# Chapter 2
# Time Series

## 2.1. Two workhorses

This chapter describes two tractable models of time series: Markov chains and first-order stochastic linear difference equations. These models are organizing devices that put particular restrictions on a sequence of random vectors. They are useful because they describe a time series with parsimony. In later chapters, we shall make two uses each of Markov chains and stochastic linear difference equations: (1) to represent the exogenous information flows impinging on an agent or an economy, and (2) to represent an optimum or equilibrium outcome of agents' decision making. The Markov chain and the first-order stochastic linear difference both use a sharp notion of a state vector. A state vector summarizes the information about the current position of a system that is relevant for determining its future. The Markov chain and the stochastic linear difference equation will be useful tools for studying dynamic optimization problems.

## 2.2. Markov chains

A stochastic process is a sequence of random vectors. For us, the sequence will be ordered by a time index, taken to be the integers in this book. So we study discrete time models. We study a discrete-state stochastic process with the following property:

MARKOV PROPERTY: A stochastic process $\{x_t\}$ is said to have the *Markov property* if for all $k \geq 1$ and all $t$,

$$\text{Prob}\,(x_{t+1}|x_t, x_{t-1}, \ldots, x_{t-k}) = \text{Prob}\,(x_{t+1}|x_t)$$

We assume the Markov property and characterize the process by a *Markov chain.* A time-invariant Markov chain is defined by a triple of objects, namely, an $n$-dimensional state space consisting of vectors $e_i, i = 1, \ldots, n$, where $e_i$ is an $n \times 1$ unit vector whose $i$th entry is 1 and all other entries are zero; an $n \times n$ *transition matrix* $P$, which records the probabilities of moving from one value of the state to another in one period; and an $(n \times 1)$ vector $\pi_0$ whose $i$th element is the probability of being in state $i$ at time 0: $\pi_{0i} = \text{Prob}(x_0 = e_i)$. The elements of matrix $P$ are

$$P_{ij} = \text{Prob}\,(x_{t+1} = e_j | x_t = e_i)$$

For these interpretations to be valid, the matrix $P$ and the vector $\pi$ must satisfy the following assumption:

ASSUMPTION M:

a. For $i = 1, \ldots, n$, the matrix $P$ satisfies

$$\sum_{j=1}^{n} P_{ij} = 1. \tag{2.2.1}$$

b. The vector $\pi_0$ satisfies

$$\sum_{i=1}^{n} \pi_{0i} = 1.$$

A matrix $P$ that satisfies property (2.2.1) is called a *stochastic matrix.* A stochastic matrix defines the probabilities of moving from each value of the state to any other in one period. The probability of moving from one value of the state to any other in *two* periods is determined by $P^2$ because

$$\begin{aligned}
&\text{Prob}\,(x_{t+2} = e_j | x_t = e_i) \\
&= \sum_{h=1}^{n} \text{Prob}\,(x_{t+2} = e_j | x_{t+1} = e_h)\,\text{Prob}\,(x_{t+1} = e_h | x_t = e_i) \\
&= \sum_{h=1}^{n} P_{ih} P_{hj} = P_{ij}^{(2)},
\end{aligned}$$

where $P_{ij}^{(2)}$ is the $i, j$ element of $P^2$. Let $P_{i,j}^{(k)}$ denote the $i, j$ element of $P^k$. By iterating on the preceding equation, we discover that

$$\text{Prob}\,(x_{t+k} = e_j | x_t = e_i) = P_{ij}^{(k)}.$$

The unconditional probability distributions of $x_t$ are determined by

$$\begin{aligned}
\pi'_1 &= \text{Prob}\,(x_1) = \pi'_0 P \\
\pi'_2 &= \text{Prob}\,(x_2) = \pi'_0 P^2 \\
& \\
\pi'_k &= \text{Prob}\,(x_k) = \pi'_0 P^k
\end{aligned}$$

where $\pi'_t = \text{Prob}(x_t)$ is the $(1 \times n)$ vector whose $i$th element is $\text{Prob}(x_t = e_i)$

