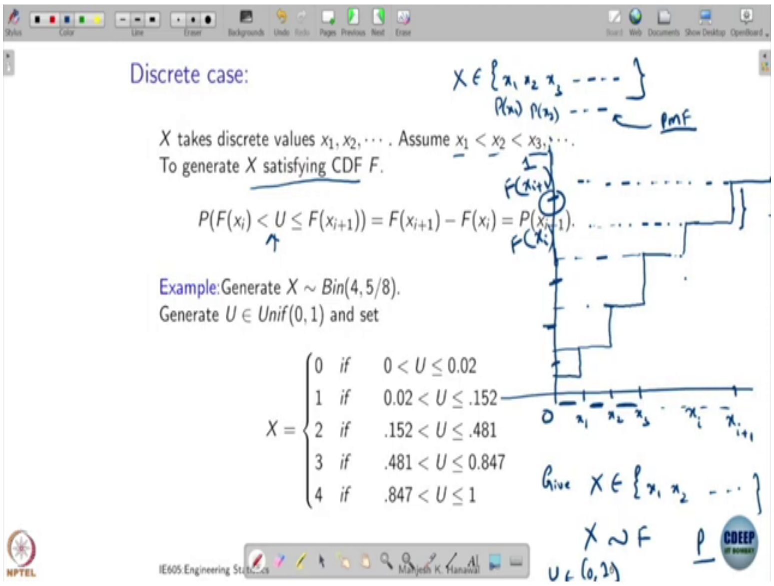
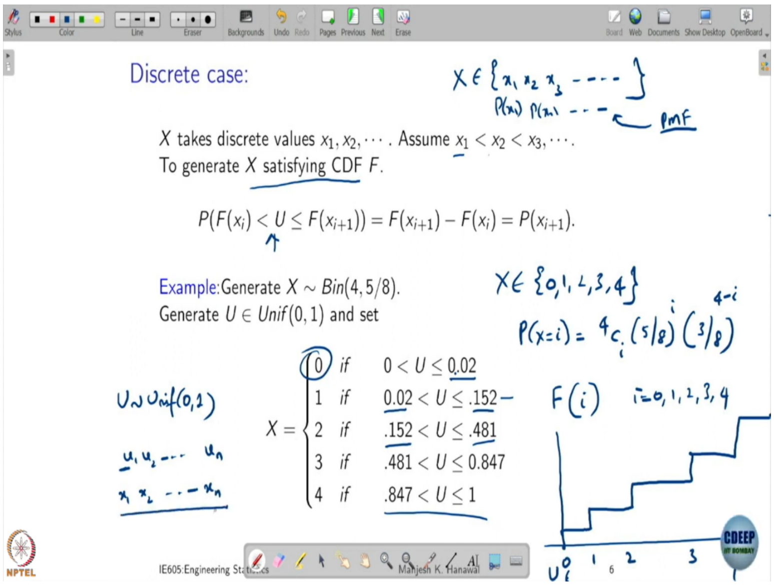
Engineering Statistics
Professor Manjesh Hanawal
Industrial Engineering and Operations Research
Indian Institute of Technology Bombay
Lecture 28
Generating random sample- Indirect Method

(Refer Slide Time: 00:14)





However, if you are interested in generating a discrete random variable there is actually a simpler method. Let us see how that how does that work. Suppose let us say you are X takes values x1, x2, x3 like that and it is a discrete random variable and assume that x1 is the smallest and x2 is the next smallest and x3 is the next smallest like this the possible values taken by this my random variable are arranged in increasing order and suppose X has a CDF of F and I want to find a probability that my random variable U takes value between Fxi and Fxi plus 1.

Let us visualize this. Let us assume that my values are x1, x2, x3 like this and this being a discrete random variable my jumps will happen at this points only. Let us jump will happen somewhere here and now what I am asking, I have a random variable U, U is taking value between Fxi and Fxi plus 1. Let us maybe what I will do is let us do this do this one and like this. This is xi where this jump is happening.

Let us do this and this is xi and at this point is xi plus 1. So, what is this xi? This is the value. Now, I am asking, what is this point here? This is Fxi. And what is this point here? This is xi plus 1. Now I am asking U is some uniform random variable with parameters 0, 1. What is the probability that it lies in this value? So, probability that U is between F of xi and F of x minus y, I know that this is nothing but F of xi plus 1 minus F of xi for a uniform random variable. Everybody agree with me?

