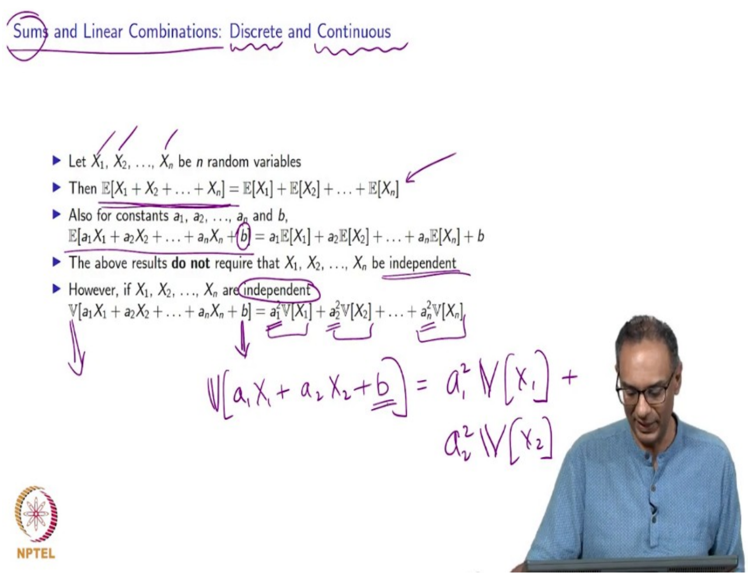
Decision Making Under Uncertainty
Prof. Natarajan Gautam
Department of Industrial and Systems Engineering
Texas A&M University, USA

Lecture – 08
Multiple Random Variables: Discrete and Continuous

Now, we come back to discrete and continuous random variables and look at what happens there are multiple random variables, not just one; so far, we have been looking at just one random variable. Now, let's see what happens when you had multiple of them.

(Refer Slide Time: 00:31)

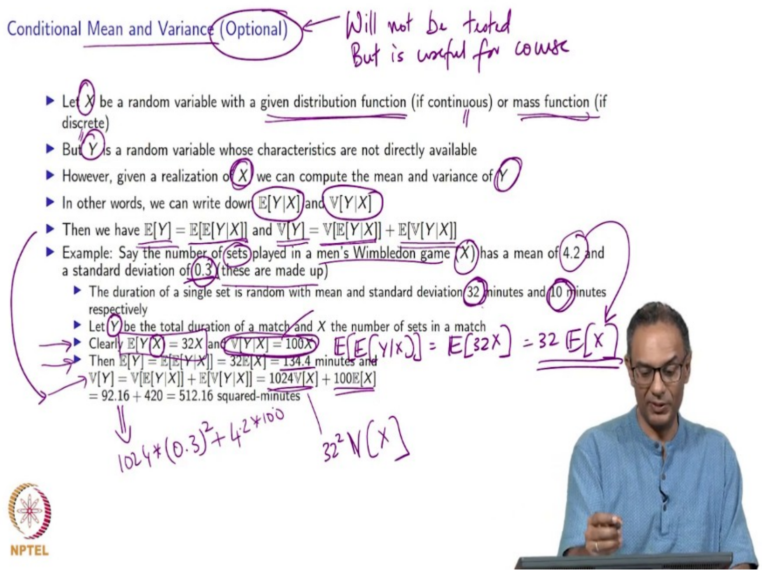


So, there are some important things because we are going to see a lot of these multiple random variables. So, let's say a little bit about it. So, the first thing we will talk about is sums and linear combinations. Again, I want to emphasize we include both discrete and continuous. It appears like we spoke a lot about continuous last few slides and now, we are back to discrete and continuous.

