On the efficient influence function and the efficient score function

Andries Lenstra
department of econometrics
Vrije Universiteit, Amsterdam[1]
31 January, 1997

Efficient influence functions and efficient score functions are representations of derivatives of maps. The invariance of the former under changes of parametrization, and the formula relating the efficient influence function for the projection of a parameter to the efficient score function, turn out to be immediate results of diagram chasing, rather than to follow from formal calculations.

1 Introduction: matrices and maps

Let H be a real Hilbert space. For the sake of freedom of movement, we will extend matrix multiplication to Hilbert matrices, i.e. matrices of which all entries are members of H. To this end we define

$$\begin{aligned} \alpha\cdot\beta &:= \alpha\beta, \text{ if } \alpha,\, \beta \in \mathbb{R}, \\ &:= \alpha\beta, \text{ if } \alpha \in \mathbb{R},\, \beta \in H, \\ &:= \beta\alpha, \text{ if } \alpha \in H,\, \beta \in \mathbb{R}, \\ &:= \langle \alpha, \beta \rangle, \text{ if } \alpha,\, \beta \in H, \end{aligned}$$

and

$$A\cdot B := \left(\sum_{i=1}^{m} \alpha_{hi}\cdot\beta_{ij} \right)_{\substack{1\le h\le l \\ 1\le j\le n}}$$

for matrices $A = (\alpha_{hi})_{\substack{1\le h\le l \\ 1\le i\le m}}$ and $B = (\beta_{ij})_{\substack{1\le i\le m \\ 1\le j\le n}}$. One will check:

[1] This research was supported by The Netherlands Foundation for Scientific Research (NWO). Typesetting: TEX by Donald Knuth, diagrams: Paul Taylor.

1.1 Theorem. *This matrix multiplication is associative as soon as the number of Hilbert matrices is less than or equal to* 2, or '$\sharp H \leq 2$', i.e.

$$A \cdot (B \cdot C) = (A \cdot B) \cdot C$$

if not all matrices A, B, or C are Hilbert matrices. □

To every matrix, maps will be associated via

$$A(v) := A \cdot v$$

for compatible matrices A and v. The matrix A is said to **represent** the map $v \mapsto A \cdot v$. In particular, for $v_0, v_1, \ldots, v_n \in H$ one has

$$\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} (v_0) = \begin{pmatrix} \langle v_1, v_0 \rangle \\ \vdots \\ \langle v_n, v_0 \rangle \end{pmatrix},$$

if we identify single entry matrices with the entry—as we always shall.

The next five theorems give some aspects of the relations between maps and matrices. With 1.1 we have:

1.2 Theorem. *If* $\sharp H \leq 2$, *then matrix multiplication corresponds to composition of maps*, because, if A, B, and v are compatible and not all Hilbert, then

$$\begin{aligned} (A \cdot B)(v) &= (A \cdot B) \cdot v \\ &\overset{\sharp H \leq 2}{=} A \cdot (B \cdot v) \\ &= A\big(B(v)\big) \\ &= (A \circ B)(v), \end{aligned}$$

so that

$$A \cdot B \overset{\text{[as a map]}}{=} A \circ B$$

for every suitable (relatively to A and B) domain of map B ('as a map' is redundant, as the right side is not a priori a matrix). □

Let H^n denote the set of all $n \times 1$ matrices (or **vectors**) of which the entries, or components, are members (also 'vectors') from H.

1.3 Theorem. *If two vectors in* H^n *represent the same map of* H *into* $\mathbb{R}^n$, *then these vectors are equal.*

Proof. The difference of the maps applied to one component, $v_i - w_i$ say, of the difference of the vectors gives $\langle v_i - w_i, v_i - w_i \rangle = 0$, so $v_i = w_i$. □

Note that this is not necessarily true for vectors in H^n and maps of a subspace of H into $\mathbb{R}^n$.

The following theorem is only invoked once, and then for sheer esthetical reasons.

1.4 Theorem. *Every bounded linear map of H into $\mathbb{R}^n$ has a unique representation by a vector from H^n.*

Proof. For each of the projections $\mathbb{R}^n \to \mathbb{R}$, the composition $H \to \mathbb{R}^n \to \mathbb{R}$ is equal to $v_0 \mapsto \langle v_i, v_0 \rangle$ for a $v_i \in H$ by the Riesz-Fréchet representation theorem. The unicity follows from 1.3. □

Vice versa, every linear map of $\mathbb{R}^n$ into H has a unique matrix representation by the transpose v^T of a vector v from H^n (take $v_i :=$ the image of the i^{th} unit vector in $\mathbb{R}^n$). The set H^n can be equipped with a norm: $\|(v_1, \ldots, v_n)^T\| := \sqrt{\sum_{i=1}^n \|v_i\|^2}$; with this norm it is a Banach space. For $r \in \mathbb{R}^n$ we have

$$\|v^T \cdot r\|_H \leq \|r\|_{\mathbb{R}^n} \|v\|_{H^n} < \infty,$$

so every linear map of $\mathbb{R}^n$ into H is bounded. We even have:

1.5 Theorem. *The map $f \mapsto v$, v^T the matrix representation of f, is a linear homeomorphism of the Banach space $L(\mathbb{R}^n, H)$ of all bounded linear maps $\mathbb{R}^n \to H$ and the Banach space H^n.*

Proof. The linearity and bijectivity being clear, we only have to check boundedness. For $f \mapsto v$ this follows from: if $\|f\|_{L(\mathbb{R}^n,H)} = 1$, then $\sup_{\|r\|=1} \|f(r)\| = \sup_{\|r\|=1} \|v^T \cdot r\| = 1$, so $\|v_i\| \leq 1$ for $i = 1, 2, \ldots, n$ and $\|v\| \leq \sqrt{n}$. For the boundedness of $v \mapsto f$, the closed graph theorem would suffice—or a look at the inequality above the theorem. □

