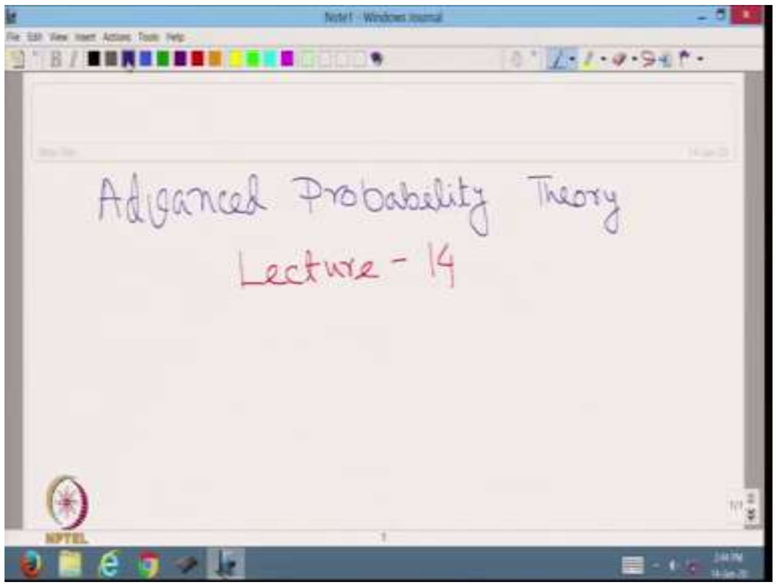
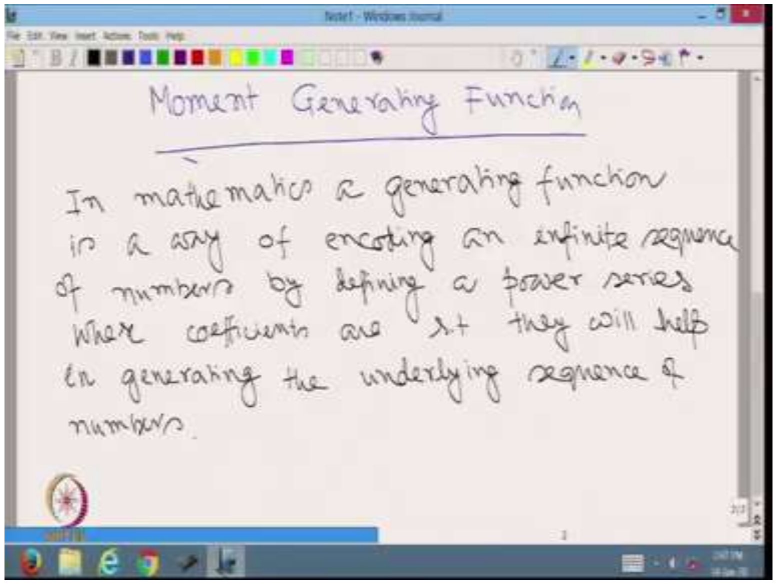
Advanced Probability Theory
Professor Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology, Delhi
Lecture 14

(Refer Slide Time: 00:21)



Welcome students to lecture number 14 in the MOOC course, on Advanced Probability Theory, as I said at the end of the last class that today we will start Moment Generating Function of a probability distribution.

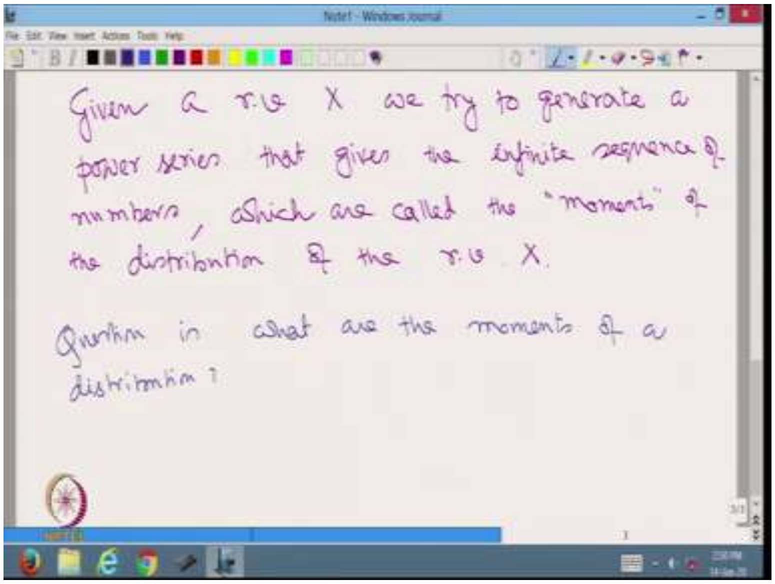
(Refer Slide Time: 00:40)

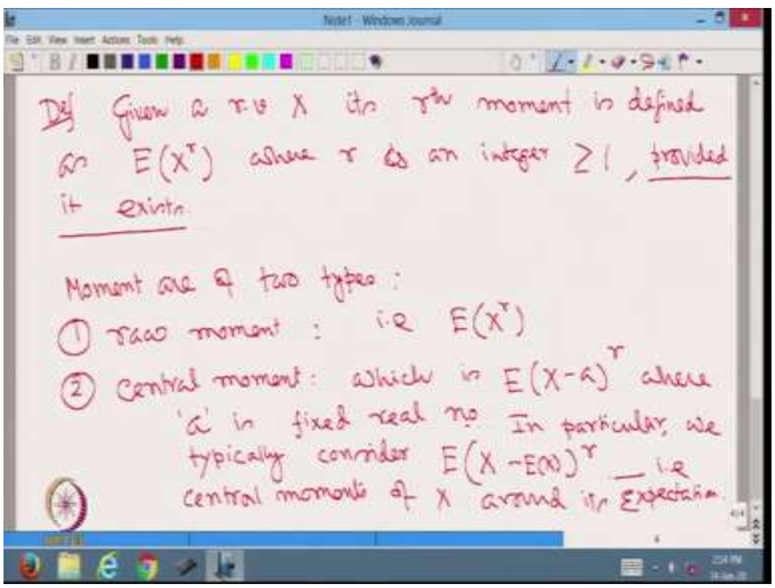


So, question is, what is moment generating function? In mathematics a generating function is a way of encoding an infinite sequence of numbers by defining a power series whose coefficients are such that they will help in generating the underlying sequence of numbers. Therefore, the theory is, if we have to remember, an infinite sequence of numbers, a0, a1, a2, up to infinity then we will develop a power series such that from that power series, we shall be able to generate these numbers.

Therefore with respect to our distribution if we can define something to be the moments, which are which essentially is an infinite sequence of numbers, then if we can generate a power series, such that the coefficients of that power series will give us these infinite number of moments of a distribution that, that function will be called the moment generating function.

