PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

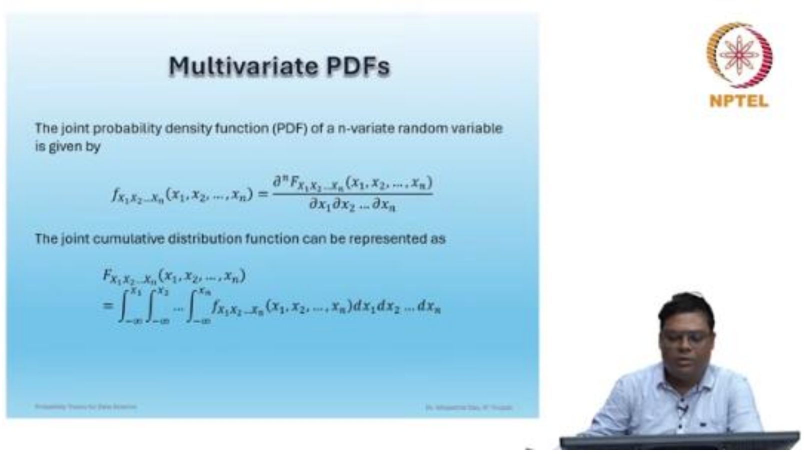
Indian Institute of Technology Tirupati

Week - 10

Lecture - 48

Multivariate Probability Density Function and Independence

We will discuss multivariate probability density functions. After completing the discussion on multivariate distribution functions, we also covered multivariate probability mass functions and their properties, including how to determine these functions. Similarly, we will now learn how to define a joint probability density function for multivariate random variables. To recap, when we discussed bivariate random variables, specifically bivariate continuous random variables, we considered two random variables, X_1 and X_2. In that case, we defined the cumulative distribution function as the probability that $X_1 \leq x_1$, and $X_2 \leq x_2$.

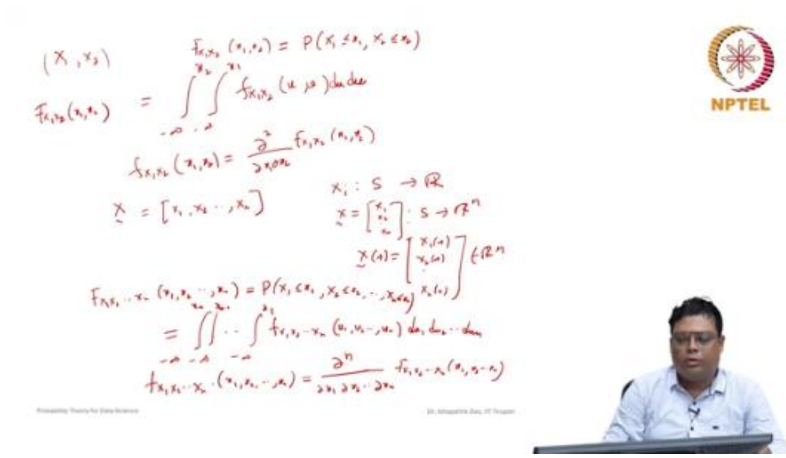


This is the cumulative distribution function for bivariate continuous random variables, and we assumed that it is differentiable with respect to both X_1 and X_2. Based on this assumption, there exists a function of X_1 and X_2 that represents the joint cumulative distribution function. This function, known as the joint probability density function, can also be obtained by taking the partial derivatives of the cumulative distribution function with respect to X_1 and X_2, provided that these derivatives exist for all values of X_1 and X_2. When extending this concept to multivariate random variables, we consider variables such as X_1, X_2, and so on, up to Xn. These are collectively represented as a random vector, often written as X, where each component, such as Xi, is a measurable function that maps the sample space to the real numbers.

