Probability and Statistics
Prof. Somesh Kumar
Department of Mathematics
Indian Institute of Technology, Kharagpur

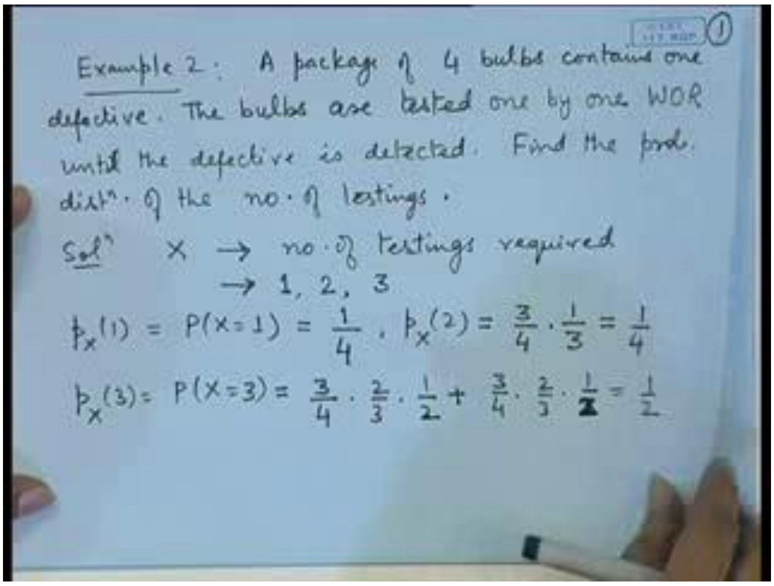
Lecture – 15
Probability Distribution of a Random Variable – II

In the last lecture, I have introduced the concept of random variables. So, the motivation for defining random variables is that most of the random experiments and their outcomes if we look at the complete description of that, it may be quite complex. Whereas, we may not be interested in the full description of the phenomena, we may be interested in a particular numerical phenomena connected with the random experiment. Therefore, we define random variable as a real valued function defined on the sample space.

Now, once we have a probabilities space omega b p and if we are defining a random variable x on omega then naturally the probability which are associated with the points in the sample space get translated to the values taken by the random variable. This correspondence between the probability allotments with the values of the random variable is called probability distribution. In the previous lecture I introduce the concept of accumulative distribution function, we showed that it is having a one to one corresponds with the probability distribution of a random variable and at the same time it gives almost all the information about the random variable.

Now, random variables are of different types; we have the cases where the random variable can take a finite number of values or countably infinite number of values or uncountably infinite values which we say that line in an interval. So, the first we called a discrete random variable. The allotment of the probabilities for a discrete random variable we call a probability mass function. So, let me give one more example of a probability mass functions consider the following.

(Refer Slide Time: 02:18)



A package of 4 bulbs contains one defective. The bulbs are tested one by one without replacement until the defective is detected. Find the probability distribution of the number of testings. So, let us look at x is the random variable which denotes the number of testings. Once we outline a random variable first thing is we should look at the set of values that the random variable can take. Here if we test it once that itself may be a defective or if that is not so second one may be defective or if that is not, so the third one may be defective or if that is not, so the forth one will be defective.

