Semiparametric Models and Efficiency

Jianqiao Wang

October 14, 2021

Overview

1. Tangent Sets and Regularity
2. Efficiency

For a statistical model $\{P \in \mathcal{P}\}$ on a sample space $\mathcal{X}$, a one-dimensional model $\{P_t\}$ is a *smooth submodel* at P if

- $P_0 = P$,
- $\{P_t; t \in N_\epsilon \equiv (-\epsilon, \epsilon)\} \subset \mathcal{P}$ for some $\epsilon > 0$ and
- for some measurable "tangent" function $g : \mathcal{X} \mapsto \mathbb{R}$,

$$\int \left[\frac{(dP_t(x))^{1/2} - (dP(x))^{1/2}}{t} - \frac{1}{2} g(x)(dP(x))^{1/2} \right]^2 \to 0, \qquad (1)$$

as $t \to 0$.

Lemma 11.11 forces the g in (1) to be contained in $L_2^0(P)$, the space of all functions $h : \mathcal{X} \mapsto \mathbb{R}$ with $Ph = 0$ and $Ph^2 < \infty$.

Tangent Sets

A tangent set $\dot{\mathcal{Q}}_P$ *represents* a submodel $\mathcal{Q} \subset \mathcal{P}$ at P if the following hold:

- For every smooth one-dimensional submodel $\{P_t\}$ for which

$$P_0 = P \text{ and } \{P_t : t \in N_\epsilon\} \subset \mathcal{Q} \text{ for some } \epsilon > 0 \tag{2}$$

  and for which (1) holds for some $g \in L_2^0(P)$, we have $g \in \dot{\mathcal{Q}}_P$
- For every $g \in \dot{\mathcal{Q}}_P$, there exists a smooth one-dimensional submodel $\{P_t\}$ such that (1) and (2) both hold.

Tangent Sets

Score functions for finite dimensional submodels can be represented by tangent sets corresponding to one-dimensional submodels.

To see this, let $\mathcal{Q} = \{P_\theta; \theta \in \Theta, \Theta \subset \mathbb{R}^k\} \subset \mathcal{P}$. Let $\theta_0 \in \Theta$ be the true value of the parameter, i.e. $P = P_{\theta_0}$. Suppose that the members P_θ of $\mathcal{Q}$ all have densities p_θ dominated by a measure μ, and that

$$\dot{\ell}_{\theta_0} \equiv \frac{\partial}{\partial\theta} \log p_\theta \bigg|_{\theta=\theta_0},$$

where $\dot{\ell}_{\theta_0} \in L_2^0(P)$, $P\|\dot{\ell}_\theta - \dot{\ell}_{\theta_0}\|^2 \to 0$ as $\theta \to \theta_0$.
The tangent set $\dot{\mathcal{Q}}_P \equiv \{h'\dot{\ell}_{\theta_0} : h \in \mathbb{R}^k\}$ contains all the information in the score $\dot{\ell}_{\theta_0}$, and $\dot{\mathcal{Q}}_P$ represents $\mathcal{Q}$.

Tangent Sets

Thus one-dimensional submodels are sufficient to represent all finite-dimensional submodels. Since semiparametric efficiency is assessed by examining the information for the worst finite-dimensional submodel, one-dimensional submodels are sufficient for semiparametric models in general, including models with infinite-dimensional parameters.

Tangent Sets

Now if $\{P_t : t \in N_\epsilon\}$ and $g \in \dot{\mathcal{P}}_P$ satisfy (1), then for any $a \geq 0$, everything will also hold when ϵ is replaced by ϵ/a and g is replaced by ag. Thus we can usually assume, without a significant loss in generality, that a tangent set $\dot{\mathcal{P}}_P$ is a *cone*, i.e., a set that is closed under multiplication by nonnegative scalars.

