M339D: November 6th, 2024.

The Normal Approximation to the Binomial (deMoivre-Laplace)

Consider a sequence of binomial random variables

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

Then,

- $\mathbb{E}[Y_n] = \underline{np}$
- $\text{Var}[Y_n] = \underline{np(1-p)} \Rightarrow SD[Y_n] = \sqrt{np(1-p)}$

$$\boxed{\boxed{\frac{Y_n - np}{\sqrt{np(1-p)}}} \xrightarrow[n\to\infty]{\mathcal{D}} N(0,1)} \quad \bigstar$$

Usage:

- Look @ "large" n (rule of thumb: $np \geq 10$ and $n(1-p) \geq 10$)
- $$\mathbb{P}[a < Y_n \leq b] = $$
$$= \mathbb{P}\left[\frac{a-np}{\sqrt{np(1-p)}} < \underbrace{\boxed{\frac{Y_n - np}{\sqrt{np(1-p)}}}}_{\approx N(0,1) \sim Z} \leq \frac{b-np}{\sqrt{np(1-p)}}\right]$$
$$\approx \mathbb{P}\left[\frac{a-np}{\sqrt{np(1-p)}} < Z \leq \frac{b-np}{\sqrt{np(1-p)}}\right]$$
$$= N\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - N\left(\frac{a-np}{\sqrt{np(1-p)}}\right)$$

$\boxed{N \dots \text{cumulative dist'n f'tion of } Z \sim N(0,1)}$

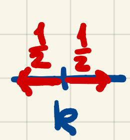
- In Statistics: We usually use

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

- $\boxed{\text{In M362K:}}$ continuity correction

$$\mathbb{P}[Y_n = k] = \mathbb{P}[k - \tfrac{1}{2} < Y_n \leq k + \tfrac{1}{2}]$$

w/ $k = 0, 1, \dots, n$



University of Texas at Austin

Problem Set 12

The normal approximation to the binomial.

Problem 12.1. 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 is bad

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

$\mathbb{P}[Y \geq 650] = ?$

$n \cdot p = 640 > 10, \quad n(1-p) = 360 > 10$

$\mu_Y = \mathbb{E}[Y] = np = 640$

$\sigma_Y = \sqrt{np(1-p)} = \sqrt{1000 \cdot 0.64 \cdot 0.36} = 15.18$

$$\mathbb{P}[Y \geq 650] = \mathbb{P}\Big[\underbrace{\frac{Y-640}{15.18}}_{\approx N(0,1) \sim Z} \geq \frac{650-640}{15.18}\Big] =$$

$$= \mathbb{P}[Z \geq 0.66] = 1 - 0.7454 = 0.2546$$

$\square$

Just for Laughs:

$\mathbb{P}[Y \geq 650] = 1 - \mathbb{P}[Y \leq 649]$

$= 1 - \text{pbinom}(649, \text{size} = 1000, \text{prob} = 0.64)$

$= 0.2663257$

Problem 12.2. 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: $np = 6000(0.22) = 1320 > 10$

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

$\mu_Y = np = 1320$

$\sigma_Y = \sqrt{np(1-p)} = \sqrt{1320(0.78)} = 32.08738$

$$\mathbb{P}[Y \leq 1400] = \mathbb{P}\Big[\underbrace{\frac{Y-1320}{\sigma_Y}}_{\approx N(0,1) \sim Z} \leq \frac{1400-1320}{\sigma_Y}\Big]$$

$$\mathbb{P}[Z \leq 2.49] = 0.9936$$

$\square$

For fun:

pbinom(1400, 6000, 0.22) = 0.9963818

Problem 12.3. 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)?

$\rightarrow$: $Y \ldots$ # of Hs $\qquad Y \sim Binomial(n = 10000, p = 0.5)$

$n - Y \ldots$ # of tails

$$\mathbb{P}[200 \leq Y - (n-Y) \leq 500] = ?$$

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

$$= \mathbb{P}[10200 \leq 2Y \leq 10500]$$

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

$np = n(1-p) = 5000 > 10$

$\mu_Y = 5000$

$\sigma_Y = \sqrt{5000 \cdot 0.5} = \sqrt{2500} = 50$

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

$$\approx \mathbb{P}[2 \leq Z \leq 5] = N(5) - N(2) = 1 - 0.9772$$

$$= 0.0228$$

$\square$