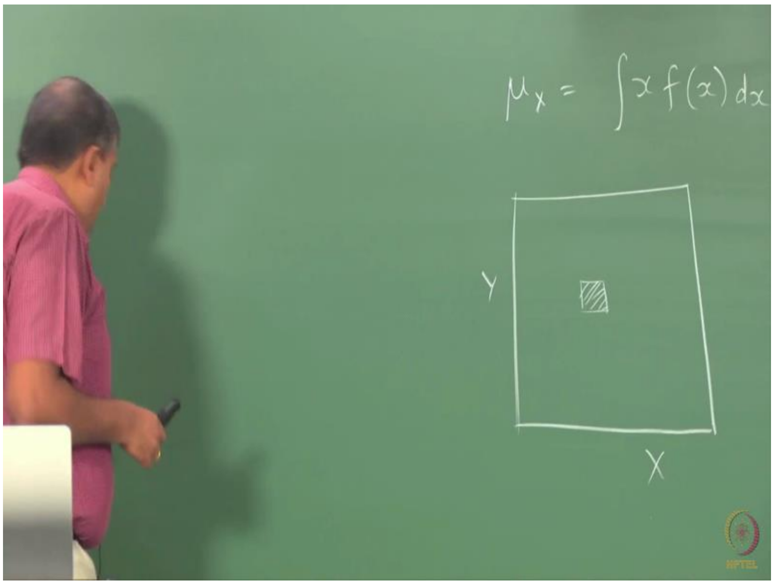
Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture - 11
Lecture 06A - Probability and Statistics Review (Part 2)-1

Very good morning, before we proceed further, there were a couple of questions after the end of the class. So, for the benefit of everyone, I am going to discuss very briefly those 2 questions. So, one of the questions was, well it has got to do with the definition of mean and variance in general moments.

(Refer Slide Time: 00:34)



Let us take the mean definition which is of course, the expected value and it is nothing but the first moment of the density function and the question was well since we said that it is hard to estimate PDFs in practice, so, how do we use this definition in practice and the answer to that question is well we do not use this definition in practice.

What I mean by in practice is when we have data; this definition is not so useful. Although you know you can come up with empirical versions of this and go on to derive an estimator, but this is not the starting point for estimating mean from data. We formulate the estimation problem in a completely different manner; we pose it as an optimization problem and come up with answers to estimating mean.

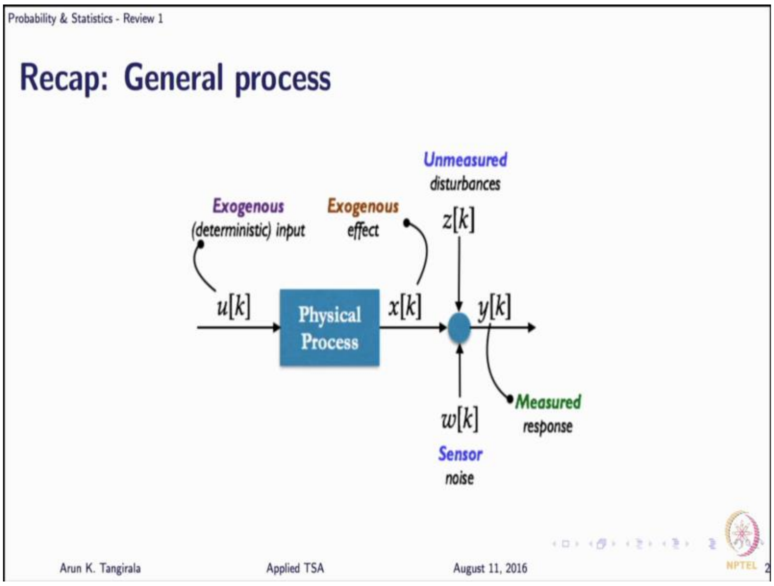
Then the question is why do we study this? Why do not we just plunge into the practical world straight away? Well the answer to that question is let us say I figured out a way of estimating mean and we all know sample mean is one way of estimating this mean, how do I know how good the estimator is? How can I answer that question whether my estimator has produced a value? That means an estimate that is very close to the truth. Any estimator is good if it can produce estimates that are close to the truth, we have not defined what is meant by closeness and so on, but qualitatively speaking. To be able to answer that question we need to have a definition for the truth and the beauty of this entire theories it is defining the truth for us. Once we have the truth defined in proper terms then we can derive not just 1 or 2, but 100 different estimators and assess the goodness of those estimators.

