PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

Week - 10

Lecture - 49

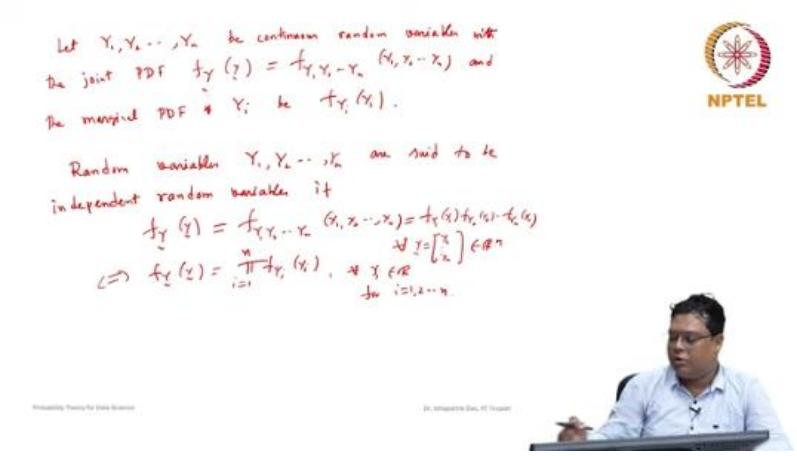
Moments of a Multivariate Random Variable

Let Y_1, Y_2, ..., Y_n be continuous random variables with the joint probability density function, denoted as f(y). This can be explicitly written as $f(y_1, y_2, ..., y_n)$. The marginal probability density function of Y_i is denoted as $f_{\gamma i}(y_i)$. If the random variables Y_1, Y_2, ..., Y_n are independent, their joint probability density function can be expressed as the product of their marginal probability density functions. This is similar to the discrete case, where the joint probability mass function is the product of the marginal probability mass functions.

In simplified notation, the joint probability density function can be written as: $f(y_1, y_2, ..., y_n) = f_{\gamma 1}(y_1) * f_{\gamma 2}(y_2) * ... * f_{\gamma n}(y_n)$.

It is important to note that this holds for all values of y_i, where i = 1, 2, ..., n. If this condition is satisfied, we say that Y_1, Y_2, ..., Y_n are independent continuous random variables.

In summary, for continuous random variables, the joint probability density function can be represented as the product of their marginal probability density functions when the random variables are independent, just as in the discrete case where the joint probability mass function is the product of the marginal probability mass functions.



Let Y_1, Y_2, ..., Y_n be continuous random variables, where each Y_i follows a normal distribution with mean μ_i and variance σ^2. For simplicity, let $\sigma^2 = 1$, though it can take any value. If the means μ_i are different for each i, then the random variables are not identically distributed. The probability density function (PDF) of each Y_i, a normally distributed random variable, is given by:

$f_{\gamma i}(y_i) = (1 / \sqrt{(2\pi)}) * e^{(- (y_i - \mu_i)^2 / 2\sigma^2)}$, where $\sigma^2 = 1$.

Here, Y_i can take any value between minus infinity and plus infinity for i = 1 to n. Now, if Y_1, Y_2, ..., Y_n are independent random variables, the joint probability density function (PDF) of Y_1, Y_2, ..., Y_n can be obtained by the product of their marginal probability density functions. By definition of independence, the joint PDF is:

$f_{\gamma}(y) = f_{\gamma 1}(y_1) * f_{\gamma 2}(y_2) * ... * f_{\gamma n}(y_n)$.

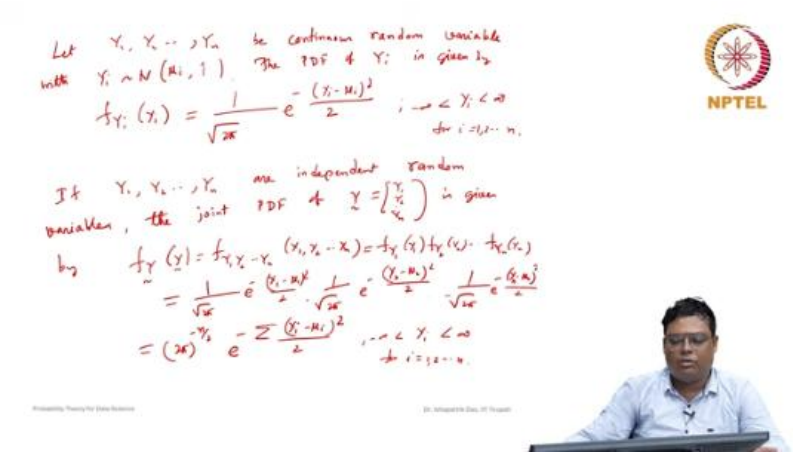
Substituting the expression for each marginal PDF:

