Bayes theorem with a few simple example of Bayesian inference

Iman Mahyaeh
(Dated: October 12, 2025)

We briefly discuss Bayes' Theorem and give a few example of Bayesian inference.

BAYES' THEOREM

The way that I understand the Bayes theorem is as follows. Let's say we want to calculate the joint probability of $p(x, y)$, meaning the probability of having both x and y **simultaneously**. These are two stochastic/random variables that somehow are or could be related to each other. For example, x shows whether it is raining or not, and y shows whether your neighbor picks up their umbrella. Note that the adverb simultaneously does not mean they should definitely occur at the same *time*; their *actual* time of occurrence could be different. Here *simultaneously* simply means *logical AND operator*. For instance, one could ask about the joint probability of attending a class during term and passing its exam which comes in the end.

To understand the theorem I usually go back to the frequentist approach of probability and consider an example. It is handy to have an example at hand and understand all in that context. Let's consider the following example:

There is a city with two areas (E, W). Each household has one car. Cars could have 3 colors (R, G, B). Let's say there are 100 households, 40 of which are in area E and 60 are in area W. To present the cars' colors we can use a tree structure:

- Area E (40 households)
  - Color R: 14
  - Color G: 9
  - Color B: 17
- Area W (60 Households)
  - Color R: 25
  - Color G: 16
  - Color B: 19

We can, of course, look at the city in a different way. We can start the tree by first dividing the households by their cars' color and then by the area they live in:

- Color R (39 cars)
  - Area E: 14
  - Area W: 25
- Color G (25 cars)
  - Area E: 9
  - Area W: 16
- Color B (36 cars)
  - Area E: 17
  - Area W: 19

We can now ask, how many people live in area E **and** (the logical AND, **simultaneously**) have a blue car? We can answer it in two ways:

1. Either first select people who live in area E, which are 40. And then see that among those 17 households have a blue car.

2. Or first select the households with blue cars which are 36, and then select those who live in area E among them. Also lead to 17!

So,

$$\boxed{\text{The order we do the counting/selection does not matter.}}$$

Now let's go back to the joint probability. Let's ask what is the probability of randomly choosing a household in area E with a blue car? Well, the result is 0.17, because 17 out of 100 household fulfill both conditions. We can write this as,

$$p(\text{area} = E, \text{color} = B) = \frac{17}{100} \ . \tag{1}$$

Let's calculate it using probabilities. Again we can do it in two ways.
We can first ask ourselves what is the probability of choosing a household from the area E. It is 40 out of 100,

$$p(\text{area} = E) = \frac{40}{100} \ . \tag{2}$$

We can now ask that what is the probability of having a blue car among the households in region E? Or in a bit more technical term **given** a household in region E, what is the probability of having a blue car? We know this from the first tree above:

$$p(\text{color} = B|\text{area} = E) = \frac{17}{40} \ . \tag{3}$$

Ok We can now simply use the product rule, or essentially logic. We know 40% live in area E **of which** 42.5%(17/40) have a blue car. So, the probability of living in E **AND** having a blue car must be $0.4 \times 0.425 = 0.17$. This way of thinking corresponds to traversing the first tree. So we can write,

$$p(\text{area} = E, \text{color} = B) = p(\text{color} = B|\text{area} = E)\ p(\text{area} = E) \ . \tag{4}$$

In the second way, we can traverse the second tree. As we can see the probability of having a blue car is 0.36. The probability of **living in area E among the people with a blue car** is $17/36 \approx 0.422$. So of $o.36$ with a blue car 0.472 live in area E. Hence the probability of having a blue car and living in E must be $0.36 \times (17/36) = 0.17$. As it should be! For this way of thinking we can write:

$$p(\text{area} = E, \text{color} = B) = p(\text{area} = E|\text{color} = B)\ p(\text{color} = B) \ . \tag{5}$$

As we discussed both ways should give us the same result and they do.

$$\begin{aligned} p(\text{area} = E, \text{color} = B) &= p(\text{color} = B|\text{area} = E)\ p(\text{area} = E) \\ &= p(\text{area} = E|\text{color} = B)\ p(\text{color} = B) \ . \end{aligned} \tag{6}$$

This was an example. We can now write this for any two random variables as:

$$p(x, y) = p(x|y)\ p(y) = p(y|x)\ p(x) \ . \tag{7}$$

And what is usually called the **Bayes rule** is derived from the Eq. (7):

$$\boxed{p(x|y) = \frac{p(y|x)p(x)}{p(y)} \ . \qquad \text{Bayes Rule.} \qquad (8)}$$

There are some naming convention which would make more sense as we look at a our examples.