So, to summarize we are looking at these definitions for 2 reasons, one for theoretical analysis. So, suppose I am given the theoretical description of a random phenomenon then I can analyze whatever I want to analyze about that phenomenon using these definitions and two, when I move to the world of data and I am going to work with estimators, I will be able to answer clearly whether one estimator is better than the other, if yes under what conditions and so on. At the moment it may be hard for most of you to be able to see that, but hang on to this theoretical definitions that we are looking at. In fact, you will be glad that you are in the world of theory when you move into the world of estimation that wall you know those definitions are easier to work with.

But again I keep saying this truth is one estimators are many god is one realizations are many, likewise truth is one, the definition of mean is fixed, you can have 100 different ways of estimating this mean, I can have sample mean for example, I can have sample median, I can have sample mode, I can build up 100 different estimators to estimate this quantity mu. So, do not think that because I am working with mean, I should only use sample mean, no I can use sample median also to estimate to get an estimate of mu, there is nothing preventing me from doing that. That is the first question; hopefully that has answered the question satisfactorily.

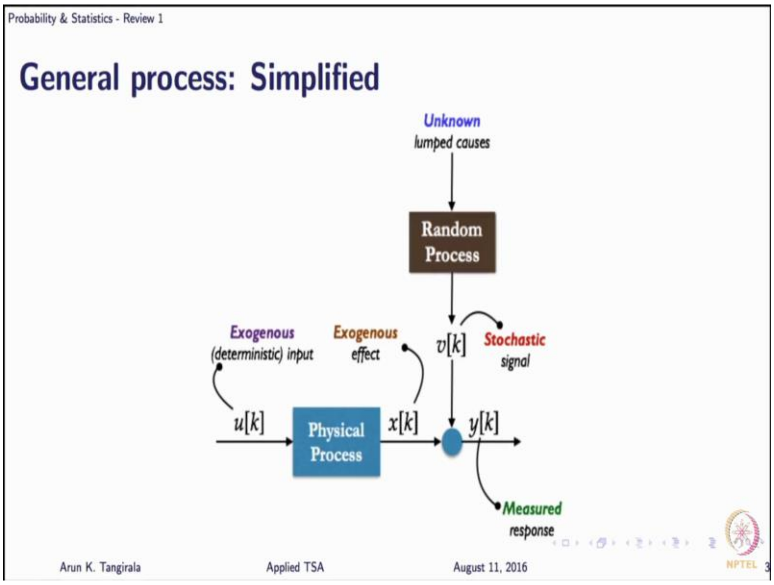
The second question was that came out of the central limit theorem and maybe I should just take you back to the other presentation that we had, one of the starting slides that we had, let me take you back quickly to that.

(Refer Slide Time: 04:22)



If you recall in a couple of lectures ago, we started off with this general this description of a general process and then gradually we converged to the notion of random process. At this point, we said well apart from the response of the physical process, you have effects of sensor noise and unmeasured disturbances and so on.

(Refer Slide Time: 04:45)



