

ST2334 Summary Notes

AY23/24 Sem 1, github.com/gerteck

1. Basic Probability Concepts

- **Sample Space:** S All possible outcomes of stat. expt.
- **Null Event:** Event that contains no element, empty set, \emptyset
- **Axioms of Probability:**
For any event X , $0 \leq P(X) \leq 1$. $P(S) = 1$.
If $A \cap B = \emptyset$ (Mut Excl), $P(A \cup B) = P(A) + P(B)$.
- Finite sample space with equally likely outcomes: $P(A) = (\frac{\# \text{sample points } A}{\# \text{total sample points } S})$. (e.g. birthday problem)

Event Operation & Relationships

- **Event Operations:** Union, Intersection, Complement.
- **Event Relationships:** Contained: $A \subset B$
Equivalence: $A \subset B$ with $A \supset B \rightarrow A = B$
Mutually Exclusive: $A \cap B = \emptyset$.
- **De Morgan's Law:** $(A \cup B)' = A' \cap B'$ and $(A \cap B)' = A' \cup B'$

Counting Methods

- Multiplication Principle: (Sequential Events)
- Addition Principle: (Pairwise Disjoin sets)
- **Permutation:** ${}_nP_r = \frac{n!}{(n-r)!}$
- **Combination:** $\binom{n}{r} = \frac{n!}{(n-r)!r!}$

Conditional Probability

- Understand conditional as reduced sample space.

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A|B)P(B)}{P(A)}$$

Independence

$$A \perp B \leftrightarrow P(A \cap B) = P(A)P(B)$$
$$A \perp B \leftrightarrow P(A|B) = P(A)$$

Law of Total Probability

- **Partition:** If A_1, \dots, A_n mutually exclusive, $\bigcup_{i=1}^n A_i = S$, then A_1, \dots, A_n are partitions.
- If A_1, \dots, A_n are partitions of S , then for any event B :

$$P(B) = \sum_{i=1}^n P(B \cap A_i) = \sum_{i=1}^n P(B|A_i)P(A_i)$$

Bayes' Theorem

Let A_1, \dots, A_n be partitions of S . For any event B :

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^n P(B|A_i)P(A_i)}$$

For when $n = 2$, $\{A, A'\}$ becomes a partition of S .

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A')P(B|A')}$$

2. Random Variables

A function X , which assigns a real number to every $s \in S$ is called a random variable.

- **Range space:** $R_x = \{x|x = X(s), s \in S\}$
- Likewise, the set $X \in A$, for A being a subset of R , is also a subset of $S : s \in S : X(s) \in A$.

Probability Distribution

Two main types of RV used in practice: discrete and continuous.

- Probability assigned to each possible X
- Given RV X with range of R_x :
Discrete: Numbers in R_x are finite or countable
Continuous: R_x is interval

(Discrete) Probability Mass Function $f(x)$:

$$f(x) \begin{cases} P(X = x), & \text{for } x \in R_x \\ 0, & \text{for } x \notin R_x \end{cases}$$

1. $f(x_i) = P(X = x_i) \geq 0$ for $x_i \in R_x$
2. $f(x_i) = 0$ for $x_i \notin R_x$
3. $\sum_{i=1}^{\infty} f(x_i) = 1$ (PSum = 1)
4. $\forall B \subseteq \mathbb{R}, P(X \in B) = \sum_{x_i \in B \cap R_x} f(x_i)$

(Continuous) Probability Density Function $f(x)$:

- Given R_x is interval. Quantifies probability that X is in some range.
- $p.f.$ must satisfy:
 1. $f(x) \geq 0$, $f(x) = 0$ for $x \notin R_x$
 2. No need $f(x) \leq 1$ (Concerned with area)
 3. $\int_{R_x} f(x)dx = 1$ (Integration over $R_x = 1$)
 4. $\forall a, b$ s.t. $a \leq b$, $P(a \leq X \leq b) = \int_a^b f(x)dx$
- **Note:** $P(X = x_0) = \int_{x_0}^{x_0} f(x)dx = 0$
- Hence, to check if a function is a pdf,
 1. $f(x) \geq 0$ for $x \in R_x$, $f(x) = 0$ for $x \notin R_x$
 2. $\int_{R_x} f(x)dx = 1$.