(Refer Slide Time: 03:32)



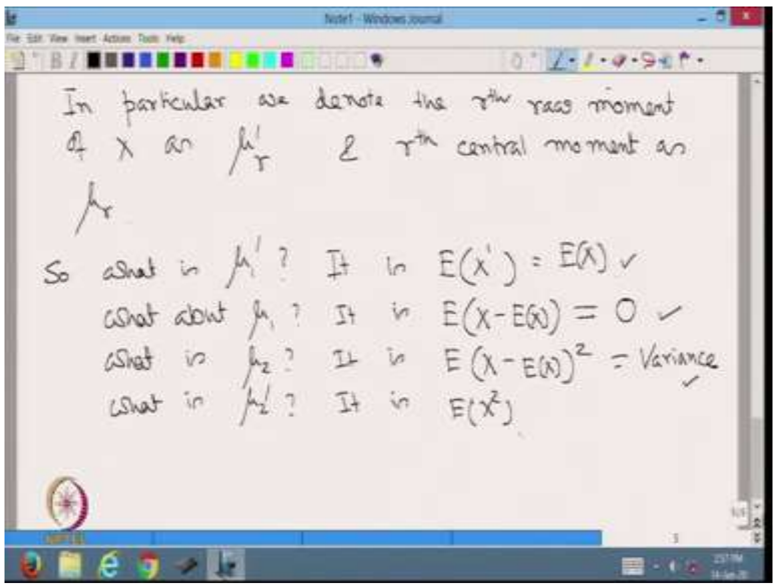


Therefore, given a random variable x, we try to generate a power series that gives the infinite sequence of numbers which are called the moments of the distribution of the random variable x. Question is what are the moments of a distribution.

So definition given a random variable x its r^{th} moment is defined as expected value of x to the power r, where r is an integer greater than or equal to 1 provided it exists So, this is very important, because there will be distributions whose moments or some specific moments may not exist moments are of two types, raw moment that is expected value of x to the power r and the other are central moments which is expected value of x minus a whole to the power r, where a is it fixed real number.

In particular, we typically consider expected value of x minus expectation of x whole to the power r that is central moments of x around its expectation. So I hope you understand the concept.

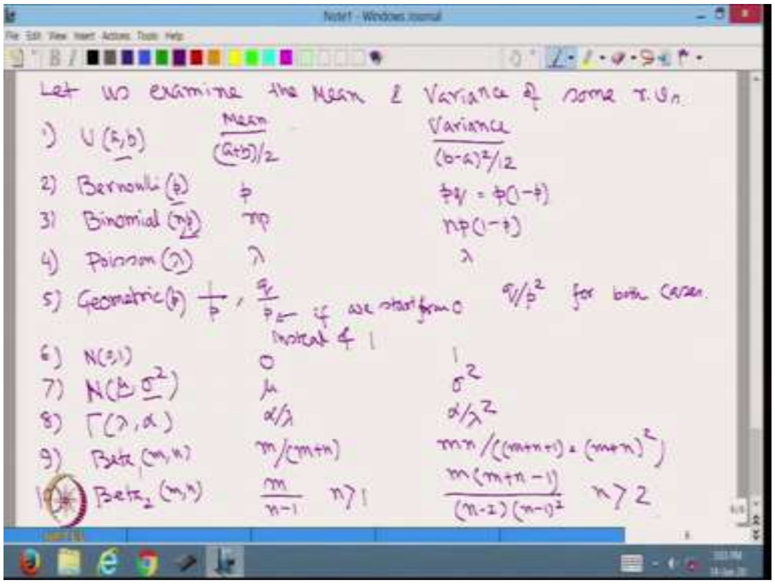
(Refer Slide Time: 8:05)



In particular we denote the rth raw movement of x as mu prime r and rth central moment as mu r. So, what is mu 1 prime it is expected value of x to the power 1 is equal to expected value of x. And over the last few classes, we have seen many examples of expectation of a random variable. In particular, we looked at random variables or certain standard distributions. What about mu 1? It is expected value of x minus expected value of x is equal to 0.

Therefore, mu 1 or the first central moment of a random variable is 0. What is mu 2? It is expected value of x minus expected value of x whole square which is variance. In fact, we have seen variance of different standard random variables in our last few classes. What is mu 2 prime? It is expected value of x square okay. Hence, the variance and expectation of random variables are very-very important, because they give us certain information with respect to the variable.

(Refer Slide Time: 10:51)



Now, let us examine the, and variance of some random variables, we have already seen them, but for your ready reference, I am enlisting them uniform a b. So, mean and variance let us write uniform a b mean is equal to a plus b by 2 and variance is equal to b minus a whole square by 12, 2. Bernoulli p expected value is p, variance is equal to p q is equal to p into 1 minus p, binomial np expected value is np and variance is equal to n into p into 1 minus p.

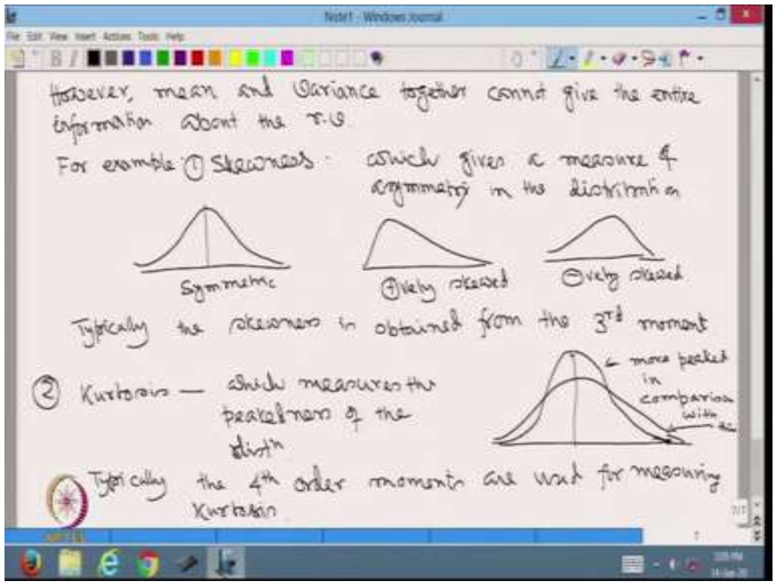
Poisson, lambda mean is equal to lambda and variance is also is equal to lambda, geometric there we have done it is 1 by p, but it is going to be q by p. If we start from 0 instead of 1 and variance is going to be q by p square for both cases. Normal 01 mean is equal to 0 variance is equal to 1 normal mu sigma square, mean is equal to mu, variance is equal to sigma square gamma lambda alpha mean is equal to alpha over lambda and variance is equal to alpha over lambda square.

Beta 1 m comma n mean is equal to m upon m plus n and variance is equal to, I did not calculate it, but I am giving you the answer. You please verify this by actually evaluating or computing the variance of beta m n from the first principle, it is going to be m plus n plus 1 into m plus n whole square and 10 beta to m comma n, it is going to be m upon n minus 1.

This we have computed in the class when m greater than 1 and it is m into m plus n minus 1 upon n minus 2 into n minus 1 whole square if n is greater than 2. I suggest that you keep these values in your finger tips for easy recollection and use of this.

