Probability II

Patrick Breheny

February 15

The law of total probability

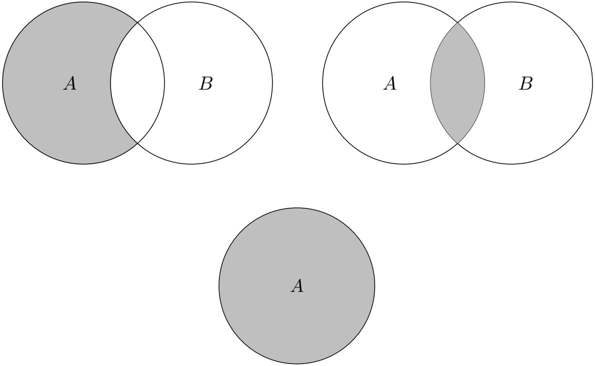
- A rule related to the addition rule is called *the law of total probability*, which states that if you divide A into the part that intersects B and the part that doesn't, then the sum of the probabilities of the parts equals $P(A)$
- In mathematical notation,

$$P(A) = P(A \cap B) + P(A \cap B^C)$$

- This is really a special case of the addition rule, in the sense that $(A \cap B) \cup (A \cap B^C) = A$ and the two events are mutually exclusive

The law of total probability: a picture

Again, the logic behind the law of total probability is clear when you see a Venn diagram:



The law of total probability in action

- In the gestational age problem, suppose we want to determine the probability of low birth weight (L) given that the gestational age was greater than 37 weeks (E^C)
- We can use the formula for calculating conditional probabilities along with the complement rule and law of total probability to solve this problem:

$$\begin{aligned} P(L|E^C) &= \frac{P(L \cap E^C)}{P(E^C)} \\ &= \frac{P(L) - P(L \cap E)}{1 - P(E)} \end{aligned}$$

The law of total probability in action (cont'd)

$$\frac{P(L) - P(L \cap E)}{1 - P(E)} = \frac{0.051 - 0.031}{1 - 0.142}$$
$$= 2.3\%$$

- Note that the unconditional probability, 5.1%, is in between the two conditional probabilities (2.3% and 21.8%)

Using tables instead of equations

- People often find it conceptually easier to see probability relationships by constructing tables
- For example, given the information that P(early labor) is .142, P(low birth weight) is .051, and P(both) is .031, we could construct the following table:

	Birth weight		
	< 2500g	≥ 2500g	Total
Early labor	31	111	142
Full term	20	838	858
Total	51	949	1000

- The probability of low birth weight given full term delivery is then $20/858 = 2.3\%$
- The probability of low birth weight or early labor is $(20 + 31 + 111)/1000 = 16.2\%$

Introduction

- Conditional probabilities are often easier to reason through (or collect data for) in one direction than the other
- For example, suppose a woman is having twins
- Obviously, if she were having identical twins, the probability that the twins would be the same sex would be 1, and if her twins were fraternal, the probability would be 1/2
- But what if the woman goes to the doctor, has an ultrasound performed, learns that her twins are the same sex, and wants to know the probability that her twins are identical?

Introduction (cont'd)

- So, we know $P(\text{Same sex}|\text{Identical})$, but we want to know $P(\text{Identical}|\text{Same sex})$
- To flip these probabilities around, we can use something called *Bayes' rule*:

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

Introduction (cont'd)

- To apply Bayes' rule, we need to know one other piece of information: the (unconditional) probability that a pair of twins will be identical
- The proportion of all twins that are identical is roughly 1/3
- Now, letting A represent the event that the twins are identical and B denote the event that they are the same sex, $P(A) = 1/3$, $P(B|A) = 1$, and $P(B|A^C) = 1/2$
- Therefore,

$$\begin{aligned} P(A|B) &= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^C)P(B|A^C)} \\ &= \frac{\frac{1}{3}(1)}{\frac{1}{3}(1) + \frac{2}{3}(\frac{1}{2})} \\ &= \frac{1}{2} \end{aligned}$$

Meaning behind Bayes' rule

- Let's think about what happened
- Before the ultrasound, $P(\text{Identical}) = \frac{1}{3}$
- This is called the *prior* probability
- After we learned the results of the ultrasound, $P(\text{Identical}) = \frac{1}{2}$
- This is called the *posterior* probability

Bayesian statistics

- In fact, this prior/posterior way of thinking can be used to establish an entire statistical framework rather different in philosophy than the one we have presented so far in this course
- In this way of thinking, we start out with an idea of the possible values of some quantity θ
- This distribution of possibilities $P(\theta)$ is our prior belief about the unknown; we then observe data D and update those beliefs, arriving at our posterior beliefs about the unknown, $P(\theta|D)$
- Mathematically, this updating is derived from Bayes' rule, hence the name for this line of inferential reasoning: *Bayesian statistics*

Bayesian statistics (cont'd)

- One clear advantage of Bayesian statistics is that it is a much more natural representation of human thought
- For example, with confidence intervals, we can't say that there is a 95% probability that the effect of the polio vaccine is between 1.9 and 3.5; with Bayesian statistics, we *can* make statements like this, because the statement reflects our knowledge and beliefs about the polio vaccine
- The scientific community has not, however, widely embraced the notion of subjective beliefs as the basis for science; the long-run frequency guarantees of p-values and confidence intervals have generally proved more marketable
- Bayesian statistics is certainly worth being aware of and is widely used and accepted in many fields – it will not, however, be the focus of this course

