BNP NETWORK

NICOSIA

april 2022

Lecture 1

Bayesian Uncertainty Quantification

a review

Aad van der Vaart
TU Delft

April 2022

Bayesian Uncertainty Quantification

a review (not a tutorial?)

Questions, examples, results on whether the spread of a posterior distribution gives a reliable indication of the possible error of its center as estimate of a parameter.

LECTURE 1

Parametric Bernstein-von Mises

Classical Nonparametric BvM

LECTURE 2

Semiparametric BvM

Nonparametric Smoothing

LECTURE 3

Weak Bernstein-von Mises

Credible Bands

Not : high-dimensional and variable selection

Not : much on rates

Parametric Bernstein-von Mises

Bernstein-von Mises

$\theta \sim \Pi$ on $\mathbb{R}^d$, smooth positive density

$X_1 \dots X_n \mid \theta \overset{iid}{\sim} p_\theta$, $\theta \mapsto \sqrt{p_\theta}$ differentiable in L_2

density

$\dot{\ell}_\theta = \frac{\partial}{\partial \theta} \log p_\theta$ $\qquad$ $i_\theta = \mathrm{Cov}_\theta(\dot{\ell}_\theta(X_1))$

score function $\qquad$ Fisher information

Bernstein-von Mises

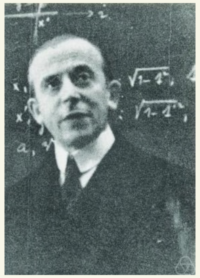
$\theta \sim \Pi$ on $\mathbb{R}^d$, smooth positive density

$X_1 \dots X_n \mid \theta \overset{iid}{\sim} p_\theta$, $\theta \mapsto \sqrt{p_\theta}$ differentiable in L_2

density

$\dot{\ell}_\theta = \frac{\partial}{\partial \theta} \log p_\theta$ $\quad$ $i_\theta = \mathrm{Cov}_\theta(\dot{\ell}_\theta(X_1))$

score function $\quad$ Fisher information



Le Cam: "called this way because it was discovered by Laplace"

Bernstein-von Mises

Le Cam: "called this way because it was discovered by Laplace"

$\theta \sim \Pi$ on $\mathbb{R}^d$, smooth positive density

$X_1 \dots X_n \mid \theta \overset{iid}{\sim} p_\theta$, $\theta \mapsto \sqrt{p_\theta}$ differentiable in L_2

density

$\dot{\ell}_\theta = \frac{\partial}{\partial \theta} \log p_\theta$ $\qquad i_\theta = \mathrm{Cov}_\theta(\dot{\ell}_\theta(X_1))$

score function $\qquad$ Fisher information

$\forall \varepsilon > 0 \ \exists$ tests φ_n with $P_{\theta_0}^n \varphi_n \to 0$, $\sup_{\theta : \|\theta - \theta_0\| > \varepsilon} P_\theta^n (1 - \varphi_n) \to 0$

Bernstein-von Mises

Le Cam: "called this way because it was discovered by Laplace"

$\theta \sim \Pi$ on $\mathbb{R}^d$, smooth positive density

$X_1 \dots X_n \mid \theta \overset{iid}{\sim} p_\theta$, $\theta \mapsto \sqrt{p_\theta}$ differentiable in L_2

density

$\dot{\ell}_\theta = \frac{\partial}{\partial \theta} \log p_\theta$ $\qquad i_\theta = Cov_\theta(\dot{\ell}_\theta(X_1))$

score function $\qquad$ Fisher information

$\forall \varepsilon > 0 \; \exists$ tests φ_n with $P_{\theta_0}^n \varphi_n \to 0$, $\sup_{\theta : \|\theta - \theta_0\| \geq \varepsilon} P_\theta^n (1-\varphi_n) \to 0$

THM For any $\hat{\theta}_n$ with $\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n i_{\theta_0}^{-1} \dot{\ell}_{\theta_0}(X_i) + o_{P_{\theta_0}}(1)$

