Business statistics
Prof. M.K. Barua
Department of Management Studies
Indian Institute of Technology-Roorkee

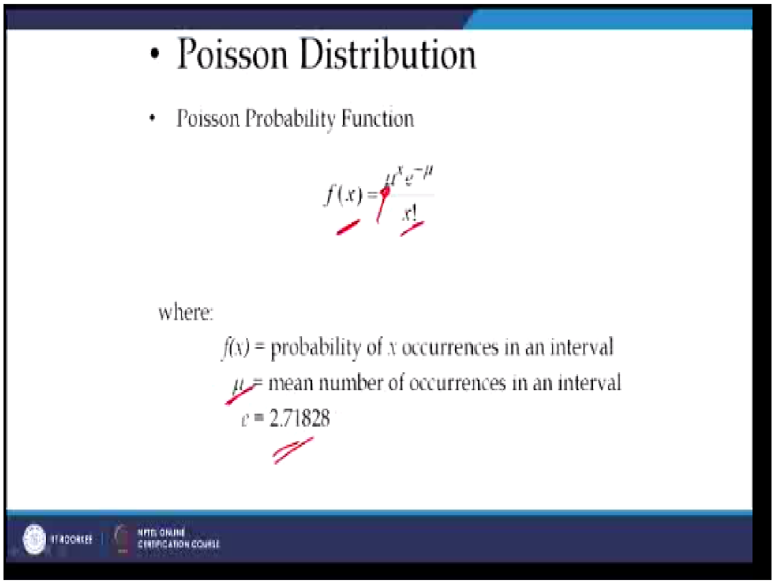
Lecture-17
Probability Distributions-Part 2

Good morning friends, as you are aware in previous class we were discussing about probability distributions and we have seen two discrete distributions. The first one was binomial distribution and we have seen that in binomial distribution the probability of outcome remains fixed and you will always have only two outputs. For example when you toss a fair coin you will always have either heads or tails right.

So, the outcome would always be only 2 right, for example failure or success whether it will rain or it will not rain and so on right. So, the you will have only 2 outcomes and probability also remains fixed. We have also seen Poisson distribution which is again a discrete distribution and poisson distribution is a distribution wherein we try to find out probability of rare events. For example let us say what is the probability that there will be an accident in let us say any atomic energy plant.

So, it is a rare event, so you can find out probability of such a rare event using poisson distribution. Poisson distribution also helps in knowing the probability of accidents happening at the intersection, what is the probability of the vehicle coming at a toll booth, what is the probability of phone calls going from a switching board system. So, all these processes can be defined by poisson distribution.

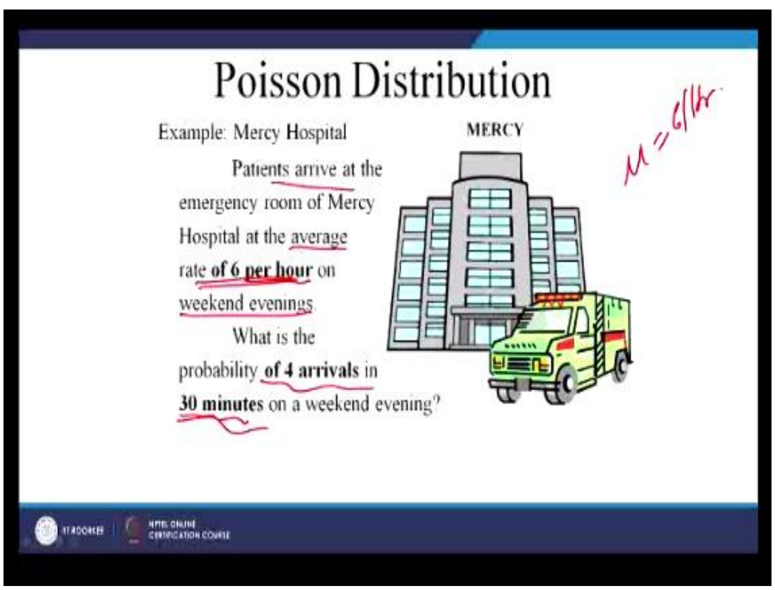
(Refer Slide Time: 02:26)



And we have seen poisson probability function $f(x) = \frac{\mu^x e^{-\mu}}{x!}$

So, we know that this is mean number of occurrences in a given interval and e=2.71828. So, let us workout couple of examples I take to poisson distribution. let us take this example, a example of Mercy hospital.

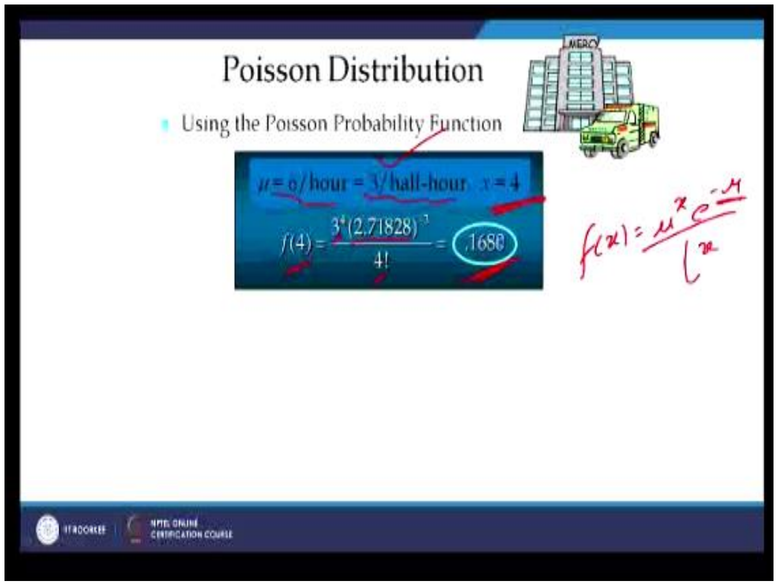
(Refer Slide Time: 02:56)



So, we know that patients are arrived at the emergency room of mercy hospital at the average rate of 6 per hour on weekend evenings. So, we know μ right this 6 per hour right what is the probability of 4 arrivals in 30 minutes on weekend evening. Now this what we have to calculate, now if you look at this arrival rate is even in terms of per hour while we have to find out probability of 4 arrivals in 30 minutes.

So, we need to convert this arrival in terms 6 hour or how many arrivals per 30 minutes then you can apply the probability function formula.

(Refer Slide Time: 03:52)



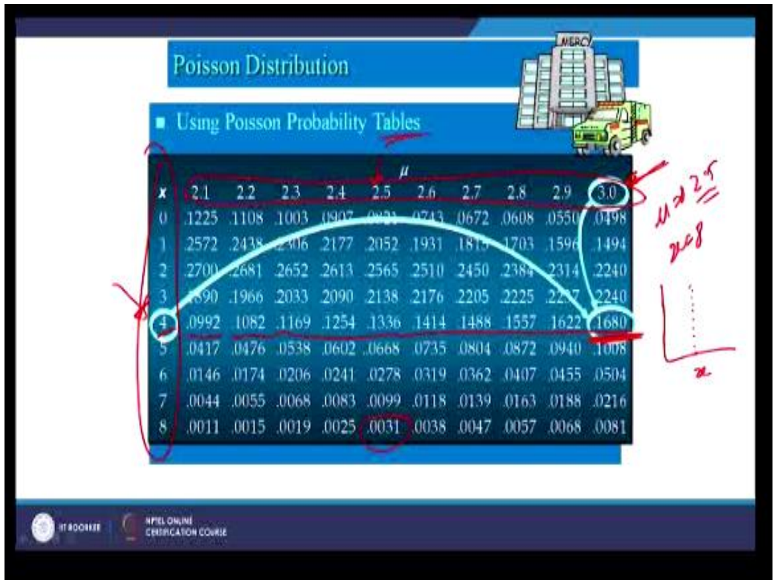
So, let us look at this, so this 6 per hour in other words 3 in 30 minutes time, isn't it. So, this is your mean and axis what is the probability that at the weekend you will have 4 patients ok.