And now if you look into the difference between F of xi plus 1 and Fi, how much is this jump? The mass that is the point xi plus 1 have. So, that is the point. So, this is going to be exactly equals to xi plus 1. So, we can exploit this property that if you are generating uniform random variable, the probability that it takes value between F of xi and F of x plus 1 is exactly close to probability of xi plus 1. And we can use this property to generate random variable which has this probability mass function.

Let us say this is like now this is P of x1, P of x2 like this is my probability mass function and I want my x1 to appear with probability Px1, x2 to appear with probability Px 2 like that. Now, what we are saying is we can get the probability using an uniform random variable in this fashion. Everybody agree like how this is working? What is the process? Now, this is the property we understood the property we want to exploit it.

Suppose have been given I have been given X and I have been told that it takes value x1, x2 like this and I have been asked to generate samples satisfying this PDF sorry this CDF. This is a discrete the CDF. So, once you give me a CDF do I know the probability mass function? Yes, right if I know F, I know P also.

Now what I will do is I take a uniform random variable generate a sample and see whether where it falls. It falls between which of these intervals? Whether it falls here, here, here and if it happens to now I will check basically these intervals like which of these intervals it will follow like I can divide this into intervals like this.

So, what is the range of my CDF value? What is the value my CDF can take? 0 to 1. So, this value is maximum going from 1, 0 to 1 and I have divided them based on the jumps and let us say this is like xi plus 1 and this is xi this is F of xi and now, you generate U and say between it is going to be between 0, 1 see that where it falls. It falls in this category in this category in this category and in this category and if it falls in this category in this block, you are going to take your value x to be xi plus 1 and then that just samples you got they are going to follow your required CDF.

Now, let us work out this in this particular example we have here. Is the basic idea clear how this is being generated? Now, I want to generate X which is binomial with parameter 4 and 5 by 8. So, what are the possible outcomes of this binomial random variable? It is going to take what values it is going to take 0, 1, 2, 3, 4 and this is a binomial random variable I know its probability mass function.

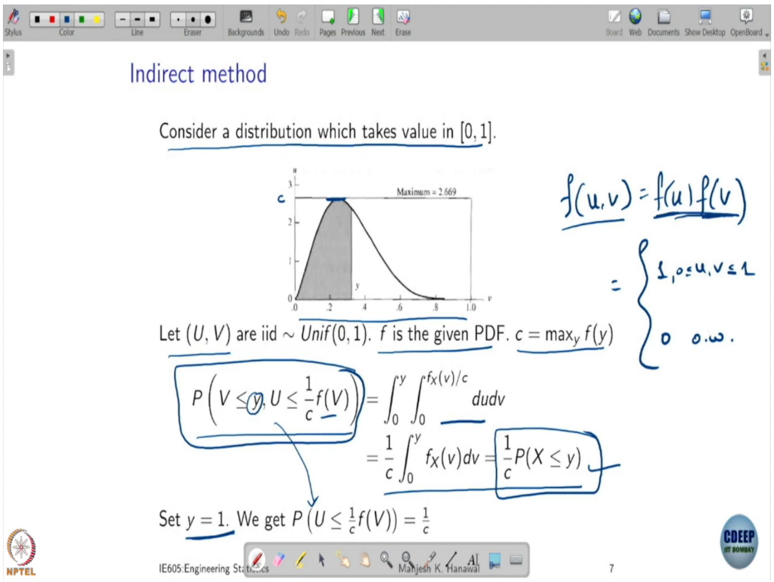
So, what is the probability that it is going to take value i? Four choose i, 5 by 8 i, 3 by 8, 4 minus i and from this can I found its probability sorry, cumulative density function? I can find out what is F. It is taking value i, i equals to 0, 1, 2, 3, 4. So, I basically have so it is like this, this like this, maybe I will just call it 0, 1, 2, 3, 4 and this is 1.

