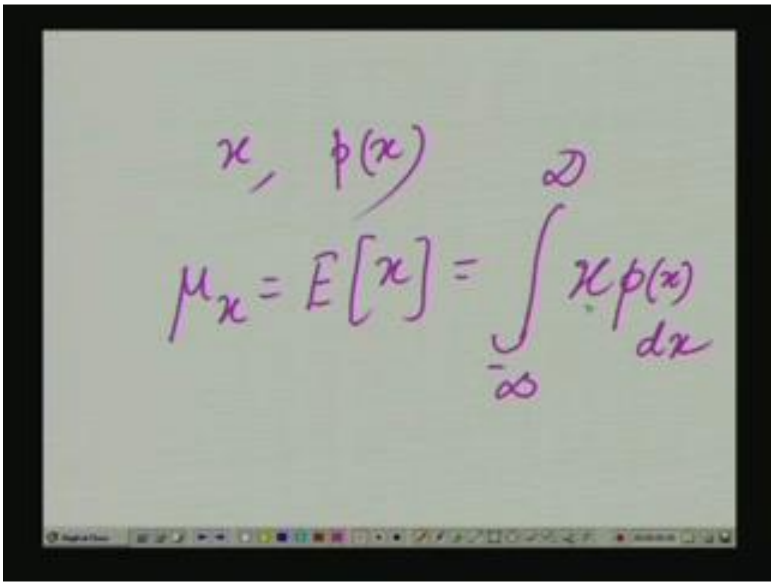
Probability and Random Variables
Prof. Dr. M. Chakraborty
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

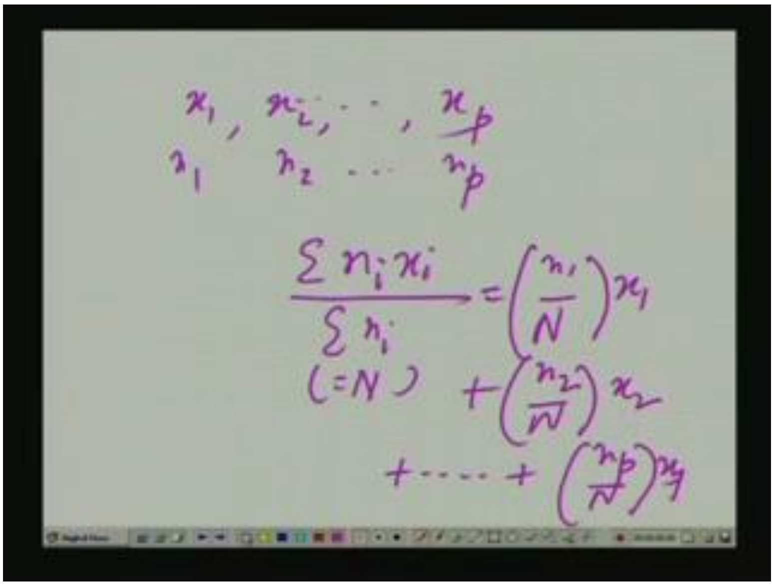
Lecture - 9
Mean and Variance of a Random Variable

(Refer Slide Time: 01:08)



So, today we discuss some important concepts like mean and variance. We are given a random variable x and its probability density function is px x for the time being is continuous random variable. Then we define it is mean mu x as also expected value it is an operator E actually, means expectation it is an operator it is just a question of notation expected value of x as. Now, you can I mean how this formula comes maybe we can give a frequency interpretation.

(Refer Slide Time: 01:57)



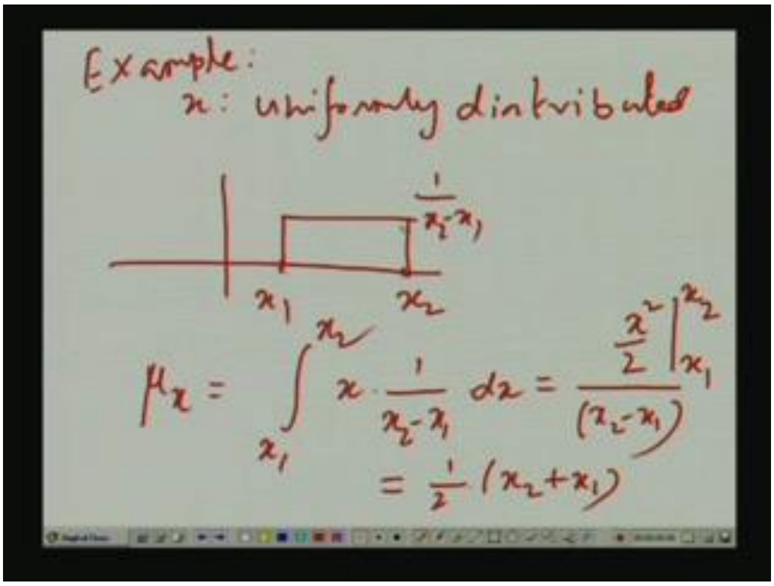
Suppose, in the discrete case suppose I mean x to x some values like you know x1 or x2 up to say xp this is n1 times. I mean if you observe x may be on n1 location, you get x1 on n2 location, you get x2 dot dot dot dot on np location you get xp. So, what is the arithmetic average net value is summation ni xi divided by summation ni this is the total number of trials.

If you can you can call this as N in that case, what you have is this n1 by N x1 plus n2 by N x2 plus dot dot dot plus np by N xp right. Now, what is n1 by N if out of a trial of you know very large number of trials equal to N. If n1 on any 1 location, you get x1 N1 by N as you have seen earlier it approximates the probability of x taking the value x1. So, you are multiplying x1 by its probability.

Similarly, you are multiplying x2 by its probability and so on and so forth and you are adding. So, this as per the discrete case, but you can easily extend this concept to the continuous case there this summation becomes continuous summation which is nothing but the integral and instead of the discrete probabilities you put the continuous probability function that is the probability density function.

So, you get px x dx or x px dx integral over the range of x which is in the most general case from minus infinity to infinity. So, it gives you the average we will call it a statistical average or mean or expected value of the random variable x. Take an example.

(Refer Slide Time: 04:08)



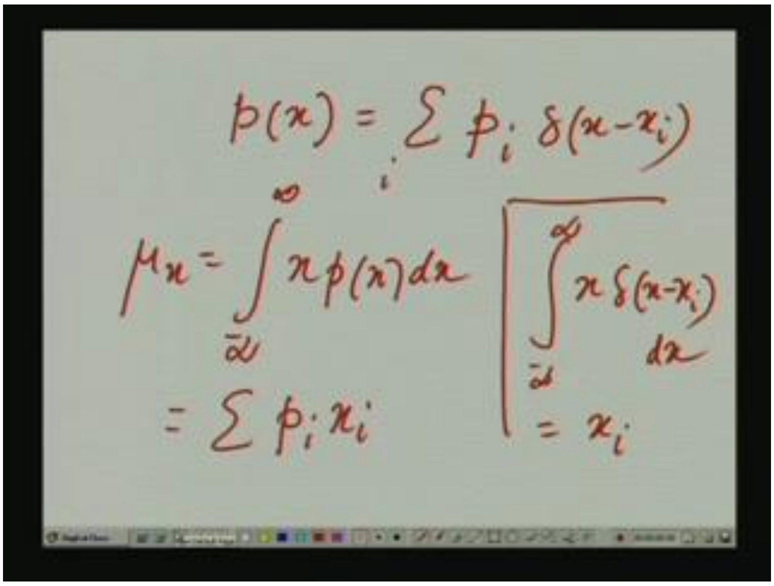
Suppose, x is uniformly distributed this is an example. May be there are 2 values x2 x1 uniformly distributed means, within this range it has a constant probability and outside this range its probability is 0, right. Since, the total probability has to be equal to 1 if this width is x2 minus x1 the area has to be 1. So, this is height will be 1 by x 2 minus x1.

