Stochastic Processes

CHAPTER 1

1.1 Introduction

The theory of stochastic processes mainly originated from the needs of physicists. It began with the study of physical phenomena as random phenomena changing with time. Let t be a parameter assuming values in a set T, and let $X(t)$ represent a random or stochastic variable for every $t \in T$. The family or collection of random variables $\{X(t), t \in T\}$ is called a stochastic process. The parameter or index t is generally interpreted as time and the random variable $X(t)$ as the state of the process at time t. The elements of T are time points, or epochs, and T is a linear set, denumerable or nondenumerable. If T is countable (or denumerable), then the stochastic process $\{X(t), t \in T\}$ is said to be a discrete-parameter or discrete-time process, while if T is an interval of the real line, then the stochastic process is said to be a continuous-parameter or continuous-time process. For example, $\{X_n, n = 0, 1, 2, \ldots\}$ is a discrete-time and $\{X(t), t \geq 0\}$ is a continuous-time process. The set of all possible values that the random variable $X(t)$ can assume is called the state space of the process; this set may be countable or noncountable. Thus, stochastic processes may be classified into these four types:

(i) discrete-time and discrete state space,

(ii) discrete-time and continuous state space,

(iii) continuous-time and discrete state space, and

(iv) continuous-time and continuous state space.

A discrete state space process is often referred to as a chain. A process such as (i) is a discrete-time chain, and a process such as (iii) is a continuous-time chain.

A stochastic process—that is, a family of random variables—thus provides description of the evolution of some physical phenomenon through time. Queueing systems provide many examples of stochastic processes. For example, $X(t)$ might be the number of customers that arrive before a service counter by time t; then $\{X(t), t \geq 0\}$ is of the type (iii) above. Again W_n might be the queueing time of the nth arrival; then $\{W_n, n = 0, 1, 2, \ldots\}$ is of the type (ii) above.

Stochastic processes play an important role in modeling queueing systems. Certain stochastic processes are briefly discussed in this chapter.

1.2 Markov[1] Chains

1.2.1 Basic ideas

Suppose that we observe the state of a system at a discrete set of time points $t = 0, 1, 2, \ldots$. The observations at successive time points define a set of random variables (RVs) $X_0, X_1, X_2, \ldots$. The values assumed by the RVs X_n are the states of the system at time n. Assume that X_n assumes the finite set of values $0, 1, \ldots, m$; then $X_n = i$ implies that the state of the system at time n is i. The family of random variables (RVs) $\{X_n, n \geq 0\}$ is a stochastic process with discrete parameter space $n = 0, 1, 2, \ldots$ and discrete state space $S = \{0, 1, \ldots, m\}$.

Definition 1.1. A stochastic process $\{X_n, n \geq 0\}$ is called a Markov chain, if for every $x_i \in S$,

$$\begin{aligned} Pr\{X_n = x_n \mid X_{n-1} = x_{n-1}, \ldots, X_0 = x_0\} \\ = Pr\{X_n = x_n \mid X_{n-1} = x_{n-1}\}, \end{aligned} \tag{1.2.1}$$

provided the first member (LHS) is defined. Equation (1.2.1) indicates a kind of dependence between the RVs X_n; intuitively, it implies that given the present state of the system, the future is independent of the past. The conditional probability

$$Pr\{X_n = k \mid X_{n-1} = j\}, \quad j, k \in S$$

is called the *transition probability* from state j to state k. This is denoted by

$$p_{jk}(n) = Pr\{X_n = k \mid X_{n-1} = j\}. \tag{1.2.2}$$

[1] A. A. Markov (1856–1922)

The Markov chain will be called (temporally) *homogeneous* if $p_{jk}(n)$ does not depend on n—that is,

$$Pr\{X_n = k \mid X_{n-1} = j\} = Pr\{X_{n+m} = k \mid X_{n+m-1} = j\}$$

for $m = -(n-1), -(n-2), \ldots, 0, 1, \ldots$. In such cases we denote $p_{jk}(n)$ simply by p_{jk}. The transition probability p_{jk}, which is the probability of transition from state j to state k in one step—that is, from step $n-1$ to next step n (or from step $n+m-1$ to step $n+m$)—is called one-step transition probability; the transition probability

$$Pr\{X_{r+n} = k \mid X_r = j\}$$

from state j to state k in n steps (from state j in step r to state k in step $r+n$) is called the n-step transition probability. We denote

$$p_{jk}^{(n)} = Pr\{X_{r+n} = k \mid X_r = j\} \tag{1.2.3}$$

so that $p_{jk}^{(1)} = p_{jk}$. Define

$$p_{jk}^{(0)} = \begin{cases} 1, & k = j \\ 0, & k \neq j. \end{cases}$$

Then (1.2.3) is defined for $n = 0, 1, 2, \ldots$. Denote

$$\pi_j = Pr\{X_0 = j\} \quad \text{and} \quad \boldsymbol{\pi}(0) = \{\pi_0, \pi_1, \ldots, \pi_m\};$$

$\boldsymbol{\pi}(0)$ is the initial probability vector. We have

$$\begin{aligned} Pr\{X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n\} \\ &= Pr\{X_0 = x_0, \ldots, X_{n-1} = x_{n-1}\} \\ &\quad \times Pr\{X_n = x_n \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}\} \\ &= Pr\{X_0 = x_0\} p_{x_0 x_1} p_{x_1 x_2} \cdots p_{x_{n-1} x_n}. \end{aligned} \tag{1.2.4}$$

Thus, given p_{jk} and $\boldsymbol{\pi}(0)$, the joint probability given by (1.2.4) can be determined.

The matrix $\boldsymbol{P} = (p_{jk}),\ j, k \in S$ is called the transition matrix or *transition probability matrix* (TPM) of the Markov chain. $\boldsymbol{P}$ is a nonnegative square matrix with unit row sums—that is, $0 \leq p_{jk} \leq 1, \sum_k p_{jk} = 1$ for every $j \in S$.

A nonnegative square matrix P with unit row sums is called a *stochastic matrix*.

It can be easily shown that P^n is also a stochastic matrix and that

$$\left(p_{jk}^{(n)}\right) = P^n. \tag{1.2.5}$$

That is,

$$p_{jk}^{(2)} = \sum_{r \in s} p_{jr} p_{rk} \quad \text{for every } j,\ k \in S$$

$$p_{jk}^{(n)} = \sum p_{jr}^{(n-1)} p_{rk},$$

more generally,

$$\begin{aligned} p_{jk}^{(m+n)} &= \sum_r p_{jr}^{(m)} p_{rk}^{(n)} \\ &= \sum_r p_{jr}^{(n)} p_{rk}^{(m)}, \quad r \in S. \end{aligned} \tag{1.2.6}$$

Equation (1.2.6) is a special case of the *Chapman-Kolmogorov equation*. It is satisfied by transition probabilities of a Markov chain.

Remark: To every stochastic matrix $P = (p_{ij}), i, j = 0, 1, \ldots$ there exists a homogeneous Markov chain $\{X_n, n = 0, 1, \ldots\}$ with state space $S = \{0, 1, \ldots\}$ and one-step transition probability $p_{ij}, i, j \in S$.

That is, to every stochastic matrix P, there corresponds a Markov chain $\{X_n\}$ for which P is the unit-step transition matrix.

Then $P^2 = (p_{ij}^{(2)})$ is also stochastic; it is the two-step transition matrix for the chain $\{X_n, n = 0, 1, \ldots\}$. However, not every stochastic matrix is the two-step transition matrix of a Markov chain.

1.2.2 Classification of states and chains

1.2.2.1 Finite homogeneous chain

Let $\{X_n, n \geq 0\}$ be a finite homogeneous Markov chain having TPM $P = (p_{jk})$ and state space S, and let i, j, k be arbitrary states of S.

State i is said to lead to state j (or state j is said to be accessible from state i) and is denoted by $i \rightarrow j$, if there exists an integer $m(\geq 1)$ such that $p_{ij}^{(m)} > 0$. If no such integer exists, then we say that i does not lead to j: denote this by $i \not\rightarrow j$. Two states are said to communicate with each other if $i \rightarrow j$ and $j \rightarrow i$; this is denoted by $i \leftrightarrow j$. The relations $i \rightarrow j$ and $i \leftrightarrow j$ are transitive.

One way to classify the state of a chain is given below.

If $i \rightarrow j$ but $j \not\rightarrow i$, then index i is said to be inessential. If $i \rightarrow j$ implies $i \leftrightarrow j$ for at least one j, then i is said to be essential. All essential states can be grouped into a number of essential classes such that all states belonging to an essential class communicate with one another, but cannot lead to a state outside the class. An essential class is closed—that is, if i, j, k belong to an essential class and l is another state outside the class, then $i \leftrightarrow j \leftrightarrow k$, but $i \not\rightarrow l$, $j \not\rightarrow l$, $k \not\rightarrow l$ (though $l \rightarrow i$, j, or k).

Inessential states, if any, can be grouped into a number of inessential classes such that all states belonging to an inessential class communicate with all states in the class. A finite homogeneous Markov chain has at least one essential class of states, whereas a Markov chain with denumerable number of states may not necessarily have any essential class.

A Markov chain is said to be *irreducible* if it contains exactly one essential class of states; in this case every state communicates with every other state of the chain. A nonirreducible (or reducible) chain may have more than one essential class of states as well as some inessential classes.

Suppose that $i \to i$—that is, there exists some $m(\geq 1)$ such that $p_{ii}^{(m)} > 0$. The greatest common divisor of all such m for which $p_{ii}^{(m)} > 0$ is called the period $d(i)$ of the state i. If $d(i) = 1$, then state i is said to be *aperiodic* (or *acyclic*), and if $d(i) > 1$, state i is said to be *periodic* with period $d(i)$. If $p_{ii} > 0$, then clearly state i is aperiodic.

For an irreducible Markov chain, either all states are aperiodic or they are periodic having the same period $d(i) = d(j) = \cdots$. Thus, irreducible Markov chains can be divided into two classes: aperiodic and periodic. An irreducible Markov chain is said to be *primitive* if it is aperiodic and *imprimitive* if it is periodic.

A Markov chain whose essential states form a single essential class and are aperiodic is said to be *regular*. Such a chain may have some inessential indices as well. Note that the transitions between states of the essential class of a regular chain form a submatrix P_1 that is stochastic. The TPM P of a regular chain can be written in canonical form,

$$P = \begin{pmatrix} P_1 & 0 \\ R_1 & Q \end{pmatrix} \tag{1.2.7}$$

where the stochastic submatrix P_1 corresponds to transitions between states of essential class, the square matrix Q to transitions between inessential states, and the rectangular matrix R_1 to transitions between inessential states and essential states (see Seneta (1981) for details).

1.2.2.2 Ergodicity property

We shall now discuss an important concept: the ergodicity property. The classification given above is enough for this discussion so far as finite chains are concerned.

Before considering ergodicity, we shall describe another concept: invariant measure.

If $\{X_n\}$ is a Markov chain with TPM P, and if there exists a probability vector $\boldsymbol{V} = (v_1, v_2, \ldots)$ (i.e., $0 \leq v_i \leq 1, \sum v_i = 1$) such that

$$\boldsymbol{V}P = \boldsymbol{V},$$

then $\boldsymbol{V}$ is called an *invariant measure* (or *stationary distribution*) of the Markov chain $\{X_n\}$ or with respect to the stochastic matrix P.

If there exists $\boldsymbol{V}$ such that $\boldsymbol{V}P \leq \boldsymbol{V}$, then $\boldsymbol{V}$ is called a subinvariant measure of the chain with TPM P. Our interest lies in the types of Markov chains that possess invariant measures.

For a finite, irreducible Markov chain with TPM P, an invariant measure exists and is unique—that is, there is a unique probability vector $\boldsymbol{V}$ such that

$$VP = \boldsymbol{V}, \qquad \boldsymbol{Ve} = 1, \tag{1.2.8}$$

where $\boldsymbol{e} = (1, 1, \ldots, 1)$ is a column vector with all its elements equal to unity. We next discuss the limiting behavior of chains.

1.2.2.3 Ergodic theorems

Theorem 1.1. Ergodic Theorem for Primitive Chains
Let $\{X_n, n \geq 0\}$ be a finite irreducible aperiodic Markov chain with TPM P and state space S. Then, as $n \to \infty$,

$$P^n \to \boldsymbol{eV} \tag{1.2.9}$$

elementwise, where $\boldsymbol{V}$ is the unique stationary distribution (or invariant measure) of the chain.

Further, the rate of approach to the limit is geometrically fast—that is, there exist positive constants $a, b, 0 < b < 1$ such that $\varepsilon_{ij}^{(n)} \leq ab^n$, where $p_{ij}^{(n)} = v_j + \varepsilon_{ij}^{(n)}$.

The above theorem implies that, for every $j \in S$,

$$\lim_{n\to\infty} p_{ij}^{(n)} \to v_j \tag{1.2.10}$$

exists and is independent of the initial state i and the quantities $v_j's$ are given by the solution of the matrix equation

$$VP = \boldsymbol{V} \quad with\ \boldsymbol{Ve} = 1. \tag{1.2.11}$$

The limiting probability distribution of the chain tends to an equilibrium distribution that is independent of the initial distribution. This tendency is known as ergodicity or ergodic property.

