Discrete case

Cond. pmf $\quad p_{X|Y}(x|y) = \dfrac{p(x,y)}{p_Y(y)}$

Cond. exp $\quad E[g(X)|Y=y] = \sum g(x)\, p_{X|Y}(x|y)$

Cont. case

Cond. density $\quad f_{X|Y}(x|y) = \dfrac{f(x,y)}{f_Y(y)}$

Cond. exp $\quad E[g(X)|Y=y]$

$$= \int_{\mathbb{R}} g(x)\, f_{X|Y}(x|y)\, dx$$

Simple example of continuous conditioning (1)

Density: Let (X, Y) be a random vector with density

$$\frac{e^{-\frac{x}{y}} e^{-y}}{y} \mathbf{1}_{(0,\infty)}(x) \, \mathbf{1}_{(0,\infty)}(y) \qquad \left(\Rightarrow \begin{array}{l} X \geq 0 \\ Y \geq 0 \end{array} \text{ a.s.}\right)$$

Question: Compute

$$\mathbf{P}(X > 1 | \, Y = y) = \mathbb{E}[\mathbb{1}_{(1,\infty)}(X) | Y = y]$$

$$= \int_1^\infty f_{X|Y}(x|y) \, dx$$

Marginal for Y : Take $y > 0$

$$f_Y(y) = \int_{\mathbb{R}} f(x,y)\,dx$$

Density of $\mathcal{E}(\frac{1}{y})$

$$= \int_{\mathbb{R}} \frac{e^{-y}}{y} e^{-\frac{x}{y}} \mathbb{1}_{(0,\infty)}(x)\,dx$$

$$= e^{-y} \int_0^{\infty} \frac{1}{y} e^{-\frac{x}{y}}\,dx$$

$$= e^{-y}$$

$(\Rightarrow Y \sim \mathcal{E}(1))$

Cond. density Take $x,y > 0$

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{\cancel{e^{-y}}\, e^{-\frac{x}{y}}}{\cancel{e^{-y}}\, y}$$

$$= \frac{1}{y} e^{-\frac{x}{y}}$$

$Z \sim \mathcal{E}(\lambda) \Rightarrow \mathbb{P}(Z \geq a) = e^{-\lambda a}$

Summary For $x, y > 0$

$$f_{X|Y}(x|y) = \frac{1}{y} e^{-\frac{x}{y}}$$

$$\Rightarrow \mathcal{L}(X \mid Y = y) \sim \mathcal{E}\left(\frac{1}{y}\right)$$

Cond. probability

$Z \sim \mathcal{E}\left(\frac{1}{y}\right)$

$$\mathbb{P}(X > 1 \mid Y = y) = \mathbb{P}(Z > 1)$$

$$= e^{-\frac{1}{y}}$$

Differences between baby / non baby cond. expectation

(i) We condition on a σ-algebra $\mathcal{F}$

$\mathbb{E}[X \mid \mathcal{F}]$

(ii) $\mathbb{E}[X \mid \mathcal{F}]$ is a r.v.

(iii) We define $\mathbb{E}[g(X) \mid \mathcal{F}]$, then conditional distributions

(iv) Interpretation: "Best possible" approx.

Simple example of continuous conditioning (2)

Marginal distribution of Y: We have

$$\begin{aligned} f_Y(y) &= \int_0^\infty f(x,y)\,dx \\ &= \frac{e^{-y}}{y}\left(\int_0^\infty e^{-\frac{x}{y}}\,dx\right)\mathbf{1}_{(0,\infty)}(y) \\ &= e^{-y}\,\mathbf{1}_{(0,\infty)}(y) \end{aligned}$$

Conditional density: For $y > 0$ we have

$$f_{X|Y}(x|\,y) = \frac{f(x,y)}{f_Y(y)} = \frac{e^{-\frac{x}{y}}}{y}\,\mathbf{1}_{(0,\infty)}(x)$$

Namely $\mathcal{L}(X|\,Y=y) = \mathcal{E}(\frac{1}{y})$

Simple example of continuous conditioning (3)

Conditional probability:

$$\begin{aligned}\mathbf{P}(X > 1 \,|\, Y = y) &= \int_1^\infty f_{X|Y}(x|\,y)\,dx \\ &= \int_1^\infty \frac{e^{-\frac{x}{y}}}{y}\,dx \\ &= e^{-\frac{1}{y}}\end{aligned}$$

Cond. expectation in the continuous case

Definition 5.

