Semiparametric posterior limits

Bas Kleijn, KdV Institute for Mathematics

based on joint work with P. Bickel, B. Knapik, M. Chae and Y. Kim
Part I

Regularity, efficiency and semiparametric bias
Example I Regression with symmetric errors

Question
Observe \( i.i.d. \ X_1, \ldots, X_n, \ X_i = \theta + e_i \) (or \( Y_i = \theta X_i + e_i, \ etcetera \)) with a symmetrically distributed error. Density for \( X \)'s is,

\[
p_{\theta_0, \eta_0}(x) = \eta_0(x - \theta_0),
\]

where \( \eta \in H \) is a symmetric Lebesgue density on \( \mathbb{R} \). We assume that \( \eta \) is smooth and that the Fisher information for location is non-singular.

Adaptivity Stein (1956), Bickel (1982)
For inference on \( \theta_0 \) it does not matter whether we know \( \eta_0 \) or not!

\[
\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{P_{\theta_0, \eta_0}^{-w.}} N(0, I_{\theta_0, \eta_0}^{-1})
\]

where \( I_{\theta_0, \eta_0} \) is the Fisher information.
Example II Domain boundary estimation

Question
Observe sample $X_1, \ldots, X_n$ i.i.d. $-P_{\theta_0, \eta_0}$ with continuous density

$$p_{\theta_0, \eta_0}(x) = \eta_0(x - \theta_0), \quad \eta_0(y) = 0, \text{ if } y < 0.$$  

and $\tilde{\lambda}_0 = \eta_0(0) > 0$. Estimate $\theta_0$ with $\eta_0 \in H$, an unknown nuisance.

Model
Define $\mathcal{L} = C_S[0, \infty]$ (cont. $f : [0, \infty] \to \mathbb{R}$ such that $|f| \leq S$)

$$\mathcal{L} \to H : \hat{\ell} \mapsto \eta, \quad \eta(x) = Z_{\hat{\ell}}^{-1} e^{-\alpha x} \int_0^x \hat{\ell}(t) \, dt, \quad (Z_{\hat{\ell}} \text{ normalizes})$$

$\eta \in H$ is monotone decreasing, differentiable and log-Lipschitz.
Example III Partial linear regression

Model
Consider \textit{i.i.d.} sample $X_1, \ldots, X_n$, with $X = (Y, U, V) \in \mathbb{R}^3$, related as,

$$Y = \theta_0 U + \eta_0(V) + e$$

where $e \sim N(0, 1)$ independent of $(U, V) \sim P$, $\theta_0 \in \mathbb{R}$, $\eta_0 \in H$.

Question
When can we estimate parameter of interest $\theta_0$ in the presence of the unknown nuisance parameter $\eta_0$?

Efficiency
There exist estimators $\hat{\theta}_n$ for $\theta_0$ under $P_0 = P_{\theta_0, \eta_0}$ such that,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{P_{\theta_0, \eta_0}} N(0, \tilde{I}^{-1}_{\theta_0, \eta_0})$$

where $\tilde{I}_{\theta_0, \eta_0}$ is the efficient Fisher information.
Likelihood expansions \textbf{LAN}  
– local asymptotic normality –

Definition (Le Cam (1960))

There is a $\hat{\ell}_{\theta_0} \in L_2(P_{\theta_0})$ with $P_{\theta_0} \hat{\ell}_{\theta_0} = 0$ s.t. for any $(h_n) = O_{P_{\theta_0}}(1)$,

$$
\prod_{i=1}^{n} \frac{p_{\theta_0} + n^{-1/2}h_n}{p_{\theta_0}}(X_i) = \exp \left( h_n \Delta'_{n,\theta_0} - \frac{1}{2} h_n^T I_{\theta_0} h_n + o_{P_{\theta_0}}(1) \right),
$$

where $\Delta'_{n,\theta_0}$ is given by,

$$
\Delta'_{n,\theta_0} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{\ell}_{\theta_0}(X_i) \xrightarrow{P_{\theta_0}^{-w.}} N(0, I_{\theta_0}),
$$

and $I_{\theta_0} = P_{\theta_0} \hat{\ell}_{\theta_0} \hat{\ell}_{\theta_0}^T$ is the Fisher information.
Likelihood expansions LAE  
– local asymptotic exponentiality –

