PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

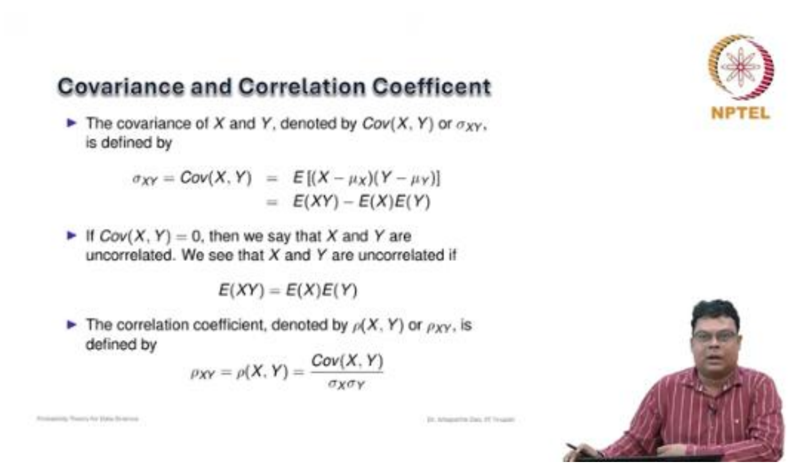
Week - 08

Lecture - 42

Association Between Two Random Variables

If X and Y are independent random variables, then the covariance between X and Y is defined as $E[(X - \mu_X) * (Y - \mu_Y)]$.

Consider the case where X and Y are continuous random variables. The discrete case will follow a similar pattern, but instead of integration, we will use summation. For continuous random variables, the integration is carried out from $-\infty$ to $+\infty$.



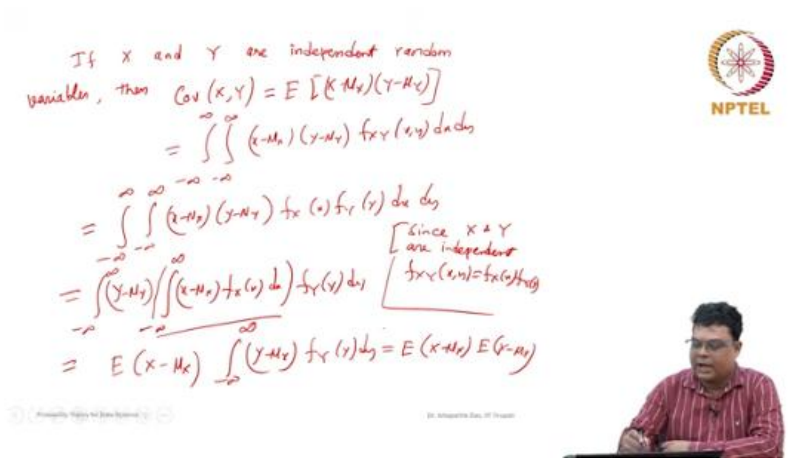
The expected value is computed as $E[(X - \mu_X) * (Y - \mu_Y)]$ with the joint probability density function $f_{XY}(x, y)$ integrated over both variables.

If X and Y are independent, the joint probability density function $f_{XY}(x, y)$ can be written as the product of the marginal probability densities of X and Y. This means $f_{XY}(x, y) = f_X(x) * f_Y(y)$. This allows us to separate the variables in the integration process.

When we perform the integration with respect to X and Y separately, we can first evaluate the part of the integral involving X. Since the term $(X - \mu_X)$ is multiplied by the marginal probability density $f_X(x)$, we can treat it as $E[X - \mu_X]$, and the result will be a real number that can be factored out of the integral.

After performing the integration for both X and Y, we find that the covariance of X and Y simplifies to the product of $E[(X - \mu_X)]$ and $E[(Y - \mu_Y)]$.

So, here we discuss a case where X and Y are considered as a bivariate continuous random variable, with the joint probability density function $f_{XY}(x, y)$. This is the assumption we make for the computation.



If X and Y are discrete, we can write them as a discrete bivariate random variable with a probability mass function P_{XY}. Since X and Y take discrete values, the notation used is P_{XY} for the probability mass function.

