PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

Week - 11

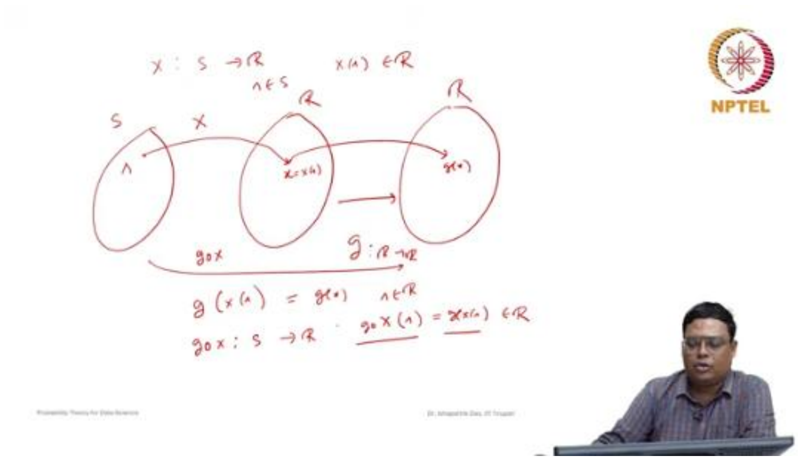
Lecture - 53

Transformation of Random Variables

Whenever we define a random variable, it is a function from a sample space (S) to the real numbers (R). However, it is not just any function; it is a measurable function. This means that if X is a random variable, it is a measurable function from the sample space S to R. For any element s belonging to the sample space S, the value of X(s) will be a real number. Moreover, if you consider a Borel set, the inverse of that set under the function X must belong to the sigma field of the sample space.

The sigma field consists of a class of subsets of S, ensuring that X is a measurable function. Now, consider another function g, which maps real numbers to real numbers (i.e., $g : R \to R$). If we apply this function to the random variable X, then g(X(s)) will be a value in the real numbers. This is a composite function, where g is applied to the outcome of the random variable X. To put it simply, if X maps a sample space element s to a real number, then g(X(s)) gives us another value in the real numbers.

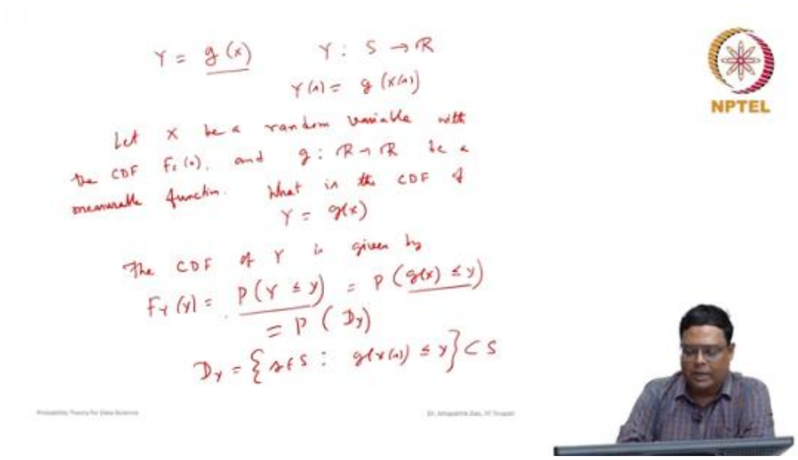
The composite function, denoted as $g \circ X$, is a function from the sample space S to R, where for any element s in S, the value of g(X(s)) is a real number. If the function g is measurable, meaning it satisfies the conditions for measurability, then the composite function $g \circ X$ will also be a measurable function. In particular, if g is continuous, it will be a measurable function, and the composite function will be measurable as well. This composite function $g \circ X$ represents another random variable. Let $Y = g(X)$.



This means that Y is a random variable, and it is a measurable function from the sample space S to the real numbers (R), defined by $Y(s) = g(X(s))$. This defines a new random variable. We have already used this concept to find the expected values of Y by using the probability distribution function of X. However, if we are interested in finding the expected value or variance of Y with respect to its own probability distribution (either the probability mass function for discrete variables or the probability density function for continuous variables), we need to focus on the distribution of Y itself. So, to find this, suppose X is a random variable with the cumulative distribution function (CDF) $F_X(x)$.