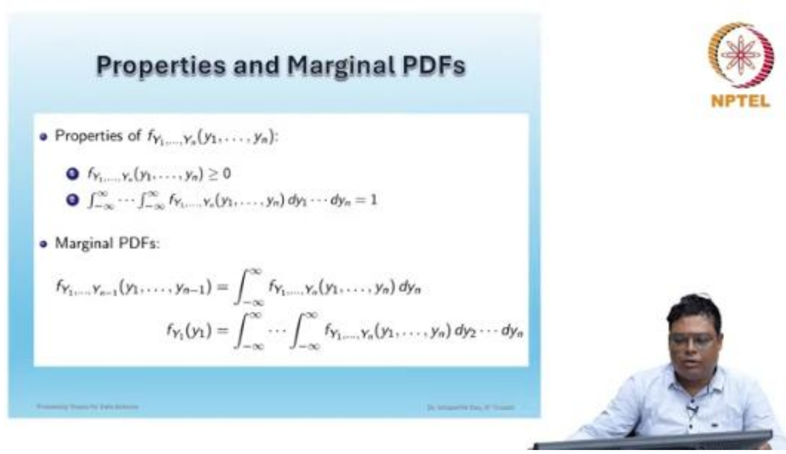
In this representation, X is a random vector that acts as a function mapping the sample space to an n-dimensional real number space. For example, the vector X can be thought of as containing the components X_1, X_2, and so on, up to Xn, as elements in this space. When all the individual random variables in this vector are continuous, the random vector is referred to as a multivariate continuous random variable. For these variables, we assume the existence of a joint density function. The distribution function for this random vector represents the probability that $X_1 \leq x_1$, $X_2 \leq x_2$, and so on, up to $Xn \leq xn$.

This is known as the multivariate cumulative distribution function. Similarly, when we extend the concept to multivariate random variables, considering up to n random variables such as X_1, X_2, and so on up to Xn, we assume that these variables are continuous random variables. For such cases, there exists a function, often referred to as the joint probability density function. This function, denoted as f, represents the joint density of the variables X_1, X_2, and Xn. The joint probability density function is defined over the range of all variables, from negative infinity to their respective values.

For example, for variables X_1, X_2, and Xn, the joint density can be expressed as an integral over these variables, where intermediate variables u_1, u_2, and so on are used within the integration limits. This function is known as the joint probability density function, and it is denoted by $f(X_1, X_2, ..., Xn)$. This function can also be obtained by taking the partial derivatives of the cumulative distribution function with respect to all variables X_1 through Xn, assuming these derivatives exist. Similar to the univariate and bivariate cases, the multivariate case also includes specific properties of the joint probability density function. These properties are important for understanding and working with multivariate continuous random variables.



So, this property is written here, We now consider the representation of the multivariate probability density function. The joint probability density function for an n-variable random variable is defined as a function that can be used to compute probabilities over the range of all variables. The corresponding joint cumulative distribution function has already been discussed. Next, we will explore the properties of the joint probability density function, along with the concept of the marginal probability density function.

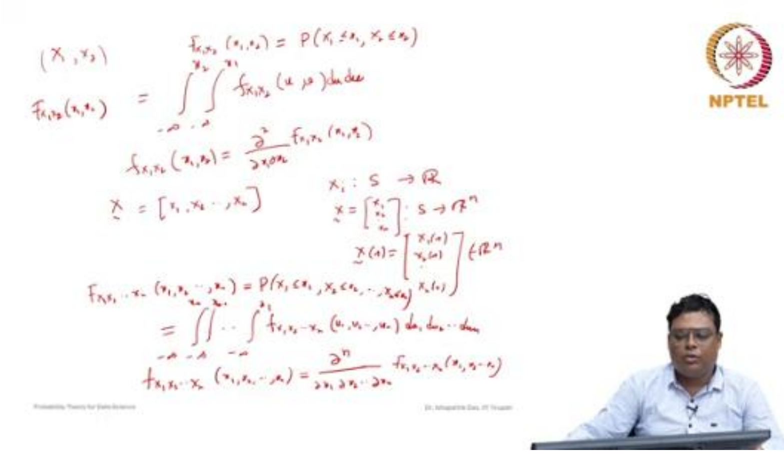


Specifically, we will discuss how to derive the marginal probability density function from the joint probability density function. Many of the properties we will discuss are extensions of those seen in univariate and bivariate cases. One important property of the cumulative distribution function is that it is non-decreasing. To explain this, let’s assume we fix some of the variables while varying one of them. For instance, if we have variables X_1, X_2, ..., Xn, and we fix X_2 through Xn while increasing X_1, the cumulative distribution function will not decrease as X_1 increases.

This property holds because, in the multivariate case, while we cannot fully order all variables in n-dimensional space, we can establish an order by fixing certain coordinates. As a result, the cumulative distribution function is non-decreasing in each coordinate. Furthermore, the partial derivatives of the cumulative distribution function with respect to any variable will always be greater than or equal to zero. This means that the partial derivative of the cumulative distribution function with respect to X_1, X_2, ..., Xn corresponds to the joint probability density function, and it is non-negative for all values of the variables.

