Topics in Computational Economics

Lecture 9

John Stachurski

NYU 2016

Preliminary Comments

HW was very good on average

See

http://nbviewer.jupyter.org/github/jstac/quantecon_nyu_2016_homework/tree/master/hw_set_6/

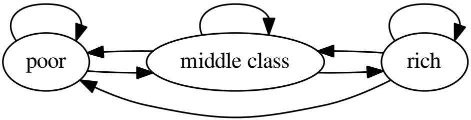
Small hint: When you prove something, you probably need to be using all the assumptions

Today's Lecture

- Stochastic kernels and Markov chains
- Distribution dynamics
- Aperiodicity and irreducibility
- Stability
- Ergodicity

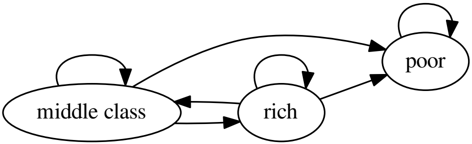
Prequel 1: Directed Graphs

A **directed graph** is a nonempty set of **nodes** $S = \{x, y, \ldots, z\}$ and a set of **directed edges** $(x, y) \in S \times S$



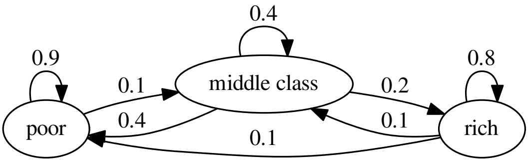
- y is called **accessible** from x if $y = x$ or $\exists$ a sequence of directed edges leading from x to y
- The graph is called **strongly connected** if y is accessible from x for all $x, y \in S$

Another example



- poor is accessible from middle
- rich is not accessible from poor
- Graph is not strongly connected

We can also attach numbers to the edges of a directed graph



The resulting graph is called a **weighted directed graph**

Interpretation will be given later

Prequel 2: Schauder's Fixed Point Theorem

Let V be Banach space and let $T\colon V \to V$

Theorem. (Schauder, 1930) If

1. C is a convex compact subset of V
2. T is continuous and invariant on C

then T has at least one fixed point in C

(T is **invariant on** C if $x \in C$ implies $Tx \in C$)

Remarks: If V is finite dimensional, then

- compact $=$ closed and bounded
- Schauder $=$ Brouwer

Prequel 3: $\ell_1(S)$ and the Space of Distributions

Let S be any discrete (countable/finite) set

- e.g., $\{0, \dots, N\}$, $\{0\} \cup \mathbb{N}$, $\mathbb{Z}$, $\mathbb{Z}^d$, etc.
- called the **state space** in what follows

The set of **distributions** on S is denoted $\mathscr{P}(S)$ and defined as all $\phi \in \mathbb{R}^S$ such that

- $\phi(x) \geqslant 0$ for all $x \in S$
- $\sum_{x \in S} \phi(x) = 1$

Think of $\phi(x)$ as probability of hitting x

Let's embed $\mathscr{P}(S)$ in a space with a metric that will work well for certain problems we are interested in

Recall the Banach space $(\ell_1(S), \| \cdot \|_1)$, where

$$\|h\|_1 := \sum_{x \in S} |h(x)| \quad \text{and} \quad \ell_1(S) := \left\{ h \in \mathbb{R}^S : \|h\|_1 < \infty \right\}$$

For the rest of **this lecture** we write

- $\| \cdot \|$ instead of $\| \cdot \|_1$ to simplify notation
- $h \leqslant g$ if $h(x) \leqslant g(x)$ for all $x \in S$

