1-p)$ or decreases by 1 with probability p from every state $i > 0$. At $i = 0$, the state remains unchanged with probability p. We often refer to a transition from i to $(i+1)$ as a "birth," and from i to $(i-1)$ as a "death." This is a simple version of what is known as a discrete-time *birth–death chain*. We will have the opportunity to explore its continuous-time version in some depth later in this chapter.

Before doing any analysis, let us argue intuitively about the effect the value of p should have on the nature of this chain. Suppose we start the chain at state 0. If $p < 1/2$, the chain tends to drift towards larger and larger values of i, so we expect state 0 to be transient. If $p > 1/2$, on the other hand, then the chain always tends to drift back towards 0, so we should expect state 0 to be recurrent. Moreover, the larger the value of p, the faster we expect a return to state 0, on the average; conversely, as p approaches $1/2$, we expect the mean recurrence time for state 0 to increase. An interesting case is that of $p = 1/2$. Here, we expect that a return to state 0 will occur, but it may take a very long time. In fact, it turns out that the case $p = 1/2$ corresponds to state 0 being null recurrent, whereas if $p > 1/2$ it is positive recurrent.

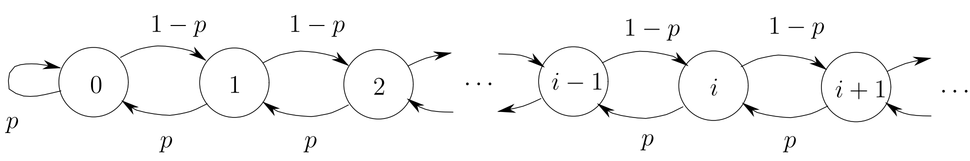


Figure 7.9: State transition diagram for Example 7.10.

Let us now try to verify what intuition suggests. Recalling (7.30), observe that

$$\rho_0 = P[T_{00} < \infty] = p + (1-p) \cdot P[T_{10} < \infty] \tag{7.32}$$

In words, starting at state 0, a return to this state can occur in one of two ways: in a single step with probability p, or, with probability $(1-p)$, in some finite number of steps consisting of a one-step transition to state 1 and then a return to 0 in T_{10} steps. Let us set

$$q_1 = P[T_{10} < \infty] \tag{7.33}$$

In addition, let us fix some state $m > 1$, and define for any state $i = 1, \ldots, m-1$,

$$q_i(m) = P[T_{i0} < T_{im}] \quad \text{for some } m > 1 \tag{7.34}$$

Thus, $q_i(m)$ is the probability that the chain, starting at state i, visits state 0 before it visits state m. We also set $q_m(m) = 0$ and $q_0(m) = 1$. We will now try to evaluate $q_i(m)$ as a function of p, which we will assume to be $0 < p < 1$. This will allow us to obtain $q_1(m)$, from which we will finally obtain q_1, and hence ρ_0.

Taking a good look at the state transition diagram of Fig. 7.9, we observe that

$$q_i(m) = p \cdot q_{i-1}(m) + (1-p) \cdot q_{i+1}(m) \tag{7.35}$$

The way to see this is similar to the argument used in (7.32). Starting at state i, a visit to state 0 before state m can occur in one of two ways: from state $(i-1)$ which is entered next with probability p, or from state $(i+1)$ which is entered next with probability $(1-p)$. Then, adding and subtracting the term $(1-p)q_i(m)$ to the right-hand side of (7.35) above, we get

$$q_{i+1}(m) - q_i(m) = \frac{p}{1-p}[q_i(m) - q_{i-1}(m)]$$

For convenience, set

$$\beta = \frac{p}{1-p} \tag{7.36}$$

We now see that

$$\begin{aligned} q_{i+1}(m) - q_i(m) &= \beta \cdot \beta[q_{i-1}(m) - q_{i-2}(m)] \\ &= \ldots = \beta^i [q_1(m) - q_0(m)] \end{aligned} \tag{7.37}$$

and by summing over $i = 0, \ldots, m-1$, we get

$$\sum_{i=0}^{m-1} q_{i+1}(m) - \sum_{i=0}^{m-1} q_i(m) = [q_1(m) - q_0(m)] \sum_{i=0}^{m-1} \beta^i$$

which reduces to

$$q_m(m) - q_0(m) = [q_1(m) - q_0(m)] \sum_{i=0}^{m-1} \beta^i \tag{7.38}$$

Recalling that $q_m(m) = 0$ and $q_0(m) = 1$, we immediately get

$$q_1(m) = 1 - \frac{1}{\sum_{i=0}^{m-1} \beta^i} \tag{7.39}$$

We would now like to use this result in order to evaluate q_1 in (7.33). The argument we need requires a little thought. Let us compare the number of steps T_{12} in moving from state 1 to state 2 to the number of steps T_{13}. Note that to get from 1 to 3 we must necessarily go through 2. This implies that $T_{13} > T_{12}$. This observation extends to any T_{1i}, T_{1j} with $j > i$. In addition, since to get from state 1 to 2 requires at least one step, we have

$$1 \le T_{12} < T_{13} < \cdots \tag{7.40}$$

and it follows that $T_{1m} \ge m - 1$ for any $m = 2, 3, \ldots$ Therefore, as $m \to \infty$ we have $T_{1m} \to \infty$. Then returning to the definition (7.34) for $i = 1$,

$$\lim_{m\to\infty} q_1(m) = \lim_{m\to\infty} P[T_{10} < T_{1m}] = P[T_{10} < \infty] \tag{7.41}$$

The second equality above is justified by a basic theorem from probability theory (see Appendix I), as long as the events $[T_{10} < T_{1m}]$ form an increasing sequence with $m = 2, 3, \ldots$, which in the limit gives the event $[T_{10} < \infty]$; this is indeed the case by (7.40).

Combining the definition of q_1 in (7.33) with (7.39) and (7.41), we get

$$q_1 = P[T_{10} < \infty] = \lim_{m\to\infty} \left[1 - \frac{1}{\sum_{i=0}^{m-1} \beta^i}\right] = 1 - \frac{1}{\sum_{i=0}^{\infty} \beta^i}$$

Let us now take a closer look at the infinite sum above. If $\beta < 1$, the sum converges and we get

$$\sum_{i=0}^{\infty} \beta^i = \frac{1}{1-\beta}$$

which gives $q_1 = \beta$. Recall from (7.36) that $\beta = p/(1-p)$. Therefore, this case corresponds to the condition $p < 1 - p$ or $p < 1/2$. If, on the other hand, $\beta \ge 1$, that is, $p \ge 1/2$, we have $\sum_{i=0}^{\infty} \beta^i = \infty$, and obtain $q_1 = 1$.

We can now finally put it all together by using these results in (7.32):

1. If $p < 1/2, q_1 = \beta = p/(1-p)$, and (7.32) gives

   $$\rho_0 = 2p < 1$$

   which implies that state 0 is transient as we had originally guessed.
2. If $p \ge 1/2, q_1 = 1$, and (7.32) gives

   $$\rho_0 = 1$$

   and state 0 is recurrent as expected. We will also later show (see Example 7.13) that when $p = 1/2$ (the point at which ρ_0 switches from 1 to a value less than 1) state 0 is in fact null recurrent.

Observing that the chain of Fig. 7.9 is irreducible (as long as $0 < p < 1$), we can also apply Theorem 7.5 to conclude that in case 1 above all states are transient, and hence the chain is said to be transient. Similarly, in case 2 we can conclude that all states are recurrent, and, if state 0 is null recurrent, then all states are null recurrent.

Periodic and Aperiodic States

Sometimes the structure of a Markov chain is such that visits to some state i are constrained to occur only in a number of steps which is a multiple of an integer $d \geq 2$. Such states are called periodic, and d is called the *period*. In order to provide a formal definition, consider the set of integers

$$\{n > 0 : p_{ii}^n > 0\}$$

and let d be the *greatest common divisor* of this set. Note that if 1 is in this set, then $d = 1$. Otherwise, we may have $d = 1$ or $d > 1$. For example, for the set $\{4,\ 8,\ 24\ \}$ we have $d = 4$.

Definition. A state i is said to be periodic if the greatest common divisor d of the set $\{n > 0 : p_{ii}^n > 0\}$ is $d \geq 2$. If $d = 1$, state i is said to be *aperiodic*. ♦

Example 7.11
Consider the Markov chain of Fig. 7.10 with transition probability matrix

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

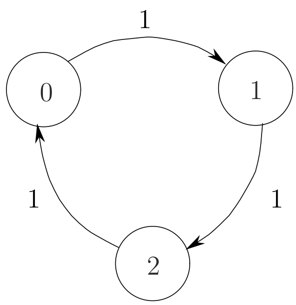


Figure 7.10: Periodic states with $d = 3$.

This is a simple example of a Markov chain whose states are all periodic with period $d = 3$, since the greatest common divisor of the set $\{n > 0 : p_{ii}^n > 0\} = \{3, 6, 9, \ldots\}$ is 3 for all $i = 0, 1, 2$.

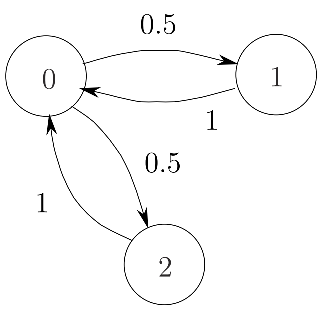


Figure 7.11: Periodic states with $d = 2$.

Another example of a Markov chain whose states are all periodic with $d = 2$ is shown in Fig. 7.11, where

$$\mathbf{P} = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

In this case, $\{n > 0 : p_{ii}^n > 0\} = \{2, 4, 6, \ldots\}$ for all $i = 0, 1, 2$, and the greatest common divisor is 2.

Finally, consider the Markov chain of Fig. 7.12, where

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \end{bmatrix}$$

In this case, $\{n > 0 : p_{ii}^n > 0\} = \{1, 2, 3, 4, \ldots\}$ for all $i = 0, 1, 2$, and $d = 1$. Thus, all states are aperiodic.

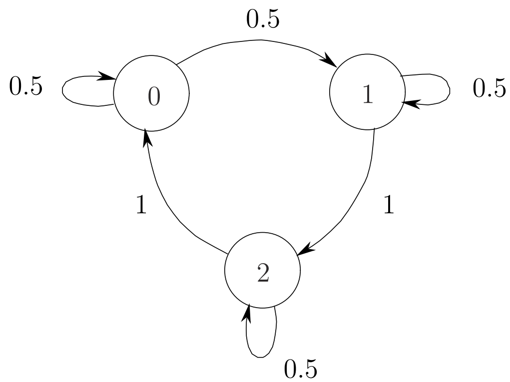


Figure 7.12: Aperiodic states.

It is easy to see that if $p_{ii} > 0$ for some state i of a Markov chain, then the state must be aperiodic, since 1 is an element of the set $\{n > 0 : p_{ii}^n > 0\}$ and hence $d = 1$. This is the case in the chain of Fig. 7.12. It is also possible to show the following (see, for example, Chap. 2 of Hoel et al., 1972):

Theorem 7.7 If a Markov chain is irreducible, then all its states have the same period. ♦

It follows that if $d = 1$ for any state of an irreducible Markov chain, then all states are aperiodic and the chain is said to be aperiodic. On the other hand, if any state has period $d \geq 2$, then all states have the same period and the chain is said to be periodic with period $d \geq 2$.

Summary of State Classifications

There is of course good reason why we have gone through the process of defining new concepts to classify states in a Markov chain. This reason will become clear in the next section, when we derive stationary state probabilities for certain types of chains. We now summarize the state classifications described above as shown in Fig. 7.13. A state i is recurrent if the probability ρ_i of ever returning to i is 1. A recurrent state i is positive recurrent if the expected time required to return to i, M_i, is finite. Finally, state i is aperiodic if the period d_i of the number of steps required to return to i is 1. As we will see, irreducible Markov chains consisting of positive recurrent aperiodic states possess particularly attractive properties when it comes to steady state analysis.

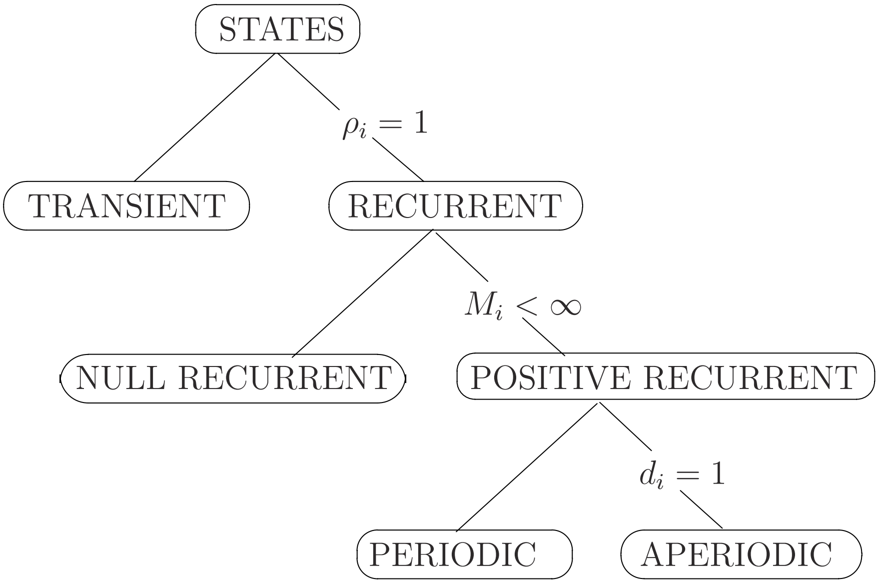


Figure 7.13: Summary of state classifications.

7.2.9 Steady State Analysis

In Sect. 7.2.7, we saw how to address questions of the type: What is the probability of finding a Markov chain at state i at time k? In transient analysis we limited ourselves to given finite numbers of steps over which the chain is observed. In steady state analysis we extend our inquiry to questions such as: What is the probability of finding a Markov chain at state i in the long run? By "long run" we mean that the system we are modeling as a Markov chain is allowed to operate for a sufficiently long period of time so that the state probabilities can reach some fixed values which no longer vary with time. This may or may not be achievable. Our study, therefore, centers around the quantities

$$\pi_j = \lim_{k \to \infty} \pi_j(k) \tag{7.42}$$

where, as defined in (7.18), $\pi_j(k) = P[X_k = j]$, and the existence of these limits is not always guaranteed. Thus, we need to address three basic questions:

1. Under what conditions do the limits in (7.42) exist?
2. If these limits exist, do they form a legitimate probability distribution, that is, $\sum_j \pi_j = 1$?
3. How do we evaluate π_j?

If π_j exists for some state j, it is referred to as a *steady-state, equilibrium,* or *stationary state probability.* Accordingly, if π_j exists for all states j, then we obtain the *stationary state probability vector*

$$\boldsymbol{\pi} = [\pi_0, \pi_1, \ldots]$$

It is essential to keep in mind that the quantity reaching steady state is a state probability – not a state which of course remains a random variable.

