CS30 (Discrete Math in CS), Summer 2021 : Lecture 15

Topic: Probability: Bayes Rule

Disclaimer: These notes have not gone through scrutiny and in all probability contain errors. Please discuss in Piazza/email errors to deeparnab@dartmouth.edu

---

- **Bayes Rule.** Put simply, Bayes' rule is the following observation: for any two events $\mathcal{A}$ and $\mathcal{B}$ which each occur with non-zero probability

$$\mathbf{Pr}[\mathcal{B} \mid \mathcal{A}] = \frac{\mathbf{Pr}[\mathcal{A} \mid \mathcal{B}] \cdot \mathbf{Pr}[\mathcal{B}]}{\mathbf{Pr}[\mathcal{A}]} \tag{Bayes Rule}$$

The proof is trivial after we substitute the formula of conditional probability.

We can expand it slightly more using the law of total probability to get

$$\mathbf{Pr}[\mathcal{B} \mid \mathcal{A}] = \frac{\mathbf{Pr}[\mathcal{A} \mid \mathcal{B}] \cdot \mathbf{Pr}[\mathcal{B}]}{\mathbf{Pr}[\mathcal{A} \mid \mathcal{B}] \cdot \mathbf{Pr}[\mathcal{B}] + \mathbf{Pr}[\mathcal{A} \mid \neg\mathcal{B}] \cdot (1 - \mathbf{Pr}[\mathcal{B}])} \tag{Bayes Rule - Opened up}$$

Why is this a big deal? We will look at a couple of examples. But in a nutshell, it states that to answer what is the probability of event $\mathcal{B}$ given event $\mathcal{A}$, if we know (a) the total probability of event $\mathcal{B}$, and if (b) the probability of event $\mathcal{A}$ is *easier* to figure out, then we can get our answer. The main applications come in when $\mathcal{A}$ is an "outcome" and $\mathcal{B}$ is a "hypothesis"; $\mathbf{Pr}[\mathcal{B}]$ is a "*prior* belief" on the hypothesis, and $\mathbf{Pr}[\mathcal{B} \mid \mathcal{A}]$ is our "*posterior* belief" given we see the outcome event $\mathcal{A}$.

Examples.

  - *Bob is communicating with Alice over a noisy channel. It has been empirically established that the noisy channel flips every transmitted bit with probability* 0.1. *Given that Bob is equally likely to be sending a* 0*-bit or a* 1*-bit, what is the probability that Bob indeed sent a* 1 *when Alice receives a* 1*? That is, how* confident *is Alice?*

    Let $\mathcal{B}$ be the event that Bob sent a 1. Let $\mathcal{A}$ be the event that Alice received a 1. We are asked to calculate the conditional probability $\mathbf{Pr}[\mathcal{B} \mid \mathcal{A}]$.
    What is that we know? We know that $\mathbf{Pr}[\mathcal{B}] = 0.5$ (Bob is equally likely to send a 0 or a 1). We know $\mathbf{Pr}[\mathcal{A} \mid \mathcal{B}] = 0.9$ (the channel *doesn't flip* the sent 1 with probability 0.9). Finally, we also know that $\mathbf{Pr}[\mathcal{A} \mid \neg\mathcal{B}] = 0.1$ (the channel does flip a sent 0 with probability 0.1). Thus, we can substitute in (Bayes Rule - Opened up) to get

$$\mathbf{Pr}[\mathcal{B} \mid \mathcal{A}] = \frac{\mathbf{Pr}[\mathcal{A} \mid \mathcal{B}] \cdot \mathbf{Pr}[\mathcal{B}]}{\mathbf{Pr}[\mathcal{A} \mid \mathcal{B}] \cdot \mathbf{Pr}[\mathcal{B}] + \mathbf{Pr}[\mathcal{A} \mid \neg\mathcal{B}] \cdot (1 - \mathbf{Pr}[\mathcal{B}])} = \frac{(0.9) \cdot (0.5)}{(0.9) \cdot (0.5) + (0.1) \cdot (0.5)} = 0.9$$

> **Exercise:** *Does this seem what you expected? What happens if the chance that Bob sends a* 1 *is actually* 60% *instead of* 50%*. Do you think the answer will change?*

Indeed, let us do this exercise. Nothing changes, except $\mathbf{Pr}[\mathcal{B}] = 0.6$ now. Substituting again, we get

$$\mathbf{Pr}[\mathcal{B} \mid \mathcal{A}] = \frac{(0.9) \cdot (0.6)}{(0.9) \cdot (0.6) + (0.1) \cdot (0.4)} = \frac{0.54}{0.58} \approx 0.931$$

So Alice is *more* confident that Bob sent a 1 when Bob sends more ones. Even when the channel is the same. Is this what you expected?

