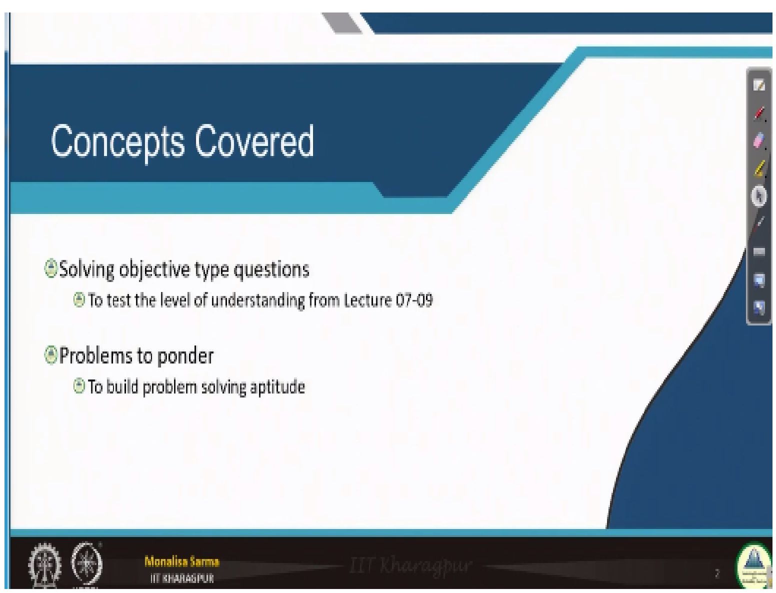
Statistical Learning for Reliability Analysis
Prof. Monalisa Sarma
Subir Chowdhury of Quality and Reliability
Indian Institute of Technology - Kharagpur

Lecture – 11
Tutorial on Discrete Probability Distributions

So, welcome once again. So, we have discussed discrete probability distribution, we have seen different type of discrete probability distribution uniform probability distribution, binomial probability distribution, multinomial than we have seen hyper geometric distribution, negative binomial geometric distribution and Poisson distribution, we have this discuss all this discrete probability distribution. Now, we will be doing a tutorial session on this distribution, the thing is here, I have problems catering to all the distribution.

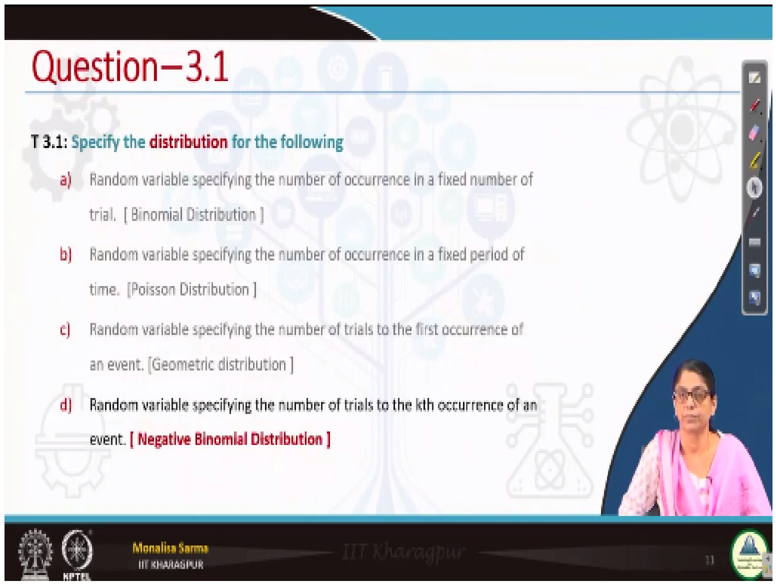
But what I want is that before immediately jumping to the solution, because I have the solution also here, immediately jumping into the solution, what a request is that once you read the question, and then from the question, you try to find out what is the random variable here and it will fall in which distribution then once you are sure that it will fall in this distribution, then you can check for the answer check for a solution, that way your learning will be proper.

(Refer Slide Time: 01:20)



So, first we will be solving few objective type questions which will help you in understanding the concepts and then we will be solving some problems. So, first to discuss the objective type questions.

(Refer Slide Time: 01:33)



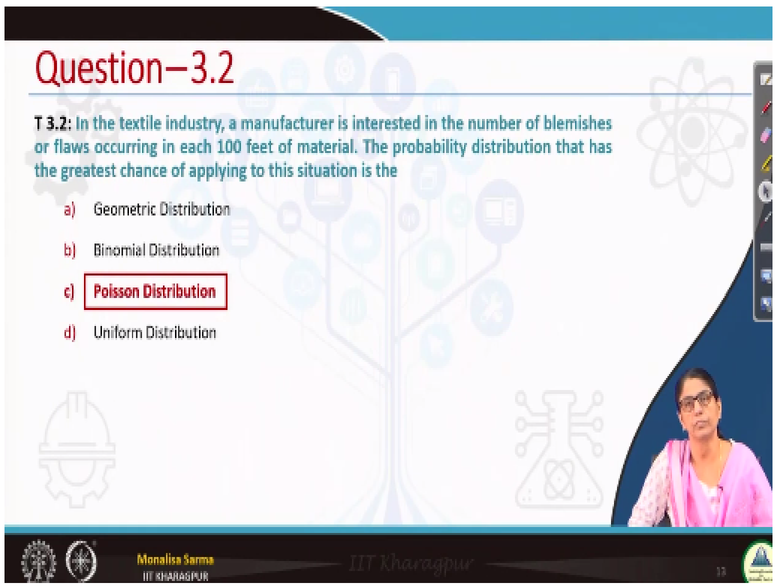
So, here I will give certain statements from this we need to find out what is the distribution for the following like first one is random variable, if my random variable specifies the number of occurrence in a fixed number of trials, random variable specifying the number of occurrences in a fixed number of trials what is that? This is binomial distribution, we have seen binomial distribution weather number of occurrences we are interested in finding out the number of successes and our trials are fixed.

Which are x, which is actually the trials are independent as well, is not it? So, this is a binomial distribution. Now, the second is random variable specifying the number of occurrences in a fixed period of time and a fixed period of time, what are the number of total number of occurrences probability of a number of 0 occurrence, probability of number 1 occurrence in a fixed period of time, or in a fixed region of space we have seen that is not it? What is that? That is your Poisson distribution.

The next random variable specifying the number of trials to the first occurrence of an event, so here the number of trials are not fixed, but what is fixed, we want to find out the random variable specifying it specifies the number of trials and we want to find out in what number of trials we will get the first occurrence that is maybe the first success or maybe the first failure, what is that that is geometric distribution.

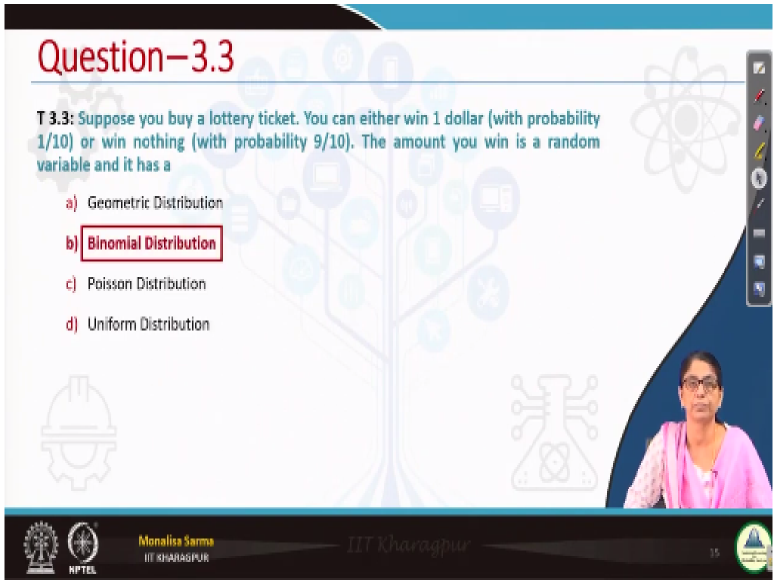
Next random variable specifying the number of trials to kth occurrence of an event. Here also my random variables is a number of trials and what I am interested in finding out the kth occurrence of an event, maybe the kth success of an event, kth failure of an event so, what is that that is negative binomial distribution.

