Normal Approximation to Binomial

De-Moivre's [Central limit Theorem]

$X \sim \text{Binomial}(n, p)$

$\underbrace{E[X] = np}_{\text{Centered}}$ $\quad$ $\underbrace{Var[X] = np(1-p)}$

Spread $\doteq$ $SD[X] = \sqrt{Var[X]}$

Theorem :- $\alpha, \beta \in \mathbb{R}$

$$P\left(\alpha < \frac{X - np}{\sqrt{np(1-p)}} < \beta\right) \xrightarrow[n \to \infty]{} \underbrace{\int_{\alpha}^{\beta} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}\, dy}_{I}$$

- Useful tool :- to approximate Binomial Prob.
- Computation :- Numerical tables for I

Question :-

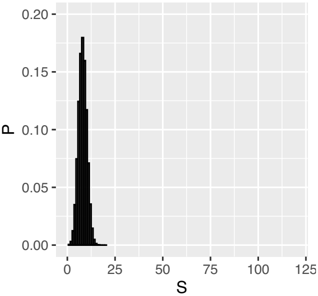
- When do we use the above approximation?
- Are there other approximations?
- How to use approximations effectively?

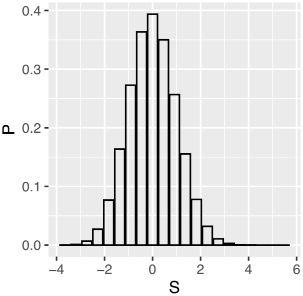
Binomial(20, 0.4)

$P = \text{Binomial}(k, 20, 0.4)$

Scaled $\sqrt{20(0.4)(0.6)}$

Centered $20(0.4)$



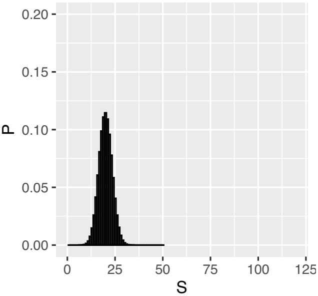


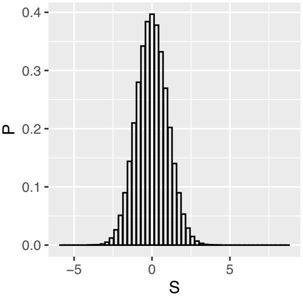
Binomial(50, 0.4)

P= Binomial(k, 20, 0.4)

Scaled $\sqrt{50(0.4)(0.6)}$

Centered $50(0.4)$



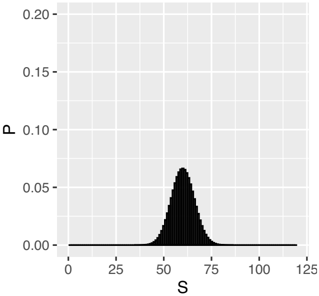


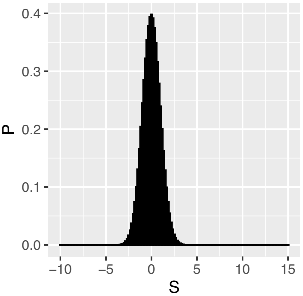
Binomial(150, 0.4)

$P = \text{Binomial}(k, 150, 0.4)$

Scaled $\sqrt{150(0.4)(0.6)}$

Centered $150(0.4)$

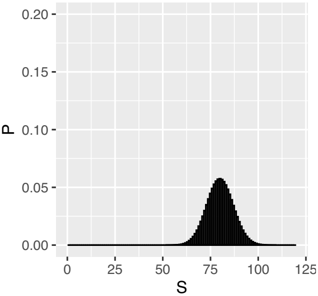


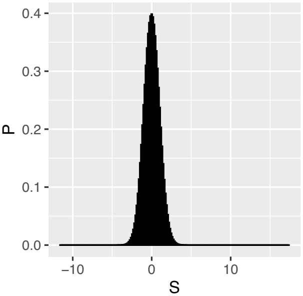


Binomial(200, 0.4)

P = Binomial(k, 200, 0.4)

Scaled $\sqrt{200(0.4)(0.6)}$
Centered $200(0.4)$





Fact:- • For Binomial (n, p) Approximation by Integral $\int_{*}^{*} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \, dy$

is good if $np(1-p)$ is "large".

