PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

Week - 09

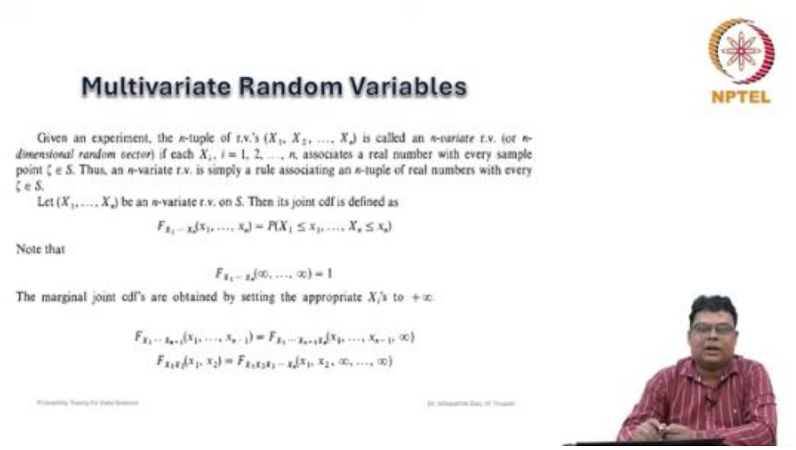
Lecture - 47

Multivariate Random Variables

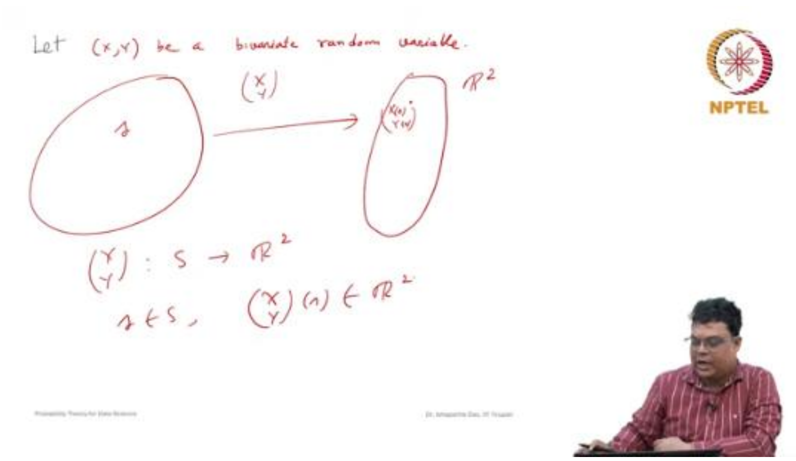
Next, we will discuss how to extend the concept of bivariate random variables to multivariate random variables. If you have clearly understood how we extended the concept from univariate to bivariate random variables, then it is straightforward to extend it further. Instead of two variables, we will discuss cases with n random variables. The concept of n random variables is important because many random phenomena cannot be explained using only one or two random variables. Multiple random variables are often needed. In this part, we will explore multivariate random variables. In many situations, a single random variable cannot properly explain a particular phenomenon, so we need to consider more than one random variable.

For example, when discussing air pollution, a single random variable like particulate matter (PM 2.5 or PM 10) is insufficient. To fully explain the situation, we need to account for other factors, such as the amount of carbon monoxide, carbon dioxide, nitric oxide, sulfur dioxide (SO_2), and other pollutants. This is essential for understanding the greenhouse effect, particulate contamination, and similar phenomena. Multivariate responses are common in many fields, including social and natural sciences, clinical trials, special experiments, ecology, econometrics, and epidemiology. For these reasons, we need to extend the concept of bivariate random variables to handle any number of random variables.

