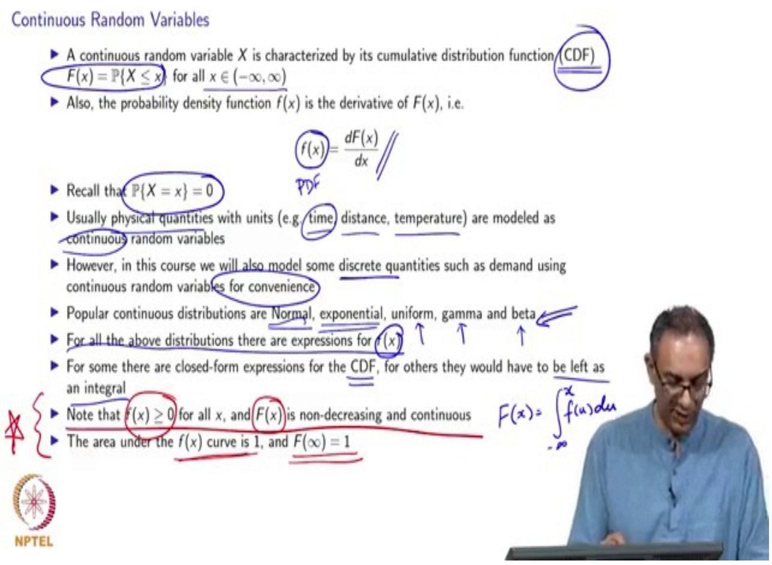
Decision Making Under Uncertainty

Prof. Natarajan Gautam
Department of Industrial and System Engineering
Texas A&M University, USA

Lecture – 06
Continuous Random Variables: Characteristics and examples

The next topic is continuous random variables. We will look at the characteristics and examples first and then we will talk about expected value and variance. I do want to say if you have never seen the topic of continuous random variables, I would recommend spending some time on this maybe either through Khan Academy or looking through these slides and the video multiple times, because this is a topic from my experience, I have seen that students find this extremely hard. So, let's see how this goes.

(Refer Slide Time: 00:44)



So, a continuous random variable X is typically characterized by what is called CDF, the Cumulative Distribution Function. Why do we need something like this? Because, earlier we had used the probability mass function, but here the mass is 0 for all values of little x; therefore, the probability mass function is meaningless.

So, we look at something like the cumulative distribution function the CDF. So, the CDF is defined for all values of x between negative infinity and infinity. Now, there is another parameter called the PDF, the Probability Density Function which is nothing, but the derivative of the CDF with respect to x.

Now, whenever we model, we will see this continuous random variable a lot in this course. The reason for that is most physical quantities like time, distance, temperature and almost any other thing with the unit is usually modeled as a continuous random variable. A discrete random variable usually are things that are number; the number of people who are waiting in a queue, the number of you know vehicles that are produced in a factory and so on.

However, if it is a physical quantity, like how much time does it take? For example, how much time is this lecture going to take; that is a continuous random variable. Although, we do discretize it because we are going to be measuring in terms of seconds; we discretize, but typically things like time and distance and temperature and so on are usually modeled as continuous random variables.

Now, that's what I was saying a little while ago, even some discrete quantities that you know that are typically discrete are sometimes, for convenience, modeled as continuous. You may wonder, "Why would anybody want to do that?". Continuous is much harder. Well, turns out that a lot of times, unless you have very special functions, the summation signs do not work out as nicely as the integral; in fact, the integrals work out beautifully in many of these examples.

So, let's look at some of those. So, before we do that, I want to give you some names of some special continuous distributions. One is the normal distribution; you have all probably heard of the normal distribution; we will go to talk about it in some detail. We will also talk about the exponential distribution in some detail. We will not talk about the uniform distribution. However, we have already said a little bit about it; this is the random variable that we generated using the calculator between 0 and 1; that is the uniform distribution between 0 and 1.

We will not talk about the gamma distribution much, except in an example. And, beta distribution is coming in two flavors: we will see one flavor in this topic and another flavor in a different topic. When you get there, I will say a little bit more about it. There are many other distributions like chi-square, t-distribution, lognormal distribution and so on.

For all those distributions, I am able to write down the little $f(x)$, which is the probability density function. The reason I do not want to define something in terms of the PDF is that it is not easy for us to write down a meaning for it.

