A Primer in Econometric Theory

Lecture 6: Further Topics in Probability

John Stachurski
Lectures by Akshay Shanker

March 26, 2017

Overview

Lecture covers key results and concepts on the theory of Stochastic processes

- Stationarity and ergodicity
- Markov processes
- Martingales
- Simulation techniques

Stochastic Processes

A **stochastic process** on $\mathbb{R}^K$ is a sequence of random vectors $\{\mathbf{x}_t\}_{t \geq 0}$ defined on a *common* probability space $(\Omega, \mathscr{F}, \mathbb{P})$

Common probability space pins down the joint distribution across t

Probabilities such as

$$\mathbb{P}\{\mathbf{x}_t \geq 0 \text{ for all } t\} \tag{1}$$

or

$$\mathbb{P}\{\mathbf{x}_t \leq -10 \text{ for some } t\} \tag{2}$$

are well-defined by the joint probability distribution $\mathbb{P}$

IID sequence is a stochastic process; structure of IID sequences makes them useful in statistical settings

Law of large numbers tells us sample average is a good approximation of expectation

Not all processes IID. Are there restrictions we can put on stochastic processes so they resemble IID sequences in *some way*?

Stationarity

A stochastic process $\{\mathbf{x}_t\}_{t \geq 0}$ is called **stationary** if the distribution of any subset of these random vectors is unaffected by shifting forward in time

$$\mathcal{L}(\mathbf{x}_{t_1}, \ldots, \mathbf{x}_{t_k}) = \mathcal{L}(\mathbf{x}_{t_1+m}, \ldots, \mathbf{x}_{t_k+m}) \tag{3}$$

for any $m \in \mathbb{N}$ and any sequence of integers $t_1, \ldots, t_k$

Stationarity preserves ID part of IID

Example. An IID process $\{\mathbf{x}_t\}$ is stationary, since independence implies that the laws on both sides of (3) are just k products of the marginal $\mathcal{L}(\mathbf{x}_1)$

Example. Let x_1 be a draw from some arbitrary law P on $\mathbb{R}$ and let $x_{t+1} = x_t$ for all t. For any collection of integers $t_1, \ldots, t_k$ and any Borel sets $B_1, \ldots, B_k$, we have

$$\mathbb{P}\{x_{t_1} \in B_1, \ldots, x_{t_k} \in B_k\} = \mathbb{P}\{x_1 \in \cap_{i=1}^k B_k\} = P(\cap_{i=1}^k B_k)$$

If we replace each x_{t_i} with x_{t_i+m}, we get the same number. Hence $\{x_t\}$ is stationary

Example. A **random walk** on $\mathbb{R}$ is a stochastic process $\{x_t\}$ where $x_t = \sum_{j=1}^{t} w_j$ for some IID zero-mean process $\{w_j\}$. Let $\sigma^2 := \operatorname{var} w_t$. If $\sigma > 0$, then $\{x_t\}$ is not stationary, since $\operatorname{var} x_t = t\sigma^2$. In particular, $\mathcal{L}(x_t)$ depends on t

Stationarity alone isn't strong enough to give us properties like the LLN and CLT

Example. For the process $\{x_t\}$ in example 7, the LLN fails whenever P is nondegenerate. Indeed $\bar{x}_T := \frac{1}{T} \sum_{t=1}^{T} x_t = x_1$, and hence $\mathcal{L}(\bar{x}_T) = \mathcal{L}(x_1) = P$ for all t. In particular, $\bar{x}_T$ does not converge in probability to any constant

Ergodicity

A stationary stochastic process $\{\mathbf{x}_t\}$ on $\mathbb{R}^K$ is **ergodic** if the LLN holds; that is, if

$$\frac{1}{T}\sum_{t=1}^{T} h(\mathbf{x}_t) \xrightarrow{p} \mu_h := \mathbb{E}\, h(\mathbf{x}_t) \quad \text{as} \quad T \to \infty \tag{4}$$

for any $\mathscr{B}$-measurable $h\colon \mathbb{R}^N \to \mathbb{R}$ such that $\mathbb{E}\,|h(\mathbf{x}_t)| < \infty$

Informally, for an ergodic process, cross-sectional and time series averages coincide

Example. Suppose $\mathbf{x}_t = x_t$ is a binary variable with $x_t = 1$ if an individual is employed and zero otherwise. The term $\frac{1}{T}\sum_{t=1}^{T} x_t$ is the fraction of time spent employed over $1, \ldots, T$.

If the employment process is ergodic, then $\frac{1}{T}\sum_{t=1}^{T} x_t \to \mathbb{E}\, x_t = \mathbb{P}\{x_t = 1\}$.

Suppose we observe time t employment outcomes $x_t^1, \ldots, x_t^N$ of individuals $1, \ldots, N$ who all follow the same model and are sufficiently independent. Then the cross sectional average $\frac{1}{N}\sum_{n=1}^{N} x_t^n$ will be close to $\mathbb{E}\, x_t = \mathbb{P}\{x_t = 1\}$.

Stochastic Recursive Sequences

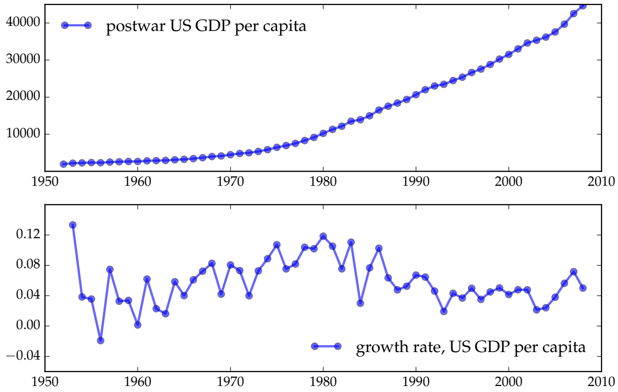


Figure: Postwar US GDP per capita. Source: Penn World Tables

The scalar Gaussian AR(1) model can come close to the time-series of US GDP growth rates

$$x_{t+1} = b + a x_t + c w_{t+1} \tag{5}$$

with $\{w_t\} \overset{\text{IID}}{\sim} \text{N}(0,1)$ and x_0 given

Here a, b, and c are parameters, while x_t is called the **state variable**

Equation for Gaussian AR(1) process an example of **stochastic difference equation**

