However, simple calculus shows that $1/\mu s$ is the Laplace transform of the function $h(t) = t/\mu$, and thus by the uniqueness of transforms, we obtain

$$m_D(t) = t/\mu.$$

(ii) For a delayed renewal process, upon conditioning on $S_{N(t)}$ we obtain, using (3.5.3),

$$P\{Y_D(t) > x\} = P\{Y_D(t) > x | S_{N(t)} = 0\}\overline{G}(t) + \int_0^t P\{Y_D(t) > x | S_{N(t)} = s\}\overline{F}(t-s)\, dm_D(s).$$

Now

$$P\{Y_D(t) > x | S_{N(t)} = 0\} = P\{X_1 > t + x | X_1 > t\} = \frac{\overline{G}(t+x)}{\overline{G}(t)}$$

$$P\{Y_D(t) > x | S_{N(t)} = s\} = P\{X > t + x - s | X > t - s\} = \frac{\overline{F}(t+x-s)}{\overline{F}(t-s)}.$$

Hence,

$$P\{Y_D(t) > x\} = \overline{G}(t+x) + \int_0^t \overline{F}(t+x-s)\, dm_D(s).$$

Now, letting $G = F_e$ and using part (i) yields

$$\begin{aligned} P\{Y_D(t) > x\} &= \overline{F}_e(t+x) + \int_0^t \overline{F}(t+x-s)\, ds/\mu \\ &= \overline{F}_e(t+x) + \int_x^{t+x} \overline{F}(y)\, dy/\mu \\ &= \overline{F}_e(x). \end{aligned}$$

(iii) To prove (iii) we note that $N_D(t+s) - N_D(s)$ may be interpreted as the number of renewals in time t of a delayed renewal process, where the initial distribution is the distribution of $Y_D(s)$. The result then follows from (ii).

3.6 Renewal Reward Processes

A large number of probability models are special cases of the following model. Consider a renewal process $\{N(t), t \geq 0\}$ having interarrival times $X_n, n \geq 1$ with distribution F, and suppose that each time a renewal occurs we receive a reward. We denote by R_n the reward earned at the time of the nth renewal. We shall assume that the $R_n, n \geq 1$, are independent and identically distributed.

However, we do allow for the possibility that R_n may (and usually will) depend on X_n, the length of the nth renewal interval, and so we assume that the pairs (X_n, R_n), $n \geq 1$, are independent and identically distributed. If we let

$$R(t) = \sum_{n=1}^{N(t)} R_n,$$

then $R(t)$ represents the total reward earned by time t. Let

$$E[R] = E[R_n], \qquad E[X] = E[X_n].$$

THEOREM 3.6.1

If $E[R] < \infty$ and $E[X] < \infty$, then

(i) with probability 1,

$$\frac{R(t)}{t} \to \frac{E[R]}{E[X]} \qquad \text{as } t \to \infty,$$

(ii)

$$\frac{E[R(t)]}{t} \to \frac{E[R]}{E[X]} \qquad \text{as } t \to \infty.$$

Proof To prove (i) write

$$\frac{R(t)}{t} = \frac{\sum_{n=1}^{N(t)} R_n}{t} = \left(\frac{\sum_{n=1}^{N(t)} R_n}{N(t)}\right)\left(\frac{N(t)}{t}\right).$$

By the strong law of large numbers we obtain that

$$\frac{\sum_{n=1}^{N(t)} R_n}{N(t)} \to E[R] \qquad \text{as } t \to \infty,$$

and by the strong law for renewal processes

$$\frac{N(t)}{t} \to \frac{1}{E[X]} \qquad \text{as } t \to \infty.$$

Thus (i) is proven.

