following comparative statics question. Suppose that the function g in (1) depends on a vector of parameters θ, so $y_t = g(x_t, z_t; \theta)$, where $\theta \in \Theta \subset \mathbf{R}^n$. Suppose, too, that for every $\theta \in \Theta$, there exists a unique invariant measure $\lambda^*(\cdot; \theta)$. We want to find conditions under which the expected value of a function under the invariant measure, $\int f(s)\lambda^*(ds; \theta)$, is continuous in θ, for every function f in a suitable class. This topic is addressed in Section 12.5.

In Chapter 13 we look at economic applications of the results developed in Chapters 11 and 12. Analyzing these models involves developing arguments that establish the desired properties of the transition function P from properties of the functions Q, g, and ϕ.

Finally, in Chapter 14 we look at the long-run properties of the time series generated by systems governed by Markov processes. Suppose that it can be established that for the system of interest, the sequence of long-run average probabilities $\{(1/T)\Sigma_{t=0}^{T-1}\lambda_t\}$ converges, in some sense, to an invariant measure λ^*. Then it seems reasonable to expect that for any real-valued function f in some suitable class, the sample average of the sequence of real numbers $\{f(s_t)\}$ converges, in the ordinary sense, to the expected value of f with respect to this invariant measure. That is, we expect that

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} f(s_t) = \int f(s)\lambda^*(ds),$$

holds for most of the sample paths—that is, for most of the sequences of realizations $\{s_t\}$—that one might obtain. Convergence statements of this type are called laws of large numbers. In Chapter 14 we develop such a law for Markov processes.

11.1 Markov Chains

In this section we discuss convergence for the case where the state space S consists of a finite number of elements, $S = \{s_1, \ldots, s_l\}$, and $\mathcal{S}$ consists of all subsets of S. A Markov process on a finite state space is called a ***Markov chain.*** Although the models discussed in Chapter 9 do not have finite state spaces, beginning with this case is useful, since many of the types of behavior possible for general Markov processes can be illustrated with a finite state space. Hence Markov chains are useful for illustrating various kinds of "bad" behavior and for examining the assumptions needed to rule them out. Moreover, many of the arguments used to establish existence, uniqueness, and convergence results for finite chains have close parallels in the general case. Since finite Markov chains are easy to analyze in a self-contained way, this is how we begin. Section 11.4 then parallels this one very closely, establishing analogous results for general state spaces.

The rest of this section is organized as follows. First, the notation is set and several terms are defined. Then we present five examples that illustrate various types of behavior. The analysis then proceeds to a more general level, and three results are established. Theorem 11.1 proves the existence of at least one invariant distribution and convergence of the sequence of long-run average probabilities, from any initial state, to one of these; Theorem 11.2 provides a necessary and sufficient condition for the uniqueness of the invariant distribution; and Theorem 11.4 provides a necessary and sufficient condition for the convergence of the sequence of probabilities, not just the averages, to this unique limit. Theorem 11.4 is, for our purposes, the most important of these results, and the reader may wish to skip to it immediately after studying the examples.

When S is the finite set $\{s_1, \ldots, s_l\}$, a probability measure on $(S, \mathcal{S})$ is represented by a vector p in the l-dimensional unit simplex: $\Delta^l = \{p \in \mathbf{R}^l: p \geq 0 \text{ and } \Sigma_{i=1}^l p_i = 1\}$. (All vectors here are row vectors.) Similarly, a transition function P is represented by an $l \times l$ matrix $\Pi = [\pi_{ij}]$, where $\pi_{ij} = P(s_i, \{s_j\})$. Note that the elements of Π are nonnegative and that each row sum is unity. Any such matrix is called a ***Markov matrix*** or a ***stochastic matrix.*** For $i = 1, \ldots, l$, let $e_i \in \Delta^l$ denote the vector with a one in the ith position and zeros elsewhere.

First consider one-step transitions. If the current state is s_i, the probability distribution over next period's state is given by the ith row of Π, $\pi_{i\cdot} = (\pi_{i1}, \ldots, \pi_{il})$. That is, if the probability distribution over the current state is e_i, then the distribution over next period's state is $e_i\Pi = \pi_{i\cdot}$. More generally, if $p \in \Delta^l$ is the probability distribution over the state in period t, then the distribution over the state in period $t + 1$ is given by $\hat{p} = (\hat{p}_1, \ldots, \hat{p}_l)$, where

$$\hat{p}_j = \sum_{i=1}^{l} p_i \pi_{ij}, \quad j = 1, \ldots, l.$$

Since $\Sigma_j \hat{p}_j = \Sigma_i p_i \Sigma_j \pi_{ij} = 1$, it follows that $\hat{p} \in \Delta^l$. In matrix notation then, $\hat{p} = p\Pi$, and we have shown that if Π is an $l \times l$ Markov matrix and $p \in \Delta^l$, then $p\Pi \in \Delta^l$.

