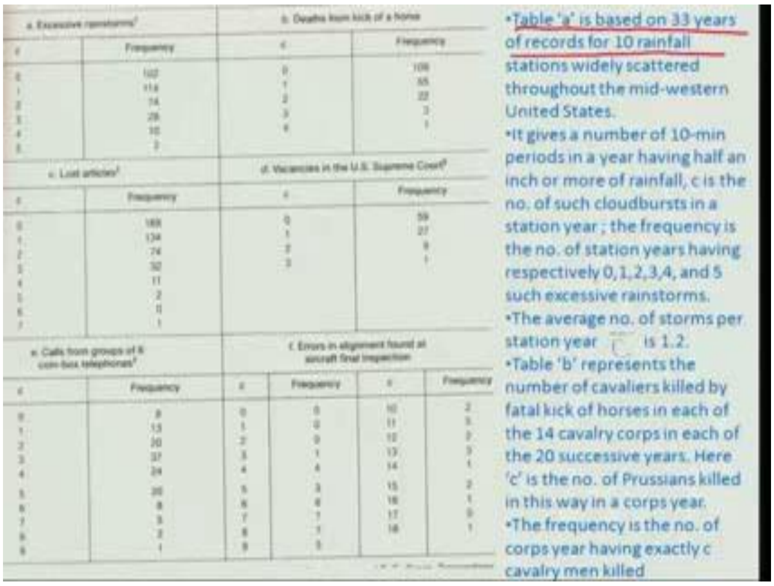
Manufacturing System Technology - II
Prof. Shantanu Bhattacharya
Department of Mechanical Engineering or
Industrial and Production Engineering
Indian Institute of Technology, Kanpur

Lecture – 25 & 26

(Refer Slide Time: 00:37)



Hello and welcome to this Manufacturing Systems Technology part two - modules 25 and 26. We were discussing in the last module about binomial distribution the way we calculate mean and standard deviation and then we also try to find out that how that can be related to the fraction defective or the fraction acceptable in acceptance sampling plan. We are going to start today looking at some of the other kind of data bases which are available and then trying to see with the binomial or some other distribution would work in that particular case. This right here shows three different tables actually six different tables where you have different situations recorded in terms of some value and a frequency.

Let us say table a, so table a is based on thirty three year of records for ten rainfall stations. So, and this is widely scatted throughout the mid-western United States. And it gives a minimum number of ten minute periods in a year having half an inch or more of rainfall, c is the number of such cloudbursts in station year; the frequency years having the respectively the 0, 1, 2, 3, 4 and 5 such excessive rainstorms. So, here for example, this excessive rainstorm has been categorized to be a let say 10 minute period in a particular year having half an inch or more of rainfall. So, these are basically c is basically represented of how many such periods are available. So, you have zero such periods, the frequencies about hundred and two. You have one such period of ten minutes or more having half an inch or more of rainfall. So, this is actually in terms of the number of the stations which are present throughout the united state the mid-western United States where this data has been taken times of the years. So, you have station years.

