Chapter XIII
Some Elements of the Theory of Markov Processes and Their Convergence to Equilibrium

Special examples of Markov processes, such as random walks in discrete time and Brownian motion in continuous time, have occurred many times in preceding chapters as illustrative examples of martingales and Markov processes. There has also been an emphasis on their recurrence and transience properties. Moreover general discrete parameter Markov processes, also called *Markov chains*, were introduced in Chapter IX, and their important *strong Markov property* is derived in Chapter XI. In the present chapter, we begin afresh and somewhat differently with a focus on the existence of, and convergence to, a unique steady state distribution.

Suppose that $\mathbf{X} = \{X_0, X_1, X_2, \dots\}$ is a (discrete parameter) sequence of random variables on a probability space $(\Omega, \mathcal{F}, P)$ taking values in a measurable space $(S, \mathcal{S})$. The **Markov property** refers to the special type of statistical dependence that arises when the conditional distribution of the *after-n sequence* $\mathbf{X}^{n+} = \{X_n, X_{n+1}, \dots\}$ given $\sigma(X_0, X_1, \dots, X_n)$ coincides with that given $\sigma(X_n)$. If the sequence $\mathbf{X}$ has the Markov property then we refer to it as a **Markov chain** with state space $(S, \mathcal{S})$. The **initial distribution** μ of the initial state X_0 is a probability on $(S, \mathcal{S})$, and the **one-step transition probabilities** are defined by

$$p_n(x, B) = P(X_{n+1} \in B | X_0, \dots, X_n), x \in S, B \in \mathcal{S}, \text{on } [X_n = x], n \geq 0. \tag{13.1}$$

The case in which these transition probabilities do not depend explicitly on n is referred to as that of **homogeneous** or **stationary transition probabilities**. Unless otherwise specified, we only consider Markov chains with homogeneous transition probabilities in this chapter.

Suppose $\mathbf{X}$ is a Markov chain having stationary one-step transition probabilities $p(x, B) \equiv p_n(x, B)$. Then, when the initial distribution is μ,

$$P(X_n \in B) = \int_S p^{(n)}(x, B)\mu(dx), \quad B \in \mathcal{S}, \tag{13.2}$$

© Springer International Publishing AG 2016

R. Bhattacharya and E.C. Waymire, *A Basic Course in Probability Theory*,
Universitext, DOI 10.1007/978-3-319-47974-3_XIII

where $p^{(n)}(x, B)$ is the n**-step transition probability** defined recursively as $p^{(1)}(x, B) = p(x, B)$, and

$$p^{(n+1)}(x, B) = \int_S p^{(n)}(y, B)p(x, dy), \quad B \in \mathcal{S}, x \in S, (n = 1, 2, \dots). \tag{13.3}$$

Given an initial distribution μ and a transition probability $p(x, B)$, $(x \in S, B \in \mathcal{S})$, a canonical construction of the Markov chain on the sequence space $(S^\infty, \mathcal{S}^{\otimes\infty})$ is discussed in Chapter IX. We denote this distribution by Q^μ, and write Q^x in place of Q^{δ_x}, $x \in S$. The Markov property may then be stated as: The conditional distribution of $\mathbf{X}^{n+}$ given $\sigma(X_0, X_1, \dots, X_n)$ is Q^{X_n}. That is, on the subset $[X_n = x]$, this conditional distribution is Q^x, namely, the distribution of the Markov chain starting at $x \in S$.

Definition 13.1 A probability π on $(S, \mathcal{S})$ is said to be an **invariant distribution** if

$$\int_S p(x, B)\pi(dx) = \pi(B), \quad \forall B \in \mathcal{S}. \tag{13.4}$$

This Definition 13.1 says that if X_0 has distribution π then so does X_1 and, by iteration, X_n has distribution π for all $n \geq 1$. In fact the initial distribution π makes the Markov chain a *stationary process* in the sense that the process $\mathbf{X}^{n+}$ has the same distribution as $\mathbf{X}$ for each $n \geq 1$; Exercise 11.

Two of the most familiar examples of Markov chains are the following:

Example 1 *(Independent Sequence)* Let $X_1, X_2, \dots$ be an i.i.d. sequence of S-valued random variables with common distribution π, and let X_0 be an S-valued random variable, independent of this sequence, and having distribution μ. Then $\mathbf{X} = \{X_0, X_1, X_2, \dots\}$ is a Markov chain with initial distribution μ and one-step transition probabilities $p(x, B) = \pi(B)$, $B \in \mathcal{S}$, $x \in S$. Clearly π is the unique invariant distribution defined by (13.4).

Example 2 *(General Random Walk on $\mathbb{R}^k$)* Let $\{Y_n : n \geq 1\}$ be an i.i.d. sequence with common distribution π on $\mathbb{R}^k$, and let Y_0 be an $\mathbb{R}^k$-valued random variable independent of this sequence. These define the *displacements* of the random walk. The position process for the random walk is defined by $X_0 = Y_0$, $X_n = Y_0 + Y_1 + \cdots + Y_n$, $n \geq 1$. Then $\mathbf{X} = \{X_0, X_1, X_2, \dots\}$ is a Markov chain with initial distribution that of Y_0 and transition probabilities $p(x, B) = \pi(B - x)$, $x \in S$, $B \in \mathcal{B}(S)$. This Markov chain has no invariant probability if $\pi(\{0\}) < 1$.

