9

Markov Chains

9.1 Objectives

This chapter introduces an underappreciated topic in the social sciences that lies directly in the intersection of matrix algebra (Chapters 3 and 4) and probability (Chapter 7). It is an interesting area in its own right with many applications in the social sciences, but it is also a nice reinforcement of important principles we have already covered. Essentially the idea is relevant to the things we study in the social sciences because Markov chains specifically condition on the current status of events. Researchers find that this is a nice way to describe individual human decision making and collective human determinations.

So Markov chains are very practical and useful. They model how social and physical phenomena move from one state to another. The first part of this chapter introduces the mechanics of Markov chains through the kernel. This is the defining mechanism that "moves" the Markov chain around. The second part of the chapter describes various properties of Markov chains and how such chains can differ in how they behave. The first few properties are elementary, and the last few properties are noticeably more advanced and may be skipped by the reader who only wants an introduction.

9.2 Defining Stochastic Processes and Markov Chains

Markov chains sound like an exotic mathematical idea, but actually the principle is quite simple. Suppose that your decision-making process is based only on a current state of affairs. For example, in a casino wagering decisions are usually dictated only by the current state of the gambler: the immediate confronting decision (which number to bet on, whether to take another card, etc.) and the available amount of money. Thus these values at a previous point in time are irrelevant to future behavior (except perhaps in the psychological sense). Similarly, stock purchase decisions, military strategy, travel directions, and other such trajectories can often successfully be described with Markov chains.

What is a Markov chain? It is a special kind of stochastic process that has a "memoryless" property. That does not help much, so let us be specific. A **stochastic process** is a consecutive series of observed random variables. It is a random variable defined in exactly the standard way with a known or unknown distribution, except that the order of events is recordable. So for some state space, Θ, the random variable θ is defined by $\theta^{[t]} \sim F(\theta),\ t \in T$, where t is some index value from the set T. Actually it is almost always more simple than this general definition of indexing since it is typical to define T as the positive integers so $t = 0, 1, 2, 3, \ldots$. The implication from this simplification is that time periods are evenly spaced, and it is rare to suppose otherwise.

The state space that a stochastic process is defined on must be identified. This is exactly like the support of a probability mass function (PMF) or probability density function (PDF) in that it defines what values $\theta^{[t]}$ can take on any point in time t. There are two types of state spaces: discrete and continuous. In general, discrete state spaces are a lot more simple to think about and we will therefore focus only on these here.

Suppose that we had a cat locked in a classroom with square tiles on the floor. If we defined the room as the state space (the cat cannot leave) and each square tile as a discrete state that the cat can occupy, then the path of the cat walking throughout the room is a stochastic process where we record the grid numbers of the tiles occupied by the cat over time. Now suppose that the walking decisions made by the cat are governed only by where it is at any given moment: The cat does not care where it has been in the past. To anyone who knows cats, this seems like a reasonable assumption about feline psychology. So the cat forgets previous states that it has occupied and wanders based only on considering where it is at the moment. This property means that the stochastic process in question is now a special case called a Markov chain.

More formally, a **Markov chain** is a stochastic process with the property that any specified state in the series, $\theta^{[t]}$, is dependent only on the previous state, $\theta^{[t-1]}$. But wait, yesterday's value ($\theta^{[t-1]}$) is then also conditional on the day before that's value ($\theta^{[t-2]}$), and so on. So is it not then true that $\theta^{[t]}$ is conditional on every previous value: $0, 1, 2, \ldots, t-1$? Yes, in a sense, but *conditioned on* $\theta^{[t-1]}$, $\theta^{[t]}$ *is independent of all previous values*. This is a way of saying that all information that was used to determine the current step was contained in the previous step, and therefore if the previous step is considered, there is no additional information of importance in any of the steps before that one. This "memoryless" property can be explicitly stated in the probability language from Chapter 7. We say that $\theta^{[t]}$ is *conditionally independent* on all values previous to $\theta^{[t-1]}$ if

$$p(\theta^{[t]} \in A|\theta^{[0]}, \theta^{[1]}, \ldots, \theta^{[t-2]}, \theta^{[t-1]}) = p(\theta^{[t]} \in A|\theta^{[t-1]}),$$

where A is any identified set (an event or range of events) on the complete state space (like our set of tiles in the classroom). Or, more colloquially, "a Markov chain wanders around the state space remembering only where it has been in the last period." Now that *does* sound like a cat! A different type of stochastic process that sometimes gets mentioned in the same texts is a **martingale**. A martingale is defined using expectation instead of probability:

$$E(\theta^{[t]} \in A|\theta^{[0]}, \theta^{[1]}, \ldots, \theta^{[t-2]}, \theta^{[t-1]}) = \theta^{[t-1]},$$

meaning that the expected value that the martingale is in the set A in the next period is the value of the current position. Note that this differs from the Markov chain in that there is a stable iterative process based on this expectation rather than on Markovian probabilistic exploration.

Also, since the future at time $t+1$ and the past at time t are independent given the state at time t, the Markovian property does not care about the direction of time. This seems like a weird finding, but recall that time here is not a physical characteristic of the universe; instead it is a series of our own construction. Interestingly, there are Markov chains that are defined to work backward through "time," such as those in *coupling from the past* (see Propp and Wilson 1996).

In general interest is restricted here to discrete-time, homogeneous Markov chains. By **discrete time**, we simply mean that the counting process above, $t = 0, 1, \dots, T$, is recordable at clear, distinguishable points. There is a corresponding study of continuous-time Markov processes, but it is substantially more abstract and we will not worry about it here. A **homogeneous** Markov chain is one in which the process of moving (i.e., the *probability* of moving) is independent of current time. Stated another way, move decisions at time t are independent of t.

★ **Example 9.1: Contraception Use in Barbados**. Ebanks (1970) looked at contraception use by women of lower socio-economic class in Barbados and found a stable pattern in the 1950s and a different stable pattern emerged in the late 1960s. This is of anthropological interest because contraception and reproduction are key components of family and social life in rural areas. His focus was on the stability and change of usage, looking at a sample from family planning programs at the time. Using 405 respondents from 1955 and another 405 respondents from 1967, he produced the following change probabilities where the row indicates current state and the column indicates the next state (so, for instance, the probability of moving from "Use" at the current state to "Not Use" in the next state for 1955 is 0.52):

1955	Use	Not Use
Use	0.48	0.52
Not Use	0.08	0.92

1967	Use	Not Use
Use	0.89	0.11
Not Use	0.52	0.48

The first obvious pattern that emerges is that users in the 1950s were nearly equally likely to continue as to quit and nonusers were overwhelming likely to continue nonuse. However, the pattern is reversed in the late 1960s, whereby users were overwhelmingly likely to continue and nonusers were equally likely to continue or switch to use.