So, $f(4) = \dfrac{3^4(2.71828)^{-3}}{4!} = .1680$

So, e is 2.74, µ is 3 right and divided by factorial x right, so this is x, so you will get answer is 0.16.

So, what was the question the question was what is the probability of having 4 patients arriving on in 30 minutes time in weekend evening, so that probability is approximately 17% ok.

(Refer Slide Time: 05:01)

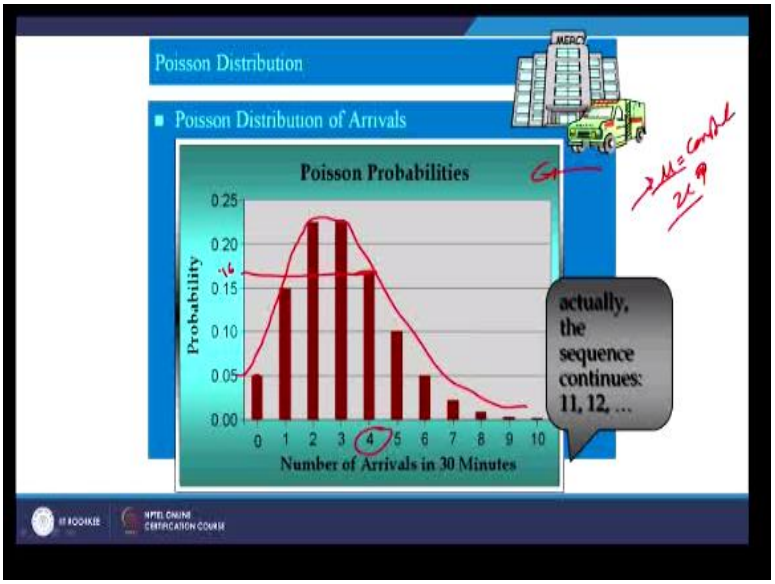


x	2.1	2.2	2.3	2.4	2.5	2.6	2.7	2.8	2.9	3.0
0	.1225	.1108	.1003	.0907	[illegible]	.0743	.0672	.0608	.0550	.0498
1	.2572	.2438	.2306	.2177	.2052	.1931	.1815	.1703	.1596	.1494
2	.2700	.2681	.2652	.2613	.2565	.2510	.2450	.2384	.2314	.2240
3	.1890	.1966	.2033	.2090	.2138	.2176	.2205	.2225	.2237	.2240
4	.0992	.1082	.1169	.1254	.1336	.1414	.1488	.1557	.1622	.1680
5	.0417	.0476	.0538	.0602	.0668	.0735	.0804	.0872	.0940	.1008
6	.0146	.0174	.0206	.0241	.0278	.0319	.0362	.0407	.0455	.0504
7	.0044	.0055	.0068	.0083	.0099	.0118	.0139	.0163	.0188	.0216
8	.0011	.0015	.0019	.0025	.0031	.0038	.0047	.0057	.0068	.0081

Now the same answer can be obtained using poisson distribution tables, so if you look at any book of statistics you will get poisson distribution tables similar to binomial distribution tables. So, here our question is what is the probability of 4 patients arriving during weekend given that mean was 3 right. So, these are x values and these are mean values right, so for mean =3, x=4 probability in this right, this what we obtained using formula.

Let us take the same example let us say mean arrival rate is 2.5 and what is the probability that 8 patients will arrive at weekend evening. So, x=8, so what would be the answer just look at this column 2.5 and 8 is this, this is the probability 0.0031. Similarly you can find out answer to any question for which μ is known and x is known.

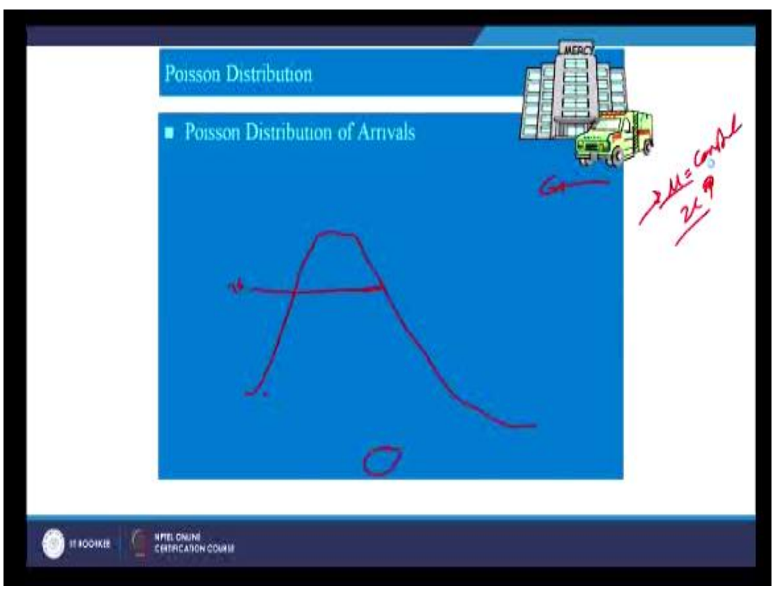
(Refer Slide Time: 06:27)



Now this is how the answer to the question would like if you draw probabilities. So, the probability of 4 patients arriving on weekend this 0.16 right this is 0.16 is not it, probability of 0 patient arriving is 0.05, 10 patients arriving is very near to 0 right. So, there is a pattern over here was the pattern the probabilities first increases right which reaches to the maximum value and then starts decreasing.

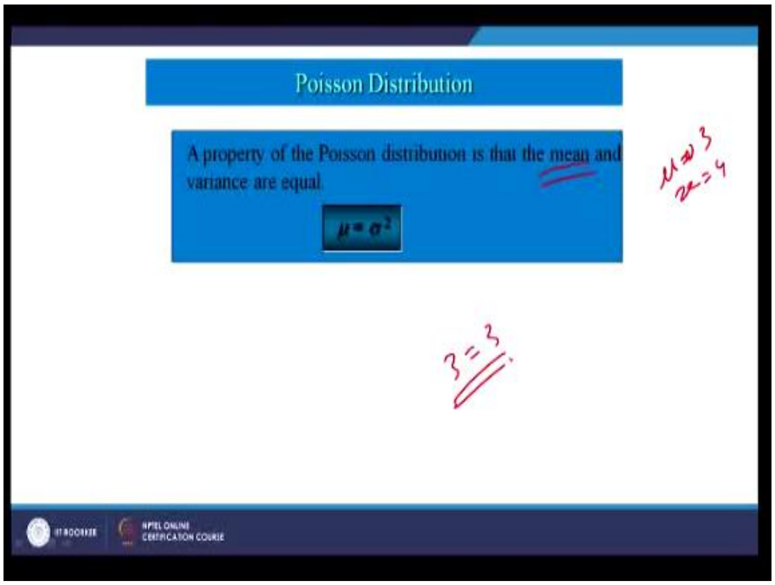
So, when you keep the mean constant and if you change the x if you increase x then this the pattern. So, there is a relationship between how the pattern changes with μ and x. let us say if you keep the x constant and change the μ what will happen.

