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 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.

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
3. Birthday Problem
4. Matching Problem (inclusion-exclusion principle)
5. n balls going into m boxes (all are distinguishable)
6. Multinomial Coefficients e.g. assign A, B, C, D, to different students $\rightarrow$ anagram problem
7. Pairing Problem

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

8. Partition of integers $\rightarrow \binom{n}{n+k-1}$ where n is the sum of integer and k is the number of partitions

Basics of Probability

Axioms

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

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

Bayes Rule

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

Law of Total Probability

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

Part 2 - Discrete and Continuous Random Variables

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

Special notes on Discrete Distributions

Binomial Distribution

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

- Approximation method $\Rightarrow$ if n is large, p very small and $np < 10$.
  - use Poisson (np), otherwise preferably $p \approx \frac{1}{2}$
  - use Normal ($np, np(1-p)$)
- 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.

Negative Binomial

$$X \backsim NB(r,p) \quad x = r, r+1, \cdots$$
$$r = ...$$
$$p = \text{probability of success}$$

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

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

Geometric Distribution

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

Variation B: Same setting, now you draw 10 times.
Question: E[# different types of coupons]?
Answer:

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

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

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

Special Notes on some Continuous Distributions

Normal Distribution

$$\textbf{Normal: } X \backsim N(\mu, \sigma^2) \Rightarrow Z = \frac{X-\mu}{\sigma} \backsim N(0,1) \text{ with CDF } P(Z \le z) = \Phi(z)$$
$$\Phi(-x) = 1 - \Phi(x)$$

Chi-Square

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

Exponential distribution

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 $X_i \backsim_{\text{i.i.d.}} exp(\lambda) \Rightarrow exp(n\lambda)$

CDF in General

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

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

- **Discrete:** "Left open, right closed" $\Rightarrow$ if you flip the sign (from $<$ to $\le$) 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)

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

Density Transformation

- Use CDF: Computer $P(Y \leq y) = P(g(x) = y)$
- **1-dim:** If Y is monotonically increasing or decreasing: $Y = g(x)$ $\boxed{f_Y(y) = f_X(x(y)) \cdot |x'(y)|}$
- **2-dim:** Joint Density:

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

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

Joint Disrtibtuion

Discrete

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

Continuous

$$F_{X,Y}(x, y) = f_X(x)f_Y(y) = \frac{\partial}{\partial x \partial y} F_{X,Y}(x, y)$$

- **Marginal Density/PMF**:

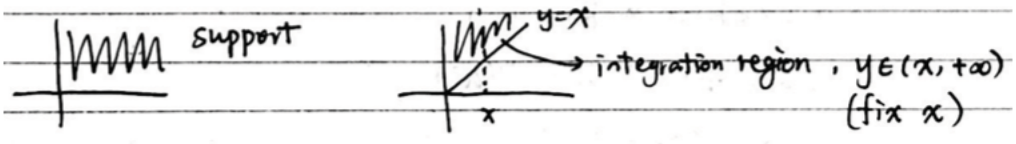
Continuous: $f_X(x) = \int_x f_{X,Y}(x, y)dy$ *and* $f_Y(y) = \int_y f_{X,Y}(x, y)dx$

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

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

- Use joint pdf to compute probability

e.g. $P(X<Y)=\int\limits_{0}^{\infty}\int\limits_{x}^{\infty} f(x,y)dydx$ 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

Conditional distribution

$$\textbf{Discrete} \qquad P_{X|Y=y}(x|y)=\frac{P_{X,Y}(x,y)}{P_Y(y)}=\frac{P(X=x,Y=y)}{P(Y=y)}$$
$$\Rightarrow \sum_y P_{X,Y}(x,y)=\sum_y P_{X|Y=y}(x|y)\cdot P_Y(y)$$
$$\textbf{Continuous} \qquad f_{X|Y=y}(x|y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}$$
$$\Rightarrow f_X(x)=\int_y f(x,y)dy=\int_y f_{X|Y=y}(x|y)\cdot f_Y(y)dy$$

Conditional Expectation

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

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

Ordered Statistics

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

$$F_{\max(X_i)}(t)=P(\max X_i\le t)=P(X_1\le t)\cdot P(X_2\le t)\cdots P(X_n\le t)$$
$$=[F_X(t)]^n \qquad \boxed{f_{\max X_i}(t)=nF(t)^{n-1}f_X(t)}$$
$$F_{\min(X_i)}(t)=1-P(\min x_i\ge t)=1-P(X_1\ge t)\cdot P(X_2\ge t)\cdots P(X_n\ge t)$$

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

General: j-th order statistic

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

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) $= nVar(X_i) + n(n-1)Cov(X_i, X_j) \qquad i \neq j$
2. **Covariance:**

$$Cov(X, Y) = E[XY] - E[X]E[Y]$$
$$Cov(X, c) = 0 \qquad c \text{ is a constant}$$
$$Cov(X + Y, Z) = Cov(X, Z) + Cov(Y, Z)$$
$$Cov(cX, dZ) = cd \cdot Cov(X, Z)$$

3. **Correlation Coefficient:**

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

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.