Conditional Expectation

10.1 Conditional Expectation

10.1.1 Definition, existence and uniqueness

Definition 10.1. *Given σ-fields $\mathcal{G} \subseteq \mathcal{F}$ and a r.v $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, we define $\mathbb{E}(X|\mathcal{G})$ as a r.v Y s.t.*

1. *Y is in $L^1(\Omega, \mathcal{G}, \mathbb{P})$.*
2. *$\mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(Y\mathbb{1}_A)\ \forall\ A \in \mathcal{G} \ \Leftrightarrow\ \mathbb{E}((X-Y)Z) = 0$ for all bounded $\mathcal{G}$-measurable function Z.*

Lemma 10.2. *Conditional expectation, if exists, is unique a.s.*

Proof. Suppose Y_1, Y_2 are conditional expectations of X given $\mathcal{G}$. Take $W := Y_1 - Y_2$ which is $\mathcal{G}$ measurable. Then $\mathbb{E}(W\mathbb{1}_A) = 0$ for all $A \in \mathcal{G}$. Take $A = \{W > \varepsilon\} \in \mathcal{G}$. Then,

$$0 = \mathbb{E}(W 1_{W>\varepsilon}) > \varepsilon\, \mathbb{P}(W > \varepsilon) \implies \mathbb{P}(W > \varepsilon) = 0$$

Similarly, taking $A = \{W < -\varepsilon\} \in \mathcal{G}$ we get

$$0 = \mathbb{E}(W 1_{W<-\varepsilon}) < -\varepsilon\, \mathbb{P}(W < -\varepsilon) \implies \mathbb{P}(W < -\varepsilon) = 0$$

Therefor $P(W \in [-\varepsilon, \varepsilon]) = 1$ for all $\varepsilon > 0$. Hence $\mathbb{P}(W = 0) = 1$ and $Y_1 = Y_2$ a.s. ■

We will prove that

Theorem 10.3. *Conditional expectation exists.*

We will prove this later.

10.1.2 Properties of Conditional Expectation

Let $\mathcal{G} \subseteq \mathcal{F}$ be σ-fields, then $\mathbb{E}(\cdot \mid \mathcal{G}) : L^1(\mathcal{F}) \to L^1(\mathcal{G})$ where $L^p(\mathcal{H}) := L^p(\Omega, \mathcal{H}, \mathbb{P}), p \geqslant 1$ for a sub σ-field $\mathcal{H}$ of $\mathcal{F}$.

(i) **Positive**: $X \geqslant 0$ a.s. $\implies \widehat{X} := \mathbb{E}(X \mid \mathcal{G}) \geqslant 0$ a.s.

Proof: Using $\mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(\widehat{X}\mathbb{1}_A)\ \forall\ A \in \mathcal{G}$; for $A = \{\widehat{X} < 0\}$, we get $\mathbb{E}(\widehat{X}\mathbb{1}_{\widehat{X}<0}) = 0$ which implies $\widehat{X} \geqslant 0$ a.s.

(ii) **Linear**: $\mathbb{E}(X + Y \mid \mathcal{G}) = \mathbb{E}(X \mid \mathcal{G}) + \mathbb{E}(Y \mid \mathcal{G})$ a.s. and $\mathbb{E}(cX \mid \mathcal{G}) = c\,\mathbb{E}(X \mid \mathcal{G})$ a.s. for $c \in \mathbb{R}$.

(iii) **Contractive on** $L^p(\mathcal{F}) \to L^p(\mathcal{G})$: For $X \in L^p(\mathcal{F}), p \geqslant 1$ we have $\mathbb{E}(X \mid \mathcal{G}) \in L^p(\mathcal{G})$ and $\|\mathbb{E}(X \mid \mathcal{G})\|_p \leqslant \|X\|_p$.

Proof for for $p = 1$: Let $\widehat{X} = \mathbb{E}(X \mid \mathcal{G})$, we have $\mathbb{E}(X \mathbb{1}_A) = \mathbb{E}(\widehat{X} \mathbb{1}_A)$ for all $A \in \mathcal{G}$. Taking $A_1 = \{\widehat{X} \geqslant 0\}$, $A_2 = \{\widehat{X} < 0\}$, we have,

$$\mathbb{E}\,|\widehat{X}| = \mathbb{E}(\widehat{X} \mathbb{1}_{\widehat{X} \geqslant 0}) - \mathbb{E}(\widehat{X} \mathbb{1}_{\widehat{X} < 0}) = \mathbb{E}(X \mathbb{1}_{\widehat{X} \geqslant 0}) - \mathbb{E}(X \mathbb{1}_{\widehat{X} < 0}) \leqslant \mathbb{E}\,|X|\,.$$

