PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

Week - 07

Lecture - 36

Numerical Examples on Bivariate Discrete Random Variables and the Concept of Joint Probability

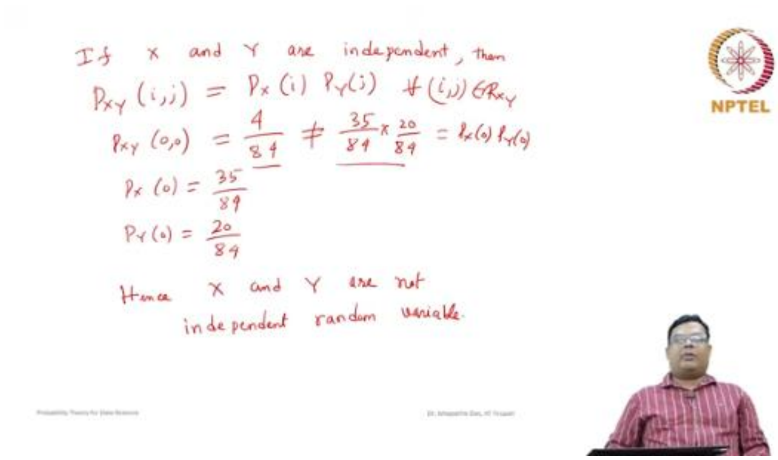
Find the range of (X, Y) and the joint probability mass function of (X, Y). We have already completed the marginal probability mass functions of X and Y, and we discussed them. Next, we will check whether X and Y are independent and how to determine that.

If X and Y are independent, then the joint probability mass function should be equal to P_X(i) * P_Y(j) for all i, j belonging to the range of (X, Y). So that means it cannot be like that.

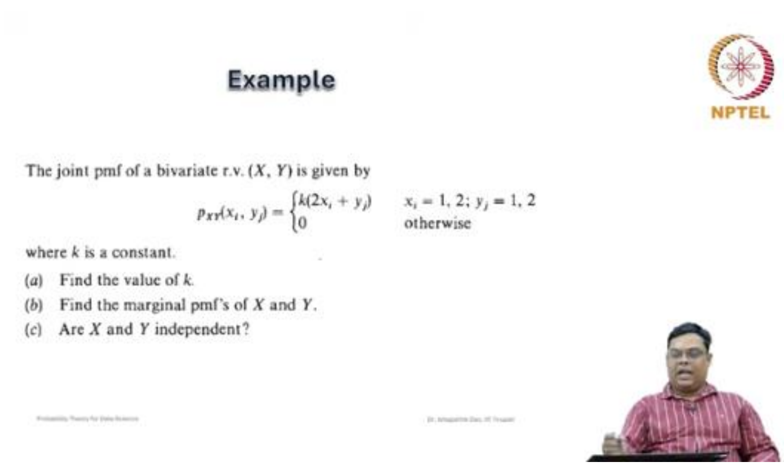
Yes, now, not only the range of (X, Y) should be checked, but for independence, it should be true for any values of i and j. Now, here you can see that for any two values you take, such as (0, 0), you can check if it is correct. For example, P_XY(0, 0) is in the range of (X, Y). P_XY(0, 0) = 4/84. Now, what is P_X(0)? P_X(0) = 35/84, and P_Y(0) = 20/84.

You can check that this is not equal to P_X(0) * P_Y(0). Why is it not equal? You can check numerically by examining the left-hand side, but just by looking at it, you can see that the right-hand side contains a factor of 5, while the left-hand side does not have a factor of 5. This is why they cannot be equal. You can also verify this numerically, which confirms the argument.

Therefore, it must be true for any values of i and j, but we can see that it is not true when both i and j are 0. Hence, X and Y are not independent random variables. So, X and Y are not independent random variables. We discussed one numerical example for the joint probability mass function. Now, let's do another numerical example.