Cumulative Distribution Function

Describes distribution of a RV X : cumulative distribution function (cdf), applicable for discrete or continuous RV.

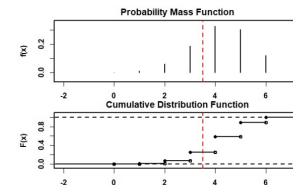
$$F(x) = P(X \leq x)$$

$F(x)$ is non-decreasing and $0 \leq F(x) \leq 1$

- Probability fn & cumulative distribution fn have one-to-one correspondence. For any probability fn given, the cdf is uniquely determined, vice versa.

CDF Discrete RV: Step Function $F(x)$

$$F(x) = \sum_{t \in R_x; t \leq x} f(t)$$

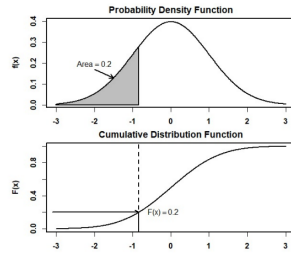


- $P(a \leq X \leq b) = P(X \leq b) - P(X < a)$
- $P(a \leq X \leq b) = F(b) - F(a-)$
- $P(a \leq X \leq b) = F(b) - \lim_{x \rightarrow a-} F(x)$
- $0 \leq f(x) \leq 1$
- c.d.f has to be **right continuous** (• —)

CDF Continuous RV: $F(x)$

$$F(x) = \int_{-\infty}^x f(t)dt$$

$$\text{impt} : f(x) = \frac{d(F(x))}{dx}$$



- $P(a \leq X \leq b) = P(a < X < b) = F(b) - F(a)$
- $0 \leq f(x)$.
e.g. $f(x) = 3x^2$ is a valid p.f. since $\int_{R_x} f(x)dx = 1$

Expectation μ & Variance σ

Expectation of Random Variable: μ

• Mean of discrete RV:

$$\mu = E(X) = \sum_{x \in R_x} x_i f(x_i)$$

- **E.g.:** X discrete RV with p.m.f. $f(x)$ and range R_X
 $\mu = E(g(x)) = \sum_{x \in R_x} g(x)f(x)$

• Mean of continuous RV:

$$\mu = E(X) = \int_{x \in R_x} xf(x)dx$$

- **E.g.:** X continuous RV with p.d.f. $f(x)$ and range R_X
 $\mu = E(g(x)) = \int_{x \in R_x} g(x)f(x)dx$

• Properties of Expectation:

- $E(aX + b) = aE(X) + b$
- Linearity of expectation: $E(X + Y) = E(X) + E(Y)$

Variance of Random Variable: σ

$$\sigma_X^2 = V(X) = E[(X - \mu_X)^2]$$

• Variance of discrete RV:

$$V(X) = \sum_{x \in R_x} (x - \mu_X)^2 f(x)$$

• Variance of continuous RV:

$$V(X) = \int_{x \in R_x} (x - \mu_X)^2 f(x)dx$$

- $V(X) \geq 0$ and $V(X) = 0$ when X is a constant
- $V(aX + b) = a^2V(X)$
- **alt. form:** $V(X) = E(X^2) - (E(X))^2$
- **Standard Deviation:** $\sigma_X = \sqrt{V(X)}$

3. Joint Distributions

- Consider more than 1 RV simultaneously,
- Given sample space S . Let X and Y be functions mapping $s \in S \rightarrow \mathbb{R}$: (X, Y) is 2D random vector.

