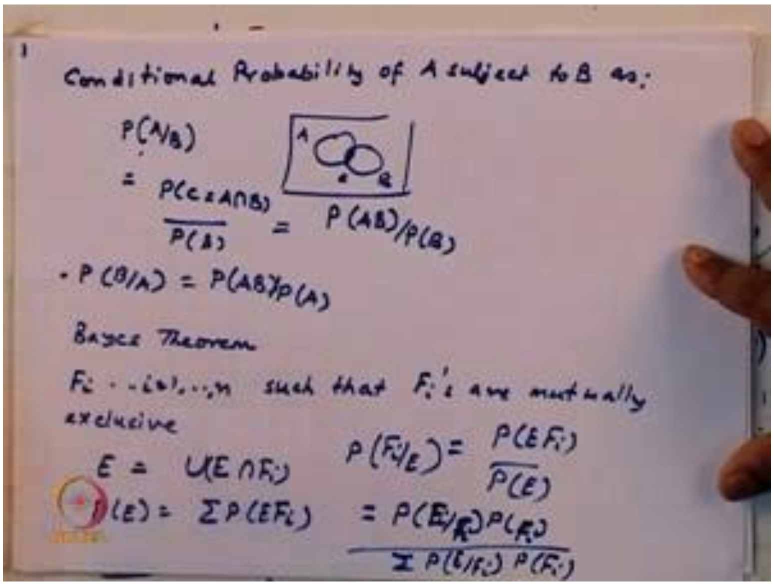
Introduction to Biostatistics
Prof. Shamik Sen
Department of Bioscience and Bioengineering
Indian Institute of Technology, Bombay

Lecture – 19
Random variables, Probability mass function, and Probability density function

Hello and welcome to today's lecture. So, we will continue with probability and random variables in today's class, but briefly I wanted to mention what was covered in class previous class and one of the most important concepts we had brought up in last class was the idea of conditional probability right.

(Refer Slide Time: 00:39)

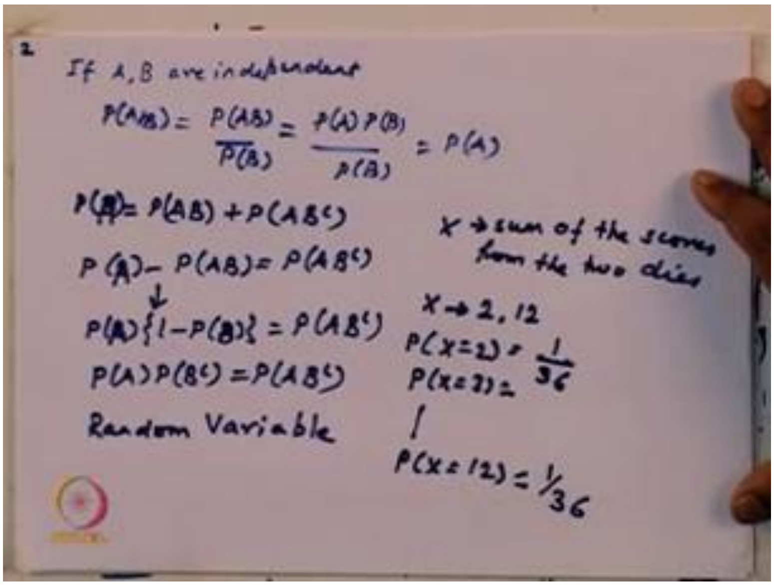


So, you write conditional probability of A subject to B as P of A given B right and what we had found that, if you have two events A and B like this right. So, P of A given B is basically the shaded zone, that this C. It is probability of C, which is equal to A intersection B by probability of B. So, this is probability of A B by probability of B. Similarly, you can have probability of B given A as probability of A B by probability of A. Another important idea, we discussed was Bayes Theorem, this also applies the idea of conditional probability. So, if you have events F i; i equal to 1 dot, dot n such that, these are F is or mutually exclusive.

