PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

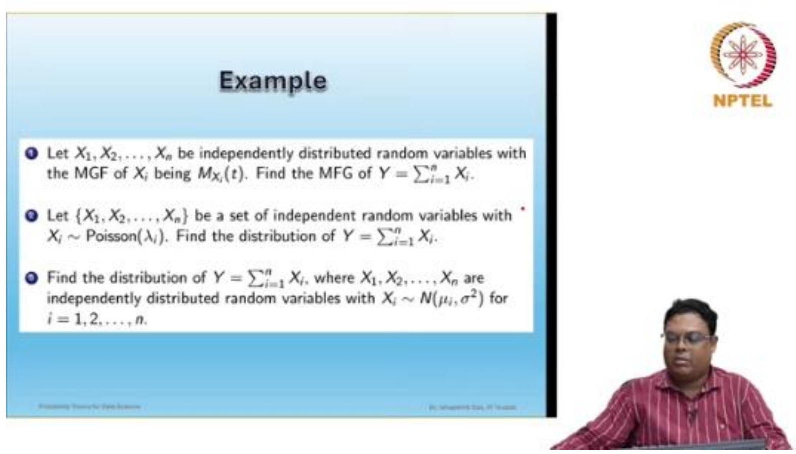
Indian Institute of Technology Tirupati

Week - 12

Lecture - 61

Moment Generating Functions for the Transformation of Random Variables

Let us discuss this problem. Let $X_1, X_2, ..., X_n$ be independently distributed random variables with moment generating function $M\_X_i(t)$. Find the moment generating function of Y, where $Y = \Sigma_{i=1}^n X_i$. Let $X_1, X_2, ..., X_n$ be independent random variables, with the moment generating function of X_i denoted as $M\_X_i(t)$ for i = 1 to n. This is the general case.



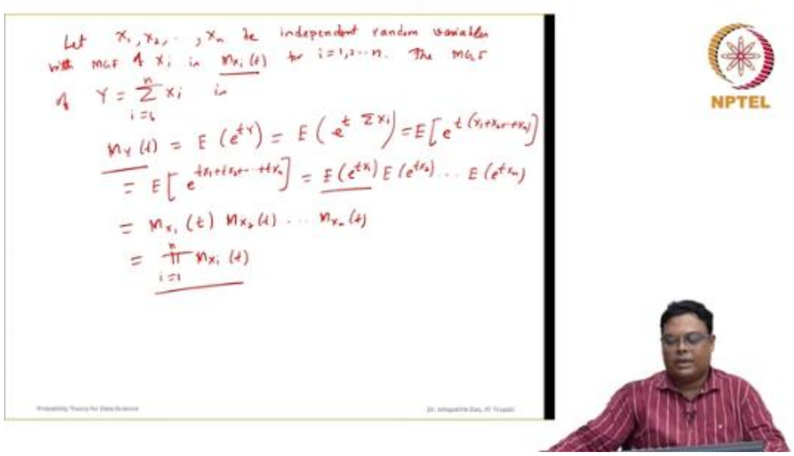
In a particular case, we can discuss some of the $M\_X_i(t)$ values. Let $Y = \Sigma_{i=1}^n X_i$. We need to find the moment generating function of Y, assuming these values are all known. So, what is the definition of the moment generating function of Y? $M_Y(t)$ is nothing but the expected value of e^(tY). This is equal to the expected value of $e^{(t * \Sigma_{i=1}^n X_i)}$, because $Y = \Sigma_{i=1}^n X_i$. So, if we explicitly write it, we get the expected value of $e^{(t(X_1 + X_2 + ... + X_n))}$.

So, if you simplify this, it becomes the expected value of e^{tX_1}, plus the expected value of e^{tX_2}, and so on, up to e^{tX_n}. This is equivalent to the expected value of e^{tX_1} * expected value of e^{tX_2} * ... * expected value of e^{tX_n}. Now, this function can be written as $M_{X_1}(t) * M_{X_2}(t) * \ldots * M_{X_n}(t)$.

