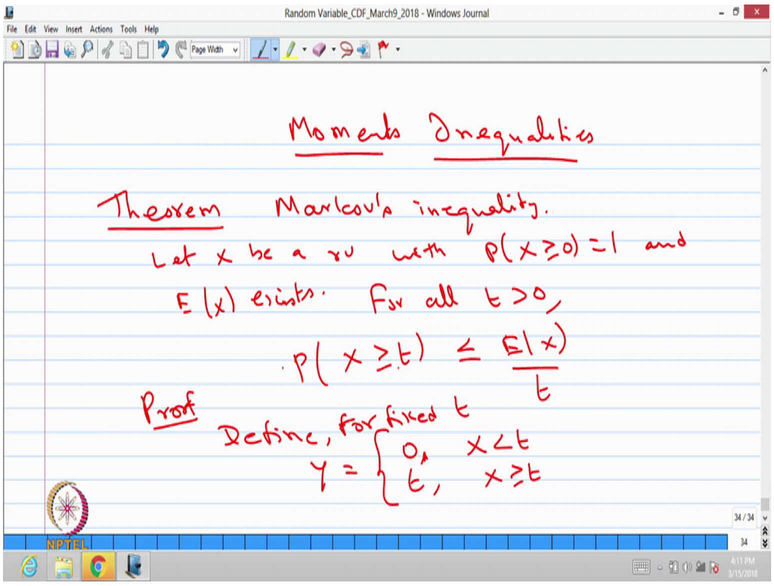
Introduction to Probability Theory and Stochastic Processes
Prof. S. Dharmaraja
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture - 15

So, even though we have done only one example for finding the nth order moment, there are many more problems we may need to find out the nth order moment when we started discussing the some standard distribution. So, for time being we will stop it with only one example of finding nth order moment, later we will do the similar problems when after we discussing the standard distributions.

(Refer Slide Time: 00:40)



So, now we are moving into the next topic that is called moments inequalities. It is a very important topic in the sense sometimes it is very difficult to find the distribution, but you may have the moments with you. That means, whenever the distribution of the random variable is not known to you, at the same time we know the fewer moments for example, first order moment or second order moment or the nth order moment.

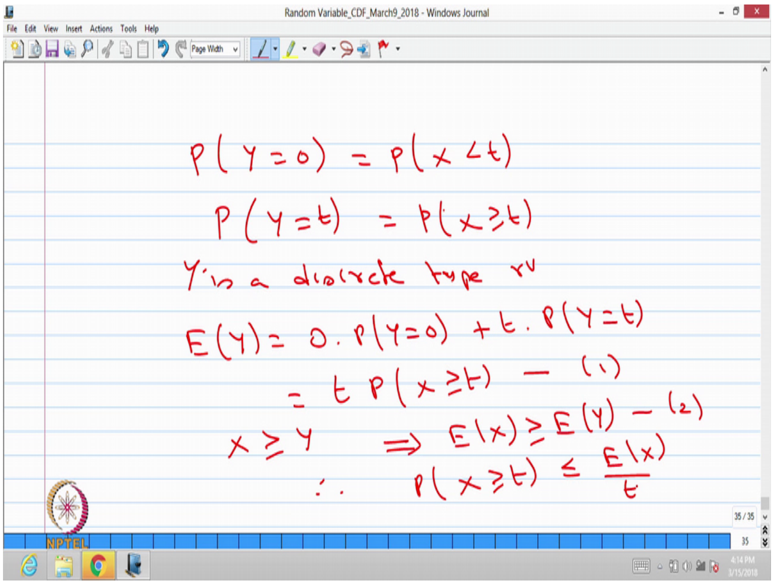
Then it is possible to find out the probability of the some events not exactly in the form of lower bound or upper bound. I am repeating the statement, when we know the distribution of the random variable. That means if it is a continuous type you know the probability density function, if it is a discrete type you know the probability mass function of the random variable you can always check whether the moments exist first. After checking the moments exist then you can find the moments either it first order moment or second order moment or any nth order moment, that is if you know the distribution of the random variable. But sometimes you may not know the distribution of the random variable you may know the moments as much as possible. That means, till the nth order that n could be 1 or 2 or whatever be the number positive integer. In that case you can find the lower bound or upper bound for the probability of some events that is what we are going to do.