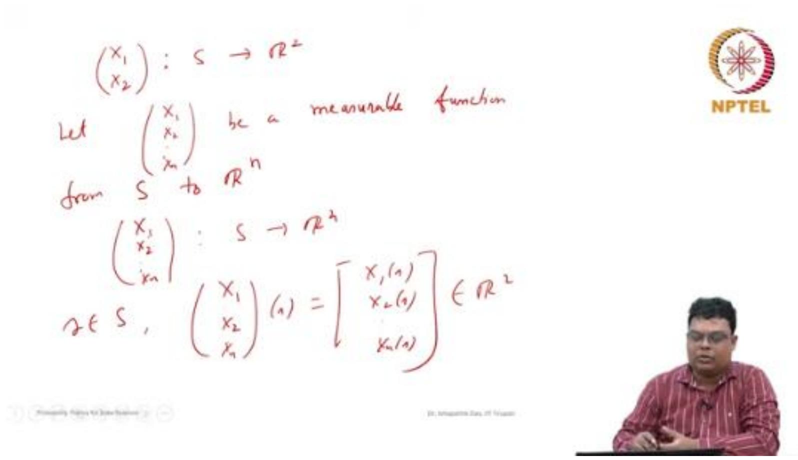
The extension to multivariate random variables is very straightforward. If you have understood bivariate random variables, we will now discuss how to extend the concept further. How to extend it from bivariate random variables to multivariate random variable cases. Let's first discuss the bivariate random variable case. Let X and Y be a bivariate random variable. How is it defined?



For a random experiment, we have a sample space S, and a bivariate random variable is defined by a measurable function X, Y. For each element s in S, we find a point in R^2, specifically (X(s), Y(s)), which is a vector in R^2. Therefore, X and Y are bivariate random variables, and they are functions from S to R^2. Each element $s \in S$, and the function is measurable, mapping elements to R^2. Let us consider, instead of X and Y, the random variables $X_1, X_2, ..., X_n$ to extend the concept of bivariate random variables to n variables.



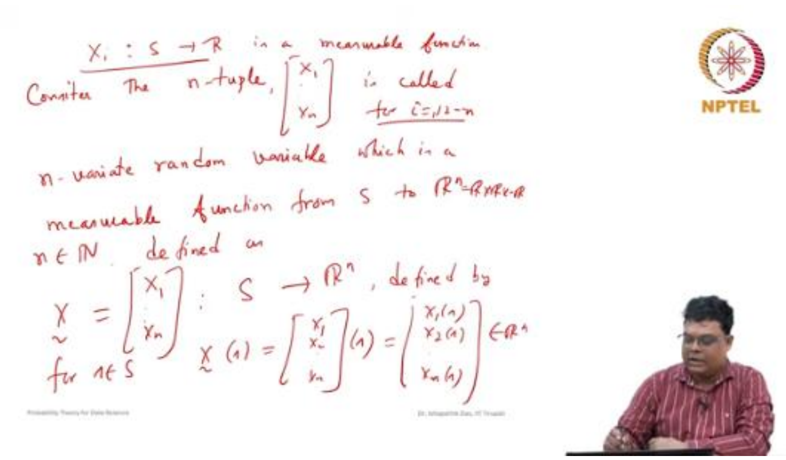
This is a measurable function from the sample space S to R^n. Let $X_1, X_2, ..., X_n$ be a measurable function from S to R^n. For each element $s \in S$, $(X_1(s), X_2(s), ..., X_n(s))$ is defined, where each $X_1, X_2, ..., X_n$ are random variables. Therefore, we can extend the bivariate random variable concept to define n-tuple random variables. $X_1, X_2, ..., X_n$ is called an n-variate random variable or n-dimensional random variable if each X_i associates a real number with every sample space element $s \in S$. Therefore, we can extend the bivariate random variable concept to define n-tuple random variables. $X_1, X_2, ..., X_n$ is called an n-variate random variable or n-dimensional random variable if each X_i associates a real number with every sample space element $s \in S$.



Thus, an n-variate random variable is simply a rule that associates an n-tuple of real numbers with every element $s \in S$ in the sample space S. Each of the random variables X_1, X_2, ..., Xn is a measurable function from S to real numbers. For simplicity, we can consider the n-tuple as a column vector. The n-tuple, consisting of X_1, X_2, ..., Xn, is called an n-variate random variable, which is a measurable function from the sample space S to R^n. This is an extension of the concept of bivariate random variables (R^2), where R represents real numbers and n is a natural number.

If n = 1, it represents a univariate random variable; if n = 2, it represents a bivariate random variable; if n = 3, it is a trivariate random variable, and so on. For any number n, the n-variate random variable is defined in the same manner. We can describe the n-variate random variable as a vector, which is a measurable function from the sample space S to R^n. This function assigns values to each random variable in the vector, such that each random variable X_i is associated with a real number. Therefore, X_1, X_2, ..., Xn are random variables, and each of them corresponds to a real number. The entire vector will be an element of R^n, making it a measurable function from S to R^n.