(Refer Slide Time: 03:23)



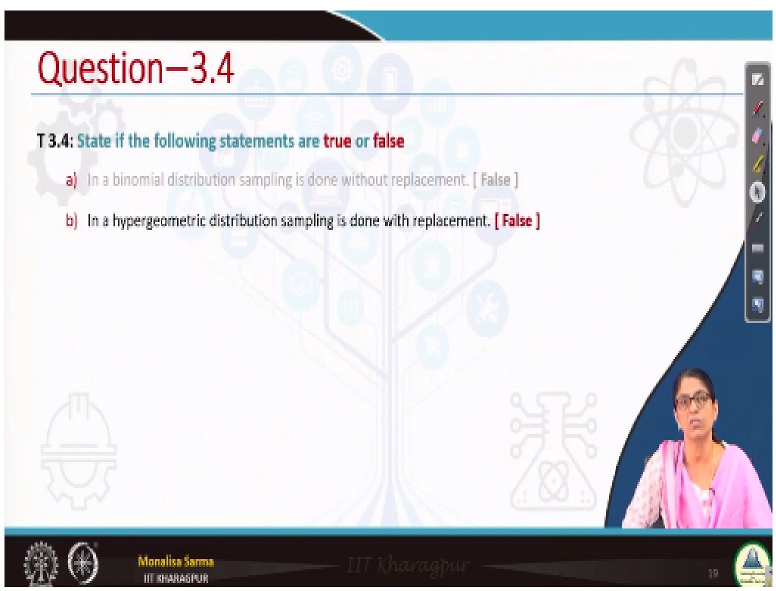
Again, another one question in a textile industry, a manufacturer is interested in the number of blemishes or flaws occurring in each 100 feet of material, the probability distribution that has the greatest chance of applying to this distribution. So, here it is number of blemishes of flaws in a particular length of material. So, what is that this is undoubtedly Poisson distribution.

(Refer Slide Time: 03:52)



Then suppose, you buy a lottery ticket, you can either win 1 dollar with probability of 1 / 10 or win nothing means you can win or you can lose that is where you are not winning is lose means not you did not win anything, you can say the success or a failure with probability 9 / 10. The amount you win is a random variable and it has a what distribution? That is a binomial distribution.

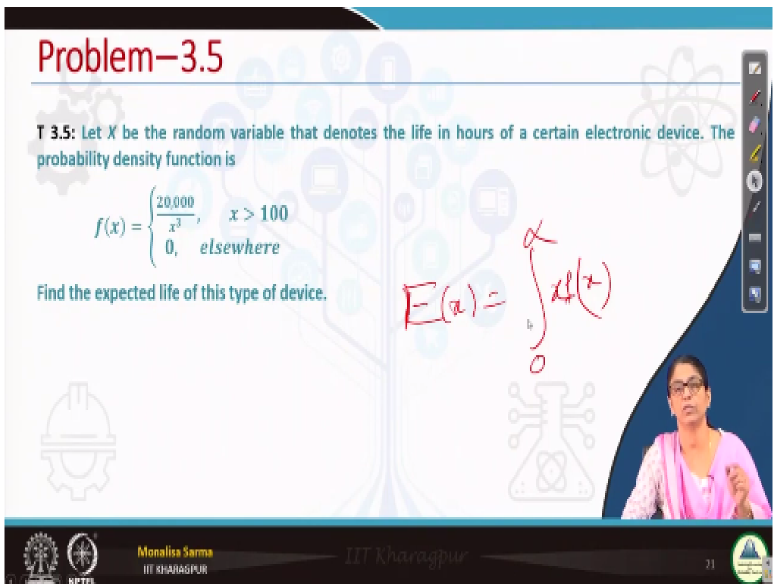
(Refer Slide Time: 04:20)



Then some statements to find out whether it is true or false. In a binomial distribution sampling is done without replacement. In a binomial distribution, do we do the sampling without replacement? No, we do the sampling with replacement in a deck of cards, we pick a card and we put it back. So that if we do without replacement, then it will be dependent it will not be trials will not be independent, so it is false. In a hypergeometric distribution sampling is done with replacement.

On the contrary, in hypergeometric distribution sampling is done without replacement. So, this is again false. But those were very simple objective type questions. Now, we will come to some problems.

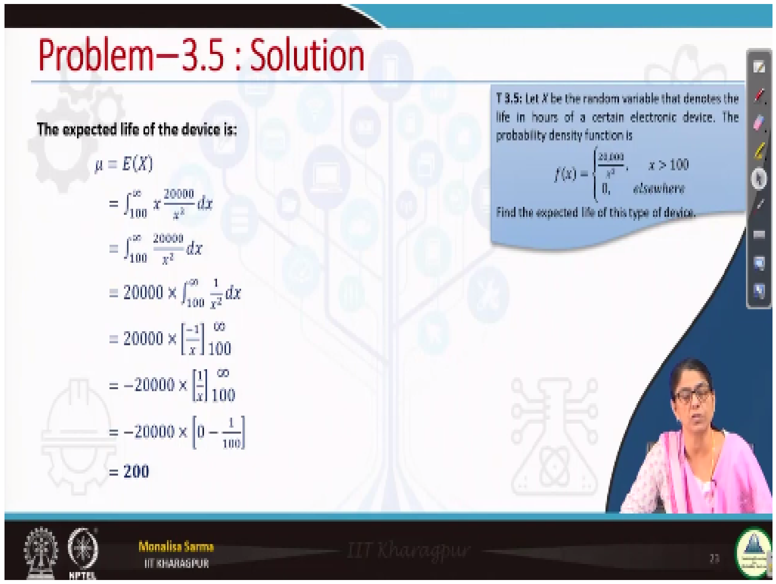
(Refer Slide Time: 05:06)



So, first question, if X is the random variable that denotes the life in hours of a certain electronic device, the probability density function, the probability density function is given here that is f(x) in a particular range where x > 100 it is given 20000 / x^3 and in other portion this means for values of x <= 100 it is 0, we have to find the expected life of this type of device. What is the formula for expected value you remember formula for expected value is E(x) = ∑ ∫ xf(x).

Since this is a continuous random variable, continuous random variable we use integration for discrete we use summation. So, e (x) is x / f(x), x f(x) so, it will be from 0 to infinity or you can say it from -∞ to +∞ as well. So, it is basically we do not take 0 it is from -∞ to +∞.

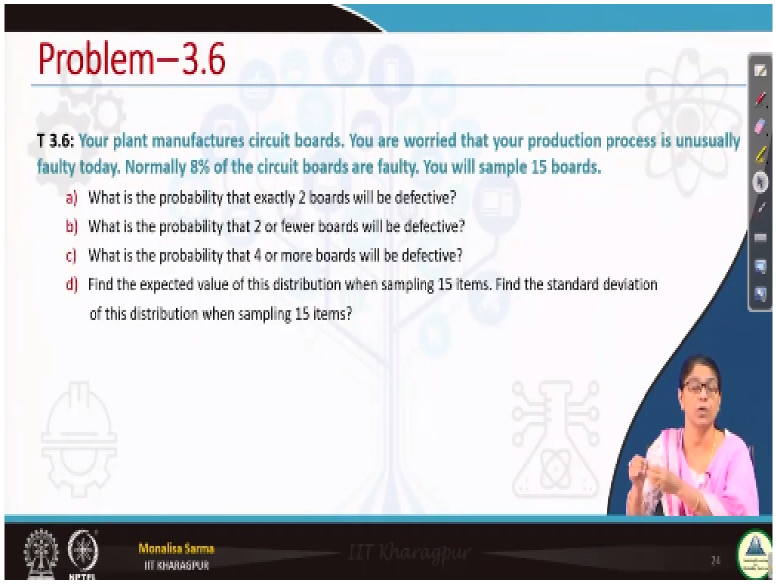
(Refer Slide Time: 06:20)



So, how do we find out E of x is here why we have specified from 100 to ∞ because an elsewhere it is given 0 so, we know that will be 0 only y instead of taking this we are taking it from 100 above x > 100 it is given so, we are taking from 100 to ∞, you can write it from 0 to ∞ also no issue but from 0 to ∞ from 0 to 100 it will be 0. So, and then it is simple question of simple integrating it.

