

# AMS 553.430 - Introduction to Statistics

Lectures by Dwijavanti P Athreya  
Notes by Kaushik Srinivasan

Johns Hopkins University  
Fall 2018

<b>Lecture 0 (2018-08-30)</b>	<b>1</b>	<b>Lecture 3 (2018-09-10)</b>	<b>20</b>
<b>Lecture 1 (2018-08-30)</b>	<b>16</b>	<b>Lecture 4 (2018-09-12)</b>	<b>23</b>
<b>Lecture 2 (2018-09-05)</b>	<b>17</b>		

## Introduction

Math 553.430 is one of the most important courses that is required/recommended for the engineering-based majors at Johns Hopkins University.

These notes are being live-Texed, through I edot for Typos and add diagrams requiring the *TikZ* package separately. I am using Texpad on Mac OS X.

I would like to thank Zev Chonoles from The University of Chicago and Max Wang from Harvard University for providing me with the inspiration to start live-Texing my notes. They also provided me the starting template for this, which can be found on their personal websites.

Please email any corrections or suggestions to [ksriniv4@jhu.edu](mailto:ksriniv4@jhu.edu).

## Lecture 0 (2018-08-30)

### Introduction to Probability (553.420) Review

#### Part 1 - Counting

- ① Multiplication rule (Basic Counting Principle)
- ② Combinations/Permutations
  - Sampling with or without replacement.  $\Rightarrow$  Inclusion-Exclusion Principle

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \qquad {}^nP_k = \frac{n!}{(n-k)!}$$

- ③ Birthday Problem
- ④ Matching Problem (inclusion-exclusion principle)
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  - $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$
  - etc...
- ⑤  $n$  balls going into  $m$  boxes (all are distinguishable)

**Example.**  $n$  balls numbered  $1, 2, \dots, n$ .  $n$  boxes labelled  $1, 2, \dots, n$ . Distribute the balls into the boxes, one in each box.  $M_i$  = ball  $i$  is in box  $i$
- ⑥ Multinomial Coefficients e.g. assign A, B, C, D, to different students  $\rightarrow$  anagram problem
  - $n$  distinct objects into  $r$  distinct groups

$$\frac{n!}{n_1!n_2!n_3! \dots n_r!} = \binom{n}{n_1, n_2, n_3, \dots, n_r}$$

- ⑦ Pairing Problem

$$2n \text{ people, paired up } \begin{cases} \text{ordered: } \binom{2n}{2, 2, \dots, 2} & \text{e.g. different courts for players} \\ \text{unordered: } \frac{\binom{2n}{2, 2, \dots, 2}}{n!} \end{cases}$$

- ⑧ Partition of integers  $\rightarrow n$ : sum of integer,  $r$ : number of partitions

$$\binom{n+r-1}{r-1} = \binom{n+r-1}{n}$$

## Basics of Probability

### Axioms

- ①  $0 \leq P(A) \leq 1, \forall A$
- ②  $P(\Omega) = 1 \rightarrow$  where  $\Omega$  is the sample space
- ③ Countable additivity
  - if  $A_1, \dots, A_n$  are mutually exclusive, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P(A_1) + P(A_2) + \dots = \sum_{i=1}^{\infty} P(A_i)$$

$$\Rightarrow P(A) = 1 - P(A^c)$$

$$P(A) = \frac{|A|}{|\Omega|}$$

### Conditional Probability

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

### Law of Total Probability

$$P(A) = \sum_j P(A|B_j)P(B_j) = \sum P(A \cap B_j) \quad \underbrace{\bigcup_j B_j = \Omega}_{\text{partition of } \Omega}$$

### Bayes Rule

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_j P(A|B_j)P(B_j)} \quad \underbrace{\bigcup_j B_j = \Omega}_{\text{partition of } \Omega}$$

### Independent events

If we have events  $A_1, A_2, \dots, A_n$ , then

$$P(A_1 \cap A_2 \cap A_3 \dots A_n) = P(A_1) \cdot P(A_2) \cdot P(A_3) \cdot \dots \cdot P(A_n)$$

## Introduction to Discrete and Continuous Random Variables

**Random Variable** - a real valued function defined on the sample space of an experiment  $X : \Omega \rightarrow \mathbb{R}, \forall \omega \in \Omega, X(\omega) \in \mathbb{R}$

Function	Discrete	Continuous
Probability Function	PMF: $P(X = x)$	PDF: $f_x(x)$
Probability Distribution	$\sum_x P(X = x) = 1$	$\int_x f_x(x)dx = 1$
Expectation	$E[X] = \sum_x xP(X = x)$	$E[X] = \int_x xf(x)dx$
Variance	$Var[X] = E[X^2] - (E[X])^2$	$Var[X] = E[X^2] - (E[X])^2$

## Law of the Unconscious Statistician (LOTUS)

$$\text{1-dim} \quad E[g(x)] = \sum_x g(x)P(X = x) \quad \Bigg/ \quad E[g(x)] = \int_x g(x)f(x)dx$$

$$\text{2-dim} \quad E[g(X, Y)] = \sum_y \sum_x g(x, y)P(X = x, Y = y) \quad \Bigg/ \quad E[g(X, Y)] = \int_y \int_x g(x, y)f(x, y)dxdy$$

## Discrete Distributions

- |                          |                                |
|--------------------------|--------------------------------|
| 1. Bernoulli( $p$ )      | 4. Geometric( $p$ )            |
| 2. Binomial( $n, p$ )    | 5. Negative Binomial( $n, p$ ) |
| 3. Poisson ( $\lambda$ ) | 6. Hypergeometric( $N, M, n$ ) |

## Bernoulli Distribution

$X$  is a random variable with Bernoulli( $p$ ) distribution

$$X \sim \text{Bernoulli}(p)$$

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

## Binomial Distribution

A sum of i.i.d. (identical, independent distribution) Bernoulli( $p$ ) R.V.

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

Support :  $x \in \{0, 1, \dots, n\}$

$n$  : sample size       $p$  : probability of success

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

$$E[X] = np \qquad \qquad \qquad Var(X) = np(1 - p)$$

- Approximation methods  $\Rightarrow$

- if  $n$  is large,  $p$  very small and  $np < 10$ .  $\Rightarrow$  use Normal  $(np, np(1-p))$
- $p \approx \frac{1}{2} \Rightarrow$  Use Poisson  $(\lambda = np)$
- Mode:
  - if  $(n+1)p$  integer, mode =  $(n+1)p$  or  $(n+1)p - 1$ .
  - if  $(n+1)p \notin \mathbb{Z}$  mode is  $\lfloor (n+1)p \rfloor$
  - **Proof:** consider  $\frac{P(X=x)}{P(X=x-1)}$  going below 1.

