Stat 260 Lecture Notes
Set 6 - Conditional Probabilities

Example 1: Rolling a 6-sided die.

Suppose we roll a standard 6-sided die and record the number that is facing up. What is the probability of rolling a 3?

Now suppose we are told that an odd number was rolled. Now what is the probability of rolling a 3?

Instead, suppose we were told that an even number was rolled. What is the probability of rolling a 3?

Idea: Knowing extra information can change the probability of an outcome. When we know extra info it is called a *conditional probability*. The "given" part of the event is the extra info we are told.

Formula: probability of A given B (i.e. the probability that event A will occur given that we know event B has occurred).

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

Notice we could also talk about the probability of B given A:

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

Where does the formula come from?

Example 2: A maufacturer wants to see if using a clear coating of paint on their product is connected to if the product rusts.

	rust present (R)	no rust present ($\overline{R}$)	
clear coating used (C)	0.03	0.12	
no clear coating used ($\overline{C}$)	0.17	0.68	

If we know that a randomly selected component has a clear coating, what is the probability that it has rust present?

If we know that a randomly selected component does not have a clear coating, what is the probability that it has rust present?

Recall: The formula for conditional probability:

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

From these we get the multiplication rule:

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

We can **partition** an event A by looking at where it overlaps with event B: event A can overlap with event B, or it can overlap with event $\overline{B}$. That is, the events $A \cap B$ and $A \cap \overline{B}$ partition the event A. Note too that $A \cap B$ and $A \cap \overline{B}$ are disjoint events, so we can say

$$P(A) = P(A \cap B) + P(A \cap \overline{B}).$$

If we then use the multiplication rule we can say that

$$P(A) = P(A \cap B) + P(A \cap \overline{B}) = P(B) \cdot P(A|B) + P(\overline{B}) \cdot P(A|\overline{B}).$$

This is best illustrated on a tree diagram.

Example 3: Balding men and heart attacks.

A survey was taken of middle aged men and it was found that 28% of them are balding. Of those who are balding, there is an 18% chance that they will have a heart attack in the next 10 years. For those who are not balding, there is an 11% chance that they will have a heart attack within the next 10 years. What is the probability that a middle aged man will have a heart attack in the next 10 years?

The **law of total probability** says that to find $P(A)$, we add up the probabilities of a partition of A (i.e. add up all disjoint cases that arrive at A). In symbols:

$$P(A) = \sum_{i=1}^{\infty} P(A \cap B_i) = \sum_{i=1}^{\infty} P(B_i) \cdot P(A|B_i)$$

This is the same thing we did when we added together the different cases from the tree branches in the last example.

Bayes' Theorem puts together the conditional probability formula along with the multiplication rule. Using tree diagrams for these questions are very useful!

Example 4: TV sets.

TV sets are made at production plants A, B, and C. Suppose 50% are made at plant A, 30% are made at plant B, and 20% are made at plant C. Quality control finds that:

- 1% of plant A TVs are defective.
- 2% of plant B TVs are defective.
- 2% of plant C TVs are defective.

Given that a randomly selected TV is defective, what is the probability that it was produced at plant C?

Suppose we have events A_1, A_2, …, A_n, which are then followed by event B or $\overline{B}$. We can write Bayes' Theorem in symbols as:

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{P(A_1) \cdot P(B|A_1)}{\sum_{i=1}^{n} P(A_i) \cdot P(B|A_i)}$$

You don't need to write the symbolic form in your solution, it is much more important to know how to come up with this from using the conditional probability formula and a tree diagram like we did in Example 4.

Sometimes we perform diagnostic tests and the results shown are wrong.

There are 4 options for outcomes in a diagnostic test:

- The condition actually occurs and the test indicates positive for the condition occurring.
  This is a **true positive**, and no error occurs here.
- The condition actually occurs and the test indicates negative for the condition occuring.
  This is a **false negative**, this is an error.
- The condition does not actually occur and the test indicates positive for the condition occurring.
  This is a **false positive**, this is an error.
- The condition does not actually occur and the test indicates negative for the condition occurring.
  This is a **true negative**, and no error occurs here.

Example 5: Kidney transplants.

A patient receives a kidney transplant. Suppose that 35% of kidney transplants are rejected. During the healing process the patient is tested to see if they are rejecting the kidney. For this particular test the false positive rate is 8% and the false negative rate is 20%. What is the probability that the patient is rejecting the kidney if their test result is positive (i.e. the test indicates they are rejecting the kidney)?

Remember:

- false positive rate $= P(+|\overline{D})$
- false negative rate $= P(-|D)$
- true positive rate $= P(+|D)$. This is also sometimes called the sensitivity.
- true negative rate $= P(-|\overline{D})$. This is also sometimes called the specificity.