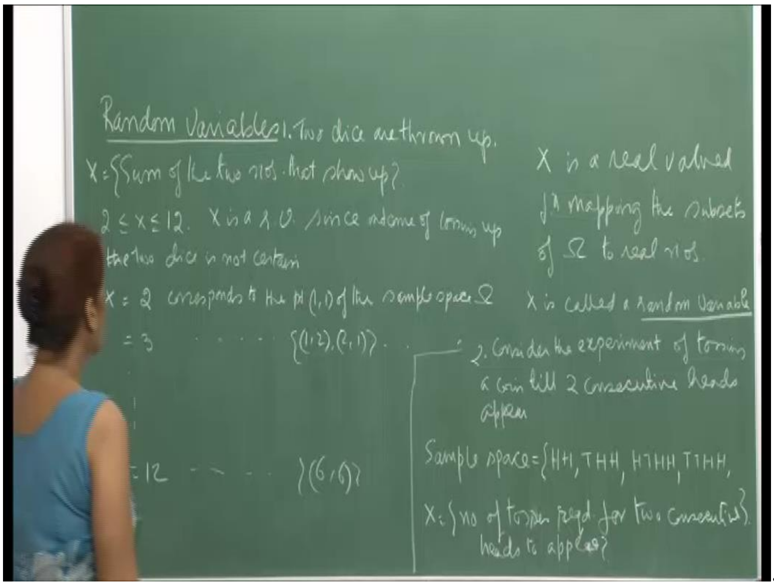
Introduction to Probability Theory and its Applications

Prof. Prabha Sharma
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur

Lecture - 5
Discrete Random Variables and their Distributions

(Refer Slide Time: 00:19)



I will now introduce the concept of random variables. So, the whole idea is let me begin by giving some examples first. Suppose two dice are thrown up, and they are fair dice. So, you can say that may be two fair dice are thrown up. Now, let X denote the sum of two numbers that show up, that means the two numbers whatever 3 4, 1 2 whatever the two number show up when I throw the two dice, I add up the sum, I add up the two numbers and I denote that sum by X. Now, we can see that the number the values of X will vary from 2 to 12.

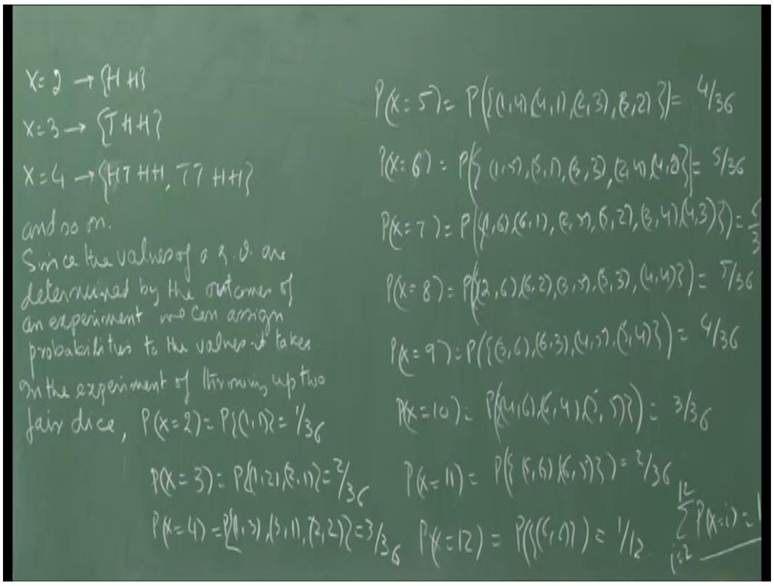
Now, since the outcome of tossing of the two dice is not certain, so that is a random phenomena, because any phase can show up, so any number can show up. Therefore, you see that the value of X is dependent on the tossing of the coin and whatever the outcome. So therefore, this is what we mean by a random variable, so this is actually now you can see that. Before I start calling it a random variable I should explain more, see when X is equal to 2 when it corresponds to the point 1 comma 1 of the sample space. I had earlier in my lecture shown you, when we were talking about when I introduce the concept of sample spaces.

I had shown you that omega the sample space will contain 36 pairs of such points i j, where i varies from 1 to 6 and j varies from 1 to 6. So, when X is equals 2, this actually corresponds to the outcome 1 comma 1 of the sample space both the phases much show one each, and therefore the sum will be 2. Similarly, X is 3 then it can be either 1 comma 2 or 2 comma 1, so both of them add up to 3 and so on. So, essentially trying to say, so one can now give a formal definition of random variable that is, so X is a real valued function that maps the sample space omega into real line.

And we call, so these X what we have described so through examples, this X is called a random variable and subsets, because these are all subsets this is the singleton, this is the subset containing two points of omega and so. This is the again the number when X is 12, it will again corresponds to the singleton 6 comma 6. So, such a function which is a real valued function, and when it maps the subsets of the sample space corresponding to an experiment to real numbers, we will say that it is a random variable. Now, if you again taken other example considered the experiment of tossing a coin, till two consecutive heads appear. So I toss a coin and unless I get two heads consecutively I will not stop.