The **positive cone** of $\ell_1(S)$ is the set

$$\ell_1^+(S) := \{h \in \ell_1(S) : h \geqslant 0\}$$

With this notation, we can express the set of distributions as follows

$$\mathscr{P}(S) = \{\phi \in \ell_1^+(S) : \|\phi\| = 1\}$$

Thus $\mathscr{P}(S) =$ positive cone $\cap$ unit sphere of $\ell_1(S)$

Reminder: If $S = \{x_1, \ldots, x_N\}$, then $\ell_1(S)$ is isometrically isomorphic to $\mathbb{R}^N$

$$\ell_1(S) \ni h = \begin{pmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{pmatrix} \quad \leftrightarrow \quad \begin{pmatrix} h_1 \\ \vdots \\ h_N \end{pmatrix} \in \mathbb{R}^N$$

$$\|h\| = \sum_{i=1}^{n} |h(x_i)| \quad \leftrightarrow \quad \sum_{i=1}^{n} |h_i| = \text{ a norm on } \mathbb{R}^N$$

Recall: All norms on $\mathbb{R}^N$ are **equivalent**

Hence topological properties of $\ell_1(S)$ identical to Euclidean space when S finite

- Compact $\iff$ closed and bounded, etc.

In the finite setting, $\mathscr{P}(S)$ "is" the **unit simplex** in $\mathbb{R}^N$

- A convex, compact subset of $\mathbb{R}^N$
- And therefore a convex, compact subset of $(\ell_1(S), \|\cdot\|)$

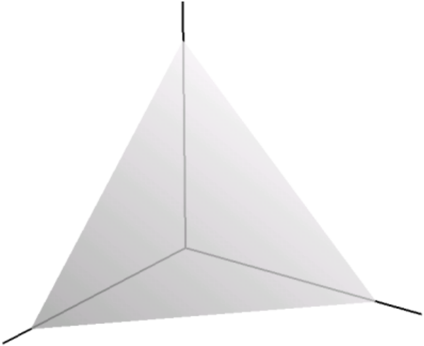
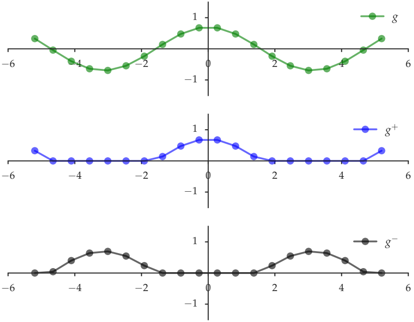


Figure: The unit simplex in $\mathbb{R}^3$

Aside: Given $g \in \ell_1(S)$, let $g^+ := \max\{g, 0\}$ and $g^- := \max\{-g, 0\}$



Ex. Show that, for all $g \in \ell_1(S)$ we have

$$g = g^+ - g^- \quad \text{and} \quad |g| = g^+ + g^-$$

Prequel 4: Markov Operators

A linear map $P\colon \ell_1(S) \to \ell_1(S)$ satisfying

1. $g \geqslant 0 \implies gP \geqslant 0$
2. $g \geqslant 0 \implies \|gP\| = \|g\|$

is called a **Markov operator** on S

In other words, $P\colon \ell_1(S) \to \ell_1(S)$ is a Markov operator if it is

1. **linear**
2. **positive** (i.e., invariant on the positive cone) and
3. **norm preserving on the positive cone**

- Argument written to the left by convention

Properties of Markov Operators

Fact. If P is a Markov operator on S, then P is invariant on $\mathscr{P}(S)$

In other words,

$$\phi \in \mathscr{P}(S) \quad \Longrightarrow \quad \phi P \in \mathscr{P}(S)$$

Proof: Fix $\phi \in \mathscr{P}(S)$

Since $\phi \geqslant 0$ and $\|\phi\| = 1$, we have

1. $\phi P \geqslant 0$ and
2. $\|\phi P\| = \|\phi\| = 1$

Hence $\phi P \in \mathscr{P}(S)$ as claimed

Fact. If $g \in \ell_1(S)$, then $\|gP\| \leqslant \|g\|$

Proof: For $g \in \ell_1(S)$, we have

$$\begin{aligned}
\|gP\| &= \|(g^+ - g^-)P\| \\
&= \|g^+P - g^-P\| \\
&\leqslant \|g^+P\| + \|g^-P\| \\
&= \|g^+\| + \|g^-\| \\
&= \sum_{x \in S} g^+(x) + \sum_{x \in S} g^-(x) \\
&= \sum_{x \in S} |g(x)| = \|g\|
\end{aligned}$$

