Chapter 3

Markov Processes

Let X_t be a random vector that we call the *state* at t, which we take to be a complete description of the position at time t of a system of interest. We can construct a consistent sequence of probability distributions Pr_ℓ for a sequence of random vectors

$$X^{[\ell]} \doteq \begin{bmatrix} X_0 \\ X_1 \\ \vdots \\ X_\ell \end{bmatrix}$$

for all nonnegative integers ℓ by specifying the following two elementary components of a *Markov process*: (i) a probability distribution for X_0, and (ii) a time-invariant distribution for X_{t+1} conditional on X_t for $t \geq 0$. All other probabilities are functions of these two distributions. By creatively defining the state vector X_t, a Markov specification is flexible enough to include many models used in applied research.

3.1 Components

Assume a state space $\mathcal{X}$ and a transition distribution $P(dx^*|x)$. For example, $\mathcal{X}$ could be $\mathbb{R}^n$ or some subset of $\mathbb{R}^n$. The transition distribution P is a conditional probability measure for each choice $X_t = x$ in the state space, so it satisfies $\int_{\{x^* \in \mathcal{X}\}} P(dx^*|x) = 1$ for every x in the state space. If in addition we specify a marginal distribution Q_0 for the initial state x_0 over $\mathcal{X}$, then we have completely specified all joint distributions for the stochastic process $\{X_t, t = 0, 1, \ldots\}$.

The notation $P(dx^*|x)$ denotes a conditional probability measure; integration is over x^* and conditioning is captured by x. Specifically, x^* is a possible realization of the next period state and x is a realization of the current period state. The conditional probability measure $P(dx^*|x)$ assigns conditional probabilities to the next period state given that the current period state is x. Often, but not always, the conditional distributions have densities against a common distribution $\lambda(dx^*)$ to be used to integrate over states. That lets us use a *transition density* to represent the conditional probability measure.

A first-order vector autoregression is an example of a Markov process. Here $Q_0(x)$ is a normal distribution with mean μ_0 and covariance matrix Σ_0 and $P(dx^*|x)$ is a normal distribution with mean Ax and covariance matrix BB' for a square matrix A and a matrix B with full column rank.[1] These conditions imply the representation

$$X_{t+1} = AX_t + BW_{t+1},$$

where W_{t+1} is a multivariate standard normally distributed random vector that is independent of X_t.

Another example is a discrete-state Markov chain in which Q_0 can be represented as a row vector and $P(dx^*|x)$ can be represented as a matrix with one row and one column for each possible value of the state x. The rows are vectors of probabilities of next period's state conditioned on a realized value of this period's state. We study both examples more later.

An important object is a one-step conditional expectation operator to be applied to functions of a Markov state. Let $f : \mathcal{X} \to \mathbb{R}$. For bounded f, define:

$$\mathbb{T}f(x) = E\left[f(X_{t+1})|X_t = x\right] = \int_{\{x^* \in \mathcal{X}\}} f(x^*)P(dx^*|x).$$

Iterating on $\mathbb{T}$ allows us to form conditional expectations over longer horizons:

$$\mathbb{T}^j f(x) = E\left[f(X_{t+j})|X_t = x\right],$$

a manifestation of the Law of Iterated Expectations.

[1] When BB' is singular, a density may not exist with respect to Lebesgue measure. Singularity occurs when we convert a higher-order vector autoregression into a first-order process.

Remark 3.1.1. *Instead of beginning with a conditional probability measure $P(dx^*|x)$, we can start with a conditional expectation operator $\mathbb{T}$ that maps a space of bounded functions into itself. We can construct a conditional probability measure $P(dx^*|x)$ from the operator $\mathbb{T}$ provided that $\mathbb{T}$ is a) well defined on the space of bounded functions, b) preserves the bound, c) maps nonnegative functions into nonnegative functions, and d) maps the unit function into the unit function.*

3.2 Stationarity

By choosing the distribution of the initial state X_0 appropriately, we can construct a stationary Markov process .

Definition 3.2.1. *A* ***stationary distribution*** *for a Markov process is a probability measure Q over a state space $\mathcal{X}$ that satisfies*

$$\int_{\{x\in\mathcal{X}\}} P(dx^*|x)Q(dx) = Q(dx^*).$$

We will sometimes refer to a stationary density q. A density is always relative to a measure. With this in mind, let λ be a measure on the state space $\mathcal{X}$ to be used to integrate over possible Markov states. Then a density q is a nonnegative (Borel measurable) function of the state for which $\int q(x)\lambda(dx) = 1$.

