Advanced Mathematical Methods for Chemistry

Prof. Madhav Raganathan
Department of Chemistry
Indian Institute of Technology, Kanpur

Module - 11
Lecture - 04
Error Estimates, Least Square Fit, Correlation Functions

This will be the 4th lecture of module 11. I am going to talk about Error Estimates and we are going to talk about Least Square Fit and then finally I will the last topic that I will do in module 11 is that of Correlation Functions. So, I will try to cover all these very briefly and very quickly. And some of these will be familiar to you, so I will be able to go fairly fast.

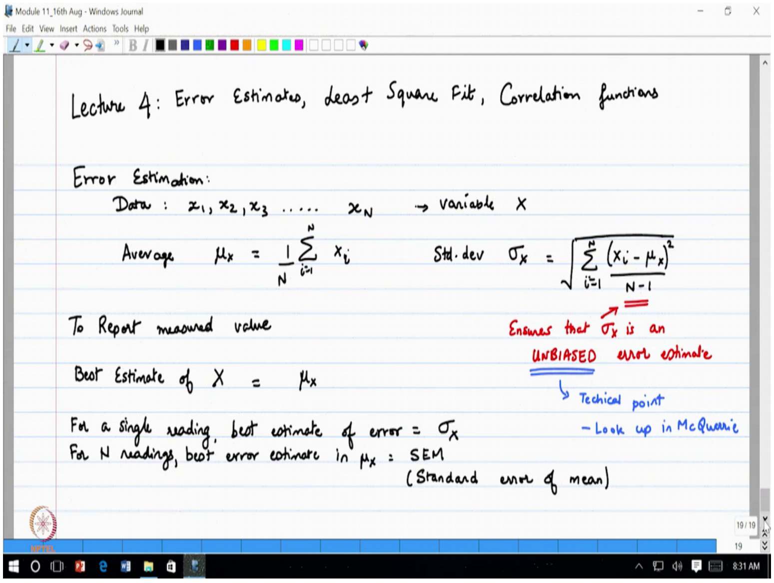
So, we will start with the idea of Error Estimates. And, this is something that you have seen many times, but you might not appreciate some of the intricacies of it. So, you might of seen for example, that let us say there is an exam and all of you gets marks in the exam. And then what you ask for what is the average score of the whole class; you might ask the standard deviation of the marks, you might ask things like that, ok.

Now, now also for example you might see various data reported in books or other places where they right some value and then they give some plus minus sum number. So, they will say that they will give a range of values instead of giving one precise value. And what we will see is that these are actually related each other and what will also see is that there is an underlying assumption that things are following Gaussian distribution.

So, we already saw in the case of the central limit theorem that if you have a very large number of variables then the distribution tends to look like a Gaussian random variable. And so it is this idea that is underneath the use of standard deviation as an error estimate ok.

So, we will just go quickly and first talk about error estimates. So, let me start.

(Refer Slide Time: 02:25)



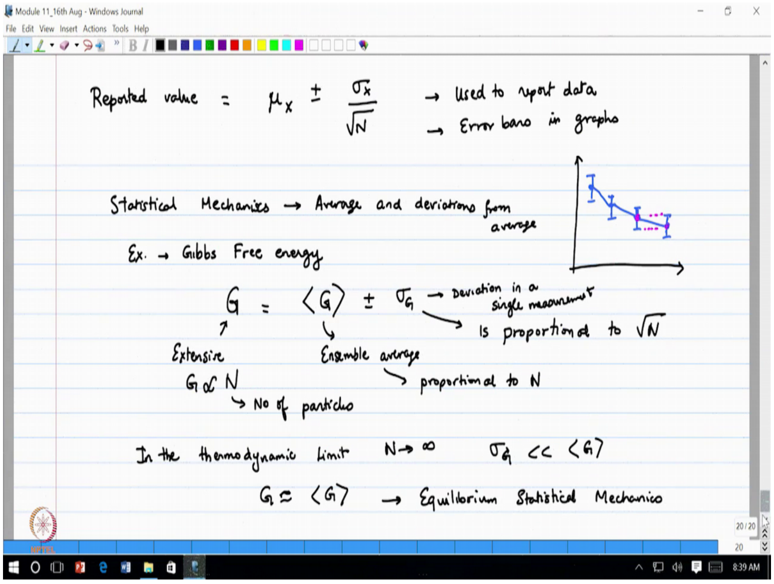
Error estimation: so suppose you have data x 1, x 2, x 3 up to x N there are N data and the data is of some variable x. Then what you say is that you define the average mu x is equal to sum over i equal to 1 to N x i divided by N. And you define the standard deviation sigma of x as square root of sum over i equal to 1 to N x i minus mu x square divided by N minus 1. So, this is the definition of the average in the standard deviation.

