Lecture: Stochastic processes

Thomas Bourany

Macro 3 – Fall Semester – 2017

Content

- In this lecture, we will cover probability theory, stochastic processes and Markov chains.
- Some mathematical concepts may be abstract, but we will try to link as much as possible the definitions to economic concepts and applications.
- *These definitions are for your culture, they allow us to be consistent and rigorous in the use of models.*
- *This aims at improving your understanding of some mathematical tools that you may encounter in articles and seminars.*
- Do not hesitate to ask questions!

Probability theory – Introduction

- Let $(\Omega, \mathscr{F}, P)$ be a probability space, and $(S, \mathscr{P}(S))$ – or $(\mathbb{R}, \mathscr{B}_{\mathbb{R}})$ a measurable space.
  - The set of all possible outcomes, or *sample space* Ω is attached to a collection $\mathscr{F}$ of sets (parts of Ω). This collection includes all the potential events.
  - $\mathscr{F}$ is a σ-algebra, it is intuitively the set of all information available. If an event/outcome A is not in $\mathscr{F}$, this means it can not happen.
  - The rules defining a σ-algebra are the following: (i) $\Omega \in \mathscr{F}$, (ii) $A \in \mathscr{F} \Rightarrow A^c \in \mathscr{F}$, and (iii) $A_n \in \mathscr{F}, \forall n \Rightarrow \cup_{n\geq 1} A_n \in \mathscr{F}$
  - All these events have a probability $\mathbb{P}$, (i.e. you can "measure" how frequent the outcome will be).
  - The rules of σ-algebra imply that if you can measure $\mathbb{P}(A)$ or $\mathbb{P}(A_n)$, you can also measure $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ or $\mathbb{P}(\cup_n A_n)$ ($\leq \Sigma_n \mathbb{P}(A_n)$)

Probability theory – Introduction

- A random variable $X : \Omega \to S$ is a measurable function from the set of possible outcomes Ω to a set S.
  - "Measurable function", intuitively, means that each value of the function X (in S) corresponds to an event included in $\mathcal{F}$.
  - *Example*: A dice maps the hazard to a set $\{1, \ldots, 6\}$. A financial asset maps the hazard to the set $\mathbb{R}$ (positive or negative return).
  - In practice, we do not focus so much on Ω.
- We call **law** (or distribution) of a random variable X the measure P_X given by $P_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(\omega \in \Omega \; s.t. \; X(\omega) \in A)$
  - The measure P_X is the "image measure" of $\mathbb{P}$ via the application X
- From this law, if the random variable is real (maps into $\mathbb{R}$), we can compute the usual things:
  - Expected value: $\mathbb{E}(X) = \int_\Omega X(\omega)\mathbb{P}(d\omega) = \int_\mathbb{R} x \, P_X(dx) = \int_\mathbb{R} x f(x) \, dx$ (the last equality holding *only* if the r.v. X has a *p.d.f.*)
  - Moments: $\mathbb{E}(X^2), \mathbb{E}(X^3), \ldots,$ and $Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$

Conditional expectation

- Given $(\Omega, \mathscr{F}, P)$, and $\mathscr{G} \subset \mathscr{F}$ a sub-σ-algebra, we can define the **conditional expectation** of X *with respect to* $\mathscr{G}$, denoted $\mathbb{E}(X|\mathscr{G})$
- *Careful!!* Conditional expectation is a mathematical object that tends to be greatly misunderstood by economists (and students in mathematics).
- We define it as any variable Y checking the two following conditions:
  - (i) Y is $\mathscr{G}$-measurable
  - (ii) $\forall A \in \mathscr{G}$, $\mathbb{E}(X\,\mathbb{1}_A) = \mathbb{E}(Y\,\mathbb{1}_A)$
- Therefore $\mathbb{E}(X|\mathscr{G})$ is a random variable ! It is *not* a number !
  - The two defining properties can be intuitively translated as follow:
  - (i) For all the different values of Y (and thus $\mathbb{E}(X|\mathscr{G})$), there exists a corresponding event in $\mathscr{G}$ (if there is not such event, then Y is not $\mathscr{G}$-measurable)
  - (ii) Along each of the events A in the information set $\mathscr{G}$, the value of Y (and thus $\mathbb{E}(X|\mathscr{G})$) is the same as the averaged value of X.