Now, the sample space can be if it happens in two tosses then I will get the outcome will be H H null stop here, but if you does not show head in the first trial, then it may show in the second and third, so that means this will happen in three trial, three tossing of the coin and so on. So this will continue, and this may not have finite process this may not be a finite process, because you may continue throwing up to a tails. So now, X is number of toss is required for two consecutive heads to appear.

(Refer Slide Time: 04:01)

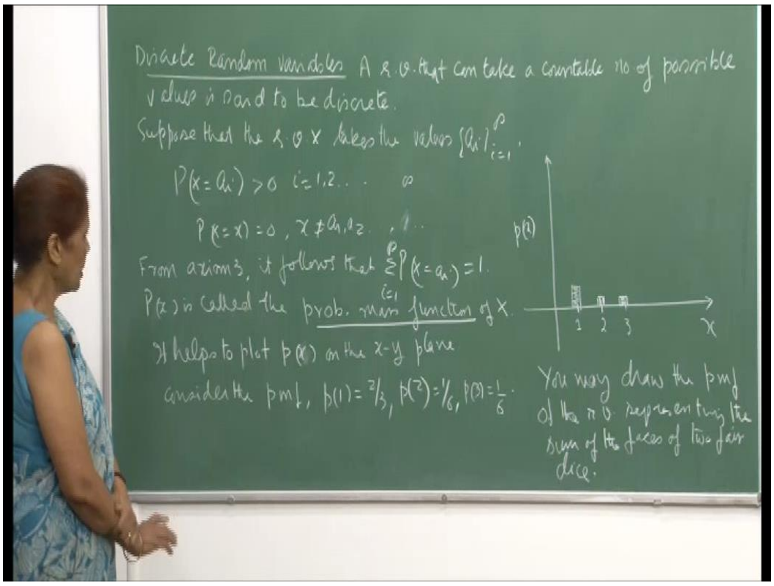


So, you see here for example in this case X will be equal to 2, then it corresponds to the outcome H H. If X is 3 then it corresponds to the outcome T H H. First toss gives you tail, and then the other next two tosses give you heads. Then X equals to 4 will correspond to this, and so you can go and writing different values of X, what will be corresponding subsets of the sample space omega. So, again I will reiterate the same thing that since the value of a random variables are determine by the outcome of an experiment. We can assign probabilities to the values it takes, because the probabilities associated with the outcomes of the sample space.

And since random variable is mapping subsets or this event from the sample space to real numbers, I can assign probabilities. So, for example in throwing up of two fair dice, see even probability X is equal to 2 would be then in that case, probability of the singleton 1 comma 1 which is 1 by 36, because every pair each of the 36 pairs is equally likely, since the two dice of fair. Then when probability of X equals to 3 would be corresponding to the probability of the subset containing the pairs 1comma 2 and 2 comma 1, so this will be 2 by 36. So, just for your benefit, in fact you can complete you for completed the table by yourself.

So, this I will shown you for all different values of that X can take from 2 to 12, what are the corresponding subsets, and then the associated probabilities. And since, one of the X must take one of the numbers, since X must take one of the numbers from 2 to 12, since you are throwing up two dice. So, two numbers will show up and there sum will be one of the numbers from 2 to 12. So, then these are all possible around events that can take place. And so, probability X equals to I from to i vary in from 2 to 12 should be 1, so this is the probability mass function, as we are I going to formally define the concept of probability mass function now.

(Refer Slide Time: 06:17)



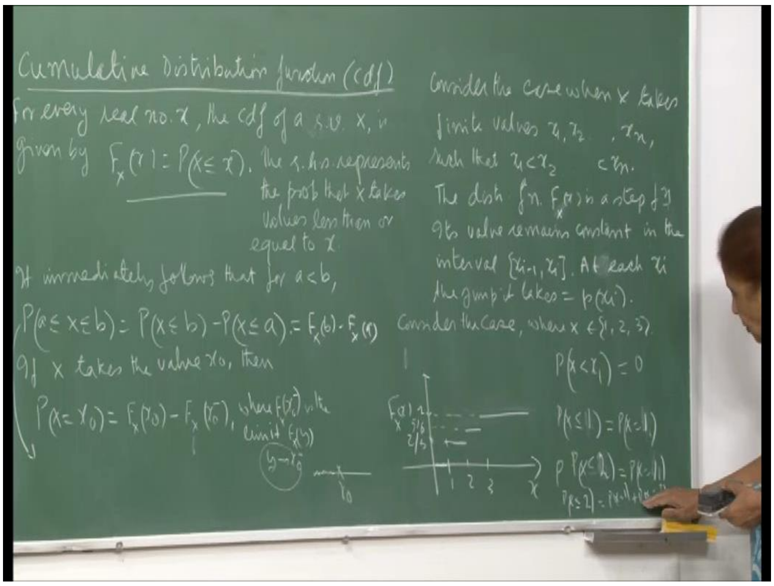
So, the random variables can be of different types, let me first consider the case when X is the discrete random variable. So, as the names suggest, it means that it takes countable number of possible values. And let say that the random variable takes the values a i, i varying from 1 to infinity. Now, of course when I say countable countably infinite or countabley finite, it can be either case, but the values are sort of you can enumerate the values that it will take. Now, so therefore we will say that probability X equal to a i is positive, because these are the values that it takes and therefore they will be positive probability associated with each of the these numbers.

