Asymptotic properties of Bayesian nonparametrics and semiparametrics

J. Rousseau

CEREMADE, Université Paris-Dauphine & ENSAE - CREST

Spring School, Slyt

Outline

Outline

- **Sampling model and prior models**

• $X^n|\theta \sim P_\theta$ on $\mathcal{X}_n$ with $\theta \in \Theta$

• θ : unknown $\rightarrow$ random variable . Π = prior proba on $(\Theta, \mathcal{A})$

- **joint, marginal and posterior distributions**

• Joint $(X^n, \theta) \sim P_\theta \times \Pi$

• Posterior : $\Pi(d\theta|X^n)$ If dominated model $f_\theta = dP_\theta/d\mu$

$$\Pi(d\theta|X^n) = \frac{f_\theta(X^n)\Pi(d\theta)}{m(X^n)}, \quad m(X^n) = \int_\Theta f_\theta(X^n)\Pi(d\theta)$$

• Marginal of X^n : $m(X^n)$

Examples

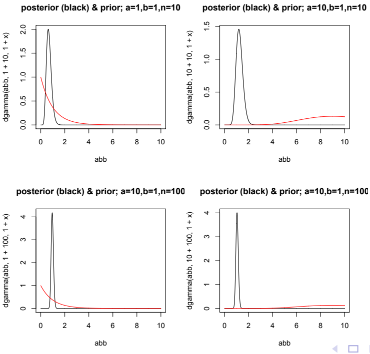
- **Parametric**

Poisson model : $X^n = (X_1, \cdots, X_n)$, $X_i \sim \mathcal{P}(\theta)$

Prior on $\theta > 0$ $\Gamma(a, b)$

• Posterior

$$\Pi(\theta|X^n) \equiv \Gamma(a+n, b+n\bar{X}_n), \quad \bar{X}_n = \sum_i X_i/n$$



What do we observe ?

- Posterior concentration : posterior shrinks towards $\theta_0 = 1$

What do we observe ?

- Posterior concentration : posterior shrinks towards $\theta_0 = 1$
- Prior becomes less and less influential as $n \uparrow$.

What do we observe ?

- Posterior concentration : posterior shrinks towards $\theta_0 = 1$
- Prior becomes less and less influential as $n \uparrow$.
- Asymptotic normality of the posterior : BvM

What do we observe ?

- Posterior concentration : posterior shrinks towards $\theta_0 = 1$
- Prior becomes less and less influential as $n \uparrow$.
- Asymptotic normality of the posterior : BvM
- General features in regular parametric models

What do we observe ?

- Posterior concentration : posterior shrinks towards $\theta_0 = 1$
- Prior becomes less and less influential as $n \uparrow$.
- Asymptotic normality of the posterior : BvM
- General features in regular parametric models
- How can we extend these results in large dimensional models ?

Outline

1. Bayesian statistics
   - General setup
   - Scope of the talk
2. Bayesian nonparametrics
3. On the consistency and posterior concentration rates
   - Definitions
4. Some general Theorems
   - General Theorems on consistency
5. Mixture models for smooth densities
   - Approximative properties of some exponential Kernels
   - Posterior concentration rates I
6. Empirical Bayes
   - Driving example
   - Change of measure
7. Application to DP mixtures of Gaussians
8. Semi - parametric : BvM
   - Semi-parametric Bayesian methods
   - BvM in the parametric case

Scope of the talk

- **General notions on Bayesian nonparametrics**

Some typical prior models
output of posteriors

- **Posterior concentrations and consistency**

Regular priors
empirical Bayes

- **Semi-parametric inference : BvM**

Some positive results
Some negative results
Understanding credible regions

First : posterior distribution = more than point estimation

- **What can we do with the posterior distribution ?**
  - Point estimators : Loss function : $\ell : \Theta \times \mathcal{D} \to \mathbb{R}^+$
    Bayes estimator

$$\delta^\pi(X^n) = \text{argmin}_\delta E^\pi\left[\ell(\theta, \delta)|X^n\right]$$

e.g. $\ell(\theta, \theta') = \|\theta - \theta'\|_2^2$ then $\delta^\pi(X^n) = E^\pi\left(\theta|X^n\right)$.

First : posterior distribution = more than point estimation

- **What can we do with the posterior distribution ?**
  - Point estimators : Loss function : $\ell : \Theta \times \mathcal{D} \to \mathbb{R}^+$
    Bayes estimator

    $$\delta^\pi(X^n) = \text{argmin}_\delta E^\pi \left[\ell(\theta, \delta) | X^n\right]$$

    e.g. $\ell(\theta, \theta') = \|\theta - \theta'\|_2^2$ then $\delta^\pi(X^n) = E^\pi(\theta | X^n)$.
  - Credible regions : measure of uncertainty

    $$C_\alpha : \Pi\left(\theta \in C_\alpha | X^n\right) \geq 1 - \alpha$$

First : posterior distribution = more than point estimation

- **What can we do with the posterior distribution ?**
  - Point estimators : Loss function : $\ell : \Theta \times \mathcal{D} \to \mathbb{R}^+$
    Bayes estimator

    $$\delta^\pi(X^n) = \text{argmin}_\delta E^\pi\left[\ell(\theta, \delta)|X^n\right]$$

    e.g. $\ell(\theta, \theta') = \|\theta - \theta'\|_2^2$ then $\delta^\pi(X^n) = E^\pi(\theta|X^n)$.
  - Credible regions : measure of uncertainty

    $$C_\alpha : \Pi(\theta \in C_\alpha|X^n) \geq 1 - \alpha$$

  - Risk estimation

    $$\hat{R} = E^\pi\left(\ell(\theta, \hat{\delta}_n)|X^n\right)$$

First : posterior distribution = more than point estimation

- **What can we do with the posterior distribution ?**
  - Point estimators : Loss function : $\ell : \Theta \times \mathcal{D} \to \mathbb{R}^+$
    Bayes estimator

    $$\delta^\pi(X^n) = \text{argmin}_\delta E^\pi\left[\ell(\theta,\delta)|X^n\right]$$

    e.g. $\ell(\theta,\theta') = \|\theta - \theta'\|_2^2$ then $\delta^\pi(X^n) = E^\pi(\theta|X^n)$.
  - Credible regions : measure of uncertainty

    $$C_\alpha : \Pi\left(\theta \in C_\alpha|X^n\right) \geq 1 - \alpha$$

  - Risk estimation

    $$\hat{R} = E^\pi\left(\ell(\theta,\hat{\delta}_n)|X^n\right)$$

  - testing : e.g.

    $$\Pi(\Theta_0|X^n) > \Pi(\Theta_1|X^n) \quad \Leftrightarrow \quad \text{accept} \quad \Theta_0$$

Questions

- What can we say about

$$E_{\theta_0}\left[\ell(\theta_0, \hat{\delta}^\pi(X^n))\right]?$$

- What can we say about

$$P_{\theta_0}\left[\theta_0 \in C_\alpha\right]?$$

- What can we say about

$$P_\theta[\Pi(\Theta_0|X^n) > \Pi(\Theta_1|X^n)]?$$

Questions

- What can we say about

$$E_{\theta_0}\left[\ell(\theta_0, \hat{\delta}^{\pi}(X^n))\right]?$$

Standard using posterior concentration rates

- What can we say about

$$P_{\theta_0}\left[\theta_0 \in C_\alpha\right]?$$

Difficult

- What can we say about

$$P_\theta[\Pi(\Theta_0|X^n) > \Pi(\Theta_1|X^n)]?$$

Some partial results

Bayesian nonparametrics

- **Setup** Θ is infinite dimensional.
- **Examples**

• Regression function : $Y_i = f(X_i) + \epsilon_i$, $f : \mathbb{R}^d \to \mathbb{R}$

$$\Theta = L_2$$

• Density estimaton $Y_i \overset{iid}{\sim} f$

$$\Theta = \mathcal{F} = \{f : \mathbb{R}^d \to \mathbb{R}^+, \int f = 1\}$$

• classification , spectral density , intensity , conditional density, etc . . .

Examples of priors : Gaussian process priors

- **Gaussian process priors** $(\Theta, \|.\|)$ Banach space (e.g. L_2)
$\theta = f$

$$f \sim GP(0, K), \quad \Rightarrow (f(r_1), \cdots, f(r_q)) \sim \mathcal{N}(0, (K(r_i, r_j))_{i,j \leq q})$$

K : drives the smoothness of f.

- $K(r, s) = \min(s, t)$: Brownion – Non statio., non smooth

- **Serie representation** [Karhunen Loeve expansion] : Hilbert Space

$$f = \sum_{i=1}^{\infty} \theta_i \phi_i, \quad (\phi_i)_i = \text{BON } \mathbb{H} \quad \theta_i \overset{ind}{\sim} \mathcal{N}(0, \tau_i^2), \quad \tau_i \downarrow 0$$

• good for curves in $\mathbb{R}$ – not so good for densities , etc.

Examples of priors : Gaussian process priors

- **Gaussian process priors** $(\Theta, \|.\|)$ Banach space (e.g. L_2)
$\theta = f$

$$f \sim GP(0, K), \quad \Rightarrow (f(r_1), \cdots, f(r_q)) \sim \mathcal{N}(0, (K(r_i, r_j))_{i,j \leq q})$$

K : drives the smoothness of f.

  - $K(r, s) = \min(s, t)$: Brownion – Non statio., non smooth
  - $K(r, s) = e^{-a(r-s)^2}$: exponential kernel – statio. , smooth

- **Serie representation** [Karhunen Loeve expansion] : Hilbert Space

$$f = \sum_{i=1}^{\infty} \theta_i \phi_i, \quad (\phi_i)_i = \text{BON } \mathbb{H} \quad \theta_i \overset{ind}{\sim} \mathcal{N}(0, \tau_i^2), \quad \tau_i \downarrow 0$$

• good for curves in $\mathbb{R}$ – not so good for densities , etc.

