Probability I

Patrick Breheny

February 13

Probability

- People talk loosely about *probability* all the time: "What are the chances the Hawkeyes will win this weekend?", "What's the chance of rain tomorrow?"
- For scientific purposes, we need to be more specific in terms of defining and using probabilities

Events

- A *random process* is a phenomenon whose outcome cannot be predicted with certainty
- An *event* is a collection of outcomes
- Examples:

Random process	Event
Flipping a coin	Obtaining heads
Child receives a vaccine	Child contracts polio
Roll a die	Die shows 1 or 2
10 children receive vaccine	At least 1 child contracts polio

Long-run frequency

- The probability of heads when flipping a coin is 50%
- The probability of rolling a 1 on a 6-sided die is 1/6
- Everyone agrees with these statements, but what do they really mean?
- The probability of an event occurring is defined as the fraction of time that it would happen if the random process occurs over and over again under the same conditions
- Therefore, probabilities are always between 0 and 1

Long-run frequency (cont'd)

- Probabilities are denoted with a $P(\cdot)$, as in $P(\text{Heads})$ or $P(\text{Child develops polio})$ or "Let H be the event that the outcome of a coin flip is heads. Then $P(H) = 0.5$"
- Example:
  - The probability of being dealt a full house in poker is 0.0014
  - If you were dealt 100,000 poker hands, how many full houses should you expect?
  - $100,000(0.0014) = 140$
- Note: It is important to distinguish between a probability of .0014 and a probability of .0014% (which would be a probability of .000014)

Long-run frequency (cont'd)

- This works both ways:
  - For the polio data, 28 per 100,000 children who got the vaccine developed polio
  - The probability that a child in our sample who got the vaccine developed polio is 28/100,000=.00028
- Of course, what we really want to know is not the probability of a child in our sample developing polio, but the probability of a child in the population developing polio – we're getting there

Listing the ways

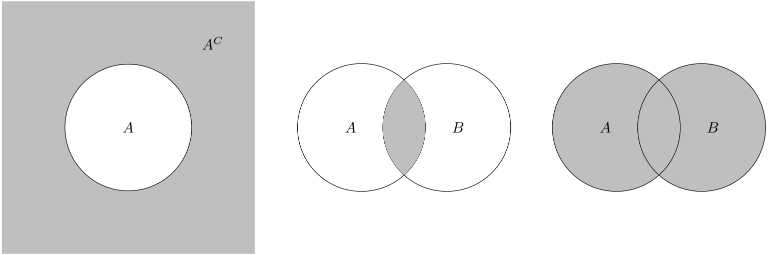
- When trying to figure out the probability of something, it is sometimes helpful to list all the different ways that the random process can turn out
- If all the ways are equally likely, then each one has probability $\frac{1}{n}$, where n is the total number of ways
- Thus, the probability of the event is the number of ways it can happen divided by n
- This is useful if the number of possibilities is small (e.g., the possible numbers that you could roll with a die) and not so useful if that number is large (e.g., all possible 5-card hands in poker)

Intersections, unions, and complements

- More complicated events can often be thought of being derived from simpler events:
  - Rolling a 2 or 3
  - Patient who receives a therapy is relieved of symptoms and suffers from no side effects
- The event that A does not occur is called the *complement* of A and is denoted A^C
- The event that both A and B occur is called the *intersection* and is denoted $A \cap B$
- The event that either A or B occurs is called the *union* and is denoted $A \cup B$

Venn diagrams

These relations between events can be represented visually using *Venn diagrams*:



Introduction

- Let event A denote rolling a 2 and event B denote rolling a 3
- What is the probability of rolling a 2 or a 3 ($A \cup B$)?
- It turns out to be

$$\frac{1}{6} + \frac{1}{6} = \frac{2}{6}$$

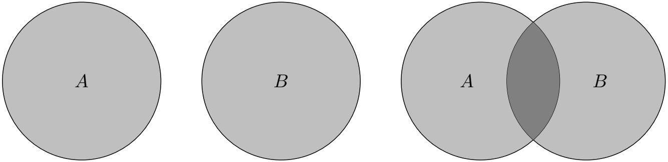
- On the surface, then, it would seem that $P(A \cup B) = P(A) + P(B)$
- However, this is not true in general

A counterexample

- Let A denote rolling a number 3 or less and B denote rolling an odd number
- $P(A) + P(B) = 0.5 + 0.5 = 1$
- Clearly, however, we could roll a 4 or a 6, which is neither A nor B
- What's wrong?

