2

Mathematical Preliminaries

2.1 Introduction

This chapter provides basic background materials needed in the subsequent chapters of the book. It briefly reviews and summarizes related results of random processes, including Markov chains in both discrete time and continuous time, martingales, Gaussian processes, diffusions, and switching diffusions.

Throughout the book, we work with a probability space $(\Omega, \mathcal{F}, P)$. A collection of σ-algebras $\{\mathcal{F}_t\}$, for $t \geq 0$ or $t = 1, 2, \ldots$, or simply $\mathcal{F}_t$, is called a filtration if $\mathcal{F}_s \subset \mathcal{F}_t$ for $s \leq t$. The $\mathcal{F}_t$ is complete in the sense that it contains all null sets. A probability space $(\Omega, \mathcal{F}, P)$ together with a filtration $\{\mathcal{F}_t\}$ is termed a *filtered probability space*, denoted by $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, P)$.

2.2 Discrete-Time Markov Chains

Working with discrete time $k \in \{0, 1, \ldots\}$, consider a sequence $\{x_k\}$ of $\mathbb{R}^r$ vectors. If for each k, x_k is a random vector (or an $\mathbb{R}^r$-valued random variable), we call $\{x_k\}$ a stochastic process and write it as x_k, $k = 0, 1, 2, \ldots$, or simply x_k if there is no confusion. A stochastic process is wide-sense (or covariance) stationary, if it has finite second moments, a constant mean, and a covariance that depends only on the time difference. The ergodicity of a stationary sequence $\{x_k\}$ refers to the convergence of the sequence

$$\frac{x_1 + x_2 + \cdots + x_n}{n}$$

to its expectation in the almost sure or some weak sense; see Karlin and Taylor [78, Theorem 5.6, p. 487] for a strong ergodic theorem of a stationary process. A stochastic process x_k is adapted to a filtration $\{\mathcal{F}_k\}$, if for each k, x_k is an $\mathcal{F}_k$-measurable random vector.

Suppose that α_k is a stochastic process taking values in $\mathcal{M}$, which is at most countable (i.e., it is either finite $\mathcal{M} = \{1, 2, \ldots, m_0\}$ or countable $\mathcal{M} = \{1, 2, \ldots\}$). We say that α_k is a Markov chain if

$$\begin{aligned} p^{ij}_{k,k+1} &= P(\alpha_{k+1} = j | \alpha_k = i) \\ &= P(\alpha_{k+1} = j | \alpha_0 = i_0, \ldots, \alpha_{k-1} = i_{k-1}, \alpha_k = i), \end{aligned}$$

for any $i_0, \ldots, i_{k-1}, i, j \in \mathcal{M}$.

Given i, j, if $p^{ij}_{k,k+1}$ is independent of time k, i.e., $p^{ij}_{k,k+1} = p^{ij}$, we say that α_k has stationary transition probabilities. The corresponding Markov chains are said to be stationary or time-homogeneous or temporally homogeneous or simply homogeneous. In this case, let $P = (p^{ij})$ denote the transition matrix. Denote the n-step transition matrix by $P^{(n)} = (p^{ij,(n)})$, with

$$p^{ij,(n)} = P(x_n = j | x_0 = i).$$

Then $P^{(n)} = (P)^n$. That is, the n-step transition matrix is simply the matrix P to the nth power. Note that

(a) $p^{ij} \geq 0$, $\sum_j p^{ij} = 1$, and

(b) $(P)^{k_1+k_2} = (P)^{k_1}(P)^{k_2}$, for $k_1, k_2 = 1, 2, \ldots$

The last identity is commonly referred to as the Chapman–Kolmogorov equation. In this book, we work with Markov chains with finite state spaces. Thus we confine our discussion to such cases. Certain algebraic properties of Markov chains will be used in the book, some of which are listed next.

Suppose that A is an $r \times r$ square matrix. Denote the collection of eigenvalues of A by Λ. Then the spectral radius of A, denoted by $\rho(A)$, is defined by $\rho(A) = \max_{\lambda \in \Lambda} |\lambda|$. Recall that a matrix with real entries is said to be a positive matrix if it has at least one positive entry and no negative entries. If every entry of A is positive, we call the matrix strictly positive. Similarly, for a vector $x = (x^1, \ldots, x^r)$, by $x \geq 0$, we mean that $x^i \geq 0$ for $i = 1, \ldots, r$; by $x > 0$, we mean that all entries $x^i > 0$.

Let $P = (p^{ij}) \in \mathbb{R}^{m_0 \times m_0}$ be a transition matrix. Clearly, it is a positive matrix. Then $\rho(P) = 1$; see Karlin and Taylor [79, p. 3]. This implies that all eigenvalues of P are on or inside the unit circle.

For a Markov chain α_k, state j is said to be accessible from state i if $p^{ij,(k)} = P(\alpha_k = j|\alpha_0 = i) > 0$ for some $k > 0$. Two states i and j, accessible from each other, are said to communicate. A Markov chain is irreducible if all states communicate with each other. For $i \in \mathcal{M}$, let $d(i)$ denote the period of state i, i.e., the greatest common divisor of all $k \geq 1$ such that $P(\alpha_{k+n} = i|\alpha_n = i) > 0$ (define $d(i) = 0$ if $P(\alpha_{k+n} = i|\alpha_n = i) = 0$ for all k). A Markov chain is called aperiodic if each state has period one. According to Kolmogorov's classification of states, a state i is recurrent if, starting from state i, the probability of returning to state i after some finite time is 1. A state is transient if it is not recurrent. Criteria on recurrence can be found in most standard textbooks of stochastic processes or Markov chains.

