Efficiency and Testing

Yu Gu

October 28, 2021

Outline

Asymptotic linearity

An estimator sequence T_n for a parameter $\psi(P)$ is *asymptotically linear* if there exists an *influence function* $\check{\psi}_P : \mathcal{X} \mapsto \mathbb{R}^k$ such that

$$\sqrt{n}\,(T_n - \psi(P)) - \sqrt{n}\mathbb{P}_n\check{\psi}_P \xrightarrow{\mathrm{P}} 0.$$

Regularity

An estimator sequence 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 $t_n = O(n^{-1/2})$,

$$\sqrt{n}\,(T_n - \psi\,(P_{t_n})) \overset{P_{t_n}}{\rightsquigarrow} Z,$$

for some tight Borel random element Z.

Asymptotic Efficiency

An estimator sequence T_n is *asymptotically efficient* at P if it is regular at P with limiting distribution

$$\sqrt{n}\,(T_n - \psi(P)) \rightsquigarrow N(0, P\tilde{\psi}_P\tilde{\psi}_P^T),$$

where $\tilde{\psi}_P : \mathcal{X} \mapsto \mathbb{R}^k$ is the *efficient influence function*.

Differentiability

$\psi : \mathcal{P} \mapsto \mathbb{R}^k$ is differentiable at P relative to the tangent set $\dot{\mathcal{P}}_P$ if, for every smooth one-dimensional submodel $\{P_t\}$ with tangent $g \in \dot{\mathcal{P}}_P$,

$$\left.\frac{d\psi(P_t)}{dt}\right|_{t=0} = \dot{\psi}_P(g)$$

for some bounded linear operator $\dot{\psi}_P : \dot{\mathcal{P}}_P \mapsto \mathbb{R}^k$.

Equivalent definition of efficiency

Theorem 1

Let the parameter $\psi : \mathcal{P} \mapsto \mathbb{R}^k$ be differentiable at P relative to the tangent space $\dot{\mathcal{P}}_P$ with efficient influence function $\tilde{\psi}_P$. A sequence of estimators T_n is efficient at P relative to $\dot{\mathcal{P}}_P$ if and only if it is asymptotically linear with influence function $\tilde{\psi}_P$.

---

* *See Lemma 25.23 (pp. 367-368) of van der Vaart (1998) for the proof.*

Remarks

- So far we have obtained useful results on efficient estimators of Euclidean parameters.
- A natural question is how to extend these results to more general parameter spaces in semiparametric models.

Composite parameter

- For example, in survival analysis, the full composite parameter is usually $\psi = (\beta, \Lambda) \in \Omega$.
- Define $\mathcal{H} = \big\{(h_1, h_2) : h_1 \in \mathbb{R}^k, h_2 \in D[0,\tau] \cap BV[0,\tau]\big\}$, equipped with norm $\|h\|_{\mathcal{H}} = \|h_1\| + \|h_2\|_{BV}$. For any $1 \le r < \infty$, define $\mathcal{H}_r = \{h \in \mathcal{H} : \|h\|_{\mathcal{H}} \le r\}$.
- ψ can be viewed as an element of $\ell^\infty(\mathcal{H}_r)$ if we define

$$\psi(h) = h_1^T\beta + \int_0^\tau h_2(s)d\Lambda(s), \quad h \in \mathcal{H}_r, \quad \psi \in \Omega.$$

- $\mathcal{H}_r$ is sufficiently rich to extract out all components of ψ. Thus, Ω becomes a subset of $\ell^\infty(\mathcal{H}_r)$ with norm $\|\psi\|_{(r)} = \sup_{h\in\mathcal{H}_r} |\psi(h)|$.

Note

For the rest of the section, we consider more general parameter spaces of the form $\ell^\infty(\mathcal{H})$. In this case, any (efficient) influence function is assumed to be a stochastic process indexed by $\mathcal{H}$.

Under this setting, ψ is differentiable $\Rightarrow \dot{\psi}_P(g)(h) = P[\tilde{\psi}_P(h)g]$ for all $h \in \mathcal{H}$.