## Poisson Distribution

$$\begin{aligned}
 X &\sim \text{Poisson}(\lambda) \\
 x &\in \{0, 1, \dots\} \\
 \lambda &: \text{parameter} \\
 P(X=x) &= \frac{e^{-\lambda} \lambda^x}{x!} \\
 E[X] &= \lambda & \text{Var}(X) &= \lambda
 \end{aligned}$$

- Approximations
  - if  $n$  is large  $\Rightarrow$  Normal $(\lambda, \lambda)$
- Sums of Poisson
 

Let  $X \sim Po(\lambda)$     $Y \sim Po(\mu)$     $\Rightarrow$     $X + Y \sim Po(\mu + \lambda)$

## Negative Binomial

$$\begin{aligned}
 X &\sim NB(r, p) \\
 \text{Support : } x &= \{r, r+1, \dots\} \\
 r &= \text{the } r\text{th success} \\
 p &= \text{probability of success} \\
 P(X=k) &= \binom{k+r-1}{k} \cdot (1-p)^r \cdot p^k
 \end{aligned}$$

A sum of i.i.d Geometric( $p$ ) R.V.

■  $a^{\text{th}}$  head before  $b^{\text{th}}$  tail

**Example.** A coin has probability  $p$  to land on a head,  $q = 1 - p$  to land on a tail.

$P[5^{th} \text{ tail occurs before the } 10^{th} \text{ head}]?$

$$\left\{ \begin{array}{l} = P[5^{th} \text{ tail occurs before or on the } 14^{th} \text{ flip}] \\ = P[\text{Neg Binomial}(5, q) = 5, 6, 7, \dots, 14] \\ = \sum_{x=5}^{14} \binom{x-1}{4} q^5 p^{x-5} \end{array} \right. \quad (\text{or}) \quad \left\{ \begin{array}{l} = P[\text{at least } 5 \text{ tails in } 14 \text{ flips}] \\ = P[\text{binom}(14, q) = 5, 6, 7, \dots, 14] \\ = \sum_{x=5}^{14} \binom{14}{x} q^x p^{14-x} \end{array} \right.$$

## Geometric Distribution

$$\begin{array}{l} X \sim \text{Geometric}(p) \\ \text{Support : } x \in \{1, 2, \dots\} \\ p : \text{probability of success} \\ P(X = r) = (1 - p)^{(r-1)} \cdot p \\ \text{prob for 1st success on } r\text{th trial} \\ E[X] = \frac{1}{p} \qquad \qquad \qquad \text{Var}(X) = \frac{1-p}{p^2} \end{array}$$

**Example. ■** Coupon Question

Variation A:  $N$  different types of coupons  $\rightarrow P(\text{ get a specific type}) = \frac{1}{N}$

Question:  $E[\text{draws to get } 10 \text{ different coupons}]?$

Answer:

$$X = X_1 + X_2 + \dots + X_{10} \qquad X_i = \# \text{ draws to get the } i\text{th distinct coupon type}$$

$$\boxed{X_i \sim \text{Geo}(p_i)} \qquad p_i : \text{prob to get a new coupon} \leftarrow \text{success, given that we have } i-1 \text{ types of coupons}$$

Hence,  $E[X_1] = 1$

$$E[X_2] = \frac{1}{p_2} = \frac{1}{\frac{N-1}{N}} = \frac{N}{N-1}$$

$$E[X_3] = \frac{1}{p_3} = \frac{1}{\frac{N-2}{N}} = \frac{N}{N-2}$$

$\vdots$

$$E[X_{10}] = \frac{1}{p_{10}} = \frac{1}{\frac{N-9}{N}} = \frac{N}{N-9}$$

$$\text{So, } E[X] = E[X_1] + E[X_2] + \dots + E[X_{10}] = E\left[\sum_{i=1}^{10} X_i\right] = 1 + \frac{N}{N-1} + \frac{N}{N-2} + \dots + \frac{N}{N-9}$$

Variation B: Same setting, now you draw 10 times.

Question:  $E[\# \text{ different types of coupons}]?$

Answer:

$$X = I_1 + I_2 + \dots + I_N$$

$$I_i \begin{cases} 1 & \text{if we have this type of coupon} \\ 0 & \text{o/w} \end{cases}$$

$$\begin{aligned}
E[I_i] &= P(\text{we draw coupon } i \text{ in } 10 \text{ draws}) \\
&= 1 - P(\text{we don't have coupon } i) \quad \text{we use binomial distribution where } 1 - P(N = 0) \\
&= 1 - \left(\frac{N-1}{N}\right)^{10}
\end{aligned}$$

$$E[X] = E\left[\sum_{i=1}^N I_i\right] = NE[I_i] = \boxed{N\left[1 - \left(\frac{N-1}{N}\right)^{10}\right]}$$

## Hypergeometric Distribution

$$\begin{aligned}
X &\sim \text{Hyp}(N, M, n) \\
N &\in \{0, 1, 2, \dots\} \quad M \in \{0, 1, \dots, N\} \quad n \in \{0, 1, \dots, N\} \\
\text{Support : } k &\in \{\max(0, n + M - N), \min(n, M)\} \\
N &\text{ is the population size} \quad K \text{ is the no. of success states in the population} \\
n &\text{ is the no. of draws (i.e. quantity drawn in each trial)} \\
k &\text{ is the no. of observed successes} \\
P(X = k) &= \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}
\end{aligned}$$

## Continuous Distributions

### Uniform Distribution

$$\begin{aligned}
X &\sim \text{Unif}(a, b) \\
f_X(x) &= \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{o/w} \end{cases} \\
E[X] &= \frac{a+b}{2} \quad \quad \quad \text{Var}(X) = \frac{(b-a)^2}{12}
\end{aligned}$$



## Normal Distribution

$$\begin{aligned} X \sim N(\mu, \sigma^2) &\Rightarrow Z = \frac{X - \mu}{\sigma} \sim N(0, 1) \text{ with CDF } P(Z \leq z) = \Phi(z) \\ \Phi(-x) &= 1 - \Phi(x) \\ \text{Support: } x &\in (-\infty, \infty) \\ f_X(x) &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ E[X] &= \mu \qquad \qquad \qquad \text{Var}(X) = \sigma^2 \end{aligned}$$

- Sums and differences of Normal R.V.

$$\begin{array}{cc} X_1 \sim N(\mu, \sigma^2) & X_2 \sim N(\mu, \sigma^2) \\ Y_1 = X_1 + X_2 & Y_2 = X_1 - X_2 \\ \underbrace{Y_1 \sim N(2\mu, 2\sigma^2)}_{\text{has } \mu} & \underbrace{Y_2 \sim N(0, 2\sigma^2)}_{\text{doesn't have } \mu} \end{array}$$

- The sum and difference of Normal R.V. are Normal R.V.
- Any Linear Combination of Independent Normal R.V. is a Normal R.V.

