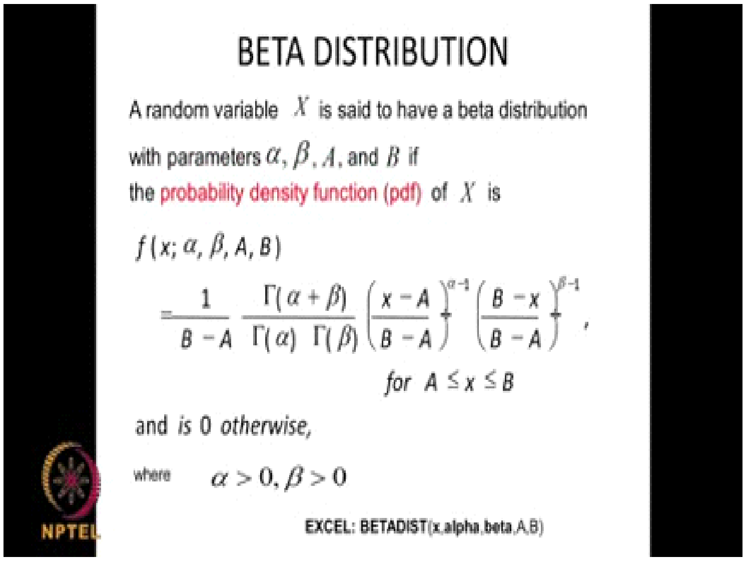
Biostatistics and Design of Experiments
Prof. Mukesh Doble
Department of Biotechnology
Indian Institute of Technology, Madras

Lecture - 30
Exponential / Hypergeometric distributions

Welcome to the course on Biostatistics and Design of Experiments. We will talk about two more distributions; one is called the exponential distribution; the other one is called the Hypergeometric distribution. All these are very useful in biology as well. Yesterday, we talked about the beta distribution, if you remember. Beta distribution, let me recollect again.

(Refer Slide Time: 00:35)



BETA DISTRIBUTION

A random variable X is said to have a beta distribution with parameters α, β, A, and B if the probability density function (pdf) of X is

$$f(x;\alpha,\beta,A,B) = \frac{1}{B-A}\,\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\left(\frac{x-A}{B-A}\right)^{\alpha-1}\left(\frac{B-x}{B-A}\right)^{\beta-1},$$

for $A \le x \le B$

and is 0 otherwise,

where $\alpha > 0, \beta > 0$

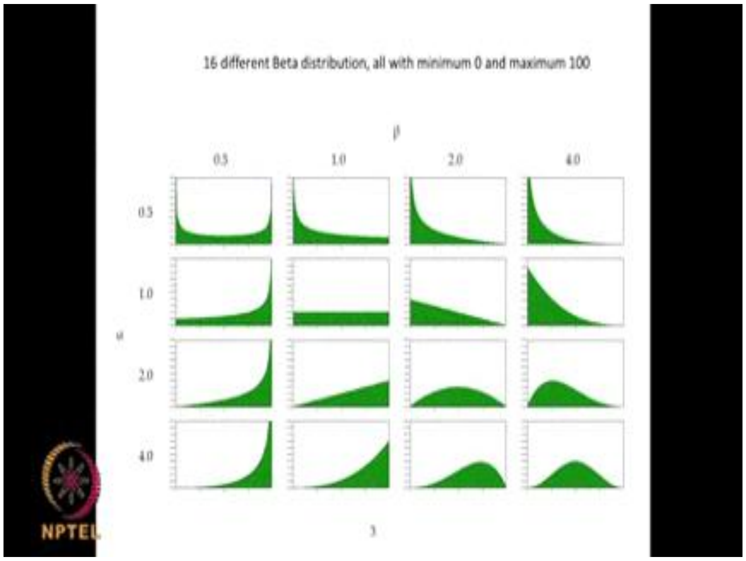
EXCEL: BETADIST(x,alpha,beta,A,B)

The probability density function f for x is given by this particular formula. We have the

$$f(x;\alpha,\beta,A,B) = \frac{1}{B-A}\,\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\left(\frac{x-A}{B-A}\right)^{\alpha-1}\left(\frac{B-x}{B-A}\right)^{\beta-1},$$

. So here, we have 4 parameters A, B; so, your x will lie between A and B and α, βare 2 parameters which are always > zero. So, we can always have A as 0 and B as 1, so, x may be lying between 0 and 1. So, that gives you a simplified version of the beta probability density function. Of course, Excel also has a BETADIST function, as you can see, you know, BETADIST, x, α, β, A, B; these are the 4 parameters. So, it calculates and lets you know what is the probability density function.

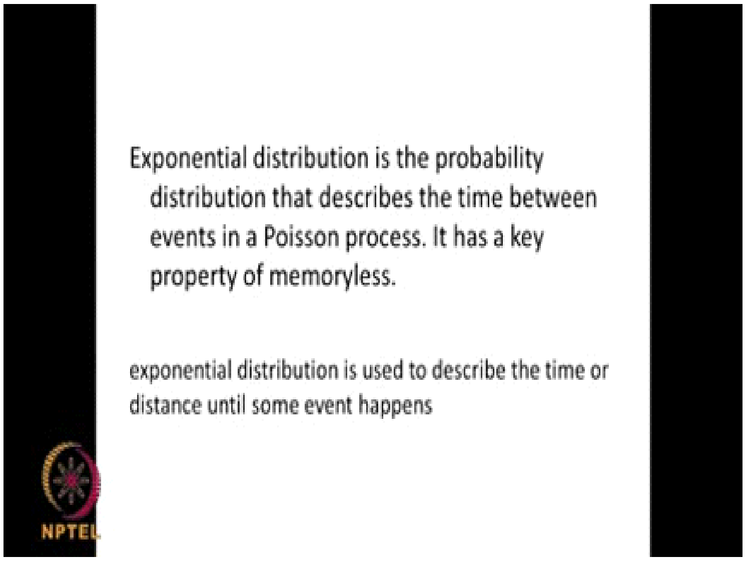
(Refer Slide Time: 01:35)



16 different Beta distribution, all with minimum 0 and maximum 100

Why do I say beta distribution is very important, we can get different types of shapes of graphs using different values of α, and β. So, that is the main advantage, especially if you are a modeling and simulating, simulation person, and if you want to generate different types of graphs, then this particular function is very, very useful. As you can see, we can get exponentially rising, linearly rising, similarly linearly dropping, exponentially dropping, and then, curves which gives you a maximum and minima. So, all these, we see in biology also. So, if one is interested in modeling in biological systems, modeling bio transformations, modeling bio reactors, then beta distribution is very useful. All I have to do is, manipulate this α, and β values, and you will get, end up getting different types of functions. Now, let us look at some more distributions.

