A Quick Overview of Conditional Expectations

Definition: Let $f \in L^1(X, \mu)$ on a finite measure space $(X, \mathcal{A}, \mu)$. A *conditional expectation* of f with respect to a sub-σ-algebra $\mathcal{B} \subseteq \mathcal{A}$ is a $\mathcal{B}$-measurable μ-integrable function $\mathbb{E}_{\mathcal{B}}(f)$ such that

$$\int gf \, d\mu = \int g\mathbb{E}_{\mathcal{B}}(f) \, d\mu \tag{†}$$

for all bounded $\mathcal{B}$-measurable functions g.

For any positive measurable function φ, we will define $\mathbb{E}_{\mathcal{B}}(\varphi) = \lim_{n\to\infty} \mathbb{E}_{\mathcal{B}}(\min(\varphi, n))$.

Theorem: Conditional expectations exist and are unique μ-almost everywhere. They also satisfy the following:

(i) $\mathbb{E}_{\mathcal{B}}(f) = f$ μ-a.e. for all $\mathcal{B}$-measurable μ-integrable functions f;

(ii) $\mathbb{E}_{\mathcal{B}}(f) \geq 0$ μ-a.e. if $f \geq 0$ μ-a.e.;

(iii) if $(f_n)_n$ is an increasing sequence of positive μ-integrable functions converging almost everywhere to a μ-integrable function f, then $\mathbb{E}_{\mathcal{B}}(f_n) \to \mathbb{E}_{\mathcal{B}}(f)$ μ-a.e.;

(iv) if $(f_n)_n \to f$ is a sequence of integrable functions converging μ-a.e. to an integrable function f, and $|f_n| \leq F$ for some $F \in L^1(X, \mu)$, then $\lim_{n\to\infty} \mathbb{E}_{\mathcal{B}}(f_n) = \mathbb{E}_{\mathcal{B}}(f)$;

(v) if $f_n \leq F$ μ-a.e. for some $F \in L^1(X, \mu)$, we have $\limsup_{n\to\infty} \mathbb{E}_{\mathcal{B}}(f_n) \leq \mathbb{E}_{\mathcal{B}} \limsup_{n\to\infty} f_n$ μ-a.e.;

(vi) if $f_n \geq G$ μ-a.e. for some $G \in L^1(X, \mu)$, then $\mathbb{E}_{\mathcal{B}}(\liminf_{n\to\infty} f_n) \leq \liminf_{n\to\infty} \mathbb{E}_{\mathcal{B}}(f_n)$;

(vii) if $(X, \mathcal{A}, \mu)$ is a probability space, and $(\mathcal{B}_n)_n$ is a filtration on $\mathcal{A}$, then for every $f \in L^1(X, \mu)$, we have that $\mathbb{E}_{\mathcal{B}_n}(f) \to \mathbb{E}_{\mathcal{B}_\infty}(f)$ pointwise almost everywhere.

Proof. If

$$\nu(E) = \int f 1_E \, d\mu,$$

then we observe that $\nu \ll \mu$. Define $\mathbb{E}_{\mathcal{B}}(f) := \frac{d\nu}{d\mu}$. By bootstrapping and the properties of the Radon–Nikodym derivative, we obtain the conditional expectation properties, and since Radon–Nikodym derivatives are unique, it follows that the conditional expectation itself is unique.

(i) Radon–Nikodym derivatives are unique.

(ii) If ν is a positive measure, then $\frac{d\nu}{d\mu}$ is necessarily a positive function.

(iii) If $(f_n)_n \to f$ with $f_n, f \in L^1(X, \mu)$ is increasing, it follows that $g_n = \mathbb{E}_{\mathcal{B}}(f_n)$ is almost everywhere increasing. So, $\lim_{n\to\infty} g_n =: g$ is $\mathcal{B}$-measurable and μ-integrable by Fatou's lemma, so by dominated convergence, the defining property (†), and the uniqueness of Radon–Nikodym derivatives, we obtain that $g = \mathbb{E}_{\mathcal{B}}(f)$.

(iv) Use dominated convergence along with the defining property (†).

(v) This emerges from the reverse Fatou's Lemma on finite measure spaces along with the defining property (†).

(vi) This emerges from Fatou's Lemma along with the defining property (†).

