Probability and sets

S520

August 30, 2016

Probability and sets

Trosset ch. 2.1

Mathematical probability is based on an extension of set theory called measure theory. Measure theory is important if you want to do a Ph.D. in Statistics, but for now, set theory will do.

We use S to denote the universe of possible objects. For a set A, the **complement** of A (written A^c) is the set of objects *not* in A.

The **union** of sets A and B, denoted $A \cup B$, is the set of all objects in either A or B (or both.) We can think of this as "A *or* B happens."

The **interesection** of sets A and B, denoted $A \cap B$, is the set of all objects in both A or B. We can think of this as "both A *and* B happen."

When we draw a Venn diagram of two sets, we write down the probabilities for:

- A but not B: $P(A \cap B^c)$
- Both A and B: $P(A \cap B)$
- B but not A: $P(A^c \cap B)$
- Neither A nor B: $P(A^c \cap B^c)$

These four probabilities must of course sum to 1.

Conditional probability

Trosset ch. 3.4

Sometimes we want to find the probability of something happening **if** or **given that** some other event happens. These are called **conditional probabilities**.

Example. I put ten balls in an urn. Of these ten balls:

- 2 have a red star
- 3 have a red cross
- 3 have a blue star
- 2 have a blue cross

Suppose I draw a red ball. What is the **conditional** probability it has a star?

The key ideas here are that once I know the ball is red, - I can ignore all the blue balls - I can treat the remaining red balls as equally likely.

Out the five red balls, two have a star. So the proability is 2/5.

The notation for "conditional on" or "given" is a vertical line:

P(star | red) = 0.4

Now suppose I draw a ball with a star. What is the probability it's a star's red?

Answer: 4 balls have stars; of these, 2 are red. The probability is thus 2/4.

P(red | star) = 0.5

Note 1: We could have asked this question in the following equivalent ways:

- What's the conditional probability the ball I draw is red given that it has a star?
- If I draw a ball with a star, what's the probability it's red?

Note 2: In general, P(A|B) is not P(B|A). The proportion of stars that are red doesn't have to be anything like the proportion of reds that are stars. Another example: The percentage of men who are drug dealers is low, but the percentage of drug dealers who are men is high.

Example. I put six balls labeled A, B, C, D, E, F in an urn. I draw two balls, without replacement. Given that one of the balls is an E, what is the conditional probability both are vowels?

We can just list all the possible pairs of balls that include an E (ignoring order):

EA EB EC ED EF

Only one of these has two vowels, so the required conditional probability is 1/5.

Example. I put six balls labeled A, B, C, D, E, F in an urn. I draw two balls, **with** replacement. Given that at least one of the balls is an E, what is the conditional probability both are vowels?

When we're sampling with replacement, we need to take order into account. The possible samples with at least one E are:

AE BE CE DE EA EB EC ED EF FE

There are eleven of these, and three have two vowels (AE, EA, EE.) So the required conditional probability is 3/11.

Here's the formal definition of conditional probability:

If A and B are events, and $P(B) > 0$, then the conditional probability of A given B is

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

If we make the intersection probability the subject, we get the following *multiplication rule*:

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

Similarly, if $P(A) > 0$, then the conditional probability of B given A is

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

and hence we get another multiplication rule:

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

Example. *I roll a fair die. Given that I get an odd number, what's the probability I get 3 or more?*

We could answer this using counting, but let's try to use the definition.

$$P(\text{at least 3}|\text{odd}) = \frac{P(\text{at least 3 and odd})}{P(\text{odd})}$$

Now, the probability of odd is 1/2. "At least 3 and odd" means 3 or 5, which has probability 1/3. So the required conditional probability is (1/3)/(1/2) = 2/3.

Tree diagrams

*Trosset pp. 60–61**

When dealing with a sequence of decisions, we can often find probabilities straightforwardly using a **tree diagram**.

- For the first decision, draw a branch for each outcome and label it with the probability of that outcome.
- For subsequent decisions, draw a branch for each outcome and label it with the *conditional* probability of that outcome, given the decisions up to that point.
- To find the probability of a sequence of decisions, multiply the probabilities along the branches corresponding to that sequence.

Example. *A class of ten students contains eight men and two women (Trish and Lita.) I select three students without replacement. What's the probability at least one woman is chosen?*

Here are the ways that at least one woman can be chosen:

MMW, MWM, MWW, WMM, WMW, WWM

(Note there's no WWW branch, since there are only two women.) By drawing out the tree, we can easily find the probabilities of each branch:

$$P(MMW) = 8/10 \times 7/9 \times 2/8 = 112/720$$
$$P(MWM) = 8/10 \times 2/9 \times 7/8 = 112/720$$
$$P(MWW) = 8/10 \times 2/9 \times 1/8 = 16/720$$
$$P(WMM) = 2/10 \times 8/9 \times 7/8 = 112/720$$
$$P(WMW) = 2/10 \times 8/9 \times 1/8 = 16/720$$
$$P(WWM) = 2/10 \times 1/9 \times 8/8 = 16/720$$

So the probability of at least one woman is

$$\frac{112 + 112 + 16 + 112 + 16 + 16}{720} = \frac{384}{720} = \frac{8}{15}$$

Easier way: This is a case where it's easier to find the probability something *doesn't* happen, and then take one minus that probability.

