Conditional probability

February 6, 2009

This handout recaps the class of 1/30—definition of conditional probability and the Prosecutor’s Fallacy.

1 Recap of 1/30

The motivation for defining conditional probabilities is that, often, it makes sense to ask what happens given that something else has already happened. Sometimes, asking for conditional probabilities gives a much clearer picture of what is truly happening. For example, if one is asked what is the probability the ten of you show up in Holmes 247 at the same time, it doesn’t seem very likely. On the other hand, conditioned on the fact that you ten have registered for the class, it is very likely you will be in Holmes 247 MWF. Or if you are on a jury, you consider what is the probability the accused is guilty given the testimony presented in court.

We defined the conditional probability of event $A$ given $B$ as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$ (1)

The above equation implies that

$$P(A \cap B) = P(A|B)P(B).$$ (2)

HINT: write each conditional probability as the ratio of regular probabilities.

To recap the example we did in class, we let $\Omega$ be the set of all chips produced at a hypothetical factory. The factory uses two processes—a high quality process and a medium quality process. Let us associate a variable $Q$ with each chip: $Q = 1$ indicates the chip has been manufactured with the high quality process, and $Q = 0$ indicates the chip has been manufactured with the medium quality process. A chip could be defective or not, and we indicate it with another variable $D$ ($D = 1$ indicates the chip is defective, $D = 0$ indicates the chip is not defective).

To make up numbers, we assumed that a high quality process produces chips which are defective with probability only .01, while the medium quality process produces chips which are defective with probability .1. As you answered in class, these statements mean that $P(D = 1|Q = 1) = .01$ and $P(D = 1|Q = 0) = .1$.

Furthermore we said the chances the high quality process is used is .25, namely $P(Q = 1) = .25$. Please note that this is different from saying 1/4 of the chips produced used the high quality process, though in class, we pointed this out but then pretended not to make the distinction. The distinction is explained in the handout “Probability and Empirical Frequency” posted on Feb 1.

Now, using equation (2)

$$P(D = 1, Q = 1) = P(D = 1|Q = 1)P(Q = 1) = .01 \times .25 = .0025,$$ (3)

$$P(D = 1, Q = 0) = P(D = 1|Q = 0)P(Q = 0) = .1 \times .75 = .075.$$ (4)

Note again that when we write $P(D = 1, Q = 1)$, we mean the probability a chip is associated with $D = 1$ and $Q = 1$. Likewise for $P(D = 1, Q = 0)$. 

1
What if one is asked $P(D = 1)$? A defective chip can only be manufactured using one of two processes—the high or medium quality process. We make up two categories of defective chips, those associated with $Q = 1$ and those with $Q = 0$. When we write $P(D = 1)$, we mean $P(D = 1, Q = 0$ or $D = 1, Q = 1$—we drop the term corresponding to $Q$ in the notation simply because so long as $D = 1$, we do not care what $Q$ is.

But to compute $P(D = 1)$, we want to consider these two cases separately. These two cases are clearly disjoint (since no chip can be associated with both $Q = 0$ and $Q = 1$), and together they make up all defective chips, namely all chips associated with $D = 1$. Therefore,

$$P(D = 1) = P(D = 1, Q = 0 \text{ or } D = 1, Q = 1) = P(D = 1, Q = 0) + P(D = 1, Q = 1) = .075 + .0025 = .0775$$

where we use (3) to write values for $P(D = 1, Q = 0)$ and $P(D = 1, Q = 1)$.

This is a general principle. Suppose you have to compute the probability of some event $A$. Suppose the outcomes have another variable $B$ associated with it. Let the value of the variable $B$ be among $B_1, \ldots, B_m$, but each outcome in $A$ is associated with exactly one of $B_1, \ldots, B_m$. Then,

$$P(A) = P(A, B = B_1 \text{ or } B_2 \cdots \text{ or } B_m)$$

$$= P(A, B = B_1) + P(A, B = B_2) + \ldots + P(A, B = B_m)$$

$$= \sum_{i=1}^{m} P(A, B = B_i).$$

Note the last line is just a notation for the previous line—once you get used to it, it is more transparent.

From (2), for all $i$,

$$P(A, B = B_i) = P(A|B = B_i)P(B_i),$$

therefore,

$$P(A) = \sum_{i=1}^{m} P(A, B = B_i) = \sum_{i=1}^{m} P(A|B = B_i)P(B_i).$$

### 2 Prosecutor’s Fallacy

The notion of conditional probabilities has often been misinterpreted. In courts of law, given some testimony, the jury or the judge has to decide on the probability the accused is guilty or not guilty. Let $G$ denote the guilt of the accused—$G = 0$ means the accused is not guilty while $G = 1$ means the accused is guilty. Let $T$ denote the testimony present. The jury decides if $P(G = 0|T)$ is negligible or not. According to the law in most countries, if $P(G = 0|T)$ is not very small, then there is reasonable doubt on the accused being guilty and the accused is acquitted. It is often very easy to get this wrong, as this example will show.

In a town Kangaroo of 50,000 there has been a murder. The police find hair and blood of the perpetrator at the crime scene and the forensics team isolates the DNA of the perpetrator. The police then go door to door till they find a match with the DNA at the crime scene. The testimony that is supposed to damning is the DNA match.

The defence is flummoxed. They bring a whole line of witnesses who claim the accused is very non-violent, an upstanding member of the community who even runs an orphanage and a dog shelter.

The new young prosecutor Overly Eager is thrilled he has what he thinks is a slam dunk case. He gets the town doctor who comes over and says the probability that someone innocent has matching DNA with the crime scene is only 1/10000. In his closing argument, he argues that 1/10000 is such a small number that it really doesn’t matter what the friends of the accused say.

You, Probabistically Suave, are on the jury. The judge asks the jury to deliberate. He adds that if you have reasonable doubt, you acquit the accused but that 1/10000 does not constitute reasonable doubt.

What should your decision be in the Kangaroo court?

There are several stories of this nature, a very famous one happening in France towards the end of the 19th century. A legionary of the French army, Alfred (?) Dreyfus was accused of being a German spy. The testimony was a letter written by him, which the prosecution argued was a code.
The prosecutor drew a grid with 1cm spacing on the paper, and showed that there were 4 cases of one sequence of syllables aligning up. The prosecutor’s case was that the probability of even a single alignment was tiny (to make up for my memory lapse, let us just say it is .1). Therefore, the prosecutor argued that the 4 alignments could have happened by chance was only $1^4 = .0001$, a very small number—thus indicating it must have been deliberate and therefore that the letter was some sort of a code.

The court convicted Dreyfus on this evidence. Do you agree with the prosecutor’s argument from just a probabilistic perspective?