

We have seen two versions of the Bayes rule-- one involving two discrete random variables, and another that involves two continuous random variables. But there are many situations in real life when one has to deal simultaneously with discrete and continuous random variables. For example, you may want to recover a discrete digital signal that was sent to you, but the signal has been corrupted by continuous noise so that your observation is a continuous random variable.

So suppose that we have a discrete random variable  $K$ , and another continuous random variable,  $Y$ . In order to get a variant of the Bayes rule that applies to this situation, we will proceed as in the more standard cases. We will use the multiplication rule twice to get two alternative expressions for the probability of two events happening. We will equate those expressions, and from these, derive a version of the Bayes rule.

So we will look at the probability that the discrete random variable takes on a certain numerical value, and, simultaneously, the continuous random variable takes a value inside a certain small interval. So here,  $\delta$  is a positive number, which we will take to be very small. And in fact, we will be interested in the limiting case as  $\delta$  goes to 0.

So now we use the multiplication rule. The probability of two events is equal to the probability of the first event times the conditional probability of the second event given that the first event has occurred. But we know that we can use the multiplication rule in any order, so the probability of two events happening can also be written as the probability that the second event occurs times the conditional probability that the first event occurs, given that the second event has occurred. So these two expressions that we obtain from the multiplication rule have to be equal. Let us rewrite those expressions using PMF notation and PDF notation.

What do we have here? The probability that a discrete random variable takes on a certain value-- that's just the PMF of this random variable evaluated at a particular point. And what do we have here? The probability that the random variable,  $Y$ , a continuous random variable, takes values inside an interval is always equal to the PDF of that random variable times the length of this interval. And this is an approximate equality.

However, because here we're talking about the probability of being in a small interval conditioned on a

certain event, we should be using a conditional PDF. It's the conditional PDF conditioned on the random variable, capital  $K$ , and conditioned on the specific event that this discrete random variable takes on a certain value, little  $k$ . Let us do a similar notation change for the second expression. Here we have the probability-- the unconditional probability-- that  $Y$  takes a value inside a small interval, and when  $\delta$  is small, this is approximately equal to the PDF of the random variable  $Y$  times the length of the interval.

And what do we have here? The probability that a discrete random variable takes on a certain value, that just corresponds to the PMF of that the random variable. However, we're talking about a conditional probability given that a random variable  $Y$  takes a value that's approximately equal to a certain little  $y$ .

So this is a notation that we have not used before, but its meaning should be unambiguous at this point. But just by arguing by analogy to what we have been doing all along, it's a PMF of a discrete random variable. But it is a conditional PMF. It describes to us the probability distribution of the discrete random variable  $K$  when the random variable  $Y$ , which happens to be a continuous one, takes on a specific value.

So we can cancel the  $\delta$ s from both sides, and we have that this expression is approximately equal to that expression, and this approximate equality is more and more exact as we send  $\delta$  to 0. But  $\delta$  has already disappeared from here, so we can set these two expressions equal to each other. At this point, now, we can take this term and move it to the other side of the equality so it will go to the denominator. And we obtain this version of the Bayes rule.

It gives us the conditional probability of a random variable  $K$  given that a certain continuous random variable  $Y$  has taken on a specific value. So this version is useful if we have a continuous noisy observation,  $Y$ , on the basis of which we're trying to say something, to make inferences about the discrete random variable  $K$ . And in order to apply the Bayes rule, we need to know the unconditional distribution of the random variable  $K$ , and we also need to have a model of the noisy observation-- a model of that observation under each possible conditional universe. So for any possibility for the random variable  $K$ , we need to know the distribution of the random variable  $Y$ .

Or, alternatively, we can take this term and send it to the denominator of the other side, and we get a different version of the Bayes rule. This version of the Bayes rule applies if we're trying to make an inference about a continuous random variable  $Y$ , given that we know the value of a certain related

observation,  $K$ , of a random variable, capital  $K$ .

In both versions of the Bayes rule, there's also a denominator term which needs to be evaluated. This term gets evaluated similar to the cases that we have considered earlier, and they are determined by using a suitable version of the total probability theorem. This is a version of the total probability theory that we have already seen. We have a conditional density of  $Y$  under different scenarios for the random variable capital  $K$ , and we get the density of  $Y$  by considering the conditional densities and weighing them according to the probabilities of the different discrete scenarios.

This version of the total probability theorem is something that we have not proved so far, and we have not seen it. On the other hand, it's not hard to derive.

If we fix the value of  $k$ , this is a density, and therefore it must integrate to 1. So the integral of this ratio, with respect to  $y$ , has to be equal to 1. Now, there's no  $y$  in the denominator, so the integral of the numerator divided by the denominator has to be equal to 1, which means that the denominator must be equal to the integral of the numerator when we integrate overall  $y$ 's, and this is just what this expression is saying.

So what we will do next will be to consider one example for each one of these two cases of the Bayes rule that we have just derived.