Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

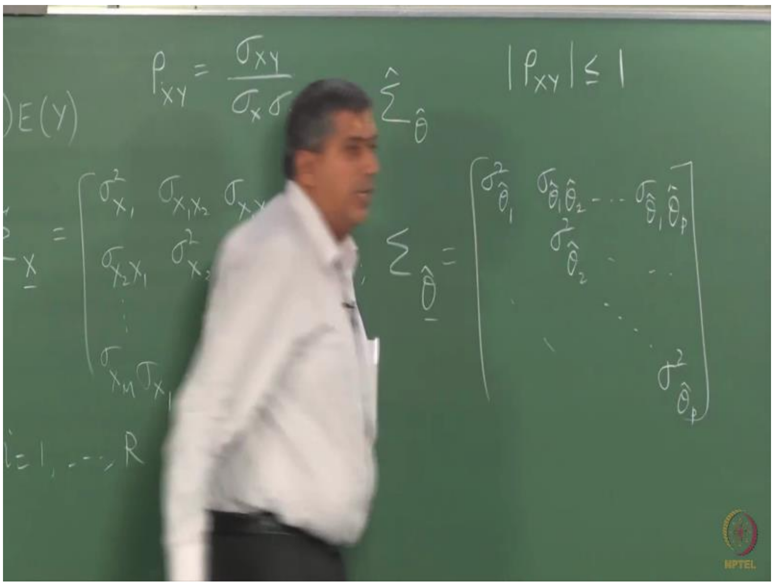
Lecture - 15
Lecture 07B - Probability and Statistics Review (Part 2)-5

So the correlation is denoted as rho is Greek symbol which is as I said standardized covariance; the numerator containing covariance and the denominators containing the product of standard deviations. I must not have mentioned this earlier, but standard deviations are nothing but the positive square root of the variance and you are all familiar with that. Straightaway tells me when covariance is 0 correlations is 0 and that mapping is established.

Of course, we have introduced this measure we need to be assured that it is a bounded measure. And it turns out that this correlation is bounded above in magnitude by 1 which is the thing that we are seeking it is. Obvious from this definition that we have addressed the first issue which is sensitivity to the choice of units; that is no issue.

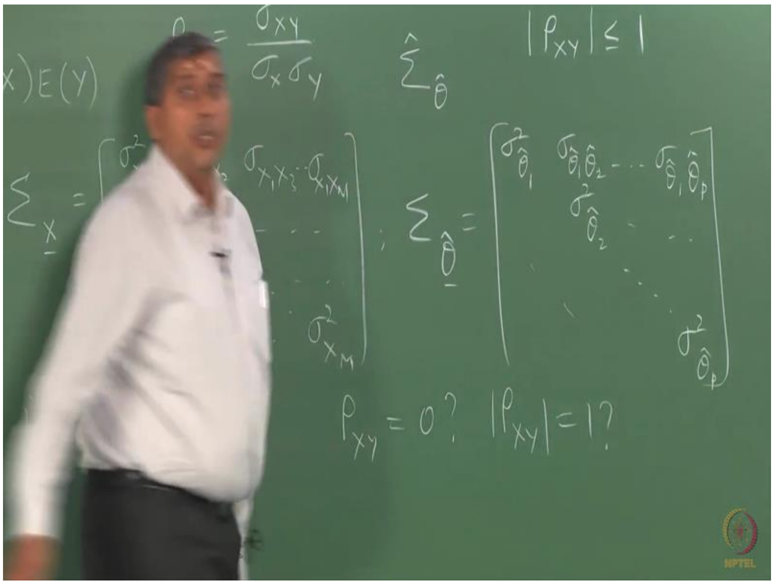
The second issue is a boundedness which is also assured and I am not going to go over a proof of this, you can prove this using Chebyshev's inequality you can find this almost everywhere on the net. So, I am going to skip that.

(Refer Slide Time: 01:26)



But it is a very very important thing to remember particularly when you are answering certain questions in your exam do not give me absurd answers for correlation like 300 255.7 and so on if you do. So, please say that you have redefined your correlation and please spell the definition. So, correlation as we have defined cannot exceed magnitude of 1 in under any circumstances.

