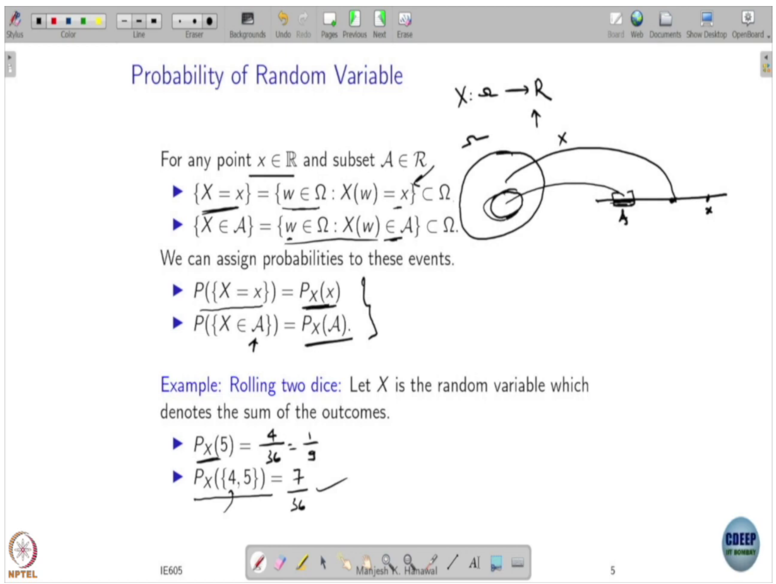
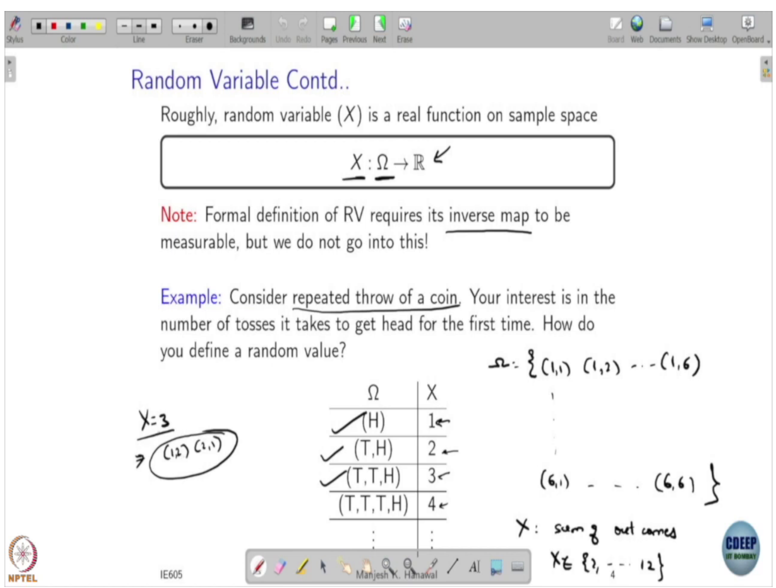
Engineering Statistics
Professor. Manjesh Hanawal
Department of Industrial Engineering and Operation Research
Indian Institute of Technology, Bombay
Lecture 06
Random Variables and Cumulative Density Function

Now, let us try to make this more formal. Now, how to define probability of a random variable on a point and probability of a random variable on a set.

(Refer Slide Time: 00:37)





Let us, take a point x belongs to R. So, I can always recall that X is a map from omega to R. So, range is an entire real line. So, I can always ask the question, what is that X is going to take a value x, small x. So, just again just think that this is a your real line all the way from minus infinity to plus infinity and this is your sample space omega, what X is doing is putting every point in the sample space to a point on the real line.

Now, for this I may ask, if I take a particular point, what is the probability of taking this particular point on my real line? So, that is like, what is X equals to x, what does this mean? This is all omegas in capital omega my sample space such that X(omega) is that is small x. So, if I am going to say this is my point, it may happen that multiple points in this sample space could get mapped to the same x. I just showed you the example here.

If I want X equals to 3, both these 1, 2 and 2, 1 they are getting mapped to the same X equals to 3. So, I am just saying that all those points in capital omega, which will get mapped to that point X, you are interested in and that is natural A subset of capital omega. And now, instead of a point, you may want to ask take this particular region, what is the probability that X is going to lie in this region and A is that region.

Now, if you want to ask what is the probability that X is in that region, this is set of all those omegas such that X of omega belongs to that subset is that set A, that is all maybe like this bunch of points here, they will all go to some points in this region. And all these points is what those so omega this part is capturing and that is another subset of omega.