Our study of the steady state behavior of discrete-time Markov chains greatly depends on whether or not we are dealing with an irreducible or a reducible chain. Thus, in the next two sections we consider the two cases separately.

7.2.10 Irreducible Markov Chains

The main objective of our study is the determination of the limits in (7.42) if they exist. Recalling the recursive equation (7.20):

$$\boldsymbol{\pi}(k+1) = \boldsymbol{\pi}(k)\mathbf{P}, \qquad k = 0, 1, \ldots$$

it is often the case that after a long time period (i.e., large values of k) we have $\boldsymbol{\pi}(k+1) \approx \boldsymbol{\pi}(k)$. In other words, as $k \to \infty$ we get $\boldsymbol{\pi}(k) \to \boldsymbol{\pi}$, where $\boldsymbol{\pi}$ is the stationary state probability vector. This vector, if it exists, also defines the stationary probability distribution of the chain. Then, if indeed in the limit as $k \to \infty$ we get $\boldsymbol{\pi}(k+1) = \boldsymbol{\pi}(k) = \boldsymbol{\pi}$, we should be able to obtain $\boldsymbol{\pi}$ from (7.20) by solving a system of linear algebraic equations

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

where the elements of $\boldsymbol{\pi}$ satisfy $\pi_j \geq 0$ and $\sum_j \pi_j = 1$.

The first important observation is that the presence of periodic states in an irreducible Markov chain prevents the existence of the limits in (7.42). We will use the example of the periodic chain of Fig. 7.11 to illustrate this point. Recall that the transition probability matrix for this chain is

$$\mathbf{P} = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

and let $\boldsymbol{\pi}(0) = [1, 0, 0]$, that is, it is known that the chain is initially at state 0. We can calculate $\mathbf{P}^2, \mathbf{P}^3, \ldots$ to get

$$\mathbf{P}^2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 0.5 & 0.5 \end{bmatrix} \qquad \mathbf{P}^3 = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$\mathbf{P}^4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 0.5 & 0.5 \end{bmatrix} \cdots$$

and, in general,

$$\mathbf{P}^{2k} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 0.5 & 0.5 \end{bmatrix}, \qquad \mathbf{P}^{2k-1} = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

for all $k = 1, 2, \ldots$ Hence, from (7.21) we have

$$\boldsymbol{\pi}(2k) = [1, 0, 0], \qquad \boldsymbol{\pi}(2k-1) = [0, 0.5, 0.5]$$

Thus, the state probability $\pi_0(k)$ oscillates between 1 and 0 depending on the number of steps k (even or odd). The same is true for the other two state probabilities, which oscillate between 0 and 0.5. It follows that

$$\lim_{k \to \infty} \pi_j(k)$$

cannot exist for $j = 0, 1, 2$.

Fortunately, as long as an irreducible Markov chain is not periodic the limit of $\pi_j(k)$ as $k \to \infty$ always exists. We state without proof this important result (see, for example, Asmussen, 1987).

Theorem 7.8 In an irreducible aperiodic Markov chain the limits

$$\pi_j = \lim_{k\to\infty} \pi_j(k)$$

always exist and they are independent of the initial state probability vector. ◆

Note that Theorem 7.8 guarantees the existence of π_j for all states j, but it does not guarantee that we can form a legitimate stationary state probability distribution satisfying $\sum_j \pi_j = 1$.

Next, let us recall Theorem 7.5, where it was stated that an irreducible Markov chain consists of states which are all positive recurrent or they are all null recurrent or they are all transient. Intuitively, if a state is transient it can only be visited a finite number of times; hence, in the long run, the probability of finding the chain in such a state must be zero. A similar argument holds for null recurrent states. We now present the two fundamental results allowing us to answer the questions we raised earlier: Under what conditions can we guarantee the existence of stationary state probability vectors and how do we determine them? The proofs require some rather elaborate convergence arguments and are omitted (the reader is again referred to Asmussen, 1987).

Theorem 7.9 In an irreducible aperiodic Markov chain consisting of transient states or of null recurrent states

$$\pi_j = \lim_{k\to\infty} \pi_j(k) = 0$$

for all states j, and no stationary probability distribution exists. ◆

Theorem 7.10 In an irreducible aperiodic Markov chain consisting of positive recurrent states a unique stationary state probability vector $\boldsymbol{\pi}$ exists such that $\pi_j > 0$ and

$$\pi_j = \lim_{k\to\infty} \pi_j(k) = \frac{1}{M_j} \tag{7.43}$$

where M_j is the mean recurrence time of state j defined in (7.31). The vector $\boldsymbol{\pi}$ is determined by solving

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P} \tag{7.44}$$

$$\sum_{\text{all } j} \pi_j = 1 \tag{7.45}$$

◆

It is clear that aperiodic positive recurrent states are highly desirable in terms of reaching steady state. We term such states *ergodic*. A Markov chain is said to be ergodic if all its states are ergodic.

Note that a combination of Theorems 7.6 and 7.10 leads to the observation that every *finite* irreducible aperiodic Markov chain has a unique stationary state probability vector $\boldsymbol{\pi}$ determined through (7.44) and (7.45). In this case, obtaining $\boldsymbol{\pi}$ is simply a matter of solving a set of linear equations. However, solving (7.44) and (7.45) in the case of an infinite state space is certainly not an easy task.

Remark. The fact that $\pi_j = 1/M_j$ in (7.43) has an appealing physical interpretation. The probability π_j represents the fraction of time spent by the chain at state j at steady state. Thus, a short recurrence time for j ought to imply a high probability of finding the chain at j. Conversely, a long recurrence time implies a small state probability. In fact, as M_j increases one can see that π_j approaches 0; in the limit, as $M_j \to \infty$, we see that $\pi_j \to 0$, that is, j behaves like a null recurrent state under Theorem 7.9.

Example 7.12

Let us consider the Markov chain of Example 7.3 shown in Fig. 7.4. Setting $\alpha = 0.5$ and $\beta = 0.7$ we found the transition probability matrix for this chain to be:

$$\mathbf{P} = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0.35 & 0.5 & 0.15 \\ 0.245 & 0.455 & 0.3 \end{bmatrix}$$

This chain is clearly irreducible. It is also aperiodic, since $p_{ii} > 0$ for all states $i = 0, 1, 2$ (as pointed out earlier, $p_{ii} > 0$ for at least one i is a sufficient condition for aperiodicity). It is also easy to see that the chain contains no transient or null recurrent states, so that Theorem 7.10 can be used to determine the unique stationary state probability vector $\boldsymbol{\pi} = [\pi_0, \pi_1, \pi_2]$. The set of equations (7.44) in this case is the following:

$$\begin{aligned} \pi_0 &= 0.5\pi_0 + 0.35\pi_1 + 0.245\pi_2 \\ \pi_1 &= 0.5\pi_0 + 0.5\pi_1 + 0.455\pi_2 \\ \pi_2 &= 0\pi_0 + 0.15\pi_1 + 0.3\pi_2 \end{aligned}$$

These equations are not linearly independent: One can easily check that multiplying the first and third equations by -1 and adding them gives the second equation. This is always the case in (7.44), which makes the normalization condition (7.45) necessary in order to solve for $\boldsymbol{\pi}$. Keeping the second and third equation above, and combining it with (7.45), we get

$$\begin{aligned} 0.5\pi_0 - 0.5\pi_1 + 0.455\pi_2 &= 0 \\ 0.15\pi_1 - 0.7\pi_2 &= 0 \\ \pi_0 + \pi_1 + \pi_2 &= 1 \end{aligned}$$

The solution of this set of equations is:

$$\pi_0 = 0.399, \quad \pi_1 = 0.495, \quad \pi_2 = 0.106$$

It is interesting to compare the stationary state probability vector $\boldsymbol{\pi} = [0.399, 0.495, 0.106]$ obtained above with the transient solution $\boldsymbol{\pi}(3) = [0.405875, 0.496625, 0.0975]$ in (7.16), which was obtained in Example 7.5 with initial state probability vector $\boldsymbol{\pi}(0) = [1, 0, 0]$. We can see that $\boldsymbol{\pi}(3)$ is an approximation of $\boldsymbol{\pi}$. This approximation gets better as k increases, and, by Theorem 7.10, we expect $\boldsymbol{\pi}(k) \to \boldsymbol{\pi}$ as $k \to \infty$.

Example 7.13 (Steady-state solution of birth–death chain)

Let us come back to the birth–death chain of Example 7.10. By looking at Fig. 7.9, we can see that the transition probability matrix is

$$\mathbf{P} = \begin{bmatrix} p & 1-p & 0 & 0 & 0 & \cdots \\ p & 0 & 1-p & 0 & 0 & \cdots \\ 0 & p & 0 & 1-p & 0 & 0 \\ 0 & 0 & p & 0 & 1-p & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

Assuming $0 < p < 1$, this chain is irreducible and aperiodic (note that $p_{00} = p > 0$). The system of equations $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ in (7.44) gives

$$\pi_0 = \pi_0 p + \pi_1 p$$
$$\pi_j = \pi_{j-1}(1-p) + \pi_{j+1}p, \qquad j = 1, 2, \ldots$$

From the first equation, we get

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

From the second set of equations, for $j = 1$ we get

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

and substituting for π_1 from above we obtain π_2 in terms of π_0:

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

Proceeding in similar fashion, we have

$$\pi_j = \left(\frac{1-p}{p}\right)^j \pi_0, \qquad j = 1, 2, \ldots \tag{7.46}$$

Summing over $j = 0, 1, \ldots$ and making use of the normalization condition (7.45), we obtain

$$\pi_0 + \sum_{j=1}^{\infty} \pi_j = \pi_0 + \pi_0 \sum_{j=1}^{\infty} \left(\frac{1-p}{p}\right)^j = 1$$

from which we can solve for π_0:

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

where we have replaced the summation index j by i so that there is no confusion in the following expression which we can now obtain from (7.46):

$$\pi_j = \frac{\left(\frac{1-p}{p}\right)^j}{\sum_{i=0}^{\infty} \left(\frac{1-p}{p}\right)^i}, \qquad j = 1, 2, \ldots \tag{7.47}$$

Now let us take a closer look at the infinite sum above. If $(1-p)/p < 1$, or equivalently $p > 1/2$, the sum converges,

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

and we have the final result

$$\pi_j = \frac{2p-1}{p}\left(\frac{1-p}{p}\right)^j, \qquad j = 0, 1, 2, \ldots \tag{7.48}$$

Now let us relate these results to our findings in Example 7.10:

3. Under the condition $p < 1/2$ we had found the chain to be transient. Under this condition, the sum in (7.47) does not converge, and we get $\pi_j = 0$; this is consistent with Theorem 7.9 for transient states.
4. Under the condition $p \geq 1/2$ we had found the chain to be recurrent. This is consistent with the condition $p > 1/2$ above, which, by (7.48), yields stationary state probabilities such that $0 < \pi_j < 1$.
5. Finally, note in (7.48) that as $p \to 1/2, \pi_j \to 0$. By (7.43), this implies that $M_j \to \infty$. Thus, we see that state 0 is null recurrent for $p = 1/2$. This was precisely our original conjecture in Example 7.10.

From a practical standpoint, Theorem 7.10 allows us to characterize the steady state behavior of many DES modeled as discrete-time Markov chains. The requirements of irreducibility and aperiodicity are not overly restrictive. Most commonly designed systems have these properties. For instance, one would seldom want to design a reducible resource-providing system which inevitably gets trapped into some closed sets of states.[1] Another practical implication of Theorem 7.10 is the following. Suppose that certain states in a DES are designated as "more desirable" than others. Since π_j is the fraction of time spent at j in the long run, it gives us a measure of system performance: Larger values of π_j for more desirable states j imply better performance. In some cases, maximizing (or minimizing) a particular π_j represents an actual design objective for such systems.

Example 7.14

Consider a machine which alternates between an UP and a DOWN state, denoted by 1 and 0 respectively. We would like the machine to spend as little time as possible in the DOWN state, and we can control a single parameter β which affects the probability of making a transition from DOWN to UP. We model this system through a Markov chain as shown in Fig. 7.14, where β ($0 \leq \beta \leq 2$ so that the transition probability 0.5β is in [0, 1]) is the design parameter we can select. Our design objective is expressed in terms of the stationary state probability π_0 as follows:

$$\pi_0 < 0.4$$

The transition probability matrix for this chain is

$$\mathbf{P} = \begin{bmatrix} 1 - 0.5\beta & 0.5\beta \\ 0.5 & 0.5 \end{bmatrix}$$

Using (7.44) and (7.45) to obtain the stationary state probabilities, we have

$$\begin{aligned} \pi_0 &= (1 - 0.5\beta)\pi_0 + 0.5\pi_1 \\ \pi_1 &= 0.5\beta\pi_0 + 0.5\pi_1 \\ \pi_0 + \pi_1 &= 1 \end{aligned}$$

Once again, the first two equations are linearly dependent. Solving the second and third equations for π_0, π_1 we get

$$\pi_0 = \frac{1}{1+\beta}, \qquad \pi_1 = \frac{\beta}{1+\beta}$$

[1] A supervisory controller S of the type considered in Chap. 3 could be synthesized, if necessary, to ensure that the controlled DES S/G (now modeled as a Markov chain) satisfies these requirements. One would rely upon the notions of marked states and nonblocking supervisor for this purpose.

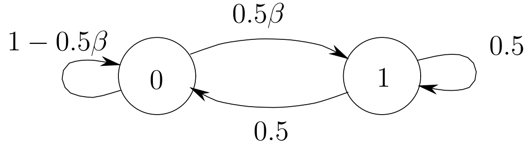


Figure 7.14: Markov chain for Example 7.14.

Since the design objective is that $\pi_0 < 0.4$, we obtain the following requirement for the parameter β:

$$\frac{1}{1+\beta} < 0.4 \quad \text{or} \quad \beta > 1.5$$

Thus, given the additional constraint $0 \leq \beta \leq 2$, we can select any β such that $1.5 < \beta \leq 2$.

7.2.11 Reducible Markov Chains

In a reducible Markov chain the steady state behavior is quite predictable: The chain eventually enters some irreducible closed set of states S and remains there forever. If S consists of two or more states, we can analyze the steady state behavior of S as in the previous section. If S consists of a single absorbing state, then the chain simply remains in that state (as is the case with state 0 in the gambler's ruin chain of Fig. 7.5). The only remaining problem arises when the reducible chain contains two or more irreducible closed sets. Then, the question of interest is: *What is the probability that the chain enters a particular set S first?* Clearly, if the chain enters S first it remains in that set forever and none of the other closed sets is ever entered.

