Bayes theorem with a few simple example of Bayesian inference

Iman Mahyaeh
(Dated: October 4, 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.

$$\begin{aligned} p(\text{area} = E, \text{color} = B) &= p(\text{color} = B|\text{area} = E) \ p(\text{area} = E) && (6) \\ &= p(\text{area} = E|\text{color} = B) \ p(\text{color} = B) \ . && (7) \end{aligned}$$

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

$$p(x, y) = p(x|y) \ p(y) = p(y|x) \ p(x) \ . \tag{8}$$

And what is usually called the **Bayes rule** is derived from the Eq. (8):

$$\boxed{p(x|y) = \frac{p(y|x)p(x)}{p(y)} \ . \qquad \text{Bayes Rule.} \qquad (9)}$$

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. (9) 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. (9) will become clear in the following examples.

EXAMPLES

Rain and wet ground

a

infection and test

b

infer parameter

c