(Refer Slide Time: 07:50)



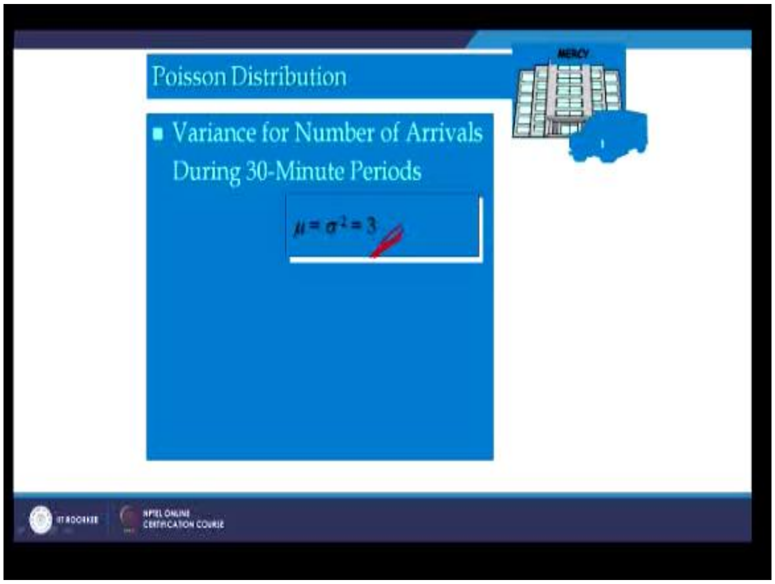
Let us look at the table, if you keep the x value constant let us say 4 and if you change μ let us say initial it is 2.1. So, this is probability when you increase μ 2.2 this is 0.10, so probability has corner right it is going on. So, in this case when you keep the x constant and change the μ or increase the μ the probabilities also increases ok. So, how this the distribution will look like, so this is your x value, let us say which is constant 4. And in μ changes then the probability will also change right, you will get a distribution like this ok.

(Refer Slide Time: 08:57)



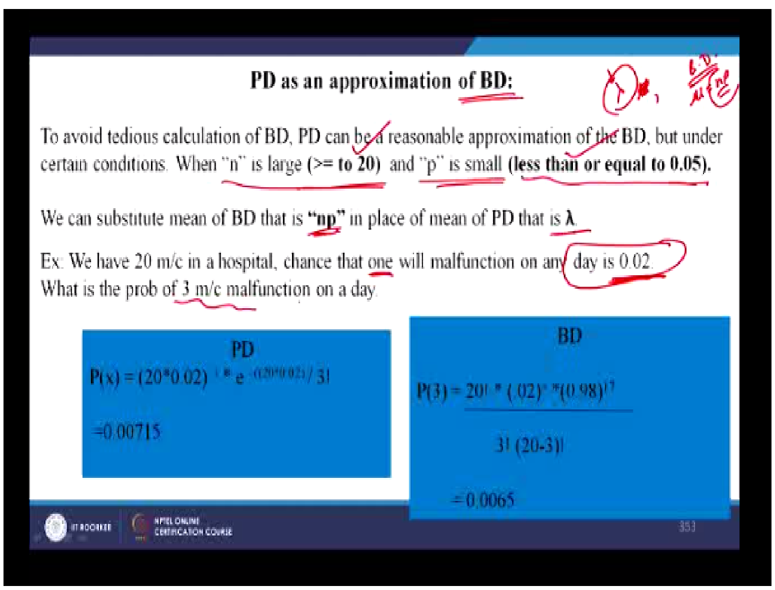
Let us look at next slide, so one of the properties of poisson distribution is that mean and variance are same. So, what would be the answer to the Mercy hospital question which we have just solved, what was the mean there, mean was 3, x was 4 right, so mean is 3 and variance is also 3 right.

(Refer Slide Time: 09:22)



So, this is the answer right, variance for number of arrivals during 30 minutes period is 3 which is equal to mean as well. Now we have seen poisson distribution and binomial distribution but binomial distribution has got very tedious calculation, you need to calculate a lot. You can approximate poisson distribution of binomial distribution.

(Refer Slide Time: 09:54)



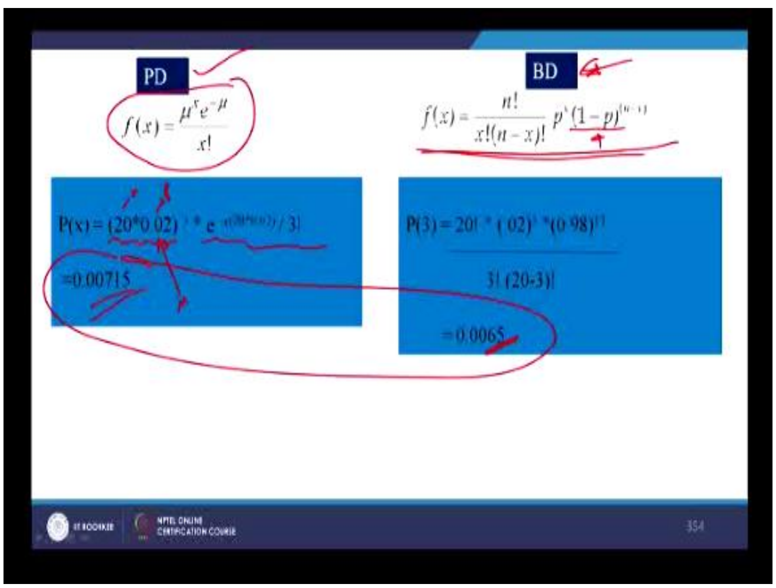
So, instead of using binomial distribution you can use poisson distribution to solve a question. But there are certain conditions to be fulfilled, this can happen only if n is very large value, large means either 20 or more than 20. And p is small value either equal to or less than 0.05, if these conditions meet then only you can approximate poisson distribution for binomial distribution.

So, we know that in Poisson distribution you have got mean and what is the mean of binomial distribution, what is the mean, mean it is np right.

So, if these 2 conditions are fulfilled when we can replace this is mean of Poisson distribution lambda right. So, we can replace lambda/np if we need these 2 conditions right, so this is what I have written over here. So, we can substitute mean for binomial distribution is np in place of mean of poisson distribution that is lambda. Let us take an example in a hospital there are 20 dialysis machines and at a given point of time or on a given day the probability that the machine will not work is 0.02.

Probability that one of the machines not work properly it is 0.02, what is the probability of 3 machines malfunctioning on a day. So, this is this information is known to us on the basis of past data, so what is the probability of 3 machines malfunctioning on a day.

(Refer Slide Time: 11:59)



So, if you have let us say let us look at this, so we know that poisson distribution is this and we know the binomial distribution is this. So, what is the question the probability that there would be 3 machines not functioning on a given day, so μ equal to this np right.

So, $P(\text{x}) = (20 \times 0.02)^3 \times e^{-(20\times 0.02)/3!}$

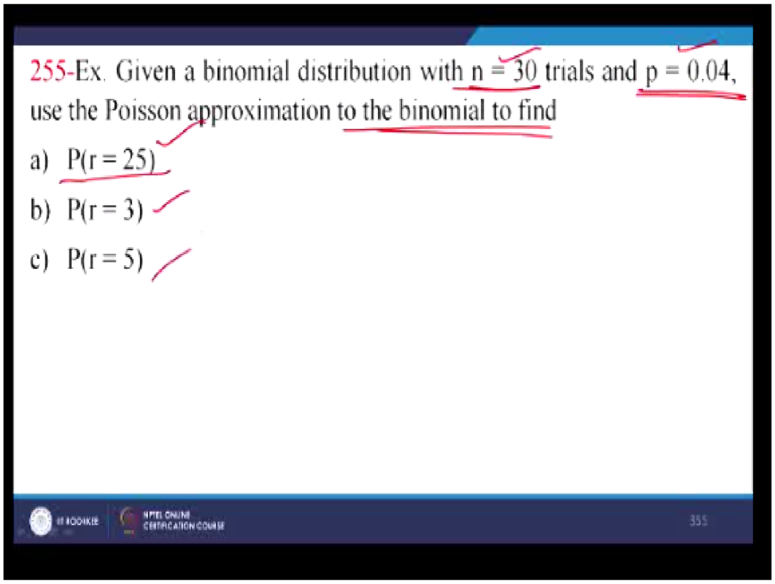
Now the answer is 0.0071, now this same as similar answer can be obtained using binomial distribution. In the absence of poisson distribution we would have used binomial distribution, in what is binomial distribution

$$f(x) = \frac{n!}{x!(n-x)!} P^x q^{(n-x)}$$

So answer is 0.0065 just see almost same answers right. So, what I am trying to say is that you can use poisson distribution instead of binomial distribution to avoid its tedious calculation if two conditions are met.

