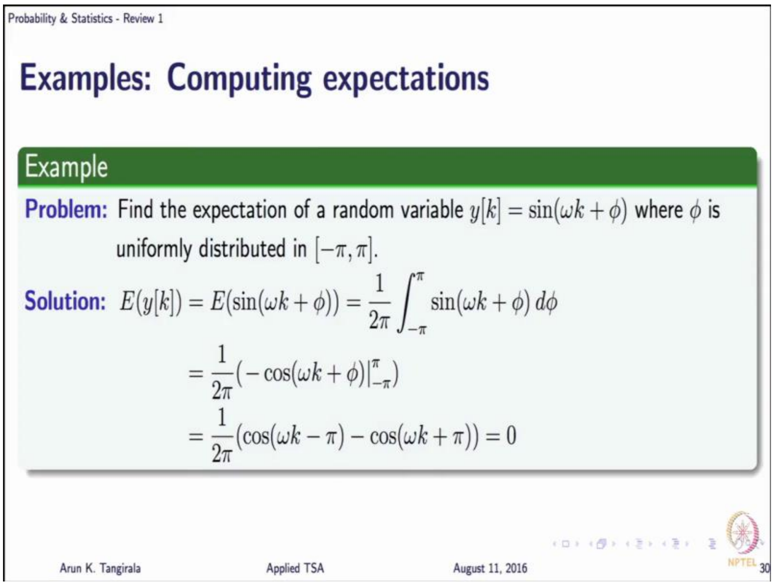
**Applied Time-Series Analysis**
**Prof. Arun K. Tangirala**
**Department of Chemical Engineering**
**Indian Institute of Technology, Madras**

**Lecture – 10**
**Lecture 05B - Probability and Statistics Review 1D (Part 1)-4**

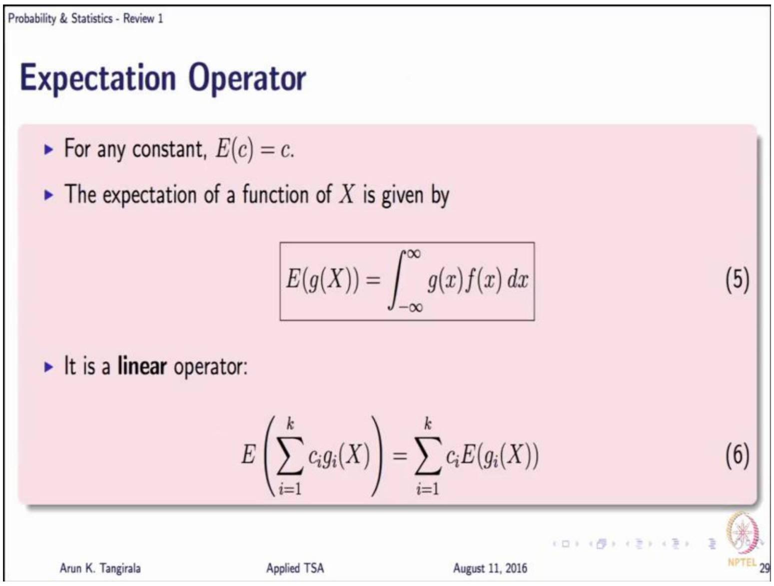
(Refer Slide Time: 00:13)




Let me go through this simple example to make sure that you understand the notion of a mean or the expectation very well that you know how to compute theoretically expectation. Now we would like to compute the expectation of a random variable. Although I say random variable, I have time instant here, but you know why we have the time instant, we are looking at the kth instant of a random signal. So, at any instant we say that the random signal is a random variable and therefore, we have here y k equals sin omega k plus phi.

As I explained yesterday you should first find out what is the source of randomness, I am saying that this is a random variable, in this case the source of randomness is a phase and the distribution or sorry the density function or the distribution of phi is also given. The randomness in phi propagates to y and that is what makes why random there is nothing random about the amplitude of the sinusoid, nor the frequency and there is nothing random about time here alright. So, as I show you in the solution there the expectation of y k you use a formula always follow a systematic procedure you will the chances of going wrong are very low.

