L13: Even More Probability

Today's objectives

- Introducing tree diagrams and frequencies (tables) as tools
- The prosecutor's fallacy
- Bayes Theorum
- Review key probability rules

Tools for complex proability calculations: Trees and Absolutes

Example: Unintended pregnancies by maternal age group (pg 257)

The question: Birth certificates show that approximately 9% of all births in the US are to teen mothers (aged 15-19), 24% to young-adult mothers (ages 20-24) and the remaining 67% to adult mothers (aged 25-44). A survey found that only 23% of births to teen mothers are intended. Among births to young adult women, 50% are intended, and among women aged 25-44 75% are intended

Define events using probability notation

The first step in probability questions is to translate the written text into probability statements.
Define our notation: Let M denote the age of the mother and B denote whether the birth was intended.

$P(M=T) + P(M=YA) + P(M=OA)$

- $P(M = \text{teen}) = 0.09$
- $P(M = \text{young adult}) = 0.24$
- $P(M = \text{older adult}) = 0.67$

$= 1$

- $P(B = \text{intended} \mid M = \text{teen}) = \boxed{0.23}$
- $P(B = \text{intended} \mid M = \text{young adult}) = 0.5$
- $P(B = \text{intended} \mid M = \text{older adult}) = 0.75$

Example: Unintended pregnancies by maternal age group (pg 257)

What if we want to know the probability that any given live birth in the U.S. is unintended?

- rewrite this question as a probability statment

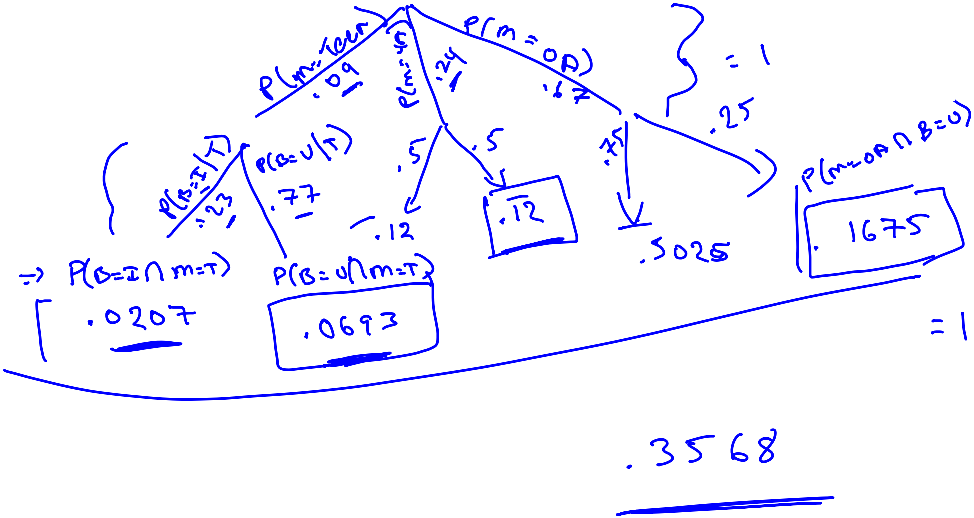
We will cover two strategies for answering this question: - Using tree diagrams - Using absoulte frequencies (not covered in your book)

Tree diagrams

Tree diagrams

- Tree diagrams can be used to perform complex probability calculations
- Tree diagrams place conditional probabilities down the branch of the tree and multiply them to obtain the probability of two (or more) events occurring.

The tree diagram for these calculations



Getting to the answer

What if we want to know the probability that any given live birth in the U.S. is unintended?
We add up all of the "branches" that contain Unintended pregnancy:

$P(B{=}\text{unintended}) = P(B = \text{unintended} \cap M = \text{teen}) +$
$P(B{=}\text{unintended} \cap M{=} \text{younger adult}) +$
$P(B = \text{unintended} \cap M = \text{older adult}) = 35.7\%$

$P(B{=}\text{unintended}) = 0.0693 + 0.12 + 0.1675 = 35.7\%$

Absolute frequencies

Absolute frequencies

- Another method for thinking about these kinds of complex probabilities is to use absolute frequencies
- We make a table of the probabilities in an imaginary population
- this may be more intuitive for most people.

Calculations using absolute frequencies

Pretend there are 1000 women. Given that 9%, 24%, and 67% of the mothers are teens, younger, and older mothers (respectively) this means that out of the 1000:

$1000 \times .09 = 90$

- 90 are teens,
- 240 are younger mothers, and,
- 670 are older mothers.