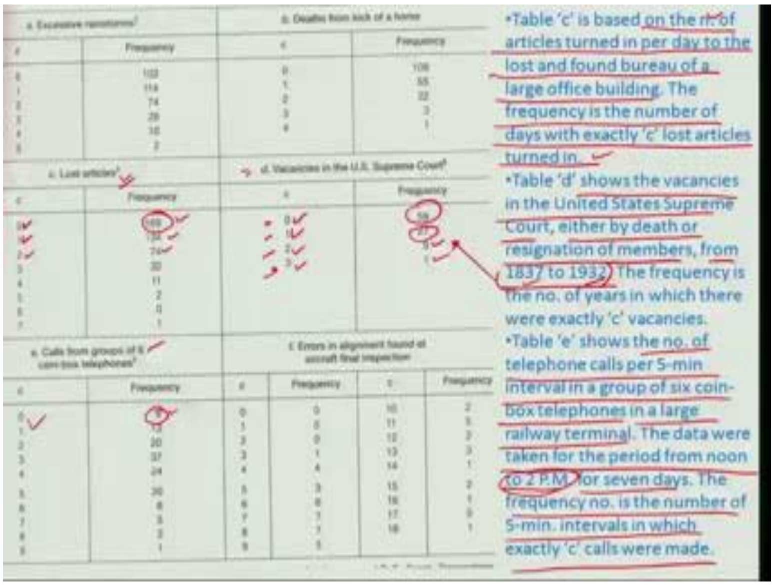
So, in all these station years which are being in question; these 102 station years have been recorded where there are zero such cloudbursts incidents; 114 station years have been recorded where there are one such cloudbursts incident; 74 station years have been recorded where there are two such you know indications or two such cloudbursts incidents and so on and so forth. And as the number goes higher, you see five you know such incidents have been recorded in at least two station years. So, station years means the number of stations times the number of years. So, if it is one year periods, so it is typically the number of stations and so on and so forth. So, the average number of storms per station year c dash is actually 1.2 which can be computed here by looking at you know the c times of the frequencies or the sigma c f divided by sigma f and that means you know on an average there about 1.2 such cloudbursts incidents of 10 minutes durations at least having half an inch or more of rainfall which is recorded in almost all the stations in one particular year period the mid-western United States that is what this data represents.

Similar kind of data is represented here in table b which shows deaths from kick of horse and you can see here that the table b here represents the number of cavaliers killed by fatal kick of horses in each of the 14 cavalry corps in each of the twenty successive years. So, there is time period which is there. There are 14 such cavalry corps which are used for recording such data and then b represents the sort of number of you know cavaliers killed by the horse kicks you know which could be fatal, and this twenty success of years of data. Just has in this particular case it was 33 years data. So, 33 year times the number of stations would be really the total number of subject points which would be used in the case of excessive rain storms.

So, c here represents the number of Prussian which are killed in this way in corps year. So, basically if you have 14 cavalries and you know you are monitoring all those 14 cavalries across 20 years period, if how many numbers of corps actually get killed by if the horse kicks is recorded as c, and that shows that you know zero such incidence of killing because of fatal kicks of horse as occur 109 times. This time say indicate the cavalry years. So, therefore, across all the 14 cavalries, there about 109 such occurrences in 20 successive years where there has been no death. There has been at least one death and 65 such you know 65 times among the 14 cavalries in the 20 successive years. Similarly two deaths in you know at least repeated twenty two times in the fourteen cavalries for twenty successive years and so on and so forth and that is how you have recorded this particular data.

So, what we are trying to see here is that the recording of this whole data is by taking into consideration at least two parameters which are about certain ranges like in the first case it was number of station times the number of years. And it is about thirty three years across which all these number of stations have been monitored. So, in fact the data size is large because of this, and you have made a bridge this statement by saying station years, so that you can keep record all the frequencies in terms of the number of such incidence which are occurring. But so in the case of you know death from kick of horse, you basically monitoring 14 cavalries across over 20 years, so it is large data said, but you are abridging that to small table here, and recording the number of deaths because of such cavalry years. So, this is in generally is way of representing or abridging larger data base by a smaller table.

(Refer Slide Time: 06:14)

