Review of posterior consistency & convergence rates

Debdeep Pati & Anirban Bhattacharya

Department of Statistical Science, Duke University

Working group meeting presentation
Overview of the slides

- Priors on density space
- Notions of neighborhood and distances
- Consistent tests
- Weak and strong posterior consistency - main conditions & applications
- Notion of rates of posterior convergence
- Main conditions
- Examples
Basic notations

- NP Bayes - priors on infinite dimensional space (density, regression function, conditional density etc)
- Examples - Dirichlet process, Gaussian process, Levy process etc
- Today - posterior consistency & rates in density estimation
- \( \mathcal{X} \) - complete separable metric space (\( \mathbb{R} \) for our discussion), \( \mathcal{B} \) Borel \( \sigma \)-field on \( \mathcal{X} \)
- \( \mathcal{F} \) space of densities on \( (\mathcal{X}, \mathcal{B}) \) w.r.t. some dominating measure
- \( Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} f \in \mathcal{F}, f \sim \Pi \)
Main questions

- The posterior distribution is the random measure

\[
\Pi(B \mid y^n) = \frac{\int_B \prod_{i=1}^{n} f(y_i) d\Pi(f)}{\int_{\mathcal{F}} \prod_{i=1}^{n} f(y_i) d\Pi(f)}
\]

where \( B \) is a m'ble subset of \( \mathcal{F} \) and \( y^n = (y_1, \ldots, y_n) \)

- Assume data sampled i.i.d. from \( f_0 \in \mathcal{F} \)

- \( Qn: \) does the posterior concentrate on arbitrary small neighborhoods of \( f_0 \) as \( n \to \infty ? \) If so, at what rate? For which neighborhoods?

- First, need notions of distances and neighborhoods on density spaces
Distances & nbds on density space

- Weak convergence - \( f_n \rightarrow f \) weakly if for any bounded continuous function \( \phi \), \( \int \phi f_n \rightarrow \int \phi f \)
- A weak nbd \( W_\epsilon(f_0) = \{ f \in \mathcal{F} : |\int \phi f - \int \phi f_0| < \epsilon \} \)
- Strong or \( L_1 \) convergence - \( f_n \rightarrow f \) in \( L_1 \) if \( \int |f_n - f| \rightarrow 0 \)
- A strong nbd \( S_\epsilon(f_0) = \{ f \in \mathcal{F} : \int |f - f_0| = ||f - f_0||_1 < \epsilon \} \)
- Also, \( KL(f_0, f) = \int f_0 \log(f_0/f) \), \( h^2(f, f_0) = \int (\sqrt{f} - \sqrt{f_0})^2 \)
- A KL nbd \( KL_\epsilon(f_0) = \{ f \in \mathcal{F} : KL(f_0, f) < \epsilon \} \)
- Entropy of \( \mathcal{F}_0 \subset \mathcal{F} := \log N(\epsilon, \mathcal{F}_0, || \cdot ||_1) \) is log min. number of balls of radius \( \epsilon \) in the metric \( d \) required to cover \( \mathcal{F}_0 \).
- Interplay among these distances crucial, list of common inequalities in appendix
Weak / strong neighborhood / consistency

\[ K_\epsilon(f_0) = \{ f : f \in \mathcal{F}, KL(f_0, f) < \epsilon^2/4 \} \]

Weak consistency:
\[ \Pi_{\mathcal{X}}(W_\epsilon(f_0) \, | \, y^n) \to 0 \text{ a.s.} \]

Strong consistency:
\[ \Pi_{\mathcal{X}}(S_\epsilon(f_0) \, | \, y^n) \to 0 \text{ a.s.} \]

\[ W_\epsilon(f_0) = \{ f : f \in \mathcal{F}, \left| \int_{\mathbb{R}} \{ g(x)f(x) - g(x)f_0(x) \} dx \right| < \epsilon, g : \mathbb{R} \to \mathbb{R} \text{ bounded continuous} \} \]

\[ S_\epsilon(f_0) = \{ f : f \in \mathcal{F}_d, \| f - f_0 \|_1 < \epsilon \} \]
Posterior consistency

