CHAPTER 10:

CONDITIONAL EXPECTATION II:

ABSTRACT DEFINITION AND EXAMPLES.

Goals:
- formal definition of conditional expectation
- particular cases helping to develop an intuition.

Setup. $(\Omega, \mathcal{F}, P)$ probability space. $\cdot\ L^p = \{ X : \Omega \to \mathbb{R} \quad E(|X|^p) < \infty \}$.

$\mathcal{G} \subset \mathcal{F}$ sub σ-algebra.

$\mathcal{G}$ represents some information about an element $\omega \in \Omega$.

- $\mathcal{G} = \sigma(\{\Omega_1, \dots, \Omega_n\})$ where $\Omega = \Omega_1 \sqcup \dots \sqcup \Omega_n$ partition.

  $\hookrightarrow$ information = the element Ω_i where ω is.

- $\mathcal{G} = \sigma(Y)$ where Y r.v.

  $\hookrightarrow$ information = value of $Y(\omega)$ is given.

- $\mathcal{G} = \sigma(Y_1, \dots, Y_k)$ where Y_i are r.v.s.

  $\hookrightarrow$ information = values of $Y_i(\omega)$ for all i.

- $\mathcal{G} = \{ B \in \mathcal{G} \}$ general σ-algebra.

  $\hookrightarrow$ information = we know $1_B(\omega)$ for every $B \in \mathcal{G}$.

If X r.v., $X \in L^1$. $\left(= \{ Z : \Omega \to \mathbb{R} \text{ s.t. } \int |Z| dP < \infty \} \right)$

- "Best guess" about X without any information: $E(X)$.
- Best guess of X knowing some information $\mathcal{G}$?

1 THEORETICAL BACKGROUND.

Thm: Let $X \subset L^1$. Let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra.

There exists a random variable $X' \in L^1$ s.t.

(i) X' is $\mathcal{G}$-measurable.

(ii) $\forall A \in \mathcal{G}$ $E(X 1_A) = E(X' 1_A)$

Furthermore, if X'' is another r.v satisfying (i) and (ii), then $X' = X''$ a.s.

Rk: (i) + (ii) $\Rightarrow$ $X' \in L^1$

(choose $A = \{X' \geq 0\}$ and $A = \{X' \leq 0\}$ in (ii))

Proof. [Existence].

Case 1 $X \geq 0$.

Let Q be the measure on $(\Omega, \mathcal{G})$ defined by

$$\forall A \in \mathcal{G} \quad Q(A) = E(X 1_A)$$

Since $X \in L^1$ $Q(\Omega) = E(X) < \infty$, Q is a finite meas.

Let $\tilde{P} = P|_{\mathcal{G}}$. For every $A \in \mathcal{G}$

$$\tilde{P}(A) = 0 \Rightarrow P(A) = 0 \Rightarrow E(X 1_A) = 0 \Rightarrow Q(A) = 0.$$

By Radon-Nikodym Theorem (applied to $(E, \mathcal{E}) = (\Omega, \mathcal{G})$, $\mu = \tilde{P}$, $\nu = Q$), there exists $X' \geq 0$, $\mathcal{G}$-measurable s.t. $\forall A \in \mathcal{G}$ $Q(A) = \int_A X' \, d\tilde{P}$ $\left(= \int_A X' \, dP \right)$.

i.e. $\forall A \in \mathcal{G}$ $E(X 1_A) = E(X' 1_A)$.

Taking $A = \Omega$ yields $E(X') = E(X) < \infty$.

<u>Case 2</u> : $X \in L^1$ general.

$X = X_+ - X_-$ where $X_+ \geq 0$ and $X_- \geq 0$

Take $X' = X_+' - X_-'$ where X_+', X_-' are $\mathcal{G}$-meas. and satisfy $\forall A \in \mathcal{G}$

$$E(X_\pm' 1_A) = E(X_\pm 1_A)$$

[Uniqueness].

Let X', X'' satisfying (i), (ii).