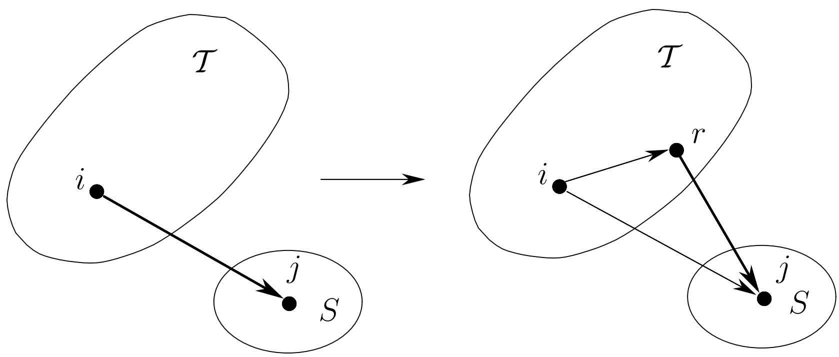


Figure 7.15: Eventual transition from a transient state to an irreducible closed set S. *State j is entered from i in one of two ways: in a single step, or by visiting some state $r \in \mathcal{T}$ first. In the latter case, the process repeats from the new state r.*

Let $\mathcal{T}$ denote the set of transient states in a reducible Markov chain, and let S be some irreducible closed set of states. Let $i \in \mathcal{T}$ be a transient state. We define $\rho_i(S)$ to be the probability of entering the set S given that the chain starts out at state i, that is,

$$\rho_i(S) \equiv P[X_k \in S \text{ for some } k > 0 \mid X_0 = i] \tag{7.49}$$

The event $[X_k \in S \text{ for some } k > 0 \mid X_0 = i]$ can occur in one of two distinct ways:

1. The chain enters S at $k = 1$, that is, the event $[X_1 \in S \mid X_0 = i]$ occurs, or

2. The chain visits some other state $r \in \mathcal{T}$ at $k = 1$ and then eventually enters S, that is, the event $[X_1 \in \mathcal{T}$ and $X_k \in S$ for some $k > 1 \mid X_0 = i]$ occurs.

This is illustrated in Fig. 7.15, where the thinner arrows represent one-step transitions. We can therefore write

$$\begin{aligned} \rho_i(S) = P[X_1 \in S \mid X_0 = i] \\ + P[X_1 \in \mathcal{T} \;\; and \;\; X_k \in S \text{ for some } k > 1 \mid X_0 = i] \end{aligned} \tag{7.50}$$

The first term in (7.50) can be rewritten as the sum of probabilities over all states $j \in S$, that is,

$$P[X_1 \in S \mid X_0 = i] = \sum_{j \in S} P[X_1 = j \mid X_0 = i] = \sum_{j \in S} p_{ij} \tag{7.51}$$

Similarly, the second term in (7.50) can be rewritten as the sum of probabilities over all states $r \in \mathcal{T}$:

$$\begin{aligned} &P[X_1 \in \mathcal{T} \;\; and \;\; X_k \in S \text{ for some } k > 1 \mid X_0 = i] \\ &\quad = \sum_{r \in \mathcal{T}} P[X_1 = r \;\; and \;\; X_k \in S \text{ for some } k > 1 \mid X_0 = i] \end{aligned} \tag{7.52}$$

Each term in this summation can be written as follows:

$$\begin{aligned} &P[X_1 = r \;\; and \;\; X_k \in S \text{ for some } k > 1 \mid X_0 = i] \\ &\quad = P[X_k \in S \text{ for some } k > 1 \mid X_1 = r, X_0 = i] \cdot P[X_1 = r \mid X_0 = i] \\ &\quad = P[X_k \in S \text{ for some } k > 1 \mid X_1 = r, X_0 = i] \cdot p_{ir} \end{aligned} \tag{7.53}$$

By the memoryless property,

$$\begin{aligned} &P[X_k \in S \text{ for some } k > 1 \mid X_1 = r, X_0 = i] \\ &\quad = P[X_k \in S \text{ for some } k > 1 \mid X_1 = r] \end{aligned}$$

and by setting $n = k - 1$, we get

$$\begin{aligned} P[X_k \in S \text{ for some } k > 1 \mid X_1 = r] &= P[X_n \in S \text{ for some } n > 0 \mid X_0 = r] \\ &= \rho_r(S) \end{aligned}$$

Thus, (7.53) becomes

$$P[X_1 = r \;\; and \;\; X_k \in S \text{ for some } k > 1 \mid X_0 = i] = \rho_r(S) \cdot p_{ir}$$

and hence (7.52) becomes

$$P[X_1 \in \mathcal{T} \;\; and \;\; X_k \in S \text{ for some } k > 1 \mid X_0 = i] = \sum_{r \in \mathcal{T}} \rho_r(S) \cdot p_{ir} \tag{7.54}$$

Finally, using (7.51) and (7.54) in (7.50), we get

$$\rho_i(S) = \sum_{j \in S} p_{ij} + \sum_{r \in \mathcal{T}} \rho_r(S) \cdot p_{ir} \tag{7.55}$$

In general, the solution of (7.55) for the unknown probabilities $\rho_i(S)$ for all $i \in \mathcal{T}$ is not easy to obtain. Moreover, if the set of transient states $\mathcal{T}$ is infinite, the solution of (7.55) may not even be unique. If, however, the set $\mathcal{T}$ is finite, it can be formally shown that the set of equations in (7.55) has a unique solution.

Example 7.15
The four-state Markov chain of Fig. 7.16 consists of a transient state set $\mathcal{T} = \{1, 2\}$ and the absorbing states 0 and 3. Suppose the initial state is known to be $X_0 = 1$. We now apply (7.55) to determine the probability $\rho_1(0)$ that the chain is absorbed by state 0 rather than state 3. Thus, we obtain the following two equations with two unknown probabilities, $\rho_1(0)$ and $\rho_2(0)$:

$$\rho_1(0) = p_{10} + \sum_{r=1}^{2} \rho_r(0) \cdot p_{1r} = 0.3 + 0.3\rho_1(0) + 0.4\rho_2(0)$$

$$\rho_2(0) = p_{20} + \sum_{r=1}^{2} \rho_r(0) \cdot p_{2r} = 0.2\rho_1(0) + 0.5\rho_2(0)$$

Solving for $\rho_1(0)$ we get

$$\rho_1(0) = 5/9$$

It follows that $\rho_1(3) = 4/9$, which can also be explicitly obtained by solving a similar set of equations in $\rho_1(3)$ and $\rho_2(3)$.

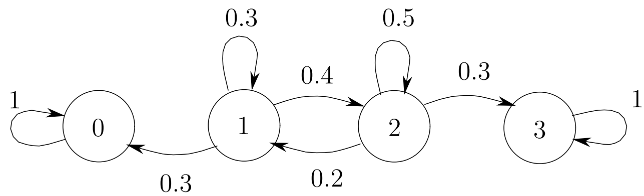


Figure 7.16: Markov chain for Example 7.15.

7.3 CONTINUOUS-TIME MARKOV CHAINS

In the case of a continuous-time Markov chain, the Markov (memoryless) property is expressed as:

$$\begin{aligned} &P[X(t_{k+1}) = x_{k+1} \mid X(t_k) = x_k, X(t_{k-1}) = x_{k-1}, \ldots, X(t_0) = x_0] \\ &\quad = P[X(t_{k+1}) = x_{k+1} \mid X(t_k) = x_k] \end{aligned} \tag{7.56}$$

for any $t_0 \leq t_1 \leq \cdots \leq t_k \leq t_{k+1}$. Thus, if the current state x_k is known, the value taken by $X(t_{k+1})$ depends only on x_k and not on any past state history (no state memory). Moreover, the amount of time spent in the current state is irrelevant in determining the next state (no age memory).

The study of continuous-time Markov chains parallels that of discrete-time chains. We can no longer, however, use the (one-step) transition probability matrix $\mathbf{P}$ introduced in (7.15), since state transitions are no longer synchronized by a common clock imposing a discrete-time structure. Instead, we must make use of a device measuring the *rates* at which various events (state transitions) take place. This will allow us to specify a model and then proceed with the analysis.

7.3.1 Model Specification

In the discrete-time case, a Markov chain model consisted of a state space $\mathcal{X}$, an initial state probability distribution $p_0(x)$ for all $x \in \mathcal{X}$, and a transition probability matrix $\mathbf{P}$. In continuous time, as pointed out above, we cannot use $\mathbf{P}$, since state transitions may occur at any time. Instead, we need to specify a matrix $\mathbf{P}(\tau)$ whose (i, j)th entry, $p_{ij}(t)$, is the probability of a transition from i to j within a time interval of duration t for all possible t. Clearly, this is a much more challenging task, since we need to specify all values of the functions $p_{ij}(t)$. As we will see, one solution to this problem is offered by introducing information regarding the *rate* at which various state transitions (events) may occur at any time.

In the next few sections, we introduce the basic definitions and notation leading to the specification and analysis of a continuous-time Markov chain model. In most cases, the definitions are similar to those of the discrete-time case.

7.3.2 Transition Functions

Recall our definition of the n-step transition probability $p_{ij}(k, k+n)$ in (7.4). A special case of $p_{ij}(k, k+n)$ was the one-step transition probability $p_{ij}(k) = p_{ij}(k, k+1)$, based on which the matrix $\mathbf{P}$ was defined.

Our starting point in the continuous-time case is the analog of $p_{ij}(k, k+n)$ defined in (7.4). In particular, we define time-dependent transition probabilities as follows:

$$p_{ij}(s,t) \equiv P[X(t) = j \mid X(s) = i], \qquad s \leq t \tag{7.57}$$

We will refer to $p_{ij}(s, t)$ as a *transition function*; it is a function of the time instants s and t. We reserve the term "transition probability" for a different quantity, which will naturally come up in subsequent sections.

The transition functions in (7.57) satisfy a continuous-time version of the Chapman-Kolmogorov equation (7.6). To derive this equation, we proceed by conditioning the event $[X(t) = j|X(s) = i]$ on $[X_u = r]$ for some u such that $s \leq u \leq t$. Then, using the rule of total probability, (7.57) becomes

$$p_{ij}(s,t) = \sum_{\text{all } r} P[X(t) = j \mid X(u) = r, X(s) = i] \cdot P[X(u) = r \mid X(s) = i] \tag{7.58}$$

By the memoryless property (7.56),

$$P[X(t) = j \mid X(u) = r, X(s) = i] = P[X(t) = j \mid X(u) = r] = p_{rj}(u,t)$$

Moreover, the second term in the sum in (7.58) is simply $p_{ir}(s, u)$. Therefore,

$$p_{ij}(s,t) = \sum_{\text{all } r} p_{ir}(s,u)p_{rj}(u,t), \qquad s \leq u \leq t \tag{7.59}$$

which is the continuous-time *Chapman-Kolmogorov equation*, the analog of (7.6).

To rewrite (7.59) in matrix form, we define:

$$\mathbf{H}(s,t) \equiv [p_{ij}(s,t)], \qquad i,j = 0,1,2,\ldots \tag{7.60}$$

and observe that $\mathbf{H}(s, s) = \mathbf{I}$ (the identity matrix). We then get from (7.59):

$$\mathbf{H}(s,t) = \mathbf{H}(s,u)\mathbf{H}(u,t), \qquad s \leq u \leq t \tag{7.61}$$

As in the discrete-time case, the Chapman-Kolmogorov equation is one of the most general relationships we can derive.

7.3.3 The Transition Rate Matrix

Let us consider the Chapman-Kolmogorov equation (7.61) for time instants $s \leq t \leq t + \Delta t$, where $\Delta t > 0$. Thus, we have:

$$\mathbf{H}(s, t + \Delta t) = \mathbf{H}(s, t)\mathbf{H}(t, t + \Delta t)$$

Subtracting $\mathbf{H}(s, t)$ from both sides of this equation gives

$$\mathbf{H}(s, t + \Delta t) - \mathbf{H}(s, t) = \mathbf{H}(s, t)[\mathbf{H}(t, t + \Delta t) - \mathbf{I}]$$

where $\mathbf{I}$ is the identity matrix. Dividing by Δt and taking the limit as $\Delta t \to 0$, we get

$$\lim_{\Delta t \to 0} \frac{\mathbf{H}(s, t + \Delta t) - \mathbf{H}(s, t)}{\Delta t} = \mathbf{H}(s, t) \lim_{\Delta t \to 0} \frac{\mathbf{H}(t, t + \Delta t) - \mathbf{I}}{\Delta t} \tag{7.62}$$

Note that the left-hand side of (7.62) is the partial derivative of $\mathbf{H}(s, t)$ with respect to t, provided of course that the derivatives of the transition functions $p_{ij}(s, t)$ (the elements of $\mathbf{H}(s, t)$) actually exist. Let us also define

$$\mathbf{Q}(t) \equiv \lim_{\Delta t \to 0} \frac{\mathbf{H}(t, t + \Delta t) - \mathbf{I}}{\Delta t} \tag{7.63}$$

which is called the *Transition Rate Matrix* of the Markov chain, or the *Infinitesimal Generator* driving the transition matrix $\mathbf{H}(s, t)$. Then, (7.62) reduces to the matrix differential equation

$$\frac{\partial \mathbf{H}(s, t)}{\partial t} = \mathbf{H}(s, t)\mathbf{Q}(t), \qquad s \leq t \tag{7.64}$$

which is also referred to as the *forward Chapman-Kolmogorov equation* in continuous time. In similar fashion, choosing time instants $s \leq s + \Delta s \leq t$ instead of $s \leq t \leq t + \Delta t$, we can obtain the *backward Chapman-Kolmogorov equation*

$$\frac{\partial \mathbf{H}(s, t)}{\partial s} = -\mathbf{Q}(s)\mathbf{H}(s, t) \tag{7.65}$$

Concentrating on the forward equation (7.64), under certain conditions which the matrix $\mathbf{Q}(s)$ must satisfy, a solution of this equation can be obtained in the form of a matrix exponential function

$$\mathbf{H}(s, t) = \exp\left[\int_s^t \mathbf{Q}(\tau) d\tau\right] \tag{7.66}$$

where $\exp[\mathbf{A}t] = e^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \mathbf{A}^2 t^2/2! + \ldots$

As we will soon see, if the transition rate matrix $\mathbf{Q}(t)$ is specified, a complete model of a continuous-time Markov chain is obtained (along with its state space $\mathcal{X}$ and initial state probability distribution $p_0(x)$ for all $x \in \mathcal{X}$). However, if we are to use $\mathbf{Q}(t)$ as the basis of our models, we should be able to identify the elements $q_{ij}(t)$ of this matrix with actual physically measurable quantities. Thus, we will address the question: What exactly do the entries of $\mathbf{Q}(t)$ represent?

To simplify our discussion, we will (as in the discrete-time case) limit ourselves to homogeneous Markov chains. The definition and implications of homogeneity for continuous-time chains are presented in the next section.