Fact. If P is a Markov operator on S, then

$$\|gP - hP\| \leqslant \|g - h\| \quad \text{for all} \quad g, h \in \ell_1(S)$$

We say that P is **nonexpansive** on $\ell_1(S)$

Proof: Given $g, h \in \ell_1(S)$, linearity and the contraction property from the last slide give

$$\|gP - hP\| = \|(g - h)P\| \leqslant \|(g - h)\| = \|g - h\|$$

One implication is that P is continuous on $\ell_1(S)$

Indeed, if $g_n \to g$ in $\ell_1(S)$, then

$$\|g_n P - gP\| \leqslant \|g_n - g\| \to 0$$

Stochastic Kernels

A **stochastic kernel** on S is a function $p \colon S \times S \to \mathbb{R}$ such that

$$p(x, \cdot) \in \mathscr{P}(S) \quad \text{for all} \quad x \in S$$

In other words,

1. $p(x, y) \geqslant 0$ for all $(x, y) \in S \times S$
2. $\sum_{y \in S} p(x, y) = 1$ for all $x \in S$

Intuition:

1. We have one distribution $p(x, \cdot)$ for each point $x \in S$
2. $p(x, y)$ is the probability of moving from x to y in one step

Matrix Representation

There are some alternative representations of stochastic kernels

When S is finite, we can represent p by a matrix

$$P = \begin{pmatrix} p(x_1, x_1) & \cdots & p(x_1, x_N) \\ \vdots & & \vdots \\ p(x_N, x_1) & \cdots & p(x_N, x_N) \end{pmatrix}$$

Comments:

1. Square, nonnegative, rows sum to one
2. Distributions are rows, stacked vertically

A matrix satisfying the properties in 1 is called a **Markov matrix**

Example. (Hamilton, 2005)

Estimates a statistical model of the business cycle based on US unemployment data

Markov matrix:

$$P_H := \begin{pmatrix} 0.971 & 0.029 & 0 \\ 0.145 & 0.778 & 0.077 \\ 0 & 0.508 & 0.492 \end{pmatrix}$$

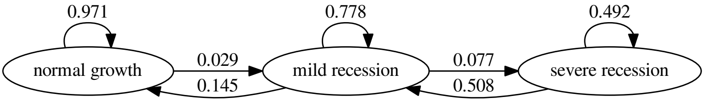
- state 1 $=$ normal growth
- state 2 $=$ mild recession
- state 3 $=$ severe recession

Length of the period $=$ one month

Digraph Representation

Another way to represent a finite stochastic kernel is by a weighted directed graph

Example. Here's Hamilton's business cycle model as a digraph



- set of nodes is S
- no edge means $p(x, y) = 0$

Example. International growth dynamics study of Quah (1993)

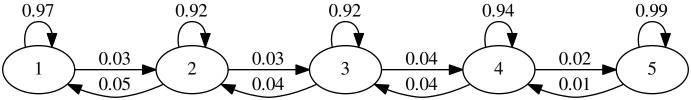
State = real GDP per capita relative to world average

States are 0–1/4, 1/4–1/2, 1/2–1, 1–2 and 2–∞

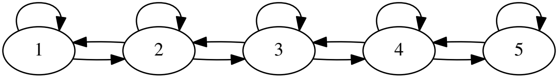
$$P_Q = \begin{pmatrix} 0.97 & 0.03 & 0.00 & 0.00 & 0.00 \\ 0.05 & 0.92 & 0.03 & 0.00 & 0.00 \\ 0.00 & 0.04 & 0.92 & 0.04 & 0.00 \\ 0.00 & 0.00 & 0.04 & 0.94 & 0.02 \\ 0.00 & 0.00 & 0.00 & 0.01 & 0.99 \end{pmatrix}$$