Let $A = \{X' > X''\} \overset{(i)}{\in} \mathcal{G}$. By (ii), we have

because $\{X' > X''\} = \bigcup_{a \in \mathbb{Q}} \{X' > a\} \cap \{a > X''\}$

$$E(X' 1_A) = E(X'' 1_A),$$

ie

$$E\Big(\underbrace{(X' - X'') 1_{(X' - X'') > 0}}_{\geq 0}\Big) = 0.$$

This implies $(X' - X'') 1_{X' - X'' > 0} = 0$ a.s.

ie $X' - X'' \leq 0$ a.s.

Equivalently, considering $A = \{X' < X''\}$, we show that $X' - X'' \geq 0$ a.s. $\blacksquare$

2 DEFINITION.

The theorem above justifies the following definition.

<u>Def.</u> Let $X \in L^1$ be a real r.v., $\mathcal{G} \subset \mathcal{F}$ σ-algebra.

<u>The conditional expectation of X given $\mathcal{G}$</u> is the random variable $E(X|\mathcal{G}) \in L^1$ satisfying

(P1) $E(X|\mathcal{G})$ is $\mathcal{G}$ measurable.

(P2) $\forall A \in \mathcal{G}$ $E(X 1_A) = E(E(X|\mathcal{G}) 1_A)$.

<u>Rk:</u> • $E(X|\mathcal{G})$ is well defined in $L^1/\sim$.

$\uparrow$ a.s. equality

• In the definition above, since we do not take quotient, the r.v. $E(X|\mathcal{G})$ is only well defined up to a a.s. set: Any random variable X' satisfying (P1) and (P2) above is called a version of conditional expectation. When we write $E(X|\mathcal{G})$ in a statement, this statement should be true for any version of $E(X|\mathcal{G})$.

• (P1, P2) is called the characteristic property of $E(X|\mathcal{G})$.

Notation: • For $A \in \mathcal{F}$, we write

$$P(A \mid \mathcal{G}) = E(1_A \mid \mathcal{G}).$$

"conditional probability of A"

• For Y a r.v. with values in a measured space $(E, \mathcal{E})$, we write

$$E(X \mid Y) := E(X \mid \sigma(Y)).$$

Proposition: (Equivalent formulations of P2).

Let $X \in L^1$, $\mathcal{G} \subset \mathcal{F}$ σ-algebra. Let $X' \in L^1$, $\mathcal{G}$ measurable.

Let $\mathcal{C} \subset \mathcal{F}$ be a π-system such that $\sigma(\mathcal{C}) = \mathcal{G}$.

The following are equivalent

[P2-0] $\forall A \in \mathcal{C}$ $E(X 1_A) = E(X' 1_A)$

[P2-1] $\forall A \in \mathcal{G}$ $E(X 1_A) = E(X' 1_A)$

[P2-2] $\forall Z$ $\mathcal{G}$-meas. bounded $E(XZ) = E(X'Z)$

Proof: 0 ⇒ 1 Dynkin Lemma.

1 ⇒ 2 Without loss of generality, we can assume $Z \geq 0$ (otherwise consider $Z + c$, where c large cst).

Let $Z_n \uparrow Z$ where Z_n simple.

By linearity (P2-1) implies that

$$E(X Z_n) = E(X' Z_n),$$

and the result follows from dominated convergence.

2 ⇒ 1) choose $Z = 1_A$.

3 σ- ALGEBRA GENERATED BY A PARTITION

Assume $\Omega = \bigsqcup_{i \in I} \Omega_i$ (disjoint union) where I finite or coutable.

Let $\mathcal{G} = \sigma(\{\Omega_i, i \in I\})$

$$= \left\{ \{ \bigcup_{i \in J} \Omega_i \}, \quad J \subset I \right\}.$$

<u>Prop</u>. Let $X \in L^1$. We have

$$E(X \mid \mathcal{G}) = \sum_{i \in I} E(X \mid \Omega_i) 1_{\Omega_i} \quad \text{a.s.}$$

(For almost all $\omega \in \Omega_i$ $E(X \mid \mathcal{G})(\omega) = E(X \mid \Omega_i)$.
(By convention, we set $E(X \mid \Omega_i) 1_{\Omega_i} = 0$ ..))