Another key property, similar to the bivariate case, is that the integral of the joint probability density function over the entire space equals one.
This reflects the fact that the total probability across all possible outcomes is one. For example, consider two points, A and B, in n-dimensional space, where A has coordinates (a_1, a_2, ..., an) and B has coordinates (b_1, b_2, ..., bn). If each $a_i \leq b_i$ for all i from 1 to n, the probability that X_1 falls between a_1 and b_1, X_2 falls between a_2 and b_2, and so on, up to Xn falling between an and bn, can be found by integrating the joint probability density function over these ranges. This integration would be performed from a_1 to b_1 for X_1, a_2 to b_2 for X_2, and so on, up to an to bn for Xn. This process essentially calculates the probability within an n-dimensional region, which, in higher dimensions, is analogous to a cube.

For instance, in the two-dimensional case, the region would be a rectangle defined by the intervals [a_1, b_1] and [a_2, b_2]. In three dimensions, it would be a cube, and in n dimensions, it becomes a hypercube. These are some of the fundamental properties of the joint probability density function. Additionally, many other properties follow as natural extensions of the univariate and bivariate cases. These are some of the fundamental properties of the joint probability density function. Additionally, many other properties follow as natural extensions of the univariate and bivariate cases. Next, we will discuss how to find the marginal probability density function if the joint probability density function is known.



Suppose we have the joint probability density function for a multivariate random variable represented by the vector $X = (X_1, X_2, ..., Xn)$. This function provides the joint probabilities for all values of $X_1, X_2, ..., Xn$ in n-dimensional space, R^n. Now, if we want to determine the probability density function for one of the variables or any subset of the variables, we calculate the marginal probability density function. For a set of n random variables, there are $2^n - 1$ possible non-empty subsets. For instance, the subsets might include individual variables like X_1 or X_2, pairs like (X_1, X_2), or larger subsets.

The process of finding the probability density function for any subset of variables is referred to as finding the marginal density. To illustrate, let’s consider finding the marginal probability density function of X_1. From our knowledge of bivariate cases, this is done by integrating out the other variables. Specifically, we integrate the joint probability density function over all variables except X_1. Mathematically, this involves integrating the joint density over $X_2, X_3, ..., Xn$ across their entire range, from negative infinity to positive infinity.

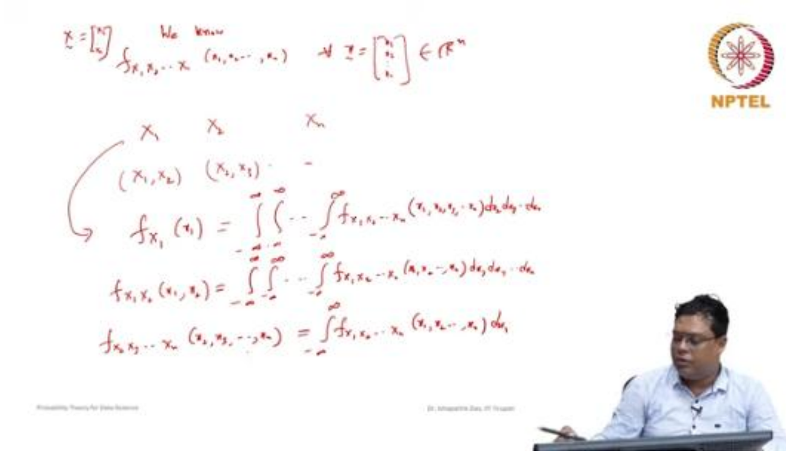
For example, the marginal density function of X_1 is calculated as: $\int ... \int f(X_1, X_2, ..., Xn)\, dX_2\, dX_3 ... dXn$.

Integrate the joint density over $X_2, X_3, ..., Xn$, while keeping X_1 fixed. The resulting function depends only on X_1, as the other variables have been integrated out. Similarly, if we want to find the marginal probability density function for a subset like (X_1, X_2), we integrate the joint density over the remaining variables, such as X_3, X_4, ..., Xn. In this case: $\int \ldots \int f(X_1, X_2, \ldots, X_n)\, dX_3\, dX_4 \ldots dX_n$.

