PROBABILITY THEORY FOR DATA SCIENCE

Prof. Ishapathik Das

Department of Mathematics and Statistics

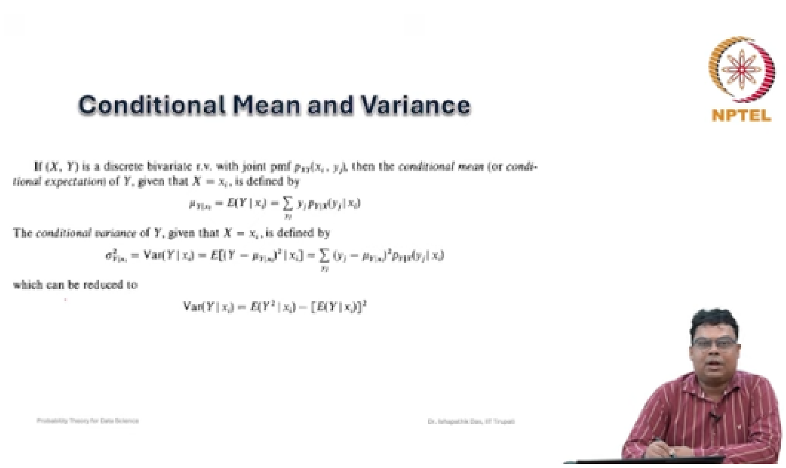
Indian Institute of Technology Tirupati

Week - 09

Lecture - 44

Conditional Mean and Variance for Discrete Random Variables

We will discuss how to find the conditional mean and conditional variance. Previously, we explored how to determine the conditional probability mass function for discrete cases and the conditional probability density function for continuous random variables. Now, let us focus first on discrete random variables. Consider two discrete bivariate random variables, X and Y. These variables have a joint probability mass function, P(X, Y), that defines the probability of specific values for X and Y occurring together.



The marginal probability mass function of X describes the probability of each value of X occurring, and the marginal probability mass function of Y describes the probability of each value of Y occurring. Let X and Y be discrete bivariate random variables with a joint probability mass function, P(X = xi, Y = yj), that defines the probability of X taking a value xi and Y taking a value yj. The marginal probability mass function of X provides the probability of each value of X, $P(X = x_i)$, and the marginal probability mass function of Y provides the probability of each value of Y, $P(Y = y_j)$. If the range of X is finite, it will go from 1 to some n, and if the range of Y is finite, it will go from 1 to some m. Otherwise, the ranges of X and Y may be countably infinite.

The conditional probability mass function of Y given that $X = x_i$ is defined for any value of y_j in the range of Y and for a particular value of x_i in the domain of X. It is calculated as the joint probability of X and Y taking the values x_i and y_j, respectively, divided by the marginal probability of X taking the value x_i. This is expressed as:

$P(Y = y_j \mid X = x_i) = P(X = x_i, Y = y_j) / P(X = x_i)$.

This assumes that $P(X = x_i) \neq 0$, as dividing by zero is undefined. This is why x_i must be within the range of X, as only in this case will $P(X = x_i)$ be greater than zero. Otherwise, $P(Y = y_j \mid X = x_i) = 0$.

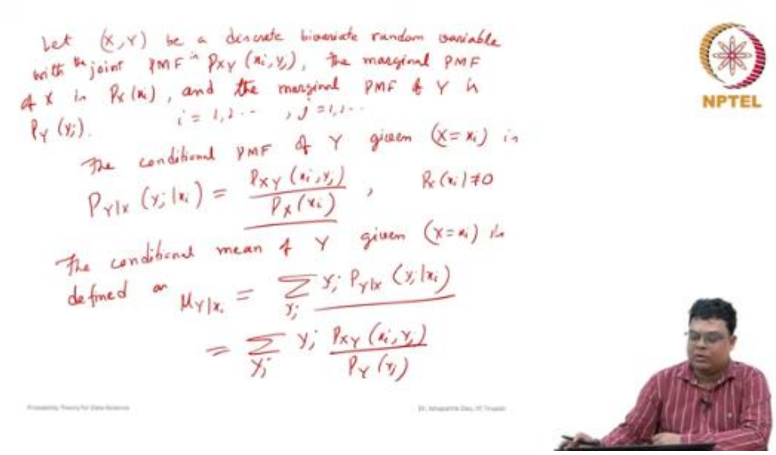
This defines the conditional probability mass function. The conditional mean of Y given that $X = x_i$ is defined as the expected value of Y under this condition. The conditional mean, or conditional expectation, is denoted as $E(Y \mid X = x_i)$ or $\mu_{Y|x_i}$. It represents the expected value of Y given that $X = x_i$. The conditional mean is calculated as the summation of all values of y_j, each multiplied by the conditional probability mass function of Y given $X = x_i$:

$E(Y \mid X = x_i) = \Sigma [y_j * P(Y = y_j \mid X = x_i)]$, where the summation is over all possible values of y_j.

To simplify, the conditional probability mass function can be expressed as the joint probability of X and Y taking specific values, divided by the marginal probability of X taking the value x_i. Using this, you can first calculate the conditional probability mass function and then compute the conditional mean as needed.

The variance, denoted as $\sigma^2_{Y|x_i}$, represents the variance of Y given x_i. It is calculated as the expected value of the square of the difference between Y and the conditional mean of Y given x_i:

$\sigma^2_{Y|x_i} = E[(Y - E(Y \mid X = x_i))^2 \mid X = x_i]$.



This formula can be simplified in the same way as before. Let's write it down here. The conditional variance of Y given X = xi is defined as the variance of Y given xi. This is the variance of Y given xi, which is the expected value of $(Y - E(Y \mid X = xi))^2$. It can be represented as:

$\sigma^2\_Y|xi = \Sigma [(yj - E(Y \mid X = xi))^2 * P(Y = yj \mid X = xi)]$,

where the summation is over all possible values of yj. Each value of yj is subtracted from the conditional mean of Y given xi, squared, and then multiplied by the conditional probability mass function of Y given X = xi.

This expression can be simplified as the expected value of Y^2 given xi, minus the square of the expected value of Y given xi:

$\sigma^2\_Y|xi = E(Y^2 \mid X = xi) - (E(Y \mid X = xi))^2$.

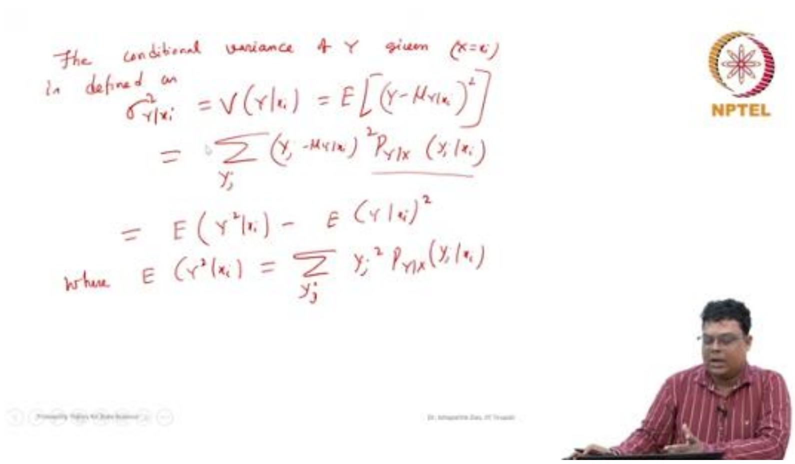
We have not yet defined the notation, but the expected value of Y given xi can be represented with a double notation. The expected value of Y^2 given xi is:

$E(Y^2 \mid X = xi) = \Sigma [yj^2 * P(Y = yj \mid X = xi)]$,

where the summation is over all possible values of yj. This is the formula for the conditional variance of Y given X = xi. Essentially, the conditional mean and conditional variance can be understood by treating the probability mass function as the conditional probability mass function. By applying the usual formulas for mean and variance, you can calculate the conditional mean and conditional variance.

Now, you might ask how we can define the conditional mean and variance in the opposite direction. If you want to find the conditional mean of X given that $Y = y_j$, you simply swap X and Y. That is, you replace X with Y and Y with X. So, the conditional mean of X given $Y = y_j$ is the expected value of X given y_j, which is denoted as:

$E(X \mid Y = y_j)$.