Let $v \in H^n$. Then for the **Gram matrix** $v \cdot v^T$ we find:

1.6 Theorem.

$$\begin{aligned} v \cdot v^T \textit{ invertible } &\Leftrightarrow v \cdot v^T \textit{ positive definite} \\ &\Leftrightarrow v^T : \mathbb{R}^n \to H \textit{ injective} \end{aligned}$$

Proof. If for all $a \in \mathbb{R}^n$, $a \neq (0 \cdots 0)^T$, one has $(v \cdot v^T) \cdot a \neq (0 \cdots 0)^T \in \mathbb{R}^n$, then for the same a one has $v^T \cdot a \neq 0 \in H$, as $(v \cdot v^T) \cdot a \overset{\sharp H \leq 2}{=} v \cdot (v^T \cdot a)$. Then, from $v^T \cdot a \neq 0$ it follows that $\|v^T \cdot a\|^2 = (a^T \cdot v) \cdot (v^T \cdot a) \overset{\sharp H \leq 2}{=} a^T \cdot \big((v \cdot v^T) \cdot a\big) > 0$, which in turn implies that $(v \cdot v^T) \cdot a \neq (0 \cdots 0)^T \in \mathbb{R}^n$. □

Let $[v]$ denote the linear span of the components of v. It is a Hilbert space, because it is a finite dimensional, and therefore closed subspace of H.

1.7 Theorem. *If $v^T : \mathbb{R}^n \to H$ is injective, then the inverse $(v^T)^{-1} : [v] \to \mathbb{R}^n$ on the image $[v]$ of v^T is represented by the vector $(v \cdot v^T)^{-1} \cdot v \in ([v])^n$.*

Proof. We only have to check that $(v \cdot v^T)^{-1} \cdot v$ is a left inverse of v^T:

$$\begin{aligned} \big((v \cdot v^T)^{-1} \cdot v\big) \circ v^T &\overset{\text{argum.s real}}{=} \big((v \cdot v^T)^{-1} \cdot v\big) \cdot v^T \\ &\overset{\sharp H \leq 2}{=} (v \cdot v^T)^{-1} \cdot (v \cdot v^T), \end{aligned}$$

i.e. the map of $\mathbb{R}^n$ into itself on the left side is the same as the map of $\mathbb{R}^n$ into itself represented by the matrix $(v \cdot v^T)^{-1} \cdot (v \cdot v^T)$; the latter map is $\mathrm{id}_{\mathbb{R}^n}$. □

2 Introduction: differentiation

Let X and Y be Banach spaces and U an open set in X. The map $f : U \to Y$ is **differentiable** at a point $x \in U$, if there exists a bounded linear map $f'(x) : X \to Y$, the **derivative of** f **at** x, such that

$$f(x+h) - f(x) = \big(f'(x)\big)(h) + o(\|h\|) \quad \|h\| \to 0.$$

Note that, if it exists, $f'(x)$ is unique. One will check that an equivalent formulation is the following:

$$\begin{aligned} &\Big\|\frac{f(x + \varepsilon_n h_n) - f(x)}{\varepsilon_n} - \big(f'(x)\big)(h_n)\Big\|_Y \to 0 \quad n \to 0 \\ &\text{for all } (\varepsilon_n)_{n=1}^{\infty} \text{ in } \mathbb{R} \setminus \{0\} \text{ with } \varepsilon_n \to 0 \\ &\text{and all bounded } (h_n)_{n=1}^{\infty} \text{ in } X. \end{aligned}$$

Now let V be an arbitrary set in X, not necessarily open, and $x \in V$. Then for every continuous map $\gamma : (-1,1) \to V \subset X$ with $0 \mapsto x$ that is differentiable at 0 (a **curve**), the representation $\big(\gamma'(0)\big)(1)$ of the derivative at 0 is a member of X; it is called a **tangent vector of** V **at** x. The closed linear span of all tangent vectors of V at x is the **tangent space of** V **at** x, notation $\dot{V}$. So if V is open, its tangent space at x is the whole space X.

Let $f: V \to Y$ be a map of V into Y and suppose there exists a bounded linear map $f'(x): \dot{V} \to Y$ such that

$$\|\frac{f(x+\varepsilon_n h_n) - f(x)}{\varepsilon_n} - \big(f'(x)\big)(h_0)\|_Y \to 0 \quad n \to 0$$

for all $(\varepsilon_n)_{n=1}^{\infty}$ in $\mathbb{R} \setminus \{0\}$, $h_0 \in \dot{V}$, and $(h_n)_{n=1}^{\infty}$ in X with $\varepsilon_n \to 0$, $h_n \to h_0$ and $x + \varepsilon_n h_n \in V$.

Then $f'(x)$ *is the only bounded linear map of* $\dot{V}$ *into* Y *with this property*, because for all tangent vectors $h_0 \in \dot{V}$ there are sequences $(\varepsilon_n)_{n=1}^{\infty}$ and $(h_n)_{n=1}^{\infty}$ as above. Namely, if $h_0 = \big(\gamma'(0)\big)(1)$, then one could take $\varepsilon_n := 1/n$, $h_n := n(\gamma(1/n) - x)$. The map f is said to be **differentiable at** x **along the tangent space** and $f'(x)$ is the **derivative of** f **at** x **along** $\dot{V}$.

Two observations connect the two kinds of differentiability that we have exposed here. The first is obvious from the equivalence of ‘bounded’ to ‘continuous’ for linear maps, and the boundedness of convergent sequences.

