Essentials of Data Science with R Software- 1
Professor Shalabh
Department of Mathematics and Statistics
Indian Institute of Technology, Kanpur
Lecture 46
Examples in Bivariate Probability Distribution Functions

Hello friends, welcome to the course Essentials of Data Science with R Software- 1, in which we are trying to understand the basics of probability theory and statistical inference. So, you can recall that in the last lecture we had understood different types of concepts related to the bivariate probability density function.

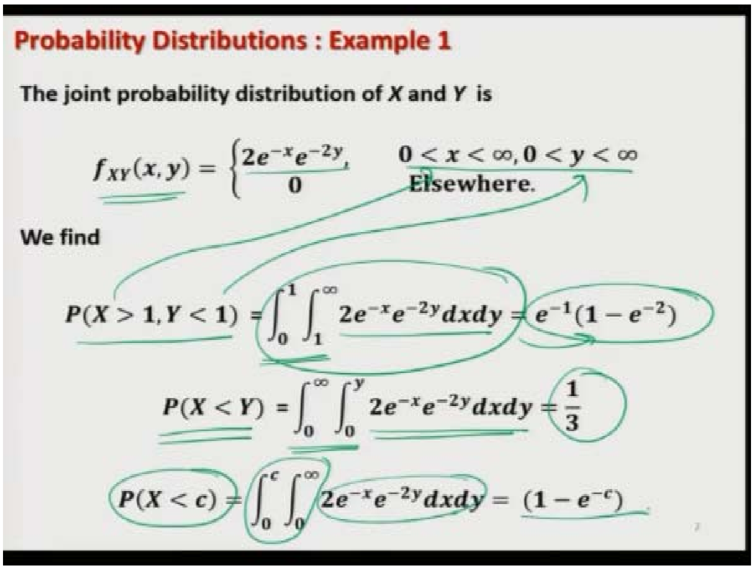
So, now in this lecture I will try to take some examples and I will try to show you the theoretical application that how are you going to compute it. Now, one thing I would like to address here today before I go into the examples, this example which I am trying to take they are based on integration.

So, definitely when you are trying to deal with the bivariate or multivariate case it is very important that you should have a good knowledge of the multiple integration and that is a very common topic which is taught in all the colleges, so that part you must know because they are going to be some complicated probabilities where the ranges of the variables are not independent but they are dependent on each other, the range of x depends on the range of y, range of y depends on the range of z and so on.

So, there the same complication that you used to find that when you are trying to find out the multiple integrals, how you have to choose the limits of the integral will also come here. So, in case if you understand the multiple integrals then you can very easily solve this problem, first thing. Second thing in the real life you will have either the data on which you have to find out such calculations, so for that we already have done in the R software and the third thing is this if you really want to solve the integral, then there are several numerical techniques which are available in the R software, you can use and you can compute this integral for a given set of data.

So, well I am not going to handle that part because there are many rules and it depends on you what type of rule you want to use and what are your limits in which you want to integrate, what are the type of function which you want to integrate but in principle those methodologies are available and they can be executed in the R software, so that I will leave up to you that how much you want to train yourself in such a topic. So, now let me try to take a couple of example and try to show you that how you can compute different types of probabilities, conditional probabilities, marginal probabilities etcetera.

(Refer Slide Time: 2:59)