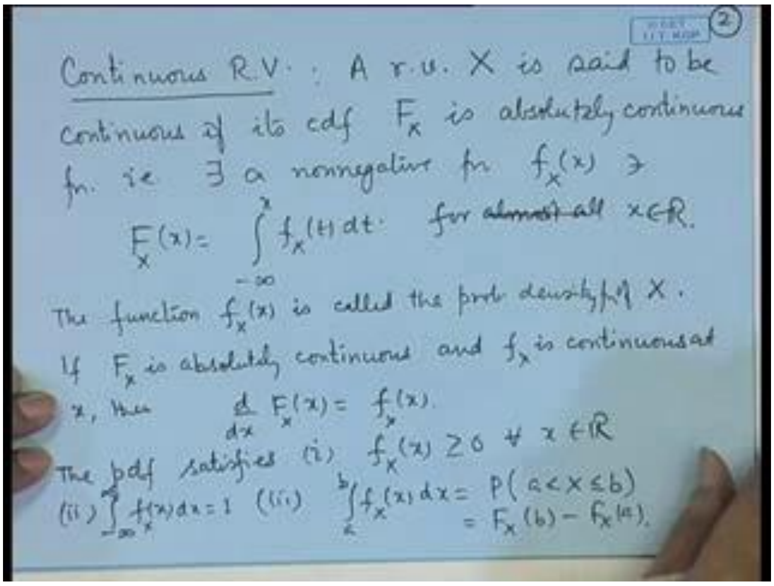
However, we should look at here that if the third one is found to be not defective, since there are only 4 bulbs in the package, so the forth one will be automatically treated as the defective. So, we in reality do not need the four testings; we need only 1, 2 or 3 testings.

So, what is the probability of x is equal to 1; now there are four bulbs out of which one is defective. So, the probability that the first testings gives you a defective will be 1 by 4; what is a probability that the second one is defective so; that means, the first one must not be defective, now there are only 3 bulbs left out of that one is defective. So, the probability of that is 3 by 4 into 1 by 3 that is equal to 1 by 4. If we look at p x 3 that is probability of x equal to 3, now by the direct logic it should be simply 1 minus the probability of x equal to 1 and probability x equal to 2, so it should be half. Let us see how it can be evaluated directly also.

So the first one is not defective, second one is not defective, and the last one is defective but this is equal to 1 by 4, we should also take into a count the possibilities that the third one is non defective because either it is defective or not defective in both the cases, the outcome is known that the forth one is will be defective or not, so it is actually equal to half. So, this is the probability distribution of the random variable x which is the number of testings required.

Next let us denote define the probability distribution of a continuous random variable. So, what is a continuous random variable? We say that a random variable x is continuous if its CDF is an absolutely continuous function.

(Refer Slide Time: 06:26)

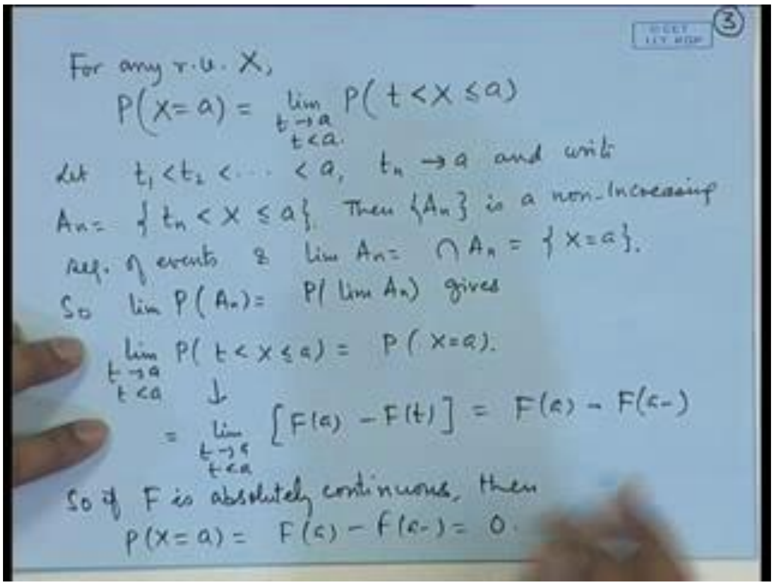


A random variable x is set to be continuous if it is cumulative distribution function F is absolutely continuous function that is there exist a non negative function is small F x such that capital F x is equal to minus infinity to x F t d t for almost all x, basically it should be for all x belonging to R.

The function F x is called the probability density of density function of x or pdf. If F is absolutely continuous and F is continuous at x then d by d x of F x is equal to a small F x. In the pdf are the probability density function satisfies; one that it is a non negative function. Secondly, the integral over the full region is equal to 1. And if we integrate the density over an interval a to b, it is denoting the probability of a less than or equal to x less than or equal to b that is basically F x b minus F x a.

