AMS 553.430 - Introduction to Statistics

Lectures by Dwijavanti P Athreya
Notes by Kaushik Srinivasan

Johns Hopkins University
Fall 2018

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 Ti*k*Z 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.

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

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

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 $1, 2, \cdots, n$. n boxes labelled $1, 2, \cdots, n$. 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

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

8. Partition of integers $\longrightarrow$ n: sum of integer, $\quad r$: number of partitions

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

Basics of Probability

<u>Axioms</u>

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

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

$\Rightarrow P(A) = 1 - P(A^c)$
$P(A) = \dfrac{|A|}{|\Omega|}$

Conditional Probability

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

Law of Total Probability

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

Bayes Rule

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

Independent events

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

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

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)

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

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

Discrete Distributions

1. Bernoulli(p)
2. Binomial(n,p)
3. Poisson (λ)
4. Geometric(p)
5. Negative Binomial(n,p)
6. Hypergeometric(N,M,n)

Bernoulli Distribution

X is a random variable with Bernoulli(p) distribution

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

Binomial Distribution

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

$$X \sim Binomial(n,p)$$
$$\text{Support}: x \in \{0,1,\cdots n\}$$
$$n: \text{sample size} \qquad p: \text{probability of success}$$
$$P(X=k)=\binom{n}{k}p^k(1-p)^{(n-k)}$$
$$E[X]=np \qquad Var(X)=np(1-p)$$

- 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 $\dfrac{P(X=x)}{P(X=x-1)}$ going below 1.

Poisson Distribution

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

- Approximations
  - if n is large $\Longrightarrow$ Normal(λ, λ)
- Sums of Poisson
  Let $X \sim Po(\lambda) \quad Y \sim Po(\mu) \quad \Longrightarrow \quad X + Y \sim Po(\mu + \lambda)$

Negative Binomial

$$X \backsim NB(r, p)$$
$$\text{Support} : x = \{r, r+1, \ldots\}$$
$$r = \text{the rth success}$$
$$p = \text{probability of success}$$
$$P(X=k) = \binom{k+r-1}{k} \cdot (1-p)^r \cdot p^k$$

A sum of i.i.d Geometric(p) R.V.
■ a^{th} head before b^{th} tail

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

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

$$\begin{cases} = P[\text{5th tail occurs before or on the 14th flip}] \\ = P[\text{Neg Binomial}(5,q) = 5,6,7,\cdots,14] \\ = \sum\limits_{x=5}^{14} \binom{x-1}{4} q^5 p^{x-5} \end{cases} \quad \text{(or)} \quad \begin{cases} = P[\text{at least 5 tails in 14 flips}] \\ = P[binom(14,q) = 5,6,7,\cdots,14] \\ = \sum\limits_{x=5}^{14} \binom{14}{x} q^x p^{14-x} \end{cases}$$

Geometric Distribution

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

Example. ■ Coupon Question

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

Question: E[draws to get 10 different coupons]?

Answer:

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

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

$$\text{Hence, } E[X_1] = 1$$
$$E[X_2] = \frac{1}{p_2} = \frac{1}{\frac{N-1}{N}} = \frac{N}{N-1}$$
$$E[X_3] = \frac{1}{p_3} = \frac{1}{\frac{N-2}{N}} = \frac{N}{N-2}$$
$$\vdots$$
$$E[X_{10}] = \frac{1}{p_{10}} = \frac{1}{\frac{N-9}{N}} = \frac{N}{N-9}$$

So, $\quad E[X] = E[X_1] + E[X_2] + \cdots + E[X_{10}] = E[\sum\limits_{i=1}^{10} X_i] = 1 + \frac{N}{N-1} + \frac{N}{N-2} + \cdots + \frac{N}{N-9}$

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

Question: E[# different types of coupons]?

Answer:

$$X = I_1 + I_2 + \cdots + I_N$$
$$I_i \begin{cases} 1 & \text{if we have this type of coupon} \\ 0 & \text{o/w} \end{cases}$$

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

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

Hypergeometric Distribution

$$X \sim Hyp(N, M, n)$$

$$N \in \{0, 1, 2, \ldots\} \quad M \in \{0, 1, \ldots, N\} \quad n \in \{0, 1, \ldots, N\}$$

$$\text{Support} : k \in \{\max(0, n + M - N), min(n, M)\}$$

N is the population size $\quad$ K is the no. of success states in the population

n is the no. of draws (i.e. quantity drawn in each trial)

k is the no. of observed successes

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{M-k-1}}{\binom{N}{n}}$$

Continuous Distributions

Uniform Distribution

$$X \sim Unif(a, b)$$

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & o/w \end{cases}$$

$$E[X] = \frac{a+b}{2} \qquad\qquad Var(X) = \frac{(b-a)^2}{12}$$

Normal Distribution

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

- Sums and differences of Normal R.V.