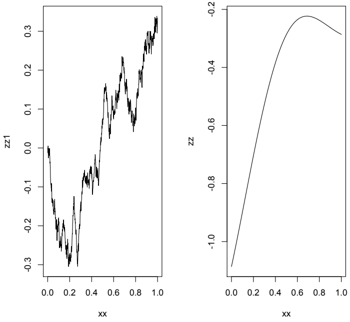


FIG.: Gaussian processes : left : Brownian motion, right : exponential

Other priors on curves in $\mathbb{R}$: hierarchical modelling

- **Splines, basis expansions**

$$f = \sum_{i=1}^{K} \theta_i \phi_i, \quad (\phi_i)_i = \text{Base } \mathbb{H} \quad \theta_i/\tau_i \overset{iid}{\sim} g(.)$$

• Choice of K , of τ_i of g ?

- K random : $K \sim \Pi_K$; then $\tau_i = \tau$ is enough – g flexible

$\longrightarrow$ more flexible - adaptation to the smoothness

Other priors on curves in $\mathbb{R}$: hierarchical modelling

- **Splines, basis expansions**

$$f = \sum_{i=1}^{K} \theta_i \phi_i, \quad (\phi_i)_i = \text{Base } \mathbb{H} \quad \theta_i/\tau_i \stackrel{iid}{\sim} g(.)$$

• Choice of K , of τ_i of g ?

- K random : $K \sim \Pi_K$; then $\tau_i = \tau$ is enough – g flexible
- $\tau_i = \tau(1+i)^{-\alpha-1/2}$, $K = +\infty$, $g = \mathcal{N}$
  either $\tau \sim \pi_\tau$ or $\alpha \sim \pi_\alpha$ or EB

$\longrightarrow$ more flexible - adaptation to the smoothness

Nonparametric mixture models

- **Density modelling**

$$f_{P,\sigma}(x) = \int_\Theta g_{\theta,\sigma}(x) dP(\theta), \quad P = \text{proba}$$

e.g.

$$g_{\theta,\sigma} = \mathcal{N}(.|\theta,\sigma), \quad \text{or } \mathcal{N}(.|\mu,\tau^2), \quad \theta = (\mu,\tau^2)$$

- **Prior** $P \sim \Pi_P$ and $\sigma \sim \pi_\sigma$
- **Examples of** Π_P

• finite mixtures :

$$P = \sum_{j=1}^{K} p_j \delta_{(\theta_j)}, \quad K \sim \Pi_K, \ (p_1, \cdots, p_k)|K = k \sim \pi_{p|k}, \quad \theta_j \overset{iid}{\sim} \pi_\theta$$

• Dirichlet Process and co.

Dirichlet Process : $P \sim DP(M, G)$

- **Sethuraman representation**

$$P = \sum_{i=1}^{\infty} p_j \delta_{(\theta_j)}, \quad \theta_j \overset{iid}{\sim} G,$$

$$p_j = V_j \prod_{i<j} (1 - V_i), \quad V_j \overset{iid}{\sim} Beta(1, M) : \text{ stick breaking}$$

- **Partition property** $\forall (B_1, \cdots, B_k)$ partition

$$(P(B_1), \cdots, P(B_k)) \sim \mathcal{D}(MG(B_1), \cdots, MG(B_k))$$

- **Nice clustering properties** Chinese restaurant process.

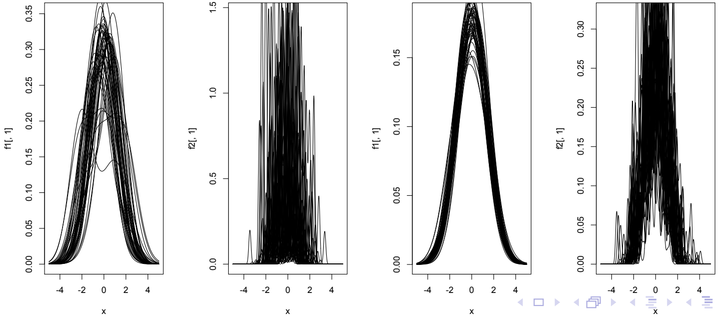
Why mixtures ?

▶ **Mixtures of Gaussians**

$$f_{P,\sigma}(x) = \int_{\mathbb{R}^d} \phi_\sigma(x-\mu)dP(\mu), \quad P = \text{proba}$$

- Analytic
- If f_0 *ordinary smooth*

$$\exists P_\sigma, \ s.t. \quad f_{P_\sigma,\sigma} \to f_0, \quad \sigma \to 0$$



Remarks – towards asymptotic properties

- Can we assess the impact of hyperparameters ?

Remarks – towards asymptotic properties

- Can we assess the impact of hyperparameters ?
- Are some hyperparameters more influencial than others ?

Remarks – towards asymptotic properties

- Can we assess the impact of hyperparameters ?
- Are some hyperparameters more influencial than others ?
- Understand how the prior model acts as an approximation tool for the curve of interest ?

Outline

Posterior consistency and concentration rates

$$X^n = (X_1, ..., X_n) \sim P_\theta, \theta \in \Theta, \quad \theta \sim \Pi$$

- **Consistency** $d(\theta_1, \theta_2)$ = distance (or loss), $\theta_0 \in \Theta$ *the posterior is consistent* at θ_0 iff $\forall \epsilon > 0$ P_{θ_0} a.s. or in proba.

$$\Pi\left[A_\epsilon | X^n\right] = 1 + o(1), \quad A_\epsilon = \{\theta \in \Theta; d(\theta_0, \theta) < \epsilon\}$$

- **Concentration rates** *the posterior concentrates* at the rate at least ϵ_n at θ_0 iff

$$E^n_{\theta_0}\left[\Pi\left[A_{\epsilon_n} | X^n\right]\right] = 1 + o(1), \quad \epsilon_n \downarrow 0$$

• It depends on $d(.,.)$ and on Π and θ_0

Outline

Doob

If Θ and $\mathcal{X}$ are separable, $X^n = (X_1, \cdots, X_n) \overset{iid}{\sim} P_\theta$ the posterior in consistent (a.s.) on a set of probability 1 wrt Π : i.e.

$$\exists \Theta_0 \subset \Theta, \textit{ s.t. } \Pi(\Theta_0) = 1 \textit{ and } \forall \theta_0 \in \Theta_0, \textit{ the posterior is consistent at } \theta_0, P_{\theta_0}^{\infty} \textit{ a.s.}$$

- **Not enough** What is Θ_0 ?

Schwartz and Barron

Under the two types of conditions :

- Kullback-Leibler support : $\forall \epsilon > 0$

  $$\Pi\left[K_\infty(\theta_0,\theta) < \epsilon\right] > 0, \quad K_\infty(\theta_0,\theta) = \limsup_n n^{-1}\left(\ell_n(\theta_0) - \ell_n(\theta)\right)$$

  e.g. iid data

  $$K_\infty(\theta_0,\theta) = K(\theta_0,\theta) = \int f_{\theta_0} \log(f_{\theta_0}/f_\theta) d\mu$$

Then the posterior is consistent at θ_0 a.s.

- **If Kullback-Leibler only** : weak consistency
- **Kullback-Leibler condition** Not necessary but important

Schwartz and Barron

Under the two types of conditions :

- Kullback-Leibler support : $\forall \epsilon > 0$

  $$\Pi\left[K_\infty(\theta_0,\theta) < \epsilon\right] > 0, \quad K_\infty(\theta_0,\theta) = \limsup_n n^{-1}\left(\ell_n(\theta_0) - \ell_n(\theta)\right)$$

  e.g. iid data

  $$K_\infty(\theta_0,\theta) = K(\theta_0,\theta) = \int f_{\theta_0}\log(f_{\theta_0}/f_\theta)d\mu$$

- Existence of tests

  $$\exists \phi_n \in [0,1]; \quad E^n_{\theta_0}[\phi_n] = o(1), \quad \sup_{\theta:d(\theta_0,\theta)>\epsilon} E^n_\theta\left[1-\phi_n\right] = o(1)$$

Then the posterior is consistent at θ_0 a.s.

- **If Kullback-Leibler only** : weak consistency
- **Kullback-Leibler condition** Not necessary but important

Concentration rates : Ghosal, Van der Vaart, Walker and co

$$\Pi\left[d(f_0, f) \leq \epsilon_n | X^n\right] = 1 + o_p(1), \quad \epsilon_n \downarrow 0$$

- **Kullback-Leibler condition :** $\exists c > 0$

$$\Pi\left[\{K_n(\theta_0, \theta) \leq n\epsilon_n^2; V(\theta_0, \theta) \leq n\epsilon_n^2\}\right] \geq e^{-cn\epsilon_n^2}$$

$$K_n(\theta_0, \theta) = E_{\theta_0}\left(\ell_n(\theta_0) - \ell_n(\theta)\right) \quad V(\theta_0, \theta) = E_{\theta_0}^n((\ell_n(\theta_0) - \ell_n(\theta))^2)$$

- **sieve** $\exists \Theta_n \subset \Theta$, s.t. $\Pi(\Theta_n^c) \leq e^{-(c+2)n\epsilon_n^2}$.
- **Tests** $\exists \phi_n$ s.t. if $A_{M\epsilon_n} = \{\theta; d(\theta_0, \theta) \leq M\epsilon_n\}$

$$E_{\theta_0}^n[\phi_n] = o(1), \qquad \sup_{\theta \in A_{M\epsilon_n}^c \cap \Theta_n} E_f^n(1 - \phi_n) \leq e^{-(c+2)n\epsilon_n^2}$$

Proof of Ghosal & VdV.