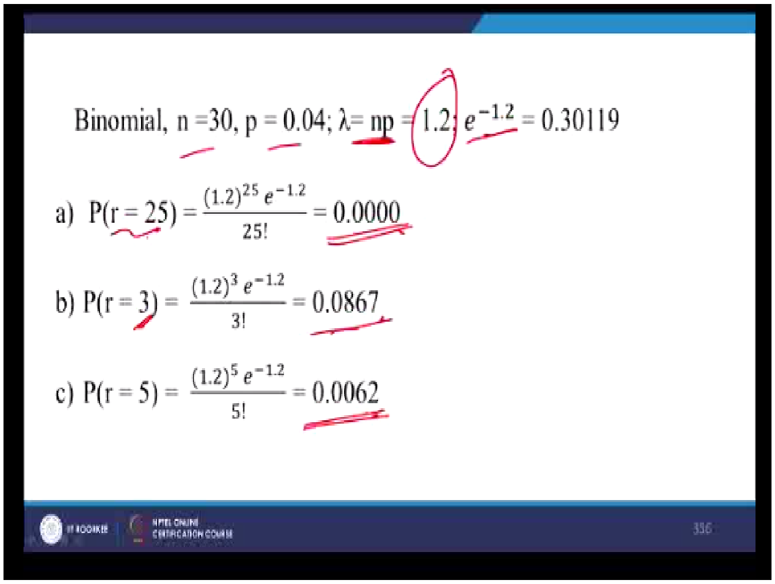
And those conditions are what the n has to be equal to or greater than 20 and p has to be less than or equal to 0.05, let us look at couple of examples.

(Refer Slide Time: 14:13)



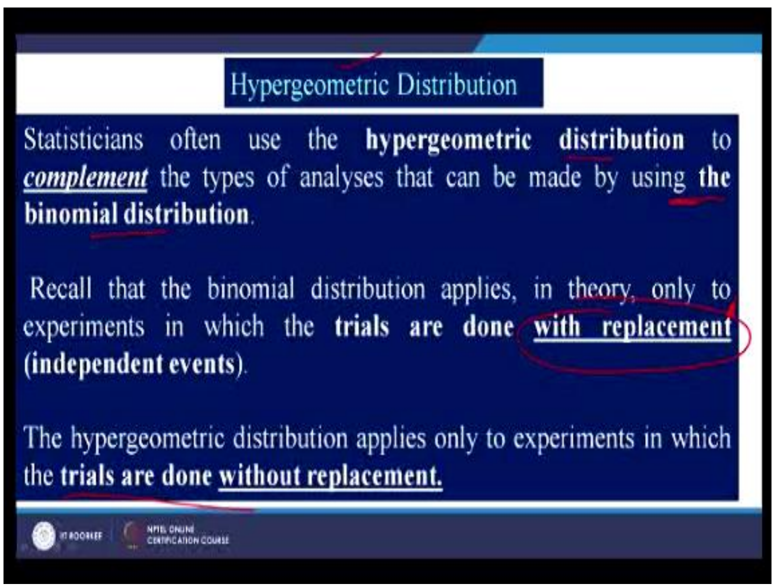
So, for a given binomial distribution n=30, p=0.04 use poisson distribution to find the answer right. Since this is a case where n is large p is small we can approximate binomial distribution as poisson distribution, so what is the probability of success 25, 3 and 5 right.

(Refer Slide Time: 14:42)



So, this is n, p; λ is equal to what n and p right, so this is 30*0.04 this becomes 1.2, so e to the power λ. So, either you call it λ or μ one and the same thing ok. So, answer to the first part is 0 right, so the probability that there would be r value =25 is 0 r is 3=0.08 and r=5 is 0.0062. So, this how you can approximate poisson distribution for binomial distribution.

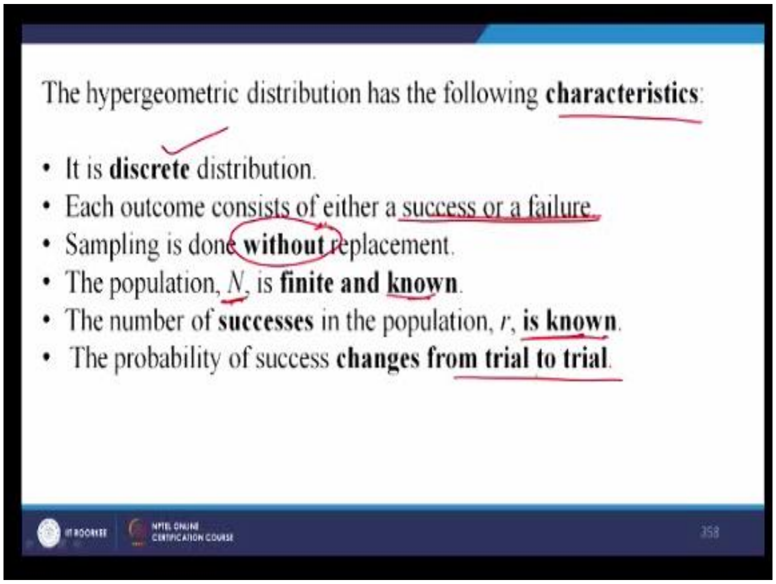
(Refer Slide Time: 15:43)



Let us look at one more distribution this is hyper geometric distribution, now this is somewhat similar to binomial distribution with few changes. So, statisticians often use the hyper geometric distribution to complement the analysis that can be made using binomial distribution. So, we know that in binomial distribution we perform our experiments with replacement ok, so always we perform experiment with replacement.

So, let us say there are 4 chips in a box we will take out 1 chit and next time when we perform the experiment we will keep the chit again and again will take out one more chit or 2 chits whatever is the question. So, here the experiment is are the trails are done with replacement in binomial distribution but in case of hyper geometric distribution they are done without replacement, the trials are done without replacement, so this is the basic difference between these 2.

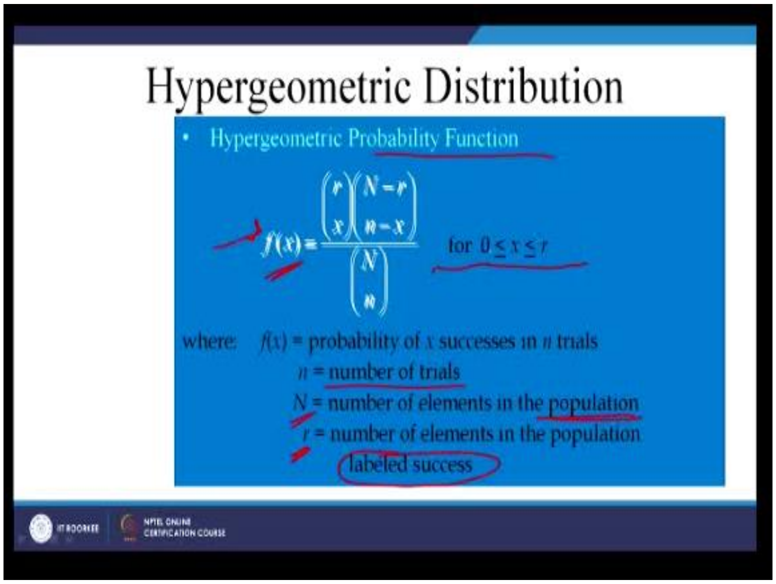
(Refer Slide Time: 17:01)



There are certain characteristics of hyper geometric distribution, so of course it is discrete distribution. Each outcome consists of either success or failures similar to what we have seen in case of binomial distribution. Sampling is done this is the only difference or this is one of the differences right, the population is finite and known. The number of success in probability is known and the probability of success changes from trial to trial while in previous case in case of binomial distribution this remains fixed ok.