The following are some basic issues concerning invariant probabilities.

- Existence; not always, $S = \{1, 2, \dots\}$, $p(x, \{x + 1\}) = 1$, $x = 1, 2 \dots$. (also see Exercise 1)
- Uniqueness; not always, $S = \{1, 2\}$, $p(x, \{x\}) = 1$, $x = 1, 2$. (also see Exercise 5).
- Convergence; not always $S = \{1, 2\}$, $p(1, \{2\}) = p(2, \{1\}) = 1$. (also see Example 3 and Exercises 3).

- Rates of convergence; e.g., exponential versus algebraic bounds on an appropriate metric? (see Theorem 13.1 below, Exercise 3(d)).

The following theorem provides a benchmark result that eliminates the obstructions captured by the counterexamples. It covers a broad range of examples but is far from exhaustive.

Theorem 13.1 *(Doeblin Minorization)* Assume that there is a nonzero measure λ on $(S, \mathcal{S})$ and an integer $N \geq 1$ such that

$$p^{(N)}(x, B) \geq \lambda(B), \quad \forall x \in S, B \in \mathcal{S}.$$

Then, there is a unique invariant probability π such that

$$\sup_{x \in S} \sup_{B \in \mathcal{S}} |p^{(n)}(x, B) - \pi(B)| \leq (1-\delta)^{[\frac{n}{N}]}, \quad n = 1, 2, \ldots, \tag{13.5}$$

where $\delta = \lambda(S)$.

Proof Notice that if $\lambda(S) = 1$ then, considering that the minorization inequality applies to both B and B^c, it follows that $p^{(N)}(x, B) = \lambda(B), x \in S, B \in \mathcal{S}$ is the invariant probability; use (13.3) to see $p^{(n)}(x, B) = \lambda(B)$ does not depend on $x \in S$, and both sides of (13.5) are zero. Now assume $\delta = \lambda(S) < 1$. Let d denote the total variation metric on $\mathcal{P}(S)$. Then recall Proposition 1.9 of Chapter I that $(\mathcal{P}(S), d)$ is a complete metric space and $d_1(\mu, \nu) := \sup\{|\int_S f d\mu - \int_S f d\nu| : f \in \mathbb{B}(S), |f| \leq 1\} = 2d(\mu, \nu)$, for all $\mu, \nu \in \mathcal{P}(S)$. Define $T^* : \mathcal{P}(S) \to \mathcal{P}(S)$ by $T^*\mu(B) = \int_S p(x, B)\mu(dx)$, $B \in \mathcal{B}(S)$. One may use (13.5) to write

$$p^{(N)}(x, B) = \delta\gamma(B) + (1-\delta)q(x, B), \tag{13.6}$$

where $\gamma(B) := \frac{\lambda(B)}{\delta}$, and $q(x, B) := \frac{p^{(N)}(x,B) - \lambda(B)}{1-\delta}$ are both probability measures. It follows that for all measurable f, $|f| \leq 1$, and $\mu, \nu \in \mathcal{P}(S)$,

$$\begin{aligned}
&\int_S f(y)T^{*N}\mu(dy) - \int_S f(y)T^{*N}\nu(dy) \\
&= \int_S \int_S f(y)p^{(N)}(x, dy)\mu(dx) - \int_S \int_S f(y)p^{(N)}(x, dy)\nu(dx) \\
&= (1-\delta)[\int_S \int_S f(y)q(x, dy)\mu(dx) - \int_S \int_S f(y)q(x, dy)\nu(dx). \qquad (13.7)
\end{aligned}$$

This implies $d_1(T^{*N}\mu, T^{*N}\nu) \leq (1-\delta)d_1(\mu, \nu)$. Iterating this one obtains (by induction)

$$d_1(T^{*Nk}\mu, T^{*Nk}\nu) \leq (1-\delta)^k d_1(\mu, \nu), \quad k \geq 1. \tag{13.8}$$

Next observe that $\forall \mu \in \mathcal{P}(S)$, the sequence $\{T^{*Nk}\mu : k \geq 1\}$ is Cauchy for the metric d_1 since $T^{*N(k+r)}\mu = T^{*Nk}(T^{*Nr}\mu)$, and therefore has a limit π which is the

unique invariant probability. Take $\mu(\cdot) = p(x, \cdot)$, and $\nu = \pi$ in (13.8) to complete the proof. ■

The following is a simple consequence.

Corollary 13.2 Suppose that $S = \{1, 2, \ldots, M\}$ is a finite set and $\{X_0, X_1, \ldots\}$ is a Markov chain on S with one-step transition probabilities $P(X_{n+1} = j | X_n = i)$ given by the transition probability matrix $p = ((p_{ij}))_{i,j \in S}$. If there is an N such that all entries of $p^N = ((p_{ij}^{(N)}))_{1 \le i,j \le M}$ are positive, then there is a unique invariant probability π on S, and $p_{i\cdot}^{(n)}$ converges to π exponentially fast and uniformly for all $i \in S$.

