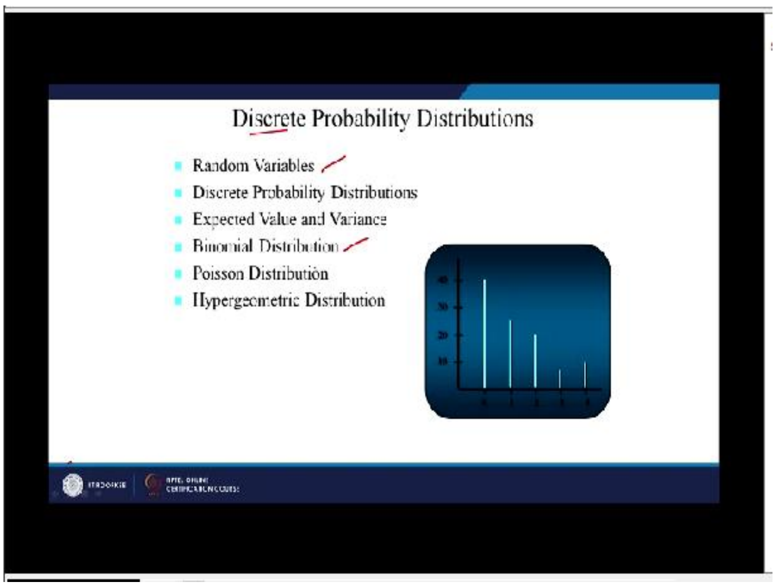
Business Statistics
Prof. M. K. Barua
Department of Management Studies
Indian Institute of Technology-Roorkee

Lecture-16
Probability Distributions-part I

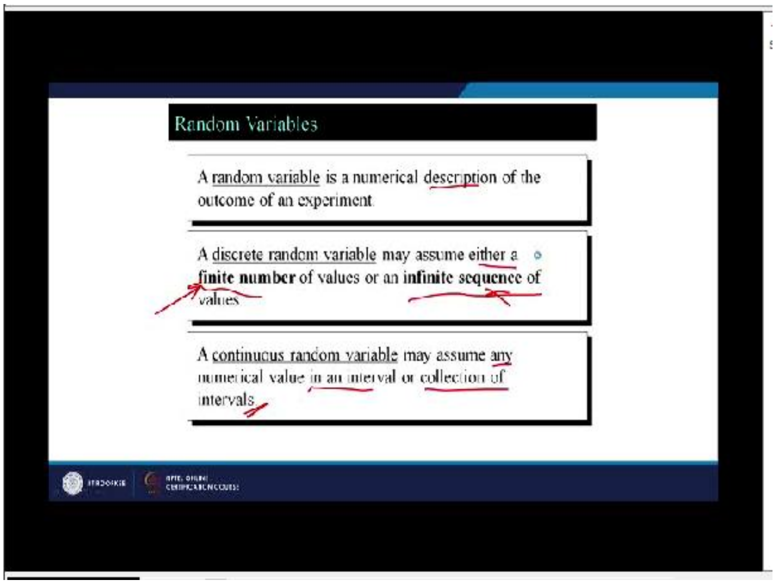
Good afternoon friends as you are aware in previous class we discussed different types of probabilities and we worked out couple of examples as well. So in today's session we are going to talk about probability distributions, so largely you have 2 types of probability distributions; you got discrete and continuous. And in discrete distributions you have got again different types of discrete distribution and in continuous as well there are different types of distributions.

(Refer Slide Time: 01:04)



So, let us start with this particular topic which is on discrete probability distributions, so will see different types of discrete probability distributions apart from definition of random variables to different types of discrete distributions like binomial, poisson, hypergeometric distribution.

(Refer Slide Time: 01:24)

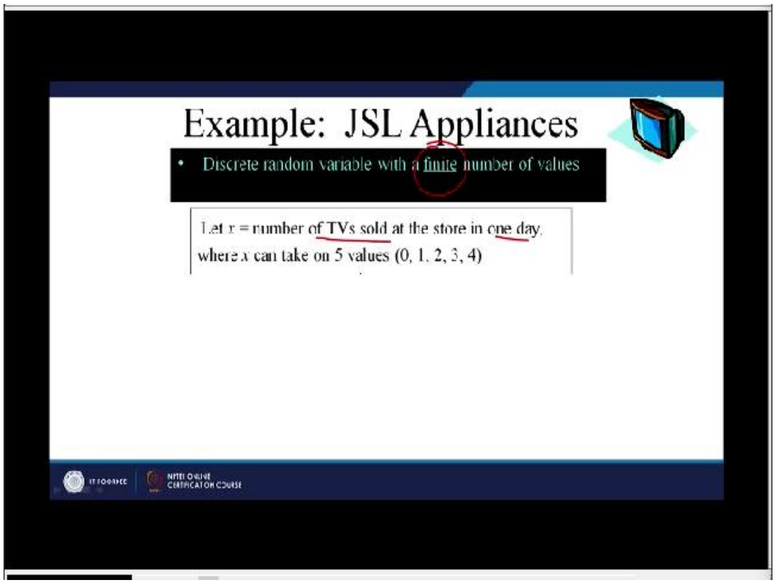


So, let us look at the definitions random variable is numerical description of the outcome of an experiment, so if someone asks you what is the random variable? Then you can say that a random variable is a variable which is, in which variable can take any random value as an output of any experiment. So, just to give you an example, let say if you throw a die, so what is the probability that on surface 5 dots appear?

Now it just 1/6 right and we do not how many dots will appear, so either 1, 2, 3, 4, or 4, 5or 6 right. So, it may take any value right, so you can have again discrete random variable and continuous random variable. So discrete variable may assume either a finite number or infinite number. So if there is an upper limit then will say it is a finite discrete random variable otherwise infinite ok.

And you can have a continuous random variable as well where any numerical value is value in an interval or collection of intervals, so that is a continuous random variable, will see more in more these and definitions in detail.

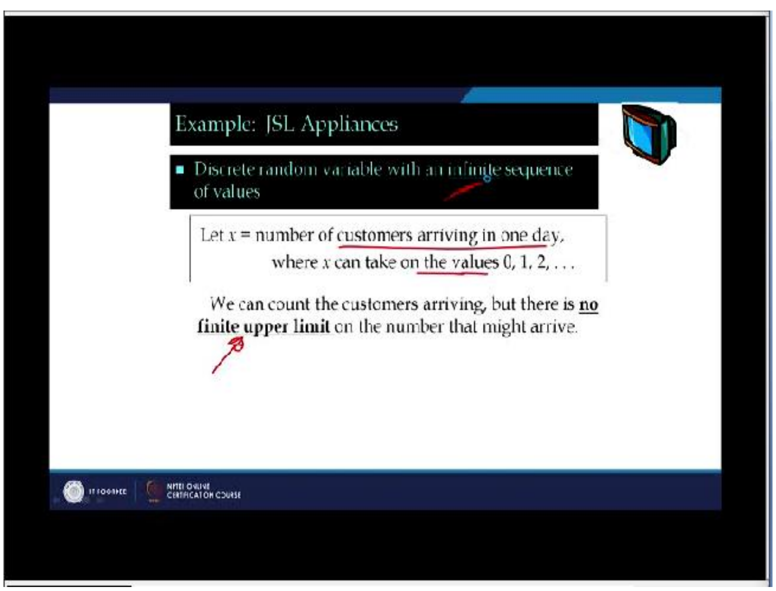
(Refer Slide Time: 02:53)



So, let us look at what is a discrete random variable which is finite, now in this case as I said it is finite, so it is infinite right. So finite means there is some upper limit, so, let say x number of TV sold at a store in 1 day, so you will have some number right 1, 2, 3, 100, 200, 500 and so on right that is finite. So, let say number of homes in a city, let us say you can have 500 homes, 600 home, 651 and so on right.

