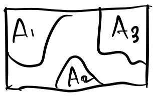
Summary of Week 1

Monday

- Basic Defs: Sample space / outcomes / Events / Prob. model



$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3)$

- Random variables: Variables take on values defined by Prob. model (Distribution)

Wednesday

- Conditioning and independence: How to think about relationships between r.v.
- Joint Prob and Marginalization: Prob dist. w/ multiple variables

Friday

- Colab notebooks / Python Basics
- Monte Carlo Simulation
- translating probability statements to code

This week

- Expectation variance
- Binomial, Normal distribution (Sums)
- CLT, LLN

Expectation (Ch. 3.1 ER) Skim 3.2

Sample average $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ where X_i iid = independent and indentically distributed

If $X_i \sim \text{Bernoulli}(q)$

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \approx \frac{\#\{X_i = 0\}}{N} \times 0 + \frac{\#\{X_i = 1\}}{N}$$

$$= (1-q) \times 0 + q \times 1 = q$$

In general

$$\bar{X} = \sum_{x \in S_x} x \frac{\#\{X_i = x\}}{N} \approx \sum_{x \in S_x} x P(X = x)$$

This motivates the definition of expectation

Expected Value (Def 3.1.1/3.1.2)

$$\mathbb{E}[X] = \sum_{x \in S_X} x P(X = x)$$

Examples

$$X \sim \text{Bernoulli}(q) \implies \mathbb{E}[X] = q$$

(Note in 3.1.8 they use a slightly different definition of Geometric where — see below)

$$X \sim \text{Geometric}(q) \implies \mathbb{E}[X] = 1/q$$

does this make intuitive sense? Can you test with simulations?

Skip EX. 3.1.9 – 3.1.13

Properties of Expectation

1) Expectation of function (theorem 3.1.1)

If X is r.v. and $g: S_X \to \mathbb{R}$ is a function, then

$$\mathbb{E}[g(X)] = \sum_{x \in S_X} g(x) P(X = x)$$

Key point here is that we don't need to determine probabilities of $g(X)$ to determine $\mathbb{E}[g(X)]$

2) Expectation is linear (Theorem 3.1.2)

X, Y r.v. then $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

a, b constants $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$

Example (two defs of geometric)

Let $X \sim \text{Geometric}(q)$. In the Book they define

$Y = X - 1 = \#$ of tails before the first heads

$$\Rightarrow \mathbb{E}[Y] = \mathbb{E}[X] - 1 = \frac{1}{q} + 1 = \frac{1-q}{q}$$

Theorem 3.13 in ER

Let X and Y be two independent r.v.

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

Note:

$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ does not imply X, Y independent

Conditional Expectations Ch 3.5 ER

Conditional Expectation (Def 3.5.2)

for two r.v. X and Y, the conditional expectation given $Y=y$ is

$$\mathbb{E}[X \mid Y=y] = \sum_{x \in S_X} x \, P(X \mid Y=y)$$

$$= \sum_{x \in S_X} x \, \frac{P(X=x, Y=y)}{P(Y=y)}$$

Note $h(y) = \mathbb{E}[X \mid Y=y]$ is a function of y and we can evaluate it at a random value of y to obtain a random variable $h(Y)$

— see Theorem 3.5.2 and HW

Example

		X: 0	1	2
Y	0	0.1	0.2	0.15
	1	0.35	0.05	0.15

look at avg. within this column

$$\mathbb{E}[Y \mid X=1] = 0 \times \frac{P(X=1, Y=0)}{P(X=1)} + 1 \times \frac{P(X=1, Y=1)}{P(X=1)}$$

$$= 1 \, \frac{0.05}{0.2 + 0.05} = \frac{0.05}{0.25} = 0.2$$

Binomial Distribution (EX. 2.3.3) watch 1Brown 3Blue!

+ $X_i \sim \text{Bernoulli}(q)$ $i = 1, 2, 3, \ldots, N$

and assume independent!

note: two parameters

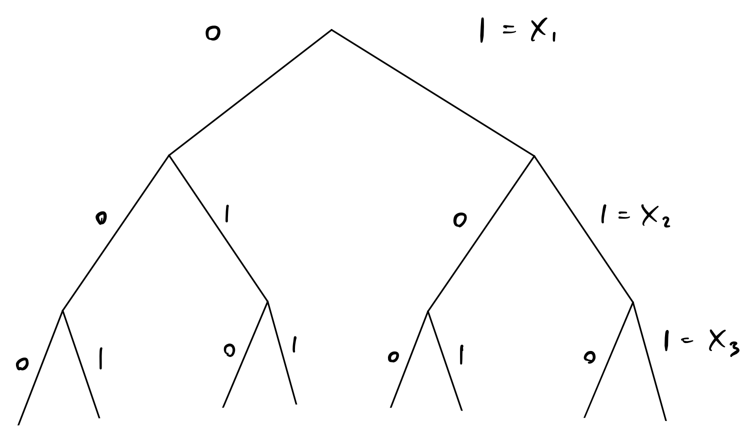
Let $S = \sum_{i=1}^{N} X_i$ then $S \sim \text{Binomial}(q, N)$

