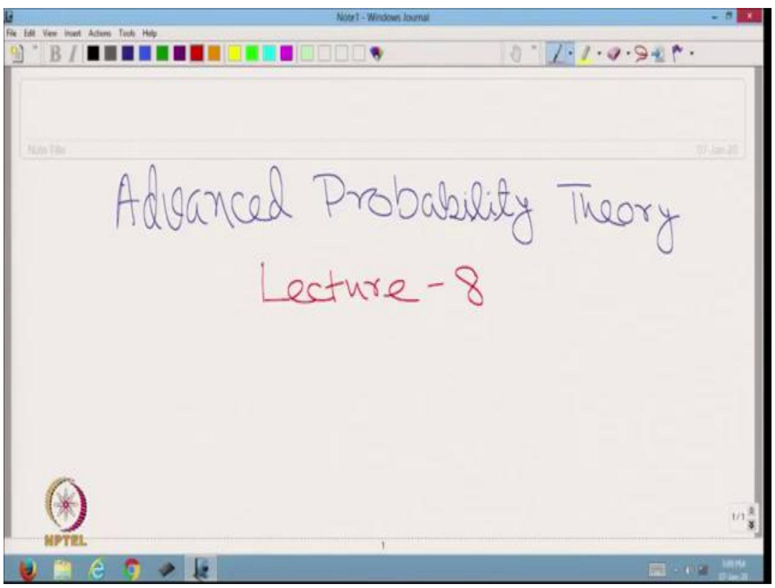
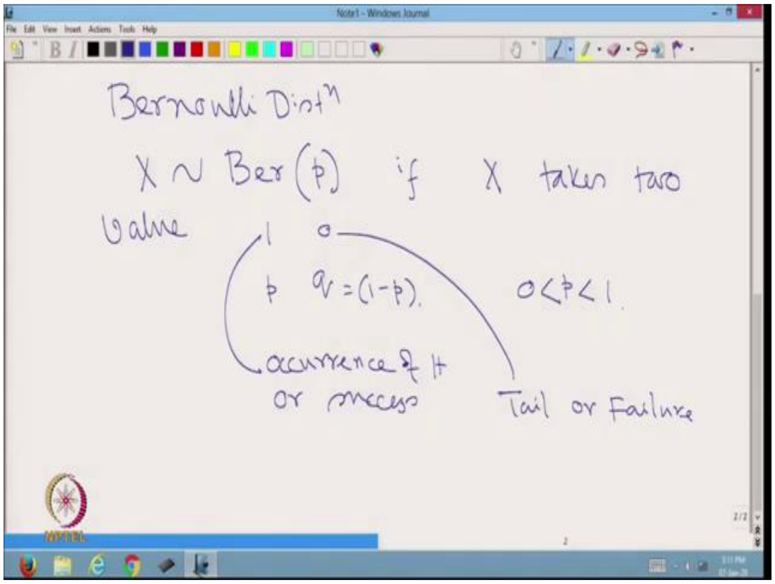
Advanced Probability Theory
Professor Niladri Chatterjee
Department of Mathematics,
Indian Institute of Technology, Delhi
Lecture 8

Welcome students. The MOOC's lecture series on Advanced Probability Theory. This is lecture number eight. In the last lecture, if you remember, we have started discussions on discrete random variables. In particular, we have studied binomial distribution and Poisson distribution. In today's lecture, we shall look at a few more discrete random variables or discrete distributions, which are very important from mathematical as well as application point of view.

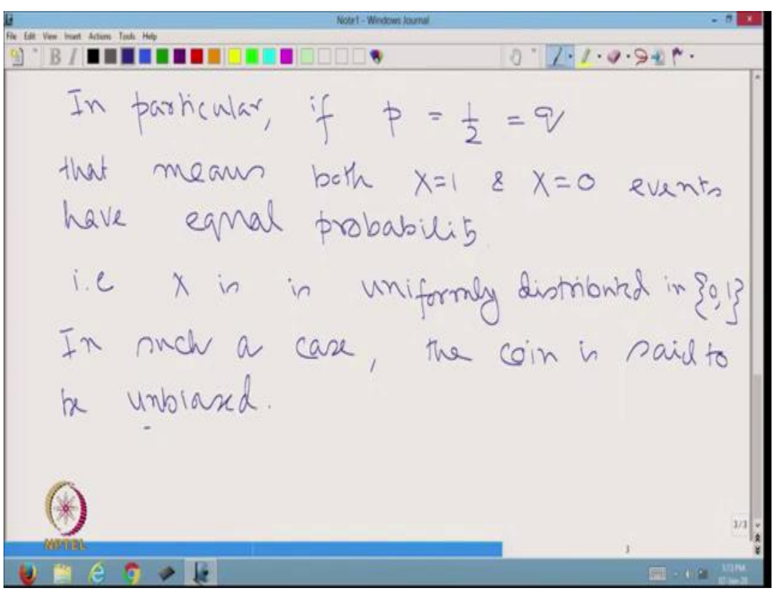
(Refer Slide Time: 1:00)





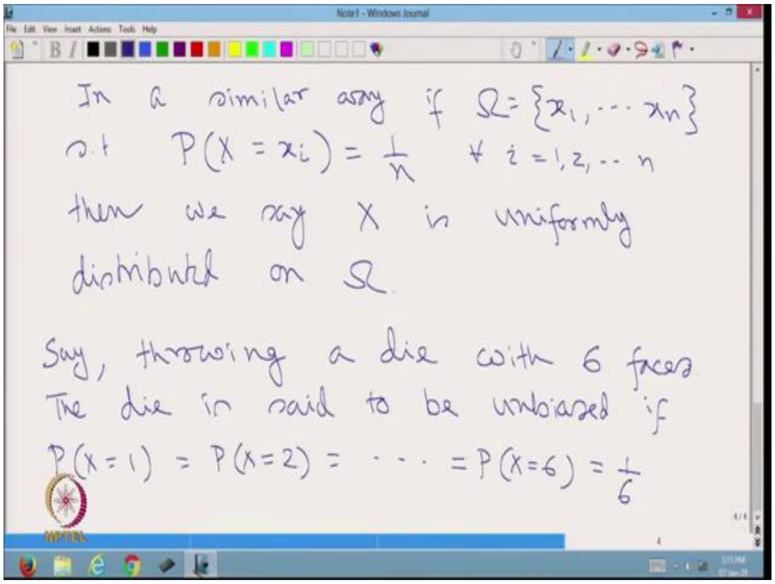
So, if you remember, we started with Bernoulli distribution that is x is distributed as Bernoulli p if x takes 2 values, one with probability p and 0 with probability q, which is equal to 1 minus p, 0 less than p less than 1. So, this we have used to model the toss of a coin, whereby 1 we may occurrence of head or a success and by 0 we mean tail or failure. This we have discussed already.

(Refer Slide Time: 2:20)



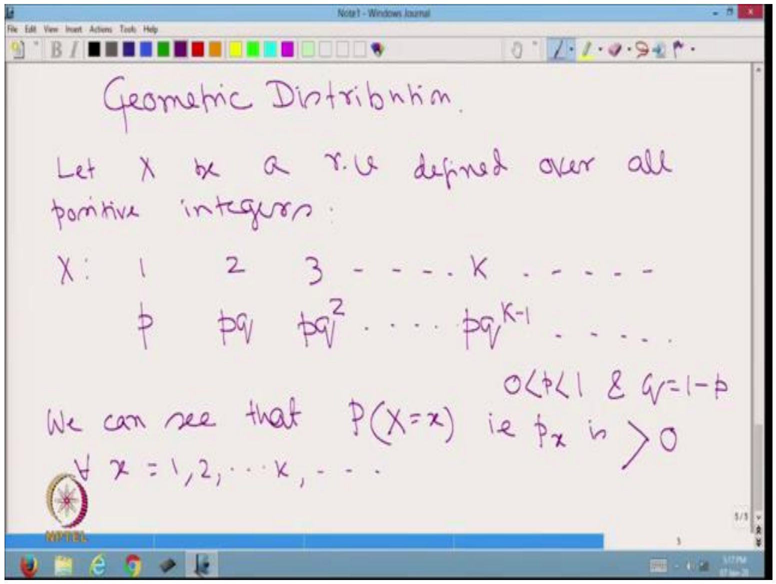
In particular, if p is equal to half is equal to q that means both x is equal to 1, and x is equal to 0. These two events have equal probability or in other words, x is uniformly distributed in 0 and 1, this is a set of two values and in such a case the coin is called, the coin mean the coin that is being used for tossing unbiased.

