**Math 170E: Winter 2023**

Lecture 26, Fri 17th Mar

The Central Limit Theorem and approximations for discrete distributions

## Last time:

- Let $X_1, X_2 \ldots,$ be random variables.

- We say that $X_n \to X$ in distribution as $n \to \infty$ if the CDFs

$$F_{X_n}(x) \to F_X(x) \text{ as } n \to \infty$$

for all $x \in \mathbb{R}$ where $F_X(x)$ is continuous at $x$.

- Theorem **~~Proposition~~ 5.14: (Limiting MGF determines the distribution)**

Let $X_1, X_2, \ldots$ and $X$ be random variables. Suppose that for some $h > 0$ and all $t \in (-h, h)$, we have

$$M_{X_n}(t) \to M_X(t) \quad \text{as } n \to \infty.$$

Then, $X_n \to X$ in distribution as $n \to \infty$.

## Example 11:

- Let $\lambda > 0$ and $X_n \sim \text{Binomial}(n, \frac{\lambda}{n})$
- Show that $X_n$ converges in distribution and find its limit.

$$M_{X_n}(t) = \left(1 - \frac{\lambda}{n} + \frac{\lambda}{n}e^t\right)^n \text{ for all } t \in \mathbb{R}.$$

$$\log M_{X_n}(t) = n \log\left(1 - \frac{\lambda}{n} + \frac{\lambda}{n}e^t\right)$$

$$= n \log\Big(1 + \underbrace{\frac{\lambda}{n}(e^t - 1)}_{\to 0 \text{ as } n \to +\infty}\Big).$$

$\lambda, t$ are fixed

FACT from 31A: $\displaystyle\lim_{x \to 0^+} \frac{\log(1+x)}{x} = 1.$ $\leadsto$ $\dfrac{\log(1+x)}{x} = \dfrac{x + O(x^2)}{x} = 1 + O(x) \to 1$.

$x$ small

$$= \underbrace{\cancel{n} \cdot \frac{\lambda}{\cancel{n}}(e^t - 1)}_{\to \lambda(e^t - 1)} \cdot \underbrace{\frac{\log\left(1 + \frac{\lambda}{n}(e^t - 1)\right)}{\frac{\lambda}{n}(e^t - 1)}}_{\to 1 \text{ as } n \to +\infty}$$

$$\longrightarrow \lambda(e^t-1) \text{ as } n\to+\infty.$$

$$\Rightarrow \quad M_{X_n}(t) \longrightarrow \underbrace{e^{\lambda(e^t-1)}}_{\text{MGF of Poisson}(\lambda)} \text{ as } n\to+\infty.$$

So by the Theorem above,

$$X_n \to X \text{ in distribution as } n\to+\infty.$$

where $X \sim \text{Poisson}(\lambda)$.

$$\lim_{n\to+\infty} \mathbb{P}(X_n\in A) = \mathbb{P}(X\in A).$$

↳ If n is big, $\mathbb{P}(X_n\in A) \approx \mathbb{P}(X\in A)$.

**Example 12:**

$$Y_n \longrightarrow Y \text{ "=" } \sum_{j=1}^{\infty} a_j X_j$$

- Let $X_j \sim \mathcal{N}(0,1)$, $j \in \mathbb{N}$, be independent and $\{a_j\}_{j=1}^{\infty}$ be real numbers.
- Suppose that

$$\lim_{n\to\infty} \sum_{j=1}^{n} a_j^2 = \sum_{j=1}^{\infty} a_j^2 < +\infty.$$

e.g. $a_j = 1/j$

$$\sum_{j=1}^{\infty} a_j = \sum_{j=1}^{\infty} \frac{1}{j^2} = \frac{\pi^2}{6}.$$

Show that $Y_n = \sum_{j=1}^{n} a_j X_j$ converges in distribution and determine the distribution of the limit.

