Introduction To Adaptive Signal Processing

Prof. Mrityunjoy Chakraborty
Department of Electronics and Electrical Communication Engineering
Indian Institute of Technology, Kharagpur

Lecture No # 01
Introduction to Adaptive Filters

Okay, welcome to this course. And you know what it is adaptive signal processing. It is a very enjoyable course and through this course you will learn many things which are useful in several other subjects. In communication and signal processing, image processing, machine learning, deep learning what not. So, I presume towards the end of this course you will really be satisfied with lot of useful knowledge. Okay, my name is M Chakraborty, Professor Mrityunjoy Chakraborty, Head of Electronics and Electrical Communication Engineering at IIT Kharagpur.

So, I will be pleased to you know bring to you this interesting topic adaptive signal processing. Now, core of the adaptive signal processing is adaptive filter. So, development of adaptive filter in its application that constitutes the story of adaptive signal processing. Now, what is adaptive filter? You have done one digital signal processing course somewhere you know in your previous studies; I am sure.

So, you have studied FIR filters, IIR filters and you can design them to have things like you know I mean low pass filter, high pass filter, band pass filter like that FIR, IIR. In the case of IIR filter you may have studied Butterworth filter, Chebyshev filter and those kinds of things. Those filters are very good, but their problem is once you design all the parameters that is the filter coefficients are fixed. And you construct the filter using these coefficients and use it in practice it will work beautifully all that is fine. But suppose the purpose the environment in which this filter is supposed to work or its goal it changes after sometime.

No way to change the filter you know you must throw it out, you must construct a new one. But in a case of adaptive filter, filter coefficients are not fixed they are continuously adapted, continuously changed from clock cycle to clock cycle by some nice adaptation relation or formula. By some training method or by some adaptation method or by some adjustment method these are very all terms which are equivalent training, adjustment, adaptation. So, that at the current moment it is doing the best job for us ok. So, that is the story of adaptive filter we will I mean I will consider adaptive filter applications and its general structure not at the very beginning.

But may be after some lectures you know once I cover a very basic adaptive filter called least mean square LMS, and then I will present many applications. Because then you will be able to appreciate those applications. So, right now if I bring in the applications and then bring in the theory and development and all that you know algorithm development, I mean you want to appreciate that much. So, you give me the freedom to start the way I want that is I will start with basics and then develop some very basic adaptive filter. And then we will put it into applications that beautiful applications will come, but I am telling you applications come in various contexts.

Ok communication, control, signal processing, image processing, machine learning, deep learning, even micro engineering so many you know. To give you a just an idea about what is an adaptive filter. Suppose you are given an analog signal x_a, a for analog, $x_a(t)$ which has a sinusoid component ok.

$$x_a(t) = A\cos(\Omega_c t + \phi) + z(t)$$

Ω_c is the frequency analog frequency, c for carrier. I mean I just I am just putting it Ω_c its unit is radian per second. So, radian per second and time is second.

So, second into second cancels whole angle must be radian you get by radian. So, radian per second is an angular frequency analog angular frequency may be plus of phase phi. This is a signal, but this is suppose added with some noise, white band noise, may be white noise ok. And noise power is not trivial not small I mean quite significant. You want to first process it digitally.

So, you will sample it may be at sample period, sampling period equal to may be capital T. So, nth sample of $x_a(t)$ will be $x_a(nT)$, T, 2 T, 3 T, 4 T like that is why at those points we will have sample. And this I denote by $x(n)$. So, I have $x(n)$. So, before I do anything I am converting into discrete time form and then I will consider this problem how to eliminate the noise as much as possible ok.

So, it is

$$x(n) = A\cos(\Omega_c nT + \Phi) + z_a(nT)$$

which is $x_a(nT)$ which I call $x(n)$. So, omega c small t is that is the time point I am taking the sample. So, $x_a(nT)$ may be $+\Phi\ + z_a(nT)$ let me put z_a if an analog to distinguish between analog and discrete time all right. Then little bit of DSP we know, $\Omega_c T = \omega_c$ which is digital frequency. This is basic of DSP basics in DSP you must be knowing, I am sure.

Digital frequency, its unit is radian per second and here second. So, second and second cancels radian. So, you can make that assumption let me call it ω_c, c for carrier. So,

$$x(n) = A\cos(\omega_c n + \Phi) + z(n)$$

So, this is one discrete time sequence cosine sequence and this I call $z(n)$ like $x_a(nT)$ was called $x(n)$, $z_a(nT)$ is called $z(n)$.