Definition

There exists a $\tilde{\lambda}_{\theta_0} > 0$ such that for any bounded, stochastic $(h_n)$,

$$
\prod_{i=1}^{n} \frac{p_{\theta_0 + n^{-1} h_n}(X_i)}{p_{\theta_0}} = \exp\left( h_n \tilde{\lambda}_{\theta_0} + o_{P_{\theta_0}}(1) \right) 1\{h_n \leq \Delta'_{n,\theta_0}\},
$$

where $\Delta'_{n,\theta_0}$ satisfies,

$$
\lim_{n \to \infty} P^n_{\theta_0}(\Delta'_{n,\theta_0} > u) = e^{-\tilde{\lambda}_{\theta_0} u}, \quad \text{(that is $\Delta'_{n,\theta_0} \xrightarrow{P_{\theta_0}-w.} \text{Exp}_0^{+},\tilde{\lambda}_{\theta_0}$)}
$$

for all $u > 0$. (Ibrigimov and Has’minskii (1981)).
Regular estimation and efficiency I
– definition and convolution theorem –

Definition
An estimator sequence $\hat{\theta}_n$ for a parameter $\theta_0$ is said to be regular, if for every $h_n = O(1)$, with $\theta_n = \theta_0 + n^{-1/2}h_n$

\[
\sqrt{n}(\hat{\theta}_n - \theta_n) \xrightarrow{P_{\theta_n}-w.} L_{\theta_0}
\]

for some $(h_n)$-independent limit distribution $L_{\theta_0}$.

Theorem 8.1 (Hájek, 1970)
Assume that the model is LAN at $\theta_0$ with non-singular Fisher information $I_{\theta_0}$. Suppose $\hat{\theta}_n$ is a regular estimator for $\theta_0$ with limit $L_{\theta_0}$. Then there exists a probability kernel $M_{\theta_0}$ such that $L_\theta = N(0, I_{\theta_0}^{-1}) \ast M_{\theta_0}$. 


Regular estimation and efficiency II
– asymptotic linearity and asymptotic bias –

Definition

Given an asymptotic estimation problem with \( i.i.d. - P_0 \) data and non-singular Fisher information \( I_0 \), an influence function \( \Delta_n \) is,

\[
\Delta_n = I_0^{-1} \Delta_n' = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_0^{-1} \dot{\ell}_{\theta_0}(X_i) \xrightarrow{P_0-\text{w.}} N(0, I_0^{-1})
\]

**Theorem 9.1** (Fisher, Cramér, Rao, Le Cam, Hájek)

An estimator \( \hat{\theta}_n \) is **efficient** if and only if it is asymptotically linear:

\[
\sqrt{n}(\hat{\theta}_n - \theta_0) = \Delta_{n, \theta_0} + o_{P_0}(1),
\]

for some influence function \( \Delta_{n, \theta_0} \xrightarrow{P_\theta_0-\text{w.}} N(0, I_{\theta_0}^{-1}) \).

Note the asymptotic bias, it equals zero because \( P_{\theta_0} \dot{\ell}_{\theta_0} = 0 \).
Semiparametric bias

An estimator $\hat{\theta}_n$ for $\theta_0$ is regular but asymptotically biased if,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \tilde{\Delta}_{n,\theta_0,\eta_0} + \mu_{n,\theta_0,\eta_0} + o_P(1),$$

with $\tilde{\Delta}_{n,\theta_0,\eta_0} \overset{P_0-w.}{\longrightarrow} N(0, \tilde{I}_{\theta_0,\eta_0}^{-1})$ and $\mu_{n,\theta_0,\eta_0} = O(1)$ or worse. Typically,

$$\left| \mu_{n,\theta_0,\eta_0} \right| \leq \sqrt{n} \sup_{\eta \in D_n} \left| \tilde{I}_{\theta_0,\eta_0}^{-1} P_{\theta_0,\eta_0} \tilde{\ell}_{\theta_0,\eta_0} \right|$$

where $D_n$ describes some form of localization for $\eta \in H$ around $\eta_0$.

