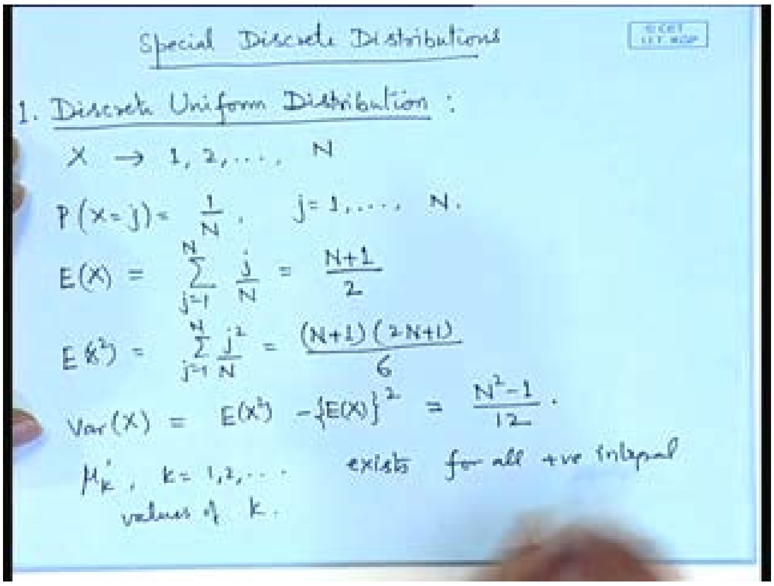
Probability and Statistics

Prof. Dr. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Module No. #01
Lecture No. #10
Special Distributions-I

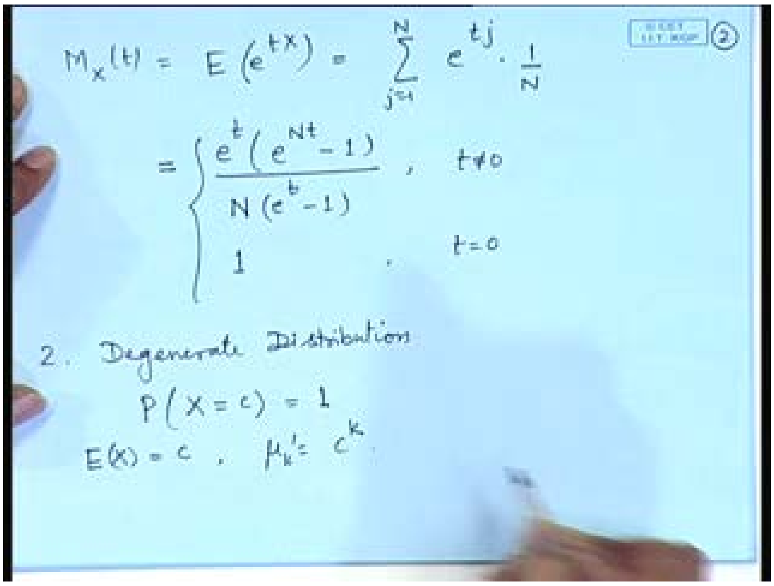
(Refer Slide Time: 00:33)



Today, we will introduce various special distributions which are encountered in various physical and other kinds of experiments. So, broadly we will categorize them into two parts- one is special discrete distributions and another is special continuous distributions. So, the first distribution that I will be taking up is the discrete uniform distribution. So, as the name suggest it allots uniform weights to various points. So, a random variable X, it takes value say 1 to N and probability that X is equal to say, j is equal to 1 by N for j is equal to 1 to N that means, this probability distribution, it gives equal weight 1 by N to N distinct integer valued points. This type of distribution arises in the classical theory for example, if we are considering a coin tossing and we are looking at head as occurrence of 1 and occurrence of tail as 0, then probability of X equal to 0 and probability of X equal to 1 both is equal to half. Suppose we are considering tossing of a die, so, there are 1, 2, 3, 4, 5, 6, as the possibilities each with probability 1 by 6. So, in all such cases discrete uniform distribution is applicable.

Let us look at some of the features such as mean, variance, etcetera. So, let us look at mean. Mean of this distribution is sigma j by N, j is equal to 1 to n, so, that becomes sigma of j is N into N plus 1 by 2 therefore, you get the mean as N plus 1 by 2, which is appropriate because it is something like a middle point of the distribution. If you want to calculate the variance, we can use expectation of X square, that is equal to sigma j square by N, j is equal to 1 to N, so sum of j square is N into N plus 1 into 2N plus 1 by 6, so, expectation of X square turns out to be N plus 1 into 2N plus 1 by 6. And therefore, variance of X is equal to expectation of X square minus expectation of X whole square that is, this quantity minus this square, so, after simplification it turns out to be N square minus 1 by 12. We may also calculate its third moment, fourth moment etcetera. Since every time it is a finite sum the moments of all positive integral orders will exist that means, we can say mu k prime for k equal to 1, 2 and so on exists for all positive integral values of k.

(Refer Slide Time: 03:34)