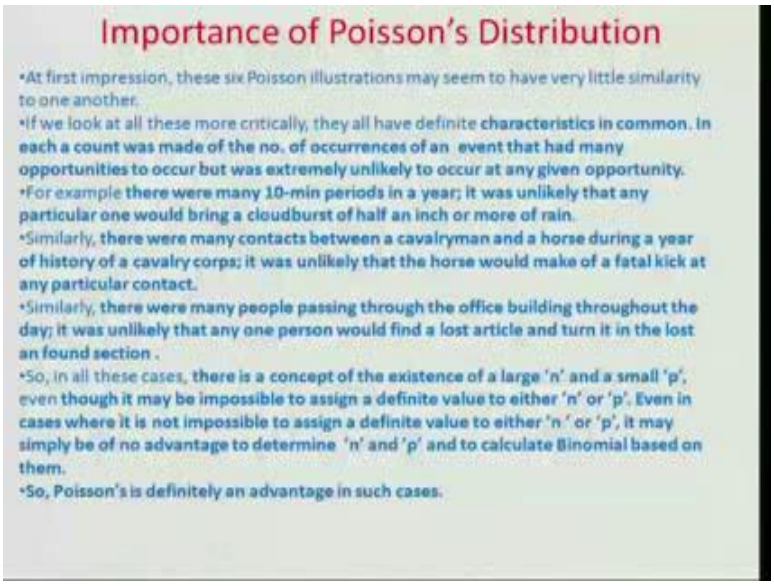


Similarly, you can have the table c of lost articles which is based on the number of articles turned in per in day to lost and found bureau of large office building. The frequency is number of days with exactly c lost articles turned in. So, what it means is that you know if supposing there are several numbers of articles turned per day to lost and found burro the for certain period of time may be let say it is months time or it is years' time, this has been recorded as exactly zero articles are return exactly hundred and sixty nine time across the whole you know number of such returns per day times the or may be the total numbers which are returned back. So, at least zero last articles are turned in hundred and sixty nine times across the time period in which you are recording times the number of articles turned per day. Similarly one for 134 times, two for 174 times and so on and so forth, again large data being recorded in an abridged manner.

Table d, for example, shows the vacancies of the United States Supreme Court, either by death or resignation of members from 1837 to 1932. So, it records about close to 79 years or 95 years of subsequently you know so there has zeroes such cases of vacancies coming in this United State, Supreme Court and or either by the death or resignation of the members at least 59 times among this 1837 to 1932 - 95 years period. Then there at least one such case here about 27 times across this whole number of years 97 years then about two such times in per year let say at least nine such cases and then three per year; obviously, for about one case or so, where this is exactly the number of vacancies which are created on yearly basis given this whole ninety seven period. Again very large data constricted into small tabular form.

Table e similarly shows the number of telephone calls per five minutes intervals in group of six coin box telephones in large railway terminal. So, the data were taken from for the period from noon to 2 pm for seven days. The frequency number is the number of five minutes interval in which exactly c calls were made. So, exactly zero cause were made. So, exactly zero cause were made eight times from the six coin box telephones for exactly five minutes you know per call kind of an interval or so for telephone calls made per five minutes interval and such eight such calls were there the frequency were exactly zero of the occurrence of telephone calls across this whole duration from 2 pm onwards in all the six telephone boxes. Similarly the frequency number, so there at least thirteen 13 numbers of at least one calls made from all the six boxes in the five minute interval at least twenty numbers of two calls made from the six boxes on the per five minutes interval for certain day from noon to 2 pm and so on and so forth, so again a very large congregation of data being recorded here as frequency table.

(Refer Slide Time: 09:45)



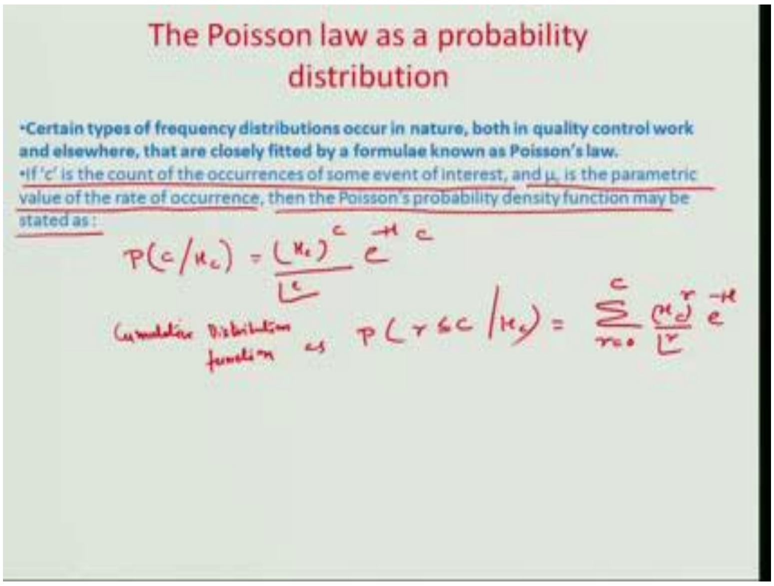
So, the idea here is that in such cases it is very peculiar two or three characteristics of recording this data and mostly all these cases are represented by certain distribution just as the binomial distribution was earlier well discussed in the case of acceptance sampling by something called a Poisson's distribution. So, the six Poisson illustrations may seem to have very little similarity to one another, but I will just record in this few statements what are the kind of similarities which will be there. So, let us look at the more critically all the five illustrations for example, the rainfall problem, the you know the fatal kicks problem, the number of telephone calls problem, the number of seats or in the Supreme Court and number of vacancies in the United States, Supreme Courts that problem or the number of lost articles returned back to building problem. So, they all have the definite characteristics in common in each a count was made of the number of occurrences of an event that had many opportunities to occur, but was extremely unlikely to occur at any given opportunity that is why the very, very large database which is in question.