Applying the same argument again, we find that if the probability distribution over the state in period t is $p \in \Delta^l$, then the probability distribution over the state two periods ahead is $(p\Pi)\Pi = p(\Pi \cdot \Pi) = p\Pi^2$. Continuing by induction, we find that the n-step transition probabilities are given by the matrix Π^n, for $n = 0, 1, 2, \ldots$, where we take Π^0 to be the identity matrix. It is straightforward to verify that if Π is a Markov matrix, then so is Π^n, for $n = 2, 3, \ldots$. Thus, if the initial state is s_i, then the probability distribution over states n periods ahead is given by $e_i\Pi^n$, the ith row of Π^n. To understand the long-run behavior of the system, then, we must study the behavior of the sequence $\{\Pi^n\}_{n=0}^{\infty}$. Before beginning the formal analysis, it is useful to look at examples that illustrate various possibilities.

Example 1 Let $l = 2$, and suppose that

$$\Pi = \begin{bmatrix} 3/4 \\ 1/4 \end{bmatrix}$$

A set $E \subseteq S$ is called an ***ergodic set*** if $p(s_i, E) = 1$, for $s_i \in E$, and if no proper subset of E has this property. In this example, the only ergodic set is S itself. Moreover, it is easy to verify by direct calculation that

$$\Pi^2 = \begin{bmatrix} 5/8 \\ 3/8 \end{bmatrix}$$

$$\Pi^3 = \begin{bmatrix} 9/16 & 7/16 \\ 7/16 & 9/16 \end{bmatrix},$$

$$\Pi^4 = \begin{bmatrix} 17/32 & 15/32 \\ 15/32 & 17/32 \end{bmatrix}.$$

Clearly the sequence $\{\Pi^n\}$ is converging, and pretty rapidly, and

$$\lim_{n \to \infty} \Pi^n = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}.$$

Notice how we use the word *converge* here: we say that the sequence of $l \times l$ matrices $\{\Pi^n\}$ converges to Q if each of the l^2 sequences of elements converges. In this example the probability distribution over the state converges to $p^* = (1/2, 1/2)$ for all initial probability distributions p_0. Note, too, that if the probability distribution over the initial state is p^*, then it is also p^* in every successive period. A vector with this property is called an ***invariant distribution***. Note that each row of the limit matrix in this example is the invariant distribution, p^*.

Example 2 Let $l = 3$, and suppose that Π has the form

$$\Pi = \begin{bmatrix} 1-\gamma & \gamma/2 & \gamma/2 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix},$$

where $\gamma \in (0, 1)$. If the system starts out in state s_1, then in the next period with probability $(1 - \gamma)$ it stays in that state, and with probability γ it leaves. Given that it leaves, it is equally likely to go to state s_2 or s_3. Note that if it leaves the state s_1, it cannot return. A state is called ***transient*** if there is a positive probability of leaving and never returning. If the initial state is s_2 or s_3, the situation is similar to that in Example 1; here $E = \{s_2, s_3\}$ is the only ergodic set.

By direct calculation we find that in this example

$$\Pi^n = \begin{bmatrix} (1-\gamma)^n & \delta_n/2 & \delta_n/2 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix},$$

where $\delta_n = 1 - (1 - \gamma)^n$. Since $\gamma \in (0, 1)$, it follows that

$$\lim_{n\to\infty} \Pi^n = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{bmatrix}.$$

Thus, with probability one the system eventually leaves the state s_1 and enters the ergodic set. Note that in this case, as in Example 1, $\{\Pi^n\}$ converges, and each row of the limit matrix is an invariant distribution.

Example 3 Next suppose that Π has the form

$$\Pi = \begin{bmatrix} 0 & \Pi_1 \\ \Pi_2 & 0 \end{bmatrix}$$

where Π_1 and Π_2 are $k \times (l - k)$ and $(l - k) \times k$ Markov matrices respectively, each with strictly positive elements. Then

$$\Pi^{2n} = \begin{bmatrix} (\Pi_1\Pi_2)^n & 0 \\ 0 & (\Pi_2\Pi_1)^n \end{bmatrix} \quad \text{and}$$

$$\Pi^{2n+1} = \begin{bmatrix} 0 & (\Pi_1\Pi_2)^n\Pi_1 \\ (\Pi_2\Pi_1)^n\Pi_2 & 0 \end{bmatrix}$$

for even- and odd-numbered transitions respectively. If the system begins in a state in the set $C_1 = \{s_1, \ldots, s_k\}$, then after any even number of steps it is back in the set C_1, and after any odd number of steps it is in the set $C_2 = \{s_{k+1}, \ldots, s_l\}$. If the system starts out in C_2, the reverse is true. In this example, as in the first, there is only one ergodic set, all of S, but that ergodic set has ***cyclically moving subsets.*** Obviously $\{\Pi^n\}$ does not converge in this case, but the two subsequences for odd and even steps do.