Now, one thing I want you to notice that the mean and variance for each of the random variables are actually helping in identifying the parameters of the distribution, right? So, whatever parameters we are giving, they are coming in some form or other in the mean and variance of the corresponding distribution.

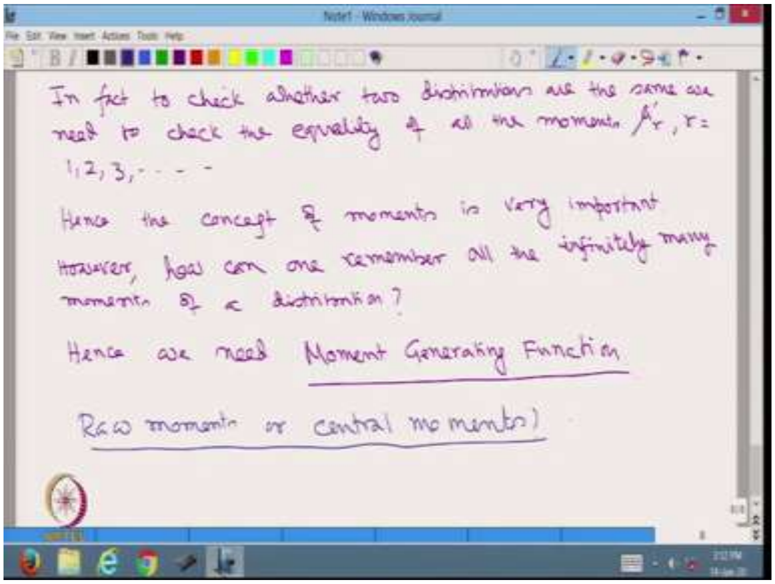
(Refer Slide Time: 15:37)



However, mean and variance together cannot give the entire information about the random variable. For example skewness which gives a measure of asymmetry in the distribution. Say for example, a normal 01 distribution is symmetric around its mean, On the other hand therefore its skewness is 0. On the other hand, a distribution of this form is called positively skewed and the distribution of this form is called negatively skewed.

Hence, typically the skewness is obtained from the third moment. Another important property so let us call it 1, let us call it 2 is called kurtosis which measures the peakedness distribution. For example, this one is more peaked in comparison with this one. How to measure that? Typically the 4th order moments are used for measuring kurtosis, okay.

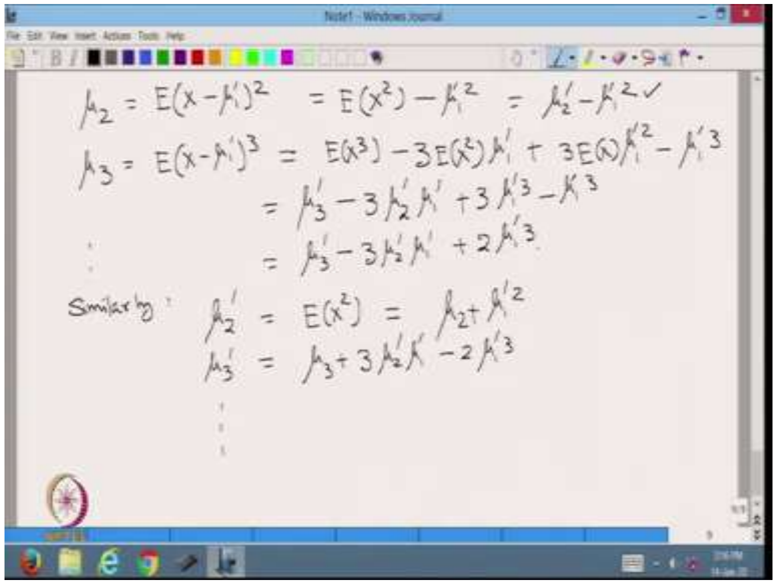
(Refer Slide Time: 19:45)



In fact to check whether two distributions are the same we need to check the equality of all the moments mu r prime, r is equal to 1, 2, 3 up to infinity. Hence the concept of moments is very important, however how can one remember all the infinitely many moments of a distribution? To answer this is something that I started with that we need moment generating function.

Now the question is raw moments or central moments? Often students think which one is more important, actually they are very closely related as we see now.

(Refer Slide Time: 22:26)



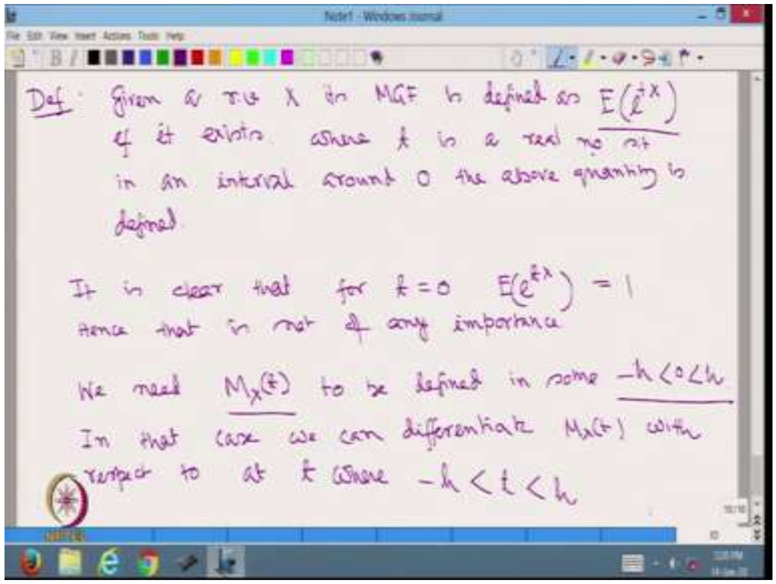
So, consider mu 2, mu 2 is equal to expectation of x minus mu 1 prime whole square, because mu 1 prime gives us the expectation is equal to expected value of x square minus mu 1 prime square. This we know that variance of x is equal to expectation of x square minus expectation of x whole square is equal to mu 2 prime minus mu 1 prime square.

Similarly mu 3 is equal to expected value of x minus mu 1 prime whole cube is equal to expected value of x cube minus 3 expected value of x square into mu 1 prime plus 3, expected value of x into mu 1 prime square minus mu 1 prime q is equal to mu 3 prime minus 3 mu 2 prime into mu 1 prime plus 3 mu 1 prime cube minus mu 1 prime cube is equal to mu 3 prime minus 3 mu 2 prime mu 1 prime plus 2 mu 1 prime cube.

In a similar way, we can express all the central moments with the help of its lowerer equal degree raw moments. Similarly, mu 2 prime is equal to expected value of x square is equal to mu 2 plus mu 1 prime square. This we get from this equation, and mu 3 prime is equal to mu 3 plus 3 mu 2 prime mu 1 prime minus 2 mu 1 prime whole cube.

Similarly, for higher order moments, so, the purpose of this is to establish that if we have all the raw moments we can get corresponding central moments and vice versa okay. So, once we understand the importance of moments, we need a mechanism to compute the different moments corresponding to a random variable.

