S&DS 241 Lecture 24

Central limit theorem

B-H: 10.3

Galton board

https://www.youtube.com/watch?v=6YDHBFVIvIs

Galton board

https://www.youtube.com/watch?v=6YDHBFVIvIs

- This is explained by de Moivre-Laplace CLT for binomials (Lec 16):
- The distribution is bell-shaped is natural: there are very few ways to reach the extreme and much more ways to be moderate; but why Gaussian arises is perhaps surprising.
- Universality of Gaussian

Sir Francis Galton on CLT '1889

"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."

Recall LLN

Let $X_1, \ldots, X_n$ be iid with mean μ and variance σ^2. Let $S_n = X_1 + \cdots X_n$ and $\overline{X}_n = \frac{S_n}{n}$.

- Law of Large Numbers (LLN): $\overline{X}_n$ converges to μ in the sense that:

$$P(|\overline{X}_n - \mu| > \epsilon) \xrightarrow{n\to\infty} 0, \quad \text{for any } \epsilon > 0$$

  that is, $\overline{X}_n$ becomes increasingly concentrated near μ.

Recall LLN

Let $X_1, \dots, X_n$ be iid with mean μ and variance σ^2. Let $S_n = X_1 + \cdots X_n$ and $\overline{X}_n = \frac{S_n}{n}$.

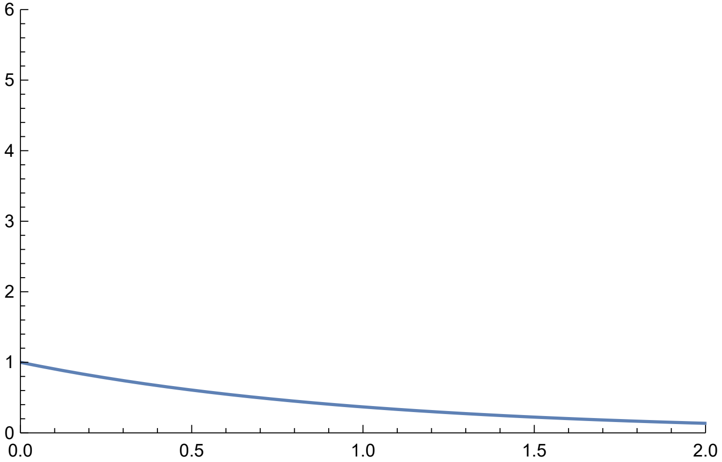
- Law of Large Numbers (LLN): $\overline{X}_n$ converges to μ in the sense that:

$$P(|\overline{X}_n - \mu| > \epsilon) \xrightarrow{n \to \infty} 0, \quad \text{for any } \epsilon > 0$$

  that is, $\overline{X}_n$ becomes increasingly concentrated near μ.
- Central Limit Theorem (CLT) tells us the shape.
- Let's look at an example: $X_i \sim \text{Expo}(1)$.

PDF of $\overline{X}_n$

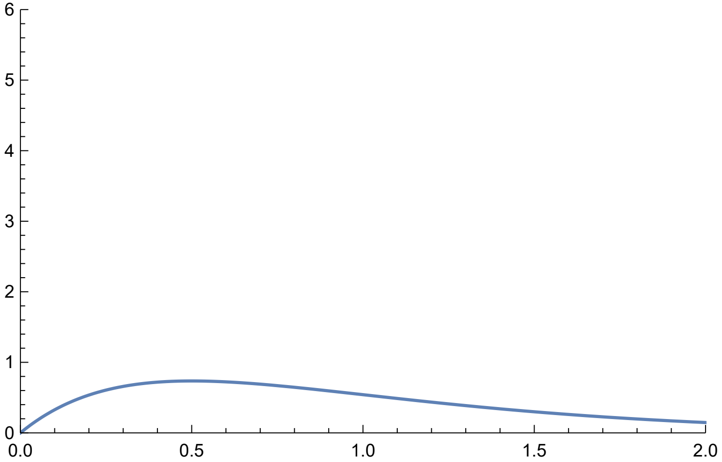
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 1$

PDF of $\overline{X}_n$

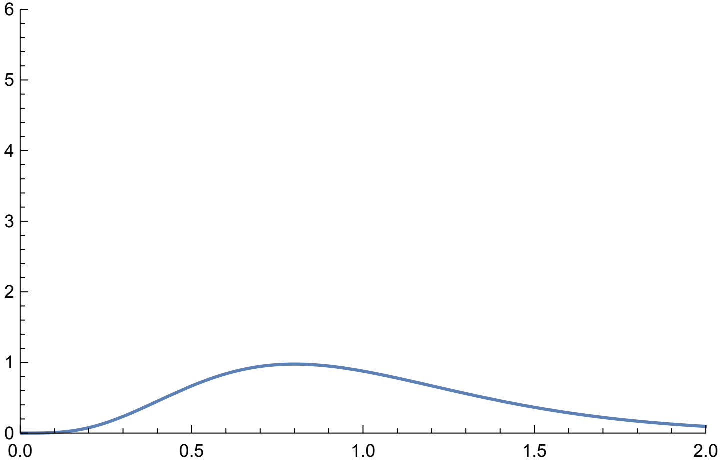
$X_1, \dots, X_n$ are iid Expo(1):



$n = 2$

PDF of $\overline{X}_n$

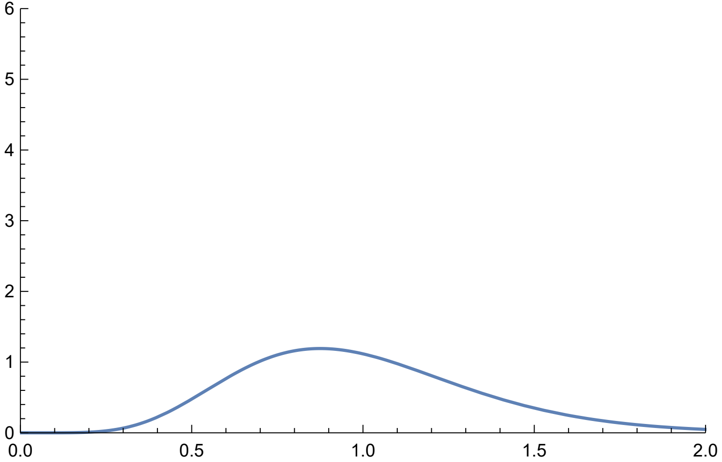
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 5$