### *2.2.1. Stationary distributions*

Unconditional probability distributions evolve according to

$$\pi'_{t+1} = \pi'_t P. \tag{2.2.2}$$

An unconditional distribution is called *stationary* or *invariant* if it satisfies

$$\pi_{t+1} = \pi_t,$$

that is, if the unconditional distribution remains unaltered with the passage of time. From the law of motion (2.2.2) for unconditional distributions, a stationary distribution must satisfy

$$\pi' = \pi' P \tag{2.2.3}$$

or

$$\pi' (I - P) = 0.$$

Transposing both sides of this equation gives

$$(I - P')\, \pi = 0, \tag{2.2.4}$$

which determines $\pi$ as an eigenvector (normalized to satisfy $\sum_{i=1}^{n} \pi_i = 1$) associated with a unit eigenvalue of $P'$.

The fact that $P$ is a stochastic matrix (i.e., it has nonnegative elements and satisfies $\sum_j P_{ij} = 1$ for all $i$) guarantees that $P$ has at least one unit eigenvalue, and that there is at least one eigenvector $\pi$ that satisfies equation (2.2.4). This stationary distribution may not be unique because $P$ can have a repeated unit eigenvalue.

*Example 1.* A Markov chain

$$P = \begin{bmatrix} 1 & 0 & 0 \\ .2 & .5 & .3 \\ 0 & 0 & 1 \end{bmatrix}$$

has two unit eigenvalues with associated stationary distributions $\pi' = [1 \quad 0 \quad 0]$ and $\pi' = [0 \quad 0 \quad 1]$. Here states 1 and 3 are both ***absorbing*** states. Furthermore, any initial distribution that puts zero probability on state 2 is a stationary distribution. See exercises *2.10* and *2.11*.

*Example 2.* A Markov chain

$$P = \begin{bmatrix} .7 & .3 & 0 \\ 0 & .5 & .5 \\ 0 & .9 & .1 \end{bmatrix}$$

has one unit eigenvalue with associated stationary distribution $\pi'$ $[0 \quad .6429 \quad .3571]$. Here states 2 and 3 form an *absorbing subset* of the state space

### *2.2.2. Asymptotic stationarity*

We often ask the following question about a Markov process: for an arbitrary initial distribution $\pi_0$, do the unconditional distributions $\pi_t$ approach a stationary distribution

$$\lim_{t \to \infty} \pi_t = \pi_\infty,$$

where $\pi_\infty$ solves equation (2.2.4)? If the answer is yes, then does the limit distribution $\pi_\infty$ depend on the initial distribution $\pi_0$? If the limit $\pi_\infty$ is independent of the initial distribution $\pi_0$, we say that the process is *asymptotically stationary with a unique invariant distribution.* We call a solution $\pi_\infty$ a *stationary distribution* or an *invariant distribution* of $P$.

We state these concepts formally in the following definition:

DEFINITION: Let $\pi_\infty$ be a unique vector that satisfies $(I - P')\pi_\infty = 0$. If for all initial distributions $\pi_0$ it is true that $P^{t\prime}\pi_0$ converges to the same $\pi_\infty$, we say that the Markov chain is asymptotically stationary with a unique invariant distribution.

The following theorems can be used to show that a Markov chain is asymptotically stationary.

THEOREM 1: Let $P$ be a stochastic matrix with $P_{ij} > 0 \; \forall (i,j)$. Then $P$ has a unique stationary distribution, and the process is asymptotically stationary.

THEOREM 2: Let $P$ be a stochastic matrix for which $P_{ij}^n > 0 \; \forall (i,j)$ for some value of $n \geq 1$. Then $P$ has a unique stationary distribution, and the process is asymptotically stationary.

The conditions of theorem 1 (and 2) state that from any state there is a positive probability of moving to any other state in one (or $n$) steps.

