Isaac Newton Institute, Cambridge, March 27th, 2008

The semiparametric Bernstein-Von Mises theorem

Bas Kleijn, Korteweg-de Vries institute,
University of Amsterdam

The parametric Bernstein-Von Mises theorem

– setting the stage –

Smooth parametric model

Θ open subset of $\mathbb{R}^d$; $\mathscr{P} = \{P_\theta : \theta \in \Theta\}$, dominated by some σ-finite μ: $p_\theta = dP_\theta/d\mu$, differentiable in suitable sense

Frequentist asymptotics

The sample $X_1, X_2, \ldots$ is distributed *i.i.d.*-P_0; $P_0 = P_{\theta_0}$, for some $\theta_0 \in \Theta$

Continuous, positive prior

Prior Π with Lebesgue-density π, continuous on neighbourhood of θ_0 and positive at θ_0, $\pi(\theta_0) > 0$

Localization

We 'localize' the parameter: centre and rescale by $\sqrt{n}$, $h = \sqrt{n}(\theta - \theta_0) \in \mathbb{R}^d$

The parametric Bernstein-Von Mises theorem

Recall that for any best-regular estimator $\widehat{\theta}_n$ for θ_0,

$$\sqrt{n}\|\widehat{\theta}_n - \theta_0\| = \Delta_{n,\theta_0} + o_{P_0}(1), \quad \text{where} \quad \Delta_{n,\theta_0}(X) = I_{\theta_0}^{-1}\mathbb{G}_n\dot{\ell}_{\theta_0}$$

Theorem 1. (Bernstein-Von Mises)
Let $\mathscr{P}$ be locally asymptotically normal at θ_0 with non-singular Fisher-information I_{θ_0}. Assume that for every sequence of radii $M_n \to \infty$,

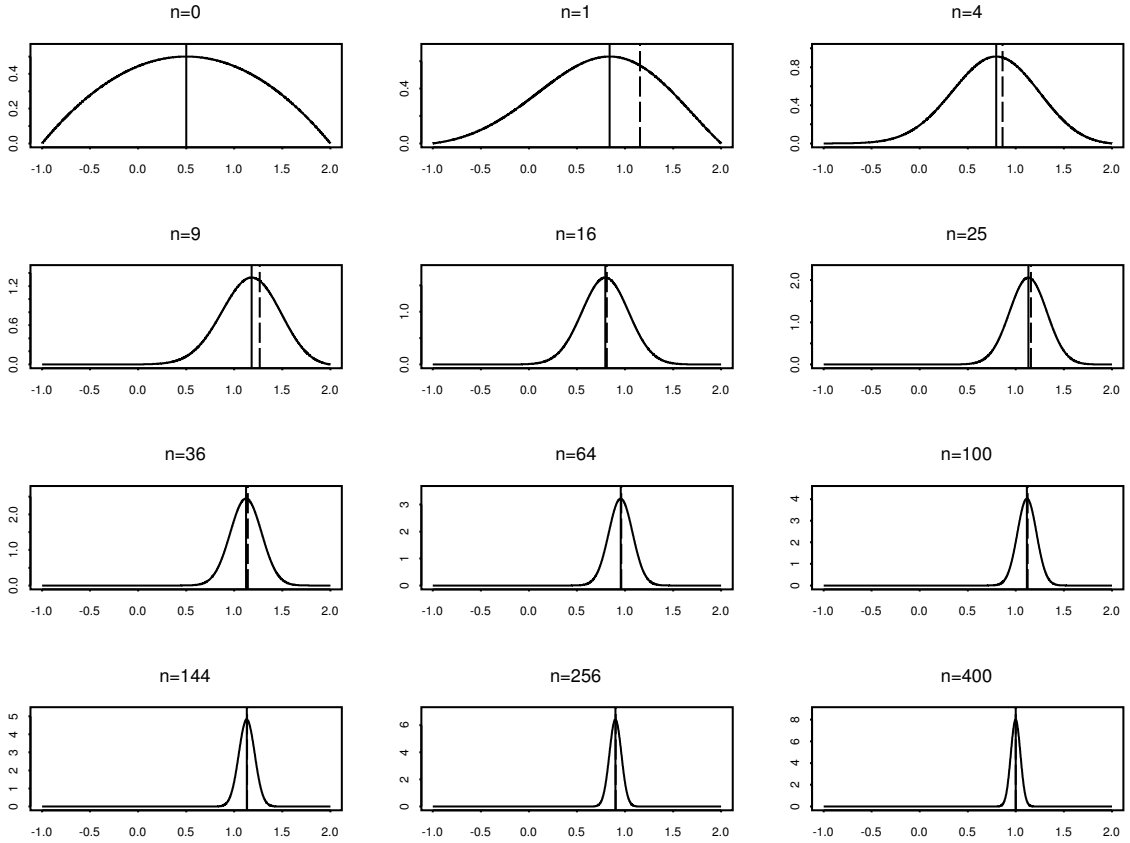
$$\Pi\Big(\|h\| \leq M_n \;\Big|\; X_1, \ldots, X_n \Big) \xrightarrow{P_0} 1 \tag{1}$$