It is the sum of all possible values of x_i, each weighted by the conditional probability mass function of X given $Y = y_j$. Now, for the conditional variance of X given $Y = y_j$, this is defined as the variance of X given y_j. The conditional variance of X is the expected value of $(X - E(X \mid Y = y_j))^2$. This can be expressed as:

$\sigma^2\_X|y_j = \Sigma\ [(x_i - E(X \mid Y = y_j))^2 * P(X = x_i \mid Y = y_j)]$,

where the summation is over all possible values of x_i. Each x_i is subtracted from the conditional mean of X given y_j, then squared, and multiplied by the conditional probability mass function of X given $Y = y_j$. This expression is similar to the earlier one, but with x_i and y_j swapped. Note that in this case, y_j is fixed, just as x_i was fixed in the earlier cases.

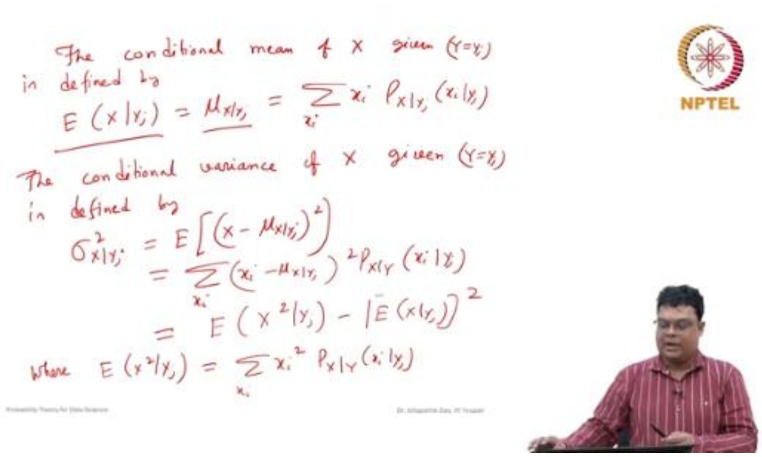
For computational purposes, calculating this directly can be complicated, especially with numerical examples, as it can be computationally intensive. Therefore, we use the simplified formula that we have already shown, which applies to general random variable cases. For the conditional probability mass function, the process is similar. It is essentially the second-order raw moment of Y^2, but in this case, it is the second-order raw moment of

X^2 given yj, minus the expected value of X given yj, squared. The expected value of X given yj has already been denoted as $E(X \mid Y = yj)$.

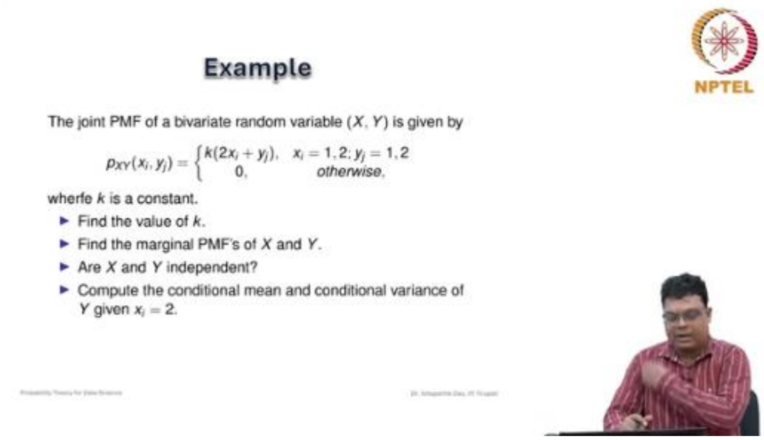
The expected value of X^2 given yj is defined as:

$E(X^2 \mid Y = yj) = \Sigma\, [xi^2 * P(X = xi \mid Y = yj)]$,

where the summation is over all possible values of xi. I hope this clarifies the concept. For further understanding, we will need to work through a numerical example. Let us do one numerical example. Before moving on to the continuous cases, let's complete the discussion by working through a numerical example. In this example, the joint probability mass function of a bivariate random variable (X, Y) is given by $Pxy(xi, yj) = k * (2xi + yj)$.



This example has already been covered, but we are repeating it to work through additional problems related to computing the conditional mean and conditional variances. Although this problem has been done before, reviewing it again will help clarify the concepts further. You can go back and check the previous solution, but we will work through it again here for better understanding. The joint probability mass function of X and Y is given by Pxy(xi, yj), not the conditional probability mass function as previously mentioned. The joint probability mass function of X and Y is given by $Pxy(xi, yj) = k(2xi + yj)$, where $xi \in \{1, 2\}$ and $yj \in \{1, 2\}$.

