Notes on conditional measures and expectation

Noga Labin

February 1, 2024

1 Conditional expectation

1.1 theorem

Let $(X, \mathcal{B}, \mu)$ be a probability space, and let $\mathcal{A} \subseteq \mathcal{B}$ be a sub σ-algebra. Then there is a map -

$$E(\cdot|\mathcal{A}) : L^1(X, \mathcal{B}, \mu) \to L^1(X, \mathcal{A}, \mu)$$

called the conditional expectation, with for any $f \in L^1(X, \mathcal{B}, \mu)$, the image function $E(f|\mathcal{A})$ holds for any $A \in \mathcal{A}$, $\int_A E(f|\mathcal{A}) d\mu = \int_A f d\mu$, and $E(f|\mathcal{A})$ is unique up to a null set.

1.2 example

if $\mathcal{A} = \sigma(A_1, A_2, ..., A_n)$ the σ-algebra generated by a finite partition, then

$$E(f|\mathcal{A})(x) = \frac{1}{\mu(A_i)} \cdot \int_{A_i} f d\mu \quad \text{(for } i \text{ s.t. } x \in A_i)$$

1.3 example

let $X = [0, 1]^2$, with Lebesgue measure, and $\mathcal{A} = \mathcal{B} \times \{\phi, [0, 1]\}$ then,

$$E(f|\mathcal{A})(x_1, x_2) = \int_0^1 f(x_1, t) dt$$

1.4 proof of theorem

1.4.1 existence

suppose $f \geq 0$, then $\mu_f(B) = \int_B f d\mu$ is an absolutely continuous finite measure on $(X, \mathcal{B})$ and $\mu_f|_{\mathcal{A}}$ is absolutely continuous relatively to $\mu|_{\mathcal{A}}$.

therefor there is a Radon-Nikodim derivative $g \in L^1(X, \mathcal{A}, \mu)$ which satisfies:

$$\forall A \in \mathcal{A}; \quad \int_A f d\mu = \mu_f(A) = \int_A g d\mu$$

this g is (a.e.) $E(f|\mathcal{A})$ and satisfies property 1.

for a general f, we decomposite it to a positive and negative part and reach the same conclusion.

1.4.2 uniqueness

let there be g_1, g_2 that satisfy property 1. then $A = \{x \in X : g_1(x) > g_2(x)\} \subseteq \mathcal{A}$ and

$$\int_A g_1 d\mu = \int_A f d\mu = \int_A g_2 d\mu$$

thus $\int_A (g_1 - g_2) d\mu = 0 \Rightarrow \forall \epsilon > 0; \mu(\{x \in X : g_1(x) - g_2(x) > \epsilon\}) = 0 \Rightarrow \mu(A) = 0$

and similarly, $\mu(g_1 < g_2) = 0$, therefor $g_1 = g_2$ a.e.

1.5 properties

1. $E(\cdot|\mathcal{A})$ is a positive linear operator. (if $f \geq 0$ a.e. then $E(f|\mathcal{A}) \geq 0$ a.e.)

2. for $f \in L^1(X, \mathcal{B}, \mu)$ and $g \in L^\infty(X, \mathcal{A}, \mu)$, $E(gf|\mathcal{A}) = gE(f|\mathcal{A})$ a.e.

3. if $\mathcal{A}' \subseteq \mathcal{A}$ a sub σ-algebra, then $E(E(f|\mathcal{A})|\mathcal{A}') = E(f|\mathcal{A}')$

4. if $f \in L^1(X, \mathcal{A}, \mu)$, then $E(f|\mathcal{A}) = f$ a.e.

5. for any $f \in L^1(X, \mathcal{B}, \mu)$, $|E(f|\mathcal{A})| \leq E(|f||\mathcal{A})$

1.5.1 proof of properties

1. linearity stems from uniqueness, and linearity of integrals. positiveness- if $f \geq 0$, and there is $A \in \mathcal{A}$ with $\mu(A) > 0$ s.t. $E(f|\mathcal{A})|_A < 0$, then $\int_A E(f|\mathcal{A}) d\mu < 0 \leq \int_A f d\mu$ in contradiction to the definition.

