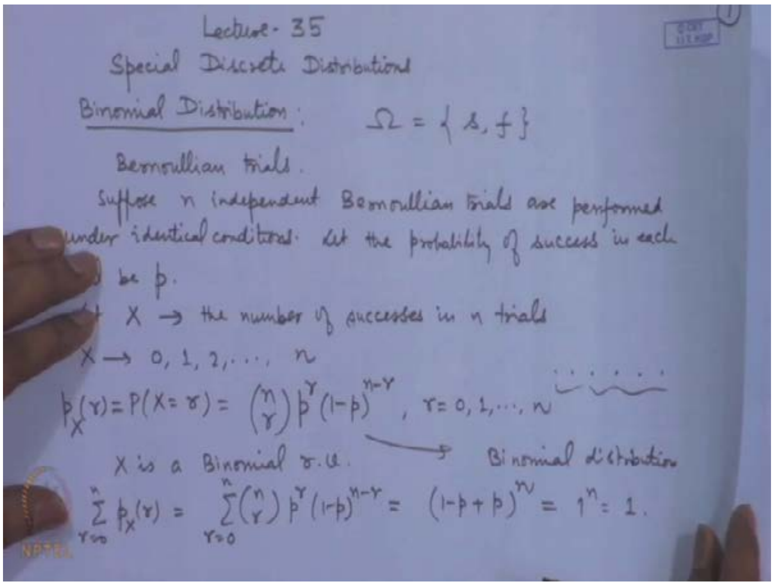
Advanced Engineering Mathematics

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

Lecture No. # 36
Special Discrete Distributions

Today, we will introduce a special discrete distributions, these distributions are the once which have been used quite frequently in practice and they arise in various natural phenomena.

(Refer Slide Time: 00:39)



So, firstly, let me start with binomial distribution. Now in the binomial distribution, we are considering random experiments in which the outcomes can be described into two types of outcomes - one is called success and one is called failure. So, my sample space consists of two outcomes, now this type of situation arises, for example, let us consider a game of target hitting. So, at each trial, we may hit the target or we may not hit the target. Suppose, we are considering treatment of a disease then for each patient, the patient may get cured or he may not get cured. A student appears in a qualifying examination, he may qualify or he may not qualify. In a game of say tennis at each hit of the ball by a player by the racket, the shot may be a winner or it may not be a winner.

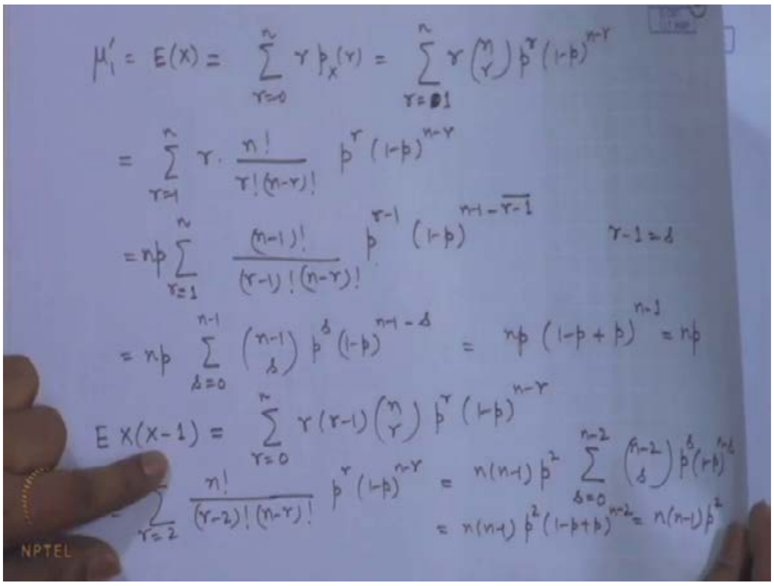
So, this type of situation then can described very large number of random experiments, where we are finally interested only in success or failure. So, if we say that the experiments are conducted under independent and identical condition then they are these are called independent and identically conducted Bernoullian trials Bernoullian trails. So, suppose we say that n independent Bernoullian trials are performed under identical conditions.

Let the probability of success in each trial be p. Now, let us consider the number of successes, the number of successes in n trials. Then what are the possible value that X can take? X can take values 0, 1, 2 and so on. So, what is the probability that say X is equal to r then we are having n trials out of this r of them are success. So, the probability of r successes will be p to the power r and then remaining n minus r will be failures. So, the probability of failure for each trail is 1 minus p, in n minus r trial, it will be 1 minus p to the power n minus r. Now out of these n trials, any of the r trials can be success. So, this can be selected in n c r ways, and therefore the probability mass function of this random variable X is given by n c r p to the power r 1 minus p to the power n minus r, here r can take values 0, 1 to n.

So, if we follow a usual notation for the probability mass function, we will write it as p X r that is the probability that X is equal to r. Then X is called binomial random variable and this is called a binomial distribution. Now this Bernoullian trials are named after the mathematician Bernoulli - James Bernoulli and because he was the first one would described this experiment. Now the name binomial distribution as come because of the use of the binomial coefficient and actually in order to evaluate that, suppose we consider the sum of this probabilities then that is equal to sigma n c r p to the power r 1 minus p to the power n minus r; r is equal to 0 to n. Then this is nothing but the sum of 1 minus p plus p to the power n, and therefore this can be consider as 1 to the power n that is equal to 1.

Let us look at the properties of this binomial distribution in the last lecture I introduce various characteristic of a distribution, for example, mean of a distribution, variance of a distribution or in general moments of a distribution. So, in the moments we had considered non central and central moments. So, based on that I had introduced the concept of a measure of symmetry or a skewness and measure of kurtosis that is the peakedness of a distribution.