<u>Proof</u>. Since $E(X \mid \mathcal{G})$ $\mathcal{G}$- measurable; we have

$$E(X \mid \mathcal{G}) = \sum_{i \in I} \lambda_i 1_{\Omega_i} \quad \text{a.s.} \qquad (\text{excercise}).$$
$(\lambda_i \in \mathbb{R})$

To identify λ_i, we use (P_2). For every $i \in I$

$$\lambda_i P(\Omega_i) = E\left(E(X \mid \mathcal{G}) 1_{\Omega_i} \right)$$
$$\overset{P2}{=} E\left(X 1_{\Omega_i} \right)$$

■

<u>Example:</u> $\Omega = [0,1]$ $\mathcal{F} = \mathcal{B}([0,1])$ $P = \mathrm{Leb}_{[0,1]}$.

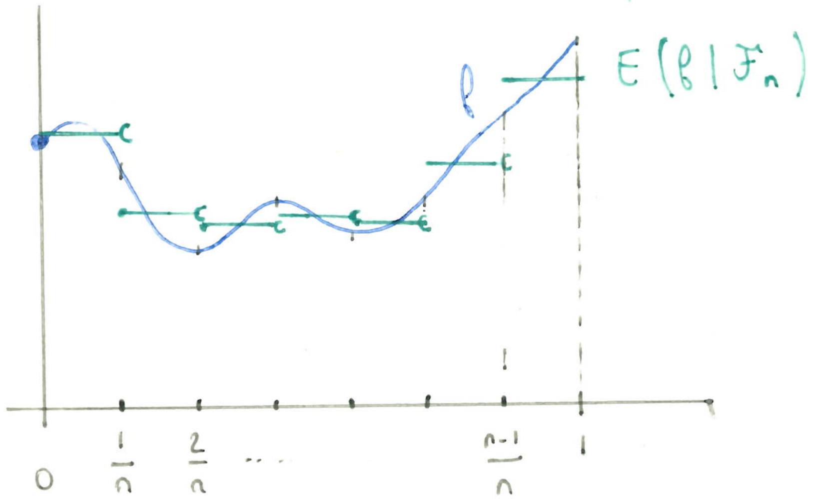
$f : [0,1] \longrightarrow \mathbb{R}$ continuous.

$$\Omega = \underbrace{[0, \tfrac{1}{n}[}_{I_1} \cup \underbrace{[\tfrac{1}{n}, \tfrac{2}{n}[}_{I_2} \cup \dots \cup \underbrace{[\tfrac{n-2}{n}, \tfrac{n-1}{n}[}_{I_{n-1}} \cup \underbrace{[\tfrac{n-1}{n}, 1]}_{I_n}$$

$$\mathcal{F}_n = \sigma(\{I_1, \dots, I_n\})$$

$$E(f \mid \mathcal{F}_n) = \sum_{i=1}^{n} \alpha_i \, 1\!\!1_{I_i}$$

where $\alpha_i = E(f \mid I_i) = n \int_{\frac{i-1}{n}}^{\frac{i}{n}} f(x)\, dx$.



"On each interval I_i we replace f by its expectation on I_i"

4 DISCRETE R.V.S

Prop. Let $X \in L^1$, Y r.v. in a measured space $(E, \mathcal{E})$.
Assume that there exists $D \subset E$ finite or countable s.t. $Y \in D$ a.s. Then

$$E(X \mid Y) = \sum_{y \in D} E(X \mid Y=y) \mathbb{1}_{Y=y},$$

(with the convention $E(X \mid Y=y) \mathbb{1}_{Y=y} = 0$ if $P(Y=y)=0$)

Rk: We recover the formula in Chapter 9.