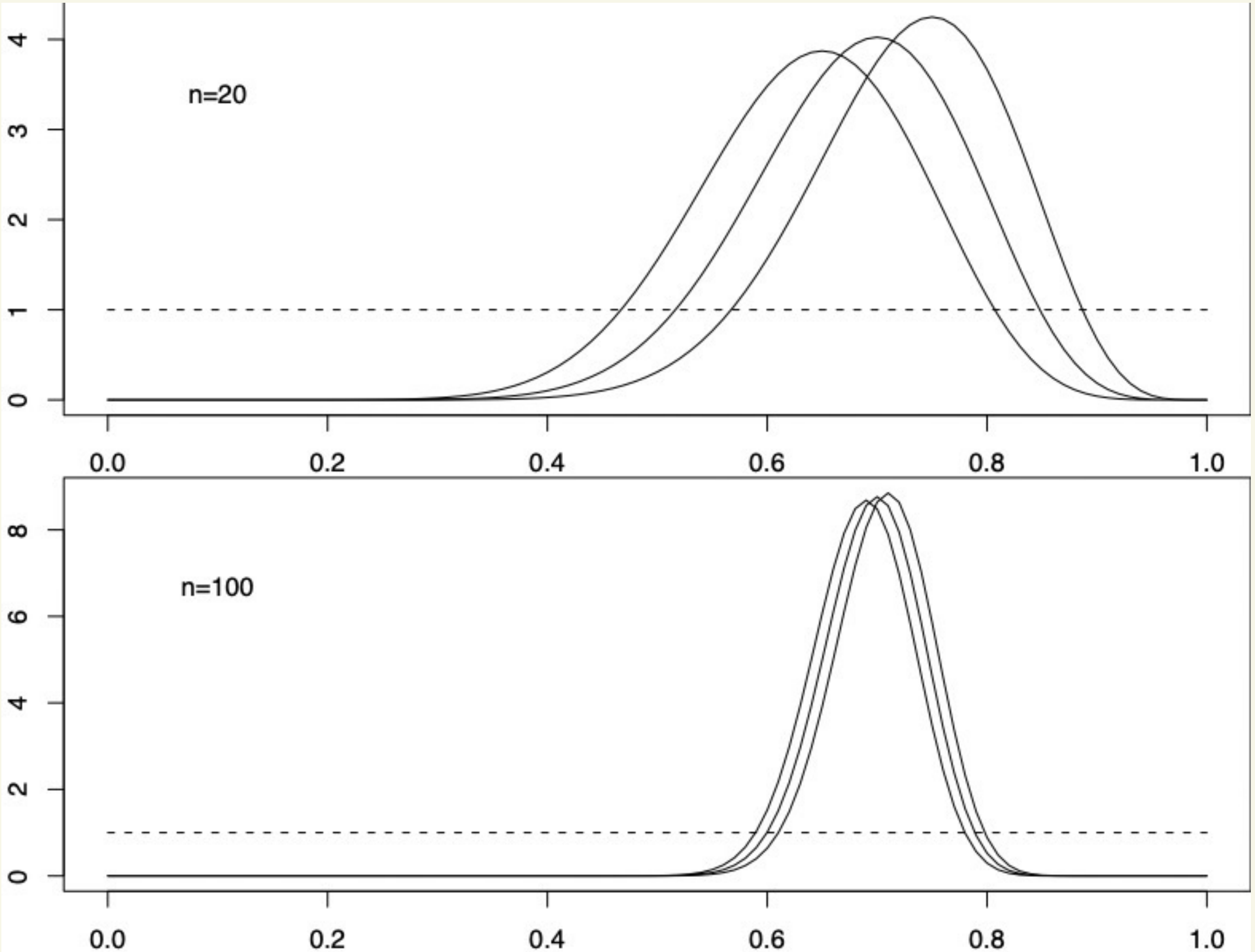
$$\left\| \Pi(\theta \in \cdot \mid X_1 \dots X_n) - N\left(\hat{\theta}_n, \frac{1}{n} i_{\theta_0}^{-1}\right) \right\| \xrightarrow{P_{\theta_0}^n} 0$$

total variation

$\|P - Q\| = \sup_B |P(B) - Q(B)|$

- Test condition automatic if $\theta \in \Theta$ compact and θ identifiable
- $\hat{\theta}_n$ can be taken MLE under further conditions.
- Generalises to LAN models.

$\theta \sim$ uniform $[0,1]$
$X|\theta \sim$ binom(n, θ)



THM For any $\hat{\theta}_n$ with $\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} i_{\theta_0}^{-1} \dot{\ell}_{\theta_0}(X_i) + o_p(1)$

$$\left\| \Pi(\theta \in \cdot \mid X_1 \cdots X_n) - N\left(\hat{\theta}_n, \tfrac{1}{n} i_{\theta_0}^{-1}\right) \right\| \xrightarrow{P_{\theta_0}^n} 0$$

total variation

Informal $\quad \sqrt{n}(\theta - \hat{\theta}_n) \mid X_1 \cdots X_n \overset{TV}{\rightsquigarrow} N(0, i_{\theta_0}^{-1})$, in $P_{\theta_0}^n$-probability.

Compare to $\quad \sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow N(0, i_{\theta_0}^{-1})$, under θ_0.

convergence in distribution

THM For any $\hat{\theta}_n$ with $\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} i_{\theta_0}^{-1} \dot{\ell}_{\theta_0}(X_i) + o_p(1)$

$$\left\| \Pi(\theta \in \cdot \mid X_1 \cdots X_n) - N\left(\hat{\theta}_n, \tfrac{1}{n} i_{\theta_0}^{-1}\right) \right\| \xrightarrow{P_{\theta_0}^n} 0$$

total variation

Informal $\quad \sqrt{n}(\theta - \hat{\theta}_n) \mid X_1 \cdots X_n \overset{TV}{\rightsquigarrow} N(0, i_{\theta_0}^{-1})$, in $P_{\theta_0}^n$-probability.