So, in this case what is mu x. Mu x is x into 1 by x2 minus x1 this is the probability density function times dx and x is varying over the range from x1 to x2 this is constant you can take it out. So, it becomes x2 minus x1 if you integrate x, you get x square by 2. So, x square by 2 from x2 to x1. So, x2 square minus x1 square you factorize 1 factor cancels from top and bottom and finally, you get this quantity x2 plus x1.

So, just the average of these 2 values x1 plus x2. So, if it so happens that you know x2 is on this side x1 is on this side and they have the same value same magnitude. Then, the mean will be origin that is if x2 and x1 if x2 equal to minus x1 or x1 equal to minus x2. Obviously, this will be 0, otherwise the mean is at the centre point between x1 and x2. This for uniform you can try with other cases also.

So, well we get the formula for mean for a continuous random variable, but you can easily extend it to the discrete case. In the case of discrete random variable x will take either some value x1 or x2 or x3 like that. Some discrete values, not continuous values and for each value you have got some probability density function.

(Refer Slide Time: 07:04)

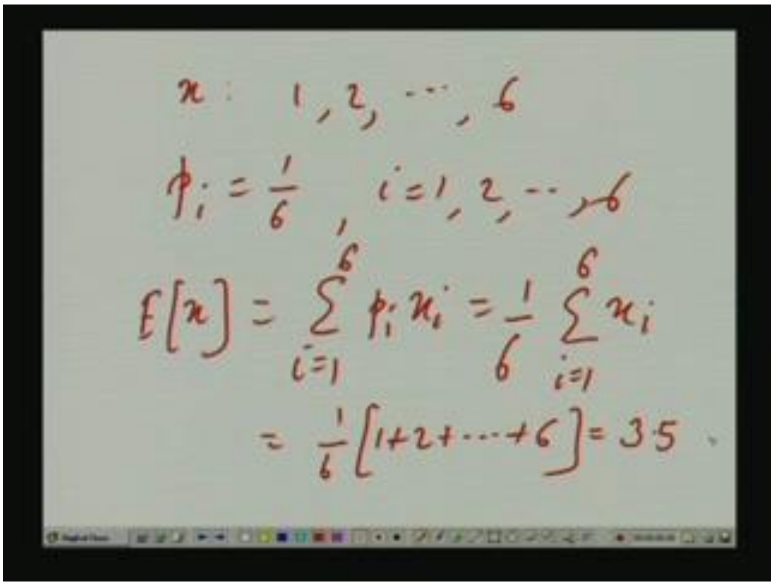


So, for a discrete case if x is discrete, then px is nothing but this we have seen already pi delta x minus xi. So, x can take either value x1, then the probability density has an impulse with string p1 or x can take x2. So, probability density is an impulse there with string p2 and likewise. So, I am summing it over the range i. So, for various discrete values of i I mean we have got impulses.

So, x can take i1 of those discrete values right. In this case obviously, mu x is nothing but as before x px dx from minus infinity to infinity you substitute p of x by this. But as you know, if you do this integral any integral x delta x minus say xi dx minus infinity to infinity what do you get, is x this integral becomes xi. The standard result the impulse has 0 value everywhere, except at x equal to xi where it is actually pointing to infinity.

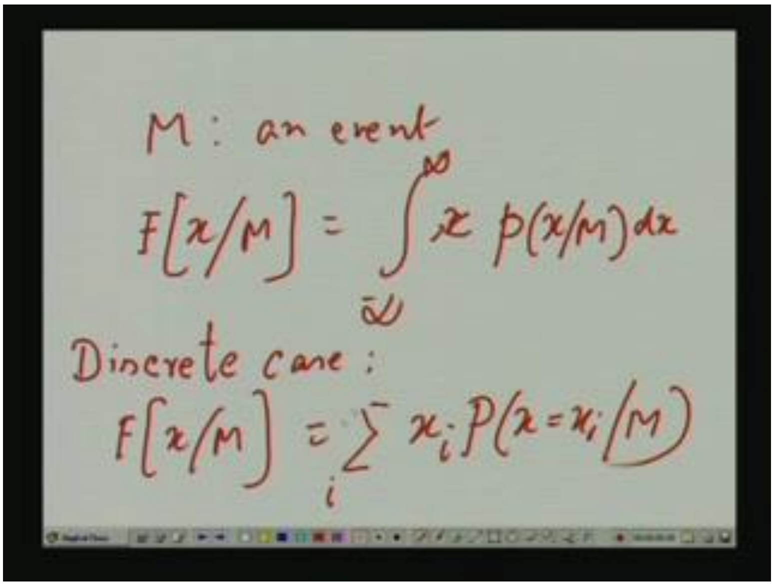
But the integral will be 1 using that concept. Actually, this is a fairly well established result in signal theory. If you can usually, look in to those books on signal assistant this is given clearly there this integral is xi. So, if you really make use of it substitute p of x by this expression here and sum it up you get simply these value pi xi.

(Refer Slide Time: 09:00)



Which means, suppose x takes these values either 1 or 2 dot dot dot up to 6 6 values and each has probability pi equal to say 1 by 6, i equal to 1, 2 up to 6. That is probability of x taking value 11 by 6 probability of x taking value 2 is again, 1 by 6 so on and so forth. What is E of x, then simply pi xi i equal to 1 to 6, but pi is common for all of them 1 by 6 which I can take out xi that is 1 plus 2 plus dot dot dot plus 6 how much is this. This is equal to 3.5 yes. So, this is the mean.

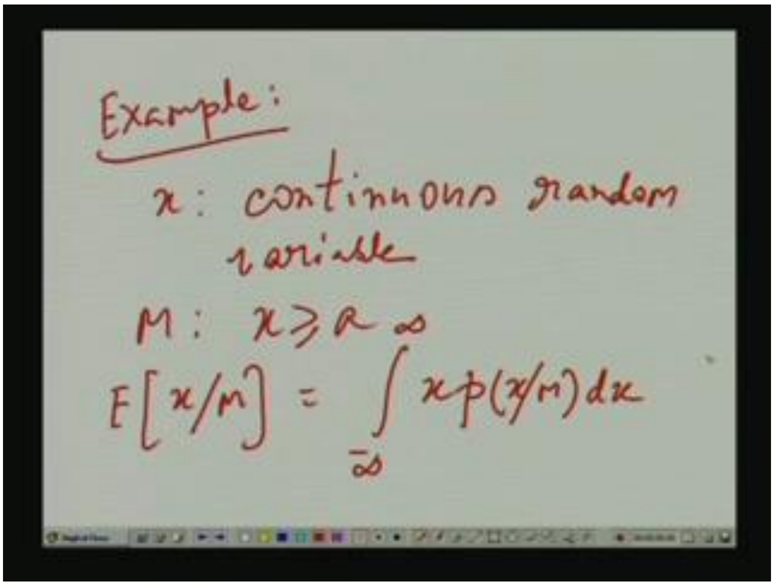
(Refer Slide Time: 10:55)



Then, we can again extend this case further, to the two conditional mean. That is suppose, there is an event there is some event M; M an event. Now, subject to the condition that this event M has occurred what is the mean of the random variable x. So, previously it was unconstrained there was not constraint like this, but now there is constraint. That we are saying there is an event M that has occurred, subject to this constraint what is the expected value of x. Obviously, in such case we should write like this is nothing but x simple this is a conditional probability density. And in the discrete case very simple here x is a continuous random variable.

