PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

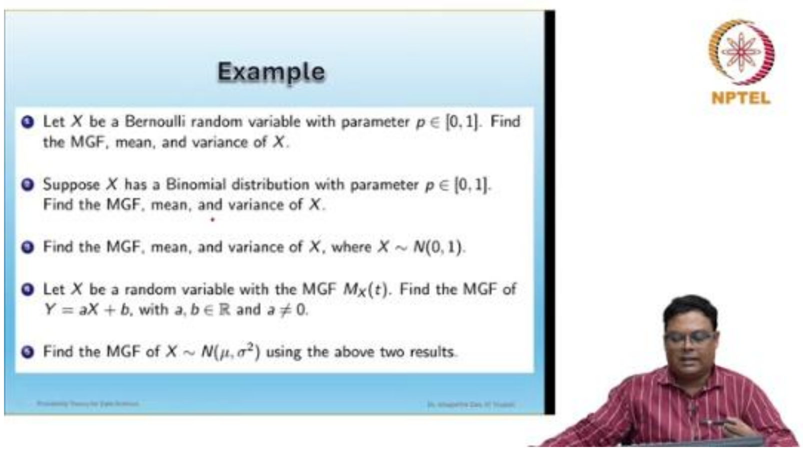
Indian Institute of Technology Tirupati

Week - 12

Lecture - 60

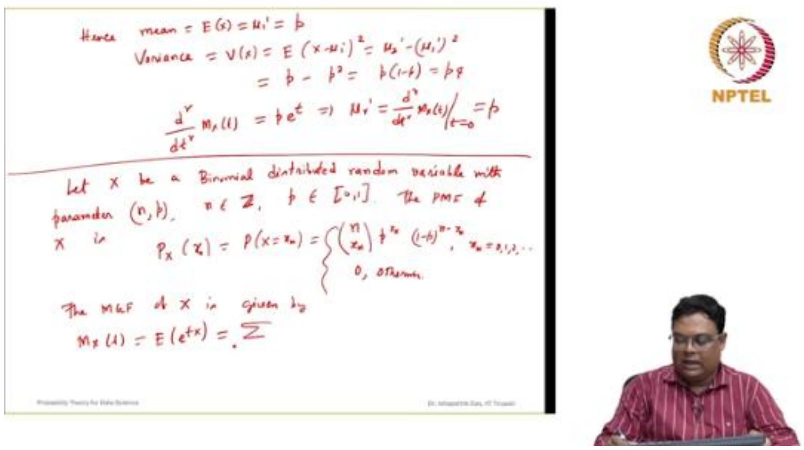
Example of Moment Generating Functions

Next, we will discuss the binomial distribution, which is an extension of the Bernoulli distribution. In this case, we will find the moment-generating function, mean, and variance of the binomial distribution. Let X be a binomial random variable with parameter p, where p ∈ [0, 1]. We will now determine the moment-generating function, mean, and variance of X. To find the moment-generating function (MGF) of a binomial distribution, let's go through the process step by step.



Let X be a binomially distributed random variable with parameters n and p. Here, n ∈ ℕ (n ≥ 1) representing the number of trials, and p ∈ [0, 1]. The probability mass function (PMF) of X gives the probability of X taking a particular value, and it is determined by the binomial distribution formula. The value of X can range from 0 to n, meaning it can take values such as 0, 1, 2, ..., n.

Now, the moment-generating function of X is defined as the expected value of e^(tX). For discrete random variables, this is calculated by taking the sum of e^(t * x_k), each multiplied by P(X = x_k). You sum over all the possible values that X can take, which range from 0 to n in the case of the binomial distribution. This sum gives the moment-generating function for the binomial distribution. If the random variable X can take values from 0 to n, we sum over all possible values of X, which range from 0 to n. The moment-generating function (MGF) is calculated by taking the expected value of e^(tX), where each possible value of X is weighted by its corresponding probability from the binomial distribution.