Conditional expectation - Important properties

- If Z is $\mathscr{G}$-measurable, then $\mathbb{E}(Z|\mathscr{G}) = Z$
  - Z is a $\mathscr{G}$-measurable random variable, (that implies that Z cannot have more information than $\mathscr{G}$), therefore it implies that the information contained in Z is redundant with the information of $\mathscr{G}$ and Z averaged on all events of $\mathscr{G}$ equal Z.
- If Z is independent of $\mathscr{G}$, then $\mathbb{E}(Z|\mathscr{G}) = \mathbb{E}(Z)$
  - Independence is here the opposite of measurability: no info contained in $\mathscr{G}$ can inform us on the value of W, therefore averaging W on each event of $\mathscr{G}$ ends up as the same thing as averaging over the whole space, i.e. $\mathbb{E}(Z)$ (which is the only value of the random variable $\mathbb{E}(Z|\mathscr{G})$).
- (Law of iterated expectations), $\mathscr{H} \subset \mathscr{G}$ (both sub-σ-algebra) $\Rightarrow \mathbb{E}(\mathbb{E}(Z|\mathscr{G})|\mathscr{H}) = \mathbb{E}(Z|\mathscr{H})$
  - If the info in $\mathscr{H}$ is smaller than the info in $\mathscr{G}$ (which is smaller than the info in $\mathscr{F}$), then averaging w.r.t. $\mathscr{H}$ ends up taking only the smaller set of info available (and it doesn't change anything if you have a more (or less) refined variable inside the sign $\mathbb{E}(\cdot|\mathscr{H})$
  - As a result, $\mathbb{E}(\mathbb{E}(Z|\mathscr{G})) = \mathbb{E}(Z)$

Great theorems of convergence - Intro

- If $(X_n)_{n\leq 0}$ is a sequence of random variables, we need to analyze the convergence toward a limit. The question of the nature of convergence is at the heart of statistics (to attest the quality of estimators and C.I.).
- There exists 4 main modes of convergences:
  - Convergence "Almost-surely" ("the proba of converging is one")
  - Convergence in mean (or L^p) ("the difference fades out in norm L^p/moment of order p")
  - Convergence in probability ("the proba of diverging tends towards zero")
  - Convergence in distribution ("the law/c.d.f. tends towards another law/c.d.f.)

Great theorems of convergence - I

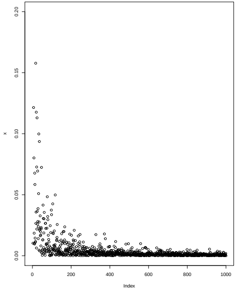
- A sequence of random variables $(X_n)_{n\geq 0}$ converges *"Almost-surely"* toward X if there exists an event A with proba one ($\mathbb{P}(A) = 1$) where, $\forall \omega \in A, \lim_{n\to\infty} X_n(\omega) = X(\omega)$

Said differently,

$$\mathbb{P}\left(\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right) = 1$$

After some fluctuations of the sequence, we are (almost-) sure that X_n won't fall too far from X

Example: convergence a.s. of $X_n \sim \mathscr{E}(\lambda = n)$
$X_n \to_{p.s.} X = 0$



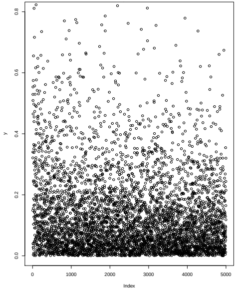
Great theorems of convergence - II

- A sequence of random variables $(X_n)_{n\geq 0}$ converges *"in probability"* toward X if, for all $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\Big(\omega \in \Omega : |X_n(\omega) - X(\omega)| > \varepsilon\Big) = 0$$

The proba that the sequence X_n falls far away from X is decreasing in n (but it can potentially be strictly positive)

