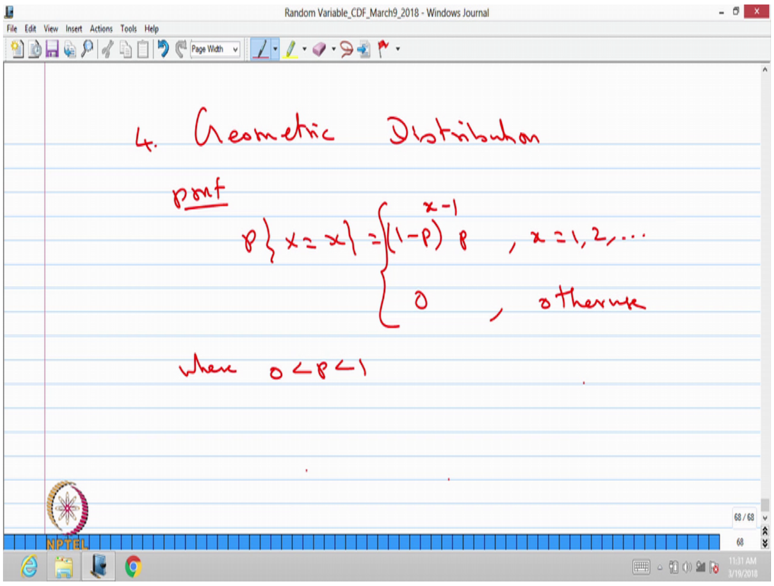
**Introduction to Probability Theory and Stochastic Processes**
**Prof. S. Dharmaraja**
**Department of Mathematics**
**Indian Institute of Technology, Delhi**

**Lecture – 19**

So, till now we have discussed a constant random variable and Bernoulli random variable, binomial random variable or binomial distribution.

(Refer Slide Time: 00:15)




Now, we are moving into fourth one, that is geometric distribution. This is also common discrete type distribution, whenever a random variable which is a discrete type random variable whose probability mass function is of the form, probability of x takes the value small x, that is 1 minus p power x minus 1 times p, where x takes the value 1 2 and so on otherwise, it is 0.

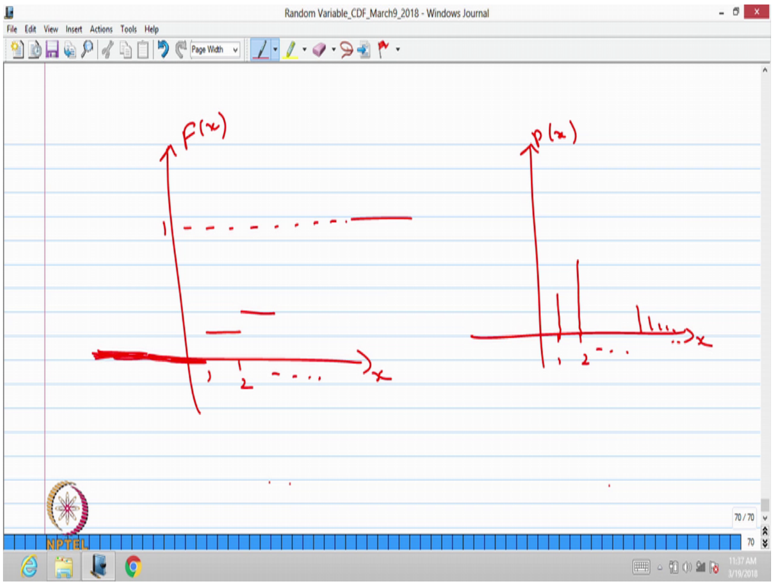
Then we call this random variable x is a geometry distributed random variable here also, the p lies between 0 to 1. There is the connection between Bernoulli distribution with the geometric distribution, that connection is a whenever you have a Bernoulli trials, the occurrence of the first trial in which you get the success that follows geometric distribution.

Suppose you have a random experiment with infinitely many Bernoulli trials in it, and each Bernoulli trial has a random variable, which is a Bernoulli distributed random variable with the probability of success p. This capital X is nothing but the trial in which you are getting the first success, that probability is you are not getting the success x minus 1 times, and the xth trial you are getting the first success therefore, it is 1 minus p power x minus 1 into p all are consecutive, and all the Bernoulli trials are independent.

So, whenever you have a n independent or sequence of a independent Bernoulli trials, the first success in the nth trial that becomes the geometric distribution. So, the difference between Bernoulli binomial and geometric, the Bernoulli distribution has a only 2 jumps the CDF has only 2 jumps, and the binomial distribution has a only n plus 1 jumps, the geometric distribution has a countably infinite jumps.

