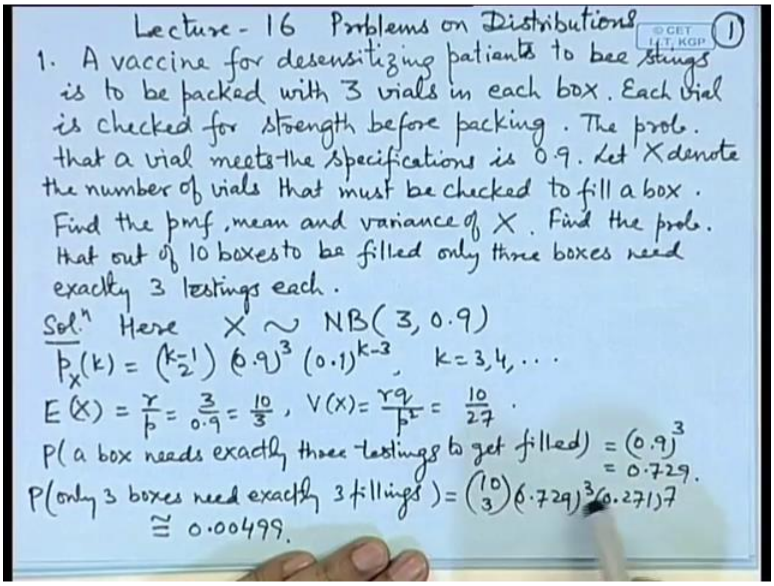
Probability and Statistics

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

Lecture – 31
Problems on Special Distributions – I

We have discussed various special discrete and continuous distributions which arise frequently in practice.

(Refer Slide Time: 00:25)



Today we will look at various applications of these distributions and this will be explained through certain problems. Let us look at the first problem; a vaccine for desensitizing patients to bee stings is to be packed with 3 vials in each box. Each vial is checked for strength before packing. The probability that a vial meets the specifications is 0.9. Let X denote the number of vials that must be checked to fill a box. Find the probability mass function mean and variance of X. Find the probability that out of 10 boxes to be filled only 3 boxes need exactly 3 testings each.

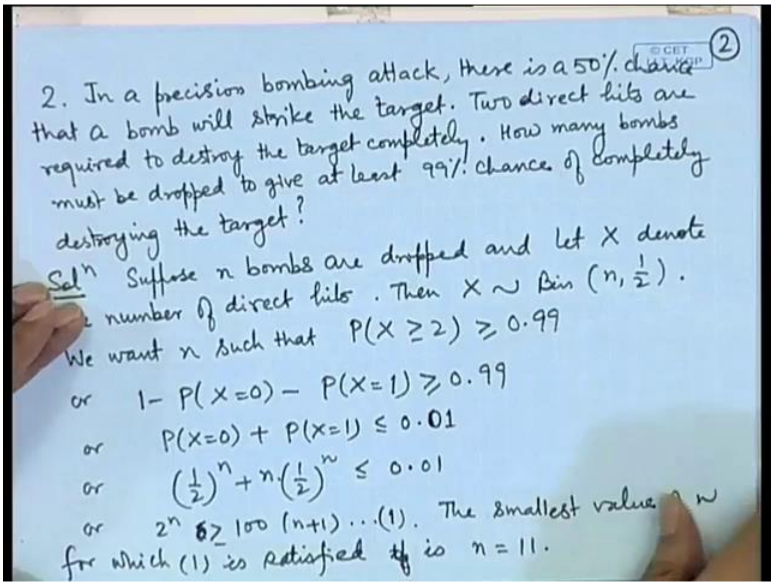
Now, let us look at the setup of this problem. So, in each box we are packing 3 vials, but the vial has to be checked. So, it may meet the specification or it may not meet that specification, we are assuming that all the vials are having identical probability of meeting the specification and each checking is done independently. Under these assumptions the vial meeting a specification or not becomes a Bernoullian trail. So, this is a sequence of Bernoullian trials. Now we keep on checking till 3 vials meet that specification and then we pack it in a box. So, this is negative binomial sampling.