Note that (see Karlin and Taylor [79, p. 4]) if P is a transition matrix for a finite-state Markov chain, the multiplicity of the eigenvalue 1 is equal to the number of recurrent classes associated with P. A row vector $\pi = (\pi^1, \ldots, \pi^{m_0})$ with each $\pi^i \geq 0$ is called a stationary distribution of α_k if it is the unique solution to the system of equations

$$\begin{aligned} &\pi P = \pi, \\ &\sum_i \pi^i = 1. \end{aligned}$$

As demonstrated in [79, p. 85], for i in an aperiodic recurrent class, if $\pi^i > 0$, which is the limit of the probability of starting from state i and then entering state i at the nth transition as $n \to \infty$, then for all j in this class of i, $\pi^j > 0$, and the class is termed positive recurrent or strongly ergodic. The following theorem, concerning the spectral gaps, will be used in the asymptotic expansions.

Theorem 2.1. *Let $P = (p^{ij})$ be the transition matrix of an irreducible aperiodic finite-state Markov chain. Then there exist constants $0 < \lambda < 1$ and $c_0 > 0$ such that*

$$|(P)^k - \overline{P}| \leq c_0 \lambda^k \quad \text{for} \quad k = 1, 2, \ldots,$$

where $\overline{P} = \mathbb{1}_{m_0} \pi$, $\mathbb{1}_{m_0} = (1, \ldots, 1)' \in \mathbb{R}^{m_0 \times 1}$, and $\pi = (\pi^1, \ldots, \pi^{m_0})$ is the stationary distribution of α_k. This implies, in particular,

$$\lim_{k \to \infty} P^k = \mathbb{1}_{m_0} \pi.$$

Suppose that α_k is a Markov chain with transition probability matrix P. One of the ergodicity conditions of Markov chains is the Doeblin's condition (see Doob [49, Hypothesis D, p. 192]; see also Meyn and Tweedie [115, p. 391]). Suppose that there is a probability measure μ with the property that for some positive integer n, $0 < \delta < 1$, and $\Delta > 0$, $\mu(A) \leq \delta$ implies that $P^n(x, A) \leq 1 - \Delta$ for all $x \in A$. In the above, $P^n(x, A)$ denotes the transition probability starting from x reaches the set A in n steps. Note that if α_k is a finite-state Markov chain that is irreducible and aperiodic, then the Doeblin's condition is satisfied.

In the subsequent chapters, we often need to treat nonhomogeneous systems of linear equations. Given an $m_0 \times m_0$ irreducible transition matrix P and a vector G, consider

$$F(P - I) = G, \tag{2.1}$$

where F is an unknown vector. Note that zero is an eigenvalue of the matrix $P - I$ and the null space of $P - I$ is spanned by $\mathbb{1}_{m_0}$. Then by the Fredholm alternative (see Lemma 14.36), (2.1) has a solution iff $G\mathbb{1}_{m_0} = 0$, where $\mathbb{1}_{m_0} = (1, \dots, 1)' \in \mathbb{R}^{m_0 \times 1}$.

Define $Q_c = (P - I \vdots \mathbb{1}_{m_0}) \in \mathbb{R}^{m_0 \times (m_0+1)}$. Consider (2.1) together with the condition $F\mathbb{1}_{m_0} = \sum_{i=1}^{m_0} F_i = \widehat{F}$, which may be written as $FQ_c = G_c$ where $G_c = (G \vdots \widehat{F})$. Since for each t, (2.12) has a unique solution, it follows that $Q_c(t)Q_c'(t)$ is a matrix with full rank; therefore, the equation

$$F[Q_c Q_c'] = G_c Q_c' \tag{2.2}$$

has a unique solution, which is given by $G_c Q_c' [Q_c Q_c']^{-1}$. This observation will be used later in this book.

2.3 Discrete-Time Martingales

Many applications involving stochastic processes depend on the concept of martingale. The definition and properties of discrete-time martingales can be found in Breiman [27, Chapter 5], Chung [38, Chapter 9], and Hall and Heyde [67] among others. This section provides a brief review.

Definition 2.2. Suppose that $\{\mathcal{F}_n\}$ is a filtration, and $\{x_n\}$ is a sequence of random variables. The pair $\{x_n, \mathcal{F}_n\}$ is a martingale if for each n,

(a) x_n is $\mathcal{F}_n$-measurable;

(b) $E|x_n| < \infty$;

(c) $E(x_{n+1}|\mathcal{F}_n) = x_n$ w.p.1.

It is a supermartingale (resp. submartingale) if (a) and (b) in the above hold, and

$$E(x_{n+1}|\mathcal{F}_n) \le x_n \quad (\text{resp. } \ E(x_{n+1}|\mathcal{F}_n) \ge x_n) \quad \text{w.p.1.}$$

In what follows if the sequence of σ-algebras is clear, we simply say that $\{x_n\}$ is a martingale.