2. this property clearly holds if g is an indicator, thus also for any simple function. the general case follows by dominated convergence theorem.

3. this property stems from uniqueness, because $E(E(f|\mathcal{A})|\mathcal{A}')$ satisfies
   $\forall A \in \mathcal{A}'$; $\int_A E(E(f|\mathcal{A})|\mathcal{A}')d\mu = \int_A E(f|\mathcal{A})d\mu = \int_A f d\mu$ which is the definition of $E(f|\mathcal{A})$.

4. this property stems from uniqueness as well,
   because f also satisfies the definition if $f \in L^1(X, \mathcal{A}, \mu)$.

5. give $f \in L^1(X, \mathcal{B}, \mu)$ we may find $g \in L^\infty(X, \mathcal{A}, \mu)$ with
   $\forall x \in X, |g(x)| = 1$ s.t. $|E(f|\mathcal{A})| = g \cdot E(f|\mathcal{A})$ then by property 2,
   $|E(f|\mathcal{A})| = E(g \cdot f|\mathcal{A})$ so for any $A \in \mathcal{A}$;
   $\int_A |E(f|\mathcal{A})|d\mu = \int_A E(g \cdot f|\mathcal{A})d\mu = \int_A g \cdot f d\mu \leq \int_A |g \cdot f|d\mu = \int_A E(|f||\mathcal{A})d\mu$.

2 Conditional measure

The examples we have seen before seem to suggest that the quantity $E(f|\mathcal{A})(x)$ should be an average of the function f over a part of the measure space, where the part used in the averaging.
in 'nice enough' spaces, this property is true, and can be reflected by the existence of a measure $\mu_x^{\mathcal{A}}$ with the property

$$\forall f \in L^1_\mu; \quad \int_A E(f|\mathcal{A})(x) = \int f d\mu_x^{\mathcal{A}}$$

2.1 example

lets reconsider example 1.2, it is already articulated in this way, if $\mathcal{A} = \sigma(A_1, A_2, ..., A_n)$ the σ-algebra generated by a finite partition, then

$$E(f|\mathcal{A})(x) = \tfrac{1}{\mu(A_i)} \cdot \int_{A_i} f d\mu \quad \text{(for } i \text{ s.t. } x \in A_i)$$

meaning $\mu_x^{\mathcal{A}}(B) = \mu(B \cup A_i)$ when $(x \in A_i)$

2.2 example

lets consider the conditions in example 1.3, then $\quad \mu^{\mathcal{A}}_{(x_1,x_2)} = \delta_{x_1} \times m_{[0,1]}$

meaning the average on each horizontal line by the one dimensional Lebesgue measure.

2.3 theorem

Let $(X, \mathcal{B}, \mu)$ be a borel probability space, where X is a borel subset of a compact metric space $\bar{X}$, and $\mathcal{A} \subseteq \mathcal{B}$ a σ-algebra. then there exists an $\mathcal{A}$ measurable $X' \subseteq X$ with $\mu(X \backslash X') = 0$ and a system $\{\mu_x^{\mathcal{A}} : x \in X'\}$ of measures on X, referred to as conditional measures where $\mu_x^{\mathcal{A}}$ is a probability measure on X with

$$\forall f \in L^1(X, \mathcal{B}, \mu); \quad E(f|\mathcal{A})(x) = \int f d\mu_x^{\mathcal{A}}, \text{ } \mu\text{-almost everywhere}$$

and this property uniquely determines $\mu_x^{\mathcal{A}}$ for a.e. $x \in X$.