So, equivalently I have been given in discrete time domain one sequence which has a sinusoidal component which is by signal, but this is added with some good amount of noise and this is a wide band signal. Your purpose is to filter out this noise as much as possible. So, ideally what you should do you should design a band pass filter of frequency response like this at this magnitude response like this. At ω_c and at $-\omega_c$ you will have a pass band very narrow pass band.

So, around ω_c and at $-\omega_c$ there is a pass band very narrow pass band. If you pass this signal through this this cosine signal whose frequency is ω_c that will pass through, but noise which is wide noise which is wide band noise I mean this spread over the entire frequency range from $-\pi$ to π you know indeed DSP, this is the range and this becomes periodic whatever you have here that gets repeated. So, you see from $-\pi$ to π. So, what the entire range I mean this noise will have it represents. So, filter band pass filter will chop off will chop off most of the noise it will allow only the noise frequencies in this which are within the pass band.

So, obviously, most of the noise power will go and fine. So, you are happy with the signal which will go through and very little noise will filter through this because pass band is very narrow and this goes on for some time. Now, suppose after some time signal Ω_c frequency, signal frequency Ω_c changes to some another frequency $\Omega_c{}'$ means analog signal ω_c which is $\Omega_c T$ this change to some new digital frequency ω_c' which is alright then this will be outside the pass band may be ω_c' is here.

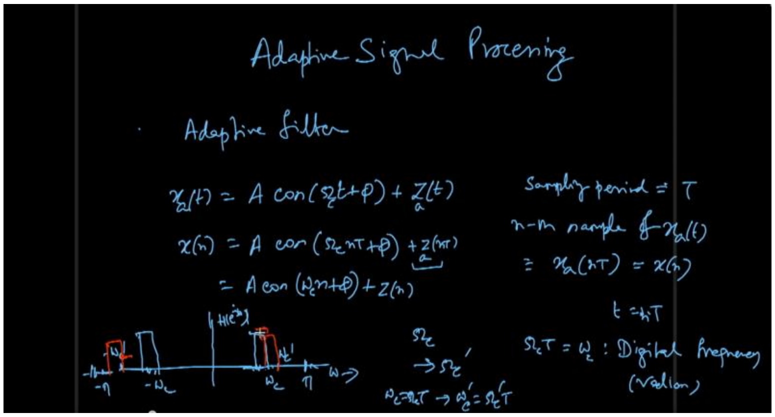
$$\Omega_c \rightarrow \Omega_c{}'$$

$$\omega_c = \Omega_c T \rightarrow \omega_c' = \Omega_c' T$$

So, then the signal also will be filtered out by this band pass filter because this center frequency the frequency of the sinusoid is outside the pass band and I am gone. So, my filter which is working fine now it will fail to work in future when the frequency changes, but if it is an adaptive filter what it will do this band pass filter if it is adaptive, it will not stop at this it will continuously monitor the input data and it will track the variation of the input frequency.

So, if the frequency changes it will also change its pass band by some formula. So, that new pass band is here ok. So, pass band will shift there will be an adaptation mechanism training mechanism adjustment mechanism by which the coefficients of the filter will change band pass filter. So, that if the frequency changes from input frequency changes digital frequency I mean Ω_c to Ω_c' the filter band parameters also will change. So, that new pass band will be here and here.

(Refer Slide time: 10:50)



So, the signal again will pass through ok. This is a very famous example this is called adaptive line because if you take the frequency I mean if you take the discrete time Fourier transform of this this will have two impulses at ω_c and $-\omega_c$ those are like lines and this will be flat quite noise. So, what this it will do it will enhance the lines that is signal will be enhanced or effectively noise will be reduced and it will be done adaptively. So, this example is that of adaptive line enhancer alright. So, this is gives you an idea about what is adaptive filter.