So, let me start with first result as a theorem that is called Markov's inequality. What this Markov inequality says? Let X be a random variable, let X be a random variable with probability of X greater than or equal to 0 that is 1. I am saying that is a non negative random variable and the expectation of X exist. What the theorem says for all t greater than 0, the probability of X is greater than or equal to t that is always has the upper bound that is expectation of X divided by t the theorem says you do not know the distribution of the random variable which is a non negative random variable. That means the probability of X greater than or equal to 0 is 1. And you know that the expectation exist and you know the expectation value also that is E of X, the value of mean or the expectation is known to you.

Then for all t greater than 0 the probability of X is greater than or equal to t is less than or equal to expectation of X divided by t, that is nothing but the probability of right tail the right tail probability has the upper bound which is a mean divided by small t for every t, t greater than 0 this can be proved easily.

The proof is as follows I am going to define a another random variable that is Y which takes a value 0 for X is less than t and this value is going to be t when X is greater than or equal to t and this definition is for fixed t. I am defining the random variable Y for fixed t either it takes a value 0 whenever X is less than t or it takes a value small t when X is greater than or equal to t. I have not said X is a discrete type random variable or continuous type random variable or mixed type random variable, I said only it is a non negative random variable. That means, probability of X is greater than or equal to 0 is 1. The way I define a Y as a function of X it takes a value 0 or t for fixed t therefore, you can say the probability of Y is equal to 0 that is same as probability of X is lesser than t.

(Refer Slide Time: 06:29)

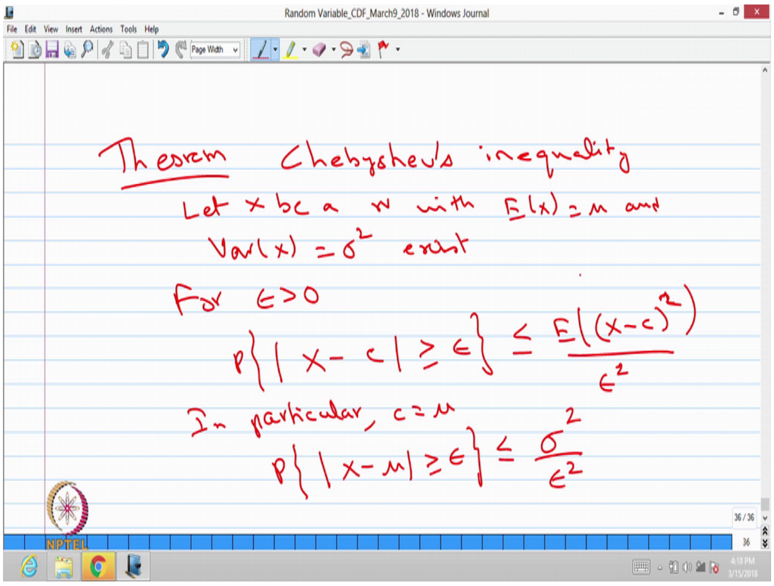


Similarly, you can say the probability of Y is same as small t that is same as probability of X is greater than or equal to t; that means, the random variable Y has a mass at 0 and small t this is for fixed t. So, in material of the random variable X is a discrete or continuous or mixed the Y is going to be a discrete type random variable, Y is a discrete type random variable with the values taken 0 and t whose probability mass function is same as probability of X is less than t and probability of X is greater than or equal to t.

Now, one can find the expectation of Y, that is nothing but since it is a discrete type random variable 0 times probability of Y is equal to 0 plus t times probability of Y is equal to t. This is same as t times probability of Y is equal to t that is same as probability of X is greater than or equal to t.

We know that X is always greater or equal to Y or fixed t X is always greater or equal to Y this implies the expectation of X is greater than or equal to expectation of y. So, I am going to use the result 1 and the result 2, expectation of Y is same as t times the probability of X is greater than or equal to t and expectation of X is greater than or equal to expectation of Y therefore, by using the result equation 1 and 2. By using equation 1 and the inequality 2 we get probability of X is greater than or equal to t, which is lesser than expectation of X divided by t. That means, the right tail probability for some fixed t which has the upper bound mean divided by the t, where t is greater than 0. So, this inequality is very important in the sense without knowing the distribution as long as you know the mean or expectation one can find the upper bound for the right tail probability.