$$U_n^c = \{d(\theta, \theta_0) > M\epsilon_n\},\ S_n = \{K_n(\theta_0, \theta) \le n\epsilon_n^2;\ V(\theta_0, \theta) \le n\epsilon_n^2$$

$$\begin{aligned}
E_{\theta_0}\left[\Pi\left(U_n^c | X^n\right)\right] &= E_{\theta_0}\left[\frac{\int_{U_n^c} e^{\ell_n(\theta)-\ell_n(\theta_0)} d\pi(\theta)}{\int_{\Theta} e^{\ell_n(\theta)-\ell_n(\theta_0)} d\pi(\theta)}\right] := E_{\theta_0}\left[\frac{N_n}{D_n}\right] \\
&\le E_{\theta_0}[\phi_n] + P_{\theta_0}^n\left[D_n < e^{-2n\epsilon_n^2}\pi(S_n)\right] \\
&\quad + \frac{e^{2n\epsilon_n^2}}{\pi(S_n)} E_{\theta_0}^n[N_n(1-\phi_n)] \\
&\le E_{\theta_0}\left[\phi_n\right] + \frac{\int_{S_n} P_{\theta_0}\left[\ell_n(\theta) - \ell_n(\theta_0) < -2n\epsilon_n^2\right] d\pi(\theta)}{\pi(S_n)} \\
&\quad + \frac{e^{2n\epsilon_n^2}}{\pi(S_n)} \int_{U_n^c \cap \Theta_n} E_\theta\left[1-\phi_n\right] d\pi(\theta) + \frac{e^{2n\epsilon_n^2}}{\pi(S_n)} \Pi(\Theta_n^c)
\end{aligned}$$

Hellinger or L1 distance for iid

- If $A_{\epsilon_n} = \{f, d(f_0, f) \leq \epsilon_n\}$ with
  $d(f_0, f) = |f - f_0|_1 = \int |f - f_0|$ (L1) or
  $d(f_0, f) = h(f_0, f) = \|\sqrt{f} - \sqrt{f_0}\|_2$ (Hell.)
  tests exist under entropy conditions $\rightarrow$ cf exo

Hellinger or L1 distance for iid

- If $A_{\epsilon_n} = \{f, d(f_0, f) \leq \epsilon_n\}$ with
  $d(f_0, f) = |f - f_0|_1 = \int |f - f_0|$ (L1) or
  $d(f_0, f) = h(f_0, f) = \|\sqrt{f} - \sqrt{f_0}\|_2$ (Hell.)
  tests exist under entropy conditions $\rightarrow$ cf exo
- Variants : There are variants of this result but same ideas.

Mixture models for smooth densities

- **Mixture model**

$$\psi_{P,\sigma}(x) = \int \phi(x|\theta,\sigma)dP(\theta)$$

- **Observations** $X_i \sim f_0$, i.i.d $i = 1, ..., n$ and f_0 is smooth,

$$f_0 \notin \{\psi_{P,\sigma}, P \in \mathcal{P}, \sigma \in \mathcal{S}\}$$

- **Can we still use the mixture model ?** sometimes yes.
- **Wu & Ghosal (08)** General conditions for **Kullback-Leiber** (ϵ) to hold.

Outline

location-Gaussian mixtures

$$\psi_p(x|\mu,\sigma) = e^{-\frac{|x-\mu|^2}{2\sigma^2}}\sigma^{-1}, \quad \theta = \mu, \quad \alpha = \sigma,$$

- **General idea :** if f C^o $\lim_{\sigma\to 0}\int \psi(x|\mu,\sigma)f(\mu)d\mu = f(x)$.

Theorem

(Kruijer et al. 2010)

If $\log f_0$ *is locally β-Hölder on $\mathbb{R}$ + other conds. $\exists g_\beta$ density s.t.*

$$K(f_0, \psi_{g_\beta,\sigma}) = O(\sigma^{2\beta}), \quad V(f_0, f_{g_\beta}) = O(\sigma^{2\beta})$$

(de Jong et van Zanten) *If* f_0 *β-Hölder ,*

$$\|f_0 - \psi_{g_\beta,\sigma}\|_\infty = O(\sigma^\beta)$$

$$f_{g_\beta,\sigma}(x) = \int \psi_p((x-\mu)/\sigma)\sigma^{-1}g_\beta(x)dx$$

Outline

Prior on P, σ

$$\psi_{P,\sigma}(x) = \int e^{-\frac{|x-\mu|^2}{2\sigma^2}} \sigma^{-1} dP(\mu), \quad d\Pi(P,\sigma)?$$

- **discrete mixtures**
  - Dirichlet Process $P \sim DP(\alpha, G_0)$

$$P(\mu) = \sum_{j=1}^{\infty} p_j \delta_{(\mu_j)}, \quad \text{Sethuraman}$$

Prior on P, σ

$$\psi_{P,\sigma}(x) = \int e^{-\frac{|x-\mu|^2}{2\sigma^2}} \sigma^{-1} dP(\mu), \quad d\Pi(P,\sigma)?$$

- **discrete mixtures**
  - Dirichlet Process $P \sim DP(\alpha, G_0)$

$$P(\mu) = \sum_{j=1}^{\infty} p_j \delta_{(\mu_j)}, \quad \text{Sethuraman}$$

  - Finite mixtures

$$P(\mu) = \sum_{i=1}^{k} p_j \delta_{(\mu_j)}, \quad d\Pi(P) = p(k)\pi_k(p_1, ..., p_k)\pi(\mu_1)....\pi(\mu_k)$$

$$p(k) \approx e^{-k(\log k)^r}, \quad \pi(\mu) \approx e^{-c|\mu|^a}, a > 0$$

Prior on P, σ

$$\psi_{P,\sigma}(x) = \int e^{-\frac{|x-\mu|^2}{2\sigma^2}} \sigma^{-1} dP(\mu), \quad d\Pi(P,\sigma)?$$

- **discrete mixtures**
  - Dirichlet Process $P \sim DP(\alpha, G_0)$

$$P(\mu) = \sum_{j=1}^{\infty} p_j \delta_{(\mu_j)}, \quad \text{Sethuraman}$$

  - Finite mixtures

$$P(\mu) = \sum_{i=1}^{k} p_j \delta_{(\mu_j)}, \quad d\Pi(P) = p(k)\pi_k(p_1, ..., p_k)\pi(\mu_1)....\pi(\mu_k)$$

$$p(k) \approx e^{-k(\log k)^r}, \quad \pi(\mu) \approx e^{-c|\mu|^a}, a > 0$$

  - Prior on σ

$$\sigma \sim IG(a_1, a_2)$$

Result : adaptive concentration rate

Theorem

If $\log f_0$ *is locally β-Hölder + conds , then*

$$P^{\pi}\left[d(f,f_0) \leq Cn^{-\beta/(2\beta+1)}(\log n)^t | X^n\right] = 1 + o_p(1)$$

- **Adaptive minimax** :

Result : adaptive concentration rate

Theorem

If $\log f_0$ *is locally* β*-Hölder + conds , then*

$$P^{\pi}\left[d(f, f_0) \leq C n^{-\beta/(2\beta+1)} (\log n)^t | X^n\right] = 1 + o_p(1)$$

- **Adaptive minimax** :
  - *Minimax rate* :

$$\inf_{\hat{f}} \sup_{f \in \mathcal{C}} r_n^{-1} E_f^n \left[d(\hat{f}, f)\right] \approx cste$$

If $\mathcal{C}$: β-Hölder densities and $d = L_1$ then $r_n = n^{-\beta/(2\beta+1)}$

Result : adaptive concentration rate

Theorem

If $\log f_0$ *is locally β-Hölder + conds , then*

$$P^{\pi}\left[d(f,f_0) \leq Cn^{-\beta/(2\beta+1)}(\log n)^t | X^n\right] = 1 + o_p(1)$$

- **Adaptive minimax** :
  - *Minimax rate* :

$$\inf_{\hat{f}} \sup_{f\in\mathcal{C}} r_n^{-1} E_f^n\left[d(\hat{f},f)\right] \approx cste$$

    If $\mathcal{C}$: β-Hölder densities and $d = L_1$ then $r_n = n^{-\beta/(2\beta+1)}$
  - *Adaptive* : $\hat{f}$ does not depend on β but still attains r_n (or nearly) Here the prior does not depend on β

Links with Kernel estimation

- **Kernel estimation**

$$\hat{f}(x) = \psi_{P_n,\sigma} = \int \psi(x|\mu,\sigma)dP_n(\mu), \quad P_n(\mu) = \frac{1}{n}\sum_{i=1}^{n} \delta_{(X_i)}(\mu)$$

The best you can do : $\|f - \psi_{f,\sigma}\| \rightarrow$ not adpative.
e.g. Gaussian mixtures f is β-Hölder.

$$\|f - \psi_{f,\sigma}\|_\infty = O(\sigma^{2\wedge\beta}), \quad \text{suboptimal if } \beta > 2$$

- **Here** It is enough to find f_β ($\neq f$)

$$\|f - \psi_{f_\beta,\sigma}\|_\infty = O(\sigma^\beta)$$

Some open questions

- **Location - scale mixtures**

$$f_P = \int \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) dP(\mu,\sigma)$$

$\rightarrow$ Consistency to any continuous and positive density
$\rightarrow$ Suboptimal rates for $\beta-$ Hölder : Sharp ? Why ? contradicts practice

- **Lower bounds ?**

Empirical Bayes : data dependent prior

- **Setup** prior model $\Pi(d\theta|\lambda)$, $\lambda \in \Lambda$

e.g. $\theta \in \mathbb{R}$, $\Pi(d\theta|\lambda) \equiv \mathcal{N}(\mu_0, \tau_0^2)$ & $\lambda = (\mu_0, \tau_0^2)$.

- **How to select λ ?**
  - Prior information : informative prior

Empirical Bayes : data dependent prior

- **Setup** prior model $\Pi(d\theta|\lambda)$, $\lambda \in \Lambda$

e.g. $\theta \in \mathbb{R}$, $\Pi(d\theta|\lambda) \equiv \mathcal{N}(\mu_0, \tau_0^2)$ & $\lambda = (\mu_0, \tau_0^2)$.

- **How to select λ ?**
  - Prior information : informative prior
  - Hierarchical $\lambda \sim Q$: Hierarchical Bayes. But Q ?