To simplify the expression, we observe that the sum is similar to a binomial expansion. Using the binomial expansion formula, we can rewrite the sum in a more compact form. This allows us to express the moment-generating function as a single expression involving the parameters p and t. The resulting moment-generating function is the sum of terms involving p raised to different powers, and it can be simplified into a form similar to a binomial expansion. After performing this simplification, we obtain the moment-generating function as a function of p, t, and n.

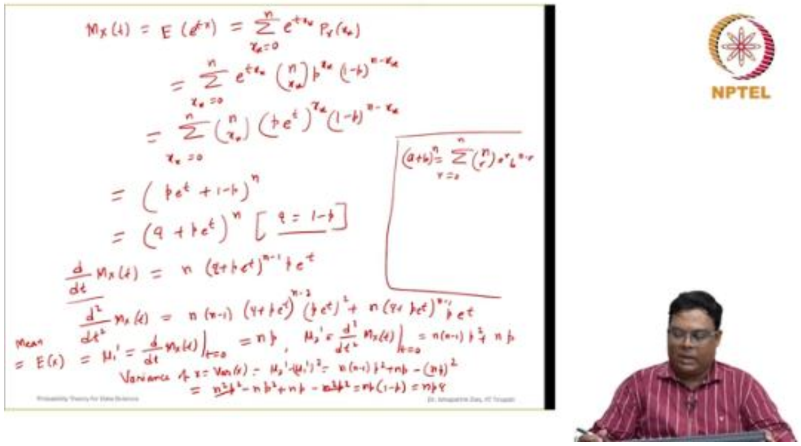
Now, to find the moments of the distribution, we differentiate the moment-generating function with respect to t. Differentiating the MGF will give us a new function, and by evaluating this function at t = 0, we can obtain the first moment or mean of the distribution. Similarly, higher-order derivatives can give us higher moments, such as the variance. Next, suppose we want to differentiate the moment-generating function again, specifically the second derivative. When we differentiate the moment-generating function for the second time, the result involves multiplying terms that include the parameters from the binomial distribution.

For the second derivative, the terms are adjusted by the power of p and the exponential function $e^{(t)}$, with their respective derivatives. After simplifying these terms, we can express them in terms of the moment-generating function, and then evaluate at $t = 0$. When we evaluate the first derivative at $t = 0$, we obtain the mean of the binomial distribution, which is given by $n * p$. For the second derivative, when we evaluate at $t = 0$, we obtain a value that corresponds to the second moment, which involves p^2. This helps us calculate the second moment, and from this, we can compute other statistical properties such as the variance.

Thus, after differentiating and evaluating at $t = 0$, we can derive both the mean (1st moment) and the second moment (which helps in computing the variance). The expressions simplify to terms involving the parameter p, and the final results give us the moments of the binomial distribution. After taking the second derivative of the moment-generating function, we obtain the second moment, which is needed to calculate the variance. The second derivative involves terms with the parameters of the binomial distribution. Simplifying the expression, we can find the second moment.

The mean is found by evaluating the 1st derivative of the moment-generating function at $t = 0$, and the second moment is similarly found by evaluating the 2nd derivative at $t = 0$. With both moments, we can calculate the variance by subtracting the square of the mean from the second moment. After simplification, we find that the variance of a binomial distribution is $n * p * (1 - p)$. Calculating the variance directly from its definition can be more complicated, as it involves finding the second moment and performing additional steps. However, using the moment-generating function makes this process simpler.

By differentiating the moment-generating function and evaluating it at $t = 0$, we can easily determine both the 1st and 2nd moments and subsequently calculate the mean and variance. In conclusion, using the moment-generating function simplifies the calculation of both the mean and variance of the binomial distribution. Now, we will discuss another example. We previously discussed discrete cases. Let's consider a continuous random variable and find the moment-generating function, mean, and variance of X, where X is a standard normal random variable.