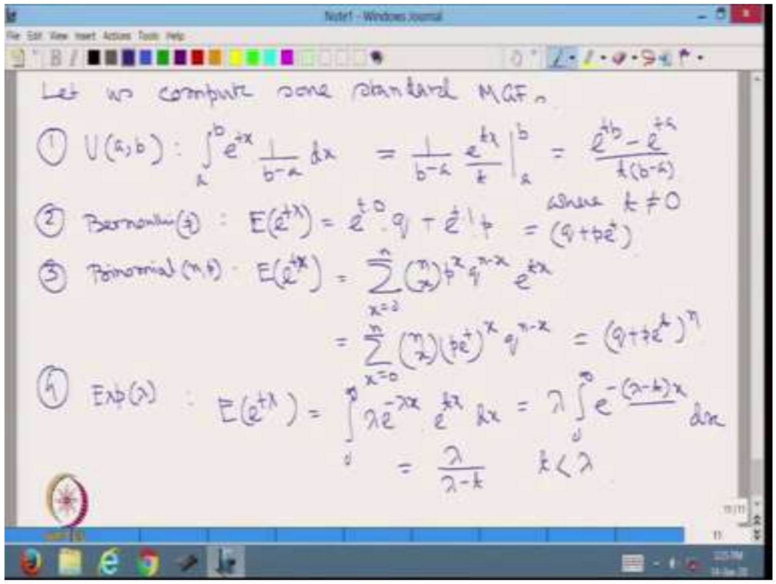
(Refer Slide Time: 25:42)



So, definition given a random variable x its moment generating function is defined as expected value of e to the power tx where if it exists. Sometimes we will see that this quantity does not exist for a random variable. Therefore, in such a case, we cannot define the moment generating function. Therefore, if it exists where t is a real number such that in an interval around 0 the above quantity is defined.

It is very clear that for t is equal to 0 expected value of e to the power t x is equal to 1. Hence, that is not of any importance. We need the moment generating function of x at a point t. So, this is the standard notation for a moment generating function to be defined in some minus h less than 0 less than plus h, what is the purpose? In that case we can differentiate M xt with respect to t at t where minus h less than t less than h.

(Refer Slide Time: 28:53)



So, let us compute from standard moment generating functions uniform a, b therefore, we are looking at integration a to b, e to the power t x into 1 minus b minus a dx because we are looking at expectation of e to the power t x is equal to 1 upon b minus a into e to the power t x upon t, b a is equal to e to the power t b minus e to the power t a upon t into b minus a, where t not equal to 0.

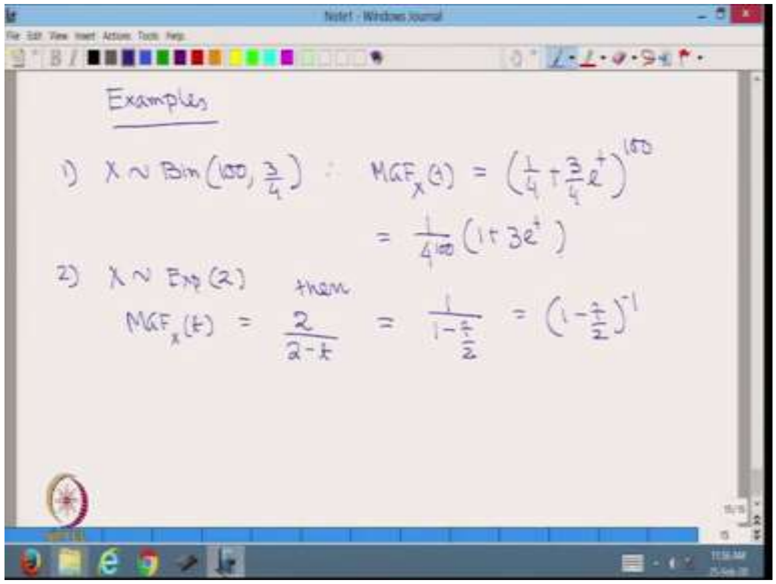
2, Bernoulli p therefore MGF is equal to expected value of e to the power t x is equal to e to the power t into 0 multiplied by the probability of 0 which is q plus e to the power t into 1 multiplied by the probability of 1 that is p.

Therefore, this is equal to q plus p, e to the power t, binomial n coma p expected value of e to the power t x is equal to sigma x is equal to 0 to n, n c x p to the power x q to the power n minus x e to the power t x is equal to sigma x is equal to 0 to n n cx p to the power t whole to the power x into q to the power n minus x which you can understand that we can write as q plus p e to the power t whole to the power n.

Let us consider exponential distribution with lambda therefore we are looking at expected value of e to the power t x is equal to 0 to infinity lambda e to the power minus lambda x into e to the power t x, dx is equal to lambda into 0 to infinity e to the power minus lambda minus t x, dx is equal to lambda upon lambda minus t. But, we know that lambda minus t has to be positive.

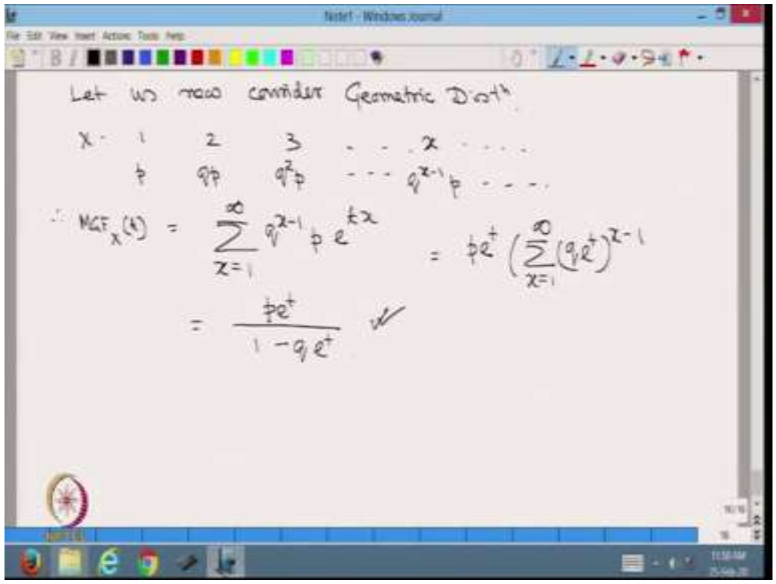
Therefore, it should be that t is less than lambda then only it will converge. Let us first see a few examples.

(Refer Slide Time: 32:54)



Suppose x is distributed as binomial with 100 and 3 by 4 therefore MGF of x at t is equal to q plus p e to the power t whole to the power n is equal to 1 upon 4 to the power 100 into 1 plus 3 e to the power t. Similarly, if x is exponential with parameter 2 then MGF of x at t is equal to lambda upon lambda minus t, which is equal to 1 upon 1 minus t by 2 is equal to 1 minus t by 2 whole to the power minus 1.