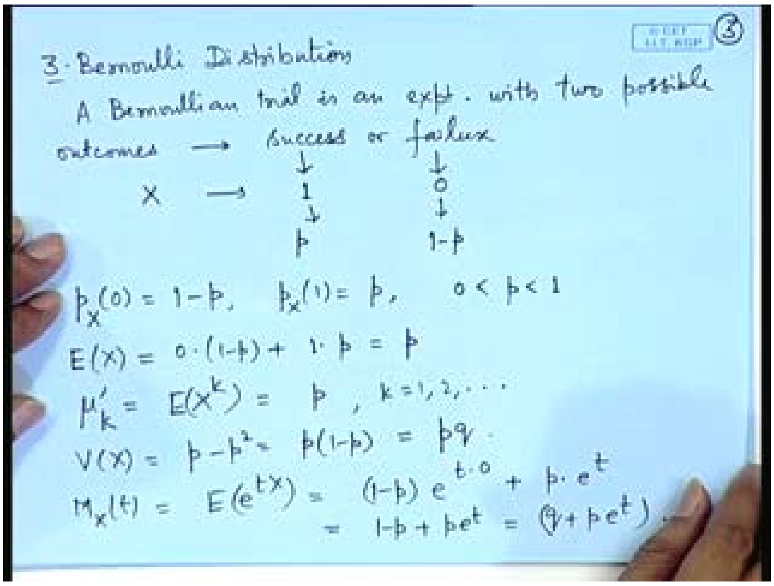


We may also look at the moment generating function of this distribution M X t, that is equal to expectation of e to the power tx, so, it is equal to sigma e to the power tj and each is with probability 1 by N, j is equal to 1 to N. Now, if we look at this, this is e to the power t plus e to the power 2t plus e to the power Nt, which is a finite geometric series; so, this sum you can evaluate and it is turning out to be e to the power t e to the power Nt minus 1 divided by e to the power t minus 1, and we have this N in the denominator- now, this expression is valid for all t. Of course, when we say t is equal to 0 then this is not defined because the denominator becomes zero and the numerator also becomes zero, so, in that case we write separately; so, this is for t not equal to 0 and for t equal to 0, we specifically write it as 1. So, the higher order moments of this distribution can also be derived from the expression for the moment generating function because you can consider a Maclaurin series expansion around t is equal to 0, or we can consider derivatives of m X t of various order and put t equal to 0 to get the moment of that particular order.

Another trivial kind of distribution is degenerate distribution. The degenerate distribution arises when we are sure about a particular event to occur. So, probability X is equal to say c is equal to 1 that means, the random variable takes only one value with probability 1. Therefore, expectation will be c, and moment of any particular order can also be calculated for example, mu k prime will be equal to c to the power k.

A third type of discrete distribution arises in experiments which are called bernoullian trials. So, several times in the real life we are interested in phenomena from a particular point of view such as, we look at only whether a particular event has occurred or it has not occurred. For example- if we appear in a competitive examination, so, whether we qualify or we do not qualify; if a medicine is taken to cure a disease then the outcome recorded may be that whether the disease is cured or it is not cured. So, generally we call it as success failure trials. We make the assumption that the trials are conducted independently under identical conditions so that the probability of success is considered to be fixed.

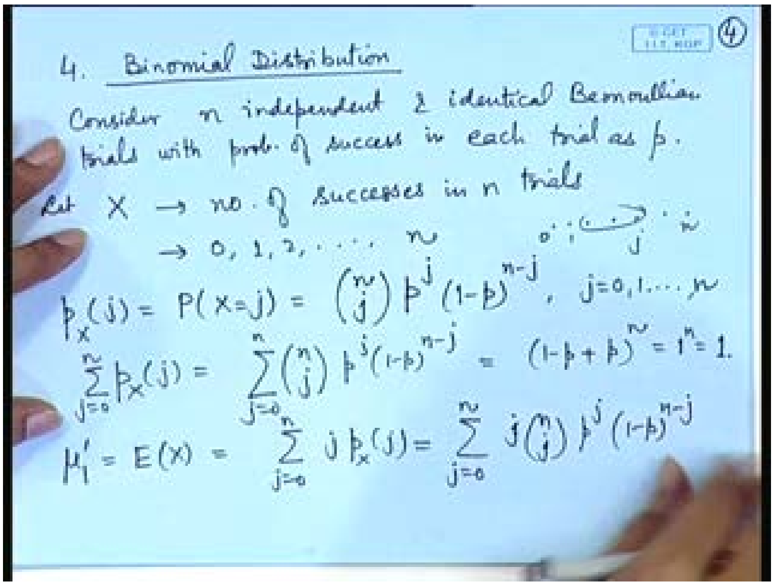
(Refer Slide Time: 06:43)



So, if we consider the outcome of one trial, then it is known as Bernoulli distribution. So, a Bernoulli trial is an experiment with two possible outcomes that is, success or failure. The random variable X will associate the value 1 with success and the value 0 with failure, and the probability with success is say p and the probability with the failure is say 1 minus p. So, the distribution is described by px equal to 0 is 1 minus p and px equal to 1 is p. So, this is also a 2 point distribution. So, if we look at our discrete uniform distribution, this was an N point distribution, for N is equal to 2 it is like a Bernoulli distribution however, here the probability of both success and failure would have been same whereas, in the Bernoulli distribution it can be different number, say p and 1 minus p, where in general p is a number between 0 and 1- if you take the extreme case that say, p is equal to 1 or p is equal to 0, then this is reducing to a degenerate distribution.

Let us consider some of the properties of this distribution say, its mean. So, mean is 0 into 1 minus p plus 1 into p, that is equal to p. So, if in a single trial the probability of success is p, then on the average, the average value of the distribution should be equal to the success probability that is, p. Now, since here the values taken are only 0 and 1 and any powers of 0 and 1 are also same that means, in general if I calculate the moment of kth order, expectation of X to the power k that will be equal to again p for k equal to 1,2 and so on. Therefore, if we look at say variance of this distribution that is equal to p minus p square that is equal to p into 1 minus p, 1minus p many times we write as q also, so it is pq. The moment generating function is equal to expectation of e to the power tx, that is equal to 1 minus p into e to the power t into 0 plus p into e to the power t, that is equal to 1minus p plus pe to the power t, or q plus pe to the power t. It is obvious that moments of all orders exist here and they can be evaluated using the relationship between central and non central moments.