Differentiability

For an arbitrary model parameter $\psi : \mathcal{P} \mapsto \mathbb{D}$, consider the general setting where $\mathbb{D}$ is a Banach space $\mathbb{B}$. In this case, we say ψ is *differentiable at P relative to the tangent set* $\dot{\mathcal{P}}_P$ if, for every smooth one-dimensional submodel $\{P_t\}$ with tangent $g \in \dot{\mathcal{P}}_P$, $d\psi(P_t)/(dt)|_{t=0} = \dot{\psi}_P(g)$ for some bounded linear operator $\dot{\psi}_P : \dot{\mathcal{P}}_P \mapsto \mathbb{B}$.

Efficient Influence Function

Suppose $\dot{\mathcal{P}}_P$ is a linear space. The Riesz representation theorem yields that for every $b^* \in \mathbb{B}^*$, $b^*\dot{\psi}_P(g) = P[\tilde{\psi}_P(b^*)g]$ for some operator $\tilde{\psi}_P : \mathbb{B}^* \mapsto \overline{\text{lin}}\dot{\mathcal{P}}_P$.

Note that for any $g \in \dot{\mathcal{P}}_P$ and $b^* \in \mathbb{B}^*$, we have $b^*\dot{\psi}_P(g) = \langle g, \dot{\psi}_P^*(b^*)\rangle$, where $\langle \cdot, \cdot \rangle$ is the inner product on $L_2^0(P)$ and $\dot{\psi}_P^*$ is the adjoint of $\dot{\psi}_P$.

Thus the operator $\tilde{\psi}_P$ is precisely $\dot{\psi}_P^*$. In this case, $\tilde{\psi}_P$ is the *efficient influence function*.

Efficient Influence Function

We now present a way of verifying an efficient influence function:

Proposition (18.2)

Assume $\psi : \mathcal{P} \mapsto \ell^\infty(\mathcal{H})$ is differentiable at P relative to the linear tangent set $\dot{\mathcal{P}}_P$, with bounded linear derivative $\dot{\psi}_P : \dot{\mathcal{P}}_P \mapsto \ell^\infty(\mathcal{H})$. Then $\tilde{\psi}_P : \mathcal{H} \mapsto L_2^0(P)$ is an efficient influence function if and only if the following both hold:

- *$\tilde{\psi}_P(h)$ is in the closed linear span of $\dot{\mathcal{P}}_P$ for all $h \in \mathcal{H}$.*
- *$\dot{\psi}_P(g)(h) = P[\tilde{\psi}_P(h)g]$ for all $h \in \mathcal{H}$ and $g \in \dot{\mathcal{P}}_P$.*

Regularity

An estimator sequence $\{T_n\}$ for a parameter $\psi(P)$ is *asymptotically linear* if there exists an *influence function* $\check{\psi}_P : \mathcal{X} \mapsto \mathbb{B}$ such that $\sqrt{n}(T_n - \psi(P)) - \sqrt{n}\mathbb{P}_n\check{\psi}_P \xrightarrow{P} 0$.

The estimator T_n is *regular* at P relative to $\dot{\mathcal{P}}_P$ if for every smooth one-dimensional submodel $\{P_t\} \subset \mathcal{P}$ and every sequence t_n with $t_n = O(n^{-1/2})$, $\sqrt{n}(T_n - \psi(P_{t_n})) \overset{P_n}{\rightsquigarrow} Z$, for some tight Borel random element Z, where $P_n \equiv P_{t_n}$.

Regularity

The following theorem provides a way to establish regularity.

Theorem (18.1)

Assume

(1) T_n *and* $\psi(P)$ *are in* $\ell^\infty(\mathcal{H})$.

(2) ψ *is differentiable at* P *relative to the tangent space* $\dot{\mathcal{P}}_P$ *with efficient influence function* $\tilde{\psi}_P : \mathcal{H} \mapsto L_2^0(P)$.

(3) T_n *is asymptotically linear for* $\psi(P)$, *with influence function* $\check{\psi}_P$.

(4) *For each* $h \in \mathcal{H}$, *let* $\check{\psi}_P^\bullet$ *be the projection of* $\check{\psi}_P(h)$ *onto* $\dot{\mathcal{P}}_P$.

