Conditional Expectation

Robert L. Wolpert
Institute of Statistics and Decision Sciences
Duke University, Durham, NC, USA

Let μ and λ be two positive bounded measures on the same meaurable space $(\Omega, \mathcal{F})$. We call μ and λ *equivalent*, and write $\mu \equiv \lambda$, if they have the same null sets— so, if they were probability measures, the notion of "*a.s.*" would be the same for both. More generally, we call λ *absolutely continuous* (AC) w.r.t. μ, and write $\lambda \ll \mu$, if $\mu(A) = 0$ implies $\lambda(A) = 0$, *i.e.*, if every μ-null set is also λ-null. We call μ and λ *mutually singular*, and write $\mu \perp \lambda$, if for some set $A \in \mathcal{F}$ we have $\mu(A^c) = 0$ and $\lambda(A) = 0$, so μ and λ are "concentrated" on disjoint sets.

For example— if $\lambda(A) = \int_A f(x)\mu(dx)$ for some non-negative function $f \in L_1(\Omega, \mathcal{F}, \mu)$ then $\lambda \ll \mu$; if $f > 0$ μ-*a.s.*, then also $\mu(A) = \int_A f(x)^{-1}\lambda(dx)$ and $\mu \equiv \mu$. If for some other measure ν and some $f, g \in L_1(\Omega, \mathcal{F}, \nu)$ with

$$\mu(A) = \int_A f(x)\nu(dx) \qquad \lambda(A) = \int_A g(x)\nu(dx)$$

then $\mu \perp \lambda$ if $f(x)g(x) = 0$ for ν-a.e. $x \in \Omega$.

Theorem 1 (Lebesgue Decomposition) *There exist a unique pair of measures λ_a, λ_s on $(\Omega, \mathcal{F})$ and a unique function $Y \in L_1(\Omega, \mathcal{F}, \mu)$ such that:*

$$\begin{aligned} \lambda &= \lambda_a + \lambda_s \\ \lambda_a &\ll \mu, \qquad \lambda_s \perp \mu \\ \lambda_a(A) &= \int_A Y(\omega)\mu(d\omega), \qquad A \in \mathcal{F}. \end{aligned}$$

Corollary 1 (Radon-Nikodym Theorem) *Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space, $X \in L_1(\Omega, \mathcal{F}, \mathsf{P})$, and $\mathcal{G} \subset \mathcal{F}$ a sub-σ-algebra. Then there exists a unique $Y \in L_1(\Omega, \mathcal{F}, \mathsf{P})$, which we will denote $Y = \mathsf{E}[X \mid \mathcal{G}]$ and call the "conditional expectation of X, given $\mathcal{G}$," that satisfies:*

$$\mathsf{E}(Y - X)1_G = 0, \qquad G \in \mathcal{G}$$

Proof. First take X to be non-negative, $X \geq 0$. Define a measure λ on $\mathcal{G}$ (*not* on all of $\mathcal{F}$) given by

$$\lambda(G) = \mathsf{E}X\,1_G = \int_G X(\omega)\,\mathsf{P}(d\omega).$$

This is bounded (since $X \in L_1(\Omega, \mathcal{F}, \mathsf{P})$) and positive (since $X \geq 0$), so by RN we can write $\lambda = \lambda_a + \lambda_s$ with $\lambda_a \ll \mathsf{P}$, $\lambda_s \perp \mathsf{P}$, and $\lambda_a(G) = \int_G Y\,d\mathsf{P}$ for some $Y \in L_1(\Omega, \mathcal{G}, \mathsf{P})$. But $\lambda \ll \mathsf{P}$ by construction, so $\lambda_s = 0$ and the Corollary follows.

For general X, consider separately the positive and negative parts $X_+ = \max(X, 0)$ and $X_- = \max(-X, 0)$ and set $Y = Y_+ - Y_-$. ☐

Example: If $\mathcal{G} = \sigma\{\Lambda_n\}$ for any finite or countable partition $\{\Lambda_n\} \subset \mathcal{F}$ (so $\Lambda_m \cap \Lambda_n = \emptyset$ for $m \neq n$ and $\Omega = \cup \Lambda_n$), then for any $X \in L_1(\Omega, \mathcal{F}, \mathsf{P})$,

$$\mathsf{E}[X \mid \mathcal{G}] = \sum 1_{\Lambda_n} \mathsf{E}_{\Lambda_n}[X] = \sum 1_{\Lambda_n}(\omega) \frac{1}{\mathsf{P}[\Lambda_n]} \mathsf{E}[X\, 1_{\Lambda_n}],$$

constant on partition elements and equal there to the P-weighted average value of X.

