Lecture 2 : Conditional Expectation II

MATH275B - Winter 2012 *Lecturer: Sebastien Roch*

References: [Wil91, Chapter 9], [Dur10, Section 5.1].

1 Conditional expectation: definition, existence, uniqueness

1.1 Definition

DEF&THM 2.1 *Let $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ a sub σ-field. Then there exists a (a.s.) unique $Y \in \mathcal{L}^1(\Omega, \mathcal{G}, \mathbb{P})$ s.t.*

$$\mathbb{E}[Y; G] = \mathbb{E}[X; G], \ \forall G \in \mathcal{G}.$$

Such Y is called a version of $\mathbb{E}[X \,|\, \mathcal{G}]$.

1.2 Proof of uniqueness

Let Y, Y' be two versions of $\mathbb{E}[X \,|\, G]$ such that w.l.o.g. $\mathbb{P}[Y > Y'] > 0$. By monotonicity, there is $n \geq 1$ with $G = \{Y > Y' + n^{-1}\} \in \mathcal{G}$ such that $\mathbb{P}[G] > 0$. Then, by definition,

$$0 = \mathbb{E}[Y - Y'; G] > n^{-1}\mathbb{P}[G] > 0,$$

which gives a contradiction.

1.3 Proof of existence

There are two main approaches:

1. First approach: Radon-Nikodym theorem. Read [Dur10, Section A.4].
2. Second approach: Hilbert space method.

We begin with a definition.

DEF&THM 2.2 *Let $X \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ a sub σ-field. Then there exists a (a.s.) unique $Y \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$ s.t.*

$$\Delta \equiv \|X - Y\|_2 = \inf\{\|X - W\|_2 \,:\, W \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})\},$$

and, moreover,

$$\langle Z, X - Y \rangle = 0,\ \forall Z \in \mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P}).$$

Such Y is called an orthogonal projection of X on $\mathcal{L}^2(\Omega, \mathcal{G}, \mathbb{P})$.

We give a proof for completeness.
Proof: Take (Y_n) s.t. $\|X - Y_n\|_2 \to \Delta$. Remembering that $L^2(\Omega, \mathcal{G}, \mathbb{P})$ is complete we seek to prove that (Y_n) is Cauchy. Using the parallelogram law

$$2\|U\|_2^2 + 2\|V\|_2^2 = \|U - V\|_2^2 + \|U + V\|_2^2,$$

note that

$$\|X - Y_r\|_2^2 + \|X - Y_s\|_2^2 = 2\|X - \frac{1}{2}(Y_r + Y_s)\|_2^2 + 2\|\frac{1}{2}(Y_r - Y_s)\|_2^2.$$

The first term on the LHS is at least Δ^2 so we have what we need. Let Y be the limit of (Y_n).

Note that for any $Z \in L^2(\Omega, \mathcal{G}, \mathbb{P})$ and $t \in \mathbb{R}$

$$\|X - Y - tZ\|_2^2 \geq \|X - Y\|_2^2,$$

so that, expanding and rearranging, we have

$$-2t\langle Z, X - Y \rangle + t^2\|Z\|_2^2 \geq 0,$$

which is only possible if the first term is 0.

Uniqueness follows from the parallelogram law again. ∎

We return to the proof of existence of the conditional expectation. We use the standard machinery. The previous theorem implies that conditional expectations exist for indicators and simple functions. Now take $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and write $X = X^+ - X^-$, so we can assume $X \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})^+$ w.l.o.g. Using the staircase function

$$X^{(r)} = \begin{cases} 0, & \text{if X = 0} \\ (i-1)2^{-r}, & \text{if } (i-1)2^{-r} < X \leq i2^{-r} \leq r \\ r, & \text{if } X > r, \end{cases}$$

we have $0 \leq X^{(r)} \uparrow X$. Let $Y^{(r)} = \mathbb{E}[X^{(r)} \,|\, \mathcal{G}]$. Using an argument similar to the proof of uniqueness, it follows that $U \geq 0$ implies $\mathbb{E}[U \,|\, \mathcal{G}] \geq 0$. Using linearity , we then have $Y^{(r)} \uparrow Y \equiv \limsup Y^{(r)}$ which is measurable in $\mathcal{G}$. By (MON)

$$\mathbb{E}[Y; G] = \mathbb{E}[X; G],\ \forall G \in \mathcal{G}.$$

2 Examples

EX 2.3 *If* $X \in \mathcal{L}^1(\mathcal{G})$*, then* $\mathbb{E}[X \,|\, \mathcal{G}] = X$ *a.s. trivially.*