Now we should note we will emphasis gain that the N minus 1 is there to ensures that sigma x is an unbiased error estimate. Now I am not going to talk about meaning of the word unbiased, but you can this is a technical point and you can refer and read in various books- technical point, ok.

So, this if you are interested you can look up in McQuarrie; what the meaning of this unbiased estimators. I will not bother about that, but I will just say that I need N minus 1 when you are defining the standard deviation. So, now suppose you have to report value. So, what you say is that the best; so if you want to report a value, so to report measured value. So, you imagine that you have done taken N readings got all these values and you want to report measure value. So, what you say is that the best estimate of X is equal to mu x. Which seems very obvious, so the best estimate for x that you can get is the average value of x. And that is what you usually report, but actually the it is a technical point that this is an estimator of x.

So now, if you ask; so for a single reading best estimate of error is equal to sigma x for N readings best error estimate in mu x; so the question is how close to you think mu x is to the actual answer. So, the best error estimates. Now what you would expect? So, this is called the standard error of mean. So, it is error in mu x. So, when you report mu x as a reading what is the best estimate of the error of mu x.

(Refer Slide Time: 06:30)



So, the standard error of mean, this works out to sigma x divided by square root of N. So, what it means is that you are reported value is equal to mu x plus minus sigma x divided by square root of N. So, often you will find that people give the error estimate in the reported value and that is usually the standard error of mean.

So, the point is that if you take any one reading then the error in that one reading is sigma x, but if you report an average the error in that average is sigma x divided by square root of N. Therefore, if you make more measurements then you expect the error in the average should go down. So, this is how you know data is reported; so use to report data. Also used show error bars in graphs. So, suppose you have a graph and so often what you will find is that your graph will have various points, and each of these points is dot by an average over lot of quantities.

So, what is done is that you might do a fitting of the graph, but what is also done is that you show errors on the; you show this error bars on each of the readings. And usually to if two error bars and two readings overlap then you say that you know you cannot distinguish these two readings within the error bars. For example, this point here and this point here let me show it in a slightly different color; this point here and this point here, ok.

So, there error bars overlap. So, you would say that you cannot actually distinguish, because what you are saying it is this point plus minus so much this plus minus so much, so there is a good chance that they are the same. So, you really cannot distinguish these two within the error bars. So, what you do in such a case to make more measurements so the error bars go down.

So, it is a very common practice that you will see often, this reporting of error bars. Also what you will see in you will also see some of these in statistical mechanism where what you will find is that the error in average value that becomes negligible.

