A local surjection theorem

Dedicated to Yann Brenier, on his birthday (or whereabouts)

Ivar Ekeland, Eric Séré

CEREMADE, Université Paris-Dauhine

January 11, 2017

A question

Let Ω_1 and Ω_2 be metric spaces with their respective Borel σ-algebras. Let $u : \Omega_1 \to \Omega_2$ be some map. Consider the corresponding map $u^\#$ between probability spaces

$$\begin{aligned} u\# \quad &: \quad \mathcal{P}(\Omega_1) \to \mathcal{P}(\Omega_2) \\ \int_{\Omega_1} \varphi \; d(u\#\mu) \quad &= \quad \int_{\Omega_2} \varphi \circ u \; d\mu \end{aligned}$$

- Is $u^\#$ differentiable, and in what sense ?

A question

Let Ω_1 and Ω_2 be metric spaces with their respective Borel σ-algebras. Let $u : \Omega_1 \to \Omega_2$ be some map. Consider the corresponding map $u^{\#}$ between probability spaces

$$\begin{aligned} u\# &: \mathcal{P}(\Omega_1) \to \mathcal{P}(\Omega_2) \\ \int_{\Omega_1} \varphi \; d(u\#\mu) &= \int_{\Omega_2} \varphi \circ u \; d\mu \end{aligned}$$

- Is $u^{\#}$ differentiable, and in what sense ?
- Is $u^{\#}$ a (local) surjection ?

The statistician's view

Consider a D-dimensional random variable X on (Ω, P) and a K-dimensional map $u : R^D \to R^K$. The number:

$$T := \int_\Omega u\left(X\left(\omega\right)\right) dP\left(\omega\right)$$

is called a statistic . It depends only on the law $\mu := X^{\#}P$ of X. In fact,

$$T = \int_{R^D} u d\mu = \int_{R^K} I \; d\left(u^{\#}\mu\right) = \mathrm{bar}\left\{u^{\#}\mu\right\}$$

Examples:

the mean	$T = \int_\Omega X dP$	$u\left(x\right) = x$
the q-th quantile	$P\left(X \geq T\right) = q$	$u\left(x\right) = 1_{x \geq T}$
Mann-Whitney	$P\left[X_1 \geq X_2\right]$	$u\left(x_1, x_2\right) = 1_{x_1 \geq x_2}$

The central limit theorem

Unfortunately, X cannot be observed in general: all we have are empirical distributions $(x_1, ..., x_N)$. The relation with the true distribution is given by the CLT:

$$\mu_N \quad : \quad = \frac{1}{N} \sum_{n=1}^{N} \delta_{x_n} \text{ (a random distribution on } R^K\text{)}$$

$$\sqrt{N} \left(\mu_N - E[\mu] \right) \quad \rightsquigarrow \quad \mathcal{N}\left(0, \sigma^2\right) \text{ (convergence in law)}$$

von Mises calculus

- What happens to the law of large numbers when the statistic is nonlinear ?

von Mises calculus

- What happens to the law of large numbers when the statistic is nonlinear ?
- (von Mises, 1936) Suppose $u^{\#}$ is differentiable

$$
\begin{aligned}
T(\mu_N) &= T\left(\mu + \frac{1}{\sqrt{N}}\sqrt{N}(\mu_N - \mu)\right) \\
&\simeq T(\mu) + \frac{1}{\sqrt{N}} DT(\mu)\,\mathcal{N}(0, \sigma^2)
\end{aligned}
$$

von Mises calculus

- What happens to the law of large numbers when the statistic is nonlinear ?
- (von Mises, 1936) Suppose $u^{\#}$ is differentiable

$$
\begin{aligned}
T(\mu_N) &= T\left(\mu + \frac{1}{\sqrt{N}}\sqrt{N}(\mu_N - \mu)\right) \\
&\simeq T(\mu) + \frac{1}{\sqrt{N}} DT(\mu)\, \mathcal{N}(0, \sigma^2)
\end{aligned}
$$

- So $\sqrt{N}(T(\mu_N) - T(\mu))$ is asymptotically normal.

von Mises calculus

- What happens to the law of large numbers when the statistic is nonlinear ?
- (von Mises, 1936) Suppose $u^{\#}$ is differentiable

$$
\begin{aligned}
T\left(\mu_N\right) &= T\left(\mu + \frac{1}{\sqrt{N}}\sqrt{N}\left(\mu_N - \mu\right)\right) \\
&\simeq T\left(\mu\right) + \frac{1}{\sqrt{N}} DT\left(\mu\right)\mathcal{N}\left(0, \sigma^2\right)
\end{aligned}
$$

