

Math 4261: Mathematical Statistics I

Lecture 6: LLN and CLT

Ref: Blitzstein–Hwang Ch. 10; HMC 5.1–5.2

Many ideas and figures credited to Yihong Wu

Fall 2026

Outline

Recap

Modes of convergence

Law of large numbers

Central limit theorem

Outline

Recap

Modes of convergence

Law of large numbers

Central limit theorem

Recap: Sampling distributions

For random data $X_1, \dots, X_n$, a **statistic** $T(X_1, \dots, X_n)$ is a function of the data. Its distribution is the **sampling distribution**.

For instance, the sample mean

$$\overline{X}_n = \frac{X_1 + \dots + X_n}{n}.$$

Often, we can compute certain properties of the distribution: for instance for i.i.d. observations with mean μ and variance σ^2

$$\mathbb{E}[\overline{X}_n] = \mu, \quad \text{Var}(\overline{X}_n) = \frac{\sigma^2}{n}.$$

More generally, we studied tools for calculating the **exact** (MGFs), and for approximating **tail probabilities** related to $T(X_1, \dots, X_n)$.

We continue this study today with **large sample approximations**.

Recap: Exact distributions using MGFs

The **moment generating function** is $M_X(t) = \mathbb{E}[e^{tX}]$. If it is finite on an interval around 0, it determines the distribution.

For independent $X_1, \dots, X_n$, and constants a, b ,

$$M_{X_1 + \dots + X_n}(t) = \prod_{i=1}^n M_{X_i}(t), \quad M_{aX+b}(t) = e^{bt} M_X(at).$$

Examples

$$X \sim \text{Bern}(p) : M_X(t) = 1 - p + pe^t,$$

$$X \sim \text{Bin}(n, p) : M_X(t) = (1 - p + pe^t)^n,$$

$$Z \sim \mathcal{N}(0, 1) : M_Z(t) = e^{t^2/2},$$

$$Y \sim \chi_n^2 : M_Y(t) = (1 - 2t)^{-n/2}, \quad t < 1/2.$$

Recap: Tail bounds by MGFs (Gaussian case)

Let $Z \sim \mathcal{N}(0, 1)$. We showed

$$\mathbb{P}[|Z| \geq t] \leq 2e^{-t^2/2}.$$

Example

For $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, with $\sigma > 0$, and every $\epsilon > 0$,

$$\mathbb{P}[|\bar{X}_n - \mu| \geq \epsilon] \leq 2 \exp\left(-\frac{n\epsilon^2}{2\sigma^2}\right).$$

In particular, $n \geq \frac{2\sigma^2}{\epsilon^2} \log \frac{2}{0.05}$ guarantees accuracy within ϵ with probability at least 95%.

Outline

Recap

Modes of convergence

Law of large numbers

Central limit theorem

Why study sequences of statistics?

As the sample size changes, the statistic and its sampling distribution change:

$$T_1 = T_1(X_1), \quad T_2 = T_2(X_1, X_2), \quad \dots \quad T_n = T_n(X_1, \dots, X_n).$$

For an estimator $\hat{\theta}_n$, there are two basic statistical questions:

1. Does $\hat{\theta}_n$ get close to the target θ as n grows?
2. Does the distribution of an appropriately standardized version of $\hat{\theta}_n$ approach a useful limiting shape?

These questions prompt us to first define **convergence in probability** and **convergence in distribution**.

Convergence in probability

Definition

Let $X_1, X_2, \dots$ and X be random variables defined on the same sample space. We say that X_n **converges in probability** to X if, for every $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}[|X_n - X| \geq \epsilon] = 0.$$

We write $X_n \xrightarrow{p} X$.

Equivalently,

$$\lim_{n \rightarrow \infty} \mathbb{P}[|X_n - X| < \epsilon] = 1.$$

Thus the probability mass of the difference $X_n - X$ becomes concentrated near 0.

Example: Gaussian tail bound

(Revisited). Suppose $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$. From our previous calculation, for every $\epsilon > 0$,

$$\mathbb{P} \left[|\overline{X}_n - \mu| \geq \epsilon \right] \leq 2 \exp \left(-\frac{n\epsilon^2}{2\sigma^2} \right) \longrightarrow 0.$$

Therefore

$$\boxed{\overline{X}_n \xrightarrow{p} \mu.}$$

The **Gaussian assumption** gave us more than convergence: it gave a finite sample bound on the error probability.

The LLN will give the same qualitative statement under much more general assumptions.

Consistency

Let $\hat{\theta}_n$ be an estimate of a parameter θ based on a sample of size n . $\hat{\theta}_n$ is a **consistent estimator** of θ if for any $\epsilon > 0$, as $n \rightarrow \infty$,

$$\mathbb{P}[|\hat{\theta}_n - \theta| > \epsilon] \longrightarrow 0$$

Equivalently, we write

$$\hat{\theta}_n \xrightarrow{p} \theta.$$

Intuitively, we say that with high probability, the estimator is eventually arbitrarily close to its target.

Convergence in distribution

Definition

Let $X_1, X_2, \dots$ and X be random variables, with CDFs F_{X_n} and F_X . We say that X_n **converges in distribution** to X if

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

at every point x where F_X is continuous. We write $X_n \xrightarrow{d} X$.

The distribution of X is called the **limiting** or **asymptotic distribution** of the sequence X_n .

It lets us approximate probabilities such as

$$\mathbb{P}[X_n \leq x] \approx F_X(x)$$

when n is large and x is a continuity point.

Example: convergence in prob. vs dist.