(Refer Slide Time: 3:52)



In a similar way, if omega is equal to x1, x2, xn such that probability X is equal to xi is equal to 1 by n for all i is equal to 1, 2 up to n. Then we say X is uniformly distributed on omega. Say for example, you are throwing a die with 6 faces. The die is said to be unbiased if probability X is equal to 1 is equal to probability X is equal to 2 is equal to probability X is equal to 6 is equal to 1 by 6.

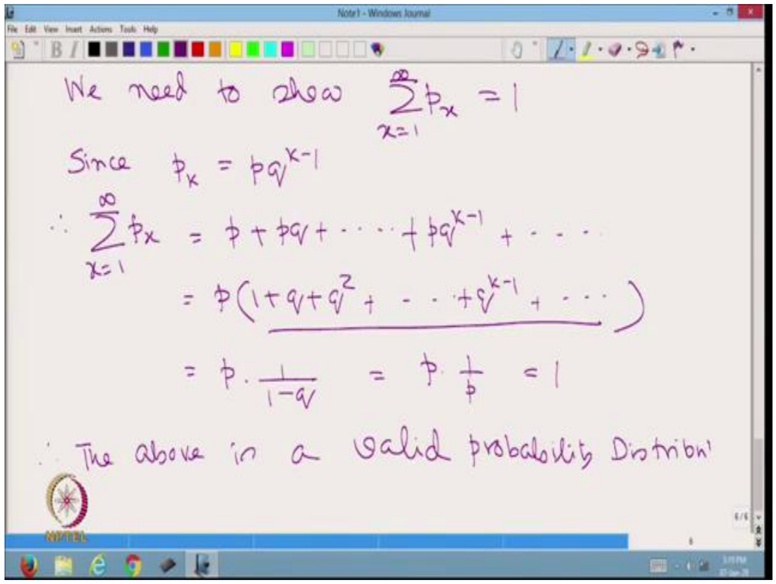
(Refer Slide Time: 5:41)



Now, let us study under distribution, which is called Geometric Distribution. Let X be a random variable defined over all positive integers. So, x takes the values 1, 2, 3, K and it goes to infinity, probability X is equal to 1 is p, probability X is equal to 2 is pq, probability X is equal to 3 is pq square like that probability X is equal to K is p q to the power K minus 1 here, the first question is it a valid probability distribution?

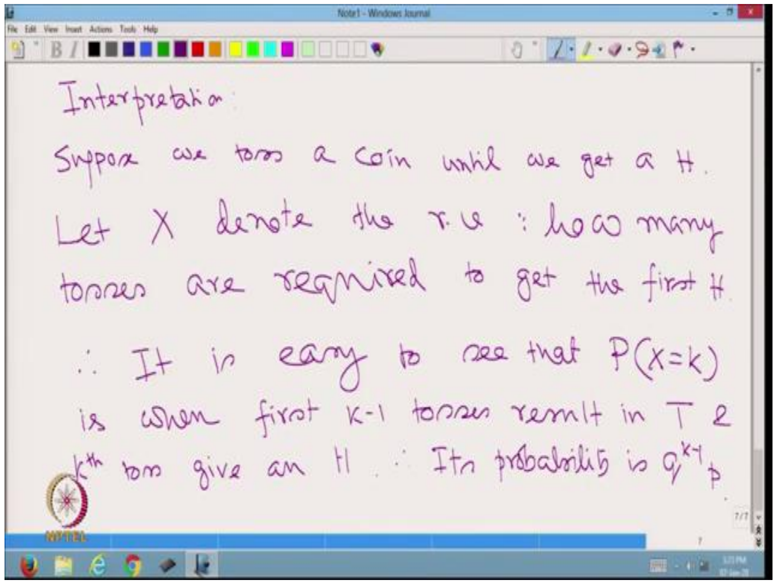
So, let us assume that 0 less than p less than 1 and q is equal to 1 minus p. Therefore, we can see that probability X is equal to X which we may denote as px is greater than 0 for all x is equal to 1 to K up to infinity, because p is positive, q is positives, so their product is positive.

(Refer Slide Time: 7:57)



Next we need to show sigma px, x is equal to 1 to infinity is equal to 1, that is very easy to show since pk is equal to p q to the power K minus 1. Therefore, sigma px x is equal to 1 to infinity is equal to p plus pq plus the p q to the power K minus 1 up to infinity is equal to p into 1 plus q plus q square plus q to the power K minus 1 plus up to infinity. Now, this is a geometric series. Therefore, we know that its sum is going to be 1 upon 1 minus q is equal to p into 1 upon p is equal to 1. Therefore, the above is a valid probability distribution.

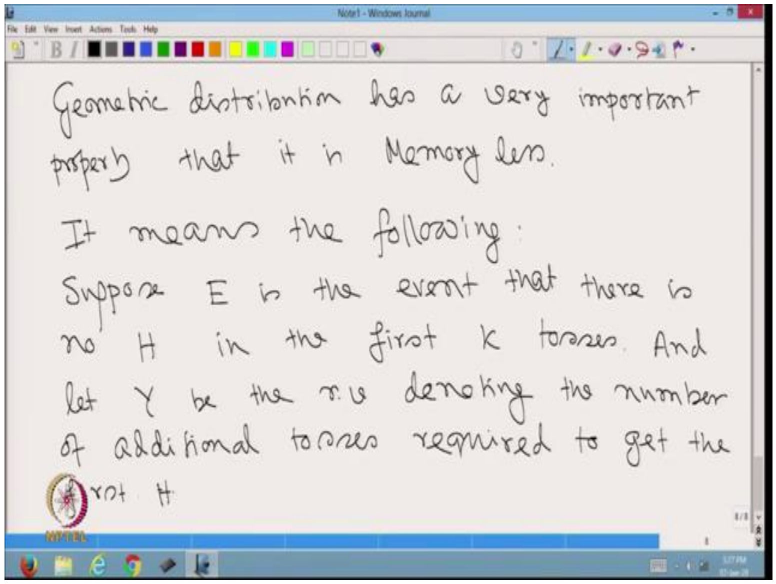
(Refer Slide Time: 9:32)



So, its interpretation. Suppose, we toss a coin until we get a Head. So, how many tosses are required? Let X denote the random variable, how many tosses are required to get the first Head. So, it may take 1, when the very first toss we get head and obviously its probability is p, if we get Tail at the first toss that probability is q and then a head in the second toss then that probability is p. Therefore, we need two tosses to get the first head and its probability is therefore p q.

