Long-run Uncertainty and Value

Lars Peter Hansen  
University of Chicago and NBER

Thomas J. Sargent  
New York University and NBER

Jose Scheinkman  
Princeton University and NBER

July 10, 2009
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contents</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>ix</td>
</tr>
<tr>
<td>Preface</td>
<td>xi</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>xiii</td>
</tr>
<tr>
<td><strong>1 Stochastic Processes</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Constructing a Stochastic Process: I</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Stationary Stochastic Processes</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Invariant Events and the Law of Large Numbers</td>
<td>4</td>
</tr>
<tr>
<td>1.4 Constructing a Stochastic Process: II</td>
<td>7</td>
</tr>
<tr>
<td>1.5 Stationarity reconsidered</td>
<td>9</td>
</tr>
<tr>
<td>1.6 Limiting Behavior</td>
<td>10</td>
</tr>
<tr>
<td>1.7 Finite-state Markov chain</td>
<td>14</td>
</tr>
<tr>
<td>1.8 Vector Autoregression</td>
<td>16</td>
</tr>
<tr>
<td><strong>2 Additive Functionals</strong></td>
<td>19</td>
</tr>
<tr>
<td>2.1 Construction</td>
<td>19</td>
</tr>
<tr>
<td>2.2 Limited dependence</td>
<td>20</td>
</tr>
<tr>
<td>2.3 Martingale Extraction</td>
<td>22</td>
</tr>
<tr>
<td>2.4 Cointegration</td>
<td>24</td>
</tr>
<tr>
<td>2.5 Identifying Shocks with Long-Run Consequences</td>
<td>25</td>
</tr>
<tr>
<td>2.6 Central Limit Theory</td>
<td>30</td>
</tr>
<tr>
<td>2.7 Growth-rate Regimes</td>
<td>33</td>
</tr>
<tr>
<td>2.8 A Quadratic Model of Growth</td>
<td>34</td>
</tr>
<tr>
<td><strong>3 Multiplicative Functionals</strong></td>
<td>37</td>
</tr>
</tbody>
</table>
List of Figures

1.1 This figure depicts the state evolution as a function of sample point $\omega$ induced by the transformation $S$. The oval shaped region is the collection $\Omega$ of all sample points. .......................... 2

2.1 The top panel plots the logarithm of consumption (smooth blue series) and logarithm of corporate earnings (choppy red series). The bottom panel plots the difference in the logarithms of consumption and corporate earnings. .............................. 26

2.2 This figure plots the response of the logarithm of consumption to a permanent shock (solid) and to a temporary shock (dashed line). The permanent shock is identified as the increment to the common martingale component of the logarithm of consumption and the logarithm of corporate earnings. The figure comes from Hansen et al. (2008). ........................................ 28

2.3 This figure plots posterior histograms for the magnitudes of the short-term and long-term responses of the logarithm of consumption to the shocks. The magnitude is measured as the absolute value across the contributions from the two shocks. The upper panel depicts the histogram for the immediate response and the lower panel depicts the histogram for the long-term limiting response. The figure comes from Hansen et al. (2008). .......................... 29

2.4 This figure plots the impulse responses for the logarithm of output obtained by estimating a three-variate vector autoregression. The shocks are identified as in Fisher (2006). The top panel gives the response to an investment-specific technology shock, the middle panel gives the response to neutral technology shock and the bottom panel gives the response to a transient shock that is uncorrelated with the other two shocks. .......................... 31
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5</td>
<td>This figure plots the impulse responses for the logarithm of hours obtained by estimating a three-variate vector autoregression. The shocks are identified as in Fisher (2006). The top panel gives the response to an investment-specific technology shock, the middle panel gives the response to neutral technology shock and the bottom panel gives the response to a transient shock that is uncorrelated with the other two shocks.</td>
<td>32</td>
</tr>
<tr>
<td>3.1</td>
<td>This figure illustrates a particular additive functional and its multiplicative counterpart.</td>
<td>38</td>
</tr>
<tr>
<td>3.2</td>
<td>The top panel plots a simulation of the multiplicative process studied in example 3.2.2, while the bottom panel plots a simulation of the Cecchetti et al. (2000) process for annual aggregate U.S. mentioned in example 3.2.3.</td>
<td>43</td>
</tr>
<tr>
<td>4.1</td>
<td>The logarithmic function is the concave function in this plot. This function is zero when evaluated at unity. By forming averages using the two endpoints of the straight line below the logarithmic function, we are lead to some point on the line segment depending upon the weights used in the averaging. Jensen’s inequality in this case is just a statement that the line segment must lie below the logarithmic function.</td>
<td>52</td>
</tr>
<tr>
<td>4.2</td>
<td>The vertical axis is implied asymptotic bound on the decay rate for the mistake probabilities for distinguishing between two models. The bound is scaled by 100 and thus expressed as a percent. This bound is reported for alternative choices of $\alpha$ as depicted on the horizontal axis. The difference between the two curves reflects the choice of models in the numerator and denominator of the likelihood ratio. The best bound is given by the maximum of the two curves. The maximizing value for each curve is the same and given a maximizing choice $\alpha^<em>$ of one curve, the other curve is maximized at $1 - \alpha^</em>$.</td>
<td>57</td>
</tr>
</tbody>
</table>
Preface

This manuscript started off as the Toulouse Lectures given by Lars Peter Hansen. Our aim is to explore connections among topics that relate probability theory to the analysis of dynamic stochastic economic systems. Martingale methods have been a productive way to identify shocks with long-term consequences to economic growth and to characterize long-run dependence among macroeconomic time series. Typically they are applied by taking logarithms of time series such as output or consumption in order that growth can be modeled as accumulating linearly over time, albeit in a stochastic fashion. Martingale methods applied in this context have a long history in applied probability and applied time series analysis. We review these methods in the first part of this monograph. In the study of valuation, an alternative martingale approach provides a notion of long-term approximation. This approach borrows insights from large deviation theory, initiated in part to study the behavior of likelihood ratios of alternative time series models. We show how such methods provide characterizations of long-term model components and long-term contributions to valuation. Large deviation theory and the limiting behavior of likelihood ratios has also been central to some formulations of robust decision making. We develop this connection and build links to recursive utility theory in which investors care about the intertemporal composition of risk. Our interest in “robustness” and likelihood ratios is motivated by our conjecture that the modeling of the stochastic components to long-term growth is challenging for both econometricians and the investors inside the models that econometricians build.

More technical developments of some of these themes are given in Hansen and Scheinkman (1995), Anderson et al. (2003), Hansen and Scheinkman (2009) and Hansen (2008).
Acknowledgements

The author wishes to thank Jarda Borovicka, Adam Clemens, Serhiy Kokak, Junghoon Lee and Kenji Wada for their comments and assistance. These notes give some of the underlying support material for the Toulouse Lectures with the listed title.
Chapter 1

Stochastic Processes

In this chapter, we describe two ways of constructing stochastic processes. The first is one that is especially convenient for stating and proving limit theorems. The second is more superficial in the sense that it directly specifies objects that are outcomes in the first construction. However, the second construction is the one that is most widely used in modeling economic time series. We shall use these constructions to characterize limiting behavior both for stationary environments and for environments with stochastic growth.

1.1 Constructing a Stochastic Process: I

We begin with a method of constructing a stochastic process that is convenient for characterizing the limit of points of time series averages.¹ This construction works with a deterministic transformation $S$ that maps a state of the world $\omega \in \Omega$ today into a state of the world $S(\omega) \in \Omega$ tomorrow. The state of the world itself is not observed. Instead, a vector $X(\omega)$ that contains incomplete information about $\omega$ is observed. We assign probabilities over states of the world $\omega$, then use the functions $S$ and $X$ to deduce a joint probability distribution for a sequence of $X$’s. In more detail:

- The probability space is a triple $(\Omega, \mathfrak{F}, Pr)$, where $\Omega$ is a set of sample points, $\mathfrak{F}$ is an event collection (sigma algebra), and $Pr$ assigns

¹A good reference for the material in this section and the two that follow it is Breiman (1968).
CHAPTER 1. STOCHASTIC PROCESSES

Figure 1.1: This figure depicts the state evolution as a function of sample point \( \omega \) induced by the transformation \( S \). The oval shaped region is the collection \( \Omega \) of all sample points.

probabilities to events.

- The (measurable) transformation \( S: \Omega \to \Omega \) used to model the evolution over time has the property that for any event \( \Lambda \in \mathcal{F} \),

\[
S^{-1}(\Lambda) = \{ \omega \in \Omega : S(\omega) \in \Lambda \}
\]

is an event. Notice that \( S(\omega) \) is a deterministic sequence of states of the world in \( \Omega \).

- The vector-valued function \( X: \Omega \to \mathbb{R}^n \) used to model observations is Borel measurable. That is for any Borel set \( b \) in \( \mathbb{R}^n \),

\[
\Lambda = \{ \omega \in \Omega : X(\omega) \in b \} \in \mathcal{F}.
\]

In other words, \( X \) is a random vector.
1.2. STATIONARY STOCHASTIC PROCESSES

- The stochastic process \( \{X_t : t = 1, 2, \ldots\} \) used to model a sequence of observations is constructed via the formula:

\[
X_t(\omega) = X[S_t(\omega)]
\]

or

\[
X_t = X \circ S^t.
\]

The stochastic process \( \{X_t : t = 1, 2, \ldots\} \) is a sequence of \( n \)-dimensional random vectors, and the probability measure \( \Pr \) allows us to make probabilistic statements about the joint distribution of successive components of this sequence. It will sometimes be convenient to extend this construction to date zero by letting \( X_0 = X \).

While this construction of a stochastic process may at first sight appear special, it is not, as the following example illustrates.

**Example 1.1.1.** Let \( \Omega \) be a collection of infinite sequences of elements of \( \mathbb{R}^n \). Specifically, \( \omega = (r_0, r_1, \ldots) \), \( S(\omega) = (r_1, r_2, \ldots) \) and \( x(\omega) = r_0 \). Then \( X_t(\omega) = r_t \).

1.2 Stationary Stochastic Processes

A state of a dynamic system is a complete description of its current position. The current state summarizes all information that can be gleaned from the past that is pertinent to forecasting the future. A stationary or steady state remains invariant as time passes. In a stochastic dynamic system, a stationary state is a probability distribution. In a stochastic steady state, for any finite \( \tau \) the probability distribution of the composite random vector \( [X_{t+1}', X_{t+2}', \ldots, X_{t+\ell}'] \) does not depend on \( t \).

For a given \( \mathcal{S} \), we can restrict the probability measure \( \Pr \) to induce stationarity.

**Definition 1.2.1.** The transformation \( \mathcal{S} \) is **measure-preserving** if

\[
\Pr(\Lambda) = \Pr(\mathcal{S}^{-1}(\Lambda))
\]

for all \( \Lambda \in \mathcal{G} \).

**Proposition 1.2.2.** When \( \mathcal{S} \) is measure-preserving, the distribution function for \( X_t \) is identical for all \( t \geq 0 \).
CHAPTER 1. STOCHASTIC PROCESSES

Given \( X \), form a vector

\[
X^{[\ell]}(\omega) = \begin{bmatrix} X_1(\omega) \\ X_2(\omega) \\ \vdots \\ X_\ell(\omega) \end{bmatrix},
\]

Apply Proposition to \( X^{[\ell]} \) and conclude that the joint distribution function for \( (X_{t+1}, X_{t+2}, \ldots, X_{t+\ell}) \) is independent of \( t \) for \( t = 0, 1, \ldots \). The fact that this property holds for any choice of \( \ell \) is equivalent to a statement that the process \( \{X_t : t = 1, 2, \ldots\} \) is stationary.\(^2\) Thus, the restriction that \( \Pr \) be measure-preserving implies that the stochastic process \( \{X_t : t = 1, 2, \ldots\} \) is stationary.

Example 1.2.3. Suppose that \( \Omega \) contains two states, \( \Omega = \{\omega_1, \omega_2\} \). Consider a transformation \( S \) that maps \( \omega_1 \) into \( \omega_2 \) and \( \omega_2 \) into \( \omega_1 \): \( S(\omega_1) = \omega_2 \) and \( S(\omega_2) = \omega_1 \). Since \( S^{-1}(\omega_2) = \omega_1 \) and \( S^{-1}(\omega_1) = \omega_2 \), for \( S \) to be measure preserving we must have \( \Pr(\omega_1) = \Pr(\omega_2) = .5 \).

Example 1.2.4. Suppose that \( \Omega \) contains two states, \( \Omega = \{\omega_1, \omega_2\} \), and that \( S(\omega_1) = \{\omega_1\} \) and \( S(\omega_2) = \omega_2 \). Since \( S^{-1}(\omega_2) = \omega_2 \) and \( S^{-1}(\omega_1) = \omega_1 \), \( S \) is measure preserving for \( \Pr(\omega_1) \geq 0 \) and \( \Pr(\omega_1) + \Pr(\omega_2) = 1 \).