(Refer Slide Time: 02:35)



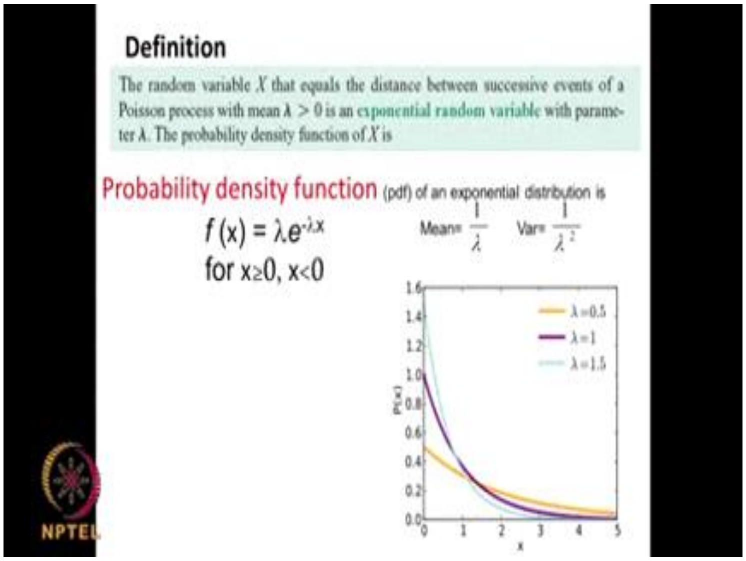
Exponential distribution is the probability distribution that describes the time between events in a Poisson process. It has a key property of memoryless.

exponential distribution is used to describe the time or distance until some event happens

The one is called the exponential distribution. Actually, exponential distribution, it is related to Poisson distribution. If you recollect, Poisson distribution gives you events; like, there are 3 road accidents in the city in a period of 1 month; there are ten infant mortalities in the South India over a period of 6 months. So, that sort of information it gives; events, number of events. But, the exponential distribution gives you the time between the events; what is the time it will take for the next event to happen, that sort of times it gives. And, another important thing is, it has a key property of memoryless; that means, the previous information and the new information are not correlated at all, actually. Just because a previous information has happened, or a event has happened, that does not mean the succeeding event will be dependent on the previous event.

So, for example, road accidents; there is no correlation that, because a road accident did not happen yesterday, or happened yesterday, there will not be another road accident tomorrow, or the… So, there is no correlation on the events. So, it is used to describe the time or distance until some events happen; like I say, as I said, you know, how long it will take for the next road accident to happen in the metro city of India; that sort of information it can give. So, it is an useful and as we can see, it is related to Poisson distribution.

(Refer Slide Time: 04:15)



Definition

The random variable X that equals the distance between successive events of a Poisson process with mean $\lambda > 0$ is an exponential random variable with parameter λ. The probability density function of X is

Probability density function (pdf) of an exponential distribution is

$f(x) = \lambda e^{-\lambda x}$

for $x \geq 0$, $x < 0$

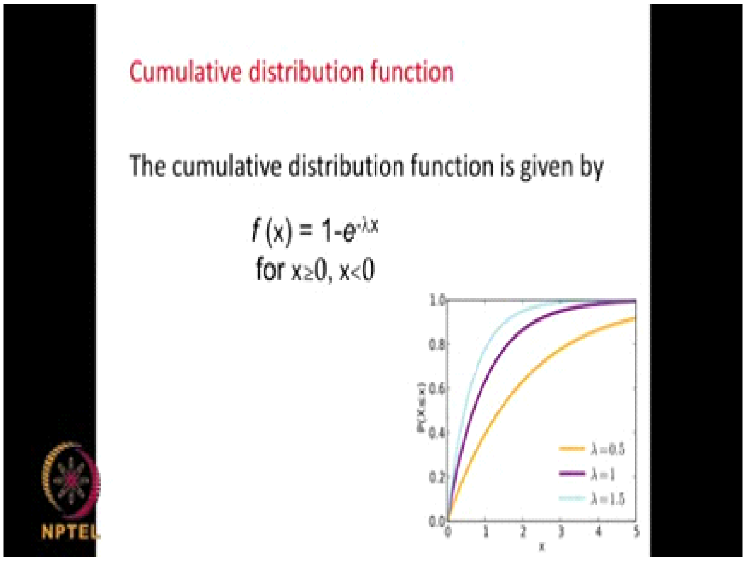
Mean= $\frac{1}{\lambda}$ Var= $\frac{1}{\lambda^2}$

So, the probability density function comes up with a relation like this; you have

$$\lambda e^{-\lambda x}$$

. If you remember, in Poisson distribution we had only one term λ, right?. So exactly, this also has only one term λ here, and then, x is the probability density function for x; x > 0, equal to 0, and generally, the graphs look like this. They will be falling down for different values of λ. So, the mean for this function is $1 / \lambda$, and the variance is given by $1 / \lambda^2$. So, the random variable x that equals distance between successive events of a Poisson process with the mean $\lambda > 1$ is an exponential random variable and this is the probability density function. For different a values of x, as you can see, the graphs gives you the probability density function.

(Refer Slide Time: 05:19)



Cumulative distribution function

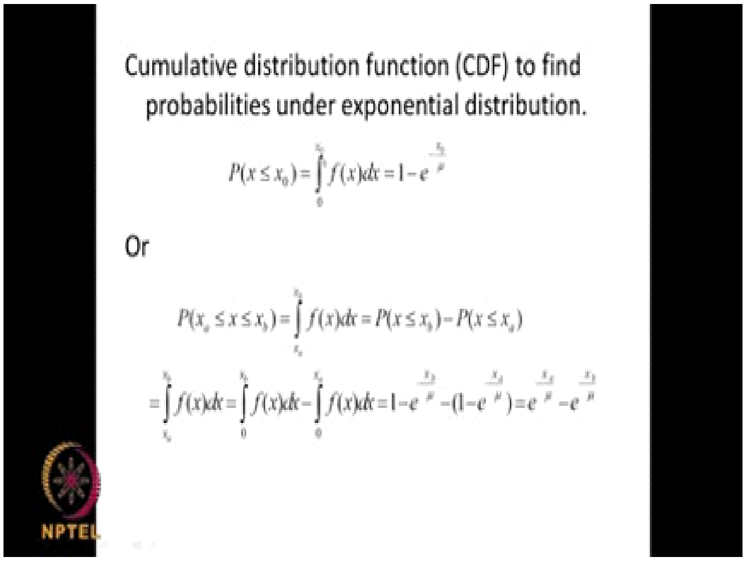
The cumulative distribution function is given by

$f(x) = 1\text{-}e^{-\lambda x}$
for $x \geq 0$, $x<0$

.