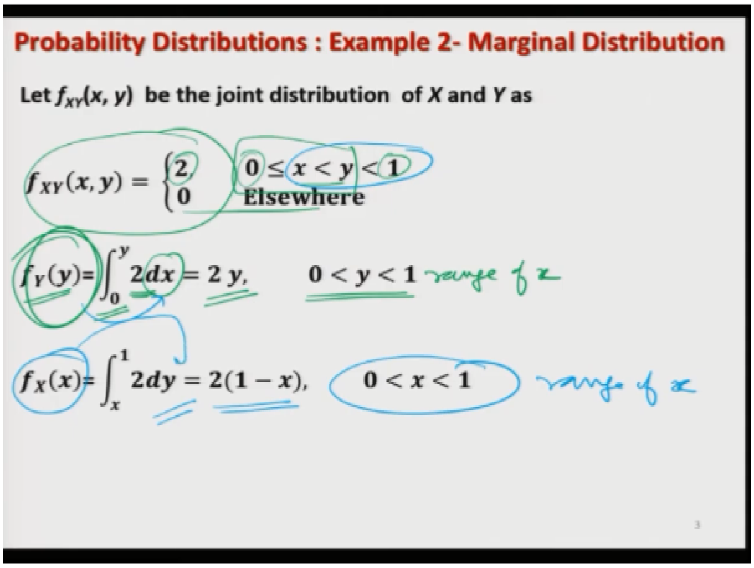
So, now we begin, so let me try to take here the first example in which the joint probability density function of the two random variable X, Y is given like this $f_{XY}(x,y) = \begin{cases} 2e^{-x}e^{-2y}, & 0 < x < \infty, \ 0 < y < \infty \\ 0 & \text{Elsewhere.} \end{cases}$, well before I move forward you have to concentrate on my here this pen in this slide because we are speaking all the mathematical sentences is a little bit time consuming and difficult.

So, I will try to indicate the expression which I am going to consider through my pen and definitely I will not be interested in giving you the complete solution of the integral because that is pretty simple thing. So, I will try to give you here the answers which you can solve, which you can obtain after solving the integrations.

So, now so in this case my X and Y they are lying between 0 and infinity and suppose I want to find out the probability X greater than 1 and Y less than 1. So, now you have to look into the range of here x and here y and you have to choose the limits 0 to 1 and 1 to infinity and then you have to write down here this joint PDF and you can solve this integral this will give you this value, $e^{-1}(1-e^{-2})$.

Now, in case if you want to find out the probability x less than y then the limits are going to be 0 to infinity and 0 to y and then you can write down this PDF and solve this multiple integral you will get here a value 1 by 3. So, my objective in this example is basically to show you that if you want to compute different types of probabilities, how you can formulate them, well once I can formulate this integral after that there is no more statistic, there is only mathematics and if mathematics becomes complicated, then you have numerical integration techniques, in case if you want to find out the probability that X is less than c, c some constant then this will be the integral over the limit 0 to c and 0 to infinity this PDF over here and then if you try to solve it you will get the answer 1 minus 1 upon c, pretty simple example.

(Refer Slide Time: 5:12)

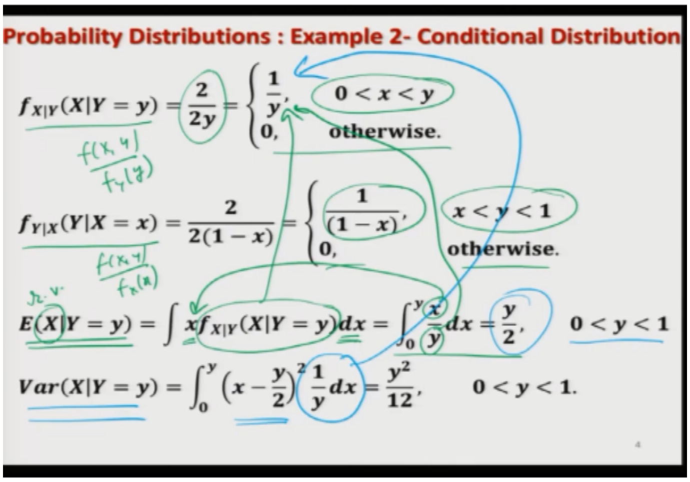


Now, let me try to take one more example that the joint PDF of X and Y is $f_{XY}(x,y) = \begin{cases} 2, & 0 \le x < y < 1 \\ 0 & \text{Elsewhere} \end{cases}$, so in this case if you want to find out the marginal probability density function of y this can be obtained here say f(y), where now you have to integrate over the range of here x, so range of your x you can see here this is like this between 0 and y, so this integration is between 0 to y and then PDF and then you have to integrate with respect to here x, because you want to find out the marginal density function of y.

And if you try to solve this integral you will get here twice of y but now you have to specify here the range of y not of x because this is a PDF whose corresponding random variable here is y, you can see here and now I try to find out the marginal probability density function of x, so this will be denoted by here f of x and you can see here that this value is going to be integral over the range x to 1.