**Theorem 10.1** (approximate, see Schick (1986), Klaassen (1987))

An efficient estimator for $\theta_0$ exists if and only if there exists an estimator $\hat{\Delta}_n$ for the influence function, whose asymptotic bias vanishes at a rate strictly faster than $\sqrt{n}$,

$$P^n_{\theta_n,\eta} \hat{\Delta}_n = o(n^{-1/2}),$$
Part II

Semiparametric posterior limits
Parametric Bernstein-von Mises theorem

**Theorem 12.1** (Le Cam (1953), $h = \sqrt{n}(\theta - \theta_0)$)

Let $\mathcal{P} = \{P_\theta : \theta \in \Theta \subseteq \mathbb{R}^d\}$ with thick prior $\Pi_{\Theta}$ be LAN at $\theta_0$ with non-singular $I_{\theta_0}$. Assume that for every sequence of radii $M_n \to \infty$,

$$\Pi\left( \|h\| \leq M_n \mid X_1, \ldots, X_n \right) \xrightarrow{P_0} 1$$

Then the posterior converges to normality as follows

$$\sup_B \left\| \Pi\left( h \in B \mid X_1, \ldots, X_n \right) - N_{\Delta_n,\theta_0,I_{\theta_0}^{-1}}(B) \right\| \xrightarrow{P_0} 0$$

Another, more familiar form of the assertion,

$$\sup_B \left\| \Pi\left( \theta \in B \mid X_1, \ldots, X_n \right) - N_{\hat{\theta}_n,(nI_{\theta_0})^{-1}}(B) \right\| \xrightarrow{P_0} 0$$

for any efficient $\hat{\theta}_n$. 

12
Consistency under $\sqrt{n}$-perturbation

Given $\rho_n \downarrow 0$ we speak of consistency under $n^{-1/2}$-perturbation at rate $\rho_n$, if for all $h_n = O_{P_0}(1)$.

$$\Pi_n\left(D(\theta, \rho_n) \mid \theta = \theta_0 + n^{-1/2}h_n; X_1, \ldots, X_n\right) \xrightarrow{P_0} 1$$
Integral local asymptotic normality
– Graph/Heuristics –

reparametrize \((\theta, \zeta) \mapsto (\theta, \eta^*(\theta) + \zeta)\)
Likelihood expansions ILAN

– Integral local asymptotic normality –

Definition
Given a nuisance prior \( \Pi_H \), the localized integrated likelihood is,

\[
s_n(h) = \int_H \prod_{i=1}^n \frac{p_{\theta_0 + n^{-1/2}h, \eta}(X_i)}{p_{\theta_0, \eta_0}} (X_i) d\Pi_H(\eta),
\]

Definition
\( s_n \) is said to have the ILAN property, if for every \( h_n = O_{P_0}(1) \)

\[
\log \frac{s_n(h_n)}{s_n(0)} = h_n^T \tilde{\Delta}'_{n, \theta_0, \eta_0} - \frac{1}{2} h_n^T \tilde{I}_{\theta_0, \eta_0} h_n + o_{P_0}(1),
\]

where the efficient \( \tilde{\Delta}'_{n, \theta_0, \eta_0} \) is given by

\[
\tilde{\Delta}'_{n, \theta_0, \eta_0} = \frac{1}{\sqrt{n}} \sum_{i=1}^\infty \tilde{\ell}_{\theta_0, \eta_0} \xrightarrow{P_{\theta_0, \eta_0}^{-w.}} N(0, \tilde{I}_{\theta_0, \eta_0})
\]
Semiparametric Bernstein-von Mises theorem

**Theorem 16.1** (Bickel and BK (2012))

Let \( \mathcal{P} = \{P_{\theta,\eta} : \theta \in \Theta, \eta \in H\} \) with thick prior \( \Pi_{\Theta} \) and nuisance prior \( \Pi_{H} \). Assume **ILAN at** \( P_{\theta_0,\eta_0} \) with non-singular \( \tilde{I}_{\theta_0,\eta_0} \). Assume that for every sequence of radii \( M_n \to \infty \),

\[
\Pi \left( \|h\| \leq M_n \mid X_1, \ldots, X_n \right) \xrightarrow{P_0} 1
\]

