ECON2125/4021/8013

Lecture 14

John Stachurski

Semester 1, 2015

Announcements

No tutorials on Friday

Consultation times over the break =

- Qingyin Ma: 3:00-5:00 Fridays (as usual)
- Guanlong Ren: 4:00-6:00 Thursdays (changed)
- John S: 9:00-11:00 Mondays (as usual)

More solved exercises coming

- One more set, on probability
- Will appear on GitHub site by Monday evening

Comments on Exam Questions

Mainly small proofs/arguments requiring only a few steps of logic

In general, good answers will

- include relevant definitions
- use relevant facts from slides
- avoid long and difficult calculations — there's probably an easier way

Use of external theorems is discouraged

- You won't need them
- Don't tell me it's true because you saw it in a book

Sample question, worth five marks:

Q: Let $\mathbf{A}$ be any matrix. Show that the symmetric matrix $\mathbf{A}'\mathbf{A}$ is nonnegative definite

What is a good answer to this question?

A1 I love Kung Fu

- Mark: $0/5$
- Why: Irrelevant

A2 $N \times N$ symmetric matrix $\mathbf{B}$ is nonnegative definite if $\mathbf{x}'\mathbf{B}\mathbf{x} \geq 0$ for any $N \times 1$ vector $\mathbf{x}$. I don't know the rest.

- Mark: 2/5
- Why: Gave the relevant definition

A3 $N \times N$ symmetric matrix $\mathbf{B}$ is nonnegative definite if $\mathbf{x}'\mathbf{B}\mathbf{x} \geq 0$ for any $N \times 1$ vector $\mathbf{x}$. Strictly concave functions have unique minima. A set is a collection of objects. Sharks continue to swim while sleeping.

- Mark: 1/5
- Why: One relevant definition cancelled out by other noise

A4. By definition, an $N \times N$ symmetric matrix $\mathbf{B}$ is nonnegative definite if

$$\mathbf{x}'\mathbf{B}\mathbf{x} \geq 0 \text{ for any } N \times 1 \text{ vector } \mathbf{x} \qquad (\star)$$

Let $\mathbf{B} := \mathbf{A}'\mathbf{A}$ and fix any such $\mathbf{x}$. By the rules of transposes we have

$$\mathbf{x}'\mathbf{B}\mathbf{x} = \mathbf{x}'\mathbf{A}'\mathbf{A}\mathbf{x} = (\mathbf{A}\mathbf{x})'(\mathbf{A}\mathbf{x}) \geq 0$$

Here last equality holds because, for any vector $\mathbf{y}$,

$$\mathbf{y}'\mathbf{y} = \sum_{n=1}^{N} y_n^2 \geq 0$$

This confirms $(\star)$

- Mark: 5/5
- Why: Correct and crystal clear

Further Comments

Assessable topics for midterm exam = lecture slides 1–14

Who will mark and with what expectations?

I will

- write all of mid-term and final exams
- write solutions as guidelines, discuss with tutors

Tutors will

- do most of the actual marking, based on my solutions and guidelines

Background: Convergence and Continuity

Loosely speaking, a sequence $\{x_n\}$ converges to $x \in \mathbb{R}$ if

$$x_n \text{ gets "arbitrarily close" to } x \text{ as } n \to \infty$$

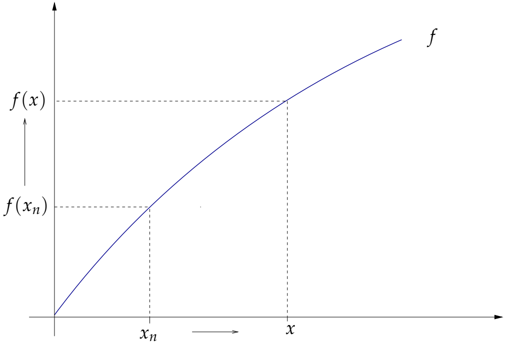
Example. If $x_n = 2 + 1/n$ then $x_n \to 2$ as $n \to \infty$