$f_{\gamma 1}(y_1) = (1 / \sqrt{(2\pi)}) * e^{(-(y_1 - \mu_1)^2 / 2)}$,
$f_{\gamma 2}(y_2) = (1 / \sqrt{(2\pi)}) * e^{(-(y_2 - \mu_2)^2 / 2)}$,
...
$f_{\gamma n}(y_n) = (1 / \sqrt{(2\pi)}) * e^{(-(y_n - \mu_n)^2 / 2)}$.

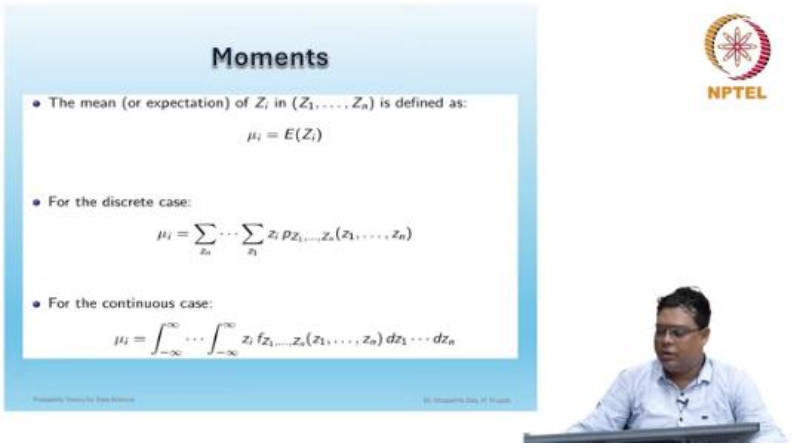
Multiplying these expressions together, we get:

$f_{\gamma}(y) = (1 / (2\pi)^{(n / 2)}) * e^{(- \Sigma (y_i - \mu_i)^2 / 2)}$,
where the summation is taken over i = 1 to n, and Y_i can take any value from minus infinity to plus infinity for each i.

This is the joint probability density function for independent normal random variables. If the random variables are not independent, this expression would not represent the correct joint probability density function. In future discussions, we will explore how to find the joint PDF when the random variables are not independent, and how the multivariate distribution changes in such cases. Now, we will discuss the moments of multivariate random variables.



We have already discussed moments for the bivariate case, and the concept extends straightforwardly to the multivariate case. Let us first define the mean of multivariate random variables. Let $X_1, X_2, ..., X_n$ be a set of multivariate discrete random variables with a joint probability mass function. If the variables are independent, we can find their joint probability mass function by using the product of their marginal distributions. However, if they are not independent, we need to know the joint probability mass function explicitly.



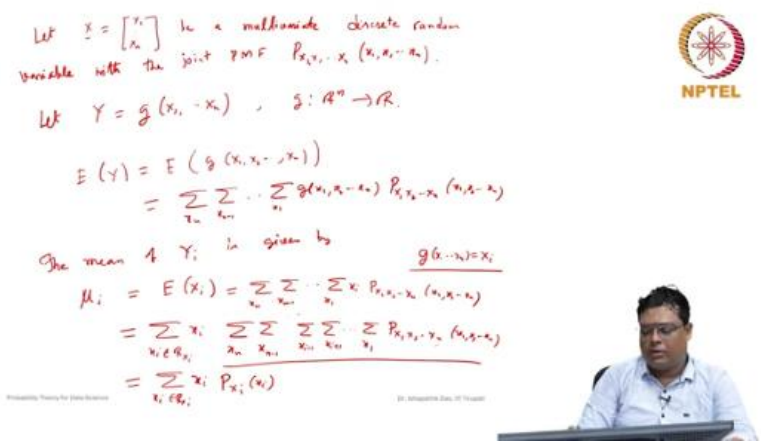
Let us assume the joint probability mass function is given by $P(x_1, x_2, ..., x_n)$. Now, consider a random variable Y, which is a function of the multivariate random variables $X_1, X_2, ..., X_n$. Specifically, $Y = g(X_1, X_2, ..., X_n)$, where g is a function that maps $\mathbb{R}^n$ to $\mathbb{R}$. This transformation will be explained in a later chapter, but for now, we focus on finding the expected value of Y. The expected value of Y, denoted as E[Y], is defined as the sum of all possible values that Y can take, weighted by their respective probabilities: $E[g(X_1, X_2, ..., X_n)] = \Sigma (g(x_1, x_2, ..., x_n) * P(x_1, x_2, ..., x_n))$.