So, we have an integral and probability density function there x takes only discrete values. So, we have put a summation discrete summation. And xi summation over i and the probability of x taking xi subject to M, these are discrete probabilities. The summation multiplying by I mean this discrete probability multiplies the corresponding discrete value of x. And you sum over all the possibilities of these discrete values. So, sum over i.

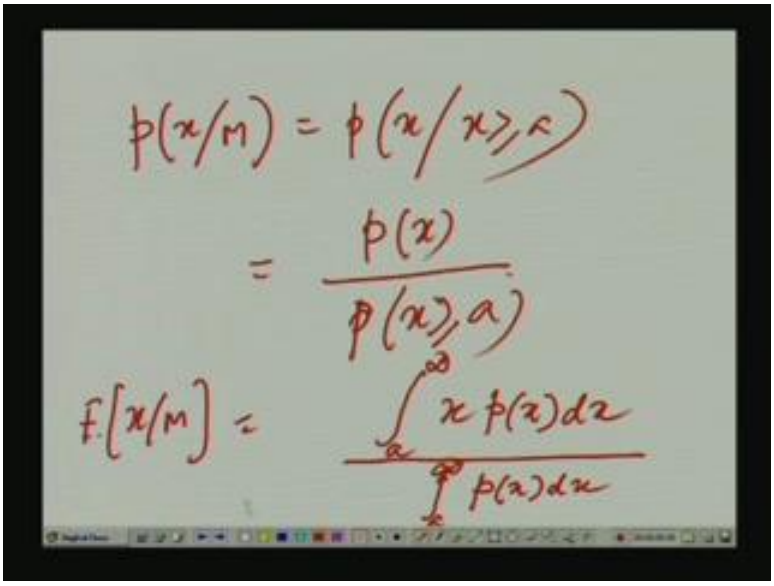
(Refer Slide Time: 13:24)



As an example suppose, x: continuous random variable and given that event M is nothing but x taking values greater than equal to a. So, x is constrained to take values greater than equal to some number a subject to this constraint. What is the mean of x. Physically it means, that we are observing x but in our observation we are not considering all possibility of x.

So, We are only counting those cases, where x takes numerical value greater than equal to a. And out of those observations what will be the mean, expected value of x. So that means, as before this is general formula first I am writing.

(Refer Slide Time: 15:15)

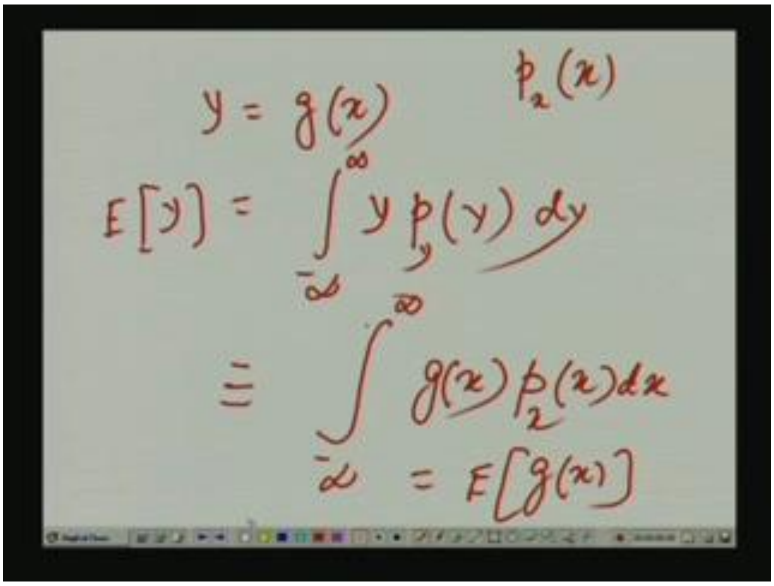


Now, what is p x by M here that is p x by this incident what is this. We know that, this can be written as px divided by this this is the probability density of x, divided by probability density of the total the probability of x taking value greater than equal to a. This is the probability of x taking value greater than equal to a. The event x greater than equal to a has this probability right.

So, this is the probability of x taking value greater than equal to a. We divide the probability density of x by this we get that. So, in this case what is, then it transferred to be x this probability px dx and this is a constant. This is the probability of net probability of x taking value greater than equal to a. So, this will go in the denominator and here x now, takes value not about the entire range, but from a to infinity right.

Then, what is this quantity this is also the total probability of x taking value from a to infinity. So, this will be the conditional mean of x. Now, we come to an interesting topic which is little trickier than this.

(Refer Slide Time: 18:06)

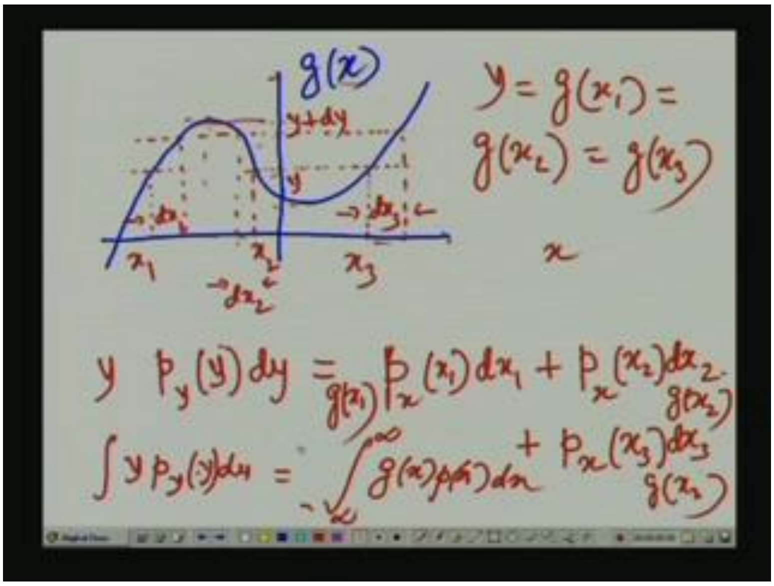


So, that suppose, x is the random variable and a function is given sum not considering the random variable as such. But I am considering another random variable y which is related to x by a function g is a continuous function g of x. Question is, what is the mean of y. Now, by the previous discussion we can always say that this will be given by y p y just put a subscript y here.

