Conditional probability

Math 477

September 25, 2014

1 Definition and examples

1.1 Motivating example

Suppose we toss a coin twice. What is the probability that we get 2 tails? From the above, it's $\frac{1}{4}$. Suppose, however, that you know the additional information that the first toss is a tail. We ask the same question: what is the probability that we get 2 tails? Clearly it's no longer $\frac{1}{4}$, because for you, the set of *all possible events* have changed. Namely, the outcomes $\{HH\}, \{HT\}$ are no longer possible.

Concretely, the set of all possible outcomes now are:

$$\{TT\}, \{TH\}.$$

Thus the probability that you get 2 tails is $\frac{1}{2}$. We say: the probability that we get 2 tails, *conditioned on* the first toss being a tail, is $\frac{1}{2}$.

1.2 Conditional probability

Definition 1.1. *Let A, B be events. If $P(A) > 0$, the probability of B conditioned on A, or B given A, denoted $P(B|A)$, is defined as:*

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

The interpretation is that we have already had the knowledge that A happened. So the probability of the event B happening, given that A has happened, should be calculated as given in the definition.

Example 1.2. *We toss a die. What is the probability that we get a 6, given that we know the toss is even?*

Ans: Let A be the event that we get an even toss, B the event that we get a 6 (when you get used to this, you don't have to explicitly name out the events). Then $P(A) = 1/2$, $P(A \cap B) = P(B) = 1/6$. Thus $P(B|A) = 1/3$.

1.3 How to tell you're dealing with a conditional probability question

The key signal to using conditional probability, rather than probability is *the persence of the word if in a question.* Note that many questions *may not spell out the conditional probability key word* for you, assuming the understanding of the context. Consider the following example:

Example 1.3. *Joe is 80 % sure he places his key in one of the two pockets of his jacket. He is 40 % sure he places in his right pocket and 40 % sure he places it in his left pocket.* **If** *he didn't find the key in his left pocket, what is the probability it is in the other pocket?*

Ans: Let L be the event the key is in the left pocket, R the event the key is in the right pocket. Then we're asked for $P(R|L^c)$. We have

$$P(R|L^c) = \frac{P(RL^c)}{P(L^c)} = \frac{P(R)}{1 - P(L)} = \frac{2}{3}.$$

Remark: The question may be a bit confusing, as we did NOT use the information that Joe is 80 % sure he places the key in one of the pockets. Turns out this information is redundant, because we can deduce it from the information:

$$P(\text{ In jacket}) = P(\text{ Left pocket}) + P(\text{ Right pocket}) = .4 + .4 = .8$$

Alternative question:

Example 1.4. *Joe is 80 % sure he places his key in one of the two pockets of his jacket. If the key is in his jacket, he is 30 % sure he places in his right pocket and 70 % sure he places it in his left pocket.* **If** *he didn't find the key in his left pocket, what is the probability it is in the other pocket?*

2 The multiplication rule

Let $E_1, E_2, \cdots, E_n$ be events. Observe that by definition of conditional expectation:

$$P(E_1E_2\cdots E_n) = P(En|E_1E_2\cdots E_{n-1})P(E_1E_2\cdots E_{n-1}).$$

Now

$$P(E_1E_2\cdots E_{n-1}) = P(E_{n-1}|E_1E_2\cdots E_{n-2})P(E_1E_2\cdots E_{n-2}).$$

Repeating this procedure we have the following **multiplication rule**

$$P(E_1E_2\cdots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\cdots P(E_n|E_1E_2\cdots E_{n-1}).$$

The multiplication allows us to compute the intersection of event, or events of the type A *and* B. Usually you intuitively "multiply" "probability" of A with "probability" of B to get probability of A and B, even in situations where these events may not be independent. Surprisingly you still get the right result. The reason is accidentally you have used the multiplication rule without realizing it, and the "probability" you used is actually conditional probability. The following examples will demonstrate.

