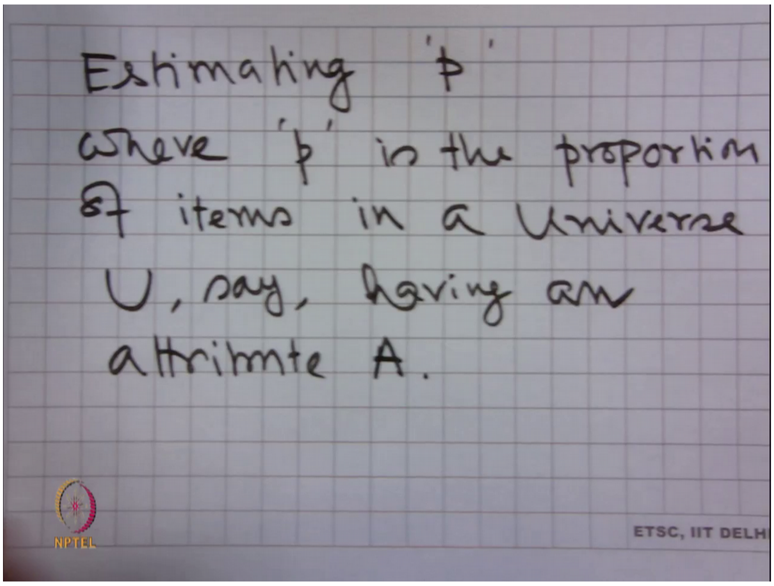
Statistical Inference
Prof. Niladri Chatterjee
Department of Mathematics
Indian Institute of Technology, Delhi

Lecture – 04
Statistical Inference

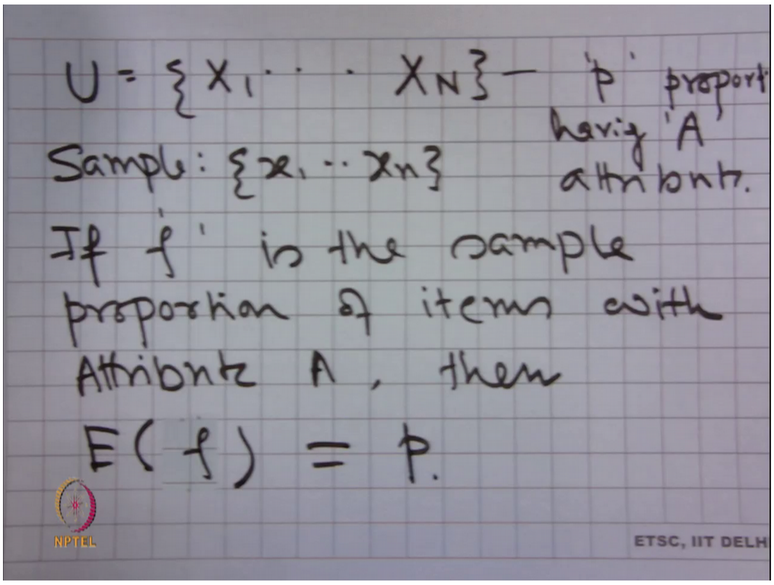
Welcome students to the 4th lecture on the MOOC's series on Statistical Inference.

(Refer Slide Time: 00:40)



If you remember in the last class I was talking about estimating p where p is the proportion of items in universe use say having an attribute A. So, this was the problem that we have done towards the end of the last class.

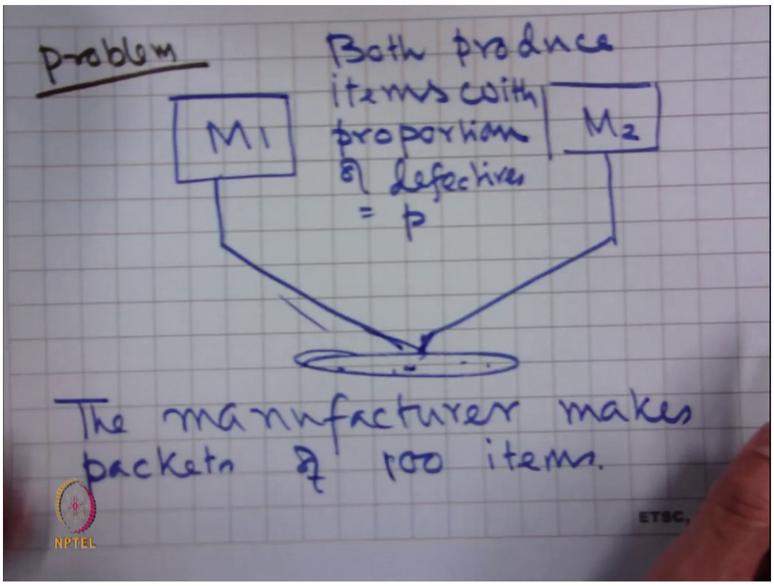
(Refer Slide Time: 01:36)



And what we have found that if universe is equal to X 1 X 2 X N of which pre proportion having A attribute. We take a sample x 1 x 2 x n if f is the sample proportion of items with attribute A, then expected value of f is equal to p. This we have found out in the last class.

Today I start with that with a slightly more complicated problem.

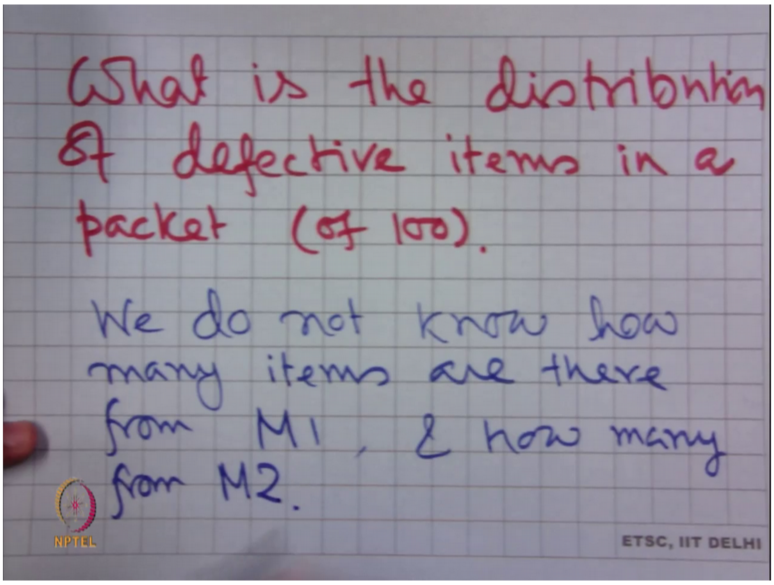
(Refer Slide Time: 02:54)



Suppose there are machines M 1 and M 2 both produce items with proportion of defectives is equal to p. So, M 1 produce items and M 2 produce items, they are all coming to the common pool, and the manufacturer makes packets of 100 items.