Definition 3.2.2. *A* ***stationary density*** *for a Markov process is a probability density q with respect to a measure λ over the state space $\mathcal{X}$ that satisfies*

$$\int P(dx^*|x)q(x)\lambda(dx) = q(x^*)\lambda(dx^*).$$

There are various sets of conditions that imply existence of a stationary distribution. Given a transition distribution P, one widely used sufficient condition is that the Markov process be *time reversible*, which means that

$$P(dx^*|x)Q(dx) = P(dx|x^*)Q(dx^*) \tag{3.1}$$

for some probability distribution Q on $\mathcal{X}$. Because a transition density satisfies $\int_{\{x\in\mathcal{X}\}} P(dx|x^*) = 1$,

$$\int_{\{x\in\mathcal{X}\}} P(dx^*|x)Q(dx) = \int_{\{x\in\mathcal{X}\}} P(dx|x^*)Q(dx^*) = Q(dx^*),$$

as required by Definition 3.2.1 for Q to be a stationary distribution. Restriction (3.1) imposes that forward and backward transition distributions coincide, implying that the process is time reversible. Time reversibility is special, so later we will explore other sufficient conditions for the existence of stationary distributions.[2]

Remark 3.2.3. *When a Markov process is initialized at a stationary distribution, we can construct the process $\{X_t : t = 1, 2, ...\}$ with a measure-preserving transformation $\mathbb{S}$ of the type featured in section 2.2.*

3.3 $\mathcal{L}^2$ and Eigenfunctions

Given a stationary distribution Q, form the space of functions $\mathcal{L}^2$

$$\mathcal{L}^2 = \{f : \mathcal{X} \to \mathbb{R} : \int f(x)^2 Q(dx) < \infty\}.$$

It can be shown that $\mathbb{T} : \mathcal{L}^2 \to \mathcal{L}^2$. On $\mathcal{L}^2$, a well defined norm is

$$\|f\| = \left[\int f(x)^2 Q(dx)\right]^{1/2}.$$

We now study eigenfunctions of the conditional expectation operator $\mathbb{T}$ that generalizes the concept of an eigenvector of a matrix.

Definition 3.3.1. *A function $\tilde{f} \in \mathcal{L}^2$ that solves $\mathbb{T} f = f$ is called an eigenfunction of $\mathbb{T}$ associated with a unit eigenvalue.*

An eigenfunction of $\mathbb{T}$ generalizes an eigenvector of a matrix. An eigenfunction $\tilde{f}$ associated with a unit eigenvalue is constant through time.

Proposition 3.3.2. *Suppose that $\tilde{f}$ is an eigenfunction of $\mathbb{T}$ associated with a unit eigenvalue. Then $\{\tilde{f}(X_t) : t = 0, 1, ...\}$ is constant over time with probability one.*

[2]Numerical Bayesian statistical analysis often computes a posterior probability distribution by iterating to convergence a reversible Markov process whose stationary distribution is that posterior distribution.

Proof.

$$E\left[\tilde{f}(X_{t+1})\tilde{f}(X_t)\right] = \int (\mathbb{T}\tilde{f})(x)\tilde{f}(x)Q(dx) = \int \tilde{f}(x)^2 Q(dx) = E\left[\tilde{f}(X_t)^2\right].$$

Then because Q is a stationary distribution,

$$\begin{aligned} E\left([\tilde{f}(X_{t+1}) - \tilde{f}(X_t)]^2\right) &= E\left[\tilde{f}(X_{t+1})^2\right] + E\left[\tilde{f}(X_t)^2\right] \\ &\quad -2E\left[\tilde{f}(X_{t+1})\tilde{f}(X_t)\right] \\ &= 0. \end{aligned}$$

□

3.4 Ergodic Markov Processes

Chapter 2 used the ergodic property to define a special kind of 'statistical model' that is affiliated with a Law of Large Numbers that implies that limit points are constant across sample points $\omega \in \Omega$. Section 2.7 described how other thatstatistical modelsthat lack the ergodic property are building blocks of more general probability specifications that we can use to express what it means not to know a statistical model.[3] We now explore ergodicity in the context of Markov processes.

Given Proposition 3.3.2, time-series averages of an eigenfunction $\mathbb{T}\tilde{f} = \tilde{f}$ are invariant over time, so

$$\frac{1}{N}\sum_{t=1}^{N} \tilde{f}(X_t) = \tilde{f}(X).$$

However, when $\tilde{f}(x)$ varies across sets of states x that occur with positive probability under Q, the time series average $\frac{1}{N}\sum_{t=1}^{N} \tilde{f}(X_t)$ can differ from $\int \tilde{f}(x)Q(dx)$. This happens when the variation of $\tilde{f}(X_t)$ along a sample path for $\{X_t\}$ conveys an inaccurate impression of variation across the stationary distribution $Q(dx)$. See example 3.6.2 below. We can rule out divergences in patterns of variation across time by imposing an assumption about the eigenfunction equation $\mathbb{T}f = f$.

[3]Unknown parameters manifest themselves as unknown statistical models.

Proposition 3.4.1. *When a unique solution to the eigenvalue equation*

$$\mathbb{T}f = f$$

is a constant function (with Q measure one), then it is possible to construct the process $\{X_t : t = 0, 1, 2, ...\}$ by using a transformation $\mathbb{S}$ that is measure preserving and ergodic.