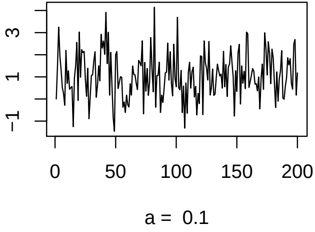
The process $\{x_t\}$ it defines is called a stochastic process or **stochastic recursive sequence**

A realization of the process is called a **time series**

The dynamics of $\{x_t\}$ depends on the parameters:

- If a is outside the interval $(-1, 1)$, the series tend to diverge
- If $|a| < 1$, then the opposite is true

For example, for $a = 0.9$, after an initial burn in period where the series is affected by the initial condition x_0, the process settles down to random motion within a band (between about 5 and 15 in this case)



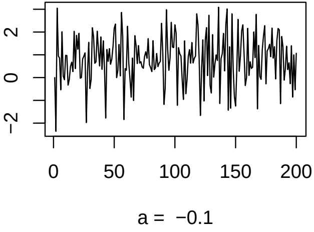


Figure: Dynamics of the linear AR(1) model

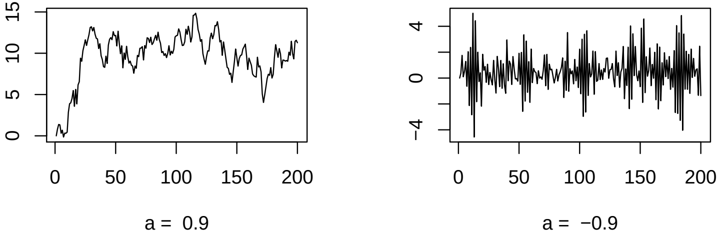


Figure: Dynamics of the linear AR(1) model

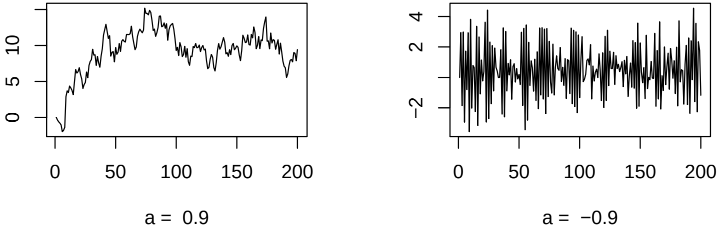


Figure: Dynamics of the linear AR(1) model

Generalise the Gaussian AR(1) model by removing assumption that shocks are normal – simply called AR(1) process

Generalize the scalar AR(1) model to $\mathbb{R}^K$, yielding the vector AR(1) model, or VAR(1) model

$$\mathbf{x}_{t+1} = \mathbf{b} + \mathbf{A}\mathbf{x}_t + \mathbf{C}\mathbf{w}_{t+1} \tag{6}$$

The sequence $\{\mathbf{w}_t\}$ is IID and satisfies $\mathbb{E}\,\mathbf{w}_t = \mathbf{0}$ and $\mathbb{E}\,[\mathbf{w}_t \mathbf{w}_t'] = \mathbf{I}$

If $\mathbf{w}_t$ is multivariate normal, then (6) is called the Gaussian VAR(1).

The vector $\mathbf{x}_t$ is called the **state vector**

AR(p) model where x_{t+1} is a linear function of previous p states

AR(2) process has dynamics

$$x_{t+1} = b + ax_t + \gamma x_{t-1} + w_{t+1}$$

We can reformulate AR(2) as a first-order model

Define y_t via $y_t = x_{t-1}$; the process can be expressed as

$$x_{t+1} = b + ax_t + \gamma y_t + w_{t+1}$$

$$y_{t+1} = x_t$$

and in matrix form as

$$\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix} = b \begin{pmatrix} 1 \\ 0 \end{pmatrix} + \begin{pmatrix} a & \gamma \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x_t \\ y_t \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} w_{t+1}$$

This is a special case of the VAR(1) model in (6)

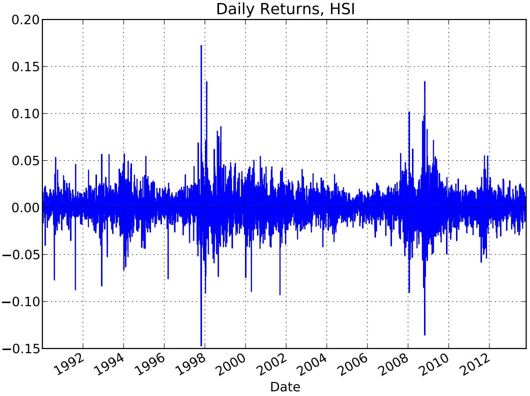


Figure: Daily returns on the Hang Seng Index. Source: Yahoo! Finance

Bursts of volatility in the Hang-Seng Index difficult to replicate without *non-linearities*

Need to model time-varying volatility in asset returns: consider pth order **autoregressive conditional heteroskedasticity model** (ARCH(p) model)

$$x_{t+1} = (\alpha_0 + \alpha_1 x_t^2)^{1/2} w_{t+1} \quad \text{with} \quad \{w_t\} \overset{\text{IID}}{\sim} \mathrm{N}(0,1) \tag{7}$$

and $\alpha_0 > 0$, $\alpha_1 \geq 0$

Model contains time varying volatility $\sigma_{t+1}^2 = \alpha_0 + \alpha_1 x_t^2$

Still better fits to asset returns data can be obtained via
Generalized ARCH models

The GARCH(1,1) process

$$x_t = \sigma_t w_t$$

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 x_t^2 + \alpha_2 \sigma_t^2$$

