Introduction to Empirical Processes and Semiparametric Inference Lecture 07: Metric Spaces

Michael R. Kosorok, Ph.D.

Professor and Chair of Biostatistics

Professor of Statistics and Operations Research

University of North Carolina-Chapel Hill

Outer Expectation

For an arbitrary probability space $(\Omega, \mathcal{A}, P)$, consider an arbitrary map $T : \Omega \mapsto \bar{\mathbb{R}}$, where $\bar{\mathbb{R}} \equiv [-\infty, \infty]$.

The outer expectation of T, denoted E^*T, is the infimum over all EU, where $U : \Omega \mapsto \bar{\mathbb{R}}$ is measurable, $U \geq T$, and EU exists.

For EU to exist, it must not be indeterminate, although it can be $\pm\infty$, provided the sign is clear.

Since T is not necessarily a random variable, the proper term for E^*T is *outer integral*.

Nevertheless, we will use the term outer expectation in deference to its connection with the classical notion of expectation.

We analogously define inner expectation: $E_*T = -E^*[-T]$.

The following lemma verifies the existence of a minimal measurable majorant $T^* \geq T$:

LEMMA 1. *For any $T : \Omega \mapsto \bar{\mathbb{R}}$, there exists a minimal measurable majorant $T^* : \Omega \mapsto \bar{\mathbb{R}}$ with*

(i) $T^* \geq T$*;*

(ii) For every measurable $U : \Omega \mapsto \bar{\mathbb{R}}$ with $U \geq T$ a.s., $T^* \leq U$ a.s.

For any T^* satisfying (i) and (ii), $E^*T = ET^*$, provided ET^* exists; the last statement is true if $E^*T < \infty$.

Analagous results holds for the existence of a maximal measurable minorant T_* for which $E_*T = ET_*$.

An important special case of outer expectation is outer probability.

The outer probability of an arbitrary $B \subset \Omega$, denoted $P^*(B)$, is the infimum over all $P(A)$ such that $A \supset B$ and $A \in \mathcal{A}$.

The inner probability of an arbitrary $B \subset \Omega$ is defined to be $P_*(B) = 1 - P^*(\Omega - B)$.

The following lemma gives the precise connection between outer/inner expectations and outer/inner probabilities:

LEMMA 2. *For any* $B \subset \Omega$,

(i) $P^*(B) = E^* \mathbf{1}\{B\}$ *and* $P_*(B) = E_* \mathbf{1}\{B\}$;

(ii) there exists a measurable set $B^* \supset B$ *so that* $P(B^*) = P^*(B)$; *for any such* B^*, $\mathbf{1}\{B^*\} = (\mathbf{1}\{B\})^*$;

(iii) For $B_* \equiv \Omega - \{\Omega - B\}^*$, $P_*(B) = P(B_*)$;

(iv) $(\mathbf{1}\{B\})^* + (\mathbf{1}\{\Omega - B\})_* = 1$.

Proof of (iv):

- We have $(\mathbf{1}\{\Omega - B\})_* = (1 - \mathbf{1}\{B\})_*$ by complementarity.
- Then $(1 - \mathbf{1}\{B\})_* = -(\mathbf{1}\{B\} - 1)^*$ by definition of inner probability.
- Finally, for a constant c and an aribitrary map T, $(T + c)^* \leq T^* + c$ and, since $T^* \leq (T + c)^* - c$, we also have $T^* + c \leq (T + c)^*$; hence $(T + c)^* = T^* + c$, and thus

$$-(\mathbf{1}\{B\} - 1)^* = 1 - (\mathbf{1}\{B\})^*,$$

and the desired conclusion follows.$\square$

LEMMA 3. *Let $S, T : \Omega \mapsto \mathbb{R}$ be arbitrary maps; the following statements are true almost surely, provided the statements are well-defined:*

(i) $S_* + T^* \le (S+T)^* \le S^* + T^*$, *with all equalities if S is measurable;*

(ii) $S_* + T_* \le (S+T)_* \le S_* + T^*$, *with all equalities if T is measurable;*