So now, let's say there are n random variables X_1, X_2 and so on till X_n; that's all I am saying. these are n random variables, could be discrete, could be continuous, does not matter, it could be a combination if you would like. Then, the expected value of the sum of the random variables equals the sum of the expected values, $E[X_1+X_2+\ldots+X_n]=E[X_1]+E[X_2]+\ldots+E[X_n]$. So, that's what this result this is. It's a very powerful result and in fact, it is extraordinarily useful. We won't do much of the usefulness of this in this course, but I do want to say that is very important. So, it is the first result on sums. You also have a result for linear combination. So, if you take the expected value of a linear combination, $E[a_1 X_1 + a_2 X_2 + \ldots + a_n X_n + b]$, then what happens is that a_1 and then a_2 comes outside and so on. So, the expected value gets the nice linear combination style, $E[a_1 X_1 + a_2 X_2 + \ldots + a_n X_n + b] = a_1 E[X_1] + a_2 E[X_2] + \ldots + a_n E[X_n] + b$. So, in these results what is really crucial is the random variables do not even have to be independent. So, that's an extraordinarily powerful result. Now, we use this quite a bit both in this course as well as in life. Now, if they are independent, then actually it gives you a much stronger result which is for variance as well as this. If they are not independent, this result is not true for variance. Notice that this b is gone; that's because the variance of a constant is 0, okay.

So, the variances also have the similar property except remember that the variance is random variable squared. So, all the prefixes a_1, a_2 and so on need to be squared. Basically, for example, if I were looking at the variance of $[a_1 X_1 + a_2 X_2 + b]$, then that is equal to $a_1^2 V[X ¿¿ 1] + a_2^2 V[X ¿¿ 2] ¿¿$, where $V[X ¿¿ 1] ¿$ denotes the variance of X_1 and $V[X ¿¿ 2] ¿$ denotes the variance of X_2; that's it, the b goes away and vanishes because b is not something which varies, okay. So, the square of the expected value equals the second moment and so on.

(Refer Slide Time: 03:21)



Now, what we will see next is this conditional mean and variance. And, notice the word "optional" here. Now, this is important; that means, this will not be tested. So, this topic will not appear in a test; however, these results that look quite nasty is going to be part of the lecture. So, it will not be tested, but is useful for the course. So, if you are wondering how certain results are derived clearly, you would need to learn this. So, I would recommend listening to this, but not worrying. So, how I would interpret this "optional" is, it's not well I will stop my recording and then go to the next one; no, don't do that. I want you to listen to this completely and then not be too worried, okay. So, that's all I am saying.

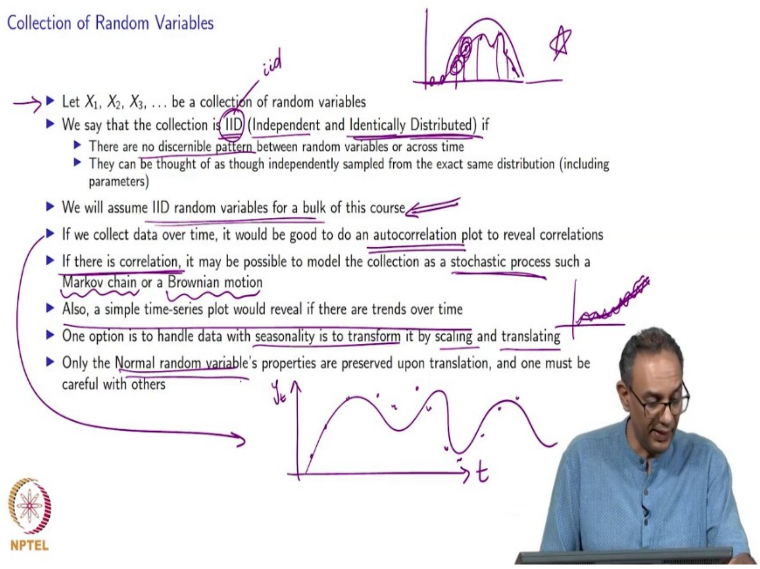
So now, let's do the following. So, let's say we have X as a random variable whose distribution function is given if it is continuous or the probability mass function is given if it is discrete. And, Y is a random variable whose characteristics is not known. I do not know the distribute function or the mass function.

But, if we know something about X, we can say something about Y. So, that is the only good thing about it. So, in other words, I could compute the expected value of Y given X, $E[Y \vee X]$ and the variance of Y given X, $V[Y \vee X]$; not to worry, we will do a little example in a short while. Then if you want to compute the expected value of Y, it is written as the expected value of the expected value of Y given X, $E[Y]=E[E[Y \vee X]]$. We will see an example; it will become a little bit clearer. Likewise; variance of Y is the variance of the expected value of Y given X plus the expected value of the variance of Y given X, $V[Y]=V[E[Y|X]]+E[V[Y \vee X]]$. So, that is important.