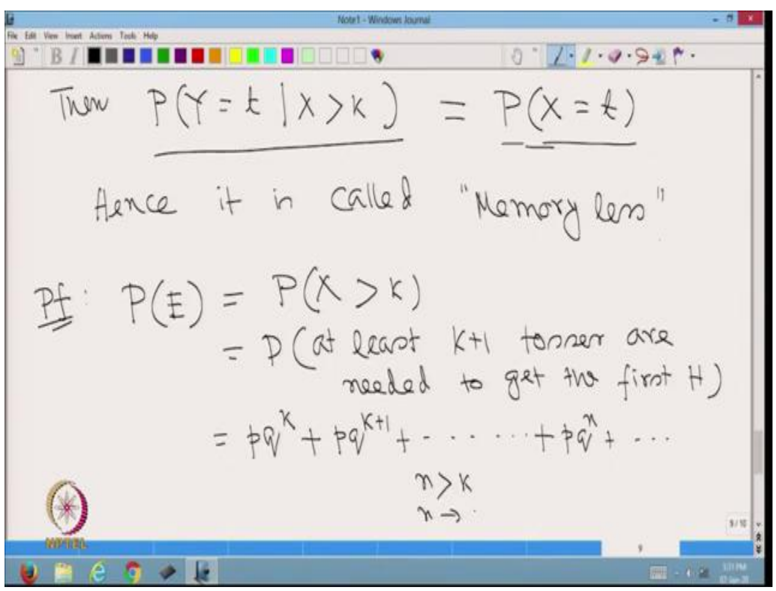
Similarly, if the value of X is K that means that we are tossing the coin K times to get the first head implies that first q minus 1 tosses are going to be tail and the Kth toss is going to be head. Therefore, it is easy to see that probability X is equal to K is when first K minus 1 tosses resulting tail and Kth toss gives them H therefore, its probability is q to the power K minus 1 into p. So, that is the basic interpretation of Geometric distribution.

(Refer Slide Time: 12:24)



Now, geometric is very important, has a very important property that it is memory less. What does it mean? It means the following: Suppose E is the event that there is no head in the first K tosses. And let Y be the random variable denoting the number of additional tosses required to get the first head.

(Refer Slide Time: 14:33)

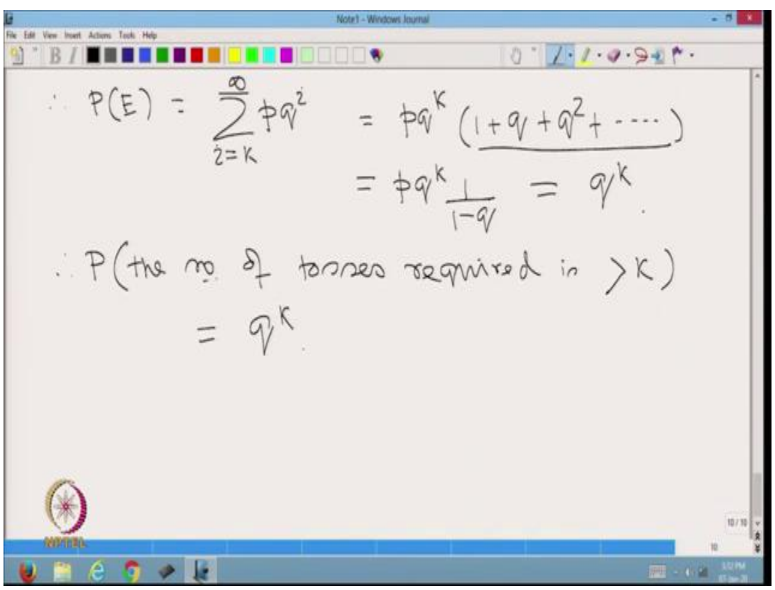


Then probability Y is equal to t given x is greater than k is same as probability X is equal to t. So, let me explain this expression, given that no head has occurred in the first K tosses. Therefore, X has to be greater than K and Y denotes the number of additional tosses needed to get the first head. So, it says that the probability of having another t tosses to get the first head, given that X is greater than k.

That is same as from the beginning we need t tosses to get the first head or in other words, that we have already consumed K many tosses, it has no effect on the probability of what is going to happen subsequently. Hence, it is called memory less. So, let us prove that so, what is the probability of event E? It is probability X is greater than K is equal to probability at least K plus 1 tosses are needed to get the first head.

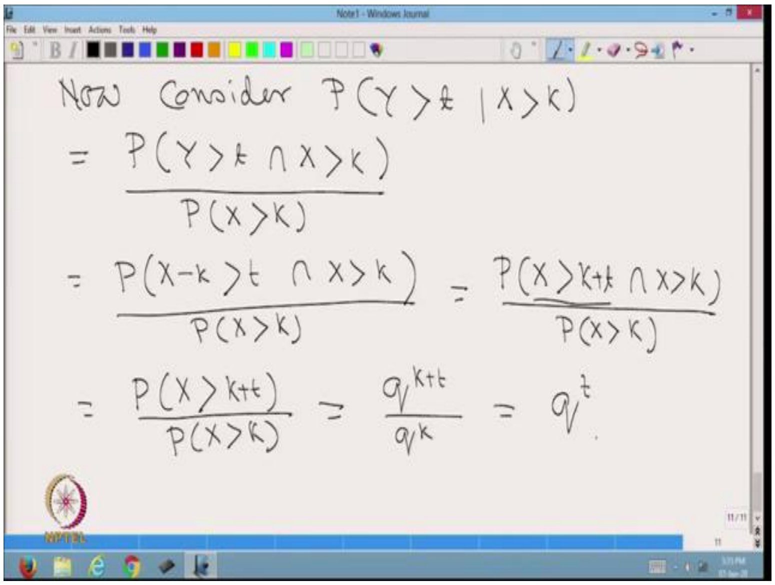
Is equal to pq to the power K that we need K plus 1 tosses to get the first head plus pq to the power K plus 1 denoting that we need K plus too many tosses to get the first head, where n is greater than K and n is going to infinity.

(Refer Slide Time: 17:20)



Therefore, probability of E is equal to sigma pq to the power i, i is equal to K to infinity is equal to pq to the power K into 1 plus q plus q square up to infinity is equal to pq to the power K, this sum we already know is equal to 1 upon 1 minus q which is equal to q to the power K. Therefore, probability the number of tosses required is greater than K is equal to q to the power K.

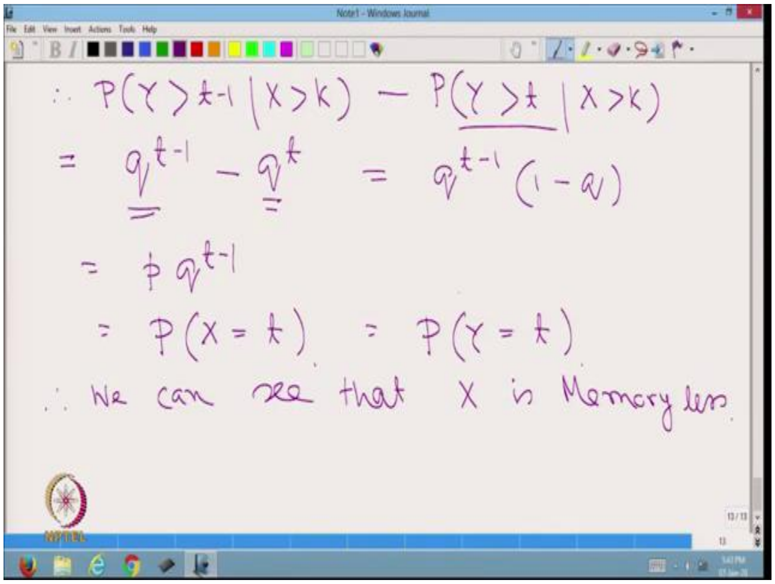
(Refer Slide Time: 18:26)

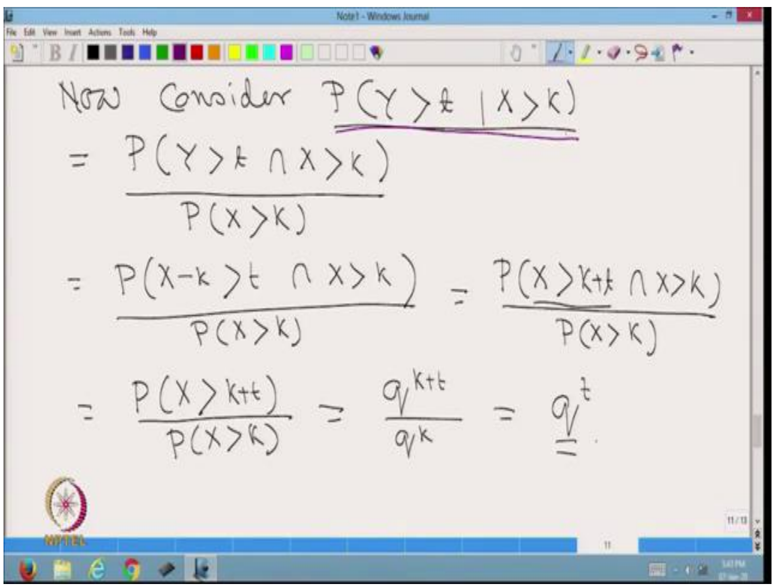


Now, let us consider probability Y is greater than t given that X is greater than K. This conditional probability we know is equal to probability Y greater than t and X is greater than K upon probability X greater than K is equal to probability X minus K greater than t and X greater than K upon probability X greater than K is equal to probability X greater than K plus t and X greater than K divided by probability X greater than K. Now, probability X greater than K plus t and X greater than K, these means they are intersession is this event.

