Probability Method in Civil Engineering

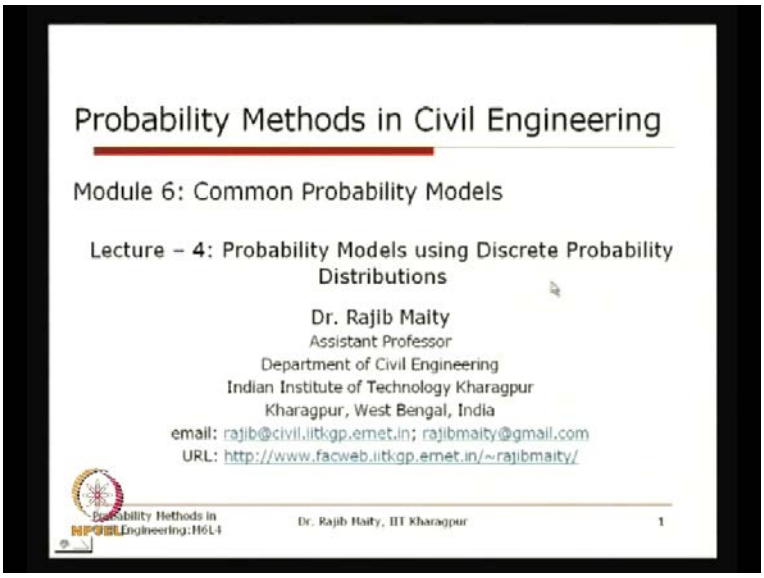
Prof. Dr. Rajib Maity
Department of Civil Engineering
Indian Institute of Technology, Kharagpur

Lecture No. # 34

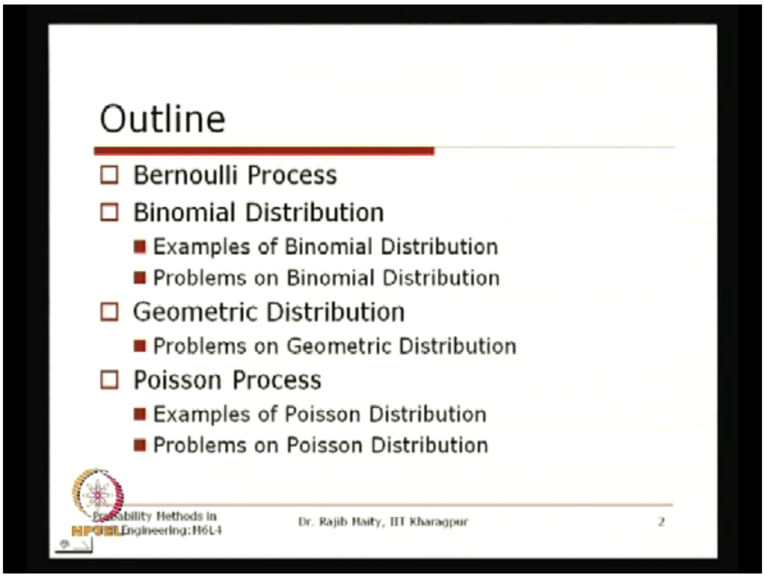
Probability Models using Discrete Probability Distributions

Hello and welcome to this lecture. In this module, we are discussing about common probability models and you know so far in the last couple of lectures, we have discussed about different continuous probability distributions and we have seen their applications in different problems of civil engineering and in this lecture we will start the discrete probability distributions. So, this discrete probability distributions, that we have discussed earlier in some module, we will just use those distributions and we will see their applications to handle different problems in, for the different problems in civil engineering. So, our today's lecture, we will deal with the probability models using discrete probability distributions and we will take couple of discrete probability distribution for example.

(Refer Slide Time: 01:01)



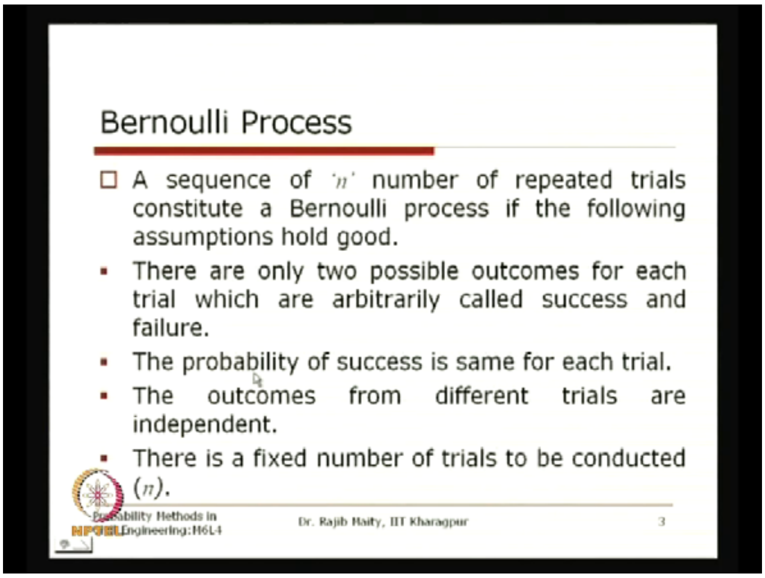
(Refer Slide Time: 01:13)



We will go first, we will take that Bernoulli distribution, that binomial distribution, which is you know that, there are certain process which we tell, that it is the Bernoulli process, there are based on some assumptions. So, we will discuss those things, first of all discuss that Bernoulli process and based on that, we will go to that binomial distribution, even though this binomial distribution is quite simpler, but simpler than other distribution that we have dealt with, but so, far as application in different problems is there, it is really an important distribution for different civil engineering problem. Particularly, where the outcome is dichotomous means, that there is only two possibilities are there. So, at that type of problem can be dealt with this binomial distribution.

Then, we will go one after another, that geometric distribution and we will discuss about its basic properties and then we will take this distribution to solve some of the problems in civil engineering and similarly after that, we will go to that Poisson process and the Poisson process is also, there are certain assumptions are there, we will discuss that one and the Poisson distribution, we will take to discuss its application to various, to handle various problems in civil engineering. So, we will start with this Bernoulli process first.

(Refer Slide Time: 02:39)



This Bernoulli's process are based on some of these assumptions and these are actually you know that these properties and these assumptions are discussed in details in the earlier module, when we are discussing about that common probability models. So, earlier means that in module 3, when we are discussing some of this standard probability distribution, there we have discussed this thing in details. Here, we will just quickly go through these properties and we will directly go to that problem, which is basically the focus of this entire module.

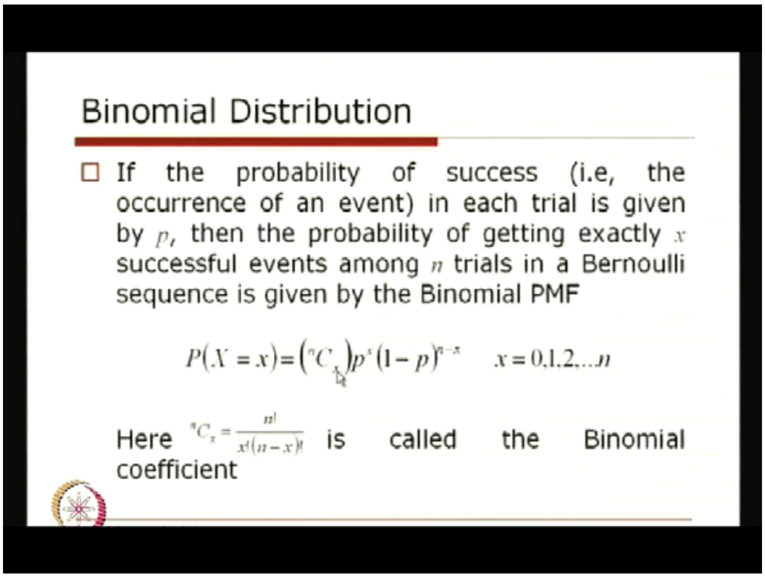
So, a sequence of n number of repeated trials constitute a Bernoulli process, if the following assumptions hold good. The first is there are only two possible outcome for each trial, which are arbitrarily called success and failure. So, as I was telling that outcome is dichotomous means, there are two possible outcomes are there. So, and this outcomes can be arbitrarily called as a success and failure, as also I have mentioned earlier this success and failure is nothing to do with their linguist meaning. So, any one of the outcome can be termed as success and another one can be failure, even that tossing a coin you know that, there are two possible outcome one is head and other one is tail, now if my goal is to get the head or that I am just going to means do some analysis, that when I am going to get the first head or something like that. That means, that getting a head might be I can term as success.

Similarly, that breaching that embankment of a canal, then I can say that the breaching event is the success which obviously, may not be the case from the social point of view.

So, these terminologies the success and failure in a Bernoulli process is completely arbitrarily can be assigned. The second assumption is the probability of success is same for each trial. So, the event that I am designate as success, that probability of that success should be same for each trial. So, for example, for tossing a coin, that getting a for a, if the coin is fair, then getting the head, if it is I can term as success then getting the head probability is equals to 0.5 and that 0.5 will not change for whatever trial I go for.

Third one is that, outcome for from different trials are independent or this means that whatever I got earlier trial, the outcome of this earlier trials, should not have any influence to the outcome of this present trial. So, the individual trials are, successive trials are independent to each other. And the last, there is a fixed number of trials to be conducted. So, this n, so, how many trials I am going to conduct, that is known, that is defined prior to the this process starts and that is designated as that n.

(Refer Slide Time: 05:58)

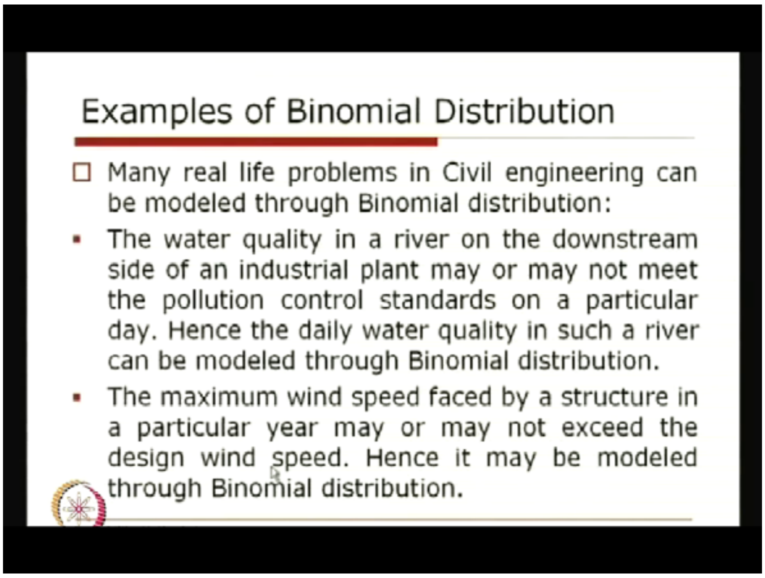


So, based on this assumption now, if we want to want to designate one random variable, that random variable is that, if the probability of the success, that is the occurrence of a particular event which we can call as success, that in each trial is given by p, as I was telling. So, this p will not change, then the probability of getting exactly x successful event among n trials in a Bernoulli sequence. Now, we have to see here that this n trials this is predefined as was the fourth assumption of this Bernoulli's process and the random number, that we are talking here is that the probability of getting exactly x successful events. So, this x, now you can say that x can take any number, any discrete number starting from 0 up to n. So, in an n trials, I can get no success, that means, x equals to 0 or exactly 1 success, 2 success, similarly I can get all n can be the successful trial. So, the random number, that we are getting is the number of success out of n trial.