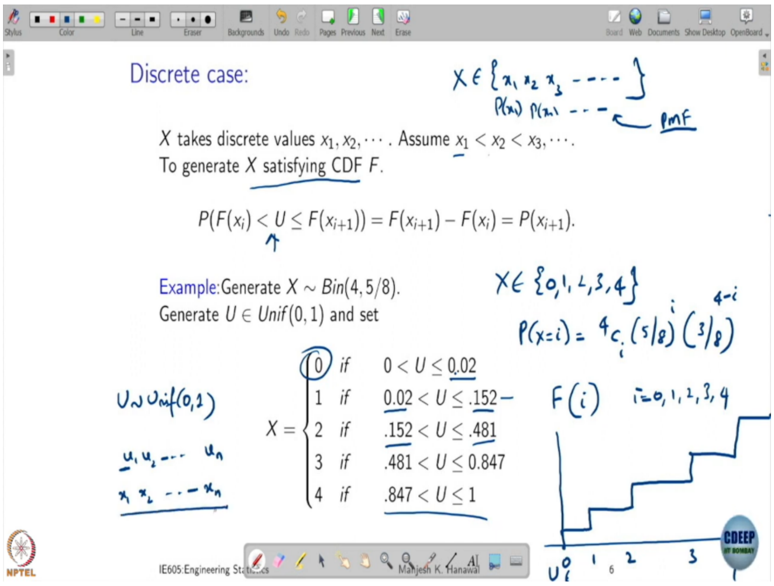
Now, what I am going to do is I generate uniform random variable. If it happens to between 0 to 0.02 I will take the value of x to be 0. If it happens to be between 0.02, 0.152 I will take 1 you know how this 0.02, 0.12 are coming from CDF values, this is a probability that it is going to take value 0 and this is CDF at 0 this is CDF at 1, CDF at 1 CDF at 2 like this. If it is going to be between 0 to 0.02 you take 0 if it is between 0.02, 0.15 to take 1 like that and if it is going to be between 0.847 it is going to be like this.

So, if you let us say you have n samples let us say u1, u2 up to un you take u1, you see which of these conditions satisfy. Can more than one condition satisfy here? No it only one condition satisfy and whichever it is, it is going to be one of them call that as x1 and similarly next you take u2. See which one are these conditions satisfied to call that x 2 like that. Now, you take un and see which condition is satisfy.

Now this x1, xn you have they are going to have what CDF? Binomial with parameter 4 and 5 by 8, Is that clear? Once it is a discrete case, it is pretty easy to generate. You do not need to even worry about inverting this F, F could be complicated. I mean, if it is very hard to invert even this discrete like this, functions, so you can just apply this method.

(Refer Slide Time: 12:01)





Now, we will just talk about one indirect method. How to do it for the indirect method? So, we need to apply some tricks here. When it is so discrete methods, we are pretty sure like how to do using this. Only thing is now continuous case. Continuous case, the only simple things we did like gamma generation, chi square with even number of degrees of freedom and we also talked about a Gaussian distribution.

But for the continuous case, the PDF can be arbitrary. How to generate that? Now, let us initially let us take my distribution to be taking only value in the interval 01. And let us assume that it looks like this, its curve looks like this. So, it is the range is only between 01. Now, let us again try to see that we can generate this distribution using some uniform random variables.

I am now starting with two uniform random variables event I which are independent. And let us say f is given PDF. And f is given and I will find out the maximum value of that PDF. So, in this case, if you take this PDF, the maximum value is here. This is what I call it as C. Now let us see I am interested in evaluating this function probability that V is less than or equal to y and capital U is less than or equal to 1 by c of f of V, and this is for a given y you take some value y and now try to evaluate this probability.

Now, how to evaluate the probability? Since U and V are independent uniform. What is the joint distribution? What is the possible value that joint distribution can take? Now, U and V because they are independent this is nothing but f of U and f of V. Because U and V are independent. And now they are also uniform. I know that this is going to take value 1 this is going to take value 1 if my UV are in the interval of 01.

So, this joint distribution can also take value is between 01 only, it cannot take anything else because both f of u and f of v takes value only between 01 or I would say not value, but it is exactly equal to 1. Uniform distribution, its PDF is constant 1 in the when both u and v, they are less than or equals to 1 or maybe I should write it like this and 0 otherwise. So, I have just used that property and now I am integrating it. First one is between 0 to y and the second one is between 0 to f of x v by c.

Now, if you simplify this, you will get that this is nothing but 1 by c is anyway constant. And this is 1 by c times probability that x is going to be less than or equal to y. Everybody agrees with this calculation? I am just simply using the property of uniform random variable here, nothing else. So, if you have not digested just again verify this.

