Let us now overview and summarize the contents of the sixth unit in this class. Actually, this unit consisted of three rather separate topics. So these are topics that go deeper into the general subject. But they were not related with each other. For this reason, lets us talk about each one of them separately.

The first topic was derived distributions. The idea is that we have a random variable X whose distribution is known. And we want to somehow calculate the distribution of Y. For continuous random variables, the way that we do that is by first finding the CDF of the random variable Y and then differentiating to find the PDF of Y. But if the function g turns out to be monotonic, then we also have a direct formula that takes us straight to the answer.

Then we saw that for the special case of a linear function, the formula is rather simple. And what we do is that we take the PDF of X, scale it, and shift it. And that gives us the PDF of Y. Finally, the same methodology applies even if we are dealing with a function of multiple random variables. Once more, what we would do would be to calculate the CDF of the random variable Z and then differentiate.

Moving further, a very interesting and important special case of a function of two variables is when we take those two variables and add them. If those two variables X and Y are also independent, then there are nice formulas for the distribution of Z. In the discrete case, there's a formula.

And there's an analogous formula in the continuous case where, as usual, the PMFs get replaced by PDFs and sums get replaced by integrals. This operation is called the convolution, convolution of two PMFs, or convolution of two PDfs. And there are also some nice mechanical ways of organizing the calculations that are involved here. Finally, an important fact which is derived by using this convolution formula is the following-- that if we're dealing with two independent normals, then their sum will also be a normal random variable.

Moving to the second topic, we defined the covariance of two random variables using this formula. And we interpreted it that it captures in some way whether X and Y deviate from their means in a coordinated way or not. So when X is higher than the mean, does Y also tend to be higher? Well, the covariance is trying to measure a phenomenon of this kind.

Besides interpretations, we also looked at some algebraic properties of the covariance. We saw that it is linear in each one of the arguments. And also, we saw that covariances are useful when we want to find the variance of the sum of a collection of random variables when those random variables are dependent. So instead of getting just the sum of the variances, we also get a bunch of cross terms involving the covariances between the different random variables.

We then introduced the correlation coefficient, which is a scaled version of the covariance that takes into account the standard deviations. And what happens is that the correlation coefficient is a dimensionless quantity that is guaranteed to be always less than or equal to 1 in magnitude.

Finally, the last topic was to revisit the subject of conditioning. And we looked at the conditional expectation and the conditional variance. And instead of having lowercase symbols here, we now introduced uppercase symbols. This is a change in notation. But more important, conceptually, we started looking at these quantities as random variables.

The conditional expectation is determined by Y. And since Y is random, we can think of this quantity as being random by itself. Since they're random variables, they have expectations, variances, and so on. So we started looking at some of their properties. And for example, the expected value of the conditional expectation turns out to be the same as the unconditional expectation.

Now this law of iterated expectations is just a version of the total expectation theorem, but written in a more abstract form. It accomplishes the same divide and conquer approach to problems. But it has an aesthetic appeal the way it is written because it doesn't involve any assumptions whether X and Y are discrete, continuous, and so on.

There's also a formula-- the law of total variance-- that relates the conditional variance to the unconditional one. But unlike what one might expect, there's also an additional term that appears here. And we saw an interpretation of these two terms in the context of a few examples that hopefully shed light into what we're dealing with. It's a decomposition of the total amount of uncertainty into uncertainties of two different types.

Finally, a key application of these two laws that we have in our hands was in the context of calculating or doing something about random variables that are the sum of a random number of independent random variables. And we saw that the expected value of the sum is equal to the product of two expectations. We have the expectation of how many terms we are adding times the expected value of each one of the terms. And as far as the variance is concerned, we applied the law of total variance to get an expression of a certain type.

So the topics that we covered in this unit start to become pretty sophisticated. And by this point, we are ready to dive into pretty deep and mature applications of probability.