$\boxed{\text{Conditional Prob}}$

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.

Define the events

E - event of the two dice summing to 11.

F - blue die has the value of 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

Conditional probability can be used to split up the probability of an event conditioned by the occurrence of some other 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 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 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".

Truel

For discussion on the Truel see
http://www.mathgoespop.com/2009/10/martin-gardner-and-the-three-way-duel.html

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. 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).

Bayes Theorem and medical tests

Bayes theorem is used to evaluate probabilities for medical tests. Here is a questions as an example

There is a medical test that has been designed to screen for a disease that affects 5 in 1000 people. Suppose the false positive rate is 3% and the false negative rate is 1%.

What is the probability that a randomly chosen person who tests positive actually has the disease?

Solution

We first need to clearly understand what false positive and false negative mean. Both these terms are used in detection systems and it is terminology that is commonly used for classifiers in machine learning as well.

False positive - the test says the person has the disease when the person actually does not have it. The word 'when' is the same as the word 'given'.

False negative - the test says the person does not have the disease when the person actually does have it. The word 'when' is the same as the word 'given'.

Now to solve this question, let us define some events

- Let A be the the event that the person tests positive for the disease.
- B_1 be the event that a randomly chosen person has the disease.
- B_2 be the event that a randomly chosen person does not have the disease.

Now the question is asking $P(B_1|A)$.

$$P(B_1|A) = \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)}$$

Now $P(B_1)$ is just the probability that a randomly selected person has a disease. So 0.005. Also $P(B_2)$ is just the probability that a randomly selected person does not have the disease. So 0.995.

$P(A|B_1)$ = 1 - probability of a false negative = 1- 0.01= 0.99

$P(A|B_2)$ = probability of a false positive = 0.03

Plugging all these values back into the original

$$\begin{aligned} P(B_1|A) &= \frac{0.99 * 0.005}{0.099 * 0.005 + 0.03 * 0.995} \\ &= 0.1422 \\ &= 14.22\% \end{aligned}$$

What is the probability that a randomly chosen person who tests negative does not actually have disease

This is now asking us to compute $P(B_2|A^c)$ which can again be broken by Bayes theorem as

$$P(B_2|A^c) = \frac{P(A^c|B_2)P(B_2)}{P(A^c|B_2)P(B_2) + P(A^c|B_1)P(B_1)}$$

Similar to our previous reasoning

$P(A^c|B_2)$ = 1 - probability of a false positive = 1 - 0.03 = 0.97

$P(A^c|B_1)$ = probability of a false negative = 0.01

Plugging these values into the above equation we get a probability of 99.99%.

Therefore we have a very high probability that someone who tests negative actually does not have the disease. We have a low probability that someone who is testing positive actually has the disease. The rationale behind this type of test is that as a screening test it is more important to make sure that the negative results are truly negative.

Those that test positive are asked to take a test where the probability of detecting the disease when the person has the disease is much higher. But those tests might be expensive and the cheaper screening test ensures that not all of the population has to pay for those expensive tests.

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.