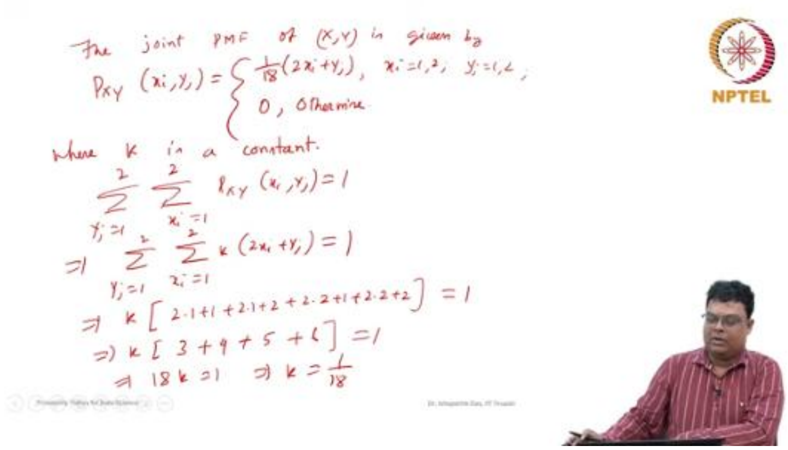


Otherwise, the probability is 0, where k is a constant. We need to find the value of k. To find k, we use the properties of the joint probability mass function. Since k is a probability mass function, it must always be greater than or equal to 0. We find k using the property that the sum of all probabilities must equal 1.

So, we sum over all possible values of xi (1 and 2) and yj (1 and 2). This gives us: k * Σ(2xi + yj) over all xi and yj values should equal 1. Expanding this, we have: k * [(2(1) + 1) + (2(1) + 2) + (2(2) + 1) + (2(2) + 2)]. This simplifies to:

k * (3 + 4 + 5 + 6) = 1. Thus, k * 18 = 1, which implies k = 1/18. Therefore, the probability mass function is 1/18.

Now, to find the marginal probability mass function of X and Y, the probability mass function of X is given by Px(xi) for xi. By definition, this is the sum over all yj values, from 1 to 2, of Pxy(xi, yj). However, it will be non-zero only for xi equal to 1 or 2; otherwise, the value will be 0.

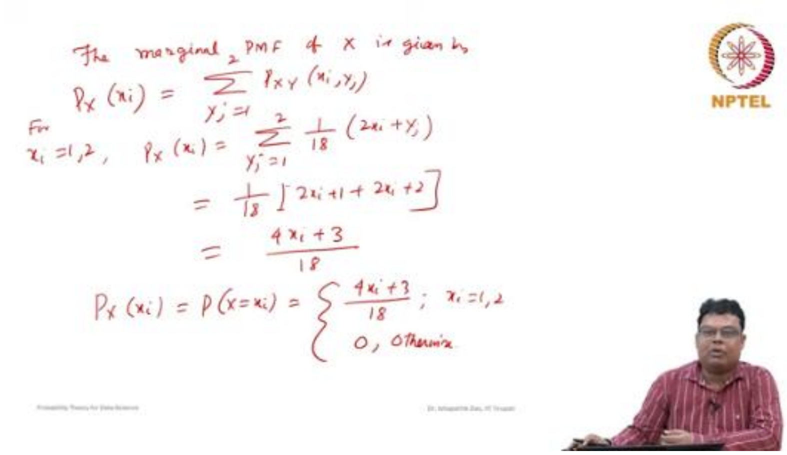


$P_X(x_i) = \sum P_{X\gamma}(x_i, y_j)\ \forall\ y_j \in \{1, 2\}$. This is $(1/18) * (2x_i + y_j)$. We can calculate this as: $(1/18) * (2x_i + 1) + (2x_i + 2)$, which simplifies to: $(1/18) * (4x_i + 3)$.

Therefore, the marginal probability mass function of X is given by $P_X(x_i) = (4x_i + 3) / 18$ for $x_i \in \{1, 2\}$, and 0 otherwise. You can check if this is correct by $\sum$ over $x_i = 1$ and 2, which gives $7/18 + 11/18 = 18/18$, confirming that this is a valid probability mass function.