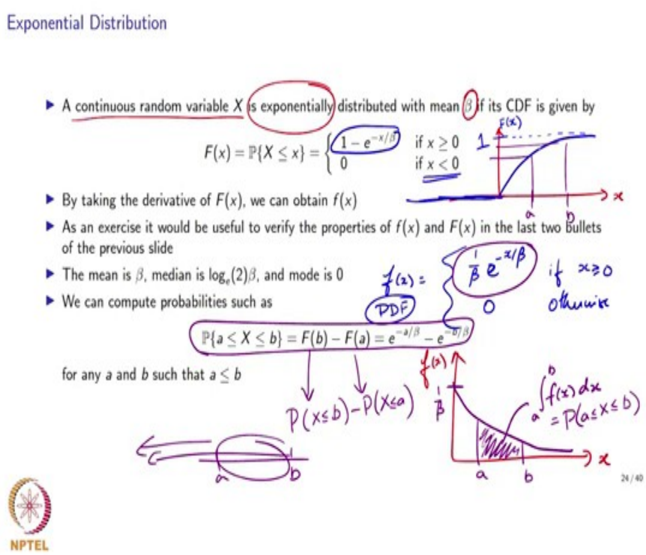
So, it is easier to write down the meaning of the CDF, the Cumulative Distribution Function, which we can write down as a probability the random variable is less than or equal to a value x. Now, the CDF is not available for all random variables. In fact, for the normal and the gamma distribution in general, it is not possible to write down the cumulative distribution function.

So, we will have to leave them as an integral. So, if you think about it, the cumulative distribution function is just the $\int_{-\infty}^{x} f(u)\,du$ because it is the area to the left of x. So, we will look at an example and I will make that a little more clearer.

Now, there are two properties that are important and I want to emphasize that. These two are crucial properties. The first property is that this probability density function is always a number that is greater than or equal to 0 for every value of x, $f(x) \geq 0 \, for \, all \, x$. Secondly, the CDF, $F(x)$, is non-decreasing and continuous.

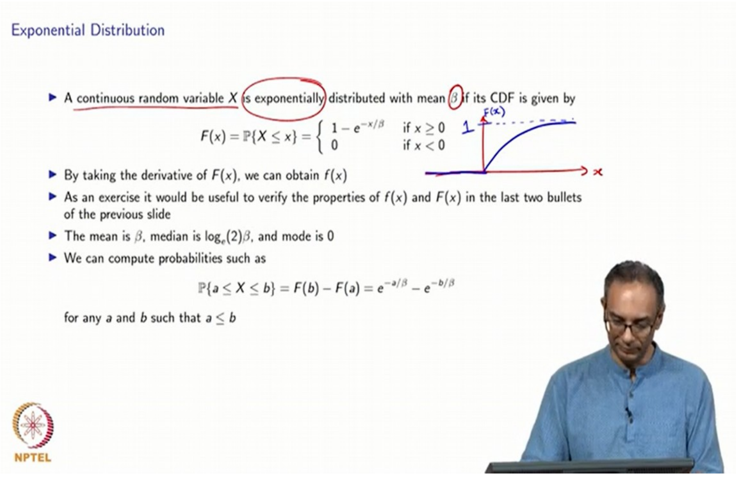
These are very important properties and the area under the $f(x)$ curve is 1 and that means, the value of the CDF at infinite equals 1; so, that's the same thing. So, when we look at an example, this will become a little bit more obvious.

(Refer Slide Time: 05:35)



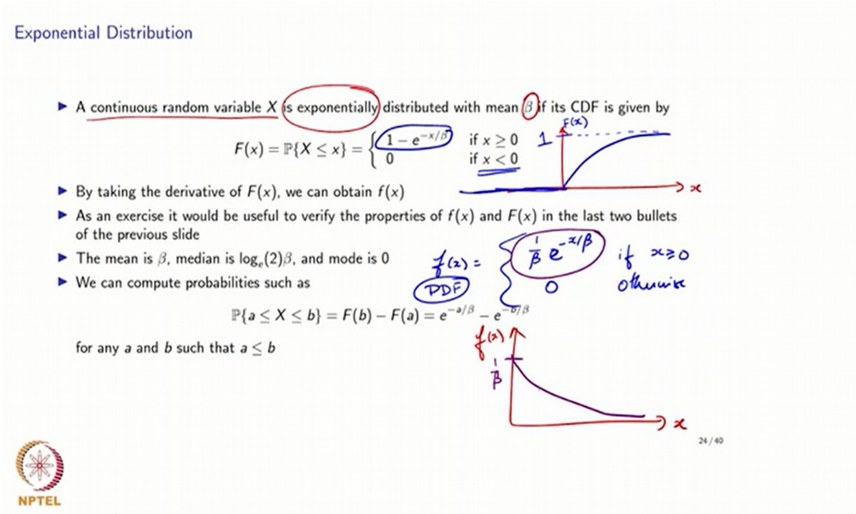
Now, let's look at this very special distribution called the exponential distribution. The exponential distribution is a very popular one in the field of operations research. A lot of us usually deal with the exponential distribution. So, this is one of the two special distributions that we will see in this course. So, a continuous random variable x is called exponential with mean β. Sometimes, the parameter of the exponential is written differently, such as 1 over the mean; for this course alone, I am choosing to use the mean itself as the parameter because most software including Octave that we will see and R and so on, use the mean as a parameter. So, we are going to go ahead and use that.