The joint probability mass function of a bivariate random variable (X, Y) is given here. Here, this function is given, but the constant k is not provided. We need to find the value of k. The question is to find the value of k, the marginal probability mass functions of X and Y, and to determine whether X and Y are independent. I hope you understand how to solve this problem.



I will now go ahead and solve it. So, let's write down the probability mass function. The joint probability mass function of (X, Y) is given by P_XY. It is represented as P_XY(x_i, y_j), which is equal to k times (2x_i + y_j). The value of x_i can be 1 or 2, and y_j can also be 1 or 2; it is 0 otherwise. Here, k is a constant.

The first question is to find the value of k. How can we find it? We will use the fact that, since it is a joint probability mass function, it will satisfy all the properties of a joint probability mass function. So, here you can see that this will always be greater than or equal to 0, so k has to be greater than 0.

The second property is that if you take the sum of all the possible values in the range of x_i and y_j, it should be equal to 1. We will use this property: P_XY(x_i, y_j), and we will first try to find the value that makes this equal to 1. Now, this implies that P_XY(x_i) is the double sum, where x_i ranges from 1 to 2 and y_j ranges from 1 to 2. This sum should be equal to 1. So, P_XY(1, 1) + P_XY(1, 2) + P_XY(2, 1) + P_XY(2, 2) should be equal to 1.

Now, we calculate the individual terms:

P_XY(1, 1) = k * (2 * 1 + 1) = k * 3

P_XY(1, 2) = k * (2 * 1 + 2) = k * 4

P_XY(2, 1) = k * (2 * 2 + 1) = k * 5

P_XY(2, 2) = k * (2 * 2 + 2) = k * 6

Now, we need to sum these and set it equal to 1:

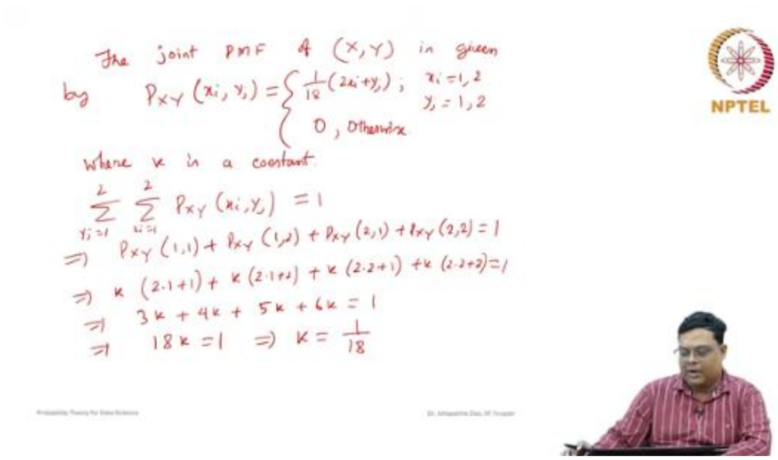
k * 3 + k * 4 + k * 5 + k * 6 = 1

k * (3 + 4 + 5 + 6) = 1

k * 18 = 1

This implies that k = 1 / 18. Now that we know the value of k, we can replace it in the probability mass function, where k = 1 / 18.

Now, the next question is that, hopefully, you understood how to find k. By using the properties, we get this equation with one unknown, and we solve this equation to find k. Now, let's find the marginal probability mass function of X and Y. So, let's first find the marginal probability mass function of X.



We know the formula: the marginal probability mass function of X, P_X of x_i, where x_i can take the values 1 and 2, is the summation over all possible values of y_j, which range from 1 to 2, of $P_{XY}(x_i, y_j)$. So, then you can see that this is nothing but P_{XY} fixing x_i. This is $P_{XY}(x_i, 1) + P_{XY}(x_i, 2)$. So, this is equal to $(1 / 18) * (2x_i + 1) + (1 / 18) * (2x_i + 2)$.

