Estimation of Signals and Systems

Prof. S. Mukhopadhyay
Electrical Engineering Department
Indian Institute of Technology, Kharagpur

Lecture - 05
Mean and Variance

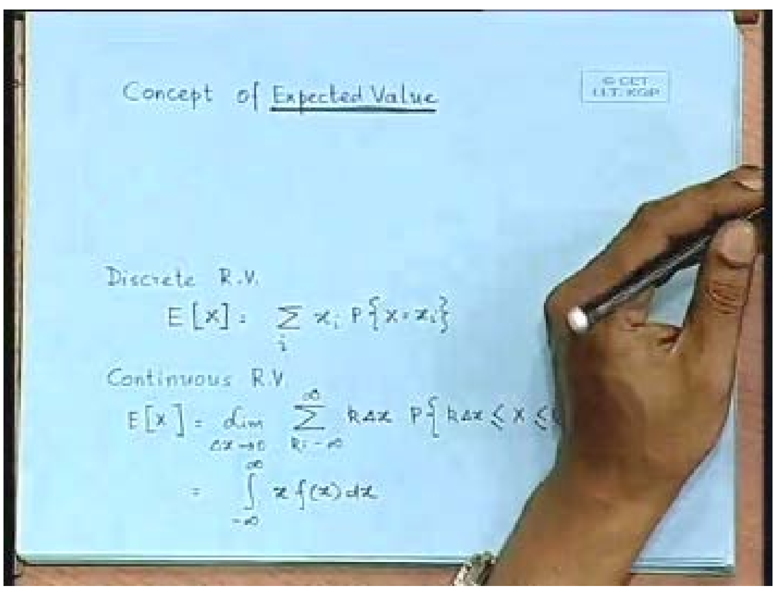
Good afternoon; so far we have seen various kinds of distributions, distribution functions of jointly distributed random variables, we have also seen what are conditional distributions. But sometimes you know random variables being random variables that they can take any value over a certain range and according to the distribution functions. It becomes very useful to have you know certain single numbers which will indicate that... which will indicate what is known as number 1 what is known as that this while they are random; what is the kind of an kind of an average property of these random variables, on an average it takes what value; or that is that is sometimes called measure of central tendency that is if you take all numbers then on an average what kind of values that numbers takes number 1 number 1.

Number 2 is the... that is not enough. because for example; pictorially it becomes interesting to see that suppose you have a random variable so you know it can take suppose it... these are you know the number of samples, so so the random variables take values say say in this region. So, if you take their average then you can say that it is let us say around the point a b, this is b this is a so this... all these dots are random but you can say that they are on an average spread out around the point a and b. this is this is this is useful to know.

Secondly, it is it is also useful to know that suppose you have these black dots which are spread out around the point a b and you also have these red dots which are spread out still around a and b but much more widespread around the point a and b, so it is not just enough to know whether they are spread out from the point a and b but they are also... it is also important to know that is in general how much they are spread out from the point a and b. So this is sometimes called... so we need to know this that this point a and b and you know what we what we say we need to know what is known as a measure of dispersion, so how much they are dispersed around the point a and b. This this you have seen in in our school statistics also. So so in general we need to... these two quantities become very important numbers to characterize let us say us single random variable and we are going to talk about these two things today they are called the mean and the variance. So lecture 5 is about a mean and the variance.

So firstly the mean is also sometimes called an expected value. Now what do we mean by this expected value?

(Refer Slide Time: 00:04:30 min)



I mean the random variable is random. If you make an experiment today you will get one value, if you make an experiment tomorrow you will get another value so what do you mean by an expected value? So what we mean by an by an expected value is actually in the sense of... in the sense of average; what sort of average? Suppose we we do... suppose we do a large number of experiments, suppose we do N experiments okay let us consider... let us considered a random variable x... a random variable x which can take values either a or b or c.

So suppose we do N experiments that is we pick x n times and of them N a times we get a, N b times we get b and N c times we get c; obviously N a plus N b plus N c is equal to N. Now if you say that what do we mean by an expected value of x? What we mean is that if we just added of these values as we picked up over these N experiments and then simply made an average what would be the value? So obviously the value would be a into N a plus b into N b plus c into N c by N: this is our expected value. So what we get is a into N a by N and plus b into N b by N and plus c into N c by N.

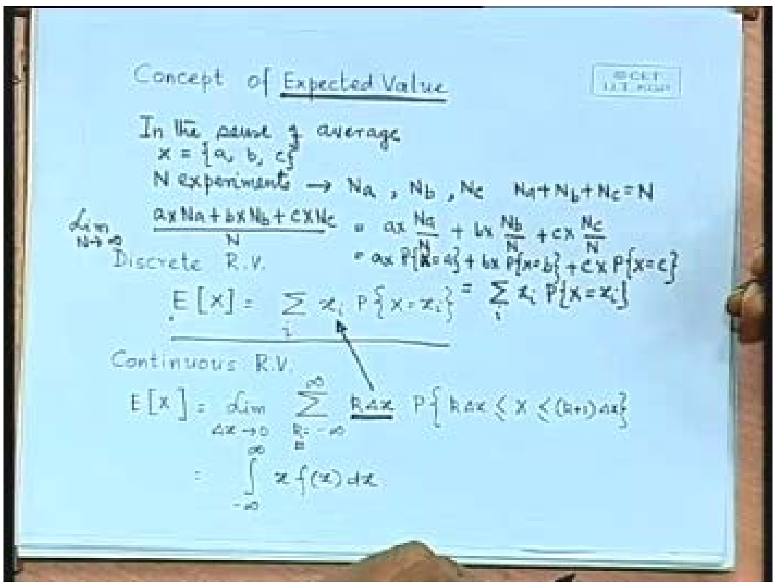
Now if we take this... if we take the limit that N tends to infinity then what do we get; we get a into probability of getting an a that probability that x is equal to a plus b into probability that x equal to b plus c into probability that x equal to c; so so so so the average is sigma; you can say that over all i x i... now we are saying that x takes values over an index at i; for example: x 1 is a, x 2 is b, and x 3 is c so x and then probability multiplied by probability that x is equal to x i, if you sum these then we get what is known as an expected value. So for a so for a discreet random variable which is which is which is easier to understand.