So, at least something is common in the data structure that in each a count was made of the number of occurrences again of an event that had many opportunities to occur, but was extremely unlikely to occur at any given opportunity. For example, there were ten minutes period in a year it was unlikely that any particular one would bring a cloudburst of half an inch or more of rain. Similarly, there were many contacts between a cavalryman and a horse during a year of history of a cavalry corps; it was unlikely that the horse would make of a fatal kick at any particular contact. Similarly, there were many people passing through the office building throughout the day; it was highly unlikely that any one person would find a lost article and turn it in to the lost and found section. So, all of them are very, very unlikely you know to occur and their probability of occurrence is quite low. So, Poisson's distribution is generally applied to such a kind of data set where the probability of occurrence of an event is quite low, and you need really an extensively large data base to somehow make the probability quite realistic.

So, in these all cases there is a concept of existence of large n; that means, the overall sample size that you are examining or the data structure you are examining is a very, very high value and a very, very small probability of occurrence in that. So, the p is very small even though it may be impossible to assign a definite value to either n or p. Even in cases where it is not impossible to assign a definite value to either n or p, it may simply be of no advantage to determine n and p and to calculate the binomial based on them. So, this is a case where you are n and p are so buzzer that the n is on the higher side and p is on the extreme low side that you do not have a really an upper limit of the largeness of the data size that is needed, so that the probability even if it is very, very small can come into some realistic value and it can be at least recorded as occurrences of the event; otherwise the events are extremely unlikely in nature.

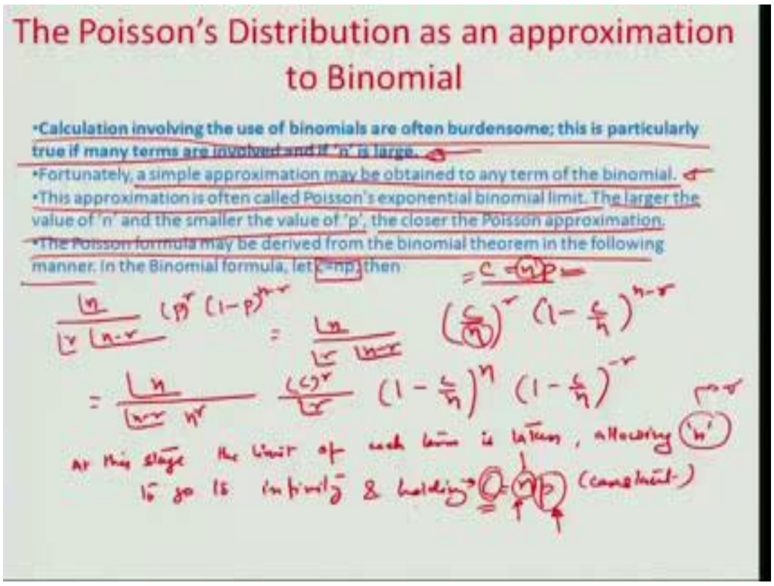
So, in such cases Poisson's distribution is definitely an advantage over and above the binomial cases. Binomial cases straight forward cases where the acceptance percentage or the probability of a sample to be good or bad is quite reasonably high number 10 percent, in fact, what we took in the binomial distribution is a very, very high number. Here this data structure particularly in if the acceptance rate is higher, can be a little bit lower also, but still you can have good accuracy to predict etcetera. But in cases when this p is extremely low, and the event is highly unlikely to occur, obviously, the data structure would increase.