So, you can calculate the cumulative distribution also. All you need to do is, it is given by this particular relation, $1\text{-}e^{-\lambda x}$, for values of, different values of x. And because it is cumulative, it keeps building up, as shown in this figure; whereas the probability density function, sorry, probability density function keeps falling as a function of x.

(Refer Slide Time: 05:49)



Cumulative distribution function (CDF) to find probabilities under exponential distribution.

$$P(x \leq x_0) = \int_0^{x_0} f(x)dx = 1 - e^{-\frac{x_0}{\mu}}$$

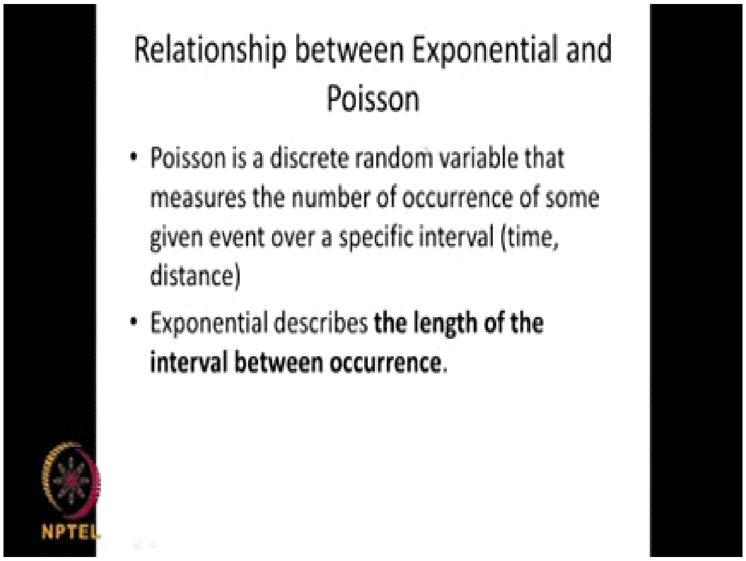
Or

$$P(x_a \leq x \leq x_b) = \int_{x_a}^{x_b} f(x)dx = P(x \leq x_b) - P(x \leq x_a)$$

$$= \int_{x_a}^{x_b} f(x)dx = \int_0^{x_b} f(x)dx - \int_0^{x_a} f(x)dx = 1 - e^{-\frac{x_b}{\mu}} - (1 - e^{-\frac{x_a}{\mu}}) = e^{-\frac{x_a}{\mu}} - e^{-\frac{x_b}{\mu}}$$

So, cumulative distribution function is nothing, but integral of 0 to X_0, for $x \leq X_0$, so, $1-e^{-\frac{x_0}{\mu}}$. So, I can integrate it and do the substitutions, and you can end up having this type of relationship; that gives you the cumulative distribution function.

(Refer Slide Time: 06:12)

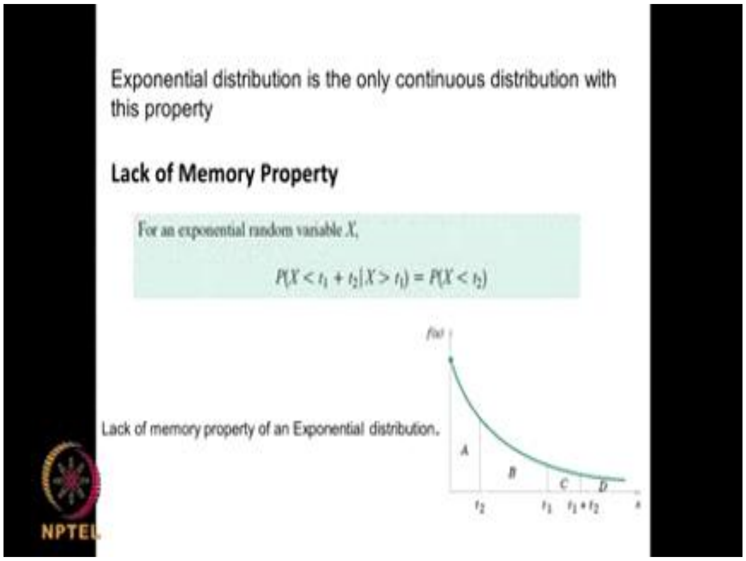


Relationship between Exponential and Poisson

- Poisson is a discrete random variable that measures the number of occurrence of some given event over a specific interval (time, distance)
- Exponential describes **the length of the interval between occurrence.**

So, what is the relation between exponential and Poisson; as I mentioned, Poisson is a discrete random variable that measures the number of occurrences of some event, whereas exponential gives, exponential distribution gives you the length of the interval between occurrences; how long it will take for a particular event to take place; whereas Poisson distribution tells you how many events; that is why they are both inter-related.

(Refer Slide Time: 06:38)



Exponential distribution is the only continuous distribution with this property

Lack of Memory Property

For an exponential random variable X,

$$P(X < t_1 + t_2 | X > t_1) = P(X < t_2)$$

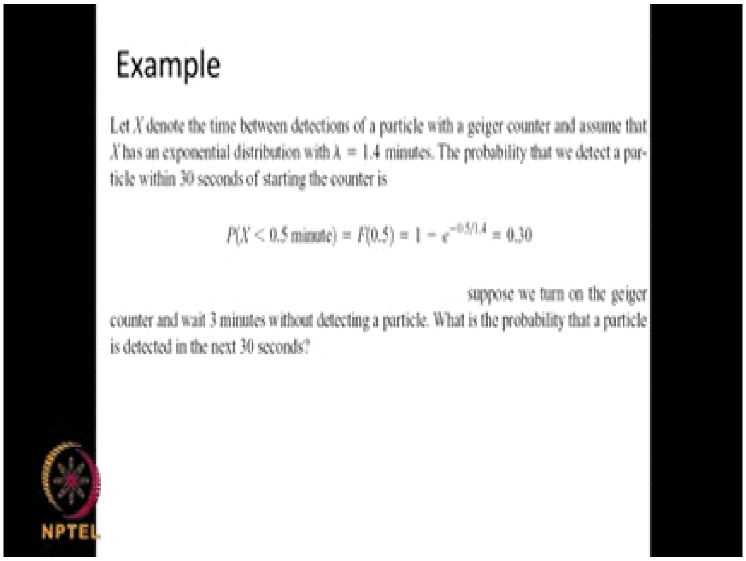
Lack of memory property of an Exponential distribution.