- Dependence

- $Y_2 = X_1 - X_2$  density does not depend on  $\mu$ . But density of  $X_1 + X_2$  does.
- Key idea is used in Data Reduction

## Exponential distribution

$$\begin{aligned} X &\sim \text{Exp}(\lambda) \\ \text{Support: } x &\in [0, \infty) \\ f_X(x) &= \lambda e^{-\lambda x} \\ E[X] &= \frac{1}{\lambda} \qquad \qquad \qquad \text{Var}(X) = \frac{1}{\lambda^2} \end{aligned}$$

**Lack of memory property:**  $P(X \geq s + t | X \geq t) = P(X \geq s)$

- $M = \min \text{ of } \text{exp}(\lambda) \text{ and } \text{exp}(\mu) \Rightarrow M \sim \text{exp}(\lambda + \mu)$
- $M = \min \text{ of } X_1, X_2, \dots, X_n, \text{ where } X_i \sim_{\text{i.i.d.}} \text{exp}(\lambda) \Rightarrow \text{exp}(n\lambda)$

## Gamma Distribution

$$X \sim \text{Gamma}(\alpha, \beta)$$

$$\text{Support: } x \in [0, \infty)$$

$$F_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$

$$E[X] = \frac{\alpha}{\beta}$$

$$\text{Var}(X) = \frac{\alpha}{\beta^2}$$

$$\textbf{Gamma Function: } \Gamma(z) = (z-1)! = \int_0^\infty x^{z-1} e^{-x} dx$$

$$\Gamma(n) = (n-1)!$$

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

- Sums of Gamma

$$- \underset{\text{ind}}{\text{Gamma}(s, \lambda)} + \text{Gamma}(s, \lambda) = \text{Gamma}(s+t, \lambda)$$

## Beta Distribution

$$X \sim \text{Beta}(\alpha, \beta)$$

$$\text{Support: } x \in [0, 1]$$

$$f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$$

$$E[X] = \frac{\alpha}{\alpha + \beta}$$

$$\text{Var}(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

- Gamma to Beta

$$X \sim \text{Gamma}(\alpha_1, \beta) \quad Y \sim \text{Gamma}(\alpha_2, \beta)$$

$$\text{Then transformation } U = \frac{X}{X+Y} \sim \text{Beta}(\alpha_1, \alpha_2) \quad (\text{Use } X = UV, Y = V - UV)$$

## Chi-Square

**Chi-Square:**  $\chi_n^2$  is Chi-square with degrees of Freedom  $n$

$$\chi_n^2 = Z_1^2 + Z_2^2 + \cdots + Z_n^2 \quad \text{where } Z_i \sim \text{standard normal. } Z_i \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right)$$

$$\Rightarrow \chi_n^2 = n \text{ i.i.d. } Z_i \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right)$$

$$= \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$$

## CDF in General

- $F_x(t) = P(X \leq t)$

$$= \sum_{x \leq t} P(X = x) \quad \text{discrete}$$

$$= \int_{-\infty}^t f(x) dx \quad \text{continuous}$$

- **Discrete:** "Left open, right closed"  $\Rightarrow$  if you flip the sign (from  $<$  to  $\leq$ ) in the left, you flip the sign of  $a$  (from  $a$  to  $a^-$ )

$$- P(a < x \leq b) = F(b) - F(a)$$

$$- P(a \leq x \leq b) = F(b) - F(a^-)$$

$$- P(a < x < b) = F(b^-) - F(a)$$

$$- P(a \leq x < b) = F(b^-) - F(a^-)$$

- **Continuous:** (because a point doesn't have a mass)

$$P(a \leq x \leq b) = \int_a^b f(x) dx = F(b) - F(a)$$

## Integration by Recognition

$$1 = \int_{-\infty}^{\infty} \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} dx \quad \sigma\sqrt{2\pi} = \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} dx \quad (\text{normal dist.})$$

## Joint Distribution

### Discrete

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

$$\text{Indep} \Rightarrow P_X(x)P_Y(y)$$

### Continuous

$$F_{X,Y}(x, y) = F_X(x)f_Y(y)$$

$$= \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$$

- **Marginal Density/PMF:**

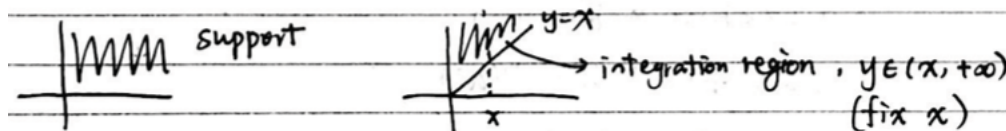
$$\text{Continuous:} \quad f_X(x) = \int_x f_{X,Y}(x, y) dy \quad \text{and} \quad f_Y(y) = \int_y f_{X,Y}(x, y) dx$$

*\* the bounds for  $y$  in the integration can depend on  $x$ , and vice versa*

$$\text{Discrete:} \quad P_X(x) = \sum_y P(X = x, Y = y) \quad \text{and} \quad P_Y(y) = \sum_x P(X = x, Y = y)$$

- Use joint pdf to compute probability

e.g.  $P(X < Y) = \int_0^{\infty} \int_x^{\infty} f(x, y) dy dx$  assume  $x > 0, y > 0$



- **Independence:** If  $X, Y$  are independent, then

**Continuous:**  $f(x, y) = f_X(x)f_Y(y)$

**Discrete:**  $P(X = x, Y = y) = P(X = x)P(Y = y)$

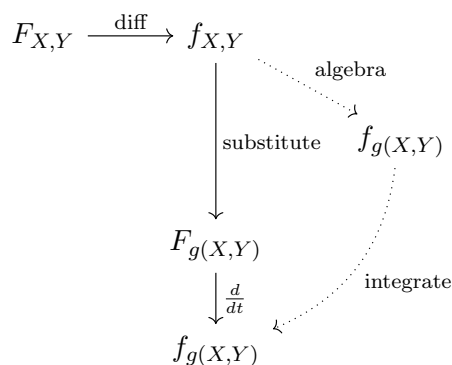
- **Convolution:** assume  $X, Y$  are independent

**Discrete:**  $P_{X+Y}(a) = \sum_y P_X(a - y)P_Y(y) = \sum_x P_X(x)P_Y(a - x)$

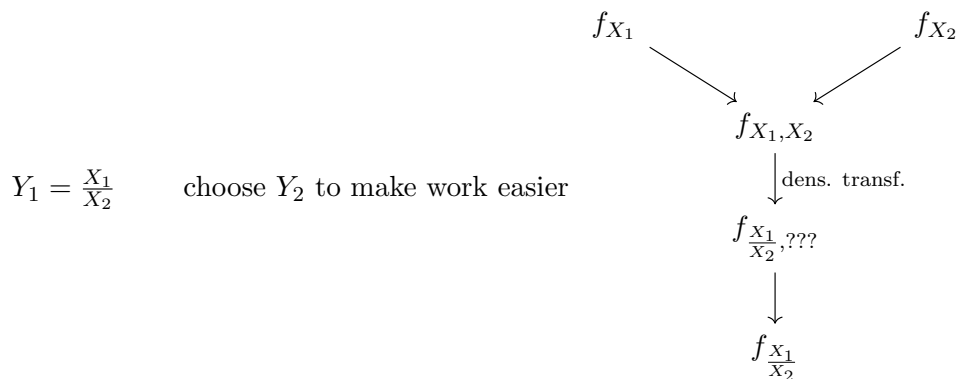
**Continuous:**  $f_{X+Y}(a) = \int_y f_X(a - y)f_Y(y)dy = \int_y f_X(x)f_Y(a - x)dx$