To prove (ii) we first note that since $N(t) + 1$ is a stopping time for the sequence $X_1, X_2, \ldots$, it is also a stopping time for $R_1, R_2, \ldots$. (Why?) Thus, by Wald's equation,

$$E\left[\sum_{i=1}^{N(t)} R_i\right] = E\left[\sum_{i=1}^{N(t)+1} R_i\right] - E[R_{N(t)+1}]$$
$$= (m(t)+1)E[R] - E[R_{N(t)+1}]$$

and so

$$\frac{E[R(t)]}{t} = \frac{m(t)+1}{t}E[R] - \frac{E[R_{N(t)+1}]}{t},$$

and the result will follow from the elementary renewal theorem if we can show that $E[R_{N(t)+1}]/t \to 0$ as $t \to \infty$. So, towards this end, let $g(t) = E[R_{N(t)+1}]$. Then conditioning on $S_{N(t)}$ yields

$$g(t) = E[R_{N(t)+1} | S_{N(t)} = 0]\bar{F}(t)$$
$$+ \int_0^t E[R_{N(t)+1} | S_{N(t)} = s]\bar{F}(t-s)\, dm(s).$$

However,

$$E[R_{N(t)+1} | S_{N(t)} = 0] = E[R_1 | X_1 > t],$$
$$E[R_{N(t)+1} | S_{N(t)} = s] = E[R_n | X_n > t - s],$$

and so

(3.6.1) $$g(t) = E[R_1 | X_1 > t]\bar{F}(t) + \int_0^t E[R_n | X_n > t-s]\bar{F}(t-s)\, dm(s).$$

Now, let

$$h(t) = E[R_1 | X_1 > t]\bar{F}(t) = \int_t^\infty E[R_1 | X_1 = x]\, dF(x),$$

and note that since

$$E|R_1| = \int_0^\infty E[|R_1| \,|\, X_1 = x]\, dF(x) < \infty,$$

it follows that

$$h(t) \to 0 \text{ as } t \to \infty \quad \text{and} \quad h(t) \le E|R_1| \text{ for all } t,$$

and thus we can choose T so that $|h(t)| < \varepsilon$ whenever $t \ge T$. Hence, from (3.6.1),

$$\frac{|g(t)|}{t} \le \frac{|h(t)|}{t} + \int_0^{t-T} \frac{|h(t-x)|\, dm(x)}{t} + \int_{t-T}^t \frac{|h(t-x)|\, dm(x)}{t}$$
$$\le \frac{\varepsilon}{t} + \frac{\varepsilon m(t-T)}{t} + E|R_1| \frac{m(t) - m(t-T)}{t}$$
$$\to \frac{\varepsilon}{EX} \quad \text{as } t \to \infty$$

by the elementary renewal theorem. Since ε is arbitrary, it follows that $g(t)/t \to 0$, and the result follows.

Remarks If we say that a cycle is completed every time a renewal occurs, then the theorem states that the (expected) long-run average return is just the expected return earned during a cycle, divided by the expected time of a cycle.

In the proof of the theorem it is tempting to say that $E[R_{N(t)+1}] = E[R_1]$ and thus $1/tE[R_{N(t)+1}]$ trivially converges to zero. However, $R_{N(t)+1}$ is related to $X_{N(t)+1}$, and $X_{N(t)+1}$ is the length of the renewal interval containing the point t. Since larger renewal intervals have a greater chance of containing t, it (heuristically) follows that $X_{N(t)+1}$ tends to be larger than an ordinary renewal interval (see Problem 3.3), and thus the distribution of $R_{N(t)+1}$ is not that of R_1.

Also, up to now we have assumed that the reward is earned all at once at the end of the renewal cycle. However, this is not essential, and Theorem 3.6.1 remains true if the reward is earned gradually during the renewal cycle. To see this, let $R(t)$ denote the reward earned by t, and suppose first that all returns are nonnegative. Then

$$\frac{\sum_{n=1}^{N(t)} R_n}{t} \le \frac{R(t)}{t} \le \frac{\sum_{n=1}^{N(t)} R_n}{t} + \frac{R_{N(t)+1}}{t}$$

and (ii) of Theorem 3.6.1 follows since

$$\frac{E[R_{N(t)+1}]}{t} \to 0.$$

Part (i) of Theorem 3.6.1 follows by noting that both $\sum_{n=1}^{N(t)} R_n/t$ and $\sum_{n=1}^{N(t)+1} R_n/t$ converge to $E[R]/E[X]$ by the argument given in the proof. A similar argument holds when the returns are nonpositive, and the general case follows by breaking up the returns into their positive and negative parts and applying the above argument separately to each.