(Refer Slide Time: 06:25)



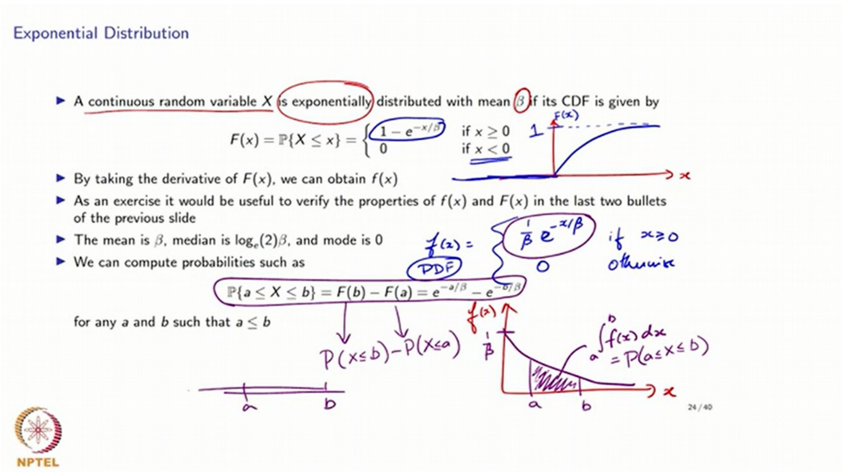
If you plot a graph of the CDF versus x, this is how it looks like. The CDF is at 0, continuous; then, it moves over and goes towards 1. So, this is my $F(x)$ versus x. So, $F(x)$ increases and hits 1 asymptotically. Now, $F(x)=0$ when $x<0$. When $x \geq 0$, the value of x takes this exponential form; hence, the name exponential distribution. So, it is $1-e^{\frac{-x}{\beta}}$ when $x \geq 0$. So, that is the functional form. So, if you take the derivative of $F(x)$, you will get little $f(x)$, which is the probability density function. So, I take the derivative of that. I would get $f(x)=\frac{1}{\beta}e^{\frac{-x}{\beta}}$ if $x \geq 0$; it is 0, otherwise. All I did is I took the derivative. I took the first derivative of uppercase, $F(x)$.

(Refer Slide Time: 07:52)



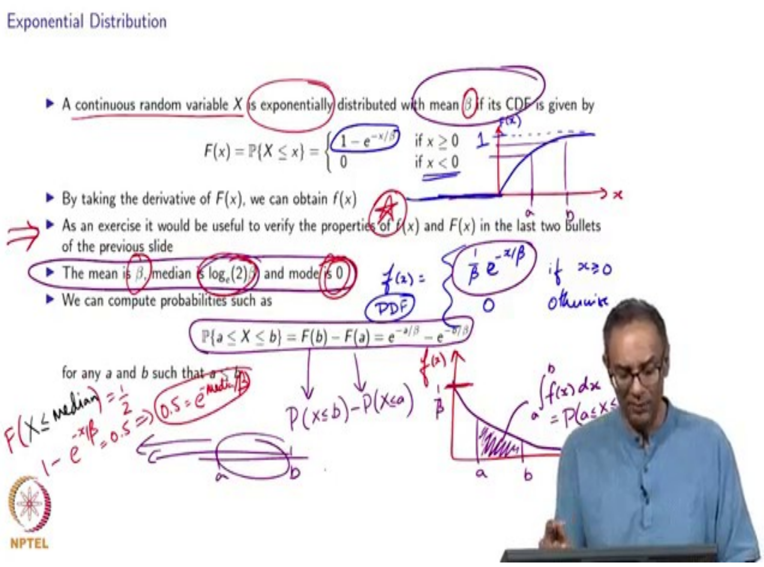
So, this is my PDF, Probability Density Function. So, if you plot the graph of the PDF, you will get the following. If you plot the graph of the PDF, the little $f(x)$ will asymptotically go toward 0. At $x=0$, this value is $\frac{1}{\beta}$, alright; if you take this equation and plug in $x=0$, then $e^{0}=1$ and you get $\frac{1}{\beta}$.

