

Conditional expectation

Samy Tindel

Purdue University

Probability Theory 2 - MA 539

Taken from *Probability: Theory and examples*
by R. Durrett

Outline

- 1 Definition
- 2 Examples
- 3 Conditional expectation: properties
- 4 Conditional expectation as a projection
- 5 Conditional regular laws

Outline

- 1 Definition
- 2 Examples
- 3 Conditional expectation: properties
- 4 Conditional expectation as a projection
- 5 Conditional regular laws

Formal definition

Definition 1.

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} :

- 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: to be seen after the examples.

Uniqueness: If it exists, the conditional expectation is unique.

Proof of uniqueness

Aim: Let Y, Y' satisfying (i) + (ii).

\hookrightarrow Let us show $Y = Y'$ a.s

General property: For all $A \in \mathcal{F}$, we have $\mathbf{E}[Y \mathbf{1}_A] = \mathbf{E}[Y' \mathbf{1}_A]$.

Particular case: Let $\epsilon > 0$, and set

$$A_\epsilon \equiv (Y - Y' \geq \epsilon).$$

Then $A_\epsilon \in \mathcal{F}$, and thus

$$0 = \mathbf{E}[(Y - Y') \mathbf{1}_{A_\epsilon}] \geq \epsilon \mathbf{E}[\mathbf{1}_{A_\epsilon}] = \epsilon \mathbf{P}(A_\epsilon)$$

$$\Rightarrow \mathbf{P}(A_\epsilon) = 0.$$

Proof of uniqueness (2)

Set A_+ : Let

$$A_+ \equiv (Y - Y' > 0) = \bigcup_{n \geq 1} A_{1/n}.$$

We have $n \mapsto A_{1/n}$ increasing, and thus

$$\mathbf{P}(A_+) = \mathbf{P}\left(\bigcup_{n \geq 1} A_{1/n}\right) = \lim_{n \rightarrow \infty} \mathbf{P}(A_{1/n}) = 0.$$

Set A_- : In the same way, if

$$A_- = \{Y - Y' < 0\}$$

we have $P(A_-) = 0$.

Proof of uniqueness (3)

Conclusion: We obtain, setting

$$A_{\neq} \equiv \{Y \neq Y'\} = A_+ \cup A_-,$$

that $\mathbf{P}(A_{\neq}) = 0$, and thus $Y = Y'$ a.s.

Absolute continuity

Definition 2.

Let μ, ν two σ -finite measures on (Ω, \mathcal{F}) .

We say that $\nu \ll \mu$ (μ is absolutely continuous w.r.t ν) if

$$\mu(A) = 0 \implies \nu(A) = 0 \text{ for all } A \in \mathcal{F}.$$

Radon-Nykodym theorem

Theorem 3.

Let

- μ, ν σ -finite measures on (Ω, \mathcal{F}) , such that $\nu \ll \mu$.

Then there exists $f \in \mathcal{F}$ such that for all $A \in \mathcal{F}$ we have

$$\nu(A) = \int_A f d\mu.$$

The function f :

- Is called Radon-Nykodym derivative of μ with respect to ν
- Is denoted by $f \equiv \frac{d\nu}{d\mu}$.
- We have $f \geq 0$ μ -almost everywhere
- $f \in L^1(\mu)$.

Conditional expectation: existence

Hypothesis: We have

- A σ -algebra $\mathcal{F} \subset \mathcal{F}_0$.
- $X \in \mathcal{F}_0$ such that $\mathbf{E}[|X|] < \infty$.
- $X \geq 0$.

Defining two measures: we set

- 1 $\mu = P$, measure on (Ω, \mathcal{F}) .
- 2 $\nu(A) \equiv \mathbf{E}[X \mathbf{1}_A] = \int_A X d\mathbf{P}$.

Then ν is a measure (owing to Beppo-Levi).

Conditional expectation: existence (2)

Absolute continuity: we have

$$\begin{aligned}\mathbf{P}(A) = 0 &\Rightarrow \mathbf{1}_A = 0 \quad P\text{-a.s.} \\ &\Rightarrow X \mathbf{1}_A = 0 \quad P\text{-a.s.} \\ &\Rightarrow \nu(A) = 0\end{aligned}$$

Thus $\nu \ll P$

Conclusion: invoking Radon-Nykodym, there exists $f \in \mathcal{F}$ such that, for all $A \in \mathcal{F}$, we have $\nu(A) = \int_A f d\mathbf{P}$.