Proof Define $\lambda(B) = \sum_{j \in B} \lambda(\{j\})$,where $\lambda(\{j\}) = \min_{i \in S} p_{ij}^{(N)}$, $j \in S$, and the empty sum is defined to be zero. Then for each $B \subset S$,

$$p^{(N)}(i, B) = \sum_{j \in B} p_{ij}^{(N)} \ge \lambda(B).$$

The uniform exponential convergence follows from (13.5). ■

Example 3 *(Simple Symmetric Random Walk with Reflection)* Here $S = \{0, 1, \ldots, d-1\}$ for some $d > 2$, and $p_{i,i+1} \equiv p(i, \{i+1\}) = \frac{1}{2} = p(i, \{i-1\}) \equiv p_{i,i-1}$, $1 \le i \le d-2$, and $p_{0,1} = p_{d-1,d-2} = 1$. The unique solution to (13.4) is $\pi(\{0\}) = \pi(\{d-1\}) = \frac{1}{2(d-1)}$, and $\pi(\{i\}) = 1/(d-1)$, $1 \le i \le d-2$. However, the hypothesis of Corollary 13.2, (or that of Theorem 13.1), does not hold. Indeed $p_{ij}^{(N)} = 0$ if N and $|i-j|$ have opposite parity; also see Exercise 4 in this regard.

Example 4 *(Fluctuation-Dissipation Effects)* Let $\theta \in (0, 1)$ and let $\varepsilon_1, \varepsilon_2, \ldots$ be an i.i.d. sequence of Gaussian mean zero, variance σ^2 random variables. Define a Markov process on $S = \mathbb{R}$ by $X_{n+1} = \theta X_n + \varepsilon_{n+1}$, $n = 0, 1, 2, \ldots$ for an initial state $X_0 = x \in S$. Then

$$X_n = \sum_{j=0}^{n-1} \theta^j \varepsilon_{n-j} =^{dist} \sum_{j=0}^{n-1} \theta^j \varepsilon_j, n = 1, 2, \ldots.$$

In particular, the limit distribution is Gaussian with mean zero and variance $\frac{1}{1-\theta^2}\sigma^2$.

Remark 13.1 Theorem 13.1, Corollary 13.2, and Example 4 concern so-called *irreducible Markov chains*, in the sense that for each $x \in S$ there is a positive integer $n = n(x)$ such that the n-step transition probability $p^{(n)}(x, B)$ is positive for every $B \in \mathcal{S}$ such that $\lambda(B) > 0$, for some nonzero reference measure λ on $(S, \mathcal{S})$. On the other hand, the Markov chain in Example 3 is not irreducible.

While the time asymptotic theory for irreducible Markov processes is quite well-developed, there are important examples for which irreducibility is too strong an hypothesis. The following example is presented to illustrate some useful theory in cases of non-irreducible Markov processes.

Example 5 *(A Fractional Linear Dynamical System; Products of Random Matrices)* Let $S = [0, \infty)$ and let $A_n, B_n, n = 1, 2, \ldots$ be an i.i.d. sequence positive random variables with $\mathbb{E} \log A_1 < 0$. Define a Markov chain on S by $X_0 = x \in S$, and

$$X_{n+1} = \frac{A_{n+1} X_n}{A_{n+1} X_n + B_{n+1}}, n = 0, 1, 2 \ldots.$$

Then $\pi = \delta_0$ is the unique invariant distribution. To see this observe that the composition of two fractional linear maps $\alpha_1 \circ \alpha_2(x) = \alpha_1(\alpha_2(x))$, $\alpha_j(x) = \frac{a_j x}{a_j x + b_j}$, $x \geq 0$, $j = 1, 2$, may be identified with the result of matrix multiplications of the two matrices $\begin{pmatrix} a_j & 0 \\ a_j & b_j \end{pmatrix}$, $j = 1, 2$, to compute the composite coefficients.[1] In particular, X_n may be identified as an n-fold matrix product whose diagonal entries are each distributed as $\prod_{j=1}^n A_j = \exp\{n \frac{\sum_{j=1}^n \log A_j}{n}\} \sim \exp\{n \mathbb{E} \log A_1\} \to 0$ almost surely, and hence in distribution, as $n \to \infty$. The upper off-diagonal entry is zero, and the lower off-diagonal entry is $\sum_{j=1}^n \prod_{i=1}^{j-1} B_i \prod_{i=j}^n A_i$. Since $\mathbb{E} \log A_1 < 0$, one has $\frac{d}{dh} \mathbb{E} A_1^h = \mathbb{E} A_1^h \log A_1 = \mathbb{E} \log A_1 < 0$ at $h = 0$, and $\mathbb{E} A_1^h = 1$ at $h = 0$. One may choose sufficiently small $h \in (0, 1)$ such that $\mathbb{E} A_1^h < 1$. For such a choice one then has from sublinearity that $\mathbb{E}\big(\sum_{j=1}^n \prod_{i=1}^{j-1} B_i \prod_{i=j}^n A_i\big)^h \leq n(\mathbb{E} A_1^h)^n = n e^{n \log \mathbb{E} A_1^h} \to 0$ as $n \to \infty$. In fact by this, Chebyshev's inequality and the Borel–Cantelli argument $\sum_{j=1}^n \prod_{i=1}^{j-1} B_i \prod_{i=j}^n A_i \to 0$ a.s. as $n \to \infty$, as well.