(Refer Slide Time: 01:36)




What I do is here I use this earlier expression in equation 5, expectation of g of X, what is X in this example that we are looking at?

Student: (Refer Time: 01:45).

That is a phase phi, phi is the random variable, the function is a sinusoid and I simply plug in this formula as much as I hate to use the term I simply plug in that integral plug in the g of X and f of X into the integral in the equation 5, what is f of X?

Student: (Refer Time: 02:06).

1 by 2 pi because it is uniformly dispute that is the loveliest and cutest distribution that you can ever have really, but in certain situations no, you would prefer to have a Gaussian 1.

Yes, go through the simple integral, I will not test your integrating abilities and therefore, I have done the solution for you, it turns out to be 0. Now this value of 0 that has come out of the solution you may think it was so obvious because the average of a sine wave is 0 correct, but the average of the sine sinusoid being 0 is something that we are thinking along in time, here we are not integrating in time, if you notice we are integrating across the outcome space of phi. It is true that the average of sin in time is 0. So, to understand that when you go back, change this example slightly or problem slightly definition that this phi is no longer uniformly distributed in minus pi to pi, but rather 0 to pi, alright. So, phi has a uniform distribution in 0 to pi and then work out this solution find out what is the average of this y of k, intuitively what do you expect? If the distribution is no longer in minus pi to pi, but 0 to pi still uniformly distributed what would you get? Will the answer change? So, what would be the answer?

Student: (Refer Time: 03:47).

1 minus?

Student: (Refer Time: 03:50).

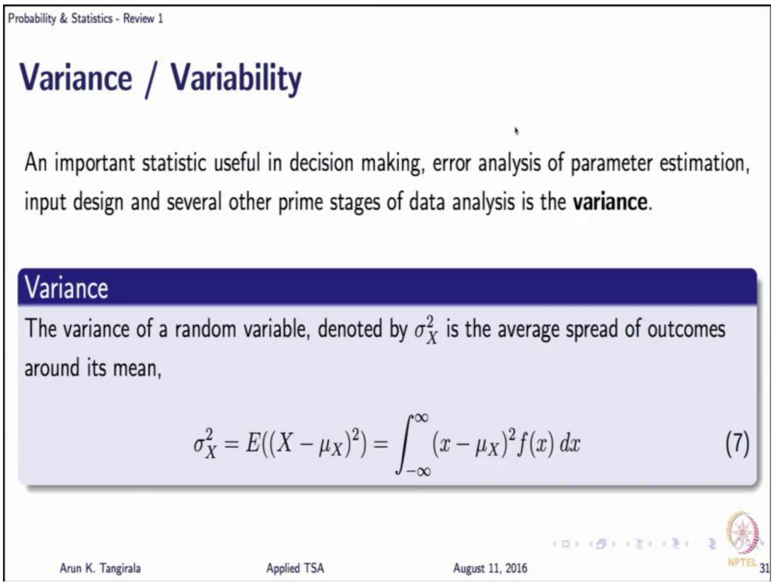
Let me ask you, do you expect a nonzero answer or not.

Student: (Refer Time: 03:57).

Yes. So, that is different from your time average of the sin wave.

The example here is to try and tell you again reinforce that time averages are different from on some lay averages. So, be careful if you actually follow things systematically, you will not go wrong, first find out what is the source of randomness and start from there and compute your expectation. We will revisit this example when we discuss the notion of stationarity for now will go past and talk about the second moment.

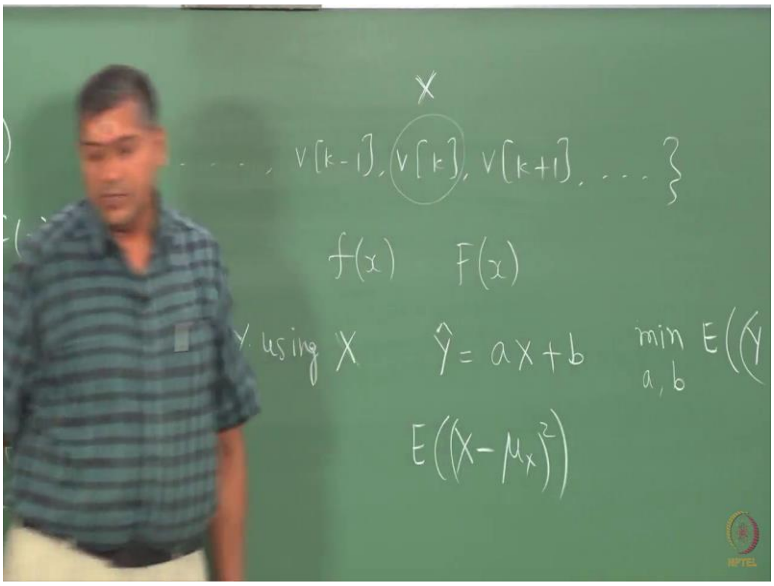
(Refer Slide Time: 04:31)



