Advanced Econometric Methods EMET3011/8014

Lecture 11

John Stachurski

Semester 1, 2011

Announcements/Reminders

- Please get a fresh copy of the course notes PDF
- Assignment 2 due tonight (midnight)
- SELT evaluations starting

Today's Lecture

- Time Series Models
- Markov Models
- Martingales
- Stochastic Stability

Convergence of Random Matrices

We need to cover matrix/vector

- Convergence in probability
- Convergence in distribution
- LLN/CLT

Involves bootstrapping scalar definitions/results

Convergence in Probability

Recall definition of scalar convergence in probability:

$$x_n \xrightarrow{p} x \quad \text{if} \quad \mathbb{P}\{|x_n - x| > \epsilon\} \to 0 \quad \text{for any} \quad \epsilon > 0$$

Let

- $\{\mathbf{X}_n\}$ be a sequence of random $I \times J$ matrices
- $\mathbf{X}$ be a random $I \times J$ matrix

$\mathbf{X}_n$ converges to $\mathbf{X}$ **in probability** if every element of $\mathbf{X}_n$ converges to the corresponding element of $\mathbf{X}$ in probability in the scalar sense

That is,

$$\mathbf{X}_n \xrightarrow{p} \mathbf{X} \text{ whenever } x_{ij}^n \xrightarrow{p} x_{ij} \text{ for all } i \text{ and } j$$

In vector case, definition becomes

$$\begin{pmatrix} x_1^n \\ \vdots \\ x_K^n \end{pmatrix} \xrightarrow{p} \begin{pmatrix} x_1 \\ \vdots \\ x_K \end{pmatrix} \quad \text{whenever} \quad x_k^n \xrightarrow{p} x_k \text{ for all } k$$

Fact. For vectors $\{\mathbf{x}_n\}$ and $\mathbf{x}$, the following statements are equivalent

- $\mathbf{x}_n \xrightarrow{p} \mathbf{x}$
- $\|\mathbf{x}_n - \mathbf{x}\| \xrightarrow{p} 0$

Exercise: Prove it. (Solved exercise in course notes)

Fact. For conformable matrices, the following is true:

1. If $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$ and $\mathbf{X}_n$ and $\mathbf{X}$ are invertible, then $\mathbf{X}_n^{-1} \xrightarrow{p} \mathbf{X}^{-1}$
2. If $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$ and $\mathbf{Y}_n \xrightarrow{p} \mathbf{Y}$, then

$$\mathbf{X}_n + \mathbf{Y}_n \xrightarrow{p} \mathbf{X} + \mathbf{Y} \quad \text{and} \quad \mathbf{X}_n \mathbf{Y}_n \xrightarrow{p} \mathbf{X}\mathbf{Y}$$

3. If $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$ and $\mathbf{A}_n \to \mathbf{A}$, then

$$\mathbf{X}_n + \mathbf{A}_n \xrightarrow{p} \mathbf{X} + \mathbf{A}, \quad \mathbf{X}_n \mathbf{A}_n \xrightarrow{p} \mathbf{X}\mathbf{A} \quad \text{and} \quad \mathbf{A}_n \mathbf{X}_n \xrightarrow{p} \mathbf{A}\mathbf{X}$$

Remark: $\mathbf{A}_n \to \mathbf{A}$ means $a_{ij}^n \to a_{ij}$ for all i and j

Fact. If $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$ and $\mathbf{a}$ is constant and conformable, then

$$\mathbf{a}'\mathbf{X}_n\mathbf{a} \xrightarrow{p} \mathbf{a}'\mathbf{X}\mathbf{a}$$

Proof: Let $\mathbf{a}$ be fixed and let $\mathbf{X}_n \xrightarrow{p} \mathbf{X}$

From previous slide, we know that if $\mathbf{A}_n \to \mathbf{A}$, then

$$\mathbf{X}_n\mathbf{A}_n \xrightarrow{p} \mathbf{X}\mathbf{A} \quad \text{and} \quad \mathbf{A}_n\mathbf{X}_n \xrightarrow{p} \mathbf{A}\mathbf{X}$$

$$\therefore \quad \mathbf{X}_n\mathbf{a} \xrightarrow{p} \mathbf{X}\mathbf{a}$$

$$\therefore \quad \mathbf{a}'\mathbf{X}_n\mathbf{a} \xrightarrow{p} \mathbf{a}'\mathbf{X}\mathbf{a}$$