The entire vector will be an element of R^n, making it a measurable function from S to R^n. Now, analogous to how we define the bivariate random variable and bivariate distribution function, we will define the multivariate distribution function. Specifically, we are discussing the n-variate cumulative distribution function (CDF). Let X_1, X_2, ..., Xn be an n-variate random variable. For simplicity, let us denote this as X_1, X_2, ..., Xn.

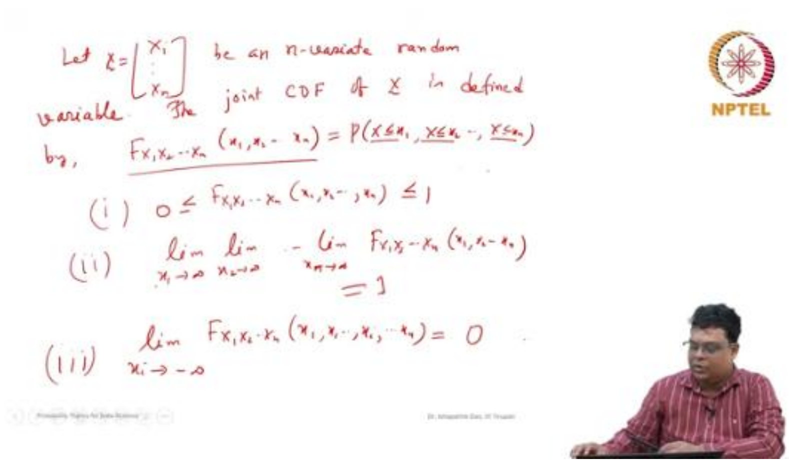


The joint CDF, or cumulative distribution function, for this n-variate random variable is defined similarly to how we defined the bivariate case. Recall that in the bivariate case, the CDF of X and Y, denoted as $F(x, y)$, is the probability that $X \leq x$ and $Y \leq y$. If we extend this concept to n variables, we replace Y with X_2 and X with X_1, so it becomes the probability that $X_1 \leq x_1$, and $X_2 \leq x_2$, and so on. Thus, for n random variables, the joint CDF is defined as the probability that $X_1 \leq x_1$, $X_2 \leq x_2$, and so on, up to $Xn \leq xn$. This can be written as:

$$F(X_1, X_2, ..., Xn) = P(X_1 \leq x_1, X_2 \leq x_2, ..., Xn \leq xn)$$

The CDF is a joint cumulative distribution function that represents the probability of all these events occurring together. It is similar to the bivariate case, but with more variables. Similar to the bivariate case, some important properties of the multivariate CDF are the same. However, we won't go into all the details here as they can get a bit complicated. We have already discussed most of the important properties in the bivariate case. For example, the CDF is right-continuous because it is a probability distribution function. Additionally, the value of the CDF will always be between 0 and 1. When any of the variables tends to infinity, the CDF will tend to 1 because it represents the probability of the whole sample space. To explain this more clearly, if all variables (X_1, X_2, ..., Xn) tend to infinity, the probability will eventually encompass the entire sample space, and thus the CDF will equal 1.

Similarly, if any of the variables (say Xi) tends to negative infinity, the CDF will approach 0 because this corresponds to a null event (the event that never happens). These properties are similar to those in the bivariate case. We will only focus on the properties that are most useful for our future discussions and applications. Now, let's discuss how to obtain the marginal probability distribution functions. For example, in the bivariate case, suppose X_1 and X_2 are bivariate random variables.