So, what we understand is the first moment is ironically, it is also incidentally, it is also the first thing that you discuss and those are the is a most important moment I should say that is useful in many decision making exercises.

But there is also an equally and sometimes people would argue that more perhaps more important is this variance which is the second central moment. So, when I am visiting a new city I am not only interested in knowing the average temperature, but I better know the range of temperatures that I expect to see because that is going to really help me in my decision making. And once again you see here what we define is they say the measure as the second central moment.

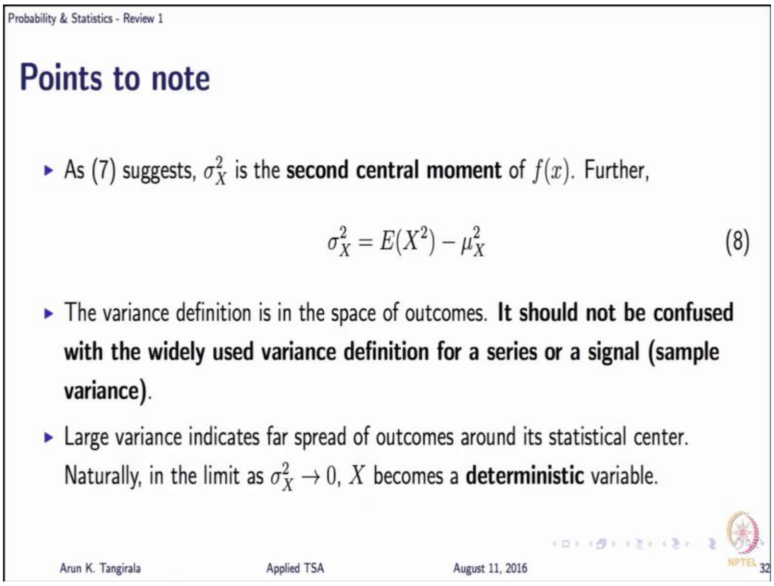
(Refer Slide Time: 05:30)




If you look at the definition carefully we have expectation of X minus mu X to the whole square if you look at the definition carefully, what it is actually giving you is the average square distance of the outcome from the center and therefore, it is a measure of the spread of outcomes how far your outcomes are spread about the center the farther they are spread the larger is a variance or variability as it is also known. And for the deterministic variable, what would be the variance 0 because there is only one outcome and that is about it. So, the entire expression works out to be 0, this is called a second central moment and once again we use the expression that we saw in 5 to arrive at the expression for the variance.

And again you can give the same interpretations. So, but you now start to see that the expectation operator appears in almost every sphere of statistical data analysis. You can also rewrite this variance in a different way sometimes this expression is useful.

(Refer Slide Time: 06:43)

