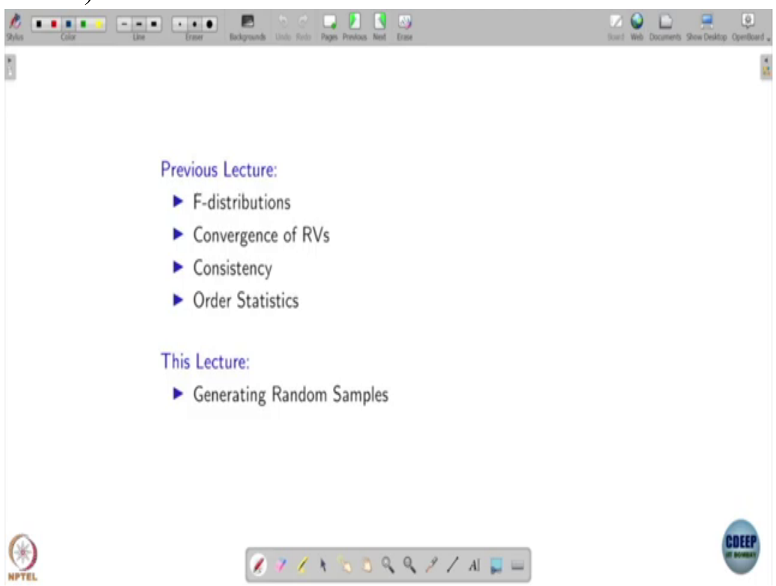
Engineering Statistics
Professor Manjesh Hanawal
Industril Engineering and Operation Research
Indian Institute of Technology Bombay
Lecture 27
Generating Random Sample-Direct method

(Refer Slide Time: 00:21)

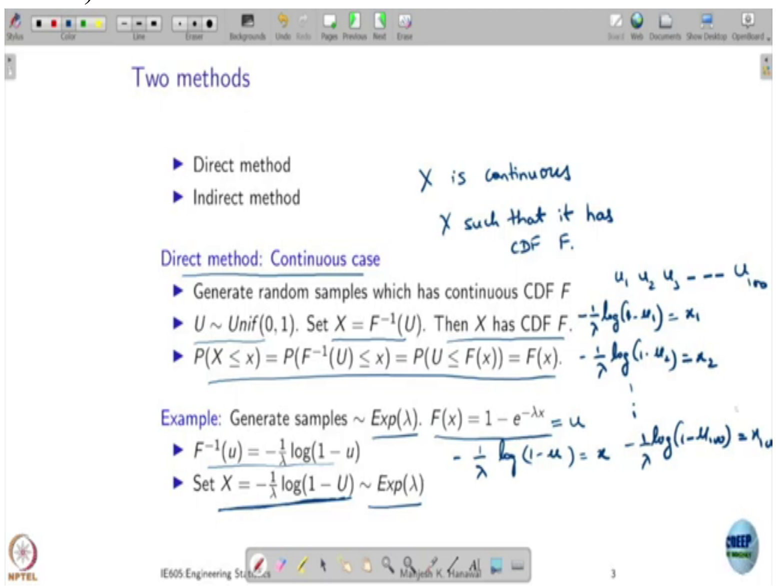


So, so far we have discussed this F distributions, we talked about convergence of random variables and we talked about consistency and then talked about order statistics. In the last class we talked about order statistics and discussed how to find distribution of an order statistic when it is discrete. We just said that let us say if my random variable is discrete that is it takes values in some finite set or some countable set then if you are interested in knowing what is the smallest value, the distribution of the smallest value, distribution of the second smallest value, distribution of the second value or the distribution of the maximum value.

Like that, we discussed one way to find this distribution. If you recall, we defined some Bernoulli random variable we took the summation and based on that we did some computations. So, that idea can be extended even when the random variable is continuous. I just did not discuss that, but the expression of that expression for that is given in the slides and you people can verify how to get the expression by going through the computations yourself or refer to the book for the computations the all the computations are given in the book.

So, today we will go ahead and just cover one more topic about generating random samples. So, this actually we already did, like this topic of generating random sample we have discussed earlier. Do anybody remember when was this? Yeah, we try to find samples or generate sample according to a given distribution using uniform samples by defining inverse of that given CDF function. Now, let us see if we can little go beyond that.