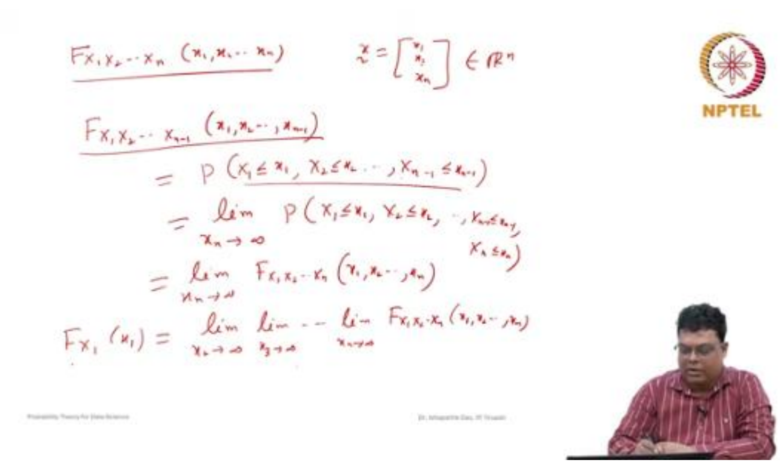


In this case, marginal means a subset of the random variables, such as X_1 or X_2. These subsets represent the variables we are interested in. If we consider three variables, the subsets of these variables could be $\{X_1, X_2\}$, $\{X_1, X_3\}$, or $\{X_2, X_3\}$. In general, for any random variables X_1, X_2, X_3, etc., the subsets of these variables form possible combinations. To find the joint distribution function for these subsets, we refer to it as the marginal distribution.

For instance, suppose we have the joint probability distribution function for $X_1, X_2, ..., Xn$. This joint probability mass function is known to us for any values of $X_1, X_2, ..., Xn$ in the space R^n. Now, if we want to find the marginal probability mass function of only $X_1, X_2, ..., Xn_{-1}$ (a subset), we are interested in the joint probability distribution of just $X_1, X_2, ..., Xn_{-1}$, without considering the nth random variable. This can be represented as the probability that $X_1 \leq x_1$, $X_2 \leq x_2$, and so on, up to $Xn_{-1} \leq xn_{-1}$. This probability can be thought of as the intersection of the event with the sample space S.

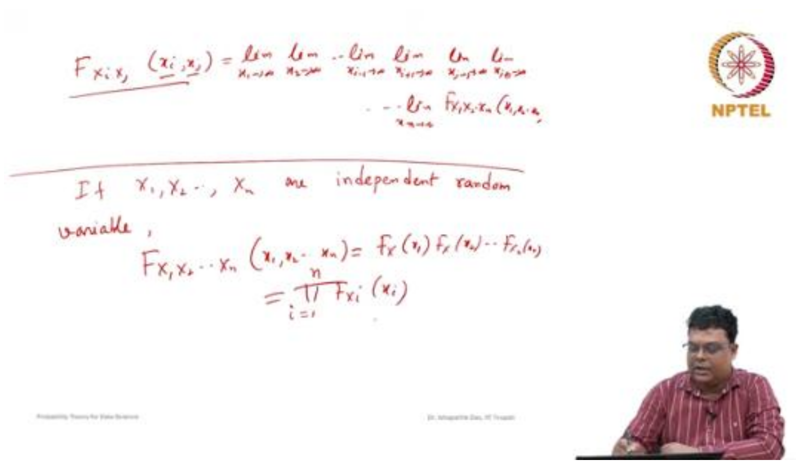
Therefore, this probability is the same as taking the limit of the probability as the nth variable tends to infinity. In other words, the probability that $X_1 \le x_1$, $X_2 \le x_2$, ..., $X_{n-1} \le x_{n-1}$, and $X_n \le x_n$ becomes the same as the limit of the joint cumulative distribution function as X_n tends to infinity. As X_n goes to infinity, this becomes an almost certain event, covering the entire sample space. Thus, the marginal distribution of $X_1, X_2, ..., X_{n-1}$ can be found by taking the limit of the joint cumulative distribution function of $X_1, X_2, ..., X_n$ as X_n tends to infinity.

Similarly, for any subset of variables, you can find the marginal distribution by taking the limit of the other variables as they tend to infinity.
For example, if you have a subset of variables, say $\{X\alpha_1, X\alpha_2, ..., X\alpha_u\}$, you would take the limit of the remaining variables as they tend to infinity to obtain the marginal distribution for this subset. In particular cases, if you want to find the marginal distribution for just one variable, say X_1, you would take the limit of all other variables ($X_2, X_3, ..., X_n$) tending to infinity. This would allow you to compute the marginal probability distribution function of X_1. To find the marginal probability distribution function of X_1, you can follow a similar approach for other cases. For example, suppose you want to find the marginal distribution function for two random variables, X_i and X_j, where $i \neq j$.