- $p(x)$: *Prior* or *Initial Belief*
- $p(y|x)$: *Likelihood*
- $p(x|y)$: *Posterior* or *Updated Belief*
- $p(y)$: *Marginal likelihood*

Some remarks:

- $p(x)$ is a probability distribution, in the sense that it is summed up or integrated to 1 ($\sum_x p(x) = 1$). The same for $p(y)$.
- $p(y|x)$ **is** a probability distribution **over** y. This is the probability of having y **given** some information about x. Because y must take a value anyway and given the new information the corresponding distribution should still sum up to 1. But $p(y|x)$ is **not** a distribution over x. The same story for $p(x|y)$.
- The reason for the name of $p(x)$ and $p(x|y)$ as *prior* and *posterior* or *initial* and *updated* will become clear later on. But the general mindset is that, you have a belief or *initial* knowledge, $p(x)$, and after getting more info, doing experiments or collecting data you *update* your belief about x and reach $p(x|y)$.
- Usually $p(y)$ does not get that much attention, since there is no need for it. Sometimes it is called the *normalization* factor. You could see that the denominator in the right hand side of Eq. (8) is the sum of its nominator, $p(y) = \sum_x p(y|x)p(x)$ by the rules of statistic, logic or common sense. And hence the name normalization. You could also reach the same conclusion by summing both sides over x: $\sum_x p(x|y) = 1$. Hence we have $\sum_x \frac{p(y|x)p(x)}{p(y)} = \frac{1}{p(y)} \sum_x p(y|x)p(x) = 1$ which is the same as what we had earlier.

These names and how we should use the Bayes rule as written in Eq. (8) will become clear in the following examples.

EXAMPLES

Wet ground and raining

This is the example 1.19 in Ref. [1].

Let's consider two random variables which could be true (1) or false (0). The first variable, r, shows if it has rained before you woke up ($r = 1$) or not ($r = 0$). The second variable, w shows whether the ground is wet ($w = 1$) or not ($w = 0$). We know that "if it has rained, the ground has to be wet". So $r = 1 \Rightarrow w = 1$.

$$p(w = 1|r = 1) = 1 \ , \qquad p(w = 0|r = 1) = 0 \ . \tag{9}$$

Now let's say it is the morning an you have just woken up. You open the curtains and see the ground is wet. What can you **infer**? Has it rained?

$$\begin{aligned} p(r = 1|w = 1) &= \frac{p(w = 1|r = 1)\ p(r = 1)}{p(w = 1)} \\ &= \frac{p(r = 1)}{p(w = 1)} \ . \end{aligned} \tag{10}$$

But we know that in general the probability of having a wet ground is less than. So $p(w = 1) < 1$. Hence we conclude that,

$$p(r = 1|w = 1) > p(r = 1) \ , \tag{11}$$

which is really nice and intuitive. Before opening the curtains you had some *initial belief* about the rain over night represented by $p(r)$. Now that you see the ground is wet, you *update* your belief and reach $p(r = 1|w = 1)$ believing that it is more likely that there has been a rain. Well, you can not be sure since there could be other reasons for a wet ground like your neighbors' sprinklers.

Infection and test

This is the example 2.3 in [2]. We add a bit to it.

Many recall the Covid-19 pandemic. During the pandemic many got tested when they had symptoms. One, however, can ask what is the probability of having the disease if the test came out positive.