General results

Theorem 2

Let T_n be an estimator sequence for a parameter $\psi : \mathcal{P} \mapsto \ell^\infty(\mathcal{H})$, where ψ is differentiable at P relative to the tangent space $\dot{\mathcal{P}}_P$ with efficient influence function $\tilde{\psi}_P : \mathcal{H} \mapsto L_2^0(P)$. Let $\mathcal{F} = \{\tilde{\psi}_P(h) : h \in \mathcal{H}\}$. Then the following are equivalent:

(a) *T_n is efficient at P relative to $\dot{\mathcal{P}}_P$ and at least one of the following holds:*
   (i) *T_n is asymptotically linear.*
   (ii) *$\mathcal{F}$ is P-Donsker for some version* of $\tilde{\psi}_P$.*

(b) *For some version of $\tilde{\psi}_P$, T_n is asymptotically linear with influence function $\tilde{\psi}_P$ and $\mathcal{F}$ is P-Donsker.*

(c) *T_n is regular and asymptotically linear with influence function $\check{\psi}_P$ such that $\{\check{\psi}_P(h) : h \in \mathcal{H}\}$ is P-Donsker and $\check{\psi}_P(h) \in \dot{\mathcal{P}}_P$ for all $h \in \mathcal{H}$.*

---

* *Two stochastic processes X and $\widetilde{X}$ are versions of each other if $X(h) = \widetilde{X}(h)$ almost surely for every $h \in \mathcal{H}$.*

Interpretation

- $(a) \Leftrightarrow (b)$ indicates that if T_n is efficient, only one of (i) or (ii) in (a) is required and the other will follow.

- $(c) \Rightarrow (a)$ gives a simple method for establishing efficiency of T_n, which requires only that
  - T_n be asymptotically linear
  - with an influence function that is contained in a Donsker class
  - for which the individual components $\check{\psi}_P(h)$ are contained in the tangent space for all $h \in \mathcal{H}$.

Interpretation (cont.)

- The requirement that $\mathcal{F}$ is P-Donsker collapses to requiring that $\|\tilde{\psi}_P\|_{P,2} < \infty$ when $\mathcal{H}$ is finite.
- However, such a requirement is not needed in the statement of Theorem 1 since it automatically follows from the required differentiability of ψ when $\psi(P) \in \mathbb{R}^k$.
- This follows since the Riesz representation theorem assures us that $\tilde{\psi}_P$ is in the closed linear span of $\dot{\mathcal{P}}_P$, which is a subset of $L_2(P)$.

Deep results

The following theorem tells us that pointwise efficiency implies uniform efficiency under weak convergence.

Theorem 3

Let T_n be an estimator sequence for a parameter $\psi : \mathcal{P} \mapsto \ell^\infty(\mathcal{H})$, where ψ is differentiable at P relative to the tangent space $\dot{\mathcal{P}}_P$ with efficient influence function $\tilde{\psi}_P : \mathcal{H} \mapsto L_2^0(P)$. The following are equivalent:

(a) *T_n is efficient for $\psi(P)$.*

(b) *$T_n(h)$ is efficient for $\psi(P)(h)$, for every $h \in \mathcal{H}$, and $\sqrt{n}(T_n - \psi(P))$ is asymptotically tight under P.*

Deep lemma

The proof of this theorem makes use of the following deep lemma:

Lemma 4

Suppose that $\psi : \mathcal{P} \mapsto \mathbb{D}$ *is differentiable at* P *relative to the tangent space* $\dot{\mathcal{P}}_P$ *and that* $d' T_n$ *is asymptotically efficient at* P *for estimating* $d'\psi(P)$ *for every* d' *in a subset* $\mathbb{D}' \subset \mathbb{D}^*$ *which satisfies*

$$\|d\| \leq c \sup_{d' \in \mathbb{D}', \|d'\| \leq 1} |d'(d)|, \tag{1}$$

for some constant $c < \infty$. *Then* T_n *is asymptotically efficient at* P *provided* $\sqrt{n}(T_n - \psi(P))$ *is asymptotically tight under* P.

Proof of Theorem 2

- $(a) \Rightarrow (b)$ is obvious.
- Assume (b), let $\mathbb{D} = \ell^\infty(\mathcal{H})$ and $\mathbb{D}'$ be the set of all coordinate projections $d \mapsto d(h)$ for every $h \in \mathcal{H}$.
- Since the uniform norm on $\ell^\infty(\mathcal{H})$ is trivially equal to $\sup_{d' \in \mathbb{D}'} |d'(d)|$ and $\|d'\| = 1$ for every $d' \in \mathbb{D}'$, Condition (1) is easily satisfied.
- All of the conditions in the lemma are satisfied by the assumptions in (b). Hence, T_n is efficient. □

Corollary of Lemma 4

The following corollary of Lemma 4 provides a simple connection between marginal and joint efficiency on product spaces:

Corollary 5

Suppose that $\psi_j : \mathcal{P} \mapsto \mathbb{D}_j$ is differentiable at P relative to the tangent space $\dot{\mathcal{P}}_P$, and suppose that $T_{n,j}$ is asymptotically efficient at P for estimating $\psi_j(P)$, for $j = 1, 2$. Then $(T_{n,1}, T_{n,2})$ is asymptotically efficient at P for estimating $(\psi_1(P), \psi_2(P))$.