So, this will positive and for all of the values of X it will be 0, that means when x is not equal to a 1 a 2 up to these thing, whatever the number of a i, then the probability at points which are not equal to any of these values is 0. Now from axiom 3, since X must take one of the values a i, therefore when you add up the probabilities 1 to infinity, all these probabilities must add up to 1 by your axiom 3. Now, this function which assigns probabilities to different values of a random variables that is called and in the case when it is a discrete random variable, we call it the probability mass function.

And as we have already seen the any probability function must satisfy 3 axioms. So here, in short we also write pmf for probability mass function, all the time you can are go and writing these 3 letters. So, normally I will be referring to it as pmf. Now, it helps to plot p x on the x y plane, so consider the probability mass function p (1) is 2 by 3, p (2) is 1 by 6 and p (3) is 1 by 6, so the random variables here is taking the values 1, 2, 3 and these are the probabilities associated with it. So, you can draw a bar chart, so the idea is that you I know erect rectangles, so the height here is, this height is 2 by 3 and this height will be 1 by 6. So, that means the rectangle bar is centered at the value 1 and the height is 2 by 3, so this is the idea.

Similarly, here this height of the bar is 1 by 6 and same as 3. Now, you I have already was to give you one other random variable, which shows giving you the sum of the numbers that show up when you throw to a fair dice. So, you can try to draw the bar chart for that random variable that will be a big one, because the values will takes from 2 to 12. Now, so having defined random variable or the discrete case. Let me now associate some other functions with it so cumulative distribution function, so probability mass function we have already defined.

(Refer Slide Time: 09:28)



Now, this is the cumulative distribution function, which again I will be referring to as cdf in the short form. So for every real number x the cdf of a random variable X is given by this equation. So, capital F x of x is the notation for the cumulative density function, and this is actually the probability of X less than or equals to the real number x. And so here this right hand side is actually the probability that your random variable X takes all values less than or equals to x, this is the idea. It immediately follows that if you want to compute the probability X greater than a and less than or equal to b.

Then this is this can be written as probability X less than or equal to b, minus probability X less than or equal to a. Sees that is why this is X greater than a, because you are subtracting for all sample points for which this capital X is less than or equals to a. So, the probability of that, so therefore this becomes X greater than a. And this in terms of our cumulative distribution function, we write F X (b) minus F X (a). Now, all along the notation for this probability is cumulative distribution function, and if some place by mistake I call it as cumulative density function, then that has to be ignored the proper terminology is cumulative distribution function.

If you want to find out the probability of X lying in an interval, then you can define this in terms of, why should have actually said this is as F X (b) minus F X (a). So, you can immediately see, because here this is actually the equal to the event that X is less than b and X is greater than a. So, X greater than a implies opposite of X less than a, so therefore minus X less than or equal to equals to a, because then this gives you the values of. If you feel that you still need from explanation you can do it for yourself to say that, see for basically what you are saying is that from X less than or equal to b, you want to subtract the because a is less than b.

So, you want to subtract all values of X less than or equals to a, then you will get the this thing that X lies between a and b. Now, if X takes the value x naught see, because after all X is a discrete random variable, then the notation would be that F X (x) naught, and then F X (x) naught minus, so this actually says that your approaching the value x naught from the left. So from values which are less than x naught, so where F X (x)naught minus is the limiting value of F X (y), where y goes to x naught minus. And this notation means that, if you are number here is x naught, then your approaching they number x naught from the left of x naught minus this.

So, and this we will be clear in a minute, because so now let me just spell out the values here, so essentially what I am saying here is that. If x 1, x 2, x n are the let say the finite n values that the random variable x takes, then an x 1 is less than x 2 is less than x n, see than what is happening is? That the distribution function F x is the step function, because you see you say probability X less than x 1. Then this is 0 in this case, because it is not taking any values less than x 1, and x 1 is the smallest value here. So this probability 0, but then if you want to say probability X less than or equals to x 1. If you now do this, then what happens?

This is equals to probability X equals to x 1, because it is not going to take any other value any value less than x 1, only value in this region that X will take is equals to x 1. So, this is actually equals to, so in this case because is the first value this is equal to x 1. And so this is what happens? Now, so that means up to values less than just less than x 1, your when if you want to draw the graph, which I am showing you here, I am drawing the graph fine. So, let me show you the graph for this first, you sees what is happening is that here the random variables taking the value 1. So, before that the value will be 0. So therefore, if I want to draw the cdf the cumulative density function for that random variable, then you see it is 0.