Comments

- We'll give a more careful definition in a later lecture
- "Close" means that $|x_n - x|$ is small

A function $f\colon \mathbb{R} \to \mathbb{R}$ is continuous at x if, for any $\{x_n\}$ with $x_n \to x$, we have

$$f(x_n) \to f(x)$$



Independence

Random variables $X_1, \ldots, X_N$ are called **independent** if, for all $(x_1, \ldots, x_N) \in \mathbb{R}^N$,

$$\mathbb{P}\{X_1 \leq x_1, \ldots, X_N \leq x_N\} = \prod_{n=1}^{N} \mathbb{P}\{X_n \leq x_n\}$$

Equivalently, if $X_1, \ldots, X_N$ are RVs with

- marginal distributions $\Phi_1, \ldots \Phi_N$
- joint distribution F

then independent iff

$$F(x_1, \ldots, x_N) = \prod_{n=1}^{N} \Phi_n(x_n)$$

An infinite sequence $\{X_n\}$ is called independent if any finite subset is independent

If all marginals of the X_n's are the same, they are called **identically distributed**

$$\Phi_1 = \cdots = \Phi_N = \Phi$$

"Independent and identically distributed" usually abbreviated to IID

If $\{X_n\}$ is IID with common cdf Φ we write $\{X_n\} \overset{\text{IID}}{\sim} \Phi$

The joint distribution of $X_1, \ldots, X_N$ is then

$$F(x_1, \ldots, x_N) = \prod_{n=1}^{N} \Phi(x_n)$$

Fact. For $X_1, \dots, X_N$ with marginal densities $\phi_1, \dots, \phi_N$ and joint density p,

$$X_1, \dots, X_N \text{ independent} \iff p(x_1, \dots, x_N) = \prod_{n=1}^{N} \phi_n(x_n)$$

Example. If X_1, X_2 are RVs with $p(x_1, x_2) = \phi_1(x_1)\phi_2(x_2)$, then

$$\begin{aligned} F(x_1, x_2) &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} p(s,t)\, dt\, ds \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \phi_1(s)\phi_2(t)\, dt\, ds \\ &= \int_{-\infty}^{x_1} \phi_1(s) ds \int_{-\infty}^{x_2} \phi_2(t) dt = \Phi_1(x_1)\Phi_2(x_2) \end{aligned}$$

Fact. If $X_1, \dots, X_N$ are independent, then

$$\mathbb{E}\left[\prod_{n=1}^{N} X_n\right] = \prod_{n=1}^{N} \mathbb{E}\left[X_n\right]$$

It follows that if X and Y are independent, then $\mathrm{cov}[X, Y] = 0$

Proof: If X and Y are RVs with $\mathbb{E}[X] = \mu_X$, $\mathbb{E}[Y] = \mu_Y$ then

$$\begin{aligned}
\mathrm{cov}[X, Y] &= \mathbb{E}\left[(X - \mu_X)(Y - \mu_Y)\right] \\
&= \mathbb{E}\left[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y\right] \\
&= \mathbb{E}\left[XY\right] - \mu_X\mu_Y \\
&= \mathbb{E}\left[X\right]\mathbb{E}\left[Y\right] - \mu_X\mu_Y = 0
\end{aligned}$$

The converse is not generally true

However,

$$\text{multivariate normal \& zero covariance} \implies \text{independence}$$

Indeed, suppose

- $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$
- $\text{cov}[X_i, X_j] = 0$ unless $i = j$