(Refer Slide Time: 06:38)



Now in the context of these distributions I will calculate these characteristics and see how they look like. Let us consider, so mu 1 prime that is the mean or the expectation of this random variable. So, that will be equal to by definition r p X r,r is equal to 0 to n that is equal to sigma r n c r p to the power r 1 minus p to the power n minus r,r is equal to 0 to n.

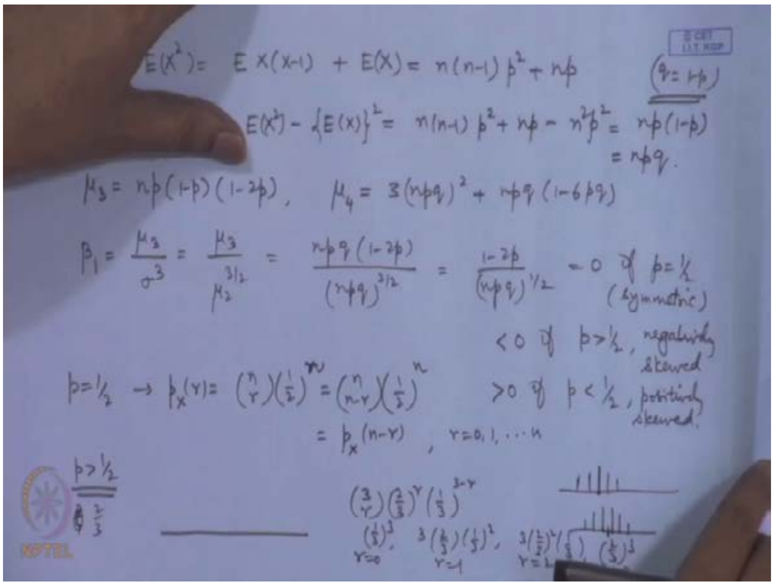
Now, you can observe that corresponding to r is equal to 0 this term is actually 0. So, basically this starts from 1 we can write like that it is from 1 to n. Now what we do? We have noticed here that when we calculated the sum we actually interpreted this as a binomial sum. So, if you interpret this as a binomial sum then we should be able to interpret this part also as a binomial sum then only we can actually evaluated. So, in order to do that we expand this factorial, this combination term in the factorials n factorial divided by r factorial n minus r factorial p to the power r 1 minus p to the power n minus r; r is equal to 1 to n. Now this r and r minus 1 you can adjust here so, this you can write as n minus 1 factorial divided by r minus 1 factorial n minus r factorial p to the power r minus 1, 1 minus p to the power n minus 1 minus r minus 1.

So, what I have done I have add this to this r minus 1 here and then this term I have written n minus 1 factorial; that means, I have separated out n and I have also taken out p from here and this is from r is equal to 1 to n. So, if I substitute r minus 1 is equal to say s then this will become n p s is equal to 0 to n minus 1 n minus 1 c s p to the power s 1 minus p to the power n minus 1 minus s, this is nothing but the binomial expansion of 1 minus p plus p to the power n minus 1, so this is again 1 and we get it as equal to n p. So, the mean of a binomial random variable is n p, now let us look at the physical interpretation of this, in each trial the probability of success is p, so in n trials the expected number of successes is n times p. So, that is justified here.

Now in order to calculate higher order moments, for example, if I want to calculate the variance of this distribution I need expectation of X square. Now, the way we have done the calculation for expectation of X we have actually ((is plate)) the combination term and cancelled out one of the term in the factorials. Therefore, it is beneficial if I calculate the so called factorial moments. So, let us look at expectation of X into X minus 1. So, that is equal to sigma r into r minus 1 n c r p to the power r into 1 minus p to the power n minus r; r is equal to 0 to n. Once again you notice here that corresponding to r is equal to 0 and r is equal to 1 this term will vanish. So, we can write it as r is equal to 2 to n; n factorial divided by r minus 2 factorial n minus r factorial p to the power r 1 minus p to the power n minus r.

So, if we follow a scheme similar to this we can express it as n into n minus 1 p square sigma s is equal to 0 to n minus 2 n minus 2 c s p to the power s 1 minus p to the power n minus s. So, this is nothing but n into n minus 1 p square into 1 minus p plus p to the power n minus 2, now this term becomes 1, so we get n into n minus 1 p square. So, this is expectation of X into X minus 1

(Refer Slide Time: 11:11)



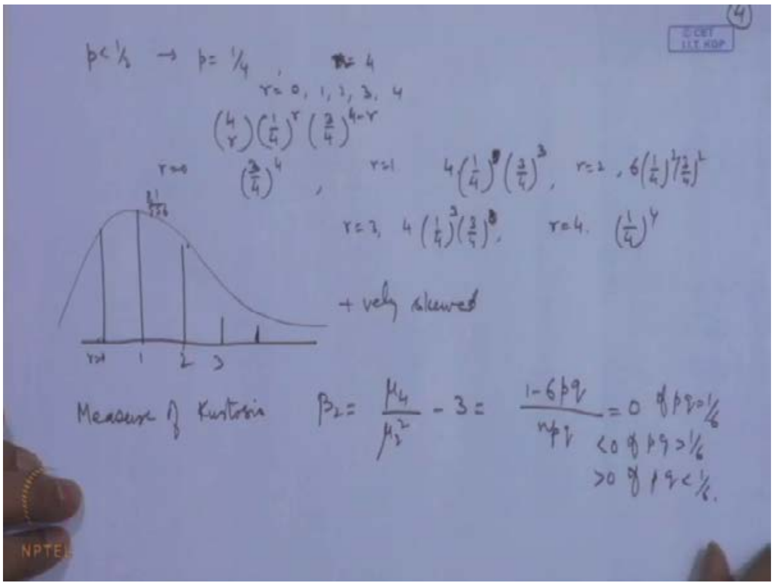
Now, we can get mu 2 prime that is equal to expectation X is square expressed as expectation of X into X minus 1 plus expectation of X, because expectation X will cancel out here, Now these 2 terms we have evaluated this is equal to n into n minus 1 p square and expectation X is equal to n p. So, we get the second non central moment of a binomial distribution, now this we can use to calculate the variance of the binomial distribution that is mu 2 that is variance of X. And we have defined it as expectation of X square minus expectation of X whole X square, now this is equal to n into n minus 1 p square plus n p minus n square p square. So, here n square p square cancels out we are getting the term as n p in to 1 minus p in most of the practical applications we use the notation 1 minus p is equal to q. So, this is also written as n p q where q is defined as 1 minus p. Now in a similar way we can calculate higher order moments for example, expectation X q for that we will calculate the third factorial moment that is expectation X into X minus 1 into X minus 2 and so on. So, without getting in to the technical details of this I will give the final expression for this.