**MGF:** we can use this  $M_{X+Y}(t) = M_X(t)M_Y(t) \rightarrow$  then identify dist of  $X+Y$  from mgf

- **Density Transformation:**



$X_1$  &  $X_2$  are indep r.v.  $\Rightarrow$  want to find density of  $\frac{X_1}{X_2}$



## Density Transformation

For density transformation e.g. finding pdf of  $U = X + Y$

- Convolution
- MGF
- Jacobian
- CDF Transformation

- Use CDF: Computer  $P(Y \leq y) = P(g(x) = y)$

- **1-dim:** If Y is monotonically increasing or decreasing:  $Y = g(x)$   $f_Y(y) = f_X(x(y)) \cdot |(x^{-1})'(y)|$

- **2-dim:** Joint Density:

$$(X, Y) \rightarrow (U, V) \quad U = h_1(X, Y) \quad V = h_2(X, Y)$$

$$f_{U,V}(u, v) = f_{X,Y}(x(u, v), y(u, v)) \cdot |J|$$

$$\text{where } J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} \quad \text{determinant}$$

- if  $Z = X + Y$  (2-dim  $\rightarrow$  1-dim) use CDF. Compute  $P(Z \leq z) = P(X + Y \leq z)$ . Integrate  $f(x, y)$  over this region.

## Sterling's Formula

$$n! \approx \sqrt{2\pi n} \cdot \left(\frac{n}{e}\right)^n$$

This is only really useful when  $n$  is large, when factorials are represented as ratios.

## Conditional distribution

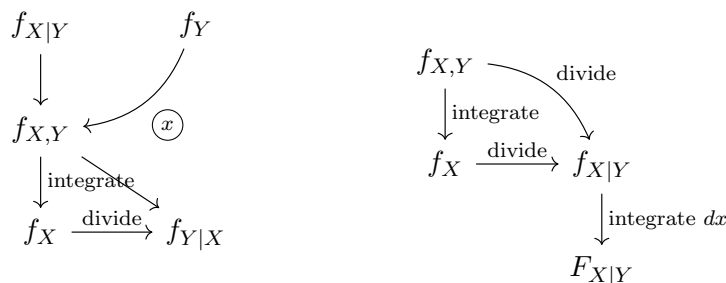
$$\text{Discrete} \quad P_{X|Y=y}(x|y) = \frac{P_{X,Y}(x, y)}{P_Y(y)} = \frac{P(X = x, Y = y)}{P(Y = y)}$$

$$\Rightarrow \sum_y P_{X,Y}(x, y) = \sum_y P_{X|Y=y}(x|y) \cdot P_Y(y)$$

$$\text{Continuous} \quad f_{X|Y=y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

$$\Rightarrow f_X(x) = \int_y f(x, y) dy = \int_y f_{X|Y=y}(x|y) \cdot f_Y(y) dy$$

$$F_{X|Y}(x|y) = \int_{-\infty}^x f_{X|Y}(x|y) dx$$



## Conditional Expectation

$$E[X|Y = y] = \sum_x xP(X = x|Y = y)$$

$$E[X|Y = y] = \int_x xf(x|y)dx$$

$E[X|Y]$  : compute  $E[X|Y = y]$  first, replace  $y$  with  $Y$

### • Properties:

- $E[aU + bV|Y = y] = aE[U|Y = y] + bE[V|Y = y]$  *LOTUS*
- If  $g(Y) = X$  then  $E[X|Y = y] = X$
- If  $X$  and  $Y$  are independent, then  $E[X|Y = y] = E[X]$

## Conditional Variance

$$\boxed{Var(X|Y) = E[(X - E[X|Y])^2]} \quad (\text{conditional variance})$$

$$\boxed{Var(X|Y) = E[X^2|Y] - (E[X|Y])^2} \quad (\text{unconditional variance})$$

## Ordered Statistics

Consider  $X_1, X_2, \dots, X_n$   $X_{(j)}$  = j-th smallest

$$F_{\max(X_i)}(t) = P(\max X_i \leq t) = P(X_1 \leq t) \cdot P(X_2 \leq t) \cdots P(X_n \leq t)$$

$$= [F_X(t)]^n \quad \boxed{f_{\max X_i}(t) = nF(t)^{n-1}f_X(t)}$$

$$F_{\min(X_i)}(t) = 1 - P(\min x_i \geq t) = 1 - P(X_1 \geq t) \cdot P(X_2 \geq t) \cdots P(X_n \geq t)$$

$$= 1 - [1 - F_X(t)]^n \quad \boxed{f_{\min X_i}(t) = n[1 - F(t)]^{n-1}f_X(t)}$$

**General:**  $j$ -th order statistic

$$f_{x(j)}(t) = \binom{n}{j-1, 1, n-j} F_X(t)^{j-1} \cdot f_X(t) \cdot [1 - F_X(t)]^{n-j}$$

**As Beta distribution:** Let  $U_1, U_2, \dots, U_N \sim i.i.d.$  Uniform(0, 1) and let  $1 \leq j \leq N$   
 $U_{(j)}$  = jth smallest in  $U_{(1)}, U_{(2)}, \dots, U_{(N)}$  (ordered statistics). Then,

$$U_{(j)} \sim \text{Beta}(j, N - j + 1)$$

$$E[U_{(j)}] = \frac{j}{N+1}$$

## Expectation and Variance

### Law of Total Expectation:

$$E[X] = E[E[X|Y]]$$

### Law of Total Variance:

$$Var(X) = E[Var(X|Y)] + Var[E(X|Y)]$$

## Expectation

① linearity of expectation

② How to compute

(a) LOTUS or definition (use density to integrate)

(b) MGF:  $M^{(n)}(0) = E[X^n]$  or by recognition

(c)  $E[X^2] = Var[X] + E[X]^2$

(d) Tail probability X is non-neg R.V. ( $x > 0$ ) then  $E[X] = \sum_{t=0}^{\infty} P(X \geq t)$  or  $= \int_0^{\infty} P(X \geq t) dt$

## Variance

①  $Var(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^n Var(X_i) + \sum_{i \neq j} Cov(X_i, X_j)$

if  $X_i, X_j$  identical (not independent)  $= nVar(X_i) + n(n-1)Cov(X_i, X_j)$   $i \neq j$

