20200123

Probability Continued: Conditional Probability

Let A and B be two events on the same sample space (S, P) (from now on, when I say "Let A and B be two events", I will always mean "on the same sample space" unless I indicate otherwise).

We can speak of A **occurring** - we mean that the experiment underlying the sample space has been executed once, and the observed outcome of the experiment is an element of A. The point here is that for an event to occur, there must have been a sampling operation resulting in an outcome.

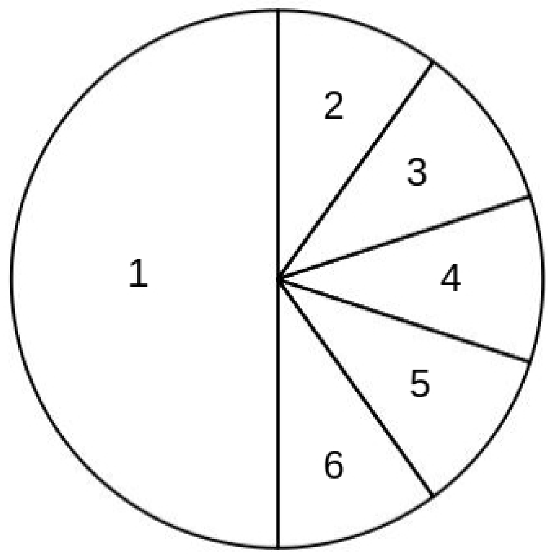
If the experiment is executed and we are told the outcome is some specific value such as k, then we can determine whether or not A occurred simply by checking to see if $k \in A$. But suppose we are only given partial information about the outcome such as "B occurred" ... can we determine if A occurred or not? If we can't be sure that A did or did not occur, can we assign a probability value to A based on the partial information we have?

It's really important to understand this question. We're saying "Suppose we conduct the experiment many many many times. We ignore all the outcomes except the ones in B. Out of the ones we keep, what is the expected ratio of the number of times A occurred to the number of outcomes we are looking at?"

Example: Suppose $S = \{1, 2, 3, 4, 5, 6\}$ and

$$P(1) = 0.5, P(2) = P(3) = P(4) = P(5) = P(6) = 0.1$$

A popular way to visualize the probability function is with a pie chart:



Let $A = \{1, 2, 3\}$ We can see immediately that $P(A) = \frac{1}{2} + \frac{1}{10} + \frac{1}{10} = \frac{7}{10}$

Now let's consider some possibilities for B

Example 1: Suppose B = {2,3}

If we know B occurred, the outcome was either 2 or 3. These are both in A so in this situation, we **know** A occurred. We could phrase this as "given that B occurred, the probability that A occurred is 1".

Example 2: Suppose B = {4, 6}

If we know B occurred, the outcome was either 4 or 6. Neither of these is in A so in this situation we know A **did not** occur. We could phrase this as "given that B occurred, the probability that A occurred is 0".

Example 3: Suppose B = {1, 2, 4}

If we know B occurred, the outcome was either 1, 2, or 4. Two of those are in A and the other is not. It's at this point we can easily go wrong. To many people it seems pretty reasonable to say "given that B occurred, the probability that A occurred is $\frac{2}{3}$ " ... but what "seems pretty reasonable" is not correct.

Here's why it's wrong. If we think about the probabilities of the individual outcomes, 1 has much higher probability than anything else. So out of all the outcomes in our hypothetical sequence of experiments, the number of times we say "the outcome is 1 or 2, so B occurred" will be much greater than the number of times we say "the outcome is 4, so B occurred" - and for all the outcomes in the first category, it will also be true that A occurred. So the number of times we will say "B occurred, but A didn't occur" is very small. If we know that B occurred, the probability that A occurred should be higher than $\frac{2}{3}$. But what should the value be?

The total probability for B is $P(1) + P(2) + P(4) = 0.7$

The total probability for $A \cap B$ is $P(1) + P(2) = 0.6$

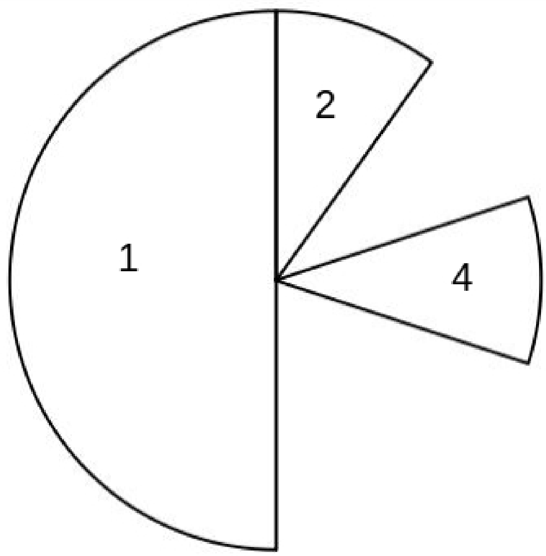
So at this point we can cut to the conclusion and assert that

$$P(A \text{ given } B) = \frac{P(A \cap B)}{P(B)} = \frac{0.6}{0.7} = \frac{6}{7}$$

