Lecture 3: Conditional Probability and Bayes' Theorem

Recall: At the end of Friday's lecture, I was trying to do an example with three sets, and it didn't work. Looking back at my notes, this example really just didn't make any sense and your first instincts were correct: more information is required.

Today I'll begin with a similar example. I take a survey, and find the following results:

- 53% of people are married.
- 67% of people are citizens.
- 36% of people have a college education.
- 31% are married citizens.
- 23% are married with a college education.
- 20% are citizens with a college education.
- Everybody surveyed was married, was a citizen, or had a college education.

How many people are married citizens with a college education?

Calling the group of married people A, the citizens B and the college educated people C, we use the above information to find that $p(A \cup B) = p(A) + p(B) - p(A \cap B) = 89\%$. Similarly, $p(A \cup C) = 66\%$ and $p(B \cup C) = 83\%$. Continuing in a similar vein gives the Venn Diagram in Figure 1.

Today, we discuss conditional probability.

This asks "What is the probability of one event given that some other event happens?"

For example, what is the probability that someone married is a citizen?

Married people make up 53% of the total, and 31% are married citizens, so the probability that a random married person is a citizen is $31/53 = .58$.

For another example, recall that the sample space for tossing three distinct coins is $\{hhh, hht, hth, thh, htt, tht, tth, ttt\}$ and that the probability that exactly one of these coins lands heads is $3/8$.

Suppose we don't care about the ttt result, and every time we get this outcome we will retoss the coins. We have reduced the sample space to only seven outcomes,

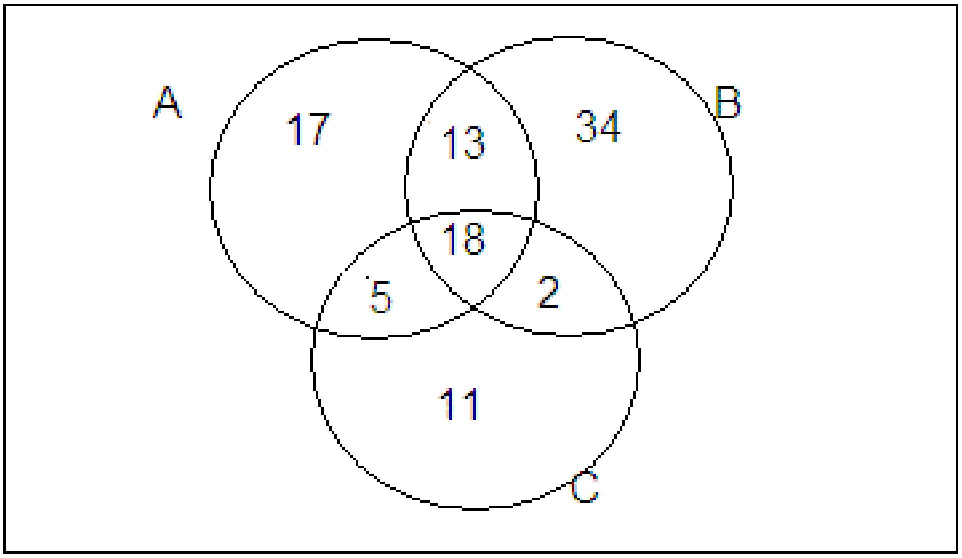


Figure 1: A Venn diagram for the example with A representing married people, B representing citizens, and C representing persons with a college education.

those where at least one heads occurs. Now, the probability that exactly one heads occurs is 3/7.

In general, for any events A and B where $p(B) \neq 0$ (B occurs some of the time) we say $p(A|B)$ is the *conditional probability* of A given B. Also,

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

We can think of this as the portion of B which is also in A.

From the formula from conditional probability, we develope the following product rule:

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

Look again at three coin tosses. Let A be the event where the first coin lands heads and B the event where the second coin lands heads. Then $p(A) = p(B) = 1/2$ and $p(A \cap B) = 1/4$. This gives $p(A|B) = 1/2$, which is $p(A)$.

When it happens that $p(A|B) = p(A)$, we say that A and B are *independent events*. One way to think of this is that whether or not B occurs has no effect on whether or not A occurs. The product rule in this case simplifies to

$$p(A \cap B) = p(A)p(B)$$

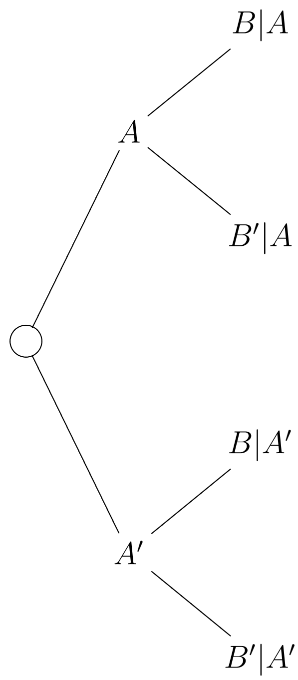
Note that this holds if and only if the events A and B are independent.