Example 3.6(a) ***Alternating Renewal Process.*** For an alternating renewal process (see Section 3.4.1) suppose that we earn at a rate of one per unit time when the system is on (and thus the reward for a cycle equals the on time of that cycle). Then the total reward earned by t is just the total on time in $[0, t]$, and thus by Theorem 3.6.1, with probability 1,

$$\frac{\text{amount of on time in } [0, t]}{t} \to \frac{E[X]}{E[X] + E[Y]},$$

where X is an on time and Y an off time in a cycle. Thus by Theorem 3.4.4 when the cycle distribution is nonlattice the limiting probability of the system being on is equal to the long-run proportion of time it is on.

Example 3.6(b) ***Average Age and Excess.*** Let $A(t)$ denote the age at t of a renewal process, and suppose we are interested in computing

$$\lim_{t\to\infty} \int_0^t A(s)\,ds/t.$$

To do so assume that we are being paid money at any time at a rate equal to the age of the renewal process at that time. That is, at time s we are being paid at a rate $A(s)$, and so $\int_0^t A(s)\,ds$ represents our total earnings by time t. As everything starts over again when a renewal occurs, it follows that, with probability 1,

$$\frac{\int_0^t A(s)\,ds}{t} \to \frac{E[\text{reward during a renewal cycle}]}{E[\text{time of a renewal cycle}]}.$$

Now since the age of the renewal process a time s into a renewal cycle is just s, we have

$$\text{reward during a renewal cycle} = \int_0^X s\,ds = \frac{X^2}{2},$$

where X is the time of the renewal cycle. Hence, with probability 1,

$$\lim_{t\to\infty} \frac{\int_0^t A(s)\,ds}{t} = \frac{E[X^2]}{2E[X]}.$$

Similarly if $Y(t)$ denotes the excess at t, we can compute the average excess by supposing that we are earning rewards at a rate equal to the excess at that time. Then the average value of the excess will, by Theorem 3.6.1, be given by

$$\begin{aligned}\lim_{t\to\infty} \int_0^t Y(s)\,ds/t &= \frac{E[\text{reward during a renewal cycle}]}{E[X]}\\ &= \frac{E\left[\int_0^X (X-t)\,dt\right]}{E[X]}\\ &= \frac{E[X^2]}{2E[X]}.\end{aligned}$$

Thus the average values of the age and excess are equal. (Why was this to be expected?)

The quantity $X_{N(t)+1} = S_{N(t)+1} - S_{N(t)}$ represents the length of the renewal interval containing the point t. Since it may also be expressed by

$$X_{N(t)+1} = A(t) + Y(t),$$

we see that its average value is given by

$$\lim_{t\to\infty} \int_0^t X_{N(s)+1}\,ds/t = \frac{E[X^2]}{E[X]}.$$

Since

$$\frac{E[X^2]}{E[X]} \geq E[X]$$

(with equality only when $\text{Var}(X) = 0$) we see that the average value of $X_{N(t)+1}$ is greater than $E[X]$. (Why is this not surprising?)

Example 3.6(c) Suppose that travelers arrive at a train depot in accordance with a renewal process having a mean interarrival time μ. Whenever there are N travelers waiting in the depot, a train leaves. If the depot incurs a cost at the rate of nc dollars per unit time whenever there are n travelers waiting and an additional cost of K each time a train is dispatched, what is the average cost per unit time incurred by the depot?

If we say that a cycle is completed whenever a train leaves, then the above is a renewal reward process. The expected length of a cycle is the expected time required for N travelers to arrive, and, since the mean interarrival time is μ, this equals

$$E[\text{length of cycle}] = N\mu.$$

If we let X_n denote the time between the nth and $(n+1)$st arrival in a cycle, then the expected cost of a cycle may be expressed as

$$\begin{aligned}E[\text{cost of a cycle}] &= E[cX_1 + 2cX_2 + \cdots + (N-1)cX_{N-1}] + K\\ &= \frac{c\mu N(N-1)}{2} + K\end{aligned}$$

Hence the average cost incurred is

$$\frac{c(N-1)}{2} + \frac{K}{N\mu}.$$

3.6.1 A Queueing Application

Suppose that customers arrive at a single-server service station in accordance with a nonlattice renewal process. Upon arrival, a customer is immediately served if the server is idle, and he or she waits in line if the server is busy. The service times of customers are assumed to be independent and identically distributed, and are also assumed independent of the arrival stream.

