ECON-GA 1025 Macroeconomic Theory I

Lecture 6

John Stachurski

Fall Semester 2018

Today's Lecture

- Markov chains continued
- deterministic linear dynamics
- vector autoregressions

Markov Chains: Probabilistic Properties

Let Π be a stochastic kernel on X and let x, y be states

We say that y is **accessible** from x if $x = y$ or

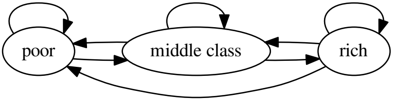
$$\exists\, k \in \mathbb{N} \text{ such that } \Pi^k(x, y) > 0$$

Equivalent: Accessible in the induced directed graph

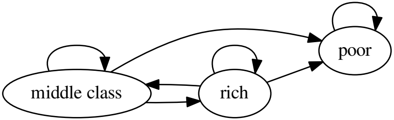
A stochastic kernel Π on X is called **irreducible** if every state is accessible from any other

Equivalent: The induced directed graph is strongly connected

Irreducible:



Not irreducible:



Aperiodicity

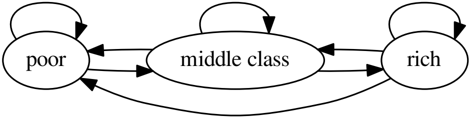
Let Π be a stochastic kernel on X

State $x \in$ X is called **aperiodic** under Π if

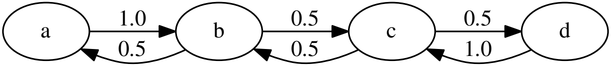
$$\exists\, i \in \mathbb{N} \text{ such that } k \geqslant i \implies \Pi^k(x, x) > 0$$

A stochastic kernel Π on X is called **aperiodic** if every state in X is aperiodic under Π

Aperiodic?

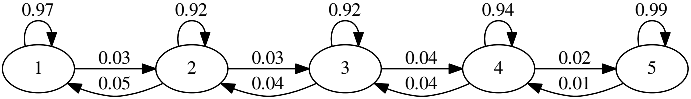


Aperiodic?



Stability of Markov Chains

Recall the distributions generated by Quah's model



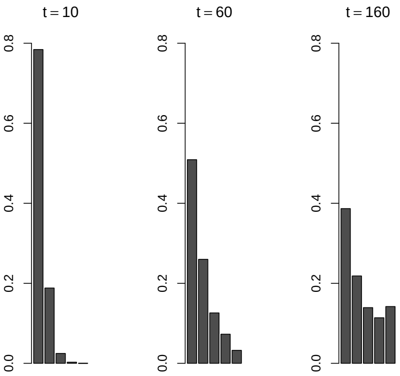


Figure: $X_0 = 1$

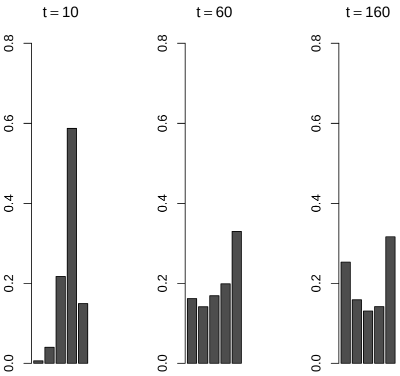


Figure: $X_0 = 4$

What happens when $t \to \infty$?

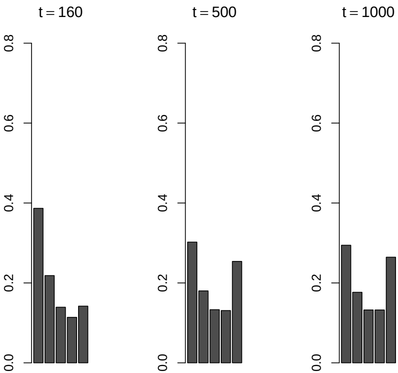


Figure: $X_0 = 1$

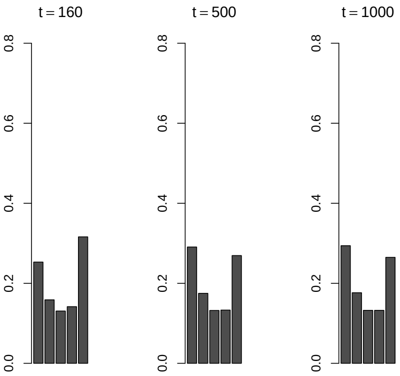


Figure: $X_0 = 4$

At $t = 1000$, the distributions are almost the same for both starting points

This suggests we are observing a form of stability

- is $(\mathcal{P}(\mathsf{X}), \Pi)$ globally stable?

Not all stochastic kernels are globally stable

Example. Let $\mathsf{X} = \{0, 1\}$ and consider the periodic Markov chain

$$\Pi = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Ex. Show $\psi^* = (0.5, 0.5)$ is stationary for Π

Ex. Show that

$$\delta_0 \Pi^t = \begin{cases} \delta_1 & \text{if } t \text{ is odd} \\ \delta_0 & \text{if } t \text{ is even} \end{cases}$$

Conclude that global stability fails

Proving Stability

Fact. The operator Π is always nonexpansive:

$$\|\varphi\Pi - \psi\Pi\|_1 \leqslant \|\varphi - \psi\|_1 \quad \forall\, \varphi, \psi \in \mathcal{P}(\mathsf{X})$$