Another ergodic theorem is stated below.

Theorem 1.2. *Let $\{X_n, n \geq 0\}$ be a finite k state regular Markov chain (i.e., having a single essential class of aperiodic states) and having TPM P. Let $\boldsymbol{V}_1$ be the stationary distribution corresponding to the primitive submatrix P_1 (corresponding to the transitions between the states of the essential aperiodic class).*

Let $\boldsymbol{V} = (V_1, 0)$ be a $1 \times k$ vector. Then, as $n \to \infty$,

$$P^n \to \boldsymbol{eV} \tag{1.2.9a}$$

elementwise.

$\boldsymbol{V}$ is the unique stationary distribution corresponding to the matrix P and the rate of approach to the limit in (1.2.9a) is geometrically fast.

If C denotes the single essential class with states $j, j = 1, 2, \ldots, m (m < k)$, $j \in C$, and $\boldsymbol{V}_1 = (v_1, v_2, \ldots, v_j, \ldots, v_m)$ is given by the solution of

$$\boldsymbol{V}_1 P_1 = \boldsymbol{V}_1, \qquad \boldsymbol{V}_1 \boldsymbol{e} = 1, \tag{1.2.12}$$

then the above result implies that, as $n \to \infty$,

$$\begin{aligned} \lim p_{ij}^{(n)} &\to v_j, \qquad j \in C, \quad \text{and} \\ \lim p_{ij}^{(n)} &\to 0, \qquad j \notin C. \end{aligned} \tag{1.2.13}$$

The above theorem asserts that regularity of the chain is a sufficient condition for ergodicity. It is also a necessary condition.

Example 1.1. Consider the Markov chain having state space $S = \{0, 1, 2\}$ and TPM P

$$P = \begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{3}{4} & \frac{1}{4} & 0 \end{pmatrix}$$

This is a finite irreducible chain. Its invariant measure $\boldsymbol{V} = (v_1, v_2, v_3)$ is given by the solution of

$$\boldsymbol{V}P = \boldsymbol{V},$$

which leads to

$$\begin{aligned} v_1 &= \frac{1}{2}v_2 + \frac{3}{4}v_3 \\ v_2 &= \frac{1}{3}v_1 + \frac{1}{4}v_3 \\ v_3 &= \frac{2}{3}v_1 + \frac{1}{2}v_2. \end{aligned}$$

As $\boldsymbol{Ve} = v_1 + v_2 + v_3 = 1$, one of the above equations is redundant. We get

$$\boldsymbol{V} = (v_1, v_2, v_3),$$

where

$$v_1 = \frac{21}{53}, \quad v_2 = \frac{12}{53}, \quad v_3 = \frac{20}{53}.$$

Thus,

$$P^n \to \boldsymbol{eV}.$$

That is, as $n \to \infty$, for all $i = 1, 2, 3$,

$$\lim p_{i1}^{(n)} = \frac{21}{53}, \qquad \lim p_{i2}^{(n)} = \frac{12}{53}, \qquad \lim p_{i3}^{(n)} = \frac{20}{53}.$$

Example 1.2. Consider the two-state Markov chain with TPM

$$P = \begin{pmatrix} 1-p & p \\ p & 1-p \end{pmatrix}, \quad 0 \le p \le 1.$$

The equation $VP = V, (V = (v_1, v_2))$ leads to

$$v_1 = v_2 = \frac{1}{2}, \quad \text{for all } p,$$

so that the invariant distribution is $(\frac{1}{2}, \frac{1}{2})$ for all p. We have, for $p \neq 0$

$$P^n = e\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2}(1-2p)^n\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

For $p = 0$, the chain consists of two absorbing states 0 and 1—that is, it consists of two essential classes, C_1 and C_2, with members 0 and 1, respectively. The chain is decomposable. For $p = 1$,

$$\begin{aligned} P^n &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{when } n \text{ is even and} \\ &= \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{when } n \text{ is odd.} \end{aligned}$$

The chain is periodic.

Thus, even though invariant distribution exists for all $p, 0 \leq p \leq 1$, as $n \to \infty$,

$$\begin{aligned} &\lim p_{i0}^{(n)} = \frac{1}{2} \\ \text{and} \quad &\lim p_{i1}^{(n)} = \frac{1}{2} \end{aligned}$$

exist only when $p \neq 0, 1$.

Remarks: We can make two deductions: (1) Existence of stationary distribution (i.e., existence of a solution of $VP = V, Ve = 1$) does not necessarily imply existence of the limiting distribution

$$\left\{\lim_{n\to\infty} p_{ij}^{(n)}\right\}$$

(2) The example also shows how much the transient behavior can vary over the class of transient matrices for a given equilibrium distribution.

See Whitt (1983) for a discussion on this topic.

1.2.2.4 Markov chain having a denumerably infinite number of states

So far we have considered homogeneous Markov chains with a finite number of states. Now we shall discuss homogeneous chains having a denumerably infinite number of states. We shall denote the state space by $S = \{0, 1, 2, \ldots\}$ instead of the more general

$$S = \{\ldots, -3, -2, -1, 0, 1, 2, \ldots\}$$

Notions already defined (e.g., accessibility, communication, and periodicity) and subsequent definitions of essential and inessential states, essential classes, inessential classes, irreducible chains, and primitive chains will remain valid for a chain with a denumerable number of states. Only the definition of a regular chain cannot be carried over to this case. For whereas in a finite chain there is at least one essential class (and so the states of a finite chain may constitute exactly one essential class or more than one essential class), there may not be any essential class in the case of a chain with a denumerably infinite number of states. For example, consider the chain $\{X_n, n \geq 0\}$ with $S = \{0, 1, 2, \ldots\}$ and TPM $P = (p_{ij})$ where

$$\begin{aligned} p_{ij} &= 1 \quad j = i + 1 \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

This denumerable chain does not possess any essential class at all. Each state is inessential, and no two states communicate. Whereas consideration of classification of states into essential and inessential classes was adequate for dealing with limiting distribution for finite chains, a more sensitive classification of states would be required in the present case of chains with a denumerable infinity of states.

1.2.2.5 Transience and Recurrence

Define

$$\begin{aligned} &\{f_{ij}^{(n)}\}, i, j = 1, 2, \ldots, n, \\ &f_{ij}^{(0)} = 0, \qquad f_{ij}^{(1)} = p_{ij}, \text{ and} \\ &f_{ij}^{(k+1)} = \sum_{\substack{r \\ r \neq j}} p_{ir} f_{rj}^{(k)}, \qquad k \geq 1. \end{aligned} \tag{1.2.14}$$

The quantity $f_{ij}^{(k)}$ is the probability of transition from state i to state j in k steps, without revisiting the state j in the meantime. (It is called *taboo* probability—j being the *taboo* state.) Here $\{f_{ij}^{(k)}\}$ gives the distribution of the first passage time from state i to state j. We can write

$$f_{ij}^{(n)} = Pr\{X_n = j, X_r \neq j, \;\; r = 1, 2, \ldots, n-1 \mid X_0 = i\}$$

The relation (1.2.14) can also be written as

$$\begin{aligned} p_{ij}^{(n)} &= \sum_{r=0}^{n} f_{ij}^{(r)} p_{jj}^{(n-r)}, \quad n \geq 1 \\ &= \sum_{r=0}^{n} f_{ij}^{(n-r)} p_{jj}^{(r)}. \end{aligned} \tag{1.2.15}$$

The relations (1.2.14) and (1.2.15) are known as *first entrance formulas*.

Let

$$P_{ij}(s) = \sum_n p_{ij}^{(n)} s^n, \qquad F_{ij}(s) = \sum_n f_{ij}^{(n)} s^n, \quad |s| < 1.$$

Then from the convolution structure,

$$\begin{aligned} P_{ij}(s) &= P_{ii}(s) F_{ij}(s), \quad j \neq i \\ P_{ii}(s) - 1 &= P_{ii}(s) F_{ii}(s). \end{aligned} \tag{1.2.16}$$

Definition 1.2. A state i is said to be *persistent* if $F_{ii} = F_{ii}(1-0) = 1$ and is said to be *transient* if $F_{ii}(1-0) < 1$.

A persistent state is null or nonnull based on whether $\mu_{ii} = F'_{ii}(1) = \infty$ or $<\infty$, respectively.

Equivalent criteria of persistence and recurrence are as follows.

An index i is persistent *iff* (if and only if)

$$\sum_n p_{ii}^{(n)} = \infty$$

and is transient *iff*

$$\sum_n p_{ii}^{(n)} < \infty.$$

The relationship between these two types of classification of states and chain can be given as follows.

An inessential state is transient and a persistent state is essential. In the case of a finite chain, i is transient *iff* it is inessential; otherwise it is nonnull persistent.

All the states of an irreducible chain, whether finite or denumerable, are of the same type: all transient, all null persistent, or all nonnull persistent.

A finite Markov chain contains at least one persistent state. Further, a finite irreducible Markov chain is nonnull persistent. The ergodic theorem for a Markov chain with a denumerable infinite number of states is stated below.

Theorem 1.3. General Ergodic Theorem

Let P be the TPM of an irreducible aperiodic (i.e., primitive) Markov chain with a countable state space S (which may have a finite or a denumerably infinite number of states). If the Markov chain is transient or null persistent, then for each $i, j \in S$,

$$\lim_{n\to\infty} p_{ij}^{(n)} \to 0. \tag{1.2.17a}$$

If the chain is nonnull persistent, then for each $i, j \in S$,

$$\lim_{n\to\infty} p_{ij}^{(n)} = v_j \tag{1.2.17b}$$

exists and is independent of i. The probability vector $V = (v_1, v_2, \ldots)$ *is the unique invariant measure of P—that is,*

$$VP = V, \qquad Ve = 1; \tag{1.2.18a}$$

and further if μ_{jj} *is the mean recurrence time of state* j, *then*

$$v_j = (\mu_{jj})^{-1}. \tag{1.2.18b}$$

The result is general and holds for a chain with a countable state space S. In case the chain is finite, irreducibility ensures nonnull persistence, so that irreducibility and aperiodicity (i.e., primitivity) constitute a set of sufficient conditions for ergodicity of a finite chain. The sufficient conditions for ergodicity ($\lim p_{ij}^{(n)} = v_i$) *for a chain with a denumerably infinite number of states involve, besides irreducibility and aperiodicity, nonnull persistence of the chain. For a chain with a denumerably infinite number of states, the number of equations given by (1.2.18a) will be infinite. It would sometimes be more convenient to find V in terms of the generating function of* $\{v_j\}$ *than to attempt to solve Eqn.(1.2.18a) as such. We shall consider two such Markov chains that arise in queueing theory. See the Note below.*

Example 1.3. Consider a Markov chain with state space $S = \{0, 1, 2, \ldots\}$ having a denumerable number of states and having TPM

$$\boldsymbol{P} = \begin{bmatrix} p_0 & p_1 & p_2 & p_3 & \cdots \\ p_0 & p_1 & p_2 & p_3 & \cdots \\ 0 & p_0 & p_1 & p_2 & \cdots \\ 0 & 0 & p_0 & p_1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{1.2.19}$$

where $\sum_k p_k = 1$. Let

$$P(s) = \sum_k p_k s^k \quad \text{and} \quad V(s) = \sum_k v_k s^k, \quad |s| < 1$$

be the probability-generating functions (PGF) of $\{p_k\}$ and $\{v_k\}$, respectively. Clearly, the chain is irreducible and aperiodic; since it is a denumerable chain, we need to consider transience and persistence of the chain to study its ergodic property.