Here ergodicity is a property that obtains relative to a stationary distribution of the Markov process. If there are multiple stationary distributions, it is possible that there is a unique constant solution to the eigenvalue problem for one stationary distribution, but that non constant solutions exist for other stationary distributions. Example 2.3.4 is an instance in which any assignment of probabilities constitutes a stationary distribution, but ergodicity prevails only if we assign probability one to one of the two states. (Also see example 3.6.3.)

Invariant events for a Markov process

Consider an eigenfunction $\tilde{f}$ of $\mathbb{T}$ associated with a unit eigenvalue. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a bounded Borel measurable function. Since $\{\tilde{f}(X_t) : t = 0, 1, 2, ...\}$ is invariant over time, so is $\left\{\phi\left[\tilde{f}(X_t)\right] : t = 0, 1, 2, \ldots\right\}$ and it is necessarily true that

$$\mathbb{T}(\phi \circ \tilde{f}) = \phi \circ \tilde{f}.$$

Therefore, given an eigenfunction $\tilde{f}$ associated with a unit eigenvalue, we can construct other eigenfunctions.[4] For example,

$$\phi[\tilde{f}(x)] = \begin{cases} 1 & \text{if } \tilde{f}(x) \in \mathfrak{b} \\ 0 & \text{if } \tilde{f}(x) \notin \mathfrak{b} \end{cases} \tag{3.2}$$

for some Borel set $\mathfrak{b}$ in $\mathbb{R}$. It follows that

$$\Lambda = \{\omega \in \Omega : \tilde{f}[X(\omega)] \in \mathfrak{b}\}$$

is an invariant event in Ω. Note that

$$\tilde{\mathfrak{b}} = \left\{x : \tilde{f}(x) \in \mathfrak{b}\right\}$$

[4]This construction also works for unbounded functions ϕ provided that $\phi \circ \tilde{f}$ is square integrable under the Q measure.

which is a Borel set. Then we can also represent Λ as

$$\Lambda = \left\{ \omega \in \Omega : X(\omega) \in \tilde{\mathfrak{b}} \right\}. \tag{3.3}$$

For Markov processes, all invariant events can be represented in terms of the location of the initial state X. See Doob (1953), Theorem 1.1, page 460. Thus, associated with an invariant event is a Borel set in $\mathcal{X}$. Let $\mathfrak{J}$ denote the collection of Borel subsets of $\mathcal{X}$ for which Λ constructed as in (3.3) is an invariant event. If $b \in \mathfrak{J}$, then the indicator function

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathfrak{b} \\ 0 & \text{if } x \notin \mathfrak{b} \end{cases} \tag{3.4}$$

satisfies

$$\mathbb{T} f = f$$

with Q probability one. If the probability of Λ is neither zero nor one, then the function f is nonnegative, strictly positive on a set of positive Q measure, and zero on a set with strictly positive Q measure. When the Markov process $\{X_t\}$ is not ergodic, there exist bounded eigenfunctions with unit eigenvalues that are not constant with Q measure one.

To assure that the eigenfunction $\tilde{f}$ is constant with Q measure one, we can appeal to a property of the *resolvent operator* $\mathbb{M}$ associated with a constant discount factor $0 < \delta < 1$:

$$\mathbb{M} f(x) = (1 - \delta) \sum_{j=0}^{\infty} \delta^j \mathbb{T}^j f.$$

Notice that if $\tilde{f}$ is an eigenfunction of $\mathbb{T}$ associated with a unit eigenvalue, then it is also an eigenfunction of $\mathbb{M}$ associated with a unit eigenvalue.

Proposition 3.4.2. *Suppose that for any $f \geq 0$ such that $\int f(x) Q(dx) > 0$, $\mathbb{M} f(x) > 0$ for all $x \in \mathcal{X}$ with Q measure one. Then any solution $\tilde{f}$ to $\mathbb{T} f = f$ is necessarily constant with Q measure one.*

Proof. Consider an eigenfunction $\tilde{f}$ associated with a unit eigenvalue. The function $f = \phi \circ \tilde{f}$ necessarily satisfies:

$$\mathbb{M} f = f$$

for any ϕ of the form (3.2). If such an f also satisfies $\int f(x) Q(dx) > 0$, then f is one with Q probability one. Since this holds for any Borel set $\mathfrak{b}$ in $\mathbb{R}$, $\tilde{f}$ must be constant with Q probability one. □

Proposition 3.4.2 supplies a sufficient condition for ergodicity. A more general sufficient condition is that there exists an integer $m \geq 1$ such that

$$\mathbb{T}^m f(x) > 0$$

for any $f \geq 0$ such that $\int f(x)Q(dx) > 0$ on a set with Q measure one.