Empirical Bayes : data dependent prior

- **Setup** prior model $\Pi(d\theta|\lambda)$, $\lambda \in \Lambda$
e.g. $\theta \in \mathbb{R}$, $\Pi(d\theta|\lambda) \equiv \mathcal{N}(\mu_0, \tau_0^2)$ & $\lambda = (\mu_0, \tau_0^2)$.
- **How to select λ ?**
  - Prior information : informative prior
  - Hierarchical $\lambda \sim Q$: Hierarchical Bayes. But Q ?
  - use data : $\hat{\lambda}(X^n)$: empirical Bayes : double use of the data

Examples of ways of choosing $\hat{\lambda}$ and examples

- **Maximum marginal likelihood estimate**

$$\hat{\lambda}_n = \text{argmax}_\lambda m(X^n|\lambda), \quad m(X^n|\lambda) = \int_\Theta f_\theta^n(X^n) d\Pi(\theta|\lambda)$$

- **Others** Moment - types estimate

$$X_1, \ldots, X_n|(F, \sigma) \overset{\text{i.i.d.}}{\sim} p_{F,\sigma}(\cdot) := \int \phi(\cdot|\mu, \sigma^2)\, \mathrm{d}F(\mu).$$

$$\theta = (F, \sigma), \quad \text{Prior} : F \sim DP(\alpha \mathcal{N}(\lambda, \tau^2)), \quad \sigma \sim \pi_\sigma$$

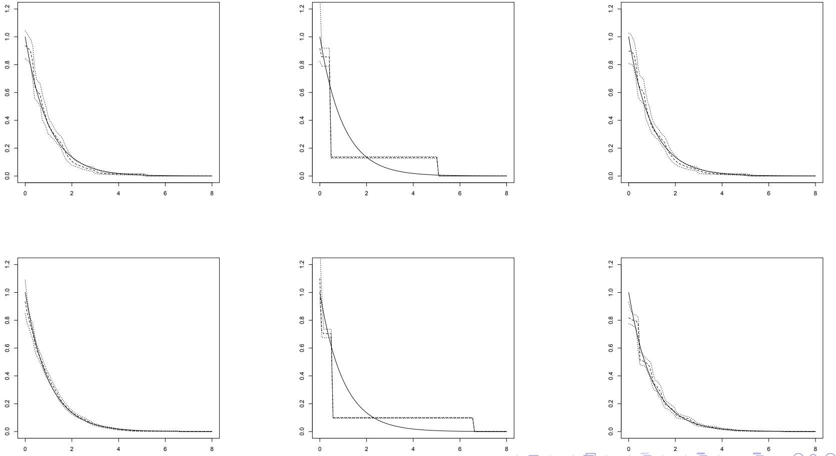
$$\hat{\lambda}_n = \bar{X}_n, \quad \hat{\tau}_n^2 = S_n^2, \max X_i - \min X_i$$

see e.g. Green & Richardson

Outline

Driving example : Poisson inhomogeneous monotone intensity estimation

Strategy 1 (Empirical) Strategy 2 (γ fixed) Strategy 3 (hierarchica



Dealing with data dependent priors

- Theory : so far fully Bayes
- How to adapt to data dependent priors ? ▶ **Ghosal and Van der Vaart 's proof** : Fubini

$$E_{\theta_0}\left[\Pi\left(U_n^c|X^n\right)\right] = E_{\theta_0}\left[\frac{\int_{U_n^c} e^{\ell_n(\theta)-\ell_n(\theta_0)}d\pi(\theta)}{\int_\Theta e^{\ell_n(\theta)-\ell_n(\theta_0)}d\pi(\theta)}\right] := E_{\theta_0}\left[\frac{N_n}{D_n}\right]$$

$$\leq E_{\theta_0}[\phi_n] + P_{\theta_0}^n\left[D_n < e^{-2n\epsilon_n^2}\pi(S_n)\right]$$

$$+ \frac{e^{2n\epsilon_n^2}}{\pi(S_n)} E_{\theta_0}^n[N_n(1-\phi_n)]$$

$$\leq E_{\theta_0}[\phi_n] + \frac{\int_{S_n} P_{\theta_0}\left[\ell_n(\theta)-\ell_n(\theta_0) < -2n\epsilon_n^2\right]d\pi(\theta)}{\pi(S_n)}$$

$$+ \frac{e^{2n\epsilon_n^2}}{\pi(S_n)}\int_{U_n^c\cap\Theta_n} E_\theta\left[1-\phi_n\right]d\pi(\theta) + \frac{e^{2n\epsilon_n^2}}{\pi(S_n)}\Pi(\Theta_n^c)$$

Difficulty for $\pi\left(U_n^c|X^n;\hat\lambda\right) = o_p(1)$

- **If** $P_{\theta_0}\left[\hat\lambda_n \in \mathcal{K}_n\right] = 1 + o(1)$

$$\pi\left(U_n^c|X^n;\hat\lambda\right) \le \sup_{\lambda\in\mathcal{K}_n} \pi\left(U_n^c|X^n;\lambda\right) = o_p(1)?, \quad U_n = \{\theta, d(\theta_0,\theta) \le \epsilon_n\}$$

- **Non dominated models** $\lambda \to \Pi(d\theta|\lambda)$: not dominated $\Rightarrow$ cannot study

$$\frac{\pi(\theta|\lambda)}{\pi(\theta|\lambda')}$$

Outline

Change of measure - $\sup_{\lambda \in \mathcal{K}_n} \pi(B_n | X^n; \lambda) = o_p(1)$?

- **A key tool** For all λ, λ'

$$\theta \sim \pi(\cdot|\lambda) \Rightarrow \psi_{\lambda,\lambda'}(\theta) \sim \pi(\cdot|\lambda')$$

- **Important class of examples** Mixtures (parametric or NP)

$\theta = (P, \phi)$

$$f_{P,\phi}(x) = \int K_\phi(x|z)dP(z) = \sum_j p_j K_\phi(x|z_j), \; P \sim DP(MG(\cdot|\lambda)), \; \phi \sim \pi_\phi$$

$$\psi_{\lambda,\lambda'}(f_{P,\phi})(x) = \sum_{j=1}^{\infty} p_j K_\phi(x|G^{-1}(G(z_j|\lambda)|\lambda')) \\ = f_{P',\phi}, \quad P' \sim DP(M, G(\cdot|\lambda'))$$

A general Theorem

$$\sup_{\lambda' \in \mathcal{K}_n} \pi(U_n^c | X^n \lambda') = \sup_{\lambda' \in \mathcal{K}_n} \frac{\int_{U_n^c} p^{(n)}_{\psi_{\lambda,\lambda'}(\theta)}(x^n) d\pi(\theta|\lambda)}{\int_{\Theta} p^{(n)}_{\psi_{\lambda,\lambda'}(\theta)}(x^n) d\pi(\theta|\lambda)} := \frac{N_n}{D_n} = o(1)$$

- **KL support condition** : $\mathcal{K}_n = \cup_{i=1}^{N_n(u_n)} B(\lambda_i, u_n)$

$$\sup_{\lambda \in \mathcal{K}_n} \sup_{\theta \in \tilde{B}_n} P^{(n)}_{\theta_0} \left\{ \inf_{\|\lambda' - \lambda\| \le u_n} \ell_n(\psi_{\lambda,\lambda'}(\theta)) - \ell_n(\theta_0) < -n\epsilon_n^2 \right\} = o(N_n(u_n)^{-1})$$

- **tests** : Let $dQ^{\theta}_{\lambda,n}(x) = \sup_{\|\lambda' - \lambda\| \le u_n} p^{(n)}_{\psi_{\lambda,\lambda'}(\theta)}(x) d\mu(x)$,

$$E^{(n)}_{\theta_0}(\phi_n) = o(1), \quad \sup_{\lambda \in \mathcal{K}_n} \sup_{d(\theta,\theta_0) > \epsilon_n} \int_{\mathcal{X}^n} (1 - \phi_n) dQ^{\theta}_{\lambda,n}(x^n) \le e^{-Kn\epsilon_n^2}$$

$$\log N_n(u_n) = o(n\epsilon_n^2)$$

Example i.i.d

- **Typically** For all $\theta \in \Theta_n$

$$\sup_{|\gamma-\gamma'|\leq u_n} |\ell_n(\psi_{\gamma,\gamma'}(\theta)) - \ell_n(\theta)| \leq u_n \sum_i h_{n,\gamma}(X_i)$$

and

$$P_0\left(h_{n,\gamma}(X) > n^H\right) = o(1/n)$$

Then replace $\mathcal{X}$ by $\mathcal{X} \cap \{h_{n,\gamma}(X) \leq n^H\}$ and $u_n \leq n^{-H-1}$

- Θ_n^c
  - Non data dependent priors : $\pi(\Theta_n^c) \leq e^{-cn\epsilon_n^2}$

Example i.i.d

- **Typically** For all $\theta \in \Theta_n$

$$\sup_{|\gamma-\gamma'|\leq u_n} |\ell_n(\psi_{\gamma,\gamma'}(\theta)) - \ell_n(\theta)| \leq u_n \sum_i h_{n,\gamma}(X_i)$$

and

$$P_0\left(h_{n,\gamma}(X) > n^H\right) = o(1/n)$$

Then replace $\mathcal{X}$ by $\mathcal{X} \cap \{h_{n,\gamma}(X) \leq n^H\}$ and $u_n \leq n^{-H-1}$

- Θ_n^c
  - Non data dependent priors : $\pi(\Theta_n^c) \leq e^{-cn\epsilon_n^2}$
  - Data dependent priors

$$\int_{\Theta_n^c} Q_{\gamma,n}^{\theta}(\mathcal{X}^n)\pi(d\theta|\gamma) \leq e^{-cn\epsilon_n^2}$$