So, and to indicate that this is the actually probability density for the random value y it takes the value y. This is the random variable and this is the value. And as long as, we have got 2 random variables: x and y. I will reserve this for the probability density for the random variable of x; x is the variable this is the value. This way we will be able to differentiate between the 2: this is 1 function, this is 1 function there are not same.

Now, coming back to this this will be the mean. But that means, that suggests that in order to calculate this mean. First we have to find out this probability density function which could be cumbersome is it necessary answer is no. We can so that, this is actually given by this will be same as identical to the expected value of gx. That is which you also call the expected value called expected value of g of x; G of x and y are same. So, it is the expected of g of x, this can be shown it is not difficult.

(Refer Slide Time: 20:42)



So, let us take it up just consider a general kind of g of x for the time being say I am having this is my y and this is y plus dy this point could be x1 and this is dx1. This point is x2 and this this distance is dx2 this point is x3 this is x axis of course and this distance is dx3. So, in such a case what is the probability of y taking value between y and y plus dy, that is for what is the probability of y falling in this range.

So, that is same as the probability of x falling in this range, in this range and in this range right. Now, what is the probability of x falling in this range that is since, it is very thin, infinitely small, this is probability density function at these value that is at x equal to x1 multiplied by this. That means, this is the probability of y falling in this range between y to y plus dy this will be equal to first px x1 dx1 plus px this will be the probability of x falling within this range, this will be the probability of x falling within this range like that.

So, you can easily see 1 more thing y is same as g of x1 or g of x2 or g of x3. So, if I multiply this side by y this side also is multiplied by y, but y is this, y is this, y is this right y is g x1 y is g x2 y is g x3. So, these the while multiplying this I multiplied by g of x1, while multiplying this I multiplied by g of x2 and while multiplying this I multiplied by g of x3 they all are same equal to y.

Now, actually here I have to now integrate. What the entire range of y, which means I have to now integrate here. Now, as you see this is very interesting as I integrate as this strip between y and y plus dy as this strip increases becomes wider and wider. What is happening, these also x2 this is moving to this side, this tip is moving to this side and in this case. Suppose, I hold y for the time you hold y here and pull it out pull this y plus dy up.

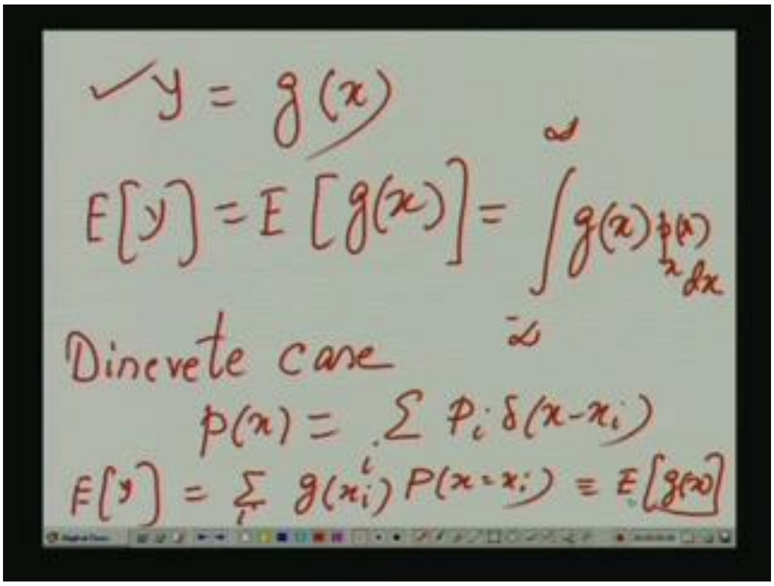
In that case, this point will be going towards this side and this point will be going towards this side and ultimately when y is here that time they will meet here. And as y goes up further, on this side we will not have anything it will come from this side. And now suppose y is pulled down of course, as y goes up as I expand this on this side also this segment widens widens widens. And whole as y goes up further and further, on this side on this left hand side there is no intersection anymore.

So, contribution comes only from this side as I go above. These goes on widening if y plus dy is here either, larger section here if it goes up further, up I will have still further area and like that. And as y is pulled down when, it comes to this minimum point as is pulled down this x3 moves left toward, x2 moves right toward and at this point they are meeting together right.

So, as y goes below that, then contribution comes only from the left hand side. So, actually the integrals are non-overlapping and as I pull y form minus infinity to infinity those 3 integrals they are mutually exclusive, but they cover the entire x axis. In that case, integral will be same as if I do this integral this will be nothing but gx px dx about the entire range .

Because, this will give rise to 1 integral, this 1 integral, this will integral there are domains are mutually exclusive as you are saying. And as y is pulled from minus infinity to infinity we get 3 distinct domains: 1 from this, 1 from this side, but they are non-overlapping, but they cover the entire x axis. So, all the all of them put under 1 integral and we can write like this. I am writing the same formula again, in the next page.

(Refer Slide Time: 29:15)

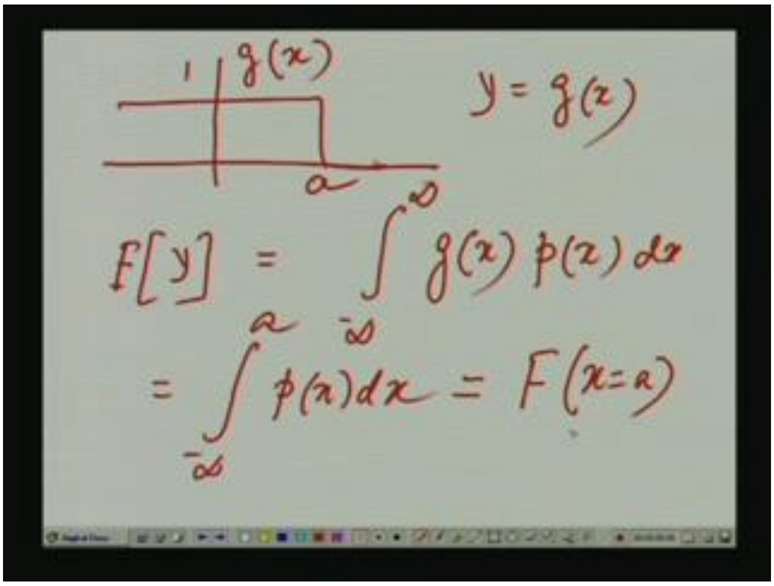


So, if I in conclusion if y is given as g of x, then E y which also can be written as E gx is equal to this is very useful and interesting result it is widely used dx. This was for the continuous case. So, if x is discrete, you can easily extend this to the discrete case that is suppose. So, I will consider the discrete case now. So, this function is given again. But x unlike here x is taking values discrete values.

So, either, x1 or may be x2 or may be x3 or like that each has its own probability. So, we have got in this case px is nothing but pi delta x minus xi standard case. Say x can be x1 or x2 or x3 whatever, they are the discrete values it takes. In this case, if I define random variable y as a function of x I mean as a function gx. Then, what is the probability there what is the mean of y.

Now, you can easily, get that from this concept in this case Ey will be what here we had a continuous summation. Now, it will a discrete summation over i. And, then this function and x will take values xi I could be 1, I could be 2, I could be 3 and each will be multiplied by the corresponding probability x taking xi that will be the mean of y or equivalently you can write E of gx. Let us, consider an example.