(Refer Slide Time: 09:41)

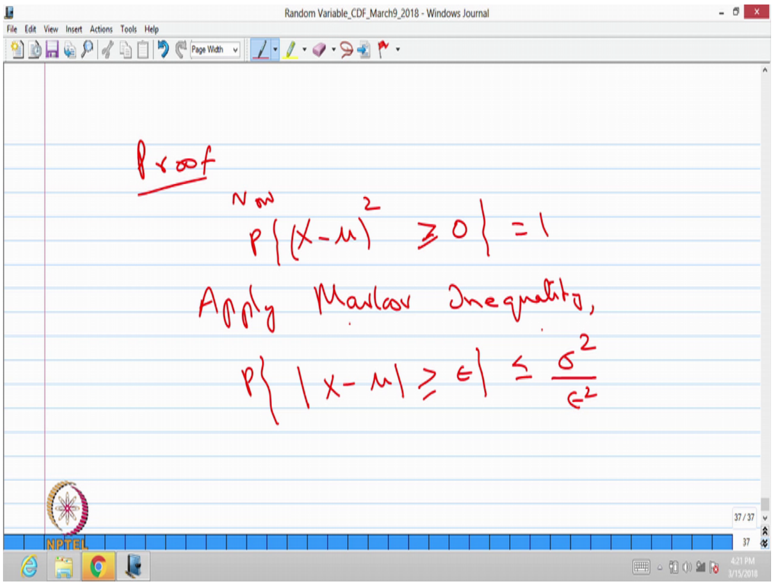


The second inequality I am going to give it as a theorem that is called Chebyshev's inequality. What this theorem says let X be a random variable with both mean and variance exists in the earlier inequality only mean exist. We are not making the restriction of second order moment exist, now we are making a variance also exist, but here we are not making a condition it is a non negative random variable. These are all the two changes with the Markov inequality and a Chebyshev's inequality.

What this theorem says for epsilon greater than 0 the probability of absolute of X minus some constant c which is greater than or equal to epsilon, this probability has the upper bound that is expectation of X minus some constant c the whole square divided by epsilon square, where c is a constant. In particular the c is same as mean of the random variable mu, then the above result probability of absolute of X minus mu greater than or equal to epsilon less than or equal to when c becomes a mu that is a mean or expectation of the random variable X the expectation of X minus mu whole square that is nothing but the variance of the random variable X that is sigma that is same as sigma square divided by epsilon square.

The previous result is a right tail probability whereas, this one is the tail probabilities absolute of X minus mu greater than or equal to epsilon that means, it is a tail probabilities has the upper bound sigma square by epsilon square for epsilon greater than 0. This also can be proved easily.

(Refer Slide Time: 12:43)

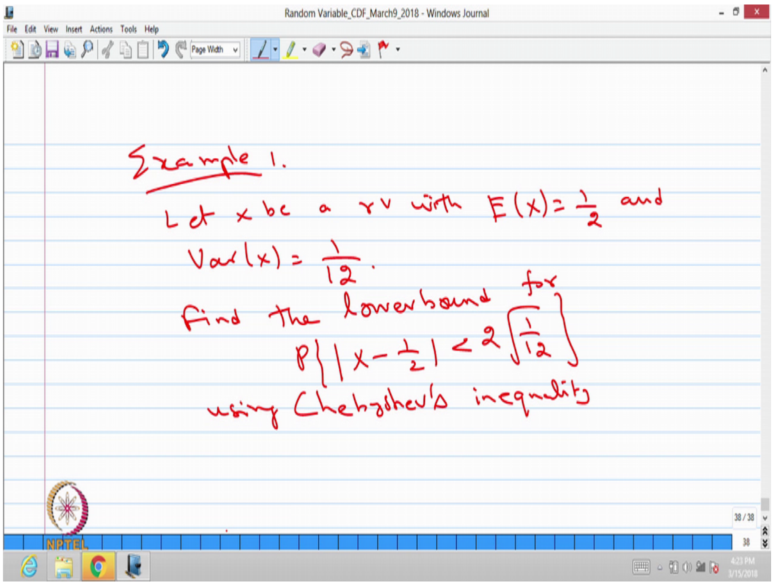


Proof: I can use the result of Markov inequality to prove the Chebyshev's inequality that is a X is a random variable whose mean and the variance exist. Now I can use X minus the mu whole square is a random variable not only it is a random variable X minus mu whole square is a non negative random, variable immaterial of a the random variable x. So, X minus mu the whole square which is going to be greater than or equal to 0 whose probability is going to be 1. The probability of X minus mu whole square greater than or equal to 0 is 1 that is a non negative random variable. Therefore, I can apply the Markov inequality because I treat X minus mu whole square as a random variable, whose mean is exists and since the variance of X is exist that is same as mean of X minus mu whole square.