The transitions are over a one year period

Quah's income dynamics model as a weighted directed graph:



Dropping labels gives the directed graph



Markov Operator Representation

Given stochastic kernel p on S, let $P\colon \ell_1(S) \to \ell_1(S)$ be defined by

$$(gP)(y) = \sum_{x \in S} p(x,y)g(x) \qquad (y \in S)$$

Fact. P is a Markov operator on $\ell_1(S)$

Proof: Evidently $g \in \ell_1^+(S)$ implies $gP \in \ell_1^+(S)$

Moreover, if $g \in \ell_1^+(S)$, then

$$\|gP\| = \sum_{y \in S} (gP)(y) = \sum_{x \in S} \sum_{y \in S} p(x,y)g(x) = \sum_{x \in S} g(x) = \|g\|$$

Ex. Show that P is linear

Aside from those created by stochastic kernels, what other kind of Markov operators can we come up with?

Answer: None

Ex. Let

1. P be a Markov operator on $\ell_1(S)$
2. p be defined by

$$p(x,y) := (\delta_x P)(y) \qquad ((x,y) \in S \times S)$$

Show that p is a stochastic kernel on S

Conclusion: The is a one to one correspondence between the set of Markov operators on S and the set of stochastic kernels on S

From Stochastic Kernels to Markov Chains

Let

1. S be a discrete set
2. $\{X_t\}_{t=0}^{\infty}$ be an S-valued stochastic process

$\{X_t\}_{t=0}^{\infty}$ is called a **Markov chain** on S if there exists a stochastic kernel p on S such that

$$\mathbb{P}[X_{t+1} = y \mid X_0, X_1, \ldots, X_t] = p(X_t, y) \quad \text{for all} \quad t \geqslant 0,\ y \in S$$

In this case we say that $\{X_t\}_{t=0}^{\infty}$ is **generated by** p

If $X_0 \sim \psi$, then ψ is called the **initial condition**

Linking Marginals

By the law of total probability we have

$$\mathbb{P}\{X_{t+1} = y\} = \sum_{x \in S} \mathbb{P}\{X_{t+1} = y \,|\, X_t = x\} \cdot \mathbb{P}\{X_t = x\}$$

Letting ψ_t be the distribution of X_t, this becomes

$$\psi_{t+1}(y) = \sum_{x \in S} p(x, y) \psi_t(x) \qquad (y \in S)$$

If P is the Markov operator induced by p, we can write this as

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

Thus the Markov operator **updates the distribution** of the state

Interpretation of Trajectories

Since $\psi_{t+1} = \psi_t P$ for all $t \geqslant 0$, we have

$$\psi_t = \psi_{t-1} P = \psi_{t-2} P P = \psi_{t-2} P^2 = \cdots = \psi_0 P^t$$

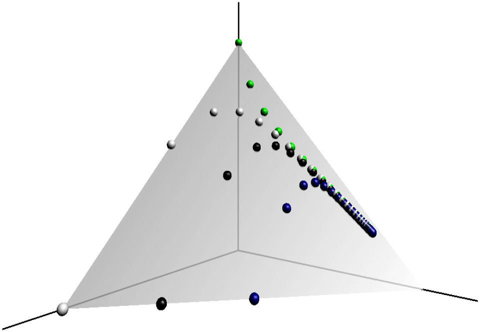
As a result,

- $X_0 \sim \psi \implies X_t \sim \psi P^t$
- $X_0 = x \implies X_t \sim \delta_x P^t$

In practice, for finite models, updates like $\psi_{t+1} = \psi_t P$ are done with matrix multiplication

- Distributions are treated as **row vectors**

Some of trajectories in $\mathscr{P}(S)$ under Hamilton's business cycle model:



