Conditional Expectation

김지수 (Jisu KIM)

확률론 2 (Probability Theory 2), 2025 2nd semester (fall)

The lecture note is largely from the lecture notes from Prof. Alessandro Rianldo's "Advanced Probability Overview".

Definition. Suppose a probability space $(\Omega, \mathcal{F}_0, P)$, and a σ-field $\mathcal{F} \subset \mathcal{F}_0$ is given. Let $X : (\Omega, \mathcal{F}_0) \to (\mathbb{R}, \mathcal{R})$ be a random variable that is $\mathcal{F}_0$ measurable with $\mathbb{E}\,|X| < \infty$. The conditional expectation of X given $\mathcal{F}$, $\mathbb{E}(X|\mathcal{F})$, is a random variable Y such that

(i) $Y \in \mathcal{F}$, i.e., is $\mathcal{F}$ measurable.

(ii) for all $A \in \mathcal{F}$, $\int_A X dP = \int_A Y dP$.

Any Y satisfying (i) and (ii) is said to be a version of $\mathbb{E}(X|\mathcal{F})$.

Lemma ([1, Lemma 4.1.1])**.** *If Y satisfies (i) and (ii), then it is integrable.*

Proof. Letting $A = \{Y > 0\} \in \mathcal{F}$, using (ii) twice, and then adding,

$$\int_A Y\, dP = \int_A X\, dP \leq \int_A |X|\, dP,$$

$$\int_{A^c} -Y\, dP = \int_{A^c} -X\, dP \leq \int_{A^c} |X|\, dP.$$

So we have $\mathbb{E}|Y| \leq \mathbb{E}|X|$. □

Uniqueness.

If Y' also satisfies (i) and (ii), then

$$\int_A Y\, dP = \int_A Y'\, dP \quad \text{for all } A \in \mathcal{F}.$$

Taking $A = \{Y - Y' \geq \varepsilon > 0\}$, we see

$$0 = \int_A X - X\, dP = \int_A Y - Y'\, dP \geq \varepsilon P(A) \quad \Rightarrow \quad P(A) = 0.$$

Since this holds for all ε, we have $Y \leq Y'$ a.s., and interchanging the roles of Y and Y', we have $Y = Y'$ a.s. Technically, all equalities such as $Y = E(X|\mathcal{F})$ should be written as $Y = E(X|\mathcal{F})$ a.s., but we have ignored this point in previous chapters and will continue to do so.

Repeating the last argument gives:

Theorem ([1, Theorem 4.1.2])**.** *If $X_1 = X_2$ on $B \in \mathcal{F}$, then $E(X_1|\mathcal{F}) = E(X_2|\mathcal{F})$ a.s. on B.*

Proof. Let $Y_1 = E(X_1|\mathcal{F})$ and $Y_2 = E(X_2|\mathcal{F})$. Taking $A = \{Y_1 - Y_2 \geq \varepsilon > 0\}$, we see

$$0 = \int_{A\cap B} (X_1 - X_2)\, dP = \int_{A\cap B} (Y_1 - Y_2)\, dP \geq \varepsilon P(A \cap B) \quad \Rightarrow \quad P(A \cap B) = 0.$$

$\square$

Below is a simple property that extends from expectations to conditional expectations. It will be used to prove the existence of conditional expectations.

Theorem ([1, Theorem 4.1.9 (b)])**.** *Monotonicity. If $X \leq Y$, then*

$$\mathbb{E}(X|\mathcal{F}) \leq \mathbb{E}(Y|\mathcal{F}).$$

Proof. Suppose that both $\mathbb{E}(X|\mathcal{C})$ and $\mathbb{E}(Y|\mathcal{C})$ exist. Let

$$C_0 = \{\infty > \mathbb{E}(X|\mathcal{C}) > \mathbb{E}(Y|\mathcal{C})\}, \quad C_1 = \{\infty = \mathbb{E}(X|\mathcal{C}) > \mathbb{E}(Y|\mathcal{C})\}.$$

Then, for $i = 0, 1$,

$$0 \leq \int_{C_i} \big(\mathbb{E}(X|\mathcal{C}) - \mathbb{E}(Y|\mathcal{C})\big)\, dP = \int_{C_i} (X - Y)\, dP \leq 0.$$

It follows that all terms in this string are 0 and $P(C_i) = 0$ for $i = 0, 1$. Since $C_0 \cup C_1 = \{\mathbb{E}(X|\mathcal{C}) > \mathbb{E}(Y|\mathcal{C})\}$, the result is proven. $\square$

Existence of Conditional Expectation

We could prove that versions of conditional expectations exist by the Radon–Nikodym theorem. However, the "modern" way to prove the existence of conditional expectations is through the theory of Hilbert spaces.