But as a student, I found this jump a bit hard to accept so I spent quite a while trying to see why it is true. Like many academics, having thought something through I now feel compelled to share it with you. If you are completely comfortable with the formula I just stated, feel free to skip over this next bit.

Remember *we are given that B occurred*, so instead of the set of possible outcomes for this particular sample being S, it is just B.

This means that if we look at the pie chart for the outcomes, it now looks like this:



But that doesn't make any sense – the probabilities have to add up to 1. Given that B has occurred, the probabilities of 3, 5 and 6 having occurred all drop to 0, and the probabilities of 1, 2 and 4 having occurred all go up. In fact they all go up proportionally so their relative sizes stay the same – ie. they are all multiplied by the same value, so that they now total to 1.

We need to find x such that $\frac{1}{2} * x + \frac{1}{10} * x + \frac{1}{10} * x = 1$

That simplifies to $\left(\frac{1}{2} + \frac{1}{10} + \frac{1}{10}\right) * x = 1$

$$\frac{7}{10} * x = 1$$

$$x = \frac{10}{7}$$

So given that B has occurred,

$$P(1) = \frac{1}{2} * \frac{10}{7} = \frac{5}{7}$$

$$P(2) = \frac{1}{10} * \frac{10}{7} = \frac{1}{7}$$

$$P(4) = \frac{1}{10} * \frac{10}{7} = \frac{1}{7}$$

Out of these three possible outcomes, the ones that correspond to A occurring are (as we have already noted) the outcomes in $A \cap B$, ie $\{1, 2\}$ ... and the sum of their "revised" probabilities is $\frac{5}{7} + \frac{1}{7} = \frac{6}{7}$

If you go back through the steps, you will see that when we found the value of x we needed to "scale up" the probabilities of 1, 2 and 4, it turned out that x was just $\frac{1}{P(B)}$. So when we multiplied all the probabilities of values in B by x we were really just dividing by $P(B)$. Since we are only interested in the outcomes in A, we can focus on $A \cap B$

Our answer can be written as $\displaystyle\sum_{\text{all } v \in A \cap B} \frac{P(v)}{P(B)}$

which gives

$$\frac{\sum_{\text{all } v \in A \cap B} P(v)}{P(B)}$$

which gives

$$\frac{P(A \cap B)}{P(B)}$$

which is exactly the formula I just pulled out of thin air a couple of pages back!

If you are ***still*** not completely satisfied with the explanation of why the probabilities of 1, 2 and 4 all go up when we know B has occurred (and as a student, I had my doubts about it), here is yet another explanation.

Remember that we are supposing that a long sequence of experiments has been conducted. We are looking only at the ones for which it is true that B occurred, and we are asking what proportion of those also correspond to A.

Suppose the start of the sequence of outcomes looks like this (I just wrote a Python program to generate this sequence):

4 1 1 5 1 5 6 5 6 1 3 1 1 2 5 1 5 1 4 6 1 1 1 1 3 1 4 4

We can see that about half of the outcomes are 1, which is what we expect. But if we look at just the outcomes for which "B occurred" is a true statement, the sequence reduces to

4 1 1 1 1 1 1 2 1 1 4 1 1 1 1 1 4 4

so the number of samples has reduced from 28 to 18. The proportion of 1's is now $\frac{13}{18}$ which is **very** close to $\frac{5}{7}$. The proportions of 2's and 4's also go up, though this example doesn't show that as nicely. With a longer sequence we would see consistent increases for 1's, 2's and 4's

OK, hopefully at this point we are all comfortable with the statement

the probability of A, given that B has occurred $= \dfrac{P(A \cap B)}{P(B)}$

Now for some notation: instead of writing out "the probability of A, given that B has occurred" we simply write **P(A | B)** (The poor old vertical bar gets **another** meaning on top of all the meanings it already has.) So our first three examples can be summarized as

$$A = \{1,2,3\} \;\; B = \{2,3\} \qquad P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{2}{10}}{\frac{2}{10}} = 1$$

$$A = \{1,2,3\} \;\; B = \{4,6\} \qquad P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0}{\frac{2}{10}} = 0$$

$$A = \{1,2,3\} \;\; B = \{1,2,4\} \qquad P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{6}{10}}{\frac{7}{10}} = \frac{6}{7}$$

Let's consider another sample space. Suppose we have a barn containing a cow, a moose, a horse, a llama, a hippo, a camel and a bear. Our sampling experiment is to open the barn door and observe which four animals come out first. We assume that all outcomes are equally probable (you can quite reasonably question the reality of this assumption, but we'll go with it for now).