We get mu 3 as equal to n p into 1 minus p into 1 minus 2 p and similarly the forth central moment mu 4 trans out to be equal to 3 n p q square plus n p q into 1 minus 6 p q. Now if we consider say beta 1 that is the measure of skewness that is mu 3 by series notation we can also write sigma square for the variance then this is mu 3 by sigma q bar mu 3 divided by mu 2 to the power 3 by 2. So, 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. Now note here that if p is equal to half then this is equal to 0, this is corresponding to symmetry that is the distribution is symmetric it is less than 0 if p is greater than half; that means, it is negatively skewed distribution. If it is greater than 0, p is less than half that is positively skewed. Let us look at the physical explanation for this, the probability mass function of the binomial distribution is n c r p to the power r 1 minus p to the power n minus r.

Certainly if p is equal to half, see p is equal to half then you will get p X r is equal to simply n c r half to the power r that is equal to n c n minus r half to the power sorry half to the power n that is equal to p X n minus r that means the probability for X is equal to r and probability X equal to n minus r is same for r is equal to 0, 1 to n. That means, the distribution is symmetric about the mid value here. It could be n by 2 or it could be mid value of n, there could be two middle points also.

For example, we may have this situation. So, in the art case it is symmetric around this midpoint and if it is even then it is symmetric about the midpoint here. If p is greater than half if p is greater than half then your initial probabilities they will be… So, let us write down some particular cases suppose I take n is equal to 3, so 3 c r and say p is equal to I take 3 by 4 or set 2 by 3. So, then this becomes 2 by 3 to the power r 1 by 3 to the power 3 minus r, now correspondent to r is equal to 0 this value will become 1 by 3 q; then next value will become 3 2 by 3 into 1 by 3 square then 3 2 by 3 square 1 by 3 and 2 by 3 cube. You can see here this is for r is equal to 0 r is equal to 1 r is equal to 2 and r is equal to 3. So, you can notice here the probabilities are this 2 by 3 cube is much bigger than 1 by 3 cube. So, 1 by 3 cube is here, then 3 into you are getting 6 by 27. So, that is somewhere here, then next value is 3 into 412 by 27 and then you are having 8 by 27.

(Refer Slide Time: 17:26)



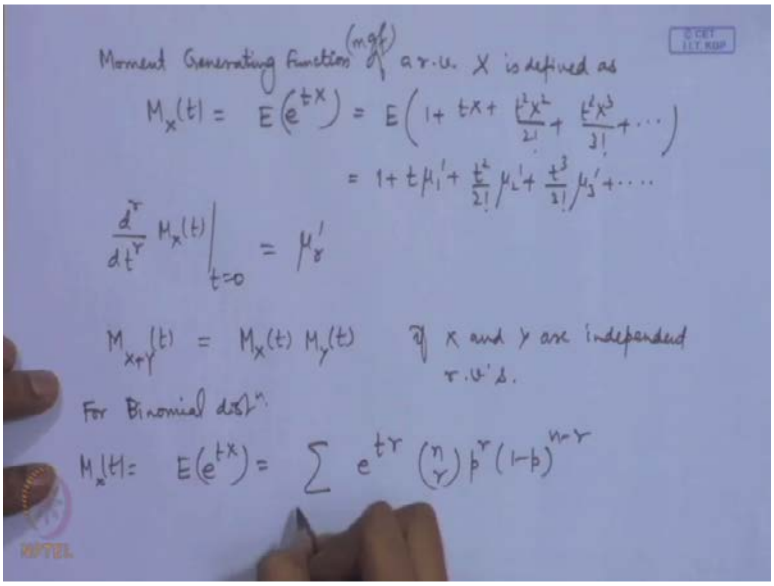
So, 0, 1, 2, 3. So, this distribution is if you say try to join by this is negatively skewed. Similarly if I take p less than half if I take p less than half suppose I take p is equal to say 1 by 4 and then, let us write down and I take the case r is equal to say 4 sorry n is equal to 4, then you will have r contain values 0, 1, 2, 3, 4. Let me write down the probabilities here 4 c r 1 by 4 to the power r 3 by 4 to the power 4 minus r. Now corresponding to r is equal to 0, this will be 1 by 4 to the power 4 corresponding to r is equal to 1 this will be 4 1 by 4 cube 3 by 4 corresponding to r is equal to 2 this will be 4 c 2 that is 6 1 by 4 square 3 by 4 square corresponding to r is equal to 3. This will become 4 1 by 4 into 3 by 4 cube corresponding to corresponding to r is equal to 4 this will become I think I made a mistake here corresponding to r is equal to 0 this will be 3 by 4 to the power 4 corresponding to r is equal to 1 this will be 4 1 by 4 3 by 4 cube corresponding to r is equal to 2; it is 6 into 1 by 4 square 3 by 4 square corresponding to r is equal to 3 this is 4 into 1 by 4 cube 3 by 4 and corresponding to r is equal to 4 this will become 1 by 4 to the power 4.