So, exponential distribution is a continuous distribution and it has no memory term attached to it; that means, just because an event has happened, that does not mean that will have a bearing on the next event; like the road accidents, there are no bearing on that at all. So, for a exponential random variable X, we have

$$P(X < t_1 + t_2 | X > t_1) = P(X < t_2)$$

, is given by P (X < t 2); that means, there is no, whether you take a time on January first, or whether we take time on February first, the time for the next event to happen will be independent; whether I start my starting point as January first, or I start my starting point as February first. So, that is the lack of memory, which is characteristic of this exponential distribution.

(Refer Slide Time: 07:32)



Example

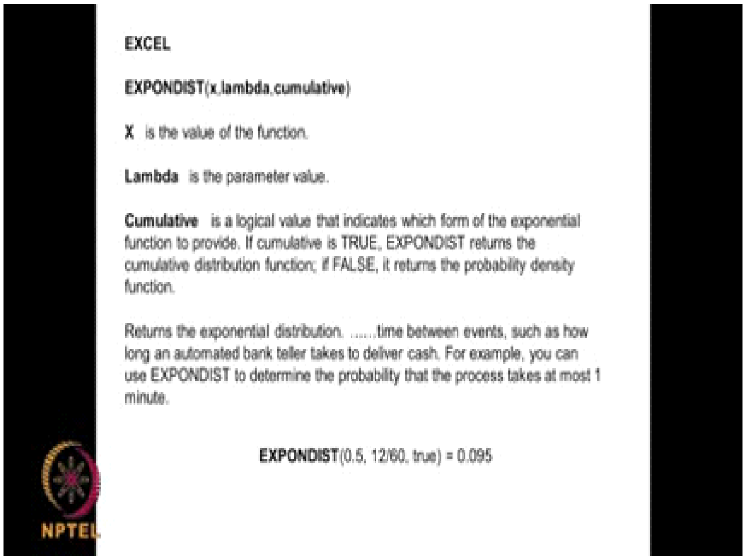
Let X denote the time between detections of a particle with a geiger counter and assume that X has an exponential distribution with $\lambda = 1.4$ minutes. The probability that we detect a particle within 30 seconds of starting the counter is

$$P(X < 0.5 \text{ minute}) = F(0.5) = 1 - e^{-0.5/1.4} = 0.30$$

suppose we turn on the geiger counter and wait 3 minutes without detecting a particle. What is the probability that a particle is detected in the next 30 seconds?

So, we will look at some examples.

(Refer Slide Time: 07:43)



EXCEL

EXPONDIST(x,lambda,cumulative)

X is the value of the function.

Lambda is the parameter value.

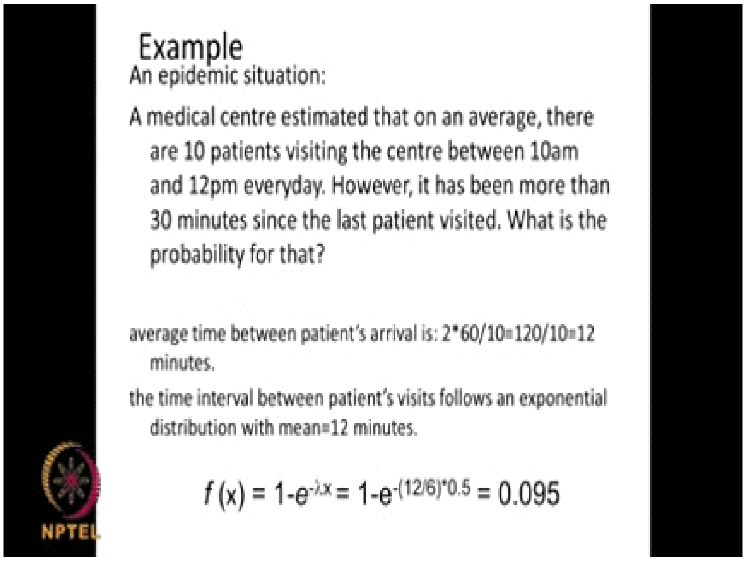
Cumulative is a logical value that indicates which form of the exponential function to provide. If cumulative is TRUE, EXPONDIST returns the cumulative distribution function; if FALSE, it returns the probability density function.

Returns the exponential distribution. ......time between events, such as how long an automated bank teller takes to deliver cash. For example, you can use EXPONDIST to determine the probability that the process takes at most 1 minute.

EXPONDIST(0.5, 12/60, true) = 0.095

Before that, let me also, sorry, let me also talk about the Excel. There is an Excel EXPONDIST, which is given by x , lambda , cumulative; lambda is the parameter value; x is the value of the function; cumulative, if it is true, it gives you a summation; cumulative, if it is, if I give it as false, then it will return the probability density function at that particular point, actually. So, let us look at one simple problem.

(Refer Slide Time: 08:09)



Example

An epidemic situation:

A medical centre estimated that on an average, there are 10 patients visiting the centre between 10am and 12pm everyday. However, it has been more than 30 minutes since the last patient visited. What is the probability for that?

average time between patient's arrival is: 2*60/10=120/10=12 minutes.

the time interval between patient's visits follows an exponential distribution with mean=12 minutes.

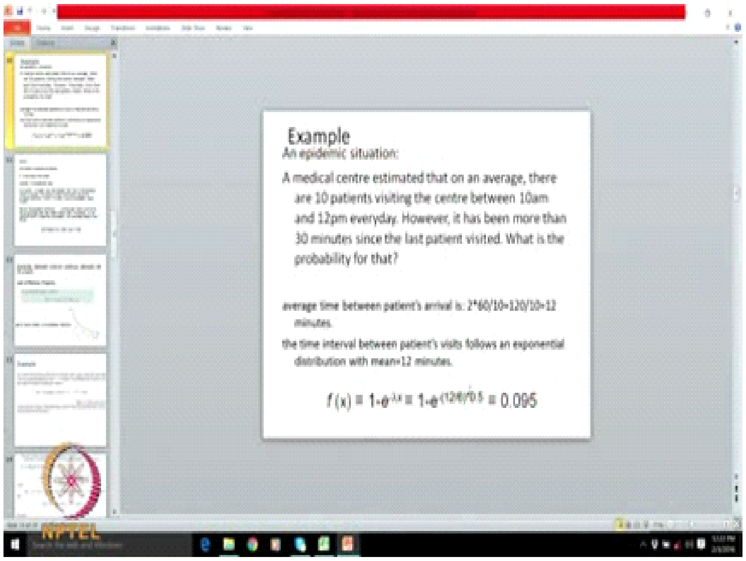
$$f(x) = 1-e^{-\lambda x} = 1-e^{-(12/6)*0.5} = 0.095$$