Let $Z \sim \mathcal{N}(0, 1)$, and define

$$X_n = \begin{cases} Z, & n \text{ odd,} \\ -Z, & n \text{ even.} \end{cases}$$

Does $X_n \xrightarrow{d} Z$? Does $X_n \xrightarrow{p} Z$?

By symmetry of std. Gaussian $-Z \stackrel{d}{=} Z$.

so $X_n \stackrel{d}{=} Z \quad \forall n \Rightarrow X_n \xrightarrow{d} Z$.

However, $\forall \varepsilon > 0$, $P[|X_n - Z| \geq \varepsilon] = \begin{cases} 0 & n \text{ odd} \\ \underbrace{P[2|Z| \geq \varepsilon]}_{2[1 - \Phi(\frac{\varepsilon}{2})]} & n \text{ even} \end{cases}$

Hence $X_n \not\xrightarrow{p} Z$ fixed number

Relating the two notions of convergence

Theorem

If X_n converges to X in probability, then X_n converges to X in distribution:

$$X_n \xrightarrow{p} X \implies X_n \xrightarrow{d} X.$$

The converse is not true in general, as the preceding example shows.

Important exception: if the limit is a **constant c** , then

$$X_n \xrightarrow{d} c \iff X_n \xrightarrow{p} c.$$

Continuous mapping theorem

Often we might know the convergence of some statistic T_n , but we want to understand the convergence of $g(T_n)$ for some function g . A useful tool is

Continuous mapping theorem

Let $g(x)$ be a continuous function of x . As $n \rightarrow \infty$,

(a) If $T_n \xrightarrow{d} Z$, then $g(T_n) \xrightarrow{d} g(Z)$.

(b) If $T_n \xrightarrow{p} Z$, then $g(T_n) \xrightarrow{p} g(Z)$.

Example: Fair coin or not?

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Bern}(\frac{1}{2})$. We have $\mathbb{E}[X_1] = \frac{1}{2}$ and $\text{Var}[X_1] = \frac{1}{4}$. The **CLT** gives us

$$\sqrt{4n} \cdot \left(\bar{X}_n - \frac{1}{2} \right) \xrightarrow{d} \mathcal{N}(0, 1).$$

What is the asymptotic distribution of the statistic

$$T = 4n \left(\bar{X}_n - \frac{1}{2} \right)^2 ?$$

Motivation: if we compute T from n observed coin tosses and find that it is too large compared to typical values from this distribution, this provides evidence that the coin may be unfair. (Will see more in hypothesis testing).

Example: Fair coin or not?

By the CLT,

$$\sqrt{4n} \cdot \left(\overline{X}_n - \frac{1}{2} \right) \xrightarrow{d} \mathcal{N}(0, 1).$$

By the Continuous Mapping Theorem, applied with $g(x) = x^2$,

$$T = 4n \left(\overline{X}_n - \frac{1}{2} \right)^2 \xrightarrow{d} \chi_1^2,$$

where χ_1^2 is the distribution of Z^2 when $Z \sim \mathcal{N}(0, 1)$.

Two statistical goals

Suppose we use $\hat{\theta}_n$ to estimate a fixed parameter θ . **Consistency** was defined as

$$\hat{\theta}_n \xrightarrow{p} \theta.$$

A stronger, more informative statement describes the limiting distribution of the remaining sampling error.

Asymptotic normality

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, V).$$

This parallels the relationship between LLN and CLT. Indeed, these will be very helpful in establishing such statistical statements.

Outline

Recap

Modes of convergence

Law of large numbers

Central limit theorem

Motivation

Let $X_1, \dots, X_n$ be n independent and identically distributed (i.i.d.) r.v.'s with common mean μ and variance σ^2 . Let

$$\text{Sample mean} = \overline{X}_n = \frac{X_1 + \dots + X_n}{n}$$

Intuition: for fair coin flips

$$X_i = \begin{cases} 1 & i^{\text{th}} \text{ flip heads} \\ 0 & i^{\text{th}} \text{ flip tails} \end{cases}$$

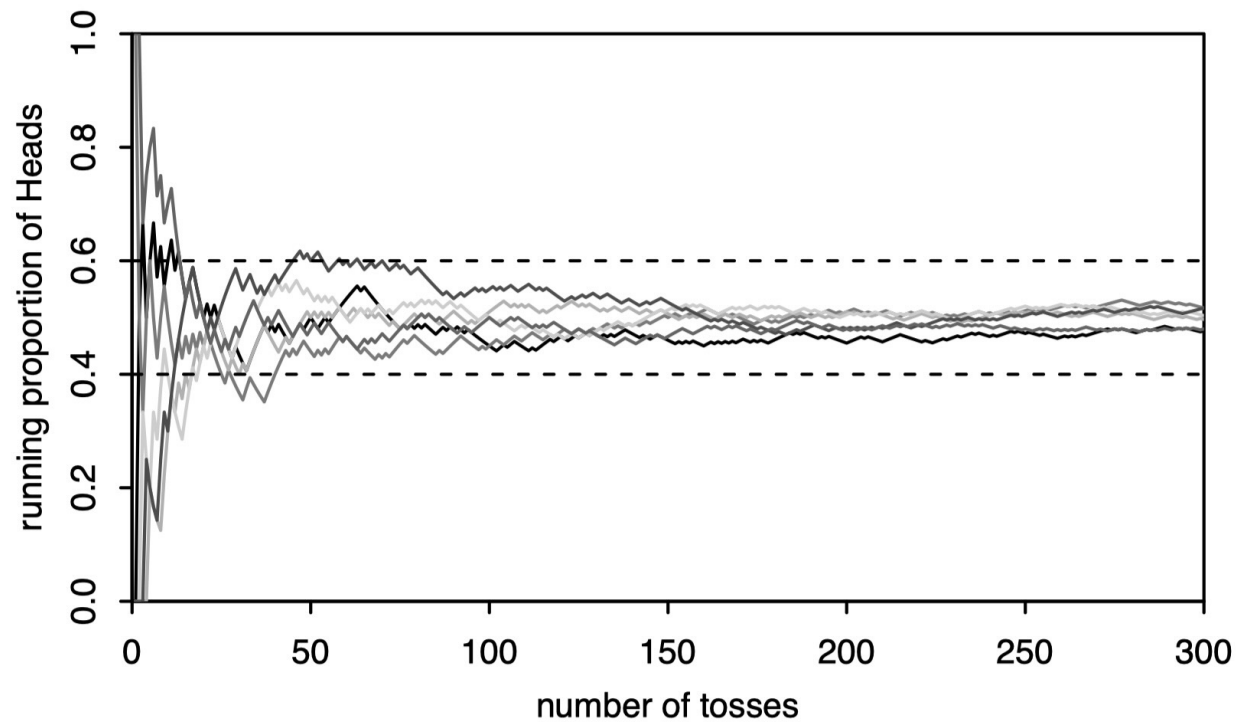
$\overline{X}_n$ = fraction of heads in n flips