### *2.2.3. Expectations*

Let $\overline{y}$ be an $n \times 1$ vector of real numbers and define $y_t = \overline{y}' x_t$, so that $y_t = \overline{y}_i$ if $x_t = e_i$. From the conditional and unconditional probability distributions that we have listed, it follows that the unconditional expectations of $y_t$ for $t \geq 0$ are determined by $E y_t = (\pi_0' P^t) \overline{y}$. Conditional expectations are determined by

$$E\left(y_{t+1} | x_t = e_i\right) = \sum_j P_{ij} \overline{y}_j = (P\overline{y})_i \tag{2.2.5}$$

$$E\left(y_{t+2} | x_t = e_i\right) = \sum_k P_{ik}^{(2)} \overline{y}_k = \left(P^2 \overline{y}\right)_i \tag{2.2.6}$$

and so on, where $P_{ik}^{(2)}$ denotes the $(i,k)$ element of $P^2$. Notice that

$$\begin{aligned} E\left[E\left(y_{t+2} | x_{t+1} = e_j\right) | x_t = e_i\right] &= \sum_j P_{ij} \sum_k P_{jk} \overline{y}_k \\ = \sum_k \left(\sum_j P_{ij} P_{jk}\right) \overline{y}_k &= \sum_k P_{ik}^{(2)} \overline{y}_k = E\left(y_{t+2} | x_t = e_i\right) \end{aligned}$$

Connecting the first and last terms in this string of equalities yields $E[E(y_{t+2}|x_{t+1})|x_t] = E[y_{t+2}|x_t]$. This is an example of the "law of iterated expectations." The law of iterated expectations states that for any random variable $z$ and two information sets $J, I$ with $J \subset I$, $E[E(z|I)|J] = E(z|J)$. As another example of the law of iterated expectations, notice that

$$E y_1 = \sum_j \pi_{1,j} \overline{y}_j = \pi_1' \overline{y} = \left(\pi_0' P\right) \overline{y} = \pi_0' \left(P \overline{y}\right)$$

and that

$$E\left[E\left(y_1|x_0=e_i\right)\right]=\sum_i \pi_{0,i}\sum_j P_{ij}\overline{y}_j=\sum_j\left(\sum \pi_{0,i}P_{ij}\right)\overline{y}_j=\pi_1'\overline{y}=Ey_1$$

### 2.2.4. Forecasting functions

There are powerful formulas for forecasting functions of a Markov process. Again, let $\overline{y}$ be an $n \times 1$ vector and consider the random variable $y_t = \overline{y}' x_t$. Then

$$E\left[y_{t+k}|x_t=e_i\right]=\left(P^k\overline{y}\right)_i$$

where $(P^k\overline{y})_i$ denotes the $i$th row of $P^k\overline{y}$. Stacking all $n$ rows together, we express this as

$$E\left[y_{t+k}|x_t\right]=P^k\overline{y}. \tag{2.2.7}$$

We also have

$$\sum_{k=0}^{\infty}\beta^k E\left[y_{t+k}|x_t=\overline{e}_i\right]=\left[\left(I-\beta P\right)^{-1}\overline{y}\right]$$

where $\beta \in (0,1)$ guarantees existence of $(I-\beta P)^{-1}=(I+\beta P+\beta^2P^2+\cdots)$.

One-step-ahead forecasts of a sufficiently rich set of random variables characterize a Markov chain. In particular, one-step-ahead conditional expectations of $n$ independent functions (i.e., $n$ linearly independent vectors $h_1,\ldots,h_n$) uniquely determine the transition matrix $P$. Thus, let $E[h_{k,t+1}|x_t=e_i]=(Ph_k)_i$. We can collect the conditional expectations of $h_k$ for all initial states $i$ in an $n \times 1$ vector $E[h_{k,t+1}|x_t]=Ph_k$. We can then collect conditional expectations for the $n$ independent vectors $h_1,\ldots,h_n$ as $Ph=J$ where $h=[\,h_1 \;\; h_2 \;\; \ldots \;\; h_n\,]$ and $J$ is the $n \times n$ matrix consisting of all conditional expectations of all $n$ vectors $h_1,\ldots,h_n$. If we know $h$ and $J$, we can determine $P$ from $P=Jh^{-1}$.