What we mean by an expected value of x is that it is simply x i it is the probability of x equal to x i. In other words, we every time we pick up the random variable, we we we perform the experiment, we actually note it down suppose in a copy and and at the end of a large number of experiments we simply sum whatever we have copied in our book and then divide by the number of experiment that is my expected value correct.

Now what happens when we have a continuous random variable?
Obviously in a continuous random variable the number of values that x can take namely the excise is is infinite. So the so the so the indexed set is actually infinite so so so we have to... actually this summation now we have to perform over very very small intervals and then this summation will finally become an integral okay. So so we write we write

this exactly like the random variable, we we want to write this summation for also a continuous as variable and we write that if we take very small intervals so limit of delta x tending to 0 and we... so obviously if if we take very small intervals then the number of intervals will become infinity. So when we are calculating the probability over the real line naturally we are going to have an infinity number of such such intervals so K indicates that that interval index... so we have K is equal to minus infinity to plus infinity, this K delta x (Refer Slide Time: 9:26) now becomes the value, now becomes is equal to x i.

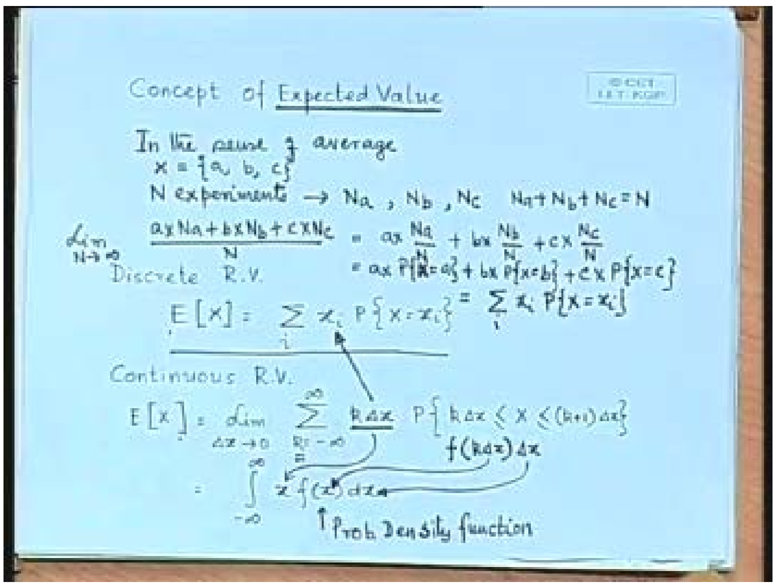
(Refer Slide Time: 9:32)



So in so in the Kth interval what is the value of x? It is K delta x. So what is the what is the probability... so when we take this... so this takes the value of x i and now we want to calculate the probability that it that it falls between K delta x and K plus 1 delta x; what is this probability; this probability is nothing but f of K delta x where f is the probability density function into delta x. So now this K delta x f of (K delta x) delta x.

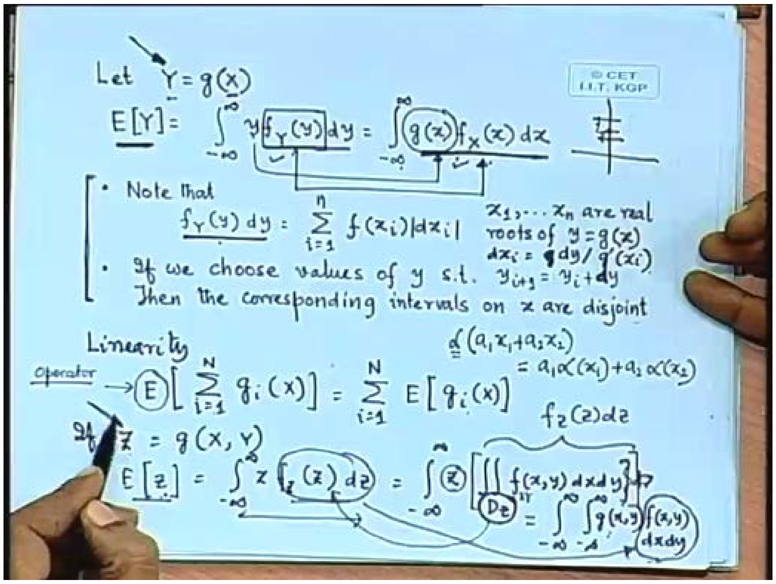
This sum if you do from K is equal to minus infinity to plus infinity and if you now taking taking delta x in the limit if you replace K delta x by the value x this K delta x by the value x and this delta x (Refer Slide Time: 10:28) by the differential dx then we will get the... and then as we take delta extends to 0 this sum will become the limit so we will get the the definition of the expected value of a continuous random variable which is nothing but minus infinity to plus infinity; x, f(x) dx where f(x) is the probability density function okay. So this is the... this is a very important quantity and often of great interest in in various kinds of analysis so this is called the expected value of a random variable capital X and is denoted as E of capital X.

(Refer Slide Time: 11:23)



Now let us see... so so this is the expected value of a random variable. Now, suppose we have a function of a random variable or in other words, suppose we have a random variable Y which is the function of a of a random variable x... we have we have already seen that if we have a random variable x and if we have Y as another random variable which is constructed by taking a function of x then we have already seen how to calculate the distribution functions, density functions of Y.

(Refer Slide Time: 00:11:31 min)



So now again if you want to find out the expectation of Y then it will be nothing but minus infinity plus infinity y, now this is the density function of y dy. This is the old old expected value definition applied to Y. Now we want to write it in terms of x. Because actually in many cases, actually y is a derived quantity, the the original random variable is x, so in all probability we will know the density or the distribution function that is whatever probabilistic description we will have we will in general have it for x, right. So so we have to now express the expected value of y in terms of the probabilistic description of x and the function g right.