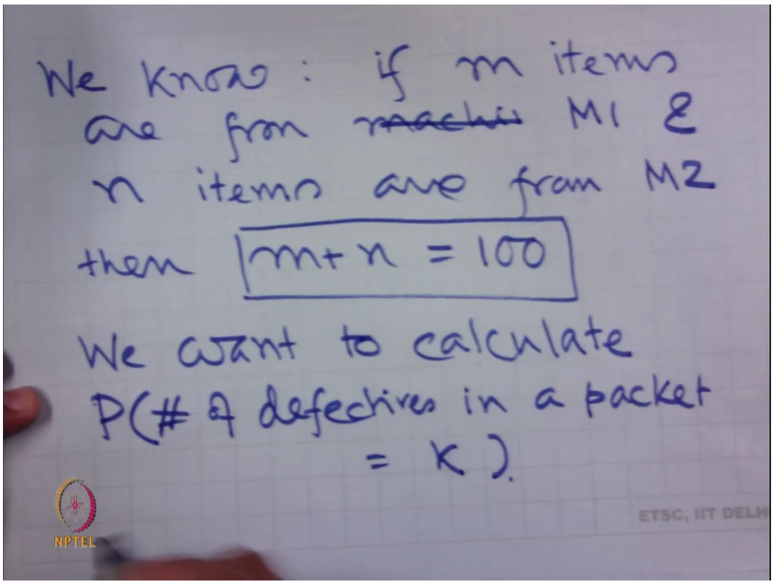
So, when items being produced by machine 1 and items being produced by machine 2 are coming to the common pool from there where the manufacturer is taking 100 items, and make packets, he wants to know what is the distribution of defective items in a packet; packet means, packet of 100.

(Refer Slide Time: 04:47)



Suppose we want to find out this distribution. The problem is we do not know how many items are there from machine one, and how many from machine 2.

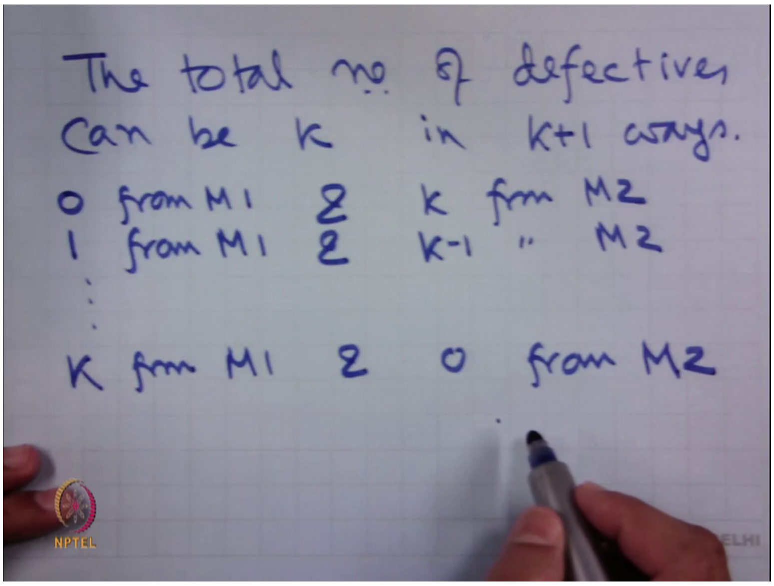
(Refer Slide Time: 06:01)



Only thing that we know if m items are from machine M 1 and n items are from M 2, then m plus n is equal to 100 this much we know.

So, we want to calculate probability number of defectives in a packet is equal to k.

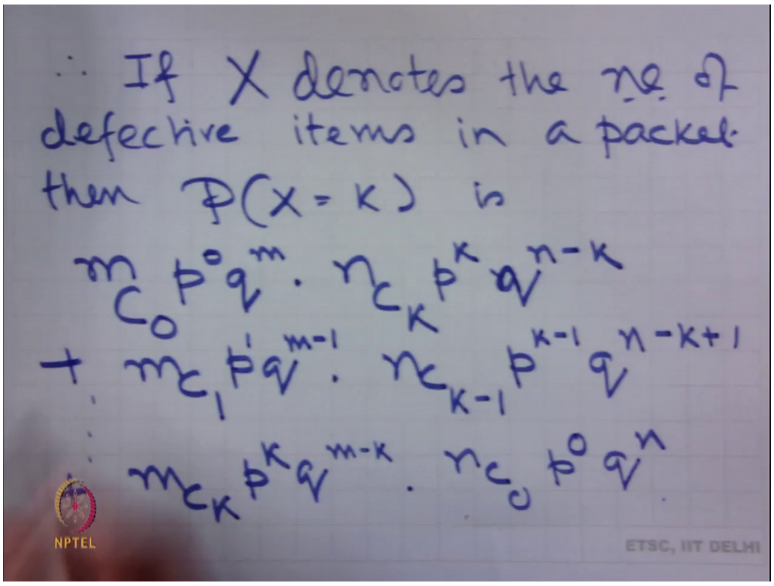
(Refer Slide Time: 07:21)



Let us note that the total number of defectives can be k in k plus 1 ways, 0 from M 1 and k from M 2 1 from M 1 and k minus 1 from M 2 up to k from M 1 and 0 from M 2.

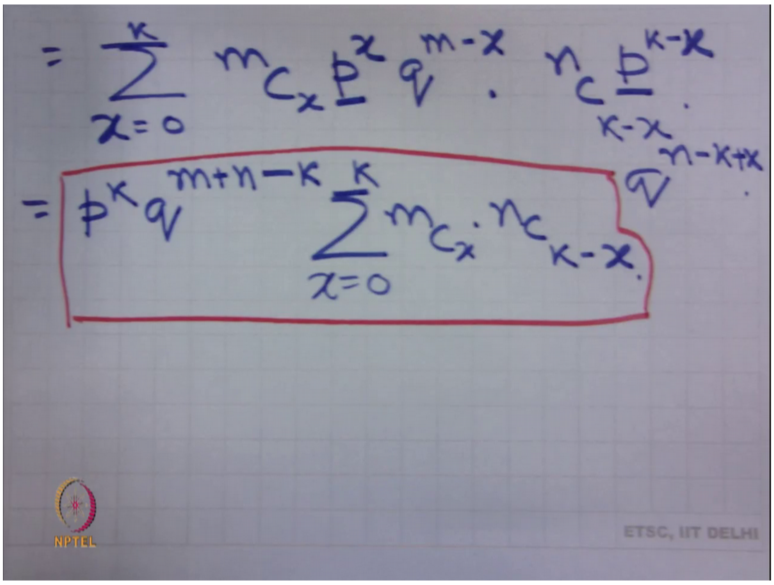
So, the probability that there are k many defective items in a packet can be computed by adding these k plus 1 individual events, and some of them may have probability 0.

(Refer Slide Time: 08:44)



So, let us see how to work out that therefore, if x denotes the number of defective items in a packet, then probability x is equal to k is; number of ways of choosing 0 defectives from machine 1 into number of ways of choosing k defectives from machine 2, plus n number of ways of choosing one defective out of m from machine 1, multiplied by number of ways of choosing k minus 1 defectives from machine 2 up to m C k p to the power k.