1.3 Invariant Events and the Law of Large Numbers

In this subsection, we describe a Law of Large Numbers that tells us that time series averages converge when \( S \) is measure-preserving. We use the concept of an invariant event to understand possible limit points and how they are related to a conditional mathematical expectation.

Definition 1.3.1. An event \( \Lambda \) is invariant if \( \Lambda = S^{-1}(\Lambda) \).

Notice that if \( \Lambda \) is an invariant event and \( \omega \in \Lambda \), then \( S^t(\omega) \in \Lambda \) for \( t = 0, 1, \ldots, \infty \).

\(^2\)Some people call this property ‘strict stationarity’ to differentiate it from notions that require only that some moments of the joint distribution be independent of time.
1.3. INVARIANT EVENTS AND THE LAW OF LARGE NUMBERS

Let $\mathcal{I}$ denote the collection of invariant events. Among the invariant events is the entire space, $\Omega$, and the null set, $\emptyset$. Like $\mathcal{F}$, this event collection is a sigma algebra. We are interested in constructing the conditional expectation $E(X|\mathcal{I})$ as a random vector. Consider first the case in which the invariant events are unions of a finite partition $\Lambda_j$ (along with the empty set). A finite partition consists of finite nonempty events $\Lambda_j$ such that $\Lambda_j \cap \Lambda_k \neq \emptyset$ for $j \neq k$ and the union of all $\Lambda_j$ is $\Omega$. We assume that each member of the partition is itself an invariant event. The expectation condition on the event $\Lambda_j$ is given by:

$$\frac{\int_{\Lambda_j} X \, dPr}{Pr(\Lambda_j)}.$$ 

This construction is applicable when $\omega \in \Lambda_j$. We extend this construction to the entire partition by

$$E(X|\mathcal{I})(\omega) = \frac{\int_{\Lambda_j} X \, dPr}{Pr(\Lambda_j)} \text{ if } \omega \in \Lambda_j.$$ 

Thus the conditional expectation $E(X|\mathcal{I})$ is constant within a partition and varies across partitions. This same construction extends directly to countable partitions.

There is an alternative way to think about a conditional expectation does not make reference to a partition but instead uses least squares when $X$ has a finite second moment. Let $Z$ be an $n$-dimensional measurement function such that $Z_t(\omega) = Z[S_t(\omega)]$ is time invariant (does not depend on calendar time). In the special case in which the invariant events are constructed from a finite partition, $Z$ can vary across partitions but remains constant within a partition.\(^3\) Let $\mathcal{Z}$ denote the collection of all such random vectors or measurement functions and solve the following least squares problem:

$$\min_{Z \in \mathcal{Z}} E[|X - Z|^2]$$

where we now assume that $E|X|^2 < \infty$. The solution to the least squares problem is $\hat{Z} = E(X|\mathcal{I})$. An implication of least squares is that

$$E \left[ \left( X - \hat{Z} \right) Z' \right] = 0$$

\[^3\]More generally, $Z$ is measurable with respect to $\mathcal{I}$. 

for $Z$ in $\mathcal{Z}$ so that the vector $X - \bar{Z}$ of regression errors must be orthogonal to any vector $Z$ in $\mathcal{Z}$.

There is a more general measure-theoretic way to construct a conditional expectation. This construction extends the orthogonality property of least squares. Provided that $E|X| < \infty$, $E(X|\mathcal{F})$ is essentially a unique random variable that for any invariant event $\Lambda$ satisfies

$$E \left( [X - E(X|\mathcal{F})] 1_\Lambda \right) = 0$$

where $1_\Lambda$ is the indicator function equal to one on the set $\Lambda$ and zero otherwise.

The following states a key Law of Large Numbers.

**Theorem 1.3.2.** (Birkhoff) Suppose that $\mathcal{S}$ is measure preserving.

i) For any $X$ such that $E|X| < \infty$

$$\frac{1}{N} \sum_{t=1}^{N} X_t \to E(X|\mathcal{F})$$

with probability one;

ii) for any $X$ such that $E|X|^2 < \infty$,

$$E \left[ \left( \frac{1}{N} \sum_{t=1}^{N} X_t - E(X|\mathcal{F}) \right)^2 \right] \to 0.$$

**Definition 1.3.3.** The transformation $\mathcal{S}$ is **ergodic** if all invariant events have probability zero or one.

From a probabilistic standpoint, when $\mathcal{S}$ is ergodic the invariant events are equivalent to the entire sample space $\Omega$, which has probability one, or the empty set $\emptyset$, which has probability measure zero. This notion of ergodicity is a restriction on $\mathcal{S}$ and $Pr$ that implies that conditioning on the invariant events is inconsequential.

**Proposition 1.3.4.** Suppose that $\mathcal{S}$ is ergodic. Then $E(X|\mathcal{F}) = E(X)$. 
Under ergodicity, the limit points of time series averages are the corresponding unconditional expectations. More generally, time series averages can only reveal expectations conditioned on the invariant events.

Consider again Example 1.2.3. Suppose that the measurement vector is

\[ X(\omega) = \begin{cases} 1 & \omega = \omega_1 \\ 0 & \omega = \omega_2 \end{cases} \]

Then it follows directly from the specification of \( S \) that

\[ \frac{1}{N} \sum_{t=1}^{N} X_t(\omega) \to \frac{1}{2} \]

for both values of \( \omega \). The limit point is the average across states. For Example 1.2.4, \( X_t(\omega) = X(\omega) \) and hence the sequence is time invariant and equal to the time series average. The time series average equals the average across states only when one of the two states is assigned probability measure one. Theorem 1.3.2 covers the convergence in general and Proposition 1.3.4 covers the case in which the probability assignment is degenerate. These two examples are included merely for illustration, and we will explore much richer specifications of stochastic processes.

1.4 Constructing a Stochastic Process: II

Instead of specifying \( X[\ell] \) as in section 1.1, we could directly specify a collection of joint distributions \( \hat{P}r_\ell \) for \( \ell \geq 1 \). But we must make sure that \( \hat{P}r_{t+1} \) is consistent with \( \hat{P}r_t \) in the sense that both of these joint distributions assign the same probabilities for the same events, namely, \( \{ X_\ell^* \in b \} \) for (Borel) sets \( b \). If this consistency condition is satisfied, then the famed Kolmogorov Extension Theorem guarantees that there exists a space \((\Omega, \mathcal{F}, Pr)\) and a stochastic process \( \{X_t : t = 1, 2, \ldots\} \) constructed as in section 1.1. Applied model builders typically use a direct specification approach.
A Markov process is an important tool for directly constructing a joint distribution \( \hat{Pr}_{t, \ell} \), \( \ell \geq 0 \). Consider a state space \( \mathcal{X} \) and a transition distribution \( T(dx^*|x) \). The transition distribution \( T \) is a conditional probability measure for each choice \( X_t = x \) in the state space, so it satisfies \( \int T(dx^*|x) = 1 \) for every \( x \) in the state space. There is an associated conditional expectation function. If in addition we specify a marginal \( Q_0 \) distribution for the initial state \( x_0 \) over \( \mathcal{X} \), then we have completely specified all joint distributions for the stochastic process.

The notation \( T(dx^*|x) \) denotes a conditional probability measure where the integration is over \( x^* \) and the 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 \( T(dx^*|x) \) 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 \( \lambda(dx^*) \) used to add up or integrate over states. In such cases we can use a transition density to represent the conditional probability measure. One example is that of first-order vector autoregression. In this case \( T(dx^*|x) \) is a normal distribution with mean \( Ax \) and covariance matrix \( BB' \) for a square matrix \( A \) and a matrix \( B \) with full column rank.\(^4\) In this we may write

\[
X_{t+1} = AX_t + BW_{t+1}
\]

where \( W_{t+1} \) is a multivariate standard normally distributed random vector that is independent of \( X_t \). Another example is that of a discrete-state Markov chain in which \( T(dx^*|x) \) can be represented as a matrix, one row for each realized value of the state \( x \). The row entries give the vector of probabilities conditioned on this realized values. Both of these examples will be developed in more detail later.

An important object for us is a one-step conditional expectation operator that we apply to functions of a Markov state. Let \( f : \mathcal{X} \to \mathbb{R} \). For bounded \( f \), define:

\[
\mathcal{T} f(x) = E[f(X_{t+1})|X_t = x] = \int f(x^*)T(dx^*|x)
\]

\(^4\)When \( BB' \) is singular, a density may not exist with respect to Lebesgue measure. Such singularity occurs when we convert a higher-order vector autoregression into a first-order process.
Iterating on $T$ allows us to form expectations over longer time periods:

$$T^j f(x) = E [f(X_{t+j})|X_t = x].$$

This is a statement of the Law of Iterated Expectations for our Markov setting.

**Remark 1.4.1.** Instead of beginning with a conditional probability distribution, we could start with a conditional expectation operator $T$ mapping a space of functions into itself. Provided that this operator 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; we can construct a conditional probability measure $T(dx^*|x)$ from the operator $T$.

### 1.5 Stationarity reconsidered

We construct Markov processes that are stationary by appropriately choosing distributions of the initial state $x_0$.

**Definition 1.5.1.** A *stationary distribution* for a Markov process is a probability measure $Q$ over the state space $X$ that satisfies

$$\int T(dx^*|x)Q(dx) = Q(dx^*).$$

We will sometimes make reference to a stationary density $q$. A density is always relative to a some measure. With this in mind, let $\lambda$ be a measure on the state space $X$ used to add up or integrate over alternative Markov states. Then a density $q$ is a nonnegative (Borel measurable) function of the state for which $\int q(x)\lambda(dx) = 1$.

**Definition 1.5.2.** A *stationary density* for a Markov process is a probability density $q$ with respect to a measure $\lambda$ over the state space $X$ that satisfies

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

The following example of a reversible Markov process occurs sometimes in simulations that implement Bayesian estimation.
Example 1.5.3. Suppose that
\[ T(dx^*|x)q(x)\lambda(dx) = T(dx|x^*)q(x^*)\lambda(dx^*). \]
Because a transition density satisfies \( \int T(dx|x^*) = 1 \), notice that
\[ \int T(dx^*|x)q(x)\lambda(dx) = \int T(dx|x^*)q(x^*)\lambda(dx^*) = q(x^*)d\lambda(dx^*). \]
Thus, \( q \) is a stationary density.

When the Markov process is initialized according to a stationary distribution, we can construct the process \( \{X_t : t = 1, 2, \ldots\} \) with a measure-preserving transformation \( S \) of the type featured in the first method of constructing a stochastic process that we described in section 1.1.

Given a stationary distribution \( Q \), form the space of functions \( L^2 \) defined as
\[ L^2 = \{ f : X \to \mathbb{R} : \int f(x)^2 Q(dx) < \infty \} \]
It can be shown that \( T : L^2 \to L^2 \). On this space there is a well defined norm given by:
\[ \|f\| = \left[ \int f(x)^2 Q(dx) \right]^{1/2} \]

1.6 Limiting Behavior

When the Markov process is not periodic, we are interested in situations when
\[ \lim_{j \to \infty} T^j f(x) = r \tag{1.1} \]
for some \( r \in \mathbb{R} \) where the convergence is either pointwise in \( x \) or define using the \( L^2 \) norm. This limit restricts the long-term forecasts eventually not to depend on the current Markov state. (See Meyn and Tweedie (1993) for a comprehensive treatment of this convergence.) Let \( Q \) be a stationary distribution. Then it is necessarily true that
\[ \int T^j f(x)Q(dx) = \int f(x)Q(dx) \]
for all \( j \). Thus
\[
\mathbf{r} = \int f(x)Q(dx).
\]

Thus the limiting forecast is necessary the expectation under a stationary distribution. Notice that here we have not assumed that the stationary density is unique, although we did presume that the limit point is a number and not a random variable.

One reason we are interested in limit (1.1) is that when a stationary distribution exists, this limit implies the convergence of:
\[
\lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} \left\| \mathbb{T}^j f(x) - \int f(x)Q(dx) \right\| = 0.
\]

This suggests that limit point for a time series version of the Law of Large Numbers is \( \int f(x)Q(dx) \). Also, if \( \int f(x)Q(dx) = 0 \) and the convergence is sufficiently fast, then
\[
\lim_{N \to \infty} \sum_{j=1}^{N} \mathbb{T}^j f(x)
\]
is a well-defined function of the Markov state. Under stationarity we can represent the limit in first case as the \( \int f(x)Q(dx) \), and a necessary condition for the second limit is that \( \int f(x)Q(dx) = 0 \).