So, an epidemic is happening. So, when there is an epidemic, you will have patients coming regularly to clinics. So, that could be modeled using an exponential distribution function. Their time of arrival could be modeled using an exponential distribution function. So, a medical center, what it does is, it estimated that, on an average, there are 10 patients visiting the center between 10 am and 12 pm; in that 2 hours, 10 patients come in. However, it has been more than 30 minutes since the last patient visited, what is the probability for that? So, for 30 minutes nobody came. So, what is the probability of that? So, average time for each patient's arrival, it is 2 hours, because 10 am to 12. So, 2 * 60, because we are converting into minutes, divided by 10; so, that gives you 12 minutes. So, average time for each patient to arrive is 12 minutes; that is, lambda. Now, we want to look at time interval between patients visit follows an exponential distribution, with the mean of 12 minutes, right?. So, we can get

$1-e^{-\lambda x}$

; lambda is nothing, but 12 minutes, and so, we are converting that into hours; then, take 0.5. So, that gives you 0.095. So, actually, this should read as 60, 12 / 60, because we are converting that into hours, multiplying by 0.5, 0.5 is your x, 30 minutes. So, this is equal to 0.095, ok?.

(Refer Slide Time: 09:50)



Example

An epidemic situation:

A medical centre estimated that on an average, there are 10 patients visiting the centre between 10am and 12pm everyday. However, it has been more than 30 minutes since the last patient visited. What is the probability for that?

average time between patient's arrival is: 2*60/10=120/10=12 minutes.

the time interval between patient's visits follows an exponential distribution with mean=12 minutes.

$$f(x) = 1\text{-}e^{-\lambda x} = 1\text{-}e^{-(12/60)*0.5} = 0.095$$

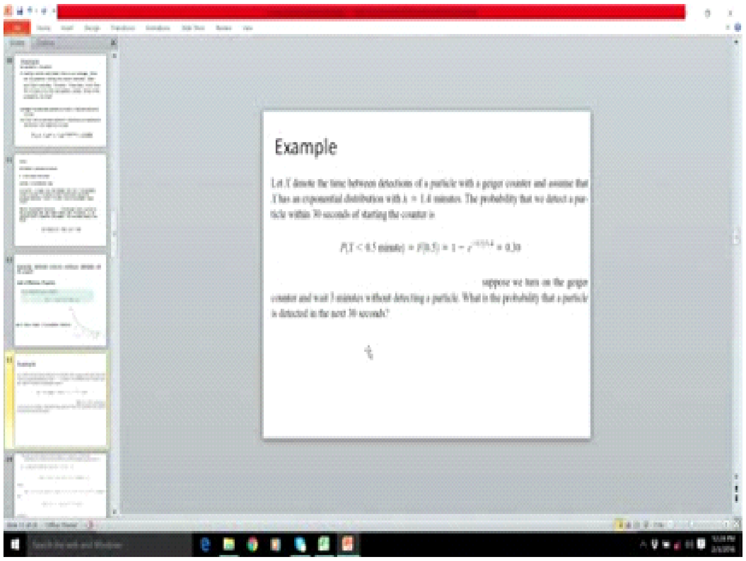
So, we can do the same thing here, using the Excel function. So, Excel function is EXPONDIST; Excel function EXPONDIST x lambda , cumulative; either it could be true, or it could be false. So, let us look at the Excel function. So, we go to the Excel function. We say, we say EXPONDIST , 0.5, that is 30 minutes, 12 / 60, because we are converting that 12 minutes into 60. So, the cumulative, so, we could call it true or false here. We look at both cases. 0.095. So, the probability for 30 minutes nobody came to the clinic because of the illness is 0.095. We put a cumulative here, because, when we say for 30 minutes nobody came, it could have been 1 minute, 2 minutes, 3 minutes, 4 minutes and so on, actually, ok?.

So, that is why we put it as true. So, the same thing is obtained from this equation also,

$$1\text{-}e^{-(12/60)*0.5} = 0.095$$

, wherein the 12 is converted into hours; that is why we are dividing by 60, multiplied by 0.5. So, that gives you 0.095. So, we get the same answer using the Excel command also. So, it is simple. Now, it lacks memory. So, that is most important property of exponential distribution. Let us look at another problem. X denotes the time between the detection of a particle with a Geiger counter, and then, we can assume that X is an exponential distribution with lambda = 1.5 minutes. The probability that we detect a particle within 30 seconds of starting the counter... So, the probability, 30 seconds, within that time, is given by 0.3. So, that is the time it takes for that to come back; it takes about 0.3.

(Refer Slide Time: 12:22)



Example

Let X denote the time between detections of a particle with a geiger counter and assume that X has an exponential distribution with λ = 1.4 minutes. The probability that we detect a particle within 30 seconds of starting the counter is

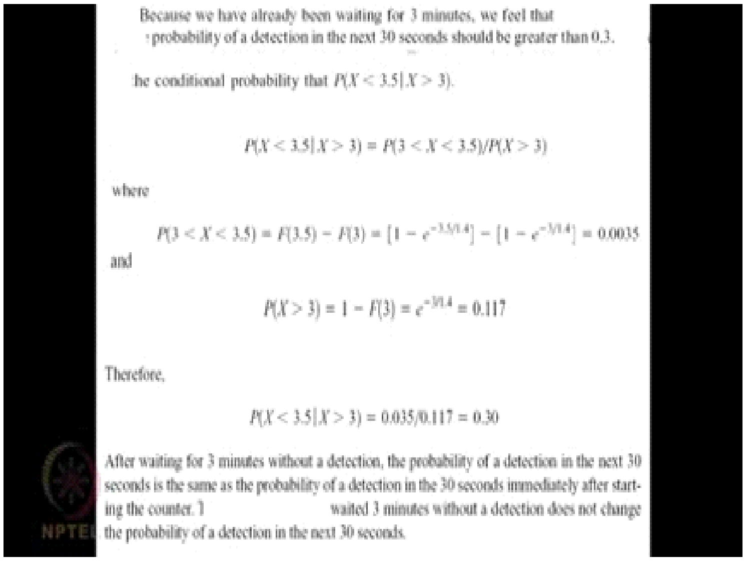
$$P(X < 0.5 \text{ minute}) = F(0.5) = 1 - e^{-0.5/1.4} = 0.30$$

suppose we turn on the geiger counter and wait 3 minutes without detecting a particle. What is the probability that a particle is detected in the next 30 seconds?