(iii) $(S-T)^* \ge S^* - T^*$;

(iv) $|S^* - T^*| \le |S-T|^*$;

(v) $\left(\mathbf{1}\{T > c\}\right)^* = \mathbf{1}\{T^* > c\}$, *for any $c \in \mathbb{R}$;*

(vi) $\left(\mathbf{1}\{T \ge c\}\right)_* = \mathbf{1}\{T_* \ge c\}$, *for any $c \in \mathbb{R}$;*

(vii) $(S \vee T)^* = S^* \vee T^*$;

(viii) $(S \wedge T)^* \le S^* \wedge T^*$, *with equality if S is measurable.*

We next present an outer-expectation version of the unconditional Jensen's inequality (without proof):

LEMMA 4. **(Jensen's inequality)** *Let* $T : \Omega \mapsto \mathbb{R}$ *be an arbitrary map, with* $E^*|T| < \infty$, *and assume* $\phi : \mathbb{R} \mapsto \mathbb{R}$ *is convex. Then*

(i) $E^*\phi(T) \geq \phi(E_*T) \vee \phi(E^*T)$;

(ii) if ϕ *is also monotone,* $E_*\phi(T) \geq \phi(E_*T) \wedge \phi(E^*T)$.

The following outer-expectation version of Chebyshev's inequality is also useful:

LEMMA 5. **(Chebyshev's inequality)** *Let* $T : \Omega \mapsto \mathbb{R}$ *be an arbitrary map, with* $\phi : [0, \infty) \mapsto [0, \infty)$ *nondecreasing and strictly positive on* $(0, \infty)$*. Then, for every* $u > 0$*,*

$$P^* \left(|T| \geq u\right) \leq \frac{E^* \phi(|T|)}{\phi(u)}.$$

Proof:

- Since $|T| \geq u$ implies $\phi(|T|) \geq \phi(u)$,

$$(\mathbf{1}\{|T| \geq u\})^* \leq (\mathbf{1}\{\phi(|T|) \geq \phi(u)\})^* .$$

- Since $\mathbf{1}\left\{[\phi(|T|)]^* \geq \phi(u)\right\}$ is a measurable majorizor of $\mathbf{1}\left\{\phi(|T|) \geq \phi(u)\right\}$, we have

$$(\mathbf{1}\{\phi(|T|) \geq \phi(u)\})^* \leq \mathbf{1}\left\{[\phi(|T|)]^* \geq \phi(u)\right\},$$

  by definition of minimal measurable majorant.

- Puting these together, we have

$$(\mathbf{1}\{|T| \geq u\})^* \leq \mathbf{1}\left\{[\phi(|T|)]^* \geq \phi(u)\right\},$$

  and the desired conclusion follows from the standard Chebyshev inequality.$\square$

Perfect Maps

Consider composing a map $T : \Omega \mapsto \mathbb{R}$ with a measurable map $\phi : \tilde{\Omega} \mapsto \Omega$, defined on some probability space, to form

$$T \circ \phi : \left(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}\right) \mapsto \mathbb{R},$$

where $\phi : \left(\tilde{\Omega}, \tilde{\mathcal{A}}, \tilde{P}\right) \mapsto (\Omega, \mathcal{A}, P)$.

Denote T^* as the minimal measurable cover of T for $\tilde{P} \circ \phi^{-1}$.

It is not hard to see that

$$T^* \circ \phi \geq (T \circ \phi)^* \geq T \circ \phi.$$

The map ϕ is *perfect* if

$$(T \circ \phi)^* = T^* \circ \phi,$$

for every bounded $T : \Omega \mapsto \mathbb{R}$.

This property of perfectness ensures that

$$P^*(\phi \circ A) = \left(\tilde{P} \circ \phi^{-1}\right)^* (A)$$

for every set $A \subset \Omega$, since

$$\begin{aligned} P^*(\phi \circ A) &= \tilde{P}^*(\mathbf{1}\{A\} \circ \phi) = \tilde{P}(\mathbf{1}\{A\} \circ \phi)^* \\ &= \tilde{P}\left(\mathbf{1}\{A\}^* \circ \phi\right) = \left(\tilde{P} \circ \phi^{-1}\right)^* \mathbf{1}\{A\} \\ &= \left(\tilde{P} \circ \phi^{-1}\right)^* (A). \end{aligned}$$

Product Probability Spaces

An important example of a perfect map is a coordinate projection in a product probability space.

