Chapter 4. Efficiency in Semiparametric models

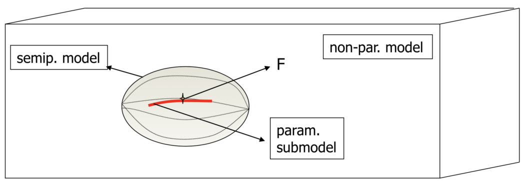
March 25, 2022

Outline

Outline

1 Regular parametric submodels
2 Tangent sets and tangent spaces
3 Calculation of the maximal tangent space in the non-parametric model
4 Regular estimators
5 Pathwise differentiable parameters, semiparametric C-R bound
6 Some examples of calculation of gradients and bounds
7 Representation of the set of gradients
8 Representation of the set of influence functions of regular asymptotically linear estimators
9 The convolution theorem for estimation in semiparametric models

Definition 4.1. Given a semiparametric model $\mathcal{F}$, a regular parametric model is a submodel of $\mathcal{F}$ at F iff
a) F in the submodel,
b) the family of distributions allowed by the parametric model is included in $\mathcal{F}$ and,
c) the parametric model is regular.



Example. Suppose that $\mathcal{F}$ is the non-parametric model and F is the $N(0,1)$ distribution. Then

$$\mathcal{F}_{\text{sub}} = \{N(\theta,1) : \theta \in \mathbb{R}\},$$

and

$$\mathcal{F}'_{\text{sub}} = \left\{N\left(\theta_1, \theta_2^2\right) : \theta_1 \in \mathbb{R}, \theta_2 \in \mathbb{R}^+\right\},$$

are regular parametric submodels of $\mathcal{F}$ through F.

Here is another regular parametric submodel:

Let

$$k(u) = 2\left(1 + e^{-2u}\right)^{-1},$$

and

$$f^*(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right).$$

Define

$$f(x;\theta) \equiv f^*(x) k\{\theta h(x)\} c(\theta),$$

where $c(\theta) = \left[\int f^*(x) k\{\theta h(x)\} dx\right]^{-1}$. Then,

$$\mathcal{F}^*_{\text{sub}} = \{f(x;\theta) : \theta \in (-\varepsilon, \varepsilon)\}$$

is a reg. parametric submodel of the non-parametric model at the $N(0,1)$ distribution that goes through $N(0,1)$ at $\theta = 0$.

The model

$$\mathcal{F}'''_{\text{sub}} = \{f(x-\theta) : f(u) \text{ is the logistic density function}, \theta \in \mathbb{R}\}$$

is not a reg. parametric submodel through F equal to $N(0,1)$ because even though the family of distr. $\mathcal{F}'''_{\text{sub}}$ is included in those allowed by $\mathcal{F}$, F is not a member of $\mathcal{F}'''_{\text{sub}}$.

Definitional convention: From now on, all parametric submodels we refer to are regular parametric submodels.

Outline

Let $\mathcal{A}$ be a collection of reg. parametric submodels of a semip. model $\mathcal{F}$ through F :

$$\mathcal{A} = \{\mathcal{F}_{\text{sub}} : \mathcal{F}_{\text{sub}} \text{ is a reg. parametric submodel of } \mathcal{F} \text{ through } F\}.$$

The tangent set of model $\mathcal{F}$ with respect to $\mathcal{A}$ at F is defined as

$$\bigcup_{\mathcal{F}_{\text{sub}} : \mathcal{F}_{\text{sub}} \in \mathcal{A}} \Lambda_{\mathcal{F}_{\text{sub}}}(F),$$

where $\Lambda_{\mathcal{F}_{\text{sub}}}(F)$ is the tangent space of $\mathcal{F}_{\text{sub}}$ at F.

The tangent space of model $\mathcal{F}$ with respect to $\mathcal{A}$ at F is defined as the closure of the linear span of the tangent set,

$$\Lambda_{\mathcal{F}}(F) \equiv \overline{\left[\bigcup_{\mathcal{F}_{\text{sub}} : \mathcal{F}_{\text{sub}} \in \mathcal{A}} \Lambda_{\mathcal{F}_{\text{sub}}}(F) \right]}.$$

Notational remark

Note that although not explicit in the notation, the tangent space is defined with respect to a collection $\mathcal{A}$ of parametric submodels.

On occasion, if the semiparametric model is indexed by θ we will write

$$\Lambda_{\mathcal{F}}(\theta) \equiv \Lambda_{\mathcal{F}}(F_\theta).$$

Also, when clear from the context, we may even eliminate the subscript $\mathcal{F}$.

A general remark on the calculation of the tangent space

As a general rule, to construct the tangent space we take an "educated guess"at this space guided by the restrictions that the scores must satisfy. We deduce these restrictions from the restrictions of the semiparametric model.

Verifying that the conjectured space is indeed the tangent space is a technical exercise which consists of

a) showing that the scores of regular submodels are in the postulated set (usually, this is the "easy" step) and,

b) exhibiting, for any given element of the postulated set, a sequence a reg. parametric submodels such that linear combinations of the scores in the sequence tend to the given element (this is the cumbersome step).

Outline

Example 4.1. Suppose that $\mathcal{A}$ is the class of all reg. parametric submodels of the non-parametric model $\mathcal{F}$.

Since the non-parametric model does not impose any restrictions on the distributions, then the set of scores of all reg. parametric submodels, should not have any restriction beyond the requirement that they be mean zero and with finite variance.

So, it is natural to conjecture that the tangent space at a given law F is equal to $\mathcal{L}_2^0(F)$.

Since by definition, $\Lambda_{\mathcal{F}}(F) \subseteq \mathcal{L}_2^0(F)$, to show that

$$\Lambda_{\mathcal{F}}(F) = \mathcal{L}_2^0(F),$$

we must prove

$$\mathcal{L}_2^0(F) \subseteq \Lambda_{\mathcal{F}}(F).$$

To do so, it suffices to show that if X has distr. F, then for any $g(X)$ with mean zero and finite variance under F, there exists a reg. parametric submodel with score at the truth equal to $g(X)$.

We will now exhibit such reg. parametric submodel.

Let $f(x)$ be the density of F. Let $g(X)$ be in $\mathcal{L}_2^0(F)$.

Define the one dimensional submodel,

$$\mathcal{F}_{\mathsf{sub},g} = \{f(x;\theta) : \theta \text{ in } (-\varepsilon, \varepsilon)\},$$

where

$$f(x;\theta) = f(x)k\{\theta g(x)\}c(\theta),$$

with $c(\theta) = \left\{\int f(x)k(\theta g(x))dx\right\}^{-1}$ and $k(u) = 2\left(1 + e^{-2u}\right)^{-1}$.

Invoking Lemma 3.1, it can be checked that the submodel

$$\mathcal{F}_{\mathsf{sub},g} = \{f(x;\theta) : \theta \text{ in } (-\varepsilon, \varepsilon)\}$$

is a reg. parametric submodel of $\mathcal{F}$ through F with score that can be computed as

$$s_\theta(x;\theta) = \left.\frac{\partial}{\partial\theta}\log f(x;\theta)\right|_{\theta=0}.$$

Also, $k'(0) = k(0) = 1$.

Then,

$$\frac{d}{d\theta}\log f(x;\theta) = \frac{f(x)g(x)k'\{\theta g(x)\}c(\theta) + f(x)k\{\theta g(x)\}c'(\theta)}{f(x)k\{\theta g(x)\}, c(\theta)}$$

and

$$\frac{d}{d\theta}\log f(x;\theta)\bigg|_{\theta=0} = \frac{f(x)g(x)k'(0)c(0) + f(x)k(0)c'(0)}{f(x)k(0)c(0)}.$$

But

$$c(\theta)' = -\left[\int f^*(x)k\{\theta g(x)\}dx\right]^{-2}\int f^*(x)k\{\theta g(x)\}' g(x)dx,$$

So $c'(0) = 0$ because $g(X)$ has mean zero and $k'(0) = 1$. Furthermore, $c(0) = 1$ because densities integrate to 1. Thus, since $k(0) = k'(0) = 1$, then

$$\frac{d}{d\theta}\log f(x;\theta)\bigg|_{\theta=0} = g(x).$$

The preceding parametric submodel

$$f(x;\theta) = f(x)k\{\theta g(x)\}c(\theta)$$

is not the only submodel with score equal to $g(x)$ at $\theta = 0$.