For general $p > 1$, use $\mathbb{E}(XZ) = \mathbb{E}(\widehat{X}Z)$ for any bounded $\mathcal{G}$-measurable Z. The r.v. $Z_n = |\hat{X}|^{p-1} \cdot (\mathbb{1}_{0 < \hat{X} \leqslant n} - \mathbb{1}_{0 < -\hat{X} \leqslant n})$ is bounded and $\mathcal{G}$-measurable. Thus $\mathbb{E}(\hat{X} Z_n) = \mathbb{E}(X Z_n)$ implies that

$$\mathbb{E}\,|\hat{X} \mathbb{1}_{|\hat{X}| \leqslant n}|^p = \mathbb{E}(X Z_n) \leqslant \mathbb{E}(|X| \cdot |\hat{X} \mathbb{1}_{|\hat{X}| \leqslant n}|^{p-1}) \leqslant (\mathbb{E}\,|X|^p)^{1/p} (\mathbb{E}\,|\hat{X} \mathbb{1}_{|\hat{X}| \leqslant n}|^p)^{1-1/p}$$

where we used Hölder inequality in the last line. Thus we have

$$\mathbb{E}\,|\hat{X} \mathbb{1}_{|\hat{X}| \leqslant n}|^p \leqslant \mathbb{E}\,|X|^p \text{ for all } n \geqslant 1.$$

Taking limit $n \to \infty$ we get the results.

(iv) **William's Tower Property**: Let $\mathcal{G} \subseteq \mathcal{H}$. Suppose $\mathbb{E}(\cdot \mid \mathcal{G})$ and $\mathbb{E}(\cdot \mid \mathcal{H})$ are well defined, then $\mathbb{E}(\mathbb{E}(X \mid \mathcal{H}) \mid \mathcal{G}) = \mathbb{E}(X \mid \mathcal{G})$.

(v) If X is $\mathcal{G}$-measurable then $\mathbb{E}(X \mid \mathcal{G}) = X$ a.s.

(vi) **Projection**: $\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{G}) = \mathbb{E}(X \mid \mathcal{G})$, which follows from Tower property.

(vii) **Monotone**: $X \geqslant Y$ a.s. $\implies$ $\mathbb{E}(X \mid \mathcal{G}) \geqslant \mathbb{E}(Y \mid \mathcal{G})$ a.s.

(viii) **Conditional MCT**: $X_n \geqslant 0, X_n \uparrow X \implies \mathbb{E}(X_n \mid \mathcal{G}) \uparrow \mathbb{E}(X \mid \mathcal{G})$ a.s.

(ix) $\mathbb{E}(X \mid \{\emptyset, \Omega\}) = \mathbb{E}\,X$.

(x) **Jensen's Inequality**: If ϕ is convex and $\mathbb{E}\,|\phi(X)| < \infty$, $\phi(\mathbb{E}(X \mid \mathcal{G})) \leqslant \mathbb{E}(\phi(X) \mid \mathcal{G})$ a.s.

(xi) **Cauchy-Schwartz and Hölder Inequality**: $|\mathbb{E}(XY \mid \mathcal{G})| \leqslant (\mathbb{E}(|X|^p \mid \mathcal{G}))^{1/p} (\mathbb{E}(|Y|^q \mid \mathcal{G}))^{1/q}$ a.s. for $1/p + 1/q = 1, p, q \geqslant 1$ and $X \in L^p(\mathcal{F}), Y \in L^q(\mathcal{F})$.

(xii) Suppose $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_\infty = \sigma\big(\cup_{n \geqslant 0} \mathcal{F}_n\big) \subseteq \mathcal{F}$, and $X \in L^p(\mathcal{F})$, $p \geqslant 1$. Then $\mathbb{E}(X|\mathcal{F}_n) \overset{\text{a.s.}/L^p}{\longrightarrow} \mathbb{E}(X|\mathcal{F}_\infty)$. In particular, if $\mathcal{F}_\infty = \mathcal{F}$ then $\mathbb{E}(X \mid \mathcal{F}_n) \overset{\text{a.s.}/L^p}{\longrightarrow} X$.