Example: convergence in probability of $X_n \sim \mathscr{E}(\lambda = log(n))$ and $X_n \rightarrow_{\mathbb{P}} X = 0$ and $X_n \nrightarrow_{p.s.} 0$



Great theorems of convergence - III

- A sequence of random variables $(X_n)_{n\geq 0}$ converges *"in mean p"* or in $L^p(\Omega, \mathcal{F}, \mathbb{P})$ toward X if

$$\lim_{n\to\infty} \mathbb{E}\Big(|X_n - X|^p \Big) = 0$$

- A sequence of random variables $(X_n)_{n\geq 0}$ converges *"in law* or *in distribution* toward X if, for all continuous and bounded functions f

$$\lim_{n\to\infty} \mathbb{E}\big(f(X_n)\big) = \mathbb{E}\big(f(X)\big)$$

Careful! At the contrary of the three previous theorems (focusing on the convergence of random variables), here it is about the convergence of a sequence of ***laws***! (i.e. $P_{X_1}, P_{X_2} \cdots \to P_X$). This is much weaker!

Great theorems of convergence - IV

- ***Law of Large Numbers:*** If $\mathbb{E}(|X|) < \infty$, and if $\mathbb{E}(X) = \mu$, then:

$$\lim_{n\to\infty} \frac{X_1 + \cdots + X_n}{n} = \mu$$

This convergence is ***almost sure*** *(strong law of large numbers) and* ***in probability*** *(weak law of large numbers).*

- ***Central Limit Theorem:*** Let $(X_n)_{n\geq 1}$ a sequence of real random variables, ***i.i.d.***, with moments of second order $\mathbb{E}(X^2) < \infty$, and noting $S_n = \sum_{i=1}^{n} X_i$ and $\sigma^2 = \text{Var}(X)$, then:

$$\lim_{n\to\infty} \sqrt{n}\Big(\frac{S_n}{n} - \mu\Big) \cong \mathcal{N}(0, \sigma^2)$$

This convergence is ***in law****, and that intuitively implies that any sum of r.v. falls "normally" around its mean μ, with a variance σ^2 and at a speed of convergence $\sqrt{n}$.*

Two important theorems for optimization

- ***Jensen inequality***
  Let $X : (\Omega, \mathscr{F}, \mathbb{P}) \rightarrow \mathbb{R}$ be a real random variable and ϕ be a real convex function. Suppose X and $\varphi(X)$ are integrable (i.e. $\mathbb{E}(|X|) < \infty$ and $\mathbb{E}(|\phi(X)|) < \infty$). Then:

$$\phi(\mathbb{E}(X)) \leq \mathbb{E}(\phi(X))$$

This holds also for conditional expectations, for any $\mathscr{G} \subset \mathscr{F}$ sub-σ-algebra:

$$\phi(\mathbb{E}(X|\mathscr{G})) \leq \mathbb{E}(\phi(X|\mathscr{G}))$$

Two important theorems for optimization

- ***Interchanging differentiation and expectation:***
  On $(\Omega, \mathscr{F}, \mathbb{P})$, and I an interval in$\mathbb{R}$, let define $\varphi : I \times \Omega \to \mathbb{R}$ be a measurable function. If it satisfies:
  1. For every $x \in I$, the random variable $\varphi(x, \cdot)$ is integrable,
  2. $\frac{\partial \varphi(x,\omega)}{\partial x}$ exists at every $x \in I$
  3. There exists Y an integrable random variable such that, $\forall x \in I : \left| \frac{\partial \varphi(x,\omega)}{\partial x} \right| \leq Y(\omega)$

  Then, the function $\Phi(x) = \mathbb{E}(\varphi(x, \cdot))$ is well defined and differentiable at every $x \in I$, with:

$$\Phi'(x) = \mathbb{E}\left(\frac{\partial \varphi(x, \cdot)}{\partial x} \right)$$

