Chris Piech
CS109
Handout #7
April 4, 2016

# Conditional Probability

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.

## 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. The definition for calculating conditional probability is:

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

This equation implies that: $P(EF) = P(E \mid F)\ P(F)$ 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:

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

In the case where the sample space has equally likely outcomes:

$$P(E|F) = \frac{\#\text{ outcomes in E consistent with F}}{\#\text{ outcomes in S consistent with F}} = \frac{|EF|}{|SF|} = \frac{|EF|}{|F|}$$

## Bayes Theorem

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. There are a lot of things called Bayes Theorem. Here are the two most common equations:

Most Common Version:

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

Expanded Version:

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

## Data Science Application

Machine Learning (sometimes called Data Science) is the love child of: probability, data and computers. Sometimes machine learning involves complex algorithms. But often it's just the core ideas of probability applied to large datasets.

As an example let us consider Netflix, a company that has thrived because of well thought out machine learning. One of the primary probabilities that they calculate is the probability that a user will like a given movie given no other information about the user. We call this the prior.

Let $E$ be the event that a user likes a given movie. We can approximate $P(E)$ using the definition of probability from Friday's class:

$$P(E) = \lim_{n \to \infty} \frac{n(E)}{n}$$

Using this definition, we can approximate the probability by counting the number of users who liked the movie divided by the number of users who watched the movie. Since the number of users on Netflix is huge, this is a good approximation.

Another probability that Netflix cares about is the conditional probability of a user liking a movie Life is Beautiful (call that event E) given that they have already liked another movie Amelie (Event F). Combining the definition of conditional probability with the limit approximation of probability we get:

$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{\frac{\text{people who liked both}}{\text{people who watched both}}}{\frac{\text{people who liked amelie}}{\text{people who watched amelie}}}$$

## Bayes Theorem Example

For an example of Bayes Theorem (fully worked out and with intuition) see the demo on the course website: cs109.stanford.edu/demos/naturalBayes.html

*Disclaimer: This handout was made fresh just for you. Did you notice any mistakes? Let Chris know and he will fix them.*