(Refer Slide Time: 32:02)

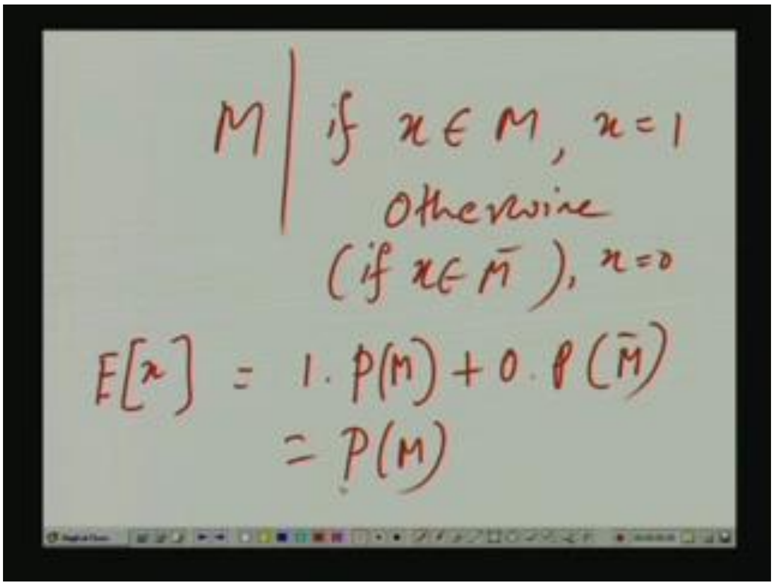


Suppose, g of x is given like this its value is 1 up to some number say a to the left of a it is value is 1, to the right of a its value is 0. So, x is a random variable and I define y as gx. So, y is another random variable in fact, it takes only 2 values either 1 or 0 and the way it takes this values is given by this function. So, what is the mean of y here, same as I mean I will use the same formula gx multiplied by px I am dropping the subscript x here now.

Because, again i have got only 1 random variable. So, I can as well drop it without any confusion this, but gx is 1 up to this and then 0. So, the integral simply becomes this minus infinity to a and there the value is 1 px dx. And all of you can remember this is nothing but the probability distribution; probability distribution means, the total probability of x taking values less than equal to a.

So, that is x taking values from minus infinity up to a what is the total probability, that is called the probability distribution function. So, this is an interesting thing it shows that any probability distribution function also, can be shown as a expected value. This is a probability distribution function, but it can always be shown as a expected value by constructing a function like this. 1 more example you can take another example.

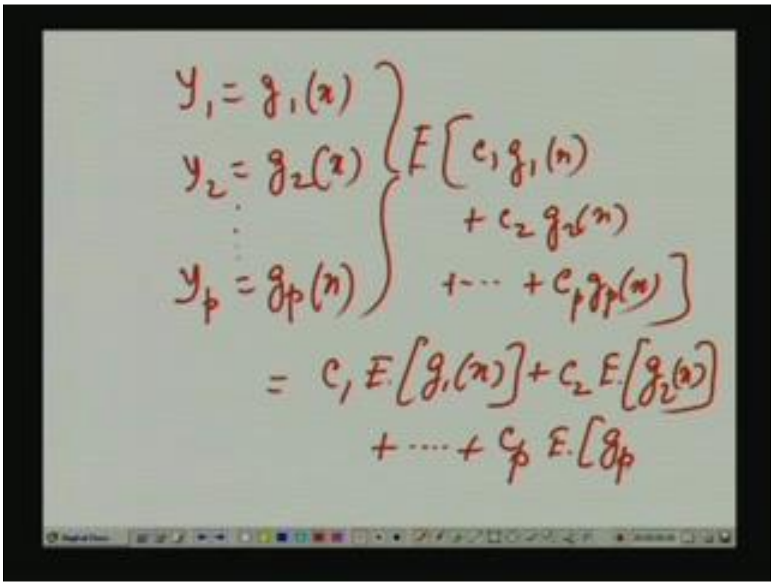
(Refer Slide Time: 34:40)



Suppose, there is an event M. So, if x belongs to M that is if so happens that you observe x. And it is member of M, then x takes the value. if x this, then x takes the value 1 otherwise, that is if x is a member of M bar, then x is 0. In this case, what is the expected value it is a discrete case x takes value either 1 or 0. So, mean value will be one into the corresponding probability p of M and 0 into the probability p of M bar which is p of M.

Then, this shows that the probability also can be shown as a as an expected value by a construction like this. Now, you consider some properties of this expected value. We can see 1 thing that suppose, we have got several functions.

(Refer Slide Time: 36:22)

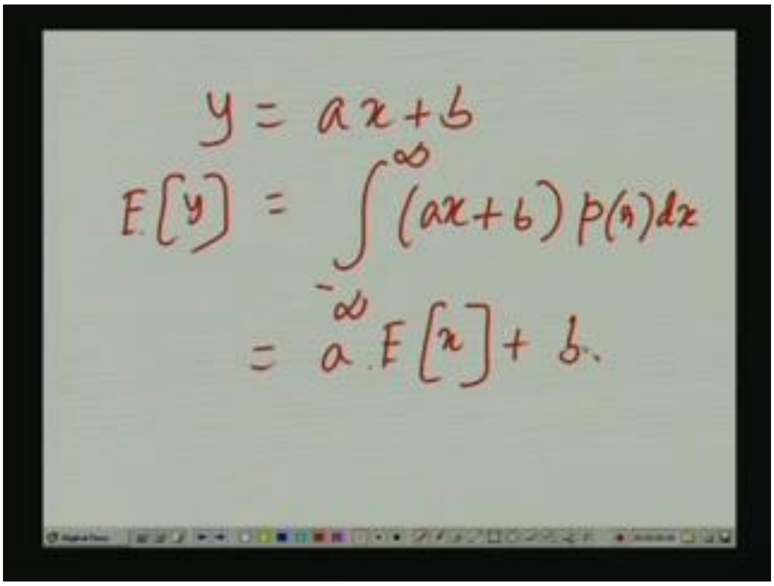


So, that is y1 as g1 x g1 is function y2 as g2 x dot dot dot yp as gp x. Instead of having just 1 function suppose if I construct a linear combination c1 g1 x plus c2 g2 x plus dot dot dot cp gp x what is the expected values of these. Obviously, the whole thing will be multiplied by the probability density p of x integrate multiplied by dx, then integrate it, but integral is linear.

So, the entire integral can be broken down in to separate integrals 1 involving this, another involving this, dot dot dot another involving this. Integral involving this will give rise to the expected values of g one x or y1 this will give rise to the expected value of g2 x or y2 so on and so forth. So, expected value of such a thing will be nothing but , so this is a linearity property.