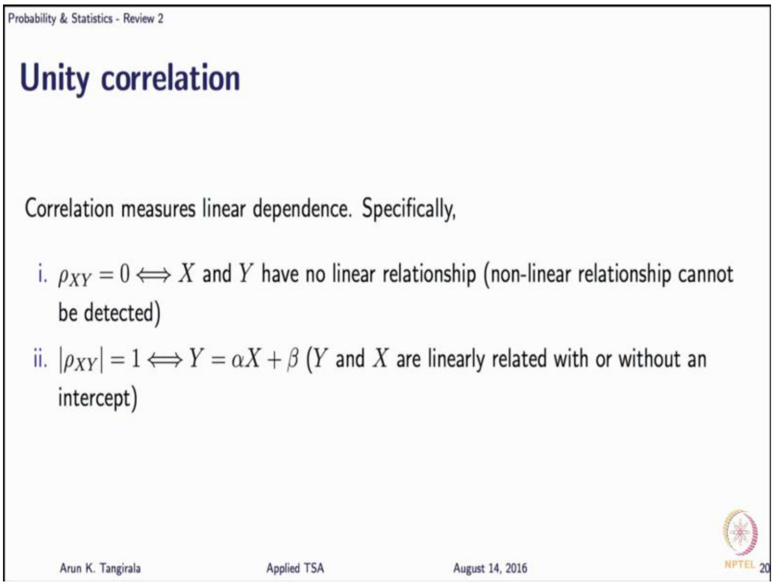
(Refer Slide Time: 01:58)



Now, there are two things that we want to ask that is what does this mean right and likewise what does unity correlation. And then of course, the intermediate values. We have already said that when there is no covariance, I mean though no covariance or no correlation there is no linear dependence. We will be interested in this part now which will kind of help us partially answer or at least convince ourselves of this answer that we have given to the first question.

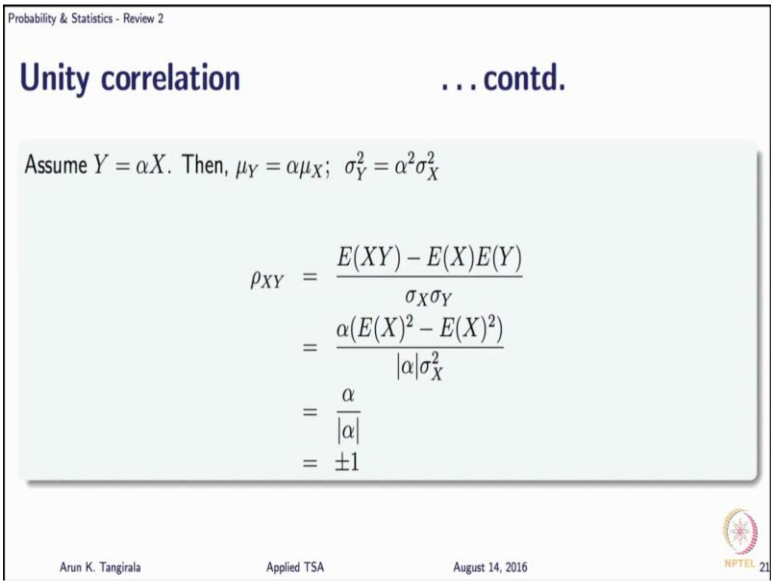
So, let us look at the case where correlation hits a value of 1. And it hits a value of 1 when y and x are perfectly linearly related, well with a slight relaxation on that a fine relation. When y is alpha x or you can say y is alpha x plus beta or it can be the other way around to it could be x equals alpha y does not matter; remember correlation or covariance does not worry about the direction of relation.

(Refer Slide Time: 03:12)



So, let us look at this case of y equals alpha x plus beta with some abuse of terminology let us call this linear, although it is not strictly linear.