Irreducibility

We now discuss a property of a stochastic process called *irreducibility* that is a useful tool to verify ergodicity. We do this by introducing a probability measure $\widetilde{Q}$ over $\mathcal{X}$ that we do not require to be a stationary probability measure. We need not even assume that a stationary probability measure exists.

We begin by recalling that the conditional expectation operator $\mathbb{T}$ and the resolvent $\mathbb{M}$ both map bounded functions into bounded functions. For a probability measure $\widetilde{Q}$ over $\mathcal{X}$ that is not stationary, we can still consider an eigenfunction problem

$$\mathbb{T}f = f$$

for bounded functions f. Suppose that f is an indicator function for a Borel set $\mathfrak{b}$. Then $\mathbb{T}f$ is the conditional probability that X_{t+1} is in $\mathfrak{b}$ conditioned on X_t being in $\mathfrak{b}$. When such an indicator function f is an eigenfunction with a unit eigenvalue, the conditional probability is one whenever X_t itself is in $\mathfrak{b}$ and it is zero otherwise. Thus, $\{f(X_t)\}$ is invariant over time.

Definition 3.4.3. *The process $\{X_t\}$ is said to be irreducible with respect to $\widetilde{Q}$ if for any $f \geq 0$ such that $\int f(x)\widetilde{Q}(dx) > 0$, $\mathbb{M}f(x) > 0$ for all $x \in \mathcal{X}$ with $\widetilde{Q}$ measure one.*

Proposition 3.4.4. *When $\widetilde{Q}$ is a stationary distribution and $\{X_t\}$ is irreducible with respect to $\widetilde{Q}$, the process is necessarily ergodic.*

Proof. By imitating the proof of proposition 3.4.2, we can establish that irreducibility rules out bounded eigenfunctions that are not constant with $\widetilde{Q}$ measure one. □

Periodicity

Next, in the spirit of example 2.B.1, we study a notion of periodicity of a stationary Markov process.[5] To define periodicity of a Markov process, for a given positive integer p we construct a new Markov process by sampling an original process every p time periods. This is sometimes called 'skip-sampling' with sampling interval p.[6] With a view toward applying Proposition 3.3.2 to $\mathbb{T}^p$, solve

$$\mathbb{T}^p f = f \tag{3.5}$$

for a function $\tilde{f}$. We know from Proposition 3.3.2 that for an $\tilde{f}$ that solves (3.5), $\{\tilde{f}(X_t) : t = 0, p, 2p, \ldots\}$ is invariant and so is $\{\tilde{f}(X_t) : t = 1, p+1, 2p+1, ...\}$. The process $\tilde{f}(X_t)$ is periodic with period p or np for any positive integer n.

Definition 3.4.5. *The* periodicity *of an irreducible Markov process* $\{X_t\}$ *with respect to* $\widetilde{Q}$ *is the smallest positive integer* p *such that there is a solution to equation* (3.5) *that is not constant with* $\widetilde{Q}$ *measure one. When there is no such integer* p*, we say that the process is* aperiodic.

Result 3.4.6. *Consider a counterpart of the resolvent operator* $\mathbb{M}$ *constructed by sampling at interval given by positive integer* p*:*

$$\mathbb{M}_p f(x) = (1-\delta)\sum_{j=0}^{\infty} \delta^j \mathbb{T}^{pj} f. \tag{3.6}$$

Provided that $\mathbb{M}_p f(x) > 0$ *with* $\widetilde{Q}$ *measure one and all* $p \geq 0$ *for any* $f \geq 0$ *such that* $\int f(x)Q(dx) > 0$*, the Markov process is aperiodic.*

3.5 Limits of Multi-Period Forecasts

When a Markov process is aperiodic, there are interesting situations in which

$$\lim_{j\to\infty} \mathbb{T}^j f(x) = \mathbf{r} \tag{3.7}$$

[5]Our definition of periodicity is confines to a stationary distribution. Actually, periodicity can be defined more generally. We limit our treatment of periodicity to specifications of transition probabilities for which there exist stationary distributions for convenience here.

[6]See Hansen and Sargent (1993) and Hansen and Sargent (2013, ch. 14).

for some $\mathbf{r} \in \mathbb{R}$, where convergence is either pointwise in x or in the $\mathcal{L}^2$ norm. Limit (3.7) asserts that long-term forecasts do not depend on the current Markov state. (See Meyn and Tweedie (1993) for a comprehensive treatment of this type of convergence.) Let Q be a stationary distribution. Then it is necessarily true that

$$\int \mathbb{T}^j f(x) Q(dx) = \int f(x) Q(dx)$$

for all j. Thus,

$$\mathbf{r} = \int f(x) Q(dx),$$

so that the limiting forecast is necessarily the mathematical expectation under a stationary distribution. Although here we have assumed that the limit point is a number and not a random variable, we have not assumed that the stationary distribution is unique.