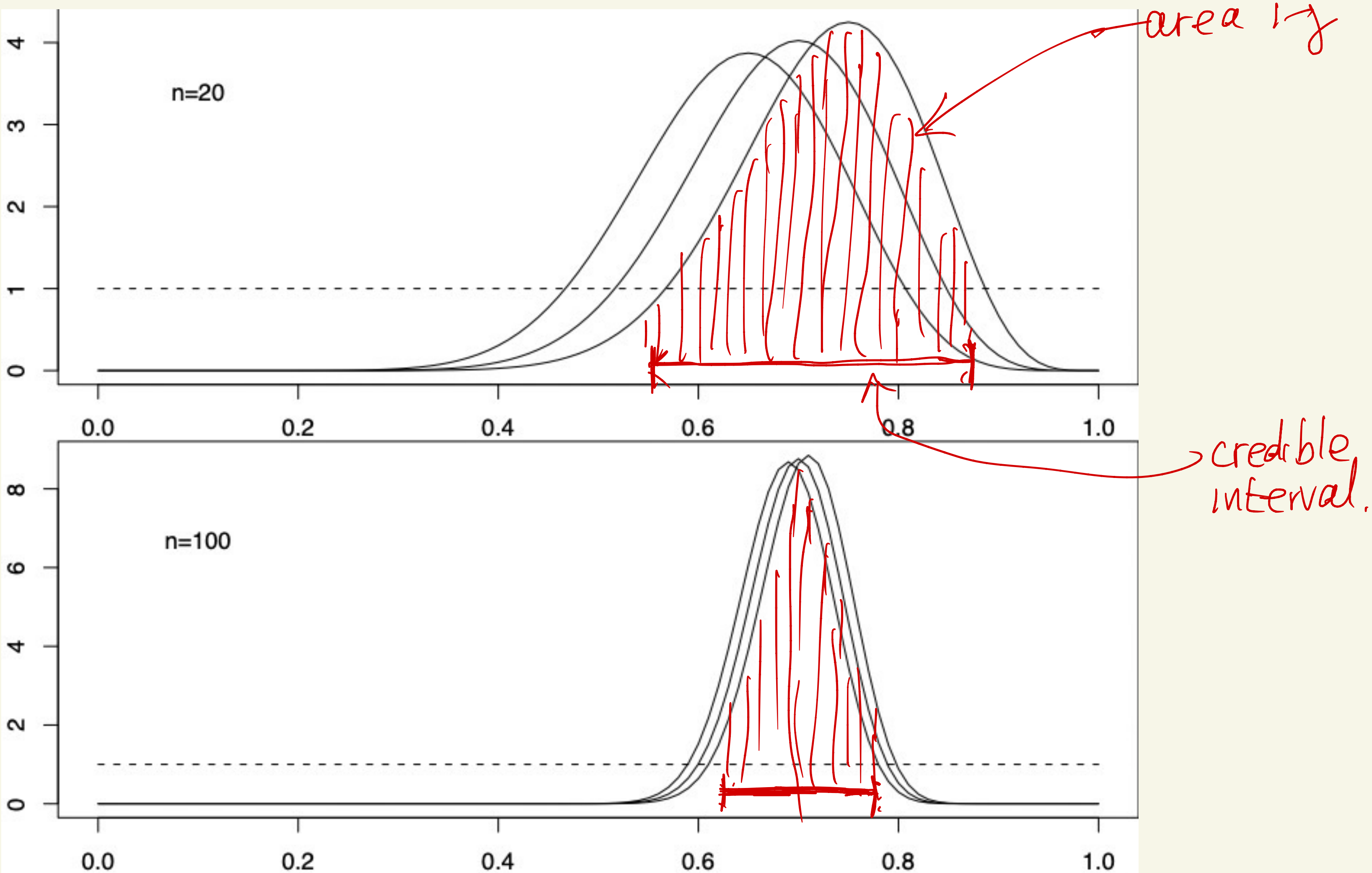
Compare to $\quad \sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow N(0, i_{\theta_0}^{-1})$, under θ_0.

convergence in distribution

$$\sqrt{n}(\theta - \hat{\theta}_n) \mid X_1 \cdots X_n \overset{\mathcal{D}}{\approx} \sqrt{n}(\hat{\theta}_n - \theta_0) \mid \theta_0$$

bootstrap-like

$\theta \sim \text{uniform}[0,1]$
$X|\theta \sim \text{binom}(n, \theta)$



"$100(1-\gamma)\%$ of the times we take data and compute posterior, the credible interval will cover the true parameter."

Credible Sets

$$\left\| \Pi(\theta \in \cdot \mid X_1 \cdots X_n) - N\left(\hat{\theta}_n, \frac{1}{n} i_{\theta_0}^{-1}\right) \right\| \xrightarrow{P_{\theta_0}^n} 0$$

COR $\Pi(\theta \in C_n^\gamma \mid X_1 \cdots X_n) = 1-\gamma \implies N\left(\hat{\theta}_n, \frac{1}{n} i_{\theta_0}^{-1}\right)(C_n^\gamma) \xrightarrow{P_{\theta_0}^n} 1-\gamma$

Credible Sets

$$\left\| \Pi(\theta \in \cdot \mid X_1 \cdots X_n) - N\left(\hat{\theta}_n, \tfrac{1}{n} i_{\theta_0}^{-1}\right) \right\| \xrightarrow{P_{\theta_0}^n} 0$$

COR $\Pi(\theta \in C_n^\gamma \mid X_1 \cdots X_n) = 1-\gamma \implies N\left(\hat{\theta}_n, \tfrac{1}{n} i_{\theta_0}^{-1}\right)(C_n^\gamma) \xrightarrow{P_{\theta_0}^n} 1-\gamma$

This does NOT imply $P_{\theta_0}^n(\theta_0 \in C_n^\gamma) \to 1-\gamma$

For central C_n^γ it does.

Credible Sets

$$\left\| \Pi(\theta \in \cdot \mid X_1 \cdots X_n) - N\left(\hat{\theta}_n, \tfrac{1}{n} i_{\theta_0}^{-1}\right) \right\| \xrightarrow{P_{\theta_0}^n} 0$$

COR $\Pi(\theta \in C_n^\gamma \mid X_1 \cdots X_n) = 1-\gamma \implies N\left(\hat{\theta}_n, \tfrac{1}{n} i_{\theta_0}^{-1}\right)(C_n^\gamma) \xrightarrow{P_{\theta_0}^n} 1-\gamma$

This does NOT imply $P_{\theta_0}^n(\theta_0 \in C_n^\gamma) \to 1-\gamma$

For central C_n^γ it does.