So, you can write some event E as simply the union of E intersection F i. So, in that case you can write P of E summation as P of E F i. So, if I am asked to compute probability of

F i given E, I can write this as probability of E F i by probability of E and I can change the extent of conditioning, I can write probability of F i given E into probability of E by sorry. This is E given F i into probability of F i and summation of probability of E given F i into probability of F i. So, using the idea of P of A given B you can calculate P of B given A.

(Refer Slide Time: 03:13)

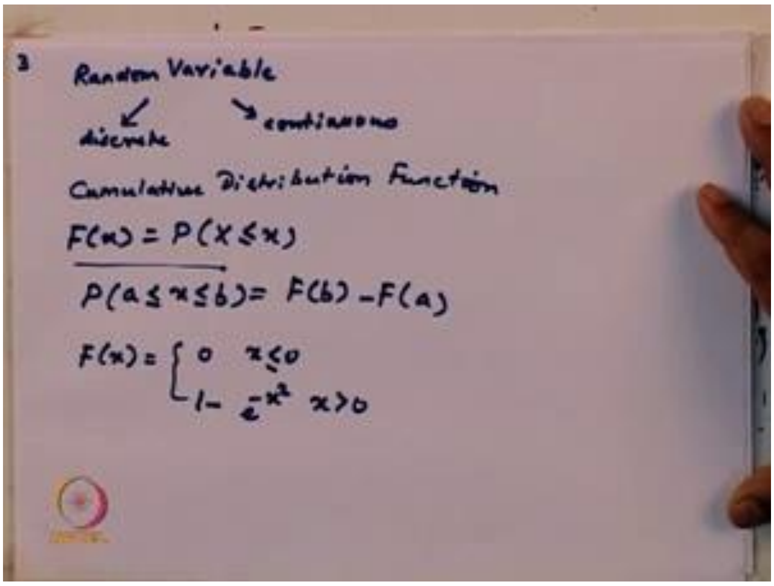


So, if A and B are independent then, probability of A given B. So, does not mean anything. So, I can, if I write it as probability of A B by probability of B and probability of A B for independent events is simply probability of A into probability of B, by probability of B is simply the same as probability of A. I take a different pen, one more thing.

So, let us say if A and B are independent I can write B is equal to A B. So, probability of B is equal to probability of A B plus quality of A B complement right. So, I can rewrite this. So, I can take it to the other side and I can simplify this as P B 1 minus P of A, sorry this is P A sorry, sorry, sorry this is A 1 minus P B is equal to P of A B complement and this is nothing, but P of A into probability of B complement is equal to P A B complement. So, this shows, if A and B are independent then A and B complement are also independent. So, towards the end of last lecture, we had discussed about random variable. So, and we had specifically discussed in the case of our, you know toss of a die twice. You can define a random variable X which is sum of the scores from the two dice.

So, if X is the score. So, I can have any X will be varying from 2 to 12 and we had derived that, we had computed each of the probabilities X equal to 2 onwards, X equal to 3 so on and so forth, twelve. So, for example, probability of X equal to 2 is nothing, but 1 by 36 and probability X equal to 12 is also 1 by 36 because 6 and 6 is the only one which corresponds to P of X equal to 2.

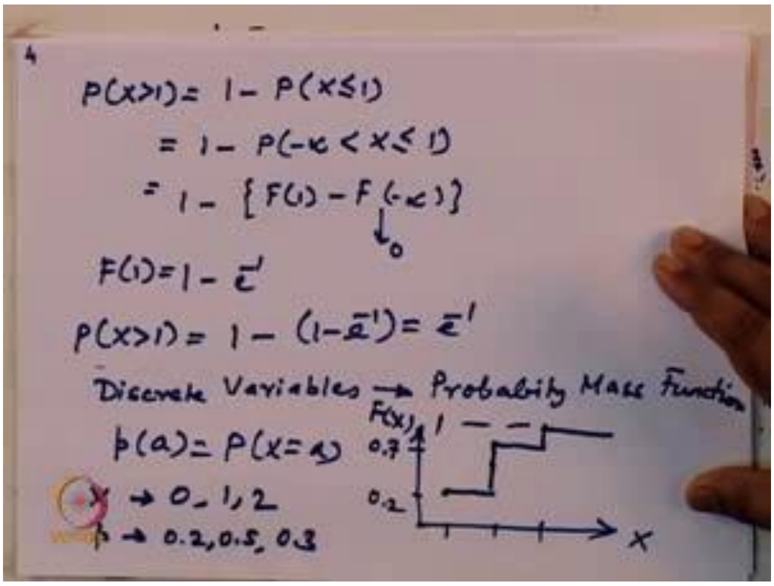
(Refer Slide Time: 06:06)