So, then finally what we are getting is $(1 / 18) * (4x_i + 3)$. This is $(4x_i + 3)$. We will write it properly because if x_i is 1 or 2, then this is true; otherwise, this is 0. So, the P_{XY} probability mass function will be written as $(1 / 18) * (4x_i + 3)$ whenever x_i is equal to 1 or 2; this is equal to 0 otherwise. So, note that it has to be a probability mass function. If we make any computational mistake, you can check it.

If you take the sum of these values, the total probability should equal 1. Because when x_i is equal to 1, the sum is 7; when x_i is equal to 2, it is 4 * 2, which is 8. 8 + 1 is 9, and 9 + 7 is 16. So, the total is 18, and (18 / 18) = 1. That is why we make the correct computation.

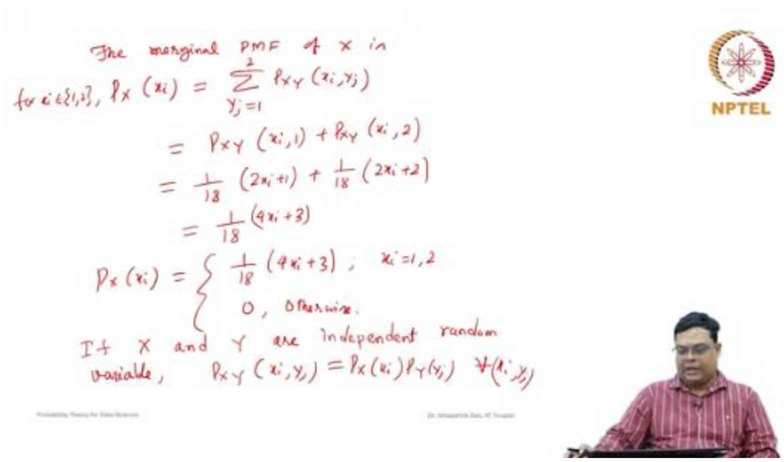
Next, we have to find the marginal probability mass function for Y. Similarly, we can find the marginal probability mass function of Y, P_Y of y_j. This is the summation of $P_{XY}(x_i, y_j)$ as x_i goes from 1 to 2. This is the marginal probability mass function. So, for y_j, fixing x_i, this is nothing but $P_{XY}(1, y_j) + P_{XY}(2, y_j)$. By definition, this is $(1 / 18) * (2x_i + y_j)$, so $(2 * 1 + y_j) + (2 * 2 + y_j)$.

This is nothing but $(1 / 18) * (2y_j + 6)$. So, this is for y_j belonging to 1 and 2; otherwise, this value is 0. So now we will write the marginal probability mass function. Finally, we have to properly write it. This is nothing but $(1 / 18) * (2y_j + 6)$, whenever y_j is equal to 1 or 2; this is equal to 0 otherwise.

So, note that it has to be a probability mass function. Then, the sum will be correct. If we make any computational mistake, we can check it. When y_j is equal to 1, this is (2 * 1 + 6), which equals 8. When y_j is equal to 2, it is (2 * 2 + 6), which equals 10, and 10 + 8 equals 18. (18 / 18) equals 1. So, this is the marginal probability mass function.

Next, we will discuss whether X and Y are independent random variables. If they are independent, then the joint probability $P_{XY}(x_i, y_j)$ should equal the product of the individual probabilities, $P_X(x_i)$ and $P_Y(y_j)$, for all (x_i, y_j). This must be true.

Now, let us check whether this is true or not. Suppose x_i is equal to 1 and y_j is equal to 1. What will happen? I have written it here, so let's move to the next page. We need to check the value of P_{XY} for (x_i, y_j). It should be equal to $P_X(x_i)$ * $P_Y(y_j)$.



Is this correct? Let's consider x_i as 1 and y_j as 1. What is $P_{XY}(1, 1)$? It is (1 / 18) multiplied by (2 * x_i + y_j). This represents the joint distribution, the joint probability mass function, which is (2 * 1 + 1).