(xiii) $\big(\widehat{X}_n = \mathbb{E}(X|\mathcal{F}_n)\big)_{n \geqslant 0}$ is adapted to $\big(\mathcal{F}_n\big)_{n \geqslant 0}$ and $\mathbb{E}(\widehat{X}_n \mid \mathcal{F}_{n-1}) = \widehat{X}_{n-1}$.

10.1.3 Proof of Theorem 10.3: Construction of $\mathbb{E}(\cdot \mid \mathcal{G})$

Here are a number of proofs for Theorem 10.3.

(***Measure theoretic proof***). We will use the following theorem from measure theory.

Theorem 10.4 (Lebesgue-Radon-Nikodym derivative). *Let μ and λ be two finite positive measures on $(\Omega, \mathcal{F})$ such that $\mu \ll \lambda$ (μ is absolutely continuous w.r.t. λ, i.e., $\lambda(A) = 0 \implies \mu(A) = 0$). Then there exists a measurable function $f \in L^1(\Omega, \mathcal{F}, \lambda)$ s.t.,*

$$f = \frac{d\mu}{d\lambda} \Leftrightarrow \mu(A) = \int_A f d\lambda \;\forall\; A \in \mathcal{F}.$$

Assume that $X \geqslant 0$. Consider Theorem 10.4 with $\lambda := \mathbb{P}$ and $\mu(A) := \int_A X d\,\mathbb{P}$ for all $A \in \mathcal{G}$. λ, μ are positive finite measures on $(\Omega, \mathcal{G})$ and $\mu \ll \lambda$. Then by Theorem 10.4 there exists $f \in L^1(\Omega, \mathcal{G}, \mathbb{P})$ s.t.

$$\mu(A) = \int_A f d\,\mathbb{P} \implies \mathbb{E}(X 1_A) = \mathbb{E}(f 1_A)\ \forall A \in \mathcal{G}.$$

So f is the conditional expectation. In general $X = X^+ - X^-$ and we define

$$\mathbb{E}(X \mid \mathcal{G}) = \mathbb{E}(X^+ \mid \mathcal{G}) - \mathbb{E}(X^- \mid \mathcal{G}).$$ ■

(***Functional analysis proof***). We will use the following lemma on orthogonal projection in Hilbert spaces.

Lemma 10.5. *Let $K \subseteq H$ be a close subspace of Hilbert space H. Then for all $X \in H$, there exists a unique decomposition $X = Y + Z$ s.t $Y \in K$ and $Z \in K^\perp$.*

Assume $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$ which is a Hilbert space. We will use shorthand notation $L^2(\mathcal{F})$ for $L^2(\Omega, \mathcal{F}, \mathbb{P})$. By Lemma 10.5 there exists a unique decomposition $X = Y + E$ such that $Y \in L^2(\mathcal{G})$ and $E \in L^2(\mathcal{G})^\perp$. So,

$$\forall\, Z \in L^2(\mathcal{G}), \quad \mathbb{E}((X - Y)Z) = 0 \implies Y = \mathbb{E}(X \mid \mathcal{G})$$

This proof can be generalized to $X \in L^1$. ■

(***Hands on proof***). (i) Assume that $|\mathcal{G}| < \infty$. Then $\mathcal{G} = \sigma(C_1, \ldots, C_k)$ such that $\{C_1, C_2, \ldots, C_k\}$ is a disjoint partition of Ω. Now any $\mathcal{G}$-measurable r.v. Z is of the form, $Z = \sum_{i=1}^k a_i \mathbb{1}_{C_i}$ for some $a_1, a_2, \ldots, a_k \in \mathbb{R}$. Then the conditional expectation is,

$$\mathbb{E}(X|\mathcal{G}) = Y := \sum_{i=1}^k \frac{\mathbb{E}(X\mathbb{1}_{C_i})}{\mathbb{E}(1_{C_i})} \mathbb{1}_{C_i},$$

since for all $\mathcal{G}$-measurable r.v. Z,

$$\mathbb{E}(YZ) = \sum_{i=1}^k a_i \sum_{j=1}^k \frac{\mathbb{E}(X\mathbb{1}_{C_j})}{\mathbb{E}(\mathbb{1}_{C_j})} \mathbb{E}(\mathbb{1}_{C_j}\mathbb{1}_{C_i}) = \sum_{i=1}^k a_i\, \mathbb{E}(X\mathbb{1}_{C_i}) = \mathbb{E}(XZ).$$