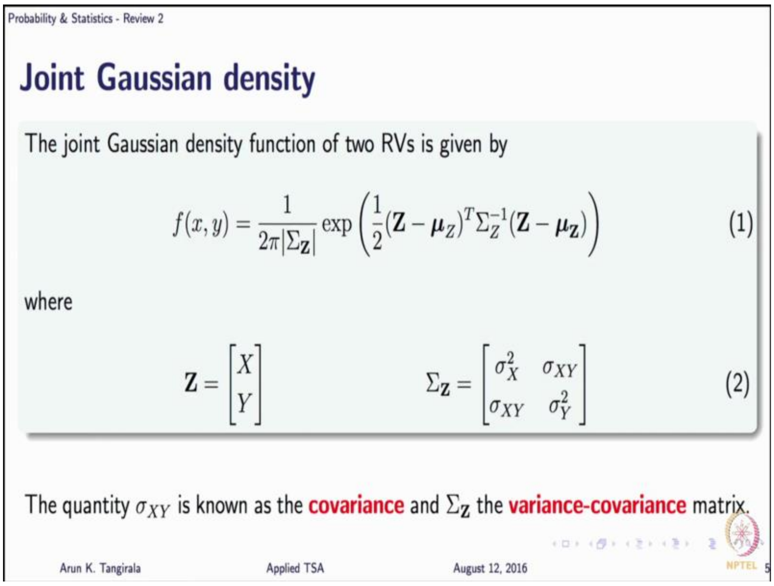
And then we said we lump the effects of unmeasured disturbances and sensor noise into a single effect and the question was whether this lumping of these 2 effects or more effects into a single quantity known as noise or you can say a stochastic signal is the motivation, the central limit theorem. Well, not exactly that well the one of the common assumptions on the distribution of the stochastic signal is that it is a Gaussian distributed signal, well jointly Gaussian distributed signal. So, the question was whether the central limit theorem is a reason for assuming this for this common assumption on the stochastic signal being jointly Gaussian distributed, and answer is no. It is not the reason.

There is another reason which comes from a different theorem called the DLT. It is called the degree limiting theorem which means that you assume whatever is simple and that can get your degree very quickly, so that is I mean I have just made it up, but the reason for assuming Gaussian distribution is not on the stochastic signal is not because the 2 effects that we have the unmeasured disturbances and the sensor noises add up and produce a random variable with Gaussian distribution, that is not the reason. The reason is it is mathematically convenient, it will get you answers quickly and then you can graduate. So, those are the reasons. The mathematics of the entire estimation theory or whatever your analysis is becomes highly tractable and easy to work with when you start when you assume Gaussian distributed noise.

But there is a body of literature on non Gaussian noises and so on; it does not mean that only people work only with Gaussian distributed stochastic signals. So, to put it short, the reason for assuming Gaussian distributions in general for random signals is not because always it is a central limit theorem that comes into picture, it is the degree limiting theorem or the paper publishing theorem and so on. So, I want to publish a paper quickly, I want to get my work done, I want to be able to sleep. So, I assume Gaussian distribution. I hopefully that answers the questions.

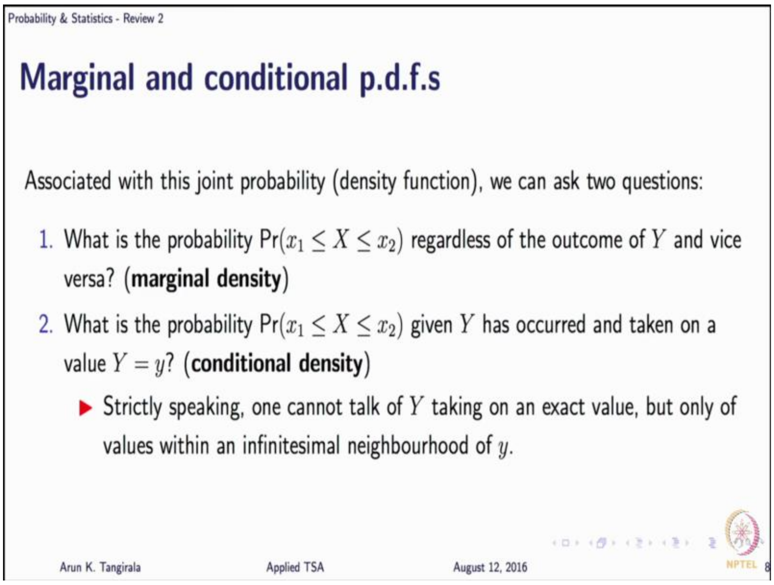
So, let us get on to the multivariate rather bivariate distribution, unless there are a few more or a few other questions pertaining to yesterday's lecture, are there any questions? So, yesterday we started looking at 2 random variables and we said that joint density function comes into play when you are dealing with 2 random variables and we were looking at an example of the joint Gaussian density and we will learn shortly what the notion of marginal density, you should also be familiar, remember this is a review lecture.