Since $\boldsymbol{\Sigma}$ is the variance covariance matrix, this means $\boldsymbol{\Sigma}$ must be diagonal

$$\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)$$

Using our facts about diagonal matrices we have

$$p(\mathbf{x}) = (2\pi)^{-N/2} \det(\boldsymbol{\Sigma})^{-1/2} \exp\left\{ -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \right\}$$

$$= \frac{1}{(2\pi)^{N/2}\prod_{n=1}^{N}\sigma_n} \exp\left\{ -\frac{1}{2}\sum_{n=1}^{N}(x_n-\mu_n)^2\sigma_n^{-2} \right\}$$

$$= \prod_{n=1}^{N} \frac{1}{(2\pi)^{1/2}\sigma_n} \exp\left\{ \frac{-(x_n-\mu_n)^2}{2\sigma_n^2} \right\}$$

$$= \prod_{n=1}^{N} \phi_n(x_n) \quad \text{where } \phi_n = \text{density of } N(\mu_n, \sigma_n^2)$$

Hence independent

Convergence in Probability

Let

- $\{X_n\}$ be a sequence of RVs
- X another RV or a constant

The sequence $\{X_n\}$ converges to X **in probability** if

$$\forall\,\delta > 0, \qquad \mathbb{P}\{|X_n - X| > \delta\} \to 0 \quad \text{as} \quad n \to \infty$$

We write $X_n \xrightarrow{p} X$

Example. If $X_n \sim N(\alpha, 1/n)$, then $X_n \xrightarrow{p} \alpha$

We can see this visually

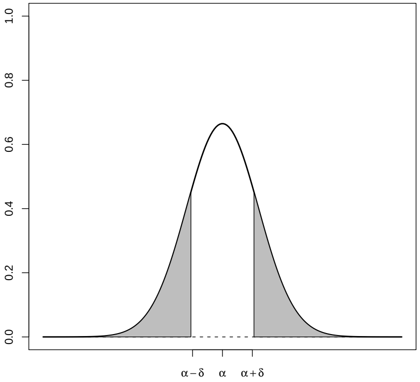


Figure : $\mathbb{P}\{|X_n - \alpha| > \delta\} \to 0$

We can see this visually

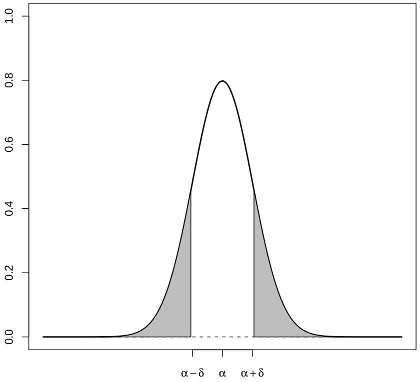


Figure : $\mathbb{P}\{|X_n - \alpha| > \delta\} \to 0$

We can see this visually

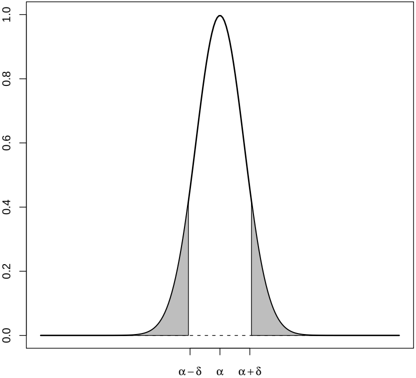


Figure : $\mathbb{P}\{|X_n - \alpha| > \delta\} \to 0$

Let's also check it formally, using this well known fact

Fact. If Y is any nonnegative random variable and $\theta > 0$, then

$$\mathbb{P}\{Y \geq \theta\} \leq \frac{\mathbb{E}[Y]}{\theta}$$

Proof: We have

$$Y \geq Y \mathbb{1}\{Y \geq \theta\} \geq \theta \mathbb{1}\{Y \geq \theta\}$$

Hence, by monotonicity and linearity of $\mathbb{E}$,

$$\mathbb{E}[Y] \geq \mathbb{E}[\theta \mathbb{1}\{Y \geq \theta\}] = \theta \mathbb{E}[\mathbb{1}\{Y \geq \theta\}] = \theta \mathbb{P}\{Y \geq \theta\}$$