Now, this is simple notation, we are going to say now, we can ask, we know that X = x is some event. If 5, if it is a event, I can ask the probability of that event. Now, probability that X equals to x, I usually simply denoted is in this way, P with a subscript denoting that random variable and inside the point x that I have.

And if I am interested in knowing this X belongs to some subset A, I will just write like this. If you understand this, just let us quickly do this examples. Let us, X is a random variable which denotes sum of the outcomes. Sum of the outcomes. Now, what is P of, what is the probability that my random variable X takes value 5?

Students: 1/9.

Professor Manjesh Hanawal: 4/36 or 1/9. Everybody agrees or anybody has different answer. So, 5 can happen in how many ways?

Student: (())(04:06).

Professor Manjesh Hanawal: 1 4, 4 1, 2 3, 3 2 and now what is the probability that my X takes value in the set 4 and 5?

Student: 7/36.

Professor Manjesh Hanawal: 7 by

Students: 7/36

Professor Manjesh Hanawal: 7 by?

Students: 7/36.

Professor Manjesh Hanawal: So, now I am asking X to take either value 4 or 5. You already know it takes 5 in 4 ways. And how many ways it can take value 4?

Students: 3 ways.

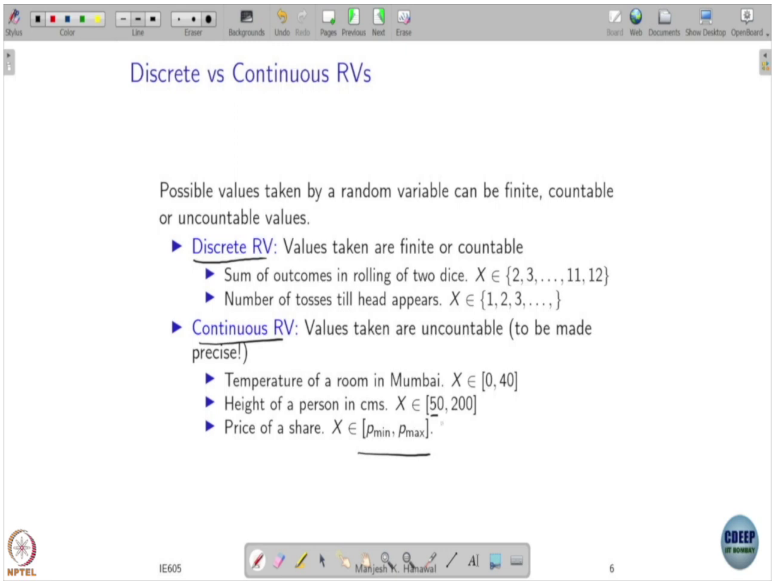
Professor Manjesh Hanawal: 3 ways, 2 2, 3 1.

Students: 1 3.

Professor Manjesh Hanawal: 1 3. So, already 4 has an additional 3, so this is a value. So, now see that, now random variables we have but that random variable is just a phase for us. Now, like what we have done is omega and the events, we have kept them aside now we are defining things in terms of the random variables. But they are connected to all those events.

Now, I am basically now saying, I have the random variable given to me, I am asking the probability of that random variable taking some value, but that is all coming from the basic events itself. Probability X equals to x means that probability X equals to means that means this is corresponding to some event. Once I know that event, I know the probability of that event so in that way, when I am asking some random variable it is connected to the events and on that events, I know probabilities that is indirectly I am saying that I will now have different probabilities on the random variables itself.

(Refer Slide Time: 06:02)



Now, there are different types of random variables we can think of depending on the type of range called discrete and random variables. So, whenever I mean, we have two kind of things obviously, two classifications can happen: discrete random variables and continuous random variables.