The complement of "at least one woman" is "no women". There's only one way to get no women: three men in a row.

$$P(MMM) = 8/10 \times 7/9 \times 6/8 = 336/720 = 7/15 P(\text{at least one woman}) = 1 - P(MMM) = 8/15$$

Finally, we could've also solved this problem using combinations if we wanted to.

Example ctd. *Given that at least one woman is chosen, what's the probability Trish is chosen?*

Write down the definition of this conditional probability:

$$P(\text{Trish chosen}|\text{at least one woman chosen}) = \frac{P(\text{Trish chosen and at least one woman chosen})}{P(\text{at least one woman chosen})}$$

We already calculated the denominator. Now let's look closely at the numerator:

$$P(\text{Trish chosen and at least one woman chosen})$$

But if Trish is chosen, the part after the "and" (at least one woman chosen) is automatically satisfied. So this is equivalent to P(Trish chosen). Since we're choosing three out of ten people, the probability Trish is chosen is 3/10.

The answer is thus

$$P(\text{Trish chosen}|\text{at least one woman chosen}) = \frac{3/10}{8/15} = 9/16$$

We're now ready to do a real-life example.

Example. Screening mammograms are used to test for breast cancer among women with no symptoms of the disease. However, their use is controversial, in particular for women in the forties: different organizations give different recommendations.

The Breast Cancer Surveillance Consortium collects data on women who get screening mammograms. For women aged 40-44 who got screening mammograms, around 0.27% had breast cancer. Out of those women who had breast cancer, 73.4% tested positive on the mammogram (the rest were false negatives.) Out of those women who didn't have breast cancer, 87.7% tested negative (the rest were false positive.)

Suppose we randomly select a 40-44 year-old woman who tested positive on a screening mammogram. What's the probability that she actually has breast cancer?

To draw the tree, we first need an unconditional probability. The one we have is P(breast cancer). So draw two branches for "BC" and "No BC". Then for each of these outcomes, draw two more branches, "positive" and "negative", giving four ending branches. The probabilities for each branch are:

$$\begin{aligned}
P(BC, +) &= .0027 \times .734 = .0020 \\
P(BC, -) &= .0027 \times .266 = .0007 \\
P(\text{No } BC, +) &= .9973 \times .123 = .1227 \\
P(\text{No } BC, -) &= .9973 \times .877 = .8746
\end{aligned}$$

In other words, out of 10,000 women aged 40-44 who take the screening mammogram, you'd expect 20 to get true positives, 7 to get false negatives, 1227 to get false positives, and the rest to get true negatives.

This gives a quick estimate of our answer. Out of 10,000 women, we'd expect 1247 to test positive: 20 true positives and 1227 false positives. Out of women who test positive, the fraction that actually have breast cancer is thus about 20/1247, or 1.6%.

More formally,

$$\begin{aligned}
P(BC|+) &= \frac{P(+ \cap BC)}{P(+)} \\
&= \frac{P(+ \cap BC)}{P(BC \cap +) + P(\text{No } BC \cap +)} \\
&= \frac{.0020}{.1247} \\
&1.59
\end{aligned}$$

Note this is very different from $P(+|BC)$, which was 73.4%. The formal way of changing from $P(A|B)$ to $P(B|A)$ and vice versa is called **Bayes' theorem** (Trosset p. 62–64), though usually it's clearer to draw a tree.

Finally, note that the conditional probability alone is not enough to make a reasoned decision about whether or not to get a mammogram – you'd also need to know the benefits and costs of true positives and false positives. If you are a 40-44 year-old woman, this is probably not a decision you should make based solely on information from your (male) statistics lecturer.

Independent events

Trosset pp. 64-69

Two events A and B are independent if and only if

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

Some math shows this definition implies:

$$P(A) = P(A|B) = P(A|B^c)$$
$$P(B) = P(B|A) = P(B|A^c)$$

(assuming the event conditioned on doesn't have zero probability.)

In other words, the probability of A is the same whether or not B occurs (or if we don't know whether or not B occurred.) Similarly, the probability of B is the same whether or not A occurs.

Example. *I roll a fair die. Are the events {odd} and {at least five} independent?*

$$\begin{aligned} P(\text{odd}) &= 1/2 \\ P(\text{at least } 5) &= 1/3 \\ P(\text{odd and at least } 5) &= 1/6 \end{aligned}$$

Since $1/2 \times 1/3 = 1/6$, the events are independent.

Note: In probability, we require this to be an exact equality. Much later, in chapter 13, we'll learn how to use real (messy) data to test for independence.