Range spc: $R_{X,Y} = \{(x, y) | x = X(s), y = Y(s), s \in S\}$

• Discrete 2D RV:

of possible values of $(X(s), Y(s))$ finite / countable

• Continuous 2D RV:

of possible values of $(X(s), Y(s))$ assume any value in some region of the Euclidean space \mathbb{R}^2

- If both X and Y are discrete/continuous, then (X, Y) is discrete/continuous respectively.

Joint Probability Function

• Joint Probability (mass) function, 2D discrete RV:

$$f_{X,Y}(x, y) = P(X = x, Y = y)$$

- $f_{X,Y}(x, y) \geq 0$ for any $(x, y) \in R_{X,Y}$
- $f_{X,Y}(x, y) = 0$ for any $(x, y) \notin R_{X,Y}$
- $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P(X = x_i, Y = y_j) = 1$
- Let $A \subseteq R_{X,Y}$.
 $P((X, Y) \in A) = \sum \sum_{(x,y) \in A} f_{X,Y}(x, y)$

• Joint Probability (density) function, 2D cont. RV:

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{X,Y}(x, y)dydx$$

- $f_{X,Y}(x, y) \geq 0$ for any $(x, y) \in R_{X,Y}$
- $f_{X,Y}(x, y) = 0$ for any $(x, y) \notin R_{X,Y}$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx dy = 1$
or equivalently:
– $\int \int_{(x,y) \in R_{X,Y}} f_{X,Y}(x, y)dx dy = 1$

Marginal Probability Function

Marginal distribution of X is individual distribution of X , ignoring the value of Y . “Projection” of 2D function $f_{X,Y}(x, y)$ to 1D function.

Let (X, Y) be 2D RV with joint probability function $f_{X,Y}(x, y)$:

$$\text{If } Y \text{ is discrete, } f_X(x) = \sum_y f_{X,Y}(x, y)$$

$$\text{If } Y \text{ is continuous, } f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$$

- $f_Y(y)$ defined similarly

- $f_X(x)$ is a p.f., satisfies all properties of prob. fn.

Conditional Distribution

Let (X, Y) be 2D RV with joint probability function $f_{X,Y}(x, y)$. Then $\forall x$ s.t. $f_X(x) > 0$: (X marg prob fn.)

Conditional probability function of Y given $X = x$:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

- Intuition: Distribution of Y given $X = x$
- Only defined for x s.t. $f_X(x) > 0$
- $f_{Y|X}(y|x)$ is a p.f. if we fix x , satisfies prop. of prob.fn.
- But, $f_{Y|X}(y|x)$ is not a p.f. for x : No need for sum / integral over $x = 1$. Hence,
If $f_X(x) > 0$: $f_{X,Y}(x, y) = f_X(x)f_{Y|X}(y|x)$
If $f_Y(y) > 0$: $f_{X,Y}(x, y) = f_Y(y)f_{X|Y}(x|y)$
- **Probability $Y \leq y$, Average Y given $X = x$**
- $P(Y \leq y | X = x) = \int_{-\infty}^y f_{Y|X}(y|x)dy$
- $E(Y | X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy$

Independent Random Variables

$$X \perp Y : \forall x, y, f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

- Necessary condition: $R_{X,Y}$ must be a product space.
i.e. $R_{X,Y} = \{(x, y) | x \in R_X; y \in R_Y\} = R_X \times R_Y$
Else, dependent.

Properties of Independent RV

For X, Y independent RV:

- If $A, B \subseteq \mathbb{R}$, then events $X \in A$ and $Y \in B$ are independent:

$$P(X \in A; Y \in B) = P(X \in A)P(Y \in B)$$

$$P(X \leq x; Y \leq y) = P(X \leq x)P(Y \leq y)$$

- Then, $g_1(X)$ and $g_2(Y)$ are **independent**, for arbitrary g .
- **Conditional distribution** given Independence:

$$f_X(x) > 0 \rightarrow f_{Y|X}(y|x) = f_Y(y)$$

$$f_Y(y) > 0 \rightarrow f_{X|Y}(x|y) = f_X(x)$$

To check independence

1. $R_{X,Y}$ is a product space. i.e. R_X does not depend on Y , vice versa. (e.g. $0 < y < x$ is NOT a product space)
2. Additionally, $f_{X,Y}(x, y) = \text{some } C * g_1(x)g_2(y)$
where g_1 depends on x only and g_2 depends on y only.