Let $X_1, X_2, \ldots$ denote the interarrival times between customers; and let $Y_1, Y_2, \ldots$ denote the service times of successive customers. We shall assume that

$$E[Y_i] < E[X_i] < \infty. \tag{3.6.2}$$

Suppose that the first customer arrives at time 0 and let $n(t)$ denote the number of customers in the system at time t. Define

$$L = \lim_{t\to\infty} \int_0^t n(s)\, ds/t.$$

To show that L exists and is constant, with probability 1, imagine that a reward is being earned at time s at rate $n(s)$. If we let a cycle correspond to the start of a busy period (that is, a new cycle begins each time an arrival finds the system empty), then it is easy to see that the process restarts itself each cycle. As L represents the long-run average reward, it follows from Theorem 3.6.1 that

$$\begin{aligned} L &= \frac{E[\text{reward during a cycle}]}{E[\text{time of a cycle}]} \\ &= \frac{E\left[\int_0^T n(s)\, ds\right]}{E[T]}. \end{aligned} \tag{3.6.3}$$

Also, let W_i denote the amount of time the ith customer spends in the system and define

$$W = \lim_{n\to\infty} \frac{W_1 + \cdots + W_n}{n}.$$

To argue that W exists with probability 1, imagine that we receive a reward W_i on day i. Since the queueing process begins anew after each cycle, it follows that if we let N denote the number of customers served in a cycle, then W is the average reward per unit time of a renewal process in which the cycle time

is N and the cycle reward is $W_1 + \cdots + W_N$, and, hence,

$$\begin{aligned} W &= \frac{E[\text{reward during a cycle}]}{E[\text{time of a cycle}]} \\ &= \frac{E\left[\sum_{i=1}^N W_i\right]}{E[N]}. \end{aligned} \tag{3.6.4}$$

We should remark that it can be shown (see Proposition 7.1.1 of Chapter 7) that (3.6.2) implies that $E[N] < \infty$.

The following theorem is quite important in queueing theory.

THEOREM 3.6.2

Let $\lambda = 1/E[X_i]$ denote the arrival rate. Then

$$L = \lambda W.$$

Proof We start with the relationship between T, the length of a cycle, and N, the number of customers served in that cycle. If n customers are served in a cycle, then the next cycle begins when the $(n + 1)$st customer arrives; hence,

$$T = \sum_{i=1}^N X_i.$$

Now it is easy to see that N is a stopping time for the sequence $X_1, X_2, \ldots$ since

$$\begin{aligned} N = n \leftrightarrow X_1 + \cdots + X_k &< Y_1 + \cdots + Y_k, \qquad k = 1, \ldots, n-1 \\ \text{and}\quad X_1 + \cdots + X_n &> Y_1 + \cdots + Y_n \end{aligned}$$

and thus $\{N = n\}$ is independent of $X_{n+1}, X_{n+2}, \ldots$. Hence by Wald's equation

$$E[T] = E[N]E[X] = E[N]/\lambda$$

and so by (3.6.3) and (3.6.4)

$$L = \lambda W \frac{E\left[\int_0^T n(s)\, ds\right]}{E\left[\sum_{i=1}^N W_i\right]}. \tag{3.6.5}$$

But by imagining that each customer pays at a rate of 1 per unit time while in the system (and so the total amount paid by the ith arrival is just W_i), we see

$$\int_0^T n(s)\,ds = \sum_{i=1}^{N} W_i = \text{total paid during a cycle,}$$

and so the result follows from (3.6.5).