A general Theorem : comments

$$\pi\left(d(\theta,\theta_0) \leq \epsilon_n | x^n, \hat{\lambda}_n\right) = o_{p_0}(1)$$

- If $\mathcal{K}_n = \{\lambda; \epsilon_n(\lambda) \leq M_n \epsilon_n^*\}$, then

$$\epsilon_n \leq M_n \epsilon_n^*$$

$$\Downarrow$$

Oracle posterior concentration rates

- BUT : need to know $\mathcal{K}_n$ e.g. MMLE

Application to DP mixtures of Gaussians

- **Model** $x^n = (x_1, \cdots, x_n)$ iid f
- **prior on** f **: DPM Gaussian**

$$f_{P,\sigma}(x) = \int_{\mathbb{R}} \phi_\sigma(x-\mu)dP(\mu), \quad P \sim DP(A\mathcal{N}(\mu_0, \tau^2)), \quad \sigma \sim \pi_\sigma$$

- **Choice for** μ_0, τ^2 **?** $\lambda = (\mu_0, \tau^2)$ Two cases :

$$\hat{\mu}_0 = \bar{x}_n, \quad \hat{\tau} = s_n, \quad \text{or } \hat{\mu}_0 = \bar{x}_n, \quad \hat{\tau} = \max_i x_i - \min_i x_i$$

- **Change of measure**

$$\psi_{\lambda,\lambda'}(f_P)(x) = \sum_{j=1}^{\infty} p_j \phi_\sigma(x - \mu_j + \Delta_j), \quad \Delta = \mu_j \left(\frac{\tau'}{\tau} - 1\right) - \mu_0 \tau' + \mu_0'$$

Then
$\psi_{\lambda,\lambda'}(f_P) \sim DPM(A\mathcal{N}(\mu_0', \tau'))$, when $P \sim DP(A\mathcal{N}(\mu_0, \tau))$

Results for DP mixtures of Gaussians

$$f_{P,\sigma}(x) = \int_{\mathbb{R}} \phi_\sigma(x-\mu)dP(\mu), \quad P \sim DP(A\mathcal{N}(\mu_0,\tau^2)), \quad \sigma \sim \pi_\sigma$$

Theorem

Under same conditions as in fully Bayes $\exists a > 0$ *such that if* $\mathcal{K}_n \subset [a_1, a_2] \times [\tau_1, (\log n)^q]$, *if* $f_0 \in \mathcal{H}_{\text{loc}}(\alpha)$

$$\pi\left(\|f_{P,\sigma} - f_0\|_1 > (\log n)^a n^{-\alpha/(2\alpha+1)} | \mathbf{x}^n\right) = o_{p_0}(1)$$

• Applies to $\hat{\lambda}_n = (\bar{x}_n, s_n)$ and $(\bar{x}_n, \max_i x_i - \min_i x_i)$: in the latter loss in $\log n$

• $(\bar{x}_n, \max_i x_i - \min_i x_i)$: acts like a non informative prior

Some examples of transformations

- Gaussian processes

$$\sum_j \theta_j \phi_j, \quad \theta_j \overset{ind}{\sim} \mathcal{N}(0, \tau_j^2), \tau_j = \tau j^{-\alpha - 1/2}$$

  - $\lambda = \tau$

$$\psi(\theta_j) = \frac{\tau'}{\tau} \theta_j$$

- Splines : $\sum_{j=1}^K \theta_j B_j, \quad \theta_j \overset{iid}{\sim} \tau g$

Some examples of transformations

- Gaussian processes

$$\sum_j \theta_j \phi_j, \quad \theta_j \overset{ind}{\sim} \mathcal{N}(0, \tau_j^2), \tau_j = \tau j^{-\alpha-1/2}$$

  - $\lambda = \tau$

$$\psi(\theta_j) = \frac{\tau'}{\tau}\theta_j$$

  - $\lambda = \alpha$

$$\psi(\theta_j) = j^{\alpha-\alpha'}\theta_j$$

- Splines : $\sum_{j=1}^{K} \theta_j B_j, \quad \theta_j \overset{iid}{\sim} \tau g$

Some examples of transformations

- Gaussian processes

$$\sum_j \theta_j \phi_j, \quad \theta_j \overset{ind}{\sim} \mathcal{N}(0, \tau_j^2), \tau_j = \tau j^{-\alpha-1/2}$$

  - $\lambda = \tau$

$$\psi(\theta_j) = \frac{\tau'}{\tau}\theta_j$$

  - $\lambda = \alpha$

$$\psi(\theta_j) = j^{\alpha-\alpha'}\theta_j$$

- Splines : $\sum_{j=1}^{K} \theta_j B_j, \quad \theta_j \overset{iid}{\sim} \tau g$
  - K : no need for transformation

Some examples of transformations

- Gaussian processes

$$\sum_j \theta_j \phi_j, \quad \theta_j \overset{ind}{\sim} \mathcal{N}(0, \tau_j^2), \tau_j = \tau j^{-\alpha - 1/2}$$

  - $\lambda = \tau$

$$\psi(\theta_j) = \frac{\tau'}{\tau} \theta_j$$

  - $\lambda = \alpha$

$$\psi(\theta_j) = j^{\alpha - \alpha'} \theta_j$$

- Splines : $\sum_{j=1}^K \theta_j B_j, \quad \theta_j \overset{iid}{\sim} \tau g$
  - K : no need for transformation
  - $\tau \rightarrow$

$$\psi_{\tau,\tau'}(\theta_j) = \frac{\tau'}{\tau} \theta_j$$

Partial conclusion on posterior concentration rates

- Generic tools to obtain ϵ_n

$$\pi(\{d(\theta_0, \theta) \leq \epsilon_n\} | X^n) \to 1$$

using

$$\psi_{\gamma,\gamma'} : \Theta \to \Theta$$

Partial conclusion on posterior concentration rates

- Generic tools to obtain ϵ_n

$$\pi(\{d(\theta_0, \theta) \leq \epsilon_n\}|X^n) \rightarrow 1$$

  using

$$\psi_{\gamma,\gamma'} : \Theta \rightarrow \Theta$$

- Enough prior mass on KL nighbourhoods of θ_0 + tests

Partial conclusion on posterior concentration rates

- Generic tools to obtain ϵ_n

$$\pi(\{d(\theta_0, \theta) \leq \epsilon_n\}|X^n) \to 1$$

  using

$$\psi_{\gamma,\gamma'} : \Theta \to \Theta$$

- Enough prior mass on KL nighbourhoods of θ_0 + tests
- extension to data dependent priors – even for MMLE

Outline

Semi-parametric Bayesian methods : setup

- **Infinite dimensional :** $\dim(\Theta) = +\infty$
- **Parameter of interest :** $\Psi(\theta) \subset \mathbb{R}^d$
- **Examples :**
  - $\theta = (\psi, \eta)$, $\psi \in \mathbb{R}^d$, $\dim(\eta) = +\infty$: ex. Cox model ; partial linear regression, semi - parametric HMMs, mixtures

$$\Psi(\theta) = \psi$$

Semi-parametric Bayesian methods : setup

- **Infinite dimensional :** $\dim(\Theta) = +\infty$
- **Parameter of interest :** $\Psi(\theta) \subset \mathbb{R}^d$
- **Examples :**
  - $\theta = (\psi, \eta)$, $\psi \in \mathbb{R}^d$, $\dim(\eta) = +\infty$: ex. Cox model ; partial linear regression, semi - parametric HMMs, mixtures

$$\Psi(\theta) = \psi$$

  - $\theta =$ curve f, (density, regression, spectral density)

$$\Psi(\theta) = \psi(f), \quad \text{functional}$$

ex : $\psi(f) = F(x) = \int \mathbb{1}_{u \le x} f(t)dt$, $\psi(f) = \int f^2(u)du$, $\psi(f) = f(x_0)$

Marginal posterior

$$\Pi(\psi(\theta) \in A_n | X^n)??$$

- **Regular models**

$$\exists \hat{\psi}, \text{ s.t. } \sqrt{n}(\hat{\psi} - \psi(\theta_0)) \to \mathcal{N}(0, v_0)$$

What about Bayesian approaches ?

$$\Pi(d(\psi, \psi(\theta_0)) \leq M_n n^{-1/2} | X^n) \to 1, \quad \forall M_n \uparrow +\infty?$$

More ? : asymptotic normality : BvM

$$\Pi(\sqrt{n}(\psi - \hat{\psi}) \in A | X^n) \to \mathbb{P}(\mathcal{N}(0, v_0) \in A)?$$

Outline

Bernstein Von Mises : i.i.d parametric

• Observations : for $i = 1, ..., n$ $X_i :\sim f(|\theta)$, i.i.d $\theta \in \Theta$.
A priori : $d\Pi(\theta) = \pi(\theta)d\theta =$ prior distribution
$\longrightarrow$ posterior density

$$\pi(\theta|X^n) = \frac{\pi(\theta)f(X^n|\theta)}{m(X^n)}, \quad X^n = (X_1, ..., X_n)$$

▶ **Bernstein Von Mises :**
When n goes to infinity, the posterior distribution of θ close to a Normal with mean $\hat{\theta}$ and variance $V_{\theta_0}(\hat{\theta})$ under P_{θ_0}.

$$\sqrt{n}(\theta - \hat{\theta}) \approx \mathcal{N}(0, V_{\theta_0}(\hat{\theta}))$$

• regular models : $\hat{\theta} =$ MLE, $V_{\theta_0}(\hat{\theta}) = I(\theta_0)^{-1} =$ Inv. Fisher information Matrix

illustration :

$X_i \sim P(\lambda)$, and $\pi(\lambda) = \Gamma(a, b)$ then

$$\pi(\lambda|X^n) = \Gamma(a+n\bar{X}_n, b+n), \quad a = 10, b = 1, \quad \lambda_0 = 1, \quad n = 1, 10, 100$$

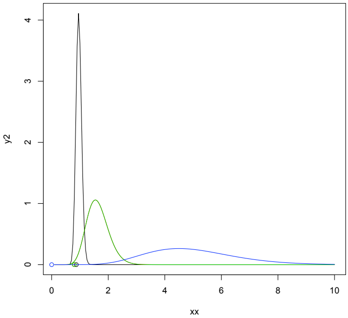


FIG.: posterior, n=1 = blue, n=10=green, n=100=black.

