PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

Indian Institute of Technology Tirupati

Week - 08

Lecture - 39

Conditional Probability Mass Function

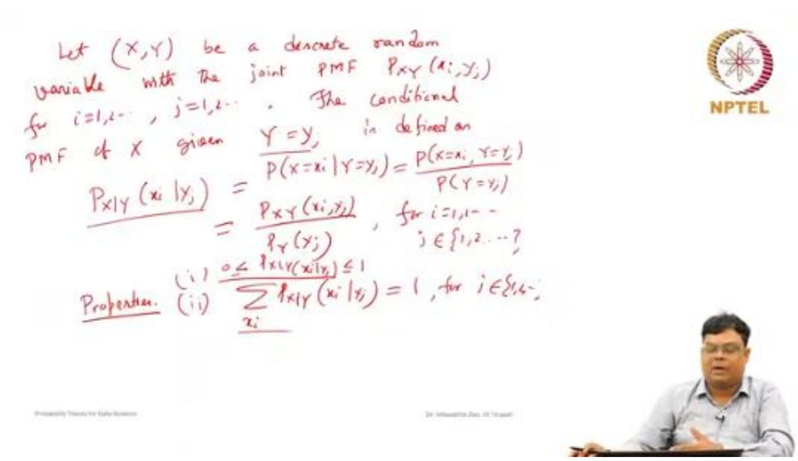
Let (X, Y) be a discrete random variable with the joint probability mass function Pxy(xi, yj), where i = 1, 2, and j = 1, 2, and so on. So, the range of Rxy may be finite or countably infinite, but we represent it as i = 1, 2, and j = 1, 2, and so on. The conditional probability mass function of X, given that Y takes some value, is defined as follows: Suppose yj is a fixed value, observed as yj. Given yj, we can define the conditional probability mass function of X given Y.

It is the probability that X = xi, given that Y = yj. So, yj is fixed, and we are considering different values that xi can take. The probability is defined as the intersection of X = A and Y = yj, divided by the probability that Y = yj. This is written as P(X = A ∩ Y = yj) / P(Y = yj). So, this is essentially the joint probability mass function divided by the marginal probability mass function.

Here, for i = 1, 2, and so on, and j is fixed, but it can take any value. That's why we can write it this way. So, now it is important to note that for a given yj, where j can be any value, we are finding the probability mass function. It will be a probability mass function, so the properties of the probability mass function must be satisfied. What are these properties?

The property is that the summation of P(X = xi | Y = yj), where xi is fixed for a particular yj, and when you sum over all values of xi, the result should be equal to 1. So, it should be clear that this is not the sum over yj. If you vary yj while fixing xi, this may not be equal to 1. This applies to the particular values of Y, where yj ∈ Y. For example, if we write j ∈ {1, 2, ..., n}, the range is finite.

However, it could also be infinite. For a particular value of j, this sum will be equal to 1, and this is a probability mass function. This is the second property, which is written here. The first property is that P(X | Y), or P(xi | yj), will always be ≥ 0 and ≤ 1. These properties are given in a similar way.



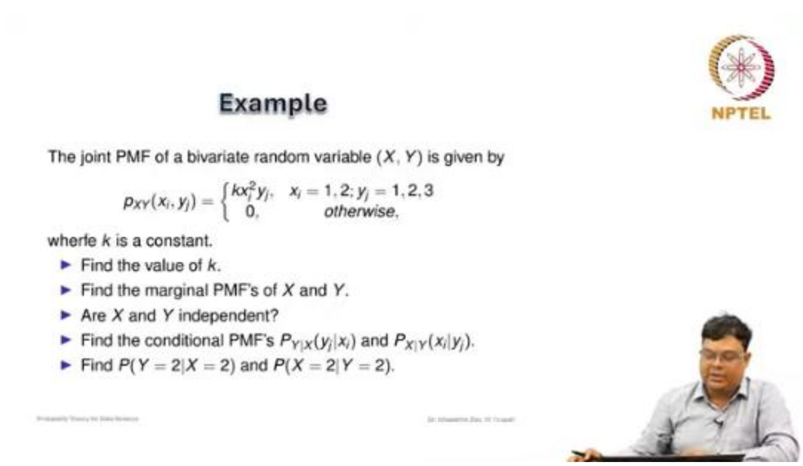
The conditional probability of X given yj is written here. Similarly, the conditional probability of Y given X can also be defined. The conditional probability, similarly, the conditional probability mass function of Y given X, is defined as the probability that Y = yj given that X = xi. Here, xi is a fixed value, and we are finding the conditional probability mass function. By definition, the probability of X = xi and Y = yj is divided by the probability of X = xi.