Remarks

(1) The proof of Theorem 3.6.2 does not depend on the particular queueing model we have assumed. The proof goes through without change for any queueing system that contains times at which the process probabilistically restarts itself and where the mean time between such cycles is finite. For example, if in our model we suppose that there are k available servers, then it can be shown that a sufficient condition for the mean cycle time to be finite is that

$$E[Y_i] < kE[X_i] \quad \text{and} \quad P\{Y_i < X_i\} > 0.$$

(2) Theorem 3.6.2 states that the

$$\text{(time) average number in "the system"} = \lambda \cdot \text{(average time a customer spends in "the system").}$$

By replacing "the system" by "the queue" the same proof shows that the

$$\text{average number in the queue} = \lambda \cdot \text{(average time a customer spends in the queue),}$$

or, by replacing "the system" by "service" we have that the

$$\text{average number in service} = \lambda E[Y].$$

3.7 Regenerative Processes

Consider a stochastic process $\{X(t), t \geq 0\}$ with state space $\{0, 1, 2, \ldots\}$ having the property that there exist time points at which the process (probabilistically) restarts itself. That is, suppose that with probability 1, there exists a time S_1 such that the continuation of the process beyond S_1 is a probabilistic replica of the whole process starting at 0. Note that this property implies the existence of further times $S_2, S_3, \ldots$ having the same property as S_1. Such a stochastic process is known as a *regenerative process.*

From the above, it follows that $\{S_1, S_2, \ldots,\}$ constitute the event times of a renewal process. We say that a cycle is completed every time a renewal occurs. Let $N(t) = \max\{n: S_n \leq t\}$ denote the number of cycles by time t.

The proof of the following important theorem is a further indication of the power of the key renewal theorem.

THEOREM 3.7.1

If F, the distribution of a cycle, has a density over some interval, and if $E[S_1] < \infty$, then

$$P_j \equiv \lim_{t\to\infty} P\{X(t) = j\} = \frac{E[\text{amount of time in state } j \text{ during a cycle}]}{E[\text{time of a cycle}]}.$$

Proof Let $P(t) = P\{X(t) = j\}$. Conditioning on the time of the last cycle before t yields

$$P(t) = P\{X(t) = j \mid S_{N(t)} = 0\}\overline{F}(t) + \int_0^t P\{X(t) = j \mid S_{N(t)} = s\}\overline{F}(t-s)\,dm(s).$$

Now,

$$P\{X(t) = j \mid S_{N(t)} = 0\} = P\{X(t) = j \mid S_1 > t\},$$
$$P\{X(t) = j \mid S_{N(t)} = s\} = P\{X(t-s) = j \mid S_1 > t-s\},$$

and thus

$$P(t) = P\{X(t) = j, S_1 > t\} + \int_0^t P\{X(t-s) = j, S_1 > t-s\}\,dm(s).$$

Hence, as it can be shown that $h(t) \equiv P\{X(t) = j, S_1 > t\}$ is directly Riemann integrable, we have by the key renewal theorem that

$$P(t) \to \int_0^\infty P\{X(t) = j, S_1 > t\}\,dt / E[S_1].$$

Now, letting

$$I(t) = \begin{cases} 1 & \text{if } X(t) = j, S_1 > t \\ 0 & \text{otherwise,} \end{cases}$$

then $\int_0^\infty I(t)\,dt$ represents the amount of time in the first cycle that $X(t) = j$. Since

$$E\left[\int_0^\infty I(t)\,dt\right] = \int_0^\infty E[I(t)]\,dt = \int_0^\infty P\{X(t) = j, S_1 > t\}\,dt,$$

the result follows.

Example 3.7(A) ***Queueing Models with Renewal Arrivals.*** Most queueing processes in which customers arrive in accordance to a renewal process (such as those in Section 3.6) are regenerative processes with cycles beginning each time an arrival finds the system empty. Thus for instance in the single-server queueing model with renewal arrivals, $X(t)$, the number in the system at time t, constitutes a regenerative process provided the initial customer arrives at $t = 0$ (if not then it is a *delayed* regenerative process and Theorem 3.7.1 remains valid).

From the theory of renewal reward processes it follows that P_j also equals the *long-run proportion* of time that $X(t) = j$. In fact, we have the following.

PROPOSITION 3.7.2

For a regenerative process with $E[S_1] < \infty$, with probability 1,

$$\lim_{t\to\infty} \frac{[\text{amount of time in } j \text{ during } (0, t)]}{t} = \frac{E[\text{time in } j \text{ during a cycle}]}{E[\text{time of a cycle}]}.$$

Proof Suppose that a reward is earned at *rate* 1 whenever the process is in state j. This generates a renewal reward process and the proposition follows directly from Theorem 3.6.1.