\hookrightarrow We set $f = \mathbf{E}[X|\mathcal{F}]$.

Outline

- 1 Definition
- 2 Examples**
- 3 Conditional expectation: properties
- 4 Conditional expectation as a projection
- 5 Conditional regular laws

Easy examples

Example 1: If $X \in \mathcal{F}$, then $\mathbf{E}[X|\mathcal{F}] = X$.

Definition: We say that $X \perp\!\!\!\perp \mathcal{F}$ if
 \Leftrightarrow 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: $X \perp\!\!\!\perp \mathbf{1}_A$.

Example 2: If $X \perp\!\!\!\perp \mathcal{F}$, 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 $A \in \mathcal{F}$,

$$\mathbf{E}[X \mathbf{1}_A] = \mathbf{E}[X] \mathbf{E}[\mathbf{1}_A] = \mathbf{E}\left[\mathbf{E}[X] \mathbf{1}_A\right].$$

Discrete conditional expectation

Example 3: We consider

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

Then

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

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 $(\alpha_j)_{j \geq 1}$,

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

(ii) It is enough to verify (1) for $A = \Omega_n$ 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 $\mathcal{F} = \sigma(B) = \{\Omega, \emptyset, B, B^c\}$
- 2 We have

$$\mathbf{P}(A|\mathcal{F}) = \mathbf{P}(A|B) \mathbf{1}_B + \mathbf{P}(A|B^c) \mathbf{1}_{B^c}.$$

Dice throwing

Example: We consider

- $\Omega = \{1, 2, 3, 4, 5, 6\}$, $A = \{4\}$, $B = \text{"even number"}$.

Then

$$\mathbf{P}(A|\mathcal{F}) = \frac{1}{3} \mathbf{1}_B.$$

Conditioning a r.v by another r.v

Definition: Let X and Y be two random variables such that $X \in L^1(\Omega)$. We set

$$\mathbf{E}[X|Y] = \mathbf{E}[X|\sigma(Y)].$$

Criterion to determine if $A \in \sigma(Y)$:

We have $A \in \sigma(Y)$ iff

$$A = \{\omega; Y(\omega) \in B\}, \quad \text{or} \quad \mathbf{1}_A = \mathbf{1}_B(Y)$$

Criterion to determine if $Z \in \sigma(Y)$:

Let Z and Y be two random variables. Then

$$Z \in \sigma(Y) \quad \text{iff we can write} \quad Z = U(Y), \quad \text{with} \quad U \in \mathcal{B}(\mathbb{R}).$$

Conditioning a r.v by a discrete r.v

Example 4: Whenever X and Y are discrete random variables
 \hookrightarrow Computation of $\mathbf{E}[X|Y]$ can be handled as in example 3.

More specifically:

- Assume $Y \in E$ with $E = \{y_i; i \geq 1\}$
- Hypothesis: $\mathbf{P}(Y = y_i) > 0$ for all $i \geq 1$.

Then $\mathbf{E}[X|Y] = h(Y)$ with $h : E \rightarrow \mathbb{R}$ defined by:

$$h(y) = \frac{\mathbf{E}[X \mathbf{1}_{(Y=y)}]}{\mathbf{P}(Y = y)}.$$

Conditioning a r.v by a continuous r.v

Example 5: Let (X, Y) couple of real random variables with measurable density $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$. We assume that

$$\int_{\mathbb{R}} f(x, y) dx > 0, \quad \text{for all } y \in \mathbb{R}.$$

Let $g : \mathbb{R} \rightarrow \mathbb{R}$ a measurable function such that $g(X) \in L^1(\Omega)$. Then $\mathbf{E}[g(X)|Y] = h(Y)$, with $h : \mathbb{R} \rightarrow \mathbb{R}$ defined by:

$$h(y) = \frac{\int_{\mathbb{R}} g(x) f(x, y) dx}{\int_{\mathbb{R}} f(x, y) dx}.$$

Heuristic proof

Formally one can write:

$$\mathbf{P}(X = x|Y = y) = \frac{\mathbf{P}(X = x, Y = y)}{\mathbf{P}(Y = y)} = \frac{f(x, y)}{\int f(x, y)dx},$$

Integrating against this density we get:

$$\begin{aligned}\mathbf{E}[g(X)|Y = y] &= \int g(x)\mathbf{P}(X = x|Y = y) dx \\ &= \frac{\int g(x)f(x, y)dx}{\int f(x, y)dx}.\end{aligned}$$

Rigorous proof

Strategy: Check (i) and (ii) in the definition for the r.v $h(Y)$.