Then the posterior converges marginally to normality as follows

\[
\sup_B \left| \Pi \left( h \in B \mid X_1, \ldots, X_n \right) - N_{\Delta_{n,\theta_0,\eta_0},\tilde{I}_{\theta_0,\eta_0}^{-1}} (B) \right| \xrightarrow{P_0} 0
\]

**Both ILAN and \( \sqrt{n} \)-consistency** are sensitive to semiparametric bias!
Example I Regression with symmetric errors

**Theorem 17.1** (Minwoo Chae, Yongdai Kim and BK (201?))

Let $X_1, \ldots, X_n$ be i.i.d.-$P_{\theta_0,\eta_0}$, i.e. $X_i = \theta_0 + e_i$ with $e$ distributed as a symmetric normal location mixture $\eta_0$ from $H$ of the form,

$$
\eta(x) = \int \phi(x - z) \, dF(z)
$$

(where $F$ is symmetric and $\phi$ denotes the standard normal density). With **thick prior** $\Pi_{\Theta}$ and **nuisance prior** $\Pi_{H}$ that has full weak support, the posterior converges marginally to normality

$$
\sup_B \left| \Pi( h \in B \mid X_1, \ldots, X_n ) - N_{\tilde{\Delta}_{n,\theta_0,\eta_0}, \tilde{I}_{\theta_0,\eta_0}^{-1}}(B) \right| \xrightarrow{P_0} 0
$$

where $\tilde{\ell}_{\theta_0,\eta_0}(X) = \hat{p}_{\theta_0,\eta_0}/p_{\theta_0,\eta_0}(X)$ and $\tilde{I}_{\theta_0,\eta_0} = P_0 \tilde{\ell}_{\theta_0,\eta_0}^2$. 
Example II Domain boundary estimation

Nuisance prior
Let $S > 0$, $W = \{W_s : s \in [0, 1]\}$ BM on $[0, 1]$, $Z \sim N(0, 1)$, indept. of $W$. Let $\Psi : [0, \infty) \mapsto [0, 1]$, $t \mapsto (2/\pi) \arctan(t)$. Define $\hat{\ell} \sim \Pi$ by,

$$\hat{\ell}(t) = S \Psi(|Z + W\Psi(t)|).$$

Then $C_S[0, \infty] \subset \text{supp}(\Pi)$.

**Theorem 18.1** (BK and B. Knapik (201?))

Let $X_1, \ldots, X_n$ be i.i.d.-$P_{\theta_0, \eta_0}$. Endow $\Theta = \mathbb{R}$ with prior *thick at* $\theta_0$ and $H$ with prior $\Pi$ like above. Then,

$$\sup_B \left| \Pi(h \in B \mid X_1, \ldots, X_n) - \text{Exp}^{-\Delta_{n, \theta_0, \tilde{\lambda}_0}(B)} \right| \xrightarrow{P_0} 0$$

where $\Delta_{n, \theta_0} = n(X_1) - \theta_0$ and $\tilde{\lambda}_0 = \eta_0(0)$. 

18
Example III Partial linear regression

Model specification
Observe i.i.d.-$P_0$ sample $(U_i, V_i, Y_i) \ (i \geq 1)$, $Y = \theta_0 U + \eta_0(V) + e$. where $e \sim N(0, 1)$ independent of $(U, V) \sim P$, $PU = 0$, $PU^2 = 1$, $PU^4 < \infty$, $P(U - E[U|V])^2 > 0$, $P(U - E[U|V])^4 < \infty$. For given $\alpha > 0$, $M > 0$, define Sobolev ball $H_{\alpha, M} = \{\eta \in C^{\alpha}[0, 1] : \|\eta\|_{\alpha} < M\}$.