Example. *Are the events "it will rain tomorrow" and "it will rain in two days" independent?*

No. When it rains, it often rains for several days. So if it rains tomorrow, your probability that it'll rain the following day should go up. If it doesn't rain tomorrow, your probability that it'll rain the following day should go down. Since your probability will change conditional on what happens tomorrow, the events aren't independent.

See Trosset example 3.15 for more examples where one can determine dependence using common sense.

Random variables

Trosset ch. 3.5

A random variable is a way of represent the result of an experiment using real numbers as possible outcomes. That is, it's a variable in the algebra sense (i.e. an unknown number), but it's random (there may be no way of knowing what number it's going to be until you actually do the experiment.) Random variables are denoted by capital letters: most commonly X.

Examples

- I toss a coin. Let X be a random variable that takes the value 1 if it's heads and 0 if it's tails.
- I toss four coins. Let X be a random variable representing the number of heads.
- I roll a die. Let X be a random variable that take a value equal to the number facing upward.
- I roll ten dice. Let X be a random variable representing the sum of these ten dice.
- I roll ten dice. Let X be a random variable representing the number of squares among these ten dice.
- I generate a random number between 0 and 1. Let X be a random variable representing this number.

- I randomly select a U.S. adult. Let X be a random variable that takes the value 1 if they intend to vote for Hillary Clinton, 0 if they intend to vote for Donald Trump, and 0.5 otherwise.
- I randomly select a U.S. adult. Let X be a random variable representing their income.

Trosset ch. 3.5 has a formal definion of random variables, which you should read carefully if you're a statistics or math grad student. The important thing in there is Definition 3.5:

The **cumulative distribution distribution (CDF)** of a random variable X, written $F(y)$, is defined as:

$$F(y) \equiv P(X \leq y)$$

for all real numbers y.

Example. I toss a fair coin. Let X be a random variable that takes the value 1 if it's heads and 0 if it's tails. What's the CDF?

Note that the *only* possible outcomes are 0 and 1. When there are a countable number of possible outcomes, the CDF will be a step function, with jumps that occur at the possible outcomes.

Let's break the real line into pieces and consider what happens in each section?

- *What's the CDF when y is less than 0?* That is, if y is less than zero, what's the probability that the random variable X takes a value less than or equal to y?

Well, it's impossible: the smallest that X can be is zero, so it can *never* be less than or equal to y. The CDF is thus zero.

- *What's the CDF when y is at least 1?* That is, if y is at least 1, what's the probability that the random variable X takes a value less than or equal to y?

It's certain: X is always less than or equal to 1, so it's always less than or equal to y. The CDF is 1. (This is still true when y is exactly 1.)

- *What's the CDF when y is in between 0 and 1?* Then X is less than y if and only if $X = 0$. For a fair coin, this happens with probability 0.5. (Note that the CDF is 0.5 for $y = 0$, but *not* for $y = 1$.)

We can thus write down the CDF formally:

$$F(y) = \begin{cases} 0 & y < 0 \\ 0.5 & 0 \leq y < 1 \\ 1 & y \geq 1 \end{cases}$$

The great thing about the CDF is that once you have it, you can work out the probability of *any* event (aside from weird infinity stuff.)

Example. Let X be a random variable with CDF

$$F(y) = \begin{cases} 0 & y < 0 \\ y & 0 \leq y < 0.5 \\ 1 & y \geq 0.5 \end{cases}$$

- *What's $P(Y \leq 0.2)$?*

The CDF is the "less than or equal to" probability, so $P(Y \leq 0.2)$ is exactly the same as $F(0.2)$. Choosing the correct piece of the function, we get $F(0.2) = 0.2$.

- *What's $P(Y > 0.2)$?*

We can get the "greater than" probability by taking one minus the "less than or equal to" probability. That is, $P(Y > 0.2) = 1 - P(Y \leq 0.2) = 1 - 0.2 = 0.8$.

- *What's $P(0.2 < Y \leq 0.7)$?*

To find an "in-between" probability, use subtraction.

$$P(0.2 < Y \leq 0.7) = P(Y \leq 0.7) - P(Y \leq 0.2) = 1 - 0.2 = 0.8.$$

- *What's $P(Y = 0.5)$?*

This can also be done with subtraction, but it needs a little finesse.

Firstly, what's $P(Y \leq 0.5)$? Plugging in to the third piece of function gives $F(0.5) = 1$.

Secondly, what's $P(Y \leq 0.4999\ldots)$? (If you are mathematically pedantic, we're asking: what's the limit of $F(y)$ when we approach $y = 0.5$ from below.) Well, $F(0.49) = 0.49$, and $F(0.499) = 0.499$, and so on. So as y gets closer to 0.5 from below, $F(y)$ approaches a limit of 0.5.

The probability that Y is exactly 0.5 is thus $F(0.5) - F(0.499\ldots) = 1 - 0.5 = 0.5$.

Note that for any y-value *other* than 0.5, this procedure gives a probability of zero, since the limit from below is the same as the actual value of $F(y)$. So where did the other 0.5 of probability go? We resolve this paradox in chapter 5.