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

  - 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 μ. But density of $X_1 + X_2$ does.
  - Key idea is used in Data Reduction

Exponential distribution

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

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

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

Gamma Distribution

$$X \sim Gamma(\alpha, \beta)$$
$$\text{Support:} \quad x \in [0, \infty)$$
$$F_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$$
$$E[X] = \frac{\alpha}{\beta} \qquad Var(X) = \frac{\alpha}{\beta^2}$$
$$\textbf{Gamma Function:} \quad \Gamma(z) = (z-1)! = \int_0^\infty x^{z-1} e^{-x} dx$$
$$\Gamma(n) = (n-1)!$$
$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

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

Beta Distribution

$$X \sim Beta(\alpha, \beta)$$
$$\text{Support:} \quad x \in [0, 1]$$
$$f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$$
$$E[X] = \frac{\alpha}{\alpha + \beta} \qquad Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

- **Gamma to Beta**

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

Chi-Square

$$\textbf{Chi-Square:} \ \chi_n^2 \text{is Chi-square with degrees of Freedom } n$$
$$\chi_n^2 = Z_1^2 + Z_2^2 + \cdots + Z_n^2 \qquad \text{where } Z_i \backsim \text{standard normal.} Z_i \backsim Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$$
$$\Rightarrow \chi_n^2 = n \text{ i.i.d. } Z_i \backsim Gamma\left(\frac{1}{2}, \frac{1}{2}\right)$$

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

CDF in General

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

$$\begin{aligned} &= \sum_{x \leq t} P(X = x) && \text{discrete} \\ &= \int_{-\infty}^{t} f(x)dx && \text{continuous} \end{aligned}$$

- **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

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

Joint Distribution

Discrete

$$P_{X,Y}(x,y) = P(X = x, Y = y)$$
$$\text{Indep} \Rightarrow P_X(x)P_Y(y)$$

Continuous

$$\begin{aligned} F_{X,Y}(x,y) &= f_X(x)f_Y(y) \\ &= \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y) \end{aligned}$$

- **Marginal Density/PMF**:

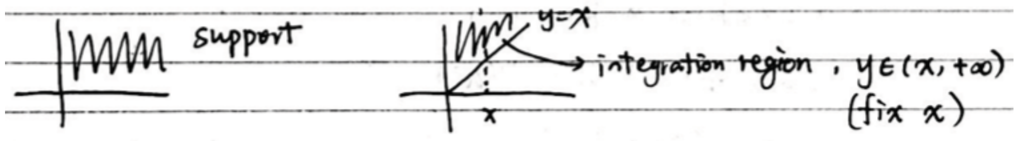
Continuous: $f_X(x) = \int_x f_{X,Y}(x,y)dy \quad and \quad f_Y(y) = \int_y f_{X,Y}(x,y)dx$

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

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

- 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

$$\textbf{Continuous:}\qquad f(x,y)=f_X(x)f_Y(y)$$
$$\textbf{Discrete:}\qquad P(X=x,Y=y)=P(X=x)P(Y=y)$$

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

$$\textbf{Discrete:}\qquad P_{X+Y}(a)=\sum_y P_X(a-y)P_Y(y)=\sum_x P_X(x)P_Y(a-x)$$
$$\textbf{Continuous:}\qquad f_{X+Y}(a)=\int_y f_X(a-y)f_Y(y)dy=\int_y f_X(x)f_Y(a-x)dx$$

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

- **Density Transformation:**

$$\begin{tikzcd}
F_{X,Y} \arrow[r, "\text{diff}"] & f_{X,Y} \arrow[d, "\text{substitute}"] \arrow[r, dotted, "\text{algebra}"] & f_{g(X,Y)} \arrow[ddl, dotted, bend left, "\text{integrate}"] \\
 & F_{g(X,Y)} \arrow[d, "\frac{d}{dt}"] & \\
 & f_{g(X,Y)} &
\end{tikzcd}$$