Conjecture 19.1 (Bickel and BK (2012))

Let $\alpha > 1/2$ and $M > 0$ be given. Assume that $\eta_0$ as well as $v \mapsto E[U|V = v]$ are in $H_{\alpha, M}$. Let $\Pi_\Theta$ be thick. Choose $k > \alpha - 1/2$ and define $\Pi_{\alpha, M}^k$ to be the distribution of $k$ times integrated Brownian motion started at random, conditioned on $\|\eta\|_{\alpha} < M$. Then,

$$\sup_A \left| \Pi \left( h \in A \bigg| X_1, \ldots, X_n \right) - N_{\tilde{\Delta}_{n, \theta_0, \eta_0}, \tilde{\Gamma}_{\theta_0, \eta_0}^{-1}}(A) \right| \xrightarrow{P_0} 0,$$

where $\tilde{\ell}_{\theta_0, \eta_0}(X) = e(U - E[U|V])$ and $\tilde{\Gamma}_{\theta_0, \eta_0} = P(U - E[U|V])^2$. 

19
The rug from under our feet!
– Why systematic answers are so hard –

Proving ILAN with a product prior $\Pi_{\Theta} \times \Pi_H$, one has to go through a condition of the form

$$\int_{D_n} \prod_{i=1}^n \frac{p_{\theta_0,\eta}(X_i)}{p_0} d\Pi_{H+\delta_n}(\eta) = \left(1 + o_{P_0}(1)\right) \int_{D_n} \prod_{i=1}^n \frac{p_{\theta_0,\eta}(X_i)}{p_0} d\Pi_H(\eta)$$

where $\Pi_{H+\delta}(B) := \Pi_H(B + \delta)$ and $\delta_n = \eta^*(\theta_n) - \eta_0$.

For Gaussian prior $\Pi_H$ approximate $\delta_n$ in RKHS and use the Cameron-Martin theorem to obtain the RN-derivative $d\Pi_{H+\delta_n}/d\Pi_H$ explicitly.

For other priors, linearize and estimate $\delta_n$ with $\hat{\delta}_n(\theta) = \theta \hat{B}_n$ and choose a nuisance prior $\Pi_H$ on $H$ and then translate $\Pi_H$ empirically

$$\Pi_H(\eta \in B | \theta) = \Pi_H(\eta - \hat{\delta}_n \in B)$$
Posterior asymptotic normality
– Analogy/Heuristics –

Parametric posterior

The posterior density \( \theta \mapsto \frac{\prod_{i=1}^{n} p_{\theta}(X_{i}) \, d\Pi(\theta)}{\int_{\Theta} \prod_{i=1}^{n} p_{\theta}(X_{i}) \, d\Pi(\theta)} \)

with LAN requirement on the likelihood.

Semiparametric analog

The marginal posterior density \( \theta \mapsto \frac{\int_{H} \prod_{i=1}^{n} p_{\theta,\eta}(X_{i}) \, d\Pi_{H}(\eta) \, d\Pi_{\Theta}(\theta)}{\int_{\Theta} \int_{H} \prod_{i=1}^{n} p_{\theta,\eta}(X_{i}) \, d\Pi_{H}(\eta) \, d\Pi_{\Theta}(\theta)} \)

with integral LAN requirement on \( \Pi_{H} \)-integrated likelihood.
Marginal convergence at rate $\sqrt{n}$

**Theorem 22.1** (*Marginal parametric rate*)

Let $\Pi_\Theta$ and $\Pi_H$ be given. Assume that there exists a sequence $(H_n)$ of subsets of $H$, such that the following two conditions hold:

(i) The nuisance posterior concentrates on $H_n$

$$\Pi\left(\eta \in H \setminus H_n \mid X_1, \ldots, X_n\right) \xrightarrow{P_0} 0$$

(ii) For every $M_n \to \infty$,

$$\sup_{\eta \in H_n} P^n_0 \Pi\left( n^{1/2} \|\theta - \theta_0\| > M_n \mid \eta, X_1, \ldots, X_n\right) \to 0$$

Then for every $M_n \to \infty$

$$\Pi\left( n^{1/2} \|\theta - \theta_0\| > M_n \mid \eta, X_1, \ldots, X_n\right) \xrightarrow{P_0} 0$$