What is this? The expected value of e^{tX_1} is nothing but the moment generating function of X_1. Similarly, it is the moment generating function of X_2, and so on for X_n. Since the moment generating functions of all these variables are known, the moment generating function of the sum of these random variables can be represented as the product of the moment generating functions from i = 1 to n, $M_{X_i}(t)$.

If you know the moment generating function of each independent random variable, you can simply multiply those moment generating functions to find the moment generating function of the sum of these random variables. This is one property of moment generating functions. Hopefully, you have understood it.

Next, we will discuss an example. Let $X_1, X_2, \ldots, X_n$ be a set of independent random variables, where each X_i follows a Poisson distribution with parameter λ_i. As we discussed earlier, if you have n random variables with Poisson(λ_i) distributions and you want to find the distribution of Y, where $Y = \sum_{i=1}^{n} X_i$, how can we find that? So, first, what will we do to find the distribution of Y?



First, we need to know the moment generating function of the Poisson distribution with parameter λ. So, in the general case, let's find that. Let X be a Poisson random variable with parameter λ, where λ > 0. First of all, let's write down the probability mass function because we need it to find the moment generating function. The probability mass function of X is given by p_X(k), where k represents the integer values.

The probability mass function is e^(-λ) * λ^k / k!, where k can take values 0, 1, 2, and so on, and 0 otherwise. Then we can find the moment generating function from the definition. The moment generating function of X is given by M_X(t), which is equal to the expected value of e^(tX). By definition, since it is a discrete random variable, this is the summation from k = 0 to ∞ of e^(t * k) * p_X(k). So, the summation from k = 0 to ∞ of e^(t * k), then p_X(k) is e^(-λ) * λ^k / k!, for k = 0, 1, 2, and so on.

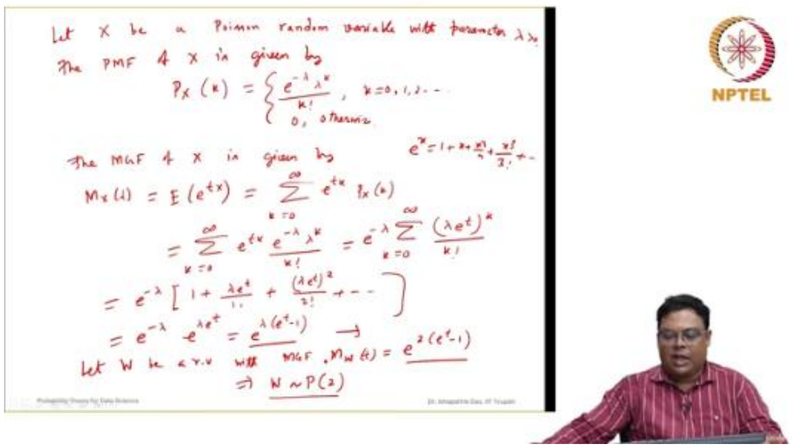
Sorry, let me clarify why I'm writing this. Essentially, this is the probability mass function. Now, we need to find the solution. So, to find the simplified form, how can we do that? Since λ and e^λ are constants, we have e^(-λ), with k ranging from 0 to ∞. Now, we can write λ * e^t raised to the power of k / k!. So, this simplifies to e^(-λ) * [1 + λ * e^t / 1! + (λ * e^t)^2 / 2! + ...]. This is an infinite series. We know that this expression resembles e^x, which is written as 1 + ax + x^2 / 2! + x^3 / 3!, and so on.

So, that is why this is nothing but e^(-λ) * e^(λ * e^t). Hence, we can write it as e^(-λ) * e^(λ * (e^t - 1)). This holds for any $t \in \mathbb{R}$. So, this is the moment generating function of a Poisson random variable with parameter λ. Hence, you can remember that this is the form of the moment generating function of a Poisson random variable.