Next, we will find the marginal probability mass function of Y. We have already worked with these values earlier, so you can refer to that or recalculate them if needed. For the conditional probability mass function, we will compute the conditional mean and conditional variance using the formulas.

We have to first find the conditional probability mass function. Now, the marginal probability mass function of Y is given by the probability of $Y = y_j$. This is $\sum P_{X\gamma}(x_i, y_j)\ \forall\ x_i$. The function is non-zero only when $y_j \in \{1, 2\}$; otherwise, the value will be 0. The probability of $Y = y_j$ is the $\sum P_{X\gamma}(x_i, y_j)\ \forall\ x_i$.

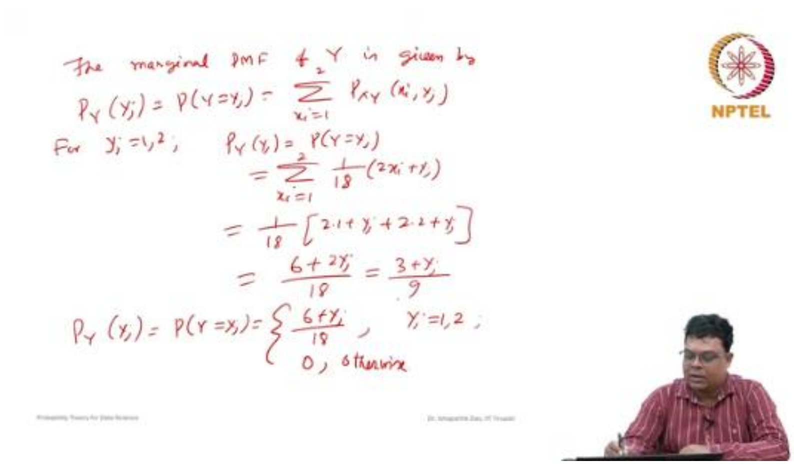


This is calculated as (1/18) * ($2x_i + y_j$). For $x_i = 1$, we get $2(1) + y_j$, and for $x_i = 2$, we get $2(2) + y_j$. This gives us:

For $x_i = 1$: $2 + y_j$,

For $x_i = 2$: $4 + y_j$.

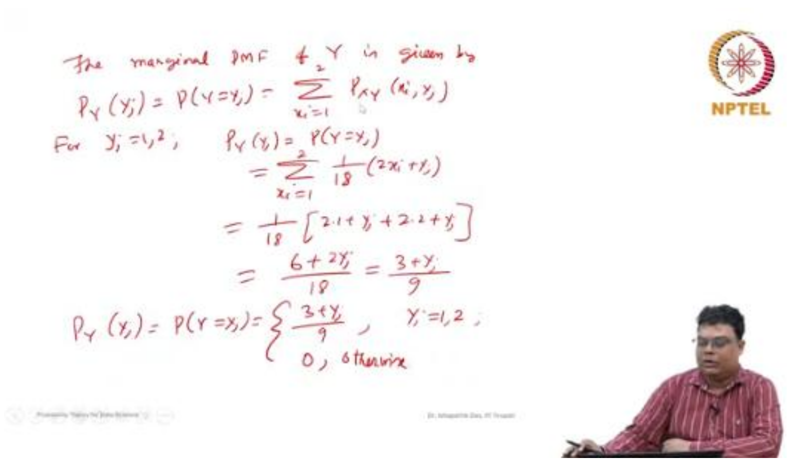
Thus, the marginal probability mass function of Y is given by: $P(y_j) = (6 + y_j) / 18$, for $y_j \in \{1, 2\}$, and 0 otherwise. You can verify this by checking the probabilities. For $y_j = 1$, $P(y_j) = (6 + 1) / 18 = 7/18$, and for $y_j = 2$, $P(y_j) = (6 + 2) / 18 = 8/18$. This confirms that the marginal probability mass function is correct.



So, the marginal probability mass function of Y is $(3 + y_j) / 9$. I made a mistake earlier, it should be $2y_j$, but now it is $3 + y_j$. Let me check if this is indeed a valid probability mass function. If you sum the probabilities for y_j, the total should equal 1. When $y_j = 1$, the value is 3 + 1, which is 4.