Testing and screening

- A common application of Bayes' rule in biostatistics is in the area of diagnostic testing
- For example, older women in the United States are recommended to undergo routine X-rays of breast tissue (*mammograms*) to look for cancer
- Even though the vast majority of women will not have developed breast cancer in the year or two since their last mammogram, this routine screening is believed to save lives by catching the cancer while it is relatively treatable
- The application of a diagnostic test to asymptomatic individuals in the hopes of catching a disease in its early stages is called *screening*

Terms involved in screening

- Let D denote the event that an individual has the disease that we are screening for
- Let $+$ denote the event that their screening test is positive, and $-$ denote the event that the test comes back negative
- Ideally, both $P(+|D)$ and $P(-|D^C)$ would equal 1
- However, diagnostic tests are not perfect

Terms involved in screening (cont'd)

- Instead, there are always *false positives*, patients for whom the test comes back positive even though they do not have the disease
- Likewise, there are *false negatives*, patients for whom the test comes back negative even though they really do have the disease
- Suppose we test a person who truly does have the disease:
  - $P(+|D)$ is the probability that we will get the test right
  - This probability is called the *sensitivity* of the test
  - $P(-|D)$ is the probability that the test will be wrong (that it will produce a false negative)

Terms involved in screening (cont'd)

- Alternatively, suppose we test a person who does not have the disease:
  - $P(-|D^C)$ is the probability that we will get the test right
  - This probability is called the *specificity* of the test
  - $P(+|D^C)$ is the probability that the test will be wrong (that the test will produce a false positive)

Terms involved in screening (cont'd)

- The accuracy of a test is determined by these two factors:
  - Sensitivity: $P(+|D)$
  - Specificity: $P(-|D^C)$
- One final important term is the probability that a person has the disease, regardless of testing: $P(D)$
- This is called the *prevalence* of the disease

Values for mammography

- According to an article in *Cancer* (more about this later),
  - The sensitivity of a mammogram is 0.85
  - The specificity of a mammogram is 0.80
  - The prevalence of breast cancer is 0.003
- With these numbers, we can calculate what we really want to know: if a woman has a positive mammogram, what is the probability that she has breast cancer?

Using Bayes' rule for diagnostic testing

- Applying Bayes' rule to this problem,

$$\begin{aligned} P(D|+) &= \frac{P(D)P(+|D)}{P(D)P(+|D) + P(D^C)P(+|D^C)} \\ &= \frac{.003(.85)}{.003(.85) + (1 - .003)(1 - .8)} \\ &= 0.013 \end{aligned}$$

- In the terminology of Bayes' rule, the prior probability that a woman had breast cancer was 0.3%
- After the new piece of information (the positive mammogram), that probability jumps to 1.3%

Why is the probability so low?

- So according to our calculations, for every 100 positive mammograms, only one represents an actual case of breast cancer
- Why is this so low?

	BC	Healthy
+	2.5	199.4
−	0.5	797.6
Total	3	997

- Way more false positives than true positives because the disease is so rare

Controversy (Part 1)

- Because $P(D|+)$ is so low, screening procedures like mammograms are controversial
- We are delivering scary news to 99 women who are free from breast cancer
- On the other hand, we may be saving that one other woman's life
- These are tough choices for public health organizations

Two studies of mammogram accuracy

- In our example, we calculated that the probability that a woman has breast cancer, given that she has a positive mammogram, is 1.3%
- The numbers we used (sensitivity, specificity, and prevalence) came from the article
  Hulka B (1988). Cancer screening: degrees of proof and practical application. *Cancer*, 62: 1776-1780.
- A more recent study is
  Carney P, et al. (2003). Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography. *Annals of Internal Medicine*, 138: 168-175.

Comparing the two studies

	Hulka (1988)	Carney (2003)
Sensitivity	.85	.750
Specificity	.80	.923
Prevalence	.003	.005
$P(D\|+)$	1.3%	4.7%

- It would seem, then, that radiologists have gotten more conservative in calling a mammogram positive, and this has increased $P(D|+)$
- However, the main point remains the same: a woman with a positive mammogram is much more likely *not* to have breast cancer than to have it

Controversy (Part 2)

- Based on these kinds of calculations, in 2009 the US Preventive Services Task Force changed its recommendations:
  - It is no longer recommended for women under 50 to get routine mammograms
  - Women over 50 are recommended to get mammograms every other year, as opposed to every year
- Of course, not everyone agreed with this change, and much debate ensued (my Google search for USPSTF "breast cancer screening" controversy returned over 75,000 hits)

Summary

- The probability of an event is the fraction of time that it happens (under identical repeated conditions)
- Know the meaning of complements (A^C), intersections ($A \cap B$), and unions ($A \cup B$)
- Addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- If (and only if!) A and B are mutually exclusive, we can ignore $P(A \cap B)$ in the addition rule
- Complement rule: $P(A^C) = 1 - P(A)$

Summary

- To determine $P(A \cap B)$, use the multiplication rule:

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

- If (and only if!) two events are independent, you can ignore conditioning on A and directly multiply the probabilities
- To determine conditional probabilities,

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

- To "flip" a conditional probability around, use Bayes' rule:

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

- Know the terms: sensitivity, specificity, prevalence