Similarly, you can show that the covariance of X and Y, where X and Y are discrete bivariate random variables with a probability mass function, and X and Y are independent, can be expressed as:

$$Cov(X, Y) = \Sigma \left((X - \mu_X) * (Y - \mu_Y) * P_{XY}(x, y) \right)$$

The joint probability mass function can be represented by the product of their marginal values. If the values are finite, the summation will be from 1 to n, but if the values are countably infinite, the summation can extend beyond that.

Therefore, we can show that the covariance of X and Y is the summation over the possible values of X and Y:

$$Cov(X, Y) = \Sigma ((X - \mu_X) * (Y - \mu_Y) * P_{XY}(x, y))$$

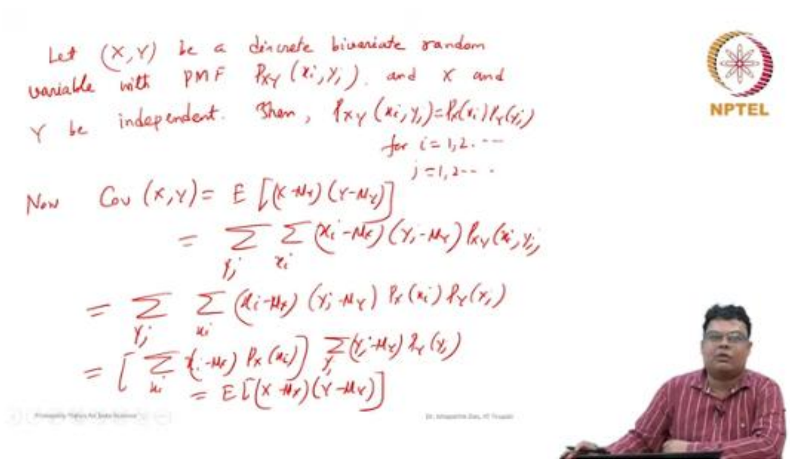
This is simply the expected value of $(X - \mu_X)$ multiplied by $(Y - \mu_Y)$. In both cases, whether continuous or discrete, we find that if X and Y are independent random variables, then the covariance of (X, Y) can be written as:

$$Cov(X, Y) = E[(X - \mu_X) * (Y - \mu_Y)]$$

This can be expressed as:

$$E[(X - \mu_X)] * E[(Y - \mu_Y)]$$

Now, what is $(X - \mu_X)$?



By the properties of expected values, the expected value of $(X - \mu_X)$ is 0 because μ_X is a constant. The expected value of a constant is just the constant itself, so the expected value of μ_X is simply μ_X, and since the expected value of a constant is 0, this results in 0.

Similarly, the expected value of $(Y - \mu_Y)$ will be the expected value of Y minus the expected value of μ_Y. This is nothing but $\mu_Y - \mu_Y$, which equals 0.

Hence, the covariance will be 0. Therefore, if X and Y are independent random variables, their covariance will be 0.

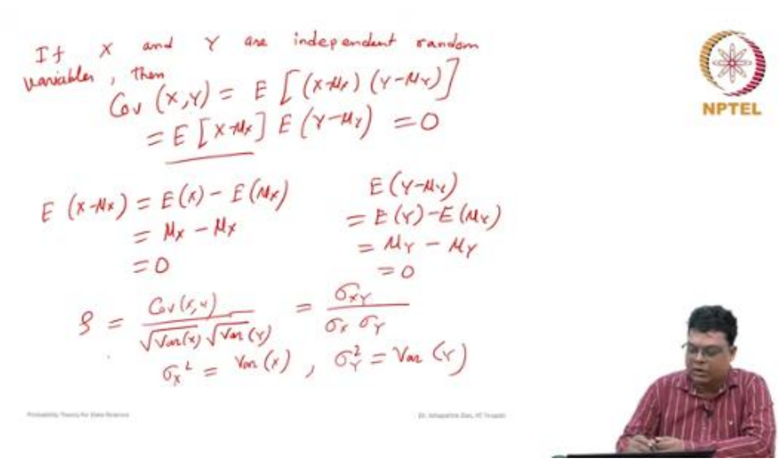
However, the converse is not always true. For example, if the covariance between two random variables is 0, we cannot necessarily conclude that X and Y are independent. There are many cases where the covariance is 0, but the variables may not be independent. Therefore, we cannot say that X and Y are independent just because their covariance is 0. However, if X and Y are independent, we can confidently say that their covariance is 0.