Last time: $X_j \sim N(\mu_j, \sigma_j^2)$.

$$\Rightarrow Y_n \sim N\left(\sum_{j=1}^{n} \mu_j a_j, \sum_{j=1}^{n} \sigma_j^2 a_j^2\right).$$

$\mu_j = 0$, $\sigma_j^2 = 1$ for all $j$

$$X_j \sim N(0,1) \Rightarrow Y_n \sim N\left(0, \sum_{j=1}^{n} a_j^2\right).$$

$$\Leftrightarrow M_{Y_n}(t) = \exp\left(\frac{1}{2}\left(\sum_{j=1}^{n} a_j^2\right) t^2\right).$$

$$\to \underbrace{\exp\left(\frac{1}{2}\left(\sum_{j=1}^{\infty} a_j^2\right)t\right)}_{\text{MGF of } \mathcal{N}\left(0,\ \sum_{j=1}^{\infty} a_j^2\right)\text{ r.v.}}.$$

By the Theorem,

$$Y_n \to Y \text{ in distribution as } n \to +\infty$$

where $Y \sim \mathcal{N}\left(0, \sum_{j=1}^{\infty} a_j^2\right)$.

↪ A representation of (periodic) <u>Brownian motion</u>

→ Math finance

→ physics

↪ interface growth

Biology.

$$B(t) = \sum_{j=1}^{\infty} X_j \cdot \frac{\sin\left(\left(j-\frac{1}{2}\right)\pi t\right)}{\left(j-\frac{1}{2}\right)\pi}, \quad X_j \sim \mathcal{N}(0,1).$$

- Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables with mean $\mu$

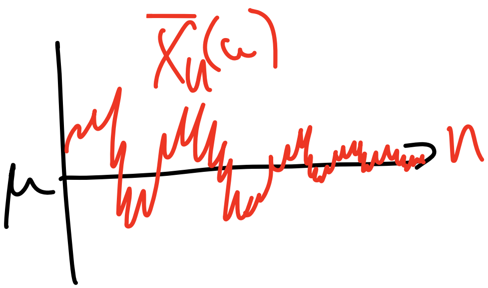
- The Weak Law of Large Numbers says that

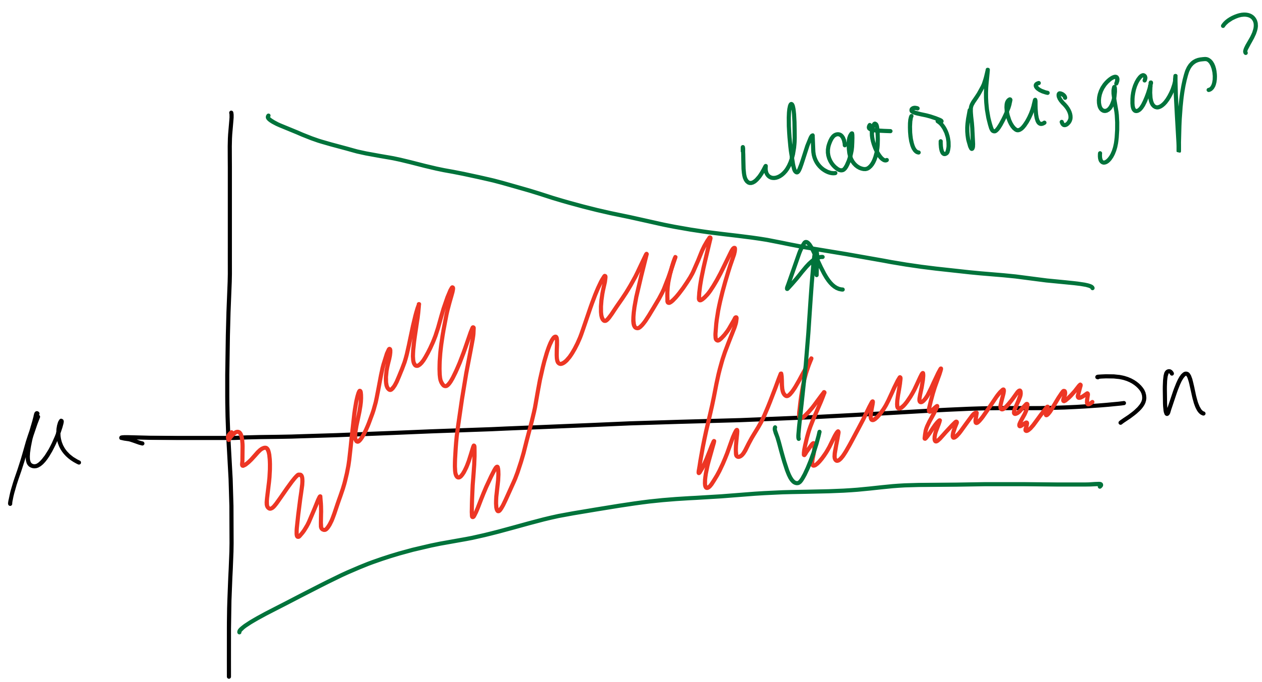
$$\overline{X}_n = \frac{1}{n}\sum_{j=1}^{n} X_j \to \mu \quad \text{in probability as } n \to \infty$$