Therefore, this is going to be probability X greater than K plus t divided by probability X greater than K. Now, probability X greater than K, we have already calculated to be q to the power K. Therefore, probability X greater than K plus t is going to be q to the power K plus t divided by q to the power K is equal to q to the power t. So, probability Y greater than t given X is greater than K, that probability is coming out to be q to the power t.

(Refer Slide Time: 20:37)

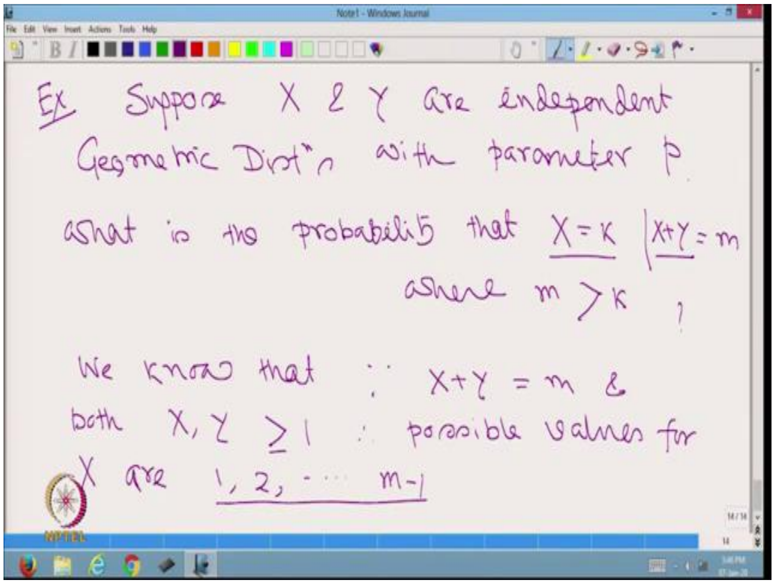




Therefore, probability Y greater than t minus 1 given X greater than K minus probability Y greater than t given X greater than K is equal to q to the power t minus 1 minus q to the power t. This is because we have seen that probability Y greater than t given X greater than K is q to the power t.

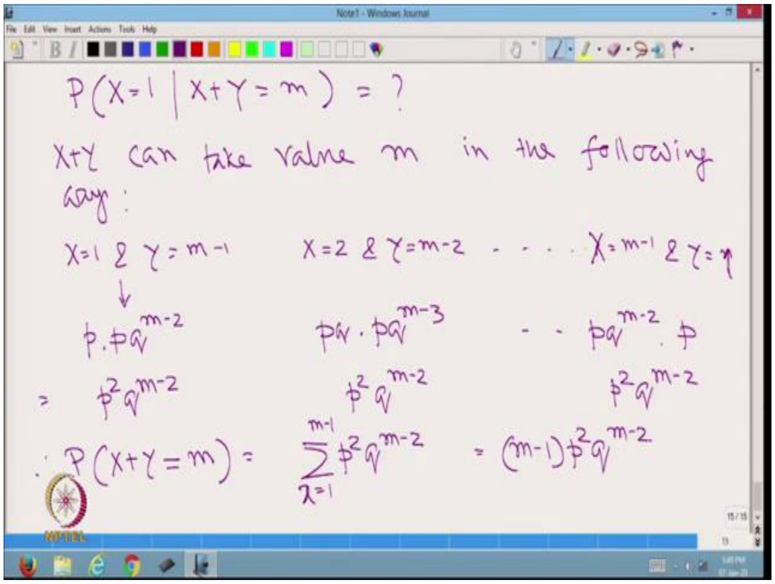
Therefore, Y greater than t minus 1 given X greater than K is q to the power t minus 1, in a very similar way, this is q to the power t. Therefore, this is q to the power t minus 1 to 1 minus q is equal to pq to the power t minus 1 which is equal to probability X is equal to t is equal to probability Y is equal to t. Therefore, we can see that X is memory less.

(Refer Slide Time: 22:03)



Another example suppose, X and Y are independent geometric distribution with parameter P. What is the probability that X is equal to K given that X plus Y is equal to m where m is greater than K? So, we are given that the sum of two geometric random variables X and Y is m, given that what is the probability that X is equal to K? We know that since, X plus Y is equal to m and both X and Y are greater than equal to 1, therefore, possible values for X are 1, 2, 3 up to m minus 1. So, X can take one of the possible values between 1, 2, 3 up to m minus 1.

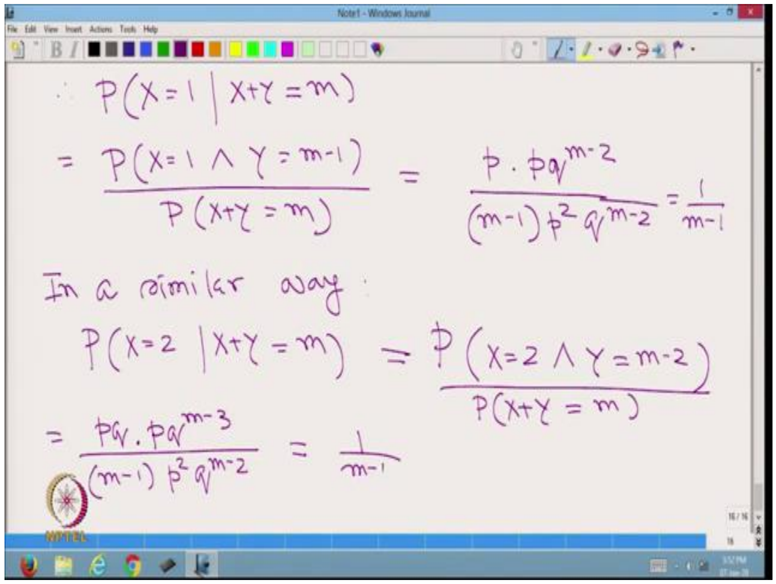
(Refer Slide Time: 24:14)

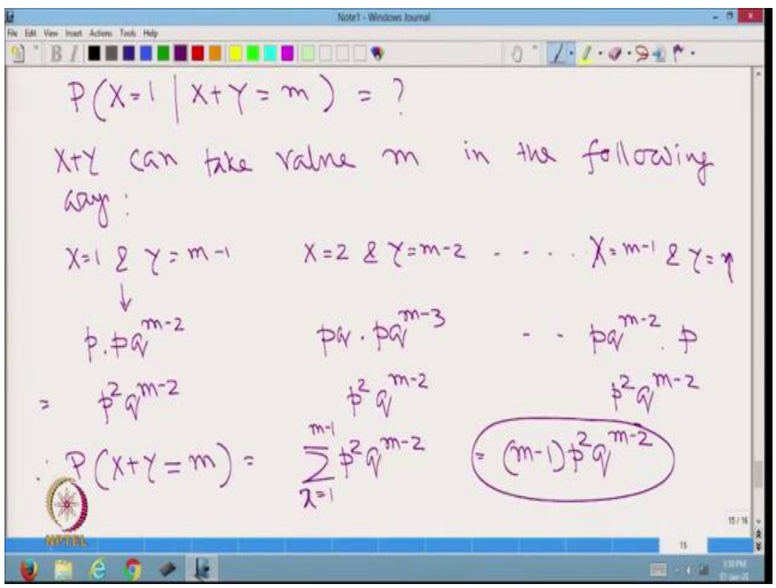


So, probability X is equal to 1, given X plus Y is equal to m is equal to how much? X plus Y can take value in the following ways. X is equal to 1, and Y is equal to m minus 1, X is equal to 2, and y is equal to m minus 2, up to X is equal to m minus 1 and Y is equal to 1. So, what is this probability? It is p multiplied by pq to the power m minus 2, this is pq multiplied by pq to the power m minus 3.

