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.

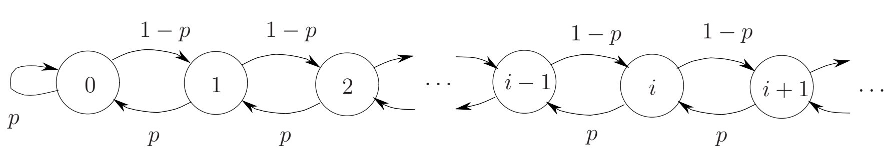

Figure 7.9: State transition diagram for Example 7.10.

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

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

and it follows that $T_{1m} \geq m - 1$ 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, $\beta \geq 1$, that is, $p \geq 1/2$, 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 $p \geq 1/2, q_1 = 1$, 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.

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:

1. 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.
2. 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 $0 < \pi_j < 1$.
3. 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

$$\pi_0 = (1 - 0.5\beta)\pi_0 + 0.5\pi_1$$
$$\pi_1 = 0.5\beta\pi_0 + 0.5\pi_1$$
$$\pi_0 + \pi_1 = 1$$

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.