Convergence in Distribution

Let $\{F_n\}_{n=1}^{\infty}$ and F be cdfs on $\mathbb{R}^K$

F_n converges to F **weakly** if, for any $\mathbf{s}$ such that F is continuous at $\mathbf{s}$, we have

$$F_n(\mathbf{s}) \to F(\mathbf{s}) \quad \text{as} \quad n \to \infty$$

Let $\{\mathbf{x}_n\}_{n=1}^{\infty}$ and $\mathbf{x}$ be random vectors, where

- $\mathbf{x}_n \sim F_n$
- $\mathbf{x} \sim F$

Definition: $\mathbf{x}_n$ converges **in distribution** to $\mathbf{x}$ if F_n converges weakly to F

In symbols: $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$

Note:

- True: If $x_n^k \xrightarrow{p} x^k$ for all k, then $\mathbf{x}_n \xrightarrow{p} \mathbf{x}$
- False: If $x_n^k \xrightarrow{d} x^k$ for all k, then $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$

Convergence of marginals does not imply convergence of joint

(Unless independent, when joint is just product of the marginals)

However,

- True: If $\mathbf{a}'\mathbf{x}_n \xrightarrow{p} \mathbf{a}'\mathbf{x}$ for any vector $\mathbf{a}$, then $\mathbf{x}_n \xrightarrow{p} \mathbf{x}$
- True: If $\mathbf{a}'\mathbf{x}_n \xrightarrow{d} \mathbf{a}'\mathbf{x}$ for any vector $\mathbf{a}$, then $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$

Second result called the **Cramer-Wold device**

Let's prove that $\mathbf{a}'\mathbf{x}_n \xrightarrow{p} \mathbf{a}'\mathbf{x}$ for any $\mathbf{a}$ implies $\mathbf{x}_n \xrightarrow{p} \mathbf{x}$

Proof: If condition holds, then $\mathbf{e}_k'\mathbf{x}_n \xrightarrow{p} \mathbf{e}_k'\mathbf{x}$ for all k

$$\therefore \quad x_n^k \xrightarrow{p} x^k \quad \text{for all } k \quad \text{(elementwise convergence)}$$

$$\therefore \quad \mathbf{x}_n \xrightarrow{p} \mathbf{x} \quad \text{(vector convergence, by definition)}$$

How about the proof that $\mathbf{a}'\mathbf{x}_n \xrightarrow{d} \mathbf{a}'\mathbf{x}$ for any $\mathbf{a}$ implies $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$?

Tricky, or involves characteristic functions

The next two results are used routinely in econometric theory

Continuous mapping theorem.

- If $g \colon \mathbb{R}^K \to \mathbb{R}^J$ continuous and $\mathbf{x}_n \stackrel{d}{\to} \mathbf{x}$, then $g(\mathbf{x}_n) \stackrel{d}{\to} g(\mathbf{x})$
- If $g \colon \mathbb{R}^K \to \mathbb{R}^J$ continuous and $\mathbf{x}_n \stackrel{p}{\to} \mathbf{x}$, then $g(\mathbf{x}_n) \stackrel{p}{\to} g(\mathbf{x})$

Slutsky's theorem.

Let $\mathbf{x}_n$ and $\mathbf{x}$ be random vectors in $\mathbb{R}^K$, let $\mathbf{Y}_n$ be random matrices, and let $\mathbf{C}$ be a constant matrix

If $\mathbf{Y}_n \stackrel{p}{\to} \mathbf{C}$ and $\mathbf{x}_n \stackrel{d}{\to} \mathbf{x}$, then

$$\mathbf{Y}_n \mathbf{x}_n \stackrel{d}{\to} \mathbf{C}\mathbf{x} \quad \text{and} \quad \mathbf{Y}_n + \mathbf{x}_n \stackrel{d}{\to} \mathbf{C} + \mathbf{x}$$

whenever matrices/vectors are conformable

Vector LLN and CLT

Scalar LLN / CLT extend to the vector case in a natural way

Theorem. Let

1. $\{\mathbf{x}_n\}$ be an IID sequence in $\mathbb{R}^K$
2. $\boldsymbol{\mu} := \mathbb{E}[\mathbf{x}_n]$
3. $\Sigma := \mathrm{var}[\mathbf{x}_n]$

For this sequence we have

$$\bar{\mathbf{x}}_N := \frac{1}{N}\sum_{n=1}^{N} \mathbf{x}_n \xrightarrow{p} \boldsymbol{\mu} \quad \text{and} \quad \sqrt{N}\,(\bar{\mathbf{x}}_N - \boldsymbol{\mu}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma)$$