This is pq to the power m minus 2 multiplied by p is equal to, if you note that all of them are p square q to the power minus 2, this is also p square q to the power m minus 2 and this is also p square q to the power minus 2. Therefore, probability X plus Y is equal to m has the probability sigma p square q to the power m minus 2, x is equal to 1 to m minus 1 is equal to m minus 1 into p square q to the power m minus 2.

(Refer Slide Time: 26:36)

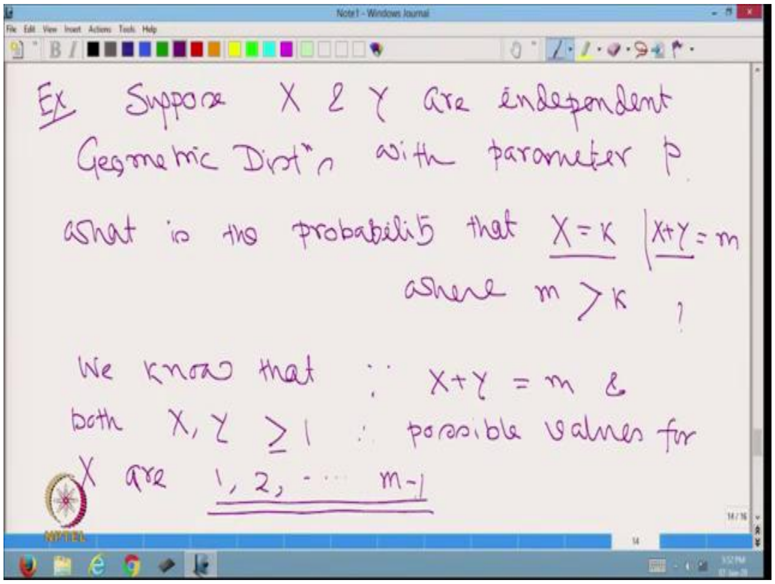




Therefore, probability X is equal to 1 given X plus Y is equal to m is equal to probability X is equal to 1, and Y is equal to m minus 1 divided by probability X plus Y is equal to m is equal to p multiplied by p into q to the power m minus 2 divided by this probability m minus 1 into p square into q minus 2. So, m minus 1 into p square into q to the power m minus 2 is equal to 1 upon m minus 1.

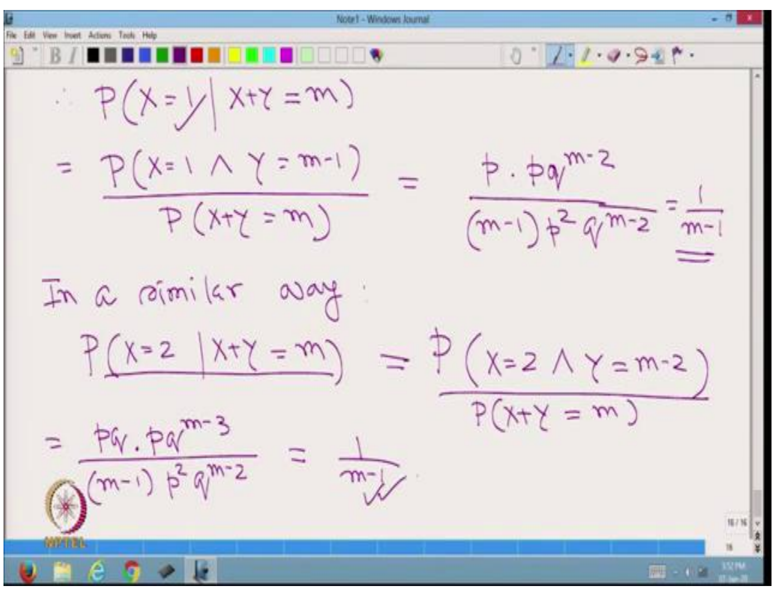
In a similar way probability X is equal to 2 given X plus Y is equal to m is same as probability X is equal to 2 and Y is equal to m minus 2 upon probability X plus Y is equal to m is equal to pq multiplied by pq to the power m minus 3 upon m minus 1 into p square into q to the power m minus 2 is equal to 1 upon m minus 1.

(Refer Slide Time: 28:49)



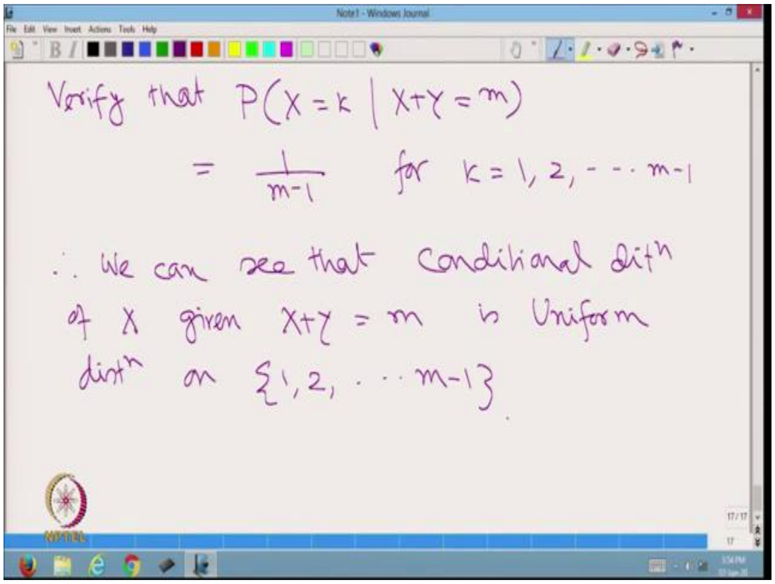
We have observed that X can take values between 1, 2, 3 up to m minus 1.

(Refer Slide Time: 28:52)



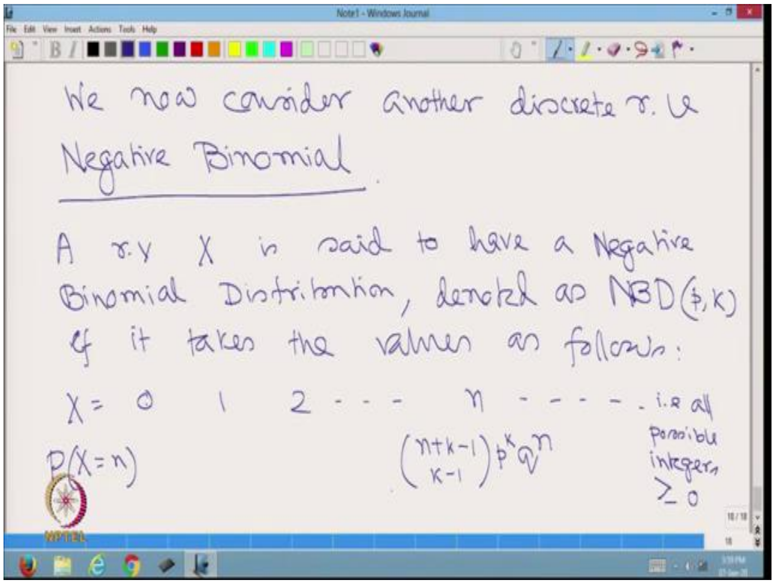
Now, we have seen that probability X is equal to 1 given X plus Y is equal to m is 1 upon m minus 1, probability X is equal 2 given X plus Y is equal to m that is also 1 upon m minus 1.