It turns out that... this is nothing that minus infinity plus infinity g(x), this y becomes g(x); now we want to express this integral in terms of x because x is the original random variable and what is fy (y) dy that is what is the probability that y y becomes I mean y stays between y and y plus dy that is... since we are considering all possible values of y so obviously we are we are also covering all possible values of x. So all you need to do is that you you take an interval of x and you just find out the corresponding value of y so in this way rather than varying y you vary x, for each value of x you consider g(x) and then probably....... so it is the same thing right. Only thing is that... that is the a same thing.

Note that we have not we here (Refer Slide Time: 13:50) we have not exactly replaced this integral by this one right because if you want it to be fy by (y) dy then as we have seen before that you have to actually find out that what are the values of x that gives you the same value of y g(x) and then you have to take intervals around those x's; that is if if if if y is equal to g(x) has a root x i then you have to consider the intervals around those x i and you have to take intervals of of of length dy by g dash x i and all that, but then finally it will turn out that if you are taking a variation of all possible y's then then that is equivalent to taking the variation of all possible x's right. So these are the... you know rather somewhat mathematical arguments which which actually show that this integral and this integral are equivalent.

Now it turns out... by having noted that it turns out that the... now this is you can sometimes say that this is an expectation operator. So you take a random variable x and if you operate it by the expectation operator you get a constant which is called expectation of phase right so it is a value like a function.

Now it turns now the first thing we are we are trying to note is that this operator is actually linear. What do you mean by linear? What we mean by linear is that if we take expectation of sum of quantities then that is equal to this sum of the expectation of these quantities. So the expectation operator is actually so expectation of sums is equal to sum of expectation right. So in the sense of... in that sense it is linear. I mean it is just like you know for any for any linear operator L (a 1 x 1 plus a 2 x 2) can be written as... this is the basic superposition theorem idea is a 1 into L of (x 1) plus a 2 into L of (x 2) so so this is a superposition theorem which any linear operator satisfies; in a similar sense this operator is called linear.

Now now let us consider a more complex function. Suppose z is a function of both x and y that is x, y and z are actually... x and y are as collection of jointly distributed variables and z is a function of x and y right. So so now what is what is the expectation of z? So previously we had a function in one dimension y is equal to g(x), now we are considering a function of two dimensions. So z is equal to g of (x and y) both, we will do exactly the same in the sense that previously we had taken the integral over the x axis, now we have to take an integral over a region because now we have a two dimensional space over which z takes values. So so the basic defining integral is the same: minus infinity to infinity z fz (z) dz that is the basic definition of an of an expectation operator. Now now if we take this fz (z) dz then that is that is that is what is the probability that z is between z and z plus dz.

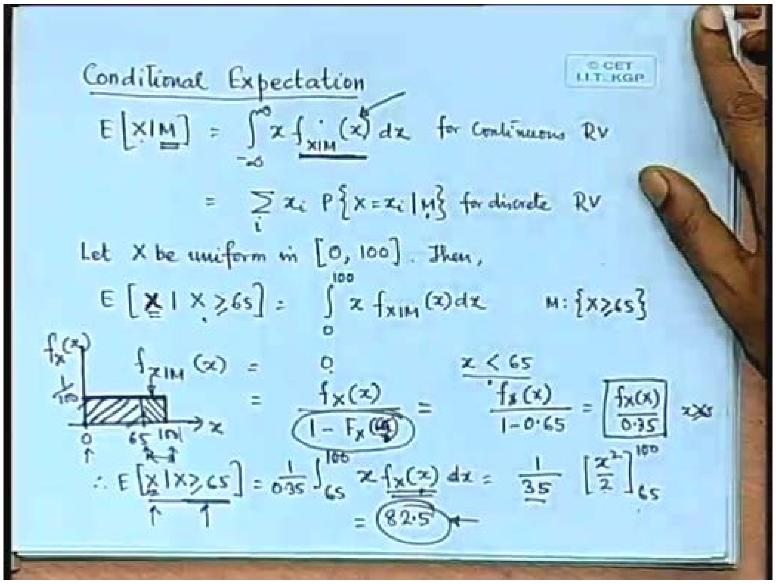
So now we have to do... we have to exactly compute the region of x and y for which z takes the value of z plus dz between z and z plus dz and now so we have to we have to find out that that what is the probability that under the given random variable descriptions that is the under the joint probability distribution function of x and y what is the probability that x and y take values in that region dz for which z takes values between z and z plus dz.

So now we have to basically do an area integral. So exactly that is what we are doing. So so we have we have a double integral because there are two axis and we have a we have a joint density function f xy of x and y dx dy and we are computing it over the region dz for which z takes the value between z and z plus dz and it turns out that that is nothing but... again by by exactly similar arguments that it is nothing but minus infinity to infinity minus infinity g(x, y) f x y dx dy right.

So so again since z we are varying between all possible values minus infinity plus infinity so that is equivalent to sweeping x and y over the whole region minus infinity plus infinity and then for each x and y pair that is for each each each x and y square simply pick up the value of z which it takes while it takes in that square and then sum the integral. It is it is one and the same thing. So this is how you evaluate the expectation operator which is I mean the expectation of a quantity z which is a function of two random variables which are jointly distributed.

Now let us come to what is known as conditional expectation.

(Refer Slide Time: 00:19:23 min)



You know this is that I mean obviously the expectation of a random variable becomes totally different if some facts are given. For example, I mean one of the simplest examples is that what is the expectation that a family has both sons that a that a that a couple... among all the couples having two children what is the probability that a couple has two sons okay, it is is obviously one fourth assuming that the probability of that that I mean the the the the the probability of having a daughter or a son is actually high. Okay so so the probability of having both sons is actually one fourth. But what is the probability of having of a couple having two sons already given that it has one son. Now it is half.

So you see that the probability of a couple having two sons changes whether you give the information that it it it already has a son or you do not give it right. Why does it change; because now you will be looking for looking for couples having two sons, you will obviously... since you have already given that that that the couple already have one son so you will be looking among couples who already have one son rather than all the possible couples right.