(Refer Slide Time: 13:48)



So, let us now look at the characteristic property of what is the Poisson's distribution. So, let say if c is the count of occurrences of some event of interest, and mu c is the parametric value of the rate of occurrence, then the Poisson's probability density function may be stated as simply probability of count of occurrences c with the probability of let say mean probability of mu c occurrence can be represented by mu c to the power of c by factorial c e to the power of minus mu c. And as such the cumulative distribution function in such a case as the probability of r to be less then equal to c given a certain occurrence rate of this particular you know count of occurrences can be represented as sigma r varying between zero to c mu c to the power of r divided by factorial r e to the power of minus mu c. So that is how you mention the formulation - the Poisson's formulation for calculation sake. I am going derive this out from the condition that had been just illustrated earlier that what is a difference between binomial in this particular case.

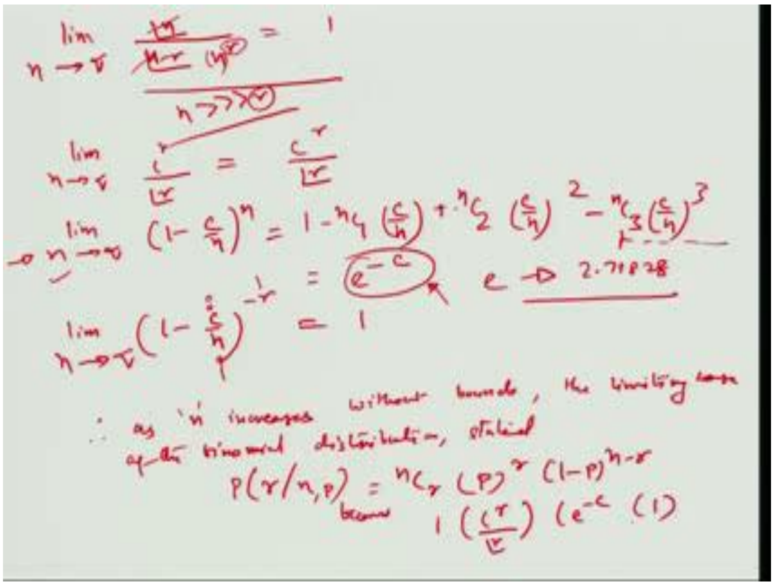
(Refer Slide Time: 15:22)



So, calculations involving the use of binomials are often burdensome; this is particularly true if many terms are involved. And if n is particularly large as in the case of the distributions that you have been seeing where the occurrences are highly unlikely. Fortunately a simple approximation may be obtained to any term of the binomial and this approximation is also called Poisson's exponential binomial limit the larger the value of n the smaller the value of p the closer the Poisson's approximation. The Poisson formula may be derived from the binomial theorem in the following manner. So, let say if I look at the chances of occurrences of an event for a certain let say number of chances of occurrence r the probability that occurrence can be provided by the binomial term n factorial by r factorial n minus r factorial times of the probability of n of the event to the power r plus probability of the even not occurring to the power of n minus r.

So, if I wanted to simply substitute you know the c p equal to n c is equal to n p in this particular expression; obviously, the p is very, very smaller; n is very, very large as the cases that you have discussed before. So, this expression can be again written as factorial n by factorial r times of factorial n minus r and value of p which is c by n to the power r and 1 minus c by n to the power of n minus r. So, I can again do some algebraic manipulation here by recording this as del the factorial n by factorial n minus r factorial r and what I can probably you know do is that will just swipe over. So, this can come from the next term. So, this is n to the power of r and the c to the power of r can be divided by factorial r. So, basically I am just taking this term and putting it here in the interest of just simplifying the expression. And then we have obviously 1 minus c by n to the power of n and times of one minus c by n to the power of minus r. So, at this stage, the limit of each term is taken allowing n to go to infinity, because obviously, the n in the Poisson distribution as you know is very, very large, and holding the c equal to n p as constants. So obviously, you can find out that the chances of occurrence are really dependent on these two values, and the probability is very, very low highly unlikely and that is why n goes to infinity. So, to you can have the number of occurrences of a particular kind.