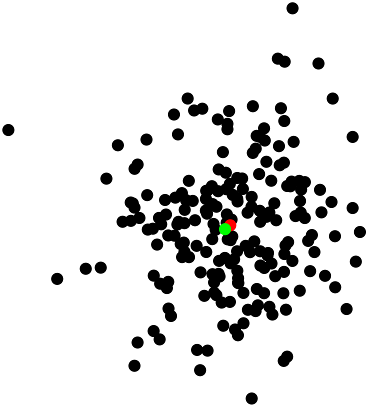


Figure: LLN, vector case

Proof of vector LLN: Fix constant vector $\mathbf{a}$ and let $y_n := \mathbf{a}'\mathbf{x}_n$

Since $\{y_n\}$ is IID (why?), the scalar LLN yields

$$\frac{1}{N}\sum_{n=1}^{N} y_n \stackrel{p}{\to} \mathbb{E}\left[y_n\right] = \mathbb{E}\left[\mathbf{a}'\mathbf{x}_n\right] = \mathbf{a}'\mathbb{E}\left[\mathbf{x}_n\right] = \mathbf{a}'\boldsymbol{\mu}$$

But

$$\frac{1}{N}\sum_{n=1}^{N} y_n = \frac{1}{N}\sum_{n=1}^{N} \mathbf{a}'\mathbf{x}_n = \mathbf{a}'\left[\frac{1}{N}\sum_{n=1}^{N} \mathbf{x}_n\right] = \mathbf{a}'\bar{\mathbf{x}}_N$$

Since $\mathbf{a}$ was arbitrary, we have

$$\mathbf{a}'\bar{\mathbf{x}}_N \stackrel{p}{\to} \mathbf{a}'\boldsymbol{\mu} \quad \text{for any} \quad \mathbf{a} \in \mathbb{R}^K$$

$$\therefore \quad \bar{\mathbf{x}}_N \stackrel{p}{\to} \boldsymbol{\mu}$$

Exercise: Verify the vector CLT

Hints:

- Similar to the vector LLN proof
- Use the Cramer-Wold device

Some Common Time Series Models

First model: Linear Gaussian scalar AR(1) model

$$x_{t+1} = \alpha + \rho x_t + w_{t+1} \quad \text{with} \quad \{w_t\} \overset{\text{IID}}{\sim} \mathcal{N}(0, \sigma^2)$$

The initial condition x_0 is some given RV

The RV x_t is called the **state variable**

The distribution of x_t will be denoted by Π_t

Generalizations of the Gaussian scalar AR(1) model:

Without normal shocks, becomes the scalar AR(1) model

In $\mathbb{R}^K$, becomes the vector AR(1) model, or VAR(1):

$$\mathbf{x}_{t+1} = \mathbf{a} + \mathbf{\Lambda} \mathbf{x}_t + \mathbf{w}_{t+1} \quad \text{with } \{\mathbf{w}_t\} \overset{\text{IID}}{\sim} \phi \text{ and } \mathbf{x}_0 \text{ given}$$

Here

- $\mathbf{a}$ is a $K \times 1$ column vector
- $\mathbf{\Lambda}$ is a $K \times K$ matrix
- ϕ is some distribution on $\mathbb{R}^K$
- $\mathbf{x}_t$ is called the **state vector**

If $\phi = \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ then called the Gaussian VAR(1)

Another generalization of scalar AR(1) is scalar AR(2)

$$x_{t+1} = \alpha + \rho x_t + \gamma x_{t-1} + w_{t+1}$$

We can reformulate it as a VAR(1) by setting $y_t = x_{t-1}$, so that

$$\begin{aligned} x_{t+1} &= \alpha + \rho x_t + \gamma y_t + w_{t+1} \\ y_{t+1} &= x_t \end{aligned}$$

or

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

Message: It's enough to understand dynamics of first order models

Nonlinear Models

The previous examples are all linear models

Linear models are simple, but

- cannot always capture the dynamics we observe in data
- many theoretical modeling models are nonlinear

Next we introduce several popular nonlinear models

Example: ARCH(1)

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

The idea:

Model returns on a given asset by $x_t = \sigma_t w_t$ where

- $\{w_t\} = \text{IID}$ shocks
- $\sigma_t =$ time-varying volatility component

Dynamics of σ_t specified by $\sigma_{t+1}^2 = \alpha_0 + \alpha_1 x_t^2$

Combining equations leads to ARCH(1) above

Example 2: Smooth transition threshold autoregression (STAR)