Common mean $\mu = \mathbb{E}X_i = \mathbb{P}[i^{\text{th}} \text{ flip heads}] = \frac{1}{2}$. We expect

$$\overline{X}_n \approx \frac{1}{2}$$

if n is large enough.

Motivation



History

LLN first proved by [James Bernoulli](#) in 1700's.

The term 'Loi des grands nombres' first appeared in:

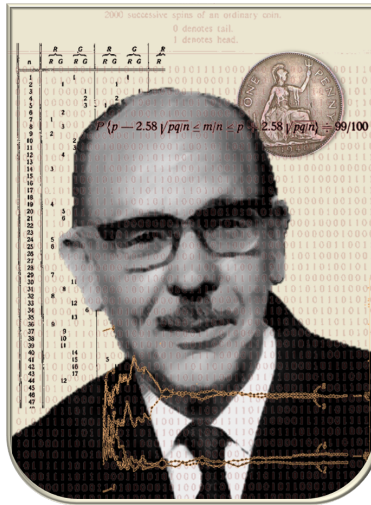
*‘Things of every kind of nature are subject to a universal law which one may well call the **Law of Large Numbers**. It consists in that if one observes large numbers of events of the same nature depending on causes which are constant and causes which vary irregularly, . . . , one finds that the proportions of occurrence are almost constant . . . ’*

Poisson (1837)¹

¹ A Tricentenary history of the Law of Large Numbers—E. Seneta, *Bernoulli*, 2013

Historical experiment

John Kerrich was a mathematician interned in Nazi-occupied Denmark in the 1940s. He performed simple experiments using coins and ping-pong balls to demonstrate the empirical validity of some fundamental laws of probability².



Most famous was the demonstration of LLN.
He tossed a coin **10000 times** and recorded **5067 heads**.

²Published after second world war in a book *Experimental Introduction to the Theory of Probability*

Law of large numbers (LLN)

Informal statement: $\overline{X}_n$ gets closer and closer to the expectation μ as n grows large.

sample (empirical) average $\approx$ population average (expectation)

Precise asymptotic statement: For any $\epsilon > 0$,

$$\mathbb{P}\left[\left|\overline{X}_n - \mu\right| > \epsilon\right] \xrightarrow{n \rightarrow \infty} 0$$

Precise non-asymptotic statement: For any $\epsilon > 0$,

$$\mathbb{P}\left[\left|\overline{X}_n - \mu\right| > \epsilon\right] \leq \frac{\sigma^2}{n\epsilon^2}$$

Examples of LLN

- ▶ **Parking lot:** 500 spots, 600 permits issued. Suppose each person drives to work with probability 80%. Then we typically expect around **480 cars**.
(Apparently not accurate during career fairs or game days).
- ▶ **Home owner insurance:** Liability of each policy

$$X = \begin{cases} \$100k & \text{w.p. 0.1\% (major)} \\ \$50k & \text{w.p. 0.1\% (substantial)} \\ \$10k & \text{w.p. 1\% (minor)} \\ 0 & \text{else} \end{cases}$$

Then $\mathbb{E}[X] = \$250$ is a fair price. The insurance company may set the premium to be \$400 to guarantee a profit typically.

Law of large numbers (LLN)

$$\mathbb{P} [|\bar{X}_n - \mu| > \epsilon] \leq \frac{\sigma^2}{n\epsilon^2}$$

$$\mathbb{E} \bar{X}_n = \mathbb{E} \frac{X_1 + \dots + X_n}{n} = \frac{n\mu}{n} = \mu$$

$$\text{Var} \bar{X}_n = \text{Var} \left[\frac{X_1 + \dots + X_n}{n} \right] = \frac{1}{n^2} n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

$$\mathbb{P} [|\bar{X}_n - \mu| > \epsilon] \stackrel{\text{Chebyshev}}{\leq} \frac{\text{Var} \bar{X}_n}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

$$\begin{aligned} & \text{Var}[X_1 + \dots + X_n] \\ \Rightarrow &= \text{Var} X_1 + \dots + \text{Var} X_n \end{aligned}$$

► Instead of assuming independence, **uncorrelatedness** suffices

Example: HMC §5.1.1

A statistical application of the LLN:

Theorem

Let $X_1, \dots, X_n$ be i.i.d. with finite variance and mean μ . Then the sample mean $\hat{\mu}_n = \overline{X}_n$ is a **consistent estimator** of the population mean:

$$\hat{\mu}_n \xrightarrow{p} \mu.$$

Example: Sample variance (Rice §5.2, Example B)

Suppose that **repeated independent unbiased measurements**, $X_1, \dots, X_n$, of a quantity are made. If n is large, LLN says that $\bar{X}_n$ will be close to the true value μ of the quantity.