7.3.4 Homogeneous Markov Chains

In a homogeneous discrete-time Markov chain the n-step transition probability $p_{ij}(k, k+n)$ is independent of k. In the continuous-time case we define a chain to be homogeneous if all transition functions $p_{ij}(s,t)$ defined in (7.57) are independent of the absolute time instants s,t, and depend only on the difference $(t-s)$. To make this more clear, we can rewrite (7.57) as

$$p_{ij}(s, s+\tau) = P[X(s+\tau) = j \mid X(s) = i]$$

Then, homogeneity requires that, for any time s, $p_{ij}(s, s+\tau)$ depends only on τ. We therefore denote the transition function by $p_{ij}(\tau)$:

$$p_{ij}(\tau) = P[X(s+\tau) = j \mid X(s) = i] \tag{7.67}$$

It follows that the matrix $\mathbf{H}(s, s+\tau)$ defined in (7.60) is also only dependent on τ. To distinguish this case from the general one, we will use the symbol $\mathbf{P}(\tau)$, and we have

$$\mathbf{P}(\tau) \equiv [p_{ij}(\tau)], \qquad i,j = 0,1,2,\ldots \tag{7.68}$$

where, similar to (7.3), we have

$$\sum_{\text{all } j} p_{ij}(\tau) = 1 \tag{7.69}$$

Looking at the definition of $\mathbf{Q}(t)$ in (7.63), note that in the homogeneous case we have $\mathbf{H}(t, t+\Delta t) = \mathbf{P}(\Delta t)$, and hence the transition rate matrix is independent of t, that is,

$$\mathbf{Q}(t) = \mathbf{Q} = \text{constant} \tag{7.70}$$

Finally, the forward Chapman-Kolmogorov equation (7.64) becomes

$$\frac{d\mathbf{P}(\tau)}{d\tau} = \mathbf{P}(\tau)\mathbf{Q} \tag{7.71}$$

with the following initial conditions (assuming that a state transition from any i to $j \neq i$ cannot occur in zero time):

$$p_{ij}(0) = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases} \tag{7.72}$$

The solution of (7.71) is of the form

$$\mathbf{P}(\tau) = e^{\mathbf{Q}\tau} \tag{7.73}$$

From this point on, we will limit our discussion to *homogeneous* Markov chains (unless explicitly stated otherwise).

7.3.5 State Holding Times

As in the discrete-time case, let $V(i)$ denote the amount of time spent at state i whenever this state is visited, which we also refer to as the *state holding time*. It is a fundamental property of continuous-time Markov chains that *the distribution of the state holding time* $V(i)$, $P[V(i) \leq t], t \geq 0$, *is exponential.* Thus,

$$P[V(i) \leq t] = 1 - e^{-\Lambda(i)t}, \qquad t \geq 0 \tag{7.74}$$

where $\Lambda(i) > 0$ is a parameter generally dependent on the state i. This is a direct consequence of the memoryless property (7.56) and should not come as a surprise given the results of the previous chapter; in particular, we saw in Sect. 6.8.3 that a GSMP with a Poisson clock structure reduces to a Markov chain and inherits the memoryless property of the Poisson process.

We will use a simple informal argument to justify (7.74). Suppose the chain enters state i at time T. Further, suppose that at time $T + s \geq T$, no state transition has yet occurred. We consider the conditional probability

$$P[V(i) > s + t \mid V(i) > s]$$

and observe that it can be rewritten as

$$P[V(i) > s + t \mid X(\tau) = i \text{ for all } T \leq \tau \leq T + s]$$

Now, by the memoryless property (7.56), the information "$X(\tau) = i$ for all $T \leq \tau \leq T + s$" can be replaced by "$X(T + s) = i$." That is, the chain behaves as if state i had just been entered at time $T + s$, and the probability of the above event is the same as that of the event $[V(i) > t]$. Therefore,

$$P[V(i) > s + t \mid V(i) > s] = P[V(i) > t]$$

We have already seen in Theorem 6.1 of Chap. 6 that the only probability distribution satisfying this property is the exponential, which immediately implies (7.74).

The interpretation of the parameter $\Lambda(i)$ follows from our discussion of the GSMP with a Poisson clock structure in Chap. 6. In particular, we saw in (6.58) that $\Lambda(i)$ is the sum of the Poisson rates of all active events at state i. In the case of a Markov chain, an "event" is identical to a "state transition," so "interevent times" are identical to "state holding times." Let us, therefore, define e_{ij} to be events generated by a Poisson process with rate λ_{ij} which cause transitions from state i to state $j \neq i$. If two or more underlying "physical" events can cause such a transition, we do not care to distinguish them, and simply refer to any one of them as "e_{ij}." Then, as long as e_{ij} is a feasible event at state i, $\Lambda(i)$ is given by (6.58):

$$\Lambda(i) = \sum_{e_{ij} \in \Gamma(i)} \lambda_{ij} \tag{7.75}$$

where $\Gamma(i)$ is the set of feasible events at state i. In other words, the parameter $\Lambda(i)$, which fully characterizes the holding time distribution for state i, may be thought of as the sum of all Poisson rates corresponding to feasible events at state i.

7.3.6 Physical Interpretation and Properties of the Transition Rate Matrix

We now attempt to gain insight into the meaning of the elements of the transition rate matrix $\mathbf{Q}$. Let us go back to the forward Chapman-Kolmogorov equation (7.71), and write down the individual scalar differential equations:

$$\frac{dp_{ij}(\tau)}{d\tau} = p_{ij}(\tau)q_{jj} + \sum_{r \neq j} p_{ir}(\tau)q_{rj} \tag{7.76}$$

We first concentrate on the case $i = j$. Thus, (7.76) is written as

$$\frac{dp_{ii}(\tau)}{d\tau} = p_{ii}(\tau)q_{ii} + \sum_{r \neq i} p_{ir}(\tau)q_{ri}$$

Setting $\tau = 0$ and using the initial conditions (7.72) we get

$$\left.\frac{dp_{ii}(\tau)}{d\tau}\right|_{\tau=0} = q_{ii} \tag{7.77}$$

which can be rewritten as

$$-q_{ii} = \left.\frac{d}{d\tau}[1 - p_{ii}(\tau)]\right|_{\tau=0}$$

Here, $[1 - p_{ii}(\tau)]$ is the probability that the chain leaves state i in an interval of length τ. Thus, we see that $-q_{ii}$ is the *instantaneous rate* at which a state transition out of i takes place. It describes the propensity the chain has to leave state i.

Some additional insight into the meaning of $-q_{ii}$ is obtained by returning to (7.74) and attempting to link q_{ii} to $\Lambda(i)$. Suppose the state is i at some time instant t and consider the interval $(t, t + \tau]$, where τ can be made arbitrarily small to guarantee that at most one event (and hence state transition) can take place in $(t, t + \tau]$. Now suppose that no state transition occurs in this interval. Thus,

$$p_{ii}(\tau) = P[V(i) > \tau] = e^{-\Lambda(i)\tau}$$

Taking derivatives of both sides, we get

$$\frac{dp_{ii}(\tau)}{d\tau} = -\Lambda(i)e^{-\Lambda(i)\tau}$$

Evaluating both sides at $\tau = 0$ and using (7.77) yields:

$$-q_{ii} = \Lambda(i) \tag{7.78}$$

In other words, $-q_{ii}$ represents the *total event rate* characterizing state i. This event rate also captures the propensity of the chain to leave state i, since a large value of $\Lambda(i)$ indicates a greater likelihood of some event occurrence and hence state transition out of i.

Next, returning to (7.76) for the case $j \neq i$, we obtain similar to (7.77):

$$q_{ij} = \left.\frac{dp_{ij}(\tau)}{d\tau}\right|_{\tau=0} \tag{7.79}$$

Here the interpretation of q_{ij} is the *instantaneous rate* at which a state transition from i to j takes place. Moreover, similar to (7.78), we have

$$q_{ij} = \lambda_{ij} \tag{7.80}$$

which is the Poisson rate of the event e_{ij} causing transitions from i to j.

From (7.78), since $\Lambda(i) > 0$, it follows that $q_{ii} < 0$. Therefore, all diagonal elements of the matrix $\mathbf{Q}$ must be negative. From (7.80), on the other hand, we see that $q_{ij} > 0$, if the event e_{ij} is feasible in state i; otherwise, $q_{ij} = 0$. Thus, all off-diagonal terms must be nonnegative. In addition, since $\sum_j p_{ij}(\tau) = 1$ in (7.69), differentiating with respect to τ and setting $\tau = 0$ we get

$$\sum_{\text{all } j} q_{ij} = 0 \tag{7.81}$$

7.3.7 Transition Probabilities

We are now in a position to see how the transition rate matrix $\mathbf{Q}$ specifies the Markov chain model. Let us define the *transition probability*, P_{ij}, of a continuous-time Markov chain as follows. Suppose state transitions occur at random time instants $T_1 < T_2 < \cdots < T_k < \cdots$ The state following the transition at T_k is denoted by X_k. We then set

$$P_{ij} = P[X_{k+1} = j \mid X_k = i] \tag{7.82}$$

Let us recall (6.67) of Chap. 6, where we derived the probability distribution of events at a given state for a GSMP with Poisson clock structure. In particular, using our Markov chain notation, given that the state is i, the probability that the next event is e_{ij} (therefore, the next state is j) is given by $\lambda_{ij}/\Lambda(i)$. Using (7.78) and (7.80) we now see that the transition probability P_{ij} is expressed in terms of elements of $\mathbf{Q}$:

$$P_{ij} = \frac{q_{ij}}{-q_{ii}}, \qquad j \neq i \tag{7.83}$$

Moreover, summing over all $j \neq i$ above and recalling (7.81), we get $\sum_{j \neq i} P_{ij} = 1$. This implies that $P_{ii} = 0$ as expected, since the only events defined for a Markov chain are those causing actual state transitions.

In summary, once the transition rate matrix $\mathbf{Q}$ is specified, we have at our disposal a full model specification:

- The state transition probabilities P_{ij} are given by (7.83)
- The parameters of the exponential state holding time distributions are given by the corresponding diagonal elements of $\mathbf{Q}$, that is, for state i, $-q_{ii} = \sum_{j \neq i} q_{ij}$.

Conversely, note that a Markov chain can also be specified through:

- Parameters $\Lambda(i)$ for all states i, characterizing the exponential state holding time distributions.
- Transition probabilities P_{ij} for all pairs of states $i \neq j$.

Example 7.16

In this example, we provide the physical description of a simple DES, and obtain a continuous-time Markov chain model.

The DES is a queueing system with a total capacity of two customers (Fig. 7.17). There are two different events that can occur: an arrival a, and a departure d. However, the system is designed so that service is only provided to two customers simultaneously (in manufacturing, for example, this situation arises in *assembly* operations, where a machine combines two or more parts together). Using our standard DES notation from previous chapters, we have a state space

$$\mathcal{X} = \{0, 1, 2\}$$

an event set

$$\mathcal{E} = \{a, d\}$$

and feasible event sets

$$\Gamma(x) = \{a\} \text{ for } x = 0, 1, \qquad \Gamma(2) = \{a, d\}$$

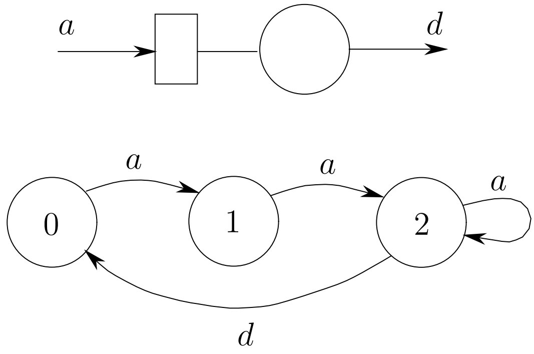


Figure 7.17: Queueing system and state transition diagram for Example 7.16.
Service in this system is provided only when two customers are present, at which time they are combined together.

The state transition diagram for the system is shown in Fig. 7.17. Departure events are only feasible at state 2 and result in emptying out the system. Arrivals are feasible at state 2, but they cause no state transition, since the capacity of the system is limited to two customers.

By assuming that events a and d both occur according to Poisson processes with rates λ and μ respectively, we obtain a Markov chain model of this system. The transition rate matrix $\mathbf{Q}$ is given by

$$\mathbf{Q} = \begin{bmatrix} -\lambda & \lambda & 0 \\ 0 & -\lambda & \lambda \\ \lambda & 0 & \lambda \end{bmatrix}$$

The Markov chain model can also be specified through the transition probabilities P_{ij} for all $i \neq j, i, j = 0, 1, 2$. By looking at the state transition diagram in Fig. 7.17, we immediately see that

$$P_{01} = P_{12} = P_{20} = 1$$

which is consistent with (7.83). In addition, we can specify the parameters of the exponential state holding time distributions as follows:

$$\Lambda(0) = -q_{00} = \lambda, \qquad \Lambda(1) = -q_{11} = \lambda, \qquad \Lambda(2) = -q_{22} = \mu$$

Example 7.17
A warehouse keeps in stock one unit of each of two types of products, P_1 and P_2. A truck periodically comes to take away one of these products. If both are in stock, the truck gives preference to P_1. In this DES, let $x_i = 1$ if P_i is present in the warehouse, $i = 1, 2$, and $x_i = 0$ otherwise. Thus, the state space is

$$\mathcal{X} = \{(x_1, x_2) : x_1 = 0, 1, x_2 = 0, 1\}$$

Let us assume that when $x_i = 0$, there is a Poisson rate λ_i with which P_i is replaced. Let us also assume that when a unit of P_1 is present at the warehouse, the truck arrives to pick it up with a Poisson rate μ_1; if only a unit of P_2 is present, then this rate is μ_2. We can now write a transition rate matrix $\mathbf{Q}$ for this model, labeling states

0, 1, 2 and 3 to correspond to $(0,0),(0,1),(1,0),(1,1)$ respectively:

$$\mathbf{Q} = \begin{bmatrix} -(\lambda_1+\lambda_2) & \lambda_2 & \lambda_1 & 0 \\ \mu_2 & -(\lambda_1+\mu_2) & 0 & \lambda_1 \\ \mu_1 & 0 & -(\lambda_2+\mu_1) & \lambda_2 \\ 0 & \mu_1 & 0 & -\mu_1 \end{bmatrix}$$

Note that $q_{10} = \mu_2$ is due to the fact that the only time the truck picks up a P_2 unit is at state (0,1).

We can now use (7.83) to determine all the transition probabilities. For example,

$$P_{01} = \frac{\lambda_2}{\lambda_1+\lambda_2}, \quad P_{20} = \frac{\mu_1}{\lambda_2+\mu_1}, \quad P_{31} = 1, \quad P_{03} = 0$$

In addition, state holding times are characterized by the exponential distribution parameters specified through (7.78). For example, $\Lambda(0) = \lambda_1 + \lambda_2$, and $\Lambda(3) = \mu_1$.

