PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

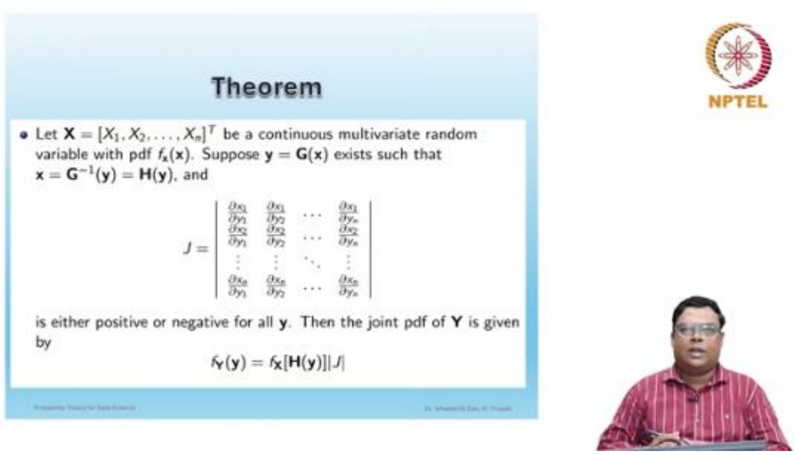
Week - 11

Lecture - 55

Transformation of Multivariate Random Variables

Now we will discuss the case of transforming a random variable into a multivariate random variable. We have completed some examples for univariate random variables. Now, we will discuss multivariate random variables and how similar concepts may be applied.

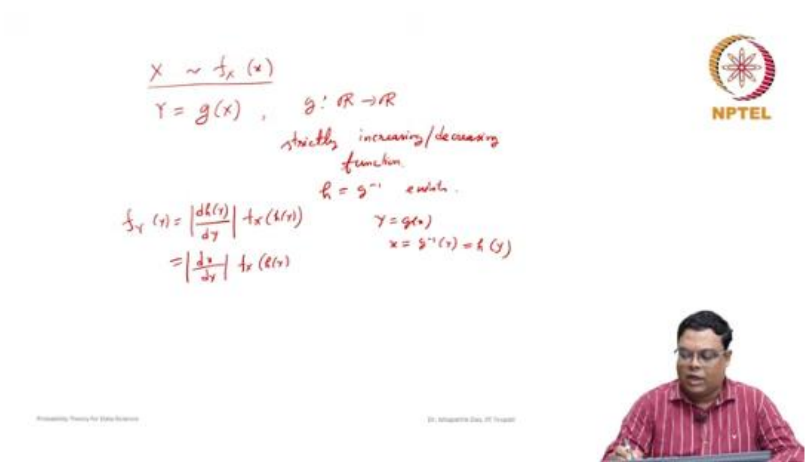
Let's revisit the univariate case. Whenever we have a random variable X with a probability density function f(X), which is a continuous random variable, and we want to find the transformation of the random variable Y = g(X), where g is a bijective function from R to R, and it is strictly increasing or strictly decreasing, we assume that g has an inverse.



If we consider g: R → R and it is strictly monotone (increasing or decreasing), then the inverse function h = g^(-1) exists. Under these conditions, if X is a continuous random variable with probability density function f(X), we can find the probability density function of Y, denoted as f_Y(y). This can be calculated using the formula: f_Y(y) = |d/dy h(y)| f_X(h(y)). Here, h(y) is the inverse function of g, where y = g(x) and x = g^(-1)(y).

This is why we sometimes write dx/dy f_X(h(y)).