Proof:

$$\|\varphi\Pi - \psi\Pi\|_1 = \sum_y \left| \sum_x \Pi(x,y)[\varphi(x) - \psi(x)] \right|$$

$$\leqslant \sum_y \sum_x \Pi(x,y)|\varphi(x) - \psi(x)|$$

$$= \sum_x \sum_y \Pi(x,y)|\varphi(x) - \psi(x)| = \|\varphi - \psi\|_1$$

With some more conditions we might be able to apply this result:

Theorem. If (M, ρ) is a compact metric space and $T \colon M \to M$ is a strict contraction, then (M, T) is globally stable

- **strict contraction** means $\rho(Tx, Ty) < \rho(x, y)$ when $x \neq y$
- a variation on the Banach CMT

X is finite, so $\mathcal{P}(\mathsf{X})$ is compact

We just need to boost nonexpansiveness to strict contractivity

Lemma. If $\Pi(x, y) > 0$ for all x, y, then Π is a strict contraction on $\mathcal{P}(\mathsf{X})$ under the metric d_1

The proof uses two lemmas:

Fact. If $\varphi, \psi \in \mathcal{P}(\mathsf{X})$ and $\varphi \neq \psi$, then

$$\exists\, x, x' \in \mathsf{X} \text{ such that } \varphi(x) > \psi(x) \text{ and } \varphi(x') < \psi(x')$$

Fact. If $g \in \mathbb{R}^{\mathsf{X}}$ and $\exists\, x, x' \in \mathsf{X}$ s.t. $g(x) > 0$ and $g(x') < 0$, then

$$|\sum_{y \in \mathsf{X}} g(y)| < \sum_{y \in \mathsf{X}} |g(y)|$$

Ex. Prove both

Under the conditions of the theorem, if $\varphi \neq \psi$, then

$$
\begin{aligned}
\|\varphi\Pi - \psi\Pi\|_1 &= \sum_y \left| \sum_x \Pi(x,y)\varphi(x) - \sum_x \Pi(x,y)\psi(x) \right| \\
&= \sum_y \left| \sum_x \Pi(x,y)[\varphi(x) - \psi(x)] \right| \\
&< \sum_y \sum_x |\Pi(x,y)[\varphi(x) - \psi(x)]| \\
&= \sum_y \sum_x \Pi(x,y)|\varphi(x) - \psi(x)| \\
&= \sum_x \sum_y \Pi(x,y)|\varphi(x) - \psi(x)| = \|\varphi - \psi\|_1
\end{aligned}
$$

We have prove the following:

Proposition. If $\Pi \gg 0$, then $(\mathcal{P}(\mathsf{X}), \Pi)$ is globally stable

But this condition is rather strict

- Hamilton's matrix fails it
- Quah's matrix fails it

$$\Pi_Q = \begin{pmatrix} 0.97 & 0.03 & 0.00 & 0.00 & 0.00 \\ 0.05 & 0.92 & 0.03 & 0.00 & 0.00 \\ 0.00 & 0.04 & 0.92 & 0.04 & 0.00 \\ 0.00 & 0.00 & 0.04 & 0.94 & 0.02 \\ 0.00 & 0.00 & 0.00 & 0.01 & 0.99 \end{pmatrix}$$

Let's see if we can do better

Preliminary observation:

Fact. If $(\mathcal{P}(\mathsf{X}), \Pi^i)$ is globally stable for some $i \in \mathbb{N}$, then $(\mathcal{P}(\mathsf{X}), \Pi)$ is also globally stable

Recall: If

1. dynamical system (M, g^i) is globally stable for some $i \in \mathbb{N}$
2. g is continuous at the fixed point of g^i

then (M, g) is also globally stable

Moreover, $\psi \mapsto \psi\Pi$ is everywhere continuous as already discussed

Theorem. If X is finite and Π is both aperiodic and irreducible, then Π is globally stable

Proof: It suffices to show that

$$\forall\, x, y \in \mathsf{X} \times \mathsf{X}, \quad \exists\, i_{x,y} \in \mathbb{N} \text{ s.t. } k \geqslant i_{x,y} \implies \Pi^k(x, y) > 0$$

Indeed, if this statement holds, then

$$i := \max\{i_{x,y}\} \quad \implies \quad \Pi^i(x, y) > 0 \quad \text{for all} \quad (x, y) \in \mathsf{X} \times \mathsf{X}$$

Implies that

- $(\mathcal{P}(\mathsf{X}), \Pi^i)$ is globally stable
- and hence $(\mathcal{P}(\mathsf{X}), \Pi)$ is globally stable

So fix $x, y \in \mathsf{X} \times \mathsf{X}$ and let's try to show that

$$\exists\, i = i_{x,y} \in \mathbb{N} \text{ s.t. } k \geqslant i \implies \Pi^k(x, y) > 0$$

Since Π is irreducible, $\exists\, j \in \mathbb{N}$ such that $\Pi^j(x, y) > 0$

Since Π is aperiodic, $\exists\, m \in \mathbb{N}$ such that

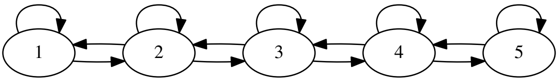
$$\ell \geqslant m \implies \Pi^\ell(y, y) > 0$$

Picking $\ell \geqslant m$ and applying the Chapman–Kolmogorov equation, we have

$$\Pi^{j+\ell}(x, y) = \sum_{z \in \mathsf{X}} \Pi^j(x, z)\Pi^\ell(z, y) \geqslant \Pi^j(x, y)\Pi^\ell(y, y) > 0$$