So, that random number is distributed through a probability, it is a discrete probability function which is the binomial function, binomial probability mass function. And this binomial probability mass function is shown here, that is probability of that X, exactly equals to x equals to n combination C. So, this is a combination means, that out of n events, how many ways you can select x multiplied by p, which is the probability of success power x multiplied by 1 minus p power n minus x, x the power is n minus x.

Now, if we see here, there are x success and the probability of success is p and there are, so, if there are x success out of n, that means, obviously, there are n minus x failure and if there are only two possible outcome and the probability of success is p and obviously, probability of failure will be 1 minus p, and as these are independent, that is why we can multiply that their probability, so, this is that is why that this p power x multiplied by 1 minus p power n minus x. This factor that is n choose x or n combination x, is this you know that this is n factorial divided by x factorial n minus x which is this is also known as the binomial coefficient.

(Refer Slide Time: 08:37)



Now, even though this distribution I was telling, may be may look simpler, compared to the other distribution that we have dealt with, but a it is having a very wide application to handle different problems in civil engineering, particular whenever we see that there are two possible outcome of any particular event or any particular phenomenon, that we are trying to model there I can see, if I can define that there are only 2 possibilities of the outcome, then we will see that whether the other assumptions are satisfied of this Bernoulli's process or not, then we can model that kind of problems by this binomial distribution.

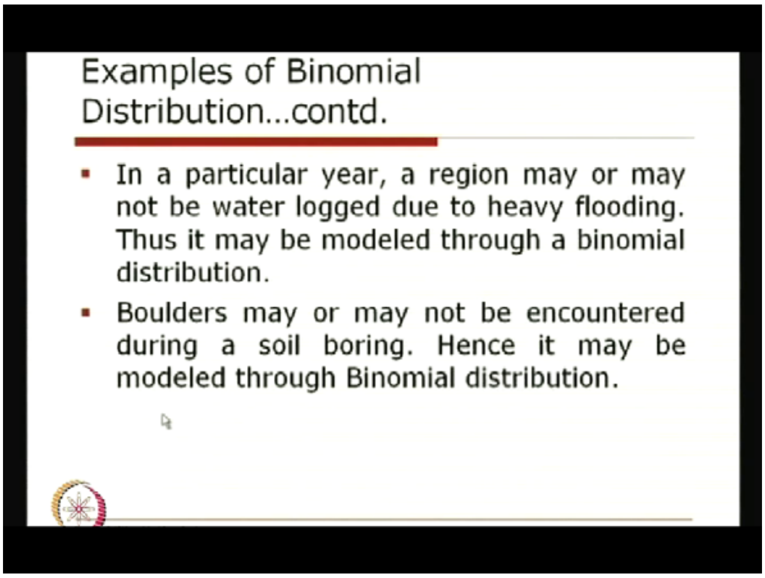
So, here are two examples are given here, say for example, that the water quality in a river on a downstream side of an industrial plant may or may not meet the population, that the pollution control standard on a particular day. Hence the daily water quality in such a river can be modeled through the binomial distribution. So, you see here that the random variable that the event that we are talking about is that may or may not meet the pollution control standard. So, we are not, so, in this type of problem, we are not interested, what is the level of this pollution, which can be a continuous random variable, but here we are just talking about the two events means, it may meet the pollution controls standard or it may not meet the control pollution control standard.

Now, these are daily basis, as it is written that on a particular day, so, these are daily basis, so, every day I will just see I will take the data, either it is this has made the standard or this has not made the standard. So, there are two possible outcome and obviously, these are independent from each other, the other assumptions if we satisfy, then this kind of problem can be handled through this binomial distribution.

Similarly, the maximum wind speed faced by a structure in a particular year, may or may not exceed the design wind speed. Hence this also can be modeled through the binomial distributions. So, each and every structure or any structure when it designed, there are several loads that we consider, one is that wind load, which is also one of the most important load, while designing the structure. So, that we assume some design wind speed for a particular, that depends on various thing, its location, then the type of the structure and etcetera.

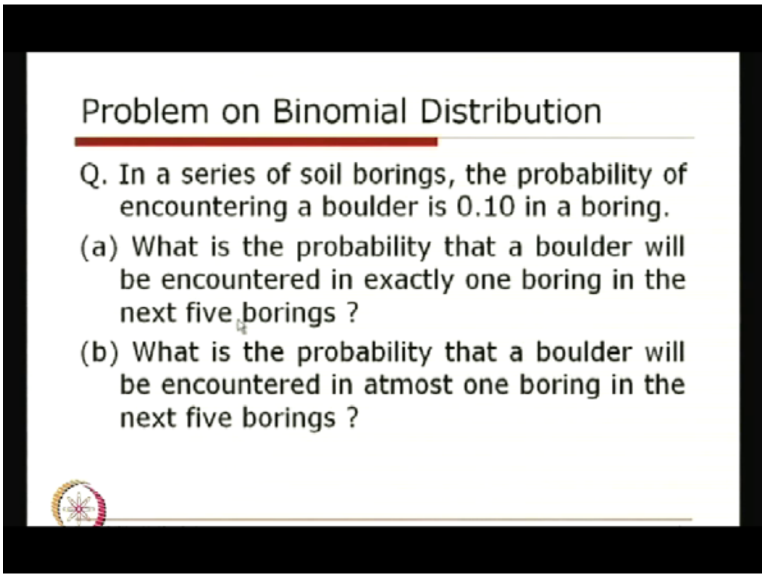
So, that design, speed once we decide and based on that we design the structure, now the question is that in a particular year, whether the designs wind speed may or may not exceed. So, each and every year, one year I will take and I will take either of these two information whether it has exceeded at some event or not. So, two outcomes are there and we can we can model this one also through the binomial distribution. Similarly, you can we can we can think of many other cases, that is after a rainstorm has occurred in a city, certain critical junction, whether that will be waterlogged or not. So, there are two things, it may waterlog or it may not waterlogged. So, like that kind of thing on a highway stretch. So, whether there is an some accident or not. So, that kind of thing. So, here you remember, that we are not interested about the quantity only we are interested that yes or no, that kind of information. So, that is the way, the random variable is is defined in Bernoulli's distribution or the binomial distribution.

(Refer Slide Time: 12:26)



So, if we take one example, that in a particular year, a region may or may not be water logged due to the heavy flooding. So, thus it may be modeled through the binomial distribution or through a geo-technical problem, if we say that while we go for this soil boring, through that boring there may or may not be some boulders can be encountered for a for a particular boring. So, that boulders may or may not be encountered during a soil boring. So, hence this also can be modeled through the binomial distribution. So, we will take one of such problem, some numerical problem to see these things.

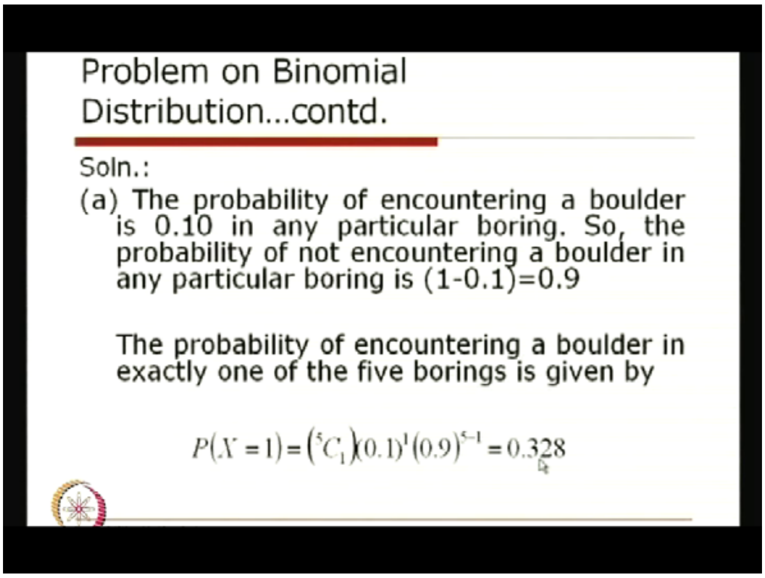
(Refer Slide Time: 13:08)



So, on the problem states, in a series of soil borings, the probability of encountering a boulder is 0.1, in a boring. So, here you see that, this is the, so, there are two events we may encounter a boulder or we may not encounter a boulder. So, we may encounter a boulder is if we arbitrarily term as success, then the probability of the success is known, it was declared that, success is 0.1, the probability of that success is 0.1.

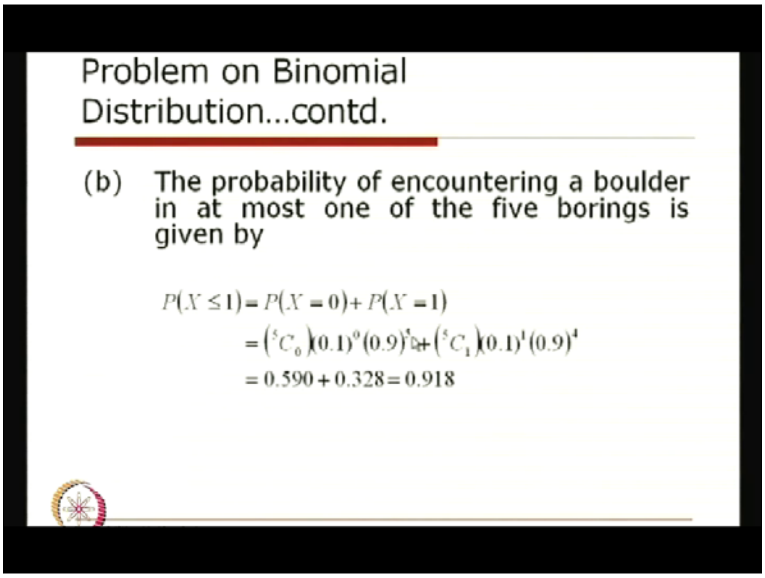
Now, what is the probability, that a boulder will be encountered in exactly 1 boring in the next 5 borings. So, whatever the information given, so, number of, so, the fourth assumption of this distribution is that, of this process that Bernoulli process is that, total number of trial should be known beforehand. So, here it is written that the next 5 borings. So, there are total 5 borings are there and we are looking for the probability that out of this 5, exactly 1 boring should encounter that boulder. So, what is the probability of that particular event? So, we know the probability of success, we know the n, n equals to 5 here. So, what is the probability to get that exactly 1 boring should encounter a boulder. Next version is, what is the probability that a boulder will be encountered in at most 1 boring in the next 5 borings. So, now, the language is slightly changed. So, one is that exactly 1 boring we are we are interested in the previous question and now here the at most 1 boring. So, at most 1 boring means it may be 0 or it may be 1. So, we have to just sum up these two probabilities, one is for the 0 encounter means a case number of cases to encounter a boulder is 0 and number of cases encountered a boulder is 1. So, if we add these two, then we will get the thing at most one boring in the next 5 borings.

