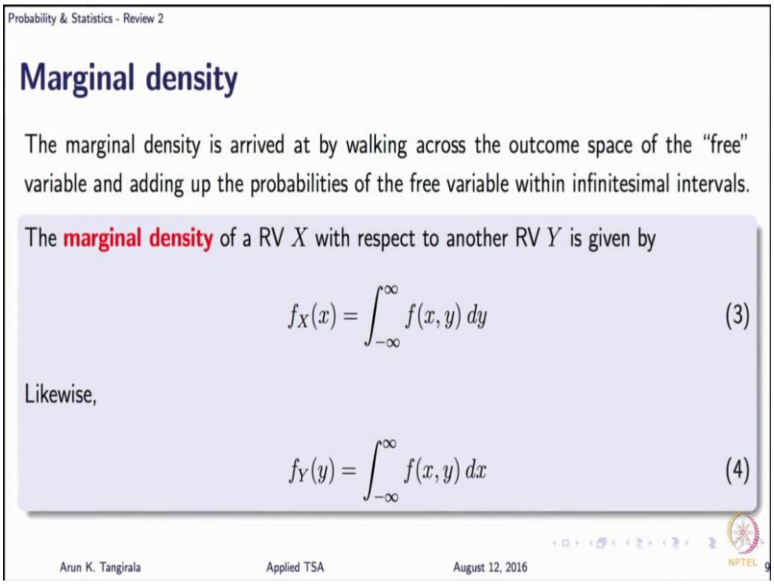
Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

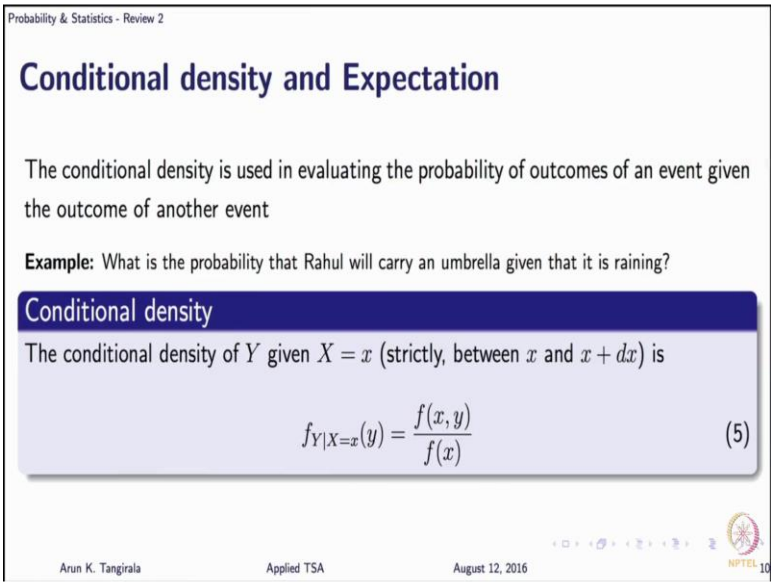
Lecture - 12
Lecture 06B - Probability and Statistics Review (Part 2)-2

(Refer Slide Time: 00:11)



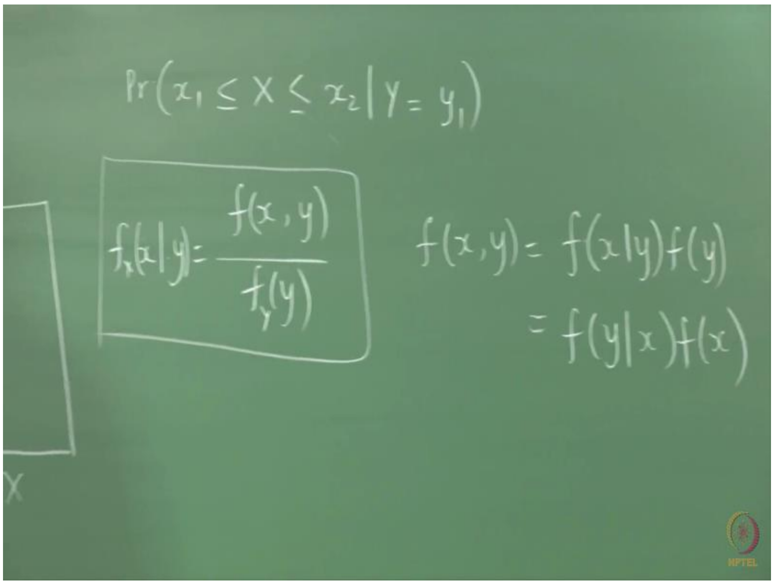
We will move along and talk more of conditional densities and conditional expectation. It is a very important to understand this notion of conditional expectation because it will make frequent appearances in the prediction theory and throughout the course as well.