$$F_{g^T\theta}(y \mid X_1 \cdots X_n) := \Pi(g^T\theta \le y \mid X_1 \cdots X_n)$$

normal quantiles

THM $F_{g^T\theta}^{-1}(\gamma \mid X_1 \cdots X_n) = g^T\hat{\theta}_n - \Phi^{-1}(\gamma)\sqrt{\dfrac{g^T i_{\theta_0}^{-1} g}{n}} + o_{P_{\theta_0}^n}\left(\dfrac{1}{\sqrt{n}}\right).$

Consequently:

$$P_{\theta_0}^n\Big(\underbrace{F_{g^T\theta}^{-1}\left(\tfrac{\gamma}{2} \mid X_1 \cdots X_n\right) \le g^T\theta_0 \le F_{g^T\theta}^{-1}\left(1-\tfrac{\gamma}{2} \mid X_1 \cdots X_n\right)}_{\text{credible interval}} \Big) \to 1-\gamma$$

Classical Nonparametric BvM

Ferguson
Lo
Hjort
Kim & Lee
James
Franssen & vdV

Dirichlet Process

empirical measure

$\mathbb{P}_n(A) = \frac{1}{n} \#(1 \le i \le n : X_i \in A)$

$$P \sim DP(\alpha)$$

$$X_1 \cdots X_n \mid P \overset{iid}{\sim} P$$

$$\Rightarrow \quad P \mid X_1 \cdots X_n \sim DP(\alpha + n\mathbb{P}_n) \quad , \quad \mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$$

Dirichlet Process

finite measure

$$P \sim DP(\alpha)$$

$$X_1 \cdots X_n \mid P \overset{iid}{\sim} P$$

empirical measure

$\mathbb{P}_n(A) = \frac{1}{n} \#(1 \le i \le n : X_i \in A)$

$$\Rightarrow \quad P \mid X_1 \cdots X_n \sim DP(\alpha + n\mathbb{P}_n) \quad , \quad \mathbb{P}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

THM For any A

$$\left\| \mathcal{L}(P(A) \mid X_1 \cdots X_n) - N\left(\mathbb{P}_n(A), \frac{P_0(A) P_0(A^c)}{n}\right) \right\|_{TV} \xrightarrow{P_{F_0}^n} 0$$

Dirichlet Process

finite measure

$P \sim DP(\alpha)$

$X_1 \cdots X_n \mid P \overset{iid}{\sim} P$

empirical measure $\mathbb{P}_n(A) = \frac{1}{n}\#(1 \le i \le n : X_i \in A)$

$\Rightarrow \quad P \mid X_1 \cdots X_n \sim DP(\alpha + n\mathbb{P}_n) \quad , \quad \mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$

THM For any A

$$\left\| \mathcal{L}(P(A) \mid X_1 \cdots X_n) - N\left(\mathbb{P}_n(A), \frac{P_0(A)P_0(A^c)}{n}\right) \right\|_{TV} \xrightarrow{P_{F_0}^n} 0$$

$$\sqrt{n}\,(P(A) - \mathbb{P}_n(A)) \mid X_1 \cdots X_n \approx \sqrt{n}\,(\mathbb{P}_n(A) - P_0(A)) \mid P_0$$

bootstrap-like

Dirichlet Process

$P \sim DP(\alpha)$

$X_1 \cdots X_n \mid P \overset{iid}{\sim} P$

$\Rightarrow \quad P \mid X_1 \cdots X_n \sim DP(\alpha + n\mathbb{P}_n)$, $\mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$

empirical measure
$\mathbb{P}_n(A) = \frac{1}{n}\#(1 \le i \le n : X_i \in A)$

THM For any A

$$\left\| \mathcal{L}(P(A) \mid X_1 \cdots X_n) - N\left(\mathbb{P}_n(A), \frac{P_0(A)P_0(A^c)}{n}\right) \right\|_{TV} \xrightarrow{P^n_{F_0}} 0$$

proof 1 Use $P(A) \mid X_1 \cdots X_n \sim \underbrace{Be(\alpha(A) + n\mathbb{P}_n(A), \alpha(A^c) + n\mathbb{P}_n(A^c))}$

mean $\frac{\alpha(A) + n\mathbb{P}_n(A)}{\alpha(\mathfrak{X}) + n} \approx \mathbb{P}_n(A)$

variance $\approx \frac{\mathbb{P}_n(A)\mathbb{P}_n(A^c)}{n}$.

proof 2 Use $P(A) \mid X_1 \cdots X_n \sim P(A) \mid n\mathbb{P}_n(A)$

and BvM for $n\mathbb{P}_n(A) \mid P(A) \sim bin(n, P(A))$. $\square$

Dirichlet Process

(Lo, 1991)

empirical measure

$\mathbb{P}_n(A) = \frac{1}{n}\#(1 \le i \le n : X_i \in A)$

$$P \sim DP(\alpha)$$

$$X_1 \cdots X_n \mid P \overset{iid}{\sim} P$$

$$\Rightarrow \quad P \mid X_1 \cdots X_n \sim DP(\alpha + n\mathbb{P}_n), \quad \mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$$

$X_1 \cdots X_n$ real, $F(t) = P[-\infty, t]$ cdf of P

Dirichlet process $\quad (F(t) : t \in \mathbb{R}) \mid X_1 \cdots X_n$

Empirical cdf $\quad (\mathbb{F}_n(t) : t \in \mathbb{R})$

Dirichlet Process

(Lo, 1991)

empirical measure
$\mathbb{P}_n(A) = \frac{1}{n}\#(1\le i\le n : X_i \in A)$

$$P \sim DP(\alpha)$$
$$X_1 \cdots X_n \mid P \overset{iid}{\sim} P$$
$$\Rightarrow \quad P \mid X_1 \cdots X_n \sim DP(\alpha + n\mathbb{P}_n), \qquad \mathbb{P}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$$

