Basics of Conditional Probability

Elliot Pickens
eeepickense@gmail.com

November 15, 2020

1 Intro

I've recently been looking back on some old probability notes for a project I hope to embark on during the coming weeks and started thinking: "maybe I should reproduce these notes and put them up online." There are obviously thousands upon thousands of explanations of all aspects of probability out there and I doubt this one will be much better than any of those, but even if this document just helps a few others (or my lonesome self) I think it's a worthwhile endeavor.

Today I'm going to start off what may or may not develop into a series on the fundamentals of probability. I'll be skipping the foundational counting/combinatorial rules and jumping right into conditional probability.

2 Conditional Probability

One deceptively hard question to answer is how exactly the probability of one event informs another. It's the sort of question we ask ourselves all the time - and often more often than not get quite wrong. That's in part due to the incredible uncertainty present is each and every one of our daily lives, but also because we often fail to employ the rules of conditional probability when talking about the world around us.

Conditional probability is the probability that some event A will occur given an event B has already occurred. The difficulty in coming up with a way to calculate this value (denoted $Pr(A|B)$) is that we can't arrive at it by simply multiply probabilities together. Instead we have to consider how one event updates our chances of the second taking place. One way to imagine this is by considering how both A and B lie within some arbitrary sample space. We can think of A and B as blobs floating at fixed positions within the space and potentially overlapping one another. The definition of $Pr(A|B)$ comes directly from that overlap. Since event B has already happened we know that we're stationed somewhere within the blob defined by B, so then our question becomes: "if we're already in B then what's the probability that we're in the portion of B that overlaps with A." The overlap of two events is their intersection or $Pr(A \cap B)$ and the probability of being in B is $Pr(B)$, so the definition of conditional probability becomes the ratio of the intersection over the totality space taken up by B or:

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)} \tag{1}$$

A Quick Example

What's the probability that we've drawn a card with a numeric value given we know we've drawn a spade? Well we know that there are 13 spades in a deck and 52 cards total so chance of drawing a spade or $Pr(spade)$ is $\frac{13}{52}$. We also know that 9 of the 13 spades have numeric value so $Pr(spade \cap numeric) = \frac{9}{52}$. Therefore by the definition of conditional probability we laid out above the chances of drawing a numeric card given we know we have a spade is $\frac{9}{52}/\frac{13}{52} = \frac{9}{13}$, which just so happens to be equal to the probability of drawing a numeric card in general since we have equal chances of drawing any one face.

Now that we've pinned down the basics of conditional probability let's lay out a few specifics.

First off we should make sure to remember that conditional probabilities are at their core still probabilities. That is to say that the basic rules we would follow elsewhere still apply here. For example, compliments work just as they would regularly, leading us to $Pr(A^c|B) = 1 - Pr(A|B)$.[1]

Another thing that's interesting to consider and potentially insightful is that sometimes we can solve for more complex conditional probabilities by splitting them up into a series of smaller calculations. We can sum this up using the following equation describing how things work with a intersection of many events:

$$\begin{aligned}
Pr(A_1 \cap A_2 \cap ... \cap A_n|B) &= \frac{Pr(A_1 \cap A_2 \cap ... \cap A_n \cap B)}{Pr(B)} && (2)\\
&= \frac{Pr(A_1 \cap B)}{Pr(B)} \frac{Pr(A_1 \cap A_2 \cap B)}{Pr(A_1 \cap B)} ... \frac{Pr(A_1 \cap A_2 \cap ... \cap A_n \cap B)}{Pr(\cap A_2 \cap ... \cap A_{n-1} \cap B)} && (3)\\
&= Pr(A_1|B)Pr(A_2|A_1 \cap B)...Pr(A_n|A_1 \cap A_2 \cap ... \cap A_{n-1} \cap B) && (4)
\end{aligned}$$

Possibly more importantly we can take advantage of the subtle ambiguity associated with $Pr(A|B)$ in equation 1 to get the following result:

$$\begin{aligned}
Pr(A \cap B) &= Pr(A|B)Pr(B) && (5)\\
&= Pr(B|A)Pr(A) && (6)
\end{aligned}$$

The simple equality above holds a great deal of power. Much of which will be explained in later posts specifically addressing Bayes' Theorem, but its utility shows up even in small details like the interaction between conditional probability and partitions.

If we have a series of events that can be used to form a partitions of a space then we can use the equality in equation 6 to find the probability of any other event within the space using the sum:

$$Pr(A) = \sum_{i=1}^{n} Pr(A|B_i)Pr(B_i) \tag{7}$$

[1] This can be shown by considering that since B has occurred we have effectively reduced our sample space to just the portion of the total space that contains B

Since we know from basic union and intersection rules that:

$$A = (B_1 \cap A) \cup (B_2 \cap A) \cup ... \cup (B_n \cap A) \tag{8}$$

There are a number of other small results that can be similarly derived, like a conditional probability based version of the law of total probability, but I'm not particularly interested in those right now, so I'm going to leave them out of this write up.

2.1 Independence and Conditional Probability

First let's begin with a definition of independence. We can say that two events are independent if:

$$Pr(A \cap B) = Pr(A)Pr(B) \tag{9}$$

Independence is at the core of many probability calculations. Without it multiplicative methods for calculating things like the probability of rolling snake eyes, or tails coming up four times in a row when flipping a coin would not longer work.

These sorts of events also hold an interesting relationship to conditional probability. If we once again ask the question: "what is the probability of event A given that event B has occurred," but with the caveat that both events are independent we arrive at a much different result. By combining equations 1 and 9 we get:

$$\begin{align} Pr(A|B) &= \frac{P(A \cap B)}{Pr(B)} \tag{10} \\ &= \frac{Pr(A)Pr(B)}{Pr(B)} \tag{11} \\ &= Pr(A) \tag{12} \end{align}$$

This result is a second, conditional probability based definition of independence. It allows us to say that two events are independent if and only if $Pr(A|B) = Pr(A)$ and $Pr(B|A) = Pr(B)$.

We can use this result to further prove that a collection of n events is independent if an only if the equation below holds for any given pair of disjoint subsets:

$$Pr(A_{i_1} \cap ... \cap A_{i_k} | A_{i_{k+1}} \cap ... \cap A_{i_n}) = Pr(A_1 \cap ... \cap A_k) \tag{13}$$

We can also rework equation 4 to define the conditional independence of a set given another event as:

$$\begin{align} Pr(A_{i_1} \cap ... \cap A_{i_k} | B) &= Pr(A_{i_1}|B)Pr(A_{i_2}|A_{i_1} \cap B)...Pr(A_{i_k}|A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_{k-1}} \cap B) \tag{14} \\ &= Pr(A_{i_1}|B)...Pr(A_{i_k}|B) \tag{15} \end{align}$$

3 Wrapping Up

I'm going to end this document here for now. I might come back to it and add things sometime in the future, but to be completely honest I doubt I will. I hope this short run down of conditional probability was helpful to someone out there. Also, I wrote this as a first draft and I'm putting it out that way, so if anyone happens to read through this and notice any errors please feel free to contact me via the email at the top of the document.

4 Acknowledgments

These notes were based on *Probability and Statistics (Fourth Edition)* by DeGroot & Schervish