So, we can use the Excel function also to get the same answer. So, we can put in EXPONDIST into x, into lambda, into cumulative, to get the same answer also. And so, the probability is 0.3. Now, suppose we turn on the Geiger counter and wait for 3 minutes without detecting a particle, in this situation, we detect a particle within 30 seconds of starting the counter. So, that gives you 0.3. So, we can put in the same thing here, using the Excel command also, sorry.

It is given as 30 seconds. So, we need to convert that into minute; that will be 0.5. Then, next one is 1.4 minutes, comma; then, we say true. So, we get EXPONDIST 1.0. So, we get it as 0.69. So, when we do it as 1 - this, and that will come to 0.30. So, this is how you deal. So, it is a probability of 0.3 that we detected a particle within 30 seconds, or within 0.5 minutes. Now, we turn on the Geiger counter, and wait for 3 minutes without detecting a particle. What is the probability that a particle is detected in the next 30 seconds? So, you see, wait for 3 minutes, and then we want to look at whether particles come within that 30 seconds, or we start our Geiger counter, and within 30 seconds we detect a particle. So, just because we have not seen a particle for 3 minutes, do you think the probability will increase? No, according, if you start doing these calculations, we will see, we will end up again with the same answer; we will end up again with the same answer, which means, even after waiting for 3 minutes without detecting a particle, the probability of a deduction in the next 30 seconds, is the same as the probability of deduction in the 30 seconds immediately after starting the counter.

(Refer Slide Time: 15:04)



Because we have already been waiting for 3 minutes, we feel that probability of a detection in the next 30 seconds should be greater than 0.3.

he conditional probability that $P(X < 3.5 | X > 3)$.

$$P(X < 3.5 | X > 3) = P(3 < X < 3.5)/P(X > 3)$$

where

$$P(3 < X < 3.5) = F(3.5) - F(3) = [1 - e^{-3.5/1.4}] - [1 - e^{-3/1.4}] = 0.0035$$

and

$$P(X > 3) = 1 - F(3) = e^{-3/1.4} = 0.117$$

Therefore,

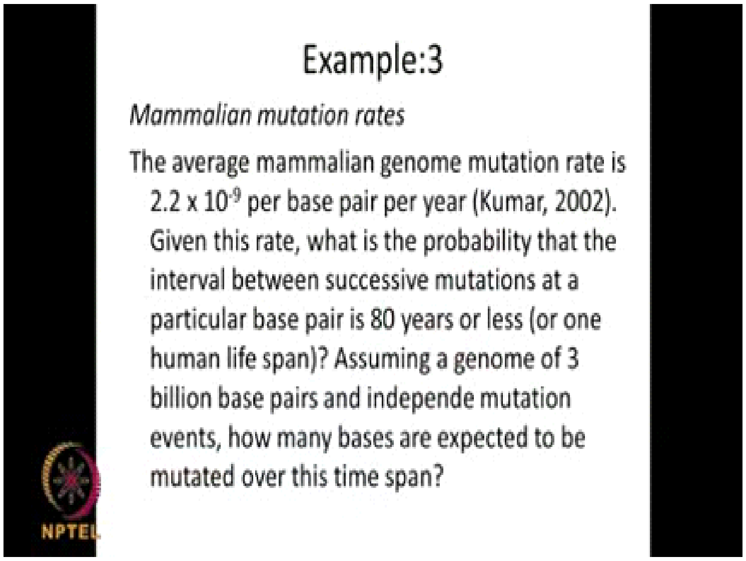
$$P(X < 3.5 | X > 3) = 0.035/0.117 = 0.30$$

After waiting for 3 minutes without a detection, the probability of a detection in the next 30 seconds is the same as the probability of a detection in the 30 seconds immediately after starting the counter. 1 waited 3 minutes without a detection does not change the probability of a detection in the next 30 seconds.

So, whether I wait for 3 minutes, and then look for a particle within 30 seconds, or I start my Geiger counter, and look for a particle within 30 seconds, you will get the same probability. Why, because, as I said, exponential distribution is memory-less. So, it does not consider, that because it has not seen a particle for a very, very long time, that the probability will increase. So, that is the beauty of this particular distribution. Do you understand this problem? So initially, we start the Geiger counter, and within 30 seconds we see a particle. So, we will say the probability is, say 0.3. Then, we start the Geiger counter, but we wait for 3 minutes without detecting a particle, then what is the probability that we will be detect a particle within that 30 seconds, that is 3 minutes, and between that 3 minutes and 3 minutes 30 seconds, you will come, get the probability as the same; there will not be any difference at all, because of the characteristic of this exponential distribution. It is not that the probability will increase, because we have not seen a particle for a long time.

So, the probability will not change, because unlike other distribution, because we have not seen a event occurring, probability will not change at all in the exponential distribution.

(Refer Slide Time: 16:37)



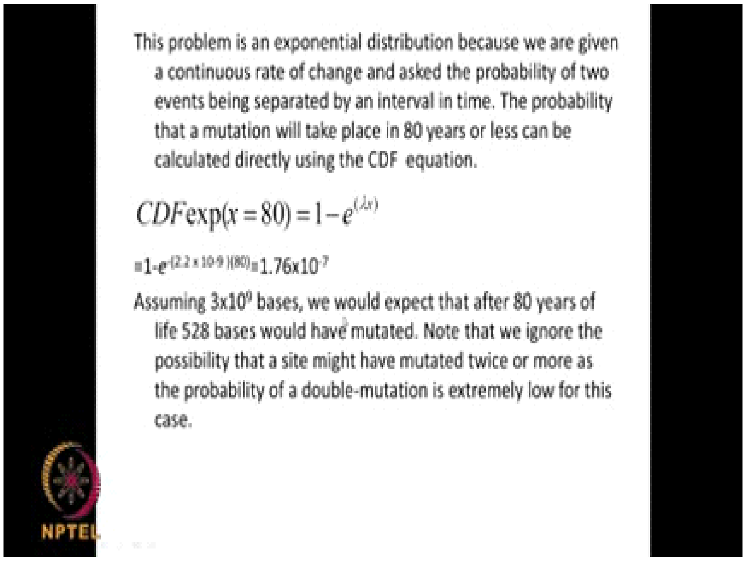
Example:3

Mammalian mutation rates

The average mammalian genome mutation rate is 2.2 x 10^{-9} per base pair per year (Kumar, 2002). Given this rate, what is the probability that the interval between successive mutations at a particular base pair is 80 years or less (or one human life span)? Assuming a genome of 3 billion base pairs and independe mutation events, how many bases are expected to be mutated over this time span?