It can be shown that the states of the chain (which are all of the same type because of the irreducibility of the chain) are transient, persistent null or persistent nonnull according to

$$P'(1) > 1, \qquad P'(1) = 1, \qquad P'(1) < 1,$$

respectively (see Prabhu, 1965). Assume that $P'(1) < 1$, so that the states are persistent nonnull; then from (1.2.18a), we get

$$v_k = p_k v_0 + p_k v_1 + p_{k-1} v_2 + \cdots + p_0 v_{k+1}, \quad k \geq 0. \tag{1.2.20}$$

Multiplying both sides of (1.2.20) by s^k and adding over $k = 0, 1, 2, \ldots$, we get

$$\begin{aligned} V(s) &= v_0 P(s) + v_1 P(s) + v_2 s P(s) + \cdots + v_{k+1} s^k P(s) + \cdots \\ &= P(s)[v_0 + (V(s) - v_0)/s]; \end{aligned}$$

whence

$$V(s) = \frac{v_0(1-s)P(s)}{P(s) - s}.$$

Since $V(1) = 1$, we have

$$\begin{aligned} 1 = \lim_{s\to 1} V(s) &= v_0 \left[\lim_{s\to 1} \frac{(1-s)P(s)}{P(s)-s} \right] \\ &= v_0 \left[\frac{1}{1 - P'(1)} \right]. \end{aligned}$$

Thus,

$$V(s) = \frac{(1 - P'(1))(1-s)P(s)}{P(s) - s}. \tag{1.2.21}$$

Example 1.4. Consider a Markov chain with state space $S = \{0, 1, 2, \ldots\}$ and having TPM

$$P = \begin{bmatrix} h_0 & g_0 & 0 & 0 & 0 & \cdots \\ h_1 & g_1 & g_0 & 0 & 0 & \cdots \\ h_2 & g_2 & g_1 & g_0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \tag{1.2.22}$$

where $h_i = g_{i+1} + g_{i+2} + \cdots, i \geq 0, g_i > 0$, and $\sum_{i=0} g_i = 1$. Here $p_{i0} = h_i, i \geq 0$,

$$\begin{aligned} p_{ij} &= g_{i+1-j}, \quad i+1 \geq j \geq 1, \quad i \geq 0 \\ &= 0, \quad i + 1 < j. \end{aligned}$$

The chain is irreducible and aperiodic. It can be shown that it is persistent non-null when $\alpha = \sum j g_j > 1$. Then the chain is ergodic and $v_j = \lim_{n\to\infty} p_{ij}^{(n)}$ exist and are given as a solution of (1.2.18a); these lead to

$$v_0 = \sum v_r h_r \tag{1.2.23a}$$

$$v_j = \sum v_{r+j-1} g_r, \quad j \geq 1 \tag{1.2.23b}$$

$$\sum_{j=0} v_j = 1 \tag{1.2.23c}$$

Let $G(s) = \sum_r g_r s^r$ be the PGF of $\{g_r\}$. Denote the displacement operator by E so that

$$E^r(v_k) = v_{k+r}, \quad r = 0, 1, 2, \ldots$$

Then we can write (1.2.23b) in symbols as

$$\begin{aligned} E(v_{j-1}) &= v_j = \sum g_r E^r(v_{j-1}), \quad j \geq 1 \quad \text{or} \\ \left\{E - \sum g_r E^r\right\} v_j &= 0, \quad j \geq 0 \quad \text{or} \\ \{E - G(E)\} v_j &= 0, \quad j \geq 0. \end{aligned} \tag{1.2.24}$$

The characteristic equation of the above difference equation is given by

$$r(z) \equiv z - G(z) = 0. \tag{1.2.25}$$

It can be shown that when $\alpha = G'(1) > 1$, there is exactly one real root of $r(z) = 0$ between 0 and 1. Denote this root by r_0 and the other roots by $r_1, r_2, \ldots, |r_i| > 1, i \geq 1$. The solution of (1.2.24) can thus be put as

$$v_j = c_0 r_0^j + \sum_{i=j} c_i\, r_i^j, \quad j \geq 0,$$

where the c's are constants. Since $\sum v_j = 1$,

$$\begin{aligned} &c_i \equiv 0 \quad \text{for } i \geq 1, \\ &v_j = c_0 r_0^j, \quad j \geq 0, \quad \text{and} \\ &c_0 = 1 - r_0 \end{aligned}$$

so that

$$v_j = (1 - r_0) r_0^j, \quad j \geq 0, \tag{1.2.26}$$

r_0 being the root lying between 0 and 1 of (1.2.25) (provided $\alpha = G'(1) > 1$). The distribution is geometric.

Notes:

(1) The equation $VP = V$ is quite well known in the matrix theory. It follows from the well-known Perron-Frobenius theorem of matrix theory that there exists a solution $V = (v_1, v_2, \ldots)$ of the matrix equation $VP = V$ subject to the constraints $v_i \geq 0, \sum v_i = 1$.

(2) When the order of P is not large, the equations can be solved fairly easily to get $V = (v_1, v_2, \ldots)$. When the order of P is large (infinite), the number of equations is also large (infinite) and the solution of the equations becomes troublesome. In Example 1.3 we considered and obtained the solution in terms of the generating function $V(s) = \sum v_j s'$. This method may not always be applicable.

See also Remarks (4) in Section 1.3.5.

1.3 Continuous-Time Markov Chains

We shall now consider continuous-time Markov chains—that is, Markov processes with discrete state space. Let $\{X(t), 0 \le t < \infty\}$ be a Markov process with countable state space $S = \{0, 1, 2, \ldots\}$. We assume that the chain is temporally homogeneous. The transition probability function given by

$$p_{ij}(t) = Pr\{X(t+u) = j \mid X(u) = i\}, \quad t > 0, \quad i, j \in S, \tag{1.3.1}$$

is then independent of $u \ge 0$. We have for all t,

$$0 \le p_{ij}(t) \le 1, \quad \sum_j p_{ij}(t) = 1, \quad \text{for all } j \in S.$$

Denote the matrix of transition probabilities by

$$P(t) = (p_{ij}(t)), \quad i, j \in S.$$

Setting $p_{ij}(0) = \delta_{ij}$, the initial condition can be put as

$$\boldsymbol{P}(0) = \boldsymbol{I}.$$

Denote the probability that the system is at state j at time t by

$$\pi_j(t) = Pr\{X(t) = j\};$$

the vector $\boldsymbol{\pi}(t) = \{\pi_1(t), \pi_2(t), \ldots\}$ is the probability vector of the state of the system at time t; $\boldsymbol{\pi}(0)$ is the initial probability vector. Now

$$\begin{aligned}\pi_j(t) &= \sum_i Pr\{X(t+u) = j \mid X(u) = i\} Pr\{X(u) = i\} \\ &= \sum_i p_{ij}(t)\ Pr\{X(0) = i\} \\ &= \sum_i p_{ij}(t)\pi_i(0).\end{aligned} \tag{1.3.2}$$

Thus, given initial probability vector $\boldsymbol{\pi}(0)$ and the transition functions $p_{ij}(t)$, the state probabilities can be calculated and the probabilistic behavior of the system can be completely determined. The matrix form of (1.3.2) is

$$\pi(t) = \pi(0)P(t). \tag{1.3.3}$$

1.3.1 Sojourn time

The time taken (or the waiting time) for change of state from state i is a random variable—say, τ_i; that is, the sojourn time at state i is τ_i. Then

$$\begin{aligned}&Pr\{\tau_i > s + t \mid X(0) = i\} \\ &\quad = Pr\{\tau_i > s + t \mid X(0) = i, \quad \tau_i > s\} \\ &\qquad \times Pr\{\tau_i > s \mid X(0) = i\}, \quad t \ge 0.\end{aligned} \tag{1.3.4}$$

Denote

$$\bar{F}_i(u) = Pr\{\tau_i > u \mid X(0) = i\}, \quad u \geq 0.$$

Then (1.3.4) can be written as

$$\bar{F}_i(t+s) = \bar{F}_i(t)\bar{F}_i(s), \quad s, t \geq 0;$$

$\bar{F}(.)$ is right continuous; the only right continuous solution of the functional equation is

$$\bar{F}_i(u) = e^{-a_i u}, \quad u \geq 0, \quad a_i > 0 \text{ is a constant.} \tag{1.3.5}$$

That is, sojourn time τ_i at state i is exponential with parameter a_i. Further, the sojourn times τ_i and τ_j are independent. We have, for $t \geq 0$, $T \geq 0$,

$$p_{ij}(T+t) = \sum_k p_{ik}(T)p_{kj}(t), \quad i, j, k \in S \tag{1.3.6}$$

or, in matrix form,

$$P(T+t) = P(T)P(t), \tag{1.3.7}$$

which is called the *Chapman-Kolmogorov equation.*

1.3.2 Transition density matrix or infinitesimal generator

Denote the right-hand derivative at $t = 0$, by

$$\begin{aligned} q_{ij} &= \lim_{h\to 0} \frac{p_{ij}(h) - p_{ij}(0)}{h} = \lim_{h\to 0} \frac{p_{ij}(h)}{h}, \quad i \neq j \quad \text{and} \\ q_{ii} &= \lim_{h\to 0} \frac{p_{ii}(h) - p_{ii}(0)}{h} = \lim_{h\to 0} \frac{p_{ii}(h) - 1}{h}; \end{aligned} \tag{1.3.8}$$

write $-q_{ii} = q_i$. It is to be noted that $q_{ij}, i \neq j$ is always finite. While $q_i(\geq 0)$ always exists and is finite when S is finite, q_i may be infinite when S is denumerably infinite. Writing $\boldsymbol{Q} = (q_{ij})$, we can denote (1.3.8) in matrix notation as

$$Q = \lim_{h\to 0} \frac{P(h) - \boldsymbol{I}}{h}.$$

From (1.3.8) it follows that, for small h,

$$\begin{aligned} p_{ij}(h) &= hq_{ij} + o(h), \quad i \neq j, \\ p_{ii}(h) &= hq_i + o(h), \end{aligned} \tag{1.3.9}$$

where $o(h)$ is used as a symbol to denote a function of h that tends to zero more rapidly than h; that is, $o(h)/h \to 0$ as $h \to 0$. Again,

$$\sum_j p_{ij}(h) = 1, \quad \text{or}$$

$$\sum_{j \neq i} p_{ij}(h) + p_{ii}(h) - 1 = 0; \quad \text{whence we get}$$

$$\sum_{j \neq i} q_{ij} + q_{ii} = 0$$

or

$$\sum_{j \neq i} q_{ij} = q_i. \tag{1.3.10}$$

The matrix $Q = (q_{ij})$ is called the *transition density matrix* or *infinitesimal generator* or *rate matrix* or simply Q-matrix. The Q-matrix is such that (i) its diagonal elements are negative and off-diagonal elements are positive, and (ii) each row sum is zero. Let $S = \{0, 1, 2, \ldots, m\}$ be a finite set, then

$$Q = \begin{bmatrix} -q_0 & q_{01} & \cdots & q_{0m} \\ q_{10} & -q_1 & \cdots & q_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ q_{m0} & q_{m1} & \cdots & -q_m \end{bmatrix}$$

1.3.3 Limiting behavior: ergodicity

The states of a continuous-time Markov chain admit of a classification similar to those of a discrete-time chain. A state j is said to be *accessible* or *reachable* from state $i (i \to j)$ if, for some $t > 0$, $p_{ij}(t) > 0$. States i and j communicate if $i \to j$ and $j \to i$. A continuous-time Markov chain is said to be *irreducible* if every state can be reached from every other state (or if each pair of states communicates).

Let α_{ij} denote the (first) entrance time from state i to state j without visiting j in the meantime and let $F(.)$ denote its DF—that is,

$$\begin{aligned} F_{ij}(t) &= Pr\{\alpha_{ij} < t\}, \quad t > 0, \\ &= 0, \quad t \leq 0. \end{aligned}$$

A state i is called *persistent* if

$$\lim_{t \to \infty} F_{ii}(t) = 1$$

and *transient* otherwise.

Criteria of transience and persistence can be expressed in terms of $p_{ij}(t)$ as follows:

State i is transient *iff*

$$\int_0^\infty p_{ii}(t)dt < \infty.$$

If state i, is *null-persistent*, then

$$\lim_{t\to\infty} p_{ii}(t) = 0,$$

and if state i is nonnull-persistent, then

$$\lim_{t\to\infty} p_{ii}(t) > 0.$$

From the Chapman-Kolmogorov equation (1.3.7) we get

$$\begin{aligned} p_{ij}(h+t) &= \sum_k p_{ik}(h)\,p_{kj}(t) \\ &= \sum_{k\neq i} p_{ik}(h)\,p_{kj}(t) + p_{ii}(h)\,p_{ij}(t) \end{aligned}$$

so that

$$\frac{p_{ij}(h+t) - p_{ij}(t)}{h} = \sum_{k\neq i} \frac{p_{ik}(h)}{h} p_{kj}(t) + \left(\frac{p_{ii}(h)-1}{h}\right) p_{ij}(t).$$

Taking the limit as $h \to 0$ and assuming that the order of the operations of taking the limit and summation can be interchanged, we get

$$\begin{aligned} \lim_{h\to 0} \frac{p_{ij}(h+t) - p_{ij}(t)}{h} &= \sum_{k\neq i} \left[\lim_{h\to 0} \frac{p_{ik}(h)}{h}\right] p_{kj}(t) \\ &\quad + \left[\lim_{h\to 0} \frac{p_{ii}(h)-1}{h}\right] p_{ik}(t) \end{aligned}$$

or

$$p'_{ij}(t) = \sum_{k\neq i} q_{ik}\,p_{kj}(t) + q_i\,p_{ij}(t), \tag{1.3.11}$$

which is another form of the Chapman-Kolmogorov (backward) equation; it is in terms of the elements of the Q-matrix. In matrix notation, we get

$$P'(t) = QP(t). \tag{1.3.11a}$$

Again, from (1.3.7) we get

$$\begin{aligned} p_{ij}(t+h) &= \sum_k p_{ik}(t)\,p_{kj}(h) \\ &= \sum_{k\neq j} p_{ik}(t)\,p_{kj}(h) + p_{ij}(t)\,p_{ij}(h). \end{aligned}$$

Assuming that the operations of limit and summation are interchangeable and proceeding as above, we get

$$p'_{ij}(t) = \sum_{k \neq i} p_{ik}(t) q_{kj} + q_j p_{ij}(t), \tag{1.3.12}$$

which is the Chapman-Kolmogorov *forward* equation. In matrix notation,

$$P'(t) = P(t)Q. \tag{1.3.12a}$$

Using (1.3.3), we can also put (1.3.11a) and (1.3.12a) in the form

$$\frac{d}{dt}\{\pi(t)\} = Q\pi(t) = \pi(t)Q. \tag{1.3.13}$$

1.3.4 Transient solution

Consider a finite $(m+1)$ state chain, with given rate matrix Q. Solving (1.3.11a) or (1.3.12a), we get $\boldsymbol{P}(t) = \boldsymbol{P}(0)e^{Q(t)}$, with $P(0) = I$; we have

$$\boldsymbol{P}(t) = e^{Qt} = I + \sum_{n=1}^{\infty} \frac{Q^n t^n}{n!} \tag{1.3.14}$$

or

$$\pi(t) = \pi(0)\left(I + \sum_{n=1}^{\infty} \frac{Q^n t^n}{n!}\right). \tag{1.3.15}$$

Assume that the eigenvalues d_i of Q are all distinct, $d_i \neq d_j, i, j = 0, 1, \ldots, m$. Let D be the diagonal matrix having $d_0, d_1, \ldots, d_m$ as its diagonal elements. Then there exists a nonsingular matrix H (whose column vectors are right eigenvectors of Q) such that Q can be written in the canonical form

$$Q = HDH^{-1}.$$

Then

$$Q^n = HD^nH^{-1}$$

and, substituting in (1.3.14), we get

$$P(t) = H\Lambda(t)H^{-1} \quad \text{and}$$

$$\pi(t) = \pi(0)P(t), \tag{1.3.16}$$

where $\Lambda(t)$ is the diagonal matrix with diagonal elements $e^{d_i t}, i = 0, 1, \ldots, m$.