So, this is also discrete type random variable. So, let us discuss the CDF and the probability mass function of this random variable geometrative distributor.

(Refer Slide Time: 03:01)



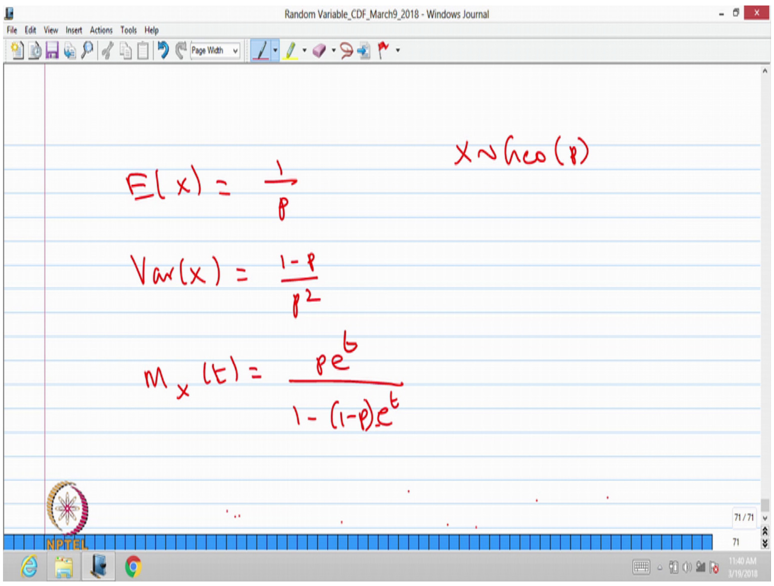

So, it has the jump points 1 2 and so on, therefore, at till x is equal to 1, it has the value 0 at x is equal to 1, it has a first jump at x is equal to 2, it has a second jump and so on. At infinity it touches 1, at infinity it touches a 1.

So that means, of the CDF has a countably infinite jumps with the jump points 1 2 and so on, and jump values are 1 minus p power x minus 1 into p. And if you draw the probability mass function at x is equal to 1, it has some height, at x is equal to 2, it has some other height and so on. Then it will be keep decreasing, then it land up at countably infinite points if you had all the heights that is going to be 1.

So, this is a probability mass function, and this is the CDF of the geometric distribution. The way I have explained through the data, suppose you have a data with the cumulative distribution, it is keep increasing at countably infinite number of points, land up to be a some finite value. Or the probability are the histogram of the data that has a some heights keep increasing and going down, and it has a countably infinite points in which is it has this values, then you can conclude the data could follows a geometric distribution.

So, in the statistics, we get this type of graphs first from the data, in the probability theory course, we started with the probability mass function then the CDF and so on, in a theoretical way we study where as in the statistics we start from the data, then we conclude what could be the distribution of those data. So, one can discuss the mean variance and the MGF for this geometric distribution also.

(Refer Slide Time: 05:33)

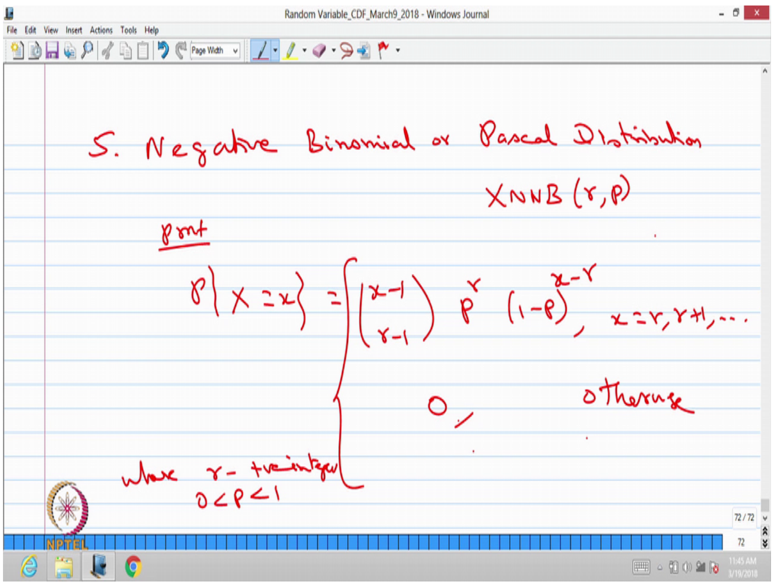



So, the mean for geometric distribution is going to be 1 divided by p, and the variance of a geometric distribution is going to be 1 minus p divided by p square, and one can get MGF of geometric distribution; that is p times e power t divided by 1 minus 1 minus p times e power t.

