3. Conditional Probability

Chris Piech and Mehran Sahami

May 2017

1 Introduction

It is that time in the quarter (it is still week one) when we get to talk about probability. Again we are going to build up from first principles. We will heavily use the counting that we learned earlier this week.

2 Conditional Probability

In English, a conditional probability states "what is the chance of an event E happening given that I have already observed some other event F". It is a critical idea in machine learning and probability because it allows us to update our beliefs in the face of new evidence.

When you condition on an event happening you are entering the universe where that event has taken place. Mathematically, if you condition on F, then F becomes your new sample space. In the universe where F has taken place, all rules of probability still hold!

The definition for calculating conditional probability is:

Definition of Conditional Probability
The probability of E given that (aka conditioned on) event F already happened:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}$$

Lets turn to our visualization friend to get an intuition for why this is true. Again consider events E and F which have outcomes that are subsets of a sample space with 50 equally likely outcomes, each one drawn as a hexagon:

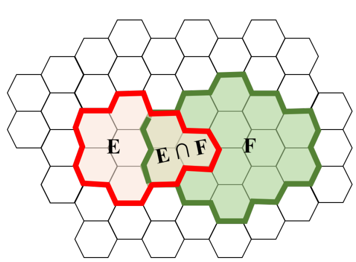

Figure 1: Conditional Probability Intuition

Conditioning on F means that we have entered the world where F has happened (and F, which has 14 equally likely outcomes, has become our new sample space). Given that event F has occured, the conditional probability that event E occurs is the subset of the outcomes of E that are consistent with F. In this case we can visually see that those are the three outcomes in $E \cap F$. Thus we have the:

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{3/50}{14/50} = \frac{3}{14} \approx 0.21$$

Even though the visual example (with equally likely outcome spaces) is useful for gaining intuition, conditional probability applies regardless of whether the sample space has equally likely outcomes!

The Chain Rule
The definition of conditional probability implies that:

$$P(E \cap F) = P(E|F)P(F)$$

which we call the Chain Rule. Intuitively it states that the probability of observing events E and F is the probability of observing F, multiplied by the probability of observing E, given that you have observed F. Here is the general form of the Chain Rule:

$$P(E_1 \cap E_2 \cap \cdots \cap E_n) = P(E_1)P(E_2|E_1)\ldots P(E_n|E_1 \ldots E_{n-1})$$

3 Independence Revisited

Our new understanding of conditional probability can give us a fresh insight into the independence. It also introduces a new concept: conditional independence.

First, lets consider the definition of conditional probability for independent events:

$$\begin{aligned} P(E|F) &= \frac{P(E \cap F)}{P(F)} && \text{Definition of conditional probability} \\ &= \frac{P(E)P(F)}{P(F)} && \text{Since } E \text{ and } F \text{ are independent} \\ &= P(E) && \text{Since the } P(F) \text{ terms cancel out} \end{aligned}$$

Similarly $P(F|E) = P(E)$. Not only does this give us a new formula when working with independent events, it gives another angle for understanding what independence means. If two events are independent, knowing that one event has occurred gives you no additional information that the other event will occurred. The do not affect one another.

3.1 Conditional Independence

Conditioning on any event can have a dramatic effect independence relationships of any pair of events. Events that were previously independent can become dependent. Events that were previously independent can become dependent.

Two events E and F are called conditionally independent given G, if

$$P(E \cap F|G) = P(E|G)P(F|G)$$

Or, equivalently:

$$P(E|F,G) = P(E|G)$$

3.2 Breaking Independence

Let's say a person has a fever if they either have malaria or have an infection. We are going to assume that getting malaria and having an infection are independent: knowing if a person has malaria does not tell us if they have an infection. Now, a patient walks into a hospital with a fever. Your belief that the patient has malaria is high and your belief that the patient has an infection is high. Both explain why the patient has a fever.

Now, given our knowledge that the patient has a fever, gaining the knowledge that the patient has malaria will change your belief the patient has an infection. The malaria explains why the patient has a fever, and so the alternate explanation becomes less likely. The two events (which were previously independent) are dependent when conditioned on the patient having a fever..

4 Law of Total Probability

Let us reconsider the idea that everything is a potato or not a potato. This suggests that

$$P(E) = P(E \cap F) + P(E \cap F^C)$$

The Law of Total Probability
If we combine our above observation with the chain rule, we get a very useful formula:

$$P(E \cap F) = P(E|F)P(F) + P(E|F^C)P(F^C)$$

5 Bayes Theorem

Bayes Theorem is one of the most ubiquitous results in probability for computer scientists. Very often we know a conditional probability in one direction, say $P(E|F)$, but we would like to know the conditional probability in the other direction. Bayes Theorem provides a way to convert from one to the other. We can derive Bayes Theorem by starting with the definition of conditional probability:

$$P(E|F) = \frac{P(F \cap E)}{P(F)}$$

Now we can expand $P(F \cap E)$ using the chain rule, which results in Bayes Theorem.

Bayes Theorem
The most common form of Bayes Theorem is:

$$P(E|F) = \frac{P(F|E)P(E)}{P(F)}$$

In the case where the demonimator is not known (the probability of the event you were initially conditioning on), you can expand it using the law of total probability:

$$P(E|F) = \frac{P(F|E)P(E)}{P(F|E)P(E) + P(F|E^C)P(E^C)}$$

A common scenareo for applying the Bayes Rule Tool is when you want to know the probability of something "unobservable" given an "observed" event. For example, you want to know the probability that a student understands a concept, given that you observed them solving a particular problem. It turns out it is much easier to first estimate the probability that a student can solve a problem given that they understand the concept and then to apply Bayes Therom.

6 Conditional Paradigm

As we noted before, when you condition on an event you enter the universe where that event has taken place. In that new universe all the laws of probability hold. Thus, as long as you condition consistently on the same event, every one of the tools we have learned still apply.