(Refer Slide Time: 03:22)

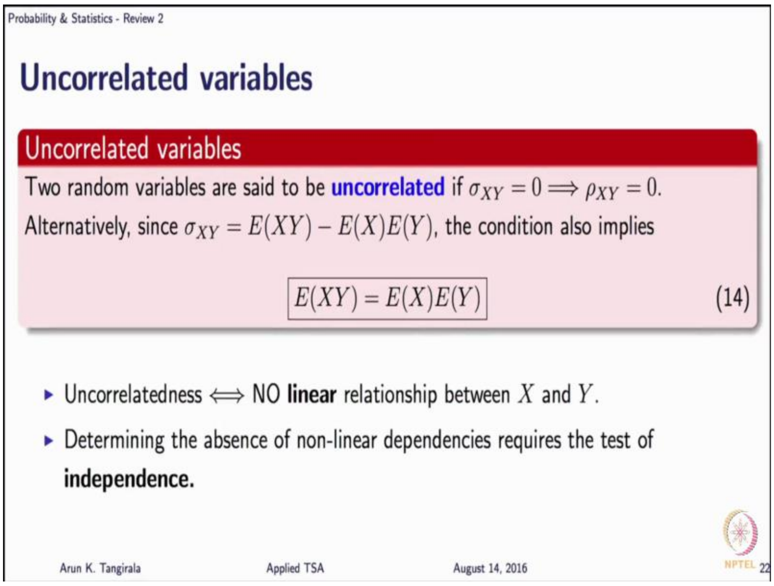


And this is something that you must have also seen in the NPTEL course, but let us if you just quickly go over it and define and derive the value of correlation for this kind of a situation that is y equals alpha x plus beta. Let us assume x to be what I have done here is have ignored beta, because you can also actually prove when beta is not 0. For simplicity let beta be 0 then mu y is alpha mu x and sigma square y is alpha square sigma square x. Remember we have talked about the scaled random variables, how do the mean and variance scale we use those results here and then you plug in all the expressions therein. And straight away you see that correlation takes on a value of plus or minus 1, depending on the sin of alpha. And negative alpha would mean that as x increases y would decrease and a positive would mean that x and y actually vary in the same direction.

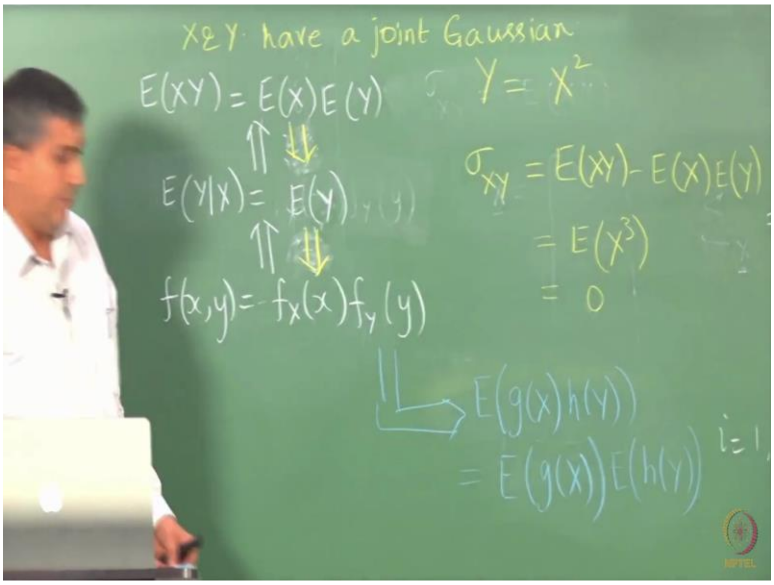
So, it is clear that 1 y equals alpha x or even alpha x plus beta you can show this for alpha x plus beta as well that correlation hits a value of unity in magnitude. So, now I know that whenever I see correlation value of 1 y and x are linearly related. But the bad news is that you will never be able to see this value of correlation hitting one in practice. Even if y and x you know were perfectly related you would have noise or some other disturbances coming in which will take down the value of correlation below unity; how much the correlation will drop below unity depends on the extent of uncertainties and so on. But at least I know now that when correlation values are very close to unity; that means, the linear model will do a very good job of predicting one variable given the other. So, that is the interpretation.