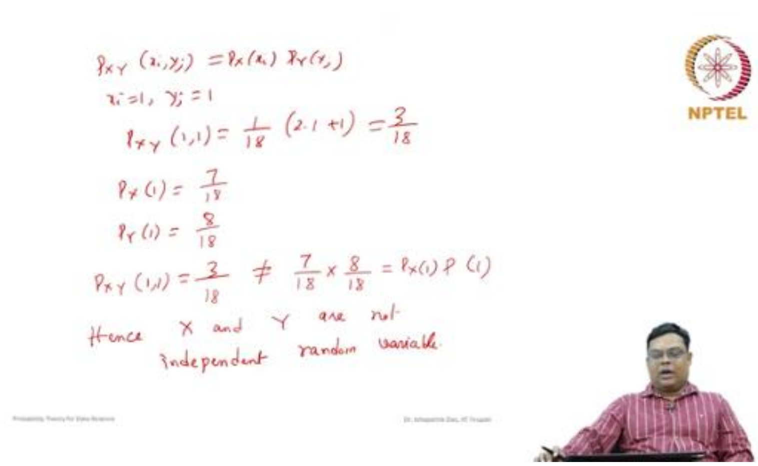
Sorry, this is not (2 * 2). It should be (2 * x_i + 1), which equals 2 + 1, or 3. So, the result is (3 / 18). Now, what about $P_X(1)$? $P_X(1)$ is (1 / 18) multiplied by (4 * 1 + 3), which equals (7 / 18). So, this is simply (7 / 18).

As for $P_Y(1)$, what is its value? $P_Y(1)$ is computed here. $P_Y(1)$ is (2 * 1), which is 2, plus 6, giving us 8. So, $P_Y(1)$ is (8 / 18). This is simply (8 / 18).

So, now you can see that this is not equal to $P_{XY}(1, 1)$. It is (3 / 18). (3 / 18) cannot be equal to (7 / 18) multiplied by (8 / 18), which is actually equal to $P_X(1)$ multiplied by $P_Y(1)$. So, $P_{XY}(1, 1)$, (3 / 18) cannot be equal to (7 / 18) * (8 / 18). You can see this numerically as well. On the right-hand side, there is a factor of 7, but on the left-hand side, after doing all the cancellations, you cannot see any factor of 7.

So, that is why it cannot be equal. If it were true for all (x_i, y_j), then they would be independent random variables. But if it is not true for at least one x_i and y_j, then we cannot say they will be independent random variables. Hence, X and Y are not independent random variables. I hope you have understood the two examples we studied and discussed the joint bivariate probability mass function here.

Next, we will discuss the joint probability density function. When X and Y are two continuous random variables, we use the joint probability density function. If both X and Y are continuous random variables, then it is called the joint continuous bivariate random variable. For bivariate continuous random variables, we discuss the joint probability density function. The joint probability density function assumes that X and Y are described by this cumulative distribution function.

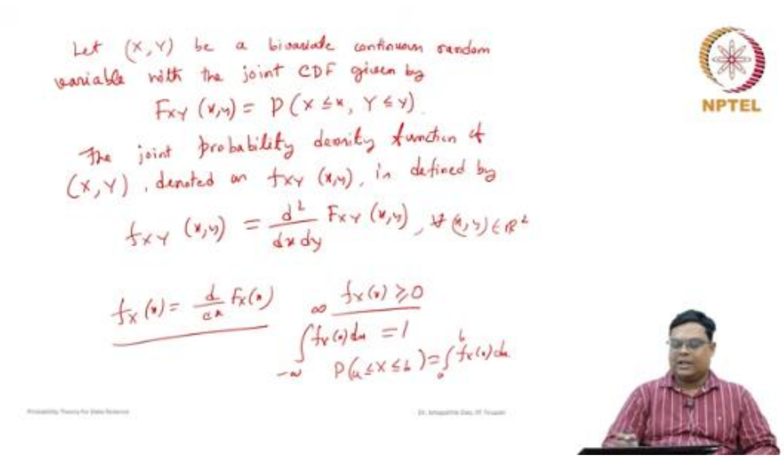


This is differentiable. Let (X, Y) be a bivariate continuous random variable with a cumulative distribution function, where the joint cumulative distribution function is given by $F_{XY}(x, y)$, representing the probability that $X \le x$ and $Y \le y$. The joint probability density function of (X, Y), denoted as $f_{XY}(x, y)$, is defined as the double derivative of $F_{XY}(x, y)$ with respect to dx and dy.