(Refer Slide Time: 08:33)



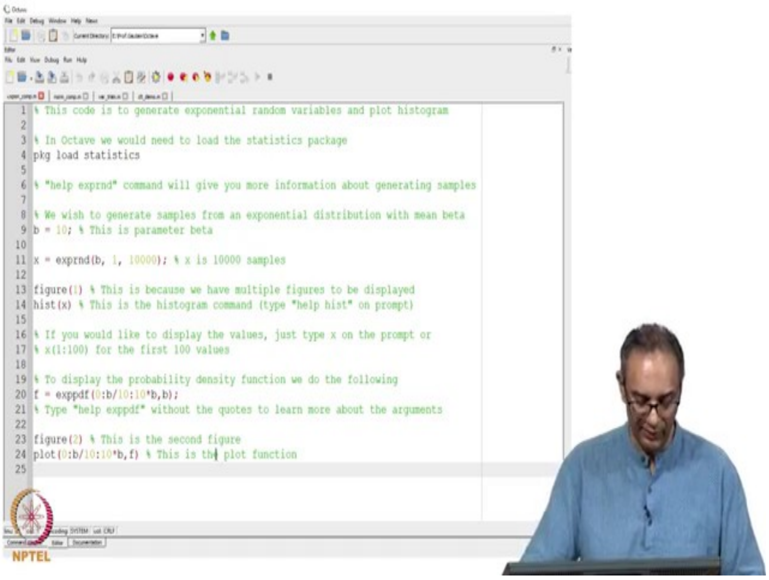
Now, if I want to compute the probability that x is in between two values a and b, I would do that using the area under this curve. So, this area is $\int_a^b f(x)\,dx$ and that is nothing, but the probability that the random variable takes a value between a and b. Now, you could also compute that as $P(X \leq b) - P(X \leq a)$. It's the same thing, right; probability that it is between a and b is equal to probability that is less than b minus the probability less than a. So, you could have as well have taken two values a and b on this graph and then, written it like this. That would have also worked for you. Whichever way you would like to compute, you can compute.

(Refer Slide Time: 09:48)



Now, there are three things I want to talk about. We will derive them in the next lecture, but I want to say a little bit because we have the graph here. The first thing is the mean as said here is β. So, that we know. I haven't told you how to compute it or why it's β; we will see that later. The median is not terribly hard to show. If you think about it, to compute the median, $P(X \leq Median) = 0.5$. So, that means, $1 - e^{\frac{-x}{\beta}} = 0.5$; that means, $0.5 = e^{\frac{-Median}{\beta}}$. Now, if you use that and take the logarithm to the base e which is a natural logarithm, you get this result because you would get natural log of 0.5, with a negative sign outside it becomes log 2 multiplied by β. So, that is the median. And, the mode is 0 because that is the value that has the highest mass. So, that's where the mode is 0. So, notice that this is a very special distribution where the mean and the median and the mode are all three different numbers; this is something that is useful. Now, another thing that I do want to say is that it is useful to check the properties we saw in the previous page, that is these two properties towards the end that are marked with a star, okay. These properties are $f(x) \geq 0 \, for \, all \, x$ and $F(x)$ is non-decreasing and continuous; the area under the $f(x)$ curve is 1 and $F(\infty) = 1$. So, it's important to verify those properties with a star and I am not going to do that here. But, I would recommend highly recommend, especially if you have never seen this topic, that you go check that those properties are actually satisfied. So, make sure that you check that the first and the second results are all valid. Now, before I go to the normal distribution, what I want to do is, I want to do a little demo. So, I am going to move to the software called Octave.

(Refer Slide Time: 12:05)



So, I am going to use a software called Octave. So, if you have used MATLAB, this is basically MATLAB which is available for free, for anyone to use. So, I decided to use that particular software. I want to go over the program. Like I said, I am going to make this available to you all.

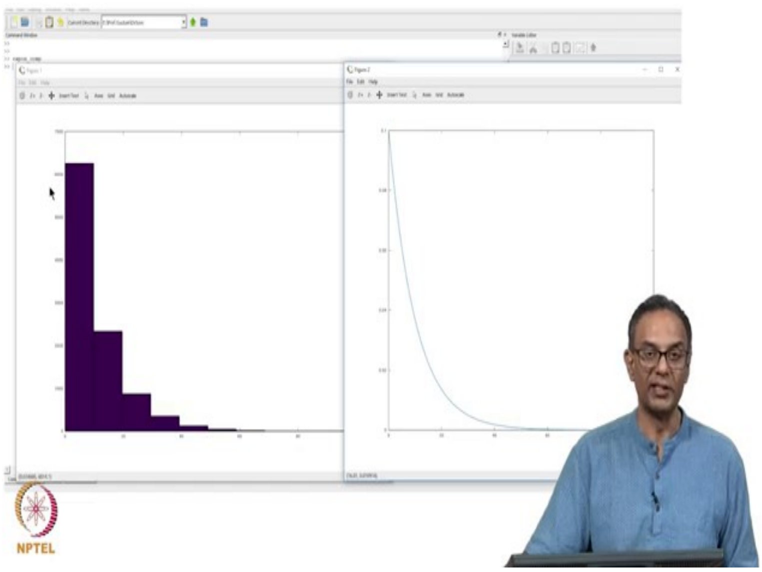
So, this code essentially generates exponential random variables and then, it plots a histogram; that's what this one does. What we need to do is, first we need to load a package called statistics by typing "pkg load statistics"; obviously, we are going to do some statistical analysis. So, you have to do that; that's the first command. If you want to learn a little bit more about generating exponential random variables; you want to type "help exprnd".

