

ECON2125/4021/8013

Lecture 15

John Stachurski

Semester 1, 2015

Announcements

- This week's Thursday lecture will be shifted to Friday
 - 9am on 23/04/2015 to 10am on 24/04/2015
 - Same location
 - To let people focus on exam preparation
- Preliminary date for final exam is June 11
 - Still subject to change

Convergence in Distribution

Let

- $\{F_n\}_{n=1}^{\infty}$ be a sequence of cdfs
- F be any cdf

We say that $\{F_n\}_{n=1}^{\infty}$ **converges weakly** to F if

$$F_n(x) \rightarrow F(x) \quad \text{as } n \rightarrow \infty$$

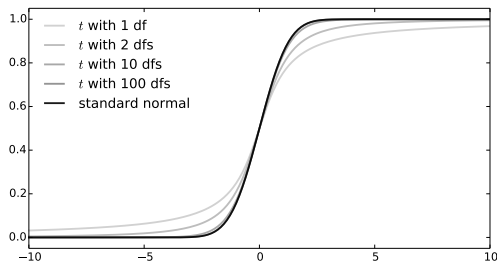
for any x such that F is continuous at x

- In essence, F_n gets close to F when n is large

Example. Student's t -density with n degrees of freedom is

$$p_n(x) := \frac{\Gamma(\frac{n+1}{2})}{(n\pi)^{1/2}\Gamma(\frac{n}{2})} \left(1 + n^{-1}x^2\right)^{-(n+1)/2}$$

It's well known that the corresponding cdfs F_n converge weakly to the standard normal cdf



We say that $\{X_n\}_{n=1}^{\infty}$ converges to X **in distribution** if

1. $X_n \sim F_n$,
2. $X \sim F$ and
3. $F_n \rightarrow F$ weakly

In this case we write $X_n \xrightarrow{d} X$

- In short, the distribution of X_n converges to that of X

Fact. If $X_n \xrightarrow{p} X$, then $X_n \xrightarrow{d} X$

Example. If X is any RV and $X_n := X + \frac{1}{n}$ then $X_n \xrightarrow{d} X$

Proof: Let F and F_n be the cdfs of X and X_n respectively

Observe that, $\forall x \in \mathbb{R}$,

$$F_n(x) = \mathbb{P} \left\{ X + \frac{1}{n} \leq x \right\} = \mathbb{P} \left\{ X \leq x - \frac{1}{n} \right\} = F \left(x - \frac{1}{n} \right)$$

Suppose that F is continuous at x

Since $x - \frac{1}{n} \rightarrow x$, we have

$$F \left(x - \frac{1}{n} \right) \rightarrow F(x)$$

(By the def of continuity — more on this later)

Fact. If $g: \mathbb{R} \rightarrow \mathbb{R}$ is continuous, then

$$1. X_n \xrightarrow{d} X \implies g(X_n) \xrightarrow{d} g(X)$$

$$2. X_n \xrightarrow{p} X \implies g(X_n) \xrightarrow{p} g(X)$$

Remark: This fact is called the **continuous mapping theorem**

Example. If α is constant and $X_n \xrightarrow{d} X$, then

- $X_n + \alpha \xrightarrow{d} X + \alpha$
- $\alpha X_n \xrightarrow{d} \alpha X$
- etc.

The Central Limit Theorem

Let $\{X_i\}_{i=1}^{\infty} \stackrel{\text{iid}}{\sim} F$ with

- $\mu := \mathbb{E}[X_i] = \int xF(dx)$
- $\sigma^2 := \text{var}[X_i] = \int (x - \mu)^2 F(dx)$, assumed finite

Fact. In this setting we have

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2) \quad \text{as } n \rightarrow \infty$$

- $\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$
- \xrightarrow{d} means the cdf of LHS \rightarrow weakly to the $N(0, \sigma^2)$ cdf

Proof: Omitted

Alternative version: Under the same conditions we have

$$\sqrt{n} \left\{ \frac{\bar{X}_n - \mu}{\sigma} \right\} \xrightarrow{d} N(0, 1)$$

To see this let $Y \sim N(0, \sigma^2)$, so that $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} Y$

Applying the continuous mapping theorem gives

$$\sqrt{n} \left\{ \frac{\bar{X}_n - \mu}{\sigma} \right\} \xrightarrow{d} \frac{Y}{\sigma}$$