Suppose there is a random variable, W, with moment generating function M_W(t). Let W be a random variable with moment generating function M_W(t). We do not know the distribution, but we know the moment generating function. The moment generating function M_W(t) is e^(t - 1). When you compare this with the standard form, it looks very similar, except for the change from λ to 2. Hence, by comparing these, we can conclude that W follows a Poisson distribution with parameter 2, based on the uniqueness theorem of the moment generating function. So, it cannot be different. If W is a Poisson random variable with parameter 2 (λ = 2), then this moment generating function looks like this. There cannot be any other distribution with the same moment generating function, because if two different random variables have the same moment generating function, they must have the same distribution. So, by the uniqueness theorem, we can say that W must have a Poisson distribution with parameter 2.

This is the usefulness of the moment generating function for identifying a distribution. Now, let's look at this problem. We need to find the distribution of Y. First of all, X_1, X_2,

..., X_n are a set of independent random variables, with each X_i following a Poisson distribution with parameter λ_i. Let's go through it.



Let X_1, X_2, ..., X_n be independent random variables, where each X_i has a Poisson distribution with parameter λ_i for i = 1 to n. So, here $\lambda_i > 0$. Therefore, $\lambda_i > 0$ for i = 1 to n. These are independent random variables. Now, we know that the moment generating function of X_i is what we just computed. We replace λ with λ_i here, as the parameter is λ_i. The moment generating function of X is given by $M_X(t)$, which represents the expected values. Here, we directly write it, and we won't compute it. So, the moment generating function of X_i is given by $M_{X_i}(t)$. The parameter for this is λ_i, so it will be $e^{(\lambda_i)} * (e^t - 1)$.

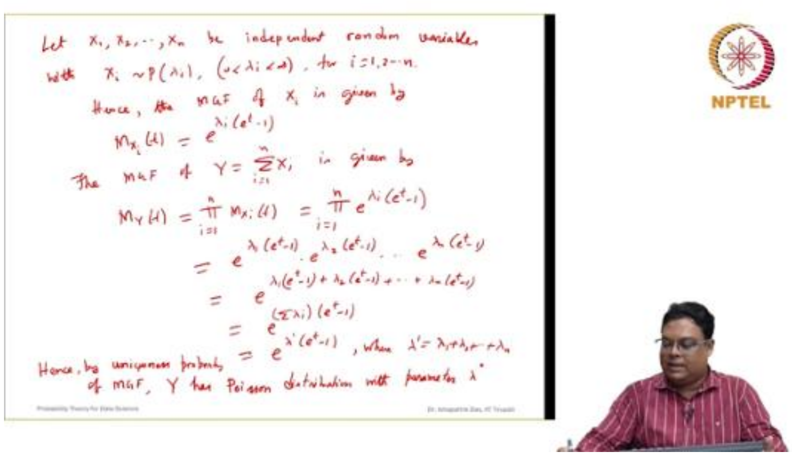
Now, we will use this theorem. Hence, the moment generating function of Y is the sum of X_i, since they are independent, with i running from 1 to n. This is given by $M_Y(t)$. So, these things we have already done. In this theorem, we can see that whenever X_1, X_2, ..., X_n are independent random variables, the moment generating function for each X_i in the general case will be the product of their individual moment generating functions. This means that the moment generating function of Y will be the product of the moment generating functions of each X_i, from i = 1 to n.

So, the moment generating function of each X_i is equal to an expression involving λ_i and t, which includes a term with $e^{(\lambda_i)} * (e^t - 1)$. This product, I believe, is understood. We have used these kinds of notations several times before, or we can explicitly write them out. Now, this expression is simply $e^{(\lambda_1 * (e^t - 1))}$, then $e^{(\lambda_2 * (e^t - 1))}$, and so on, up to $e^{(\lambda_n * (e^t - 1))}$. I have just explicitly written this out.

Essentially, this is the summation of these terms. This is $\lambda_1 * e^{(t - 1)}$, plus $\lambda_2 * e^{(t - 1)}$, and so on, up to $\lambda_n * e^{(t - 1)}$. Since it is a multiplication in the power, it will become a summation. So, this is the summation of $\lambda_i * e^{(t - 1)}$, which is the common form. This is summation 1.