2.1 Theorem. *Let* X *and* Y *be Banach spaces and* U *open in* X, $V \subset U$. *Let* $x \in V$. *If* $f: U \to Y$ *is differentiable at* x, *then* $f|_V: V \to Y$ *is differentiable at* x *along the tangent space* $\dot{V}$ *of* V *at* x *with derivative* $(f|_V)'(x) = f'(x)|_{\dot{V}}$. □

In particular, differentiability at x implies differentiability at x along X. On the other hand, as to the convergence of bounded sequences: in $\mathbb{R}^n$ every bounded sequence has a convergent subsequence. These facts lead to the second observation:

2.2 Theorem. *For* $X = \mathbb{R}^n$ *differentiability at* x *and differentiability at* x *along* X *are equivalent.* □

The first differentiability is often referred to as **Fréchet** differentiability, differentiability along X is known as **Hadamard** differentiability. Differentiability along a tangent space is strong enough to support the chain rule; we formulate it and leave the proof as a straightforward check.

2.3 Theorem (chain rule). *Let X, Y and Z be Banach spaces and $V \subset X$, $W \subset Y$. If $f\colon V \to W$ is differentiable at $x \in V$ along the tangent space $\dot V$ of V at x, $g\colon W \to Z$ is differentiable at $f(x)$ along the tangent space $\dot W$ of W at $f(x)$, and $\big(f'(x)\big)(\dot V) \subset \dot W$, then $g \circ f$ is differentiable at x along $\dot V$ and $(g \circ f)'(x) = g'\big(f(x)\big) \circ f'(x)$.* □

Take $f : 0 \mapsto 0$, $f := \sum_{k=1}^{\infty} \frac{1}{k} 1_{[\frac{1}{k+1}, \frac{1}{k})}$ on $(0,1)$, and $f(x) = -f(-x)$ for $x \in (-1,0)$, $W := \{\frac{1}{k} : k \in \mathbb{Z}\} \cup \{0\}$, and $g\colon w \mapsto |w|$ in order to see why we imposed the condition $\big(f'(x)\big)(\dot V) \subset \dot W$. For a continuous $f\colon V \to W$, the chain rule and 2.2 make sure that the image $\big(f'(x)\big)\Big(\big(\gamma'(0)\big)(1)\Big)$ of a tangent vector in $\dot V$ is the tangent vector $\big((f \circ \gamma)'(0)\big)(1)$ of the curve $f \circ \gamma$. We see:

2.4 Theorem. *The condition $\big(f'(x)\big)(\dot V) \subset \dot W$ always holds if $f\colon V \to W$ is continuous.* □

In the sequel, every time we come across a derivative along a tangent space, we could before, if we wished, have come across an ordinary derivative of which the former is the restriction to the tangent space. So the notion of differentiating along a tangent space, and the corresponding chain rule, for that matter, could have been avoided there altogether. Not to be avoided, then, would be the use of the following theorem, which is obvious from 2.1 and the unicity of derivatives along a tangent space.

2.5 Theorem. *Let X and Y be Banach spaces and U open in X, $V \subset U$. Let $x \in V$. If $f\colon U \to Y$ and $g\colon U \to Y$ are differentiable at x and $f|_V = g|_V$, then $f'(x)|_{\dot V} = g'(x)|_{\dot V}$.* □

If $f\colon U \to Y$ is differentiable at every point of U, then $x \in U \mapsto f'(x)$ defines a map of U into the Banach space $L(X,Y)$ of all bounded linear maps $X \to Y$. For the case $X = \mathbb{R}^n$ and $Y = H$, 1.5 says this map is continuous precisely if the corresponding matrix valued map is continuous.

2.6 Theorem (mean value). *Let X and Y be Banach spaces, U open in X, and $x_0 \in U$. Let $f\colon U \to Y$ be differentiable at every point of U. Then for every segment $[x_0, z_0] := \{x_0 + t(z_0 - x_0) : t \in [0,1]\}$ in U with $z_0 \neq x_0$ we have*

$$\frac{\|f(z_0) - f(x_0)\|}{\|z_0 - x_0\|} \leq \sup_{\xi \in [x_0, z_0]} \|f'(\xi)\|_{L(X,Y)}.$$

Proof. For every segment $[x,z]$ in U with $z \neq x$ and $y \in (x,z)$ we have: if both the difference quotients $\|f(x)-f(y)\|/\|x-y\|$ and $\|f(y)-f(z)\|/\|y-z\|$ are smaller than s, then also

$$\frac{\|f(x)-f(z)\|}{\|x-z\|} \leq \frac{\|x-y\|}{\|x-z\|}\frac{\|f(x)-f(y)\|}{\|x-y\|} + \frac{\|y-z\|}{\|x-z\|}\frac{\|f(y)-f(z)\|}{\|y-z\|} < s$$

by virtue of $\|x-y\| + \|y-z\| = \|x-z\|$. This argument shows that, if the required inequality would not hold, i.e. if there were an $\varepsilon > 0$ with $\frac{\|f(z_0)-f(x_0)\|}{\|z_0-x_0\|} = \sup_{\xi\in[x_0,z_0]} \|f'(\xi)\|_{L(X,Y)} + \varepsilon$, then there is a sequence of segments such that each segment is one of the two halves of the former segment and has a difference quotient which is greater than or equal to $\sup_{\xi\in[x_0,z_0]} \|f'(\xi)\|_{L(X,Y)} + \varepsilon$. Let y_0 belong to all these segments. From the definition of the differentiability of f at y_0, then, it follows that (if $y_0 \in (x_0,z_0)$; else this argument needs a slight but obvious change) there is a subsegment $[y_0-h_0, y_0+h_0] \subset [x_0,z_0]$, $h_0 \neq 0$, such that

$$\frac{\|f(y_0+h)-f(y_0)\|}{\|h\|} < \|f'(y_0)\| + \frac{\varepsilon}{2}$$

for all h with $y_0+h \in [y_0-h_0, y_0+h_0]$. In particular, one could take h_1, h_2, not both equal to 0, such that $[y_0-h_1, y_0+h_2] \subset [y_0-h_0, y_0+h_0]$, and $[y_0-h_1, y_0+h_2]$ belongs to the above sequence, in order to obtain a contradiction, from the initial argument if $h_1, h_2 \neq 0$. □

