Intro to Bayes' Theorem

Elliot Pickens
eeepickense@gmail.com

November 20, 2020

1 Intro

I recently wrote a post laying out some commonly used theorems central to conditional probability. If you haven't already taken a look at that post you might want to briefly skim through it before reading this one, because today I will be taking a Bayes' Theorem.

2 Bayes' Theorem

Previously we saw that conditional probability can be defined using the following two equations:

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

$$P(A|B)P(B) = P(B|A)P(A) \tag{2}$$

To create the mostly commonly seen statement of Bayes' Theorem we simply rearrange equation 2 such that we isolate a single conditional probability on one side of the inequality to get:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \tag{3}$$

This particular statement of Bayes' Theorem is simple and straight to the point, but it is often not the first one you'll see in a textbook. Instead it's common to present a version of Bayes' Theorem that constructs the denominator using the conditional law of total probability. Assuming we have a set of n events that partition our sample space, this version can be written as:

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} Pr(A|B_i)Pr(B_i)}^{1} \tag{4}$$

It may seem a little odd that this version is the one most frequently presented first, given that at a glance it looks more complex than equation 3. And it is admittedly more difficult to write out, but let's consider what happens when we actually put it to use. It's very possible (and really quite likely) that we don't know the exact value of $P(A)$. Without having $P(A)$ in hand we have to calculate it, and we can do so using the sum in equation 4.[2]

[1] It is also possible to present equation 4 using an integral instead of a sum. This is useful when working with random variables.

[2] For a more detailed dive into conditional probability and conditional law of total probability check out my previous post on the subject, or check out the books I list in the acknowledgments.

Are We Holding a Fair Coin?

Let's say we have a two coins: a fair coin and a coin with two heads. We randomly select one of them without looking and then flip it a single time revealing a head. What are the chances that we are holding a fair coin? We can write this out using Bayes' Theorem as $P(Fair\ Coin|Heads) = \frac{P(Heads|Fair\ Coin)P(Fair\ Coin)}{P(Heads)}$. We know that $P(Heads|Fair\ Coin) = 1/2$ and $P(Fair\ Coin) = 1/2$, but what is $P(Heads)$? To solve for that value we turn to the conditional law of total probability to get $P(Heads) = \sum_{i\in\{Fair,\ Unfair\}} Pr(Heads|Coin_i)Pr(Coin_i) = \frac{1}{2} * \frac{1}{2} + 1 * \frac{1}{2} = \frac{3}{4}$. Putting it all together we get $P(Fair\ Coin|Heads) = \frac{P(Heads|Fair\ Coin)P(Fair\ Coin)}{P(Heads)} = (\frac{1}{2} * \frac{1}{2})/\frac{3}{4} = \frac{1}{3}$. Not too hard! That's part of the beauty of Bayes' Theorem, but this is only the beginning!

Before we move past the example above let's consider what happens if we change things up a little and flip the coin twice. How do we determine the probability of event B when our A is composed of multiple events? It's possible to approach this is to simply doing the same calculation (with an $A_1 \cap A_2$ in place of A), but when that's too difficult we can turn to the conditional version of Bayes' Theorem.

$$P(B_j|A_1 \cap A_2) = \frac{P(A_2|B_j \cap A_1)P(B_j|A_1)}{\sum_{i=1}^{n} Pr(A_2|B_i \cap A_1)Pr(B_i|A_1)} \tag{5}$$

With equation 5 we can break down a long combination of events into a sequence that both simplifies our calculations and allows us to change our beliefs over time. To explain just how that update takes place let's work things out when we add that second coin flip to our example.

How Does a Second Flip Alter Our Beliefs?

Imagine we flip the coin a second time and once again see a head. Does this change how certain we are that we're holding a fair coin? Previously we worked out that after seeing one head our chances of having a fair coin were 1 in 3. Using that value and the conditional formulation of Bayes' Theorem we can compute the shift as follows:

$$\begin{aligned} P(Fair|Head_1 \cap Head_2) &= \frac{P(H_2|F \cap H_1)P(F|H_1)}{P(H_2|F \cap H_1)P(F|H_1) + P(H_2|Not\ F \cap H_1)P(Not\ F|H_1)} && (6)\\ &= \frac{(1/2)(1/3)}{(1/2)(1/3) + (1)(2/3)} && (7)\\ &= \frac{(1/6)}{(5/6)} && (8)\\ &= \frac{1}{5} && (9) \end{aligned}$$

The great thing about this method is that it both makes things easier and gets at one of the most powerful aspects of Bayes' Theorem (and Bayesian statistics as a whole): the ability to update your beliefs overtime. As new data (like new coin flips) come in we can repeatedly use our previous probability calculations to compute our new beliefs in an online manner.

Now that I've presented the basics let's talk about a little bit of vocab. When talking about Bayes' Theorem it is common to break it down into three pieces:

- The prior: $P(B)$
- The likelihood: $P(A|B)$
- The posterior: $P(B|A)$

We'll slowly develop a better and more intuitive understanding of each of these values as we grow the depth of our knowledge of Bayesian Statistics, but it's still useful to establish a few not-so-mathematical definitions. First off we have our *prior*. As the name suggests this is our previous, or initial belief. The second piece we have is the *likelihood*, which is a little harder to describe intuitively, but given it's relation to the event of interest (B) and the prior we can roughly say it's the "weight" given to new data. That is to say the likelihood helps us say how we are going to update our prior to give us the third piece to our puzzle: the *posterior*. The posterior is our new, updated belief, and the end result of our calculation.

In the case of the fair coin example, without any additional data we know nothing and thus assume an *uninformative* prior where we have a 50/50 chance of holding a fair coin. We then calculate a posterior value of 1/3 using that prior and a likelihood. That posterior then becomes our new prior, which we can use as our starting point should we want to update our beliefs and get yet another posterior value.

In the future we'll take these three concepts and reconfigure them to shift our belief system away from isolated probability values towards random variables and distributions. This shift will give our problem solving abilities an incredible latitude that will allow us to tackle nearly any statistical problem we're interested in, but first we need to revisit the subject of my next post: random variables.

3 Acknowledgments

These notes were based on the following texts:

- *Probability and Statistics (Fourth Edition)* by DeGroot & Schervish
- *A First Course in Bayesian Statistical Methods* by Peter D. Hoff
- *Introduction to Bayesian Statistics-Wiley (Third Edition)* by William M. Bolstad & James M. Curran