So, the reason I put all these up in green is because I put them inside a percentage sign. So, all this does not get read by the program; this is just for comments for you all. Now, we want to generate samples from an exponential distribution with parameter β; that's what we want to do.

And in this case, I am picking $\beta = 10$ and what I do is, I generate 10000 samples from an exponential distribution with parameter β. And, 1 means 1 row; 10000 means 10000 columns; that's what I am going to generate.

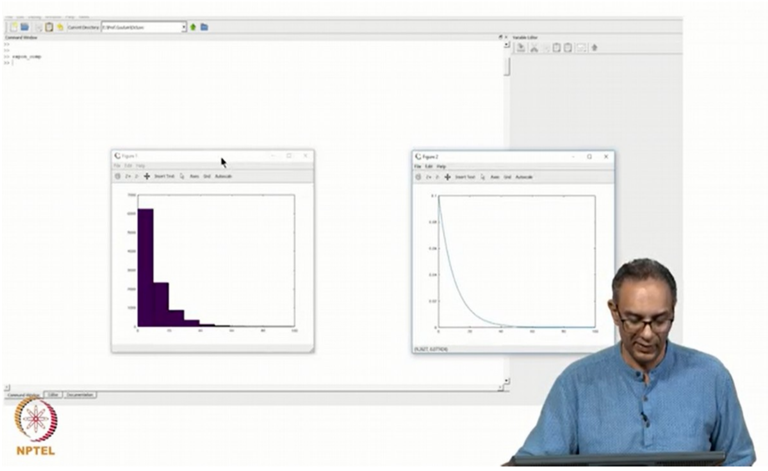
Then, what we will do is, we will draw the histogram. We will draw the histogram and the reason I wrote "figure(1)" is because we are going to do two things; we are going to both draw the histogram as well as the PDF little later. So, that's why I want to call something as figure 1 and then, histogram is histogram of x. So, what it does is, it takes the values of x and then, it computes the histogram. Now, I am going to also display a few of the values and I will tell you later, why it is crucial for you to do this exercise and then, I will also plot the PDF, okay. So, this one, "f=exppdf(0:b/10:10*b,b)", plots the PDF of various values of the exponential random variable. So, I plot the PDF, okay.

(Refer Slide Time: 14:17)



So, let's run this. I go to the command window and if you see the name of this file, it is called expon_comp; that means, exponential computation. So, I type expon_comp; I hit enter. It draws two figures and I am going to make them a little bit big, so that you can see them clearly.

(Refer Slide Time: 14:35)

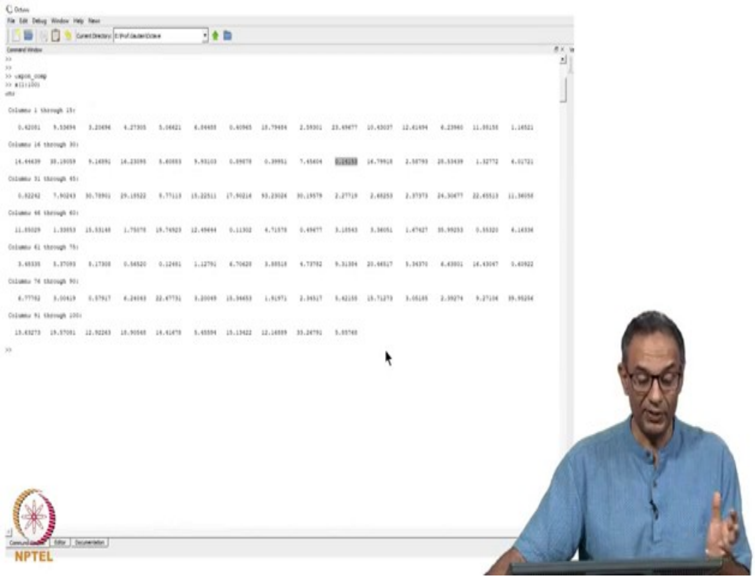


The figure on the left is randomly generated values. We have generated random numbers, 10000 of them, and then, drew a histogram. Octave necessarily puts it down 10 bins, and in each bin, it tells you how many there are. So, on the first bin between say 0 and 10, it has a little over 6000 out of the 10000 are small; like I said, there are many small values and as the numbers become larger, you get fewer and fewer.

So, if you plot the histogram, this is roughly how it looks like. Now, notice that the PDF looks exactly like the histogram. This is not a surprise; now, all distributions have this property. So, when you have data and you are taking data historical data and then, you plot the histogram and the histogram looks like this, what you would do say is, “Oh! Perhaps, this one would fit an exponential distribution. So, that’s the reason I wanted to show this graph, is because you look at this and say, “Haha! This exponential distribution, if I were to sample from an exponential distribution, I get my histogram like this”.