So, in notation we are use x follows geometric with the parameter, when we say x tilde a the geo within bracket p; that means, this is a geometric distribution with the parameter p, whose probability mass function is 1 minus p power x minus 1 into p where x takes a value one and so on. You can always create a another random variable in which the probability mass function start from 0 onwards instead of one onwards, then that random variable is call it as a modified geometric distribution. In the real world problem sometimes you come across the possible values are 1 2 and so on, or sometimes at the values start from 0 1 2 and so on.

So, you can use the correct probability mass function so, that summation is one. I am not going for the derivation the same derivation what we have done it for the binomial you can use the same thing.

(Refer Slide Time: 07:20)



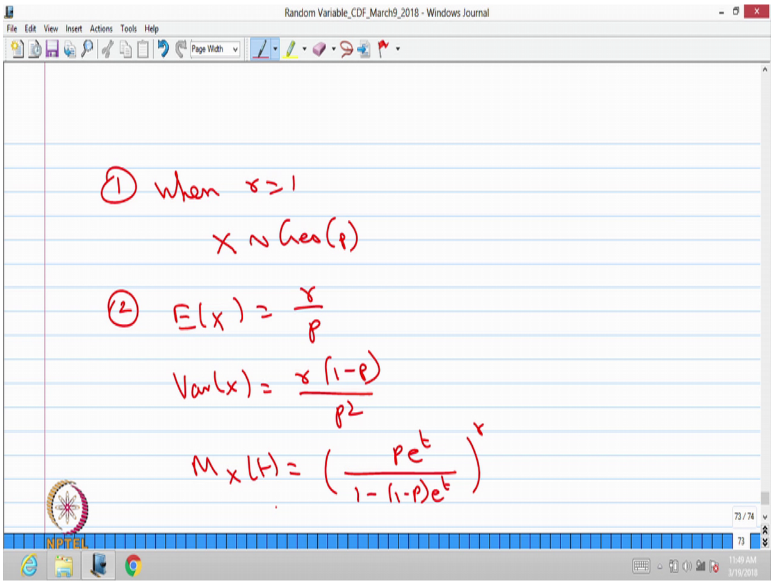

Now, we will move into the fifth one, that is negative binomial there is another name for this distribution, that is called Pascal distribution, a discrete type random variable is said to be a negative binomial or Pascal distribution. Whenever the probability mass function is of the form probability of X takes the value small x is x minus 1 see r minus 1 multiplied by p power r and 1 minus p power x minus r; where x takes the value r r plus 1 and so on, otherwise it is 0.

Here r is positive integer, and p lies between 0 to 1; that means, whenever you supply the value of r and p, you know the distribution of a this random variable. We use a notation

X follows a negative binomial NB, if the parameters r come up p, this is also related to the Bernoulli distributed random variable in the form of capital X denotes in the x trail, we are getting first time rth success follows negative binomial, whenever each trails are Bernoulli and they are independent.

So, whenever you have a independent Bernoulli trials, obtaining first time rth success that follows a negative binomial distribution with the probability of success is small p, and the probability of failure is 1 minus p.

(Refer Slide Time: 09:49)



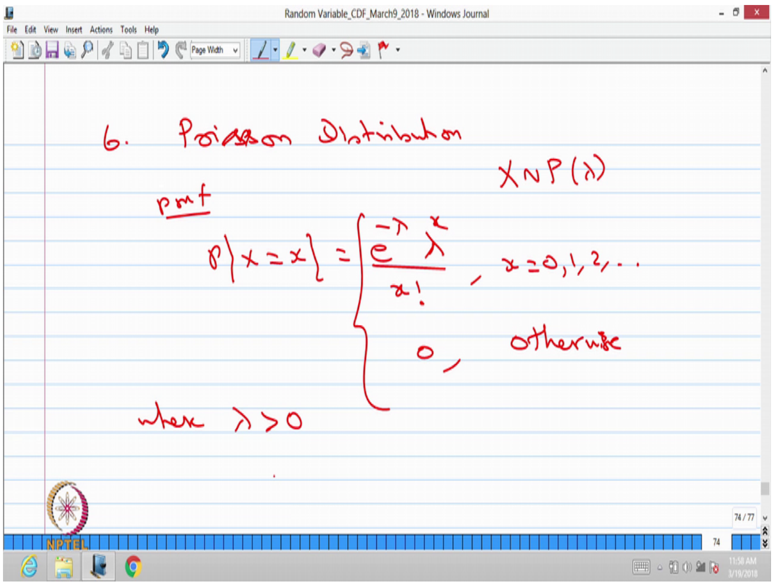

When r is equal to 1 the same random variable x follows a geometric distribution with a parameter p. When the rth success that is when it is a first success in the xth trail, then that follows a geometric distribution. Therefore, geometric distribution is a special case of negative binomial or Pascal distribution with the parameter r is equal to 1.

