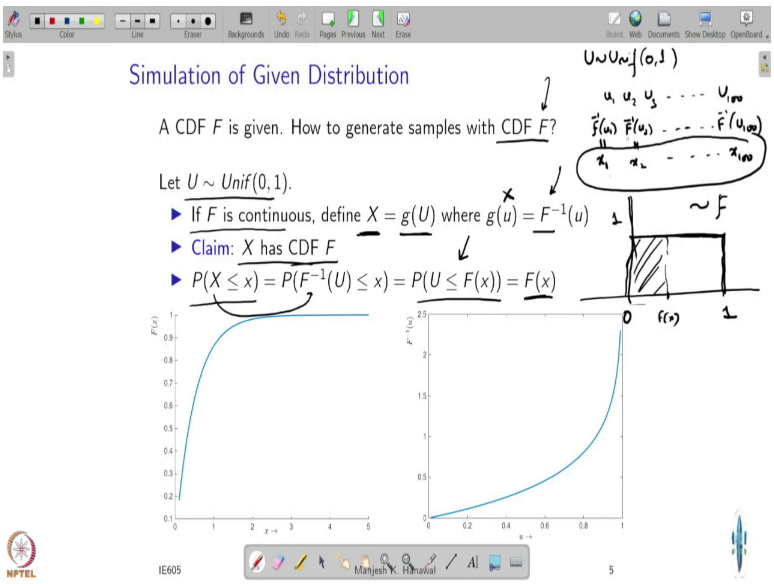
Engineering Statistics
Professor Manish Hanawal
Department of Industrial Engineering and Operation Research
Indian Institute of Technology, Bombay
Week 3
Lecture 11
Generating RVs, Joint Distribution of RVs

When we have to deal with certain models, it is not that all the time data is available to us, somebody has already provided data. Sometimes we have to simulate data itself. And you need to simulate the systems when you have to simulate the system you need to, you are going to make some models, the models will assume certain distributions, and that distributions will be characterized by certain CDFs. And according to using those CDFs, you need to generate data. Now, the question is how to generate data. This is, we briefly discussed last time.

(Refer Slide Time: 00:50)



Suppose, let us say some CDF is given to you. Our model has assumed some distribution, which comes with a certain CDF. And we want to simulate that. Now for that, we are going to do that assuming uniform distribution. Let us say I have access to uniform distribution. Now, for time being assume that your F is continuous, we know that if your function, your CDF function, it is already monotone by definition, or like it is by property, and now it is continuous.

And if it is continuous, and monotone, its inverse function is well defined. It is also going to be one to one. And that is what it looks like here, the, on the x axis, x can take minus infinity to plus infinity, here y axis is [0,1]. But on the inverse map, the axis this is between 01 and the y axis is between it can be minus infinity to plus infinity. Now, what I will do is, we discussed functions of random variables last time.

Now, what I will be looking is a function of this uniform random variable, I will apply some g function on my uniform random variable, and I am going to get a new random variable. But I will not going to be using any arbitrary g function here, I will be using particular g functions, which is given by this F inverse function, you know that F is given to me. And I will be able to find out its F inverse. And I am going to use g function as that F inverse.

Now the claim is, if I define, apply this transformation on my uniform function, the new random variable I am going to get it has a CDF of F. Why is that? Let us say I want to find CDF of my random variable X, that is probability that let $X \leq x$, but replace X by F inverse U, that is by definition. That is what we have defined F equals to F inverse U, g is F inverse U.

And now, because F inverse is a one to one map, I can write it in this fashion probability that U is less than or F equals to U. And now, you can go and compute, we know that U is uniform. So, it will have this nature, this is between 01 and this is 1, this is my uniform function. Now, if you compute what is the probability and F(x) so maybe I made a little wrong here, so this is like my uniform function between 01.

And now if we are going to ask this uniform random variable is going to take value less than or equal to F(x), somewhere F(x) is here. This is exactly close to F(x), because area under this curve is going to be 1. If you are going to look into area under this portion, because the height is 1 F(x) it is going to be exactly equal to F(x). And now you see that the CDF of this has exactly F(x). Now, how to, this is nice property.