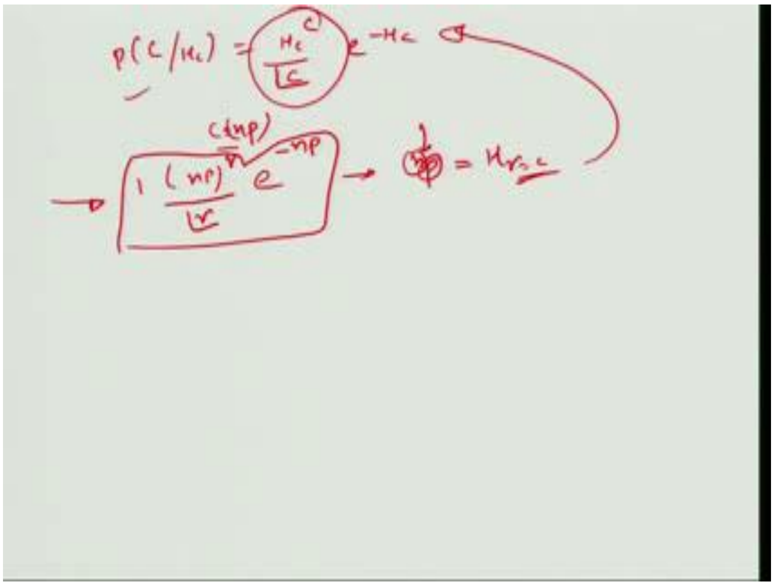
(Refer Slide Time: 18:46)



So, let us just put this limit here and see what happens to this expression. So, the limit of n going to infinity sig factorial n by factorial n minus r times of n to the power of r; if n is very, very high is actually equal to 1, for obvious reasons because n being very, very high is also very, very higher in comparison to r. And so you are basically left with n factorial over n factorial n, because the r being much smaller even the n to the power r will not have significant contribution in this particular case actually. So, the other issue is the limit of n going to infinity c to the power of r by factorial r. So, this value is independent of n. So, it would as such remains c to the power r by factorial r. The other term if n is limited to infinity is one minus c by n to the power of n which can be recorded as 1 minus n c 1 c by n plus n c 2 c by n square minus n c 3 c by n cube plus and so on and so forth. And the limit of n tends to infinity of this particular expression where n very large actually, so in fact, this turns out to be e to the power of minus c where value of e is 2.71828.