Next period volatility depends on its own lagged state as well as x_t

Markov Processes

Processes we have studied so far in this lecture are Markov processes

Let $S \subset \mathbb{R}^K$ and let $\mathscr{B}(S)$ be the Borel subsets of S

The primitive of a discrete time Markov process on S is a **stochastic kernel** or **transition probability function** $Q \colon S \times \mathscr{B}(S) \to [0, 1]$ such that

1. $Q(\mathbf{s}, \cdot)$ is a probability measure over $\mathscr{B}(S)$ for all $\mathbf{s} \in S$ and
2. $g(\mathbf{s}) := Q(\mathbf{s}, B)$ is $\mathscr{B}$-measurable for each $B \in \mathscr{B}(S)$

S is called the **state space** of the model

Q and some initial distribution P_0 generate a Markov process $\{\mathbf{x}_t\}$ via

1: draw $\mathbf{x}_0$ from P_0
2: **for** t in $0, 1, 2, \ldots$ **do**
3: draw $\mathbf{x}_{t+1}$ from the distribution $Q(\mathbf{x}_t, \cdot)$
4: **end for**

The generated sequence $\{\mathbf{x}_t\}$ also called a **sample path** for Q

The canonical stochastic difference equation for a **first-order Markov process** takes the form

$$\mathbf{x}_{t+1} = G(\mathbf{x}_t, \mathbf{w}_{t+1}) \quad \text{with} \quad \mathcal{L}(\mathbf{x}_0) = P_0 \tag{8}$$

The sequence of $\mathbb{R}^M$-valued shocks $\{\mathbf{w}_t\}_{t\geq 1}$ is IID has common distribution Ψ

G is a given $\mathscr{B}$-measurable function

Initial condition $\mathbf{x}_0$ and the shocks $\{\mathbf{w}_t\}_{t\geq 1}$ also independent

Stochastic difference equation specifies the process $\{x_t\}$ and stochastic kernel

$$Q(\mathbf{s}, B) = \mathbb{P}\{G(\mathbf{s}, \mathbf{w}_{t+1}) \in B\} \qquad (\mathbf{s} \in S,\ B \in \mathscr{B}(S))$$

Recalling that Ψ is the distribution of $\mathbf{w}_{t+1}$, the stochastic kernel can also be written as

$$Q(\mathbf{s}, B) = \Psi\left\{\mathbf{w} \in \mathbb{R}^M \ : \ G(\mathbf{s}, \mathbf{w}) \in B\right\}$$

Example. For the ARCH(1) model of (7), the stochastic kernel is

$$Q(s, B) = \mathbb{P}\{(\alpha_0 + \alpha_1 s^2)^{1/2} w_{t+1} \in B\}$$

when w_{t+1} is standard normal

Repeated substitution gives

$$\mathbf{x}_1 = G(\mathbf{x}_0, \mathbf{w}_1)$$

$$\mathbf{x}_2 = G(G(\mathbf{x}_0, \mathbf{w}_1), \mathbf{w}_2)$$

$$\mathbf{x}_3 = G(G(G(\mathbf{x}_0, \mathbf{w}_1), \mathbf{w}_2), \mathbf{w}_3)$$

and so on...

The state vector $\mathbf{x}_t$ can be written as a function of $\mathbf{x}_0$ and the shocks $\mathbf{w}_1, \ldots, \mathbf{w}_t$ for any t; there exists a function H_t such that

$$\mathbf{x}_t = H_t(\mathbf{x}_0, \mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_t) \tag{9}$$

Example. For the scalar linear AR(1) process (5) we can write the right-hand side of (9) explicitly as

$$x_t = b \sum_{k=0}^{t-1} a^k + \sum_{k=0}^{t-1} a^k c w_{t-k} + a^t x_0 \tag{10}$$

This is called the **moving average representation** of x_t

Fact. (7.2.1) For the process (8), the pair $\mathbf{x}_t$ and $\mathbf{w}_{t+k}$ are independent for all $k \geq 1$

If $Q(\mathbf{s}, \cdot)$ is absolutely continuous for all $\mathbf{s}$, then define the corresponding **stochastic density kernel** or the **transition density**

$$q(\mathbf{s}, \cdot) := \text{ the density of } Q(\mathbf{s}, \cdot) \text{ for all } \mathbf{s} \in S$$

The function $q(\mathbf{s}, \cdot)$ is the conditional density of $\mathbf{x}_{t+1}$ given $\mathbf{x}_t = \mathbf{s}$

Heuristically, $q(\mathbf{s}, \mathbf{s}') \, \mathrm{d}\mathbf{s}'$ is the probability of transitioning from $\mathbf{s}$ to $\mathbf{s}'$ in one step

Example. Recall the ARCH(1) process, with stochastic kernel

$$Q(s, B) = \mathbb{P}\{(\alpha_0 + \alpha_1 s^2)^{1/2} w_{t+1} \in B\}$$

The transition density $q(s, \cdot)$ is the density of $y = (\alpha_0 + \alpha_1 s^2)^{1/2} w_{t+1}$ when $\mathcal{L}(w_{t+1}) = \mathrm{N}(0, 1)$. Hence

$$q(s, s') = \frac{1}{\sqrt{2\pi\sigma_s^2}} \exp\left\{-\frac{(s')^2}{2\sigma_s^2}\right\} \quad \text{where} \quad \sigma_s^2 := \alpha_0 + \alpha_1 s^2$$

Consider the generic process $\mathbf{x}_{t+1} = G(\mathbf{x}_t, \mathbf{w}_{t+1})$, where $\mathcal{L}(\mathbf{x}_0) = P_0$ and $\{\mathbf{w}_t\}$ is IID on $\mathbb{R}^M$ with common distribution Ψ

The **marginal distribution of the state vector** is $\mathcal{L}(\mathbf{x}_t)$

Since $\mathbf{x}_t$ is a well-defined random vector, $\mathcal{L}(\mathbf{x}_t)$ is also well-defined, and we denote it by

$$P_t(B) = \mathbb{P}\{\mathbf{x}_t \in B\} \qquad (B \in \mathscr{B}(S))$$

Example. Consider the Gaussian AR(1) process

Suppose $\mathcal{L}(x_0) = \mathrm{N}(\mu_0, \sigma_0^2)$, where μ_0 and σ_0 are given constants and x_0 is independent of the shock process $\{w_t\}$