$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$

② **Covariance:**

$$Cov(X, Y) = E[XY] - E[X]E[Y]$$

$$Cov(X, c) = 0 \quad c \text{ is a constant}$$

$$Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$$

$$Cov(cX, dZ) = cd \cdot Cov(X, Z)$$

$$Cov(aX + b, cY + d) = ac \cdot Cov(X, Y) \quad a, b, c, d \text{ are constants}$$

$$Cov(X, Y) = 0 \quad \text{If } X \perp Y \text{ (independent)}$$

③ **Correlation Coefficient:**

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X, Y)}{\sigma_x \sigma_y}$$

## MGFs

Let  $X$  be a random variable. Then

$$M_X(t) = E[e^{tX}]$$

it can also be written as:

$$\begin{aligned} &= E\left[\sum_{j=0}^{\infty} \frac{(tX)^j}{j!}\right] \\ &= E\left[\sum_{j=0}^{\infty} \left(\frac{X^j}{j!} \cdot t^j\right)\right] \\ &\boxed{M_X^{(n)}(0) = E[X^n]} \end{aligned}$$

If  $X$  and  $Y$  are independent, then

$$\begin{aligned} M_{X+Y}(t) &= E[E^{(X+Y)t}] \\ &= E[e^{tX}]E[e^{tY}] \\ &= M_X(t)M_Y(t) \end{aligned}$$

## Limit Theorems

### Markov's Inequality

For any non-negative random variable  $X$

$$P(X \geq a) \leq \frac{E(X)}{a} \quad (\text{for any } a > 0)$$

*Proof.* Let  $X \geq 0$  a random variable and let  $a > 0$ . Define new random variable from  $X$  as  $Y_a$

$$\begin{aligned} Y_a &= \begin{cases} 0 & \text{if } X < a \\ a & \text{if } X \geq a \end{cases} \\ 0 \leq Y_a \leq X &\implies \underbrace{E[Y_a]}_{a \cdot P(X \geq a)} \leq E[X] \\ E[Y_a] &= 0 \cdot P(Y_a < a) + a \cdot P(X \geq a) \\ E[Y_a] = a \cdot P(X \geq a) \leq E[X] &\implies \boxed{P(X \geq a) \leq \frac{E(X)}{a}} \end{aligned}$$

■

### Chebyshev's Inequality

For any random variable  $Y$  with mean  $\mu_y$  and variance  $\sigma_y^2$

$$P(|Y - \mu_y| \geq c) \leq \frac{\sigma_y^2}{c^2} \quad (\text{for any } c > 0)$$



*Proof.*

$$P(|Y - \mu_y| \geq c) = P(\underbrace{|Y - \mu_y|^2}_{=X} \geq c^2)$$

$$P(|Y - \mu_y|^2 \geq c^2) \leq \frac{E[|Y - \mu_y|^2]}{c^2} = \frac{\sigma_y^2}{c^2}$$

■

This is the same as

$$- P(|Y - \mu_y| \geq k\sigma_y) \leq \frac{1}{k^2}$$

$$- P(|Y - \mu_y| \leq k\sigma_y) \geq \underbrace{1 - \frac{1}{k^2}}_{\text{very conservative}}$$

## Central Limit Theorem

$$\sum_{i=1}^n X_i \sim \mathcal{N}(n\mu_n, n\sigma_x^2)$$

$$\frac{1}{n} \sum_{i=1}^n X_i \sim \mathcal{N}\left(\mu_n, \frac{\sigma_x^2}{n}\right)$$

## Weak Law of Large Numbers

If  $X_1, X_2, \dots$  are *i.i.d.* with a mean  $\mu$

$$\text{then } \lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0$$

## Strong Law of Large Numbers

$$X \xrightarrow{p} \mu_X \quad \text{as } n \rightarrow \infty$$

$$Pr\left(\lim_{n \rightarrow \infty} \bar{X}_n = \mu\right) = 1$$

## Lecture 1 (2018-08-30)

### Survey Sampling

We have a population of objects under study (people, animals, places, etc.). We will consider a single numerical measurement associated to object  $i$ :  $x_i$

**Example.**  $N = 5000$ ,  $x_i$  = height of person  $i$ , Population size =  $N$ . We denote population measurements  $\{x_1, x_2, \dots, x_N\}$

Compute population quantities:

- population total  $\tau = \sum_{i=1}^N x_i$
- population mean  $\mu = \frac{\tau}{N} = \frac{\sum_{i=1}^N x_i}{N}$

**Note:**  $\tau$  and  $\mu$  are population parameters, their computation depends on all the population data.

**Question.** How to estimate  $\tau$  and  $\mu$  based on a sample of observation from this population?

Classical Answer: Choose a "random" sample of objects and associated measurements denoted  $\{x_1, x_2, \dots, x_n\}$ . *Note:* capital  $X_i$  denote random variables.

Whiter "Random"? Two types of ways to sample:

– without replacement

– with replacement

**Claim 1.** If  $X_i$  are drawn without replacement, then the distribution of  $X_1$  and  $X_2$  are identical. Is this true? **In fact, it is**  $\Rightarrow$  They are **NOT** independent but they are identically distributed.

$$P(\text{Ace in Pos 1}) = P(\text{Ace in Pos 2}) = \frac{4}{52}$$

### Combinatorial Approach

"well-shuffled deck"  $\leftrightarrow$  all  $52!$  rearrangements of the card are equally likely. How many rearrangements have ace at pos 1?  $4 \cdot 51!$

$$P(A_1) = \frac{4 \cdot 51!}{52!} = \frac{4}{52} = P(A_2) = P(A_{19}) = P(A_{36})$$

**Question.** If  $X_1$  and  $X_2$  are identically distributed, then how do they differ between corresponding draws with replacement?

**Answer.** Independence. We can have Random Variables that are identically distributed and not independent. Note if independent,  $P(A_2|A_1) = P(A_2)$ .

**with replacement**

$$P(A_1) = \frac{4}{52}, \quad P(A_2) = \frac{4}{52}$$
$$P(A_2|A_1) = \frac{4}{52}$$

**without replacement**

$$P(A_1) = \frac{4}{52}, \quad P(A_2) = \frac{4}{52}$$
$$P(A_2|A_1) = \frac{3}{51}$$

We can see from this that depending on sampling method, we gain or lose independence. In the finite population sampling method, we have  $1, \dots, N$  objects we care about.

**Loss of Independence** when choosing sampling method is important.

## Lecture 2 (2018-09-05)

Finite Population sampling – without replacement. Mean/expected value and variance of  $\bar{X}$

Suppose our population is given by  $\{x_1, \dots, x_N\} = \{1, 2, 2, 7, 8, 9\}$  where

$$N = 6, \quad x_1 = 1 \quad x_2 = 2 \quad x_3 = 2 \quad x_4 = 7 \quad x_5 = 8 \quad x_6 = 9$$

Could also describe it by counting.