First order scalar version has dynamics

$$x_{t+1} = g(x_t) + w_{t+1}$$

where g is of the form

$$g(s) := (\alpha_0 + \rho_0 s)(1 - \tau(s)) + (\alpha_1 + \rho_1 s)\tau(s)$$

Here τ is a smooth cdf on $\mathbb{R}$, so

- Small s implies $g(s) \approx \alpha_0 + \rho_0 s$
- Large s implies $g(s) \approx \alpha_1 + \rho_1 s$

Markov Models

Perhaps the simplest generalization of an IID process

Simple enough to have neat theory, but general enough to accommodate many different classes of models

First order discrete time Markov process has the form

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

where

- $\{\mathbf{w}_t\}$ IID with common distribution ϕ
- $\mathbf{x}_0$ independent of $\{\mathbf{w}_t\}$

We let

$$\Pi_t(\mathbf{s}) := \mathbb{P}\{\mathbf{x}_t \leq \mathbf{s}\}$$

Taking model $\mathbf{x}_{t+1} = G(\mathbf{x}_t, \mathbf{w}_{t+1})$ and iterating,

$$\begin{aligned}
\mathbf{x}_1 &= G(\mathbf{x}_0, \mathbf{w}_1) \\
\mathbf{x}_2 &= G(G(\mathbf{x}_0, \mathbf{w}_1), \mathbf{w}_2) \\
\mathbf{x}_3 &= G(G(G(\mathbf{x}_0, \mathbf{w}_1), \mathbf{w}_2), \mathbf{w}_3) \\
&\text{etc., etc.}
\end{aligned}$$

Message: $\mathbf{x}_t$ can be written as a function of $\mathbf{x}_0$ and $\mathbf{w}_1, \ldots, \mathbf{w}_t$

In other words, for each t, there exists a function H_t such that

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

This expression clarifies the fact that

- $\mathbf{x}_t$ as a well-defined random variable
- $\mathbf{x}_t$ and $\mathbf{w}_{t+j}$ are independent for every t and $j > 0$

Example: Scalar linear AR(1) process

$$x_{t+1} = \alpha + \rho x_t + w_{t+1}$$

Recall that, in general, there is an H_t such that

$$x_t = H_t(x_0, w_1, w_2, \ldots, w_t)$$

In linear AR(1) case, there is a neat expression for H_t:

$$x_t = \alpha \sum_{k=0}^{t-1} \rho^k + \sum_{k=0}^{t-1} \rho^k w_{t-k} + \rho^t x_0$$

Exercise: Check it

Transition density of a Markov process is

$$p(\cdot \mid \mathbf{s}) := \text{ conditional density of } \mathbf{x}_{t+1} \text{ when } \mathbf{x}_t \text{ equals } \mathbf{s}$$

Example: What is the transition density for

$$x_{t+1} = g(x_t) + w_{t+1} \quad \text{with} \quad \{w_t\} \overset{\text{IID}}{\sim} \text{ density } \phi$$

Answer:

$$p(t \mid s) = \phi(t - g(s))$$

Reason: If $y := g(s) + w$ and $w \sim \phi$, then density of y is

$$\psi(t) = \phi(t - g(s))$$

Claim: If $y := g(s) + w$ and $w \sim \phi$, then density of y is

$$\psi(t) = \phi(t - g(s))$$

Proof: Let

- Φ be the cdf corresponding to ϕ (i.e., cdf of w)
- Ψ be the cdf corresponding to ψ (i.e., cdf of y)

$$\begin{aligned}\Psi(t) &:= \mathbb{P}\{y \leq t\} \\ &= \mathbb{P}\{g(s) + w \leq t\} \\ &= \mathbb{P}\{w \leq t - g(s)\} = \Phi(t - g(s))\end{aligned}$$

$$\therefore \quad \psi(t) = \frac{d}{dt}\Psi(t) = \frac{d}{dt}\Phi(t - g(s)) = \phi(t - g(s))$$

Martingales

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

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

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

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

Then $\{\mathcal{F}_t\}$ is a filtration

Let

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

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

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

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

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

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

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

Proof: We know that

- m_t is $\mathcal{F}_t$-measurable
- If $j \geq 0$, then $\mathcal{F}_t \subset \mathcal{F}_{t+j}$
- 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 $\{\mathcal{F}_t\}$
- having finite first moment