Combining the moving average representation of x_t we have $P_t = \mathrm{N}\left(\mu_t, \sigma_t^2\right)$

$$\text{where} \quad \mu_t := b \sum_{k=0}^{t-1} a^k + a^t \mu_0 \;\; \text{and} \;\; \sigma_t^2 := \sum_{k=0}^{t-1} a^{2k} c^2 + a^{2t} \sigma_0^2$$

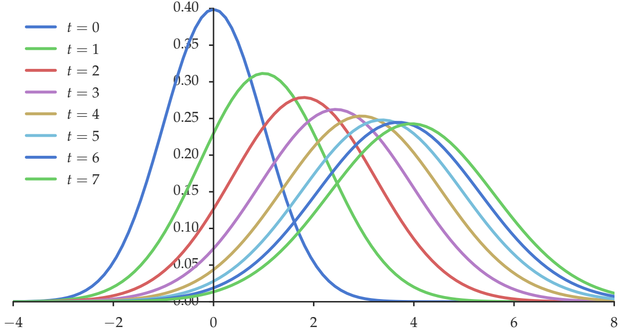


Figure: Sequence of marginal densities for the Gaussian AR(1) process when $\mu_0 = 0$, $\sigma_0 = b = c = 1$ and $a = 0.8$

For general Markov processes, tracking evolution of first two moments is more complicated

Fact. (7.2.2) The marginal distributions of a Markov process with stochastic kernel Q obey the recursion

$$P_{t+1}(B) = \int Q(\mathbf{s}, B) P_t(\mathrm{d}\mathbf{s}) \qquad (B \in \mathscr{B}(S),\ t \geq 0) \tag{11}$$

Fact. (7.2.3) If $Q(\mathbf{s}, \cdot)$ is absolutely continuous for all $\mathbf{s} \in S$, then P_t is absolutely continuous for all $t \geq 1$. Letting p_t be the corresponding density, the sequence $\{p_t\}$ satisfies

$$p_{t+1}(\mathbf{s}') = \int q(\mathbf{s}, \mathbf{s}') p_t(\mathbf{s}) \, \mathrm{d}\mathbf{s} \qquad (B \in \mathscr{B}(S), \ t \geq 1) \tag{12}$$

where q is the transition density corresponding to Q

Fact. (7.2.4) Let $\{\mathbf{x}_t\}$ be a Markov process with transition density q and initial density p_0. Under the conditions of fact 7.2.3 in ET, the joint density p_T of $\mathbf{x}_0, \ldots, \mathbf{x}_T$ is

$$p_T(\mathbf{s}_0, \ldots, \mathbf{s}_T) = p_0(\mathbf{s}_0) \prod_{t=0}^{T-1} q(\mathbf{s}_t, \mathbf{s}_{t+1}) \tag{13}$$

Proof. First, factor the joint density into

$$p_T(\mathbf{s}_0, \ldots, \mathbf{s}_T) = p_0(\mathbf{s}_0) \prod_{t=0}^{T-1} p(\mathbf{s}_{t+1} \mid \mathbf{s}_t, \ldots, \mathbf{s}_0)$$

This is the expression for the joint density of a general stochastic process. With first order Markov property, $p(\mathbf{s}_{t+1} \mid \mathbf{s}_t, \ldots, \mathbf{s}_0)$ reduces to $p(\mathbf{s}_{t+1} \mid \mathbf{s}_t) = q(\mathbf{s}_t, \mathbf{s}_{t+1})$

Stationarity of Markov Processes

Marginals $\{P_t\}$ of Gaussian AR(1) process can change over time, but difference between successive distributions diminishing over time

Recall $P_t = \mathrm{N}(\mu_t, \sigma_t^2)$, where μ_t and σ_t^2. Provided $|a| < 1$, sequences converge

$$\mu_t \to \mu_\infty := \frac{b}{1-a} \quad \text{and} \quad \sigma_t^2 \to \sigma_\infty^2 := \frac{c^2}{1-a^2}$$

Let P_∞ denote the limiting distribution

$$P_\infty = \mathrm{N}(\mu_\infty, \sigma_\infty^2) = \mathrm{N}\left[\frac{b}{1-a}, \frac{c^2}{1-a^2}\right] \tag{14}$$

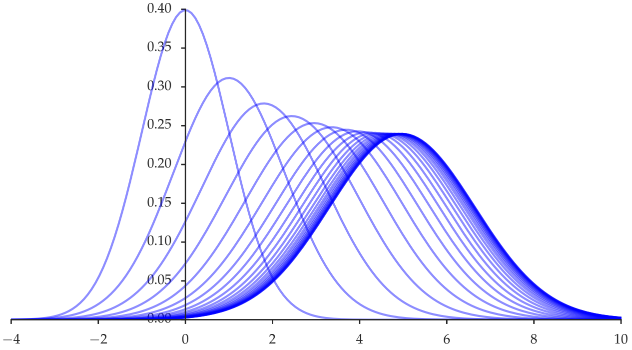


Figure: Convergence of marginal densities, Gaussian AR(1) case

Limiting distribution is an example of a stationary distribution

A **stationary distribution** for a Markov process on S with stochastic kernel Q is any distribution P_∞ on S satisfying

$$P_\infty(B) = \int Q(\mathbf{s}, B) P_\infty(\mathrm{d}\mathbf{s}) \qquad \text{for all } B \in \mathscr{B}(S)$$

Comparing with the recursion $P_{t+1}(B) = \int Q(\mathbf{s}, B) P_t(\mathrm{d}\mathbf{s})$, the definition of stationarity of P_∞ means precisely

$$\mathcal{L}(\mathbf{x}_t) = P_\infty \implies \mathcal{L}(\mathbf{x}_{t+1}) = P_\infty$$