Now, the continuous random variable is quite different from the discrete random variable. So, the first thing that we noticed here is that in the discrete random variable we had a probability mass function. So, p x i; it denoted the probability that the random variable x takes values x I; here at the point x, we have a density function F x. Now this is one should not get miss lead and say that this is probability that capital x is equal to small x. In fact, in the discrete case the mass function defines the probability of that point in the continuous case, it does not define in fact probability of every point is 0.

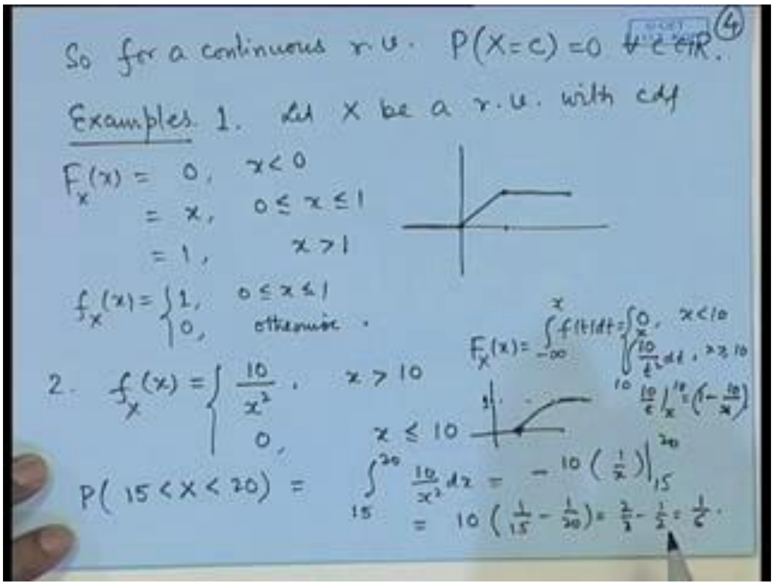
(Refer Slide Time: 10:12)



So, that is a important observation to proof that let us see, firstly that for any random variable x probability that x is equal to a is equal to limit probability t less than x less than or equal to a t tends to a where t is less than a. So, if we denote let us proof this is statement let us consider say t 1 less than t 2 and so on; t n goes to a and write a n to be the set the t n is less than x less than or equal to a; then this is a non increasing and limit of a n will be equal to intersection of a n which is basically equal to x is equal to a.

So, limit of probability of a n is equal to probability of limit a n this statement is equivalent to limit t tend him to a t less than a probability t less than x less than or equal to a is equal to probability x is equal to a. Now this is actually equal to limit F a minus F t where t tends to a t is less than a which is exactly equal to F a minus the left hand limit at a. So, if F is absolutely continuous function then probability x equal to a will be equal to F a minus; F a minus is equal to 0.

(Refer Slide Time: 12:49)