Here, it is assumed that this double derivative exists for the cumulative distribution function, and this applies to all (x, y) belonging to $\mathbb{R}^2$. The derivative is defined as shown.

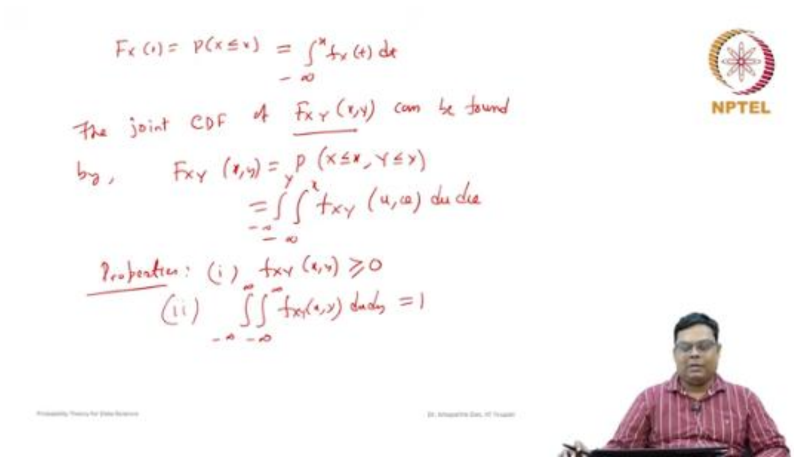
Now, analogous to the univariate case, when X is a univariate random variable, we define $f_X(x)$ as the derivative of the cumulative distribution function. For a continuous random variable, this derivative exists, and some of the properties are also discussed. One of the properties is that, because it is a non-decreasing function, $f_X(x)$, the derivative, will always be ≥ 0.

If you take the integral of $f_X(x)\,dx$ over the entire real line, it should be equal to 1. Also, to find the probability over an interval, this is simply the integral from a to b of $f_X(x)\,dx$. These are some of the properties. Similarly, for this integration, we will discuss some of the properties. Additionally, if you are given a probability density function $f_X(x)$, you can find the cumulative distribution function using this relationship: $F_X(x) = P(X \leq x) = \int$ from $-\infty$ to x $f_X(t)\,dt$.



Similarly, the joint CDF of F_{XY} can be found by $F_{XY}(x, y)$, which is the probability that $X \leq x$ and $Y \leq y$. This is equal to $\int$ from $-\infty$ to x and $\int$ from $-\infty$ to y of $f_{XY}(u, v)\,du\,dv$. This is analogous to the univariate case, where we find the joint cumulative distribution function from the probability density function. For continuous random variables, we often find it more convenient to represent them by their density functions. That is why we first introduce and discuss the density function, and then, if required, we find the cumulative distribution function from the probability density function using this formula.

So, what are the properties? Here are some of the properties. One property is that $f_{XY}(x, y)$ should be ≥ 0 because f_{XY} is non-decreasing for every x, so its derivative will be ≥ 0. The second property is that if you take the integral over the entire $\mathbb{R}^2$, from $-\infty$ to $+\infty$ for both x and y, this should be equal to 1. These are some of the properties. Similar to the univariate case, f_XY is always continuous, except at a finite number of points.



It is also a piecewise continuous function. Additionally, an important property is that if you want to find the probability over an interval, we can determine that as well. Suppose a < b and c < d, where a, b, c, and d are real numbers. You can also say a ≤ b. Now, how can we find the probability?