$X_1 \cdots X_n$ real, $F(t) = P(-\infty, t]$ cdf of P

Dirichlet process $\quad (F(t) : t \in \mathbb{R}) \mid X_1 \cdots X_n$

Empirical cdf $\quad (\mathbb{F}_n(t) : t \in \mathbb{R})$

Brownian bridge

THM $\quad (\sqrt{n}(F(t) - \mathbb{F}_n(t)) : t \in \mathbb{R}) \mid X_1 \cdots X_n \rightsquigarrow (B \circ F_0(t) : t \in \mathbb{R})$, in $P_{F_0}^n$-probability

compare $\quad (\sqrt{n}(\mathbb{F}_n(t) - F_0(t)) : t \in \mathbb{R}) \rightsquigarrow (B \circ F_0(t) : t \in \mathbb{R})$, under F_0^n.

Donsker's theorem

Formal statement

$$\sup_{h \in BL_1} \left| E\left(h(\sqrt{n}(F - \mathbb{F}_n)) \mid X_1 \cdots X_n\right) - Eh(B \circ F_0) \right| \xrightarrow{F_0^n} 0$$

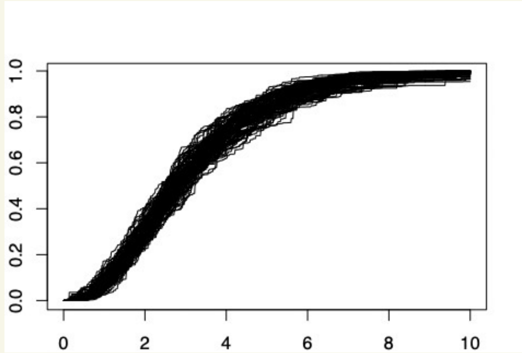
functions $h : \ell^\infty(\mathbb{R}) \to [0,1]$ with $|h(F) - h(G)| \le \|F - G\|_\infty$

Credible Bands

or $ECF(X_1 \cdots X_n)$

Determine ξ_γ with $\Pi\big(\underbrace{F : \|F - \mathbb{F}_n\|_\infty \le \xi_\gamma}_{C_n^\gamma} \mid X_1 \cdots X_n\big) = 1-\gamma$

THM $P_{F_0}^n\left(F_0 \in C_n^\gamma\right) \to 1-\gamma$

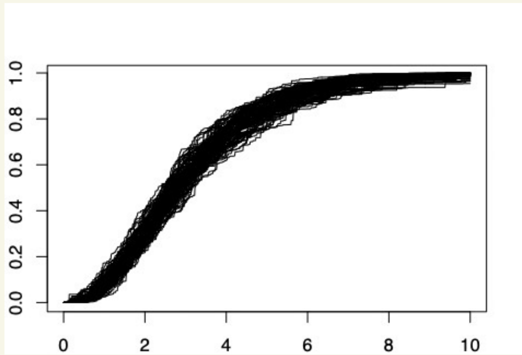


Credible Bands

or $E(F|X_1 \cdots X_n)$

$$\text{Determine } \xi_\gamma \text{ with } \Pi\big(\underbrace{F : \|F - \mathbb{F}_n\|_\infty \leq \xi_\gamma}_{C_n^\gamma} \,|\, X_1 \cdots X_n\big) = 1-\gamma$$

THM $P^n_{F_0}(F_0 \in C_n^\gamma) \to 1-\gamma$



- could determine ξ_γ by simulation (BB)
- could use weighted norm instead
- slightly wider than pointwise credible intervals, based on $\pi(F(t)|X_1 \cdots X_n)$, for fixed t.

	width
pointwise	$\approx \frac{2}{\sqrt{n}} \cdot 1.96 \sqrt{F_0(t)(1-F_0(t))}$
uniform	$\approx \frac{1}{\sqrt{n}} \, 1.358$ (KS-table)

Dirichlet Process

(James, 2008)

$$P \sim DP(\alpha)$$
$$X_1 \cdots X_n \mid P \overset{iid}{\sim} P$$
$$\Rightarrow \quad P \mid X_1 \cdots X_n \sim DP(\alpha + n\mathbb{P}_n) \quad , \quad \mathbb{P}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$$

empirical measure
$\mathbb{P}_n(A) = \frac{1}{n}\#(1 \le i \le n : X_i \in A)$

Take collection $\mathcal{F}$ of functions $f : \mathcal{X} \to \mathbb{R}$

Identify $P \longleftrightarrow (f \to Pf = \int f\,dP)$
$\mathcal{F} \to \mathbb{R}$

$$P \in \ell^\infty(\mathcal{F}) = \{ z : \mathcal{F} \to \mathbb{R} \text{, bounded} \} \qquad \|z\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |z(f)|$$

DEF $\mathcal{F}$ is P_0-Donsker if $\sqrt{n}(\mathbb{P}_n - P_0) \rightsquigarrow \mathbb{G}_{P_0}$ in $\ell^\infty(\mathcal{F})$

tight, Gaussian
$E(\mathbb{G}_{P_0} f)^2 = P_0(f - P_0 f)^2$

THM If $\mathcal{F}$ is P_0-Donsker

$$\sqrt{n}(P - \mathbb{P}_n) \mid X_1 \cdots X_n \rightsquigarrow \mathbb{G}_{P_0} \text{ , in } P_0^n\text{-probability}$$