(Refer Slide Time: 09:58)

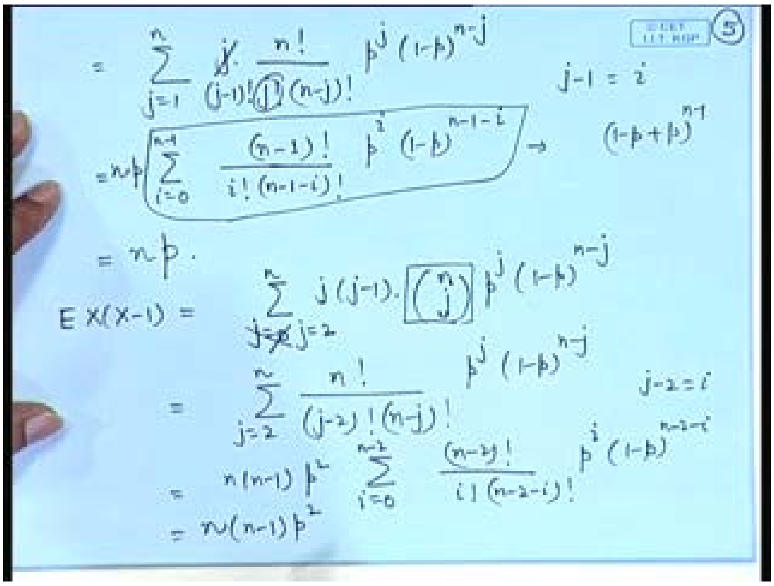


Now, an immediate generalization of the Bernoulli distribution is the so called Binomial distribution. So, consider n independent and identical Bernoullian trials with probability of success in each trial as say, p. Let us denote the X as the number of successes in n trials. Then what are the possible values of X? X can take value 0, 1, 2 and n, etcetera. What is the probability that X is equal to say, j? That is, probability of X equal to... So, if you have j successes, then in each trial the probability of success is p, so, p into p j times. Now, it is obvious that if we are looking at total number of trails as n trials and we are fixing some of these j trials as success, then the remaining n minus j must be failures, the probability of a failure is 1 minus p therefore, the probability of n minus j failures will be 1 minus p to the power n minus j. Now, out these n trials these j trials can be selected in ncj possible ways. Therefore, the probability mass function of the binomial distribution is ncj p to the power j into 1 minus p to the power n minus j. Here we have made use of the independence of the trials because the probabilities of individual trials whether success or failure have been multiplied. Since the binomial coefficients ncj occurs here that is why it is known as binomial distribution.

This distribution has wide applicability, as I mentioned we may be looking at the number of successes in solving a certain multiple choice question paper; suppose, we are looking at number for successful hits in a basketball game by a particular team, we may be interested in looking at the number of patients treated successfully following a particular line of treatment and so on. So, whenever we are interested in dividing particular phenomena only as a success or failure then this particular distribution is quite useful.
Let us look at various properties. Now, whenever we write down a probability mass function we should ensure that the sum of the probability mass function over the required range must be equal to 1. So, if we consider the sum of the binomial probabilities, this is equal to 1 minus p plus p to the power n which is equal to 1 to the power n, that is equal to 1- so, this is basically the binomial expansion of 1 minus p plus p to the power n.

Now, this fact can be used for evaluation of the moments. For example, if we are calculating mu1prime that is, expectation of X, this is equal to sigma j px j is equal to 0 to n, that is equal to sigma j ncj p to the power j 1 minus p to the power n minus j. Now, if we look at this sum and if we compare it with the sum of the distribution taken in the previous statement, it is required here that I should make this particular term as a binomial coefficient term; here it is j into ncj, so somehow it has to be adjusted, so for this we make certain observation- first thing is that corresponding to j is equal to 0 this term vanishes.

(Refer Slide Time: 14:43)

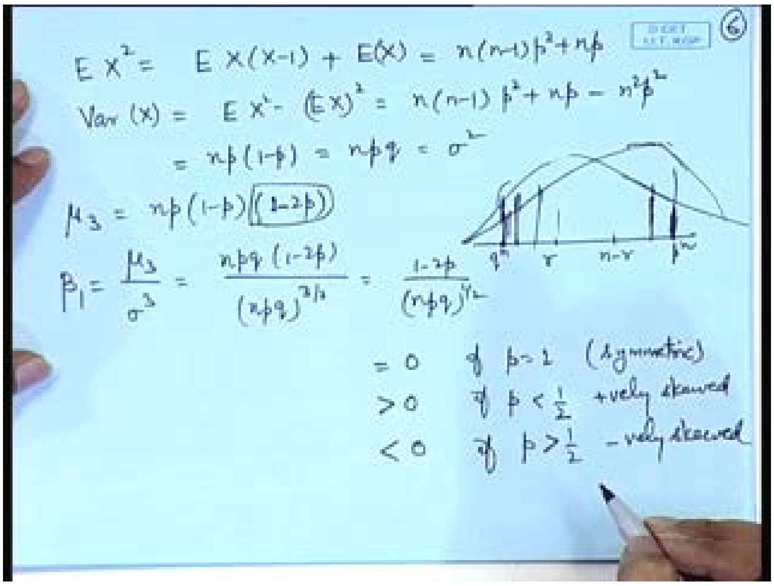


So, in effect this sum is actually considered from j is equal to 1 to n, j, and this ncj we can write as n factorial divided by j factorial n minus j factorial p to the power j into 1 minus p to the power n minus j obviously, this j and j factorial we can adjust the term and we can write this as j minus 1 factorial. Then, this suggests that we can replace j minus 1 as equal to say, i , and then this becomes sigma i is equal to 0 to n minus 1. Now, if we are putting this term as i , then this we have to adjust, we can write it as n minus 1 minus i factorial and therefore, in the numerator in place of N factorial we can consider n minus 1 factorial and n can be kept out, so, this is p to the power i and 1p will be here 1 minus p to the power n minus 1 minus i. So, if you look at this term, it is the expansion of 1 minus p plus p to the power n minus 1, which is 1 and therefore, this summation, reduces to np. So, the mean of the binomial distribution is np, which is understandable, because if we say that the probability of success in one trial is p, then out of n trials, what is the expected number of successes? It must be n into the probability of success in each trial that is, np.

Now, this particular way of deriving the moment of a binomial distribution suggests that if we want to look at say, mu2prime, then here I will get j square, now, in this particular expansion 1j was cancelled, so, if I have j square another j cannot be cancelled therefore, it suggests that it will may be beneficial to consider factorial moments.