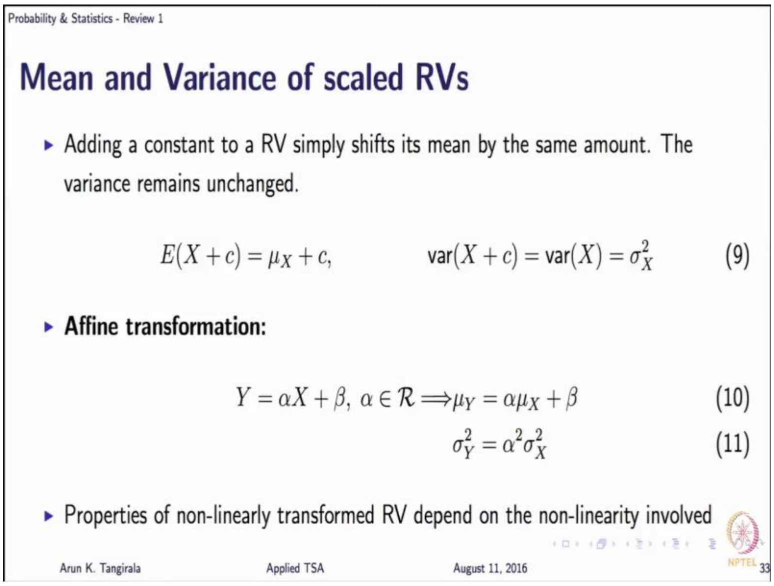



What you see in equation 8? This expectation of X square minus mu square expectation of X square is simply the second moment. So, the second central moment is a difference between the second moment and the square of the first moment and once again I would like to say that do not confuse this definition of variance with the sample variance definition that is used for signals which is computed across time, this variance is also computed across the outcome space and this is sometimes people call this as a population variance earlier one as apartment the first moment as a population mean.

But remember it is very simple to remember this as theoretical definitions all we are learning right now is theory, we will later on learn how to estimate mean how to estimate variance and so on, this is the time has not yet come. Generally variance is also used as in fairly crude sense as a measure of randomness that is how much uncertainty is that as a measure of uncertainty or a quantification of uncertainty larger the variability you say well larger the variances randomness is because variance is a measure of the spread of outcomes and because if variance is large; that means, outcomes are spread far away from the center and therefore, more say in some literal sense the uncertainty.

Very often we do perform some operations on random variables and therefore, it is important to know how when these operations are performed and variance for example, scale.

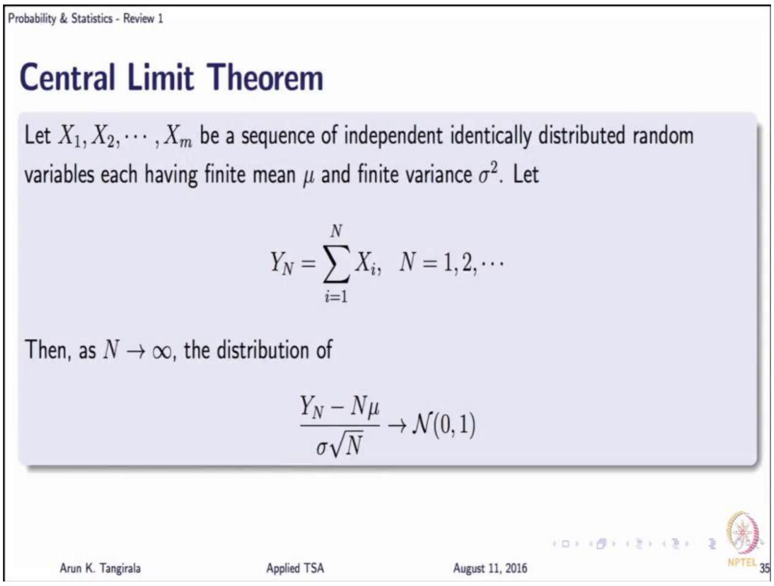
(Refer Slide Time: 08:22)




For example I may add some constant to a random variable, in that case the mean is shifted by that amount, but the variance remains unperturbed; obviously, because variance is always a measure with reference to the center and you can also have affine transformations such as if I am given a random variable X I may construct a new random variable y which is alpha X plus beta. In that case again the mean is shifted appropriately, but the variance is now alpha square sigma square X this is something that you will use from time to time in different theoretical exercises.

What happens when X is nonlinearly transformed well it depends on the kind of non-linear transformation that you induce? So much about the mean and variance any questions on mean and variance from the other hall, now before we move along it may be good to know certain properties of Gaussian distributed random variables, but I will and have postponed this discussion to a bit later on because it requires the understanding of what is known as a correlation and then uncorrelated normal variables and so on. We will come back to this slide when the time comes.