This is known. Let g be a measurable function from R to R. The goal is to determine the cumulative distribution function (CDF) of Y. To find the CDF of Y, we need to compute the probability that Y is less than or equal to a certain value y. The CDF of Y, denoted $F_Y(y)$, is given by the probability that $Y \leq y$.

We can express this as: $F_Y(y) = P(Y \leq y) = P(g(X) \leq y)$. This can be rewritten as the probability that X belongs to the set $D(y)$, where $D(y)$ is defined as all the elements in the sample space S such that $g(X(s)) \leq y$. So, the set $D(y)$ consists of all s in the sample space for which $g(X(s))$ is less than or equal to y. At this point, we are interested in understanding the relationship between the CDF of Y and the set $D(y)$, but we do not explicitly know the distribution of Y yet.

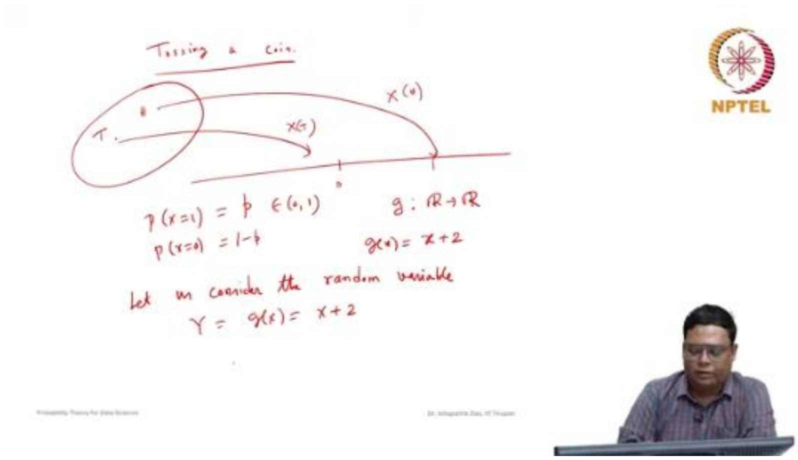


We can find the probability, but in certain cases, we need to consider examples to illustrate how to compute it. Let's discuss one such example. Suppose we are tossing a coin, and we define the sample space S as either heads or tails. So, the sample space consists of "head" or "tail." Now, let's define a random variable X such that X("head") = 1 and X("tail") = 0.

If we consider a transformation function g(X), where g is a function from R to R, we can calculate the new probability distribution. In this case, the random variable X is a discrete random variable, and we can represent it using a probability mass function. For example, the probability that X = 1 ("heads") is p, where $p \in [0, 1]$, and the probability that X = 0 ("tails") is 1 - p. So, the probability mass function is given by: P(X = 1) = p, P(X = 0) = 1 - p.

Next, we introduce a transformation. Let's consider the transformation $g: R \to R$, defined by g(X) = X + 2. This is a simple transformation where we add 2 to the value of X. Now, we define a new random variable Y as Y = X + 2. We want to find the probability mass function of Y.

To do so, we first need to determine the range of Y. So, what are the possible values that Y can take? The sample space consists of "head" and "tail," and each outcome is mapped to a real number. Under the random variable X, the value for "head" is 1 and for "tail," it is 0.



When we apply the transformation function g(X), which is defined as g(X) = X + 2, we can find the corresponding values for Y. For instance, if X = 1 (head), then g(1) = 1 + 2 = 3. If X = 0 (tail), then g(0) = 0 + 2 = 2. Therefore, for the random variable Y, we have:

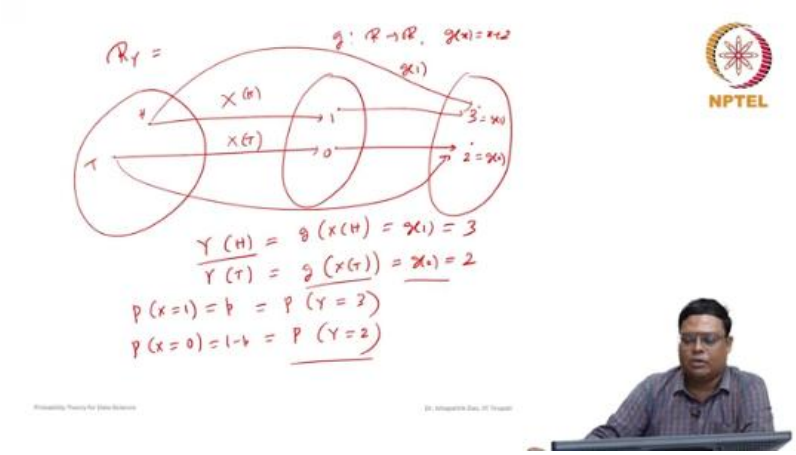
Y(head) = g(X(head)) = g(1) = 3
Y(tail) = g(X(tail)) = g(0) = 2

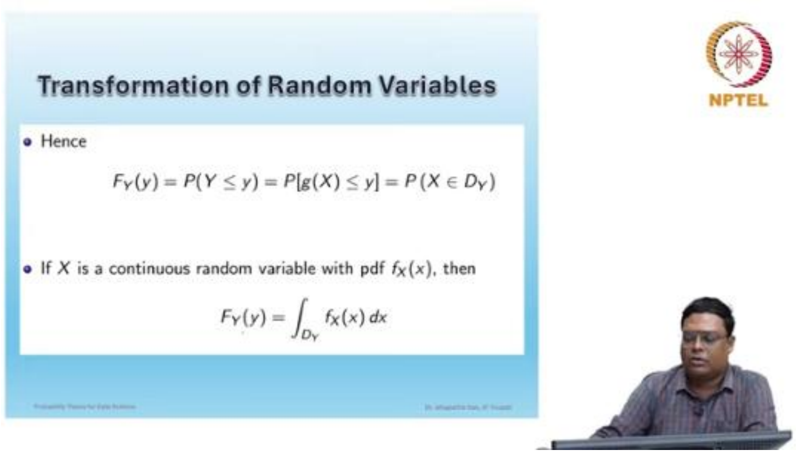
The probability of getting "head" is p, and the probability of getting "tail" is 1 - p. For the new random variable Y, we have:
P(Y = 3) = P(X = 1) = p (because X = 1 corresponds to "head")
P(Y = 2) = P(X = 0) = 1 - p (because X = 0 corresponds to "tail")

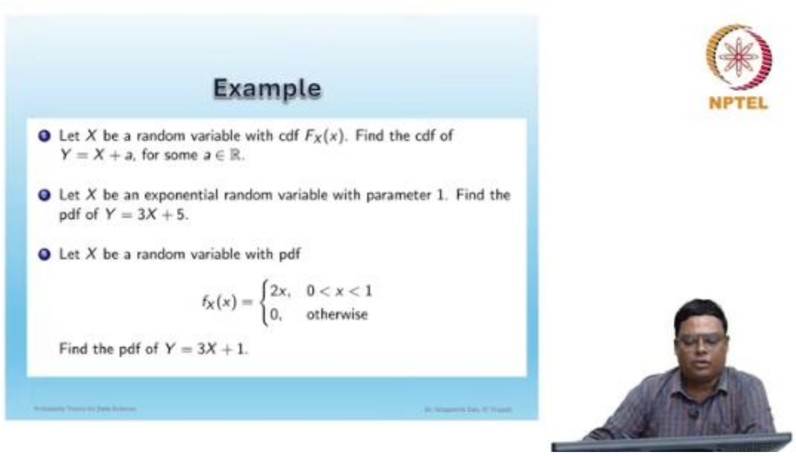
Therefore, the probability mass function of Y is:
P(Y = 3) = p
P(Y = 2) = 1 - p

Thus, the transformation has changed the probability mass function. From here, we can also calculate the cumulative distribution function (CDF) of Y, if needed. This is a simple example to illustrate how transformations affect the probability mass function of a random variable.