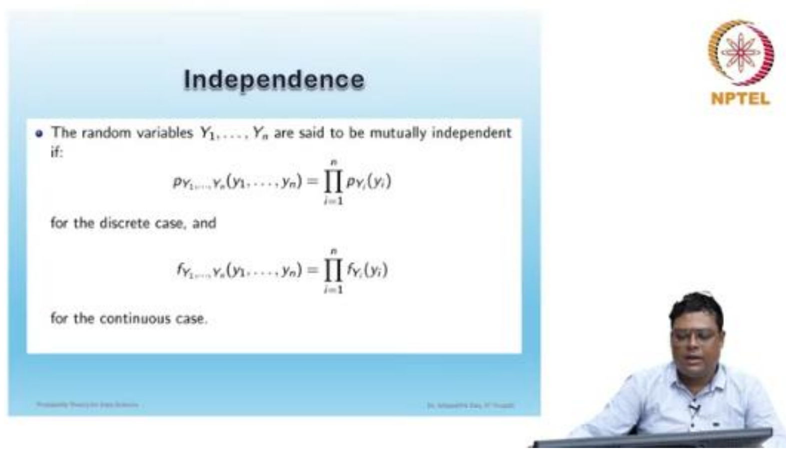
X_1 and X_2 are kept fixed, while the integration is performed over X_3 through Xn. The result is a function of X_1 and X_2, representing their marginal density. This process can be generalized for any subset of the random variables. For example, if we want to find the marginal density function for variables X_2 and X_3, we integrate the joint density over all other variables, such as X_1, X_4, ..., Xn. The integration effectively removes the influence of the unwanted variables, leaving the marginal density function for the variables of interest.

In summary, to find the marginal probability density function for any subset of variables:

1. Identify the variables of interest that you wish to keep.
2. Integrate the joint probability density function over all other variables that are not part of the subset.
3. The result is the marginal density function for the selected subset. This method provides a systematic way to extract the probability density for any subset of variables from the joint density. Hopefully, this explanation clarifies the concept of finding marginal probability density functions when the joint probability density function is known. The discussion here explains how to find the marginal probability density function from the joint probability density function. In the example, the variables Y_1, Y_2, ..., Y_{n-1} are retained, while Yn is eliminated.

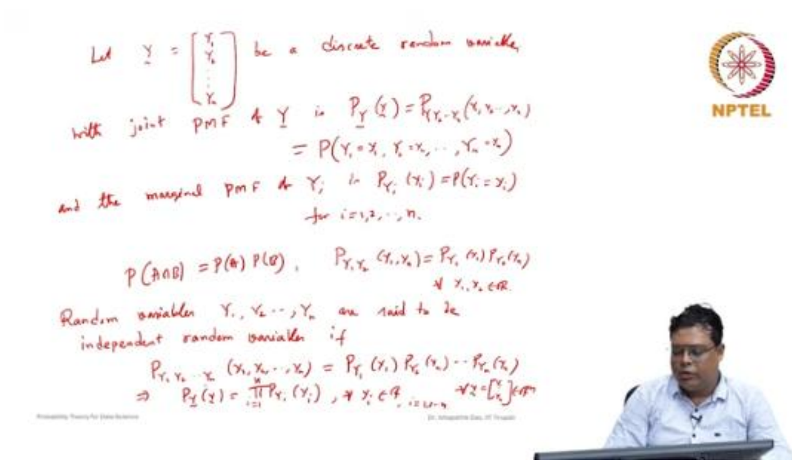


To achieve this, the joint probability density function is integrated with respect to Yn. Similarly, if we wish to keep only Y_1 and eliminate all other variables, we integrate the joint probability density function with respect to Y_2, Y_3, ..., Yn. This gives the marginal probability density function for Y_1. This process shows how the marginal probability density function can be obtained by integrating the joint probability density function over the unwanted variables. Now, we will discuss how to determine whether a multivariate random variable consists of independent random variables.



This is an important concept, especially for topics like statistical inference, where we frequently deal with independent and identically distributed random variables in the context of random samples. Let Y_1, Y_2, ..., Y_n be random variables. We denote this as a vector $Y = (Y_1, Y_2, ..., Y_n)$. Suppose these are discrete random variables with a joint probability mass function (PMF) represented as $P(Y = y)$. Explicitly, this is $P(Y_1 = y_1, Y_2 = y_2, ..., Y_n = y_n)$, which gives the probability that Y_1 takes the value y_1, Y_2 takes the value y_2, and so on up to Y_n taking the value y_n. The marginal probability mass function of Y_i is denoted as $P(Y_i = y_i)$, which is the probability that the random variable Y_i takes the value y_i for each i from 1 to n.