The probability mass function can be visualized once you are getting r minus 1 success out of x minus 1 bernoulli trials, that follows a binomial distribution followed by the rth success; that means, x minus 1 see r minus 1 p power r minus 1, 1 minus p power x minus 1 minus r plus 1 that can be treated as r minus success getting out of x minus 1 trials, which follows a binomial distribution multiplied by the rth success getting in the xth trail. Therefore, x can be r that means, you may get the rth success in the x trail itself, or you may get rth success in r plus 1th trail and so on.

So, that is a interpretation of a the probability mass function p of x equal to small x where x takes the value r r plus 1 and so on. So, this is also discrete type random variable, and CDF has a countably infinite jumps. So, I am not going to draw the CDF of a negative binomial, but one can visualize the CDF has a countably infinite jumps of this discrete type random variable.

For this random variable also, one can find the mean variance and so on, the mean of this random variable is nothing but r divided by p, and variance of this negative binomial that is r times 1 minus p divided by p square. You can verify when you put r is equal to 1 it has to be a same as half geometric distribution. And the MGF of negative binomial or Pascal distribution (Refer Time: 12:40) p times e power t divided by 1 minus 1 minus p times e power t whole power r, when r is equal to 1 that is same as the MGF of geometric distribution, I am not going for the derivation, but one can drive and you can get the same as 0.

(Refer Slide Time: 13:07)

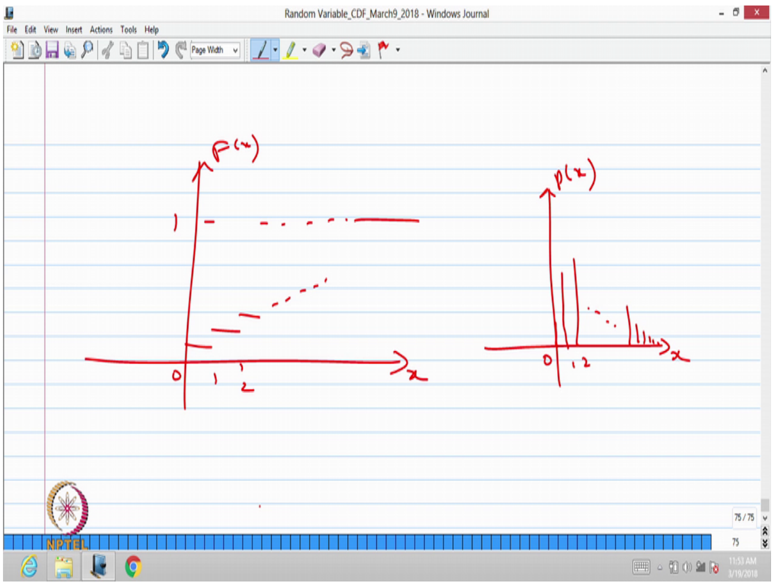



So, the next distribution is Poisson distribution is a very important distribution, because this connects a probability with a stochastic process in the different level. A discrete type random variable is said to be a Poisson distributed random variable, if the probability mass function of this random variable is going to be of the form e power minus lambda lambda power x divided by x factorial where x takes value 0 1 2 and so on, otherwise 0. Here the lambda has to be strictly greater than 0 it is a constant.

So, whenever any discrete type random variable whose probability mass function of this form, e power minus lambda lambda power x by x factorial; where x takes a value 0 1 2 and so on otherwise the probability mass function is must be 0, then that random variable is call it as a Poisson distributed random variable.

You can verify in this probability mass function, this is always greater or equal to 0, and if you make a summation over x starting from 0 to infinity, e power minus lambda is out and e power minus lambda is outside then the summation, and that summation quantity becomes e power lambda. And since the lambda is satirically greater than 0; Therefore, this quantity is going to be 1 therefore, this is a probability mass function.

