Probability Theory Lecture 12

Padmini Mukkamala

Budapest University of Technology and Economics

November 29, 2020

Overview

Joint Conditional Mass Function

Discrete case: Joint Conditional PMF

For two random variables X, Y, the joint conditional probability mass function of X conditioned on a specific value of $Y = y$ where $p_Y(y) \neq 0$, is given by,

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

We must check that this function $p_{X|Y}(\cdot|y)$ is a proper probability mass function for the variable X. We can do so by checking,

$$\sum_x p_{X|Y}(x|y) = 1$$

Further properties

We notice that if X and Y are independent random variables, then,

$$p_{X|Y}(x|y) = p_X(x)$$

Joint Conditional CDF

The joint conditional cumulative distribution function, then, can be computed as follows:

$$F_{X|Y}(a|y) = \sum_{x \le a} p_{X|Y}(x|y)$$

Example - 1

This is example $4a$ from the book: Consider the following joint pmf table.

$X \backslash Y$	0	1	$p_X(x)$
0	0.4	0.2	0.6
1	0.1	0.3	0.4
$p_Y(y)$	0.5	0.5	1

Table: Joint pmf of X, Y

What is the joint conditional pmf $p_{X|Y}(x|1)$?

Example - 2

Let X the sum of two independent dice rolls, and let Y denote the outcome of the first dice roll. What is the conditional pmf of $p_{Y|X}(y|5)$ and $p_{Y|X}(y|7)$?
What is $p_{Y|X}(y|x)$, where now we try to find the joint pmf as a function of x?

Example - 2 cont.

Y	$p_{Y\|X}(y\|5)$
1	$\frac{1}{4}$
2	$\frac{1}{4}$
3	$\frac{1}{4}$
4	$\frac{1}{4}$
5	0
6	0

Y	$p_{Y\|X}(y\|7)$
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$

Table: Conditional pmf of $p_{Y|X}(y|5)$ and $p_{Y|X}(y|7)$

Example - 2 cont.

We can check that the joint pmf, as a function of x can be denoted as: For $2 \leq x \leq 7$,

$$p_{Y|X}(y|x) = \begin{cases} \frac{1}{x-1} & 1 \leq y \leq x-1 \\ 0 & \text{otherwise} \end{cases}$$

and for $7 < x \leq 12$,

$$p_{Y|X}(y|x) = \begin{cases} \frac{1}{13-x} & x-6 \leq y \leq 6 \\ 0 & \text{otherwise} \end{cases}$$

Example - 3

Let a fair dice be rolled n times. Let $X \sim Bin(n, \frac{1}{6})$ denote the number of times 4 appears in these n dice rolls. Also, let $Y \sim Bin(n, \frac{1}{3})$ be the number of times a perfect square $(1, 4)$ appears in these n dice rolls. What is the distribution of $p_{X|Y}(x|y)$?

Example - 3

Let a fair dice be rolled n times. Let $X \sim Bin(n, \frac{1}{6})$ denote the number of times 4 appears in these n dice rolls. Also, let $Y \sim Bin(n, \frac{1}{3})$ be the number of times a perfect square $(1, 4)$ appears in these n dice rolls. What is the distribution of $p_{X|Y}(x|y)$? $Bin(y, \frac{1}{2})$.

What is the distribution of $p_{Y|X}(y|x)$?

Example - 3

Let a fair dice be rolled n times. Let $X \sim Bin(n, \frac{1}{6})$ denote the number of times 4 appears in these n dice rolls. Also, let $Y \sim Bin(n, \frac{1}{3})$ be the number of times a perfect square $(1, 4)$ appears in these n dice rolls. What is the distribution of $p_{X|Y}(x|y)$? $Bin(y, \frac{1}{2})$.

What is the distribution of $p_{Y|X}(y|x)$? $Bin(n - x, \frac{1}{5})$.

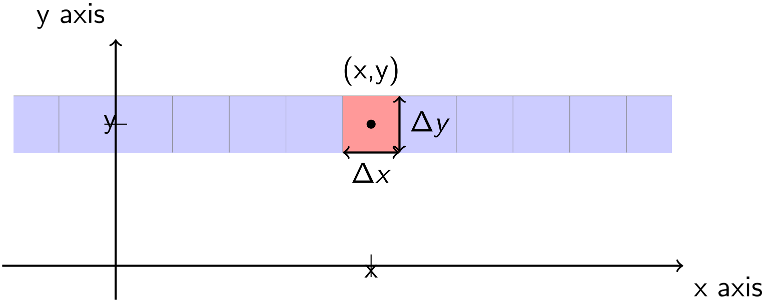
Continuous Case: Joint Conditional PDF

For two continuous random variables X, Y with joint pdf given by $f_{X,Y}(x, y)$, for the values of $Y = y$ where the density function $f_Y(y) \neq 0$, there the joint conditional probability density function is given by,

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Proof: We look at a small rectangle of side lengths Δx and Δy in the neighborhood of the point (x, y). In the picture, we need the conditional probability of the small red rectangle, conditioned on the the fact that we are in the big blue rectangle. To get density, this will eventually be divided by Δx and we'll take the limit as $\Delta x \to 0$.