Finally, we will need a statement concerning tangent spaces of subsets:

2.7 Theorem. *Let X be a Banach space and $x \in O \subset V \subset X$. If the set O is open in V, then the tangent space of O at x is equal to the tangent space of V at x.*

Proof. If $\gamma\colon(-1,1) \to V$, $\gamma\colon 0 \mapsto x$, is a curve in V, then the openness of O in V guarantees the existence of a $\delta > 0$ with $\gamma\big((-\delta,\delta)\big) \subset O$. Let α be one of the many continuous maps $(-1,1) \to (-\delta,\delta)$ with $\alpha|_{(-\delta/2,\delta/2)} = \mathrm{id}_{(-\delta/2,\delta/2)}$. Then $\gamma\circ\alpha$ is a curve in O, so $\big((\gamma\circ\alpha)'(0)\big)(1)$ is a tangent vector of O at x, while $\big((\gamma\circ\alpha)'(0)\big)(1) = \big(\gamma'(0)\big)(1)$. □

3 Introduction: the inverse mapping theorem

The derivative reflects the local behaviour of the map. Bijectivity of the derivative implies—under a continuity condition—local bijectivity of the map (the 'inverse mapping theorem') and injectivity of the derivative implies—if one adds a condition on its range—local injectivity of the map. Of the last truth, the next theorem is a version that is adapted to our needs; the extra condition has disappeared behind the finite dimension of the domain space $\mathbb{R}^n$.

3.1 Theorem. *Let H be a Hilbert space, U open in $\mathbb{R}^n$, and $x_0 \in U$. Let $f: U \to H$ be differentiable at every point of U and let $x \in U \mapsto f'(x)$ be continuous at x_0, while $f'(x_0)$ is injective. Then there is an open set $U_1 \subset U$, $x_0 \in U_1$, such that $f|_{U_1}: U_1 \to f(U_1)$ is a homeomorphism of which the inverse $(f|_{U_1})^{-1}: f(U_1) \to U_1$ is differentiable at every point of $f(U_1)$ along the tangent space of $f(U_1)$ at the point.*

Proof. We follow [Lang], p. 16. Let $f'(x_0)$ be represented by v^T for $v \in H^n$, so that the image $\big(f'(x_0)\big)(\mathbb{R}^n)$ is equal to $[v]$. The closed graph theorem (or inspection: $(\lambda_1, \ldots, \lambda_n)^T \in \mathbb{R}^n \mapsto \| \sum \lambda_i v_i \|^{-1}$ is defined and continuous on the unit ball in $\mathbb{R}^n$, hence it has a maximum) tells us that $f'(x_0)$ has a bounded inverse on $[v]$, so that it is a linear homeomorphism of $\mathbb{R}^n$ and $[v]$. By the continuity of the projections $H \to [v]$ and $H \to [v]^\perp$, the map $(u, u_\perp) \in [v] \times [v]^\perp \mapsto u + u_\perp \in H$ has a bounded inverse and is a linear homeomorphism, too.

Let $\beta: H \to [v] \times [v]^\perp$ be the composition of the inverse ι of this map, and a translation of $[v] \times [v]^\perp$, such that $\beta: f(x_0) \mapsto (0,0) \in [v] \times [v]^\perp$. Let $\alpha: [v] \to \mathbb{R}^n$ be the composition of a translation of $[v]$, and $\big(f'(x_0)\big)^{-1}$, such that $\alpha: 0 \in [v] \mapsto x_0 \in \mathbb{R}^n$. Let $\tilde{U} := \alpha^{-1}(U)$ and $\tilde{f}: \tilde{U} \subset [v] \to [v] \times [v]^\perp$ be the composition $\tilde{f} := \beta \circ f \circ \alpha$. Then $\tilde{U}$ is open and $\tilde{f}$ is differentiable at every point of $\tilde{U}$, while $u \in \tilde{U} \mapsto \tilde{f}'(u)$ is continuous at 0 (by the Chain Rule $\tilde{f}'(u) = \iota \circ f'\big(\alpha(u)\big) \circ \big(f'(x_0)\big)^{-1}$; further, $(f,g) \in L(X,Y) \times L(Y,Z) \mapsto g \circ f \in L(X,Z)$ is continuous for Banach spaces X, Y, and Z), the derivative $\tilde{f}'(0)$ at 0 is the map $u \in [v] \mapsto (u,0) \in [v] \times [v]^\perp$, and $\tilde{f}: 0 \mapsto (0,0)$.

Now define $\phi: \tilde{U} \times [v]^\perp \to [v] \times [v]^\perp$ by

$$\phi: (u, u_\perp) \in \tilde{U} \times [v]^\perp \mapsto \tilde{f}(u) + (0, u_\perp) \in [v] \times [v]^\perp.$$

Then $\phi(u,0) = \tilde{f}(u)$ and $\phi(0,0) = (0,0)$; moreover it is easily seen that ϕ is everywhere differentiable, with

$$\phi'(u, u_\perp): (h_1, h_2) \in [v] \times [v]^\perp \mapsto \big(\tilde{f}'(u)\big)(h_1) + (0, h_2) \in [v] \times [v]^\perp,$$

so that $\phi'(0,0)$ is equal to $\mathrm{id}_{[v] \times [v]^\perp}$ and $(u, u_\perp) \mapsto \phi'(u, u_\perp)$ is continuous at $(0,0)$ (inspection gives $\|\phi'(u, u_\perp) - \phi'(0,0)\| = \|\tilde{f}'(u) - \tilde{f}'(0)\|$).