Then the posterior converges to normality as follows

$$\sup_B \Big| \Pi\Big(h \in B \;\Big|\; X_1, \ldots, X_n \Big) - N_{\Delta_{n,\theta_0}, I_{\theta_0}^{-1}}(B) \Big| \xrightarrow{P_0} 0 \tag{2}$$

BvM-type posterior behaviour in graphs

Posterior density on a normal location model



Simulated for sample sizes $n = 0, 1, 4, 9, \ldots, 256, 400$

Posterior asymptotic normality and the MLE

If the MLE $\widehat{\theta}_n$ is best-regular

$$\sup_B \left| N_{\sqrt{n}(\widehat{\theta}_n - \theta_0), I_{\theta_0}^{-1}}(B) - N_{\Delta_{n,\theta_0}, I_{\theta_0}^{-1}}(B) \right| \xrightarrow{P_0} 0$$

Then the BvM-assertion can be rewritten

$$\sup_B \left| \Pi\Big(\theta \in B \;\Big|\; X_1, \ldots, X_n \Big) - N_{\widehat{\theta}_n, n^{-1} I_{\theta_0}^{-1}}(B) \right| \xrightarrow{P_0} 0$$

From this, we see that if the BvM-assertion holds,

Bayesian point-estimators and MLE coincide up to $o_{P_0}(1)$

Credible sets are (MLE-based) confidence regions asymptotically

Of practical importance: it is relatively easy to compute (or rather, simulate) posterior distributions. (*e.g.* Markov-chain Monte-Carlo)

LAN and sLAN differentiability

A model $\mathscr{P}$ is locally asymptotically normal (LAN) at θ_0, if for every converging sequence $h_n \to h$,

$$\log \prod_{i=1}^{n} \frac{p_{\theta_0+h_n/\sqrt{n}}}{p_{\theta_0}}(X_i) = h^T \mathbb{G}_n \dot{\ell}_{\theta_0} - \tfrac{1}{2} h^T I_{\theta_0} h + o_{P_0}(1) \qquad (3)$$

Here $\dot{\ell}_{\theta_0} \in L_2(P_{\theta_0})$, $P_0 \dot{\ell}_{\theta_0} = 0$ and $I_{\theta_0} = P_0 \dot{\ell}_{\theta_0} \dot{\ell}_{\theta_0}^T$ is the Fisher information.

LAN is equivalent to differentiability in quadratic mean.

A model $\mathscr{P}$ is stochastically locally asymptotically normal (sLAN) at θ_0, if for every random sequence $h_n = O_{P_0}(1)$,

$$\log \prod_{i=1}^{n} \frac{p_{\theta_0+h_n/\sqrt{n}}}{p_{\theta_0}}(X_i) = h_n^T \mathbb{G}_n \dot{\ell}_{\theta_0} - \tfrac{1}{2} h_n^T I_{\theta_0} h_n + o_{P_0}(1) \qquad (4)$$

Under continuity requirements, sLAN is equivalent to uniform LAN.

Semiparametric Bernstein-Von Mises theorem

– setting the stage –

Semiparametric model

Θ open subset of $\mathbb{R}^d$, H normed space; $\mathscr{P} = \{P_{\theta,\eta} : \theta \in \Theta, \eta \in H\}$, dominated by some σ-finite μ: $p_{\theta,\eta} = dP_{\theta,\eta}/d\mu$

Frequentist asymptotics

Sample $X_1, X_2, \ldots$ is *i.i.d.*-P_0; $P_0 = P_{\theta_0,\eta_0}$, for some $\theta_0 \in \Theta$, $\eta_0 \in H$