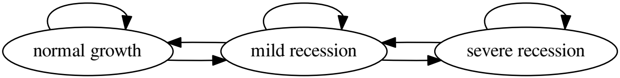
QED

Example. Quah's stochastic kernel is both irreducible and aperiodic



And therefore globally stable

Same with Hamilton's business cycle model



```
In [1]: import quantecon as qe

In [2]: P = [[0.971 , 0.029 , 0],
   ...:       [0.145 , 0.778 , 0.077],
   ...:       [0 , 0.508 , 0.492]]

In [3]: mc = qe.MarkovChain(P)

In [4]: mc.is_aperiodic
Out[4]: True

In [5]: mc.is_irreducible
Out[5]: True

In [6]: mc.stationary_distributions
Out[6]: array([[ 0.8128 ,  0.16256,  0.02464]])
```

A Weaker Set of Conditions

Let Π be a stochastic kernel on (finite set) X

Theorem. The following statements are equivalent:

1. Π^k has a strictly positive column for some $k \in \mathbb{N}$
2. For any $x, x' \in$ X, there exists a $k \in \mathbb{N}$ and a $y \in$ X such that

$$\Pi^k(x, y) > 0 \text{ and } \Pi^k(x', y) > 0$$

3. $(\mathcal{P}(\mathsf{X}), \Pi)$ is globally stable

Intuition for sufficiency

We know a stationary distribution exists, just need to prove convergence

Suppose that, for any $x, x' \in \mathsf{X}$, there exists a $k \in \mathbb{N}$ and a $y \in \mathsf{X}$ such that

$$\Pi^k(x, y) > 0 \text{ and } \Pi^k(x', y) > 0$$

Wherever we are now, we can meet up again

Hence no one is stuck at a local attractor

Initial conditions don't matter in the long run

Hence $(\mathcal{P}(\mathsf{X}), \Pi)$ is globally stable

Application: Inventory Dynamics

Let X_t = inventory of a product, obeys

$$X_{t+1} = \begin{cases} (X_t - D_{t+1})^+ & \text{if } X_t > s \\ (S - D_{t+1})^+ & \text{if } X_t \leqslant s \end{cases}$$

Assume $\{D_t\} \overset{\text{IID}}{\sim}$ the geometric distribution, say

A Markov chain on $\mathsf{X} := \{0, 1, \ldots, S\}$ with kernel

$$\Pi(x, y) = \begin{cases} \mathbb{P}\{(x - D_{t+1})^+ = y\} & \text{if } x > s \\ \mathbb{P}\{(S - D_{t+1})^+ = y\} & \text{if } x \leqslant s \end{cases}$$

Proposition The pair $(\mathcal{P}(\mathsf{X}), \Pi)$ is globally stable

Proof: Suppose that $D_{t+1} \geqslant S$

Then

$$0 \leqslant X_{t+1} \leqslant (S - D_{t+1})^+ = 0$$

Hence $\mathbb{P}\{D_{t+1} \geqslant S\} > 0$ implies $\Pi(x, 0) > 0$ for all x

Moreover $\mathbb{P}\{D_{t+1} \geqslant S\} > 0$ holds for the geometric distribution

Hence $(\mathcal{P}(\mathsf{X}), \Pi)$ is globally stable

The Law of Large Numbers

Fix $h \in \mathbb{R}^{\mathsf{X}}$ and let $\{X_t\}$ be a Markov chain generated by Π

Theorem. If X is finite and Π is globally stable with stationary distribution ψ^*, then

$$\mathbb{P}\left\{ \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} h(X_t) = \sum_{x \in \mathsf{X}} h(x)\psi^*(x) \right\} = 1$$

Intuition: $\{X_t\}$ "almost" identically distributed for large t

Also, stability means that initial conditions die out — a form of long run independence

An approximation of the IID property used in the classical LLN

LLN provides a new **interpretation** for the stationary distribution

Using the LLN with $h(x) = \mathbb{1}\{x = y\}$, we have

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{1}\{X_t = y\} \to \sum_{x \in \mathsf{X}} \mathbb{1}\{x = y\} \psi^*(x) = \psi^*(y)$$

Turning this around,

$$\psi^*(y) \approx \text{ fraction of time that } \{X_t\} \text{ spends in state } y$$

This is **not** always valid **unless** the chain in question is stable

Deterministic Linear Models

Linear vector valued dynamic models — workhorse of macro

- Often used as a building block for more complex models
- Even nonlinear models can often be mapped into linear systems (at cost of higher dimensionality)

Our generic (deterministic) linear model specification on $\mathbb{R}^n$ is

$$x_{t+1} = Ax_t + b \tag{1}$$

where

- x_t is $n \times 1$, a vector of **state variables**
- A is $n \times n$ and b is $n \times 1$

Maps to the dynamical system $(\mathbb{R}^n, g)$ with $g(x) = Ax + b$

When is it stable?

Example. Consider $n = 1$ and $g(x) = ax + b$ for scalars a and b

If $a \neq 1$, then g has a unique fixed point $x^* = b/(1-a)$

Moreover, iterating backwards,

$$x_t = a^t x_0 + b \sum_{i=0}^{t-1} a^i$$

- Converges to $b/(1-a)$ whenever $|a| < 1$
- Hence $(\mathbb{R}, g)$ is globally stable whenever $|a| < 1$

In the general n dimensional case,

$$x_t = Ax_{t-1} + b = A(Ax_{t-2} + b) + b = A^2 x_{t-2} + Ab + b = \cdots$$

Leads to

$$x_t = g^t(x_0) = A^t x_0 + \sum_{i=0}^{t-1} A^i b \tag{2}$$

Does this sequence converge as t gets large?

Does g have a fixed point?