(vii) We start by showing L^1 convergence. Observe that, if we let $\mathcal{B}_\infty = \sigma(\bigcup_n \mathcal{B}_n)$, then we have that $\mathbb{E}_{\mathcal{B}_n}(f) = \mathbb{E}_{\mathcal{B}_n}\mathbb{E}_{\mathcal{B}_\infty}(f)$. If $f = 1_B$ and $\varepsilon > 0$, then there is some N and $C \in \mathcal{B}_N$ such that $\mu(B \triangle C) < \varepsilon$ since such a set exists in the algebra generated by all the $\mathcal{B}_n$. Let $g = 1_C$. Then, $\mathbb{E}_{\mathcal{B}_n} g = g$ for all $n \geq N$. Since

$$f - \mathbb{E}_{\mathcal{B}_n}(f) = (f - g) + \mathbb{E}_{\mathcal{B}_n}(g - f),$$

and $\mathbb{E}_{\mathcal{B}}\colon L^1(X,\mathcal{A},\mu)\to L^1(X,\mathcal{B},\mu)$ is a contractive linear map, we have that

$$\begin{aligned}\left\|\mathbb{E}_{\mathcal{B}_n}(f)-\mathbb{E}_{\mathcal{B}_n}(g)\right\|_{L^1} &\leq \|f-g\|_{L^1}\\ &= \varepsilon,\end{aligned}$$

so the claim is true for indicators. Bootstrapping yields the desired result.

Next, we show that a version of Chebyshev's inequality holds. That is, we have that

$$\mu\left(\left\{x : \sup_n\left|\mathbb{E}_{\mathcal{B}_n}(f)\right| > c\right\}\right) \leq \frac{1}{c}\int_X |f|\,d\mu.$$

For nonnegative $f_n \coloneqq \mathbb{E}_{\mathcal{B}_n}(f)$, we let $E = \{x : \sup_n f_n(x) > c\}$. If

$$E_j = \Big\{x : f_j(x) > c \text{ and } f_i(x)\leq c \text{ for all } i = 1,\dots,j-1\Big\},$$

then E_j are measurable, disjoint, and their union is E. Since $E_j\in\mathcal{B}_j$ and $\left(\mathcal{B}_j\right)_j$ is a filtration, we have that f_j is $\mathcal{B}_j$-measurable. In particular, using monotone convergence and the defining property (†), we have

$$\begin{aligned}\int_X f\,d\mu &\geq \int_E f\,d\mu\\ &= \sum_{j=1}^{\infty}\int_{E_j} f\,d\mu\\ &= \sum_{j=1}^{\infty}\int_{E_j} f_j\,d\mu\\ &\geq c\sum_{j=1}^{\infty}\mu\Big(E_j\Big)\\ &= c\mu(E).\end{aligned}$$

Define

$$\psi(x) = \limsup_{n\to\infty}\left|\mathbb{E}_{\mathcal{B}_n}(f)(x) - f(x)\right|.$$

Let $\varepsilon>0$. For sufficiently large n, there is a $\mathcal{B}_n$-measurable integrable function g such that $\|f-g\|_{L^1} < \varepsilon^2$. Since $\mathbb{E}_{\mathcal{B}_m}(g) = g$ for all $m\geq n$, we have

$$\begin{aligned}\psi(x) &\leq \limsup_{n\to\infty}\left|\mathbb{E}_{\mathcal{B}_n}(f-g)(x)\right| + \limsup_{n\to\infty}\left|\mathbb{E}_{\mathcal{B}_n}(g)(x)-g(x)\right| + |f(x)-g(x)|\\ &= \limsup_{n\to\infty}\left|\mathbb{E}_{\mathcal{B}_n}(f-g)(x)\right| + |f(x)-g(x)|.\end{aligned}$$

We have

$$\begin{aligned}\mu\left(\left\{x : \limsup_{n\to\infty}\left|\mathbb{E}_{\mathcal{B}_n}(f-g)(x)\right| > \varepsilon\right\}\right) &\leq \mu\left(\left\{x : \sup_n\left|\mathbb{E}_{\mathcal{B}_n}(f-g)(x)\right| > \varepsilon\right\}\right)\\ &\leq \frac{1}{\varepsilon}\|f-g\|_{L^1}\\ &< \varepsilon,\end{aligned}$$

and

$$\mu(\{x : |f(x)-g(x)| > \varepsilon\}) \leq \frac{1}{\varepsilon}\|f-g\|_{L^1}$$

$$< \varepsilon,$$

we have that

$$\mu(\{x : \psi(x) > 2\varepsilon\}) \leq 2\varepsilon,$$

so that $\psi \equiv 0$ almost everywhere.

□

Conditional Measures

When $f = 1_A$ is the indicator function on some $A \in \mathcal{A}$, then we define $\mathbb{E}_{\mathcal{B}}(1_A) =: \mu(A|\mathcal{B})$ as the *conditional measure* of A with respect to the sub-σ-algebra $\mathcal{B}$.