(Refer Slide Time: 10:40)

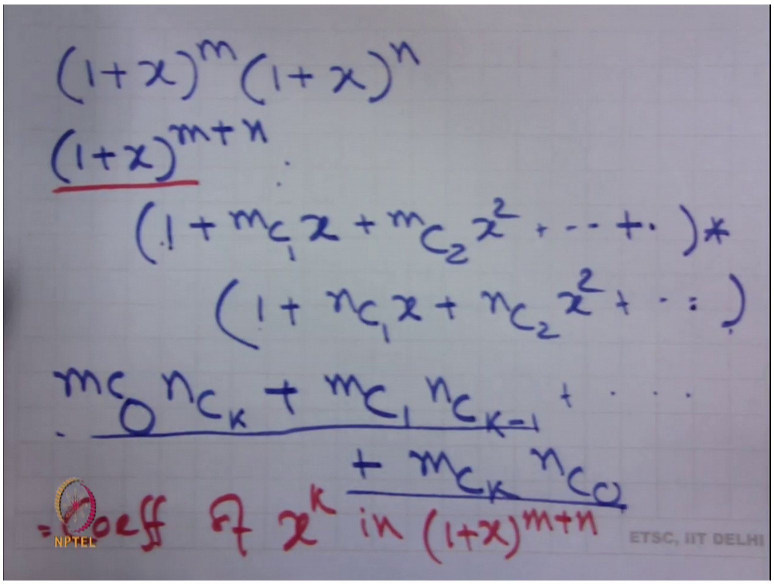


This we can write it as sigma x is equal to 0 to k m C x p to the power x q to the power m minus x multiplied by n C k minus x p to the power k minus x into q to the power n minus k plus x; is equal to, now if you look at it, this comes out to be p to the power k which we can take out of the summation.

This comes out to be q to the power m plus n minus k so that is also independent of x. So, we can take it out of the summation, and then what we get x is equal to 0 to k m C x into n C k minus x. So, that is the expression that we get

So now we need to find out this term. What is summation over x from 0 to k m C x multiplied by n C k minus x?

(Refer Slide Time: 12:35)

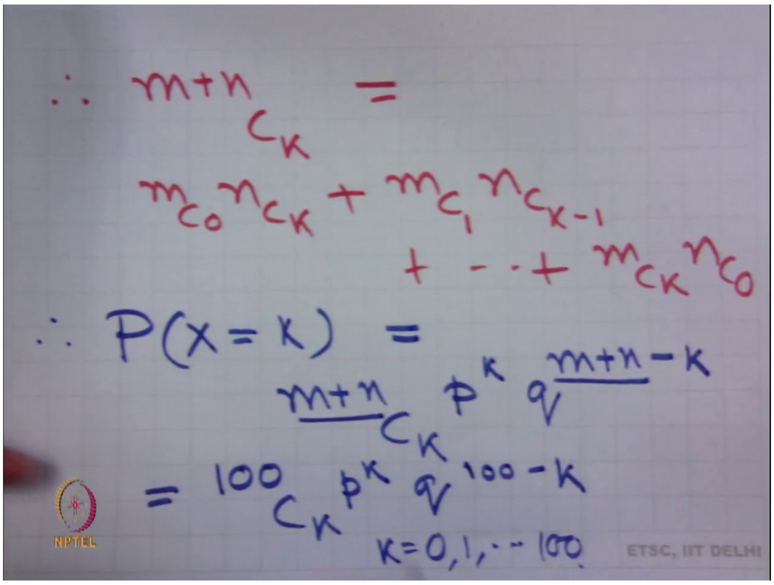


To do that let us consider 1 plus x whole to the power n into 1 plus x whole to the power n. So, this term is 1 plus x whole to the power m plus n.

Now, let us expand both of them binomially. Therefore, if we consider the power of x to the power 0 0 here which is mc 0, and multiplied with power of x to the power k here that is n C k plus mc 1 into power of x to the power k minus 1 here which is n C k minus 1, and like that power of the coefficient of x to the power k multiplied by the coefficient of x to the power 0 here. So, that is what we are getting as m C k multiplied by n C 0

So, this is precisely the summation that we are talking about. If you look at it, it is mc 0 into n C k plus mc 1 into n C k minus 1 up to m C k into nc 0. So, this is the summation that we get. And what it is? So, this is coeff of x to the power k in 1 plus x whole to the power m plus n.

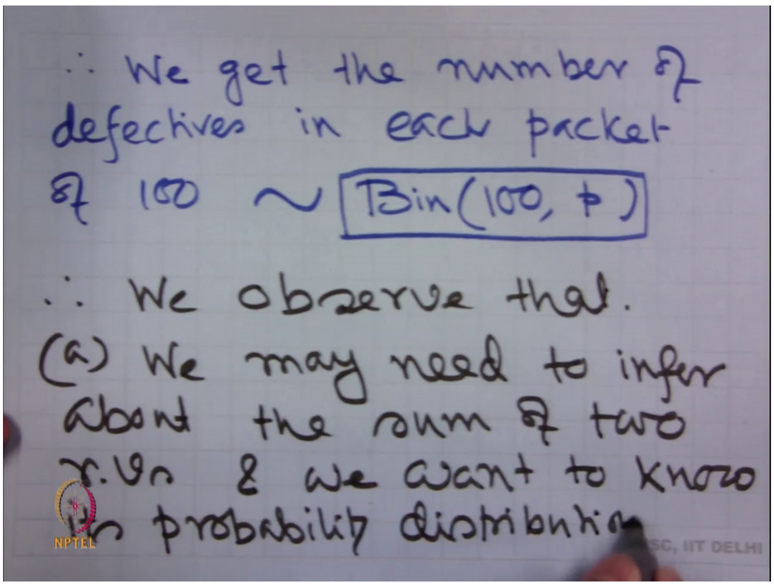
(Refer Slide Time: 15:12)



Therefore m plus n C k is equal to mc 0 into n C k plus mc 1 into n C k minus 1 plus up to m C k into nc 0. Therefore, the final probability is equal to so, this was the term that we have obtain. Therefore, final of probability x is equal to k is equal to m plus n c k multiplied by p to the power k q to the power m plus n minus k.

Since m plus n is constant equal to 100, we can write it as 100 ck p to the power k q to the power 100 minus k. And this is true for k is equal to 0 1 up to 100.

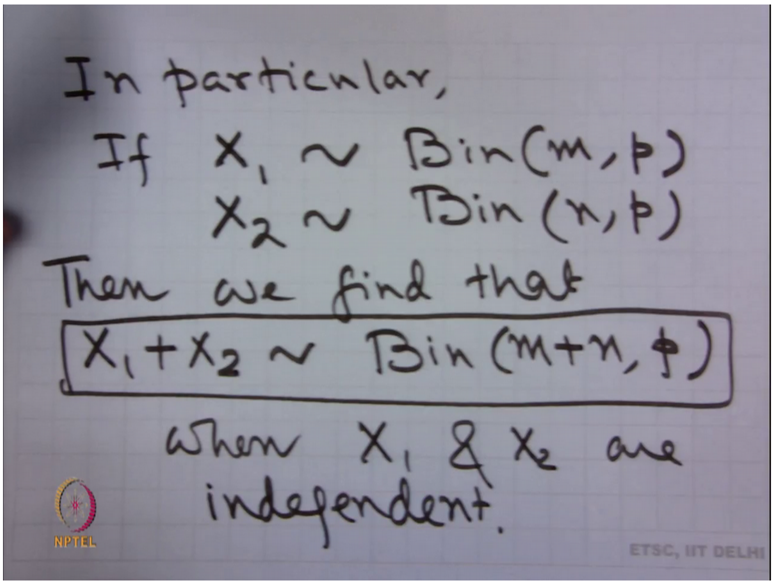
(Refer Slide Time: 16:53)



Therefore, what we get the number of defectives in each packet of 100 follow binomial with 100 p. So, we get the distribution of the number of defectives in each packet.