Distinct Value	frequency
$\varphi_1 = 1$	$n_1 = 1$
$\varphi_2 = 2$	$n_2 = 2$
$\varphi_3 = 7$	$n_3 = 1$
$\varphi_4 = 8$	$n_4 = 1$
$\varphi_5 = 9$	$n_5 = 1$

Possible sample of size  $n = 6$ , where we sample without replacement

$$X_1 = 7 \quad X_2 = 2 \quad X_3 = 8 \quad X_4 = 9 \quad X_5 = 1 \quad X_6 = 2$$

Sample here is the same as population as  $\bigcirc n=N$

Same thing with replacement

$$X_1 = 9 \quad X_2 = 9 \quad X_3 = 9 \quad X_4 = 9 \quad X_5 = 9 \quad X_6 = 9$$

Typically  $N$  is large and  $n \ll N$

Recall population parameters

$$\mu = \frac{\sum_{i=1}^N X_i}{N} \quad \tau = N\mu = \sum_{i=1}^N X_i$$

Next,  $\sigma^2$  (population variance)

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (\sigma^2 \text{ is pop. variance})$$

Alternatively, we can also express  $\sigma^2$  as

$$\begin{aligned} \sigma^2 &= \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} = \frac{\sum_{i=1}^N (x_i^2 - 2\mu x_i + \mu^2)}{N} \\ &= \frac{\sum_{i=1}^N x_i^2}{N} - \frac{2\mu}{N} \underbrace{\sum_{i=1}^N x_i}_{\mu} + \frac{N\mu^2}{N} \\ &= \frac{\sum_{i=1}^N x_i^2}{N} - 2\mu^2 + \mu^2 \end{aligned}$$

$$= \underbrace{\left( \frac{1}{N} \sum_{i=1}^N x_i^2 \right)}_{\text{2nd moment}} - \mu^2 = \mu^{(2)} - \mu^2$$

**Define:**  $\mu^{(k)} = \frac{1}{N} \sum_{i=1}^N x_i^k$

## Sample Mean $\bar{X}$ as an estimator

A function of the sample data for the population  $\mu$ .

*Note:* If the sample is random ( $X_1, \dots, X_n$  are R.Vs), then  $\bar{X}$  is **random**!

Questions:

- ① How is  $\bar{X}$  distributed? - in theory, if we know ①, then we know the answers ② & ③ too.
- ② What is  $E[\bar{X}]$ ?
- ③ What is  $Var(\bar{X})$ ?

Let's address ②

Consider  $E[\underbrace{X_1}_{\text{first draw}}]$

possible values for  $X_1 = \{x_1, \dots, x_N\}$

$$P(X_1 = x_k) = \frac{1}{\binom{N}{1}} = \frac{1}{N}$$

**e.x.**  $\{\underbrace{1}_{x_1}, \underbrace{2}_{x_2}, \underbrace{2}_{x_3}, \underbrace{7}_{x_4}, \underbrace{7}_{x_5}, \underbrace{9}_{x_6}\}$  gives every separate entry a unique ticket even if they are the same

$$E[X_1] = \frac{1}{N} \sum_{k=1}^N x_k = \mu = E[X_2] \quad (\text{b/c } X_1 \text{ \& } X_2 \text{ are identically dist.})$$

In sampling without replacement  $X_i$  &  $X_j$  are still identically distributed, but they are not independent.

In sampling with replacement,  $X_i$  &  $X_j$  are *i.i.d.*

Note that whether or not  $X_1, \dots, X_n$  are independent,

$$E\left[\sum_{i=1}^N X_i\right] = \sum_{i=1}^N E[X_i]$$

*Note:* The sample mean is equal to expected population mean regardless of sampling with or without replacement.

$$\begin{aligned} E[\bar{X}] &= E\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n} \sum_{i=1}^n E[X_i] \\ &= \frac{n\mu}{n} = \mu \end{aligned}$$

Since  $E[\bar{X}] = \mu$ , we say  $\bar{X}$  is an unbiased estimator for  $\mu$ . **BUT**  $\underbrace{\bar{X}}_{\text{R.V.}} \neq \underbrace{\mu}_{\text{constant}}$

Let's address ③

## Sampling with replacement.

**Theorem.** *Sampling from finite population with replacement*

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

*Proof.* Here  $X_1, \dots, X_n$  are *i.i.d.*. In general,  $X_i$ 's are R.V. and  $a_i$ 's are constants

$$\text{Var}\left(\sum_i a_i X_i\right) = \sum_i \sum_j a_i a_j \text{cov}(X_i, X_j)$$

If  $X_1, \dots, X_N$  are independent,  $\text{Cov}(X_i, X_j) = 0$ ! Hence

$$\begin{aligned} \text{Var}(\bar{X}) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \underbrace{\text{Var}(X_i)}_{\text{a constant}} \\ &\boxed{\text{Var}(\bar{X}) = \frac{\text{Var}(X_i)}{n} = \frac{\sigma^2}{n}} \end{aligned}$$

■

We need to compute  $\text{Var}(X_i)$ . Observe that  $\text{Var}(X_i)$  are same for all: *Why?* because they are identical.

Also notice  $\frac{\text{Var}(X_i)}{n}$  decreases with  $n$ .

Observe that for all finite  $n$ ,  $\text{Var}(\bar{X})$  is not 0 unless  $\text{Var}(X_i) = 0$ !

*Note:*  $\text{Var}(X_i) = E[(X_i - E(X_i))^2] = E[(X_i - \mu)^2] = \frac{1}{N} \sum (x_i - \mu)^2 = \sigma^2$

So  $\text{Var}(X_i) = 0$  **iff** all  $X_i \equiv \mu$

**Lemma.**  $\bar{X}$  is consistent for  $\mu$ , i.e.  $\forall \delta > 0$ , the  $P(|\bar{X} - \mu| > \delta) \rightarrow 0$  as  $n \rightarrow \infty$

For this Lemma, we need to Prove Chebyshev's Inequality, which is

$$P(|Z - E(Z)| > \delta) \leq \frac{\text{Var}(Z)}{\delta^2}$$

Use this identity!

$$\begin{aligned} E[\bar{X}] &= \mu, & \text{Var}(\bar{X}) &= \frac{\sigma^2}{n} \\ P(|\bar{X} - E(\bar{X})| > \delta) &\leq \frac{\text{Var}(\bar{X})}{\delta^2} = \frac{\sigma^2}{n\delta^2} \rightarrow 0 \quad \text{as } n \rightarrow \infty \end{aligned}$$

## Lecture 3 (2018-09-10)

### Sampling without replacement

$Var(\bar{X})$  = when sampling without replacement

**Theorem.** *Sampling from finite population without replacement*

$$Var(\bar{X}) = \frac{\sigma^2}{n} \underbrace{\left[ \frac{N-n}{n-1} \right]}_{FPN} \quad (\text{finite population correction})$$

Points to Note - In sample without replacement,

- If  $n = N$ ,  $Var(\bar{X}) = 0$
- If  $n = 1$ ,  $Var(\bar{X}) = \frac{\sigma^2}{n} = \sigma^2$ , same as with replacement
- Check: for  $n > 1$ , how does  $\frac{N-n}{N-1}$  relate to 1? The  $Var(\bar{X})$  is always less without replacement

*Proof.* Start

①

$$Var(\bar{X}) = Var\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_i \sum_j Cov(X_i, X_j)$$

\left( \text{When sampling with replacement, } Cov(X\_i, X\_j) = 0 \text{ if } i \neq j \right)

In sampling without replacement, we cannot assert that  $Cov(X_i, X_j) = 0$  and we'll compute it explicitly.

$$\begin{aligned} \text{Recall} \quad Cov(X_i, X_j) &= E[X_i X_j] - \underbrace{E[X_i]E[X_j]}_{\mu^2} \\ \mu^2 \leftarrow \text{as identical but not independent} &= E[X_i X_j] - \mu^2 \end{aligned}$$

② To calculate  $E[X_i X_j]$ , let us list distinct values in population

**Example.**  $\left\{ \underbrace{5}_{x_1}, \underbrace{5}_{x_2}, \underbrace{8}_{x_3}, \underbrace{11}_{x_4}, \underbrace{8}_{x_5}, \underbrace{17}_{x_6}, \underbrace{9}_{x_7} \right\}$  Let  $n_l = \#$  of times  $\zeta_l$  appears in population.