So, because you can see here this y is lying between x and 1 and then if you try to write down here this PDF you solve it you will get here twice of 1 minus x and now you have to specify here the range of here x, without any problem. So, this is what you have to keep in mind that when you are trying to find out the marginal probability of y you have to integrate with respect to x and when you are trying to find out the marginal probability density function of x you have to integrate with respect to y.

(Refer Slide Time: 7:06)



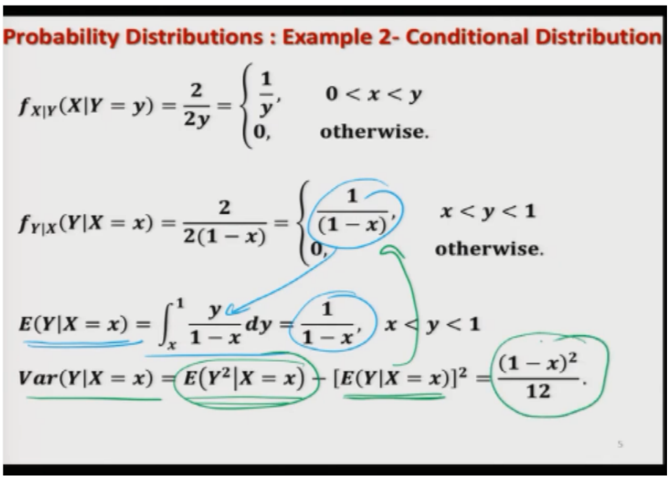
Now, I try to find out the conditional probability density function. So, for that you already have obtained here the joint probability density function, you have obtained here the marginal probability density function, so you can define for example, the conditional distribution function of x given y as here f(x, y)/ f(y) like this. So, you have obtained all these values you simply have to substitute and you can obtain it here 2 upon 2y that is equal to 1 upon y, if x is lying between 0 and y and 0 otherwise and similarly, you can find out the conditional density function of y given x by here f(x, y) divided by the marginal of your x which is here 1 upon 1 minus x if y is lying between x and 1 and 0 otherwise.

Now, in case if you want to find out the conditional mean and conditional variance and that you can find for the conditional distribution of x given y as well as for y given x. So, just for the sake of example of I try to take here the conditional expectation of x given y, so that is simply going to be the integral with respect to here x and here it will be dx, that is what you have to keep in mind, why? Because your x here is the random variable.

So, in case if you just try to write down the value of the conditional PDF that you have obtained and then this is here and then if you try to solve this $\int_0^y \frac{x}{y} dx$, x is coming from here and this y is coming from here and if you try to solve this you can obtain this integral without any problem as y by 2 where y will be lying between 0 and 1.

Now, in case if you try to find out the variance of x given y, so that you can find out here directly, you already have found the conditional expectation which is here y by 2, so I can write down here $\left(\frac{x-y}{2}\right)^2$, but now this PDF will be coming from this conditional PDF of x given y, that is what you have to keep in mind.

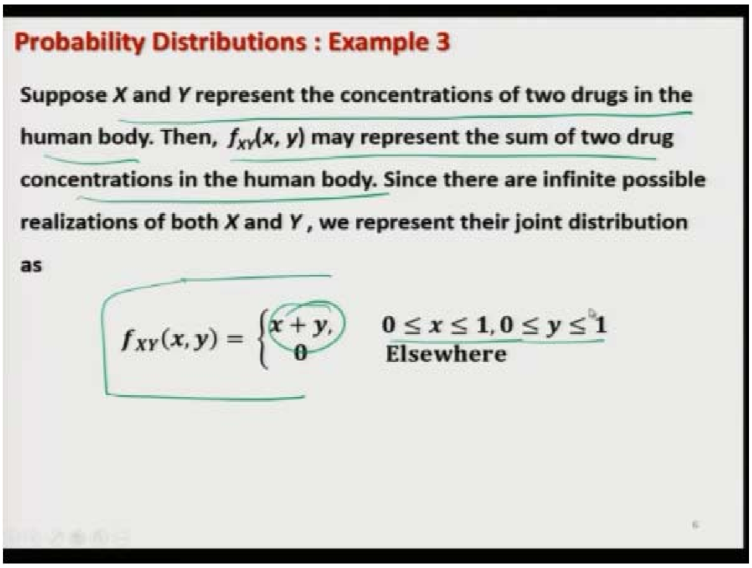
(Refer Slide Time: 9:30)



And similarly, if I ask you to find out this conditional mean and conditional variance for y given x you can very easily find them out here like this one, you can see here, here I am trying to find out in the next slide, this is E(Y|x), so here you have to use this integral exactly in the same way but you now you have to use the conditional PDF this one here and if you try to solve this integral you will get here 1/(1 – x) and similarly if you want to find out the conditional variance of y given x this is very simple.