r.v.

$$\frac{\overline{X}_n(\omega) - \mu}{\sigma/\sqrt{n}} = Z_n(\omega).$$

- What can we say about the fluctuations of $\overline{X}$ about $\mu$?



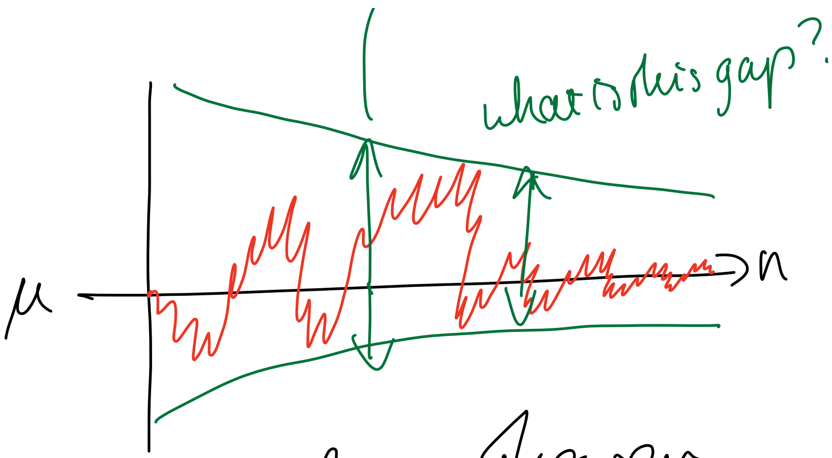
- Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables with mean $\mu$
- The Weak Law of Large Numbers says that

$$\overline{X}_n = \frac{1}{n} \sum_{j=1}^{n} X_j \to \mu \quad \text{in probability as } n \to \infty$$

- What can we say about the fluctuations of $\overline{X}_n$ about $\mu$?
- **Simple answer:** If the $X_j$ have variance $\sigma^2$, we have

$$\mathrm{var}(\overline{X}_n) = \mathbb{E}\big[(\overline{X}_n - \mu)^2\big] = \frac{\sigma^2}{n}.$$

$\approx \sigma/\sqrt{n}$.




Central limit Theorem:

**Theorem 5.20 (The Central Limit Theorem)**

Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables with finite mean $\mu$ and variance $\sigma^2$. Then,

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \to \mathcal{N}(0,1) \quad \text{in distribution as } n \to \infty$$

~~**Proof:**~~

If $n$ is large,

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \approx Z, \quad Z \sim \mathcal{N}(0,1).$$

$$\Leftrightarrow \quad \overline{X}_n \approx \mu + \frac{\sigma}{\sqrt{n}} Z$$

$$\mathbb{P}(a \le \overline{X}_n \le b) = \mathbb{P}\left( \frac{a-\mu}{\sigma/\sqrt{n}} \le \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \le \frac{b-\mu}{\sigma/\sqrt{n}} \right)$$

$$\approx \mathbb{P}\left( \frac{a-\mu}{\sigma/\sqrt{n}} \le Z \le \frac{b-\mu}{\sigma/\sqrt{n}} \right).$$

conv. in distribution

**Proof:** Idea: Show that $M_{\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}}(t) \to e^{\frac{1}{2}t^2} \leftarrow$ MGF of $N(0,1)$ r.v. as $n \to +\infty$.