So so you see that this is this is called conditional expectation. So the expectation... this is the expectation of a of a of a random variable when certain facts have already given certain facts are known they are they are not random right, so if certain facts are already given what is the expectation obviously that is going to be different, so so now we are going to talk about that. So this is exactly in the similar fashion that if what is the expectation of x given an event M given some relationship some information about x through the event M. So obviously it is going to be the conditional density function now. Previously it was the ordinary density function. But now it is going to be x, conditional density function of x because this actually indicates the likelihood of x occurring given that given the information M right.

So, for a for for a [22:00.....] discrete random variable it is going to be x i probability they are x is equal to x i given M. So always we are going to consider that fact M which is the condition given right. Let us take an example that suppose x is uniform, suppose we are talking about the the persons we have we have we are just assuming that that people must be aged between 0 and 100 and we are actually ignoring the people who are more than 100 years old actually the number is very small.

So if we assume that the number of... that the probability distribution function of persons is is uniform that is there are between there are equal number of persons between the age 0 and 1e and 1 and 2 and 3 and so on right, so then if somebody ask that what is the probability that a person’s age is greater than 65, what is the probability? So you have... what what is the probability density function? The the probability density function is like this that is 1 by 100, the interval of age is 0 to 100, total probability must be 1 and this has to be uniform so therefore this is 1 by 100 so this is the probability density function.

So what is the probability that that a that if you pick a random person from the street what is the probability that his age is going to be greater than 65? So obviously it is going to be this one (Refer Slide Time: 23:34). So it is going to be 0.35 correct 100 minus 65 into 1 by 100.

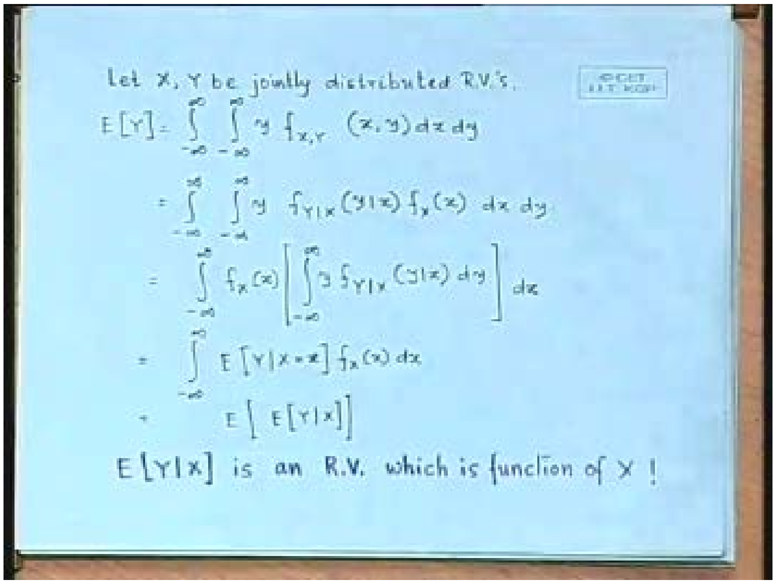
Now what is the conditional density function; what is the conditional density function given the fact that let us say x is greater than 65? That is okay.

So in that case obviously since it is given that x is greater than 65 so therefore x so f of (x) when when x is less than 65 is 0 because when it is given that the x is greater than 65 therefore there cannot be there cannot be any... we cannot choose any person whose age is less than 65 so will get 0 here. So all the people will get are now from this region. So obviously so obviously now the the probability density function is going to be this divided by 0.35 because now over this region only this this conditional density function is going to be 1. So the so the conditional density function is going to be fx x by 0.35 right.

So now if if somebody asks me that if you if you pick if you pick a a large number of persons from persons who are aged greater than 65 what is the average age that you are going to find that you are going to find, so that is going to be the expected value of x when x is greater than 65 already given. So it is a conditional expectation of x given x greater than 65 and then if we just put this this this conditional density function then what we will get is nothing but 82.5 which is nothing but the midpoint between 65 and 100.

So, since the probability... this is also obvious, I mean since this since the probability density function is uniform so obviously the the the mean is at the middle. So if you if you are considering people between 65 and 100 it is obvious that the that the mean is going to be at the middle of 65 and 100 which is nothing but 82.5 right. So this is called conditional expectation.

(Refer Slide Time: 00:25:43 min)

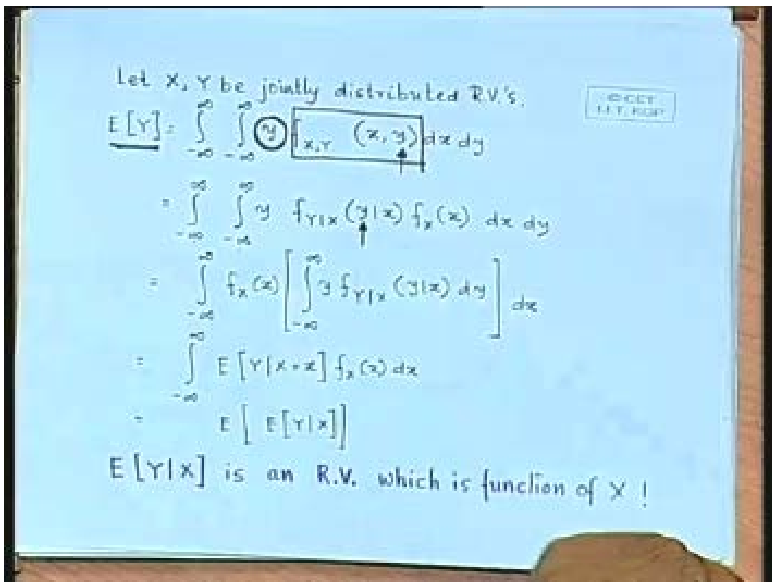


Now we want to show that a particular I mean an interesting... what is the conditional distribution that is we we want to compute conditional distribution mathematically right. So previously we had we had talked of... we have just now found conditional distribution of a of a single random variable. Now we have now we are going to find out if the conditional distribution of a random variable y which is jointly distributed with x through the distribution function f xy okay. So obviously what what is going to be it what is what is expected what is what is what is the expectation of y, we are now talking about only continuous random variables. So now what is what is the expectation of y?