For example, suppose that $k = l - k = 2$ and that Π_1 and Π_2 are both equal to Π of Example 1. Then as n increases,

$$\Pi^{2n} \rightarrow \begin{matrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{matrix} \quad \text{and}$$

$$\Pi^{2n+1} \rightarrow \begin{matrix} 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \\ 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \end{matrix}$$

for even- and odd-numbered transitions. Thus $\{\Pi^n\}$ does not converge.

On the other hand, the following average does:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \Pi^n = \begin{matrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \end{matrix}$$

Note, too, that each row of this limit matrix, $p^* = (1/4, 1/4, 1/4, 1/4)$, is an invariant distribution.

Example 4 Next suppose that Π has the form

$$\Pi = \begin{bmatrix} \Pi_1 & 0 \\ 0 & \Pi_2 \end{bmatrix},$$

where Π_1 and Π_2 are Markov matrices of dimension $k \times k$ and $(l - k) \times (l - k)$ respectively, each with strictly positive elements. The n-step transitions are then given by

$$\Pi^n = \begin{bmatrix} \Pi_1^n & 0 \\ 0 & \Pi_2^n \end{bmatrix}$$

Thus, if the system starts out in the set $E_1 = \{s_1, \ldots, s_k\}$, then it stays in that set forever. The same is true if the system starts out in the set $E_2 = \{s_{k+1}, \ldots, s_l\}$. In this case, then, there are two ergodic sets. Clearly the sequence $\{\Pi^n\}$ converges if and only if $\{\Pi_1^n\}$ and $\{\Pi_2^n\}$ both converge. Suppose that Π_1 and Π_2 are both as specified in Example 1. Then

$$\lim_{n\to\infty} \Pi^n = \begin{matrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{matrix}$$

Thus there are two invariant distributions, $p_1^* = (1/2, 1/2, 0, 0)$ and $p_2^* = (0, 0, 1/2, 1/2)$, and the system converges to one or the other, depending on the initial state. Note, too, that all convex combinations of p_1^* and p_2^* are invariant distributions as well.

Example 5 Finally, consider a case where there are three states, and

$$\Pi = \begin{bmatrix} 1-\gamma & \alpha\gamma & \beta\gamma \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

where $\alpha, \beta, \gamma \in (0, 1)$ and $\alpha + \beta = 1$. Here, as in Example 2, s_1 is a transient state; however, there are two ergodic sets: $E_1 = \{s_2\}$ and $E_2 = \{s_3\}$. If the system starts out in state s_1 and leaves, the conditional probability of going to E_1 is α and to E_2 is β. If the state is s_2 or s_3, it remains constant forever after.

The n-step transition matrix in this case is

$$\Pi^n = \begin{bmatrix} (1-\gamma)^n & \alpha\delta_n & \beta\delta_n \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

where $\delta_n = 1 - (1 - \gamma)^n$. Since $\gamma \in (0, 1)$, it then follows that

$$\lim_{n\to\infty} \Pi^n = \begin{bmatrix} 0 & \alpha & \beta \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

In this case $\{\Pi^n\}$ converges; each row of the limit matrix is an invariant distribution; and the row corresponding to the transient state is a convex combination of the rows corresponding to the ergodic sets.

These five examples illustrate all possible types of limiting behavior for finite Markov chains. We will establish this fact in the remainder of the section, studying the existence and uniqueness of an ergodic set, the existence and uniqueness of an invariant distribution, and the convergence of the sequences $\{(1/n)\Sigma_{k=0}^{n-1}\Pi^k\}_{n=1}^{\infty}$ or $\{\Pi^k\}_{k=0}^{\infty}$ or both. Theorems 11.1, 11.2, and 11.4 deal with these questions under successively stronger assumptions about Π.

Let $S = \{s_1, \ldots, s_l\}$, let the stochastic matrix $\Pi = [\pi_{ij}]$ define the transition probabilities, and let $\Pi^n = [\pi_{ij}^{(n)}]$ denote the powers of Π. [Note that $\pi_{ij}^{(n)}$ is *not* in general equal to $(\pi_{ij})^n$.] Our first result requires no further restrictions on Π.

THEOREM 11.1 *Let $S = \{s_1, \ldots, s_l\}$ be a finite set, and let t stochastic matrix Π define transition probabilities on S. Then*

a. *S can be partitioned into $M \geq 1$ ergodic sets and a transient set.*
b. *The sequence $\{(1/n)\sum_{k=0}^{n-1}\Pi^k\}_{n=1}^{\infty}$ converges to a stochastic matrix Q. That is, $\lim_{n\to\infty}(1/n)\sum_{k=0}^{n-1}p_k = p_0 Q$, for any sequence $\{p_k\} = \{p_0\Pi^k\}$ where $p_0 \in \Delta^l$.*
c. *Each row of Q is an invariant distribution, so p_0Q is an invariant distribution for each $p_0 \in \Delta^l$; and every invariant distribution for Π is a convex combination of the rows of Q.*

Proof of (a). Call j a *consequent* of i if $\pi_{ij}^{(n)} > 0$ for some $n \geq 1$. Call s_i a *transient* state if it has at least one consequent j for which $\pi_{ji}^{(n)} = 0$, all $n \geq 1$. Thus, a state is transient if and only if there is a positive probability of not returning to it. Call a state i *recurrent* if for every j that is a consequent of i, i is also a consequent of j.