Let $Y_j = \frac{X_j-\mu}{\sigma}$. Then $(Y_j)_{j=1}^{\infty}$ are i.i.d and

"change of variables" easier to see what is happening

- $E[Y_j] = \frac{1}{\sigma}(E[X_j]-\mu) = 0$
- $\mathrm{var}(Y_j) = \frac{1}{\sigma^2}\mathrm{var}(X_j-\mu) = \frac{1}{\sigma^2}\mathrm{var}(X_j) = \frac{\sigma^2}{\sigma^2} = 1.$

**Note:**
$$\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} = \frac{\frac{1}{n}\sum_{j=1}^{n}(X_j-\mu)}{\sigma/\sqrt{n}} = \frac{\sqrt{n}}{n}\sum_{j=1}^{n}\left(\frac{X_j-\mu}{\sigma/\sqrt{n}}\right)$$
$$= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}Y_j = \bar{Y}_n/{\tfrac{1}{\sqrt{n}}}.$$

**Goal:** $\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} \to N(0,1)$ is equivalent to $\bar{Y}_n/\tfrac{1}{\sqrt{n}} \to N(0,1)$

in distribution

Let $Y_j$ have MGF $M(t)$. Then

$$M_{\bar{Y}_n}(t) = [M(t/n)]^n$$

So $M_{\bar{Y}_n/(1/\sqrt{n})}(t) = E[e^{\sqrt{n}t\bar{Y}_n}] = M_{\bar{Y}_n}(\sqrt{n}t) = [M(t/\sqrt{n})]^n$.

It is convenient to take a logarithm!

$$\log M_{\bar{Y}_n/(1/\sqrt{n})}(t) = n\log[M(t/\sqrt{n})].$$

Goal: Show $n\log[M(t/\sqrt{n})] \to \log(e^{\frac{1}{2}t^2}) = \frac{1}{2}t^2$ as $n \to \infty$

Taylor's Theorem

$$M(t) = \underbrace{M(0)}_{} + \underbrace{M'(0)}_{}t + \frac{1}{2}\underbrace{M''(0)}_{}t^2 + \underbrace{R(t)}_{\text{Remainder}}$$

$= 1.$ $\quad = \mathbb{E}[Y_j] = 0$ $\quad = \mathbb{E}[Y_j^2] = 1.$

Since $1 = \mathrm{var}(Y_j) = \mathbb{E}(Y_j^2) - \mathbb{E}[Y_j]^2 = \mathbb{E}[Y_j^2]$

$|R(t)| \leq C|t|^3$ for some constant $C>0$.

So
$$M(t) = 1 + \frac{1}{2}t^2 + R(t)$$

$$\Rightarrow \log M_{\bar{Y}_n/\sqrt{n}}(t) = n \log\left[M(t/\sqrt{n})\right]$$
$$= n\log\left[1 + \frac{1}{2}(t/\sqrt{n})^2 + R(t/\sqrt{n})\right]$$
$$= n\log\Big(1 + \underbrace{\frac{t^2}{2n} + R(t/\sqrt{n})}_{X.}\Big)$$

$|R(t/\sqrt{n})| \leq \underbrace{\frac{Ct^3}{n^{3/2}}}_{\text{good compared to } n^1 \text{ outside log.}}$

**Taylor's Theorem:** $\log(1+X) = X + r(X)$

where $|r(x)| \le C_1 |x|^2$ for $x > 0$ & some constant $C_1 > 0$.

$$\Rightarrow \log M_{\bar{Y}_n/(1/\sqrt{n})}(t) = n\left(\frac{t^2}{2n} + R(t/\sqrt{n}) + r\left(\frac{t^2}{2n} + R(t/\sqrt{n})\right)\right)$$

$$= \frac{t^2}{2} + nR(t/\sqrt{n}) + nr\left(\frac{t^2}{2n} + R(t/\sqrt{n})\right).$$

For fixed $t$

- $|nR(t/\sqrt{n})| \le n \cdot \frac{Ct^3}{n^{3/2}} = \frac{Ct^3}{n^{1/2}} \to 0$ as $n \to +\infty$
- $\left|nr\left(\frac{t^2}{2n} + R(t/\sqrt{n})\right)\right| \le nC_1\left(\frac{t^2}{2n} + R(t/\sqrt{n})\right)^2$

$$= \underbrace{\frac{C_1}{n}\Big(\underbrace{\frac{t^2}{2} + \underbrace{nR(t/\sqrt{n})}_{\to 0}}_{\to 0}\Big)^2}_{\to 0 \text{ as } n \to \infty}$$

Therefore

$$\lim_{n \to +\infty} \log M_{\bar{Y}_n/(1/\sqrt{n})}(t) = \frac{1}{2}t^2$$

i.e. $\lim_{n \to +\infty} M_{\bar{Y}_n/(1/\sqrt{n})}(t) = e^{\frac{1}{2}t^2}$.