These are the marginal values of your table

Pregnancy	Teens	Younger	Older	Total
Intended				
Unintended				
Total	90	240	670	1000

Calculations using absolute frequencies

Now, *conditional* on being a teen, 23% of the pregnancies are intended. This means that 90*23% $=$ 20.7 teen mothers had intended pregnancies. We can calculate these joint probabilities for each age group

Calculations using absolute frequencies

$P(I \mid T) = .23$

- 90 are teens, $90 \times 23\% = 20.7$ teens with intended pregnancies (69.3 teens with unintended pregnancies).
- 240 are younger mothers, $240 \times 50\% = 120$ younger mothers with intended pregnancies.
- 670 are older mothers, $670 \times 75\% = 502.5$ older mothers with intended pregnancies.

Pregnancy	Teens	Younger	Older	Total
Intended	20.7	120	502.5	
Unintended	69.3			
Total	90	240	670	1000

Calculations using absolute frequencies

Now since we know the total and we know how many in each group had intended pregnancies, we can find the marginal total of Intended pregnancies.

Pregnancy	Teens	Younger	Older	Total
Intended	20.7	120	502.5	643.2
Unintended				356.8
Total	90	240	670	1000

$P(U) = \frac{356.8}{1000} = 35.7$

Thus the number of total intended pregnancies is $20.7 + 120 + 502.5 = 643.2$. Therefore, approximately 64% of all pregnancies are intended. Subtracting from 1000 we now know that 356.8 pregnancies were unintended - so 35.7% of pregnancies were unintended.

The prosecutor's fallacy

Statistics is Everywhere

from - Slate.com article, August 2018:

Consider this case from 1964: In the wake of a robbery in Los Angeles, someone reported seeing a blond woman with a ponytail commit the deed before jumping into a yellow getaway car driven by a bearded and mustachioed black man.

Statistics is Everywhere

Malcolm Collins and his wife Janet Collins, who matched the descriptions but could not be identified by the eyewitnesses, were initially convicted of the crime. The conviction hinged largely on the testimony of a mathematics instructor at a local state college who made the damning assertion that the probability of the Collins' innocence was 1 in 12 million. He came to that number by simply multiplying together some rough guesses for the probabilities of each of the separate attributes: About 1 in 4 men had a mustache. About 1 in 10 women had a ponytail. And so forth.

Prosecutors' fallacy

but:

- Even if a concordance of all those attributes (blond hair, ponytail, yellow car, etc.) is unlikely, innocence isn't necessarily equally unlikely.

Prosecutors' fallacy and conditional probability

consider the following:

- $P(\text{evidence}|\text{innocence})$
- $P(\text{innocence}|\text{evidence})$

Prosecutors often point to low values of $P(\text{evidence}|\text{innocence})$ for example, "it would be very unlikely for us to have this evidence if the defendant were innocent"

But as we saw in the Collins case, a low value of $P(\text{evidence}|\text{innocence})$ does not necessarily imply a low value of $P(\text{innocence}|\text{evidence})$

Prosecutors' fallacy

A more revealing statistic was computed in the appeal: The probability that there were two couples in the Los Angeles area that both matched the description was 40 percent.

Fortunately:

The Collins case had a relatively happy ending. Their case was appealed, the system corrected its statistical misstep, and the case became a standard example in legal pedagogy for the misuse of statistical evidence.

Bayes' theorem

Bayes' theorem

If we know the conditional probability for B given A: $P(B|A)$ and we want to know the opposite - the conditional probability of A given B: $P(A \mid B)$ we can use Bayes' theorem to get the probability we want.

Bayes' theorem (simple version)

Suppose that A and A^c are disjoint events whose probabilities are not 0 and add exactly to 1. That is, any outcome has to be exactly in one of these events. Then if B is any other event whose probability is not 0 or 1,

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

6 min video about Baye's theorem

Bayes' theorem (simple version)

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

How did we end up with the formula?

1. First recall that $P(A|B) = \frac{P(A\&B)}{P(B)}$ by conditional probability rule
2. Also: $P(B|A) = P(A\&B)/P(A)$, which implies $P(A\&B) = \boxed{P(B|A) \times P(A)}$
3. Plugging (2) into (1): $\mathrm{P(A|B)} =$

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

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

Bayes' theorem (simple version)

3. $P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$
4. Now think about P(B): $P(B) = P(B\&A) + P(B\&A^c) = P(B|A) \times P(A) + P(B|A^c) \times P(A^c)$
5. Plug in (4) into (3):

$$P(A|B) \quad \frac{P(B|A) \times P(A)}{P(B|A) \times P(A) + P(B|A^c) \times P(A^c)}$$

Bayes' theorem (simple version)

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B|A) \times P(A) + P(B|A^c) \times P(A^c)}$$

Bayes' theorem example: HIV testing

suppose we randomly select an individual living in South Africa

- we test the individual for HIV
- the prevalence of HIV in South Africa is 20%
- the sensitivity of our testing method is 85%, ie $P(T+|HIV+) = 0.85$
- the specificity of our testing method is 60%, ie $P(T-|HIV-) = 0.60$
- what is the positive predictive value of this testing method?