Proof: By applying the proposition of Section 3 to the partition

$$\Omega = \{Y \notin D\} \cup \bigcup_{y \in D} \{Y=y\},$$

we get

$$E(X \mid Y) = \sum_{y \in D} E(X \mid Y=y) \mathbb{1}_{Y=y}. \quad \blacksquare$$

5 CONDITIONAL DENSITIES

<u>Def.</u> (density). Let $f : \mathbb{R}^d \to [0, \infty]$. A random variable X in $\mathbb{R}^d$ has density f if

$$\boxed{\forall A \subset \mathbb{R}^d \text{ meas. } P(X \in A) = \int_A f(x)\,dx.}$$

<u>Rk:</u> • X has a density iff $\mu_X \ll Leb_{\mathbb{R}^d}$, and in this case its density is $f = \frac{d\,Leb}{d\mu_X}$.

(Radon_Nikodym derivative).

• If X has density f, we have for every $\Psi : \mathbb{R}^d \to \mathbb{R}$ meas. ≥ 0 or s.t. $E(|\Psi(X)|) < \infty$.

$$E(\Psi(X)) = \int_{\mathbb{R}^d} \Psi(x) f(x)\,dx.$$

• If a vector $(X, Y) \in \mathbb{R}^k \times \mathbb{R}^\ell$ has density f, then X has density f_X, given by

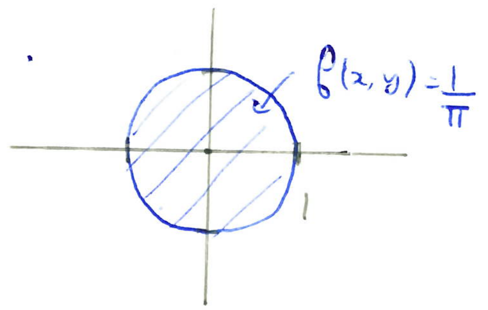
$$\forall x \in \mathbb{R}^k \quad f_X(x) = \int_{\mathbb{R}^\ell} f(x, y)\,dy$$

and Y has density f_Y, given by

$$\forall y \in \mathbb{R}^\ell \quad f_Y(y) = \int_{\mathbb{R}^k} f(x, y)\,dx.$$

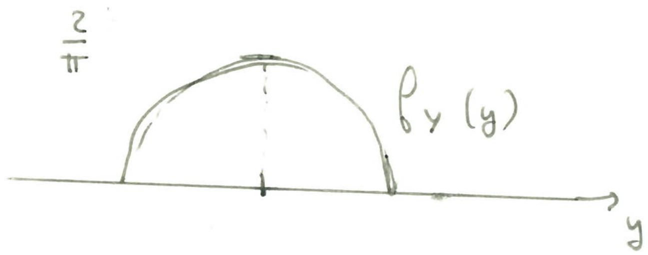
Example. Uniform point in a circle.

Let $f(x,y) = \frac{1}{\pi} 1_{x^2+y^2 \leq 1}$



Let $(X,Y) \in \mathbb{R}^2$ with density f. Density of Y:

$$f_Y(y) = \frac{1}{\pi} \int_{-1}^{1} 1_{x^2+y^2 \leq 1} \, dx = \frac{2}{\pi} \sqrt{1-y^2}$$



Prop. Let $k, \ell \geq 1$, X random variable in $\mathbb{R}^k$, Y random variable in $\mathbb{R}^\ell$. Assume that (X,Y) has a density $f : \mathbb{R}^k \times \mathbb{R}^\ell \to [0,\infty]$.
Let $\Psi : \mathbb{R}^k \longrightarrow \mathbb{R}$ s.t. $\Psi(X) \in L^1$. Then

$$E(\Psi(X) \mid Y) = \varphi(Y), \text{ where}$$

$$\varphi(y) = \int_{\mathbb{R}^k} \Psi(x) \cdot \frac{f(x,y)}{f_Y(y)} \, dy$$

Convention: $\varphi(y) = 0$ if $f_Y(y) = 0$.

Rk: • This is the continuous analog of: if X takes value in D and Y in D' (at most countable).

$$E(\Psi(X) \mid Y) = \varphi(Y) \quad \text{a.s.}$$

where $\varphi(y) = \sum_x \Psi(x) \underbrace{P(X = x \mid Y = y)}_{= \frac{P(X = x, Y = y)}{P(Y = y)}}$.