Thus we observe the following we may need to infer about the sum of 2 random variables, and we want to know it is probability distribution.

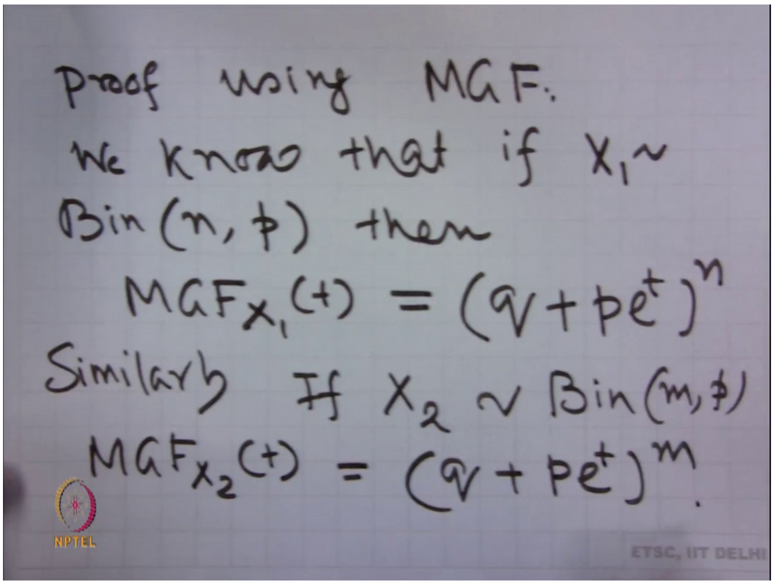
(Refer Slide Time: 18:39)



In particular,, if X 1 is a random variable which is binomial with parameters m and p and X 2 is a random variable which is binomial with n comma p, then we find that X 1 plus X 2 is binomial m plus n comma p; when X 1 X 2 are independent.

So, this is one result we have established. We can prove it in a slightly tricky way if we make use of moment generating function.

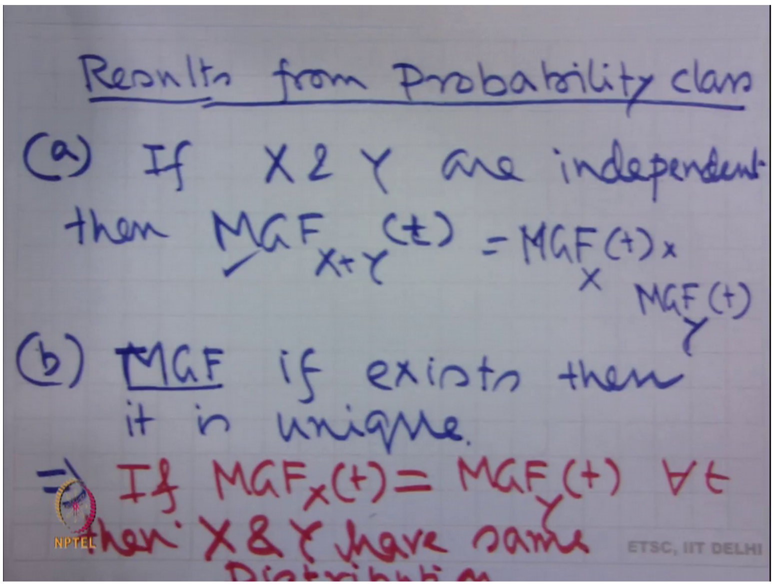
(Refer Slide Time: 19:58)



We know that if X 1 is binomial with n comma p, then MGF of X 1 is equal to; if you remember in the first class I have given you the formula it is q plus p e to the power t whole to the power n. Similarly, if X 2 follows binomial with m comma p then MGF of X 2 at t is equal to q plus p e to the power t whole to the power m.

Now, you must have seen 2 results in your probability class.

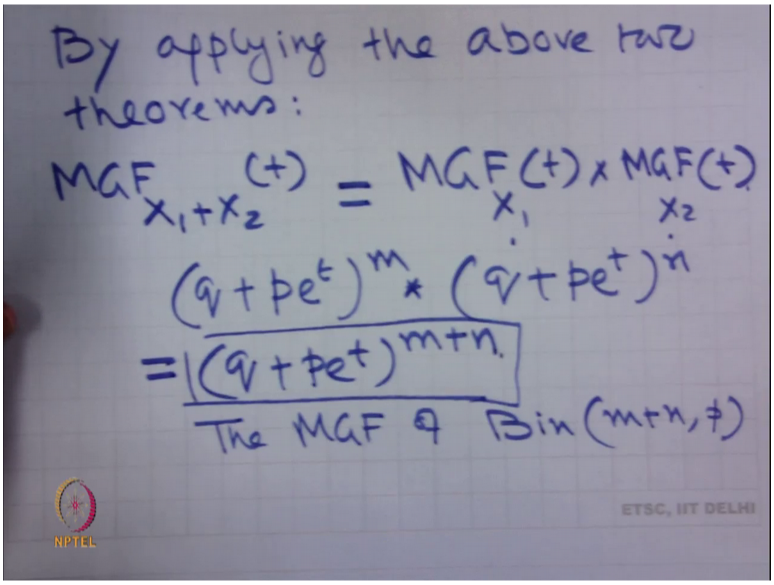
(Refer Slide Time: 21:27)



So, you must have seen these 2 results in your probability class that if x and y are independent then moment generating function of x plus y t is equal to the product of their individual moment generating functions. And the second result is that moment generating function if exists then it is unique.

There are distributions for each moment generating function may not exist, but if it exists then it is unique. What does it mean? It means that if MGF of X t is equal to MGF of y t for all t, then x and y have same distribution. That is if x and y are 2 random variables having the same moment generating function for all t. Then they have the same distribution.