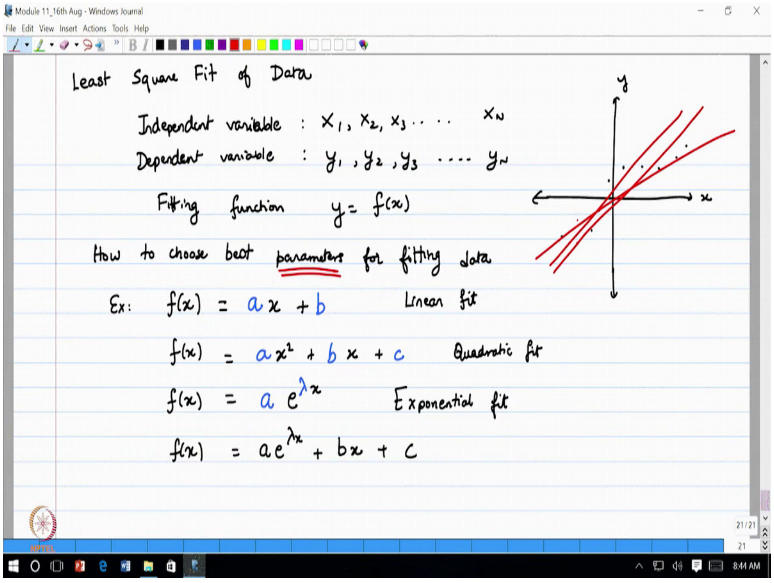
So, in statistical mechanics we calculate average and deviations from average. But what we will find is that for an intensive variable this. So, suppose you have energy; so for example you might you might calculate the Gibbs free energy. So, G, this is proportional to N, so this is extensive variable proportional to N various number of articles. Now in statistical mechanics we do something called ensemble average. So, G is written as the ensemble average show in square brackets, so plus minus b sigma of G.

So, the standard deviation of G ok; so, it turns out that this standard deviation is proportional to square root of N. So, then this is proportional to square root of N, whereas, this is proportional to N. So, this is deviation in single measurement. So, basically we can write: so in the thermodynamic limit N tending to infinity so sigma G is less than less than average value of G. Sigma G will comes much smaller, since this is only proportional to square root of N, whereas this is proportional to N.

So, for example if N is 10 raise to 22 then sigma G will be of the order of 10 raise to 11 which is much smaller than 10 raise 23. So, basically you can write G is approximately equal to average value of G; this is one of the things that you do in equilibrium statistical mechanics. Or you can say statistical thermodynamics.

So, the idea is that thte deviations are very small. So, your G becomes approximately equal to average value of G.

(Refer Slide Time: 12:52)



So, now the next topic that I want to do is Least Square Fit of Data. This is something that that you have seen many times when you often do it in your lab. For example, if you have a table of data, so you have let us say the independent variable. So, this is x 1, x 2, x 3 up to x N and you have a dependent variable that you measure y 1 y 2 y 3 to y N. And now you want to fitting function, we have a fitting function y equal to f of x.

So, the fitting function might be linear, it might be quadratic, it might be anything, might be exponential anything. So, typically what you think of is that you have x and you have y and each point, so you have x 1 y 1; so x 1 y 1 might be let us say here, x 2 y 2 might be here, x 3 y 3 might be here x 4 y 4 might be here and so on. You might have all these points, ok.

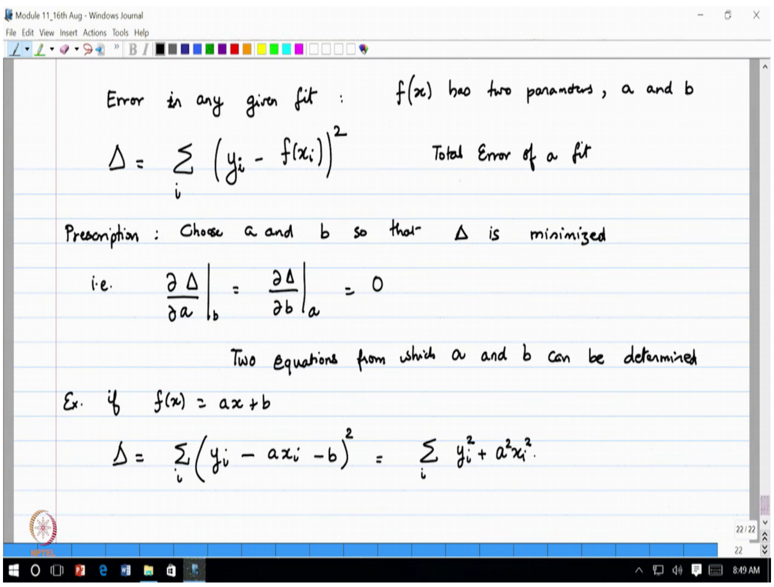
So, now what you want to do is you want to fit some curve to this and you might fit a function, you might fit a straight line, you might fit some function that has shape like this or you might fit a quadratic function something like a parabola, you might fit any function. Now the question is how to choose best parameters for fitting data.