It may be noted that in the general case when the eigenvalues of the matrix Q are not necessarily distinct, Q can still be expressed in the canonical form $Q = SZS^{-1}$ and $P(t)$ can be obtained as above.

The transient solution can be obtained as given above. While an analytical solution can be obtained, especially when m is small, it becomes difficult when m is large. For such cases, numerical methods have been put forward (Grassman, 1977; Gross and Miller, 1984a,b). See Section 1.6.3.

For many stochastic systems such as queueing systems and reliability systems, computation of the vector $\pi(t)$ transient probabilities is useful. It is especially important when convergence to steady state is slow.

1.3.5 Alternative definition

A continuous-time Markov chain with state space $S = \{0, 1, 2, \ldots\}$ can be defined in another way as follows (Ross, 1980). It is a stochastic process such that (i) each time it enters state i, the time it spends in that state before making a transition to another state $j(\neq i) \in S$—that is, sojourn time in state i is an exponential RV with mean $1/a_i(a_i$ depends on i but not on $j)$; and (ii) when the process leaves state i, it enters another state $j(\neq i)$, with some probability say, p_{ij} (which depends on both i and j), such that, for all i.

$$p_{ii} = 0, \quad 0 \leq p_{ij} \leq 1$$
$$\sum_j p_{ij} = 1, \quad j \in S.$$

Thus, a continuous-time Markov chain is a stochastic process such that (i), its transition from one state to another state of the state space S, is as in a discrete-time Markov chain and (ii) the sojourn in a state i (holding time in state i before moving to another state) is an exponential RV whose parameter depends on i but not on the state next visited. The sojourn times in different states must be independent random variables with exponential distribution.

1.3.5.1 Relationship between p_{ij} and $p_{ij}(t)$

We have

$$p_{ij}(h) = ha_i p_{ij} + o(h),$$

since $p_{ij}(h)$ is the probability that the state of the process changes from i to j in an infinitesimal interval h. Thus,

$$\lim_{h\to o} \frac{p_{ij}(h)}{h} = a_i p_{ij},$$

but by definition LHS equals q_{ij}, so that

$$q_{ij} = a_i p_{ij}. \tag{1.3.17}$$

Again, $1 - p_{ii}(h)$ is the probability that the state of the system changes from state i to some other state in the interval h, so that

$$\begin{aligned} 1 - p_{ii}(h) &= a_i h \sum_j p_{ij} + o(h) \\ &= a_i h + o(h). \end{aligned}$$

Thus,

$$\lim_{h\to o} \frac{1 - p_{ii}(h)}{h} = a_i;$$

but by definition LHS equals q_i so that

$$a_i = q_i. \tag{1.3.18}$$

Thus, the Q matrix can also be written as

$$Q = \begin{bmatrix} -a_0 & a_o p_{01} & \cdots & a_0 p_{0m} \\ a_1 p_{10} & -a_1 & \cdots & a_1 p_{1m} \\ \cdots & & \cdots & \\ a_m p_{m0} & a_m p_{m1} & \cdots & -a_m \end{bmatrix} \tag{1.3.19}$$

We have the corresponding ergodicity property.

Theorem 1.4. Ergodic Theorem
If a Markov chain $\{X(t), t \in T\}$ is irreducible, then all the states are of the same type.

In case they are all transient or null-persistent, then

$$\lim_{i\to\infty} p_{ij}(t) = 0, \quad i, j \in S.$$

In case they are nonnull persistent, then

$$\lim_{i\to\infty} p_{ij}(t) = u_j \tag{1.3.20}$$

exists and is independent of the initial state i. Further, $\boldsymbol{U} = (u_1, u_2, \ldots), (\boldsymbol{Ue} = 1)$ is a probability distribution and is given by the solution of

$$q_j u_j + \sum_{i\neq j} u_i q_{ij} = 0, \quad i, j \in S, \quad or$$

$$\sum_i u_i q_{ij} = 0, \quad or \tag{1.3.21}$$

$$\boldsymbol{UQ} = \boldsymbol{0}, \quad \boldsymbol{Ue} = 1. \tag{1.3.22}$$

We now consider the alternative definition of the continuous time chain. Using (1.3.17) and (1.3.19), we get from (1.3.21)

$$a_j u_j = \sum_{i\neq j} a_i p_{ij} u_i, \quad j \in S, \quad with \quad \boldsymbol{Ue} = 1, \tag{1.3.23}$$

from which $\boldsymbol{U} = (u_1, u_2, \ldots)$ can be obtained.

Remarks

(1) If an irreducible chain is finite, then

$$\lim_{t\to\infty} p_{ij}(t) = u_j, \quad i, j \in S$$

exists. If the chain has a denumerable state space and if it is nonnull persistent, then u_j exists.

(2) When u_j exists, it can be interpreted as the long-run proportion of time the system is in state j.

(3) Equations (1.3.23) have an interesting interpretation.

When the process is in state j, it leaves that state at rate a_j, and u_j is the long-run proportion of time it is in state j, so that $a_j u_j$ = rate at which the process *leaves* state j. Again, when the process is in state i, the rate of transition into state j is $a_i p_{ij} = q_{ij}$, so that

$$\sum_{i\neq j} a_i p_{ij} u_i = \text{ rate at which the process } \textit{enters} \text{ state } j.$$

Thus, Eqn. (1.3.23) can be interpreted as follows: In the long run, the two rates are equal—that is, the rate at which the process enters a state j equals the rate at which it leaves the state j (for each $j \in S$). As the two rates balance each other for every state, Eqn. (1.3.23) are also known as *balance equations.*

The balance equations, as interpreted above, have very useful applications in queueing systems in particular and in stochastic systems in general.

(4) The equations $\boldsymbol{VP} = \boldsymbol{P}$, with $\sum v_i = 1$, (1.2.12) and (1.2.18a) for discrete-time Markov chains written as $\boldsymbol{VQ} = \boldsymbol{O}$, where $\boldsymbol{Q} = \boldsymbol{P} - \boldsymbol{I}$ is a matrix with zero row sums (with $\boldsymbol{Ve} = 1$). Then the equations have the same form as in case of continuous-time Markov chains (see Eqn. (1.3.22)). For methods of numerical solutions of Markov Chains, see Stewart (1994). In queueing context (where such equations occur) besides matrix methods (Kaufman, 1983), the graph-theoretic approach (for $\boldsymbol{Q}$ having certain structural properties) has recently been put forward (Tang and Yeung (1999), who also point out the merit and limitation of their approach).

Example 1.5. Two-State Process

Suppose that a system can be in two states: operating and nonoperating or under repair (denoted by 0 and 1, respectively). Suppose that the lengths of the operating and nonoperating periods are independent exponential RVs with parameters a and b, respectively. Let $X(t)$ be the state of the process at time t. $\{X(t), t \geq 0\}$ is a Markov process with state spaces $S = \{0, 1, \}$.

We have, because of exponential distribution,

$$\begin{aligned}
p_{01}(h) &= Pr\{\text{change of state from operating to} \\
&\qquad \text{nonoperating in an infinitesimal interval } h\} \\
&= ah + o(h), \quad \text{and so} \\
p_{00}(h) &= 1 - ah + o(h) \\
p_{10}(h) &= bh + o(h), \quad \text{and} \\
p_{11}(h) &= 1 - bh + o(h)
\end{aligned}$$

so that the $\boldsymbol{Q}$-matrix is

$$\boldsymbol{Q} = \begin{pmatrix} -a & a \\ b & -b \end{pmatrix}. \tag{1.3.24}$$

The Chapman-Kolmogorov forward equation $\boldsymbol{P}'(t) = \boldsymbol{P}(t)\boldsymbol{Q}$ gives, for $i = 0, 1$.

$$\begin{aligned}
p'_{i0}(t) &= -ap_{i0}(t) + bp_{i1}(t) \\
p'_{i1}(t) &= ap_{i0} - bp_{i1}(t).
\end{aligned}$$

Again,

$$p_{i0}(t) = 1 - p_{i1}(t)$$

Assume that $p_{00}(0) = 1$. Solving, we get

$$\begin{aligned}
p_{00}(t) &= \frac{b}{a+b} + \frac{a}{a+b}e^{-(a+b)t}, \\
p_{01}(t) &= \frac{a}{a+b} - \frac{a}{a+b}e^{-(a+b)t}, \\
p_{11}(t) &= \frac{a}{a+b} + \frac{b}{a+b}e^{-(a+b)t}, \\
\text{and} \quad p_{10}(t) &= \frac{b}{a+b} - \frac{b}{a+b}e^{-(a+b)t}.
\end{aligned}$$

As $t \to \infty$,

$$\begin{aligned}
p_{00}(t) &\to \frac{b}{a+b}, \quad p_{01}(t) \to \frac{a}{a+b} \quad \text{and} \\
p_{10}(t) &\to \frac{b}{a+b}, \quad p_{11}(t) \to \frac{a}{a+b}.
\end{aligned}$$

These limiting probabilities can also be obtained by using (1.3.21). We have, as $t \to \infty, \lim p_{ij}(t) = u_j$; then from (1.3.21) we get

$$\begin{aligned}
q_0 u_0 = u_1 q_{10} &\Rightarrow a u_0 = b u_1 \\
&\Rightarrow u_1 = \frac{a}{b} u_0
\end{aligned}$$

and since $u_0 + u_1 = 1$,

$$\lim_{t\to\infty} p_{i0}(t) = u_0 = \frac{b}{a+b}, \quad i = 0, 1$$

and $$\lim_{t\to\infty} p_{i1}(t) = \frac{a}{a+b}, \quad i = 0, 1.$$

1.4 Birth-and-Death Processes

The class of all continuous-time Markov chains has an important subclass formed by the birth-and-death processes. These processes are characterized by the property that whenever a transition occurs from one state to another, then this transition can be to a neighboring state only. Suppose that the state space is $S = \{0, 1, 2, \ldots, i, \ldots\}$, then transition, whenever it occurs from state i, can be only to a neighboring state $(i - 1)$ or $(i + 1)$.

A continuous-time Markov chain $\{X(t), t \in T\}$ with state space $S = \{0, 1, 2, \ldots\}$ and with rates

$$\begin{aligned} q_{i,i+1} &= \lambda_i \text{ (say)}, \quad i = 0, 1, \ldots, \\ q_{i,i-1} &= \mu_i \text{ (say)}, \quad i = 1, 2, \ldots, \\ q_{i,j} &= 0, \quad j \neq i \pm 1, \quad j \neq i, \quad i = 0, 1, \ldots, \quad \text{and} \\ q_i &= (\lambda_i + \mu_i), \quad i = 0, 1, \ldots, \quad \mu_0 = 0, \end{aligned}$$

is called

(i) a *pure birth process*, if $\mu_i = 0$ for $i = 1, 2, \ldots$,

(ii) a *pure death process*, if $\lambda_i = 0, i = 0, 1, \ldots$, and

(iii) a *birth-and-death-process* if some of the λ_i's and some of the μ_i''s are positive.

Using (1.3.12) we get the Chapman-Kolmogorov forward equations for the birth-and-death process.