Let us do a little example. This concept will become a little bit clearer. So, let us say we have a Men's Wimbledon game and the number of sets played, let us say is the random variable X. So typically, you have 3 sets, 4 sets or 5 sets. So, let us say the mean number of sets is 4.2 and the standard deviation is 0.3. These numbers are completely made up; this is not taken from any Wimbledon site.

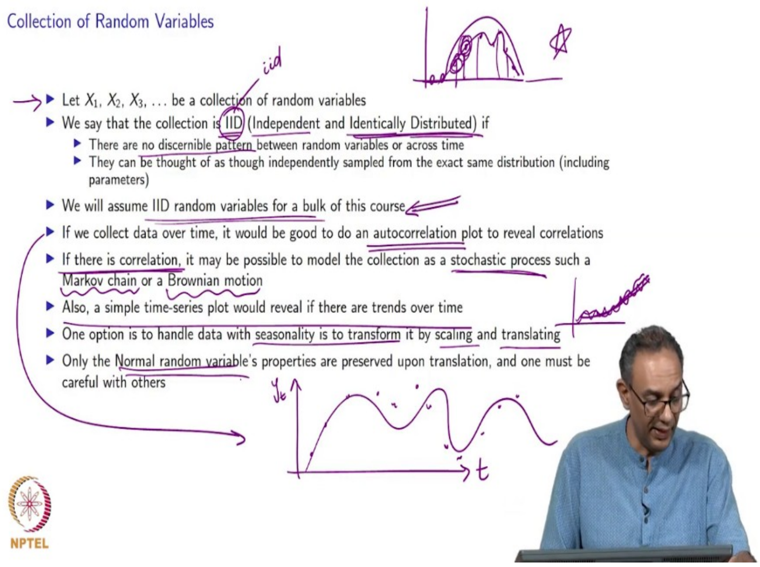
Now, the duration of a single set is also known to have a mean of 32 minutes and a standard deviation of 10 minutes. Remember, X is the number of sets. So, what you want is what is the mean and standard deviation or the mean and variance of the entire match? So, match has X sets so, and match totally takes Y amount of time. So therefore, we know how to compute these. So, let's look at this, $E[Y|X]=32X$; why is that the case? Because each game takes a X, I mean it takes 32 minutes on average. So, if you are given X, if you are given the number of games is equal to X, if that is given, then the expected value of Y which is a total duration is 32 times the number of games which is X, and then the expected value of that. So, the total duration is, $E[Y]=E[Y|X]=32E[X]$. So, the expected value of; the expected value of Y given X is expected value of 32 X, which is 32 times the expected value of X; here, $E[X]$ is 4.2. Likewise, the variance of Y given X is this. So, we get this in a very similar fashion. If we know there are X sets and each set has a standard deviation of 10 minutes, then $V[Y|X]=10^2X=100X$; so, that is going to be the variance if you give an X. Now, the expected value of Y is the expected value of the expected value of Y given X and that is $32E[X]$ which is equal to 134.4 minutes. And, the variance of Y uses this formula here, $V[Y]=V[E[Y|X]]+E[V[Y \vee X]]$. So, we have those two numbers here from this one and 1024 is basically the variance of the expected value of Y given X which is given here. Expected value Y given X is, $E[Y|X]=32X$; So, the first term is, $V[E[Y|X]]=32^2V[X]=1024V[X]$. Now, if you plugged in the numbers, the first term is 92.16 which is basically 1024 multiplied by 0.3^2 because this is the standard deviation. The second term is, $E[V[Y|X]]=100E[X]=420$ which is 4.2 times 100. Then, $V[Y]=92.16+420=512.16\,squared\,minutes$; so, that is the variance. So, that is how we calculate expected value by conditioning and the variance by conditioning. Now, like I said once more, this is an optional.