7.3.8 State Probabilities

Similar to the discrete-time case, we define state probabilities as follows:

$$\pi_j(t) \equiv P[X(t) = j] \tag{7.84}$$

Accordingly, we have a *state probability vector*

$$\boldsymbol{\pi}(t) = [\pi_0(t), \pi_1(t), \ldots] \tag{7.85}$$

This is a row vector whose dimension is specified by the dimension of the state space of the chain (not necessarily finite).

A continuous-time Markov chain model is completely specified by the state space $\mathcal{X}$ and the transition matrix $\mathbf{P}(\tau)$, and an initial state probability vector

$$\boldsymbol{\pi}(0) = [\pi_0(0), \pi_1(0), \ldots]$$

which provides the probability distribution of the initial state of the chain $X(0)$. By (7.73), however, $\mathbf{P}(\tau) = e^{\mathbf{Q}\tau}$. Therefore, the specification of $\mathbf{P}(\tau)$ is immediately provided by the transition rate matrix $\mathbf{Q}$.

7.3.9 Transient Analysis

Our main objective is to determine the probability vector $\boldsymbol{\pi}(t)$ given a Markov chain specified by its state space $\mathcal{X}$, transition rate matrix $\mathbf{Q}$, and initial state probability vector $\boldsymbol{\pi}(0)$. Starting out with the definition of $\pi_j(t)$ in (7.84), let us condition the event $[X(t) = j]$ on the event $[X(0) = i]$, and use the rule of total probability to obtain

$$\begin{aligned} \pi_j(t) = P[X(t) = j] &= \sum_{\text{all } i} P[X(t) = j \mid X(0) = i] \cdot P[X(0) = i] \\ &= \sum_{\text{all } i} p_{ij}(t)\pi_i(0) \end{aligned}$$

Recalling the definition of the transition matrix in (7.68), we can rewrite this relationship in matrix form:

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)\mathbf{P}(t) \tag{7.86}$$

and since $\mathbf{P}(t) = e^{\mathbf{Q}t}$, the state probability vector at time t is given by

$$\boldsymbol{\pi}(t) = \boldsymbol{\pi}(0)e^{\mathbf{Q}t} \tag{7.87}$$

Therefore, in principle, one can always obtain a solution describing the transient behavior of a chain characterized by $\mathbf{Q}$ and an initial condition $\boldsymbol{\pi}(0)$. However, obtaining explicit expressions for the individual state probabilities $\pi_j(t), j = 0, 1, 2, \ldots$ is far from simple, as we shall see later on in this chapter. The time functions $\pi_j(t)$ are fairly complicated, even for the simplest Markov chain models of DES.

Note that by differentiating (7.87) with respect to t, we can obtain the differential equation

$$\frac{d\boldsymbol{\pi}(t)}{dt} = \boldsymbol{\pi}(t)\mathbf{Q} \tag{7.88}$$

which is of the same form as the forward Chapman-Kolmogorov equation (7.71). In practice, it is often useful to write (7.88) as a set of scalar differential equations

$$\frac{d\pi_j(t)}{dt} = q_{jj}\pi_j(t) + \sum_{i \neq j} q_{ij}\pi_i(t) \tag{7.89}$$

for all $j = 0, 1, \ldots$ Note that we have separated the $i = j$ element from the sum above in order to emphasize the fact that the summation term is independent of $\pi_j(t)$.

Clearly, if the matrix $\mathbf{Q}$ is specified, then (7.89) can be directly obtained. We have seen, however, that Markov chains are often more efficiently described in terms of state transition diagrams. In such diagrams, nodes represent states and a directed arc from node i to node j represents a transition from i to j. We obtain a state transition rate diagram by associating to an arc (i, j) the transition rate q_{ij}. This is illustrated in Fig. 7.18 for a typical state j. Such a diagram provides a simple graphical device for deriving (7.89) by inspection. This is done as follows.

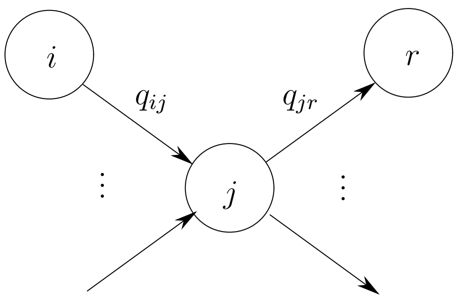


Figure 7.18: Probability flow balance for state j.

We view $\pi_j(t)$ as the level of a "probability fluid" residing in node j and taking on values between 0 (empty) and 1 (full). The transition rate q_{ij} represents the "probability flow rate" from i to j. Hence, the flow from i to j at time t is given by $q_{ij}\pi_i(t)$. We can then write simple flow balance equations for the "probability fluid" level of state j by evaluating the total incoming flow at j and the total outgoing flow from j. Referring to Fig. 7.18, we see that

$$\text{Total flow into state } j = \sum_{i \neq j} q_{ij}\pi_i(t)$$

$$\text{Total flow out of state } j = \sum_{r \neq j} q_{jr}\pi_j(t)$$

Therefore, the net probability flow rate into state j is described by

$$\frac{d\pi_j(t)}{dt} = \sum_{i \neq j} q_{ij}\pi_i(t) - \left[\sum_{r \neq j} q_{jr}\right] \pi_j(t)$$

From (7.81), we have

$$\sum_{r \neq j} q_{jr} = -q_{jj}$$

so that the flow balance equation above becomes

$$\frac{d\pi_j(t)}{dt} = q_{jj}\pi_j(t) + \sum_{i \neq j} q_{ij}\pi_i(t)$$

which is precisely (7.89). Thus, the state transition rate diagram contains the exact same information as the transition rate matrix $\mathbf{Q}$.

Example 7.18

We return to the Markov chain model in Example 7.16. To obtain the state transition rate diagram shown in Fig. 7.19, observe that the probability flow out of state 0 is only due to events a occurring with rate λ, whereas the probability flow into state 0 is only due to events d occurring at state 2 with rate μ. Similarly, at state 1 we have an incoming flow from 1 with rate λ, and an outgoing flow due to a events with rate λ. Finally, state 2 receives a probability flow from state 2 with rate λ, and has an outgoing flow due to d events with rate μ. The state transition rate diagram also corresponds directly to the transition rate matrix $\mathbf{Q}$ derived in Example 7.16:

$$\mathbf{Q} = \begin{bmatrix} -\lambda & \lambda & 0 \\ 0 & -\lambda & \lambda \\ \mu & 0 & -\mu \end{bmatrix}$$

To determine the state probability vector $\boldsymbol{\pi}(t) = [\pi_0(t), \pi_1(t), \pi_2(t)]$ at any time t, we need to obtain the differential equations (7.89). By inspection of the state transition rate diagram in Fig. 7.19, we can easily write down the following probability flow balance equations:

$$\frac{d\pi_0(t)}{dt} = \mu\pi_2(t) - \lambda\pi_0(t)$$
$$\frac{d\pi_1(t)}{dt} = \lambda\pi_0(t) - \lambda\pi_1(t)$$
$$\frac{d\pi_2(t)}{dt} = \lambda\pi_1(t) - \mu\pi_2(t)$$

If an initial condition $\boldsymbol{\pi}(0) = [\pi_0(0), \pi_1(0), \pi_2(0)]$ is specified, these equations can be solved for $\pi_0(t), \pi_1(t)$, and $\pi_2(t)$. However, obtaining an explicit solution, even for this simple example, is not a trivial matter.

Example 7.19

We now reconsider the Markov chain model in Example 7.17. First, we obtain a state transition rate diagram as shown in Fig. 7.20, which corresponds to the transition rate

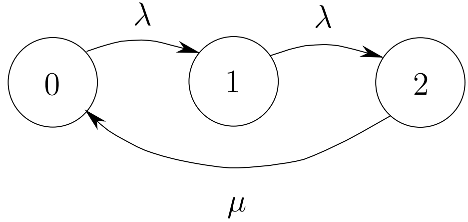


Figure 7.19: State transition rate diagram for Example 7.18.

matrix $\mathbf{Q}$:

$$\mathbf{Q} = \begin{bmatrix} -(\lambda_1+\lambda_2) & \lambda_2 & \lambda_1 & 0 \\ \mu_2 & -(\lambda_1+\mu_2) & 0 & \lambda_1 \\ \mu_1 & 0 & -(\lambda_2+\mu_1) & \lambda_2 \\ 0 & \mu_1 & 0 & -\mu_1 \end{bmatrix}$$

Recall that states (0,0), (0,1), (1,0), (1,1) are labeled 0,1,2,3 respectively. Also, recall that the truck always prioritizes P_1 units. Thus, from both (1,0) and from (1,1) there are probability flows with rate μ_1 to states (0,0) and (0,1) respectively.

By inspection of Fig. 7.20, we can now write down probability flow balance equations:

$$\frac{d\pi_0(t)}{dt} = \mu_2\pi_1(t) + \mu_1\pi_2(t) - (\lambda_1+\lambda_2)\pi_0(t)$$
$$\frac{d\pi_1(t)}{dt} = \lambda_2\pi_0(t) + \mu_1\pi_3(t) - (\lambda_1+\mu_2)\pi_1(t)$$
$$\frac{d\pi_2(t)}{dt} = \lambda_1\pi_0(t) - (\lambda_2+\mu_1)\pi_2(t)$$
$$\frac{d\pi_3(t)}{dt} = \lambda_1\pi_1(t) + \lambda_2\pi_2(t) - \mu_1\pi_3(t)$$

Once again, we see that obtaining an explicit solution for these equations is not an easy task.

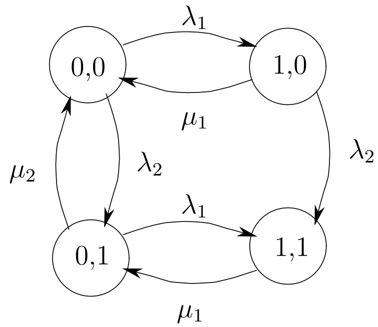


Figure 7.20: State transition rate diagram for Example 7.19.

7.3.10 Steady State Analysis

In the previous section, we saw that the behavior of a continuous-time Markov chain is completely described by the set of differential equations (7.89) or, in matrix form, (7.88). Although the solution to these equations can always be written in the matrix exponential form

(7.87), obtaining explicit closed-form expressions for the state probabilities $\pi_0(t), \pi_1(t), \ldots$ is generally a difficult task. Fortunately, in practice we are often only interested in the steady state behavior of a system; this is much easier to analyze than the transient behavior. In other words, we turn the system on, let it run for some time, and then examine its performance in the "long run." As in the discrete-time case, by "long run" we generally mean that the system has operated sufficiently long to allow all state probabilities to reach some fixed values, no longer varying with time. Of course, there is no guarantee that such values are indeed achievable. Therefore, our study centers around the existence and determination of the limits:

$$\pi_j = \lim_{t \to \infty} \pi_j(t) \tag{7.90}$$

As in the discrete-time case, we need to address three basic questions:

1. Under what conditions do the limits in (7.90) exist?
2. If these limits exist, do they form a legitimate probability distribution, that is, $\sum_j \pi_j = 1$?
3. How do we evaluate π_j?

If π_j exists for some state j, it is referred to as a *steady-state, equilibrium,* or *stationary state probability.* Accordingly, if π_j exists for all states j, then we obtain the *stationary state probability vector*

$$\boldsymbol{\pi} = [\pi_0, \pi_1, \ldots]$$

If the limits in (7.90) exist, it follows that as $t \to \infty$ the derivative $d\boldsymbol{\pi}/dt \to 0$, since $\boldsymbol{\pi}(t)$ no longer depends on t. We therefore expect that at steady state the differential equation (7.88):

$$\frac{d\boldsymbol{\pi}(t)}{dt} = \boldsymbol{\pi}(t)\mathbf{Q}$$

reduces to the algebraic equation

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

The analysis of the steady state behavior of continuous-time Markov chains is very similar to that of the discrete-time case, presented in Sects. 7.2.9 through 7.2.11. The concepts of irreducibility and recurrence are still valid, and results paralleling Theorems 7.9 and 7.10 can be formally obtained. Thus, we will limit ourselves here to stating the most important result for our purposes, pertaining to positive recurrent irreducible chains (see also Asmussen, 1987).

Theorem 7.11 In an irreducible continuous-time Markov chain consisting of positive recurrent states, a unique stationary state probability vector $\boldsymbol{\pi}$ exists such that $\pi_j > 0$ and

$$\pi_j = \lim_{t \to \infty} \pi_j(t)$$

These limits are independent of the initial state probability vector. Moreover, the vector $\boldsymbol{\pi}$ is determined by solving:

$$\boldsymbol{\pi}\mathbf{Q} = 0 \tag{7.91}$$

$$\sum_{\text{all } j} \pi_j = 1 \tag{7.92}$$

♦

As in the discrete-time case, it is possible to show that every *finite* irreducible Markov chain has a unique stationary state probability vector $\boldsymbol{\pi}$ determined through (7.91) and (7.92). Then, obtaining $\boldsymbol{\pi}$ is a matter of solving a set of linear algebraic equations. However, in the case of an infinite state space, it is generally a difficult task to solve (7.91) and (7.92). Fortunately, for a large class of Markov chains with a special structure, this task is somewhat simplified. This class of chains, called *birth–death chains*, is the subject of the next section.

Finally, note that (7.91) in scalar form is the set of equations (7.89) with $d\pi_j(t)/dt = 0$. In other words, we can still use the idea of probability flow balancing at steady state, where all the "probability fluid" levels occupying the nodes of the state transition rate diagram have all reached equilibrium. Thus, we have

$$q_{jj}\pi_j + \sum_{i \neq j} q_{ij}\pi_i = 0 \tag{7.93}$$

for all states $j = 0, 1, \ldots$ As in the discrete-time case, these equations are not linearly independent. This necessitates including the normalization condition (7.92) in order to obtain the unique stationary state probabilities $\pi_0, \pi_1, \pi_2, \ldots$

Example 7.20

Consider the system modeled and analyzed in Examples 7.16 and 7.18, whose state transition rate diagram is shown in Fig. 7.19. By inspection, we write down the probability flow balance equations at steady state, which are precisely the set of equations (7.91) or (7.93):

$$\begin{aligned}\mu\pi_2 - \lambda\pi_0 &= 0\\ \lambda\pi_0 - \lambda\pi_1 &= 0\\ \lambda\pi_1 - \mu\pi_2 &= 0\end{aligned}$$

The first two equations give

$$\pi_2 = \frac{\lambda}{\mu}\pi_0, \qquad \pi_1 = \pi_0$$

One can then see that the third equation is redundant. The solution is obtained by using the two equations above in conjunction with the normalization condition (7.92):

$$\pi_0 + \pi_1 + \pi_2 = 1$$

which becomes

$$2\pi_0 + \frac{\lambda}{\mu}\pi_0 = 1$$

Hence, we get

$$\pi_0 = \pi_1 = \frac{\mu}{2\mu + \lambda}, \qquad \pi_2 = \frac{\lambda}{2\mu + \lambda}$$

7.4 BIRTH–DEATH CHAINS

Continuous-time Markov chains provide a rich class of models for many stochastic DES of interest. Although we have seen how to analyze both the transient and steady-state behavior of these chains, obtaining explicit expressions for the state probabilities is not an easy task in practice. *Birth–death chains* form a limited class of Markov chains whose special structure facilitates this task, while still providing a sufficiently rich modeling framework.