If we are willing to consider these observed (empirical) probabilities as enduring and stable indications of underlying behavior, then we can "run" a Markov chain to get anticipated future behavior. This is done very mechanically by treating the 2×2 tables here as matrices and multiplying repeatedly. What does this do? It produces expected cell values based on the probabilities in the last iteration: the Markovian property. There are (at least) two interesting things we can do here. Suppose we were interested in predicting contraception usage for 1969, that is, two years into the future. This could be done simply by the following steps:

$$\begin{bmatrix} 0.89 & 0.11 \\ 0.52 & 0.48 \end{bmatrix} \times \begin{bmatrix} 0.89 & 0.11 \\ 0.52 & 0.48 \end{bmatrix} = \underbrace{\begin{bmatrix} 0.85 & 0.15 \\ 0.71 & 0.29 \end{bmatrix}}_{1968}$$

$$\begin{bmatrix} 0.85 & 0.15 \\ 0.71 & 0.29 \end{bmatrix} \times \begin{bmatrix} 0.89 & 0.11 \\ 0.52 & 0.48 \end{bmatrix} = \underbrace{\begin{bmatrix} 0.83 & 0.17 \\ 0.78 & 0.22 \end{bmatrix}}_{1969}.$$

This means that we would *expect* to see an increase in nonusers converting to users if the 1967 rate is an underlying trend. Secondly, we can test Ebanks' assertion that the 1950s were stable. Suppose we take the 1955 matrix of transitions and apply it iteratively to get a predicted distribution across the four cells for 1960. We can then compare it to the actual distribution seen in 1960, and if it is similar, then the claim is supportable (the match will not be exact, of course, due to sampling considerations). Multiplying the 1955 matrix four times gives

$$\begin{bmatrix} 0.48 & 0.52 \\ 0.08 & 0.92 \end{bmatrix}^4 = \begin{bmatrix} 0.16 & 0.84 \\ 0.13 & 0.87 \end{bmatrix}.$$

This suggests the following empirical distribution, given the marginal numbers of users for 1959 in the study:

1959-1960 predicted	Use	Not Use
Use	7	41
Not Use	46	311

which can be compared with the actual 1960 numbers from that study:

1959-1960 actual	Use	Not Use
Use	27	21
Not Use	39	318

These are clearly dissimilar enough to suggest that the process is not Markovian as claimed. More accurately, it can perhaps be described as a martingale since the 1955 actual numbers are $\left[\begin{smallmatrix} 24 & 26 \\ 28 & 327 \end{smallmatrix}\right]$. What we are actually seeing here is a question of whether the 1955 proportions define a transaction kernel for the Markov chain going forward. The idea of a transaction kernel is explored more fully in the next section.

9.2.1 The Markov Chain Kernel

We know that a Markov chain moves based only on its current position, but using that information, how does the Markov chain decide? Every Markov chain is defined by two things: its state space (already discussed) and its **transition kernel**, $K()$. The transition kernel is a general mechanism for describing the probability of moving to other states based on the current chain status. Specifically, $K(\theta, A)$ is a defined probability measure for all θ points in the state space to the set $A \in \Theta$: It maps potential transition events to their probability of occurrence.

The easiest case to understand is when the state space is discrete and $\mathbf{K}$ is just a matrix mapping: a $k \times k$ matrix for k discrete elements in that exhaust the allowable state space, A. We will use the notation θ_i, meaning the ith state of the space. So a Markov chain that occupies subspace i at time t is designated $\theta_i^{[t]}$.

Each individual cell defines the probability of a state transition from the first term to all possible states:

$$\mathbf{K} = \begin{bmatrix} p(\theta_1, \theta_1) & p(\theta_1, \theta_2) & \dots & p(\theta_1, \theta_{k-1}) & p(\theta_1, \theta_k) \\ p(\theta_2, \theta_1) & p(\theta_2, \theta_2) & \dots & p(\theta_2, \theta_{k-1}) & p(\theta_2, \theta_k) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ p(\theta_{k-1}, \theta_1) & p(\theta_{k-1}, \theta_2) & & p(\theta_{k-1}, \theta_{k-1}) & p(\theta_{k-1}, \theta_k) \\ p(\theta_k, \theta_1) & p(\theta_k, \theta_2) & \dots & p(\theta_1, \theta_{k-1}) & p(\theta_k, \theta_k) \end{bmatrix}.$$

The first term in $p()$, constant across rows, indicates where the chain is at the current period and the column indicates potential destinations. Each matrix element is a well-behaved probability, $p(\theta_i, \theta_j) \geq 0, \ \forall i, j \in A$. The notation here can be a little bit confusing as it looks like a *joint distribution* in the sense of Chapter 7. This is an unfortunate artifact, and one just has to remember the different context. The rows of $\mathbf{K}$ sum to one and define a conditional PMF because they are all specified for the same starting value and cover each possible destination in the state space.

We can also use this kernel to calculate state probabilities for arbitrary future times. If we multiply the transition matrix (kernel) by itself j times, the result,

$\mathbf{K}^j$, gives the **j-step transition matrix** for this Markov chain. Each row is the set of transition probabilities from that row state to each of the other states in exactly j iterations of the chain. It does not say, however, what that sequence is in exact steps.

★ **Example 9.2: Campaign Contributions**. It is no secret that individuals who have contributed to a Congress member's campaign in the past are more likely than others to contribute in the next campaign cycle. This is why politicians keep and value donor lists, even including those who have given only small amounts in the past. Suppose that 25% of those contributing in the past to a given member are likely to do so again and only 3% of those not giving in the past are likely to do so now. The resulting transition matrix is denoted as follows:

$$\begin{array}{ccc} & & \overbrace{\begin{array}{cc} \theta_1 & \theta_2 \end{array}}^{\text{current period}} \\ \text{last period} \left\{ \begin{array}{c} \theta_1 \\ \theta_2 \end{array} \right. & & \left[\begin{array}{cc} 0.97 & 0.03 \\ 0.75 & 0.25 \end{array} \right] \end{array},$$

where θ_1 is the state for no contribution and θ_2 denotes a contribution. Notice that the rows necessarily add to one because there are only two states possible in this space. This articulated kernel allows us to ask some useful questions on this candidate's behalf. If we start with a list of 100 names where 50 of them contributed last period and 50 did not, what number can we expect to have contribute from this list? In Markov chain language this is called a starting point or starting vector:

$$S_0 = \left[\begin{array}{cc} 50 & 50 \end{array} \right];$$

that is, before running the Markov chain, half of the group falls in each category. To get to the Markov chain first state, we simply multiply the initial state by the transition matrix:

$$S_1 = \begin{bmatrix} 50 & 50 \end{bmatrix} \begin{bmatrix} 0.97 & 0.03 \\ 0.75 & 0.25 \end{bmatrix} = \begin{bmatrix} 86 & 14 \end{bmatrix} = S_1.$$