In this case, you would take the limit of all other variables except X_i and X_j as they tend to infinity. Specifically, you would take the following limits: X_1 tends to infinity, X_2 tends to infinity, and so on for all other variables except for X_i and X_j. This would allow you to compute the marginal cumulative distribution function for X_i and X_j. The other variables are taken to infinity to remove their influence on the joint distribution. It's important to note that infinity is not a real number but a notation indicating that the variables are tending to extreme values.

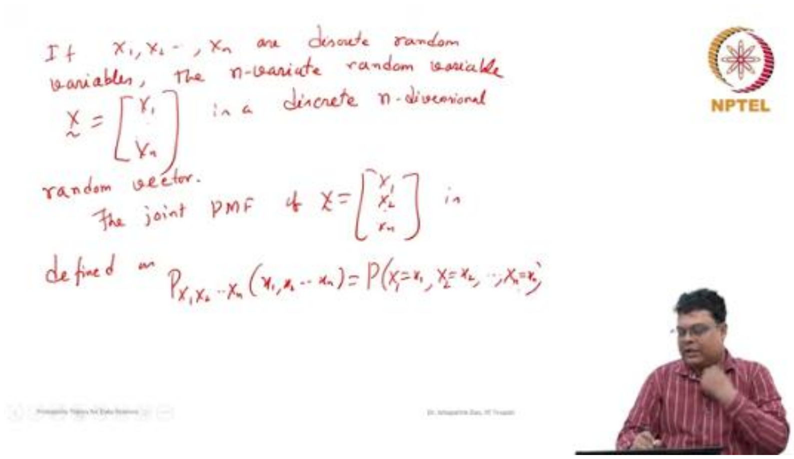
Thus, the marginal cumulative distribution function is obtained by taking these limits, rather than directly replacing values with infinity. This concept is similar to the previous extension, where we take the limits of the variables we are not interested in, leaving only the subset of random variables that we want to analyze. Now, let's discuss how to determine if $X_1, X_2, \ldots, X_n$ are independent random variables. Similar to the bivariate case, if the joint cumulative distribution function of $X_1, X_2, \ldots, X_n$ can be expressed as the product of their marginal distribution functions, then we can conclude that the random variables are independent. For example, if the joint cumulative distribution function of $X_1, X_2, \ldots, X_n$ can be written as the product of their individual marginal distribution functions, then $X_1, X_2, \ldots, X_n$ are independent random variables with respect to the distribution function.



Now, let's discuss the case where all the random variables are discrete random variables, such as $X_1, X_2, \ldots, X_n$. In the bivariate case, we discussed the joint probability mass function. Similarly, we will now discuss the joint probability mass function for discrete random variables $X_1, X_2, \ldots, X_n$. When $X_1, X_2, \ldots, X_n$ are discrete random variables, we refer to them as an n-variate random variable or an n-dimensional random vector. In this case, the probability mass function applies to the discrete random variables $X_1, X_2, \ldots, X_n$.

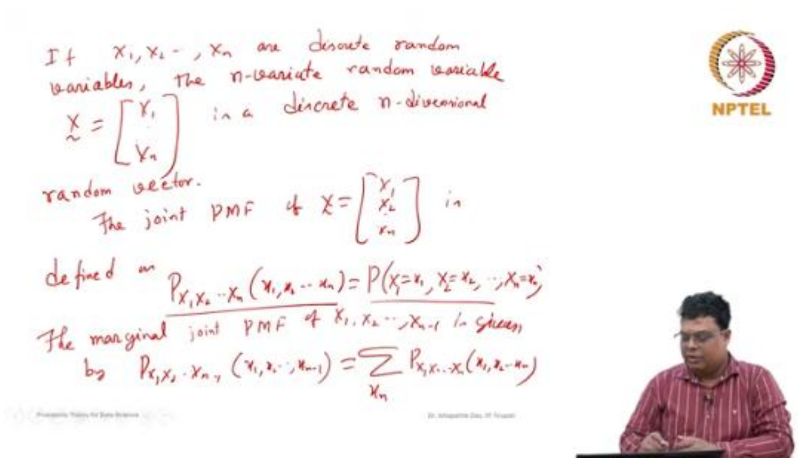
You may remember that for two random variables, X and Y, we define the joint probability mass function as the probability that $X = x_1$ and $Y = y_2$.