(Refer Slide Time: 00:29)



As I have written here, the conditional density of Y here, in this case have written on the slide for Y, but on the board I have written for X, does not matter, you should understand.

(Refer Slide Time: 00:43)



Now, this offers a nice way of writing the joint density as a product of the conditional densities and the marginal density. Now I have stopped writing the subscripts, you have to understand that this is actually the marginal density of Y and obviously, by straightforward extension, one can even write it this way and we will use this expression later on when we talk of maximum likelihood estimation and so on.

Now, from this also stems the definition of independence which will talk about shortly, but first let us look at the notion of a conditional expectation, all right.

Yesterday we reviewed the notion of expectation which was to be called as the unconditional expectation that is only given information about X, what is the best prediction of X? It is expectation. Now you can call it as unconditional expectation, why it is unconditional? Because it is not conditioned on any other phenomenon, so let me give you an example, suppose you were to ask, what would happen to the outcome of the ongoing cricket match or some other event in Rio Olympics.

For example, we know that Saina and Sindhu have made it through their opening games and or if you go to cricket, 3rd day has been washed out in the India - West Indies test match; 3rd test match. So, what is a prediction of the outcome of this test match, I know that not all of your interest in cricket, but it is one word that you cannot escape if you are in India, you have to hear it at least some 100 1000 times in your life. So, what is your prediction?

Let us say before the match began, what was your prediction?

Student: India wins.

India wins, there is some percentage and so on each channel has its own statistical calculations and so on. So, this is based on; what is your prediction based on?

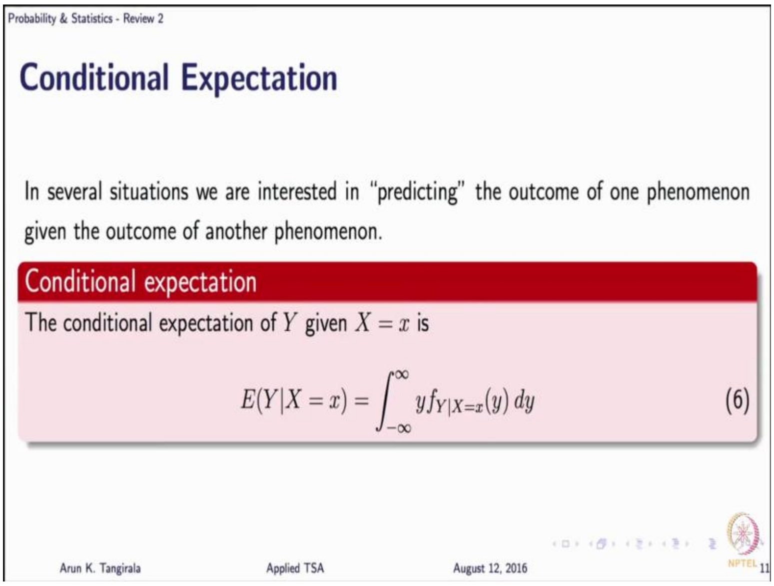
You say India wins, that is your prediction, it can go wrong too, hope not, but it can go wrong, what is it based on? History, let us assume that just you pulled off the history and he said you know if previously it used to be if it wins one then the next one is a loss and so on, but looks like the pattern has changed now. But now there used to be also this pattern I would say that the wickets that fall in a cricket match follows a Poisson distribution that is if one person gets out somehow 2 3 people get out in a row. Reminds me of the you know state transport buses in A.P. maybe I do not know if it is true for I have not traveled so much by buses in Chennai, but when I used to go to school, at that time I was not possoned by the Poisson distribution in the school days, but the general observation was that if a bus arrives then there are like 2 3 buses arriving with them together as if they have discussed and made sure that they all come together and after that there is no bus for about another half an hour so, you are doomed if your actually missed that point.

So, likewise here in cricket also, if 1 person falls and there is at least 2 3 wickets falling in a row maybe you can study the distribution apparently if you go to cricinfo, there is some enormous amounts of data that you can spend your life time in analyzing and there is a lot of money in that. So, I am giving you some ideas for startups, all right. So, let us get back to the game. So, it is based on history. Now the game begins like it began couple of days ago. As this core starts to roll out your prediction would change I mean need not, but generally speaking it would change, that is your conditional expectation, but an unconditional expectation is you are not given what is happening it regards to the even now the score is your another random variable.

