4/2/19 **AMS131** David Draper

Class website:
- typed lecture notes
- handwritten notes
- scanned lecture notes

Send an email to DD to be the official notetaker
- scan today's notes
- send all math classes previously taken with your grade

Discussion sections are required
- Happy to have people attend other sections
- Info will be identical
- One quiz per section (10 sections – 10 quizzes)

All tests will be take-home and open-notes

Everything is due in pdf form to Canvas

Every quiz will be due the Tuesday after they were assigned

There will be 3 take-home tests and case studies so we start with a real-world problem.

Process:

Start with problem → Come up with methods to solve the problem → Generalize the methods

One problem can have more than one possible answer based on what your assumptions are.

Office Hours will be sent out soon.

This class is cumulative. If at any moment you feel like you are falling behind, go to more office hours and sections.

You can have help from MSI and LSS. Sign ups are on Friday.

The pdf of the class textbook is attached to the class website which can only be seen when you log in.

- There are no typos or mistakes
- We'll go over chapters 1-7
- Download and Read Ch. 1 tonight

Classes are recorded with Webcast

- webcast.ucsc.edu
- U: ams-131-1
- P: uncertainty-quantification

Tomorrow the first lecture should be available.

Pace from the summer class is like drinking from a fire hose. It's still fast this quarter.

Sections start tomorrow morning.

Permission codes will be given out to students on the waitlist but not everyone will get one.

3.1 The Meaning of Probability

Case study (genetics): Tay-Sachs (T-S) disease
aka storage disease

Symptoms: accumulation of ganglioside

Cause: Lack of a Hex-A enzyme

Carriers have one gene (H) that operates normally and one gene (h) that does not.

If Hex level is 100% of normal you're a non-carrier
If it's about 50% of normal you're a carrier.

If a man and woman both are carriers, what is the possibility they will have one or more T-S babies?

Meaning of Probability

Frequentist approach: repeatable phenomena under identical conditions (with independent repetitions)

P(A) of an event is the long-run relative frequency with which A would occur in the repetitions.

Bayesian approach: A can be any T/F proposition you want (not restricted to repeatable phenomena)

P(A) is a numerical measure of the weight of evidence in favor of the statement that A is true.

Bayesian is more general but is a lot harder. We'll focus on frequentist.

<u>The Genetic Story</u>

A = { 1 or more T-S babies in a family of 5 children of 2 parents, both of whom are carriers (Hh)}

Ask what's the relative frequency of 1 or more T-S babies

Punnet Square: Father's Genes

Mother's Genes	H	h
H	(H, H)	(H, h)
h	(H, h)	(h, h)

If you have the genetic makeup (H, H) you'll have 100% of the normal level of Hex A (normal)

If you have the genetic makeup (H, h) you'll have 50% of the normal level of Hex A (carrier)

If you have the genetic makeup (h, h) you'll have 0% of the normal level of Hex A (T-S baby)

Only possibilities

Equally likely model (ELM): If you can enumerate all the ways the repeatable phenomenon you're thinking about can come out in such a way that all of these possible outcomes are equally likely then for any event A

$$P(A) = \frac{\text{\# of outcomes favorable to A}}{\text{total \# of possible outcomes}}$$

Ex: Ask deer in UCSC if they're diseased

$$\text{deer responses: } N\left\{\begin{bmatrix} \text{no} \\ \text{yes} \\ \vdots \\ \text{no} \\ \text{no} \end{bmatrix}\right. \longrightarrow \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

qualitative or categorical   quantitative or numerical

Add up the ones to count the sick deer. There are 800 deer in UCSC total.

$$\text{Mean or average} = \frac{\text{sick deer}}{\text{all deer}}$$

Take a sample of n deer where $n = 113$.

$$\begin{array}{c} y_1 \\ y_2 \\ y_3 \end{array} \left.\begin{bmatrix} 1\text{s} \\ \& \\ 0\text{s} \end{bmatrix}\right\} n = \text{sample size}$$

Assume mean $y = \theta = \frac{2}{113}$

$$\text{1 row for each individual}\left\{\underbrace{\begin{bmatrix} y_1 & \\ y_2 & \\ y_3 & \end{bmatrix}}_{\text{1 column for each variable}}\right.$$

We want our sampling method to be unbiased

Population : Sample and unsample

Goal: Make sample and unsample as similar as possible in all relevant ways

Simplest method: Choose sampled individuals at random
- Random Sampling can't achieve perfect similarity every time, but
    - a) if we imagine repeating random sampling many (M) times and averaging results, the average will move toward achieving perfect similarity as the # of repetitions increases
    - b) As sample size (n) increases, it becomes harder for (sample, unsample) to differ by a lot along relevant dimensions

2 kinds of at-random

Picking a deer at random
- every deer has an equally likely chance to get chosen at first draw

At random w/ replacement : independent identically-distributed (IID) sampling

At random w/out replacement: simple random sampling

If $n=1$, IID = SRS

If $n \ll N$, IID = SRS

IID has easier math but SRS is more informative

pop　　　　　Sample　　　　　ELM? = yes

$$\begin{bmatrix}1\\2\\9\end{bmatrix} \xrightarrow{\text{at random}} \begin{bmatrix} Y \end{bmatrix} \Big\} n=1 \qquad P(Y=2) = \frac{1}{3}$$

favorable # EOs

total # EOS

elemental outcomes (EOs)

$$\begin{pmatrix} 1 \text{ or more} \\ \text{T-S} \end{pmatrix} = \begin{pmatrix} \text{exactly } 1 \\ \text{T-S} \end{pmatrix} \text{ or } \begin{pmatrix} \text{exactly } 2 \\ \text{T-S} \end{pmatrix} \text{ or } \begin{pmatrix} \text{exactly } 5 \\ \text{T-S} \end{pmatrix}$$

$$P(A \text{ or } B) \stackrel{?}{=} P(A) \; ? \; P(B)$$

$$\begin{pmatrix} 1 \text{ or more} \\ \text{T-S} \end{pmatrix} \neq \begin{pmatrix} \text{exactly } 0 \\ \text{T-S} \end{pmatrix}$$

$$P(\text{not } A) \; ? \; P(A)$$

$$\begin{pmatrix} \text{exactly } 0 \\ \text{T-S} \end{pmatrix} = \begin{pmatrix} \text{not T-S} \\ \text{on 1st} \\ \text{baby} \end{pmatrix} \text{ and } \begin{pmatrix} \text{not T-S} \\ \text{on 2nd} \\ \text{baby} \end{pmatrix} \text{ and} \ldots \begin{pmatrix} \text{not T-S} \\ \text{on 5th} \\ \text{baby} \end{pmatrix}$$

$$P(A \text{ and } B) \stackrel{?}{=} P(A) \; ? \; P(B)$$