We will now use the Bayes rule in an important application that involves a discrete unknown random variable and a continuous measurement. Our discrete unknown random variable will be one that takes the values plus or minus 1 with equal probability. And the measurement will be another random variable, Y, which is equal to the discrete random variable, but corrupted by additive noise that we denote by W. So what we get to observe is the sum of K and W.

This is a common situation in digital communications. We're trying to send one bit of information whether K is plus 1 or minus 1, but the observation that we're making is corrupted by a communication channel, by some noise that is present in the channel, and on the basis of the value of Y that we will observe, we will try to guess what was sent. The assumption that we will make about the noise is that it is a standard normal random variable.

So suppose that we observed a specific value for the random variable Y. We want to make a guess about the random variable capital K. Of course, there's no way to guess with complete certainty. The only thing that we can say is to determine how likely it is that a 1 was sent as opposed to how likely it is that a minus 1 was sent. How do we approach such a problem? Well, we use the version of the Bayes rule that we have already developed, which is this formula that gives us the conditional probabilities that we want.

And in particular, here, we're asking a question about the conditional probability that K takes the value of 1 given that a value of y has been observed. This is what we want to calculate.

So let us look at the various terms involved here and see what each term is. First, we need the prior probability of K. This is simple. The prior probabilities are 1/2 for k equal to minus 1 or plus 1, because we said that the two possibilities are equally likely.

Then we need the conditional density of Y given K. So what does this assumption mean? It means that Y is a standard normal random variable to which we add the value of K. So if K is equal to 1, we're taking a standard normal, and we add a value of plus 1. So Y, given that K is equal to plus 1, is going to be a standard normal plus 1.

What does that do? If we take a standard normal and add a constant to it, that changes the mean and

makes the mean equal to 1, and does not change the variance. On the other hand, if K happens to be equal to minus 1, then the observation that we see is going to be a standard normal plus a value of minus 1, and that changes the mean to become minus 1, but with a variance of 1.

So if we are to plot the density of Y, that density, of course, will depend on what the value of K was. And if K is equal to 1, then we will obtain a normal that has a mean of 1, so it's centered here. But if K is equal to minus 1, then our observation will be a normal with unit variance, but centered at minus 1.

So if we are to write this in terms of symbols, the distribution of Y is normal with variance equal to 1. So the PDF is given by this form, e to the minus 1/2 y minus the mean of Y. But given the value of K, the mean of Y is equal to k, plus or minus 1, depending on what k is.

So this is the PDF of a normal with unit variance and mean equal to k. And it corresponds, when you set k equal to 1, it corresponds to this graph. When you set K equal to minus 1, it corresponds to that graph.

Let us continue with the next term in this expression. We need the term in the denominator, which is obtained by taking a sum over the different choices of k. There are 2 choices, and each choice has a probability of 1/2. From the first choice, we have 1/2 times the density of Y when k is equal to minus 1. And when k is equal to minus 1, we obtain this expression.

And we have another term that corresponds to the case where k is equal to plus one, in which case we have this expression here. Once more, this expression here corresponds to this normal with a mean of minus 1. This expression here corresponds to a normal with a mean of plus 1, which is this graph here.

So at this point, we have in our hands expressions for everything that is involved here, and we can just apply the formula and carry out a fair amount of algebra. There are some very nice simplifications that happen along the way, and we end up with an answer that has the following form. It's 1 divided by 1 plus e to the minus 2 y. And this gives us the probability that a 1 was sent.

Let us try to make sense of this expression. Let's see what it looks like by plotting it as a function of y.

So what we're plotting here is this expression. OK, if y is very large, as y goes to plus infinity, this term disappears, and we obtain a 1. If, on the other hand, y is very, very negative-- so y goes to minus infinity-- here we get to e to the infinity, which is a very large number. So this ratio is going to converge

to 0. So we have a graph that starts at 0. It actually rises monotonically, and in the limit, converges to 1. If y is equal to 0, then this term is 1, and we obtain a 1/2.

Let us interpret this plot. If y is very large, it is much more likely that y is coming out of this distribution so that K is equal to 1. So the probability that K is equal to 1, if we obtain this observation, is almost 1. We have almost certainty.

If, on the other hand, y is very, very negative, then it is much more likely that what we're seeing is coming from this distribution so that K is equal to minus 1. And in that case, the probability that K was 1 is going to be approximately 0. Finally, if y is 0, then we're just in the middle of the two possibilities, and by symmetry, either choice of K is equally likely. Therefore, the posterior probability that K is equal to 1, given that Y was equal to 0-- that probability is 1/2. When Y is equal to 0, it's equally likely that either signal was sent.

This example is a prototype of the kind of calculations that are done in the analysis of communication systems. This is the simplest model of communication of a single bit in the presence of additive noise, but of course, there can also be more complicated models in which we have more complicated signals that are sent, and more complicated models of the noise. But the general principles of the analysis are always of this kind. We're using the Bayes rule, and we need to write down the different terms that are involved.