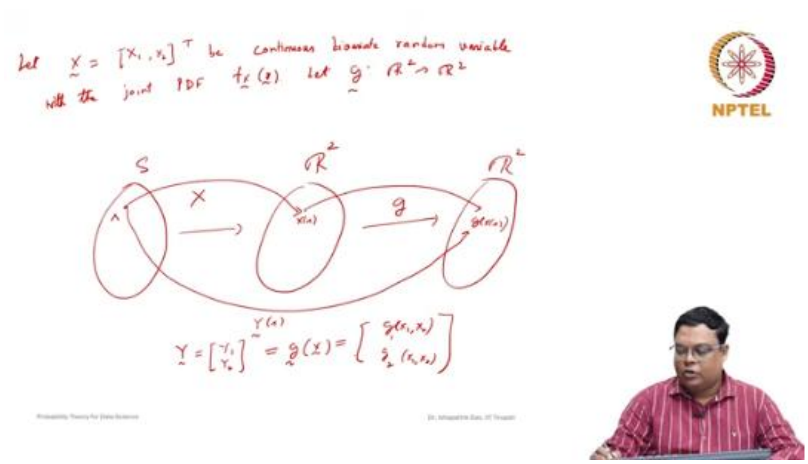
Now, the question arises: is this applicable to cases where we consider multivariate random variables? Let's simplify the case and consider a bivariate random variable. Let X1 and X2 be a bivariate continuous random variable, and we denote it as X. Let X1 and X2 be a bivariate continuous random variable, and we denote it as X. Let X be a continuous bivariate random variable with the joint probability density function f(X).



This function is known. In general, we write f(X), and in specific examples, we know the form of the density function. Let g: $R^2 \rightarrow R^2$. Since X is a bivariate random variable, it is a measurable function from the sample space S to R^2. Earlier, we discussed a composite function from S to R, where the function was measurable from R to R.

Now, since it is a bivariate random variable, we need to consider the case where X is from S to R^2. Some values will be taken here, and then the point under g will transform into g(Xs). This is a composite function, which we refer to as y(s). This transformation y will be a vector because Y is a vector, as it is a bivariate random variable. Therefore, g is a multivariate (specifically, bivariate) function.

We can write it as g tilde. This represents the function $g(X_1, X_2)$, where $g_1(X_1, X_2)$ and $g_2(X_1, X_2)$ are the components of the vector. Both g_1 and g_2 are functions from R^2 to R, making g a function from R^2 to R^2. However, this function has properties that we need, similar to what we discussed in the univariate case. Specifically, we assume that the function is strictly increasing, interpretable, and either strictly increasing or decreasing (monotone).



Let g be a function such that g is a bivariate function, meaning it maps from R^2 to R^2. For any x belonging to R^2, where x is a vector (x_1, x_2), the vector $y = g(x)$ also belongs to R^2. We take $g_1(x)$ and $g_2(x)$, which are components of this vector, satisfying the condition that the inverse of g exists. The inverse is denoted as h, so we can write $x = g^{-1}(y) = h(y)$. In the univariate case, we discussed the derivative being either strictly greater than 0 or strictly less than 0.

However, since g is a bivariate function, we need to consider the derivatives in terms of both x_1 and x_2. This is where the Jacobian comes into play. The Jacobian is a matrix of partial derivatives, and its determinant is denoted as the Jacobian determinant. The Jacobian matrix is defined as:

$$\begin{matrix} \partial x_1/\partial y_1 & \partial x_1/\partial y_2 \\ \partial x_2/\partial y_1 & \partial x_2/\partial y_2 \end{matrix}$$

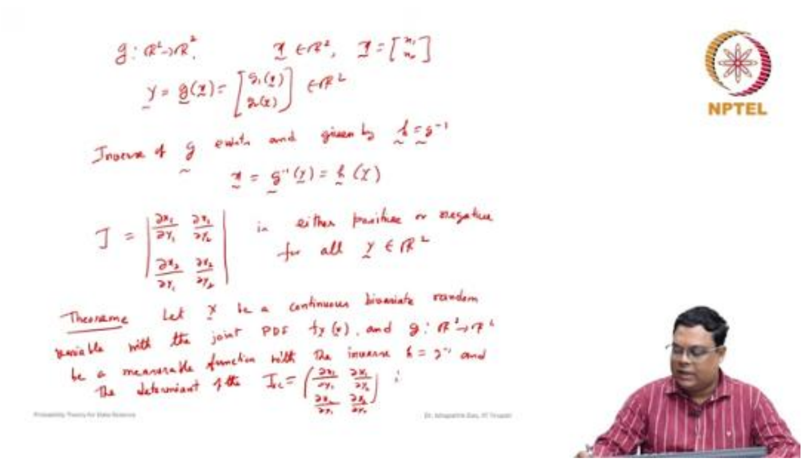
This matrix is used to calculate the determinant, which will indicate whether the function is invertible. The determinant of the Jacobian matrix will be either positive or negative for all values of y belonging to R^2 (or for the subset of R^2 that corresponds to the range of Y). It cannot be both positive and negative.