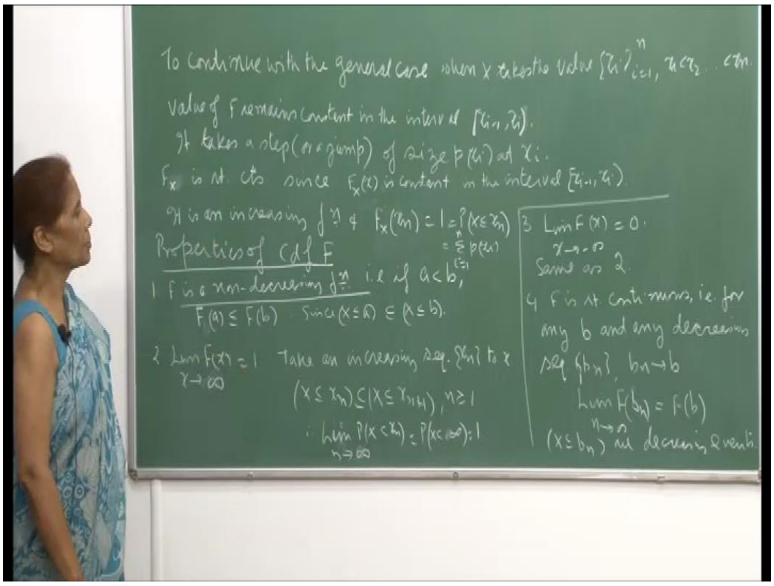
And it is 0 till at the point 1 it takes a jump, because at point 1 when x 1 is 1 this will be equal to probability of X equal to 1 which is 2 by 3. So therefore, the function from 0 will take the jump, and so it will be this thing. And then you see, for probability X less than or equal to less than x 2 if I do this. Then here it is again continuous to take this only, because less than x 2 or if you want to write 1 here in this case and this is 2. Then as long as X is strictly less than 2 the number there is no other probability, because X is only continuous to take the value 1, and here also it takes the value 1.

So, as long as X is strictly less than 2, that means as long as I am here some way here and not taking the value 2, my value of the cumulative density function remains constant. So, this is like a step function, it continuous to be the same value as probability X equals to 1. And then the moment I say probability X less than or equal to 2, then this will be probability X equal to 1 plus probability X equal to 2. Because now, for X less than or equal to 2 there are two possibilities X can be equal to 1, on X can be equal to 2. So, then the two probabilities will get added up, so that means it will be 2 by 3 plus 1 by 6 which will be 4 by 6 was will be 5 by 6.

And you see so from 2 by 3 it takes a jump, and at 2 the value now becomes 5 by 6. And so the jump, and you can see that this jump that it takes is equals to the probability of that discrete random variable at that point, that means probability X equals to 2. So, till up to this point this was probability X equals to 1, the moment I said probability X less than or equal to 2, it become probability X equals to 1 plus probability X equal to 2. So, the value of the cumulative distribution function jump by the probability of X equal to 2. And finally, when you talk about probability X less than 3 again, it will continue to be this, because there will be no other value of X here. And then the moment you make it less than or equal to it will take a jump, the figure is not accurate it will be this.

And so I here again, the moment I say equal less than or equals to probability of X equals to 3 will get added to it, and so again the jump will be by the probability, so that means this will be a discrete function or and a jump function or a step function whatever you can call. And the point of discontinuities or the point of jumps that it takes will correspond to the values of the random variable, and the amount of jump that the functions takes will be equal to the probability of the equal to the probability of that of the random variable, taking that particular value where you considering the jump.

(Refer Slide Time: 17:54)



To continue with the general case, when X takes the values x i, i varying how 1 to n, and x 1 is less than x 2 less than x n. Then value of f remains constant in the interval x i minus 1 comma x i, because it takes the value x i minus 1. And after that it does not take any other value when a variable, so the cumulative density function remains the same. And you see this is the sign this say that is a closed at this end, and this is the open that means in this interval the values began from x i minus 1 to all values which are less than x i.

So therefore, for all these values your cumulative density function remains constant, which I will interpreter saying that because it is constants in this interval. Therefore, it is right continuous, because I am approaching from here, if I approach from the right hand side that means larger value then x i minus 1. Then I will approach x i minus 1 this value of the function remains the same it is a constant, so it continuous and it attains the value at x i minus 1. So, the same value and therefore the function is right continuous, and what we just saw is that it takes a step or a jump equal to the size of the probability x i. So that means at x i, it takes a jump which is equal to the probability of the random variable at that point x i.

And also we have see that it is an increasing function, and since this will be when you want to compute this, see this is equal to probability X less than or equal to x n. And by this mean it has taken all the values for X less than or equal to x n means it has taken all the values x 1, x 2, x n. So all the probabilities has been added up, actually so this is nothing but summation x i, i varying from 1 to n, and therefore it must add up to 1. So, this is what, now let me formulize properties of the cumulative distribution function. So, as we have seen that the function has to be a increasing function. So that means what do we mean by, when we say a function is increasing, that is the a is less than b.