Hahn-Banach theorem

The proof of this corollary makes use of the following theorem:

Theorem 6 (Hahn-Banach theorem)

If $\mathbb{X}$ is a normed space and $x \in \mathbb{X}$, then

$$\|x\| = \sup\left\{|f(x)| : f \in \mathbb{X}^* \text{ and } \|f\| \le 1\right\}.$$

Moreover, this supremum is attained.

Proof of Corollary 5

- Let $\mathbb{D}'$ be the set of all maps $(d_1, d_2) \mapsto d_j^* d_j$ for $d_j^* \in \mathbb{D}_j^*$ and j equal to either 1 or 2.
- By the Hahn-Banach theorem, $\|d_j\| = \sup \left\{ |d_j^*(d_j)| : \|d_j^*\| = 1, d_j^* \in \mathbb{D}_j^* \right\}$.
- Thus the product norm $\|(d_1, d_2)\| = \|d_1\| \vee \|d_2\|$ satisfies Condition 1 of Lemma 4 with $c = 1$.
- Hence the desired conclusion follows. □

Remarks

- Marginal efficiency implies joint efficiency even though marginal weak convergence does not imply joint weak convergence.
- Consider the setting where $\psi_j(P) \in \mathbb{R}$ for $j = 1, 2$. If $T_{n,j}$ is efficient for $\psi_j(P)$, then the limiting distribution of $\sqrt{n}(T_{n,j} - \psi_j(P))$ is $N(0, P\tilde{\psi}_{j,P}\tilde{\psi}_{j,P}^T)$, for $j = 1, 2$.
- Thus the limiting joint distribution will be the optimal bivariate Gaussian distribution.
- The preceding theorem can be viewed as an infinite-dimensional generalization of this finite-dimensional phenomenon.

Outline

Optimality of tests

- We will test the hypothesis

$$H_0 : \psi(P) \leq 0 \quad vs. \quad H_1 : \psi(P) > 0 \tag{2}$$

  for a one-dimensional parameter $\psi(P)$.

- Null hypotheses of the form $H_0 : \psi(P) \leq \psi_0$ can trivially be rewritten in the form given in (2) by replacing $P \mapsto \psi(P)$ with $P \mapsto \psi(P) - \psi_0$.

- We want to show the basic conclusion that a test based on an asymptotically efficient estimator for $\psi(P)$ will, in a meaningful way, be asymptotically optimal.

Local asymptotic power

- For a given model $\mathcal{P}$ and measure P on the boundary of the null hypothesis where $\psi(P) = 0$, we are interested in studying the "local asymptotic power" in a neighborhood of P.
- These neighborhoods are of size $1/\sqrt{n}$ and are the appropriate magnitude when considering sample size computation for $\sqrt{n}$ consistent parameter estimates.

Example

Consider the univariate normal setting where the data are i.i.d. $N(\mu, \sigma^2)$. A natural choice for testing $H_0 : \mu \leq 0$ vs. $H_1 : \mu > 0$ is the indicator of whether

$$T_n = \sqrt{n}\frac{\bar{x}}{s_n} > z_{1-\alpha},$$

where

- $\bar{x}$ and s_n are the sample mean and standard deviation from an i.i.d. sample $X_1, \ldots, X_n$,
- z_q is the qth quantile of a standard normal,
- α is the size of the test.

Example (cont.)

- For any $\mu > 0$, T_n diverges to infinity with probability 1.
- However, if $\mu = k/\sqrt{n}$ for some finite k, then $T_n \rightsquigarrow N(k/\sigma, 1)$.
- Thus we can derive non-trivial power functions only for shrinking "contiguous alternatives" in a $1/\sqrt{n}$ neighborhood of zero.

General contiguous alternatives

- In general, we study the performance of tests under contiguous alternatives defined by one-dimensional submodels.
- For a given element g of a tangent set $\dot{\mathcal{P}}_P$, let $t \mapsto P_{t,g}$ be a one-dimensional submodel which is differentiable in quadratic mean at P with tangent g along which ψ is differentiable, i.e.,

$$\frac{\psi\left(P_{t,g}\right) - \psi(P)}{t} \to P\left[\tilde{\psi}_P g\right]$$

as $t \downarrow 0$.