So we would expect to get contributions from 14 off of this list. Since incumbent members of Congress enjoy a repeated electoral advantage for a number of reasons, let us assume that our member runs more consecutive races (and wins!). If we keep track of this particular list over time (maybe they are especially wealthy or influential constituents), what happens to our expected number of contributors? We can keep moving the Markov chain forward in time to find out:

$$\text{Second state:} \quad S_2 = \begin{bmatrix} 86 & 14 \end{bmatrix} \begin{bmatrix} 0.97 & 0.03 \\ 0.75 & 0.25 \end{bmatrix} = \begin{bmatrix} 94 & 6 \end{bmatrix}$$

$$\text{Third state:} \quad S_3 = \begin{bmatrix} 94 & 6 \end{bmatrix} \begin{bmatrix} 0.97 & 0.03 \\ 0.75 & 0.25 \end{bmatrix} = \begin{bmatrix} 96 & 4 \end{bmatrix}$$

$$\text{Fourth state:} \quad S_4 = \begin{bmatrix} 96 & 4 \end{bmatrix} \begin{bmatrix} 0.97 & 0.03 \\ 0.75 & 0.25 \end{bmatrix} = \begin{bmatrix} 96 & 4 \end{bmatrix}.$$

We rounded to integer values at each step since by definition donors can only give or not give. It turns out that no matter how many times we run this chain forward from this point, the returned state will always be $[96, 4]$, indicating an overall 4% donation rate. This turns out to be a very important property of Markov chains and the subject of the next section.

In fact, for this simple example we could solve directly for the steady state $S = [s_1, s_2]$ by stipulating

$$\begin{bmatrix} s_1 & s_2 \end{bmatrix} \begin{bmatrix} 0.97 & 0.03 \\ 0.75 & 0.25 \end{bmatrix} = \begin{bmatrix} s_1 & s_2 \end{bmatrix}$$

and solving the resulting two equations for the two unknowns:

$$[0.97s_1 + 0.75s_2, 0.03s_1 + 0.25s_2] = [s_1, s_2]$$

$$0.97s_1 + 0.75s_2 = s_1 \qquad 0.03s_1 + 0.25s_2 = s_2$$
$$s_1 = \frac{7500}{78} \qquad s_2 = \frac{300}{78}$$

(using $s_1 + s_2 = 100$), where the difference from above is due to rounding.

9.2.2 The Stationary Distribution of a Markov Chain

Markov chains can have a **stationary distribution**: a distribution reached from iterating the chain until some point in the future where all movement probabilities are governed by a single probabilistic statement, regardless of time or position. This is equivalent to saying that when a Markov chain has reached its stationary distribution there is a single marginal distribution rather than the conditional distributions in the transition kernel.

To be specific, define $\pi(\theta)$ as the stationary distribution of the Markov chain for θ on the state space A. Recall that $p(\theta_i, \theta_j)$ is the probability that the chain will move from θ_i to θ_j at some arbitrary step t, and $\pi^t(\theta)$ is the corresponding marginal distribution. The stationary distribution satisfies

$$\sum_{\theta_i} \pi^t(\theta_i)p(\theta_i, \theta_j) = \pi^{t+1}(\theta_j).$$

This is very useful because we *want* the Markov chain to reach and describe a given marginal distribution; then it is only necessary to specify a transition kernel and let the chain run until the probability structures match the desired marginal.

★ **Example 9.3: Shuffling Cards**. We will see here if we can use a Markov chain algorithm to shuffle a deck of cards such that the marginal distribution is *uniform*: Each card is equally likely to be in any given position. So the objective (stationary distribution) is a uniformly random distribution in the deck: The probability of any one card occupying any one position is $1/52$.

The suggested algorithm is to take the top card and insert it uniformly randomly at some other point in the deck, and continue. Is this actually a Markov chain? What is the stationary distribution and is it the uniform distribution as desired? Bayer and Diaconis (1992) evaluated a number of these shuffling algorithms from a technical perspective.

To answer these questions, we simplify the problem (without loss of generality) to consideration of a deck of only three cards numbered $1, 2, 3$. The sample space for this setup is then given by

$$A = \{[1, 2, 3], [1, 3, 2], [2, 1, 3], [2, 3, 1], [3, 1, 2], [3, 2, 1]\},$$

which has $3! = 6$ elements from the counting rules given in Chapter 7. A sample chain trajectory looks like

$$\begin{gathered}[1{,}3, 2]\\ [3{,}1, 2]\\ [1{,}3, 2]\\ [3{,}2, 1]\\ \vdots\end{gathered}$$

Looking at the second step, we took the 3 card off the top of the deck and picked the second position from among three. Knowing that the current position is currently $[3, 1, 2]$, the probabilities and potential outcomes are given by

Action	Outcome	Probability
return to top of deck	$[3, 1, 2]$	$\frac{1}{3}$
put in middle position	$[1, 3, 2]$	$\frac{1}{3}$
put in bottom position	$[1, 2, 3]$	$\frac{1}{3}$

To establish the potential outcomes we only need to know the current position of the deck and the probability structure (the kernel) here. Being aware of the position of the deck at time $t = 2$ tells us everything we need to know about the deck, and having this information means that knowing that the position of the deck at time $t = 1$ was $[1, 3, 2]$ is irrelevant to calculating the potential outcomes and their probabilities in the table above. So once the current position is established in the Markov chain, decisions about where to go are conditionally independent of the past.

It should also be clear so far that not every position of the deck is immediately reachable from every other position. For instance, we cannot move directly from $[1, 3, 2]$ to $[3, 2, 1]$ because it would require at least one additional step.

The transition kernel assigns positive (uniform) probability from each state to each reachable state in one step and zero probability to all other states:

$$K = \begin{bmatrix} 1/3 & 0 & 1/3 & 1/3 & 0 & 0 \\ 0 & 1/3 & 0 & 0 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\ 0 & 0 & 1/3 & 1/3 & 0 & 1/3 \end{bmatrix}.$$

Let us begin with a starting point at $[1, 2, 3]$ and look at the marginal distribution after each application of the transition kernel. Mechanically, we do this by pre-multiplying the transition kernel matrix by $[1, 0, 0, 0, 0, 0]$ as the starting probability (i.e., a deterministic decision to begin at the specified point. Then we record the result, multiply it by the kernel, and continue. The first 15 iterations produce the following marginal probability vectors:

Iteration	p([1,2,3])	p([1,3,2])	p([2,1,3])	p([2,3,1])	p([3,1,2])	p([3,2,1])
1	0.222222	0.111111	0.222222	0.222222	0.111111	0.111111
2	0.185185	0.148148	0.185185	0.185185	0.148148	0.148148
3	0.172839	0.160493	0.172839	0.172839	0.160493	0.160493
4	0.168724	0.164609	0.168724	0.168724	0.164609	0.164609
5	0.167352	0.165980	0.167352	0.167352	0.165980	0.165980
6	0.166895	0.166438	0.166895	0.166895	0.166438	0.166438
7	0.166742	0.166590	0.166742	0.166742	0.166590	0.166590
8	0.166692	0.166641	0.166692	0.166692	0.166641	0.166641
9	0.166675	0.166658	0.166675	0.166675	0.166658	0.166658
10	0.166669	0.166663	0.166669	0.166669	0.166663	0.166663
11	0.166667	0.166665	0.166667	0.166667	0.166665	0.166665
12	0.166667	0.166666	0.166667	0.166667	0.166666	0.166666
13	0.166666	0.166666	0.166666	0.166666	0.166666	0.166666
14	0.166666	0.166666	0.166666	0.166666	0.166666	0.166666
15	0.166666	0.166666	0.166666	0.166666	0.166666	0.166666

Clearly there is a sense that the probability structure of the marginals have converged to a uniform pattern. This is as we expected, and it means that the shuffling algorithm will *eventually* produce the desired marginal probabilities. It is important to remember that these stationary probabilities are not the probabilities that govern movement at any particular point in the chain; that is still the kernel. These are the probabilities of seeing one of the six events at any arbitrary point in time unconditional on current placement.