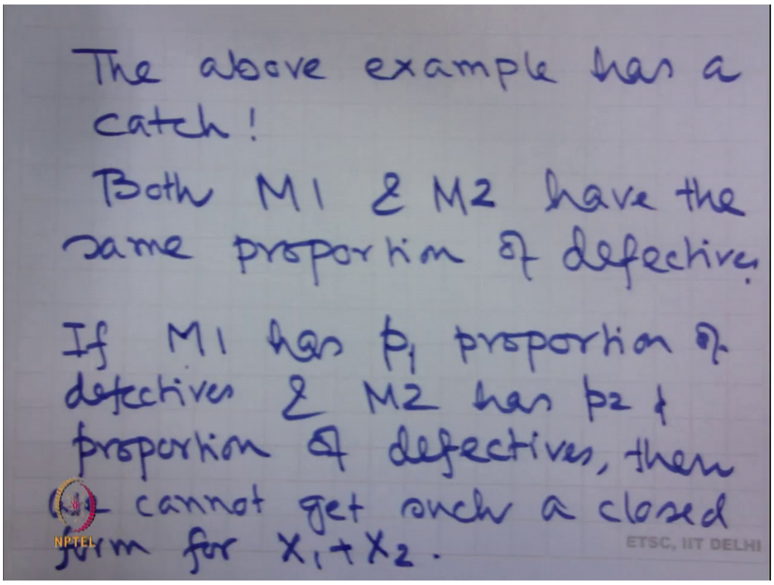
(Refer Slide Time: 23:57)



By applying these 2 theorem, we can see that moment generating function of X 1 plus X 2 at t is equal to moment generating function of X 1 at t into mg f of X 2 at t, and when they are binomial, then we can write it as q plus p e to the power t whole to the power m multiplied by q plus p e to the power t whole to the power n is equal to q plus p e to the power t whole to the power m plus n with respect to the problem that I was dealing with.

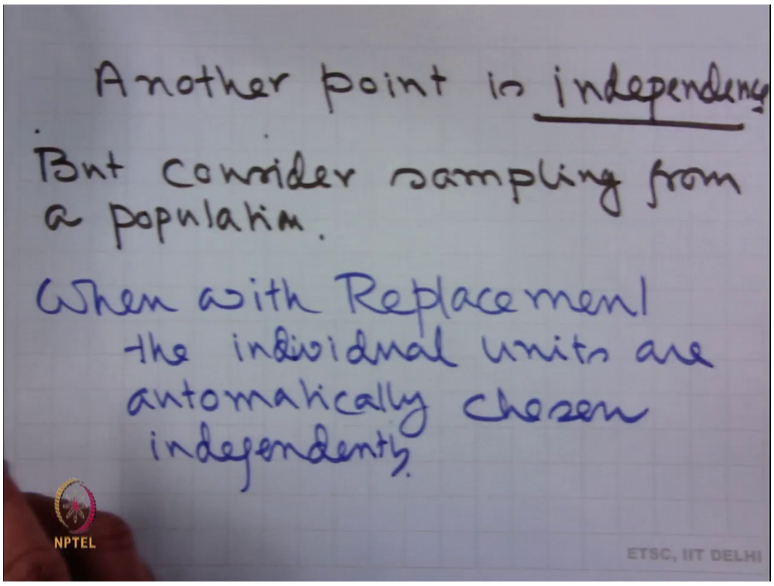
And this is precisely the moment generating function of binomial m plus np. So, from here also we can prove that if X 1 which is a binomial random variable telling you the number of defectives in a sample m from machine M 1 and X 2 is the number of defectives in a sample size n from machine M 2, then X 1 plus X 2 will have a distribution which is binomial m plus n p. So, the results that I have shown you earlier by explicit summation can also be proved using moment generating function. Subsequently, in many examples I may use these results directly.

(Refer Slide Time: 26:32)



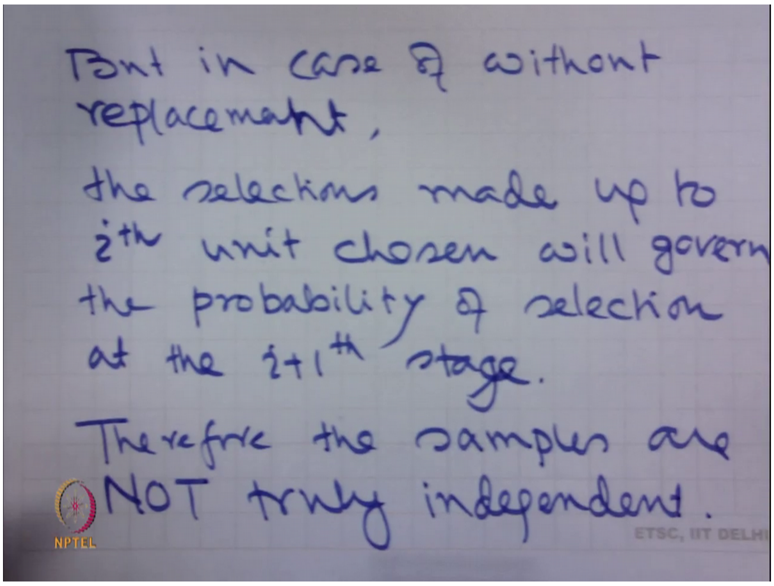
Now, let us examine the example what is that catch? That both M 1 and M 2 have the same proportion of defectives, if if M 1 has p one proportion of defectives and M 2 has p to proportion of defectives, then we cannot get such a closed form. Therefore, when we deal with summation of random variables, we will try to see cases so, that we can get a closed form for the distribution of the summation of the random variables.

(Refer Slide Time: 28:35)



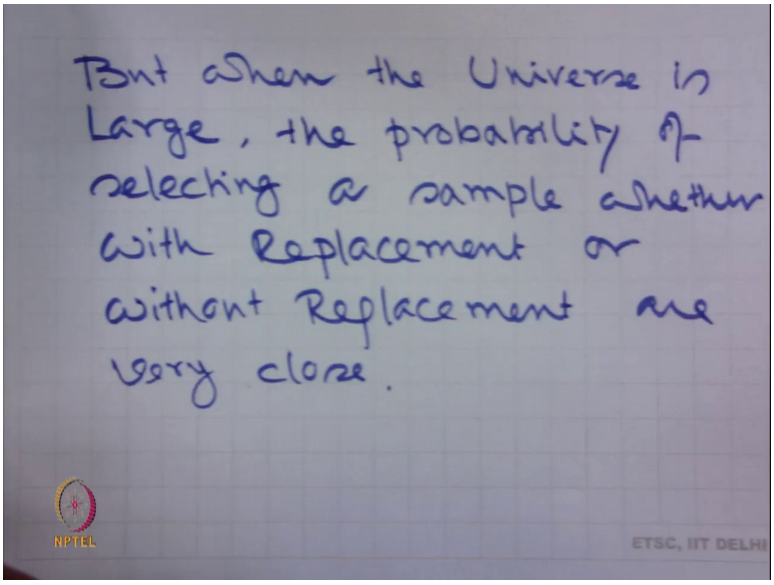
Another point is independence; when we are collecting samples from 2 different machines as we have done in the previous example, we know that they are going to be independent anyway. But suppose we are looking at sampling from a population. If the sampling is with replacement, then effectively each of the selection of items are independent.