Let X be a normal (0, 1) random variable. The probability density function (PDF) of X is given by the formula, as it is a continuous random variable. To find the moment-generating function (MGF), we use the probability density function. The probability density function of X is equal to $1 / \sqrt{2\pi} * e^{-X^2 / 2}$. This is valid for X ranging from $-\infty$ to $+\infty$.

To find the moment-generating function, we need to calculate the expected value of e^{tX}. The moment-generating function of X, denoted as $M_X(t)$, is defined as the expected value of e^{tX}. This expression becomes the integral from $-\infty$ to $+\infty$ of $e^{tX} * 1 / \sqrt{2\pi} * e^{-X^2 / 2}$. Our goal is to evaluate this integral. We can treat $1 / \sqrt{2\pi}$ as a constant factor outside the integral.

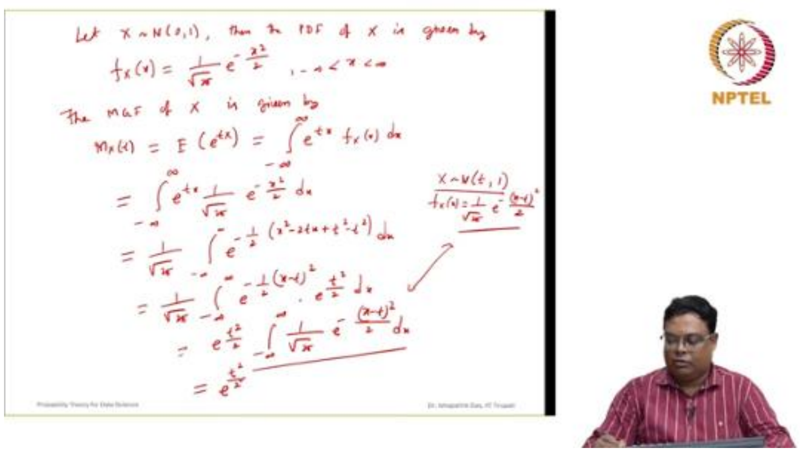
Now, we need to focus on solving the integral part, which involves the exponential terms. Now, we have the expression where we complete the square by rewriting the term as $X^2 - 2tX$. To make it a perfect square, we add and subtract t^2. This simplifies the expression, and now we have the form we can use for integration. So, we can rewrite the expression as $1 / \sqrt{2\pi} * e^{-(1/2) * (X - t)^2}$.

The additional term involving t^2 becomes $e^{t^2 / 2}$, and this can be treated as a constant during the integration process. So, the integral from $-\infty$ to $+\infty$ of $1 / \sqrt{2\pi} * e^{-(X - t)^2 / 2}$ dx, now what will be the value of this function integration? So, note that if you have, suppose instead of normal (0, 1), suppose if you have normal (t), where t is any real number, so t1. So then, what will be the density function?

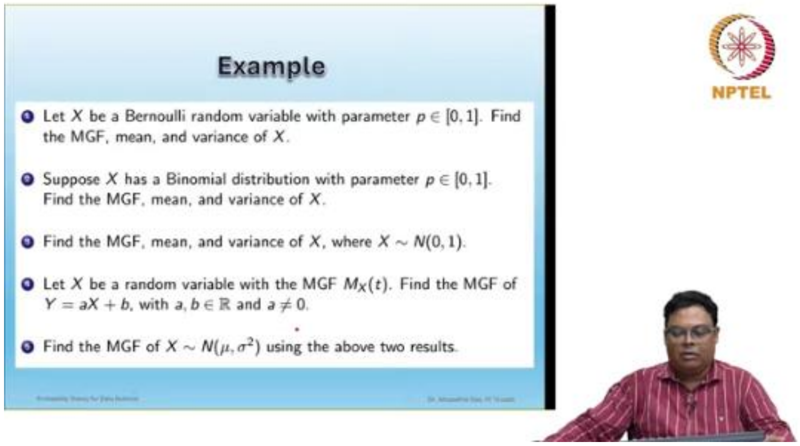
So, the density function of $f_X(x)$ will be $1 / \sqrt{2\pi} * e^{-(X - \mu)^2 / 2\sigma^2}$. This is the mean μ and variance σ^2.

