

If you remember our discussion from a long time ago, we said that much of this class consists of variations of a few basic skills and ideas, one of which is the Bayes rule, the foundation of inference.

So let's look here at the Bayes rule again and its different incarnations.

In a discrete setting we have a random variable with a known PMF but whose values are not observed.

Instead we observe the value of another random variable, call it Y , which has some relation with X .

And we will use the value of Y to make some inferences about X . The relation between the two random variables is captured by specifying the conditional PMF of Y given any value of X . Think of X as an unknown state of the world and of Y as a noisy observation of X . The conditional PMF tells us the distribution of Y under each possible state of the world.

Once we observe the value of Y we obtain some information about X . And we use this information to make inferences about the likely values of X . Mathematically, instead of relying on the prior for X , we form some revised beliefs.

That is, we form the conditional [PMF] of X given the particular observation that we have seen.

All this becomes possible because of the Bayes rule.

We have seen the Bayes rule for events.

But it is easy to translate into PMF notation.

We take the multiplication rule.

And we use it twice in different orders to get two different forms-- or two different expressions-- for the joint PMF.

We then take one of the terms involved here and send it to the other side.

We obtain this expression, which is the Bayes rule.

What [do] we have here?

We want to calculate the conditional distribution of X which we typically call the posterior.

And to do this we rely on the prior of X as well as on the model that we have for the observations.

The denominator requires us to compute the marginal of Y . But this is something that is easily done because we have the joint available.

The numerator, this expression here, is just the joint PMF.

And using the joint PMF you can always find the marginal PMF.

Essentially, we're using here the total probability theorem.

And we're using the pieces of information that were given to us, the prior and the model of the observations.

When we're dealing with continuous random variables the story is identical.

We still have two versions of the multiplication rule.

By sending one term-- this term-- to the other side of the equation we get the Bayes rule.

And then we use the total probability theorem to calculate the denominator term.

So as far as mathematics go, the story is pretty simple.

It is exactly the same in the discrete and the continuous case.

This story will be our stepping stone for dealing with more complex models and also when we go into more detail on the subject of inference later in this course.