So, we may consider say expectation of X into X minus 1, that will be equal to j into j minus 1 ncj p to the power j 1 minus p to the power n minus j, j is equal to 0 to n. Now, like in the previous case we can observe that this term is vanishing for j is equal to 0 and j is equal to 1, so, this we can cancel and we can consider it as j is equal to 2 to n. Once we write that then in the expansion of ncj the j factorial term in the denominator can be adjusted with j into j minus 1and we get here n factorial divided by j minus 2 factorial n minus j factorial p to the power j into 1 minus p to the power n minus j, j is equal to 2 to n; so, obviously, we can substitute j minus 2 is equal to i and this gives us n into n minus 1 sigma i is equal to 0 to n minus 2 n minus 2 factorial divided by i factorial n minus 2 minus i factorial p to the power i 1 minus p to the power n minus 2 minus I, so, p square term has come out. So, the second factorial moment is n into n minus 1 p square.

(Refer Slide Time: 18:19)



So, we may utilize this to calculate expectation of X square- because expectation of X square is expectation of X into X minus 1 plus expectation of X- and we substitute these terms, so it is n into n minus 1 p square plus np. Now, that helps us to calculate variance of X, that is equal to expectation of X square minus expectation of X whole square, that is equal to n into n minus 1p square plus np minus n square p square. So, here you can observe that the term n square p square cancels out, we are left with np minus np square which is np into 1 minus p, or npq. Obviously, if p is a number between 0 and 1, then npq is going to be less than np. So, in the binomial distribution in particular we have the average less than the variance value. We may also look at the higher order moments to discuss the properties of Skewness and kurtosis etcetera.

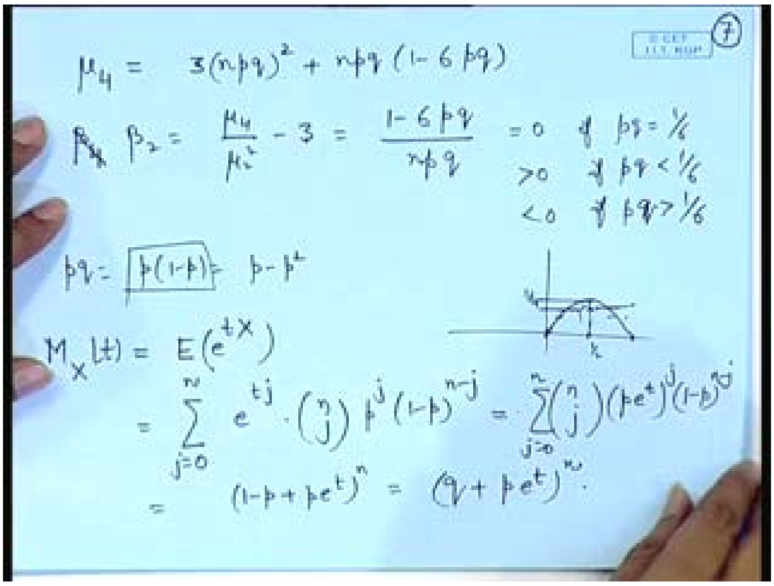
So, let us consider third moment. Now, the coefficient of Skewness beta1 is mu3 divided by sigma cube, so this is sigma square here. Now, if we are considering mu3, this is the third central moment, now, third central moment can be expressed in terms of the first three non central moments, now, that means, we need mu3prime. Now, mu3prime, expectation of X cube will require the third factorial moment. So, we can use it because we can follow the similar type of calculations, the third factorial moment that is, expectation of X into X minus 1 into X minus 2, if we follow the same logic, it will come out to be n into n minus 1 into n minus 2p cube.

And also if we look at say, measure of kurtosis, it will require mu4. Now, mu4 will require mu4prime, etcetera, and mu 4 prime will require the fourth factorial moment and that will be equal to n into n minus 1 into n minus 2 into n minus 3p to the power 4. So, after doing certain algebraic simplifications we can obtain mu3 as np into 1 minus p into 1 minus 2p. Now, you can easily see here that the term np into 1 minus p, because it is the variance term it is always non negative whereas, if you look at this term this will determine the symmetry of this distribution. Obviously, if we look at the value p is equal to half, then this is 0, then this is a symmetric distribution, which is alright because if we consider the binomial probabilities, it is starting from 1 minus p to the power n and goes up to p to the power n the second 1 is nc1 p into q to the power n minus 1, the last, but one is nc n minus 1 which is again same as nc1 p to the power n minus 1 into q. So, if p and q are same, then the rth term will be same as n minus rth term and therefore, it will be a symmetric distribution.

So, if we write down beta 1 that is, mu3 divided by sigma cube, that is equal to npq into 1 minus 2 p divided by npq to the power 3 by 2, that is equal to 1 minus 2p divided by npq to the power half. Then, this is equal to 0 if P is equal to half that means symmetric distribution, so binomial distribution will be symmetric. So, which is obvious also because in a binomial distribution we are looking at the probabilities of success and failures, and if in each individual trial the probability of success and failure is the same for example, if you are tossing a fair coin, then the distribution must be symmetric; this is greater than 0 if p is less than half.

Now, naturally this value is q to the power n, this value is p to the power n etcetera. So, if p is less than half, then this values will be higher corresponding to the values on this side, this values will be smaller for example, p to the power n will become less than q to the power n, so it will be a positively skewed distribution. On the other hand, if I have p greater than half, then this will become less than 0. So, if p is greater than half, then these values will become higher and these values will become lower. So, the shape of the distribution will be something like this, so it will become negatively skewed. So, this third moment, third central moment clearly gives the information about the Skewness of the distribution.

