M339D: March 31st, 2023.

The Normal Distribution.

We completely specify any normal distribution by its mean and its variance (or its standard deviation).

We write $\boxed{X \sim \text{Normal}(\text{mean} = \mu_X, \text{variance} = \sigma_X^2)}$

X can be written as a linear transform of a standard normal Z:

$$\boxed{X = \mu_X + \sigma_X \cdot Z}$$

We can check:

- $\mathbb{E}[X] = \mathbb{E}[\mu_X + \sigma_X \cdot Z] \overset{\text{linearity}}{=} \mu_X + \sigma_X \underbrace{\mathbb{E}[Z]}_{=0} = \mu_X$

- $\text{Var}[X] = \text{Var}[\mu_X + \sigma_X \cdot Z]$

deterministic (added, so does not affect the variance)

$$= \text{Var}[\sigma_X Z] = \sigma_X^2 \cdot \underbrace{\text{Var}[Z]}_{=1} = \sigma_X^2$$

The Normal Approximation to the Binomial

(deMoivre-Laplace)

Consider a sequence of binomial random variables:

$$Y_n \sim \text{Binomial}(n = \#\text{ of trials},\ p = \text{probab. of success})$$

Then, $\mathbb{E}[Y_n] = n \cdot p$

$\mathrm{Var}[Y_n] = n \cdot p \cdot (1-p) \Rightarrow SD[Y_n] = \sqrt{np(1-p)}$

$$\boxed{\frac{Y_n - np}{\sqrt{np(1-p)}} \xRightarrow{\mathcal{D}} N(0,1)}$$

Usage:

- Look @ "large" n (rule of thumb: $np \geq 10$ and $n(1-p) \geq 10$).
- $\mathbb{P}[a < Y_n \leq b] =$

$$= \mathbb{P}\left[\frac{a-np}{\sqrt{np(1-p)}} < \underbrace{\frac{Y_n - np}{\sqrt{np(1-p)}}}_{\approx N(0,1) \sim Z} \leq \frac{b-np}{\sqrt{np(1-p)}}\right]$$

$$= \mathbb{P}\left[\frac{a-np}{\sqrt{np(1-p)}} < Z \leq \frac{b-np}{\sqrt{np(1-p)}}\right]$$

$$= N\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - N\left(\frac{a-np}{\sqrt{np(1-p)}}\right)$$

N... cumulative dist'n f'tion of $N(0,1)$, i.e., $N(x) = \mathbb{P}[Z \leq x]$

- In statistics: we usually use

$$\boxed{Y_n \approx \text{Normal}(\text{mean} = np,\ sd = \sqrt{np(1-p)})}$$

- In M362K: continuity correction.

UNIVERSITY OF TEXAS AT AUSTIN

Problem Set 11

The normal approximation to the binomial.

Problem 11.1. According to the Pew research center, 64% of Americans say that social media have a mostly negative effect on things (see https://pewrsr.ch/3dsV7uR). You take a sample of 1000 randomly chosen Americans. What is the approximate probability that at least 650 of them say that social media have a mostly negative effect on things?

→: Y... a r.v. denoting the # of surveyed people who claim that social media are negative

$Y \sim Binomial(n = 1000, p = 0.64)$

$n \cdot p = 640 \geq 10$ and $n(1-p) = 360 \geq 0$ ✓

$\mu_Y = \mathbb{E}[Y] = n \cdot p = 640$

$\sigma_Y = \sqrt{n \cdot p \cdot (1-p)} = \sqrt{640(0.36)} = 15.18$

$\mathbb{P}[Y \geq 650] = \mathbb{P}\left[\frac{Y-640}{15.18} \geq \frac{650-640}{15.18}\right]$

$\approxeq \mathbb{P}[Z \geq 0.66] = 1 - N(0.66)$

$= 1 - 0.7454 = 0.2546$ $\square$

Problem 11.2. According to a Gallup survey, only 22% of American young adults rate their mental health as *excellent*:

https://www.gallup.com/education/328961/why-higher-education-lead-wellbeing-revolution.aspx.

You sample 6000 randomly chosen American young adults. What is the approximate probability that at most 1400 of them rate their mental health as *excellent*?

→ : Y... # of sampled people who said excellent

$Y \sim Binomial(n = 6000, p = 0.22)$

Check: $n \cdot p = 6000(0.22) = 1320 \geq 10$

$n(1-p) = 6000(0.78) = 4680 \geq 10$

$\mu_Y = n \cdot p = 1320$

$\sigma_Y = \sqrt{np(1-p)} = \sqrt{1320(0.78)} = 32.09$

$$\mathbb{P}[Y \leq 1400] = \mathbb{P}\left[\frac{Y-1320}{32.09} \leq \frac{1400-1320}{32.09}\right]$$

$$\approx \mathbb{P}[Z \leq 2.49] = N(2.49) = 0.9936$$

$\square$

Problem 11.3. You toss a fair coin 10,000 times. What is the approximate probability that the number of *Heads* exceeds the number of *Tails* by between 200 and 500 (inclusive)?

→ : $Y \ldots$ # of Heads

$Y \sim Binomial(n = 10000, p = 0.5)$

$$\mathbb{P}[200 \leq \underbrace{Y}_{\#Hs} - (\underbrace{10000 - Y}_{\#Ts}) \leq 500] =$$

$$= \mathbb{P}[200 \leq 2Y - 10000 \leq 500]$$

$$= \mathbb{P}[10\,200 \leq 2Y \leq 10\,500]$$

$$= \mathbb{P}[5100 \leq Y \leq 5250] = ?$$

$np = n(1-p) = 5000 \geq 10$

$\mu_Y = 5000$

$\sigma_Y = \sqrt{5000(0.5)} = 50$

$$\mathbb{P}[5100 \leq Y \leq 5250] \approx \mathbb{P}\left[\frac{5100-5000}{50} \leq Z \leq \frac{5250-5000}{50}\right]$$

$$= N(5) - N(2) = 1 - 0.9772 = 0.0228 \quad \square$$