Here, $P(x_1, x_2, ..., x_n)$ is the joint probability mass function, and $g(x_1, x_2, ..., x_n)$ represents the transformation applied to the variables. Now, let's focus on the mean of the random variable X_i. The expected value of X_i, denoted as μ_i, is given by: $\mu_i = E[X_i] = \Sigma (x_i * P(x_1, x_2, ..., x_n))$.

In this case, we treat $g(x_1, x_2, ..., x_n)$ as equal to x_i, and we sum over all possible values of $x_1, x_2, ..., x_n$, except for x_i, which we are focusing on. Since X_i is independent of the other random variables $X_1, ..., X_{i-1}, X_{i+1}, ..., X_n$, we can separate the sum for X_i from the others. Thus, the expected value of X_i simplifies to: $\mu_i = \Sigma (x_i * P(x_i))$, where $P(x_i)$ is the marginal probability mass function of X_i.

This is the standard definition of the mean of a random variable. We can use this general definition to calculate the mean in univariate cases, as well as to find other moments. Similarly, we can extend this approach to find the variance of X_i using the general definition of expected value and the corresponding transformations. Similarly, we can extend this approach to find the variance of X_i using the general definition of expected value and the corresponding transformations. The variance (σ^2) of a random variable X_i is

given by the expected value of $(X_i - \mu_i)^2$, where μ_i is the mean of X_i. To compute the variance, we can use the general formula for the expected value of a function.



Let's consider the function $g(x_1, x_2, ..., x_n) = (X_i - \mu_i)^2$. To find the variance, we will apply this function to the general formula for the expected value: $E[(X_i - \mu_i)^2] = \Sigma (g(x_1, x_2, ..., x_n) * P(x_1, x_2, ..., x_n))$.

Here, $g(x_1, x_2, ..., x_n) = (X_i - \mu_i)^2$, and $P(x_1, x_2, ..., x_n)$ is the joint probability mass function. We can separate out the i-th variable (X_i) from the sum, since it is independent of the other random variables.

Thus, the formula simplifies to: $E[(X_i - \mu_i)^2] = \Sigma (x_i - \mu_i)^2 * P_{xi}(x_i)$, where $P_{xi}(x_i)$ is the marginal probability mass function of X_i.

This is similar to the previous calculation for the mean. We can further simplify this expression using the formula for variance: $Var(X_i) = E[X_i^2] - (E[X_i])^2$.

The expected value of X_i^2 is denoted as μ_{2i}' and the expected value of X_i is μ_{1i}. So, the variance of X_i can be written as: $Var(X_i) = \mu_{2i}' - (\mu_{1i})^2$.

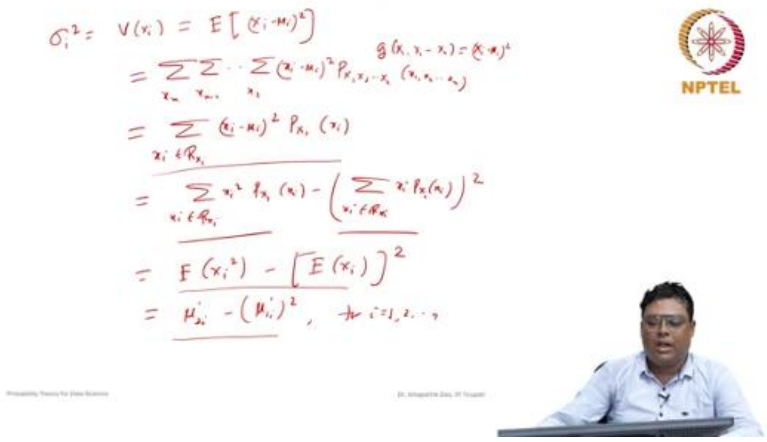
This is the general definition of the variance for the i-th random variable.