What about the other way around, can I show this that this is an if and only if condition; yes, but will not go into that. Now, very soon we will also look at the consequences of correlation that is under what situations correlation values take on values between 0 and 1 in magnitude. But before we do that let us step back and now put in perspective the notion of correlation and independence.

(Refer Slide Time: 06:07)



(Refer Slide Time: 06:14)



When two variables are uncorrelated as you see on the slide; and we have seen this in the last class as well when sigma x y 0 or even when rho is 0 expectation of x y is a product of expectations. And when two variables are independent then you have the joint density being factorizable into product of marginals. And there was one more condition that we talked about when it comes to when it came to dependence between x and y which was based on conditional expectations. And that was; so let me write this here slightly between, so this is the notion of independence. And there lies an intermediate result which is that of expectations and conditional expectations.

So, here you have uncorrelatedness and then here you have whatever in between when there is no name to this you can say that it is conditional expectation is the same as unconditional one I think that made a mistake here. You should correct me feel free, sometimes mistakes are deliberate sometimes it is just the flow. And then you have independence here. So, the independence is the strongest one and if you look at the direction of implications independence implies this which in turn implies; that means, your independence condition implies both this. Remember, what we mean by independence is absolutely no relation between x and y, you cannot fit a model so forget one of the other variables for prediction. You just use individual histories to make your predictions

Now, this traffic is in general not there and this traffic is also in general, that also does not exist. So, it is one way implication that which means when two variables are uncorrelated for example, it does not necessarily means they are independent. And likewise if two variables are uncorrelated it does not necessarily mean that the conditional expectations are identical.

So, if you are able to show this then you do not have to show any of these. From a modeling viewpoint what this means is; if there is no relation; that means no model can be built linear or non-linear does not matter. But, when you have determined that no linear model will work it does not mean that non-linear model also do not work. Let me give you a very simple example that will hopefully convince you further. So, let us say here I have y equals x square, you do not want to have white noise all the time some colour noise will help. So, y equals x square, so this is the relation and let us say and let us assume that x has a Gaussian distribution, let us assume that you know it is a 0 mean just for simplicity.

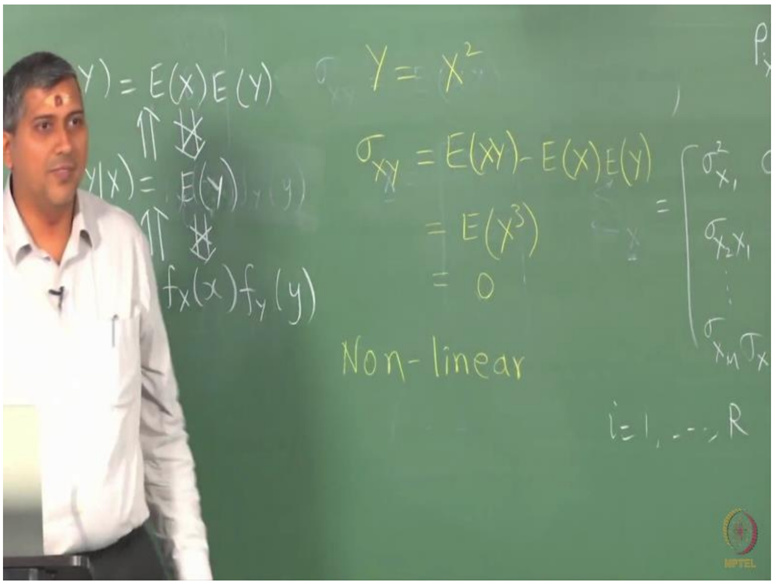
Now, is this a linear relation or a non-linear relation, right? Now if I want to compute the covariance between x and y what would be the answer, how would I do it and what is the answer. You are given that x is 0 mean for convenience do not assume variance of x is 0 what would be the covariance, how do you compute it? X is a Gaussian distributed variable you cannot compute.