And therefore, if we consider X as the number of vials which are needed to fill a box. That means, the first time 3 vials are correctly meeting that specification then the distribution of X is negative binomial with r is equal to 3 and p is equal to 0.9; that means, the probability mass function of this will be k minus 1 c 2, 0.9 cube 0.1 to the power k minus 3, which is the probability mass function of a binomial negative binomial distribution with parameter r is equal to 3 and p is equal to 0.9, the values of k are 3 4 and so on. As we know the mean of a negative binomial distribution is r by p. So, that is 3 by 0.9 that is 10 by 3, and variance is r q by p square which after a simplification becomes 10 by 27.

Find the probability that out of 10 boxes to be filled only 3 boxes need exactly 3 testings. So now what is the probability that 1 box needs exactly 3 testings; that means, the first 3 vials which are checked all of them need the specification, so this is corresponding to k equal to 3 term here which is giving 0.9 cube that is p cube. Now, the second part of this problem is that each box may need exactly 3 testing or it may not need exactly 3 testing. So, if total numbers of boxes are 10, a particular box may need 3 testing or may not need 3 testing. So, it again becomes like a Bernoullian trial with probability of success p as equal to 0.9 cube that is 0.729.

So, out of 10 boxes 3 boxes will need 3 testings, it is the binomial probability of X is equal to 3, where n is equal to 10 and p is equal to 0.729. So, by applying the formula n c x p to the power x q to the power n minus x, we get 10 c 3 0.729 cube 0.27, 1 to the power 7 which is approximately 0.00499. You can see here we have to make the assumption of independence and identical nature of the trials, so that we can apply the concept of binomial or negative binomial distribution here.

(Refer Slide Time: 05:05)



Let us look at another application: in a precision bombing attack there is a 50 percent chance that a bomb will strike the target. Two direct hits are required to destroy the target completely. How many bombs must be dropped to give at least 99 percent chance of completely destroying the target?

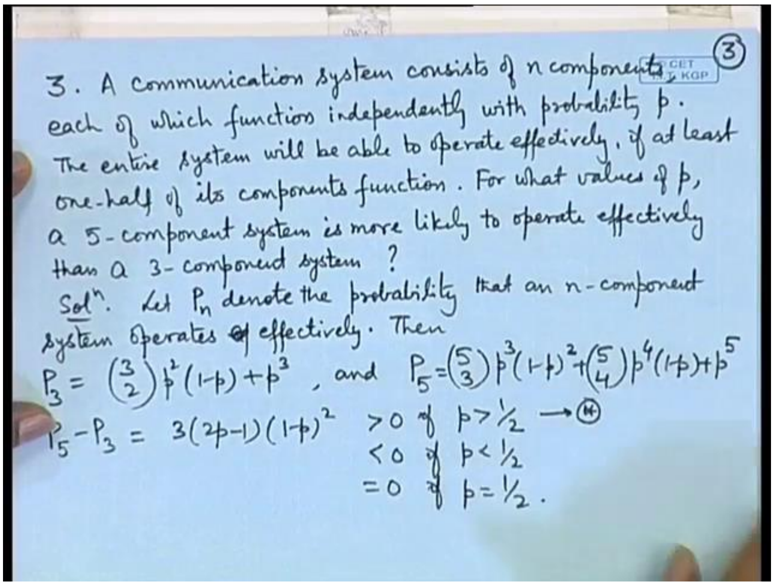
So now, here you can see a particular bomb may hit the target or it may not hit the target. So, each trial can be considered as; that means, hitting of the throwing of a bomb or attacked by a bomb that be can be considered as a Bernoullian trail. So, it may strike the target or it may not a strike the target, we assume that the attack by the bombs is independently done and it is identical in nature that is the probability of striking a same for all. So, you will have; if I consider out of n bombs, X bombs are the direct hits then the distribution of X will be binomial n, and here the probability of a striking is half because there is a 50 percent chance so it becomes binomial n half. So, we want that there is at least 99 percent chance of destroying the target.

Since we need at least two hitting that means, what is the probability that X is greater than or equal to 2. So, we want this probability to be greater than or equal to 0.99; that means, what should be the value of n for which this condition is satisfied. So, we apply the formula of the binomial probability mass function here, the probability of X greater than or equal to 2 is having many terms, so we consider the complementation of this event that is probability X is equal to 0 and probability X equal to 1. So, 1 minus this must be greater than or equal to 0.99.

So, after simplification it becomes probability X equal to 0, plus probability X equal to 1 is less than or equal to 0.01 and now here p is equal to half; so this is n c 0, P to the power 0, 1 minus P to the power n, which is half to the power n and this is n c 1, P into 1 minus P to the power n minus 1; since P and 1 minus P both are half so it is again half to the power n, n c 1 is n. So, the term is n plus 1 divided by 2 to the power of n. So, after simplification this condition is equivalent to that 2 to the power n greater than or equal to 100 into n plus 1, what is the smallest value of n for which this is satisfy?

