

Probability theory II Facts

Taeyoung Chang

Textbook : Rick Durrett 「 Probability : Theory and Examples 」 5th edition

Last Update : September 18, 2021

Contents

1	Conditional Expectation	2
---	-------------------------	---

1 Conditional Expectation

- Projection Thm for Hilbert Space

- If E is a Hilbert space and $M \subset E$ is closed and convex, then for any $y \in E$,
 \exists a unique $w \in M$ s.t. $\|y - w\| = d(y, M) := \inf\{\|y - v\| : v \in M\}$.
Denote it as $w = \text{proj}_M y$ i.e. w is a projection of y onto M .
- If E is a Hilbert space and $M \subset E$ is a closed vector subspace, then for any $y \in E$,
 - \exists a unique decomposition $y = w + v$ with $w = \text{proj}_M y \in M$ and $v \in M^\perp$
 - For $w \in M$, $w = \text{proj}_M y \Leftrightarrow \langle y - w, z \rangle = 0 \quad \forall z \in M$

* $\mathcal{L}^2 := \{ \text{Random Variable } X : E(X^2) = \int X^2 dP < \infty \}$

✓ If $X \in \mathcal{L}^2$ then $E|X| < \infty$ i.e. every element of \mathcal{L}^2 is integrable.

★ Trick : $|X| \leq X^2 + \frac{1}{4}$

✓ \mathcal{L}^2 is a vector space

★ Trick : inequality $(aX + bY)^2 \leq 2(a^2X^2 + b^2Y^2)$

- \mathcal{L}^2 is a Hilbert space with inner product $\langle X, Y \rangle = E(XY)$

★ Trick : Cauchy seq. having a subseq. converging to a point converges to the point.

- Lemma for proving \mathcal{L}^2 is a complete normed space.

- If $\{X_n\} \subset \mathcal{L}^2$ and $\|X_n - X_{n+1}\| \leq 2^{-n} \quad \forall n \in \mathbb{N}$ then $\exists X \in \mathcal{L}^2$ s.t. $X_n \rightarrow X$ a.s. and $\|X_n - X\| \rightarrow 0$ i.e. $X_n \rightarrow X$ in \mathcal{L}^2 .

★ Lemma : If a random variable Z satisfies $Z \geq 0$ and $E(Z) < \infty$ then $Z < \infty$ a.s.

* For $X \in \mathcal{L}^2$, $\mathcal{L}^2(X) := \{g(X) \mid g : \mathbb{R} \rightarrow \mathbb{R} \text{ is a Borel function, } E[(g(X))^2] < \infty\}$

✓ For $X \in \mathcal{L}^2$, $\mathcal{L}^2(X)$ is a vector subspace of \mathcal{L}^2 .

- For $X \in \mathcal{L}^2$, $\mathcal{L}^2(X)$ is a closed vector subspace of \mathcal{L}^2 so that $\mathcal{L}^2(X)$ is also a Hilbert space.

* Geometric definition for conditional expectation

- For $X, Y \in \mathcal{L}^2$, define $E[Y|X] = \text{Proj}_{\mathcal{L}^2(X)} Y$
- $E[Y|X] = g(X)$ a.s. for some Borel function g
- $\|Y - E[Y|X]\| = \min_{h(X) \in \mathcal{L}^2(X)} \|Y - h(X)\|$
i.e. $E[(Y - E[Y|X])^2] \leq E[(Y - h(X))^2] \quad \forall h(X) \in \mathcal{L}^2$
- For $g(X) \in \mathcal{L}^2(X)$, $g(X) = E[Y|X] \Leftrightarrow \langle Y - g(X), h(X) \rangle = 0 \quad \forall h(X) \in \mathcal{L}^2$
 $\Leftrightarrow E[(Y - g(X))h(X)] = 0 \quad \forall h(X) \in \mathcal{L}^2$

- Elementary properties of conditional expectation from geometric definition

- If $X, Y, Z \in \mathcal{L}^2$ then the followings are true.
 - $E[c|X] = c$ a.s. $\forall c \in \mathbb{R}$
 - $E[\alpha Y + \beta Z|X] = \alpha E[Y|X] + \beta E[Z|X] \quad \forall \alpha, \beta \in \mathbb{R}$
 - $E[Y|X] = E[Y]$ if X and Y are independent.

iv. $E[g(X)Y|X] = g(X)E[Y|X]$ if g satisfies $g(X) \in \mathcal{L}^2(X)$ and $\sup_x |g(x)| < \infty$

v. $E[E[Y|X]] = E[Y]$

✓ In fact, the additional assumption about boundedness of g in (iv) is not necessary. We will see later.