(Refer Slide Time: 08:00)



So, the joint Gaussian density is given on the slide for you and my point was yesterday in general when 2 variables have a joint Gaussian distribution; yes, the marginals will also work out to be; that means, the individual PDFs will work out to be Gaussian, but not the other way around. If the individual distributions are so, called marginals work out to be Gaussian, it does not mean necessarily that the joint is also Gaussian unless only under some special conditions as I said when the 2 variables are uncorrelated. So, today will also talk about correlation.

(Refer Slide Time: 08:28)

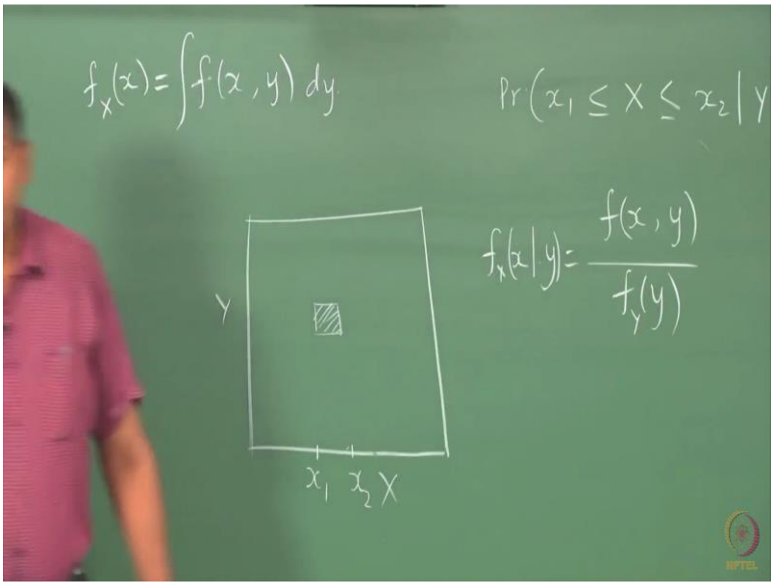


Let us move on and ask 2 burning questions that are of interest in when it comes to bivariate analysis and the first question is - what is the probability? So, the joint density function allows us to compute the probability of the 2 variables taking on values within a cell, but quite often I am interested in knowing what is the probability of 1 variable x falling within an interval of outcomes given the other has actually taken some value in its own space. And that is as standard thing for example, when there are many examples suppose you have Rahul carrying an umbrella as one event and you know it is raining outside as another event. So, what is the probability that Rahul carries an umbrella given that it is raining outside, that is different from I mean it is not necessarily going to be identical to the probability that Rahul will carry an umbrella and that it is raining outside.

This takes us to the notion of marginal densities, what we mean by marginal densities? Is that here you; now in dealing with the in the 2 dimensional plane. So, here is your x and this is your y. Basically I have drawn the on the y axis, you have the space of outcomes for y x axis as a space of outcomes for x and f of x comma y, the joint density allows us to calculate the probability that x and y will take on values in this cell, but now we are asking, I do not care, a real limit does not matter, sorry! The earlier thing that I said was conditional density, I am really sorry! The going back to the earlier example I would like to know what is the probability that Rahul carries an umbrella regardless of whether it is raining or not, I slightly jumped the gun a bit.

What is the probability that Rahul carries an umbrella regardless of whether it is raining or not? In other words I would like to know what is the probability that x takes on values within a certain interval regardless of the outcome for y, I really do not care; obviously, then to answer that question I should be able to I should first get this information or the corresponding density function.

(Refer Slide Time: 11:02)



To answer this question here, I need the joint density that is fine. Now, I am interested in knowing what is the probability that x takes on values within some interval x1 and x2 regardless of what y is doing? So, in other words I am ignoring in some sense the second dimension, you can think of it that way or I am projecting now a 2 dimensional function to a 1 dimensional function and the way that is done is by integrating this density function along the space of y because you are actually now condensing a 2 dimensional function to a single dimensional function and since we are interested in this question of course, you can also pose the other question, let us stick to this then one obtains what is known as the marginal density. This marginal density will allow me to answer the question that we have on the screen which is what is the probability that x takes on values within that interval x1 and x2 regardless of what happens to y.