- Basic idea: posterior probability of an arbitrary nbd around \( f_0 \) goes to 1 as \( n \to \infty \)

- Weak consistency: \( \prod(W_\epsilon(f_0) \mid y^n) \to 1 \) a.s. \( f_0 \)

- Strong consistency: \( \prod(S_\epsilon(f_0) \mid y^n) \to 1 \) a.s. \( f_0 \)

- Early result by Doob (1948): posterior consistent a.e. on prior support, not useful to check consistency at a particular density

- Breakthrough result by Schwartz (1965)
Let $f_0 \in F$ and $U$ be some nbd of $f_0$

Intuitively, should be able to separate $f_0$ from $U^c$ - formalized through consistent tests

A test function $\phi_n(y^n)$ is a non-negative measurable function bounded by 1

Suppose testing $H_0 : f = f_0$ vs $H_1 : f \in U^c$

$\phi_n(y^n)$ can be thought of as a randomized decision rule so that $\phi_n(y^n) = I(\text{Rejection region}|y^n)$

A sequence of test functions said to be uniformly consistent if both probabilities of type I and II errors converge to 0 as $n$ increases
Exponentially consistent & unbiased tests

- \{\phi_n(y^n)\} is uniformly exponentially consistent if there exist constants \(C, \beta > 0\) such that
  \[
  \mathbb{E}_{f_0}[\phi_n(y^n)] \leq C \exp(-n\beta)
  \]
  \[
  \sup_{f \in \mathcal{U}_c} [1 - \phi_n(y^n)] \leq C \exp(-n\beta)
  \]

- \{\phi_n(y^n)\} is strictly unbiased if
  \[
  \mathbb{E}_{f_0}[\phi_n(y^n)] < \inf_{f \in \mathcal{U}_c} [\phi_n(y^n)]
  \]

- The two notions above are equivalent (Hoeffding’s inequality)
- Unbiased tests often easier to construct
Theorem
Let $\Pi$ be a prior on $\mathcal{F}$ and $f_0 \in \text{KL}(\Pi)$. If there exist a sequence of exponentially consistent tests for $H_0 : f = f_0$ vs $H_1 : f \in U^c$, then $\Pi(U \mid y^n) \to 1$ a.s. $P_{f_0}^\infty$

- Note $f_0 \in \text{KL}(\Pi)$ means for any $\epsilon > 0$, $\Pi(\text{KL}_\epsilon(f_0)) > 0$
- Loosely speaking, Schwartz's theorem states large KL support + model identifiability condition $\implies$ posterior consistency
- The KL distance related to likelihood ratios, since $(1/n) \sum_{i=1}^n \log\{f_0(Y_i)/f(Y_i)\} \to \text{KL}(f_0, f)$ by SLLN
Why Schwartz’ s theorem works?

Exp. Cons. sequence of tests for $f = f_0$ vs. $f \in U^c$ makes likelihood ratio small

$$\Pi(U^c \mid Y^n) = \frac{\int_{U^c} \prod_{i=1}^{n} \frac{f(Y_i)}{f_0(Y_i)} \Pi(df)}{\int_{\mathcal{F}} \prod_{i=1}^{n} \frac{f(Y_i)}{f_0(Y_i)} \Pi(df)} \rightarrow N_n$$

$$\int_{\mathcal{F}} e^{-\sum_{i=1}^{n} \log \frac{f_0(Y_i)}{f(Y_i)}} \Pi(df) \rightarrow \int_{\mathcal{F}} e^{-nKL(f_0;f)} \Pi(df)$$

$f_0 \in KL(\Pi) \Rightarrow \lim \inf e^{n\beta} D_n = \infty, \forall \beta > 0.$
Turns out that the exponentially consistent test criterion is difficult to verify
Need easy to verifiable conditions specific to neighborhoods