In more general cases, similar methods can be applied to find the distribution of transformed random variables. So, for example, for continuous cases, we need to do this kind of indication. So, how we will do that? Let us discuss some of the examples. For continuous cases, we need to follow this kind of procedure. It is similar to what we discussed earlier.



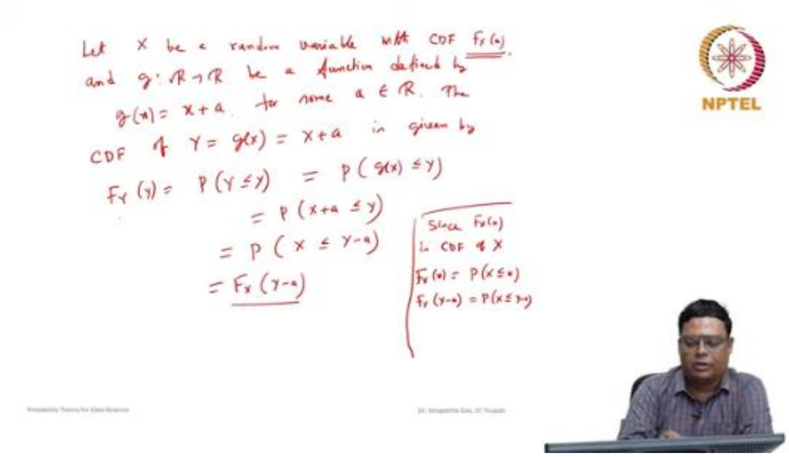


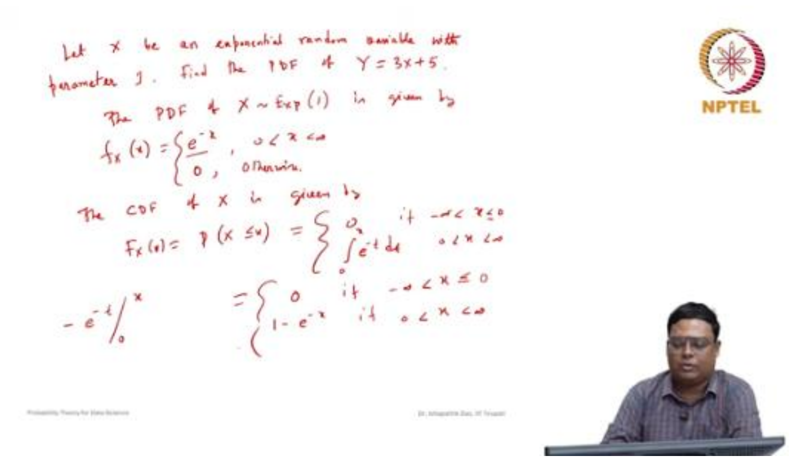


Let X be a random variable with a cumulative distribution function (CDF), and let g: R → R be a function defined as g(x) = x + a, where a is a fixed value belonging to the real numbers. For example, a could be 2, 3, or 0. If a is 0, the function becomes the identity function, g(x) = x, meaning no transformation occurs, and the distribution of X remains unchanged. However, for any non-zero transformation, we need to find the cumulative distribution function (CDF) of the transformed random variable Y, where Y = g(X).

The CDF of Y, denoted as F_Y(y), is the probability that Y is less than or equal to y. Since Y = g(X), this probability can be written as the probability that g(X) is less than or equal to y. Since g(x) = x + a, this becomes the probability that x + a is less than or equal to y. Rearranging this inequality, we get x ≤ y - a. Now, using the CDF of X, which is F_X(x), we know that F_X(x) is the probability that X is less than or equal to x.