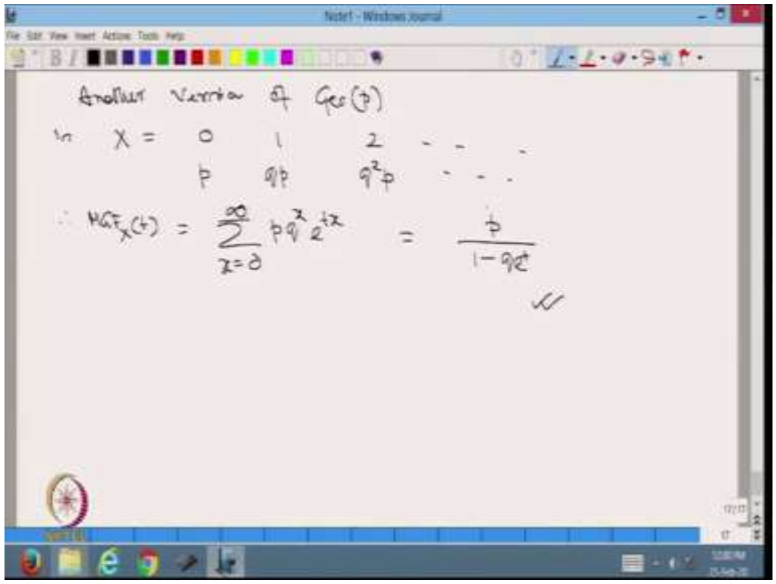
(Refer Slide Time: 34:13)



Let us now consider geometric distribution. So, we have x taking values 1, 2, 3 up to infinity with probabilities p, q p, q square p, q to the power x minus 1 p etc. Therefore, moment generating function of x at t is equal to sigma x is equal to 1 to infinity q to the power x minus 1 into p e to the power p x is equal to p e to the power t into sigma x is equal to 1 to infinity q e to the power t whole to the power x minus 1 is equal to p e to the power t upon 1 minus q e to the power t.

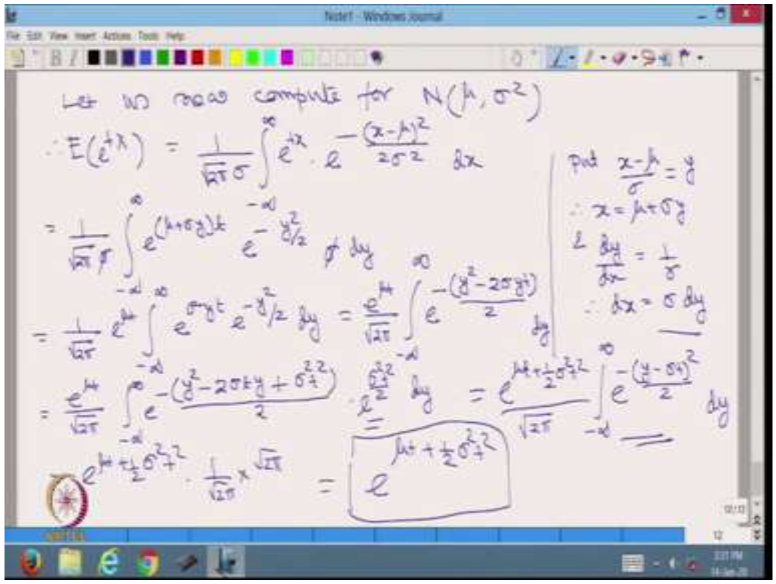
So, this is the moment generating function for the geometric distribution that we have considered.

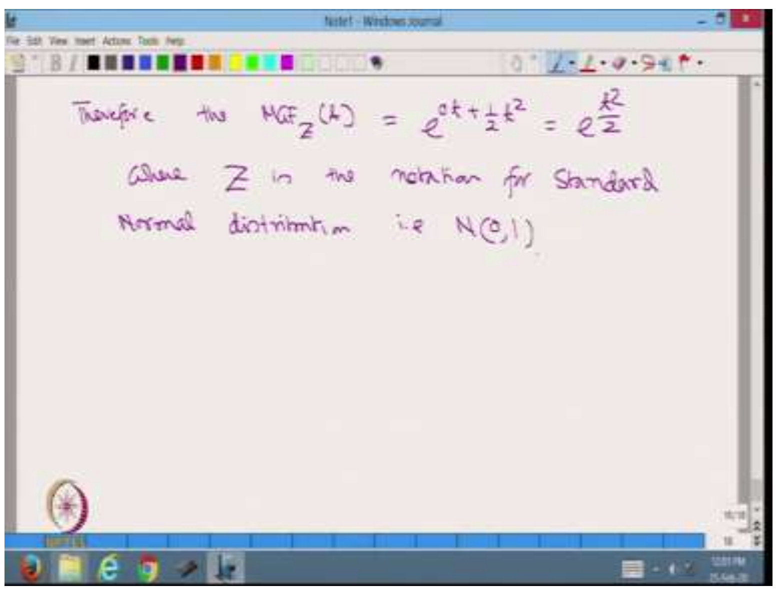
(Refer Slide Time: 35:48)



However, another version of geometric with p is x takes the value 0, 1, 2 etc with probabilities p, q p, q square p etc. Therefore MGF of x at t is equal to be sigma x is equal to 0 to infinity p q to the power x into e to the power t x is equal to as you can understand is going to be p upon 1 minus q e to the power t. So, in some book you may find that the moment generating function of geometric is given like this.

(Refer Slide Time: 36:52)





Let us now compute for normal mu comma sigma square. Therefore, expected value of e to the power t x is equal to 1 over root over 2 phi sigma minus infinity to infinity e to the power t x into e to the power minus x minus mu whole square upon 2 sigma square dx put x minus mu upon sigma is equal to y. Therefore, x is equal to mu plus sigma y and dy dx is equal to 1 by sigma. Therefore, dx is equal to sigma dy.

So, when we put those values, we get it 1 over root over 2 Pi sigma minus infinity to infinity e to the power mu plus sigma y multiplied by t e to the power minus y square by 2 into sigma dy. So, this sigma cancels with this.