**Ergodicity**

In section 1.3 we defined the concept of ergodicity for stationary stochastic processes. We now explore this concept when the process of interest is Markovian.

**Definition 1.6.1.** A function \( \tilde{f} \in \mathcal{L}^2 \) that solves the equation \( \mathbb{T}f = f \) is called an eigenfunction associated with a unit eigenvalue.

An eigenfunction of \( \mathbb{T} \) is a generalization of an eigenvector of a matrix. Notice that if \( \tilde{f} \) is used in calculation (1.1), the \( \tilde{f} \) is necessarily constant.

**Proposition 1.6.2.** Suppose that \( \tilde{f} \) is an eigenfunction of \( \mathbb{T} \) associated with a unit eigenvalue. Then \( \{ \tilde{f}(X_t) : t = 1, 2, \ldots \} \) is constant over time with probability one.
Proof.

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

Then because \( Q \) is a stationary distribution,

\[
E \left( [\hat{f}(X_{t+1}) - \hat{f}(X_{t})]^2 \right) = E \left[ \hat{f}(X_{t+1})^2 \right] + E \left[ \hat{f}(X_{t})^2 \right] - 2E \left[ \hat{f}(X_{t+1})\hat{f}(X_{t}) \right] = 0.
\]

\( \square \)

Obviously, time series averages of an such an eigenfunction \( T\hat{f} = \hat{f} \) do not move either, so

\[
\frac{1}{N} \sum_{t=1}^{N} \hat{f}(X_{t}) = \hat{f}(X).
\]

However, the time series average \( \frac{1}{N} \sum_{t=1}^{N} \hat{f}(X_{t}) \) differs from \( \int \hat{f}(x)Q(dx) \) when \( \hat{f}(x) \) is not constant across states \( x \) that occur with probability one under \( Q \). This happens when the variation of \( \hat{f}(X_{t}) \) along a sample path for \( \{X_{t}\} \) conveys an inaccurate impression of its variation across the stationary distribution \( Q(dx) \). See example 1.7.2 below. This possibility leads us to use eigenfunctions to state a sufficient condition for ergodicity.

When \( f \) is an indicator function of a Borel set \( b \) in \( X \) and \( Tf = f \), then

\[ \Lambda = \{ \omega \in \Omega : X \in b \} \]

is an invariant event in \( \Omega \) under the corresponding probability measure \( Pr \) and transformation \( S \). For Markov processes, all invariant events can be represented like this, a result that is not obvious. A reference for it is Doob (1953), Theorem 1.1, page 460.

**Proposition 1.6.3.** When the only solution to the eigenvalue equation

\[ Tf = f \]

is a constant function (with \( Q \) measure one), then it is possible to construct the process \( \{X_{t} : t = 0, 1, \ldots\} \) using a transformation \( S \) that is measure preserving and ergodic.
1.6. LIMITING BEHAVIOR

Notice here that ergodicity is a property that obtains only relative to a stationary distribution for the Markov process. When there are multiple stationary distributions, a constant solution to the eigenvalue problem can be the only one that works for one stationary distribution, but non-constant solutions can exist for other stationary distributions. For instance, consider example 1.2.4. Although any assignment of probabilities constitutes a stationary distribution, we get ergodicity only when we assign probability one to one of the two states. Also see example 1.7.3.

**Sufficient Conditions for Ergodicity**

While finding nondegenerate eigenfunctions associated with a unit eigenvalue often gives a way to establish that a process is not ergodic, it can be difficult to establish ergodicity directly using Proposition 1.6.3. There are convenient sufficient conditions, including the drift conditions discussed in Meyn and Tweedie (1993). We explore one set of sufficient conditions in this subsection.

Let $Q$ be a stationary distribution. Form the *resolvent* operator:

$$Rf(x) = (1 - \delta) \sum_{j=1}^{\infty} \delta^j T^j f$$

associated with some constant discount factor $0 < \delta < 1$. To study periodic components of processes, we introduce sampling at an interval $p$. If we sample a periodic process of period $p$, we want functions of the Markov state to be invariant, so

$$T^p f = f$$

for some function $f$ that is nondegenerate and not constant with probability one. This leads us to consider a sampled counterpart of the resolvent operator $R$:

$$R_p f(x) = (1 - \delta) \sum_{j=1}^{\infty} \delta^j T^{pj} f.$$ 

A set of sufficient conditions for

$$\lim_{j \to \infty} T^j f(x^*) \to \int f(x) dQ(x)$$

(1.2)
for each $x \in X$ and each bounded $f$ is:

**Condition 1.6.4.** Suppose that the stationary Markov process satisfies:

(i) For any $f \geq 0$ such that $\int_{X} f(x)Q(dx) > 0$, $\mathbb{R}_p f(x) > 0$ for all $x \in X$ and all $p \geq 0$. (The Markov process is $(Q)$ irreducible and aperiodic.)

(ii) $T$ maps bounded continuous functions into bounded continuous functions. (The Markov process satisfies the Feller property.)

(iii) The support of $Q$ has a nonempty interior in $X$.

(iv) $TV(x) - V(x) \leq -1$ outside a compact subset of $X$ for some nonnegative function $V$.

Sufficient condition (i) may be hard to check, but it suffices to show that there exists an $m$ such that for any $f \geq 0$ such that $\int_{X} f(x)Q(dx) > 0$

$$T^m f(x) > 0$$

for all $x \in X$. Given this property holds for $T^n$ it must also hold true for $pT^{m}$ for any $p$. Condition (iv) is the drift condition for stability. It is constructive provided that we can establish the inequality for a conveniently chosen function $V$. Heuristically, this drift condition says that outside a compact subset of the state space, the conditional expectation must push inward. The choice of $-1$ as a comparison point is made 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.

### 1.7 Finite-state Markov chain

Suppose that $X$ consists of $n$ states. We may label these state in any of a variety of ways but suppose that state $x_j$ is the coordinate vector of all zeros except in position $j$ where there is a one. Let $T$ be an $n \times n$ 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 $T$ are all nonnegative and this matrix must satisfy:

$$T1_n = 1_n.$$

---

5Restriction 1.2 is stronger than ergodicity. It rules out periodic processes, although we know that periodic processes can be ergodic.
1.7. FINITE-STATE MARKOV CHAIN

where $1_n$ is an $n$-dimensional vector of ones.

Let $q$ be an $n$-dimensional vector of probabilities. Then stationarity requires that

$$q'T = q'$$

That is, $q$ is a row eigenvector of $T$ associated with a unit eigenvalue.

We use a vector $f$ to represent a function from the state space to the real line, where each coordinate of $f$ gives the value of the function at the corresponding coordinate vector. Consider column eigenvectors of $T$ associated with a unit eigenvalue. Suppose that the only solutions to

$$Tf = f$$

are of the form $f \propto 1_n$. Then we can construct a process that is stationary and ergodic by initializing the process with density $q$.

We can weaken this condition to allow nonconstant right eigenvectors. A weaker condition is that the eigenvector and stationary distribution satisfy

$$\min_r \sum_{i=1}^n (f_i - r)^2 q_i = 0$$

Notice that we are multiplying by probabilities, so that if $q_i$ is zero, the contribution of $f_i$ to the least squares objective can be neglected, which allows for non-constant $f$’s, albeit in a limited way.

Three examples illustrate these concepts.

**Example 1.7.1.** We now recast Example 1.2.3 as a Markov chain with transition matrix $T = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. This chain has a unique invariant distribution $q = [0.5, 0.5]'$ and the invariant functions are $[r, r]'$ for any scalar $r$. Therefore, the process initiated from the stationary distribution is ergodic.

**Example 1.7.2.** Next we recast Example 1.2.4 as a Markov chain with transition matrix $T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. 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 $[r_1, r_2]'$ and for any scalars $r_1, r_2$. Therefore, the process is not ergodic when
\( \pi \in (0, 1) \), for note that if \( r_1 \neq r_2 \) the resulting invariant function will fail to be constant across states receive positive probability according to a 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 \( \bar{y} \) of the state.

**Example 1.7.3.** A chain with transition matrix
\[
T = \begin{bmatrix}
0.8 & 0.2 & 0 \\
0.1 & 0.9 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]
has a continuum of stationary distributions \( \pi \left[ \frac{1}{3}, \frac{2}{3}, 0 \right]' + (1 - \pi) \left[ 0, 0, 1 \right]' \) for \( \pi \in [0, 1] \) and invariant functions \( \left[ r_1 \ r_1 \ r_2 \right]' \) and for any scalars \( r_1, r_2 \). With any stationary distribution associated with \( \pi \in (0, 1) \), the chain is not ergodic because some invariant functions are not constant with probability one under such a stationary distribution. But for stationary distributions associated with \( \pi = 1 \) or \( \pi = 0 \), the chain is ergodic.

### 1.8 Vector Autoregression

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 and \( B \) has full column rank. Then
\[
W_{t+1} = (B'B)^{-1}B'(X_{t+1} - AX_t),
\]
so we can recover the shock vector \( W_{t+1} \) from \( X_{t+1} \) and \( X_t \). 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 \( \mu \) of a stationary distribution satisfies
\[
\mu = A\mu.
\]
Because we have assumed that $A$ is a stable matrix, the only $\mu$ that solves $(A - I)\mu = 0$ is $\mu = 0$. Thus, the mean of the stationary distribution is $\mu = 0$.

Let $\Sigma_t$ be the covariance matrix for $X_t$. Then

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

For $\Sigma_t = \Sigma$ to be invariant over time, it must be true that

$$\Sigma = A\Sigma A' + BB'.$$

Because $A$ is a stable matrix, this equation has a unique solution for a positive semidefinite matrix $\Sigma$. This is the covariance matrix of the stationary distribution.

Suppose that $\Sigma_0 = 0$ (a matrix of zeros). Then

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

The limit of this sequence is:

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

which we have seen is the covariance matrix for the stationary distribution. 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 $\mu$ and covariance matrix $\Sigma$.

To verify ergodicity, suppose that the covariance matrix $\Sigma$ of the stationary distribution has full rank. Then restriction (iii) of Condition 1.6.4 is satisfied. Furthermore, $\Sigma_t$ has full rank for some $t$, which guarantees that the process is irreducible and aperiodic (restriction (i)). Let $V(x) = |x|^2$. Then

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

Thus

$$\mathbb{T}V(x) - V(x) = x'(A' A - I)x + \text{trace}(B'B).$$
That $A$ is a stable matrix implies that $A'A - I$ is negative definite, so that drift restriction (iv) of Condition 1.6.4 is satisfied for $|x|$ sufficiently large.\(^6\)

We now show how to extend this example to obtain a nonzero mean for the stationary distribution. Partition the Markov state as:

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

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

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

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

$$X^{[2]}_{t+1} = X^{[2]}_t = \ldots = X^{[2]}_0$$

and hence is invariant. Let $\mu^{[2]}$ denote the mean of $X^{[2]}_t$ for any $t$. In a stationary distribution we require that the mean $\mu^{[1]}$ of the first component of the state vector satisfy:

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

Hence

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

Imitating a previous argument, the covariance matrix, $\Sigma^{[11]}$ for this component satisfies:

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

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

Since $\{X^{[2]}_t : t = 0, 1, \ldots\}$ is invariant, the process $\{X_t : t = 0, 1, \ldots\}$ is only ergodic when the variance $\Sigma^{[22]}$ is zero. Otherwise, the limit points for the Law of Large Numbers (Theorem 1.3.2) should be computed by conditioning on $X^{[2]}_0$.

\(^6\)The Feller property can also be established.
Chapter 2

Additive Functionals

For economic applications, it is too limiting to consider only time series models that are stationary because we are interested in processes that display stochastic counterparts to geometric growth or, equivalently, arithmetic growth in logarithms. We suggest a convenient construction of such a process and produce a convenient decomposition into a time trend, a martingale and a stationary component.

2.1 Construction

Let \( \{X_t\} \) be a stationary Markov process. We now build functionals of this process by accumulating the impact of the Markov process over time.

Definition 2.1.1. A process \( \{Y_t : t = 0, 1, \ldots\} \) is said to be an additive functional if it can be represented as

\[
Y_{t+1} - Y_t = \kappa(X_{t+1}, X_t),
\]

or equivalently

\[
Y_t = \sum_{j=1}^{t} \kappa(X_j, X_{j-1})
\]

where we initalize \( Y_0 = 0 \).

The initialization, \( Y_0 = 0 \) is imposed for convenience, but it does allow us to construct \( Y_t \) as a function of only the underlying Markov process
between date zero and $t$. Adding a nonzero initial condition will have obvious consequences for the results in this chapter.