EX 2.4 *If* $\mathcal{G} = \{\emptyset, \Omega\}$*, then* $\mathbb{E}[X \,|\, \mathcal{G}] = \mathbb{E}[X]$*.*

EX 2.5 *Let* $A, B \in \mathcal{F}$ *with* $0 < \mathbb{P}[B] < 1$*. If* $\mathcal{G} = \{\emptyset, B, B^c, \Omega\}$ *and* $X = \mathbb{1}_A$*, then*

$$\mathbb{P}[A \,|\, \mathcal{G}] = \begin{cases} \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}, & \text{on } \omega \in B \\ \frac{\mathbb{P}[A \cap B^c]}{\mathbb{P}[B^c]}, & \text{on } \omega \in B^c \end{cases}$$

3 Conditional expectation: properties

We show that conditional expectations behave the way one would expect. Below all Xs are in $\mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G}$ is a sub σ-field of $\mathcal{F}$.

3.1 Extending properties of standard expectations

LEM 2.6 (cLIN) $\mathbb{E}[a_1 X_1 + a_2 X_2 \,|\, \mathcal{G}] = a_1 \mathbb{E}[X_1 \,|\, \mathcal{G}] + a_2 \mathbb{E}[X_2 \,|\, \mathcal{G}]$ *a.s.*

Proof: Use linearity of expectation and the fact that a linear combination of RVs in $\mathcal{G}$ is also in $\mathcal{G}$. ■

LEM 2.7 (cPOS) *If* $X \geq 0$ *then* $\mathbb{E}[X \,|\, \mathcal{G}] \geq 0$ *a.s.*

Proof: Let $Y = \mathbb{E}[X \,|\, \mathcal{G}]$ and assume $\mathbb{P}[Y < 0] > 0$. There is $n \geq 1$ s.t. $\mathbb{P}[Y < -n^{-1}] > 0$. But that implies, for $G = \{Y < -n^{-1}\}$,

$$\mathbb{E}[X; G] = \mathbb{E}[Y; G] < -n^{-1}\mathbb{P}[G] < 0,$$

a contradiction. ■

LEM 2.8 (cMON) *If* $0 \leq X_n \uparrow X$ *then* $\mathbb{E}[X_n \,|\, \mathcal{G}] \uparrow \mathbb{E}[X \,|\, \mathcal{G}]$ *a.s.*

Proof: Let $Y_n = \mathbb{E}[X_n \,|\, \mathcal{G}]$. By (cLIN) and (cPOS), $0 \leq Y_n \uparrow$. Then letting $Y = \limsup Y_n$, by (MON),

$$\mathbb{E}[X; G] = \mathbb{E}[Y; G],$$

for all $G \in \mathcal{G}$. ■

LEM 2.9 (cFATOU) *If* $X_n \geq 0$ *then* $\mathbb{E}[\liminf X_n \,|\, \mathcal{G}] \leq \liminf \mathbb{E}[X_n \,|\, \mathcal{G}]$ *a.s.*

Proof: Note that, for $n \geq m$,

$$X_n \geq Z_m \equiv \inf_{k \geq m} X_m \uparrow \in \mathcal{G},$$

so that $\inf_{n\geq m} \mathbb{E}[X_n \,|\, \mathcal{G}] \geq \mathbb{E}[Z_m \,|\, \mathcal{G}]$. Applying (cMON)

$$\mathbb{E}[\lim Z_m \,|\, \mathcal{G}] = \lim \mathbb{E}[Z_m \,|\, \mathcal{G}] \leq \lim \inf_{n \geq m} \mathbb{E}[X_n \,|\, \mathcal{G}].$$

■

LEM 2.10 (cDOM) *If $X_n \leq V \in \mathcal{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $X_n \to X$ a.s., then*

$$\mathbb{E}[X_n \,|\, \mathcal{G}] \to \mathbb{E}[X \,|\, \mathcal{G}]$$

Proof: Apply (cFATOU) to $W_n = 2V - |X_n - X| \geq 0$

$$\mathbb{E}[2V \,|\, \mathcal{G}] = \mathbb{E}[\liminf W_n] \leq \liminf \mathbb{E}[W_n \,|\, \mathcal{G}] = \mathbb{E}[2V \,|\, \mathcal{G}] - \liminf \mathbb{E}[|X_n - X| \,|\, \mathcal{G}].$$