PDF of $\overline{X}_n$

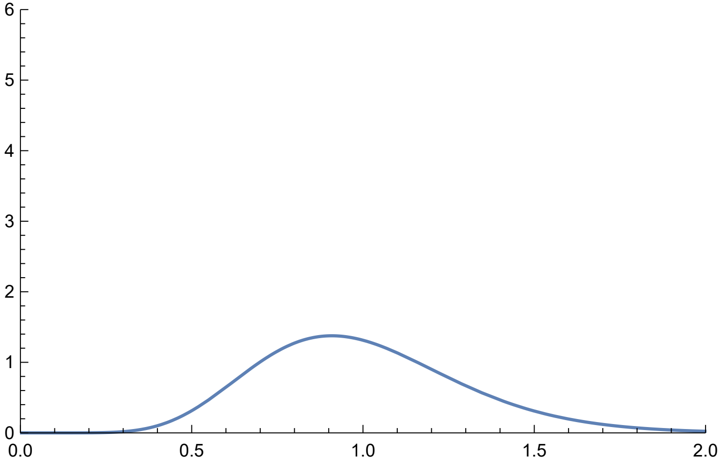
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 8$

PDF of $\overline{X}_n$

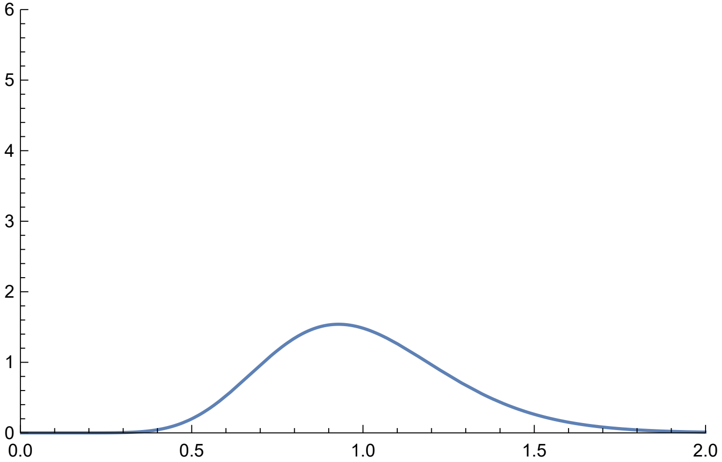
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 11$

PDF of $\overline{X}_n$

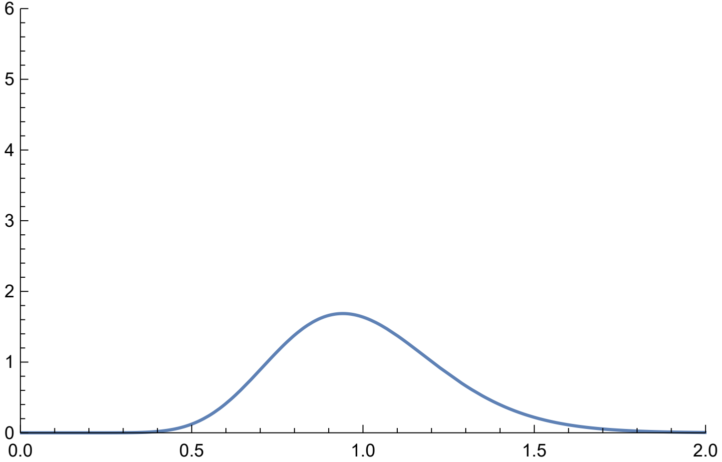
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 14$

PDF of $\overline{X}_n$

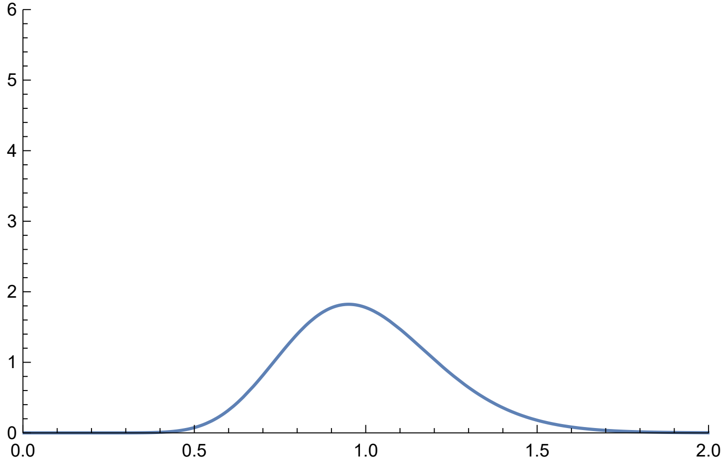
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 17$

PDF of $\overline{X}_n$

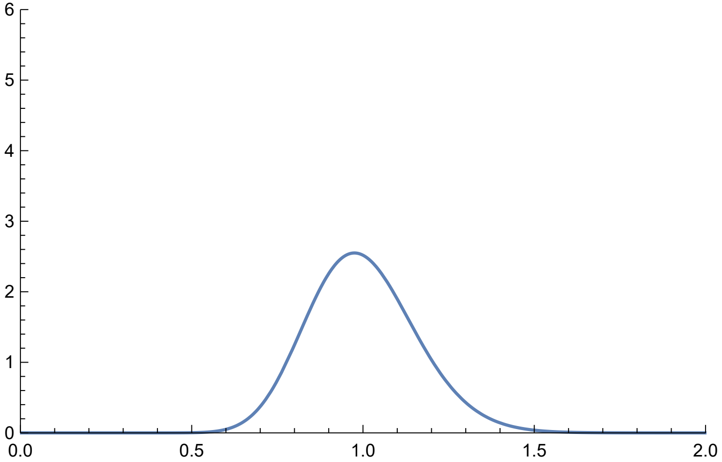
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 20$

PDF of $\overline{X}_n$

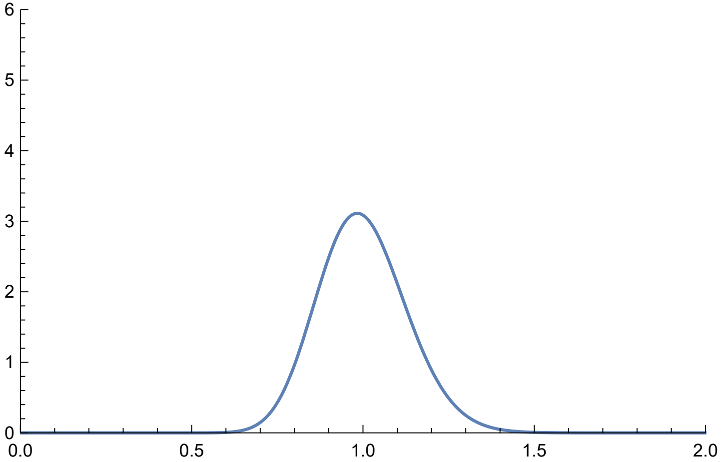
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 40$