(Refer Slide Time: 29:12)



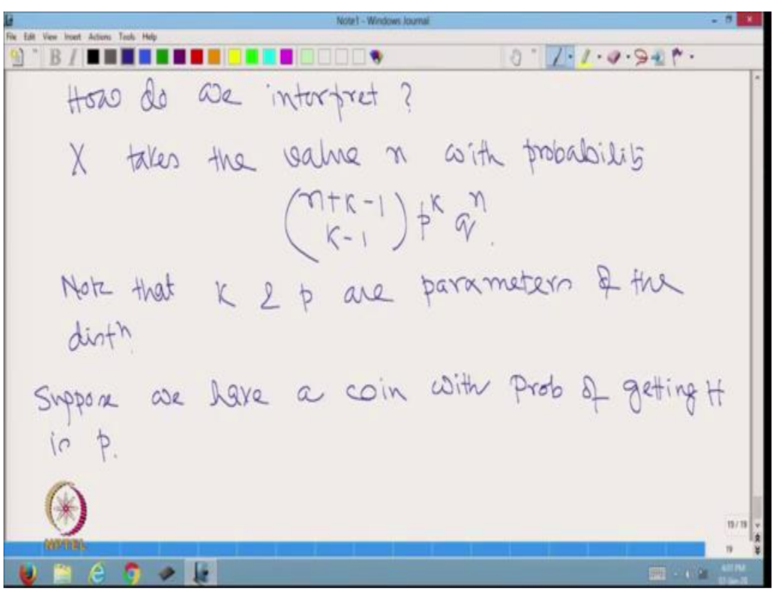
Verify that probability X is equal to k given X plus Y is equal to m is equal to 1 upon m minus 1, for k is equal to 1, 2, 3 up to m minus 1. Therefore, we can see conditional distribution of X given X plus Y is equal to m is uniform distribution on 1, 2 up to m minus 1. So, that is another interesting property for geometric distribution.

(Refer Slide Time: 30:21)



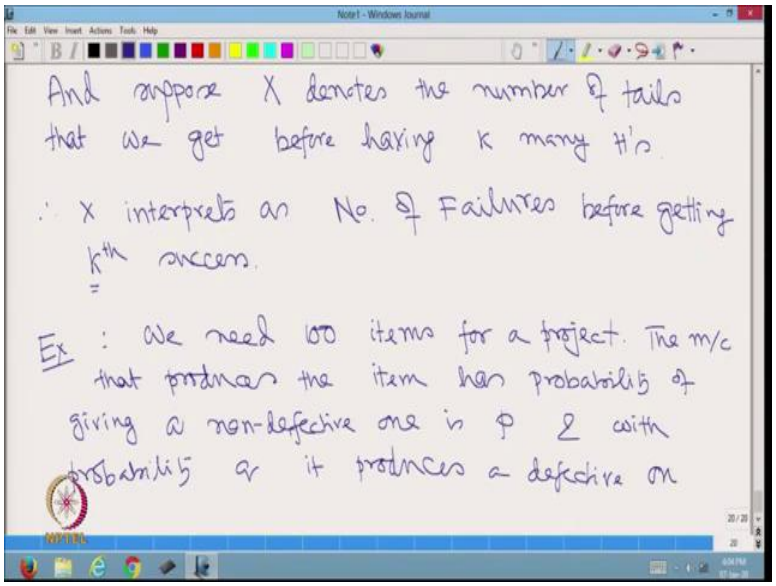
Now, let me move further. We now consider another discrete random variable, namely Negative Binomial. So, a random variable X is said to have a negative binomial distribution denoted as NBD, with two parameters p and k as follows. So, X takes the values 0, 1, 2 then up to infinity that is all possible integers greater than equal to 0. So, X takes the value of 0, 1, 2 up to n, all possible integers and the probability X is equal to n is n plus k minus 1, see k minus 1 p to the power k q to the power n.

(Refer Slide Time: 32:53)



What does it mean? So, X takes the value n with probability n plus k minus 1, see k minus 1, p to the power k q to the power n. Note that k and p are parameters of the distribution. So, suppose we have a coin with probability of getting head is p.

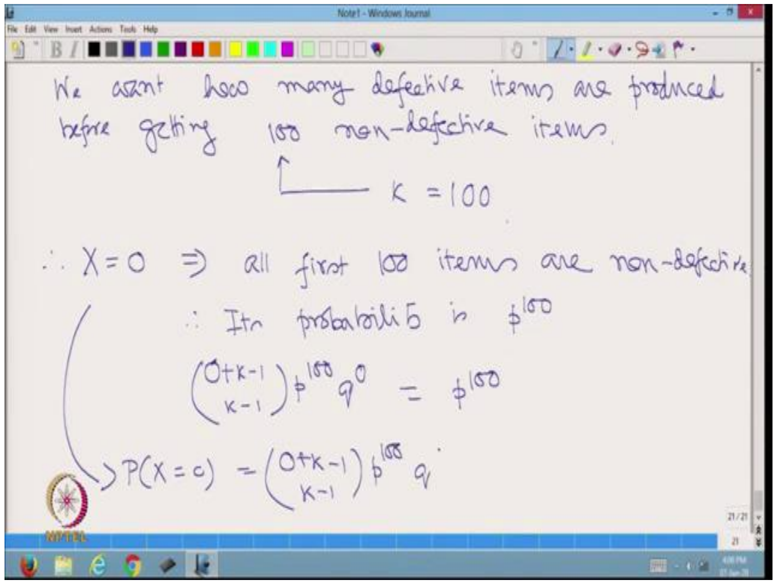
(Refer Slide Time: 34:16)



And suppose X denotes the number of failures or number of tails that we get before having k many heads. So, we have a coin, we keep on tossing and we keep noting the outcomes, we keep doing that until we get k heads and X denotes the number of tails that we have got in that process.

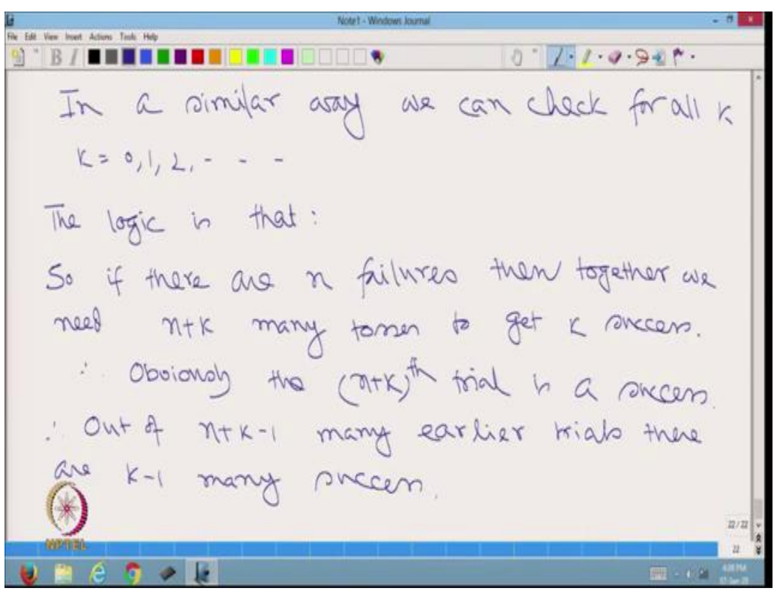
Therefore, X interprets as number of failures before getting kth success so k is parameter of the distribution, so that has been fixed. Say, for example, we need 100 items for a project. The machine that produces the item has probability of giving a non-defective one is p and with probability q it produces a defective one.

