M339D: April 1st, 2024.

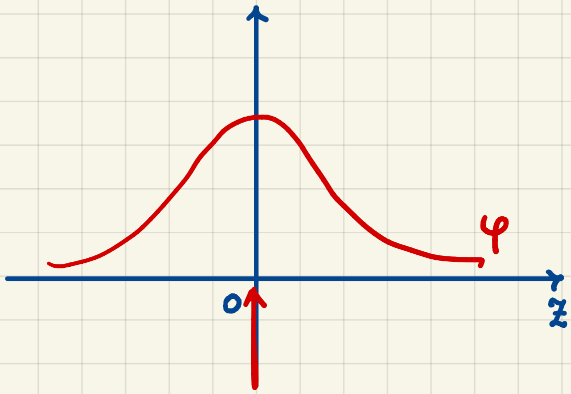
Standard Normal Distribution.

We say that a random variable Z has the

standard normal distribution

if its probability density function (pdf) has the following form:

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



- symmetric about the vertical axis, i.e., $\varphi(z) = \varphi(-z)$, i.e., 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 providing its mean μ_X and its variance (or its standard deviation σ_X).

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:

linearity of expectation

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

a deterministic shift which doesn't 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.

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

Then,

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

LLN

$$\frac{Y_n}{n} \xrightarrow[n\to\infty]{} p$$

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

Usage:

- Look @ "large" n (rule of thumb: $n \cdot p \geq 10$ and $n(1-p) \geq 10$)

- $\mathbb{P}[a < Y_n \leq b] =$

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

$\approx N(0,1) \sim Z$

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

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

N... cumulative distribution function of $Z \sim N(0,1)$

- In statistics: We usually use

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

- In M362K: continuity correction

$\overset{\leftarrow}{\tfrac{1}{2}} \ \overset{\rightarrow}{\tfrac{1}{2}}$

$\mathbb{P}[Y_n = k] =$

w/ $k \in \{0, 1, \dots, n\}$