So, we can check it and it turns out that n is equal to 11 first time satisfies this condition. So, at least we have to drop 11 bombs so that the target is completely destroyed with probability greater than or equal to 0.99. So, here if you see once again we have made the assumption that the trails; that means, the dropping of the bombs are a striking of the target etcetera is considered as Bernoullian trial, that is independence and identical nature of the trails has been considered here.

(Refer Slide Time: 08:37)

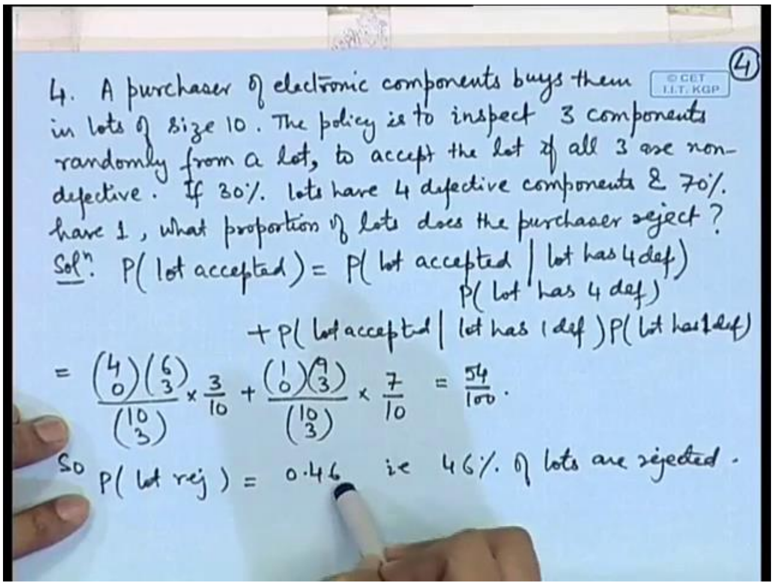


A communication system consists of n components each of which function independently with probability p. So, once again functioning of components is a Bernoullian trial because the component may fail or it may not fail. So, if it is working the working probability is p; the entire system will be able to operate effectively, if at least 1 half of its component function; that means, suppose there are 10 components, then at least 5 should function then the system will be operating effectively. For what values of p, a 5 component system is more likely to operate effectively than a 3 component system?

So, we have to calculate the probability of a 5 component system working effectively and a 3 component system operating effectively. So, let us use a notation P n let it be the probability that an n-component system operates effectively. So, P 3 will denote the probability that 3 component systems is working effectively; that means, at least 2 or 3; that means, either 2 or 3 of the components are working correctly. So, out of 3 2 are working. So, 3 c 2, p square, 1 minus p and all the 3 are working so p cube. So, the probability of P 3 is given by this. In a similar way a 5 component system will be operating effectively if either 3 or 4 or all 5 components are working properly. So, the probabilities of them are given by these, so we add up so P 5 is given by this. We have to check that whether a 5 component system is more effective or 3 component. So, we consider the difference P 5 minus P 3, which after certain simplification is equal to 3 into 2 p minus 1 into 1 minus p square.

Now, here you observe 1 minus p square is a positive term, because p is a number between 0 and 1. So, this term is positive if p is greater than half; that means, a 5 component system is more effective if the probability of each system working effectively is more than 50 percent so this answers the question here; we can also see that this is less than 0 if p is less than half. That means, if each component fails with a probability which is less than 50 percent, then if we add more components in the system it is actually reducing the probability of operating effectively and of course, if p is equal to half then both the systems have the same probability of operating effectively. So, these are some of the applications of the binomial and negative binomial distribution etcetera.

