Achieving Local Asymptotic Bounds and Extensions

John Duchi

Stats 300b – Winter Quarter 2021
Outline

- Regular estimands and estimators
- Hájek’s convolution theorem
- Optimality (achieving the local asymptotic minimax bound) for regular estimands
- Semi (non)-parametric efficiency (quite cursory)

Reading:

Regular estimands

- parametric family \( \{ P_\theta \}_{\theta \in \Theta}, \Theta \subset \mathbb{R}^d \)
- estimand \( \psi(\theta), \psi : \mathbb{R}^d \to \mathbb{R}^k \) of interest
- estimand is *regular* at \( \theta_0 \) if it is differentiable at \( \theta_0 \), derivative \( \dot{\psi}(\theta_0) \in \mathbb{R}^{k \times d} \)

**more generality possible:** (but we won’t do this)

- sequence of estimands \( \psi_n : \mathbb{R}^d \to \mathbb{R}^k \)
- estimands are *regular* for rate \( r_n \to \infty \) if

\[
    r_n(\psi_n(\theta_0 + h/r_n) - \psi_n(\theta_0)) \to \dot{\psi}(\theta_0)h
\]

for any \( h \), where \( \dot{\psi}(\theta_0) \in \mathbb{R}^{k \times d} \)
Regular estimator

Definition
An estimator sequence $T_n$ is regular at $\theta_0$ for estimating $\psi(\theta_0)$ if for each $h \in \mathbb{R}^d$,

$$\sqrt{n} \left( T_n - \psi(\theta_0 + h/\sqrt{n}) \right) \xrightarrow{d} Z_{\theta_0}$$

where $Z_{\theta_0}$ is a random vector
Regular estimator examples

Example (Typical asymptotically linear estimators)
Let family \( \{ P_\theta \}_{\theta \in \Theta} \) be QMD at \( \theta_0 \) with score \( \ell_\theta_0 \) and Fisher information \( I_\theta_0 \). If

\[
\hat{\theta}_n - \theta_0 = P_n I_\theta_0^{-1} \ell_\theta_0 + o_{P_{\theta_0}} (1/\sqrt{n})
\]

then it is regular (even more)

Example (the delta method and regular estimands)
Let setting be as above. Then \( \psi(\hat{\theta}_n) \) is regular for \( \psi(\theta_0) \).
Hájek Convolution Theorem

Theorem (Hájek)

Let $T_n$ be a regular estimator sequence for $\theta_0$ in an LAN model $\{P_\theta\}_{\theta \in \Theta}$ with information $I_{\theta_0}$. Then

$$\sqrt{n}(T_n - \theta_0) \xrightarrow{d} Z_{\theta_0} + V_{\theta_0}$$

where $Z_{\theta_0} \sim \mathcal{N}(0, I_{\theta_0})$ and $V_{\theta_0}$ are independent

▶ almost-everywhere extensions exist (see Theorem 8.9 in van der Vaart)
Achieving the local asymptotic minimax bound

**Theorem**

Let \( \hat{\theta}_n \) be any estimator of \( \theta_0 \) in LAN family with

\[
\hat{\theta}_n - \theta_0 = I_{\theta_0}^{-1} P_n \ell_{\theta_0} + o_{P_{\theta_0}} \left( 1/\sqrt{n} \right).
\]

Then for any bounded continuous \( L \),

\[
\lim_{c \to \infty} \lim_{n \to \infty} \sup_{\| h \| \leq c} \mathbb{E}_{\theta_0 + h/\sqrt{n}} \left[ L \left( \sqrt{n} (\hat{\theta}_n - (\theta_0 + h/\sqrt{n})) \right) \right] = \mathbb{E}[L(Z)]
\]

where \( Z \sim \mathcal{N}(0, I_{\theta_0}^{-1}) \).
Corollary (to local asymptotic minimax bound)

Let \( \{P_\theta\}_{\theta \in \Theta} \) be LAN at \( \theta_0 \) with Fisher information \( I_{\theta_0} \) and \( \psi : \mathbb{R}^d \to \mathbb{R}^k \) be differentiable at \( \theta_0 \). If \( L : \mathbb{R}^k \to \mathbb{R} \) is symmetric, quasiconvex, bounded, and Lipschitz continuous, then there exist prior \( \pi_c \) supported on \( \{h : \|h\| \leq c\} \) such that

\[
\lim_{c \to \infty} \liminf_n \inf_{\hat{\psi}_n} \int \mathbb{E}_{\theta_0 + h/\sqrt{n}} \left[ L \left( \sqrt{n}(\hat{\psi}_n) - \psi(\theta_0 + h/\sqrt{n}) \right) \right] d\pi_c(h) \\
\geq \mathbb{E}[L(Z)] \quad \text{for } Z \sim \mathcal{N} \left( 0, \dot{\psi}(\theta_0)I_{\theta_0}^{-1}\dot{\psi}(\theta_0)^T \right)
\]
Best-regular estimators and the delta method

estimator $T_n$ of $\theta_0$ satisfying
$$T_n = \theta_0 + l_{\theta_0}^{-1} P_n \hat{\ell}_{\theta_0} + o_{P_{\theta_0}}(1/\sqrt{n})$$
is best regular