Random variables $Y_1, Y_2, ..., Y_n$ are said to be independent if their joint probability mass function can be expressed as the product of their marginal probability mass functions. Specifically, $P(Y_1 = y_1, Y_2 = y_2, ..., Y_n = y_n) = P(Y_1 = y_1) \times P(Y_2 = y_2) \times ... \times P(Y_n = y_n)$. This must hold true for all values of $y_1, y_2, ..., y_n$ that belong to their respective domains (usually the set of real numbers). In a simplified notation, the independence condition can be written as: $P(Y = y) = \prod P(Y_i = y_i)$, where the product is taken over i from 1 to n, and y represents the vector $(y_1, y_2, ..., y_n)$. In summary, $Y_1, Y_2, ..., Y_n$ are independent random variables if their joint probability mass function equals the product of their marginal probability mass functions for all values in their domain.

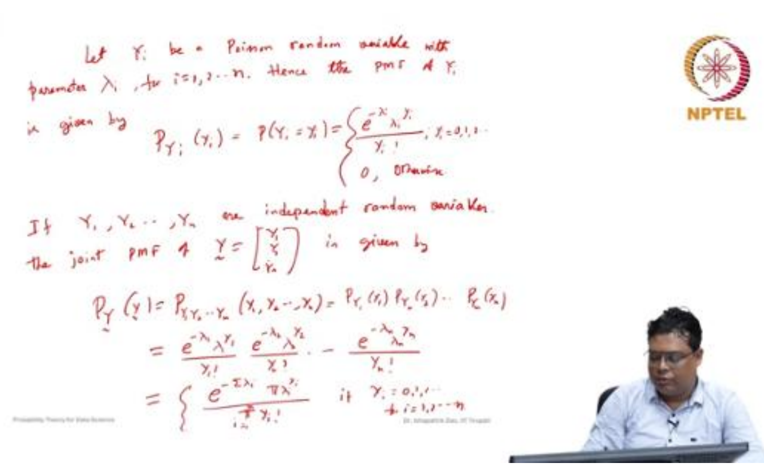


Now, let's consider an example to find the joint probability mass function. If the random variables are not independent, we may not know the joint probability mass function unless it is provided. Suppose we have a Poisson random variable. Let Y_i be a Poisson random variable with parameter λ_i, for i = 1 to n. Note that λ_i are different, so these random variables are not identically distributed, as identically distributed random variables would have the same distribution.

For example, if all Y_i's had the same Poisson distribution with parameter λ, then they would be identically distributed. But here, we have different λ_i values. If Y_i is a Poisson random variable with parameter λ_i, its probability mass function (PMF) is given by: $P(Y_i = y_i) = (e^{-\lambda_i} * \lambda_i^{y_i}) / (y_i!)$. This is valid for integer values of y_i (i.e., $y_i = 0, 1, 2, 3, \ldots$). Now, if $Y_1, Y_2, \ldots, Y_n$ are independent random variables, we can find their joint probability mass function.

Since they are independent, the joint PMF of $Y_1, Y_2, \ldots, Y_n$ is the product of their individual PMFs: $P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n) = P(Y_1 = y_1) * P(Y_2 = y_2) * \ldots * P(Y_n = y_n)$. By applying the Poisson PMF for each Y_i, the joint probability mass function becomes: $P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n) = (e^{-\lambda_1} * \lambda_1^{y_1} / y_1!) * (e^{-\lambda_2} * \lambda_2^{y_2} / y_2!) * \ldots * (e^{-\lambda_n} * \lambda_n^{y_n} / y_n!)$. This simplifies to:

$P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n) = e^{-\Sigma\lambda_i} * \Pi(\lambda_i^{y_i} / y_i!)$. This is the joint probability mass function for independent Poisson random variables. If Y_i takes integer values, the probability mass function is nonzero for integer values of $y_1, y_2, \ldots, y_n$, and zero otherwise. Thus, the joint probability mass function can be written as: $P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n) = e^{-\Sigma\lambda_i} * \Pi(\lambda_i^{y_i} / y_i!)$ for $y_i \in \{0, 1, 2, \ldots\}$.



So, $P_\gamma(y)$ is equal to $e^{-\Sigma\lambda_i} * \Pi(\lambda_i^{y_i} / y_i!)$, where $y_i \in \{0, 1, 2, \ldots\}$ for $i = 1, 2, \ldots, n$. This is valid for integer values of y_i. Otherwise, it is 0. This represents the joint probability mass function of independent Poisson random variables. To summarize, when we know the random variables are independent, we can find their joint probability mass function if we know their marginal probability mass functions. This joint probability mass function is applicable to independent Poisson random variables. Next, we will discuss the case when the random variables are continuous. In this case, we define independence with respect to their probability density functions.