Notice that if (3.7) is satisfied, then any function f that satisfies

$$\mathbb{T} f = f$$

is necessarily constant with probability one. Also, if $\int f(x) Q(dx) = 0$ and convergence is sufficiently fast, then

$$\lim_{N \to \infty} \sum_{j=0}^{N} \mathbb{T}^j f(x) \tag{3.8}$$

is a well-defined function of the Markov state. We will study the limiting function (3.8) in depth in chapter 4.

A set of sufficient conditions for

$$\lim_{j \to \infty} \mathbb{T}^j f(x^*) \to \int f(x) Q(dx) \tag{3.9}$$

for each $x^* \in \mathcal{X}$ and each bounded f is:[7]

Condition 3.5.1. *Suppose that the stationary Markov process satisfies:*

[7] Restriction 3.9 is stronger than ergodicity. It rules out periodic processes, although we know that periodic processes can be ergodic.

(i) For any $f \geq 0$ *such that* $\int f(x)Q(dx) > 0$, $\mathbb{M}_p f(x) > 0$ *for all* $x \in \mathcal{X}$ *with* Q *measure one and all positive integers* $p \geq 0$*, where the operator* $\mathbb{M}_p$ *is defined in* (3.6).

(ii) $\mathbb{T}$ *maps bounded continuous functions into bounded continuous functions, i.e., the Markov process is said to satisfy the Feller property.*

(iii) The support of Q *has a nonempty interior in* $\mathcal{X}$.

(iv) $\mathbb{T}V(x) - V(x) \leq -1$ *outside a compact subset of* $\mathcal{X}$ *for some nonnegative function* V.

We encountered condition (i) in our section 3.4 discussion of Markov processes that are ergodic and aperiodic. Condition *(*iv) is the *drift condition* for stability that requires that we find a function V that satisfies the requisite inequality. Heuristically, the drift condition says that outside a compact subset of the state space, the conditional expectation pushes inward. The choice of -1 as a comparison point is made only for convenience, since we can always multiply the function V by a number greater than one. Thus, -1 could be replaced by any strictly negative number. In section 3.7, we will apply condition 3.5.1 to verify ergodicity of a vector autoregression (a VAR for short).

3.6 Finite-State Markov Chains

Suppose that $\mathcal{X}$ consists of possible n states. We can label these states in a variety of ways, but for now we suppose that state x_j is the coordinate vector consisting entirely of zeros except in position j, where there is a one. Let $\mathbb{P}$ be an n by n transition matrix, where entry i, j is the probability of moving from state i to state j in a single period. Thus, the entries of $\mathbb{P}$ are all nonnegative and

$$\mathbb{P}\mathbf{1}_n = \mathbf{1}_n,$$

where $\mathbf{1}_n$ is an n-dimensional vector of ones.

Let $\mathbf{q}$ be an n-dimensional vector of probabilities. Stationarity requires that

$$\mathbf{q}'\mathbb{P} = \mathbf{q}', \tag{3.10}$$

so that $\mathbf{q}$ is a row eigenvector (also called a left eigenvector) of $\mathbb{P}$ associated with a unit eigenvalue.

We use a vector $\mathbf{f}$ to represent a function from the state space to the real line, where each coordinate of $\mathbf{f}$ gives the value of the function at the corresponding coordinate vector. Then the conditional expectation operator $\mathbb{T}$ can be represented as a matrix that equals $\mathbb{P}$:

$$E\left(\mathbf{f} \cdot X_{t+1} | X_t = x\right) = (\mathbb{T}\mathbf{f}) \cdot x = x'\mathbb{P}\mathbf{f}.$$

Consider column eigenvectors (also called right eigenvectors) of $\mathbb{P}$ associated with a unit eigenvalue. Suppose that the only solutions to

$$\mathbb{T}\mathbf{f} = \mathbf{f}$$

are of the form $\mathbf{f} \propto \mathbf{1}_n$, where $\propto$ means 'proportional to'. Then we can construct a process that is stationary and ergodic by initializing the process with density $\mathbf{q}$ determined by equation (3.10).

We can weaken this sufficient condition for stationarity and ergodicity to allow nonconstant right eigenvectors. A weaker condition is that there exists a real number $\mathbf{r}$ such that the right eigenvector $\mathbf{f}$ and the stationary distribution $\mathbf{q}$ satisfy

$$\min_{\mathbf{r}} \sum_{i=1}^{n} (\mathbf{f}_i - \mathbf{r})^2 \mathbf{q}_i = 0.$$

Notice that if $\mathbf{q}_i$ is zero, the contribution of $\mathbf{f}_i$ to the least squares objective can be neglected. This allows for non-constant $\mathbf{f}$'s, albeit in a limited way.

Three examples illustrate these concepts.