Clearly Y/σ is normal, with

$$\mathbb{E} \left[\frac{Y}{\sigma} \right] = \frac{1}{\sigma} \mathbb{E}[Y] = 0 \quad \text{and} \quad \text{var} \left[\frac{Y}{\sigma} \right] = \frac{1}{\sigma^2} \text{var}[Y] = 1$$

Discussion: The CLT tells us about distribution of \bar{X}_n when

- sample is IID
- n large

Informally,

$$\sqrt{n}(\bar{X}_n - \mu) \approx Y \sim N(0, \sigma^2)$$

$$\therefore \bar{X}_n \approx \frac{Y}{\sqrt{n}} + \mu \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Thus, \bar{X}_n approximately normal, with

- mean equal to μ , and
- variance $\rightarrow 0$ at rate proportional to $1/n$

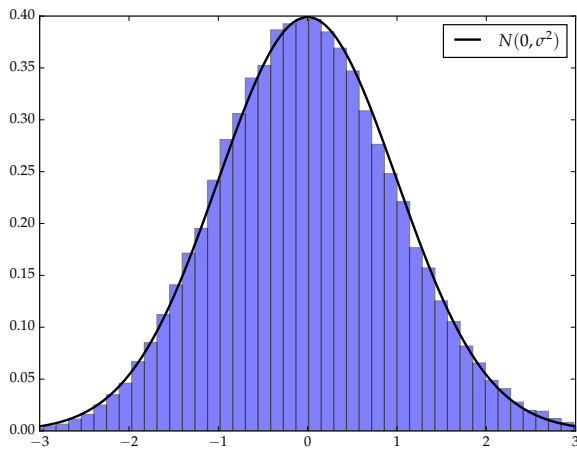
Illustrating the CLT

We can illustrate the CLT with simulations by

1. choosing an arbitrary cdf F for X_n and a large value for n
2. generating independent draws of $Y_n := \sqrt{n}(\bar{X}_n - \mu)$
3. using these draws to compute some measure of their distribution, such as a histogram
4. comparing the latter with $N(0, \sigma^2)$

We do this for

- $F(x) = 1 - e^{-\lambda x}$ (exponential distribution)
- $n = 250$



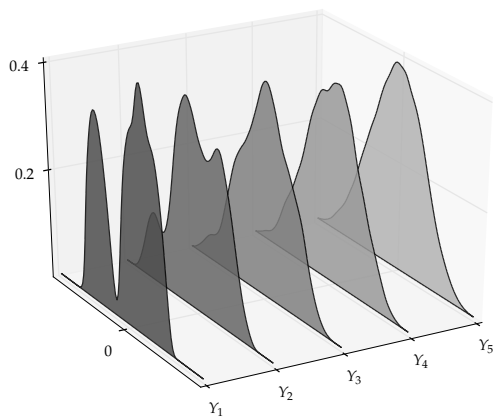
Another way we can illustrate the CLT:

Numerically compute the distributions of

1. $Y_1 = \sqrt{1}(\bar{X}_1 - \mu) = X_1 - \mu$
2. $Y_2 = \sqrt{2}(\bar{X}_2 - \mu) = \sqrt{2}(X_1/2 + X_2/2 - \mu)$
3. $Y_3 = \dots$

The distribution of each Y_n can be calculated once the distribution F of X_n is specified

The next figure shows these distributions for arbitrarily chosen F



Conditional Expectation

Let X and Y be two random variables

To economize on notation we overload the p symbol by writing

- $p(x, y)$ for the joint density
- $p(y \mid x)$ for the conditional density of y given x , etc.

Example. If on a computer we draw

1. $X \sim U[0, 1]$
2. and then $Y \sim N(\mu, \sigma^2)$ with μ set to X

then

$$p(y \mid x) = p(y \mid X = x) = N(x, \sigma^2)$$

The **conditional expectation** of Y given X is then defined as

$$\mathbb{E}[Y | X] = \int y p(y | X) dy$$

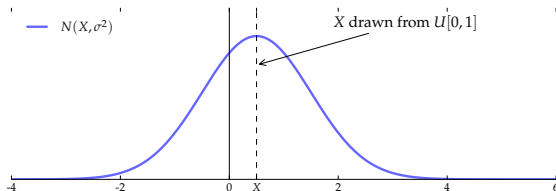
- Notation: Here and below, by convention, $\int := \int_{-\infty}^{\infty}$

The right hand side contains X , so it is a random variable!