Now let $X_n \sim N(\alpha, 1/n)$ as before, fix $\delta > 0$

Observe that

$$\{|X_n - \alpha| > \delta\} = \{(X_n - \alpha)^2 > \delta^2\}$$

As a result, we have

$$\begin{aligned}
\mathbb{P}\{|X_n - \alpha| > \delta\} &= \mathbb{P}\{(X_n - \alpha)^2 > \delta^2\} \\
&\leq \frac{\mathbb{E}\left[(X_n - \alpha)^2\right]}{\delta^2} \\
&= \frac{1}{n\delta^2} \\
&\to 0
\end{aligned}$$

Sample Averages

It's often said that diversified portfolios are "less risky"

For example, let

- X_n be the payoff from holding asset n
- $\mathbb{E}[X_n] = \mu$
- $\mathrm{var}[X_n] = \sigma^2$
- $\mathrm{cov}[X_j, X_k] = 0$ when $j \neq k$

If we hold just X_1 expected payoff is μ and variance is σ^2

If hold $Y = X_1/2 + X_2/2$ then mean is still μ but variance is

$$\mathrm{var}[Y] = \mathrm{var}\left[X_1/2 + X_2/2\right] = \frac{\sigma^2}{4} + \frac{\sigma^2}{4} = \frac{\sigma^2}{2}$$

More generally, if $\bar{X}_N := \frac{1}{N}\sum_{n=1}^{N} X_n$ then

$$\mathbb{E}\left[\bar{X}_N\right] = \mathbb{E}\left[\frac{1}{N}\sum_{n=1}^{N} X_n\right] = \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}\left[X_n\right] = \mu$$

but

$$\mathbb{E}\left[(\bar{X}_N - \mu)^2\right] = \mathbb{E}\left\{\left[\frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)\right]^2\right\}$$

$$= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\,(X_i - \mu)(X_j - \mu)$$

$$= \frac{1}{N^2}\sum_{i=1}^{N}\mathbb{E}\,(X_i - \mu)^2 = \frac{\sigma^2}{N}$$

Hence for this portfolio

1. the mean stays the same
2. variance of the portfolio goes to zero with N

Note the key step

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\,(X_i-\mu)(X_j-\mu)=\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}\,(X_i-\mu)^2$$

depends crucially on lack of correlation

If correlation is present the same argument doesn't work

Example. The subprime crisis

Law of Large Numbers

The next fact is called the (weak) law of large numbers (LLN)

Fact. Let $\{X_n\} \overset{\text{IID}}{\sim} F$. If

$$\int_{-\infty}^{\infty} |x| F(dx) < \infty$$

then

$$\frac{1}{N} \sum_{n=1}^{N} X_n \xrightarrow{p} \mu \quad \text{as} \quad N \to \infty$$