Example 2.1. *An urn has 12 balls, 8 of which are white and the rest are blue. We select 4 balls without replacement. What is the conditional probability that the 1st and 3rd ball are white, given that there are 3 white balls selected among the 4? Do the same problem where we select the balls with replacement.*

Ans: We're looking for $P(W_1W_3|3W)$. By definition

$$P(W_1W_3|3W) = \frac{P(W_1W_3, 3W)}{P(3W)}.$$

We have

$$P(3W) = \frac{\binom{8}{3}\binom{4}{1}}{\binom{12}{4}}.$$

Now

$$P(W_1W_3, 3W) = P(W_1W_2W_3B_4) + P(W_1B_2W_3W_4).$$

We calculate 1 term in the RHS above for demonstration.

$$P(W_1B_2W_3W_4) = P(W_4|W_1B_2W_3)P(W_3|W_1B_2)P(B_2|W_1)P(W_1).$$

We can calculate all the terms in the RHS above.

$$\begin{aligned}
P(W_1) &= \frac{8}{12}; \\
P(B_2|W_1) &= \frac{4}{11}; \\
P(W_3|W_1B_2) &= \frac{7}{10}; \\
P(W_4|W_1B_2W_3) &= \frac{6}{9}.
\end{aligned}$$

Example 2.2. *Again consider the problem of 10 people picking their hats at random. We want to compute $P(E_1E_2)$. This answer is $\frac{8!}{10!} = \frac{1}{90}$. You may also reason the probability that the first person gets his right hat is $\frac{1}{10}$. After the first person gets his right hat, then the probability the second person gets his right hat is $\frac{1}{9}$. Thus the probability that they both get their right hats is $\frac{1}{9}\frac{1}{10} = \frac{1}{90}$. You have used the multiplication rule:*

$$P(E_1) = 1/10, P(E_2|E_1) = 1/9, P(E_1E_2) = P(E_2|E_1)P(E_1) = 1/91/10 = 1/90.$$

2.1 Application

The multiplication rule illustrates an important idea in computing probability of an event: by conditioning on other events, the problem might become more tractable or easier to handle. Of course this doesn't always have to be the case, but at least it provides us with some alternatives in solving a problem.

We illustrate this technique by considering the problems of n people randomly picking their hats. Now the question is what is the probability of *exactly* k people having correct hats?

Let F be the event that none of the $k+1$th to nth people got their right hat. Using the same notation as before, we compute $P(E_1E_2\cdots E_kF)$. Then the answer we want will be $\binom{n}{k}P(E_1E_2\cdots E_kF)$ since out of n people we choose k to have their correct hats, and the rest to have incorrect hats. The probability of any such event is equal.

First of all, note that

$$P(E_1E_2\cdots E_kF) = P(F|E_1E_2\cdots E_k)P(E_1E_2\cdots E_k).$$

How to compute $P(F|E_1E_2\cdots E_k)$? Conditioning on the event that the first k get their right hats, you can easily see that the probability of the rest of $n-k$ getting the wrong hats is the same as (without any conditioning) the probability of $n-k$ people getting their wrong hats, which we found out to be

$$\sum_{j=2}^{n-k}(-1)^j\frac{1}{j!}.$$

Now we just have to compute $P(E_1E_2\cdots E_k)$. Using the multiplication rule, we see that

$$P(E_1E_2\cdots E_k)=P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\cdots P(E_k|E_1E_2\cdots E_{k-1}).$$

Now $P(E_1)=\frac{1}{n}$ as we have explained. Given that the first person got his correct hat, the probability the second person got his correct hat is just $\frac{1}{n-1}$. Similarly, given the first 2 got their correct hats, the probability the 3rd got his correct has is $\frac{1}{n-2}$. Thus

$$P(E_1E_2\cdots E_k)=\frac{1}{n(n-1)\cdots(n-k)}=\frac{(n-k)!}{n!}.$$

So the original probability is

$$\begin{aligned}P(\text{exactly k correct hats}) &= \binom{n}{k}P(E_1E_2\cdots E_kF)=\binom{n}{k}\frac{(n-k)!}{n!}\sum_{j=2}^{n-k}(-1)^j\frac{1}{j!}\\ &= \frac{\sum_{j=2}^{n-k}(-1)^j\frac{1}{j!}}{k!}\end{aligned}$$

