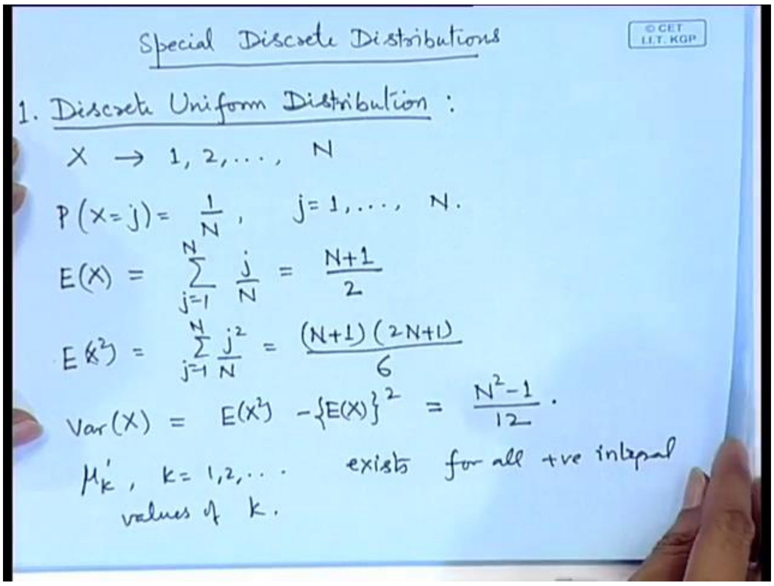
Probability and Statistics
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 19
Special Discrete Distributions – I

Today we will introduce various special distributions which are encountered in various physical and other kinds of experiments.

(Refer Slide Time: 00:38)

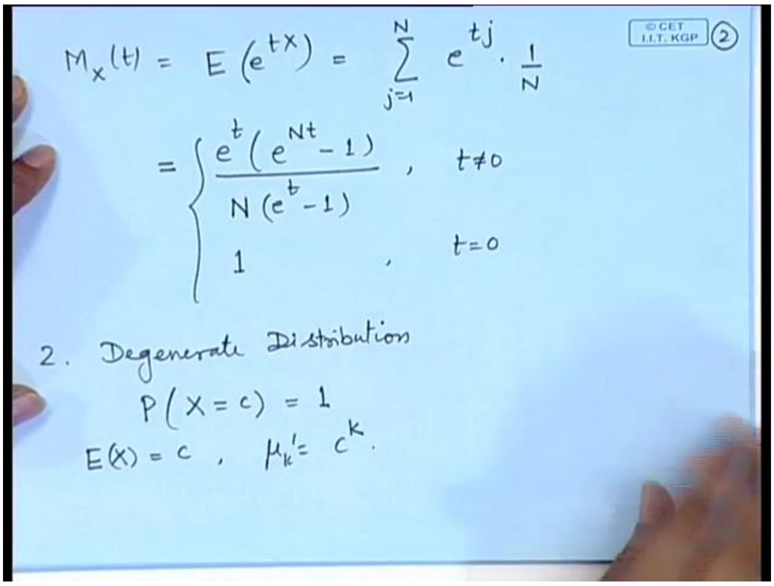


So, broadly we will categorize them into two parts; one is Special Discrete Distributions and another is Special Continuous Distributions. So, the first distribution that I will be taking up is the discrete uniform distribution. So as the name suggest it allots uniform weights to various points. So, a random variable x it takes value say 1, 2 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 2 N plus 1 by 6, so expectation of X square turns out to be N plus 1 into 2 N 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 movement, forth movement etcetera. Since every time, it is a finite sum, the moments of all positive integral orders will exist; that means, you 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: 03:37)