where

$$\mu := \mathbb{E}[X_n] = \int_{-\infty}^{\infty} x F(dx)$$

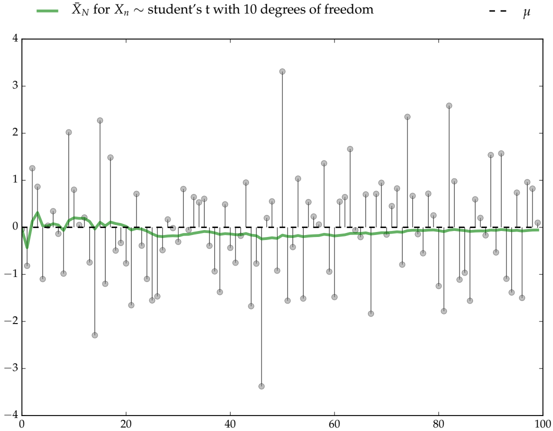


Figure : Example. Student t distributed RVs

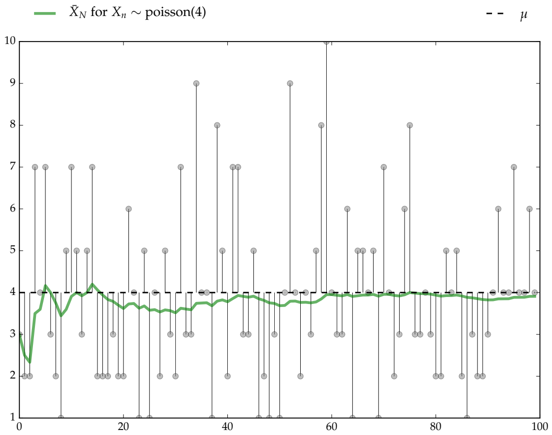


Figure : Example. Poisson distributed RVs

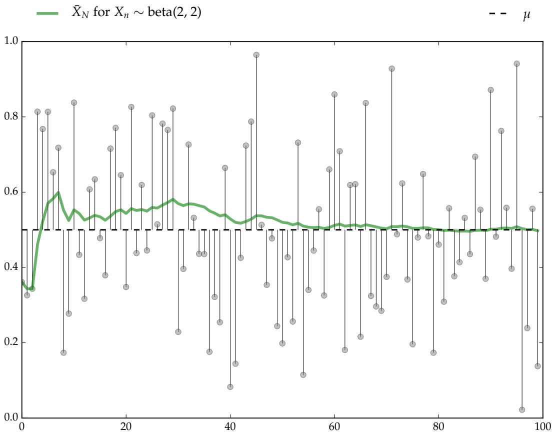


Figure : Example. Beta distributed RVs

Proof of the LLN for the case $\mathrm{var}[X_n] = \sigma^2 < \infty$

We saw before that

$$\mathbb{E}\left[(\bar{X}_N - \mu)^2\right] = \frac{\sigma^2}{N}$$

and

$$\mathbb{P}\{(\bar{X}_N - \mu)^2 > \delta^2\} \leq \frac{\mathbb{E}\left[(\bar{X}_N - \mu)^2\right]}{\delta^2}$$

Therefore

$$\mathbb{P}\{|\bar{X}_N - \mu| > \delta\} = \mathbb{P}\{(\bar{X}_N - \mu)^2 > \delta^2\} \leq \frac{\sigma^2}{N\delta^2}$$

$$\therefore \quad \mathbb{P}\{|\bar{X}_N - \mu| > \delta\} \to 0 \quad \text{as} \quad N \to \infty$$

The LLN is more general than it looks

Fact. If $\{X_n\} \overset{\text{IID}}{\sim} F$ and

$$h \colon \mathbb{R} \to \mathbb{R} \quad \text{with} \quad \int |h(x)| F(dx) < \infty$$

then

$$\frac{1}{N} \sum_{n=1}^{N} h(X_n) \overset{p}{\to} \int h(x) F(dx)$$

Proof: Apply LLN to $Y_n := h(X_n)$

Example. Set $h(x) = x^2$ to get

$$\frac{1}{N} \sum_{n=1}^{N} X_n^2 \overset{p}{\to} \int x^2 F(dx) \quad \text{as} \quad N \to \infty$$

Example. I have a model that tells me the distribution of household wealth in 5 years will be equal to the distribution of

$$Y = \log(\cos(X+1)^2 + \exp(X)^{1/2} + 5)$$

where

$$X \sim N(0,1)$$

Since Y is a well defined RV it has a cdf

$$G(y) := \mathbb{P}\{Y \leq y\}$$

I want to know $\mathbb{E}[Y] = \int yG(dy)$ but how to calculate it?

Easiest way is simulation

```
In [1]: import numpy as np

In [2]: X = np.random.randn(1e6)  # 10^6 N(0,1) draws

In [3]: temp = np.cos(X+1)**2 + np.sqrt(np.exp(X))

In [4]: Y = np.log(temp + 5)

In [5]: np.mean(Y)
Out[5]: 1.8837663629867571
```

This is a sample mean of 10^6 draws Y_n

The sample mean is close to the true mean by the LLN

Use same idea to get variance, standard deviation, median, etc.