All this means that f has been transformed into a map (ϕ) to which Lang's proof of the Inverse Mapping Theorem can be applied (cf. [Lang], p. 13)—at least, if our continuity condition, i.e. continuity at $(0,0)$, is as good as Lang's continuity condition, i.e. continuity everywhere. In Lang's proof, continuity is used more often than a superficial look would reveal; the only occasion, however, where the continuity at points different from 0 is involved, is the application of the Mean Value Theorem, for the proof of which Lang needs continuity everywhere. As 2.6 has demonstrated, for that theorem the mere existence of the derivative suffices.

The Inverse Mapping Theorem now guarantees the existence of open sets $\tilde{U}_1 \subset \tilde{U}$ and $\tilde{U}_2 \subset [v]^\perp$ such that $(0,0) \in \tilde{U}_1 \times \tilde{U}_2$, the image $\phi(\tilde{U}_1 \times \tilde{U}_2)$ is open, the map ϕ is invertible on $\phi(\tilde{U}_1 \times \tilde{U}_2)$, and its inverse ϕ^{-1} is differentiable at all points of $\phi(\tilde{U}_1 \times \tilde{U}_2)$. Let $\tilde{f}^{-1} := \big((u,0) \mapsto u\big) \circ \phi^{-1}|_{\phi(\tilde{U}_1 \times 0)}$; this map is continuous on $\phi(\tilde{U}_1 \times 0) = \tilde{f}(\tilde{U}_1)$, with the property

$$\tilde{f}^{-1} \circ \tilde{f}(u) = \tilde{f}^{-1}\big(\phi(u,0)\big) = u \quad \text{for all } u \in \tilde{U}_1.$$

So $\tilde{f}|_{\tilde{U}_1}$ is a homeomorphism indeed; moreover, by 2.1, 2.3, and 2.4, its inverse $\tilde{f}^{-1}$ is differentiable at every point of $\tilde{f}(\tilde{U}_1)$ along the tangent space of $\tilde{f}(\tilde{U}_1)$ along the point. Then the identity $f = \beta^{-1} \circ \tilde{f} \circ \alpha^{-1}$, together with 2.1, 2.3, and 2.4, implies that $\alpha(\tilde{U}_1)$ fulfils the requirements of the set U_1. □

Of the next result, the surprise is the proof, which does not exploit the continuity of $(f|_{U_1})^{-1}$. In particular, no curves are lifted from $f(U_1)$ to U_1.

3.2 (continuation). *The tangent space T of $f(U_1)$ at $f(x_0)$ is equal to $\big(f'(x_0)\big)(\mathbb{R}^n)$.*

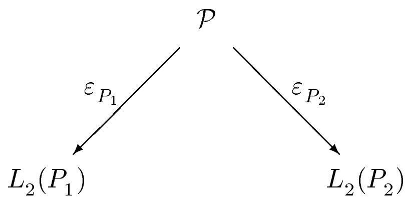
Proof. By 2.1 or 2.2, 2.4, and the continuity of f, we only have to ascertain that T lies within $\big(f'(x_0)\big)(\mathbb{R}^n)$. As $f|_{U_1} \circ (f|_{U_1})^{-1} = \mathrm{id}_{f(U_1)}$, and the 2.4-condition is already fulfilled by $\dot{W}$ being maximal, namely $\mathbb{R}^n$, 2.3 gives $f'(x_0) \circ \big((f|_{U_1})^{-1}\big)'\big(f(x_0)\big) = \mathrm{id}_T$. □

Now 2.3 and 1.7 give:

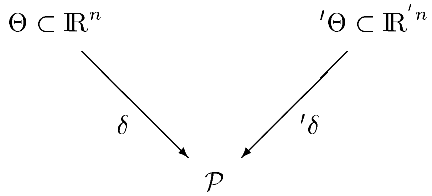
3.3 (end). *If v^T represents $f'(x_0)$, then the derivative of $(f|_{U_1})^{-1}$ at $f(x_0)$ along the tangent space of $f(U_1)$ is represented by $(v \cdot v^T)^{-1} \cdot v$.* □

4 The efficient influence function

In statistics one tries to deduce properties of lotteries on the base of outcomes of the latter. To this end, a lottery is seen as an unknown member P of a set $\mathcal{P}$ of probability distributions on some probability space, and the desired property, or **parameter**, as the value in P of a map $\nu\colon \mathcal{P} \to \mathbb{R}^m$. Structures on $\mathcal{P}$ may help to express one's beliefs or to facilitate inferencing. A common belief concerning $\mathcal{P}$ implies the existence of a family $(\varepsilon_P)_{P\in\mathcal{P}}$ of **local embeddings**



i.e. maps of $\mathcal{P}$ to the Hilbert spaces $L_2(P)$ with $\varepsilon_P(P) = 0 \in L_2(P)$ for all $P \in \mathcal{P}$, and of $(\varepsilon_P)_{P\in\mathcal{P}}$-compatible **parametrizations**, i.e. bijective maps



from open sets $\Theta \subset \mathbb{R}^n$ to $\mathcal{P}$ that co-operate with $(\varepsilon_P)_{p\in\mathcal{P}}$ in the following way: for every $P \in \mathcal{P}$

(i) there is a sequence of open balls around $\delta^{-1}(P)$ whose radii decrease to zero, such that every open ball has an open image in $\varepsilon_P(\mathcal{P}) \subset L_2(P)$ under the composition $\varepsilon_P \circ \delta$, and

