Probability Refresher

Chapter 5: Conditioning

Tabea Rebafka

September 2022

Master X–HEC 2022–23

Outline

Conditional distributions I

Recall that for events A, B such that $\mathbb{P}(B) > 0$, the conditional probability of A given B is defined by

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Conditional distributions II

Definition

- Let (X, Y) be a pair of **discrete** random variables. The **conditional distribution** of X given $Y = y$ is given by

$$p(x|y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}, \qquad \text{if } \mathbb{P}(Y = y) > 0,$$

  and $p(x|y) = 0$ if $\mathbb{P}(Y = y) = 0$.

- Let (X, Y) be a random vector with joint density $f_{(X,Y)}(x, y)$. For $y \in \mathbb{R}$, the **conditional density** of X given $Y = y$ is defined as the function

$$f_{X|Y=y}(x) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)}, \qquad \text{if } f_Y(y) \neq 0$$

  and $f_{X|Y=y}(x) = 0$ if $f_Y(y) = 0$, where f_Y denotes the marginal density of Y.

Conditional distributions III

Proposition

For any fixed y,

- the function $x \mapsto p(x|y)$ is the probability mass function of a discrete random variable **(discrete case)** .
- the function $x \mapsto f_{X|Y=y}(x)$ is a probability density function **(continuous case)**. $\int f_{X|Y=y}(x)\,dx = \int \frac{f_{(X,Y)}(x,y)}{f_Y(y)}\,dx = \frac{1}{f_Y(y)} \int f_{X,Y}(x,y)\,dx = 1$, with $\int f_{X,Y}(x,y)\,dx = f_Y(y)$

If X and Y are independent, then

- $p(x|y) = \mathbb{P}(X = x)$ for all x **(discrete case)** .
- $f_{X|Y=y}(x) = f_X(x)$ for all x **(continuous case)**.

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

Conditional distributions IV

$\log = \ln$

Exercise

- Pick X uniformly in $[0, 1]$, and then pick Y uniformly in $[0, X]$.
- What is the conditional density $f_{Y|X=x}$ of Y given $X = x$?
- Determine the marginal density f_Y of Y.

$X \sim \mathcal{U}[0,1]$

$(Y|\{X=x\} \sim \mathcal{U}[0,x] \quad \forall x)$

$Y|X \sim \mathcal{U}[0,X] \rightsquigarrow f_{Y|X=x}(y) = \frac{1}{x}\mathbb{1}_{[0,x]}(y) \quad \forall x \in [0,1]$

$$f_{Y|X=x}(y) = \frac{f_{(X,Y)}(x,y)}{f_X(x)} \Rightarrow f_{(X,Y)}(x,y) = f_X(x) f_{Y|X=x}(y) = \mathbb{1}_{[0,1]}(x)\frac{1}{x}\mathbb{1}_{[0,x]}(y)$$

$0 < x < 1 \qquad 0 < y < x$

$$f_Y(y) \overset{y \in [0,1]}{=} \int f_{(X,Y)}(x,y)\,dx = \int_y^1 \frac{1}{x}\,dx = \log x \Big|_y^1 = -\log y.$$

Conditional distributions V

Relation between the conditional densities $f_{X|Y}$ and $f_{Y|X}$:

Proposition (Bayes' formula)

We have

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

Conditional expectation I

What is the best prediction that can be made on X, given the value of another random variable Y?

Conditional expectation II

Definition

- Let (X, Y) be a pair of **discrete** random variables. The **conditional expectation** of X given Y is defined by

$$\mathbb{E}[X|Y] = \sum_x x p(x|Y).$$

- Let (X, Y) be a random vector with **continuous** distribution. The **conditional expectation** of X given Y is defined by