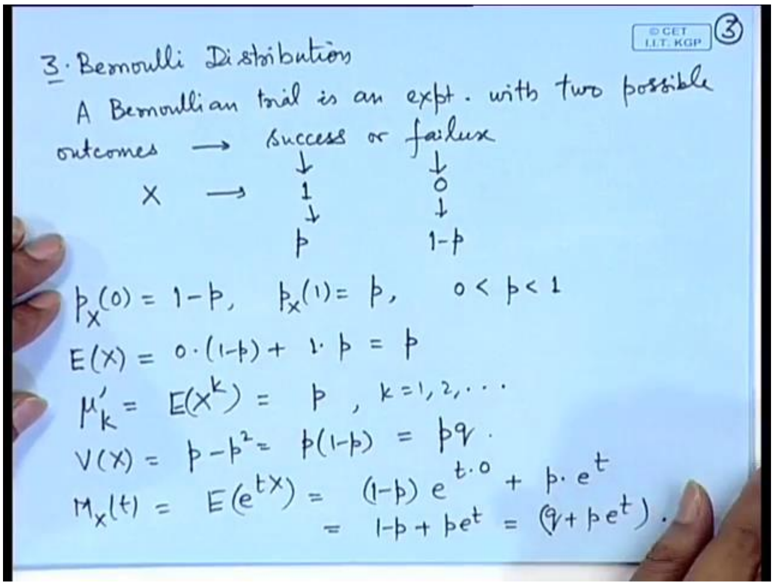
We may also look at the moment generating function of this distribution M x t; that is equal to expectation of E to the power t x. So, it is equal to sigma e to the power t j and each is with probability 1 by N j is equal to 1 to n. Now if you look at this is e to the power t plus e to the power 2 t plus e to the power N t, which is a finite geometric series. So, the sum you can evaluate and it is turning out to be e to the power t, e to the power N t 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 0 and the numerator also becomes 0. 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 denoting function because we 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 say 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 variability 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 maybe that whether the disease is cured or it is not cured. So, generally we call it as a 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: 06:45)



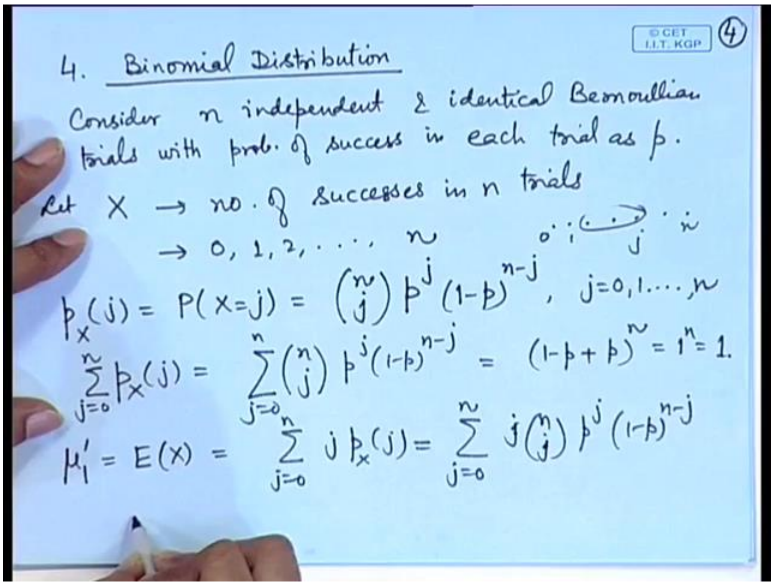
So, if we consider the outcome of one trial then it is known as Bernoulli distribution. So, a Bernoullian trial is an experiment with two possible outcomes that is success or failure. The random variable x will associate the value 1 with success and the value 0 with failure and the probability with success is say p and the probability with the failure is say 1 minus p. So, the distribution is described by p x equal to 0 is 1 minus p and p x equal to 1 is p.

So, this is also a 2 point distribution, so if you 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 summed different numbers 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 it is 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 you look at say variance of this distribution that is equal to p minus p square that is equal to p into 1 minus p; 1 minus p many times we write as q also; so it is p q. The moment generic function is equal to expectation of E to the power t x; 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 1 minus p plus p e to the power t or q plus p e 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: 10:01)



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.