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

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

We say that $\{m_t\}$ is a **martingale difference sequence** (MDS) with respect to $\{\mathcal{F}_t\}$ if

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

Exercise: If $\{m_t\}$ is a martingale w.r.t. $\{\mathcal{F}_t\}$, then $d_t := m_t - m_{t-1}$ is a MDS w.r.t. $\{\mathcal{F}_t\}$

Example 1 (Random walk):

Let $\{\eta_t\}$ be IID with $\mathbb{E}[\eta_1] = 0$

Let $m_t := \sum_{j=1}^{t} \eta_j$ and let $\mathcal{F}_t := \{\eta_1, \ldots, \eta_t\}$

In this case, $\{m_t\}$ is is a martingale with respect to $\{\mathcal{F}_t\}$

That $\{m_t\}$ is adapted to $\{\mathcal{F}_t\}$ follows from the definitions

Moreover,

$$
\begin{aligned}
\mathbb{E}[m_{t+1} \mid \mathcal{F}_t] &= \mathbb{E}[\eta_1 + \cdots + \eta_t + \eta_{t+1} \mid \mathcal{F}_t] \\
&= \mathbb{E}[\eta_1 \mid \mathcal{F}_t] + \cdots + \mathbb{E}[\eta_t \mid \mathcal{F}_t] + \mathbb{E}[\eta_{t+1} \mid \mathcal{F}_t] \\
&= \eta_1 + \cdots + \eta_t + \mathbb{E}[\eta_{t+1}] \\
&= \eta_1 + \cdots + \eta_t \\
&= m_t
\end{aligned}
$$

Example 2 (Consumption, Robert Hall, 1978, JPE):

Consider an Euler equation of the form

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

where

- u is a utility function
- $r_t =$ interest rate and $\rho =$ a discount factor
- $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot \mid \mathcal{F}_t]$, where $\mathcal{F}_t =$ information set at t

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

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

In this case, consumption is a martingale with respect to $\{\mathcal{F}_t\}$

Dynamics

Questions:

- When is a given stochastic process "stable"?
- How can it be "stable" and stochastic at the same time?
- When does the LLN hold?
- When does the CLT hold?

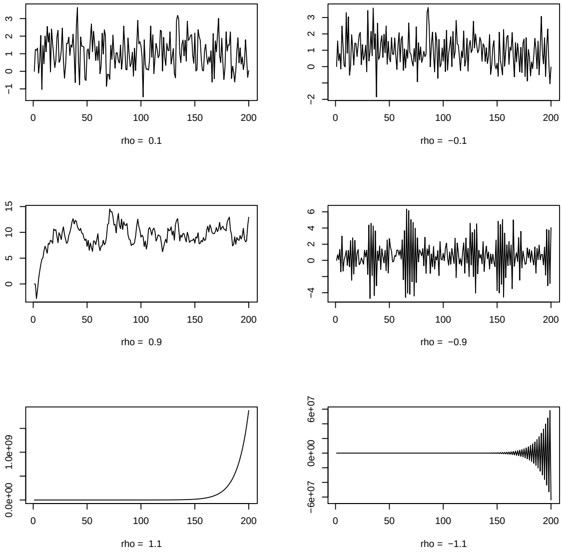
Let's start informally, focusing on the Gaussian scalar AR(1) process

$$x_{t+1} = \alpha + \rho x_t + w_{t+1} \qquad \{w_t\}_{t=0}^{\infty} \stackrel{\text{IID}}{\sim} \mathcal{N}(0,1)$$

Dynamics are sensitive to the value of the coefficient ρ

- If $\rho \in (-1, 1)$, then "stable"
- If $\rho \notin (-1, 1)$, then time series diverge

Some different scenarios are illustrated on next slide


rho = 0.1
rho = −0.1
rho = 0.9
rho = −0.9
rho = 1.1
rho = −1.1

Let's investigate this stability/instability analytically

Recall that

$$x_t = \alpha \sum_{k=0}^{t-1} \rho^k + \sum_{k=0}^{t-1} \rho^k w_{t-k} + \rho^t x_0 \tag{2}$$

From (2) we see that

- x_t normally distributed (assume x_0 normal or constant)
- $\mu_t := \mathbb{E}\left[x_t\right] = \alpha \sum_{k=0}^{t-1} \rho^k + \rho^t \mu_0$
- $\sigma_t^2 := \mathrm{var}[x_t] = \sum_{k=0}^{t-1} \rho^{2k} + \rho^{2t} \sigma_0^2$