(Refer Slide Time: 37:13)



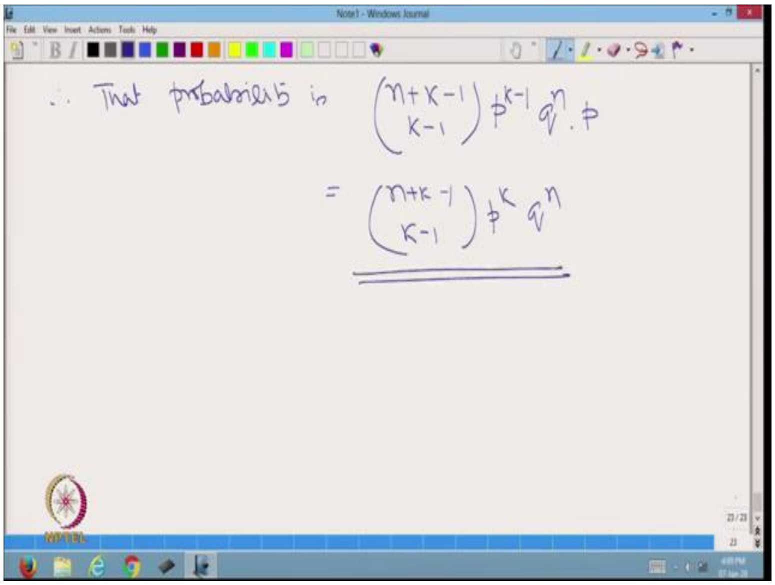
We want to know how many defective items are produced before getting 100 non-defective items. So, value of K is equal to 100. Therefore, X is equal to 0 implies all first 100 items are non-defective and its probability is p to the power 100. So, by our definition, 0 plus K minus 1, see K minus 1, p to the power 100 q to the power 0 is equal to p to the power 100. Therefore, probability X is equal to 0 is equal to 0 plus K minus 1 see K minus 1 p to the power 100, q to the power 0.

(Refer Slide Time: 40:58)



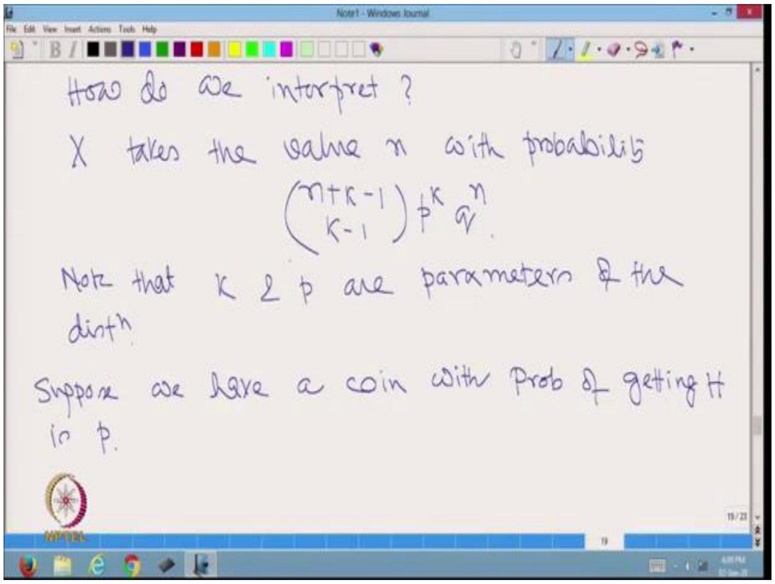
In a similar way we can check for all K. K is equal to 0, 1, 2 up to infinity. The logic is that so if there are n failures then together we need n plus K many tosses to get K success which you can say as H. Therefore, obviously the n plus Kth trial is a success. Therefore, out of n plus K minus 1 many earlier trials there are K minus 1 many success.

(Refer Slide Time: 40:59)



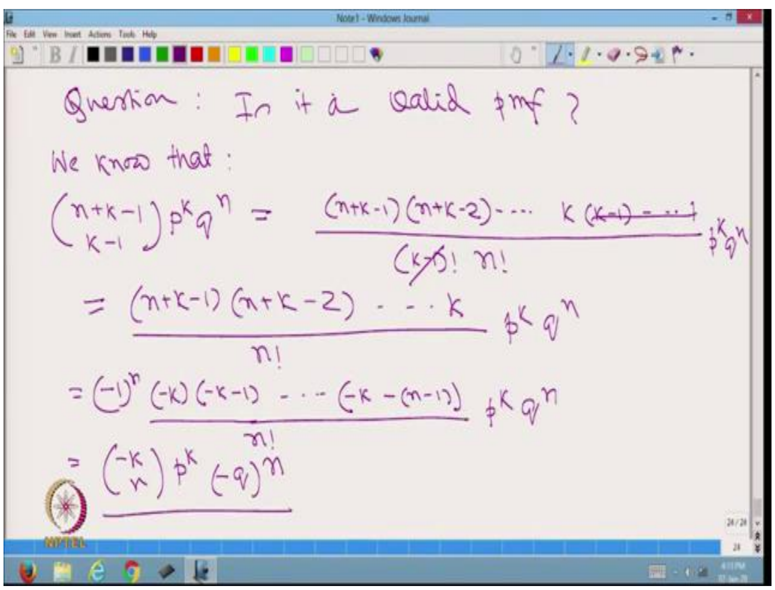
Therefore, that probability is out of the total n plus K minus 1, you are choosing K minus 1, all of them have been a success that has the probability p to the power K minus 1. The remaining ones are failures that has probability q to the power n finally, the last one is a success. Therefore, this is coming out to be n plus K minus 1 see K minus 1 p to the power K q to the power n.

(Refer Slide Time: 41:53)



So, that is how we get these probabilities as we have indicated here.

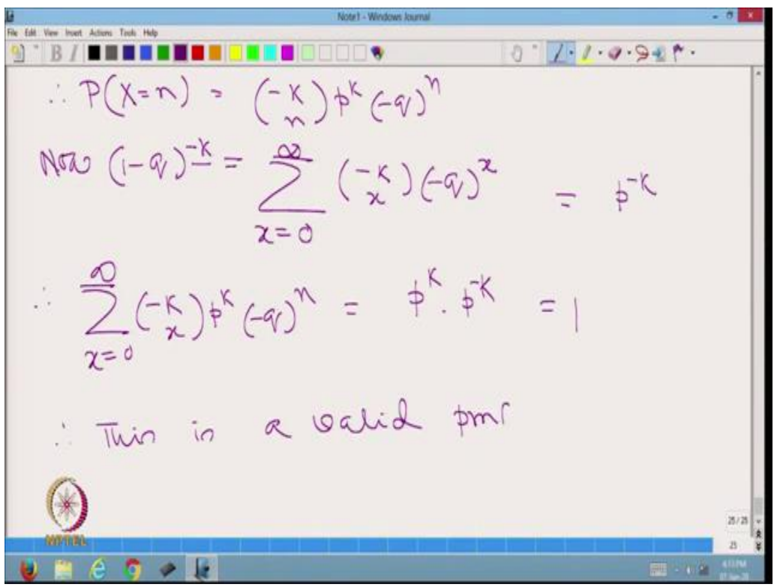
(Refer Slide Time: 41:58)



Question is, is it a valid pmf? We know that n plus k minus 1 see k minus 1 p to the power k q to the power n is equal to n plus k minus 1 into n plus k minus 2 up to k, k minus 1 up to 1 divided by k minus 1 factorial, n factorial into p to the power k q to the power n, which is equal to if we cancel this part we get n plus k minus 1 into n plus k minus 2 up to k divided by n factorial p to the power k q to the power n.