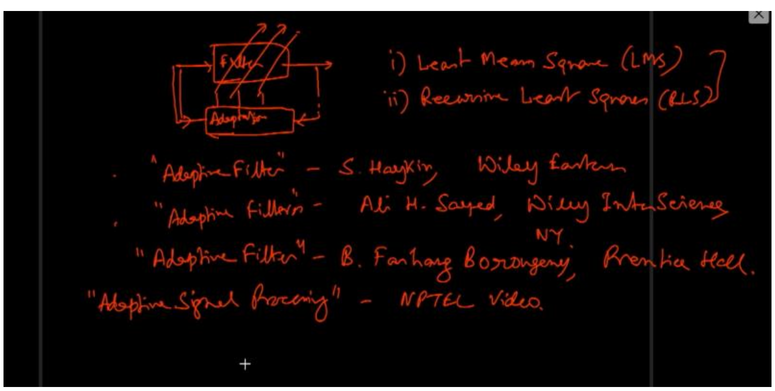
So, there will be an adaptation mechanism that is required because if the external environment changes like the frequency incoming frequency changes I should be able to process correctly ok. So, I should have an adjustment mechanism adaptation mechanism by which my parameters can be changed to suit to current requirements. So, the adaptive filter there is a filtering part in parallel there is an adaptation part alright. So, there is actually a filtering part and there is an adaptation part. So, there is an adaptation algorithm which works on the filter coefficients and adjusts.

So, symbolically a schematically we show like this tuning. So, at every clock it is calculating new parameters for the filter and changing. So, that change is indicated by this and adaptation algorithm requires input information also output information also ok. This is a typical structure of adaptive filter both the filter part and adaptation part. So, adaptation part is what is our target there are two very famous class of adaptive filter algorithms, one is called least mean square LMS, another is recursive least squares RLS alright.

We will study these and these, but again we will just give an introduction because this is a very huge topic many algorithms under each of them category each category has come up their analysis is quite involved and all that. So, we will not be able to cover all, but I my purpose is to introduce you to these two and through that journey also make you knowledgeable about many related things which you find useful in other subjects like signal processing, image processing, communication and as I told deep learning, machine learning and all that alright. There are two there are many books available one is adaptive filter I cannot remember the exact title whether it is introduction to the adaptive filter or adaptive filter theory and practice and all kind of thing, but essentially adaptive filter. This is by no other person that good old Simon Haykin, I think it is a Wiley book, Wiley Eastern available in India, another is a more involved book by Ali H. Sayed It is again Wiley Interscience New York available in the internet, I think New York.

There is another book of mine I mean which I like adaptive filter by an ex Iranian friend of mine his name is B. Farhang Boron Geni, Prentice Hall. But apart from this please see my video lectures related you will find some useful information here. But most importantly it is very important in any of my subjects that I teach to attend to my lectures and understand there because many things I say in the class which are not found in any of the books ok they are very useful alright.

(Refer Slide time: 16:13)



Now so how to go about the subject the basis of this subject is you know two things one is probability and random variable because the signals that we come across they are random in nature ok.

So, how to characterize them and all some basic properties and you know features and all you have to study like probability density of one variable, two variable, multiple variables, correlation matrix, covariance matrix, their properties all those ok. Random process these things I will cover I will take quite some time then I will go to what is called optimum filter and then make it adaptive filter by LMS and then I will take up the applications I will show many examples then I will come back I will do some analysis of the LMS algorithm that it should finally converge to the best filter you know if the adaptation should take you to the best filter and then I will take another category which is recursive risk squares again there are lot of many I will take the very basic one. So, that will be the flow of this course roughly twenty lectures, but I think this is a very good foundation for many related subjects including even neural network. Neural network also is a kind of adaptive filter only thing is in adaptive filter input is a sequence, time sequence, comes in time. In neural network we also have filtering kind of thing, linear combination we also have error pressure of threading, but input is not a time sequence, input is just a stationary data a block ok difference, but many things are common between the two.

So, in with that background I did start basic probability of random variables ok. So, we will consider real and continuous random variables. Ok, real and continuous like suppose you are measuring the temperature day time temperature call it T. So, T will take real value it cannot be complex, but T is not fixed any time in anywhere any part of the city or any time any during the day and night it will have fluctuating values ok different values. So, it is random it is real because it is not complex and it is not discrete it is not the temperature will be either 30 degree or 31 or 32 it can be anything.

So, on a real axis if it is a real axis suppose the origin is 0 degrees here it is continuous it can be here any way it is not discrete that it will be either 1 degree or 2 degree or 3 degrees ok it can be. So, this kind of random variables are called continuous and of course, real random variable. More formally whenever I say there is a random variable real continuous any random variable it means along with the random variable say x we denote the variable by X, always there is an experiment or trial or observation process ok by which we observe the value of this like a Bezier using a thermometer the temperature and find a value. So, that moment whatever be the value you get you assign that value to this. So, there is always in the background an observation or experiment and notion and notion of experiment or trial or observation along with a random variable that is it is not fixed every time I observe ok or I experiment and find the value give it to this.

