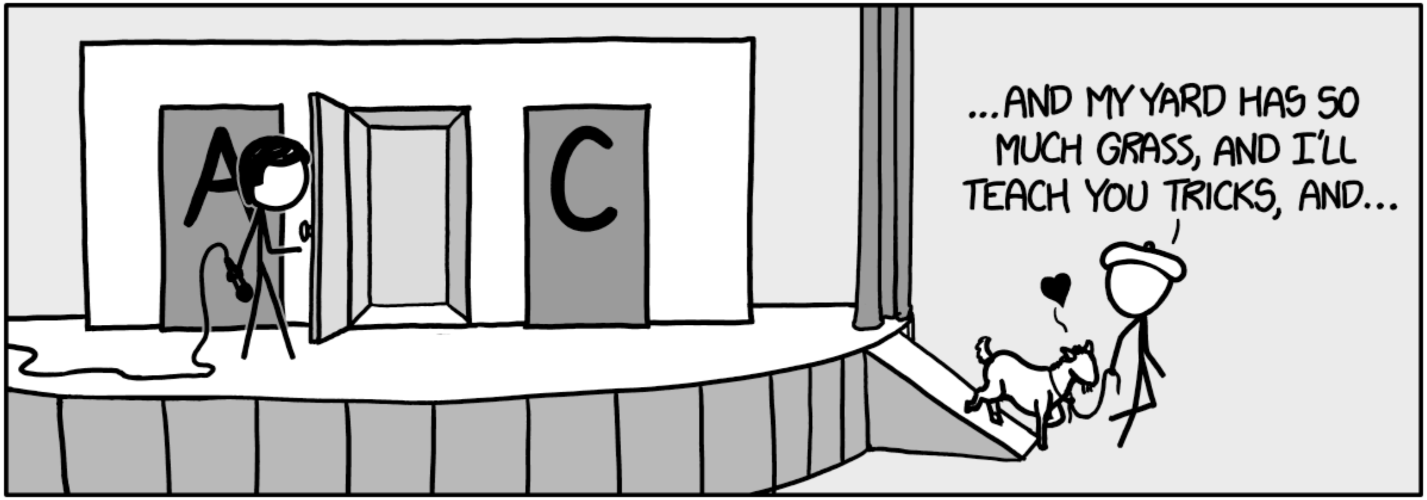
Stat 88: Probability and Mathematical Statistics in Data Science



https://xkcd.com/1282/

Lecture 6: 2/1/2021

Finish up chapter 2, Random variables & their distributions

Sections 2.4, 3.1, 3.2

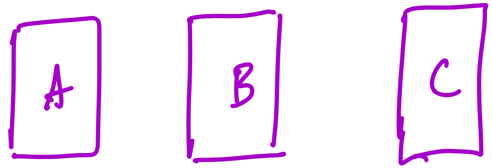
Let's make a deal!: The Monty Hall Problem

There are 3 doors, A, B, C, behind one is a new car (a Ferrari, say), and behind the other two are goats.

Now suppose you are the contestant, and you choose door A. Then Monty Hall opens one of the other two doors, say B, to show you a goat!

A : car behind A

He asks you if you want to switch to C or stick with your original choice A, you say…?



$P(A) = 1/3$

$P(A^c) = \frac{2}{3}$

$P(\text{car behind } C \mid \text{host opened door } B \text{ to show a goat})$

Agenda

- Examples from last time, talk about use and interpretation of Bayes' rule
  - Binge drinking
  - Disease, prevalence, base rate, base rate fallacy

- 3.1, 3.2, 3.3
- Review counts, permutations and combinations, $\binom{n}{k}$
- Success and failure
- Random variables
- The binomial distribution

Examples

- $P(B_{52} \mid R_{21}R_{35}) = 26/50$ (2 cards been pulled, both red, 26 black cards left)

If you pull out 2 red cards,

50 cards & 26 B cards left

So $P(B \mid 2\text{ reds}) = \frac{26}{50}$

Important bit of symmetry.