To show that S can be partitioned as claimed, we will begin by showing that S has at least one recurrent state. Suppose the contrary. Then $\pi_{ii}^{(n)} \neq 1$, $i = 1, \ldots, l$ (otherwise s_i would be recurrent). Since s_1 is transient, there exists a state—call it s_2—and $N \geq 1$ such that $\pi_{12}^{(N)} > 0$ and $\pi_{21}^{(n)} = 0$, $n = 1, 2, \ldots$. Then since $\pi_{22}^{(n)} \neq 1$, $\pi_{21}^{(n)} = 0$, $n = 1, 2, \ldots$, and s_2 is transient, there exists a state—call it s_3—and $N' \geq 1$ such that $\pi_{23}^{(N')} > 0$ and $\pi_{32}^{(n)} = 0$, $n = 1, 2, \ldots$. Moreover, since $0 = \pi_{32}^{(n+N)} \geq \pi_{31}^{(n)}\pi_{12}^{(N)}$, $n = 1, 2, \ldots$, it follows that $\pi_{31}^{(n)} = 0$, $n = 1, 2, \ldots$. Continuing by induction, we conclude that $\pi_{ll} \neq 1$ and $\pi_{lj} = 0, j = 1, \ldots, l-1$, which contradicts the fact that Π is a stochastic matrix.

Next we will show that if the state s_i is recurrent and j is a consequent of i, then i is a consequent of j and s_j is also recurrent. Suppose that s_i is recurrent and that j is a consequent of i. Since s_i is recurrent, not transient, it follows that $\pi_{ji}^{(N)} > 0$ for some $N \geq 1$, so i is a consequent of j. Next, suppose that k is a consequent of j. Then $\pi_{jk}^{(L)} > 0$, for some $L \geq 1$. But this implies that $\pi_{ik}^{(N+L)} \geq \pi_{ij}^{(N)}\pi_{jk}^{(L)} > 0$, so k is a consequent of i. Since s_i is recurrent, it then follows that $\pi_{ki}^{(K)} > 0$, for some $K \geq 0$. Hence $\pi_{kj}^{(K+N)} \geq \pi_{ki}^{(K)}\pi_{ij}^{(N)} > 0$, so j is a consequent of k. Since k was an arbitrary consequent of j, it follows that j is recurrent.

Hence the set S can be partitioned as follows. First, let F be the set of all transient states. Then partition the recurrent states into disjoint sets $E_1, E_2, \ldots, E_M$, by assigning two states to the same set if and only if they are consequents of each other. Since there is at least one recurrent state, there is at least one such set. Moreover, by construction, $s_i \in E_m$ implies that $\sum_{j\in E_m}\pi_{ij} = 1$ and that no subset of E_m has this property. Hence each set E_m is ergodic. Note that once the state enters one of the ergodic sets, it remains in that set forever.

Proof of (b). Next consider the average probabilities over n-step horizons. Let $p_0 = (p_{01}, \ldots, p_{0l})$ be a probability distribution over states in period 0. Then $p_k = p_0\Pi^k$ is the probability distribution over states in period k, for $k = 1, 2, \ldots$; and the average over these distributions for the first n periods is given by

$$\frac{1}{n}\sum_{k=0}^{n-1} p_k = \frac{1}{n}\sum_{k=0}^{n-1} p_0\Pi^k = p_0\left[\frac{1}{n}\sum_{k=0}^{n-1}\Pi^k\right]$$

Define $A^{(n)} = (1/n)\Sigma_{k=0}^{n-1}\Pi^k$, and note that since it is an average of stochastic matrices, each $A^{(n)}$ is itself a stochastic matrix. We will first characterize the behavior of the long-run average probabilities by examining the behavior of $A^{(n)}$ as $n \to \infty$ and then will show that if S is a finite set, the sequence $\{A^{(n)}\}$ converges.

First we will show that there exists a subsequence—call it n_k—such that $\{A^{(n_k)}\}$ converges. To see this, note that each of the sequences $\{a_{ij}^{(n)}\}_{n=1}^{\infty}$, $i, j = 1, \ldots, l$, lies on the compact interval [0, 1]. Hence there exists a subsequence of the n's—call it n'—for which $\{a_{11}^{(n')}\}$ converges. Then, by the same reasoning, there exists a subsequence of the n''s, call it n'', for which $\{a_{12}^{(n'')}\}$ also converges, etc. Since there are only a finite number of elements to consider, continuing by induction establishes the desired conclusion. Note, too, that this argument establishes that every subsequence of $\{A^{(n)}\}$ contains a convergent subsequence.