- Extending the definition from \mathcal{L}^2 to all integrable functions

$$E[\{Y - E[Y|X]\}I(X \in A)] = 0 \quad \forall A \in \mathcal{B}(\mathbb{R}) \quad \because I(X \in A) \in \mathcal{L}^2(X)$$

$$\int_{(X \in A)} Y dP = \int_{(X \in A)} E[Y|X] dP \quad \forall A \in \mathcal{B}(\mathbb{R})$$

$$\int_B Y dP = \int_B E[Y|X] dP \quad \forall B \in \sigma(X)$$

- $E[Y|X] \in \sigma(X)$ and $\int_B Y dP = \int_B E[Y|X] dP \quad \forall B \in \sigma(X)$. Such r.v. is unique in the sense that if any r.v. Z satisfies $Z \in \sigma(X)$ and $\int_B Y dP = \int_B Z dP \quad \forall B \in \sigma(X)$ then $Z = E[Y|X]$ a.s. provided $E|Y| < \infty$
- From the theory on \mathcal{L}^2 space, we get geometric understanding about conditional expectation. But now, from the equation above, we can guess that definition for conditional expectation may be extended to all integrable random variables.

- Proof for the uniqueness mentioned above

- (Ω, \mathcal{F}, P) : a prob. space. $Y \in \mathcal{F}$ and $E|Y| < \infty$. $\mathcal{G} \subset \mathcal{F}$ is a sub σ -field. If X is a random variable satisfying (a) $X \in \mathcal{G}$ (b) $\int_A Y dP = \int_A X dP \quad \forall A \in \mathcal{G}$ then
 - X is integrable
 - Such X is unique in the sense that if there is another X' then $X = X'$ a.s.
 - ★ Trick : For any r.v. Z , $(Z > 0) = \bigcup_{\epsilon > 0} (Z \geq \epsilon) = \bigcup_{n \in \mathbb{N}} (Z > \frac{1}{n})$
 - ★ Lemma : For any \mathcal{F} -measurable and integrable X and Y , if $\int_A X dP = \int_A Y dP \quad \forall A \in \mathcal{F}$ then $X = Y$ a.s.

- Radon-Nikodym Thm

- If μ, ν are σ -finite measures on (Ω, \mathcal{F}) and $\nu \ll \mu$ ($\mu(A) = 0 \Rightarrow \nu(A) = 0 \quad \forall A \in \mathcal{F}$) then \exists a \mathcal{F} -measurable nonnegative function g s.t. $\nu(A) = \int_A g d\mu \quad \forall A \in \mathcal{F}$. The function g is unique in the sense that if h is another such function then $g = h$ μ -a.e.

- * Definition of conditional expectation

- $(\Omega, \mathcal{F}_0, P)$: a prob. space. $\mathcal{F} \subset \mathcal{F}_0$: a sub σ -field.
 X is a random variable s.t. $X \geq 0$, $X \in \mathcal{F}_0$ and $E|X| < \infty$. Then \exists a unique r.v. Y s.t. $Y \geq 0$, $Y \in \mathcal{F}$ and $\int_A X dP = \int_A Y dP \quad \forall A \in \mathcal{F}$. Such Y is unique in the sense that if another Y' exists then $Y = Y'$ a.s.
- $Y = E[X|\mathcal{F}]$ is said to be conditional expectation of X given \mathcal{F}
 - ★ Applying Radon Nikodym thm to measures $P|_{\mathcal{F}}$ and Q on (Ω, \mathcal{F}) where Q is defined by $Q(A) = \int_A X dP \quad \forall A \in \mathcal{F}$. Note that $Q \ll P|_{\mathcal{F}}$ and Q is a finite measure.
- We can extend the definition to general integrable r.v. X
 $Y = E[X|\mathcal{F}]$ is a unique random variable s.t. $Y \in \mathcal{F}$ and $\int_A X dP = \int_A Y dP \quad \forall A \in \mathcal{F}$.
 $E[X|\mathcal{F}]$ is also integrable and the uniqueness is in the sense of a.s. equivalence relation.
 $Y = E[X|\mathcal{F}]$ can be derived by $Y = Y_1 - Y_2$ where $Y_1 = E[X^+|\mathcal{F}]$ and $Y_2 = E[X^-|\mathcal{F}]$

* Conditional expectation given a random variable

– X : integrable r.v. For a random variable Y , define $E[X|Y] := E[X|\sigma(Y)]$

✓ Y need not be integrable.