Product prior

Prior Π_Θ with Lebesgue-density π, continuous on neighbourhood of θ_0 and positive at θ_0, $\pi(\theta_0) > 0$; prior Π_H on H, no pointmasses

Localization and reparametrization

Again we 'localize' the parameter of interest $h = \sqrt{n}(\theta - \theta_0) \in \mathbb{R}^d$; we translate the nuisance along a least-favourable submodel $\zeta = \eta - \eta^*(\theta)$

Central question

Define $\tilde{\Delta}_n(X) = \tilde{I}^{-1}_{\theta_0,\eta_0} \mathbb{G}_n \tilde{\ell}_{\theta_0,\eta_0}$ where $\tilde{\ell}_{\theta_0,\eta_0}$ is the efficient score

BvM-assertion for the marginal posterior of the parameter of interest

$$\sup_B \left| \Pi\Big(h \in B \;\Big|\; X_1, \ldots, X_n \Big) - N_{\tilde{\Delta}_n, \tilde{I}^{-1}_{\theta_0,\eta_0}}(B) \right| \xrightarrow{P_0} 0. \tag{5}$$

So the central question is

"Under which conditions does the BvM-assertion hold?"

or rather, "Can we prove (5) under 'reasonable' conditions?"

‘Reasonable’ conditions

Reasonable conditions

Conditions on Π_Θ (*e.g.* like in parametric case)

Continuity of a θ-dependence (*e.g.* likelihood, Π_H-integrals)

Bounds, uniform in θ (*e.g.* Glivenko-Cantelli property)

Differentiability or (s)LAN of the model (*e.g.* along least-favourable direction)

Too strong conditions

(Strong) conditions on Π_H (*e.g.* rate-related constraints or worse)

Continuity of a η-dependence (*e.g.* almost-sure continuity of likelihood)

Bounds, uniform in η (*e.g.* GC- or Donsker-type log-likelihood classes)

Higher-order Taylor (or LAN) expansions (*e.g.* likelihood, scores *etcetera*)

Also, simplicity of conditions leads to transparency of formulation

A useful analogy: the integrated likelihood

The parametric posterior density $d\Pi(\theta|X_1,\ldots,X_n)$

$$\prod_{i=1}^{n} p_\theta(X_i)\, d\Pi(\theta) \Big/ \int_\Theta \prod_{i=1}^{n} p_\theta(X_i)\, d\Pi(\theta)$$

The marginal semiparametric posterior density $d\Pi(\theta|X_1,\ldots,X_n)$

$$\int_H \prod_{i=1}^{n} p_{\theta,\eta}(X_i)\, d\Pi_H(\eta)\, d\Pi_\Theta(\theta) \Big/ \int_\Theta \int_H \prod_{i=1}^{n} p_{\theta,\eta}(X_i)\, d\Pi_H(\eta)\, d\Pi_\Theta(\theta)$$

Define the integrated likelihood

$$h \mapsto s_n(h) = \int_H \prod_{i=1}^{n} \frac{p_{\theta_0+n^{-1/2}h,\eta}}{p_0}(X_i)\, d\Pi_H(\eta)$$

Analogous conditions

The parametric BvM-proof can stay (almost) intact:
replace likelihoods by integrated likelihoods!

For parametric BvM, the likelihood is required to be LAN: $h_n \to h$

$$\log \prod_{i=1}^{n} \frac{p_{\theta_0+h_n/\sqrt{n}}}{p_{\theta_0}}(X_i) = h^T \mathbb{G}_n \dot{\ell}_{\theta_0} - \tfrac{1}{2} h^T I_{\theta_0} h + o_{P_0}(1)$$

Now, we require integrated sLAN (ILAN): $h_n = O_{P_0}(1)$

$$\log s_n(h_n) = \log s_n(0) + h_n^T \mathbb{G}_n \tilde{\ell}_{\theta_0,\eta_0} - \tfrac{1}{2} h_n^T \tilde{I}_{\theta_0,\eta_0} h_n + o_{P_0}(1)$$

In addition, $\sqrt{n}$-rate for the marginal posterior is required

Semiparametric posterior asymptotic normality

Theorem 2. (Marginal posterior asymptotic normality)
Suppose that for every $n \geq 1$, $h \mapsto s_n(h)$ *is continuous at* $h = 0$ *and that* $h \mapsto s_n(h)$ *satisfies the ILAN property. Furthermore, assume that for every sequence of radii* $M_n \to \infty$,

$$\Pi\Big(\|h\| \leq M_n \;\Big|\; X_1, \ldots, X_n \Big) \xrightarrow{P_0} 1 \tag{6}$$