Theorem: weak
If $f_0 \in KL(\Pi)$, the posterior is weakly consistent at $f_0$. 
Specialized conditions for weak and strong consistency

- Turns out that the **exponentially consistent** test criterion is difficult to verify
- Need **easy to verifiable** conditions specific to neighborhoods

**Theorem: weak**
If $f_0 \in KL(\Pi)$, the posterior is weakly consistent at $f_0$.

**Theorem: strong (Ghosal et al. 1999)**
If $f_0 \in KL(\Pi)$ and there exists a sequence of subsets $\mathcal{F}_n \subset \mathcal{F}$ such that for any $\epsilon > 0$

1. $\log N(\epsilon, \mathcal{F}_n, \| \cdot \|_1) \approx o(n)$
2. $\Pi(\mathcal{F}^c_n) \leq e^{-cn}$

then the posterior is $L_1$-consistent at $f_0$. 
Weak consistency

**Weak consistency:** If $U_\phi$ is a weak neighborhood of $f_0$, for a bounded conts. function $\phi$

\[ U_\phi = \left\{ f : \left| \int \phi f - \int \phi f_0 \right| < \epsilon \right\} \]

Choose the test function to be $\phi$ since

- **Type I error:** $E_{f_0} \{ \phi(Y_1) \} = \int \phi f_0$ and
- **Power:** $\inf_{f \in U_\phi} \int \phi f \geq \int \phi f_0 + \epsilon$

$\Rightarrow$ existence of unbiased sequence of tests

**KL condition suffices for weak consistency**
**Strong consistency:** If $U$ is a strong nhbd. of $f_0$ i.e.

$$U = \{ f : \| f - f_0 \|_1 < \epsilon \}$$

Trivial to construct exponential consistent tests for

$H_0 : f = f_0$ & $H_1 : f \in C$

How do we do it?

$$C = \{ f : \| f - f_1 \|_1 \leq \| f_1 - f_0 \|_1 / 2 \}$$

Take $B = \{ y : f_1(y) > f_0(y) \}$ and $\Phi = I_B$

Then $E_{f_1}(\Phi) \geq E_{f_0}(\Phi) + \| f_1 - f_0 \|_1 / 2$
Why Ghosal et al 1999 works?

For $L_1$ consistency, we need:

\[ \Pi(\mathcal{F}_n^c) < e^{-n\beta_1}, \]
\[ \log N(\epsilon, \|\cdot\|, \mathcal{F}_n) \approx o(n) \]
Example: Density estimation using DPM

- $Y_1, Y_2, \ldots, \sim f_0 \in \mathcal{F}$, want to estimate $f_0$

- We specify $\Pi$ by
  
  $Y_i \sim N(\mu_i, \sigma_i^2), (\mu_i, \sigma_i^2) \mid \mathcal{P} \sim \mathcal{P}, \mathcal{P} \sim DP(\alpha G_0)$, $G_0$ a distribution on $\mathbb{R} \times \mathbb{R}^+$, $\pi_h$ are constructed by stick-breaking $Beta(1, \alpha)$ variates.

- Induced density of
  
  $Y_i, f(y_i) = \sum_{h=1}^{\infty} \pi_h N(y_i, \mu_h, \sigma_h^2), (\mu_h, \sigma_h)^2 \sim G_0$

- Under what conditions on $f_0$ and $G_0$ do we have weak and strong posterior consistency?
Weak cons. in DPM (Ghosal et al. 1999; Tokdar 2006)

Split $KL(f_0, f)$ into 2 parts

$$KL(f_0, f) = \int_{\mathbb{R}} f_0(y) \log \frac{f_0(y)}{\tilde{f}(y)} dy + \int_{\mathbb{R}} f_0(y) \log \frac{\tilde{f}(y)}{f(y)} dy.$$ 