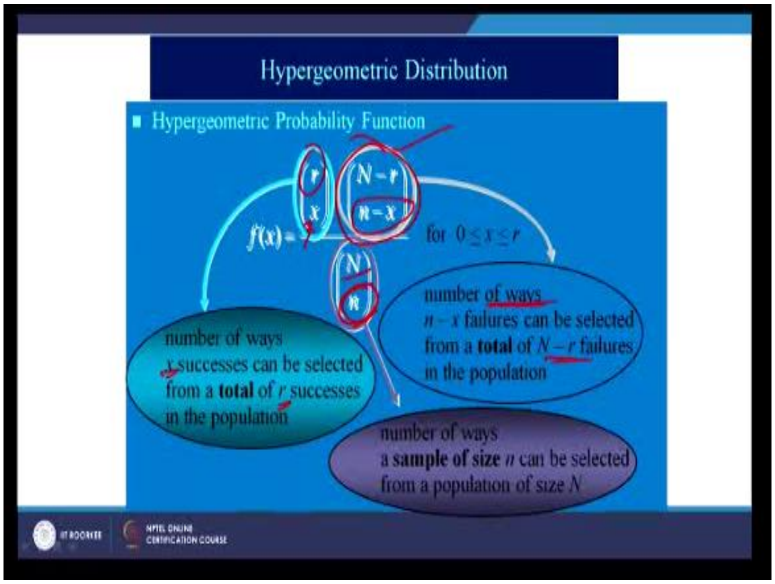
So, these are couple of characteristics of hyper geometric distribution, so if you look at hyper geometric distribution the probability function is given by $f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$

(Refer Slide Time: 17:56)



So, small n is number of trials capital n is the population total population or number of elements in population, r is number of elements in the population labeled as success. So, this you always try to find out r ok and for this probability function is valid for x = x is ranging between 0 and r.

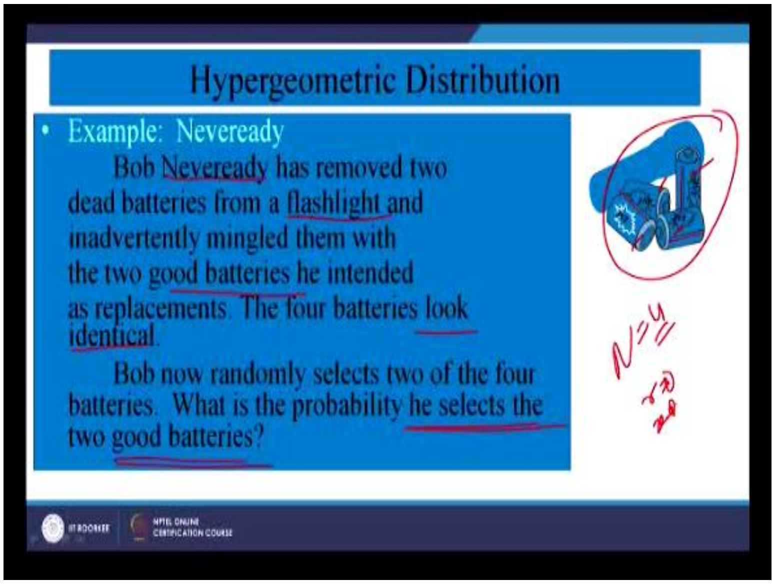
(Refer Slide Time: 18:42)



Now if you look at this is nothing but the number of ways x successes can be selected from total r successes. So, there are total r successes and we have to find out number of ways of x successes similarly this portion of the formula deals with the N-r failures out of n- x total failures right. So, the number of ways n-x failures can be selected out of N-r failures, similarly number of ways of sample size n which can be selected out of n.

So, all these 3 are you need not remember these points right, so we want to select x successes out of r. we want to select n-x failures our of N-r you want to select small n emphasize out of capital N right.

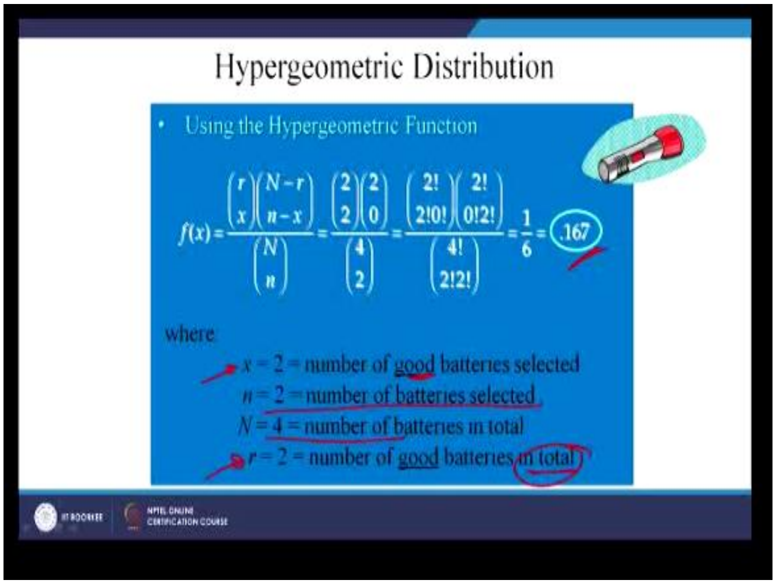
(Refer Slide Time: 19:59)



Let us look at an example on hyper geometric distribution, Bob Neveready has removed 2 dead batteries from a flashlight this is flashlight. And he inadvertently mingled with the 2 good batteries, he intended as replacement, the 4 batteries look identical. So, you can just think of these are 4 batteries they look alike, so 2 of them are dead batteries and 2 of them are god batteries.

Now Bob randomly selects 2 of the 4 batteries, what is the probability that he selects 2 good batteries right. So, that is the question. Now you can solve this question, so if you look at how many total batteries N=4 right is not it. So, you need to find out what is r, what is x is not it, what is N-r and n-x is not it.

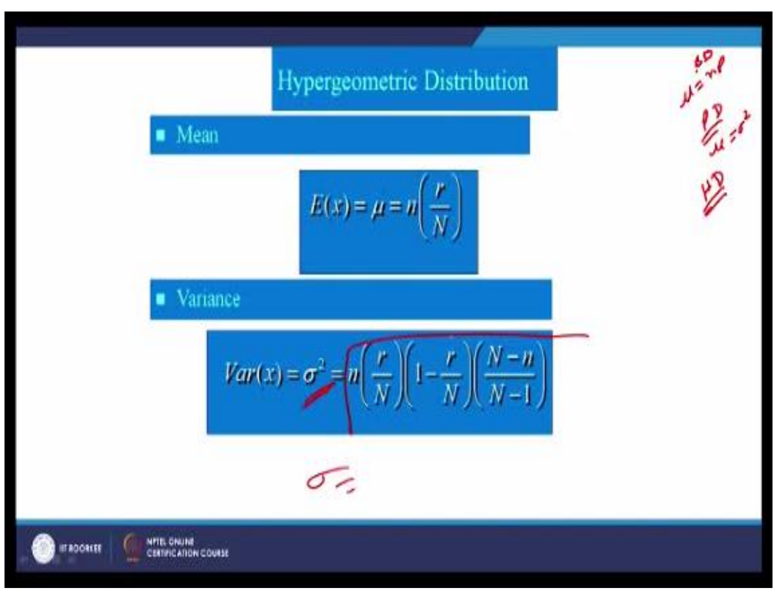
(Refer Slide Time: 21:06)



So, let us look at this, so x is number of good batteries right, r is number of good batteries in total. So, we have to select this ^rCx right, so this is number of good batteries selected, number of good batteries in total. Of course small n number of batteries selected, so you the Bob has selected 2 batteries out of 4 batteries. So, if you put all these values in this formula you will get answer = 0.167, so this how you can solve the question like this.