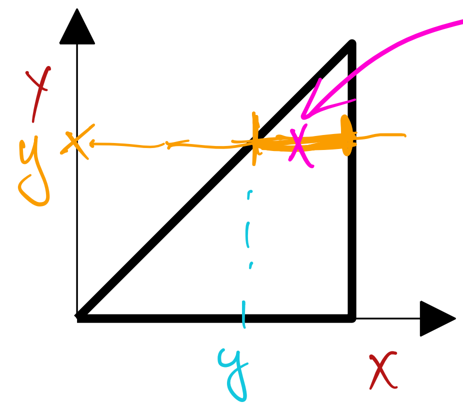
$$\mathbb{E}[X|Y] = \int_x x f_{X|Y}(x|Y) \mathrm{d}x.$$

Conditional expectation III

- In both cases, $\mathbb{E}[X|Y]$ is a function of Y, and therefore is a random variable.
- For any integrable function ϕ,

$$\mathbb{E}[\phi(X,Y)|Y] = \begin{cases} \sum_x \phi(x,Y)p(x|Y), & \text{in the discrete case} \\ \int_x \phi(x,Y)f_{X|Y}(x|Y)\mathrm{d}x, & \text{in the continuous case} \end{cases}$$

Conditional expectation IV



$$\mathbb{E}[X|Y=y] = \mathbb{E}[U(y,1)] = \frac{1+y}{2}$$

Exercise

- Let (X, Y) have uniform distribution on the unit triangle with density

$$f_{(X,Y)}(x,y) = 2\mathbb{1}_{0\leq y\leq x\leq 1}.$$

- Compute $\mathbb{E}[X|Y]$.

$$\mathbb{E}[X|Y] = \frac{1+Y}{2}$$

Conditional expectation V

Properties of conditional expectations

Let X, X', Y be random variables. In both the discrete and the continuous case, we have

1. **(Linearity)** $\mathbb{E}[aX + X'|Y] =$ $a\,\mathbb{E}[X|Y] + \mathbb{E}[X'|Y]$ for any constant a.
2. **(Averaging)**

   random variable of the form $g(Y)$

   $$\mathbb{E}[X] = \mathbb{E}\left[\mathbb{E}[X|Y]\right] = \begin{cases} \sum_y \mathbb{E}[X|Y=y]p_Y(y), & \text{discrete case} \\ \int_y \mathbb{E}[X|Y=y]f_Y(y)\mathrm{d}y, & \text{continuous case} \end{cases}$$

3. **('Taking out what is known')** For any function g,

   $$\mathbb{E}[g(Y)X|Y] = g(Y)\,\mathbb{E}[X|Y]$$

   In particular, $\mathbb{E}[g(Y)|Y] = g(Y)$.
4. **(Independence)** If X is independent of Y, then $\mathbb{E}[X|Y] =$ $\mathbb{E}[X]$

Conditional expectation VI

Exercise

Pick X uniformly in $[0,1]$, and then pick Y uniformly in $[0,X]$. Compute $\mathbb{E}[Y]$.

$$\mathbb{E}[Y] = \mathbb{E}\Big[\underbrace{\mathbb{E}[Y|X]}\Big] = \mathbb{E}\left[\frac{X}{2}\right] = \frac{1}{2}\mathbb{E}[X]$$

$\sim \mathcal{U}[0,X]$

$$= \frac{1}{4}.$$

Conditional expectation VII

- For some random variable $Y \in L^2$, denote the vector space

$$L^2(Y) = \left\{\text{random variables of the form } g(Y) \text{ such that } \mathbb{E}[g(Y)^2] < +\infty\right\},$$

equipped with the inner product $\langle X_1, X_2 \rangle = \mathbb{E}[X_1 X_2 | Y]$.

- Then the conditional expectation $\mathbb{E}[X|Y]$ coincides with the orthogonal projection of X onto $L^2(Y)$.

Theorem

Let $X, Y \in L^2$. The conditional expectation $\mathbb{E}[X|Y]$ is the best predictor of X among all possible predictors which are functions of Y, i.e.

$$\mathbb{E}\left[(X - \mathbb{E}[X|Y])^2\right] \leq \mathbb{E}\left[(X - g(Y))^2\right]$$

for every function g such that $g(Y) \in L^2$.