The correlation coefficient, denoted by ρ, is defined as:

$\rho(X, Y) = Cov(X, Y) / (\sigma_X * \sigma_Y)$

In terms of notation, the variance of X is denoted as σ^2_X, and the variance of Y is denoted as σ^2_Y. The correlation coefficient measures the strength and direction of the linear relationship between X and Y.

If X and Y are independent random variables, their covariance will be 0, indicating no linear relationship. Similarly, the correlation coefficient will also be 0 if X and Y are independent.

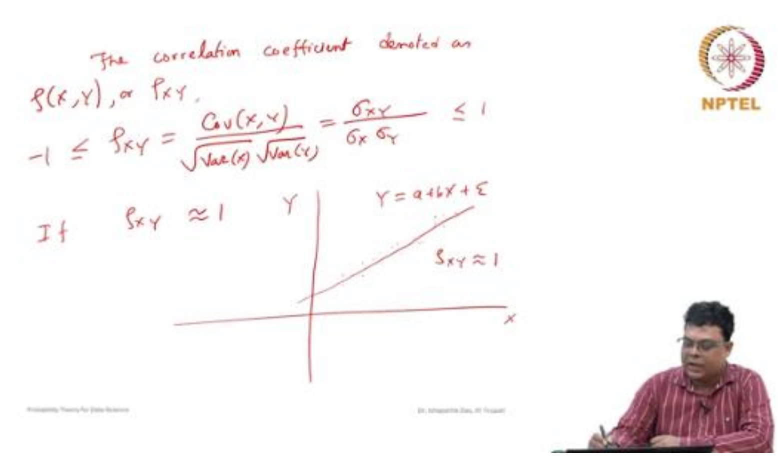


The correlation coefficient is always unitless because it is divided by the standard deviations of X and Y. It can be shown using the Cauchy-Schwarz inequality that the correlation coefficient, denoted as $\rho(X, Y)$, always has a lower bound of -1 and an upper bound of 1.

The significance of the correlation coefficient lies in its value. If $\rho(X, Y)$ is close to 1, it suggests a strong linear relationship between X and Y. For example, if you create a scatter plot of the data by plotting X versus Y, and the data points closely resemble a straight line, this indicates that as X increases, Y also increases in a linear manner.

So, Y is related to X in some way, possibly as a linear equation with a constant term and a coefficient for X, plus some error term because the data doesn't lie exactly on a straight line. The error is small, meaning that most of the points are close to the straight line. If the scatter plot looks like this, the correlation coefficient between X and Y will be approximately 1.

Now, if we observe that the data is arranged in a way that resembles a straight line, but in the opposite direction, it means that as X increases, Y decreases. This is also on the straight line.

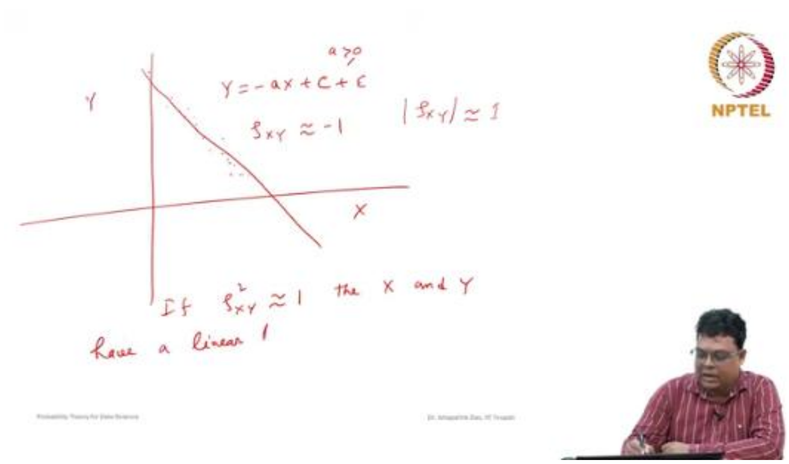


So, Y will be some negative number times X, where the constant "a" is a positive number or greater than or equal to zero. This results in an equation like Y = a negative number * X + C. If a = 0, then Y = C. However, we want to consider the more general case where a > 0. In this case, the equation would look like Y = a negative value * X + C, with some error.

