

module_prob4_lecture2

Fri, 2/18 12:51PM 14:47

SUMMARY KEYWORDS

expectation, random variable, $x - \mu$, covariance, measure, relationship, product, μ , average, x tilde, interested, variance, crime rate, unemployment rate, similarly, formula, multiply, tilde, sigma, red dots

SPEAKERS

Sumon Majumdar



Sumon Majumdar 00:05

So the last module on discrete variables, we talked about two important measures about the random variable, one was expectation, and the other was variance. So continue with that introducing another third useful measure, which is called the covariance and a related measure called correlation. So, so far, we have looked at expectation, which tends to measure the average value of the mean value of the random variable. And this was denoted by μ of X . And the other measure was variance, which is var of X tilde or σ^2 of X . And this measure measures the dispersion of the spread of the random pair. So, now, the new measure which is called covariance, so, this is tries to measure the relationship between to run. So, if you have random variables X and Y , what you may be interested in is that as X becomes bigger does Y also go up. So, for example, you may be interested in the relationship between greenhouse emissions and temperature. Or you may be interested in the number of hours that one studies and the mark on ones in the course. Or you may be interested, for example, unemployment rate and the crime rate does as unemployment rate goes up is crime rates, also more likely to increase? So, this sort of relationships, the measurement between the measurement of the relationship between two random variables X and Y is given by a measure called covariance. So, there are lots of examples where it's useful, for example, apart from the ones that I've also already say, is, you may be interested in the relationship between height and weight of of people, right? The people who are taller, are they also more likely to be way more. So, more precisely what we want to mean is that people who are taller than the average, are their weights also more likely to be above average? Or, or let's say, in cities, where the unemployment rate is higher than average, or in periods with unemployment rate is higher than average, is the crime rate also more likely to be evolved? So what's the basic idea of how do we measure covariance? So let me show that in terms of a graph, or in terms of a plot, let's hear, so let's first focus on the blue dots. Right so, the blue dots are so particular blue dot, is, a particular value of X and a particular value of one. So if we look at the blue dots, there is obviously an upward sloping relationship between red that means whenever X value is high, that means the Y value is also more likely to be high. Or when X is low, it's remember this is the negative part. So when X is likely to be low, Y is also likely to be low. So this is the negative one. And similarly here, when X is more likely to be positive, Y is also positive. So how do we actually capture this relationship? So one way to do that is to look at a product of X and Y . So

see here, whenever X is positive, Y is also positive. So that means if we take the product of X and Y , it's going to be positive. Here to read whenever X is negative, X is negative, Y is also negative. So if we take the product here, it's the product of two negative numbers, that's also going to be positive. Now the opposite case is the red dots, right? So in the red dots, you see, there's a downward sloping relationship. In this red dots whenever X is positive, right, the Y value is negative. Or vice versa, that like if you look at a point like here, the X is negative but the Y is positive. So for these cases, in the case of the red dots, if you take the product of X times Y , that's always going to be. And this is the idea that we'll take to, to construct a measure of this relationship. So in particular, now we take the same X and Y 's, right? But now suppose I put in the mean, right, so this is the mean value of X , that's given by the green line here. And this is the mean value of y , which is the, the average value of Y . So if we take a point, like here, so this point, the X is greater than it's been, right, and the Y is also greater than its me. Similarly, if you see this blue dot here, the X and the Y , they're both below their means. So among the blue dots, right, whenever the X value is above its mean. The Y value is also likely to be above it's mean. On the other hand, here, too, right? Like, whenever X is below its mean value, the Y is also more likely to be below. So if we look at the product of X minus μ of X , right? So this tells me whether x is above its mean, or below, it's B , right? And similarly, Y minus μ of Y , which tells you whether Y is above its mean, or below its mean, right? So whenever the product of these two, right, in cases of the blue dot, this product is always positive, not always red. So there are some cases like here where the X is above its mean, but the Y is below, right? But on average, is this product greater than zero, right? Then it means there is a positive relationship. On the other hand, if you look at the red dots, right, whenever X is above, its mean, right? So this is a value where X is greater than its mean. But the Y is below that for the red dots, if we look at the same thing, X minus μ of X times Y minus we will fly this product is. So that means if we look at this product, right, that's going to give us a good idea about whether this is a positive relationship between X and Y , or is it a negative relationship between. and covariance is exactly that. It's the expectation of X minus μ of X times Y minus μ . So that's the definition of covariance. It's the expectation of X minus μ of X times Y minus μ . And this is denoted by σ_{xy} , right? So this is the Greek letter sigma. And how would you calculate actually this covariance, it's the same idea that we were using so far to calculate expectation. Remember, expectation of an any random variable is we take the value of the random variable multiplied by its probability, and then added up across all possible values. Here, there are two random variables X and Y . So what we do is we take X particular value of X subtracted from it, mean μ_x , take a particular value of Y , subtract it from its own mean new μ_y , right? Multiply those two, and then multiply this whole thing by the probability of X and Y . And we do this over all possible values of X and Y . And if we do that, then we get the expectation of X minus μ_x times Y minus μ_y . In other words, the covariance so once more the covariance the definition of it is X minus μ_x times Y minus μ_y . And what it measures is the relationship between two random variable Now, let me give you an easier formula like we did for variance. Remember, for variance, we have this easier formula that it is the expectation of X squared minus expectation of X whole thing squared. So can we get a slightly easier formula for the covariance as well. And we can, basically, what we're going to do is see this is going to be the product of two terms, right? So let's open up those brackets. Right, so if we are doing the product of these two, so it will be x times y minus this x times this, so minus $\tilde{x}$ times Y , then I'm going to multiply this and this. So that's minus μ_x times $\tilde{y}$. And then last thing, I'm going to multiply this and this, it's going to be plus μ_x . Now you know that if you have to take expectation of this, this, this and this, we can do it separately. So that means we can do this as expectation of $\tilde{x}$ - $\tilde{y}$. So that's this first here, minus the second one, which is this here, which is expectation of $\tilde{x}$ times one minus this third one, which is expectation of new X times $\tilde{y}$. And lastly, this here, which is expectation of new X . And I'm going to use one of the properties of expectation that we learned in the last module. See this

here, new effects is a constant μ of y is a constant expectation of a constant is just constant. Similarly, if we look at the third term, this μ of X , this is a constant, so I can just pull it outside and multiplied this expectation of y tilde. And similarly, here, this μ of Y is a constant, I can pull it outside, and this becomes expectation. And we're nearly there. See, this is going to be expectation of X tilde Y tilde, I'm just leaving it as it is minus this is μ of Y . And this is expectation of X . So that's μ of X . Then I go to the third term, so that's minus μ x . And this is expectation of Y , that's just going to be Y . And the last bit last term here is UX times Y . Now see, lots of these terms are, we can cancel on some of these terms, we can cancel out this with this. So that means the formula for covariance is just going to be expectation of X tilde Y tilde minus μ of Y is expectation of Y times expectation of X . So this here is going to be my easier formula for computing covariance. And I'm going to show it to you later in an example exactly how to use this formula. So, so, what we have just derived is this easier formula for covariance, which is expectation of $X Y$ minus expectation of X times expectation of Y . So let me stop this clip here. So two things to remember one is covariance measures the relationship between two random variables. And secondly, the formula for it is given by expectation of XY minus expectation of X times expectation of Y .