Definition. An inner product space is a vector space V with an inner product $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$, a function that satisfies

- **symmetry:** $\langle u, v \rangle = \langle v, u \rangle$,
- **bilinearity (part 1):** $\langle u_1 + u_2, v \rangle = \langle u_1, v \rangle + \langle u_2, v \rangle$,
- **bilinearity (part 2):** for real λ, $\langle \lambda u, v \rangle = \lambda \langle u, v \rangle$,
- **positivity:** $\langle u, u \rangle > 0$ for all $u \neq 0$, and $\langle u, u \rangle = 0$ iff $u = 0$.

An inner product provides a norm, namely $\|u\| = \sqrt{\langle u, u \rangle}$ and a metric $d(u, v) = \|u - v\|$. These facts follow from some simple properties of inner products.

Proposition. *Let V be a vector space with an inner product $\langle \cdot, \cdot \rangle$. Then*

1. ***Parallelogram law:*** *for all $u, v \in V$,*

$$\|u\|^2 + \|v\|^2 = \tfrac{1}{2}\big(\|u + v\|^2 + \|u - v\|^2\big).$$

2. ***Cauchy–Schwarz inequality:*** *for all $u, v \in V$,*

$$|\langle u, v \rangle| \leq \|u\| \|v\|,$$

with equality if and only if u and v are collinear.

3. ***Triangle inequality:*** *for all $u, v \in V$,*

$$\|u+v\| \leq \|u\| + \|v\|.$$

Definition. A complete inner product space is a Hilbert space.

Example. Let $V = L^2(\Omega, \mathcal{F}, \mu)$. Define $\langle f, g\rangle = \int fg\, d\mu$. This is an inner product that produces the norm $\|\cdot\|_2$. Note that L^2 is complete.

We prove existence of conditional expectations using orthogonal projection in Hilbert spaces. The following theorem is a basic result in Hilbert space theory

Theorem (Hilbert projection theorem)**.** *Let $\mathcal{H}$ be a Hilbert space and let $C \subset \mathcal{H}$ be a closed vector space of $\mathcal{H}$. For every $x \in \mathcal{H}$, there exists a unique $m \in C$ such that*

$$\|x - m\| = \inf_{c \in C} \|x - c\|\,.$$

Further, $x - m$ is orthogonal to C.

Now, we can prove the existence of conditional expectations.

Theorem (Existence of conditional expectation)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let Y be a random variable. Let $\mathcal{C}$ be a sub-σ-field of $\mathcal{F}$. If $\mathbb{E}(Y)$ exists, then there exists a version of $\mathbb{E}(Y|\mathcal{C})$.*

Proof. It is easy to see that $L^2(\Omega, \mathcal{C}, P)$ is a closed linear subspace. If $Y \in L^2(\Omega, \mathcal{F}, P)$, let Y_0 be the projection of Y into $L^2(\Omega, \mathcal{C}, P)$. By Hilbert projection theorem,

$$\mathbb{E}([Y - Y_0]X) = 0 \quad \text{for all } X \in L^2(\Omega, \mathcal{C}, P),$$

in particular for $X = I_C$ for arbitrary $C \in \mathcal{C}$.

If $Y > 0$ but not in L^2, define $Y_n = \min\{Y, n\}$. Then $Y_n \in L^2$. Let $Y_{0,n}$ be a version of $\mathbb{E}(Y_n|\mathcal{C})$, and assume that $Y_{0,n} \leq Y_{0,n+1}$ for all n, which is allowed by monotonicity of conditional expectation. Let $Y_0 = \lim_{n\to\infty} Y_{0,n}$. For each $C \in \mathcal{C}$,

$$\lim_{n\to\infty} \mathbb{E}(I_C Y_n) = \mathbb{E}(I_C Y), \quad \lim_{n\to\infty} \mathbb{E}(I_C Y_{0,n}) = \mathbb{E}(I_C Y_0),$$

and

$$\mathbb{E}(I_C Y_n) = \mathbb{E}(I_C Y_{0,n}), \quad \forall n.$$

It follows that $\mathbb{E}(I_C Y) = \mathbb{E}(I_C Y_0)$ for all $C \in \mathcal{C}$ and Y_0 is a version of $\mathbb{E}(Y|\mathcal{C})$.