Outline

Applications of BVM

1. Construction of HPD regions

$$C_\alpha^\pi = \{\theta; \pi(\theta|X^n) \geq k_\alpha\}; \quad P^\pi\left[C_\alpha^\pi|X^n\right] = 1-\alpha$$

Then

$$C_\alpha^\pi \approx \{\theta; (\theta-\hat{\theta})^t J_n(\theta-\hat{\theta}) \leq \mathcal{X}_d^{-1}(1-\alpha)\}$$

close to the highest likelihood frequentist confidence region.

Applications of BVM

1. Construction of HPD regions

$$C_\alpha^\pi = \{\theta; \pi(\theta|X^n) \geq k_\alpha\}; \quad P^\pi\left[C_\alpha^\pi|X^n\right] = 1-\alpha$$

Then

$$C_\alpha^\pi \approx \{\theta; (\theta-\hat{\theta})^t J_n(\theta-\hat{\theta}) \leq \mathcal{X}_d^{-1}(1-\alpha)\}$$

close to the highest likelihood frequentist confidence region.

2. α credible regions C_α^π for θ are asymptotically α-confidence regions

$$P_\theta[\theta \in C_\alpha^\pi] = \alpha + o(1)$$

Applications of BVM

1. Construction of HPD regions

$$C_\alpha^\pi = \{\theta; \pi(\theta|X^n) \geq k_\alpha\}; \quad P^\pi\left[C_\alpha^\pi|X^n\right] = 1 - \alpha$$

   Then

$$C_\alpha^\pi \approx \{\theta; (\theta - \hat{\theta})^t J_n(\theta - \hat{\theta}) \leq \mathcal{X}_d^{-1}(1 - \alpha)\}$$

   close to the highest likelihood frequentist confidence region.

2. α credible regions C_α^π for θ are asymptotically α-confidence regions

$$P_\theta[\theta \in C_\alpha^\pi] = \alpha + o(1)$$

3. Approximation of estimators

Outline

Types of conditions required

Theorem

Then

$$\sqrt{n}(\theta - \hat{\theta}) \approx \mathcal{N}(0, I(\theta_0)^{-1})$$

- **Extensions to**

• Non regular models (sometimes)

• Non iid

Types of conditions required

Theorem

1. *If* $\Theta \subset \mathbb{R}^d$

Then

$$\sqrt{n}(\theta - \hat{\theta}) \approx \mathcal{N}(0, I(\theta_0)^{-1})$$

▶ **Extensions to**

- Non regular models (sometimes)
- Non iid

Types of conditions required

Theorem

1. *If* $\Theta \subset \mathbb{R}^d$
2. *If* $f(.|\theta)$ *regular (Positive Fisher, LAN)*

Then

$$\sqrt{n}(\theta - \hat{\theta}) \approx \mathcal{N}(0, I(\theta_0)^{-1})$$

▶ **Extensions to**

• Non regular models (sometimes)

• Non iid

Types of conditions required

Theorem

1. *If* $\Theta \subset \mathbb{R}^d$
2. *If* $f(.|\theta)$ *regular (Positive Fisher, LAN)*
3. *If* $\forall \epsilon > 0, \exists \delta > 0$ *s.t.*

$$\lim_{M\to\infty} limsup_n P^\pi \left[|\theta - \theta_0| > Mn^{-1/2} | X^n \right] = o_p(1)$$

Then

$$\sqrt{n}(\theta - \hat{\theta}) \approx \mathcal{N}(0, I(\theta_0)^{-1})$$

- **Extensions to**

• Non regular models (sometimes)

• Non iid

Types of conditions required

Theorem

1. *If* $\Theta \subset \mathbb{R}^d$
2. *If* $f(.|\theta)$ *regular (Positive Fisher, LAN)*
3. *If* $\forall \epsilon > 0$, $\exists \delta > 0$ *s.t.*

$$\lim_{M \to \infty} limsup_n P^\pi \left[|\theta - \theta_0| > M n^{-1/2} | X^n \right] = o_p(1)$$

4. $\pi(\theta_0) > 0$ *and* C^o *at* θ_0

Then

$$\sqrt{n}(\theta - \hat{\theta}) \approx \mathcal{N}(0, I(\theta_0)^{-1})$$

- **Extensions to**

• Non regular models (sometimes)

• Non iid

Why does it work ?

- **Taylor expansion** of log-likelihood : $l_n(\theta)$ around $\hat{\theta}$ (LAN)

$$l_n(\theta) = \log f(X^n|\theta), \quad \hat{\theta} = \text{ post mean or normalized score}$$

$$\begin{aligned} \pi(\theta|X^n) &\propto e^{l_n(\theta) - l_n(\hat{\theta}) + \log(\pi(\theta)) - \log(\pi(\hat{\theta}))} \\ &\propto e^{-\frac{(\theta-\hat{\theta})J_n(\theta-\hat{\theta})}{2}(1+o_P(1))} \quad \text{when } |\theta - \hat{\theta}| = o_P(1) \end{aligned}$$

$$J_n = D^2 l_n(\theta)|_{\theta=\hat{\theta}}$$

- **Integrate the approximation**

Extension to nonparametric models

- **Control of the LAN rest** uniformly compared $n\|\theta - \theta_0\|_2^2$
- **Continuity of the prior density**

• Spokoiny 2014 for increasing dimensions

• Castillo & Nickl, 2014 for weaker versions (weaker topologies)

Outline

General BVM theorem : framework

- **Model :** $X^n|\theta \sim f_\theta^n$ where $\theta \in \Theta$ infinite dimensional
  π : prior on θ
- **Parameter of interest :** $\psi(\theta)$
- **Aim :** Asymptotic posterior distribution of $\psi(\theta)$:
  - Normality ?

General BVM theorem : framework

- **Model :** $X^n|\theta \sim f_\theta^n$ where $\theta \in \Theta$ infinite dimensional
  π : prior on θ
- **Parameter of interest :** $\psi(\theta)$
- **Aim :** Asymptotic posterior distribution of $\psi(\theta)$:
  - Normality ?
  - Centering ? Variance ?

General BVM theorem : framework

- **Model :** $X^n|\theta \sim f_\theta^n$ where $\theta \in \Theta$ infinite dimensional
  π : prior on θ
- **Parameter of interest :** $\psi(\theta)$
- **Aim :** Asymptotic posterior distribution of $\psi(\theta)$:
  - Normality ?
  - Centering ? Variance ?
  - ex : Linear functional . $\theta = f$ and $\psi(f) = \int \psi(u) f(u) du$
    But not only

Outline

1. Model = LAN. under $f_0^n = f_{\theta_0}^n$ (truth)

$$\log f_\theta^n(X^n) - \log f_0^n(X^n) = -\frac{n\|\theta - \theta_0\|_L^2}{2} + \sqrt{n}W_n(\theta - \theta_0) + R_n(\theta, \theta_0)$$

with $W_n(u) \sim \mathcal{N}(0, \|u\|_L^2)$ and $u \to W_n(u)$ linear.
• White noise $dX(t) = f(t)dt + dW(t)/\sqrt{n}$ ($\Leftrightarrow X_i = \theta_i + n^{-1/2}\epsilon_i$, $i \in \mathbb{N}$)

$$\ell_n(\theta) - \ell_n(\theta_0) = \frac{-n\|\theta - \theta_0\|_2^2}{2} + \sqrt{n}\sum_i (\theta_i - \theta_{0i})\epsilon_i$$

$$\|\theta - \theta_0\|_L^2 = \sum_{i=1}^{\infty} (\theta_i - \theta_{0i})^2$$

LAN condition, Ex 2

- Density $X_i \sim f$ i.i.d $\theta = \log f$

$$\ell_n(\theta) - \ell_n(\theta_0) = \sum_i \theta(X_i) - \theta_0(X_i) = -\frac{n\|\theta - \theta_0\|_L^2}{2} + \sqrt{n}\mathbb{G}_n(\theta - \theta_0) + R_n(\theta)$$

$$\|\theta - \theta_0\|_L^2 = \int f_0(x)\left(\log f(x) - \log f_0(x)\right)^2 dx - \left(\int f_0(\log f - \log f_0)\right)^2$$

- auto- regression $Y_i = f(Y_{i-1}) + \epsilon_i$, $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$

$$\|\theta - \theta_0\|_L^2 = \int_{\mathbb{R}} q_{f_0}(x)(f(x) - f_0(x))^2 dx$$

context again

2. Concentration : $\exists A_n \subset \Theta$

$$P^{\pi}\left[A_n | X^n\right] = 1 + o_p(1)$$

typically

$$A_n \subset \{d(\theta_0, \theta) \leq \epsilon_n\}, \quad \epsilon_n \downarrow 0$$

3.Smoothness of ψ

$$\psi(\theta) = \psi(\theta_0) + < \theta - \theta_0, \dot{\psi}_0 >_L + < \theta - \theta_0, \ddot{\psi}_0(\theta - \theta_0) >_L + r(\theta, \theta_0)$$

when $\|\theta - \theta_0\|_L \leq \epsilon_n$.
2 regimes

- Linear : $\ddot{\psi}_0 = 0$

context again

2. Concentration : $\exists A_n \subset \Theta$

$$P^\pi \left[A_n | X^n\right] = 1 + o_p(1)$$

typically

$$A_n \subset \{d(\theta_0, \theta) \leq \epsilon_n\}, \quad \epsilon_n \downarrow 0$$

3. Smoothness of ψ

$$\psi(\theta) = \psi(\theta_0) + < \theta - \theta_0, \dot{\psi}_0 >_L + < \theta - \theta_0, \ddot{\psi}_0(\theta - \theta_0) >_L + r(\theta, \theta_0)$$

when $\|\theta - \theta_0\|_L \leq \epsilon_n$.