So, if you plot these values here r is equal to 0, you see this is 3 by 4 to the power 4 that is it t 1 by 256. So, some varier corresponding to r is equal to 1, you can see here the value is now 3 cube that is 27 into 4 that is 108 by 256 corresponding to r is equal to 2 it is 54 by 256, that is it is coming down corresponding to 3 this is standing out to be 3 by 256 corresponding to r is equal to 4 it is 1 by 256. So, it is there is a steep decline. So, this is positively skewed. So, this is conform from this calculations here, the exact that is if p is equal to 0 you have a symmetric distribution, if p is greater than half the distribution is negatively skewed, if p is less than half the distribution is positively skewed.

We also look at mu 4, here mu 4 is equal to 3 n p q square plus n p q into 1 minus X p q, this gives as the major of kurtosis, that is beta 2 that is equal to mu 4 by mu 2 square minus 3. So, that terms out to be 6 1 minus 6 p q divided by n p q. Naturally this is equal to 0, if p q equal to 1 by 6 it is less than 0 if p q is greater than 1 by 6 it is greater than 0 if p q is less than 1 by 6. Now this will give a quadratic inequality and therefore, you can solve that to get the range for negative and positive values.

In the context of the characteristic of distribution, there is one important function which is called the moment generating function of a distribution.

(Refer Slide Time: 21:32)

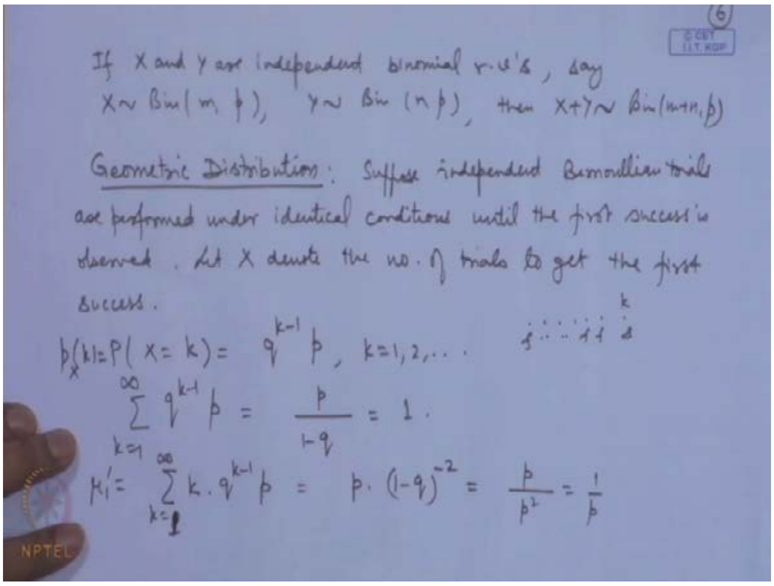


So, moment generating function of a random variable X is defined as M X t, that is equal to expectation of e to the power t X; why is it called moment generating function? Because suppose this exist, then we can consider the expansion of this in Maclaurin series, and apply this expectation linearly then you get t expectation X that is mu 1 prime plus t square by 2 factorial mu 2 prime plus t cube by 3 factorial mu 3 prime and so on. That means, the series that we are getting is an infinite power series in t with the coefficient of t to the power k by k factorial as the k eth non-central moment.

And if I consider say r eth derivative of the moment generating function at t equal to 0, that is the r eth non-central moment. That is why it is called the moment generating function of a random variable or m g f? It has some important properties, for example, a moment generating function you neatly determines a distribution; that means, two distributions - different distributions will have different m g f, and it is also very useful in derivation of certain distributions. For example, if I have 2 random variables which are independent, and if I am considering moment generating function of a sum, then it is equal to moment generating function of the product, if X and Y are independent random variables.

Now, let us look at this in context of the binomial distribution. For the binomial distribution, moment generating function that is equal to expectation e to the power t X that is equal to sigma e to the power t r n c r p to the power r into 1 minus p to the power n minus r; r is equal to 0 to n. This we can write as sigma n c r 1 minus p to the power n minus r p e to the power t to the power r, r is equal to 0 to n which is nothing but the expansion of 1 minus p plus p e to the power t to the power n, that is q plus p e to the power t to the power n.

(Refer Slide Time: 24:50)



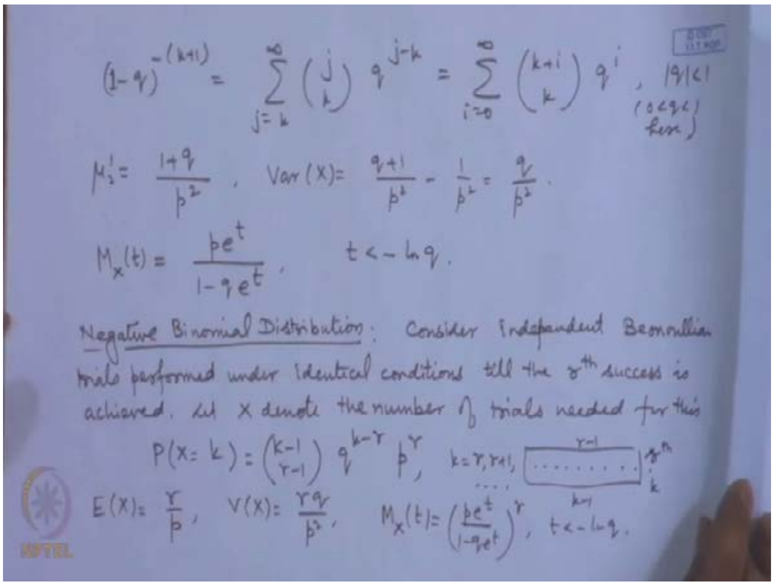
So, the moment generating function of a binomial distribution with parameters n and p can be written as q plus p e to the power t whole to the power n. See, if I say if X and Y are independent binomial random variables say X follows binomial m p and Y follows binomial n p then X plus Y follows binomial m plus n p, this you can easily prove using the moment generating function. Now in the Bernoullian trials we can also look at in a different way, we introduce what is called a geometric distribution? Now, what is the geometric distribution? Suppose independent Bernoullian trails are performed under identical conditions until the first success is observed.