– *Arithmophobia is a quality of life debilitating condition and should be detected as early as possible. Fortunately, the pharmaceutical company* HAYSTEAM *have come up with a test. It's not perfect. It has a* ***false positive rate*** *of* $\mathsf{fp} = 1\%$*; that is, on* 1% *of the healthy population, the test detects the condition, It also has a* ***false negative rate*** *of* $\mathsf{fn} = 2\%$*. Again, this means that* 2% *of the afflicted population go undetected. It is assumed around* 1% *of the population may be suffering from Arithmophobia.*

  You take the test and unfortunately it comes positive (the test says you have Arithmophobia). What is the probability that you actually do?

  What a story! But such situations abound. Easy-peasy if you know Bayes rule and know how to set things up.

  $\mathcal{A}$ be the event that you have the affliction. Now, you don't know whether you do or not (that's why, presumably, you take the test). Before taking the test, you just look at the statistics and believe that you are as likely as anyone else to have this condition. Since 40% of the population have it, you conclude (reasonably)

$$\mathbf{Pr}[\mathcal{A}] =: p_A = 0.01$$

  $\mathcal{P}$ be the event that the test comes out positive on you. We are really interested in figuring out $\mathbf{Pr}[\mathcal{A} \mid \mathcal{P}]$. We will do so by applying Bayes rule.

  First, is $\mathbf{Pr}[\mathcal{P} \mid \mathcal{A}]$ easy? That is, if you did have the affliction, what is the probability that the test would catch it? The answer is $(1 - \mathsf{fn})$; you would be tested positive unless we got a *false negative*. Thus,

$$\mathbf{Pr}[\mathcal{P} \mid \mathcal{A}] = (1 - \mathsf{fn}) = 0.98$$

  How about $\mathbf{Pr}[\mathcal{P} \mid \neg\mathcal{A}]$? This is precisely the *false positive* rate. So,

$$\mathbf{Pr}[\mathcal{P} \mid \neg\mathcal{A}] = \mathsf{fp} = 0.01$$

  Now to apply Bayes rule,

$$\mathbf{Pr}[\mathcal{A} \mid \mathcal{P}] = \frac{\mathbf{Pr}[\mathcal{P} \mid \mathcal{A}] \cdot \mathbf{Pr}[\mathcal{A}]}{\mathbf{Pr}[\mathcal{P} \mid \mathcal{A}] \cdot \mathbf{Pr}[\mathcal{A}] + \mathbf{Pr}[\mathcal{P} \mid \neg\mathcal{A}] \cdot (1 - \mathbf{Pr}[\mathcal{A}])}$$

  which simplifies to

$$\mathbf{Pr}[\mathcal{A} \mid \mathcal{P}] = \frac{(1 - \mathsf{fn})p_A}{(1 - \mathsf{fn})p_A + \mathsf{fp}(1 - p_A)} \tag{FN-FP}$$

  which equates to around 0.497. Thus, if the test comes positive, then the chances you have the affliction gets close to 49.7%. That is, after one positive test (with the given rates), your chances of being afflicted is more like a toss of a coin (and not 99% or 98%). (Of course, you should still be worried since 0.497 is way larger than 1%, but it is not a certainty yet.)

  Exercise: *Suppose you take the test again and it again comes positive. What is the probability* $\mathbf{Pr}[\mathcal{A} \mid \mathcal{P}]$ *now?*

There are two ways to do this. One is the long way. One is the short way.

Let's do the short way first. After the first test was positive, your "belief" that you have the affliction, that is, $\mathbf{Pr}[\mathcal{A}]$ is no longer 0.01. Rather, it is 0.497. Therefore, if we just plug it into (FN-FP), we get that your new belief is

$$\mathbf{Pr}[\mathcal{A} \mid \mathcal{P}] = \frac{(0.98) \cdot (0.497)}{(0.98) \cdot (0.497) + (0.01)(0.503)} = 0.989$$

That is, if the second test also comes positive, your belief that you have the disease shoots up to 0.989.

The longer way now. Let $\mathcal{P}$ now be the event that both tests are positive. And $\mathcal{A}$ is back to being the event of the original chances of being afflicted. So, $\mathbf{Pr}[\mathcal{A}] = 0.01$.