So, for random variables, a random variable we can either have it discrete or continuous. So, just you know the idea of sum of two scores is an example of a discrete variable, but you can have any variable which is continuous. For example, let us say lifetime of a given electronic device which can take on any value. So, which can have its essentially a continuous variable. So, we can define something called cumulative distribution function.

So this is typically represented of F of x and this is defined by probability of X less or equal to x. So, if we know F x then, we can compute probability of any probability let us say of x between a and b and this is nothing, but F of b minus F of a.

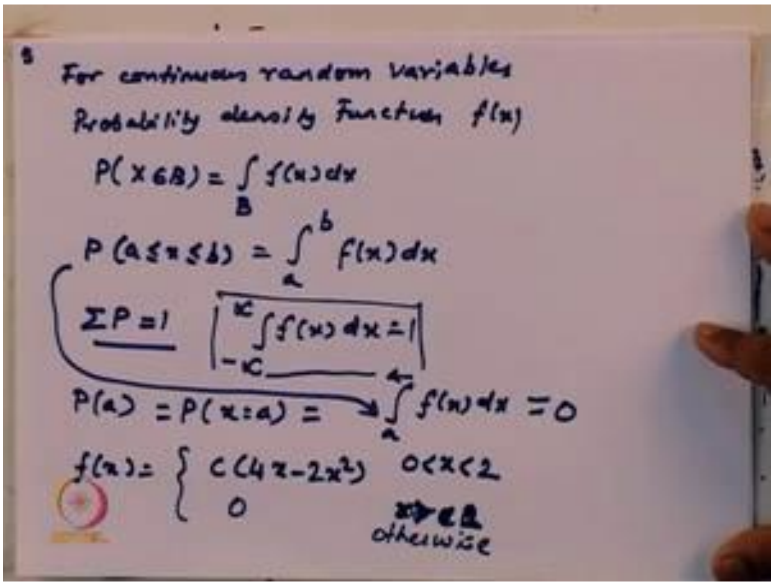
(Refer Slide Time: 07:39)



So, let us take a simple case. So, imagine I define F of x as follows; greater than 1, 1 minus probability X less equal to 1. So, this is simply F of 1 minus F of minus infinity. So, you know for all these values this is nothing, but 0 and F of 1 is given by 1 minus e to the power minus 1. So, probability of X greater than 1 this becomes 1 minus, so 1 minus e to the power minus 1 is equal to e to the power minus 1. So, this is how you can calculate probability, if you are given a cumulative distribution function.

So, I can for discrete variables. So, I can have something called a probability mass function and probability mass function is given by p of a is probability of X equal to a. So, if X is a random variable which, takes the value 0, 1, 2 and the probabilities, let us assume is 0.2, 0.5, 0.3. From here I can compute the cumulative distribution function. the cumulative distribution function, so you have X. So, probability, so this value is equal to 0.2, this value is equal to 0.5. So, you have 0.7 here and from here you have the value of 1. So, this is how your F of X will look like. This value is 0.2, this value is 0.7, and this value is 1.