So, in other words f of x; so what are these parameters? For example: so you can take example f of x is equal to a times x plus b then a and b are parameters. So, we call this linear fit and a and b are the parameters. Other example you can have f of x is equal to let us say you can fit into any function let us say you can fit into so x square plus x; so you can fit let us see into a x square plus b x plus c- this is a quadratic fit. And here there are three parameters: a b c. And you can go on, you can you can also do other functions. You can do a for example, f of x is equal to A, let me show the parameters in colors. So, a and you have e raise to lambda x. This is exponential fit.

Actually you can take any function, now exponential we have represented by two parameters, you can also have exponential, you can also have something like this f of x is equal to let us say I will not use colors. So, let us say a e to the lambda x plus b x plus c. So, you have a sum of exponential and a linear and you know you have many more here you have 4 parameters. So, the question is; how do you choose the best values of these parameters.

So for example if you choose one value of a b c you might get something like this you might take another value of a b c might get something like this, might take a third value might get something like this. So, what is the best value of these parameters to be chosen? So, each value of these parameters will give you a different ways in different line and how to choose the best values of these parameters. So, the answer is you use a least square fit procedure, so I will explain the procedure.

(Refer Slide Time: 17:36)



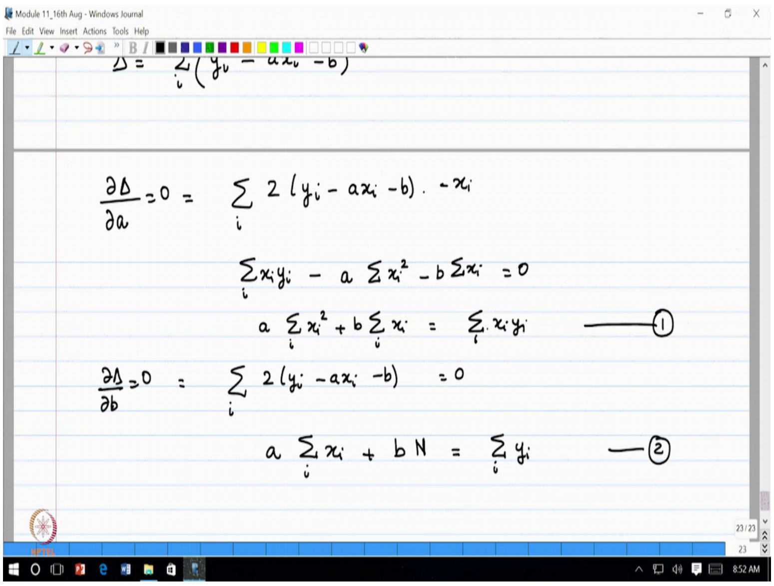
So, what you do is you define error as error in any given fit. That means, let me take the example f of x has two parameters a and b, ok. So, what you do is in any fit you fix these two parameters and you define the error. So, the error is; so y i is the value of y minus f of x i, ok. So, x i is the actual data; actual data point x i. So, what we have here x 1 and y 1.

So, what you are calculating is y 1 minus f of x 1. Now if it was a perfect fit then this difference should be 0, but what you will find is any fit you want exactly fit with that point there might be slide deviation. So, you define the error as this and sum over i; this i will call it delta, delta equal to. So, this is total square error of any fit. And notice that since f has these two parameters a and be delta depends on a and b, ok.

Now, the prescription choose a and b so that delta is minimized. So that is, dou delta by dou a at fixed b is equal to dou delta by dou b at fixed a equal to 0. So, that is the prescription for the choice. And basically this method when you set dou delta by dou a equal to 0 you will get; so this will give you two equations from which a and b can be determined. Because there are two equations, two unknowns a and b and a and b can be determined from this.

For example, if f of x equal to a x plus b then what you have is delta is equal to sum over i y i minus a x i minus b whole square. Now, what you can do is you can expand this sum over i y i square plus a square x i square well ok you do not actually expand it, we will just leave it like this. Now let us calculate.