(Refer Slide Time: 14:57)



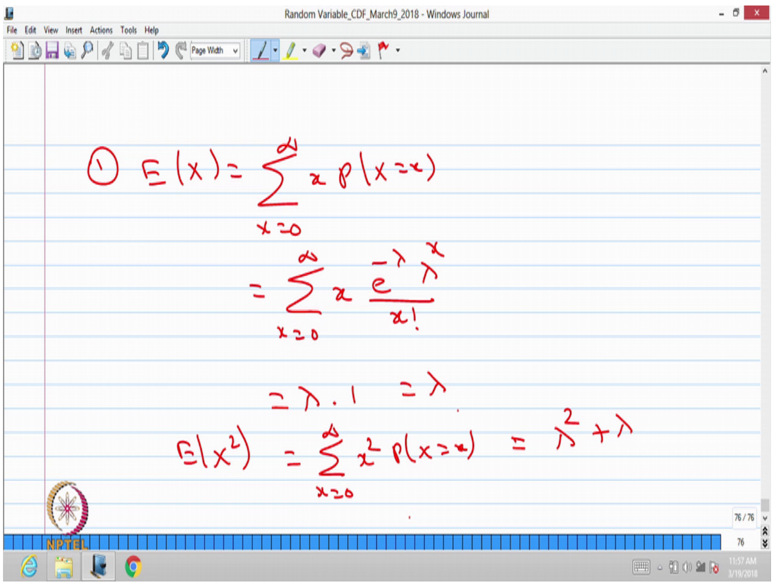

One can draw the CDF for this Poisson distributed, x takes value 0 it as a jump at x equal to 1, it has a another jump x equal 2, it has a another jump and so on countably infinity jumps. It touches one at infinity, similarly, if you draw the probability mass function of Poisson distribution 0, it has some height, and one it has another height, 2 it may have another height and so on, and it will be keep decreasing at countably infinite number of points.

So, this is a one diagram in which the lambda value is; so, that it is keep increasing decreasing, or there is a possibility it may have a at x is equal to 0, it to have a tallest then it may keep going down. And model possibility the summation of probability mass at the countably infinite number of points it is going to be 1. So, the same conclusion, if the data has cumulative distribution graph, or the histogram look like this CDF form or probability mass function form, then one can concluded that data for his Poisson distribution.

And there is another relation with Poisson distribution with binomial and Bernoulli, if you have a n independent Bernoulli distributed random variable, that summation becomes a binomial, when the n becomes very large, and the p probability of success is very small one can prove the limiting case of a n tends to infinity, and p is very small, then the binomial distribution will tends to Poisson distribution. For binomial distribution the n is always finite quantity, and the p is probability of success in any one Bernoulli trial, and all such n Bernoulli trails are with the probability of success p same, as and all are independent therefore, you are getting the binomial distribution.

But for larger n also when p is very small, then the limiting case of the binomial distribution goes to Poisson distribution, therefore, you have a countably infinite jumps in the CDF, one can visualize the limiting case of a binomial distribution is Poisson distribution. So, that is the connection between Bernoulli distribution, binomial distribution and Poisson distribution.

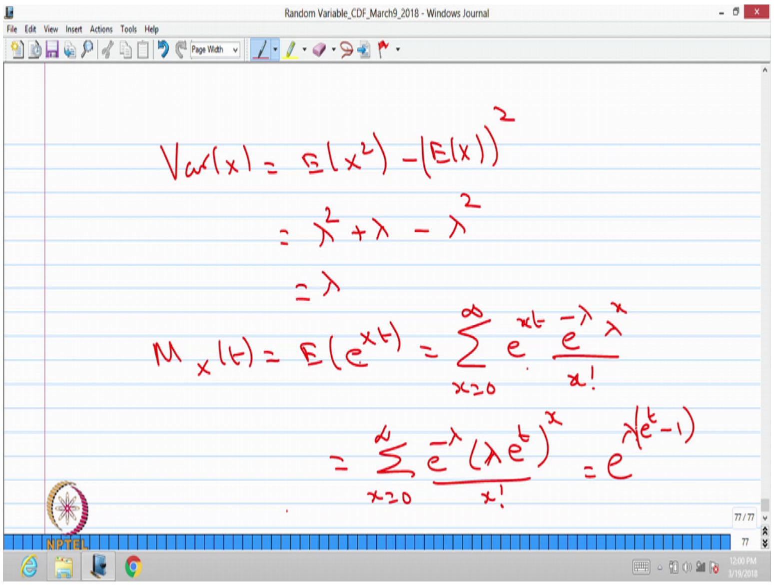
(Refer Slide Time: 17:45)




One can get the mean for Poisson distribution, that is nothing but a summation, x times probability of X is equal to x, where x takes a value 0 to infinity. This is nothing but x times probability of X is equal to x is e power minus lambda lambda power x by x factorial when x take a value 0 to infinity. X factorial and x cancel so, you will get x minus 1 factorial, you can take one lambda outside, the remaining quantity becomes one.

Therefore the mean is going to be lambda, similarly one can find E of X square E of X square that is nothing but summation x square times probability of x equal to x, where x takes a value 0 to infinity, the similar way one can compute. So, you can get the answer that is a lambda square plus lambda. If you do the little simplification by substituting probability of x equal to x that is e power minus lambda lambda power x by x factorial to the simplification, you will get lambda square plus lambda.