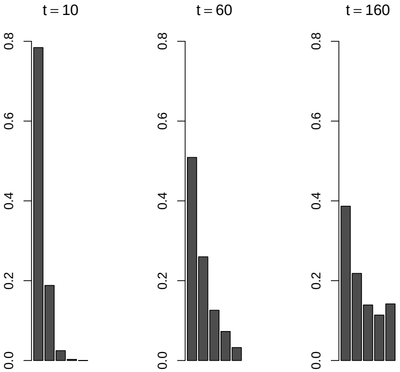


Figure: Distributions from Quah's stochastic kernel, $X_0 = 1$

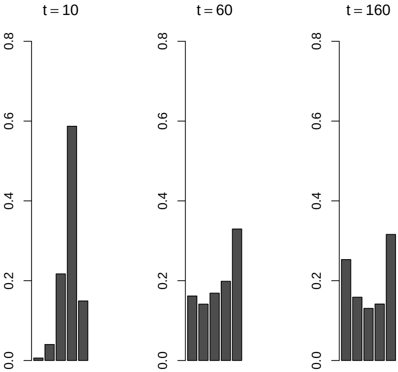


Figure: Distributions from Quah's stochastic kernel, $X_0 = 4$

Higher Order Kernels

Let p be a stochastic kernel and let $\{p^k\}$ be defined inductively by

$$p^1 := p \quad \text{and} \quad p^{k+1}(x,y) := \sum_{z \in S} p(x,z) p^k(z,y)$$

Called the **k-step stochastic kernel**

Fact. If p is a stochastic kernel, then so is p^k for all k

Fact. If $\{X_t\}$ is generated by p, then, for any $k \in \mathbb{N}$, we have

$$p^k(x,y) = \mathbb{P}\{X_k = y \,|\, X_0 = x\} \qquad (x, y \in S)$$

Chapman-Kolmogorov Equations

The kernels $\{p^k\}$ satisfy the **Chapman–Kolmogorov relation**

$$p^{j+k}(x,y) = \sum_{z \in S} p^k(x,z) p^j(z,y) \qquad ((x,y) \in S \times S)$$

Proof: Let $X_0 = x$ and let $y \in S$ be given

By the law of total probability, we have

$$\begin{aligned} p^{j+k}(x,y) &= \mathbb{P}\{X_{j+k} = y\} \\ &= \sum_{z \in S} \mathbb{P}\{X_{j+k} = y \mid X_k = z\} \mathbb{P}\{X_k = z\} \\ &= \sum_{z \in S} p^k(x,z) p^j(z,y) \end{aligned}$$

Stationary Distributions

Let p be a stochastic kernel on S

If $\psi^* \in \mathscr{P}(S)$ satisfies

$$\psi^*(y) = \sum_{x \in S} p(x,y)\psi^*(x) \quad \text{for all} \quad y \in S$$

then ψ^* is called **stationary** or **invariant** for p

Equivalent: ψ^* is a fixed point of the induced Markov operator

Interpretation:

$$X_t \sim \psi^* \implies X_{t+1} \sim \psi^*$$

Existence

Theorem (Krylov–Bogolyubov). If S is finite then p has at least one stationary distribution

Proof: Let P be the Markov operator induced by p and recall that

- $(\ell_1(S), \|\cdot\|)$ is a Banach space
- P is continuous on $(\ell_1(S), \|\cdot\|)$
- $\mathscr{P}(S)$ is a convex subset of $(\ell_1(S), \|\cdot\|)$
- P maps $\mathscr{P}(S)$ into itself

If S is finite, then $\mathscr{P}(S)$ is also compact

Existence of a fixed point follows from the Schauder fixed point theorem

Failure of Existence

Existence can fail when S is infinite

Intuition: Probability mass can diverge

Example. Let $\{X_t\}$ be defined by $X_{t+1} = X_t + 1$ on $\mathbb{Z}$

Ex. Show that

1. The corresponding stochastic kernel is

$$p(x, y) = \mathbb{1}\{x = y - 1\}$$

2. No stationary distribution exists

Probabilistic Properties

Let p be a stochastic kernel on S and let x, y be states

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

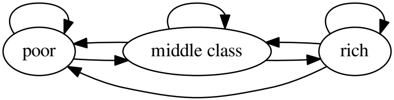
$$\exists\, k \in \mathbb{N} \text{ such that } p^k(x, y) > 0$$