Stochastic process – definition

- A stochastic process is a sequence of random variables X_t indexed (and ordered) by their time $t \in T$.
  - t is a index of time: it can be countable ($t \in \mathbb{N}$) and the time is discrete, or it can be uncountable ($t \in \mathbb{R}$) and the time is continuous. With the use of stochastic calculus, most models in finance are in continous-time.
- Given a probability space $(\Omega, \mathscr{F}, P)$, we define a **filtration** $(\mathscr{F}_t)_{t\geq 0}$ as a increasing sequence of sub-σ-algebras. $\mathscr{F}_0 \subset \mathscr{F}_{t_1} \subset \mathscr{F}_{t_2} \subset \cdots \subset \mathscr{F}$
  - In econ or finance, we often call $(\mathscr{F}_t)_{t\geq 0}$ the *information set*, as the knowledge of what can happen (i.e. the set of events that can be measured) grows over time.

Stochastic process – definition

- A sequence of random variables $(X_t)_{t\geq 0}$ is a **stochastic process**
  - We can pose $(\mathscr{F}_t)_{t\geq 0} \equiv \sigma(X_s : 0 \leq s \leq t)$, which is a filtration generated by the stochastic process (or canonical filtration).
- A stochastic process is **adapted** w.r.t. $(\mathscr{F}_t)_{t\geq 0}$, if $\forall t$, X_t is $\mathscr{F}_t$-measurable.
  - If X_t is $\mathscr{F}_{t-1}$-measurable, then the knowledge of X_t can be predicted by the information in $\mathscr{F}_{t-1}$
  - If X_t is not $\mathscr{F}_t$-measurable, it often means that X_t contains more (or different) information than $\mathscr{F}_t$
  - It implies that if $(X_t)_t$ is adapted, the knowledge of X_t does not give you *more* information than the information set $\mathscr{F}_t$ (in particular you can't predict the future).
  - A stochastic process is always adapted to its canonical filtration.
- A stochastic process is said to be **predicable**, if $\forall t \in \mathbb{N}$, X_t is $\mathscr{F}_{t-1}$-measurable.

Stochastic process – Examples

- A sequence of deterministic variables (constant accross Ω), such as $X_t = t$ is a stochastic process, but quite boring.
- A sequence of random variables $(X_t)_{t \geq 0}$ which are all following the same law (for example $X_t \sim \mathcal{N}(0, 1)$) is also a stochastic process, but not some much interesting neither.
- Researchers in probability are looking for processes that "behave well", whose law may vary over time or have constant properties over time *and* that are simple to study.
- Two "simple" processes are i) **martingales**, and ii) **Markov process**
  - The former are used in mathematical finance (to model the behavior of an asset price, a portfolio or a derivative) and the latter are used in macro models (with heterogenous agents, especially in the context of the Bellman algorithm) and in biology (multiplication of cells, growth of tumors, evolution of population of species)

Stochastic processes and conditional expectation

- In economics, the conditional expectation w.r.t. a σ-algebra from a filtration $(\mathscr{F}_t)_{t\geq 0}$ is a crucial tool. It is denoted compactly by the operator $\mathbb{E}_t$
$$\mathbb{E}_t(X) \equiv \mathbb{E}(X|\mathscr{F}_t)$$
Note: $\mathbb{E}_t(X)$ is ***not*** a number, but rather a function of the different shocks present in $\mathscr{F}_t$ (in economics: TFP shocks – aggregate or idiosyncratic – or policy shocks)
- Law of iterated expectations rewrites :
$$\mathscr{F}_t \subset \mathscr{F}_{t+1} \text{(both sub-}\sigma\text{-algebra)} \Rightarrow \mathbb{E}(\mathbb{E}(X|\mathscr{F}_{t+1})|\mathscr{F}_t) = \mathbb{E}(X|\mathscr{F}_t)$$
or in short: $\mathbb{E}_t(\mathbb{E}_{t+1}(X)) = \mathbb{E}_t(X)$
- For a stochastic processes evolving over time:
  - If X_t is adapted, then $\mathbb{E}_t(X_t) = X_t$
  - If ε_t is idiosyncratic, i.i.d., mean zero and not predictable, then $\mathbb{E}_t(\varepsilon_{t+1}) = \mathbb{E}(\varepsilon_{t+1}) = 0$.
  - If X_t is adapted, but not Y_t, $\mathbb{E}_t(X_t\, Y_t) = X_t\ \mathbb{E}_t(Y_t)$