In other words, if $\Pi_t =$ distribution of x_t, then

$$\Pi_t = \mathcal{N}(\mu_t, \sigma_t^2) = \mathcal{N}\left(\alpha \sum_{k=0}^{t-1} \rho^k + \rho^t \mu_0, \; \sum_{k=0}^{t-1} \rho^{2k} + \rho^{2t} \sigma_0^2 \right)$$

If $|\rho| \geq 1$ then mean and variance diverge, while if $|\rho| < 1$, then

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

In this case, can show that

$$\Pi_t = \mathcal{N}(\mu_t, \sigma_t^2) \xrightarrow{d} \Pi_\infty := \mathcal{N}(\mu_\infty, \sigma_\infty^2) := \mathcal{N}\left(\frac{\alpha}{1-\rho}, \frac{1}{1-\rho^2} \right)$$

Note: limit does <u>not</u> depend on μ_0 and σ_0^2

Next two slides illustrate convergence of

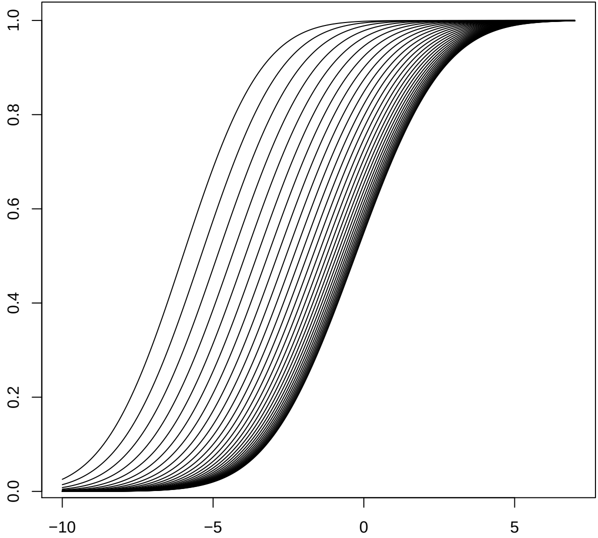
- $\Pi_t \to \Pi_\infty$ (cdfs)
- $\pi_t \to \pi_\infty$ (densities)

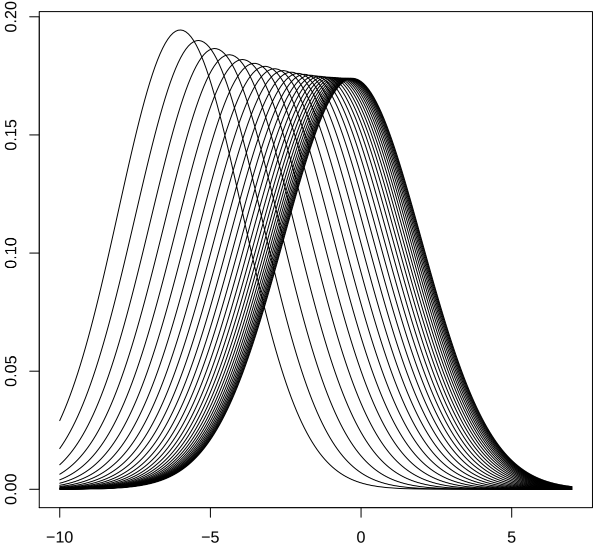
Parameters: $\alpha = 0$ and $\rho = 0.9$

Initial distribution $\Pi_0 := \mathcal{N}(\mu_0, \sigma_0^2)$ with

- $\mu_0 = -6$
- $\sigma_0^2 = 4.2$

Convergence is from left to right


1.0
0.8
0.6
0.4
0.2
0.0
−10
−5
0
5


0.20
0.15
0.10
0.05
0.00
−10
−5
0
5

Assume that $\rho \in (-1, 1)$

Distribution $\Pi_\infty = \mathcal{N}(\mu_\infty, \sigma_\infty^2)$ exists and has a special property:

- If $\Pi_0 = \Pi_\infty$, then $\Pi_t = \Pi_\infty$ for all t

Exercise: Check it by verifying that

- $x_t \sim \Pi_\infty$ implies $x_{t+1} \sim \Pi_\infty$

The distribution Π_∞ is called a **stationary distribution**

Note in particular: If $x_0 \sim \Pi_\infty$, then $\{x_t\}$ is identically distributed