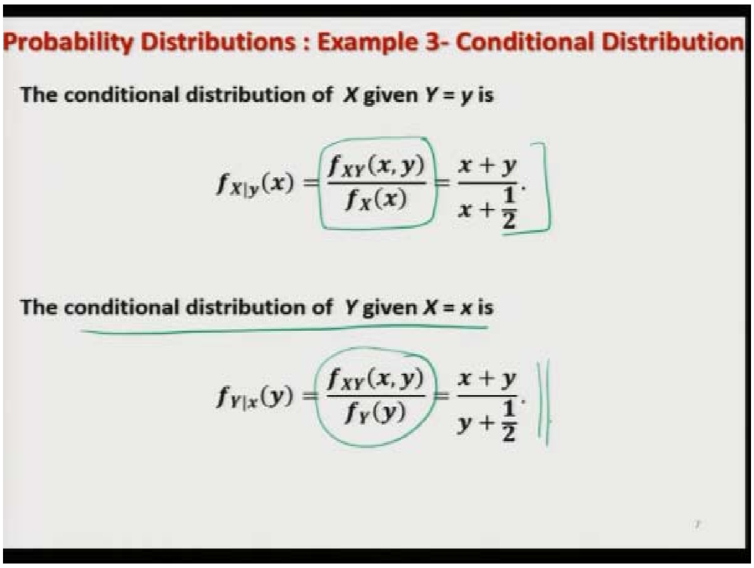
You can either obtain it directly or you can also obtain the $E(Y^2)$ and E(y) using the conditional distribution of y given x, this one so this is essentially the E(Y|X=x) and this is the $[E(Y|X=x)]^2$ and if you simply find them simple integration and then substitute them, solve them you will get here this answer $(1 - x)^2/12$, that is not difficult I am promising you these are very simple exercises for you.

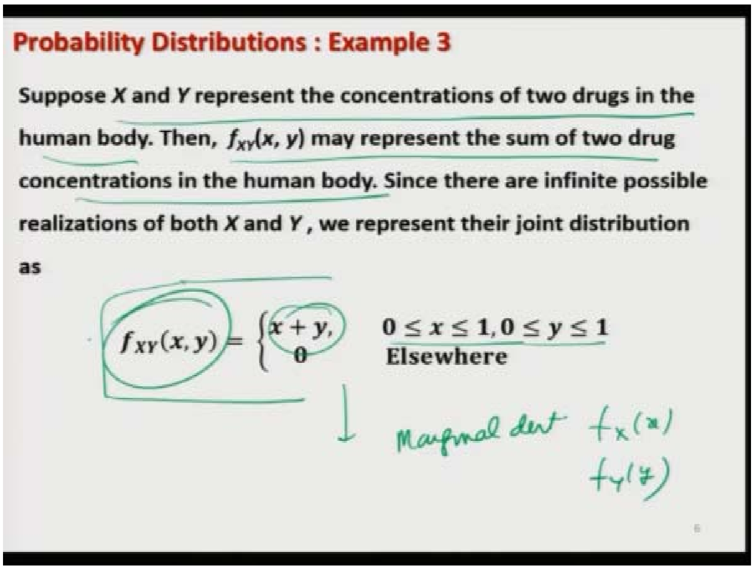
(Refer Slide Time: 10:42)



Now, I come to third example where there are two random variable X and Y they are trying to represent the concentration of the two drugs in the human body and this f(x, y) is suppose representing the sum of two drug concentration in the human body and since, there are infinite possible combination in which this x and y can be executed and all this x and y can take the values we can represent their joint PDF as here like this. So, this is given to us that $f_{XY}(x,y) = \begin{cases} x+y, & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{Elsewhere} \end{cases}$, you can imagine that how many combination will be there for the values of 0 and 1.