Let X denote the number of trials to get the first success then what is the probability of X equal to say k. Now the trails are performed and all of them are failure till the X eth trial. So, if I am saying, this is k eth trail where the success is observed, before that all of them are failures; that means, k minus 1 failure are there. So, you have q to the power k minus 1 into p, where k can take values 1, 2 and so on. So, this is the probability mass function of this distribution; this is known as the geometric distribution. The reason is that, if I consider, the sum q to the power k minus 1 p for k equal to 1 to infinity, this is nothing but the infinite geometric series that is equal to p by 1 minus q that is p by p that is equal to 1.

Let us look at the mean of this, that is equal to sigma k q to the power k minus 1 p k equal to 1 to infinity. Now this is the nothing but the infinite arithmetic geometric series and the sum of this is simply p into 1 minus q to the power minus 2 that is equal to p divided by p square that is equal to 1 by p. That means, if in each trial the probability of successes p then the number of trials - expected number of trials needed for the first success will be 1 by p. So, it is something like if you consider a coin tossing experiment and in the coin tossing experiment, if the coin is unbiased then the probability of head is half. So, expected number of trials needed to get the first head that will be 1 by half that will be equal to 2; that means, on the average two trials will quite to get the first head.

Similarly, suppose I am considering a fair die and the probability of say observing a 6 for the first time. So, that will be 1 by 6, now what is the expected number of trials needed to get the first success? first time head is coming sorry first time 6 is coming. So, that will become 1 by 1 by 6 that is equal to 6, that is on the average 6 trails will be required to get a particular face. Now the moments of the geometric distribution can be calculated using this type of provision. So, I will give the general formula for that actually.

(Refer Slide Time: 29:16)



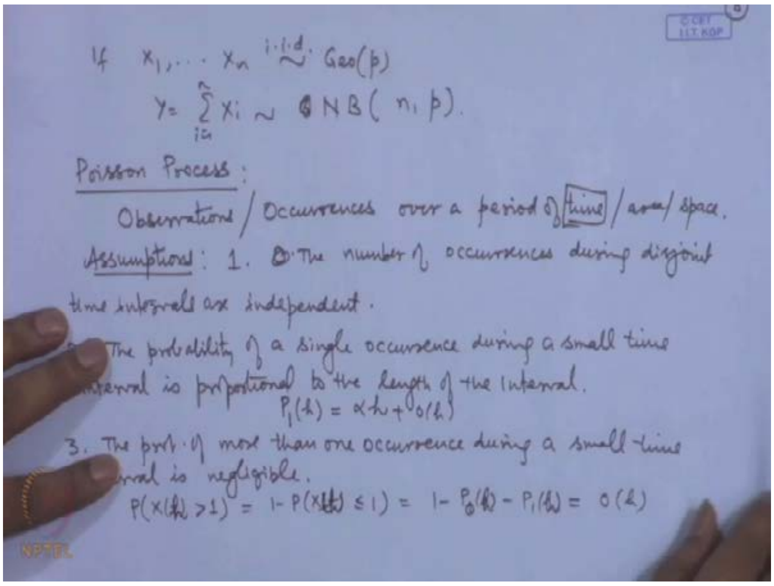
If we consider say 1 minus say q to the power minus k plus 1 that is equal to sigma j c k q to the power j minus k for j is equal to k to infinity which can also we expressed as r is equal to 0 to infinity k plus i c k q to the power i, where q is between minus and 1. Here of course, q will (( )) between 0 to 1, using this we can calculate the second moment. So, this turns out to be 1 plus q by p square and the variance will turn out to be q plus 1 by p square minus 1 by p square that is equal to q by p square. The moment generating function of geometric distribution is p e to the power t divided by 1 minus q e to the power t, where t is less than minus log of q.

Now, here in the geometric distributions we are conducting the Bernoullian trails till we get the first success. Now in place of the first success, we need a specified number of success, it could be various kind of experiments where for example, you consider certain machinery which has several identical components which are part of that, and the machine will work if a specified number suppose I say five of them are working. So, suppose total number of components are ten or fifteen etcetera and suppose five of them are working, then the system works and the system will fail if a specified number fails. So, for example, first time when four components fail or first time two components fail. So, in place of first time success or first time failure, if you look at first time r eth the success or r eth failure this is the generalization of the geometric distribution and it is called negative binomial distribution.

So, let me introduced this one, negative binomial distribution. Consider independent Bernoullian trials performed under identical conditions till the r eth success is achieved. So, let X denote the number of trials needed for this, then what is the probability that X is equal to say k. Now you see here the trials are getting perform and on the k eth trial. So, this is first time r the success is observed; that means, before to this there are k minus 1 trials out of this k minus 1 trials r minus 1 success should be there, that mean this can be done in k minus 1 choose r minus one base. So, now you have out of this k minus r failures will be there. So, q to the power k minus r and p to the power r, because r successes r their their k can take values r plus 1 and so on. So, this is called negative binomial distribution; the mean of negative binomial distribution is r by p. The variance of negative binomial distribution is r q by p square, the moment generating function of negative binomial distribution is p e to the power t divided by 1 minus q e to the power t whole to the power r for t less than minus l n q.