- So $\sqrt{N}\left(T\left(\mu_N\right) - T\left(\mu\right)\right)$ is asymptotically normal.
- For the q-quantile, and a law μ with density f, the correction term is computed to be $\mathcal{N}\left(0, \sigma_0^2\right)$ with:

$$\sigma_0 = \frac{q\left(1-q\right)}{f\left(T\right)^2}$$

The analyst's point of view

The problem is to validate this computation. Let us try it on a simple statistic:

$$T(X) = \int_0^1 u(X(t))\,dt$$

where u is C^∞ and satisfies $|u'(t)| \leq b$. Then $|u(t)| \leq a + bt$, and $T : L^1(dt) \to L^1(dt)$, so that $u^\# dt$ is absolutely continuous wrt Lebesgue measure. We have

$$DT(X)\,Y = \lim_{\varepsilon \to 0} \int_0^1 \frac{u(X + \varepsilon Y) - u(X)}{\varepsilon} dP$$

Using the mean value theorem and the estimate, we find:

$$DT(X)\,Y = \int_0^1 u'(X(t))\,Y(t)\,dt = \left\langle u'(X), Y \right\rangle_{\langle L^\infty,\ L^1\rangle}$$

Types of differentiablity

The nonlinear integral map $T : L^1 \rightarrow R$ given by:

$$T(X) = \int_0^1 u(X(t))\, dt$$

- is Gâteaux-differentiable (the restriction to any straight line is differentiable). This is NOT enough to ensure the validity of the von Mises calculus

Types of differentiablity

The nonlinear integral map $T : L^1 \rightarrow R$ given by:

$$T(X) = \int_0^1 u(X(t))\, dt$$

- is Gâteaux-differentiable (the restriction to any straight line is differentiable). This is NOT enough to ensure the validity of the von Mises calculus
- is NOT C^1, unless u is linear. Indeed, we have $DT(X) = u'(X) \in L^\infty$, and the map $X \rightarrow u'(X)$ from L^1 to L^∞ cannot be continuous, unless it is constant (Nemitsky, Krasnoselskii)

Types of differentiablity

The nonlinear integral map $T : L^1 \to R$ given by:

$$T(X) = \int_0^1 u(X(t))\, dt$$

- is Gâteaux-differentiable (the restriction to any straight line is differentiable). This is NOT enough to ensure the validity of the von Mises calculus
- is NOT C^1, unless u is linear. Indeed, we have $DT(X) = u'(X) \in L^\infty$, and the map $X \to u'(X)$ from L^1 to L^∞ cannot be continuous, unless it is constant (Nemitsky, Krasnoselskii)
- this is a general phenomenon . For instance, a map $X \to \int u(X)$ from L^2 into itself cannot be C^2, unless it is exactly quadratic. It follows that one cannot define directly a Morse theory for integral functionals, which causes technical difficulties in variational problems (Smale, IE)

Hadamard differentiability

Definition

Let $T : E \longrightarrow F$ be a G-differentiable map between Banach spaces. The following are equivalent:

- $DT(X)Y = \lim_{\varepsilon \to 0} \varepsilon^{-1}(T(X + \varepsilon Y) - T(X))$ exists uniformly for $Y \in K$, where K is a compact subset of F
- for every $Y(\varepsilon) \longrightarrow Y(0)$, we have $\lim_{\varepsilon \to 0,} \varepsilon^{-1}(T(X + \varepsilon Y(\varepsilon)) - T(X)) = DT(X)Y$
- for every C^1 curve $c : [0, 1] \longrightarrow X$, the curve $T \circ c : [0, 1] \longrightarrow X$ is differentiable

Definition

If these conditions are satisfied, T is called Hadamard-differentiable

Statistics are H-differentiable

$$T(X) = \int_0^1 u(X(t))\,dt \text{ defines } T : L^1 \to R$$

Theorem

If u is C^1 and $|u'| \leq a$, the map T is H-differentiable

Proof.

Consider a C^1 curve X_s in L^1, and differentiate:

$$s \to T(X_s) = \int_0^1 u(X_s(t))\,dt$$

Use Lebesgue's dominated convergence theorem to conclude □

Validation of von Mises

Theorem

The von Mises calculus is valid for the statistic T

Proof.

$$\begin{aligned}\frac{1}{N}\sum u(x_n) &= \int u(X) + \frac{1}{N}\sum\left(u(x_n) - \int u(X)\right) \\ &= \int u(X) + \frac{1}{N}\sum\left(\int (u(x_n) - u(X))\right) \\ &\simeq \int u(x) + \frac{1}{\sqrt{N}}\left(\int u'(X)\sqrt{N}\left(\frac{\sum x_n}{N} - X\right)\right)\end{aligned}$$

where we have used Prokhorov's theorem in the last ste □

Why local surjection theorems ?