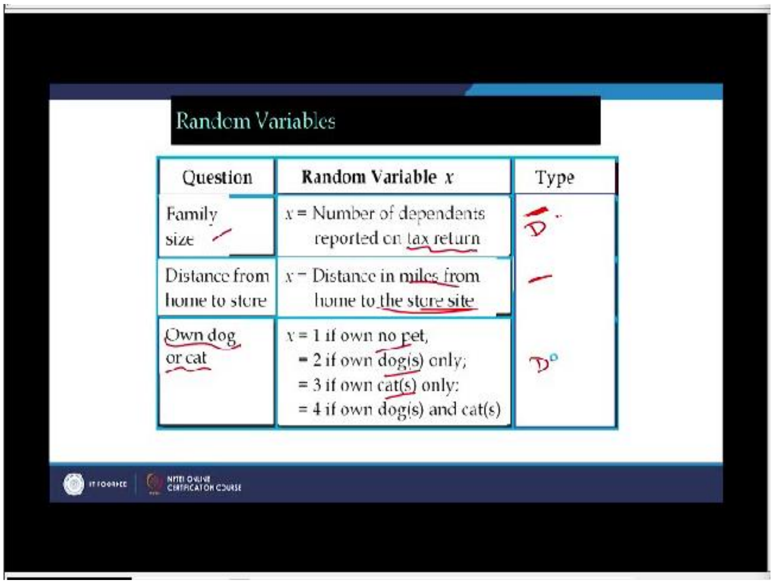
So that is a finite number of values ok, so let say number of pet animals in a house it can be 1, 2, 3 is not it, so it should be a finite.

(Refer Slide Time: 04:00)



When I say infinite so let say customer number of customers arriving in one day at a TV shop, so X can take value 0 1 2 3 and so on right. So, there is there is no upper limit, so will call it infinite is not it, so this is a case of discrete random variable which is infinite

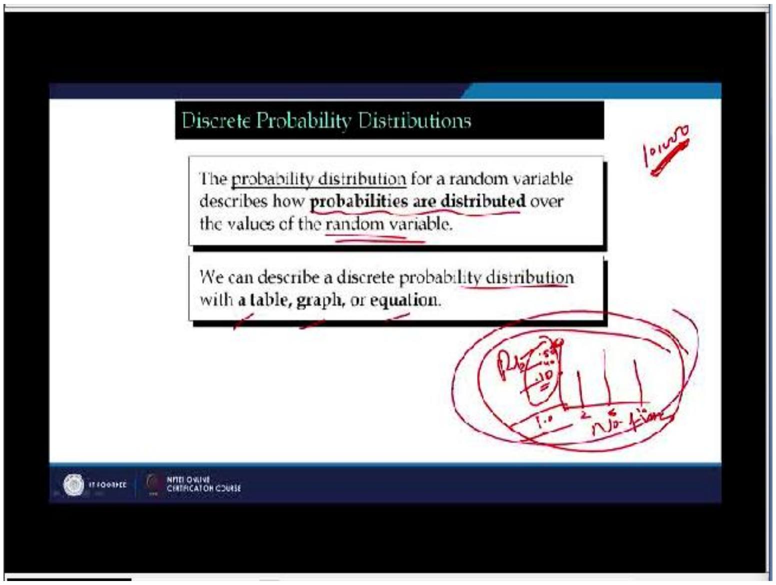
(Refer Slide Time: 04:23)



Next look at this example, so family size is what type of variables, it is finite or infinite, so number of dependent reported on tax return. So, can you have this infinite or first of all you tell me whether it's a discrete or continuous right. So, this is basically discrete type right, distance in miles from home to store site, now the distance can have some non-integer value as well, distance can be 30.33 kilometer is not it.

So, this is not a discrete right let say the number of dogs are gets owned by a family right, so 1 pet right 2 dogs, 3 cats and so on. So, what is this again a continuous or discrete it is to be discrete on right, so let us look at probability distribution what is probability distribution.

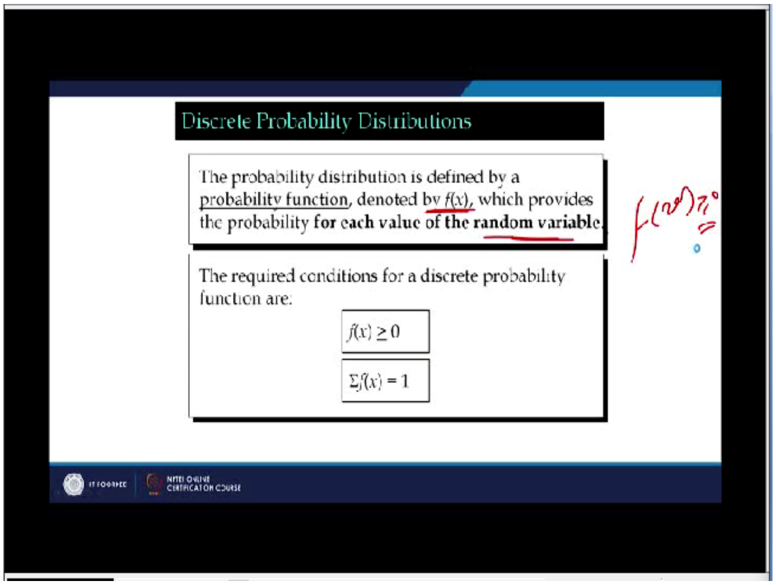
(Refer Slide Time: 05:53)



Probability distribution is basically representation of probabilities of some variable, so how probabilities are distributed over values of some random variable. Let me give you an example, let say you are contesting for an election. And let say there are 10,000 voters in that particular city or in that particular district ok, so the total number of voters are 10000. Now you want to know how many votes you would get and the probability is associated with those number of votes.

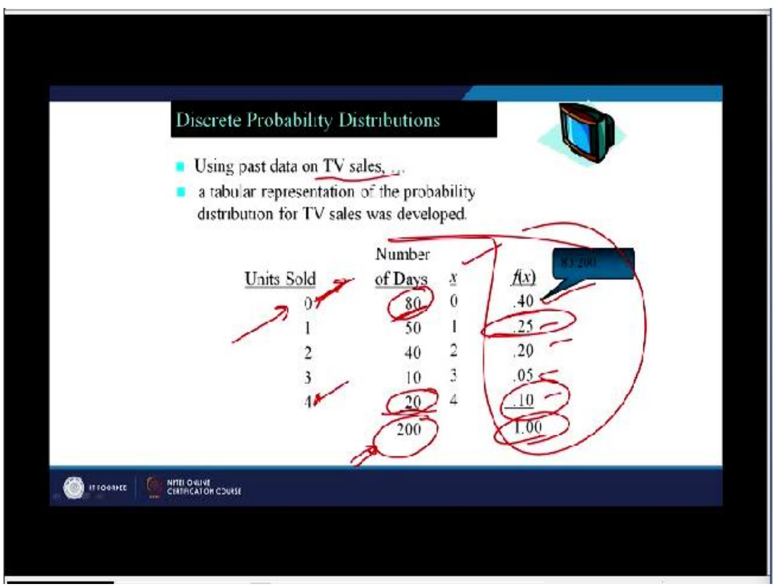
So let say these are number of votes on x-axis right, on y-axis you got probabilities right ok. So let say you will get 2000 votes means probabilities 0.1, 0.10 you will get nexus 6000 votes probabilities 0.50 and let say you will get 10000 votes its probabilities 0.40. So, this is nothing but a probability distribution ok, keep in mind that the sum of these probabilities should be 1. In this case 0.5, 0.4, and 0.1 this is equal to 1. So we can describe a discrete probability distribution with table graph or equation rights, so this is a graph is not it.