So, its value changes fluctuate experiment to experiment, but there is an experiment or observation trial that notion of that goes with that and if it is continuous real then it will take in all experiment's real values only not complex and if it is continuous the values will be continuous on the real axis it will not be taking specific discrete values that is the only thing all right. This is the meaning of random variable continuous real random variable. Now suppose this is my real axis, this is the origin X will take any value, I mean it is not necessary that it will take all the values from minus infinity to infinity, it may be concentrated in certain zone. For instance, if you are measuring temperature, you know it can be from may be 10 degrees centigrade to 40 degrees centigrade it cannot be 1000 degrees centigrade. So, whenever you measure your measured values will may be concentrated in some zone, but there will be real and there will be continuous they will not take pre specified discrete values only there are continuous on these axes ok.

So, suppose X now I want to find out probability, I must define probability. So, suppose I mean I have a position some x this much value and x to some small portion dx I take. So, $x\ to\ x + dx$ and dx is very small, then the probability that X in my experiments will fall here, that is probability that X will be less than equal to this point greater than equal to this point. That should be proportional to this width because if it is larger so obviously, transits will go up of X falling here if it is smaller.

So, it will be smaller. So, as long as dx is very small I can say this probability is linearly proportional to this width dx.

$$Prob(x \leq X \leq x + dx)\alpha\ dx$$

So, it is proportional to dx that means, it will be equal to some proportionality constant times dx. So, that proportionally constant is given by P and I put a subscript X because my random variable is X not Y or Z.

$$Prob(x \leq X \leq x + dx) = P_X dx$$

Because there may be a situation where I may have very probably random variables X, Y, Z to distinguish between them, I put a subscript not only at dx this is a proportionality constant ok. But not only that this P_X will depend on where this slot is located that is at point x because if I now take here the same dx, the probability of X falling here may not be same as this. It will depend on like as you tell do as I told you day time temperature maybe this concentrated between 10 degrees to 40 degrees if it is 100 degrees.

So, obviously, there will be hardly any chance ok which means probability of X falling in a slot like this will depend on where this slot is located here or here or here or here or here. So, this is located at the coordinate x. So, this will be this proportionally constant will vary, it will be linearly proportional to dx, but proportionally constant cannot be same it will be changed depending on where is this slot located ok. So, it will be a function of small x ok these are proportionally constant and that is called probability density function of x alright. Therefore, probability of X falling between two points some x_2, x_1 that will be what that is suppose I got x, this is 0, this much is x_1, this much is x_2.

So, probability of the X will be here will be what I must sum up the probability of probability in this slot in this slot in this each of widths dx. So, probability of X falling in this slot of width dx, in this slot of width dx, again in this and so on. So, I must add since dx is infinitely small this total probability will be nothing, but integral. It will not be a discrete summation, it will be integral, that you know integral as a definite integral as a limit of a sum, if dx is very small and goes to 0, this summation will go to become a integral. So, this will be nothing, but this thing summed, but summation will not be discrete sum will now be this is the probability.

$$Prob(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} P_X(x) dx$$

Therefore, a question comes probability of X taking value from $+\infty$ that is the extreme point to the right, to the extreme point to the left $-\infty$, that will be just this input change the limits, but you understand in every experiment X will take a value from $-\infty$ to ∞. So, that is an event called certainty, that is an event is bound to occur, there is X taking value from $-\infty$ to ∞ in each trial.

$$Prob(-\infty \leq X \leq +\infty) = \int_{-\infty}^{+\infty} P_X(x)dx$$

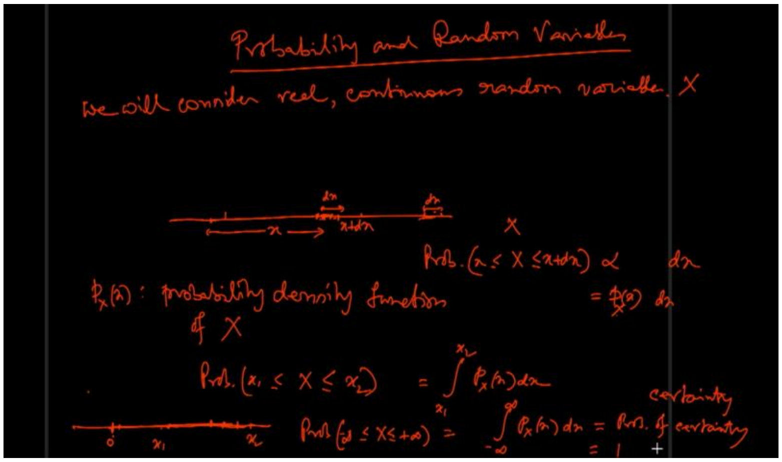
So, probability of that is this. So, this is nothing, but probability of certainty. Now certainty is a global event, global parameter, sun rises in east.