The error may vary, but it is small. Most of the data will be very close to this straight line or fall on it. If the data follows this pattern, the correlation coefficient, ρ, between X and Y will be close to -1. Essentially, if you consider the square of ρ or the absolute value of ρ, and it is close to 1, then X and Y are related by a straight line, indicating a linear relationship between them.

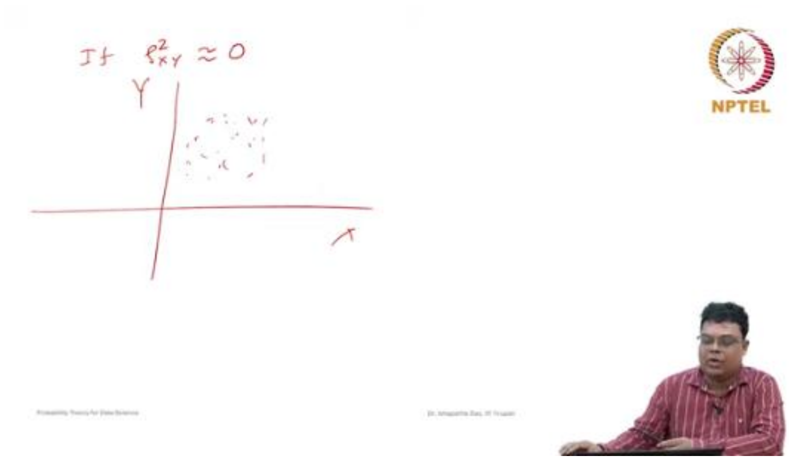
If the relationship between X and Y looks like Y = a negative value * X + C, the data will include some error. The error may vary slightly, but it is small. Most of the data will be very close to, or fall directly on, this straight line. If the data follows this pattern, the correlation coefficient, ρ, between X and Y will be close to -1. Essentially, if you consider either the square of ρ or the absolute value of ρ, and it is close to 1, then X and Y are related by a straight line. This indicates a linear relationship between X and Y.

If the square of ρ is close to 1, then X and Y have a linear relationship. If the square of the correlation coefficient, ρ, between X and Y is close to 1, then X and Y have a linear relationship.



Now, if the correlation coefficient, ρ, is close to 0, or if the square of the correlation coefficient, ρ^2, between X and Y is close to 0, then there is no linear relationship. The relationship may be non-linear, such as a circular pattern, where sometimes Y increases as X increases, and at other times Y decreases as X increases. There is no consistent straight-line relationship. In this case, the correlation coefficient, ρ, will be close to 0. These are the interpretations of the covariance and correlation coefficient that we learned.

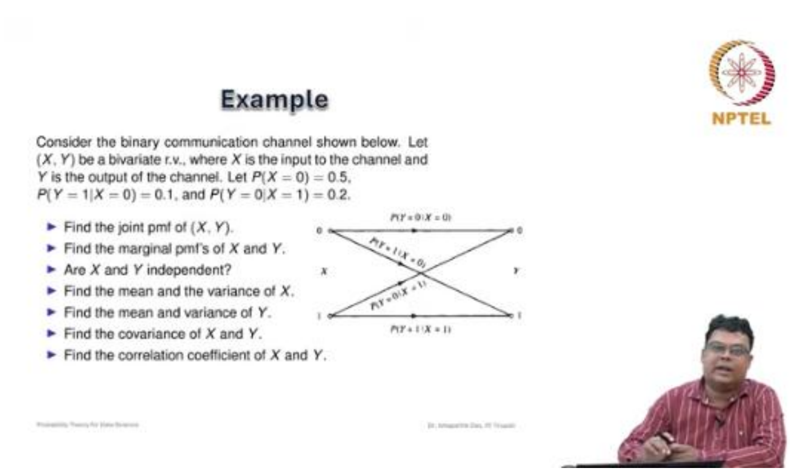
Let us discuss an example. Here, you can see that the covariance of X and Y, denoted as Cov(X, Y), is defined as the expected value of the product of $(X - \mu_X)$ and $(Y - \mu_Y)$, which we have already presented.