So we can apply Theorem 5.14 above to conclude $\bar{Y}_n/(1/\sqrt{n}) \to Z$ in distribution where $Z \sim N(0,1)$, as $n \to +\infty$.

- The CLT says that for $X_1, X_2, \ldots$ i.i.d., we have

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \to \mathcal{N}(0,1) \quad \text{in distribution as } n \to \infty$$

$\mu + \frac{\sigma}{\sqrt{n}} Z \sim \mathcal{N}(\mu, \sigma^2/n).$

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

- Recall that if $Z \sim \mathcal{N}(0,1)$, then $\sigma Z + \mu \sim \mathcal{N}(\mu, \sigma^2)$

$\overline{X}_n \approx \underbrace{\mu + \frac{\sigma}{\sqrt{n}} Z}.$

- So the CLT roughly says

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

when $n$ is sufficiently large. Note $\mathbb{E}[\overline{X}_n] = \mu$ and $\operatorname{var}(\overline{X}_n) = \frac{\sigma^2}{n}$.

- Also

$$S_n = \sum_{j=1}^{n} X_j \approx \mathcal{N}(n\mu, n\sigma^2)$$

when $n$ is sufficiently large.

$\mathbb{P}(a \le S_n \le b) \approx \mathbb{P}\left(\frac{a - n\mu}{\sqrt{n\sigma^2}} \le Z \le \frac{b - n\mu}{\sqrt{n\sigma^2}}\right)$

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

- $\Longrightarrow$ can approximate probabilities of sums of i.i.d. r.v.s

Two case: (i) continuous, (ii) discrete

**Example 13:** $\mathbb{E}[X_j] = 1/2$, $\text{var}(X_j) = 1/12$. $\overline{X}_n \approx \mu + \frac{\sigma}{\sqrt{n}} Z$.

- Let $X_1, \ldots, X_{12}$ be independent, Uniform((0, 1)) random variables
- Use the Central Limit Theorem to approximate $\mathbb{P}(\overline{X}_n \leq \frac{1}{4})$.

Mean & variance of $\overline{X}_n$:

$$\Rightarrow \mathbb{E}[\overline{X}_n] = 1/2 \,, \quad \text{var}(\overline{X}_n) \overset{n=12}{=} \frac{\text{var}(X_j)}{n} = \frac{1}{12^2} \,, \Rightarrow$$

$$\Rightarrow \sigma/\sqrt{n} = \frac{1}{\sqrt{12}} \frac{1}{\sqrt{12}} = \frac{1}{12}.$$

$$\mathbb{P}(\overline{X}_n \leq 1/4) \overset{\text{by CLT}}{\approx} \mathbb{P}\left(\frac{1}{2} + \frac{1}{12} Z \leq 1/4\right)$$

$$= \mathbb{P}\left(\frac{1}{12} Z \leq -1/4\right)$$

$$= \mathbb{P}(Z \leq -3)$$

$$= \Phi(-3) = 1 - \Phi(3) = 1 - 0.9987$$

$$\approx 0.0013.$$

**Example 14:**

- Let $X \sim \text{Gamma}(25, \frac{1}{2})$. $\longrightarrow \propto \underline{x^{25-1} e^{-}}$
- Using the Central Limit Theorem, approximate $\mathbb{P}(X \leq 13)$.

Exercise: $X = \underbrace{\sum_{j=1}^{25} X_j}_{= S_{25}}$, $X_j \sim \text{Exp}(1/2)$. independent.

$\mathbb{E}[S_n] = 25 \times 1/2 = \frac{25}{2}$., $\text{var}(S_n) \overset{\text{indep}}{=} n\sigma^2 = 25 \times (1/2)^2 = \frac{25}{4}$.