This is certainty all human beings will die certainty. all human beings were born one day certainty. So, they are basically same event, they are manifesting in defined form, but they basically mean the same thing that is certainty. That is something that is certain to happen. So, certainty will have a you know constant will have a fixed probability whether sun rises in the east, probability of that or all human beings will die, probability of that.

Basically probability of only same event that is certainty that is something which is certain to occur. So, that is basically a constant independent of whether I am talking in terms of sun or in terms of human beings dying certainty. So, this is nothing, but probability of certainty and that we can take to be a we can normalize to a value of our choice this is probability of certainty which we normalize to be 1. If we have taken 100 also does not matter, but globally it is taken to be 1, that certainty is an event which will always occur and we take its probability to be 1, which means this integral when $-\infty$ is here and $+\infty$ here. This is the probability of X falling from $-x$ to x that is equal to 1 that we take all right.

$$\int_{-\infty}^{+\infty} P_X(x)dx = 1$$

(Refer Slide Time: 31:19)



Now suppose you are given I am writing here now suppose given a $f(X)$, X is the variable, x is the actual value ok. Now every time X you find out by an experiment you plug in the value here you get something next time another X plug in here you get something. So, this function also is random its value changes fluctuate randomly depending on experiment. So, what is the average value of this function? Average value means X falling here ok because you have this slot the dx is so small whether X is here or here or here or here or here dx is small. So, I will assume it will take a constant value f at the boundary whatever be the value same value whether I am here because dx is infinitely small.

So, function does not change much hardly changes. So, function value at that point which is x ok that point which is x that value will continue inside also ok that is the thing, but that will not always occur it is not that whenever I measure X I will get this I will get this with a chance. So, what I do I multiply by a chance factor. So, if the chance probability of X falling here is high, value will be this will be you know multiplied by a weight appropriately high. that is that this value will occur with more power. If the chance of X falling here is very less, then the chance factor which is this you know that probability of X falling here that will be less.

So, I will be multiplying this by this. So, its value will go down because chances are less ok. So, I will be applying a weight factor multiplication factor on this which is the chance factor ok. The chance of I mean having this value f of this much x which is same as whatever we have inside ok. Let me go to the next page I draw again this much is your x. So, this width is dx ok function is $f(X)$ takes the same value $f(x)$ whether here or here or here or here because dx is so small its value does not change.

So, when I conducted experiment, I observe X. So, I find out $f(X)$. So, that will be equal to this value with a chance what is the chance was chance factor was this much that is the chance of X falling here. So, the value will not be this because I am not certain that I will get this. So, I have to multiply by the chance I have to use weighting, I have to weight this by a chance factor. So, I weight this by a chance factor then again you can take next slot.

$$f(X) = \int_{-\infty}^{\infty} f(x)P_X(x)dx$$

Again, $f(X)$ falling I mean taking value from here it will be multiplied by another chance factor f of you know and it will go on. So, if I want to take a weighted average this function into weight again another value into weight and if I add them. Basically, this will be integral because dx is infinitely small. So, if I add them what I am doing is actually you can understand one thing suppose I gave you some variables x_1, x_2 or maybe not x_1, I gave you some values maybe $y_1, y_2, \dots y_n$ and I have scalars $c_1, c_2, \dots c_n$. So, I multiply y_1 by c_1 weight, weighting using a weighting factor on y_1 c_1, then $y_2 c_2$, $y_n c_n$ and then divide by $c_1 + c_2 + \cdots + c_n$ this is called weighted average.

$$\frac{c_1 y_1 + c_2 y_2 + \cdots + c_n y_n}{c_1 + c_2 + \cdots + c_n}$$

All of us know this. So, it was discrete same thing I am doing I am taking this function multiplied by a weight factor and this I am doing over all the slots. So, I am adding here. So, adding here adding here it was a discrete sum here is an integral. So, I am adding. So, if we divide it by sum of the weights are these weighting factors it was discrete sum now it is continuous sum ok.