When $y_j = 2$, the value is 3 + 2, which is 5. Adding these gives 4 + 5, which equals 9. So, the sum is 9, and dividing by 9 gives 1, confirming that this is the correct marginal probability mass function for y_j. Now, are X and Y independent? If X and Y are independent, then $P_{x\gamma}(x_i, y_j)$ should be equal to $P_x(x_i) * P_\gamma(y_j)$ $\forall$ x_i, y_j values. This should hold true $\forall$ possible values of x_i and y_j.

To clarify, let's consider some specific values. Suppose $x_i = 1$ and $y_j = 1$. What is $P_{x\gamma}(1,1)$? The joint probability $P_{x\gamma}(1,1)$ is $(1/18) * (2x_i + y_j)$. Substituting the values, we get 2 * 1 + 1, which equals 3.

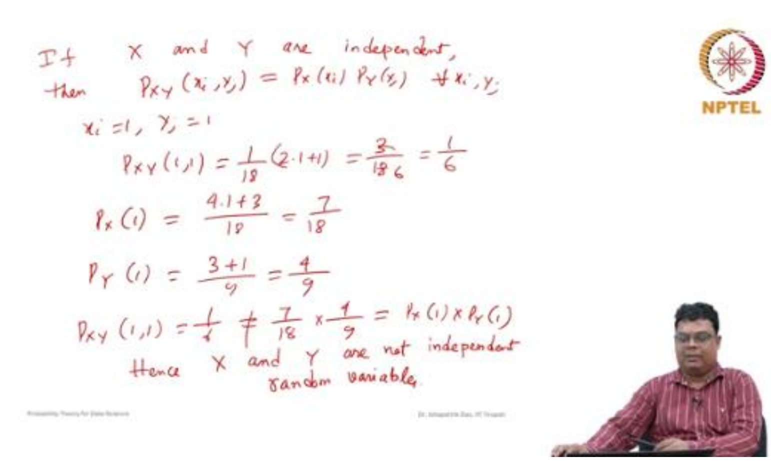


So, $P_{x\gamma}(1,1) = 3/18$. Now, what about $P_x(1)$? $P_x(1) = (4x_i + 3) / 18$. For $x_i = 1$, this becomes 4 * 1 + 3 = 7. So, $P_x(1) = 7/18$.

Next, let's find $P_\gamma(1)$. $P_\gamma(1) = (3 + y_j) / 9$. For $y_j = 1$, this becomes 3 + 1 = 4. So, $P_\gamma(1) = 4/9$.

To check for independence, we need to see if $P_{x\gamma}(1,1) = P_x(1) * P_\gamma(1)$. $P_x(1) * P_\gamma(1) = (7/18) * (4/9) = 28/162 = 14/81$. However, $P_{x\gamma}(1,1) = 3/18 = 1/6$.

Clearly, $1/6 \neq 14/81$, so the two sides are not equal. Therefore, we conclude that X and Y are not independent random variables. Now, we need to compute the conditional mean and conditional variance of Y given that X = 2. To compute these, we need the conditional probability mass function. We have already discussed the formulas for conditional mean and variance.



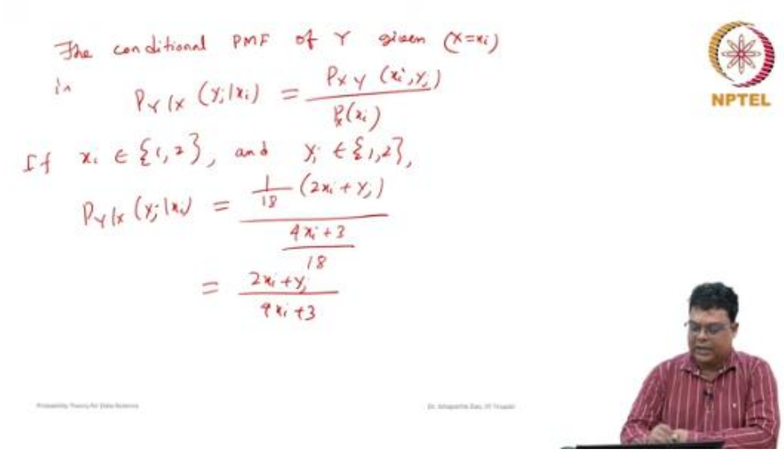
For this, we need the conditional probability mass function of Y given X. The conditional probability mass function of Y given X = x_i, for specific values of x_i and y_j, is defined as: $P(Y = y_j \mid X = x_i) = P(x_i, y_j) / P(X = x_i)$.