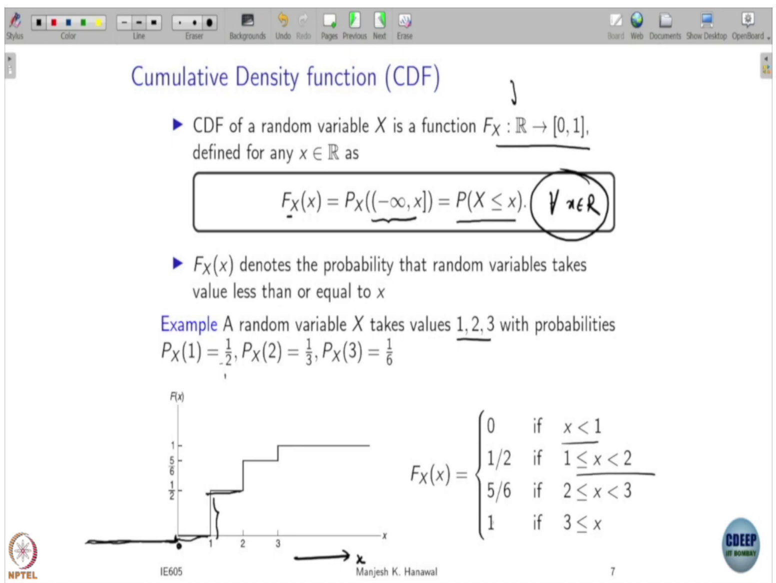
Whenever the values taken are finite or countable, we are going to call them discrete random variables and we have already seen many discrete random variables. The sum of the outcomes of rolling a two dice in that case, X can take values 2, 3 all the way up to 12. So, this is in fact, like, it is going to take finite values, if you are interested in a random variable, where you are going to toss still a head appears. So, in that case, X can go from 1, 2, 3, 4, 5, 6 all the way up to infinity. Here it is, infinity but still countable. So, we are going to treat it as a discrete random variable.

On the other hand, if it does not belong to either case, we are going to call it a continuous random variable. That means there are uncountably many points values my random variable can take. So, for example, if I am interested in temperature of a room, temperature of the room if let us say I am interested in degrees Celsius, and it could be anything between let us say 0 to 40. All values between 0 to 40 can happen. So, it is a continuous thing here.

And similarly, the height of a person, maybe you are height varies from between 50 centimeters to 200 centimeters, any value in this range is possible. It is continuous. Like that, and it will be given some min and max range, if it is going to take all possible values in that range, we are going to call it a continuous random variable.

Now, based on what we have defined, now we want to aggregate start aggregating this information. Random variable gives you a nice way of representing translating everything on the numbers. Now, on those numbers, we may want to start asking questions. Whether my random variable is going to take value less than this or more than this, or it is going to be in this interval, or it is not going to be in this interval, we can ask all these questions. They all correspond to different, different events. But now, instead of talking about events, we will just talk about random variables.

(Refer Slide Time: 08:44)



The first thing we are going to look into is something called cumulative density function CDF. So, CDF is again a function, which is a map from real numbers to interval [0, 1]. Notice that it is range is entire real line and its domain is [0, 1] interval. So, if you have a random variable X, and I want to compute CDF at a given point, X, that is defined as probability of X taking the value between minus infinity to x, which is same as saying that probability that X is going to take value less than or equals to small x. So, what CDF is saying is and this should be this is for all x belongs to R.

So, CDF of a random variable X at point small x is nothing but that random probability of that random variable taking value less than or equal to that value small x. Now, an example. Let us say I have a random variable X which takes only 3 values 1, 2, 3. And the probabilities of their probabilities I have given to you, probability that X takes value 1 is half. Probability that it takes value 2 is 1/3. And the value it takes value 3 is given 1/6. Now, we want to find out its cumulative density, cumulative density function.

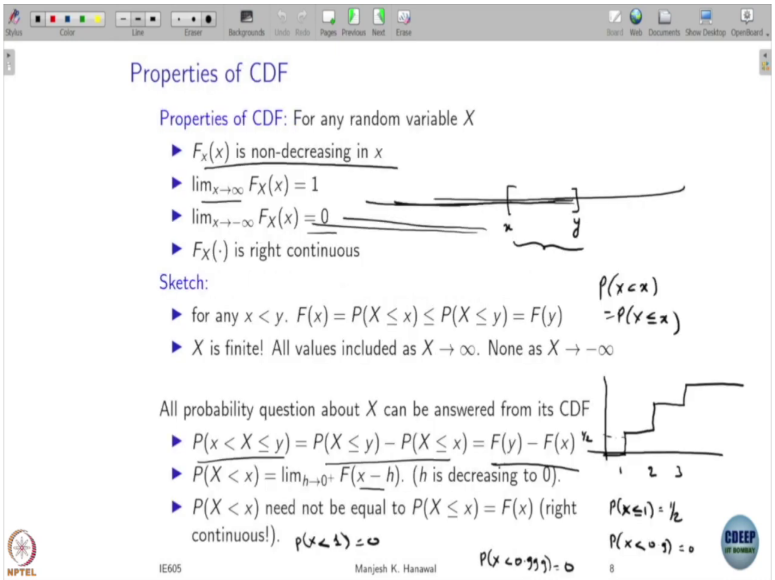
Let us, plot how does it look for this particular example. So, I have to plot like x is on the x axis, and y is my F(x). So, now let us look into till point 0. Let us take I mean, I have to define it for all possible x. Let us start at. Let us looking let us start looking into what happens at x = 0. What is its value?