(Refer Slide Time: 09:58)

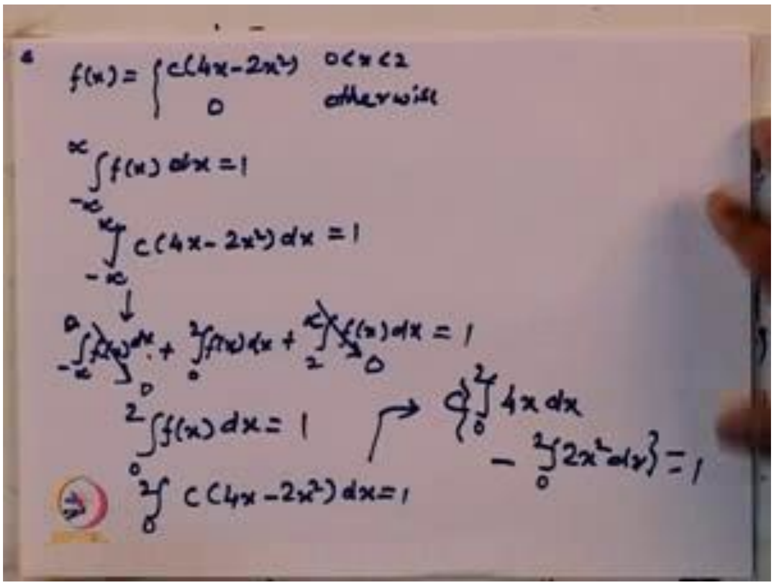


So, for continuous random variables, I can write what is called a probability density function. So, probability density function let us say it is indicated by x and I define the probability of X belonging to some regime B is simply the integral of f x d x over this interval B. So, if you have probability of let us say a less equal to x less is equal to b, I will simply integrate a to b f x d x. So, if this is true and then how do I define the total probability, I must have the function? So, you have summation of P must be equal to 1 right. So, what I essentially have is minus infinity to infinity f of x d x must always be equal to 1.

At the same time, what is probability of a? Is probability of x equal to a. So, I can simplify from this equation, I can write it is simply a to a f x d x and what is a to a f x d x? It is nothing but 0. So, in for a continuous random variable any integral or probability of an absolute value is always 0 versus you must always have this which has to be followed by the probability density function.

So, let us assume you have your given apologies function f x is defined as follows. C, so x greater than 2, you have the following function given f x is C times 4 x minus 2 x square, for 0 less than x this is 2 and it is 0 otherwise. So, we can actually say, 0 otherwise because x can have both positive and negative values. We want to calculate what is the value of C.

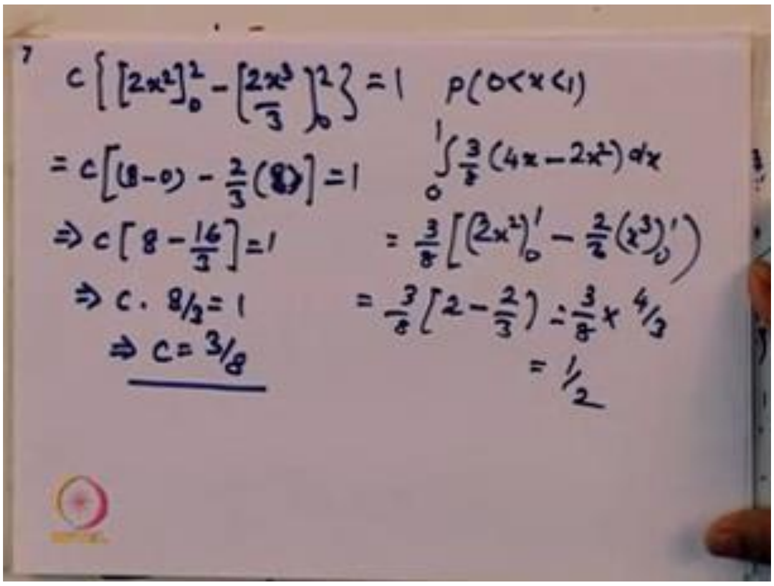
(Refer Slide Time: 12:34)



So, I have been given f x. So, I must have this constraint satisfied at all conditions. I can plug the values here. So, I have C into 4 x minus 2 x square d x is equal to 1 from minus infinity to infinity. So, I can break this limit into minus infinity to 0. I am not writing the entire thing, I can break the interval from 0 to 2 plus 2 to infinity f x d x, f x d x, f x d x this is 1.