(Refer Slide Time: 07:52)



The probability distribution is defined by a probability function, so you can denoted by f(x) which provides the probability for each value of selected random variable. So, it can be either 0 or more than 0 or equal to 1. But it can never be less than 1 ok.

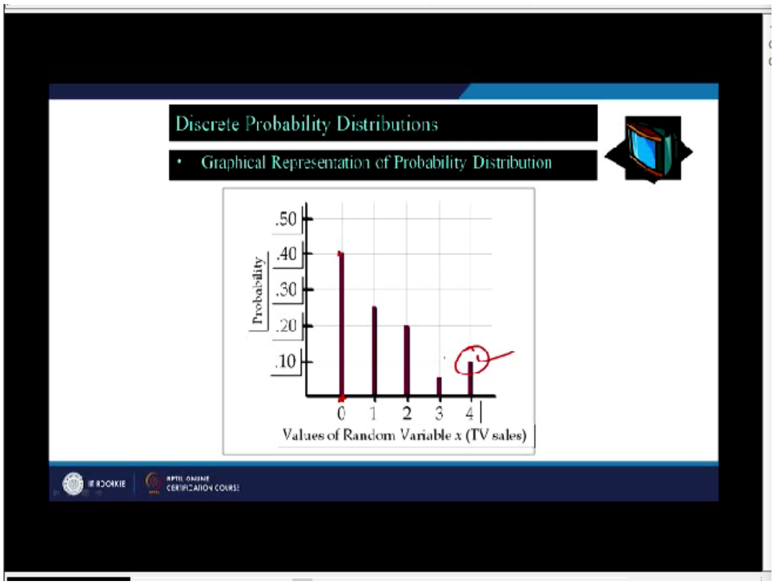
(Refer Slide Time: 08:23)



Let’s look at this example, so there is a TV retail outlet and the owner of the shop has collected some data over number of days, the number of units, the number of TV units sold right. So the shop owner collected data of last 200 days and he found that on 80 days the number of units sold were 0. On 20 days, he sold 4 number of units, so this is x number of units sold and f(x) is can be calculated by just dividing this 80 by this total number of days.

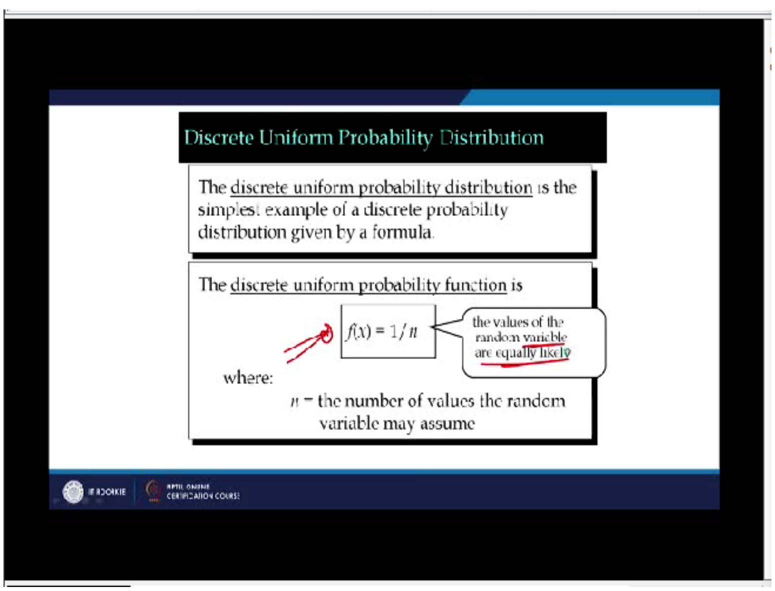
So, this is 0.40 is the f(x) similarly 0.25, 0.20, 0 5and 0.10, so this total has to be 1 right. So, this is nothing but a probability distribution right. So what the probability that 0 unit of TV sold was 40% right and so on.

(Refer Slide Time: 09:53)



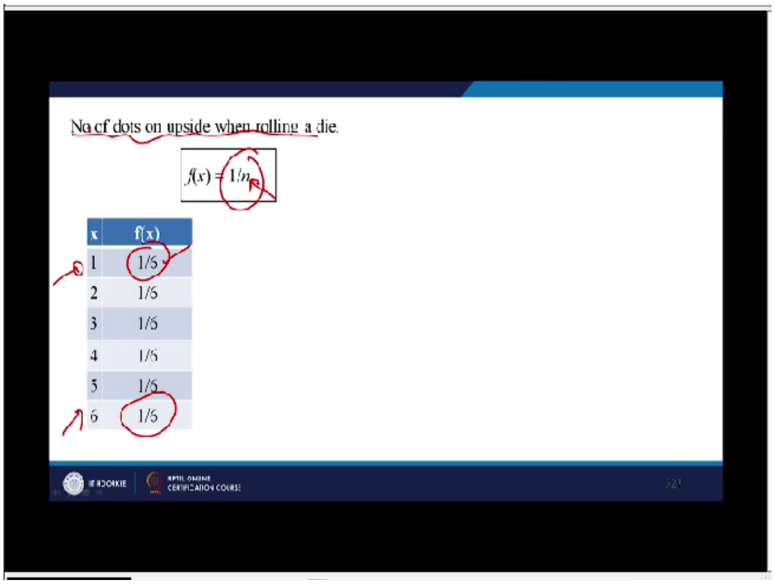
So, the same information can be put in this graphical representation form, so, 0 unit of TV sold 0.40, then 1 unit previous slide for 1 unit it is 25%, for 4 units it is 10% right. So you can see all this is nothing but a probability distribution, discrete probability distribution.

(Refer Slide Time: 10:32)



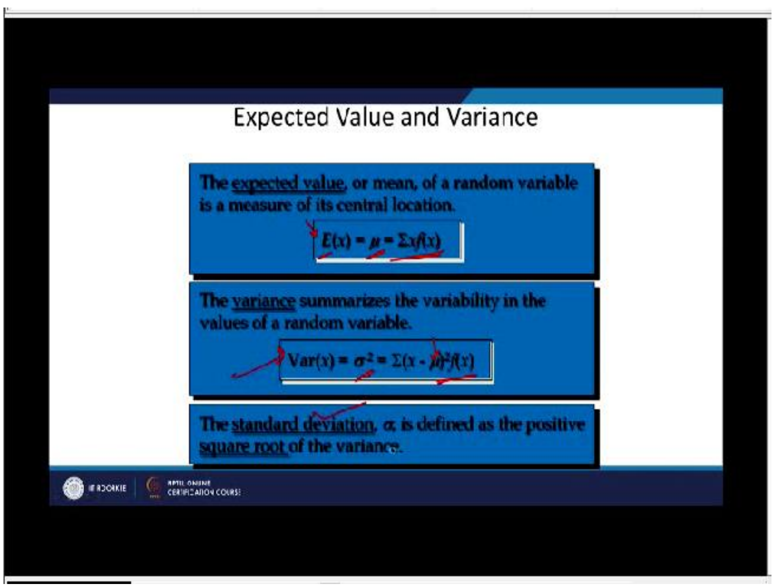
So let us look at uniform probability distribution, uniform probability distribution is simplest of the discrete probability distributions. So this is how you represent uniform probability distribution f(x) = 1/n, here n is the value of random variables which are equally likely ok.