Then the marginal posterior for θ *converges to asymptotic normality as follows*

$$\sup_B \Big| \Pi\Big(h \in B \;\Big|\; X_1, \ldots, X_n \Big) - N_{\tilde{\Delta}_n, \tilde{I}^{-1}_{\theta_0,\eta_0}}(B) \Big| \xrightarrow{P_0} 0 \tag{7}$$

Integrated local asymptotic normality

– setting the stage –

Least-favourable submodel

Existence $\theta \mapsto \eta^*(\theta)$ minimizing $\eta \mapsto -P_0 \log(p_{\theta,\eta}/p_0)$ in an open U_0 around θ_0.

Reparametrization

Reparametrize nuisance $\zeta(\theta, \eta) = \eta - \eta^*(\theta)$ so that $D(\theta, \rho) = \{\eta : \|\eta - \eta^*(\theta)\| < \rho\} = \{\zeta : \|\zeta\| < \rho\} = B(\rho)$

Consistency under $\sqrt{n}$-perturbation

For all $\rho > 0$, $M > 0$

$$\sup_{\|h\| \le M} \Pi\big(D^c(\theta, \rho) \;\big|\; \theta = \theta_0 + n^{-1/2} h \,;\, X_1, \ldots, X_n \big) \overset{P_0-a.s.}{\longrightarrow} 0$$

Stochastic local asympotic normality

Denote by g_ζ the score along θ-direction at (θ_0, ζ) and $g_\eta = g_{\zeta(\theta_0,\eta)}$

Integrated local asymptotic normality

Theorem 3. (Integrated local asymptotic normality)
Suppose that the model is sLAN and that consistency under $\sqrt{n}$-perturbation obtains at rate $(\rho_n)_{n\geq 1}$. Furthermore, assume that Π_H has no point-masses and that

$$\sup_{\{\eta:\|\eta-\eta_0\|\leq\rho_n\}} \left\| P_{\theta_0,\eta} g_{\eta_0} \right\| = o(n^{-1/2}),$$
$$\sup_{\{\eta:\|\eta-\eta_0\|\leq\rho_n\}} \left\| P_{\theta_0,\eta} (g_\eta - g_{\eta_0})(g_\eta - g_{\eta_0})^T \right\| = o(1)$$

Then, for every $h_n = O_{P_0}(1)$,

$$\log s_n(h_n) = \log s_n(0) + h_n^T \mathbb{G}_n \tilde{\ell}_{\theta_0,\eta_0} - \tfrac{1}{2} h_n^T \tilde{I}_{\theta_0,\eta_0} h_n + o_{P_0}(1) \qquad (8)$$

Consistency under $\sqrt{n}$-perturbation

Theorem 4. (Consistency under $\sqrt{n}$-perturbation)
Assume that the following three conditions hold

(i) *For every* $\eta \in H$, $M > 0$

$$\sup_{\|h\| \le M} \left| \left(\mathbb{P}_n - P_0 \right) \log \frac{p_{\theta_0 + n^{-1/2} h, \eta}}{p_0} \right| \xrightarrow{P_0 - a.s.} 0 \tag{9}$$

(ii) *For every* $\epsilon > 0$ *there exists an open* U *around* θ_0 *such that,*

$$\inf_{\theta \in U} \Pi_H \left(\eta \in H \, : \, -P_0 \log \frac{p_{\theta,\eta}}{p^*_\theta} \le \epsilon^2 \right) > 0 \tag{10}$$

Consistency under $\sqrt{n}$-perturbation (continued)

(iii) For every open V around (θ_0, η_0), there exists a test sequence $(\psi_n)_{n\geq 1}$ such that

$$P_0^n \psi_n \to 0, \qquad \sup\Big\{ P_{\theta,\eta}^n (1-\psi_n) \,:\, (\theta,\eta) \in V^c \Big\} \to 0 \tag{11}$$

Then, for every $\rho > 0$, $M > 0$,

$$\sup_{\|h\|\leq M} \Pi\Big(D^c(\theta,\rho) \;\Big|\; \theta = \theta_0 + n^{-1/2} h \,;\, X_1, \ldots, X_n \Big) \overset{P_0 - a.s.}{\longrightarrow} 0 \tag{12}$$

