Probability II

Patrick Breheny

February 16

Balls in urns

- Imagine a random process in which balls are placed into an urn and picked out at random, so that each ball has an equal chance of being drawn
- Statisticians love these examples because lots of problems can be thought of in terms of balls and urns
- For example, imagine an urn that contains 1 red ball and 2 black balls
- Let R denote drawing a red ball; what is $P(R)$?

Balls in urns (cont'd)

- Now, imagine we draw a ball, put it back in the urn, and draw a second ball (this method of drawing balls from the urn is called *sampling with replacement*)
- What is the probability of drawing two red balls?
- *i.e.*, letting R_i denote that the i^{th} ball was red, what is $P(R_1 \cap R_2)$?
- It turns out that this probability is:

$$\frac{1}{3}\left(\frac{1}{3}\right) = \frac{1}{9} \approx 11\%$$

- On the surface, then, it would seem that $P(A \cap B) = P(A) \cdot P(B)$
- Once again, however, this is not true in general

Balls in urns (cont'd)

- Suppose we don't put the 1st ball back after drawing it (this method of drawing balls from the urn is called *sampling without replacement*)
- Now, it is impossible to draw red balls; instead of 11%, the probability is 0
- Why doesn't multiplying the probabilities work?
- Because the outcome of the first event changed the system; after R_1 occurs, $P(R_2)$ is no longer 1/3, but 0
- When we draw without replacement, $P(R_i)$ depends on what has happened in the earlier draws

Conditional probability

- The notion that the probability of an event may depend on other events is called *conditional probability*
- The conditional probability of event A given event B is written as $P(A|B)$
- For example, in our ball and urn problem, when sampling without replacement:
  - $P(R_2) = \frac{1}{3}$
  - $P(R_2|R_1) = 0$
  - $P(R_2|R_1^C) = \frac{1}{2}$

The multiplication rule

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

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

- Alternatively, if we know $P(B)$ and $P(A|B)$,

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

- This rule is always true for any two events

Calculating conditional probabilities

- The multiplication rule also helps us calculate conditional probabilities
- Rearranging the formula, we have

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

- Similarly,

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

Gestational age example

- Recall our earlier example, where the probability that a child's gestational age is less than 37 weeks is 14.2%, the probability that his or her birth weight is less than 2500 grams is 5.1%, and the probability of both is 3.1%
- What is the probability that a child's birth weight will be less than 2500 grams, given that his/her gestational age is less than 37 weeks?

$$\begin{aligned} P(\text{Low weight}|\text{Early labor}) &= \frac{P(\text{Low weight and early labor})}{P(\text{Early labor})} \\ &= \frac{.031}{.142} \\ &= 21.8\% \end{aligned}$$

Note that this is much higher than the unconditional probability of 5.1%

Independence

- Note that sometimes, event B is completely unaffected by event A, and $P(B|A) = P(B)$
- If this is the case, then events A and B are said to be *independent*
- This works both ways – all the following are equivalent:
    - $P(A) = P(A|B)$
    - $P(B) = P(B|A)$
    - A and B are independent
- Otherwise, if the probability of A depends on B (or vice versa), then A and B are said to be *dependent*

Dependence and independence

Scientific questions often revolve around whether or not two events are independent, and if they are dependent, how dependent are they?

Event A	Event B
Patient survives	Patient receives treatment
Student is admitted	Student is male
Person develops lung cancer	Person smokes
Patient will develop disease	Mutation of a certain gene

Independence and the multiplication rule

- Note that if A and B are independent, **and only if they are independent**, then the multiplication rule reduces to $P(A \cap B) = P(A)P(B)$
- This form is often much easier to work with, especially when more than two events are involved:
- For example, consider an urn with 3 red balls and 2 black balls; what is the probability of drawing three red balls?
- With replacement (draws are independent):

$$P(\text{Three red balls}) = \left(\frac{3}{5}\right)^3 = 21.6\%$$

Independence and the multiplication rule (cont'd)

- On the other hand, when events are dependent, we have to use the multiplicative rule several times:

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

  and so on

- So, when our draws from the urn are not independent (sampled without replacement):

$$P(\text{Three red balls}) = \frac{3}{5} \cdot \frac{2}{4} \cdot \frac{1}{3} = 10\%$$

Independent versus mutually exclusive

- It is important to keep in mind that "independent" and "mutually exclusive" mean very different things
- For example, consider drawing a random card from a standard deck of playing cards
  - A deck of cards contains 52 cards, with 4 suits of 13 cards each
  - The 4 suits are: hearts, clubs, spades, and diamonds
  - The 13 cards in each suit are: ace, king, queen, jack, and 10 through 2
- If event A is drawing a queen and event B is drawing a heart, then A and B are independent, but not mutually exclusive
- If event A is drawing a queen and event B is drawing a four, then A and B are mutually exclusive, but not independent
- It is impossible for two events to be both mutually exclusive and independent

Genetics

- Independent events come up often in genetics
- A brief recap of genetics to make sure that we're all on the same page:
  - Humans have two copies of each gene
  - They pass on one of those genes at random to their child
  - Certain diseases manifest symptoms if an individual contains at least one copy of the harmful gene (these are called *dominant* disorders)
  - Other diseases manifest symptoms only if an individual contains two copies of the harmful gene (these are called *recessive* disorders)