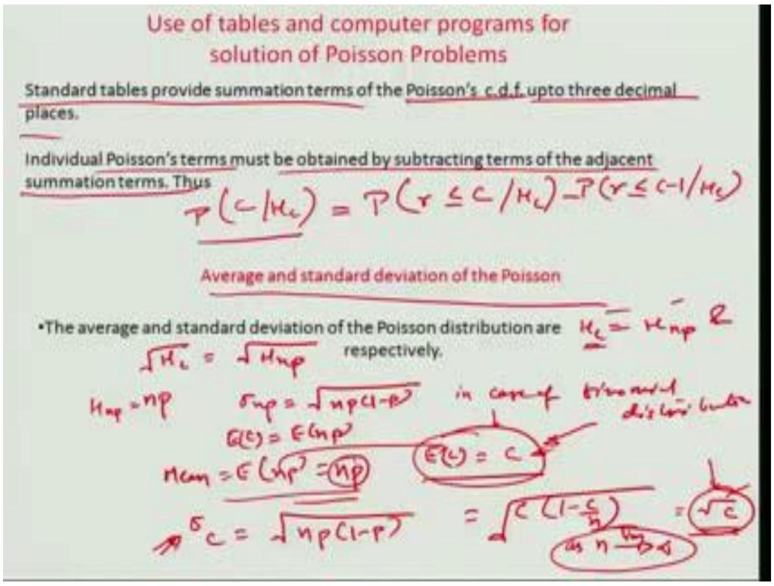
Similarly, for limit n turning to infinite one minus c by n to the power of minus r is actually recorded as one, because n being infinity this part would obviously be zero. And therefore, the r value being small this can be recorded as only one. So, I am not really going to delve into the process of proving how this limit comes out because it is actually understood to have been done at a different course module earlier covered earlier. So but if there are any questions regarding this n if you have any single query could send and then if need be this prove can be sent online later on. So, the basic idea is that with all this limits on board as n increases without any bound, the limiting case of the binomial distribution stated as the probability of r occurrences given a large sample size n, and the probability of one occurrences p small is actually given by n c r p to the power of r 1 minus p to the power of n minus r. And this becomes actually equal to in this particular case, the limit at n equal to infinity of n by n minus r factorial n to the power of r, which is one times of c to the power of r by r factorial times of e to the power of minus c times of one, so that is what is the limiting case of the Poisson's distribution which says that, p of exactly c occurrences given a certain rate at which the occurrence is happening is given as mu c to the power of c factorial c to the power of minus mu c.

(Refer Slide Time: 22:33)



So, this is nothing, but initially saying that the c can be recorded as n p in the expression given earlier here. And certainly this can be changed to one times of n p to the power of r by factorial r e to the power of minus n p. So, n p s you know in the binomial distribution if the n where significantly large and p is significantly small is basically the mu of the distribution mu of r and r being n p is basically again you know this occurrence level c particularly in this particular case which is actually giving out this Poisson distribution. So, the expected numbers of occurrence obviously are probably determined by mu c to the power of c factorial c to the power of minus mu c.

(Refer Slide Time: 23:58)



So, the standard tables provide summation terms of Poisson's cumulative density function up to three decimal places; I will share these tables when we do some numerical examples later on. Individual Poisson's terms must be obtained by subtracting terms of the adjacent summation series. So, thus any particular let say probability function of c occurrences given the mean occurrence level is mu c is actually given by the cumulative density function of any r less than c given mu c minus probability density function of any r less than or equal to c minus one given mu c term. The average and standard of the Poisson's can also be calculated, so the average in the Poisson's distribution are basically average and the standard deviation of the Poisson distribution are as you trace saw earlier mu c is actually mu of n p.

And in fact the root of mu c can also be represented as root of mu n p, the same reason mu np from the binomial distribution was represented as n p and sigma n p from the binomial distribution was written represented as n p times of 1 minus p add may detail proves in the earlier binomial distribution case, in case of binomial distribution. In this particular case, it is a limiting case where E c is equal to n p E of n p, so obviously, the mean in the Poisson's distribution then would be the mathematical expectation of n p which is the mathematical expectation in the in the binomial terms is again the value n p. So, if the occurrences very, very large here c is n p. So, it basically becomes c, so that is the mean of the binomial distribution.

And the sigma of the binomial distribution in this particular case being n p times 1 minus p can be represented as c times 1 by c one minus c by n whole under the root. Or in other words, you can represent this as root of c as the size of the n is very, very large. So, the mean of the binomial distribution is the number of occurrence c on the standard deviation Poisson's distribution and the standard deviation of occurrence of standard you know the particular distribution Poisson's distribution for a very, very large sample size and limited to infinity which is actually the case when you are applying the Poisson's distribution is root of c, so that is about it.

So, seeing the binomial, you have also seen cases where the occurrences are very high and highly unlikely and there is a huge database, so that represented by the Poisson's distribution. This particular module will close here and the interest of time, but in the next module, we would like to see again another case of another distribution called the normal distribution which can be later on very easily applied particularly to some of the acceptance sampling problems, where the distributions slightly changes then what you saw in the acceptance sampling case earlier where binomial rules are applied.

Thank you as of now.