This holds for $x_i \in \{1, 2\}$ and $y_j \in \{1, 2\}$. The joint probability mass function is non-zero only when $x_i \in \{1, 2\}$ and $y_j \in \{1, 2\}$.

In this case, we are finding the conditional probability mass function for specific values of x_i and y_j. If x_i is fixed (e.g., $x_i = 2$), we compute the conditional probability for each $y_j \in \{1, 2\}$. The joint probability mass function is given as: $P(x_i, y_j) = (1/18) * (2x_i + y_j)$.

This is divided by the marginal probability mass function of X, which we have already computed as: $P(X = x_i) = (4x_i + 3) / 18$.

So, the conditional probability mass function is: $P(Y = y_j \mid X = x_i) = [(1/18) * (2x_i + y_j)] / [(4x_i + 3) / 18]$.

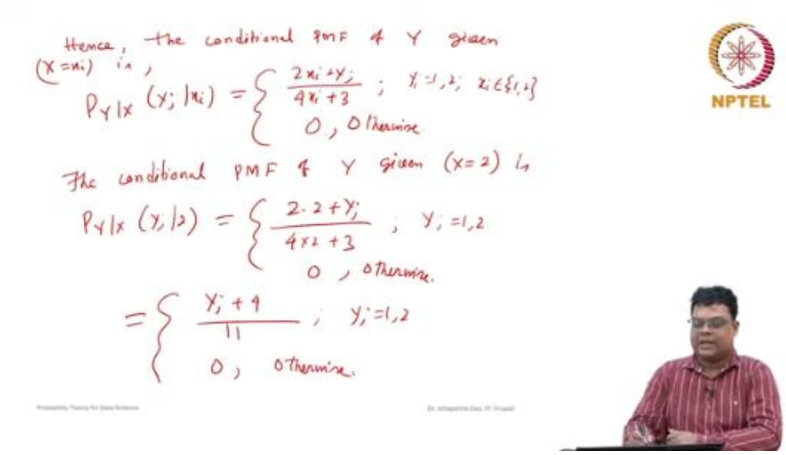


After canceling, we get the simplified expression: the conditional probability mass function of Y given X = x_i is the ratio of ($2x_i + y_j$) to ($4x_i + 3$). Thus, the conditional probability mass function of Y given X = x_i is ($2x_i + y_j$) / ($4x_i + 3$). This holds when $y_j \in \{1, 2\}$, and for the specific values of $x_i \in \{1, 2\}$; it is zero for other cases.

Now, to compute the conditional mean and conditional variance of Y given X = 2, we substitute $x_i = 2$ into the formula for the conditional probability mass function of Y. Therefore, the conditional probability mass function of Y given X = 2 is ($4 + y_j$) / 11.

This expression holds when $y_j \in \{1, 2\}$, and is zero otherwise. Finally, simplifying the expression for the conditional probability mass function, we get ($y_j + 4$) / 11 when $y_j \in \{1, 2\}$, and 0 for other values.

This shows that the conditional probability mass function is clearly defined, where x_i acts as a constant parameter and y_j is the variable. When $x_i = 2$, the expression becomes a probability mass function for y_j. For $y_j = 1$, the probability is (4 + 1) / 11 = 5/11, and for y_j = 2, the probability is (4 + 2) / 11 = 6/11.



Adding these probabilities gives 5/11 + 6/11 = 11/11 = 1, confirming the correctness of the computation.

Now, for the next step, we need to compute the conditional mean and conditional variance of Y given that $x_i = 2$. To do this, we use the definition of conditional mean, which is the sum of the possible values of y_j multiplied by their respective probabilities. For the conditional mean, we multiply each value of y_j by its probability. For $y_j = 1$, the probability is 5/11, and for $y_j = 2$, the probability is 6/11.

The conditional mean will be the sum of these values, which is the expected value of Y given $x_i = 2$.

Similarly, the conditional variance can be computed using the formula for variance, which involves squaring the difference between each value of y_j and the conditional mean, and then multiplying by their respective probabilities. This will give the conditional variance of Y given $x_i = 2$.