A linear combination of two additive functionals $\{Y_t^{[1]}\}$ and $\{Y_t^{[2]}\}$ is an additive functional. If $\kappa_1$ is used to construct the first process and $\kappa_2$ the second process, then $\kappa = \kappa_1 + \kappa_2$ can be used to construct the sum of the two processes.

**Example 2.1.2.**

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

where $\{W_{t+1} : t = 1, 2, \ldots\}$ is an iid sequence of multivariate normally distributed random vectors and $B$ has full column rank. Premultiply by $B'$ and obtain:

$$B'X_{t+1} - B'AX_t = B'BW_{t+1}.$$  

Then

$$W_{t+1} = (B'B)^{-1} (B'X_{t+1} - B'AX_t).$$

Form

$$\kappa(X_{t+1}, X_t) = \mu(X_t) + \sigma(X_t)W_{t+1}.$$  

Then $\mu(X_t)$ is the conditional mean of $Y_{t+1} - Y_t$ and $|\sigma(X_t)|^2$ is the conditional variance. When $\sigma$ depends on the Markov state, this is referred to as a stochastic volatility model.

Let $\mathcal{F}_t$ be the information set (sigma algebra) generated by $X_0, X_1, \ldots, X_t$.

**Definition 2.1.3.** A process $\{Y_t : t = 0, 1, \ldots\}$ is an additive martingale provided that $E[\kappa(X_{t+1}, X_t)|X_t] = 0$.

Note that $E[\kappa(X_{t+1}, X_t)|X_t] = 0$ implies the usual martingale restriction

$$E[Y_{t+1}|\mathcal{F}_t] = Y_t, \text{ for } t = 0, 1, \ldots.$$  

The process $\{Y_t : t = 0, 1, \ldots\}$ defined in example (2.1.2) is evidently a martingale if $\mu(X_t) = 0$.

### 2.2 Limited dependence

In what follows we will have cause to limit the dependence of the Markov process. One way to achieve this is to restrict the corresponding conditional
expectation operator. In chapter 1, we introduced the space $L^2$. We now explore an important subspace given by:

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

Thus functions in $Z$ have mean zero under the stationary distribution. As in chapter 1 we use the norm $\|f\| = \left[ \int f(x)^2Q(dx) \right]^{1/2}$ on $L^2$ and hence on $Z$.

**Definition 2.2.1.** The conditional expectation operator $T$ is a strong contraction (on $Z$) if there exists a $0 < \rho < 1$ such that

$$\|Tf\| \leq \rho\|f\|$$

for all $f \in Z$.

When the conditional expectation operator is a strong contraction, the operator $(I - T)^{-1}$ is well defined (and bounded) and given by the geometric sum:

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

The geometric series is well defined under the weaker restriction that $T^m$ is a strong contraction for some $m$. We will use this series to extract martingale components of additive functionals.

**Example 2.2.2.** Consider the Markov chain setting of subsection 1.7. The conditional expectation can be represented using a transition matrix $T$. We have seen that to obtain a stationary density, we should solve

$$q^T = q$$

for a nonnegative vector $q$ such that $q \cdot 1_n = 1$. If the only column eigenvector of $T$ associated with a unit eigenvalue is a constant over states $i$ for which $q_i > 0$, then the process is ergodic. If in addition, the only eigenvector of $T$ with unit norm (this includes complex eigenvalues), is constant over states $i$ for which $q_i > 0$, then $T^m$ will be a strong contraction for some $m$.

When this property is satisfied the underlying process is said to be $\rho$-mixing.
In addition to imposing ergodicity, this rules out periodic components that can be forecasted perfectly.

2.3 Martingale Extraction

Additive processes have additive martingales embedded within them that capture all long-run variation. In this section, we show how to extract these martingales and suggest ways they can be used.

For a Markov process \( \{X_t : t = 0, 1, \ldots\} \), consider the following special algorithm that applies to a special type of an additive process for which \( \kappa(x^*, x) = f(x) \) with \( \int f(x)Q(dx) = 0 \).

**Algorithm 2.3.1.** Suppose that \( f \in \mathcal{Z} \) and

\[
Y_{t+1} - Y_t = f(X_t).
\]

Thus \( \kappa(x^*, x) = f(x) \). Solve the equation \( g(x) = f(x) + Tg(x) \) for \( g \). The solution is:

\[
g(x) = (I - T)^{-1} f(x) = \sum_{j=0}^{\infty} T^j f(x) = \sum_{j=0}^{\infty} E[f(X_{t+j})|X_t = x], \tag{2.2}
\]

where \( I \) is the identity operator, a legitimate calculation provided that the infinite sum on the right-hand side of \( (2.2) \) is finite. The function \( g \) gives the best forecast today of the long-term limit of the additive functional \( \{Y_t : t = 0, 1, \ldots\} \) with the argument being the current Markov state. A sufficient condition for the sum on the right-hand side of \( (2.2) \) to be finite is that \( T^m \) be a strong contraction for some \( m \). Evidently, \( (I - T)g(x) = f(x) \).

Let

\[
\kappa_1(x^*, x) = g(x^*) - g(x) + f(x)
\]

and note that \( (I - T)g(x) = f(x) \) implies that

\[
\kappa_1(x^*, x) = g(x^*) - Tg(x).
\]

---

2This result follows from Gelfand's Theorem. Let \( \mathcal{Z} \) be the \( n - 1 \) dimensional space of vectors that are orthogonal to \( q \). Then \( T \) maps \( \mathcal{Z} \) into itself. The spectral radius of this transformation is the maximum of the absolute values of the eigenvalues of the induced transformation. Gelfand's formula shows that the spectral radius governs the asymptotic decay of the transformation applied \( m \) times as \( m \) gets large implying the strong contraction property for any \( \rho \) larger than the spectral radius.
2.3. MARTINGALE EXTRACTION

Thus, $\kappa_1(X_{t+1}, X_t)$ is the forecast error in forecasting $g(X_{t+1})$ given $X_t$, and in particular

$$E[\kappa_1(X_{t+1}, X_t)|X_t] = 0.$$ 

Therefore,

$$Y_{t1} = \sum_{j=1}^{t} f(X_{j-1})$$

$$= \sum_{j=1}^{t} \kappa_1(X_j, X_{j-1}) - g(X_t) + g(X_0)$$

is a martingale.

This algorithm is a martingale counterpart to a more general construction of Gordin (1969) for stationary processes. We now use this algorithm as a component of a more general martingale extraction algorithm.

**Algorithm 2.3.2.** Let $\{X_t : t = 0, 1, \ldots\}$ be a stationary, ergodic Markov process. Let $Y_t, t = 0, 1, \ldots$ be an additive process. Perform the following steps.

1. Compute the conditional expectation of the growth rate $E[\kappa(X_{t+1}, X_t)|X_t = x] = \bar{f}(x)$ and form the deviation from the conditional mean

$$\kappa_2(X_{t+1}, X_t) = \kappa(X_{t+1}, X_t) - \bar{f}(X_t).$$

Note that $E[\kappa_2(X_{t+1}, X_t)|X_t = x] = 0$.

2. Compute the deviation of the conditional mean from the unconditional mean of the growth rate $\nu = \int \bar{f}(x)q(x)d\lambda(x)$, namely, $f(x) = \bar{f}(x) - \nu$ and apply algorithm 2.3.1 to form $g$ and $\kappa_1$ in the decomposition $f(x) = \kappa_1(x^*, x) - g(x^*) - g(x)$.

3. Note that

$$\kappa(x^*, x) = \kappa_2(x^*, x) + \bar{f}(x)$$

$$= \kappa_2(x^*, x) + f(x) + \nu$$

$$= \kappa_2(x^*, x) + \kappa_1(x^*, x) - g(x^*) + g(x) + \nu.$$ 

\(^3\)See also Hall and Heyde (1980).
(iv) It follows that

\[ Y_t = t \nu + \left[ \sum_{j=1}^{t} \kappa_a(X_j, X_{j-1}) \right] - g(X_t) + g(X_0) \]

where \( \kappa_a(X^*, X) = \kappa_1(X^*, X) + \kappa_2(X^*, X) \) and \( E[\kappa_a(X_{j+1}, X_j)|X_j] = 0 \).

Thus, we have established that

**Proposition 2.3.3.** Suppose that \( \{Y_t : t = 0, 1, \ldots\} \) is an additive functional, \( \mathbb{T}^n \) is a strong contraction on \( \mathbb{Z} \) for some \( n \) and \( E[\kappa(X_{t+1}, X_t)^2] < \infty \). Then

\[
Y_t = t \nu + \sum_{j=1}^{t} \kappa_a(X_j, X_{j-1}) + g(X_0) - g(X_t).
\]

The three terms in the decomposition are each additive processes initialized at zero. The first is a linear time trend, the second is an additive martingale, and the third is a stationary process translated by a term that is time invariant.

The remainder of this section describes some applications.

### 2.4 Cointegration

Linear combinations of two additive processes are additive. Specifically, form

\[ Y_t = r_1 Y_t^{[1]} + r_2 Y_t^{[2]} \]

where \( Y_t^{[1]} \) is built with \( \kappa_1 \) and \( Y_t^{[2]} \) is built with \( \kappa_2 \). Thus, we can build

\[
Y_t = r_1 Y_t^{[1]} + r_2 Y_t^{[2]} = \sum_{j=1}^{t} [r_1 \kappa_1(X_j, X_{j-1}) + r_2 \kappa_2(X_j, X_{j-1})]
\]

The martingale of Proposition 4.2 for \( \{Y_t : t = 0, 1, \ldots\} \) is the corresponding linear combination of the martingales for the two components.
2.5. IDENTIFYING SHOCKS WITH LONG-RUN CONSEQUENCES

Engle and Granger (1987) call two processes \textit{cointegrated} if there exists a linear combination of these processes that is stationary. This occurs when

\[
\begin{align*}
\mathbf{r}_1\nu_1 + \mathbf{r}_2\nu_2 &= 0 \\
\mathbf{r}_1\kappa_a + \mathbf{r}_2\kappa_{a2} &= 0
\end{align*}
\]

where the \(\nu\)'s and \(\kappa\)'s correspond to the first two components of the representation in Proposition 4.2. These two zero restrictions imply that the time trend and martingale component for the linear combination \(Y_t\) are both zero.\(^4\) It is of particular interest when \(\mathbf{r}_1 = 1\) and \(\mathbf{r}_2 = -1\). In this case the two component additive processes \(Y_t^{[1]}\) and \(Y_t^{[2]}\) share a common growth component.

Figure 2.1 plots the logarithms of two macro time series, nondurable consumption and corporate earnings. Both grow over time and arguably these series should grow together and hence have common trend and martingale components. The second panel plots the difference logarithms of the macro time series.

2.5 Identifying Shocks with Long-Run Consequences

Suppose that the Markov state \(\{X_t: t = 0, 1, \ldots\}\) follows the first-order VAR

\[
X_{t+1} = AX_t + BW_{t+1}
\]

where \(A\) has stable eigenvalues. Let

\[
Y_{t+1} - Y_t = \kappa(X_{t+1}, X_t) = D \cdot X_t + \nu + F \cdot W_{t+1}.
\]

where \(D\) and \(F\) are vectors with the same dimensions as \(X_t\) and \(W_{t+1}\), respectively.

We use this model to illustrate the four-step construction in algorithm 2.3.2.

(i) Form the conditional growth rate

\[
\bar{f}(x) = D \cdot x + \nu
\]

\(^4\)The vector \((\mathbf{r}_1, \mathbf{r}_2)\) is the cointegration vector and is only determined up to scale.
Additive Macroeconomic Processes

Figure 2.1: The top panel plots the logarithm of consumption (smooth blue series) and logarithm of corporate earnings (choppy red series). The bottom panel plots the difference in the logarithms of consumption and corporate earnings.
2.5. IDENTIFYING SHOCKS WITH LONG-RUN CONSEQUENCES

and the deviation
\[ \kappa_2(X_{t+1}, X_t) = F \cdot W_{t+1}. \]

(ii) Remove the unconditional mean:
\[ f(x) = D \cdot X_t + \nu - \nu = D \cdot X_t. \]

Here we are using that the unconditional mean of \( X \) is 0 because \( A \) is a stable matrix.

