Chapter 3: Probability

S520

These notes are written to accompany Trosset chapter 3.

The rules of the game

Read: Trosset 3.1–3.2

Definitions

Sample space: A "universe" of possible *outcomes* for the experiment in question. Exactly one of the possible outcomes will happen.

Event: A subset of the sample space. It may be made up of one outcome, or many, or none.

Mathematical probability

Mathematical probability is essentially just a set of rules for putting numbers on events. For every event E, we want to assign a probability $P(E)$. In particular, we take the following three rules as axioms (Trosset p. 48):

1. If E is an event, then $0 \leq P(E) \leq 1$.
2. $P(S) = 1$.
3. If $\{E_1, E_2, E_3, \ldots\}$ is a countable collection of pairwise disjoint events, then

$$P\left(\cup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i)$$

Remember that "pairwise disjoint" means for any pair of events, the probability that they both happen is zero. Axiom 3 says that the probability that at least one of a collection of pairwise disjoint happens is the sum of the probabilities of the individual events.

Note that the math doesn't say anything about how you get your probabilities – it could be entirely subjective. So the following set of probabilities:

- P(I become Chief Justice of the Supreme Court) = 99%
- P(I do not become Chief Justice of the Supreme Court) = 1%

is perfectly fine as far as mathematical probability is concerned. As statisticians, however, we need to make sure our probabilities are connected with reality. There are two basic ways to do this:

- Sensible assumptions, e.g. equally likely outcomes.
- Data.

The goal for repeatable events is that events that have a probability of, say, 30%, should happen on 30% of repetitions in the long run. (How probability should interpreted for one-off events is controversial.)

We can use the rules above to make all kinds of crazy new rules. For example:

Complement rule. The probability that event A *doesn't* happen is $1 - P(A)$.

Proof. The events "A" and "Not A" are disjoint. So the probability that at least one of these two events happens is the sum of the two individual probabilities. But "at least of A and Not A" covers all possible outcomes, so it's certain to happen. That is, by Axiom 2:

$$P(A) + P(\text{Not } A) = 1$$

so

$$P(\text{Not } A) = 1 - P(A).$$

Venn diagrams

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.

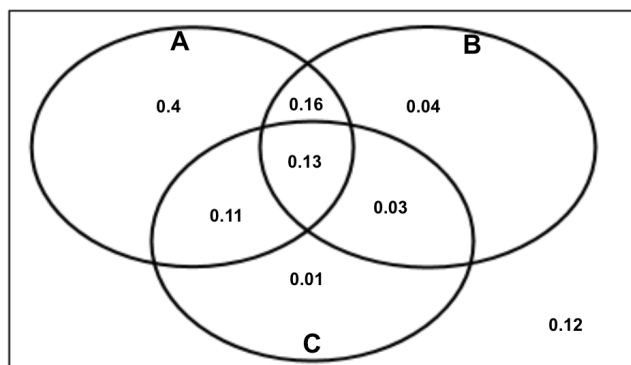**Example (Trosset chapter 3.7 exercise 1)**

(a) Venn diagram:

Figure 1: Venn diagram for Exercise 3.7.1.

Note:

- Mark the probability in EACH disjoint section of the figure.
- Don't forget to write the probability of the event that includes all outcomes not in sets A, B or C. In this example, that is 0.12.

- Before calculate the probabilities, check the sum of all the probabilities. If it is not equal to 1, there must be something wrong in the Venn diagram.

(b) From the Venn diagram, $P(A \cap B \cap C^c) = 0.16$.

(c) From the Venn diagram, $P(A \cap B^c \cap C^c) = 0.4$.

(d) From the Venn diagram, 0.12.

(e) The event $A^c \cap (B \cup C)$ includes all outcomes in the union of B and C, but not in A. From the Venn diagram, $0.04 + 0.03 + 0.01 = 0.08$.

Trosset 3.3

Equally likely outcomes

When we consider a set of *outcomes* for an experiment, we mean a set such that exactly one of the outcomes must occur. Suppose there are N equally likely outcomes. Since their probabilities must add to 1, the probability of any particular outcome is $1/N$.

This seems trivial, but we can use this to turn some quite complicated situations into counting problems.

Example

My top six Pokemon in Pokemon Go are Vaporeon, Hypno, Golbat, Flamethrower Growlithe, Persian, and Body Slam Growlithe.