PDF of $\overline{X}_n$

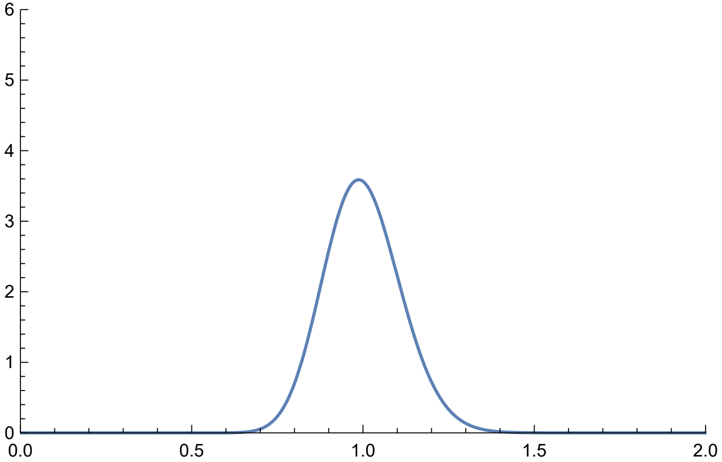
$X_1, \dots, X_n$ are iid Expo(1):



$n = 60$

PDF of $\overline{X}_n$

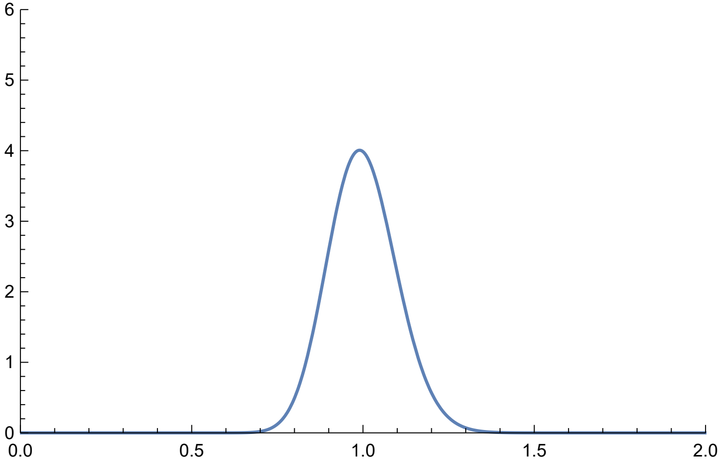
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 80$

PDF of $\overline{X}_n$

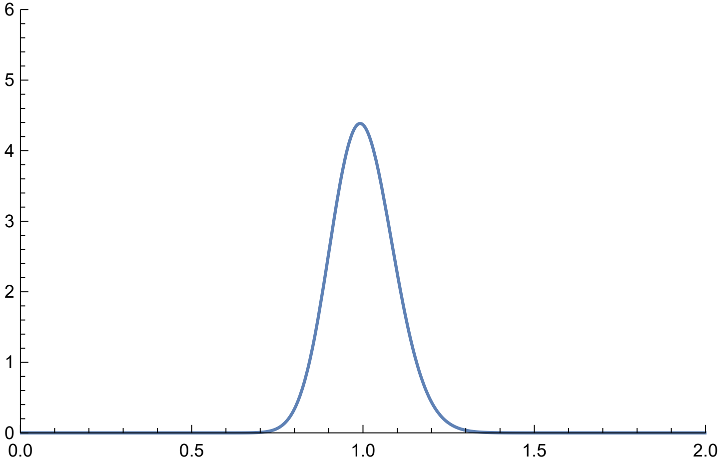
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 100$

PDF of $\overline{X}_n$

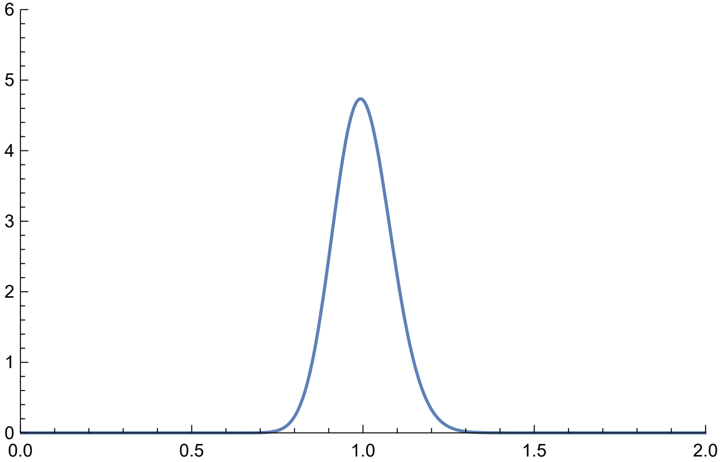
$X_1, \dots, X_n$ are iid Expo(1):



$n = 120$

PDF of $\overline{X}_n$

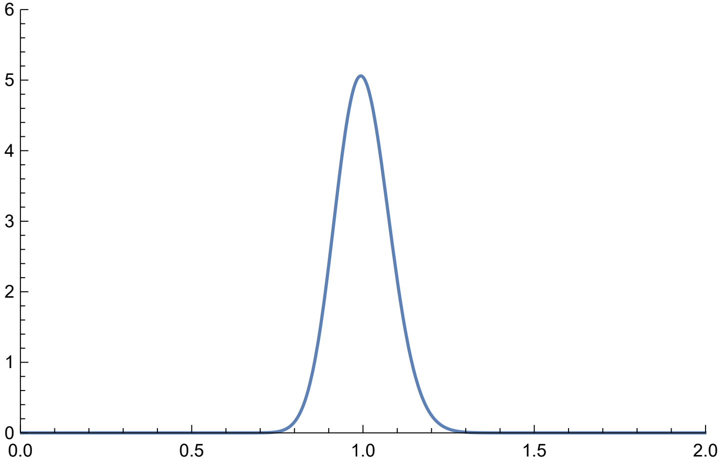
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 140$

PDF of $\overline{X}_n$

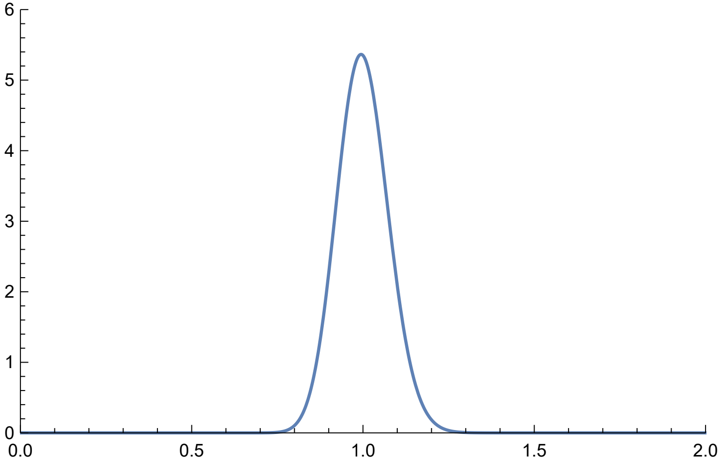
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 160$