The assumption here is that, since it is a probability mass function, the range of X will be non-zero when X = xi. It can be represented as Pxy(xi, yj) / Px(xi), the marginal probability mass function. This is the joint probability mass function for j, which can take any value, such as 1, 2, and so on. Now, these properties have been discussed because it is a probability mass function. Since it is a probability, it must be ≥ 0 and ≤ 1.

It will satisfy all the properties of a probability mass function. For example, if you take the sum of P(Y = yj | X = xi) for a particular value of xi, and fix this xi, then the sum over yj should be equal to 1, where i ∈ {1, 2, ...}. This is given here. You can see that this condition is written because it is a probability mass function. This part has to be non-zero, meaning it must be > 0.

Similarly, this part also has to be > 0, so Px(xi) must be > 0. Otherwise, this will be undefined. This is the concept of conditional probability. Let us discuss the conditional probability mass function and how it is defined. It will be clearer when we work through some numerical examples. These examples are provided.

For instance, the joint probability mass function of a bivariate random variable X and Y is given by the function $k * xi^2 * yj$, where xi ∈ {1, 2}, and yj ∈ {1, 2, 3}. The function equals 0 otherwise. Let us write down the joint probability mass function. The joint probability mass function of the bivariate discrete random variable (X, Y) is given by Pxy(xi, yj). This is equal to $k * xi^2 * yj$, where xi ∈ {1, 2}, and yj ∈ {1, 2, 3}.



These are the values for xi and yj, and it's 0 otherwise. So, $k * xi^2 * yj$ for xi ∈ {1, 2}, and yj ∈ {1, 2, 3}, and 0 otherwise. So, this will be 0 otherwise. Now, k is a constant. We need to find the values of k.

First, we have to determine the values of k. We'll take a similar approach as before, as we have solved a similar problem earlier. We will use the properties of the probability mass function. The probability mass function must be ≥ 0, which means the constant k must be positive. If $k \leq 0$, the probability mass function would not satisfy the requirement of being non-negative, making it invalid.

Additionally, a non-positive k would fail to meet the property that the sum of all probabilities equals 1, which is a fundamental property of any probability mass function. Therefore, $k > 0$. Furthermore, the sum of the probabilities over all values of $x_i \in \{1, 2\}$ and $y_j \in \{1, 2, 3\}$ must equal 1 to ensure that the probability mass function is properly normalized.

So now, if you write k into xi, where $x_i = 1$ and $y_j = 2$, then $k * x_i^2 * y_j$ becomes:

$k * (1^2 * 1 + 2^2 * 1 + 1^2 * 2 + 2^2 * 2 + 1^2 * 3 + 2^2 * 3)$.

This is the total sum, and it should be equal to 1.

So now, let us simplify this:

$(1^2 * 1) + (2^2 * 1) + (1^2 * 2) + (2^2 * 2) + (1^2 * 3) + (2^2 * 3)$

$= 1 + 4 + 2 + 8 + 3 + 12$

$= 30$.

Thus, we have:

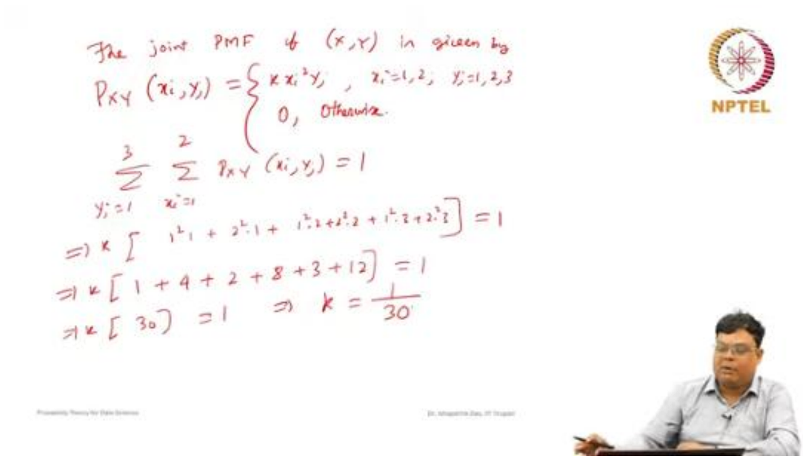
$k * 30 = 1$.