As I already told you one of the class you will have to brush up your knowledge on integration and differentiation for different problems that we will be encountering in this course, we will not go for very complicated integration, but simple integration you will definitely need it also integration by parts also we will need it, so please brush up your knowledge on that.

(Refer Slide Time: 07:13)

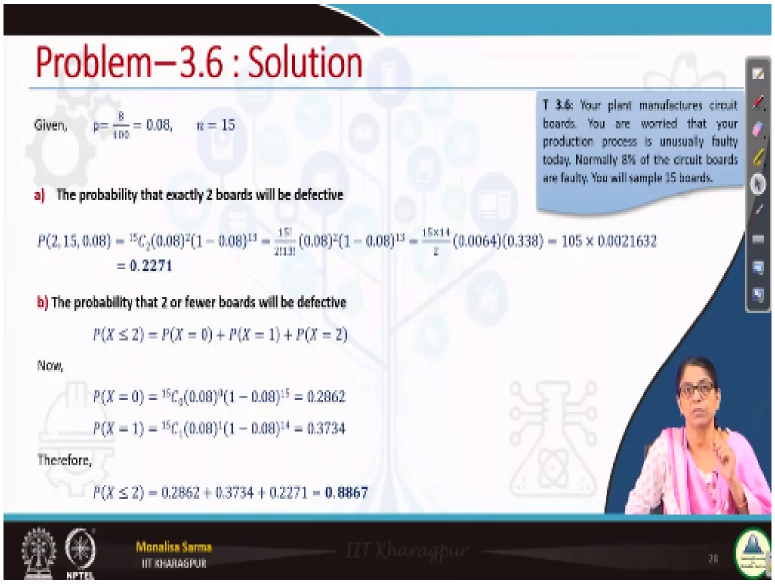


So, that is all now, next question. The earlier question is the simple question of continuous probability distribution where we are trying to find out the expected value. Now see this vision your plant manufactures circuit boards you are worried that your production process is unusually faulty today normally 8% of the circuit boards are faulty you will sample 15 boards. 8% of the circuit boards are faulty that means my probability of faulty is given, probability of fault is 8 / 100.

So, and whatever we will find out, we will sample total 15 boards now, the question is what is the probability that exactly 2 boards will be defective? Now, first one, this is a question of which probability distribution? It is definitely a question of binomial probability distribution, you are just what you are doing in a production line, you are just finding out one whether it is faulty or not, you are keeping it back taking one keeping it back. So, it is not that we are taking it and you are doing it finding it out till destruction, and then you cannot keep it back.

So, it is not that sort of sampling process, you are just picking a circuit board and you are testing it on a probably again picking another testing and putting it back. So, it is definitely a question of binomial probability distribution.

(Refer Slide Time: 08:31)

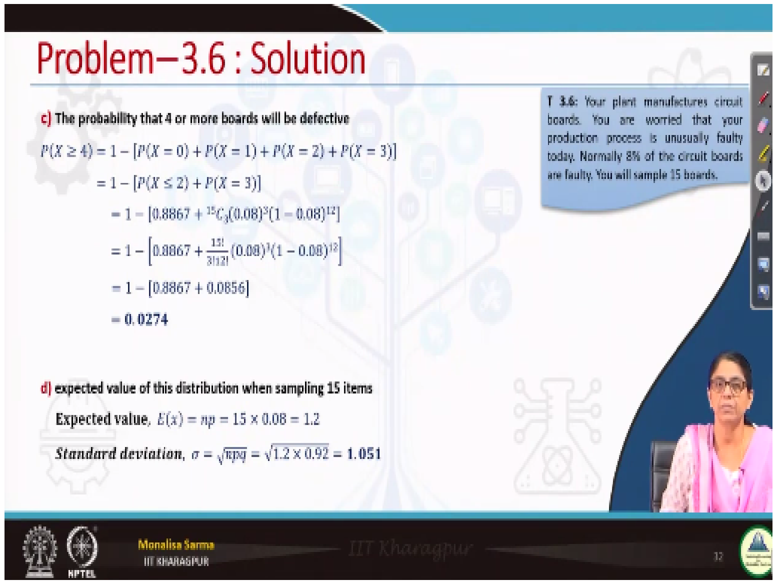


So, probability of exactly 2 boards it is what is my n? n is 15 so 15 C 2 we have this formula just put it in a formula that is all, once you know it is binomial distribution, you know the values of the parameter you know what is n, what is p, what is x then just putting it in a formula. Similarly, the next question probability that 2 or fewer boards will be defective 2 or fewer that means, we have to find 0 defective, 1 defective and 2 defective.

So, for 0 defective, 1 defective and 2 defective in fact, instead of doing it manually, you can consult the table also which I have discussed in the class that is binomial distribution table because binomial distribution table give us the cumulative value here we are interested in this question problem we are interested in a cumulative less than equals to 2 means from 0 to 2. Here I have solved it without solving the consulting table, but you can very well consult a table in the table if this particular probability value is given.

But, in a table there it has some limited number of probability values actually all this probability it can take any value. So, table it does not contain any probability values. If first you find out whether it is probability concerning probability value equals to 0.08 if it is there, then very well write it on the table.

(Refer Slide Time: 09:57)

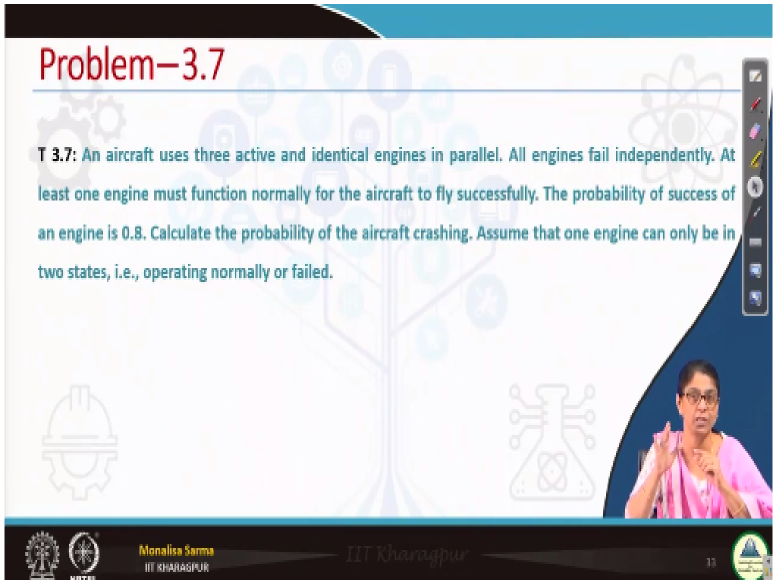


Then next question, what is the probability that 4 or more boards will be defective 4 or more that means probability that 4 will be defective total there are 15 boards, 4 or more will be defective means 4 may be defective 4 + 5 + 6 + up to it will go to 15. So, it will be well if it is in a table we can find out but if it is not at the table, then it will be very difficult to find out from 4 to 15. And the table also we do not get it we get it from 0 to 4, we will not get it the other way around. So, one way is that finding it out 1 minus of 0 to 4.

It greater equal to 4 that means 1 minus of 0 to 3 so that means 0 defective, 1 defective 2 defective and 3 defective 1 minus of that will give us the data is equal to 4. You will see it here, probability that 1 minus probability that x = 0, probability of x = 1, probability of x = 2, probability of x = 3 and probability of x = 0 and x = 1 these 2 we have found out in our last question. So, we found out this value that is where we will not solve it again for x = 2 and 3 we will find it out.

Sorry we find it out till this portion on the 3 is remaining with 3 we will find it out and we will get the value. Then expected value binomial probability distribution what is the expected value? Expected value is n x p what is n here? n is 15, p is what? p is 0.08. Standard deviation is √ np, variances npq, mean standard deviation is √ npq.