3.7.1 The Symmetric Random Walk and the Arc Sine Laws

Let $Y_1, Y_2, \ldots$ be independent and identically distributed with

$$P\{Y_i = 1\} = P\{Y_i = -1\} = \tfrac{1}{2},$$

and define

$$Z_0 = 0, \qquad Z_n = \sum_{i=1}^{n} Y_i.$$

The process $\{Z_n, n \geq 0\}$ is called the symmetric random walk process.

If we now define X_n by

$$X_n = \begin{cases} 0 & \text{if } Z_n = 0 \\ 1 & \text{if } Z_n > 0 \\ -1 & \text{if } Z_n < 0, \end{cases}$$

then $\{X_n, n \geq 1\}$ is a regenerative process that regenerates whenever X_n takes value 0. To obtain some of the properties of this regenerative process, we will first study the symmetric random walk $\{Z_n, n \geq 0\}$.

Let

$$u_n = P\{Z_{2n} = 0\} = \binom{2n}{n}\left(\tfrac{1}{2}\right)^{2n}$$

and note that

$$u_n = \frac{2n-1}{2n}u_{n-1}. \tag{3.7.1}$$

Now let us recall from the results of Example 1.5(E) of Chapter 1 (the ballot problem example) the expression for the probability that the first visit to 0 in the symmetric random walk occurs at time $2n$. Namely,

$$P\{Z_1 \neq 0, Z_2 \neq 0, \ldots, Z_{2n-1} \neq 0, Z_{2n} = 0\} = \frac{\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}}{2n-1} = \frac{u_n}{2n-1}. \tag{3.7.2}$$

We will need the following lemma, which states that u_n—the probability that the symmetric random walk is at 0 at time $2n$—is also equal to the probability that the random walk does not hit 0 by time $2n$.

state chosen by the interarrival distribution. Hence it seems that the reversed process is just the *excess* or *residual life* process.

Thus letting P_i denote the probability that an interarrival is i, $i \geq 1$, it seems likely that

$$P^*_{1i} = P_i, \qquad P^*_{i,i-1} = 1, \qquad i > 1.$$

Since

$$P_{i1} = \frac{P_i}{\sum_{j=i}^{\infty} P_j} = 1 - P_{i,i+1}, \qquad i \geq 1,$$

for the reversed chain to be as given above we would need from (4.7.5) that

$$\frac{\pi_i P_i}{\sum_{j=i}^{\infty} P_j} = \pi_1 P_i$$

or

$$\pi_i = \pi_1 P\{X \geq i\},$$

where X is an interarrival time. Since $\sum_i \pi_i = 1$, the above would necessitate that

$$1 = \pi_1 \sum_{i=1} P\{X \geq i\}$$
$$= \pi_1 E[X],$$

and so for the reversed chain to be as conjectured we would need that

$$\pi_i = \frac{P\{X \geq i\}}{E[X]}. \tag{4.7.6}$$

To complete the proof that the reversed process is the excess and the limiting probabilities are given by (4.7.6), we need verify that

$$\pi_i P_{i,i+1} = \pi_{i+1} P^*_{i+1,i},$$

or, equivalently,

$$P\{X \geq i\}\left[1 - \frac{P_i}{P\{X \geq i\}}\right) = P\{X \geq i+1\},$$

which is immediate.

Thus by looking at the reversed chain we are able to show that it is the excess renewal process and obtain at the same time the limiting distribution (of both excess and age). In fact, this example yields additional insight as to why the renewal excess and age have the same limiting distribution.

The technique of using the reversed chain to obtain limiting probabilities will be further exploited in Chapter 5, where we deal with Markov chains in continuous time.

4.8 Semi-Markov Processes

A semi-Markov process is one that changes states in accordance with a Markov chain but takes a random amount of time between changes. More specifically consider a stochastic process with states $0, 1, \ldots$, which is such that, whenever it enters state i, $i \geq 0$:

(i) The next state it will enter is state j with probability P_{ij}, $i, j \geq 0$.
(ii) Given that the next state to be entered is state j, the time until the transition from i to j occurs has distribution F_{ij}.