Continuous Case: Joint Conditional PDF



Continuous Case: Joint Conditional PDF

So the conditional probability of the pink rectangle, conditioning on the blue rectangle is: $P(X \in [x - \frac{\Delta x}{2}, x + \frac{\Delta x}{2}] | Y = y) = \lim_{\Delta y \to 0} \frac{f_{X,Y}(x,y)\Delta x \Delta y}{f_Y(y)\Delta y} = \frac{f_{X,Y}(x,y)\Delta x}{f_Y(y)}$

Then, $f_{X|Y}(x|y) = \lim_{\Delta x \to 0} \frac{P(X \in [x - \frac{\Delta x}{2}, x + \frac{\Delta x}{2}] | Y = y)}{\Delta x} = \lim_{\Delta x \to 0} \frac{1}{\Delta x} \frac{f_{X,Y}(x,y)\Delta x}{f_Y(y)} = \frac{f_{X,Y}(x,y)}{f_Y(y)}$

Continuous Case: Joint Conditional CDF

For the joint conditional cumulative distribution function, we have to note that y in $f_{X|Y}(x|y)$ is treated as a constant, and this is entirely a function of x, so,

$$F_{X|Y}(a|y) = \int_{-\infty}^{a} f_{X|Y}(x|y)dx$$

Examples

Let $f_{X,Y}(x,y) = \begin{cases} \frac{12}{5}y(2-x-y) & 0<x<1, 0<y<1 \\ 0 & \text{otherwise} \end{cases}.$

What is $f_{Y|X}(y|x)$?

Examples cont.

Let $f_{X,Y}(x,y) = \begin{cases} \frac{e^{-\frac{x}{y}} e^{-y}}{y} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}$.

What is $F_{X|Y}(x|y)$?

Use this to find $P(X > 1|Y = y)$.

Conditional Expectation

Definition

The conditional expected value of a discrete random variable Y, conditioned on $X = c$ is as follows:

$$E(Y|X = c) = \sum_y y p_{Y|X}(y|c)$$

while if Y is continuous, then its given by,

$$E(Y|X = c) = \int_{-\infty}^{\infty} y f_{Y|X}(y|c) dy$$

Regression

When we find the conditional expectation $E(Y|X = x)$ where instead of a constant c, we find the expected value in terms of variable x, we notice that then $E(Y|X = x)$ is a function of x.
This function is called the **Regression** of the variable Y in terms of X.
We will see (after proving the Law of Total Expectation,) that the regression of Y in terms of X is the best possible 'guess' for Y as a function of X, according to one error heuristic.
Here Y is called the 'dependent' variable, while X is called the 'independent' variable.

Example - 1

Again example 4*a* from the book: Consider the following joint pmf table.

$X \backslash Y$	0	1	$p_X(x)$
0	0.4	0.2	0.6
1	0.1	0.3	0.4
$p_Y(y)$	0.5	0.5	1

Table: Joint pmf of X, Y

What is $E(X|Y=1)$?

Example - 2

Let X the sum of two independent dice rolls, and let Y denote the outcome of the first dice roll. What is $E(Y|X=5)$ and $E(Y|X=7)$?

Example - 2

Let X the sum of two independent dice rolls, and let Y denote the outcome of the first dice roll. What is $E(Y|X=5)$ and $E(Y|X=7)$? $\frac{5}{2}, \frac{7}{2}$.
What is the regression function of Y in terms of X, i.e., what is $E(Y|X)$?

Example - 2

Let X the sum of two independent dice rolls, and let Y denote the outcome of the first dice roll. What is $E(Y|X=5)$ and $E(Y|X=7)$? $\frac{5}{2}, \frac{7}{2}$.
What is the regression function of Y in terms of X, i.e., what is $E(Y|X)$? $\frac{X}{2}$.

Examples cont.

Let $f_{X,Y}(x,y) = \begin{cases} \frac{e^{-\frac{x}{y}} e^{-y}}{y} & 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise} \end{cases}.$

What is $E(X|Y = y)$?

Properties of Regression

- For any function $h(X)$, $E(h(X)Y|X) = h(X)E(Y|X)$. In particular, $E(h(X)|X) = h(X)$.
- $E(aY + b|X) = aE(Y|X) + b$.
- $E((Y - g(X))^2)$ is **minimum** for $g(X) = E(Y|X)$.

Recap: Example - 2

Let X the sum of two independent dice rolls, and let Y denote the outcome of the first dice roll. Recall that the regression function of Y in terms of X, i.e., $E(Y|X)$ is $\frac{X}{2}$.
What is $E(E(Y|X))$?