What is the correct generalization of the condition $|a| < 1$ from the scalar case?

Fact. If $r(A) < 1$, then $(\mathbb{R}^n, g)$ is globally stable with steady state

$$x^* = \sum_{i=0}^{\infty} A^i b \tag{3}$$

Proof: $r(A) < 1 \implies (I - A)^{-1} = \sum_{i=0}^{\infty} A^i$

Hence x^* in (3) is the unique solution to $x = Ax + b$

Regarding stability, given x_0 and y_0 in $\mathbb{R}^n$,

$$\|x_t - y_t\| = \|A^t(x_0 - y_0)\| \leqslant \|A^t\| \cdot \|x_0 - y_0\|$$

- But $\|A^t\| \to 0$, so $\|x_t - y_t\| \to 0$
- Taking $y_0 = x^*$ completes the proof

Example. The Samuelson **multiplier–accelerator model**

Consumption obeys

$$C_t = \alpha Y_{t-1} + \gamma$$

Aggregate investment increases with output growth:

$$I_t = \beta(Y_{t-1} - Y_{t-2})$$

Letting G be a constant level of government spending and using the accounting identity

$$Y_t = C_t + I_t + G$$

Combining equations gives

$$Y_t = (\alpha + \beta) Y_{t-1} - \beta Y_{t-2} + G + \gamma \tag{4}$$

This is **not** a first order system

But we **can** map it to the first order framework by taking

$$x_t := \begin{pmatrix} Y_t \\ Y_{t-1} \end{pmatrix}, \quad A := \begin{pmatrix} \alpha + \beta & -\beta \\ 1 & 0 \end{pmatrix}, \quad \text{and} \quad b := \begin{pmatrix} G + \gamma \\ 0 \end{pmatrix}$$

We can recover (4) from the first entry in

$$x_{t+1} = A x_t + b$$

Stability depends on $r(A)$

Step 1: Solve $\det(A - \lambda I) = 0$

Letting $\rho_1 := a + b$ and $\rho_2 := -b$, the two solutions are the roots of

$$\lambda^2 - \rho_1 \lambda - \rho_2 = 0$$

Hence

$$\lambda_i = \frac{\rho_1 \pm \sqrt{\rho_1^2 + 4\rho_2}}{2} \qquad i = 1, 2$$

If both are interior to the unit circle in $\mathbb{C}$, then $r(A) < 1$

In the next fig, $\alpha = 0.6$ and $\beta = 0.7$, so $r(A) \approx 0.837$

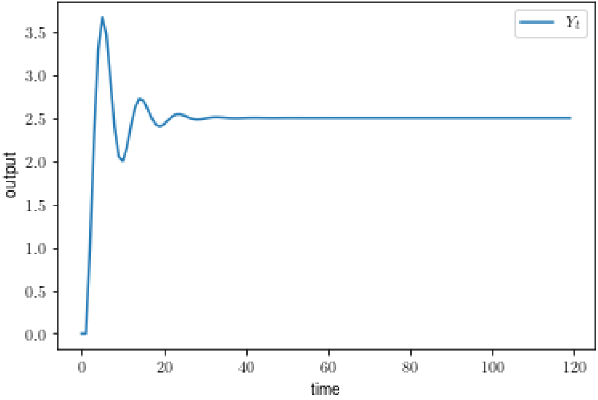


Figure: Time series of output

Adding Stochastic Components

Now we wish to add shocks to the model — get closer to data

Before that let's review some building blocks:

- Conditional expectations
- Martingales
- Martingale difference sequences

Let Y and $\mathscr{G} := \{X_1, \ldots, X_k\}$ be random variables with finite second moments

Problem: Predict Y given $\mathscr{G}$

- In this context, $\mathscr{G}$ is called an **information set**

Thus, we seek a function $f : \mathbb{R}^k \to \mathbb{R}$ such that

$$\hat{Y} := f(X_1, \ldots, X_k) \text{ is a good predictor of } Y$$

"Good" defined to mean that $\mathbb{E}[(\hat{Y} - Y)^2]$ is small

Thus, we seek $\hat{f}$ that solves

$$\hat{f} = \operatorname*{argmin}_{f} \mathbb{E}[(Y - f(X_1, \ldots, X_k))^2]$$

Fact. There exists an (almost everwhere) unique $\hat{f}$ in the set of Borel measurable functions from $\mathbb{R}^k$ to $\mathbb{R}$ that solves

$$\hat{f} = \operatorname*{argmin}_f \mathbb{E}[(Y - f(X_1, \ldots, X_k))^2]$$

We call the resulting variable

$$\hat{Y} := \hat{f}(X_1, \ldots, X_k)$$

the **conditional expectation** of Y given $\mathscr{G}$

Common alternative notation:

$$\mathbb{E}_{\mathscr{G}} Y =: \mathbb{E}[Y \mid \mathscr{G}] =: \mathbb{E}[Y \mid X_1, \ldots, X_k]$$

The definition extends to RVs with finite **first** moment — details omitted

We say that Y is **$\mathscr{G}$-measurable** if there exists a Borel measurable function f such that $Y = f(X_1, \ldots, X_k)$

- Meaning: Y is perfectly predictable given the data in $\mathscr{G}$

Fact. Let X and Y be random variables with finite first moment, let α and β be scalars, and let $\mathscr{G}$ and $\mathscr{H}$ be information sets

The following properties hold:

1. $\mathbb{E}_{\mathscr{G}}[\alpha X + \beta Y] = \alpha \mathbb{E}_{\mathscr{G}} X + \beta \mathbb{E}_{\mathscr{G}} Y$
2. If $\mathscr{G} \subset \mathscr{H}$, then $\mathbb{E}_{\mathscr{G}}[\mathbb{E}_{\mathscr{H}} Y] = \mathbb{E}_{\mathscr{G}} Y$ and $\mathbb{E}[\mathbb{E}_{\mathscr{G}} Y] = \mathbb{E} Y$
3. If Y is independent of the variables in $\mathscr{G}$, then $\mathbb{E}_{\mathscr{G}} Y = \mathbb{E} Y$
4. If Y is $\mathscr{G}$-measurable, then $\mathbb{E}_{\mathscr{G}} Y = Y$
5. If X is $\mathscr{G}$-measurable, then $\mathbb{E}_{\mathscr{G}}[XY] = X \mathbb{E}_{\mathscr{G}} Y$

Let

- $Y = (Y_1, \ldots, Y_m)$ be a vector
- $\mathscr{G}$ be an information set

The **(vector valued) conditional expectation** of Y given $\mathscr{G}$ is just the vector containing the conditional expectation of each element

Thus, written as column vectors,

$$\mathbb{E}_{\mathscr{G}} \begin{pmatrix} Y_1 \\ \vdots \\ Y_m \end{pmatrix} = \begin{pmatrix} \mathbb{E}_{\mathscr{G}} Y_1 \\ \vdots \\ \mathbb{E}_{\mathscr{G}} Y_m \end{pmatrix}$$

(Same as ordinary unconditional expectation for vectors)

Martingales

A **filtration** is an increasing sequence of information sets $\{\mathscr{G}_t\}_{t \geqslant 0}$

- Increasing in set inclusion, so that $\mathscr{G}_t \subset \mathscr{G}_{t+1}$ for all $t \geqslant 0$

Example. If $\{\xi_t\}_{t \geqslant 0}$ is a stochastic process, then the **filtration generated by** $\{\xi_t\}_{t \geqslant 0}$ is

$$\mathscr{G}_t = \{\xi_0, \ldots, \xi_t\} \qquad t \geqslant 0$$

A stochastic process $\{\eta_t\}$ is said to be **adapted** to filtration $\mathscr{G}_t$ if η_t is $\mathscr{G}_t$-measurable for all t

- time t value is revealed by time t information.

A stochastic process $\{w_t\}_{t \geqslant 1}$ taking values in $\mathbb{R}^n$ is called a **martingale** with respect to a filtration $\{\mathscr{G}_t\}$ if

- $\mathbb{E}\|w_t\|_1 < \infty$ and
- $\{w_t\}_{t \geqslant 1}$ is adapted to $\{\mathscr{G}_t\}$
- and

$$\mathbb{E}[w_{t+1} \,|\, \mathscr{G}_t] = w_t, \quad \forall\, t \geqslant 1$$

Example. Consider a scalar **random walk** $\{w_t\}$ defined by

$$w_t = \sum_{i=1}^{t} \xi_i, \qquad \{\xi_t\} \text{ is IID with } \mathbb{E}[\xi_t] = 0$$

This process is a martingale with respect to the filtration generated by $\{\xi_t\}$, since

1. adapted
2. satisfies

$$\mathbb{E}[w_{t+1} \,|\, \mathscr{G}_t] = \mathbb{E}[w_t + \xi_{t+1} \,|\, \mathscr{G}_t] = \mathbb{E}[w_t \,|\, \mathscr{G}_t] + \mathbb{E}[\xi_{t+1} \,|\, \mathscr{G}_t]$$

The martingale property now follows (why?)

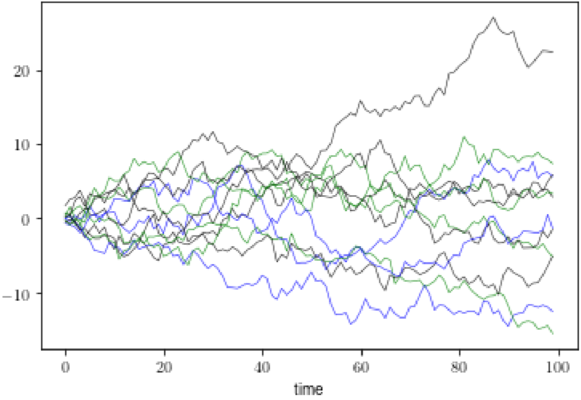


Figure: Twelve realizations of a random walk

A stochastic process $\{w_t\}_{t \geqslant 1}$ in $\mathbb{R}^n$ is called a **martingale difference sequence** (MDS) with respect to a filtration $\{\mathscr{G}_t\}$ if

- $\mathbb{E}\|w_t\|_1 < \infty$ and
- $\{w_t\}_{t \geqslant 1}$ is adapted to $\{\mathscr{G}_t\}$
- and

$$\mathbb{E}[w_{t+1} \mid \mathscr{G}_t] = 0, \quad \forall\, t \geqslant 1$$

Example. If $\{v_t\}$ is a martingale with respect to $\{\mathscr{G}_t\}$ then $w_t := v_t - v_{t-1}$ is an MDS with respect to $\{\mathscr{G}_t\}$

Proof: For any t,

$$\mathbb{E}[w_{t+1} \,|\, \mathscr{G}_t] = \mathbb{E}[v_{t+1} - v_t \,|\, \mathscr{G}_t]$$

$$= \mathbb{E}[v_{t+1} \,|\, \mathscr{G}_t] - \mathbb{E}[v_t \,|\, \mathscr{G}_t] = v_t - v_t = 0$$