(Refer Slide Time: 30:26)



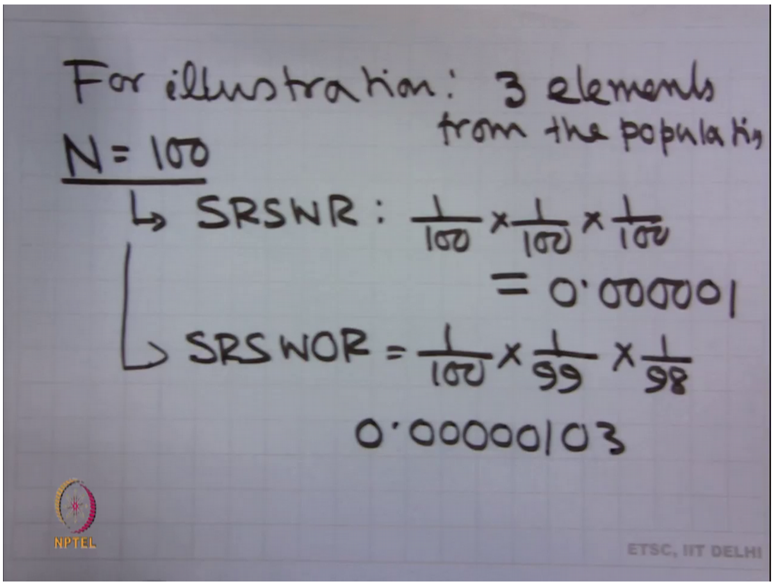
But if it is without replacement; obviously, the selection made up to ith unit chosen will govern the probability of selection at the i plus 1 th stage. This is very clear, because it is without replacement therefore, the items that I have chosen till the i th stage then when I am selecting the I plus 1 th element; obviously, it will depend upon the items that I have selected so far. Therefore, they are not truly independent.

(Refer Slide Time: 32:19)



But consider that the universe is very large, the probability of selection of a sample whether with replacement or without replacement are very close.

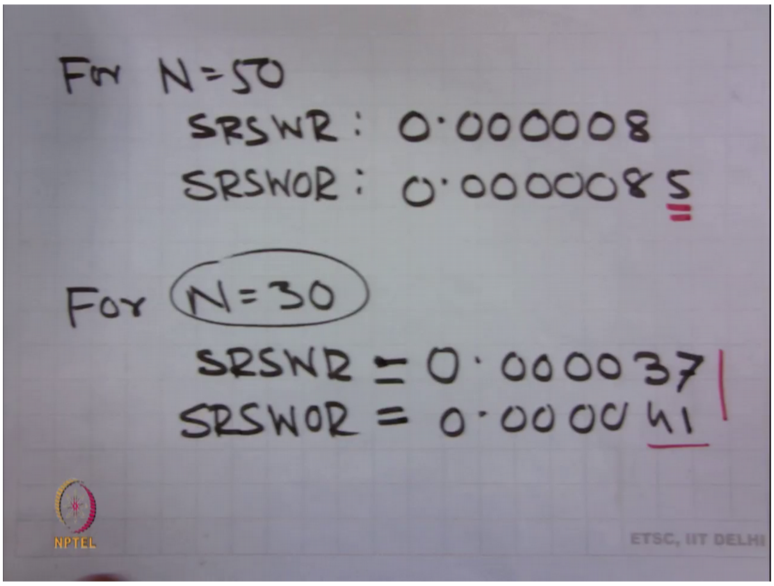
(Refer Slide Time: 33:22)



So, for your benefit I have calculated some of these probabilities. Consider n is equal to 100, and I am choosing 3 elements from the population. So, when n is equal to 100 under SRSWR that is simple random sampling with replacement, the probability of a triplet is 1 by 100 into 1 by 100 into 1 by 100, which is equal to 0.00001. Under SRSWOR this is going to be 1 by 100 into 1 by 99 into 1 by 98. This probability is going to be

0.00000103. So, if we look at their difference we find that the difference is coming in the 7th and 8h places of decimals.

(Refer Slide Time: 35:35)

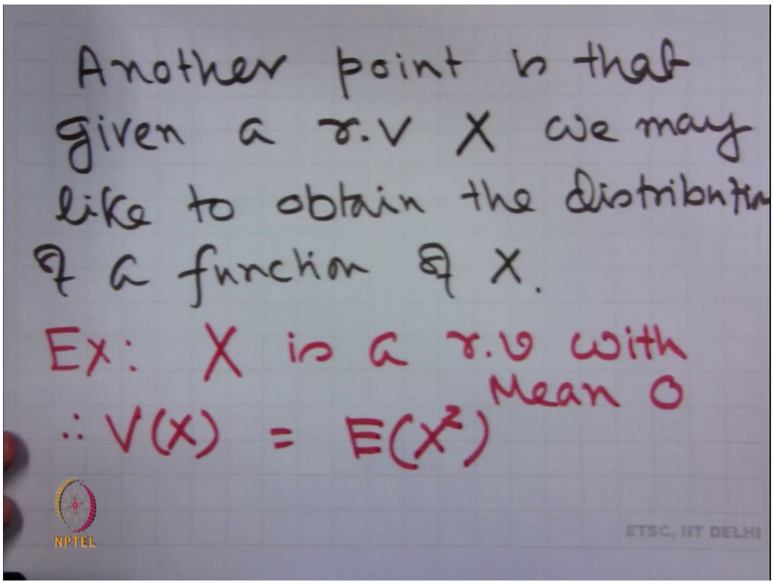


Similarly, for n is equal to 50 we get SRSWR 0.000008 SRSWOR it is going to be 0.0000085. So, again the difference is coming at the 7th decimal place. If we make N even smaller then what we are getting for SRSWR this is 0.000037 and for SRS WOR, this is equal to 0.000041.

So, even when the size of the population is as small as 30, we find that the probabilities are very, very close, they are differing only in the fifth decimal place so, that tells us something. So, that tells us that if the population is even reasonably large we may consider each sample to be taken from the population to be independent of the other samples to be taken.

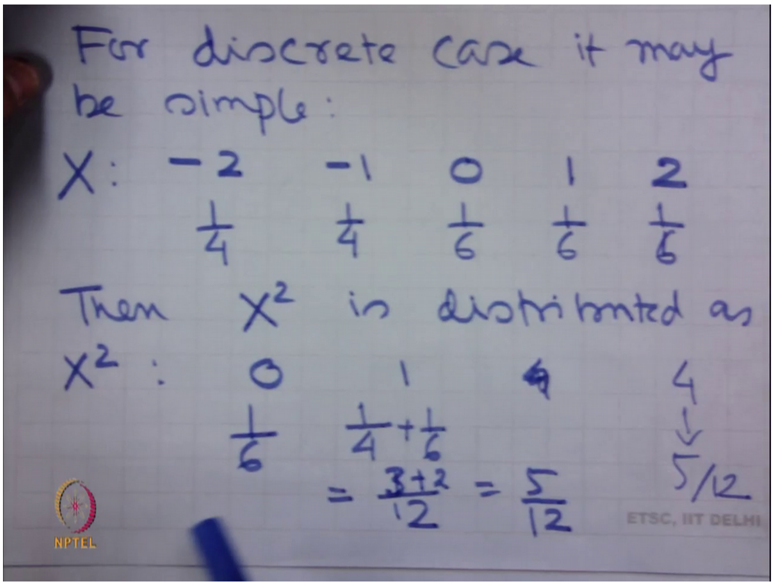
So, what is the advantage? The advantage is that we can use the independence of probability by multiplying the individual probabilities to obtain the probability of selection of a particular sample of size n.