Recap: Example - 2

Let X the sum of two independent dice rolls, and let Y denote the outcome of the first dice roll. Recall that the regression function of Y in terms of X, i.e., $E(Y|X)$ is $\frac{X}{2}$.
What is $E(E(Y|X))$?
$= E(\frac{X}{2})$, which by linearity is,
$= \frac{2E(Y)}{2} = E(Y)$.

Proof sketch of Law of Total Expectation

The previous example is not a coincidence. In general, $E(Y|X)$ is a function of X, so lets call it $g(X)$. Taking its expected value we get,

$E(g(X)) = \sum_x g(x)p_X(x)$. But we know $g(X) = E(Y|X)$. So,

$E(E(Y|X)) = \sum_x E(Y|X = x)p_X(x)$

$= \sum_x \left(\sum_y y p_{Y|X}(y|x)\right) p_X(x) = \sum_x \left(\sum_y y \frac{p_{X,Y}(x,y)}{p_X(x)}\right) p_X(x)$

Let X and Y be finite random variables, then its easy to see that we can rearrange the above terms to get,

$= \sum_x \sum_y y p_{X,Y}(x, y) = E(Y)$

Note: the above proof, with more care, can be modified to work for infinite RVs and continuous RVs.

Law of Total Expectation

Law of Total Expectation

For any two random variables X and Y, the following is always true:

$$E(Y) = E(E(Y|X))$$

It is important to note that in the RHS above, the outer expected value is taken in terms of the random variable X, while the inner one in terms of Y where Y here has the conditional distribution (either $p_{Y|X}$ or $f_{Y|X}$).

Proof of Property 3

Consider the problem of finding a function $g(X)$ that would best approximate the random variable Y. One heuristic would be if $g(X)$ would minimize $E((Y - g(X))^2)$. (Recall, in linear regression we wanted a linear function $\beta X + \alpha$ that would minimize the above least squares error).

Proof of Property 3 cont.

Using Law of Total expectation, we can write this as,

$$E(E((Y - g(X))^2 | X = x))$$

where the outer expectation is over the variable X and the inner over the random variable Y conditioned on $X = x$. We recall Steiner's inequality that, $E((X - c)^2)$ is minimized when $c = E(X)$. Then the above expected value is minimized when,

$$g(X) = E(Y|X = x)$$

So regression function, $E(Y|X = x)$ can be thought of as the best approximation of Y as a function of X (best in terms of least squared errors).

Example - 1

Suppose there are two machines in a factory such that the first is responsible for the production of 40% of the items, while the second for 60%. We further know that the average lifetime of items produced by the machines is 10 and 5 years respectively. What is the average lifetime of items produced at the factory? Note: it is incorrect to think it will be 7.5.

Example - 1

Suppose there are two machines in a factory such that the first is responsible for the production of 40% of the items, while the second for 60%. We further know that the average lifetime of items produced by the machines is 10 and 5 years respectively. What is the average lifetime of items produced at the factory?
Note: it is incorrect to think it will be 7.5.
It is 7 years.

Example - 2

Suppose the number of people arriving per day at a certain store has an average of 100, while the mean spending of the customers at the store is 10 dollars. What is the expected amount of money spent in the store in a day if spending of each customer is independent of the number of people?

Example - 2

Suppose the number of people arriving per day at a certain store has an average of 100, while the mean spending of the customers at the store is 10 dollars. What is the expected amount of money spent in the store in a day if spending of each customer is independent of the number of people?
Let X_i be the money spent by the i^{th} customer in the store. What we want to find is $E(\sum_i X_i)$. But the limit of the summation is itself a random variable! Let N be the number of customers on a given day. Then,
$E(\sum_i X_i) = E(E((\sum_{i=1}^{N} X_i)|N)) = E(10N) = 1000.$

Example

Suppose that soap bubbles blown by a kid burst after time T, where $T \sim Exp(R)$, where R is the radius of the soap bubble. So, larger bubbles burst sooner. If we know that $R = 1, 2$ or 3 with probability $\frac{1}{3}$ for each value, then what is $P(A)$ where A is the event that the next soap bubble bursts within a minute?

Example

Suppose that soap bubbles blown by a kid burst after time T, where $T \sim Exp(R)$, where R is the radius of the soap bubble. So, larger bubbles burst sooner. If we know that $R = 1, 2$ or 3 with probability $\frac{1}{3}$ for each value, then what is $P(A)$ where A is the event that the next soap bubble bursts within a minute?
we notice that for given R, $P(A) = 1 - e^{-R}$, then we can think that since each of the three radius are equally likely,
$P(A) = \frac{1}{3}(1 - e^{-1}) + \frac{1}{3}(1 - e^{-2}) + \frac{1}{3}(1 - e^{-3})$
What if instead R had density function $f_R(x) = 2x, 0 < x < 1$?

Law of Total Probability

Law of Total Probability

For any event A,

$$P(A) = \int_{-\infty}^{\infty} P(A|X = x) f_X(x) dx$$

Example cont.

Returning to A the event of a soap bubble bursting in 1 minute,
$P(A) = \int_0^1 2x(1 - e^{-x})dx$

The End