But how close $\bar{X}_n$ is depends not only on n but on the **variance of the measurement error** σ^2 . This too can be estimated by LLN.

Estimator of variance: $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$

Expand:

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \underbrace{\frac{1}{n} \sum_{i=1}^n 2X_i \bar{X}}_{2\bar{X} \cdot \underbrace{\frac{1}{n} \sum_{i=1}^n X_i}_{\bar{X}}} + \underbrace{\frac{1}{n} \sum_{i=1}^n \bar{X}^2}_{\bar{X}^2} \\ &= \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X} \end{aligned}$$

Example: Sample variance (Rice §5.2, Example B)

$$\text{Var } \bar{X}_n = \frac{\sigma^2}{n} \leftarrow \text{want to est. this from data}$$

$$\text{Var } X = \underbrace{\mathbb{E}[X^2]}_{P \uparrow \text{LLN}} - \underbrace{(\mathbb{E}X)^2}_{P \uparrow \text{LLN}}$$

Hence

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \left(\frac{1}{n} \sum_{i=1}^n X_i \right)^2$$

i.e.

$$\hat{\sigma}^2 \xrightarrow{P} \text{Var } X$$

Continuous mapping theorem

$\mathcal{H}$

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \mathbb{E}X,$$

then for any cont. func. g ,

$$g\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \xrightarrow{P} g(\mathbb{E}X)$$

Outline

Recap

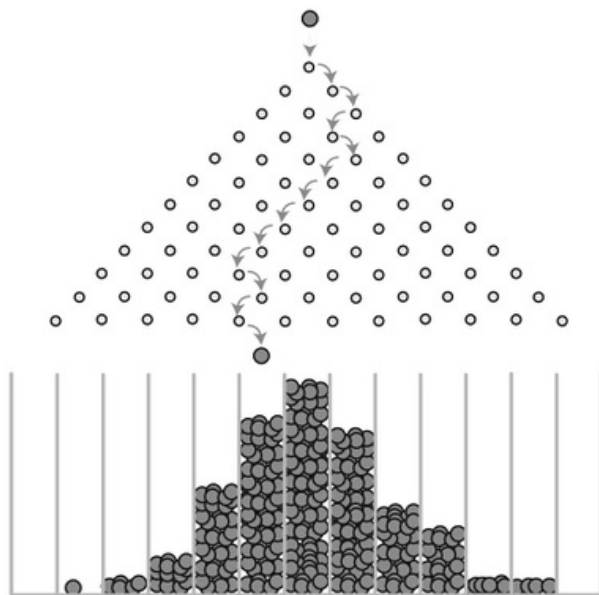
Modes of convergence

Law of large numbers

Central limit theorem

Motivation

Galton board:



- ▶ **Bell-shape** is natural. Many ways to be moderate, few ways to be extreme.
- ▶ The fact that many natural phenomena display **Gaussian** behavior is more surprising.

History

*‘I know of scarcely anything so apt to impress the imagination as the wonderful form of cosmic order expressed by the **law of frequency of error**. A savage, if he could understand it, would worship it as a god. ... Let a large sample of chaotic elements be taken and marshalled in order of their magnitudes, and then, however wildly irregular they appeared, an unexpected and most beautiful form of regularity proves to have been present all along. The larger the mob, the greater the apparent anarchy, the more perfect is its sway. It is the supreme law of unreason. ’*

Francis Galton (1889)

Example: Expo(1)

Let $X_1, \dots, X_n$ i.i.d. Expo(1). Let

$$\textcolor{red}{S}_n = X_1 + \dots + X_n, \quad \overline{X}_n = \frac{S_n}{n}$$

Center and standardize $\overline{X}_n$

► $\mathbb{E}[\overline{X}_n] = \mu, \text{ Var}[\overline{X}_n] = \frac{\sigma^2}{n}$