For $i, j = 1, 2, \ldots$,

$$p'_{ij}(t) = -(\lambda_j + \mu_j)p_{ij}(t) + \lambda_{j-1}p_{i,j-1}(t) + \mu_{j+1}p_{i,j+1}(t) \quad (1.4.1)$$

and $$p'_{i0}(t) = -\lambda_0 p_{i0}(t) + \mu_1 p_{i,1}(t). \quad (1.4.2)$$

The boundary conditions are

$$p_{i,j}(0+) = \delta_{ij}, \quad i, j = 0, 1, \ldots. \quad (1.4.3)$$

Denote

$$P_j(t) = Pr\{X(t) = j\}, \quad j = 0, 1, \ldots, t > 0$$

and assume that at time $t = 0$, the system starts at state i, so that

$$P_j(0) = Pr\{X(0) = j\} = \delta_{ij}, \tag{1.4.4}$$

then

$$P_j(t) = p_{ij}(t),$$

and the forward equations can be written as

$$P_j'(t) = -(\lambda_j + \mu_j)P_j(t) + \lambda_{j-1}P_{j-1}(t) + \mu_{j+1}P_{j+1}(t), \; j = 1, 2, \ldots, \tag{1.4.5}$$

$$P_0'(t) = -\lambda_0 P_0(t) + \mu_1 P_1(t). \tag{1.4.6}$$

Suppose that all the λ_i's and μ_i's are nonzero. Then the Markov chain is irreducible. It can be shown that such a chain is non-null persistent and that the limits

$$\lim_{t\to\infty} p_{ij}(t) = p_j$$

exist and are independent of the initial state i. Then Eqs. (1.4.5) and (1.4.6) become

$$0 = -(\lambda_j + \mu_j)p_j + \lambda_{j-1}p_{j-1} + \mu_{j+1}p_{j+1}, \quad j \geq 1 \tag{1.4.7}$$

$$0 = -\lambda_0 p_0 + \mu_1 p_1. \tag{1.4.8}$$

Define

$$\pi_j = \frac{\lambda_0\lambda_1 \ldots \lambda_{j-1}}{\mu_1\mu_2 \ldots \mu_j}, \quad j \geq 1, \quad \text{and}$$

$$\pi_0 = 1; \tag{1.4.9}$$

then the solution of the above can be obtained by induction. We have from (1.4.8)

$$p_1 = \left(\frac{\lambda_0}{\mu_1}\right) p_0 = \pi_1 p_0$$

and assuming $p_k = \pi_k p_0, k = 1, 2, \ldots, j$, we get from (1.4.7)

$$p_{j+1}\mu_{j+1} = \lambda_i \pi_j p_0, \quad \text{or}$$

$$p_{j+1} = \pi_{j+1} p_0.$$

Thus, if $\sum_{k=0}^{\infty} \pi_k < \infty$, then

$$p_j = \frac{\pi_j}{\sum \pi_k}, \quad j \geq 0. \tag{1.4.10}$$

Incidentally, $\sum \pi_k < \infty$ is a sufficient condition for the birth-and-death process to have all the states non-null persistent (and therefore for the process to be ergodic).

This process is of particular interest in queueing theory as several queueing systems can be modeled as birth-and-death processes. As an example, we consider the simple queue.

1.4.1 Special case: M/M/1 queue

For this queueing model

$$\lambda_j = \lambda, \quad i = 0, 1, 2, \dots \quad \text{and}$$

$$\mu_i = \mu, \quad i = 1, 2, \dots, \mu_0 = 0;$$

then $\pi_j = (\lambda/\mu)^j$ and $\sum \pi_k < \infty$ *iff* $(\lambda/\mu) < 1$, and then

$$\sum \pi_k = 1/[1 - (\lambda/\mu)]$$

$$p_j = [1 - (\lambda/\mu)](\lambda/\mu)^j, \quad j = 0, 1, 2, \dots \tag{1.4.11}$$

1.4.2 Pure birth process: Yule-Furry process

If $\mu_i = 0, i \geq 0$, then we get a pure birth process; further, if $\lambda_i = i\lambda$, for all i, we get the Yule-Furry process for which

$$P_j'(t) = -j\lambda P_j(t) + (j-1)\lambda P_{j-1}(t), \quad j \geq 1, \quad \text{and}$$

$$P_0'(t) = 0 \tag{1.4.12}$$

1.5 Poisson Process

If $\mu_i = 0, i \geq 0, \lambda_i = \lambda$ for all i, then we get what is known as the homogeneous Poisson process with parameter λ. It is a pure birth process with constant rate λ.

The Poisson process can be used as a model of a large class of stochastic phenomena and is thus extremely useful from the point of view of application.

The Chapman-Kolmogorov forward equations are

$$P_j'(t) = -\lambda[P_j(t) - P_{j-1}(t)], \quad j \geq 1 \tag{1.5.1a}$$

$$P_0'(t) = -\lambda P_0(t). \tag{1.5.1b}$$

Let the boundary condition be

$$P_j(0) = \delta_{ij}.$$

The Eq. (1.5.1a) can be solved in a number of ways. Let us consider the method of generating function. Define

$$P(s,t)=\sum_{j=0}^{\infty} P_j(t)s^j, \tag{1.5.2}$$

(when the RHS converges); then

$$P(s,0)=s^i. \tag{1.5.3}$$

Assuming the validity of term-by-term differentiation, we get from (1.5.2)

$$\begin{aligned}\frac{\partial}{\partial t}P(s,t) &= \sum_{j=0}^{\infty}\frac{\partial}{\partial t}\{P_j(t)\}s^j\\ &= P_0'(t)+\sum_{j=1}^{\infty}P_j'(t)s^j.\end{aligned}$$

Multiplying (1.5.1a) by s^j and adding over $j=1,2,3,\dots$, we get

$$\frac{\partial}{\partial t}P(s,t)-P_0'(t)=-\lambda[P(s,t)-P_0(t)-sP(s,t)].$$

Using (1.5.1b), we have

$$\frac{\partial}{\partial t}P(s,t)=P(s,t)[\lambda(s-1)].$$

Solving, we get

$$\begin{aligned}P(s,t) &= Ce^{\lambda(s-1)t}\\ &= s^i e^{\lambda(s-1)t};\end{aligned} \tag{1.5.3a}$$

whence

$$\begin{aligned}P_j(t) &= \text{coeff. of } s^j \text{ in } \quad P(s,t)\\ &= e^{-\lambda t}\frac{(\lambda t)^{j-i}}{(j-i)!}, \quad j=i,i+1,\dots,\\ &= 0, \quad j=0,1,\dots,i-1.\end{aligned} \tag{1.5.4}$$

Since the Poisson process is a Markov chain $\{X(t), t\in(0,\infty)\}$ with stationary transition probabilities, we have

$$\begin{aligned}Pr\{X(t+s)-X(s)=k \mid X(s)=i\} &= Pr\{X(t+s)=i+k \mid X(s)=i\}\\ &= \frac{(\lambda t)^k}{k!}e^{-\lambda t}, \qquad i,k=0,1,\dots;\ t,s\ge 0.\end{aligned} \tag{1.5.5}$$

We have defined the Poisson process as a birth process with constant birth rate. It can be introduced as a renewal process (as we shall see later in Section 1.7). A third way of defining the Poisson process is given below.

Let $N(t)$ denote the number of occurrences of a specified event in an interval of length t (i.e., during the time period, say, from 0 to t). Let

$$P_n(t) = Pr\{N(t) = n\}, \quad n = 0, 1, \ldots$$

We make the following postulates:

(1) *Independence.* The number of events occurring in two disjoint intervals of time are independent—that is, if $t_0 < t_1 < t_2, \ldots$, then the increments $N(t_1) - N(t_0), N(t_2) - N(t_1), \ldots$ are independent RVs.

(2) *Homogeneity in time.* The RV $\{N(t+s) - N(s)\}$ depends on the length of the interval $(t+s) - s = t$ and not on s or on the value of $N(s)$.

(3) *Regularity or orderliness.* In an interval of infinitesimal length h, the probability of *exactly one* occurrence is

$$P_1(h) = \lambda h + o(h)$$

and the probability of two or more occurrences is

$$\sum_{k=2}^{\infty} P_k(h) = o(h).$$

It follows that $P_0(h) = 1 - \lambda h + o(h)$. From the assumption of independence, we get

$$P_0(t+h) = P_0(t)P_0(h) = P_0(t)[1 - \lambda h + o(h)]$$

so that

$$\lim_{h \to o} \frac{P_0(t+h) - P_0(t)}{h} = -\lambda P_0(t) + \lim_{h \to 0} \frac{o(h)}{h}$$

or

$$P_0'(t) = -\lambda P_0(t).$$

We have

$$\begin{aligned} P_j(t+h) &= P_j(t)P_0(h) + P_{j-1}(t)P_1(h) + \sum_{r=2}^{\infty} P_{j-r}(t)P_r(h) \\ &= P_j(t)[1 - \lambda h + o(h)] + P_{j-1}(t)[\lambda h + o(h)] + o(h) \end{aligned}$$

so that

$$\lim_{h \to 0} \frac{P_j(t+h) - P_j(t)}{h} = -\lambda P_j(t) + \lambda P_{j-1}(t) + \lim_{h \to 0} \frac{o(h)}{h}$$

$$\text{or} \qquad P_j'(t) = -\lambda[P_j(t) - P_{j-1}(t)], \quad j \geq 1.$$

Thus, we get the same Chapman-Kolmogorov equations as given in (1.5.1a).

If $P_j(0) = \delta_{ij}$, then $P_j(t)$ is given by (1.5.4). When $P_j(0) = \delta_{0j}$ we get

$$P_j(t) = e^{-\lambda t}\frac{(\lambda t)^j}{j!}, \quad j = 0, 1, 2, \ldots \tag{1.5.6}$$

Thus, $N(t)$ follows Poisson distribution with parameter λt—that is, $\{N(t), t \geq 0\}$ is a Poisson process with parameter λ (or rate λ). We have $E[N(t)] = \lambda t$ and var $[N(t)] = \lambda t$. We shall state below some important properties of the Poisson process. For proof see works on stochastic processes such as Karlin and Taylor (1975), Medhi (1994), and Ross (1980, 1983). For an account of the Poisson process, see Kingman (1993); also Serfozo (1990).

1.5.1 Properties of the Poisson process

(1) *Additive Property.* Sum of n independent Poisson processes with parameter $\lambda_i, i = 1, 2, \ldots, n$ is a Poisson process with parameter $\lambda_1 + \lambda_2 + \cdots + \lambda_n$.

(2) *Decomposition Property.* Suppose that $N(t)$ is the number of occurrences of a specified event and that $\{N(t), t \geq 0\}$ is a Poisson process with parameter λ. Suppose further that each occurrence of the event has a probability p of being recorded, and that recording of an occurrence is independent of other occurrences and also of $N(t)$. If $M(t)$ is the number of occurrences so recorded, then $\{M(t), t \geq 0\}$ is also a Poisson process with parameter λp; if $M_1(t)$ is the number of occurrences not recorded, then $\{M_1(t), t \geq 0\}$ is a Poisson process with parameter $\lambda(1-p)$. Further, $\{M(t), t \geq 0\}$ and $\{M_1(t), t \geq 0\}$ are independent.
The above implies that a *random selection* of a Poisson process yields a Poisson process.
In fact, a Poisson process can be decomposed into any number of independent Poisson processes—that is, a Poisson process is *infinitely divisible.*

(3) *Interarrival Times.* The interarrival times (i.e., the intervals) between two successive occurrences of a Poisson process with parameter λ are IID RVs that are exponential with mean $1/\lambda$.

(4) *Memoryless Property of Exponential Distribution.* Exponential distribution possesses what is known as a *memoryless* or *Markovian* property and is the only continuous distribution to possess this property. It may be stated as follows. Suppose that X has exponential distribution with mean $1/\lambda$; then

$$Pr\{X \geq x + y \mid X \geq x\} = Pr\{X \geq y\}$$

is independent of x, for the LHS equals

$$\begin{aligned}\frac{Pr\{X \geq x + y \text{ and } X \geq x\}}{Pr\{X \geq x\}} &= \frac{Pr\{X \geq x + y\}}{Pr\{X \geq x\}} \\ &= \frac{e^{-\lambda(x+y)}}{e^{-\lambda x}} \\ &= e^{-\lambda y} = Pr\{X \geq y\}.\end{aligned}$$

If the interval between two occurrences is exponentially distributed, then the memoryless property implies that the interval to the next occurrence is statistically independent of the time from the last occurrence and has exponential distribution with the same mean.

If τ is an arbitrary epoch in the interval (t_i, t_{i+1}) between the ith and $(i+1)$th occurrences of a Poisson process with parameter λ, then the distribution of the interval $(t_{i+1} - \tau)$ is independent of the elapsed time $(\tau - t_i)$ since the last occurrence and is exponential with mean $1/\lambda$.

In the queueing context, arrivals (or service completions) may be taken as occurrences; so in the case of the Poisson-exponential process, the above remarkable property leads to easily tractable and mathematically agreeable results.

(5) *Randomness Property.* Given that exactly one event of a Poisson process $\{N(t), t \geq 0\}$ has occurred by epoch T, then the time interval γ in $[0, T]$ in which the event occurred has uniform distribution in $[0, T]$. In other words,

$$Pr\{t < \gamma \leq t + dt \mid N(T) = 1\} = \frac{dt}{T}, \quad 0 < t < T.$$

This is also expressed by saying that an event of a Poisson process is a purely *random* event. The Poisson process is sometimes called a completely random process.

The preceding result holds in a more general case. This is stated below.

If an interval of length T contains exactly m occurrences of a Poisson process, then the joint distribution of the epochs at which these events occurred is that of m points uniformly distributed over an interval of length T. The result holds in case of the (more general) birth process of which the Poisson process forms a special class.

1.5.2 Generalization of the Poisson process

There are several directions in which the classical Poisson process can be generalized.