Specifically, let T be a real valued map defined on $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2, P_1 \times P_2)$ which only depends on the first coordinate of $\omega = (\omega_1, \omega_2)$.

T^* can then be computed by just ignoring Ω_2 and thinking of T as a map on Ω_1.

More precisely, suppose $T = T_1 \circ \pi_1$, where π_1 is the projection on the first coordinate.

The following lemma shows that $T^* = T_1^* \circ \pi_1$, and thus coordinate projections such as π_1 are perfect.

LEMMA 6. *A coordinate projection on a product probability space with product measure is perfect.*

We will see other examples of perfect maps later on in Chapter 7.

Now we consider Fubini's theorem for maps on product spaces which may not be measurable.

There is no generally satisfactory version of Fubini's theorem that will work in all nonmeasurable settings of interest, and it is frequently necessary to establish at least some kind of measurability to obtain certain key empirical process results.

The version of Fubini's theorem we now present basically states that repeated outer expectations are always less than joint outer expectations.

Let T be an arbitrary real map defined on the product space $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \times \mathcal{A}_2, P_1 \times P_2)$.

We write $E_1^* E_2^* T$ to mean outer expectations taken in turn.

For fixed ω_1, let $(E_2^* T)(\omega_1)$ be the infimum of $\int_{\Omega_2} U(\omega_2) dP_2(\omega_2)$ taken over all measurable $U : \Omega_2 \mapsto \bar{\mathbb{R}}$ with $U(\omega_2) \geq T(\omega_1, \omega_2)$ and for which $\int_{\Omega_2} U(\omega_2) dP_2(\omega_2)$ exists.

Next, $E_1^* E_2^* T$ is the outer integral of $E_2^* T : \Omega_1 \mapsto \mathbb{R}$.

Repeated inner expectations are analogously defined.

The following version of Fubini's theorem gives bounds for this repeated expectation process:

LEMMA 7. **(Fubini's theorem)** *Let T be an arbitrary real valued map on a product probability space; then $E_* T \leq E_{1*} E_{2*} T \leq E_1^* E_2^* T \leq E^* T$.*

Linear Operators

A *linear operator* is a map $T : \mathbb{D} \mapsto \mathbb{E}$ between normed spaces with the property that $T(ax + by) = aT(x) + bT(y)$ for all scalars a, b and any $x, y \in \mathbb{D}$.

When the range space $\mathbb{E}$ is $\mathbb{R}$, then T is a *linear functional*.

When T is linear, we will often use Tx instead of $T(x)$.

A linear operator $T : \mathbb{D} \mapsto \mathbb{E}$ is a *bounded linear operator* if

$$\|T\| \quad \equiv \quad \sup_{x \in \mathbb{D} : \|x\| \leq 1} \|Tx\| < \infty. \tag{1}$$

Here, the norms $\| \cdot \|$ are defined by the context.

We have the following proposition:

PROPOSITION 1. *For a linear operator $T : \mathbb{D} \mapsto \mathbb{E}$ between normed spaces, the following are equivalent:*

(i) T is continuous at a point $x_0 \in \mathbb{D}$;

(ii) T is continuous on all of $\mathbb{D}$;

(iii) T is bounded.

For normed spaces $\mathbb{D}$ and $\mathbb{E}$, let $B(\mathbb{D}, \mathbb{E})$ be the space of all bounded linear operators $T : \mathbb{D} \mapsto \mathbb{E}$.

This structure makes the space $B(\mathbb{D}, \mathbb{E})$ into a normed space with norm $\| \cdot \|$ defined in (1).