Double counting

- With a Venn diagram, we can get a visual idea of what is going wrong:



- When we add $P(A)$ and $P(B)$, we count $A \cap B$ twice
- Subtracting $P(A \cap B)$ from our answer corrects this problem

The addition rule

- In order to determine the probability of $A \cup B$, we need to know:
  - $P(A)$
  - $P(B)$
  - $P(A \cap B)$
- If we're given those three things, then we can use the *addition rule*:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- This rule is always true for any two events

Mutually exclusive events

- So why did $P(A \cup B) = P(A) + P(B)$ work when A was rolling a 2 and B was rolling a 3?
- Because $P(A \cap B) = 0$, so it didn't matter whether we subtracted it or not
- A special term is given to the situation when A and B cannot possibly occur at the same time: such events are called *mutually exclusive*

Mutually exclusive events, example

- According to the National Center for Health Statistics, the probability that a randomly selected woman who gave birth in 1992 was aged 20-24 was 0.263
- The probability that a randomly selected woman who gave birth in 1992 was aged 25-29 was 0.290
- Are these events mutually exclusive?
- Yes, a woman cannot be two ages at the same time
- Therefore, the probability that a randomly selected woman who gave birth in 1992 was aged 20-29 was $0.263+0.290=.553$

Example: Failing to use the addition rule

- In the 17th century, French gamblers used to bet on the event that in 4 rolls of the die, at least one "ace" would come up (an ace is rolling a one)
- In another game, they rolled a pair of dice 24 times and bet on the event that at least one double-ace would turn up
- The Chevalier de Méré, a French nobleman, thought that the two events were equally likely

Example: Failing to use the addition rule

- His reasoning was as follows: letting A_i denote the event of rolling an ace on roll i and AA_i denote the event of rolling a double-ace on roll i

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3 \cup A_4) &= P(A_1) + P(A_2) + P(A_3) + P(A_4) \\ &= \frac{4}{6} = \frac{2}{3} \end{aligned}$$

$$\begin{aligned} P(AA_1 \cup AA_2 \cdots) &= P(AA_1) + P(AA_2) + \cdots \\ &= \frac{24}{36} = \frac{2}{3} \end{aligned}$$

Example: Failing to use the addition rule

- Is the Chevalier using the addition rule properly?
- Are A_1 and A_2 mutually exclusive?
- No; it is possible to get an ace on roll #1 and roll #2, so you have to subtract $P(A_1 \cap A_2)$, $P(A_1 \cap A_3)$, ...
- We'll calculate the real probabilities a little later

Using the addition rule correctly

- An article in the *American Journal of Public Health* reported that in a certain population, the probability that a child's gestational age is less than 37 weeks is 0.142
- The probability that his or her birth weight is less than 2500 grams is 0.051
- The probability of both is 0.031
- Can we figure out the probability that either event will occur?
- Yes: 0.142 + 0.051 - 0.031 = 0.162

The complement rule

- Because an event must either occur or not occur, $P(A) + P(A^C) = 1$
- Thus, if we know the probability of an event, we can always determine the probability of its complement:

$$P(A^C) = 1 - P(A)$$

- This simple but useful rule is called the *complement rule*
- Example: If the probability of getting a full house is 0.0014, then the probability of not getting a full house must be $1 - 0.0014 = 0.9986$

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: Cystic fibrosis

- Cystic fibrosis is an example of a recessive disorder
- Letting C denote the normal version of the gene and c the disease-causing version of the gene, the possible outcomes of an individual inheriting cystic fibrosis genes are

$$CC \qquad Cc \qquad cC \qquad cc$$

- If all these possibilities are equally likely (as they would be if the individual's parents had one copy of each version of the gene), then the probability of having the disease (cc) is 1/4

Alternative solution

- We can derive the same result using the laws of probability
- Again, supposing that an unaffected man and woman both have one copy of the normal gene and one copy of the harmful gene, and letting M/F denote the transmission of the harmful gene from the mother/father,

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

- This might seem pedantic in this case, but the equations are useful for more complicated cases

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?
- Again, we could count, or use 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)

In probability, there is often more than one way to arrive at a solution; 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