(iii) Where \( g(x) = (I - T)^{-1} f(x) = D'(I - A)^{-1}x \), form
\[ \kappa_1(x^*, x) = f(x) + g(x^*) - g(x) = D \cdot x + D'(I - A)^{-1}(Ax + Bw^*) - D'(I - A)^{-1}x \]
\[ = [B'(I - A')^{-1}D] \cdot w^*. \]

(iv) It follows that \( \kappa_a = \kappa_1 + \kappa_2 \) is
\[ \kappa_a(X_{t+1}, X_t) = [F + B'(I - A')^{-1}D] \cdot W_{t+1}. \quad (2.3) \]

Blanchard and Quah (1989) use formula (2.3) in conjunction with a version of the decomposition in proposition 4.2 to identify a supply shock or a technology shock. In their application, the growth rate of output is one of the components of \( X_t \), and it is assumed that only supply shocks or technology shocks have long-run consequences. Then \( F + B'(I - A')^{-1}D \) identifies the linear combination of \( W_{t+1} \) that is the technology shock. This idea has been extended to include more than one shock with long-run consequences by Shapiro and Watson (1988) and Fisher (2006).

Similarly, for Beveridge and Nelson (1981), \( [F + B(I - A')^{-1}D] \cdot W_{t+1} \) is the permanent shock in a permanent-transitory decomposition of a univariate time series model. When \( \{W_{t+1} : t = 0, 1, \ldots\} \) is a univariate process, permanent and transitory shocks are necessarily (perfectly) correlated, but in a multivariate setting, transitory shocks can be restricted to be uncorrelated with permanent shocks.

Figure 2.2 shows how the logarithm of aggregate consumption responds to a shock identified as the common martingale component to the logarithm of consumption and the logarithm of corporate earnings. The immediate response is less than half of the long-term response. The long-term response is also a measure of the magnitude of the martingale component to the
logarithms of consumption and earnings. This figure illustrates a long-term consumption risk of the type featured by Bansal and Yaron (2004). The short-term impact to shocks is measured with much more accuracy as is evident in Figure 2.3. We take this evidence to be that there could be a long-term risk component to consumption, but that it is poorly measured.

Fisher (2006) used a production-based model to identify two shocks with permanent consequences. One of the shocks is investment specific and the other is a neutral shock. The investment specific shock is identified as the martingale increment to the relative price of investment goods and the neutral shock is the component of the martingale increment to output that is uncorrelated with the investment shock. Aggregate hours are included in the vector autoregression implying that there are three innovations used to generate new information each time period. Thus the matrix $B$ has rank three. A third shock is constructed to be uncorrelated with the two

![Impulse Responses of Consumption](image)

Figure 2.2: This figure plots the response of the logarithm of consumption to a permanent shock (solid) and to a temporary shock (dashed line). The permanent shock is identified as the increment to the common martingale component of the logarithm of consumption and the logarithm of corporate earnings. The figure comes from Hansen et al. (2008).
Figure 2.3: This figure plots posterior histograms for the magnitudes of the short-term and long-term responses of the logarithm of consumption to the shocks. The magnitude is measured as the absolute value across the contributions from the two shocks. The upper panel depicts the histogram for the immediate response and the lower panel depicts the histogram for the long-term limiting response. The figure comes from Hansen et al. (2008).
technology shocks.

The results are reported in Figures 2.4 and 2.5. As featured by Fisher (2006), the investment specific technology shock is an important contributor to the variation in hours and output. It takes many time periods for the contribution to output to be pronounced. The long-term impact is negative because the shock is normalized to have a positive impact on the relative price of output. While the neutral shock contributes to output variability in an important way, this same shock accounts for only a modest amount of the variation in hours relative to the other shocks.

2.6 Central Limit Theory

Consider an additive martingale process \( \{Y_t : t = 0, 1, \ldots\} \) whose increments \( Y_{t+1} - Y_t \) are stationary, ergodic,\(^5\) martingale differences:

\[
E (Y_{t+1} - Y_t | \mathcal{F}_t) = 0.
\]

Billingsley (1961) shows that this process obeys a central limit theorem:

\[
\frac{1}{\sqrt{t}} Y_t \implies N(0, E[(Y_{t+1} - Y_t)^2])
\]

where \( \implies \) means convergence in distribution. This central limit theorem looks standard except that the terms in the sum (the increments in the additive process) are martingale differences rather than iid.

Gordin (1969) extends this result to allow for temporally dependent increments. We can regard Gordin’s result as an application of Proposition 4.2. Under the assumptions of this proposition:

\[
\frac{1}{\sqrt{t}} Y_t \implies N(0, \sigma^2)
\]

provided that \( \eta = 0 \). The variance used in the central limit approximation is

\[
\sigma^2 = \lim_{t \to \infty} \frac{1}{t} \text{variance}(Y_t) = E ( [\kappa_a(X_j, X_{j-1})]^2 )
\]

a long-term concept that takes account of the temporal dependence of the increments.

\(^5\)Ergodicity can be dispensed with if we replace the variance by \( E[(Y_1 - Y_0)^2] \).
Output Responses to Three Shocks

Figure 2.4: This figure plots the impulse responses for the logarithm of output obtained by estimating a three-variate vector autoregression. The shocks are identified as in Fisher (2006). The top panel gives the response to an investment-specific technology shock, the middle panel gives the response to neutral technology shock and the bottom panel gives the response to a transient shock that is uncorrelated with the other two shocks.
Figure 2.5: This figure plots the impulse responses for the logarithm of hours obtained by estimating a three-variate vector autoregression. The shocks are identified as in Fisher (2006). The top panel gives the response to an investment-specific technology shock, the middle panel gives the response to neutral technology shock and the bottom panel gives the response to a transient shock that is uncorrelated with the other two shocks.
Corollary 2.6.1. (Gordin (1969)) Under the assumptions of Proposition 4.2,

\[ \frac{1}{\sqrt{t}} Y_t \Longrightarrow N(0, \sigma^2) \]

where \( \sigma^2 = E \left( \kappa_a(X_j, X_{j-1}) \right)^2 \).

By way of illustration, we return to the first-order VAR example with \( \nu = 0 \):

\[
\begin{align*}
X_{t+1} &= A X_t + B W_{t+1} \\
Y_{t+1} - Y_t &= D \cdot X_t + F \cdot W_{t+1}
\end{align*}
\]

The variance that appears in this Central Limit Theorem is that of the martingale increment:

\[ \sigma^2 = [F + B'(I - A')^{-1}D] \cdot [F + B'(I - A')^{-1}D]' \]

This differs from both the conditional variance \(|F|^2\) of \( Y_{t+1} \) and the unconditional variance, \( D'\Sigma D + |F|^2 \), of \( Y_{t+1} - Y_t \) where \( \Sigma \) is the covariance matrix of \( X_t \) in the implied stationary distribution

\[
\Sigma = \sum_{j=0}^{\infty} (A)^j BB'(A^j)'.
\]

Since linear combinations of additive functionals are additive, Corollary 2.6.1 has a direct extension for multivariate counterparts to additive processes. The corollary can be applied to any linear combination of a vector of additive processes.

2.7 Growth-rate Regimes

Suppose that \( \{Z_t\} \) evolves according to an n-state Markov chain with transition matrix \( \mathbb{T} \). In addition suppose that \( \mathbb{T} \) has only one unit eigenvalue. The realized values of \( Z_t \) are the coordinate vectors in \( \mathbb{R}^n \). Let \( \mathbf{q} \) be corresponding row eigenvector restricted so that \( \mathbf{q} \cdot \mathbf{1}_n = 1 \):

\[ \mathbf{q}' \mathbb{T} = \mathbf{q}' \]

\( ^6 \)Hall and Heyde (1980) show how to extend this approach to functional counterparts to the Central Limit Theorem.
Consider an additive functional satisfying
\[ Y_{t+1} - Y_t = D \cdot Z_t + Z'_t F W_{t+1}, \]
where \( \{W_t\} \) is an iid sequence of multivariate standard normally distributed random vectors. The date \( t \) composite state vector is:
\[ X_t = \begin{bmatrix} Z_t \\ W_t \end{bmatrix}. \]
This is a model with discrete changes in the conditional mean and the conditional volatility of the process \( \{Y_t\} \).

First compute
\[ \kappa_2(X_{t+1}, X_t) = Z'_t F W_{t+1}, \]
and
\[ \nu = D \cdot q, \]
Let \( f(x) = f \cdot z \) where
\[ f = D - \nu 1_n. \]
Then \( g(x) = g \cdot z \) where
\[ g = (I - T)^{-1} f \]
and \( \kappa_1(x^*, x) = f \cdot z + g \cdot z^* - g \cdot z. \) Then
\[ Y_t = t \nu + \left[ \sum_{j=1}^{t} \kappa_a(X_j, X_{j-1}) \right] - g \cdot Z_t + g \cdot Z_0 \]
where \( \kappa_a = \kappa_1 + \kappa_2. \) Notice that in this example the martingale increment has a continuous and a discrete component:
\[ \kappa_a(X_{t+1}, X_t) = \underbrace{Z'_t F W_{t+1}}_{\text{continuous}} + \underbrace{g \cdot Z_{t+1} - g \cdot Z_t + f \cdot Z_t}_{\text{discrete}}. \]

### 2.8 A Quadratic Model of Growth

Suppose that \( \{X_t\} \) follows the first-order autoregression:
\[ X_{t+1} = AX_t + BW_{t+1} \]
2.8. A QUADRATIC MODEL OF GROWTH

where $A$ has stable eigenvalues, $B'B$ is nonsingular and $\{W_{t+1}\}$ is a sequence of independent and identically normally distributed random variables with mean zero and covariance matrix $I$. Consider an additive functional $\{Y_t\}$ given by

$$Y_{t+1} - Y_t = \epsilon + D \cdot X_t + \frac{1}{2} X_t' H X_t + F \cdot W_{t+1} + X_t' G W_{t+1}$$

where $H$ is a symmetric matrix. First compute

$$\kappa_2(X_{t+1}, X_t) = F \cdot W_{t+1} + X_t' G W_{t+1}.$$  

Next compute

$$\nu = \epsilon + E \left( \frac{1}{2} X_t' H X_t \right) = \epsilon + \frac{1}{2} \text{trace}(H \Sigma)$$

where $\Sigma$ the covariance matrix in a stochastic steady state given by formula (2.4), and

$$f(x) = D \cdot x + \frac{1}{2} x' H x - \frac{1}{2} \text{trace}(H \Sigma).$$

Recall that $g - \mathbb{T} g = f$ and guess that

$$g(x) = \hat{D} \cdot x + \frac{1}{2} x' \hat{H} x - \frac{1}{2} \text{trace} \left( \hat{H} \Sigma \right).$$

This gives rise to the following three relations:

$$\hat{D} - A' \hat{D} = D,$$

$$\hat{H} - A' \hat{H} A = H.$$  

(2.5)

It may be verified that

$$\hat{H} = \sum_{j=0}^{\infty} (A^j)' H (A^j),$$

$$\hat{D} = (I - A')^{-1} D.$$  

Since $\Sigma = B B' + A \Sigma A'$,

$$\text{trace} \left( \hat{H} \Sigma \right) = \text{trace} \left( \hat{H} B B' \right) + \text{trace} \left( \hat{H} A \Sigma A' \right) = \text{trace} \left( B' \hat{H} B \right) + \text{trace} \left( A' \hat{H} A \Sigma \right) = \text{trace} \left( B' \hat{H} B \right) + \text{trace} \left[ \left( \hat{H} - H \right) \Sigma \right].$$
where the last equality follows from (2.5). Thus

\[ \text{trace} \left( B' \hat{H} B \right) = \text{trace} \left( H \Sigma \right). \tag{2.6} \]

The increment in the martingale component to the additive functional is

\[
\kappa_a(X_{t+1}, X_t) = F \cdot W_{t+1} + X_t' G W_{t+1} + \left( B' \hat{D} \right) \cdot W_{t+1} \\
+ \frac{1}{2} X_{t+1}' \hat{H} X_{t+1} + \frac{1}{2} X_t' \left( H - \hat{H} \right) X_t - \nu \\
= \left( F + B' \hat{D} \right) \cdot W_{t+1} + X_t' \left( G + A' \hat{H} \right) W_{t+1} \\
+ \frac{1}{2} W_{t+1}' B' \hat{H} B W_{t+1} - \frac{1}{2} \text{trace}(H \Sigma) \\
= \left( F + B' \hat{H} \right) \cdot W_{t+1} + X_t' \left( G + A' \hat{H} \right) W_{t+1} \\
+ \frac{1}{2} W_{t+1}' B' \hat{H} B W_{t+1} - \frac{1}{2} \text{trace}(B' \hat{H} B)
\]

where the last equality follows from (2.6).
Chapter 3

Multiplicative Functionals

If \( Y \) is a Gaussian random variable with mean \( \mu \) and variance \( \sigma^2 \), then \( M = \exp(Y) \) is said to have a log normal distribution and the logarithm of its mean satisfies

\[
\log EM = \mu + \frac{\sigma^2}{2}.
\] (3.1)

A variety of formulas in this chapter will remind us of (3.1) as we use log normal distributions to create models of stochastic geometric growth.