(i) If $h \in \mathbb{B}(\mathbb{R})$, we have seen that $h(Y) \in \sigma(Y)$.

(ii) Let $A \in \sigma(Y)$ Then

$$A = \{\omega; Y(\omega) \in B\} \implies \mathbf{1}_A = \mathbf{1}_B(Y)$$

Thus

$$\begin{aligned} \mathbf{E}[h(Y)\mathbf{1}_A] &= \mathbf{E}[h(Y)\mathbf{1}_B(Y)] \\ &= \int_B \int_{\mathbb{R}} h(y)f(x, y)dx dy \\ &= \int_B dy \int_{\mathbb{R}} \left\{ \frac{\int g(z)f(z, y)dz}{\int f(z, y)dz} \right\} f(x, y)dx \\ &= \int_B dy \int g(z)f(z, y)dz = \mathbf{E}[g(X)\mathbf{1}_B(Y)]. \end{aligned}$$

Weird example

Example 6: We take

- $\Omega = (0, 1)$, $\mathcal{F}_0 = \mathcal{B}((0, 1))$ and $P = \lambda$.

We set $X(\omega) = \cos(\pi\omega)$, and

$$\mathcal{F} = \{A \subset (0, 1); A \text{ or } A^c \text{ countable}\}.$$

Then $\mathbf{E}[X|\mathcal{F}] = 0$.

Proof

Strategy: Check (i) and (ii) in the definition for the r.v $Y = 0$.

(i) Obviously $0 \in \mathcal{F}$.

(ii) Let $A \in \mathcal{F}$, such that A is countable. Then

$$\mathbf{E}[X \mathbf{1}_A] = \int_A \cos(\pi x) dx = 0.$$

Similarly, if $A \in \mathcal{F}$ is such that A^c is countable, we have

$$\mathbf{E}[X \mathbf{1}_A] = \int_0^1 \cos(\pi x) dx - \int_{A^c} \cos(\pi x) dx = 0,$$

which ends the proof.

Weird example: heuristics

Intuition: One could think that

- 1 We know that $\{x\}$ occurred for all $x \in [0, 1]$
- 2 $\{x\} \in \mathcal{F}$
- 3 Thus $\mathbf{E}[X|\mathcal{F}] = X$.

Paradox: This is wrong because $X \notin \mathcal{F}$.

Correct intuition: If we know $\omega \in A_i$ for a finite number of $A_i \in \mathcal{F}$ then nothing is known about X .

Outline

- 1 Definition
- 2 Examples
- 3 Conditional expectation: properties**
- 4 Conditional expectation as a projection
- 5 Conditional regular laws

Linearity, expectation

Proposition 4.

Let $X \in L^1(\Omega)$. Then

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

Proposition 5.

Let $\alpha \in \mathbb{R}$, and $X, Y \in L^1(\Omega)$. Then

$$\mathbf{E}[\alpha X + Y|\mathcal{F}] = \alpha \mathbf{E}[X|\mathcal{F}] + \mathbf{E}[Y|\mathcal{F}] \quad \text{a.s.}$$

Proof

Strategy: Check (i) and (ii) in the definition for the r.v

$$Z \equiv \alpha \mathbf{E}[X|\mathcal{F}] + \mathbf{E}[Y|\mathcal{F}].$$

Verification: we have

(i) Z is a linear combination of $\mathbf{E}[X|\mathcal{F}]$ and $\mathbf{E}[Y|\mathcal{F}]$

$$\Leftrightarrow Z \in \mathcal{F}.$$

(ii) For all $A \in \mathcal{F}$, we have

$$\begin{aligned} \mathbf{E}[Z \mathbf{1}_A] &= E\left\{(\alpha \mathbf{E}[X|\mathcal{F}] + \mathbf{E}[Y|\mathcal{F}]) \mathbf{1}_A\right\} \\ &= \alpha E\left\{\mathbf{E}[X|\mathcal{F}] \mathbf{1}_A\right\} + E\left\{\mathbf{E}[Y|\mathcal{F}] \mathbf{1}_A\right\} \\ &= \alpha \mathbf{E}[X \mathbf{1}_A] + \mathbf{E}[Y \mathbf{1}_A] \\ &= \mathbf{E}[(\alpha X + Y) \mathbf{1}_A]. \end{aligned}$$

Monotonicity

Proposition 6.