Now, you notice here the moment generating function of geometric and moment generating function of the negative binomial this is power r of this. So, we can establish a relationship that.

(Refer Slide Time: 34:39)



If say X 1, X 2, X n are independent and identically distributed geometric random variables, then Y that is equal to sigma X i; i is equal to 1 to n, that will follow negative binomial n p distribution, that is the sum of independent geometric random variable is a binomial negative binomial distribution. Now it can be easily explained also from physical phenomena. If I am considering X 1 here, now X 1 is what? It is the number of trials needed for the first time a success is observed, X 2 is also the number of trials needed for the first time a success is observed, X n is the number of trials needed for the first time success. If I consider X 1 plus X 2 plus X n, what does it denote? It will represent the number of trials needed for the first time n eth success is observed. And therefore, they should be negative binomial, because I am considering identical Bernoullian trials performed under independently, and therefore this will become negative binomial random variable.

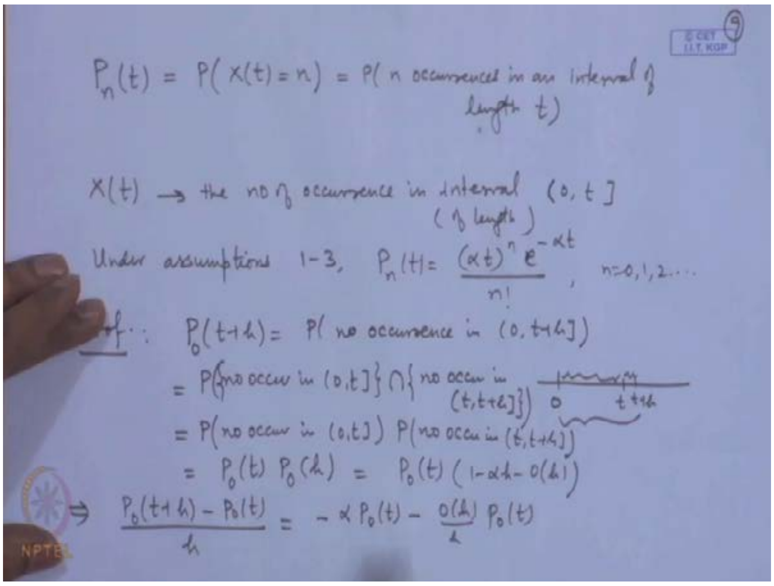
Now let me introduce another important discrete distribution that is known as Poisson distribution. So, we introduce what is called a Poisson process? So, when we observe certain phenomena such as the number of accidents occurring at a particular traffic junction over a period of time. Suppose, we observe the number of telephone calls recorded at a telephone junction. Suppose we record the number of earthquakes in a geographical region over a period of time. Suppose we observe the number of say astronomical events are say comet the observing of a comet etcetera in a space over a period of time. Many of these events satisfies certain assumptions, these assumptions are called assumptions of a Poisson process.

So, first thing that we notice here, that we are observing events over a period of time, over area, over space. So, for convenience we consider observations occurrences over a period of time area space etcetera, for convenience we will restrict our attention to time. So, we make the following assumptions, the first assumption is that the number of occurrences during disjoint time intervals are independent. Now when we change the time to area then over different geographical regions they will be independent or if you are observing over a space then over the different regions of a space they will be independent. So, when we say time. So, it means that suppose we are observing number of traffic accidents occurring. So, if we consider say time between eleven o'clock to twelve o'clock and we consider a time between four o'clock to five o'clock then the number of accidents observed during 11 to 12 or 4 to 5, they will be independent.

Suppose we are observing say a phenomena such as earthquake then the number of earthquake occurring over say Asian region may be independent of the number of earthquakes occurring over say European region etcetera. The probability of a single occurrence during a small time interval is proportional to the length of the interval. Once again if we replace the time by area then it will be proportional to the area of that region, if we replace it by region in the space then it will be proportional to the volume of that a region etcetera and the third assumption is that the probability of more than one occurrence during a small time interval is negligible; that means, more than one occurrence during a small time interval is negligible or it is very very small or it can be ignore.

Let me introduce some notation to express this.

(Refer Slide Time: 40:56)



So, let us use the notation say P n t now if I denote X t to be the random variable that is the number of occurrences in interval say 0 to t. We have kept the intentionally one side opened and one side closed. So, this means interval of length t if we are making the assumption of the stationarity, that is the independent intervals then 0 to t is the same as a to a plus t or from any point. If I am starting and if I am considering length t then it is same. So, we can consider without loss of general t 0 to t, then we use the notation P and t that probability of X t is equal to n; that means, in the interval of length t there are n occurrences - that is probability of n occurrences in an interval of length t now in the framework of this P n t notation. We can expressed this assumptions the probability of a single occurrence during a small time interval, single occurrence then means P 1 and that small thing will may use an notation h; that means, one occurrence in an interval of length h is proportional to the length of the interval. So, proportional means we can use the notation alpha h that is the constant of proportional t is taken to the alpha.