Solving this gives:

$k = 1 / 30$.

So, the probability mass function now is:

$P_{xy}(x_i, y_j) = (1/30) * x_i^2 * y_j$, where $x_i \in \{1, 2\}$ and $y_j \in \{1, 2, 3\}$, and 0 otherwise.



So, this is the probability mass function we found. Find the marginal probability mass function of X and Y. Now, we need to find the marginal probability mass function. How do we find the marginal probability mass function? By definition, to find the marginal probability mass function of X, we calculate the probability mass function of X. To do this, we need to find $P_x(x_i)$, where $x_i \in \{1, 2\}$. This is equal to the summation over $y_j \in \{1, 2, 3\}$ of $P_{xy}(x_i, y_j)$.

This is simply (1/30). Then, k = 1/30.

Next, we calculate $x_i^2 * 1 + x_i^2 * 2 + x_i^2 * 3$. This value is for $x_i \in \{1, 2\}$; otherwise, this value is 0. So, we can write for $x_i \in \{1, 2\}$; otherwise, this value is 0. So, x_i^2 becomes $(1^2 + 2^2) = 1 + 4 = 5$, and multiplying 5 by 1/30 gives 5/30 = 1/6. This simplifies to $(x_i^2 / 5) = 1/6$.

Now, we need to properly write the probability mass function. The marginal probability mass function of X is given by $P_x(x_i)$. This is $x_i^2 / 5$, whenever $x_i \in \{1, 2\}$; otherwise, this is 0.

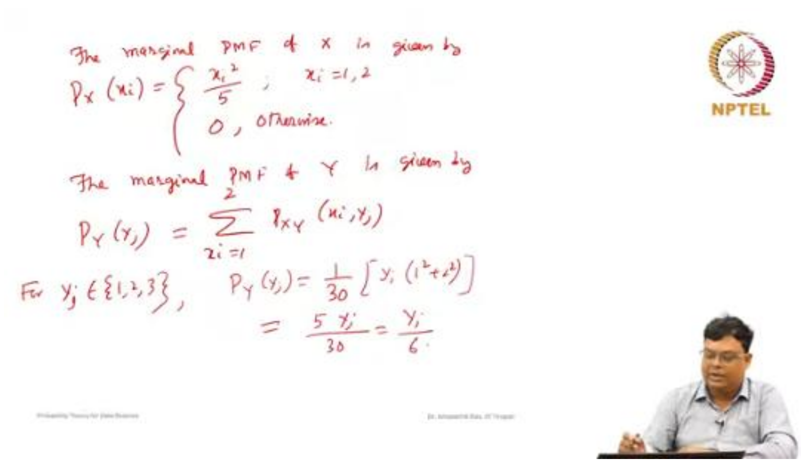
Similarly, we will find the marginal probability mass function of Y. The probability mass function of Y is given by $P_y(y_j)$. This is equal to the summation over $x_i \in \{1, 2\}$ of $P_{xy}(x_i, y_j)$.

Here, we will take the sum over the other variable, fixing y_j. So, now this will be for $y_j \in \{1, 2, 3\}$. $P_y(y_j)$ can be written as the summation over x_i.

This is (1/30) because, otherwise, this value will be 0 for any other values of yj, where yj ∈ {1, 2, 3}. It is (1/30) * xi^2 * yj. So, (1/30) * yj * xi^2. Then, yj will come in, and we have $(1^2 + 2^2)$ * yj = xi^2 * yj.

Finally, we get (5 * yj) / 30, which simplifies to yj / 6.

So, we will write this value properly. Finally, what we got is Py(yj), the marginal probability mass function of Y. This is equal to yj / 6, where yj ∈ {1, 2, 3}. We can also write it as yj ∈ {1, 2, 3}, and 0 otherwise. You have to write it properly, showing where it is non-zero, what the value is, and where it is actually equal to zero.

