Conditional Probability and Bayes' Rule

STA 198: Introduction to Health Data Science

Yue Jiang

May 27, 2020

Announcements

- HW/Lab 01 will be graded and returned by this weekend (Saturday morning, US EST) on Gradescope (linked to Sakai)
- Sakai is the official grade source - make sure it is accurate
- Regrade requests will be open for 48 hours (technically a bit longer: they'll be open until Monday, June 01, 11:59pm US EST) on Gradescope
- HW/Lab 02 will be due next Tuesday, June 02, at 11:59pm US EST; HW will be released after class today

The pace of the class is starting to pick up. Make sure you aren't waiting until the last minute to work on assignments, and please utilize office hours/the course Piazza page/reach out to me or the TAs for help.

Conditional probability

The probability an event will occur when another event has already occurred. The conditional probability of event A given event B is

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

Examples come up all the time in the real-world:

- *Given* that a mammogram comes back positive, what is the probability that a woman has breast cancer?
- *Given* that a 68-year old man has suffered four previous heart attacks, what is the probability he die in the next five years?
- *Given* that a patient has a mutation in the *CFTR* gene, what is the probability their offspring will have cystic fibrosis?

Gunter et al. (2017) study

ORIGINAL RESEARCH — Annals of Internal Medicine

Coffee Drinking and Mortality in 10 European Countries

A Multinational Cohort Study

	Died		
Coffee drinking	Yes	No	Total
None	1039	5438	6477
Med-Low	4440	29712	34152
High	3601	24934	28535
Total	9080	60084	69164

Define the events $A =$ died and $B =$ non-coffee drinker

- Marginal probability $P(A)$
- Joint probability $P(A \cap B)$
- Conditional probability $P(A|B)$

Independence and the multiplicative rule

We can rewrite the definition of conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \implies \underbrace{P(A \cap B) = P(A|B) \times P(B)}_{\text{Multiplicative Rule}}$$

What does the multiplicative rule mean in plain English?

Events A and B are said to be independent when

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

or equivalently, $P(A|B) = P(A)$ or $P(B|A) = P(B)$

Independent vs. disjoint events

Since for two independent events $P(A|B) = P(A)$ or $P(B|A) = P(B)$, knowing that one event has occurred tells us nothing more about the probability of the other occurring.

For two disjoint events A and B, knowing that one has occurred tells us that the other definitely has not occurred: $P(A \cap B) = 0$

Conditional probabilities and independence

ORIGINAL RESEARCH

Annals of Internal Medicine

Coffee Drinking and Mortality in 10 European Countries

A Multinational Cohort Study

What was the probability a randomly selected person in the study...

	Died		
Coffee drinking	Yes	No	Total
None	1039	5438	6477
Med-Low	4440	29712	34152
High	3601	24934	28535
Total	9080	60084	69164

- ...died?
- ...died, given they were a non-coffee drinker?

In this study, were dying and coffee drinking independent events? How might we check?

The law of total probability

Suppose we partition B into mutually disjoint events $B_1, B_2, \cdots B_k$ that comprise the entire sample space. Then the law of total probability states that the probability of event A is

$$\begin{aligned} P(A) &= P(A \cap B_1) + P(A \cap B_2) + \cdots + P(A \cap B_k) \\ &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_k)P(B_k) \end{aligned}$$

How did we get the last equality?

The law of total probability in action

ORIGINAL RESEARCH

Annals of Internal Medicine

Coffee Drinking and Mortality in 10 European Countries

A Multinational Cohort Study

	Died		
Coffee drinking	Yes	No	Total
None	1039	5438	6477
Med-Low	4440	29712	34152
High	3601	24934	28535

What was the probability a randomly selected person died?

The law of total probability in action

In an introductory statistics course, 50% of students were first years, 30% were sophomores, and 20% were upperclassmen.

80% of the first years didn't get enough sleep, 40% of the sophomores didn't get enough sleep, and 10% of the upperclassmen didn't get enough sleep.

What is the probability that a randomly selected student in this class didn't get enough sleep? (remember to show your work in order to get partial credit!)

Bayes' rule

What is the probability that a random person...

Coffee drinking	Died: Yes	Died: No	Total
None	1039	5438	6477
Med-Low	4440	29712	34152
High	3601	24934	28535
Total	9080	60084	69164

- ...was a high coffee drinker, given that they died?
- ...died, given that they were a high coffee drinker?

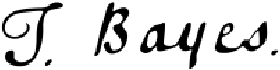
These are $P(High|Died)$ and $P(Died|High)$.
Are these two probabilities the same?

Bayes' rule

(A real portrait of him doesn't seem to exist)

We can use Bayes' rule to "reverse" the order of conditioning. By definition:

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



Bayes' rule

Using the definition of conditional probability, the law of total probability, and the multiplicative rule, we have

$$
\begin{aligned}
P(A|B) = \frac{P(A \cap B)}{P(B)} &= \frac{P(B|A)P(A)}{P(B)} \\
&= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}
\end{aligned}
$$

Bayes' rule

What is the probability that a random person in the EPIC study...

	Died		
Coffee drinking	Yes	No	Total
None	1039	5438	6477
Med-Low	4440	29712	34152
High	3601	24934	28535
Total	9080	60084	69164

- ...was a high coffee drinker, given that he died?
- ...died, given that he was a high coffee drinker?

Verify these results using Bayes' rule