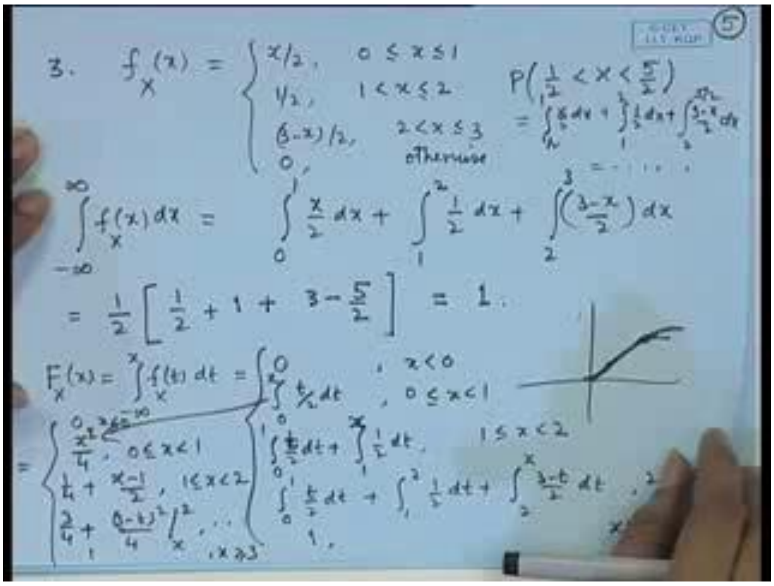


So, for a continuous random variable probability that x is equal to some value c is 0 for all c. Let us consider some examples here, let x be a random variable with cdf F x is equal to 0 for x less than 0 is equal to x; for 0 less than or equal to x less than or equal to 1; it is 1; for x greater than 1. Quite obviously, this is a absolutely continuous function and in fact you can plot it for x less than 0; it is 0, for 0 to 1 it is x and there after it is 1. So, the density function is given by 1 for 0 less than or equal to x less than or equal to 1 and it is 0 for x lying outside the interval 0 to 1.

Let us consider some more examples suppose f x is given by 10 by x square; for x greater than 10 and 0 for x less than or equal to 10; I want to find out what is the probability 15 is less than x less than 20 then it will be equal to integral from 15 to 20; 10 by x square d x that is equal to minus 10; 1 by x from 15 to 20 that is equal to 10; 1 by 15 minus 1 by 20. So, 2 by 3 minus half that is equal to 1 by 6 if you want to find out the cdf F then it is minus infinity to x f t; d t.

Now, here the positive value of the density starts from the point 10 therefore, it will be 0 for x less than 10 and thereafter it will become 10 by; integral of 10 by t square. So, that will give me from 10 to x for x greater than or equal to 10. So, it is equal to 10 by t from x to 10 that is equal to 1 minus 10 by x; this is for x greater than or equal to 10. So, if you want you can plot this function up to x equal to 10 it is 0 and at x equal to after that it becomes 1 minus 10 by x. So, the function is increasing because 10 by x is decreasing function as x tends to infinite, this goes to 0; that means, it goes to 1. In fact, if you look at the derivative of this it becomes like this basically a concave function this is suppose 1 here.

(Refer Slide Time: 17:24)



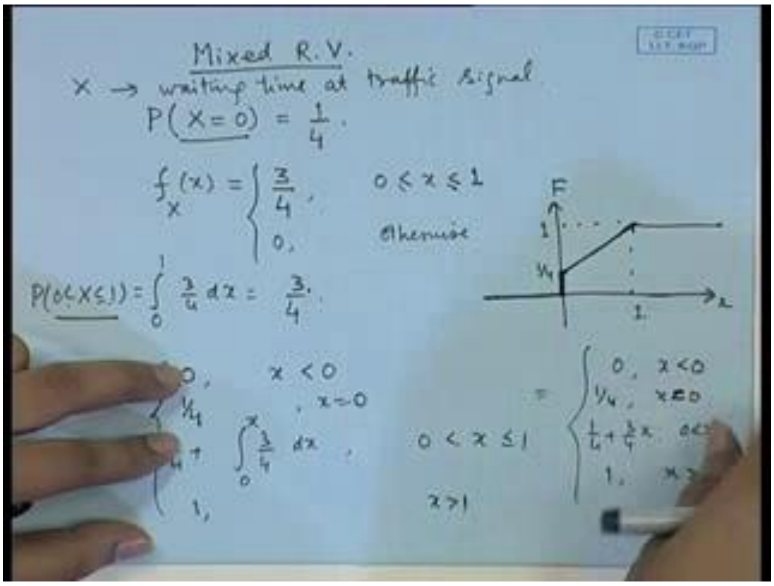
Let me take one more application of this, suppose x is a continuous random variable with the density function say x by 2 for 0 less than or equal to x less than or equal to 1; it is half for 1 less than x less than or equal to 2, it is 3 minus x by 2 for 2 less than x less than or equal to 3, it is 0 at all other points. First thing is we should notice whether it is a proper probability distribution then the integral of f x; d x over the full range must be equal to 1; obviously, it is non-negative function. So, this will be equal to integral 0 to 1; x by 2 d x plus integral 1 to 2; 1 by 2 d x plus integral 2 by 2 to 3, 3 minus x by 2 d x.