So, when I reverse engineer; so, I had data which looks like this. When I plotted the histogram, I can say, “Maybe, my population is exponentially distributed”. So, that would be a good guess.

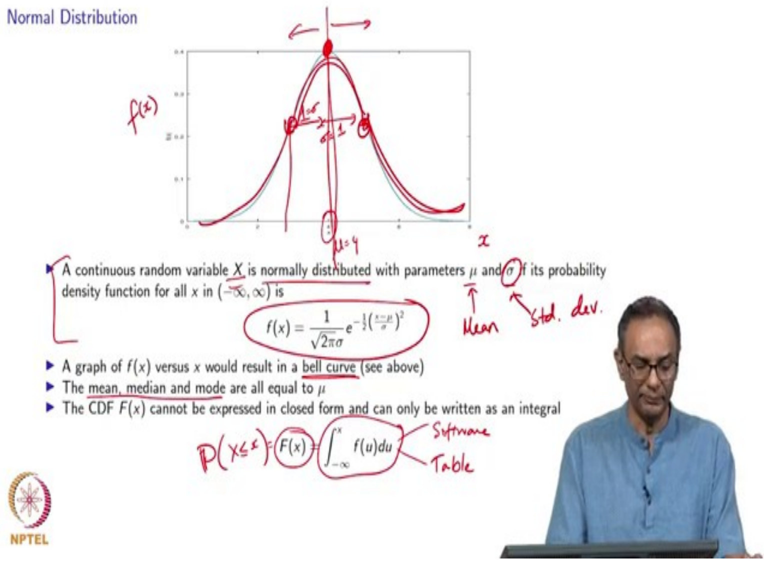
(Refer Slide Time: 15:56)



Now, I do want to also do the other thing. I showed you to do, x(1:100). So, let's see how the first 100 numbers look like. I know, this is a little bit small on the screen and you cannot see it very well. But, notice that there are some really tiny numbers like 0.42 and there are somewhat large numbers like 38; remember, the average is 10. So, remember the β was 10. So, on average, it is 10, but you see numbers much larger than 10 and much smaller than 10. So, if you look around, you see some numbers like 20, 33, 15 and some small numbers like 0.5, 0.3, 0.14 and so on. So, there are some small numbers; there are some large numbers. One of the things to remember is, exponential has a large amount of variability.

And, if you have not done something like this, I would recommend trying out the exponential distribution by generating these numbers and get a feel for how the numbers look like. Notice that when you look at the picture, there are some large and small ones. So, what I just did is this. I wrote "x(1:100)"; it tells me the first 100 values. I could write any other number as well and it tells me; you can notice how varying the numbers are. So now, let's continue with where we left off. So, we were here looking at the exponential distribution.

(Refer Slide Time: 17:16)

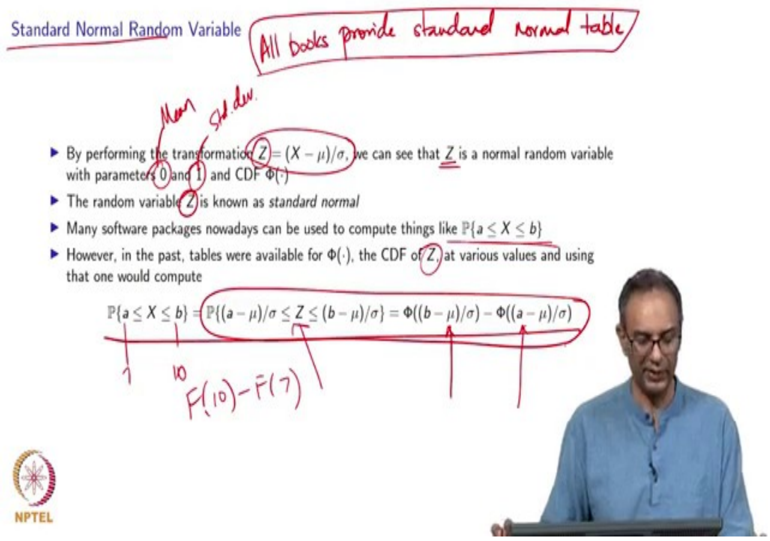


The next distribution that we are interested is the normal distribution. This is probably the most popular statistical distribution called the normal distribution. So, this is the famed bell curve. So, we see, this is a bell curve; that we see above is the probability mass function; this $f(x)$ is not very clear, this is $f(x)$ and this is x. If you the plot $f(x)$ versus x, it looks like this.