Corollary (Delta-method and best-regular estimators)

If $\psi : \mathbb{R}^d \rightarrow \mathbb{R}^k$ is differentiable at $\theta_0$ and $T_n$ is best regular, then $\psi(T_n)$ achieves local asymptotic minimax bound
Nonparametric efficiency

- family $\mathcal{P}$ of distributions on $\mathcal{X}$
- parameter $\theta : \mathcal{P} \rightarrow \mathbb{R}^d$
- lower bound based on model subfamilies

**Definition**
A collection $\{P_h\}_{h \in \mathbb{R}^k} \subset \mathcal{P}$ (often take $\|h\| \leq \epsilon$) is a *quadratic mean differentiable* subfamily (QMD) at $P_0 \in \mathcal{P}$ if there exists a score $g : \mathcal{X} \rightarrow \mathbb{R}^k$, $g \in L^2(P_0)$, with

$$
\int \left( \sqrt{dP_h} - \sqrt{dP_0} - \frac{1}{2} g^T h \sqrt{dP_0} \right)^2 = o(\|h\|^2)
$$
Subfamily examples

Example (Parametric families)
If $\mathcal{P} = \{P_\theta\}_{\theta \in \Theta}$ is a (usual) QMD family, we have score $\ell_\theta$

Example (Nonparametric families)
Let $\phi : \mathbb{R} \to \mathbb{R}_+$ be bounded, differentiable at 0, with $\phi(0) = \phi'(0) = 1$. For $g \in L^2(P_0)$, $g : \mathcal{X} \to \mathbb{R}^k$ with $P_0g = 0$, model family $\{P_h\}$ with

$$dP_h(x) = c(h)\phi(h^Tg(x))dP_0(x)$$

is QMD at $P_0$ with score $g$
The “basic” idea

- Look at parameter differentiable relative to submodel with score $g$,
  $$\theta(P_h) = \theta(P_0) + \dot{\theta}_{P_0}(g)h + o(\|h\|)$$
  where $\dot{\theta}_{P_0}(g) \in \mathbb{R}^{d \times k}$

- Corollary on page 15–8 suggests asymptotic lower bound
  $$\mathbb{E}\left[L \left(\sqrt{n}(\hat{\theta}_n - \theta(P))\right)\right] \geq \mathbb{E}[L(Z)],$$
  $$Z \sim \mathcal{N}\left(0, \dot{\theta}_{P_0}(g)(P_0gg^T)^{-1}\dot{\theta}_{P_0}(g)^T\right)$$

- Choose “worst” sub-model $g$
Tangent sets

Definition
The tangent set $\mathcal{P}_P$ to $\mathcal{P}$ at $P$ is the collection of score functions $g : \mathcal{X} \to \mathbb{R}^d$ as we vary QMD subfamilies.

- Often just one-dimensional subfamilies (higher-dimensional easier for us).
- Always a subset of $L^2(P_0) = \{ g : \mathcal{X} \to \mathbb{R}^d \mid \int \|g\|_2^2 dP_0 < \infty \}$.
- Often linear, so becomes a tangent space.
influence functions and derivatives

interested in “appropriately smooth” functions of distribution

Definition
A parameter \( \theta : \mathcal{P} \rightarrow \mathbb{R}^d \) is differentiable at \( P_0 \) relative to tangent set \( \dot{\mathcal{P}}_{P_0} \) if for each QMD submodel \( \{ P_h \} \) with score \( g \in \dot{\mathcal{P}}_{P_0} \), \( g : \mathcal{X} \rightarrow \mathbb{R}^k \), if \( t_n \rightarrow 0 \) and \( h_n \rightarrow h \in \mathbb{R}^d \) imply

\[
\frac{\theta(P_{tn}h_n) - \theta(P_0)}{t_n} \rightarrow D_{P_0}(g^T h)
\]

for a continuous linear mapping \( D_{P_0} : L^2(P_0) \rightarrow \mathbb{R}^d \)
Influence functions

**Observation**

*There exists a mean-zero influence function* \( \dot{\theta}_0 : \mathcal{X} \to \mathbb{R}^d \), \( \dot{\theta}_0 \in L^2(P_0) \), *such that*

\[
D_{P_0}(f) = \int \dot{\theta}_0(x)f(x)dP_0(x)
\]

*for each* \( f : \mathcal{X} \to \mathbb{R} \) *with* \( P_0f^2 < \infty \)
Influence function examples

Example (Parametric families)
For parametric family \( \{ P_{\theta} \}_{\theta \in \Theta} \), influence function is \( I_{\theta_0}^{-1} \ell_{\theta_0} \).