Perhaps the simplest example of a discrete-time martingale is the sum $x_n = \sum_{j=1}^n y_j$ of a sequence of i.i.d. random variables $\{y_n\}$ with zero mean. It is readily seen that

$$\begin{aligned} E[x_{n+1}|y_1,\ldots,y_n] &= E[x_n + y_{n+1}|y_1,\ldots,y_n] \\ &= x_n + Ey_{n+1} = x_n \quad \text{w.p.1.} \end{aligned}$$

The above equation illustrates the defining relation of a martingale.

If $\{x_n\}$ is a martingale, we can define $y_n = x_n - x_{n-1}$, which is known as a martingale difference sequence. Suppose that $\{x_n, \mathcal{F}_n\}$ is a martingale. Then the following properties hold.

(a) Suppose $\varphi(\cdot)$ is an increasing and convex function defined on $\mathbb{R}$, if for each positive integer n, $E|\varphi(x_n)| < \infty$, then $\{\varphi(x_n), \mathcal{F}_n\}$ is a submartingale.

(b) Let τ be a stopping time with respect to $\mathcal{F}_n$ (i.e., an integer-valued random variable such that $\{\tau \le n\}$ is $\mathcal{F}_n$-measurable for each n). Then $\{x_{\tau\wedge n}, \mathcal{F}_{\tau\wedge n}\}$ is also a martingale.

(c) The martingale inequality (see Kushner [96, p. 3]) states that for each $\lambda > 0$,

$$\begin{aligned} &P\left(\max_{1\le j\le n} |x_j| \ge \lambda\right) \le \frac{1}{\lambda} E|x_n|, \\ &E \max_{1\le j\le n} |x_j|^2 \le 4E|x_n|^2, \quad \text{if } \ E|x_n|^2 < \infty \ \text{ for each } \ n. \end{aligned} \tag{2.3}$$

(d) The Doob's inequality (see Hall and Heyde [67, p.15]) states that for each $p > 1$,

$$E^{1/p}|x_n|^p \le E^{1/p}\left(\max_{1\le j\le n} |x_j|\right)^p \le qE^{1/p}|x_n|^p,$$

where $p^{-1} + q^{-1} = 1$;

(e) The Burkholder's inequality (see Hall and Heyde [67, p.23]) is: For $1 < p < \infty$, there exist constants K_1 and K_2 such that

$$K_1 E\left|\sum_{j=1}^n y_j^2\right|^{p/2} \le E|x_n|^p \le K_2 E\left|\sum_{i=j}^n y_j^2\right|^{p/2},$$

where $y_n = x_n - x_{n-1}$.

Consider a discrete-time Markov chain $\{\alpha_n\}$ with state space $\mathcal{M}$ (either finite or countable) and one-step transition probability matrix $P = (p^{ij})$. Recall that a sequence $\{f(i) : i \in \mathcal{M}\}$ is P-harmonic or right-regular

(Karlin and Taylor [79, p. 48]), if (a) $f(\cdot)$ is a real-valued function such that $f(i) \geq 0$ for each $i \in \mathcal{M}$, and (b)

$$f(i) = \sum_{j \in \mathcal{M}} p^{ij} f(j) \quad \text{for each} \quad i \in \mathcal{M}. \tag{2.4}$$

If the equality in (2.4) is replaced by $\geq$ (resp. $\leq$), $\{f(i) : i \in \mathcal{M}\}$ is said to be P-superharmonic or right superregular (resp. P-subharmonic or right subregular). Considering $f = (f(i) : i \in \mathcal{M})$ as a column vector, (2.4) can be written as $f = Pf$. Similarly, we can write $f \geq Pf$ for P-superharmonic (resp. $g \leq Pf$ for P-subharmonic). Likewise, $\{f(i) : i \in \mathcal{M}\}$ is said to be P left regular, if (b) above is replaced by

$$f(j) = \sum_{i \in \mathcal{M}} f(i) p^{ij} \quad \text{for each} \quad j \in \mathcal{M}. \tag{2.5}$$

Similarly, left superregular and subregular functions can be defined.

The following paragraph reveals the natural connection between a martingale and a discrete-time Markov chain. Following the idea presented in Karlin and Taylor [78, p. 241], let $\{f(i) : i \in \mathcal{M}\}$ be a bounded P-harmonic sequence. Define $x_n = f(\alpha_n)$. Then $E|x_n| < \infty$. Moreover, owing to the Markov property,

$$\begin{aligned} E(x_{n+1} | \mathcal{F}_n) &= E(f(\alpha_{n+1}) | \alpha_n)) \\ &= \sum_{j \in \mathcal{M}} p^{\alpha_n, j} f(j) \\ &= f(\alpha_n) = x_n \quad \text{w.p.1.} \end{aligned}$$

Therefore, $\{x_n, \mathcal{F}_n\}$ is a martingale. Note that if $\mathcal{M}$ is finite, the boundedness of $\{f(i) : i \in \mathcal{M}\}$ is not needed.