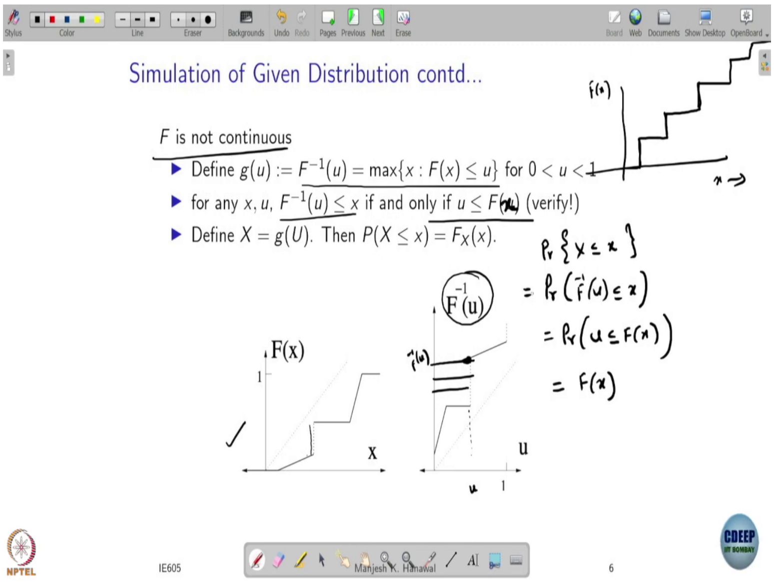
Now, the question is how to generate data as per my CDF F. Suppose U is your uniform and somehow assume that somebody generates data according to this uniform distribution. Let us call this data as U1, U2, U3, up to let us say you have been generated 100 samples, these are actual data, data generated as per uniform distribution. Now, I know this F, what I will do is on these data points, I will apply F inverse U1, F inverse U2 and F inverse U100.

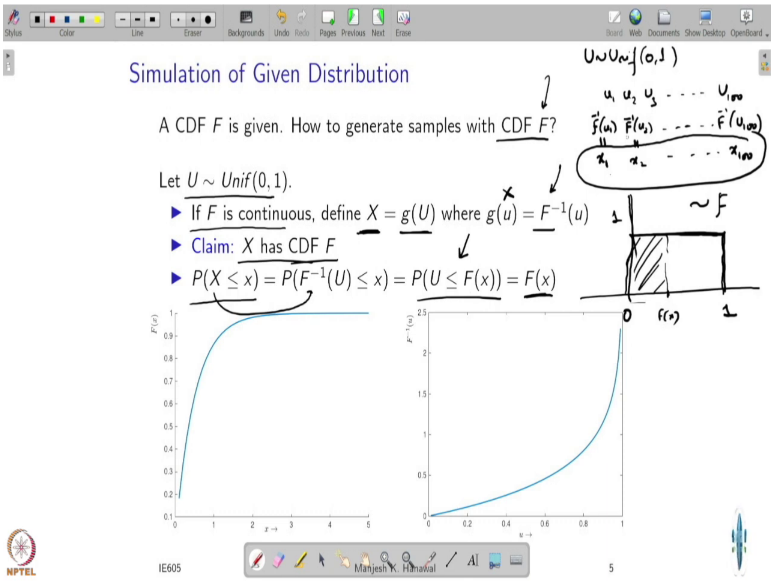
Let us call this point, this is my x value now, that is what we have said. Now, this is what we are calling it as x. Let us call this as x1, let us call this as x2, and let us call this x100. Now, we have 100 data points. Now, these data points are coming as for they are following my CDF of F. Is that clear?

Now, you see that even though uniform distributions samples are coming from uniform distribution, but whichever CDF F is given to you, I have used that and now I have transferred the samples to new samples, those new samples are now as per my required CDF. Now tomorrow like later, you will see that if you are going to call a function in Python, let us say generate normal samples, normal I mean Gaussian distribution with certain mean and variance generate 100 samples how it is going to do that, it is going to do like this, it is going to generate certain uniform samples.