1.5.2.1 Poisson cluster process (compound Poisson process)

One of the postulates of the Poisson process is that at most one event can occur at a time. Now suppose that several events (i.e., a cluster of events) can occur simultaneously at an epoch of occurrence of a Poisson process $N(t)$ and that the number of events X_i in the ith cluster is a RV, X_is having independent and identical distributions

$$Pr\{X_i = j\} = p_j, \quad j = 1, 2, \ldots.$$

Then $M(t)$, the total number of events in an interval of length t, is given by

$$M(t) = \sum_{i=1}^{N(t)} X_i.$$

The stochastic process $\{M(t), t \geq 0\}$ is called a *compound Poisson process.* Its PGF is given by

$$G[P(s)] = \exp\{\lambda t[P(s) - 1]\},$$

where $P(s)$ is the PGF of X_i and $G(s)$ is the PGF of $N(t)$. We have

$$\begin{aligned} Pr\{M(t) = m\} &= \sum_{k=0}^{m} [Pr\{N(t) = k\} Pr\{X_i = m\}] \\ &= \sum_{k=0}^{m} e^{-\lambda t} \frac{(\lambda t)^k}{k!} p_m^{k*}, \end{aligned}$$

where p_m^{k*} is the probability associated with a k-fold convolution of X_i with itself.

We have

$$\begin{aligned} E\{M(t)\} &= \lambda t E\{X_i\} \quad \text{and} \\ \text{var}\{M(t)\} &= \lambda t E\left\{X_i^2\right\}. \end{aligned}$$

The compound Poisson process is useful in modeling queueing systems with batch arrival/batch service, exponential interarrival/service time, and independent and identical batch-sized distribution.

1.5.2.2 Nonhomogeneous Poisson process

The parameter λ in the classical Poisson process is assumed to be a constant, independent of time. Generalizations of the Poisson process arise when λ is assumed to be (i) a nonrandom function of time $\lambda(t)$ and (ii) a random variable.

Here it is assumed that the probability that arrival occurs between time t and time $t + \Delta t$, given that n arrivals occurred by time t, is equal to $\lambda(t) \Delta t + o(\Delta t)$, while the probability that more than one arrival occurs is $o(\Delta t)$. The resulting process is the so-called nonhomogeneous Poisson process $\{N(t), t \geq 0\}$. It can be shown that

$$\begin{aligned} p_n(t) &= Pr\{N(t) = n\} \\ &= \exp\left\{-\int_0^t \lambda(x) dx\right\} \frac{\left[\int_0^t \lambda(x) dx\right]^n}{n!}, \quad n \geq 0. \end{aligned}$$

1.5.2.3 Random variation of parameter

Here we assume that λ is a random variable having PDF $f(\lambda), 0 \leq \lambda \leq \infty$. Thus,

$$p_n(t) = Pr\{N(t) = n\} = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^n}{n!} f(\lambda) d\lambda.$$

The case when the parameter λ of a Poisson process is a random function of time $\lambda(t)$ (and so is itself a stochastic process) leads to a doubly stochastic Poisson process. There are several situations where such generalizations of Poisson process may be realistic.

1.5.2.4 Truncated process

A simple generalization is truncation of the infinite domain of the Poisson process. This case arises in modeling a queueing system with waiting space limited to n; so arrivals that occur when the waiting space is full are not permitted and are lost to the system. This will be involved only in scaling the Poisson probabilities by a suitable scale factor.

1.5.3 Role of the Poisson process in probability models

The Poisson process and its associated exponential distribution possess many agreeable properties that lead to mathematically tractable results when used in probability models. Its importance is also due to the fact that occurrences of events in many real-life situations do obey the postulates of the Poisson process, and thus its use in probability modeling is considered realistic. An arrival process to a queueing system is often taken to be Poisson.

Consider an event and an interval of time during which the occurrences of the event happen. Suppose that the interval is subdivided into a large number of subintervals (say, n), and that p_i is the probability of occurrence of the event in the ith subinterval. Suppose further that the events occur independently of one another and that $\lambda = p_1 + \cdots + p_n$, while the largest of p_i tends to 0. Then the number of occurrences of the event in the interval tends in the limit to a Poisson distribution with mean λ. The Poisson distribution thus gives an adequate description of the cumulative effect of a large number of events, such that occurrence of an event in a small subinterval is improbable.

There are other contexts arising out of extreme value theory as well as information theory that provide justification of using the Poisson process in modeling.

Note: Rego and Szpankowski (1989) show that there is an equivalence between using entropy maximization with a two-moment constraint and assumption of exponential distribution in a certain queueing context.

1.6 Randomization: Derived Markov Chains

Let $\{X(t), t \geq 0\}$ be a continuous-time Markov chain with transition matrix $\boldsymbol{Q}$ and countable state space S. Assume that $X(t)$ is *uniformizable*—that is, the diagonal elements of $\boldsymbol{Q}$ are uniformly bounded. Let

$$\alpha = \sup_i q_i < \infty.$$

Then there exists a discrete-time Markov chain $\{Y_n, n = 0, 1, \ldots\}$ with state space S and TPM $P = (p_{ij})$ such that

$$\boldsymbol{P} = \frac{Q}{\lambda} + I, \tag{1.6.1}$$

where λ is any real number not less than α. Since $P(t) = e^{Qt}$(Eqn. 1.3.14), we have

$$\begin{aligned} P(t) &= e^{Qt} \\ &= e^{\lambda(P-I)t} \\ &= e^{-\lambda t} e^{\lambda P t} \\ &= e^{-\lambda t} \sum_{n=0}^{\infty} \frac{\lambda^n t^n}{n!} P^n, \end{aligned}$$

so that elementwise

$$p_{ij}(t) = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{\lambda^n t^n}{n!} p_{ij}^{(n)}, \quad t \geq 0, \quad i, j \in S. \tag{1.6.2}$$

We shall have from the above

$$\pi(t) = \pi(0) P(t) = \pi(0) e^{-\lambda t} \sum_{n=0}^{\infty} \frac{\lambda^n t^n}{n!} P^n \tag{1.6.3a}$$

or elementwise

$$\pi_j(t) = \pi(0) e^{-\lambda t} \sum_{n=0}^{\infty} \frac{\lambda^n t^n}{n!} p_{ij}^{(n)}. \tag{1.6.3b}$$

Another interesting fact is that there exists a Poisson process $\{N(t), t \geq 0\}$ with parameter λ such that Y_n and $N(t)$ are independent and that $\{X(t), t \geq 0\}$ and $\{Y_{N(t)}, t \geq 0\}$ are probabilistically identical—that is, we can write

$$X(t) \equiv Y_{N(t)}.$$

The converse also holds: if $X(t) = Y_{N(t)}$, then $Q = \lambda(P - I)$.

1.6.1 Markov chain on an underlying Poisson process (or subordinated to a Poisson process)

The above method of construction leads from a Markov process $\{X(t), t \geq 0\}$ to a derived Markov chain $\{Y_{N(t)}, t \geq 0\}$ by randomization of operational time through events of a Poisson process. For, we can obtain $p_{ij}(t)$ in terms of $p_{ij}^{(n)}$ by conditioning over the number of occurrences of the Poisson process $N(t)$ in $(0, t)$. Conditioning over the number of occurrences of the Poisson process with parameter λ over $[0, 1]$, we get

$$\begin{aligned} p_{ij}(t) &= Pr\{X(t) = j \mid X(0) = i\} \\ &= \sum_{n=0}^{\infty} Pr\{X(t) = j \mid X(0) = i, N(t) = n\} \\ &\quad \times Pr\{N(t) = n \mid X(0) = i\}. \end{aligned}$$

Now

$$Pr\{N(t) = n \mid X(0) = i\} = e^{-\lambda t}\left[\frac{(\lambda t)^n}{n!}\right]$$

and $Pr\{X(t) = j \mid X(0) = i, N(t) = n\}$ is the probability that the system goes from state i to state j in time t during which n Poisson occurrences took place. (That is, n transitions took place. Here, the time interval t is replaced by number of transitions). Thus,

$$Pr\{X(t) = j \mid X(0) = i, N(t) = n\} = p_{ij}^{(n)}.$$

Hence, we have

$$p_{ij}(t) = e^{-\lambda t}\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!}p_{ij}^{(n)}. \tag{1.6.4}$$

1.6.2 Equivalence of the two limiting forms

Let $\{Y_n, n \geq 0\}$ be an irreducible and aperiodic chain with finite state space S and TPM P. Then from the ergodic theorem (Theorem 1.1) we get that

$$\lim_{n\to\infty} p_{ij}^{(n)} = v_j, \quad i, j \in S$$

exists and is independent of i, and $V = \{v_1, v_2, \ldots\}$ is the invariant distribution given by

$$VP = V, \qquad Ve = 1. \tag{1.6.5}$$

The Markov process $\{X(t) = Y_{N(t)}, t \geq 0\}$ is also aperiodic and irreducible and has the same state space S. From the ergodic theorem (Theorem 1.4), we get that

$$\lim_{t\to\infty} p_{ij}(t) = u_j$$

exists and is independent of i. Further, $\boldsymbol{U} = \{u_1, u_2, \ldots\}$ is a probability vector and $\boldsymbol{U}$ is given as the solution of

$$\begin{aligned} \boldsymbol{UQ} &= \boldsymbol{0}, \quad \boldsymbol{Ue} = 1, \\ \boldsymbol{UQ} &= 0 \leftrightarrow \boldsymbol{U}[\lambda(P - I)] = \boldsymbol{0} \\ &\leftrightarrow \boldsymbol{UP} = \boldsymbol{U}. \end{aligned} \tag{1.6.6}$$

Thus, from (1.6.5) and (1.6.6), we get

$$\boldsymbol{U} \equiv \boldsymbol{V}.$$

In other words,

$$\lim_{t\to\infty} p_{ij}(t) = \lim_{n\to\infty} p_{ij}^{(n)}, \quad i, j \in S. \tag{1.6.7}$$

1.6.3 Numerical method

The numerical method is a subject in itself. We discuss the importance of the randomization technique in numerical analysis. This method of construction gives very useful formulas for computation of $p_{ij}(t)$ or $\pi_j(t)$—that is, transient probabilities of a uniformizable Markov process $\{X(t), t \geq 0\}$. Ross (1980) calls this method *uniformization*, though *randomization* appears to be a more generally used term. What is generally done in computational work is to choose a truncation point N and to set N to bound the error of truncation ε as follows. From (1.6.2),

$$p_{ij}(t) = \sum_{n=0}^{N} e^{-\lambda t}\frac{(\lambda t)^n}{n!} p_{ij}^{(n)} + \sum_{n=N+1}^{\infty} e^{-\lambda t}\frac{(\lambda t)^n}{n!} p_{ij}^{(n)}, \tag{1.6.8}$$

where N is so chosen that the second term is less than or equal to the desired control error ε. It would ensure that $p_{ij}(t)$ would be accurate to within ε. The same holds for the computation of $\pi_j(t) = Pr\{X(t) = j\}$.

Algorithms for computation have been developed by Grassman (1977) and Gross and Miller (1984a,b). These algorithms have been shown to be useful for computation of transient probabilities of many stochastic systems such as queueing, inventory, reliability, and maintenance systems. (Refer to Gross and Miller (1984a) for their SERT algorithm).

1.7 Renewal Processes

1.7.1 Introduction

We noted that the interarrival (or interoccurrence) times between successive events of a Poisson process are IID exponential random variables. A possible generalization is obtained by removing the restriction of exponential distribution and by considering that the interarrival times are IID random variables with an arbitrary distribution. The resulting process is called a renewal process.

Definition 1.3. Let X_n be the interval between the $(n-1)th$ and nth events of a counting process $\{N(t), t \geq 0\}$. Let $\{X_n, n = 1, 2, \ldots\}$ be a sequence of nonnegative IID random variables having distribution function F with mean μ. Then $\{N(t), t \geq 0\}$ is said to be a *renewal process* generated or induced by the distribution F. Assume that $N(t)$ is independent of X_i.

The discrete time process $\{X_n, n = 1, 2, \ldots\}$ also represents the same renewal process. Let

$$S_0 = 0, \qquad S_n = X_1 + \cdots + X_n, \quad n \geq 1.$$

Then

$$N(t) = \sup\{n : S_n \leq t\}. \tag{1.7.1}$$

If $S_n = t$ for some n, then a renewal is said to occur at time t. Thus, S_n gives the epoch of nth renewal. We have $F_n(x) = Pr\{S_n \leq x\}$, and $F_n = F^{n*}$ where F^{n*} is the n-fold convolution of F with itself. Assume that $E\{X_i\} = \mu$ exists and is finite. The function $M(t) = E\{N(t)\}$ is called the *renewal function* (which is a nonrandom function of t). When it exists, the derivative $M'(t) = m(t)$ is called the *renewal density* (*not* a PDF). The distribution of $N(t)$ is given by

$$\begin{aligned} p_n(t) &= Pr\{N(t) = n\} \\ &= Pr\{N(t) \geq n\} - Pr\{N(t) \geq (n+1)\} \\ &= Pr\{S_n \leq t\} - Pr\{S_{n+1} \leq t\} \\ &= F_n(t) - F_{n+1}(t). \end{aligned} \tag{1.7.2}$$

It can be easily verified that for X_n exponential, $\{N(t), t \geq 0\}$ is a Poisson process. The average number of renewals by time t equals

$$\begin{aligned} M(t) &= \sum_{n=0}^{\infty} n p_n(t) \\ &= \sum_{n=1}^{\infty} F_n(t) = \sum_{n=1}^{\infty} F^{n*}(t) \\ &= F(t) + \sum_{n=1}^{\infty} F^{(n+1)*}(t). \end{aligned} \tag{1.7.2a}$$

Now,

$$\sum_{n=1}^{\infty} F^{(n+1)*}(t) = \sum_{n=1}^{\infty} \int_0^t F^{n*}(t-x)dF(x)$$
$$= \int_0^t \left\{ \sum_{n=1}^{\infty} F^{n*}(t-x) \right\} dF(x)$$

assuming the validity of interchange of summation and integration operations. Thus,

$$M(t) = F(t) + \int_0^t M(t-x)dF(x). \tag{1.7.3}$$

The above is known as the *fundamental equation of renewal theory.*

Renewal theorems involving limiting behavior of $M(t)$ are interesting as well as important from the point of view of applications. (For details refer to any work on stochastic processes, such as Çinlar (1975), Karlin and Taylor (1975), Medhi (1994), and Ross (1983).)