Let $X, Y \in L^1(\Omega)$ such that $X \leq Y$ almost surely. We have

$$\mathbf{E}[X|\mathcal{F}] \leq \mathbf{E}[Y|\mathcal{F}],$$

almost surely.

Proof: Along the same lines as proof of uniqueness for the conditional expectation. For instance if we set

$$A_\varepsilon = \{\mathbf{E}[X|\mathcal{F}] - \mathbf{E}[Y|\mathcal{F}] \geq \varepsilon > 0\},$$

then it is readily checked that

$$\mathbf{P}(A_\varepsilon) = 0.$$

Monotone convergence

Proposition 7.

Let $\{X_n; n \geq 1\}$ be a sequence of random variables such that

- $X_n \geq 0$
- $X_n \nearrow X$ almost surely
- $\mathbf{E}[X] < \infty$.

Then

$$\mathbf{E}[X_n | \mathcal{F}] \nearrow \mathbf{E}[X | \mathcal{F}].$$

Proof

Strategy: Set $Y_n \equiv X - X_n$. We are reduced to show $Z_n \equiv \mathbf{E}[Y_n|\mathcal{F}] \searrow 0$.

Existence of a limit: $n \mapsto Y_n$ is decreasing, and $Y_n \geq 0$
 $\Leftrightarrow Z_n$ is decreasing and $Z_n \geq 0$.
 $\Leftrightarrow Z_n$ admits a limit a.s, denoted by Z_∞ .

Aim: Show that $Z_\infty = 0$.

Proof (2)

Expectation of Z_∞ : we will show that $\mathbf{E}[Z_\infty] = 0$. Indeed

- X_n converges a.s. to X .
- $0 \leq X_n \leq X \in L^1(\Omega)$.

Thus, by dominated convergence, $\mathbf{E}[X_n] \rightarrow \mathbf{E}[X]$.

We deduce:

- $\mathbf{E}[Y_n] \rightarrow 0$
- Since $\mathbf{E}[Y_n] = \mathbf{E}[Z_n]$, we also have $\mathbf{E}[Z_n] \rightarrow 0$.
- By monotone convergence, we have $\mathbf{E}[Z_n] \rightarrow \mathbf{E}[Z_\infty]$

This yields $\mathbf{E}[Z_\infty] = 0$.

Conclusion: $Z_\infty \geq 0$ and $\mathbf{E}[Z_\infty] = 0$

$\Leftrightarrow Z_\infty = 0$ almost surely.

Cauchy-Schwarz inequality

Proposition 8.

Let $X, Y \in L^2(\Omega)$. Then

$$\mathbf{E}^2[X Y | \mathcal{F}] \leq \mathbf{E}[X^2 | \mathcal{F}] \mathbf{E}[Y^2 | \mathcal{F}] \quad \text{a.s.}$$

Proof:

For all $\theta \in \mathbb{R}$, we have

$$\mathbf{E}[(X + \theta Y)^2 | \mathcal{F}] \geq 0 \quad \text{a.s.}$$

Thus almost surely we have: for all $\theta \in \mathbb{Q}$,

$$\mathbf{E}[(X + \theta Y)^2 | \mathcal{F}] \geq 0,$$

Proof

Expansion: For all $\theta \in \mathbb{Q}$

$$\mathbf{E}[Y^2|\mathcal{F}]\theta^2 + 2\mathbf{E}[XY|\mathcal{F}]\theta + \mathbf{E}[X^2|\mathcal{F}] \geq 0.$$

Recall: If a polynomial satisfies $a\theta^2 + b\theta + c \geq 0$ for all $\theta \in \mathbb{Q}$
 \hookrightarrow then we have $b^2 - 4ac \leq 0$

Application: Almost surely, we have

$$E^2[XY|\mathcal{F}] - \mathbf{E}[X^2|\mathcal{F}]\mathbf{E}[Y^2|\mathcal{F}] \leq 0.$$

Jensen's inequality

Proposition 9.

Let $X \in L^1(\Omega)$, and $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ such that $\varphi(X) \in L^1(\Omega)$ and φ **convex**. Then

$$\varphi(\mathbf{E}[X|\mathcal{F}]) \leq \mathbf{E}[\varphi(X)|\mathcal{F}] \quad \text{a.s.}$$

Proposition 10.