This difference may be a little bit subtle. Recall that there are three unavailable outcomes and three equal probability outcomes for each position of the deck. So the marginal distribution above *cannot* be functional as a state to state transition mechanism. What is the best guess as to the unconditional state of the deck in 10,000 shuffles? It is equally likely that any of the six states would be observed. But notice that this question ignores the state of the deck in $9,999$ shuffles. If we do not have this information, then the marginal distribution (if known) is the best way to describe outcome probabilities because it is the long-run probability of the states once the Markov chain is in its stationary distribution.

9.3 Properties of Markov Chains

Markov chains have various properties that govern how they behave as they move around their state spaces. These properties are important because they determine whether or not the Markov chain is producing values that are useful to our more general purpose.

9.3.1 Homogeneity and Periodicity

A Markov chain is said to be homogeneous at some step t if the transition probabilities at this step do not depend on the value of t. This definition implies that Markov chains can be homogeneous for some periods and **non-homogeneous** for other periods. The homogeneity property is usually important in that Markov chains that behave according to some function of their age are usually poor theoretical tools for exploring probability statements of interest.

A related, and important, property is the **period** of a Markov chain. If a Markov chain operates on a deterministic repeating schedule of steps, then it is said to be a Markov chain of **period-n**, where n is the time (i.e., the number of steps) in the reoccurring period. It seems fairly obvious that a periodic Markov chain is not a homogeneous Markov chain because the period implies a dependency of the chain on the time t.

9.3.1.1 A Simple Illustration of Time Dependency

As an illustration of homogeneity and periodicity, consider a simple Markov chain operating on a discrete state space with only four states, $\theta\colon\ 1, 2, 3, 4$, illustrated by

$$\begin{array}{cccc} \theta_1 & \theta_2 & \theta_3 & \theta_4 \\ \boxed{1} & \boxed{2} & \boxed{3} & \boxed{4} \end{array}$$

The movement of the Markov chain at time t is governed by the rule

$$p(s=1)=\frac{1}{10} \qquad p(s=2)=\frac{2}{10}$$
$$p(s=3)=\frac{3}{10} \qquad p(s=4)=\frac{4}{10},$$

where s is the number of steps that the chain moves to the right. The state space wraps around from 4 back to 1 so that $s=4$ means that the chain returns to the same place. Does this chain actually have the Markovian property? Certainly it does, because movement is dictated only by the current location and the stipulated kernel. Is it periodic or homogeneous? Since there is no repetition or dependency on t in the kernel it is clearly both aperiodic and nonhomogeneous.

How could this Markov chain be made to be periodic? Suppose that the kernel above was replaced by the cycling rule:

$$[1,2,3,4,4,3,2,1,\ 1,2,3,4,4,3,2,1,\ 1,2,3,4,4,3,2,1,\ldots],$$

then the period would be 8 and the chain would repeat this pattern forever. While the path of the chain depends on the time in the sense that the next deterministic step results from the current deterministic step, we usually just call this type of chain periodic rather than nonhomogeneous. This is because periodicity is more damaging to stochastic simulation with Markov chains than nonhomogeneity. To see the difference, consider the following kernel, which gives a nonhomogeneous Markov chain that is *not* periodic:

$$\theta <[t+1] = runif[(t-1){:}t] \bmod 4 + 1,$$

where the notation $runif[(t-1){:}t]$ here means a random uniform choice of integer between $t-1$ and t. So this chain does not have a repeating period, but it is clearly dependent on the current value of t. For example, now run this chain for 10 iterations from a starting point at 1:

period	0	1	2	3	4	5	6	7	8	9	10
$runif[(t-1){:}t]$	-	2	2	3	5	6	7	7	9	9	10
θ_t	1	2	3	3	4	2	3	4	4	2	2

There is a built-in periodicity to this chain where lower values are more likely just after t reaches a multiple of 4, and higher values are more likely just before t reaches a multiple of 4.

There is one last house-keeping detail left for this section. Markov chains are generally implemented with computers, and the underlying random numbers generated on computers have two characteristics that make them not truly random. For one thing, the generation process is fully discrete since the values are created from a finite binary process and normalized through division. This means that these are *pseudo-random* numbers and necessarily rational. Yet, truly random numbers on some defined interval are irrational with probability one because the irrationals dominate the continuous metric. In addition, while we call these values random or more accurately pseudorandom numbers, they are not random at all because the process that generates them is completely deterministic. The algorithms create a stream of values that is not random in the indeterminant sense but still *resembles a random process*. The deterministic streams varyingly lack systematic characteristics (Coveyou 1960): the time it takes to repeat the stream exactly (the period) and repeated patterns in lagged sets within the stream. So, by necessity, algorithmic implementations have to live with periodicity, and it is worth the time and energy in applied settings to use the available random number generator with the longest period possible.

9.3.2 Irreducibility

A *state* A in the state space of a Markov chain is **irreducible** if for every two substates or individual events θ_i and θ_j in A, these two substates "communicate." This means that the Markov chain is irreducible on A if every reached point or collection of points can be reached from every other reached point or collection of points. More formally,

$$p(\theta_i, \theta_j) \neq 0, \; \forall \theta_i, \theta_j \in A.$$

As an example, consider the following kernel of a **reducible** Markov chain:

$$K = \begin{array}{c c} & \begin{array}{cccc} \theta_1 & \theta_2 & \theta_3 & \theta_4 \end{array} \\ \begin{array}{c} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{array} & \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix} \end{array}.$$

The chain determined by this kernel is reducible because if is started in either θ_1 or θ_2, then it operates as if has the transition matrix:

$$K_{1,2} = \begin{array}{c c} & \begin{array}{cc} \theta_1 & \theta_2 \end{array} \\ \begin{array}{c} \theta_1 \\ \theta_2 \end{array} & \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \end{pmatrix} \end{array},$$

and if it is started in either θ_3 or θ_4, then it operates as if it has the transition matrix:

$$K_{3,4} = \begin{array}{c c} & \begin{array}{cc} \theta_3 & \theta_4 \end{array} \\ \begin{array}{c} \theta_3 \\ \theta_4 \end{array} & \begin{pmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix} \end{array}.$$

Thus the original Markov chain determined by K is *reducible* to one of two forms, depending on where it is started because there are in each case two permanently unavailable states. To provide a contrast, consider the Markov chain determined by the following kernel, K', which is very similar to K but is irreducible:

$$K = \begin{array}{c} \\ \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{array} \begin{array}{c} \begin{array}{cccc} \theta_1 & \theta_2 & \theta_3 & \theta_4 \end{array} \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 \\ 0 & \frac{3}{4} & 0 & \frac{1}{4} \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix} \end{array}.$$

This occurs because there is now a two-way "path" between the previously separated upper and lower submatrices.