Student: (Refer Time: 10:31).

Probably be able to give answer next week.

So, I emphasized an important point x is a Gaussian distributed random variable. When you are in confusion I think we used to follow this tricks even our school days. I do not know anything a write the formula, some of the students I still see. If I ask a question all the theory that comes to the mind is written on answer sheet. Basically the instructor or the examiner is being asked to take whatever is relevant and apply to the problem and get the answer. The student is demonstrating the ability to remember the associated theory. Which is not a bad idea, but we would like to see application of that theory. No, seriously what do you do in times of crisis you will just throw out whatever theory you know and say you know take it or leave it this is what I know. Especially as you mature and you are getting closer to your under graduation you become more and more bold and practical.

(Refer Slide Time: 11:44)



So, expectation of x y, we just start from that expression right minus product of expectations. Now can you proceed further, what happened? What is it?

Student: (Refer Time: 12:14).

Why is e of x cube 0 where is e of x cube coming from? From the first term right. Second term is 0, I have said assume mean to be 0 that is why I said let us keep it simple. X is a Gaussian distributed random variable with 0 mean. And what you are looking at is a third moment of x and skewness. This is called skewness. Skewness is a measure of the symmetry of the pdf. And Gaussian distribution is a symmetric one; I mean pdf is a symmetric one. Therefore, you can say the skewness is 0, so the answer is 0. You would also always like the answers to be 0 not the marks. And this is a good thing here, what is it tell us about the relationship between y and x?

Student: (Refer Time: 13:07).

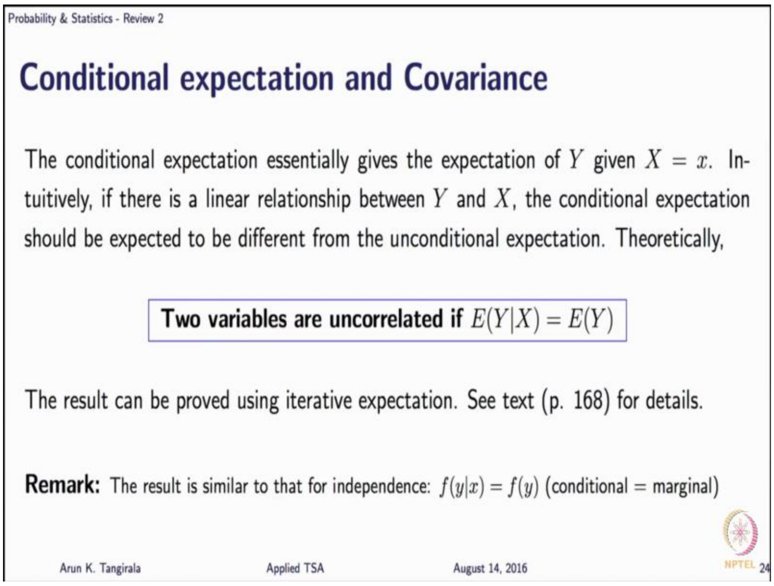
No linear relationship right, but if you add it, so it is nice what you have said is nice no linear relationship, but the reality it is non-linear. Do you see the non-linear relationship y equals x square is a non-linear one. So, just because two variables are uncorrelated it does not mean that there is no non-linear relationship. And do not you take this example a bit further you say no always it means quadratic. So, it is very simple example to drive home this point.

Now the good news which is always got to do with Gaussian; the good news is that if x and y have a joint Gaussian distribution then this traffic opens up, otherwise there is a stop sign. The moment it sees that x and y are joint Gaussian the gates open, and you can prove this theoretically. If x and y have a joint Gaussian distribution I can write that then these gates open up one implies the other. That is why again you can see why Gaussian distributions are very popular, they have some very very nice properties. Of course, here it is a joint Gaussian that we are looking here.