Again what are you going to do; you are going to make number of experiments and then from each experiment you are going to take a value of y and then and then simply sum them. So how do you how do you... and in doing this you have to consider the whole range of x and y right so it is you are given the joint density function, so what is the probability that a y... if you if you perform an experiment you you will get y so so that will be obviously f xy x y dx dy that is so so this is the probability that you you will get a value y here (Refer Slide Time: 27:25).

Now I mean x is a free variable right. So now you first integrate it for all possible x and then you integrate it for all possible y then you will get the expectation of y. So if now let us try to rearrange the integral. So we are writing at minus infinity plus infinity minus infinity to plus infinity what is f xy what is this that is a if we write what is what is f xy x y; that is what is the probability that x falls between x and x plus dx and y falls between y and y plus dy that is nothing but f xy dx dy.

(Refer Slide Time: 28:12)



So we can also write it as what is the probability that y is this is this is nothing but Bayes' rule okay. So we are now writing the joint occurrence of x and y as y given x that is the probability of y given x and then the probability of x; we can always write it. So in doing so we have express the total probability as a conditional probability of y given x multiplied by the probability f(x) okay.

So now we can... one of these integrals is actually over x, the other integral is over y. So now we are saying that we take one integral inside and we sort of bracketed these terms so now this is expectation y f y given x dy. So, as if this is the conditional expectation of y this is this is nothing but the conditional expectation y given capital X is equal to small x. So you are saying that first you find out what is the expectation of y if if given let us x is equal to 1, then you find out what is the expectation of y if given x equal to 2, then we find out what is the expectation of y given x equal to 3 and so on. So each time find out the the conditional expectation of y condition on a given value of x then multiply weight each of these by the probability of x which is a same thing. So that is what we are doing. So first we are calculating the the conditional expectation and then we are and then we are weighting it because then we have because we have to multiplied by the probability of x, so the probability of x is f x (x) dx so so then we are weighting it using fx (x) dx right.

So what you actually get is this one. So you get expectation of y given capital X is equal to x which is this integral, multiplied by fx (x) into dx; now this looks like what? this looks like simply like previously just like minus infinity to infinity remember g(x) f(x) d(x); if y was is equal to so g(x) f(x) d(x) if you remember that if y is equal to g(x) we we derived so then expectation of y is equal to minus infinity plus infinity g(x) f(x) d(x).

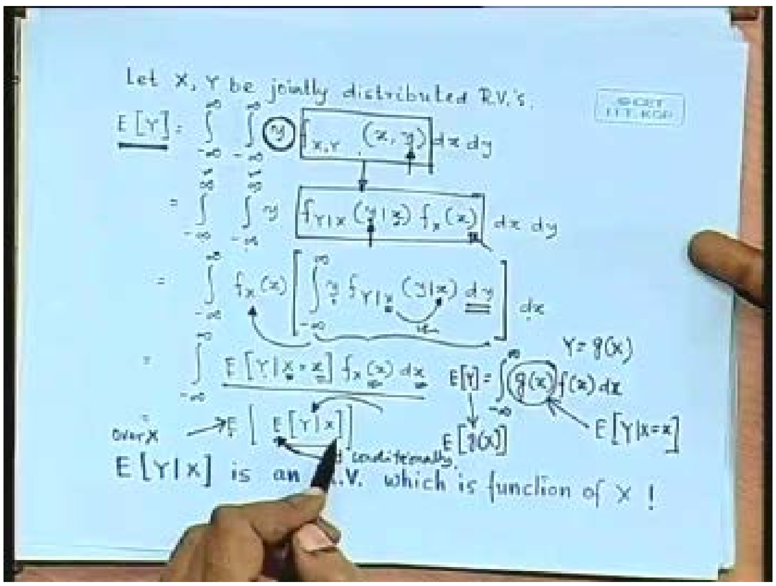
So now what is this g(x)?
This g(x) itself is becoming expectation of Y given X is equal to x. So what is this g function? This this this g function is nothing but the expectation operator. So now we are we are... what are the calculating, we are actually calculating expectation of g(x) because we have calculating expectation y, now y is equal to g(x) so we are calculating nothing but expectation of g(x). But g(x) happens to be this so we are actually calculating expectation of expectation of y units.

So this is very interesting and sometimes so useful that if you want to calculate the unconditional mean of y it becomes sometimes very useful to first calculate a conditional mean as a function of x and then taken expectation over x. So so so remember that this expectation is over x while this expectation is is over y conditionally with x.

So in other words, what I wanted to say is that this expectation operator at least the at least the conditional expectation itself is like a function of x. So when you write expectation of y given x it is a function of a random variable x simply because the

expectation value depends on x. So if you if you take x greater than 65 you will get one value, if you take x greater than 55 you will get another value. so obviously this this expectation of y given x greater than 65 is going to be different from y given x is greater than 55. So in that sense it is a function of x. So so the so the conditional expectation itself is actually a random variable which is a function of an which is a function of x
Right. So this is...this is interesting.

(Refer Slide Time: 32:15)



Next we can actually again we can we can we can extend this same thing for for functions of two random variables. For example; if you have z is equal to g(x, y) so you have two random variables x and y and z is a function of a random variable. Now what is expectation of z? So obviously expectation of z is expectation of z given x, y same thing same rule we are applying but the only thing is that this expectation is now over x and y both. So this expectation is now over x and y.

So exactly you see what that is precisely what I am doing, this is my... this is my conditional expectation that is this term is this term (Refer Slide Time: 33:41) and then I am taking joint expectation over x and y so I have added the joint distribution I am multiplying by dx dy and I am integrating over minus infinity minus infinity plus infinity. We can also write it like this.