The conditional expectation is a **contraction** in $L^p(\Omega)$ for all $p \geq 1$

Proof: According to Jensen's inequality,

$$X \in L^p(\Omega) \Rightarrow \mathbf{E}[X|\mathcal{F}] \in L^p(\Omega)$$

and

$$\mathbf{E}\{|\mathbf{E}[X|\mathcal{F}]|^p\} \leq \mathbf{E}[|X|^p]$$

Successive conditionings

Theorem 11.

Let

- Two σ -algebras $\mathcal{F}_1 \subset \mathcal{F}_2$.
- $X \in L^1(\Omega)$.

Then

$$\mathbf{E} \{ \mathbf{E}[X | \mathcal{F}_1] | \mathcal{F}_2 \} = \mathbf{E}[X | \mathcal{F}_1] \quad (2)$$

$$\mathbf{E} \{ \mathbf{E}[X | \mathcal{F}_2] | \mathcal{F}_1 \} = \mathbf{E}[X | \mathcal{F}_1]. \quad (3)$$

Proof

Proof of (2): We set $Z \equiv \mathbf{E}[X|\mathcal{F}_1]$. Then

$$Z \in \mathcal{F}_1 \subset \mathcal{F}_2.$$

According to Example 1, we have $\mathbf{E}[Z|\mathcal{F}_2] = Z$, i.e. (2).

Proof of (3): We set $U = \mathbf{E}[X|\mathcal{F}_2]$.

\hookrightarrow We will show that $\mathbf{E}[U|\mathcal{F}_1] = Z$, via (i) and (ii) of Definition 1.

(i) $Z \in \mathcal{F}_1$.

(ii) If $A \in \mathcal{F}_1$, we have $A \in \mathcal{F}_1 \subset \mathcal{F}_2$, and thus

$$\mathbf{E}[Z\mathbf{1}_A] = \mathbf{E}[X\mathbf{1}_A] = \mathbf{E}[U\mathbf{1}_A].$$

Conditional expectation for products

Theorem 12.

Let $X, Y \in L^2(\Omega)$, such that $X \in \mathcal{F}$. Then

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

Proof: We use a 4 steps methodology

Proof

Step 1: Assume $X = \mathbf{1}_B$, with $B \in \mathcal{F}$

We check (i) and (ii) of Definition 1.

(i) We have $\mathbf{1}_B \mathbf{E}[Y|\mathcal{F}] \in \mathcal{F}$.

(ii) For $A \in \mathcal{F}$, we have

$$\begin{aligned} \mathbf{E}\{(\mathbf{1}_B \mathbf{E}[Y|\mathcal{F}]) \mathbf{1}_A\} &= \mathbf{E}\{\mathbf{E}[Y|\mathcal{F}] \mathbf{1}_{A \cap B}\} \\ &= \mathbf{E}[Y \mathbf{1}_{A \cap B}] \\ &= \mathbf{E}[(\mathbf{1}_B Y) \mathbf{1}_A], \end{aligned}$$

and thus

$$\mathbf{1}_B \mathbf{E}[Y|\mathcal{F}] = \mathbf{E}[\mathbf{1}_B Y|\mathcal{F}].$$

Proof (2)

Step 2: If X is of the form

$$X = \sum_{i \leq n} \alpha_i \mathbf{1}_{B_i},$$

with $\alpha_i \in \mathbb{R}$ and $B_i \in \mathcal{F}$, then, by linearity we also get

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

Step 3: If $X, Y \geq 0$

\hookrightarrow There exists a sequence $\{X_n; n \geq 1\}$ of simple random variables such that

$$X_n \nearrow X.$$

Then applying the monotone convergence we end up with:

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

Proof (3)

Step 4: General case $X \in L^2$

\hookrightarrow Decompose $X = X^+ - X^-$ and $Y = Y^+ - Y^-$, which gives

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

by linearity.

Conditional expectation and independence

Theorem 13.

Let

- X, Y two independent random variables
- $\alpha : \mathbb{R}^2 \rightarrow \mathbb{R}$ such that $\alpha(X, Y) \in L^1(\Omega)$

We set, for $x \in \mathbb{R}$,

$$g(x) = \mathbf{E}[\alpha(x, Y)].$$

Then

$$\mathbf{E}[\alpha(X, Y)|X] = g(X).$$

Proof: with 4 steps method applied to α .