$\tilde{f} = \int \phi\left(\frac{y-\mu}{\sigma}\right)\tilde{G}(\mu, \sigma)$

$f = \int \phi\left(\frac{y-\mu}{\sigma}\right)G(\mu, \sigma)$

$T_1$

Impose tail conditions on $f_0$ to approximate $f_0$ by compactly supported $\tilde{G}$ needed to construct density $\tilde{f}$

$T_2$

Verify $\tilde{f}$ in the weak support of $\Pi$ and also $T_2$ is arbitrarily small
Constructing $\tilde{f}$: approximation idea

知 $\int_{\mathbb{R}} \frac{1}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) f_0(\mu) d\mu \to f_0(y)$ as $\sigma \to 0$

欲使支持 compactly supported $\tilde{G}_n$ 使得

$\int_{\mathbb{R}} \frac{1}{\sigma} \phi\left(\frac{y-\mu}{\sigma}\right) d\tilde{G}_n(\mu, \sigma) \to f_0(y)$

令 $d\tilde{G}_n(\mu, \sigma) \propto \delta_{\sigma_n}(\sigma) f_0(\mu) I_{[-n,n]}(\mu)$

with $\sigma_n \to 0$
Handling $T_1$

A1. $f_0$ is nowhere zero, continuous and bounded by $M < \infty$.
A2. $|\int_\mathbb{R} f_0(y) \log f_0(y) dy| < \infty$.
A3. $|\int_\mathbb{R} f_0(y) \log \frac{f_0(y)}{\psi(y)} dy| < \infty$,
where $\psi(y) = \inf_{t \in [y-1, y+1]} f_0(t)$.
A4. $\exists \eta > 0$ such that $\int_\mathcal{Y} |y|^{2(1+\eta)} f_0(y) dy < \infty$.

Under (A1)-(A4), using a compactly supported sequence $\tilde{G}_n$,

$$f_n(y) = \int \frac{1}{\sigma} \phi \left( \frac{y - \mu}{\sigma} \right) d\tilde{G}_n(\mu, \sigma)$$

approximates $f_0(y)$ and makes $T_1$ arbitrarily small as $n \to \infty$. Choose $\tilde{f} = f_{n_0}$ for large enough $n_0$. 
Handling \( T_2 \)

Find a weak nhbd \( W \) of \( \tilde{G}_{n_0} \) such that for \( G \in W \), \( T_2 \) is small.

\[
T_2 = \int_{\mathbb{R}} f_0(y) \log \left\{ \frac{\phi * \tilde{G}_{n_0}(y)}{\phi * G(y)} \right\} dy < \epsilon
\]

\[
W = \{ G : \left| \phi * \tilde{G}_{n_0}(y) - \phi * G(y) \right| < \epsilon \}
\]

What are the pieces left?
Need to ensure that a DP assigns some mass at \( W \)

TRUE if \( G_0 \) has full support
Strong consistency in DPM (Sieve construction)

How do we construct a sieve $\mathcal{F}_n$ such that

1. $\log N(\mathcal{F}_n, \| \cdot \|, \epsilon) = o(n)$
2. $\Pi(\mathcal{F}_n^c) = O(e^{-n})$

Ghosal et al. 1999 restrictive in terms of applicability

An alternative (Pati, Dunson & Tokdar, 2011)

$\mathcal{F}_n$ resembles finite mixtures

$$\sum_{h=1}^{m_n} \pi_h \frac{1}{\sigma_h} \phi\left(\frac{y-\mu_h}{\sigma_h}\right)$$

1. First few atoms are in a compact set
2. Tail sum is small
Strong consistency in DPM (Sieve construction)

\[ \Theta_{a_n, h_n, l_n} = \{ (\mu, \sigma) : |\mu| \leq a_n, l_n \leq \sigma \leq h_n \}. \]

\[ F_n = \left\{ f : f(y) = \sum_{h=1}^{\infty} \pi_h \frac{1}{\sigma_h} \phi \left( \frac{y - \mu_h}{\sigma_h} \right), \{ (\mu_h, \sigma_h) \}_{h=1}^{m_n} \in \Theta_{a_n, h_n, l_n}, \sum_{h \geq m_n + 1} \pi_h \leq \epsilon \right\}. \]
Sieve construction (Contd.)