If we let $Z(t)$ denote the state at time t, then $\{Z(t), t \geq 0\}$ is called a *semi-Markov process.*

Thus a semi-Markov process does not possess the Markovian property that given the present state the future is independent of the past. For in predicting the future not only would we want to know the present state, but also the length of time that has been spent in that state. Of course, at the moment of transition, all we would need to know is the new state (and nothing about the past). A Markov chain is a semi-Markov process in which

$$F_{ij}(t) = \begin{cases} 0 & t < 1 \\ 1 & t \geq 1. \end{cases}$$

That is, all transition times of a Markov chain are identically 1.

Let H_i denote the distribution of time that the semi-Markov process spends in state i before making a transition. That is, by conditioning on the next state, we see

$$H_i(t) = \sum_j P_{ij} F_{ij}(t),$$

and let μ_i denote its mean. That is,

$$\mu_i = \int_0^\infty x\, dH_i(x)$$

If we let X_n denote the nth state visited, then $\{X_n, n \geq 0\}$ is a Markov chain with transition probabilities P_{ij}. It is called the *embedded* Markov chain of the semi-Markov process. We say that the semi-Markov process is *irreducible* if the embedded Markov chain is irreducible as well.

Let T_{ii} denote the time between successive transitions into state i and let $\mu_{ii} = E[T_{ii}]$. By using the theory of alternating renewal processes, it is a simple matter to derive an expression for the limiting probabilities of a semi-Markov process.

PROPOSITION 4.8.1

If the semi-Markov process is irreducible and if T_{ii} has a nonlattice distribution with finite mean, then

$$P_i \equiv \lim_{t\to\infty} P\{Z(t) = i \mid Z(0) = j\}$$

exists and is independent of the initial state. Furthermore,

$$P_i = \frac{\mu_i}{\mu_{ii}}.$$

Proof Say that a cycle begins whenever the process enters state i, and say that the process is "on" when in state i and "off" when not in i. Thus we have a (delayed when $Z(0) \neq i$) alternating renewal process whose on time has distribution H_i and whose cycle time is T_{ii}. Hence, the result follows from Proposition 3.4.4 of Chapter 3.

As a corollary we note that P_i is also equal to the long-run proportion of time that the process is in state i.

Corollary 4.8.2

If the semi-Markov process is irreducible and $\mu_{ii} < \infty$, then, with probability 1,

$$\frac{\mu_i}{\mu_{ii}} = \lim_{t\to\infty} \frac{\text{amount of time in } i \text{ during } [0, t]}{t}.$$

That is, μ_i/μ_{ii} equals the long-run proportion of time in state i.

Proof Follows from Proposition 3.7.2 of Chapter 3.

While Proposition 4.8.1 gives us an expression for the limiting probabilities, it is not, however, the way one actually computes the P_i. To do so suppose that the embedded Markov chain $\{X_n, n \geq 0\}$ is irreducible and positive recurrent, and let its stationary probabilities be $\pi_j, j \geq 0$. That is, the $\pi_j, j \geq 0$, is the unique solution of

$$\pi_j = \sum_i \pi_i P_{ij},$$

$$\sum_j \pi_j = 1,$$

and π_j has the interpretation of being the proportion of the X_n's that equals j. (If the Markov chain is aperiodic, then π_j is also equal to $\lim_{n\to\infty} P\{X_n = j\}$.) Now as π_j equals the proportion of transitions that are into state j, and μ_j is the mean time spent in state j per transition, it seems intuitive that the limiting probabilities should be proportional to $\pi_j\mu_j$. We now prove this.

THEOREM 4.8.3

Suppose the conditions of Proposition 4.8.1 and suppose further that the embedded Markov chain $\{X_n, n \geq 0\}$ is positive recurrent. Then

$$P_i = \frac{\pi_i \mu_i}{\sum_j \pi_j \mu_j}.$$

Proof Define the notation as follows:

$Y_i(j)$ = amount of time spent in state i during the jth visit to that state, $i, j \geq 0$.

$N_i(m)$ = number of visits to state i in the first m transitions of the semi-Markov process.