Let Q be the limit of some convergent subsequence $\{A^{(n_k)}\}$:

$$\lim_{k\to\infty}\frac{1}{n_k}\sum_{m=0}^{n_k-1}\Pi^m = Q$$

Then pre- and postmultiplying by Π, we find that

$$\lim_{k\to\infty}\frac{1}{n_k}\sum_{m=1}^{n_k}\Pi^m = Q\Pi = \Pi Q$$

Since the two averages in these equations differ only by the terms Π^0/n_k and Π^{n_k}/n_k, both of which go to zero as $k \to \infty$, the two limits are equal. Hence $Q = Q\Pi = \Pi Q$. This fact in turn implies that $Q = Q\Pi^n = \Pi^n Q$, all n.

We have defined Q to be the limit of the subsequence $\{A^{(n_k)}\}$. From the remaining terms in $\{A^{(n)}\}$ we can extract another convergent subsequence; call its limit A. Then since $Q = Q\Pi^n = \Pi^n Q$, $n = 1, 2, \ldots$, it follows that $Q = QA = AQ$, and, with the roles of Q and A reversed, that $A = AQ = QA$. Hence $A = Q$. Since the choice of subsequences converging to A and Q was arbitrary, it follows that

$$Q = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \Pi^k,$$

and hence that

$$p_0 Q = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} p_0 \Pi^k, \quad \text{all } p_0 \in \Delta^l.$$

Proof of (c). Finally we will show that each of the rows of Q is an invariant distribution and that every invariant distribution is a convex combination of these rows.

As shown above $Q\Pi = Q$; that is $\Sigma_{k=1}^{l} q_{ik}\pi_{kj} = q_{ij}$, $i, j = 1, \ldots, l$. Hence each row of Q is an invariant distribution. Conversely, suppose that $r = (r_1, \ldots, r_l)$ is an invariant distribution. Then

$$\sum_{i=1}^{l} r_i \pi_{ij}^{(n)} = r_j, \quad j = 1, \quad, l;\ n = \ , 2, .$$

so

$$\frac{1}{N} \sum_{n=0}^{N-1} \left[\sum_{i=1}^{l} r_i \pi_{ij}^{(n)} \right] = \sum_{i=1}^{l} r_i \left[\frac{1}{N} \sum_{n=0}^{N-1} \pi_{ij}^{(n)} \right] = r_j,$$

$$j = 1, \ldots, l;\ N = \ , 2, .$$

Taking the limit as $N \to \infty$, we obtain $\Sigma_{i=1}^{l} r_i q_{ij} = r_j$, $j =$ $\ .\ , l$, so r is a convex combination of the rows of Q. ■

This theorem applies to all of the five examples above, and they illustrate the variety of behavior that is consistent with its conclusions. There may be one ergodic set or more than one, and in addition there may be a transient set; the sequence $\{\Pi^k\}$ may or may not converge, but the sequence of averages $\{A^{(n)}\} = \{(1/n)\Sigma_{k=0}^{n-1} \Pi^k\}$ necessarily converges; and the rows of the limiting matrix Q are invariant distributions. For the case of finite chains, then, given any initial probability distribution, the long-run average probabilities over states converge. In particular, if the initial state is s_i, then the long-run average probabilities are given by the ith row of the matrix Q. If p_0 is an initial probability distribution over states, then the long-run average probabilities are given by p_0Q. Note that we have not ruled out the possibility of multiple ergodic sets or cyclically moving subsets within any ergodic set.

There is a close connection between the M ergodic sets and the M invariant distributions described in Theorem 11.1. To see it note that we can, without loss of generality, order the states so that all of the transient states come first, and the states in each ergodic class appear in a block. With this ordering of the states, the transition matrix Π takes the (almost block diagonal) form

$$\begin{array}{c|ccccc}
 & F & E_1 & E_2 & \cdots & E_M \\
\hline
F & R_{00} & R_{01} & R_{02} & \cdots & R_{0M} \\
E_1 & 0 & R_{11} & 0 & \cdots & 0 \\
E_2 & 0 & 0 & R_{22} & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
E_M & 0 & 0 & 0 & \cdots & R_{MM}
\end{array}$$

Note that each matrix $R_{11}, \ldots, R_{MM}$ is a stochastic matrix; hence Theorem 11.1 applies to each one, and we can define

$$Q_j = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} R_{jj}^n, \quad j = \qquad , M.$$

On the other hand, R_{00} is not a stochastic matrix; otherwise F would be an ergodic set. That is, R_{00} has at least one row sum that is strictly less than one.

