ECON-GA 1025 Macroeconomic Theory I

Lecture 5

John Stachurski

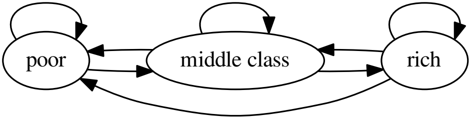
Fall Semester 2018

Today's Lecture

- Stochastic kernels and Markov chains on **finite** sets
- Distribution dynamics
- High performace computing tips

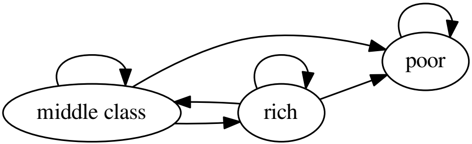
Prequel 1: Directed Graphs

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



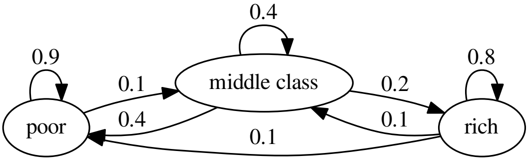
- y is called **accessible** from x if $y = x$ or $\exists$ a sequence of arcs leading from x to y
- The graph is called **strongly connected** if y is accessible from x for all $x, y \in \mathsf{X}$

Another example



- 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: Brouwer's Fixed Point Theorem

Theorem. (Brouwer–Hadamard, 1910) If

1. C is a convex compact subset of $\mathbb{R}^d$
2. T is a continuous self-map on C

then T has at least one fixed point in C

Prequel 3: The Space of Distributions

Let X be any finite set with elements $x_1, \ldots, x_n$

As usual, on $\mathbb{R}^{\mathsf{X}}$ we adopt the pointwise partial order

- $h \leqslant g$ if $h(x) \leqslant g(x)$ for all $x \in \mathsf{X}$

The set of **distributions** on X is denoted $\mathcal{P}(\mathsf{X})$ and defined as all $\varphi \in \mathbb{R}^{\mathsf{X}}$ such that

- $\varphi \geqslant 0$
- $\sum_{x \in \mathsf{X}} \varphi(x) = 1$

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

Metrizing $\mathcal{P}(\mathsf{X})$

As usual,

- $\|h\|_1 := \sum_{x \in \mathsf{X}} |h(x)|$
- $d_1(g, h) := \|g - h\|_1$

Thus,

$$\mathcal{P}(\mathsf{X}) = \{h \in \mathbb{R}^{\mathsf{X}} : h \geqslant 0 \text{ and } \|h\|_1 = 1\}$$

$\mathcal{P}(\mathsf{X})$ also called the **unit simplex** in $\mathbb{R}^n$

- A convex, compact subset of $\mathbb{R}^n$

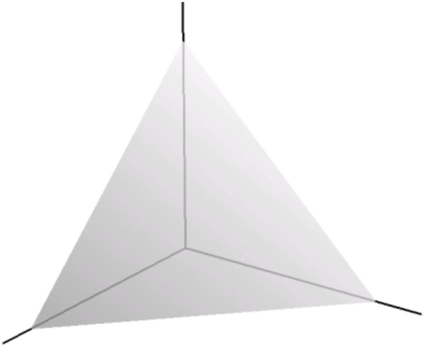
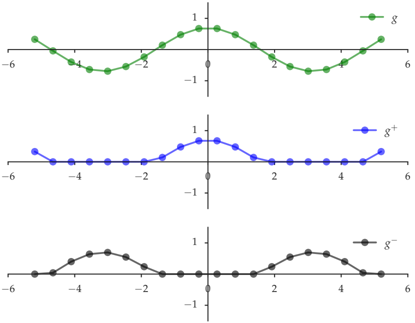


Figure: The unit simplex in $\mathbb{R}^3$

Notation: $g^+ := g \vee 0$ and $g^- := (-g) \vee 0$



Ex. Show that, for all $g \in \ell_1(\mathsf{X})$ we have

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

Stochastic Kernels

A **stochastic kernel** on X is a function $\Pi \colon \mathsf{X} \times \mathsf{X} \to \mathbb{R}$ such that

$$\Pi(x, \cdot) \in \mathcal{P}(\mathsf{X}) \quad \text{for all} \quad x \in \mathsf{X}$$

In other words,

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

Intuition:

1. We have one distribution $\Pi(x, \cdot)$ for each point $x \in \mathsf{X}$
2. $\Pi(x, y)$ is the probability of moving from x to y in one step

Matrix Representation

There are some alternative representations of stochastic kernels

When X is finite, we can represent Π by a matrix

$$\Pi = \begin{pmatrix} \Pi(x_1, x_1) & \cdots & \Pi(x_1, x_n) \\ \vdots & & \vdots \\ \Pi(x_n, x_1) & \cdots & \Pi(x_n, x_n) \end{pmatrix}$$