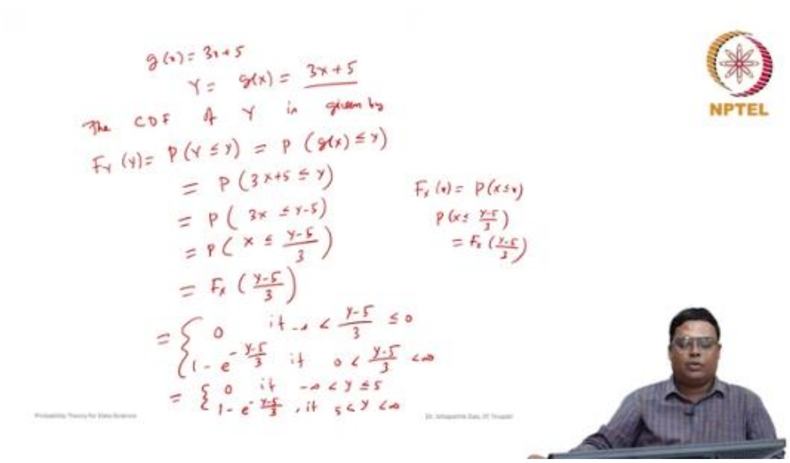
Therefore, we can express the CDF of Y as F_X(y - a). This is the general approach to finding the CDF of a transformed random variable Y. Let's consider an example. Let X be an exponential random variable with a parameter of 1. The task is to find the probability density function (PDF) of Y = 3X + 5. To solve this, start by determining the cumulative distribution function (CDF). For X, which follows an exponential distribution, the CDF is zero for values less than or equal to zero.

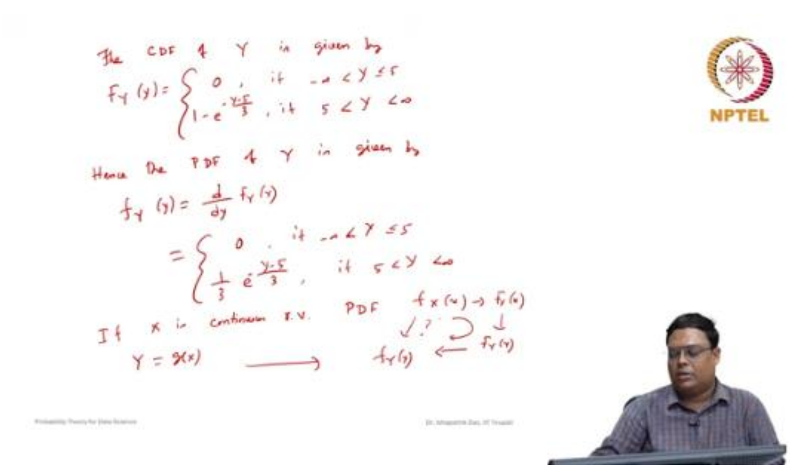




For values greater than 0, the CDF is obtained by integrating the exponential PDF from 0 to a given value. Next, consider the transformation Y = 3X + 5. To find the CDF of Y, rewrite the inequality "Y ≤ y" in terms of X, which gives "X ≤ (y - 5) / 3." For values of Y ≤ 5, the CDF is 0 because there is no density in this range. For values of Y > 5, the CDF of Y corresponds to the CDF of X evaluated at (y - 5) / 3.

Finally, the PDF of Y is obtained by differentiating its CDF with respect to Y. For Y ≤ 5, the PDF is 0. For Y > 5, the PDF is derived from the transformation, resulting in a scaled exponential decay function. This completes the derivation of the PDF for Y = 3X + 5. Let us first go through this example, and then we will discuss how we can utilize this concept.



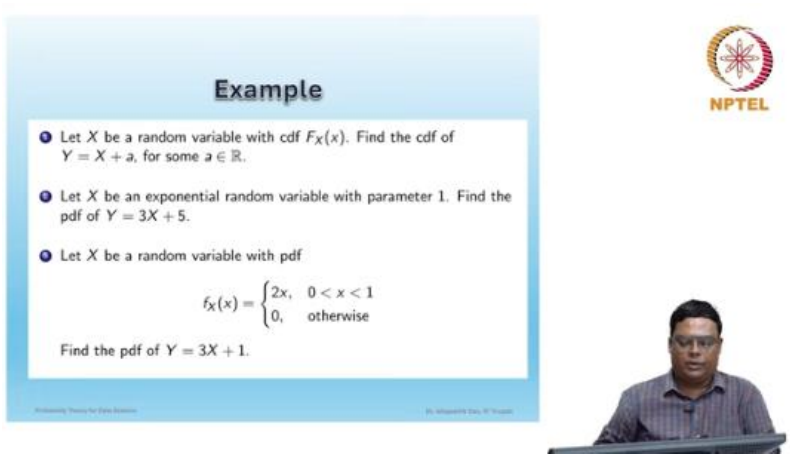


Consider the random variable X with a given probability density function. In this example, let X be a random variable with a probability density function between 0 and 1, where it is 2x, and 0 otherwise. The task is to find the probability density function. In this case, the result is as follows: 0 if $y \leq 5$, and 1 - e^(...) if $y > 5$. Now, to find the probability density function, we need to note that since this is a continuous random variable, the probability density function can be found by differentiating the cumulative distribution function.