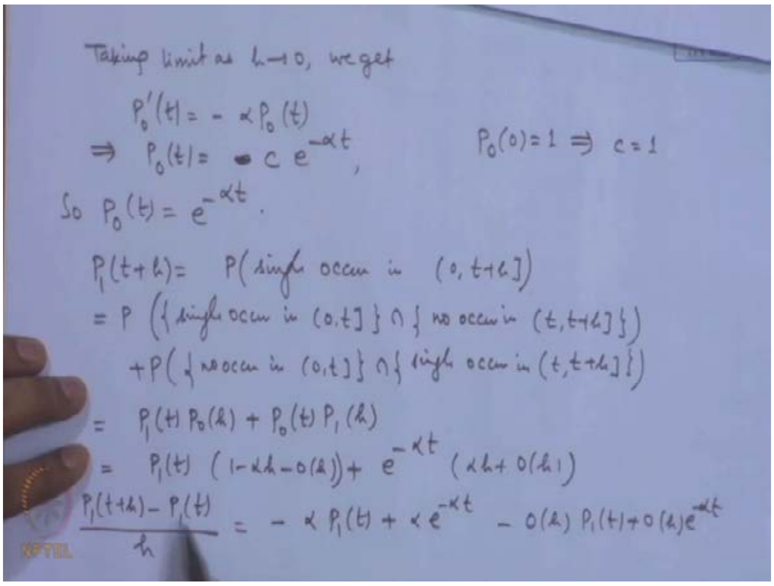
Similarly probability of more than 1 occurrence, now more than 1 occurrence means probability of X t is greater than 1 that can be written as 1 minus probability of X t is less. It is actually equal to 0 more than 1 occurrence yeah more than 1 occurrence is greater than one. So, this is less than or equal to 1 that is equal to 1 minus P 1 0 and P 1 1 this is negligible, now for negligible t we use a notation and again we are considering small time interval. So, X h X h. So, P 1 h P 0 h etcetera that is equal to I am assuming negligible. So, we are using small o notation here, that it is equal to o h and here also we may introduced o h here. So, these are the assumption that we are having 1 minus P 0 h minus P 1 h is negligible and P 1 h is equal to alpha h, and we have just added a negligible amount here. It will not mechanic different to this.

So, under assumptions 1 to 3 P n t is equal to alpha t to the power n e to the power minus alpha t by n factorial n is equal to 0 1 2 and so on, that is the number of occurrences in a Poisson process. So, this is the occurrences which satisfy these assumptions they are called occurrences in a Poisson process. And then tire phenomena is called a Poisson process and we are deriving the distribution of the number of occurrences in a Poisson process as alpha t to the power n e to the power minus alpha t by n factorial what is alpha, here alpha is the constant of proportional t that we have assumed here, let me give a prove of this we start with 0 let us consider say P 0 t plus h.

Now P 0 t plus h means no occurrence in the interval 0 to t plus h. Now this we can write as now let us consider on the scale suppose this is t 0 this is t and t plus h is here if I say that there is no occurrence in 0 to t plus h, this mean that there is no occurrence in 0 to t there is no occurrence in t to t plus h that means. We can say it is no occurrence in 0 to t and So, we can write it as the event intersection no occurrence in t to t plus h.

Now if we look at this interval this interval is disjoint from this interval using the first assumption these two events are independent. So, we can write it as probability of no occurrence in 0 to t into probability of no occurrence in t to t plus h. Now this is P 0 t and this is an interval of length h the starting point may be t, but the length is… So, we can use the notation P 0 h. Now P 0 h we have expression here 1 minus P 0 h minus P 1 h is equal to o h and P 1 h is alpha h plus o h. So, if we substitute here we get P 0 h is equal to 1 minus alpha h minus o h 2 times o h is same as writing once. So, this is equal to P 0 t 1 minus alpha h minus o h. So, we can write it as P 0 t plus h minus P 0 t divided by h that is equal to minus alpha P 0 t minus o h by h P 0 t in this 1 if you take the limit as h standing to 0.

(Refer Slide Time: 47:29)

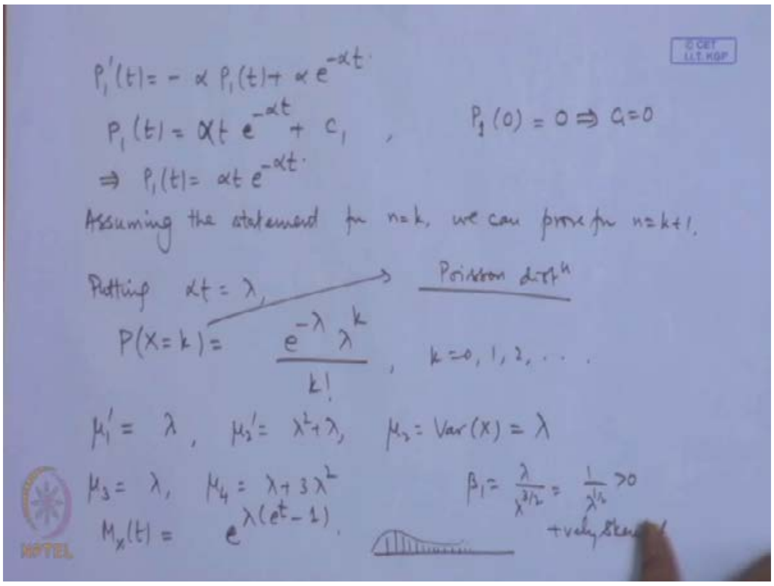


Taking limit as h tends to 0, we get P 0 prime t is equal to minus alpha P 0 t this is nothing but a first order differential equation which is just like a variable separable. So, if we simplify this you get P 0 t is equal to minus a constant times e to the power minus alpha t. Now this constant can be determine by the initial condition that p 0 0 equal to 1. So, if you substitute this here we get c is equal to 1. So, the ((solution)) is equal to P 0 t is equal to e to the power minus alpha t.

Now if we look at this expression, here P and t here if you put n is equal to 0 we get e to the power minus alpha t; that means, we have prove this statement for n is equal to 0. Now in a similar way we can prove for 1 and then so on, if you consider say P 1 t plus h that is equal to probability of single occurrence in the interval 0 to t plus h. Once again, let us look at this interval if we say from 0 to t plus h there is 1 occurrence then that 1 occurrence can be in 0 to t or it could be from t to t plus h. So, we can ((exclude)) this event as probability of single occurrence in 0 to t and no occurrence in t 2 t plus h plus probability of no occurrence in 0 to t and single occurrence in t 2 t plus h.