(Refer Slide Time: 03:07)



So, on that front we are going to discuss two methods called direct and indirect method. So, direct method is what we already know. So, in direct method also there could be different possibilities and one possibility we already know. Suppose, let us say I have been given x which is discrete or let us say X is continuous. Now, I have been asked to generate X such that it has CDF of F. Now, you know how to do this.

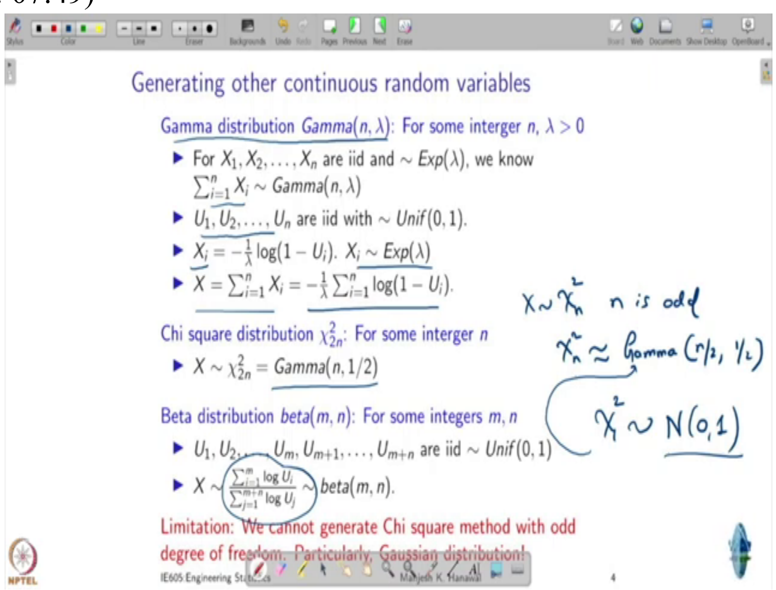
You had to start with a uniform random variable and set your X to be F inverse of U. Once you do this, you already know that X has a CDF of F. And this is argument for that, we have already discussed this. Next, let us look into a specific example. Suppose, I have been told that I want to generate a random variable that follows an exponential distribution with parameter lambda. So, for this I already know that the CDF I want is given by this expression, everybody agree that the CDF that I am interested in is F of X equals to 1 minus e to the power lambda X.

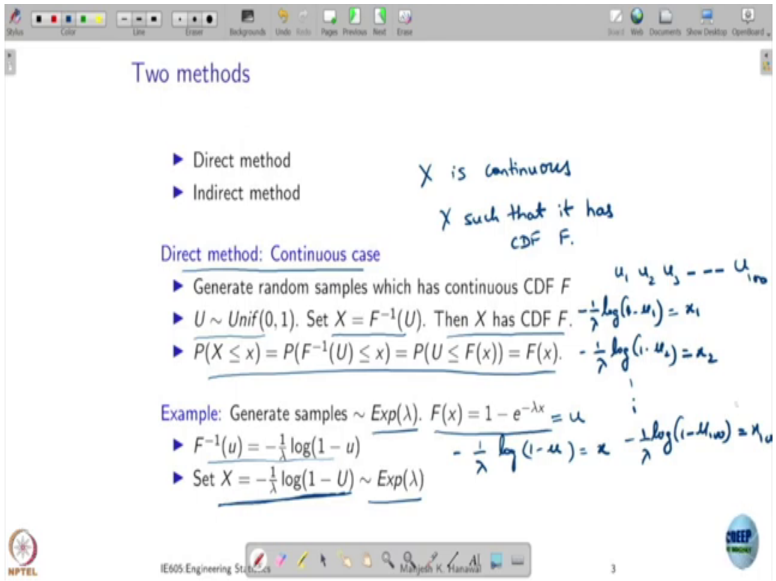
This is a CDF for my exponential distribution. Now, I need to define the F inverse for this function. So, what is the F inverse for this function? Suppose, let us say you call this some value U, yeah, then you can write F inverse of U like this. All you need to do is you just take it, it becomes 1 minus U, then the X that gives me this particular U is, can be written as log of 1 by lambda with a minus.