(Refer Slide Time: 10:01)




The last thing that I want to talk about in the review of the univariate random variables although occasionally we have talked about on the random variable y is this very familiar central limit theorem which basically tells us, there are various versions of central limit theorem, if you look at the original version of central limit theorem that came about in 60s. It says essentially if you take a bunch of random variables and you simply add them up in some proportion or in fact, in the purest in the simplest case you are simply add adding them up no weighted summation and you construct a new random variable y which comes out of this summation. Then the distribution of this resulting random variable tends to be Gaussian as you add more and more random variables into this summation, and as you can see the distribution is mentioned in the form of a standardized, the result is stated in the form of in the standardized form now the beauty of this result is why which is the sum of the random variables has a Gaussian distribution regardless of the distribution of X.

I am not stating this because I have to mention the central limit theorem, you will see the applications of central limit theorem later on for example, in deriving the distribution of what we call as sample mean. So, you can quickly draw some from your memory you can recall that the sample mean is also a sum of observations divided by of course, the number of observations the 1 over n does not appear here in the expression for y, but that is the only referring factor otherwise the sample mean and what you see as y look very much similar. When we enter the world of estimation theory we will have at some point in time talk about distributions of estimates at that point in time the central limit theorem comes handy in telling us how the sample mean is distributed at the moment will not get into that.

In general when you are dealing with linear estimators the central limit theorem comes very handy in deriving the distributions of those estimates at the moment will not worry about it, but it is also useful in a general sense that you if you consider a random variable any random variable let us say you are looking at the stock market index although it is a not a correct example there are a number of factors that affect the stock market index. If you assume that those factors add up linearly then as the number of factors increase for that stock market index or any other random variable then you can expect that random variable to follow a Gaussian distribution.

This is also one of the reasons why in the absence of any a priori knowledge nobody has told me what distribution it follows generally the Gaussian distribution is assumed because you assume that maybe that random variable is actually may be constructed as a sum of many many random variables which may not be the case, but given no knowledge it is your free to assume anything. So, I assume the one that simplifies your situation. So much about the CLT, we will revisit the CLT later on when we talk about distributions of estimates with this we closer review, do not close your notebooks, we will now get into the world of bivariate analysis everything is ok, yes.

Student: (Refer Time: 13:35) the random variables that we are dealing with (Refer Time: 13:40) we have the same mean and variables.

Definitely, I mean I have given a very nice version of central limiting; there are results that that deal with deviations from this statement. So, in this we assume all this excess are coming from identical it is a good point that you brought up because I did not emphasize an important aspect of CLT which is the independence. The factors of the random variables that you are adding up should be independent, not only having identical distribution, but also independent - independence is something that we have not talked about we are going to talk about it shortly. But, to answer your question whether, what happens when these random variables have different means.

Even in some weak, sense the resulting random variable will have a Gaussian distribution, but you can still calculate the average of y.

(Refer Slide Time: 14:39)

For example, if you say Y is expression that we had is let me put the subscript n you are adding up n random variables. Even when these Xs have different means you can simply say that expectation of Y N is nothing but sigma of expectation of X n, you can still calculate the mean, the concern should be whether still Y will follow Gaussian distribution as n goes to infinity. Yes, in some loose ends yes, but the number of observations that are required to make Y reach the distribution of Y reach a Gaussian 1 maybe different from the case where the number of factors that you require to make why go to a Gaussian distribution when all of them have the same means. So, we call this as a convergence rate.

How quickly does a distribution of Y converge to a Gaussian and there are lot of studies on that. What are the impacts of the deviations from the central limit theorem on the convergence of the distribution of Y or convergence of Y in distributional sense, how quickly does Y tend to be Gaussian? As a simple example, you can try this out when you go back in (Refer Time: 16:02), you can generate for example, you can draw X from a uniform distribution may be about 5 observations add them up.

And repeat this because you will have to consider the distribution of Y, repeat this exercise, put it through a kind of a Monte Carlo simulation, you will find that maybe the moment you add about 5 random variables from uniform distribution, Y is also already showing Gaussian shapes, but you draw Y from X from some other distribution. Let us say chi-square or maybe in some beer distribution it may take more number of random variables to help y converge to Gaussian in distribution.

Student: What will be a random 2 kinds of distributions?