## *2.2.5. Invariant functions and ergodicity*

Let $P, \pi$ be a stationary $n$-state Markov chain with the state space $X = [e_i, i = 1, \ldots, n]$. An $n \times 1$ vector $\overline{y}$ defines a random variable $y_t = \overline{y}' x_t$. Thus a random variable is another term for "function of the underlying Markov state."

The following is a useful precursor to a law of large numbers:

**Theorem 2.2.1.** *Let $\overline{y}$ define a random variable as a function of an underlying state $x$, where $x$ is governed by a stationary Markov chain $(P, \pi)$. Then*

$$\frac{1}{T} \sum_{t=1}^{T} y_t \rightarrow E\left[y_\infty | x_0\right] \tag{2.2.8}$$

*with probability* 1

Here $E[y_\infty | x_0]$ is the expectation of $y_s$ for $s$ very large, conditional on the initial state. We want more than this. In particular, we would like to be able to replace $E[y_\infty | x_0]$ with the constant unconditional mean $E[y_t] = E[y_0]$ associated with the stationary distribution. To get this requires that we strengthen what is assumed about $P$ by using the following concepts. First, we use

**Definition 2.2.1.** A random variable $y_t = \overline{y}' x_t$ is said to be *invariant* if $y_t = y_0, t \geq 0$, for any realization of $x_t, t \geq 0$.

Thus, a random variable $y$ is invariant (or "an invariant function of the state") if it remains constant while the underlying state $x_t$ moves through the state space $X$.

For a finite-state Markov chain, the following theorem gives a convenient way to characterize invariant functions of the state.

**Theorem 2.2.2.** *Let $P, \pi$ be a stationary Markov chain. If*

$$E\left[y_{t+1} | x_t\right] = y_t \tag{2.2.9}$$

*then the random variable $y_t = \overline{y}' x_t$ is invariant.*

*Proof.* By using the law of iterated expectations, notice that

$$\begin{aligned}
E\left(y_{t+1} - y_t\right)^2 &= E\left[E\left(y_{t+1}^2 - 2 y_{t+1} y_t + y_t^2\right) | x_t\right] \\
&= E\left[E y_{t+1}^2 | x_t - 2 E\left(y_{t+1} | x_t\right) y_t + E y_t^2 | x_t\right] \\
&= E y_{t+1}^2 - 2 E y_t^2 + E y_t^2 \\
&= 0
\end{aligned}$$

where the middle term on the right side of the second line uses that $E[y_t|x_t] = y_t$, the middle term on the right side of the third line uses the hypothesis (2.2.9), and the third line uses the hypothesis that $\pi$ is a stationary distribution. In a finite Markov chain, if $E(y_{t+1} - y_t)^2 = 0$, then $y_{t+1} = y_t$ for all $y_{t+1}, y_t$ that occur with positive probability under the stationary distribution. ∎

As we shall have reason to study in chapters 16 and 17, *any* (not necessarily stationary) stochastic process $y_t$ that satisfies (2.2.9) is said to be a *martingale.* Theorem 2.2.2 tells us that a martingale that is a function of a finite-state stationary Markov state $x_t$ must be constant over time. This result is a special case of the martingale convergence theorem that underlies some remarkable results about savings to be studied in chapter 16.[1]

Equation (2.2.9) can be expressed as $P\overline{y} = \overline{y}$ or

$$(P - I)\overline{y} = 0, \tag{2.2.10}$$

which states that an invariant function of the state is a (right) eigenvector of $P$ associated with a unit eigenvalue.