So, this will give me that X for which will use this particular value U that I am interested in. So, now I know this is my inverse function. Now, all I have to do is take your X to be minus 1 by lambda log of 1 minus U. If you take this, we already know that this is going to follow exponential distribution with parameter lambda. Everybody agree that if I do like this, it is going to give me this X if I do like this x is going to have exponential distribution with parameter lambda.

So, by this what I mean here is what does this mean suppose, lets say, U1, U2, U3 and the how generated, let us say some 100 samples. And now, what I am doing is I am now doing 1 minus lambda log of 1 minus U 1 and then 1 minus lambda log of 1 minus U2 and I am calling this as X1 and calling this for X2 like this and 1 minus lambda log of 1 minus U100 this has X100. So, whatever the new samples I got X1 X2 X100 they are following the exponential distribution with parameter lambda.

(Refer Slide Time: 07:49)





So, exponential was simple. Now, let us see using a simple method, what else we can generate? Let us see, is it possible for us to generate gamma distribution with parameter n and lambda. For some integer n and some non zero value of lambda. Now, given that I know how to generate samples from exponential distribution just as we did in this example, let us try to see that we can leverage this.

So, I know something about exponential, how to generate exponential distribution easily. But if I can do that, maybe I should attempt to connect my gamma distribution with the exponential distributions and try to exploit that relation. But we know that there exists a relation between gamma distribution and exponential distribution. Suppose you have X1 X2 Xn these are all iid exponential distribution with parameter lambda. Then we already know that their summation is gamma distributed with parameter n lambda.

So, then, what is one way to generate gamma distribution? Tell louder, just generate n exponentially distributed random variable and then add then it will become a gamma. But to generate exponential distribution, how I did? I generated uniform distribution and then use it. So, I am going to just use this. So, use of this iid random variable, first I am going to generate them following making them follow exponential distribution using my uniform random variables.

So, I am going to generate uniform random variables U1 U2 upto U n and set Xi is equals to minus 1 by lambda log 1 minus u. And I know that if I do like that Xi is exponentially distributed with parameter lambda. And now I have this Xi which are exponentially distributed lambda, just add them then I am going to get this X to be gamma distributed. So, if I had to directly write this in terms of my uniform random variable, I have to do is generate my uniform random variables.

And then add them after taking the log and doing by scaling with 1 minus lambda then whatever I get is directly gamma distributed with parameter and n lambda. You see that I am basically used, I am playing around with the properties. And then applying these functions of random variable properties to extract all these desired properties. Now, I know anybody has any question on how to generate this gamma distribution now?

Now, once I know gamma, maybe I can also go and generate a chi square distribution because gamma distribution and chi square distributions are related. So, I know that chi square distribution with the 2 n degrees of freedom is nothing but a gamma distribution with parameter n and half. So, to generate a gamma distribution with parameter n and half what should I do? Exponential with what parameter? Half.

So I need to generate n number of exponentially distributed random variable with parameter half and then simply add them, that will directly give me chi square distribution. And now such relationship be exploited to even generate other distributions. Like suppose, if you want to generate beta distributions with parameter m and n, how you are going to do that? You are going first exploit the relation that X, If I am going to take log of Uis where I is running from m.

That is basically what I am trying to do here is I am taking the average of two sums. In the numerator, I am taking the sum of m uniform random variable ask applying the log function and then the denominator I am taking sum of all the m plus n uniform random variables after applying the log function. We know that if I do this transformation on the uniform random variable it is already beta distributed with parameter m and n.

So, all you readily got like just the directly by taking uniform random variable and applying this transformation you readily got beta distribution. So, this method is nice, we are basically exploiting the fact that the relation between one distribution with the other. And as long as we could represented one of them, generate one of those random variables through a uniform random variable like we are done.

Because if it for example, here gamma distribution depends on exponential and I know how to generate exponential using uniform random variable. So, you either gamma chi square or beta, everything can be generated using only uniform random variables. And uniform is something simple, which we can assume to be available to us, and then just use these properties to get this distribution. But now is this trick works all the time?

Maybe it may not work all the time. One simple example I given is if I want to generate let us say X is X1 square where n is odd. Now, in this case X1 square is nothing but gamma with what parameters?

Student: n by 2.

Professor: n by 2 and this we know if I had to write it as a sum of exponential random variables I have to add n by 2 exponential, but when n is odd, can I do that? No, you cannot add 10 by 2 number of exponential random variables when n is odd.