The previous two examples are illustrations of Markov processes that arise as iterations of i.i.d. random maps, or so-called *random dynamical systems*.[2]

Example 6 *(Ehrenfest urn model)* The following model for heat exchange was introduced by P. and T. Ehrenfest in 1907, and later by Smoluchowski in 1916, to explain an apparent paradox that threatened to destroy the basis of Boltzmann's kinetic theory of matter. In the kinetic theory, heat exchange between two bodies in contact is a random process involving the exchange of energetic molecules, while in thermodynamics it is an orderly irreversible progression toward an equilibrium state in which the (macroscopic) temperatures of two bodies in contact become (approximately) equal. The main objective of kinetic theory was to explain how the larger scale thermodynamic equilibrium could be achieved, while allowing for statistical recurrence of the random process. In fact, Zermelo argued forcefully that recurrence would contradict thermodynamic irreversibility. However, Boltzmann was of the view that the time required by the random process to pass from the equilibrium state to one of

[1] The authors thank our colleague Yevgeniy Kovchegov for suggesting this example to illustrate products of random matrices. Such examples as this, including the positivity constraints, arise naturally in the context of mathematical biology.

[2] A comprehensive treatment of such Markov processes can be found in Bhattacharya, R., and M. Majumdar (2007). Limit distributions of products of random matrices has been treated in some generality by Kaijser, T.(1978): A limit theorem for Markov chains on compact metric spaces with applications to random matrices, *Duke Math. J.* **45**, 311–349; Kesten, H. and F. Spitzer (1984): Convergence in distribution of products of random matrices, *Z. Wahrsch. Verw. Gebiete* **67** 363–386.

macroscopic nonequilibrium would be so large that such recurrence would be of no physical significance. Not all physicists were convinced of this reasoning.

So enter the Ehrenfests. Suppose that $2d$ balls labelled $1, 2, \dots 2d$ are distributed between two boxes A and B at time zero. At each instant of time, a ball label is randomly selected, independently of the number of balls in either box, and that ball is moved from its current box to the other box. Suppose that there are initially Y_0 balls in box A, and let Y_n denote the number of balls in box A at the nth stage of this process. Then one may check that $Y = \{Y_0, Y_1, \dots\}$ is a Markov chain on the state space $S = \{0, 1, 2, \dots, 2d\}$ with one-step transition probabilities $p(y, y+1) = \frac{2d-y}{2d}$, $p(y, y-1) = \frac{y}{2d}$, $p(y, z) = 0$ otherwise. Moreover, Y has a unique invariant probability π with mean d, given by the binomial distribution with parameters $\frac{1}{2}, 2d$, i.e.,

$$\pi_j = \binom{2d}{j} 2^{-2d}, \quad j = 0, 1, \dots, 2d.$$

Viewing the average state d of the invariant distribution π as thermodynamic equilibrium, the paradox is that, as a result of recurrence of the Markov chain, the state $j = 0$ of extreme disequilibrium is certain to eventually occur. The paradox can be resolved by calculating the average length of time to pass from $j = d$ to $j = 0$ in this kinetic theoretical model.[3]

The following proposition provides a general framework for such calculations.

Proposition 13.3 *(Birth–Death Markov Chain with Reflection)* Let $Y = \{Y_n : n = 0, 1, 2, \dots\}$ be a Markov chain on the state space $S = \{0, 1, \dots, N\}$ having stationary one-step transition probabilities $p_{i,i+1} = \beta_i$, $i = 0, 1, \dots, N-1$, $p_{i,i-1} = \delta_i$, $i = 1, 2, \dots, N$, $p_{0,1} = p_{N,N-1} = 1$, and $p_{ij} = 0$ otherwise, where $0 < \beta_i = 1 - \delta_i < 1$. Let

$$T_j = \inf\{n \geq 0 : Y_n = j\}, \quad j \in S,$$

denote the first-passage time to state $j \in S$. Then

$$m_i = \mathbb{E}_i T_0 = \sum_{j=1}^{i} \frac{\beta_j \beta_{j+1} \cdots \beta_{N-1}}{\delta_j \delta_{j+1} \cdots \delta_{N-1}} + \sum_{j=1}^{i} \sum_{k=j}^{N-1} \frac{\beta_j \cdots \beta_{k-1} \beta_k}{\delta_j \delta_{j+1} \cdots \delta_k \beta_k}, \quad 1 \leq i \leq N-1.$$

Proof The idea for the proof involves a scale-change technique that is useful for many Markov chains that do not skip over adjacent states; including one-dimensional diffusions having continuous paths. Specifically, one relabels the states $j \to u_j$ by

[3]The original calculations of the Ehrenfests and Smoluchowski were for the mean recurrence times. Such calculations are easily made from the general mean return-time formula $\mathbb{E}_i \tau_i = \frac{1}{\pi_i}$, where $\tau_i = \inf\{n \geq 1 : Y_n = i\}$, $i \in S$, for irreducible, ergodic Markov chains. In particular, using the formula for π and Stirling's formula, $\mathbb{E}_0 \tau_0 \sim 2^{20,000}$, $\mathbb{E}_d \tau_d \sim 100\sqrt{\pi}$, for the same numerical values for the number of balls and transition rate; e.g., see Kac (1947): Random walk and the theory of Brownian motion, *Am. Math. Monthly*, **54**(7), 369–391. The mean-return time formula and more general theory can be found in standard treatments of discrete parameter Markov processes.

