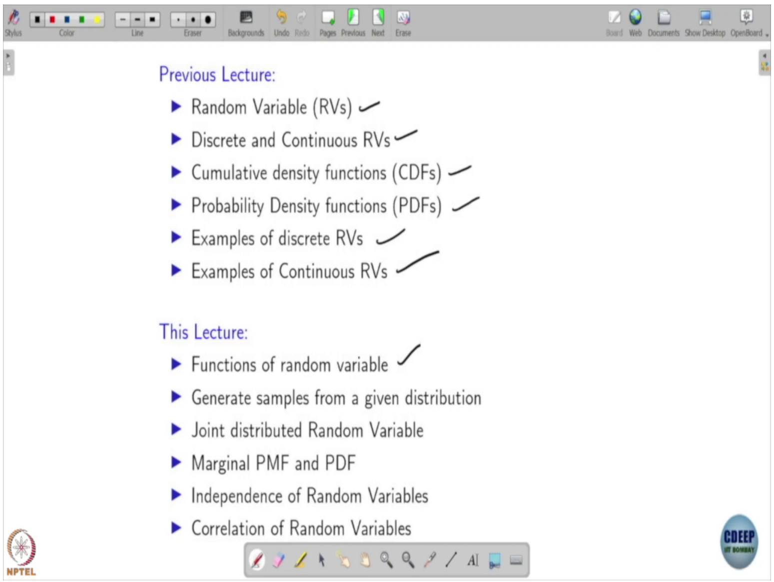
Engineering Statistics
Professor. Manjesh Hanawal
Indian Engineering and Operations Research
Indian Institute of Technology, Bombay
Lecture 10
Function of Random Variables

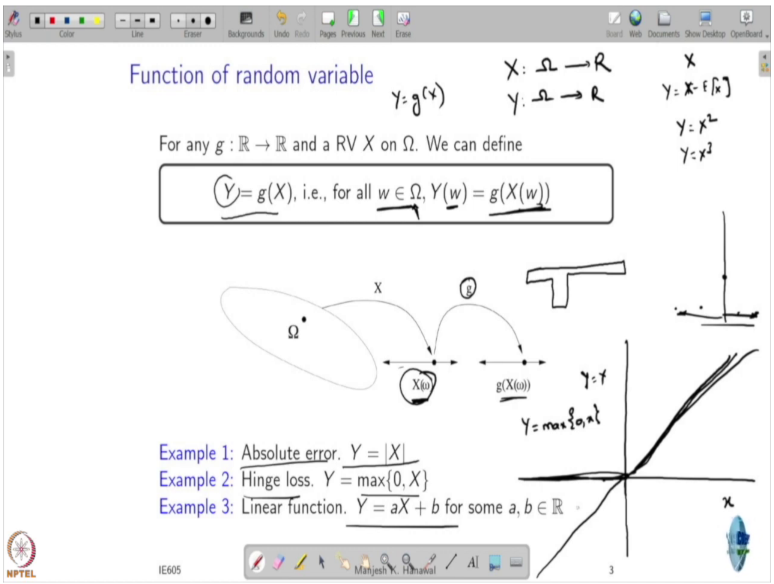
Next move on to functions of random variables.

(Refer Slide Time: 00:22)



And I do not know how much of that points I will be able to cover today. But let us try to cover as much as possible because as I said all of this, you would have also learned in your IE 621 course.

(Refer Slide Time: 00:37)



We will now consider functions of random variables. Random variable, as I said itself is a function. So X, we said that is a function from Ω to R, real line it is a function. But now we want to concern function of this random variable itself. In fact, like you already see one version of this, like you, we just said, right X. And we defined V equals X minus expectation of X. Now, can we think Y is a function of X, Y is already a function of right like in this, you are basically subtracting constant amount. That is one example.

So, we have to like that, you may be interested in another function, maybe Y equals X square, maybe Y equals x cube, all are functions have random variables. Now, let us take g to be any function. And I am going to define Y = g(X). What I mean by this is, and it is still going to be under the same sample space Ω only, underlying sample space remains the same. That means, if the new random variable is still on the sample space Ω only, but now.