2.4 Riesz representation theorem for measures

Let X be a locally compact Hausdorff space and ψ a positive linear functional on $C_c(X)$ (compactly supported continuous functions). Then there exists a Borel σ-algebra Σ on X and a unique positive Borel measure μ on X such that-

$$\forall f \in C_c(X); \, \psi(f) = \int f d\mu$$

with the additional properties:

1. $\mu(K) < \infty$ for any compact K
2. μ is regular (both inner and outer)
3. (X, Σ, μ) is a complete measure space

we will not prove this theorem today, but we will use it.

2.5 proof of theorem (2.3)

we will proof it for a compact X, and if it is not we can simply complete it and make $\mu(\bar{X} \backslash X) = 0$

The idea is to use Riesz representation theorem on $E(\cdot|\mathcal{A})$, but since it is only defined linear and positive a.e. so we must do some work to create a fitting defined function with a single uniform null set to set aside.

2.5.1 Existence

Let $\mathcal{F} = \{f_0 = 1, f_1, f_2, f_3, ...\} \subseteq C(X)$ be a dense vector space over $\mathbb{Q}$, for each $i \in \mathbb{N}$ choose $g_i = E(f_i, \mathcal{A}), g_0 = 1$ now-

- $g_i(x) \geq 0$ a.e. if $f_i \geq 0$
- $|g_i(x)| \leq ||f||_\infty$ a.e.
- if $f_i = \alpha f_j + \beta f_k$, then $g_i = \alpha g_j + \beta g_k$ a.e.

let N be the union of all sets where those properties do not hold, it is null.

for any $x \notin N$, define $\Delta_x(f_i) = g_i(x)$, a positive $\mathbb{Q}$-linear map from $\mathcal{F}$ to $\mathbb{R}$, with $||\Delta_x|| \leq 1$ thus it extends uniquely to a continuous linear positive functional $\Delta_x : C(X) \to \mathbb{R}$

by the Riesz representation theorem for measures, there exists a measure $\mu_x^{\mathcal{A}}$ on X with

$$\forall f \in C(X);\ \Delta_x(f) = \int f d\mu_x^{\mathcal{A}}$$

moreover, $\Delta_x(1) = 1$ so $\mu_x^{\mathcal{A}}$ is a probability measure.

by dominated convergence and some added construction, we can extend this for any $f \in L^1(X, (\mathcal{B}), \mu)$, as shown thoroughly in page 139 in Einsiedler-Ward.

2.5.2 Uniqueness

if we have to sets of measures that satisfy the definition on a countable dense subset $\{f_n\} \in C(X)$, then for each $n \in \mathbb{N}$, and a.e. x

$$\int f_n d\rho_x = E(f_n|\mathcal{A})(x) = \int f_n d\nu_x$$

so there is a null set of all objects for which there is an n that this does not hold for, thus by dominated convergence

$$\forall f \in C(X), \forall x \notin N;\ \int f d\rho_x = \int f d\nu_x$$

hence $\forall x \notin N;\ \rho_x = \nu_x$.

2.6 definition

An atom, is the smallest set in a σ-algebra containing a point, and is marked

$$[x]_{\mathcal{A}} = \bigcap_{x \in A \in \mathcal{A}} A$$

notice how this is not necessarily maesurable because it might be an uncountable intersection, but in a countably generated σ-algebra, any atom is measurable.

2.7 simple claim

if $\mathcal{A}$ is countably generated, then $\forall x \in X' \quad \mu([x]_{\mathcal{A}}) = 1$

moreover, for any $x, y \in X'$; $[x]_{\mathcal{A}} = [y]_{\mathcal{A}} \Rightarrow \mu_x^{\mathcal{A}} = \mu_y^{\mathcal{A}}$.

proof:

we simply know $\mu([x]_{\mathcal{A}}) = E(\chi_{[x]_{\mathcal{A}}} | \mathcal{A})(x) = \chi_{[x]_{\mathcal{A}}}(x) = 1$.

and for x, y as mentioned, any measurable set that contains one contains the other, thus all indicators have an equal integral under the two measures, so they must be equal.