As pointed out in Karlin and Taylor [78], one of the widely used ways of constructing martingales is through the utilization of eigenvalues and eigenvectors of a transition matrix. Again, let $\{\alpha_n\}$ be a discrete-time Markov chain with transition matrix P. Recall that a column vector f is a right eigenvector of P associated with an eigenvalue $\lambda \in \mathbb{C}$, if $Pf = \lambda f$. Let f be a right eigenvector of P satisfying $E|f(\alpha_n)| < \infty$ for each n. For $\lambda \neq 0$, define $x_n = \lambda^{-n} f(\alpha_n)$. Then $\{x_n\}$ is a martingale.

2.4 Continuous-Time Martingales and Markov Chains

Denote the space of $\mathbb{R}^r$-valued continuous functions defined on $[0, T]$ by $C([0, T]; \mathbb{R}^r)$, and the space of functions that are right continuous with left-hand limits endowed with the Skorohod topology by $D([0,T];\mathbb{R}^r)$; see Definition 14.2. Consider $x(\cdot) = \{x(t) \in \mathbb{R}^r : \ t \geq 0\}$. If for each $t \geq 0$, $x(t)$ is an $\mathbb{R}^r$ random vector, we call $x(\cdot)$ a continuous-time stochastic process and write it as $x(t)$, $t \geq 0$, or simply $x(t)$ if there is no confusion.

A process $x(\cdot)$ is *adapted* to a filtration $\{\mathcal{F}_t\}$, if for each $t \geq 0$, $x(t)$ is an $\mathcal{F}_t$-measurable random variable; $x(\cdot)$ is *progressively measurable* if for each $t \geq 0$, the process restricted to $[0,t]$ is measurable with respect to the σ-algebra $\mathcal{B}[0,t] \times \mathcal{F}_t$ in $[0,t] \times \Omega$, where $\mathcal{B}[0,t]$ denotes the Borel sets of $[0,t]$. A progressively measurable process is measurable and adapted, whereas the converse is not generally true. However, any measurable and adapted process with right-continuous sample paths is progressively measurable.

For many applications, we often need to work with a *stopping time*. A stopping time τ on $(\Omega, \mathcal{F}, P)$ with a filtration $\{\mathcal{F}_t\}$ is a nonnegative random variable such that $\{\tau \leq t\} \in \mathcal{F}_t$, for all $t \geq 0$. A stochastic process $\{x(t) : \ t \geq 0\}$ (real or vector valued) is said to be a martingale on $(\Omega, \mathcal{F}, P)$ with respect to $\{\mathcal{F}_t\}$ if:

(a) For each $t \geq 0$, $x(t)$ is $\mathcal{F}_t$-measurable,

(b) $E|x(t)| < \infty$, and

(c) $E[x(t)|\mathcal{F}_s] = x(s)$ w.p.1 for all $t \geq s$.

If we only say that $x(\cdot)$ is a martingale without specifying the filtration $\mathcal{F}_t$, $\mathcal{F}_t$ is taken to be the natural filtration $\sigma\{x(s) : \ s \leq t\}$. If there exists a sequence of stopping times $\{\tau_n\}$ such that $0 \leq \tau_1 \leq \tau_2 \leq \cdots \leq \tau_n \leq \tau_{n+1} \leq \cdots$, $\tau_n \to \infty$ w.p.1 as $n \to \infty$, and the process $x^{(n)}(t) := x(t \wedge \tau_n)$ is a martingale, then $x(\cdot)$ is a local martingale.

A jump process is a right-continuous stochastic process with piecewise-constant sample paths. Let $\alpha(\cdot) = \{\alpha(t) : t \geq 0\}$ be a jump process defined on $(\Omega, \mathcal{F}, P)$ taking values in $\mathcal{M}$. Then $\{\alpha(t) : t \geq 0\}$ is a Markov chain with state space $\mathcal{M}$, if

$$P(\alpha(t) = i|\alpha(r) : r \leq s) = P(\alpha(t) = i|\alpha(s)),$$

for all $0 \leq s \leq t$ and $i \in \mathcal{M}$, with $\mathcal{M}$ being either finite or countable.

For any $i, j \in \mathcal{M}$ and $t \geq s \geq 0$, let $p^{ij}(t,s)$ denote the transition probability $P(\alpha(t) = j|\alpha(s) = i)$, and $P(t,s)$ the matrix $(p^{ij}(t,s))$. We name $P(t,s)$ the transition matrix of the Markov chain $\alpha(\cdot)$, and postulate that

$$\lim_{t \to s^+} p^{ij}(t,s) = \delta^{ij},$$

where $\delta^{ij} = 1$ if $i = j$ and 0 otherwise. It follows that for $0 \le s \le \varsigma \le t$,

$$\begin{cases} p^{ij}(t,s) \ge 0, \ i,j \in \mathcal{M}, \\ \displaystyle\sum_{j\in\mathcal{M}} p^{ij}(t,s) = 1, \ i \in \mathcal{M}, \\ p^{ij}(t,s) = \displaystyle\sum_{k\in\mathcal{M}} p^{ik}(\varsigma,s)p^{kj}(t,\varsigma), \ i,j \in \mathcal{M}. \end{cases}$$