So, because if X is a normal distribution with mean μ and variance σ^2, and this is the density function, then this integration from $-\infty$ to $+\infty$, because it is a density function of a normal distribution, will be equal to 1. So, here you can see that this is exactly the same density function. So, that is why this will be equal to 1. Finally, what we got from using this property is that, as we have already shown when we discussed the normal distribution, for any mean μ and variance σ^2, whenever you use a density function, the integration will be equal to 1. We have already demonstrated this.

Therefore, instead of computing this integral again, we will use this property of the density function, and that is why we conclude that this integral is equal to 1. So, we simplify this, and this becomes the moment-generating function of a standard normal random variable. Now, we need to find the mean and variance of this random variable. As you might recall, when we discussed the normal distribution, the computation of mean and variance was a bit tricky. Now, let's proceed with what we will do next.



We know that the moment-generating function (MGF) of a standard normal random variable, where X is distributed as normal (0, 1), is given by $e^{(t^2 / 2)}$. This is the moment-generating function. To find the mean, we first differentiate the moment-generating function with respect to t. The moment-generating function is $e^{(t^2 / 2)}$. Differentiating this with respect to t, we get $t * e^{(t^2 / 2)}$. Next, if we differentiate again, we get $e^{(t^2 / 2)} + t * e^{(t^2 / 2)} * (2t / 2)$.



Simplifying this, we get $e^{(t^2 / 2)} + e^{(t^2 / 2)} * (1 + t^2)$. So, the second-order moment comes from the second derivative of the moment-generating function. From that, we can find the first-order moment, which is the expected value of X. To do this, we differentiate the moment-generating function with respect to t and then evaluate it at t = 0. This process gives us the mean or expected value of X.

So, we evaluate these things at t = 0. When t = 0, the result is 0, which gives the mean of a standard normal variate as 0. Now, for the second-order raw moment, denoted as μ_2' (mu_2'), which is the expected value of X^2, we find it by taking the second derivative of the moment-generating function and then evaluating it at t = 0. After doing this, we find that the second-order moment is 1. This is because when we substitute t = 0, the t^2 term becomes 0, and we are left with the value of 1.

So, this becomes 1. Therefore, the variance of X, denoted as σ^2 (sigma squared), is calculated as the second moment (μ_2') minus the square of the first moment (μ_1'). Since $\mu_1' = 0$, the variance simplifies to just μ_2', which is 1. Hence, the mean of the standard normal variate is 0, and the variance is 1. So, this is very simple. Once we know the moment-generating function, it becomes easy to compute the mean and variance of the random variable.

This was one example. Now, let us move on to discuss some other properties of the moment-generating function. Let's look at the next property. Let X be a random variable with the moment-generating function M_X(t), and we want to find the moment-generating function of Y, where Y is a transformation of the random variable, specifically Y = aX + b. Let's proceed with finding this.

So, let X be a random variable with the moment-generating function $M_X(t)$. This is the general function. For a particular random variable, suppose X is normally distributed with mean 0 and variance 1, then the moment-generating function will be $e^{(-t^2 / 2)}$. But in general, we write $M_X(t)$, and let $Y = aX + b$, where a and b are real numbers. Here, $a \neq 0$ because if $a = 0$, Y would just be a constant, and b would be any real number.

So, a is any real number, not equal to 0. Now, to find the moment-generating function of Y, we can express it in terms of the moment-generating function of X. The moment-generating function, often abbreviated as MGF, of Y is denoted by $M_Y(t)$. So, by definition, the moment-generating function of Y, denoted as $M_Y(t)$, is nothing but the expected value of $e^{(tY)}$. So, now, Y is the expected value of $e^{(t(aX + b))}$.