If Y takes both positive and negative values, write $Y = Y^+ - Y^-$. If one of the means $\mathbb{E}(Y^+)$ or $\mathbb{E}(Y^-)$ is finite then the probability is 0 that both $\mathbb{E}(Y^+|\mathcal{C}) = \infty$ and $\mathbb{E}(Y^-|\mathcal{C}) = \infty$. Then $\mathbb{E}(Y^+|\mathcal{C}) - \mathbb{E}(Y^-|\mathcal{C})$ is a version of $\mathbb{E}(Y|\mathcal{C})$. $\square$

The following result summarizes what we have learned about the existence and uniqueness of conditional expectation.

Corollary. *If $Y \in L^2(\Omega, \mathcal{F}, P)$ and $\mathcal{C}$ is a sub-σ-field of $\mathcal{F}$. Let $Z \in L^2(\Omega, \mathcal{C}, P)$. Then the following are equivalent.*

1. *$Z = \mathbb{E}(Y|\mathcal{C})$.*

2. *$\mathbb{E}(XZ) = \mathbb{E}(XY)$ for all $X \in L^2(\Omega, \mathcal{C}, P)$.*

3. *Z is the orthogonal projection of Y into $L^2(\Omega, \mathcal{C}, P)$.*

Examples

Intuitively, we think of $\mathcal{F}$ as describing the information we have at our disposal: for each $A \in \mathcal{F}$, we know whether or not A has occurred. $E(X|\mathcal{F})$ is then our "best guess" of the value of X given the information we have.

Example ([1, Example 4.1.3])**.** If $X \in \mathcal{F}$, $\mathbb{E}(X|\mathcal{F}) = X$; i.e., if we know X then our "best guess" is X itself. A special case of this example is $X = c$, where c is a constant.

Example ([1, Example 4.1.4])**.** If X is independent of $\mathcal{F}$, then $\mathbb{E}(X|\mathcal{F}) = \mathbb{E}(X)$. This is at the other extreme of having no information; if you don't know anything about X, then the best guess is the mean $\mathbb{E}X$. To check the definition, note that $EX \in \mathcal{F}$ so (i) holds. To verify (ii), we observe that if $A \in \mathcal{F}$, then since X and $\mathbf{1}_A \in \mathcal{F}$ are independent, [1, Theorem 2.1.3] implies

$$\int_A X\,dP = E(X\mathbf{1}_A) = EX \cdot E(\mathbf{1}_A) = \int_A EX\,dP.$$

Example ([1, Example 4.1.5])**.** Let $\Omega_1, \Omega_2, \cdots$ be a countable partition of Ω into disjoint sets, and let $\mathcal{F} = \sigma(\Omega_1, \Omega_2, \cdots)$ be the σ-field generated by these sets. Then

$$E(X|\mathcal{F})(\omega) = \sum_{i=1}^{\infty} c_i I(\omega \in \Omega_i),$$

where

$$c_i = \begin{cases} \frac{\mathbb{E}(X;\Omega_i)}{P(\Omega_i)}, & if\ P(\Omega_i) > 0, \\ arbitrary, & if\ P(\Omega_i) = 0. \end{cases}$$

In words, the information in Ω_i tells us which element of the partition our outcome lies in and given this information, the best guess for X is the average value of X over Ω_i.

Definition. Conditional expectation given random variable is a special case of the conditional expectation given a σ-field, being defined as

$$E(X|Y) = E(X|\sigma(Y)),$$

where $\sigma(Y)$ is the σ-field generated by Y.

Definition. Conditional expectation given $Y = y$, i.e. $E(X|Y = y)$.