In general,

- $\mathbb{E}[Y | X]$ is the “best predictor of Y given X ”
- A rule that maps X into a prediction of Y
- And therefore a function of X
- And therefore random

Example. As before we draw $X \sim U[0, 1]$ and then $Y \sim N(X, \sigma^2)$



We want a rule that maps X to a prediction of Y

Intuition suggests that the best guess of Y given X is just X

Let's make sure this checks out

$$\mathbb{E}[Y | X] = \int y p(y | X) dy$$

For this case we saw that

$$p(y | x) = N(x, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ \frac{-(y - x)^2}{2\sigma^2} \right\}$$

$$\therefore p(y | X) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ \frac{-(y - X)^2}{2\sigma^2} \right\}$$

$$\therefore \mathbb{E}[Y | X] = \int y \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ \frac{-(y - X)^2}{2\sigma^2} \right\} dy$$

This is just the mean of $N(X, \sigma^2)$, which is X

Also intuitive: when X and Y are independent, X is no help in predicting Y

- the same as predicting Y with no information

Since $\mathbb{E}[Y] = \text{best guess of } Y \text{ with no information}$, this suggests

$$\mathbb{E}[Y | X] = \mathbb{E}[Y]$$

The conjecture checks out too, since for this case we have

$$p(y | X) = \frac{p(y, X)}{p(X)} = \frac{p(y)p(X)}{p(X)} = p(y)$$

Hence

$$\mathbb{E}[Y | X] = \int yp(y | X)dy = \int yp(y)dy = \mathbb{E}[Y]$$

Sometimes we want to compute the conditional expectation of a function $f(X, Y)$ depending on both X and Y

Example. Suppose that

- Y is the payoff from a foreign asset, random
- $r(X)$ is an exchange rate, depending on some random X
- return in domestic currency is $f(X, Y) = r(X)Y$

What is the expectation of $f(X, Y)$ given X ?

The general definition is

$$\mathbb{E}[f(X, Y) \mid X] = \int f(X, y)p(y \mid X)dy$$

For the preceding example this gives

$$\mathbb{E} [r(X)Y \mid X] = \int r(X) y p(y \mid X) dy$$

Since $r(X)$ doesn't depend on y it can pass out of the integral

Hence

$$\mathbb{E} [r(X)Y \mid X] = r(X) \int y p(y \mid X) dy$$

That is,

$$\mathbb{E} [r(X)Y \mid X] = r(X) \mathbb{E} [Y \mid X]$$

This is a general rule — when conditioning on X , RVs depending only on X can be passed out of the expectation

The Multivariate Case

We can condition on X_1, \dots, X_K using

$$\begin{aligned} p(y \mid \mathbf{x}) &= p(y \mid x_1, x_2, \dots, x_K) \\ &= p(y \mid X_1 = x_1, X_2 = x_2, \dots, X_K = x_K) \end{aligned}$$

Then we set

$$\begin{aligned} \mathbb{E}[Y \mid \mathbf{X}] &:= \int y p(y \mid \mathbf{X}) dy \\ &= \int y p(y \mid X_1, X_2, \dots, X_K) dy \end{aligned}$$

- \mathbf{X} can be a matrix: we condition on all X_{ij} in \mathbf{X}

We can also extend the definition the case where \mathbf{X} and \mathbf{Y} are matrices

Given

$$\mathbf{Y} = \begin{pmatrix} Y_{11} & \cdots & Y_{1K} \\ \vdots & \vdots & \vdots \\ Y_{N1} & \cdots & Y_{NK} \end{pmatrix}$$

we set

$$\mathbb{E}[\mathbf{Y} \mid \mathbf{X}] = \begin{pmatrix} \mathbb{E}[Y_{11} \mid \mathbf{X}] & \cdots & \mathbb{E}[Y_{1K} \mid \mathbf{X}] \\ \vdots & \vdots & \vdots \\ \mathbb{E}[Y_{N1} \mid \mathbf{X}] & \cdots & \mathbb{E}[Y_{NK} \mid \mathbf{X}] \end{pmatrix}$$

We have provided some intuition for the following key facts

Fact. If \mathbf{X} , \mathbf{Y} and \mathbf{Z} are random matrices and \mathbf{A} and \mathbf{B} are constant matrices, then, assuming conformability,

