1. Introduction

A **Markov process** is a random process in which the future is independent of the past, given the present. Markov processes, named for Andrei Markov are among the most important of all random processes. In a sense, they are the stochastic analogs of differential equations and recurrence relations, which are of course, among the most important deterministic processes.

Basic Theory

The Markov Property

Suppose that $\boldsymbol{X} = \{X_t : t \in T\}$ is a random process on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ where X_t is a random variable taking values in S for each $t \in T$. We think of X_t as the **state** of a system at time t. The **state space** S is usually either a countable set or a "nice" region of $\mathbb{R}^k$ for some k. The **time space** T is either $\mathbb{N}$ or $[0, \infty)$.

For $t \in T$, let $\mathscr{F}_t$ denote the σ-algebra of events generated by $\{X_s : (s \in T) \text{ and } (s \leq t)\}$. Intuitively, $\mathscr{F}_t$ contains the events that can be defined in terms of X_s for $s \leq t$. In other words, if we are allowed to observe the random variables X_s for $s \leq t$, then we can tell whether or not a given event in $\mathscr{F}_t$ has occurred.

The random process $\boldsymbol{X}$ is a **Markov process** if the following property (known as the **Markov property**) holds: For every $s \in T$ and $t \in T$ with $s < t$, and for every $H \in \mathscr{F}_s$ and $x \in S$, the conditional distribution of X_t given H and $X_s = x$ is the same as the conditional distribution of X_t just given $X_s = x$:

$$\mathbb{P}(X_t \in A \mid H, X_s = x) = \mathbb{P}(X_t \in A \mid X_s = x), \quad A \subseteq S$$

In the statement of the Markov property, think of s as the present time and hence t is a time in the future. Thus, x is the present state and H is an event that has occurred in the past. If we know the present state, then any additional knowledge of events in the past is irrelevant in terms of predicting the future.

The complexity of Markov processes depends greatly on whether the time space or the state space are discrete or continuous. In this chapter, we assume both are discrete, that is we assume that $T = \mathbb{N}$ and that S is countable (and hence the state variables have discrete distributions). In this setting, Markov processes are known as **Markov chains**. The theory of Markov chains is very beautiful and very complete.

The Markov property for a Markov chain $\boldsymbol{X} = (X_0, X_1, \ldots)$ can be stated as follows: for any sequence of states $(x_0, x_1, \ldots, x_{n-1}, x, y)$,

$$\mathbb{P}(X_{n+1} = y \mid X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x) = \mathbb{P}(X_{n+1} = y \mid X_n = x)$$

Stopping Times and the Strong Markov Property

Suppose that $\boldsymbol{X} = (X_0, X_1, X_2, \ldots)$ is a Markov chain with state space S. As before, let $\mathscr{F}_n$ is the σ-algebra generated by $(X_0, X_1, \ldots, X_n)$ for each $n \in \mathbb{N}$. A random variable T taking values in $\mathbb{N} \cup \{\infty\}$ is a **stopping time** or a **Markov**

time for $\boldsymbol{X}$ if $\{T = n\} \in \mathscr{F}_n$ for each $n \in \mathbb{N}$. Intuitively, we can tell whether or not $T = n$ by observing the chain up to time n. In some sense, a stopping time is a random time that does not require that we see into the future.

1. Suppose that T is a stopping time for $\boldsymbol{X}$. Show that for $n \in \mathbb{N}$,

   a. $\{T \le n\} \in \mathscr{F}_n$
   b. $\{T > n\} \in \mathscr{F}_n$

The quintessential example of a stopping time is the **hitting time** to a nonempty set of states A:

$$T_A = \min\{n \ge 1 : X_n \in A\}$$

where as usual, we define $\min(\varnothing) = \infty$. This random variable gives the first positive time that the chain is in A.