(Refer Slide Time: 09:24)



Now, what comes next is not optional; this is required for the course. We will use it several times during this course; it's an important topic. So now, let's say X_1, X_2, X_3 and so on, this is a long set of random variables and we say that this collection is IID which means Independent and Identically Distributed. Some books write this in lowercase, iid; that's also perfectly legitimate. What is the meaning of IID? Well, that means there is no pattern. The random variables go all over the place; there is no pattern; it is not like the numbers increase and decrease. And, they could also be thought of as samples from the same distribution; so, that is the identical distribution. So, they have to be independent; that means, each sample is not dependent on the previous sample or the ones before that and so on. And then, they are also identically distributed.

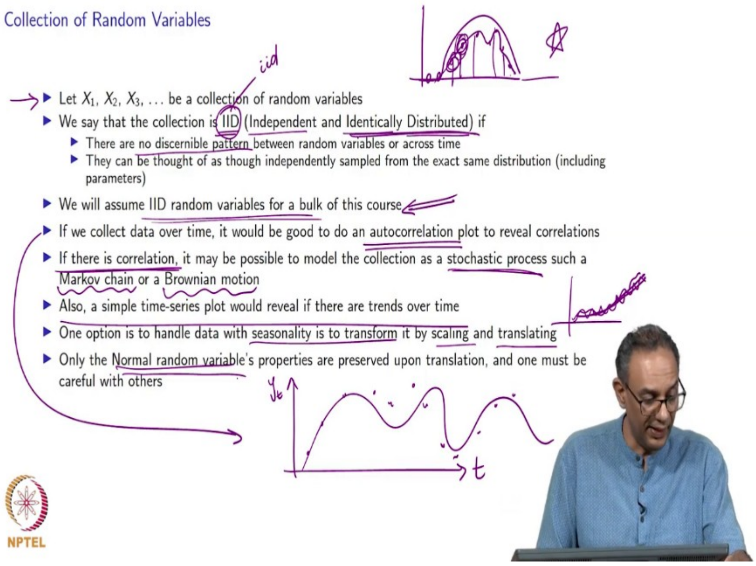
(Refer Slide Time: 10:17)



So, recall the example that we had on the solar power. So, it kind of looks like this and then, the random quantity, let's say every hour you took and this was good and then, the values start to go like this, like this, like this. These numbers are not independent or identically distributed as a matter of fact. They are not independent because this number probably depends on the time of the day. And, this number is expected to be smaller in general than this number and so on. Also, they are not identically distributed; these guys are all different distributions. So, that is an example where something is not. So, you do see some patterns here. So, that is one of these situations and also, they are not from the same distribution. So, what we want is IID. So, that is one special type of collection of random variables; for most of this course, we will assume that random variables are IID.

So, that's an important characteristic. So, I may not come back to this. I might just say the random variables are IID. I just want you to think that; that means, there is no pattern and they are identically distributed.

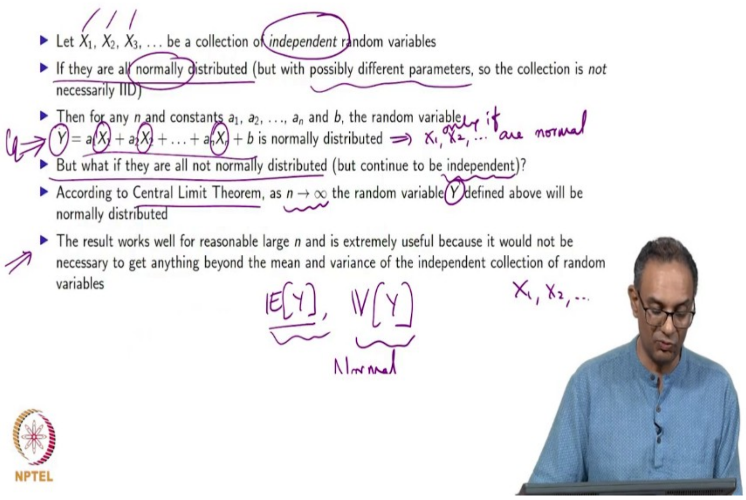
(Refer Slide Time: 11:23)



So now, this point is important and I want to pick a little picture. So, whenever you collect data and like I said, in practice, one would collect a lot of data and let's say you collect a time series and then, you collect some points. So, let me just call this y_t. So, the first point is here, second point is here, the third is here, the fourth is here, the fifth is here, the sixth is here and then, comes down over time. So, if you see that, there is a pattern; then, you would say, "Oh! Wait a second. I am not sure this is IID". However, if you don't, then say maybe the values are IID. Now, it's hard to look, get a pattern looking at data like this; that's difficult. Therefore, what one normally does is, use what is called an autocorrelation plot. So, you do this plot and see if correlations are revealed and we are not going to talk about that in this course, but a lot of standard packages typically have the autocorrelation function; go ahead and plug it in and see if there are correlations.