And once you say Gaussian it is not exactly what is the CDF of that Gaussian and it will find the inverse of that and then apply that inverse function on those uniform samples and whatever the values it gets gives you as output, take this as your Gaussian samples.

(Refer Slide Time: 07:07)





So, last time we come to this stage where we are saying that Okay, now, what if F is not continuous and we know that F need, if F is not continuous, always the case when it is a discrete random variable, discrete random variable will have CDF like this. And then there is a jump everywhere and jump everywhere there is a discrete point is there now how to handle that case.

Now, I have particularly put one example here look into this, here, there is a jump at this point. If you look into its inverse, it looks look into this, but this function is not unique, this function is not well-defined. Why is that? For example, let us say but I take one particular point here. At this point u, what is the value I should be assigning, should I be assigning this value, this value, this value or which value? There is ambiguity here.

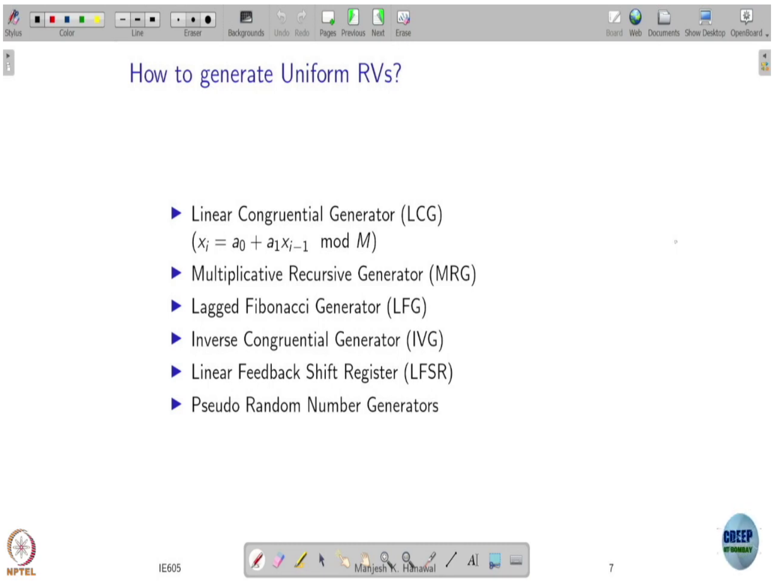
I need to properly define this what is that F inverse function. This things did not arise when it was a continuous function, but the moment it is a discrete that question arises. So, we need to appropriately defined so that is what we said last time, we are going to define in such a way that F of, F inverse u is going to take max of F of x less than equals to u.

Now, if you define like that, now, you can you have basically to this point u here you have uniquely given this particular value as the value to be assigned, this is what F inverse u if this is just a minute, if this is u, then F inverse u exactly this point, not any of this intermediary points. So, because of that your F inverse becomes well defined.

And now you can check this, we said this is x then you can check that if F inverse at $u \leq x$, this is only going to happen if $u \leq F(x)$. So, with this because of this if and only if condition, I can write it as if u ≤ F(x) and this already gives me F(x). So, and this is what probability that my X ≤ x. So, now you see that even for the discontinued case, these things works.

Any question on this simulation of data as per given CDF? So, one obvious question that should come to your mind is who will give me uniform samples, this is all assuming under the, assuming that somebody has provided uniform samples, but uniform is another distribution, that distribution like how we are going to generate. If that is there, we are saying okay, everything, every other distribution we will be able to generate, but who will provide.