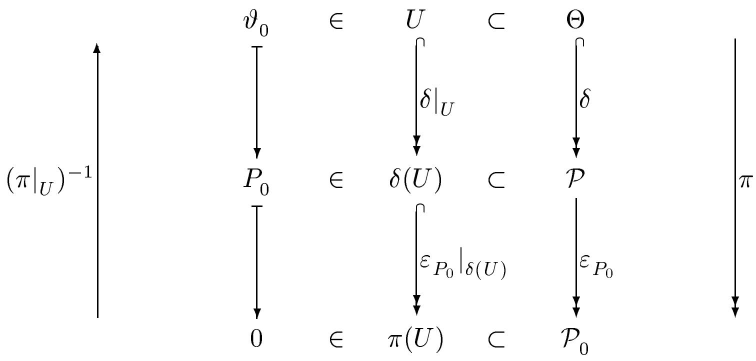
(ii) this composition $\varepsilon_P \circ \delta$ is differentiable at every point of such an open ball, while

(iii) $\vartheta \mapsto (\varepsilon_P \circ \delta)'(\vartheta)$ is continuous at $\delta^{-1}(P)$, and

(iv) at this point the derivative $(\varepsilon_P \circ \delta)'\big(\delta^{-1}(P)\big)$ is injective.

Let $\mathcal{P}$ have local embeddings $(\varepsilon_P)_{P\in\mathcal{P}}$. *From now on 'parametrization' means '$(\varepsilon_P)_{P\in\mathcal{P}}$-compatible parametrization'.* Let $\delta\colon \Theta \subset \mathbb{R}^n \to \mathcal{P}$ be a parametrization. Let $P_0 \in \mathcal{P}$, $\vartheta_0 := \delta^{-1}(P_0)$, and $\pi := \varepsilon_{P_0} \circ \delta$; let the matrix representation of the derivative $\pi'(\vartheta_0)$ at ϑ_0 be denoted by $\dot{l}^T$ for $\dot{l} \in \big(L_2(P_0)\big)^n$ and the image $\pi(\Theta)$, or $\varepsilon_{P_0}(\mathcal{P})$, in $L_2(P_0)$ by $\mathcal{P}_0$. Then (i)–(iv) and 3.1 provide us with an open ball $U \subset \Theta$ around ϑ_0 such that

$\pi(U)$ is open in $\mathcal{P}_0$ and $\pi|_U\colon U \to \pi(U)$ is a homeomorphism (this makes $\varepsilon_{P_0}|_{\delta(U)}\colon \delta(U) \to \pi(U)$ bijective) of which the inverse $(\pi|_U)^{-1}$ is differentiable at 0 along the tangent space of $\pi(U)$ at 0. According to 2.7, this tangent space is equal to $\dot{\mathcal{P}}_0$, the tangent space of $\mathcal{P}_0$ at 0, and by 3.2 we have $\dot{\mathcal{P}}_0 = [\dot{l}]$. The figure below might be helpful.

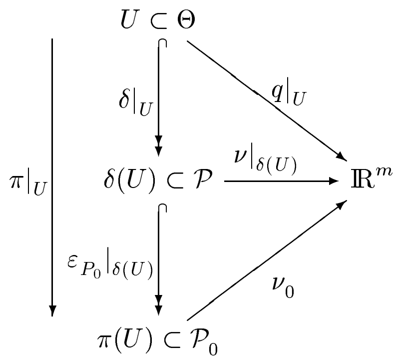


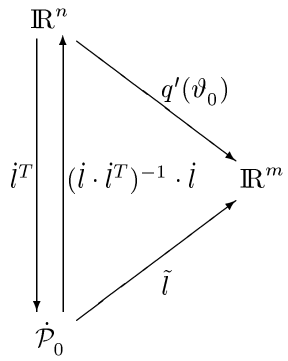
Now let $\nu\colon \mathcal{P} \to \mathbb{R}^m$ give a parameter, the open ball U around ϑ_0 be as above, and let us believe one more thing, namely that for this particular parametrization δ the composition $q := \nu \circ \delta$, $q : \Theta \subset \mathbb{R}^n \to \mathbb{R}^m$, is everywhere differentiable. Then the chain rule provides us with the differentiability of $\nu_0 := q \circ (\pi|_U)^{-1}$ at 0 along $\dot{\mathcal{P}}_0$. We claim that *this differentiability and the derivative at* 0 *are not only invariant under the choice of* U*, but also under the choice of the parametrization, and for any choice of the latter the map that corresponds to* q *is everywhere differentiable.*

Namely, suppose $'\delta\colon {'\Theta} \subset \mathbb{R}^n \to \mathcal{P}$ is a parametrization (Brouwer's invariance of dimension rules out the possibility $n \neq {'n}$); then we observe that the corresponding $'\nu_0 := \nu \circ {'\delta} \circ ({'\pi}|_{'U})^{-1}$ coincides with ν_0 on the open set $\pi(U) \cap {'\pi}({'U}) \subset \mathcal{P}_0$, as on this set both ν_0 and $'\nu_0$ coincide with $\nu \circ$ the inverse of a restriction of ε_{P_0}. Thus $'\nu_0$ is also differentiable at 0 along $\dot{\mathcal{P}}_0$, with the same derivative as ν_0, and $'q := \nu \circ {'\delta}$, too, is everywhere differentiable, because on $'U$ it coincides with $'\nu_0 \circ {'\pi}|_{'U}$ for any choice of P_0, which suffices by virtue of 2.2.

All this means that, after an invocation of 1.4, we can safely call the vector $\tilde{l}$ in $(\dot{\mathcal{P}}_0)^m \subset \big(L_2(P_0)\big)^m$ representing the derivative $\nu_0'(0)\colon \dot{\mathcal{P}}_0 \to \mathbb{R}^m$ of ν_0 at 0 along $\dot{\mathcal{P}}_0$ the **efficient influence function for** ν **in** P_0.