So, that is basically I am asking what is the probability that my value random variable X takes value less than or equal to 0. Is it taking anything less than or equal to 0? No, because it is only taking value 1, 2, 3. So, cumulative, should not that point is going to be 0. So, in fact, I should stress this back. So, it is going to remain 0 till this point. And now, if I going take any point between 0 to 1, it is still going to remain 0.

And when I include this point 1, when I make these x equals to 1, then this half will come into picture, in the sum, probability x is less than or equals to x, then x is 1, what can happen? With probability half it can take 1 and that is included. So, there is a jump of amount half here. But now between 1 and 2, can my random variable X take any value between 1 and 2? No, because it only taking 1, 2, 3 here. So, till 2, this is going to remain flat. And at 2 it is again going to make a jump.

And like that, when 3 it is going to make a jump and after that it is going to stay at 1, it cannot go beyond 1. So, this is a graphical way, another way of representing the same thing is you be more descriptive, let us say x is less than equals to 1 it is 0, between 1 and 2, it is half. Between 2 and strictly less than 3, it is 5 by 6. And for x greater than or equal to 3, it is going to be 1, I represent like this.

(Refer Slide Time: 12:37)



Now, because of this definition itself, CDF has certain properties, the very natural property that emerges says F(x) is non-decreasing in x. Everybody agree with this? Everybody see, it has to be non-decreasing. I am saying non-decreasing instead of increasing because it can remain flat. That is why we are saying non-decreasing can remain flat at some points. And now if I let x goes to infinity, what does F(x) of X, when x goes to infinity?

Students: 1.

Professor Manjesh Hanawal: This is because when I am letting x here to infinity, x is I am letting x, I am interested in let here x take any value, that means I am interested in the whole space outcomes space. So, that is why it is going to be 1. On the other hand, when I let x go to minus infinity I am letting x to be minus infinity here I am I want x to be less than or equal to minus infinity that means I am not considering any possible outcomes of x.

That means it becomes a null set and it becomes a 0 and that is why it is 0. Another property that emerges is something called right continuity. All of you know what a continuous function is? Anyone here who does not know, what is a continuous function. All of you know, what is the right continuous function? Anyone here, who do not know what is the right continuous function?

Student: (())(14:31).

Professor Manjesh Hanawal: You do not know. So fine. I mean, you are the only one who do not know talk to me later. And now, I mean, all this straight forward properties. Nothing I mean, you these are straightforward, and you actually go and actually prove them. Even though these are all pretty intuitive in this case. You can just go and apply the axioms we have and demonstrate that they are all indeed true.

Now, actually, probability is like, ultimately, CDF is capturing probabilities, it is basically cumulate, accumulating the probabilities. Now we can ask the questions like, what is the probability that X is going to take a value and X is between x and y. This is nothing but probability that X is going to take value less than or equal to y and minus probability that X is less than or equal to x. Like just think of like this, like I have this and I want to ask the question, what is the probability that my random variable is going to take value between this.

So, one thing is you consider all the values below this. And from this, you subtract the value below x, then you will exactly get this portion. And that is what it is doing. So, subtract take everything between y. And now take everything below x to subtract this will get and but this is nothing but F(y) - F(x) as per our definition.

So, this probability is nothing, once we know the CDF and you want to know the probability that your random variable lies between those two points, all you need to do is, so take the difference of your CDF at those two points. And our actual definition of the probability here, if you notice F(x) is X less than or equals to minus x, X less than or equals to x.

But then, what if I want to compute what is the probability that X is strictly less than x. We can compute it using the limiting notion. You just take F(x - h) and let h goes to 0 from the top side, you are approaching h in a positive direction and compute it and you will get it and since F(x) is defined for all possibilities, this is fine.

And F(x) and probability that $X \leq x$ need not be equals to probability that $X = x$. This may not be the case. So, just let us see the example. Let us say this is 1, 2, and 3 here. And there is a jump here. Now, in this case, let us say there is a jump of half here. So, in this case, probability that $X \leq 1$ is half. Now, what is the probability that $X < 0.9$?

Student: 0.

Professor Manjesh Hanawal: 0. What is the probability that X is less than, let us say 0.999. So, what is the probability that $X < 1$. It is still going to be 0. But $X \leq 1$, it is going to be half. So, notice that here, this amount of the jumps are actually corresponding to the probability those points.

So, here, the jump is like half that corresponds probability that $X = 1$ and here the jump, this much of jump is corresponding to 1/3, and that is a probability of $X = 2$. And here the jump is 1/6, and that is the probability that $X = 3$.