Let

- (X, Y) couple of continuous random variables
- Joint density f
- Marginal densities f_X, f_Y, y such that $f_Y(y) > 0$
- $f_{X|Y}(x|\,y)$ conditional density

Then the conditional exp. of X given $Y = y$ is defined by

$$\mathbf{E}\left[X|\,Y = y\right] = \int_{\mathbb{R}} x\, f_{X|Y}(x|\,y)\, dx$$

Example of continuous conditional expectation (1)

Density: Let (X, Y) be a random vector with density

$$\frac{e^{-\frac{x}{y}} e^{-y}}{y} \mathbf{1}_{(0,\infty)}(x) \mathbf{1}_{(0,\infty)}(y)$$

Question: Compute

$$\mathbf{E}\left[X \middle| Y = y\right]$$

Example of continuous conditional expectation (2)

Conditional density: For $y > 0$ we have seen that

$$f_{X|Y}(x|\,y) = \frac{f(x,y)}{f_Y(y)} = \frac{e^{-\frac{x}{y}}}{y}\,\mathbf{1}_{(0,\infty)}(x)$$

Namely $\mathcal{L}(X|\,Y = y) = \mathcal{E}(\frac{1}{y})$

Conditional expectation: We have

$$\begin{aligned} \mathbf{E}\left[X|\,Y = y\right] &= \int_{\mathbb{R}} x\, f_{X|Y}(x|\,y)\,dx \\ &= \int_0^\infty x\,\frac{e^{-\frac{x}{y}}}{y} \\ &= y \end{aligned}$$

Outline

Formal definition

Big σ-algebra

Definition 6.

We are given a probability space $(\Omega, \mathcal{F}_0, \mathbf{P})$ and

- A σ-algebra $\mathcal{F} \subset \mathcal{F}_0$.
- $X \in \mathcal{F}_0$ such that $\mathbf{E}[|X|] < \infty$.

Conditional expectation of X given $\mathcal{F}$:

$\mathbb{E}[X|\mathcal{F}] = Y$

- Denoted by $\mathbf{E}[X|\mathcal{F}]$
- Defined by: $\mathbf{E}[X|\mathcal{F}]$ is the $L^1(\Omega)$ r.v Y such that
  (i) $Y \in \mathcal{F}$.
  (ii) For all $A \in \mathcal{F}$, we have

$$\mathbf{E}[X\mathbf{1}_A] = \mathbf{E}[Y\mathbf{1}_A],$$

or otherwise stated $\int_A X \, d\mathbf{P} = \int_A Y \, d\mathbf{P}$.

Remarks

Notation: We use the notation $Y \in \mathcal{F}$ to say that a random variable Y is $\mathcal{F}$-measurable.

Interpretation: More intuitively

- $\mathcal{F}$ represents a given information
- Y is the best prediction of X given the information in $\mathcal{F}$.

Existence and uniqueness:
To be seen after the examples.

Outline

Setting $(\Omega, \mathcal{F}_0, \mathbb{P})$, $\mathcal{F} \subset \mathcal{F}_0$

Example 1 If $X \in \mathcal{F}$

Recall: Y is the "best approx" of X as a r.v. $Y \in \mathcal{F}$

Intuitively: If $X \in \mathcal{F}$, the best approx should be X

Hyp: $\boxed{X \in \mathcal{F}.}$ Set $Y = X$. Then

(i) $Y \in \mathcal{F}$

(ii) If $A \in \mathcal{F}$

$$\mathbb{E}[\underbrace{Y \, \mathbb{1}_A] = \mathbb{E}[X}_{Y=X} \, \mathbb{1}_A]$$

Thus $\boxed{X = \mathbb{E}[X \mid \mathcal{F}]}$

Easy examples (1)

Example 1: Assume

$$X \in \mathcal{F}.$$

Then

$$\mathbf{E}[X|\mathcal{F}] = X$$

Independence of a r.v and a σ-field

Definition 7.