So, if you integrate each of these terms; it becomes in fact half can be kept outside then you will get half plus 1 plus here 3 minus x square by 2. So, it becomes 9 minus 4 that is 5 by 2, so clearly this is equal to 1. If we want to write down the cumulative distribution function of this then minus infinite to x f t; d t, so for x less than 0 we are integrating only 0, so this will be equal to 0. In the other regions for 0 to 1, we have to integrate the term t by 2 d t from 0 to x. If we look at x lying between; so this is for x less than 0 for x between 0 to 1; we have this term for x between 1 and 2, we have 0 to 1; x by 2 or t by 2 d t plus 1 to 2 half d t; that is 1 to x, it will be 0 to 1 t by 2 d t plus 1 to 2 half d t plus 3 minus t by 2 d t from 2 to x and it is equal to 1; for x greater than or equal to 3.

So if we simplify these terms; we will get here let us look at this one; it is equal to x square by 4, here up to 1; it has been integrated therefore, this will be 1 by 4 plus x minus 1 by 2. Once again here if we have integrated up to 2 then this value would have become 3 by 4 plus the value of this term that is 3 minus t whole square by 4; from x to 2 which will be some value; it is 1 for x greater than or equal to 3; 0 for x less than 0. So, here you can see this function will be having different values; x square means it is a convex function up to 1, then there after it is a line and there after again it is a parabola kind of function which is actually in the reverse, so it will go to 1.

However in each case, if you differentiate it will get the density function. If you want to find out probability of a certain set; suppose I say what is a probability that x lie between half to 5 by 2 then it will be integral of x by 2 from half to 1 plus integral of half from 1 to 2 plus integral of 3 minus x by 2 from 2 to 5 by 2 which can be evaluated after certain calculations. There is also a possibility that random variable may be partly discrete or partly continuous; these are known as mixed random variables.

(Refer Slide Time: 22:39)



Let me consider some example consider for example, a person traveling to his office everyday by car. On the way to the office there is a traffic crossing, it may happen that some of the days, the traffic crossing has a green signal will be approaches the traffic light and he crosses without waiting. On the other days there is a red light and he as to wait for a certain amount of period, in other words he may have if I say x denotes the waiting time at traffic signal, it may happen that probability that x equal to 0 is say 1 by 4; that means, 25 percent of the time he is able to go without waiting. On the other hand if he as to wait then it is a continuous distribution he may have to wait for 0 to 1 minute.

If we see the total allotment of the probabilities, in this case the probability of 0 less than x less than or equal to 1 is actually equal to 3 by 4. So, probability x equal to 0 and probability of x lying between 0 to 1 is 3 by 4, so if you add up these two; it is actually becoming equal to 1, but here the distribution is not for totally discrete or totally continuous random variable. We can also look at the cdf of this particular random variable; at the 4 x equal to 0; it is 0, at x equal to 0 it is becoming 1 by 4 and there after it is becoming 1 by 4 plus integral 3 by 4 d x from 0 to x and of course, it is equal to 1 for x greater than 1. So, if we expanded it is becoming 0, x less than 0; 1 by 4 for x less than equal to 0 and here it is becoming 3; 1 by 4 plus 3 by 4 x for 0 less than x less than or equal to 1, it is 1 for x greater than or equal to 1.

You notice here that the function is not continuous at x equal to 0 and therefore in fact, there is a jump point, the right hand limit and the value at x equal to 0 is 1 by 4 whereas, the left hand limit at 0 is 0. So, there is a jump of size o1 by 4 at 0. So, that shows that discrete nature of the random variable at x equal to 0; whereas, after 0; 0 to 1 it is a continuous random variable. So this is an example of a mixed random variable, if we plot this particular distribution cdf then up to 0, it is 0; at 0 it is becoming 1 by 4 then from 1 by 4 to 1 it is increasing and it is 1.

So in this case it is a continuous random variable, there is a jump here of size 1 by 4. So, we will look into these things in the next lecture.

Thank you.