So, then that times this simple method does not work. And yeah and what is the relation between chi square distribution and gamma distribution did we study that? Anybody recall what is the distribution between or like relation between Gaussian and the chi square distribution?

Student: Square of this.

Professor: I want to generate chi square distribution with one degrees of freedom. What is the relation between chi square distribution of one degrees of freedom and Gaussian distribution?

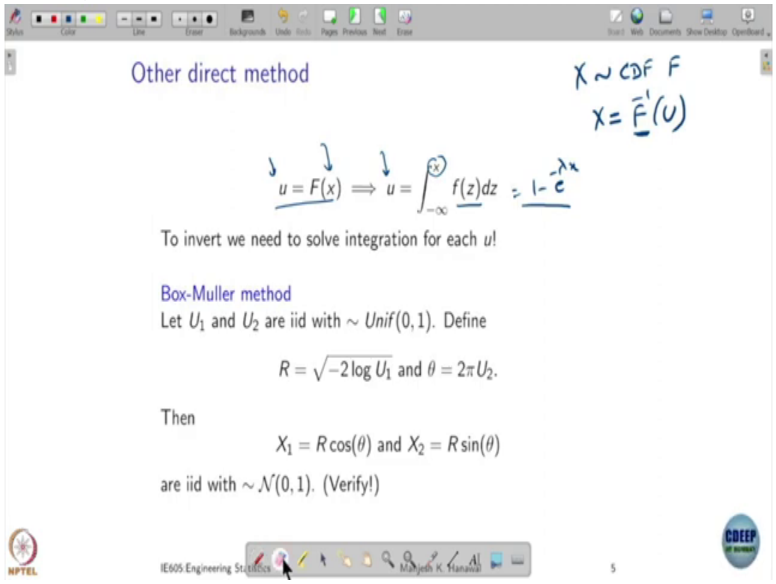
Student: Square root of Chi-square.

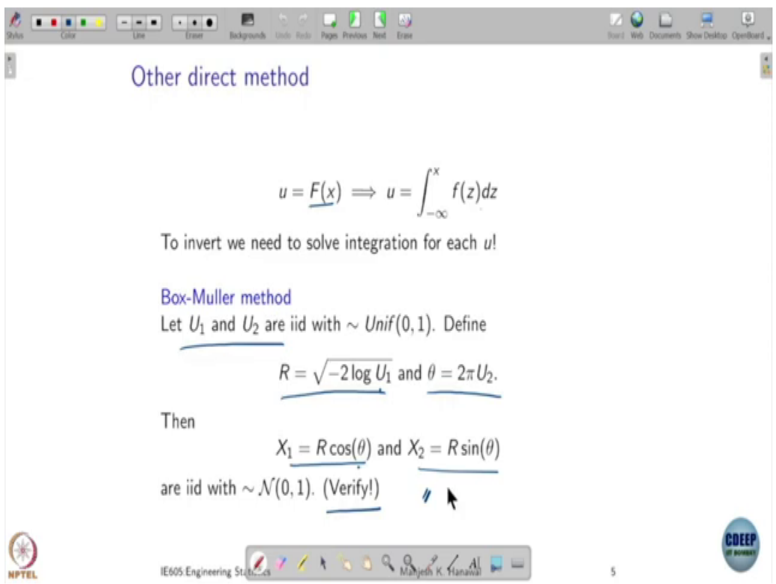
Professor: This is this is normal with what? What parameters?

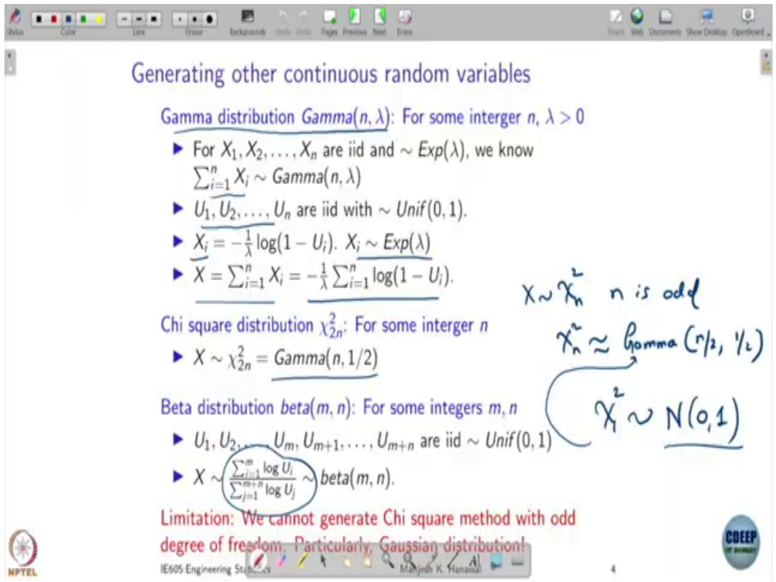
Student: 0, 1.