Even if $x_0 \sim \Pi_\infty$ not true

$$\Pi_t \to \Pi_\infty \text{ as } t \to \infty \text{ regardless of } \Pi_0$$

- This property called **global stability**
- Informally: $\{x_t\}$ is "asymptotically identically distributed"

LLN for the AR(1) Model

LLN does not hold in general

Example: Consider special case $x_{t+1} = x_t$ with $x_0 \sim \mathcal{N}(0,1)$

Process $\{x_t\}$ is identically distributed, because

$$x_t = x_{t-1} = \cdots = x_0$$

$$\therefore \quad \mathbb{P}\{x_t \leq s\} = \mathbb{P}\{x_{t-1} \leq s\} = \cdots = \mathbb{P}\{x_0 \leq s\}$$

However, LLN fails because $\mathbb{E}[x_t] = 0$ while

$$\bar{x}_T := \frac{1}{T}\sum_{t=1}^{T} x_t = \frac{1}{T}\sum_{t=1}^{T} x_0 = x_0 \sim \mathcal{N}(0,1)$$

Recall IID LLN: If $\{x_t\}$ is IID, then

$$\bar{x}_T := \frac{1}{T}\sum_{t=1}^{N} x_t \xrightarrow{p} \mathbb{E}[x_t] \quad \text{as} \quad T \to \infty$$

Key to proof: $\operatorname{var}[\bar{x}_T] \to 0$ as $T \to \infty$, because

$$\begin{aligned}\operatorname{var}\left[\frac{1}{T}\sum_{t=1}^{T} x_t\right] &= \frac{1}{T^2}\left[\sum_{t=1}^{T}\operatorname{var}[x_t] + 2\sum_{t<m}\operatorname{cov}[x_t, x_m]\right] \\ &= \frac{\sigma^2}{T} + \frac{2}{T^2}\sum_{t<m}\operatorname{cov}[x_t, x_m]\end{aligned}$$

- IID implies zero autocorrelation implies $\operatorname{var}[\bar{x}_T] \to 0$
- But LLN may still hold if $\operatorname{cov}[x_t, x_{t+j}] \approx 0$ when j large

Correlations dying out over time closely related to global stability

To see why, consider the AR(1) model, where

$$x_j = \alpha \sum_{k=0}^{j-1} \rho^k + \sum_{k=0}^{j-1} \rho^k w_{j-k} + \rho^j x_0$$

Shifting forward, we get

$$x_{t+j} = \alpha \sum_{k=0}^{j-1} \rho^k + \sum_{k=0}^{j-1} \rho^k w_{t+j-k} + \rho^j x_t$$

Global stability is equivalent to $|\rho| < 1$

Means x_{t+j} and x_t are “asymptotically independent/uncorrelated”

Hence, global stability implies

- “asymptotically identically distributed”
- “asymptotically uncorrelated/independent”

In other words

- global stability $\implies$ “asymptotically almost IID”

This provides enough for the LLN to hold

Next slide illustrates with $\alpha = 0$ and $\rho = 0.8$

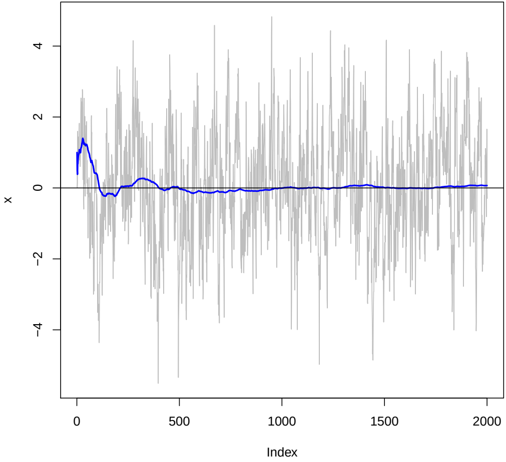
If the LLN holds we would have

$$\bar{x}_T \xrightarrow{p} \mu_\infty = \frac{\alpha}{1-\rho} = \frac{0}{0.2} = 0$$

The grey line is a simulated time series of the process

The blue line is the plot of $\bar{x}_T$ against T

We see that $\bar{x}_T$ is converging to zero as expected


x
4
2
0
−2
−4
0
500
1000
1500
2000
Index

Markovian Dynamics

In the last few slides we studied the dynamics of scalar AR(1) processes

Discussion was relatively informal

Let's now study the general case

And consider more formal definitions / results

Recall the general Markov process

$$\mathbf{x}_{t+1} = G(\mathbf{x}_t, \mathbf{w}_{t+1})$$

where

- $\{\mathbf{w}_t\}$ is IID with common density ϕ
- $\mathbf{x}_0$ independent of $\{\mathbf{w}_t\}$