Thus, the probability density function of Y is given by the following: it is 0 when $y \leq 5$. When differentiating the cumulative function, it gives a value of (1 / 3) * e^(...) when $y > 5$. This is the probability density function of Y. This illustrates the method of finding the probability density function. It is important to note that if X is a continuous random variable, we aim to find its probability density function.

For X, the probability density function is defined. Now, if we apply a transformation to X, we must first find the cumulative distribution function of X, apply the transformation, and then differentiate to find the probability density function of the transformed variable Y. This is the method we used. While it works, it can be a little complicated. If there were a direct way to find the probability density function of Y from the known density function of X, it would save us the extra steps.

In most cases, we are more interested in the probability density function of continuous random variables rather than their cumulative distribution function. Therefore, a direct method would be helpful because it would eliminate the need to find the cumulative distribution function during the transformation process. Let us look at another example to understand this concept further. So let us first do this example and then we will discuss how we can utilize this concept. Let us start with this example and then discuss how we can utilize this concept.

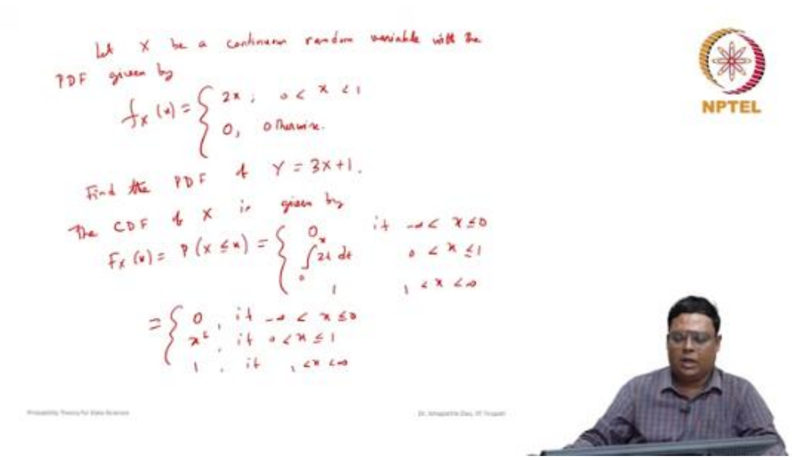


Consider a random variable X with a given probability density function. In this example, X is a random variable with a probability density function defined as 2x for values between 0 and 1, and 0 otherwise. The task is to find the probability density function of Y, where Y is defined as 3x + 1. To solve this, we use the method involving the transformation of random variables. First, we need to find the cumulative distribution function of X.

X is a continuous random variable, and its probability density function is given as 2x for values between 0 and 1, and 0 otherwise. To find the cumulative distribution function of X, denoted by F(x), we calculate the probability that $X \leq x$. Since the density is nonzero only between 0 and 1, we can determine the cumulative distribution function as follows: $F(x) = 0$ for $X \leq 0$, as there is no density for negative values. For values of X between 0 and 1, the cumulative distribution function is the integral from 0 to x of the density function, which is the integral of 2t dt.

Solving this gives t^2, so $F(x) = x^2$ in this range. For $X > 1$, the cumulative distribution function is 1, as the total probability is 1. Thus, the cumulative distribution function of X is: $F(x) = 0$ if $X \leq 0$, $F(x) = x^2$ if $0 < X \leq 1$, and $F(x) = 1$ if $X > 1$.