So, now if I am going to define Y like this, Y is again a mapping from Ω to R where Y = g(X). But what does this mean is, my new random variable on any Ω is g(X(ω)). So first, I will compute X(ω) and then I am going to apply my function g. So, what does this mean is, basically this is like a composition of functions. So, I have my omega space, every point I first mapped on the real line using my random variable X.

And now from there, I am further mapping it using my function g. So, that is why g is applied on this point X(ω). And this is like a composition of two functions. And this is very useful, especially when you try to do optimization in machine learning. You have to use such random variables and their functions.

One thing is like absolute error. Let us say X is the quantity you measure, but that quantity could be positive or negative, but you do not care whether it is a positive or error what you, what you bothers you, how much from the target I am away. So, there you may want to consider the absolute value.

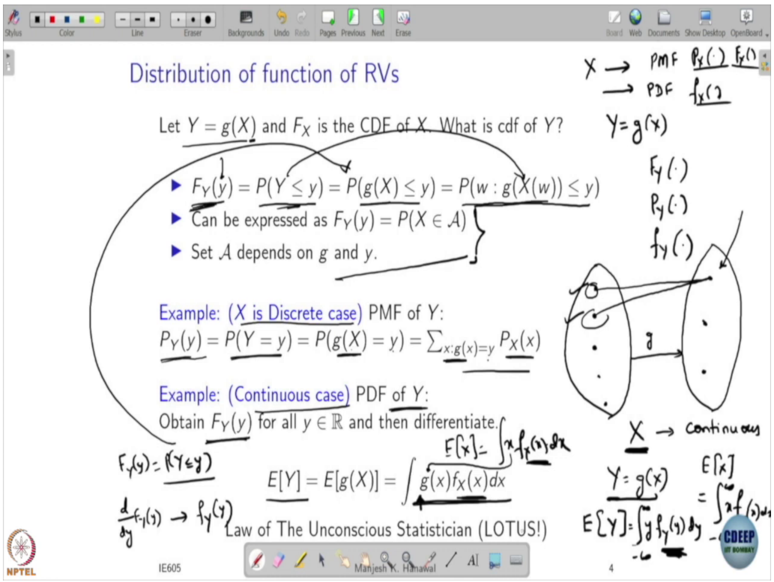
Like I said, the previous example right like I have this a curl and you have some target here. You want to hit this target, you are exactly should fall here, but you may fall short or you may fall when you may go beyond the point, what matters is how much from this point how much you are away, how much you are, whether you are short, or over shooting does not matter. And in that case, you want to consider the absolute value.

And like this is what I am using this machine learning terminologies here absolute error, and other something called hinge loss we use in machine learning, which is like you truncate the negative values. You take max{0, X}, suppose let us say. How to do this max of 0, X. Let us say, you have something like this, this is my y. No just a minute, so let us say I have this line with 45 degree slope. So, what is the slope of this line is going to be 1.

So, if I have this Y here, and I am going to take it as a f(X) here, this Y is simply x, this Y simply X, that is why it is a linear line. But I may want to truncate the negative values. So, then what I can do is, I can redefine my function as Y equals to max of 0, X. In this case, this value get truncated, all this becomes 0 for me, so it is going to be like this now.

So, whenever you want to discard your negative values of your X, you just put a max, now that is what you can think of X is Y is some function of X now and other simple things, we will consider R like linear function, maybe Y is simply aX + b for some constant a and b. Now you see that motivation, like, I may have to in certain cases deal with functions of random variables. There is an underlying function X, but I may not be interested in that itself, but some function of this. And that time, I have to deal with such functions.

(Refer Slide Time: 07:05)



Now in that case, suppose let us say you have X. And it is a PMF. If it is a discrete random variable, or let us say, it is PDF, it is a continuous random variable as the case may be, you know this. Now, you have been defined a new case. So, this was like a P(X). And this was like F(X). Now you have to face with, or you may have their CDFs, which is like F(x).

Now, you have to deal with this case, now, I want to come up with the CDF of this new random variable, or I want to come up with its PMF or if it is a continuous, I may have to come up with this PDF. Now how to find this based on what I know about X. How to do this, for that, we will simply use the knowledge and the relationships. Now, suppose Y is my new random variable, which is a function of X. And now I want to find its PDF. I will take some value Y. And I am trying to come find out PDF, CDF of Y at points y.