(Refer Slide Time: 11:05)



So, let us take an example, so you want to know the number of dots, number of dots on upside when you roll a die. So, f(x) is 1/n is not it, so n can be either 1, 2, 3, 4 or 5 or 6 right, so there are 6 possibilities right. So, this 1/6, so you can get 1 dot its probabilities 1/6 again you will get 6 dots on its surface again it is 1/6 right.

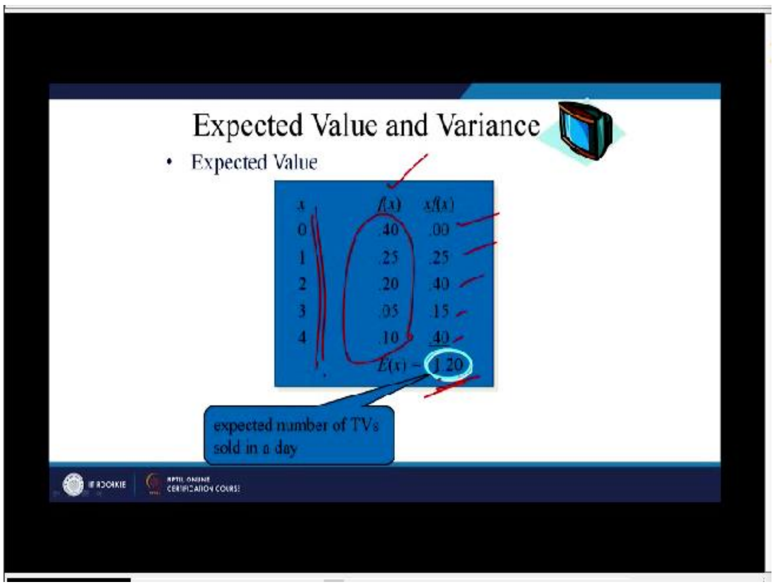
(Refer Slide Time: 11:40)



So, this is nothing but uniform distribution let us look at expected value and variance , expected variance nothing but it is mean value of random variable which measures central tendency. So this is E(x); E is stands for expected value is equal to mean and this is E(x) summation of x*f(x) and variance is this the variance. So $(x-\mu)^2 \times f(x)$ and what is a μ, μ is mean right.

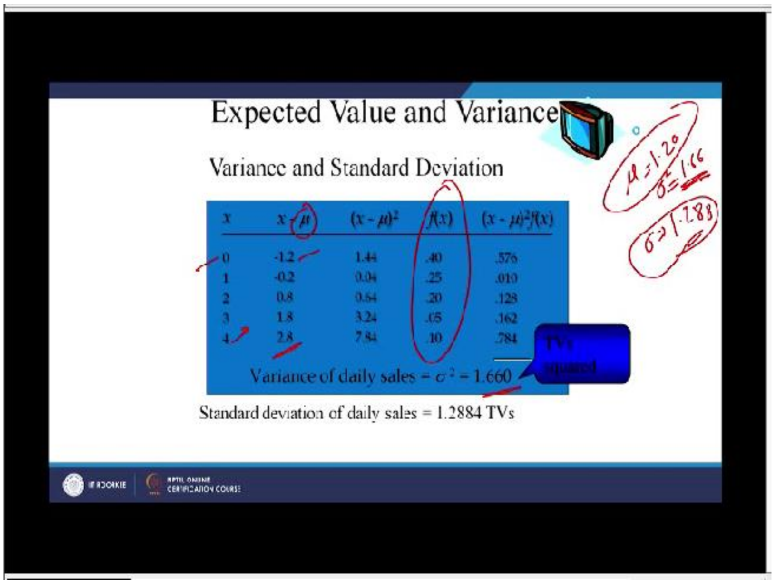
So, once you know variance you can easily calculate standard deviation by take positive square root of variance. So, will calculate all these values, so for the same example right.

(Refer Slide Time: 12:31)



So, this what we have seen our example, so f(x) is given for these number of units sold and this total is 1 right is not it. So, x*f(x) is 0.25, 0.40 and so on right. So this total is 1.20 right, so E, this is the mean value right 1.20 right.

(Refer Slide Time: 13:03)

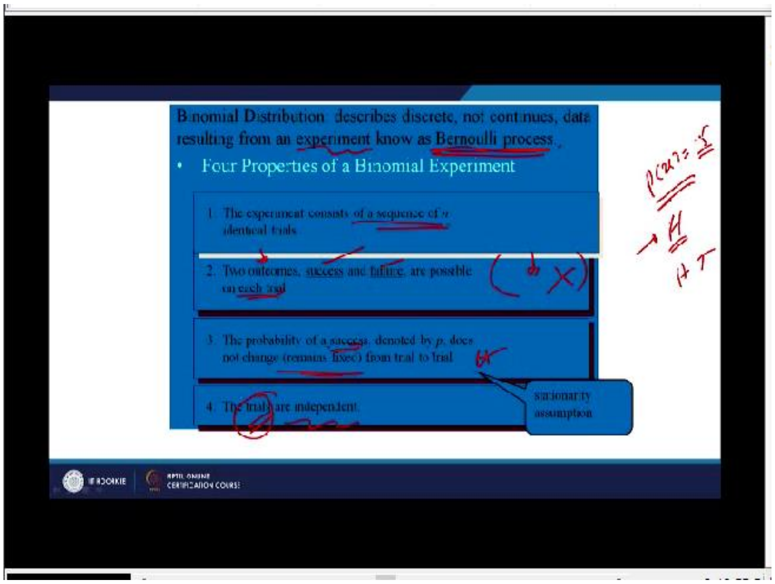


x	x - μ	(x - μ)²	f(x)	(x - μ)²f(x)
0	-1.2	1.44	.40	.576
1	-0.2	0.04	.25	.010
2	0.8	0.64	.20	.128
3	1.8	3.24	.05	.162
4	2.8	7.84	.10	.784

So, once mean is calculated you can easily calculate variance and standard deviation, so this is mean is given right you have calculated in previous slide which was what 1.20 right, so 0-1.20 is -1.20, 4-1.20 which is 2.8 right. So take square of this f(x) is known multiply this two and summation of this, so the variance of daily TV sales is 1.66. so, we are saying that the mean is 1.20, variance is mean is this variance is 1.66 and if you know variance you just take positive under root of it right.

So, this would become 1.28 isn't it, so this nothing but standard deviation, so this is a situation where standard deviation is more than μ ok? So, this is expected value and variance. Now let us look at another type of discrete probability distribution. So, we have seen one which was uniform the second one is binomial distribution, now binomial distribution is a discrete one and it is not a continuous one.