We may assert that the sub-σ-algebra $\mathcal{B}$ is produced by a measurable mapping $\pi\colon X \to (Y, \mathcal{E})$. For a positive μ, we may find for any $A \in \mathcal{A}$ a $\mathcal{B}$-measurable function $\mu(A, \cdot)$ such that

$$\mu(A \cap B) = \int_B \mu(A, x)\, d\mu(x)$$

for all $B \in \mathcal{B}$. Since $x \mapsto \mu(A, x)$ is $\mathcal{B}$-measurable, there is a $\mathcal{E}$-measurable function $y \mapsto \mu^y(A)$ on Y such that $\mu(A, x) = \mu^{\pi(x)}(A)$. If $\nu = \mu \circ \pi^{-1}$ is the pushforward measure of μ with respect to π on Y, then we may write

$$\mu\Big(A \cap \pi^{-1}(E)\Big) = \int_E \mu^y(A)\, d\nu(y)$$

for all $E \in \mathcal{E}$. If $E = Y$, then we have

$$\mu(A) = \int \mu^y(A)\, d\nu(y).$$

That is, we have a Fubini-type theorem where we may compute $\mu(A)$ by computing the conditional measures on the level sets $\pi^{-1}(\{y\})$ and integrating with respect to the measure ν.

Definition: Let $(X, \mathcal{A}, \mu)$ be a measure space, and let $\mathcal{B} \subseteq \mathcal{A}$ be a sub-σ-algebra. We say a function $\mu^{\mathcal{B}}\colon \mathcal{A} \times X \to \mathbb{R}$ is a *regular conditional measure* on $\mathcal{A}$ with respect to $\mathcal{B}$ if

(i) for every x, $A \mapsto \mu^{\mathcal{B}}(A, x)$ is a measure on $\mathcal{A}$;

(ii) for every $A \in \mathcal{A}$, the function $x \mapsto \mu^{\mathcal{B}}(A, x)$ is $\mathcal{B}$-measurable and $|\mu|$-integrable;

(iii) one has

$$\mu(A \cap B) = \int \mu^{\mathcal{B}}(A, x)\, d|\mu|(x)$$

for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$.

Definition: A system of regular conditional measures μ^y generated by a $\mathcal{A}$-$\mathcal{E}$-measurable map $\pi\colon X \to Y$ is a function $(A, y) \mapsto \mu^y(A)$ on $\mathcal{A} \times Y$ such that, $\mu^y(\cdot)$ is a measure on $\mathcal{A}$ for fixed y and $y \mapsto \mu^y(A)$ is $\mathcal{E}$-measurable and $\pi_*|\mu|$-integrable for fixed A. Furthermore, for all $A \in \mathcal{A}$ and $E \in \mathcal{E}$, it satisfies

$$\mu\Big(A \cap \pi^{-1}(E)\Big) = \int \mu^y(A)\, d\pi_*|\mu|.$$

If for $\pi_*|\mu|$-almost every point $y \in Y$, we have $\pi^{-1}(\{y\}) \in A$, and μ^y is concentrated on $\pi^{-1}(\{y\})$, then we say $\{\mu^y : y \in Y\}$ is a system of *proper* conditional measures.

Lemma: Let $\mathcal{A}$ be countably generated. Then, regular conditional measures on $\mathcal{A}$ are essentially unique.

Furthermore, if μ is a probability measure, then $\mu^{\mathcal{B}}(\cdot, x)$ is a probability measure for μ-a.e. x.

Proof. Let $\mu_1^{\mathcal{B}}$ and $\mu_2^{\mathcal{B}}$ be two such conditional measures. Let $\mathcal{R} = \{A_n : n \in \mathbb{N}\}$ be a countable algebra generating $\mathcal{A}$. We have that for every $A_n \in \mathcal{R}$, there is $Z_n \in \mathcal{B}$ such that $\mu_1^{\mathcal{B}}(A_n, x) = \mu_2^{\mathcal{B}}(A_n, x)$ for all $x \in X \setminus Z_n$, and $\left|\mu\right|(Z_n) = 0$. Take $Z = \bigcup_n Z_n$.

If μ is a probability measure, then $\mu^{\mathcal{B}}(A_n, x)$ is positive μ-a.e. since its integral over every $B \in \mathcal{B}$ is positive, and $\mu^{\mathcal{B}}(X, x) = 1$ for μ-a.e. x, so $\mu^{\mathcal{B}}(\cdot, x)$ is a probability measure on $\mathcal{R}$, so it is a probability measure on $\mathcal{A}$. □

Our next goal is to understand exactly when regular conditional measures exist.