(Refer Slide Time: 11:36)



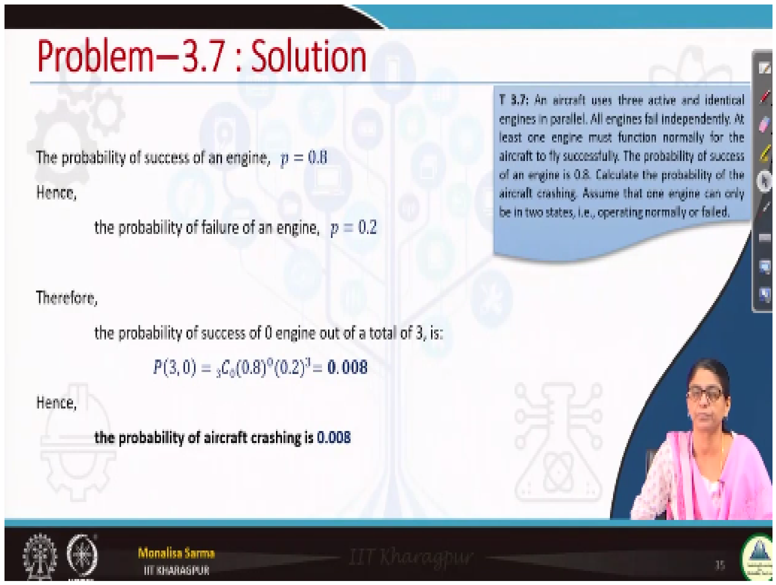
Now, one more question an aircraft uses 3 active and identical engineers parallel, all engine can fail independently it uses 3 identical engine at least 1 engineers function normally for the aircraft to fly successfully. It usually happens in a normal case in aircraft we usually have 3 different machines and at least 1 machine should perform so that aircraft will fly normally. So, at least once a function the probability of success of an engine is 0.8.

Calculate the probability of the aircraft crashing, when the aircraft will crash when there is not a single machine is successfully operating if all the machines are failing then the aircraft will crash even if 1 machine is working this aircraft will not crash. So, when we have to find out probability of the aircraft crashing that means we have to find a probability of 0 success, is not it? When there is not a single successful machine then it will crash so we will have to find a probability of 0 success.

Assume that 1 engine can only be in order of 2 states operating normally or fail. Why this question is why this point is given assume that engine can only be in 2 states operating normally or fail there are some machines where there are 3 different states that you do not need to know at this point just the 3 different states means failing as 1 state and working maybe through 2 different like some like circuit all of you may know circuit short circuit failure, open circuit failure 2 different failure modes 1 success and 2 different failure modes.

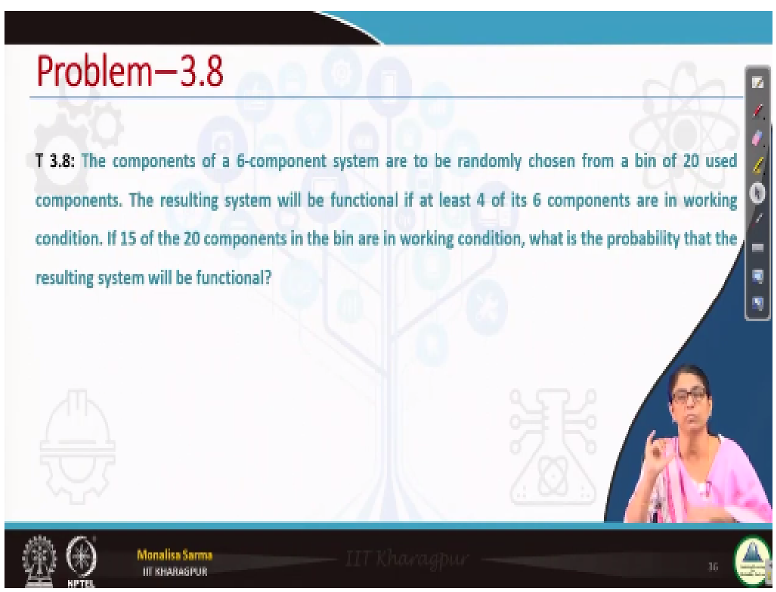
So that is why this particular statement is given. So, it is it has 2 states operating or fail but that is not very that is redundant information for description for your knowledge I have just mentioned it.

(Refer Slide Time: 13:23)



So, now how to find out a probability of success is given probability of success is 0.8 we have to find out 0 success in a total of 3, simple it is simply a case of binomial distribution very simple case of binomial distribution.

(Refer Slide Time: 13:38)



So, now distribution those components of 6, 6 component system are to be randomly chosen from a bin of 20 used component. 6 component system we have to choose randomly from a bin of 20 components the resulting system will be functional if at least 4 of its 6 components are in working condition. From this 20 we have to form a 6 components system there is a system which is consisting of 6 component system we have to pick from this 20.

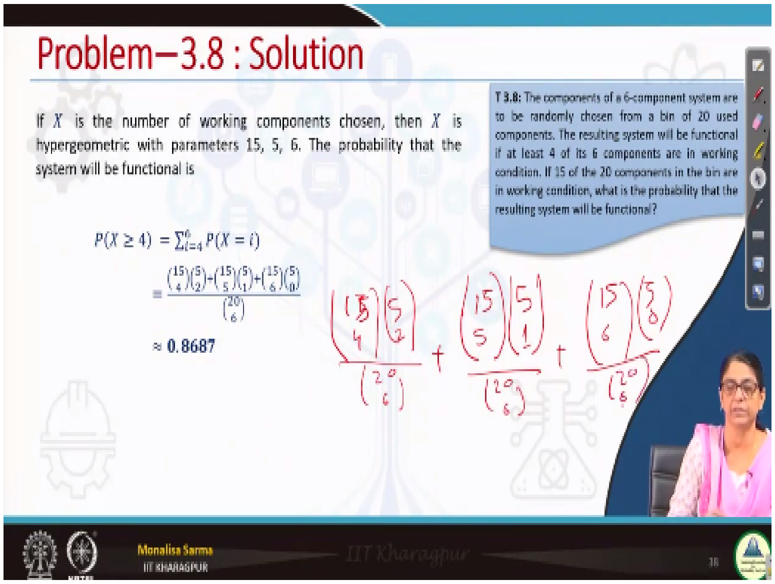
And form gradually will pick 1, will pick the second one, will pick the third one and that way we have to form a 6 component system. That directly implies we are bringing a component means we are not putting it back there. We are not replacing it back isnt that directly implied, because we are bringing a component and we are making the system and this will take this whole system and take 6 components one by one we bought 6 components. So, directly implies that we are not replacing it back.

So now next, the resulting system will be functional if at least 4 of its 6 components are in working condition for it to function, at least 4 of the 6 components should be in working condition. If 15 of the 20 components in a bin are in working condition in the bin there are 20 components out of this 20 we know that 15 are in working condition, 5 bad components are there. In a lot of 20 there are 5 bad components are there 15 good components are there.

And for the system that I have made the 6 component system picking component from the 20 components, I should get at least 4 working component than only my system will work. Now, what is the probability that the resulting system will be functional? Resulting system will be functional if I pick 4 good components, if I pick 5 good components, if I pick 6 good components, then my resulting system will work.

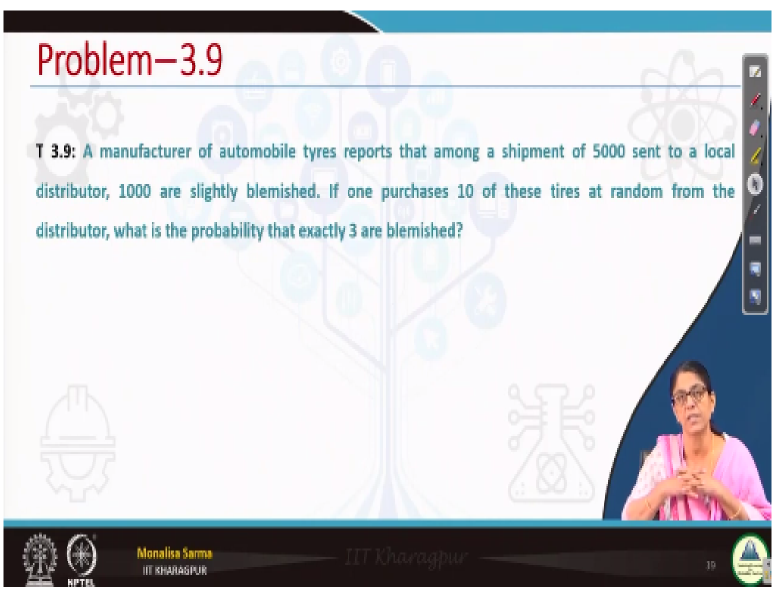
So, I will have to take these 3 scenarios picking 4 good components, picking 4 good components, picking 5 good components and picking 6 good components then only my system will work. So, definitely it is a question of hypergeometric distribution we are not putting it back and we have 2 different categories one is 20 components are working condition out of 20, 20 components out of these 20 15 are in working condition and is from 20 how many we have to pick we have to pick 6.

