$\boxed{\textbf{More Conditional Prob}}$

Conditional Probability

Often times, given the occurrence of an event, the probability of another event changes. The general formula for conditional probability is

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

Example - rolling a red die and a blue die. What is the probability that the two numbers on the dice sum to at least 11 given the information that the blue die has the value 5.

$P(E \cap F) = \frac{1}{36}$; the only way this happens is if the red die has the value 6.

$P(F) = \frac{6}{36}$; when the blue die is 6, the red die could be anything.

Therefore the conditional probability is $P(E|F) = 1/6$.

Using conditional probability to reason about an event

A student knows 80% of the material on a true-false exam. If the student knows the material, she has a 95% chance of getting it right. If the student does not know the material she just guesses and as expected has just a 50% chance of getting it right.

What is the probability of getting the question right?

In this type of problem we can basically define the following events

R - the event of getting the question right

K - the event of knowing the question

$P(R) = P(R \cap K) + P(R \cap K^c)$

$P(R \cap K) = P(R|K)P(K) = 0.95 * 0.8 = 0.76$

$P(R \cap K^c) = P(R|K^c)P(K^c) = 0.5 * 0.2 = 0.1$

Adding those you get a probability of 86% to get a question right.

Independence

The following conditions are independent and correspond to the criteria for declaring 2 events to be independent.

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

As a really small example (more in next lecture) consider the event of rolling two dice of different color (red and yellow). We want to show the event "total number of dots on top is odd" is independent of the event "the red die has an odd number of dots on top".

recitation questions

1. Your neighbour has 2 children. You learn that he has a son, Joe. What is the probability that Joe's sibling is a brother?

   Consider the sample space in this question. You know there are 2 children so the possibilities are $\{BB, GG, BG, GB\}$. The reason we want to include both GB and BG is because one represents that the first born is a boy and the other represents the first boy is a girl.

   The event E that the neighbour has a son Joe is the set $E = \{BB, BG, GB\}$.

   The event F that the neighbour has two sons, which is the same thing as saying that Joe has a brother is just the set $F = \{BB\}$

   $P(F|E) = \frac{P(F \cap E)}{P(E)} = \frac{1/4}{3/4} = \frac{1}{3}$

2. Your neighbour has 2 children. He picks one of them at random and decides to visit your house. The kid he brings is his son named Joe. What is the probability that Joe's sibling is a brother?

   The sample space does not change but what has changed is the way in which the gathering of information has occurred. The father has done a random sampling before you find out information. What changes?

   Let the event E' be your neighbour randomly chose a child and that happened to Joe. Event E' does imply the event E that we stated before. If the father randomly picks Joe then you do get the information that your neighbour has a son. However the other way round does not work. If you know that he has a son Joe, that does guarantee that he random picks Joe to bring over to your house.

$$\begin{aligned} P(E') = P(E'|\{BB\})P(\{BB\}) + P(E'|\{BG\})P(\{BG\}) + P(E'|\{GB\})P(\{GB\}) \\ + P(E'|\{GG\})P(\{GG\}) \end{aligned}$$

The last term is 0 so can be eliminated right away. For the others, if there is a girl and a boy there is a 0.5 chance of bringing the boy over.

$P(E') = 1 * \frac{1}{4} + 0.5 * \frac{1}{4} + 0.5 * \frac{1}{4}$

Therefore the conditional probability

$P(F|E') = \frac{P(F \cap E')}{P(E')} = \frac{P\{BB\}}{P(E')} = \frac{1/4}{1/2} = \frac{1}{2}$

Bayes Theorem

Suppose that F and X are events from a common sample space and $P(F) \neq 0$ and $P(X) \neq 0$. Then

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

where F^c is just the event of F not happening.

Example - Taken from UWash notes

In Orange County, 51% of the adults are males. (It doesn't take too much advanced mathematics to deduce that the other 49% are females.) One adult is randomly selected for a survey involving credit card usage.

1. Find the prior probability that the selected person is a male.
2. It is later learned that the selected survey subject was smoking a cigar. Also, 9.5% of males smoke cigars, whereas 1.7% of females smoke cigars (based on data from the Substance Abuse and Mental Health Services Administration). Use this additional information to find the probability that the selected subject is a male.

Let us use some notation to simplify the analysis

M = male . Thereby M^c = female.

S = cigar smoker. And of course this means S^c = not a cigar smoker.

$P(M) = 0.51$

The second part is basically asking us to compute $P(M|S)$.

$$P(M|S) = \frac{P(M)P(S|M)}{P(S|M^c)P(M^c) + P(S|M)P(M)}$$

We are given $P(S|M)$ since 9.5% males smoke. Similarly $P(S|M^c)$ has been provided since 1.7% females smoke.

$$P(M|S) = \frac{0.51 * 0.095}{0.51 * 0.095 + 0.49 * 0.017} = 0.853$$

So there is an 85.3% chance that it is a male if you observe cigars! Stands to reason. Men smoking is a far more likely event than women smoking (based on the data).

Application - Bayesian Machine Learning

The primary goal of machine learning is learning models of data. The Bayesian framework for machine learning states that you start out by enumerating all reasonable models of the data and then assign what is called a prior probability $P(M)$ to each of these models. Then you go out and collect some data (that part has become easy these days!).

Apply Bayes Theorem to get

$$P(M|D) = \frac{P(D|M)P(M)}{P(D)}$$

Once you have observed the data D, you evaluate how probable the data was under each of these models to compute $P(D|M)$. Multiplying this probability by the prior and renormalizing gives you what is called the posterior probability, $P(M|D)$, which encapsulates everything that you have learned from the data regarding the models under consideration.

How do you compare two models M and M'? We compute their relative probability given the data: $P(M)P(D|M)$ and $P(M')P(D|M')$.

This leads to a very common technique called MAP estimation which stands for maximum a posteriori. Taken from wikipedia - MAP.