(Refer Slide Time: 14:34)

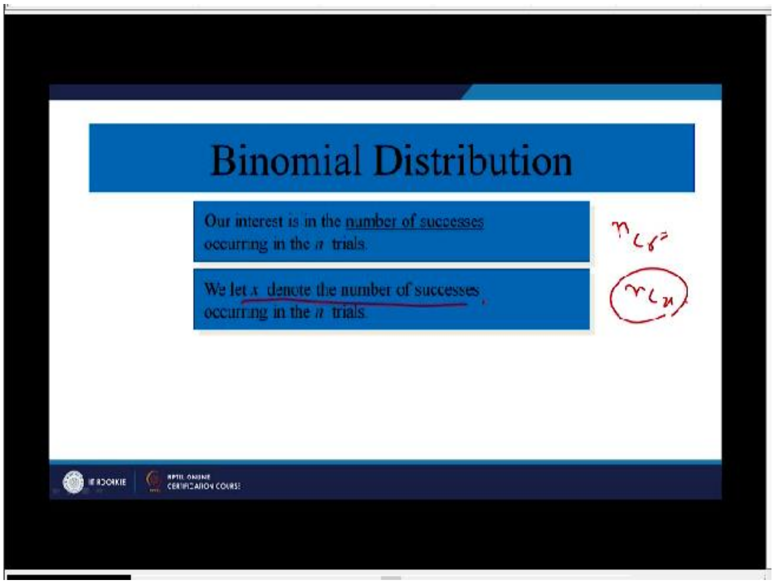


So, the data resulting from an experiment is known as Bernoulli process and there are certain properties of binomial experiment. So, these are the properties, so the first is the experiment consist of sequence of n initial, n identical trials. And there would be always two outcomes of an experiment, so when you throw a fair coin you will always have 2 outcomes right. So, throwing of a fair coin is nothing but a Bernoulli process right.

Because it has not only 2 outcomes, so the outcomes cannot be more than 2 that is the property of base. So, you can have outcome as success failure now what about throwing of a die is that this an example of this that is not an example of this. So, you will always have 2 outcomes whenever there is a trial right. the probability of success denoted by p does not change it remains fixed.

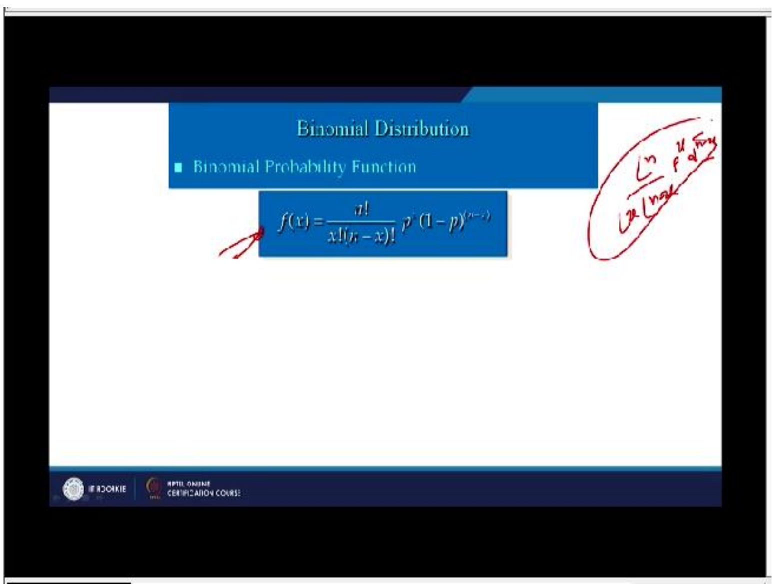
So when you throw a die the probability of success is always 0.5, now the success can be either heads or tails right. So, this is the second, the third probability right, so this Bernoulli process and what is binomial distribution it is a discrete one and it represents data which are output of a Bernoulli process and the trials are independent is the fourth condition. Because if you toss a fair coin first time let say you are getting heads right, second trial you do not know. You may get heads right, so the second trial is independent of first, so the are independent of each other.

(Refer Slide Time: 16:56)



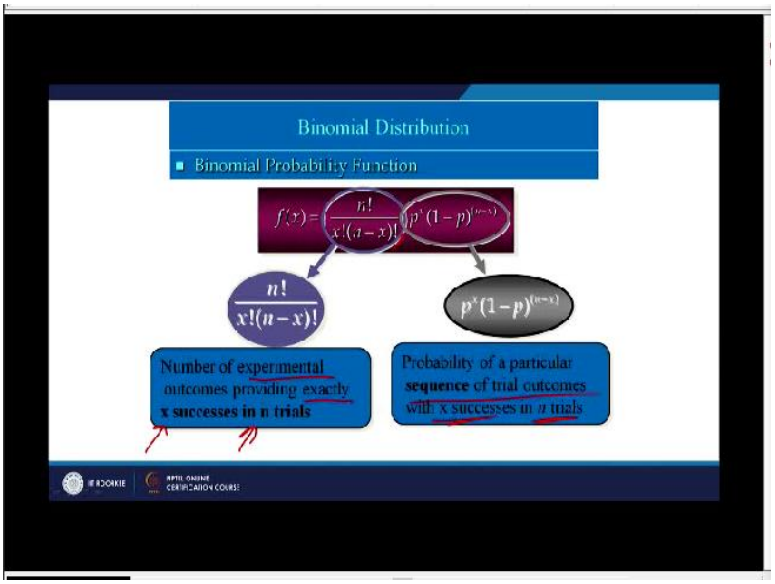
So, what we want in binomial distribution is we want to know how many success in n number of trials. So, it is let say either equal it $^{n}C_{r}$ or you can have it ^{n}Cx whatever you want right, so let x denote the number of successes occurring in n trials right. So, let say if you toss a Coin for 15 times what is the probability the 3 times you will get heads right.

(Refer Slide Time: 17:36)



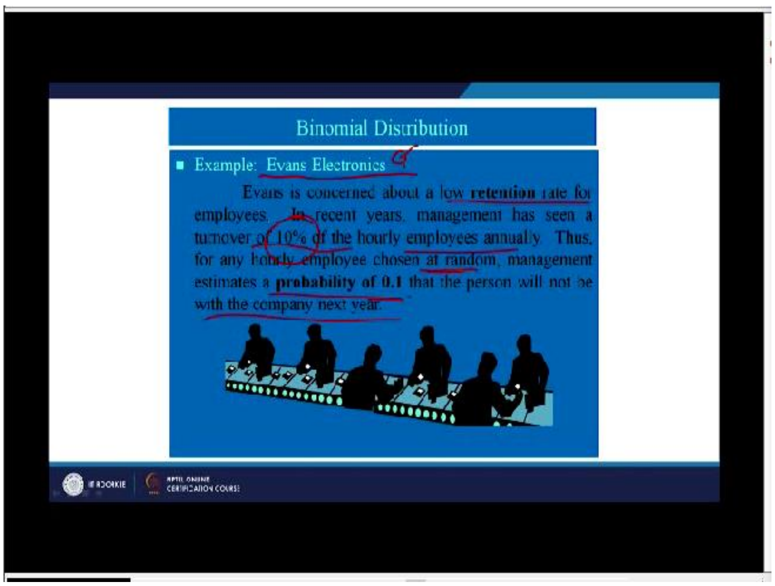
So, binomial probability function is this = $\dfrac{n!}{x!(n-x)!}p^{x}(1-p)^{(n-x)}$

(Refer Slide Time: 17:52)



So, this how you can have formula like this, so this portion is nothing but the number of experimental outcomes providing exactly x successes in n trials right. And this is probability of a particular sequence of trial outcomes x successes in n trials right.