an increasing sequence $0 = u_0 < u_1 < \cdots < u_N = 1$ determined by the requirement that the probabilities of reaching one boundary before another, starting in-between, is proportional to the respective distance to the boundary, as in the examples of simple symmetric random walk on $\mathbb{Z}$, and one-dimensional standard Brownian motion. That is,

$$\psi(i) = P(Y \text{ reaches } 0 \text{ before } \mathrm{N} | \mathrm{Y}_0 = \mathrm{i}) = \frac{\mathrm{u_N} - \mathrm{u_i}}{\mathrm{u_N} - \mathrm{u_0}}, \quad \mathrm{i} \in \mathrm{S}.$$

Since

$$\psi(i) = \beta_i \psi(i+1) + \delta_i \psi(i-1), 1 \le i \le N-1,$$

and $\psi(0) = 1$, $\psi(N) = 0$, one has

$$u_{i+1} - u_i = \frac{\delta_i}{\beta_i}(u_i - u_{i-1}) = \frac{\delta_1 \cdots \delta_i}{\beta_1 \cdots \beta_i}(u_1 - u_0). \tag{13.9}$$

Thus, one obtains the appropriate *scale function*

$$u_{j+1} = 1 + \sum_{i=1}^{j} \frac{\delta_1 \cdots \delta_i}{\beta_1 \cdots \beta_i}, \quad 1 \le j \le N-1.$$

The transformed Markov chain u_Y is said to be on *natural scale*. Now write $m(u_j) = m_j$, $j \in S$.

$$\{m(u_{j+1}) - m(u_j)\}\beta_j - \{m(u_j) - m(u_{j-1})\}\delta_j = -1, \quad 1 \le j \le N-1, \tag{13.10}$$

with boundary conditions

$$m(u_0) = m(0) = 0, \quad m(u_N) - m(u_{N-1}) = 1.$$

Using (13.9), one has

$$\frac{m(u_{j+1}) - m(u_j)}{u_{j+1} - u_j} - \frac{m(u_j) - m(u_{j-1})}{u_j - u_{j-1}} = -\frac{\beta_0 \beta_1 \cdots \beta_{j-1}}{\delta_1 \delta_2 \cdots \delta_j}, 1 \le j \le N-1.$$

Summing over $j = i, i+1, \ldots, N-1$ and using the boundary conditions, one has

$$(u_N - u_{N-1})^{-1} - \frac{m(u_i) - m(u_{i-1})}{u_i - u_{i-1}} = -\sum_{j=i}^{N-1} \frac{\beta_0 \beta_1 \cdots \beta_{j-1}}{\delta_1 \delta_2 \cdots \delta_j}, \quad 1 \le i \le N-1.$$

This and (13.10) lead to

$$m(u_i) - m(u_{i-1}) = \frac{\beta_i \beta_{i+1} \cdots \beta_{N-1}}{\delta_i \delta_{i+1} \cdots \delta_{N-1}} + \sum_{j=i}^{N-1} \frac{\beta_i \cdots \beta_{j-1} \beta_j}{\delta_i \cdots \delta_j \beta_j}, \quad 1 \le i \le N-1.$$

The factor β_j/β_j was introduced to accommodate the term corresponding to $j = i$. The asserted formula now follows by summing over i, using $m(u_0) = 0$. ■

In the application to the Ehrenfest model one obtains

$$m_d = \sum_{j=1}^{d} \frac{(2d-j)!(j-1)!}{(2d-1)!} + \sum_{j=1}^{d} \sum_{k=j}^{2d-1} \frac{(2d-j)!(j-1)!}{(2d-k)!k!} = \frac{2^{2d}}{2d}(1 + O(\frac{1}{d})),$$

in the limit as $d \to \infty$. For $d = 10,000$ balls and an exchange rate of one ball per second, it follows that $m_d = 10^{6000}$ years. The companion calculation of the mean time to thermodynamic equilibrium from a state far away,

$$\tilde{m}_0 = \mathbb{E}_0 T_d \le d + d \log d + O(1), d \to \infty, \tag{13.11}$$

is left as Exercise 6. For the same numerical values one obtains from this that $\tilde{m}_0 \le 29$ h. In particular, it takes about a day on average for the system to reach thermodynamic equilibrium from a state farthest away, but it takes an average time that is inconceivably large for the system to go from a state of thermodynamic equilibrium to the same state far from equilibrium.

We saw that Brownian motion is an example of a continuous parameter Markov process having continuous sample paths. More generally, any right-continuous stochastic process $\mathbf{X} = \{X(t) : t \ge 0\}$ having independent increments has the *Markov property* since for $0 \le s < t$, the conditional distribution of $X(t) = X(s) + X(t) - X(s)$ given $\sigma(X(u) : 0 \le u \le s)$ is the same as that given $\sigma(X(s))$. In view of the independence of $X(t) - X(s)$ and $\sigma(X(u) : 0 \le u \le s)$, the former is the distribution of $x + X(t) - X(s)$ on $[X(s) = x]$. If the Markov process is *homogeneous*, i.e., the conditional distribution of $X(t+s)$ given $\sigma(X(s))$ does not depend on s, then this distribution is the *transition probability* $p(t; x, dy)$ on $[X(s) = x]$, namely the distribution of $X(t)$ when $X(0) = x$. Exercise 12.

The following is another example of a continuous parameter Markov process.