Conditional expectation VIII

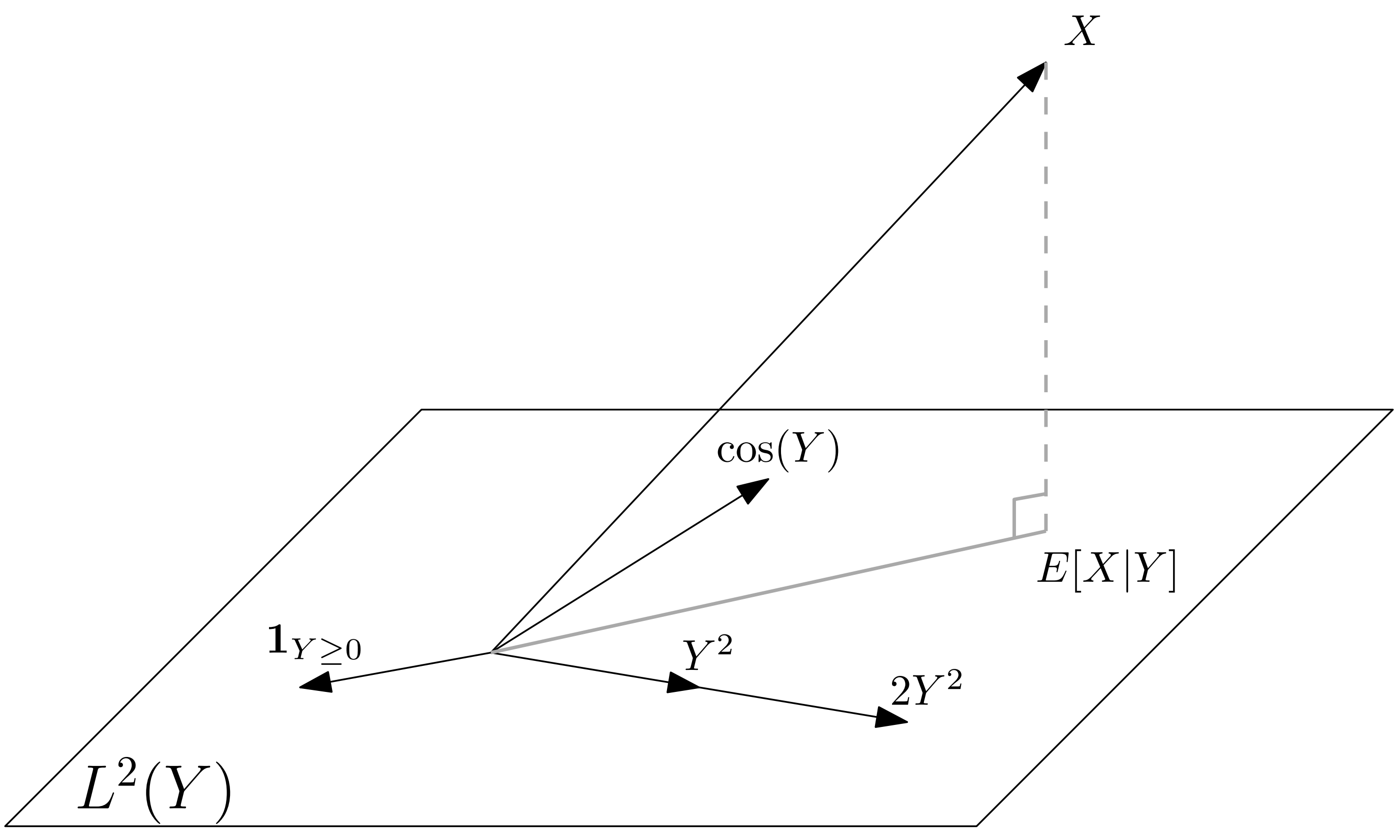


Illustration of the conditional expectation $\mathbb{E}[X|Y]$ as the orthogonal projection of X onto $L^2(Y)$.