For \( f_1, f_2 \in \mathcal{F}_n, \|f_1 - f_2\|_1 \leq \int_{X} \sum_{h=1}^{m_n} \pi_h^{(1)} \left| \phi_{\mu_h^{(1)}, \sigma_h^{(1)}}(y) - \phi_{\mu_h^{(2)}, \sigma_h^{(2)}}(y) \right| \, dy \)

\[ + \sum_{h=1}^{m_n} \left| \pi_h^{(1)} - \pi_h^{(2)} \right| + 2\varepsilon. \]

\textcolor{cyan}{balls needed=} \( N(\Theta_{a_n, h_n, l_n, \epsilon, \|\cdot\|}) \leq d_1(\frac{a_n}{l_n}) + d_2 \log \frac{h_n}{l_n} + 1. \)

\textcolor{brown}{balls needed=} \( N(\mathcal{F}_n, 4\epsilon, \|\cdot\|_1) \leq \left\{ d_1(\frac{a_n}{l_n}) + d_2 \log \frac{h_n}{l_n} + 1 \right\} m_n \)
Strong consistency (Choice of $m_n, a_n, l_n, h_n$)

1. If $G_0 = N_p(\mu;\mu_0, \Sigma_0) \times IG(\sigma^2; a, b)$, then $a_n = O(\sqrt{n}), l_n = O\left(\frac{1}{\sqrt{n}}\right), h_n = e^n$.

2. $P(\sum_{h=m_n+1}^{\infty} \pi_h > \epsilon) \leq e^{-m_n \log m_n}, m_n = O\left(\frac{n}{\log n}\right)$

3. With these choices of $m_n, a_n, l_n, h_n$, given any $\xi > 0$,

$$\log(N(F_n, 4\epsilon, \|\cdot\|_1)) = o(n),$$

4. $\Pi(F_n^c) \leq O(e^{-n})$
Once we have consistency, natural to ask whether we can characterize how fast the posterior concentrates.

In posterior consistency, we consider a fixed ball of radius $\epsilon$ around $f_0$.

Let the ball around $f_0$ shrink with $n$ as fast as possible so that it still captures most of the posterior mass.

The minimum possible such sequence $\epsilon_n$ such that
\[
\mathbb{E}\{\Pi(f : d(f, f_0) \geq M\epsilon_n \mid y^n}\} \to 0
\]
is called the rate of convergence of the posterior.
Main rate theorem

Ghosal, Ghosh & van der Vaart (2000)

Suppose that for a sequence \( \epsilon_n \to 0 \) with \( n\epsilon_n^2 \to \infty \), a constant \( C > 0 \) and sets \( F_n \subset F \), one has

\[
\log N(\epsilon_n, F_n, d) \leq C_1 n\epsilon_n^2
\]

\[
\Pi(F_n^c) \leq C_3 \exp\{-n\epsilon_n^2(C_2 + 4)\}
\]

\[
\Pi\left(f_{\mu,\sigma} : \int f_0 \log \frac{f_0}{f_{\mu,\sigma}} \leq \epsilon_n^2, \int f_0 \log \left(\frac{f_0}{f_{\mu,\sigma}}\right)^2 \leq \epsilon_n^2\right) \geq C_4 \exp\{-C_2 n\epsilon_n^2\}.
\]

Then, for sufficiently large \( M \), \( \mathbb{E}\{\Pi(f : d(f, f_0) \geq M\epsilon_n \mid y^n}\} \to 0 \)