LLN applies to probabilities as well

Example. Given any $x \in \mathbb{R}$,

$$\frac{1}{N}\sum_{n=1}^{N} \mathbb{1}\{X_n \leq x\} \xrightarrow{p} F(x)$$

Proof: Let $h(s) := \mathbb{1}\{s \leq x\}$

We then have

$$\mathbb{E}\left[h(X_n)\right] = \mathbb{E}\left[\mathbb{1}\{X_n \leq x\}\right] = \mathbb{P}\{X_n \leq x\} = F(x)$$

Hence, by the previous fact,

$$\frac{1}{N}\sum_{n=1}^{N} \mathbb{1}\{X_n \leq x\} = \frac{1}{N}\sum_{n=1}^{N} h(X_n) \xrightarrow{p} \mathbb{E}\left[h(X_n)\right] = F(x)$$

Failure of the LLN

We discussed how the LLN can fail when there's correlation

In fact the LLN can still work if correlations die out sufficiently quickly

- e.g., $\mathrm{cov}[X_j, X_{j+k}] \to 0$ quickly as $k \to \infty$

The other important assumption is

$$\int_{-\infty}^{\infty} |x| F(dx) < \infty$$

Conversely, with very heavy tailed distributions the LLN can fail

- Individual extreme observations dominate the average

Example. Recall the Cauchy distribution

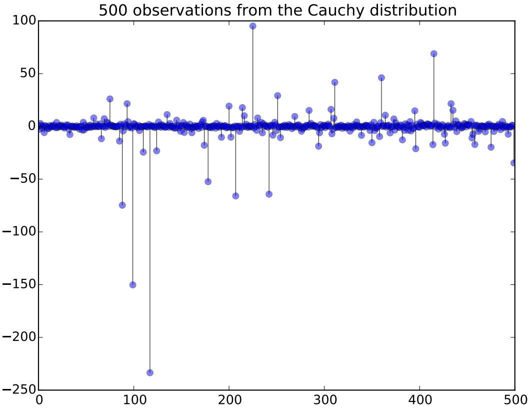
$$F(x) = \arctan(x)/\pi + 1/2 \quad \text{and} \quad p(x) = \frac{1}{\pi(1+x^2)}$$

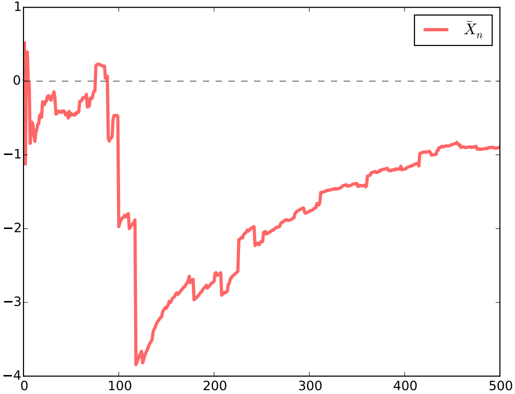
In this case it's known that

$$\int_{-\infty}^{\infty} |x| F(dx) := \int_{-\infty}^{\infty} |x| p(x) dx = \infty$$

In fact for Cauchy samples the LLN always fails

(Proof omitted)


500 observations from the Cauchy distribution
100
50
0
−50
−100
−150
−200
−250
0
100
200
300
400
500



$\bar{X}_n$
1
0
−1
−2
−3
−4
0
100
200
300
400
500

Convergence in Distribution

Let

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

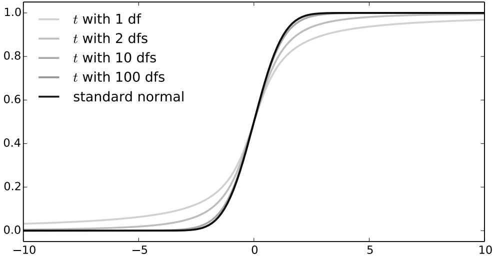
We say that $\{F_n\}_{n=1}^{\infty}$ **converges weakly** to F if, for any x such that F is continuous at x,