Now, if you notice, the top point is where the mean is. So, the mean is 4, $\mu=4$. Another thing that is important is that the standard deviation is 1; so, this distance is 1. So, $\sigma=1$. So, how do I know what is the standard deviation? Well, the standard deviation occurs at the points of inflection; this is where the figure goes from going like this to like this. So, it changes from convex to concave. So, the point where that happens is called the point of inflection and that happens at a distance σ from the mean. Coming back, a continuous random variable X is normally distributed with two parameters, μ and σ; this is most popularly used for normal. Typically, μ is the mean and σ is the standard deviation. For that to be true, the probability density function should look like this. So, $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$. So, this is a very well-known bell curve. We will see a lot of the normal distribution throughout this course. Now, the nice feature about the normal distribution is that the median, the mean and the mode are all equal to μ. That's because the highest probability, the mode happens at the same value μ; we know it's the mean, it's symmetric about this. So, if you put a little mirror here, what you see on the right is a reflection of what is on the left since it is symmetric, it is also the median and it also happens to be the mean. Now, σ is the standard deviation.

Unfortunately, the CDF, $F(x)$ is the probability that the random variable X is less than or equal to little x; that quantity cannot be written in closed form. That means, I cannot write a formula; I can only write it as an integral. We will be computing this in two ways. One is using a software; that's what we all do these days; but you can also use a table and that's what most books use because some of the books were written many years ago when our software was not as prevalent. So, I will try and touch both styles.

(Refer Slide Time: 20:19)

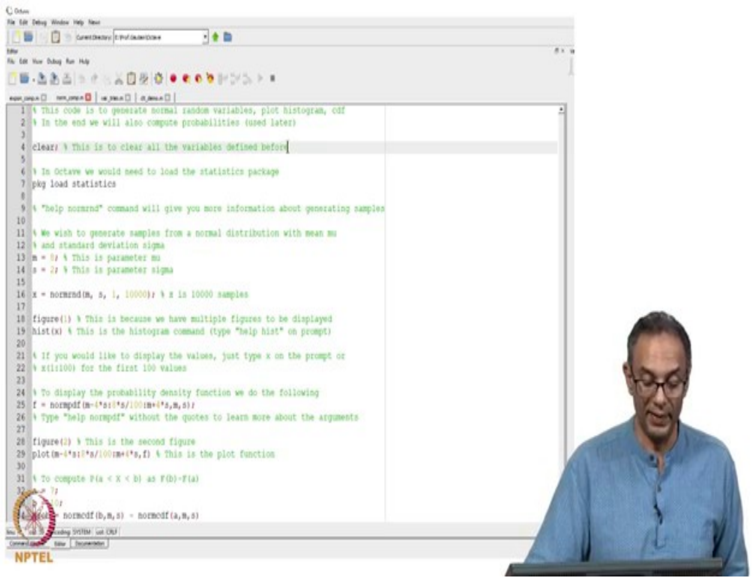


In particular, the tables are for standard normal random variables. So, all books provide what is called the standard normal table. So, what we do is, we perform a transformation; we subtract μ from X and divide it by σ, that is $Z=\frac{X-\mu}{\sigma}$. What it lets us do is, if you subtract μ, then the mean or the expected value, the mean of Z becomes 0, and the standard deviation becomes equal to 1.

We will see those topics later because we have not talked about the properties of mean and standard deviation; so, we will get to that later. So, turns out that Z is a very special normal random variable, where the mean is 0 and the standard deviation is 1. We like that because now, for this guy, we can provide a table. So, if you look at any textbook, there will be tables for what is called the standard normal random variable; this Z is called the standard normal.

So, there are many software like the Octave that we will see very soon, which will actually compute the probability, $P(X \leq b)$. Like I said earlier, historically this was computed using this z table. I am going to show you both using Octave. So, I am going to go back and use Octave.

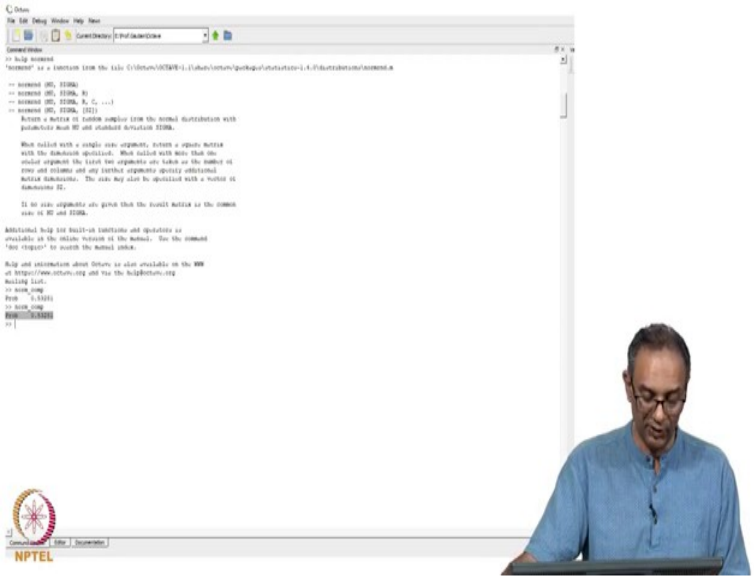
(Refer Slide Time: 21:54)



Now, I am using the second program; this is called norm_comp. So, this program is to generate, like we did before with exponential; now, we are going to generate normal random variables; we plot the histogram just like we did for exponential. And then, we will compute the cumulative distribution function and then, we will also compute the probabilities that I talked about. I said we used it later. So now, what I decided was to present everything together.