If the covariance of X and Y is 0, then we say that X and Y are uncorrelated. This occurs when the expected value of the product of X and Y, E(XY), equals the product of their individual expected values, E(X) * E(Y). In this case, their covariance is 0, indicating there is no linear relationship between them. As a result, the correlation coefficient, ρ(X, Y), would also be close to 0. We have discussed that if the correlation coefficient is very close to 0, then there is no linear relationship.

So, we say that they are uncorrelated. These are some of the interpretations of the covariance and correlation coefficients.

Now, let us discuss a numerical example to find the correlation coefficient and covariance. Consider a binary communication channel, as shown in the graph below. Let (X, Y) represent a bivariate unknown variable, where X is the input to the channel.



X is controlled by a switch: if the switch is on, X = 1; if the switch is off, X = 0. Y represents the output of the channel, which depends on the value of X. The relationship between X and Y can be likened to an AND gate or a NOR gate, where the input affects the output in some way. Now, X has a P(X=1) = 0.5 probability of being 1, which means it behaves like an unbiased coin toss. If the result is heads, the switch is turned on (X = 1), and if it's tails, the switch is turned off (X = 0).

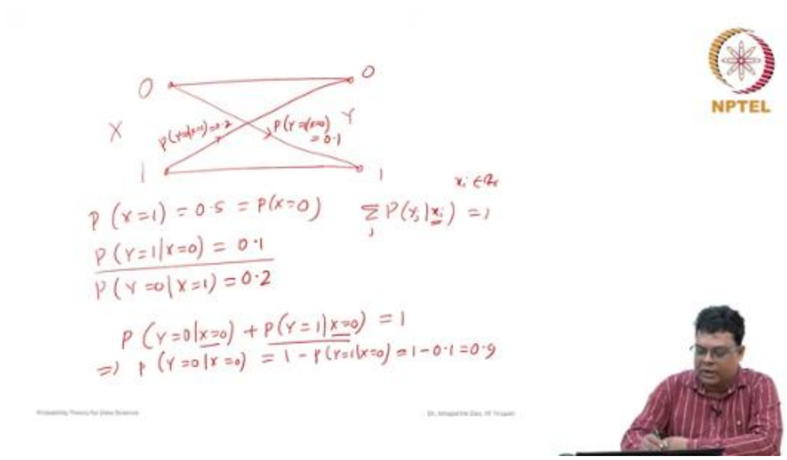
Therefore, this represents a binary system where X can be 0 or 1. When X = 1, Y can take either the value 0 or 1, with certain probabilities. It will not always be 0 or always be 1, as there is some variability in the output. These are the Y values, where 1 represents Y = 1 and 0 represents Y = 0. It is given that P(X = 1) = 0.5, which is the same as P(X = 0).

This is like tossing an unbiased coin, which is why both probabilities are equal. When X = 0, meaning the switch is off, Y can still be 1. The probability of Y = 1, given X = 0, is P(Y = 1 | X = 0) = 0.1. Similarly, the probability of Y = 0, given X = 1, is P(Y = 0 | X = 1) = 0.2. When X is on, there is a possibility that Y can be off. This probability is P(Y = 0 | X = 1) = 0.2, and the probability of Y = 1, given X = 0, is P(Y = 1 | X = 0) = 0.1. These are the given probabilities.

Now, we need to find the joint probability mass function, P(X, Y). Some of the conditional probability mass functions are provided, but not all values are given. We need to find the remaining values using the properties of the conditional probability mass function.

One important property is that for any conditional probability mass function, if we sum the conditional probabilities of Y, given a fixed value of X, the result will be 1. If you fix X = 0, the possible values Y can take are 0 and 1. Therefore, if you take the sum of P(Y = 1 | X = 0) and P(Y = 0 | X = 0), the sum should equal 1. It is important to fix the value of X before taking the sum. If you don't fix X, this property will not hold, and the sum will not equal 1.