One random variable is the outcome of the game another random variable is the score that is happening. So, if I before the match begin, I do not have any information about it. So, I have an unconditional kind of expectation based on whatever history that I have now this course starts to pour in and I start making conditional expectations.

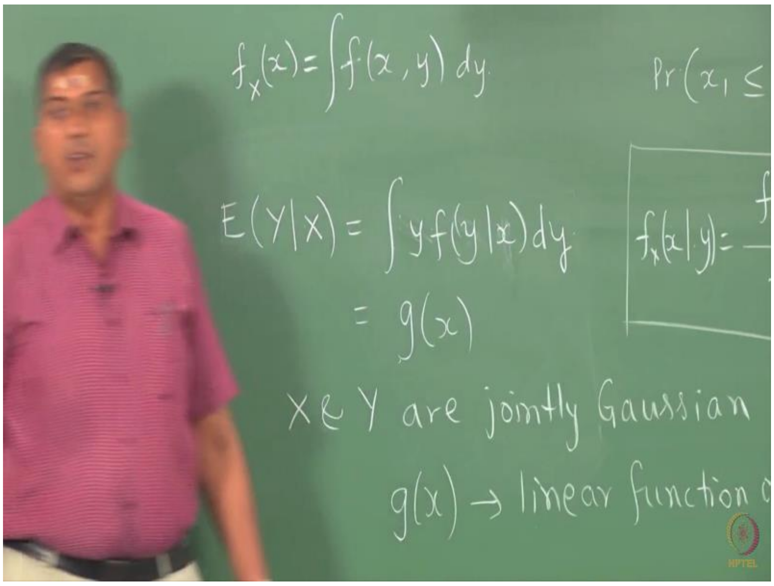
The conditional expectation in general need not be the same as unconditional expectation, but you can see yesterday, we argued intuitively that expectation is the best prediction in the absence of any other information, best in the sense of mean square minimum mean square error.

(Refer Slide Time: 05:48)



It turns out that this conditional expectation that you see on the screen is also the best prediction of Y, imagine that you are predicting Y earlier without X, now you are predicting Y given X. So, the condition prediction is also the best prediction and this is one of the milestone results in prediction theory. Given two random variables, if I have 2 random variables Y and X and I want to predict Y using X, generally speaking you think of a neural network some 100 neural networks and you have deep learning deeper learning and so on, deeper learning is yet to come, but right now, we have gone from learning to deep learning and so on, you can build any complicated function of X, the conditional expectation will beat it all.

(Refer Slide Time: 06:37)



They cut therefore, you should be well versed with the conditional expectations you can write it this way well. On the screen have said X equals x, but generally we will stick to this notation and it is defined using the conditional density there as the same way it is the first moment sorry, say why it is the first moment of the conditional density, integrated or what?

Student: (Refer Time: 07:08).

You are not divided on this, sure. Because what you are doing is now you are saying X is anchored at some point, now you are walking across the outcome space of I and taking the statistical average, all right which means that this condition expectation is a function of X, very good so, it is some function of X.

You can build any complicated function of X to predict Y given X the conditional expectation will beat them all in what sense? In the minimum mean square error senses among all the functions of X that will predict Y given X the conditional expectation has the lowest mean square error. And it is one of the most beautiful results in prediction theory, unfortunately it is one of the most difficult results to use in practice always nature is not so kind to us, we have some fantastic results in theory, but unfortunately we cannot use them.

Why do you think we cannot use them, I mean we can use it, but it is very difficult to do this, imagine now extending this to the multivariate case, what generally prevents us from using this function; this result in practice for prediction purposes, somewhere here.

Student: (Refer Time: 08:38).

Sorry, exactly. So, they again the same, somebody was also saying conditional densities, joint densities are very difficult to come by very difficult to obtain, it is not easy maybe in the bivariate it is easier to estimate, but the moment you move to maybe a 10 variable case maybe 20 years down the line when we have you know even more advances in computation power and also theory will be in a better position to use this result, but then we will have something else to trouble us no worries.