(Refer Slide Time: 37:58)



Another important point is that given a random variable x we may like to obtain the distribution of function of x. For example, x is a random variable with mean 0. Therefore, what is the variance of x is equal to expected value of x square. So, given that distribution of x we are trying to find out the expectation of a function namely square of the random variable. For discrete case this may be simple.

(Refer Slide Time: 39:29)



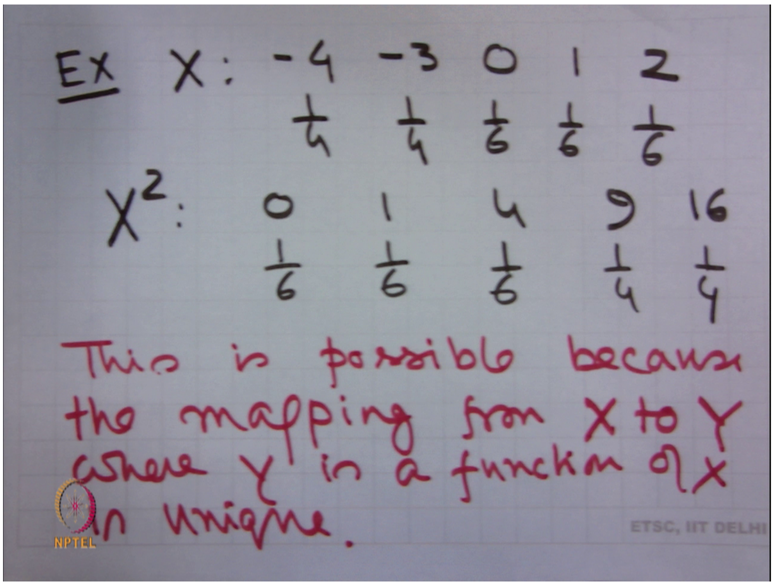
For example, consider x which takes values minus 2, minus 1, 0 1 and 2, with probabilities 1 by 4, 1 by 4, 1 by 6, 1 by 6, 1 by 6. Then how is x squared distributed? X square can take values only 3 possibilities 0 1 and 4. Probability of 0 is x is equal to 0 therefore, it is 1 by 6.

Probability x squared is equal to 1, that we can get in 2 different ways when x is minus 1 whose probability is 1 by 4, and x is plus 1 whose probability is 1 by 6. Therefore, probability x square is equal to 1 is 1 by 4 plus 1 by 6 which is equal to 5 by 12.

Similarly, probability x squared is equal to 4 will coming out to be 5 by 12. Therefore, from the distribution of x, we can get the distribution of x square to be 1 by 6, 5 by 12 for 1 and 5 by 12 for 4. But now suppose I ask you given this distribution of x square can you find the distribution of x then you see that you cannot. Because this tells me that probability x is equal to minus 1 plus probability x is equal to plus 1 that sum is 5 by 12. But there is no way I know the decomposition of this into the 2 values which are the respective probabilities of x is equal to minus 1 and x is equal to plus 1.

The question is why we cannot do it in this case? But we can do in some other case.

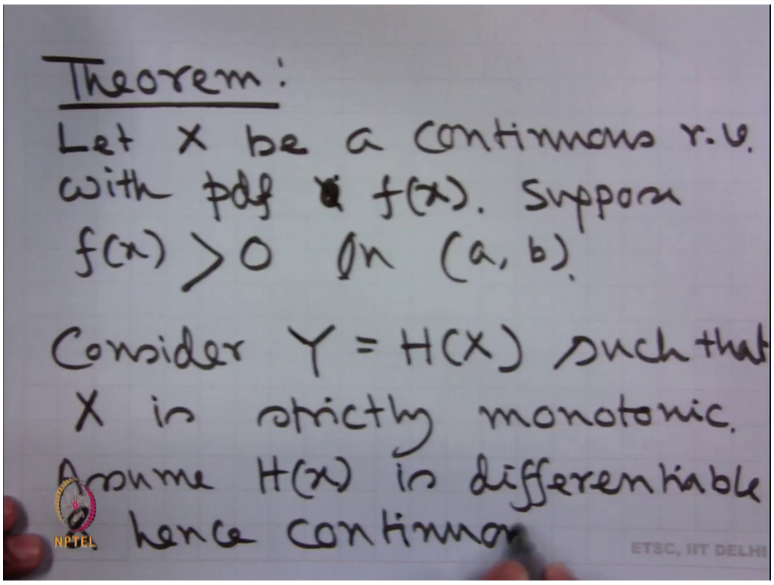
(Refer Slide Time: 42:48)



Say for example, x takes values minus 4 minus 3 0 1 2, and suppose their probabilities are 1 by 4, 1 by 4, 1 by 6, 1 by 6, 1 by 6. Then x square can take values 0 1 4 9 and 16 with probabilities 1 by 6, 1 by 6, 1 by 6, 1 by 4 and 1 by 4.

In this case, I can get the probabilities of different values of x. This we can do because the mapping from x to the function y; where y is the function of x is unique. Given x I know the x square, and given x square I know which value of x has produced this value for x squared. Thus when the mapping is unique it helps us to compute the pdf of a function hx of the random variable x from the distribution of the original random variable namely x.

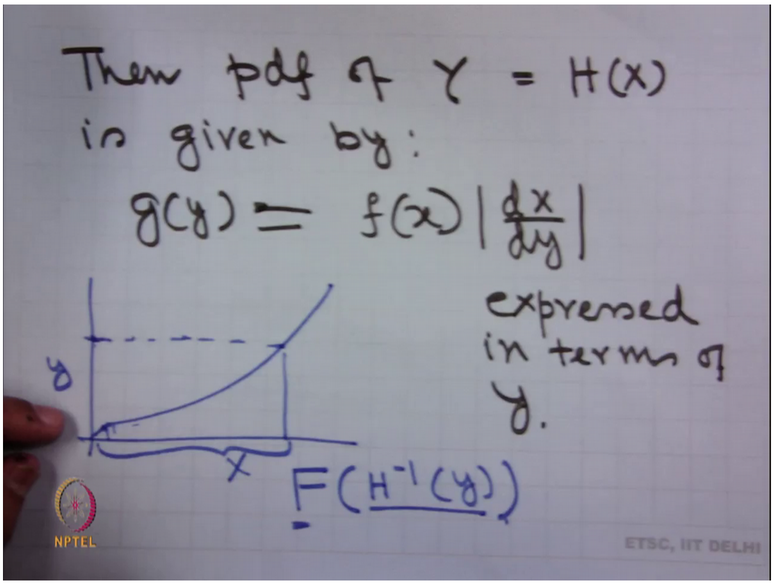
(Refer Slide Time: 44:59)



I appeal to the following theorem which many of you must have seen; Let x be a continuous random variable with pdf f of x, and suppose f x is greater than 0 on an interval a b.

