Notes 13 : Conditioning

Math 733-734: Theory of Probability *Lecturer: Sebastien Roch*

References: [Wil91, Sections 0, 4.8, 9, 10], [Dur10, Section 5.1, 5.2], [KT75, Section 6.1].

1 Conditioning

1.1 Review of undergraduate conditional probability

1.1.1 Conditional probability

For two events A, B, the conditional probability of A given B is defined as

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}.$$

We assume $\mathbb{P}[B] > 0$.

1.1.2 Conditional expectation

Let X and Z be RVs taking values $x_1, \ldots, x_m$ and $z_1, \ldots, z_n$ resp. The conditional expectation of X given $Z = z_j$ is given as

$$y_j \equiv \mathbb{E}[X \mid Z = z_j] = \sum_i x_i \mathbb{P}[X = x_i \mid Z = z_j].$$

We assume $\mathbb{P}[Z = z_j] > 0$.

As motivation for the general definition, we make the following observations:

- We can think of the conditional expectation as a RV $Y = \mathbb{E}[X \mid Z]$ defined as follows:
$$Y(\omega) = y_j, \text{ on } G_j \equiv \{\omega : Z(\omega) = z_j\}.$$
- Then Y is $\mathcal{G}$-measurable where $\mathcal{G} = \sigma(Z)$.

- On sets in $\mathcal{G}$, the expectation of Y agrees with the expectation of X, that is,

$$\begin{aligned} \mathbb{E}[Y; G_j] &= y_j \mathbb{P}[G_j] \\ &= \sum_i x_i \mathbb{P}[X = x_i \,|\, Z = z_j] \mathbb{P}[Z = z_j] \\ &= \sum_i x_i \mathbb{P}[X = x_i, Z = z_j] \\ &= \mathbb{E}[X; G_j]. \end{aligned}$$

  This is also true for all $G \in \mathcal{G}$ by summation.

1.2 Conditional expectation: definition, existence, uniqueness

1.2.1 Definition

DEF&THM 13.1 *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 $Y \in \mathcal{L}^1(\Omega, \mathcal{G}, \mathbb{P})$ (note the $\mathcal{G}$-measurability) s.t.*

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

Such Y is called a version of $\mathbb{E}[X \,|\, \mathcal{G}]$. (E.g., see example above.)

1.2.2 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.

1.2.3 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. (Gives a more geometric perspective.)

We begin with a definition. Let $\langle U, V \rangle = \mathbb{E}[UV]$.

DEF&THM 13.2 *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. $\|X - Y_n\|_2 \to \Delta$. Recalling that $L^2(\Omega, \mathcal{G}, \mathbb{P})$ as a Hilbert space 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 RHS is at least $2\Delta^2$ by definition of Δ, so taking limits $r, s \to +\infty$ we have what we need.

Let Y be the limit of (Y_n) in $L^2(\Omega, \mathcal{G}, \mathbb{P})$. Note that

$$\Delta \leq \|X - Y\|_2 \leq \|X - Y_n\|_2 + \|Y_n - Y\|_2 \to \Delta.$$

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

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

so that, expanding and rearranging, we have

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

which is only possible *for every* $t \in \mathbb{R}$ if the first term is 0.

Uniqueness follows from the parallelogram law and the definition of Δ. ■

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 (see LEM 13.8 below), it follows that $U \geq 0$ implies $\mathbb{E}[U \,|\, \mathcal{G}] \geq 0$ for a simple function U. Using linearity (which is immediate from the definition), we then have $Y^{(r)} \uparrow Y \equiv \limsup Y^{(r)}$ which is measurable in $\mathcal{G}$. By (MON)

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

1.2.4 Examples

EX 13.3 *If* $X \in \mathcal{L}^1(\mathcal{G})$ *then* $\mathbb{E}[X \,|\, \mathcal{G}] = X$ *a.s. trivially.*

EX 13.4 *If* $\mathcal{G} = \{\emptyset, \Omega\}$*, then* $\mathbb{E}[X \,|\, \mathcal{G}] = \mathbb{E}[X]$.

EX 13.5 *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*

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

Intuition about conditional expectation sometimes breaks down:

EX 13.6 *On* $(\Omega, \mathcal{F}, \mathbb{P}) = ((0,1], \mathcal{B}(0,1], \mathrm{Leb})$*, let* $\mathcal{G}$ *be the* σ*-field of all countable and co-countable subsets of* $(0,1)$*. Then* $\mathbb{P}[G] \in \{0,1\}$ *for all* $G \in \mathcal{G}$ *and*

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

so that $\mathbb{E}[X \,|\, \mathcal{G}] = \mathbb{E}[X]$*. Yet,* $\mathcal{G}$ *contains all singletons and we seemingly have full information, which would lead to the wrong guess* $\mathbb{E}[X \,|\, \mathcal{G}] = X$.