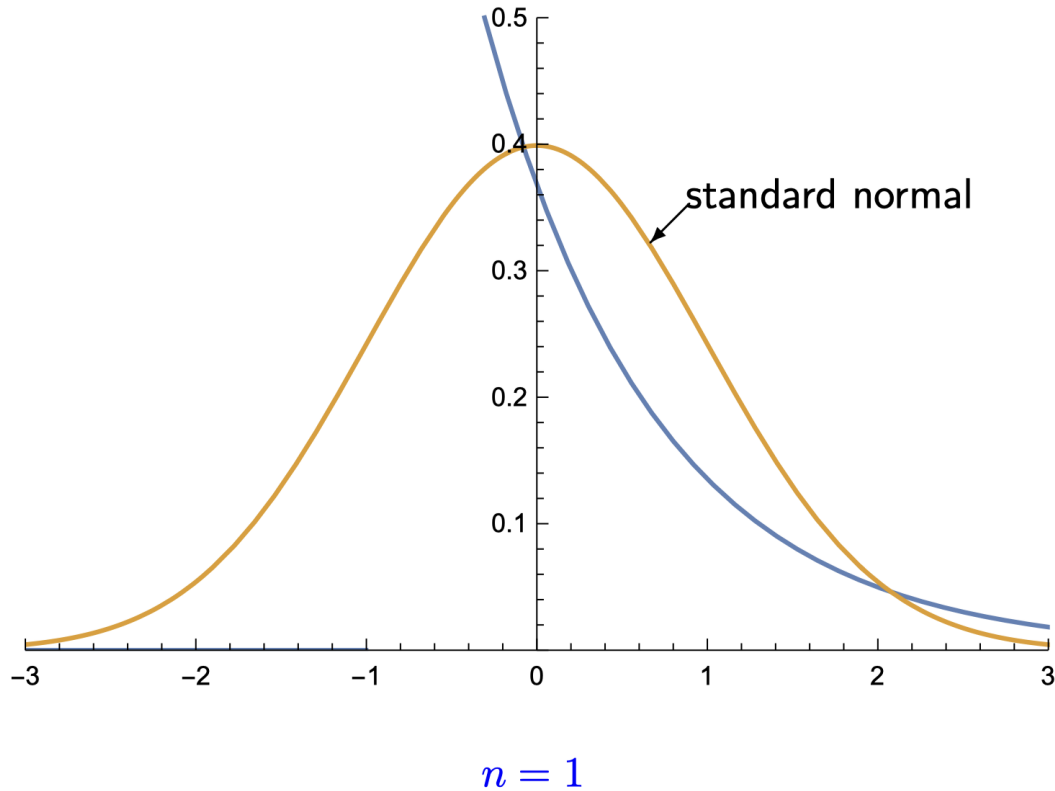
► So instead consider the **standardized** S_n :

$$\frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} = \frac{\textcolor{red}{S}_n - n\mu}{\sqrt{n\sigma^2}}.$$

Example: Expo(1)

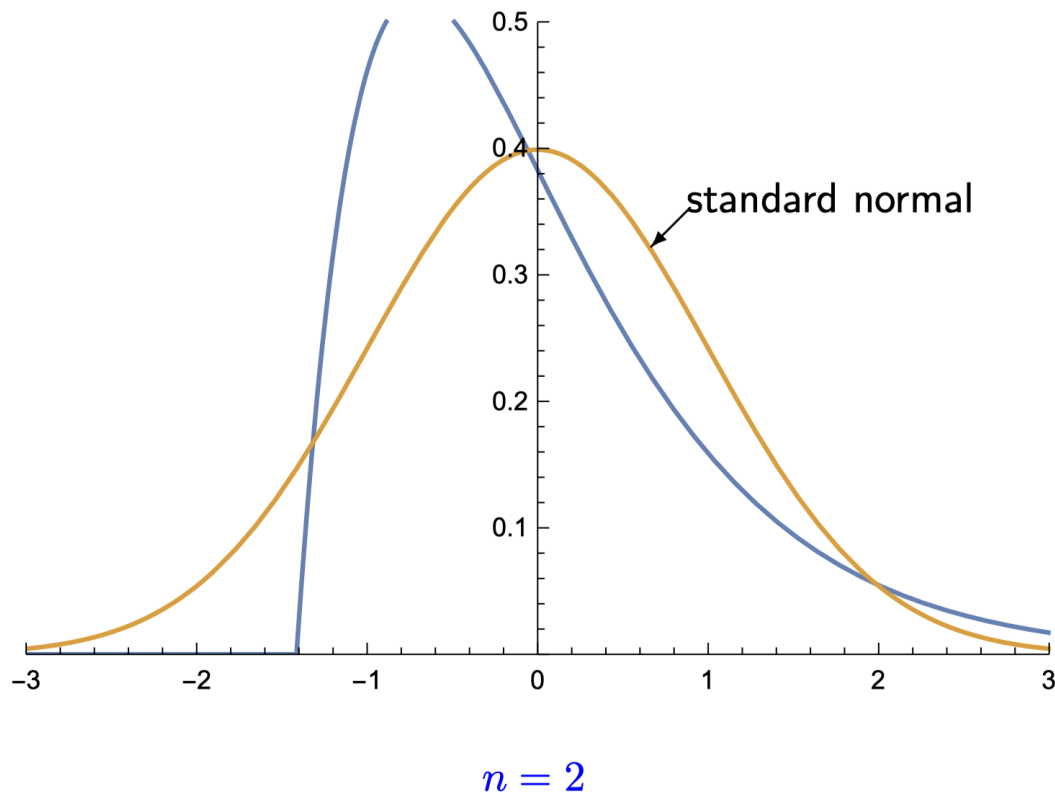
Here $\mu = \sigma^2 = 1$

$X_1, \dots, X_n$ i.i.d. Expo(1). PDF of $(S_n - n)/\sqrt{n}$:



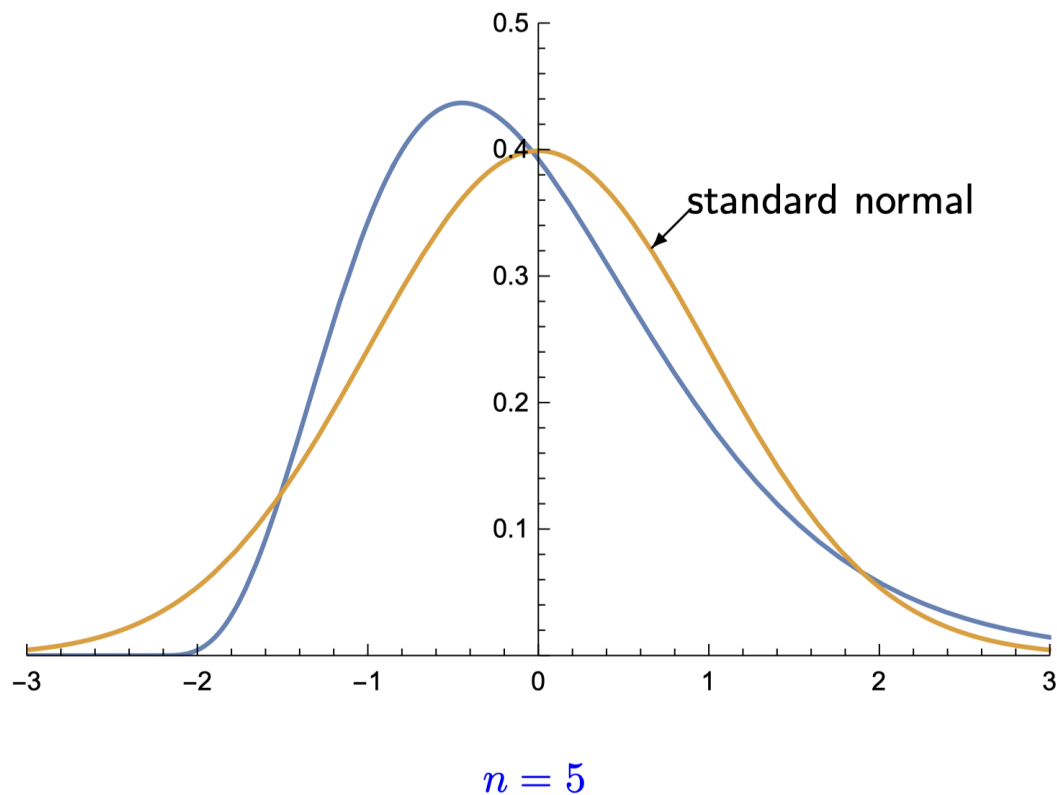
Example: Expo(1)

$X_1, \dots, X_n$ i.i.d. Expo(1). PDF of $(S_n - n)/\sqrt{n}$:



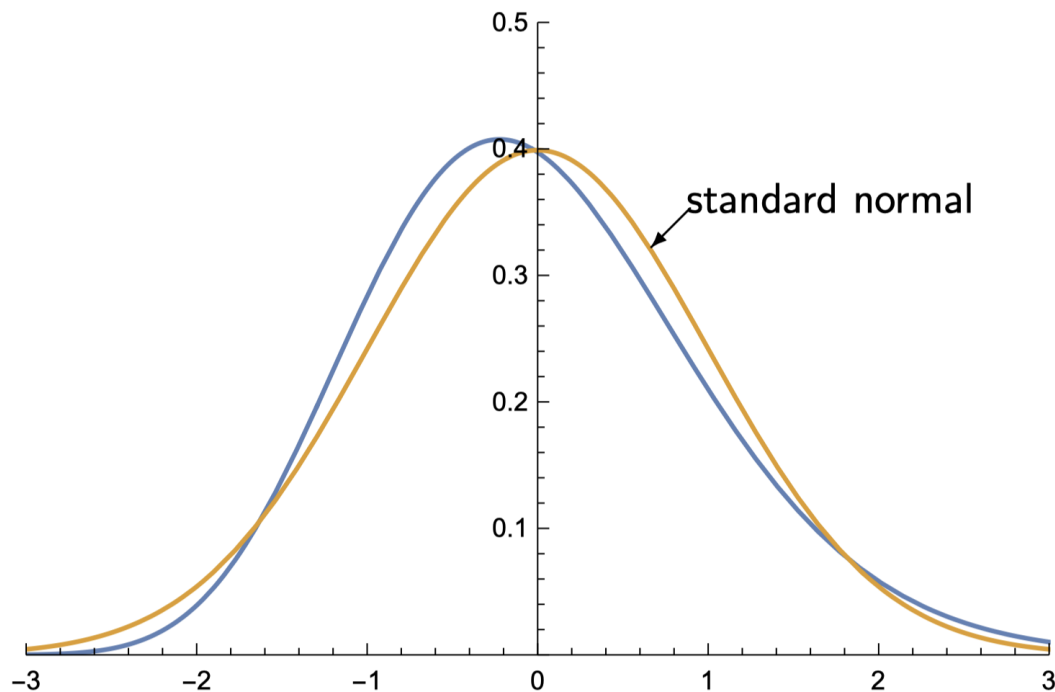
Example: Expo(1)

$X_1, \dots, X_n$ i.i.d. Expo(1). PDF of $(S_n - n)/\sqrt{n}$:



Example: Expo(1)