This can be rewritten as the expected value of $e^{(t(aX))} * e^{(t(b))}$. Since $e^{(t(b))}$ is a real number and t is a real number, we can treat it as a constant and pull it out of the expectation. So, it will be outside, $e^{(t * b)}$, and now we have the expected value of $e^{(t * a * X)}$. Now, what is $M_X(t)$? Since $M_X(t)$ is the moment-generating function of X, it is the expected value of $e^{(t * X)}$. We assume that this moment-generating function exists for some values of t.

Thus, $M_X(t)$, a function of t, is nothing but the expected value of $e^{(t * X)}$. Hence, $e^{(t * a * X)}$, where a is a real number, can be written as $e^{(t * b)}$. So, this is $M_X(ta)$, because $M_X(t)$ means that if you replace t with ta, it becomes $e^{(t * aX)}$. That's why we have written it as $M_X(ta)$. Now, using this concept, this is the general case: If you know the moment-generating function of X, you can replace t with ta and find this function.

Then, you can directly find the moment-generating function of Y as well. So, here you can see the next example. Suppose X is a normal distribution with mean μ and variance σ^2. Using the above two results, like you know the homogeneous function of X and in this transformation, you can find the moment-generating function of X when it is normal with mean μ and variance σ^2. Now, if you take this transformation, So, let $Y = \mu + \sigma * X$, where X is a random variable.

Here, we assume that X follows a normal distribution with mean μ and variance σ^2. Alternatively, instead of X and Y, you could consider a standard normal distribution, where X follows a normal distribution with mean 0 and variance 1. If you consider $Y = \mu + \sigma * X$, we want to find the moment-generating function of Y. But what is the distribution of Y? Earlier, we discussed that any linear combination of a normal distribution, when we talked about the normal distribution, will again result in a normal distribution. So, in this case, Y will follow a normal distribution with mean μ and variance σ^2.

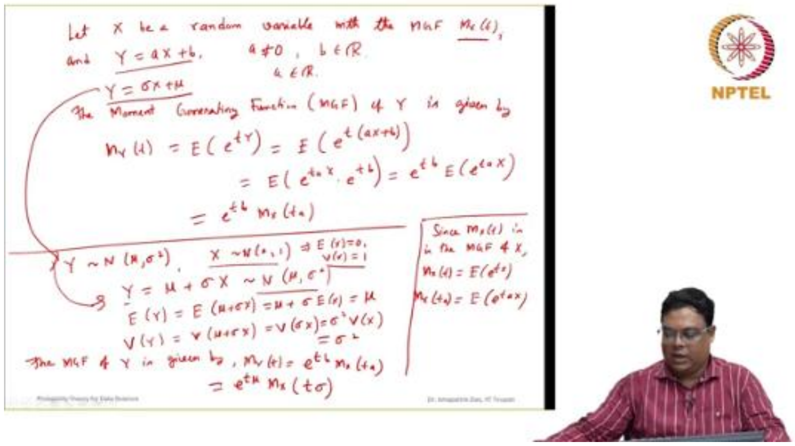
So, we know that this is a normal distribution, and we just have to find the mean and variance. So, to find the mean, we calculate the expected value of Y. The expected value of Y is the expected value of $\mu + \sigma * X$. Since μ is a constant, this becomes $\mu + \sigma * E[X]$. Now, since X is a normal distribution with mean 0, the expected value of X is 0. Therefore, the expected value of Y is simply μ.