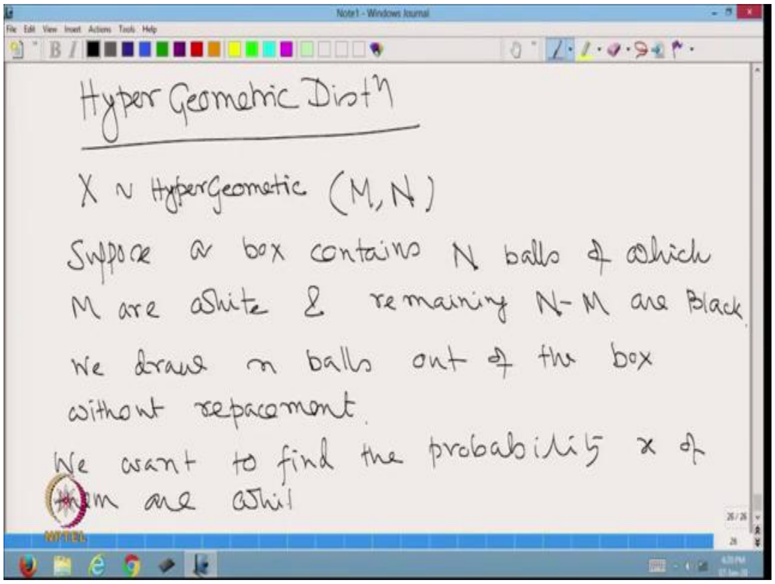
Is equal to minus 1 to the power n, note that there are n many terms here, k plus 0, k plus 1 up to k plus n minus 1. So, we take minus 1 to the power n common and then we get minus k, minus k minus 1 up to minus k minus n minus 1 divided by n factorial into p to the power k q to the power n and this we can write it as minus k see n p to the power k minus q to the power n. You have to understand binomial expansion with negative coefficients, which I am sure you have done in your class 11 or 12 algebra. So, this is the expansion that we are getting for the kth term.

(Refer Slide Time: 44:41)



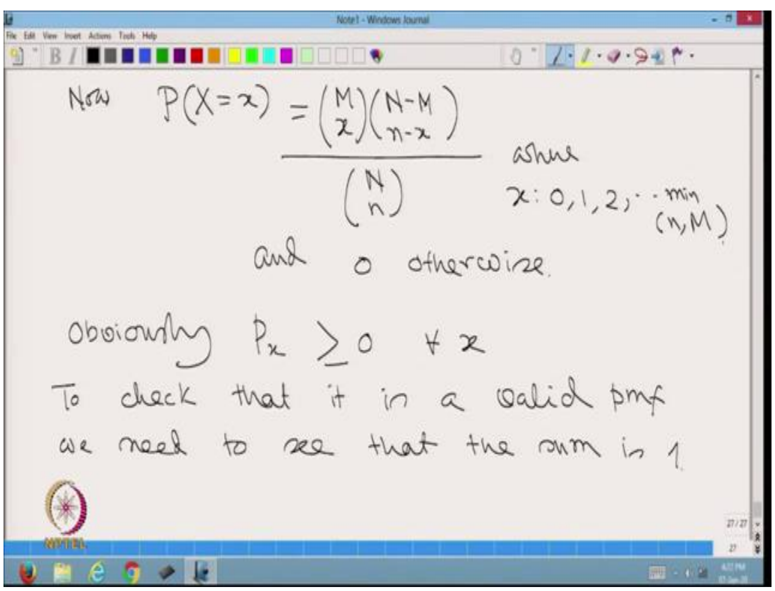
Therefore, probability X is equal to n, we can write it as minus k see n p to the power k minus q to the power n, now 1 minus q whole to the power minus k is equal to sigma x is equal to 0 to infinity, note that it is a negative coefficient. Therefore, there will be infinitely many terms minus k see x minus q whole to the power x, which is equal to p to the power minus k. Therefore, sigma x is equal to 0 to infinity minus k see x p to the power k minus q to the power x is equal to p to the power k into p to the power minus k is equal to 1 therefore, this is a valid pmf.

(Refer Slide Time: 46:10)



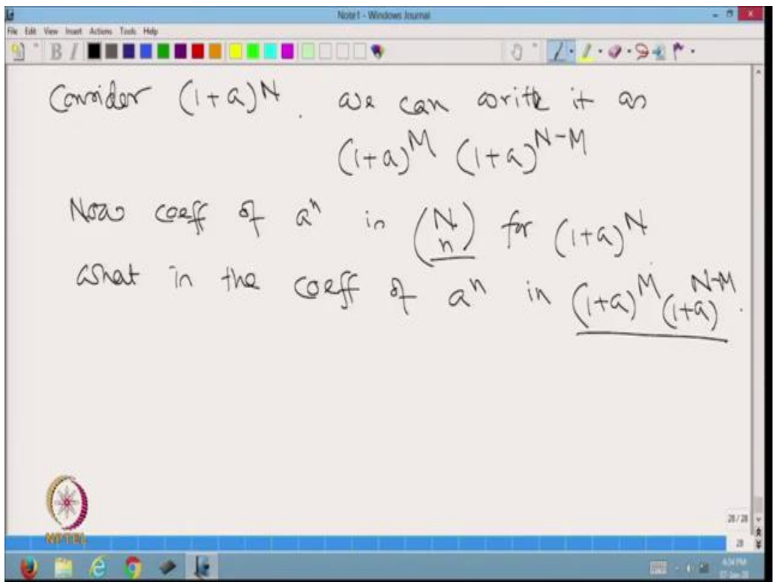
Before we stop, I give you another discrete random distribution, Hyper Geometric Distribution. Here, it has two parameters m and n. So, suppose a box contains n balls of which m are white and remaining n minus m are black. We draw n balls out of the box without replacement, we want to find the probability x of them are white.

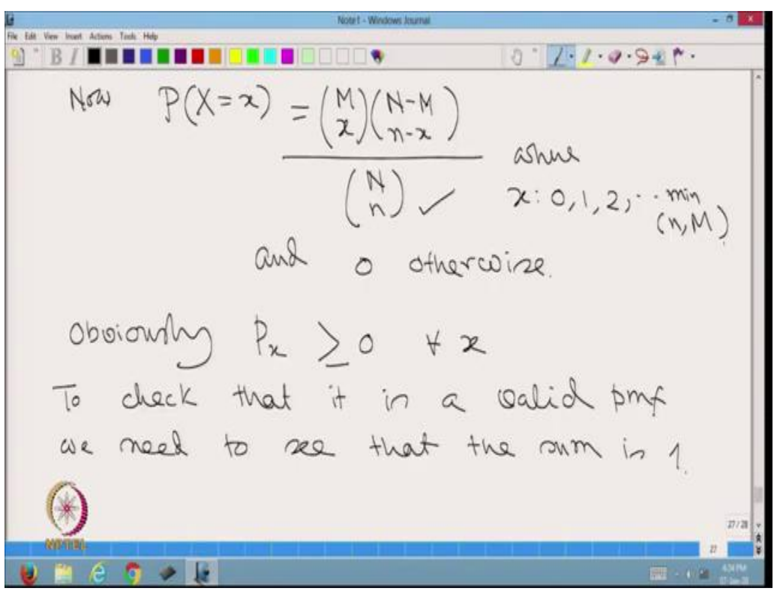
(Refer Slide Time: 48:07)



Now, probability X is equal to x is there are m white balls out of them x can be chosen in M see x ways out of the remaining N minus M black balls. Therefore, I choose n minus x are blacks, and total number of selection is N see n where x can take the value 0, 1, 2 up to a minimum of n and M and 0 otherwise. Obviously, these terms Px is greater than equal to 0 for all x to show it is a valid pmf we need to see that the sum is 1.

(Refer Slide Time: 49:42)





I give you a hint, considered 1 plus a whole to the power n, we can write it as 1 plus a whole to the power m into 1 plus a whole to the power N minus M. Now, coefficient of a to the power n is N see n for1 plus a whole to the power N. What is the coefficient of a to the power n in 1 plus a whole to the power M into 1 plus a whole to the power N minus M? I want you to calculate that and see that this actually is giving the sum of the probabilities that we get from the Hyper Geometric distribution.

Therefore, that divided by N see n is going to give us 1 and N see n is the denominator here. So, effectively, I want you to check from here that the coefficient of a to the power n is actually the sum of these terms for x is equal to 0 to minimum over n M. Therefore, that will give us a pmf.

Okay friends, I stopped here today. So, in this last two lectures, we have covered different discrete distributions, namely, Bernoulli, Binomial, Poisson, Geometric, Negative binomial and Hyper Geometric. In the next class, I shall introduce you to the concept of Continuous Random Variables. Thank you.