Then the following are equivalent:

(a) *The class* $\mathcal{F} \equiv \{\check{\psi}_P(h) : h \in \mathcal{H}\}$ *is P-Donsker and* $\check{\psi}_P^\bullet(h) = \tilde{\psi}_P(h)$ *almost surely for all* $h \in \mathcal{H}$.

(b) T_n *is regular at* P.

Notation

- Model parameter: $\psi : \mathcal{P} \mapsto \ell^\infty(\mathcal{H})$.
- Bounded linear operator: $\dot{\psi}_P : \dot{\mathcal{P}}_P \mapsto \ell^\infty(\mathcal{H})$.
- Efficient influence function: $\tilde{\psi}_P : \mathcal{H} \mapsto L_2^0(P)$.
- Influence function: $\check{\psi}_P : \mathcal{X} \mapsto \ell^\infty(\mathcal{H})$.
- Projection of influence function onto tangent space $\dot{\mathcal{P}}_p$: $\check{\psi}_P^\bullet$.

Proof of Theorem 18.1

Assume $\mathcal{F}$ is P-Donsker. Let P_t be any smooth one-dimensional submodel with tangent $g \in \dot{\mathcal{P}}_P$, and let t_n be any sequence with $\sqrt{n}t_n \to k$, for some finite constant k. Then $P_n = P_{t_n}$ satisfies

$$\int \left[\frac{(dP_{t_n}(x))^{1/2} - (dP(x))^{1/2}}{t_n} - \frac{1}{2} g(x)(dP(x))^{1/2} \right]^2 \to 0$$

i.e.

$$\int \left[\sqrt{n}\{(dP_n(x))^{1/2} - (dP(x))^{1/2}\} - \frac{1}{2} g(x)(dP(x))^{1/2} \right]^2 \to 0$$

Proof of Theorem 18.1

By Theorem 11.12,

$$\sqrt{n}\mathbb{P}_n\check{\psi}_P(\cdot) \overset{P_n}{\rightsquigarrow} \mathbb{G}\check{\psi}_P(\cdot) + P[\check{\psi}_P(\cdot)g] = \mathbb{G}\check{\psi}_P(\cdot) + P[\check{\psi}_P^{\bullet}(\cdot)g],$$

where $\mathbb{G}$ is a tight Brownian bridge. The last equality follows from the fact that $g \in \dot{\mathcal{P}}_P$.

Let $Y_n = \|\sqrt{n}(T_n - \psi(P)) - \sqrt{n}\mathbb{P}_n\check{\psi}_P\|_{\mathcal{H}}$, and note that $Y_n \xrightarrow{P} 0$ by the asymptotic linearity assumption. Without loss of generality, assume that the measurable sets for P_n and P^n (applied to the data $X_1, \ldots, X_n$) are both the same for all $n \geq 1$.

Proof of Theorem 18.1

By Theorem 11.14, $Y_n \xrightarrow{P_n} 0$. Combining this with the differentiability of ψ, we obtain that

$$\begin{aligned}\sqrt{n}(T_n - \psi(P_n))(\cdot) =& \sqrt{n}(T_n - \psi(P))(\cdot) - \sqrt{n}(\psi(P_n) - \psi(P))(\cdot) \\ \overset{P_n}{\rightsquigarrow}& \mathbb{G}\check{\psi}_P(\cdot) + P[(\check{\psi}_P^{\bullet}(\cdot) - \tilde{\psi}_P(\cdot))g], \end{aligned} \tag{3}$$

in $\ell^\infty(\mathcal{H})$.

Proof of Theorem 18.1

Suppose now that T_n is regular ((b) holds) and we don't know wheter $\mathcal{F}$ is P-Donsker. The regularity of T_n implies that $\sqrt{n}(T_n - \psi(P)) \rightsquigarrow Z$ for some tight process Z. The asymptotic linearity of T_n now forces $\sqrt{n}\mathbb{P}_n\check{\psi}_P(\cdot) \rightsquigarrow Z$, which yields that $\mathcal{F}$ is P-Donsker.