How many outcomes are there? It should be clear that the answer is $\binom{7}{4} = 35$, so (under our assumption that each outcome has equal probability) each outcome has probability = $\frac{1}{35}$

Let event A = {all outcomes that include the horse and the moose} It's not hard to see that there are exactly 10 such outcomes, so P(A) = $\frac{10}{35}$

Let event B = {all outcomes that don't include the cow and don't include the llama} Again we can see that there are exactly 5 such outcomes, so P(B) = $\frac{5}{35}$

Now what is P(A | B)? To apply the formula we need to know $P(A \cap B)$, which we can easily compute. There are precisely 3 outcomes that include the horse and the moose, and exclude the cow and llama: {horse, moose, hippo, camel} {horse, moose, hippo, bear} and {horse, moose, camel, bear}

So $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{3}{5}$

Here is a way to make sense of this: If we don't know exactly which output occurred but we know it is in B, we are limited to just 5 possible outcomes. Out of those, there are 3 outcomes that are elements of A. So the probability that the (unknown) outcome is in A is $\frac{3}{5}$

A last note: Since we now know that $P(A|B) = \frac{P(A \cap B)}{P(B)}$, we can write

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

We can also compute $P(B|A) = \frac{P(A \cap B)}{P(A)}$, which turns around to give

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

Combining these two equations gives the result

$$P(A|B) * P(B) = P(B|A) * P(A)$$

We can use this! Suppose we know $P(A) = 0.3$, $P(B) = 0.6$ and $P(B|A) = 0.5$ ... we can compute

$$P(A|B) = \frac{0.5 * 0.3}{0.6} = 0.25$$

Exercise: Is it possible to have events A and B where $P(A) = 0.3$, $P(B) = 0.1$ and $P(B|A) = 0.5$?

Independent Events

Our intuitive understanding of the word "independent" is "not controlled or influenced", and this is exactly the sense in which we use the word when discussing probability. We say that event A is **independent** of event B if the probability that A has occurred is not affected by whether B has occurred or not.

For example, suppose our experiment consists of tossing a red die and a blue die (both 6-sided, both fair) Each outcome consists of an ordered pair (r,b) where r is the value of the red die and b is the value of the blue die. We intuitively accept that the two dice are independent of each other so the probability of an event such as "the red die is showing 3" should not be affected by whether or not an event such as "the blue die is showing 4" has occurred.

Consider these two events: A: { the sum of the two dice = 7 } and B: { the blue die is showing either 2 or 3 }

The possible outcomes can be seen in this table, in which the first number in each pair represents the red die, and the second represents the blue die:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

Here the outcomes that constitute A are coloured green and yellow, and the outcomes that constitute B are coloured blue and yellow.

We can see that P(A) $= \frac{6}{36} = \frac{1}{6}$

Now, what is the probability that A has occurred, if we know that B has occurred? (In other words, what is P(A | B)?) B contains 12 outcomes, and 2 of those are in A ... so the probability that A has occurred, given that B has occurred, is $\frac{2}{12} = \frac{1}{6}$ which is exactly P(A) !!! Knowing that B has occurred does not increase or decrease the probability that A has occurred. This becomes our mathematical way of expressing the idea that A is independent of B.

In notation, this would mean that P(A | B) = P(A) which we can process as follows

$$
\begin{aligned}
P(A|B) &= P(A) \\
\frac{P(A \cap B)}{P(B)} &= P(A) \\
P(A \cap B) &= P(A) * P(B)
\end{aligned}
$$

Note that we can go one step further:

$$
\begin{aligned}
P(A \cap B) &= P(A) * P(B) \\
\frac{P(A \cap B)}{P(A)} &= P(B)
\end{aligned}
$$

which is just the notational way of saying that B is independent of A. In other words, **A is independent of B if and only if B is independent of A.**

Thus we say **A and B are independent events if** $P(A \cap B) = P(A) * P(B)$