Now, if there is correlation, what do you do? It is not hopeless. If there is correlation, you can do one of two things. Basically, one common thing you can do is to model it as a stochastic process such as a Markov chain which we will look at in the very last part of the course; or a Brownian motion which we will not see in this course, but a lot of times for example, the stock prices are modeled as a geometric Brownian motion. This is something that that one does when there is correlation. Another thing that one can do is to model things as a multivariate distribution and that is the second way of doing this; it's little bit beyond the scope of this course; so, I will not talk about those. Another thing that I would say is always plot a time series plot to see if there are trends over time. Sometimes, some values will slowly start to increase over time, it is like seeing trends going up and then, in which case, you want to catch those. So, whenever you get a time series data, plot it and see if there are trends. Now, this is a very useful tip that I would like to give you all. Now, if one has the option of doing this time series which I would highly recommend, then if you can transform the data by scaling and translating, then it is possible to handle; otherwise, time varying data is not very easy to handle. So, the most common way of handling time varying data or trends is to transform or scale or translate in some way. So, for example, if you look at this, if you plot like this along this curve things about that curve things look to be somewhat IID. So, that's roughly what people do. One careful thing is the normal random variable works really well for scaling and translating; not all random variable scale and translate nicely. Especially, translation is a big problem for many; scaling as well, many random variables are okay with that. So, you have to be a little bit careful. You cannot just say, we will do scaling or translating. I want you to be somewhat careful.

(Refer Slide Time: 14:31)

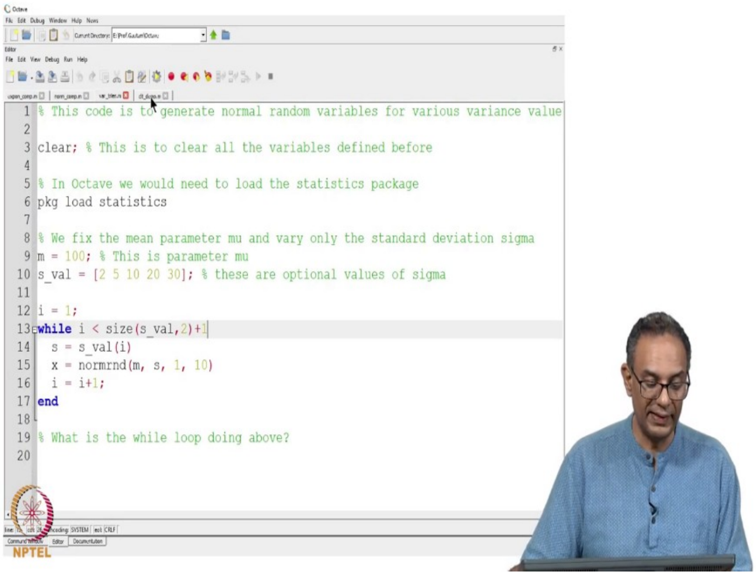


Let's move on. The last topic which we will also do a little demo, is the topic of central limit theorem. Now, this is an important topic because two of our results that will come in the future use the CLT, Central Limit Theorem. Now again, we have a set of random variables, but we are only assuming that they are independent; we do not want IID. IID would be great, but it is not necessary; independent is good enough. So, X_1, X_2, X_3 are all independent random variables. Now, if they are all normally distributed even with different parameters, then life is easy. So, we don't have, need IID, but they are independent and normally distributed; then, no problem. I have a random variable Y which is equal to $a_1 X_1 + a_2 X_2 + \ldots + a_n X_n + b$. Then, this Y itself automatically is normal; for this requires X_1, X_2 and so on are normal. This is only if they are normal. Otherwise, what happens? Well, turns out that it's not too bad.