Therefore, now I can apply Markov inequality because X minus mu whole square is greater than or equal to 0 whose probability is 1 and expectation of X minus mu whole square that is same as variance of the X that is also exist therefore, this is going to be probability of modulus of X minus mu greater than or equal to epsilon that is going to be less than or equal to sigma square by epsilon square. I am skipping in between two steps you a play first X minus mu whole square and take the absolute; therefore, you will get for epsilon greater than 0 probability of absolute of X minus mu greater than or equal to epsilon which has the upper bound sigma square waves epsilon square by using the Markov inequality.

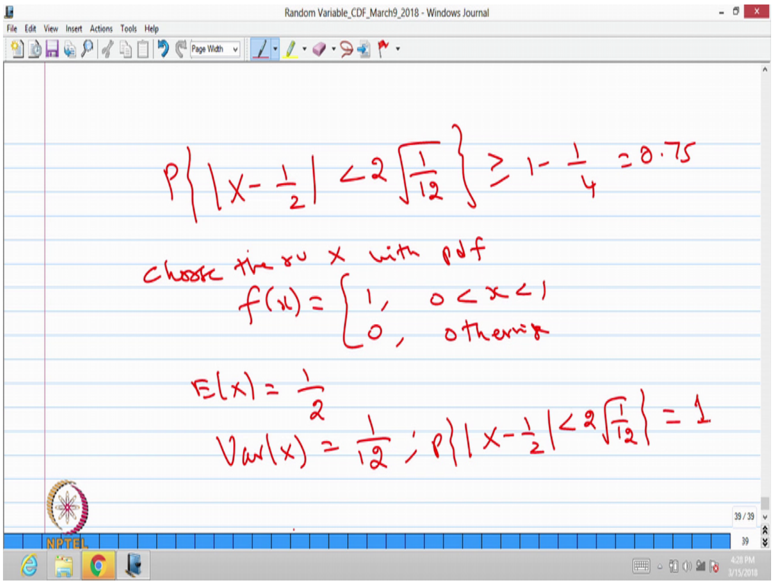
Now, we will go for one simple example how one can apply the Chebyshev's inequality.

(Refer Slide Time: 15:38)



That is example 1, let X be a random variable with mean of the random variable is 1 by 2 and variance is 1 divided by 12. Let us find the lower bound for the event probability of the event that is absolute of X minus 1 by 2 less than or less than 2 times square root of 1 by 12. Find the lower bound for probability of this using Chebyshev's inequality, using Chebyshev's inequality find the lower bound for the probability absolute of X minus 1 by 2 less than 2 square root 1 by 12. Because we have separate random variable mean and variance you do not know the distribution of this random variable.

(Refer Slide Time: 17:34)



One can find the lower bound by using the Chebyshev's inequality. That is probability of absolute of X minus 1 by 2, 1 by 2 is a mean less than 2 times square root of 1 by 12 which is greater than or equal to. I have given the Chebyshev's inequality for the tail probability and this is the probability of negation of tail probability therefore, the inequality changes and the value becomes 1 minus of variance divided by epsilon square here the epsilon is a 2 times square root of 1 by 12. So, if you do the simplification you will get 1 minus 1 by 4 that is nothing but 0.75.

So, what this Chebyshev's inequality say is the negation of tail probability has a lower limit 0.75. And you know that the probability lies between 0 to 1. The lower bound 0.75 means if you choose the distribution or if you know the distribution of the random variable X, the exact probability will be lies between 0.75 to 1, whereas without knowing the distribution of the random variable X by knowing the mean is 1 by 2 and the variance is 1 by 12, one can conclude the lower bound for this probability is 0.75.

Let us fit some distribution for this random variable X whose mean is going to be 1 by 2 and the variance is going to be 1 by 12. Then we will find out what is the exact probability of this event. Let me repeat by using a Chebyshev's inequality without knowing the distribution we got the lower bound for this probability of the event that is 0.75. Suppose we choose the random variable X with some distribution in which the mean is going to be 1 by 2 and the variance is going to be 1 by 12, one can find the exact probability of this event, ok.