This is simply $e^{(\lambda' * (e^t - 1))}$, where λ' is the sum of λ_1, λ_2, and up to λ_n. Now, if you compare this with the Poisson distribution, we do not yet know the distribution of Y. It looks very similar to this. Now, by the uniqueness property of the moment generating function, we can conclude that Y follows a Poisson distribution with parameter λ', which is equal to the summation of λ_i values.

Therefore, we found that for X_1, which follows a Poisson distribution with parameter λ_1, and X_2, which follows a Poisson distribution with parameter λ_2, we have shown that $X_1 + X_2$ follows a Poisson distribution with parameter $\lambda_1 + \lambda_2$, by directly finding the probability mass function. In the general case, if you consider $X_1, X_2, \ldots, X_n$ as independent random variables with Poisson distributions, we found that if $X_1, X_2, \ldots, X_n$ are independent random variables, where each X_i follows a Poisson distribution with parameter λ_i, and $\lambda_i > 0$ and $< \infty$ for i = 1 to n, then $Y = X_1 + X_2 + \ldots + X_n$ follows a Poisson distribution with parameter $\lambda_1 + \lambda_2 + \ldots + \lambda_n$. This is the general result we have proved here using the moment generating function. Hopefully, you have followed and understood these concepts.



Next, we will discuss more examples and explore how the moment generating function can be useful. We will find the distribution of Y, which is the summation of X_i, in similar types of problems where $X_1, X_2, ..., X_n$ are independently distributed. Consider random variables where X_i follows a normal distribution with mean μ_i and variance σ^2 for i = 1 to n. We have already discussed that the summation of independent random variables, X_i, results in a normal distribution. Here, we will use the moment generating function to prove this.

So, how can we do that? Let us recall what the moment generating function is. Let $X_1, X_2, ..., X_n$ be independent random variables, where each X_i follows a normal distribution with mean μ_i and variance σ^2. We can consider different σ values for each i, but here it is given that X_i follows a normal distribution with mean μ_i and variance σ^2 for i = 1 to n. Recall that if X follows a normal distribution with mean μ and variance σ^2, the moment generating function of X is given by $e^{(\mu t)} * e^{(\frac{1}{2}\sigma^2 t^2)}$.

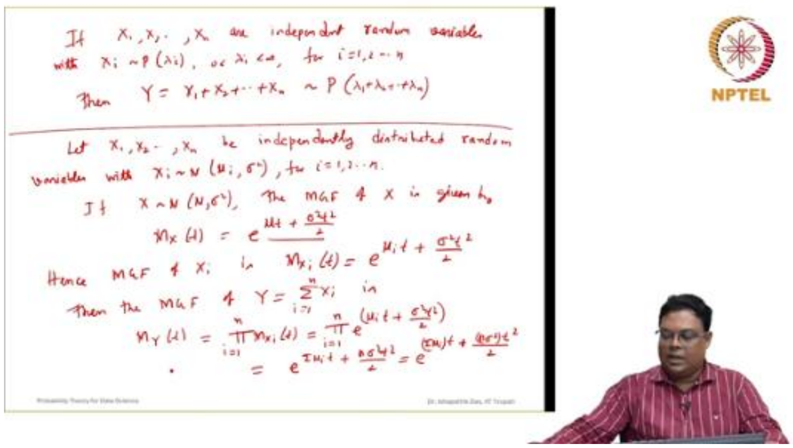
Let us go through this again. When Y is normally distributed with mean μ and variance σ^2, the moment generating function of Y is $e^{(t\mu)} * e^{(\frac{1}{2}\sigma^2 t^2)}$. This is what we have written here. Hence, the moment generating function of X_i is $M_{X_i}(t)$, which is equal to the moment generating function of X_i with the mean μ_i. So, it is $e^{(\mu_i t)} * e^{(\frac{1}{2}\sigma^2 t^2)}$, where σ^2 is the same for all i.