If we need to compute the variance for all random variables, we can apply the same logic to each variable individually.

Now, in the case of multivariate random variables, we are also interested in how different random variables relate to each other. For example, if we have two random variables X_i and X_j, we define their covariance to measure how they are related. The covariance between X_i and X_j is denoted as σ_{ij}, and it is calculated as: $Cov(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)]$.

This measures the relationship between the two variables, indicating whether they vary together (positive covariance) or in opposite directions (negative covariance).

For bivariate cases, we discussed covariance in detail, and we will continue to use the same concept when dealing with multivariate random variables. By definition, the covariance between two random variables X_i and X_j is given by: $Cov(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)]$, where μ_i and μ_j are the means of X_i and X_j, respectively.



This is the covariance for any i and j, where $i, j \in \{1, 2, ..., n\}$. Note that i and j can be equal, in which case the covariance simplifies to: $Cov(X_i, X_i) = E[(X_i - \mu_i)^2]$, which is simply the variance of X_i (σ_i^2).

So, when $i \neq j$, the result is the covariance between X_i and X_j. This can also be written as: $Cov(X_i, X_j) = E[X_i * X_j] - \mu_i * \mu_j$.

We also discussed earlier that covariance is not a unit-less quantity, which means that we cannot directly compare the covariance of different sets of random variables. To address this, we need a standardized measure that can be compared across different random variables. This is where the correlation coefficient comes in.

The correlation coefficient between two random variables X_i and X_j, denoted as ρ_{ij}, is defined as: $\rho_{ij} = Cov(X_i, X_j) / (\sigma_i * \sigma_j)$, where σ_i and σ_j are the standard deviations (the square root of the variance) of X_i and X_j, respectively.

The correlation coefficient is unit-less, making it comparable across different pairs of random variables. It ranges from -1 to 1, where: $\rho_{ij} = 1$ indicates a perfect positive correlation.
$\rho_{ij} = -1$ indicates a perfect negative correlation. $\rho_{ij} = 0$ indicates no correlation.

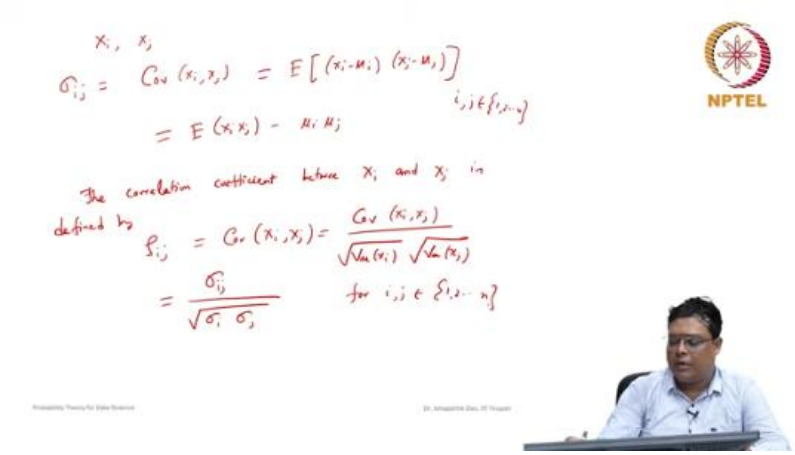
This provides a more standardized way to understand the relationship between two random variables.