In simple terms, a birth–death chain is one where transitions are only permitted to/from neighboring states. Since we have adopted a state space $\mathcal{X} = \{0, 1, 2, \ldots\}$, such transitions have a simple appealing interpretation if we think of the state as representing a certain "population" level (e.g., customers, jobs, messages): A transition from state i to state $i+1$ is a *birth*, whereas a transition from i to $i-1$ is a *death*. Obviously, no death can occur at state $i = 0$. Similarly, if the state space is finite, so that $\mathcal{X} = \{0, 1, 2, \ldots, K\}$ for some $K < \infty$, no birth can occur at state $i = K$. A birth–death chain can be defined for discrete or continuous time. A simple version of a discrete-time birth–death chain was considered in Examples 7.10 and 7.13. In what follows, we will limit ourselves to the continuous-time case.

Definition. A *birth–death chain* is a continuous-time Markov chain whose transition rate matrix $\mathbf{Q}$ satisfies

$$q_{ij} = 0 \qquad \text{for all } j > i+1 \text{ and } j < i-1 \tag{7.94}$$

♦

It is customary to define a *birth rate*

$$\lambda_j = q_{j,j+1} > 0, \qquad j = 0, 1, \ldots \tag{7.95}$$

and a death rate

$$\mu_j = q_{j,j-1} > 0, \qquad j = 1, 2, \ldots \tag{7.96}$$

where we normally assume all birth and death parameters above to be strictly positive. Note, however, that we do not necessarily constrain the chain to be homogeneous, since these rates may depend on the states.

By (7.81), we know that all elements in the jth row of $\mathbf{Q}$ must add to 0. Therefore, the diagonal elements q_{jj} are immediately obtained from (7.94) through (7.96) as follows:

$$\begin{aligned} q_{jj} &= -(\lambda_j + \mu_j) \qquad \text{for all } j = 1, 2, \ldots \\ q_{00} &= -\lambda_0 \end{aligned}$$

The general form of the transition rate matrix $\mathbf{Q}$ for a birth–death chain is as follows:

$$\mathbf{Q} = \begin{bmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \ldots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 & 0 & \ldots \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & 0 & 0 \\ 0 & 0 & \mu_3 & -(\lambda_3+\mu_3) & \lambda_3 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \tag{7.97}$$

The special structure of this model is clearly seen in the state transition rate diagram of Fig. 7.21, corresponding to $\mathbf{Q}$ above. We may think of births and deaths as events which can occur with generally different rates at various states. A birth or death at state $j > 0$ causes a transition to $j+1$ or $j-1$ respectively. Similarly, a transition to state $j > 0$ can only occur through a birth at state $j-1$ or a death at state $j+1$. State $j = 0$ represents a "boundary case" where only a birth can occur.

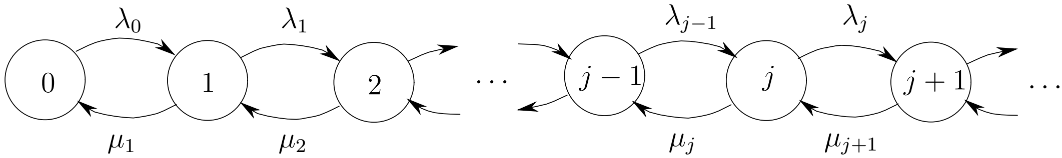


Figure 7.21: State transition rate diagram for a general birth–death chain.

The differential equations characterizing the state probabilities $\pi_0(t), \pi_1(t), \pi_2(t), \ldots$ of a birth–death chain can be obtained from the general case (7.94) by exploiting the special structure of $\mathbf{Q}$. In particular,

$$\frac{d\pi_j(t)}{dt} = -(\lambda_j + \mu_j)\pi_j(t) + \lambda_{j-1}\pi_{j-1}(t) + \mu_{j+1}\pi_{j+1}(t), \qquad j = 1, 2, \ldots \tag{7.98}$$

$$\frac{d\pi_0(t)}{dt} = -\lambda_0\pi_0(t) + \mu_1\pi_1(t) \tag{7.99}$$

Even though the solution of (7.98) and (7.99) is easier to obtain than the general case (7.89), it is still an elaborate task. Fortunately, the steady-state analysis of these equations is much simpler to handle, and forms the basis of much of the elementary queueing theory we will develop in the next chapter. In the next few sections, we examine some simple cases for which transient solutions can be obtained and then discuss the steady-state solution of the birth–death chain.

7.4.1 The Pure Birth Chain

In the pure birth chain, we have $\mu_j = 0$ for all $j = 1, 2, \ldots$ Equations (7.98) and (7.99) become

$$\frac{d\pi_j(t)}{dt} = -\lambda_j\pi_j(t) + \lambda_{j-1}\pi_{j-1}(t), \qquad j = 1, 2, \ldots \tag{7.100}$$

$$\frac{d\pi_0(t)}{dt} = -\lambda_0\pi_0(t) \tag{7.101}$$

The solution of (7.101) is easily obtained as an exponential function

$$\pi_0(t) = e^{-\lambda_0 t} \tag{7.102}$$

where we assume that $\pi_0(0) = 1$, that is, the initial state is 0 with probability 1. The solution of (7.100) for $j = 1, 2, \ldots$ is more difficult to obtain. Given (7.101), the following recursive expression allows us to evaluate $\pi_1(t), \pi_2(t), \ldots$:

$$\pi_j(t) = e^{-\lambda_j t}\lambda_{j-1}\int_0^t \pi_{j-1}(\tau)e^{\lambda_j\tau}d\tau, \qquad j = 1, 2, \ldots \tag{7.103}$$

By differentiating (7.103) it is easy to verify that this is indeed the solution of (7.100). It is not a very simple expression, which leads us to suspect (and this is in fact the case) that the general solution of (7.98) and (7.99) is far from simple.

7.4.2 The Poisson Process Revisited

Let us consider the pure birth chain with *constant birth rates*, that is,

$$\lambda_j = \lambda \qquad \text{for all } j = 0, 1, 2, \ldots$$

Then, (7.100) and (7.101) reduce to

$$\frac{d\pi_j(t)}{dt} = -\lambda\pi_j(t) + \lambda\pi_{j-1}(t), \qquad j = 1, 2, \ldots \tag{7.104}$$

$$\frac{d\pi_0(t)}{dt} = -\lambda\pi_0(t) \tag{7.105}$$

Returning to our discussion of the Poisson counting process in Sect. 6.2 in Chap. 6, we can immediately recognize that (7.104) is the same as (6.33), which characterizes the counter $N(t)$ taking values in $\{0, 1, 2, \ldots\}$. This is not surprising since counting events with an initial count of 0 is the same as counting births with an initial population of 0. In addition, we have seen that the Poisson process is inherently characterized by exponentially distributed interevent times with parameter λ; so is the birth chain whose only event is a "birth" occurring at rate λ.

In summary, the Poisson counting process may be obtained in one of two ways. From first principles, as a counting process satisfying the basic assumptions **(A1)** through **(A3)** of Sect. 6.2; or as a special case of a Markov chain (pure birth chain with constant birth rate λ). This observation serves to reinforce the strong connection between (a) the memoryless property, (b) the exponential distribution for interevent (or state holding) times, and (c) stationary independent increments in a stochastic process. All three characterize a class of stochastic processes for which the Poisson process is a fundamental building block.

The solution of (7.104) and (7.105) is of course the well-known by now Poisson distribution

$$\pi_j(t) = \frac{(\lambda t)^j}{j!}e^{-\lambda t}, \qquad t \geq 0, \qquad j = 0, 1, 2, \ldots \tag{7.106}$$

which was also derived in (6.34). Note that for $j = 0$ we obtain $\pi_0(t)$ in (7.102) with $\lambda_0 = \lambda$.

7.4.3 Steady State Analysis of Birth–Death Chains

We will now concentrate on the steady state solution of the birth–death chain, that is, the solution of (7.98) and (7.99) when $d\pi_j(t)/dt = 0$, if it exists:

$$-(\lambda_j + \mu_j)\pi_j + \lambda_{j-1}\pi_{j-1} + \mu_{j+1}\pi_{j+1} = 0, \qquad j = 1, 2, \ldots \tag{7.107}$$

$$-\lambda_0\pi_0 + \mu_1\pi_1 = 0 \tag{7.108}$$

Recall that in addition to these equations the normalization condition (7.92) must also be satisfied, that is $\sum_j \pi_j = 1$.

Ignoring for the time being the issue of existence, we can obtain a general solution to these equations by proceeding as follows.

Step 1. From (7.108), we obtain a relationship between π_1 and π_0:

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

Step 2. From (7.107) with $j = 1$, we get

$$\mu_2\pi_2 = -\lambda_0\pi_0 + (\lambda_1 + \mu_1)\pi_1$$

and using the result of *Step 1* above we obtain a relationship between π_2 and π_0:

$$\pi_2 = \frac{\lambda_0\lambda_1}{\mu_1\mu_2}\pi_0$$

Step 3. Repeat *Step 2* for all $j = 2, 3, \ldots$ to obtain a relationship between π_j and π_0:

$$\pi_j = \left(\frac{\lambda_0 \cdots \lambda_{j-1}}{\mu_1 \cdots \mu_j}\right)\pi_0 \tag{7.109}$$

This is the general solution we are seeking if π_0 can be determined; this is done next.

Step 4. Add all π_j in (7.109) and use the normalization condition (7.92) to obtain π_0:

$$\pi_0 + \pi_0 \sum_{j=1}^{\infty} \left(\frac{\lambda_0 \cdots \lambda_{j-1}}{\mu_1 \cdots \mu_j}\right) = 1$$

which gives

$$\pi_0 = \frac{1}{1 + \sum_{j=1}^{\infty}\left(\frac{\lambda_0 \cdots \lambda_{j-1}}{\mu_1 \cdots \mu_j}\right)} \tag{7.110}$$

This completes the derivation of the steady-state solution, which is provided by the combination of (7.109) and (7.110). These two equations are essential in the study of simple queueing systems and we will make extensive use of them in the next chapter.

The existence of this solution is closely related to the behavior of the sum

$$S_1 = 1 + \sum_{j=1}^{\infty}\left(\frac{\lambda_0 \cdots \lambda_{j-1}}{\mu_1 \cdots \mu_j}\right) \tag{7.111}$$

appearing in the denominator of the expression for π_0 in (7.110). If $S_1 = \infty$, then $\pi_0 = 0$. We can immediately see that in this case the chain cannot be positive recurrent, and Theorem 7.11 does not apply. It can be formally shown that if $S_1 = \infty$ then the chain is either null recurrent or transient. Intuitively, we can see in (7.111) that S_1 tends to blow up when birth rates are larger than death rates, that is, the chain drifts towards ∞ and states near 0 are never revisited (or take infinite expected time to be revisited). Conversely, when death rates are larger than birth rates, then the chain drifts back towards state 0, and the steady-state value π_0 is positive.

Another critical sum, which determines whether the chain is transient or not, is the following:

$$S_2 = \sum_{j=1}^{\infty}\left(\frac{\mu_1 \cdots \mu_j}{\lambda_1 \cdots \lambda_j}\right)\frac{1}{\lambda_j} \tag{7.112}$$

This is a good point to look back at Example 7.10, where we considered a simple discrete-time version of the birth–death chain. In that case, we saw in (7.39) that if a sum of the form of S_2 above is less than ∞ then that chain is transient. Carrying out an analysis similar to that of Example 7.10, it can be shown that $S_2 < \infty$ is a necessary and sufficient condition for our birth–death chain to be transient. Intuitively, we see that S_2 in (7.112) remains finite when birth rates tend to be larger than death rates, and therefore the chain drifts toward ∞.

Most practical DES we care to design and analyze are positive recurrent in nature. Thus, the case of interest for our purposes corresponds to birth and death rates so selected that $S_1 < \infty$, $S_2 = \infty$. It can be seen that this situation arises whenever we have

$$\frac{\lambda_j}{\mu_j} < 1 \tag{7.113}$$

either for all $j = 0, 1, \ldots$, or for all $j > j'$ for some finite positive integer j'. Thus, we generally require that death rates eventually overtake birth rates to reach some form of "stability." In queueing systems, births correspond to customer arrivals requesting service, and deaths correspond to service completions. It naturally follows that service completion rates must exceed arrival rates if our system is to handle customer demand; otherwise, queue lengths tend to grow without bound. Moreover, if the empty state $j = 0$ is never revisited, this implies that the system server is operating above capacity. Therefore, condition (7.113) represents a fundamental "rule of thumb" to remember in designing DES modeled as birth–death chains. We shall have much more to say on this issue when we discuss queueing systems in the next chapter.

In summary, a birth–death chain may be thought of as a stochastic DES with state space $\mathcal{X} = \{0, 1, \ldots\}$ and event set $\mathcal{E} = \{a, d\}$, where a is a birth (arrival) and d is a death (departure). In general, birth and death rates are state-dependent, so the underlying clock uses variable speeds as described in Sect. 6.9 of Chap. 6. The interevent times of this DES are always exponentially distributed with parameter $(\lambda_j + \mu_j)$ when the state is $j = 1, 2, \ldots$ and λ_0 when $j = 0$. The framework used in this chapter serves to provide solutions for the state probabilities $\pi_j(t)$. However, as pointed out in Sect. 7.2.4, the compact transition matrix representation we have used does hide some of the underlying dynamic behavior of an event-driven system. A case in point arises when one compares the state transition diagram in Fig. 7.17 with the state transition rate diagram in Fig. 7.19 for the same DES. The fact that several a events may occur while in state 2 is hidden from the representation of Fig. 7.19. While this has no effect on our effort to determine the state probability vector $\pi(t)$, it does limit us in studying other useful properties of the DES. This is something that one should keep in mind and will be further discussed in later chapters.

7.5 UNIFORMIZATION OF MARKOV CHAINS

Continuous-time Markov chain models allow for state transitions (events) to occur at any (real-valued) time instant, and are therefore viewed as more "realistic" for a large number of applications. On the other hand, there are also applications where discrete-time models are more convenient (e.g., the gambling problem in Example 7.4). Discrete-time models are also often easier to set up and simpler to analyze. It is natural, therefore, to explore how to convert a continuous-time model into a discrete-time model without distorting the information that the former contains. As we will see in Chap. 9, this conversion process is particularly useful in the study of "controlled" Markov chains.