So, the first step is, I clear everything; now, this is a fairly decent practice I would recommend putting “clear” because sometimes what will happen is, the variables that you defined earlier might creep in right now and mess you up. Again, we need to load the statistics package. Again, if you want to learn more about the normal random variable, type “help normnd”.

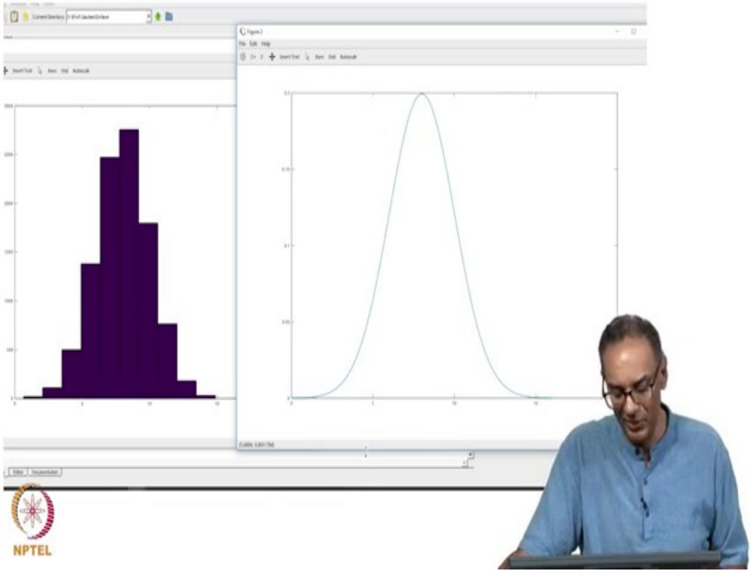
(Refer Slide Time: 22:44)



So, you can go here and then type, “help normrnd” and then hit enter; it will give you a little help menu of the normal random variable. So, it will do that for you. So, now what I am going to do is, I am going to generate some samples from a normal distribution. So, Octave using the statistical package, will generate that for me. I am going to pick μ as 8 andσ as 2. And, I am going to ask it to generate 10000 samples in 1 row, and I am going to ask it to use mean μ of m and standard deviation σ of s.

Again, we are going to have two figures. I am going to try to close the other two figures so that we do not get confused which figures we are talking about. I am going to compute the PDF and I am also going to draw the histogram. So, histogram here and PDF here. Then, once we have time, I am going to do the last one towards the end, is to compute the probability. I will come to that later; so, first let’s do that. Like I said, the title of this is called norm_comp.

(Refer Slide Time: 23:56)



So, I type here, "norm_comp" and I hit enter. And, it draws me two graphs and I am going to blow it up a little bit; so, this is my normal. So, notice again how the normal random variable, the PDF of the normal looks a lot like the histogram we collected. So, we collected samples plot of the histogram; it looks a lot like the normal. Now, if you generated another 10000, it might look slightly different from this, but overall, it has this shape. So, whenever you see a shape like this, you kind of know looking at your PDF, the bell curve, you will say, "Well! Maybe, we should fit a normal distribution".

So, that is what I want to say about normal. So, you can play with this and try different values of μ and σ if you would like. I also recommend just even trying the same program multiple times which will give you a different output; so, if you did this the second time, it will give you a different set of different figures possibly. The right figure, of course, should always be the same; the left run could be slightly different. So, that 's something to think about.

Now, this is one thing I do want to talk about; this gives me the probability. So, if I want to compute the probability that the random variable is between a and b, then what I do is, say picking a as 7, b as 10, and I compute the probability that x is between a and b, $P(a \leq X \leq b)$, by computing the CDF at b minus the CDF at a, that is, $F(b)-F(a)$. Now, this is exactly what we had in the formula before and I did my computations and it gives me that that probability is equal to 0.53281.

So, if I go here and I am going to hit view, this computation at the very end is basically what we did in that. So, we said, what is the probability that X is between 7 and 10 and we computed it as $F(10) - F(7)$. Instead, I could have alternatively, we will see this later in the course; I could have alternatively computed by converting it into the Z random variable which is also normal with mean 0 and standard deviation 1, and then, use the table to compute these. I could have done that as well.

However, we decided to just use a software per se, but I would recommend that, if you have a table and do not have a software with you, to go ahead and make this transformation. When we do deal with normal random variables, we will visit this in the future.

Thank you.