That also you can study so there are lot of results that are available, but there are then additional restrictions constraints and so on, we do not generally go into that that is usually.

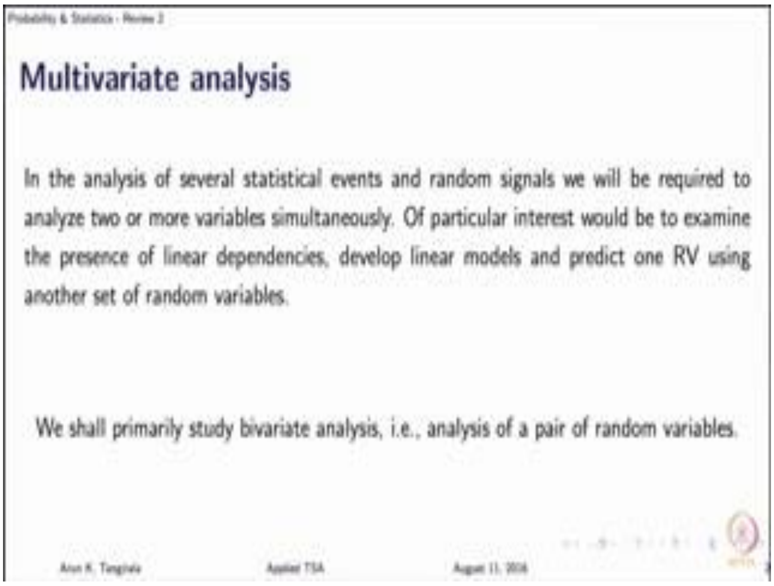
Student: Why is it always Gaussian distribution?

Sorry.

Student: Noise is a.

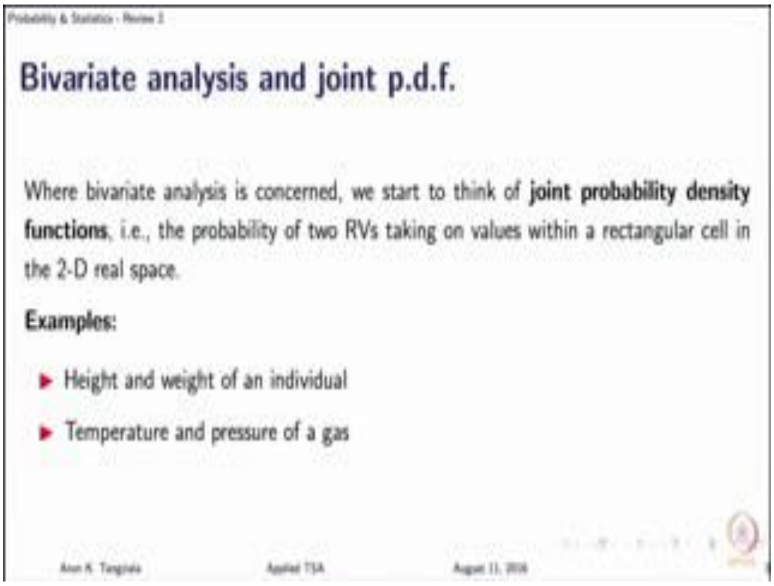
We will come to that, right now we are not talking of random signals, we are only talking of random variables when we come to the signals, we will answer that part of your question.

(Refer Slide Time: 17:19)