So, let us look at another example. The average mammalian genome mutation rate is 2.2 * 10 power minus 9 per base pair per year. This is taken from this reference. So, this rate is given. What is the probability that the interval between the successive mutations at a particular base pair is 80 years or less? We assume that as the human life span. Assuming a genome of 3 billion base pairs and independent mutation events, how many bases are expected to be mutated over this time span? So, we assume it as independent mutation, and we will assume it as base pairs, billion pair. So, this is the average mutation and we are given 80 years. So, we want to look at, what is the probability that the interval between successive mutations is about 80 years or less. So, what do we do?

(Refer Slide Time: 17:30)



This problem is an exponential distribution because we are given a continuous rate of change and asked the probability of two events being separated by an interval in time. The probability that a mutation will take place in 80 years or less can be calculated directly using the CDF equation.

$$CDF\exp(x=80)=1-e^{(\lambda x)}$$

$=1\text{-}e^{(2.2 \times 10\text{-}9)(80)}=1.76\text{x}10^{-7}$

Assuming $3\text{x}10^9$ bases, we would expect that after 80 years of life 528 bases would have mutated. Note that we ignore the possibility that a site might have mutated twice or more as the probability of a double-mutation is extremely low for this case.

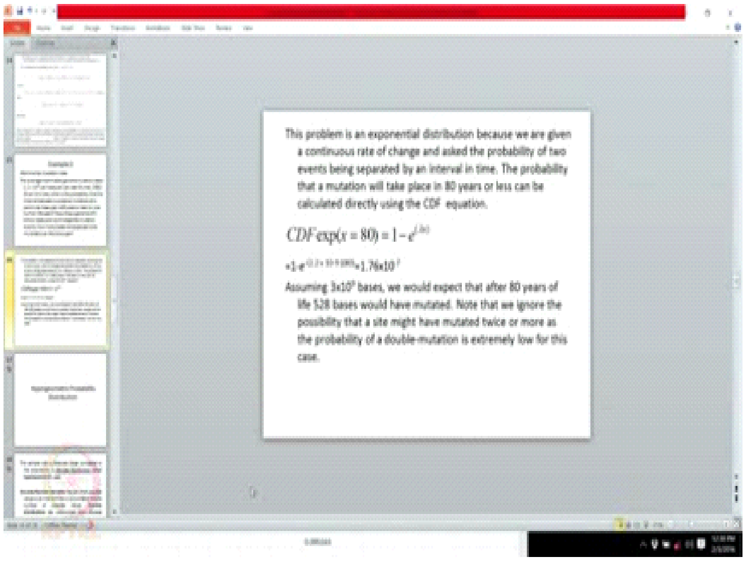
$1\text{-}e^{-\lambda x}$; x is your 80 years; lambda is

- $1\text{-}e^{-(2.2 \times 10\text{-}9)(80)}=1.76\text{x}10^{-7}$

, sorry. So, when you multiple by 80, we will end up getting

. So, the probability that a mutation will take place in 80 years or less, is given by this particular probability. So, we can use the same thing using our Excel function also, ok?.

(Refer Slide Time: 18:09)



This problem is an exponential distribution because we are given a continuous rate of change and asked the probability of two events being separated by an interval in time. The probability that a mutation will take place in 80 years or less can be calculated directly using the CDF equation.

$$CDF\exp(x=80)=1-e^{(-\lambda x)}$$

$=1\text{-}e^{-(2.2\times10\text{-}9)(80)}=1.76\text{x}10^{-7}$

Assuming $3\text{x}10^{9}$ bases, we would expect that after 80 years of life 528 bases would have mutated. Note that we ignore the possibility that a site might have mutated twice or more as the probability of a double-mutation is extremely low for this case.

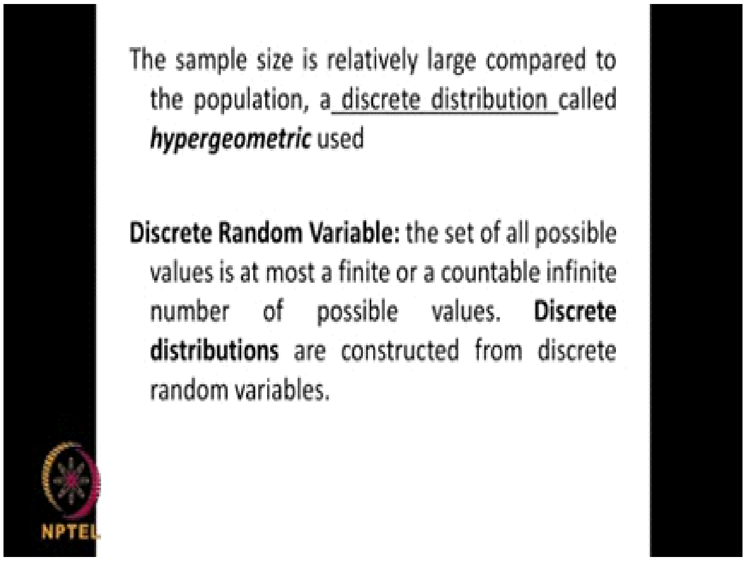
So, we have the Excel function. It has got a EXPONDIST;

$1\text{-}e^{-(2.2 \text{ x } 10\text{-}9\ )(80)}=1.76\text{x}10^{-7}$

x = 80 years * $^{-(2.2 \text{ x } 10\text{-}9\ )}$true; you get 1.76x$10^{-7}$, you understand?. So, that is the probability of a mutation taking place in 80 years or less. Now, if you assume 3 * 10 ^{9} as bases, then, we would expect that after 80 years, there will be 528 bases that should have got mutated; do you understand, how do to do this? We multiply 3 * 10 ^{9}. So, multiply 3 * 10 ^{9}with, 3 * 10 ^{9} multiplied by this particular term, that will give you 528. . So, totally 528 bases would have been mutated. So, of course, here we assume that, the possibility of a site getting mutated twice. So, every time we assume that double mutation is extremely low. So, it is only single mutation. So, by assuming that, we are able to get a probability of 1.76x10^{-7}, and we assume, if there are about 3 * 10 ^{9}, there would have been 528 bases that would have got mutated in the 80 years. So, you understand? So, this is an exponential distribution. It tells you the time between events. So, it calculates the probability of time, for a particular time to, for an event to happen.