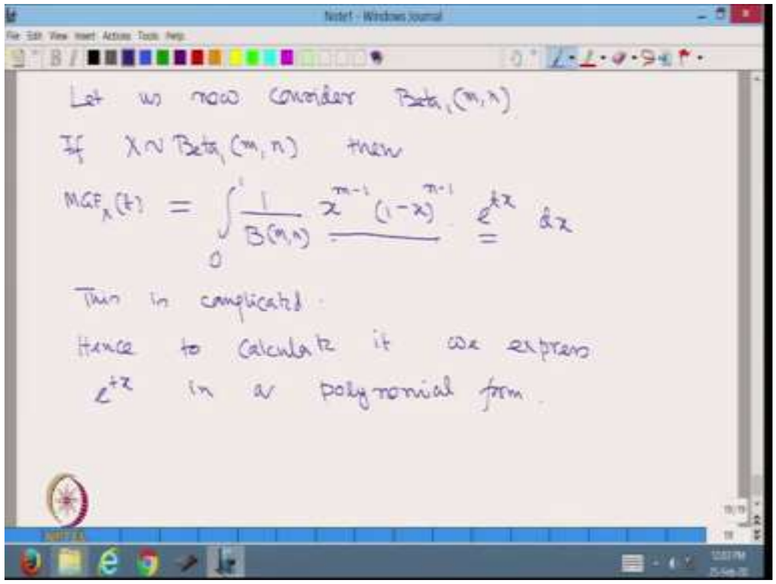
Therefore, we are getting 1 over root over 2 Pi e to the power mu t we take out of the integration because that is not involved in the integration process e to the power sigma y t e to the power minus y square by 2 dy is equal to e to the power mu t upon root over 2 Pi integration of minus infinity to infinity e to the power minus y square minus 2 sigma y t upon 2 dy is equal to e to the power mu t upon root over 2 Pi integration of minus infinity to infinity e to the power minus y square minus 2 sigma t y plus sigma square t square divided by 2.

So we have added sigma square t square with a negative sign. Therefore, to compensate this we write into e to the power sigma square t square by 2 dy. Now, this quantity if you take out we get e to the power mu t plus half sigma square t square upon the root over 2 Pi integration minus infinity to infinity e to the power minus y minus sigma t whole square by 2 dy.

And this we know is a standard normal integration by changing y minus sigma t to z we can write it as e to the power minus z square by 2 therefore that is going to give us over 2 Pi. Hence, we can write it as e to the power mu t plus half sigma square t square, 1 over root 2 Pi root over 2 Pi which is equal to e to the power mu t plus sigma square t square. So, that is the moment generating function of a normal random variable with mean is equal to mu and variance is equal to sigma square.

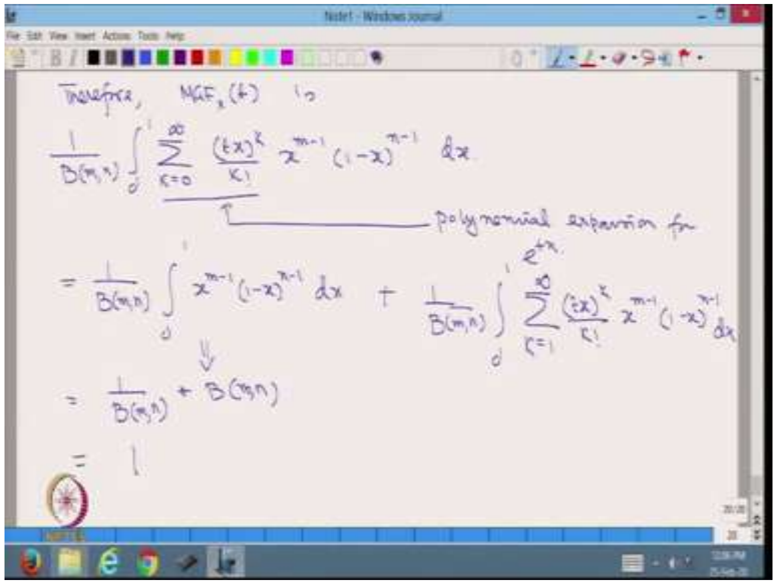
Therefore, the moment generating function of z at t is equal to e to the power 0 t plus half t square is equal to e to the power t square by 2 where z is the notation for standard normal distribution that is normal 01.

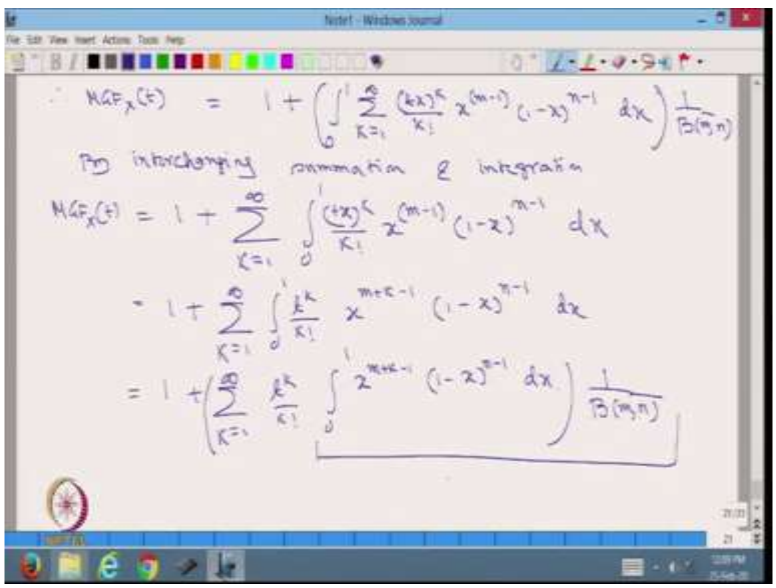
(Refer Slide Time: 42:40)



Let us now consider beta 1 m comma n. So if x is distributed as beta 1 m comma n then MGF of x at t is equal to integration 0 to 1, 1 upon beta m comma n, x to the power m minus 1, 1 minus x to the power n minus 1 into e to the power tx dx. So, this is slightly complicated as in the PDF the x coming in a polynomial form. Whereas, e to the power tx is exponential form hence to calculate it We express e to the power tx in a polynomial form.

(Refer Slide Time: 44:21)





Therefore MGF of x at t is 1 upon beta m comma n, integration 0 to 1, sigma k is equal to 0 to infinity tx to the power k upon factorial k multiplied by x to the power m minus 1, 1 minus x whole to the power n minus 1 dx as you can understand that this is the polynomial expansion for e to the power tx is equal to 1 upon beta m comma n integration 0 to 1.

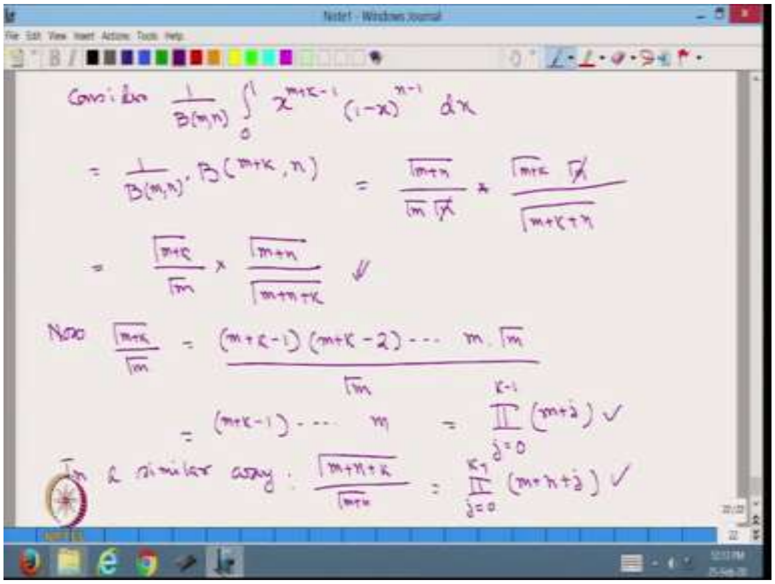
Now, let us take the term k is equal to 0 out. So, that is going to give us x to the power m minus 1 into 1 minus x whole to the power n minus 1 dx plus 1 upon beta m comma n integration 0 to 1 sigma k is equal to 1 to infinity tx to the power k upon k factorial is x to the power m minus 1, 1 minus x whole to the power n minus 1 dx.