Related to this is the idea of **hitting times**. The hitting time of a state A and a Markov chain θ is the shortest time for the Markov chain to begin in A and return to A:

$$T_A = \text{inf}[n > 0, \theta^{[n]} \in A].$$

Recall that the notation "inf" from Chapter 1 means the lowest (positive) value of n that satisfies $\theta^{[n]} \in A$. It is conventional to define T_A as infinity if the Markov chain never returns to A. From this definition we get the following important result:

An irreducible and aperiodic Markov chain on state space A will for each subspace of A, a_i, have a finite hitting time for a_i with probability one and a finite expected value of the hitting time:

$$p(T_{a_i} < \infty) = 1, \qquad E[T_{a_i}] < \infty.$$

By extension we can also say that the probability of transitioning from two substates of A, a_i and a_j, in finite time is guaranteed to be nonzero.

9.3.3 Recurrence

Some characteristics belong to *states* rather than Markov chains. Of course Markov chains operating on these states are affected by such characteristics. A state A is said to be **absorbing** if once a Markov chain enters this state it cannot leave: $p(A, A^{\complement}) = 0$. Conversely, A is **transient** if the probability of the chain not returning to this state is nonzero: $1 - p(A, A) > 0$. This is equivalent to saying that the chain will return to A for only a finite number of visits in infinite time. State A is said to be **closed** to another state B if a Markov chain on A cannot reach B: $p(A, B) = 0$. State A is clearly closed *in general* if it is absorbing since $B = A^{\complement}$ in this case (note that this is a different definition of "closed" used in a different context than that in Chapter 4).

These properties of states allow us to define an especially useful characteristic of both states and chains. If a state is closed, discrete, and irreducible, then this state and all subspaces within this subspace are called **recurrent**, and Markov chains operating on recurrent state spaces are recurrent. From this we can say something important in two different ways:

- **[Formal Definition.]** A irreducible Markov chain is called *recurrent* with regard to a given state A, which is a single point or a defined collection of points, if the probability that the chain occupies A infinitely often over unbounded time is nonzero.
- **[Colloquial Definition.]** When a chain moves into a recurrent state, it stays there forever and visits every subspace infinitely often.

There are also exactly two different, mutually exclusive, "flavors" of recurrence with regard to a state A:

- A Markov chain is **positive recurrent** if the mean time to return to A is bounded.
- Otherwise the mean time to return to A is infinite, and the Markov chain is called **null recurrent**.

With these we can also state the following properties.

Properties of Markov Chain Recurrence

$\rightarrowtail$ **Unions** If A and B are recurrent states, then $A \cup B$ is a recurrent state

$\rightarrowtail$ **Capture** A chain that enters a closed, irreducible, and recurrent state stays there and visits every substate with probability one.

★ **Example 9.4: Conflict and Cooperation in Rural Andean Communities**. Robbins and Robbins (1979) extended Whyte's (1975) study of 12 Peruvian communities by extrapolating future probabilities of conflict and cooperation using a Markov chain analysis. Whyte classified these communities in 1964 and 1969 as having one of four types of relations with the other communities: high cooperation and high conflict (HcHx), high cooperation

Table 9.1. Distribution of Community Types

Community	Type in 1964	Type in 1969
Huayopampa	HcLx	HcHx
Pacaraos	HcHx	HcHx
Aucallama	LcHx	LcLx
La Esperanza	LcLx	LcLx
Pucará	LcHx	LcLx
St. A De Caias	LcHx	LcHx
Mito	LcHx	LcLx
Virú	LcHx	LcLx
Pisaq	LcHx	LcHx
Kuyo Chico	HcLx	HcLx
Maska	HcLx	HcLx
Qotobamba	HcLx	HcLx

and low conflict (HcLx), low cooperation and high conflict (LcHx), or low cooperation and low conflict (LcLx). The interesting questions were, what patterns emerged as these communities changed (or not) over the five-year period since conflict and cooperation can exist simultaneously but not easily. The states of these communities at the two points in time are given in Table 9.1.

So if we are willing extrapolate these changes as Robbins and Robbins did by assuming that "present trends continue," then a Markov chain transition matrix can be constructed from the empirically observed changes between 1964 and 1969. This is given the following matrix, where the rows indicate 1964 starting points and the columns are 1969 outcomes:

$$\begin{array}{cc} & \begin{array}{cccc} \text{HcHx} & \text{HcLx} & \text{LcHx} & \text{LcLx} \end{array} \\ \begin{array}{c} \text{HcHx} \\ \text{HcLx} \\ \text{LcHx} \\ \text{LcLx} \end{array} & \begin{pmatrix} 1.00 & 0.00 & 0.00 & 0.00 \\ 0.25 & 0.75 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.33 & 0.67 \\ 0.00 & 0.00 & 0.00 & 1.00 \end{pmatrix} \end{array}.$$

The first thing we can notice is that HcHx and LcLx are both absorbing states as described above: Once the Markov chain reaches these states it never leaves. Clearly this means that the Markov chain is not irreducible because there are states that cannot "communicate." Interestingly, there are two noncommunicating state spaces given by the 2×2 upper left and lower right submatrices. Intuitively it seems that any community that starts out as HcHx or HcLx ends up as HcHx (upper left), and any community that starts out as LcHx or LcLx ends up as LcLx (lower right). We can test this by running the Markov chain for some reasonable number of iterations and observing the limiting behavior. It turns out that it takes about 25 iterations (i.e., 25 five-year periods under the assumptions since the 0.75 value is quite persistent) for this limiting behavior to converge to the state:

$$\begin{array}{cc} & \begin{array}{cccc} \mathbf{HcHx} & \mathbf{HcLx} & \mathbf{LcHx} & \mathbf{LcLx} \end{array} \\ \begin{array}{c} \mathbf{HcHx} \\ \mathbf{HcLx} \\ \mathbf{LcHx} \\ \mathbf{LcLx} \end{array} & \begin{pmatrix} 1.00 & 0.00 & 0.00 & 0.00 \\ 1.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 1.00 \\ 0.00 & 0.00 & 0.00 & 1.00 \end{pmatrix}, \end{array}$$

but once it does, it never changes. This is called the stationary distribution of the Markov chain and is now formally defined.

9.3.4 Stationarity and Ergodicity

In many applications a stochastic process eventually converges to a single limiting value and stays at that value permanently. It should be clear that a Markov chain *cannot* do that because it will by definition continue to move about the parameter space. Instead we are interested in the *distribution* that the Markov chain will eventually settle into. Actually, these chains do not *have* to converge in distribution, and some Markov chains will wander endlessly without pattern or prediction. Fortunately, we know some criteria that provide for Markov chain convergence.