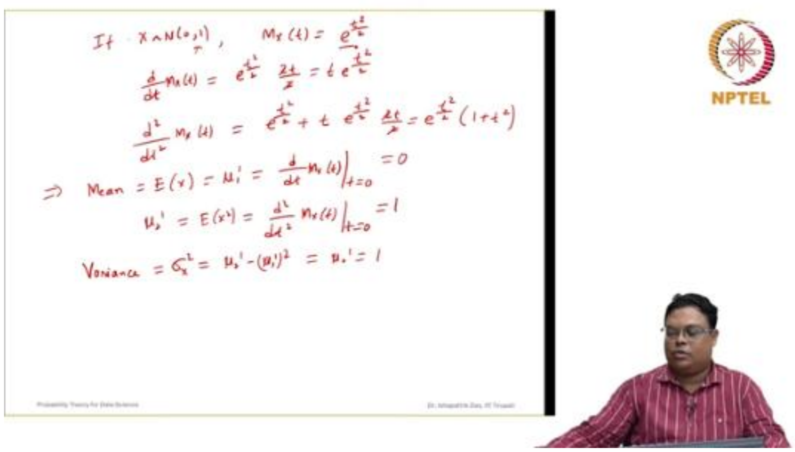
Now, for the variance of Y, the variance is the square of the coefficient of X. Since the variance of X is 1, the variance of Y will be σ^2. So, in that case, this becomes μ. Now, what will be the variance of Y? The variance of Y will be the variance of $\mu + \sigma * X$. Since σ is a constant, the property we discussed earlier applies: if you add a constant to a random variable, the variance does not change. Therefore, the variance of $\mu + \sigma * X$ is simply the variance of $\sigma * X$.

The variance of $\sigma * X$ is given by $\sigma^2 * Var(X)$. Since the variance of X is 1, this becomes σ^2. Hence, if $Y = \mu + \sigma * X$, it follows a normal distribution with mean μ and variance σ^2. Using these properties and transformations, if we know the probabilities for a standard normal variate, we can use those values from the table to find the probability of Y. So, Y will be any normal distribution, and using the transformation we discussed in earlier lectures, we can find its moment-generating function where Y is normally distributed with mean μ and variance σ^2.

Now, we know the moment-generating function of X and the property we discussed earlier. We will use this property, which tells us that the moment-generating function of Y is related to the moment-generating function of X. Specifically, if Y is a linear transformation of X, say $Y = aX + b$, then the moment-generating function of Y can be written in terms of the moment-generating function of X. In this case, the transformation is $Y = \mu + \sigma * X$, where "μ" and "σ" are constants. According to the property, the moment-generating function of Y will be the moment-generating function of X evaluated at $t * \sigma$, multiplied by an exponential factor involving $t * \mu$.

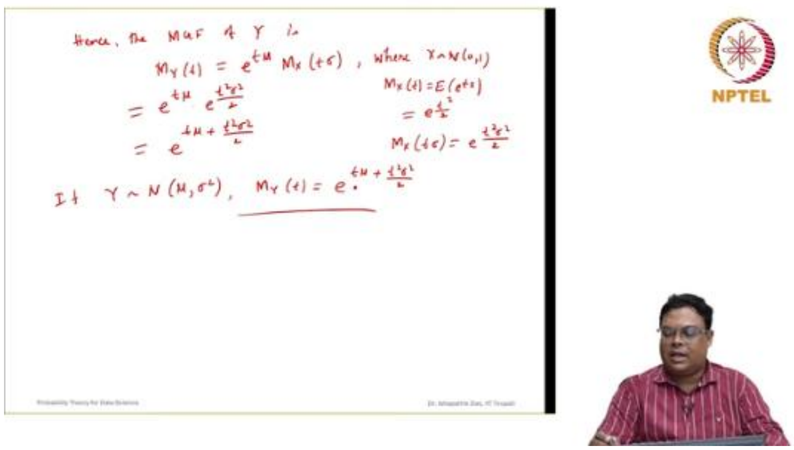
Therefore, we find that the moment-generating function of Y is equal to the moment-generating function of X evaluated at $t * \sigma$, multiplied by an exponential term involving $t * \mu$. For the standard normal distribution, we know that the moment-generating function of X is given by the expected value of $e^{(t * X)}$, which simplifies to an expression involving t^2. Using this, we conclude that the moment-generating function of Y, where Y follows a normal distribution with mean μ and variance σ^2, can be written as an exponential factor involving μ, and another exponential factor involving t^2 and σ^2. So, what we find is that the moment-generating function of Y simplifies into two exponential terms. One term involves the mean, where the exponent is t * μ.