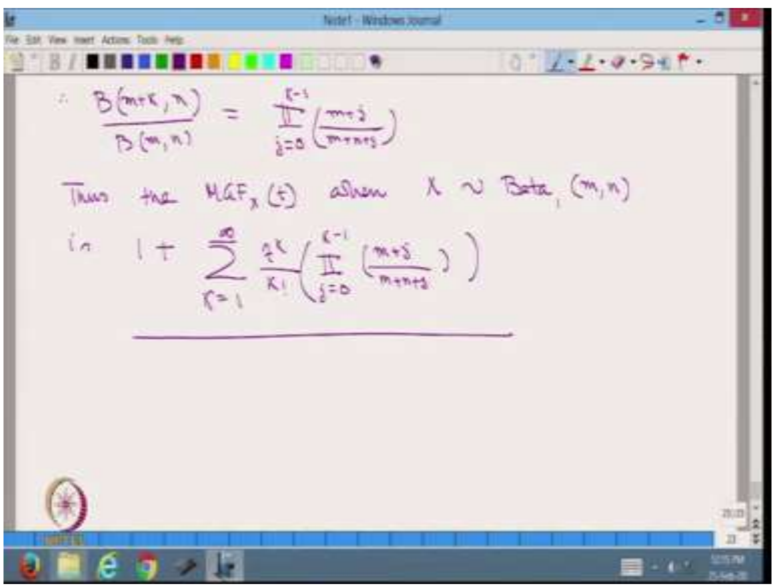
This part as you can understand is going to be 1 upon beta m, n into beta m comma n is equal to 1 therefore MGF of x at t is equal to 1 plus integration 0 to 1, sigma k is equal to 1 to infinity tx to the power k upon k factorial x to the power m minus 1, 1 minus x to the power n minus 1 dx.

Now by interchanging summation and integration we get MGF of x at t is equal to 1 plus sigma k is equal to 1 to infinity, integration of 0 to 1 tx to the power k upon factorial k x to the power m minus 1, 1 minus x whole to the power n minus 1 dx is equal to 1 plus sigma k is equal to 1 to infinity, integration 0 to 1, t to the power k upon factorial k x to the power m plus k minus 1 into 1 minus x whole to power n minus 1 dx is equal to 1 plus sigma k is equal to 1 to infinity t to the power k upon factorial k integration 0 to 1 x to the power m plus k minus 1 into 1 minus x whole to the power n minus 1 dx.

Now, note that we have a 1 upon beta a main factor from here that we have not considered yet therefore, this whole thing is also to be multiplied by 1 upon beta m comma n. Now, let us look at this part.

(Refer Slide Time: 49:02)





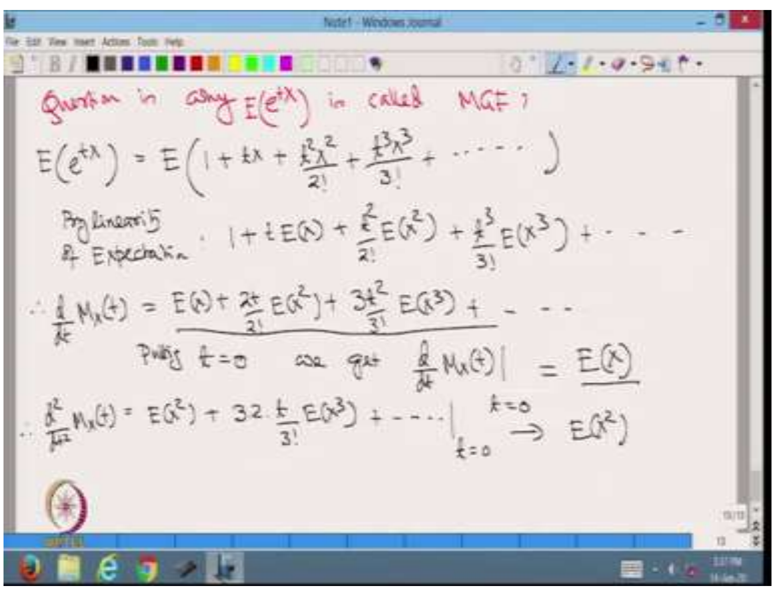
Consider 1 upon beta m comma n integration 0 to 1 x to the power m plus k minus 1, 1 minus x to the power n minus 1 dx is equal to 1 upon beta m comma n into beta m plus k comma n is equal to gamma m plus n upon gamma m gamma n multiplied by gamma m plus k into gamma n upon gamma m plus k plus n. Gamma m, gamma n and gamma n cancels out therefore, what we are left with is gamma m plus k upon gamma m multiplied by gamma m plus n upon gamma m plus n plus k.

Now, gamma m plus k upon gamma m, we can write it as m plus k minus 1 into m plus k minus 2 up to m times gamma m divided by gamma m is equal to m plus k minus 1 into up to m is equal to product of say j is equal to 0 to k minus 1 m plus j. In a similar way gamma m plus n plus k upon gamma m plus n is equal to product of j is equal to 0 to k minus 1 m plus n plus j.

Now, in this we are going to place these 2 terms therefore beta m plus k comma n upon beta m comma n is equal to product of j is equal to 0 to k minus 1, m plus j upon m plus n plus j thus the moment generating function of x at t, when x is distributed as beta 1 with m comma n is 1 plus sigma, k is equal to 1 to infinity t to the power k upon factorial k product j is equal to 0 to k minus 1 of m plus j into m plus n plus j.

So, that is the answer, which is slightly complicated, but a good exercise to work out. So, we have seen the moment generating function for several random variables given any random variable, we can try to find its MGF in a very similar way, although it is not necessary that it will always converge.