Then the value of the function at a must be less than or equal to the value of function at b, and this can be easily explained, I have already done it through example, but you see that the event X less than or equal to a is a subset of the event X less than or equal to b, because b is bigger than a. So, all the values that are that give you this event also all the points of sample space which give you this event also are here. And therefore, as we have seen from our proposition using the axioms of probability, that probability of this event must be less than or equal to probability of this event.

And that is what this represent, this represent for probability of this event, and this represent for probability of this event. Therefore, this in equality follows, and so the function is a increasing function is a non decreasing function. Then limit F (x) as x goes through infinite is 1, so we will take an increasing sequence, let say of values x n to x, increasing that means values keep on increasing. So we approach that means if you have an x here, and you are approaching x from here, so all these values are increasing and you are reaching up to x.

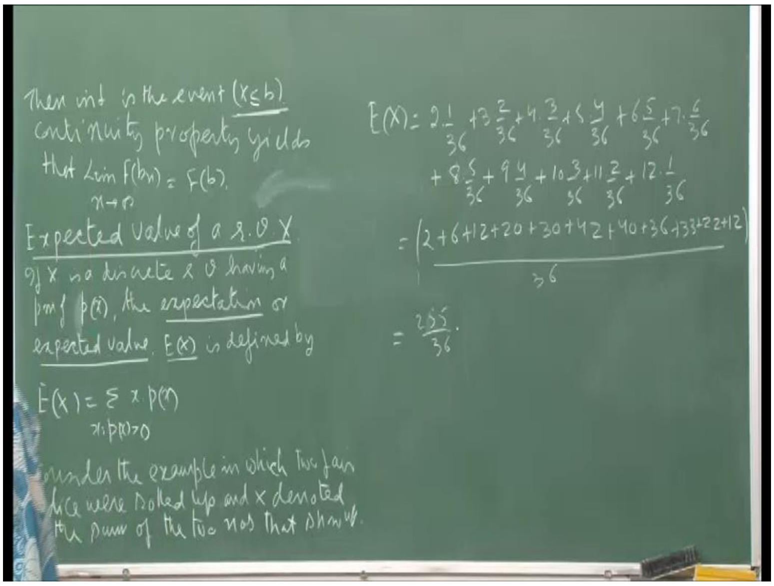
So, then again because of this property the event X less than or equal to x n, this are the subset of the event X less than or equal to x n plus 1, because x n plus 1 is bigger than x n. And therefore, limit probability X less than x n as goes to infinity. Actually, because this goes to infinity, so therefore this events the merge into X less than infinity, so all possible value of X get covered up just as. So therefore, this becomes equal to probability X less than infinity, and therefore this must be 1 because all value of X get covered up in this event, and so this must be equal to 1.

Then we say that the limit F (x) as x goes to minus infinity is 0, so the argument here is same except that here we took an increasing sequence to x n. There we will take decreasing sequence of this is as X goes to minus infinity, so limit F (x) is 0. So we can argue, because here I am say that as X goes to infinity F (x) is 1. So, when you take the decreasing sequence, events will be that means if I take decreasing sequence x 1, x n, where this is greater than this, then this is greater than x 2, this is greater than x n. So, then the event could be when n goes to infinity, because I am taking decreasing sequences.

So, then it will become empty set, so this will be probability, see you will be considering the event X less than x i. And this is bigger than probability X less than or equals to x i plus 1, so this is the whole idea, so just reverse the argument. And so therefore as you go on, there will be nothing common as n goes to infinity this will be nothing common to… Because I am saying x goes to minus infinity, and so here there will be nothing common, and so this will finally converge to this will become probability X, what shall I say here. The symbol you have to uses that you have to say that x is empty, let me this becomes probability of the empty set that is what will happen?

So therefore, the limiting value of F (x) as x goes to minus infinity must be 0, because there will be no values of x that are possible, once x goes to minus infinity, so same as to. Then f is right continuous, because any b and any decreasing sequence b, and you take a b and any decreasing sequence to b n. So, same thing I am saying your approach b from right, your approaching this of the sequence b n is coming like this from right hand side. And limit F (b n) as n goes to infinity is F (b), because same thing here X less than or equal b n or decreasing events, the sequence b n convergence to b.

(Refer Slide Time: 24:48)



So the events converge to X less than or equal to b, and so what we are trying to say that X less than or equal to b n, this event will contain the event X less than or equal to b n plus 1. And finally, will also all of them will contain the event X less than or equal to b. And so therefore, probability of X less than or equal to b n will be greater than or equal to the probability of X taking the values less than or equal b n plus 1and so on. And finally this, so now by the continuity property of the probability function P. We get that limit F b n will be equal to F (b), because this is the value you know you are taking the limit here as n goes to infinity. So therefore, because P is a continuous function probability function which is continuous, therefore this will be equal to F (b), and so these proof the right continuity of F.