Exercise 11.1 Show that when Π has the form above, Q has the form

$$
\begin{array}{c|ccccc}
 & F & E_1 & E_2 & \cdots & E_M \\
\hline
F & 0 & w_1Q_1 & w_2Q_2 & \cdots & w_MQ_M \\
E_1 & 0 & Q_1 & 0 & \cdots & 0 \\
E_2 & 0 & 0 & Q_2 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
E_M & 0 & 0 & 0 & \cdots & Q_M
\end{array}
$$

where the rows within each matrix Q_j are all identical and where the column vectors w_j sum to the unit vector. Show that the ith element in w_j is the probability of a transition, eventually, from the ith transient state to the set E_j.

Since all the rows within each matrix Q_i are identical, it follows that if the system begins in any state in the ergodic class E_i, the long-run average probabilities are the same. The same is true if the initial position is described by any probability distribution that assigns zero probability to all states outside of E_i. The M distinct rows of the matrices $Q_1, \ldots, Q_M$ correspond to the M ergodic classes. The first block of rows, those corresponding to initial states in F, are convex combinations of the others. If the system begins in a transient state, then the long-run average probabilities are determined by the probability of eventually getting into each of the various ergodic classes. Thus in the case of a finite state space we can, for an arbitrary transition matrix Π, establish the existence of at least one ergodic class and, accordingly, one invariant distribution.

To obtain sharper results, we must impose additional structure on the transition matrix Π. The next theorem provides a necessary and sufficient condition to ensure that there is a unique ergodic class. That is, it rules out cases like Examples 4 and 5.

THEOREM **11.2** *Let $S = \{s_1, \ldots, s_l\}$ be a finite set, and let the stochastic matrix Π define transition probabilities on S. Then Π has a unique ergodic set if and only if there exists a state s_j such that the following holds: for every*

$i \in \{1, \ldots, l\}$, there exists $n \geq 1$ such that $\pi_{ij}^{(n)} > 0$. In this case Π has a unique invariant distribution, call it p^*; each row of Q is equal to p^*; and for any $p_0 \in \Delta^l$, $p_0 Q = p^*$.

Proof. Suppose that the stated condition holds for some state s_j. Then s_j cannot be transient and is a consequent of every state s_i, $i = 1, \ldots, l$. Hence there is at most one ergodic set. The other claims then follow immediately from Theorem 11.1.

Conversely, suppose that there is only one ergodic class, E, and choose $s_j \in E$. Since every element of E is a consequent of j, it follows that the stated condition holds for every $s_i \in E$. Consider next any $s_i \in F$. For some $s_k \in E$ and $n \geq 1$, $\pi_{ik}^{(n)} > 0$; otherwise s_i is not transient. Since j is a consequent of k, $\pi_{kj}^{(m)} > 0$ for some $m \geq 1$. Hence $\pi_{ij}^{(n+m)} \geq \pi_{ik}^{(n)}\pi_{kj}^{(m)} > 0$. ■

Our final result provides a condition that, in addition to ensuring the uniqueness of the ergodic set, rules out cyclically moving subsets within the ergodic set. That is, it rules out cases like Example 3. Under this condition the sequence $\{\Pi^k\}$, not just the sequence of long-run averages, converges to Q. Since by Theorem 11.2 each row of Q is equal to the unique invariant distribution p^*, it then follows that the sequence of distributions $\{p_0\Pi^k\}$ converges to p^* for all $p_0 \in \Delta^l$. Moreover the convergence is at a geometric rate that is uniform in p_0.

There are many ways to establish this result; the proof we will present is based on the Contraction Mapping Theorem. The idea is to show that if for some $N \geq 1$, Π^N defines a contraction mapping on Δ^l, then the unique fixed point of this mapping is the vector p^* and the convergence of $\{p_0\Pi^k\}$ is as claimed. We will also show that the converse is true: if $\{p_0\Pi^k\} \to p^*$, all $p_0 \in \Delta^l$, then for some $N \geq 1$, Π^N defines a contraction mapping on Δ^l. Lemma 11.3 provides a sufficient condition for Π to define a contraction mapping on Δ^l; Theorem 11.4 uses this condition to obtain the conclusions described above.

To apply the Contraction Mapping Theorem we must show that Δ^l is a complete metric space, for an appropriately chosen metric, and that the transition matrix Π defines an operator taking Δ^l into itself. For the rest of this section let $\|\cdot\|_\Delta$ denote the norm on $\mathbf{R}^l$ defined by

$$\|x\|_\Delta = \sum_{j=1}^{l} |x_j|$$

Recall from Exercises 3.4c and 3.6a that with this metric $(\mathbf{R}^l, \|\cdot\|_\Delta)$ is a complete metric space and Δ^l is a closed subset. Hence $(\Delta^l, \|\cdot\|_\Delta)$ is also a complete metric space. Then as was shown above, for any $l \times l$ Markov matrix Π, the operator T^* defined by $T^*p = p\Pi$ takes Δ^l into itself.