(Refer Slide Time: 53:48)

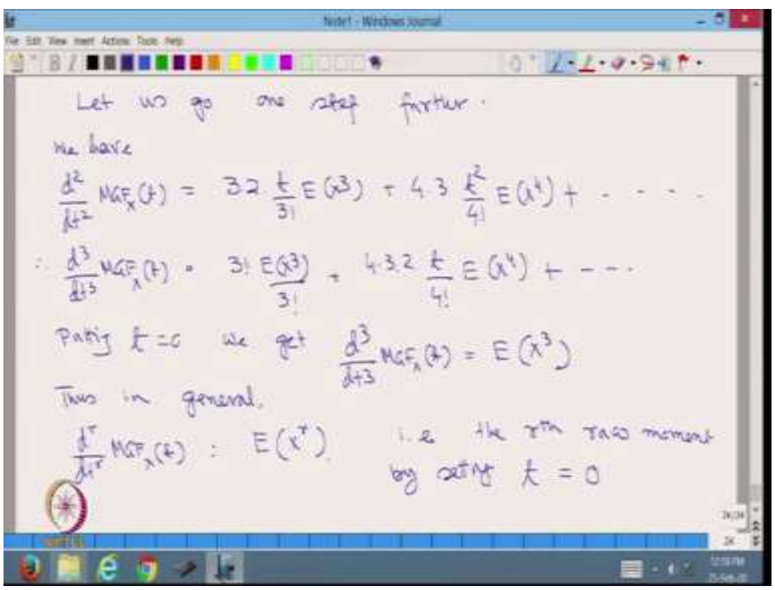


Question is why e to the power tx is its expectation is called moment generating function. So, we see that expectation of e to the power tx is equal to expectation of 1 plus tx plus t square x square upon factorial 2 plus t cube x cube upon factorial 3. So, we get an infinite series and by linearity of expectation we can write it as 1 plus t times expected value of x plus t square upon factorial 2 expectation of x square plus t cube upon factorial 3 expectation of x cube plus up to infinity.

Therefore, this is the moment generating function in general, therefore, d dt of Mx t is equal to expected value of x plus 2 t upon 2 factorial expected value of x square plus 3 t square upon 3 factorial expectation of x cube infinite sum, putting t is equal to 0 we get d dt of Mx t at t is equal to 0 is equal to expected value of x. Thus we get the first raw moment by differentiating Mx t with t and setting t is equal to 0.

Let us now take the second derivative. Therefore, d to dt square of Mx t is equal to, from here by differentiating this with respect to t we get is equal to expectation of x square plus 3 into 2 into t factorial 3 expectation of x cube plus up to infinity and taking t is equal to 0 is going to give us expected value of x square.

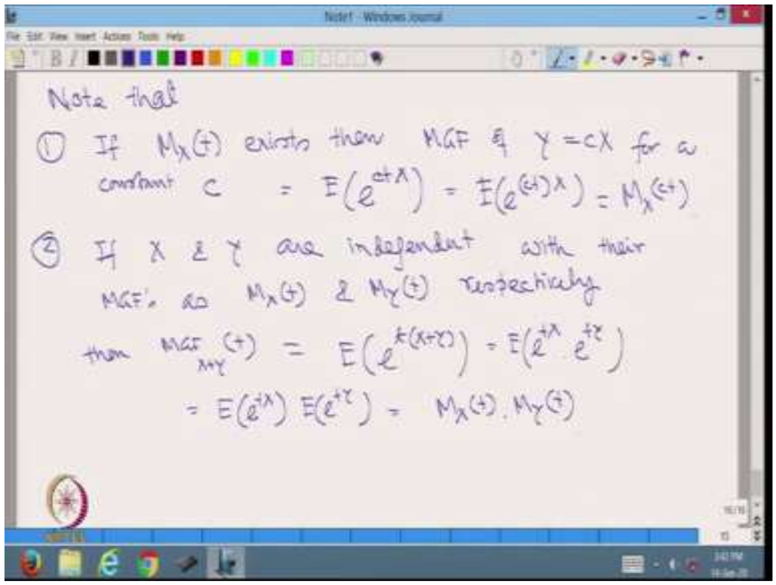
(Refer Slide Time: 57:18)

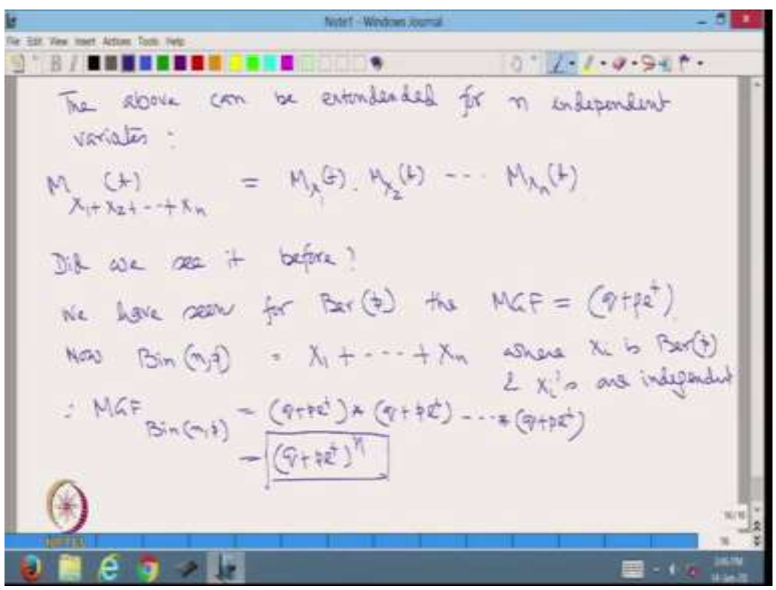


Let us go one step further, we have d2 dt 2 of moment generating function of x at t is equal to 3 into 2 into t by factorial 3 expected value of x cube plus 4 into 3 into t square upon factorial 4 expected value of x to the power 4 like that. Therefore, if we take the third derivative of the MGF at point t, we get factorial 3, expected value of x cube upon factorial 3, plus 4 into 3 into 2 into t upon factorial 4 into expected value of x to the power 4.

Like that putting t is equal to 0, we get d cube dt cube, d3 dt cube of MGF at t is equal to expected value of x cube because subsequent terms are becoming 0. Thus, in general the r th derivative of the moment generating function at t will give us expected value of x whole to the power r that is the r th raw moment by setting t is equal to 0.

(Refer Slide Time: 59:53)





Note that if Mx t exists then MGF of y is equal to cx for a constant c is equal to expected value of e to the power Ct x is equal to expected value of e to the power Ct times x is equal to moment generating function of x around the point Ct. Second thing is if x and y are independent with their moment generating functions Mx t and My t respectively.

Then moment generating function of x plus y at t is equal to expected value of e to the power t x plus y is equal to expected value of e to the power tx into e to the power ty is equal to, because they are independent, we can write it as expectation of e to the power tx into expectation of e to the power ty is equal to moment generating function of x multiplied by moment generating function of y.

This above can be extended for n independent variates that is moment generating function of x 1 plus x 2 plus xn at t is equal to Mx 1 at t into Mx 2 t into Mxn t. Did we see it before? If you remember, we have seen for Bernoulli p, the MGF is equal to q plus p e to the power t. Now, binomial n p number of heads in n independent trials of a coin is equal to actually x 1 plus x 2 plus xn where xi is Bernoulli with p and xi‘s are independent.

Therefore MGF of binomial n p is equal to q plus p e to the power t into q plus p, e to the power t n times. And if you look at the earlier part of this lecture, you can see that we have obtained that the moment generating function of a binomial random variable is indeed q plus p e to the power t whole to the power n okay friends, I stop here today, In the next class I shall discuss about different other generating functions. Okay, till then thank you.