Let $\pi_t :=$ density of $\mathbf{x}_t$

The transition density is

$$p(\cdot \mid \mathbf{s}) := \text{ conditional density of } \mathbf{x}_{t+1} \text{ when } \mathbf{x}_t \text{ equals } \mathbf{s}$$

The sequence $\{\pi_t\}_{t=0}^{\infty}$ is known to satisfy

$$\pi_{t+1}(\mathbf{t}) = \int p(\mathbf{t} \,|\, \mathbf{s}) \pi_t(\mathbf{s}) d\mathbf{s} \quad \text{for all } \mathbf{t} \in \mathbb{R}^K \tag{3}$$

Informal interpretation via law of total probability:

$$\mathbb{P}\{\mathbf{x}_{t+1} = \mathbf{t}\} = \sum_{\mathbf{s}} \mathbb{P}\{\mathbf{x}_{t+1} = \mathbf{t} \,|\, \mathbf{x}_t = \mathbf{s}\} \mathbb{P}\{\mathbf{x}_t = \mathbf{s}\}$$

Density π_∞ called **stationary** if

$$\pi_\infty(\mathbf{t}) = \int p(\mathbf{t} \,|\, \mathbf{s}) \pi_\infty(\mathbf{s}) d\mathbf{s} \quad \text{for all } \mathbf{t} \in \mathbb{R}^K$$

Interpretation: Updating does not change it

Fact. If π_∞ is stationary for the process then

- $\pi_t = \pi_\infty$ implies $\pi_{t+1} = \pi_\infty$
- $\mathbf{x}_0 \sim \pi_\infty$ implies $\{\mathbf{x}_t\}$ is identically distributed

Def. A stationary density π_∞ is called **globally stable** if

$$\pi_t \to \pi_\infty \text{ for every initial cdf } \pi_0$$

Implication:

- $\{\mathbf{x}_t\}$ is "asymptotically identically distributed"

A general stability result for $\mathbf{x}_{t+1} = g(\mathbf{x}_t) + \mathbf{w}_{t+1}$ with $\{\mathbf{w}_t\} \overset{\text{IID}}{\sim} \phi$

Theorem. (Markov Stability) If

1. $\phi =$ density with finite mean, $\phi > 0$ everywhere on $\mathbb{R}^K$
2. g is continuous
3. exist positive constants λ and L such that $\lambda < 1$ and

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

then $\exists$ unique, globally stable stationary density π_∞, and

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

for any $h \colon \mathbb{R}^K \to \mathbb{R}$

Example: Gaussian VAR(1)

$$\mathbf{x}_{t+1} = \mathbf{a} + \mathbf{A}\mathbf{x}_t + \mathbf{w}_{t+1}$$

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

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

$$\therefore \quad \|g(\mathbf{s})\| \leq \lambda\|\mathbf{s}\| + L$$

where

$$\lambda := \text{ spectral norm } := \max_{\mathbf{s} \neq 0} \frac{\|\mathbf{A}\mathbf{s}\|}{\|\mathbf{s}\|} \quad \text{and} \quad L := \|\mathbf{a}\|$$

Conditions of theorem satisfied if $\lambda < 1$

Martingale Difference LLN and CLT

Martingale difference sequences are good candidates for the LLN / CLT because they are uncorrelated

Proof of zero autocorrelation:

- Let $\{m_t\}$ be an MDS w.r.t. filtration $\{\mathcal{F}_t\}$
- Fix $t \geq 0$ and $j \geq 1$

We have

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

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

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

Theorem. If $\{m_t\}$ is an identically distributed MDS w.r.t. $\{\mathcal{F}_t\}$, then

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

If, in addition, $\gamma^2 := \mathbb{E}\left[m_t^2\right]$ is positive and finite, and

$$\frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[m_t^2 \,|\, \mathcal{F}_{t-1}\right] \xrightarrow{p} \gamma^2 \quad \text{as} \quad T \to \infty$$

then

$$\sqrt{T}\left[\frac{1}{T}\sum_{t=1}^{T} m_t\right] = T^{-1/2}\sum_{t=1}^{T} m_t \xrightarrow{d} \mathcal{N}(0, \gamma^2) \quad \text{as} \quad T \to \infty$$