And there is yet another thing to know about this independence, which is when x and y are independent we have written expectation of x times y as product of expectation, but there is yet another implication which is that expectation of product of functions of x and y are the product of the expectation of the respective functions. This is also true

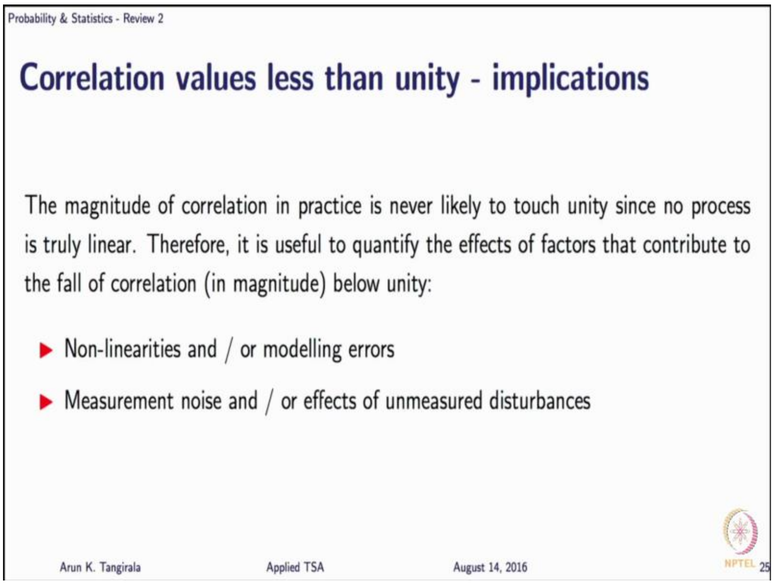
Now, that is fairly easy to see intuitively. So remember, when imagine that g of x is some x cube and h of y is y square; you are looking at some joint fifth moment of x and y. And once you write the joint fifth moment the theoretical expression it will involve the joint pdf and then because they are independent you can express a joint pdf as a product of the marginals then you can separate the double integral that you get for the joint fifth moment. And you can then see that each of these integrals is expectation of g of x and the other one is expectation of h of y. These are some very useful properties that are used in deriving a lot of theoretical results; any questions on independence, uncorrelatedness, anything.

(Refer Slide Time: 16:35)



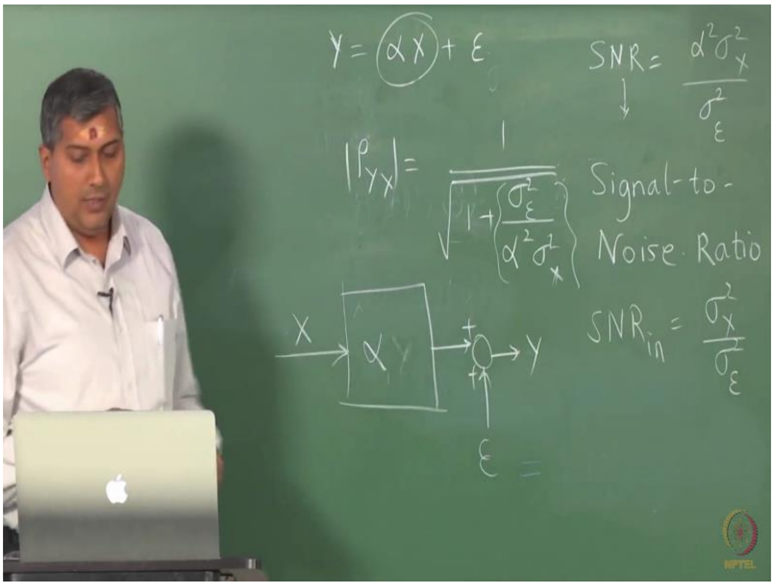
Of course, we will skip this we have already said two variables are uncorrelated if the conditional expectations are identical. And the proof of this is given in my textbook I have given the page number; will therefore go past in fact the proof uses the iterative expectation result that I have given earlier.

(Refer Slide Time: 16:46)



So, let us return now to this question of what do I take when I get correlation values between 0 and 1. Specifically, what kind of situations actually gives me correlations between 0 and 1? And to do this we will again go back to this model that we had earlier.