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

$Y_1=\frac{X_1}{X_2}$ choose Y_2 to make work easier

$$\begin{tikzcd}
f_{X_1} \arrow[dr] & & f_{X_2} \arrow[dl] \\
 & f_{X_1,X_2} \arrow[d, "\text{dens. transf.}"] & \\
 & f_{\frac{X_1}{X_2},???} \arrow[d] & \\
 & f_{\frac{X_1}{X_2}} &
\end{tikzcd}$$

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: $Y = g(x)$ $\boxed{f_Y(y) = f_X(x(y)) \cdot |(x^{-1})'(y)|}$
- **2-dim:** Joint Density:

$$(X, Y) \to (U, V) \qquad U = h_1(X, Y) \qquad V = h_2(X, Y)$$
$$f_{U,V}(u, v) = f_{X,Y}(x(u, v), y(u, v)) \cdot |J|$$
$$\text{where} \quad J = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} \quad \text{determinant}$$

- if $Z = X + Y$ (2-dim $\to$ 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

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

Continuous
$$f_{X|Y=y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$
$$\Rightarrow f_X(x) = \int_y f(x, y) dy = \int_y f_{X|Y=y}(x|y) \cdot f_Y(y) dy$$
$$F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(x|y) dx$$

$$\begin{tikzcd}
f_{X|Y} \arrow[d] & f_Y \arrow[dl, bend left, "x"] \\
f_{X,Y} \arrow[d, "\text{integrate}"] \arrow[dr] & \\
f_X \arrow[r, "\text{divide}"] & f_{Y|X}
\end{tikzcd}
\qquad
\begin{tikzcd}
f_{X,Y} \arrow[d, "\text{integrate}"] \arrow[dr, bend left, "\text{divide}"] & \\
f_X \arrow[r, "\text{divide}"] & f_{X|Y} \arrow[d, "\text{integrate } dx"] \\
 & F_{X|Y}
\end{tikzcd}$$

Conditional Expectation

$$E[X|Y=y] = \sum_x xP(X=x|Y=y)$$
$$E[X|Y=y] = \int_x xf(x|y)dx$$
$$E[X|Y] : \text{compute } E[X|Y=y] \text{ first, replace } y \text{ with } Y$$

- **Properties:**
  - $E[aU+bV|Y=y] = aE[U|Y=y] + bE[V|Y=y]$ $\boxed{LOTUS}$
  - If $g(Y) = X$ then $E[X|Y=y] = X$
  - If $\underline{X \text{ and } Y}$ are independent, then $E[X|Y=y] = E[X]$

Conditional Variance

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

Ordered Statistics

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

$$\begin{aligned} F_{\max(X_i)}(t) &= P(\max X_i \leq t) = P(X_1 \leq t) \cdot P(X_2 \leq t) \cdots P(X_n \leq t) \\ &= [F_X(t)]^n \qquad \boxed{f_{\max X_i}(t) = nF(t)^{n-1} f_X(t)} \\ F_{\min(X_i)}(t) &= 1 - P(\min x_i \geq t) = 1 - P(X_1 \geq t) \cdot P(X_2 \geq t) \cdots P(X_n \geq t) \\ &= 1 - [1 - F_X(t)]^n \qquad \boxed{f_{\min X_i}(t) = n[1 - F(t)]^{n-1} f_X(t)} \end{aligned}$$

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 $U_1, U_2, \ldots, U_N \sim i.i.d.$ Uniform$(0,1)$ and let $1 \leq j \leq N$
$U_{(j)} = $ jth smallest in $U_{(1)}, U_{(2)}, \ldots, U_{(N)}$ (ordered statistics). Then,

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

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) $= nVar(X_i) + n(n-1)Cov(X_i, X_j) \qquad i \neq j$

   $$\boxed{\text{Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)}}$$

2. **Covariance:**

$$\begin{aligned}
Cov(X, Y) &= E[XY] - E[X]E[Y] \\
Cov(X, c) &= 0 \qquad c \; is \; a \; constant \\
Cov(X + Y, Z) &= Cov(X, Z) + Cov(Y, Z) \\
Cov(cX, dZ) &= cd \cdot Cov(X, Z) \\
Cov(aX + b, cY + d) &= ac \cdot Cov(X, Y) \qquad a, b, c, d \text{ are constants} \\
Cov(X, Y) &= 0 \qquad \text{If } X \perp Y \text{ (independent)}
\end{aligned}$$