Note that there is a shorter way to solve this problem, by counting and using the result we got before. We count how many ways we can assign the right hats to the first k people and wrong hats to the last $n-k$. This is not an easy combinatorics problem to solve fresh, but using our result last time, we see that there are $(n-k)!\sum_{j=2}^{n-k}(-1)^j\frac{1}{j!}$ ways to assign wrong hats to $n-k$ people (since the probability of them getting wrong hats is $\sum_{j=2}^{n-k}(-1)^j\frac{1}{j!}$ and the sample size is $(n-k)!$, being the total number of ways we assign hats to $n-k$ people). There is only 1 way to assign correct hats to the first k people, thus the probability

$$P(E_1E_2\cdots E_kF)=\frac{(n-k)!\sum_{j=2}^{n-k}(-1)^j\frac{1}{j!}}{n!},$$

as the original sample size is $n!$. This is exactly what we got above. So conditioning may not be the shortest way to solve a problem, but it gives us some alternative techniques. You should also look at the above example as illustrating thinking in conditional probability terms.

3 $P(A|B)$ versus $P(AB)$

It is easy to get confused between $P(A|B)$ and $P(AB)$? What is their difference (in meaning) and how do we know when to use which one? In meaning, $P(A|B)$ asks for the probability that A happens, if you know B has happened. $P(AB)$ asks for the probability that A and B happen together. Thus the difference is in terms of information. In $P(AB)$ you do not know whether B has happened or not. In $P(A|B)$ you do.

(It may be helpful, in computing $P(A|B)$ to imagine B has happened before A, even though this may not be the case in reality.)

Again, if there is a word *if* then it is a signal to use conditional probability. Otherwise it would be a regular $P(AB)$ computation.

Finally to illustrate the difference, we look at the hat problem again. There we can consider $P(E_1E_2)$ and $P(E_2|E_1)$. To compute $P(E_1E_2)$, we look at the events both 1st and 2nd people getting their right hats at the same time, among the n hats. We explained that there are $\binom{n}{2}2!$ ways of choosing 2 hats out of n, with regards to order. Thus the probability $P(E_1E_2)$ is $\frac{1}{n(n-1)}$.

On the other hand $P(E_2|E_1)$ is simply $\frac{1}{n-1}$ because after the 1st person got his right hat, the problem is as if we look at $n-1$ people choosing their hats, and asking for the 2nd person (the 1st among the remaining $n-1$) choosing his right hat.

4 Bayes' rule

From the definition of conditional probability, we have

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

It is clear that

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

Thus

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

We also have

$$P(B) = P(BA) + P(BA^c) = P(B|A)P(A) + P(B|A^c)P(A^c).$$

Therefore we conclude

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}. \tag{1}$$

This formula is called the Baye's rule. At first glance this is pure mathematical manipulation. But it has an important implication: that of switching what we conditioned on. An example would illustrate what this means.

It is well-known that medical test is not 100% reliable. That is suppose you test for a disease, which has 1% chance of happening, then even if the test comes out negative, it doesn't mean you have 0% of contracting the disease. Instead, with a very small probability, it could be a false negative. Concretely, suppose that if you indeed have the disease, then there is 98% chance that the test comes out positive, and 2% negative. However, suppose you don't have the disease, there is 95% chance the test comes out negative, and 5% chance it comes out positive. Now you go for the test, and it comes out negative. What is the probability that you contract the disease?

Ans: Let A be the event that you contract the disease and B be the event that the test is positive. Then we have

$$P(B|A) = .98, P(B^c|A) = .02, P(B|A^c) = .05, P(B^c|A^c) = .95.$$

The question asks for $P(A|B^c)$. Thus you see how Bayes' rule is appropriate for the situation. Can you figure out what it is?