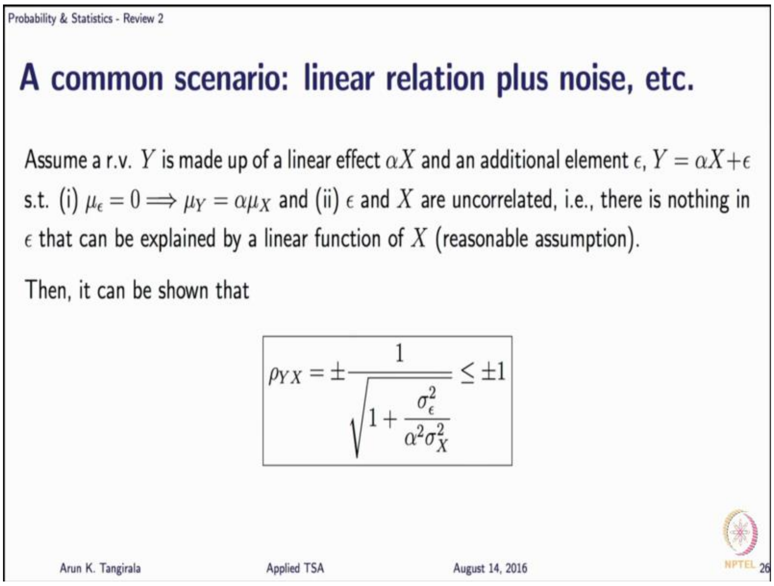
(Refer Slide Time: 17:25)



Remember earlier we wrote y equals alpha x plus beta, but now we shall write y equals alpha x plus some epsilon. Why are we writing it this way? Because we want to say that there is something in y that a linear function of x cannot explain; that epsilon signifies whatever your alpha x which is a linear function of x cannot explain, you can have a beta but that does not matter.

When do I run into these kinds of situations, whenever I have let us say measurement noise for y let us say why is a measured quantity, x is some random variable, y is a measured variable. Then epsilon could contain measurement noise or epsilon could contain effects of unmeasured disturbances that are also that are contributing to y. And on top of all of this epsilon could also contain modeling errors. What we mean by modeling errors is that maybe the true relationship between y and x is non-linear and I have only included the linear part. So, epsilon is like consolidated whatever affects of sensor noise unmeasured disturbances and modeling errors and so on; all of that is lumped into epsilon. And when this is the case you can actually show that the correlation between y and x is what I have given on the slide.

(Refer Slide Time: 18:52)



Which is one over, well let me write the magnitude square root of 1 plus sigma square epsilon by alpha square sigma square x.

Typically, I derive this result it is not difficult at all in fact, I think it is given in the text you can go and look up the text or you can derive it yourself. But let us look at the result here and try to understand what is happening. So, what you see in the denominator you see sigma square epsilon over alpha square sigma square x, of course you have 1 here.

Now, this ratio here is a very important quantity that appears in all parameter estimation problems or all signal analysis problems and so on known as the inverse of the signal to noise ratio. The signal to noise ratio is defined as alpha square sigma square x by sigma square epsilon. What is the signal that we are referring to here?

Student: (Refer Time: 20:06).

Really, as you said the signal is y. And what do we mean by y here? Y is actually alpha x plus epsilon. So, how can I signal be y?

Student: (Refer Time: 20:24).

Correct. So, what is the signal then? So, here this is called the signal to noise ratio which estimated sigma.

Student: (Refer Time: 20:49)

Alpha x part, why do we.

Student: (Refer Time: 21:02).

That is fine, but what is the signal that we are referring to somebody said alpha x y.

Student: (Refer Time: 21:12).

Ok.

Student: (Refer Time: 21:32).

Ok.

Student: (Refer Time: 21:41).