In a similar manner, we extend this concept to n random variables. The joint probability mass function of X_1, X_2, ..., Xn is defined as the probability that $X_1 = x_1$, $X_2 = x_2$, and so on up to Xn = xn. This is the joint probability mass function for X_1, X_2, ..., Xn. Now, if we know the joint probability mass function, we can proceed to find the marginal probability mass function, just as we did in the bivariate case. This is the joint probability mass function for X_1, X_2, ..., Xn. Now, if we know the joint probability mass function, we can proceed to find the marginal probability mass function, just as we did in the bivariate case. Now, if you want to find the marginal probability of X_1, you need to sum over the other variables.

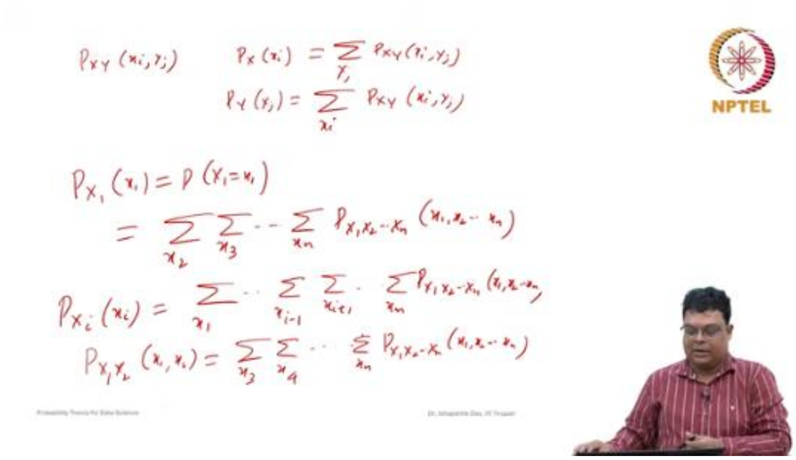


By fixing X_1, you can calculate the probability that the other variables take their respective values. This summing process is how we find the marginal probability mass function. Similarly, if you want to find the marginal probability mass function for Y_j, you will sum over the other variables, fixing Y_j. If you want to find the marginal probability mass function for a subset of the random variables, such as X_1, X_2, ..., Xn_{-1}, you will sum over the other variables. In this case, the marginal probability mass function of the subset will be derived from the joint probability mass function.

For example, if the joint probability mass function P(X_1, X_2, ..., Xn_{-1}) is known, and you want to find the marginal probability mass function for this subset, you sum over the appropriate variables. Similarly, if you are only interested in X_1, you would sum over all the other variables (X_2, X_3, ..., X_n) to find the marginal probability mass function for X_1. So, if you want to find the marginal probability mass function for any specific variable, such as X_1, you need to sum over all the other variables. For example, to find the marginal probability of $X_1 = x_1$, you would sum over the other variables (X_2, X_3, ..., X_n). This means summing the joint probability mass function $P(X_1, X_2, ..., X_n)$ over the values of all other variables.



For any i-th variable, if you want to find its marginal probability mass function, you would sum over all the variables except the i-th one. For instance, if you want to find the marginal probability mass function for X_1, you would sum over X_2, X_3, ..., X_n in the joint probability mass function $P(X_1, X_2, ..., X_n)$. To generalize, if you are looking for the marginal probability mass function for a subset, say $\{X_1, X_2\}$, you would sum over the remaining variables $\{X_3, X_4, ..., X_n\}$ in the joint probability mass function $P(X_1, X_2, ..., X_n)$. This is how you can find the marginal distribution from the joint probability mass function. The properties of the joint probability mass function are similar to those in the bivariate case because it is a probability distribution.