Marginal Distribution under Independence

- Since, $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ for independent RV, we derive marginal distribution by standardising $g_1(x)$ and $g_2(y)$.
- For discrete: $f_X(x) = \frac{g_1(x)}{\sum_{t \in R_X} g_1(t)}$
- For continuous: $f_X(x) = \frac{g_1(x)}{\int_{t \in R_X} g_1(t) dt}$

Expectation of a Random Vector

Given **2 variable function** $g(x, y)$:

- If (X, Y) is discrete:

$$E(g(X, Y)) = \sum_x \sum_y g(x, y) f_{X,Y}(x, y)$$

- If (X, Y) is continuous:

$$E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dy dx$$

- If $X \perp Y$:

$$E(XY) = E(X)E(Y)$$

- (If $X \perp Y$, follows that $cov(X, Y) = 0$). However, converse not always true.

Covariance

- For random variables X, Y :

$$cov(X, Y) = E((X - E(X))(Y - E(Y)))$$

- If (X, Y) both **discrete**:

$$cov(X, Y) = \sum_x \sum_y (x - \mu_X)(y - \mu_Y) f_{X,Y}(x, y)$$

- If (X, Y) both **continuous**:

$$cov(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f_{X,Y}(x, y) dx dy$$

- **Alt:** $cov(X, Y) = E(XY) - E(X)E(Y)$
- **Hence, for** $X \perp Y \rightarrow cov(X, Y) = 0$.
(However, converse not always true).
- **Properties of covariance:**
- $cov(aX + b, cY + d) = (ac)cov(X, Y)$
- $V(aX + bY) = a^2V(X) + b^2V(Y) + 2ab * cov(X, Y)$
- $X \perp Y \rightarrow V(X \pm Y) = V(X) + V(Y)$

4. Special Probability Distributions

- **Discrete Distributions:** Study whole classes of discrete RVs that arise frequently in applications.

Discrete Uniform Distribution

- If X has values x_1, x_2, \dots, x_k with **equal probability**

$$f(x) \begin{cases} \frac{1}{k}, & \text{for } x = x_1, x_2, \dots, x_k \\ 0, & \text{otherwise} \end{cases}$$

- **Expectation:**

$$\mu_X = E(X) = \sum_{i=1}^k x_i f_X(x_i) = \frac{1}{k} \sum x_i$$

- **Variance:**

$$\sigma_X^2 = V(X) = E(X^2) - (E(X))^2 = \frac{1}{k} \sum x_i^2 - \mu_X^2$$

Bernoulli

- **Bernoulli Trial:** Random experiment has 2 possible outcomes (success and failure).
- **Bernoulli Random Variable:** X represents number of success in a single Bernoulli Trial. X has only two possible values: 1, or 0.
- **Probability mass function:** Let $0 \leq p \leq 1$ be the probability of success in Bernoulli trial

$$f_X(x) = P(X = x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \\ 0 & \text{otherwise} \end{cases}$$

- $f_X(x) = p^x(1-p)^{1-x}$ for $x = 0$ or 1
- Bernoulli distr. is case of binomial distr. where $n = 1$.
- **Notation:** $X \sim \text{Ber}(p)$ and $q = 1 - p$

$$f_x(1) = p, f_x(0) = q$$

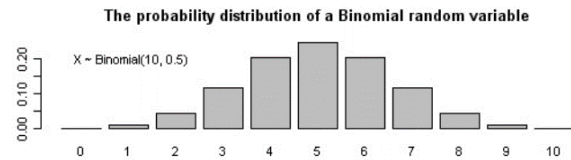
- **Expectation:** $\mu_X = E(X) = p$
- **Variance:** $\sigma_X^2 = V(X) = p(1-p)$
- **Bernoulli Process:** Sequence of repeatedly performed independent and identical Ber. trials.
- Generates sequence of **independent and identically distributed (i.i.d.)** Ber. RVs: X_1, X_2, \dots