In fact, any continuously differentiable bounded function $k(u)$ satisfying $k'(0) = k(0) = 1$ yields a reg. parametric submodel with the score $g(X)$. (You should convince yourself of this by going carefully throughout he preceding derivation).

Another submodel that it is sometimes useful in proofs is

$$f(x;\theta) \equiv f(x)\{1 + \theta g(x)\}c(\theta),$$

with $c(\theta) \equiv \left[\int f(x)\{1 + \theta g(x)\}dx\right]^{-1}$.

$f(x;\theta)$ is a density so long as $g(x)$ is bounded and θ is in $(-\varepsilon, \varepsilon)$ for a sufficiency small ε and $E_F[g(X)] = 0$.

In the homework you will check, using Lemma 3.1 that $\{f(x;\theta) : \theta \in (-\varepsilon, \varepsilon)\}$ is a reg. parametric model with score $g(X)$.

Now, because every function in $\mathcal{L}_2{}^0(F)$ can be approximated by bounded functions in $\mathcal{L}_2{}^0(F)$, i. e. for any $h(X)$ in $\mathcal{L}_2{}^0(F)$ there exist $g_1, g_2, \ldots$ such that $\|g_n - h\|_{\mathcal{L}_2(F)} \to 0$, then the tangent space of the class $\mathcal{A}'$ of parametric submodels with densities of the form as in the first display is $\mathcal{L}_2{}^0(F)$.

This example also teaches us important points.

First, we learn that there can be two different parametric submodels through the same F that have the same score. (just as two functions can have the same derivative at a point). In fact, it is possible to find infinitely many parametric submodels with the same score at F.

Second, and as a consequence of the preceeding point, a class $\mathcal{A}$ may be strictly smaller than another class and may have the same tangent space. For instance, the class of all reg. parametric submodels of the non-parametric model $\mathcal{F}$ has the same tangent space as the class $\mathcal{A}$ of parametric submodels of the form $\mathcal{F}_{\mathsf{sub},g}$ as defined as in the preceding example.

Outline

Definition 4.2. Given a collection $\mathcal{A}$ of reg. parametric submodels of a semiparametric model $\mathcal{F}$, an estimator $\widehat{\beta}_n$ is said to be a regular estimator of a parameter $\beta(F)$ in model $\mathcal{F}$ with respect to $\mathcal{A}$ at F, if it is a regular estimator at F under every parametric submodel in $\mathcal{A}$.

NOTE: by definition, if an estimator is regular wrt to a class $\mathcal{A}$ then it is regular with respect to every class $\mathcal{A}'$ contained in $\mathcal{A}$.

Example 4.2. Let $\mathcal{F}$ be the non-parametric model for a random variable X absolutely continuous with respect to the Lebesgue measure, i.e. "continuous".
For a given real valued function $b(\cdot)$, let

$$\beta(F) = E_F[b(X)].$$

Let $\mathcal{A}$ be the class of all reg. parametric submodels such that for each model in the class, say indexed by θ, through F^* at θ^*, i.e. $F^* = F_{\theta^*}$, the map

$$\theta \to E_\theta\left[b(X)^2\right],$$

is continuous in an open neighborhood of θ^*.

The empirical mean (sample average) $\widehat{\beta}_n = n^{-1}\sum_{i=1}^n b(X_i)$, satisfies

$$\sqrt{n}\left\{\widehat{\beta}_n - \beta(F)\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \{b(X_i) - \beta(F)\},$$

so it is a. linear with influence function

$$\varphi_{F^*}(X) = b(X) - E_{F^*}(b(X)).$$

Furthermore, by the assumed continuity of the map $\theta \to \mathrm{var}_\theta[b(X)]$ for every reg. parametric model in the class $\mathcal{A}$, then by Lemma 3.3,

$$\begin{aligned} \partial\beta(\theta)/\,\partial\theta^T\big|_{\theta^*} \equiv \partial E_\theta[b(X)]/\,\partial\theta^T\big|_{\theta^*} &= E_{\theta^*}\left[b(X)s_\theta\left(X;\theta^*\right)^T\right] \\ &= E_{\theta^*}\left[\varphi_{F^*}(X)s_\theta\left(X;\theta^T\right)^T\right]. \end{aligned}$$

So, by Lemma 3.5, the sample mean is a regular estimator of $\beta(F)$ in model F w.r.t $\mathcal{A}$ at F^*.

Remark: The requirement on the submodels in the class $\mathcal{A}$ of the continuity of the map $\theta \to E_\theta[b(X)^2]$ is a technical requirement that we impose to ensure that the sample average is a "regular" estimator in the submodel.

Although the class $\mathcal{A}$ is not the class of all reg. parametric submodels of the non-parametric model, nevertheless, the tangent space of $\mathcal{A}$ at F, is indeed equal to the tangent space of the non-parametric model, i.e $\mathcal{L}_2{}^0\left(F^*\right)$. This will be an important point when we discuss the efficiency bounds. To show the tangent space of $\mathcal{A}$ is $\mathcal{L}_2{}^0\left(F^*\right)$ it suffices to show that for any $g(X)$ in $\mathcal{L}_2{}^0\left(F^*\right)$ the map

$$\theta \to E_\theta\left[b(X)^2\right]$$

is continuous under the submodel

$$f^*(X)c(\theta)k(\theta g(X)),$$

where $k(\cdot)$ and $c(\cdot)$ were defined in slide 15. Because in such case $g(X)$, being the model's score, is in the tangent space.

Remark: As in this example, in nearly all inferential problems, the "maximal class" of all regular parametric submodels of a given semiparametric model is too big, as it may be impossible to find an estimator of the estimand of interest that is regular w.r.t. the maximal class.

It is because of this that we define regularity of an estimator w.r.t a class $\mathcal{A}$ of submodels of $\mathcal{F}$ that may not contain all of the reg. parametric submodels of F.

However, as in this example, in the spirit of making inferences valid in the big semiparametric model, i.e. to ensure valid inference under a large set of data generating laws, we like to find regular estimators w.r.t. a class $\mathcal{A}$ that is as big as possible.

The point about tangent sets and regularity defined with respect to a class of parametric submodels that is not maximal, is not spelled out in the book of Tsiatis.

Van der Vaart, 2000 (see sec 25.3) and van der Vaart, 1999 (see sec 1.2) does define tangent sets relative to a collection of submodels (more on this later).

In many examples, one can choose the class so that the tangent space associated with the class coincides with the maximal tangent space, i.e. the tangent space associated with the class of all reg. parametric submodels of the assumed semiparametric model.

Outline

Suppose that $\widehat{\beta}_n$ is a regular estimator of a scalar parameter $\beta(F)$ in a semiparametric model $\mathcal{F}$ w.r.t. $\mathcal{A}$ at F. Then, the asymptotic variance of $\sqrt{n}\left\{\widehat{\beta}_n - \beta(F)\right\}$ cannot be smaller than $C_{\mathcal{F}_{\text{sub}}}(F)$ for every $\mathcal{F}_{\text{sub}}$ in $\mathcal{A}$.

Consequently, the asymptotic variance of $\sqrt{n}\left\{\widehat{\beta}_n - \beta(F)\right\}$ cannot be smaller than

$$\sup_{\mathcal{F}_{\text{sub}}:\mathcal{F}_{\text{sub}}\in\mathcal{A}} C_{\mathcal{F}_{\text{sub}}}(F).$$

Regular estimators of $\beta(F)$ may or may not exist when the supremum of the C-R bounds is finite. However, one thing is sure, regular estimators don't exist if the supremum is infinite.