(Refer Slide Time: 21:29)



Calculate dou delta by dou a equal to 0 equal to. Now if I want to take the derivative with respect to a, so the derivative will go inside this sum. So, I have sum over i. Now I take the derivative with respect to a of this, so I will get twice y i minus a x i minus b into. Now derivative of this with respect to a is minus x i, ok.

So, what this is gives you is the following, now if I multiply it out and I cancel the factor of two, so I get sum over i. X i y i I am taking the minus two factor, I am cancelling the minus two factor. So, sum over x i y i, and then I will write this explicitly so minus a sum over x i square minus b sum over x i equal to 0. So, I just wrote it as in this way. In other words you will get a times sum over i x i square plus b times sum over i x i is equal to sum over i x i y i.

Now if I do dou delta by dou b equal to 0 then what I will get is something very similar. So, I will get sum over i twice y i minus a x i minus b. And now I will just get a 1. So, derivative with respect to b is just 1; so this equal to 0.

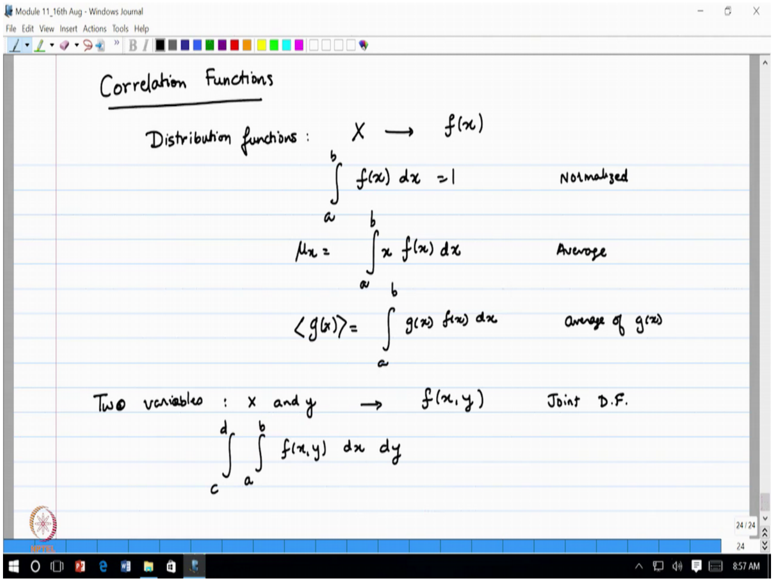
So, in other words I can write this as: so if I try to write in the same way I will get a times sum over (Refer Time: 23:28) we call this equation 1. So, what I get is a times sum over i x i plus b. Now you have sum over i of just b that is just N and this is equal to sum over i y i. So, these are the two equations we got, and we can solve for a and b. This is a simple this is just. So, you know the x i and the y i, so you know all these quantities; you know this, you know this, you know all these quantities and so you can solve for a and b.

So, this is the least square fit procedure that can be used to fit any arbitrary function in terms of choosing the right parameters for the fit. And this is something that is used regularly, and in fact it is so common that most of the plotting programs like x l Mathematica, MATLAB, etcetera, they have inbuilt feting functions which are based on least square. So, you can fit to a linear quadratic any polynomial or you can fir to any exponential, you can fit to trigonometric functions, you can fit to many different functions.

But what should be kept in mind is that you are trying to minimize square error. So, the assumption is that your data is distributed like a Gaussian random variable. So, that is an underlying assumption is also.

So, the last topic that I want to do as part of this module is that of correlation functions.