Let $T : P(\Omega) \longrightarrow R^K$ be a statistic, with $T(\mu_0) = T_0$.

- does T cover a neighbourhood of T_0 ? (local surjection)

Why local surjection theorems ?

Let $T : P(\Omega) \longrightarrow R^K$ be a statistic, with $T(\mu_0) = T_0$.

- does T cover a neighbourhood of T_0 ? (local surjection)
- in that case, does there exist a continuous map $S : R^K \longrightarrow P(\Omega)$ such that $T \circ S = I$ (continuous right-inverse)

Why local surjection theorems ?

Let $T : P(\Omega) \longrightarrow R^K$ be a statistic, with $T(\mu_0) = T_0$.

- does T cover a neighbourhood of T_0 ? (local surjection)
- in that case, does there exist a continuous map $S : R^K \longrightarrow P(\Omega)$ such that $T \circ S = I$ (continuous right-inverse)
- no answer in the statistical literature

Why local surjection theorems ?

Let $T : P(\Omega) \to R^K$ be a statistic, with $T(\mu_0) = T_0$.

- does T cover a neighbourhood of T_0 ? (local surjection)
- in that case, does there exist a continuous map $S : R^K \to P(\Omega)$ such that $T \circ S = I$ (continuous right-inverse)
- no answer in the statistical literature
- in the analysis literature, no local surjection theorem for maps which are not C^1

The G-differentiable case

Let E and F be Banach spaces, and $\Phi : E \rightarrow F$ such that $\Phi(0) = 0$.
Denote by $B_E(R) \subset E$ the ball of radius R around 0

Theorem (IE)

Assume Φ is G-differentiable and continuous. Assume that $D\Phi(x)$ has right inverse $L(x)$ for all $x \in B_E(R)$. Assume moreover that the operator $L(x) : F \rightarrow E$ is uniformly bounded on $B_E(R)$:

$$\exists M > 0 : \sup \{ \| L(x) \| \mid \|x\| \leq R \} < M$$

Then $\Phi(B_E(R))$ covers $B_F\left(\frac{R}{M}\right)$

IE Annales IHP, Analyse non linéaire, 28 (2011) 91-105

Comparison with the standard local surjection theorem

- Φ is not required to be C^1

Comparison with the standard local surjection theorem

- Φ is not required to be C^1
- the range R/M is much larger than the standard range (Newton)

Comparison with the standard local surjection theorem

- Φ is not required to be C^1
- the range R/M is much larger than the standard range (Newton)
- no uniqueness

Comparison with the standard local surjection theorem

- Φ is not required to be C^1
- the range R/M is much larger than the standard range (Newton)
- no uniqueness
- An example:

$$
\begin{aligned}
\Phi &: \mathbb{C} \to \mathbb{C} \\
\Phi(x) &= y = x^n
\end{aligned}
$$

The H-differentiable case

Theorem (IE, Eric Séré)

Assume Φ *is H-differentiable, and that* $\Phi'(x)$ *has right inverse* $L(x)$ *for all* $x \in B_E(R)$ *.More precisely, there is a constant* $a < 1$ *and, for every* (x, y) *a positive radius* $\varepsilon(x, y)$ *such that, if* $\|x - x'\| \leq \varepsilon$*, then:*

$$\left\|D\Phi\left(x'\right) L(x)\, y - y\right\| \leq a \|y\|$$

Assume moreover that the operator $L(x) : F \to E$ *is uniformly bounded on* B_R*:*

$$\exists M > 0 : \sup\left\{\|L(x)\| \;\middle|\; \|x\| \leq R\right\} < M$$

Set $r := (1 - a) RM^{-1}$*. Then there is a continuous map* $\Psi : B_F(r) \to B_E(R)$ *such that:*

$$(\Phi \circ \Psi)(y) = y$$

Proof

Let $y \in B_F(r)$ be given. Choose some $M_0 < M$ and some a' with $a < a' < 1$ such that:

$$\sup \{\|L(x)\| \mid \|x\| \leq R\} \leq \frac{1-a'}{1-a} M_0 < M$$

Consider the set $\mathcal{C}(B_r, B_R)$ of all continuous maps $G : B_F(r) \to B_E(R)$ endowed with the uniform metric. It is a complete metric space. Consider the function $f : \mathcal{C}(B_F(r), B_E(R)) \to \mathbb{R}$ defined by:

$$f(G) = \sup \{\|(F \circ G)(y) - y\| \mid y \in B_F(r)\}$$

We have:

$$\begin{aligned} f(0) &= \|y\| \leq r \\ f &\geq 0 \end{aligned}$$

So we are in a position to apply Ekeland's variational principle.