Outline

- 1 Definition
- 2 Examples
- 3 Conditional expectation: properties
- 4 Conditional expectation as a projection**
- 5 Conditional regular laws

Orthogonal projection

Definition: Let

- H Hilbert space
↔ complete vectorial space equipped with inner product.
- F closed subspace of H .

Then, for all $x \in H$

- There exists a unique $y \in F$, denoted by $y = \pi_F(x)$

Satisfying one of the equivalent conditions (i) or (ii).

- (i) For all $z \in F$, we have $\langle x - y, z \rangle = 0$.
- (ii) For all $z \in F$, we have $\|x - y\|_H \leq \|x - z\|_H$.

$\pi_F(x)$ is denoted **orthogonal projection** of x onto F .

Conditional expectation and projection

Theorem 14.

Consider

- The space $L^2(\mathcal{F}_0) \equiv \{Y \in \mathcal{F}_0; \mathbf{E}[Y^2] < \infty\}$.
- $X \in L^2(\mathcal{F}_0)$.
- $\mathcal{F} \subset \mathcal{F}_0$

Then

- 1 $L^2(\mathcal{F}_0)$ is a Hilbert space
 \hookrightarrow Inner product $\langle X, Y \rangle = \mathbf{E}[XY]$.
- 2 $L^2(\mathcal{F})$ is a closed subspace of $L^2(\mathcal{F}_0)$.
- 3 $\pi_{L^2(\mathcal{F})}(X) = \mathbf{E}[X|\mathcal{F}]$.

Proof

Proof of 2:

If $X_n \rightarrow X$ in $L^2 \Rightarrow$ There exists a subsequence $X_{n_k} \rightarrow X$ a.s.
Thus, if $X_n \in \mathcal{F}$, we also have $X \in \mathcal{F}$.

Proof of 3: Let us check (i) in our definition of projection

Let $Z \in L^2(\mathcal{F})$.

\Leftrightarrow We have $\mathbf{E}[Z X | \mathcal{F}] = Z \mathbf{E}[X | \mathcal{F}]$, and thus

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

which ensures (i) and $\mathbf{E}[X | \mathcal{F}] = \pi_{L^2(\mathcal{F})}(X)$.

Application to Gaussian vectors

Example: Let

- (X, Y) centered Gaussian vector in \mathbb{R}^2
- Hypothesis: $V(Y) > 0$.

Then

$$\mathbf{E}[X|Y] = \alpha Y, \quad \text{with} \quad \alpha = \frac{\mathbf{E}[X Y]}{V(Y)}.$$

Proof

Step 1: We look for α such that

$$Z = X - \alpha Y \implies Z \perp\!\!\!\perp Y.$$

Recall: If (Z, Y) is a Gaussian vector
 $\iff Z \perp\!\!\!\perp Y$ iff $\text{cov}(Z, Y) = 0$

Application: $\text{cov}(Z, Y) = \mathbf{E}[Z Y]$. Thus

$$\text{cov}(Z, Y) = \mathbf{E}[(X - \alpha Y) Y] = \mathbf{E}[X Y] - \alpha V(Y),$$

et

$$\text{cov}(Z, Y) = 0 \quad \text{iff} \quad \alpha = \frac{\mathbf{E}[XY]}{V(Y)}.$$

Proof (2)

Step 2: We invoke (i) in the definition of π .

\hookrightarrow Let $V \in L^2(\sigma(Y))$. Then

$$Y \perp\!\!\!\perp (X - \alpha Y) \quad \Longrightarrow \quad V \perp\!\!\!\perp (X - \alpha Y)$$

and

$$\mathbf{E}[(X - \alpha Y) V] = \mathbf{E}[X - \alpha Y] \mathbf{E}[V] = 0.$$

Thus

$$\alpha Y = \pi_{\sigma(Y)}(X) = \mathbf{E}[X | Y].$$

Outline

- 1 Definition
- 2 Examples
- 3 Conditional expectation: properties
- 4 Conditional expectation as a projection
- 5 Conditional regular laws**

Definition 15.

Let

- (Ω, \mathcal{F}, P) a probability space
- (S, \mathcal{S}) a measurable space
- $X : (\Omega, \mathcal{F}) \rightarrow (S, \mathcal{S})$ a random variable
- \mathcal{G} a σ -algebra such that $\mathcal{G} \subset \mathcal{F}$.