We say that $X \perp\!\!\!\perp \mathcal{F}$ if
$\hookrightarrow$ for all $A \in \mathcal{F}$ and $B \in \mathcal{B}(\mathbb{R})$, we have

$$\mathbf{P}((X \in B) \cap A) = \mathbf{P}(X \in B)\, \mathbf{P}(A),$$

or otherwise stated:

$$X \perp\!\!\!\perp \mathbf{1}_A$$

Hyp: Consider $\boxed{X \perp\!\!\!\perp \mathcal{F}}$. Then

$$\boxed{\mathbb{E}[X \mid \mathcal{F}] = \mathbb{E}[X]}$$

Proof: Set $Y = \mathbb{E}[X]$. Then

(i) $Y \in \mathcal{F}$ (Y is constant)

(ii) If $A \in \mathcal{F}$. Then

$$\mathbb{E}[X \mathbb{1}_A] \overset{\perp\!\!\!\perp}{=} \overbrace{\mathbb{E}[X]}^{=Y} \mathbb{E}[\mathbb{1}_A]$$

$$= \mathbb{E}[Y \mathbb{1}_A]$$

Easy examples (2)

Example 2: Assume

$$X \perp\!\!\!\perp \mathcal{F}.$$

Then

$$\mathbf{E}[X|\mathcal{F}] = \mathbf{E}[X]$$

Proof: example 2

We have

(i) $\mathbf{E}[X] \in \mathcal{F}$ since $\mathbf{E}[X]$ is a constant.

(ii) If $A \in \mathcal{F}$,

$$\mathbf{E}[X\,\mathbf{1}_A] = \mathbf{E}[X]\;\mathbf{E}[\mathbf{1}_A] = \mathbf{E}\Big[\mathbf{E}[X]\,\mathbf{1}_A\Big].$$

Discrete conditional expectation

Example 3: We consider

- $\left\{\Omega_j; j \geqslant 1\right\}$ partition of Ω such that $\mathbf{P}(\Omega_j) > 0$ for all $j \geqslant 1$.
- $\mathcal{F} = \sigma(\Omega_j; j \geqslant 1)$.

Then

$$\mathbf{E}[X|\mathcal{F}] = \sum_{j \geqslant 1} \frac{\mathbf{E}[X \, \mathbf{1}_{\Omega_j}]}{\mathbf{P}(\Omega_j)} \, \mathbf{1}_{\Omega_j} \equiv Y. \tag{1}$$

Setting: Partition $\{\Omega_j \,;\, j \geq 1\}$

$Y = \sum_{j \geq 1} \beta_j \mathbb{1}_{\Omega_j}$, $\beta_j = \dfrac{\mathbb{E}[X \mathbb{1}_{\Omega_j}]}{\mathbb{P}(\Omega_j)}$

(i) $\mathcal{F} = \sigma(\Omega_j \,;\, j \geq 1)$

$\Rightarrow \mathbb{1}_{\Omega_j} \in \mathcal{F}$ (as a r.v)

$\Rightarrow \sum_{j \geq 1} \alpha_j \mathbb{1}_{\Omega_j} \in \mathcal{F}$ $\forall (\alpha_j)_{j \geq 1}$

In particular, $\boxed{Y \in \mathcal{F}}$

$Y = \sum_{j \geq 1} \beta_j \mathbb{1}_{\Omega_j}$, $\beta_j = \frac{E[X \mathbb{1}_{\Omega_j}]}{P(\Omega_j)}$ $\Rightarrow A = \bigcup_{j \in J} \Omega_j$

(ii) $E[Y \mathbb{1}_A] \overset{(*)}{=} E[X \mathbb{1}_A] \quad \forall A \in \mathcal{F}$

Note: Since $(\Omega_j)_{j \geq 1}$ is a partition, for every $\boxed{A \in \mathcal{F}}$ we have

$$\mathbb{1}_A = \sum_{j \in J} \mathbb{1}_{\Omega_j} \qquad J \subset \{1, 2, \dots$$

$\Rightarrow$ $\boxed{\text{It is sufficient to prove } (*) \text{ for } A = \Omega_j \,,\ j = 1, \dots}$