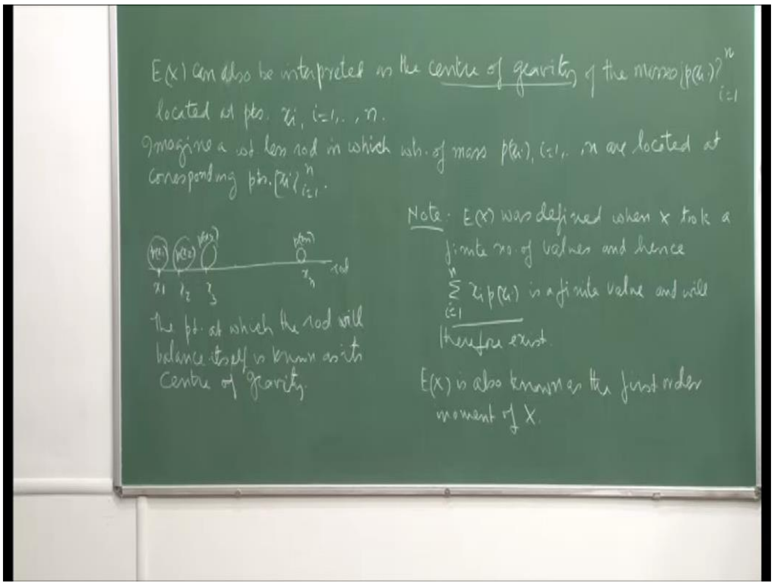
And so now we have shown all the four properties 1, 2, 3, 4 of the cumulative distribute function, and so any cumulative distribution function must satisfy all these four criteria. And in fact, these 4 conditions are necessary and sufficient for any function to be a cumulative distribution function corresponding to the random variables X. So, this is importance, so whenever you want to characterize a function which is which you says a cumulative distribution function for a random variable X, then you have to show that these 4 condition are satisfied by that function before you can take it to be the cumulative distribution function.

Now, so if you take the example that we have been referring back to all the time. This is when two dice are rolled up, and X denotes the random variable X denotes the sum of the two numbers that show up, then the expect so this is fine, so that is it. Now, I started give you the example, but first let me now define a very important term a commodity or quantity which we associative with the random variable expective value of a random variable X here. So, if X is the descriptive random variable having the probability mass function p (x), the expectation or expected values E (X) is defined by E (X) is equal to sigma x into p (x) such that, so that means you multiply x by it is corresponding probability, and then you add up.

So when you add up all this products, so which is over all x so that p (x) is positive, because if p x is 0 then the contribution here will be 0, so we take this some over all possible x is for which p x is positive. So, if you add up these values when we define this as the expected value of the random variable X. Now, so therefore we consider this example now, consider the example it which two fair dice were rolled up, and X denoted the sum of the two numbers that show up.

So, in that case had given you the table of you know for different values of X what will be the probabilities, so if you just refer to the table, then you can come see that this will be the thing, and this add up to 255 upon 36 which will be some number close to 7 little bigger than 7. So, this is the expective value of the random variable, that means in other words what you saying is that, if you sort of keep throwing the two dices, and at the numbers and then that means you take the average, so that means suppose you throw up the number 100 times throw up the two dice100 times. And then add up the numbers that show up, and then divide that by 100 that will be close to your expected value.

(Refer Slide Time: 28:57)



Now, here if you look at this expression what does it say? Now, since p x for all x such that p (x) is positive is 1, this you can say that expected x is the weighted average of the values that X takes with weights as the corresponding probabilities. So, they can be some more interesting interpretation of the expected values which are you show you right now. So, E (x) can also be interpreted as the center of gravity of the masses p (x i), i varying from 1 to n located at points x i, i varying from 1 to n that means you can imagine that the p (x i) are the masses, which you place at the point, corresponding points x i. And then you take the center of gravity of this distribution of masses is also which is same as the expectation X.

Now, see that means imagine a weightless rod in which weights of masses p (x i) are located at corresponding points x i, i varying from 1 to n, so that means imagine that this rod weightless rod, and you have placed these points the masses p x 1 at x 1, p x 2 at x 2 and so on. So, I have taken the points to be x 1, x 2, x n you know this one distribution, but it could be anywhere distributed but whatever the diagram will be the same, that p x 1 is the mass located at x 1, p x 2 is mass located at x 2 and so on. Now, the point at which the rod will balance itself is known as it is center of gravity.