(Refer Slide Time: 15:23)



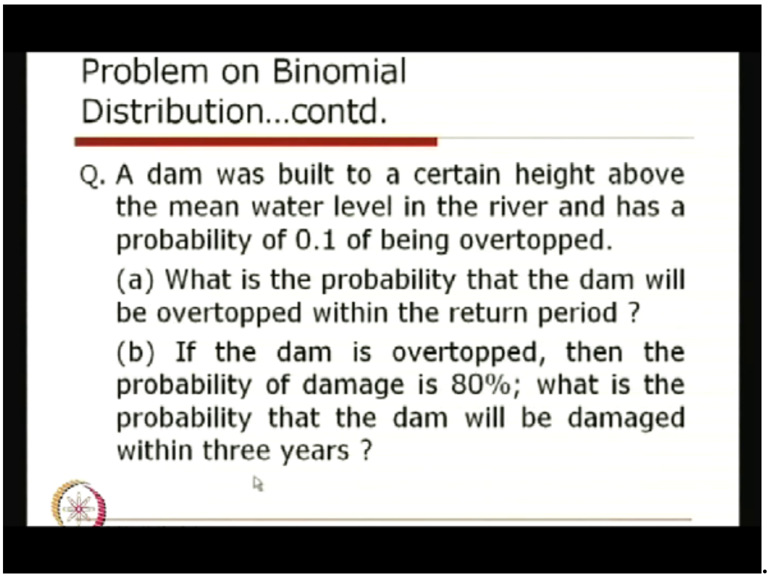
So, if we solve this one, that the probability of encountering a boulder is point 1 in any particular boring. So, the probability of not encountering a boulder in that particular boring, because you know this is the two possibilities are there. So, the other probability or the probability of failure is the 1 minus 0.1 which is 0.9. Now, the probability of encountering a boulder in exactly 1 of the 5 borings is given by P, probability of X equals to 1 is equals to 5 C 1, there is a n C x, that we use there. So, 5 C 1, 0.1 power 1 and 0.9 power 5 minus 1 which if we calculate we will get that 0.328. So, exactly 1 boring will face will encounter 1 boulder is equals to 0.328.

(Refer Slide Time: 16:17)



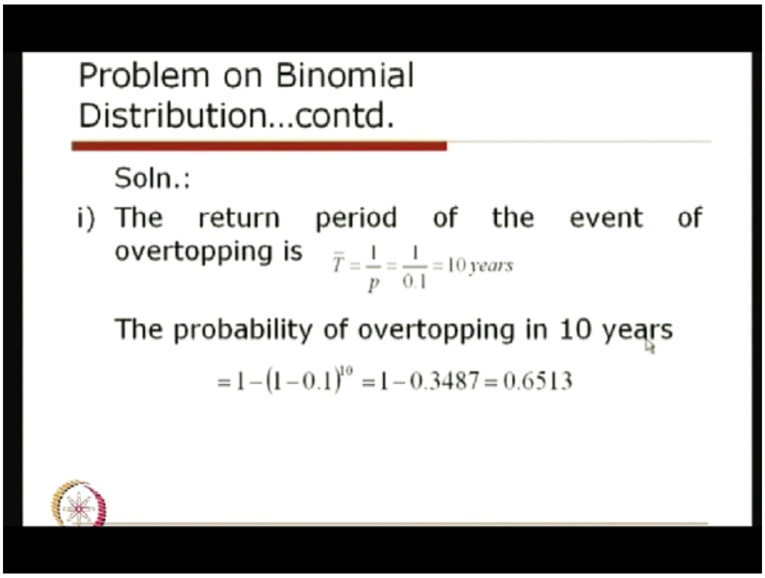
And the second was that, probability of getting the at the most 1 out of this 5 borings is that. So, we have to add two such events, that X equals to 0 and probability X equals to 1 and X equals to 1, we have just seen it is 0.328 and probability of facing 0 is 0.59. So, 5 C 0. So, n C now x is here 0, 0.1 power 0 and 0.9 power 5. So, if we get this one this is 0.918, the probability of encountering a boulder at most 1 of these 5 borings.

(Refer Slide Time: 16:57)



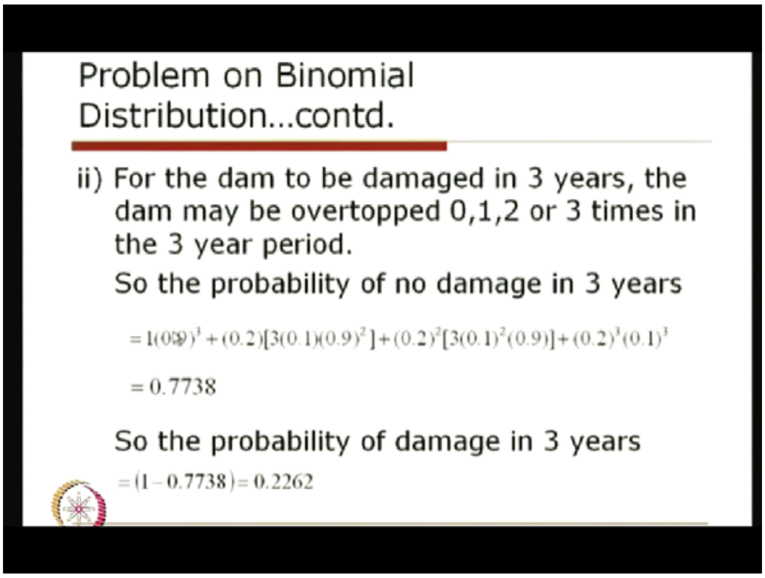
So, another example on this binomial distribution is that a dam was built to a certain height above the mean water level in the river and has a probability of 0.1 of being overtopped. Now, here you see that whether the water level of the dam will be such that, the water will overtop and for it has a probability. So, that we are now terming as a success is that whether it is overtopped or not, is that has a probability of 0.1, then what is the probability that dam will be overtop, within the return period and second is that if the dam is overtopped, then the probability of damage is 80 percentage, what is the probability that dam will be damaged within 3 years.

(Refer Slide Time: 17:51)



Now, to solve this problem, as we have seen that, the return period of the event of overtopping is T, which is the inverse of the probability of that event, that is overtopping that is 1 by 0.1, that is 10 years. So, now, the question is, the probability of overtopping in 10 years. So, if, so, the 1 minus if you see this one, that 1 minus 0.1 which is 0.9, this is the probability of not overtopping. Now, this probability power 10 is gives that it will not be overtopped in 10 successive years, now if we want to know that what is the probability of overtopping, then this probabilities should be subtracted from 1. So, by calculating, we get that the probability comes to be 0.6513, which is the probability of overtopping in 10 years; that means, the return period.

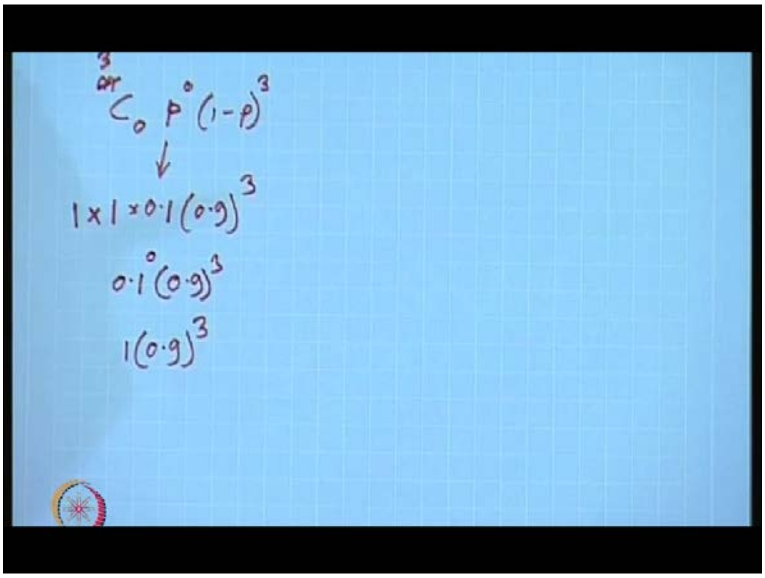
(Refer Slide Time: 18:55)



So, this is first part and the second part, which is the probability of damage in 3 years. Now, we will do this problem again that, we first will find what is the probability of no damage in 3 years. Now, if we see that how many years that it can overtop first, that is in 3 years it may so happen, that none of these 3 years is overtopped. So, the case is 0, it may so happen that only 1 year is overtopped or 2 years out of these 3 years is overtopped or it may be all 3 years, in all 3 years it can be overtopped. Now, it is given that, if it is overtopped then the probability of damage is 0.8. So, the probability of no damage in 3 years, if we want to calculate, then we have to calculate that for all these 4 cases.

Now, if we take that the first one, that is when it is overtopped for 0 cases, now if that is, it is not at all overtopped in 3 years, that means, the probability of no damage should be equals to 1 because it is not overtopped. So, there should there is no question of overtopped. So, that is why that probability is 1 now, if you just want to expand, this one then you will see this particular case. That is you know that C, that is n C how many times the success is. Now, here, a n is equals 3 years and we are saying that it is for 0 years, it is overtopped and then the probability of success is power 0 and this 1 minus p power 3, that is 3 minus 0 basically.

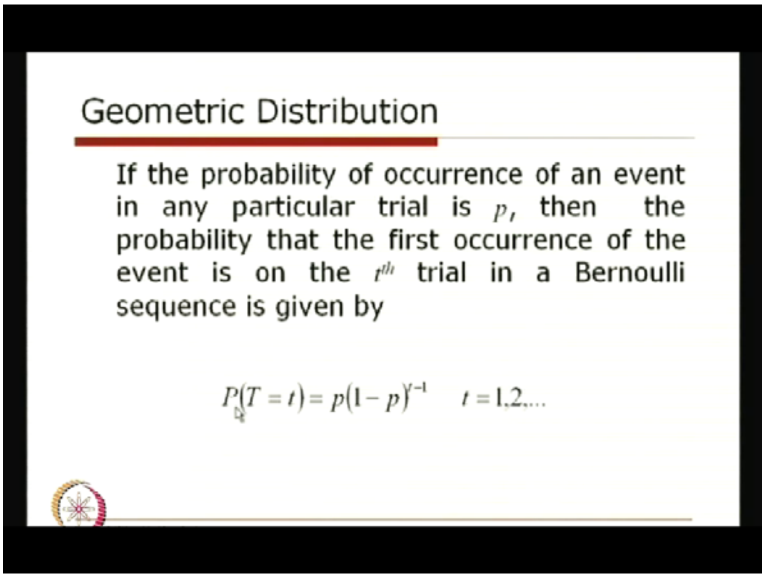
(Refer Slide Time: 20:20)



Now, you know this p is equals to your 0.1 and so, this is 0.9 power 3 and this is, obviously 1 and this should be multiplied with that chance of damage and you know that if it is not overtopped, the probability is equals to 1. So, which is basically your 0.1 power 0 multiplied by 0.9 power 3 which is, so, 1, 0.9 power 3. So, this is the first entry that you can see here in this expression and the second one is that this 3 is coming from that 3 C 1, then probability of success, which is 0.1 power 1 and 1 minus 0.1 that is 0.9 power 2. And this is the case when only 1 year out of these 3 years, the overtopped has happened.