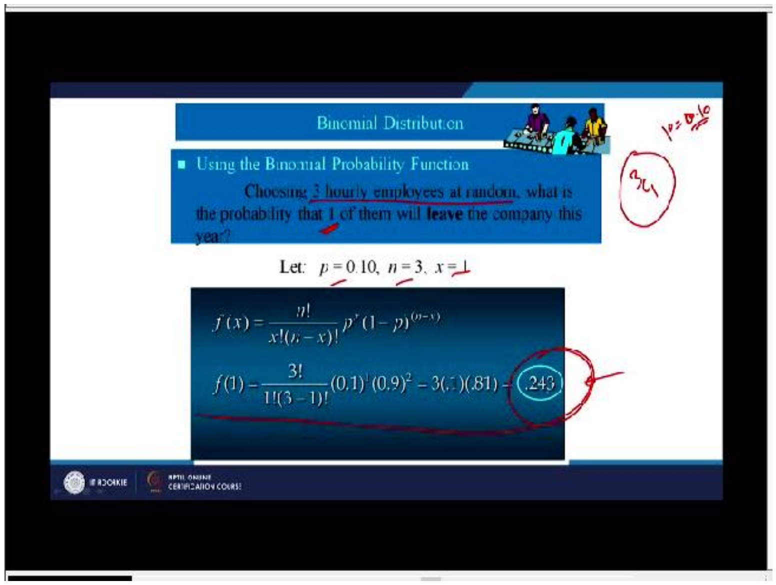
(Refer Slide Time: 18:17)



So, let us look at one question which is related on binomial distribution, so there is called Evans Electronics. And the this form is facing a problem of retaining employees and over the past few years' experience the management has seen that turnover is 10% of hourly employees. So let us say in the beginning of the year there are 100 employees and by the beginning of next year there will be only 90 employees right.

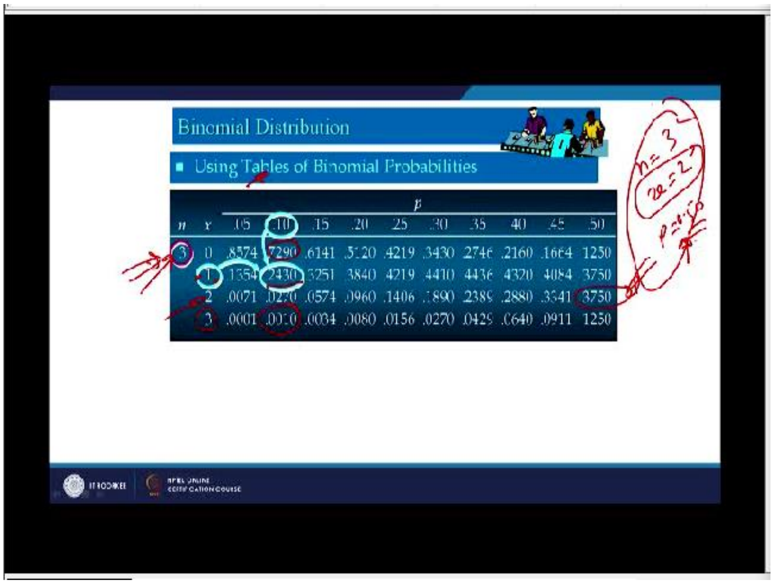
So, that is turn over it, so thus for an hourly employee chosen at random, so suppose if you choose an employee at random management estimates a probability of 10% right. So, that a person will not be with the company next year is not it, does the meaning of 10% and turnover it.

(Refer Slide Time: 19:21)



Now if you choose 3 hourly employees at random who are working with the Evan electronics what is the probability that one of them will leave the company. So, it is a case of $^3C_1$ right is not it, so you just put and we know what is p**.** P is 0.10, so you can put those values over here. So, p is known n is known x is known, so you just put this values. So the probability that if you out of 3 one hourly employee will leave is 0.2 for 3 hour 24.3% right.

(Refer Slide Time: 20:12)

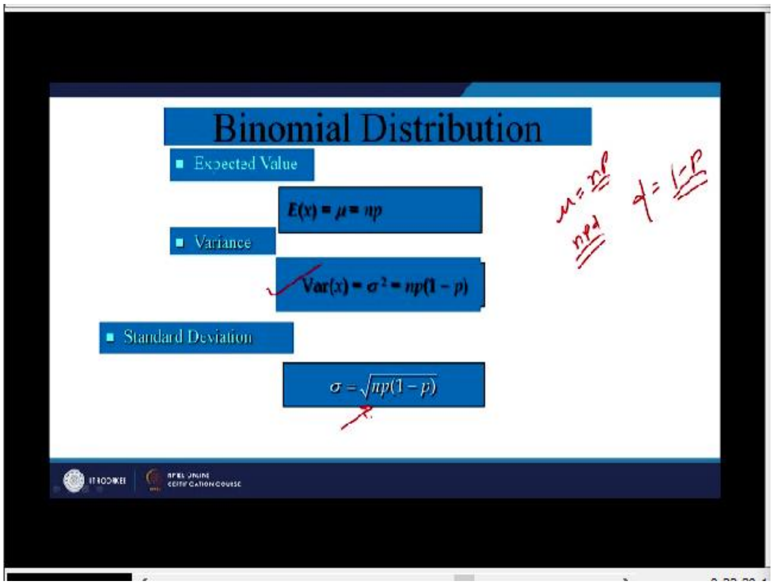


So, if you do not want to use this formula then you can solve this question using tables as well. So, if you get any good books on statistics, you will always have binomial table and this table gives you answer directly. So, we know that n is equal to 3 in our questions in our example you wanted to know what is the probability that one of them will leave the jobs, so this n this x, so probability that one will leave nobody will leave job this, probability that one will leave job is this is not it.

Probability that 3 will leave job is this, so let say if I ask you a question let say n=3, x=2 and p=0.50, so what is the probability that out of 3, 2 people will leave the job given that p is 0.50. So, it would be what this n=3, x=2. So, you have to look at here, so this the answer, so you can put these values in either that formula or you can directly get it from table. So, this is binomial distribution ok. So, keep in mind that its properties are there are only 2 output right.

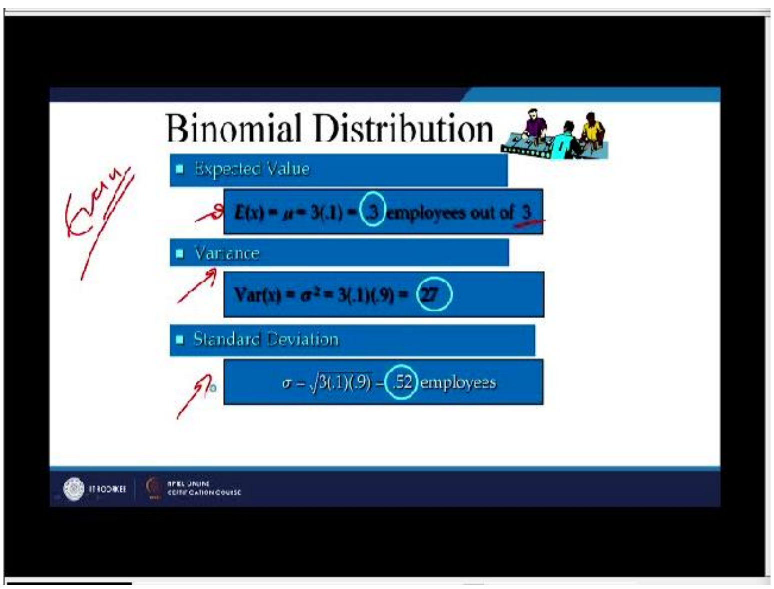
And the trials are independent these 2 important characteristics, let us look at what is the mean and variance of binomial distribution we have seen mean variance and standard deviation of uniform distribution.