Example. If $\{v_t\}$ is IID with zero mean and $\{\mathscr{G}_t\}$ is the filtration generated by $\{v_t\}$, then $\{v_t\}$ is an MDS with respect to $\{\mathscr{G}_t\}$

Ex. Check it

An MDS is additive white noise:

Fact. If $\{w_t\}$ is an MDS with respect to $\{\mathscr{G}_t\}$, then

$$\mathbb{E}[w_t] = 0 \text{ for all } t \geqslant 0$$

Ex. Check it

Fact. If $\{w_t\}$ is an MDS with respect to $\{\mathscr{G}_t\}$, then w_s and w_t are **orthogonal**, in the sense that

$$\mathbb{E}[w_s w_t'] = 0 \text{ whenever } s \neq t$$

Ex. Check it

Linear Vector Systems with Noise

Next consider

- $x_{t+1} = Ax_t + b + C\xi_{t+1}$ with x_0 given
- $\mathscr{G}_t = \{x_0, \xi_0, \xi_1, \ldots, \xi_t\}$
- $\{\xi_t\}_{t \geqslant 1}$ is an $\mathbb{R}^j$-valued MDS with respect to $\mathscr{G}_t$ satisfying

$$\mathbb{E}[\xi_t \xi_t'] = I$$

An example of a **vector autoregressive (VAR) process**

What are the dynamics of the state process $\{x_t\}$?

This is a multi-layered question so let's start with an easy component

What is the time path of the first two moments?

These are

- $\mu_t := \mathbb{E}[x_t]$
- $\Sigma_t := \text{var}[x_t] := \mathbb{E}[(x_t - \mu_t)(x_t - \mu_t)']$

Dynamics of the Mean

First, regarding μ_t, take expectations over

$$x_{t+1} = Ax_t + b + C\xi_{t+1}$$

to get

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

Fact. If $r(A) < 1$, then $\{\mu_t\}$ converges to the unique fixed point

$$\mu^* = \sum_{i=0}^{\infty} A^i b$$

regardless of μ_0

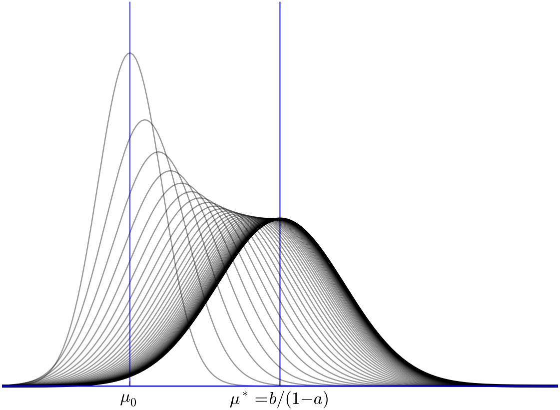


Figure: Convergence of μ_t to μ^* in the scalar model

Dynamics of the Variance

Consider again

$$x_{t+1} = Ax_t + b + C\xi_{t+1}$$

We want a similar law of motion for $\Sigma_t := \mathrm{var}[x_t]$

We will use the fact that $\mathbb{E}[x_t \xi'_{t+1}] = 0$

Ex. Show this follows from the assumptions above

By definition,

$$\begin{aligned}\operatorname{var}[x_{t+1}] &= \mathbb{E}[(x_{t+1} - \mu_{t+1})(x_{t+1} - \mu_{t+1})'] \\ &= \mathbb{E}[(A(x_t - \mu_t) + C\xi_{t+1})(A(x_t - \mu_t) + C\xi_{t+1})']\end{aligned}$$

The right hand side is equal to

$$\begin{aligned}\mathbb{E}[A(x_t - \mu_t)(x_t - \mu_t)'A'] + \mathbb{E}[A(x_t - \mu_t)\xi_{t+1}'C'] \\ + \mathbb{E}[C\xi_{t+1}(x_t - \mu_t)'A'] + \mathbb{E}[C\xi_{t+1}\xi_{t+1}'C']\end{aligned}$$

Some further manipulations (check) lead to

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

To repeat

$$\Sigma_{t+1} = g(\Sigma_t) \quad \text{where} \quad S(\Sigma) := A\Sigma A' + CC'$$

Variance is a trajectory of the dynamical system $(\mathcal{M}(n \times n), S)$

A steady state of this system is a Σ satisfying

$$\Sigma = A\Sigma A' + CC'$$

Fact. If $r(A) < 1$, then $(\mathcal{M}(n \times n), S)$ is **globally stable**

- distance is generated by the (spectral) norm on $\mathcal{M}(n \times n)$

In the proof, we use the following extension of Banach's fixed point theorem

Theorem. Let T be a self-mapping on complete metric space M such that

1. T^k is a Banach contraction mapping on M for some $k \in \mathbb{N}$
2. T is continuous on M

Then (M, T) is globally stable

Ex. Verify this based on our results for dynamical systems

Consider the **discrete Lyapunov equation**

$$\Sigma = A \Sigma A' + M$$

- all matrices are in $\mathcal{M}(n \times n)$ and Σ is the unknown

Given A and M, let ℓ be the **Lyapunov operator**

$$\ell(\Sigma) = A \Sigma A' + M$$

Ex. Show that ℓ is continuous on $\mathcal{M}(n \times n)$

Fact. If $r(A) < 1$, then $(\mathcal{M}(n \times n), \ell)$ is globally stable

Proof: Suffices to show that ℓ^k is a Banach contraction on $(\mathcal{M}(n \times n), \| \cdot \|)$ for some $k \in \mathbb{N}$