So, and that probability of no damage is again 1 minus 0.8 which is 0.2. So, that is multiplied here. Now, if it is the 2 years, out of this 3 whether it is overtopped or not and for that it is no damage will be the probability of no damage will be 0.2 square and which is again that this 3 is again that 3 C 2 0.1 power 2 and 0.9 power 3 minus 2 that is 1 and similarly the last one is the case for when all 3 years it is overtopped. Now, if we calculate this one we will get 0.7738 which is the probability of no damage in 3 years. Now, our question is to get that what is the probability of damage in 3 years which is subtracted from 1 and is equals to 0.2262.

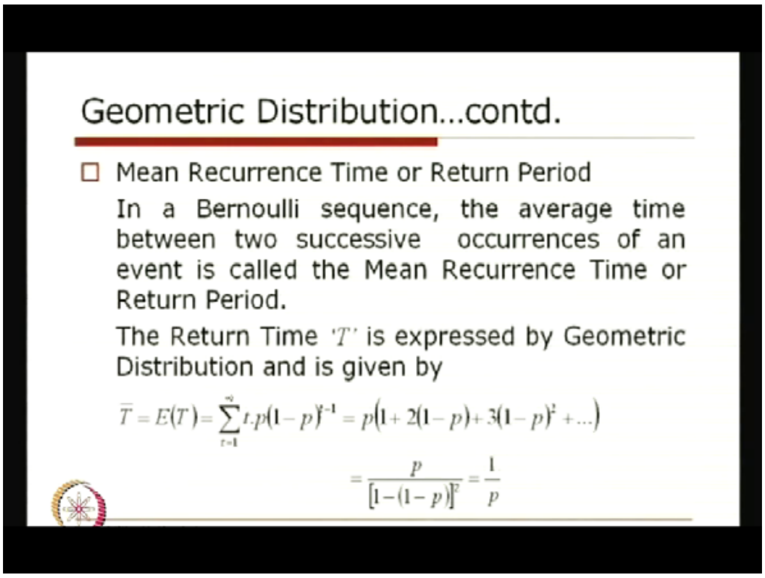
(Refer Slide Time: 22:25)



Next we will go to this geometric distribution, we will first know that, what is the difference between this binomial distribution and this geometric distribution, because the both the distribution basically is (( )) from the same Bernoulli's sequence. Now, here in this geometric distribution, it says that, if the probability of occurrence of an event in any particular trial p. Then the probability that the first occurrence of the event is on the t-eth trial in a Bernoulli sequence is given by the probability of T equals to t, p into 1 minus p power t minus 1. So, what we are looking for, what is the random variable that we are considering in this geometric distribution is that, how many trials you need to get the first success. So, success again is defined arbitrarily, that is the how many trials that we should go. So, here you can see that this probability of success and this out, so, this is power 1 and 1 minus p this is the probability of failure in t minus 1 event.

So, there are total t events are there, out of which 1 that 1 is success, the last one and this 1 minus p before that t minus 1 cases, this is a failure case. So, that we are interested to know the distribution of this T, that how many trails you need to get the first successful event. So, this random variable is modeled through this geometric distribution.

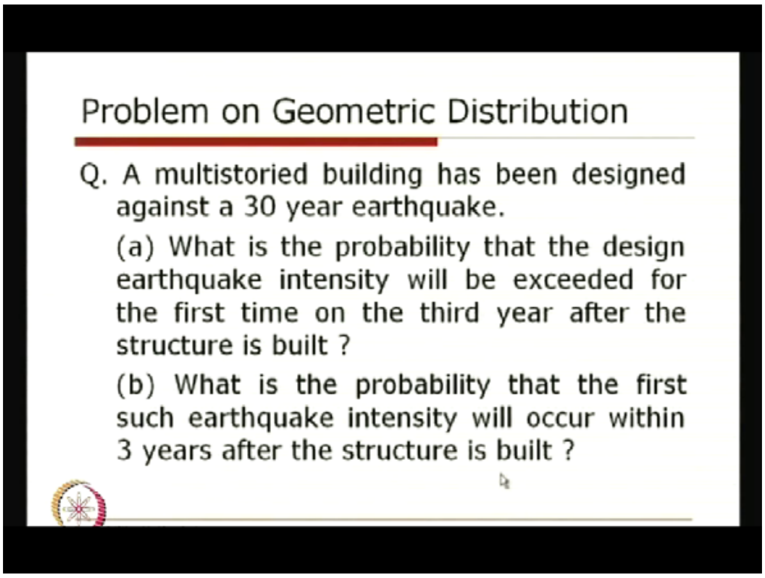
(Refer Slide Time: 24:11)



Now, this mean recurrence time or this return period, that is in a Bernoulli sequence, the average time between two successive occurrence of an event is called the mean recurrence time or return period. The return periods or return period T, this should be period, the return period T is expressed by geometric distribution and is given by the T bar, that is expectation of this T is equals to T equals to 1 to infinity t multiplied by its probability, this you know that how we get this expectation, we have discussed in earlier modules, that is if we get this one, then we will get this infinite series as this the p is taken common and this 1 plus 2 into 1 minus p plus 3 into 1 minus p square and this way and this is the expression of this infinite series is that p by 1 minus 1 minus p whole square. So, this is 1 by p.

So, the mean recurrence time of that event that how frequent that particular event, particular event means, here we are talking about that the arbitrarily selected successfully event, how frequently that is occurring is nothing, but is that 1 by p, it is probability. So, this is how that this recurrence time and the T return period and the probability is related to.

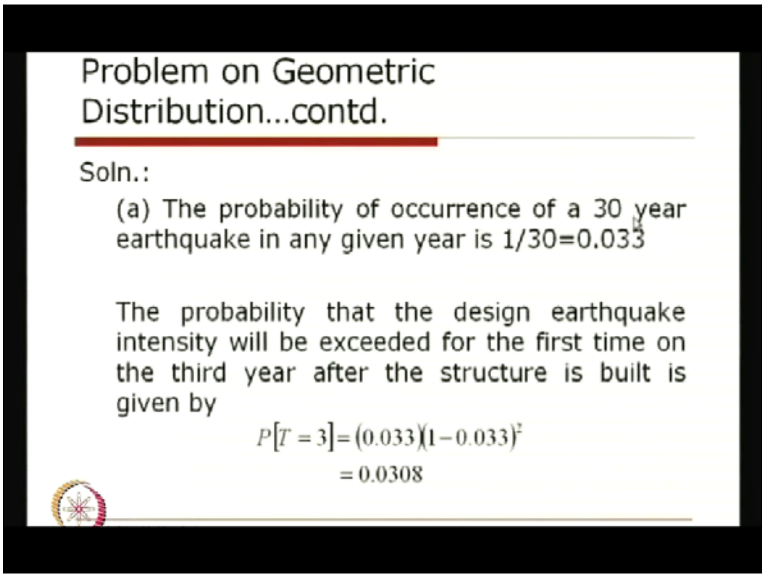
(Refer Slide Time: 25:41)



We will take one problem on this geometric distribution that a multistoried building has been designed against a 30 year earthquake, what is the probability that the design earthquake intensity will be exceeded for the first time on the third year, after the structure is built. So, here that 30 years earthquake means, this is the event, that we are talking about, that the event of the earthquake has a return period of 30 years. So, the probability of success again that having the earthquake is the term as a success. So, that the probability of the success we should get from its return period, like the way that we have discussed just now.

And the question is what is the probability that the design earthquake intensity will be exceeded for the first time on the third year after this after the structure is built. So, this is that the first event will occur at the third year itself. So, that we can model through this geometric distribution. So, the first success after some particular time stamp and here it is the third year. Second question is, what is the probability that the first such earthquake intensity will occur within 3 years after the structure is built. So, this is that on the third year and this is within the 3 years itself that design thing that the design earthquake will occur and which is we have designed for the 30 years earthquake.

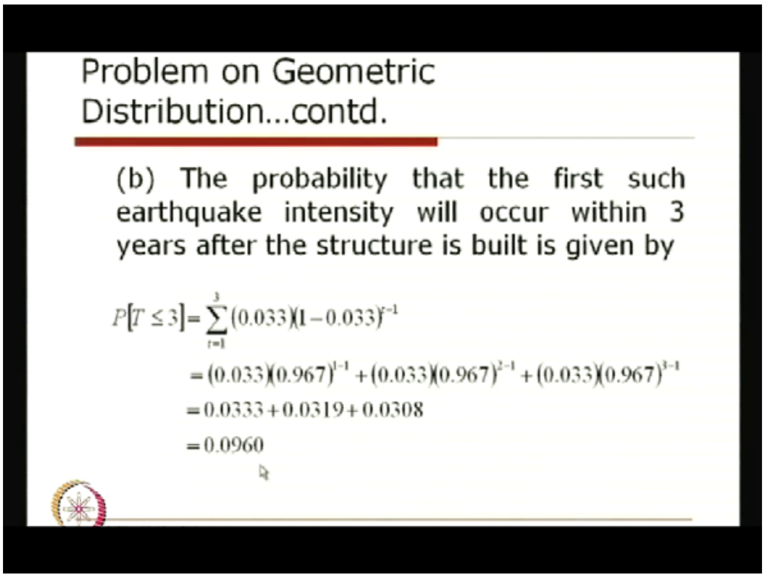
(Refer Slide Time: 27:16)



So, to solve this one, first of all we have to get what is the what is the probability of getting such event and as it is shown in the problem, that the probability of occurrence of a 30 year earthquake. So, the return period is 30 years here. So, that probability of occurrence of such event should be 1 by 30. So, which is 0.033. So, obviously, it is up to 3 decimal is considered. So, the probability that the design earthquake intensity will be exceeded for the first time on the third year after the structure is built is given by that T equals to the probability that T equals to 3 is p multiplied by 1 minus p n minus 1. So, this n minus 1 here is that 3 minus 1. So, it is 2, if we get this one, we get the probability equals to 0.0308. So, this is the probability that the earthquake will occur at the third year.

So, you can see that this is quite expected that when we design the structure for 30 years return period. So, that earthquake will occur at the third year itself. So, one tenth of the total design total design period. So, this obviously, the probability will be very less and we got the probability 0.0308.

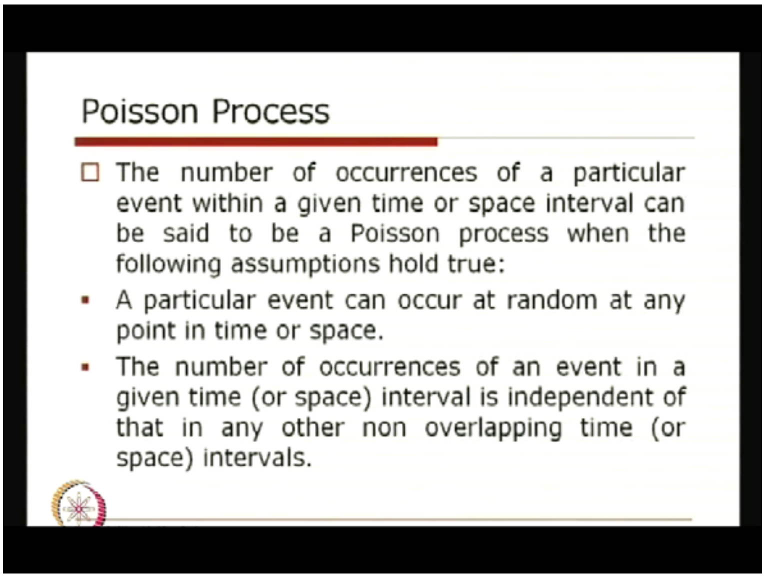
(Refer Slide Time: 28:45)



In the second was that, the probability that the first such earthquake intensity will occur within 3 years after the structure is built. Now, here within the 3 years after the structure is built means, it can happen any time between that, it may occur in the first year, it may occur in the second year and third year. And, we have seen that in the Bernoulli's process that successive events are independent. So, what we have to do is that we have to get this probability, for what is the probability for the event to be occur in the first year, what is the probability that event may occur in the second year and what is the probability that event may occur in the third year. So, these 3 probabilities we should add up. So, if this is exactly what is done. So, probability T is less than equals to 3 is that T equals to 1 to 3 and this distribution, so, 0.033 1 minus 0.033 power t minus 1. So, we can put this one as this t equals to 1 first. So, we will get this expression t equals to 2 and t equals to 3.