Distinct Value	frequency
$\zeta_1 = 5$	$n_1 = 2$
$\zeta_2 = 8$	$n_2 = 2$
$\zeta_3 = 11$	$n_3 = 1$
$\zeta_4 = 17$	$n_4 = 1$
$\zeta_5 = 9$	$n_5 = 1$

$$P[X_i = 5] = \frac{2}{7} = \frac{n_1}{N} \quad (\text{i draws identical})$$

$$\Rightarrow P[X_i = \zeta_l] = \frac{n_l}{N}$$

$$n_1 + n_2 + \dots + n_m = \sum_{j=1}^m n_j = N$$

$$E[X_i X_j] = \sum_{k=1}^m \sum_{l=1}^m \zeta_k \zeta_l \underbrace{P[X_i = \zeta_k, X_j = \zeta_l]}_?$$

$$P[X_i = \zeta_k, X_j = \zeta_l] = \underbrace{P[X_j = \zeta_l | X_i = \zeta_k]}_{\textcircled{3}} \cdot \underbrace{P[X_i = \zeta_k]}_{= \frac{n_k}{N}}$$

③ Cases for Conditional probability

$$P[X_j = \zeta_l | X_i = \zeta_k] \stackrel{\text{cases}}{=} \begin{cases} \frac{n_l}{N-1} & l \neq k \rightarrow \text{numbers are diff.} \\ \frac{n_l - 1}{N-1} & l = k \rightarrow \text{numbers are same} \end{cases}$$

④ So we have

$$E[X_i X_j] = \sum_{k=1}^m \sum_{l=1}^m \zeta_k \zeta_l P[X_i = \zeta_k, X_j = \zeta_l]$$

$$E[X_i X_j] = \sum_{k=1}^m \sum_{l=1}^m \zeta_k \zeta_l P[X_j = \zeta_l | X_i = \zeta_k] \cdot P[X_i = \zeta_k]$$

$$= \sum_k \zeta_k P[X_i = \zeta_k] \zeta_k \left( \sum_l \zeta_l P[X_j = \zeta_l | X_i = \zeta_k] \right)$$

$$= \sum_k \zeta_k P[X_i = \zeta_k] \zeta_k \left( \sum_{l \neq k} \zeta_l P[X_j = \zeta_l | X_i = \zeta_k] + \zeta_k P[X_j = \zeta_k | X_i = \zeta_k] \right)$$

$$= \sum_k \zeta_k P[X_i = \zeta_k] \zeta_k \underbrace{\left( \sum_{l \neq k} \zeta_l \frac{n_l}{N-1} + \zeta_k \frac{n_k - 1}{N-1} \right)}_{\textcircled{5}}$$

⑤ When  $l \neq k$  and we want to remove all  $l$  terms

$$\begin{aligned} \sum_{l \neq k} \zeta_l \frac{n_l}{N-1} &= \frac{1}{N-1} \sum_{l \neq k} \zeta_l n_l \\ \left( \sum_l \zeta_l n_l = \tau = n\mu \right) &\quad \text{population total} \\ &= \frac{1}{N-1} (\tau - \zeta_k n_k) \end{aligned}$$

⑥ Now Back

$$\begin{aligned}
 E[X_i X_j] &= \sum_k \zeta_k \frac{n_k}{N} \left( \frac{1}{N-1} (\tau - \zeta_k n_k) + \zeta_k \frac{n_k - 1}{N-1} \right) \\
 &= \frac{1}{N(N-1)} \sum_k \zeta_k n_k [(\tau - \cancel{\zeta_k n_k}) + \cancel{\zeta_k n_k} - \zeta_k] \\
 &= \frac{1}{N(N-1)} \sum_k \zeta_k n_k [\tau - \zeta_k] \\
 &= \frac{1}{N(N-1)} \left( \sum_k \zeta_k n_k \tau - \sum_k \zeta_k^2 n_k \right) \\
 &= \frac{1}{N(N-1)} \left[ \tau^2 - \sum_k \zeta_k^2 n_k \right]
 \end{aligned}$$

⑦ What is  $\sum_k (\zeta_k)^2 \frac{n_k}{N}$ ? Second moment  $E[X_i^2]$        $E[X_i^2] = \sigma^2 + \mu^2$

$$\begin{aligned}
 E[X_i^2] &= \sigma^2 + \mu^2 & \frac{\tau^2}{N} &= N \mu^2 \text{ as } \mu = \frac{\tau}{N} \\
 E[X_i X_j] &\Rightarrow \frac{1}{N-1} \left[ N \mu^2 - (\sigma^2 + \mu^2) \right] \\
 &= \frac{1}{N-1} [(N-1) \mu^2 - \sigma^2] = \mu^2 - \frac{\sigma^2}{N-1}
 \end{aligned}$$

$$\begin{aligned}
 \text{So } \text{Cov}(X_i, X_j) &= \mu^2 - \frac{\sigma^2}{N-1} - \mu^2 \\
 &= -\frac{\sigma^2}{N-1}
 \end{aligned}
 \tag{Cov < 0}$$

$$\text{So } \text{Cov}(X_i, X_j) = \text{Var}(X_i) = \sigma^2$$

⑧ Putting it all together

$$\begin{aligned}
 \text{Var}(\bar{X}) &= \frac{1}{n^2} \left( \sum_{i \neq j} \text{Cov}(X_i, X_j) + \sum_{i=1}^n \text{Var}(X_i) \right) \\
 &= \frac{1}{n^2} \left( \sum_{i \neq j} -\frac{\sigma^2}{N-1} + n \sigma^2 \right) \\
 &= \frac{1}{n^2} \left( \frac{-n(n-1)\sigma^2}{N-1} + \frac{\sigma^2}{n} \right) \\
 &= \frac{\sigma^2}{n} \left( 1 - \frac{n-1}{N-1} \right) \\
 &= \frac{\sigma^2}{n} \left( \frac{N-n}{N-1} \right)
 \end{aligned}$$





## Lecture 4 (2018-09-12)

- Binary data- special case.
- Approximate distance of  $\bar{X}$  when  $n$  is large but  $n \ll N$
- Estimating population Variance
- Bivariate data

Recall that population is dichotomous or binary then  $x_i = \begin{cases} 1 \\ 0 \end{cases}$