Binomial Distribution

- **Binomial RV:** counts **number of successes** in n trials in a Ber. process.
- Given n independent trials with each trial having same probability p of success:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

- **Notation:** $X \sim B(n, p)$
- $E(X) = np, V(X) = np(1-p)$



Negative Binomial Distribution (k^{th} success)

- Let X = no. of independent identical distributed Bernoulli(p) trials until k^{th} success occurs.
- **Probability mass function of X :**

$$P(X = x) = \binom{x-1}{k-1} p^k (1-p)^{x-k}$$

- **Notation:** $X \sim NB(k, p)$
- $E(X) = \frac{k}{p}$ and $V(X) = \frac{(1-p)k}{p^2}$

Geometric Distribution (till 1^{st} success)

- Let X = no. of i.i.d. Bernoulli(p) trials until 1st success occurs.

$$P(X = x) = p(1-p)^{x-1}$$

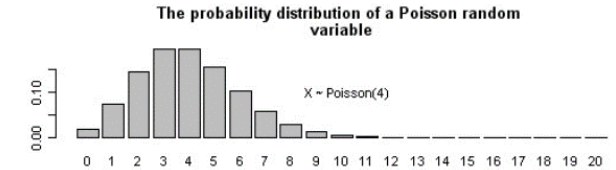
- **Notation:** $X \sim G(p)$
- $E(X) = \frac{1}{p}$ and $V(X) = \frac{1-p}{p^2}$

Poisson Distribution

- **Poisson RV:** Denotes number of events occurring in **fixed period of time or fixed region**, k = no. of occurrences.

$$P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

- **Notation:** $X \sim \text{Poisson}(\lambda)$ where $\lambda > 0$ is expected number of occurrences during given period/region
- $E(X) = \lambda$ and $V(X) = \lambda$



The number of infections X in a hospital each week has been shown to follow a Poisson distribution with a mean of 3 infections per week. What is the probability that

- there is *no* infection for a week?
- there are *less than* 4 infections for a week?

We are given that $X \sim \text{Poisson}(3)$. Then required probabilities are

- $P(X = 0) = e^{-3}$.
- $P(X \leq 3) = e^{-3} \left(1 + 3 + \frac{3^2}{2!} + \frac{3^3}{3!} \right)$.

Poisson Process

- Continuous time process, count number of occurrences within some interval of time. (given **rate** α)
- Properties of **Poisson process with rate parameter α :**
 - Expected no. of occurrences in interval length T : αT
 - No simultaneous occurrences, and no. of occurrences in disjoint intervals independent.
- **Number of occurrences in any interval T** of Poisson process follows $\text{Poisson}(\alpha T)$ distribution. (**Apply $X \sim \text{Poisson}(\alpha T)$ directly**)

Poisson Approximation of Binomial Distribution

- Let $X \sim B(n, p)$. Suppose $n \rightarrow \infty$ and $p \rightarrow 0$ s.t. $\lambda = np$ remains constant.
- Then, approximately, $X \sim \text{Poisson}(\lambda)$.

$$\lim_{p \rightarrow 0; n \rightarrow \infty} P(X = x) = \frac{e^{-np} (np)^x}{x!}$$

- Approximation is good when ($n \geq 20$ and $p \leq 0.05$), or ($n \geq 100$ and $np \leq 10$)

Special Probability Distributions

- **Continuous Distributions:** Many “natural” RVs whose set of possible values **uncountable**. Model with classes of continuous random variables.