In terms of the above notation we see that the proportion of time in i during the first m transitions, call it $P_{i=m}$, is as follows:

$$P_{i=m} = \frac{\sum_{j=1}^{N_i(m)} Y_i(j)}{\sum_i \sum_{j=1}^{N_i(m)} Y_i(j)} \tag{4.8.1}$$

$$= \frac{\frac{N_i(m)}{m} \sum_{j=1}^{N_i(m)} \frac{Y_i(j)}{N_i(m)}}{\sum_i \frac{N_i(m)}{m} \sum_{j=1}^{N_i(m)} \frac{Y_i(j)}{N_i(m)}}.$$

Now since $N_i(m) \to \infty$ as $m \to \infty$, it follows from the strong law of large numbers that

$$\sum_{j=1}^{N_i(m)} \frac{Y_i(j)}{N_i(m)} \to \mu_i,$$

and, by the strong law for renewal processes, that

$$\frac{N_i(m)}{m} \to (E[\text{number of transitions between visits to } i])^{-1} = \pi_i.$$

Hence, letting $m \to \infty$ in (4.8.1) shows that

$$\lim_{m\to\infty} P_{i=m} = \frac{\pi_i \mu_i}{\sum_j \pi_j \mu_j},$$

and the proof is complete.

From Theorem 4.8.3 it follows that the limiting probabilities depend only on the transition probabilities P_{ij} and the mean times μ_i, $i, j \geq 0$.

Example 4.8(a) Consider a machine that can be in one of three states: *good condition*, *fair condition*, or *broken down*. Suppose that a machine in good condition will remain this way for a mean time μ_1 and will then go to either the fair condition or the broken condition with respective probabilities $\frac{3}{4}$ and $\frac{1}{4}$. A machine in the fair condition will remain that way for a mean time μ_2 and will then break down. A broken machine will be repaired, which takes a mean time μ_3, and when repaired will be in the good condition with probability $\frac{2}{3}$ and the fair condition with probability $\frac{1}{3}$. What proportion of time is the machine in each state?

Solution. Letting the states be 1, 2, 3, we have that the π_i satisfy

$$\pi_1 + \pi_2 + \pi_3 = 1,$$
$$\pi_1 = \tfrac{2}{3}\pi_3,$$
$$\pi_2 = \tfrac{3}{4}\pi_1 + \tfrac{1}{3}\pi_3,$$
$$\pi_3 = \tfrac{1}{4}\pi_1 + \pi_2.$$

The solution is

$$\pi_1 = \tfrac{4}{15}, \qquad \pi_2 = \tfrac{1}{3}, \qquad \pi_3 = \tfrac{2}{5}.$$

Hence, P_i, the proportion of time the machine is in state i, is given by

$$P_1 = \frac{4\mu_1}{4\mu_1 + 5\mu_2 + 6\mu_3},$$
$$P_2 = \frac{5\mu_2}{4\mu_1 + 5\mu_2 + 6\mu_3},$$
$$P_3 = \frac{6\mu_3}{4\mu_1 + 5\mu_2 + 6\mu_3}.$$

The problem of determining the limiting distribution of a semi-Markov process is not completely solved by deriving the P_i. For we may ask for the limit, as $t \to \infty$, of being in state i at time t of making the next transition after time $t + x$, and of this next transition being into state j. To express this probability let

$$Y(t) = \text{time from } t \text{ until the next transition},$$
$$S(t) = \text{state entered at the first transition after } t.$$

To compute

$$\lim_{t\to\infty} P\{Z(t) = i, Y(t) > x, S(t) = j\},$$

we again use the theory of alternating renewal processes.

THEOREM 4.8.4

If the semi-Markov process is irreducible and not lattice, then

(4.8.2)
$$\lim_{t\to\infty} P\{Z(t) = i, Y(t) > x, S(t) = j \mid Z(0) = k\} = \frac{P_{ij} \int_x^\infty \overline{F}_{ij}(y)\, dy}{\mu_{ii}}.$$

Proof Say that a cycle begins each time the process enters state i and say that it is "on" if the state is i and it will remain i for at least the next x time units and the next state is j. Say it is "off" otherwise. Thus we have an alternating renewal process. Conditioning on whether the state after i is j or not, we see that

$$E[\text{"on" time in a cycle}] = P_{ij} E[(X_{ij} - x)^+],$$