Note: this is a Markov matrix / stochastic matrix

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

Example. (Hamilton, 2005)

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

Markov matrix:

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

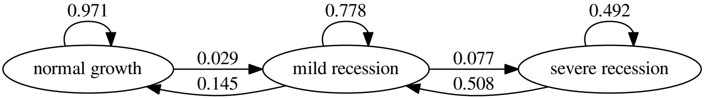
- state 1 = normal growth
- state 2 = mild recession
- state 3 = severe recession

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 X
- no edge means $\Pi(x, y) = 0$

Example. International growth dynamics study of Quah (1993)

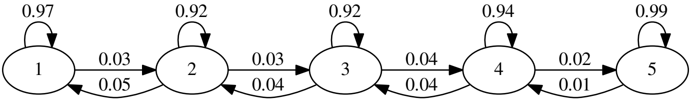
State = real GDP per capita relative to world average

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

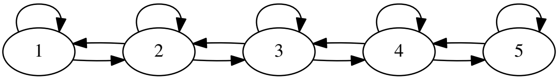
$$\Pi_Q = \begin{pmatrix} 0.97 & 0.03 & 0.00 & 0.00 & 0.00 \\ 0.05 & 0.92 & 0.03 & 0.00 & 0.00 \\ 0.00 & 0.04 & 0.92 & 0.04 & 0.00 \\ 0.00 & 0.00 & 0.04 & 0.94 & 0.02 \\ 0.00 & 0.00 & 0.00 & 0.01 & 0.99 \end{pmatrix}$$

The transitions are over a one year period

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



Dropping labels gives the directed graph



From Stochastic Kernels to Markov Chains

Let

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

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

$$\mathbb{P}\{X_{t+1} = y \mid X_0, X_1, \ldots, X_t\} = \Pi(X_t, y) \quad \text{for all} \quad t \geqslant 0, \; y \in X$$

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

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

Simulation

One technique for generating $\{X_t\}$ from a given kernel Π

For $x \in \mathsf{X}$ and $u \in (0,1)$, let

$$F(x,u) := \sum_{i=1}^{n} y_i \mathbb{1}\{q_{i-1}(x) < u \leqslant q_i(x)\}$$

where $\{y_1, \ldots, y_n\} = \mathsf{X}$ and

$$q_i(x) := \sum_{j=1}^{i} \Pi(x, y_j) \quad \text{with } q_0 = 0$$

Now X_0 is drawn from $\psi_0 \in \mathcal{P}(\mathsf{X})$ and then

$$X_{t+1} = F(X_t, U_{t+1}) \quad \text{where } \{U_t\} \overset{\text{IID}}{\sim} U(0,1) \tag{1}$$

Generates a Markov chain with stochastic kernel Π

The next exercise asks you to verify this

Ex. Conditional on $X_t = x$, show that, for each i in $1, \ldots, n$,

1. $X_{t+1} = y_i$ if and only if

$$q_{i-1}(x) < U_{t+1} \leqslant q_i(x)$$

2. This event has probability $\Pi(x, y_i)$

Conclude that X_{t+1} in (1) is a draw from $\Pi(x, \cdot)$

Linking Marginals

By the law of total probability we have

$$\mathbb{P}\{X_{t+1} = y\} = \sum_{x \in \mathsf{X}} \mathbb{P}\{X_{t+1} = y \mid X_t = x\} \cdot \mathbb{P}\{X_t = x\}$$

Letting ψ_t be the distribution of X_t, this becomes

$$\psi_{t+1}(y) = \sum_{x \in \mathsf{X}} \Pi(x, y)\psi_t(x) \qquad (y \in \mathsf{X})$$

Regarding distributions as **row** vectors, we can write this as

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

The map $\psi \mapsto \psi\Pi$ **updates the distribution** of the state

Dynamical System Representation

Think of $(\mathcal{P}(\mathsf{X}), \Pi)$ as a dynamical system

- Π is identified with the map $\psi \mapsto \psi\Pi$,

$$(\Pi\psi)(y) := \sum_{x \in \mathsf{X}} \Pi(x, y)\psi(x)$$

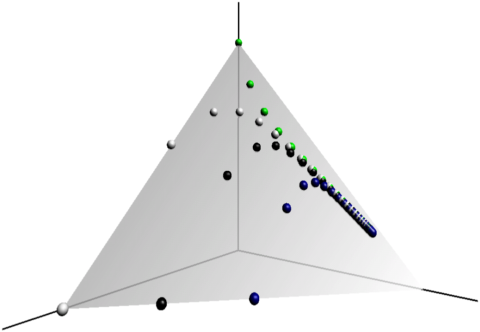
- Also called the **Markov operator** generated by kernel Π

Ex. Show that Π is a self-mapping on $\mathcal{P}(\mathsf{X})$

Interpretation of trajectories:

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

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



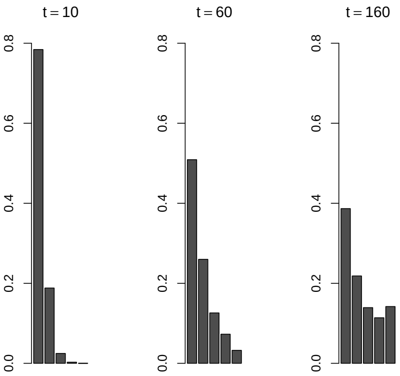


Figure: Distributions from Quah's stochastic kernel, $X_0 = 1$

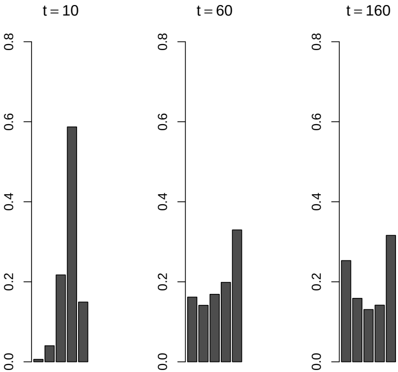


Figure: Distributions from Quah's stochastic kernel, $X_0 = 4$

Higher Order Kernels

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

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

- Called the k**-step stochastic kernel**
- We are just taking matrix powers (finite case)

Ex. Show that if Π is a stochastic matrix, then so is Π^k for all k

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

$$\Pi^k(x, y) = \mathbb{P}\{X_k = y \mid X_0 = x\} \qquad (x, y \in \mathsf{X})$$

To see why, recall that

$$\{X_t\} \text{ generated by } \Pi \text{ and } X_0 = x \implies X_k \sim \delta_x \Pi^k$$

Hence

$$\mathbb{P}\{X_k = y \mid X_0 = x\} = (\delta_x \Pi^k)(y)$$

But

$$(\delta_x \Pi^k)(y) = \Pi^k(x, y)$$

Chapman-Kolmogorov Equations

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

$$\Pi^{j+k}(x, y) = \sum_{z \in \mathsf{X}} \Pi^k(x, z) \Pi^j(z, y) \qquad ((x, y) \in \mathsf{X} \times \mathsf{X})$$

Proof: Let $X_0 = x$ and let $y \in \mathsf{X}$ be given

By the law of total probability, we have

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

Expectations

Given stochastic kernel Π and h in $\mathbb{R}^{\mathsf{X}}$, consider

$$(\Pi h)(x) = \sum_{y \in \mathsf{X}} h(y) \Pi(x, y) \qquad (x \in \mathsf{X})$$

($\Pi h =$ the product of matrix Π and column vector h)

Interpretation

$$(\Pi h)(x) = \mathbb{E}[h(X_{t+1}) \mid X_t = x]$$

More generally,

$$(\Pi^k h)(x) = \sum_{y} h(y) \Pi^k(x, y) = \mathbb{E}[h(X_{t+k}) \mid X_t = x]$$

Stationary Distributions

Let Π be a stochastic kernel on X

If $\psi^* \in \mathcal{P}(\mathsf{X})$ satisfies

$$\psi^*(y) = \sum_{x \in \mathsf{X}} \Pi(x, y) \psi^*(x) \quad \text{for all} \quad y \in \mathsf{X}$$

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

Equivalent to the above:

- $\psi^* = \psi^* \Pi$
- ψ^* is a steady state of $(\mathcal{P}(\mathsf{X}), \Pi)$

Interpretation:

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

Existence

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

Proof: Think of distribution $\varphi \in \mathcal{P}(\mathsf{X})$ as vector $(\varphi(x_i))_{i=1}^n$

- Π is a continuous map (just matrix multiplication)
- Π maps $\mathcal{P}(\mathsf{X})$ into itself
- $\mathcal{P}(\mathsf{X})$ is a closed, bounded subset of $\mathbb{R}^n$
- $\mathcal{P}(\mathsf{X})$ is also convex in $\mathbb{R}^n$

Existence of a fixed point follows from Brouwer

Computing the Stationary Distribution

Consider solving $\psi^* \Pi = \psi^*$ for ψ^*

Problem: there are trivial solutions, such as $\psi^* = 0$

To force our solution to be in $\mathcal{P}(\mathsf{X})$, let

- I be the $n \times n$ identity matrix
- $\mathbb{1}_n$ be the $1 \times n$ vector of ones, $\mathbb{1}_{n \times n}$ be the $n \times n$ matrix of ones

Ex. Show that $\psi \in \mathcal{P}(\mathsf{X})$ is stationary for Π if and only if

$$\mathbb{1}_n = \psi(I - \Pi + \mathbb{1}_{n \times n}) \tag{2}$$

Now transpose and solve for ψ' — requires that $I - \Pi + \mathbb{1}_{n \times n}$ is nonsingular