Now, this is true for any y this is true for i particularly take y equals to 1. When y equals to 1, what is the relation I am going to get? Now, concentrate on this and this relation when y equals to 1, v less than or equals to 1 that means, V can only between less than or equals to 1. So that will become marginal for me and when x is less than or equals to y, here x I am assumed that its range can be only between 01. What is the probability that x can be less than or equals to 1? It is going to be one. So, that is simply becomes 1 by x, 1 by C here.

So, because of when I take y equals to 1, this has to be true for every solution has to be true for every 1, when I take y equals to 1 what the relation I get is probability that u is less than or equals to 1 by f of v is exactly equals to 1 by v.

Student: Sir in (())(18:02) anything about distribution.

Professor: Which one?

Student: In first line we are considering (())(18:09) this question.

Professor: Yes.

Student: Anything about distribution.

Professor: No, all we are assuming is it is taking value between 01.

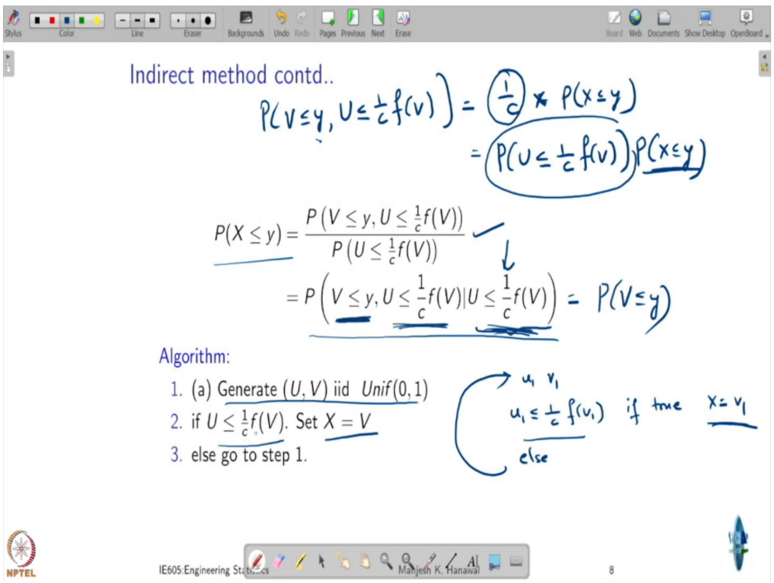
Student: Here it is looks like normal so we do not anything.

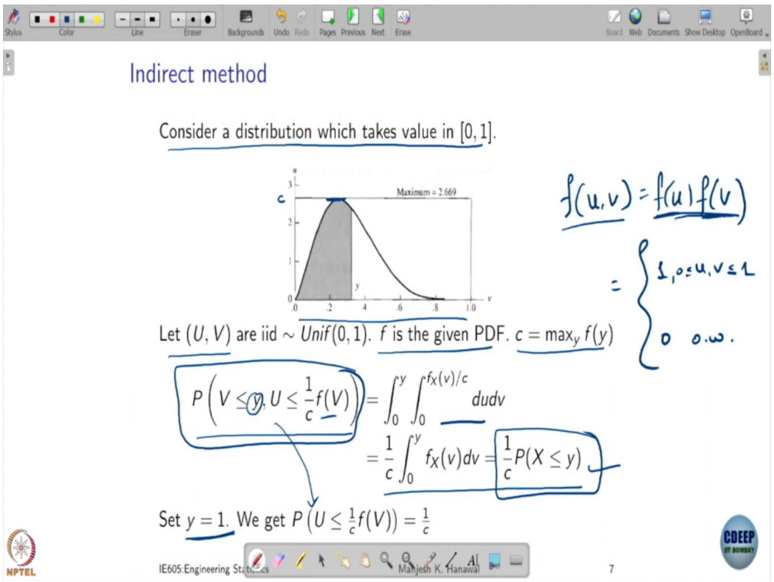
Professor: No, this just for representation you put it can be anything it could be even like something like this. It could be anything. So, from this exercise, what we know is if I have a two uniformly distributed random variable u and v, for any f whatever the underlying CDF we are talking about this relation holds that is the crucial thing that we have just derived.