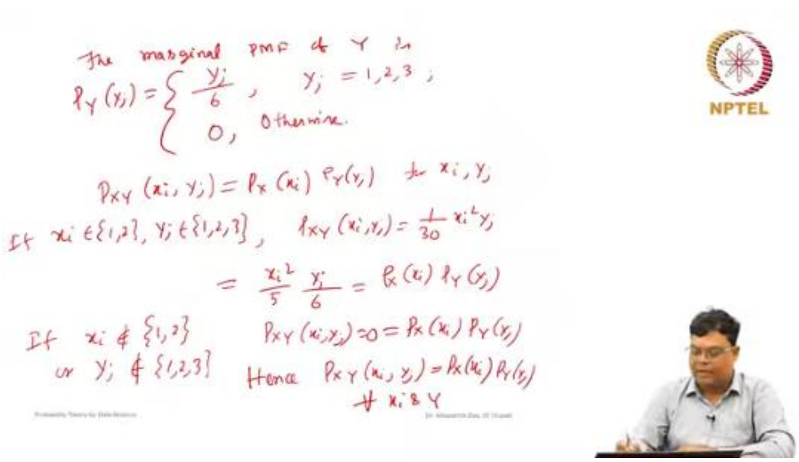


This is the marginal probability mass function. Now, the next question is: Are X and Y independent? How can we find if X and Y are independent? Suppose X and Y, or xi and yj, are independent. Then, P(xi, yj), the joint probability mass function, will be Px(xi) * Py(yj) for any values of xi and yj.

Now, you can see that if xi ∈ {1, 2}, and yj ∈ {1, 2, 3}, then Pxy(xi, yj) is simply (1/30) * xi^2 * yj. This can be represented as (xi^2 / 5) * (yj / 6). This is simply Px(xi) * Py(yj). You can see that Px(xi) is (xi^2 / 5), and Py(yj) is (yj / 6). Hence, if xi ∉ {1, 2}, or yj ∉ {1, 2, 3}, then Pxy(xi, yj) will be 0, which is equal to Px(xi) * Py(yj).

Because if xi ∉ {1, 2}, then the left-hand side, the joint probability mass function, will be 0, and Px(xi) will also be 0. Similarly, if yj ∉ {1, 2, 3}, then Py(yj) will be 0, and the left-hand side will also be 0. So, if one of them is 0, the marginals will also be 0. If neither belongs to this point, then both will be 0. In all these situations, this will hold true.

Hence, we found that $P_{xy}(x_i, y_j) = P_x(x_i) * P_y(y_j)$ for all x_i and y_j values. So, X and Y are independent random variables, which implies that X and Y are independent. Now, what is the next problem? So, X and Y are independent. Find the conditional probability mass function of Y given X, i.e., $P_y(y_j | x_i)$.



We have to find the conditional probability mass function, and then we can compute those values from the probability mass function as well. So, what will be the conditional probability mass function? By definition, the conditional probability mass function of X given some value Y is equal to y_j. So, $P(X | Y) = P_{xy}(x_i, y_j) / P_y(y_j)$.

This is $P_{xy}(x_i, y_j)$ by definition. Now, note that this will be non-zero when $y_j \in \{1, 2, 3\}$. This is nothing but, if this is true, otherwise it will be 0. It will be undefined. So, for $y_j \in \{1, 2, 3\}$, we can compute this value. Otherwise, the upper part will also be 0. Therefore, we define it as 0 because it is a probability mass function, and we must define the other values as 0.

So, the joint probability mass function, $P_{xy}(x_i, y_j)$, is nothing but $(1/30) * x_i^2 * y_j$. By the definition of $P_y(y_j)$, we know that $P_y(y_j) = y_j / 6$.

So, when we substitute this in, the y_j terms cancel out, leaving us with (x_i^2 / 5). The conditional probability mass function is defined as $P(x_i | y_j)$. This is (x_i^2 / 5), for $x_i \in \{1, 2\}$, and $y_j \in \{1, 2, 3\}$.