This is the moment generating function of X_i. Now, using this result, the moment generating function of Y is the summation of X_i, where i ranges from 1 to n. Since they are independently distributed random variables, we apply the previous result. The moment generating function of Y is given by $M_Y(t)$, which is the expected value. Because the random variables are independently distributed, we write this as the product of $M_{X_i}(t)$ for i from 1 to n.

This is equivalent to the product from i = 1 to n of $e^{(\mu_i t)} * e^{(\frac{1}{2}\sigma^2 t^2)}$. Since this is a product, it will become a sum. Notice that the σ is the same for all terms. This will result in $e^{(\sum(\mu_i t))} * e^{(n * \frac{1}{2}\sigma^2 t^2)}$. Please go through this.

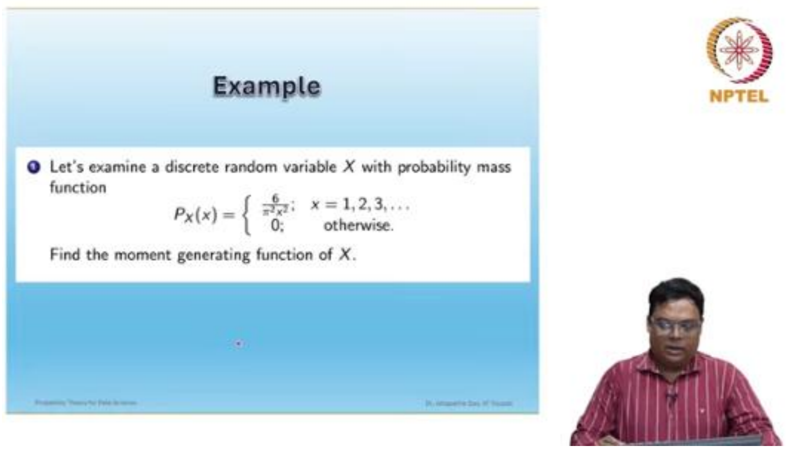
To sum it, you can express it as a product. This product, when written as a power, will be the summation of $\mu_i * t$. While each μ_i is different, the sum of n terms will be the sum of σ^2, so it becomes $n * \sigma^2 * t^2 / 2$. Hence, this looks like $e^{(\sum(\mu_i t))}$, with the coefficient of t^2 being $n * \sigma^2$, and then $t^2 / 2$. This also resembles the probability function of a random variable, with the only changes being that μ is replaced by $\sum(\mu_i)$ and σ^2 is replaced by $n * \sigma^2$.

Hence, by the uniqueness property of the moment generating function, Y, which is the summation of X_i, has a normal distribution. The mean here is $\sum(\mu_i)$, and the variance is replaced by n * σ^2. See that the variance, σ^2, is replaced by n * σ^2. So, the variance is now n * σ^2, which means Y has a normal distribution with mean = $\sum(\mu_i)$ and variance = n * σ^2. Using the moment generating function, we can conclude this.



Therefore, the sum of normal distributions is again a normal distribution, with the mean being the summation of μ_i because they are independently distributed random variables, and the variance being n * σ^2. Hopefully, you have understood this. Now, we will discuss a new topic. First of all, what we have found is that in all of these cases, the moment generating function exists. But we discussed that the moment generating function may or may not exist.

Let us now discuss an example where the moment generating function may not exist. We will go through one such example. Let us consider this example: a discrete random variable X with a probability mass function given by this. Let us consider the discrete random variable X with the probability mass function described as follows: the probability of X taking any value is given as $6 / \pi^2 * x^2$ for values x = 1, 2, 3, and so on. For all other values, the probability is 0.



To verify if this is a valid probability mass function, first note that the probability for all valid x values is always ≥ 0, satisfying the non-negativity condition. Next, we check whether the total probability sums to 1. When summing over all possible values of x from 1 to ∞, the series converges to 1 because it is a well-known convergent series.

Therefore, this is a valid probability mass function. If the moment-generating function exists for this random variable, it can be expressed as an infinite series of moments. This concept can be analyzed further to understand its properties. So, now, if the moment does not exist, then the moment-generating function cannot exist either. To examine this, the moment-generating function is defined as the expected value of $e^{(tx)}$. This involves taking the summation over all possible values of x, starting from 1 and extending to ∞. So, we have $e^{(tx)}$ multiplied by $P_X(x)$.