LEMMA 11.3 *Let Π be an $l \times l$ Markov matrix, and for $j = 1, \ldots, l$, let $\varepsilon_j = \min_i \pi_{ij}$. If $\sum_{j=1}^l \varepsilon_j = \varepsilon > 0$, then the mapping $T^*: \Delta^l \to \Delta^l$ defined by $T^*p = p\Pi$ is a contraction of modulus $1 - \varepsilon$.*

Proof. Let $p, q \in \Delta^l$. Then

$$\begin{aligned}
\|T^*p - T^*q\|_\Delta &= \|p\Pi - q\Pi\|_\Delta \\
&= \sum_{j=1}^l \left| \sum_{i=1}^l (p_i - q_i)\pi_{ij} \right| \\
&= \sum_{j=1}^l \left| \sum_{i=1}^l (p_i - q_i)(\pi_{ij} - \varepsilon_j) + \sum_{i=1}^l (p_i - q_i)\varepsilon_j \right| \\
&\leq \sum_{j=1}^l \sum_{i=1}^l |p_i - q_i|(\pi_{ij} - \varepsilon_j) + \sum_{j=1}^l \varepsilon_j \left| \sum_{i=1}^l (p_i - q_i) \right| \\
&= \sum_{i=1}^l |p_i - q_i| \sum_{j=1}^l (\pi_{ij} - \varepsilon_j) + 0 \\
& (1 - \varepsilon)\|p - q\|_\Delta \quad \blacksquare
\end{aligned}$$

THEOREM 11.4 *Let $S = \{s_1, \ldots, s_l\}$ be a finite set, and let the stochastic matrix Π define transition probabilities on S. For $n = 1, 2, \ldots$, let $\varepsilon_j^{(n)} = \min_i \pi_{ij}^{(n)}$, $j = 1, \ldots, l$, and let $\varepsilon^{(n)} = \sum_{j=1}^l \varepsilon_j^{(n)}$. Then S has a unique ergodic set with no cyclically moving subsets if and only if for some $N \geq 1$, $\varepsilon^{(N)} > 0$. In this case $\{p_0\Pi^k\}$ converges to a unique limit $p^* \in \Delta^l$, for all $p_0 \in \Delta^l$, and convergence is at a geometric rate that is independent of p_0.*

Proof. If $\varepsilon^{(N)} > 0$ then by Lemma 11.3, $T^{*N}: \Delta^l \to \Delta^l$ defined by $T^*p = p\Pi^N$ is a contraction of modulus $1 - \varepsilon^{(N)}$. Since Δ^l is a closed subset of a complete metric space, it follows from the Contraction Mapping Theorem (Theorem 3.2) that T^{*N} has a unique fixed point—call it p^*—and that

$$\|p_0\Pi^{kN} - p^*\|_\Delta \leq (1 - \varepsilon^{(N)})^k \|p_0 - p^*\|_\Delta, \quad k = 1, 2, \quad , \text{all } p_0 \in \Delta^l.$$

Conversely, suppose that $\{p_0\Pi^n\} \to p^*$, all $p_0 \in \Delta^l$. Then $\{\Pi^n\} \to (p^*, \ldots, p^*)'$, so for some $N \geq 1$ sufficiently large there is at least one column j for which $\pi_{ij}^{(N)} > 0$, all i. Then $\varepsilon^{(N)} \geq \varepsilon_j^{(N)} > 0$. ∎

Hence for the case where the state space S is finite, we have a very simple condition that is both necessary and sufficient for convergence to a unique stationary distribution. Moreover, in this case convergence is at a geometric rate that is uniform in the initial distribution.

It is worthwhile to summarize briefly the steps in the analysis above. First we defined Δ^l, the set of all probability distributions on S, and showed that a transition matrix Π defines a mapping T^* of Δ^l into itself. Specifically, we saw that if $p_0 \in \Delta^l$ is the distribution over the initial state, then $p_n = p_0\Pi^n$ is the distribution over the state n periods later, and $(1/n)\Sigma_{k=0}^{n-1} p_k = p_0A^{(n)}$ is the average of the distributions over the first n periods. The main results were

1. S can always be partitioned into $M \geq 1$ ergodic sets and a transient set, and the sequence $\{A^{(n)}\}$ always converges to a stochastic matrix Q. Moreover, the rows of Q are invariant distributions; hence for any $p_0 \in \Delta^l$, the sequence $\{(1/n)\Sigma_{k=0}^{n-1} p_0\Pi^k\} = \{(1/n)\Sigma_{k=0}^{n-1} p_k\}$ converges to p_0Q, an invariant distribution. Finally, all invariant distributions can be formed as convex combinations of the rows of Q.
2. If the additional hypothesis of Theorem 11.2 holds, then there is a unique ergodic set. In this case all of the rows of Q are identical, and the vector p^* in each row is the unique invariant distribution.
3. If the additional hypothesis of Theorem 11.4 holds, then the unique ergodic class has no cyclically moving subsets, and the sequence $\{\Pi^n\}$ converges to Q. Hence for any $p_0 \in \Delta^l$, the sequence $\{p_n\} = \{p_0\Pi^n\}$ converges to p^*. Moreover, the rate of convergence is geometric and is uniform in p_0.