PDF of $\overline{X}_n$

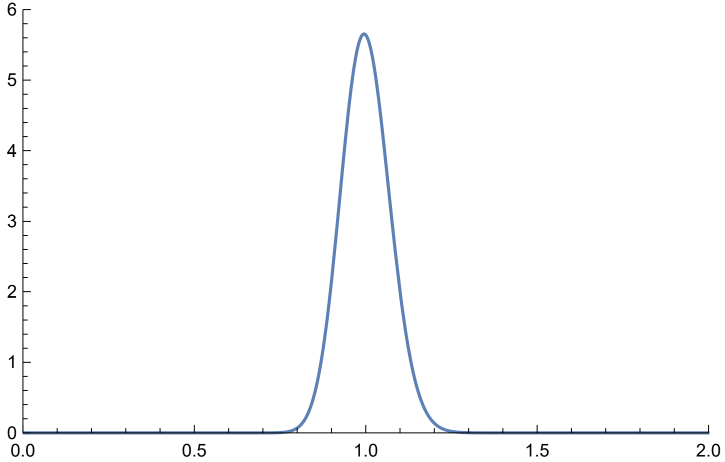
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 180$

PDF of $\overline{X}_n$

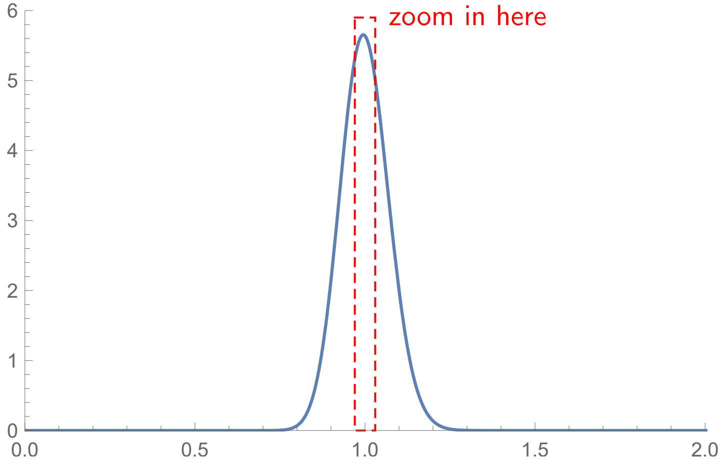
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 200$

PDF of $\overline{X}_n$

$X_1, \ldots, X_n$ are iid Expo(1):



$n = 200$

How to zoom in?

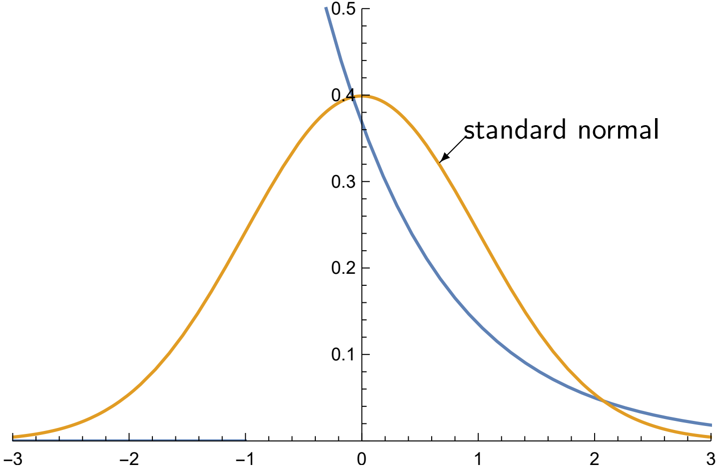
Standardize (center and normalize) $\overline{X}_n$:

- $E(\overline{X}_n) = \mu$ and $\mathrm{Var}(\overline{X}_n) = \frac{\sigma^2}{n}$.
- So let's consider the standardized version of $\overline{X}_n$:

$$\frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}}$$

PDF of $(S_n - n)/\sqrt{n}$

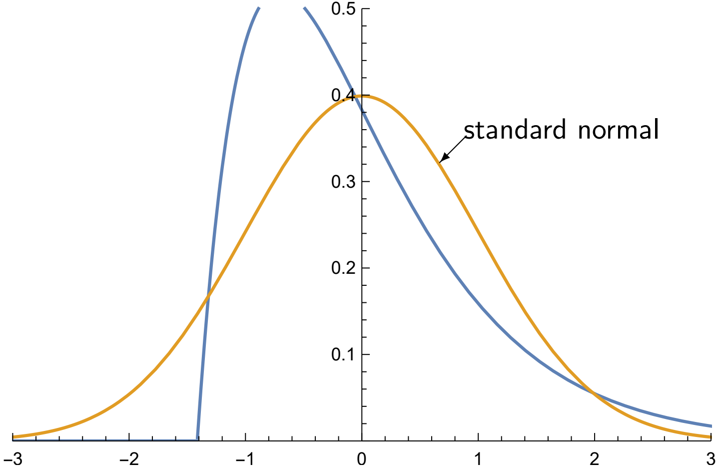
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 1$

PDF of $(S_n - n)/\sqrt{n}$

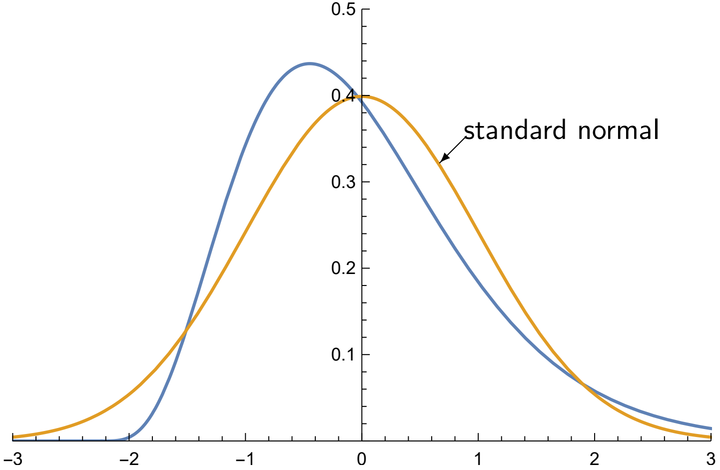
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 2$

PDF of $(S_n - n)/\sqrt{n}$

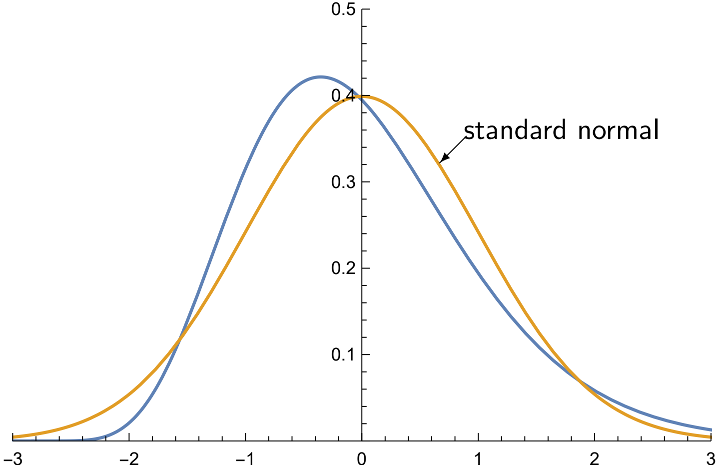
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 5$