Once again we can use the independence and we get it as P 1 t P 0 h plus p 0 t P 1 h that is equal to now P 1 t and then P 0 h we have calculated as 1 minus alpha h minus o h plus P 0 t value. We have already evaluated e to the power minus alpha t P 1 h is alpha h plus o h. So, from here I can again set of the differential equation P 1 t plus h minus P 1 t divided by h that is equal to minus alpha P 1 t plus alpha e to the power minus alpha t minus o h P 1 t plus o h e to the power minus alpha t.

(Refer Slide Time: 50:33)

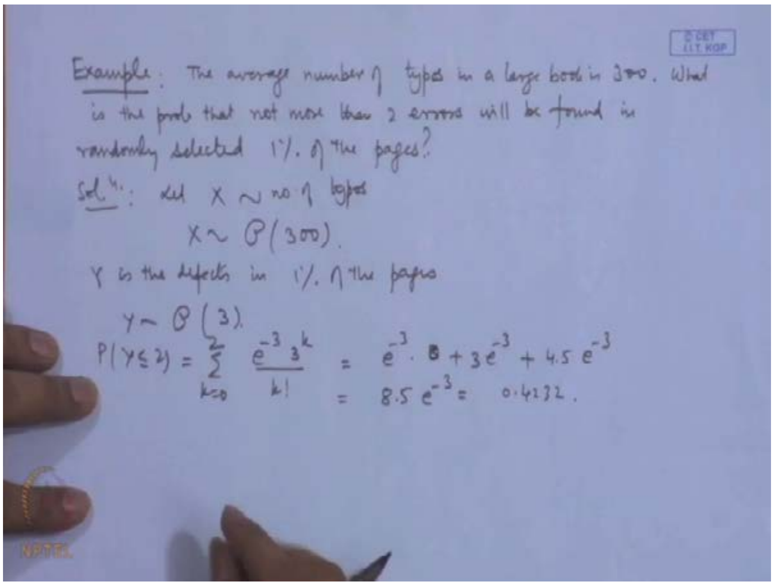


So, if I take extending to 0. We get here P 1 prime t is equal to minus alpha P 1 t plus alpha e to the power minus alpha t. This is nothing but first order linear differential equation. So, we can solve it easily and the solution turns out to be P 1 t is equal to lambda t alpha t e to the power minus alpha t plus a constant. Now, once again we can use the initial condition P 1 0 that is the probability of the single occurrence in the interval of length 0 that will be 0 if you substitute this will get c 1 is equal to 0. So, this means P 1 t is equal to alpha t e to the power minus alpha t. Now once again if we note here the general expression that we wanted to prove here in this 1 if you put n is equal to one. We get P 1 t is equal to alpha t into e to the power minus alpha t. So, we have prove this statement for n is equal to 1 also. So, assuming the statement for n is equal to k we can prove for n is equal to k plus 1. So, this result is proved.

So, now in practice ((generally)) what we do? We can substitute this alpha t as some lambda and we can write down the expression for the distribution has putting alpha t is equal to lambda. We write the distribution probability X equal to k is equal to e to the power minus lambda lambda to the power k by k factorial, for k equal to 0, 1, 2, and so on. So, in popularly this is known as the Poisson distribution that is the distribution of the number of occurrences in a Poisson process and when we are considering the interval of length t and we multiply that alpha into t then that gives the value lambda. So, this is called the Poisson distribution.

Now we just give the expressions for the mean variance etcetera; the mean of the Poisson process is equal to the parameter lambda here, mu 2 prime is equal to lambda square plus lambda and therefore, mu 2 that is the variance is again lambda. So, in a Poisson distribution mean and variance are same. Similarly if you calculate the third moment that is also lambda the fourth moment is lambda plus 3 lambda X square, the moment generating function of a Poisson distribution is lambda e to the power lambda e to the power t minus 1. If you look at the behavior of this distribution here see if you consider say beta 1 that will be equal to lambda by lambda to the power 3 by 2 that is equal to 1 by lambda to the power half this is positive. So, the Poisson distribution is somewhat positively skewed actually as lambda increases the probabilities converge towards zero.

(Refer Slide Time: 54:20)



So, this is a positively skewed distribution. Let me give one example, here the average number of typos in a large book is three hundred what is the probability that not more than two errors will be found in randomly selected one percent of the pages? Let us look at the solution of this. So, let us consider X to be the number of typos. So, X follows Poisson distribution with parameter 300 now Y is the number of defects in one percent of the pages. So, I will follow Poisson 3 now we want probability of Y less than are equal to 2 that is equal to e to the power minus 3 3 to the power k by k factorial sigma k equal to 0 to 2 that is equal to e to the power minus 3 three. So, when we put k equal to 0 I get e to the power minus 3 for k equal to 1 I will get 3 to the power minus 3 for k equal to 2 I will get 4.5 e to the power minus 3 that is equal to 8.5 e to the power minus 3 that is equal to 0.4232.

So, we can evaluate by applying the assumptions of the Poisson process here; that means, we are assuming that the number of typo graphical errors in different pages or different areas of the book they are independent the probability of a typo in a small portion is proportional to the length of or the area of that space of that page and similarly the probability of more than one typo in a small area is negligible under that assumption the Poisson process model can be applied here we can calculate this probabilities.

So, we have discussed important discrete probability distributions today. In fact, there are many more, but that one can differ to… For example, hyper geometric distribution there is a discrete uniform distribution and so on and then that is a broad class of distributions called power series distributions. So, one can look at those distributions in the following lecture I will be discussing special continuous distributions.