First, define a marginal distribution of a Markov chain. For a Markov chain operating on a discrete state space, the marginal distribution of the chain at the m step is obtained by inserting the current value of the chain, $\theta_i^{[m]}$, into the row of the transition kernel for the mth step, p^m:

$$p^m(\theta) = [p^m(\theta_1), p^m(\theta_2), \ldots, p^m(\theta_k)].$$

So the marginal distribution at the very first step of the discrete Markov chain is given by $p^1(\theta) = p^1\pi^0(\theta)$, where p^0 is the initial starting value assigned to the chain and $p^1 = p$ is a transition matrix. The marginal distribution at some (possibly distant) step for a given starting value is

$$p^n = pp^{n-1} = p(pp^{n-2}) = p^2(pp^{n-3}) = \ldots = p^n p^0.$$

Since successive products of probabilities quickly result in lower probability values, the property above shows how Markov chains eventually "forget" their starting points.

Now we are prepared to define stationarity. Recall that $p(\theta_i, \theta_j)$ is the probability that the chain will move from θ_i to θ_j at some arbitrary step t, and $\pi^t(\theta)$ is the corresponding marginal distribution. Define $\pi(\theta)$ as the stationary distribution (a well-behaved probability function in the Kolmogorov sense) of the Markov chain for θ on the state space A, if it satisfies

$$\sum_{\theta_i} \pi^t(\theta_i) p(\theta_i, \theta_j) = \pi^{t+1}(\theta_j).$$

The key point is that the marginal distribution remains fixed when the chain reaches the stationary distribution, and we might as well drop the superscript designation for iteration number and just use $\pi(\theta)$; in shorthand, $\pi = \pi p$. Once the chain reaches its stationary distribution, it stays in this distribution and moves about, or "mixes," throughout the subspace according to marginal distribution, $\pi(\theta)$, indefinitely. The key theorem is

> *An irreducible and aperiodic Markov chain will eventually converge to a stationary distribution, and this stationary distribution is unique.*

Here the recurrence gives the range restriction property whereas stationarity gives the constancy of the probability structure that dictates movement.

As you might have noticed by now in this chapter, Markov chain theory is full of new terminology. The type of chain just discussed is important enough to warrant its own name: If a chain is recurrent and aperiodic, then we call it **ergodic**, and ergodic Markov chains with transition kernel K have the property

$$\lim_{n\to\infty} K^n(\theta_i, \theta_j) = \pi(\theta_j),$$

for all θ_i and θ_j in the subspace What does this actually means? *Once an ergodic Markov chain reaches stationarity, the resulting values are all from the distribution* $\pi(\theta_i)$. The Ergodic Theorem given above is the equivalent of the strong law of large numbers but instead for Markov chains, since it states that any specified function of the posterior distribution can be estimated with samples from a Markov chain in its ergodic state because averages of sample values give strongly consistent parameter estimates.

The big deal about ergodicity and its relationship to stationarity comes from the important **ergodic theory**. This essentially states that, given the right conditions, we can collect empirical evidence from the Markov chain values in lieu of analytical calculations. Specifically

> *If* θ_n *is a positive recurrent, irreducible Markov chain with stationary distribution given by* $\pi(\theta)$*, then*
>
> $$\lim_{n\to\infty} \frac{1}{n} \sum f(\theta_n) = \sum_{\Theta} f(\theta)\pi(\theta).$$

Specifically, this means that empirical averages for the function $f()$ converge to probabilistic averages. This is the justification for using Markov chains to approximate difficult analytical quantities, thus replacing human effort with automated effort (at least to some degree!).

★ **Example 9.5: Population Migration Within Malawi**. Discrete Markov chains are enormously useful in describing movements of populations, and demographers often use them in this way. As an example Segal (1985) looked at population movements between Malawi's three administrative regions from 1976 to 1977. The Republic of Malawi is a narrow, extended south African country of 45,745 square miles wrapped around the eastern and southern parts of Lake Malawi. Segal took observed migration numbers to create a transition matrix for future movements under the assumption of stability. This is given by

$$\begin{array}{ccc} & & \text{Destination} \\ & & \begin{array}{ccc} \scriptstyle\text{Northern} & \scriptstyle\text{Central} & \scriptstyle\text{Southern} \end{array} \\ \text{Source} & \begin{array}{c} \scriptstyle\text{Northern} \\ \scriptstyle\text{Central} \\ \scriptstyle\text{Southern} \end{array} & \begin{pmatrix} 0.970 & 0.019 & 0.012 \\ 0.005 & 0.983 & 0.012 \\ 0.004 & 0.014 & 0.982 \end{pmatrix} \end{array}.$$

It is important to note substantively that using this transition matrix to make future predictions about migration patterns ignores the possibility of major shocks to the system such as pandemics, prolonged droughts, and political upheaval. Nonetheless, it is interesting, and sometimes important, to anticipate population changes and the subsequent national policy issues.

The obvious question is whether the transition matrix above defines an ergodic Markov chain. Since this is a discrete transition kernel, we need to assert that it is recurrent and aperiodic. This is a particularly simple example because recurrence comes from the lack of zero probability values in the matrix. Although the presence of zero probability values alone would not be proof of nonreccurence, the lack of any shows that all states communicate with nonzero probabilities and thus recurrence is obvious. Note that there is also no mechanism to impose a cycling effect through the cell values so aperiodicity is also apparent. Therefore this transition kernel defines an ergodic Markov chain that must then have a unique stationary distribution.

While there is no proof of stationarity, long periods of unchanging marginal probabilities are typically a good sign, especially with such a simple and well-behaved case. The resulting stationary distribution after multiplying the transition kernel 600 times is

$$\begin{array}{ccc} & \text{Destination} & \\ \text{Northern} & \text{Central} & \text{Southern} \\ \end{array}$$
$$\begin{pmatrix} 0.1315539 & 0.4728313 & 0.3956149 \end{pmatrix}.$$

We can actually run the Markov chain much longer without any real trouble, but the resulting stationary distribution will remain unchanged from this result. This is a really interesting finding, though. Looking at the original transition matrix, there is a strong inclination to stay in the same region of Malawi for each of the three regions (the smallest has probability 0.97), yet in the limiting distribution there is a markedly different result, with migration to the Central Region being almost 50%. Perhaps more surprisingly, even though there is a 0.97 probability of remaining in the Northern Region for those starting there on any given cycle, the long-run probability of remaining in the Northern Region is only 0.13.