PDF of $(S_n - n)/\sqrt{n}$

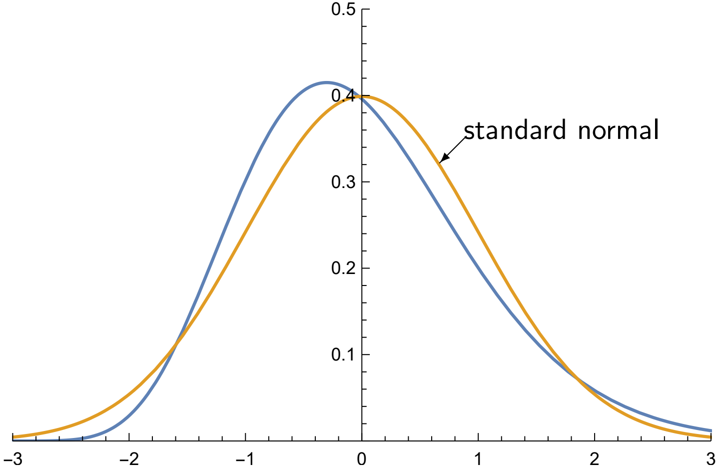
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 8$

PDF of $(S_n - n)/\sqrt{n}$

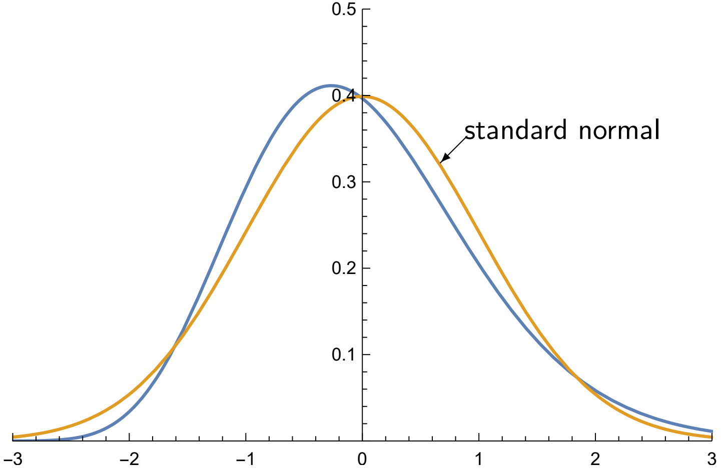
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 11$

PDF of $(S_n - n)/\sqrt{n}$

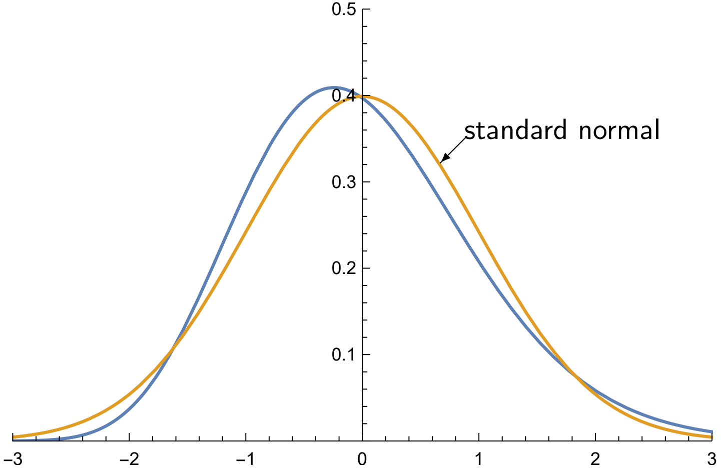
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 14$

PDF of $(S_n - n)/\sqrt{n}$

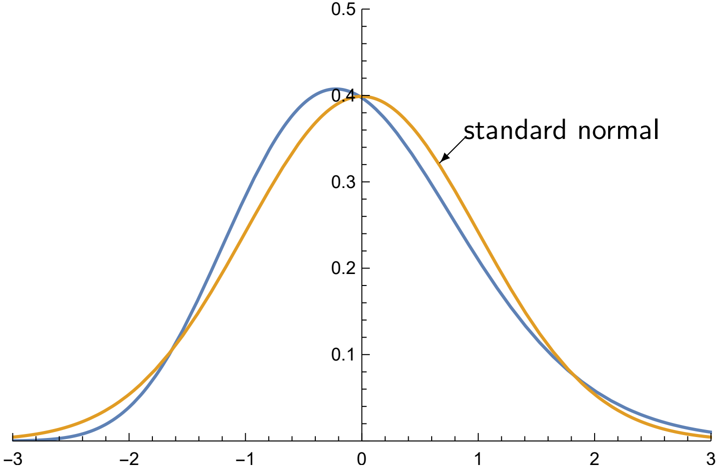
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 17$

PDF of $(S_n - n)/\sqrt{n}$

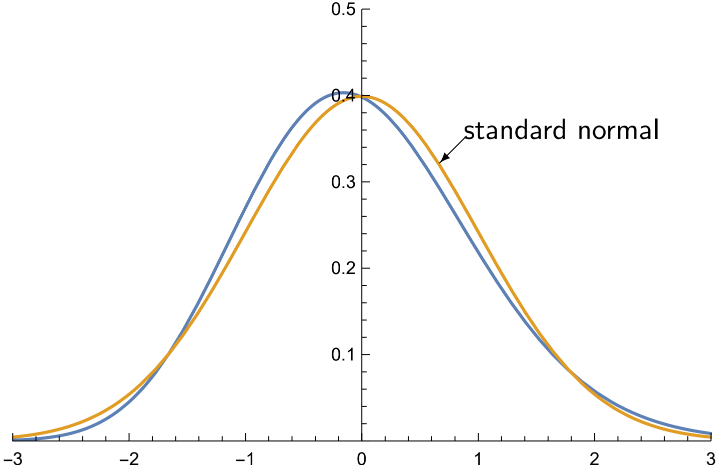
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 20$

PDF of $(S_n - n)/\sqrt{n}$

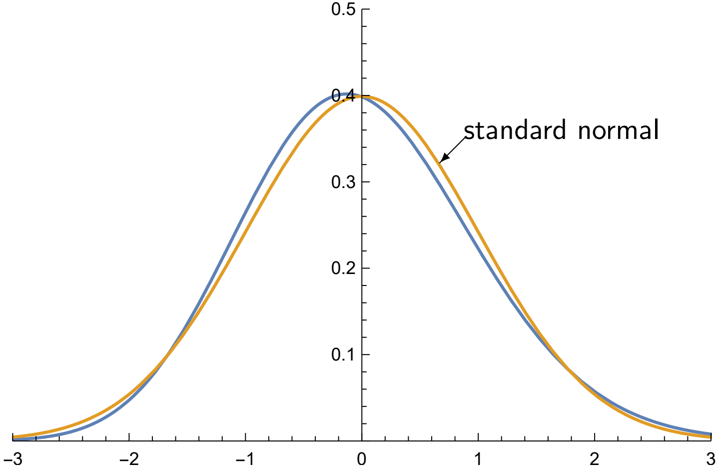
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 40$