Consider $E(X|Y)$, which is $\sigma(Y)$-measurable. Then there exists a measurable function $h : \mathbb{R} \to \mathbb{R}$ s.t. $\mathbb{E}(X|Y) = h(Y)$ (Exercise 1.3.8) Now, we can define

$$\mathbb{E}(X|Y = y) = h(y)$$

Example ([1, Example 4.1.6])**.** Suppose X and Y have joint density $f(x, y)$, and suppose for simplicity that $\int f(x, y)\,dx > 0$ for all y. If $E|g(X)| < \infty$, then

$$E(g(X)|Y) = h(Y), \quad \text{where } h(y) = \frac{\int g(x)f(x, y)\,dx}{\int f(x, y)\,dx}.$$

To "guess" this formula, note that treating the probability densities $P(Y = y)$ as if they were real probabilities,

$$P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{f(x, y)}{\int f(x, y)\,dx},$$

so, integrating against the conditional probability density, we have

$$E(g(X)|Y = y) = \int g(x)P(X = x|Y = y)\,dx.$$

To “verify” the proposed formula now, observe $h(Y) \in \sigma(Y)$ so (i) holds. To check (ii), observe that if $A \in \sigma(Y)$ then $A = \{\omega : Y(\omega) \in B\}$ for some $B \subseteq \mathbb{R}$, so

$$E(h(Y); A) = \int_B h(y)f(y)\,dy = \int_B \left(\int g(x)f(x,y)\,dx \right) dy = E(g(X) \cdot \mathbf{1}_B(Y)) = E(g(X); A).$$

Remark. To drop the assumption that $\int f(x,y)\,dx > 0$, define h by

$$h(y) \int f(x,y)\,dx = \int g(x)f(x,y)\,dx$$

(i.e., h can be anything where the denominator is 0), and observe this is enough for the proof.

Example ([1, Example 4.1.7])**.** Suppose X and Y are independent. Let φ be a function wit $\mathbb{E}|\varphi(X,Y)| < \infty$ and let $g(x) = \mathbb{E}(\varphi(x,Y))$. Then

$$\mathbb{E}(\varphi(X,Y)|X) = g(X).$$

Proof. It is clear that $g(X) \in \sigma(X)$. To check (ii), note that if $A \in \sigma(X)$ then $A = \{X \in C\}$, so using Fubini's theorem and the independence of X and Y:

$$\int_A \varphi(X,Y)\,dP = E[\varphi(X,Y) \cdot \mathbf{1}_C(X)] = \int_C E[\varphi(x,Y)]\,P_X(dx) = \int_A g(X)\,dP.$$

□

Properties

Conditional expectation has many of the same properties that ordinary expectation does.

Theorem ([1, Theorem 4.1.9])**.** *(a) Linearlity.*

$$\mathbb{E}(aX + Y|\mathcal{F}) = a\mathbb{E}(X|\mathcal{F}) + \mathbb{E}(Y|\mathcal{F}).$$

(c) Monotone convergence theorem.
If $X_n \geq 0$ and $X_n \uparrow X$ with $E|X| < \infty$, then

$$\mathbb{E}(X_n|\mathcal{F}) \uparrow \mathbb{E}(X|\mathcal{F}).$$

Remark. By applying the last result to $Y_1 - Y_n$, we see that if $Y_n \downarrow Y$ and $E|Y_1|, E|Y| < \infty$, then $E(Y_n|\mathcal{F}) \downarrow E(Y|\mathcal{F})$.

Proof. To prove (a), we need to check that the right-hand side is a version of the left. It clearly is $\mathcal{F}$-measurable. To check (ii), we observe that if $A \in \mathcal{F}$ then by linearity of the integral and the defining properties of $E(X|\mathcal{F})$ and $E(Y|\mathcal{F})$,

$$\int_A \{aE(X|\mathcal{F}) + E(Y|\mathcal{F})\}dP = a\int_A E(X|\mathcal{F})dP + \int_A E(Y|\mathcal{F})dP = a\int_A X\,dP + \int_A Y\,dP = \int_A (aX + Y)\,dP.$$

For (c): Let $Y_n = X - X_n$. It suffices to show that $E(Y_n|\mathcal{F}) \downarrow 0$. Since $Y_n \downarrow 0$, part (b) implies $Z_n := E(Y_n|\mathcal{F}) \downarrow Z_\infty$. If $A \in \mathcal{F}$, then

$$\int_A Z_n\,dP = \int_A Y_n\,dP.$$

Letting $n \to \infty$, and noting $Y_n \downarrow 0$, the dominated convergence theorem gives:

$$\int_A Z_\infty \, dP = 0 \quad \text{for all } A \in \mathcal{F}, \Rightarrow Z_\infty \equiv 0.$$