Student: So, why are we taking (()) (18.52)? In this relation it does not depend on it.

Professor: It does not depend on c like you can put any value here that is fine. Nothing specific that why we took only thing is we took c to make sure that this ratio right here is less than 1. If you take something like c to be less than 1, this ratio could exceed 1 and that is not a good case for us.

(Refer Slide Time: 19:21)





Now, what we did, if you now look into this, so in summary, what we did, we just showed that probability that u is less than or equals to v, u less than or equals to v 1 by c, f of v is 1 by c x less than or equals to y. That is what we said sorry, this is into, but we say that but we have a relation. This 1 by c is nothing but probability that u is less than or equals to 1 by c for f of v this is the relation we have got here.

Now, what I do is probability x is less than or equal to v. I take this in the denominator here, I just take this in the denominator side here and I will get this relationship and now if you look into this, this is nothing but now this is like I can treat it like a conditional probability. Probability that u is less than or equals to u and probability that u is less than or equals to 1 by c f of v given that u is less than or equals to 1 by c f of v.

Everybody agree with me? So, now I want to explore this relation. I what I was able to show what I have derived is probability that x is less than or equals to y is dependent on u and v and c in this fashion, now, let us see if we can generate an algorithm based on this simple idea. All I need is two random variables u and v which are uniform independent and you have to give me the CDF according to which you want your data.

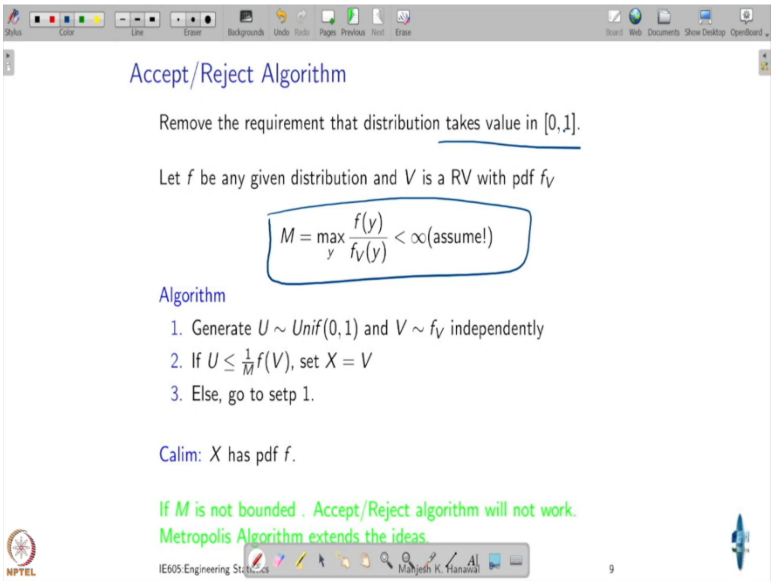
So, first step in the algorithm is generate u, v sample IID uniform distributed and then check that if this u is less than or equal to 1 by c f of v that is a conditional part you are going to check is if u is less than or equals to 1 of you, if this is true, if this condition is true, what is this value? This if this condition is true, then this joint part does not matter. What matters is only u is less than or equal to y.

If this is true, then it is simply probability that v is less than or equals to 1. So, then in that case, I can simply take this x to be equal to u. That is it. So now, you see that whenever this condition is met those samples I will take it as x, otherwise, I will go back and regenerate. J

ust let us see how this works. Suppose let us say you generated u1 and v1 and then what you are going to do is you check whether your u1 is less than or equal to 1 by c f of v1. If true, you take your x to be v1 else you go back and do this and keep doing that and in some cases, you are going to this condition becomes true and another thing see like you have asked why the c to be taken maximum this condition used to be less than or equal to 1 by c f of v this is guaranteed to be less than 1 only when you take c to be max value, if you do not take it max value this is not guaranteed to be less than 1.

And now you will get samples whenever this condition is met those samples you collect and those samples are guaranteed to be following your required CDF function and this is what called indirect method.

(Refer Slide Time: 23:54)



So, one last thing is if you want to get rid of the assumption that your PDF is between 0, 1 all you need to do is instead of c you redefine your function M and then repeat the same process. Then this will also give you a random variable which has your required distribution f but without requiring the requiring it to be taking value between 0, 1, so let us stop here.