(Refer Slide Time: 11:50)



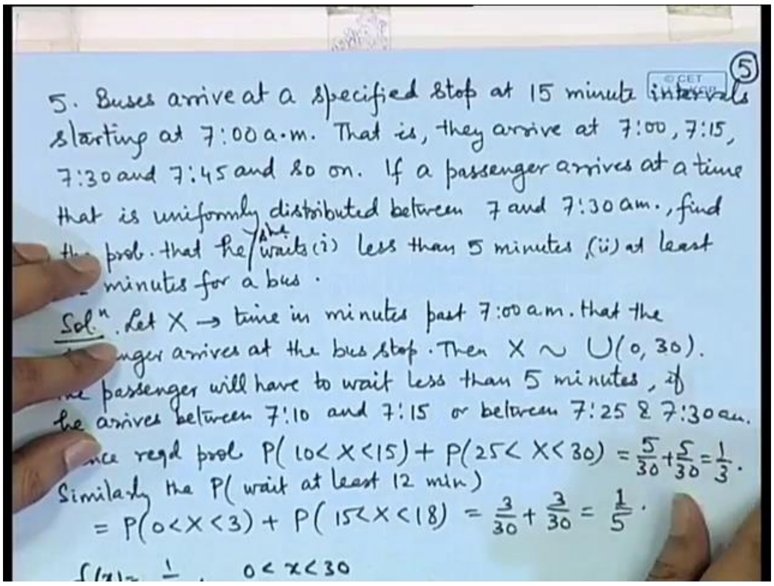
Let us look at an application of hyper geometric distribution; a purchaser of electronic components buys them in lots of size 10. So, the policy is to inspect 3 components randomly from a lot, so a lot each lot is having 10 components. So, what the purchaser will do? He randomly selects 3 components or 3 parts from that lot and he inspects them. If all the 3 are non defective he accepts the lot otherwise he rejects the lot. Now it is known to us that 30 percent of the lots have 4 defective components and 70 percent of the lots have 1 defective component.

So, in general what proportion of lots the purchaser will be rejecting? So, in place of probability of rejecting we can calculate the probability of lot getting accepted. So, the lot is accepted now this is conditional upon 2 types of possibilities, that the lot has come from the set where 4 defective components are there or where 1 defective component is there. So, we can apply the theorem of total probability, the probability that the lot is accepted given that the lot has 4 defectives into the probability that the lot has 4 defective, plus probability that lot is accepted given that lot has 1 defective multiplied by probability that lot has 1 defective term.

Now, we evaluate these probabilities. So, probability that the lot has 4 defectives is 0.3, because 30 percent of the lots have 4 defective components and similarly the probability that the lot has 1 defective that is 0.7. Now what is the probability of the lot getting accepted if the lot has 4 defective items. So, total number of items in the lot is 4 is 10, we are selecting randomly 3. So, the lot will be accepted if all the 3 checking which have been done from here are for the good ones. So, since the lot which is having 4 defective, 6 will be good, so the 3 which the purchaser has selected must be all good. So, it is 6 c 3 and out of the bad ones none of them is selected; so 4 c 0 into 6 c 3 divided by 10 c 3, which is the hyper geometric probability.

In the second case if the lot has one defective. So, out of 10; 9 are good 1 is defective and our selection all the 3 must be from good. So, it is 1 c 0 9 c 3 divided by 10 c 3. So, after simplification this turns out to be 0.54. So, probability of the lot getting rejected will be 1 minus this that is 0.46. So, 46 percent of the lots get rejected which is quite high which is understandable, because the person checks and each of the components which are checked must be all right then only will accept. So, the condition is relatively tough and therefore, almost 50 percent of the lots are actually getting rejected under these given conditions.

(Refer Slide Time: 15:38)