Definition: A family $\mathcal{K} \subseteq \mathcal{A}$ is called a compact approximating class if it satisfies the following:

- for any sequence $(K_n)_n$ of elements of $\mathcal{K}$ with $\bigcap_n K_n = \emptyset$, there exists N such that $\bigcap_{n=1}^N K_n = \emptyset$;
- for any $A \in \mathcal{A}$ and any $\varepsilon > 0$, there is $K_\varepsilon \in \mathcal{K}$ and $A_\varepsilon \in \mathcal{A}$ such that $A_\varepsilon \subseteq K_\varepsilon \subseteq A$ and $\mu(A \setminus A_\varepsilon) < \varepsilon$.

Theorem: Suppose the σ-algebra $\mathcal{A}$ is countably generated, and that there is a compact approximating class in $\mathcal{A}$. Then, for every sub-σ-algebra $\mathcal{B} \subseteq \mathcal{A}$, there is a regular conditional measure $\mu^{\mathcal{B}}$ such on $\mathcal{A}$.

More generally, let $\mathcal{A}_0$ be a sub-σ-algebra of $\mathcal{A}$ with countable generating subalgebra $\mathcal{U}$. Then, for every sub-σ-algebra $\mathcal{B} \subseteq A$, there is a regular conditional measure $\mu^{\mathcal{B}}_{\mathcal{A}_0}$ on $\mathcal{A}_0$.

Furthermore, for every $\mathcal{A}_0$-measurable μ-integrable function f, we have

$$\int f \, d\mu = \int \int f(y) \, d\mu^{\mathcal{B}}_{\mathcal{A}_0}(y, x) \, d\left|\mu\right|(x).$$

Proof. We start by considering a probability measure. Let $\mathcal{U} = \{A_n : n \in \mathbb{N}\}$ be a countable family of subsets of $\mathcal{A}$. Then, for every n, we find $C_{n,k} \in \mathcal{K}$ and $A_{n,k} \in \mathcal{A}$ such that $A_{n,k} \subseteq C_{n,k} \subseteq A$ and $\mu(A_n \setminus A_{n,k}) < 1/k$. Let $\mathcal{U}_0$ be the algebra generated by $\{A_{n,k} : n, k \in \mathbb{N}\}$ and $\{A_n : n \in \mathbb{N}\}$. By the Radon–Nikodym theorem, for every $A \in \mathcal{U}_0$, there is some positive $\mathcal{B}$-measurable function $x \mapsto p_0(A, x)$ such that $p_0(X, x) = 1$, $p_0(\emptyset, x) = 0$ for all $x \in X$ and

$$\mu(A \cap B) = \int_B p_0(A, x) \, d\mu(x).$$

for all $B \in \mathcal{B}$. Notice that $p_0(A \cup B, x) = p_0(A, x) + p_0(B, x)$ for all disjoint $A, B \in \mathcal{U}_0$ and for μ-almost every $x \in X$. Since the set of all pairs (A, B) for which this identity fails is countable as $\mathcal{U}_0$ is countable, the union of all subsets in $\mathcal{B}$ for which this equality fails is measure zero. Therefore, there is a measure zero subset $B_0 \in \mathcal{B}$ such that $A \mapsto p_0(A, x)$ is additive on $\mathcal{U}_0$ for all $x \in X \setminus B_0$.

We will show for almost every x, $p_0(A_n, x) = \sup_k p_0(A_{n,k}, x)$ for all $n \in \mathbb{N}$. Write $q_n(x) = \sup_k p_0(A_{n,k}, x)$. We will show that the set function $p_0(\cdot, x)$ is approximated on the algebra $\mathcal{U}$ by the class $\mathcal{K}$ with respect to $\mathcal{U}_0$.

Since $A_{n,k} \subseteq A_n$, we have a measure zero set $N_{n,k} \in \mathcal{B}$ such that $p_0(A_{n,k}, x) \leq p_0(A_n, x)$ for all $x \in X \setminus N_{n,k}$. Therefore, $q_n(x) \leq p_0(A_n, kx)$ for all $x \notin N_0 := \bigcup_{n,k=1}^\infty N_{n,k}$. However, since $p_0(A_{n,k}, x) \leq q_n(x)$, we have that

$$\begin{aligned} \mu(A_{n,k}) &= \int_X p_0(A_{n,k}, x) \, d\mu(x) \\ &\leq \int_X q_n(x) \, d\mu(x). \end{aligned}$$

Since $\mu(N_0) = 0$, we have that

$$\begin{aligned} \sup_k \mu(A_{n,k}) &\leq \int_X q_n(x) \, d\mu(x) \\ &\leq \int_X p_0(A_n, x) \, d\mu(x) \\ &= \mu(A_n). \end{aligned}$$

Since $\sup_k \mu(A_{n,k}) = \mu(A_n)$, we have that $q_n(x) = p_0(A_n, x)$ for all x save for some measure zero set $B_1 \in \mathcal{B}$.