The last identity is usually referred to as the Chapman–Kolmogorov equation. If the transition probability $P(\alpha(t) = j|\alpha(s) = i)$ depends only on $(t-s)$, then $\alpha(\cdot)$ is said to be stationary or it is said to have stationary transition probabilities. In this case, we define $p^{ij}(h) := p^{ij}(s+h,s)$ for any $h \ge 0$. Otherwise, the process is nonstationary. Suppose that $\alpha(t)$ is a continuous-time Markov chain with stationary transition probability $P(t) = (p^{ij}(t))$. It then naturally induces a discrete-time Markov chain. In fact, for each $h > 0$, the transition matrix $(p^{ij}(h))$ is the transition matrix of the discrete-time Markov chain $\alpha_k = \alpha(kh)$, which is called an h-skeleton of the corresponding continuous-time Markov chain in Chung [38, p. 132].

Definition 2.3 (q-Property). A matrix-valued function $Q(t) = (q^{ij}(t))$, for $t \ge 0$, satisfies the q-Property, if

(a) $q^{ij}(t)$ is Borel measurable for all $i,j \in \mathcal{M}$ and $t \ge 0$;

(b) $q^{ij}(t)$ is uniformly bounded. That is, there exists a constant K such that $|q^{ij}(t)| \le K$, for all $i,j \in \mathcal{M}$ and $t \ge 0$;

(c) $q^{ij}(t) \ge 0$ for $j \ne i$ and $q^{ii}(t) = -\sum_{j\ne i} q^{ij}(t)$, $t \ge 0$.

For any real-valued function f on $\mathcal{M}$ and $i \in \mathcal{M}$, write

$$Q(t)f(\cdot)(i) = \sum_{j\in\mathcal{M}} q^{ij}(t)f(j) = \sum_{j\ne i} q^{ij}(t)(f(j) - f(i)).$$

Let us now recall the definition of the generator of a Markov chain.

Definition 2.4 (Generator). A matrix $Q(t)$, $t \ge 0$, is an infinitesimal generator (or in short a generator) of $\alpha(\cdot)$ if it satisfies the q-Property, and for any bounded real-valued function f defined on $\mathcal{M}$

$$f(\alpha(t)) - \int_0^t Q(\varsigma)f(\cdot)(\alpha(\varsigma))d\varsigma \tag{2.6}$$

is a martingale.

Remark 2.5. Motivated by the applications we are interested in, a generator is defined for a matrix satisfying the q-Property above, where an additional condition on the boundedness of the entries of the matrix is posed. Different definitions, including other classes of matrices, may be devised as in Chung [38]. To proceed, we give an equivalent condition for a finite-state Markov chain generated by $Q(\cdot)$.

Lemma 2.6. *Let $\mathcal{M} = \{1, \ldots, m_0\}$. Then $\alpha(t) \in \mathcal{M}$, $t \geq 0$, is a Markov chain generated by $Q(t)$ iff*

$$\left(I_{\{\alpha(t)=1\}}, \ldots, I_{\{\alpha(t)=m_0\}}\right) - \int_0^t \left(I_{\{\alpha(\varsigma)=1\}}, \ldots, I_{\{\alpha(\varsigma)=m_0\}}\right) Q(\varsigma) d\varsigma \quad (2.7)$$

is a martingale.

Proof: See Yin and Zhang [158, Lemma 2.4]. □

For any given $Q(t)$ satisfying the q-Property, there exists a Markov chain $\alpha(\cdot)$ generated by $Q(t)$. If $Q(t) = Q$, a constant matrix, the idea of Ethier and Kurtz [55] can be utilized for the construction. For time-varying generator $Q(t)$, we need to use the piecewise-deterministic process approach, described in Davis [42], to define the Markov chain $\alpha(\cdot)$.

Let $0 = \tau_0 < \tau_1 < \cdots < \tau_l < \cdots$ be a sequence of jump times of $\alpha(\cdot)$ such that the random variables τ_1, $\tau_2 - \tau_1$, $\ldots$, $\tau_{k+1} - \tau_k$, $\ldots$ are independent. Let $\alpha(0) = i \in \mathcal{M}$. Then $\alpha(t) = i$ on the interval $[\tau_0, \tau_1)$. The first jump time τ_1 has the probability distribution

$$P(\tau_1 \in B) = \int_B \exp\left\{\int_0^t q^{ii}(s) ds\right\} \left(-q^{ii}(t)\right) dt,$$

where $B \subset [0, \infty)$ is a Borel set. The post-jump location of $\alpha(t) = j$, $j \neq i$, is given by

$$P(\alpha(\tau_1) = j | \tau_1) = \frac{q^{ij}(\tau_1)}{-q^{ii}(\tau_1)}.$$

If $q^{ii}(\tau_1)$ is 0, define $P(\alpha(\tau_1) = j | \tau_1) = 0$, $j \neq i$. Then $P(q^{ii}(\tau_1) = 0) = 0$. In fact, if $B_i = \{t : q^{ii}(t) = 0\}$, then

$$\begin{aligned} P(q^{ii}(\tau_1) = 0) &= P(\tau_1 \in B_i) \\ &= \int_{B_i} \exp\left\{\int_0^t q^{ii}(s) ds\right\} \left(-q^{ii}(t)\right) dt = 0. \end{aligned}$$