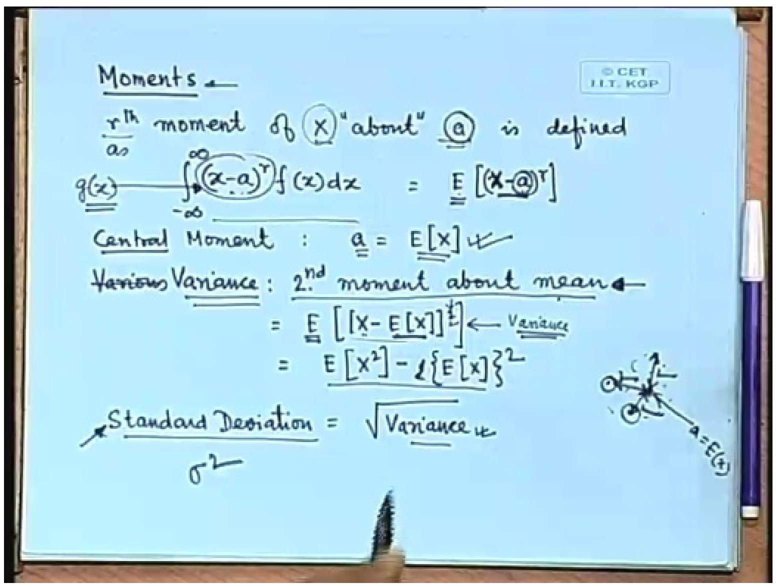
Similarly, I am just showing different forms. Suppose I... it is not always necessary that it has to be a condition both on x and y, it can be a condition on one. So for example suppose we have expectation of z given given capital X is equal to x, y is a free variable so then what do we do? So now we have to do expectation of given x and y or rather expectation of conditional expectation of z over x and y given X is equal to x. Now x is fixed we cannot evaluate this integral over x so we so this outer expectation now must be made over only y because x is fixed. So when we write it as an integral we have to write it over y okay.

Remember that this x is taking a value so this x actually freezes this x while this y is actually a free variable of integration which is the which is integrated here and this is the conditional this is the conditional expectation expectation and this takes... this is this is actually a function of this actually some function of g dash x and y so we will get it at the function of x and y. But for x we have to put the value small x, now it will become a function of y it will become a function of small y so now you have to do g dash of (x, y) f xy of (x, y) and then dy. So you have to vary y over the integral you have to keep x held at small x because you are given this and you have to evaluate at these integrate so in this way you have to evaluate it; conditional means conditional over various kinds of facts which may be stated in terms of x and y okay.

So basically the idea was to show that how to compute the expected value of a random variable under various conditions. So we have seen firstly how to calculate ordinary ordinary expected value, then we have seen how we can calculated expected value if it is the function of a random variable, we have seen how to calculate when it is a function of two random variables, we have seen calculate how to calculate if some already some constraining facts are given as conditions, so we have seen how to how to calculate conditional expectations, again normal conditional expectation, then conditional expectations of functions of one variable and function of two variables okay. So these are roughly speaking these are the various ways which cover.... you can you can obviously generalize it for n variables etc. But this is this covers the basic crux of computing of expectation.

Now we talk about... we gradually come to the concept of you know how much, so now we have an expected value something like a center of these of this distribution, now we want to somehow characterize that how big is this ball around which the value usually I mean the random variable usually takes value.

(Refer Slide Time: 00:37:04 min)



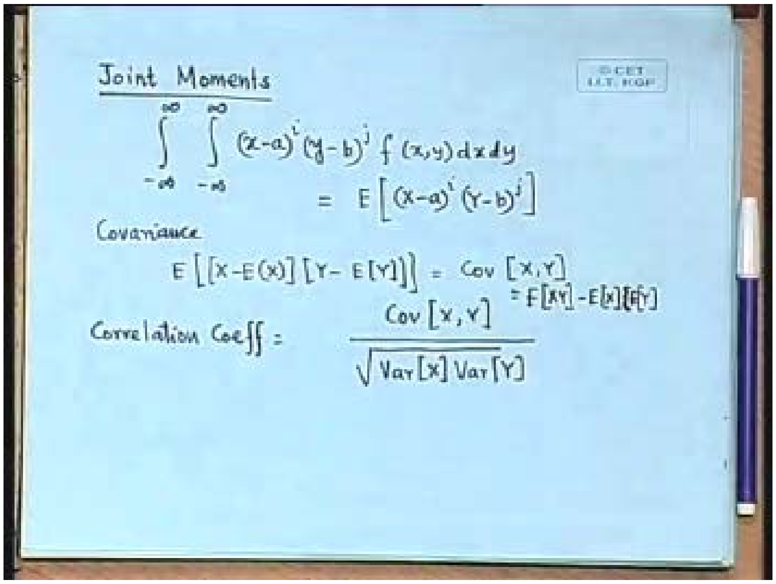
What is the... of course the ball is I mean strictly speaking it can take any value. But on an average that is I mean can we say that let us say with probability I mean what is the... what is what is the expected value of the dispersion okay. So we write express that. For doing that we first introduce what is known as a moment. What is the moment? So moment is an is an expected value of a of a certain special quantity. So what is that... so now in place of in place of g(x) we are choosing a special function so this is like g(x) you know. Only thing is that we are we are choosing a special form g(x) so what is that form? So the rth moment of x about a is defined as when you put this g(x) and you take the expected value of that g(x) then it is called... so so it is nothing but an expected value of x x minus a this random variable whole to the power r.

Now typically we can where do we choose x? Where do we choose a? So obviously what we are trying to find out here, we are trying trying to find out some measure of how this random variable is... how much it is away from a small a? So this small a number we can choose as anything and in general we are very much interested to know that in what is the expected value by which a random variable is away from the away from its expected value. So it has an expected value somewhere and it takes values around it. So for every value it has some x minus a, this this my a which is equal to expectation of x (Refer Slide Time: 39:11). So for every value that the that x takes there is a dispersion, there is a distance so we are just trying to find out that what is the expected value of this distance. Okay so that somehow indicates that on an average on an expected on on an average how much are the random variables how much away from them, so so if this expected value is large it means that the ball is large around the mean okay.