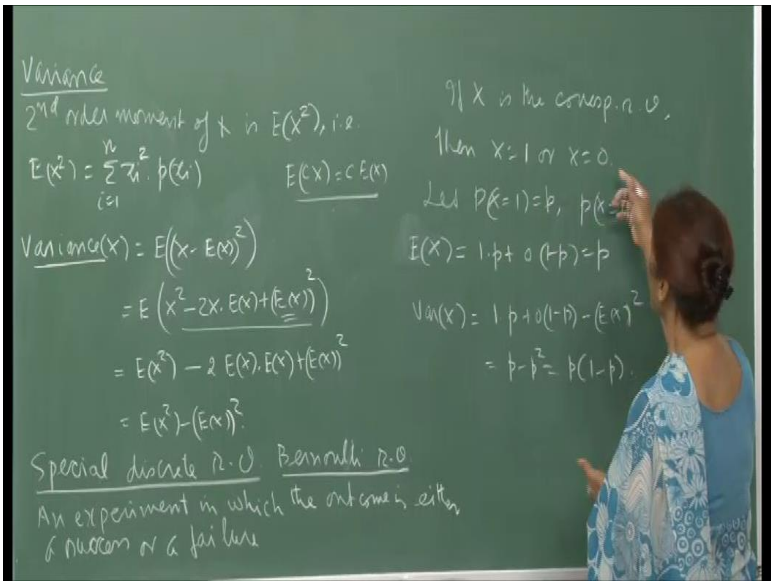
So, from this thing you can this is the notation, and so this is exactly what is given by the expression E (X). So E (X) can be the notation it can also be referred to as the center of gravity of these different masses, which are the probabilities located at the corresponding points x i. Now, note that here I have just taken when I find E (X), I took x 2 I said that x takes finite number of values and hence this quantity is a finite quantity, and therefore it is defined therefore it exist. So, wherever the different cases I will discuss them as that when we are right. Now, we also refer to E (X) as the first moment of X.

So, for when X is the discrete random variable and it is taking finite number of values. So then I can also define expectation X square which will be sigma x i square p (x i) for all i varying from 1 to n, obviously the summation is over, all those i for which p (x i) is positive. Now, even if the random variable X is taking countably infinite values, then also if then I can you know take any function of X, and I can accordingly write down the expectation of that function of a random variable. We will formally define expectation of you know g of x, when g of x is some function of a random variable X.

So, that we will take care later on, and of course for a continuous random variable also we will define in a formal way, but right now I can just say that because it is a summation sign. So, and if as long as the summation is finite number, I can define these expectations and so here for example this will be the second moment, and also the linearity of expected function, because it is a summation. So therefore, it is a linear function that means if I take c x plus d y two random variable x and y, then also I will be able to write the expectation of c x plus d y will be c times expectation of x plus d times expectation of y.

So, because of this summation thing, of course if x and y have the same to a probability mass function. Then so what I am saying is that right now, wherever if I am using the linearity of the expectation, then I am doing it this scenario when your x takes may be i finite values or countably infinite values, when wherever this summation is finite number. I can treat E as a linear function, and also I can define the expectation as a of a any function of a random variable in this way. Now, here I would like to top of an example also.

(Refer Slide Time: 34:17)



So let see, then the second moment can also be defined here, which is the second moment if this is the first moment, then the second moment would be expectation of X square, which will be summation i varying 1 to n x i square p (x i). And then very important quantity that we are associate with random variable one is of course E (X). And the second one is variance X, which is the expect that means it is now the expectation of X minus E (X) and then whole square. So, you say that this is the moment second order moment of X about it is expected value.

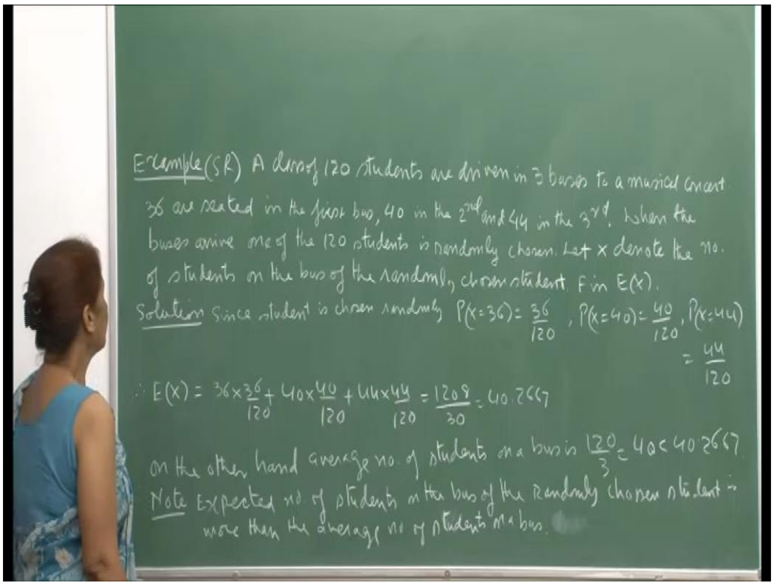
See these are all for example this is the moment E (X) is the first order moment about the origin, and this is about 0 this is also the second order moment about 0, but now here this is the second order moment about it is expected value. Now if I open up this bracket when I get this, and now when I will take E inside, because as we have seen writes definition expected value is distributive, I can take it inside the bracket. So this will be E X square minus since E (X) is finite quantity already, so they will be E (X) here, so 2 comes out constant.