- When $\psi(P) = 0$, for each g with $P[\tilde{\psi}_P g] > 0$, the submodel $\{P_{t,g}\}$ satisfies $\psi(P_{t,g}) > 0$ for all sufficiently small $t > 0$.
- Thus, we will consider power over contiguous alternatives of the form $\{P_{h/\sqrt{n},g}\}$ for $h > 0$.

Power function

Definition 7 (Power function)

For a subset $\mathcal{Q} \subset \mathcal{P}$ containing P, a power function $\pi : \mathcal{Q} \mapsto [0, 1]$ at level α is a function on probability measures that satisfies $\pi(Q) \leq \alpha$ for all $Q \in \mathcal{Q}$ satisfying $\psi(Q) \leq 0$.

$\pi(Q)$ is the probability of rejecting $H_0 : \psi(P) \leq 0$ under Q.

We say that a sequence of power functions $\{\pi_n\}$ has asymptotic level α if $\limsup_{n\to\infty} \pi_n(Q) \leq \alpha$ for every $Q \in \mathcal{Q} : \psi(Q) \leq 0$.

Main results

The following theorem provides an upper bound for the power at the alternatives $P_{h/\sqrt{n},g}$:

Theorem 8

Let $\psi : \mathcal{P} \mapsto \mathbb{R}$ be differentiable at P relative to the tangent space $\dot{\mathcal{P}}_P$ with efficient influence function $\tilde{\psi}_P$, and suppose $\psi(P) = 0$. Then, for every sequence of power functions $P \mapsto \pi_n(P)$ of asymptotic level α tests for $H_0 : \psi(P) \leq 0$, and for every $g \in \dot{\mathcal{P}}_P$ with $P\left[\tilde{\psi}_P g\right] > 0$ and every $h > 0$,

$$\limsup_{n\to\infty} \pi_n\left(P_{h/\sqrt{n},g}\right) \leq 1 - \Phi\left[z_{1-\alpha} - h\frac{P[\tilde{\psi}_P g]}{\sqrt{P[\tilde{\psi}_P^2]}}\right].$$

* *See Lemma 25.44 (pp. 384-385) of van der Vaart (1998) for the proof.*

Tests based on efficient estimators

As a consequence of the preceding theorem, a test based on an efficient estimator for $\psi(P)$ is automatically "locally uniformly most powerful": its power function attains the upper bound.

Lemma 9

Let $\psi : \mathcal{P} \mapsto \mathbb{R}$ be differentiable at P relative to the tangent space $\dot{\mathcal{P}}_P$ with efficient influence function $\tilde{\psi}_P$, and suppose $\psi(P) = 0$. Suppose the estimator T_n is asymptotically efficient at P, and, moreover, that $S_n^2 \xrightarrow{\mathrm{P}} P\tilde{\psi}_P^2$. Then, for every $h > 0$ and $g \in \dot{\mathcal{P}}_P$,

$$\lim_{n\to\infty} \mathrm{P}_{h/\sqrt{n},g}\left(\frac{\sqrt{n}T_n}{S_n} \geq z_{1-\alpha}\right) = 1 - \Phi\left(z_{1-\alpha} - h\frac{P[\tilde{\psi}_P g]}{\sqrt{P\tilde{\psi}_P^2}}\right).$$

Proof of Lemma 9

- Let $P_n \equiv P_{h/\sqrt{n},g}$. By the efficiency of T_n,

$$\sqrt{n}(T_n - \psi(P_{h/\sqrt{n},g})) \overset{P_n}{\rightsquigarrow} N(0, P\tilde{\psi}_P^2). \qquad (3)$$

- By the differentiability of ψ,

$$\frac{\psi(P_{h/\sqrt{n},g}) - \psi(P)}{h/\sqrt{n}} \to P[\tilde{\psi}_P g], \quad \text{as } n \to \infty. \qquad (4)$$

- Combining (3) and (4) yields

$$\sqrt{n}T_n \overset{P_n}{\rightsquigarrow} N\left(hP[\tilde{\psi}_P g],\, P\tilde{\psi}_P^2\right).$$

- The desired conclusion follows since $S_n^2 \overset{\mathrm{P}}{\to} P\tilde{\psi}_P^2$. $\square$