That means I have to be able to take into account all possible outcomes of y and that is obtained by walking along y and integrating f of x comma y along that space of outcomes of y, well you can think of this as a page and you are looking at a margin. So, you can think of it as a reason for being this called as a marginal density.

Marginal density is not only occur in probability, for those of you who work in signal processing, who work in slightly advanced stuff like time frequency analysis and so on, you come across the notion of joint energy distributions, joint energy densities and marginal densities as well in time and in frequency. So, density functions appear everywhere.

The second question is what I had mentioned earlier that is also of interest in practice is now I am given that y has taken on some value specific value. So, this is completely complimentary to the first question. Now y has taken on some value like now it is raining, I would like to know what is the probability that Rahul carries an umbrella which is going to be different from the general probably; I mean it is not going to be necessarily identical to the general case where Rahul carries an umbrella just like that regardless of whether its rains or not, I mean we would think that Rahul would carry an umbrella only when it rains perhaps or the probability is higher.

What happens? Let me ask you what happens if this probability that Rahul carries an umbrella when it rains is the same as a probability that Rahul carries an umbrella in general, what do you call such events?

Students: (Refer Time: 14:05).

Very good, that is exactly when you say, well you know Rahul just has this fancy of carrying an umbrella and does not matter whether it rains or shine that is what it is, but let us hope that is not the case. But in general you would love to deal with independent events because the math is easier. So, there is always this conflict between reality and the con mathematical convenience.

To answer this question which is what is the probability that x takes on values within this interval given that y has taken on some value y 1 or y you can say. So, we need the notion of conditional density, this is different from marginal density. So, now, you are trying to evaluate, what you are going to do is you are going to use this joint density, again you need a 1 dimensional functions, but this 1 dimensional function is not obtained by projecting f of x comma y completely on to the x axis, but by rather working with a slightly different procedure.

So, the conditional density of x given y is obtained using the Bayesian rule and in fact, more specifically I should use the marginal density here. You observe that I am using the subscripts x and y for marginal densities and there is a reason, what do you think is a reason? Why cannot I simply write for example, here f of x and why should I use this subscript? Any particular answer!

Student: Small x is w.

Sorry!

Student: Whatever (Refer Time: 15:56) put in the parenthesis is the dummy variable.

No.

Student: (Refer Time: 16:01).

That is not the reason, any, yes.

Student: (Refer Time: 16:06) marginal PF of x.

Ok.

Student: And you are calculating at x (Refer Time: 16:12).

How different that would be from the general density function that we have looked at in the univariate case?

Student: (Refer Time: 16:20).

Sorry.

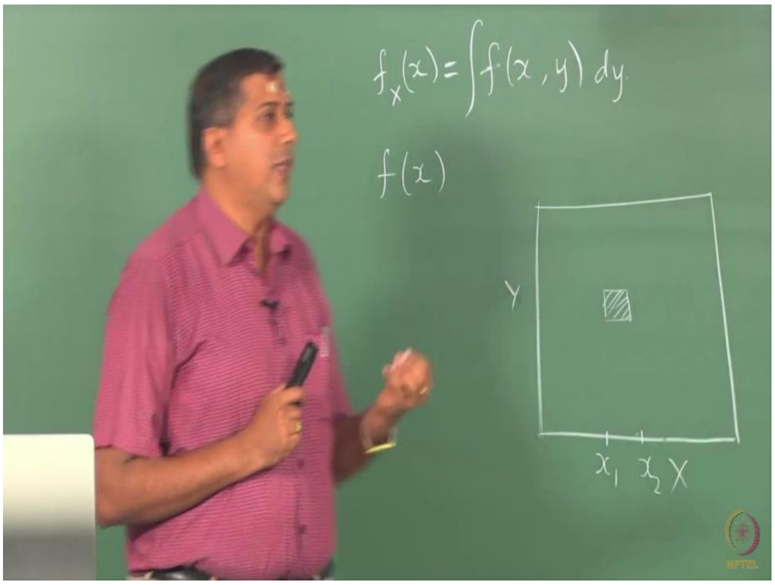
Student: It is only for that particular (Refer Time: 16:26) it is not that entire (Refer Time: 16:28).