Thus, for all $x \notin N := B_0 \cup B_1$, the additive set function $p_0(\cdot, x)$ on the algebra $\mathcal{U}_0$ is such that $\mathcal{K}$ approximates $p_0(\cdot, x)$ with respect to $\mathcal{U}_0$. This set function is countably additive on $\mathcal{U}$, so by the Carathéodory extension theorem, it extends to a countably additive measure on $\mathcal{A}_0$. If we take $\mu(\cdot, x) = \mu^{\mathcal{B}}_{\mathcal{A}_0}(\cdot, x)$, then for all $x \in N$, we define $\mu(\cdot, x) = \mu$.

We will now verify that this is our desired family of conditional measures. If $A = A_n$, then $x \mapsto \mu(A, x)$ is measurable with respect to $\mathcal{B}$. The class of all sets $A \in \sigma(\mathcal{U})$ for which this is true is a monotone class. Therefore, it coincides with $\sigma(\mathcal{U})$. If $B \in \mathcal{B}$ and $A = A_n$, then by construction, we have

$$\mu(A \cap B) = \int_B \mu(A, x)\, d\mu(x).$$

Let $B \in \mathcal{B}$ be fixed. The class $\mathcal{E}$ of all $A \in \sigma(\mathcal{U})$ for which this equation holds is monotone. If $E_j \nearrow E$ in $\mathcal{E}$, then $\lim_{j\to\infty} \mu\left(E_j, x\right) = \mu(E, x)$, so by the dominated convergence theorem, we have that $E \in \mathcal{E}$. Thus, $\mathcal{E} = \sigma(\mathcal{U})$. Since $\sigma(\mathcal{U}) = \mathcal{A}_0$, we obtain the desired result.

Finally, it suffices to show the integral equality

$$\int_X f\, d\mu = \int_X \int_X f(y)\, d\mu(y, x)\, d\mu(x)$$

for indicator functions on sets in $\mathcal{A}_0$, but it already is the case by definition. □

Finally, we may represent this result in terms of a map π producing a σ-algebra $\mathcal{B}$.

Theorem: Let μ be a (possibly signed) measure on a measurable space $(X, \mathcal{A})$, $(Y, \mathcal{E})$ a measurable space, and $\pi\colon (X, \mathcal{A}) \to (Y, \mathcal{E})$ a map measurable with respect to $\left(\mathcal{A}_\mu, \mathcal{E}\right)$, where $\mathcal{A}_\mu$ is the σ-algebra of μ-measurable subsets of X with respect to $\mathcal{A}$. Suppose $\pi(X) \in \mathcal{E}_{\pi_*|\mu|}$, where $\mathcal{E}_{\pi_*|\mu|}$ is the completion of $\mathcal{E}$ with respect to $\pi_*\left|\mu\right|$. Additionally, assume that $\mathcal{A}$ is countably generated and μ has a compact approximating class. Then, there exist regular conditional measures μ^y generated by π on $\mathcal{A}$.

Proof. Let $\mathcal{B} = \pi^{-1}(\mathcal{E})$. Then, if $\mathcal{A}_0 = \mathcal{A}$ and $\mathcal{A} = \mathcal{A}_\mu$ in the previous theorem, we find a conditional measure $\mu^{\mathcal{B}}(A, x)A$ on $\mathcal{A}$ such that the function $x \mapsto \mu^{\mathcal{B}}(A, x)$ is $\mathcal{B}$-measurable. Therefore, for every $A \in \mathcal{A}$, there is a $\mathcal{E}$-measurable function $\eta(A, y)$ such that $\mu^{\mathcal{B}}(A, x) = \eta(A, \pi(x))$.

By the hypothesis, there is $Y_0 \in \mathcal{E}$ with $Y_0 \subseteq \pi(X)$ and $\pi_*\left|\mu\right|(Y \setminus Y_0) = 0$. Thus, for all $y \in Y_0$, $\eta(A, y)$ is a measure as a function of A. Define $\mu^y(A) = \eta(A, y)$. If $y \notin Y_0$, we let $\mu^y = \mu$. For each $A \in A$, the function $y \mapsto \mu^y(A)$ is $\mathcal{E}$-measurable since $Y \setminus Y_0 \in \mathcal{E}$. These are our desired regular conditional measures. □