Lecture 4: Conditional Probability, Total Probability, Bayes's Rule

12 September 2005

1 Conditional Probability

How often does A happen if B happens? Or, if we know that B has happened, how often should we expect A?

Definition:

$$\Pr(A|B) \equiv \frac{\Pr(A \cap B)}{\Pr(B)}$$

Why? Go back to the counting rules. The probability of A is $\mathrm{Num}(A)/N$. But if we *know* B has happened, only those outcomes count, so we should replace the denominator by $\mathrm{Num}(B)$, and the numerator by $\mathrm{Num}(A \cap B)$. Now divide numerator and denominator by N to get the definition.

All of the probability rules have their conditional equivalents

$$\Pr(B|B) = 1$$

because $\Pr(B \cap B)/\Pr(B) = \Pr(B)/\Pr(B)$

$$\Pr(A|B) = 1 - \Pr(A'|B)$$

because $\Pr(B) = \Pr(A \cap B) + \Pr(A' \cap B)$; divide both sides by $\Pr(B)$ to get $1 = \Pr(A|B) + \Pr(A'|B)$.

$$\Pr(A \cup C|B) = \Pr(A|B) + \Pr(B|C) - \Pr(A \cap C|B)$$

2 Mutually Exclusive and Jointly Exhaustive Events

The events A and A' are **mutually exclusive**: if one happens, the other can't. The mathematical expression of this is that $A \cap A' = \emptyset$, so $\Pr(A \cap A') = 0$.

The events A and A' are also **jointly exhaustive**: one or the other of them must happen. Symbolically, $A \cup A' = \mathcal{S}$, and $\Pr(A \cup A') = 1$.

Because A and A' are mutually exclusive and jointly exhaustive, we say that they **partition** the sample space $\mathcal{S}$, or are a partition of the sample space. We can consider partitions with more than two events, however. We might want to partition the sample space up into any number of distinct events. A collection of events $A_1, A_2, ...A_k$ will be a partition if they are all mutually exlcusive,

$$A_i \cup A_j = \emptyset \text{ (unless } i = j)$$

and jointly exhaustive

$$\bigcup_{i=1}^{k} A_i = \mathcal{S}$$

Think back to the wind-tunnel velocity data from the second lecture. We could partition the sample space into velocities above or below the mean velocity, but we could also use a three-event partion: within one standard deviation of the mean ($\overline{v}-\sigma \leq v \leq \overline{v}+\sigma$), more than one standard deviation above the mean ($v > \overline{v}+\sigma$), and more than one standard deviation below the mean ($v < \overline{v}-\sigma$). The events $v < \overline{v} - \sigma$ and $v > \overline{v} + \sigma$ are mutually exclusive, but not jointly exhausitve. Their complements, the events $v > \overline{v} - \sigma$ and $v < \overline{v} + \sigma$, are jointly exhausitve, but not mutually exlcusive.

3 Total probability

Suppose $A_1, \ldots A_k$ are mutually exclusive and jointly exhaustive, so that $\bigcup_i A_i = \mathcal{S}$. Then

$$\bigcup_i B \cap A_i = B \cap \left(\bigcup_i A_i \right) = B \cap \mathcal{S} = B$$

But $B \cap A_i$ and $B \cap A_j$ are also mutually exclusive (unless $i = j$), because $(B \cap A_i) \cap (B \cap A_j) = B \cap A_i \cap A_j = B \cap (A_i \cap A_j)$. So

$$\begin{aligned} \Pr(B) &= \Pr\left(\bigcup_i B \cap A_i \right) = \sum_i \Pr(B \cap A_i) \\ &= \sum_i \Pr(B|A_i) \Pr(A_i) \end{aligned}$$

This is called the rule of total probability. It becomes useful when we have easy ways of calculating $\Pr(A_i)$ and $\Pr(B|A_i)$, but need to find $\Pr(B)$.

4 Bayes's Rule

Let's start by re-arranging the definition of conditional probability.

$$\Pr(A \cap B) = \Pr(A|B) \Pr(B)$$

But we can switch the roles of A and B.

$$\Pr(A \cap B) = \Pr(B|A) \Pr(A)$$

Equating the right hand sides,

$$\Pr(B|A)\Pr(A) = \Pr(A|B)\Pr(B)$$

We can then solve for $\Pr(A|B)$.