In general, $\alpha(t) = \alpha(\tau_l)$ on the interval $[\tau_l, \tau_{l+1})$. The jump time τ_{l+1} has the conditional probability distribution

$$\begin{aligned} &P(\tau_{l+1} - \tau_l \in B_l | \tau_1, \ldots, \tau_l, \alpha(\tau_1), \ldots, \alpha(\tau_l)) \\ &\quad = \int_{B_l} \exp\left\{\int_{\tau_l}^{t+\tau_l} q^{\alpha(\tau_l)\alpha(\tau_l)}(s) ds\right\} \left(-q^{\alpha(\tau_l)\alpha(\tau_l)}(t + \tau_l)\right) dt. \end{aligned}$$

The post-jump location of $\alpha(t) = j$, $j \neq \alpha(\tau_l)$ is given by

$$P(\alpha(\tau_{l+1}) = j | \tau_1, \ldots, \tau_l, \tau_{l+1}, \alpha(\tau_1), \ldots, \alpha(\tau_l)) = \frac{q^{\alpha(\tau_l)j}(\tau_{l+1})}{-q^{\alpha(\tau_l)\alpha(\tau_l)}(\tau_{l+1})}.$$

Theorem 2.7. *Suppose that the matrix $Q(t)$ satisfies the q-Property for $t \geq 0$. Then the following statements hold.*

(a) *The process $\alpha(\cdot)$ constructed above is a Markov chain.*

(b) *The process*

$$f(\alpha(t)) - \int_0^t Q(\varsigma) f(\cdot)(\alpha(\varsigma)) d\varsigma \tag{2.8}$$

is a martingale for any uniformly bounded function $f(\cdot)$ on $\mathcal{M}$. Thus $Q(t)$ is indeed the generator of $\alpha(\cdot)$.

(c) *The transition matrix $P(t, s)$ satisfies the forward differential equation*

$$\begin{aligned} &\frac{dP(t,s)}{dt} = P(t,s)Q(t),\ t \geq s, \\ &P(s,s) = I, \end{aligned} \tag{2.9}$$

where I is the identity matrix.

(d) *Assume further that $Q(t)$ is continuous in t. Then $P(t, s)$ also satisfies the backward differential equation*

$$\begin{aligned} &\frac{dP(t,s)}{ds} = Q(s)P(t,s),\ t \geq s, \\ &P(s,s) = I. \end{aligned} \tag{2.10}$$

Proof. See Yin and Zhang [158, Theorem 2.5]. □

Suppose that $\alpha(t)$, $t \geq 0$, is a Markov chain generated by an $m_0 \times m_0$ matrix $Q(t)$. The notions of irreducibility and quasi-stationary distribution are given next.

Definition 2.8 (Irreducibility).

(a) A generator $Q(t)$ is said to be weakly irreducible if, for each fixed $t \geq 0$, the system of equations

$$\begin{aligned} &\nu(t)Q(t) = 0, \\ &\sum_{i=1}^{m_0} \nu^i(t) = 1 \end{aligned} \tag{2.11}$$

has a unique solution $\nu(t) = (\nu^1(t), \ldots, \nu^{m_0}(t))$ and $\nu(t) \geq 0$.

(b) A generator $Q(t)$ is said to be irreducible, if for each fixed $t \geq 0$ the systems of equations (2.11) has a unique solution $\nu(t)$ and $\nu(t) > 0$.

By $\nu(t) \geq 0$, we mean that for each $i \in \mathcal{M}$, $\nu^i(t) \geq 0$. Similar interpretation holds for $\nu(t) > 0$. It follows from the definitions above that irreducibility implies weak irreducibility. However, the converse is not true. For example, the generator

$$Q = \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$$

is weakly irreducible, but it is not irreducible because it contains an absorbing state corresponding to the second row in Q. A moment of reflection reveals that for a two-state Markov chain with generator

$$Q = \begin{pmatrix} -\lambda(t) & \lambda(t) \\ \mu(t) & -\mu(t) \end{pmatrix}$$

the weak irreducibility requires only $\lambda(t) + \mu(t) > 0$, whereas the irreducibility requires that both $\lambda(t)$ and $\mu(t)$ be positive. Such a definition is convenient for many applications (e.g., the manufacturing systems mentioned in Khasminskii, Yin, and Zhang [85, p. 292]).

Definition 2.9 (Quasi-Stationary Distribution). For $t \geq 0$, $\nu(t)$ is termed a quasi-stationary distribution if it is the unique solution of (2.11) satisfying $\nu(t) \geq 0$.

Remark 2.10. While studying homogeneous Markov chains, the stationary distributions play an important role. In the context of nonstationary (non-homogeneous) Markov chains, they are replaced by the quasi-stationary distributions, as defined above.

If $\nu(t) = \nu > 0$, it is termed a stationary distribution. In view of Definitions 2.8 and 2.9, if $Q(t)$ is weakly irreducible, then there is a quasi-stationary distribution. Note that the rank of a weakly irreducible $m_0 \times m_0$ matrix $Q(t)$ is $m_0 - 1$, for each $t \geq 0$. The definition above emphasizes the probabilistic interpretation. An equivalent definition pinpointing the algebraic properties of $Q(t)$ is provided next. One can verify their equivalence using the Fredholm alternative; see Lemma 14.36.