Because, what if they are all not normally distributed, but they are still independent; then, what do you do? Well, turns out the central limit theorem comes to our rescue; tells me that this random variable Y that we had here, is actually still normally distributed as long as n is large, it goes to infinite. But, it works for fairly reasonably sized values of n. The question that I get asked a lot is, "What is a good value of n?". That is difficult to say. It will work for smaller values of n when the shape of the PDF looks similar to normal. When it looks farther away from normal, it will take you a little bit more. Now, what if you had a bunch of discrete random variables, it could still work; you need a little bit more. So, you really need more and more as you know, the result where it basically works only for infinite. But, as you get larger and larger values of n, it will work reasonably well.

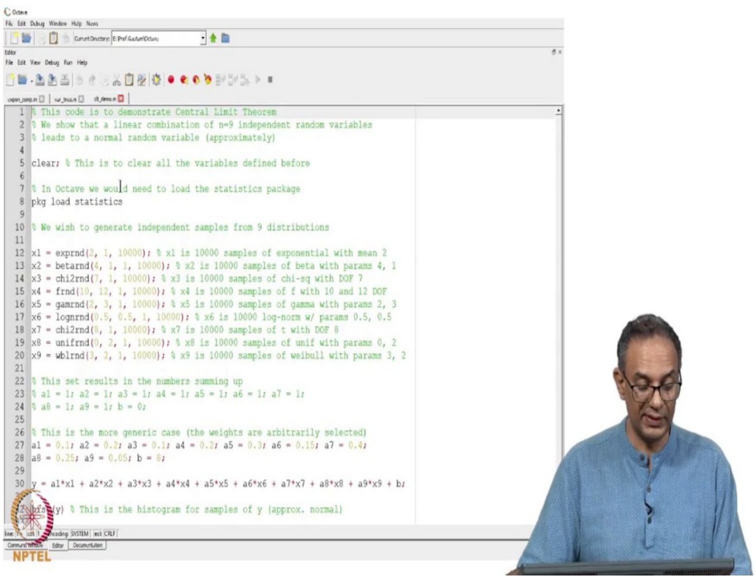
Now, turns out that this is extremely useful. And, I would say that and then, I will do a little demo in Octave. But I do want to show that this is an extremely useful result. Because now, all we need is not the entire distributions of X_1, X_2 and so on, but we just need the mean and the standard deviation. Once we have that, we can compute the expected value of Y and we can compute the variance of Y easily using the results we had before. Once I know this, that is all I need for my normal distribution. I do not really need anything beyond the mean and the variance of your X_1, X_2, X_3 and so on. So, for X_1, X_2 and so on, all I need is the mean and the variance. So, let's do a little demo of that on Octave.

(Refer Slide Time: 17:34)



So, I go back to Octave and what I am going to do is, let's not look at what we saw before; what I am going to do is, I am going to look at the last. So, notice that I have four. So, I click on the bottom, it says editor. The first one is the first program that we saw.

(Refer Slide Time: 17:45)



Then, I am going to now do the clt_demo, the Central Limit Theorem. So, this code is to demonstrate what we call as central limit theorem. So, what we are going to do is exactly what we saw in the previous slide which is a linear combination here and as always, I would like to clear and call the statistics package.

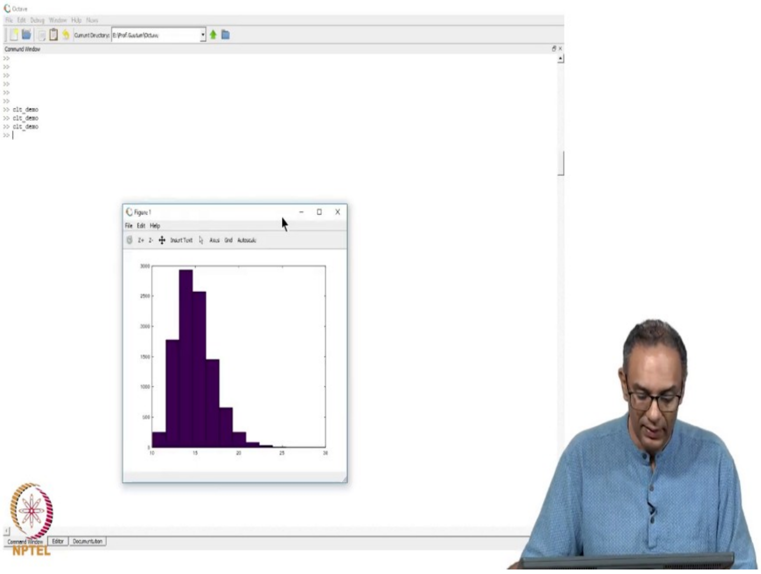
This is what I am going to do. I am going to create n equals 9 random variables. So, 9 random variables are $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9$. I am picking a bunch of different distributions; one is exponential, one is beta, chi-squared, F-distribution, gamma distribution, lognormal distribution, t-distribution, uniform, Weibull.