2. Show that T_A is a stopping time for $\boldsymbol{X}$: $\{T_A = n\} = \{X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A\}$

The **strong Markov property** states that the future is independent of the past, given the present, when the present time is a stopping time. For a Markov chain, the ordinary Markov property implies the strong Markov property.

$$\mathbb{P}(X_{T+1} = y | X_0 = x_0, X_1 = x_1, \ldots, X_T = x) = \mathbb{P}(X_{T+1} = y | X_T = x)$$

Functions and Matrices

The study of Markov chains is simplified by the use of operator notation that is analogous to operations on vectors and matrices. Suppose that $U : S \times S \to \mathbb{R}$ and $V : S \times S \to \mathbb{R}$ and that $f : S \to \mathbb{R}$ and $g : S \to \mathbb{R}$. Define $U\,V : S \times S \to \mathbb{R}$ by

$$\left(U\,V\right)\left(x, z\right) = \sum_{y \in S} U(x, y)\,V(y, z), \quad (x, y) \in S^2$$

Define $U\,f : S \to \mathbb{R}$ by

$$\left(U\,f\right)\left(x\right) = \sum_{y \in S} U(x, y)\,f(y), \quad x \in S$$

and define $f\,U : S \to \mathbb{R}$ by

$$\left(f\,U\right)\left(y\right) = \sum_{x \in S} f(x)\,U(x, y), \quad y \in S$$

Finally, we will sometimes let $f\,g$ denote the real number

$$f\,g = \sum_{x \in S} f(x)\,g(x)$$

In all of the definitions, we assume that if S is infinite then the either the terms in the sums are nonnegative or the sums converge absolutely, so that the order of the terms in the sums does not matter. We will often refer to a function $U : S \times S \to \mathbb{R}$ as a **matrix** on S. Indeed, if S is finite, then U really is a matrix in the usual sense, with rows and columns labeled by the elements of S. The product $U\,V$ is ordinary matrix multiplication; the product $U\,f$ is the product of the matrix U and the column vector f; the product $f\,U$ is the product of the row vector f and the matrix U; and the product $f\,g$ is he product of f as a row vector and g as a column vector, or equivalently, the inner product of the vectors f and g.

The sum of two matrices on S or two functions on S is defined in the usual way, as is a real multiple of a matrix or function.

In the following exercises, suppose that U, V, and W are matrices on S and that f and g are functions on S. Assume also that the sums exist.

3. Show that the **associate property** holds whenever the operations makes sense. In particular,

   a. $U\left(V W\right) = \left(U V\right) W$
   b. $U\left(V f\right) = \left(U V\right) f$ and $f\left(U V\right) = \left(f U\right) V$
   c. $f\left(U g\right) = \left(f U\right) g$,

4. Show that the **distributive property** holds whenever the operations makes sense. In particular,

   a. $U(V + W) = U V + U W$ and $(U + V) W = U W + V W$
   b. $f(U + V) = f U + f V$ and $(U + V) f = U f + V f$
   c. $U(f + g) = U f + U g$ and $(f + g) U = f U + g U$

5. The **commutative property** does not hold in general. Give examples where

   a. $U V \neq V U$
   b. $f U \neq U f$

If U is a matrix on S we denote $U^n = U U \cdots U$, the n-fold (matrix) power of U for $n \in \mathbb{N}$. By convention, $U^0 = I$, the **identity matrix** on S, defined by

$$I(x, y) = \begin{cases} 1, & x = y \\ 0, & x \neq y \end{cases}$$

If U is a matrix on S and $A \subseteq S$, we will let U_A denote the restriction of U to $A \times A$. Similarly, if f is a function on S, then f_A denotes the restriction of f to A.

Probability Functions and Matrices

If $f : S \to [0, \infty)$, then f is a discrete probability density function (or probability mass function) on S if

$$\sum_{x \in S} f(x) = 1$$

We have studied these in detail! On the other hand, if $P : S \times S \to [0, \infty)$, then P is a **transition probability matrix** or **stochastic matrix** on S if

$$\sum_{y \in S} P(x, y) = 1, \quad x \in S$$

Thus, if P is a transition matrix, then $y \mapsto P(x, y)$ is a probability density function on S for each $x \in S$. In matrix terminology, the row sums are 1.

6. Show that if P is a transition probability matrix on S then $P\,\mathbf{1} = \mathbf{1}$ where $\mathbf{1}$ is the constant function 1 on S. Thus, in the language of linear algebra, $\mathbf{1}$ is a **right eigenvector** of P, corresponding to the **eigenvalue** 1.