This is because if $\widehat{\beta}_n$ is a regular estimator of $\beta(\cdot)$ at F under a parametric submodel $\mathcal{F}_{\text{sub}}$ then by the Convolution Theorem

$$\sqrt{n}\left\{\widehat{\beta}_n - \beta(F)\right\} \to U + U^*,$$

where $U \sim N\left(0, C_{\mathcal{F}_{\text{sub}}}(F)\right)$ and U^* is indep of U, so

$$\operatorname{var}\left(U + U^*\right) \geq C_{\mathcal{F}_{\text{sub}}}(F).$$

So, a necessary condition for $\widehat{\beta}_n$ to be regular w.r.t. to a class $\mathcal{A}$ is

$$\sup_{\mathcal{F}_{\text{sub}}:\mathcal{F}_{\text{sub}}\in\mathcal{A}} C_{\mathcal{F}_{\text{sub}}}(F) < \infty.$$

We will now see that a sufficient condition on an estimand $\beta(F)$ for the previous display to be true is that the functional $F \to \beta(F)$ be pathwise differentiable at F^* wr.t. $\mathcal{A}$.

Road map to the answer

In regular parametric models, subject to regularity conditions, all unbiased estimators project into the same element of the tangent space. The variance of the projection is the CR bound.

↓

If $\beta(F) = E_F[\psi(X)]$ then the CR bound in ANY submodel is the variance of the projection $\psi(X)$ into the tangent space for the submodel. The supremum of the CR bounds for all submodels in a class is therefore less than or equal to the variance of the projection of $\psi(X)$ onto the smallest closed linear space that includes the tangent spaces for all submodels.

↓

For an arbitrary parameter $\beta(F)$ the supremum of the CR bounds is finite so long as the parameter is pathwise differentiable.

Road map to the answer of question 1

In regular parametric models, subject to regularity conditions, all unbiased estimators project into the same element of the tangent space. The variance of the projection is the CR bound.

We have already established this result in Chapter 3.

Road map to the answer of question 1

In regular parametric models, subject to regularity conditions, all unbiased estimators project into the same element of the tangent space. The variance of the projection is the CR bound.

$\downarrow$

If $\beta(F) = E_F[\psi(X)]$ then the CR bound in ANY submodel is the variance of the projection $\psi(X)$ into the tangent space for the submodel. The supremum of the CR bounds for all submodels in a class is therefore less than or equal to the variance of the projection of $\psi(X)$ onto the smallest closed linear space that includes the tangent spaces for all submodels.

Suppose that we wish to estimate the mean of a real valued function of X, i.e.

$$\beta(F) = E_F[\psi(X)],$$

for some known function $\psi(X)$ under a given semiparametric model $\mathcal{F}$ which, imposes the restriction

$$\mathrm{Var}_F[\psi(X)] < \infty$$

for F in $\mathcal{F}$ and perhaps some other restrictions.

We will now show that for the class $\mathcal{A}$ of all reg. parametric submodels $\mathcal{F}_{\mathsf{sub}}$ indexed, say, by θ with $F^* = F_{\theta^*}$ such that the map

$$\theta \to E_\theta\left[\psi(X)^2\right]$$

is continuous in a neighborhood θ^*, it holds that

$$\sup_{\mathcal{F}_{\mathsf{sub}} : \mathcal{F}_{\mathsf{sub}} \in \mathcal{A}} C_{\mathcal{F}_{\mathsf{sub}}}(F) < \infty.$$

Here is why...

1. Using what we have learned for the C-R bound of unbiased estimators in ch 3, for any $\mathcal{F}_{\mathsf{sub}}$ in $\mathcal{A}$ it holds that

$$C_{\mathcal{F}_{\mathsf{sub}}}(F) = \mathrm{var}_F\left\{\prod_F\left[\psi(X) \mid \Lambda_{\mathcal{F}_{\mathsf{sub}}}(F)\right]\right\}.$$

2. Then, from step 1 (with supremum over the class $\mathcal{A}$)

$$\sup_{\mathcal{F}_{\mathsf{sub}}} C_{\mathcal{F}_{\mathsf{sub}}}(F) \leqslant \mathrm{var}_F\left\{\Pi_F\left[\psi(X) \mid \Lambda_{\mathcal{F}}(F)\right]\right\},$$

where

$$\Lambda_{\mathcal{F}}(F) = \overline{\left[\underset{\mathcal{F}_{\mathsf{sub}}}{\cup} \Lambda_{\mathcal{F}_{\mathsf{sub}}}(F)\right]}.$$

3. Since $\Lambda_{\mathcal{F}}(F)$ is a closed linear subspace of $\mathcal{L}_2^0(F)$ (because each $\Lambda_{\mathcal{F}_{\mathsf{sub}}}(F)$ is in $\mathcal{L}_2^0(F)$) then

$$\mathrm{var}_F\left\{\Pi_F\left[\psi(X) \mid \Lambda_{\mathcal{F}}(F)\right]\right\} < \infty.$$

Road map to the answer

In regular parametric models, subject to regularity conditions, all unbiased estimators project into the same element of the tangent space. The variance of the projection is the CR bound.

$\downarrow$

If $\beta(F) = E_F[\psi(X)]$ then the CR bound in ANY submodel is the variance of the projection $\psi(X)$ into the tangent space for the submodel. The supremum of the CR bounds for all submodels in a class is therefore less than or equal to the variance of the projection of $\psi(X)$ onto the smallest closed linear space that includes the tangent spaces for all submodels.

$\downarrow$

For an arbitrary parameter $\beta(F)$ the supremum of the CR bounds is finite so long as the parameter is pathwise differentiable.

In our previous argument that the bound for the estimand $\beta(F) = E_F[\psi(X)]$ was finite the key point was that

$$C_{\mathcal{F}_{\text{sub}}}(F) = \operatorname{var}_F \left\{ \Pi_F \left[\psi(X) \mid \Lambda_{\mathcal{F}_{\text{sub}}}(F) \right] \right\}.$$

This identity says that the CR bound for $\beta(F) = E_F[\psi(X)]$ in ANY reg. parametric submodel in the class $\mathcal{A}$ is the variance of the projection of THE SAME RANDOM VARIABLE $\psi(X)$ into the tangent space for the submodel, regardless of the submodel.

This allowed us to deduce that the supremum of the CR bounds had to be less than or equal to the projection of the SAME RANDOM VARIABLE $\psi(X)$ into the tangent space $\Lambda_{\mathcal{F}}(F)$ of the semiparametric model.

For an arbitrary parameter

$$\beta : \mathcal{F} \to \mathbb{R},$$

we can reason similarly. Suppose that we can find a random variable $\psi_F(X)$ possibly depending on F, such that for all reg. parametric submodels $\mathcal{F}_{\text{sub}}$ in a given class $\mathcal{A}$

$$C_{\mathcal{F}_{\text{sub}}}(F) = \operatorname{var}_F \left\{ \Pi_F \left[\psi_F(X) \mid \Lambda_{\mathcal{F}_{sub}}(F) \right] \right\}.$$

Then, reasoning identically as in slide 35 we would conclude that

$$\sup_{\mathcal{F}_{\text{sub}}} C_{\mathcal{F}_{\text{sub}}}(F) \leq \operatorname{var}_F \left\{ \Pi_F \left[\psi_F(X) \mid \Lambda_{\mathcal{F}}(F) \right] \right\} < \infty.$$

We will see next that a sufficient condition for the existence of such $\psi_F(X)$ is that there exists $\psi_F(X)$ such that for every $\mathcal{F}_{\text{sub}}$ in $\mathcal{A}$ indexed by, say θ, and with $F^* = F_{\theta^*}$ at θ^*, the map $\theta \to \beta(\theta) \equiv \beta(F_\theta)$ is differentiable at θ^* and satisfies

$$\left. \frac{\partial}{\partial \theta^T} \beta(\theta) \right|_{\theta=\theta^*} = E_{\theta^*} \left[\psi_F(X) S_\theta(\theta^*)^T \right].$$

When such a function $\psi_F(X)$ exists the parameter $\beta(\cdot)$ is said to be pathwise differentiable or regular with respect to $\mathcal{A}$ at F and $\psi_F(X)$ is called a gradient of $\beta(\cdot)$ at F w.r.t. $\mathcal{A}$ as we define next.