Returning to the transition rate matrix $\mathbf{Q}$, recall that a diagonal element $-q_{ii}$ represents the total probability flow rate out of state i. As in (7.78), let us set $\Lambda(i) = -q_{ii}$. There is of course no reason to expect that $\Lambda(i)$ is the same for all i. Suppose, however, that we pick a *uniform* rate γ such that

$$\gamma \geq \Lambda(i) \qquad \text{for all states } i$$

and replace all $\Lambda(i)$ by γ. We can then argue that the "extra probability flow" $[\gamma - \Lambda(i)]$ at state i corresponds to "fictitious events" occurring at this rate which simply leave state i unchanged. These self-loop transitions should have no effect on the behavior of the chain: The state itself has not been affected, and, by the memoryless property, the time left until the next actual transition out of this state also remains unaffected, that is, it is still exponentially distributed with parameter $\Lambda(i)$. As a result, the new "uniformized" chain should be stochastically indistinguishable from the original one. That is, given a common initial state x_0, the two models should have identical state probabilities $\pi_i(t)$ for all t and states i.

To visualize the uniformization process, consider a state i which can have state transitions to two other states only, j and k, as illustrated in Fig. 7.22. The corresponding transition rates are q_{ij} and q_{ik}, and let $\Lambda(i) = q_{ij} + q_{ik}$. This is the actual total probability flow rate out of state i. We also know that the corresponding transition probabilities are given by

$$P_{ij} = q_{ij}/\Lambda(i) \qquad \text{and} \qquad P_{ik} = q_{ik}/\Lambda(i)$$

Now, let us define a *uniform* rate γ such that

$$\gamma = q_{ij} + q_{ik} + \nu$$

where $\nu \geq 0$ is an arbitrary rate for the fictitious events we wish to introduce. The uniformized chain is a discrete-time Markov chain where the transition probabilities for state i are given by

$$P_{ij}^U = \left(\frac{\Lambda(i)}{\gamma}\right) P_{ij} = \frac{q_{ij}}{\gamma}, \qquad P_{ik}^U = \left(\frac{\Lambda(i)}{\gamma}\right) P_{ik} = \frac{q_{ik}}{\gamma},$$

$$P_{ii}^U = 1 - \left(\frac{\Lambda(i)}{\gamma}\right) = \frac{\nu}{\gamma}$$

where P_{ii}^U is the self-loop probability in the new chain (see Fig. 7.22). Observe that in the expressions above the uniformized transition probabilities are scaled versions of the original ones with the common scaling factor $\Lambda(i)/\gamma$.

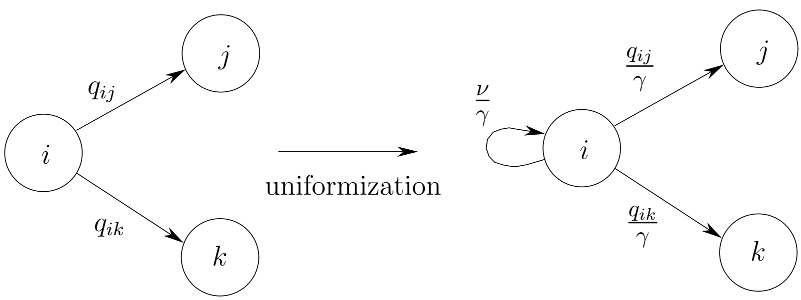


Figure 7.22: Illustrating the uniformization process.
The actual total transition rate out of state i is $\Lambda(i) = q_{ij} + q_{ik}$. We choose a uniform rate $\gamma = q_{ij} + q_{ik} + \nu$, for some $\nu \geq 0$. In the uniformized chain, we allow a self-loop with probability ν/γ, corresponding to fictitious events leaving the state unaffected. The uniformized transition probabilities are the original transition rates scaled by $1/\gamma$.

In summary, given a continuous-time Markov chain with state space $\mathcal{X}$ and transition rate matrix $\mathbf{Q}$, we can construct a stochastically equivalent uniformized discrete-time Markov chain by selecting a uniform transition rate

$$\gamma \geq \max_{i \in \mathcal{X}} \{-q_{ii}\} \tag{7.114}$$

and transition probabilities

$$P_{ij}^U = \begin{cases} \dfrac{q_{ij}}{\gamma} & \text{if } i \neq j \\ 1 + \dfrac{q_{ii}}{\gamma} & \text{if } i = j \end{cases} \tag{7.115}$$

Alternatively, if the Markov chain is specified through transition probabilities $P_{ij}, i \neq j$, and state holding time parameters $\Lambda(i)$, then the uniformization process becomes

$$\gamma \geq \max_{i \in \mathcal{X}} \{\Lambda(i)\} \tag{7.116}$$

$$P_{ij}^U = \begin{cases} \dfrac{\Lambda(i)}{\gamma} P_{ij} & \text{if } i \neq j \\ 1 - \dfrac{\Lambda(i)}{\gamma} & \text{if } i = j \end{cases} \tag{7.117}$$

Although we will not explicitly prove that the uniformized chain and the original continuous-time chain are indeed stochastically equivalent, the following example provides a verification of this fact for the steady-state probabilities of homogeneous birth–death chains.

Example 7.21 (Uniformization of a birth–death chain)
Let us consider a homogeneous birth–death chain, as shown in Fig. 7.23. Here, the birth and death rate parameters are fixed at λ and μ respectively. We can now make use of the steady-state solution in (7.109) and (7.110). The expression for π_0 becomes

$$\pi_0 = \frac{1}{1 + \sum_{j=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^j}$$

The sum in the denominator is a simple geometric series that converges as long as $\lambda/\mu < 1$. Under this assumption, we get

$$\sum_{j=1}^{\infty} \left(\frac{\lambda}{\mu}\right)^j = \frac{\lambda/\mu}{1 - \lambda/\mu}$$

and, therefore,

$$\pi_0 = 1 - \frac{\lambda}{\mu}$$

Then, from (7.109),

$$\pi_j = (1 - \lambda/\mu)(\lambda/\mu)^j, \quad j = 0, 1, \ldots \tag{7.118}$$

Now let us consider a uniformized version of this model. Looking at Fig. 7.23, observe that the total probability flow rate out of every state $i > 0$ is $\Lambda(i) = -q_{ii} = \lambda + \mu$. For state 0, we have $\Lambda(0) = \lambda$. Let us, therefore, select a uniform rate equal to the maximum of all $\Lambda(i)$:

$$\gamma = \lambda + \mu$$

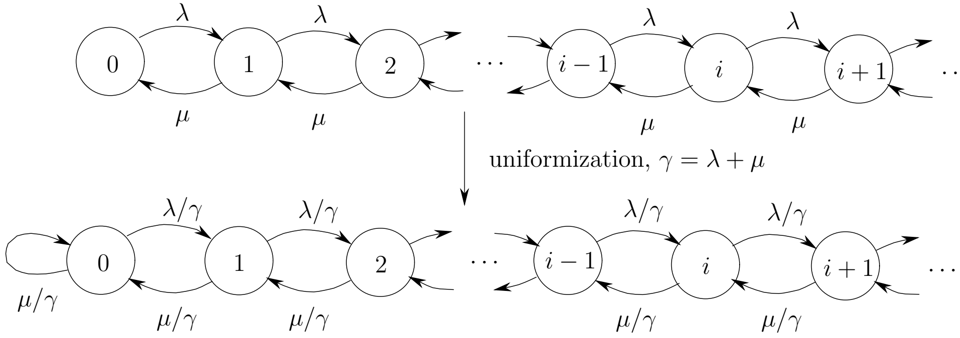


Figure 7.23: Uniformization of birth–death chain of Example 7.21.

Using (7.115), we determine the transition probabilities for the uniformized chain. For all states $i > 0$, we have:

$$P^U_{i,i+1} = \lambda/\gamma, \quad P^U_{i,i-1} = \mu/\gamma, \quad P^U_{i,i} = 0$$

and for state 0:

$$P^U_{01} = \lambda/\gamma, \quad P^U_{00} = \mu/\gamma$$

as shown in Fig. 7.23. We now recognize the discrete-time birth–death chain of Fig. 7.9 that was analyzed in Example 7.13 The steady-state probabilities were obtained in (7.48), and will be denoted here by π^U_j:

$$\pi^U_j = \frac{2p-1}{p}\left(\frac{1-p}{p}\right)^j, \qquad j = 0, 1, 2, \ldots$$

where p is the probability of a death. In our case, $p = \mu/\gamma$. Thus, substituting for $p = \mu/\gamma$ above, we get

$$\pi^U_j = \frac{2\mu - \gamma}{\mu}\left(\frac{\gamma-\mu}{\mu}\right)^j$$

and since $\gamma = \lambda + \mu$, this becomes

$$p^U_j = (1 - \lambda/\mu)(\lambda/\mu)^j, \quad j = 0, 1, \ldots$$

which is precisely what we found in (7.118) to be the steady-state probability of state j in the original chain. We have therefore shown that the steady-state probabilities of the original chain and its uniformized version are indeed identical.

Example 7.22

As a further illustration of the uniformization process, consider the Markov chain of Example 7.19 shown once again in Fig. 7.24. The state transition diagram of the uniformized chain can be obtained by inspection, and is also shown in Fig. 7.24. In this case, a convenient uniform rate is

$$\gamma = \lambda_1 + \lambda_2 + \mu_1 + \mu_2$$

Then, all transition probabilities in the uniform version are obtained by dividing the original transition rates by γ, and by introducing self-loops as indicated in the diagram.

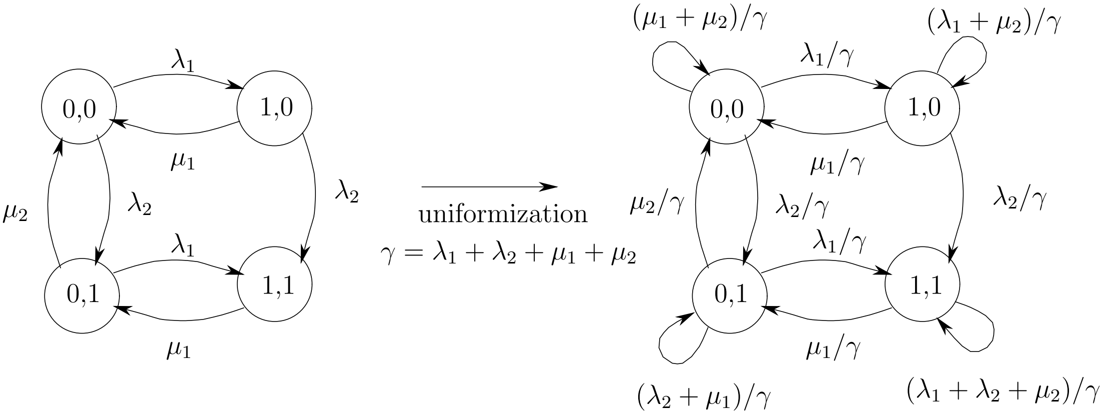


Figure 7.24: Uniformization of chain of Example 7.22.

SUMMARY

- The *n-step transition probabilities* of a discrete-time Markov chain, denoted by $p_{ij}(k, k+n) = P[X_{k+n} = j \mid X_k = i]$, satisfy the *Chapman-Kolmogorov equation*

$$p_{ij}(k, k+n) = \sum_{\text{all } r} p_{ir}(k, u) p_{rj}(u, k+n), \qquad k \leq u \leq k+n$$

- We obtain a *homogeneous* Markov chain whenever the transition probability $p_{ij}(k)$ is independent of k for all i, j. Most of the results in this chapter apply to this case only.

- A consequence of the memoryless property is the fact that state holding times in discrete-time Markov chains are geometrically distributed with parameter p_{ii}.

- A Markov chain is *irreducible* if any state is reachable from any other state by some sequence of transitions. Otherwise, the chain is *reducible.* If the chain visits a state from which it can never again leave, that state is called absorbing.

- States can be classified as *transient, positive recurrent*, or *null recurrent.* A transient state is revisited with probability less than 1, whereas a recurrent state is revisited with probability 1. The expected recurrence time for a positive recurrent state is finite, whereas for a null recurrent state it is infinite.

- States can also be classified as *periodic* or *aperiodic.* If visits to some state are constrained to occur only in a number of steps which is a multiple of an integer greater than or equal to 2, that state is said to be periodic.

- In an irreducible aperiodic Markov chain consisting of positive recurrent states, a unique stationary state probability vector $\boldsymbol{\pi}$ exists, which is obtained by solving $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ subject to the normalization condition $\sum_j \pi_j = 1$.

- For continuous-time Markov chains, the *Chapman-Kolmogorov equation* applies to *transition functions* $p_{ij}(s,t) = P[X(t) = j \mid X(s) = i]$, $s \geq t$:

$$p_{ij}(s, t) = \sum_{\text{all } r} p_{ir}(s, u) p_{rj}(u, t), \qquad s \leq u \leq t$$

In the homogeneous case, $p_{ij}(s,t)$ depends only on the difference $(t-s)$ for all i, j.

- State holding times in homogeneous continuous-time Markov chains are exponentially distributed.

- A continuous-time Markov chain may be specified either by its transition rate matrix $\mathbf{Q}$, or by defining transition probabilities P_{ij} for all pairs of states $i \neq j$ along with parameters $\Lambda(i)$ for all states i, characterizing the exponential state holding time distributions. Denoting by q_{ij} the (i,j) entry of $\mathbf{Q}$, we have $P_{ij} = -q_{ij}/q_{ii}$, and $\Lambda(i) = -q_{ii}$.

- In an irreducible continuous-time Markov chain consisting of positive recurrent states, a unique stationary state probability vector $\boldsymbol{\pi}$ exists, which is obtained by solving $\boldsymbol{\pi}\mathbf{Q} = 0$ subject to the normalization condition $\sum_j \pi_j = 1$.

- A *birth–death chain* is a continuous-time Markov chain whose transition rate matrix $\mathbf{Q}$ satisfies $q_{ij} = 0$ for all $j > i+1$ and $j < i-1$. We define the birth rate to be $\lambda_j = q_{j,j+1} > 0$, and the death rate $\mu_j = q_{j,j-1} > 0$. A homogeneous pure birth chain (i.e., when $\mu_j = 0$ and $\lambda_j = \lambda =$ constant) reduces to a Poisson process. Under certain conditions on its parameters, the stationary probabilities of a birth–death chain can be obtained in the closed form (7.109) and (7.110).

- A continuous-time Markov chain can be converted into an equivalent discrete-time chain through the *uniformization* process: A uniform transition rate γ is selected such that $\gamma \geq -q_{ii}$ for all i, and the uniformized transition probabilities are given by q_{ij}/γ for $i \neq j$, $(1 + q_{ii}/\gamma)$ for $i = j$.

PROBLEMS

7.1 The breeding habits of every family in a certain species are as follows: The family has a birth at most once a year; if at the end of a year the family has no children, there is a 50% chance of a birth in the next year; if it has one child, there is a 40% chance of a birth in the next year; if it has two children, there is a 10% chance of a birth in the next year; if it has three or more children, there is never a birth in the next year; when there is a birth, there is a 2% chance of twins, but there are never any triplets, quadruplets, etc. in this species.