7. Suppose that P and Q are transition probability matrices on S and that f is a probability density function on S. Show that

a. $P\,Q$ is a transition probability matrix.
b. P^n is a transition probability matrix for each $n \in \mathbb{N}$
c. $f\,P$ is a probability density function.

8. Suppose that f is the probability density function of a random variable X taking values in S and that $g : S \to \mathbb{R}$. Show that $f\,g = \mathbb{E}(g(X))$ (assuming that the expected value exists).

A function f on S is said to be **left-invariant** for a transition probability matrix P if $f\,P = f$. In the language of linear algebra, f is a **left eigenvector** of P corresponding to the **eigenvalue** 1.

9. Show that if f is left-invariant for P then $f\,P^n = f$ for each $n \in \mathbb{N}$.

Basic Computations

Suppose again that $\boldsymbol{X} = (X_0,\ X_1,\ X_2,\ \ldots)$ is a Markov chain with state space S. For $m \in \mathbb{N}$ and $m \in \mathbb{N}$ with $m \le n$, let

$$P_{m,n}\big(x,\ y\big) = \mathbb{P}(X_n = y \big| X_m = x), \quad x \in S,\ \ y \in S$$

The matrix $P_{m,n}$ is the transition probability matrix from time m to time n. The result in the next exercise is known as the **Chapman-Kolmogorov** equation, named for Sydney Chapman and Andrei Kolmogorov. It gives the basic relationship between the transition matrices.

10. Suppose that k, m, and n are nonnegative integers with $k \le m \le n$. Use basic properties of conditional probability and the Markov property to show that

$$P_{k,m}\,P_{m,n} = P_{k,n}$$

.

It follows immediately that all of the transition probability matrices for $\boldsymbol{X}$ can be obtained from the **one-step** transition probability matrices: if m and n are nonnegative integers with $m \le n$ then

$$P_{m,n} = P_{m,m+1}\,P_{m+1,m+2}\,\cdots\,P_{n-1,n}$$

.

11. Suppose that m and n are nonnegative integers with $m \le n$. Show that if X_m has probability density function f_m, then X_n has probability density function $f_n = f_m\,P_{m,n}$.

Combining the last two results, it follows that the distribution of X_0 (the **initial distribution**) and the one-step transition matrices determine the distribution of X_n for each n. Actually, these basic quantities determine the *joint distributions* of the process, a much stronger result.

12. Suppose that X_0 has probability density function f_0. Show that for any sequence of states $(x_0,\ x_1,\ \ldots,\ x_n)$,

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = f_0(x_0) P_{0,1}(x_0, x_1) \cdots P_{n-1,n}(x_{n-1}, x_n)$$

Computations of this sort are the reason for the term *chain* in Markov chain.

Time Homogeneous Chains

A Markov chain $\boldsymbol{X} = (X_0, X_1, X_2, \ldots)$ is said to be **time homogeneous** if the transition matrix from time m to time n depends only on the difference $n - m$ for any nonnegative integers m and n with $m \le n$. That is,

$$\mathbb{P}(X_n = y | X_m = x) = \mathbb{P}(X_{n-m} = y | X_0 = x), \quad x \in S, \ y \in S$$

It follows that there is a single one-step transition probability matrix P, given by

$$P(x, y) = \mathbb{P}(X_{n+1} = y | X_n = x), \quad x \in S, \ y \in S$$

and all other transition matrices can be expressed as powers of P. Indeed if $m \le n$ then $P_{m,n} = P^{n-m}$, and the Chapman-Kolmogorov equation is simply the law of exponents for matrix powers. From Exercise 11, if X_m has probability density function f_m then $f_n = f_m P^{n-m}$ is the probability density function of X_n. In particular, if X_0 has probability density function f_0 then $f_n = f_0 P^n$ is the probability density function of X_n. The joint distribution in Exercise 12 above also simplifies: if X_0 has probability density function f_0, then for any sequence of states $(x_0, x_1, \ldots, x_n)$,

$$\mathbb{P}(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = f_0(x_0) P(x_0, x_1) \cdots P(x_{n-1}, x_n)$$

13. Suppose that X_0 has probability density function f and that f is left-invariant for P. Show that X_n has probability density function f for each $n \in \mathbb{N}$ and hence the sequence of random variables $\boldsymbol{X}$ is **identically distributed** (although certainly **not** independent in general).