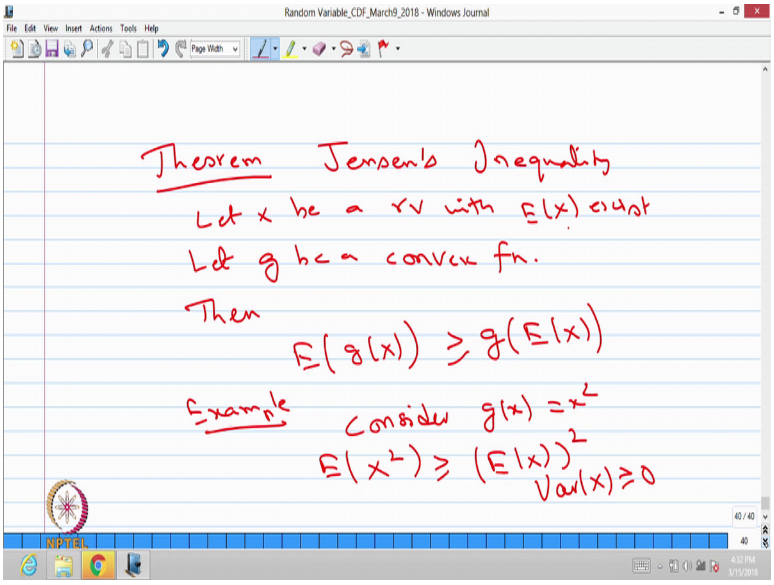
So, let us choose the random variable the distribution of the random variable X such a way that the mean is going to be 1 by 2 and the variance is going to be 1 by 12. You can make out the mean the probability density function of this random variable is 1, between 0 to 1 and 0 otherwise choose the random variable X with probability density function. That means, I am choosing a continuous type random variable with the probability density function that is f of x is a 1 between the interval 0 to 1 otherwise 0.

You can verify if we find the mean for this random variable that is from minus infinity to infinity X times f of x dx, then substitute f of x then integrate between 0 to 1 X therefore, you will get the answer 1 by 2. Similarly if you find out the variance of X for that either you can do expectation of X minus 1 by 2 whole square or you find out E of X square, you know already e of X then variance of X is E of X square minus E of X whole square either way you do you can get the answer variance of X is going to be 1 by 12. Therefore, it matches with the given problem the random variable X if you choose the probability density function these you will get the mean is 1 by 2 and the variance is 1 by 12.

Now, you can find the same event probability absolute of X minus 1 by 2 which is less than 2 times square root of 1 by 12, if you find out the probability of this event which is going to be we can do the simplification and so on you can get the answer 1. That means, exact probabilities is going to be 1 by choosing a random variable whose probability density function is this whereas, without knowing the probability density function oh sorry without knowing the distribution of the random variable you are getting the lower bound for the probability that is 0.75. Instead of this distribution if you would have chosen some other distribution you may get some other value which is also lies between 0.75 to 1 both are inclusive. So, the Markov inequality and Chebyshev's inequality are very useful to get the lower bound or upper bound for the probability of some events whenever the distribution of the random variable is unknown. If you know the distribution you can get the exact probability.

Now, we will move into the next inequality that is called Jensen's inequality as a theorem, Jensen's inequality.

(Refer Slide Time: 23:21)