Any process starting at P_∞ will be IID

Fact. (7.2.5) Let $\{\mathbf{x}_t\}$ be a Markov process with stationary distribution P_∞. If $\mathcal{L}(\mathbf{x}_0) = P_\infty$, then $\{\mathbf{x}_t\}$ is a stationary stochastic process, with $\mathcal{L}(\mathbf{x}_t) = P_\infty$ for all t

Example. Consider the scalar Gaussian linear AR(1) process from (5). If $|a| < 1$ and $\mathcal{L}(x_0) = P_\infty$, where P_∞ is as in (14), then $\{x_t\}$ is stationary

Fact. (7.2.6) If $Q(\mathbf{s}, \cdot)$ is absolutely continuous for all $\mathbf{s} \in S$ with transition density $q(\mathbf{s}, \cdot)$, then every stationary distribution P_∞ is absolutely continuous, and its density p_∞ satisfies

$$p_\infty(\mathbf{s}') = \int q(\mathbf{s}, \mathbf{s}') p_\infty(\mathbf{s}) \, \mathrm{d}\mathbf{s} \quad \text{for all } \mathbf{s}' \in S \tag{15}$$

Not every Markov process has a stationary distribution

Recall linear scalar AR(1) model $x_{t+1} = b + ax_t + cw_{t+1}$

- the marginal variance evolves according to $\sigma^2_{t+1} = a^2\sigma^2_t + c^2$

If $|a| \geq 1$, then this sequence diverges since the variance is not constant

A sufficient condition for existence

The condition we give slightly specializes the famous Krylov–Bogolyubov theorem

It uses the notion of a **coercive** function, which is any nonnegative function V on the state space S such that

$$C_\gamma := \{\mathbf{s} \in S : V(\mathbf{s}) \leq \gamma\}$$

is a compact subset of S for all $\gamma \in \mathbb{R}$

Example. If $S = \mathbb{R}^K$ and $V(\mathbf{s}) = \|\mathbf{s}\|$, then V is coercive on S. Indeed, for this function, each C_γ is a closed sphere centered on the origin. Closed and bounded subsets of $\mathbb{R}^K$ are compact.

Example. If $S = \mathbb{R}$ and $V(s) = s^2$, then V is coercive on S. To see this, just observe that $C_\gamma = [-\sqrt{\gamma}, \sqrt{\gamma}]$. This is a closed, bounded subset of S.

Theorem. (7.2.1) Let $\mathbf{x}_{t+1} = G(\mathbf{x}_t, \mathbf{w}_{t+1})$ be the process on S defined in (8). If

1. $\mathbf{s} \mapsto G(\mathbf{s}, \mathbf{w})$ is continuous for each fixed $\mathbf{w} \in \mathbb{R}^M$, and
2. there exists a coercive function V on S and constants λ and L such that $0 \leq \lambda < 1$ and

$$\mathbb{E} V[G(\mathbf{s}, \mathbf{w}_{t+1})] \leq \lambda V(\mathbf{s}) + L \qquad (\mathbf{s} \in S) \tag{16}$$

then the Markov process has at least one stationary distribution.

Condition (16) is an example of a **drift condition**. One way to understand it is to write it as

$$\frac{\mathbb{E}\, V[G(\mathbf{s}, \mathbf{w}_{t+1})]}{V(\mathbf{s})} \leq \lambda + \frac{L}{V(\mathbf{s})}$$

As we move "away from the center" of the state space, $V(\mathbf{s})$ becomes large, and the term on the right is eventually < 1

Looking at the left hand side, the value attached to the state by V is expected to decrease, suggesting movement away from the "edges" of the state space and back towards the center

Example. Consider again the ARCH(1) process $x_{t+1} = (\alpha_0 + \alpha_1 x_t^2)^{1/2} w_{t+1}$ from (7), where, as usual $\alpha_0 > 0$ and $\alpha_1 \geq 0$. Let $S = \mathbb{R}$. If $\alpha_1 < 1$, then this process has a stationary distribution on S. To see this, observe that

$$G(s, w) = (\alpha_0 + \alpha_1 s^2)^{1/2} w \qquad (s, w \in \mathbb{R})$$

is continuous in s for each fixed w. Moreover, we showed $V(s) = s^2$ is coercive on S, and

$$\mathbb{E}\, G(s, w_{t+1})^2 = (\alpha_0 + \alpha_1 s^2) \mathbb{E}\, w_{t+1}^2 = \alpha_0 + \alpha_1 s^2 \qquad (17)$$

Hence (16) holds with equality when $V(s) := s^2$, $\lambda := \alpha_1$ and $L := \alpha_0$

Theorem. (7.2.2) Let $\{\mathbf{x}_t\}$ be a Markov process on S with stochastic kernel Q. Suppose $Q(\mathbf{s}, \cdot)$ is absolutely continuous for all $\mathbf{s} \in S$ and let $q(\mathbf{s}, \cdot)$ be the corresponding transition density. If

(a) q is strictly positive and continuous on $S \times S$, and

(b) there exists a coercive function V on S and constants λ and L such that $0 \leq \lambda < 1$ and

$$\int V(\mathbf{s}')q(\mathbf{s}, \mathbf{s}')\, \mathrm{d}\mathbf{s}' \leq \lambda V(\mathbf{s}) + L \qquad (\mathbf{s} \in S)$$

then the following statements are true (continued on the next slide)

Theorem. (7.2.2)(cont.)