In the context of the previous exercise, the probability distribution on S associated with f is said to be **invariant** or **stationary** for P or for the Markov chain $\boldsymbol{X}$. Stationary distributions turn out to be of fundamental importance in the study of the limiting behavior of Markov chains.

The assumption of time homogeneity is not as restrictive as might first appear. The following exercise shows that *any* Markov chain can be turned into a homogeneous chain by enlarging the state space with a time component.

14. Suppose that $\boldsymbol{X} = (X_0, X_1, X_2, \ldots)$ is an inhomogeneous Markov chain with state space S, and let $Q_n(x, y) = \mathbb{P}(X_{n+1} = y | X_n = x)$ denote the one-step transition probability at time n, for $x \in S$ and $y \in S$. Suppose that N_0 is a random variable taking values in $\mathbb{N}$, independent of $\boldsymbol{X}$. Let $N_n = N_0 + n$ and let $Y_n = \left(X_{N_n}, N_n\right)$ for $n \in \mathbb{N}$. Show that $\boldsymbol{Y} = (Y_0, Y_1, Y_2, \ldots)$ is a homogeneous Markov chain on $S \times \mathbb{N}$ with transition probability matrix P given by

$$P((x, m), (y, n)) = \begin{cases} Q_m(x, y), & n = m + 1 \\ 0, & \text{otherwise} \end{cases}$$

From now on, unless otherwise noted, the term *Markov chain* will mean homogeneous Markov chain.

15. Suppose that $\boldsymbol{X} = (X_0, X_1, X_2, \ldots)$ is an Markov chain with state space S and transition probability matrix P. For fixed $k \in \mathbb{N}$, show that $\left(X_0, X_k, X_{2k}, \ldots\right)$ is a Markov chain on S with transition probability matrix P^k.

The following exercise also uses the basic trick of enlarging the state space to turn a random process into a Markov chain.

16. Suppose that $\boldsymbol{X} = (X_0, X_1, X_2, \ldots)$ is a random process on S in which the future depends stochastically on the last two states. Specifically, suppose that for any sequence of states $(x_0, x_1, \ldots, x_{n-1}, x, y, z)$

$$\mathbb{P}(X_{n+2} = z \mid X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = x, X_{n+1} = y) = \mathbb{P}(X_{n+2} = z \mid X_n = x, X_{n+1} = y)$$

We also assume that this probability is independent of n, and we denote it by $Q(x, y, z)$. Let $Y_n = (X_n, X_{n+1})$ for $n \in \mathbb{N}$. Show that $\boldsymbol{Y} = (Y_0, Y_1, Y_2, \ldots)$ is a Markov chain on $S \times S$ with transition probability matrix P given by

$$P((x, y), (w, z)) = \begin{cases} Q(x, y, z), & y = w \\ 0, & \text{otherwise} \end{cases}$$

The result in the last exercise generalizes in a completely straightforward way to the case where the future of the random process depends stochastically on the last k states, for some fixed $k \in \mathbb{N}$.

Suppose again that $\boldsymbol{X} = (X_0, X_1, X_2, \ldots)$ is a Markov chain with state space S and transition probability matrix P. The **directed graph** associated with $\boldsymbol{X}$ has vertex set S and edge set $\left\{(x, y) \in S^2 : P(x, y) > 0\right\}$. That is, there is a directed edge from x to y if and only if state x leads to state y in one step. Note that the graph may well have loops, since a state can certainly lead back to itself in one step.

17. Suppose that A is a nonempty subset of S. Show that

$$P_A{}^n\left(x, y\right) = \mathbb{P}(X_1 \in A, X_2 \in A, \ldots, X_{n-1} \in A, X_n = y \mid X_0 = x), \quad (x, y) \in A^2$$

That is, $P_A{}^n(x, y)$ is the probability of going from state x to state y in n steps, remaining in A all the while. In graphical terms, it is the sum of products of probabilities along paths of length n from x to y that stay inside A. Note that in general, $P_A{}^n(x, y) \neq P^n{}_A(x, y)$.