Example 3.6.1. *Recast Example 2.3.3 as a Markov chain with transition matrix* $\mathbb{P} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. *This chain has a unique stationary distribution* $q = \begin{bmatrix} .5 & .5 \end{bmatrix}'$ *and the invariant functions are* $\begin{bmatrix} \mathbf{r} & \mathbf{r} \end{bmatrix}'$ *for any scalar* $\mathbf{r}$. *Therefore, the process initiated from the stationary distribution is ergodic. The process is periodic with period two since the matrix* $\mathbb{P}^2$ *is an identity matrix and all two dimensional vectors are eigenvectors associated with a unit eigenvalue.*

Example 3.6.2. *Recast Example 2.3.4 as a Markov chain with transition matrix* $\mathbb{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. *This chain has a continuum of stationary distributions* $\pi \begin{bmatrix} 1 \\ 0 \end{bmatrix} + (1 - \pi) \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ *for any* $\pi \in [0, 1]$ *and invariant functions* $\begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix}$ *for any* *scalars* $\mathbf{r}_1, \mathbf{r}_2$*. Therefore, the process is not ergodic when* $\pi \in (0,1)$ *because if* $\mathbf{r}_1 \neq \mathbf{r}_2$ *the resulting invariant function fails to be constant across states that have positive probability under the stationary distribution associated with* $\pi \in (0,1)$*. When* $\pi \in (0,1)$*, nature chooses state* $i = 1$ *or* $i = 2$ *with probabilities* $\pi, 1-\pi$*, respectively, at time* 0*. Thereafter, the chain remains stuck in the realized time* 0 *state. Its failure ever to visit the unrealized state prevents the sample average from converging to the population mean of an arbitrary function of the state.*

Example 3.6.3. *A Markov chain with transition matrix* $\mathbb{P} = \begin{bmatrix} .8 & .2 & 0 \\ .1 & .9 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ *has a continuum of stationary distributions* $\pi \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 \end{bmatrix}' + (1-\pi) \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}'$ *for* $\pi \in [0,1]$ *and invariant functions* $\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_1 & \mathbf{r}_2 \end{bmatrix}'$ *for any scalars* $\mathbf{r}_1, \mathbf{r}_2$*. Under any stationary distribution associated with* $\pi \in (0,1)$*, the chain is not ergodic because some invariant functions are not constant with probability one. But under stationary distributions associated with* $\pi = 1$ *or* $\pi = 0$*, the chain is ergodic.*

3.7 Vector Autoregressions

When the eigenvalues of a square matrix A have absolute values that are strictly less than one, we say that A is *stable*. For a stable A, suppose that

$$X_{t+1} = AX_t + BW_{t+1},$$

where $\{W_{t+1} : t = 1, 2, ...\}$ is an iid sequence of multivariate normally distributed random vectors with mean vector zero and covariance matrix I. To complete the specification of a Markov process, we specify an initial distribution $X_0 \sim \mathcal{N}(\mu_0, \Sigma_0)$.

Let $\mu_t = EX_t$. Notice that

$$\mu_{t+1} = A\mu_t.$$

The mean μ of a stationary distribution satisfies

$$\mu = A\mu.$$

Because we have assumed that A is a stable matrix, $\mu = 0$ is the only solution of $(A - I)\mu = 0$. Thus, the mean of the stationary distribution is $\mu = 0$.

Let $\Sigma_t = E(X_t - \mu_t)(X_t - \mu_t)'$ be the covariance matrix of X_t. Then

$$\Sigma_{t+1} = A\Sigma_t A' + BB'.$$

For $\Sigma_t = \Sigma$ to be invariant over time, it must satisfy the discrete Lyapunov equation

$$\Sigma = A\Sigma A' + BB'. \tag{3.11}$$

When A is a stable matrix, this equation has a unique solution for a positive semidefinite matrix Σ.

Suppose that $\Sigma_0 = 0$ (a matrix of zeros). Then for $t \geq 1$

$$\Sigma_t = \sum_{j=0}^{t-1} A^j BB'(A^j)'.$$

The limit of this sequence is

$$\Sigma = \sum_{j=0}^{\infty} A^j BB'(A^j)',$$

which can be verified to satisfy Lyapunov equation (3.11) and thus equals the covariance matrix of the stationary distribution.[8] Similarly,

$$\mu_t = A^t \mu_0,$$

converges to zero for all $\mu_0 = EX_0$. Recall that 0 is also the mean of the stationary distribution.

The linear structure of the model implies that the stationary distribution is Gaussian with mean μ and covariance matrix Σ. To verify ergodicity, suppose that the covariance matrix Σ of the stationary distribution has full rank and verify conditions 3.5.1. Restriction (*iii*) of Condition 3.5.1 is satisfied. Furthermore, Σ_t has full rank for some t, which guarantees that

[8] To verify the asserted equality, notice that $\sum_{j=0}^{\infty} A^j BB' A^{j'} = A(\sum_{j=0}^{\infty} A^j BB' A^{j'})A' + BB'$.