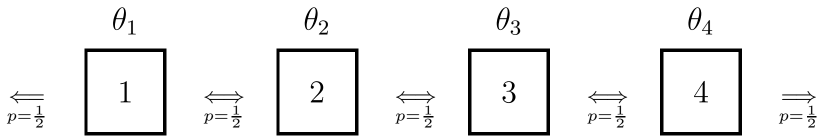
9.3.5 Reversibility

Some Markov chains are **reversible** in that they perform the same run backward as forward. More specifically, if $p(\theta_i, \theta_j)$ is a single probability from a transition kernel K and $\pi(\theta)$ is a marginal distribution, then the Markov chain is reversible if it meets the condition

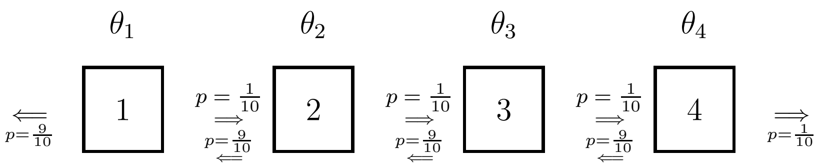
$$p(\theta_i, \theta_j)\pi(\theta_i) = p(\theta_j, \theta_i)\pi(\theta_j).$$

This expression is called both the **reversibility condition** and the **detailed balance equation**. What this means is that the *distribution* of θ at time $t+1$ conditioned on the *value* of θ at time t is the same as the *distribution* of θ at time t conditioned on the *value* of θ at time $t+1$. Thus, for a reversible Markov chain the direction of time is irrelevant to its probability structure.

As an example of reversibility, we modify a previous example where the probability of transitioning between adjacent states for a four-state system is determined by flipping a fair coin (states 1 and 4 are assumed adjacent to complete the system):



It should be clear that the stationary distribution of this system is uniform across the four events, and that it is guaranteed to reach it since it is recurrent and aperiodic. Suppose we modify the transition rule to be asymmetric from every point, according to



So what we have now is a chain that strongly prefers to move left at every step by the same probability. This Markov chain will also lead to a uniform stationary distribution, because it is clearly still recurrent and aperiodic. It is, however, clearly not reversible anymore because for adjacent θ values (i.e., those with nonzero transition probabilities)

$$p(\theta_i, \theta_j)\pi(\theta_i) \neq p(\theta_j, \theta_i)\pi(\theta_j), \; i < j$$
$$\frac{1}{10}\frac{1}{4} \neq \frac{9}{10}\frac{1}{4}$$

(where we say that $4 < 1$ by assumption to complete the system).

9.4 New Terminology

absorbing, 410

closed, 410

detailed balance equation, 417

discrete time, 395

ergodic theory, 415

ergodic, 415

hitting times, 409

homogeneous, 395

irreducible, 407

j-step transition matrix, 399

Markov chain, 394

martingale, 394

non-homogeneous, 405

null recurrent, 410

period-n, 405

period, 405

positive, 410

recurrent, 410

reducible, 408

reversibility condition, 417

reversible, 417

stationary distribution, 401

stochastic process, 393

transient, 410

transition kernel, 398

Exercises

9.1 Consider a lone knight on a chessboard making moves uniformly randomly from those legally available to it at each step. Show that the path of the knight (starting in a corner) is or is not a Markov chain. If so, is it irreducible and aperiodic?

9.2 For the following matrix, fill in the missing values that make it a valid transition matrix:

$$\begin{bmatrix} 0.1 & 0.2 & 0.3 & \\ 0.9 & & 0.01 & 0.01 \\ 0.0 & 0.0 & & 0.0 \\ & 0.2 & 0.2 & 0.2 \end{bmatrix}.$$

9.3 Using this matrix:

$$\mathbf{X} = \begin{bmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix},$$

find the vector of stationary probabilities.

9.4 Consider a discrete state space with only two events: 0 and 1. A stochastic process operates on this space with probability of transition one-half for moving or staying in place. Show that this is or is not a Markov chain.

9.5 There are many applications and famous problems for Markov chains that are related to gambling. For example, suppose a gambler bets \$1 on successive games unless she has won three in a row. In the latter case she bets \$3 but returns to \$1 if this game is lost. Does this dependency on more than the last value ruin the Markovian property? Can this process be made to depend only on a previous "event"?

9.6 One urn contains 10 black marbles and another contains 5 white marbles. At each iteration of an iterative process 1 marble is picked from each urn and swapped with probability $p = 0.5$ or returned to its original urn with probability $1 - p = 0.5$. Give the transition matrix for the process and show that it is Markovian. What is the limiting distribution of marbles?

9.7 Consider the prototypical example of a Markov chain kernel:

$$\begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix},$$

where $0 \leq p, q \leq 1$. What is the stationary distribution of this Markov chain? What happens when $p = q = 1$?

9.8 For the following transition matrix, perform the listed operations:

$$\mathbf{X} = \begin{bmatrix} 0.3 & 0.7 \\ 0.9 & 0.1 \end{bmatrix}.$$

(a) For the starting point $[0.6, 0.4]$ calculate the first 10 chain values.

(b) For the starting point $[0.1, 0.9]$ calculate the first 10 chain values.

(c) Does this transition matrix define an ergodic Markov chain?

(d) What is the limiting distribution, if it exists?

9.9 Suppose that for a Congressional race the probability that candidate B airs negative campaign advertisements in the next period, given that candidate A has in the current period, is 0.7; otherwise it is only 0.07. The same probabilities apply in the opposite direction. Answer the following questions.

(a) Provide the transition matrix.

(b) If candidate B airs negative ads in period 1, what is the probability that candidate A airs negative ads in period 3?

(c) What is the limiting distribution?

9.10 Duncan and Siverson (1975) used the example of social mobility where grandfather's occupational class affects the grandson's occupational class according to the following transition matrix:

$$\begin{array}{c} \\ \text{Upper} \\ \text{Middle} \\ \text{Lower} \end{array} \begin{array}{c} \begin{array}{ccc} \text{Upper} & \text{Middle} & \text{Lower} \end{array} \\ \begin{pmatrix} 0.228 & 0.589 & 0.183 \\ 0.065 & 0.639 & 0.296 \\ 0.037 & 0.601 & 0.361 \end{pmatrix} \end{array},$$

where the grandfather's occupational class defines the rows and the grandson's occupational class defines the columns. What is the long-run probability of no social mobility for the three classes?

Their actual application is to Sino-Indian relations from 1959 to 1964, where nine communication states are defined by categorizing each country's weekly communication with the other as high (3 or more), medium (1 or 2), or low (zero), at some point in time:

Weekly Communication Frequency

China	India	System State
Low	Low	1
Low	Medium	2
Low	High	3
Medium	Low	4
Medium	Medium	5
Medium	High	6
High	Low	7
High	Medium	8
High	High	9

This setup nicely fits the Markovian assumption since interstate communication is inherently conditional. Their (estimated) transition matrix, expressed as percentages instead of probabilities, is

$$P = \begin{array}{cc} & \begin{array}{ccccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \end{array} & \begin{pmatrix} 50 & 13 & 2 & 15 & 13 & 2 & 3 & 2 & 2 \\ 15 & 26 & 2 & 15 & 28 & 4 & 4 & 6 & 0 \\ 8 & 38 & 8 & 0 & 15 & 23 & 0 & 8 & 0 \\ 19 & 19 & 9 & 9 & 19 & 6 & 6 & 9 & 3 \\ 18 & 14 & 4 & 12 & 16 & 6 & 4 & 16 & 10 \\ 8 & 8 & 8 & 6 & 19 & 19 & 6 & 17 & 8 \\ 25 & 8 & 0 & 25 & 17 & 8 & 0 & 8 & 8 \\ 4 & 7 & 7 & 7 & 11 & 29 & 7 & 11 & 18 \\ 0 & 10 & 0 & 0 & 6 & 29 & 06 & 48 & \end{pmatrix} \end{array},$$

What evidence is there for the claim that there is "a tendency for China to back down from high levels of communication and a certain lack of responsiveness for India"? What other indications of "responsiveness" are found here and what can you conclude about the long-run behavior of the Markov chain?