Now, using the cumulative distribution function of X, we find the cumulative distribution function of Y. Here, Y is defined as $Y = 3X + 1$. This is given as $F_Y(y)$, which is the probability that $Y \leq y$. Since $Y = g(X)$, this is the probability that $3X + 1 \leq y$. Subtracting 1 from both sides, we have $3X \leq y - 1$.



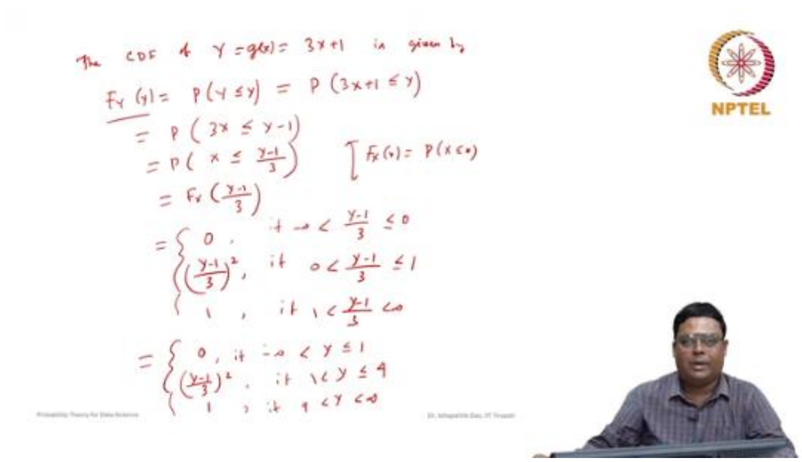
Dividing by 3, we get $X \leq (y - 1) / 3$. This is equivalent to $F_Y(y) = F_X((y - 1) / 3)$, where $F_X(x)$ is the cumulative distribution function of X.

The definition of $F_X(x)$ is:
$F_X(x) = 0$ if $x \leq 0$.
$F_X(x) = x^2$ if $0 < x \leq 1$.
$F_X(x) = 1$ if $x > 1$.

Substituting $(y - 1) / 3$ into $F_X(x)$:
$F_Y(y) = 0$ if $(y - 1) / 3 \leq 0$, meaning $y \leq 1$.
$F_Y(y) = ((y - 1) / 3)^2$ if $(y - 1) / 3 > 0$ and $(y - 1) / 3 \leq 1$, meaning $1 < y \leq 4$.
$F_Y(y) = 1$ if $(y - 1) / 3 > 1$, meaning $y > 4$.

Simplifying this, the cumulative distribution function of Y is:
$F_Y(y) = 0$ if $y \leq 1$.
$F_Y(y) = ((y - 1) / 3)^2$ if $1 < y \leq 4$.
$F_Y(y) = 1$ if $y > 4$.

But it is asked that we first find the probability density function (PDF) of Y, where $Y = 3X + 1$.



So, how do we find that? Now, we will differentiate with respect to y to get the probability density function of Y, denoted as $f_Y(y)$. Since the cumulative distribution function (CDF) is 0 when $y \leq 1$, we know that the PDF in this range is 0. So, for $y \leq 1$, the probability density function is 0. For $1 < y \le 4$, we have the term $(y - 1) / 3$, squared. Taking the derivative of this with respect to y:

First, the constant 1/9 remains. The derivative of $(y - 1)^2$ is $2(y - 1)$. So, the probability density function is $(2/9) * (y - 1)$ for $1 < y \le 4$. Now, for $y > 4$, the probability density function is constant, so its derivative will be 0. Thus, the probability density function can be written as:

$f_Y(y) = (2/9) * (y - 1)$ for $1 \le y \le 4$.
$f_Y(y) = 0$ for $y < 1$ or $y > 4$.

By verifying this with the integration from 1 to 4, we find that the total probability equals 1, confirming that the calculations are correct. In this process, we first found the CDF of X and then used it to find the CDF of Y. After differentiating the CDF of Y, we obtained the PDF of Y.

However, there is a more direct method for finding the PDF of Y if we know the PDF of X. We will discuss this method in the next section using a theorem that allows us to avoid the steps we just followed. This method we just used is the general method, but it may not be applicable to all cases. In specific cases, the method we will discuss can be more efficient. Let's move on to that now.