• $\dfrac{f(x, y)}{f_Y(y)}$ is the density of X "conditionally on $Y = y$", sometimes denoted $f_{X|Y}(x \mid y)$.

Proof. $\varphi(Y)$ is Y-meas. and for every bounded meas. $g : \mathbb{R}^\ell \to \mathbb{R}$, we have

$$E(\varphi(Y) g(Y)) = \int_{\mathbb{R}^\ell} \varphi(y) g(y) f_Y(y) \, dy$$

$$\overset{\text{Fub.}}{=} \int_{\mathbb{R}^{k+\ell}} \Psi(x) \frac{f(x,y)}{f_Y(y)} g(y) f_Y(y) \, dx \, dy$$

$$= E(\Psi(X) g(Y))$$

$\blacksquare$

GEOMETRIC INTERPRETATION OF $E(X)$ IN L^2

$L^2 = \{ X : \Omega \to \mathbb{R} \text{ measurable } E(X^2) < \infty \}$

$H = L^2 / \sim \qquad (X \sim Y \text{ if } X = Y \text{ a.s.}).$

If $X \in L^2$, write $\tilde{X} \in H$ for its equivalence class.

Given $\tilde{X}, \tilde{Y} \in H$, we can define

$$\langle \tilde{X}, \tilde{Y} \rangle = E(XY).$$

$(H, \langle \cdot, \cdot \rangle)$ Hilbert space (admitted).

Geometric interpretation of $E(X)$ and $Var(X)$.

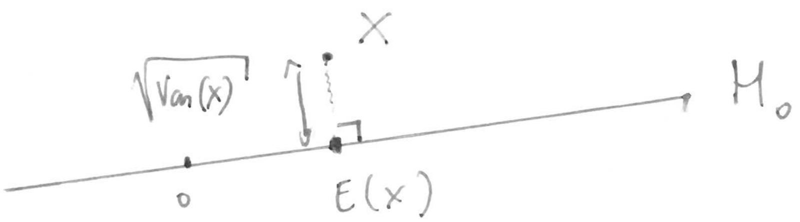
Let $H_0 = \{ \tilde{\lambda}, \lambda \in \mathbb{R} \}$ constant r.v.s.

Let $X \in L^2$. For every $\lambda \in \mathbb{R}$, we have

$$E\big((X - E(X)) \cdot \lambda\big) = 0.$$

Hence $\forall \tilde{\lambda} \in H_0$

$$\langle \tilde{\lambda}, \tilde{X} - \widetilde{E(X)} \rangle = 0$$



$E(X)$ is the orthogonal projection of X on the closed subspace H_0. In particular

$$E\left((X-\lambda)^2\right) \text{ is minimized for } \lambda = E(X).$$

GEOMETRIC INTERPRETATION OF $E(X|\mathcal{G})$.

Let $\mathcal{G} \subset \mathcal{F}$ σ-algebra. Write

$$H_{\mathcal{G}} = \{X : \Omega \to \mathbb{R} \ \ \mathcal{G}\text{-meas. } E(X^2) < \infty\} / \sim$$

One can check that $H_{\mathcal{G}}$ is a closed subspace of H.

<u>Prop.</u> Let $X \in L^2$. Among all functions $Z \in L^2$, $\mathcal{G}$-meas,

