Simulation is an important tool in the analysis of probabilistic phenomena. For example, suppose that X, Y, and Z are independent random variables, and you're interested in the statistical properties of this random variable. Perhaps you can find the distribution of this random variable by solving a derived distribution problem, but sometimes this is impossible.

And in such cases, what you do is, you generate random samples of these random variables drawn according to their distributions, and then evaluate the function g on that random sample.

And this gives you one sample value of this function. And you can repeat that several times to obtain some kind of histogram and from that, get some understanding about the statistical properties of this function. So the question is, how can we generate a random sample of a random variable whose distribution is known? So what we want is to create some kind of box that outputs numbers.

And these numbers are random variables that are distributed according to a CDF that's given to us. How can we do it? Well, computers typically have a random number generator in them. And random number generators, typically what they do is generate values that are drawn from a uniform distribution.

So this gives us a starting point. We can generate uniform random variables. But what we want is to generate values of a random variable according to some other distribution. How are we going to do it? What we want to do is to create some kind of box or function that takes this uniform random variable and generates g of U.

And we want to find the right function to use. Find a g so that the random variable, g of U, is distributed according to the distribution that we want. That is, we want the CDF of g of U to be the CDF that's given to us.

So let's see how we can do this. Let us look at the discrete case first, which is easier. And let us look at an example. So suppose that I want to generate samples of a discrete random variable that has the following PMF. It takes this value with probability 2/6, this value with probability 3/6, and this value with probability 1/6. What I have is a uniform random variable that's drawn from a uniform distribution.

What can I do? I can do the following. Let this number here be 2/6. If my uniform random variable falls in this range, which happens with probability 2/6, I'm going to report this value for my discrete random

variable. Then I take an interval of length 3/6, which takes me to 5/6.

And if my uniform random variable falls in this range, then I'm going to report that value for my discrete random variable. And finally, with probability 1/6, my uniform random variable happens to fall in here. And then I report that [value]. So clearly, the value that I'm reporting has the correct probabilities.

I'm going to report this value with probability 2/6, I'm going to report that value with probability 3/6, and so on. So this is how we can generate random samples of a discrete distribution, starting from a uniform random variable. Let us now look at what we did in a somewhat different way.

This is the x-axis. And let me plot the CDF of my discrete random variable. So the CDF has a jump of 2/6, at a point which is equal to that. Then it has another jump of size 3/6, which takes us to 5/6 at some other point. And that point here corresponds to the location of that value.

And finally, it has another jump of 1/6 that takes us to 1, at another point, that corresponds to the third value. And look now at this interval here from 0 to 1. And let us think as follows. We have a uniform random variable distributed between 0 to 1. If my uniform random variable happens to fall in this interval, I'm going to report that value.

If my uniform random variable happens to fall in this interval, I'm going to report that value. And finally, if my uniform falls in this interval, I'm going to report that value. We're doing exactly the same thing as before.

With probability 2/6, my uniform falls here. And we report this value and so on. So what's a graphical way of understanding of what we're doing? We're taking the CDF. We generate a value of the uniform.

And then we move until we hit the CDF and report the corresponding value of x. It turns out that this recipe will also work in the continuous case. Let's see how this is done. So let's assume that we have a CDF, which is strictly monotonic.

So the picture would be as follows. It's a CDF. CDFs are monotonic, but here, we assume that it is strictly monotonic. And we also assume that it is continuous.

It doesn't have any jumps. So this CDF starts at 0 and rises, asymptotically, to 1. What was the recipe that we were just discussing?

We generate a value for a uniform random variable. We move until we hit the CDF, and then report this value here for x. So what is it that we're doing? We're going from u's to x's. So we're using the inverse function.

The cumulative takes as an input an x, a value on this axis, and then reports, a value on that axis. The inverse function is the function that goes the opposite way. We start from a value on the vertical axis and takes us to the horizontal axis.

Now, the important thing is that because of our assumption that f is continuous and strictly monotonic, this inverse function is well-defined. Given any point here, we can always find one and only one corresponding x. Now, what are the properties of this method that we have been using?

If I take some number c and then take the corresponding number up here, which is going to be F_X of c, then we have the following property. My random variable X is going to be less than or equal to c if and only if my random variable X falls into this interval. But that's equivalent to saying that the uniform random variable fell in that interval.

Values of the uniform in this interval-- these are the values that give me x's that are less than or equal to c. So the event that X is less than or equal to c is identical to the event that U is less than or equal to F_X of c. So this is how I am generating my x's based on u's.

We now need to verify that the x's that I'm generating this way have the correct property, have the correct CDF. So let's check it out. The probability that X is less than or equal to c, this is the probability that U is less than or equal to F_X of c. But U is a uniform random variable.

The probability of being less than something is just that something. So we have verified that with this way of constructing samples of X based on samples of U, the random variable that we get has the desired CDF. Let's look at an example now.

Suppose that we want to generate samples of a random variable, which is an exponential random variable, with parameter 1. In this case, we know what the CDF is. The CDF of an exponential with parameter 1 is given by this formula, for non-negative x's. Now, let us find the inverse function.

If a u corresponds to 1 minus e to the minus x-- so we started with some x here and we find the corresponding u-- this is the formula that takes us from x's to u's. Let's find the formula that takes us

from u's to x's. So we need to solve this equation.

Let's send u to the other side, and let's send this term to the left hand side. We obtain e to the minus x equals 1 minus u. Let us take logarithms: minus x equals to the logarithm of 1 minus u. And finally, x is equal to minus the logarithm of 1 minus u. So this is the inverse function.

And now, what we have discussed leads us to the following procedure. I generate a random variable, U, according to the uniform distribution. Then I form the random variable X by taking the negative of the logarithm of 1 minus U. And this gives me a random variable, which has an exponential distribution.

And so we have found a way of simulating exponential random variables, starting with a random number generator that produces uniform random variables.