$\hookrightarrow \mathbb{P}(X \leq 13) = \mathbb{P}(S_{25} \leq 13)$.

$= \mathbb{P}\left( \frac{S_{25} - \frac{25}{2}}{\sqrt{25/4}} \leq \frac{13 - \frac{25}{2}}{\sqrt{25/4}} \right)$.

$\approx Z \overset{\text{CLT}}{\Leftarrow}$ where $Z \sim N(0,1)$.

$\frac{13 - \frac{25}{2}}{\sqrt{\frac{25}{4}}} = \frac{1/2}{5/2} = 1/5$

$\approx \mathbb{P}(Z \leq 1/5) = \Phi(0.2)$.

CLT: $S_n \sim N(n\mu, n\sigma^2)$

$$S_n \longrightarrow n\mu + \underbrace{\sqrt{n\sigma^2}}_{\sqrt{\mathrm{Var}(S_n)}} Z.$$

## Discrete case (using CLT)

Matters whether we have $\leq$ or $<$

$$\mathbb{P}(X \leq \ell) = \mathbb{P}(X \leq \ell + 1/2).$$

discrete & $p_X(\ell) > 0$.

## Proposition 5.21: (DeMoivre-Lapace Correction)

Let $X \sim \text{Binomial}(n, p)$. Then, we have

$\mathbb{E}(X)$

$$\mathbb{P}(X \leq \ell) \approx \Phi\left(\frac{\ell + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right),$$

$\sqrt{Var(X)}$

$$\mathbb{P}(X \geq k) \approx 1 - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right),$$

$$\mathbb{P}(k \leq X \leq \ell) \approx \Phi\left(\frac{\ell + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right).$$

**Proof:**

$X \sim Bin(n,p)$, we know that

$$X = \underbrace{\sum_{j=1}^{n} X_j}_{S_n}, \quad X_j \sim Bernoulli(p) \text{ independent}$$

$$\mathbb{E}[S_n] = np, \quad Var(S_n) = np(1-p)$$

$$\mathbb{P}(X \le \ell) = \mathbb{P}(X \le \ell + 1/2)$$

$$= \mathbb{P}\Big( \underbrace{\frac{X - np}{\sqrt{np(1-p)}}}_{\text{CLT} \approx Z \sim N(0,1)} \le \frac{\ell + 1/2 - np}{\sqrt{np(1-p)}} \Big).$$

$$\approx \mathbb{P}\Big( Z \le \frac{\ell + 1/2 - np}{\sqrt{np(1-p)}} \Big)$$

$$= \Phi\Big( \frac{\ell + 1/2 - np}{\sqrt{np(1-p)}} \Big).$$

## Example 15:

- Let $X \sim \text{Binomial}(16, \frac{1}{2})$
- Use the DeMoivre-Laplace correction to the Central Limit Theorem to approximate $\mathbb{P}(X \leq 10)$.

$$\mathbb{E}(X) = 8$$
$$\mathrm{Var}(X) = 16/4 = 4$$

$$\mathbb{P}(X \leq 10) = \mathbb{P}(X \leq 10 + 1/2)$$
$$= \mathbb{P}\left(X - 8 \leq \frac{21}{2} - 8\right)$$
$$= \mathbb{P}(X - 8 \leq 5/2)$$
$$\frac{21 - 16}{2} = \frac{5}{2}$$
$$= \mathbb{P}\left(\frac{X-8}{2} \leq 5/4\right)$$
$$\overset{\text{by CLT}}{\approx} \mathbb{P}(Z \leq 5/4) = \Phi(5/4) \approx 0.8944$$

## Example 16:

- Let $X \sim \text{Binomial}(16, \frac{1}{2})$
- Use the DeMoivre-Laplace correction to the Central Limit Theorem to approximate $\mathbb{P}(X = 10)$.

$$\mathbb{P}(X=10) = \mathbb{P}\left(10-\tfrac{1}{2} \le X \le 10+\tfrac{1}{2}\right)$$

$$= \mathbb{P}\left(3/4 \le \frac{X-8}{2} \le 5/4\right)$$

by CLT $\approx \Phi(5/4) - \Phi(3/4)$.