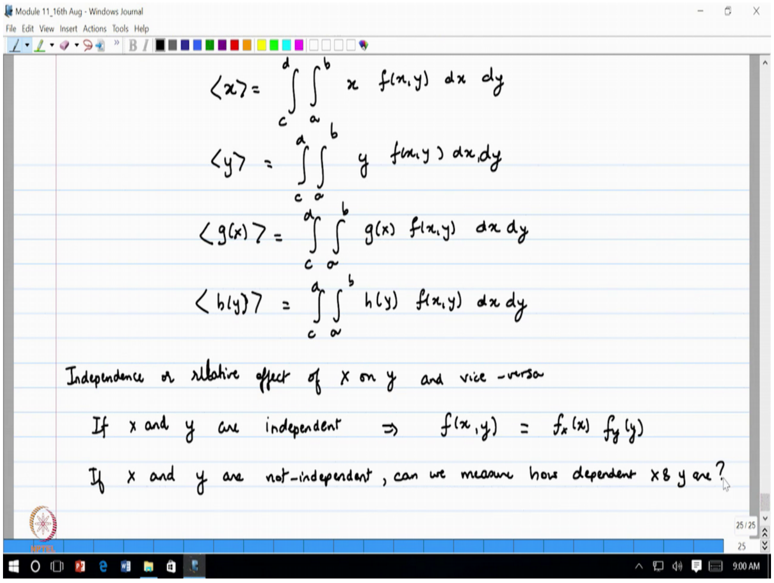
(Refer Slide Time: 25:16)



This is a little different from the error estimate, but I just wanted to mention this because it is something that is used very widely and in statistical mechanics. So, here what we said is that suppose you have; so go back to your distribution functions. So, you might have something like variable x and it might be having a distribution function f of x. And what you say is that integral f of x over whatever the range of a is let us say a to b dx equal to 1. And average value of x is integral a to b x times f of x d x; this is normalized, this is average. And you can take any arbitrary moment. So, I can take average of x raise to N as or average of any function of x g of x as integral a to b g of x f of x d x. So, this is average of g of x, ok.

Now, suppose you had two variables x and y and let us say they had joint distribution function f of x y. So, this is a joint distribution function then what you would expect that integral let us say I will say a to b f of x y dx dy and then we will have another integral c to d dx dy equal to 1 this is normalized.

(Refer Slide Time: 27:45)

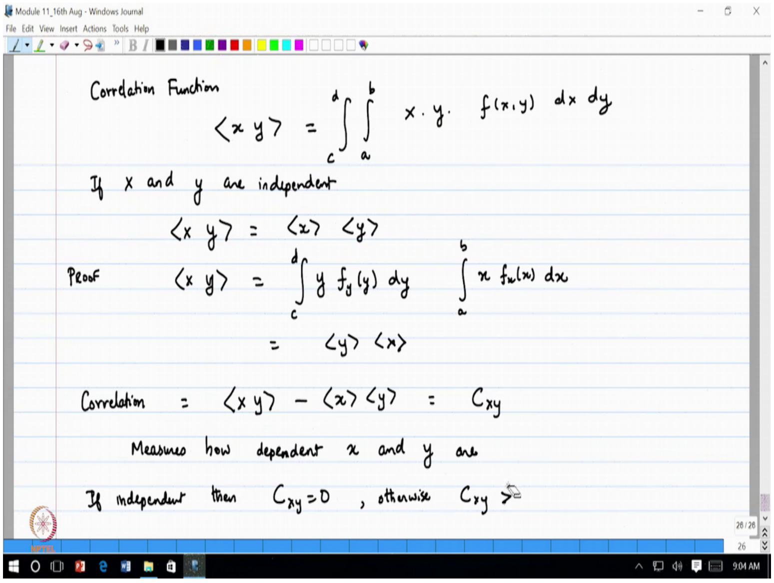


And you can ask average value of x is equal to integral c to d integral a to b x f of x y dx dy. Similarly, I can write average value of y as integral c to d, integral a to b, y f of x y dx dy. I can also write average value of any function so g of x is equal to integral c to d integral a to b g of x f of x y dx dy. Similarly I can write any function h of y; I can write average as integral c to d integral a to b h of y f of x y dx dy.

Now often your interested in looking at the independence or relative effect of x on y and vice versa. So, you want to know whether x and y are two independent variable or they are dependent variable. If x and y are independent, that implies f of x f of x y, this can be factored and if it is factored I will call it as f x of x f y of y. So, it factors into these two terms.

If x and y are not independent then you cannot have this. But the question is can we measure how dependent they are. So, can you get measure of how dependent these two variables are? This is where this is called the correlation function. So, the correlation functions is actually closely related to this, but let us motivate that.