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

Example. **Student's t-distribution** with n degrees of freedom is the distribution on $\mathbb{R}$ with density

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

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

$$X_n \sim F_n, \quad X \sim F \quad \text{and} \quad F_n \to F \text{ weakly}$$

Write as $X_n \xrightarrow{d} X$

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

$$F_n(x) = \mathbb{P}\{X + 1/n \le x\} = \mathbb{P}\{X \le x - 1/n\} = F(x - 1/n)$$

If F is continuous at x, then, since $x - 1/n \to x$,

$$F(x - 1/n) \to F(x)$$

Fact. If $g\colon \mathbb{R} \to \mathbb{R}$ is continuous and $X_n \xrightarrow{d} X$, then

$$g(X_n) \xrightarrow{d} g(X)$$

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

Fact. If α is constant, $X_n \xrightarrow{p} \alpha$ and $Y_n \xrightarrow{d} Y$, then

$$X_n + Y_n \xrightarrow{d} \alpha + Y \quad \text{and} \quad X_n Y_n \xrightarrow{d} \alpha Y$$

Remark: This fact is sometimes called **Slutsky's theorem**

The Central Limit Theorem

Let $\{X_n\}_{n=1}^{\infty} \overset{\text{IID}}{\sim} F$ with

- $\mu := \mathbb{E}[X_n] = \int x F(dx)$
- $\sigma^2 := \text{var}[X_n] = \int (x - \mu)^2 F(dx) < \infty$

Fact. In this setting we have

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

- Note we have assumed variance $< \infty$
- Here $\xrightarrow{d}$ means the cdf of LHS $\to$ weakly to the cdf of RHS

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

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

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

- sample is IID
- N large

Informally,

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

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

Thus, $\bar{X}_N$ approximately normal, with

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

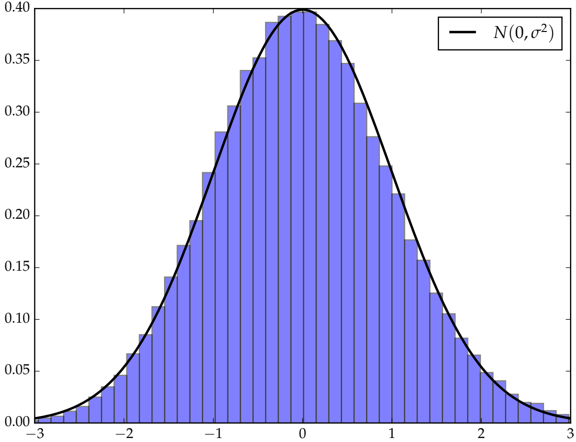
Illustrating the CLT

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)
- $N = 250$


0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−3
−2
−1
0
1
2
3
$N(0,\sigma^2)$

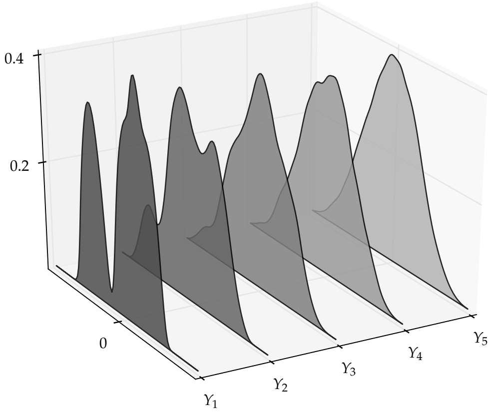
Another way we can illustrate the CLT:

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. $Y_3 = \cdots$

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


0.4
0.2
0
Y_1
Y_2
Y_3
Y_4
Y_5