Martingales

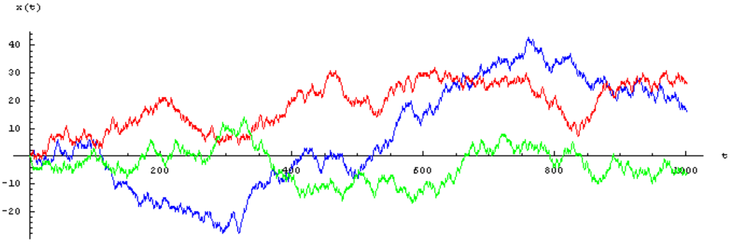
- In discrete-time, we define M_t as a ***martingale*** (resp. super-martingale, sub-martingale), w.r.t. a filtration $(\mathscr{F}_t)_{t\geq 0}$, a stochastic process verifying:
  1. $(M_t)_t$ is adapted
  2. $\forall t, \mathbb{E}\big(|M_t|\big) < \infty$
  3. $\forall t, \mathbb{E}(M_{t+1}|\mathscr{F}_t) = M_t$
     (resp. $\mathbb{E}(M_{t+1}|\mathscr{F}_t) \leq M_t$ and $\mathbb{E}(M_{t+1}|\mathscr{F}_t) \geq M_t$)

- *Intuitively, the mean of a martingale M_t is constant over time, while decreasing for a supermartingale and increasing for a submartingale.*

Martingales examples: Random walks

$$X_{t+1} = X_t + \varepsilon_t$$
$$= X_0 + \sum_{i=0}^{t} \varepsilon_i \qquad \forall t \geq 0$$

- where ε_t are i.i.d. random variable s.t. $\varepsilon_t \sim \mathscr{P}$ (any distribution, with probability mass function (if countable set) or p.d.f. (if uncoutable set) ψ.
- A random walk is called *simple* or *isotropic* if $\psi(1) = \psi(-1) = 1/2$



Martingales examples: Random walks

$$\begin{aligned} X_{t+1} &= X_t + \varepsilon_t \\ &= X_0 + \sum_{i=0}^{t} \varepsilon_i \qquad \forall t \geq 0 \end{aligned}$$

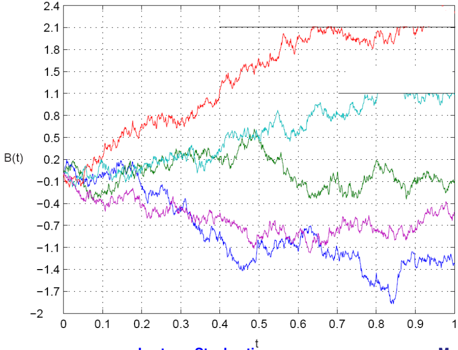
- where ε_t are i.i.d. random variable s.t. $\varepsilon_t \sim \mathscr{P}$ (any distribution, with probability mass function (if countable set) or p.d.f. (if uncoutable set) ψ.
- A random walk is called *simple* or *isotropic* if $\psi(1) = \psi(-1) = 1/2$
- If the mean of $\mathscr{P}$ is null (i.e. $\mathbb{E}(\varepsilon_t) = 0$), then the random walk is a ***martingale***. (resp. if $\mathbb{E}(\varepsilon_t) \geq 0$, then X_t is a sub-martingale, or $\mathbb{E}(\varepsilon_t) \leq 0$, then X_t is a super-martingale)

Martingales examples: Brownian motion

- This is the *"continuous-time"* stochastic process which is the closest of a random-walk.
- We define as a ***Brownian motion*** the continuous process W_t valued in $\mathbb{R}$ such that:
  1. The function $t \mapsto W_t(\omega)$ is continuous on $\mathbb{R}_+$
  2. For all $0 \leq s < t$, the increment $W_t - W_s$ is independent of $\sigma(W_u, u \leq s)$
  3. For all $t \geq s \geq 0$, $W_t - W_s$ follows the normal distribution $\mathcal{N}(0, \sigma^2)$
- The brownian motion is "standard" if $W_0 = 0$ and $\sigma = 1$.
- Here, the Brownian motion is a martingale
- It is used to model any "small" shock in a continuous-time finance/macro models.
- *By a theorem (Donsker theorem), it is possible to show that a "normal-shock"-random-walk converges in law toward a brownian motion, when time increment tends to zero.*

