Beginning of 01/08/2024

Conditional Expectation

Some quick recap on concepts and notation...

Consider a probability space $(\Omega, \mathcal{F}_0, \mathbb{P})$. Fix a sub σ-algebra $\mathcal{F} \subset \mathcal{F}_0$. Let X be a random variable on $\mathcal{F}_0$, i.e., $X : \Omega \to \mathbb{R}$ measurable w.r.t. $\mathcal{F}_0$ ($X^{-1}(A) \in \mathcal{F}_0$ for all Borel set A, i.e., for all $A \in \mathcal{R}$). Also assume $X \in L^1$, i.e., $\mathbb{E}[|X|] < \infty$.

We define the **conditional expectation of $\boldsymbol{X}$ given $\boldsymbol{\mathcal{F} \subset \mathcal{F}_0}$** to be any r.v. Y such that:

(i) $Y \in \mathcal{F}$ (measurable: $Y^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{R}$), and

(ii) $\int_A Y \, d\mathbb{P} = \int_A X \, d\mathbb{P}$ for all $A \in \mathcal{F}$. (The RHS is $\mathbb{E}[X\mathbf{1}[A]]$, a well-defined expectation).

In order for this definition to make sense, we will prove the following statements:

(1) Existence: there exists Y satisfying (i) and (ii).

(2) Integrability: any Y that satisfies (i) and (ii) is L^1 (integrable).

(3) Uniqueness: if Y and Y' both satisfy the conditions, then $Y = Y'$ a.s.

With these properties established, we denote any satisfying variable by $\mathbb{E}[X|\mathcal{F}]$. *This can also be viewed as an equivalence class in the sense that any member inside this equivalence class has the same* $\int_A Y \, d\mathbb{P}$. *Each element is called a version.*

Proof. We first prove *integrability* (**D4.1.1**). If Y satisfies (i) and (ii) then $Y \in L^1$. Let $A := \{Y > 0\} = \{\omega \in \Omega : Y(\omega) > 0\}$. Note $A \in \mathcal{F}$ since $A = Y^{-1}((0, \infty))$ and consequently $A^c \in \mathcal{F}$ by closure of complement. Then

$$\int_A |Y| d\mathbb{P} = \int_A Y \, d\mathbb{P} = \int_A X \, d\mathbb{P} \leqslant \int_A |X| \, d\mathbb{P}$$

and

$$\int_{A^c} |Y| \, d\mathbb{P} = \int_{A^c} -Y \, d\mathbb{P} = -\int_{A^c} Y \, d\mathbb{P} = -\int_{A^c} X \, d\mathbb{P} = \int_{A^c} (-X) \, d\mathbb{P} \leqslant \int_{A^c} |X| \, d\mathbb{P}.$$

Adding these two gives $\int_\Omega |Y| \, d\mathbb{P} \leqslant \int_\Omega |X| \, d\mathbb{P} < \infty$.

For *uniqueness*: suppose Y, Y' both satisfy (i) and (ii). Let

$$A_n := \{Y - Y' \geqslant 1/n\}.$$

Since Y, Y' are $\mathcal{F}$-measurable, so is there difference, so $A_n \in \mathcal{F}$ for all n. Then

$$\begin{aligned} 0 &= \int_{A_n} (X - X) \, d\mathbb{P} = \int_{A_n} X \, d\mathbb{P} - \int_{A_n} X \, d\mathbb{P} = \int_{A_n} Y \, d\mathbb{P} - \int_{A_n} Y' \, d\mathbb{P} \\ &= \int_{A_n} (Y - Y') \, d\mathbb{P} \geqslant \int_{A_n} n^{-1} \, d\mathbb{P} = n^{-1}\mathbb{P}(A_n) \geqslant 0. \end{aligned}$$

Note subtraction is well-defined since both integrals involving Y, Y' are finite by L^1 assumption. But then $P(A_n) = 0$, so $\mathbb{P}(Y > Y') = \mathbb{P}(\cup A_n) \leqslant \sum \mathbb{P}(A_n) = 0$. Identical proof for the other direction.

We'll (informally) prove existence later. □

Theorem: (D4.1.2)

If $X_1 = X_2$ on $B \in \mathcal{F}$, then $\mathbb{E}[X_1|\mathcal{F}] = \mathbb{E}[X_2|\mathcal{F}]$ on B.