The second term involves the variance, where the exponent has t^2 multiplied by ½ of the variance. When you combine these two terms, the result is μ * t + ½ * σ^2 * t^2. So, this is the moment-generating function of Y. Hence, we found that if Y follows a normal distribution with mean μ and variance σ^2, then the moment-generating function of Y, denoted M_Y(t), is equal to e^(t * μ + t^2 * σ^2 / 2). This is the moment-generating function of a normal distribution with mean μ and variance σ^2.

So, in particular, if the mean $\mu = 0$ and the variance $\sigma^2 = 1$, you can see that the moment-generating function simplifies to $e^{(t^2 / 2)}$, which is the moment-generating function of a standard normal random variable. But this is the more general case, and we didn't need to do the complicated integration. Instead, we used these two properties to directly find it. That's why we said to find the homogeneous function of X using the above two results, and we've completed that here. Hopefully, you're following along and understanding it.



You can practice finding the moment-generating function for different random variables such as the Poisson distribution, uniform distribution, exponential distribution, and gamma distribution. You can also find their moments and practice these tasks. Doing so will help make things clearer for you. Now, let's move on to an important theory. If two random variables have the same moment-generating function, $M_X(t) = M_Y(t)$, they must have the same distribution, $F_X(x) = F_Y(y)$. So, this is one of the very important theories.

We are not going to prove this theory, but it has significant applications and implications. If the moment-generating function exists, it cannot be different for multiple distributions. This means that if you know the moment-generating function, you can identify the random variable based on it. So, if you know the moment-generating function, you can determine whether the distribution is Poisson, normal, exponential, or any other known distribution. For example, if you are working with an unknown random variable and performing a transformation, you may not know the distribution at first.

But if you can compute the moment-generating function and find that it closely resembles the moment-generating function of a known distribution, such as Poisson, then you can confidently conclude that the distribution is Poisson. It cannot be any other distribution. Exactly, if two random variables have the same moment-generating function, $M_X(t) = M_Y(t)$, then they must have the same distribution, $F_X(x) = F_Y(y)$. They cannot have different distributions. This property is very useful because, by examining the moment-generating function, you can identify the distribution of a random variable. Now, let's explore some applications of this concept.

We have a series of random variables, say X_1, X_2, and so on, up to X_n. These are all independent random variables, and each of them has its own moment-generating function, $M_{X_1}(t)$, $M_{X_2}(t)$, ..., $M_{X_n}(t)$. What we want to do is figure out the moment-generating function of the sum of all these random variables, $M_{X_1 + X_2 + ... + X_n}(t)$. Since the random variables are independent, the moment-generating function of their sum is simply the product of the moment-generating functions of each individual random variable:

$$M_{X_1 + X_2 + ... + X_n}(t) = M_{X_1}(t) * M_{X_2}(t) * ... * M_{X_n}(t)$$

Now, there's an important theorem here called the uniqueness theorem. According to this theorem, if we have two random variables, say X and Y, and their moment-generating functions are identical for every value of t where the functions are defined, $M_X(t) = M_Y(t)$, then the two random variables must have the same distribution, $F_X(x) = F_Y(y)$. This means that their cumulative distribution functions are also identical, $F_X(x) = F_Y(y)$.

So, what this means is that if two random variables have the same moment-generating function, $M_X(t) = M_Y(t)$, they will have the same distribution function, $F_X(x) =$

$F_Y(y)$. Now, if we're talking about discrete random variables, this would mean they have the same probability mass function, $P(X = x) = P(Y = y)$. For continuous random variables, it means they have the same probability density function, $f_X(x) = f_Y(y)$. Essentially, these two random variables would be identical in terms of their distribution. This is the concept of uniqueness.