1. I select two of these Pokemon without replacement. What's the probability they're both Growlithes?

There are $6 \times 5 = 30$ ways of choosing two out of six objects in order. But in this question, order doesn't matter: Flamethrower-Body Slam is the same as Body Slam-Flamethrower. So divide by 2: there are 15 ways of choosing two out of six objects if order doesn't matter. Each way has probability $1/15$. That's the answer.

1. I select two of these Pokemon with replacement. What's the probability they're both Growlithes?

To use equally likely outcomes when sampling with replacement, we have to count them as if order matters. Then there are 6×6 equally likely outcomes. Of these, $2 \times 2 = 4$ of them are Growlithe-Growlithe (Flame-Flame, Flame-Body, Body-Flame, Body-Body.) So the probability is $4/36$, or $1/9$.

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
- 2 have a blue star
- 3 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 the star is 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?

In sampling with replacement questions, it's easier if we take order into account (since then we have equally likely outcomes.) The possible samples with at least one E are:

AE BE CE DE EE FE EA EB EC ED EF

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.

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

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

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

Tress lead naturally to the idea of **Bayes' Theorem** (Trosset pp. 62–64), which we omit for reasons of time.

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?*

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

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 representing 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. Formalizing the idea of random variables turns out to be exceedingly difficult – it involves the concept of **Borel sets**, and unless you are planning a Ph.D. in statistics or math, you should skip any paragraph in the textbook that mentions Borel sets. Fortunately, we can understanding what a random variable is by looking at some examples.

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.

The CDF

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

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

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:

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

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

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

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

The CDF is the "less than or equal to" probability, so $P(Y \le 0.2)$ 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, $P(Y > 0.2) = 1 - P(Y \le 0.2) = 1 - 0.2 = 0.8$.

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

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

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

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

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

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

Secondly, what's $P(Y \le 0.4999\ldots)$? (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.

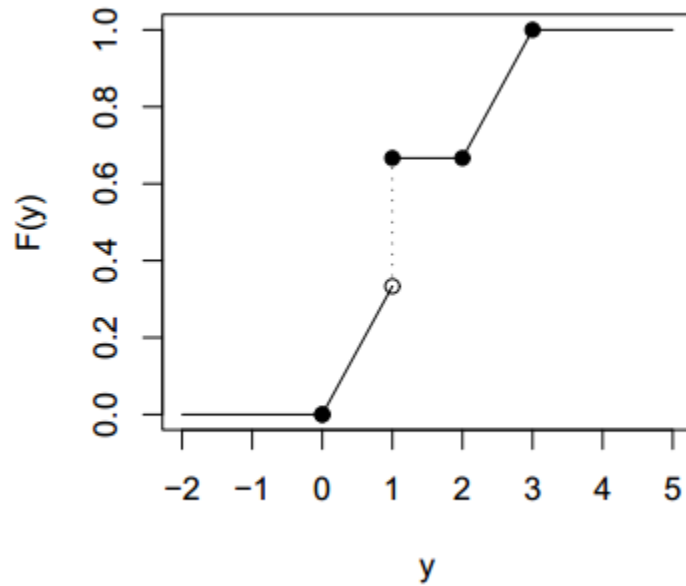**Example.** (Trosset chapter 3.7 exercise 14)

Figure 2: Graph of CDF for Exercise 3.7.14.

(a)

$$P(X > 0.5) = 1 - P(X \leq 0.5) = 1 - F(0.5) = 1 - \frac{1}{6} = \frac{5}{6}$$

(b)

$$
\begin{aligned}
P(2 < X \leq 3) &= P(X \leq 3) - P(X \leq 2) \\
&= F(3) - F(2) \\
&= 1 - \frac{2}{3} = \frac{1}{3}
\end{aligned}
$$

(c)

$$
\begin{aligned}
P(0.5 < X \leq 2.5) &= P(X \leq 2.5) - P(X \leq 0.5) \\
&= F(2.5) - F(0.5) \\
&= \frac{5}{6} - \frac{1}{6} \\
&= \frac{2}{3}
\end{aligned}
$$

(d) $P(X = 1)$ is the vertical distance of the jump at $y = 1$. This is $1/3$.

Note: When there is no jump at a certain value x, the probability of X being x is 0. For instance, $P(X = -1) = P(X = 0.5) = P(X = 1.2) = P(X = 3.3) = 0$ in this problem.

Trosset chapter 3.6 is a case study that we invite you to read at your leisure.