2 regimes

- Linear : $\ddot{\psi}_0 = 0$
- quadratic $\ddot{\psi}_0 \neq 0$

About the 2 regimes : examples

- Linear functional : $\theta = f$

$$\psi(f) = \int \psi(x) f(x) dx = \psi(f_0) + \int \psi(f - f_0)$$

- Quadratic

$$\psi(f) = \int f^2(x) dx = \psi(f_0) + 2 < f_0, f - f_0 >_2 + \|f - f_0\|_2^2, \quad \dot{\psi}_0 = 2f_0$$

If on A_n :

$$\|f - f_0\|_2^2 \leq \epsilon_n^2 = o(1/\sqrt{n})$$

then

$$\psi(f) = \int f^2(x) dx = \psi(f_0) + 2 < f_0, f - f_0 >_2 + o(1/\sqrt{n}), \quad \ddot{\psi}_0 = 0$$

else $\ddot{\psi}_0 h = 2h$

Outline

Theorem

Set

$$\theta_t = \theta - t\frac{\dot\psi_0}{\sqrt{n}} - \frac{t\ddot\psi_0(\theta-\theta_0)}{2\sqrt{n}} + \frac{t\Delta}{n}, \quad t \neq 0$$

If on A_n, $R(\theta,\theta_0) - R(\theta_t,\theta_0) + t\sqrt{n}r(\theta,\theta_0) = o(1)$ and

- **The condition**

$$\frac{\int_{A_n} p_{\theta_t}(Y^n)d\pi(\theta)}{\int_{A_n} p_{\theta}(Y^n)d\pi(\theta)} = 1 + o_p(1)$$

Then a posteriori :

$$\sqrt{n}(\psi(\theta) - \hat\psi) \approx \mathcal{N}(0, V_{0,n}), \quad \hat\psi = \psi(\theta_0) + \frac{W_n(\dot\psi_0)}{\sqrt{n}} - \frac{W_n(\Delta)}{n}$$

$$V_{0,n} = \|\dot\psi_0 - \frac{\Delta}{\sqrt{n}}\|_L^2$$

General idea

- Prove & find $\hat{\psi}$ s. t. ($A_n = \{\|\theta - \theta_0\|_L \leq \epsilon_n\}$)

$$E^\pi \left[e^{t\sqrt{n}(\psi(\theta)-\hat{\psi})} \mathbb{I}_{A_n}(f)|X^n \right] = e^{t^2 V^2/2} + o_P(1),$$

$$E^\pi \left[e^{t\sqrt{n}(\psi(\theta)-\hat{\psi})} \mathbb{I}_{A_n}(f)|X^n \right] \approx e^{t\sqrt{n}(\psi(\theta_0)-\hat{\psi})} \times$$

$$\frac{\int_{A_n} e^{-n\frac{\|\theta-\theta_0\|^2}{2}+\sqrt{n}W_n(\theta-\theta_0)+R(\theta,\theta_0)+t\sqrt{n}<\theta-\theta_0,\dot{\psi}_0>_L+t\sqrt{n}\frac{<\theta-\theta_0,\ddot{\psi}_0(\theta-\theta_0)>}{2}}}{\int_{A_n} e^{-n\frac{\|\theta-\theta_0\|^2}{2}+\sqrt{n}W_n(\theta-\theta_0)+R_n(\theta)} d\pi(\theta)}$$

$$\approx \quad e^{t\sqrt{n}(\psi(\theta_0)-\hat{\psi})+tW_n(\dot{\psi}_0)+t^2\frac{V_{0,n}}{2}} \times$$

$$\frac{\int_{A_n} e^{-n\frac{\|\theta_t-\theta_0\|_L^2}{2}+\sqrt{n}W_n(\theta_t-\theta_0)+R_n(\theta_t)} d\pi(\theta)}{\int_{A_n} e^{-n\frac{\|\theta-\theta_0\|^2}{2}+\sqrt{n}W_n(\theta-\theta_0)+R_n(\theta)} d\pi(\theta)}$$

Comments

- **LAN+ Concentration + smoothness** Usual type of condition. Posterior concentration rates (LAN norm)

Comments

- **LAN+ Concentration + smoothness** Usual type of condition. Posterior concentration rates (LAN norm)
- **The condition** Means that we can consider a *change of parameters*

$$\theta_t = \theta - t\frac{\dot\psi_0}{\sqrt{n}} - \frac{t\ddot\psi_0(\theta - \theta_0)}{2\sqrt{n}} + \frac{t\Delta}{n}, \quad s.t.$$

$$d\pi(\theta_t) = d\pi(\theta)(1 + o(1))$$

In parametric cases : $\theta' = \theta + tu/\sqrt{n}$

$$\pi(\theta') = \pi(\theta)(1 + o(1)), \quad \text{if } \pi \quad \text{is } C^o$$

In nonparametric : "holes" in π.

BvM – summary and further

- **Model and aim**

$$X^n|\theta \sim P_\theta;\ \psi(\theta) \in \mathbb{R}^d; \quad \Pi(\sqrt{n}(\psi(\theta) - \hat{\psi}) \in A|X^n) \overset{P_0}{\approx} \mathcal{N}(0, v_0)$$

$$\text{and} \quad \sqrt{n}(\hat{\psi} - \psi(\theta_0)) \approx \mathcal{N}(0, v_0)$$

- **Types of *easy* conditions** • Quadratic approximation

$$\ell_n(\theta) - \ell_n(\theta_0) = -\frac{n}{2}\|\theta - \theta_0\|_L^2 + \sqrt{n}W_n(\theta - \theta_0) + R_n(\theta, \theta_0)$$

• Smooth functional

$$\psi(\theta) = \psi(\theta_0) + < \dot{\psi}_0, \theta - \theta_0 >_L + \frac{< \ddot{\psi}_0(\theta - \theta_0), \theta - \theta_0 >_L}{2} + r(\theta)$$

• Concentration

$$\exists A_n \subset \{d(\theta, \theta_0) \leq \epsilon_n\}, \quad \Pi(A_n|X^n) = 1 + o_{p_0}(1), \quad \sup_{\theta \in A_n} |\sqrt{n}r(\theta)| = o(1)$$

The nasty condition

Under the above conditions, If , $\theta_t = \theta_0 - t\frac{\dot\psi_0}{\sqrt{n}} - \frac{t\ddot\psi_0(\theta-\theta_0)}{2\sqrt{n}} + \frac{t\Delta}{n}$

$$\& \quad \frac{\int_{A_n} p_{\theta_t}(Y^n)d\pi(\theta)}{\int_{A_n} p_{\theta}(Y^n)d\pi(\theta)} = 1 + o_p(1)$$

Then a posteriori :

$$\sqrt{n}(\psi(\theta) - \hat\psi) \approx \mathcal{N}(0, v_{0,n}), \quad \hat\psi = \psi(\theta_0) + \frac{W_n(\dot\psi_0)}{\sqrt{n}} - \frac{W_n(\Delta)}{n}$$

$$V_{0,n} = \|\dot\psi_0 - \frac{\Delta}{\sqrt{n}}\|_L^2$$

linear regime : $\ddot{\psi}_0 = 0$

$$\theta_t = \theta_0 - t\frac{\dot{\psi}_0}{\sqrt{n}}$$

$$\& \quad \frac{\int_{A_n} p_{\theta_t}(Y^n)d\pi(\theta)}{\int_{A_n} p_{\theta}(Y^n)d\pi(\theta)} = 1 + o_p(1)$$

Then a posteriori :

$$\sqrt{n}(\psi(\theta) - \hat{\psi}) \approx \mathcal{N}(0, V_{0,n}), \quad \hat{\psi} = \psi(\theta_0) + \frac{W_n(\dot{\psi}_0)}{\sqrt{n}}$$

$$V_{0,n} = \|\dot{\psi}_0\|_L^2$$

BvM

Example in linear regime

- **Model** $X_1, ..., X_n | f \sim f$ i.i.d $X_i \in [0,1]$, $\theta = \log f$
- **functionals**
  - Entropy $\psi(f) = \int_0^1 f \log f(x) dx$ & f smooth

$$\dot{\psi}_0 = \log f_0 - \psi(f_0), \quad \ddot{\psi}_0 = 0$$

- **Prior model** random histogram

$$f(x) = \sum_{j=1}^{k} \mathbb{I}_{I_j}(x) k w_j, \quad \sum w_j = 1, \quad I_j = ((j-1)/k, j/k]$$

$$(w_1, \cdots, w_k) \sim \mathcal{D}(\alpha_1, \cdots, \alpha_k)$$

Example in linear regime

- **Model** $X_1, ..., X_n | f \sim f$ i.i.d $X_i \in [0, 1]$, $\theta = \log f$
- **functionals**
  - Entropy $\psi(f) = \int_0^1 f \log f(x) dx$ & f smooth

$$\dot{\psi}_0 = \log f_0 - \psi(f_0), \quad \ddot{\psi}_0 = 0$$

  - Linear $\psi(f) = \int a(x) f(x) dx$.

$$\dot{\psi}_0 = a - \psi(f_0), \quad \ddot{\psi}_0 = 0$$

- **Prior model** random histogram

$$f(x) = \sum_{j=1}^{k} \mathbb{I}_{I_j}(x) k w_j, \quad \sum w_j = 1, \quad I_j = ((j-1)/k, j/k]$$

$$(w_1, \cdots, w_k) \sim \mathcal{D}(\alpha_1, \cdots, \alpha_k)$$

Results

$$f_0 \in \mathcal{H}(\beta), \beta > 0, \quad \|\log f_0\|_\infty < +\infty$$

$$\begin{aligned} \theta_t &= \log f_{w,k} - \frac{t\dot{\psi}_0}{\sqrt{n}} = \log f_{w,k} - \frac{t\dot{\psi}_{[k]}}{\sqrt{n}} + \frac{t}{\sqrt{n}}[\dot{\psi}_{[k]} - \dot{\psi}_0] \\ &:= \theta_{t[k]} + \frac{t}{\sqrt{n}}[\dot{\psi}_{[k]} - \dot{\psi}_0] \end{aligned}$$

and $A_{n,k} = \{f_{w,k}; h(f_{w,k}, f_{0[k]}) \lesssim \sqrt{k \log n/n}\}$

$$\ell_n(\theta_t) - \ell_n(\theta_{t[k]}) = \sqrt{n}\int(\dot{\psi}_{[k]} - \dot{\psi}_0)(f_{0[k]} - f_0) + \mathbb{G}_n(\dot{\psi}_{[k]} - \dot{\psi}_0) + o_p(1)$$

True for any $k \lesssim n/(\log n)^2$.