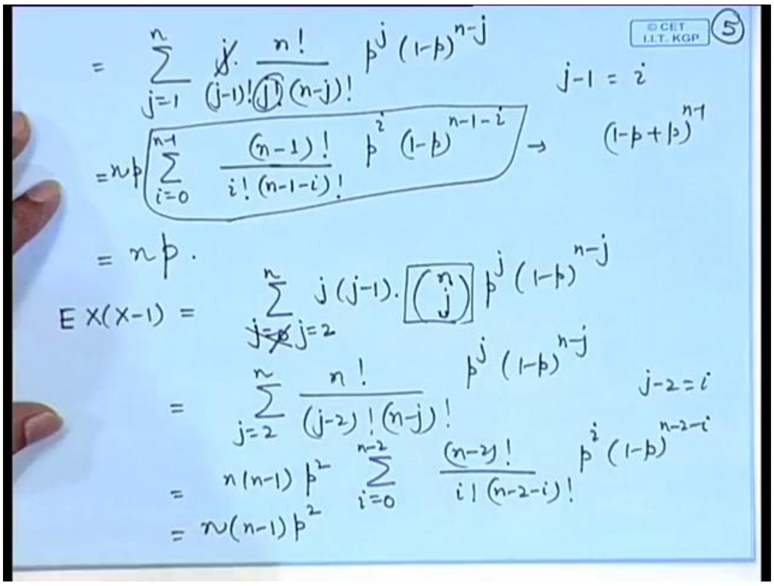
Now it is obvious that if we are looking at total number of trials as N trials and we are fixing some of these j trials as success, then the remaining n minus j must be failures. The probability of a failure is 1 minus p; therefore the probability of n minus j failures will be one minus p to the power n minus j.

Now, out of this n trials this j trials can be selected in N c j possible ways, therefore the probability mass function of the binomial distribution is N c j 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 N c j 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 of successful hits in a basket ball 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 mu 1 prime that is expectation of x, this is equal to sigma j p x; j, j is equal to 0 to n, that is equal to sigma j; n c j, 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 n c j, so somehow it has to be adjusted. So, for this we made certain observation; first thing is that corresponding to j is equal to 0 this term vanishes.

(Refer Slide Time: 14:46)



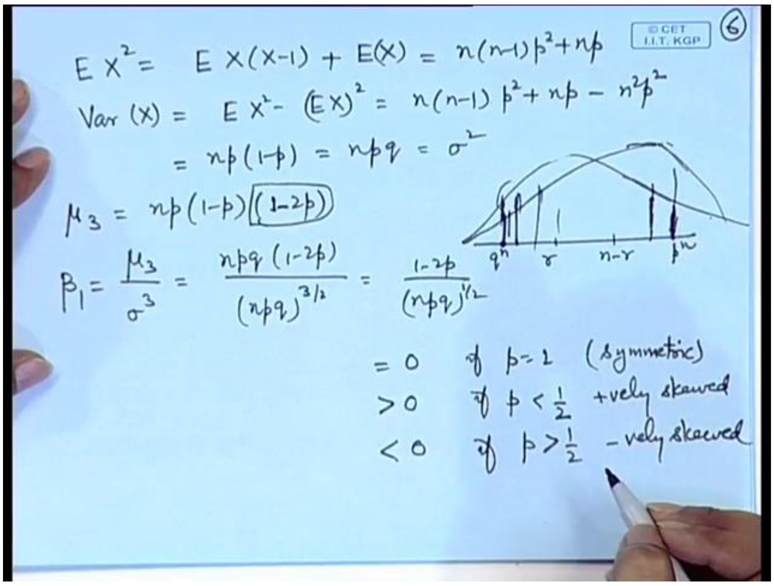
So, in effect this sum is actually considered from j is equal to 1 to n; j and this n c j we can write as n factorial divided by j factorial, n minus j factorial; p to the power j into 1 minus p to the power n minus j; obviously, this j and j factorial we can adjust the term and we can write this as 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 1 p 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 deduces to n p. So, the mean of the binomial distribution is n p 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 n p.

Now, this particular way of deriving the moment of a binomial distributions suggests that if we want to look at say mu 2 prime, then here I will get j square now in this particular expansion 1 j was canceled. So, if I have j square, another j cannot be canceled therefore it suggests that it will maybe 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; n c j 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 n c j, the j factorial term in the denominator can be adjusted with j into j minus 1 and 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: 18:21)



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 n p. 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 1 p square plus n p, minus n square p square. So, here you can observe that the term n square p square cancels out; we are left with n p minus n p square which is n p into 1 minus p or n p q; obviously, if p is a number between 0 and 1 then n p q is going to be less than n p. 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 beta 1 is mu 3 divided by sigma q, so this is sigma square here. Now if we are considering mu 3 this is a third central moment, now third central moment can be expressed in terms of the first three non central moments. Now that means we need mu three prime; now mu three prime 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 2 p q and also if we look at say measure of kurtosis, it will require mu 4, now mu 4 will require mu 4 prime 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 3, p to the power 4.