□

Theorem ([1, Theorem 4.1.10])**.** *Jensen Inequality*

If φ is convex and $E|X| < \infty$ and $E|\varphi(X)| < \infty$, then

$$\varphi(E(X|\mathcal{F})) \leq E(\varphi(X)|\mathcal{F}).$$

Proof. If φ is linear, the result is trivial, so we will suppose φ is not linear. We do this so that if we let

$$S = \{(a, b) : a, b \in \mathbb{Q}, ax + b \leq \varphi(x) \text{ for all } x\},$$

then $\varphi(x) = \sup\{ax + b : (a, b) \in S\}$ (see proof of [1, Theorem 1.6.2] for details). If $\varphi(x) \geq ax + b$, then using [1, Theorem 4.1.9] parts (a) and (b):

$$E(\varphi(X)|\mathcal{F}) \geq aE(X|\mathcal{F}) + b \quad \text{a.s.}$$

Taking the supremum over $(a, b) \in S$ gives:

$$E(\varphi(X)|\mathcal{F}) \geq \varphi(E(X|\mathcal{F})) \quad \text{a.s.}$$

□

Remark. Here we have written a.s. in the inequalities to stress that there is an exceptional set for each a, b, so we must take the supremum over a countable set.

Theorem ([1, Theorem 4.1.11])**.** *Conditional expectation is a contraction in L^p, $p \geq 1$, i.e., i.e.,*

$$\mathbb{E}(|\mathbb{E}(X|\mathcal{F})|^p) \leq \mathbb{E}|X|^p.$$

Proof. [1, Theorem 4.1.10] implies $|E(X|\mathcal{F})|^p \leq E(|X|^p|\mathcal{F})$. Taking expectations:

$$E(|E(X|\mathcal{F})|^p) \leq E(E(|X|^p|\mathcal{F})) = E|X|^p.$$

□

In the last equality, we have used an identity that is an immediate consequence of the definition (use property (ii) with $A = \Omega$):

$$E(E(Y|\mathcal{F})) = E(Y).$$

Conditional expectation also has properties like this that have no analogue for "ordinary" expectation.

Theorem ([1, Theorem 4.1.12])**.** *If $\mathcal{F} \subset \mathcal{G}$ and $E(X|\mathcal{G}) \in \mathcal{F}$, then*

$$\mathbb{E}(X|\mathcal{F}) = \mathbb{E}(X|\mathcal{G}).$$

Proof. By assumption $E(X|\mathcal{G}) \in \mathcal{F}$. To check the other part of the definition we note that if $A \in \mathcal{F} \subset \mathcal{G}$ then

$$\int_A X \, dP = \int_A E(X|\mathcal{G}) \, dP.$$

□

Theorem ([1, Theorem 4.1.13])**.** *If $\mathcal{F}_1 \subset \mathcal{F}_2$, then*
(i) $\mathbb{E}(\mathbb{E}(X|\mathcal{F}_1)|\mathcal{F}_2) = \mathbb{E}(X|\mathcal{F}_1)$
(ii) $\mathbb{E}(\mathbb{E}(X|\mathcal{F}_2)|\mathcal{F}_1) = \mathbb{E}(X|\mathcal{F}_1)$

In words, the smaller σ-field always wins.

Proof. For (i): Since $E(X|\mathcal{F}_1) \in \mathcal{F}_1 \subset \mathcal{F}_2$, this follows from [1, Example 4.1.3].

For (ii): Notice $E(X|\mathcal{F}_1) \in \mathcal{F}_1$, and if $A \in \mathcal{F}_1 \subset \mathcal{F}_2$,

$$\int_A E(X|\mathcal{F}_1)\,dP = \int_A X\,dP = \int_A E(X|\mathcal{F}_2)\,dP.$$

Hence $E(E(X|\mathcal{F}_2)|\mathcal{F}_1) = E(X|\mathcal{F}_1)$ a.s. □

The next result shows that for conditional expectation with respect to $\mathcal{F}$, random variables $X \in \mathcal{F}$ are like constants — they can be brought outside the "integral."