So typical... what is... I mean... so obviously we... very often a... very... this is a very common choice that which choose a is equal to expectation of x and since you know I mean the Euclidian norms and some squared norms are so popular and makes so much sense to us in the physical sense of the word so we often take a second moment about the... so this is a very very common choice taking the second moment about the mean.

So what we compute is X minus now because means so a is expectation of X and r is 2 so this is called the variance okay. I mean a bit of algebra I will tell you that it is nothing but the expectation of X square minus expectation of a X whole square so that is a very very standard algebra and we also very commonly define another term called the standard deviation which is nothing but the variance, the square root of variance on other standard deviation square or sometimes we we very commonly read out as sigma square is variance okay.

(Refer Slide Time: 40:53)

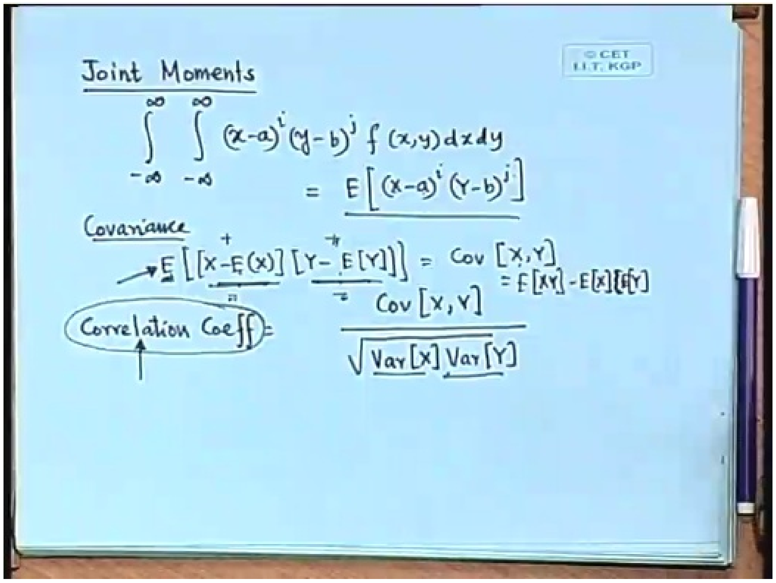


This is about a single variable. In in a similar way, suppose we have two variables then whenever we have two variables then we have to not only say... when we have two... when we have one variable we have to just just describe that how much it is... here is the expected value how much it is away from the expected value, now you have two variables. So you are going to have... so suppose you have two variables x and y so now x is going to be away from its expected value expectation of x and y is going to be away from its expected value expectation of y and this is going to happen together.

So now you you not only one to find out how much by how much x is away from expectation of x and y is away from is expectation of y but you rather want to find out that that whether these things happen together or or is it that when when y is close, x is far away and x is close, y is for away so you want to also characterize these things. So whenever you have two quantities you want to always see them as couples I mean you want to see their see their collection and you want to study the behavior of the collection of x and y but not individually x and individually y right.

So if you want to study the those properties then you have to define what is a what is a joint moment and so this is define like this which is nothing but expectation of X minus a whole to the power i Y minus b whole to the power i. This is a very very generalized definition of a joint moment.

(Refer Slide Time: 00:42:10 min)



But what we actually consider in practice is that we take we consider various types of things, we consider either... if you only want to consider that of X then we consider what is known as variance that is the old definition X minus expectation of whole square. Similarly we can find out just the variance of Y which is expectation of Y minus expectation of Y whole square, old definition.

But now we introduce a new quantity which characterizes that that when X is X is lower than expectation of X is Y also lower than expectation of X or is Y higher than expect... is Y lower than expectation of Y or is Y higher than expectation of Y on an average?
So we want to see them as couples, we want to see their collective Va Vi. So is that when this is positive this is positive or is it on and average and when this is negative with this is negative on an average or is it that sometimes this is positive this is negative (Refer Slide Time: 43:27) this is this negative this is positive, all sorts of combinations. This is also a quantity of interest. Very often this this is required to especially to find out that whether the whether the randomness of X and Y are actually related. That is if X goes high Y

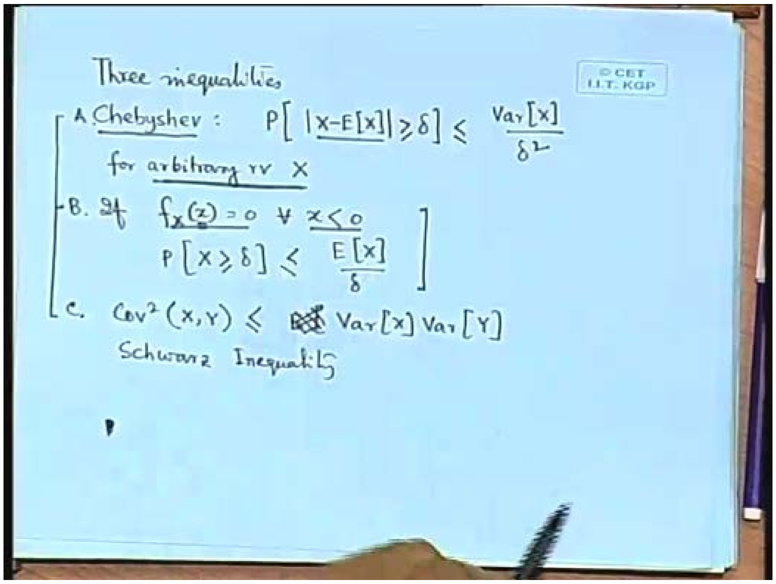
also... X is a random variable Y is also random variable but is there a connection pattern or what is called a correlation in their variation, are they somehow related? So to be able to characterize that we we often consider this joint moment which we call the covariance. So the covariance of X and Y is defined as this into this. So obviously you can find out that if they related. For example if this is plus.... whenever this is plus this is plus and this is minus this is minus then the overall sum is going to be plus so you have a positive correlation.