Now, consider Y is equal to H x, that is Y is a function of X; X is strictly monotonic. So, from x we are mapping it into a random variable Y which is strictly monotonic. Assume H x is differentiable and hence continuous for all X, then pdf the probability density function of Y is equal to H of x is given by g of y is equal to f of x multiplied by dx dy.

(Refer Slide Time: 47:02)

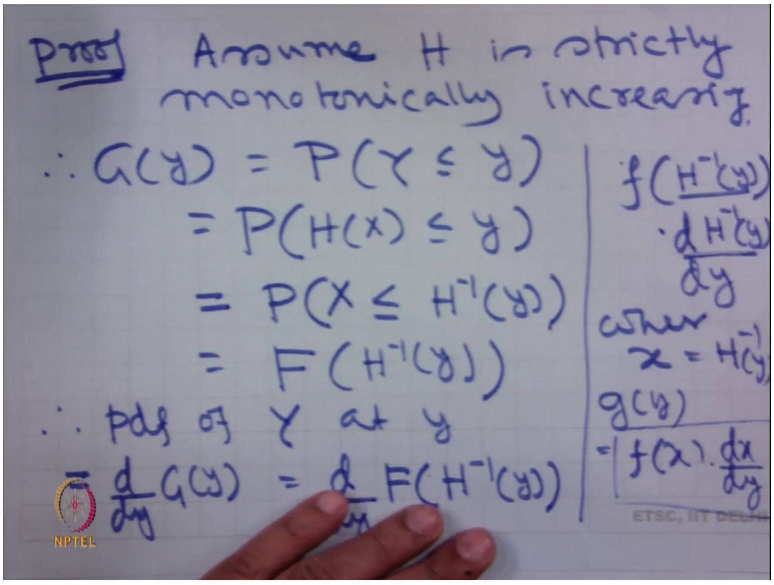


The mod value of that expressed in terms of y.

Now, why did we take strictly monotonic whether increasing or decreasing. Suppose this is x and this is a strictly monotonic function of x, then given y what is the probability that the random variable y is less than equal to a particular value y? So, that probability will come from the probability of x. So, that probability will be given by F of H inverse of Y. The strict monotonicity allows us to make the inverse of Y and to get the value of x, and therefore, we can apply the cumulative distribution function of x on that one.

So, let me give a very quick proof of the above statement.

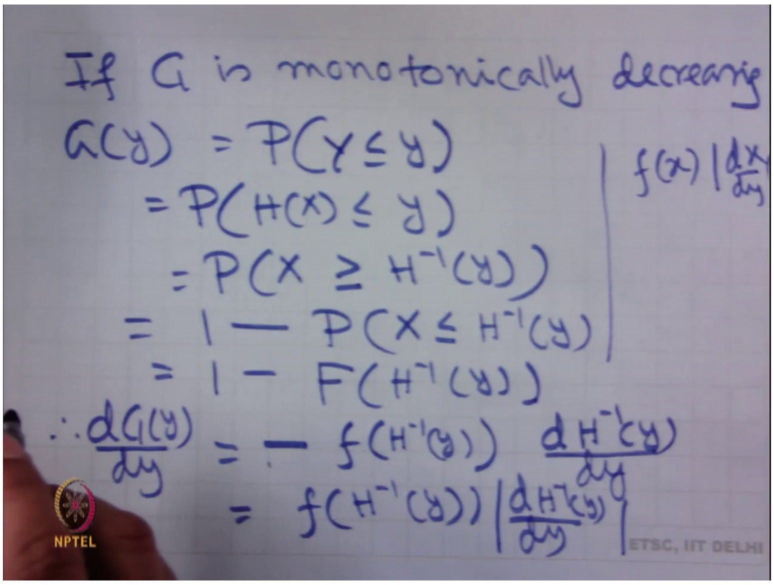
(Refer Slide Time: 49:07)



So, we have taken an increasing function. Therefore, G of y which is probability y less than equal to y is same as probability H of X less than equal to y, and since it is strictly monotonic we can get the inverse of y uniquely. And therefore, this is a fact H inverse Y. Therefore, what is the pdf of y? We get it by differentiating the cumulative distribution function.

So, this is d dy of f of H inverse y; which is multiplied by d H inverse y d y where x is equal to H inverse y. Therefore, we get it that g of y is equal to f of x H inverse y is equal to x into dx dy and this we express using y.

(Refer Slide Time: 52:00)

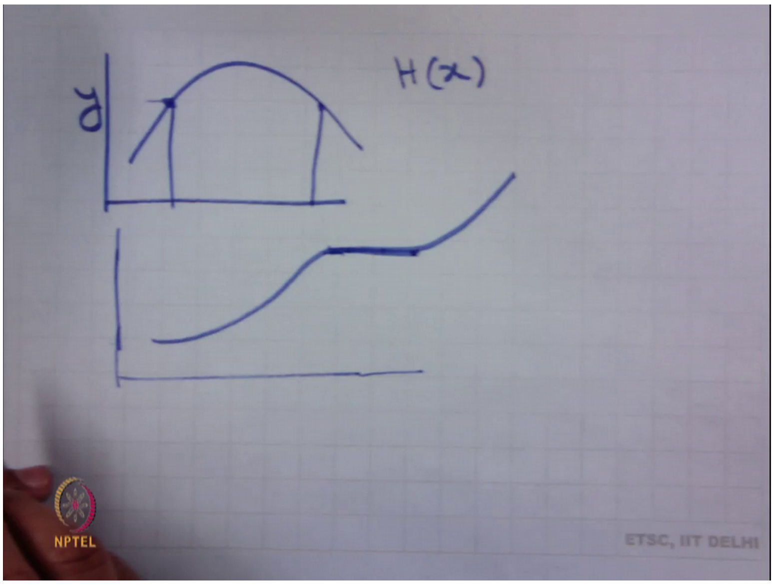


If G is monotonically decreasing, then what will happen? G of y is equal to probability Y less than equal to Y is equal to probability H of x less than equal to y is equal to probability x greater than equal to H inverse y, this is because it is decreasing.

Therefore, inequality will be reversed is equal to 1 minus probability X less than equal to H inverse y is equal to 1 minus F at H inverse y. Therefore, d G y dy is equal to minus f at H inverse y which is f into d H inverse y dy, and we know that if it is a decreasing function, then this derivative is going to be negative. Therefore, that multiplied by this minus sign will give you the modulus of the derivative, right? Or we can write it as fx into mod of dx dy.

So, if x is monotonically increasing or monotonically decreasing? We can get the inverse that is not the case.

(Refer Slide Time: 54:14)



Say for example, H of x is like this, then given a y, I can get 2 different values of x and we cannot take the inverse of y. In a similar way, if it is something like this although it is increasing, but on this range the value is fixed then also for these values of y we cannot actually get the inverse of the H function.

And therefore, in these cases we have to apply some tricks to partition it into non overlapping segments, and from there we will have to calculate the pdf of the function of x. With that I stop here. In the next class I will use this theorem, and I will show a case of this type to show how we can obtain the distribution of some functions particularly of normal variable, ok. So, with that I stop here; see you in the next class.

Thank you.