**Definition 2.2.2.** Let $(P, \pi)$ be a stationary Markov chain. The chain is said to be *ergodic* if the only invariant functions $\overline{y}$ are constant with probability 1, i.e., $\overline{y}_i = \overline{y}_j$ for all $i, j$ with $\pi_i > 0, \pi_j > 0$.

A law of large numbers for Markov chains is:

**Theorem 2.2.3.** *Let $\overline{y}$ define a random variable on a stationary and ergodic Markov chain $(P, \pi)$. Then*

$$\frac{1}{T}\sum_{t=1}^{T} y_t \rightarrow E[y_0] \tag{2.2.11}$$

*with probability* 1

This theorem tells us that the time series average converges to the population mean of the stationary distribution.

[1] Theorem 2.2.2 tells us that a stationary martingale process has so little freedom to move that it has to be constant forever, not just eventually, as asserted by the martingale convergence theorem.

Three examples illustrate these concepts.

**Example 1.** A chain with transition matrix $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ has a unique invariant distribution $\pi = [\,.5 \quad .5\,]'$ and the invariant functions are $[\,\alpha \quad \alpha\,]'$ for any scalar $\alpha$. Therefore, the process is ergodic and Theorem 2.2.3 applies.

**Example 2.** A chain with transition matrix $P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ has a continuum of stationary distributions $\gamma \begin{bmatrix} 1 \\ 0 \end{bmatrix} + (1 - \gamma) \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ for any $\gamma \in [0, 1]$ and invariant functions $\begin{bmatrix} 0 \\ \alpha \end{bmatrix}$ and $\begin{bmatrix} \alpha \\ 0 \end{bmatrix}$ for any $\alpha$. Therefore, the process is not ergodic. The conclusion (2.2.11) of Theorem 2.2.3 does not hold for many of the stationary distributions associated with $P$, but Theorem 2.2.1 does hold. Conclusion (2.2.11) does hold for one particular choice of stationary distribution.

**Example 3.** A chain with transition matrix $P = \begin{bmatrix} .8 & .2 & 0 \\ .1 & .9 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ has a continuum of stationary distributions $\gamma\,[\,\frac{1}{3} \quad \frac{2}{3} \quad 0\,]' + (1-\gamma)\,[\,0 \quad 0 \quad 1\,]'$ and invariant functions $\alpha\,[\,1 \quad 1 \quad 0\,]'$ and $\alpha\,[\,0 \quad 0 \quad 1\,]'$ for any scalar $\alpha$. The conclusion (2.2.11) of Theorem 2.2.3 does not hold for many of the stationary distributions associated with $P$, but Theorem 2.2.1 does hold. But again, conclusion (2.2.11) does hold for one particular choice of stationary distribution.

### *2.2.6. Simulating a Markov chain*

It is easy to simulate a Markov chain using a random number generator. The Matlab program `markov.m` does the job. We'll use this program in some later chapters.[2]

[2] An index in the back of the book lists Matlab programs.

### 2.2.7. The likelihood function

Let $P$ be an $n \times n$ stochastic matrix with states $1, 2, \ldots, n$. Let $\pi_0$ be an $n \times 1$ vector with nonnegative elements summing to $1$, with $\pi_{0,i}$ being the probability that the state is $i$ at time $0$. Let $i_t$ index the state at time $t$. The Markov property implies that the probability of drawing the path $(x_0, x_1, \ldots, x_{T-1}, x_T) = (\overline{e}_{i_0}, \overline{e}_{i_1}, \ldots, \overline{e}_{i_{T-1}}, \overline{e}_{i_T})$ is

$$\begin{aligned} L &\equiv \text{Prob}\left(\overline{x}_{i_T}, \overline{x}_{i_{T-1}}, \ldots, \overline{x}_{i_1}, \overline{x}_{i_0}\right) \\ &= P_{i_{T-1}, i_T} P_{i_{T-2}, i_{T-1}} \cdots P_{i_0, i_1} \pi_{0, i_0}. \end{aligned} \tag{2.2.12}$$