1. Q has a unique stationary distribution P_∞ with density p_∞.
2. $P_t \xrightarrow{w} P_\infty$ as $t \to \infty$ for all initial P_0.
3. If $h \colon S \to \mathbb{R}$ is $\mathscr{B}$-measurable and $\int |h(\mathbf{s})| p_\infty(\mathbf{s})\,\mathrm{d}\mathbf{s} < \infty$, then

$$\frac{1}{T}\sum_{t=1}^{T} h(\mathbf{x}_t) \xrightarrow{p} \int h(\mathbf{s}) p_\infty(\mathbf{s})\,\mathrm{d}\mathbf{s} \quad \text{as} \quad T \to \infty \tag{18}$$

4. If, in addition, $h^2 \leq V$, then there exists a $\sigma_h^2 \geq 0$ such that

$$\sqrt{T}\left\{\frac{1}{T}\sum_{t=1}^{T} h(\mathbf{x}_t) - \int h(\mathbf{s}) p_\infty(\mathbf{s})\,\mathrm{d}\mathbf{s}\right\} \xrightarrow{d} \mathrm{N}(0, \sigma_h^2) \tag{19}$$

as $\quad T \to \infty$

Fact. (7.2.7) Let $\mathbf{x}_{t+1} = g(\mathbf{x}_t) + \mathbf{w}_{t+1}$ be the additive shock process on $\mathbb{R}^K$ from (7.12). If $\mathbb{E}\,\|\mathbf{w}_t\|$ is finite, g is continuous, the density ψ of $\mathbf{w}_t$ is continuous and positive everywhere on $\mathbb{R}^K$, and there exist constants λ and L such that $0 \leq \lambda < 1$ and

$$\|g(\mathbf{s})\| \leq \lambda \|\mathbf{s}\| + L \quad \text{for all } \mathbf{s} \in \mathbb{R}^K \tag{20}$$

then the conditions of theorem 7.2.2 are satisfied.

Proof. Recall the transition density satisfies $q(\mathbf{s}, \mathbf{s}') = \psi(\mathbf{s}' - g(\mathbf{s}))$

Since ψ is assumed continuous and positive on $\mathbb{R}^K$, condition (a) in theorem 7.2.2 is satisfied

Condition (b) is also satisfied because, taking λ and L as in (20)

$$\int \|\mathbf{s}'\| q(\mathbf{s}, \mathbf{s}') \, \mathrm{d}\mathbf{s}'$$

$$= \mathbb{E} \, \|g(\mathbf{s}) + \mathbf{w}_{t+1}\| \leq \lambda \|\mathbf{s}\| + L + \int \|\mathbf{w}\| \psi(\mathbf{w}) \, \mathrm{d}\mathbf{w}$$

Example. Gaussian VAR(1)

$$\mathbf{x}_{t+1} = \mathbf{b} + \mathbf{A}\mathbf{x}_t + \mathbf{C}\mathbf{w}_{t+1} \tag{21}$$

In this case, $g(\mathbf{s}) = \mathbf{a} + \mathbf{A}\mathbf{s}$, and

$$\|g(\mathbf{s})\| = \|\mathbf{b} + \mathbf{A}\mathbf{s}\| \leq \|\mathbf{b}\| + \|\mathbf{A}\mathbf{s}\| \leq \|\mathbf{b}\| + \|\mathbf{A}\|\|\mathbf{s}\|$$

Thus, we can apply fact 7.2.7 to obtain asymptotic stability and LLN and CLT results whenever the matrix norm of $\mathbf{A}$ satisfies $\|\mathbf{A}\| < 1$

We have a better result

Fact. (7.2.8) For the VAR process (21), the conclusions of theorem 7.2.2 hold with $V = \| \cdot \|$ whenever $\varrho(\mathbf{A}) < 1$

Here $\varrho(\mathbf{A})$ is the spectral radius of $\mathbf{A}$.

Taking expectations across (21), the mean $\boldsymbol{\mu}_\infty$ must satisfy $\boldsymbol{\mu}_\infty = \mathbf{b} + \mathbf{A}\boldsymbol{\mu}_\infty$.

Apply the Neumann series lemma (page 69 in ET)

$$\boldsymbol{\mu}_\infty = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{b} = \sum_{i=0}^{\infty} \mathbf{A}^i \mathbf{b}$$

Let $\mathbf{\Sigma}_\infty$ be the asymptotic variance (i.e., the variance–covariance matrix of the stationary distribution); $\mathbf{\Sigma}_\infty$ must satisfy

$$\mathbf{\Sigma}_\infty = \mathbf{A}\mathbf{\Sigma}_\infty\mathbf{A}^\intercal + \mathbf{C}\mathbf{C}^\intercal$$

This is an example of a **discrete time Lyapunov equation**

Under the condition $\varrho(\mathbf{A}) < 1$, we can be solve by iteration or by using existing numerical solvers

```
using QuantEcon

A = [0.8 -0.2;
    -0.1 0.7]
C = [0.5 0.4;
     0.4 0.6]

solve_discrete_lyapunov(A, C * C')
```

Martingales

A stochastic process evolving over time such that best guess of next value given current value is current value

Let $\{\mathscr{F}_t\}$ be a sequence of information sets

The sequence $\{\mathscr{F}_t\}$ is called a **filtration** if $\mathscr{F}_t \subset \mathscr{F}_{t+1}$ for all t

Intuition: reflects idea that more information is revealed over time

Example. Let $\{\mathbf{x}_t\}$ be a sequence of random vectors, and let

$$\mathscr{F}_0 := \emptyset, \;\; \mathscr{F}_1 := \{\mathbf{x}_1\}, \;\; \mathscr{F}_2 := \{\mathbf{x}_1, \mathbf{x}_2\}, \;\; \cdots$$

Then $\{\mathscr{F}_t\}$ is a filtration called the filtration generated by $\{\mathbf{x}_t\}$

Let

- $\{m_t\}$ be a sequence of RVs (scalar stochastic process)
- $\{\mathscr{F}_t\}$ be a filtration

We say $\{m_t\}$ is **adapted** to the filtration $\{\mathscr{F}_t\}$ if m_t is $\mathscr{F}_t$-measurable for every t

Intuition: If we know $\mathscr{F}_t$ then we know m_t

Example. If $\{\mathscr{F}_t\}$ is the filtration defined by

$$\mathscr{F}_0 := \emptyset, \quad \mathscr{F}_1 := \{x_1\}, \quad \mathscr{F}_2 := \{x_1, x_2\}, \quad \mathscr{F}_3 := \{x_1, x_2, x_3\}, \quad \cdots$$

and $m_t := t^{-1} \sum_{j=1}^{t} x_j$, then $\{m_t\}$ is adapted to $\{\mathscr{F}_t\}$