Applied econometricians find it convenient to build stochastic process models after taking logarithms. Such models were on display in the previous chapter. In this chapter we use (3.1) and other formulas that show how variation in logarithms contributes to the expected growth in levels. We develop methods that allow us to explore the impact of compounding stochastic growth over alternative time horizons.

3.1 Multiplicative Growth

Definition 3.1.1. Let \( \{Y_t : t = 0, 1, ...\} \) be an additive functional parameterized by \( \kappa \), and let \( M_t = \exp(Y_t) \). We say that \( \{M_t : t = 0, 1, ...\} \) is a multiplicative functional parameterized by \( \kappa \).

Figure 3.1 plots an additive functional and its multiplicative counterpart. Previously we have seen how additive functionals can display Law of Large Numbers type behavior and Central Limit type behavior. To analyze the corresponding multiplicative functional formed through exponentiation will require an alternative approach as we will see in this and later chapters.
Additive and Multiplicative Functionals

A simple discrete-time additive functional $Y$

Note: X’s can be temporally dependent

A simple multiplicative functional $M$ made from exponentiating $Y$

Range of $Y$ grows + or -

Figure 3.1: This figure illustrates a particular additive functional and its multiplicative counterpart.

A special type of multiplicative functional will be of particular interest in this chapter.

**Definition 3.1.2.** A multiplicative functional parameterized by $\kappa$ is a multiplicative martingale if

$$E \left( \exp \left[ \kappa(X_{t+1}, X_t) \right] \right| X_t = x) = 1$$

for all $x$.

Consider a benchmark additive functional $\{Y_t : t = 0, 1, \ldots\}$ and a class of additive functionals that grow together:

$$Y_t^\diamond = Y_t + h(X_t),$$
3.1. **MULTIPLICATIVE GROWTH**

or taking exponentials:

\[ M_t^\circ = M_t \exp [h(X_t)] \]

where \( M_t^\circ = \exp(Y_t^\circ) \). Notice that

\[ Y_{t+1}^\circ - Y_t^\circ = \kappa(X_{t+1}, X_t) + h(X_{t+1}) - h(X_t) = \kappa^\circ(X_{t+1}, X_t). \]

Provided that \( E[|\kappa(X_{t+1}, X_t)|] < \infty \) and \( E|h(X_t)| < \infty \),

\[ E[\kappa(X_{t+1}, X_t)] = E[\kappa^\circ(X_{t+1}, X_t)], \]

suggesting that \( \{M_t : t = 0, 1, \ldots\} \) and \( \{M_t^\circ : t = 0, 1, \ldots\} \) have the same growth rate. As we will see, this is often the case, although the presence of randomness means that the common growth rate is not \( E[\kappa(X_{t+1}, X_t)] \).

More generally, consider

\[ M_t^\circ = M_t f(X_t) \]

for a multiplicative functional \( \{M_t : t = 0, 1, \ldots\} \) and some \( f \) that can be negative. Construct an operator

\[
M f(x) = E \left[ \exp(Y_{t+1} - Y_t) f(X_{t+1}) | X_t = x \right] \\
= E \left[ \frac{M_{t+1}}{M_t} f(X_{t+1}) | X_t = x \right]. \tag{3.2}
\]

Consider

\[
M^2 f(x) = E \left[ \exp(Y_{t+2} - Y_t) f(X_{t+2}) | X_t = x \right] \\
= E \left[ \frac{M_{t+2}}{M_t} f(X_{t+2}) | X_t = x \right].
\]

The limiting growth rate is

\[
\eta(M) = \lim_{j \to \infty} \frac{1}{j} \log E \left[ \exp(Y_{t+j} - Y_t) f(X_{t+j}) | X_t = x \right] \\
= \lim_{j \to \infty} \frac{1}{j} \log E \left[ \frac{M_{t+j}}{M_t} f(X_{t+j}) | X_t = x \right]. \tag{3.3}
\]

for \( f > 0 \). For instance, \( f \) can be unity. This limit will be the same for a large collection of \( f \), where we will have more to say about when this approximation is valid.
3.2 Four Growth Specifications

We study three alternative specifications of stochastic growth and for each one compute the limiting growth rate $\eta(M)$ in (3.3).

Finite-State Markov Growth

As in Section 1.7, consider a finite state Markov chain with transition matrix $\mathbb{P}$. Suppose that

$$Y_{t+1} - Y_t = \kappa(X_{t+1}, X_t)$$

Construct a square matrix $\mathbb{M}$ by multiplying entry $(i, j)$ of $\mathbb{P}$ by entry $(i, j)$ of a matrix $\exp(\kappa)$. All entries of $\mathbb{M}$ are thereby constructed to be nonnegative, but unlike a transition matrix, rows do not sum to one.

**Theorem 3.2.1. (Perron-Frobenius)** Suppose that there exists a $j \geq 1$ for which all entries of $\mathbb{M}^j$ are positive. Then there exists a unique nonnegative solution (up to scale) to

$$\mathbb{M}e = \exp(\eta)e$$

with all elements of $e$ being strictly positive. Moreover, $\exp(\eta)$ is the eigenvalue with largest magnitude.

The eigenvector, $e$, and associated eigenvalue, $\exp(\eta)$, determine growth over long horizons as we will now show.

- Why does $e$ have strictly positive entries?

  Consider a vector $f$ with nonnegative entries and $f \cdot 1_n > 0$. Then $\mathbb{M}^j f$ must have all strictly positive entries. If $e$ happens to be an eigenvector of $\mathbb{M}$ with eigenvalue $\exp(\eta)$, then the entries of $e$ must all be strictly positive along with the eigenvalue $\exp(\eta)$.

- How can we use $e$ to dominate growth?

  Let $f$ be any other vector. Scale $e$ so that $e \geq f$ entry by entry. Then

  $$\mathbb{M}^j f \leq \mathbb{M}^j e = (\exp(\eta))^j e.$$  

  Thus,

  $$\exp(-j\eta)\mathbb{M}^j f \leq e.$$
Similarly, we may scale \( e \) so that \(-f \leq e\). Then
\[
- \exp(-j\eta)M^j f \leq e.
\]
It follows that all other eigenvalues must have absolute values that are less than or equal to \( \exp(\eta) \).

• Why is \( \exp(\eta) \) unique and \( e \) unique up to scale?

Suppose now that \( g \geq 0 \) and \( Mg = \exp(\rho)g \). We have already argued that the entries of \( \tilde{e} \) must be strictly positive. We may scale \( e \) so that \( \bar{r}e \geq g \geq r e \) for some strictly positive numbers \( \bar{r} \) and \( r \). Thus,
\[
\bar{r} \exp(j\eta)e \geq \exp(j\rho)g \geq r \exp(j\eta)e.
\]
This inequality implies that \( \eta = \rho \). We may also find a \( r > 0 \) such that \( re - g \geq 0 \) and one entry of this difference is exactly zero. Then \( re - g \) is either an eigenvector associated with an eigenvalue \( \exp(\eta) \) or it is exactly zero. The former is not possible because we know that a nonnegative eigenvector must in fact have strictly positive entries. Thus, \( re = g \).

To summarize, there is only one eigenvector (up to scale) with nonnegative entries and in fact it must have strictly positive entries. Moreover, the growth rate of \( Mf \) for any nonnegative \( f \) is no greater than \( \eta \).

Theorem 3.2.1 allows us to compute the asymptotic growth rate \( \eta(M) \) for this process as follows. Solve
\[
\tilde{e}'M = \exp(\eta)\tilde{e}'.
\]
Thus \( \tilde{e}' \) is a row eigenvector. Normalize this vector so that \( \tilde{e} \cdot e = 1 \). Suppose that \( M \) can be decomposed as:
\[
M = FDF^{-1}
\]
where the columns of \( F \) are column eigenvectors and the rows of \( F \) are row eigenvectors. One column of \( F \) consists of \( e \) and the corresponding row of \( F^{-1} \) consists of \( \tilde{e}' \). The matrix \( D \) is a diagonal matrix with eigenvalues on the diagonal.\(^1\) Then
\[
\lim_{j \to \infty} \exp(-\eta j)M^j = e\tilde{e}'.
\]

\(^1\)Representation 3.4 does not always apply, but the limit characterized here can be obtained using other arguments. More generally, we could work with Jordan decomposition of the matrix.
CHAPTER 3. MULTIPLICATIVE FUNCTIONALS

Equation (3.5) gives a refined long-run approximation. For large $j$, the approximate expected growth is

$$M^j f \approx \exp(\eta_j) e (\tilde{e} \cdot f)$$

The right-hand side of the approximation depends on $f$ only through a scale factor. The approximate variation across states is determined by $e$ for large $j$.

**Example 3.2.2.** In this example, a two-state Markov chain has transition matrix $T = \begin{bmatrix} .9 & .1 \\ .1 & .9 \end{bmatrix}$ and $\kappa = \begin{bmatrix} .02 & -.005 \\ 0 & -.01 \end{bmatrix}$. The stationary distribution for $\{X_t\}$ is $q = \begin{bmatrix} .5 \\ .5 \end{bmatrix}$. The stationary mean of $\kappa$ is $E\kappa(X_{t+1}, X_t) = .0043$ while $\eta = .0053$. The top panel of figure 3.2 shows a simulation.

**Example 3.2.3.** In this example, a two-state Markov chain estimated by Cecchetti et al. (2000) to model annual U.S. consumption growth has transition matrix $T = \begin{bmatrix} 0.9780 & 0.0220 \\ 0.4850 & 0.5150 \end{bmatrix}$ and $\kappa = \begin{bmatrix} 0.0225 & 0.0225 \\ -0.0678 & -0.0678 \end{bmatrix}$. The stationary distribution for $\{X_t\}$ is $q = \begin{bmatrix} 0.9566 \\ 0.0434 \end{bmatrix}$. The stationary mean of $\kappa$ is $E\kappa(X_{t+1}, X_t) = .0186$ while $\eta = .0190$. The bottom panel of figure 3.2 shows a simulation.

**Log-normal Growth**

Consider next the log-normal model given previously in section 2.5. Our purpose here is to study the expected growth rate of the corresponding multiplicative process. In this example,

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

and

$$Y_{t+1} - Y_t = D \cdot X_t + \nu + F \cdot W_{t+1}$$

where $\{W_{t+1} : t = 0, 1, \ldots\}$ is an independently and identically distributed sequence of multivariate standard normal random vectors. Let

$$f(x) = \exp(U \cdot x + u).$$
Figure 3.2: The top panel plots a simulation of the multiplicative process studied in example 3.2.2, while the bottom panel plots a simulation of the Cecchetti et al. (2000) process for annual aggregate U.S. mentioned in example 3.2.3.
CHAPTER 3. MULTIPLICATIVE FUNCTIONALS

Form the operator:
\[ \mathbb{M}f(x) = E[\exp(Y_{t+1} - Y_t) \exp(U \cdot X_{t+1} + u)|X_t = x] \]
or
\[ E[\exp((D + A'U) \cdot X_t + (\nu + u) + (F + B'U)W_{t+1})|X_t = x] \).

Using the properties of log-normal distributions or else appealing to a complete-the-square argument under a Gaussian distribution shows that
\[ \mathbb{M}f(x) = \exp(\tilde{U} \cdot x + \tilde{u}) \]
where
\[ \tilde{U} = D + A'U \]
and
\[ \tilde{u} = u + \nu + \frac{|F + B'U|^2}{2}. \] (3.6)

To compute higher powers \( \mathbb{M}^j f(x) = E[\exp(Y_{t+j} - Y_t) \exp(U X_{t+j} + u)|X_t = x] \), we iterate on
\[ U_{j+1} = D + A'U_j \]
The limit solves the fixed point equation:
\[ U^o = D + A'U^o, \]
or \( U^o = (I - A')^{-1}D \). Similarly, form
\[ u_{j+1} - u_j = \nu + \frac{|F + B'U_j|^2}{2} \]
The increment in the \( u_j \) converges to
\[ \eta = \nu + \frac{|F + B'U^o|^2}{2} = \nu + \frac{|F + B'(I - A')^{-1}H|^2}{2}. \] (3.8)

Notice how the limiting growth rate \( \eta \) acquires a contribution from volatility that equals the same correction \( \hat{\kappa}(X_{t+1}, X_t) \) given by (2.3) that is associated with the martingale decomposition provided in Proposition 4.2. The variance of the martingale increment gives the pertinent volatility adjustment to growth.