(Refer Slide Time: 23:34)



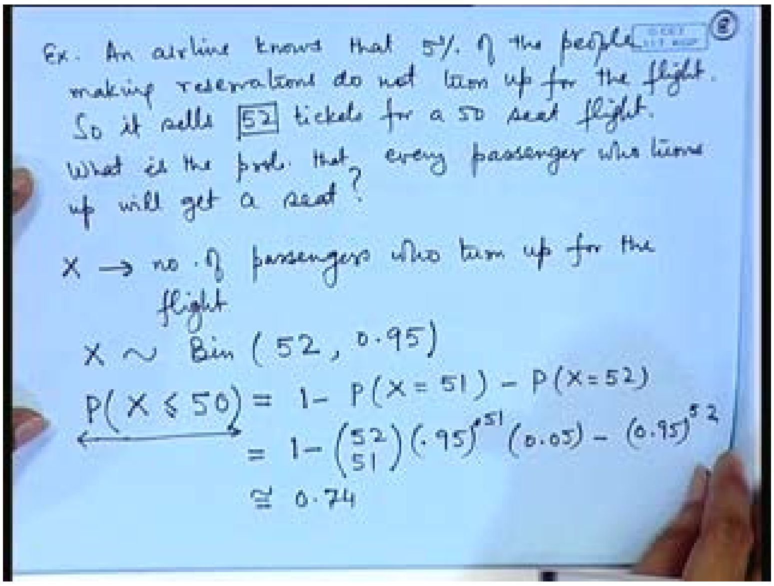
Let us also look at the fourth central moment. And as I have explained the method of calculation, you will need to calculate mu4prime, which will require the fourth factorial moment, which can be easily evaluated, after using that mu4 turns out to be 3 npq square plus npq into 1 minus 6 pq. So, if we look at the coefficient beta 4, sorry, beta 2, it is equal to mu4 by mu2 squares minus 3. So, mu 4 is this term and mu 2 is npq, so mu 2 square becomes npq whole square. So, if we consider mu4 by mu2 square this will give me 3 here and 3 minus 3 will cancel out. So, we are left with 1 minus 6 pq divided by npq. So, obviously, this is equal to 0 if pq is equal to 1 by 6, it is greater than 0 if pq is less than 1 by 6, it is less than 0 if pq is greater than 1 by 6.

Now, this p into q term is actually p into 1 minus p that is, p minus p square. Now, we know that the range of this, it is from 0 to 1 by 4, the maximum value is attained at p is equal to half. So, it is a basically a concave function p into 1 minus p, it is like this, at p is equal to 0 and p is equal to 1, it is 0, and at p is equal to half, the value is equal to 1 by 4. So, naturally it is possible that the value of p into 1 minus p can be greater than 1 by 6, equal to 1 by 6, or less than 1 by 6. So, if p into q is less than 1 by 6 that means, the values are here then, the peak of the binomial distribution is slightly higher than the normal, if pq is greater than 1 by 6, it will be slightly less than the peak of the normal distribution.

Another thing which you can observe from these coefficients that beta1 is equal to certain term and in the denominator we have square root n. So, even though it may be positively or negatively skewed, but if n becomes large the skew becomes closer to zero that means, it will become closer to a symmetric distribution.

In a similar way, if we look at the coefficient beta2 here in the denominator, we have n and therefore, as n becomes large the measure of kurtosis is closer to 0 even though pq may not be equal to 1 by 6 and therefore, it will become closer to a normal peak. We may also look at the moment generating function of the binomial distribution-expectation of e to the power tX, that is equal to e to the power tj ncj p to the power j into 1 minus p to the power n minus j, j is equal to 0 to n. Now, if we are making use of the binomial expansion, then it is clear that the term e to the power tj must be adjusted with the term p to the power j. So, this becomes ncj pe to the power t whole to the power j into 1 minus p to the power n minus j, j is equal to 0 to n. And this becomes 1 minus p plus pe to the power t whole to the power n, that is q plus pe to the power t whole to the power n. So, for all values of t this is well defined.

(Refer Slide Time: 28:01)



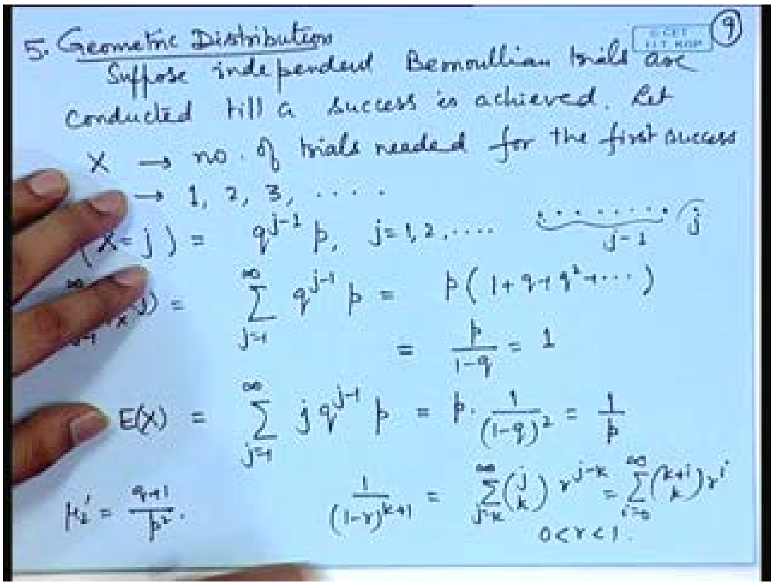
Let us take one application of the binomial distribution. An airline knows that 5 percent of the people making reservations do not turn up for the flight, so it sells fifty two tickets for a fifty seat flight, what is the probability that every passenger who turns up will get a seat? So, let us look at this. So, if a passenger makes a booking, he may not turn up for the flight with probability 0.05 and may turn up for the flight with probability 0.95. So, if I say X is the number of passengers who turn up for the flight then, this can be considered as a collection of Bernoullian trials that means, a passenger purchasing a ticket, he may turn up or he may not turn up, so, turning up may be considered as success and may not turn up as a failure, or vice versa. So, if we look at X as the number of passengers who turn up for the flight, then the distribution of X will be binomial 52- because there are total fifty two tickets that have been sold- 0.95.

So, a passenger will get a seat provided the number of people who turn up is less than or equal to 50, so, we may consider it as 1 minus probability X is equal to 51 and probability X equal to 52. So, using the form of the binomial distribution these can be easily calculated, this is 52c51 0.95 to the power 0.51 0.5 minus - so, this will become simply- 0.95 to the power 52. So, these values can be easily evaluated and this probability turns out to be approximately 0.74.