In the next figure, the left diagram gives the maps at hand; the right diagram shows matrix representations of the derivatives of the differentiable ones.





4.1 Theorem. *Let $q'(\vartheta_0)$ be the matrix representation of the derivative of q at ϑ_0. Then for the efficient influence function $\tilde{l}$ for ν in P_0 we have*

$$\tilde{l} \;=\; q'(\vartheta_0)\cdot(\dot{l}\cdot\dot{l}^T)^{-1}\cdot\dot{l}.$$

Proof. Starting with the chain rule, we obtain the map equalities

$$\begin{aligned}
\nu_0'(0) &\overset{2.3}{=} q'(\vartheta_0)\circ\big((\pi|_U)^{-1}\big)'(0)\\
&\overset{3.3}{=} q'(\vartheta_0)\circ(\dot{l}\cdot\dot{l}^T)^{-1}\cdot\dot{l}\\
&\overset{\sharp H\le 2}{=} q'(\vartheta_0)\cdot\big((\dot{l}\cdot\dot{l}^T)^{-1}\cdot\dot{l}\big)\\
&\overset{\sharp H\le 2}{=} q'(\vartheta_0)\cdot(\dot{l}\cdot\dot{l}^T)^{-1}\cdot\dot{l}.
\end{aligned}$$

So the last expression makes sense indeed and represents $\nu_0'(0)$. It is a vector in $([\dot{l}])^m = (\dot{\mathcal{P}}_0)^m$; hence it is equal to $\tilde{l}$ by 1.3. □

We see that, by postponing the definition of $\tilde{l}$, we could have avoided using 1.4.

If $'\delta{:}\,'\Theta\subset\mathbb{R}^n\to\mathcal{P}$ is another parametrization, then the derivative of $'q$ exists everywhere, as we saw above. Let $'\delta('\vartheta_0)=P_0$; from 4.1 it is immediately clear that

$$'q'('\vartheta_0)\cdot('\dot{l}\cdot{}'\dot{l}^T)^{-1}\cdot{}'\dot{l} = q'(\vartheta_0)\cdot(\dot{l}\cdot\dot{l}^T)^{-1}\cdot\dot{l};$$

cf. the derivation in [Bickel et al.] p. 23.

An outcome, or observation, of lottery $P_0\in\mathcal{P}$ is a point X of the probability space, drawn according to P_0. As $\dot{l}$ and $\tilde{l}$ are vectors of (classes of) functions (L_2 with respect to P_0) on the probability space, the values

$\dot{l}(X)$ and $\tilde{l}(X)$ have been defined. The vector $\dot{l}(X)$ is called the **score function of the observation for the parametrization δ at ϑ_0**. The matrix $\dot{l}\cdot\dot{l}^T$ is known as the **Fisher information matrix for the parametrization δ at ϑ_0**, the matrix $\tilde{l}\cdot\tilde{l}^T$ as the **information bound for ν at P_0**; remark that

$$\begin{aligned}\tilde{l}\cdot\tilde{l}^T &\overset{4.1}{=} (q'(\vartheta_0)\cdot(\dot{l}\cdot\dot{l}^T)^{-1}\cdot\dot{l})\cdot(\dot{l}^T\cdot(\dot{l}\cdot\dot{l}^T)^{-1}\cdot q'(\vartheta_0)^T)\\ &\overset{\sharp H\leq 2}{=} q'(\vartheta_0)\cdot(\dot{l}\cdot\dot{l}^T)^{-1}\cdot q'(\vartheta_0)^T.\end{aligned}$$

Deducing, or 'estimating' the property $\nu(P_0)$ of $P_0 \in \mathcal{P}$ on the base of independent outcomes $X_1,\ldots,X_N$ of (or 'a sample from') P_0 will, in general, not be perfect; the convolution theorem relates the measure of imperfection, i.e. the variance of the estimator, to the information bound $\tilde{l}\cdot\tilde{l}^T$, and optimal ('efficient') estimators to the values $\tilde{l}(X_i)$ of the efficient influence function in the points of the sample.

5 The efficient score function

Let ν be the inverse of the parametrization, $\nu : \delta(\vartheta) \mapsto \vartheta$. Then $q = \mathrm{id}_\Theta$, so that for the efficient influence function for ν we find

$$\tilde{l} = (\dot{l}\cdot\dot{l}^T)^{-1}\cdot\dot{l}$$

and for the information bound

$$\tilde{l}\cdot\tilde{l}^T = (\dot{l}\cdot\dot{l}^T)^{-1},$$

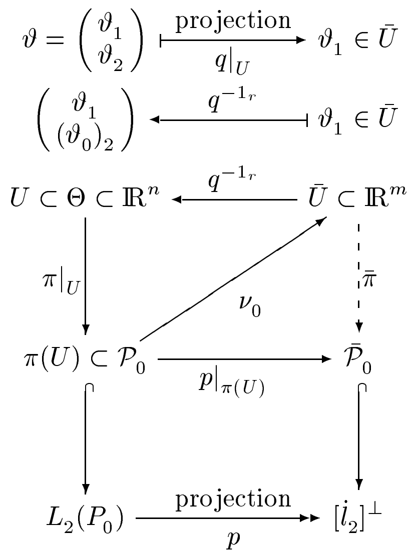
the inverse of the Fisher information matrix.