Now, this is nothing but probability that Y is less than y. This is by definition. Now, why are replaced by g of X? This is again, the relation we have, now I am interested in this basically implies the set of all those ω such that X(ω) and that value computed through function g, that is going to be less than or equals to y. Now this is like an event.

Now what is the probability of this event? This we know like, we have always defined probability on events. Now this condition, we have translated into event. And on that event, I know how to compute my probability. So, through that I have CDF of my new random variable Y and now. I mean, maybe I do not need to go into this too. I will just leave this.

Now for example, take a discrete random variable X. Now I am particularly focusing on discrete case and later I come to continuous case, if it is a discrete case how to do this, we know that in this case probability of Y taking some value, y is nothing but probability of Y = y and in this case Y is nothing but g(X), I replaced Y by g(X) and now, notice that it is g function need not be 1 to 1, okay for example

So, let us say whatever the points we have here it could be like this, like this is my domain and this is my range and this is my g function mapping it may happen that these two points could go to the same value. So, when I said g(X) = Y, there could be two points of X realisation would, could can go to the same Y. So, because of that when I have to deal with this I need to take into all those Xs which gets mapped to Y and add their probabilities, is this clear to all of you?

So, now, if I want to compute the probability of this, I know that this point and this points are getting mapped to this, to compute the probability of this I need to add the probability of this value as well as this value and that is what we are doing here. Now similarly, when we have continuous case, I now, I want to find out PDF of Y, how to go about this.

Now, let me ask me one simple question, suppose let us say you have x is sorry continuous and Y is g(X) and the g is some arbitrary function, is X also continuous or it can be discrete, it can be both that depends on the g function right, it may just happen g function can be such that, if the value is above 0, I can take it as 1, if the value of x is below 0, I will take it as minus 1. In that case Y is only taking two values 1 and minus 1.

Now, in that case, depending on what is the situation, you may want to compute either PM like you may want to accordingly compute your CDF and also go for it's a PMF or you find it says PDF function. So, now, so in that case like if you want to find P(F(Y)), what you just to do is, you know that Y is P(Y) $\leq$ y you find this value through all these relations like the relations we have here and then, if it is differentiable at point y you go and differentiate Y and this will give you F(y) at point Y and this is provided you are F(y) is differentiable at that point Y? If it is not differentiable you cannot compute that.

Fine and next suppose we have F(X), sorry X random variable and you have this random variable Y, how to compute what is the definition of, what is the expectation of Y, you know expectation of X, what is expectation of X this is nothing but x into F of X X d X which is minus infinity to plus infinity. Now, how to find expectation of Y, then this is nothing but by definition yF(y) minus infinity to plus infinity dy. This is by definition.

So, if I want to compute this first, I have to know what is my PDF of Y, then I can find it. But it is not always necessary. What we can do is instead yeah, Y you just replace by g(X). But in that case, you already know that if you have done this g of X, Y's mass is already taken account by F of X, you just, you do not need to do this F of Y, you can just use F of X here, and this will still use the expectation of Y.

So, what is the good thing about this? I only know to g here and I do not need to really go back and compute what is my F of Y in this case, because computing F of Y itself is some task right. So I do not need to compute as long as I know F(X), I can simply go and compute based on that what and do this integration and I will get expectation of Y. So, this is like, it looks like almost similar to expectation of X expression right expectation of X is X into F of X X d X.

But when I computed expectation of Y, all I did is replaced X by g of X and I did not change its CDF function. So, because of that, sometimes this is called as a law of the unconscious statistician, because you may be simply changing this X to g of X, without worrying about you actually use the F of X correctly or not. But still, you get the correct answer. While being unconscious, you are getting the right thing. So, that is what it is also called as lotus.

So, you will see that sometimes in probability, every time like you do not need to compute everything repeatedly. Like if you have some function Y, you do not need to go and repeat the story. Find its PMF or PDF, to compute its mean or variance. It is enough, even if you know x, and you have to just use some properties.