(Refer Slide Time: 19:06)



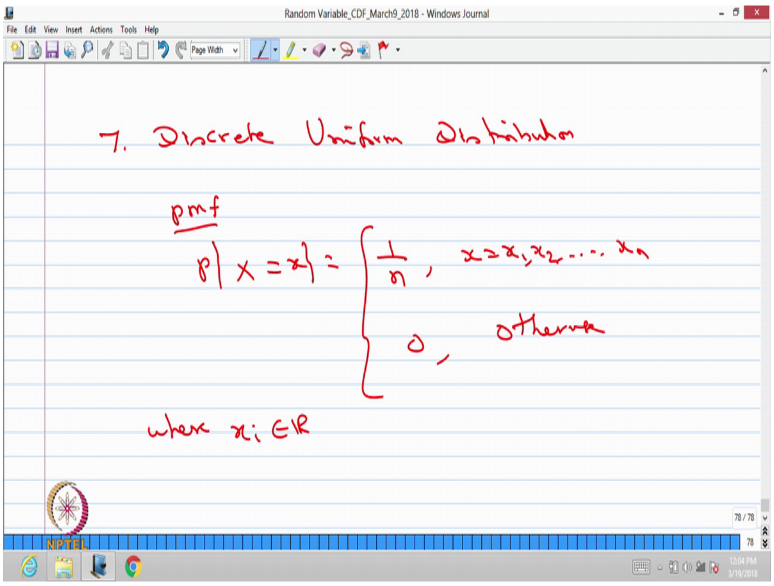

Therefore you can get the variance of X; that is E of x square minus E of X the whole square. And expectation of X square is a lambda square plus lambda and E of X is lambda therefore, that is lambda square so, simplify you will get lambda. So, these are very important result the mean and variance of the random variable is same which is lambda. So, here lambda is a parameter, because once you know the value of lambda you are known with the distribution; therefore, we use a notation x follows the capital P with the parameter lambda; that means, this is a Poisson distributed random variable with a parameter lambda.

So, once you specify the value of lambda, you are known with the distribution of this random variable. So, in Poisson distribution the important result is a mean and variance are same which is same as the parameter. Similarly, one can compute the moment generating function, because through this you can get all the moments of (Refer Time: 20:20). So, if you do the MGF calculation, it is a expectation of e power X times t then that is same as a summation e power small x t, and the probability of X takes value small x, that is e power minus lambda lambda power x by x factorial, where x takes a value from 0 to infinity. You can keep lambda and e power, sorry, you can keep a lambda power x and e power x t together. So, therefore, this is nothing but summation x is equal to 0 to infinity, e power minus lambda lambda times e power t power x i x factorial

So, if you do the little simplification, you can get the answer that is same as e power lambda times e t minus lambda so, that I can write it as minus 1. So, it is basically e power minus lambda is outside. So, this summation is nothing but e power lambda time's e power t. Therefore, it is a exponential of a lambda time's e power t minus 1 that is a MGF. So, from the MGF you can always get the by derivative you can get the expectation of x expectation of x square, then through that you can get the variance also.

So, since it is a discrete type random variable, you can go for probability generating function; so, even though I have not explained how to find out the probability generating function for all the distribution. So, you can starting from the Bernoulli binomial geometric Poisson and negative binomial, all this distribution because it is a discrete type and it takes a positive integer values. Therefore, one can go for finding the probability generating function of this stand common standard discrete type distributions.

(Refer Slide Time: 22:39)



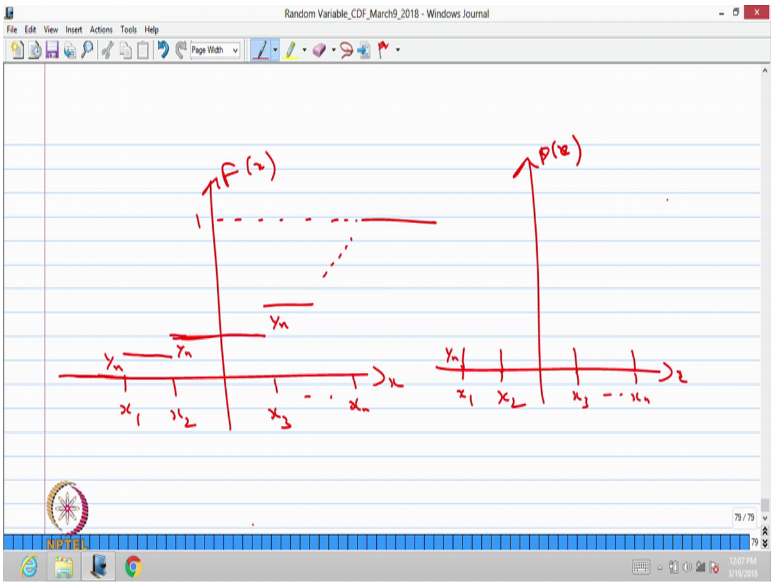
The next one that is number 7, that is a discrete uniform distribution, a discrete type random variable is said to be a discrete uniform distribution, whenever the probability mass function is of the form probability of X takes a value small x, that takes a value one divided by n, when x takes a value x 1 x 2 and so on x n, otherwise 0. Here the x is or the real numbers.