S is Model of many things

- Votes in election
- S/N is Sample mean of Bernoulli trials
- Successful free throws
- responses on Y/N survey
- Positive tests in clinical trial

By linearity $\mathbb{E}[S] = \sum_{i=1}^{N} \mathbb{E}[X_i] = Nq$

What about distribution? Take $N=3$



$S = \quad 0 \quad 1 \quad 1 \quad 2 \quad 1 \quad 2 \quad 2 \quad 3$

$\text{Prob} = (1-q)^3 \quad (1-q)^2 q \quad (1-q)^2 q \quad (1-q) q^2 \quad (1-q)^2 q \quad (1-q) q^2 \quad (1-q) q^2 \quad q^3$

By theorem 1.5.1,

# of sequences w/ 2 1s

chance to get any particular sequence of flips w/ 2 1s

$$P(S=2) = \underbrace{2}_{\text{\# of sequences w/ 2 1s}} \times \underbrace{(1-q)\,q^2}_{\text{chance to get any particular sequence of flips w/ 2 1s}}$$

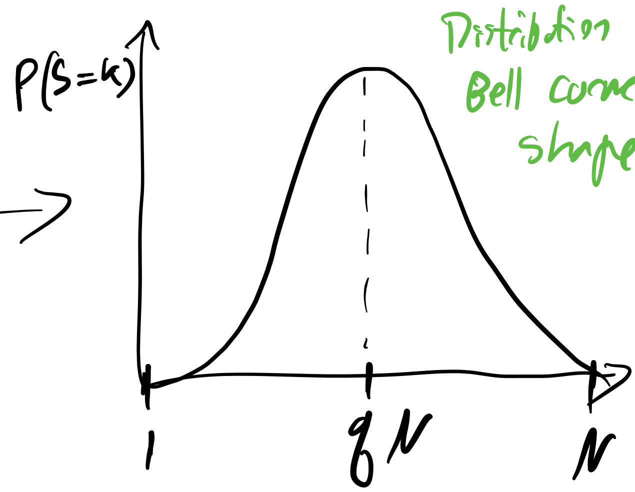
<u>In general</u>

$$P(S=k) = \underbrace{\binom{N}{k}}_{\text{\# sequences w/ k heads (don't need to know formula)}} \underbrace{(1-q)^{N-k} q^k}_{\text{chance to get sequence w/ k heads}}$$

(see Ch. 1 for details, don't need to know)

<u>Key point:</u>

many more ways to get $k \approx N/2$ heads than $k \approx 0$ or $k \approx N$ →



Distribution has Bell curve shape

useful to quantify how "spread out" distribution is

Variance and Covariance (ch 3.3 ER)

To measure spread of distribution we look at how much, on avg., it deviates from the mean

Variance (Def 3.1.1)

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

Example $X \sim \mathrm{Bernoulli}(q)$

$$\Rightarrow \mathbb{E}[(X-q)^2] = \mathbb{E}[X^2] - 2\mathbb{E}[X]q + q^2$$

$$= 0\cdot(1-q) + 1\cdot q^2 - 2q + q = q^2 - q = q(1-q)$$

$\longrightarrow$ X is least "variable" when $q \approx 0$ or $q \approx 1$

Important identity

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2$$

$$= \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Example

$$S \sim \text{Binomial}(q, N) \longrightarrow S = \sum_{i=1}^{N} X_i$$

$$Var(S) = \mathbb{E}[S^2] - \mathbb{E}[S]^2$$

$$= \mathbb{E}\left[\left(\sum_{i=1}^{N} X_i\right)^2\right] - (Nq)^2$$

don't need to know derivation

$$\left(\sum_{i=1}^{N} X_i\right)^2 = (X_1 + \cdots X_N)(X_1 + \cdots X_N)$$

$$= X_1X_1 + X_1X_2 + \cdots + X_1X_N$$
$$+ X_2X_1 + \cdots \quad X_2X_N + \cdots X_NX_N$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} X_{ij}$$

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

$$= Nq + N(N-1)q^2 = Nq + N^2q^2 - Nq^2$$

$$Var(S) = Nq + \cancel{N^2q^2} - Nq^2 - \cancel{N^2q^2} = \underbrace{Nq(1-q)}_{\text{need to know!}}$$

<u>Notice</u> $Var(S) = N\,var(X)$

where $X \sim Bernoulli(\theta)$

Always true (see theorem 3.3.4)

when $S = \sum_{i=1}^{N} X_i$, X_i iid any dist

$\Longrightarrow$ $Var(S) = N\,Var(X)$

Key point variance grows linearly in # of terms

<u>theorem 3.3.1 in ER</u>

- $Var(X) > 0$
- a, b constants then

$$Var(aX + b) = a^2\,Var(X)$$

If we look at sample mean,

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

$$\Rightarrow \text{Var}\left(\frac{1}{N} S\right) = \frac{1}{N^2} N \,\text{var}(X) = \frac{\text{Var}(X)}{N}$$

<u>Problem</u> Var(X) has different units than X

<u>Standard deviation</u>

$$\sigma_X = \sqrt{\text{var}(X)}$$

<u>Covariance (Def 3.3.3)</u>

$$\text{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right]$$

thm 3.3.1, 3.3.3, 3.3.2 are important