$$E\left((X-Z)^2\right) \text{ is minimized for } Z = E(X|\mathcal{G}).$$

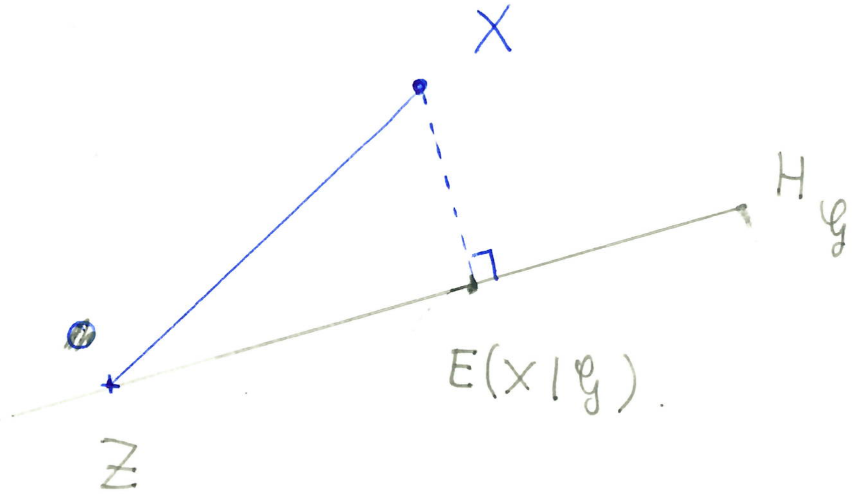
<u>Proof</u>. Let $\tilde{X}' \in H_{\mathcal{G}}$ be the orthogonal projection of $\tilde{X}$ on $H_{\mathcal{G}}$. We have

$$\begin{cases} \tilde{X}' \in H_{\mathcal{G}}, & \text{and} \\ \forall \tilde{Z} \in H_{\mathcal{G}} & \langle \tilde{X} - \tilde{X}', \tilde{Z} \rangle = 0 \end{cases}$$

This implies $\begin{cases} X' & \text{is } \mathcal{G}\text{-measurable} \\ \forall Z & \mathcal{G}\text{-measurable bounded} \quad E(XZ) = E(X'Z). \end{cases}$

Hence $X' = E(X|\mathcal{G})$ a.s. ∎

Rk : This gives a geometrical interpretation of $E(X | \mathcal{G})$ when $X \in L^2$: it is the orthogonal projection of X on $H_{\mathcal{G}}$.



(P2) is the orthogonality condition $(X - E(X | \mathcal{G})) \perp H_{\mathcal{G}}$.

7 NON NEGATIVE R.V.

Def. Let X be r.v. with values in $[0,\infty]$ a.s. Let $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra. The conditional expectation $E(X|\mathcal{G})$ is the r.v. in $[0,\infty]$ satisfying

(P1) $E(X|\mathcal{G})$ is $\mathcal{G}$-measurable

(P2) For every $A \in \mathcal{G}$ $\quad E(X 1_A) = E(E(X|\mathcal{G}) 1_A)$.

Rk: • $E(X|\mathcal{G})$ is well def. up to a.s. equality:

[Existence] Let $X_n = X 1_{X \leq n}$ for $n \geq 1$, and define

$$X_n' = E(X_n | \mathcal{G}). \text{ We have}$$

$$0 \leq X_n' \leq X_{n+1}' \quad \text{a.s.}$$

(if $U \leq V$ $\quad U, V \in L^1$ then $E(U|\mathcal{G}) \leq E(V|\mathcal{G})$ a.s.:

consider $A = \{V > U\}$ and use $E(U 1_A) \leq E(V 1_A)$.)

Set $X' = \lim_{n\to\infty} X_n'$. X' satisfies P1 as a a.s. lim of $\mathcal{G}$-measurable functions. Furthermore, by monotone cv, for every $A \in \mathcal{G}$

$$E(X' 1_A) = \lim_{n\to\infty} E(X_n' 1_A) = \lim_{n\to\infty} E(X_n 1_A) = E(X 1_A)$$

[Uniqueness] Let X', X'' satisfying P1 and P2

Let $Q < a < b$ and $A = \{X' \leq a < b \leq X''\}$

$$b.P(A) \leq E(1_A X'') = E(1_A X') \leq a P(A).$$

i.e. $(b-a) P(A) \leq 0$, which gives $P(A) = 0$

By taking the union over $a, b \in \mathbb{Q}$, we get

$$P(X' < X'') = 0.$$

Rk: P2 is equivalent to

$$\forall Z \geq 0 \ \mathcal{G}\text{-measurable } E(ZX) = E(Z E(X|\mathcal{G}))$$

(exercise)