Suppose that for some $h \in \mathcal{H}$ we have $\tilde{g}(h) \equiv \check{\psi}_P^\bullet(h) - \tilde{\psi}_P(h) \neq 0$. Since the choice of g in the arguments leading up to (3) was arbitrary, we can choose $g = a\tilde{g}$ for any $a > 0$ to yield

$$\sqrt{n}(T_n(h) - \psi(P_n)(h)) \overset{P_n}{\rightsquigarrow} \mathbb{G}\check{\psi}_P(h) + aP\tilde{g}^2. \tag{4}$$

Thus we can easily have different limiting distributions by choosing different values of a, which is contradictory with regularity of T_n. Thus (a) holds.

Proof of Theorem 18.1

Assume (a) holds. For arbitrary choices of $g \in \dot{\mathcal{P}}_P$ and constant k such that $\sqrt{n} t_n \to k$,

$$\sqrt{n}(T_n - \psi(P_n)) \overset{P_n}{\rightsquigarrow} \mathbb{G}\check{\psi}_P$$

in $\ell^\infty(\mathcal{H})$. Now relax the assumption that $\sqrt{n} t_n \to k$ to $\sqrt{n} t_n = O(1)$ and allow g to arbitrary as before. Under weaker assumption, we have that for every subsequence n', there exists a further subsequence n'' such that $\sqrt{n''} t_{n''} \to k$ for some finite k, as $n'' \to \infty$. Arguing along this subsequence, our previous arguments can all be recycled to verify that

$$\sqrt{n''}(T_{n''} - \psi(P_{n''})) \overset{P_{n''}}{\rightsquigarrow} \mathbb{G}\check{\psi}_P$$

in $\ell^\infty(\mathcal{H})$, as $n'' \to \infty$.

Proof of Theorem 18.1

Define $Z_n \equiv \sqrt{n}(T_n - \psi(P_n))$ and $Z \equiv \mathbb{G}\check{\psi}_P$. Fix $f \in C_b(\ell^\infty(\mathcal{H}))$, recall portmanteau theorem. Every subsequence n' has a further subsequence n'' such that $E^* f(Z_{n''}) \to Ef(Z)$, as $n'' \to \infty$, which implies that $E^* f(Z_n) \to Ef(Z)$, as $n \to \infty$. Then portmanteau theorem yields $Z_n \overset{P_n}{\rightsquigarrow} Z$. Thus T_n is regular.

Efficiency

We now turn our attention to the question of efficiency in estimating general Banach-values parameters.

We first present general optimality results and then characterize efficient estimators in the special Banach space $\ell^\infty(\mathcal{H})$.

We then

- consider efficiency of Hadamard-differentiable functionals of efficient parameters,
- show how to establish efficiency of estimators in $\ell^\infty(\mathcal{H})$ from efficiency of all one-dimensional components, and
- examine the related issue of efficiency in product spaces.

Efficiency

For a Borel random element Y, let $L(Y)$ denote the law of Y, and let $*$ denote the convolution operation.

Define a function $u : \mathbb{B} \mapsto [0, \infty)$ to be *subconvex* if,

- for every $b \in \mathbb{B}$, $u(b) \geq 0 = u(0)$ and $u(b) = u(-b)$,
- for every $c \in \mathbb{R}$, the set $\{b \in \mathbb{B} : u(b) \leq c\}$ is convex and closed.

A simple example of a *subconvex* function is the norm $\|\cdot\|$ for $\mathbb{B}$.

The following two theorems characterize optimality in Banach spaces.