(ii) Suppose $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \cdots \subseteq \mathcal{G} = \sigma\big(\cup_{i \geqslant 1} \mathcal{G}_i\big)$ and $X \in L^2(\mathcal{F})$. Then, let

$$\widehat{X}_n = \mathbb{E}(X \mid \mathcal{G}_n).$$

We have $\|\widehat{X}_n\|_2 \leqslant \|X\|_2$. Let $\Delta_n = \widehat{X}_n - \widehat{X}_{n-1}, n \geqslant 1$. By William's Tower property we have, $\mathbb{E}(\widehat{X}_n \mid \mathcal{G}_{n-1}) = \widehat{X}_{n-1}$. Thus for any $L^2(\mathcal{G}_{n-1})$ r.v. Y, $\mathbb{E}((\widehat{X}_n - \widehat{X}_{n-1})Y) = 0$. Thus $\{\widehat{X}_1, \Delta_2, \Delta_3, \ldots\}$ are uncorrelated. By definition of Δ_n, we have $\widehat{X}_n = \widehat{X}_1 + \Delta_2 + \Delta_3 + \cdots + \Delta_n$. This implies $S_n^2 := \|\widehat{X}_n\|_2^2 = \|\widehat{X}_1\|_2^2 + \|\Delta_2\|_2^2 + \cdots + \|\Delta_n\|_2^2 \leqslant \|X\|_2$. Thus $\widehat{X}_n$ is L^2-Cauchy, since $S_n^2 \uparrow S_\infty^2 \leqslant \mathbb{E}\, X^2$. Therefore $\widehat{X}_n \xrightarrow{\mathbf{L^2}} \widehat{X}_\infty$.

Claim: $\widehat{X}_\infty$ is $\mathcal{G}$-measurable.

Claim: $\mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(\widehat{X}_\infty \mathbb{1}_A)$ for all $A \in \cup_{n \geqslant 1} \mathcal{G}_n$. To verify the claim, note that: $\exists\ n$ s.t. $A \in \mathcal{G}_m$ for $m \geqslant n$. Also $\mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(\widehat{X}_m \mathbb{1}_A), m \geqslant n$, and as $m \to \infty$, it converges to $\mathbb{E}(\widehat{X}_\infty \mathbb{1}_A)$.

Now, $\{A : \mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(\widehat{X}_\infty \mathbb{1}_A)\}$ is a λ system and $\cup_{n \geqslant 1} \mathcal{G}_n$ is a π-system. By $\pi - \lambda$ theorem, $\mathbb{E}(X\mathbb{1}_A) = \mathbb{E}(\widehat{X}_\infty \mathbb{1}_A)$ for all $A \in \sigma(\cup_{n \geqslant 1} \mathcal{G}_n) = \mathcal{G}$. Therefore, we have, $\widehat{X}_\infty = \mathbb{E}(X \mid \mathcal{G})$ a.s.

(iii) If $\mathcal{G} = \sigma(C_1, C_2, \ldots)$ then take $\mathcal{G}_n = \sigma(C_1, C_2, \ldots, C_n), n \geqslant 1$. In general, given $X \in L^2(\mathcal{F})$, $\mathcal{G} \subseteq \mathcal{F}$, $S_* := \sup_{|\mathcal{H}|<\infty, \mathcal{H} \subseteq \mathcal{G}} \|\mathbb{E}(X \mid \mathcal{H})\|_2$ exists. Take $\mathcal{H}_i$ s.t. $\|\mathbb{E}(X|\mathcal{H}_i]\|_2 \uparrow S_*$ and work with $\mathcal{G}_i = \sigma(\mathcal{H}_1 \cup \mathcal{H}_2 \cup \cdots \cup \mathcal{H}_i), i \geqslant 1$.

(iv) Finally, from $X \in L^2(\mathcal{F})$, one can generalize to $X \in L^1(\mathcal{F})$ by approximating a non-negative r.v. by a bounded r.v. monotonically.

■

10.1.4 Lévy's 0-1 Law

We have $\mathbb{P}(A \mid \mathcal{F}_n) \xrightarrow{a.s.} \mathbb{1}_A$ whenever $\mathcal{F}_n \uparrow \mathcal{F}$.

Exercise 10.6. *Let* $X_1, X_2, \ldots$ *be independent r.v.s and* $\mathcal{F}_n = \sigma(X_1, \cdots, X_n)$, $A \in \cap_{n \geqslant 1} \sigma(X_n, X_{n+1}, \ldots)$ *then* $\mathbb{P}(A) \in \{0, 1\}$.