(Refer Slide Time: 22:11)



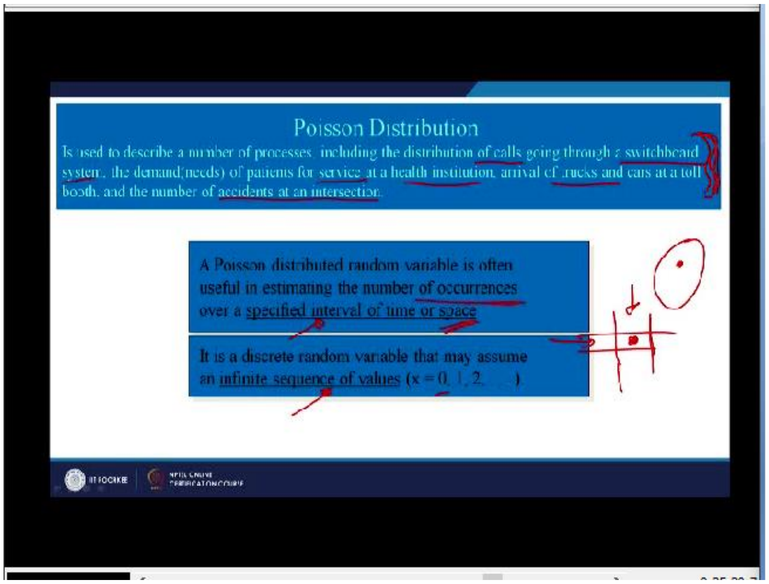
So, its mean is simple its np right, this variance is npq right, this q is nothing but 1-p isn't it and of course once variance is known standard deviation can be calculated right. So, it is very simple it is np, npq right.

(Refer Slide Time: 22:41)



So, mean for that example is, this is the example where in the expected value or mean is 3 right, so we are doing this the you are solving same example, Evan electrical example right Evan electronics example right. So, once you know mean you can calculate variance and standard deviation right. So, this how you can calculate mean and standard deviation of binomial distribution, now let us look at Poisson distribution.

(Refer Slide Time: 23:11)

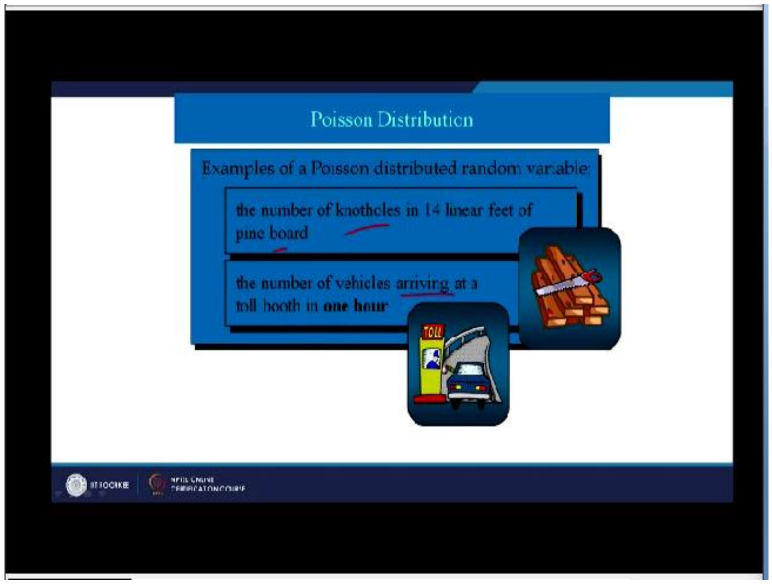


Poisson distribution is another type of probability distribution which we used to describe several processes for example the distribution of phone calls going throughout switch board system right. So, let say there is a firm and there are 1000 of employees, so what is the probability that at a particular given point of time 50 people make phone calls right or the demand of patients for service in a hospital right or let say arrival of trucks or cars at a toll booth or the number of accidents taking place at an intersection right.

So, this is an intersection, so is one road second road and this**,** so what is the probability that an accident will take place over here right. So, this is Poisson distribution and these are the processes which describe Poisson distribution, so here the and the random variable is often useful in estimating the number of occurrences over a specified interval of time or space or area right. So, **so** you can always find out let say what is the probability that a mosquito will bite you on your right hand is not it.

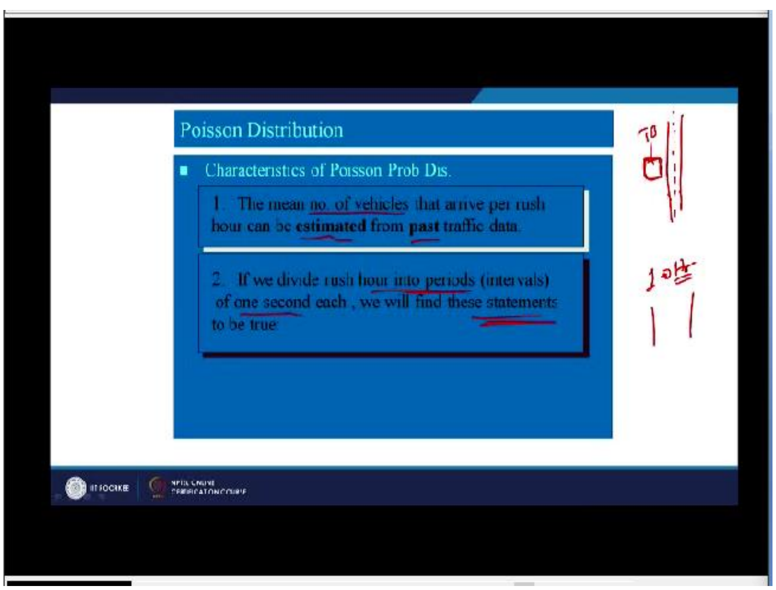
So, in a given particular area you will be finding probability of happening of an event, so it is a discrete distribution, so random variable may assume an infinite sequence of values it is a 0 to infinite right.

(Refer Slide Time: 25:21)



Let us look at this example these are couple of examples, so let say the number of knotholes in 14 linear feet of pine board right. So, these are nothing but not holes right is not it, so you may have different number of knotholes on a particular area of any let say any tree right. So, the number of vehicles arriving at a booth we already seen is not it. So, these are couple of examples right.