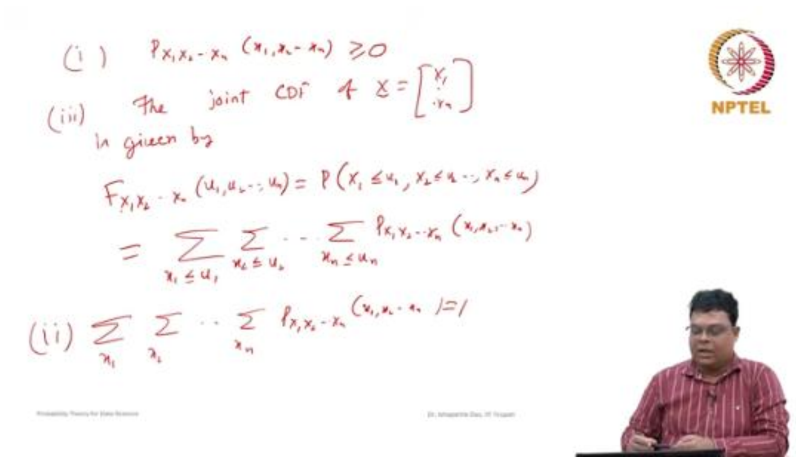


Therefore, the joint probability mass function for the random variables X_1, X_2, ..., Xn will always be greater than or equal to 0. Next, suppose you want to find the cumulative distribution function (CDF) from this joint probability mass function. The joint CDF of the random vector X_1, X_2, ..., Xn can be found by summing the probability mass function up to the values of X_1, X_2, ..., Xn. This is represented as the probability that $X_1 \leq x_1$, $X_2 \leq x_2$, and Xn ≤ xn. The sum will include all values less than or equal to the specified values of X_1, X_2, ..., Xn.

As with the bivariate case, there are some important properties: The sum of all values of the joint probability mass function will always equal 1. This is because it represents a probability distribution. The joint probability mass function is non-negative, meaning it will always be greater than or equal to 0. The third property involves considering all possible values of X_1, X_2, ..., Xn.

If you denote the values of X_1, X_2, ..., Xn as u_1, u_2, ..., un, the probability mass function for the values less than or equal to these values will be non-zero. This helps in understanding how the joint probability mass function works across different possible values. This concept is essentially an extension of the bivariate case to the multivariate case. Once you understand the bivariate case, the transition to the multivariate case becomes straightforward.

Next, we will discuss the joint probability density function (PDF) when the random variables X_1, X_2, ..., Xn are continuous. We'll explore its properties and how we can find the marginal distributions, as well as discuss the concept of independence for continuous random variables. With respect to discrete random variables, if X_1, X_2, ..., Xn are independent, then the joint probability mass function of $X_1, X_2, ..., X_n$ is the product of the individual probabilities. Specifically, this is the probability that $X_1 = x_1$, $X_2 = x_2$, and $X_n = x_n$. Since the variables are independent, the joint probability mass function is the product of the individual probabilities: $P(X_1 = x_1) * P(X_2 = x_2) * ... * P(X_n = x_n)$.



In notation, this can be written as $P(X_1 = x_1) * P(X_2 = x_2) * ... * P(X_n = x_n)$. Note that X_1, X_2, and X_n are random variables, which are measurable functions from the sample space (S) to the set of real numbers. On the other hand, x_1, x_2, and x_n are specific real numbers. This product can be simplified in a short form as the product of the individual probabilities, written as the product from i = 1 to n of $P(X_i = x_i)$. Next, we will discuss continuous random variables, focusing on multivariate random variables.

We will explore how joint probability density functions are defined for these variables. The concept is very similar to the bivariate case, and the joint probability density function for continuous variables will be an extension of that. We will also discuss the properties of these joint probability density functions, and how independence is defined for continuous random variables. It's important to note that this holds true for all values of $X_1, X_2, ..., X_n$, not just particular values.

If $X_1, X_2, \ldots X_n$ are independent discrete random variables, then

$$P_{X_1, X_2 \ldots X_n}(x_1, x_2 \ldots x_n) = P(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n)$$

$$= P(X_1 = x_1)\, P(X_2 = x_2) \ldots P(X_n = x_n)$$

$$= P_{X_1}(x_1)\, P_{X_2}(x_2) \ldots P_{X_n}(x_n)$$

$$= \prod_{i=1}^{n} P_{X_i}(x_i)$$