Examples and Applications

The Two-State Chain

Perhaps the simplest, non-trivial Markov chain has two states, say $S = \{0, 1\}$ and the transition probability matrix given below, where p and q are parameters with $0 < p < 1$ and $0 < q < 1$.

$$P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

18. Show that the eigenvalues of P are 1 and $1 - p - q$.

19. Show that $B^{-1} P B = D$ where

$$B = \begin{pmatrix} 1 & -p \\ 1 & q \end{pmatrix} \text{ and } D = \begin{pmatrix} 1 & 0 \\ 0 & 1-p-q \end{pmatrix}$$

.

20. Show that

$$P^n = B\, D^n\, B^{-1} = \frac{1}{p+q}\begin{pmatrix} q + p(1-p-q)^n & p - p(1-p-q)^n \\ q - q(1-p-q)^n & p + q(1-p-q)^n \end{pmatrix}$$

21. Show that

$$P^n \to \frac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix} \text{ as } n \to \infty$$

22. Show that the only invariant probability density function for the chain is $f = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$.

In spite of its simplicity, the two state chain illustrates some of the basic limiting behavior and the connection with invariant distributions that we will study in general in a later section.

Independent Variables and Random Walks

Suppose that $\boldsymbol{X} = (X_0, X_1, X_2, \ldots)$ is a sequence of independent random variables taking values in a countable set S, and that $(X_1, X_2, X_3, \ldots)$ are identically distributed with (discrete) probability density function f.

23. Show that $\boldsymbol{X}$ is a Markov chain on S with transition probability matrix P given by $P(x, y) = f(y)$ for $x \in S$ and $y \in S$. Show also that f is invariant for P.

As a Markov chain, the process $\boldsymbol{X}$ is not very interesting, although it is *very* interesting in other ways. Suppose now that $S = \mathbb{Z}$, the set of integers, and consider the **partial sum process** (or **random walk**) $\boldsymbol{Y}$ associated with $\boldsymbol{X}$:

$$Y_n = \sum_{i=0}^{n} X_i$$

24. Show that $\boldsymbol{Y}$ is a Markov chain on $\mathbb{Z}$ with transition probability matrix Q given by $P(x, y) = f(y - x)$ for $x \in \mathbb{Z}$ and $y \in \mathbb{Z}$.

25. Consider the special case where $f(1) = p$ and $f(-1) = 1 - p$, where $0 < p < 1$. Show that the transition probability matrix is given by

$$Q(x, x-1) = 1 - p, \; Q(x, x+1) = p, \quad x \in \mathbb{Z}$$

This special case is the **simple random walk** on $\mathbb{Z}$. When $p = \frac{1}{2}$ we have the **simple, symmetric random walk**. Simple random walks are studied in more detail in the chapter on Bernoulli Trials.

Symmetric and Doubly Stochastic Matrices

A matrix P on S is **doubly stochastic** if it is nonnegative and if the row *and* columns sums are 1:

$$\sum_{u \in S} P(x, u) = 1, \; \sum_{u \in S} P(u, y) = 1, \quad x \in S, \; y \in S$$

26. Suppose that $\boldsymbol{X}$ is a Markov chain on a finite state space S with doubly stochastic transition probability matrix P. Show that the uniform distribution on S is invariant.

27. A matrix P on S is **symmetric** if $P(x, y) = P(y, x)$ for all $(x, y) \in S^2$.

a. Show that if P is a symmetric, stochastic matrix then P is doubly stochastic.
b. Give an example of a doubly stochastic matrix that is not symmetric.

Special Models

The Markov chains in the following exercises model important processes that are studied in separate sections. We will refer to these chains frequently.

28. Read the introduction to the Ehrenfest chains and work the exercises.

29. Read the introduction to the Bernoulli-Laplace chain and work the exercises.

30. Read the introduction to the reliability chains and work the exercises.

31. Read the introduction to the branching chain and work the exercises.

32. Read the introduction to the queuing chains and work the exercises.

33. Read the introduction to random walks on graphs and work the exercises.

34. Read the introduction to birth-death chains and work the exercises.