Equivalent: Accessible in the induced directed graph

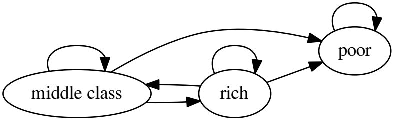
A stochastic kernel p on S is called **irreducible** if every state is accessible from any other

Equivalent: The induced directed graph is strongly connected

Irreducible:



Not irreducible:



Aperiodicity

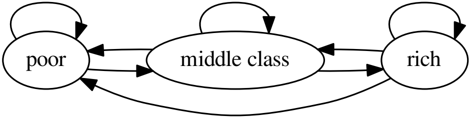
Let p be a stochastic kernel on S

State $x \in S$ is called **aperiodic** under p if

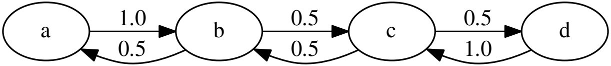
$$\exists\, n \in \mathbb{N} \text{ such that } k \geqslant n \implies p^k(x, x) > 0$$

A stochastic kernel p on S is called **aperiodic** if every state in S is aperiodic under p

Aperiodic:

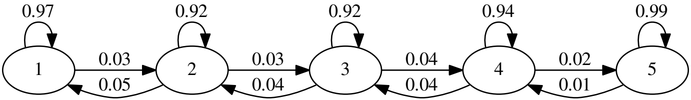


Periodic:



Stability of Markov Chains

Recall the distributions generated by Quah's model



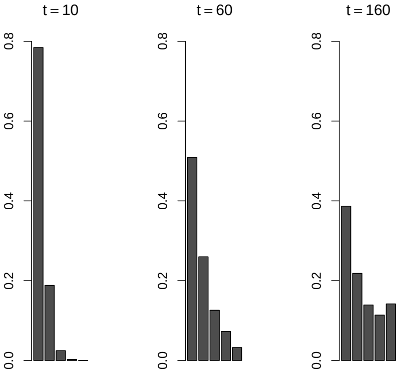


Figure: $X_0 = 1$

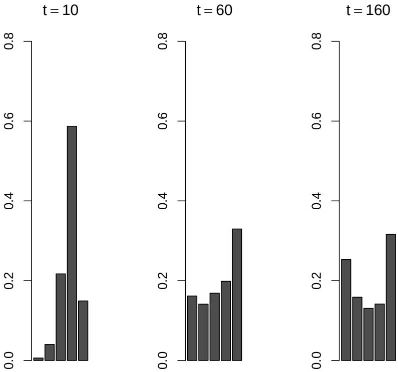


Figure: $X_0 = 4$

What happens when $t \to \infty$?

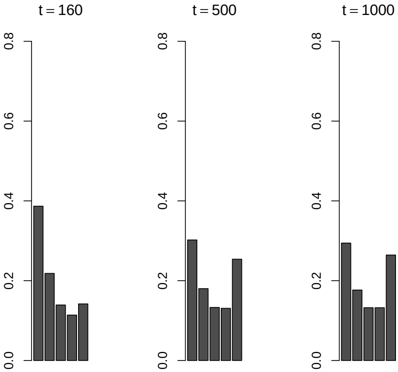


Figure: $X_0 = 1$

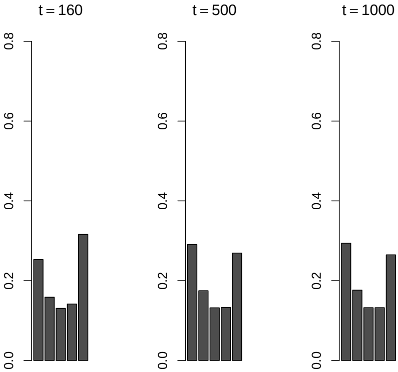


Figure: $X_0 = 4$

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

This suggests we are observing a form of stability

- How to define stability of Markov chains?
- And how to prove it?