the process is irreducible and aperiodic, i.e., restriction (*i*) is satisfied. As a candidate for $V(x)$ in condition (*iv*), take $V(x) = |x|^2$. Then

$$\mathbb{T}V(x) = x'A'Ax + \text{trace}(B'B)$$

so

$$\mathbb{T}V(x) - V(x) = x'(A'A - I)x + \text{trace}(B'B).$$

That A is a stable matrix implies that $A'A - I$ is negative definite, so that drift restriction (*iv*) of Condition 3.5.1 is satisfied for $|x|$ sufficiently large.[9]

We can extend this example to allow the mean of the stationary distribution not to be zero. Partition the Markov state as

$$x = \begin{bmatrix} x^{[1]} \\ x^{[2]} \end{bmatrix}$$

where $x^{[2]}$ is scalar. Similarly, partition the matrices A and B as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & 1 \end{bmatrix}$$
$$B = \begin{bmatrix} B_1 \\ 0 \end{bmatrix}$$

where A_{11} is a stable matrix. Notice that

$$X_{t+1}^{[2]} = X_t^{[2]} = \cdots = X_0^{[2]}$$

and hence is invariant. Let $\mu^{[2]}$ denote the mean of $X_t^{[2]}$ for any t. In a stationary distribution we require that the mean $\mu^{[1]}$ of $X_t[1]$ satisfy

$$\mu^{[1]} = A_{11}\mu^{[1]} + A_{12}\mu^{[2]}.$$

Hence

$$\mu^{[1]} = (I - A_{11})^{-1} A_{12}\mu^{[2]}.$$

Imitating our earlier argument, the covariance matrix, $\Sigma^{[11]}$ of $X_t[1]$ satisfies

$$\Sigma^{[11]} = \sum_{j=0}^{\infty} (A_{11})^j B_1(B_1)' ({A_{11}}')^j + (I - A_{11})^{-1} A_{12}\Sigma^{[22]}{A_{12}}' (I - {A_{11}}')^{-1}$$

[9]The Feller property (*ii*) of Condition 3.5.1 can also be established.

where $\Sigma^{[22]}$ is the variance of $X_t^{[2]}$ for all t. Stationarity imposes no restriction on the mean $\mu^{[2]}$ and the variance $\Sigma^{[22]}$.

Since $\{X_t^{[2]} : t = 0, 1, ...\}$ is invariant, the process $\{X_t : t = 0, 1, ...\}$ is ergodic only when the variance $\Sigma^{[22]}$ is zero. When $\{X_t : t = 0, 1, ...\}$ is not ergodic, the limit points in the Law of Large Numbers (Theorem 2.5.1) should be computed by conditioning on $X_0^{[2]}$.

3.8 Estimating Vector Autoregressions

Let Y_{t+1} be one of the entries of X_{t+1}, and consider the regression equation:

$$Y_{t+1} = \beta \cdot X_t + U_{t+1},$$

where U_{t+1} is a least squares residual. In the spirit of chapter 2, we allow the coefficient vectorβ to be random as a way of expressing prior uncertainty about β. As in chapter 2, letting $\mathfrak{J}$ be the set of invariant events, we presume that the *random vector* β is measurable with respect to $\mathfrak{J}$ which means that it is revealed by events in $\mathfrak{J}$. The first-order condition for minimizing the expected value of U_{t+1}^2 requires that the regression residual U_{t+1} be orthogonal to X_t:

$$E\left(X_t U_{t+1} | \mathfrak{J}\right) = 0.$$

Then

$$E\left(X_t Y_{t+1} | \mathfrak{J}\right) = E\left[X_t (X_t)' | \mathfrak{J}\right] \beta, \tag{3.12}$$

which uniquely pins down the regression coefficient β provided that the matrix $E\left[X_t (X_t)' | \mathfrak{J}\right]$ is nonsingular with probability one. Notice that

$$\frac{1}{N} \sum_{t=1}^{N} X_t Y_{t+1} \rightarrow E\left(X_t Y_{t+1} | \mathfrak{J}\right)$$
$$\frac{1}{N} \sum_{t=1}^{N} X_t (X_t)' \rightarrow E\left(X_t (X_t)' | \mathfrak{J}\right),$$

where convergence is with probability one. Thus, from equation (3.12) it follows that a consistent estimator of β is b_N that satisfies

$$\frac{1}{N} \sum_{t=1}^{N} X_t Y_{t+1} = \frac{1}{N} \sum_{t=1}^{N} X_t (X_t)' b_N.$$

Solving for b_N gives the familiar least squares formula:

$$b_N = \left[\sum_{t=1}^{N} X_t(X_t)'\right]^{-1} \sum_{t=1}^{N} X_t Y_{t+1}.$$

Note how statements about the consistency of b_N are conditioned on $\mathfrak{J}$.

3.9 Inventing a past again

In section 2.10, we invented an infinite past for a stochastic process. In this subsection, we invent an infinite past for a vector autoregression in a way that is equivalent with drawing an initial condition X_0 at time $t = 0$ from the stationary distribution $\mathcal{N}(0, \Sigma_\infty)$, where Σ_∞ solves the discrete Lyapunov equation (3.11) described in section 3.7, namely, $\Sigma_\infty = A\Sigma_\infty A' + BB'$.