✓ Since $E[X|Y] \in \sigma(Y)$, $E[X|Y] = g(Y)$ for some Borel function g . This coincides with the definition of conditional expectation in \mathcal{L}^2 space.

* Conditional probability

– For $A \in \mathcal{F}_0$ and a sub σ -field $\mathcal{F} \subset \mathcal{F}_0$, we define $P(A|\mathcal{F}) := E[I_A|\mathcal{F}]$

• Elementary properties of conditional expectation

– $(\Omega, \mathcal{F}_0, P)$: a prob. space. $\mathcal{F} \subset \mathcal{F}_0$: a sub σ -field. X, Y : integrable random variables

i. $E[c|\mathcal{F}] = c$

ii. $E[\psi(X)|\mathcal{F}] = \psi(X)$ given $E|\psi(X)| < \infty$

iii. If \mathcal{F} is a trivial σ -field i.e. $\mathcal{F} = \{\Omega, \emptyset\}$ then $E[X|\mathcal{F}] = E[X]$

iv. $\Omega = \bigcup_{i=1}^{\infty} \Omega_i$ is a partition of Ω with $\Omega_i \in \mathcal{F}_0$ and $P(\Omega_i) > 0 \quad \forall i \in \mathbb{N}$
 $\mathcal{F} = \sigma\{\Omega_1, \Omega_2, \dots\} = \{\bigcup_{j \in \kappa} \Omega_j : \kappa \subset \mathbb{N}\} \quad (\mathcal{F} \text{ is a } \sigma\text{-field}).$ Then we have

$$E[X|\mathcal{F}] = \sum_{i=1}^{\infty} a_i I_{\Omega_i} \quad \text{with} \quad a_i = \frac{E[X I_{\Omega_i}]}{P(\Omega_i)}$$

★ Lemma : If $Z \in \mathcal{F}$ for such \mathcal{F} , then we can write $Z = \sum_{i=1}^{\infty} c_i I_{\Omega_i}$ where $c_i \in \mathbb{R}$

v. $E[aX + bY|\mathcal{F}] = aE[X|\mathcal{F}] + bE[Y|\mathcal{F}] \quad \forall a, b \in \mathbb{R}$

vi. $X \geq 0 \Rightarrow E[X|\mathcal{F}] \geq 0 \quad a.s.$

★ Lemma : If $Z > 0$ on A with $P(A) > 0$ then $\int_A Z dP > 0$

vii. $X \leq Y \Rightarrow E[X|\mathcal{F}] \leq E[Y|\mathcal{F}] \quad a.s.$

viii. $|E[X|\mathcal{F}]| \leq E[|X||\mathcal{F}]$

• X, Y : integrable r.v's where $X \perp\!\!\!\perp Y$. $\psi : \mathbb{R}^2 \rightarrow \mathbb{R}$ Borel measurable s.t. $E|\psi(X, Y)| < \infty$
 Define $g : \mathbb{R} \rightarrow \mathbb{R}$ by $g(x) = E[\psi(x, Y)] \quad \forall x \in \mathbb{R}$. Then $E[\psi(X, Y)|\mathcal{F}] = g(X)$

✓ $g(x) = E[\psi(x, Y)] = \int \psi(x, Y) dP = \int_{\mathbb{R}} \psi(x, y) dP Y^{-1}(y) = \int_{\mathbb{R}} \psi(x, y) d\mu_Y(y) \quad \forall x \in \mathbb{R}$
 By Fubini thm in real analysis course, it is shown that g is Borel measurable & integrable.

• Conditional expectation and convergence

– $(\Omega, \mathcal{F}_0, P)$: a probability space. $\mathcal{F} \subset \mathcal{F}_0$: a sub σ -field

i. (MCT) If $X_n \geq 0$ and $X_n \nearrow X$ a.s. with $E(X) < \infty$ then $E[X_n|\mathcal{F}] \nearrow E[X|\mathcal{F}]$ a.s.

ii. (DCT) If $|X_n| \leq Y$, $E(Y) < \infty$ and $X_n \rightarrow X$ a.s. then $E[X_n|\mathcal{F}] \rightarrow E[X|\mathcal{F}]$ a.s.