9.11 Given an example transition matrix that produces a *non*irreducible Markov chain, and show that it has at least two distinct limiting distributions.

9.12 For the following matrices, find the limiting distribution:

$$\begin{bmatrix} 0.3 & 0.7 \\ 0.6 & 0.4 \end{bmatrix} \qquad \begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{bmatrix} \qquad \begin{bmatrix} 0.75 & 0.25 & 0.0 \\ 0.25 & 0.50 & 0.25 \\ 0.0 & 0.25 & 0.75 \end{bmatrix}.$$

9.13 For the following transition matrix, which classes are closed?

$$\begin{bmatrix} 0.50 & 0.50 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.50 & 0.00 & 0.50 & 0.00 \\ 0.00 & 0.00 & 0.50 & 0.50 & 0.00 \\ 0.00 & 0.00 & 0.75 & 0.25 & 0.00 \\ 0.50 & 0.00 & 0.00 & 0.00 & 0.50 \end{bmatrix}.$$

9.14 Chung (1969) used Markov chains to analyze hierarchies of human needs in the same way that Maslow famously considered them. The key point is that needs in a specific time period are conditional on the needs in the previous period and thus are dynamic moving up and down from basic to advanced human needs. Obviously the embedded assumption is that the pattern of needs is independent of previous periods conditional on the last period. In a hypothetical example Chung constructed a proportional composition of needs for some person according to

$$N = (N_{ph}, N_{sf}, N_{so}, N_{sr}, N_{sa}) = (0.15, 0.30, 0.20, 0.25, 0.10),$$

which are supposed to reflect transition probabilities from one state to another where the states are from Maslow's hierarchy: physiological, safety, socialization, self-respect, and self-actualization. Furthermore, suppose that as changes in socio-economic status occur, the composition of needs changes probabilistically according to

$$P = \begin{array}{c} \\ N_{ph} \\ N_{sf} \\ N_{so} \\ N_{sr} \\ N_{sa} \end{array} \begin{array}{c} \begin{array}{ccccc} N_{ph} & N_{sf} & N_{so} & N_{sr} & N_{sa} \end{array} \\ \begin{pmatrix} 0.20 & 0.30 & 0.35 & 0.15 & 0.00 \\ 0.10 & 0.40 & 0.20 & 0.20 & 0.10 \\ 0.10 & 0.15 & 0.35 & 0.25 & 0.15 \\ 0.05 & 0.15 & 0.20 & 0.40 & 0.20 \\ 0.00 & 0.05 & 0.15 & 0.30 & 0.50 \end{pmatrix} \end{array}.$$

Verify his claim that the system of needs reaches a stationary distribution after four periods using the starting point defined by N. How does this model change Maslow's assumption of strictly ascending needs?

9.15 Consider the following Markov chain from Atchadé and Rosenthal (2005). For the discrete state space $\Theta = \{1, 3, 4\}$, at the nth step produce the $n + 1^{st}$ value by:

- if the last move was a rejection, generate $\theta' \sim$ uniform$(\theta_n - 1 : \theta_n + 1)$;
- if the last move was an acceptance, generate $\theta' \sim$ uniform$(\theta_n - 2 : \theta_n + 2)$;
- if $\theta' \in \Theta$, accept θ' as θ_{n+1}, otherwise reject and set θ_n as θ_{n+1};

where these uniform distributions are on the inclusive positive integers and some arbitrary starting point θ_0 (with no previous acceptance) is assumed. What happens to this chain in the long run?

9.16 Markov chain analysis can be useful in game theory. Molander (1985) constructed the following matrix in his look at "tit-for-tat" strategies in international relations:

$$\begin{bmatrix} (1-p)^2 & p(1-p) & p(1-p) & p^2 \\ p(1-p) & p^2 & (1-p)^2 & p(1-p) \\ p(1-p) & (1-p)^2 & p^2 & p(1-p) \\ p^2 & p(1-p) & p(1-p) & (1-p)^2 \end{bmatrix}.$$

There are four outcomes where the rows indicate action by player C and the columns indicate a probabilistic response by player D for some stable probability p. Show that this is a valid transition matrix and find the vector of stationary probabilities. Molander modified this game to allow players the option of "generosity," which escapes the cycle of vendetta. This 4×4 transition matrix is given by

$$\begin{bmatrix} (1-p)(1-p-d+2pd & (p+c-2pc)(1-p-d+2pd)\ldots \\ \ldots p(1-p) & p(p+c-2pc) \\ (1-p)(p+d-2pd) & (p+c-2pc)(p+d-2pd)\ldots \\ \ldots(1-p)^2 & (1-p)(p+c-2pc) \\ p(1-p-d+2pd) & (1-p-c+2pc)(1-p-d-2pd)\ldots \\ \ldots p^2 & p(1-p-c+2pc) \\ p(1-p+d+2pd) & (1-p-c+2pc)(p+d-2pd)\ldots \\ \ldots p(1-p) & (1-p)(1-p-c+2pc) \end{bmatrix}$$

where c is the probability that player C deviates from tit-for-tat and d is the probability that player D deviates from tit-for-tat. Show that this matrix defines a recurrent Markov chain and derive the stationary distribution for $p = c = d = \frac{1}{2}$.

9.17 Dobson and Meeter (1974) modeled the movement of party identification in the United States between the two major parties as a Markovian. The following transition matrix gives the probabilities of *not* moving from one status to another conditional on moving (hence the zeros on the diagonal):

$$\begin{array}{c c} & \begin{array}{ccccc} SD & WD & I & WR & SR \end{array} \\ \begin{array}{c} SD \\ WD \\ I \\ WR \\ SR \end{array} & \begin{pmatrix} 0.000 & 0.873 & 0.054 & 0.000 & 0.073 \\ 0.750 & 0.000 & 0.190 & 0.060 & 0.000 \\ 0.154 & 0.434 & 0.000 & 0.287 & 0.125 \\ 0.039 & 0.205 & 0.346 & 0.000 & 0.410 \\ 0.076 & 0.083 & 0.177 & 0.709 & 0.000 \end{pmatrix} \end{array},$$

where the labels indicate SD = Strong Democrat, WD = Weak Democrat, I = Independent, WR = Weak Republican, and SR =

Strong Republican. Convert this to a transition matrix for moving from one state to another and show that it defines an ergodic Markov chain. What is the stationary distribution?