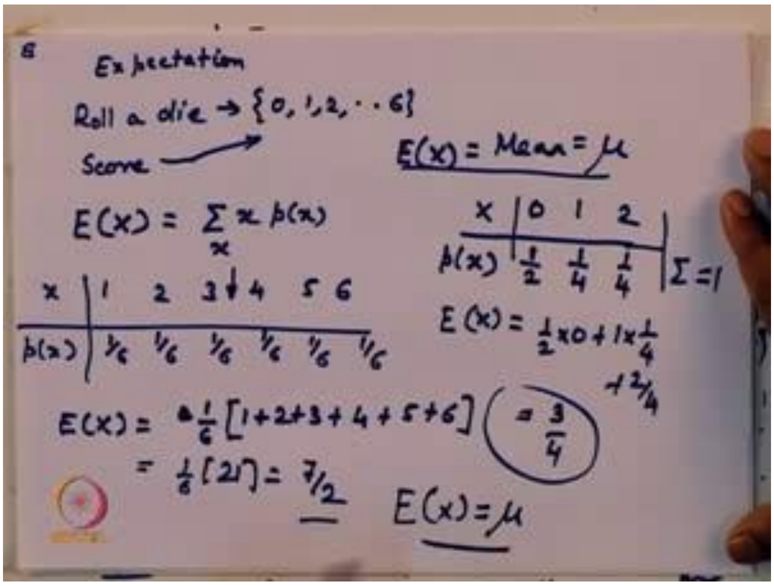
Now, for minus infinity to 0, we have know, we know that f x is 0 for every place other than this interval. So, this is 0, this is 0. So, I am left with a condition 0 to 2 f x d x is equal to 1. I can plug in the value 0 to 2 C into 4 x minus 2 x square d x equal to 1. So, this takes me to, I can take out the constancy. This is integral 0 to 2, 4 x d x minus integral 0 to 2, 2 x square d x equal to 1.

(Refer Slide Time: 14:18)



So, 4 x plus 2 x square minus equal to one. So, this gives me C into 8 minus 0 minus two third into 2, 8 equal to 1. 8 minus 16 by 3 (Refer Time: 15:02) push. So, I get C as 3 by 8. So, if let us assume this is right, we can compute, what is the chance of probability 0 less than x less than 1? I can compute that probability as 0 to 1, C is now 3 by 8, (Refer Time: 15:37), 2 x square into 1 minus equal to 3 by 8, 2 minus two third 6 minus 2, 4 by 3 equal to 3 by 8 into 4 by 3 equal to half. So, probability 0 less than x less than 1, you can calculate. So, this is how you calculate the expression for a probability density function using the constraint that minus infinity to infinity f x d x must always give you a value of 1.

(Refer Slide Time: 16:32)