Example 7 *(Ornstein–Uhlenbeck process)* The Ornstein–Uhlenbeck process provides an alternative to the Brownian motion model for the molecular diffusion of a suspended particle in a liquid. It is obtained by considering the particle's *velocity* rather than its position. Considering one coordinate, say $V = \{V(t) : t \ge 0\}$, one assumes that the motion is driven by a combination of inertial drag and the momentum provided by random bombardments by surrounding molecules. Specifically, in a small amount of time $h > 0$,

$$V(t+h) - V(t) \approx -\beta V(t)h + \sigma(B(t+h) - B(t)), \quad t \ge 0,$$

where $\beta > 0$ is a constant drag coefficient, $\sigma^2 > 0$ is the molecular diffusion coefficient, and B denotes standard Brownian motion. The frictional term embodies Stokes law from fluid dynamics which asserts that the frictional force decelerating a spherical particle of radius $r > 0$, mass m, is given by

$$\beta = \frac{6\pi r\eta}{m},$$

where $\eta > 0$ is the coefficient of viscosity of the surrounding fluid. To achieve this modeling hypothesis one may consider the integrated form in which V is specified as a process with continuous sample paths satisfying the so-called *Langevin equation*

$$V(t) = -\beta \int_0^t V(s)ds + \sigma B(t), \quad V(0) = u. \tag{13.12}$$

Theorem 13.4 For each initial state $V(0)$, there is a unique Markov process V with state space $S = \mathbb{R}$ having continuous sample paths defined by (13.12). Moreover, V is Gaussian with transition probability density

$$p(t; u, v) = \frac{1}{\sqrt{2\pi\sigma^2(1 - e^{-2\beta t})}} \exp\{-\frac{1}{2\sigma^2(1 - e^{-2\beta t})}(v - ue^{-\beta t})^2\}, \; u, v \in \mathbb{R}.$$

Proof The proof is by the Picard iteration method. First define a process $V^{(0)}(t) = u$ for all $t \geq 0$. Next recursively define $V^{(n+1)}$ by

$$V^{(n+1)}(t) = \int_0^t V^{(n)}(s)ds + \sigma B(t), \quad t \geq 0, n = 0, 1, 2, \ldots.$$

Iterating this equation for $n = 1, 2, 3$, changing the order of integration as it occurs, one arrives at the following induction hypothesis

$$V^{(n)}(t) = u \sum_{j=0}^{n} \frac{(-\beta t)^j}{j!} + \sum_{j=1}^{n-1} (-\beta)^j \sigma \int_0^t \frac{(t-s)^{j-1}}{(j-1)!} B(s)ds + \sigma B(t), \quad t \geq 0. \tag{13.13}$$

Letting $n \to \infty$ one obtains sample pathwise that

$$V(t) := \lim_{n\to\infty} V^{(n)}(t) = e^{-\beta t}u - \beta\sigma \int_0^t e^{-\beta(t-s)} B(s)ds + \sigma B(t), \quad t \geq 0.$$

In particular V is a linear functional of the Brownian motion B. That V has continuous paths and is Gaussian follows immediately from the corresponding properties of Brownian motion. Moreover, this solution is unique. To prove uniqueness, suppose that $Y = \{Y(s) : 0 \leq s \leq T\}$ is another a.s. continuous solution to (13.12) and consider

$$\Delta(t) = \mathbb{E}(\max_{0\le s\le t} |X(s) - Y(s)|^2), 0 \le t \le T.$$

Then,

$$\Delta(T) \le 2\beta^2\mathbb{E}(\int_0^T |V(s) - Y(s)|ds)^2 \le 2\beta^2 \int_0^T \Delta(s)ds. \tag{13.14}$$

Since $t \to \Delta(t)$ is nondecreasing on $0 \le t \le T$, applying this inequality to the integrand $\Delta(s)$ and reversing the order of integration yields $\Delta(T) \le (2\beta^2)^2 \int_0^T (T - s)\Delta(s)ds \le \frac{(2\beta^2 T)^2}{2}\Delta(T)$. Iterating, one sees by induction that

$$\Delta(T) \le \frac{(2\beta^2 T)^n}{n!}\Delta(T), n = 2, 3, \ldots.$$

Thus $\Delta(T) = 0$ and $Y = V$ a.s. on $[0, T]$. Since T is arbitrary this establishes the uniqueness. From uniqueness one may prove the Markov property holds for V as follows. First, let us note that the solution starting at u at time s, i.e.,

$$V^{(s,u)}(t) = u - \beta \int_s^t V^{(s,u)}(s)ds + \sigma(B(t) - B(s)), t \ge s, \tag{13.15}$$

can be obtained by Picard iteration as a unique measurable function $\theta(s, t; u, B(t) - B(s)), t \ge s$. Since $V(t), t \ge s$ is a solution starting at $u = V(s)$, i.e.,

$$V(t) = V(s) - \beta \int_s^t V(r)dr + \sigma(B(t) - B(s)), \quad 0 \le s < t,$$

it follows from uniqueness that $V(t) = \theta(s, t; V(s), B(t) - B(s)), t \ge s$. Thus, the conditional distribution of $V(t)$ given $\mathcal{F}_s = \sigma(B(r) : r \le s)$ is the distribution of $\theta(x, t; u, B(t) - B(s))$ evaluated at $u = V(s)$. Since $\sigma(V(r) : r \le s) \subset \mathcal{F}_s, s \ge 0$, this proves the Markov property.