★ Lemma : For integrable r.v. Z , we have $E[E(Z|\mathcal{F})] = E[Z]$

iii. (Continuity from below) $\{B_n\} \subset \mathcal{F}_0$ s.t. $B_n \subset B_{n+1} \quad \forall n \in \mathbb{N}$. $B := \bigcup_n B_n$

Then $P(B_n|\mathcal{F}) \nearrow P(B|\mathcal{F})$

iv. (Countable additivity) If $\{C_n\} \subset \mathcal{F}_0$ is mutually disjoint then $P(\bigcup_n C_n|\mathcal{F}) = \sum_n P(C_n|\mathcal{F})$

• Essential inequalities

i. (Markov) $P(|X| \geq c|\mathcal{F}) \leq \frac{1}{c} E[|X||\mathcal{F}] \quad \forall c > 0$

ii. (Jensen) If $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is convex then $\phi(E[X|\mathcal{F}]) \leq E[\phi(X)|\mathcal{F}]$ a.s.

★ Trick : For each $x \in \mathbb{R}$ and convex function $\phi : \mathbb{R} \rightarrow \mathbb{R}$, we have

$$\phi(x) = \sup\{ax + b : (a, b) \in S\} \text{ where } S = \{(a, b) \in \mathbb{R}^2 : ax + b \leq \phi(x) \forall x \in \mathbb{R}\}$$

iii. (Cauchy-Schwarz) For $X, Y \in \mathcal{L}^2$, we have $E^2[XY|\mathcal{F}] \leq E[X^2|\mathcal{F}]E[Y^2|\mathcal{F}]$ a.s.

- Smoothing property of conditional expectation

i. If $X \in \mathcal{F}$, $E|Y| < \infty$, and $E|XY| < \infty$ then $E[XY|\mathcal{F}] = XE[Y|\mathcal{F}]$ a.s.

✓ $E|X| < \infty$ assumption is not required.

□ If $X \in \mathcal{F}$ and $E|X| < \infty$ then $E[X|\mathcal{F}] = X$ a.s.

ii. If $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_0$ are sub σ -fields and $E|X| < \infty$ then

(a) $E[E(X|\mathcal{F}_1)|\mathcal{F}_2] = E[X|\mathcal{F}_1]$

(b) $E[E(X|\mathcal{F}_2)|\mathcal{F}_1] = E[X|\mathcal{F}_1]$

★ Lemma : If $\mathcal{F}_1 \subset \mathcal{F}_2$ then $Y \in \mathcal{F}_1 \Rightarrow Y \in \mathcal{F}_2$

✓ In short, “the smaller wins”. In view of information, it is similar to projection onto vector subspaces $S_1 \subset S_2 \subset S$ where $Proj_{S_1} Proj_{S_2} = Proj_{S_2} Proj_{S_1} = Proj_{S_1}$

- Def. of conditional expectation by Radon-Nikodym derivative agrees with def. in \mathcal{L}^2 space.

– If $E(X^2) < \infty$ then for $\mathcal{C} = \{Y : Y \in \mathcal{F}, E(Y^2) < \infty\}$,

$$E[\{X - E[X|\mathcal{F}]\}^2] = \inf_{Y \in \mathcal{C}} E[\{X - Y\}^2] \text{ and } E[X|\mathcal{F}] = \arg \min_{Y \in \mathcal{C}} E[\{X - Y\}^2]$$

★ Lemma : If $X \in \mathcal{L}^2$ then $E[X|\mathcal{F}] \in \mathcal{L}^2$

- * Independence of a random variable and a σ -field

– A random variable X and a σ -field \mathcal{F} are said to be independent if $\sigma(X)$ and \mathcal{F} are independent

- If an integrable random variable X and a σ -field \mathcal{F} are independent then $E[X|\mathcal{F}] = E[X]$

- * Conditional variance

$$Var(X|\mathcal{F}) := E[\{X - E[X|\mathcal{F}]\}^2|\mathcal{F}] = E[X^2|\mathcal{F}] - E^2[X|\mathcal{F}]$$

Conditional variance is defined for $X \in \mathcal{L}^2$