Definition 4.3. Given a semiparametric model $\mathcal{F}$, a law F^* in F, and a class $\mathcal{A}$ of reg. parametric submodels of $\mathcal{F}$, a real valued functional

$$\beta : \mathcal{F} \to \mathbb{R}$$

is said to be a pathwise differentiable or regular parameter at F^* w.r.t. $\mathcal{A}$ in model $\mathcal{F}$ iff there exists $\psi_{F^*}(X)$ in $\mathcal{L}_2(F^*)$ such that for each submodel in $\mathcal{A}$, say indexed by θ and with $F^* = F_{\theta^*}$ and score, say $S_\theta(\theta^*) = s_\theta(X; \theta^*)$ at θ^*, it holds that

$$\left. \frac{\partial}{\partial \theta^T} \beta(F_\theta) \right|_{\theta=\theta^*} = E_{F^*} \left[\psi_{F^*}(X) S_\theta(\theta^*)^T \right].$$

$\psi_{F^*}(\cdot)$ is called a gradient of β at F^* (w.r.t. $\mathcal{A}$).
If, in addition, $\psi_{F^*}(X)$ has mean zero under F^*, $\psi_{F^*}(X)$ is most commonly referred to as an influence function of the functional β at F^*. Notice that $\psi_{F^*}(X)$ is a r.v. related to a parameter whereas before we used the name inf. fcn for a r.v. related to an estimator (we will later see the connection between both concepts).

Definition 4.4. Suppose that each entry of an $\mathbb{R}$-valued parameter

$$\beta : \mathcal{F} \to \mathbb{R}^k$$

is a pathwise differentiable parameter in model $\mathcal{F}$ w.r.t. $\mathcal{A}$ at F. Then $\beta(\cdot)$ is said to be pathwise differentiable or regular at F in model $\mathcal{F}$ w.r.t. $\mathcal{A}$ and the vector

$$\psi_F(X) = \left[\begin{array}{cccc} \psi_{1,F}(X) & \psi_{2,F}(X) & \cdots & \psi_{k,F}(X) \end{array} \right]^T,$$

where $\psi_{j,F}(X)$ is a gradient of the j-th component of $\beta(\cdot)$ is called a gradient of $\beta(\cdot)$ at F in model $\mathcal{F}$ w.r.t. $\mathcal{A}$.

Now suppose that if β is a real-valued pathwise differentiable parameter wrt a class $\mathcal{A}$ of reg. parametric submodels through F^* at θ. If $\psi_{F*}(\cdot)$ is a gradient of β, then for any submodel $\mathcal{F}_{\text{sub}}$ in $\mathcal{A}$,

$$C_{\mathcal{F}_{\text{sub}}}(\theta) = \text{var}_\theta \left\{ \Pi_\theta \left[\psi_F(X) \mid \Lambda_{\mathcal{F}_{\text{sub}}}(\theta) \right] \right\},$$

because

$$\begin{aligned} C_{\mathcal{F}_{\text{sub}}}(\theta) &= \left.\frac{\partial \beta(\theta')}{\partial \theta'^T}\right|_{\theta'=\theta} \text{var}_\theta \left\{ S_\theta(\theta) \right\}^{-1} \left.\frac{\partial \beta(\theta')}{\partial \theta'}\right|_{\theta'=\theta} \\ &= E_\theta \left\{ \psi_F(X) S_\theta(\theta)^T \right\} \text{var}_\theta \left\{ S_\theta(\theta) \right\}^{-1} E_\theta \left\{ \psi_F(X) S_\theta(\theta)^T \right\}^T \\ &= \text{var}_\theta \left\{ \Pi_\theta \left[\psi_F(X) \mid \Lambda_{\mathcal{F}_{\text{sub}}}(\theta) \right] \right\}. \end{aligned}$$

Then

$$\sup_{\mathcal{F}_{\text{sub}}} C_{\mathcal{F}_{\text{sub}}}(F) \le \text{var}_F \left\{ \Pi_F \left[\psi_F(X) \mid \Lambda_{\mathcal{F}}(F) \right] \right\} < \infty.$$

Conclusion

The supremum of the C-R bounds over all regular parametric submodels in a class $\mathcal{A}$ for a scalar pathwise differentiable parameter w.r.t. $\mathcal{A}$ is finite and it is less than or equal to the variance of the projection of any gradient for the parameter into the tangent space for the model w.r.t. $\mathcal{A}$.

Definition 4.5. Given a class $\mathcal{A}$ of parametric submodelsof a semi-parametric model $\mathcal{F}$, the semiparametric Cramer-Rao bound w.r.t. $\mathcal{A}$ of a regular $\mathbb{R}^k$-valued parameter $\beta(\cdot)$ at F with gradient $\psi_F(X)$ is defined as

$$C_{\mathcal{F}}(F) \equiv \text{var}_F \left\{ \Pi \left[\psi_F(X) \mid \Lambda_{\mathcal{F}}(F) \right] \right\},$$

where $\Lambda_{\mathcal{F}}(F)$ is the tangent space of the model F w.r.t. the class $\mathcal{A}$ at F.

Outline

Example 4.3 (mean functional). Let $\mathcal{F}$ be the "non-parametric" model for a random variable X restricted only by the condition that $E_F\left[b(X)^2\right] < \infty$ for all F in $\mathcal{F}$, where $b(\cdot)$ is a given real valued function. Let $\beta(F) = E_F[b(X)]$. Let $\mathcal{A}$ be the class of all regular parametric submodels such that for each submodel, say indexed, say by θ, with $F^* = F_{\theta^*}$, the map

$$\theta \to E_\theta\left[b(X)^2\right]$$

is continuous in an open neighborhood of θ^*.

We will compute the gradient of $\beta(F)$ w.r.t. θ at F^*.

By Lemma 3.3, we have that for any submodel in the class $\mathcal{A}$ with score $s_\theta(X;\theta^*) = S_\theta(\theta^*)$,

$$\left.\frac{\partial}{\partial\theta^T}\beta(F_\theta)\right|_{\theta=\theta^*} = E_{\theta^*}\left[b(X)S_\theta(\theta^*)^T\right].$$

Then, because $E_{\theta^*}\left[S_\theta(\theta^*)\right] = 0$, we have

$$\left.\frac{\partial}{\partial\theta^T}\beta(F_\theta)\right|_{\theta=\theta^*} = E_{\theta^*}\left[\{b(X) - E_{\theta^*}[b(X)]\}S_\theta(\theta^*)^T\right],$$

from where we conclude that $b(X)$ is a gradient and $\psi_F^*(X) = b(X) - E_{\theta^*}[b(X)]$ is a mean zero gradient of $\beta(F)$ at F^*.

Note that as we have shown before, the tangent space corresponding to the class $\mathcal{A}$ is $\mathcal{L}_2^0(F^*)$, then the (non- parametric) C-R bound for estimating the mean functional is equal to the variance of

$$\psi_F^*(X) = b(X) - E_{\theta^*}[b(X)].$$

This happens because $\psi_F^*(X)$ is a gradient and

$$\psi_F^*(X) = \Pi\left[\psi_F^*(X) \mid \mathcal{L}_2^0(F^*)\right].$$

Therefore, the non-parametric C-R bound for the population mean estimand at F^* is equal to

$$C_{\mathcal{F}}(F^*) = \operatorname{var}_{F^*}[b(X)].$$

Now, suppose that F^* is the $N\left(\mu^*, \sigma^{*2}\right)$ distribution and $b(X) = x$.

The C-R bound for estimating $\beta(F) = E_F(X)$ in the non-parametric model at F^* is $\sigma^{*2} = \mathrm{var}_{F^*}(X)$ which is the same as the C-R bound for estimating the mean under the parametric normal submodel

$$\mathcal{F}^*_{\text{sub}} = \left\{N\left(\mu, \sigma^2\right) : \mu \text{ in } \mathbb{R}, \sigma > 0\right\}$$

at F^*.

Then, the normal submodel $\mathcal{F}^*_{\text{sub}}$ is a least favorable submodel.