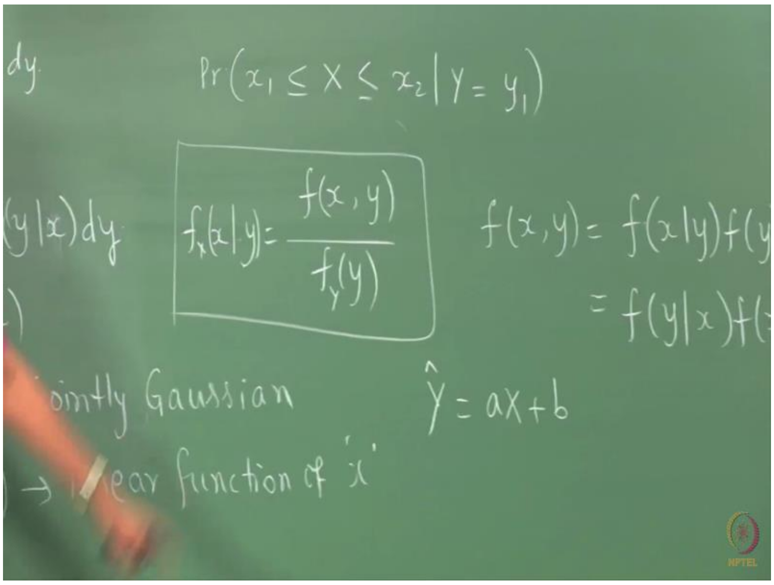
So, this result is useful nevertheless to derive a lot of other relations and we will spend a bit more time on this conditional expectation. So, let me state that result formally, now although I am not proving anything, well revisit this result when we get into predictions. The conditional expectation is the best prediction of Y given X and it is best in the minimum mean square error sense.

Now there is a very nice result which is kind of some good news, the bad news is in using this result is we do not have generally the knowledge of conditional densities or joint densities and so on, but there is a good news in some situations and that situation is when X and Y are jointly Gaussian. Earlier we had seen the joint Gaussian density function, when X and Y are jointly Gaussian, what we mean by that is X and Y have a joint Gaussian distribution then you can show theoretically that g of x is a linear function of x and that forms the backbone of many of the linear models that we study in this course.

We can prove it and you can prove it, you understood what I meant, you can prove it means it will be a part of your assignment, but it is a fairly easy problem to solve, it is a fairly easy thing to show. You start with the joint Gaussian density function and start evaluating the conditional density conditional expectation and then use a small trick and show that the conditional expectation is a linear function of X in general g of x is non-linear.

What does it mean? Look at it this way; we decide in general that given an opportunity I would like to work with linear models regardless of whether X and Y are jointly Gaussian or not why? Because of convenience it is easy to work with linear models, it is easy to estimate them, is easy to analyze them, you will realize all of that when we talk of estimation.

(Refer Slide Time: 11:41)



But then the question is given that this is the best predictor and that I decide to work with this kind of a model always at least to begin with which is a linear, it is a linear predictor, linear function of x; obviously, in general I am working with a suboptimal predictor because g of x is in general a non-linear function of X where as I have decided to work with a linear function.

Now, do I lose anything is in general I do lose the optimality, but the benefit that I have is mathematical convenience, anyway I have lot of uncertainties in life. So, why break my head, actually in starting off with a very rigorous function of X and trying to force fit that onto the data, start off with a linear model if it is doing a good job, I would do it.

Now the good news is that this result gives us some hope in many situations that if X and Y are jointly Gaussian then I have hit the jackpot; that means, would the linear predictor that I am going to fit of course, I will have to be able to estimate A and be optimally also it is assumed that there is a procedure to estimate A and B optimally, let us say I have that in place. Then your linear predictor is optimal predicted you do not have to go in search of any other predict when X and Y are jointly Gaussian. At the moment we are talking of random variables, we will stitch all of this together for the random signal at a later stage, but it is very important to have this clear in your mind, any questions on this, from the other audience, there looks like they are perfect set of people, they are understanding everything, any questions here in this hall.

So, hopefully now you understand the conditional expectation is the best prediction, it is a non-linear function of X, but when X and Y are jointly Gaussian then g of x is a linear function, but in general at least in this course and as a beginner, you would begin with linear predictors, this result here is a good source of encouragement in the sense that your linear predictor is going to give you the optimal prediction when X and Y are jointly Gaussian. What happens in a non Gaussian? Obviously, it is going to give you suboptimal predictions, we extent of loss in your optimality entirely depends on the process then on the nature of non or deviation from the Gaussianity, we do not know it is very hard to answer that question.