Example: Binge drinking & Alcohol related accidents

(This example is from the text *Intro Stats* by De Veaux, Velleman, and Bock)

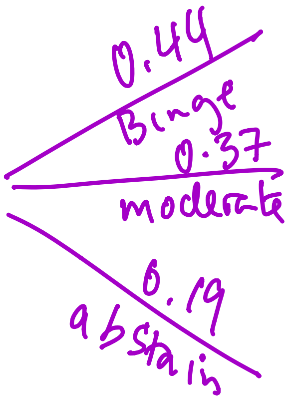
For men, binge drinking is defined as having 5 or more drinks in a row and for women as having 4 or more drinks in a row.
(The difference is because of the average difference in weight.)

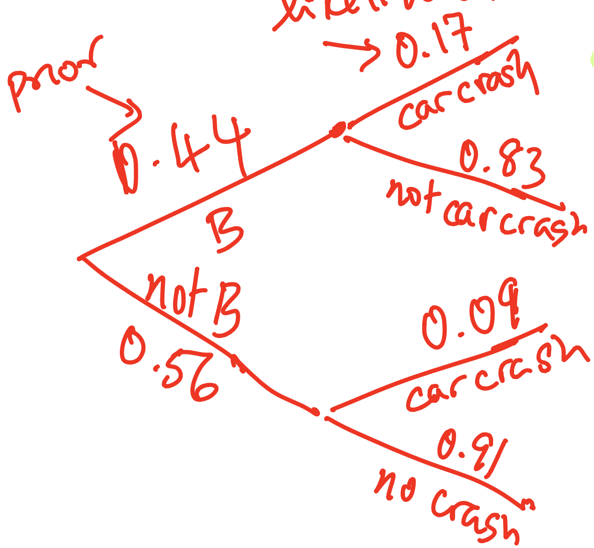
According to a study by the Harvard School of Public Health (H.Wechsler, G. W. Dowdall, A. Davenport, and W. Dejong, "*Binge Drinking on Campus: Results of a National Study*"):

- 44% of college students engage in binge drinking,37% drink moderately, and 19% abstain entirely.
- Another study, published in American journal of Health Behavior, finds that among binge drinkers aged 21 to 34, 17% have been involved in an alcohol-related automobile accident, while among nonbingers of the same age, only 9% have been involved in such accidents.
- Given that a student has been in a car crash, what is the chance that they were a binge drinker?

P(crash | binger)

Tree diagram:





(0.44)(0.17)

P(Binge & crash)

P(non bingers & crash)

$$P(\overset{B}{\text{binger}} \mid \overset{A}{\text{involved in crash}})$$

$$\left.P(B|A) = \frac{P(AB)}{P(A)} = \frac{P(B)\,P(A|B)}{P(AB) + P(AB^c)}\right\} \text{Bayes' rule.}$$

$$= \frac{(0.44)(0.17)}{(0.44)(0.17) + (0.56)(0.09)} \approx 0.597 \simeq 0.6$$

2.4: Use and interpretation of Bayes' rule

- Harvard study: 60 physicians, students, and house officers at the Harvard Medical school were asked the following question:
- "If a test to detect a disease whose ***prevalence*** is 1/1,000 has a false positive rate of 5 per cent, what is the chance that a person found to have a positive result actually has the disease, assuming that you know nothing about the person's symptoms or signs?" Lets assume the test has sensitivity 100% (P(+ test | Disease)) & specificity 95% = P(– test | no disease)

  P(+ | no disease)
- What is your guess – without any computations?
- Prevalence aka Base Rate = fraction of population that has disease. $\frac{1}{1000} = 0.1\%$
- False positive rate: fraction of positive results among people who don't have the disease
- Positive result: test is positive