In particular— let $(\Omega, \mathcal{F}, \mathsf{P})$ be the unit interval with Lebesgue measure, and let $\mathcal{G}_n = \sigma\{(i/2^n, j/2^n]\}$, $0 \leq i < j \leq 2^n$. Note that $\mathcal{G}_n \subset \mathcal{G}_m$ for $n \leq m$ and that $\mathcal{F} = \bigvee \mathcal{G}_n$. Then for any $X \in L_1(\Omega, \mathcal{F}, \mathsf{P})$,

$$X_n = \mathsf{E}[X \mid \mathcal{G}_n] = 2^n \int_{i/2^n}^{(i+1)/2^n} X\, d\mathsf{P}, \qquad i/2^n < \omega \leq (i+1)/2^n.$$

This is our first example of a *margingale*, a sequence of random variables $X_n \in L_1(\Omega, \mathcal{F}, \mathsf{P})$ with the property that $X_n = \mathsf{E}[X_m \mid \mathcal{G}_n]$ for $n \leq m$; we'll see more soon. What happens as $n \to \infty$?

Properties:

- If $X = 1_A$ and if $\mathcal{G} = \sigma\{B\}$ for some $A, B \in \mathcal{F}$, then

  $$\mathsf{E}[1_A \mid \sigma(B)](\omega) = \begin{cases} \mathsf{P}[A \cap B]/\mathsf{P}[B] & \omega \in B \\ \mathsf{P}[A \cap B^c]/\mathsf{P}[B^c] & \omega \notin B \end{cases}$$

  Thus, conditional expectation (given a σ-algebra $\mathcal{G}$) generalizes the elementary notion of contional probability (given an event B).

- More generally, If $X \in L_1$ and if $\mathcal{G} = \sigma\{G_i\}$ for some (finite or countable) measurable partition $\{G_i\} \subset \mathcal{F}$, then

  $$\mathsf{E}[X \mid \mathcal{G}](\omega) = \sum \mathbf{1}_{G_i}(\omega) \frac{1}{\mathsf{P}(G_i)} \int_{G_i} X(\omega)\, P(d\omega),$$

  the weighted average of X over the partition element that contains ω.

- If $X, Y \sim f(x, y)$ are jointly absolutely-continuous and if $\mathcal{G} = \sigma(Y)$,

  $$\mathsf{E}[X \mid \sigma(Y)] = \frac{\int x f(x, Y)\, dx}{\int f(x, Y)\, dx}.$$

  Thus, conditional expectation (given a σ-algebra $\mathcal{G}$) generalizes the elementary notion of contional expectation (given an RV Y). What if X and Y are both discrete? What if just one is discrete? What if Y is a vector?

  To prove this property, first show that any event G is $\sigma(Y)$-measurable if and only if $1_G = \phi(Y)$ a.s. for some Borel measurable ϕ (use a $\pi - \lambda$ argument), then extend from 1_G to arbitrary $\sigma(Y)$-measurable random variables.

- If $X \in L_1(\Omega, \mathcal{F}, \mathsf{P})$ and if $X \perp\!\!\!\perp \mathcal{G}$ then

$$\mathsf{E}[X \mid \mathcal{G}] = \mathsf{E}X$$

  In particular, $\mathsf{E}[X \mid \{\Omega, \emptyset\}] = \mathsf{E}X$. Thus, conditional expectation (given a σ-algebra $\mathcal{G}$) generalizes the elementary notion of expectation.

- If $X \in L_1(\Omega, \mathcal{F}, \mathsf{P})$ and if $\mathcal{G} \subset \mathcal{H} \subset \mathcal{F}$, then

$$\mathsf{E}[X \mid \mathcal{G}] = \mathsf{E}\big[\ \mathsf{E}[X \mid \mathcal{H}]\ \big|\mathcal{G}\big]$$

  This is called the "tower" property of conditional expectation.

- If $X \in L_2(\Omega, \mathcal{F}, \mathsf{P})$ and $\{Y_n\} \subset L_2(\Omega, \mathcal{F}, \mathsf{P})$ then $\mathsf{E}[X \mid \sigma\{Y_n\}]$ is the orthogonal projection of X onto the linear span of $\{Y_n\}$ in the Hilbert space $L_2(\Omega, \mathcal{F}, \mathsf{P})$. Thus, conditional expectation (given a σ-algebra $\mathcal{G}$) generalizes the notion of orthogonal projection. This is the best way to compute conditional expectations in multivariate normal examples.

- Let $\{X_n\} \overset{\text{iid}}{\sim} L_1(\Omega, \mathcal{F}, \mathsf{P})$ with means $\mu = \mathsf{E}[X_n]$ and set $S_n = \sum_{j \le n} X_j$, $\mathcal{G}_n = \sigma\{X_1, ..., X_n\}$. Then for $n < m$,

$$\mathsf{E}[S_m \mid \mathcal{G}_n] = S_n + (m-n)\mu;$$

  in particular, then S_n is another martingale if $\mu = 0$. If $\sigma^2 = \mathsf{V}X_n < \infty$, check that $(S_n - n\mu)^2 - n\sigma^2$ is a martingale.

- All the usual integration tools and inequalities— DCT, MCT, Fatou, Jensen, Hölder and Minkowski, Markov, Chebychev, *etc.*— hold for *conditional* expecatations as well.