Pitman–Yor Process

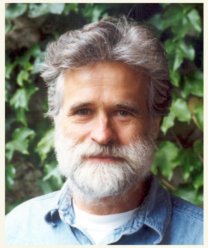
(James 2008, Franssen+vdV, 2021)

$$P=\sum_{i=1}^{\infty} W_i \delta_{\theta_i}, \quad W_i = V_i\prod_{j<i}(1-V_j), \quad V_j \overset{iid}{\sim} \mathrm{Beta}(1-\sigma, M+j\sigma), \quad \theta_j \overset{iid}{\sim} \bar{\alpha}$$

$\sigma \geq 0$

$\sigma = 0$: Dirichlet

$$X_1,\dots,X_n \mid P \overset{iid}{\sim} P$$

Pitman–Yor Process

(James, Franssen + vdV)

$$P = \sum_{i=1}^{\infty} W_i \delta_{\theta_i}, \quad W_i = V_i \prod_{j<i}(1-V_j), \quad V_j \overset{ind}{\sim} \mathrm{Beta}(1-\sigma, M+j\sigma), \quad \theta_j \overset{iid}{\sim} \bar{\alpha}$$

$\sigma \geq 0$

$\sigma = 0$: Dirichlet

$$X_1 \cdots X_n \mid P \overset{iid}{\sim} P$$

distinct values in $X_1 \cdots X_n$

empirical of distinct values.

THM $\sqrt{n}\left(P - \mathbb{P}_n - \frac{\sigma K_n}{n}(\bar{\alpha} - \tilde{\mathbb{P}}_n)\right) \mid X_1 \cdots X_n \rightsquigarrow$ Gaussian process, in P_0^n-prob

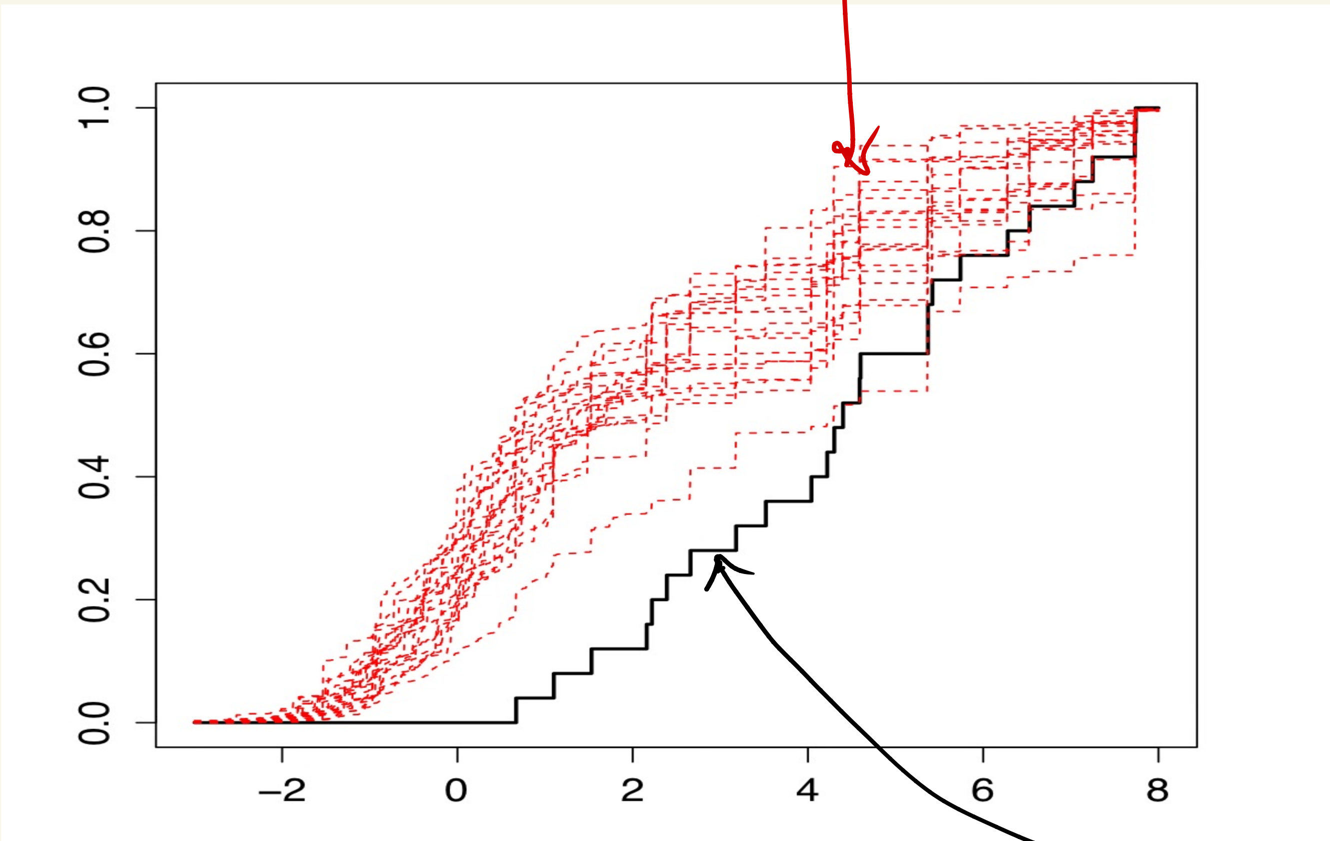
If $\sigma > 0$ then credible sets are correct only if $\frac{K_n}{\sqrt{n}} \xrightarrow{P_0} 0$

i.e. if P_0 is discrete with atoms decreasing fast.

counter-intuitive: $\sigma > 0$ creates bigger K_n than $\sigma = 0$

but does not work for continuous P_0.