So, all these distributions and some of these, exponential, beta and the uniform look very different in terms of the shape of the PDF compared to the normal. So, summing them up is truly somewhat diabolical where you will see what happens. So, now I do want to say a couple of quick things. Now, when we looked here and we wanted the sum, $a_1 X_1 + a_2 X_2 + \ldots + a_n X_n + b$. If you just wanted the sum here in this in this expression, then I will just put $a_1 = a_2 = \ldots.. = a_9 = 1$ and $b = 0$; that will give me the sum.

So, that is just another case. You may want to try when you get a chance to play with this. However, I am going to pick an arbitrary set now. I just really arbitrarily selected these values, $a_1 = 0.1, a_2 = 0.2$ and so on; $b = 8$. You can select any other value; the results do not change. I just really arbitrarily pick these numbers.

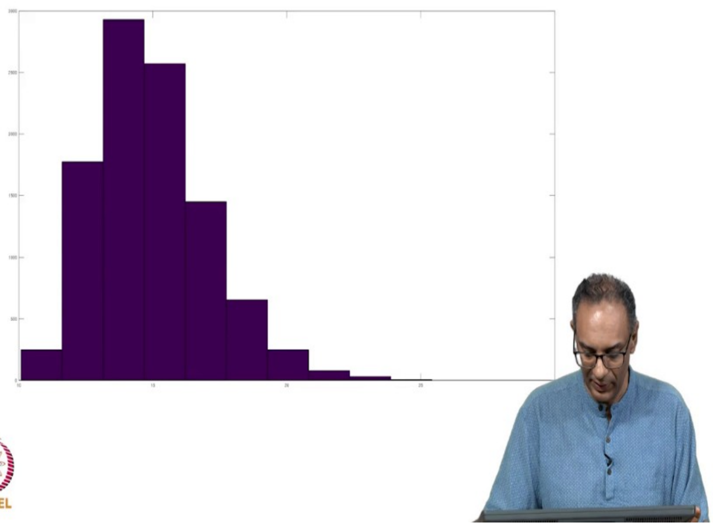
So, I write down y as the linear combination, $y = a_1 x_1 + a_2 x_2 + \ldots + a_n x_9 + b$. So essentially, what it does is it takes a randomly generated value of X_1 from exponential distribution, throws it in here, adds it to the X_2 random number and then to X_3 and so on. So now, y is again 10000 samples, and if we plot the histogram of y, my claim according to central limit theorem is that, this y guy is going to be normally distributed. Let's see if that happens. I am going to write clt_demo and see how it works out.

(Refer Slide Time: 20:01)



So, I write down clt_demo.

(Refer Slide Time: 20:11)



So, if you plot this, it looks nice and pretty. I am going to auto scale so that it looks nice and big. So, let’s do this one more time; let’s try this again; let’s do clt_demo again. So, this is when I say, it is a good idea to actually write “clear” because now, it just clears the value first and draws some picture.

Now, notice that this guy is a little bit skewed on one side. So, this could happen a lot because you are really adding up not too larger numbers. You only adding up 9 random numbers; that is not largest that we have seen. So, let's try it one more time and then, I will stop. Now see that it does have that nice shape, it's kind of normal looking. So, when I added up these random numbers, I get a nice normal looking distribution, alright. So, that brings us to the end of this topic.

Thank you.