Fact. (7.3.1) If $\{m_t\}$ is adapted to filtration $\{\mathscr{F}_t\}$, then $\mathbb{E}[m_t \mid \mathscr{F}_{t+j}] = m_t$ for any $j \geq 0$

Proof. We know

- m_t is $\mathscr{F}_t$-measurable
- If $j \geq 0$, then $\mathscr{F}_t \subset \mathscr{F}_{t+j}$
- If $\mathcal{G} \subset \mathcal{H}$, then $\mathcal{G}$-measurable implies $\mathcal{H}$-measurable
- If y is $\mathcal{H}$-measurable, then $\mathbb{E}[y \mid \mathcal{H}] = y$

The result follows

Let $\{m_t\}$ be a sequence of random variables

- adapted to a filtration $\{\mathscr{F}_t\}$
- having finite first moment

We say that $\{m_t\}$ is a **martingale** with respect to $\{\mathscr{F}_t\}$ if

$$\mathbb{E}\left[m_{t+1} \,|\, \mathscr{F}_t\right] = m_t \quad \text{for all} \quad t$$

We can simplify the expression $\mathbb{E}\left[m_{t+1} \,|\, \mathscr{F}_t\right]$ to $\mathbb{E}_t[x_{t+1}]$ when the filtration is understood

Example. Let $\{\eta_t\}$ be IID with $\mathbb{E}[\eta_1] = 0$ and define $m_t := \sum_{j=1}^t \eta_j$ and let $\mathscr{F}_t := \{\eta_1, \ldots, \eta_t\}$

The sequence $\{m_t\}$ is a martingale with respect to $\{\mathscr{F}_t\}$

From the definitions, $\{m_t\}$ will be adapted to $\{\mathscr{F}_t\}$. Moreover

$$\mathbb{E}[m_{t+1} \mid \mathscr{F}_t] = \mathbb{E}[m_t + \eta_{t+1} \mid \mathscr{F}_t] = m_t + \mathbb{E}[\eta_{t+1} \mid \mathscr{F}_t]$$

Example. Consider an Euler equation of the form

$$u'(c_t) = \mathbb{E}_t \left[\frac{1 + r_{t+1}}{1 + \varrho} \cdot u'(c_{t+1}) \right]$$

Where u is a utility function, $r_t =$ the interest rate and $\varrho =$ a discount factor. $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot \mid \mathscr{F}_t]$, where $\mathscr{F}_t =$ information set at t

Specializing to $r_{t+1} = \varrho$ and $u(c) = c - ac^2/2$

$$c_t = \mathbb{E}_t[c_{t+1}] =: \mathbb{E}[c_{t+1} \mid \mathscr{F}_t]$$

Here, consumption is a martingale with respect to $\{\mathscr{F}_t\}$

A stochastic process $\{d_t\}$ is called a **martingale difference sequence** (MDS) with respect to $\{\mathscr{F}_t\}$ if

$$\mathbb{E}\left[d_{t+1} \,|\, \mathscr{F}_t\right] = 0 \quad \text{for all} \quad t$$

If $\{m_t\}$ is a martingale with respect to $\{\mathscr{F}_t\}$, then $d_t = m_t - m_{t-1}$ is an MDS with respect to $\{\mathscr{F}_t\}$

Fact. (7.3.2) If $\{d_t\}$ is an MDS, then $\mathbb{E}\, d_t = 0$ for all t

Martingale Difference LLN and CLT

MDS are good candidates for LLN / CLT because...

Fact. (7.3.3) If $\{m_t\}$ is an MDS, then $\mathrm{cov}[m_{t+j}, m_t] = 0$ for all $j \geq 1$.

Proof. Let $\{m_t\}$ be an MDS w.r.t. filtration $\{\mathscr{F}_t\}$ and fix $t \geq 0$ and $j \geq 1$. We have

$$\mathrm{cov}[m_{t+j}, m_t] = \mathbb{E}\left[m_{t+j} m_t\right] = \mathbb{E}\left[\mathbb{E}\left[m_{t+j} m_t \,|\, \mathscr{F}_{t+j-1}\right]\right]$$

Since $t + j - 1 \geq t$ and $\{\mathscr{F}_t\}$ is a filtration,

$$\mathbb{E}\left[\mathbb{E}\left[m_{t+j} m_t \,|\, \mathscr{F}_{t+j-1}\right]\right] = \mathbb{E}\left[m_t \mathbb{E}\left[m_{t+j} \,|\, \mathscr{F}_{t+j-1}\right]\right] = \mathbb{E}\left[m_t \cdot 0\right] = 0$$

Theorem. (7.3.1) If $\{m_t\}$ is a stationary MDS with respect to some filtration $\{\mathscr{F}_t\}$, then

$$\frac{1}{T}\sum_{t=1}^{T} m_t \xrightarrow{p} 0 \quad \text{as} \quad T \to \infty$$

If, in addition, $\gamma^2 := \mathbb{E}[m_t^2] < \infty$ and $\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[m_t^2 \mid \mathscr{F}_{t-1}] \xrightarrow{p} \gamma^2$ as $T \to \infty$, then

$$T^{-1/2}\sum_{t=1}^{T} m_t \xrightarrow{d} \mathrm{N}(0, \gamma^2) \quad \text{as} \quad T \to \infty$$

Simulation

Modern scientific computing provides good routines for simulating independent observations from common distributions

But when distribution is non-standard or has no closed form solution, we have to implement our own routines

We look at:

- Inverse transforms
- Markov Chain Monte-Carlo (MCMC)

Inverse Transforms

Aim is to draw from a distribution on $\mathbb{R}$ with CDF F

The **inverse transform** method involves:

1: draw u from the uniform distribution on $[0, 1]$
2: return $F^{-1}(u)$, where F^{-1} is the quantile function of F (see (4.35) in ET)

See ET page 200 for a proof

Here's one extension; we want to simulate from a joint distribution on $\mathbb{R}^N$ defined in terms of a copula (see §5.1.5 in ET)