The limiting function \( U^o x \) solves the eigenfunction problem:
\[ \mathbb{M} \exp(U^o \cdot x) = \exp(\eta) \exp(U^o \cdot x) \]
The function \( \exp(U^o \cdot x) \) is strictly positive. The number \( \eta \) gives the asymptotic rate of growth of \( \mathbb{M}^j f \) not only for \( f(x) = \exp(U \cdot x + u) \) but for a substantially larger collection of functions \( f \).
3.2. FOUR GROWTH SPECIFICATIONS

Growth-Rate Regimes

As in section 2.7, suppose that \( \{Z_t\} \) evolves according to an \( n \)-state Markov chain with transition matrix \( T \). In addition suppose that \( T \) has only one unit eigenvalue. The realized values of \( Z_t \) are the coordinate vectors in \( \mathbb{R}^n \).

Consider the multiplicative functional \( \{\exp(Y_t) : t = 0, 1, \ldots\} \) where

\[ Y_{t+1} - Y_t = D \cdot Z_t + Z_t'FW_{t+1}, \]

and \( \{W_t\} \) is an iid sequence of multivariate standard normally distributed random vectors.

Notice that

\[ E[\exp(D \cdot Z_t + Z_t'FW_{t+1})g \cdot Z_{t+1}|Z_t = z] = \exp \left( D \cdot z + \frac{1}{2} z'F'Fz \right) z'Tg \]

By evaluating the right-hand side of this expression at each of the coordinate vectors, we see that the left-hand side can be expressed as \( z'Mg \) for some matrix \( M \). Given this construction, we can analyze this model using the methods described in subsection 3.2.

Log-Quadratic Growth

As in section 2.8 suppose that

\[ X_{t+1} = AX_t + BW_{t+1}, \]

and consider the associated process:

\[ Y_{t+1} - Y_t = \epsilon + D \cdot X_t + \frac{1}{2} X_t'HX_t + F \cdot W_{t+1} + X_t'GW_{t+1}. \]

Again \( \{W_{t+1} : t = 0, 1, \ldots\} \) is a multivariate sequence of iid standard normally distributed random vectors, \( G \) and \( H \) are matrices with the appropriate dimensions with \( H \) being symmetric. Form \( M_t = \exp(Y_t) \).

Let

\[ Mf(x) = E[\exp(Y_{t+1} - Y_t)f(X_{t+1})|X_t = x]. \]

Consider the specific case in which

\[ f(x) = \exp \left( \frac{1}{2} x'Ux + u \right) \]
where $U$ is a symmetric matrix. Then

$$Mf(x) = \exp\left(\frac{1}{2}x'\hat{U}x + \hat{u}\right)$$

where

$$\hat{U} = A'UA + H + (G + A'UB)(I - B'UB)^{-1}(G' + B'UA') \quad (3.9)$$

and

$$\hat{u} = \epsilon + u + \frac{1}{2}F'(I - B'UB)F - \frac{1}{2} \log \det(I - B'UB).$$

These formulas are obtained from an extended version of a “complete the square” argument together with the properties of the normal density. Equation (3.9) is a special case of a Riccati equation mapping $U$ into $\hat{U}$.\(^2\)

One way to compute a positive eigenfunction is to iterate starting say from $U_0 = 0$ and compute the limit of the sequence $\{U_j : j = 1, 2, \ldots\}$. Call this limit $U^o$ and then set

$$\eta = \epsilon + u + \frac{1}{2}F'(I - B'U^oB)F - \frac{1}{2} \log \det(I - B'U^oB).$$

Then $e(x) = \exp(x'U^o x)$ satisfies:

$$\mathcal{M}e(x) = \exp(\eta)e(x),$$

which states that $\exp(x'U^o x)$ is positive eigenfunction associated with the eigenvalue $\exp(\eta)$. The iterative method presumes that $I - B'U_j B$ is positive definite for every $j$. Of course, $\eta$ will only be finite provided that $I - B'U^o B$ is also positive definite.

### 3.3 Multiplicative Martingale Extraction

Return to the operator (3.2):

$$\mathcal{M}f(x) = E\left[\exp(Y_{t+1} - Y_t)f(X_{t+1})|X_t = x\right].$$

\(^2\)There are often multiple solutions to this equation, but only one of these solutions will be of us. It is limit point of iterations starting from a unit function. See Hansen and Scheinkman (2009) for a discussion. When a constant is included as a state, the coefficient on the square of this state should be set to zero at each iteration.
As in the examples, construct a strictly positive function \( e \) that solves the eigenvalue equation

\[ M e(x) = \exp(\eta) e(x). \]  

(3.10)

Following Hansen and Scheinkman (2009), use this function to build an alternative multiplicative functional

\[ M^o_t = \exp(-\eta t) M_t \frac{e(X_t)}{e(X_0)}, \]

and an associated additive process \( \{Y^o_t : t = 0, 1, \ldots\} \) such that

\[
Y^o_{t+1} - Y^o_t = -\eta + Y^o_{t+1} - Y_t + \log e(X_{t+1}) - \log e(X_t) \\
= \kappa_m(X_{t+1}, X_t)
\]

where \( \kappa_m(x^*, x) = -\eta + \kappa(x^*, x) + \log e(x^*) - \log e(x) \). By construction,

\[ E \left[ \exp(Y^o_{t+1} - Y^o_t) | X_t = x \right] = 1. \]

Equivalently,

\[ E \left( \exp [\kappa_m(X_{t+1}, X_t)] | X_t = x \right) = 1. \]

Thus, as a consequence of our construction, \( \{M^o_t : t = 0, 1, \ldots\} \) is a multiplicative martingale. This gives us a martingale factorization and hence a multiplicative analog to Proposition 4.2:

**Proposition 3.3.1.** Suppose that (3.10) has a solution. Then

\[ M_t = \exp(\eta t) M^o_t \frac{e(X_0)}{e(X_t)} \]

where \( \{M^o_t : t = 0, 2, \ldots\} \) is a multiplicative martingale with representation\(^3\)

\[
\log M^o_{t+1} - \log M^o_t = -\eta + \kappa(x^*, x) + \log e(x^*) - \log e(x) \\
= \kappa_m(x^*, x)
\]

\(^3\)Decompositions like this were suggested in Bansal and Lehmann (1997) and Alvarez and Jermann (2005) but their formal construction and link to eigenfunctions is given by Hansen and Scheinkman (2009).
This martingale extraction gives a possible refinement to the view of $\eta$ as a long term growth rate by producing a martingale component and an apparent transient net of the growth rate. In the next chapter we will see how to view the martingale $\{M_t^\eta : t \geq 0\}$ as a change of probability measure. The multiplicative martingale decomposition turns out not to be unique because there may be multiple positive solutions to the eigenvalue problem. (We have already seen that this is possible for the log-quadratic model.) By exploring the implications of this change in probability measure we will have cause to select one among the potentially multiple solutions that is of particular interest.
Chapter 4

Likelihood-ratio Processes

By regarding likelihood ratios and derivatives of log-likelihoods as martingales, this chapter describes how to characterize some of their limiting features. We apply results from previous chapters by first establishing that likelihood ratio processes are multiplicative martingales and derivatives of log-likelihoods are additive martingales.

4.1 Likelihood Processes

Suppose that a vector of “signals” are observed. Each of the signals evolves as an additive functional of the Markov process:

\[ Y_{t+1}^{[i]} - Y_t^{[i]} = \kappa_i(X_{t+1}, X_t) \]

for \( i = 1, 2, \ldots, k \). Stack these \( k \) signals into a vector \( O_{t+1} \) and form:

\[ O_{t+1} = \Psi(X_{t+1}, X_t). \]

Initially we impose the invertibility condition:

**Assumption 4.1.1.** There exists a function \( \Phi \) such that

\[ X_{t+1} = \Phi(O_{t+1}, X_t). \]

\( and \ X_0 \ is \ observed. \)

Under this restriction, given the current state and next period signals, we can infer the next period state. Thus initial state and sequence of signals
reveal the sequence of states. Subsequently, we will relax this restriction by treating the states as hidden.

Let $\tau$ denote a measure over the space $\mathcal{O}$ of admissible signals.

**Assumption 4.1.2.** $O_{t+1}$ has a conditional density $\psi(\cdot|x)$ with respect to $\tau$ conditioned on $X_t = x$.

The density $\psi$ is built from $\Psi$ and the transition distribution $T(dx^*|x)$. One possibility is

**Example 4.1.3.** $O_{t+1} = X_{t+1}$ and the Markov process $\{X_t\}$ has a transition density $\phi(x^*|x)$ relative to a measure $\lambda$. In this case $\phi = \psi$ and $\tau = \lambda$.

Another example is:

**Example 4.1.4.**

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

$$O_{t+1} = DX_t + FW_{t+1}$$

where $F$ is nonsingular. Thus

$$X_{t+1} = [A - BF^{-1}D]X_t + BF^{-1}O_{t+1},$$

which gives $\Phi$ as a linear function of $O_{t+1}$ and $X_t$. The conditional distribution for $O_{t+1}$ is normal with mean $DX_t$ and nonsingular covariance matrix $FF'$. The conditional distribution has a density against Lebesgue measure on $\mathbb{R}^m$.

The joint density of a history of observations $O_1, \ldots, O_t$ conditioned on $X_0$ is

$$L_t = \prod_{j=1}^{t} \psi(O_j|X_{j-1}).$$

The logarithm of $\{L_t : t = 0, 1, \ldots, t\}$ is an additive functional with

$$\kappa_t(x^*, x) = \log \psi(o^*|x) = \log \psi[\Psi(x^*, x)|x]$$

where $o^*$ is a realized value of $O_{t+1}$. Thus the likelihood is a multiplicative functional. The likelihood and log-likelihood are stochastic processes because of their dependence on the processes for signals and states.\footnote{Recall that we have initialized additive functionals at zero and multiplicative functionals at one when $t = 0$. For likelihoods and log-likelihoods we achieve this initialization by conditioning on $X_0$. More generally, we can introduce a density for the initial state $X_0$ which simply alters the initialization of $L_t$ and $\log L_t$.}
Consider again example (4.1.4). It follows from the formula for the normal density that
\[
\kappa_\ell(X_{t+1}, X_t) = -\frac{1}{2}(O_{t+1} - DX_t)'(FF)'^{-1}(O_{t+1} - HX_t) \\
- \frac{1}{2}\log \det(FF) - \frac{k}{2}\log(2\pi),
\]
and
\[
\log L_t = -\frac{1}{2}\sum_{j=1}^{t}(O_{t} - HX_{t-1})'(FF)'^{-1}(O_{t} - HX_{t+1}) \\
- \frac{1}{2}\log \det(FF) - \frac{kt}{2}\log(2\pi).
\]

Our construction of \( \{L_t : t = 0, 1, \ldots\} \) does not require that \( \psi \) be the density that actually generates the data. Suppose as econometricians we know the true transition density only up to an unknown parameter vector \( \theta \) so that our model is \( \psi(o^*|x, \theta) \) for \( \theta \) in a parameter space \( \mathcal{P} \). We can use the above construction for each parameter vector \( \theta \) by building
\[
\kappa_\ell(x^*, x|\theta) = \log \psi(o^*|x, \theta) = \log \psi[\Psi(x^*, x)|x, \theta]
\]
where \( \log L_0(\theta) \) is the logarithm of the initial density function for \( X_0 \), which we allow to depend on \( \theta \). We now build a log-likelihood process:
\[
\log L_t(\theta) = \frac{1}{2}\sum_{j=1}^{t}\log \phi(O_j|X_{j-1}, \theta) + \log L_0(\theta).
\]
Let \( \theta_o \) be the parameter value that generates the data.

In what follows we study likelihood ratios. These ratios are built from contributions of the form:
\[
\exp[\kappa_\ell(x^*, x|\theta) - \kappa_\ell(x^*, x|\theta_o)] = \frac{\psi(o^*|x, \theta)}{\psi(o^*|x, \theta_o)}.
\]

The conditional expectation of the incremental contribution to the likelihood ratio is:
\[
\int_{\mathcal{O}} \left[ \frac{\psi(o^*|x, \theta)}{\psi(o^*|x, \theta_o)} \right] \psi(o^*|x, \theta_o) \tau(do^*) = \int_{\mathcal{O}} \psi(o^*|x, \theta) \tau(do^*) = 1 \quad (4.2)
\]
for all \( x \in \mathcal{X} \) and all \( \theta \in \mathcal{P} \). This follows because \( \psi(o^*|x, \theta) \) is a density for every \( x \) and \( \theta \). Therefore,
Figure 4.1: The logarithmic function is the concave function in this plot. This function is zero when evaluated at unity. By forming averages using the two endpoints of the straight line below the logarithmic function, we are lead to some point on the line segment depending upon the weights used in the averaging. Jensen’s inequality in this case is just a statement that the line segment must lie below the logarithmic function.