So, this is a weighted average or average or expected value of this, but this we have seen this is equal to 1. we have seen in the previous page because this is the probability of certainty that is 1. So, this entire thing we denote by $E[f(X)]$ again I put a X here because I may have many random variables X, Y, Z and all that to distinguish, there is an alignment problem here. So, expected value average value it is a weighted average that will be same as because this is 1. So, it will turn out to be this here you put x integral variable all right.

$$E_X[f(X)] = \frac{\int_{-\infty}^{\infty} f(x)P_X(x)dx}{\int_{-\infty}^{\infty} P_X(x)dx}$$

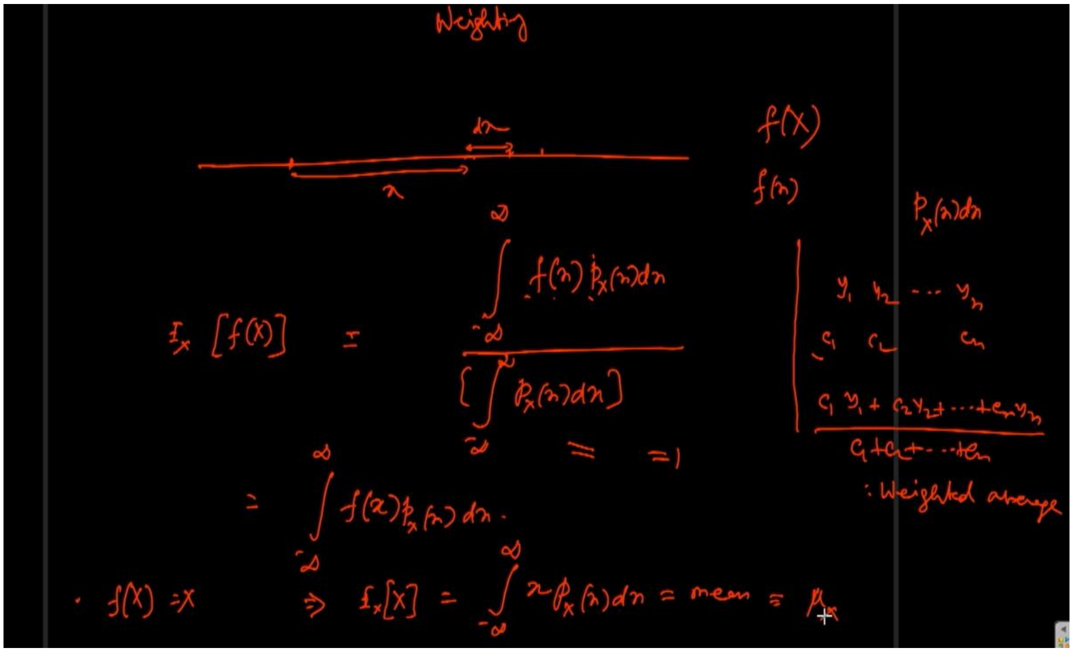
$$E_X[f(X)] = \int_{-\infty}^{\infty} f(x)P_X(x)dx$$

So, this is called expected value, expectation, $E_X[f(X)]$ average value all right. Example 1 suppose $f(X) = X$ itself very simple function $f(X)$ is nothing, but x itself, then $E_X[f(X)]$, $f(X) = X$. So, you put $E_X[X]$, x here that will be what $f(x)$. So, this was X.

So, this $f(x)$ integral will be $f(X) = X$. So, $f(x)$ is x and this remains as it is. So, x multiplied by this probability thing trans factor integral and that is called the mean because expected value of the random variable that is mean you can denote it as μ. μ of random variable with X. So, μ_X all right.

$$E_X(X) = \int_{-\infty}^{\infty} xP_X(x)dx = \mu_X$$

(Refer Slide Time: 37:48)



Then another function can be,

$$f(X) = (X - \mu_X)^2$$

So, here expected value will be $f(x)$. So, it will be

$$E_X[(X - \mu_x)^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 P_X(x)dx$$

This is called variance and denoted by σ^2 you can put X to indicate random variable X. What it means that μ was the mean this is μ_X and around that I have fluctuations. So, X will can be here may be here may be here may be here in experiment to experiment.

So, I am finding out the value minus the mean this much is the increment this much is the increment this much is the increment ok I am squaring them up. So, even if it is positive square of positive even if it is negative squaring of positive. So, every increment incremental part I am taking the power means I am squaring ok and then average because I am multiplying by the trans factor and integrating. So, it is the average power of the increment incremental part ok like it is a dc. So, if you take the waveform minus the dc that will be ac fluctuating around the origin.