(Refer Slide Time: 15:56)



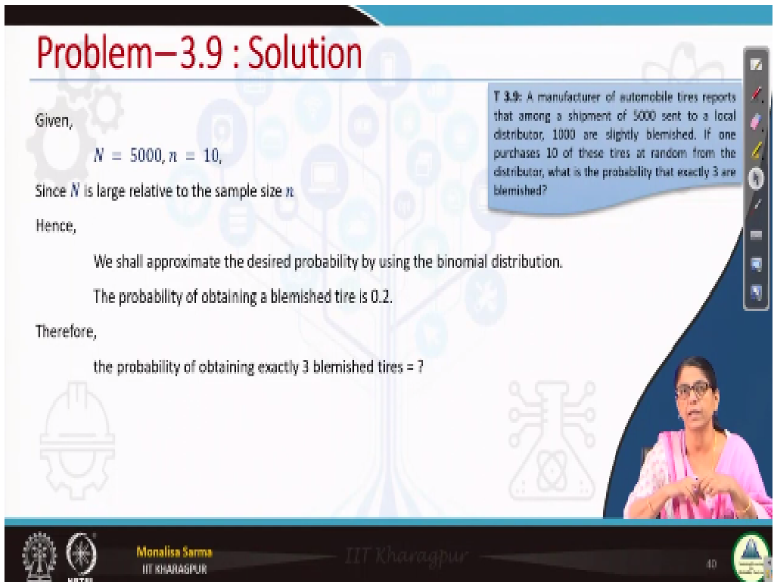
So, my sample space is 20 / 6 here I have written it together. So, basically if I write it separately it is 15 4 5 of 2 from 15 I got from 15 good ones I got 4 total I got 6 from rest 5 I got 2 and this is 20 of 6 + from 15 good got component and component I got 5 from the remaining 5 I got just 1 20 6 plus from 15 I got 6 then from remaining 5 I got 0 divided by 20 6 this I am writing it in this form here basically.

(Refer Slide Time: 16:54)



So, a manufacture automobile tyres report that among a shipment of 5000 sent to a local distributor 1000 are slightly blemished. Out of a shipment of 5000 1000 are blemished, what is the probability a blemish that means 1000 / 5000 1 by 5 if 1 purchased 10 of these tires at random from the distributor, what is the probability that exactly 3 are blemished? Fine if one purchased 10 of these tires at random, from the distributor what is the probability that exactly 3 are blemished?

(Refer Slide Time: 17:37)

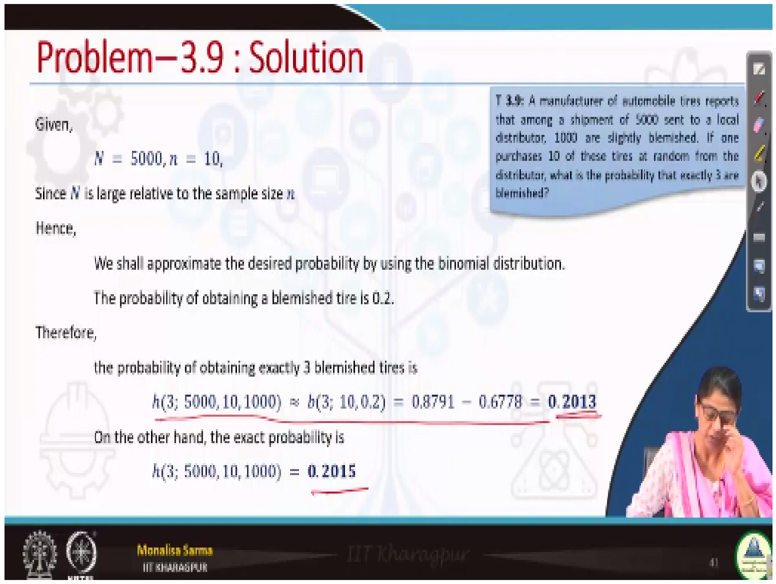


So, can you guess what is this? This is also a simple case of hypergeometric distribution, is not it?

If 1 purchase 10 of this tire, so I am purchasing 1 tire, and I am not giving it back again, is not it? Again I am purchasing another tire I am not giving it back again. So, if I am purchasing 10 tires out of 10 tires what is the probability that exactly 3 are blemished? And I know from this 5000 1000 are blemished total, there are 5000 tires out of them 1000 are blemishes.

And I am purchasing 10 one out of this 10 what is the probability that exactly 3 are blemished. So, it is directly a question of hypergeometric distribution.

(Refer Slide Time: 18:22)



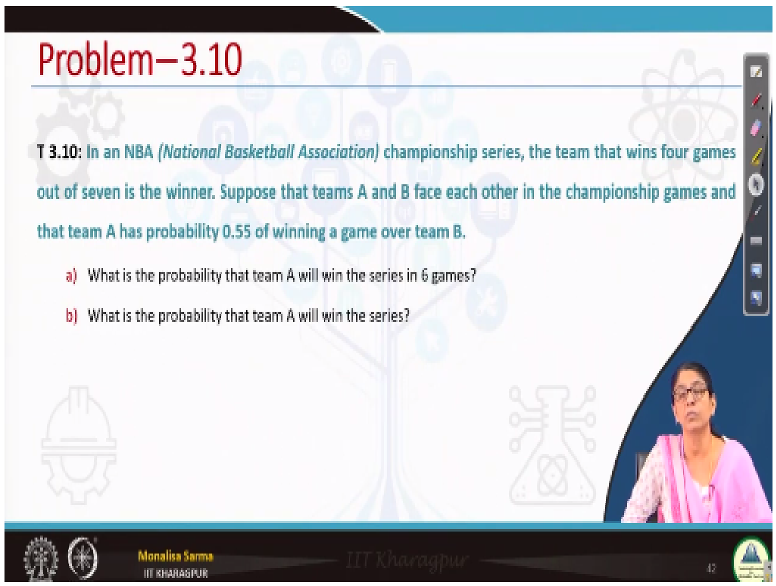
But here there is a twist here, see here; I can very well use the hypergeometric distribution here. But since N is very large relative to the sample size, as I told you from a bucket of water, if I pick one glass of water, it will hardly make it dangerous, N is very large what is N? N is 5000. And what I am picking out my sample size is only 10 out of 5000 I am picking out 10. From 5000 I am picking out 10 if I am not replacing it back also it will hardly make any difference it will not make much of a difference.

When N is very large relative to the sample size that is small and then instead of hypergeometric distribution, we can also approximate it using binomial distribution that I have already mentioned. But now the question is why we will do, when hyper it is fitting hypergeometric distribution why we will unnecessarily use binomial distribution. The main reason behind is that hypergeometric distribution the calculation is a bit tough, is not it?

We have to do from numerator and denominator the calculation is a bit complicated compared to the binomial distribution. And hypergeometric distribution you we do not have a table. So, instead of hypergeometric if we can approximate it binomial distribution, we can also consult the table and our calculation is also easier. So, if it satisfies this condition and is quite large compared to the sample size, we can approximate binomial distribution.

So, this is the question see, if we approximate if we do solve it by hypergeometric distribution, this is the thing we will get this value 0.2013 you can solve it and see. Now same thing, if we have used binomial distribution from by using binomial distribution we got this, but if we solve it by hypergeometric distribution we will get this value almost same slightly different. So, you can very well approximate by binomial this question that so, that is the objective for giving this question to bring you to your notice that how sometimes in some of the case we can approximate it using binomial distribution.