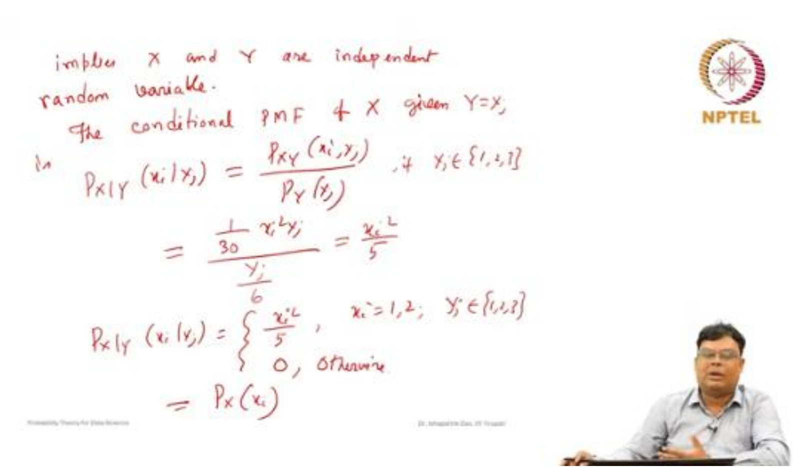
In this case, the probability mass function will be non-zero. For any other values of x_i or y_j, this value will be 0. Either $x_i \notin \{1, 2\}$, or $y_j \notin \{1, 2, 3\}$.

Note that this is the marginal probability. This is the same as the marginal probability mass function of x_i.

Since X and Y are independent random variables, the conditional probability of Y should be equal to the marginal probability of Y. So, $P(Y | X) = P_Y(y_j)$.

The probability of X taking some value should not depend on the value that Y is taking. So, the conditional probability of Y, whether it is given or not, does not matter. This is what independence means.

Similarly, we can show that the conditional probability of Y given X is the same as the probability of Y.



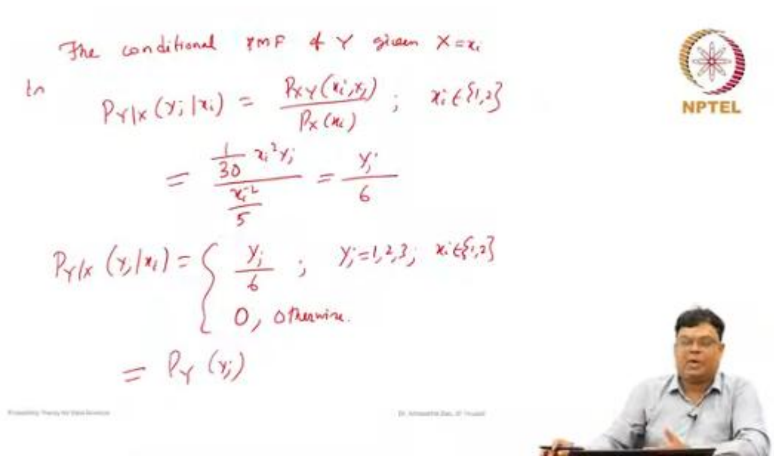
We will discuss again how the conditional probability should be the same. So now, the conditional probability mass function of Y given X, where X is equal to some value xi, is defined as P(Y | X) = Pxy(xi, yj) / Px(xi). Sorry, the probability of Y given X, with some value yj given xi, where xi is a fixed value, is defined as Pxy(xi, yj).

Note that xi must belong to {1, 2}; otherwise, Px(xi) will be 0, and the probability will be undefined. So, by definition, this is (xi^2 * yj) / Px(xi).

This is the joint probability mass function: Pxy(xi, yj) = (1/30) * xi^2 * yj. Px(xi) = (xi^2 / 5). So, this becomes (xi^2 * yj) / (xi^2 / 5). Then, this cancels out, leaving yj / 6. So, hence, the joint conditional probability mass function can be written as P(Y | X) = yj / 6, whenever yj ∈ {1, 2, 3}.

And given the condition, xi can take some value; it can be 1 or 2. This will be 0 otherwise. So, note that this function does not depend on xi. Whatever the values of xi are, whether 1 or 2, the value remains the same. It is only dependent on yj, which can be 1, 2, or 3.

This is nothing but Py(yj). So, this is the marginal probability mass function. The conditional probability mass function does not depend on xi because X and Y are independent. So, the conditional probability mass function of Y is the same as the marginal probability mass function of Y. This comes from this relationship.