There exists some $\bar{G}$ such that:

$$\begin{aligned}
f\left(\bar{G}\right) &\leq r \\
\left\|\bar{G}\right\| &\leq R\frac{M_0}{M} < R \\
\forall G,\ f\left(G\right) &\geq f\left(\bar{G}\right) - \frac{r}{R}\frac{M}{M_0}\left\|G-\bar{G}\right\|
\end{aligned}$$

We have $\frac{r}{R}M = 1-a$. If $\left(F\circ\bar{G}\right)\left(y\right) = y$ the proof is over. If not, we will move $\left(F\circ\bar{G}\right)\left(y\right)$ in the direction of y. Set:

$$\begin{aligned}
v\left(y\right) &= y - \left(F\circ\bar{G}\right)\left(y\right) \\
G_t\left(y\right) &= \bar{G}\left(y\right) + tL\left(\bar{G}\left(y\right)\right)v\left(y\right)
\end{aligned}$$

Getting first-order information

$$
\begin{aligned}
\forall t,\ f\left(G_t\right) &\geq f\left(\bar{G}\right) - \frac{1-a}{M_0}\left\|G_t - \bar{G}\right\| \\
\forall t,\ \frac{f\left(G_t\right) - f\left(\bar{G}\right)}{t} &\geq -\frac{1-a}{M_0}\left\|L\left(\bar{G}\left(y\right)\right)v\left(y\right)\right\| \\
\left|\left(Df\left(\bar{G}\right), L\left(\bar{G}\left(y\right)\right)v\left(y\right)\right)\right| &\leq \frac{1-a}{M_0}\left\|L\left(\bar{G}\left(y\right)\right)v\left(y\right)\right\|
\end{aligned}
$$

Computing the derivative

$$f\left(G_t\right) = \sup_{y \in B_F(r)} \left\| \left(F \circ G\right)(y) - y \right\|$$

By the envelope theorem:

$$\begin{aligned} \left(Df\left(\bar{G}\right), g\right) &= \left(DF\left(\bar{G}\left(y\right)\right) g\left(\bar{y}\right), \frac{F\left(\bar{G}\left(\bar{y}\right)\right) - \bar{y}}{\left\|F\left(\bar{G}\left(\bar{y}\right)\right) - \bar{y}\right\|}\right) \\ F\left(\bar{G}\left(y\right)\right) - \bar{y} &= \max_{y \in B_F(r)} \left\{F\left(\bar{G}\left(y\right)\right) - y\right\} \end{aligned}$$

The final contradiction

$$\left|\left(Df\left(\bar{G}\right), L\left(\bar{G}\left(y\right)\right) v\left(y\right)\right)\right| \leq \frac{1-a}{M_0} \left\| L\left(\bar{G}\left(y\right)\right) v\left(y\right)\right\|$$

$$\left|\left(DF\left(\bar{G}\left(y\right)\right) L\left(\bar{G}\left(y\right)\right) v\left(\bar{y}\right), \frac{F\left(\bar{G}\left(\bar{y}\right)\right)-\bar{y}}{\left\|F\left(\bar{G}\left(\bar{y}\right)\right)-\bar{y}\right\|}\right)\right| \leq \frac{1-a}{M_0} \left\| L\left(\bar{G}\left(y\right)\right) v$$

$$\left|\left(DF\left(\bar{G}\left(y\right)\right) L\left(\bar{G}\left(y\right)\right) v\left(\bar{y}\right), \frac{v\left(\bar{y}\right)}{\left\|v\left(\bar{y}\right)\right\|}\right)\right| \leq \frac{1-a}{M_0} \left\| L\left(\bar{G}\left(y\right)\right) v\left(y\right)\right\|$$

$$\left\|v\left(\bar{y}\right)\right\| \leq \frac{1-a}{M_0} M \left\|v\left(\bar{y}\right)\right\| < \left|v\left(\bar{y}\right)\right|$$

Advice from a fellow traveller:

The journey, not the arrival, matters