So, occurring and the probability of the occurrence of the event in the first year itself is 0.0333 and then it is 0.0319 and 0.0308 for the third year. So, if we add up this probabilities, we get the 0.0960. So, the probability that the first such earthquake intensity will occur within the 3 years, we are using the word within the within the 3 years after the structure is built is 0.0960 which is slightly; obviously, this will be slightly more than that what we got in the last one, which is exactly at the third year is this 0.0308, because we are including the other years also, that is first one and second one. But, obviously, if we see that there are the written period is 30 years, so, that 30 years earthquake intensity is considered in the design. So, both this probabilities will be very less and that is what we got that second case which is 0.0960.

(Refer Slide Time: 31:01)

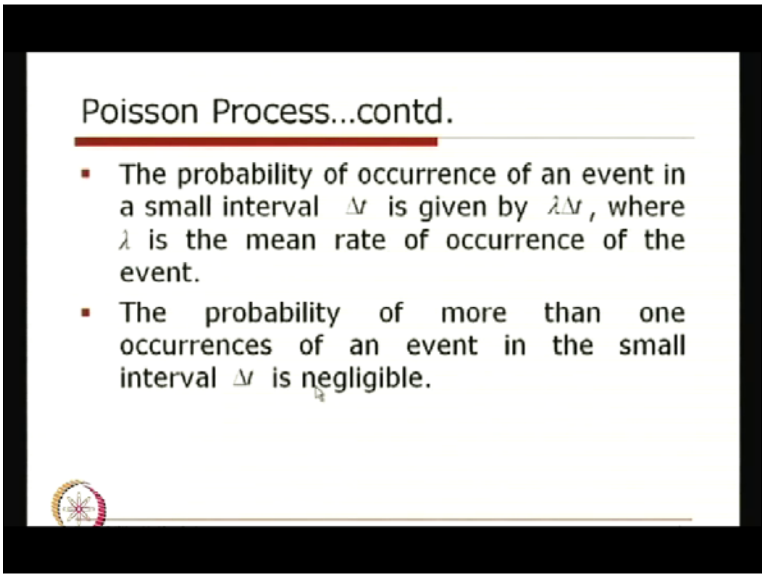


Next, we will discuss about other one that is the Poisson process. So, this is also based on some of the assumption like similar to that Bernoulli's process. So, here the number of occurrences of a particular event within a given time or space interval can be said to be a Poisson process when the following assumptions hold true. Now, we can say here is that, now the specified duration of the time or obviously, over the space or length of the highway and on so, or time is in the temporal directions and number of events over particular time over a month, over a year like that. So, now, that duration, the temporal gap or the space limit over which, that is defined and what we are interested to know is that, how many such events can occur in this, over this duration or over this special limits. And the assumptions are a particular event can occur at random at any point in time or space. So, it has no dependence of this, what has been occurred earlier.

So, if I just consider in the temporal direction. So, over suppose that we are considering a particular time from 0 to T, so, it can that particular event can occur any time within this time frame. And the number of occurrences of an event in a given time or space interval is independent of that in any other non overlapping time or in intervals. So, that is what means, what he has occurred earlier, it should be independent of what is going to occur in the next time step.

So, if we consider to suppose that from the total time span, that we are talking about this 0 to t and in between that if I just segregate this total time into two parts and this two parts are non overlapping to each other, then the number of occurrence within the first interval and the second interval should be independent to each other. So, basically there is no memory, if the process does not have any memory what has what has occurred in the last time strip or that last duration of the time.

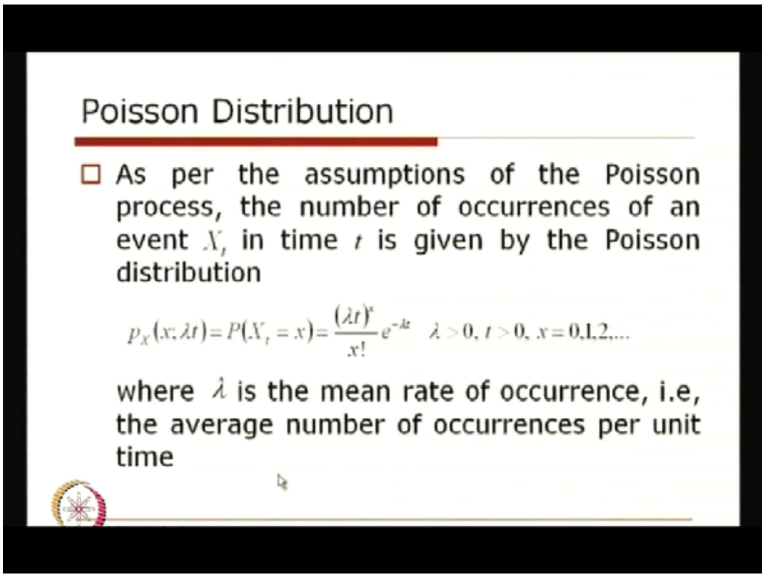
(Refer Slide Time: 33:37)



The probability of occurrence of an event in a small interval delta t is given by the lambda delta t. So, now, if we just consider a small time interval delta t over the time that we are considering and obviously, this can be the in the special direction also, we can say we can denote that one as a delta s. So, over the small interval of time or the space, that is given by the probability, that it will occur in that time interval is the lambda delta t, where this lambda is the mean rate of occurrence of that event.

So, this lambda is basically now, how we will get is the average number of occurrence of that time, number of occurrence of that event per unit time, basically that is what is shown by this parameter lambda. So, this lambda is known, before you can define a poisson process. The probability of more than one occurrence of an event in the small interval of delta t is negligible. So, in that small time, so, this is basically the infinite small, so, in this time in interval, there is the possibility of occurrence of more than one event is negligible and that is based on what we are just even that this a mean rate of occurrence of that event, which is denoted by lambda.

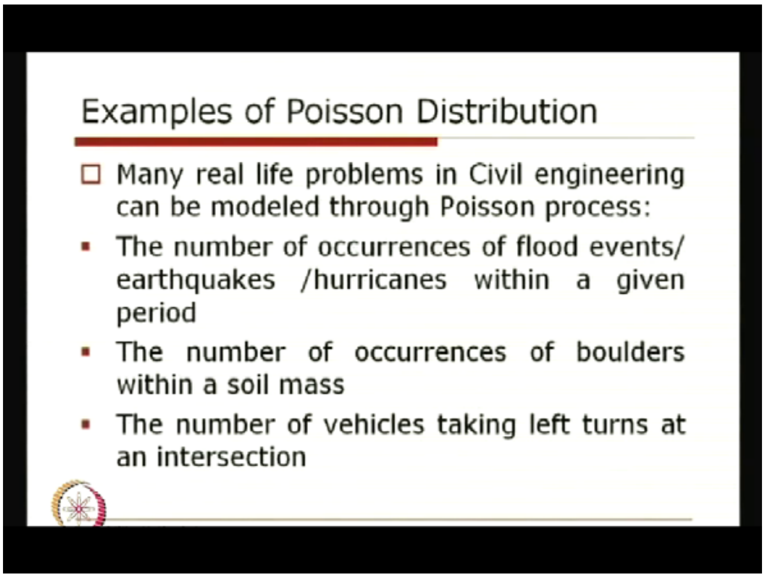
(Refer Slide Time: 35:11)



So, as per the assumptions of the Poisson process, the number of occurrences of an event X t in the time t is given by the Poisson process. So, here the random number, that we are talking about is that number of occurrences of an event X t. So, over the total time, total time from 0 to that particular time interval t. So, over this interval, what is the total number of occurrence, that we are interested here. Where this you can see that, this t can, t is a continuous variable. So, t can have a continuous length. So, at any point of time, it can happen. So, over this continuous time frame, over this continuous time or it can be space also, over this continuous special extent how many how many times that particular event can occur. So, when we are looking for this binomial distribution, there are some discrete trials over which how many times that event can occur and here in the Poisson process over a continuous time frame or continuous special extent how many times that event can occur.

So, this is the random variable of which distribution is given through this one that probability of this, this is the variable on which we are expressing, this distribution is x and the parameters is the is the lambda t, here you can see that if the lambda is the mean rate of occurrence, then this lambda t is a total number of expected that it can occur within that time interval of t. So, that X t is exactly equals to x is that lambda t power x divided by x factorial e power minus lambda t, where this parameter, there is a lambda is greater than 0, t is greater than 0 and x can take any value from 0, 1, 2 and up to infinity. So, here the support of this x, you can see that, it can start from 0 and go up to infinity, all are discrete the number of occurrences. But in the earlier that binomial distribution, it was, it can take from the 0 to that number of total number of trials considered that is up to n. So, here that what I told that this lambda is the mean rate of occurrence that is average number of occurrence per unit time.

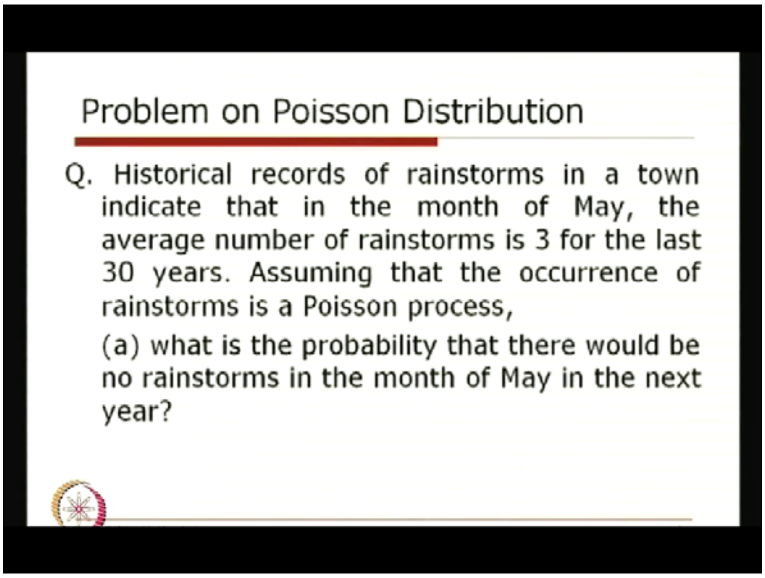
(Refer Slide Time: 37:47)



Now, many real life problems in civil engineering can be modeled through this Poisson process, first is that number of occurrences of the flood events, earthquake, hurricanes within a given period. Now, when, so, in this Poisson distribution, when we are talking about we are just talking about some discrete variable, which is the number of occurrences, but that total time frame that we define should be a continuous one, so, over a given period.