Moreover if we consider  $x_i = 1$  as a "success" and  $x_i = 0$  as a "failure", then

$$\mu = \frac{\sum_{i=1}^N X_i}{N} = \frac{\# \text{ of successes in population}}{\text{population size}} = p \quad (\text{pop}^n \text{ proportion of success})$$

$$\text{Now, } \sigma^2 = \underbrace{\frac{\sum_{i=1}^N X_i}{N}}_{\mu} - \mu^2 = p - p^2 = p(1 - p) = pq$$

$$\mu \text{ as } 1 \Rightarrow 1^2 = 1$$

$$0 \Rightarrow 0^2 = 0$$

Recall that if  $Y \sim \text{Bernoulli}(p)$ ,  $Y_i = \begin{cases} 1 & \text{w/ prob } p \\ 0 & \text{w/ prob } 1 - p \end{cases}$

$$E[Y] = p$$

$$\text{Var}(Y) = p(1 - p)$$

Last few weeks involved an analysis of  $\bar{X}$ ,  $E(\bar{X})$ ,  $\text{Var}(\bar{X})$ . Could also ask: How is  $\bar{X}$  distributed if  $n$  is large.

### Confidence Intervals - Sampling W.R.

If sampling **with replacement**, where  $X_1, \dots, X_n$  denotes sample, we know  $X_i$ 's are *i.i.d.* Hence when  $n$  is large, by CLT  $\bar{X}$  has an approximately normal distribution.

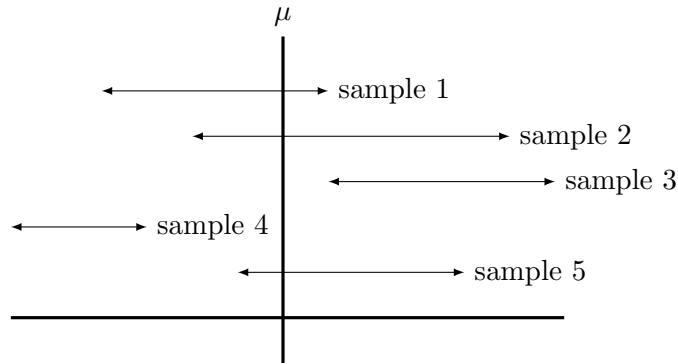
$$P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq x\right) \rightarrow \Phi(x) \quad \text{as } n \rightarrow \infty$$

When sampling with replacement, we can use this to obtain confidence intervals for  $\mu$ : Let  $\alpha \in (0, 1)$  be given.

Let  $Z_\alpha \in \mathbb{R}$  such that  $P(Z > Z_\alpha) = \alpha$  where  $Z \sim N(0, 1)$

By the Central Limit Theorem, for  $n$  large (sampling w/replacement)

$$\begin{aligned} &= P\left(-Z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq Z_{\alpha/2}\right) \\ &= P\left(\underbrace{\bar{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}_{\text{Random}} \leq \mu \leq \underbrace{\bar{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}_{\text{Random}}\right) \\ &\quad \text{Var}(\bar{X}) = 0 \quad \text{Never happens} \end{aligned}$$



In repeated sampling, approx  $(1 - \alpha)$  of intervals contain  $\mu$ , and  $(\alpha)$  frac will not.

We say  $\boxed{\bar{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}}$  is  $100(1 - \alpha)\%$  2-sided confidence interval for  $\mu$

**Problem:** This interval involved  $\sigma$  which is unknown. Observe that if  $n$  is large, then  $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$  is still approx  $N(0, 1)$  in distribution where (no population parameters)

$$\boxed{s^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2} \quad (\text{sample variance})$$

So we obtain

$$\boxed{\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}} \quad \text{as a } 100(1 - \alpha) \text{ CI for } \mu$$

In the dichotomous case,

$$\bar{X} = \frac{\# \text{ of the succession sample}}{\text{sample size}} = \hat{p}$$

$$100(1 - \alpha)\% \text{ CI for } p : \hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

## Confidence Intervals - Sampling W.o.R.

Recall now what happens when sampling **without replacement**

Here,  $X_1, X_2, \dots, X_n$  remain identically distributed, but not independent

We surmised, that if  $n \ll N$ ,  $X_i$  &  $X_j$  have an "approximate independence"

**Example 1.** Let population consist of 1000 elements. In this case:

$$\left. \begin{array}{l} \text{blue} - \textcircled{1} - 200, \quad \text{red} - \textcircled{2} - 300, \quad \text{green} - \textcircled{1} - 500 \\ P(X_1 = \textcircled{3}) = \frac{1}{2} \\ P(X_2 = \textcircled{3} | X_1 = \textcircled{3}) = \frac{499}{999} \end{array} \right\} \text{not independent, but have approximate independence.}$$

In short,  $n \ll N$ , each successive draw does not alter probabilities that much, precisely b/c removal is only of a sample # of population elements.

So if  $n \ll N$ , then even in sampling W.O.R,  $X_i$ 's retain an approximate independence. Further if  $n$  is "large" and small relative to  $N$ , (note delicate point!) then  $\bar{X}$  will still have an approx Normal distribution.

$$\frac{\bar{X} - \mu}{\sqrt{\frac{\sigma^2}{n} \left( \frac{N-n}{N-1} \right)}} \sim N(0, 1)$$

Observe  $\sigma^2$  is still unknown. We'd like to consider estimators for  $\sigma^2$

### Estimator for variance W.o.R

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Try to understand  $E[\hat{\sigma}^2]$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i^2 - 2x_i\bar{X} + \bar{X}^2)$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n x_i^2}{n} - 2\bar{X}\bar{X} + \bar{X}^2 = \frac{\sum_{i=1}^n x_i^2}{n} - \bar{X}^2$$

$$E[\hat{\sigma}^2] = E\left[\frac{\sum_{i=1}^n x_i^2}{n}\right] - E[\bar{X}^2] \quad \text{can't get } E[\bar{X}^2] \text{ from } Var(\bar{X})$$

$$Var(\bar{X}) = E[\bar{X}^2] - (E[\bar{X}])^2$$

$$\underbrace{Var(\bar{X})}_{\text{computed}} + \mu^2 = E[\bar{X}^2]$$

$$E\left[\frac{1}{n} \sum_{i=1}^n x_i^2\right] = \frac{1}{n} \sum_{i=1}^n E[X_i^2] = \sigma^2 + \mu^2$$

Combining, we get:

$$E[\hat{\sigma}^2] = \sigma^2 + \mu^2 - (Var(\bar{X}) + \mu^2)$$

$$= \sigma^2 - \left[ \frac{\sigma^2}{n} \left( \frac{N-n}{N-1} \right) \right]$$

The estimator is biased, but

$$= \sigma^2 \left( \underbrace{1 - \frac{N-n}{(n)(N-1)}}_{\text{constant, } c} \right)$$

$$= C\sigma^2$$

and thus  $\frac{\hat{\sigma}^2}{C}$  is an unbiased estimator.