In contrast, suppose that F^* is the $\log N(\mu^*, \sigma^{*2})$ distribution and, again, $b(x) = x$.

The C-R bound for estimating $\beta(F) = E_F(X)$ in the non-parametric model at F^* is

$$\{\exp(\sigma^{*2}) - 1\} \exp\left(2\mu^{*2} + \sigma^{*2}\right) = \mathrm{var}_{F^*}(X),$$

but this is not the same as the C-R bound for $\beta(F)$ at F^* under the lognormal parametric submodel

$$\mathcal{F}^*_{\text{sub}} = \left\{\log N\left(\mu, \sigma^2\right) : \mu \text{ in } \mathbb{R}, \sigma > 0\right\}.$$

(you will check this as hmw)

So $\mathcal{F}^*_{\text{sub}}$ is not least favorable for the mean functional.

Example 4.4 (median functional). Let $\mathcal{F}$ be the "non- parametric model" for a random variable X absolutely continuous with respect to the Lebesgue measure, restricted only by the requirement that the density $f(x)$ be continuous and satisfies $f(\beta(F)) \neq 0$ where $\beta(F)$ is the median functional, i.e.

$$\beta(F) \text{ solves } F(\beta) = 1/2.$$

Let $\mathcal{A}$ be the class of all regular parametric submodels indexed, say by θ, with $F^* = F_{\theta^*}$, and such that the map

$$(\theta, u) \to F_\theta(u),$$

has continuous partial derivatives in an open neighborhood of (θ^*, u^*) with $u^* = \beta(F_{\theta^*})$.

The assumption on the map $(\theta, u) \to F_\theta(u)$ is continuously differentiable and the assumption that $f(\beta(F)) \neq 0$ implies, by the Implicit Function Theorem, that the map $\theta \to \beta(F_\theta))$ is differentiable at θ^* and it holds that

$$\begin{aligned} 0 &= d\{1/2\}/\, d\theta|_{\theta=\theta^*} \\ &= dF_\theta\left(\beta\left(F_\theta\right)\right)/\, d\theta|_{\theta=\theta^*} \\ &= dF_\theta\left(\beta\left(F_{\theta^*}\right)\right)/\, d\theta|_{\theta=\theta^*} + dF_{\theta^*}(\beta)/\, d\beta|_{\beta=\beta(\theta^*)}\, d\beta\left(F_\theta\right)/\, d\theta|_{\theta=\theta^*}. \end{aligned}$$

Then,

$$\left.\frac{d\beta(\theta)}{d\theta}\right|_{\theta=\theta^*} = -\left\{\left.\frac{d}{d\beta}F_{\theta^*}(\beta)\right|_{\beta=\beta(\theta^*)}\right\}^{-1} \left.\frac{d}{d\theta}F_\theta\left(\beta\left(\theta^*\right)\right)\right|_{\theta=\theta^*}.$$

Now, with

$$b(X) \equiv \mathrm{I}\left(X \leq \beta\left(F_{\theta^*}\right)\right),$$

we have

$$\left.\frac{d}{d\theta} F_\theta\left(\beta\left(F_{\theta^*}\right)\right)\right|_{\theta=\theta^*} = \left.\frac{d}{d\theta} E_\theta[b(X)]\right|_{\theta=\theta^*},$$

and since $b(X)^2 = b(X)$, the assumptions on the map $(\theta, u) \to F_\theta(u)$ imply that the map $\theta \to E_\theta\left[b(X)^2\right]$ is continuous in a neighborhood of θ^*, so by Lemma 3.3,

$$\left.\frac{d}{d\theta} E_\theta[b(X)]\right|_{\theta=\theta^*} = E_{\theta^*}\left[b(X) S_\theta\left(\theta^*\right)\right].$$

Also,

$$\left.\frac{d}{d\beta} F_\theta(\beta)\right|_{\beta=\beta\left(F_{\theta^*}\right)} = f^*\left(\beta\left(F_{\theta^*}\right)\right).$$

Therefore from

$$\left.\frac{d\beta(\theta)}{d\theta}\right|_{\theta=\theta^*} = -\left\{\left.\frac{d}{d\beta} F_{\theta^*}(\beta)\right|_{\beta=\beta(\theta^*)}\right\}^{-1} \left.\frac{d}{d\theta} F_\theta\left(\beta\left(\theta^*\right)\right)\right|_{\theta=\theta^*},$$

we obtain

$$\left.\frac{\partial}{\partial\theta} \beta\left(F_\theta\right)\right|_{\theta=\theta^*} = -E_{\theta^*}\left[\frac{\mathrm{I}\left(X \leq \beta\left(F_{\theta^*}\right)\right)}{f^*\left(\beta\left(F_{\theta^*}\right)\right)} S_\theta\left(\theta^*\right)\right].$$

Then, $\frac{\mathrm{I}(X \leq \beta(F_{\theta^*}))}{f^*(\beta(F_{\theta^*}))}$ is a gradient at $F^* = F_{\theta^*}$ and

$$\psi_{F^*}(X) = -\frac{\mathrm{I}\left(X \leq \beta\left(F_{\theta^*}\right)\right) - 1/2}{f^*\left(\beta\left(F_{\theta^*}\right)\right)},$$

is a mean zero gradient at F^*.

It is not difficult to show that in models of the form

$$f(x;\theta) = f^*(x) c(\theta) k(\theta g(x)), \theta \in (-\varepsilon, \varepsilon),$$

where $k(u)$ is previously defined and $g(X)$ is any r.v. in $\mathcal{L}_2^0\left(F^*\right)$ it holds that the map

$$(\theta, u) \to F_\theta(u),$$

has continuous partial derivatives in an open neighborhood of $(0, u^*)$ with $u^* =$ median of F_{θ^*}. Thus the model is in the class $\mathcal{A}$. Since this model has score $g(x)$ and g is any r.v. in $\mathcal{L}_2^0\left(F^*\right)$ this shows that the tangent space wrt class $\mathcal{A}$ is $\mathcal{L}_2^0\left(F^*\right)$. Thus, the non-parametric C-R bound for the population median is equal to the variance of

$$\psi_{F^*}(X) = -\frac{\mathrm{I}\left(X \leq \beta\left(F_{\theta^*}\right)\right) - 1/2}{f^*\left(\beta\left(F_{\theta^*}\right)\right)}.$$

Thus the non-parametric CR bound for the population median is

$$\begin{aligned} C_{\mathcal{F}}\left(F^*\right) &= \operatorname{var}_{F^*}\left[\frac{\mathrm{I}\left(X \leq \beta\left(F^*\right)\right) - 1/2}{f^*\left(\beta\left(F^*\right)\right)}\right] \\ &= \frac{E_{F^*}\left[\mathrm{I}\left(X \leq \beta\left(F^*\right)\right)\right]\left\{1 - E_{F^*}\left[\mathrm{I}\left(X \leq \beta\left(F^*\right)\right)\right]\right\}}{f^*\left(\beta\left(F^*\right)\right)^2} \\ &= \frac{1/4}{f^*\left(\beta\left(F^*\right)\right)^2}. \end{aligned}$$

Both the mean and median functional solve moment equations and the CR bounds for them can be seen as special cases of the CR bound for functionals that solve moment equations derived in the next example.