Example (Nonparametric mean estimation)
For \( \theta(P) = PX \), influence function is \( \dot{\theta}_0(x) = x - P_0X \).
Influence function examples (continued)

Example (M-estimation)

Let $\ell : \mathbb{R}^d \times \mathcal{X} \to \mathbb{R}$ be convex, sufficiently smooth and integrable. Let $L(\theta) = P\ell(\theta, X)$. For $\theta(P) := \text{argmin}_\theta P\ell(\theta, X)$, influence is

$$\dot{\theta}_0(x) = - (\nabla^2 L(\theta_0))^{-1} \nabla \ell(\theta_0, x)$$
Theorem

Let $L : \mathbb{R}^d \rightarrow \mathbb{R}$ be bounded, quasiconvex, symmetric, and Lipschitz and $\mathcal{P}_{P_0} \subset L^2(P_0) \subset \mathcal{X} \rightarrow \mathbb{R}^k$ be a tangent space. Assume \( \theta \) is differentiable relative to $\mathcal{P}_{P_0}$. Then there exist priors $\pi_c$ supported on \( \{ h \in \mathbb{R}^k : \| h \| \leq c \} \) such that

$$
\lim_{c \to \infty} \lim_{n \to \infty} \inf_{\hat{\theta}_n} \int \mathbb{E}_{h/\sqrt{n}} \left[ L \left( \sqrt{n}(\hat{\theta}_n - \theta(P_h/\sqrt{n})) \right) \right] d\pi_c(h) \geq \mathbb{E}[L(Z)] \quad \text{for } Z \sim \mathcal{N} \left( 0, \text{Cov}_0(\dot{\theta}_0, g^T)(P_0gg^T)^{-1}\text{Cov}_0(g, \dot{\theta}_0^T) \right)
$$
Comments on nonparametric local minimax bound

- often consider only 1-dimensional submodels, look at

\[
\lim_{t \downarrow 0} \frac{\theta(P_{tg}) - \theta(P_0)}{t} = D_{P_0}(g)
\]

- equivalent if \( \psi(h) := \theta(P_{g^T h}) \) locally Lipschitz in \( h \) for \( g : \mathcal{X} \rightarrow \mathbb{R}^d \)
Types of differentiability: Gateaux, Hadamard, and Fréchet

Let $f : X \to V$ for Banach spaces $X$, $V$. Then $f$ is

- **Gateaux differentiable** at $x$ if directional derivatives exist:

$$f'(x; v) := \lim_{t \downarrow 0} \frac{f(x + tv) - f(x)}{t}$$

- **Hadamard (compactly) differentiable** if the directional derivative is linear, $f'(x; v) = D_x v$, and for all $v_t \to v$,

$$\lim_{t \downarrow 0} \frac{f(x + tv_t) - f(x)}{t} = D_x v$$

- **Fréchet differentiable** if

$$f(x + v) - f(x) - D_x v = o(\|v\|) \quad \text{as} \quad \|v\| \to 0.$$
Finite dimensional equivalence differentiability

Proposition

Let \( f : \mathbb{R}^n \to \mathbb{R}^k \), i.e., in finite dimensions and assume its Gateaux derivative at \( x \) exists and is linear. Then

- If \( f \) is locally Lipschitz, it is Hadamard differentiable at \( x \)
- Hadamard and Fréchet differentiability coincide.
The largest lower bound

Corollary

Assume conditions of Theorem on page 15–18. Then \( g = \dot{\theta}_0 \) maximizes the lower bound, yielding asymptotic lower bound

\[
\mathbb{E} \left[ L \left( \mathcal{N}(0, P_0 \dot{\theta}_0 \dot{\theta}_0^T) \right) \right].
\]
Achieving the bound

- regular estimator with efficient influence function

\[ \hat{\theta}_n - \theta_0 = P_n \dot{\theta}_0 + o_{P_0}(1/\sqrt{n}) \quad (1) \]

Corollary

Any regular estimator of the form (1) achieves the local asymptotic minimax lower bound.