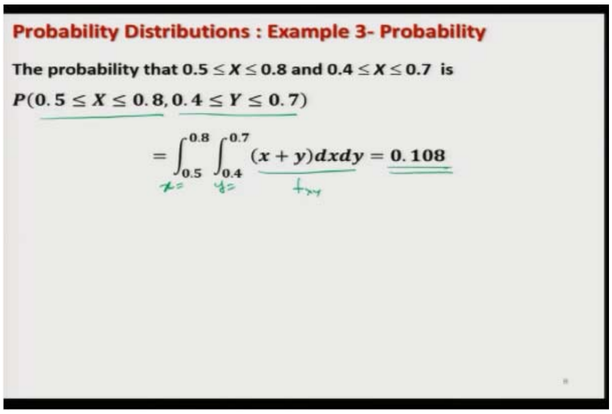
(Refer Slide Time: 11:33)





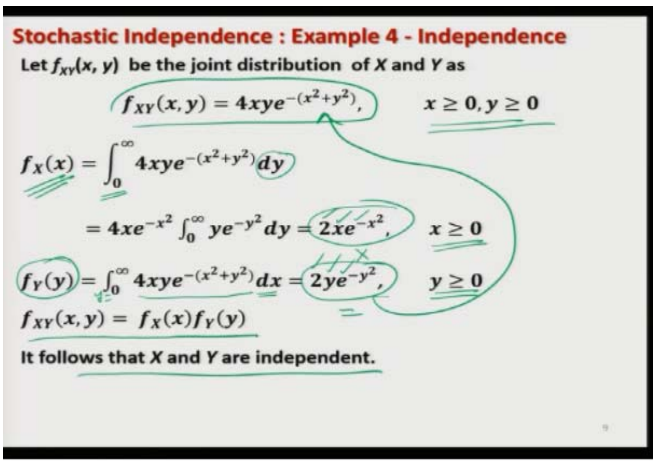
So, now we try to find out the conditional distribution of x given y, so what you have to do here, I am only giving you here the steps so that you solve yourself. Now, from this one you try to find out here the marginal like is f(x) and f(y) and then you have got here f(x,y) the joint PDF try to obtain this expression and then you can solve it, you will get here like this and similarly the conditional distribution of y given x equal to x can be obtained just by obtaining this expression f(x,y) upon f of y and you can solve it you will get here this expression.

(Refer Slide Time: 12:17)



Now, next we try to find out the joint probability. So, suppose we want to compute the probability that x lies between 0.5 and 0.8 and y lies between 0.4 and 0.7, so this is simply the multiple integral where you are trying to take the range of x to be 0.5 to 0.8 and the range of y to be 0.4 to 0.7 and then here this f(x,y), and if you try to solve this integral you will get here the value 0.108. So, you can see that it is not a very difficult thing to compute such probabilities, joint probabilities, marginal distribution, conditional distribution.

(Refer Slide Time: 12:54)

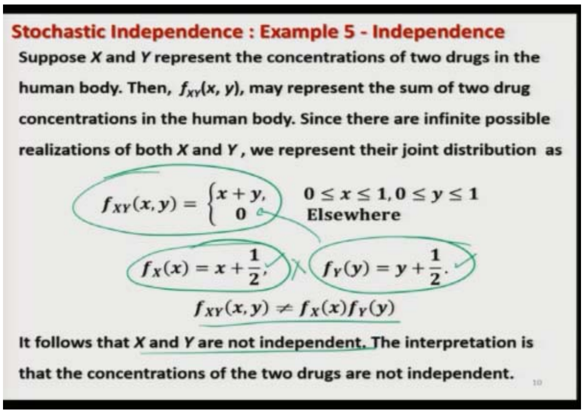


So, now in this example I try to illustrate the concept of independent that I discussed in the last lecture. So, as I discussed that there are different ways in which one can establish the independence using PDF, CDF expectation, so here let me try to take this example to show you. So, suppose a joint PDF of two random variable x and y is given by like this $f_{XY}(x,y) = 4xye^{-(x^2+y^2)}, \quad x \geq 0, y \geq 0.$