Theorem ([1, Theorem 4.1.14])**.** *If $X \in \mathcal{F}$ and $E|X| < \infty$, then*

$$\mathbb{E}(XY|\mathcal{F}) = X\mathbb{E}(Y|\mathcal{F}).$$

Proof. The right-hand side is $\mathcal{F}$-measurable, so it remains to verify property (ii). Suppose $X = \mathbf{1}_B$ with $B \in \mathcal{F}$. Then for $A \in \mathcal{F}$,

$$\int_A \mathbf{1}_B E(Y|\mathcal{F})\,dP = \int_{A\cap B} E(Y|\mathcal{F})\,dP = \int_{A\cap B} Y\,dP = \int_A \mathbf{1}_B Y\,dP.$$

This extends to simple X by linearity. If $X \geq 0$, let $X_n \uparrow X$ by simple functions and apply the monotone convergence theorem to conclude:

$$\int_A XE(Y|\mathcal{F})\,dP = \int_A XY\,dP.$$

Split X and Y into positive and negative parts for the general case. □

Theorem ([1, Theorem 4.1.15])**.** *Suppose $EX^2 < \infty$, then $\mathbb{E}(X|\mathcal{F})$ is a random variable $Y \in \mathcal{F}$ that minimizes $\mathbb{E}(X-Y)^2$ among all random variables$\in \mathcal{F}$.*

Remark. This result gives a geometric interpretation of $E(X|\mathcal{F})$. Let

$$L^2(\mathcal{F}_0) = \{Y \in \mathcal{F}_0 : EY^2 < \infty\}$$

be a Hilbert space, and let $L^2(\mathcal{F})$ be the closed subspace. Then $E(X|\mathcal{F})$ is the orthogonal projection of X onto $L^2(\mathcal{F})$, i.e., the point in the subspace closest to X.

Remark. Conditional operation generator

Let $L^1(\Omega, \mathcal{F}, P) = \{X \in \mathcal{F}|\ E|X| < \infty\}$. We can think of $E(\cdot|\mathcal{G})$ as a mapping $L^1(\Omega,\mathcal{F},P) \to L^1(\Omega,\mathcal{F},P)$

Consider $L^2(\Omega,\mathcal{F},P) = \{X \in \mathcal{F}|\ EX^2 < \infty\}$ Then $L^2(\Omega,\mathcal{F},P) \subset L^1(\Omega,\mathcal{F},P)$.

Now, when $E(\cdot|\mathcal{G})$ is viewed as $E(\cdot|\mathcal{G}) : L^2(\Omega,\mathcal{F},P) \to L^2(\Omega,\mathcal{F},P)$, then $E(\cdot|\mathcal{G})$ has the following properties:

- $E(\cdot|\mathcal{G})$ is a linear operator.

- $E(\cdot|\mathcal{G})$ is an order-preserving operator (i.e. $X \leq Y \implies E(X|\mathcal{G}) \leq E(Y|\mathcal{G})$).
- $E(\cdot|\mathcal{G})$ is a contraction operator (i.e. $E(E(X|\mathcal{G}))^2 \leq EX^2$).
- $L^2(\Omega, \mathcal{F}, P)$ is a Hilbert space with $\langle X, Y\rangle = E(XY)$, then above implies $\|E(X|\mathcal{G})\| \leq \|X\|$
- $E(\cdot|\mathcal{G})$ is an idempotent operator (i.e. $E(E(X|\mathcal{G})|\mathcal{G}) = E(X|\mathcal{G})$)
- $E(\cdot|\mathcal{G})$ is a symmetric operator (i.e. $\langle E(X|\mathcal{G}), Y\rangle = \langle X, E(Y|\mathcal{G})\rangle$)

Conclusively, $E(\cdot|\mathcal{G})$ is an orthogonal projection operator (since it is linear, idempotent and symmetric). This can be illustrated as $E(X - E(X|\mathcal{G}))^2 \leq E(X-Y)^2$ for $\forall Y \in L^2(\Omega, \mathcal{G}, P)$

Conditional Distribution

Now we introduce the measure-theoretic version of conditional probability and distribution.

Conditional probability and Regular conditional probability

For $A \in \mathcal{F}$, define

$$\Pr(A|\mathcal{C}) = \mathbb{E}(I_A|\mathcal{C}).$$

That is, treat I_A as a random variable X and define the conditional probability of A to be the conditional mean of X. We would like to show that conditional probabilities behave like probabilities. The first thing we can show is that they are additive. That is a consequence of the following result.

It follows easily from [1, Theorem 4.1.9 (a)] that

$$\Pr(A|\mathcal{C}) + \Pr(B|\mathcal{C}) = \Pr(A \cup B|\mathcal{C}) \quad \text{a.s.}$$

if A and B are disjoint. The following additional properties are straightforward. They are similar to [1, Theorem 4.1.9 (a)].