From the definition,

$$\ell^k(\Sigma) = A^k \Sigma (A^k)' + A^{k-1} M (A^{k-1})' + \cdots + M$$

Hence, for any Σ, Λ in $\mathcal{M}(n \times n)$, we have

$$\begin{aligned} \|\ell^k(\Sigma) - \ell^k(\Lambda)\| &= \left\| A^k \Sigma (A^k)' - A^k \Lambda (A^k)' \right\| \\ &= \left\| A^k (\Sigma - \Lambda)(A^k)' \right\| \\ &\leqslant \|A^k\| \cdot \|\Sigma - \Lambda\| \cdot \|(A^k)'\| \end{aligned}$$

Transposes don't change norms, so $\|(A^k)'\| = \|A^k\|$ and hence

$$\|\ell^k(\Sigma) - \ell^k(\Lambda)\| \leqslant \|A^k\|^2 \|\Sigma - \Lambda\|$$

Since $r(A) < 1$, we can find $k \in \mathbb{N}$, $\lambda < 1$ such that

$$\|\ell^k(\Sigma) - \ell^k(\Lambda)\| \leqslant \lambda \|\Sigma - \Lambda\| \quad \text{for all } \Sigma, \Lambda \in \mathcal{M}(n \times m)$$

Now apply Banach contraction mapping theorem

Note: Gives an algorithm for computing Σ^*

Application: Log Output

Kydland and Prescott (1980) study detrended log output via

$$y_{t+1} = \alpha_1 y_t + \alpha_2 y_{t-1} + \epsilon_{t+1}, \qquad \{\epsilon\} \overset{\text{IID}}{\sim} N(0, \sigma) \tag{5}$$

We can map it to our VAR framework $x_{t+1} = Ax_t + b + C\xi_{t+1}$ via

$$x_t := \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix}, \quad A := \begin{pmatrix} \alpha_1 & \alpha_2 \\ 1 & 0 \end{pmatrix}, \quad C := \begin{pmatrix} \sigma \\ 0 \end{pmatrix}$$

along with $\xi_t := \frac{1}{\sigma}\epsilon_t$

Estimated values: $\hat{\alpha}_1 = 1.386$ and $\hat{\alpha}_2 = -0.477$

Implies $r(A) \approx 0.75$

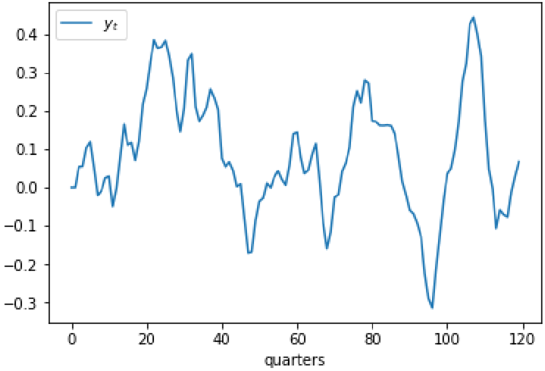


Figure: Time series of detrended log output

Distribution Dynamics: The Gaussian Case

We have obtained the moment dynamics of

$$x_{t+1} = Ax_t + b + C\xi_{t+1} \tag{6}$$

They were

- $g^t(\mu_0)$ where $g(\mu) := A\mu + b$ on $\mathbb{R}^n$
- $S^t(\Sigma_0)$ where $S(\Sigma) := A'\Sigma A + CC'$ on $\mathcal{M}(n \times n)$

Now we want to learn about the distributions themselves

That is, we wish to track $\{\psi_t\}$ where

$$\psi_t := \text{ the distribution of } x_t$$

This is straightforward if the model is Gaussian

- Gaussian distributions described by their first two moments

We can give a complete analytical description of the marginal distributions $\{\psi_t\}$

Works because

- Linear combinations of mutivariate Gaussians are Gaussian
- Our law of motion $x_{t+1} = Ax_t + b + C\xi_{t+1}$ is linear

A scalar random variable z has a (univariate) **standard normal distribution** if

$$z \stackrel{\mathscr{D}}{=} \varphi \text{ where } \varphi(s) = \sqrt{\frac{1}{2\pi}} \exp\left(\frac{-s^2}{2}\right) \qquad (s \in \mathbb{R})$$

We write $z \stackrel{\mathscr{D}}{=} N(0,1)$.

Scalar random variable x has **normal distribution** $N(\mu, \sigma)$ for some $\mu \in \mathbb{R}$ and $\sigma \geqslant 0$ if

$$x \stackrel{\mathscr{D}}{=} \mu + \sigma z, \text{ for some } z \text{ with } z \stackrel{\mathscr{D}}{=} N(0,1)$$

Note that we allow $\sigma = 0$, in which case x is a point mass on μ

A random vector x in $\mathbb{R}^n$ is called **multivariate Gaussian** with distribution $N(\mu, \Sigma)$ if

- μ is a vector in $\mathbb{R}^n$
- Σ is a positive semidefinite element of $\mathcal{M}(n \times n)$ and
- $h'x \stackrel{\mathscr{D}}{=} N(h'\mu, h'\Sigma h)$ on $\mathbb{R}$ for any $h \in \mathbb{R}^n$

If Σ is positive definite, then x has density

$$\varphi(s) = \det(2\pi\Sigma)^{-1/2} \exp\left(-\frac{1}{2}(s-\mu)'\Sigma^{-1}(s-\mu)\right)$$

Question: If x_1 and x_2 are normally distributed in $\mathbb{R}$, is $x = (x_1, x_2)$ multivariate Gaussian?