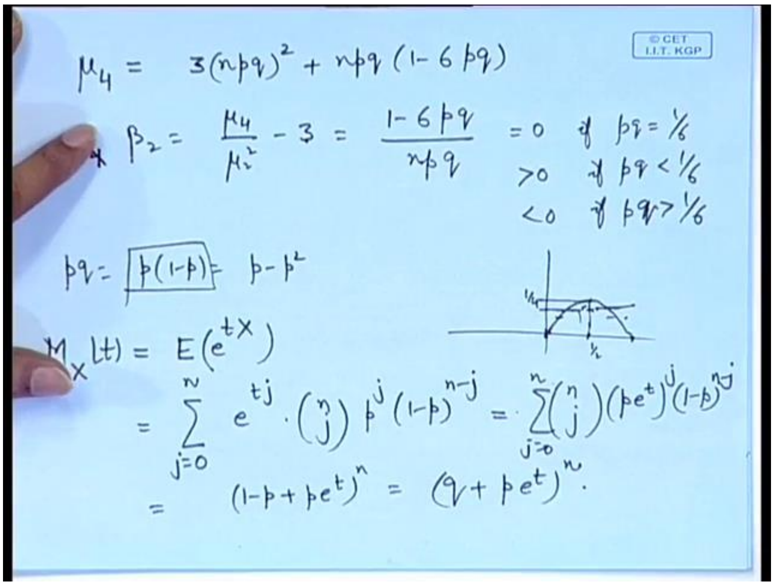
So, after doing certain algebraic simplifications; we can obtain mu 3 as n p into 1 minus p into 1 minus 2 p. Now you can easily see here that the term n p into 1 minus p because it is the variance terms it is always non-negative. Whereas, if you look at this term; this will determine the symmetry of 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 one is n c 1, p into q to the power n minus 1; the last, but 1 is n c n minus 1 which is again same as n c 1; 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 mu 3 divided by sigma cube that is equal to n p q into 1 minus 2 p divided by n p q to the power 3 by 2 that is equal to 1 minus 2 p divided by n p q 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 the in each individual trial the probability of success and failure is the same for example, if you are tossing if you are 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 central moment clearly gives the information about the skewness of the distribution.

(Refer Slide Time: 23:37)



Let us also look at the fourth central moment and as I have explained the method of calculation, you will need to calculate mu 4 prime which will require the fourth factorial moment which can be easily evaluated. After using that mu 4 turns out to be 3 n p q square plus n p q into 1 minus 6 p q. So, if you look at the coefficient beta 2, it is equal to mu 4 by mu 2 square minus 3, so mu 4 is this term and mu 2 is n p q. So, mu 2 square becomes n p q whole square. So, if we consider mu 4 by mu 2 square this will give me 3 here and 3 minus 3 will cancel out. So, we are left with 1 minus 6 p q divided by n p q. So; obviously, this is equal to 0 if p q is equal to 1 by 6, it is greater than 0 if p q is less than 1 by 6, it is less than 0 if p q 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 be slightly higher than the normal, if p q 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 beta 1 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 0; that means, it will become closer to asymmetric distribution. In a similar way, if we look at the coefficient beta 2, here in the denominator we have n and therefore, as n becomes large; the measure of kurtosis is closer to 0 even though p q 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 t x; that is equal to e to the power t j; n c j; p to the power j into 1 minus p to the power n minus j, j is equal to 0 to n.

Now, if you are making use of the binomial expansion then it is clear that the term e to the power t j must be adjusted with the term p to the power j. So, this becomes n c j; p e 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 p e to the power t whole to the power n; that is q plus p e to the power t whole to the power n, so for all values of t this is well defined.

Thank you.