Unit 2: Review of Conditional Expectations

Instructor: Quan Zhou

Proofs are omitted for the results in this section; they can be found in probability textbooks such as [2, 3, 4].

2.1 Conditional expectations

Definition 2.1. Consider a probability space $(\Omega, \mathcal{F}, \mathsf{P})$, a sub-$\sigma$-field $\mathcal{G} \subset \mathcal{F}$, and a random variable X such that $\mathsf{E}|X| < \infty$. The conditional expectation of X given $\mathcal{G}$, denoted by $\mathsf{E}[X \,|\, \mathcal{G}]$, is a random variable such that

(i) $\mathsf{E}[X \,|\, \mathcal{G}]$ is $\mathcal{G}$-measurable;

(ii) for any $A \in \mathcal{G}$, we have $\int_A X \mathrm{d}\mathsf{P} = \int_A \mathsf{E}[X \,|\, \mathcal{G}] \mathrm{d}\mathsf{P}$.

Any random variable that satisfies the above two properties is called a version of $\mathsf{E}[X \,|\, \mathcal{G}]$. For two random variables X, Y defined on the same probability space, we often write $\mathsf{E}[X \,|\, Y] = \mathsf{E}[X \,|\, \sigma(Y)]$.

Theorem 2.1. *There exists a random variable that satisfies (i) and (ii) in Definition 2.1. Further, such a random variable is essentially unique, which means that any two versions of* $\mathsf{E}[X \,|\, \mathcal{G}]$ *are equivalent almost surely.*

Proof. See, e.g., [2]. □

Example 2.1. Consider a probability space $(\Omega, \mathcal{F}, \mathsf{P})$. Let $\Omega_1, \Omega_2, \dots$ be a countable partition of the entire sample space Ω ("partition" implies "disjoint") such that $\mathsf{P}(\Omega_i) > 0$ for each i. Define a sub-σ-algebra by

$$\mathcal{G} = \sigma(\Omega_1, \Omega_2, \dots).$$

Then, one can show that the conditional expectation of a random variable X given $\mathcal{G}$ is

$$\mathsf{E}[X \,|\, \mathcal{G}](\omega) = \sum_{i \geq 1} \frac{\int_{\Omega_i} X \mathrm{d}\mathsf{P}}{\mathsf{P}(\Omega_i)} \mathbb{1}_{\Omega_i}(\omega), \quad \text{a.s.}$$

Observe that equivalently this can be expressed as, almost surely,

$$\mathsf{E}[X \,|\, \mathcal{G}](\omega) = \frac{\mathsf{E}[X \mathbb{1}_{\Omega_i}]}{\mathsf{P}(\Omega_i)}, \quad \text{if } \omega \in \Omega_i.$$

This justifies why in elementary probability, we use the following formula to calculate the conditional expectation given any $A \in \mathcal{F}$,

$$\mathsf{E}[X \mid A] = \frac{\mathsf{E}[X \mathbb{1}_A]}{\mathsf{P}(A)}. \tag{1}$$

In (1), $\mathsf{E}[X \mid A]$ is a real number, not a random variable.

Remark 2.1. Let $Y \colon (\Omega, \mathcal{F}) \to (\Lambda, \mathcal{H})$. Consider a version of $E[X \mid \sigma(Y)]$, which by definition is a mapping from Ω to $\mathbb{R}$ and should be $\sigma(Y)$-measurable. This implies that there exists a function $h \colon (\Lambda, \mathcal{H}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\mathsf{E}[X \mid \sigma(Y)](\omega) = (h \circ Y)(\omega) = h(Y(\omega))$. In statistics, we often use the notation $\mathsf{E}[X \mid Y = y]$, which is defined by $E[X \mid Y = y] = h(y)$.