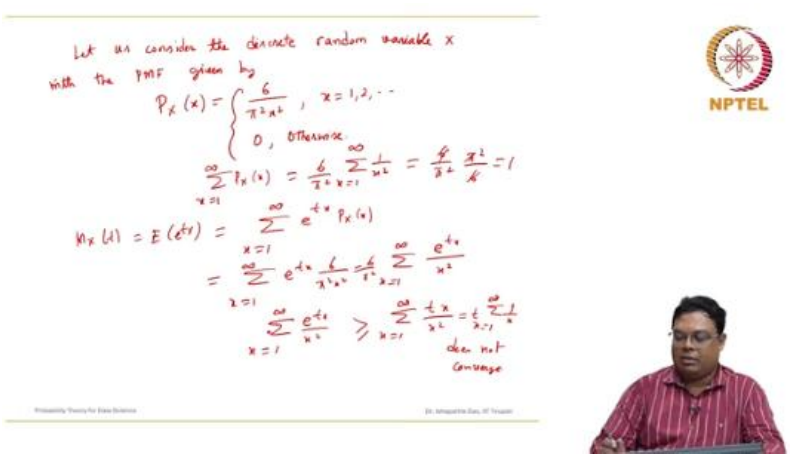
This becomes a summation starting from x = 1 up to ∞ of $e^{(tx)}$. The term tx multiplied by x is essentially $6 / \pi^2 * x^2$. Therefore, the summation is x starting from 1 to ∞ with the factor

$6 / \pi^2$ taken outside. Then, inside the summation, we have $e^{(tx)} / x^2$. So, how can we show that this may diverge?

Well, see, $1 + e^{(tx)}$ contains higher-order terms, such as $1 + tx + t^2x^2$, all of which are positive terms. Now, let's look at this sum. It's actually $\geq$ the summation from $x = 1$ to ∞ of $t * x / x^2$. This is $\geq$ the summation from $x = 1$ to ∞. So, if you take t as any positive value, you can represent tx as $1 + tx + t^2x^2 / 2!$ and so on.

So, the other terms will be positive, and we are not considering just this term. That's why this will be strictly greater than the previous sum. This sum is essentially the same as the summation from $x = 1$ to ∞ of $1 / x$, which is a divergent series when t is a non-zero value. Therefore, this is not convergent. So, this actually goes to ∞.

This sum is a divergent series and does not converge to finite values. That's why it is not convergent. So, now, in another way, we can discuss what their moments are. Basically, we consider the first-order moments, suppose μ_1'. By definition, this is the summation from $x = 1$ to ∞ of $x * P\_X(x)$.



So, this becomes the summation from $x = 1$ to ∞, $x * P\_X(x)$, which equals $6 / \pi^2 * x^2$. Since $6 / \pi^2$ is a constant, we get $6 / \pi^2$ * summation from $x = 1$ to ∞ of $1 / x$. This is a divergent series, so it does not exist finitely. So, that is why μ' does not exist, and in general,

μ_r' does not exist for any r, whether r = 1, 2, or higher. None of the first-order, second-order, or third-order moments exist in a finite value.

Hence, the moment-generating function does not exist, meaning the moment-generating function of X does not exist. So, that is one example where the moment-generating function does not exist. There may be other examples where the moment-generating function does not exist as well. Whenever the moment-generating function does not exist and you want to find a similar property, there is a different definition. This function is known as the characteristic function.

The characteristic function of a random variable X is defined similarly to the moment-generating function, with a slight modification. It is denoted as $\psi_X(t)$. Instead of multiplying by t, we multiply by j, where j is the complex number, the square root of -1. So, what is the benefit here? If we take the expected value of e^{jtx}, for example, let us consider a discrete random variable. In this case, the definition becomes the summation of all $x_k * e^{jtx_k}$, where $P_X(x) = P_X(x_k)$, whenever X is a discrete random variable.