Definition 2.11. A generator $Q(t)$ is said to be weakly irreducible if, for each fixed $t \geq 0$, the system of equations

$$\begin{aligned} &f(t)Q(t) = 0, \\ &\sum_{i=1}^{m_0} f^i(t) = 0 \end{aligned} \tag{2.12}$$

has only the trivial (zero) solution.

2.5 Gaussian, Diffusion, and Switching Diffusion Processes

A Gaussian random vector $x = (x^1, x^2, \ldots, x^r)$ is one whose characteristic function has the form

$$\phi(y) = \exp\left(\mathtt{i}\langle y, \mu\rangle - \frac{1}{2}\langle \Sigma y, y\rangle\right),$$

where $\mu \in \mathbb{R}^r$ is a constant vector, $\langle y, \mu\rangle$ is the usual inner product, $\mathtt{i}$ denotes the pure imaginary number satisfying $\mathtt{i}^2 = -1$, and Σ is a symmetric nonnegative definite $r \times r$ matrix. In the above, μ and Σ are the mean vector and covariance matrix of x, respectively.

Let $x(t)$, $t \geq 0$, be a stochastic process. It is a Gaussian process if for any $0 \leq t_1 < t_2 < \cdots < t_k$ and $k = 1, 2, \ldots$, $(x(t_1), x(t_2), \ldots, x(t_k))$ is a Gaussian vector. A random process $x(\cdot)$ has *independent increments* if for any $0 \leq t_1 < t_2 < \cdots < t_k$ and $k = 1, 2, \ldots$,

$$(x(t_1) - x(0)),\ (x(t_2) - x(t_1)),\ \ldots,\ (x(t_k) - x(t_{k-1}))$$

are independent. A sufficient condition for a process to be Gaussian is given next, whose proof can be found in Skorohod [139, p. 7].

Lemma 2.12. *Suppose that the process $x(\cdot)$ has independent increments and continuous sample paths with probability one. Then $x(\cdot)$ is a Gaussian process.*

An $\mathbb{R}^r$-valued random process for $t \geq 0$ is a Brownian motion, if

(a) $B(0) = 0$ w.p.1;

(b) $B(\cdot)$ is a process with independent increments;

(c) $B(\cdot)$ has continuous sample paths with probability one;

(d) the increments $B(t) - B(s)$ have Gaussian distribution with $E(B(t) - B(s)) = 0$ and $\mathrm{Cov}(B(t), B(s)) = \Sigma|t - s|$ for some nonnegative definite $r \times r$ matrix Σ, where $\mathrm{Cov}(B(t), B(s))$ denotes the covariance.

A process $B(\cdot)$ is said to be a standard Brownian motion if $\Sigma = I$. By virtue of Lemma 2.12, a Brownian motion is necessarily a Gaussian process. For an $\mathbb{R}^r$-valued Brownian motion $B(t)$, let $\mathcal{F}_t = \sigma\{B(s) : s \leq t\}$. Let $h(\cdot)$ be an $\mathcal{F}_t$-measurable process taking values in $\mathbb{R}^{r \times r}$ such that $\int_0^t E|h(s)|^2 ds < \infty$ for all $t \geq 0$. Using $B(\cdot)$ and $h(\cdot)$, one can define a stochastic integral $\int_0^t h(s) dB(s)$ such that it is a martingale with mean 0 and

$$E\Big|\int_0^t h(s) dB(s)\Big|^2 = \int_0^t E\Big[\mathrm{tr}(h(s)h'(s)) ds\Big].$$

Suppose that $b(\cdot)$ and $\sigma(\cdot)$ are non-random Borel measurable functions. A process $x(\cdot)$ defined as

$$x(t) = x(0) + \int_0^t b(s, x(s))ds + \int_0^t \sigma(s, x(s))dB(s) \tag{2.13}$$

is called a diffusion. Then $x(\cdot)$ defined in (2.13) is a Markov process in the sense that the Markov property

$$P(x(t) \in A|\mathcal{F}_s) = P(x(t) \in A|x(s))$$

holds for all $0 \le s \le t$ and for any Borel set A. A slightly more general definition allows $b(\cdot)$ and $\sigma(\cdot)$ to be $\mathcal{F}_t$-measurable processes. However, the current definition is sufficient for our purpose.