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

This is Bayes's rule. It lets us invert conditional probabilities, going from $\Pr(B|A)$ to $\Pr(A|B)$. It's very handy, but it's fundamentally very simple. The trickiest bit is often computing the denominator, $\Pr(B)$, but that's why we have the rule of total probability.

4.1 Homer Simpson vs. Bayes's Rule

The alarm system at a nuclear power plant is not *completely* reliable. If there is something wrong with the reactor, the probability that the alarm goes off is 0.99. On the other hand, the alarm goes off on 0.01 of the days when nothing is actually wrong. Suppose that something is wrong with the reactor only one day out of 100. What is the probability that something is actually wrong if the alarm goes off?

Let $A =$ "something is wrong with reactor"
$B =$ "alarm goes off"

$$\begin{aligned}
\Pr(B|A) &= 0.99\\
\Pr(B|A') &= 0.01\\
\Pr(A) &= 0.01
\end{aligned}$$

We desire $\Pr(A|B)$, the probability that the alarm indicates an actual fault.

$$\Pr(A|B) = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)}$$

$\Pr(B|A)$ is given (it's the reliability of the alarm when something is wrong, 0.99), as is $\Pr(A)$ (the frequency of accidents, 0.01). We need to find $\Pr(B)$, and to do so we use total probability.

$$\Pr(B) = \Pr(B|A)\Pr(A) + \Pr(B|A')\Pr(A')$$

Remember A and A' are mutually exclusive and jointly exhaustive.

$$\begin{aligned}
\Pr(B|A') &= 0.01\\
\Pr(A') &= 1 - \Pr(A) = 0.99\\
\Pr(B) &= 0.99 * 0.01 + 0.01 * 0.99 = 2 * 0.01 * 0.99\\
\Pr(B|A)\Pr(A) &= 0.99 * 0.01\\
\Pr(A|B) &= 0.99 * 0.01/(2 * 0.01 * 0.99) = 1/2
\end{aligned}$$

In other words, the alarm is right only half the time!

Now, this way of doing the problem is one many people find somewhat confusing. Here's another way, which is completely equivalent, but which many people find easier.

Think about ten thousand days of the reactor. (Any large number will do, but this way all the numbers come out even.) On how many days will the reactor have trouble? Well, that's $10000 \times \Pr(A) = 100$ days. On how many of those days when there's trouble will the alarm go off? That's $\Pr(B|A) \times 100$ days $= 99$ days. On the other hand, there are 9900 days when there's no trouble. On how many of them does the alarm go off? That's $\Pr(B|A') \times 9900 = 99$ days. So how likely is it that an alarm day is also a problem day? Well, just 1/2, because there are as many false alarm days as real alarm days.

4.1.1 Don't Worry about Your Positive Cancer Test?!

Another canonical example is medical testing. Suppose only one person in 100 has some rare disease, but the test is 99% accurate. What is the probability that someone who tests positive actually has the disease? The exact same reasoning as in the nuclear reactor example says that probability is only 50%.

In fact, suppose the disease is very rare, so that only 1 person in 1 million has it, but the test is still 99% accurate. What's the probability that someone who tests positive has it?

$$\begin{aligned}
\Pr(A|B) &= 0.99 * (10^{-6})/[0.99 * 10^{-6} + 0.01 * (1 - 10^{-6})] \\
&= 0.99 * (10^{-6})/[0.99 * 10^{-6} + 0.01 - 0.01 * 10^{-6}] \\
&= 0.99 * (10^{-6})/[0.98 * 10^{-6} + 1 * 10^{-2}] \\
&= 0.99/10,000.98 \\
&= 9.9 * 10^{-5}
\end{aligned}$$

or about 1 in 10,000. This is much higher than the baseline level, but still much less than one. In fact, this just gets worse and worse as the disease gets rarer — as $\Pr(A)$ shrinks, $\Pr(A|B)$ does too. So — should most people who think get medical tests saying they have a rare disease not worry?

Answer: No, because we don't (generally) test random samples of the population for rare diseases — we test people whom other symptoms suggest have the disease. So the relevant $\Pr(A)$ isn't the fraction of the population at large with the disease, but the fraction of people *tested* who have the disease.