On the other hand, if you have it could be the other way. For example, you will also say that there is a correlation, whenever this goes plus this goes minus and wherever this goes plus this goes minus that is also a correlation, there is a pattern among them in which case again the again the expectation value will be non-zero it will be negative. But if in general this plus minus plus plus and all the four cases occur uniformly then you can roughly we say that the variation pattern of X and Y are really you know they have no relationship; sometimes this is positive that is negative; this is negative this is positive and these occur more or less uniform then then this expectation value is going to be low, it is going to be a number which will be very close to zero. In such a case X and Y are said to be uncorrelated okay.

So if there... whether there is a you know hidden relation between X and Y is is is captured in terms of covariance, so we take covariance and the normalize it by the individual variation patterns of X and Y. Because we do not want to know that, we just want to know that is what is the degree of correlation between them that is whether they move in the same way or they move independently and random right. So these are very important quantities which are very often used and we will use them in our future treatment in this course very often. So it is so it is important to understand these... I mean first of all to get familiarized with these concepts and that then then understand that what is it that I am trying to study right.

Now we will talk about certain inequalities or or the certain relationships so which are also very often used.

(Refer Slide Time: 00:46:15 min)



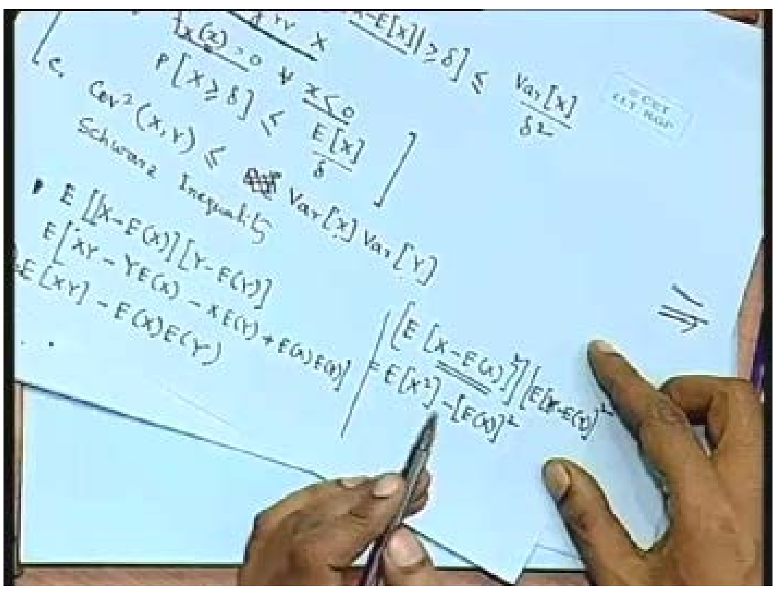
For example we have these three inequalities which are which show some properties of which relate mean and variance and then used in analysis many times so just want to introduce that. So for example you have this Chebyshev inequality which says that probability that the magnitude that the absolute value of X minus E[X] is greater than or equal to delta is less than equal to the variance of [X] by delta square. So this is this is actually very simple to prove because for any arbitrary random variable X right; so so how do you prove that? For example what is the probability that...okay let us first let us first cover these things.

So first of all if if f x (x) is equal to 0 this is there is there is there is another inequality which says that if you have only positive random variable that is the random variable does not take values which are negative for all X less than 0 and so whenever X is negative this is 0 which means that this random variable capital X takes values takes only positive values. So this random variable which takes only positive values then the probability that this is greater than equal to delta is less than expectation of [X] by delta right. And finally we have covariance square (x, y) is less than and equal to variance [X] by variance [Y] which means that the correlation value that is that is the correlation coefficient that we have considered here is always is less than or equal to 1 that is what we are saying.

So how does it happen; let us start from the back. For example expectation of X minus expectation of X into Y minus expectation of Y this is covariance, right? So let us multiply, then we get expectation of XY minus Y into expectation of (x) minus X into expectation of (y) plus expectation of (x) into expectation of (y).

If you take expectation now will get expectation of [XY] minus expectation of... this will give you minus expectation of (x) into expectation of (y). So this is your covariance (X, Y). So if you make it make it whole square what do we get? Now now let us see the other end. So variance of X is expectation of X minus expectation of (x) this into whole square into expectation of (y) minus expectation of Y whole square. What is this? This is nothing but expectation of (x) square minus expectation of (x) whole square.

(Refer Slide Time: 50:13)



Now you want to.... right right? So this is this and this is going to be multiplied by expectation of Y square minus expectation of [Y] whole square, correct? If you multiply them what do you get? you get... see if you whole square this what do you get, you get expectation of [XY] whole square minus 2 expectation of [XY] into expectation of (x) into expectation of (y) plus expectation of (x) expectation of (y) whole square and here you get if you multiply first of all you get if you multiply you will get this into this right so this so you will get expectation of X square into expectation of Y square.

Here we will get expectation of [X] whole square expectation of [Y] whole square so this term will get cancelled with this term and always note that expectation of (X square) into expectation of (Y square) is greater than or equal to expectation of (XY) whole square, this is obvious okay. So so so now using these properties you can very easily prove the prove this inequality.

Similarly by from from from simple integral formulae you can just try and then we can we can consider this later on in a in a in a tutorial class that that you can very very easily prove these properties. Also there are there are there are other simple things, today we are running out of time so I will not get in to proving this, so I just wanted to say that there are certain you know convenient inequalities which are which are used and which are very simple to prove. So I leave these as an exercise in the class which we will discuss in the tutorial class right.

So now what have we achieved today? So we today we have have introduced the concept of a mean or an expected value and we have calculated that for a for a for a number of cases, you know function of single variable, function of two variable, conditional expectations etc. We have also found out that... we have also introduced a a measure of dispersions from the mean. so we defined the we defined the variance in terms of a as a as a special case of a general moment function and we also defined the the concept of covariance which show what is the you know joint variation from this and we finally defined the quantity of correlation which is a very important quantity which which detects whether the two things are somehow related to move up and down randomly but these their whether their randomness are related, they are individually random but they are they may not be you know fully random in their interrelationships; and thirdly we saw three inequalities which as is never the easy to prove and we left it as an exercise. So that is all for today, see you in the next class. Thank you very much.