In Section 11.4 we show that all of these results have analogues in general state spaces. However, there are possibilities that arise in infinite state spaces that do not have counterparts in the finite-state Markov chains we have treated so far. We conclude this section with two examples of infinite-state Markov processes that illustrate these possibilities.

Example 6 Let $S = \{s_1, s_2, \ldots\}$, and suppose that the transition matrix Π is the infinite-dimensional identity matrix. Then there are an infinite number of ergodic sets, each of the sets $E_i = \{s_i\}$, $i = 1, 2, \ldots$; and corresponding to each is an invariant distribution, the probability vectors e_i, $i = 1, 2, \ldots$. Clearly this possibility arises whenever the state space is infinite.

Example 7 Let $S = \{s_1, s_2,$ and let

$$\Pi = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

In this case all of the states are transient. There is no ergodic set and hence no invariant distribution. Note that in this case the sequence $\{\Pi^k\}$ converges to the zero matrix: all of the probability "wanders off to infinity."

To extend the results in the theorems above to more general state spaces, then, we must find assumptions that rule out cases like these. Before turning to these issues, however, we need to discuss the notion of convergence for measures in those spaces.

11.2 Convergence Concepts for Measures

In the last section we examined conditions under which a sequence of probability distributions $\{p_n\}$ on a finite space, generated from an initial distribution p_0 by the recursive formula $p_{n+1} = p_n\Pi$, would converge to a limiting distribution p^* as $n \to \infty$. In that discussion $\{p_n\}$ and p^* were elements of $\mathbf{R}^l$, so the term "convergence" could be made precise using any of the many equivalent norms for $\mathbf{R}^l$. Our use of the particular norm $\|p\| = \Sigma_i|p_i|$ was important for establishing facts about the *rate* of convergence in Lemma 11.3 and Theorem 11.4, but not for the other results. In the rest of this chapter and the next, our concern is with sequences $\{\lambda_n\}$ of probability measures on a fixed measurable space $(S, \mathcal{S})$, and with questions about the convergence of such a sequence to a limiting measure λ. When $(S, \mathcal{S})$ is an arbitrary measurable space, there is considerable latitude in defining convergence of a sequence of probability measures. Therefore, it is useful to begin by reviewing some of the issues involved in defining convergence for measures.

Let $(S, \mathcal{S})$ be a measurable space; let $\Lambda(S, \mathcal{S})$ be the set of probability measures on $(S, \mathcal{S})$; and let $\{\lambda_n\}_{n=1}^{\infty}$ and λ be measures in $\Lambda(S, \mathcal{S})$. Two basic approaches can be used in defining convergence concepts for sequences of probability measures. The first is based on measures of sets, the second on the expected values of functions. As we show below, however, all of the standard notions of convergence can be characterized using either approach.

First consider the possibilities for defining convergence in terms of measures of sets. One criterion is simply

$$\lim_{n\to\infty} \lambda_n(A) = \lambda(A), \quad \text{all } A \in \mathcal{S}. \tag{1}$$

If (1) holds, we say that λ_n *converges setwise* to λ. This criterion can be weakened by choosing a family of sets $\mathcal{A} \subset \mathcal{S}$ and requiring only that (1) hold for sets in $\mathcal{A}$. An appropriate choice of $\mathcal{A}$ leads to a different, strictly weaker, notion of convergence. Alternatively we can strengthen the definition in (1) by requiring, in addition, some sort of uniformity in the rate of convergence.

A second way to think about convergence of measures is in terms of the behavior of the expected values of functions. Let $(S, \mathcal{S})$ and $\Lambda(S, \mathcal{S})$ be as specified above, and let $B(S, \mathcal{S})$ be the set of bounded measurable functions $f: S \to \mathbf{R}$. Then given $\{\lambda_n\}$ and λ in $\Lambda(S, \mathcal{S})$, we might want to use as a criterion the requirement that

$$\lim_{n\to\infty} \int f\, d\lambda_n = \int f\, d\lambda, \quad \text{all } f \in B(S, \mathcal{S}). \tag{2}$$

Obviously (2) implies (1), since the indicator function of each set in $\mathcal{S}$ is in $B(S, \mathcal{S})$. The converse is also true.

> ***Exercise 11.2*** Show that (1) implies (2).

Thus (2) is an alternative and completely equivalent way to characterize setwise convergence.

The criterion (2) can be weakened by choosing a smaller family of functions and requiring only that (2) hold for functions in this smaller family. As before, if the smaller family is appropriately chosen, a different, strictly weaker, definition of convergence is obtained. Alternatively, (2) can be strengthened by requiring some sort of uniformity in the rate of convergence.

As the discussion thus far should suggest, whether convergence of probability measures is defined in terms of measures of sets or in terms