Genetics example #1

- Cystic fibrosis is an example of a recessive disorder
- Suppose that an unaffected man and woman both have one copy of the normal gene and one copy of the harmful gene
- If they have a child, what is the probability that the child will have cystic fibrosis?
- Letting M/F denote the transmission of the harmful gene from the mother/father,

$$\begin{aligned} P(\text{Child has disease}) &= P(M \cap F) \\ &= P(M)P(F) \\ &= \frac{1}{2} \cdot \frac{1}{2} \\ &= 25\% \end{aligned}$$

Genetics example #2

- Huntington's disease is an example of a dominant disorder
- Suppose that a man and woman each carry one copy of the normal gene and one copy of the harmful gene; if they have a child, what is the probability that the child will have Huntington's disease?
- To solve the problem, we need to combine the rules of probability:

$$\begin{aligned} P(\text{Child has disease}) &= P(M \cup F) \\ &= P(M) + P(F) - P(M \cap F) \\ &= P(M) + P(F) - P(M) \cdot P(F) \\ &= \frac{1}{2} + \frac{1}{2} - \frac{1}{2} \cdot \frac{1}{2} \\ &= 75\% \end{aligned}$$

Genetics example #2 (cont'd)

Alternatively, we could have solved the problem using:

$$\begin{aligned} P(\text{Child has disease}) &= 1 - P(\text{Child does not have disease}) \\ &= 1 - P(M^C \cap F^C) \\ &= 1 - P(M^C)P(F^C) \\ &= 1 - .25 \\ &= 75\% \end{aligned}$$

The Chevalier de Méré, Part II

- We can also use the rules of probability in combination to solve the problem that stumped the Chevalier de Méré
- Recall that we are interested in two probabilities:
  - What is the probability of rolling four dice and getting at least one ace?
  - What is the probability of rolling 24 pairs of dice and getting at least one double-ace?

The Chevalier de Méré, Part II (cont'd)

- First, we can use the complement rule:

$$P(\text{At least one ace}) = 1 - P(\text{No aces})$$

- Next, we can use the multiplication rule:

$$\begin{aligned} P(\text{No aces}) = &P(\text{No aces on roll 1}) \\ &\cdot P(\text{No aces on roll 2}|\text{No aces on roll 1}) \\ &\cdots \end{aligned}$$

- Are rolls of dice independent?
- Yes; therefore,

$$\begin{aligned} P(\text{At least one ace}) &= 1 - \left(\frac{5}{6}\right)^4 \\ &= 51.7\% \end{aligned}$$

The Chevalier de Méré, Part II (cont'd)

- By the same reasoning,

$$\begin{aligned} P(\text{At least one double-ace}) &= 1 - \left(\frac{35}{36}\right)^{24} \\ &= 49.1\% \end{aligned}$$

- Note that this is a little smaller than the first probability, and that both are much smaller than the $\frac{2}{3}$ probability reasoned by the Chevalier

Caution

- In genetics and dice, we could multiply probabilities, ignore dependence, and still get the right answer
- However, people often multiply probabilities when events are not independent, leading to incorrect answers
- This is probably the most common form of mistake that people make when calculating probabilities

The Sally Clark case

- A dramatic example of misusing the multiplication rule occurred during the 1999 trial of Sally Clark, on trial for the murder of her two children
- Clark had two sons, both of which died of sudden infant death syndrome (SIDS)
- One of the prosecution's key witnesses was the pediatrician Roy Meadow, who calculated that the probability of one of Clark's children dying from SIDS was 1 in 8543, so the probability that both children had died of natural causes was

$$\left(\frac{1}{8543}\right)^2 = \frac{1}{73,000,000}$$

- This figure was portrayed as though it represented the probability that Clark was innocent, and she was sentenced to life imprisonment

The Sally Clark case (cont'd)

- However, this calculation is both inaccurate and misleading
- In a concerned letter to the Lord Chancellor, the president of the Royal Statistical Society wrote:

> *The calculation leading to 1 in 73 million is invalid. It would only be valid if SIDS cases arose independently within families, an assumption that would need to be justified empirically. Not only was no such empirical justification provided in the case, but there are very strong reasons for supposing that the assumption is false. There may well be unknown genetic or environmental factors that predispose families to SIDS, so that a second case within the family becomes much more likely than would be a case in another, apparently similar, family.*

The Sally Clark case (cont'd)

- There are also a number of issues, also mentioned in the letter, with the accuracy of the calculation that produced the "1 in 8543" figure
- Finally, it is completely inappropriate to interpret the probability of two children dying of SIDS as the probability that the defendant is innocent
- The probability that a woman would murder both of her children is also extremely small; one needs to compare the probabilities of the two explanations
- The British court of appeals, recognizing the statistical flaws in the prosecution's argument, overturned Clark's conviction and she was released in 2003, having spent three years in prison

The law of total probability

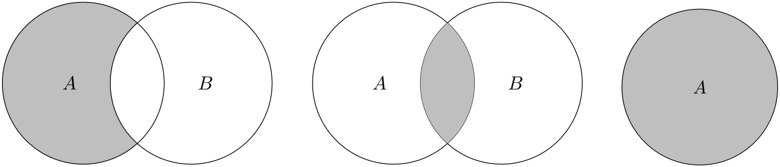
- A rule related to the addition rule is called *the law of total probability*, which states that if you divide A into the part that intersects B and the part that doesn't, then the sum of the probabilities of the parts equals $P(A)$
- In mathematical notation,

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

- 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:

$$P(L|E^C) = \frac{P(L \cap E^C)}{P(E^C)}$$

$$= \frac{P(L) - P(L \cap E)}{1 - P(E)}$$

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 20,000 hits)

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