So, it can be any n points, the probability mass function are those n points is same, which is same as one divided by n, and all other points the probability mass function is 0 such a discrete type random variable is called a discrete uniform distribution, why the word uniform because the probability mass function is same for all such n points.

So, all such n point has to be district all should be different distinct n real values in which the probability mass function is a same. And since it is a probability mass function the summation has to be 1. Therefore, the for a uniform distribution the probability mass function is 1 divided by n, then only the summation is going to be one and all are going to greater than equal to 0 1, we at those n points.

So, such as discrete type random variable is called a discrete uniform distribution.

(Refer Slide Time: 24:32)




Let me draw the sample CDF, suppose x 1 is here, x 2 is here, x 3 is here, x n is somewhere here. It need not be equi-distance, it can be any n district n points. And the

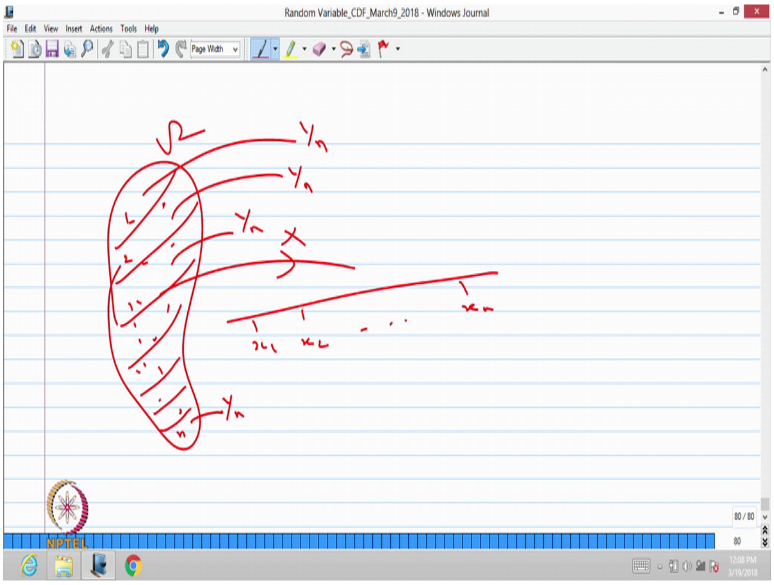
CDF I have just list out x 1 is the first value, and x 2 is greater than value x 3 is a greater than x 2 and so on. So, the CDF is 0 till x 1, at x 1 it has a jump and jump values 1 by n.

So, this height is 1 by n till x to the value is 1 by n at x 2 it as a next jump. Till x 3 it is going to be the same value and this jump is 1 by n. And x 3 it has the next jump and this jump is 1 by n. All the jump heights are same, at the point x n it has a last jump, and it becomes one.

If you see the CDF for any discrete type random variable, which has a only n jumps, and all the jump values are same, then that random variable is a discrete uniform distributed random variable. You can relate this CDF with the earlier random variable CDF. It may have a one jump or n plus 1 jumps or countably infinite jumps, but the jump values are different at different point. Whereas, here it is fixed always n jump points and always the n jump values are same which value is one divided by n that value is on by n, then that CDF is corresponding to the CDF of a discrete uniform distribution.

So, if you draw the probability mass function at those n points, the heights are going to be 1 by n same heights; that means, if you have a data in which if you draw the histogram. And all the histogram heights are same, with the n number of points or with the way you made a groups and so on, you can think of that could comes from the discrete uniform distribution. Or the data if you draw the CDF cumulative distribution and it has a same jump heights, and only finite number of jumps then it is a discrete uniform distribution.