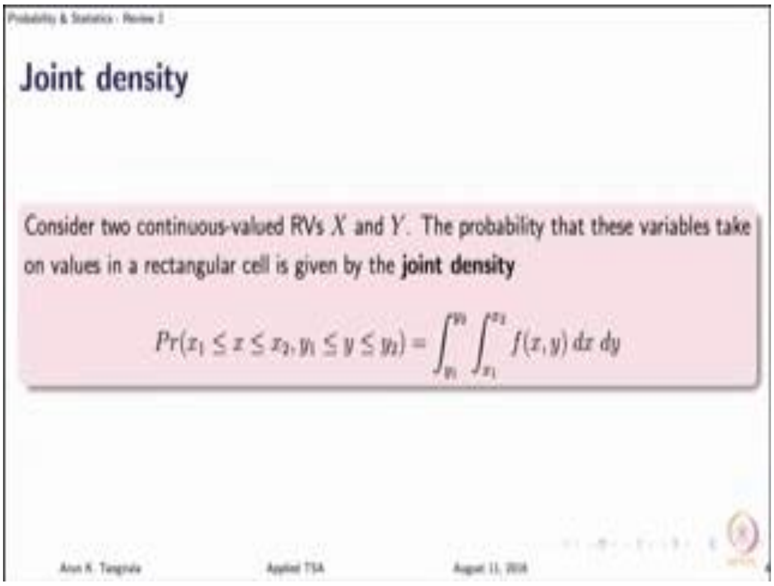

Let me at least start off on multivariate analysis very briefly maybe about 2 3 minutes and then will continue tomorrow. So, we have learned how to characterize a univariate sorry a single random variable, but very often we are dealing with more than 1 random variable; obviously, even if you take the random signal ultimately what am I going to do? I am going to actually look at the correlation for example, or look at how v k depends on its past which means that I am going to collectively analyze or jointly analyze many random variables together, until now we have focused on a single random variable. So, we need to be well versed with the theory of multivariate analysis or random you know analysis of more than one random variable. But we will start with bivariate analysis because by and large that is sufficient even for the course the bivariate analysis is more than in a occasionally we may talk of more than 1 random variable particularly in estimation we will, but in those cases the random variables are not the observations, but the parameter estimates we will talk about that later anyway.

(Refer Slide Time: 18:37)




Let us assume that I have now 2 random variables and I have to analyze them jointly and there are numerous examples that one can give for example, height and weight of an individual or temperature and pressure of a gas or maybe student grades and maybe earthquakes and maybe Timbuktu or whatever. So, there are so many things I can correlate I can jointly analyze.

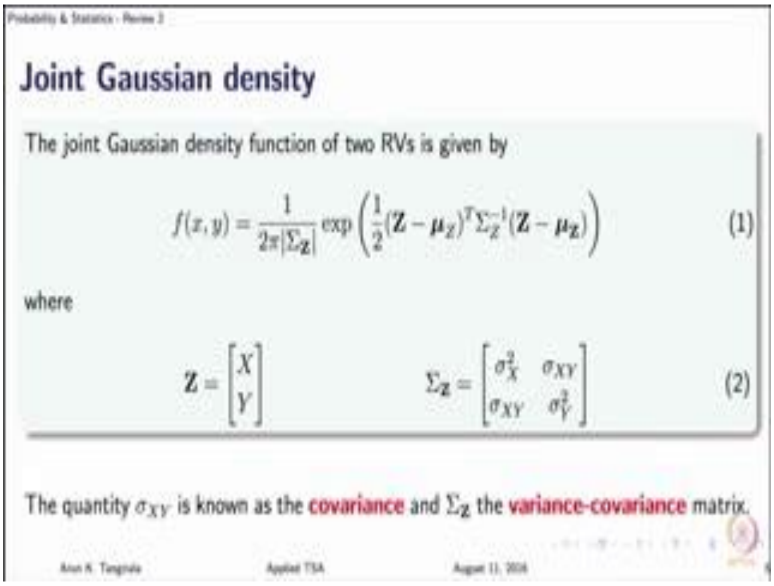
(Refer Slide Time: 18:56)




When we are jointly analyzing more than 2 or more random variables, the notion of joint density comes into picture because now the question is about probability that both these random variables take on values within a cell in the 2 dimensional space earlier we looked at the probability of the single random variable taking on values within some interval on the real line, but now we have to look at the 2 dimension space and to compute that probability we need the joint density which is given denoted by f of x comma y. So, instead of a single integral now you have double integrals a bit more scary, but it is ok.

We know by now you should know that we do not work with densities as much as we work with moments. So, very quickly we will talk about moments.