Martingales examples: Brownian motion

- We define as a ***Brownian motion*** the continuous process W_t valued in $\mathbb{R}$ such that:
  1. The function $t \mapsto W_t(\omega)$ is continuous on $\mathbb{R}_+$
  2. For all $0 \leq s < t$, the increment $W_t - W_s$ is independent of $\sigma(W_u, u \leq s)$
  3. For all $t \geq s \geq 0$, $W_t - W_s$ follows the normal distribution $\mathcal{N}(0, \sigma^2)$



Finite Markov Chains – Introduction

We will now consider **Markov chains** – a simple example of stochastic process – which are *finite* – i.e. that happen on a finite number of "states".

- S, the state space is a finite space, with n elements $\{x_1, \ldots, x_n\}$
  - S can be a real value (consumption level, growth rate) or anything else (high or low "states of the world": $h, l \in S$).
- The **Transition function**, or transition matrix we can denote $Q \equiv p(x, y)$, is a function $Q : S \times S \to [0, 1]$ such that

  (i) Each element of $Q(\cdot, \cdot)$ is non-negative

  (ii) $\sum_{y \in E} p(x, y) = 1, \forall x \in E$

  This means the rows of the matrix some to one.
- It is easy to see that if Q is a transition matrix, then it k-th power $\tilde{Q} = Q^k$ is also a transition matrix.

- A **Markov chain** X_t is a sequence of *S-valued* random variables, with transition matrix Q, if, for all $t \geq 0$, and for all $y \in S$ we have:

$$\mathbb{P}(X_{t+1} = y \,|\, X_0, X_1, \ldots, X_t) = p(X_n, y)$$

- Therefore, it satisfies the *Markov property*:

$$\mathbb{P}(X_{t+1} = y \,|\, X_0, X_1, \ldots, X_t) = \mathbb{P}(X_{t+1} = y \,|\, X_t)$$

  - In other words, to forecast the distribution of X_{t+1} on S, the only information need is the current state X_t.

Examples

- The simplest example: A worker can be either (i) unemployed or (ii) employed
  - When unemployed, he finds a job at rate α
  - When employed, he loses its job with probability β
- Therefore, the transition matrix writes:

$$Q = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

- Question (exercises?)
  - What is the average duration of unemployment?
  - Over the long-run, what fraction of time does a worker find herself unemployed?
  - Conditional on employment, what is the probability of becoming unemployed at least once over the next 12 months?

- Another example: Hamilton (2005) used the US employment data, and determined the frequency of:
  (i) Normal growth
  (ii) Mild recession
  (iii) Severe recession.
- The stochastic matrix is estimated such as:

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

  - It says that, when US are in a severe recession, there is a 50.8 probability to face a mild recession next month, and no chance at all to come back to normal growth.

- Another common example: **Random walk:**

$$X_{t+1} = X_t + \varepsilon_t$$
$$= X_0 + \sum_{i=0}^{t} \varepsilon_i \qquad \forall t \geq 0$$

  - where ε_t are i.i.d. random walk s.t. $\varepsilon_t \sim \mathscr{P}$ (any distribution) with probability mass function ψ – recall that S is finite and thus countable.