This is equal to the integral from $-\infty$ to $+\infty$ of $e^{jtx} * f_X(x)\,dx$, when X is a continuous random variable. So, now, why is this beneficial? If we add $-1 * \sqrt{(-1)}$ instead of t, it's because the moment-generating function may not exist for all random variables. What about $\psi_X(t)$? If we consider this for a continuous random variable, the discrete case will be very similar.

So, if you consider $\psi_X(t)$, the absolute value of $\psi_X(t)$ is nothing but the absolute value of the integral from $-\infty$ to $+\infty$ of $e^{jtx} * f_X(x)\,dx$. If you take the absolute value inside, then e^{jtx} and $f_X(x)$ are already positive numbers, always ≥ 0. So, the absolute value will be the same as this. So, now, what is e^{jtx}? This represents the modulus of a complex number.

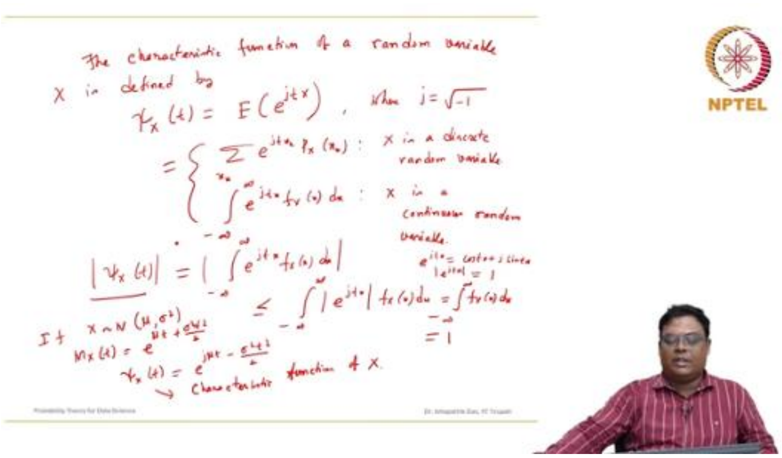
e^{jtx} is essentially the sum of $\cos(tx) + j * \sin(tx)$. Now, what is the modulus of this? It is $\sqrt{[\cos^2(tx) + \sin^2(tx)]}$, and the square root of that equals 1. So, $\cos^2(tx) + \sin^2(tx) = 1$. The square root of 1 is 1.

This value is nothing but the integral from $-\infty$ to $+\infty$ of $f_X(x)\,dx$. Since it is a density function, this value will equal 1. Therefore, the modulus of $\psi_X(t)$ will be ≤ 1. This means that it will always be convergent. It will not be ∞ or $-\infty$; it will have some finite value.

Hence, it always exists. So, other properties are very similar. Also, whenever the moment-generating function exists, if you replace t with jt, you will obtain the characteristic function. For example, in the case where X follows a normal distribution with mean 0 and variance 1, or in a more general case with mean μ and variance σ^2, the moment-generating function is simply $\mu * t + (1/2) * \sigma^2 * t^2$. So, then for $\psi_X(t)$, if you replace t with jt, it will become $e^{(j\mu t)} * e^{(-\sigma^2 * t^2 / 2)}$, because $j^2 = -1$.

So, this is the characteristic function of X. We have already discussed the moment-generating function in detail, and when the moment-generating function exists, you can simply replace t with jt to find the characteristic function of X. The properties of the characteristic function are very similar to those of the moment-generating function. If the moment-generating function does not exist, we can still use the characteristic function to find the properties, which is one of the advantages of using the characteristic function.

So, this is the concept of the moment-generating function and some of the numerical examples, their properties, and how the moment-generating function can be utilized to find the unknown distribution, especially for the transformation of a random variable and finding the distribution.



How the moment generating function is used in that process, we have also discussed. So, hopefully, you have followed and understood it. In the references, there are many numerical examples you can go through for moment generating functions and the transformation of random variables to gain better clarity. Next, we will discuss another topic: an important inequality known as Chebyshev's inequality.