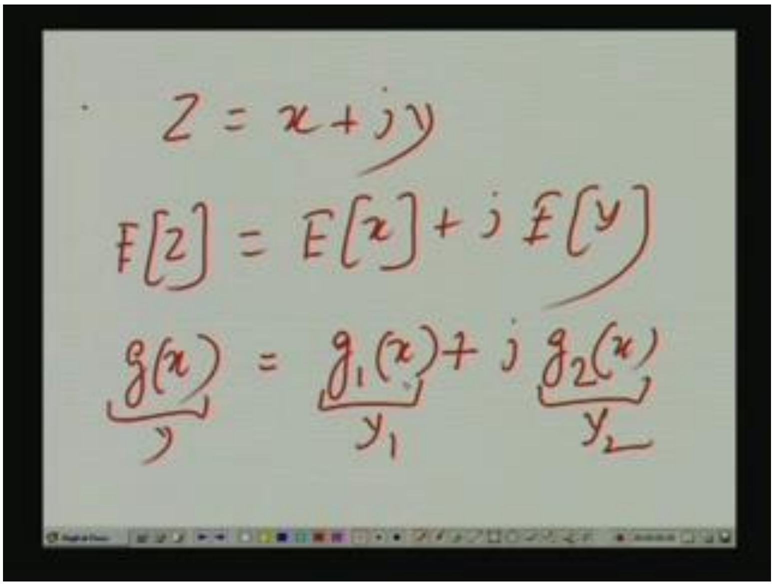
So, the expectation operator actually is a linear operator if it works on a linear combination like this. Then, it can be applied to each of them separately and the various results can be just combined like this. So, expectation operator is a linear operator this is very useful result. It will be used many times in future another thing you see.

(Refer Slide Time: 38:45)



If y is given as say ax plus b what is E of y. This will be nothing but first term a will go out you will get a times x px dx which is nothing but expected value of x. And the second term b will go out integral px dx which is equal to 1. So, you will get b so for ax plus b expected value, this expected value of ax plus b is nothing but a times expected value of x plus b. This also, is a very useful result though it may appear to be simple 1. This are very useful results. Now, I consider a complex valued random variable.

(Refer Slide Time: 40:03)

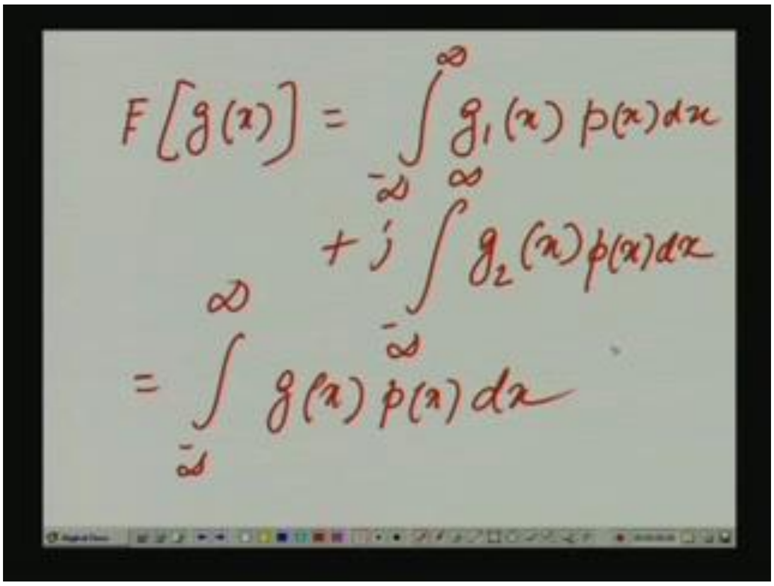


Suppose, Z is a complex valued random variable it has got 2 components: 1 real part, another imaginary part say x plus jy, then we define the mean of z as mean of x or expected value of x plus j times expected value of y. That is both x and y there are real values x has it is own probability density function, y has it is own probability density function to calculate this multiply x by px dx integrate.

Then, to calculate this by this multiply y by py dy and integrate likewise. This is the definition. Now, suppose we have got a complex valued function gx which has got 2 components: the real part is g1 x, imaginary part is j times say this is the imaginary part g2 x. So, total value is g1 x plus g2 x. You can if you it y, this you can call as y1, this you can call as y2.

So, by this definition what is the expected value of y that is nothing but E of y1 plus j in to E of y2. But what is E of y1 we have already seen that E of y1 is nothing but g1 x px dx integral. And what is E of y2 that is nothing but as you have seen a little while back g2 x multiplied by px dx integral.

(Refer Slide Time: 42:05)

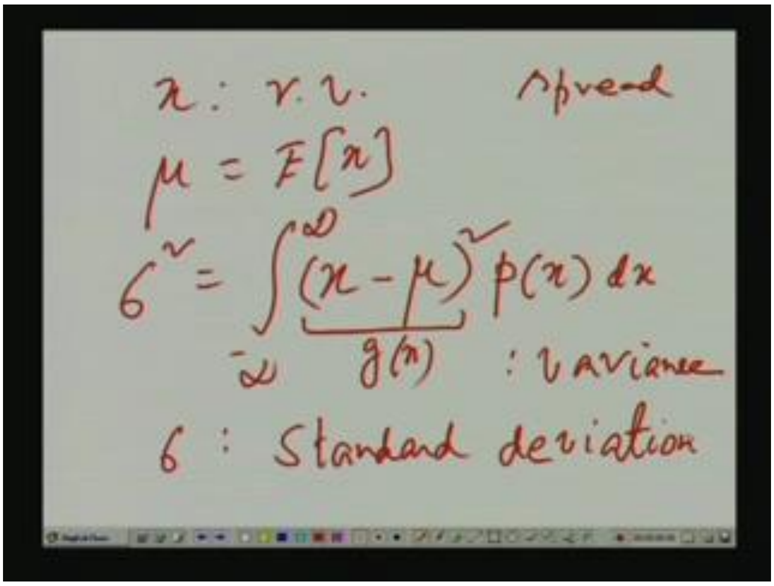


Which means, E of gx plus j and now you can combine the 2 integral. So, sincerely g1 x plus j g2 x together again you will get back gx same as before. So, whether g of x is a real valued function or a complex value function expected value of gx is given by this look here the expression is same. This takes us to a very important concept that is variance.

Suppose, x is a random variable and it has mean mu. So, when you to try measure when you to observe x it is not that always you will get mu. You will get sometimes values above mu, sometimes below mu. So, essentially a question arises how much is the spread average spread of the variable x around the mean or how much approximately is the range of variation, range of fluctuation around the mean.

Does this random variable vary within the small range around mean, then it is spread is less does it vary widely around the mean. Then, it is spread is more. How to measure this how to have a quantitative description of the spread that is given by variance.

(Refer Slide Time: 44:09)



So, x random variable r v mu here is E of x, then you find out this quantity first. Definitely, this gives you the spread this gives you the deviation around the mean. So, in our first attempt we might try to take the expected value of this. Because, x can take various values depending on each value you get some value for this deviation sometimes high, sometimes less.

So, you may try to get an expected value of this by multiplying by this by p of x dx and integrating that would give you in your opinion. I mean that, could give you an idea of the average value of the deviation, but the problem there is even though there is deviation this can be positive and negative. And your average value could turn out to be 0 or very small. And that will try, that will imply as if the spread is indeed very small, but that it is not the case actually.