- What would be the transition matrix?
- It can be shown (exercise?) that this matrix is such that:

$$p(x_t, x_{t+1}) = \psi(x_{t+1} - x_t) \qquad \forall\, x_t,\, x_{t+1} \in S$$

Recursive formulation of stochastic processes

- Recall that $\mathbb{P}(X_{t+1} = x_{t+1} \,|\, X_t = x_t) = p(x_t, x_{t+1})$
- Therefore, knowing the initial state, one can iterate over the transition matrix:

$$\mathbb{P}(X_t = x_t \,|\, X_0 = x_0) = [Q^t](x_0, x_t)$$

  - In other words, the initial condition and the transition matrix are the only determinant of the path of X_t

Marginal distribution:

- Knowing the distribution at time $X_t \sim \mathscr{P}$ (with p.m.f. ψ) and the transition matrix $Q \equiv p(x_t, x_{t+1})$, what can we say about the probability of X_{t+1}?
- The solution lies in the *law of total probabilities*:

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

- Rewriting, we get:

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

- If you express the p.m.f. ψ as a n-values *rows vector* (of probas), the n equations become matrices as:

$$\psi_{t+1} = \psi_t \, Q$$

Multi-step transition probabilities:

- Similarly, if $\psi_{t+1} = \psi_t \, Q$, therefore, we can generalize it:

$$\psi_t = \psi_0 \, Q^t$$
$$\psi_{t+m} = \psi_t \, Q^m$$

- Exercise: Using the transition matrix on recessions seen before, and considering the today's state as unknown, (you only know the distribution-vector ψ_t), what is the probability to be in a mild or severe recession in 6 months?
- Answer:

$$\mathbb{P}(recession) = \psi_{t+6} \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \psi_t \cdot Q^6 \cdot \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

Irreducibility:

- Two states x_a and x_b are said to communicate with each other if there exist positive integers m and n such that:

$$Q^m(x_a, x_b) > 0 \qquad \text{and} \qquad Q^n(x_b, x_a) > 0$$

- The stochastic/transition matrix is said **irreducible** if all states communicate, i.e. x_a and x_b in $S \times S$ can communicate.
- Question: is the ”recession matrix” irreducible?

Aperiodicity:

- The period of a state x_o is the greatest common divisor of the set of integers defined by:

$$D(x_o) \equiv \{j \geq 1 : Q^j(x_o, x_o) > 0\}$$

- Example: if $D(x) = \{3, 6, 9, \dots\}$, the period is 3
- A stochastic matrix is said **aperiodic** if the period of every state is 1, or **periodic** otherwise.

Stationary distribution:

- Some distributions are invariant under the transition matrix. We call these distribution **stationary**, if the distribution $\psi^\star$ on S is such that:

$$\psi^\star = \psi^\star Q$$

- Obviously, an immediate consequence is : $\psi^\star = \psi^\star Q^t \quad \forall t$
- Therefore if the random variable X_0 has a stationary distribution, then X_t also have this same distribution.

Stationary distribution:

- **Theorem**:
  Every stochastic matrix Q has at least one stationary distribution.
  - $\forall Q, \exists \psi^\star$, s.t. $\psi^\star = \psi^\star Q$
  - Here, the assumption that S is a finite set is a key one.
- The proof of this theorem lie in the *Brouwer fixed point theorem*
- *Note:* If Q is the identity matrix, then all distributions are stationary
- Is this stationary distribution unique?

Stationary distribution:

- **Theorem**: If the stochastic matrix Q is **irreducible** and **aperiodic**, then :
  1. Q has exactly one stationary distribution $\psi^\star$
  2. For any initial distribution ψ_0, we have $\|\psi_0 Q^t - \psi^\star\| \to 0$ when $t \to \infty$

- A stochastic matrix satisfying the conditions of the theorem is sometimes called **uniformly ergodic**
  - Note that part 1 of the theorem requires only irreducibility, whereas part 2 requires both irreducibility and aperiodicity
  - One easy sufficient condition for aperiodicity and irreducibility is that every element of Q is strictly positive (Exercise?)

Ergodicity:

- **Theorem**: Under irreducibility, another important result:

$$\frac{1}{n}\sum_{t=1}^{n} \mathbb{1}\{X_t = x\} \rightarrow_{n\rightarrow\infty} \psi^{\star}(x)$$

  - The convergence is *Almost sure*
  - The result does not depend on the initial distribution of X_0.

- The result tells us that the fraction of time the chain spends at state x converges to $\psi^{\star}(x)$ as time goes to infinity
- This convergence theorem is a special case of a **Law of large numbers** result for Markov chains