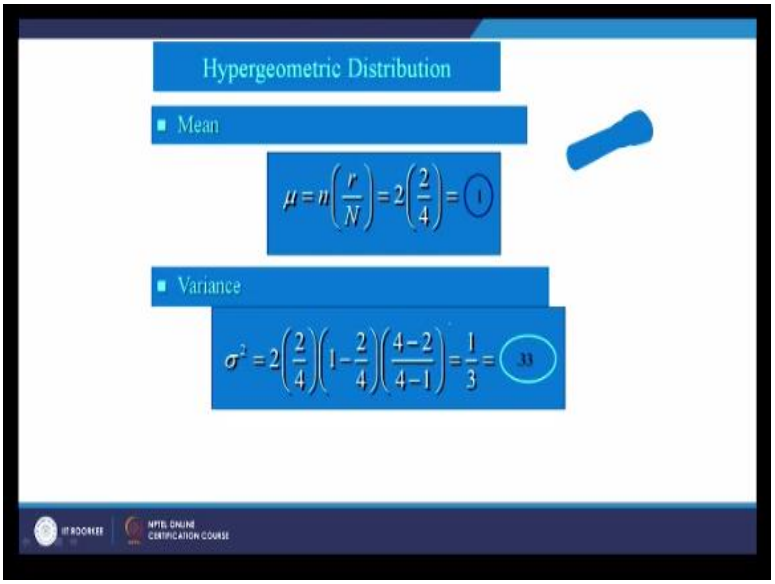
Now the mean of hyper geometric distribution, so we have seen mean of poisson distribution, mean of binomial distribution.

(Refer Slide Time: 22:10)



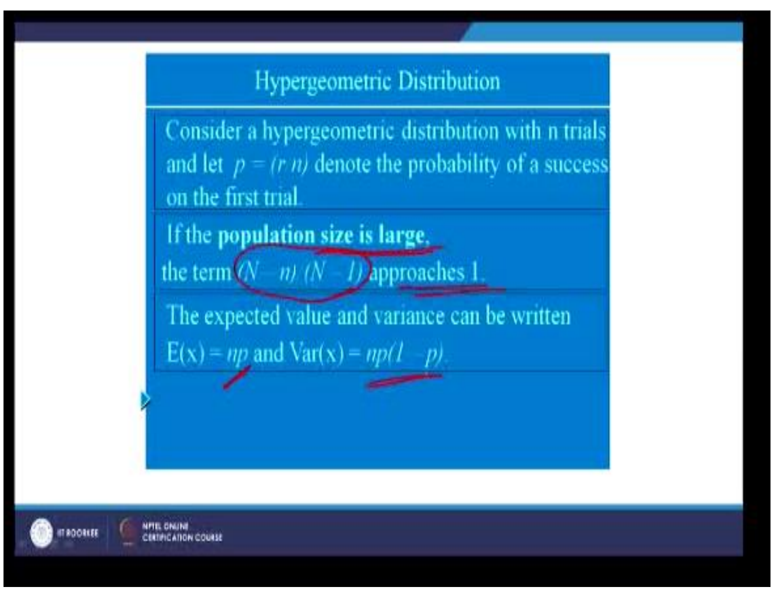
So, for binomial distribution what was the mean np right, for poisson distribution what was the mean. Mean and variance were same for poisson distribution for hyper geometric distribution mean is this right, it is $^{r}C_{N}$ right*n similarly variance. So, if I ask you to calculate standard deviation what would you do you just standard derivation would be this value is not it just take positive square root of it right.

(Refer Slide Time: 22:46)



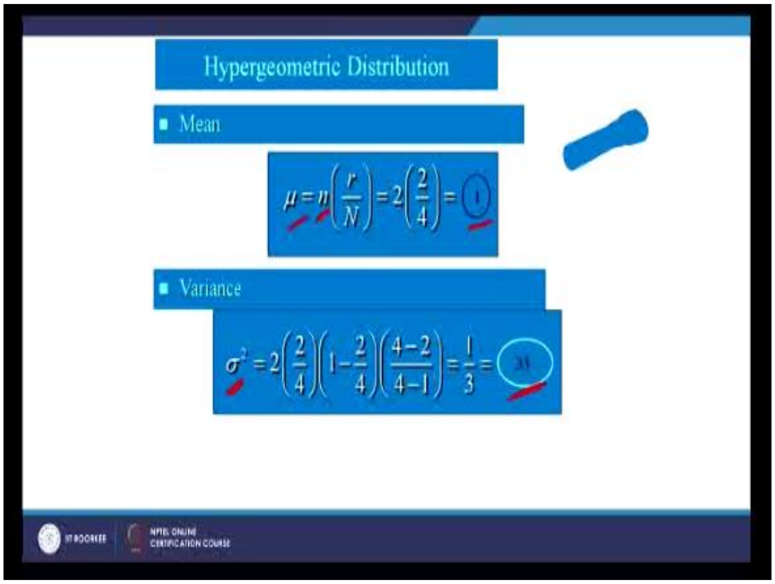
So, for the question which we have taken up wherein there were total 4 batteries 2 are dead batteries, 2 are good batteries is not it. So, mean is 1 is n and $^{r}C_{N}$ right, this is variance is 0.33.

(Refer Slide Time: 23:09)



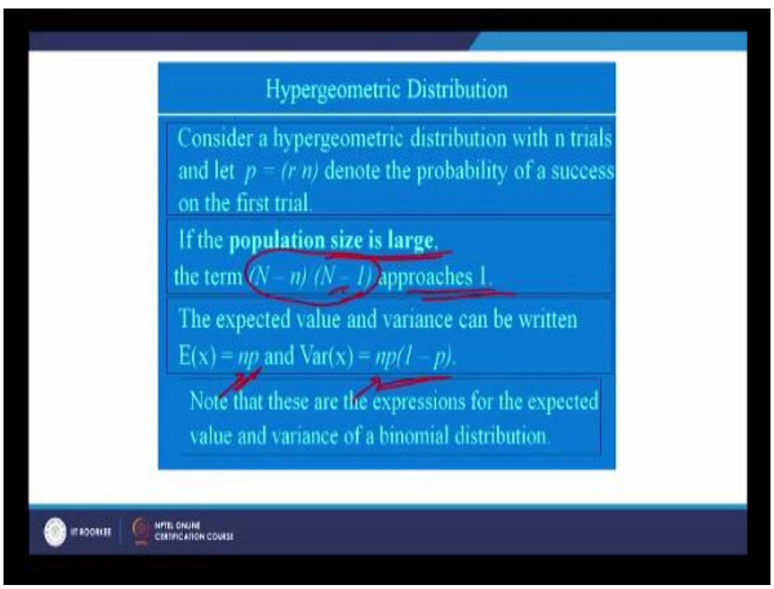
Now there are couple of features of geometric distribution, now if the population size is large then this term in the probability function u approaches 1. And in that case the mean is this and variance is this otherwise what was the formula for mean.

(Refer Slide Time: 23:39)



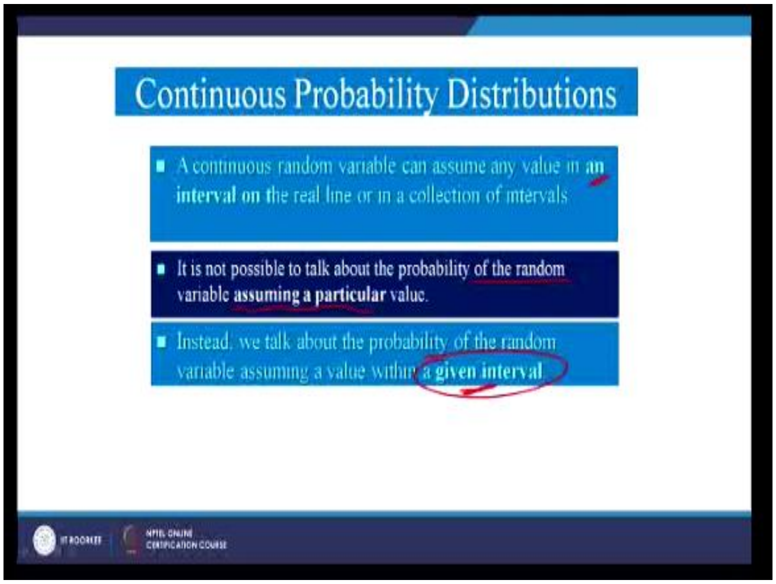
This was mean and this was variance right.