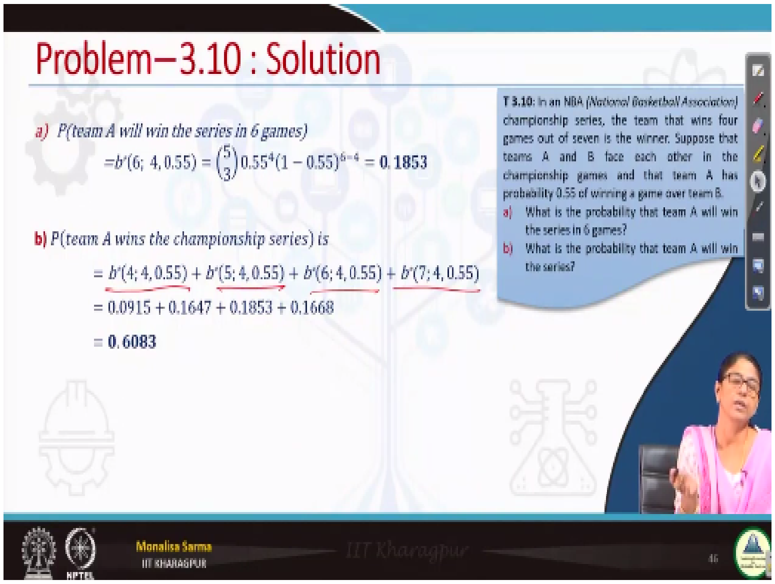
(Refer Slide Time: 20:30)



Now this in NBA National Basketball Association championship in the team that wins 4 games out of 7 is a winner. So, my number of success is 6 I need to have 4 success out of 7 games. That means number of 7 is success is fixed. Suppose the team A and B face each other in this championship games and a team A has probability of 0.55 of winning a game over team B that is probability A winning is given 0.55 what is the probability that team A will win the series in 6 games.

What is the probability that a team A will be in the series in 6 games that means the team A has for winning the series it needs to be in total 4 games, the 4th success it got in the 6th game 4th success, it is basically the question it is asking what is the probability that it got the 4th success in a 6 game. So that means it is a question of negative binomial distribution our success is fixed and number of games is the random variable.

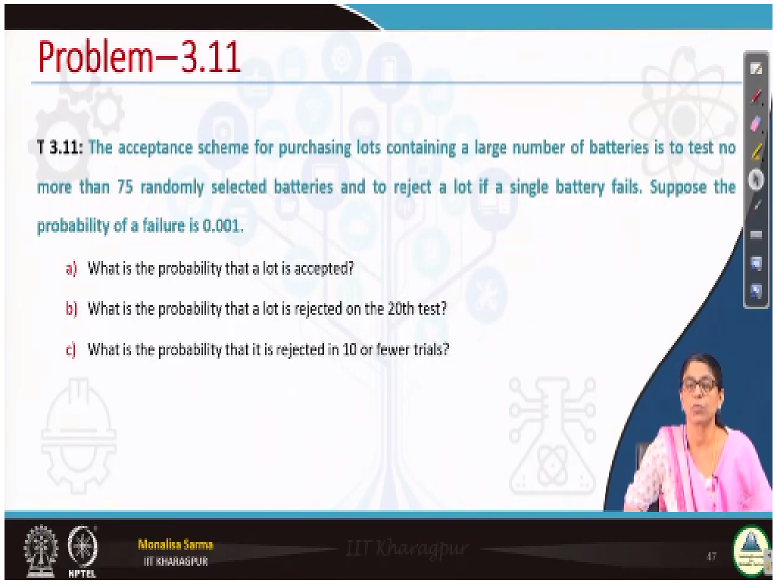
(Refer Slide Time: 21:36)



So, it is the 4th success he got in the 6th game. So that is just once we know this, you know what is n what is p, what is k, what is x than just put it in the formula. Second question, what is the probability that team A will win the series when a team A win the series for winning the series it has to win 4 games, so it can win again in the 4th game. The 4th success it can has it in the 4th game, it can have the 4th success in the 5th game, it can have the 4th success in the 6th game.

So, it will be total having the 4th success in the 4th plus having the 4th success in a 5th game plus having the 4th success in a 6th game. That is a total 7 games. So, here having the 4th success in the 4th game, having the 4th success in the 5th game, having the 4th success in the 6th game, having the 4th success in the 7th game then put it in a formula x - 1 C k – 1 x ^{k} x p $^{x-k}$.

(Refer Slide Time: 22:52)

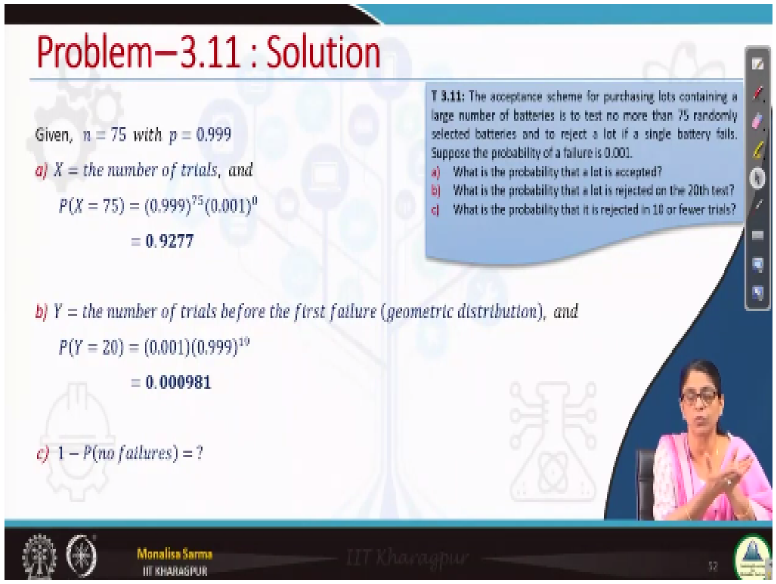


So, the acceptance sample for purchasing lots containing a large number of batteries is to test no more than 75 randomly selected batteries. And to reject a lot if a single battery fails. So, this is our acceptance sampling means is a large we want to purchase some batteries and a large number. And we will purchase the batteries for that what sampling plan we will see is we will pick at the most 75 batteries, at the most 75 batteries will fix and we will reject the lot in while picking the 75 battery if we get a single failure, we will reject a lot.

Let us see if we get a failure in the at the most 75 we will pick we will get a failure if suppose we are picking gradually 1 2 we are picking and we got the failure while picking the 10 batteries then also we will reject a lot we will go on suppose we got the first failure in that 10 to 10 also will reject a lot. So, at the most we will try till 75, till 75 if you do not get any failure then we will not see then we will assume that a lot is acceptable and we will accept it.

So, that is the question what it is asking containing a large sum is to test no more than 75 randomly selected lot is and to reject a lot if a single battery fails, suppose the probability of failure is 0.001 What is the probability that a lot is accepted when a lot is accepted? Till 75 there is no failure then a lot is accepted because it will check maximum till 75.

(Refer Slide Time: 24:27)

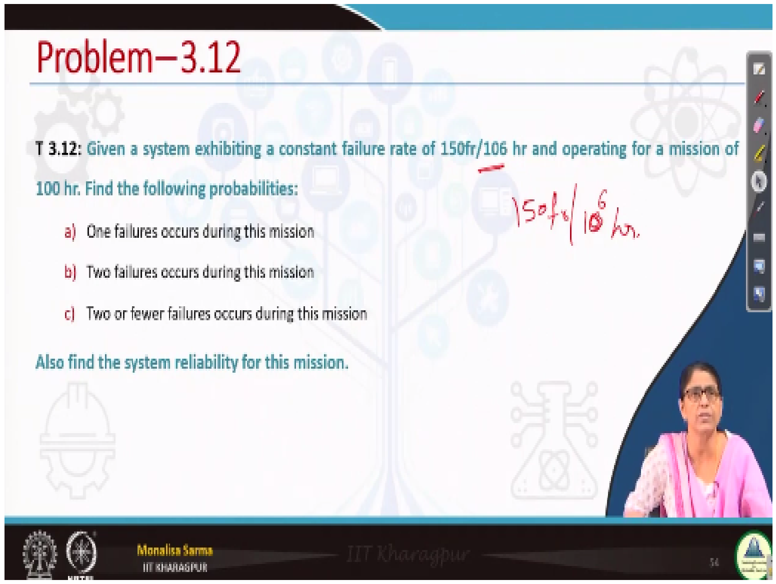


So, it is a question of geometric distribution. So, till 75 there is no failure. So, tell what is the probability of what to say probability of failure is 0.001. What is the probability of, then probability have success is 0.99 so we are using, 1 - probability of failures probability all till all 75 we got them success. So, this p to the 75 and this is q^{0}, next is what is the probability that a lot is rejected on the 20th test, lot is rejected on the 20th test means the first failure that is the first failure we got on the 20th test mean still 19 it was all success.