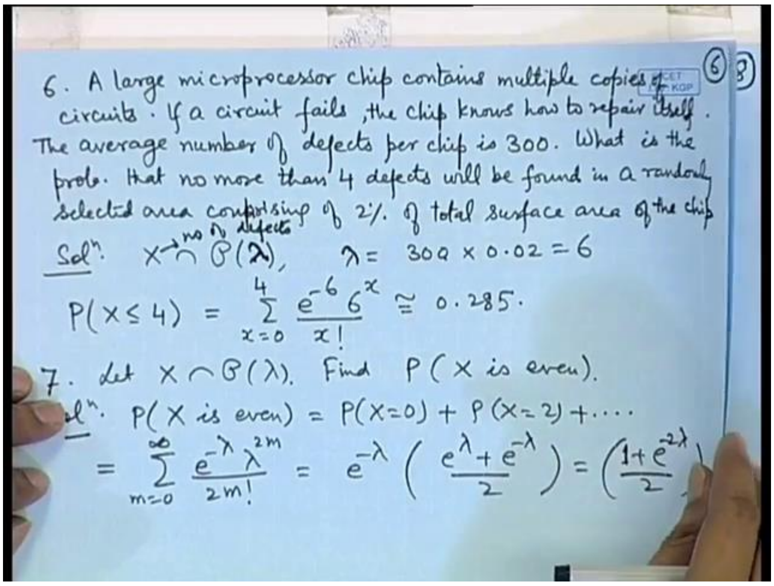
Let us look at application of continuous uniform distribution. Buses arrive at a specified stop at 15 minute intervals starting at 7 a m. That means, bus will come at 7, it will come at 7:15, 7:30, 7:45 and so on. So, if a passenger arrives at a time uniformly distributed between 7 to 7:30 a m, find the probability that he has to wait for less than 5 minutes or at least 12 minutes for a bus. That means, when he comes then he will get a bus in less than 5 minutes. So, let us consider the arrival time of the passenger. So, since the time of the passenger is between 7 to 7:30 a m, if I say X is the time in minutes past 7 a m that is the arrival time of the passenger at the bus stop, then we can say that under the given conditions X follows uniform 0, 30 where the measurement is done in the minutes; because it is between 7 to 7:30, but we are considering that starting point as 0, so between 0 to 30.

The passenger will have to wait less than 5 minutes if he arrives between 7:10 to 7:15 because the next bus is at 7:15, so if he arrives at say 7:05 then he will have to wait 10 minutes, if he comes at 7:07 then he have to wait for 8 minutes and so on. So, if he arrives between 7:10 and 7:15 he will get a bus in less than 5 minutes, similarly after this bus departs then the next bus comes at 7:30; so, if the passenger arrives between 7:25 to 7:30 then he will have to wait less than 5 minutes.

So, the required probability that the passenger has to wait for less than 5 minutes is probability that X lies between 10 and 15, or X lies between 25 to 30; since it is a uniform distribution on the interval 0 to 30 the density of x is 1 by 30, 0 less than x less than 30. So, probability of 10 less than X less than 15 will become 15 minus 10 by 30 that is 5 by 30 and similarly probability of 25 less than X less than 30 that will become 30 minus 25 by 30 that is 5 by 30. So, the total probability is 1 by 3. That means the probability that he has to wait for less than 5 minutes is one-third of the time.

In a similar way the probability that he has to wait for at least 12 minutes. So, he will have to wait for at least 12 minutes, if he arrives between 7 to 7:3 because if he arrives after 7:03 then he has to wait less than 12 minutes, because the next bus is at 7:15. Similarly if he arrives between 7:15 to 7:18 then he has to wait for more than 12 minutes. So, the required probability of waiting at least 12 minutes is that X lies between 0 to 3 or X lies between 15 to 18. So once again using the uniform density it is 3 by 30 plus 3 by 30 that is 1 by 5; that means, 20 percent of the times he has to wait for more than 12 minutes.

(Refer Slide Time: 19:24)



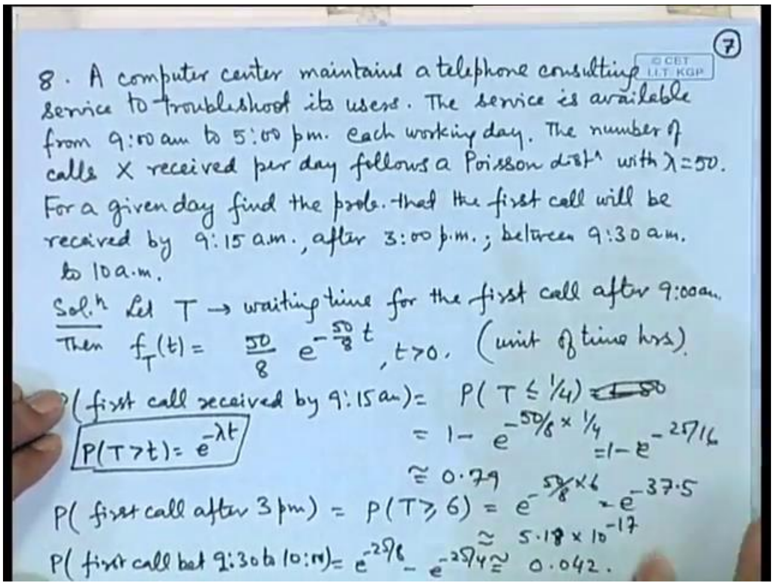
Let us look at an application of Poisson distribution. A large microprocessor chip contains multiple copies of circuits; if a circuit fails the chip knows how to repair itself the average number of defects per chip is 300. So, this is the condition that is the rate kind of thing that is one chip has nearly 300 defects, what is the probability that no more than 4 defects will be found in a randomly selected area comprising of 2 percent of the total surface area of the chip? So, here if I consider X as the number of defects in the 2 percent of the surface area, then if I assume it to be Poisson distribution of the parameter lambda, then lambda here will be because 300 is for the full area.