The probability $L$ is called the *likelihood*. It is a function of both the sample realization $x_0, \ldots, x_T$ and the parameters of the stochastic matrix $P$. For a sample $x_0, x_1, \ldots, x_T$, let $n_{ij}$ be the number of times that there occurs a one-period transition from state $i$ to state $j$. Then the likelihood function can be written

$$L = \pi_{0,i_0} \prod_i \prod_j P_{i,j}^{n_{ij}},$$

a *multinomial* distribution.

Formula (2.2.12) has two uses. A first, which we shall encounter often, is to describe the probability of alternative histories of a Markov chain. In chapter 8, we shall use this formula to study prices and allocations in competitive equilibria.

A second use is for estimating the parameters of a model whose solution is a Markov chain. Maximum likelihood estimation for free parameters $\theta$ of a Markov process works as follows. Let the transition matrix $P$ and the initial distribution $\pi_0$ be functions $P(\theta), \pi_0(\theta)$ of a vector of free parameters $\theta$. Given a sample $\{x_t\}_{t=0}^T$, regard the likelihood function as a function of the parameters $\theta$. As the estimator of $\theta$, choose the value that maximizes the likelihood function $L$.

## 2.3. Continuous-state Markov chain

In chapter 8 we shall use a somewhat different notation to express the same ideas. This alternative notation can accommodate either discrete- or continuous-state Markov chains. We shall let $S$ denote the state space with typical element $s \in S$. The *transition density* is $\pi(s'|s) = \text{Prob}(s_{t+1} = s'|s_t = s)$ and the initial density is $\pi_0(s) = \text{Prob}(s_0 = s)$. For all $s \in S, \pi(s'|s) \geq 0$ and $\int_{s'} \pi(s'|s)ds' = 1$; also $\int_s \pi_0(s)ds = 1$.[3] Corresponding to (2.2.12), the likelihood function or density over the history $s^t = [s_t, s_{t-1}, \dots, s_0]$ is

$$\pi\left(s^t\right) = \pi\left(s_t|s_{t-1}\right) \cdots \pi\left(s_1|s_0\right) \pi_0\left(s_0\right). \tag{2.3.1}$$

For $t \geq 1$, the time $t$ unconditional distributions evolve according to

$$\pi_t\left(s_t\right) = \int_{s_{t-1}} \pi\left(s_t|s_{t-1}\right) \pi_{t-1}\left(s_{t-1}\right) d\, s_{t-1}.$$

A stationary or *invariant* distribution satisfies

$$\pi_\infty\left(s'\right) = \int_s \pi\left(s'|s\right) \pi_\infty\left(s\right) d\, s,$$

which is the counterpart to (2.2.3).

Paralleling our discussion of finite-state Markov chains, we can say that the function $\phi(s)$ is invariant if

$$\int \phi\left(s'\right) \pi\left(s'|s\right) ds' = \phi\left(s\right).$$

A stationary continuous-state Markov process is said to be *ergodic* if the only invariant functions $p(s')$ are constant with probability 1 according to the stationary distribution $\pi_\infty$. A law of large numbers for Markov processes states:

**Theorem 2.3.1.** *Let $y(s)$ be a random variable, a measurable function of $s$, and let $(\pi(s'|s), \pi_0(s))$ be a stationary and ergodic continuous-state Markov process. Assume that $E|y| < +\infty$. Then*

$$\frac{1}{T}\sum_{t=1}^{T} y_t \rightarrow Ey = \int y\left(s\right) \pi_0\left(s\right) ds$$

*with probability 1 with respect to the distribution $\pi_0$.*

[3] Thus, when $S$ is discrete, $\pi(s_j|s_i)$ corresponds to $P_{s_i,s_j}$ in our earlier notation.