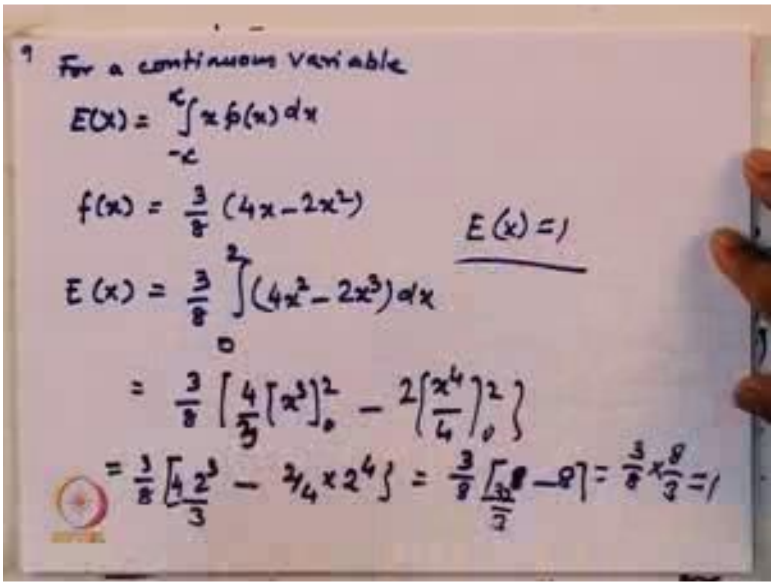


So, we introduce one more important concept today, which is the concept of expectation. So, imagine you have a diving rolled, you roll a die. So, you know the values can be anywhere between 0, 1, 2, dot, dot 6 and let us say your score is equal to whatever value you get. Roll of a die gives you the score and you are asked to compute, what is the on an average, how much do you expect to score? And that brings us to the conception of expectation. So, expectation of a random variable X is defined as E of X or expectation of X and it is defined by summation of x p x or all values of x. It is summation of x p x over all values of x. Now let us compute the expectation for the value of this for rolling of a die. So, you know that you have six values; 1, 2, 3, 4, 5, 6 are your values of x and the probability for each of them is simply 1 by 6.

So the expectation for this case will be nothing, but 1 dot 1. So, I will take out 1 by 6 in common, 1 plus 2 plus 3 plus 4 plus 5 plus 6. So, what value do I get? So, I have 3, 6, 6, 12, 16, 21; 21 by 6 I can simplify it, it comes to 7 by 2. So, if you look at your values of x what is this value 7 by 2? What is the average? So, on an average if you were to do this your average of 1, 2, 3, 4, 5, 6 is nothing, but here which is going to be 7 by 2. So, in general, it is always 2, you will see that the expectation of a value of X, E of X is nothing, but the mean mu of X is always going to be the mean. So, let us take another case. So, let us say my X values can be 0, 1, 2 and the probability values are half, one-fourth and one-fourth.

So, you see summation of probability is always is equal to 1. So, summation will give you the value of 1. Let us compute the value of expectation in this case; E of X will give you a value of half into 0 plus 1 into one-fourth plus 2 into one-fourth. So, 2 by 4 is equal to 3 by 4. So, if you do if you take these as your weights, you will see you will always your weighted average of these is nothing, but this value itself. So, you have always expectation gives you the value of mean.

(Refer Slide Time: 20:27)



So, for a continuous variable, your expectation is given by just x p x d x. So, for the probability density function that we had f x is equal to 3 by 8 into 4 x minus 2 x square, what will be the expected value? We can calculate the expected value X as. So, I can take 3 by 8 out, I can integrate. So, this is only from 0 to 2 remaining all places you have you know that the probability is 0. So, you can have 4 x square minus 2 x cubed into d x is equal to 3 by 8, is cubed by 4, x cubed 0 to 2 minus 2 x 4th by 4, 0 to 2 equal to 3 by 8. So, 2 cubed is what I get minus. So, I can take 2 by 4 into 2 to the power 4 is equal to 3 by 8, 8 minus 8 (Refer Time: 22:03) the right thing x square minus expected value, x cube by 3 sorry, sorry, sorry; this is x cube by 3, 4 by 3. This has to be 4 by 3, this is 8. So, basically it was a 32 by 3 is equal to 3 by 8 into 24, 8 by 3 is equal to 1. So, you get for this particular distribution your expected value is 1.

So, I would summarize today's lecture by saying that. So, today we started with random variables discussed, how you can use cumulative distribution function for calculating probability in a particular domain. We introduced for the random variable what is called the probability mass function or the probability density function and how using the equality that your integral minus 2 infinity to infinity f x d x is equal to 1, you can find out if there is any constants in your probability density function. And towards the end, we discussed about expectations. So, we will continue in next class. Again we will rediscuss expectation and then take it from there. With that I thank you for your attention.