Example 4.5. Let $(X, \beta) \to u(X; \beta)$ be an $\mathbb{R}^k$-valued map where β is in $\mathbb{R}^k$ and X is a random vector. Consider the model $\mathcal{F}$ comprise by all laws F such that the equation in β

$$E_F[u(X; \beta)] = 0,$$

has a unique solution, denoted with $\beta(F)$ and such that
1) the map $\beta \to E_F[u(X; \beta)]$ has continuous partial derivatives in an open neighbor of $\beta(F)$,
2) $\partial E_F[u(X; \beta)] / \left. \partial \beta^T \right|_{\beta=\beta(F)}$ is non-singular,
3)

$$E_F \left[u(X; \beta(F))^T u(X; \beta(F)) \right] < \infty.$$

Example 4.5 includes example 4.3 if we take

$$u(X; \beta) = X - \beta,$$

and example 4.4 if we take

$$u(X; \beta) = 2\mathrm{I}(X \le \beta) - 1.$$

It does include many other functionals that we care in applications. For instance, it includes the minimum distance (least squares) functional (where Y is a scalar outcome and Z is a covariate vector)

$$\beta(F) = \operatorname{argmin}_\beta E_F[Y - m(Z; \beta)]^2.$$

For this functional $X = (Y, Z^T)$ and

$$u(X; \beta) = \{\partial m(X; \beta)/\partial \beta\}\{Y - m(Z; \beta)\}.$$

Let $\mathcal{A}$ be the class of all regular parametric submodels through F^*, say, indexed by θ and with $F^* = F_{\theta^*}$, such that

a) the map

$$(\theta, \beta) \to E_\theta[u(X; \beta)]$$

has continuous partial derivatives in an open neighborhood of (θ^*, β^*) where $\beta^* = \beta(F^*)$ and

b) the map

$$\theta \to E_\theta \left[u(X; \beta^*) \, u(X; \beta^*)^T \right]$$

is continuous in an open neighborhood of θ^*.

Under condition (a) on the class $\mathcal{A}$, and the assumption of non- singularity of the Jacobian matrix of the map $\beta \to E_{F*}[u(X; \beta)]$ at β^*, the implicit function theorem implies that in a neighborhood of θ^*, the map $\theta \to \beta(\theta) \equiv \beta(F_\theta)$ is differentiable and

$$\begin{aligned} 0 &= \partial E_\theta[u(X; \beta(\theta))] / \left. \partial \theta^T \right|_{\theta=\theta^*} \\ &= \partial E_\theta \left[u\left(X; \beta\left(\theta^*\right)\right) \right] / \left. \partial \theta^T \right|_{\theta=\theta^*} \\ &\quad + \partial E_{\theta^*}[u(X; \beta)] / \left. \partial \beta^T \right|_{\beta=\beta(\theta^*)} \partial \beta(\theta) / \left. \partial \theta^T \right|_{\theta=\theta^*}. \end{aligned}$$

Under the condition (b) of the class $\mathcal{A}$, by Lemma 3.3 we have

$$\partial E_\theta \left[u\left(X; \beta\left(\theta^*\right)\right) \right] / \left. \partial \theta^T \right|_{\theta=\theta^*} = E_{\theta^*} \left[u\left(X; \beta\left(\theta^*\right)\right) S_\theta\left(\theta^*\right)^T \right],$$

from where we conclude that

$$\begin{aligned} &\partial \beta(\theta) / \left. \partial \theta^T \right|_{\theta=\theta^*} \\ &= - \left\{ \partial E_{\theta^*}[u(X; \beta)] / \left. \partial \beta^T \right|_{\beta=\beta(\theta^*)} \right\}^{-1} E_{\theta^*} \left[u\left(X; \beta\left(\theta^*\right)\right) S_\theta\left(\theta^*\right)^T \right]. \end{aligned}$$

We thus conclude that

$$\psi_{F^*}(X) = -\left\{\partial E_{F^*}[u(X;\beta)]/\ \partial\beta^T\big|_{\beta=\beta(F^*)}\right\}^{-1} u\left(X;\beta\left(F^*\right)\right),$$

is a mean zero gradient of $\beta(F)$ w.r.t. to the class $\mathcal{A}$ at F^*.

As in examples 4.3 and 4.4, it can be shown that the tangent space of class $\mathcal{A}$ is equal to $\mathcal{L}_2^0(F^*)$.

So the non-parametric CR bound is equal to

$$\begin{aligned}&\operatorname{Var}_{F^*}\left[\psi_{F^*}(X)\right]\\ =&\left\{\partial E_{F^*}[u(X;\beta)]/\ \partial\beta^T\big|_{\beta(F^*)}\right\}^{-1} V \left\{\partial E_{F*}[u(X;\beta)]/\ \partial\beta|_{\beta(F^*)}\right\}^{-1},\end{aligned}$$

where $V = \operatorname{Var}_{F^*}\left[u\left(X;\beta\left(F^*\right)\right)\right]$.

Outline

The following Lemma states that, except when the tangent set for the model is $\mathcal{L}_2^0(F)$ there are infinitely many mean zero gradients.

Lemma 4.1. Let $\beta(\cdot)$ be an $\mathbb{R}^k$-valued pathwise differentiable parameter in model $\mathcal{F}$ at F and let $\psi_F(X)$ be a gradient of $\beta(\cdot)$ at F w.r.t. $\mathcal{A}$. Then

$$\psi_F'(X) \equiv \psi_F(X) + b_F(X)$$

is a gradient of $\beta(F)$ at F w.r.t. $\mathcal{A}$ if and only if

$$b_F(X) \in \Lambda_{\mathcal{F}}(F)^{\perp},$$

where $\Lambda_{\mathcal{F}}(F)$ is the tangent space w.r.t. $\mathcal{A}$ at F and $b_F(X) \in \Lambda_{\mathcal{F}}(F)^{\perp}$ means that every component of $b_F(X)$ is orthogonal to $\Lambda_{\mathcal{F}}(F)$.

Proof of Lemma 4.1

$\Rightarrow$) Let $\psi_F(X)$ be a gradient of $\beta(.)$ at F^* and let $\psi_F'(X) = \psi_F(X) + b(X)$ where $b(X) \in \Lambda_{\mathcal{F}}^{\perp}(F^*)$. Let $s_\theta\left(X;\theta^*\right)$ be a score in a parametric submodel $f(X;\theta)$ at θ^*. Then, since $\psi_F(X)$ is a gradient of $\beta(\cdot)$ at F^*,

$$\begin{aligned}\partial\beta(\theta)/\ \partial\theta^T\big|_{\theta^*} &= E_{\theta^*}\left[\psi_F(X)s_\theta\left(X;\theta^*\right)^T\right]\\ &= E_{\theta^*}\left[\psi_F(X)s_\theta\left(X;\theta^*\right)^T\right] + E_{\theta^*}\left[b(X)s_\theta\left(X;\theta^*\right)^T\right]\\ &= E_{\theta^*}\left[\left\{\psi_F(X)+b(X)\right\}s_\theta\left(X;\theta^*\right)^T\right]\\ &= E_{\theta^*}\left[\psi_F'(X)s_\theta\left(X;\theta^*\right)^T\right],\end{aligned}$$

So, $\psi_F'(X)$ is a gradient of $\beta(F)$ at F w.r.t. $\mathcal{A}$.

Proof of Lemma 4.1

$\Leftarrow$) If $\psi'(X) = \psi_F(X) + b(X)$ is a gradient and $\psi_F(X)$ is a gradient, then, for any score $s_\theta(X;\theta^*)$,

$$\begin{aligned} E_{\theta^*}\left[b(X)s_\theta(X;\theta^*)^T\right] &= E_{\theta^*}\left[\{\psi'_F(X) - \psi_F(X)\} S_\theta(X;\theta^*)^T\right] \\ &= E_{\theta^*}\left[\psi'_F(X)S_\theta(X;\theta^*)^T\right] - E_{\theta^*}\left[\psi_F(X)s_\theta(X;\theta^*)^T\right] \\ &= \partial\beta(\theta)/\,\partial\theta^T\big|_{\theta^*} - \partial\beta(\theta)/\,\partial\theta^T\big|_{\theta^*} \\ &= 0. \end{aligned}$$

Then $b(X) \in \Lambda_{\mathcal{F}}^{\perp}(F^*)$.

Notational Remark

We will denote the set of all mean zero gradients at F for a pathwise diffen. par. $\beta(F)$ in model $\mathcal{F}$ at F with

$$\overset{0}{\mathrm{IF}}_{\mathcal{F}}(F).$$

We will show shortly that the set of all influence functions of RAL estimators of a parameter $\beta(F)$ is included $\mathrm{IF}_{\mathcal{F}}(F)$.

It is for this reason that $\overset{0}{\mathrm{IF}}_{\mathcal{F}}(F)$ is often referred to as the INFLUENCE FUNCTION SET. However, it may happen that for a given gradient there may not exist any RAL estimator with inf. function equal to that gradient, so I prefer to use different terminology for each concept.