To shift to the Gaussian case we assume that

- $\{\xi_t\}_{t \geqslant 1} \overset{\text{IID}}{\sim} N(0, I)$ and
- $x_0 \overset{\mathscr{D}}{=} N(\mu_0, \Sigma_0)$
- μ_0 is any vector in $\mathbb{R}^j$ and Σ_0 is positive semidefinite

The random vector x_0 is assumed to be independent of $\{\xi_t\}$

Under these Gaussian conditions we have

$$x_t \overset{\mathscr{D}}{=} N\left(g^t(\mu_0), S^t(\Sigma_0)\right) \text{ for all } t \geqslant 0 \tag{7}$$

Ex. Check normality using the definition of multivariate Gaussians

Proposition. If $r(A) < 1$, then under the Gaussian conditions we have

$$\psi_t \xrightarrow{w} N(\mu^*, \Sigma^*) \qquad (t \to \infty) \tag{8}$$

where

- $\xrightarrow{w}$ means **weak convergence** (convergence "in distribution")
- $\psi_t \stackrel{\mathscr{D}}{=} x_t$
- $\mu^* = \sum_{i=0}^{\infty} A^i b$ and
- Σ^* is the unique fixed point of $\Sigma := A' \Sigma A + CC'$

Equivalent to (8): the characteristic function of $N(\mu_t, \Sigma_t)$ converges pointwise to that of $N(\mu^*, \Sigma^*)$

Proof: We must show that, at any fixed $s \in \mathbb{R}^n$,

$$\lim_{t \to \infty} \exp\left(is'\mu_t - \frac{1}{2}s'\Sigma_t s\right) = \exp\left(is'\mu^* - \frac{1}{2}s'\Sigma^* s\right) \qquad (9)$$

Fixing such an s, to prove (9) it suffices to show that

$$s'\mu_t \to s'\mu^* \quad \text{and} \quad s'\Sigma_t s \to s'\Sigma^* s \quad \text{in } \mathbb{R} \text{ as } t \to \infty \qquad (10)$$

From the Cauchy–Schwarz inequality we have

$$|s'\mu_t - s'\mu^*| = |s'(\mu_t - \mu^*)| \leqslant \|s\| \cdot \|\mu_t - \mu^*\| \to 0$$

Ex. Prove the second part of (10)

Example. In the **AR(1)** case, $\{x_t\}$ is real valued and obeys

$$x_{t+1} = a x_t + b + \sigma \epsilon_{t+1}, \quad \{\epsilon_t\} \overset{\text{IID}}{\sim} N(0,1) \tag{11}$$

In this case $r(A) = |a|$

The stable case $|a| < 1$ is called **mean-reverting**

The distribution of x_t converges weakly to

$$\psi^* := N\left(\frac{b}{1-a}, \frac{\sigma^2}{1-a^2}\right) \tag{12}$$

Dynamical systems formulation

Let $\mathscr{G}$ be the set of all Gaussian distributions on $\mathbb{R}^n$

- topology $=$ weak convergence

Let Π be the operator on $\mathscr{G}$ defined by

$$\psi := N(\mu, \Sigma) \mapsto \psi\Pi := N(g(\mu), S(\Sigma))$$

Then

- Π is a self-mapping on $\mathscr{G}$
- $(\mathscr{G}, \Pi)$ is globally stable whenever $r(A) < 1$

Distribution Dynamics: The General Density Case

Let's drop the Gaussian assumptions, replace them with

- $\{\xi_t\}$ is IID on $\mathbb{R}^n$ with density φ
- C is $n \times n$ and nonsingular

Under these assumptions, each ψ_t will be a density

To prove this we use

Fact. If ξ has density φ and C is nonsingular, then $y = d + C\xi$ has density

$$p(y) = \varphi\left(C^{-1}(y - d)\right) |\det C|^{-1} \tag{13}$$

The density of x_{t+1} conditional on $x_t = x$ is therefore

$$\pi(x, y) = \varphi\left(C^{-1}(y - Ax - b)\right) |\det C|^{-1}$$

The law of total probability tells us that, for random varables (x, y) with densities,

$$p(y) = \int p(y \,|\, x) p(x) \, \mathrm{d}x$$

Hence the densities ψ_t and ψ_{t+1} are connected via

$$\psi_{t+1}(y) = |\det C|^{-1} \int \varphi\left(C^{-1}(y - Ax - b)\right) \psi_t(x) \, \mathrm{d}x$$

Suppose we introduce an operator Π from the set of densities $\mathcal{D}$ on $\mathbb{R}^n$ to itself via

$$(\psi\Pi)(y) = \int \pi(x,y)\psi(x)\,\mathrm{d}x \tag{14}$$

Then our law of motion for marginals

$$\psi_{t+1}(y) = |\det C|^{-1} \int \varphi\left(C^{-1}(y - Ax - b)\right)\psi_t(x)\,\mathrm{d}x$$

becomes

$$\psi_{t+1} = \psi_t \Pi$$

- a concise description of distribution dynamics

Comments:

- In $\psi_{t+1} = \psi_t \Pi$ we write the argument to the left following tradition (see Meyn and Tweedie, 2009)
- The set of densities $\mathcal{D}$ is endowed with the topology of weak convergence

Proposition. If $r(A) < 1$, then $(\mathcal{D}, \Pi)$ is globally stable

Moreover, if h is any function such that $\int |h(x)| \psi^*(x) \, \mathrm{d}x$ is finite, then

$$\mathbb{P} \left\{ \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} h(x_t) = \int h(x) \psi^*(x) \, \mathrm{d}x \right\} = 1$$