"+" : pos test

"–" : neg test

D : has Disease

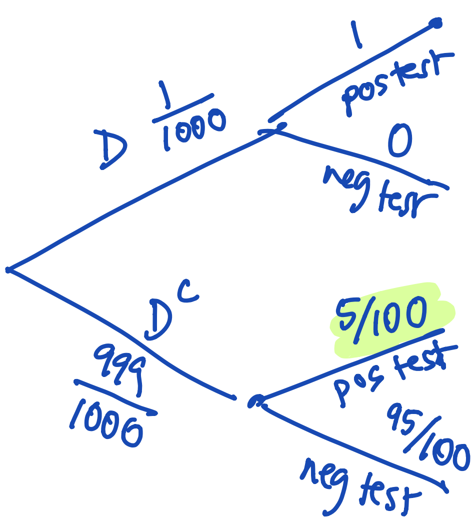
D^c : does not have disease.

Guess. P(Disease | +)

Tree diagram for disease and positive test

- P(D|pos. test) or *posterior* probability = ≈ 2%
- Recall that prior probability = 0.001 = 0.1%

Base rate.



P(D & +)

P(D & −)

$P(D^c \& +)$

$P(D^c \& -)$

$P(D|+)$

$$P(D|+) = \frac{P(+|D)\,P(D)}{P(+|D)\,P(D) + P(+|D^c)\,P(D^c)}$$

$\approx 0.0196 \approx 2\%$

Note: very small prob b/c very few of pos. tests come from people who have the diseases.

Base Rate Fallacy

- P(D|pos. test) or *posterior* probability = 2%
- Recall that prior probability = 0.001 = 0.1%

- P(+ test) = P(+ & disease) +P(+ & no disease)

$P(- | D^c) = 95\%$

- Base rate fallacy: Ignore the base rate and focus only on the likelihood. (Moral of this story: ignore the base rate at your own peril)

- Note: Want P(D|+) but most people focus on the test giving correct results for negative tests 95% of the time, that is P(no disease| neg)

- What happens to posterior probability if we change prior probability?

Read book. change base rate to 10% & the posterior prob increases ~69%

Case of Sally Clarke and SIDS

- Around 2003, Sally Clark, in a famous murder trial had two children one year apart who both died mysteriously. Sally Clarke's defence was that the babies both died of Sudden Infant Death Syndrome (SIDS)
- A = event the first child dies of SIDS
- B = event the second child dies of SIDS.
- Assumption: P(A) = P(B) = 1/8543 (based on stats, unconditional probability)

$\cancel{P(AB) = P(A)P(B)}$ was computed as $\cancel{\frac{1}{8543} \cdot \frac{1}{8543}}$

$P(AB) >$ then what was computed.

Back to counting outcomes of tosses

2^8

- Toss a coin 8 times, how many possible outcomes? 256 possible.
- What is the chance of all heads? $\frac{1}{256}$ $P(\text{not all } H) = 1 - \frac{1}{256} = \frac{255}{256}$
- If each of the students in this class present today flip a coin 8 times, what is the chance that at least 1 person gets all heads?

125

$P(\text{at least 1 person out of 125 with all H})$

$$\left(1 - \underbrace{\left(\frac{255}{256}\right)^{125}}_{\text{all 125 students not getting 8 H in 8 tosses}}\right) = P(\text{at least 1 student w/ all H})$$

Counting permutations & combinations

- Recall # of ways to rearrange n things, taking them 1 at a time is n!
- If we have only $k \leq n$ spots to fill, then $n \cdot (n-1) \cdot \ldots \cdot \big(n-(k-1)\big)$
- # of perm. of n things taken k at a time.

- If we don't care about order, then we are counting subsets, and this number is denoted by $\binom{n}{k}$, which we get by dividing: $n \cdot (n-1) \cdot \ldots \cdot \big(n-(k-1)\big)$ by $k!$

- Note: $\binom{n}{n} = 1$, $\binom{n}{0} = 1$

- Prob. of Full house =

$$P(\text{Full House}) = \frac{P(\text{pair \& a triple})}{\#\text{ of poker hands}}$$

$\boxed{2}\ \boxed{2}\quad \boxed{K}\ \boxed{K}\ \boxed{K}$