If $\Lambda_{\mathcal{F}}(F^*)$ is a tangent space wrt a class $\mathcal{A}$ under which an $\mathbb{R}^k$ valued parameter $\beta(F)$ is pathwise differentiable at F^* and if, in an abuse of notation, we let $\Lambda_{\mathcal{F}}^{\perp}(F^*)$ denote the collection of all random vectors which have mean zero and finite variance under F^* and which are uncorrelated under F^* with the r.v. in $\Lambda_{\mathcal{F}}(F^*)$, Lemma 4.1 implies that

$$\overset{0}{\mathrm{IF}}_{\mathcal{F}}(F) = \{\psi_F(X)\} + \Lambda_{\mathcal{F}}(F)^{\perp},$$

where $\psi_F(X)$ is any mean zero gradient of $\beta(\cdot)$ w.r.t. a class $\mathcal{A}$ at F^*.

Lemma 4.1 implies that for any pair of gradients $\psi'_F(X)$ and $\psi_F(X)$

$$\Pi_F\left[\psi'_F(X) - \psi_F(X) \mid \Lambda_{\mathcal{F}}(F)\right] = 0,$$

or equivalently,

$$\Pi_F\left[\psi'_F(X) \mid \Lambda_{\mathcal{F}}(F)\right] = \Pi_F\left[\psi_F(X) \mid \Lambda_{\mathcal{F}}(F)\right].$$

Definition 4.6. The projection of any gradient into the tangent space for the model $\mathcal{F}$ w.r.t. a class $\mathcal{A}$ is called the efficient influence function for the parameter $\beta(F)$ in the semiparametric model $\mathcal{F}$ w.r.t. $\mathcal{A}$ at F. We shall denote it with $\psi_{F,\mathsf{eff}}(X)$, i.e., for any gradient $\psi_F(X)$,

$$\psi_{F,\mathsf{eff}}(X) \equiv \Pi_F\left[\psi_F(X) \mid \Lambda_{\mathcal{F}}(F)\right].$$

Lemma 4.2. $\psi_{F,\text{eff}}(X)$ is a gradient.

Proof.

$$\begin{aligned}\psi_F(X) &= \Pi_F\left[\psi_F(X) \mid \Lambda_{\mathcal{F}}(F)\right] + \left\{\psi_F(X) - \Pi_F\left[\psi_F(X) \mid \Lambda_{\mathcal{F}}(F)\right]\right\} \\ &= \psi_{F,\text{eff}}(X) + \Pi_F\left[\psi_F(X) \mid \Lambda_{\mathcal{F}}(F)^{\perp}\right].\end{aligned}$$

Then

$$\psi_{F,\text{eff}}(X) = \psi_F(X) + \Pi\left[\psi_F(X) \mid \Lambda_{\mathcal{F}}^{\perp}(F)\right],$$

and since $\Pi\left[\psi_F(X) \mid \Lambda_{\mathcal{F}}^{\perp}(F)\right]$ belongs to $\Lambda_{\mathcal{F}}{}^{\perp}(F)$, then by Lemma 4.1, $\psi_{F,\text{eff}}(X)$ is a gradient. □

Corollary

$$\overset{0}{\mathrm{IF}}_{\mathcal{F}}(F) = \{\psi_{F,\text{eff}}(X)\} \oplus \Lambda_{\mathcal{F}}(F)^{\perp}.$$

Remark: We have earlier shown that in the non-parametric model the maximal tangent space is

$$\Lambda_{\mathcal{F}}(F) = \mathcal{L}_2^0(F).$$

Then $\Lambda_{\mathcal{F}}(F)^{\perp} = \{0\}$ and consequently, for any reg. parameter w.r.t. to a class $\mathcal{A}$ corresponding to the maximal tangent space

$$\begin{aligned}\overset{0}{\mathrm{IF}}_{\mathcal{F}}(F) &= \{\psi_{F,\text{eff}}(X)\} \oplus \Lambda_{\mathcal{F}}(F)^{\perp} \\ &= \{\psi_{F,\text{eff}}(X)\}.\end{aligned}$$

We conclude that in the non-parametric model for any pathwise differentiable parameter w.r.t. a class $\mathcal{A}$ corresponding to the maximal tangent space, there exists a unique mean zero gradient which is equal to the efficient influence function.

Example 4.5 (continued). In these examples we have exhibited a mean zero gradient, of the functional $\beta(F)$ at F^*, namely,

$$\psi_{F^*}(X) = -\left\{\partial E_{F^*}[u(X;\beta)]/\,\partial\beta^T\big|_{\beta=\beta(F^*)}\right\}^{-1} u\left(X;\beta\left(F^*\right)\right).$$

Because this was a gradient w.r.t. to a class of parametric submodels corresponding to the maximal tangent space in the non-parametric model, then we conclude that this is the UNIQUE mean zero gradient of $\beta(F)$ at F^*.

Technical remark on the class of parametric submodels and its tangent space

Consider two classes of reg. parametric submodels, say $\mathcal{A}$ and $\mathcal{A}'$, of a model $\mathcal{F}$ through F^* such that their tangent spaces agree and suppose that $\beta(F)$ is pathwise differentiable w.r.t. each class.

Then, it can be shown that a gradient $\psi_{F^*}(X)$ of $\beta(F)$ at F^* wrt $\mathcal{A}$ is also a gradient at F^* wrt $\mathcal{A}'$, under a stronger requirement than pathwise differentiability, namely, the requirement that the map from $\mathcal{L}_2(\mu)$ to $\mathbb{R}$:

$$\sqrt{f} \to \beta(\sqrt{f}),$$

is Frechet differentiable, i.e. there exists a continuous linear map $\dot{B}$ from $\mathcal{L}_2(F^*)$ to $\mathbb{R}$ such that

$$\beta(\sqrt{f}) - \beta\left(\sqrt{f^*}\right) = \dot{B}\left(\sqrt{f} - \sqrt{f^*}\right) + o\left(\left\|\sqrt{f} - \sqrt{f^*}\right\|^2_{\mathcal{L}_2(F^*)}\right).$$

Van der Vaart, 2000, ch 25 talks about pathwise differentiability of a parameter and its influence functions (i.e. gradients) with respect to a tangent space, rather than with respect to a class of parametric models.

Although he does not say it, he is probably implicitly making the aforementioned Frechet-differentiability assumption.

From now on, we will make this assumption and then interchangeably mention pathwise differentiability or regularity and gradients with respect to a tangent space or to a class of parametric submodels.

Outline

We have seen in chapter 3 that the set of influence functions of RAL estimators of a differentiable parameter in parametric models can be constructed by adding to one influence function any element of the orthogonal complement of the tangent space.

We derived this characterization from Lemma 3.5. Indeed, such Lemma applies also to RAL estimators of parameters in semiparametric models.

The next Lemma essentially states that the set of influence functions of RAL estimators for a pathwise differentiable parameter is included in the set of mean zero gradients for the parameter.

Lemma 4.3. Let $\mathcal{F}$ be a semiparametric model. Suppose that $\widehat{\beta}_n$ is an asymptotically linear estimator at F of a parameter

$$\beta(\cdot): \mathcal{F} \to \mathbb{R}^k,$$

with influence function $\varphi_F(X)$. Suppose that for every regular parametric submodel in a class $\mathcal{A}$ indexed by θ that goes through F at $\theta^*, \beta(\theta)$ is differentiable at θ^*.
Then $\widehat{\beta}_n$ is a regular estimator if and only if for all reg. parametric submodels in $\mathcal{A}$ with score $S_\theta(\theta^*)$ at θ^*

$$\left.\frac{\partial \beta(\theta)}{\partial \theta^T}\right|_{\theta=\theta^*} = E_{\theta^*}\left\{\varphi_F(X) S_\theta(\theta^*)^T\right\}.$$

1

[1] This is Lemma 25.23 of van der Vaart, 2000.