20 draws from posterior of $PY(\sigma=\frac{1}{2}, M=1, \bar{\alpha}=N(0,1))$



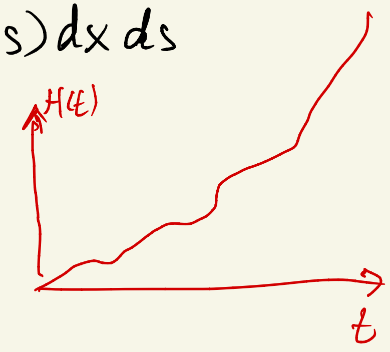
empirical cdf of sample of size 25 from $N(1,4)$

Survival Analysis

(Hjort, Kim and Lee 2001, 04)

$H \sim$ independent increment process $\nu_H(dx, ds) = \frac{1}{s}\, q(x,s)\, dx\, ds$

$T_1 \cdots T_n \mid H \overset{iid}{\sim}$ cumulative hazard H $\qquad dH = \frac{dF}{1-F_-}$



$C_1 \cdots C_n \mid G_0 \overset{iid}{\sim} G_0$

Observe $X_i = T_i \wedge C_i$, $\Delta_i = 1\{T_i \le C_i\}$, $i = 1, \dots, n.$

Survival Analysis

(Hjort, Kim and Lee 2001, 04)

$H \sim$ independent increment process $\nu_H(dx, ds) = \frac{1}{s} q(x,s)\,dx\,ds$

$T_1 \cdots T_n \mid H \overset{iid}{\sim}$ cumulative hazard $dH = \frac{dF}{1-F_-}$

$C_1 \cdots C_n \mid G_0 \overset{iid}{\sim} G_0$

Observe $X_i = T_i \wedge C_i$, $\Delta_i = 1\{T_i \le C_i\}$, $i = 1, \cdots, n$.

$0 << q_0 << \infty$ $\quad \frac{1}{2} < \alpha < 1$

THM If $\sup_{x,s} (1-s)^{\kappa} q(x,s) < \infty$, $\kappa > 0$, $\sup_{x} |q(x,s) - q_0(x)| = O(s^{\alpha})$ as $s \to 0$

then $\sqrt{n}(H - \hat{H}_n) \mid X_1 \cdots X_n, \Delta_1 \cdots \Delta_n \rightsquigarrow B \circ U_0$, in $P^{\infty}_{F_0 G_0}$

Nelson-Aalen

$\int_0^{\cdot} \frac{dH}{(1-F_-)(1-G_-)}$

...

Survival Analysis

(Hjort, Kim and Lee 2001, 04)

$H \sim$ independent increment process $\nu_H(dx, ds) = \frac{1}{s} q(x,s)\,dx\,ds$

$T_1 \cdots T_n \mid H \overset{iid}{\sim}$ cumulative hazard $dH = \dfrac{dF}{1-F_-}$

$C_1 \cdots C_n \mid G_0 \overset{iid}{\sim} G_0$

Observe $X_i = T_i \wedge C_i$, $\Delta_i = 1\{T_i \le C_i\}$, $i = 1, \ldots, n$.

$0 << q_0 << \infty$ $\quad \frac{1}{2} < \alpha < 1$

THM If $\sup_{x,s} (1-s)^{\kappa} q(x,s) < \infty$ ($\kappa > 0$), $\sup_x |q(x,s) - q_0(x)| = O(s^{\alpha})$ as $s \to 0$

then $\sqrt{n}(H - \hat{H}_n) \mid X_1 \cdots X_n, \Delta_1 \cdots \Delta_n \rightsquigarrow B \circ U_0$, in $P^{\infty}_{F_0 G_0}$

Nelson-Aalen

$\int_0^{\cdot} \dfrac{dH}{(1-F_0)_-(1-G_0)_-}$

(COUNTER) EXAMPLE $\quad \nu_H(dx, ds) = s^{-1} \dfrac{s^{a(x)-1}(1-s)^{b(x)-1}}{B(a(x), b(x))} ds\, d\Lambda(x)$

Extended Beta

Consistency holds iff $a(x) \equiv 1$.

Semiparametric BvM

Castillo

Castillo & Rousseau

Franssen & Nguyen & vdV

Ray & vdV

Nickl et al.

Semiparametric BvM

$\theta \sim \Pi$ on $\mathbb{R}^1$, $\eta \sim \Pi_\eta$ (nuisance parameter, general)

$X_1 \cdots X_n \mid \theta, \eta \overset{iid}{\sim} P_{\theta,\eta}$

Semiparametric BvM

$\theta \sim \Pi$ on $\mathbb{R}^1$, $\eta \sim \Pi_\eta$ (nuisance parameter, general)

$X_1 \cdots X_n \mid \theta, \eta \overset{iid}{\sim} P_{\theta,\eta}$

$\dot{\ell}_{\theta,\eta} = \frac{\partial}{\partial \theta} \log p_{\theta\eta}$ ordinary score

$g = \frac{\partial}{\partial t}\Big|_{t=0} \log p_{\theta,\eta_t}$ ${}_\eta\dot{\mathcal{P}}_{\theta,\eta} = \{\text{all scores for } \eta\}$

for curve $t \mapsto \eta_t$ with $\eta_0 = \eta$

$\Pi_{\theta\eta} \dot{\ell}_{\theta\eta} = \underset{g \in \overline{\text{lin}}\, {}_\eta\dot{\mathcal{P}}_{\theta\eta}}{\operatorname{argmin}} \; P_{\theta\eta}(\dot{\ell}_{\theta\eta} - g)^2$ projection onto $\overline{\text{lin}}\, {}_\eta\dot{\mathcal{P}}_{\theta\eta}$