Now, it may sound to be somewhat reasonable that if 5 percent of the people do not turn up for the flight, then the airlines sells fifty two tickets for a fifty seat flight and it may consider that we are quite safe, that there will be hardly any complaints, but you see the probability that less than or equal to fifty passengers turn up for the flight is 0.75, 0.74, you can say roughly three-fourth of the passengers or 75 percent of the passengers. So, what does it mean? That means, 25 percent of the times there will be a case where a passenger turns up for the flight and he will not get a seat, so, which is not a very good advertisement for the flight he has that means, they must not sell more tickets basically-may be they will sell only 51 ticket, that may be alright.

Now, if we consider these Bernoullian trials in a slightly different way, here what we did is that we looked at the total number of trials, how many successes are there, but we may also have a situation where we are looking at a specified number of successes or specified number of failures. For example, you consider the trials for a particular drug for a certain disease; so, generally in all medical trials the trials are done repeatedly over many subjects and if a specified number of successes are obtained, then only the medicine are approved to be applied to the general population; so, we may find out that a particular medicine is given and there is a target group, and if say fifteen people are successfully treated, then we may say that the medicine is alright; suppose hundred people are successfully treated, then we say that the medicine is successful. So, here the trials are conducted till we get a specified number of successes.

(Refer Slide Time: 33:26)

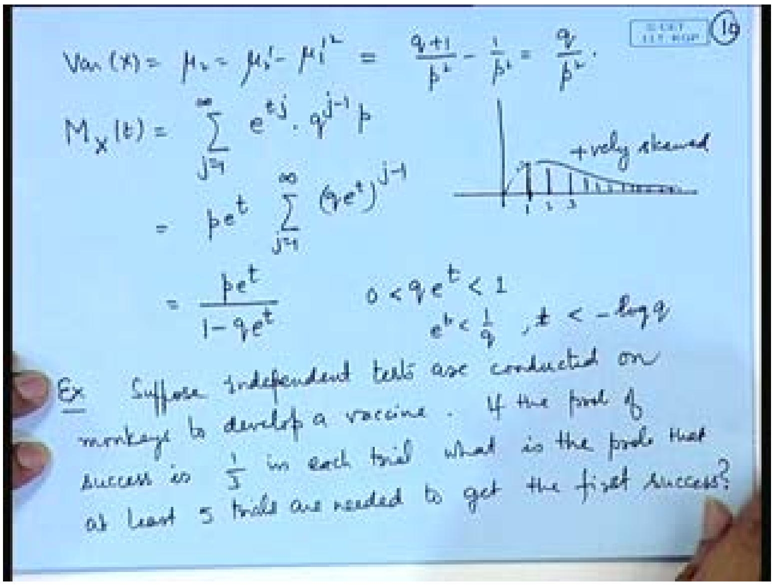


So, this way of looking at it, suppose independent Bernoullian trials are conducted till a success is achieved. Let X denote the number of trials needed for the first success, so what are the possible values of the X? You may get a success in the first trial, you may get the success in the second trial and so on. So, unlike the binomial distribution this is an infinite, but countable many valued random variables. You will have probability X is equal to j is equal to- now, we are saying that in the jth trial the success is observed that means, before that in the j minus 1 trials all of them are failures and all the trials are fixed, because first is a failure, second is a failure, up to j minus 1th it is a failure, jth one is a success- so, the probability of this becomes q to the power j minus 1 into p. So, if we look at say sigma px j, j is equal to 1 to infinity, that is equal to sigma q to the power j minus 1 into p, j is equal to 1 to infinity, that is equal to p and then you have 1 plus q plus q square, an infinite geometric series, this is equal to p divided by 1 minus q, that is equal to 1. Since here the terms are involved from a geometric series, so, that is why this is known as a geometric distribution.

So, we may look at its various characteristics such as mean say, mu1prime, expectation X, that is equal to sigma j q to the power j minus 1 p. Now, from the proof that this was a proper probability mass function we expressed the sum as an infinite geometric series that means, it is an expansion of a term of the type 1 minus q to the power minus 1. Therefore, if you want to calculate higher order moments, we have to make use of the similar term. So, this one we can write as expansion of 1 by 1 minus q square, so, it becomes 1 by p. That means, the average number of trials needed for the first success is the inverse of the probability of success in a single trial.

So, suppose we consider tossing of a coin where the probability of head may be 1 by 3 then, we require on the average three throws of the coin to get a head. Suppose we are having a die where the probability of a 6 is say 1 by 6, so, you will need on the average six trials to get the first 6. Now, if you want to calculate say mu2prime and so on, then we will need to consider the expansions of 1 minus q to the power minus 3 and so on. So, in general we will be making use of the formula 1 by 1 minus r to the power k plus 1 is equal to sigma jck r to the power j minus k, j is equal to k to infinity, or we can also write it as k plus ick, i is equal to 0 to infinity, r to the power i , where r is a- this is j minus k- here r is a number between- in fact, this is the mathematical formula- and we need to have r between minus 1 to 1, but since here q is a probability, so it has to be between 0 and 1.