Example (Probability at most 1)**.** We shall show that $\Pr(A|\mathcal{C}) \leq 1$ a.s. Let $B = \{\omega : \Pr(A|\mathcal{C}) > 1\}$. Then $B \in \mathcal{C}$, and

$$P(B) \leq \int_B \Pr(A|\mathcal{C})\, dP = \int_B I_A\, dP = P(A \cap B) \leq P(B),$$

where the first inequality is strict if $P(B) > 0$. Clearly, neither of the inequalities can be strict, hence $P(B) = 0$.

Example (Countable Additivity)**.** Let $\{A_n\}_{n=1}^{\infty}$ be disjoint elements of $\mathcal{F}$. Let

$$W = \sum_{n=1}^{\infty} \Pr(A_n|\mathcal{C}).$$

We shall show that W is a version of $\Pr\left(\bigcup_{n=1}^{\infty} A_n \,\middle|\, \mathcal{C}\right)$. Let $C \in \mathcal{C}$.

$$\mathbb{E}\big[I_C I_{\cup_{n=1}^{\infty} A_n}\big] = P\left(C \cap \bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(C \cap A_n) = \sum_{n=1}^{\infty} \int_C \Pr(A_n|\mathcal{C})\, dP$$

$$= \int_C \sum_{n=1}^{\infty} \Pr(A_n|\mathcal{C})\, dP = \int_C W\, dP,$$

where the sum and integral are interchangeable by the monotone convergence theorem.

We could also prove that $\Pr(A|\mathcal{C}) \geq 0$ a.s. and $\Pr(\Omega|\mathcal{C}) = 1$ a.s. But there are generally uncountably many different $A \in \mathcal{F}$ and uncountably many different sequences of disjoint events. Although countable additivity holds a.s. separately for each sequence of disjoint events, how can we be sure that it holds simultaneously for all sequences a.s.?

Definition (Regular Conditional Probabilities)**.** Let $\mathcal{A} \subseteq \mathcal{F}$ be a sub-σ-field. We say that a collection of versions $\{\Pr(A|\mathcal{C}) : A \in \mathcal{A}\}$ are *regular conditional probabilities* if, for a.e. $\omega \in \Omega$, $\Pr(\cdot|\mathcal{C})(\omega)$ is a probability measure on $(\Omega, \mathcal{A})$.

Rarely do regular conditional probabilities exist on $(\Omega, \mathcal{F})$, but there are lots of common sub-σ-fields $\mathcal{A}$ such that regular conditional probabilities exist on $(\Omega, \mathcal{A})$. Oddly enough, the existence of regular conditional probabilities doesn't seem to depend on $\mathcal{C}$.

Example. Continuation of [1, Example 4.1.6]. Set $\mathcal{C} = \sigma(Y)$. Under the same setup, for each $B \in \mathcal{R}$, let $A = X^{-1}(B)$, and define

$$h(y; A) = \frac{\int_B f(x,y)\, dx}{\int_{\mathbb{R}} f(x,y)\, dx}, \quad \text{for all } y.$$

Finally, define $\Pr(A|\mathcal{C})(\omega) = h(Y(\omega); A)$. Then the same calculation of [1, Example 4.1.6] shows that this is a version of the conditional mean of I_A given $\mathcal{C}$. And the dominated convergence theorem implies that $A \mapsto \Pr(A|\mathcal{C})(\omega)$ is a probability measure on $(\Omega, \sigma(X))$.

The results we have on existence of regular conditional probabilities are for the cases in which $\mathcal{A}$ is the σ-field generated by a random variable or something a lot like a random variable. Note that this is a condition on $\mathcal{A}$, not on $\mathcal{C}$. The conditioning σ-field can be anything at all. What matters is the σ-field on which the conditional probability is to be defined.

Conditional Distribution and Regular conditional distribution

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $X : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ measurable map, and $\mathcal{C} \subset \mathcal{F}$ a σ-field. If $\mathcal{A} = \sigma(X)$, conditional probabilities on $\mathcal{A}$ form a conditional distribution for X.