1.7.2 Residual and excess lifetimes

We discuss below two RVs that arise in several situations. To a given $t > 0$, there corresponds a unique $N(t)$ such that

$$S_{N(t)} \le t < S_{N(t)+1}, \tag{1.7.4}$$

that is, t lies in the interval $X_{N(t)+1}$ between $\{N(t)\}$th and $\{N(t)+1\}$th renewals.

The RV $Y(t) = S_{N(t)+1} - t$ (which is the interval between t and the renewal epoch after t) is called the *residual lifetime* or *forward-recurrence time* at t.

The RV $Z(t) = t - S_{N(t)}$ (which is the interval between t and the last renewal epoch before t) is called the *spent lifetime* or *excess lifetime* or *backward-recurrence time* at t.

Note that

$$Y(t) + Z(t) = S_{N(t)+1} - S_{N(t)} = X_{N(t)+1} \tag{1.7.5}$$

is the total life.

These RVs arise in various queueing contexts. The RV $Z(t)$ denotes the elapsed time between t and the last arrival before t or between t and the commencement of the last service before t depending on whether X_i denotes the interarrival or service time. Similarly, $Y(t)$ can be interpreted. We consider now the distribution of $Y(t)$ and $Z(t)$. We have

$$Pr\{Y(t) \le x\} = F(t+x) - \int_0^t [1 - F(t+x-y)]dM(y), \quad x > 0$$
$$= 0, \quad x \le 0. \tag{1.7.6}$$

If F is not a lattice distribution, then the limiting distribution Y of $Y(t)$ is given by

$$Pr\{Y \le x\} = \lim_{t\to\infty} Pr\{Y(t) \le x\} = \frac{1}{\mu}\int_0^x [1 - F(y)]dy, \quad x \ge 0. \quad (1.7.7)$$

Again,

$$Pr\{Z(t) \le x\} = \begin{cases} 0, & x \le 0, \\ F(t) - \int_0^{t-x} [1 - F(t-y)]dM(y), & 0 < x \le t, \\ 1, & x > t, \end{cases} \quad (1.7.8)$$

and if F is not a lattice distribution, then the limiting distribution Z of $Z(t)$ is given by

$$\begin{aligned} Pr\{Z \le x\} &= \lim_{t\to\infty} Pr\{Z(t) \le x\} \\ &= \frac{1}{\mu}\int_0^x [1 - F(y)]dy, \quad x \ge 0 \\ &= 0, x < 0. \end{aligned} \quad (1.7.9)$$

Assumption of finite mean μ ensures that the above is a proper distribution (in both the cases). Z is also called stationary excess distribution.

When these exist, the two limiting distributions Y and Z are identical. It can be easily verified that for exponential X_i, the distributions of $Y(t)$ and $Z(t)$ are again exponential with the same mean $\mu = E(X_i)$.

Suppose that $m_r = E(X_i^r)$ exist for $r = 1, 2$. Then

$$E\{Y\} = E\{Z\} = \frac{m_2}{2\mu} = \frac{m_2}{2m_1}. \quad (1.7.10)$$

If F is a lattice distribution, then the distributions of $Y(t)$ and $Z(t)$ have no limits for $t \to \infty$ except in some special cases.

Note: Higher moments of Y (or Z) can be found in terms of those of X (see Problem 1.26).

1.8 Regenerative Processes

Let $\{X(t), t \ge 0\}$ be a stochastic process with countable state space $S = \{0, 1, 2, \ldots\}$. Suppose that there exists an epoch t_1 such that the continuation of the process beyond t_1 is a probabilistic replica of the whole process starting at $0(=t_0)$. Then this implies the existence of epochs $t_2, t_3, \ldots (t_i > t_{i-1})$

having the same property. Such a process is known as a regenerative process. If $T_n = t_n - t_{n-1}, n = 1, 2, \ldots$, then $\{T_n, n = 1, 2, \ldots\}$ is a renewal process.

A renewal process is regenerative, with T_i representing the time of the ith renewal.

Another example of a regenerative process is provided by what is known as an *alternating renewal process.* Such a process can be envisaged by considering that a system can be in one of two possible states—say, 0 and 1—that is, having $S = \{0, 1\}$. Initially, it is at state 0 and remains at that state for a time Y_1, and then a change of state to state 1 occurs in which it remains for a time Z_1, after which it again goes to state 0 for a time Y_2 and then goes to state 1 for a time Z_2 and so on. That is, its movement could be denoted by $0 \to 1 \to 0 \to 1 \ldots$. The initial state could be 1, in which case the movement could be denoted by $1 \to 0 \to 1 \to 0 \ldots$.

Suppose that $\{Y_n\}, \{Z_n\}$ are two sequences of IID random variables and that Y_n and Z_n need not be independent. Let

$$T_n = Y_n + Z_n, \quad n = 1, 2, \ldots.$$

Then at time T_1 the process restarts itself, and so also at times $T_2, T_3 \ldots$. The interval T_n denotes a complete cycle, and the process restarts itself after each complete cycle. Let

$$E\{Y_n\} = E\{Y\}, \qquad E\{Z_n\} = E\{Z\}.$$

Then the long-run proportions of time that the system is at states 0 and 1 are given, respectively, by

$$p_0 = \lim_{t\to\infty} Pr\{X(t) = 0\} = \frac{E\{Y\}}{E\{Y\} + E\{Z\}} \tag{1.8.1}$$

$$\text{and} \quad p_1 = \lim_{t\to\infty} Pr\{X(t) = 1\} = \frac{E\{Z\}}{E\{Y\} + E\{Z\}}$$
$$= 1 - p_0. \tag{1.8.2}$$

1.8.1 Application in queueing theory

The results (1.8.1) and (1.8.2) have an important application in queueing theory. Consider a single-server queueing system such that an arriving customer is immediately taken for service if the server is free, but joins a waiting line if the server is busy. The system can be considered to be in two states (idle or busy) according to whether the server is idle or busy. The idle and busy states alternate and together constitute a cycle of an alternating renewal process. A busy period starts as soon as a customer arrives before an idle server and ends at the instant when the server becomes free for the first time.

The epochs of commencement of busy periods are regeneration points. Let I_n and B_n denote the lengths of nth idle and busy periods, respectively, and let

$$E\{I_n\} = E\{I\} \quad \text{and}$$
$$E\{B_n\} = E\{B\}. \tag{1.8.3}$$

Then the long-run proportion of time that the server is idle equals

$$p_0 = \frac{E\{I\}}{E\{I\} + E\{B\}}, \tag{1.8.4}$$

and the long-run proportion of time that the server is busy equals

$$p_1 = \frac{E\{B\}}{E\{I\} + E\{B\}}. \tag{1.8.5}$$

In particular, if the arrival process is Poisson with mean λt, then it follows (from its lack of memory property) that an idle period is exponentially distributed with mean $1/\lambda$—that is, $E(I) = 1/\lambda$. Then when p_0 or p_1 is known, $E(B)$ can be found.

The case of the alternating renewal process can be generalized to cover cyclical movement of more than two states. Suppose that the state space of the process $\{X(t), t \geq 0\}$ is $S = \{0, 1, \ldots, m\}$ and its movement from initial state 0 is cyclic as $0 \to 1 \to 2 \cdots m \to 0 \ldots$, and that τ_k is the duration of sojourn at state k, having mean $\mu_k = E\{\tau_k\}, k = 0, 1, \ldots, m$. Then

$$p_k = \lim_{t\to\infty} Pr\{X(t) = k\} = \frac{\mu_k}{\sum_{i=0}^{m} \mu_i}, \quad k = 0, 1, \ldots, m. \tag{1.8.6}$$

1.9 Markov Renewal Processes and Semi-Markov Processes

We shall now consider a kind of generalization of a Markov process as well as a renewal process. Let $\{X(t), t \geq 0\}$ be a Markov process with discrete countable state $S = \{0, 1, 2, \ldots\}$, and let $t_0 = 0, t_1, t_2 \cdots (t_i < t_{i+1})$ be the epochs at which transitions occur. The sequence $\{X_n = X(t_n + 0), n \geq 0\}$ forms a Markov chain, and the transition intervals $T_n = t_n - t_{n-1}, n = 1, 2, \ldots$ are distributed as independent exponential variables having means that may depend on the state of X_n.

We generalize the situation as follows: Suppose that the transitions $\{X_n, n \geq 0\}$ of the process $\{X(t), t \geq 0\}$ constitute a Markov chain but the transition intervals $T_n, n = 0, 1, \ldots$ have an independent arbitrary distribution and that the mean may depend not only on the state of X_n but also on the state of X_{n+1}.

The process $\{X(t), t \geq 0\}$ is then no longer Markovian. The two-dimensional process $\{X_n, t_n n \geq 0\}$ is called a Markov renewal process with state space S. Here

$$\begin{aligned} Pr\{X_{n+1} &= j, T_{n+1} \leq t \mid X_0 = x_0, \ldots, X_n = i, T_0, T_1, \ldots, T_n\} \\ &= Pr\{X_{n+1} = j, T_{n+1} \leq t \mid X_n = i\} \\ &= Q_{ij}(t), \quad \text{say } i, j \in S. \end{aligned} \tag{1.9.1}$$

Let

$$p_{ij} = \lim_{t\to\infty} Q_{ij}(t) \quad \text{and} \quad F_{ij}(t) = \frac{Q_{ij}(t)}{p_{ij}} \quad \text{and} \tag{1.9.2}$$

$$Y(t) = X_n \quad \text{on} \quad t_n \leq t < t_{n+1}. \tag{1.9.3}$$

Then $\{Y(t), t \geq 0\}$ is called a *semi-Markov process*, and the Markov chain $\{X_n, n \geq 0\}$ is called the *embedded Markov* chain of $\{X(t), t \geq 0\}$. $Y(t)$ gives the state of the process at its most recent transition. The chain $\{X_n, n \geq 0\}$ has TMP (p_{ij}). $F_{ij}(t) = Pr\{T_{ij} \leq t\}$ is the distribution function of T_{ij}, the conditional transition time (or sojourn time) at state i given that the next transition is to state j. If τ_k is the unconditional waiting time at state k, then $\tau_k = \sum_j p_{ij} T_{ij}$.

For example, a pure birth process is a special type of Markov renewal process having

$$\begin{aligned} Q_{ij}(t) &= 1 - e^{a_i t}, \quad j = i + 1, \\ &= 0, \quad \text{otherwise.} \end{aligned}$$

Then

$$\begin{aligned} p_{ij} &= 1, \quad j = i + 1, \\ &= 0 \quad \text{otherwise, and} \end{aligned}$$

$$F_{ij}(t) = Q_{ij}(t), \qquad T_i = T_{ij}, \quad j = i + 1.$$

A Markov renewal process becomes a Markov process when the transition times are independent exponential and are independent of the next state visited. It becomes a Markov chain when the transition times are all identically equal to 1. It reduces to a renewal process if there is only one state and then only transition times becomes relevant. Semi-Markov processes are used in the study of certain queueing systems. Let $p_k = \lim_{t\to\infty} Pr\{Y(t) = k\}$ be the long-run proportion of time the semi-Markov process is at state k. Suppose that the embedded Markov chain $\{X_n, n = 0, 1, 2, \}$ is irreducible, aperiodic, and, if denumerable, recurrent nonnull. Then the limiting probabilities

$$v_j = \lim_{n\to\infty} p_{ij}^{(n)}$$

exist and are given as the unique non-negative solution of

$$v_j = \sum_{k \in s} v_k p_{kj}, \quad j \in S.$$

Then we shall have

$$p_k = \frac{v_k \mu_k}{\sum_{j \in S} v_j \mu_j}, \tag{1.9.4}$$

where $\mu_k = E\{\tau_k\}$ is the expected sojourn time in state k. One can get this result by extending the result (1.8.6) through an intuitive argument. For a formal proof, see Medhi (1994).

Problems

1.1. The transition probability matrix of a Markov chain with three states 0, 1, 2, is given by

$$\begin{pmatrix} 0.4 & 0.5 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.3 & 0.3 & 0.4 \end{pmatrix}$$

and the initial distribution is $(0.6, 0.3, 0.1)$. Find (i) $Pr(X_2 = 3)$, and (ii) $Pr\{X_3 = 1, X_2 = 0, X_1 = 2, X_0 = 0\}$. Find the invariant measure of the chain.