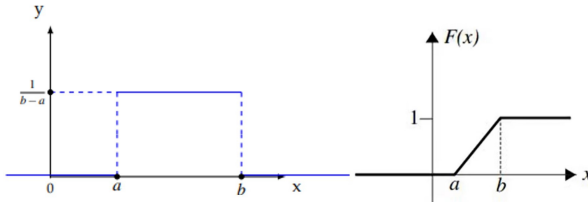
Continuous Uniform Distribution

RV X follows uniform distribution over interval (a, b) if $p.d.f.$ given by:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

- **Notation:** $X \sim U(a, b)$
- $E(X) = \frac{a+b}{2}$ and $V(X) = \frac{(b-a)^2}{12}$ (derive by integration).
- **Cumulative distr. func.** $c.d.f.$ is given by:

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$



Exponential Distribution

X follows exponential distribution with parameter $\lambda > 0$ if $p.f.$ is given by:

$$f_x(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

- **Notation:** $X \sim Exp(\lambda)$
- $E(X) = \frac{1}{\lambda}$ and $V(X) = \frac{1}{\lambda^2}$
- $c.d.f.$ is given by:

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

- Suppose X has exponential distribution with parameter $\lambda > 0$. Then for any positive numbers s and t , we have:

$$P(X > s + t | X > s) = P(X > t)$$

Normal Distribution

X follows normal distribution with mean μ and variance σ^2 if $p.f.$ is given by:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}$$

- **Notation:** $X \sim N(\mu, \sigma^2)$
- $E(X) = \mu$ and $V(X) = \sigma^2$
- $p.f.$ is bell-shaped curve and symmetric about $x = \mu$
- Total area under curve is 1
- 2 normal curves are identical in shape if they have same σ^2 . They differ in location by $\mu_1 - \mu_2$.
- As σ increases, curve becomes more spread out
- If $X \sim N(\mu, \sigma^2)$ and let $Z = \frac{X-\mu}{\sigma}$

Standardized Normal Distribution

If $X \sim N(\mu, \sigma^2)$ and let

$$Z = \frac{X - \mu}{\sigma}$$

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

- $E(Z) = 0$ and $V(Z) = 1$
- $p.f.$ is given by:

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

- Importance of standardizing normal distribution is that it allows us to use tables to find probabilities
- Let $X \sim N(\mu, \sigma^2)$. We can compute $P(x_1 < X < x_2)$ by standardization:

$$x_1 < X < x_2 \leftrightarrow \frac{x_1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{x_2 - \mu}{\sigma}$$

- $c.d.f.$ is given by:

$$\phi(z) = F_Z(z) = \int_{-\infty}^z f_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt$$

- $P(Z \geq 0) = P(Z \leq 0) = \phi(0) = 0.5$
- For any z ,
 $\phi(z) = P(Z \leq z) = P(Z \geq -z) = 1 - \phi(-z)$

- $-Z \sim N(0, 1)$
- If $Z \sim N(0, 1)$, then $\sigma Z + \mu \sim N(\mu, \sigma^2)$
- **Upper Quantile** Given α is upper-tail percentage. The α upper quantile is x_α that satisfies:

$$P(X \geq x_\alpha) = \alpha$$

e.g. The 0.05th (upper) quantile of $Z \sim N(0, 1)$ is 1.645, i.e. $z_{0.05} = 1.645$.

- $P(Z \geq z_\alpha) = P(Z \leq -z_\alpha) = \alpha$
- Upper z_α = Lower $z_{1-\alpha}$

Normal Approximation to Binomial Distribution

Let $X \sim B(n, p)$, then as $n \rightarrow \infty$:

$$Z = \frac{X - E(X)}{\sqrt{V(X)}} = \frac{X - np}{\sqrt{np(1-p)}} \sim N(0, 1)$$

- Approximation is good when $np > 5$ and $n(1-p) > 5$