Let us now compute the transition probabilities, from which we will also see that they are homogenous in time. In view of the linearity of the functional θ of Brownian motion it is clear that the conditional distribution is Gaussian. Thus, it is sufficient to compute the conditional mean and variance of $V(t)$ started at $u = V(s), s < t$. In particular, one obtains

$$p(t; u, v) = \sqrt{\frac{\beta}{\pi\sigma^2(1 - e^{-2\beta t})}} \exp\{ - \frac{\beta(v - ue^{-\beta t})^2}{\sigma^2(1 - e^{-2\beta t})}\}$$

is Gaussian with mean $ue^{-\beta t}$ and variance $\frac{\sigma^2}{2\beta}(1 - e^{-2\beta t})$. ■

Remark 13.2 A simpler construction of the Ornstein–Uhlenbeck process is given in Exercise 8 which expresses it as a functional of Brownian motion. The Markov

property is also immediate from this representation. However, the above derivation is significant because of its historic relation to physics, in particular, significant in its role as a precursor to the development of the mathematical theory of stochastic differential equations. In this regard, the Ornstein–Uhlenbeck example provides an example of a stochastic differential equation

$$dV(t) = -\beta V(t)dt + \sigma dB(t), \quad V(0) = u,$$

which, because σ is a constant, requires no special calculus to interpret. In fact, the definition is provided for (13.12) using ordinary Riemann integrals $\int_0^t V(s)ds$ of the (continuous) paths of V. The extension to more general equations of the form

$$dV(t) = \mu(V(t), t)dt + \sigma(V(t), t)dB(t), \quad V(0) = 0,$$

in one and higher dimensions is the subject of stochastic differential equations and Itô calculus to define integrals of the form $\int_0^t \sigma(V(s), s)dB(s)$ for nonconstant integrands $\sigma(V(s), s)$. K. Itô's development of a useful calculus in this regard provides a striking illustration of the power of martingale theory.

Exercise Set XIII

1. (*Unrestricted Simple Symmetric Random Walk on* $\mathbb{Z}$) Define a transition probability on $S = \mathbb{Z}$ by $p_{i,i+1} = \frac{1}{2} = p_{i,i-1}$, $i \in \mathbb{Z}$. Show that there is not an invariant probability for this Markov chain.
2. (*Uniqueness of an Invariant Probability*) (a) Suppose $\frac{1}{N}\sum_{n=1}^N p^{(n)}(x, dy)$ converges, for each $x \in S$, to a probability $\pi(dy)$ in total variation norm as $N \to \infty$. Show that π is the unique invariant probability. (b) Suppose that the convergence in (a) to $\pi(dy)$ is weak convergence of the probabilities $\frac{1}{N}\sum_{n=1}^N p^{(n)}(x, dy)$ on a metric space $(S, \mathcal{B}(S))$. Show the same conclusion as in (a) holds if the transition probability $p(x, dy)$ has the *Feller property*: Namely, for each bounded, continuous function f on S the function $x \to \int_S f(y)p(x, dy)$, $x \in S$ is continuous.
3. (*Asymmetric Simple Random Walk with Reflection*) Let $S = \{0, 1, \ldots, d-1\}$ for some $d > 2$, and for some $0 < p < 1$, define $p_{i,i+1} = p$, $p_{i,i-1} = 1 - p$, $1 \le i \le d-2$, and $p_{0,1} = 1 = p_{d-1,d-2}$. (a) Show that there is a unique invariant probability and compute it. (b) Show that $p_{ij}^{(n)} = 0$ if n and $|i - j|$ have opposite parity. (c) Show that $\tilde{p}_{i,j} := p_{i,j}^{(2)}$ defines a transition probability on each of the state spaces $S_0 = \{i \in S : i \text{ is even}\}$, and $S_1 = \{i \in \mathrm{S} : i \text{ is odd}\}$, and that the hypothesis of Corollary 13.2 holds for each of these Markov chains. (d) Show that $\frac{1}{N}\sum_{n=1}^N p_{ij}^{(n)}$ converges to the unique invariant probability π on S. Moreover, show that the convergence is exponentially fast as $N \to \infty$, and uniform over all $i, j \in S$.
4. (*Lazy Random Walk*) Suppose the transition probabilities in Exercise 3 are modified to assign positive probability $p_{ii} = \varepsilon > 0$ to each state in S while keeping

$p_{i,i+1} = p_{i,i-1} = (1-\varepsilon)/2$, $1 \le i \le d-2$, and $p_{0,1} = p_{d-1,d-2} = 1-\varepsilon$, and $p_{i,j} = 0$ if $|i-j| > 1$. Show that Doeblin's Theorem 13.1 applies to this Markov chain.