So, 300 into 0.02 that means, in point 2 percent of the surface area what is the number of defectives, for that the rate will be 300 into 0.02 that is 6; that is the rate of detecting the defects. Here we want what is the probability that not more than 4 defects will be found. So, it is probability of X less than or equal to 4 based on the Poisson probability mass function, so it is e to the power minus lambda, lambda to the power X by X factorial X is equal to 0 to 4. So, by putting lambda is equal to 6 this can be evaluated either by direct or by Poisson tables it is approximately 0.285.

Let us look at one more application of the Poisson distribution, suppose X is a Poisson lambda distribution what is the probability that X is an even number? As we know in the Poisson distribution X can take values 0, 1, 2, 3 and so on; that means any non negative integer value. So, the probability that X is even is probability X is equal to 0, probability

X is equal to 2 etcetera; that means, e to the power minus lambda, lambda to the power j by j factorial, where j is of the form 2 n. So, we can write it in this form m is equal to 0 to infinity. So, if we take out e to the power minus lambda, the infinite series is actually the sum of e to the power lambda plus e to the power minus lambda divided by 2. So, it is equal to 1 plus e to the power minus 2 lambda by 2.

(Refer Slide Time: 22:14)



A computer center maintains a telephone consulting service to troubleshoot its users, the service is available from 9 a m to 5 p m each working day, the number of calls X received per day follows a Poisson distribution with lambda is equal to 50, for a given day find the probability that the first call will be received by 9:15 a m, after 3 p m between 9:30 a m to 10 a m. So, let us look at the modeling of this problem, since here time is asked, so we can consider T as the waiting time for the first call after 9 a m. Now here we are considering the day as an 8 hour period and in 8 hour period the number of calls received is rate lambda is equal to 50.

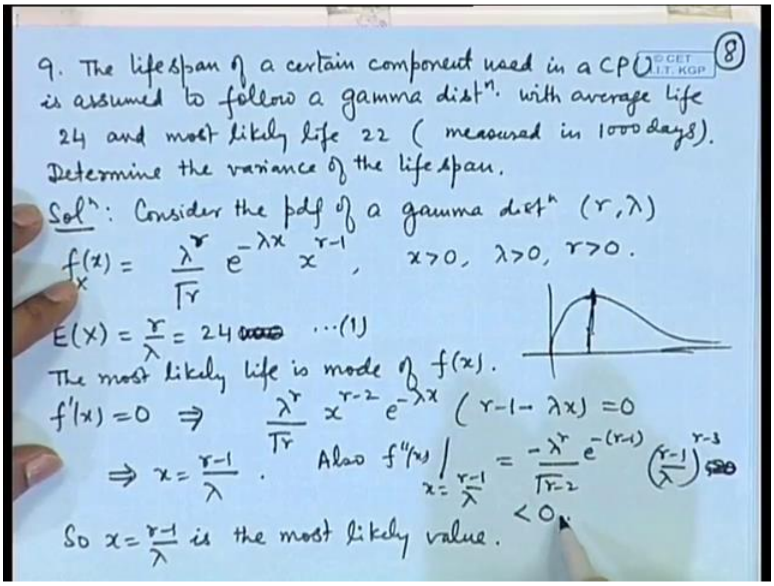
So, in 1 hour period it will be 50 by 8. So, now, if I consider T as the waiting time for the first call after 9 a m, then the distribution will be negative exponential with lambda is equal to 50 by 8, so lambda e to the power minus lambda t. What is the probability that the first call is received by 9:15? That means, in quarter of an hour; that means, what is the probability of T less than or equal to 1 by 4 this unit of measurement is in the hours.

So, now we apply the formula for the exponential distribution probability of T greater than some small t is equal to e to the power minus lambda t. So, if we use this formula probability of capital T less than or equal to 1 by 4 is 1 minus e to the power minus lambda t. So, lambda is minus 50 by 8 and time is 1 by 4. So, it is after simplification 0.79, which looks surprisingly quite high that is almost four-fifth of the probability, the reason is that in a day we are receiving roughly 50 calls.