1. $\mathbb{E} [\mathbf{AX} + \mathbf{BY} | \mathbf{Z}] = \mathbf{A}\mathbb{E} [\mathbf{X} | \mathbf{Z}] + \mathbf{B}\mathbb{E} [\mathbf{Y} | \mathbf{Z}]$
2. If \mathbf{X} and \mathbf{Y} are independent, then $\mathbb{E} [\mathbf{Y} | \mathbf{X}] = \mathbb{E} [\mathbf{Y}]$
3. If $G(\mathbf{X})$ is a matrix depending only on \mathbf{X} , then
 - $\mathbb{E} [G(\mathbf{X}) \mathbf{Y} | \mathbf{X}] = G(\mathbf{X})\mathbb{E} [\mathbf{Y} | \mathbf{X}]$
 - $\mathbb{E} [\mathbf{Y} G(\mathbf{X}) | \mathbf{X}] = \mathbb{E} [\mathbf{Y} | \mathbf{X}] G(\mathbf{X})$
4. $\mathbb{E} [\mathbf{Y} | \mathbf{Z}]' = \mathbb{E} [\mathbf{Y}' | \mathbf{Z}]$
5. $\mathbb{E} [\mathbb{E} [\mathbf{Y} | \mathbf{X}]] = \mathbb{E} [\mathbf{Y}]$

Let's just check that $\mathbb{E} [\mathbb{E} [Y | X]] = \mathbb{E} [Y]$ in the scalar case

We have

$$\begin{aligned}\mathbb{E} [\mathbb{E} [Y | X]] &= \mathbb{E} \left[\int y p(y | X) dy \right] \\ &= \int \left[\int y p(y | x) dy \right] p(x) dx \\ &= \int y \left[\int p(y | x) p(x) dx \right] dy \\ &= \int y p(y) dy = \mathbb{E} [Y]\end{aligned}$$

New Topic

ANALYSIS

Motivation

We looked at linear systems carefully, but how about nonlinear systems?

- Solving nonlinear equations
- Optimization problems

How are these problems different?

What mathematics do we need to study them?

An example problem:

Let f be a given nonlinear function

Does there exist an \bar{x} such that $f(\bar{x}) = 0$?

Examples.

- F is a profit function, $f = F'$, we're looking for stationary points of the profit function
- We want to solve an equation $g(\bar{x}) = y$ for \bar{x}
 - Set $f(x) = g(x) - y$

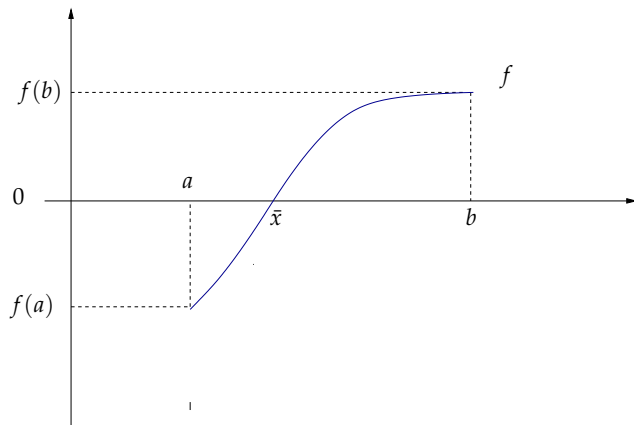


Figure : Existence of a root

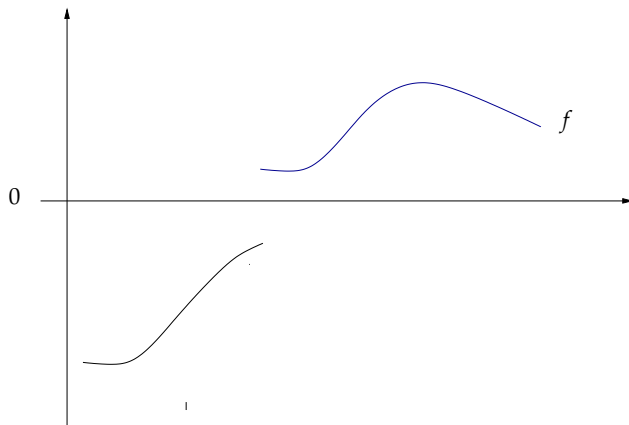


Figure : Non-existence of a root

One answer: a solution exists under certain conditions including continuity

Questions:

- So how can I tell if f is continuous?
- Can we weaken the continuity assumption?
- Does this work in multiple dimensions?
- When is the root unique?
- How can we compute it?
- Etc.

These are typical problems in analysis