Remark 2.2. Consider $\mathbb{1}_{\{X \in A\}}$ for a random variable X and $A \in \mathcal{B}(\mathbb{R})$. Let Y be another random variable. From Remark 2.1, $\mathsf{P}(X \in A \mid Y = y) := \mathsf{E}[\mathbb{1}_{\{X \in A\}} \mid Y = y] = h(y)$ for some measurable function h. Further, it can be shown that, almost surely,

$$h(y) = \lim_{\delta \downarrow 0} \mathsf{P}(X \in A \mid Y \in (y - \delta, y + \delta]). \tag{2}$$

The right-hand side is evaluated by using elementary formula for conditional probabilities. This yields a natural interpretation of $\mathsf{P}(X \in A \mid Y = y)$. For the proof of (2), see [1].

Example 2.2. Let X, Y be independent standard normal random variables, and consider $\mathsf{P}(X \in A \mid X = Y)$. In light of Remark 2.2, we may want to interpret $\mathsf{P}(X \in A \mid X = Y)$ as the limit of $\mathsf{P}(X \in A \mid B_n)$ for some sequence of events $\{B_n\}_{n \geq 1}$ that converges to $\{X = Y\}$. This will be problematic, because the limit, even if it exists, largely depends on how we construct the sequence $\{B_n\}_{n \geq 1}$. For example, we can let $U = X - Y$ and $B_n^U = \{|U| < n^{-1}\}$; we can also let $V = X/Y$ and $B_n^V = \{|V - 1| < n^{-1}\}$. But $\lim_{n \to \infty} \mathsf{P}(X \in A \mid B_n^U)$ and $\lim_{n \to \infty} \mathsf{P}(X \in A \mid B_n^V)$ are unequal in general. (You can use the formula given in Proposition 2.5 to verify that the regular conditional distribution of $X \mid U = 0$ and $X \mid V = 1$ are actually different.) This is not too surprising upon observing that $\sigma(U) \neq \sigma(V)$. Whenever we do conditioning, we should think about the σ-algebra we are conditioning on. The two random variables $E[\mathbb{1}_{\{X \in A\}} \mid \sigma(U)]$ and $E[\mathbb{1}_{\{X \in A\}} \mid \sigma(V)]$ are very different. A similar example is given by the Borel-Kolmogorov paradox.

2.2 Properties of conditional expectation

For all results below, assume the probability space $(\Omega, \mathcal{F}, \mathsf{P})$ is given.

Proposition 2.1 (Basic properties of conditional expectation)**.** *Let X, Y be integrable random variables and $\mathcal{G} \subset \mathcal{F}$ be a given sub-σ-algebra.*

(i) For $a, b \in \mathbb{R}$, $\mathsf{E}[(aX + bY) \,|\, \mathcal{G}] = a\mathsf{E}[X \,|\, \mathcal{G}] + b\mathsf{E}[Y \,|\, \mathcal{G}]$, a.s.

(ii) If $X = c$ where $c \in \mathbb{R}$, then $\mathsf{E}[X \,|\, \mathcal{G}] = c$, a.s.

(iii) If $X \geq Y$, then $\mathsf{E}[X \,|\, \mathcal{G}] \geq \mathsf{E}[Y \,|\, \mathcal{G}]$, a.s.

(iv) If $X \in \mathcal{G}$, then $\mathsf{E}[X \,|\, \mathcal{G}] = X$, a.s.

(v) $\mathsf{E}[X \,|\, \{\emptyset, \Omega\}] = \mathsf{E}[X]$.

(vi) Law of total expectation: $\mathsf{E}[\,\mathsf{E}[X \,|\, \mathcal{G}]\,] = \mathsf{E}[X]$.