$P(HIV^+) = .2$

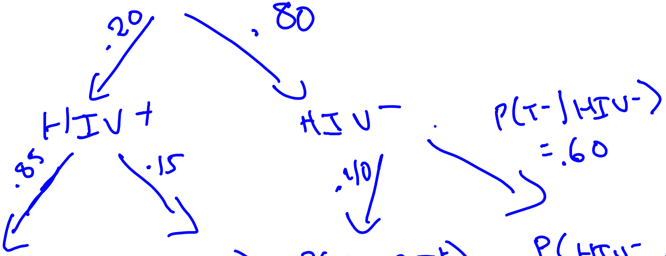
$PPV = P(HIV^+ | T^+)$

PPV as a conditional probability

$$P(T \mid HI)$$

$$P(HIV \mid T)$$

Let's look at this as a tree



$P(HIV^+ \cap T^+)$ $\quad$ $P(HIV^+ \cap T^-)$ $\quad$ $P(HIV^- \cap T^+)$ $\quad$ $P(HIV^- \cap T^-)$

$.20 \times .85$ $\quad$ $.20 \times .15$ $\quad$ $.8 \times .4$ $\quad$ $.8 \times .6$

$$S = \{\, = \boxed{.17} \quad = .03 \quad \boxed{= .32} \quad = .48 \,\} = 1$$

$$PPV = P(HIV^+ \mid T^+) = \frac{P(HIV^+ \cap T^+)}{T^+} = \frac{.17}{.32 + .17} = .3469$$

Now let's look at this in a table format

	HIV^+	HIV^-	
T^+	170	320	490
T^-	30	480	
	200	800	1000

$$PPV \quad \frac{170}{490} = .3469$$

$$P(T^+ \mid HIV^+) = 85\%$$

$$.85 \times 200 = 170$$

$$P(T^- \mid HIV^-) = .6$$

$$.6 \times 800 = 480$$

Bayes' theorem (generalized)

- Rather than only having A and A^c, suppose you had the events A_1, and $A_2, \ldots,$ through to A_k as disjoint events whose probabilities are not 0 or 1 That is, any outcome has to be exactly in one of these events. Then if B is any other event whose probability is not 0 or 1,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + ... + P(B|A_k)P(A_k)}$$

Review of probability rules

Probabilities are numbers between 0 and 1.

$0 \leq P(A) \leq 1$

The probabilities in the probability space must sum to 1.

The probabilities of an event and it's complement must sum to 1

$P(A) + P(\bar{A}) = 1$

Adding and decomposing probability

For any two events A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

For any two events A and B, $P(A) = P(A \cap B) + P(A \cap \bar{B})$

Rules for independence

Written out in probability notation, for any two events A and B, the events are independent if:

$$P(A|B) = P(A)$$

or

$$P(B|A) = P(B)$$

or

$$P(A \cap B) = P(A) * P(B)$$

Multiplication rule and conditional probability

For any two events, the probability that both events occur is given by:

$$P(A \cap B) = P(B|A) \times P(A)$$

When $P(A) > 0$, the conditional probability of B, given A is:

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

Peter Backus application of Drake's equation to dating

"Why I don't have a Girlfriend"

Number of potential girlfriends=Population of UK X P(Woman) X P(London) X P(age appropriate) X P(University education) *P(Subjectively attractive to Peter)

Or 60,975,000 X (.51) X (.13) X (.20) X (.26) X (0.05)

= 10,510

Further estimating P(woman finds peter subjectively attractive) X P(Single) X P(get along)

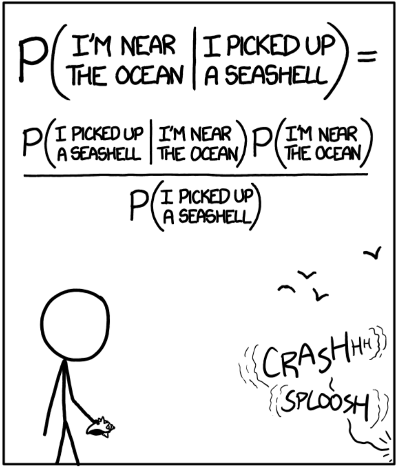
His conclusion is: "there are 26 women in London with whom I might have a wonderful relationship"

What assumption is he making in order to come to this conclusion?

full article here:

Comic Relief

From xkcd.com



STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND *DON'T* HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.