(Refer Slide Time: 25:57)

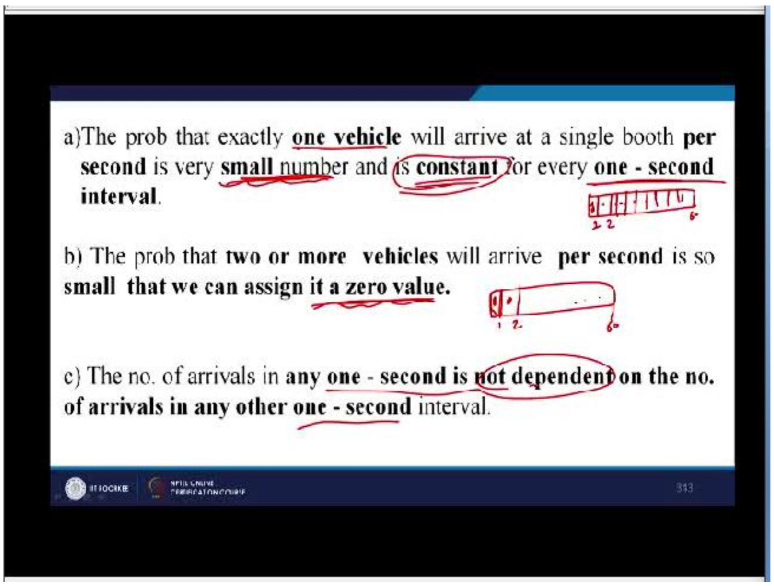


Similar to what we have seen in case of binomial distribution there are certain characteristics of poisson distribution as well. So, the mean number of vehicles that arrive per rush hour can be estimated from past data, so let say this is your road and you have but toll booth over here right.

So, these are the vehicles coming right, so you know from past data how many vehicles generally come during 1 particular hour ok.

So, we can calculate mean number of vehicles which are arriving during 1 particular hour of a day and that can be estimated from past data right. So if we divide rush hour into periods or equal intervals of 1 second each we will find these estimates true the following statements which are there in next slide true. And what are those statements, so what you are saying let's say if you get 1 hour time right.

(Refer Slide Time: 27:17)

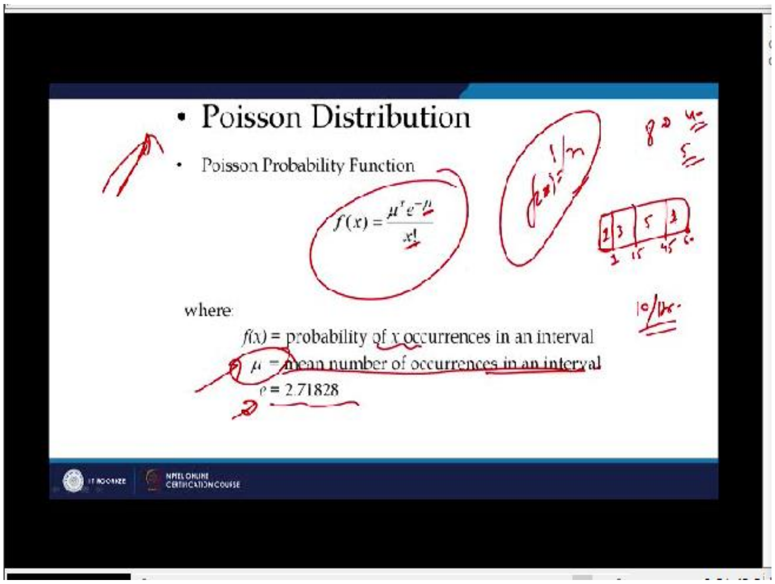


Then you can divide it into let say 60 equal parts ok, so the probability that exactly one vehicle will arrive at a single booth per second is very small number and is constant for every 1 second. So, let say this is your 1 hour period, so let us say this is first minute, second minute and so on right, so this is your 60th minute right. So, according to this probability that exactly 1 vehicle coming in first minute is a small one number.

So, it probabilities very small and that probability will remain same in all these intervals and it is that it is very small and it can be and it is it is a constant right for every 1 second interval. And the probability of 2 or more vehicles is coming in each of these intervals or this 1 second period is such a small value that you can keep it 0, is not it. So, let say this again last minute second minute and so on.

This is your 60th minute, so probability that 2 vehicles first, second and more than 2 vehicles coming in this interval that probabilities is can be assign equal to zero value. And the number of arrivals in any 1 second is independent of number of arrivals in any other seconds, so number of arrivals which are the number of vehicles coming over here or independent of number of vehicles coming in this interval right. So, that is independent right.

(Refer Slide Time: 29:07)

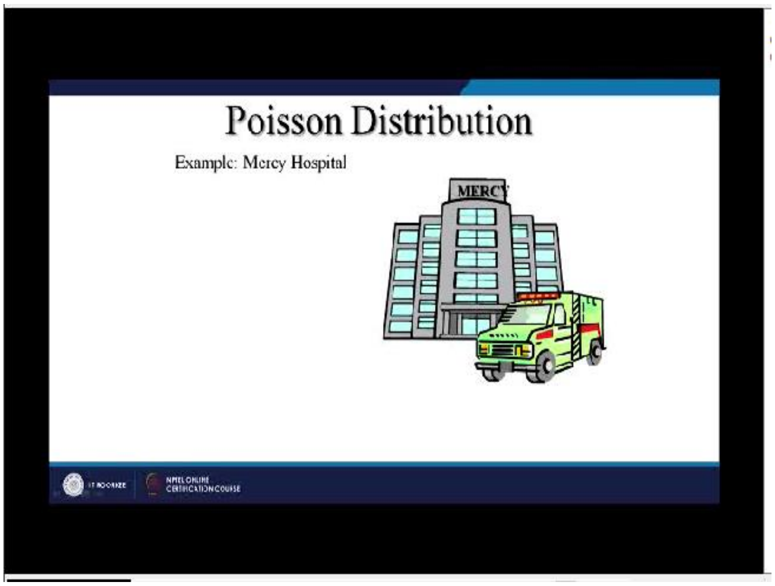


Poisson probability distribution is this, $f(x) = \frac{\mu^x e^{-\mu}}{x!}$

So, f(x) is probability of x occurrence in an interval is mean and this is e value 2.71 right. So, mean number of occurrence in and keep in mind this is mean value right ok so, let us take this example. So in an hour let say in let say this first met this is lecture 15th minute, 45th minute, and this is 60th minute.

So, let say you are getting one vehicle over here one vehicle here 3 here 5 here again one here right. So, this mean is the mean number of vehicles coming per hour, so what would be the mean, so 1 plus 3 4 4 5 9 and 10. So we will say 10 vehicles per hour ok, so let say in a day let say in 8 hours day you are getting 40 vehicles, so how many per hour is 5 per hour is not it. So, that is how you should calculate mean number of occurrences of an interval.

(Refer Slide Time: 30:37)



So, before going on to example of on Poisson distribution let me summarize what you have done today. Today we have seen probability distribution and we have define what is random variable. A variable which can take any value which as an output of an experiment ok, so if you toss a fair coin it may take either heads or tails, it is random you do not know what it will you know what will be the output right.

Then we have seen different types of distributions, we have seen uniform distribution we have seen binomial distribution, we have seen Poisson distribution and we have seen uniform distribution. So the simplest of all these distributions, so far was uniform distribution because the f(x) was just 1/n right and binomial distribution is again discrete one. And you need to ensure that in binomial distribution you are having your following Bernoulli's right characteristics of Bernoulli process right.

And then Poisson distribution which again use which is again useful for describing several process, for example the number of accidents taking place hidden intersection number of vehicles coming at a toll booth and so on. So with this let me finish today's session, will have couple of examples on by poisson distribution, in next session and we will also see continuous distribution in next class, thank you very much.