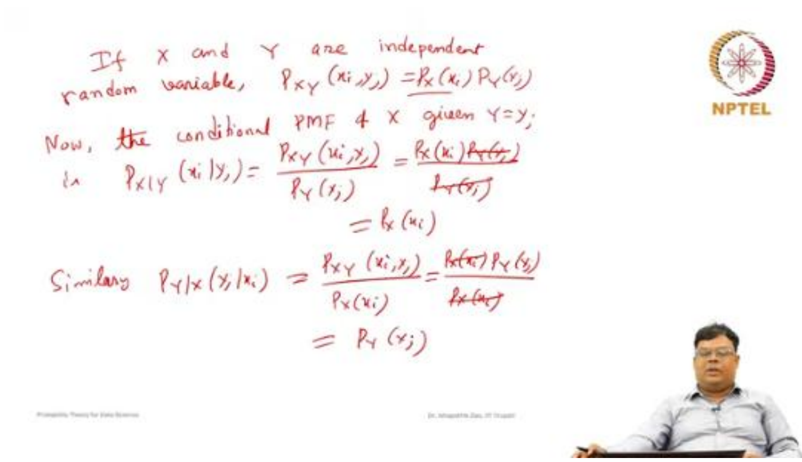
If X and Y are independent, then, based on the example, we can conclude that X and Y are independent random variables in general. So, by definition, Pxy(xi, yj) = Px(xi) * Py(yj). Now, the conditional probability mass function of X given Y, where Y = yj, is Pxy(xi, yj) / Py(yj).

Now, since they are independent, this condition is true. We can write Px(xi) * Py(yj) / Py(yj). Then this cancels, assuming that Py(yj) ≠ 0. So, this is nothing but Px(xi), which is equal to Px. Therefore, this is equal to the marginal probability mass function Px.

Similarly, it can be shown that P(Y | X), or Py(yj | xi), by definition is Pxy(xi, yj) / Px(xi). This is the definition. But since X and Y are independent, this is nothing but Px(xi) * Py(yj) / Px(xi). So, this cancels, leaving Py(yj).

So, if X and Y are independent, we do not need to compute the conditional probability mass function. We can directly say that, because they are independent, the conditional probability mass function should be equal to their marginal probability mass function.

So, now for the last problem, we will apply this result without computing it.

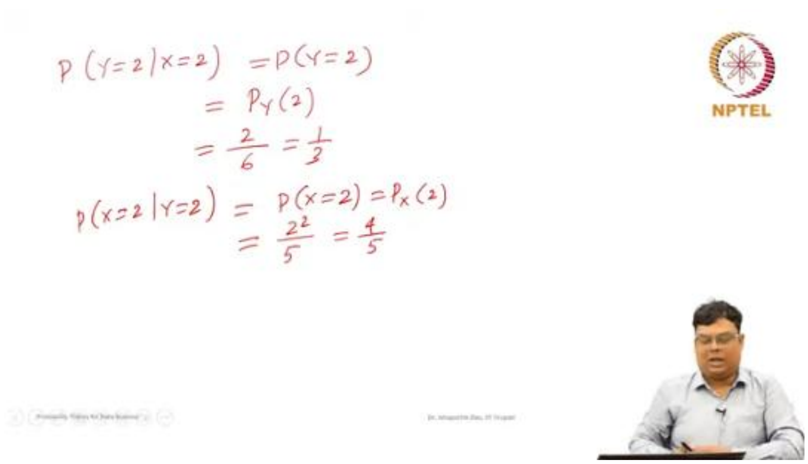


Using the definition, since it asks for the probability of Y = 2 given X = 2, the probability of Y = 2 given X = 2. Since we have seen that X and Y are independent, this will be equal to the probability of Y = 2 only. So, we do not need to compute A given B, intersection B, or any of these things. The conditional probability mass function is simply the probability of Y = 2. The probability mass function of Y is the probability of yj / 6.

So, when yj = 2, the result is 2 / 6, which simplifies to 1/3. Similarly, if you want to find the probability that X = 2 given Y = 2, this is simply the probability of X = 2 given Y = 2.

Since X and Y are independent, this is equivalent to the probability mass function of X evaluated at 2. The probability mass function of X is given as $xi^2 / 5$. So, we will write $xi^2 / 5$.

With xi = 2, this becomes $2^2 / 5$. This is equivalent to 4 / 5. So, this is the problem we discussed using the conditional probability mass function. I hope you are following along and understand the concept of how to find the conditional probability mass function when X and Y are discrete random variables. We also discussed their properties, how to check their independence with respect to conditional probability mass functions, and how to compute the respective probabilities, as demonstrated in the numerical examples. Next, we will discuss what happens if X and Y are two random variables that are continuous.



In that case, we need to discuss the joint probability density function, which we have already covered. Here, we will focus on the conditional probability density function—how it is defined, its properties, and how we can compute the related probabilities with respect to the conditional probability density function for continuous bivariate random variables.