It will make :- $\frac{X_n - np}{\sqrt{np(1-p)}} \simeq$ "reasonable"

• It does not work if $n >>$ but $p <<$ $\Rightarrow$ np is not large

Different Approximation in such cases :-

Poisson Approximation

n – large $\quad p = \frac{\lambda}{n}$ for some $\lambda > 0$ $\quad [np = \lambda]$

$k \geq 1$ let $A_k = \{k$ success in Binomial(n, p) Experiment$\}$

$$\mathbb{P}(A_k) = {}^{n}C_{k} \, p^k (1-p)^{n-k}$$

$$= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!} \; 1 \left(1-\frac{1}{n}\right) \cdots \left(1-\frac{(k-1)}{n}\right) \cdot \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k}$$

$$\xrightarrow[n \to \infty]{} \quad \frac{\lambda^k}{k!} \quad 1 \cdot 1 \quad \cdots \quad 1 \quad e^{-\lambda} \quad 1$$

$$\boxed{\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}}$$

We have shown

$$\lim_{n\to\infty} \mathbb{P}(A_k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

where $A_k = \{ k$ successes in Binomial$(n, \frac{\lambda}{n})$ experiment. $\}$

Result (Not to be Proved) $\exists\, C > 0$

$$\sum_{k=0}^{\infty} \left| \mathbb{P}(A_k) - \underbrace{\frac{\lambda^k}{k!} e^{-\lambda}}_{\text{valid distribution on } \{0,1,2,..\}} \right| \leq C \frac{\lambda}{n}$$

Poisson Distribution :- $X \sim \text{Poisson}(\lambda)$ if

$\text{Range}(X) = \{0, 1, 2, \ldots\}$

and $\mathbb{P}(X = k) = \dfrac{\lambda^k e^{-\lambda}}{k!}$

$E[X] = \text{"} n \cdot \frac{\lambda}{n} \text{"} = \lambda$

$\text{Var}[X] = \text{"} n \frac{\lambda}{n}(1 - \frac{\lambda}{n}) \text{"} = \text{"} \lambda (1 - \frac{\lambda}{n}) \text{"} \longrightarrow \lambda$

Properties: $X \sim \text{Binomial}(n, p)$
$Y \sim \text{Binomial}(m, p)$

Then $X + Y \sim$?

$X = \#$ of Head in n tosses $\qquad Y = \#$ of Head in m tosses

$\underbrace{\_\ \_\ \_\ \_\ \ldots\ \_}_{n \text{ tosses}} \quad \underbrace{\_\ \_\ \_\ \_\ \_}_{m \text{ tosses}}$

$X + Y = \#$ of Heads in $m+n$ $\sim \text{Binomial}(m+n, p)$

$\underbrace{\_\ \_\ \_\ \_\ \_\ \_\ \_\ \_\ \_\ \_}_{m+n}$

Limiting procedure $\qquad p = \frac{\lambda}{n}$

$X \sim \text{Poisson}(\mu)$ and $Y \sim \text{Poisson}(\lambda)$

$\Longrightarrow X + Y \sim \text{Poisson}(\mu + \lambda)$

Usage of Poisson (λ) :- Counts of rare events

E.g: # of atoms undergoing radio decay during a short period of time.

Think of counts as number of successes in a large numbers of trials (independent) with the chance of success on any particular trial being very small.

Sampling WITH (out) Replacement

Town: 5000 people
- 1000 people under age of 18

Select: 4 people
- How many of the 4 are likely to be under age of 18? $= X$

WITH Replacement

- success - choose a person under 18

$X \sim \text{Binomial}\left(4, \frac{1000}{5000}\right)$ Experiment

$\mathbb{P}(X=2) = {}^4C_2 \left(\frac{1}{5}\right)^2 \left(\frac{4}{5}\right)^2 \simeq 0.1536$

WITHOUT replacement

$\mathbb{P}(X=2) = ?$

\# of samples of size 4 from 5000 $= {}^{5000}C_4$

"$X=2$" .. 2 in sample are from 1000
2 in . . . .. 4000 $= ({}^{1000}C_2) \cdot ({}^{4000}C_2)$

$$P(X=2) = \frac{({}^{1000}C_2) \cdot ({}^{4000}C_2)}{{}^{5000}C_4} \simeq 0.153592$$

Reflection: large population – with & without replacement are "close" by.

Hyper Geometric (N, m, r)

Choose sample WITHOUT Replacement size r
m – Population out of N have characteristic A (say)

$$\mathbb{P}(\text{choosing } k \text{ people in } r \text{ from } m\text{–Population}) = \frac{{}^{m}C_k \cdot {}^{N-m}C_{r-k}}{{}^{N}C_r}$$

$\max\{0, m-(N-r)\} \leq k \leq \min\{m, r\}$