So, then you square up square up square up square up average that is what it is. So, basically average ac power it is called variant all right. Now I consider joint random variable suppose I have got two random variables X Y ok both continuous and real like suppose you are measuring temperature which may be X and humidity which is Y. Now temperature humidity I do not know whether there is any formula relating them, but at least we understand that they have there is some kind of relation between them temperature they are all about atmosphere. So, one influences the other and vice versa how and all we do not know, but temperature humidity.

So, every time I conduct an experiment, I observe not just one variable here I have measure 2 which and these two have an internal relation between them one is temperature which is x another is a humidity y. Both are real because the economic complex both are continuous, they cannot be discrete that humidity is either this or that or that. So, both are they may be concentrated in some area or the real axis. So, in every experiment I measure a pair therefore, now instead of 1, x_1 real axis which is for x I will have 2 all right. And now I want to come to again similar notion of probability density not of 1, but of two variables jointly they are called jointly random variable ok.

So, I go to x and then this much is supposed dx and here again I go up to 1 and suppose this much is dy. So, this coordinate is x and y. So, its coordinate is x, y, this point and this point is $x + dx$, $y + dy$ all right. So, every time I conduct the experiment X and Y together may fall within this box that is may X may lie between x to $x + dx$ simultaneously Y may lie between y to $y + dy$ or may not ok because a random. So, if they lie if X lies between x and $x + dx$ simultaneously with Y between y to $y + dy$ that means, the pair lies here ok in this small box.

$$Prob(x \leq X \leq x + dx, y \leq Y \leq y + dy) \ \alpha \ dxdy$$

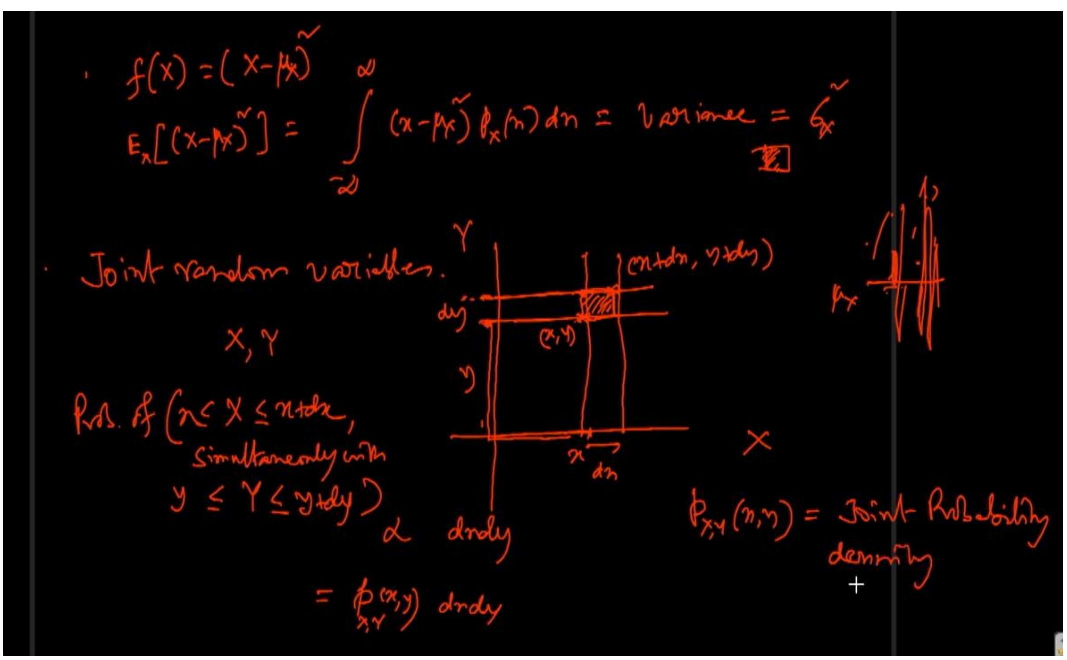
So, the probability that the pair x, y will lie within this box will be linearly proportional to the area of the box because area is large that means, my chance will go up area smaller means chance will go down as long as the area is very small, I can assume it to be linearly proportional. So, probability of X lying between $x + dx$ to x simultaneously with simultaneous is very important that is why they are joined with Y that will be linearly proportional to linearly proportional to area, area is $dx\ dy$. So, equal to a proportionality constant. So, I put X and Y here to under indicate which random variables are important are present constant $dx\ dy$. Now this constant will depend on the location of the slot because if the location is somewhere here same $dx\ dy$, but there is no guarantee this probability will be same here also because it is quite likely it may be likely that you know I mean more often you will have temperature in this zone and you know humidity in this zone hardly ever here.