(Refer Slide Time: 23:51)



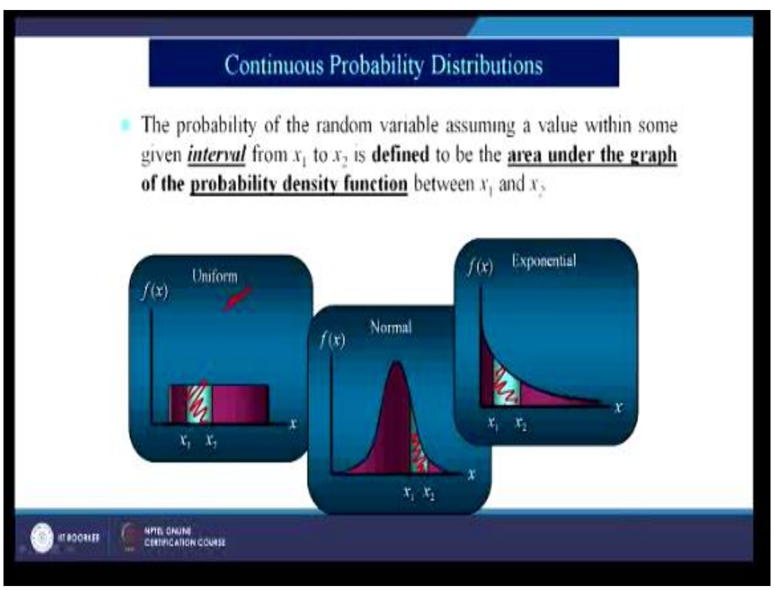
So this formula. These 2 formulae change when you change or when your population size is large then this term becomes 1 right and the formula changes to this for mean and variance. So, for we discussed about discrete probability distributions, let us look at continuous probability distribution. Now there are several continuous probability distributions we will see few of them. So, what is continuous probability distribution?

(Refer Slide Time: 24:29)



Continuous random variable can assume any value in interval on the real line or in collection of intervals. So, it is not possible to talk about the probability of the random variable assuming a particular variable value but instead we talk about the probability of the random variable assuming a value within given range or interval. So, we do not talk about the probability of random variable assuming a particular value but the probability of random variable assuming a value within given interval right.

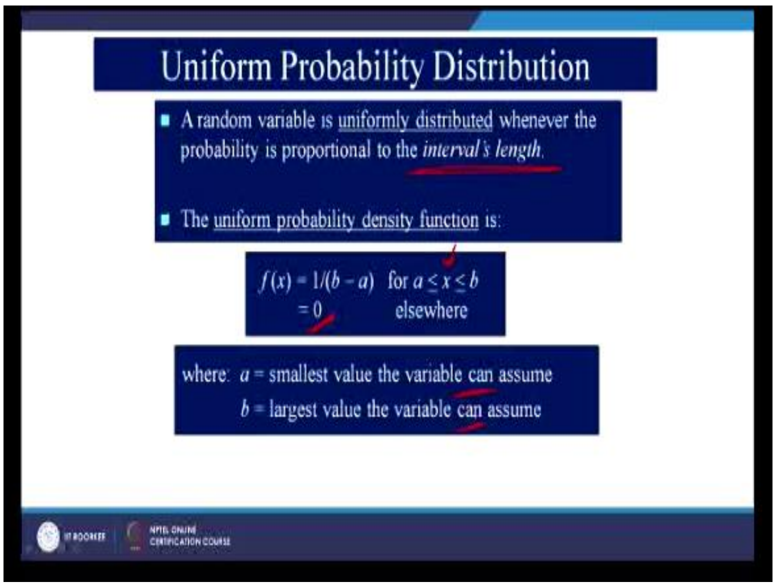
(Refer Slide Time: 25:19)



So, the probability of random variable assuming a value within some interval let us x1 to x2 is defined as the area under the graph of the probability density function between x1 and x2. So, these are different continuous probability distributions you have got a uniform probability distribution, so what is the probability between x1 and x2 is the area under the graph of the probability density function between x1 and x2 right.

Similarly, for normal distribution this is the area under curve and this for exponential distribution. So, we will see one by one first of all we will look at uniform distribution.

(Refer Slide Time: 26:14)



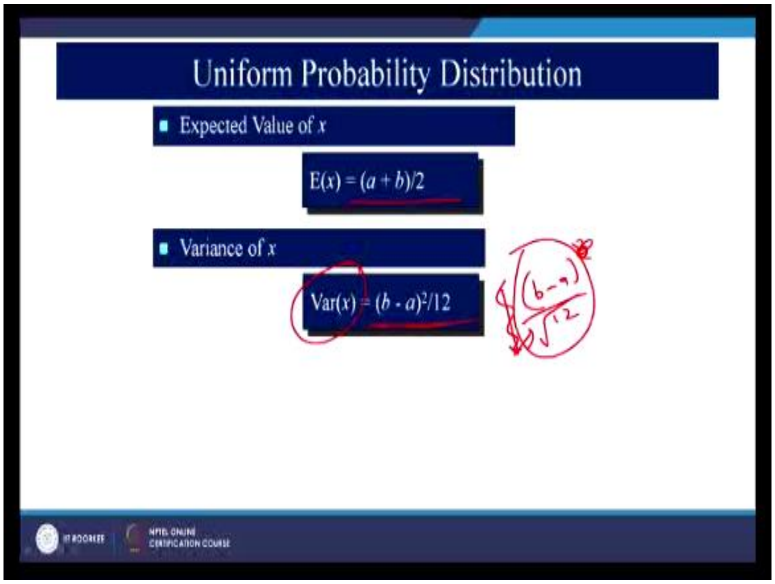
A random variable is uniformly distributed whenever the probability is proportional to the intervals length.

So, the uniform probability density function is $f(x) = 1(\mathrm{b-a})$ for $\mathrm{a \le x \le b}$

= 0 elsewhere

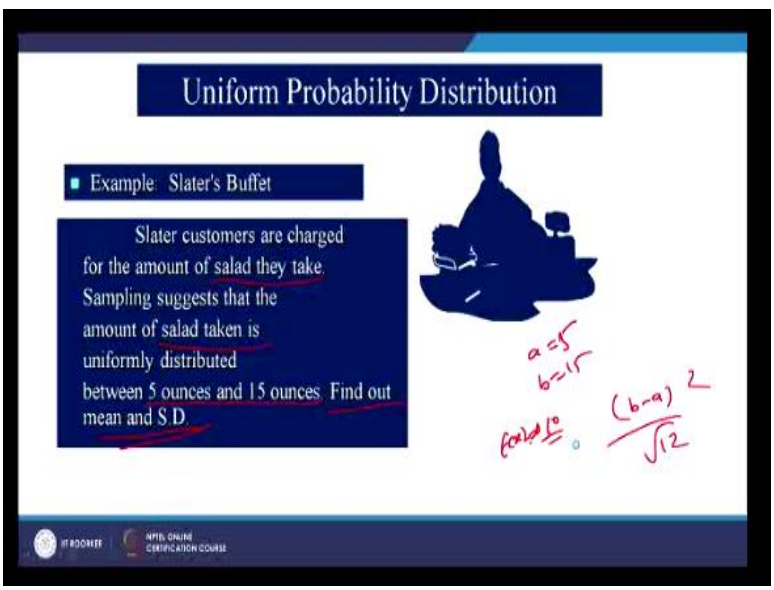
So, if x is not between a and b then f of x is o, so a is smallest value of the variable which the variable can take and this is the maximum value. Let us work out couple of examples.