1.2. A chain with $S = \{1, 2, 3, \}$ has TPM

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ p_1 & p_2 & p_3 \end{pmatrix}, \quad p_i > 0, \ \sum p_i = 1.$$

Examine the nature of the states. Find P^n.

1.3. Find the invariant measure of a chain with $S = \{0, 1, 2, \ldots, m-1\}$ and a doubly stochastic transition probability matrix.

1.4. Consider a Markov chain with $S = \{1, 2, 3, 4\}$ and TPM

$$\begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Is the chain irreducible? Verify that states 1 and 2 are recurrent. Find μ_1 and μ_2. ($\mu_1 = 5/3, \mu_2 = 5/2$).

1.5. Show that for a Markov chain with a finite state space S, the probability of staying forever among the transient states is zero.

1.6. Show that if state j is transient, then

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty \quad \text{for all} \quad i \in S.$$

1.7. Show that a transient state cannot be reached from a persistent state.

1.8. Consider a service facility having a limited waiting space for m customers, including the one being serviced. The server serves one customer, if any, at epochs $0, 1, 2, \ldots$. Assume that the number of arrivals in the intervals $(k, k+1)$ is given by an IID random variable A with $Pr(A = n) = p_n, \sum p_n = 1$. Assume further that arrivals that occur when the waiting space is full leave the system and do not return. Denote by X_n the number of customers present at time n, including the one being served, if any. Show that $\{X_n, n \geq 0\}$ is a Markov chain and find its TPM.

1.9. In what is called a *gambler's ruin problem*, consider a gambler who with capital a agrees to play a series of games with an adversary having a capital $b(a + b = c$, the total capital). The probability of the gambler winning one game (and with it, one unit of money) is p, and that of losing one unit is $q = 1 - p$. (There is no draw.) Suppose that successive games are independent. If X_n is the gambler's fortune at time n (at time of the nth game), show that $\{X_n, n = 0, 1, 2, \ldots\}$ is a Markov chain. Write down its TPM. Is the chain irreducible? Examine the nature of the states of the chain.

1.10. Consider the Markov chain with $S = \{0, 1, \ldots, m\}$, such that

$$\begin{aligned} p_{0,0} &= q, \quad p_{01} = p, \\ p_{i,i-1} &= q, \quad p_{i,i+1} = p, \quad i = 1, 2, \ldots, m-1 \\ p_{m,m-1} &= q, \quad p_{m,m} = p, \quad 0 < p, q < 1, \end{aligned}$$

where $p + q = 1$. (Each of the transitions between other pairs of states has probability 0.) Show that the chain is irreducible and aperiodic. Find the limiting distribution V.

1.11. Consider the Markov chain of Example 1.3.

(a) Use (1.2.21) to find $\sum_j j v_j$.

(b) For $p_n = q^n p, n = 0, 1, 2, \ldots, p + q = 1$, examine the nonnull persistence of the chain. Find $V(s)$ in this case. Consider the particular case $q = \rho/(1 + \rho), \rho < 1$.

1.12. Suppose that customers arrive at a bank in accordance with a Poisson process at the rate of two per minute. Find the probability that the number of customers that arrive during a 10-minute period is (i) exactly 20, (ii) greater than 20, and (iii) between 10 and 20.

1.13. Suppose that customers arrive at a certain service facility center in accordance with a Poisson process $\{N(t), t \geq 0\}$ having parameter λ. A customer can make a preliminary inquiry as to whether he or she actually needs the facility. Suppose that the proportion of customers actually needing the service facility is $p(0 < p < 1)$. If $M(t)$ gives the number of customers actually needing service, show that $\{M(t), t \geq 0\}$ is again a Poisson process with parameter λp.

1.14. Suppose that customers arrive at a service counter in accordance with a Poisson process at a rate of three per minute. Find the probability that the interval between two successive arrivals is (i) between one and three minutes and (ii) less than one-third of a minute.

1.15. Suppose that messages arrive at a telephone switch board, the interarrival time of messages being exponential with mean 10 minutes. Find the probability that the number of messages received during the five afternoon hours (1–6 P.M.) is (i) exactly 24, (ii) more than 24, and (iii) nil.

1.16. If $X_i, i = 1, \ldots, n$ are IID exponential RVs with parameter a, then show that $S_n = X_1 + \cdots + X_n$ has gamma distribution with PDF

$$\begin{aligned} f_{a,n}(x) &= \frac{a^n x^{n-1} e^{-ax}}{\Gamma(x)}, \quad x > 0 \\ &= 0, \quad x \leq 0. \end{aligned}$$

Find $E\{S_n\}$ and var $\{S_n\}$.

1.17. Suppose that X and Y are independent exponential RVs with parameters a and b, respectively. Show that

(a) $W = X + Y$ has PDF

$$\begin{aligned} f(x) &= \frac{ab(e^{-ax} - e^{-bx})}{b - a}, \quad x > 0, \quad a \neq b \\ &= 0, \quad x \leq 0. \end{aligned}$$

(b) $Z = \min(X, Y)$ is exponential with parameter $a + b$.

(c) $M = \max(X, Y)$ has PDF

$$\begin{aligned} f(x) &= ae^{-ax} + be^{-bx} - (a + b)e^{-(a+b)x}, \quad x > 0 \\ &= 0, \quad x \leq 0. \end{aligned}$$

(d)

$$Pr\{X \leq Y\} = \frac{a}{(a+b)}.$$

1.18. If X is an exponential RV, then show that

$$E\{X \mid X > y\} = y + E\{X\} \quad \text{for all} \quad y > 0;$$

that is, $E\{X - y \mid X > y\} = E(X)$, independent of y.

1.19. A piece of equipment is subject to random shocks that occur in accordance with a Poisson process with rate λ. The equipment fails due to the cumulative effect of k shocks. Show that the duration of the lifetime T of the equipment has gamma distribution with PDF $f_{\lambda,k}(x)$. Note that T is the interval between k occurrences of a Poisson process.

1.20. Consider that two independent series of events A and B occur in accordance with Poisson processes with parameters a and b, respectively. Show that the number N of occurrences of the event A between two *successive* occurrences of the event B has geometric distribution with mass function

$$Pr\{N = n\} = (1-q)q^n, \qquad q = \frac{a}{(a+b)}, \qquad n = 0, 1, 2, \ldots$$

1.21. Consider a Poisson process with parameter λ. Given that n events happen by time t, show that the PDF of the time of occurrence T_k of the kth event $(k < n)$ is given by

$$\begin{aligned} f(x) &= \frac{n!}{(k-1)!(n-k)!} \frac{x^{k-1}}{t^k} \left(1 - \frac{x}{t}\right)^{n-k}, \quad 0 < x < t, \\ &= 0, \quad x \geq t. \end{aligned}$$

Show that $E(T_k) = kt/(n+1)$.

1.22. Suppose that a queueing system has m service channels. The demand for service arises in accordance with a Poisson process with rate a, and the service time distribution has exponential distribution with parameter b. Suppose that the service system has no storage facility—that is, a demand that arises when all m channels are busy is rejected and is lost to the system. Let $X(t)$ be the number of busy service channels (number of demands) at time t. Show that $\{X(t), t \geq 0\}$ is a continuous-time Markov chain. Determine the infinitesimal generator. (See problem 1.8 for the same queue with discrete service time.)

1.23. **The three-state process.** Suppose that an automatic machine can be in three states: working (state 0), failed in mode 1 (state 1), or failed in mode 2 (state 2). Suppose that a failed machine in either mode cannot go to another failed mode. (That is, transitions from state 1 to 2 and from state 2 to 1 are not possible.) Suppose that $X(t)$ denotes the state (condition) of the machine at time t and that the failed times are IID exponential with rates $a_i, i = 1, 2$, and the repair times are IID exponential with rates $b_i, i = 1, 2$. Show that $\{X(t), t \geq 0\}$ is a continuous-time Markov chain. Find the Q-matrix. (State any assumption that you make.)

1.24. Assume that the lifetime X of a device is random having DF $F(x)$. Show that the expected remaining life of a device aged y (which has already attained age y) is given by

$$E\{X - y \mid X > y\} = \frac{\int_y^{\infty}[1 - F(t)]dt}{1 - F(y)}.$$

In particular, for the exponential lifetime, this is equal to $E(X)$, independent of y. (See problem 1.18.)

1.25. Let $\{N(t), t \geq 0\}$ be a renewal process induced by a RV X. Show that for large t

$$E\{N(t)\} \simeq \frac{t}{E(X)} \quad \text{and}$$

$$\text{var}\{N(t)\} \simeq \frac{\text{var}(X)}{[E(X)]^3}t.$$

1.26. Let Y be the stationary forward-recurrence time of a random variable X. Suppose that $m_n = E(X^n)$ exists for $n = 1, 2, \ldots$. Then show that

$$E\{Y^n\} = \frac{m_{n+1}}{(n+1)m_1}.$$

1.27. Prove that the limiting joint distribution of the residual lifetime $Y(t)$ and spent lifetime $Z(t)$ of a random variable with DF $F(x)$ and finite mean μ is given by

$$Pr\{Y(t) > y, Z(t) > z\} = \frac{1}{\mu}\int_{y+z}^{\infty}[1 - F(u)]du, \quad y > 0, \quad z > 0.$$

1.28. Renewal-reward process. Consider a renewal process $\{X_n, n = 1, 2, \ldots\}$. Suppose that renewal epochs are $t_0 = 0, t_1, \ldots$, and that $N(t)$ is the number of renewals by time t; associate a RV $Y_i (i = 1, 2, \ldots)$ with renewal epoch $t_i (i = 1, 2, \ldots)$. (A reward or cost is associated with each renewal, the amount being given by a RV Y_i for ith renewal.)
Let

$$Y(t) = \sum_{i=1}^{N(t)} Y_i.$$

Then the stochastic process $\{Y(t), t \geq 0\}$ is called a renewal-reward process. Suppose that $E(X_n) = E(X)$ and $E(Y_n) = E(Y)$ are finite.

Then show that (a) with probability 1,

$$\lim_{t\to\infty} \frac{Y(t)}{t} \to \frac{E(Y)}{E(X)};$$

and that (b)

$$\lim_{t\to\infty} \frac{E\{Y(t)\}}{t} \to \frac{E(Y)}{E(X)}.$$

The relation (b) gives the long-run average reward (cost) per unit time in terms of $E(X)$ and $E(Y)$.

References and Further Reading

Bhat, U. N. (1984). *Elements of Applied Stochastic Processes,* 2nd ed., Wiley, New York.

Çinlar, E. (1975). *Introduction to Stochastic Processes,* Prentice-Hall, Englewood Cliffs, NJ.

Grassman, W. K. (1977). Transient solutions on Markovian queueing systems. *Comp. Opns. Res.* **4**, 47–53.

Gross, D., and Miller, D. R. (1984a). The randomization technique as a modeling tool and solution procedures for transient Markov processes. *Opns. Res.* **32**, 343–361.

Gross, D., and Miller, D. R. (1984b). Multiechelon repairable-item provisioning in a time-varying environment using a randomization technique. *Naval Res. Log. Qrly.* **31**, 347–361.

Karlin, S., and Taylor, H. M. (1975). *A First Course in Stochastic Processes,* 2nd ed., Academic Press, New York.

Kaufman, L. (1983). Matrix methods for queueing problems. *SIAM J. Sci. Stat. Comp.,* **4**, 525–552.

Kingman, J. F. C. (1993). *Poisson Process,* Oxford University Press, Oxford.

Lal, R., and Bhat, U. N. (1987). Reduced systems in Markov chains and their applications in queueing theory. *Queueing Systems* **2**, 147–172; Correction **4**, 93.

Medhi, J. (1994). *Stochastic Processes,* 2nd ed. J. Wiley & Sons, New York & Wiley Eastern Ltd. (now New Age International Publishers), New Delhi; 1st ed (1982).

Prabhu, N. U. (1965). *Stochastic Processes,* Macmillan, New York.

Rego, V., and Szpankowski, W. (1989). The presence of exponentiality in entropy maximized $M/GI/1$ queues. *Comp. Opns. Res.* **16**, 441–449.

Ross, S. M. (1980). *Introduction to Probability Models,* 2nd ed., Academic Press, New York.

Ross, S. M. (1983). *Stochastic Processes,* Wiley, New York.

Seneta, E. (1981). *Non-negative Matrices and Markov Chains,* 2nd ed., Springer, New York.

Serfozo, R. F. (1990). Point Processes in *Handbooks in Operations Research and Management Science,* Vol. 2, pp. 1–94 (Eds. D. P. Heyman and M. J. Sobel), North-Holland, Amsterdam.

Stewart, W. J. (1994). *Introduction to the Numerical Solution of Markov Chains,* Princeton University Press, Princeton, NJ.

Tang, Chi-Nang, and Yeung, R. W. (1999). A Graph theoretic approach to Queueing Analysis, Part I & II. *Comm. Statist.—Stochastic Models.* **15**, 791–870.

Whitt, W. (1982). Approximating a point process by a renewal process. Two basic methods. *Opns. Res.* **30**, 125–147.

Whitt, W. (1983). Untold horrors of the waiting time: what the equilibrium distribution will never tell about the queue length process. *Mgmt. Sci.* **29**, 395–408.

Whitt, W. (2002). *Stochastic-Process Limits (An Introduction to Stochastic-Process Limits and Their Application to Queueing),* Springer-Verlag, New York.