(Refer Slide Time: 38:27)

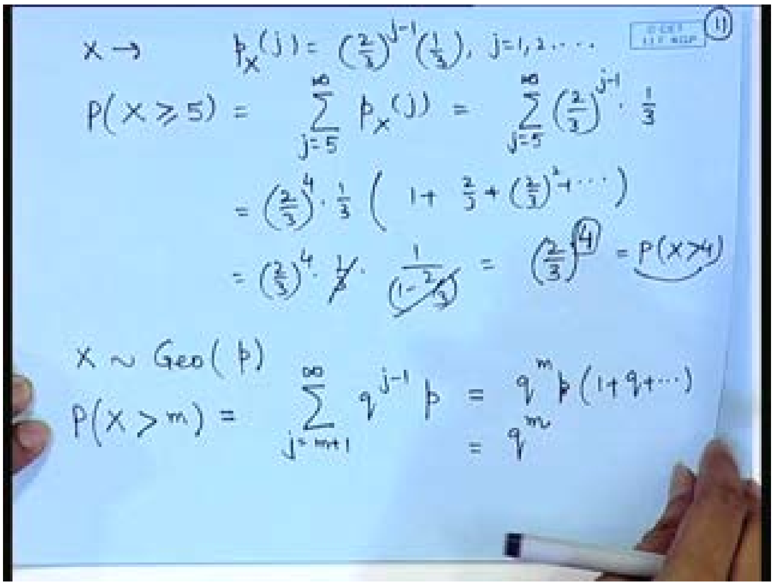


So, using this we may write other terms that is, mu2prime turns out to be q plus 1 by p square and therefore, we may calculate the variance of X that is, mu2 that is, mu2prime minus mu1prime square and it will be equal to q plus 1 by p square minus 1 by p square, that is equal to q by p square. And we may write down further higher order moments say, mu3, mu4, etcetera, to look at the shape of the distribution. However, that can be looked at from the direct description of the probabilities also. Here the probability that X is equal to 1 is p, probability that X equal to 2 is say, pq, now, since q will be between 0 and 1 therefore, this will be less than this, the probability that X equal to 3 will be q square p. So, naturally you can see that there is a decreasing trend here, so it is a positively skewed distribution.

We may also look at the moment generating function of this distribution. Expectation e to the power tj, so, it is sigma e to the power tj, q to the power j minus 1 p, j is equal to 1 to infinity. So, obviously, we can see here that this can be expressed as pe to the power t, qe to the power t to the power j minus 1, j is equal to 1 to infinity, which is nothing but the infinite geometric sum as an expansion of 1 by 1 minus qe to the power t; now, this expansion will be valid provided qe to the power t is less than 1 of course, this is positive that means, e to the power t is less than 1 by q or t is less than minus log of q, since q is a number between 0 and 1, log of q is negative and therefore, minus log q is positive. So, in a neighborhood of t is equal to 0, this moment generating function exists. So, this can be utilized to calculate all the moments of this distribution.

Let me give one example here. Suppose independent tests are conducted on monkeys to develop a vaccine, if the probability of success is 1 by 3 in each trial, what is the probability that at least five trials are needed to get the first success?

(Refer Slide Time: 42:08)

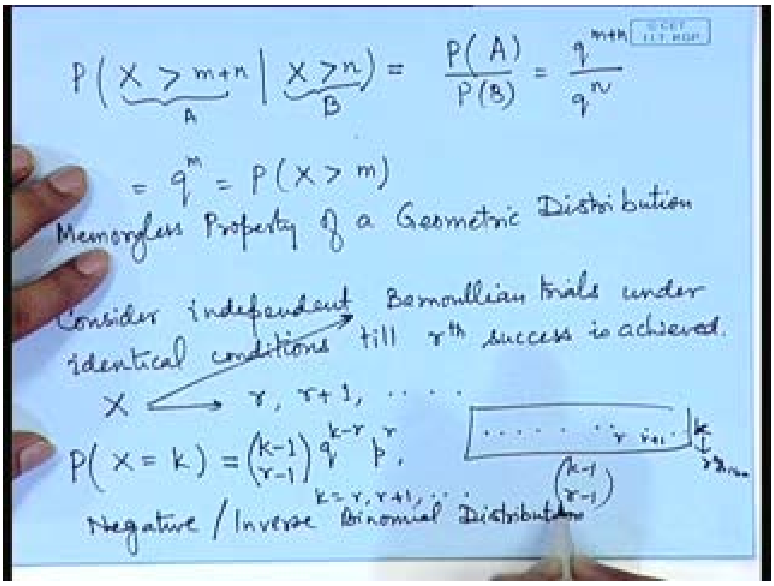


So, this is perfectly a case of geometric distribution. If I am considering X as the number of trials needed, then you will have px j is equal to 2 by 3 to the power j minus 1 into 1 by 3 for j is equal to 1, 2 and so on that means, in the first trial you will get a success with probability 1 by 3, in the second trial you will get a success with probability 2 by 3 into 1 by 3 and so on.

So, what is required here is that probability X greater than or equal to 5 that means, sigma px j, j is equal to 5 to infinity, which is equal to sigma 2 by 3 to the power j minus 1 1 by 3, j is equal to 5 to infinity. So, this is equal to- so, if we look at the first term of this series, this is 2 by 3 to the power 4 into 1 by three, so we can keep it out and thereafter it will become 1 plus 2 by 3 plus 2 by 3 square and so on- so, this becomes 2 by 3 to the power 4 1 by 3 1 divided by 1 minus 2 by 3, so this cancels out and the probability is 2 by 3 to the power 4. Since this is a discrete distribution we can also consider it as probability X greater than 4. Now, notice here probability X greater than 4 is 2 by 3 to the power 4.

So, from here we can write down somewhat more general expressions if we are considering. So, X is a geometric distribution with probability p of success in each trial, if we look at what is the probability say X greater than m, then this is equal to sigma q to the power j minus 1 p, j is equal to m plus 1 to infinity. So, this is q to the power m that is the first term here will correspond to j is equal to m plus 1, that will give us q to the power m into p, and then we will have infinite geometric series 1 plus q plus q square etcetera. So, this is 1 by 1minus q which is cancelling with q with p, so you will get q to the power m.

(Refer Slide Time: 44:36)



Let us consider probability X greater than m plus n, given that X is greater than n. Now, suppose I consider this as event A, this as event B. So, this is conditional probability of A event B, it is equal to probability of A intersection B. Now, here if you look at the sets A and B, then clearly A is a subset of B, so, this becomes probability of A divided by probability of B. Now, by the formula that we have developed just now, probability of X greater than m is of the form q to the power m therefore, these probabilities will become q to the power m plus n divided by q to the power n, and this is equal to q to the power m which is probability X greater than m. Let us look at this statement carefully, probability X greater than m denotes that the first success is observed after m trials, because we are saying X is the number of trials needed for the first success and I am saying X is greater than m that means, the there is no success observed till mth trial.