Now, if the function g satisfies these conditions, we have the following theorem:

Theorem: Let X be a continuous bivariate random variable with the joint probability density function f_X(x), and let g: $R^2 \rightarrow R^2$ be a measurable function with an inverse h = g^{-1}. Let the determinant of the Jacobian matrix, denoted as J, be either positive or negative for all y belonging to R^2. Then, the probability density function of Y, denoted f_Y(y), is given by:

$f_Y(y) = |det(J)| * f_X(h(y))$

This is the theorem.

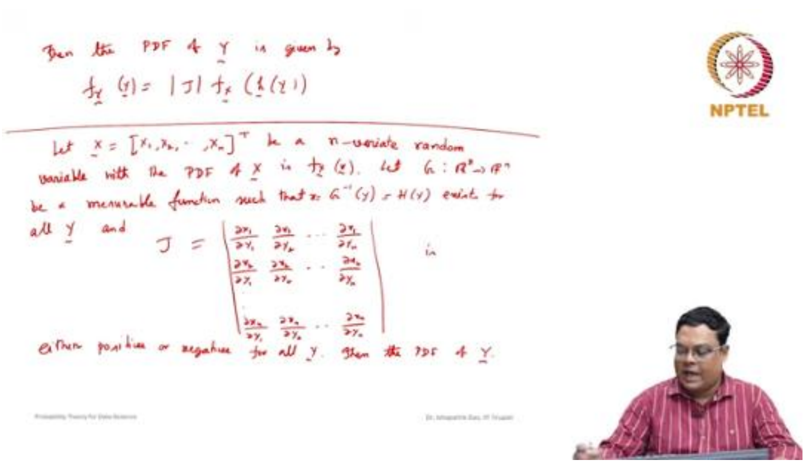


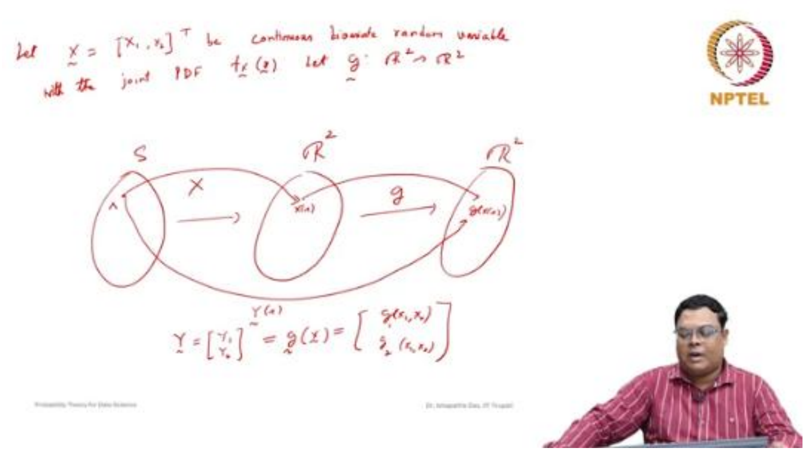
This theorem can be extended. I have just discussed it for a function of two variables in the case of bivariate random variables, but it can be extended to n variables. Let X = (x_1, x_2, ..., x_r) be a continuous multivariate random variable with the probability density function f_X(x). Suppose Y = g(X) exists such that X = g^{-1}(Y) and the determinant of the Jacobian matrix is either positive or negative for all Y. Then, the joint probability density function of Y is given by:

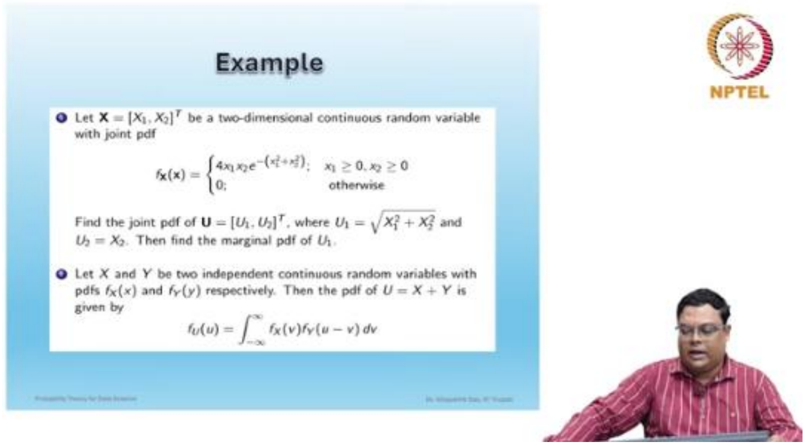
$f_Y(y) = |det(J)| f_X(g^{-1}(y))$. This theorem can be extended to n random variables. Let X = (x_1, x_2, ..., x_n) be a multivariate random variable with the probability density function f_X(x). Let g: $R^n \rightarrow R^n$ be a measurable function such that g^{-1} exists, meaning X = g^{-1}(Y) = h(Y) exists for all Y. The determinant of the Jacobian matrix J is given by:

$J = | \partial x_1/\partial y_1, \partial x_1/\partial y_2, ..., \partial x_1/\partial y_n | | \partial x_2/\partial y_1, \partial x_2/\partial y_2, ..., \partial x_2/\partial y_n | | ... | | \partial x_n/\partial y_1, \partial x_n/\partial y_2, ..., \partial x_n/\partial y_n |$ The determinant of this Jacobian matrix is either positive or negative for all Y. Therefore, the joint probability density function of Y is given by: f_Y(y) = |det(J)| f_X(h(y)), where h(Y) is the vector function representing the inverse transformation.

This is the theorem. We have already proven this theorem for the univariate case. The proof for multivariate cases will be a little more complicated, but we will discuss it. We have already discussed the univariate case, so now we will go over some numerical examples for multivariate transformations. Let us consider one example. To better understand this concept, let's consider a simple example.







Let X be a continuous bivariate random variable with the probability density function f(X), which is given by: f(X) = e^(-x_1 - x_2), where $x_1 > 0$ and $x_2 > 0$, and 0 otherwise. This is the density function. Note that we are using the following notation repeatedly: capital X_1 and X_2 for random variables, and small $\tilde{x} = (x_1, x_2, ..., xn)$ for real numbers. This function is defined from the sample space S to R^n, which is the notation we are following.

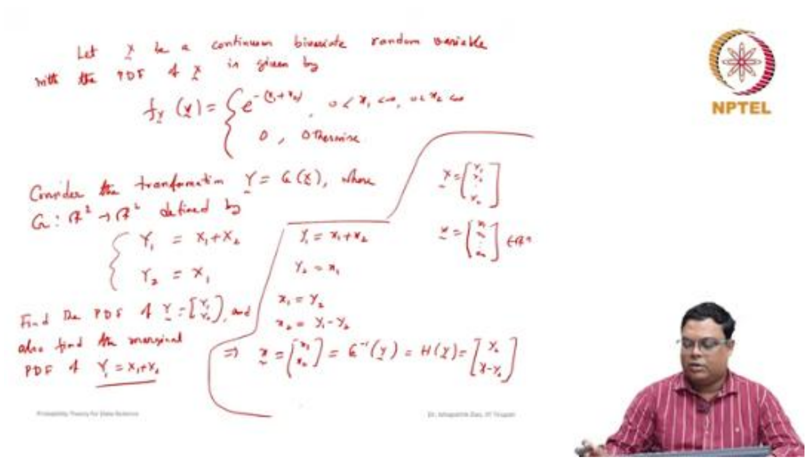
Thus, when we write $\tilde{x}$, it refers to the vector (x_1, x_2) because we are working with a bivariate random variable. Consider the transformation Y = g(X), where g is a transformation from R^2 to R^2 defined by: $Y_1 = X_1 + X_2$, $Y_2 = X_1$. Now we will check whether these satisfy the conditions. This transformation results in the equations:

$Y_1 = x_1 + x_2$, $Y_2 = x_1$. To find the inverse transformation, we solve for x_1 and x_2: $x_1 = y_2$, $x_2 = y_1 - y_2$.

Thus, the inverse of g, denoted as $g^{-1}(Y)$, is: $x = g^{-1}(y) = (x_1, x_2) = (y_2, y_1 - y_2)$. Next, we need to compute the Jacobian for this transformation. The next step is to find the probability density function of $Y = (Y_1, Y_2)$, and also the marginal probability density function of Y_1. In this case, we are interested in finding the probability density function of $X_1 + X_2$.

While the transformation might not always be explicitly given as R^2 to R^2, we are always interested in finding the probability density function for $X_1 + X_2$. According to the theorem, the transformation should be from R^2 to R^2. Therefore, we can consider a bivariate random variable transformation from R^2 to R^2 and then integrate out Y_2 from the bivariate distribution to find the marginal probability density function of Y_1. In this example, we found that: $Y_1 = X_1 + X_2$, $Y_2 = X_1$.

$Y_2 = X_1$. From here, we have found the inverse, $x = g^{-1}(y)$, which is h(y), and this is y_2, $y_1 - y_2$. Now, we want to find the Jacobian. The inverse is given by $x_1 = y_2$ and $x_2 = y_1 - y_2$.



Next, we calculate the Jacobian. The Jacobian J is given by the partial derivatives:

$J = [\partial x_1/\partial y_1\ \partial x_1/\partial y_2]\ [\partial x_2/\partial y_1\ \partial x_2/\partial y_2]$

From the transformation, we find the following values:

$\partial x_1/\partial y_1 = 0$, $\partial x_1/\partial y_2 = 1$, $\partial x_2/\partial y_1 = 1$, $\partial x_2/\partial y_2 = -1$.

Thus, the Jacobian matrix becomes:

$J = [0\ 1]\ [1\ -1]$

The determinant of this Jacobian matrix is -1, which is less than 0 for all values of y, satisfying the condition. Hence, the joint probability density function of y is given by:

$f(y) = |J| * f(x, h(y))$.

Since the determinant of the Jacobian is -1, its absolute value is 1. Now we need to determine f(x). From the given probability density function:

$f(x) = e^{-(x_1 + x_2)}$ for $0 < x_1 < \infty$ and $0 < x_2 < \infty$.

This can be written as e^{-y_1} because $x_1 + x_2 = y_1$, where $y_1 > 0$.

Also, we know that $x_1 = y_2$ and $x_2 = y_1 - y_2$. Therefore, both y_2 and $y_1 - y_2$ must be greater than 0 and less than infinity. This implies that $0 < y_2 < y_1 < \infty$.

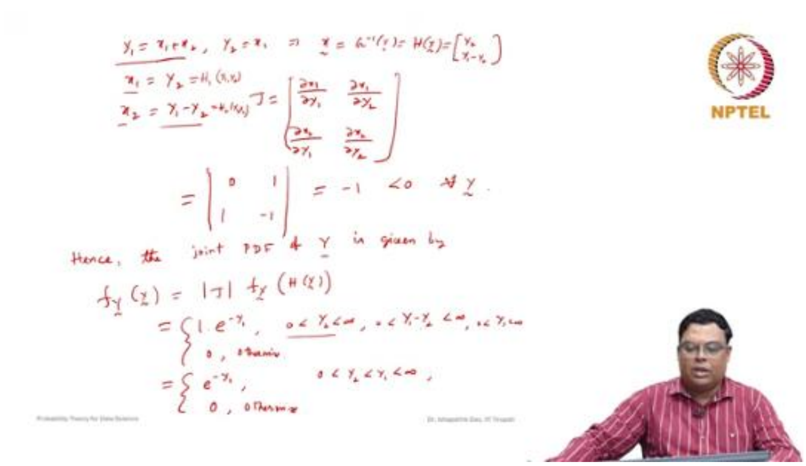
Now, we can simplify the expression:

$f(y) = e^{-y_1}$, where $0 < y_2 < y_1 < \infty$.

This is the joint probability density function of Y. Next, we are interested in the marginal probability density function of y_1. The marginal probability density function of y_1 is given by:

$f(y_1) = \int_0^{y_1} f(y_1, y_2)\, dy_2$.

Thus, the marginal probability density function of y_1 is obtained by integrating over y_2.



By definition, since we know the joint probability density function for y_1 and y_2, integrating with respect to y_2 will give us the marginal probability density function. The notation for the joint probability density function is the same as $f(y_1, y_2)$, but we simplify it for convenience. Note that when integrating, the density is non-zero within certain conditions. Specifically, we need to find the range of y_2 where it is non-zero. For this case, the range of y_2 is between 0 and y_1.

Therefore, the marginal probability density function is given by the integral:

$\int_0^{y_1} e^{(-y_1)}\, dy_2$.

Since the integrand is independent of y_2, we can factor out $e^{(-y_1)}$:

$e^{(-y_1)} \int_0^{y_1} dy_2$.

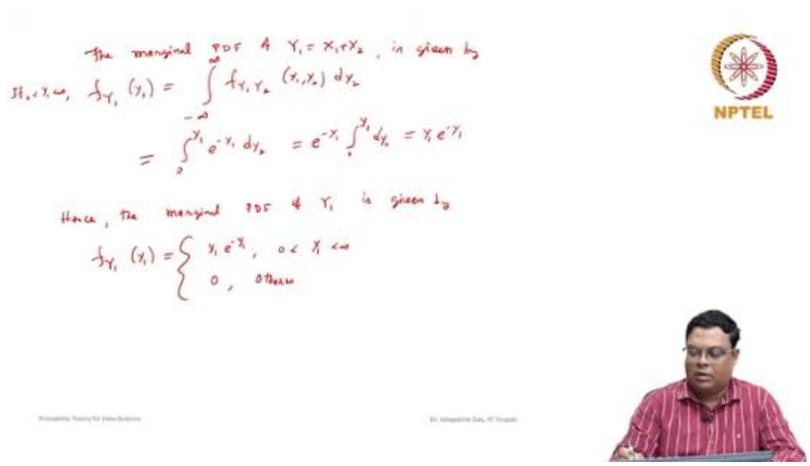
The integral of dy_2 from 0 to y_1 is simply y_1. Thus, we have:

$y_1\, e^{(-y_1)}$.

This is valid when y_1 is greater than 0 and less than infinity. If y_1 is negative, the density function is 0. Therefore, the marginal probability density function of y_1 is:

$f(y_1) = y_1\, e^{(-y_1)}$ for $0 < y_1 < \infty$, and 0 otherwise.

This is the desired density function. Hopefully, you have understood the concept of how to use transformations of random variables to find the joint probability density function for a transformed random variable. We will explore more numerical examples to understand it better. Let's consider a slightly more complicated example that requires some computation. In this case, we are given a joint density function for a two-dimensional continuous random variable.



Let x_1 and x_2 be the two random variables, with the joint probability density function given by: $f(x_1, x_2) = 4x_1x_2\, e^{(-x_1^2 + x_2^2)}$. The task is to find the joint probability density function of a transformed random variable, where the transformation is given by: $u_1 = f(x_1, x_2)$ and $u_2 = g(x_1, x_2)$. Then, we are to find the marginal probability density function of u_1. Let's now proceed to solve that.