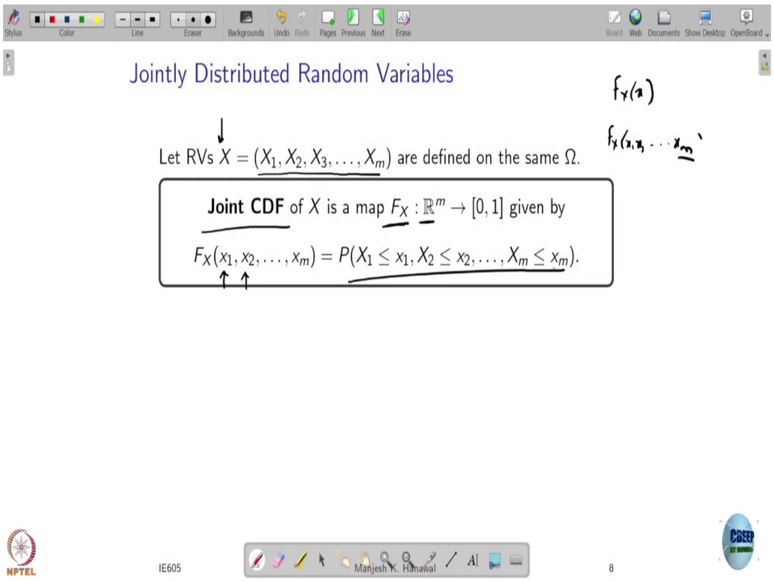
(Refer Slide Time: 10:39)



So, for that, there are many different methods, which we will not go into in this class something called these are based on some congruential generators called linear congruential generator, multiplicative recursive generators, some Fibonacci generator and all, these are based on making things very random within your machine like you take some observations and you iterate in such a way that things start becoming looking very random.

Like say like uniform is one such things, which is easy to generate uniform, because in uniform everything is equally likely. And there is no prior information they are putting. So, when you make certain things very iteratively, everything becomes kind of equally possible. And that is why it is generating often uniform is easier to some extent. And that is why we use them. And based on that we build other, generate sample for other distributions.

(Refer Slide Time: 11:44)



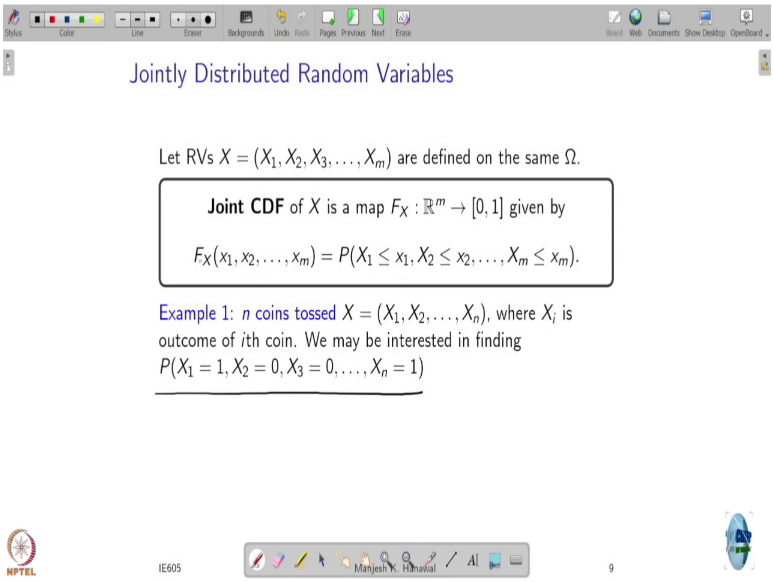
Now, with this, I want to switch to the next topic called as jointly distributed random variables. Now, we will quickly run through again this fast. Again, this is a some bunch of definitions we need to go through. Often, it is not that you have to deal with one random variable, you may have to deal with a bunch of random variables. For example, take 2 coins, coin 1 I am going to throw and coin 2 I am going to throw.