So that is a separate question we will answer that. The first part is a lot of you have gotten it correctly but you are looking at from your own application viewpoint. The signal that we are talking of in this definition is alpha x. You can essentially, why this is called signal to noise ratio is many in many applications I am interested in whatever signal is here it is a linear function, but I am interested in some truth; in knowing the truth, but typically truth comes in a corrupted form when we measure the truth. And that corruption here you can call it as epsilon, but it need not be corruption all the time as I said epsilon also contains modeling errors. But keeping that aside when you are looking when you are standing at the measurement side you would like to hear alpha x, but you are hearing alpha x plus epsilon.

So, your ability to discover this linear relationship between y and x clearly depends on this ratio. Why, because if this ratio is very high then what happens to correlation, it nears one. In fact, as this ratio hits infinity the correlation will hit one. In the limit as s n r goes to infinity the correlation goes to 1. And when correlation hits one then you have a perfect linear relationship you will be able to estimate that linear part very well.

Although we introduced s n r in the context of linear models s n r in itself is a generic concept. It is a concept that tells you how much power if you talk in terms of signal analysis how much power is contained in the truth versus uncertainty. Epsilon is your uncertain part sigma square epsilon is measure of the uncertainty. Remember we said variances are measures of uncertainty so that also answers to your answers your question as to why we do not have mu and why we have sigma squares.

Because mu’s are not measures of uncertainties; mu is only a measure of the center the measure of uncertainty per measure there are many measures a measure of uncertainty is variance. Remember if variance is 0 then there is no uncertainty. That is one of the reasons why that you have variances here. But this was not really defined up front and then it made it is way, s n r was more of a discovery in hindsight. That is people started asking what happens to my parameter estimate, how good is my parameter estimate and so on.

When you started answering those questions they started to see the appearance of this ratio. In fact, later on and estimation theory will show that the signal to noise ratio affects the precision of your estimate, how precisely you can affect you can estimate the parameter of interest. If s n r is very high then you can estimate the parameter in a more precise manner; higher the s n r better the precision.

Here higher the s n r closer the correlation to unity and your ability to fit a linear model or you can say the ability of a linear model to do a good job of predictions. There are several different ways of looking at it. So, s n r was introduced more in hindsight rather than up front. You can define as a ratio of mu of y or mu of alpha times mu x by mu epsilon, but it would be useless and also mu epsilon can be 0 remember. So, that will then become an ill defined quantity.

So, this is something that we will revisit in the parameter estimation discussion as well, but you should keep this in mind that always you will have the signal to noise ratio coming into play in every stage of your data analysis, right from your correlation to your parameter estimate. For example, if I give you y that is observations of y and observations of x and I asked you to estimate alpha we can show later on that is the precision of alpha or the variance in alpha is again related to this quantity here. And by the way this is specifically called the output signal to noise ratio, it called s n r out. You can also define s n r in; so the s n r in would be simply sigma square x over sigma square epsilon.

Now, this is based on the idea or this imagination that x is going into some process and producing alpha x here and here is your epsilon adding up to produce y. Essentially, x is going through a gain pure gain system and adding on I mean the uncertainty adds on to alpha x then you have a measurement of y. So, if you think of x as input and y as output that ratio is called the output signal to noise ratio and this ratio here is known as the input signal to noise ratio. The input signal to noise ratio is also an important quantity.

By now you must be actually kind of slowly getting into the groove that correlation has a very strong connection with linearity. And we have already shown when y and x share a linear relationship, perfect linear relationship then the correlation hits 1. When there is something more apart from a linear that is from a linear function of x then the correlation will dip. To the extent to which it will dip depends on how much is not being explained by this linear function of x. The extreme cases when alpha is 0 and purely why is epsilon then the correlation between y and x will be 0. As you can see because this will hit a value of infinity when alpha is 0 and as a result s n r will go to 0.

So, in practice now generally we think that by looking at correlation I can conclude whether the relationship is non-linear or not. Unfortunately that is not true, when correlation is below one as this example suggests it is not possible to ascertain whether the dip in the correlation is because of a non-linear relationship or because of noise or because of disturbances we do not know. All we know is that there is some epsilon, besides alpha x and that is all I can. If I want to know whether a non-linear relationship exists between y and x then I have to conduct tests of non-linearity, and that we will not get into.