1.3 Conditional expectation: properties

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

1.3.1 Extending properties of ordinary expectations

LEM 13.7 (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 13.8 (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 13.9 (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 13.10 (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 13.11 (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 13.12 (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! ■

1.3.2 More properties

The next two properties provide some insight into the interpretation of the conditional expectation.

LEM 13.13 (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

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

for all $G \in \mathcal{G}$. Use the standard machinery to conclude. ■

LEM 13.14 (Role of independence) *If X is independent of $\mathcal{H}$ then $\mathbb{E}[X \,|\, \mathcal{H}] = \mathbb{E}[X]$. In fact, if $\mathcal{H}$ is independent of $\sigma(\sigma(X), \mathcal{G})$, then*

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

Proof: By taking positive and negative parts, we can assume that $X \geq 0$. Let $H \in \mathcal{H}$ and $G \in \mathcal{G}$. Since $Y = \mathbb{E}[X \,|\, \mathcal{G}] \in \mathcal{G}$, we have

$$\mu_1(G \cap H) \equiv \mathbb{E}[X; G \cap H] = \mathbb{E}[X; G]\mathbb{P}[H] = \mathbb{E}[Y; G]\mathbb{P}[H] = \mathbb{E}[Y; G \cap H] \equiv \mu_2(G \cap H).$$

We conclude with the following lemma.

LEM 13.15 (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 $\sigma(\mathcal{G}, \mathcal{H})$. (Clearly, $\mathcal{I} \subseteq \sigma(\mathcal{G}, \mathcal{H})$ so $\sigma(\mathcal{I}) \subseteq \sigma(\mathcal{G}, \mathcal{H})$. Moreover $\mathcal{G}, \mathcal{H} \subseteq \mathcal{I} \subseteq \sigma(\mathcal{I})$ so $\sigma(\mathcal{G}, \mathcal{H}) \subseteq \sigma(\mathcal{I})$.) ■

1.3.3 Law of total probability

The following is often useful in computations.

LEM 13.16 (Tower) *We have* $\mathbb{E}[\mathbb{E}[X \,|\, \mathcal{G}]] = \mathbb{E}[X]$. *In fact, if* $\mathcal{H} \subseteq \mathcal{G}$ *is a* σ*-field*

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

(i.e., the smallest σ*-field wins).*

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

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

■

1.4 Regular conditional probability

The conditional probability of $A \in \mathcal{F}$ given $\mathcal{G}$ is

$$\mathbb{P}[A \,|\, \mathcal{G}] = \mathbb{E}[\mathbb{1}_A \,|\, \mathcal{G}].$$

For fixed A, $\mathbb{P}[A \,|\, \mathcal{G}]$ is a RV. What about the opposite? For fixed $\omega \in \Omega$, is $\mathbb{P}[\cdot \,|\, \mathcal{G}]$ a probability measure a.s.? The answer is, unfortunately, not always.

DEF 13.17 *The map* $\mu : \Omega \times \mathcal{F} \to [0, 1]$ *is a* regular conditional probability *given* $\mathcal{G}$ *if:*

- *For each* $A \in \mathcal{F}$, $\mu(\cdot, A)$ *is a version of* $\mathbb{P}[A \,|\, \mathcal{G}]$.
- *For almost every* $\omega \in \Omega$, $\mu(\omega, \cdot)$ *is a probability measure on* $\mathcal{F}$.

(They are known to exist on "nice" spaces. See [Dur10].)

EX 13.18 *Let* (X, Y) *have joint density* $f_{X,Y}$. *For simplicity, assume* $f_Y(y) \equiv \int f_{X,Y}(x, y)dx > 0$ *for all* y. *Define*

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

Then,

$$\mu(\omega, B) \equiv \mathbb{P}[X \in B \,|\, Y](\omega) \equiv \int_B f_{X|Y}(x|Y(\omega))dx,$$

is a regular conditional distribution function. Indeed, for

$$G = \{Y \in B'\} \in \mathcal{G} = \sigma(Y),$$

we have

$$\begin{aligned}
\mathbb{E}[\mathbb{1}_{X\in B};G] &= \int_B \int_{B'} f_{X,Y}(x,y)dxdy \\
&= \int_B \int_{B'} f_Y(y) f_{X|Y}(x|y)dxdy \\
&= \int_{B'} f_Y(y) \left(\int_B f_{X|Y}(x|y)dx \right) dy \\
&= \mathbb{E}[\mathbb{P}[X \in B \mid Y]; G],
\end{aligned}$$

by Fubini (where note that everything is non-negative).

References

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

[KT75] Samuel Karlin and Howard M. Taylor. *A first course in stochastic processes*. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1975.

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