Conditional probability

3/5/25

Recap

- Two events are disjoint/mutually exclusive if they do not have any overlapping outcomes
- Addition rule: $\Pr(A \cup B) =$
- Complement rule: $\Pr(A^c) =$

Probabilities with contingency tables

- As we saw in the previous class, sometimes the probabilities of events are quite clear to calculate (e.g. dice rolls or drawing cards)
- But oftentimes we have to use data to try and estimate probabilities
  - Why? Some probabilities are not known, and we use proportions from data as a proxy
- When we have two (or more) variables, we often want to understand the relationships between them (e.g. $A \cap B$)

Practice

Source: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5788283/

	Did not die	Died	Total
Does not drink coffee	5438	1039	6477
Drinks coffee occasionally	29712	4440	34152
Drinks coffee regularly	24934	3601	28535
Total	60084	9080	69164

Define events A = died and B = non-coffee drinker. Calculate/set-up the calculations for the following for a randomly selected person in the cohort:

- $\mathrm{P}(A)$
- $\mathrm{P}(A \cap B)$
- $\mathrm{P}(A \cup B^c)$

Three types of probability

Marginal and joint probabilities

1. $\mathrm{P}(A)$ is an example of a **marginal probability**, which is a probability involving a single event
    - From the contingency table, we use row totals or column totals and the overall total to obtain marginal probabilities
2. $\mathrm{P}(A \cap B)$ is an example of a **joint probability**, which is a probability involving two or more events that have yet to occur
    - From the contingency table, we use specific cells and the overall total to obtain joint probabilities

Marginal from joint

We can obtain the marginal probabilities from joint probabilities:

	Did not die	Died	Total
Does not drink coffee	5438	1039	6477
Drinks coffee occasionally	29712	4440	34152
Drinks coffee regularly	24934	3601	28535
Total	60084	9080	69164

$$\begin{aligned}
\mathrm{P}(B) &= \mathrm{P}(\text{no coffee}) \\
&= \mathrm{P}(\text{no coffee } \cap \text{ did not die}) + \mathrm{P}(\text{no coffee } \cap \text{ died}) \\
&= \mathrm{P}(B \cap A) + \mathrm{P}(B \cap A^c) \\
&= \frac{5438}{69164} + \frac{1039}{69164} \\
&= 0.0936
\end{aligned}$$

Conditional probability

3. **Conditional probability:** a probability that an event will occur *given* that another event has already occurred
   - E.g. Given that it rained yesterday, what is the probability that it will rain today?
   - It is called "conditional" because we calculate a probability under a specific condition

- $\Pr(A|B)$: probability of A given B
  - Not to be confused with the coding `|` which is "or"
  - Appears to involve two events, but we assume that the event that is conditioned on (in this case B) *has already happened*
- We can easily obtain conditional probabilities from contingency tables!

Conditional probability with contingency tables

	Did not die	Died	Total
Does not drink coffee	5438	1039	6477
Drinks coffee occasionally	29712	4440	34152
Drinks coffee regularly	24934	3601	28535
Total	60084	9080	69164

- From contingency table, we use specific cells and row or column totals to obtain conditional probabilities

Recall events A = died and B = non-coffee drinker. Write $\mathrm{P}()$ notation for the conditional probability of dying given that someone does not drink coffee, and then obtain this probability.

General multiplication rule

Conditional, joint, and marginal probabilities are related via the **general multiplication rule**:

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

- Let's see this in the coffee example!
- Very useful for finding probability that two events will happen in sequence.
  - Example: A box has three tickets, colored red, orange, yellow. We will draw two tickets randomly one-at-a-time without replacement. What is the probability of drawing the red ticket first and then the orange ticket?

Independence and conditional probabilities

- Recall, events A and B are independent when what is true about their joint probability?
- Using the general multiplication rule, what is another way to determine if events A and B are independent?
  - Why does this make sense “intuitively”?
- Using this new test of independence, are dying and abstaining from coffee independent events?

Conditional probability formula

We can re-arrange the general multiplication formula to obtain the following general formula for conditional probability. For any events A and B:

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

- Come up with a similar formula for $\mathrm{P}(B|A)$
- Note: complement rule holds for conditional probabilities if we condition on the *same* information: $\mathrm{P}(A|B) = 1 - \mathrm{P}(A^c|B)$

Law of Total Probability

- Let A be an event, then let $\{B_1, B_2, \ldots, B_k\}$ be a set of mutually exclusive events whose union comprises their entire sample space S
- Then **Law of Total Probability (LoTP)** says:

$$\Pr(A) = \Pr(A \cap B_1) + \Pr(A \cap B_2) + \ldots + \Pr(A \cap B_k)$$

- Blob picture
- We already did this in the coffee example! We said $\text{P(no coffee)} = \text{P(no coffee } \cap \text{ did not die)} + \text{P(no coffee } \cap \text{ died)}$
  - Here, the outcomes of “did not die” and “died” are the mutually exclusive events that comprise S

Tree diagram

Tool to organize outcomes and probabilities around the structure of the data. Useful when outcomes occur sequentially, and outcomes are conditioned on predecessors. Let's do an example:

- A class has a midterm and a final exam. 80% of students passed the midterm. Of those students who passed the midterm, 90% also passed the final. Of those student who did not pass the midterm, 15% passed on the final. You randomly pick up a final exam and notice the student passed. What is the probability that they passed the midterm?
- Using $\mathrm{P}()$ notation, what probability are we interested in? What probabilities do we need to calculate along the way?
- Let's construct our tree!
- In the tree diagram, where are the three types of probabilities appearing?

Bayes' Rule

Bayes' Rule

- As we saw before, the two conditional probabilities $P(A|B)$ and $P(B|A)$ are not the same. But are they related in some way?
- **Bayes' rule**:

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

- Why is this seemingly more complicated formula useful?

Bayes' Theorem (more general)

- Suppose we have a random process and have a defined event A
- Further suppose we can break up the sample space into k disjoint/mutually exclusive outcomes or events $B_1, B_2, \ldots, B_k$
- Without loss of generality, suppose we want $\mathrm{P}(B_1|A)$
- **Bayes' Theorem** states:

$$\begin{aligned}
\mathrm{P}(B_1|A) &= \frac{\mathrm{P}(A|B_1)\mathrm{P}(B_1)}{\mathrm{P}(A)} && \text{(Bayes' Rule)} \\
&= \frac{\mathrm{P}(A|B_1)\mathrm{P}(B_1)}{\mathrm{P}(A \cap B_1) + \mathrm{P}(A \cap B_2) + \ldots + \mathrm{P}(A \cap B_k)} && \text{(LoTP)} \\
&= \frac{\mathrm{P}(A|B_1)\mathrm{P}(B_1)}{\mathrm{P}(A|B_1)\mathrm{P}(B_1) + \mathrm{P}(A|B_2)\mathrm{P}(B_2) + \ldots + \mathrm{P}(A|B_k)\mathrm{P}(B_k)}
\end{aligned}$$

- Let's see how the tree diagram compares to the formula!

Example

- In Canada, about 0.35% of women over 40 will develop breast cancer in any given year. A common screening test for cancer is the mammogram, but this test is not perfect.
- In about 11% of patients with breast cancer, the test gives a *false negative*: it indicates a woman does not have breast cancer when she does have breast cancer.
- In about 7% of patients who do not have breast cancer, the test gives a *false positive*: it indicates these patients have breast cancer when they actually do not.
- If we tested a random Canadian woman over 40 for breast cancer using a mammogram and the test came back positive, what is the probability that the patient actually has breast cancer?