Use that, by definition, $|\mathbb{E}[X_n - X \,|\, \mathcal{G}]| \leq \mathbb{E}[|X_n - X| \,|\, \mathcal{G}]$. ■

LEM 2.11 (cJENSEN) *If f is convex and $\mathbb{E}[|f(X)|] < +\infty$ then*

$$f(\mathbb{E}[X \,|\, \mathcal{G}]) \leq \mathbb{E}[f(X) \,|\, \mathcal{G}].$$

Proof: Exercise! ■

3.2 Other properties

LEM 2.12 (Tower) *If $\mathcal{H} \subseteq \mathcal{G}$ is a σ-field*

$$\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}] \,|\, \mathcal{H}] = \mathbb{E}[X \,|\, \mathcal{H}].$$

In particular $\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}]] = \mathbb{E}[X]$.

Proof: Let $Y = \mathbb{E}[X \,|\, \mathcal{G}]$ and $Z = \mathbb{E}[X \,|\, \mathcal{H}]$. Then $Z \in \mathcal{H}$ and for $H \in \mathcal{H} \subseteq \mathcal{G}$

$$\mathbb{E}[Z; H] = \mathbb{E}[X; H] = \mathbb{E}[Y; H].$$

■

LEM 2.13 (Taking out what is known) *If $Z \in \mathcal{G}$ is bounded then*

$$\mathbb{E}[ZX \,|\, \mathcal{G}] = Z\mathbb{E}[X \,|\, \mathcal{G}].$$

This is also true if $X, Z \geq 0$ and $\mathbb{E}[ZX] < +\infty$ or $X \in \mathcal{L}^p(\mathcal{F})$ and $Z \in \mathcal{L}^q(\mathcal{G})$ with $p^{-1} + q^{-1} = 1$ and $p > 1$.

Proof: By (LIN), we restrict ourselves to $X \geq 0$. Clear if $Z = \mathbb{1}_{G'}$ is an indicator with $G' \in \mathcal{G}$ since

$$\mathbb{E}[\mathbb{1}_{G'} X; G] = \mathbb{E}[X; G \cap G'] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]; G \cap G'] = \mathbb{E}[\mathbb{1}_{G'} \mathbb{E}[X \mid \mathcal{G}]; G],$$

for all $G \in \mathcal{G}$. Use the standard machine to conclude. ∎

LEM 2.14 (Role of independence) *If $\mathcal{H}$ is independent of $\sigma(\sigma(X), \mathcal{G})$, then*

$$\mathbb{E}[X \mid \sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X \mid \mathcal{G}].$$

In particular, if X is independent of $\mathcal{H}$ then $\mathbb{E}[X \mid \mathcal{H}] = \mathbb{E}[X]$.

Proof: Let $H \in \mathcal{H}$ and $G \in \mathcal{G}$. Since $Y = \mathbb{E}[X \mid \mathcal{G}] \in \mathcal{G}$, we have

$$\mathbb{E}[X; G \cap H] = \mathbb{E}[X; G]\mathbb{P}[H] = \mathbb{E}[Y; G]\mathbb{P}[H] = \mathbb{E}[Y; G \cap H].$$

We conclude with the following lemma.

LEM 2.15 (Uniqueness of extension) *Let $\mathcal{I}$ be a π-system on a set S, that is, a family of subsets stable under intersection. If μ_1, μ_2 are finite measures on $(S, \sigma(\mathcal{I}))$ with $\mu_1(\Omega) = \mu_2(\Omega)$ that agree on $\mathcal{I}$, then μ_1 and μ_2 agree on $\sigma(\mathcal{I})$.*

Indeed, note that the collection $\mathcal{I}$ of sets $G \cap H$ for $G \in \mathcal{G}, H \in \mathcal{H}$ form a π-system generating $\sigma(\mathcal{G}, \mathcal{H})$. ∎

Further reading

Regular conditional probability [Dur10, Section 5.1]. π-λ theorem [Dur10, Section A.1].

References

[Dur10] Rick Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, fourth edition, 2010.

[Wil91] David Williams. *Probability with martingales*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge, 1991.