PDF of $(S_n - n)/\sqrt{n}$

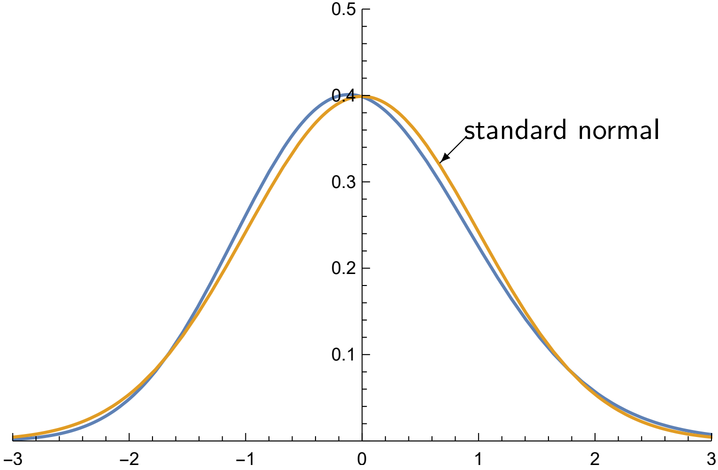
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 60$

PDF of $(S_n - n)/\sqrt{n}$

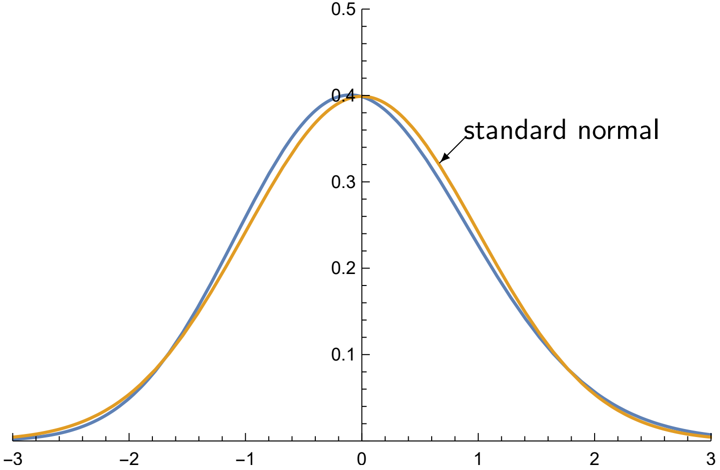
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 80$

PDF of $(S_n - n)/\sqrt{n}$

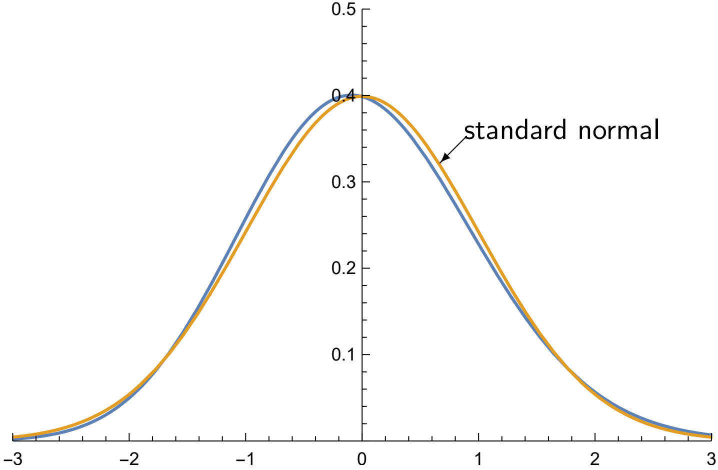
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 100$

PDF of $(S_n - n)/\sqrt{n}$

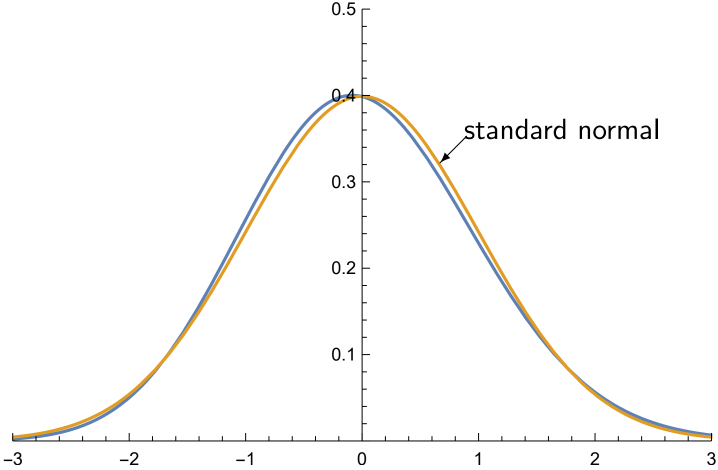
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 120$

PDF of $(S_n - n)/\sqrt{n}$

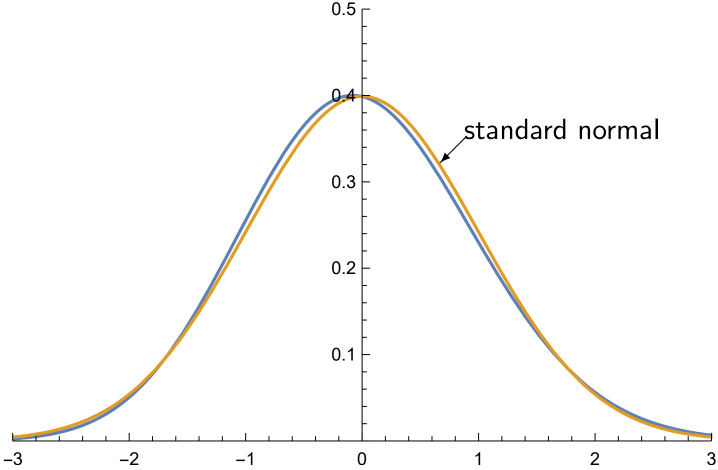
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 140$

PDF of $(S_n - n)/\sqrt{n}$

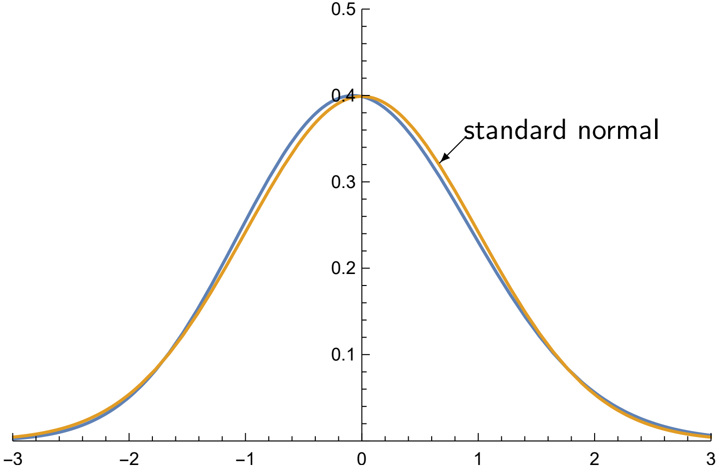
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 160$