In particular, assume F_n is a continuous CDF on $\mathbb{R}$ for $n = 1, \ldots, N$, C is a copula on $\mathbb{R}^N$ and that the joint distribution of interest is

$$F(s_1, \ldots, s_N) = C(F_1(s_1), \ldots, F_N(s_N))$$

(This is (5.22) from page 138 in ET)

If we can simulate from C, then we can simulate from F

1: draw $u_1, \ldots, u_N$ from C
2: return $(x_1, \ldots, x_N) := (F_1^{-1}(u_1), \ldots, F_N^{-1}(u_N))$

Exercise 7.6.13 asks for a proof

Markov Chain Monte Carlo (MCMC)

Markov chain Monte Carlo (MCMC) is a way of simulating from a given density π on $S \subset \mathbb{R}^K$

The idea is to construct a stochastic kernel P on S such that

1. π is a stationary distribution for P, and
2. P is sufficiently ergodic that its sample path averages converge to expectations under π

A clearer statement of 2. is

$$\frac{1}{T}\sum_{t=1}^{T} h(\mathbf{x}_t) \xrightarrow{p} \int h(\mathbf{s})\pi(\mathbf{s})\,\mathrm{d}\mathbf{s} \quad \text{as} \quad T \to \infty$$

where $\{\mathbf{x}_t\}$ is a process with kernal P — this is (18) from Theorem 7.2.2

We consider the Metropolis–Hastings method for MCMC (the other important method is the Gibbs sampler)

Start with a Markov process in the form of a stochastic density kernel $q = q(s, s')$ called the **proposal density**

Draws from the proposal density are called **proposals**

Each time we draw a proposal, we either

- accept it and move to that new state
- reject it and stay put

The probability of accepting is structured so that the chain tends to stay in regions where π puts most probability mass

For a sequence $\{x_t\}$ generated by this process:

$$\text{fraction of time spent in } B = \frac{1}{T}\sum_{t=1}^{T} \mathbb{1}\{x_t \in B\} \approx \pi(B)$$

for large T This is a version of (18) from Theorem 7.2.2 with $h = \mathbb{1}_B$

Consider first an accept-reject style process where an arbitrary function of the current state x_t and proposal y, $\alpha = \alpha(x_t, y)$, gives the probability of acceptance

The function $\alpha = \alpha(x_t, y)$ takes values in $[0, 1]$

Algorithm for generating x_{t+1} from x_t given α and proposal density q:

draw y from $q(x_t, \cdot)$
draw u independently from the uniform $[0, 1]$ distribution
if $u \leq \alpha(x_t, y)$ **then** ▷ with probability $\alpha(x_t, y)$
    set $x_{t+1} = y$
else ▷ with probability $1 - \alpha(x_t, y)$
    set $x_{t+1} = x_t$
end if

The stochastic kernel P associated with $\{x_t\}$ in algorithm has the form

$$P(s, B) = \int_B p(s, s')\, \mathrm{d}s' + (1 - \lambda(s))\mathbb{1}\{s \in B\} \tag{22}$$

for all $s \in S$ and $B \in \mathscr{B}(S)$, where

$$p(s, s') := q(s, s')\alpha(s, s') \quad \text{and} \quad \lambda(s) := \int p(s, s')\, \mathrm{d}s'$$

See ET page 202 for a proof

Now design the selection function α so the stationary distribution of P is the target density π

$$\alpha(s, s') := \min\left\{\frac{\pi(s')q(s', s)}{\pi(s)q(s, s')}, 1\right\} \tag{23}$$

The ratio inside the min is large when the new proposal s' has high probability under π relative to the current location s and small otherwise

Theorem. (7.4.1) Let P be the stochastic kernel in (22). If α is as defined (23), then π is a stationary distribution for P

For a proof, see page 204 in ET

P not always stable, we require conditions on q and π

One well-known result given in corollary 3 of Tierney (1994):

- q and π are continuous and strictly positive on S (cf. (a) of theorem 7.2.2)
- and S itself is compact (cf. (b) of theorem 7.2.2)

One disadvantage of MCMC methods is that the variates they generate are not in general independent

Sample averages from dependent samples are typically less informative than IID samples. (See example 7.1.4 in ET)

Desirable that dependence between x_t and x_{t+k} dies out quickly

Whether it does or not is determined by the choice of q

Example Julia code: set proposal density of the form $q(s, s') = \psi(s' - s)$, where ψ is a given density

If ψ is symmetric, then

- $q(s, s') = q(s', s)$
- $\alpha(s, s')$ simplifies to $\min\{\pi(s')/\pi(s), 1\}$ if $\pi(s) > 0$ and 1 otherwise

We set ψ set to the $\mathrm{N}(0, \sigma^2)$

```
function rw_metropolis(pi_density, T, init=0, sigma=1)
    xvec = Array(Float64, T)
    xvec[1] = init
    for t in 1:(T - 1)
        x = xvec[t]
        y = x + sigma * randn()
        alpha = pi_density(x) > 0 ? pi_density(y)
            / pi_density(x) : 1
        xvec[t+1] = rand() < alpha  ? y : x
    end
    return xvec
end
```

Listing 1: The random walk Metropolis–Hastings algorithm (Julia)

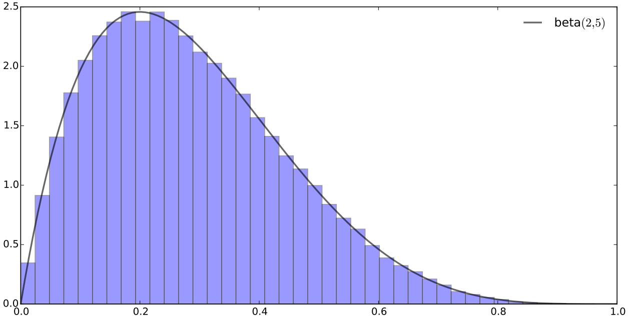


Figure: Observations from the random walk Metropolis–Hastings algorithm when π is the beta distribution