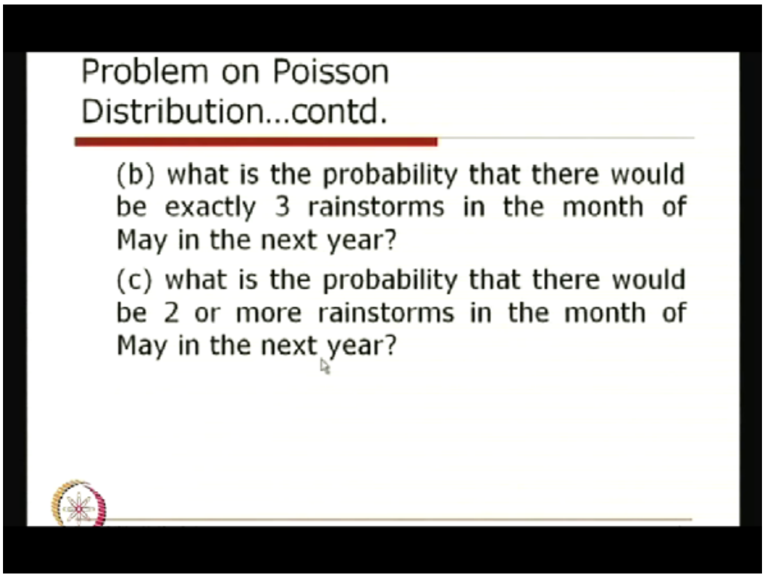
So, basically this number of occurrences can happen from 0 to infinity, because this the time frame or the special extent we are considering a as a continuous medium. The number of occurrences of the boulders within the soil mass. So, for a particular soil boring, so, how many boulders can be encountered. So, that is, so, in a soil boring means it is from the surface to certain depth. So, in this special direction at any point, at any depth that boulder can occur and any numbers of boulder can happen So, this number of occurrences of boulders within the soil mass or in the within a particular depth of soil boring that also can be modeled through this Poisson distribution. The number of vehicles taking left turn at an intersection can also be modeled through this Poisson distribution over a particular, obviously, over a particular time. So, suppose that in 1 hour, this many vehicles will take the left turn at a particular intersection. So, that is can also be modeled through the Poisson distribution.

(Refer Slide Time: 39:30)



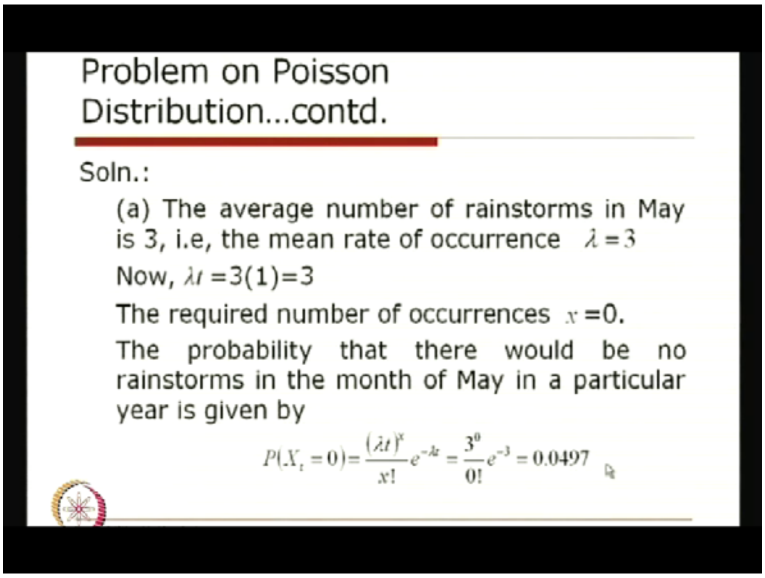
Now, if we take one example of this Poisson distribution, the historical records of rainstorm in a town indicate that, in a month of in the month of May, the average number of rainstorm is 3 for the last 30 years. So, we are considering a particular month of the year and in that month, there is an average number of rain storm for the town is 3, which is observed for this last 30 years. So, this we can say that what is the average occurrence of the event, assuming that the occurrence of the rainstorm is a Poisson process what is the probability that there would be no rainstorm in the month of May in the next year. So, that May, that is the total duration of this time, anytime the rainstorm can occur and we have seen that over that period it is there are 3 rainstorm is the is the value that we have seen from the historical record and the question is that what is the probability that there will be no rainstorm in the next year.

(Refer Slide Time: 40:44)



Other two questions are, what is the probability that there would be exactly 3 rainstorm in the month of May, in the next year, what is the probability that there will be 2 or more rainstorm in the month of May in the next year. So, basically the first question is, we have to say that exactly 0 rainstorm, second one it is exactly 3 rainstorm and third one is the, so, that summation of the probability of having 2 or 3 or 4 or 5 or any numbers can go to infinity. Now, so, to solve this third one what we can show is that what we can do is that the total probability is 1 and from that 1 we can deduct that what is the probability of the occurrence of the rainstorm, no rainstorm that is 0 event one, we can if we just sum up these two probability in case of 0 rainstorm and 1 rainstorm sum up and if we deduct from the 1 then we will get what is the probability, that it will occur 2 or more rainstorms. So, without solving you can say that the probability of this having this 0 rainfall and this exactly 3 rainfall will be much lesser compared to this last one where we are basically considering all the possibilities except that 0 and 1 case.

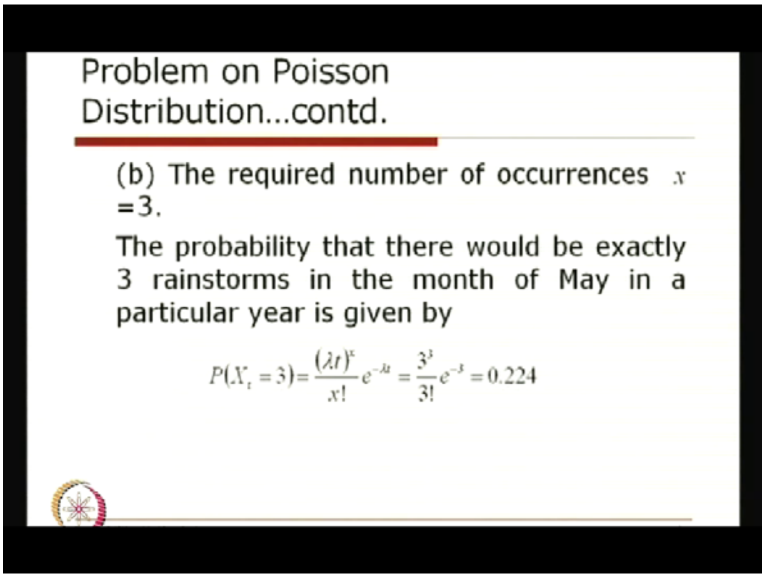
(Refer Slide Time: 42:02)



So, here you can see that the average number of the rainstorm in the month of May is 3, that is the new rate of occurrence the lambda equals to 3 and now this lambda t, when we are considering here this is a 1 month and there is average occurrence is 3. So, this lambda t is equals to 3 and the required number of occurrences here or the probability, that we are looking for the number here is x equals to 0 for the first case.

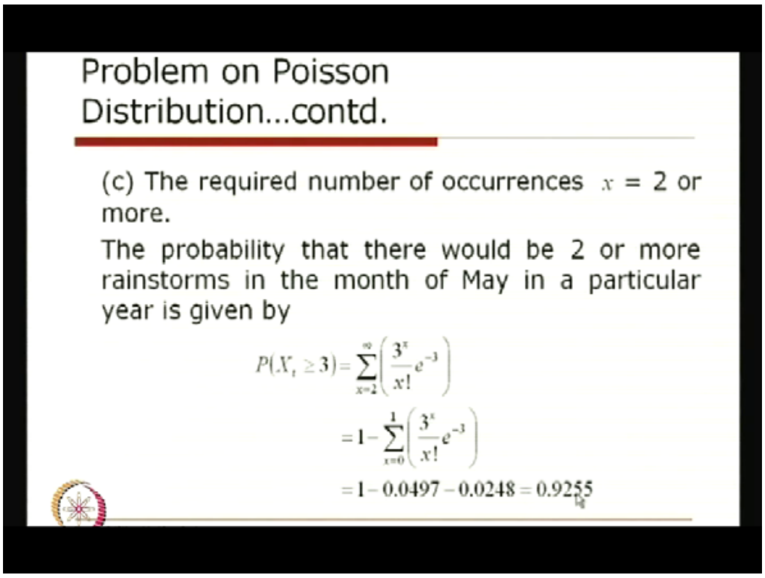
The probability that there will be no rainstorm in the month of May in a particular year is given by probability X t equals to 0 is equals to lambda t power x divided by x factorial e power lambda x and now sorry e power lambda t where this lambda t, we know as the 3 and this x equals to 0 lambda t is 3 here it is shown. So, the probability that there will be no rainstorm that is number of rainstorm is 0 is equals 0.0497.

(Refer Slide Time: 43:07)



Second the required number of occurrences x is equals to 3 here. So, number of occurrences of rainstorm is exactly equal to 3. So, we have to just put that in place of 0 we will just put 3 now in this expression and we can see that this number of occurrences here is the 0.224, which is obviously, because this is the expected number of occurrences. So, this have the higher probability, then compare to the earlier case, where we have consider the 0 of occurrences.

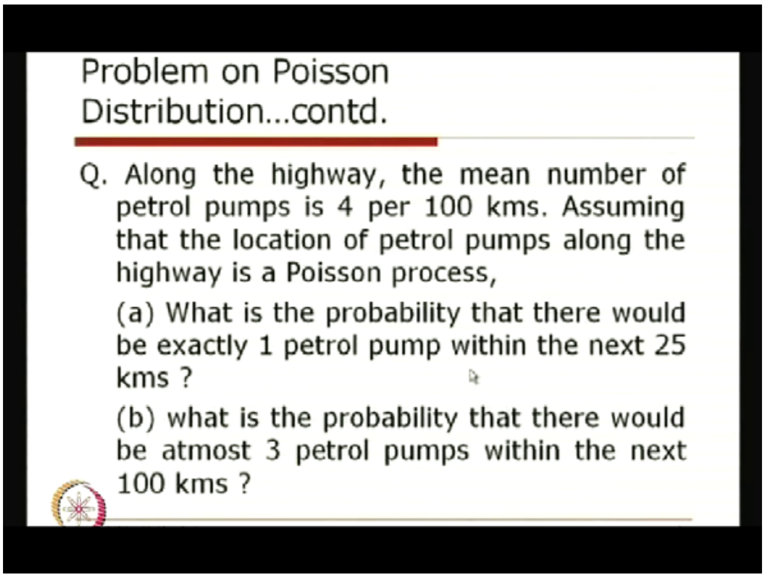
(Refer Slide Time: 43:44)



Now, the third one is that the required number of occurrences is the 2 or more. So, that the probability for the occurrences of the rainstorm occurrences x equals to 2 or more means, that includes that x equals to 2, 3, 4, 5 like this up to infinity. So, the probability that there will be a 2 or more rainstorm in the month of May in a particular year can be given by basically from x equals to 2 to infinity is, put this 3 that is the lambda t power x by x factorial e power minus lambda t. So, this what I was telling that this we can calculate just by 1 minus x equals to 0 to 1 of this one. So, the total probability, so, this from 0 to infinity, this is total is, obviously, equals to 1 from the basic principal of this probability mass function the total probability should be equals to 1.