$\tilde{\ell}_{\theta,\eta} := \dot{\ell}_{\theta,\eta} - \Pi_{\theta\eta} \dot{\ell}_{\theta\eta}$ $\tilde{i}_{\theta\eta} = \operatorname{var}_{\theta,\eta}(\tilde{\ell}_{\theta,\eta}(X))$

efficient score   efficient information

Semiparametric BvM

$\theta \sim \Pi$ on $\mathbb{R}^1$, $\eta \sim \Pi_\eta$ (nuisance parameter, general)

$X_1 \cdots X_n \mid \theta, \eta \overset{iid}{\sim} P_{\theta,\eta}$

$\dot{\ell}_{\theta,\eta} = \frac{\partial}{\partial\theta} \log p_{\theta\eta}$ ordinary score

$g = \frac{\partial}{\partial t}\Big|_{t=0} \log p_{\theta,\eta_t}$ $\quad {}_\eta\dot{P}_{\theta,\eta} = \{\text{all scores for } \eta\}$

for curve $t \mapsto \eta_t$ with $\eta_0 = \eta$

$\Pi_{\theta\eta} \dot{\ell}_{\theta\eta} = \underset{g \in \overline{\text{lin}}\, {}_\eta\dot{P}_{\theta\eta}}{\text{argmin}} \; P_{\theta\eta}(\dot{\ell}_{\theta\eta} - g)^2$ projection onto $\overline{\text{lin}}\, {}_\eta\dot{P}_{\theta\eta}$

$\tilde{\ell}_{\theta,\eta} := \dot{\ell}_{\theta,\eta} - \Pi_{\theta\eta}\dot{\ell}_{\theta\eta}$ $\quad \tilde{i}_{\theta\eta} = \text{var}_{\theta,\eta}(\tilde{\ell}_{\theta,\eta}(X))$

efficient score $\quad$ efficient information

DESIRED THM For $\hat{\theta}_n$ with $\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{i}_{\theta_0,\eta_0}^{-1} \tilde{\ell}_{\theta_0\eta_0}(X_i) + o_{P_{\theta_0\eta_0}}(1)$

$$\sqrt{n}(\theta - \hat{\theta}_n) \mid X_1 \cdots X_n \rightsquigarrow N(0, \tilde{i}_{\theta_0\eta_0}^{-1}) \text{ in } P^n_{\theta_0\eta_0}\text{-probability.}$$

Gives correct credible intervals, as before.

Bernstein-von Mises - TV

(Castillo, 2008)

THM If $\exists\ \eta \mapsto \tilde{\eta}_n(\theta,\eta)$ with

- $\sum_{i=1}^{n} \log \frac{p_{\theta,\eta}}{p_{\theta_0,\tilde{\eta}_n(\theta,\eta)}}(X_i) = \sqrt{n}(\theta-\theta_0)\mathbb{G}_n \tilde{\ell}_{\theta_0,\eta_0} - \frac{1}{2}n\tilde{I}_{\theta_0,\eta_0}(\theta-\theta_0)^2 + R_n(\theta,\eta)$, for

$$\sup_{\substack{\theta\in\Theta_n \\ \eta\in\mathcal{H}_n}} \frac{R_n(\theta,\eta)}{1+n|\theta-\theta_0|^2} \to 0$$

likelihood expansion

- $$\sup_{\substack{\theta\in\Theta_n \\ \eta\in\mathcal{H}_n}} \frac{\log \dfrac{d\Pi_\eta \circ \tilde{\eta}_n(\theta,\eta)^{-1}}{d\Pi_\eta \circ \tilde{\eta}_n(\theta_0,\eta)^{-1}}}{1+n(\theta-\theta_0)^2} \to 0$$

change of measure

- $\Pi_n(\theta\in\Theta_n, \eta\in\mathcal{H}_n \mid X_1\cdots X_n) \to 1$

$\Pi_n(\eta\in\tilde{\eta}_n(\theta,\mathcal{H}_n) \mid X_1-X_n, \theta=\theta_0) \to 1$

localisation

then desired theorem holds in TV.

$\tilde{\eta}_n$ least favourable, expansion in both θ and η, typically need rate on η.

If estimating θ does not become more difficult by not knowing η,
then likelihood expansion will only be in θ and change of measure condition disappears.

→ BvM valid under weak conditions, prior on η unimportant.

Example: estimate point of symmetry θ in symmetric density $p_{\theta,\eta}(x) = \eta(|x-\theta|)$.

Otherwise the prior on η may cause a bias in the posterior for θ
if not chosen with care (oversmoothing)

EXAMPLE - COX MODEL

(Castillo, 2008)

$\log \lambda \sim$ Riemann-Liouville (α)

($\alpha - \frac{1}{2}$ times integrated BM)

$\theta \perp\!\!\!\perp \lambda$

$T \mid Z \sim$ hazard $\lambda(t) e^{\theta Z}$

$T \perp\!\!\!\perp C \mid Z$

$X = (T \wedge C, Z, \Delta = 1_{T \le C})$

least favourable

THM If $\log \lambda_0 \in C^{\beta}[0,\tau]$, $\beta > 3/2$, $\gamma_0 \in C^{2\beta/3}[0,\tau]$

then BvM holds for $\alpha \in \left(3/2, \frac{4\beta}{3} - \frac{1}{2}\right)$