M339D: October 27th, 2023.

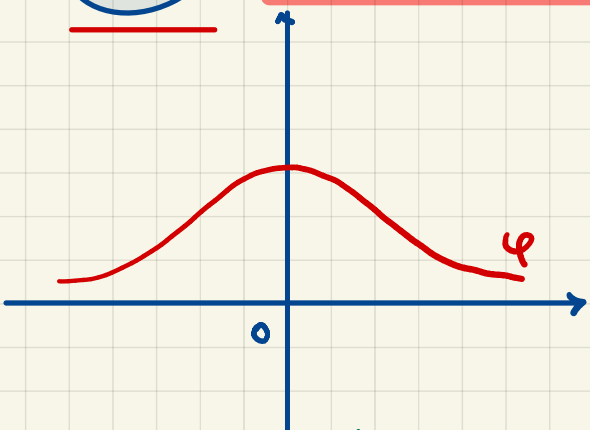
Standard Normal Distribution.

We say that a random variable Z has the

standard normal distribution

if its pdf has the following form:

$$f_Z(z) = \varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad \text{for all } z \in \mathbb{R}$$



- symmetric about the vertical axis, i.e., $\varphi(z) = \varphi(-z)$, i.e., $\boxed{\text{even}}$
- mean of $Z = 0$
- median of $Z = 0$

The cumulative distribution function of the standard normal is:

$$N(z) = \Phi(z) = \mathbb{P}[Z \leq z] = \int_{-\infty}^{z} f_Z(u)\,du$$

$$= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\,du$$

$\boxed{\text{No analytic form!}}$

There are standard normal tables!

We can use the built-in commands in R.

We write $\boxed{Z \sim N(0,1)}$

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:

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

We can check:

linearity of expectation

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

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.

(de Moivre · Laplace)

Consider a sequence of binomial random variable:

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

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

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

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

Usage: • Look @ "large" n (rule of thumb: $np \geq 10$ and $n(1-p) \geq 10$).

• $\mathbb{P}[a < Y_n \leq b] =$

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

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

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

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

• In statistics: We usually use $\boxed{Y_n \approx \text{Normal}(\text{mean} = n \cdot p,\ sd = \sqrt{np(1-p)})}$

• 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 $\boxed{1000}$ 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 negative

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

$n \cdot p = 1000(0.64) = 640 \geq 10$ and $n(1-p) = 360 \geq 10$ ✓

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

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

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

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

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

$n(1-p) = 4680 \geq 10$ ✓

$\mu_Y = n \cdot p = 1320$

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

$$\mathbb{P}[Y \leq 1400] = \mathbb{P}\left[\underbrace{\boxed{\frac{Y-1320}{32.09}}}_{\approx N(0,1)} \leq \frac{1400-1320}{32.09}\right]$$

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

□

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... # of Heads

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

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

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

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

$\vdots$