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:

$$= E\left[\sum_{j=0}^{\infty} \frac{(tX)^j}{j!}\right]$$

$$= E\left[\sum_{j=0}^{\infty} \left(\frac{X^j}{j!} \cdot t^j\right)\right]$$

$$\boxed{M_X^{(n)}(0) = E[X^n]}$$

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

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

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

$$Y_a = \begin{cases} 0 & \text{if } X < a \\ a & \text{if } X \geq a \end{cases}$$

$$0 \leq Y_a \leq X \implies \underbrace{E[Y_a]}_{a \cdot P(X \geq a)} \leq E[X]$$

$$E[Y_a] = 0 \cdot P(Y_a < a) + a \cdot P(X \geq a)$$

$$E[Y_a] = a \cdot P(X \geq a) \leq E[X] \implies \boxed{P(X \geq a) \leq \frac{E(X)}{a}}$$

■

Chebyshev's Inequality

For any random variable Y with mean μ_y and variance σ_y^2

$$P(|Y - \mu)y| \geq c) \leq \frac{\sigma_y^2}{c^2} \qquad \text{(for \underline{any} } c > 0)$$

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 $X_1, X_2, \cdots$ are *i.i.d.* with a mean μ

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

Strong Law of Large Numbers

$$X \xrightarrow{p} \mu_X \qquad \text{as } n \to \infty$$
$$Pr\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1$$

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 $\{x_1, x_2, \cdots, x_N\}$
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: τ and μ are population parameters, their computation depends on all the population data.

Question. How to estimate τ and μ based on a sample of observation from this population?

Classical Answer: Choose a "random" sample of objects and associated measurements denoted $\{x_1, x_2, \cdots, x_n\}$. *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.

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

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 $1, \ldots, N$ objects we care about.
Loss of Independence when choosing sampling method is important.

Lecture 2 (2018-09-05)

Finite Population sampling – without i 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 <u>without replacement</u>

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

Sample here is the same as population as n=N

Same thing <u>with replacement</u>

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

Typically N is large and $n << N$

Recall population parameters

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

Next, σ^2 (population variance)

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

Alternatively, we can also express σ^2 as

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

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

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 μ.
Note: If the sample is random ($X_1, \ldots, X_n$ 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 $X_1 = \{x_1, \ldots, x_N\}$

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

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

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

In sampling $\boxed{\text{without replacement}}$ X_i & X_j are still identically distributed, but they are not independent.
In sampling $\boxed{\text{with replacement}}$, X_i & X_j are *i.i.d.*
Note that whether or not $X_1, \cdots, X_n$ 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.

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

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

Let's address 3

Sampling with replacement.

Here $X_1, \cdots, X_n$ are *i.i.d.*. In general, X_is are R.V. and a_is are constants

$$Var\big(\sum a_i X_i\big) = \sum_i \sum_j a_i a_j cov(X_i, X_j)$$

If $X_1, \cdots, X_N$ are independent, $\underset{i \neq j}{Cov(X_i, X_j)} = 0$! Hence

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

$$\boxed{Var(\bar{X}) = \frac{Var(X_i)}{n}}$$

We need to compute $Var(X_i)$. Observe that $Var(X_i)$ are same for all: *Why?* because they are identical.
Also notice $\frac{Var(X_i)}{n}$ decreases with n.
Observe that for all finite n, $Var(\bar{X})$ is not 0 unless $Var(X_i) = 0$!
Note: $Var(X_i) = E[(X_i - E(X_i))^2] = E[(X_i - \mu^2)] = \frac{1}{N}\sum (x_i - \mu)^2 = \sigma^2$
<u>So</u> $Var(X_i) = 0$ **iff** all $X_i \equiv \mu$

Lemma. *bX is <u>consistent</u> for μ, i.e. $\forall \delta > 0$, the $P(|\bar{X} - \mu| > \delta) \longrightarrow 0$ as $n \to \infty$*

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

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

Use this identity!

$$E[\bar{X}] = \mu, \qquad\qquad Var(\bar{X}) = \frac{\sigma^2}{n}$$

$$P(|\bar{X} - E(\bar{X})| > \delta) \leq \frac{Var(\bar{X})}{\delta^2} = \frac{\sigma^2}{n\delta^2} \to 0 \qquad \text{as } n \to \infty$$

Sampling without replacement