A mapping $T \colon M \to M$ on metric space (M, ρ) is called **globally stable** if

1. T has a unique fixed point $x^* \in M$
2. $T^i x \to x^*$ as $i \to \infty$ for any $x \in M$

A stochastic kernel on S is called **globally stable** if the corresponding Markov operator is globally stable on $\mathscr{P}(S)$

Fact. If P is a Markov operator on S and P^n is globally stable for some $n \in \mathbb{N}$, then P is also globally stable

Proof: See Stachurski EDTC (2009), lemma 4.1.21

Or check it yourself (hint: Use the fact that P is nonexpansive)

Of course not all stochastic kernels are globally stable

Example. Let $S = \{1, 2\}$ and consider the periodic Markov chain

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

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

Ex. Show that

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

Conclude that global stability fails

Proving Stability

Recall that P is always nonexpansive:

$$\|\phi P - \psi P\| \leqslant \|\phi - \psi\| \quad \forall\, \phi, \psi \in \mathscr{P}(S)$$

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

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

(Strict contraction means $\rho(Tx, Ty) < \rho(x, y)$ whenever $x \neq y$)

For now let's stick to S is finite, so $\mathscr{P}(S)$ is compact

Then we just need the strict contraction property

To get the strict contraction property we use this result:

Theorem. If $p(x, y) > 0$ for all x, y, then the corresponding Markov operator P is a strict contraction on $(\mathscr{P}(S), \| \cdot \|)$

The proof uses two lemmas:

Fact. If $\phi, \psi \in \mathscr{P}(S)$ and $\phi \neq \psi$, then

$$\exists\, x, x' \in S \text{ such that } \phi(x) > \psi(x) \text{ and } \phi(x') < \psi(x')$$

Fact. If $g \in \ell_1(S)$ and $\exists\, x, x' \in S$ such that $g(x) > 0$ and $g(x') < 0$, then

$$\Big| \sum_{x \in S} g(x) \Big| < \sum_{x \in S} |g(x)|$$

Ex. Prove both

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

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

Theorem. If S is finite and P is both aperiodic and irreducible, then P is globally stable

Proof: It suffices to show that

$$\forall\, x, y \in S \times S, \quad \exists\, n \in \mathbb{N} \text{ s.t. } k \geqslant n \implies p^k(x, y) > 0$$

Indeed, if this statement holds, then, by choosing the largest such n over all (x, y) pairs, we obtain

$$\exists\, n \in \mathbb{N} \text{ s.t. } p^n(x, y) > 0 \quad \text{for all} \quad (x, y) \in S \times S$$

Since S is finite, this implies that P^n is globally stable

As discussed above, global stability of P^n implies global stability of P

So fix $x, y \in S \times S$ and let's try to show that

$$\exists\, n \in \mathbb{N} \text{ s.t. } k \geqslant n \implies p^k(x,y) > 0$$

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

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

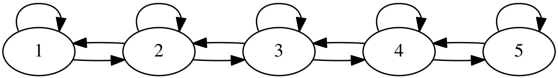
$$\ell \geqslant m \implies p^\ell(y,y) > 0$$

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

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

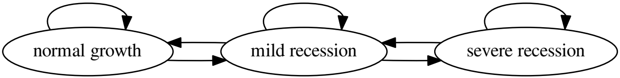
QED

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]])
```

The Law of Large Numbers

Let $h \in \mathbb{R}^S$ and let $\{X_t\}$ be a Markov chain on generated by stochastic kernel p

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

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

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

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

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