Let's figure out $\mathbf{Pr}[\mathcal{P} \mid \mathcal{A}]$. This is $(1 - \mathsf{fn}) \cdot (1 - \mathsf{fn}) = (0.98)^2$. What is the probability $\mathbf{Pr}[\mathcal{P} \mid \neg\mathcal{A}]$? This is simply $\mathsf{fp} \cdot \mathsf{fp} = (0.01)^2$. Thus, we get

$$\mathbf{Pr}[\mathcal{A} \mid \mathcal{P}] = \frac{(0.98)^2 \cdot 0.01}{(0.98)^2 \cdot 0.01 + (0.01)^2 \cdot (0.99)} = 0.989$$

Are you surprised the answers are the same? A little math shows one shouldn't be. Suppose we denote the original $\mathbf{Pr}[\mathcal{A}]$ as p_o, that is, $p_o = 0.01$, and the belief after one round being $p_n = \frac{(1-\mathsf{fn})p_o}{(1-\mathsf{fn})p_o + \mathsf{fp}(1-p_o)}$. Then, the first calculation is giving the final answer p as

$$p = \frac{(1 - \mathsf{fn})p_n}{(1 - \mathsf{fn})p_n + \mathsf{fp}(1 - p_n)}$$

while the second answer is giving

$$p = \frac{(1 - \mathsf{fn})^2 p_o}{(1 - \mathsf{fn})^2 p_o + \mathsf{fp}^2(1 - p_o)}$$

And these two are the same (I leave it as an exercise).

A more general reason is described in the supplement. To understand this, one needs to understand the notion of "conditional independence" which we leave out of this course. The interested student should read the supplement.

– *Upon inspecting the spam folder of his email, Ganesh finds that 7 emails out of the 100 emails there are actually not spam. He researches a bit and finds that his email server's spam filter claims a false positive of less than 1%. That is, given any random email, if it is not spam then there is less than 1% chance the server puts it in the spam folder. Is Ganesh's email server lying? If not, how can you help Ganesh reconcile with the fact that he is seeing around 7% non-spam (significantly higher than 1%) in his spam folder?*

  This was a question in your problem set 0, and indeed many of you figured out why Ganesh had really no need to worry. The short answer is: perhaps Ganesh actually got > 800 emails of which 93 were spam, and 707 were not spam. Of these 707, 7 were sent to the spam folder. This is indeed within the desired accuracy claimed by the mail server.

  Let us change the problem to actually get an estimate of the rate at which Ganesh does get spam. Turns out, we actually don't have enough data to answer this — we need the mail server's *false* *negative* rate as well. Suppose, this was 0.1% – that is, the chance an actual spam being sent to the real inbox is 0.1%. Now, we can address the question as follows.

Let $\mathcal{S}$ be the event that a random incoming email is spam. We are interested in finding our $p := \mathbf{Pr}[\mathcal{S}]$. This would indicate the "rate" at which Ganesh gets spam. Let $\mathcal{F}$ be the event that Ganesh's mail server marks a random email spam. What all do we know?

We know that $\mathbf{Pr}[\mathcal{F} \mid \mathcal{S}] = 1 - \mathsf{fn} = 0.999$ – given that a mail is indeed spam, it goes to the spam folder with 99.9% probability.

We also know that $\mathbf{Pr}[\mathcal{F} \mid \neg\mathcal{S}] = \mathsf{fp} = 0.01$ – given that a mail is not spam, it goes to the spam folder with 1% probability.

Finally, from the data that 7 out of 100 emails in the flag are not spam, we can also assert $\mathbf{Pr}[\mathcal{S} \mid \mathcal{F}] = \theta := 0.93$. Bayes rule tells us

$$\mathbf{Pr}[\mathcal{S} \mid \mathcal{F}] = \frac{\mathbf{Pr}[\mathcal{F} \mid \mathcal{S}] \cdot \mathbf{Pr}[\mathcal{S}]}{\mathbf{Pr}[\mathcal{F} \mid \mathcal{S}] \cdot \mathbf{Pr}[\mathcal{S}] + \mathbf{Pr}[\mathcal{F} \mid \neg\mathcal{S}] \cdot (1 - \mathbf{Pr}[\mathcal{S}])}$$

giving,

$$\theta = \frac{(1 - \mathsf{fn}) \cdot p}{(1 - \mathsf{fn}) \cdot p + \mathsf{fp} \cdot (1 - p)}$$

Rearranging and solving (do this!), we get

$$p = \frac{\theta \cdot \mathsf{fp}}{\theta \cdot \mathsf{fp} + (1 - \theta) \cdot (1 - \mathsf{fn})}$$

Substituting $\theta = 0.93$, $\mathsf{fp} = 0.01$, and $\mathsf{fn} = 0.001$, we get that Ganesh receives spam at rate

$$\frac{(0.93) \cdot (0.01)}{(0.93) \cdot (0.01) + (0.07) \cdot (0.999)} \sim 0.11738$$

Around 11.73%