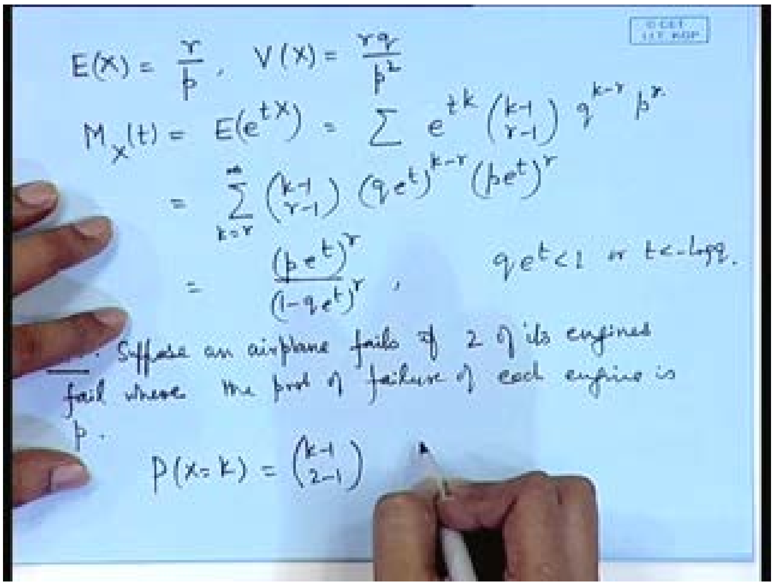
Now, if you look at the left hand side, this is denoting the conditional probability that no success is achieved till m plus nth trial, given that no success is achieved till nth trial. You look at the difference here, n to m plus n that means, m more trials are needed, and here we started from 0 and we say that m trials are needed for the first success that means, in m trials there is no success. So, this means that starting point is immaterial in a geometric distribution, this is called memory less property of a geometric distribution. So, this is quite useful when the independent trials are conducted. So, the starting point does not matter if we are interested into look at how many more trials are needed for a particular kind of either success or failure, because the distribution of success and failure can be interchanged, both will be same.

Now, a further generalization of this distribution can be considered. In the geometric distribution we considered the number of trials needed for the first success now, in place of first success we say that we need third success, seventh success, or a particular rth success. So, if we consider independent Bernoullian trials under identical conditions till rth success is achieved, this is applicable in various industrial applications, etcetera. For example, a particular mechanical system is there which is having several identical components and each component may fail or may keep on working, so, suppose it fails with probability p, and the entire component may fail if say three of the components fail, or entire system fails if five components fail. Suppose you have an airplane, so, there are four engines, the plane can fly if at least two of them are working, so, if both, at least two of them are failing, then the airplane crashes. So, we ((may)) are interested in such kind of events.

So, if we consider X is the number of trials needed for the first time rth success is achieved, then what are the possible values of X? X can take values r, r plus 1 and so on. So, if we write down probability of X is equal to say k, then what this will be? So, 1, 2, 3, 4, etcetera, say, r, r plus 1, k. So, if we are having k trials required for the rth success that means, the last one has to be a success, this is the rth success. So, the probability for that is p and before that in k minus 1 trials, you should have r minus 1 successes and remaining k minus r are the failures. So, the probability of this event will become k minus 1cr minus 1 p to the power r q to the power k minus r, for k is equal to r, r plus 1, etcetera. This is known as inverse binomial or negative binomial distribution; the name inverse binomial, etcetera., is apt here because in the binomial distribution what we do we fix the total number of trials and we see how many successes are observed, here we fix the number of successes and then we see how many trials are required to obtain that many number of successes- so, it is a, something called, so, earlier one if we call binomial sampling, then this is called inverse binomial sampling. So, this is known as negative binomial distribution.

Now, in order to see that it is a valid probability distribution we have to sum it, and if we want to calculate the moments, then the calculations which were shown for the geometric distribution, the similar type of formula will be applicable because it will require the expansions of negative powers of 1 minus q.

(Refer Slide Time: 51:19)



So, if we make use of that formula we will get expectation of X equal to r by p, if we look at say variance of X, we will get rq by p square. We may look at the moment generating function that will be expectation of e to the power tx, that is equal to sigma e to the power tk k minus 1c r minus 1 q to the power k minus r p to the power r, that is equal to k minus 1c r minus 1 qe to the power t to the power k minus r pe to the power t to the power r, k is equal to r to infinity, which is equal to pe to the power t to the power r divided by 1 minus qe to the power t to the power r, which is valid for qe to the power t less than 1 or t less than minus log of q. So, if we are looking at r is equal to 1, this is reducing to the geometric distribution.

Let us look at one application of this negative binomial distribution. Suppose an airplane fails if two of its engines fail, where the probability of failure of each engine is say p, in how many of the flights we will require- like in one flight one engine may fail, in second flight the first engine may fail and so on- so, how many flights will be required for the second engine to fail that is, the first time second engine fails?

So, if we consider probability of X equal to k here, then that will be equal to k minus 1 cr minus 1 that is, 2 minus 1 and p is the probability of failure here, so we need 2, so, p square and q to the power k minus r that is, k minus 2, for k equal to 2, 3 and so on, that is equal to k minus 1 q to the power k minus 2 p square.

So, in the, today we have mainly spent time on one particular kind of trials which are known as Bernoulli trials and in these trials we looked at from different angles. So, we looked at number of success, the distribution of the number of successes, or the number of trials needed for the particular specified number of success. So, these are all called Bernoulli a trials and we looked at various phenomena related to that.

Now, here you can consider it as the approximation of the real life situation because in reality it may not happen that each trial will be independent under identical conditions, but the real probability distributions are approximations of the real life situations and we may have to make certain assumptions to make them applicable. We can also consider finite sampling situations. So, we will be taking up hyper geometric distribution and also other kind of distributions where identical situation may not be there to describe other kind of discrete distributions. So, in the next lectures I will be taking up these issues. Thank you.