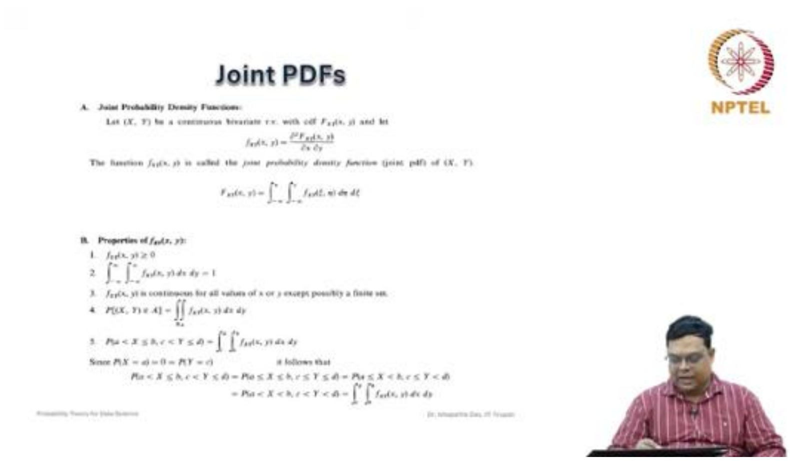
So, the probability that X ≤ b, X > a, and Y ≤ d, Y > c, can be found as follows. Since it is a continuous random variable, you can also use ≤. This probability is represented by the integral from a to b and from c to d of the probability density function fxy. So, for a two-dimensional random variable, we have to write the limits from a to b and c to d. This represents the area under the curve fxy.

It is expressed as:

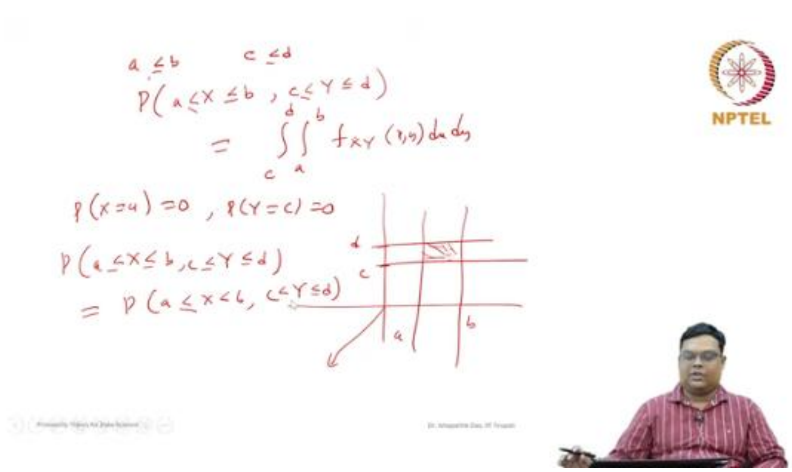
∫ (from a to b) ∫ (from c to d) fxy(dx) dy.

Now, this is the probability inside the square, from a to b. So, if you want to find the probability within the interval from a to b and c to d, this is what it looks like. Suppose this is a, this is b, this is c, and this is d. We want to find the probability of this curve, which is actually a three-dimensional curve.

So, the area, or in this case, the volume under this curve, represents the probability. If X ≤ a, and suppose Y = some value b or c, then the probability at a point is always 0. Not only that, even if you want to find a region for one coordinate, this is essentially a volume you are trying to calculate. Now, if you fix one variable as a constant and consider it as a plane, the volume you want to find will be equal to 0. So, this has already been discussed here: X = a, which equals 0.



Similarly, since X = 0, okay. So, it has already been discussed that at a point, the probability is 0. This is because, at a specific point of a random variable, the probability is always 0. If you consider this probability, it is for $Y \geq c$ and $Y \leq d$. This is the same as when you take one side as an open interval and the other side as a closed interval, or both sides as closed intervals.



The probability does not change if you modify the interval, because it is a continuous random variable. The probability at a point is always 0, which is why this is mentioned here. The joint probability density function is defined in this way. The next part is how we can find the marginal probability density function. So, suppose you already know what the joint probability density function is, and it is given to us.