PDF of $(S_n - n)/\sqrt{n}$

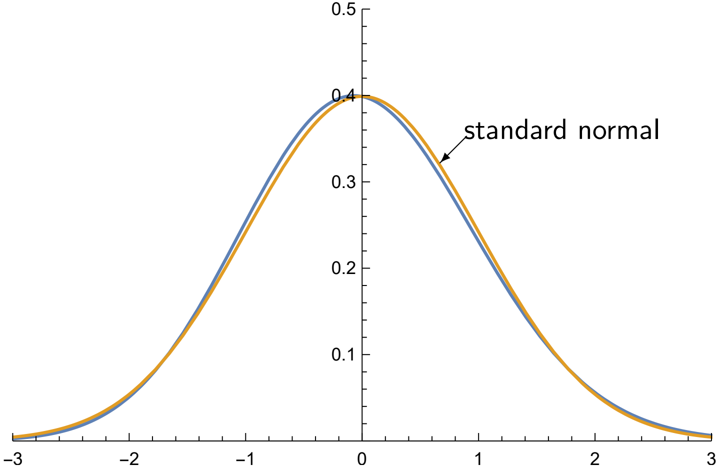
$X_1, \ldots, X_n$ are iid Expo(1):



$n = 180$

PDF of $(S_n - n)/\sqrt{n}$

$X_1, \ldots, X_n$ are iid Expo(1):



$n = 200$

Central limit theorem

Theorem (CLT)

Let $X_1, X_2, \dots$ *be iid with mean* μ *and variance* σ^2*. Let* $S_n = X_1 + \cdots X_n$*. Then* $\frac{S_n - n\mu}{\sqrt{n\sigma^2}}$ *is approximately standard normal (in the sense of CDF):*

$$P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x\right) \xrightarrow{n\to\infty} \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, \quad \forall x \in \mathbb{R}.$$

Central limit theorem

Theorem (CLT)

Let $X_1, X_2, \ldots$ *be iid with mean* μ *and variance* σ^2*. Let* $S_n = X_1 + \cdots X_n$*. Then* $\frac{S_n - n\mu}{\sqrt{n\sigma^2}}$ *is approximately standard normal (in the sense of CDF):*

$$P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x\right) \xrightarrow{n\to\infty} \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, \quad \forall x \in \mathbb{R}.$$

- CLT refines the LLN: "sample mean $\approx$ population mean + small Gaussian"
$$\overline{X}_n \approx \mu + N\left(0, \frac{\sigma^2}{n}\right)$$

Central limit theorem

Theorem (CLT)

Let $X_1, X_2, \ldots$ be iid with mean μ and variance σ^2. Let $S_n = X_1 + \cdots X_n$. Then $\frac{S_n - n\mu}{\sqrt{n\sigma^2}}$ is approximately standard normal (in the sense of CDF):

$$P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x\right) \xrightarrow{n\to\infty} \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, \quad \forall x \in \mathbb{R}.$$

- CLT refines the LLN: "sample mean $\approx$ population mean + small Gaussian"
$$\overline{X}_n \approx \mu + N\left(0, \frac{\sigma^2}{n}\right)$$
- Special case $X_i \sim \text{Bern}(p)$: de Moivre-Laplace CLT for Binomial distribution (Lec 16)

Central limit theorem

Theorem (CLT)

Let $X_1, X_2, \ldots$ be iid with mean μ and variance σ^2. Let $S_n = X_1 + \cdots X_n$. Then $\frac{S_n - n\mu}{\sqrt{n\sigma^2}}$ is approximately standard normal (in the sense of CDF):

$$P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x\right) \xrightarrow{n\to\infty} \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, \quad \forall x \in \mathbb{R}.$$

- CLT refines the LLN: "sample mean ≈ population mean + small Gaussian"
$$\overline{X}_n \approx \mu + N\left(0, \frac{\sigma^2}{n}\right)$$
- Special case $X_i \sim \text{Bern}(p)$: de Moivre-Laplace CLT for Binomial distribution (Lec 16)
- Next we prove this result using the apparatus of MGF. WLOG, assume $\mu = 0$ and $\sigma = 1$; otherwise replace X_i by $(X_i - \mu)/\sigma$.

Central limit theorem

Theorem (CLT)

Let $X_1, X_2, \dots$ be iid with mean μ and variance σ^2. Let $S_n = X_1 + \cdots X_n$. Then $\frac{S_n - n\mu}{\sqrt{n\sigma^2}}$ is approximately standard normal (in the sense of CDF):

$$P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \le x\right) \xrightarrow{n\to\infty} \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt, \quad \forall x \in \mathbb{R}.$$

- CLT refines the LLN: "sample mean $\approx$ population mean + small Gaussian"
$$\overline{X}_n \approx \mu + N\left(0, \frac{\sigma^2}{n}\right)$$
- Special case $X_i \sim \text{Bern}(p)$: de Moivre-Laplace CLT for Binomial distribution (Lec 16)
- Next we prove this result using the apparatus of MGF. WLOG, assume $\mu = 0$ and $\sigma = 1$; otherwise replace X_i by $(X_i - \mu)/\sigma$.
- Goal: show $\frac{S_n}{\sqrt{n}}$ is approximately distributed as $N(0, 1)$.

MGF

Recall moment generating function (MGF)

$$M_X(t) = E(e^{tX})$$

Example: $X \sim N(0,1)$, $M_X(t) = e^{t^2/2}$.

Three facts about MGF useful for proving CLT

- Closeness of MGFs implies closeness of distribution (not proved)

$$
\begin{aligned}
& M_X(t) \approx e^{t^2/2} \\
\implies & X \approx \text{standard normal (in the sense of CDF)}
\end{aligned}
$$

Three facts about MGF useful for proving CLT

- Closeness of MGFs implies closeness of distribution (not proved)

$$M_X(t) \approx e^{t^2/2}$$
$$\implies X \approx \text{standard normal (in the sense of CDF)}$$

- For independent X and Y (Lec 23):

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

This allows us to sidestep convolution!!

Three facts about MGF useful for proving CLT

- Closeness of MGFs implies closeness of distribution (not proved)

$$M_X(t) \approx e^{t^2/2}$$
$$\implies X \approx \text{standard normal (in the sense of CDF)}$$

- For independent X and Y (Lec 23):

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

  This allows us to sidestep convolution!!