5. (*Simple Random Walk with Absorption*) Suppose that the transition probabilities in Exercise 3 are modified so that $p_{0,0} = p_{1,1} = 1$. Show that there are two invariant probabilities $\delta_{\{0\}}$ and $\delta_{\{1\}}$, and hence infinitely many.
6. (*Ehrenfest model continued*) Calculate $\tilde{m}_0$ in (13.11) for the Ehrenfest model by the following steps:
   (i) Write $\tilde{m}(u_i) = \tilde{m}_i$, $1 \le i \le d-1$, and show that the same equations as for $m(u_i)$ apply with boundary conditions $\tilde{m}(u_0) = 1 + \tilde{m}(u_1)$, $\tilde{m}(u_d) = 0$.
   (ii) Summing over $j = 1, 3, \ldots, d-1$, show that $\tilde{m}_0 = 1 + \sum_{j=1}^{d-1} \frac{j!}{(2d-1)\cdots(2d-j)} + \sum_{j=1}^{d-1} \sum_{k=1}^{j} \frac{(j+1)j\cdots(k+2)(k+1)}{(2d-k)\cdots(2d-j)(j+1)}$
   (iii) Verify that $\tilde{m}_0 \le d + d\log d + O(1)$ as $d \to \infty$.
7. (*Stationary Ornstein–Uhlenbeck/Maxwell-Boltzmann Steady State*) (a) Show that the time-asymptotic distribution of the Ornstein–Uhlenbeck process is Gaussian with mean zero and variance $\frac{\sigma^2}{2\beta}$ regardless of the initial distribution. (b) Show that this is the unique invariant distribution of V.[4] (c) What general features do the Erhenfest model and Ornstein–Uhlenbeck diffusion have in common ? [*Hint*: Consider the conditional mean and variance of displacements of the process $v_n = Y_n - d$, $n = 0, 1, 2, \ldots$. Namely, $\mathbb{E}(v_{n+1} - v_n | v_0, \ldots, v_n)$ and $\mathbb{E}((v_{n+1} - v_n)^2 | v_0, \ldots, v_n)$.]
8. (*Ornstein–Uhlenbeck process; Time change of Brownian Motion*) Assume that $V(0)$ has the stationary distribution for the Ornstein–Uhlenbeck process. Then V can be expressed as a time-change of Brownian motion as follows: $V(t) = e^{-\beta t} B(\frac{\sigma^2}{2\beta} e^{2\beta t})$, $\quad t \ge 0$. [*Hint*: Compute the mean and variance of the Gaussian transition probability densities.]
9. (*Poisson Process*) Let $T_1, T_2, \ldots$ be an i.i.d. sequence of exponentially distributed random variables with intensity $\lambda > 0$, i.e., $P(T_1 > t) = e^{-\lambda t}$, $t \ge 0$. Define a counting process $N = \{N(t) : t \ge 0\}$ by $N(t) = \max\{n : T_1 + \cdots + T_n \le t\}$, $t \ge 0$. The random variables $T_1, T_2, \ldots$ are referred to as *interarrival times* of N. Show that N is a continuous parameter Markov process on the state space $S = \{0, 1, 2, \ldots\}$ with transition probabilities $p(t; x, y) = \frac{(\lambda t)^{y-x}}{(y-x)!} e^{-\lambda t}$, $y = x, x+1, \ldots, x = 0, 1, 2, \ldots, t \ge 0$. [*Hint*: N has independent increments.]
10. (*Dilogarithmic Random Walk*) The *dilogarithmic random walk*[5] is the multiplicative random walk on the multiplicative group $S = (0, \infty)$ defined by

[4] The invariant distribution of the Ornstein–Uhlenbeck process is referred to as the Maxwell–Boltzmann distribution. The physics of fluids requires that the variance be given by the physical parameter $\frac{\kappa T}{m}$ where κ is Boltzmann constant, T is absolute temperature, and m is the mass of the particle.

[5] This random walk plays a role in probabilistic analysis of the incompressible Navier–Stokes equations introduced by Y. LeJan, A. S. Sznitman (1997): Stochastic cascades and three-dimensional Navier–Stokes equations. *Probab. Theory Related Fields* 109, no. 3, 343–366. This particular structure was exploited in Dascaliuc, R., N. Michalowski, E. Thomann, E. Waymire (2015): Symmetry breaking and uniqueness for the incompressible Navier-Stokes equations, *Chaos*, American Physical

$M_n = R_0 \prod_{j=1}^{n} R_j, n = 1, 2, \ldots$ where R_0 is a positive random variable independent of the i.i.d. sequence $\{R_n : n \geq 1\}$ having marginal distribution given by $P(R_1 \in dr) = \frac{2}{\pi^2} \ln \frac{|1+r|}{|1-r|} \frac{dr}{r}, r > 0$. Show that (a) $\mathbb{E}R_1 = \infty$. (b) $\mathbb{E}|\ln R_1|^m < \infty$ for $m = 1, 2, \ldots$. (c) The distribution of M_n is symmetric about 1, the identity element of the multiplicative group S, and $\{M_n : n \geq 0\}$ is 1-neighborhood recurrent. [*Hint*: Show that the additive random walk $S_n = \ln M_n, n \geq 0$, is 0-neighborhood recurrent.]

11. Suppose that X_0 has an invariant distribution π in the sense of (13.4). Show that the Markov chain $\mathbf{X}$ is stationary (or translation invariant) in the sense that $\mathbf{X}^{n+}$ and $\mathbf{X}$ have the same distribution for each $n \geq 1$.
12. For a homogeneous continuous parameter Markov process show that the conditional distribution of $X(t+s)$ given $\sigma(X(s))$ on $[X(s) = x]$ is the same as the conditional distribution of of $X(t)$ given $X(0)$ on $[X(0) = x]$.

(Footnote 5 continued)
Society, 25 (7). The dilogarithmic functions are well-studied and arise in a variety of unrelated contexts.