One way to get a better understanding of probability is through tree diagrams. Tree diagrams divide the sample space into mutually exclusive events.

Starting with an example, suppose we're getting oranges from two sources, Farmer Albert and Farmer Bob. Albert has a bigger orchard, and 70% of our oranges come from him, while only 30% come from Bob. But Bob is closer to us, and more of his oranges arrive undamaged; 95% as opposed to 90% of Albert's oranges.

What is the probability that an orange arrives damaged?

Consider the following:



Here, A represents the oranges from Albert and B represents the undamaged oranges. The damaged oranges are thus $B'|A \cup B'|A' = B'$. Since the events $B'|A$ and $B'|A'$ are mutually exclusive, $p(B') = p(B'|A)p(A) + p(B'|A')p(A') = .1 \cdot .7 + .05 \cdot .3 = .085$.

What is the probability that a damaged orange is from Albert?

This is exactly $p(A|B')$. By definition, this is $p(B' \cap A)/p(B')$. By the product rule, $p(B' \cap A) = p(A)p(B'|A) = .1 \cdot .7 = .07$. From above, $p(B') = .085$ so the final probability is 82%.

This is an example of Bayes' Theorem. In this special case, it says,

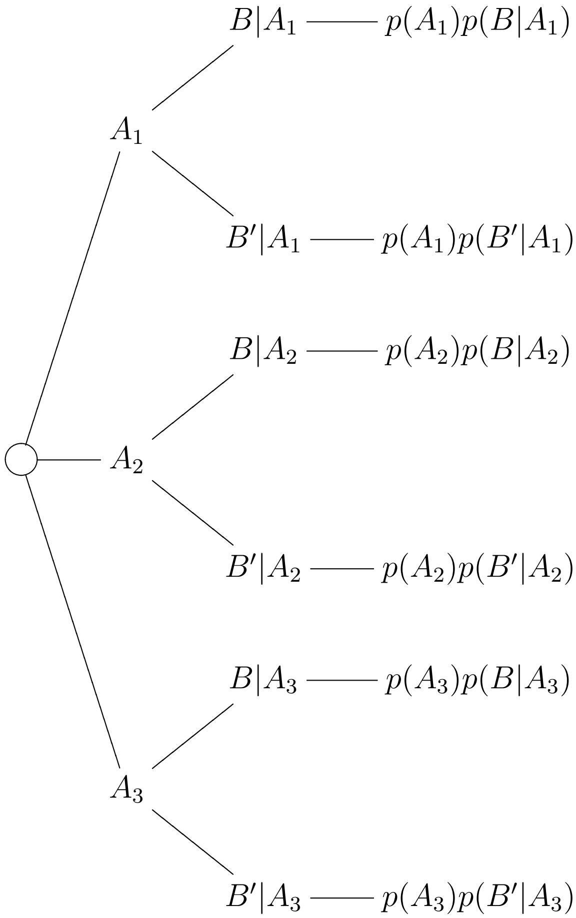
$$p(A|B) = \frac{p(A)p(B|A)}{p(A)p(B|A) + p(A')p(B|A')}.$$

The more general form allows for multiple events dividing B, say $A_1, A_2, \ldots, A_n$, all of which are mutually exclusive, and together they cover the sample space. The general form of the theorem is

$$p(A_i|B) = \frac{p(A_i)p(B|A_i)}{p(A_1)p(B|A_1) + p(A_2)p(B|A_2) + \cdots + p(A_n)p(B|A_n)}$$

For example, suppose we have three shipping methods, land, air, and sea. 2% of land cargo gets damaged, 4% of air cargo, and 12% of sea cargo. We use land 50% of the time, air 49% and sea 10%. What is the probability that cargo will arrive damaged? What is the probability that the damaged cargo came by land?

Consider A_1 the event that the cargo was shipped by land, A_2 the event that it was shipped by air, and A_3 the event that it was shipped by sea. Let B be the event that the cargo arrives damaged. The probability that cargo arrives damaged is $p(B)$ and the probability that damaged cargo comes by land is $p(A_1|B)$. We have the following tree diagram for this problem:



$$p(B) = p(A_1)p(B|A_1)+p(A_2)p(B|A_2)+p(A_3)p(B|A_3) = .5\cdot.02+.4\cdot.04+.1\cdot.12 = .038$$

So the probability that the cargo arrives damaged is 3.8%.

$$p(A_1|B) = \frac{p(A_1)p(B|A_1)}{p(B)} = \frac{.5 \cdot .02}{.038} = .263$$

So the probability that it was shipped by air is 26.3%.

Among Harvard men, 86% of fraternity house residents, 71% of fraternity members who do not reside at the fraternity house, and 45% of other men are binge drinkers. Of the male population, 10% are fraternity house residents, 15% fraternity members who don't live at the house, and 75% do not belong to a fraternity. What is the probability that a random male is a binge drinker?

This example is very similar to the previous example.

Homework Read 8.1,8.2
7.3 53,54
7.4 70,71
7.5 7-10,19,20,34,46
7.6 1-6, 25, 32,33
8.1 15,16