Now, if you try to find out here the marginal distributions of x and y, so the marginal distribution of x can be obtained here over the integration over the range 0 to infinity with respect to y for this joint PDF and if you simply try to solve this integral you will get here $2xe^{-x^2}, \quad x \geq 0$ and similarly if you try to find out the marginal distribution of y this will come out to be the range of y which is 0 to infinity the joint PDF and the integration is going to be with respect to x and if you try to solve this integral you will get this value $2ye^{-y^2} y \geq 0.$

Now, if you try to see if you try to multiply this and this together, what do you get here, you get the same thing what you have obtained here as a joint PDF, that will become here 2 into 2 4, x into y is xy and this exponential part will become exponential of minus times x square plus y square. So, in this case you can see here that fxy can be expressed as the product of marginal distribution of x and y, so it follows that x and y are independent here.

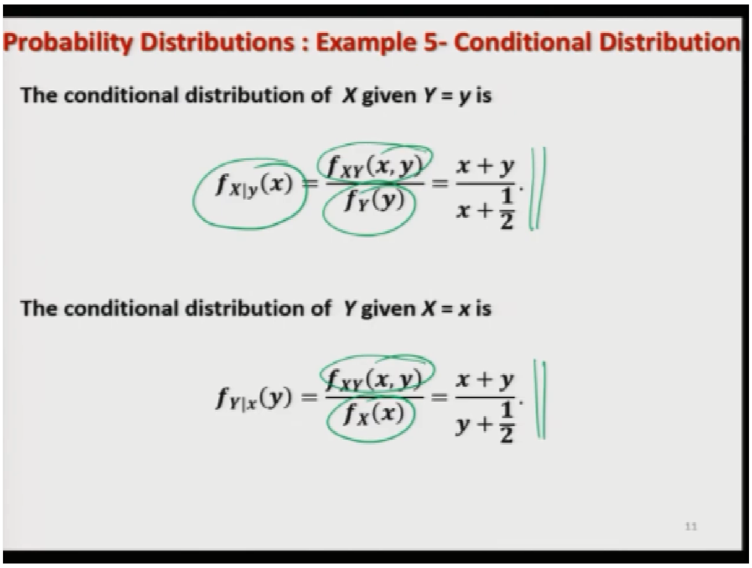
(Refer Slide Time: 14:50)



And suppose now, I try to take one more example where I can show you that it is not always that the random variables are independent. So, suppose, I try to take the same example which I just considered, where I took the $f_{XY}(x,y) = \begin{cases} x+y, & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{Elsewhere} \end{cases}$.

So, in this case if you try to find out the marginal PDF of x and y they will come out to be like this, marginal PDF of x is x plus 1 by 2, marginal PDF of y is $f_Y(y) = y + \frac{1}{2}$ if you try to multiply them together do you think that are you going to get this value x plus y? Certainly not, so in this case the joint PDF cannot be expressed as the product of the marginal PDF of x and y. So, it follows that that x and y are not interpretation, and the interpretation is that the concentration of the two drugs are not independent.

(Refer Slide Time: 15:51)



Now, you see once you have obtained the marginal from the joint PDF you can very easily compute the conditional distribution also. So, we try to compute here the conditional distribution of x given y and y given x. So, you can see here conditional distribution of x given y is simply the joint PDF divided by the marginal of y and the conditional distribution of y given x is simply the joint PDF divided by the marginal of x and if you try to substitute this value, you get here these two conditional distributions.

So, now we come to an end to this lecture, and you can see here this lecture was simply to give you an idea or a feeling that how these expressions are going to look like, and these expressions are simply the mathematical computations like double integral or multiple integral type of calculations. So, it is not difficult to do those calculation, but the more important part is you have to understand that whether you are integrating with respect to x or y and y, that is more important, once you understand this thing after that I do not think if there is any problem in their computation.

So, I will request you now you first you try to solve this integral and try to see are you getting the same answer or not what I have informed you, try to take some examples from the book and try to solve them this will make you more confident, so you try to practice it and I will see you in the next lecture till then goodbye.