We say that $\mu : \Omega \times \mathcal{S} \rightarrow [0, 1]$ is a **Conditional regular law** of X given \mathcal{G} if

- For all A , the map $\omega \rightarrow \mu(\omega, A)$ is a random variable, equal to $\mathbf{P}(X \in A | \mathcal{G})$ a.s.
- ω -a.s. $A \rightarrow \mu(\omega, A)$ is a probability measure on (S, \mathcal{S}) .

Remark:

We will always have (S, \mathcal{S}) of the form $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$, etc.

Discrete example

Poisson law case: Let

- $X \sim \mathcal{P}(\lambda)$ and $Y \sim \mathcal{P}(\mu)$
- $X \perp\!\!\!\perp Y$

We set $S = X + Y$.

Then CRL of X given S is $\text{Bin}(S, p)$, with $p = \frac{\lambda}{\lambda + \mu}$

Proof: we have seen that for $n \leq m$

$$\mathbf{P}(X = n | S = m) = \binom{m}{n} p^n (1 - p)^{m-n} \quad \text{with } p = \frac{\lambda}{\lambda + \mu}.$$

Then we consider

- $S = \mathbb{N}$, $\mathcal{G} = \sigma(S)$

and we verify that these conditional probabilities define a CRL.

Continuous example

Exponential law case: Let

- $X \sim \mathcal{E}(1)$ and $Y \sim \mathcal{E}(1)$
- $X \perp\!\!\!\perp Y$

We set $S = X + Y$.

Then CRL of X given S is $\mathcal{U}([0, S])$.

Proof: The joint density of (X, S) is given par

$$f(x, s) = e^{-s} \mathbf{1}_{\{0 \leq x \leq s\}}.$$

Let then $\psi \in \mathcal{B}_b(\mathbb{R}_+)$. Thanks to Example 5, we have

$$\mathbf{E}[\psi(X)|S] = u(S),$$

with

$$u(s) = \frac{\int_{\mathbb{R}_+^2} \psi(x) f(x, s) dx}{\int_{\mathbb{R}_+^2} f(x, s) dx} = \frac{1}{s} \int_0^s \psi(x) dx.$$

Proof

In addition, $S \neq 0$ almost surely, and thus if $A \in \mathcal{B}(\mathbb{R})$ we have:

$$\mathbf{P}(X \in A | S) = \frac{|A \cap [0, S]|}{S}.$$

Considering the state space as $= \mathbb{R}_+$, $\mathcal{S} = \mathcal{B}(\mathbb{R}_+)$ and setting

$$\mu(\omega, A) = \frac{|A \cap [0, S(\omega)]|}{S(\omega)},$$

one can verify that we have defined a **conditional regular law**.

Existence of the CRL

Theorem 16.

Let

- X a random variable on $(\Omega, \mathcal{F}_0, P)$.
- Taking values in a space of the form $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.
- $\mathcal{G} \subset \mathcal{F}_0$ a σ -algebra.

Then the CRL of X given \mathcal{G} exists.

Proof: nontrivial and omitted.

Computation rules for CRL

(1) If $\mathcal{G} = \sigma(Y)$, with Y random variable with values in \mathbb{R}^m , we have

$$\mu(\omega, A) = \mu(Y(\omega), A),$$

and one can define a CRL of X given Y as a family $\{\mu(y, \cdot); y \in \mathbb{R}^m\}$ of probabilities on \mathbb{R}^n , such that for all $A \in \mathcal{B}(\mathbb{R}^n)$ the function

$$y \mapsto \mu(y, A)$$

is measurable.

(2) If Y is a discrete r.v, this can be reduced to:

$$\mu(y, A) = \mathbf{P}(X \in A | Y = y) = \frac{\mathbf{P}(X \in A, Y = y)}{\mathbf{P}(Y = y)}.$$

Computation rules for CRL (2)

- (3) When one knows the CRL, quantities like the following (for $\phi \in \mathcal{B}(\mathbb{R}^n)$) can be computed:

$$\mathbf{E}[\phi(X)|\mathcal{G}] = \int_{\mathbb{R}^n} \phi(x) \mu(\omega, dx)$$

$$\mathbf{E}[\phi(X)|Y] = \int_{\mathbb{R}^n} \phi(x) \mu(Y, dx).$$

- (4) The CRL is not unique.
However if N_1, N_2 are 2 CRL of X given \mathcal{G}
 \hookrightarrow we have ω -almost surely:

$$N_1(\omega, A) = N_2(\omega, A) \quad \text{for all } A \in \mathcal{B}(\mathbb{R}^n).$$