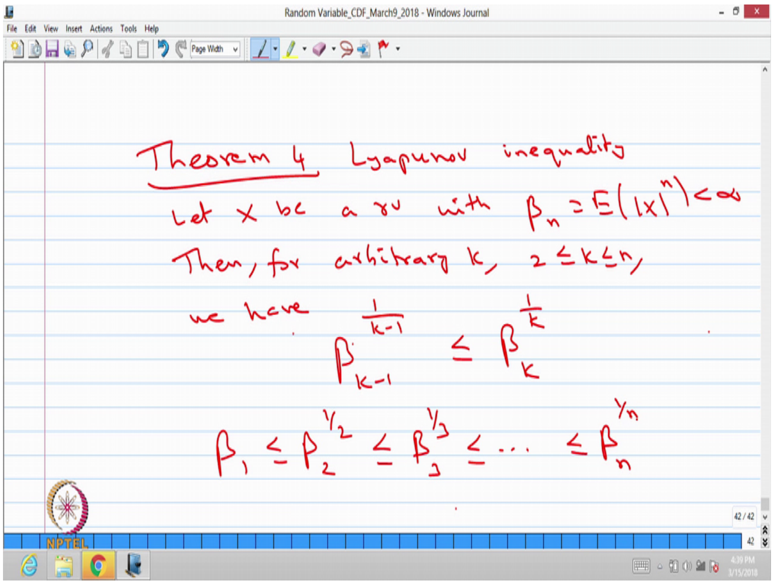


What this inequality says let X be a random variable with the expectation of X exist, let g be a convex function then what the theorem says the expectation of g of X is always greater than or equal to g of expectation of X. The Jensen's inequality says whenever the random variable whose expectation exist for a convex function g, expectation of g of X is always greater than or equal to g of expectation of X there are many real world problems one can use this inequality to get the bound for expectation whenever you interchange the expectation and the g interchanged.

Let us give a simple example for this. I am not going for the proof of this theorem, but you can go for the example. Consider g of x as x square consider g of x as x square, you apply the theorem here that means, expectation of X square greater than or equal to expectation of X the wholes square. This you know the result because this is nothing but variance of X that is always greater than or equal to 0. This implies this greater or equal to 0 that is same as variance of X is always greater than or equal to 0. So, like that you can think of any convex function to apply the Jensen's inequality.

Now, I am going for one more inequality of the nth order. The Markov inequality involves the first order, Chebyshev's inequality involves the second order, and the Jensen's inequality involves any convex function with the first order. Now, we are going for nth order moments in the form of inequality.

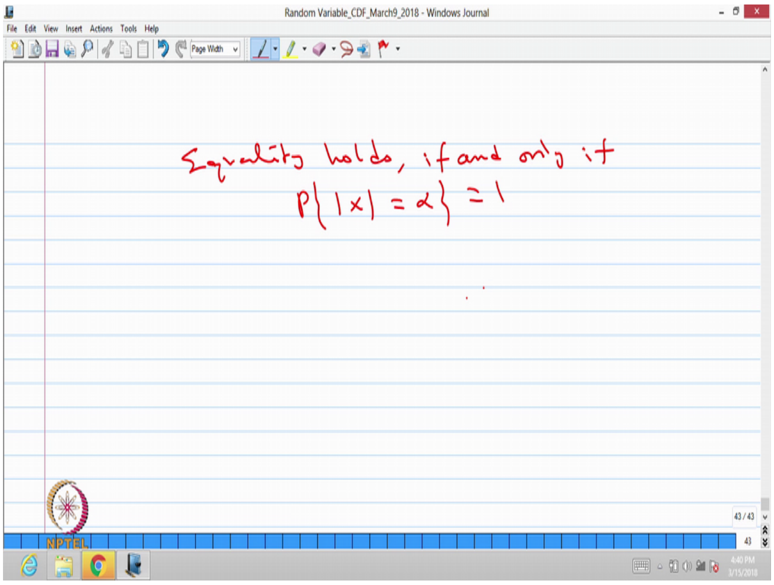
(Refer Slide Time: 26:54)



We will move into the 4th inequality as a theorem 4 it is called Lyapunov inequality. What this inequality says let X be a random variable with beta suffix n that is nothing but expectation of absolute of X power n which is a finite quantity that I denoted as a beta suffix n which exist then for arbitrary k, where k is lies between 2 to n. This theorem says the beta suffix k minus 1 power 1 divided by k minus 1 that is always less than or equal to beta suffix k power 1 divided by k.

So, this is valid whenever the beta n exist that is expectation of absolute of X power n exist for any n, n can be 1 2 3 and so on. So, if any beta suffix n exist then this inequality holds for k lies between 2 to n. That means, I am not going for the proof of this theorem; that means, the beta one always less than or equal to beta 2 power 1 by 2, that is always less than or equal to beta suffix 3 of power 1 by 3 and so on till less than or equal to beta suffix n power 1 by n.

(Refer Slide Time: 29:14)



So, the equality holds, equality holds if and only if the probability of absolute of X take some constant value that is going to be 1; equality holds if and only if. So, I am not going to give the proof of this theorem whereas, we have given the proof of Markov inequality and Chebyshev's inequality.

In the later section we are going to include, we are going to solve some problems related to the Chebyshev's inequality and Markov inequality.