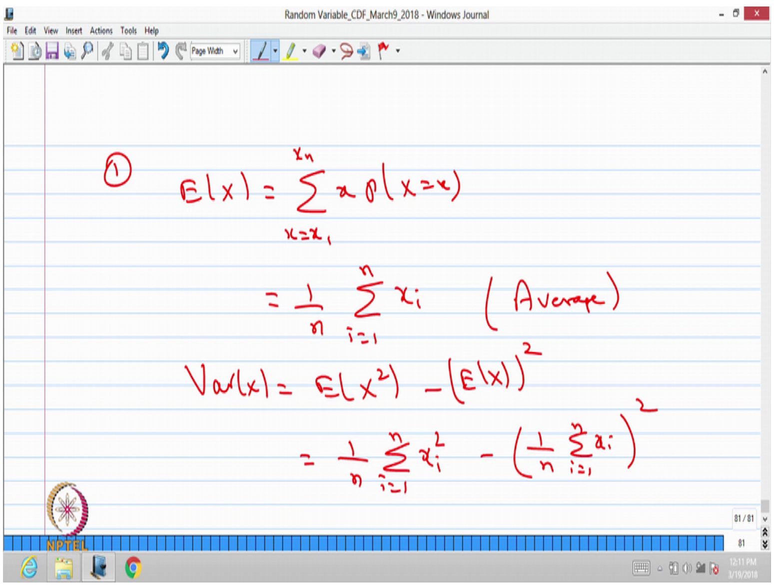
(Refer Slide Time: 27:39)




That means, you can interpret in other way also, you have omega, omega consist of finite or countably infinite, or uncountably many samples in it the way the mapping goes maps into x 1 x 2 x n, such a way partitioning omega into n pieces, and each one has a mass 1 by n. Each one is attached with one point whose probability mass function is 1 by n.

So, you partition is the first partition second partition so on. This is a nth partition whose mass is 1 by n; that means, that random variable is discrete uniform distribution.

(Refer Slide Time: 28:38)

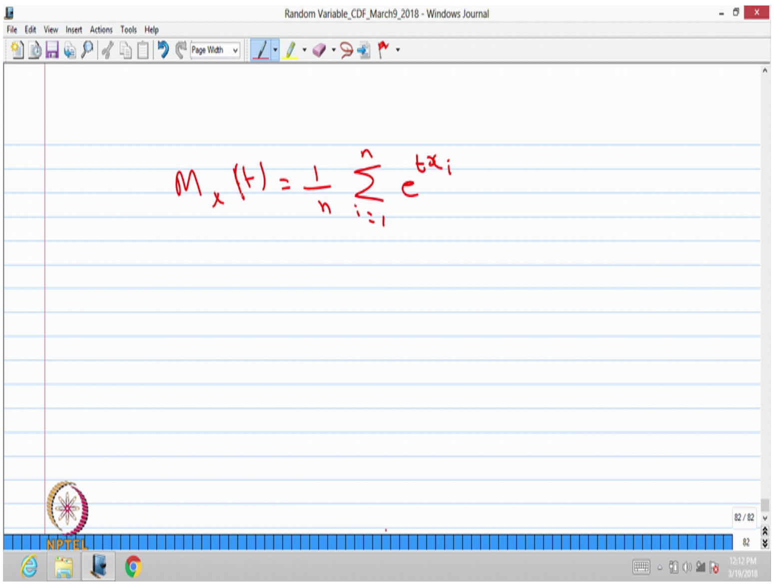


Since it is a discrete uniform distribution, you can find the mean is nothing but summation x times probability of X takes a value small x, where x takes value from x is equal to x 1 to x n. And the probability of x equal to x is 1 by n therefore, the 1 by 1 can be taken out side, and you add all the values I is equal to 1 to n.

So, the mean is nothing but me or expectation is nothing but some of those values multiplied by 1 by n. That is nothing but it is average. So, whenever the random variable is of the discrete uniform, the mean or expectation which is same as the average, we can go for finding the variance of x that is a expectation of x square minus expectation of x the whole square. So, first you compute the expectation of x square, then you substitute in this formula then you can get the variance.

So, since the probability mass function at those points is 1 by n, therefore, this is going to be one divided by n summation i is equal to 1 to n x i square minus. This is a 1 by n summation of x i, i is running from 1 to n the whole square so, if you do the simplification you can get it.

(Refer Slide Time: 30:27)




And the similar way you can get the MGF also, the MGF is going to be MGF of x, that is same as 1 by n summation e power t times x i, where i is running from 1 to n, because of the probability of mass function it x equal to x is 1 by n that can be taken out.

So, the summation e power t times x i i is equal to 1 to n with the multiplication 1 by n may give the moment generating function for the discrete uniform distributed random variable. With this we are completing a some common discrete distributions, starting from a constant Bernoulli binomial geometric negative binomial Poisson and discrete uniform distributions.

So, there are 7 distributions we have discussed they are all called common discrete type random variable; whose probability mass function and CDF mean variance MGFs are discussed.