**Theorem 4.1.5.** For each $\theta$, the likelihood ratio process \( \left\{ \frac{L_t(\theta)}{L_t(\theta_0)} : t = 0, 1, \ldots \right\} \) is a multiplicative martingale.

### 4.2 Log-Likelihoods

The logarithmic function is concave. As an application of Jensen’s Inequality, the expectation of the logarithm of a random variable cannot exceed that of the logarithm of the expectation. Thus we obtain the inequality,

\[
\int_O \left[ \log \psi(o^*|x, \theta) - \log \psi(o^*|x, \theta_0) \right] \psi(o^*|x, \theta_0) \tau(\text{do}^*) \\
\leq \log \int_O \left[ \frac{\psi(o^*|x, \theta)}{\psi(o^*|x, \theta_0)} \right] \psi(o^*|x, \theta_0) \tau(\text{do}^*) = 0.
\]  

(4.3)
Furthermore, equality obtains only if:

$$\log \psi(o^*|x, \theta) = \log \psi(o^*|x, \theta_0)$$

with probability one under the measure $\psi(o^*|x, \theta_0)\tau(\omega^*)$. Figure 4.1 illustrates the role of concavity in obtaining this bound. A consequence of this inequality is that

$$E [\log L_{t+1}(\theta) - \log L_{t+1}(\theta_0)|\mathcal{F}_t] \leq \log L_t(\theta) - \log L_t(\theta_0)$$

which implies that the log-likelihood is a supermartingale.

**Theorem 4.2.1.** For each $\theta$, the log-likelihood ratio process

$\{\log L_t(\theta) - \log L_t(\theta_0) : t = 0, 1, \ldots\}$

is an additive supermartingale.

Since the log-likelihood is an additive process, we may apply Proposition to obtain a decomposition. Since the log-likelihood ratio is a supermartingale, the coefficient $\nu$ on the time trend $t$ is necessarily negative. Maximum likelihood estimation estimates the time trend for the log-likelihood by a sample average:

$$\hat{\nu}_t(\theta) = \frac{1}{t} \log L_t(\theta),$$

and finds the maximum.

### 4.3 Score process

For simplicity suppose that $\Theta$ is an open interval of $\mathbb{R}$ containing $\theta_0$. The parameter $\theta_0$ necessarily maximizes the objective:

$$\int_{\mathcal{X}} \log \psi(o^*|x, \theta)\psi(o^*|x, \theta_0)\tau(\omega^*)$$

(4.4)

given (4.3). Suppose that we can differentiate under the integral sign. The first-order conditions imply that

$$\int_{\mathcal{X}} \left[ \frac{d}{d\theta} \log \psi(o^*|x, \theta_0) \right] \psi(o^*|x, \theta_0)\tau(\omega^*) = 0.$$  (4.5)

\[2\] The decomposition has a straightforward modification to allow for a nonzero initial condition appended to the additive functional.
Definition 4.3.1. The score process \( \{ s_t : t = 0, 1, \ldots \} \) is:

\[
s_t = \sum_{j=1}^{t} \frac{d}{d\theta} \log \psi(O_t|X_t, \theta_o).
\]

Theorem 4.3.2. The score process is an additive martingale.

Proof. This follows directly from (4.5).

Since the score process is initialized at 

\[
\frac{d}{d\theta} \log L_0(\theta_o),
\]

and not at zero, we refer to it as an additive process instead of an additive functional. Nevertheless Theorem 4.3.2 lets us use the martingale central limit theorem to characterize the large sample behavior of the score process because the initial contribution does not change as \( t \) increases. The resulting central limit approximation can be used to establish a large sample characterization of the maximum likelihood estimator in a Markov setting.

The variance:

\[
E \left( \left[ \frac{d}{d\theta} \log \psi(O_{t+1}|X_t, \theta_o) \right]^2 \right)
\]

of the martingale increment is one measure of “information” in the data about the parameter \( \theta_o \) called Fisher information, named for the statistician R.A. Fisher. The method of maximum likelihood uses the reciprocal of the Fisher information as the asymptotic variance of the associated estimator.

Consider next the extension of Fisher information when the parameterization is not multivariate. Suppose that the likelihood is parameterized on an open set \( \Theta \) in a finite dimensional Euclidean space and that the true parameter vector is \( (\theta_o, \vartheta_o) \in \Theta \). Write the multivariate counterpart to the score process as

\[
\left\{ \begin{array}{c}
\tilde{s}_{t+1} \\
\hat{s}_{t+1}
\end{array} \right\} : t = 0, 1, \ldots
\]

where \( \{ s_{t+1} : t = 0, 1, \ldots \} \) is the constructed by differentiating the log-likelihood with respect to \( \theta \) and \( \{ \tilde{s}_{t+1} : t = 0, 1, \ldots \} \) is constructed by differentiating with respect to \( \vartheta \). Estimating \( \vartheta_o \) at the same time as \( \theta_o \) makes the estimation problem more difficult relative to estimating \( \theta_o \) given
knowledge of $\theta_o$. Fisher’s measure of information corrects for this additional challenge. To construct the measure, consider the population regression:

$$s_{t+1} - s_t = \beta(\tilde{s}_{t+1} - \tilde{s}_t) + e_{t+1}$$

where $\beta$ is the population regression coefficient and $e_{t+1}$ is the population regression residual. The Fisher information is: $E(e_{t+1})^2$ which is no larger than $E(s_{t+1} - s_t)^2$. Information is lost by making the estimation problem more difficult and this is reflected in the inequality:

$$E[(e_{t+1})^2] \leq E[(s_{t+1} - s_t)^2].$$

### 4.4 Limiting Behavior of the Likelihood Ratio

Next we consider a rather different approach to measuring when an estimation problem is difficult. Instead of looking locally, consider two alternative models, perhaps associated with two distinct parameter values say $\theta_o$ and $\theta_1$. Suppose the challenge is to select the correct model given a data set. As more and more data become available, the *ex ante* probability of making a mistake becomes smaller. To characterize this behavior formally, we are led to study the tail behavior of the likelihood ratio. Specifically, we consider the probability that the likelihood ratio exceeds a given threshold. By setting the threshold, we can study the probability of making a mistaken model selection by using a likelihood ratio test. From a decision-theoretic perspective, the threshold is determined by the prior probabilities assigned to the respective models and losses associated with misclassification.

Following Chernoff (1952), we use the *large deviation method* described in the next chapter to characterize the limiting behavior of these threshold probabilities. This approach applied to likelihood ratios is based on the bound:

$$\frac{1}{\exp(\alpha r)} E \left( \left[ \frac{L_{t}(\theta_1)}{L_{t}(\theta_o)} \right]^\alpha \mid X_0 = x \right) \geq Pr \{ \log L_{t}(\theta_1) - \log L_{t}(\theta_o) \geq r \mid X_0 = x \} \quad (4.6)$$

When $\theta_o$ is the true parameter value, we have already argued that the likelihood ratio process is a multiplicative martingale. A *large deviation*
CHAPTER 4. LIKELIHOOD-RATIO PROCESSES

A bound is given

\[ \eta \left( \left[ \frac{L(\theta_1)}{L(\theta_o)} \right]^\alpha \right) = \lim_{t \to \infty} \frac{1}{t} \log E \left( \frac{L_t(\theta_1)}{L_t(\theta_o)} \right)^\alpha | X_0 = x \]

\[ \lim_{t \to \infty} \frac{1}{t} \log Pr \left\{ \log L_t(\theta_1) - \log L_t(\theta_o) \geq r | X_0 = x \right\} \quad (4.7) \]

Since the likelihood ratio is a martingale and the function \( z^\alpha \) is concave in \( z \) for \( 0 < \alpha < 1 \), \( \eta \left( \left[ L(\theta_1)/L(\theta_o) \right]^\alpha \right) \) is less than or equal to zero. This is to natural because we seek a bound on how fast probabilities decline to zero. Notice that a contribution coming from the choice of \( r \) can be ignored. It vanishes because we are studying the rate at which new information changes the threshold probabilities.

The bound works for any choice of \( 0 < \alpha < 1 \), so the left-hand side can be minimized by choice of \( \alpha \). We are minimizing a convex function that is less than or equal to zero and approaches zero at the two endpoints: \( \alpha = 0, \alpha = 1 \). Figure 4.2 depicts \(-\eta \left( \left[ L(\theta_1)/L(\theta_o) \right]^\alpha \right)\) and \(-\eta \left( \left[ L(\theta_o)/L(\theta_1) \right]^\alpha \right)\) as a function of \( \alpha \). Both functions are concave in \( \alpha \). A remarkable property is that provided that we optimize over \( \alpha \), the rate bound would be the same if we switched the roles of \( L_t(\theta_1) \) and \( L_t(\theta_o) \) by letting \( L_t(\theta_1) \) be the true data generating process.\(^3\) This invariance allows Chernoff (1952) to view the resulting rate bound as a measure of statistical discrepancy between two models. The rate measures the asymptotic decay in mistake probabilities when models are selected using relative likelihoods. There are two types of mistakes. One is to conclude that the data are generated by model two when in fact the data are generated by model one. The other is to conclude that the data are generated by model one when in fact the data are generated by model two. When the likelihood ratio is compared to a constant threshold, both mistake probabilities tend to zero at a geometric rate, and this rate is the same independent of how large the threshold is!

\(^3\)This property follows from the fact that for two positive functions \( g_1 \) and \( g_2 \)

\[ \int \left( \frac{g_2}{g_1} \right)^\alpha g_1 d\lambda = \int (g_1)^\alpha (g_2)^{1-\alpha} d\lambda = \int \left( \frac{g_1}{g_2} \right)^{1-\alpha} g_2 d\lambda.\]
Figure 4.2: The vertical axis is implied asymptotic bound on the decay rate for the mistake probabilities for distinguishing between two models. The bound is scaled by 100 and thus expressed as a percent. This bound is reported for alternative choices of $\alpha$ as depicted on the horizontal axis. The difference between the two curves reflects the choice of models in the numerator and denominator of the likelihood ratio. The best bound is given by the maximum of the two curves. The maximizing value for each curve is the same and given a maximizing choice $\alpha^*$ of one curve, the other curve is maximized at $1 - \alpha^*$.

4.5 Twisted Probability Measures

We know that a likelihood ratio is a multiplicative martingale. In many applications it is convenient to go the other direction by interpreting a multiplicative martingale as a likelihood ratio and using it to form another joint probability density for the original process $X_t, t \geq 0$.

Recall that a multiplicative martingale can be represented as $M_t = \exp(Y_t)$ where

$$Y_{t+1} - Y_t = \kappa_m(X_{t+1}, X_t)$$

and

$$E\left(\exp[\kappa_m(X_{t+1}, X_t)] | X_t = x\right) = 1.$$
We can then form a “distorted” or “twisted” probability measure and associated expectations operator:

$$M^\circ f(x) = E(\exp[\kappa_m(X_{t+1}, X_t)] f(X_{t+1}) | X_t = x)$$

Notice that $M^\circ$ maps nonnegative functions into nonnegative functions and a constant function into the same constant function. These properties allow us to call $M^\circ$ a conditional expectation operator for some Markov process. The process $\{M^\circ_t : t = 0, 1, \ldots\}$ is the likelihood ratio for this alternative probability distribution.

**Definition 4.5.1.** The Markov process $\{X_t : t = 0, 1, \ldots\}$ is said to be stochastically stable under the change of measure if

$$\lim_{j \to \infty} (M^\circ)^j f = \tilde{E} f$$

with probability measure one for each bounded function $f$ where $\tilde{E} f$ denotes the limit function restricted to be constant.

We use the notation $\tilde{E}$ to represent the limit points because, as we argued previously, the limit points are expectations under an invariant distribution for the Markov process. In this particular instance, it is the invariant distribution for the twisted specification. As discussed in Chapter 1, Meyn and Tweedie (1993) describe methods for verifying stochastic stability. When $M^\circ f$ converges for bounded functions $f$, there exists a probability measure over the Markov state space associated with $\tilde{E}$, which allows us to extend the domain of $\tilde{E}$. As a consequence, the limit in (4.8) can be extended to any function $f$ for which $\tilde{E}|f| < \infty$.

When we presented Proposition 3.3.1, we argued that the method for constructing a multiplicative martingale did not produce unique solution, but Hansen and Scheinkman (2009) show that only one such solution is stochastically stable. From stochastic stability, it follows that

$$\exp(-\eta j) \lim_{j \to \infty} M^j f(x) = e(x) \tilde{E} \left[ \frac{f(X_t)}{e(X_t)} \right]$$

provided that

$$\tilde{E} \left( \frac{|f|}{e} \right) < \infty$$
Thus, a construction that preserves stochastic stability allows us to provide a more refined characterization of limiting behavior after we scale by the growth rate.
Bibliography