Proof. Put $Y_1 = \mathbb{E}[X_1|\mathcal{F}]$ and Y_2 similarly. Let $A_n = \{Y_1 - Y_2 \geqslant 1/n\}$. Then

$$\begin{aligned} 0 &= \int_{A_n \cap B} (X_1 - X_2)\, \mathrm{d}\mathbb{P} = \int_{A_n \cap B} X_1\, \mathrm{d}\mathbb{P} - \int_{A_n \cap B} X_2\, \mathrm{d}\mathbb{P} \\ &= \int_{A_n \cap B} Y_1\, \mathrm{d}\mathbb{P} - \int_{A_n \cap B} Y_2\, \mathrm{d}\mathbb{P} = \int_{A_n \cap B} (Y_1 - Y_2)\, \mathrm{d}\mathbb{P} \\ &\geqslant n^{-1}\mathbb{P}(A_n \cap B). \end{aligned}$$

Hence $\mathbb{P}(A_n \cap B) = 0$. Therefore $\mathbb{P}(Y_1 > Y_2) = \mathbb{P}(\cup(A_n \cap B)) = 0$. Other direction analogous. □

Proof of existence of conditional variables. This proof quotes results from **Lebesgue-Radon-Nikodym** (525a, Durrett appendix A4). Let μ, ν be two σ-finite measures on $(\Omega, \mathcal{F})$ (making the entire space the union of countably many finite-measure sets). We assume that $\nu \ll \nu$ (*absolutely continuous*): if $\mu(A) = 0$ then $\nu(A) = 0$. Then there exists $f \in \mathcal{F}$ such that for all $A \in \mathcal{F}$,

$$\nu(A) = \int_A f\, \mathrm{d}\mu.$$

*Intuition: f is the density of ν. More formally, this is called the **Radon-Nikodym derivative**, written* $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$.

Going back to the proof. First suppose $X \geqslant 0$. Let $\nu(A) := \int_A X\, \mathrm{d}\mathbb{P}$. By non-negativity assumption of X, $\nu : \mathcal{F} \to [0, \infty)$. Finite because $X \in L^1$. It follows that

- ν is a measure: $\nu(\varnothing) = 0$ and countably additive via either MCT or DCT: $\nu(\sqcup A_n) = \sum \nu(A_n)$ and $\int_{\sqcup A_n} X\, \mathrm{d}\mathbb{P} = \sum \int_{A_n} X\, \mathrm{d}\mathbb{P}$.
- $\nu(\Omega) = \int_\Omega X\, \mathrm{d}\mathbb{P} < \infty$, a finite measure, and
- $\nu \ll \mathbb{P}$, clearly.

Therefore by L-R-N, there exists a random variable Y such that for all $\mathcal{A} \in \mathcal{F}$, $\int_A Y\, \mathrm{d}\mathbb{P} = \nu(A) = \int_A X\, \mathrm{d}\mathbb{P}$.

Now, for general $X \in L^1$ we sign decompose it into $X^+ = \max(0, X)$ and $X^- = \max(0, -X)$. We know

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

exist. Let $Y = Y^+ - Y^-$. It's clear that (i) is satisfied. To see (ii):

$$\int_A Y\, \mathrm{d}\mathbb{P} = \int_A Y^+\, \mathrm{d}\mathbb{P} - \int_A Y^-\, \mathrm{d}\mathbb{P} = \int_A X^+\, \mathrm{d}\mathbb{P} - \int_A X^-\, \mathrm{d}\mathbb{P} = \int_A X\, \mathrm{d}\mathbb{P},$$

where again we used the assumption that everything is L^1. □

Some examples of conditionals:

- (**D4.1.3**, *perfect information*) If $X \in \mathcal{F}$, then $\mathbb{E}[X|\mathcal{F}] = X$. Both definition checks are trivial.
- (**D4.1.4**, *no information*) If X is independent of $\mathcal{F}$, i.e. $\mathbb{P}(X \in B \cap A) = \mathbb{P}(X \in B)\mathbb{P}(A)$ for all $B \in \mathcal{R}, A \in \mathcal{F}$,

then $\mathbb{E}[X|\mathcal{F}] = \mathbb{E}X$, a constant variable. Constants are clearly $\mathcal{F}$-measurable, since the inverse image is either $\varnothing$ or Ω, so (i) passes. And for (ii), by independence,

$$\int_A X \, \mathrm{d}\mathbb{P} = \mathbb{E}[X\mathbf{1}[A]] = \mathbb{E}X \cdot \mathbb{E}[\mathbf{1}[A]] = \int_A \mathbb{E}X \, \mathrm{d}\mathbb{P}.$$