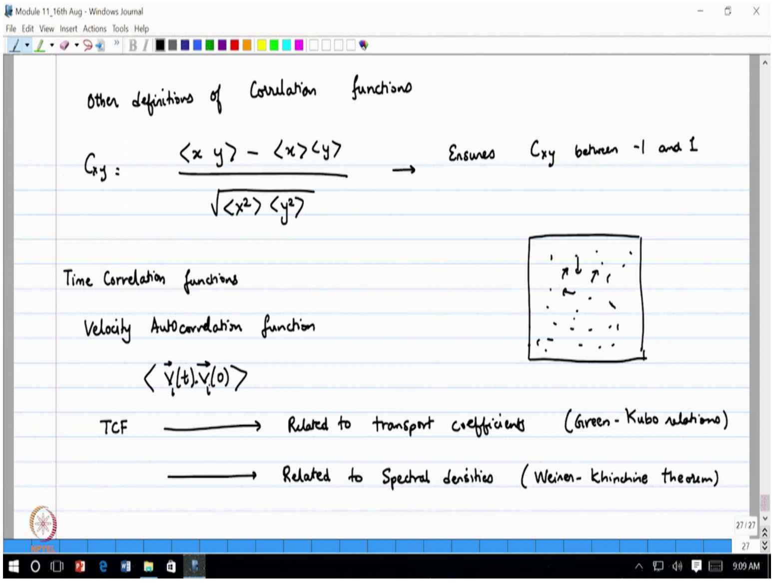
(Refer Slide Time: 30:21)



So, if x and y are independent then average value of x y can be written as average of x times average value of y. So, this is fairly straight forward to show, because average value of x y will be equal to; so the proof here integral c to dy f y of y dy and integral a to be x f x of x d x; so if they are independent and so this is average value of y times average value of x.

So, if they are independent it is this we can define correlation as average value of x y minus average value of x times average value of y, ok. So, this measures how dependent x and y are. So, if they are completely independent. So, if independent then correlation I will just call it c of x y, then c of x y equal to 0; otherwise c of x y not equal to 0. And you know this is one way to call the to estimate the correlation.

(Refer Slide Time: 32:24)



Sometimes other definitions are used, other definitions of correlation functions. So, sometimes just this is called the correlation functions, sometimes you take this and subtract the average that is called the correlation function, sometimes what is done is you define it as x times y minus x y divided by what is done is square root of x square y square. Sometime it is defined this way and so on.

So, this ensures that correlation function. So, this ensures c x y between 0 and 1 0 or other minus 1 and 1 and 1. So, there are several other definitions, but essentially this quantity in all these it is this average of x y; how different it is how average of x times average of y is the measure of this of the relation between two variables. So, this is very widely used in. So, one example is time correlation functions. For example, you might ask; suppose you have a collection of several part particles in a box suppose you have a box with lot of particles and many particles and they are like in a typical liquid and all these particles are moving with some velocity. Then you can ask; so velocity auto correlation functions. So, you take the velocity at time t of any particle and you take the velocity at time 0 and when you take their product and averages. So, this is velocity of any particle, so I can put an I if I want. So, this is the velocity of any particle and if it is a vector you can dot it the vectors. So, this is called a velocity auto correlation function and this is a very well known quantity, ok.

Now, time correlation functions are really important is for various reasons: one is that the time correlation functions I will write them as TCF these are related to transport coefficients. This is what is called Green-Kubo relations. And also time correlation functions are related to spectral densities this is what is called Weiner-Khinchine theorem.

Actually there are several time correlation functions, not just velocity you can look at time correlation functions of position of density and so on. And these time correlation functions are something that you will see very often when you are actually doing research in statistical mechanics. So, they are in fact one of the foundations of non equilibrium statistics mechanics. In any case the point is you know correlation functions are ways; correlations are ways to estimate the relation between variables. So, they are generally useful. And in particular time correlation functions can give lot of interesting information.

So, I will conclude the lecture part of module 11 here. So, in the next lecture we will recap all the things that you have learnt and then we will do a few practices problems.

Thank you.