Efficiency

Theorem (Convolution theorem)

Assume the $\psi : \mathcal{P} \mapsto \mathbb{B}$ is differentiable at P relative to the tangent space $\dot{\mathcal{P}}_P$, with efficient influence function $\tilde{\psi}_P$. Assume that T_n is regular at P relative to $\dot{\mathcal{P}}_P$, with Z being the tight weak limit of $\sqrt{n}(T_n - \psi(P))$ under P. Then $L(Z) = L(Z_0) * L(M)$, where M is some Borel random element in $\mathbb{B}$, and Z_0 is a tight Gaussian process in $\mathbb{B}$ with covariance $P[(b_1^* Z_0)(b_2^* Z_0)] = P[\tilde{\psi}_P(b_1^*)\tilde{\psi}_P(b_2^*)]$ for all $b_1^*, b_2^* \in \mathbb{B}^*$.

Efficiency

Theorem (18.4)

Assume that conditions of convolution theorem holds and that $u : \mathbb{B} \mapsto [0, \infty)$ is subconvex. Then

$$\limsup_{n\to\infty} E_* u(\sqrt{n}(T_n - \psi(P))) \geq Eu(Z_0),$$

where Z_0 is as defined in convolution theorem.

Efficiency

The previous two theorems characterize optimality of regular estimators in terms of the limiting process Z_0, which is a tight, mean zero Gaussian process with covariance obtained from the efficient influence function. This can be viewed as an asymptotic generalization of the Cramer-Rao lower bound.

We say that an estimator T_n is *efficient* if it it regular and the limiting distribution of $\sqrt{n}(T_n - \psi(P))$ is Z_0, i.e., T_n achieves the optimal lower bound.

Efficiency

The next proposition assures us that Z_0 is fully characterized by the distributions of $b^* Z_0$ for b^* ranging over all of $\mathbb{B}^*$:

Proposition (18.5)

Let X_n be an asymptotically tight sequence in a Banach space $\mathbb{B}$ and assume $b^* X_n \rightsquigarrow b^* X$ for every $b^* \in \mathbb{B}^*$ and some tight Gaussian process X in $\mathbb{B}$. Then $X_n \rightsquigarrow X$.

Proof of Proposition 18.5

Let $\mathbb{B}_1^* \equiv \{b^* \in \mathbb{B}^* : \|b^*\| \leq 1\}$ and $\tilde{\mathbb{B}} \equiv \ell^\infty(\mathbb{B}_1^*)$. Note that $(\mathbb{B}, \|\cdot\|) \subset (\tilde{\mathbb{B}}, \|\cdot\|_{\mathbb{B}_1^*})$ by letting $x(b^*) \equiv b^* x$ for every $b^* \in \mathbb{B}^*$ and all $x \in \mathbb{B}$. By Hahn-Banach theorem,

$$\|x\| = \sup_{b^* \in \mathbb{B}^*} |b^* x| = \|x\|_{\mathbb{B}_1^*}.$$

Thus, by Lemma 7.8, weak convergence of X_n in $\tilde{\mathbb{B}}$ will imply weak convergence in $\mathbb{B}$.

Since we already know that X_n is asymptotically tight in $\tilde{\mathbb{B}}$, we only need to show that all finite-dimensional distributions of X_n converge.

Proof of Proposition 18.5

Let $b_1^*, \ldots, b_m^* \in \mathbb{B}_1^*$ be arbitrary and note that for any $(\alpha_1, \ldots, \alpha_m) \in \mathbb{R}^m$,

$$\sum_{j=1}^{m} \alpha_j X_n(b_j^*) = \tilde{b}^* X_n, \text{ for } \tilde{b}^* \equiv \sum_{j=1}^{m} \alpha_j b_j^* \in \mathbb{B}^*.$$

Since we know that $\tilde{b}^* X_n \rightsquigarrow \tilde{b}^* X$, we know that

$$\sum_{j=1}^{m} \alpha_j b_j^* X_n \rightsquigarrow \sum_{j=1}^{m} \alpha_j b_j^* X.$$

Thus $(X_n(b_1^*), \ldots, X_n(b_m^*))^T \rightsquigarrow (X(b_1^*), \ldots, X(b_m^*))^T$ since $(\alpha_1, \ldots, \alpha_m) \in \mathbb{R}^m$ was arbitrary and X is Gaussian. Since $b_1^*, \ldots, b_m^*$ and m were also arbitrary, all finite-dimensional distributions of X_n converge.