Examples of functionals

$$\sqrt{n}\int(\dot\psi_{[k]}-\dot\psi_0)(f_{0[k]}-f_0)+\mathbb{G}_n(\dot\psi_{[k]}-\dot\psi_0)$$

- **Deterministic k case** : $k = K_n = \lfloor\sqrt{n}(\log n)^{-2}\rfloor$
  - Entropy : $\dot\psi = \log f_0 - \psi(f_0)$, $\beta > 1/2$: BVM Model k
- **random k case** : $k \sim \mathcal{P}(\lambda)$

Examples of functionals

$$\sqrt{n}\int(\dot\psi_{[k]}-\dot\psi_0)(f_{0[k]}-f_0)+\mathbb{G}_n(\dot\psi_{[k]}-\dot\psi_0)$$

- **Deterministic *k* case** : $k = K_n = \lfloor \sqrt{n}(\log n)^{-2} \rfloor$
  - Entropy : $\dot\psi = \log f_0 - \psi(f_0)$, $\beta > 1/2$: BVM Model k
  - Linear & $a \in \mathcal{H}(\gamma)$, $\beta + \gamma > 1$: BVM linear CDF : BVM
- **random *k* case** : $k \sim \mathcal{P}(\lambda)$

Examples of functionals

$$\sqrt{n}\int(\dot\psi_{[k]}-\dot\psi_0)(f_{0[k]}-f_0)+\mathbb{G}_n(\dot\psi_{[k]}-\dot\psi_0)$$

- **Deterministic k case** : $k = K_n = \lfloor\sqrt{n}(\log n)^{-2}\rfloor$
  - Entropy : $\dot\psi = \log f_0 - \psi(f_0)$, $\beta > 1/2$: BVM Model k
  - Linear & $a \in \mathcal{H}(\gamma)$, $\beta + \gamma > 1$: BVM linear CDF : BVM
- **random k case** : $k \sim \mathcal{P}(\lambda)$
  - entropy $\dot\psi = \log f_0 - \psi(f_0)$, $\beta > 1/2$: BVM

Examples of functionals

$$\sqrt{n}\int(\dot\psi_{[k]}-\dot\psi_0)(f_{0[k]}-f_0)+\mathbb{G}_n(\dot\psi_{[k]}-\dot\psi_0)$$

- **Deterministic k case** : $k = K_n = \lfloor \sqrt{n}(\log n)^{-2} \rfloor$
  - Entropy : $\dot\psi = \log f_0 - \psi(f_0)$, $\beta > 1/2$: BVM Model k
  - Linear & $a \in \mathcal{H}(\gamma)$, $\beta + \gamma > 1$: BVM linear CDF : BVM
- **random k case** : $k \sim \mathcal{P}(\lambda)$
  - entropy $\dot\psi = \log f_0 - \psi(f_0)$, $\beta > 1/2$: BVM
  - Linear : Risk of bias : There are counterexamples

Some explanation about bias : where it can go wrong

$$\frac{\int_{A_n} e^{-n\frac{\|\eta_t-\eta_0\|_L^2}{2}+\sqrt{n}W_n(\eta_t-\eta_0)+R_n(\eta_t)}d\pi(\eta)}{\int_{A_n} e^{-n\frac{\|\eta-\eta_0\|^2}{2}+\sqrt{n}W_n(\eta-\eta_0)+R_n(\eta)}d\pi(\eta)}$$

$$\eta_t = \eta - t\frac{\dot\psi_0}{\sqrt{n}}, \quad \eta = \log f = \log(\sum_{j=1}^{k}\omega_j k \mathbb{1}_{I_j})$$

$$\Rightarrow \eta_t \to \omega_t???$$

Need

$$\int (\dot\psi_0(x) - \dot\psi_{0[k]}(x))(f_0 - f_{0[k]})(x)dx = o(1/\sqrt{n})$$

Only ok if k large enough

Outline

White noise : quadratic functional - non smooth but regular $\beta > 1/4$

- **Model**

$$dX(t) = f(t)dt + n^{-1/2}dW(t) \quad f \in L^2([0,1])$$

$$X_i = \theta_i + n^{-1/2}\zeta_i, \quad \zeta_i \quad i.i.d \quad \mathcal{N}(0,1), \quad \theta \in \ell_2$$

- **True model** $\theta_0 \in \mathcal{S}_\beta := \{\sum_{j=1}^{\infty} j^{2\beta}\theta_j^2 < +\infty\}$
- **Prior** Given k :

$$\theta_j/\tau_j \sim g \quad j \leq k \quad \& \quad \theta_j = 0 \quad j > k$$

$$k = k_n \text{ OR } k \sim \pi$$

- **functional**

$$\psi(\theta) = \|\theta\|^2 (= \|f\|^2) = < 2\theta_0, \theta - \theta_0 > + \|\theta - \theta_0\|^2$$

Non smooth case $1/4 < \beta \leq 1/2$

$$\theta_t = \theta - \frac{2t\theta_0}{\sqrt{n}} - \frac{t(\theta - \theta_0)}{\sqrt{n}} + \frac{t\epsilon_{[k]}}{n}$$

- **So here** : we concentrate on $1/4 < \beta \leq 1/2$ (not nece. continuous f_0)

$$\sum_{j=0}^{\infty} j^{2\beta}\theta_{0j}^2 < +\infty$$

- **Deterministic K_n**

$$\theta_j/\tau_j \sim g \quad j \leq K_n \quad \& \quad \theta_j = 0 \quad j > K_n, \quad K_n = n/\log n$$

set $\hat{\psi} = \|f_0\|^2 + 2n^{-1/2}\sum_i \theta_{0i}\zeta_i$

- If g Gaussian with $\sum_{j\leq K_n}\tau_j^{-2} = o(n^{3/2})$

Then

$$\sqrt{n}(\psi(f) - \hat{\psi} - \frac{2K_n}{n}) \approx \mathcal{N}(0, 4\|f_0\|_L^2), \quad Var(\hat{\psi}) = 4\|f_0\|_L^2$$

BVM after recentering with $2K_n/n$

Non smooth case $1/4 < \beta \leq 1/2$

$$\theta_t = \theta - \frac{2t\theta_0}{\sqrt{n}} - \frac{t(\theta - \theta_0)}{\sqrt{n}} + \frac{t\epsilon_{[k]}}{n}$$

- **So here** : we concentrate on $1/4 < \beta \leq 1/2$ (not nece. continuous f_0)

$$\sum_{j=0}^{\infty} j^{2\beta}\theta_{0j}^2 < +\infty$$

- **Deterministic K_n**

$$\theta_j/\tau_j \sim g \quad j \leq K_n \quad \& \quad \theta_j = 0 \quad j > K_n, \quad K_n = n/\log n$$

set $\hat{\psi} = \|f_0\|^2 + 2n^{-1/2}\sum_i \theta_{0i}\zeta_i$

- If g Gaussian with $\sum_{j\leq K_n} \tau_j^{-2} = o(n^{3/2})$
- If $g \propto 1_{[-M,M]}$ (Unif) with $\sum_{j\leq K_n} \tau_j e^{-cn\tau_j^2} = o(1)$

Then

$$\sqrt{n}(\psi(f) - \hat{\psi} - \frac{2K_n}{n}) \approx \mathcal{N}(0, 4\|f_0\|_L^2), \quad Var(\hat{\psi}) = 4\|f_0\|_L^2$$

BVM after recentering with $2K_n/n$

Some remarks

- If $\beta > 1/2$ Always BVM even with k random

- About $2K_n/n$: In freq $\bar{\psi} = \sum_{j=1}^{K_n} Y_j^2 - K_n/n$
& Jacobian :

$$\theta_t = \theta(1 - t/\sqrt{n}) - \frac{t\theta_0}{\sqrt{n}}(2 - t/\sqrt{n}) - ...$$

- Conditions on τ_k (prior variances) : Need flat priors
if $\tau_k = k^{-\delta}$, then
  - $\delta < 1/4$ for Gaussian

Some remarks

- If $\beta > 1/2$ Always BVM even with k random

- About $2K_n/n$: In freq $\bar{\psi} = \sum_{j=1}^{K_n} Y_j^2 - K_n/n$
& Jacobian :

$$\theta_t = \theta(1 - t/\sqrt{n}) - \frac{t\theta_0}{\sqrt{n}}(2 - t/\sqrt{n}) - ...$$

- Conditions on τ_k (prior variances) : Need flat priors
if $\tau_k = k^{-\delta}$, then
  - $\delta < 1/4$ for Gaussian
  - $\delta < 1/2$ for Uniform

conclusion

- **BVM for** $\psi(\theta)$ based on : LAN + concentration + smoothness of ψ + Change or parameter
- **Change of parameter** No bias condition : This is the difficult condition
- **Global BVM** (for θ) $\Rightarrow$ BVM for smooth functionals but not necessary
- **Non smooth functionals** ($f(x_0)$, $\|f\|^2$ if $\beta < 1/2$) harder to get BVM (need larger k)
- **Different priors for different functionals ?** $\Rightarrow$ Different likelihoods ? See PAC Bayesian