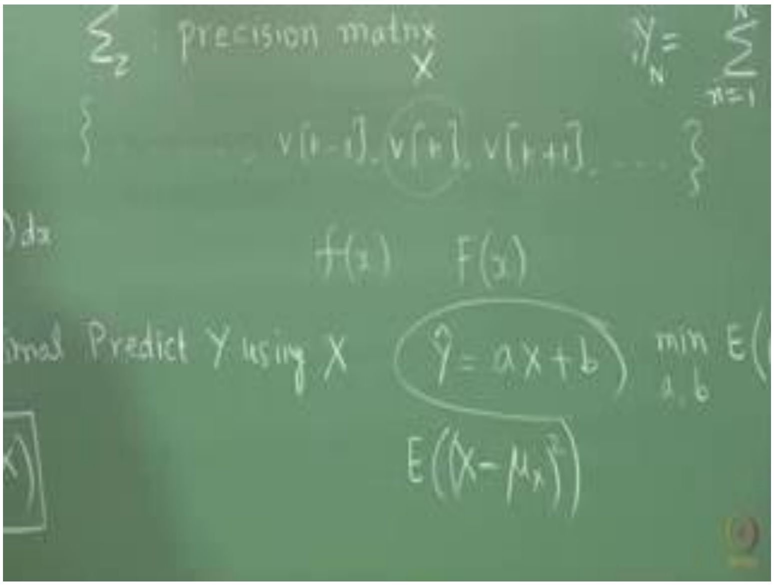
(Refer Slide Time: 19:44)




But before we do that let me give you an example of a joint Gaussian density which we will frequently work with you have seen the expression for the density of a sorry for the Gaussian density for a single random variable case yesterday. This is the expression that you see for the joint Gaussian density and you can see striking similarities with the expression for the univariate case, except that instead of sigma X in the denominator in the multiplication factor now you have determinant of the so called big sigma z and what is that? It is defined in the equation 2z is a vector of random variables now. So, slowly now we are moving from a scalar to a vector case and it is good to get used to this notation. So, z is a vector of random variables and this sigma big sigma that you see is called the variance covariance matrix we will define what is covariance and talk about correlation tomorrow. But by now you should be also familiar with this terms the diagonals of this variance covariance matrix contains what are known as variances of the individual random variables and off diagonals contains the most important quantity in data analysis which is the covariance, all right.

And it is a symmetric matrix. In fact, it is a symmetric positive definite matrix and instead of X minus mu X whole square by sigma square X in the univariate case now you have z minus mu z transpose sigma z inverse times z minus music. Now the sigma z inverse has a name it is called the precision matrix you will understand the reason for this name when we learn the estimation theory as to why this is called the precision matrix.

(Refer Slide Time: 21:23)




Now, this is a joint Gaussian density a very an important point that I want to adjourn the class with is it does not mean that this of course, it does not mean that individually X and Y have Gaussian densities that is a very important point that you should keep in mind just because X and Y jointly have a Gaussian density function it does not necessarily mean that individually X and Y have Gaussian density functions this individual density functions are known as marginal densities and will talk about that briefly and then go to independence and talk about covariance tomorrow, but it is important to keep that in mind.

That means of course, what I mean by this is if X and Y have joined Gaussian densities. So, let me correct the statement slightly, if X and Y have individually Gaussian density functions it does not mean that they have necessarily a joint Gaussian density function. But if X and Y have a joint Gaussian density function then it is necessary that individually they have Gaussian densities. So, I am just making correction to the previous statement that I made earlier I said it does not it is not necessary that X and Y should individually have a Gaussian density function it is incorrect.

When X and Y have a joint Gaussian density function then the marginals are going to be Gaussian, but not the other way round; that means, I cannot start necessarily from Gaussian distributed X and Y and then guarantee that the joint one is also going to be a Gaussian density function. It has its own implications except under special cases where

X and y are uncorrelated when X and Y are uncorrelated the sigma that you have is a diagonal matrix the sigma the off diagonal elements are 0 in which case you can just simply factorize the joint density into a product of 2 Gaussians. So, when X and Y are uncorrelated and individually have Gaussian density functions the joint density function is also a Gaussian, but in general, it is not; on the other hand when X and Y have a joint Gaussian density then individually X and Y will have Gaussian densities as well you will understand that later on through an assignment exercise and so on.

Tomorrow what we will do is, we will talk about the notion of independence covariance, the importance of correlation in the linear world, remember we said in the linear world, it is sufficient to know the first and second order moments we will see that with an example.