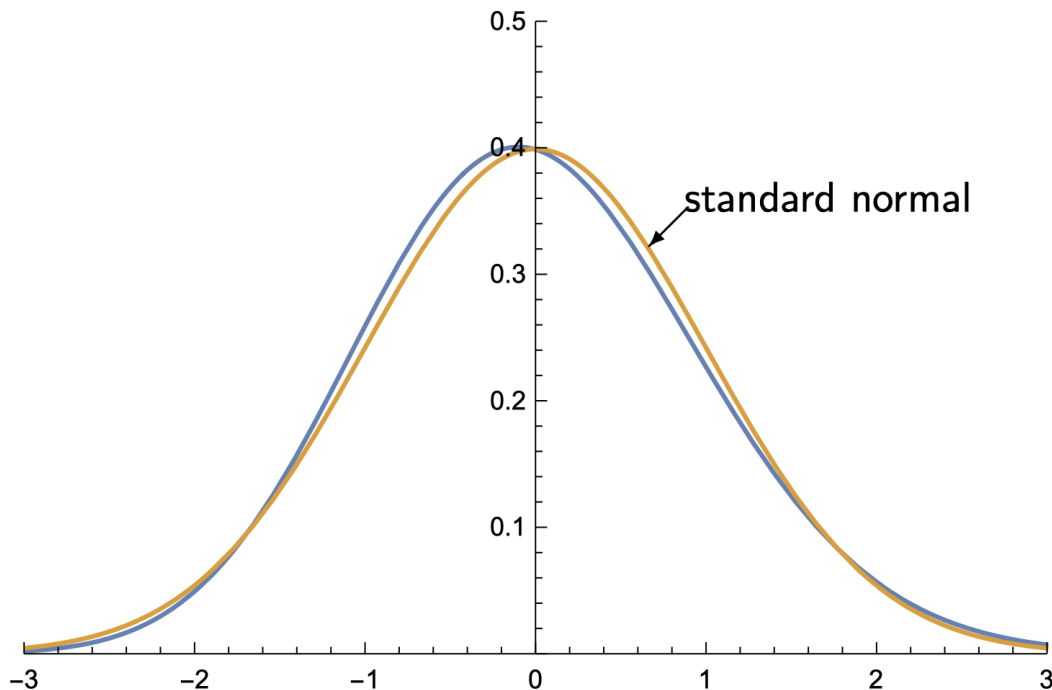
$X_1, \dots, X_n$ i.i.d. Expo(1). PDF of $(S_n - n)/\sqrt{n}$:



$n = 20$

Example: Expo(1)

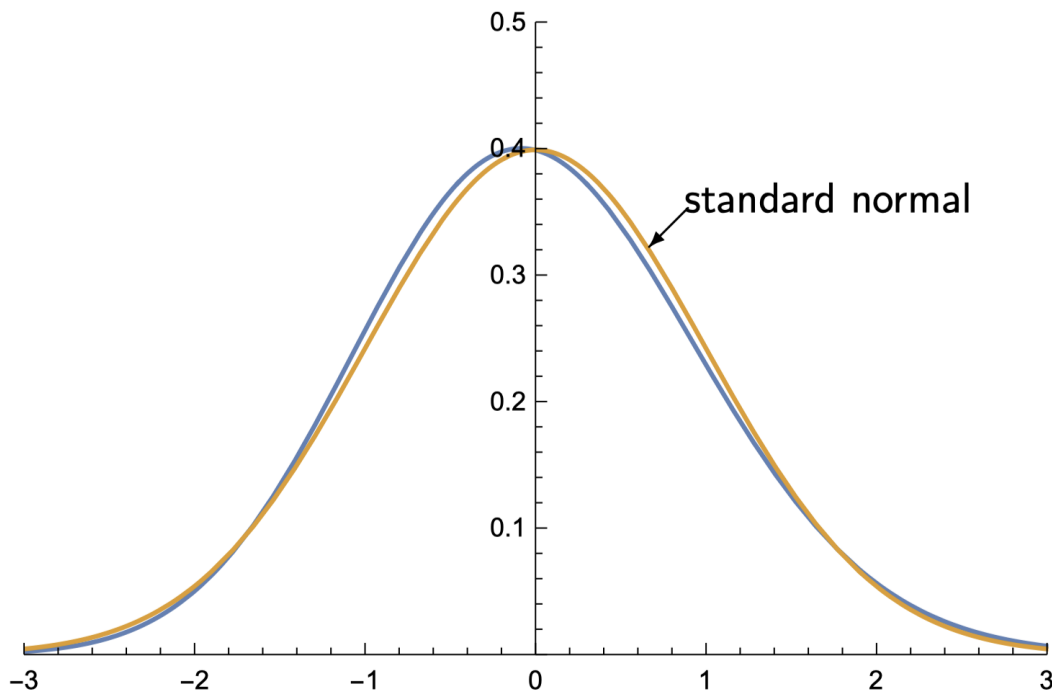
$X_1, \dots, X_n$ i.i.d. Expo(1). PDF of $(S_n - n)/\sqrt{n}$:



$n = 100$

Example: Expo(1)