Now let ν be the map $\nu : \delta(\vartheta) \mapsto \vartheta \mapsto \vartheta_1 \in \mathbb{R}^m$, the projection on the first m coordinates of ϑ. Then $q{:}\,\vartheta \mapsto \vartheta_1$ and $q'(\vartheta_0)$ is equal to the $m\times n$ projection matrix $\begin{pmatrix}1 & & 0 & \\ & \ddots & & 0\\ 0 & & 1 & \end{pmatrix}$, so that for the efficient influence function $\tilde{l}_\nu$ for ν we find

$$\begin{aligned}&\tilde{l}_\nu = \text{the first } m \text{ coordinates of } (\dot{l}\cdot\dot{l}^T)^{-1}\cdot\dot{l}, \text{ or}\\ &\tilde{l}_\nu = \tilde{l}_1,\end{aligned}$$

if the previous efficient influence function is decomposed as $\tilde{l} = \begin{pmatrix}\tilde{l}_1\\ \tilde{l}_2\end{pmatrix}$ for $\tilde{l}_1 \in (\dot{\mathcal{P}}_0)^m$, $\tilde{l}_2 \in (\dot{\mathcal{P}}_0)^{n-m}$. Let also $\dot{l} = \begin{pmatrix}\dot{l}_1\\ \dot{l}_2\end{pmatrix}$ for $\dot{l}_1 \in (\dot{\mathcal{P}}_0)^m$, $\dot{l}_2 \in (\dot{\mathcal{P}}_0)^{n-m}$, and let l^* be the vector in $\big(L_2(P_0)\big)^m$ of which every component is the projection of the corresponding component of $\dot{l}_1$ onto $[\dot{l}_2]^\perp \subset L_2(P_0)$. We will relate l^* and $\tilde{l}_\nu$.

Consider the following diagram.



For any open ball $U \subset \Theta$ around ϑ_0 as above, its image $\bar{U} := q(U)$ is an open ball around $(\vartheta_0)_1$, whose own image under q^{-1_r} is contained in U again. After inspection it will be clear that $\bar{\pi} := p \circ \pi \circ q^{-1_r}$ is differentiable on $\bar{U}$, that $\vartheta_1 \mapsto \bar{\pi}'(\vartheta_1)$ is continuous at $(\vartheta_0)_1$, that the derivative $\bar{\pi}'\big((\vartheta_0)_1\big)$ is represented by l^{*T} (just look at what happens to the unit vectors in $\mathbb{R}^m$; continuous maps with a linear and a constant part, such as projections or right inverses thereof as q^{-1_r}, have the linear part as derivative), and that this derivative is injective (if a member of $[\dot{l}_1]$ is not equal to zero, it will not have projection 0 on $[\dot{l}_2]^\perp$; if it had, it would be a member of $[\dot{l}_2]$, while $\dot{l}$ is injective).

So theorem 3.1 provides us with an open set $\bar{U}_1 \subset \bar{U}$ such that $\bar{\pi}|_{\bar{U}_1}\colon \bar{U}_1 \to \bar{\pi}(\bar{U}_1)$ has an inverse that is differentiable at 0 along the tangent space of $\bar{\pi}(\bar{U}_1)$, theorem 3.2 says this tangent space is equal to $[l^*]$, and 3.3 gives $(l^* \cdot l^{*T})^{-1} \cdot l^*$ as a representation of the derivative $\big((\bar{\pi}|_{\bar{U}_1})^{-1}\big)'(0)\colon [l^*] \to \mathbb{R}^m$.

Another derivative, $\nu_0'(0)$, also acts on $[l^*]$, because it acts on $[\dot{l}]$ (observe that $[\dot{l}]$ is the closed linear span of $[l^*]$ and $[\dot{l}_2]$). As to their relation, further diagram chasing shows

$$\nu_0'(0)|_{[l^*]} = \big((\bar{\pi}|_{\bar{U}_1})^{-1}\big)'(0) \tag{1}$$

and

$$\nu_0'(0)|_{[\dot{l}_2]} = 0, \tag{2}$$

which amounts to

$$\nu_0'(0) = \big((\bar\pi|_{\bar U_1})^{-1}\big)'(0) \circ p'(0)|_{[\dot{l}]}.$$

(Equality (2) is obvious from

$$\nu_0'(0) = q'(\vartheta_0) \circ \big((\pi|_U)^{-1}\big)'(0),$$

because the very last map maps the i^{th} component of $\dot{l}_2$ to the $(m+i)^{\text{th}}$ unit vector of $\mathbb{R}^n$, a vector that $q'(\vartheta_0)$ makes disappear; for equality (1), note that

$$p^{-1_r} := \pi|_U \circ q^{-1_r} \circ (\bar\pi|_{\bar U_1})^{-1}$$

is a right inverse of p, with

$$\nu_0 \circ p^{-1_r} = (\bar\pi|_{\bar U_1})^{-1},$$

so that

$$p'(0) \circ (p^{-1_r})'(0) = \mathrm{id}_{[l^*]}$$

and therefore

$$(p^{-1_r})'(0)|_{[l^*]} = \mathrm{id}_{[l^*]}$$

by virtue of $p'(0) = p$, as well as

$$\nu_0'(0) \circ (p^{-1_r})'(0) = \big((\bar\pi|_{\bar U_1})^{-1}\big)'(0),$$

is guaranteed by 2.3.) From 1.3 it now follows that

$$\tilde{l}_\nu = (l^* \cdot l^{*T})^{-1} \cdot l^*.$$

Remark the similarity between the formula at the beginning of this section, for the whole parameter, and this one, for part of the parameter. In the latter case, classes of distributions instead of distributions are parametrized, so locally we deal with $L_2(P_0)/[\dot{l}_2] \cong [\dot{l}_2]^\perp$ instead of with $L_2(P_0)$. The score function $\dot{l}$ is replaced by l^*; we call l^* the **efficient score function**.

6 References

[Bickel et al.] Bickel, P.J., C.A.J. Klaassen, Y. Ritov, and J.A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, Johns Hopkins University Press, Baltimore, 1993

[Lang] Lang, S., *Differential and Riemannian manifolds*, Springer-Verlag, New York etc., 1995