The provided value for P(Y = 1 | X = 0) is 0.1. Thus, P(Y = 0 | X = 0) = 1 - P(Y = 1 | X = 0) = 1 - 0.1 = 0.9. From this, we can determine the conditional probability mass function. Similarly, we will apply the same process to obtain the remaining values. If you fix X = 1, then you can take the sum of the possible values Y can take. Y can be either 1 or 0.



Since X is fixed at 1, the sum will be 1. From here, we can determine P(Y = 1 | X = 1). This is 1 - P(Y = 0 | X = 1). The probability of Y = 0, given X = 1, is P(Y = 0 | X = 1) = 0.2. Therefore, P(Y = 1 | X = 1) = 1 - 0.2 = 0.8.

These probabilities are necessary to find the joint probability mass function, P(X, Y), for all possible values of X and Y. Since X can take values 0 or 1, and Y can also take values 0 or 1, we need to find the joint probabilities for all combinations. Starting with (0, 0), we need to calculate P(X = 0, Y = 0). Some of these probabilities are already known.

For example, P(Y = 0 | X = 0) is already provided. We can express P(Y = 0 | X = 0) as the product of P(Y = 0 | X = 0) and P(X = 0). This is because the probability of the intersection of two events can be expressed as the conditional probability of one event given the other, multiplied by the probability of the second event.

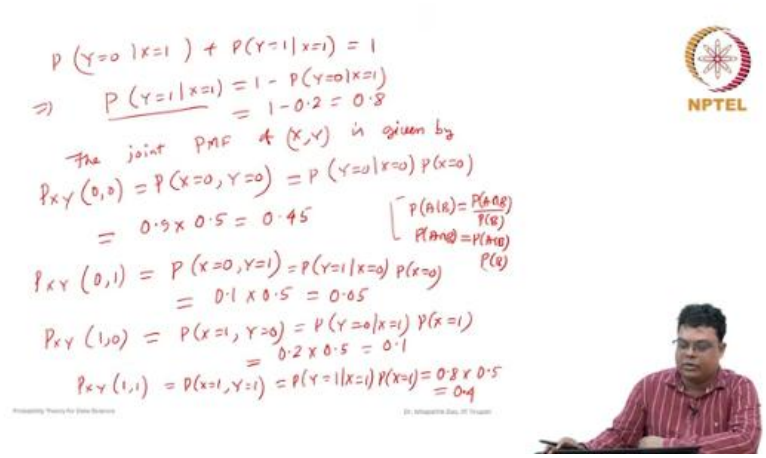
Now, we know these probabilities. For instance, P(Y = 0 | X = 0) = 0.9. We multiply this by P(X = 0), which is 0.5. This gives us 0.45. Similarly, we can compute other probabilities. For example, to find P(X = 0, Y = 1), we multiply P(Y = 1 | X = 0) by P(X = 0). P(Y = 1 | X = 0) = 0.1. This will result in 0.1 * 0.5 = 0.05.

Now, let's move on to find the other probability for when X = 1 and Y = 0. To calculate this, we take P(Y = 0 | X = 1) and multiply it by P(X = 1). P(Y = 0 | X = 1) = 0.2, so multiplying 0.2 by P(X = 1) = 0.5 gives 0.1.

Next, we calculate the probability when X = 1 and Y = 1. This is found by multiplying P(Y = 1 | X = 1) by P(X = 1). From earlier, we know P(Y = 1 | X = 1) = 0.8, and P(X = 1) = 0.5. This gives us 0.8 * 0.5 = 0.4.

Now, you can check that if you add all these values together, the sum should be 1. Adding 0.45 + 0.05 + 0.1 + 0.4 gives 1.

Therefore, we have found the joint probability mass function, which can be represented as shown. Suppose this is X and this is Y. We have values 0 and 1 for both X and Y. When X = 0 and Y = 0, P(X = 0, Y = 0) = 0.45. When X = 1 and Y = 0, P(X = 1, Y = 0) = 0.1.



For X = 0 and Y = 1, P(X = 0, Y = 1) = 0.05. Lastly, when X = 1 and Y = 1, P(X = 1, Y = 1) = 0.4. For better visibility, we represent the data this way.