Remarks on conditions for consistency

on(i) If for some U around θ_0,

$$\mathscr{G}(\eta) = \left\{ X \mapsto -P_0 \log \frac{p_{\theta,\eta}}{p_0}(X) \,:\, \theta \in U \right\}$$

is Glivenko-Cantelli for every η, then *(i)* holds.

on(ii) Note that

$$\inf_{\theta \in U} \Pi_H \left(\eta \in H \,:\, -P_0 \log \frac{p_{\theta,\eta}}{p^*_\theta} \leq \epsilon^2 \right)$$
$$\geq \Pi_H \left(\bigcap_{\theta \in U} \left\{ \eta \in H \,:\, -P_0 \log \frac{p_{\theta,\eta}}{p^*_\theta} \leq \epsilon^2 \right\} \right)$$

on(iii) If $N(\epsilon, \mathscr{P}, H) < \infty$, then the required test sequence exists

Rate of marginal convergence

Theorem 5. *Assume that* $h \mapsto s_n(h)$ *is ILAN and that* Π_Θ *is as before. Suppose that* $\{P_\theta^* : \theta \in U_0\}$ *is LAN at* θ_0 *with invertible Fisher information. Also assume that for every* $\epsilon > 0$*, there exists a test-sequence* $(\psi_n)_{n\geq 1}$ *such that:*

$$P_0^n \psi_n \to 0, \quad \sup\Big\{ P_{\theta,\eta}^n (1-\psi_n) \,:\, \|\theta - \theta_0\| > \epsilon, \eta \in H \Big\} \to 0 \tag{13}$$

Furthermore, assume that there exists a $\delta > 0$ *such that for all sequences* $(L_n)_{n\geq 1}$*,* $L_n \to \infty$*,*

$$P_0^n \left(\sup\left\{ \frac{S_n(\theta)}{S_n(\theta_0)} : \|\theta - \theta_0\| < \delta \right\} > e^{L_n^2} \right) \to 0 \tag{14}$$

Then, for every $(M_n)_{n\geq 1}$*,* $M_n \to \infty$*,*

$$P_0^n \Pi\Big(n^{1/2} \|\theta - \theta_0\| \leq M_n \,\Big|\, X_1, \ldots, X_n \Big) \xrightarrow{P_0} 1 \tag{15}$$

The semiparametric BvM theorem

Theorem 6. *Assume that Π_Θ is as before. Suppose that the model is sLAN and that various continuity requirements are met. Furthermore, assume that Π_H has no point-masses and that the following five conditions hold*

(i) Assume that

$$\sup_{\{\eta : \|\eta - \eta_0\| \leq \rho_n\}} \left\| P_{\theta_0,\eta} g_{\eta_0} \right\| = o(n^{-1/2}),$$
$$\sup_{\{\eta : \|\eta - \eta_0\| \leq \rho_n\}} \left\| P_{\theta_0,\eta} (g_\eta - g_{\eta_0})(g_\eta - g_{\eta_0})^T \right\| = o(1)$$

(ii) For every $M > 0$, $\eta \in H$,

$$\sup_{\|h\| \leq M} \left| \left(\mathbb{P}_n - P_0 \right) \log \frac{p_{\theta_0 + n^{-1/2}h,\eta}}{p_0} \right| \xrightarrow{P_0 - a.s.} 0$$

The semiparametric BvM theorem (cont.)

(iii) *For every* $\epsilon > 0$, *there exists an open* U *around* θ_0 *such that,*

$$\inf_{\theta \in U} \Pi_H \left(\eta \in H \, : \, -P_0 \log \frac{p_{\theta,\eta}}{p^*_\theta} \leq \epsilon^2 \right) > 0$$

(iv) *For every open* V *around* (θ_0, η_0), *there exists a sequence of test functions* $(\psi_n)_{n \geq 1}$ *such that*

$$P_0^n \psi_n \to 0, \qquad \sup \Big\{ P^n_{\theta,\eta}(1 - \psi_n) \, : \, (\theta, \eta) \in V^c \Big\} \to 0$$

The semiparametric BvM theorem (cont.)

(v) Assume that there exists a $\delta > 0$ *such that for all sequences* $(L_n)_{n\geq 1}$, $L_n \to \infty$,

$$P_0^n\Big(\sup\Big\{\frac{S_n(\theta)}{S_n(\theta_0)} : \|\theta - \theta_0\| < \delta\Big\} > e^{L_n^2}\Big) \to 0$$

Then the marginal posterior converges to normality as follows

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