Definition. $\mu : \Omega \times \mathcal{S} \to [0,1]$ is a *regular conditional distribution* (r.c.d.) for X given $\mathcal{C}$ if

(i) For each A, $\omega \mapsto \mu(\omega, A)$ is a version of $Pr(X \in A|\mathcal{C})$

(ii) For a.e. ω, $A \mapsto \mu(\omega, A)$ is a probability measure on $(S, \mathcal{S})$

When $S = \Omega$ and X is the identity map, μ becomes a regular conditional probability $Pr(A|\mathcal{C})$ we defined before.

Example. Continuation of [1, Example 4.1.6].

Suppose X and Y have a joint density $f(x,y) > 0$. Define:

$$\mu(y,A) = \frac{\int_A f(x,y)\,dx}{\int f(x,y)\,dx}.$$

Then $\mu(Y(\omega),A)$ is a regular conditional distribution for X given $\sigma(Y)$.

Regular conditional distributions are useful because they allow us to simultaneously compute the conditional expectation of all functions of X and to generalize properties of ordinary expectation in a more straightforward way.

Theorem. *[1, Theorem 4.1.16] Let $\mu(\omega, A)$ be a regular conditional distribution for X given $\mathcal{F}$. If $f : (S, \mathcal{S}) \to (\mathbb{R}, \mathcal{B})$ has $E|f(X)| < \infty$, then*

$$E(f(X)|\mathcal{F}) = \int f(x)\,\mu(\omega, dx) \quad a.s.$$

Proof. If $f = \mathbf{1}_A$, this follows from the definition. Linearity extends the result to simple functions. Monotone convergence extends it to nonnegative f. The general result follows by writing $f = f^+ - f^-$. □

Unfortunately, regular conditional distributions do not always exist. The existence is guaranteed when $(S, \mathcal{S})$ is nice.

Theorem. *[1, Theorem 4.1.16] If $(S, \mathcal{S})$ is nice, i.e., if there is a 1-1 map φ from S into $\mathbb{R}$ so that φ and φ^{-1} are both measurable., then there exists a regular conditional distribution.*

Radon-Nikodym Derivatives and alternative for existence

Definition. Let μ, ν be measure. We say ν is said to be absolutely continuous with respect to μ (abbreviated $\nu \ll \mu$) if $\mu(A) = 0$ implies $\nu(A) = 0$.

Theorem. *[1, Theorem A.4.8] Radon-Nikodym Theorem. Let μ and ν be σ-finite measures on $(\Omega, \mathcal{F})$. If $\nu \ll \mu$, there is a function $f \in \mathcal{F}$ such that for all $A \in \mathcal{F}$,*

$$\int_A f d\mu = \nu(A).$$

If g is another such function, then $f = g$ a.e.$[\mu]$.

Remark. In fact Radon-Nikodym Theorem holds for σ-finite signed measures as well; see [1, Section A.4].

This theorem easily gives the existence of conditional expectation $\mathbb{E}(X|\mathcal{F})$. Suppose first that $X \geq 0$. Let $\mu = P$ and

$$\nu(A) = \int_A X\,dP, \quad \text{for } A \in \mathcal{F}.$$

The dominated convergence theorem implies ν is a measure, and the definition of the integral implies $\nu \ll \mu$. The Radon-Nikodym derivative $d\nu/d\mu \in \mathcal{F}$ and for any $A \in \mathcal{F}$ has

$$\int_A X\,dP = \nu(A) = \int_A \frac{d\nu}{d\mu}\,dP.$$

Taking $A = \Omega$, we see that $d\nu/d\mu \geq 0$ is integrable, and we have shown that $d\nu/d\mu$ is a version of $\mathbb{E}(X|\mathcal{F})$.

To treat the general case, write $X = X^+ - X^-$, let $Y_1 = \mathbb{E}(X^+|\mathcal{F})$ and $Y_2 = \mathbb{E}(X^-|\mathcal{F})$. Now $Y_1 - Y_2 \in \mathcal{F}$ is integrable, and for all $A \in \mathcal{F}$ we have

$$\int_A X\,dP = \int_A X^+\,dP - \int_A X^-\,dP = \int_A Y_1\,dP - \int_A Y_2\,dP = \int_A (Y_1 - Y_2)\,dP.$$

This shows $Y_1 - Y_2$ is a version of $\mathbb{E}(X|\mathcal{F})$ and completes the proof.

References

[1] Rick Durrett. *Probability—theory and examples*, volume 49 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2019. Fifth edition.