(vii) Tower property: If $\mathcal{H}$ is another σ-algebra such that $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$, then

$$\mathsf{E}[\,\mathsf{E}[X \,|\, \mathcal{G}] \,|\, \mathcal{H}] = \mathsf{E}[\,\mathsf{E}[X \,|\, \mathcal{H}] \,|\, \mathcal{G}] = \mathsf{E}[X \,|\, \mathcal{H}], \quad \text{a.s.}$$

(viii) Suppose $\mathsf{E}|XY| < \infty$ and $Y \in \mathcal{G}$. Then $\mathsf{E}[XY \,|\, \mathcal{G}] = Y\mathsf{E}[X \,|\, \mathcal{G}]$, a.s.

Remark 2.3. By part (vi), $\mathsf{E}[\,\mathsf{E}[X \,|\, Y]\,] = \mathsf{E}[X]$ for any random variable Y, which is the non-measure theoretic version of the law of total expectation. Actually, part (vi) is just a special case of part (vii). Let $\mathcal{H} = \{\emptyset, \Omega\}$. Then, by part (v), $\mathsf{E}[\,\mathsf{E}[X \,|\, \mathcal{G}]] = \mathsf{E}[\,E[X \,|\, \mathcal{G}] \,|\, \mathcal{H}] = \mathsf{E}[X \,|\, \mathcal{H}] = \mathsf{E}[X]$, a.s.

Proposition 2.2 (Conditional expectation and independence)**.** *Let X, Y, Z be integrable random variables and $\mathcal{G} \subset \mathcal{F}$ be a given sub-σ-algebra.*

(i) If $\sigma(X)$ and $\mathcal{G}$ are independent, then $\mathsf{E}[X \,|\, \mathcal{G}] = \mathsf{E}[X]$, a.s.

(ii) Suppose X, Y are independent, and ϕ is a Borel function such that $\mathsf{E}|\phi(X, Y)| < \infty$. Define a function f by letting $f(x) = \mathsf{E}[\phi(x, Y)]$ for each $x \in \mathbb{R}$. Then, $\mathsf{E}[\phi(X, Y) \,|\, X] = f(X)$, a.s.

(iii) If $\sigma(X, Y)$ is independent of $\sigma(Z)$, $\mathsf{E}[Y \,|\, X, Z] = \mathsf{E}[Y \,|\, X]$, a.s.

Proposition 2.3 (Limits of conditional expectation)**.** *Let X and $\{X_n\}$ be integrable random variables and $\mathcal{G} \subset \mathcal{F}$ be a given sub-σ-algebra.*

(i) *MCT: If* $0 \le X_n \uparrow X$*, then* $\mathsf{E}[X_n \,|\, \mathcal{G}] \uparrow \mathsf{E}[X \,|\, \mathcal{G}]$*, a.s.*

(ii) *DCT: If* $X_n \to X$ *and* $|X_n| \le Z$ *for some integrable random variable* Z*, then* $\mathsf{E}[X \,|\, \mathcal{G}] = \lim_{n\to\infty} \mathsf{E}[X_n \,|\, \mathcal{G}]$*, a.s.*

Exercise 2.1. Let $(\mathcal{G}_n)_{n\ge 1}$ be a filtration on the probability space $(\Omega, \mathcal{F}, \mathsf{P})$. Let $(X_n)_{n\ge 1}$ be adapted to $(\mathcal{G}_n)$ and also a martingale w.r.t. $(\mathcal{G}_n)$. Define $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ for each n. Prove that (i) $\mathcal{F}_n \subset \mathcal{G}_n$ for each n, and (ii) (X_n) is a martingale w.r.t. $(\mathcal{F}_n)$.

2.3 Inequalities involving conditional expectations

Assume X is an *integrable* random variable defined on $(\Omega, \mathcal{F}, \mathsf{P})$ and $\mathcal{G} \subset \mathcal{F}$ is a given sub-σ-algebra.