Efficiency

The next theorem assures us that Hadamard differentiable functions of efficient estimators are also asymptotically efficient.

Theorem (18.6)

Assume that

- $\psi : \mathcal{P} \mapsto \mathbb{B}$ *is differentiable at* P *relative to the tangent space* $\dot{\mathcal{P}}_P$*, with derivative* $\dot{\psi}_P g$*, for every* $g \in \dot{\mathcal{P}}_P$*, and efficient influence function* $\tilde{\psi}_P$*, and takes its values in a subset* $\mathbb{B}_\phi$*.*
- $\phi : \mathbb{B}_\phi \subset \mathbb{B} \mapsto \mathbb{E}$ *is Hadamard differentiable at* $\psi(P)$ *tangentially to* $\mathbb{B}_0 \equiv \overline{lin} \dot{\psi}_P(\dot{\mathcal{P}}_P)$*.*

Then $\phi \circ \psi : \mathcal{P} \mapsto \mathbb{E}$ *is also differentiable at* P *related to* $\dot{\mathcal{P}}_P$*. If* T_n *is a sequence of estimators with values in* B_ϕ *that is efficient at* P *for estimating* $\psi(P)$*, the* $\phi(T_n)$ *is efficient at* P *for estimating* $\phi \circ \psi(P)$*.*

Proof of Theorem 18.6

Let $\phi'_{\psi(P)} : \mathbb{B} \mapsto \mathbb{E}$ be the derivative of ϕ. Note that for any $g \in \dot{\mathcal{P}}_P$ and any submodel $\{P_t\}$ with tangent g, by the differentiablity of ϕ and ψ,

$$\begin{aligned} \frac{\phi \circ \psi(P_t) - \phi \circ \psi(P)}{t} =& \frac{\phi(\psi(P) + t\frac{\psi(P_t) - \psi(P)}{t}) - \phi(\psi(P))}{t} \\ \to& \phi'_{\psi(P)} \dot{\psi}_P g \end{aligned}$$

as $t \to 0$. Thus $\phi \circ \psi : \mathcal{P} \mapsto \mathbb{E}$ is differentiable at P relative to $\dot{\mathcal{P}}_P$.

Proof of Theorem 18.6

For any chosen submodel $\{P_t\}$ with tangent $g \in \dot{\mathcal{P}}_P$, define $P_n \equiv P_{1/\sqrt{n}}$. By the efficiency of T_n, we have that

$$\sqrt{n}(T_n - \psi(P_n)) \overset{P_n}{\rightsquigarrow} Z_0,$$

where Z_0 has the optimal, mean zero, tight Gaussian limiting distribution. By the delta method,

$$\sqrt{n}(\phi(T_n) - \phi \circ \psi(P_n)) \overset{P_n}{\rightsquigarrow} \phi'_{\psi(P)} Z_0.$$

Since the choice of $\{P_t\}$ was arbitrary, we now know that $\psi(T_n)$ is regular and also that

$$\sqrt{n}(\phi(T_n) - \phi \circ \psi(P)) \rightsquigarrow \phi'_{\psi(P)} Z_0.$$

Proof of Theorem 18.6

By convolution theorem, for all $e_1^*, e_2^* \in \mathbb{E}^*$,

$$P[(e_1^* \phi'_{\psi(P)} Z_0)(e_2^* \phi'_{\psi(P)} Z_0)] = P[\tilde{\psi}_P(e_1^* \phi'_{\psi(P)})\tilde{\psi}_P(e_2^* \phi'_{\psi(P)})]$$

Thus the desired result follows from differentiability of $\phi \circ \psi$ and the definition of efficient influence function $\tilde{\psi}_P$ since for every $e^* \in \mathbb{E}^*$ and $g \in \overline{lin}\dot{\mathcal{P}}_P$,

$$P[\tilde{\psi}_P(e^* \phi'_{\psi(P)})g] = e^* \phi'_{\psi(P)} \dot{\psi}_P(g).$$