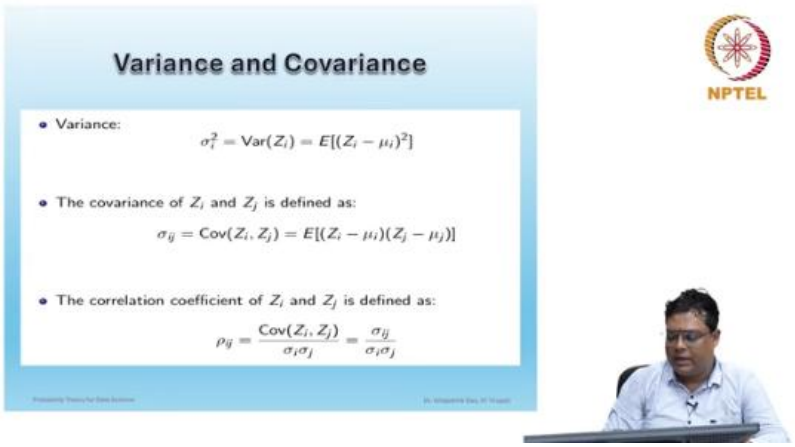
If i = j, then the correlation coefficient is 1, because the covariance between a random variable and itself is simply the variance, and when divided by the product of the standard deviations (which are the same), the result is 1. This is because: $Cov(X_i, X_i) = \sigma_i^2$, and $\rho_{ii} = Cov(X_i, X_i) / (\sigma_i * \sigma_i) = \sigma_i^2 / (\sigma_i * \sigma_i) = 1$.

So, for $i \neq j$, the correlation coefficient represents the relationship between two different random variables, X_i and X_j. In the slide, the notation was changed to Z_i and Z_j instead of X_i and X_j. The concept remains the same. The correlation coefficient between Z_i and Z_j is given by: $\rho_{ij} = Cov(Z_i, Z_j) / (\sigma_i * \sigma_j)$.

This is the same as the covariance and correlation coefficient between two random variables, just with different variable names.

For bivariate cases, we already discussed that we are considering two random variables at a time. In this case, we are looking at the correlation coefficient between two random variables, X_i and X_j. If the correlation coefficient, ρ_{ij}, is close to 1, it indicates that there is a strong positive linear relationship between the two variables.



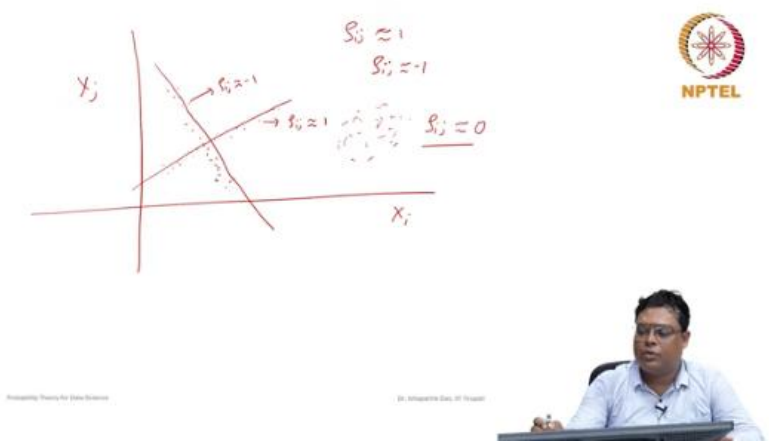


This means that as one variable increases, the other will also increase, and the data points will lie close to a straight line. If the correlation coefficient, ρ_{ij}, is close to -1, it indicates a strong negative linear relationship. In this case, as one variable increases, the other will decrease, and the data points will form a straight line with a negative slope.

The correlation coefficient is always between -1 and 1, and this property is shown by the Cauchy-Schwarz inequality. Although we won't go into the details here, it's important to note that the correlation coefficient plays a significant role in understanding the relationship between two random variables.

When the correlation coefficient is close to 1, the relationship between the variables is positive and linear, and when it is close to -1, the relationship is negative and linear. If the correlation coefficient is close to 0, this indicates that there is no linear relationship between the variables. In this case, the data may have a more complex, non-linear relationship, such as a circular pattern or fluctuations without a clear trend.

This concludes the concept of moments. We have discussed the first-order and second-order moments.



Since we are dealing with multivariate random variables, higher-order moments can be more complicated to discuss. Therefore, we focused on the first-order moment, which is the mean (μ_i), and the second-order moment, which is the variance (σ_i^2). Additionally, we explored how these moments relate to each other, specifically through covariance ($Cov(X_i, X_j)$) and the correlation coefficient (ρ_{ij}), which measure how random variables are correlated with each other.