Proposition 2.4 (Jensen's inequality for conditional expectation)**.** *If* φ *is convex and* $E|\varphi(X)| < \infty$*, then*

$$\varphi(E[X \,|\, \mathcal{G}]) \le E[\varphi(X) \,|\, \mathcal{G}], \quad \text{a.s.}$$

Theorem 2.2 (Conditional expectation as a projection)**.** *Let* $L^2(\Omega, \mathcal{G}, \mathsf{P}) = \{Y : Y \in \mathcal{G},\ EY^2 < \infty\}$*. If* $EX^2 < \infty$*, then* $\inf_{Y \in L^2(\Omega,\mathcal{G},\mathsf{P})} E(X - Y)^2$ *is attained by* $Y = E[X \,|\, \mathcal{G}]$*.*

Theorem 2.3 (Conditional expectation as a contraction)**.** *Suppose* $p \ge 1$*. If* $E|X|^p < \infty$*, then* $\|E[X \,|\, \mathcal{G}]\|_p \le \|X\|_p$*.*

Exercise 2.2. Let $X, Y_1, Y_2, \dots$ be random variables defined on the same probability space $(\Omega, \mathcal{F}, \mathsf{P})$ such that $\mathsf{E}|X| < \infty$. Let $\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$ for each n, and $X_n = \mathsf{E}[X \,|\, \mathcal{F}_n]$. Is $(X_n)_{n\ge 1}$ a martingale w.r.t. $(\mathcal{F}_n)_{n\ge 1}$?

2.4 Conditional probability distributions

Definition 2.2. Consider the probability space $(\Omega, \mathcal{F}, \mathsf{P})$, a sub-$\sigma$-field $\mathcal{G} \subset \mathcal{F}$, and a random variable $X \colon (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\mathsf{E}|X| < \infty$. The conditional probability $\mathsf{P}(X \in A \,|\, \mathcal{G})$ for any $A \in \mathcal{B}(\mathbb{R})$ is defined as

$$\mathsf{P}(X \in A \,|\, \mathcal{G}) = \mathsf{E}[\mathbb{1}_{\{X \in A\}} \,|\, \mathcal{G}].$$

Theorem 2.4. *Consider the setting of Definition 2.2. There always exists a function* $p : \Omega \times \mathcal{B}(\mathbb{R}) \to [0, 1]$*, which is called a regular conditional distribution of* X *given* $\mathcal{G}$*, such that* *(i) for each* $A \in \mathcal{B}(\mathbb{R})$*, the function* $p(\cdot, A)$ *is a version of* $\mathsf{P}(X \in A \,|\, \mathcal{G})$*;*

(ii) for P*-almost every* $\omega \in \Omega$*, the function* $p(\omega, \cdot)$ *is a probability measure on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$*.*

Remark 2.4. The above theorem is not necessarily true if X does not take values in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. There are explicit counterexamples where the regular conditional distribution does not exist.

Proposition 2.5. *Let* $Z = (X, Y)\colon (\Omega, \mathcal{F}, \mathsf{P}) \to (\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ *be a random vector with density* $f_Z = d(\mathsf{P} \circ Z^{-1})/dm^2$*. Define* $f_Y(y) = \int_{\mathbb{R}} f_Z(x, y) m(dx)$ *and* $f_{X|Y}(x, y) = f_Z(x, y)/f_Y(y)$*. Then,*

$$p(\omega, A) = \int_A f_{X|Y}(x, Y(\omega)) m(dx), \qquad \forall \omega \in \Omega,\ A \in \mathcal{B}(\mathbb{R}).$$

is the regular conditional distribution of X *given* $\sigma(Y)$*. In other words, the regular conditional distribution of* X *given* $Y = y$ *has density* $f_{X|Y}(\cdot, y)$*.*

References

[1] Patrick Billingsley. *Probability and measure.* John Wiley & Sons, 2017.

[2] Rick Durrett. *Probability: theory and examples*, volume 49. Cambridge university press, 2019.

[3] Achim Klenke. *Probability theory: a comprehensive course.* Springer Science & Business Media, 2013.

[4] Sidney Resnick. *A probability path.* Springer, 2019.