The first failure we got in a 28 test. So, first failure this all this trial 19 all success then we got the first failure here geometric distribution. What is the probability that is rejected in 10 or fewer trials now distribution what is the probability that it is rejected in 10 or fewer trials, what does that means? That means, the first failure we may get in the first trial or in the second trial, third trial that means, the first release from 1 to 10 we may get it in any that is what we have to find out what is a probability that it is rejected in 10 or fewer trials.

So, that instead of finding it that we are getting the first failure in the first trial plus getting the first failure in the second trial plus getting the first failure in the third trial up to up to 10 instead of finding that we can just find out a compliment with 1minus probability of no failure that means still 10, what is till 10 we got no failure 1 minus probability of 10 is getting failure between 1 to 10 is not it?

(Refer Slide Time: 26:27)



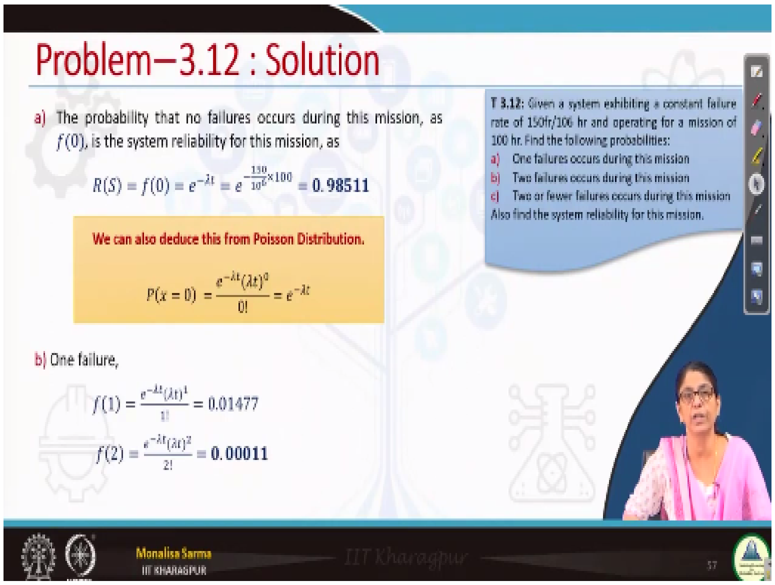
So, now, given a system exhibiting a constant failure that of 150 failure for it is not 106 hours is 10 to the power 6 hour, there is a mistake here. So, in an operating for a mission of 100 hours find a following probability. So, this here is 150 failure by 10 to the power 6, 10 over 6 hours this one this was a mistake here. Of course, when I will be giving the slide I will correct this and give it and for me to find a following probability is one failure occurs during this mission.

So, in this 100 hour time, we need to find out what is the probability of 1 failure what it is given in 10^6 hours total 150 failures this is the average failure, average failure intended for 6 hours is 150 failure and we need to find out what is the probability of 1 failure in this 100 hours 2 failures in this hour because mission time is 100 hours we have studied what is mission time in a first lecture. So, this is again a question of can you will find out? Yes, this is a question of Poisson distribution.

Where we are interested in a number of occurrences in a particular time interval or number of occurrences in a region of space, here number of occurrences in a time interval what is the time interval time interval is 100 hour. What is the λ? λ is 150 per 10^6 hours. So, what will be λ t 150 / 10^6 x 100 that is the λt so, first we are interested in 1 failure then 2 failures or 2 or fewer failures, 2 or fewer failures means 0 failure plus 1 failure plus 2 also find the system reliability for this mission.

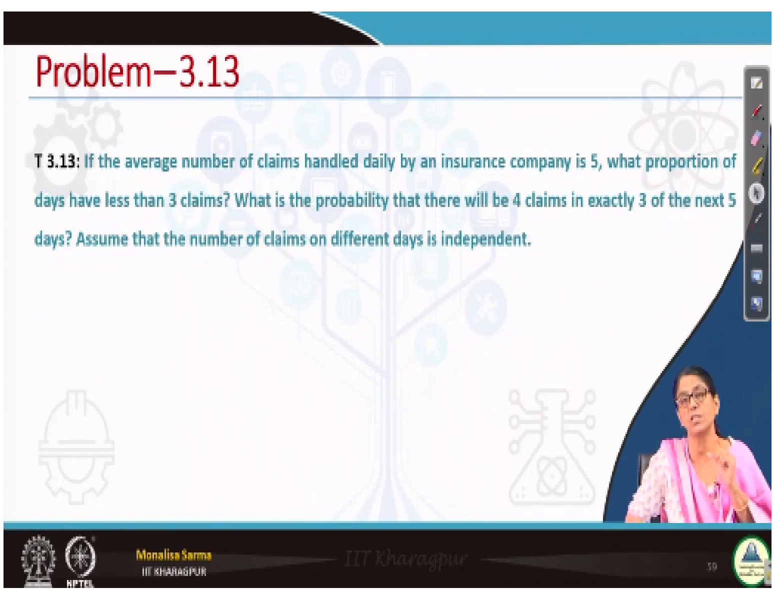
How do we find a system reliability? We talked up system level what is reliability? Reliability means failure free operation that means no failure that means 0 failures. So, what is the probability of 0 failures that is the system reliability.

(Refer Slide Time: 28:34)



First we are doing here probability finding 0 failure 0 failure is in Poisson distribution what is the formula e - λt, $\lambda t^{x\,!}$. So, what is x here x is 0 λt / 0 is 1 0 ! is 1 so, e to the power 1 - λt for zero failure. Similarly, we can do it for more than 1 failure, 2 failure and 2 or few failures is 0 1 2 summation on 0, 1 and 2, simple question of Poisson distribution.

(Refer Slide Time: 29:08)



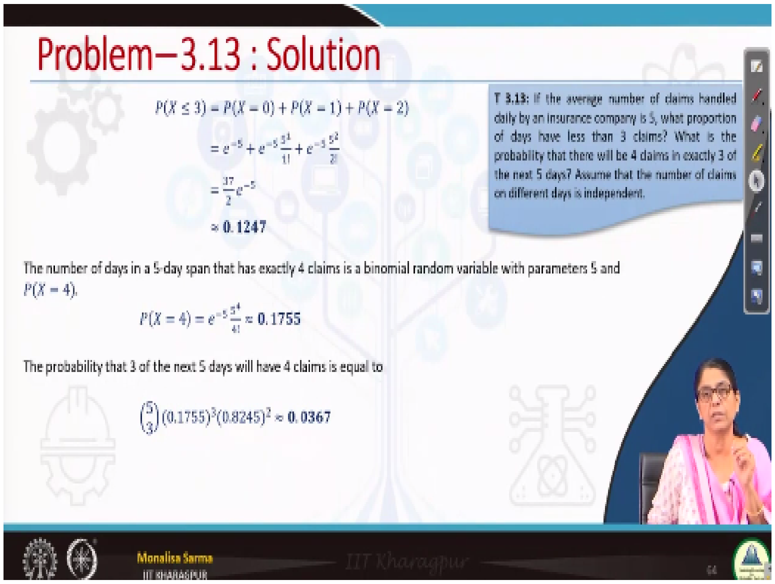
There is one more question here. This is an interesting question to see if the average number of claims handled daily by an insurance company is 5. Average number of claims is insurance companies 5 in a day, in a day average number of claims is 5 directly it is a question of Poisson distribution. What proportion of days have less than 3 claims? We have to find what proportion of days have less than 3 claims probability of less than 3, less than 3 means 0 claim, 1 claim, 2 claim that is one question.