In fact I am assuming that what I am saying is that, if you take C some constant, then this will be C times expected X, which immediately follows from the definition, because as here defining this as summation x i p (x i). So if I consider the random variable C X instead of when C X will takes the values C x i, and so here when I want to compute this will have C is present here, but since C is a constant till come out. And so this will be simply C X.

So therefore, when I take E inside, this will be to twice E X into E (X) then this is again a constant. So, therefore this is simple be E (X) square there will be no expectation. Again I am using the property that if random variable is taking constant value, then wills I this is not a random value essentially, this is a constant. So, if X C for all possible values then will be the expectation was simply be C, because this is probability 1, so the constant whenever random variable is just equal to a constant this expectation would be just that constant value. So this is this, and therefore minus 2 X square plus E (X) square, so that reduces and therefore this becomes E (X) square minus, so therefore you can also say that the variance of a random variable X is a variable of random variable X is second order moment about the origin. So, expected X square minus it is expectation whole square, so this is what we. And therefore we will go and computing this as we introduce our special random variables and so on. So, the first one of the simplest random variable that we talk about is a Bernoulli random variable, this was this is name after the swish mathematician James Bernoulli, and I think 1700 something probably he define this random variable. And you will see that it is a very basic random variable, and it we use this to build up on other special kinds of random variables. So, this describes a situation or this random variable describe the situation in which the outcome is either success or a failure, so very simple you perform an experiment and the outcome would either be a success or failure. So, for example if you toss a coin, you can say that coming of a head is a success and coming up of a tail is a failure. And so, the values that X will take, we just see we associate X equals to 1 with the success, and X equals to 0 with the failure. And then let us say that probability p X equal to 1 is p and so where of course p is a number which belongs to 0, 1.

And I am taking the open interval on both side, that means p is not 0, I am defining a meaning full random variable or a meaning full experiment, in which the outcome is either a success or a failure. And so the probability X equals to 1 will be p, and probability X equals to 0 will be 1 minus p. So, now if you compute the expectation or the first moment of this random variable, then this will be simply one in to p, because the random variables taking the values 1 or 0. So, 1 in to p plus 0 in to 1 minus p, and therefore this is equal to p the probability of a success. And the variance by this formula, so when you compute the E (X) square, so here E (X) square 1 square is 1, so 1 square in to p plus 0 square in to 1 minus p, so which again is p. So, the where second order moment is also p, and so this is p minus the first order moment first squared which is p square, so p minus p square so this is the variance. So, where you simple quantities which you can write away compute, and now we will further on use other special random variables discrete random variables, and then of course one will talk about continuous random variables also.

(Refer Slide Time: 40:14)



So, let me just illustrate interesting aspect of the expected value, here I am taken this example functions in rows, a class of 120 students are driven in 3 buses to a musical concert. Now, 36 are seated in the first bus, 40 in the second and 44 in the third, when the buses arrive one of the120 students is randomly chosen, from the group of this all get down from the bus, one student gets picked up. Let X denote the number of students on the bus of the randomly chosen student, see now we carefully understand the event that I am telling you, X is the number of students on the bus of the randomly chosen student.

So, we picked up one student randomly, and then now you want to the because again that is a random process have chosen one of the student out of 120. Then X is the random variable, which denotes the number of students on the bus, on which this randomly chosen student was sitting. So, we have to find E (X), so find the expected value. So, first of all you see that since student is chosen randomly, any student is equally likely out of the group of 120 I choose any one of them.

So, therefore the probability of the student in chosen from the first bus X equal to 36 is 36 by 120, then probability X equal to 40, because the second bus, so that will 40 upon 120, because that many students are travelling in that by the second bus, and finally probability X equal to 44 will be 44 upon 120. So, now if I want to find out the expected value of this random variable, then the expected value will be the random variable takes the value 36 into the probability of that bus being chosen, so it was 36 upon 120, then 40 into 40 upon 120 plus 44 into 44 upon 120.

Now, when you add up these numbers this comes out to be 1208 by 30 in which is 40.2667, but if you just compute the because if you look at the event that you take any bus, then the probability of being chosen is 1 by 3, it because every buses equally likely. So, if you want to compute the number of students on the expected number of students on the bus of the randomly, this is later on.

Now, on the other hand average number of students on a bus is 120 by 3, yes because I will add up, because each bus is equally likely to be chosen, so that probabilities 1 by 3, so that in to 36 plus 40 plus 44 that will be 120, so this is number is 40 now this number is less than 40.2667. And this what I want to point out here is that you see the expected number of students on the bus of the randomly chosen student is more than the average number of students on a bus.

So, just think about it and why is this happening, because you see the bus in which the largest number of people or I should say the more the number of students on a bus the more possibility of that student being chosen as be a student, it because you know the large number of students are coming from that bus, in which more students were travelling. So, the possibility of choosing that student is higher than choosing from other students. So, just think about this thing and so I thought I will end this lecture by giving you this interesting example. And so you will as we go on we will see various, various implication uses of these measures expected value variance and so on will introduce some more.