There is no interval here, I am going to integrate this along the entire outcomes phase of y, am I right? So, I am not integrating it in an interval in y, somebody has an answer, yes.

Student: (Refer Time: 16:39) case we are considering both x and y (Refer Time: 16:42) then respective y I mean then why there is a particular value, but in the second case (Refer Time: 16:46).

No, I am not comparing conditional densities with marginal density I am just asking, why are we using this subscript for marginal densities? Why cannot I let go the subscript here, why?

(Refer Slide Time: 17:00)



In other words what would be the difference between a marginal density and the individual density function itself for x, anyone from that hall, yeah?

Student: (Refer Time: 17:14) next to covariance in between x and y.

Where is covariance? We have not even spoken of covariance here.

Student: (Refer Time: 17: 25) But then I am joining the Gaussian distribution if you assume Gaussian distribution.

Yes.

Student: the only thing is that separating joint distribution function with the (Refer Time: 17:32).

Ok.

Student: And in the joint distribution the covariance between the x and y is taking care of.

We are not the perfect answer, yes, yes.

Student: x y depends on other variables.

Yes, very good. So, the main reason is that here when you look at the marginal density, it has been derived from the joint density by only condensing it along the y direction, but x may be jointly varying with many other factors and this individual density is obtained I mean imagine that you have taken all the factors that x co varies with like when you talked about covariance, co varies with and you have actually condense it along all dimensions and obtain this density.

Normally we do not think of it that way, but now that we are speaking of joint densities, you can give a different interpretation to this individual density function that you have for f of x whereas, this one has been specifically derived by projecting it only along the y dimension. They need not be equal in general. How do you verify it we do not well, all you have to do is you have to observe x by itself you do not worry about other factors at all and then estimate the PDF that will get you f of x.

Suppose you want to get the marginal density then you observe x and y jointly and then condense it and obtain your marginal density. If y is the only factor that x covariance with and no other factor then the marginal density and your individual density will coincide. In other words f of x has been obtained whether you did it or not and have been obtained by considering all the factors that x co varies with for that phenomenon whereas, marginal density is only by considering that particular another other phenomenon y.

Therefore, in the initial stages, it is important to observe the subscript, but gradually well get rid of this, it is a not needed, it is understood that we are dealing with marginal densities whenever we are dealing with more than one variables. When we are only handling a single variable is understood that we are dealing with the regular density function. So, hopefully that is clear, all right.

Now going back to the problem of conditional density, this is your conditional density that you will work with and that you will use to answer this question, what is the probability that x takes on values in a certain interval given that y has occurred and let us look at this of both bit more in detail, but before we do that, let me also say that strictly speaking with the conditional density, this is not the perfect way of writing it, why? Because we are dealing with continuous valued random variables and it does not make sense to say that y has exactly taken on a value y 1, we have already said that is kind of an ill post statement to say that a continuous valued random variable has exactly taken on some value in the continuum. It is understood implicitly here that although we write y equals y 1, it means that y has taken on some values within infinitesimal neighborhood of y 1 that is what it means.

But of course, you can ask these questions for discrete valued random variables that is not an issue, only for continuous valued random variables the strictness comes in only if you interpret this as y taking on values within an infinitesimal neighborhood of y 1. So, that is all for the rigor, but for all practical purposes we keep writing it this way.

Associated with this marginal and in fact, marginal density concept is essential to arrive at conditional density, that is why the questions were posed in that order as you can see and of our interest will be in practice a conditional density. The marginal density kind of stays in the background helps you to compute the conditional densities, but practically what is of use and even theoretically are is the conditional density, it is one of the most powerful concepts in the entire data analysis or you can say random signal processing especially in prediction theory.