(Refer Slide Time: 27:01)



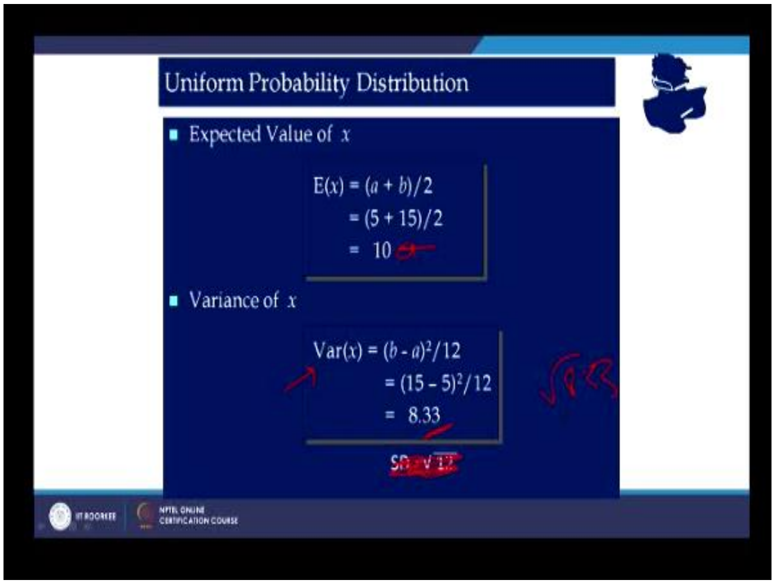
So, the mean of uniform probability distribution is (a+b)/2 and variance is (b-a) whole square/12. So, this variance if I ask you to calculate standard deviation then it becomes b-a whole square like this or you can simply remove this and this is under root of 12 is not it. So, this becomes standard deviation, so mean is this and standard deviation is this right.

(Refer Slide Time: 27:31)



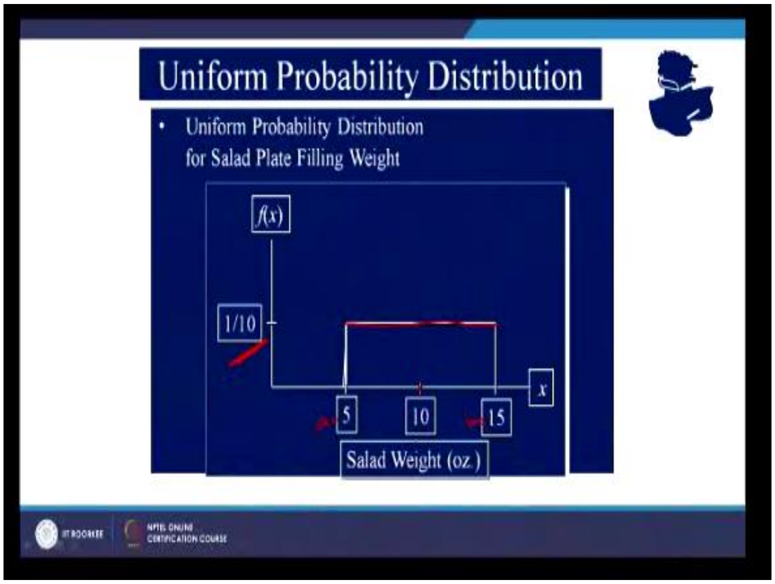
Let us take the example Slater's Buffet example, so Slater customers are charged for amount of salad they take. Sampling suggest that the amount of salad taken is uniformly distributed between 5 ounces to 15 ounces, find out mean and standard deviation for this question. So, mean is, so a is 5, b is 15 mean (a+b)/2, so mean is right, so this is mean. Of course standard deviation would be $(b-a)^2/\sqrt{12}$ that would be standard deviation.

(Refer Slide Time: 28:26)



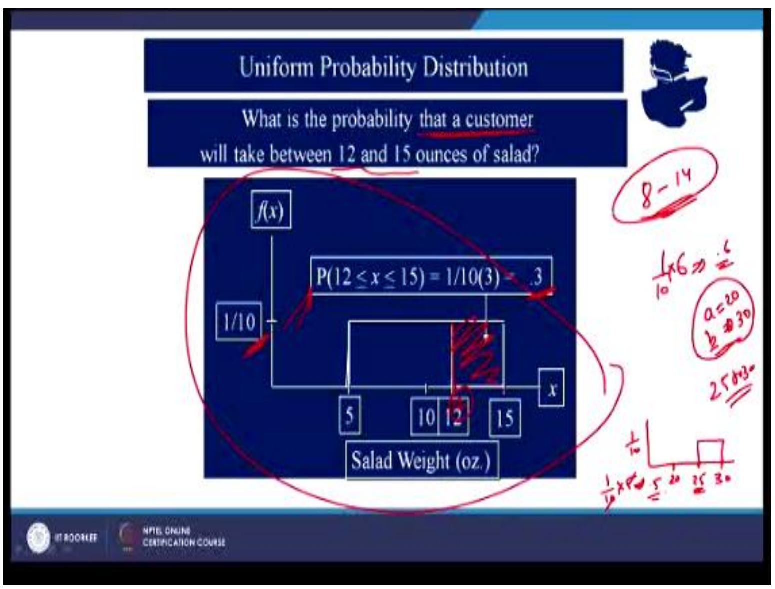
So, f of x is 1 upon b-a right, so this becomes 10, 15-5 is 10, so x is salad plate filling weight ok in terms of ounce. So, mean sis 10 this how we have already calculated let us look at variance, so $(b-a)^2/\sqrt{12}$ is the variance it is 8.33 right. And I have already talked about variance would be what on this variance and standard deviation would be under root of 8.33 right, this is just keep this right ok. So, this how you can calculate the mean and variance of uniform probability distribution.

(Refer Slide Time: 29:42)



So, we will continue with the same example, so this is f of x is 1/10, so this is that value 1/10, the value is 5, b value is 15 and this is mean 15 right. Now the question is this, now the second part of the question is the first part was we calculated mean and standard deviation.

(Refer Slide Time: 30:18)



Now the second part is what is the probability that a customer will take between 12 to 15 ounces of salad. Now this is what we have already worked out we have to find out what is the probability that he will take between 12 to 15 ounces. So, 12 is somewhere right is not it, so this 12 ok, so we have to find out this area and will multiply that value by 1/10 right. So, this 12 ok, so probability that a customer will take between 12 to 15 ounces of salad is 0.3 right.

So, this is 15-12 is 3, 3*1/10, so it becomes 0.33 ok, so if I ask you what is the probability that a customer will take between let us say 8 to 14 what would be the answer. So, we know this is 1/10 is not it and between 8 to 14 it means 6, so this is multiplied by 6 that is 0.6 isn't it. So, this very simple question, so if you are given a and b values you can easily calculate the probabilities for any given range.

Let me put the same question in different way, let us say you are waiting at the airport and for a particular flight you have observed that the mean the waiting time ranges from 20 minutes to 30 minutes right. Now the question is what is the probability that a passenger has to wait between let us say 25 to 30 hours. So, how would solve this question, so this is very simple you have got 20 this 30, so mean would 25 is not it and this is 1/10 again is not it the 1 upon b-a.

So, 1/10, so probability between 25 and 30 minutes would be 5, so 1/10*5, so this becomes this is ½ this is 0.5 right. So, this is how you can apply this formula to different situations. Now let us look at the most widely used continuous distribution, so before going for normal probability distribution let me summarize what we did so far. We have seen discrete distributions like binomial, poisson, hyper geometric and we have seen continuous distribution that is uniform probability continuous distribution.

Before taking normal probability distribution you should have thorough knowledge of discrete distributions. As far as normal distribution is concerned it can be applied to variety of situations, so we will discuss about normal probability distribution in next session, thank you.