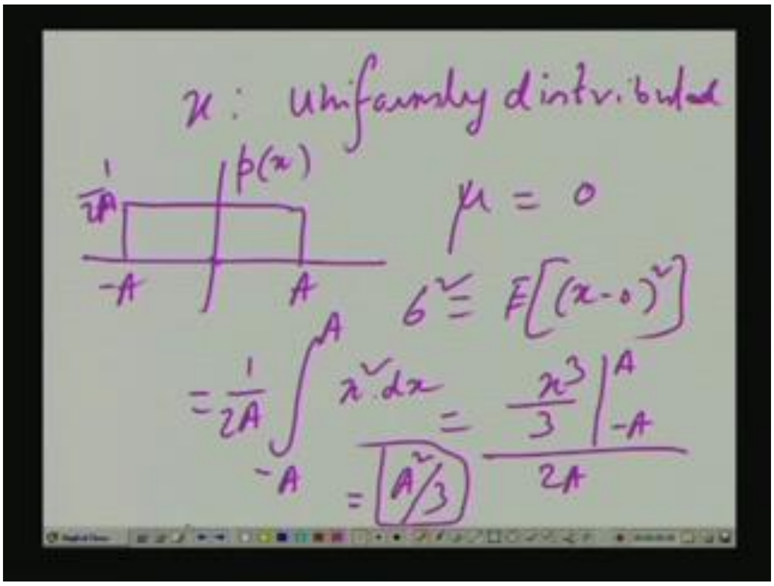
Because, spread still could be very large, but x can take values uniformly to the right of mu to the left of mu. So, x minus mu sometimes taking positive values, you get high positive values x minus mu taking sometimes negative values. In fact, highly negative values, but if you just keep x minus mu and average out that average could still be 0 or very close to 0.

So, an error an erroneous implication will come up I mean we will tend we will be you know, we will tend to think as if the standard rule of variable x is fluctuating in a very small range around the mean, but that is not really the case. Here the problem is coming, because the positive deviation and negative deviations they may cancel out, and giving an indication as if the expected value of the deviation is 0 or very small.

So, to avoid that we take the square sometimes it is called power also, because the point I mean if you use electrical engineering analogy mean means, DC x minus mu means the AC square of that so power. And this quantity whether, this deviation is positive or negative this quantity will always be positive right. So, if we take the expected value of that will give us a correct picture of the extent of the fluctuation.

The range of fluctuation or the spread average spread around the mean. So, we take this multiply by. So, this is a gx here now function gx what is the expected value of this, this multiplied by px dx. This we call sigma square; the sigma square all together is the symbol and sigma, then is the square root of this which is called actually the standard deviation. So, sigma square is the variance.

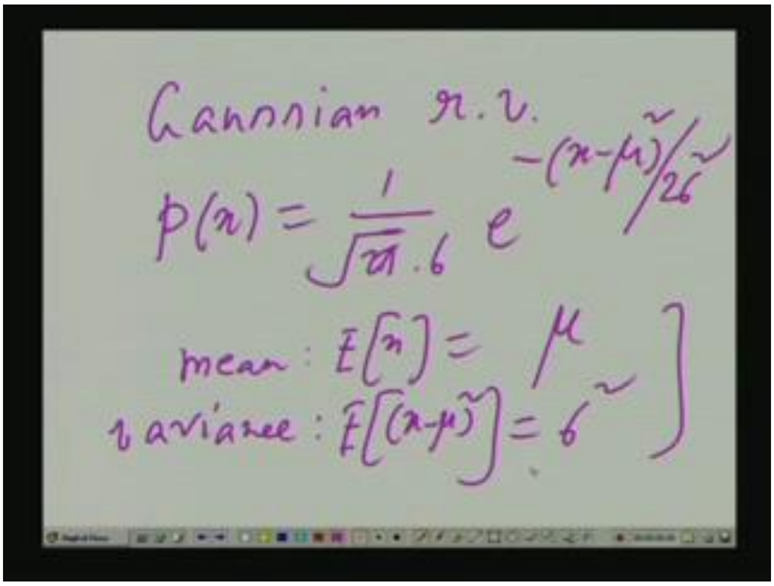
(Refer Slide Time: 48:03)



Let us take an example x: uniformly distributed to make life simple I assume that, the distribution is given like this A minus A which means this is 1 by 2A. The height is 1 by 2A, then only the area is 1. What is the variance, first what is mu I mean I do not have to recalculate you can easily understand that, mean will be mu will be 0. And what is, then sigma square will be E of x minus mean which is 0 whole square that is E of x square.

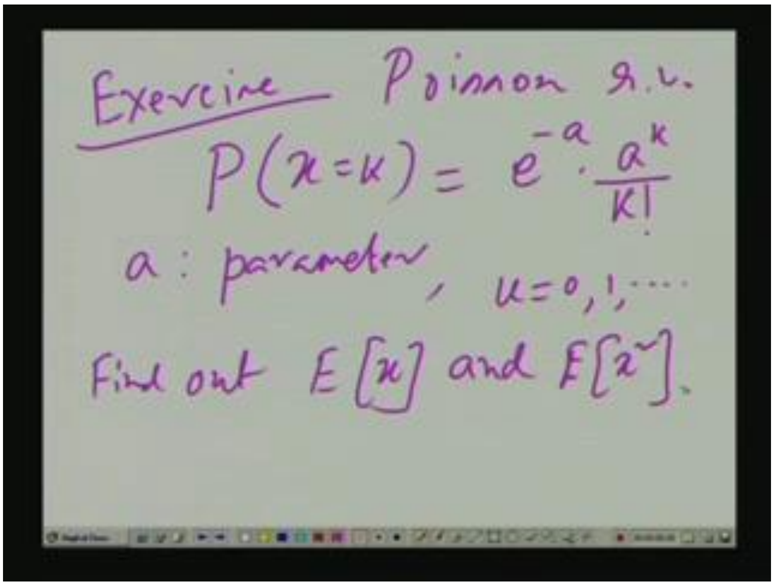
Then, we know the probability density which is a constant 1 by 2A. So, you have x square in to 1 by 2A that comes out minus infinity I mean this will be from minus A to A dx and this integral is x cube by 3. So, x cube by 3 A minus A divided by twice A. What is this, 2A cube by 3 twice A cube by 3 divided by 2A. So, essentially this turns out to be A square by 3 this is the result.

(Refer Slide Time: 50:28)



Another very important random class of random variables are the Gaussian random variable r v for random variable here p of x is given by this formula sigma. For this function mean, that is E of x will be given by this parameter mu and variance that is E of x minus mu whole square is actually, given by sigma square. I just give you an exercise and then I call off today.

(Refer Slide Time: 51:34)



Suppose I consider. So, this is an exercise I consider a Poisson random variable Poisson distributed random variable where, the probability of x taking a value k is given by let me, see e to the power minus a a to the power k divided by factorial k; a is a parameter and k can be 0, 1 dot dot dot like that. For this find out, so that is all for today in the next class we will take it up.

Thank you very much.