It has got no memory. Just because we look at some event in the month of January, or we look at some event in the month of February, the starting point does not matter. The probability of that particular event occurring, the time for the probability, is going to be the same. So, if you are looking at, say accidents on the roads. So, whether we look at it from January, how long it will take for that accident to take place, or when we look at February, how long it will take, it will be the same. So, that is the special characteristic of this exponential distribution. And, it is related to Poisson distribution, in the sense that, the Poisson distribution tells you the number of events, whereas the exponential distribution tells you the time between two events. So, this is also a very important distribution to know, especially in the area of biology, as you can see, it is quite useful. Now, let us look at another distribution, that is called the hypergeometric distribution, ok?.

(Refer Slide Time: 21:30)

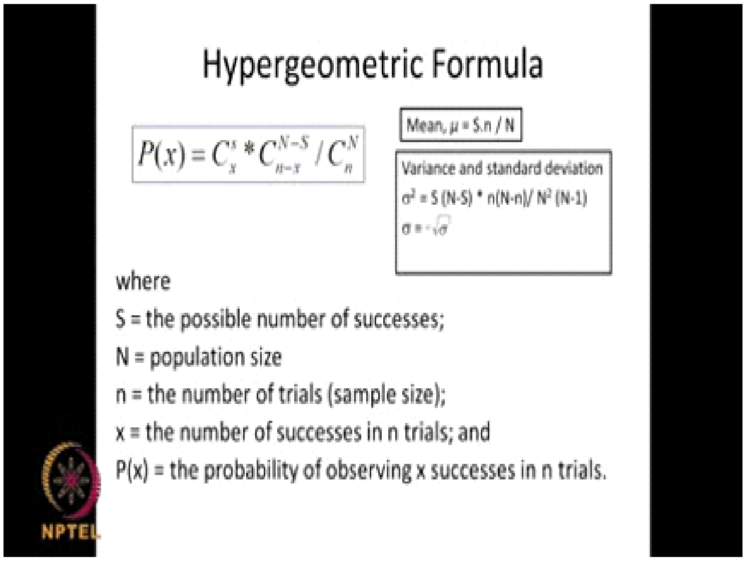


The sample size is relatively large compared to the population, a discrete distribution called ***hypergeometric*** used

Discrete Random Variable: the set of all possible values is at most a finite or a countable infinite number of possible values. **Discrete distributions** are constructed from discrete random variables.

What is hypergeometric distribution? So generally, here in a hypergeometric, the sample size is relatively large when compared to the population of a discrete distribution; that is called hypergeometric. So, the sample size is quite large, whereas normally, in normal distribution, the sample may be small, population is very very large. We take 10 students from a university and measure their heights. So, the population is huge. Whereas, in a hypergeometric distribution, the sample size is sort of considerable to the population size. So, it is not really very very small; that means, the population is not very, very large. So, the sample and the population are not very far away. Then, that sort of distribution is called the hypergeometric distribution, ok?. So, it gives you a discrete random variable, the set of all possible values, at most a finite or a countable infinite number of possible values. So, it gives you, discrete distributions are constructed from discrete random variables, actually. So, it gives you a set of all possible values, that means, a finite, or a countable infinite number. Normally, infinite number is not countable; but here, in this particular distribution, we call it infinite, but it is still countable, ok?.

(Refer Slide Time: 22:58)



Hypergeometric Formula

$P(x) = C_x^s * C_{n-x}^{N-S} / C_n^N$

Mean, $\mu = S.n / N$

Variance and standard deviation

$\sigma^2 = S\,(N\text{-}S) * n(N\text{-}n)/ N^2\,(N\text{-}1)$

$\sigma = \sqrt{\sigma}$

where

S = the possible number of successes;

N = population size

n = the number of trials (sample size);

x = the number of successes in n trials; and

P(x) = the probability of observing x successes in n trials.

So, it is given by this relationship, the probability.

$P(x) = C_x^s * C_{n-x}^{N-S} / C_n^N$

and small n, where your n is your sample size, S is the event supposed to have happened, N, capital N is your population size, small n is your sample size, x is the number of successes in n trials. When I take a sample size of small n, then, my x is the number of successes, whereas in the population of the capital N, capital S gives you the possible number of successes; do you understand? For example, if I have 100 balls in a box, out of that 10 balls are black, the capital S, if I am picking up balls, I want to pick up black balls, capital S will be 10, and n could be 100; and, when I take a sample of, say 5, the small n will be 5; and, if I get 1 black ball in that, the success x will be 1. So, this calculates the probability. So, you understand that, the capital N, the population, although we call it population, has got a finite size, unlike the normal distribution, where population is very very very large. Here, the mean μ is given by S n / capital N; S is the possible number of successes; n, small n is the number of trials; capital N is the population. So, n / N is some sort of a ratio. So, S, that gives you the μ mean; very logical to think about. Variance and standard deviations are given like this. The variance is S, that is capital S,

$$\sigma^2 = S\,(N\text{-}S) * n(N\text{-}n)/\ N^2\ (N\text{-}1)$$

. This is how the variance is given by. So, this is the mean; this is the variance. So, the hypergeometric distribution is based on the concept, that is sample size is not very very small when compared to population size, or the population size is finite. It is not infinite, unlike the other type of distribution. So, this is also very useful distribution.

We will come across in many situations, in many problems, even in biology and non biological studies as well. We will look at each one of them more in detail. So, this is the C S C x is nothing, but S factorial ÷ x factorial, ≤ S - x factorial. So, this is the symbol for that. So, S factorial in the numerator; denominator will have x factorial and S - x factorial. So, if you look at this, this will have N - S factorial in the numerator, small n - x factorial in the denominator; then, capital N -S - small n plus x factorial again in the denominator. That is how it is broken, and this will show capital N in the numerator; denominator will have small n factorial, and also, N - n factorial also in the denominator. That is how the terms look like in hypergeometric distribution. So, we can also look at problems in hypergeometric distribution and as I said, it is very useful type of analysis to carry out, especially in biological systems also. And, this distribution is quite useful in that sense, where you have a limited number of population, and sample, you take from that limited number of population, ok?.

Thank you very much. We will continue in the next class.

Key Words: Exponential distribution, Poissan distribution, hypergeometric distribution, variance, standard deviation, probability, cumulative distribution