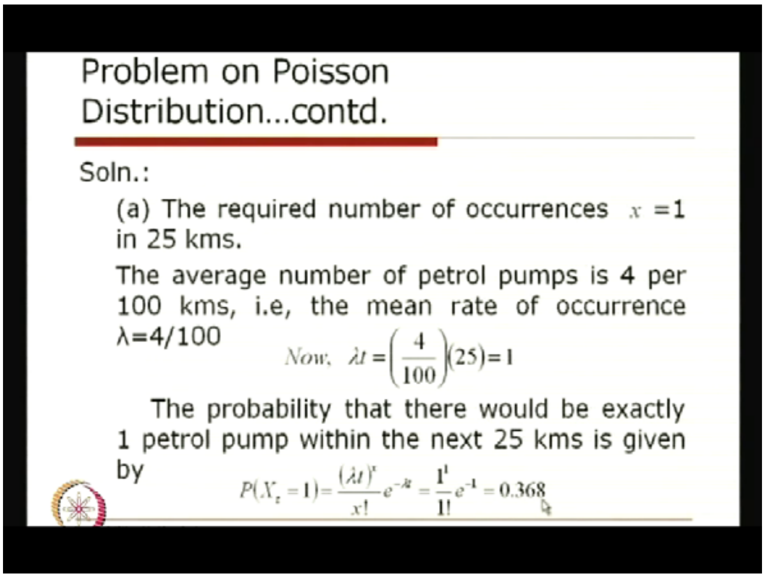
So, this 1 minus, now we can put this 0 that is we have already got this one that for x equals to 0, the probability is 0.0497 and similarly for x equals to 1, it will come 0.0248. So, that the probability that there will be 2 or more rainstorm in the month of May is 0.9255. So, you can say that this is basically taking care of almost full support of this random variables, so, which is much higher than compare to the other 2 cases, where it is 0.9255.

(Refer Slide Time: 45:19)



Now, another problem on this Poisson distribution is that along a high way, the mean number of petrol pumps is 4 per 100 kilometers. So, this is now in the special direction, earlier it was in the temporal direction now this example is on the special direction. So, over the length of 100 kilometer, on an average there are 4 petrol pumps are there. Assuming that the location of the petrol pumps along the high way is a Poisson process means, assuming that it is Poisson process means all the assumptions that we have shown earlier should be followed. So, if we can assume, if those assumptions are satisfied, then the question is, what is the probability, that there would be exactly 1 petrol pump within the next 25 kilometer. And second, what is the probability that there would be at most 3 petrol pumps within the next 100 kilometer.

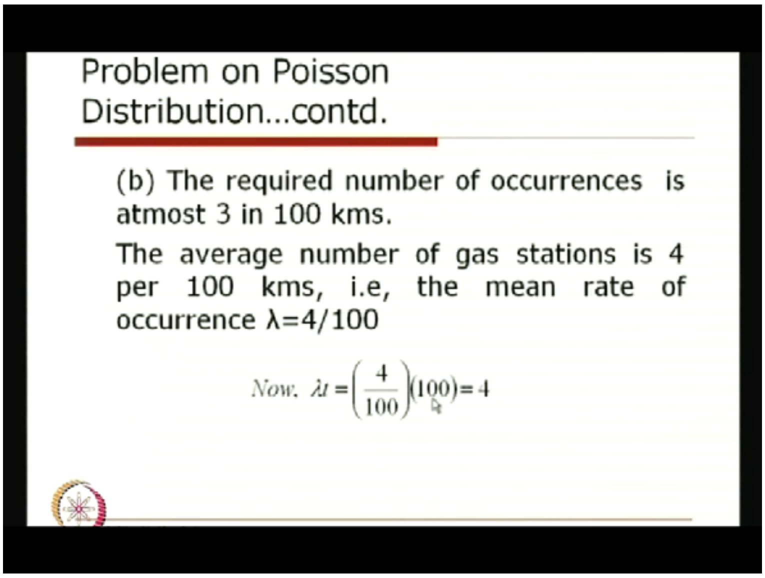
(Refer Slide Time: 46:23)



So, the first one the required number, so, required number of occurrence here is that x equals to 1 in 25 kilometer. And the average number of petrol pumps is a 4 per 100 kilometer. So, in 100 kilometer there are force, so, the mean rate of occurrences of 1 petrol pump is the 4 by 100. So, if this one is, so, 1 by 25 is the value of this lambda, now that lambda t. So, here means, even though the notation, we are using exactly similar to this distribution 1, so, here the t means, that in the special direction. So, we are interested now in this 25 kilometer. So, 25 kilometer, so, this t is here 25 and this lambda equals to the 4 petrol pump per 100 kilometer. So, it is 4 by 100. So, which is the lambda t is given the value equals to 1.

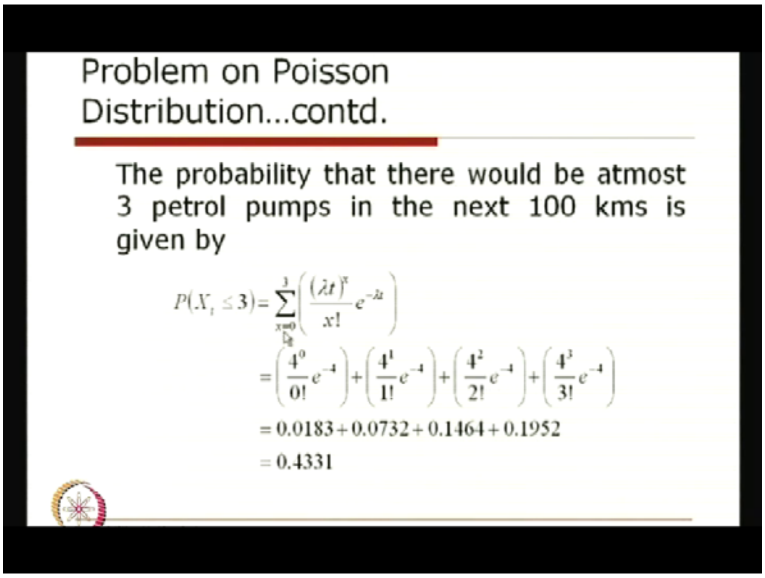
So, the probability that there would be exactly 1 petrol pump within the next 25 kilometer that means, the X t is equals to 1 with the lambda t value is again equals to 1. So, lambda t power x divided by x factorial e power minus lambda t, that means, lambda t is 1, 1 power exactly 1 petrol pump 1 power 1 factorial e power minus 1. So, the probability that we get is 0.368, that is within the next 25 kilometer the probability of getting 1 petrol pump is 0.368.

(Refer Slide Time: 47:55)



And second is that, the required number of occurrences is at most 3 in 100 kilometer. So, what is the probability, that there will be at the most 3 petrol pumps in the next 100 kilometer. So, the average number here again is this 4 by that 100. So, the lambda is obviously, 4 by 100, now if we calculate the lambda t in this case now, the t here is your 100 kilometer that we are considering, lambda is same 4 by 100. So, here the lambda t is equals to your 4.

(Refer Slide Time: 48:30)

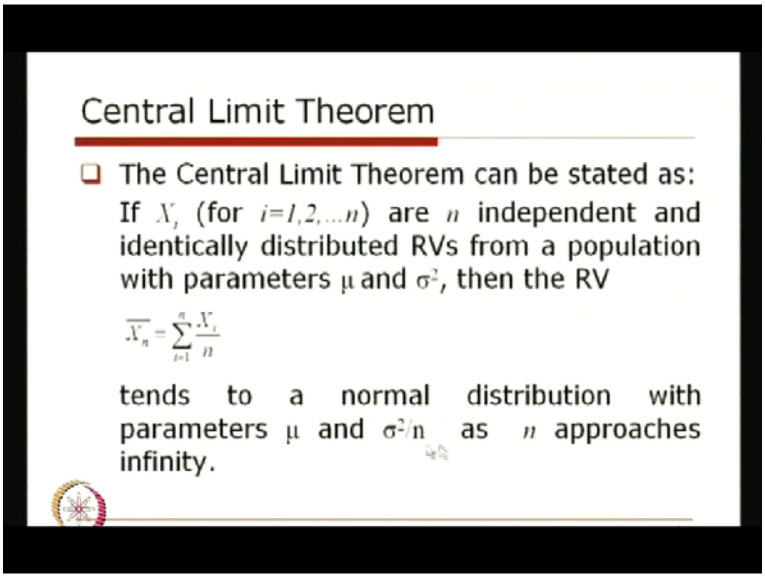


Now, if we do this one that the probability that there would be at most 3 petrol pumps in next 100 kilometer is given by that X t is less than equals to 3. So, it can happen that, there is no petrol pump, there is only 1 petrol pump, 2 petrol pump or 3 petrol pump. So, we have to sum up these probabilities, stating that x equals to 0 to 3 and this probability with the lambda t value here as we have seen in the last slide that lambda t equals to 4. So, this 4 first case is equals to 0. So, 4 power 0 by 0 factorial equal e power minus 4 similarly 4 power 1 then 4 power 2, 4 power 3 this probabilities, if we just get the first one is 0.0183, 0.0732, 0.1464, 0.1952. So, if we add up these things then we get that 0.4331. So, the probability that there will be at most 3 petrol pumps in next 100 kilometer is 0.4331.

Now, when we are, so, this type of problem that we have discussed so far, both in this Bernoulli’s distribution and for this Poisson distribution, geometry distribution. So, there are those parameters, those are the probability of the success is generally obtained from the historical record. Now, this historical record, when we are talking about basically the problems, that we are discussing may look that, if this information is supplied, then I can get this probabilities, but when we are taking that real life example, then may be those probabilities we have to obtain from the whatever the data is available to us. For example, the one example we have shown that out of this last 30 years, it has been seen that there are 3 rainstorms in the month of May. So, that last 30 years record, we can just see and that may be in the first year there are 6 rainstorms, second year there is 0, third year there is 4 like that these things. We have to see from the historical record we have to get those parameter.

In this example, that this number of occurrences of this petrol pump 4 in 100 kilometer. So, this also we have to see, that entire stretch of entire length of this highway and from there we have to find out what is the data, that is there and then, we can answer this type of probabilitical answer that, what is that at the most 3 petrol pumps in the next 100 kilometer and so, that data that we are using for this type of problem can be taken from this, whatever the historical record is available to us, in case of the temporal direction or whatever is available on ground in case of the special direction.