When $\mathbb{E}$ is a Banach space, then any convergent sequence $T_n x_n$ will be contained in $\mathbb{E}$, and thus $B(\mathbb{D}, \mathbb{E})$ is also a Banach space.

When $\mathbb{D}$ is not a Banach space, T has a unique continuous extension to $\overline{\mathbb{D}}$.

For normed spaces $\mathbb{D}$ and $\mathbb{E}$, and for any $T \in B(\mathbb{D}, \mathbb{E})$,

- $N(T) \equiv \{x \in \mathbb{D} : Tx = 0\}$ is the *null space* of T and
- $R(T) \equiv \{y \in \mathbb{E} : Tx = y \;\text{ for some } x \in \mathbb{D}\}$ is the *range space* of T.

It is clear that T is one-to-one if and only if $N(T) = \{0\}$.

We have the following two results for inverses, which we give without proof:

LEMMA 8. *Assume* $\mathbb{D}$ *and* $\mathbb{E}$ *are normed spaces and that* $T \in B(\mathbb{D}, \mathbb{E})$. *Then*

(i) T *has a continuous inverse* $T^{-1} : R(T) \mapsto \mathbb{D}$ *if and only if there exists a* $c > 0$ *so that* $\|Tx\| \geq c\|x\|$ *for all* $x \in \mathbb{D}$*;*

(ii) **(Banach's theorem)** *If* $\mathbb{D}$ *and* $\mathbb{E}$ *are complete and* T *is continuous with* $N(T) = \{0\}$*, then* T^{-1} *is continuous if and only if* $R(T)$ *is closed.*

A linear operator $T : \mathbb{D} \mapsto \mathbb{E}$ between normed spaces is a *compact operator* if $\overline{T(U)}$ is compact in $\overline{\mathbb{E}}$, where

$$U = \{x \in \mathbb{D} : \|x\| \leq 1\}$$

is the unit ball in $\mathbb{D}$ and, for a set $A \in \mathbb{D}$, $T(A) \equiv \{Tx : x \in A\}$.

The operator T is *onto* if for every $y \in \mathbb{E}$, there exists an $x \in \mathbb{D}$ so that $Tx = y$.

Later on in the class, we will encounter linear operators of the form $T + K$, where T is continuously invertible and onto and K is compact.

The following result will be useful:

LEMMA 9. *Let $A = T + K : \mathbb{D} \mapsto \mathbb{E}$ be a linear operator between Banach spaces, where T is both continuously invertible and onto and K is compact; then if $N(A) = \{0\}$, A is also continuously invertible and onto.*

The following simple inversion result for *contraction operators* is also useful.

An operator A is a contraction operator if $\|A\| < 1$.

PROPOSITION 2. *Let* $A : \mathbb{D} \mapsto \mathbb{D}$ *be a linear operator with* $\|A\| < 1$*; then* $I - A$*, where* I *is the identity, is continuously invertible and onto with inverse* $(I - A)^{-1} = \sum_{j=0}^{\infty} A^j$.

Proof:

- Let $B \equiv \sum_{j=0}^{\infty} A^j$, and note that

$$\|B\| \leq \sum_{j=0}^{\infty} \|A\|^j = (1 - \|A\|)^{-1} < \infty.$$

- Thus B is a bounded linear operator on $\mathbb{D}$.

- Since $(I - A)B = I$ by simple algebra, we have that $B = (I - A)^{-1}$, and the result follows.$\square$

Functional Differentiation

Let $\mathbb{D}$ and $\mathbb{E}$ be two normed spaces, and let $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto \mathbb{E}$ be a function.

We allow the domain $\mathbb{D}_\phi$ of the function to be an arbitrary subset of $\mathbb{D}$.

The function $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto \mathbb{E}$ is *Gâteaux-differentiable* at $\theta \in \mathbb{D}_\phi$, in the direction h, if there exists a quantity $\phi'_\theta(h) \in \mathbb{E}$ so that

$$\frac{\phi(\theta + t_n h) - \phi(\theta)}{t_n} \to \phi'_\theta(h),$$

as $n \to \infty$, for any scalar sequence $t_n \to 0$.