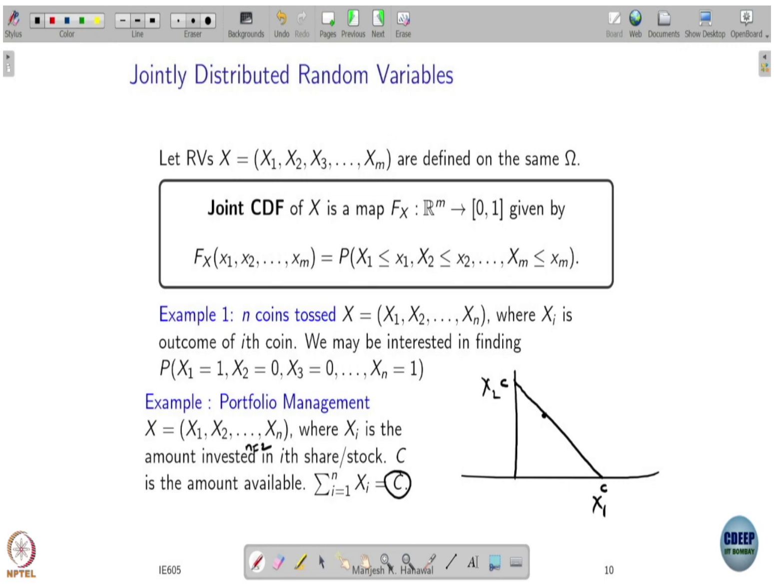
Actually, outcome of coin 1 is one random variable, outcome of the second coin is another random variable. There are 2 random variable I want to see jointly how they behave. So, now, let us say I have this bunch of random variables X1, X2, X3, and then all of them are defined on the same sample space Ω . And now we are going to define something called as joint cumulative density function, sorry, joint cumulative distribution function, which is of this random variable X.

Now onwards, notice that earlier Xi was used to denote for one random variable, now that X could be a vector, because there that could involve more than one random variable. Now, I am going to take like, it is now this CDF is now going to not single varied, but it is a depends on multiple variables. Earlier it was simply F_X(x), now it is F_X(x1, x2,..., xm) m here is the number of random variables you have.

Now, its definition is this joint probability, probability that X1 ≤ x1 and X2 ≤ x2 like that, all the way up to Xm ≤ xm. Notice that this is a joint probability we are talking about. Now, joint CDF is expressed in terms of joint probability.

(Refer Slide Time: 14:06)





So, this example, we said, suppose I toss a coin n times each outcome is like one random variable. And I may be interested in the joint probability. What is the probability that the first throw give me a value of 1, second throw gives me a value of 0 like that and the last one, give me a value of 1. This is like a joint probability I am asking. And the you is based on this are going to define your cumulative density functions. Other things, sometimes you will have to deal with coupled things. For example.

Student: When we are talking about jointly distributed random variables, are we saying x1, x2, up to xn are dependent.

Professor: Right now, we are not saying that dependent or independent we just want to understand the joint behavior.

Student: So, can we guess that x1, x2, x3, up to xn are independent but they are jointly distributed?

Professor: We are not, we have not defined what is independence here. So, far we have defined independence of events. If I give you 2 events when we say they are independent?

Student: Probability of their intersection is product of probability of individual events

Professor: Yeah, probability of their intersection that is like both events happening together is nothing but probability of, product of the probability of individuals that is the independence. But what is the independence here? We have not yet defined it, we will define it. Right now, this definition does not worry about things are dependent or independent. It is simple at this point just take it as a joint probabilities we are defining and based on that we are defining joint cumulative density function, distribution function.

Now, let us take for simplicity, this let us I am going to take this n = 2 here. Let us say I have 2 random variables X1 and X2. Hypothetically, assume that that is representing the amount you are going to put in some stock, you have decided to put your money in two stocks, stock 1 and stock 2. In stock 1, you are going to put your amount x1 and you are going to put an amount x2 in your stock.

But you have a total budget of C. The amount you are going to put x1 plus x2 has to be equals to C. now, there is a constraint. If I increase x1, x2 has to?

Student: Come down.

Professor: Come down. Similarly, if x2 goes up x1 has to, both cannot increase or both cannot decrease if there is some has to add up. So, there is a dependency. Now, to capture this dependency, we need to define joint probabilities. So, here we know that X1 and X2, if I had to plot is not this entire first quarter quadrant here, it has to be something like this. The values can only lie on this land. So, if this is like a C and this is also C only the values have to lie on this point. So, there is a kind of dependency of one on the other.