To do so we need to recall that tests are not definitive, and have a probabilistic interpretation. Given an infected sample to test does not always come back with a positive result. The test could make a mistake. We show sample with s which could be 0 meaning healthy/not infected or 1 meaning infected. We show a test's result with t and it also could be 0 or 1. The test could make a mistake and have “false negative” or “false positive”. If we give a sample with virus to the test, it makes mistake in 5% of the time and says the sample was healthy. We can show this as:

$$p(t = 1 \mid s = 1) = 0.95 \ , \qquad p(t = 0 \mid s = 1) = 0.05 \ . \tag{12}$$

In the same manner the test could make mistake if we give it a healthy sample, recognize it as an infected one:

$$p(t = 1 \mid s = 0) = 0.03 \ , \qquad p(t = 0 \mid s = 0) = 0.97 \ . \tag{13}$$

Now let's say we know that 2% of the population is infected. Hence for a random person the chance of being infected is $p(s = 1) = 0.02$ if we do not have any info. Let's add a superscript to it reflect this:

$$p^{(0)}(s = 1) = 0.02 \ . \tag{14}$$

This is our initial belief; when we don't know anything about the person we get the general knowledge. Now let's say we take a test and “update our belief” using the Bayes rule.

$$\forall t \ : \ p(s = 1 \mid t) = \frac{p(t \mid s = 1)p(s = 1)}{p(t)} \ . \tag{15}$$

So we can update our belief based on the test result. Let's say it was positive. Then we need to know $p(t = 1)$ to be able to calculate the denominator. Well, there are two ways to get a positive test result. The test could give a positive answer to an infected sample (“true positive”) or it could make a mistake and give a positive answer to a healthy sample (“false positive”). This intuition is reflected in the following sum rule:

$$\begin{aligned} p^{(0)}(t = 1) &= \sum_s p(t \mid s)p^{(0)}(s) \\ &= p(t = 1 \mid s = 1)p^{(0)}(s = 1) + p(t = 1 \mid s = 0)p^{(0)}(s = 0) \\ &= 0.95 \times 0.02 + 0.03 \times 0.98 \\ &= 0.0484 \ . \end{aligned} \tag{16}$$

So overall the test “overestimates” the percentage of the infected population, since it makes mistake. Note that $p(t = 1)$ does depend on our belief about the person and hence has the superscript $p^{(0)}(t = 1)$; this shows our belief about the test to come back positive for this person. The part $p(t \mid s)$, however, is the test property and would not change if we take more tests as we do below.

We can now go back and get an update belief for a person with positive test:

$$\begin{aligned} p^{(1)}(s = 1 \mid t = 1) &= \frac{p(t = 1 \mid s = 1)p^{(0)}(s = 1)}{p^{(0)}(t = 1)} \\ &= \frac{0.95 \times 0.02}{0.0484} \\ &\approx 0.39 \ . \end{aligned} \tag{17}$$

Given the new info we have an updated belief about this person, and it is considerably higher than what we believed before (3%).

Let's say we want to take a second test. We can now use our belief based on the first one:

$$p^{(2)}(s = 1 \mid t) = \frac{p(t \mid s = 1)p^{(1)}(s = 1)}{p^{(1)}(t)} \ . \tag{18}$$

As it is clear we need calculate $p^{(1)}(t)$. To use the sum rule we we would use updated beliefs for $p(s)$, because we are asking what is the *probability having a positive result for this person* and not for the general public or a random person:

$$\begin{aligned} p^{(1)}(t=1) &= \sum_s p(t \mid s) p^{(1)}(s) \\ &= p(t=1 \mid s=1) p^{(1)}(s=1) + p(t=1 \mid s=0) p^{(1)}(s=0) \\ &= 0.95 \times 0.39 + 0.03 \times 0.61 \\ &= 0.3888 \ . \end{aligned} \tag{19}$$

As we see the probability of having the positive result also increased! Let's assume the second test does also show positive result:

$$\begin{aligned} p^{(2)}(s=1 \mid t=1) &= \frac{p(t=1 \mid s=1) p^{(1)}(s=1)}{p^{(1)}(t=1)} \\ &= \frac{0.95 \times 0.39}{0.3888} \\ &\approx 0.95 \ . \end{aligned} \tag{20}$$

That is quite interesting! The second positive test makes us more confident that the person was actually infected, which makes sense.

Examples 2.6 in [2] is also interesting.

[1] A. Krause and J. Hübotter, Probabilistic artificial intelligence (2025), arXiv:2502.05244 [cs.AI] .
[2] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms* (Cambridge university press, 2003).