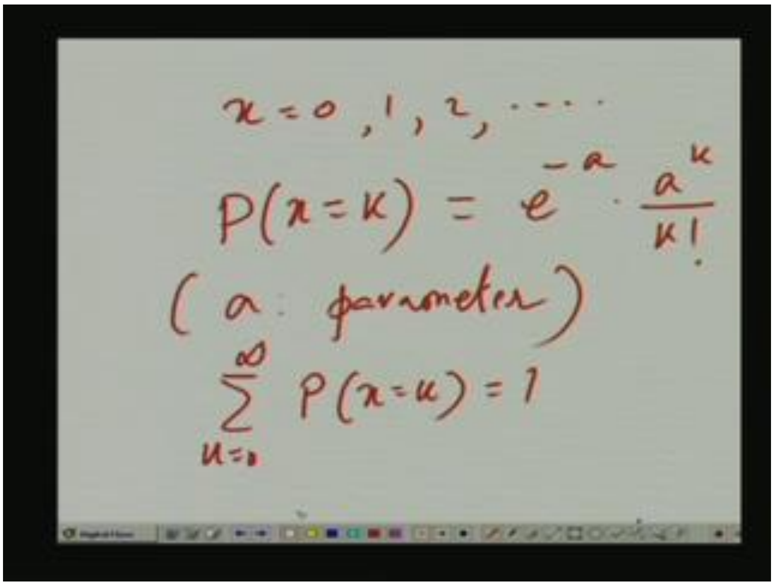
Preview of Next Lecture.

Probability and Random Variables
Prof. Dr. M. Chakraborty
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture - 10
Moments

In the last class, we gave a problem. I mean we said that x is a random variable which is Poisson distributed and it takes values it is a discrete random variable. It takes values like you know x equal to either 0 or 1 or 2 dot dot dot dot dot.

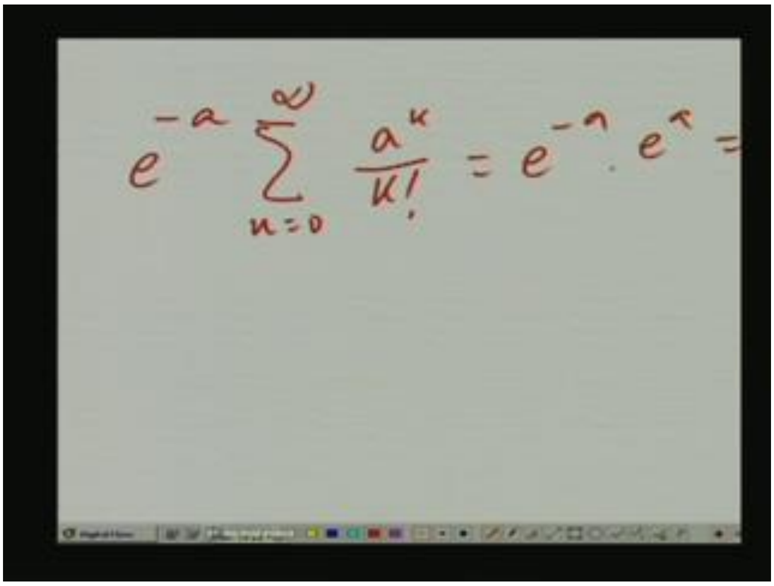
(Refer Slide Time: 53:37)



And the corresponding probabilities is given by Like, p if p equal to I mean for x equal to k the corresponding probability will be given by e to power minus a and a to the k divided by factorial k where a is a parameter. We ask you to find out the mean and variants. Or rather, we ask you to find out e of x expected value of x and expected value of x square.

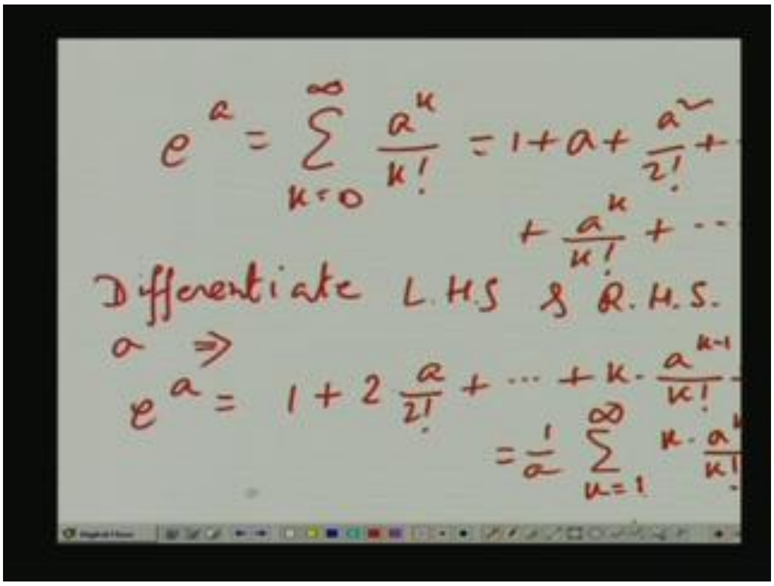
If know expected value of x square and if you know the mean obviously you can find the variants that was a question. A is called a parameter. So, we say that x is Poisson distributed random variable with parameter a. can you tell us, what is the function of this factor e to the power minus a. Actually, if you just consider this part a to the power k by factorial k, this will case for various case if you take their values and add them this gives rise to an exponential series. In fact, we should have this thing this must be equal to 1 right, because total probability is 1. Now, because of this factor this equality will be satisfied this you can see easily.

(Refer Slide Time: 55:16)



If you put that expression take this out x equal to 0 to infinity and we have got a to the power k divided by factorial k. Now this is nothing but e to power a there is a exponential series. E to the power a when expanded gives you this. So, you e to the power minus a into e the power a is equal to 1. So, just to make this equal to 1 this factor has been added that is introduced. Anyway, this is just a side issue we have to now find out the mean of x that is e of x and e of x square.

(Refer Slide Time: 56:06)



To do that, we consider this exponential series e to the power a, which I have just a while back stated is equal to a to the power k divided by factorial k k is 0 to infinity. Which is equal to 1 plus a a square by factorial 2 plus dot dot dot dot plus a to the power k by factorial k plus dot dot dot dot right. I now, differentiate these left hand side and also right hand side with respect to a. Once and then again one more time that is twice.

So, first if i differentiate LHS a very next thing from here is a very important thing called characteristic functions. It is basically, related to Fourier transform; Fourier transform of a probability density function is actually a characteristic function. Now, that I will not start now, but i am just telling little about it. Since, Fourier transforms have been very effectively used for carrying out convolution and all that.

So, I mean you can use this result in the end where suppose, you have got a summation of various random variables each having its own probability density. And you have to find out the probability density of I mean resulting sum which is also a random variable there this characteristic function I mean is very useful. Also, this characteristic function can be used very effectively to calculate moments.

The moments which we have just considered, they can be calculated for this characteristic functions very easily. So, I reserve this for the next class and I stop here today. So, we have so for covered in this chapter function of random variable and if y is gx and giving the probability density of x. What is the probability density of y, we took several examples. Then, we went in to moments we went in to another important thing that is given a function gx. What is expected value of gx that is nothing but gx in to px dx integral. We considered several examples and applications of that, that took us to this motion of moments not only mean and variance to start with, but more general in case of moments and from moments we go to characteristic function. That will complete this chapter, then we will go in to what is called joint statistics that is not only 1 random variable.

But more than 1 random variable say two random variables are present and they are related. So, we will be considering things like joint probability density and all that and do similar exercise. So, that we, so today we stop here in the next class as I said we will consider this characteristic functions.

Thank you.