Professor: Now, see now, this chi square distribution with one degrees of freedom I know that this I cannot generate with my gamma method because this degrees of freedom is odd. So, because of this I cannot even generate the Gaussian distribution in this case. Is this clear to all of you? Now, we need to overcome this what other methods we have?

(Refer Slide Time: 17:39)







So, we will just explore one other direct method now. By the way if you have to the previous methods, the previous direct method we applied, we always needed to do F inverse. If you want to generate a X according to which is has the CDF, F that I needed to find, I needed to set F equals to F inverse of U. So, you need to find what is that F inverse function? But always finding F inverse is not an easy task.

Maybe something like exponential it was easy because of its simple structure. So, in general if you have to like invert, I mean basically you need to solve such integration functions. Suppose, let us say u is F of X and I want to invert F is how you are going to get. So, I want to find out if I want to invert if you give me U, I need to found find what is that X that will give me U when I apply F on X.

So, for example, let us say you have been given to you and F is let us say is given as a integration of this. Now, to find what is that x which will give me this U I need to reverse this integration process which can be very hard. So, when F was exponentially distributed, this was easy, it was like 1 to, 1 upon e to the power lambda x. But every time you may not be dealing with exponential. You may have to deal with other distributions.

So, doing this inversion is required and that may become hard. So, that is why we have to resort we have to look for other methods than what we just dicussed. So, there is a one method called Box Miller method which is also a direct method, which is particularly useful to generate normal distribution samples according to following normal distribution. So, let us see how does this work. Suppose you have two random variables U1 and U2 which are iid and uniformly distributed.

Now, you define two values R which is a transformation on U1 and theta 2 which is a transformation on U2. Now, you define X1 to be R cos theta and X2 to be R sin theta. It so, happens that X1 and X2 are iid and also Gaussian with parameter 0 and 1. You can verify this, this exercise you people have been doing multiple times like basically I am doing this is a one transformation and this is another transformation.

And if I tell you and here can you find how this R and theta, R and theta are independent?

Student: They have functions of U1 and U2.

Professor: They have functions of U1 and U2, but U1 U2 are independent. So, R and theta are independent, will you be able to find distributions of R and theta? A joint distribution of R and theta, you should be able to. Like you know the distribution of U1, R is a function of U1 you should be able to find distribution of R and you should be able to also find distribution of theta.

And now X1 depends on both R and theta but R and theta are independent you should be also able to find the distribution of X1 and X2 here. And remain you can apply the Jacobian method that we have discussed before and you once you do that, you will see that and you find the joint distribution and when you find their marginal you will see that both X1 and X2 are gaussian distributed. So, since you know the method it is just about doing the calculations just working out the details, I am skipping and you should verify this.

So, you see that like the previous method just using this based on this uniform distribution generating Gaussian was not feasible, just based on uniform, we could not generate Gaussian distribution. But now, we just came up with another method here which is also using uniform distribution, but it is still giving us gaussian distribution. It is just like we need to do the appropriate mapping by transforming one and in an appropriate way we may get desired distribution, but what is that mapping?

That is the question and that needs to be solved. Like here this box Miller method maybe this box Miller figured out that, if I do a transformations like this, if I take R and theta in terms of U1 and U2 like this, and if I define an X 1 and X2, then he said that this transformation actually gives what I want. It is not necessarily that if I want to generate let us say tomorrow, binomial distribution, the same kind of transformation may will work. Maybe you have to find a other transformation suitable for that case. However, I noticed this here this, here we are interested in generating Gaussian and Gaussian is continuous.