Let X_k denote the number of children in such a family at the end of the kth year. Assume that $X_0 = 0$ and that no children ever die for the range of years we are interested in.

(a) What is the state space of this chain?

(b) Determine the transition probabilities $P[X_2 = j \mid X_1 = i]$ for all possible values of i, j. Is this a Markov chain? If so, is it a homogeneous Markov chain?

(c) Compute the probability that the family has one child at the end of the second year.

7.2 An electronic device driven by a clock randomly generates one of three numbers, 0, 1, or 2, with every clock tick according to the following rules: If 0 was generated last, then the next number is 0 again with probability 0.5 or 1 with probability 0.5;

if 1 was generated last, then the next number is 1 again with probability 0.4 or 2 with probability 0.6; if 2 was generated last, then the next number is either 0 with probability 0.7 or 1 with probability 0.3. Moreover, before the clock starts ticking, a number is generated from the distribution: 0 with probability 0.3, 1 with probability 0.3, 2 with probability 0.4. The device is attached to a bomb set to explode the first time the sequence $\{1, 2, 0\}$ takes place. There is a display on the device showing the most recent number generated.

(a) Draw a state transition diagram for this chain showing all transition probabilities.

(b) Compute the probability that the bomb explodes after exactly two clock ticks.

(c) Compute the probability that the bomb explodes after exactly three clock ticks.

(d) Suppose you are handed the bomb with the number 1 showing. What is the probability that the bomb explodes after exactly two clock ticks? What is the probability that it explodes after exactly three clock ticks?

7.3 The following game is played with a regular 52-card deck. A card is drawn with a player trying to guess its color (red or black). If the player guesses right, he gets 1 dollar, otherwise he gets nothing. After a card is drawn it is set aside (it is not placed back in the deck). Suppose the player decides to adopt the simple policy "guess red all the time," and let X_k denote his total earnings after k cards are drawn, with $X_0 = 0$.

(a) Is $\{X_k\}$ a Markov chain? If so, is it a homogeneous Markov chain?

(b) Compute $P[X_7 = 6 \mid X_6 = 5]$.

(c) Compute $P[X_9 = 4 \mid X_8 = 2]$.

(d) Derive an expression for $P[X_{k+1} = j \mid X_k = i]$ for all $i, j = 0, 1, 2, \ldots$, and $k = 0, 1, \ldots, 51$.

7.4 Repeat Problem 7.3 with the following modification in the game: After a card is drawn from the deck, it is replaced by a card from another deck; this card is chosen to be of the opposite color.

7.5 The following is used to model the genetic evolution of various types of populations. The term "cell" is used to denote a unit in that population. Let X_k be the total number of cells in the kth generation, $k = 1, 2, \ldots$ The ith cell in a particular generation acts independently of all other cells and it either dies or it splits up into a random number of offspring. Let Y_i be an integer-valued non-negative random variable, so that $Y_i = 0$ indicates that the ith cell dies without reproducing, and $Y_i > 0$ indicates the resulting number of cells after reproduction. If $X_k = n$ and $Y_i = 0$ for all cells $i = 1, \ldots, n$ in the kth generation, we get $X_{k+1} = 0$, and we say that the population becomes extinct.

(a) Is $\{X_k\}$ a Markov chain? If so, is it a homogeneous Markov chain?

(b) Suppose $X_0 = 1$ and the random variable Y_i can only take three values, 0, 1, or 2, each with probability 1/3. Draw a state transition diagram for the chain and calculate the transition probabilities $P[X_{k+1} = j \mid X_k = i]$ for $i = 0, 1, 2, 3$.

(c) Compute the probability that the population becomes extinct after four generations.

(d) Compute the probability that the population remains in existence for more than three generations.

7.6 A manufacturing workstation is viewed as a single-server queueing system with a buffer capacity of three parts (including one in process). In a slotted time model (as in Example 7.2), with probability 0.3 a new part arrives at the workstation in any one slot, and with probability 0.5 a part in process is done (at most one part can arrive or complete processing in a time slot). A part arriving and finding no buffering space is simply rejected and lost. A part completing processing is either removed from the workstation with probability 0.5 or it immediately goes through an inspection process that works as follows: With probability 0.5 a part is good and is removed; with probability 0.3 a part is slightly defective and it is returned to the workstation buffer for rework; and with probability 0.2 a part is hopelessly defective and it is simply scrapped. Let X_k denote the number of parts in the workstation at time $k = 1, 2, \ldots,$ and assume that $X_0 = 0$.

(a) Draw a state transition diagram for this chain and compute all transition probabilities involved.

(b) Calculate the probability that there is only one part in the workstation at the second slot.

7.7 Consider the following transition probability matrix:

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/8 & 0 & 7/8 & 0 \\ 0 & 0 & 0 & 4/5 & 1/5 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 2/3 \end{bmatrix}$$

(a) Is this an irreducible or reducible Markov chain? Is it periodic or aperiodic?

(b) Determine which states are transient and which states are recurrent.

(c) Identify all closed irreducible sets of the chain.

7.8 The weather in some area is classified as either "sunny," "cloudy," or "rainy" on any given day. Let X_k denote the weather state on the kth day, $k = 1, 2, \ldots,$ and map these three states into the numbers 0, 1, 2 respectively. Suppose transition probabilities from one day to the next are summarized through the matrix:

$$\mathbf{P} = \begin{bmatrix} 0.4 & 0.4 & 0.2 \\ 0.5 & 0.3 & 0.2 \\ 0.1 & 0.5 & 0.4 \end{bmatrix}$$

(a) Draw a state transition diagram for this chain.

(b) Assuming the weather is cloudy today, predict the weather for the next two days.

(c) Find the stationary state probabilities (if they exist) for this chain. If they do not exist, explain why this is so.

(d) If today is a sunny day, find the average number of days we have to wait at steady state until the next sunny day.

7.9 A total of N gas molecules are split between two containers so that at any time step $k = 0, 1, \ldots$ there are X_k molecules in the first container and $(N - X_k)$ molecules in the second one. At each time step, one of the N molecules is randomly selected, independently from all previous selections, and moved from its container to the other container. This type of model is known as the *Ehrenfest chain*.

(a) Determine the transition probability $P[X_{k+1} = j \mid X_k = i]$ for all $i, j = 0, 1, \ldots, N$.

(b) Draw a state transition diagram for this chain when $N = 3$.

(c) Find the stationary state probabilities (if they exist) for this chain. If they do not exist, explain why this is so.

7.10 Repeat Problem 7.9 with the following modification in the molecule movement: After a molecule is selected, it either remains in its container with probability p or it is moved to the other container with probability $1 - p$. Do part (c) by assuming that $p = 1/3$.

7.11 A gambler plays the following game: He tosses two fair dice and if a total of 5 spots show he wins 1 dollar, otherwise he loses 1 dollar. He has decided that he will quit as soon as he has either won 5 dollars or lost 2 dollars. Calculate the probability that he has won 5 dollars when he quits.

7.12 Recall the definition of the n-step transition probability $p_{ij}^n = P[X_{k+n} = j \mid X_k = i]$ for a homogeneous Markov chain in Sect. 7.2.3, and the definition of the hitting time $T_{ij} = \min\{k > 0 : X_0 = i, X_k = j\}$ in Sect. 7.2.8.

(a) Prove that

$$p_{ij}^n = \sum_{m=1}^{n} P[T_{ij} = m \mid X_0 = i] p_{jj}^{n-m}$$

(b) If j is an absorbing state, then show that

$$p_{ij}^n = P[T_{ij} \le n \mid X_0 = i]$$

7.13 Consider a Markov chain $\{X_k\}, k = 0, 1, \ldots$ with the property

$$\sum_{\text{all } j} j \cdot p_{ij} = i$$

for all states i, and assume that the state space is $\{0, 1, \ldots, N\}$ for some finite integer N. This is an example of a *martingale*, a type of chain that was briefly discussed following Example 7.4.

(a) Show that $E[X_k] = E[X_{k+1}]$ for all $k = 0, 1, \ldots$

(b) Show that states 0 and N are absorbing.

(c) Assume that $X_0 = i$, for some $i \in \{1, 2, \ldots, N - 1\}$. Show that

$$\lim_{k\to\infty} E[X_k] = NP[T_{iN} \le \infty]$$

(*Note*: You may find the results of Problem 7.12 useful.)

(d) Show that the absorption probability $P[T_{i0} < \infty]$ is given by $(1 - i/N)$.

7.14 A random number of messages A_k is placed in a transmitter at times $k = 0, 1, \ldots$ Each message present at time k acts independently of past arrivals and of other messages, and it is successfully transmitted with probability p. Thus, if X_k denotes the number of messages present at k, and $D_k \leq X_k$ denotes the number of successful transmissions at k, we can write

$$X_{k+1} = X_k - D_k + A_{k+1}$$

Assume that A_k has a Poisson distribution $P[A_k = n] = \lambda^n e^{-\lambda}/n!, n = 0, 1, \ldots,$ independent of k.

(a) If the system is at state $X_k = i$, find $P[D_k = m \mid X_k = i], m = 0, 1, \ldots$

(b) Determine the transition probabilities $P[X_{k+1} = j \mid X_k = i], i, j = 0, 1, \ldots$

(c) If $X_0 = 0$, show that X_1 has a Poisson distribution with parameter λ, and that $(X_1 - D_1)$ has a Poisson distribution with parameter $\lambda(1-p)$. Hence, determine the distribution of X_2.

(d) If X_0 has a Poisson distribution with parameter ν, find the value of ν such that there exists a stationary probability distribution for the Markov chain $\{X_k\}$.

7.15 A DES is specified as a stochastic timed automaton $(\mathcal{E}, \mathcal{X}, \Gamma, f, p_0, G)$ with

$$\begin{array}{lll}
\mathcal{E} = \{\alpha, \beta, \gamma, \delta\}, & \mathcal{X} = \{0, 1, 2, 3, 4\} & \\
\Gamma(0) = \{\alpha, \beta\}, & f(0, \alpha) = 1, & f(0, \beta) = 2 \\
\Gamma(1) = \{\beta, \gamma\}, & f(1, \beta) = 2, & f(1, \gamma) = 0 \\
\Gamma(2) = \{\beta, \gamma, \delta\}, & f(2, \beta) = 3, & f(2, \gamma) = 1, \\
 & f(2, d) = 3 & \\
\Gamma(3) = \{\alpha, \beta, \gamma, \delta\}, & f(3, \alpha) = 1, & f(3, \beta) = 4, \\
 & f(3, \gamma) = 2, & f(3, \delta) = 3 \\
\Gamma(4) = \{\gamma, \delta\}, & f(4, \gamma) = 3, & f(4, \delta) = 0
\end{array}$$

The system is initially at state 0, and the event processes for $\alpha, \beta, \gamma, \delta$ are all Poisson with parameters $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ respectively.

(a) Determine the transition probabilities P_{01}, P_{23}, P_{31}, and P_{42}.

(b) What is the average amount of time spent at state 1? At state 3?

(c) Write down the transition matrix $\mathbf{Q}$ for this Markov chain.

7.16 Consider a Markov chain $\{X(t)\}$ with state space $\mathcal{X} = \{0, 1, 2, 3, 4\}$ and transition rate matrix

$$\mathbf{Q} = \begin{bmatrix}
-\lambda & \lambda & 0 & 0 & 0 \\
0 & -(\lambda + \mu_1) & \lambda & \mu_1 & 0 \\
0 & 0 & -\mu_1 & 0 & \mu_1 \\
\mu_2 & 0 & 0 & -(\lambda + \mu_2) & \lambda \\
0 & \mu_2 & 0 & 0 & -\mu_2
\end{bmatrix}$$

(a) Draw a state transition rate diagram.

(b) Determine the stationary state probabilities (if they exist) for $\lambda = 1, \mu_1 = 3/2, \mu_2 = 7/4$.

(c) If steady state is reached and we know that the state is not 0, what is its expected value?

7.17 One third of all transactions submitted to the local branch of a bank must be handled by the bank's central computer, while all the rest can be handled by the branch's own computer. Transactions are generated according to a Poisson process with rate l. The central and the branch's computer act independently, they can both buffer an infinite number of transactions, process one transaction at a time, and they are both assumed to require transaction processing times which are exponentially distributed with rates μ_1 and μ_2 respectively. The central computer, however, is also responsible for processing other transactions, which we will assume to be generated by another independent Poisson process with rate ν. Let $\lambda = 10$ transactions per minute, $\mu_1 = 50$ transactions per minute, $\mu_2 = 15$ transactions per minute, and $\nu = 30$ transactions per minute. Finally, let $X_1(t), X_2(t)$ denote the number of transactions residing in the central and branch computer respectively (see also Problem 6.7).

(a) Do stationary probabilities exist for this process? Why or why not?
If stationary probabilities do exist, then at steady state:

(b) Calculate the probability that the branch computer is idle.

(c) Calculate the probability that there are more that three transactions *waiting* to be processed by the central computer.

(d) Calculate the probability that both computers have a single transaction in process.

7.18 Consider a pure death chain $\{X(t)\}$ such that in (7.98) and (7.99) we have $\lambda_j = 0$ for all $j = 0, 1, \ldots$ and $\mu_j = j\mu, j = 1, 2, \ldots, N$, where N is a given initial state. Show that $\pi_j(t) = P[X(t) = j], j = 1, 2, \ldots, N$, has a binomial distribution with parameters N and $e^{-\mu t}$.

7.19 Use the uniformization process to derive a discrete-time Markov chain model for $\{X(t)\}$ in Problem 7.16, and obtain the stationary probabilities (if they exist). Compare the result to part (b) of Problem 7.16.

SELECTED REFERENCES

- Asmussen, S., *Applied Probability and Queues*, Wiley, New York, 1987.
- Bertsekas, D.P., *Dynamic Programming: Deterministic and Stochastic Models*, Prentice-Hall, Englewood Cliffs, NJ, 1987.
- Bertsekas, D.P., *Dynamic Programming and Optimal Control, Vols. 1 and 2*, Athena Scientific, Belmont, MA, 1995.
- Cinlar, E., *Introduction to Stochastic Processes*, Prentice-Hall, Englewood Cliffs, NJ, 1975.
- Clarke, A.B., and R.L. Disney, *Probability and Random Processes*, Wiley, New York, 1985.
- Gallager, R.G., *Discrete Stochastic Processes*, Kluwer Academic Publishers, Boston, 1996.
- Hoel, P.G., S.C. Port, and C.J. Stone, *Introduction to Stochastic Processes*, Houghton Mifflin, Boston, 1972.

- Kleinrock, L., *Queueing Systems, Volume I: Theory*, Wiley, New York, 1975.
- Parzen, E., *Stochastic Processes*, Holden-Day, San Francisco, 1962.
- Trivedi, K.S., *Probability and Statistics with Reliability, Queuing and Computer Science Applications, Second Edition*, Wiley, New York, 2001.