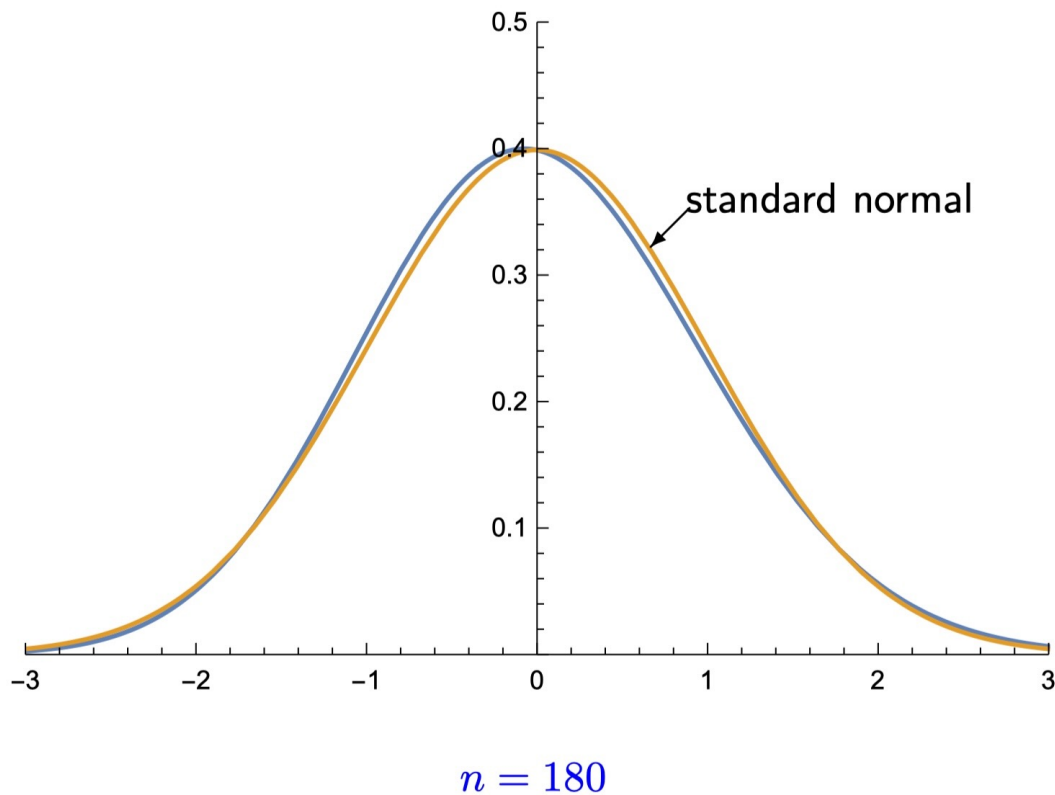
$X_1, \dots, X_n$ i.i.d. Expo(1). PDF of $(S_n - n)/\sqrt{n}$:



$n = 140$

Example: Expo(1)

$X_1, \dots, X_n$ i.i.d. Expo(1). PDF of $(S_n - n)/\sqrt{n}$:



Central limit theorem

Informal

If Y can be written as the sum of a large number of relatively small, independent random variables, then the standardized variable

$$\frac{Y - \mathbb{E}Y}{\sqrt{\text{Var } Y}} \sim \mathcal{N}(0, 1) \text{ approximately}$$

Equivalently, Y is approximately $\mathcal{N}(\mathbb{E}Y, \text{Var } Y)$ distributed.

Precise

Let $X_1, \dots, X_n$ be i.i.d. r.v.'s with mean μ and variance σ^2 . Then the standardized sum $(S_n - n\mu)/\sqrt{n\sigma^2}$ converges to $\mathcal{N}(0, 1)$ in distribution: for any x ,

$$\mathbb{P}\left[\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x\right] \xrightarrow{n \rightarrow \infty} \Phi(x)$$

CLT vs LLN

CLT *refines* the LLN. The CLT says

$$\overline{X}_n \approx \mu + \mathcal{N}\left(0, \frac{\sigma^2}{n}\right)$$

I.e. sample mean $\approx$ population mean + **small Gaussian**

- ▶ Consistent with LLN: as $n \rightarrow \infty$, the error term $\rightarrow 0$
- ▶ CLT: blow everything up by $\sqrt{n}$. Then the Gaussian error is non-negligible in the limit:

$$\sqrt{n} \cdot \overline{X}_n = \frac{S_n}{\sqrt{n}} \approx \sqrt{n}\mu + \mathcal{N}(0, \sigma^2)$$

Ways to write CLT

CLT says

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} \mathcal{N}(0, 1).$$

Equivalent forms of CLT:

$$\begin{aligned}\sqrt{n} \left(\frac{\overline{X}_n - \mu}{\sigma} \right) &\xrightarrow{d} \mathcal{N}(0, 1) \\ \sqrt{n} (\overline{X}_n - \mu) &\xrightarrow{d} \mathcal{N}(0, \sigma^2)\end{aligned}$$

These highlight the fact that $\overline{X}_n$ has fluctuations on the order of $\frac{1}{\sqrt{n}}$ (because $\text{Var}[\overline{X}_n] = \frac{\sigma^2}{n}$).

Example: Running proportion of heads

Go through
next class

Let $X_1, X_2, \dots$ be i.i.d. Bernoulli(1/2). LLN tells us

$$\bar{X}_n \xrightarrow{n \rightarrow \infty} \frac{1}{2} \quad \text{in probability}$$

CLT gives more. Use CLT to approximate the chance that $\bar{X}_n$ will be in the interval $[0.4, 0.6]$. CLT: $\bar{X}_n \stackrel{d}{\approx} \mathcal{N}(\mu, \frac{\sigma^2}{n}) = \mathcal{N}(\frac{1}{2}, \frac{1}{4n})$

$$\mu = \mathbb{E}(X) = \frac{1}{2}, \quad \sigma^2 = \text{Var } X = \frac{1}{2}(1 - \frac{1}{2}) = \frac{1}{4}$$

Suppose $n=100$, then

$$\text{SD} = \frac{1}{20}$$

$$P[0.4 \leq \bar{X}_{100} \leq 0.6] \approx P[0.4 \leq \mathcal{N}(\frac{1}{2}, \frac{1}{400}) \leq 0.6]$$

$$= P\left[\frac{0.4 - 0.5}{1/20} \leq Z \leq \frac{0.6 - 0.5}{1/20}\right]$$

$$= P[-2 \leq Z \leq 2] \approx 95\%$$

check that LLN

$$\text{giving } \leq \frac{\sigma^2}{n(0.1)^2} = \frac{1}{400(0.1)^2}$$

$$= 0.25 \text{ is two}$$

Conservative