So, basically the two probabilities will not same. So, $dx\ dy$ is same. So, proportionality constant will vary. So, that will depend on where this is located it is located at x y ok. So, it will be a function of and this thing is called joint probability density.

$$Prob(x \leq X \leq x + dx, y \leq Y \leq y + dy) = P_{X,Y}(x, y)dxdy$$

So, if somebody ask you in an interview what is probability density you have to give the entire background like if it is just only one ok. If it is only one you have to draw this line draw up to x then $x + dx$ then tells that probability of X falling here will be linearly proportional to this then the proportionality constant ok.

(Refer Slide Time: 45:10)

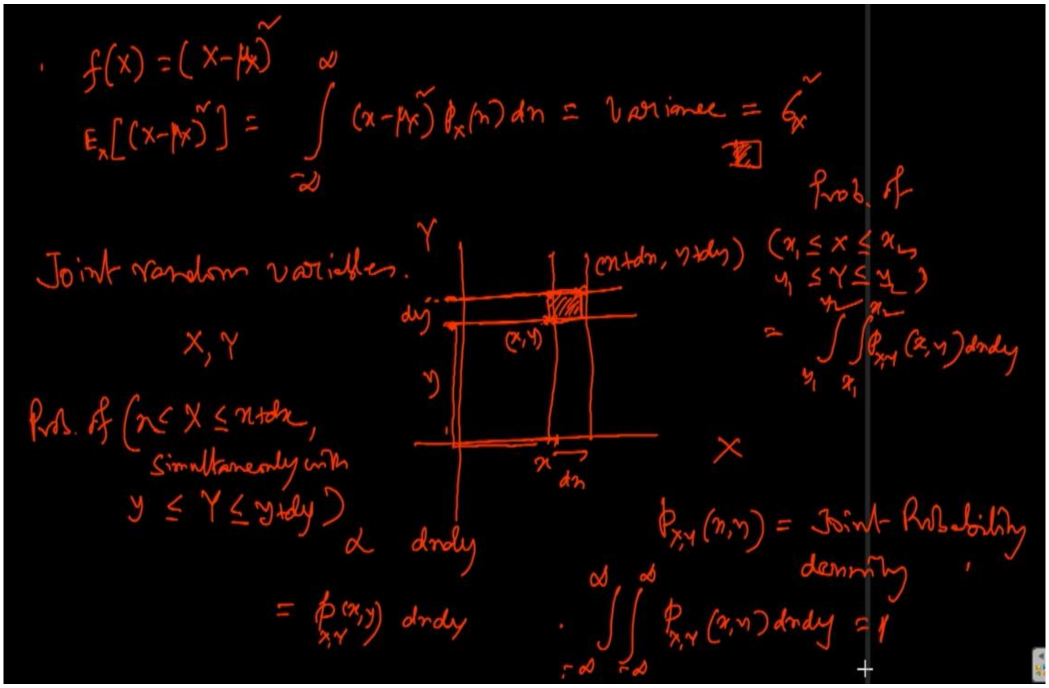


That will be this thing probability density case you have to draw this line then you have to define that this is linearly proportional to d x then the constant which is a function of the location of the slot that is called probability density it is integral will be one all those things you have to tell all right. Now, after this so, certain things will follow very simple. probability just generalization from the previous example probability of X lying between x_2 to x_1 simultaneously Y lying between y_2 to y_1 this is nothing, but I have to integrate this over that zone $x, y \ dx\ dy$ now double integral from x_1 to x_2 from y_1 to y_2 and clearly from $-\infty$ to ∞ if I do this again that is the probability of certainty, because the probability of XY lying between lying on this entire plane there is X from $-\infty$ to ∞ and simultaneously Y from $-\infty$ to ∞ that will always happen. So, probability of that is given by this formula this is again the same event which is your probability of certainty all right probability of certainty, but that is constant that is global that is always 1.

$$Prob\ of\ (x_1 \le X \le x_2, y_1 \le Y \le y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} P_{X,Y}(x,y) dxdy$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_{X,Y}(x,y) dxdy \ = 1$$

(Refer Slide Time: 47:57)



So, this will be 1 ok. So, this much is for this lecture in the next lecture I will again proceed the same way I will assume that given a function $f(x, y)$ now or X comma Y. What will be the expected value of that and then I will take some cases of that and then I will go to something very related and useful called conditional probability density all right. So, I stop here now and we meet next time. Thank you.