So, within first 15 minutes a call will be received with a substantially high probability; likewise if we calculate what is the probability that the first call is after 3 p m? Now from 9 a m to 3 p m it is 6 hours, so that means what is the probability of T greater than or equal to 6. So, if we utilize this formula it is e to the power minus lambda t that is 50 by 8 into 6 which is extremely small probability, which it must be because the rate is quite high of receiving the complaints and here we are saying that from morning 9 to 3 p m there is no call, so the probability of that event must be pretty small and similarly what is the probability that the first call is between 9:30 to 10? Since the rate is high, the first call is after 9:30 the probability must be small, and then we are saying between 9:30 to 10:10 which is further small, so it is after the calculation using this formula it is turning out to be 0.042.

So, these are some of the applications of the Poisson distribution or the Poisson process, you can notice here that in order to apply the Poisson distribution we have to look at the rate in the appropriate time interval are the area that is lambda t we have to calculate .

(Refer Slide Time: 26:02)



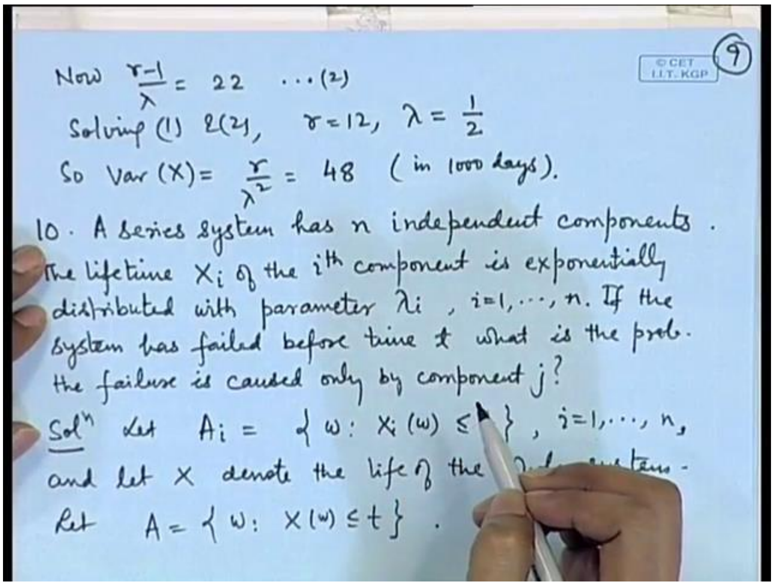
Let us look at the lifespan of a certain component used in a CPU is assumed to follow a gamma distribution with average life 24 and most likely life 22, it is measured in 1000 days so; that means, average life is 24 1000 days and most likely life is 22 1000 days find the variance of the life span? So, by standard form of a gamma distribution we have considered as a waiting time for the rth occurrence in Poisson process. So, we had the parameters r and lambda and the form of the probability density function was given by lambda to the power r by gamma r, e to the power minus lambda x, x to the power r minus 1.

So, if we take the mean of this distribution it was r by lambda that is given to be 24 the most likely life. So, by most likely life we mean the maximum value of the density function. In the discrete case it would mean that the point corresponding to the maximum probability mass function and if we take analogous value of that in the continuous case it means the mode of the distribution.

So, if we have the density function like this then the maximum value is at end at this point. Now for a gamma distribution the maximum value can be calculated. So, here the density function is given this, we can use the ordinary calculus by looking at the derivative. So, f prime x is 0 for this that is x is equal to r minus 1 by lambda and at this point we can check the second derivative it is actually negative. So, this is x is equal to r minus 1 lambda is the mode of the distribution that is the maximizing point, it is given here that r minus 1 by lambda is equal to 22. So, we have 2 equations: 1 is r by lambda is equal to 24 and another is r minus 1 by lambda is equal to 22. So, if we simplify this we get r is equal to 12 and lambda is equal to half.

So, the distribution of the lifespan of CPU of a certain component used CPU is specified as a gamma 12 and half. So, now if we want the variance of this lifespan, variance of a gamma distribution is r by lambda square, so after substitution it becomes 48 and in 1000. So, 48000 days that is (Refer Time: 29:00) the squared units.

(Refer Slide Time: 29:02)



So, we will stop here.