Next is what is the probability that there will be 4 claims in exactly 3 of the next 5 days? We have to find out 3 here we have to find what is the probability that there will be exactly 4 claims in exactly 3 of the next 5 days, out of 5 days, in 3 days we will get 4 claims. So, what is this first question directly? First question is directly a question of Poisson distribution average number is given what proportion of day we will have less than 3 claims.

Probability of less than 3 probability of 0 claim what is the probability of 0 claim what is the probability of 1 claim what is a probability of 2 claim given that my lambda is 5 per day. Now, what I need to find out in exactly in the next 5 days, in exactly 3 days what is the probability that I will be getting 4 claims? So, for that, I need to find out next portion. Next portion is you see it is a binomial distribution and binomial distribution means out of 5, 5 trials my 3 success out of 5 it is a question of as if I can put it out of 5 what is the success? 3 successes.

Now I need the probability of success. So, here success means 4 claims. Here success is what success means not success, basically, for this question, I am just drawing an analogy here to success is basically the 4 claims. So, I need to find out what is the probability of 4 claims. Probability of 4 claims how do I get it? Probability of 4 claims by using this Poisson distribution, I will be getting the probability of 4 claims. Now, this probability I will be using it in a binomial distribution. So, this question is a combination of Poisson plus binomial distribution.

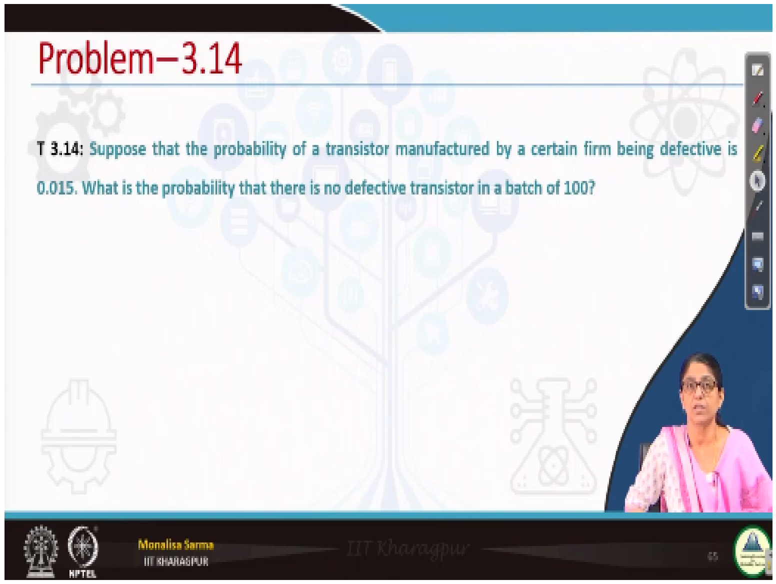
(Refer Slide Time: 31:30)



First you find a probability of x less than 3 that is for a first portion of the equation. As I told you, this, we have done probability of x = 0, x = 1, x = 2 by using the formula $e^{-\lambda t} \lambda t^{x}/x!$ then in a number of days in a 5 day span that is exactly 4 claim is a binomial random variable with parameters 5, 5 is the total number of trials that is m and p what is the p? p of x = 4 we found a probability distribution Poisson distribution.

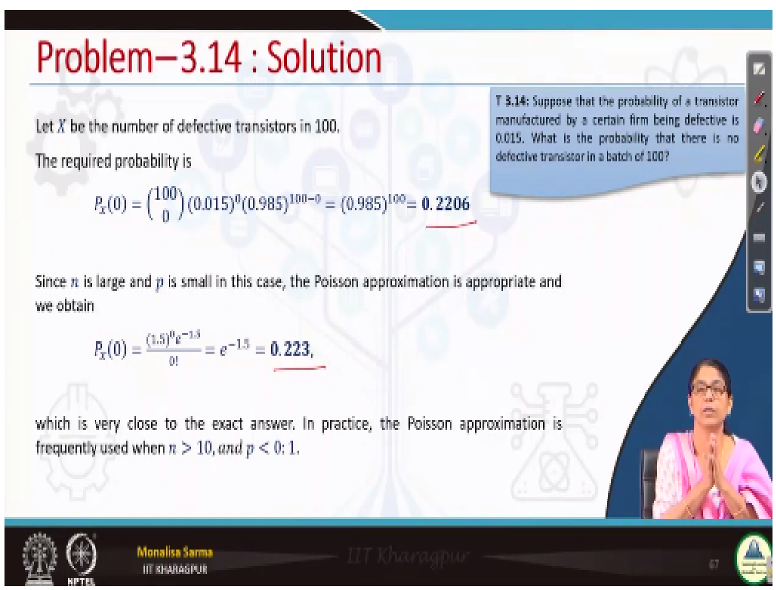
And then to find compute this probability of next 3 of the next 5 days using simple binomial distribution. So, this question we had Poisson distribution also as well as we found out the probability using Poisson distribution and then this probability we have used to calculate the value of the random variable binomial random variable.

(Refer Slide Time: 32:22)



Suppose the probability of a transistor manufactured by certain firm being defective is 0.015. What is the probability that there is no defective transistor in a batch of 100.

(Refer Slide Time: 32:34)



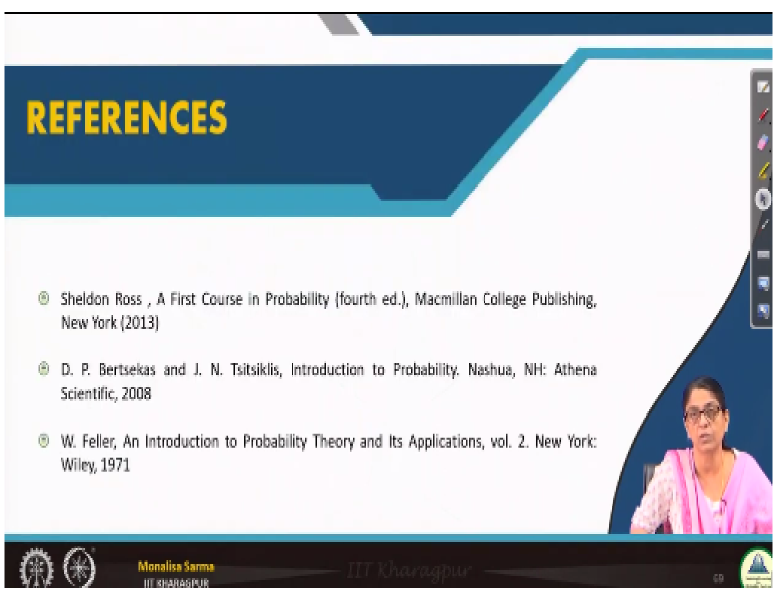
Very simple question of binomial distribution, but this question you can use, because n is very large and p is very small. Here again you can use approximation when n is very large and p is very small instead of binomial distribution, you can also use Poisson distribution. So, we have seen 2 approximation in a hypergeometric distribution can be approximated by binomial distribution when the number of population that is population size that is N is very large and sample size that is n is very small.

Then, instead of hypergeometric distribution we can use binomial distribution that we have seen. Now, again we have seen another one another one approximation in case of binomial distribution when what to say now, n is very large that is the number of trials, number of trials is very large and the probability is very small here to see if there is a batch of 100 that means, the number of trials basically it is 100, out of 100 how much we are picking up what is the probability is given? Probability is given only 0.015 p is very small.

When n is very large, p is very small instant when was under sufficient where binomial distributions would be used, instead of binomial we can approximate it by Poisson distribution as well which will not give a much of a difference in a result, you see both the result here, here we got the result is 0.2206. And here we got this not a very different high difference. So, there is a rule of term also in practice Poisson approximation is frequently used when n is greater than 10 and p is less than 0.1 and both the condition has to be satisfied.

Not that only when n is was greater than 10 we will go for Poisson and even a binomial is necessary in order both have satisfied n is greater than 10 and p is less than 0.1. So that is all and again, I request you to solve as many problems as you get which will give you give you a confidence.

(Refer Slide Time: 34:43)



So, these are the references and thank you. Thank you guys.