Thus, consider the vector autoregression

$$X_{t+1} = AX_t + BW_{t+1}$$

where A is a stable matrix, $\{W_{t+1}\}_{t=-\infty}^{\infty}$ is now a two-sided infinite sequence of i.i.d. $\mathcal{N}(0, I)$ random vectors, and t is an integer. We can represent X_t by solving the difference equation backwards to get

$$X_t = \sum_{j=0}^{\infty} A^j BW_{t-j}.$$

Then

$$E\left[X_t\,(X_t)'\right] = \sum_{j=0}^{\infty} A^j BB'\,(A_j)' = \Sigma_\infty$$

where the Σ_∞ is unique positive semidefinite solution to $\Sigma_\infty = A\Sigma_\infty A' + BB'$.

3.10 Limited Dependence

Recall the conditional expectations operator $\mathbb{T}$ defined for functions f of a Markov process having transition probability P and stationary distribution Q:

$$\mathbb{T}f(x) = E\left[f(X_{t+1})|X_t = x\right] = \int_{\{x^* \in \mathcal{X}\}} f(x^*)P(dx^*|x)$$

for which $f(X_t)$ has a finite second moment under Q. In section 3.3, we denoted this space of functions by $\mathcal{L}^2$.

An important subspace of the space $\mathcal{L}^2$ is

$$\mathcal{N} = \left\{ f \in \mathcal{L}^2 : \int f(x) Q(dx) = 0 \right\}.$$

Functions in $\mathcal{N}$ have mean zero under the stationary distribution Q. We use the norm $\|f\| = \left[\int f(x)^2 Q(dx)\right]^{1/2}$ on $\mathcal{L}^2$ and hence on $\mathcal{N}$ too.

Definition 3.10.1. *The conditional expectation operator* $\mathbb{T}$ *is said to be a* strong contraction *on* $\mathcal{N}$ *if there exists* $0 < \rho < 1$ *such that*

$$\|\mathbb{T} f\| \leq \rho \|f\|$$

for all $f \in \mathcal{N}$.

When $\mathbb{T}^m$ is a strong contraction for some positive integer m and some $\rho \in (0,1)$, the underlying process X process is said to be ρ-mixing.

Remark 3.10.2. $\mathbb{T}$ *being a strong contraction on* $\mathcal{N}$ *limits the intertemporal dependence of the Markov process* $\{X_t\}$.

Let $\mathbb{I}$ be the identity operator. When the conditional expectation operator $\mathbb{T}$ is a strong contraction, the operator $(\mathbb{I} - \mathbb{T})^{-1}$ is well defined, bounded on $\mathcal{N}$, and equal to the geometric sum:[10]

$$(\mathbb{I} - \mathbb{T})^{-1} f(x) = \sum_{j=0}^{\infty} \mathbb{T}^j f(x) = \sum_{j=0}^{\infty} E\left[f(X_{t+j}) | X_t = x \right].$$

In section 4.2, we use this geometric sum to extract martingale components of additive functionals.

Example 3.10.3. *Consider the Markov chain setting of section 3.6 with a transition matrix* $\mathbb{P}$. *A stationary density* $\boldsymbol{q}$ *is a nonnegative vector that solves*

$$\boldsymbol{q}' \mathbb{P} = \boldsymbol{q}'$$

[10] The geometric series after the first equality sign is well defined under the weaker restriction that $\mathbb{T}^m$ is a strong contraction for some integer $m \geq 1$.

and satisfies $\boldsymbol{q} \cdot \boldsymbol{1}_n = 1\boldsymbol{q}$. *If the only column eigenvector of* $\mathbb{T}$ *associated with a unit eigenvalue is constant over states* i *for which* $\mathbf{q}_i > 0$, *then the process is ergodic. If in addition the only eigenvector of* $\mathbb{P}$ *that is associated with an eigenvalue that has a unit norm (here the eigenvalue is possibly complex) is constant over states* i *for which* $\mathbf{q}_i > 0$, *then* $\mathbb{T}^m$ *is a strong contraction for some integer* $m \geq 1$.[11] *This implies that the process is ergodic. It also rules out periodic components that can be forecast perfectly.*

[11] This follows from Gelfand's Theorem, which asserts the following. Let $\mathcal{N}$ be the $n-1$ dimensional space of vectors that are orthogonal to $\mathbf{q}$. $\mathbb{T}$ maps $\mathcal{N}$ into itself. The spectral radius of $\mathbb{T}$ restricted to $\mathcal{N}$ is the maximum of the absolute values of the eigenvalues. Gelfand's Theorem asserts that the spectral radius governs the asymptotic decay factor of the $\mathbb{T}$ transformation applied m times as m gets large. Provided that the spectral radius is less than one, the strong contraction property prevails for any $\rho < 1$ that is larger than the spectral radius.