Associated with the diffusion process, there is an operator $\mathcal{L}$, known as the generator of the diffusion $x(\cdot)$, defined as follows. Let $C^{1,2}$ be the class of real-valued functions on (a subset of) $\mathbb{R}^r \times [0, \infty)$ whose first-order partial derivative with respect to t and the second-order mixed partial derivatives with respect to x are continuous. Define an operator $\mathcal{L}$ on $C^{1,2}$ by

$$\mathcal{L}f(t, x) = \frac{\partial f(t, x)}{\partial t} + \sum_{i=1}^{r} b^i(t, x)\frac{\partial f(t, x)}{\partial x^i} + \frac{1}{2}\sum_{i,j=1}^{r} a^{ij}(t, x)\frac{\partial^2 f(t, x)}{\partial x^i \partial x^j}, \tag{2.14}$$

where $A(t, x) = (a^{ij}(t, x)) = \sigma(t, x)\sigma'(t, x)$. The well-known Ito's lemma (see Gihman and Skorohod [62], Kunita and Watanabe [92], and Liptser and Shiryayev [105]) states that

$$df(t, x(t)) = \mathcal{L}f(t, x(t)) + f'_x(t, x(t))\sigma(t, x(t))dB(t),$$

or in its integral form

$$\begin{aligned} f(t, x(t)) - f(0, x(0)) &= \int_0^t \mathcal{L}f(s, x(s))ds \\ &\quad + \int_0^t f'_x(s, x(s))\sigma(s, x(s))dB(s). \end{aligned}$$

One of the consequences of the Ito's lemma is that

$$M_f(t) = f(t, x(t)) - f(0, x(0)) - \int_0^t \mathcal{L}f(s, x(s))ds$$

is a square integrable $\mathcal{F}_t$-martingale. Conversely, let $x(\cdot)$ be right continuous. Using the notation of martingale problems given by Stroock and Varadhan [143], $x(\cdot)$ is said to be a solution of the martingale problem with operator $\mathcal{L}$ if $M_f(\cdot)$ is a martingale for each $f(\cdot, \cdot) \in C_0^{1,2}$ (the class of $C^{1,2}$ functions with compact support).

Suppose that $\alpha(\cdot)$ is a continuous-time Markov chain with finite-state space $\mathcal{M} = \{1, \dots, m_0\}$ and generator $Q(t)$ and that $\alpha(\cdot)$ is independent of the standard r-dimensional Brownian motion $B(\cdot)$. Then the process $x(\cdot)$

$$x(t) = x(0) + \int_0^t b(s, x(s), \alpha(s))ds + \int_0^t \sigma(s, x(s), \alpha(s))dB(s)$$

is called a switching diffusion or system of diffusions with regime switching. The corresponding operator is defined as follows. For each $\iota \in \mathcal{M}$ and each $f(\cdot, \cdot, \iota) \in C^{1,2}$,

$$\begin{aligned} \mathcal{L}f(t, x, \iota) &= \frac{\partial f(t, x, \iota)}{\partial t} + \sum_{i=1}^{r} b^i(t, x, \iota) \frac{\partial f(t, x, \iota)}{\partial x^i} \\ &\quad + \frac{1}{2} \sum_{i,j=1}^{r} a^{ij}(t, x, \iota) \frac{\partial^2 f(t, x, \iota)}{\partial x^i \partial x^j} + Q(t) f(t, x, \cdot)(\iota), \end{aligned} \tag{2.15}$$

where $A(t, x, \iota) = (a^{ij}(t, x, \iota)) = \sigma(t, x, \iota)\sigma'(t, x, \iota)$. Similar to the case of diffusions, with the $\mathcal{L}$ defined in (2.15), for each $i \in \mathcal{M}$ and $f(\cdot, \cdot, i) \in C^{1,2}$, a result known as generalized Ito's lemma (see [19]) reads

$$\begin{aligned} df(t, x(t), \alpha(t)) &= \mathcal{L}f(t, x(t), \alpha(t)) \\ &\quad + f_x'(t, x(t), \alpha(t))\sigma(t, x(t), \alpha(t))dB(t), \end{aligned}$$

or in its integral form

$$\begin{aligned} &f(t, x(t), \alpha(t)) - f(0, x(0), \alpha(0)) \\ &\quad = \int_0^t \mathcal{L}f(s, x(s), \alpha(s))ds + \int_0^t f_x'(s, x(s), \alpha(s))\sigma(s, x(s), \alpha(s))dB(s). \end{aligned}$$

In addition,

$$M_f(t) = f(t, x(t), \alpha(t)) - f(0, x(0), \alpha(0)) - \int_0^t \mathcal{L}f(s, x(s), \alpha(s))ds$$

is a martingale. Similar to the case of diffusion processes, we can define the corresponding notion of solution of martingale problem accordingly.

2.6 Notes

A nonmeasure theoretic introduction to stochastic processes can be found in Ross [130]. The two volumes by Karlin and Taylor [78, 79] provide an introduction to discrete-time and continuous-time Markov chains. More advanced treatments can be found in Chung [38] and Revuz [127]. A book that deals exclusively with finite-state Markov chain is Iosifescu [73]. The book of Meyn and Tweedie [115] examines Markov chains and their stability. The connection between generators of Markov processes and martingales is explained in Ethier and Kurtz [55]. An account of piecewise-deterministic processes is in Davis [42]. Results on basic probability theory may be found in Chow and Teicher [37]; theory of stochastic processes can be found in Gihman and Skorohod [62]. More detailed discussions regarding martingales and diffusions can be found in Elliott [54]; in-depth study of stochastic differential equations and diffusion processes can be found in Kunita and Watanabe [92].

http://www.springer.com/978-0-387-21948-6

Discrete-Time Markov Chains
Two-Time-Scale Methods and Applications
Yin, G.; Zhang, Q.
2005, XX, 347 p., Hardcover
ISBN: 978-0-387-21948-6