Gâteaux differentiability is usually not strong enough for the applications of functional derivatives needed for Z-estimators and the delta method.

The stronger differentiability we need is *Hadamard* and *Fréchet* differentiability.

A map $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto \mathbb{E}$ is Hadamard differentiable at $\theta \in \mathbb{D}_\phi$ if there exists a continuous linear operator $\phi'_\theta : \mathbb{D} \mapsto \mathbb{E}$ such that

$$\frac{\phi(\theta + t_n h_n) - \phi(\theta)}{t_n} \quad \to \quad \phi'_\theta(h), \tag{2}$$

as $n \to \infty$, for any scalar sequence $t_n \to 0$ and any $h, \{h_n\} \in \mathbb{D}$, with $h_n \to h$, and so that $\theta + t_n h_n \in \mathbb{D}_\phi$ for all n.

Hadamard differentiability is equivalent to *compact differentiability*, where compact differentiability satisfies

$$\sup_{h\in K,\theta+th\in\mathbb{D}_\phi}\left\|\frac{\phi(\theta+th)-\phi(\theta)}{t}-\phi'_\theta(h)\right\|\to 0,\quad \text{as } t\to 0, \tag{3}$$

for every compact $K\subset\mathbb{D}$.

Consider restricting the h values to be in a set $\mathbb{D}_0\subset\mathbb{D}$.

Specifically, if in (2) it is required that $h_n\to h$ only for $h\in\mathbb{D}_0\subset\mathbb{D}$, we say ϕ is *Hadamard-differentiable tangentially* to the set $\mathbb{D}_0$.

There appears to be no easy way to refine compact differentiability in an equivalent manner.

A map $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto \mathbb{E}$ is *Fréchet-differentiable* if there exists a continuous linear operator $\phi'_\theta : \mathbb{D} \mapsto \mathbb{E}$ so that (3) holds uniformly in h on bounded subsets of $\mathbb{D}$.

This is equivalent to $\|\phi(\theta + h) - \phi(\theta) - \phi'_\theta(h)\| = o(\|h\|)$, as $\|h\| \to 0$.

Since compact sets are bounded, Fréchet differentiability implies Hadamard differentiability.

Fréchet differentiability will be needed for Z-estimator theory, while Hadamard differentiability is useful in the delta method.

The following chain rule for Hadamard differentiability will also prove useful:

LEMMA 10. **(Chain rule)** *Assume* $\phi : \mathbb{D}_\phi \subset \mathbb{D} \mapsto \mathbb{E}_\psi \subset \mathbb{E}$ *is Hadamard differentiable at* $\theta \in \mathbb{D}_\phi$ *tangentially to* $\mathbb{D}_0 \subset \mathbb{D}$*, and* $\psi : \mathbb{E}_\psi \subset \mathbb{E} \mapsto \mathbb{F}$ *is Hadamard differentiable at* $\phi(\theta)$ *tangentially to* $\phi'_\theta(\mathbb{D}_0)$*.*

Then $\psi \circ \phi : \mathbb{D}_\phi \mapsto \mathbb{F}$ *is Hadamard differentiable at* θ *tangentially to* $\mathbb{D}_0$ *with derivative* $\psi'_{\phi(\theta)} \circ \phi'_\theta$*.*

Proof:

- First,

$$\psi \circ \phi(\theta + th_t) - \psi \circ \phi(\theta) = \psi(\phi(\theta) + tk_t) - \psi(\phi(\theta)),$$

  where

$$k_t = \frac{\phi(\theta + th_t) - \phi(\theta)}{t}.$$

- Note that if $h \in \mathbb{D}_0$, then

$$k_t \to k \equiv \phi'_\theta(h) \in \phi'_\theta(\mathbb{D}_0),$$

  as $t \to 0$, by the Hadamard differentiability of ϕ.

- Now,

$$\frac{\psi(\phi(\theta) + tk_t) - \psi(\phi(\theta))}{t} \to \psi'_{\phi(\theta)}(k)$$

  by the Hadamard differentiability of ψ.$\square$