Then $E[Y \mathbb{1}_A] = \sum_{j \in J} E[Y \mathbb{1}_{\Omega_j}]$

$$= \sum_{j \in J} E[X \mathbb{1}_{\Omega_j}] = E[X \mathbb{1}_A]$$

$Y = \sum_{j \geq 1} \beta_j \mathbb{1}_{\Omega_j}$, $\beta_j = \frac{\mathbb{E}[X \mathbb{1}_{\Omega_j}]}{\mathbb{P}(\Omega_j)}$

For a given n,

$=0$ a.s. if $j \neq n$

$$\boxed{\mathbb{E}[Y \mathbb{1}_{\Omega_n}]} = \mathbb{E}[(\sum_{j \geq 1} \beta_j \mathbb{1}_{\Omega_j}) \mathbb{1}_{\Omega_n}]$$

$$= \mathbb{E}[\beta_n \mathbb{1}_{\Omega_n}]$$

$$= \beta_n \mathbb{E}[\mathbb{1}_{\Omega_n}]$$

$$= \beta_n \mathbb{P}(\Omega_n) = \frac{\mathbb{E}[X \mathbb{1}_{\Omega_n}]}{\mathbb{P}(\Omega_n)} \mathbb{P}(\Omega_n)$$

$$\boxed{= \mathbb{E}[X \mathbb{1}_{\Omega_n}]}$$

$\Rightarrow$ (ii) proved

Simple example of partition : B, B^c

$$\Rightarrow \sigma(B, B^c) = \{\emptyset, B, B^c, \Omega\} = \mathcal{F}$$

Now take $A \in \mathcal{F}_0$ and set

$$\mathbb{E}[\mathbb{1}_A \mid \mathcal{F}] \equiv \mathbb{P}(A \mid \mathcal{F})$$

We get, according to example 3

$$\boxed{\mathbb{P}(A \mid \mathcal{F})} = \frac{\mathbb{E}[\mathbb{1}_A \mathbb{1}_B]}{\mathbb{P}(B)} \mathbb{1}_B + \frac{\mathbb{E}[\mathbb{1}_A \mathbb{1}_{B^c}]}{\mathbb{P}(B^c)} \mathbb{1}_{B^c}$$

$$\boxed{= \mathbb{P}(A \mid B) \, \mathbb{1}_B + \mathbb{P}(A \mid B^c) \, \mathbb{1}_{B^c}}$$

Proof: example 3

Strategy: Verify (i) and (ii) for the random variable Y.

(i) For all $j \geq 1$, we have $\mathbf{1}_{\Omega_j} \in \mathcal{F}$. Thus, for any sequence $(\alpha_j)_{j\geq 1}$,

$$\sum_{j\geq 1} \alpha_i \mathbf{1}_{\Omega_j} \in \mathcal{F}.$$

(ii) It is enough to verify (1) for $\mathrm{A} = \Omega_n$ and $n \geq 1$ fixed. However,

$$\mathbf{E}[Y\,\mathbf{1}_{\Omega_n}] = \mathbf{E}\left\{\frac{\mathbf{E}[X\mathbf{1}_{\Omega_n}]}{\mathbf{P}(\Omega_n)}\,\mathbf{1}_{\Omega n}\right\} = \frac{\mathbf{E}[X\,\mathbf{1}_{\Omega_n}]}{\mathbf{P}(\Omega_n)}\mathbf{E}[\mathbf{1}_{\Omega_n}] = \mathbf{E}[X\,\mathbf{1}_{\Omega_n}].$$

Undergrad conditional probability

Definition: For a generic measurable set $A \in \mathcal{F}_0$, we set

$$\mathbf{P}(A|\mathcal{F}) \equiv \mathbf{E}[\mathbf{1}_A|\mathcal{F}]$$

Discrete example setting:

Let B, B^c be a partition of Ω, and $A \in \mathcal{F}_0$. Then

1. $\mathcal{F} = \sigma(B) = \left\{\Omega, \emptyset, B, B^c\right\}$
2. We have
$$\mathbf{P}(A|\mathcal{F}) = \mathbf{P}(A|B)\,\mathbf{1}_B + \mathbf{P}(A|B^c)\,\mathbf{1}_{B^c}.$$