- Taylor expansion at $t = 0$ (Lec 23):

$$M_X(t) = \underbrace{M_X(0)}_{1} + \underbrace{M'_X(0)}_{E(X)}\,t + \underbrace{M''_X(0)}_{E(X^2)}\,t^2/2 + o(t^2)$$

Proof of CLT

Asymptotic behavior of MGF of $\frac{S_n}{\sqrt{n}}$: as $n \to \infty$

$$M_{\frac{S_n}{\sqrt{n}}}(t) = M_{S_n}\left(\frac{t}{\sqrt{n}}\right)$$

Proof of CLT

Asymptotic behavior of MGF of $\frac{S_n}{\sqrt{n}}$: as $n \to \infty$

$$
\begin{aligned}
M_{\frac{S_n}{\sqrt{n}}}(t) &= M_{S_n}\left(\frac{t}{\sqrt{n}}\right) \\
&= M_{X_1}\left(\frac{t}{\sqrt{n}}\right) \times \cdots \times M_{X_n}\left(\frac{t}{\sqrt{n}}\right) \\
&= \left(M_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n
\end{aligned}
$$

Proof of CLT

Asymptotic behavior of MGF of $\frac{S_n}{\sqrt{n}}$: as $n \to \infty$

$$\begin{aligned} M_{\frac{S_n}{\sqrt{n}}}(t) &= M_{S_n}\left(\frac{t}{\sqrt{n}}\right) \\ &= M_{X_1}\left(\frac{t}{\sqrt{n}}\right) \times \cdots \times M_{X_n}\left(\frac{t}{\sqrt{n}}\right) \\ &= \left(M_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n \\ &= \left(1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \qquad E(X_1) = 0, E(X_1^2) = 1 \end{aligned}$$

Proof of CLT

Asymptotic behavior of MGF of $\frac{S_n}{\sqrt{n}}$: as $n \to \infty$

$$\begin{aligned} M_{\frac{S_n}{\sqrt{n}}}(t) &= M_{S_n}\left(\frac{t}{\sqrt{n}}\right) \\ &= M_{X_1}\left(\frac{t}{\sqrt{n}}\right) \times \cdots \times M_{X_n}\left(\frac{t}{\sqrt{n}}\right) \\ &= \left(M_{X_1}\left(\frac{t}{\sqrt{n}}\right)\right)^n \\ &= \left(1 + \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \\ &\to e^{\frac{t^2}{2}} \end{aligned}$$

$$E(X_1) = 0, E(X_1^2) = 1$$

Calculus: $\left(1 + \frac{x}{n}\right)^n \to e^x$

LLN vs CLT

- CLT is a more refined result than LLN
- LLN only requires uncorrelatedness, CLT requires independence.

Example: Poisson

Let $X \sim \text{Pois}(n)$. Then as n grows, X is approximately $N(n, n)$.
Why?

Example: Poisson

Let $X \sim \text{Pois}(n)$. Then as n grows, X is approximately $N(n, n)$. Why?

- Recall the property of Poisson distribution: if $A \sim \text{Pois}(\lambda)$ and $B \sim \text{Pois}(\mu)$, then $A + B \sim \text{Pois}(\lambda + \mu)$
- Thus we can write

  $$X = X_1 + \cdots + X_n$$

  where $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Pois}(1)$, with unit mean and variance.
- Applying CLT justifies the normal approximation

Application: stock

The price of a stock behaves independently each day, which goes up by 1% with probability 0.5, goes down by 1% with probability 0.1, or stays put with probability 0.4. Buy the stock at $\$1$ and hold for one year. What is the chance to triple the value?

Application: stock

The price of a stock behaves independently each day, which goes up by 1% with probability 0.5, goes down by 1% with probability 0.1, or stays put with probability 0.4. Buy the stock at $\$1$ and hold for one year. What is the chance to triple the value?

- The price after one year $Y = D_1 D_2 \cdots D_{365}$, where D_i are iid with

$$D_i = \begin{cases} 1.01 & \text{w.p. } 0.5 \\ 1 & \text{w.p. } 0.4 \\ 0.99 & \text{w.p. } 0.1 \end{cases}$$

- Question: $P(Y \geq 3)$

Application: stock

- To apply CLT, we take the logarithm to turn products into sums:

$$\ln Y = X_1 + X_2 + \cdots + X_{365},$$

where $X_i = \ln D_i$ are iid with

$$X_i = \begin{cases} \ln 1.01 & \text{w.p. } 0.5 \\ 0 & \text{w.p. } 0.4 \\ \ln 0.99 & \text{w.p. } 0.1 \end{cases}$$

- Then $\mu = 3.97 \times 10^{-3}$ and $\sigma^2 = 4.38 \times 10^{-5}$.
- CLT says $\ln Y$ is approximately distributed as

$$N(n\mu, n\sigma^2) = N(1.45, (0.127)^2)$$

Application: stock

- Chance of tripling:

$$P(Y \geq 3) = P(\ln Y \geq \ln 3) \overset{\text{CLT}}{\approx} 1 - \Phi\left(\frac{\ln 3 - 1.45}{0.127}\right) = 99.7\%$$

Application: stock

- Chance of tripling:

$$P(Y \geq 3) = P(\ln Y \geq \ln 3) \overset{\text{CLT}}{\approx} 1 - \Phi\left(\frac{\ln 3 - 1.45}{0.127}\right) = 99.7\%$$

$$P(Y \geq 4) = P(\ln Y \geq \ln 4) \overset{\text{CLT}}{\approx} 1 - \Phi\left(\frac{\ln 4 - 1.45}{0.127}\right) = 69\%$$

Application: stock

- Chance of tripling:

$$P(Y \geq 3) = P(\ln Y \geq \ln 3) \overset{\text{CLT}}{\approx} 1 - \Phi\left(\frac{\ln 3 - 1.45}{0.127}\right) = 99.7\%$$

$$P(Y \geq 4) = P(\ln Y \geq \ln 4) \overset{\text{CLT}}{\approx} 1 - \Phi\left(\frac{\ln 4 - 1.45}{0.127}\right) = 69\%$$

- Median return:

$$\text{CLT} \implies \text{median of } \ln Y \approx 1.45 \implies \text{median of } Y \approx e^{1.45} = 4.26$$

Application: stock

- Chance of tripling:

$$P(Y \geq 3) = P(\ln Y \geq \ln 3) \overset{\text{CLT}}{\approx} 1 - \Phi\left(\frac{\ln 3 - 1.45}{0.127}\right) = 99.7\%$$

$$P(Y \geq 4) = P(\ln Y \geq \ln 4) \overset{\text{CLT}}{\approx} 1 - \Phi\left(\frac{\ln 4 - 1.45}{0.127}\right) = 69\%$$

- Median return:

$$\text{CLT} \implies \text{median of } \ln Y \approx 1.45 \implies \text{median of } Y \approx e^{1.45} = 4.26$$

- Alternatively: on average, expect the stock $+1\%$ on $365/2$ days and -1% on $365/10$. So overall $(1.01)^{365/2}(0.99)^{365/10} \approx 4.26$