(Refer Slide Time: 51:24)

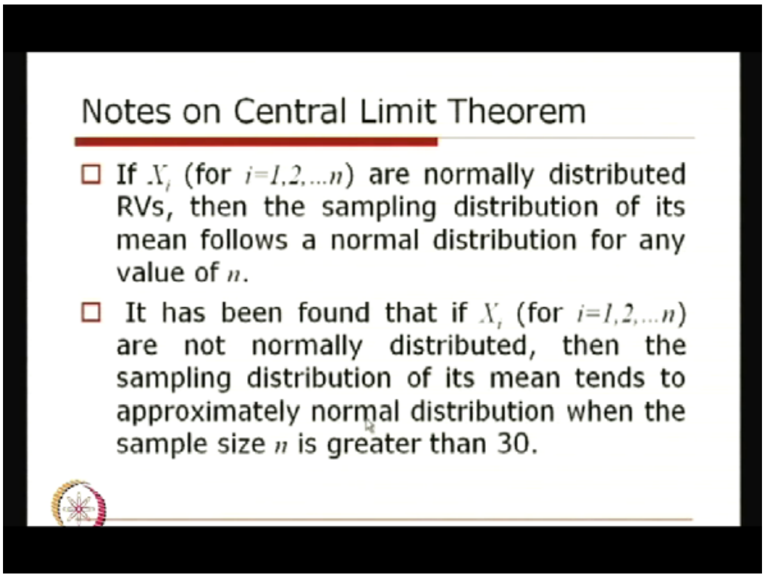


The central limit theorem can be stated as if X i for there are i equals to 1, 2, n that there are the n independent and identically distributed random variables. So, this independent and identical distribution, this we may have used earlier in other classes also, this means that this X 1, X 2, X 3, up to X n, these are the random variables, which are independent to each other and their distributions are identical. And there from a population with the parameters of this mu and sigma square. So, they are all having this parameters the mean is the mu and the variance is this sigma square.

Then, the random variable, which is the X n bar, that is the average of this, all this n random variables, which is we can express at the summation of this all X i divided by n, which is the arithmetic average. So, this is a new random variable, which we can say that it also tend this tends to a normal distribution with the parameters means, mean equals to mu and variance equals to sigma square by n, as n approaches to infinity. Now, I see here we have not defined any particular distribution of this X i' s. So, it can have any distribution, but only thing is that whatever the distribution it is having these are identical distributions that what we have defined, but what is there parent distribution, what is the distribution of their particular distribution, this X i' s, we do not know.

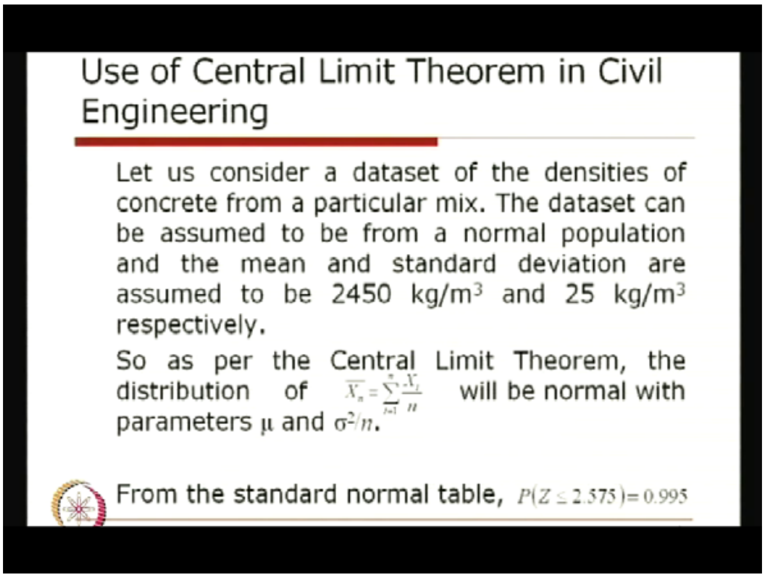
So, even though we do not know that distribution, then their mean can be a normal distribution with this parameters. Now, if we are interested, if we know that the particular random variable, that we are considering, which can be considered to be the mean of the several others random variable, even though we do not know what is the distribution of their of those random variables, we can model this mean as a normal distribution having these parameters mu and sigma square by n.

(Refer Slide Time: 53:56)



And other things if the X i, so, the X 1, X 2 up to X n are normally distributed, if we say that this is normally distributed, then the sampling distribution of its mean follow again a normal distribution for any value of n. So, if the parent distributions are normal distribution, then for any values of this n, we can say that their mean will follow the normal distribution, with the parameters that we have discussed now. Now, it has also been shown that if this X i' s are not normally distributed with means, this is the most interesting case, even though this X i' s are not normally distributed, then the sampling distribution of its mean tends to approximately normal distribution, when the sample size n is greater than 30. So, we can safely assume that, this mean of this X 1, X 2, up to X n, if that n is greater than 30, then that mean will be approximately normal distribution with the parameters, that we have discussed in the last slide with the mu and sigma square by n.

(Refer Slide Time: 54:46)

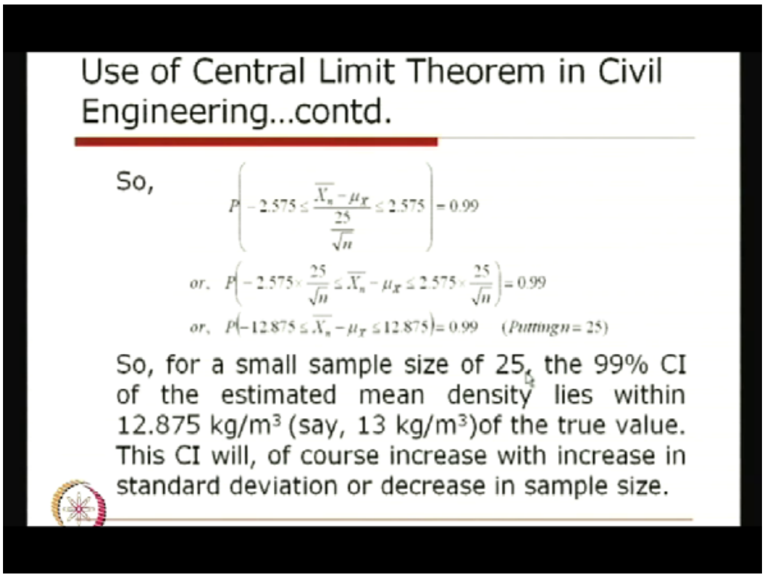


Now, let us consider, that a data set of the densities of the concrete from a particular mix, the data set can be assumed to form a normal population and the mean and standard deviations are assumed to be 2450 kg per meters cube and this standard deviation is 25 kg per meter cube. Now, if this is the thing because now you see that, I have given this as the mean and this is as the standard deviation, now because these things, we can cast several such concrete from that particular mix design and we can have their records and those records, if we get their mean and standard deviation we will get this values. So, background to this two data that is 2450 kg per meter cube and 25 kg per meter cube is that that experiment.

So, far as a central limit theorem, as per the central limit theorem, the distribution of this X n bar, that mean of their densities, that is the X i by n will be a normal distribution with parameter mu and sigma square by n is the variance. So, from the standard normal table that is probabilities that Z is less then equals to 2.575

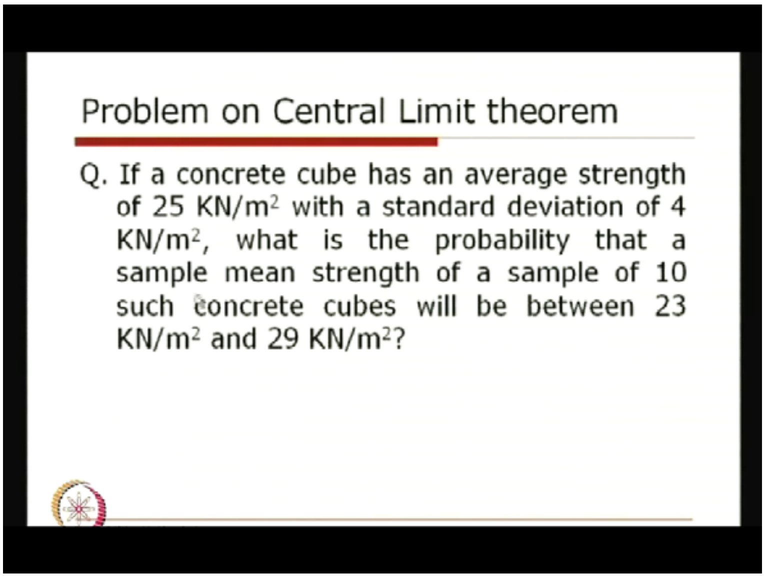
So, this 2.575, you know that we have referred to this standard normal table several times, this is respective to this 99 percentage probability, 0.99.5. So, this is a 99 percentage confidence intervals to the right hand side 0.05 and left hand side 0.05 is left.

(Refer Slide Time: 56:24)

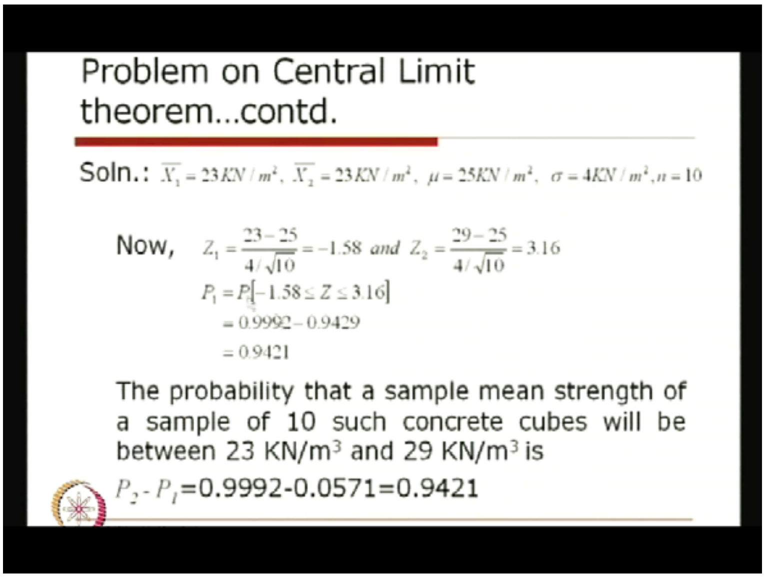


So, this one if we see then this X bar minus mu x divided by this standard deviation, that is 25 by this square root of their n, standard deviation by n, this is basically the sigma square when we are talking about this is 25 square. So, the standard deviation will be, that 25 square root of 25 square by n. So, 25 by square root n, if we take this one, then we can do this one, this one approximately it will become as this X i minus mu x to this 1 and finally, we will get the 0.99 putting this one. So, for the small sample size of 25, the 99 percentage confidence interval is estimated mean density lies within this 25 sorry 12.875 kg per meter cube, say 13 kg per meter cube of the true value, this confidence interval will of course, increase with the increase in the standard deviation or decrease in the sample size as you can see it from here. So, this is the true value. So, deviation from the true value for the 25 samples, sample size of 25, if we put that n equals to 25, then it will become as 12.875.

(Refer Slide Time: 57:34)



(Refer Slide Time: 57:56)



Quickly, we will see one example, if a concrete cube has an average strength of 25 kilo Newton per meter square, with a standard deviation of 4 kilo Newton per meter square, what is the probability that sample mean strength of a sample of 10 such concrete cubes will be between 23 kilo Newton per meter square and 29 kilo Newton per meter square. So, this we can get from there their respective means and we can reduce variate, we can get their standard deviation also is known. So, this type of problem, we have shown earlier, how to refer to the standard normal table and all. So, you can get the probability 0.9421. So, the probability that a sample mean strength of the sample of ten such concrete cubes will be between 23 kilo Newton per meter cube and 29 kilo Newton meter cube is your this 0.9421, which is the difference between these two.

So, in this module, we have seen that there are several examples have been taken and this one. So, the few continuous distribution few discrete distribution we have taken and from next module onwards, we will see some of this statistic, sample statistic based problem how to define that different distribution, how to satisfy which distribution is following and other statistical test we will follow in the next module. Thank you.