In words, Lemma 4.3 says that an asymptotically linear estimator of a parameter $\beta(F)$ is regular wrt a class $\mathcal{A}$ of parametric models with $\beta(\theta)$ differentiable iff $\beta(F)$ is a pathwise differentiable parameter at F in model $\mathcal{F}$ and its influence function is a gradient of $\beta(F)$ at F.

We can immediately deduce that the variance of the efficient influence function is a lower bound for the asymptotic variance of any RAL estimator, since

a) the variance of the limiting normal distrib. of a RAL estimator with influence function $\varphi_F(X)$ is $\operatorname{var}_F\{\varphi_F(X)\}$,

b) $\varphi_F(X)$ is a gradient (by the previous theorem) and therefore

$$\psi_{F,\text{eff}}(X) = \Pi_F\left[\varphi_F(X) \mid \Lambda_{\mathcal{F}}(F)\right],$$

and

c) projections "contract" the length of vectors.

RAL estimators in the non-parametric model

We have seen that in the non-parametric model for any pathwise differentiable parameter w.r.t. a class $\mathcal{A}$ corresponding to the maximal tangent space, there exists a unique mean zero gradient which is equal to the efficient influence function.

It follows from Lemma 4.3, that all estimators of such a parameter that are RAL w.r.t. to such class $\mathcal{A}$ must have the same influence function, namely the efficient influence function, and consequently they are all asymptotically equivalent and efficient.

Example 4.5 (continued). Recall in example 4.5 we considered an $\mathbb{R}^k$-valued map $(X, \beta) \to u(X; \beta)$ where β is in $\mathbb{R}^k$ and X is a random vector. The model $\mathcal{F}$ was comprised by all laws F such that the equation in β

$$E_F[u(X; \beta)] = 0,$$

has a unique solution, denoted with $\beta(F)$ and such that
1) the map

$$\beta \to E_F[u(X; \beta)]$$

has continuous partial derivatives in an open neighbor of $\beta(F)$,
2) $\partial E_F[u(X; \beta)] / \left.\partial \beta^T\right|_{\beta=\beta(F)}$ is non-singular,
3)

$$E_F\left[u(X; \beta(F))^T u(X; \beta(F))\right] < \infty.$$

Let $\mathcal{A}$ be the class of all regular parametric submodels through F^*, say, indexed by θ and with $F^* = F_{\theta^*}$, such that

a) the map

$$(\theta, \beta) \to E_\theta[u(X; \beta)],$$

has continuous partial derivatives in an open neighborhood of (θ^*, β^*) where $\beta^* = \beta(F^*)$ and

b) the map

$$\theta \to E_\theta\left[u(X; \beta^*) u(X; \beta^*)^T\right],$$

is continuous in an open neighborhood of θ^*.

We concluded that

$$\psi_{F^*}(X) = -\left\{\partial E_{F*}[u(X; \beta)] / \left.\partial \beta^T\right|_{\beta=\beta(F^*)}\right\}^{-1} u(X; \beta(F^*))$$

is the unique mean zero gradient of $\beta(F)$ at F^* w.r.t. to the class $\mathcal{A}$.

Is there a RAL estimator with inf. fcn equal to $\psi_{F^*}(X)$?

Under reg. Conditions: YES. In proposition 3.4 we established that under certain reg. conditions, the Z-estimator satisfying

$$n^{-1/2} \sum_{i=1}^{n} u\left(X_i; \widehat{\beta}_n\right) = o_p(1)$$

is asymptotically linear at F^* with inf. function equal to $\psi_{F*}(X)$.

It follows from Lemma 4.3 that the Z-estimator $\widehat{\beta}_n$ is not only asymptotically linear but also regular.

Note that since $\psi_{F^*}(X)$ is the unique gradient of $\beta(F)$ w.r.t. to the class $\mathcal{A}$ at F^*, then any RAL estimator of $\beta(F)$ must necessarily have the influence function equal to $\psi_{F^*}(X)$.

We thus conclude that any other RAL estimator $\widetilde{\beta}_n$ of $\beta(F)$ must have the same limiting distribution as $\widehat{\beta}_n$. In fact, it must be asymptotically equivalent to $\widehat{\beta}_n$, i.e.

$$\sqrt{n}\left\{\widehat{\beta}_n - \widetilde{\beta}_n\right\} \xrightarrow{P} 0.$$

Outline

The Convolution Theorem

Let $\mathcal{F}$ be a semiparametric model. Let

$$\beta(\cdot) : \mathcal{F} \to \mathbb{R}^k,$$

be a pathwise differentiable parameter at F with efficient influence function $\psi_{F,\text{eff}}(X)$ w.r.t. a class $\mathcal{A}$.

If $\widehat{\beta}_n$ is regular at F and the tangent set w.r.t at F is convex, then

$$\sqrt{n}\left\{\widehat{\beta}_n - \beta(F)\right\} \overset{D(F)}{\to} U + U^*,$$

where

$$U^* \sim N_k\left(0, C_{\mathcal{F}}(F)\right),$$

$C_{\mathcal{F}}(F) = \text{var}_{\mathcal{F}}\left[\psi_{F,\text{eff}}(X)\right]$ and U is independent of U^*. [1]

[1] This is theorem 25.20 of van der vaart, 2000.

Definition 4.7. Let $\beta(F)$ be a parameter such that at each F in $\mathcal{F}$ it is a pathwise differentiable w.r.t. to a class $\mathcal{A}_F$.
An estimator $\widehat{\beta}_n$ is a locally asymptotically efficient estimator of $\beta(F)$ in $\mathcal{F}$ at F^* w.r.t. $\mathcal{A}_{F^*}$, iff
a) it is a regular estimator of $\beta(\cdot)$ in model $\mathcal{F}$ at every F in $\mathcal{F}$ w.r.t. $\mathcal{A}_F$, and
b) it satisfies

$$\sqrt{n}\left\{\widehat{\beta}_n - \beta\left(F^*\right)\right\} \underset{n\to\infty}{\overset{D(F^*)}{\to}} N_k\left(0, C_{\mathcal{F}}(F^*)\right),$$

where $C_{\mathcal{F}}(F^*)$ is the semiparametric C-R bound.

Definition 4.8. Let $\mathcal{F}$ be a semip. model and $\mathcal{F}^* \subseteq \mathcal{F}$. For every F in $\mathcal{F}$, let $\mathcal{A}_F$, be a class of reg. param. submodels of c through F. Let $\beta(\cdot)$ be a pathwise differentiable w.r.t. to $\mathcal{A}_F$ at each F in $\mathcal{F}$.

An estimator $\widehat{\beta}_n$ is a locally asymptotically efficient estimator of $\beta(F)$ in $\mathcal{F}$ at $\mathcal{F}^*$ w.r.t. to $\{\mathcal{A}_F : F \text{ in } \mathcal{F}^*\}$ iff $\widehat{\beta}_n$ is regular in $\mathcal{F}$ and asymptotically efficient estimator of $\beta(F)$ in $\mathcal{F}$ at F w.r.t. $\mathcal{A}_F$ for every F in $\mathcal{F}^*$.

If $\mathcal{F}^* = \mathcal{F}$, then $\widehat{\beta}_n$ is called globally asymptotically efficient.

Locally, much less globally, efficient estimators of a pathwise differentiable parameter do not always exist.

However, if an efficient estimator exists, then the next Theorem states that it must be RAL and have influence function equal to the efficient influence function.

Note that this result states, just as in the parametric case, that as far as efficiency is concerned, we don't loose anything by restricting attention to regular and asymptotically linear estimators.

Lemma 4.4. Suppose that $\mathcal{F}$ is a semiparametric model and

$$\beta(\cdot) : \mathcal{F} \longrightarrow \mathbb{R}^k,$$

is a pathwise differentiable parameter at F with efficient influence function $\psi_{F,\mathsf{eff}}(X)$ w.r.t. to a class $\mathcal{A}$ with convex tangent set.

The estimator $\widehat{\beta}_n$ is locally asymptotically efficient at F w.r.t. $\mathcal{A}$ if and only if

$$\sqrt{n}\left\{\widehat{\beta}_n - \beta(F)\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{F,\mathsf{eff}}\left(X_i\right) + o_{p,F}(1).$$