- A more subtle interplay, roughly requires prior to be uniformly spread over the parameter space
- \( d \) usually Hellinger or \( L_1 \) metric
Application to a specific problem

- Density estimation model (Kundu & Dunson, 2011)

\[ y_i = \mu(\eta_i) + \epsilon_i, \quad \eta_i \sim U(0, 1), \]
\[ \epsilon_i \sim N(0, \sigma^2), \quad (i = 1, \ldots, n). \]

- \( f_0 \) true density, \( F_0 \) c.d.f. with \( \mu_0 = F_0^{-1} : (0, 1) \rightarrow \mathbb{R} \), induced density \( f_{\mu_0, \sigma}(y) = \)

\[
\int_0^1 \phi_\sigma(y - F_0^{-1}(t)) dt = \int_{a_0}^{b_0} \phi_\sigma(y - z)f_0(z) dz
\]

- \( f_{\mu_0, \sigma}(y) = \phi_\sigma * f_0(y) \), smoothness assumptions on \( f_0 \) imply \( d(f_0, f_{\mu_0, \sigma}) \rightarrow 0 \) as \( \sigma \rightarrow 0 \)

- \( f_0 \) compactly supported implies \( \mu_0 : [0, 1] \rightarrow [a_0, b_0] \)

- \( f_0 \) supported on \( \mathbb{R} \) implies \( |\mu_0(t)| \rightarrow \infty \) as \( t \rightarrow 0/1 \)
Prior specification

- Prior for \((\mu, \sigma) \in C([0, 1]) \otimes (0, \infty)\) induces a prior on the space of densities on \((\mathbb{R}, \mathcal{B})\).

- Intuition: \(\Pi_{\mu}\) concentrating around \(\mu_0\) and \(\Pi_{\sigma}\) around zero would imply \(f_{\mu,\sigma}\) places +ve probability to arbitrary nbds of \(f_0\).

- Induced measure \(\nu_{\mu}(B) = \tilde{\lambda}(\mu^{-1}(B)), \mu : ([0, 1], \tilde{\lambda}) \to (\mathbb{R}, \mathcal{B})\) m’ble, \(\tilde{\lambda}\) Leb. meas. on \([0, 1]\).

- Marginalizing out \(\eta_i\), induced density \(f_{\mu,\sigma}\),

\[
f_{\mu,\sigma}(y) = \int_0^1 \phi_\sigma(y - \mu(t))dt = \int \phi_\sigma(y - z)\nu_{\mu}(dz)
\]
Want mechanism to produce random (continuous) functions.

A random vector $X : (\Omega, \mathcal{E}, P) \to \mathbb{R}^k$ is Gaussian if $a^T X$ is Gaussian for any $a \in \mathbb{R}^k$

Let $X : (\Omega, \mathcal{E}, P) \to (C[0, 1], \| \cdot \|_\infty)$ be measurable

$X$ is called Gaussian if $L(X)$ is Gaussian for any linear functional $L$

For example, $L(f) = f(1/2)$, $L(f) = 2f(1/3) - f(3/4)$, ...

Clearly, for any $(t_1, \ldots, t_m)$, $\sum_{i=1}^m a_i X(t_i)$ is Gaussian for any $a \in \mathbb{R}^m$

$(X_{t_1}, \ldots, X_{t_m})$ is MVN
Specify a joint Gaussian for \((X_{t_1}, \ldots, X_{t_m})\) consistently

Let \(C(t, s)\) be a positive definite covariance kernel, i.e., \(C = (C(t_i, t_j))\) is positive definite for any \(t_1, \ldots, t_m\)

\((X_{t_1}, \ldots, X_{t_m}) \sim N(0, C)\), so that \(C(s, t) = \text{cov}(X_s, X_t)\)

Common examples: \(C(t, s) = \min(t, s)\), 
\(C(t, s) = \exp(-\kappa|t - s|)\), \(C(t, s) = \exp(-\kappa|t - s|^2)\) etc
Mercer's theorem: There exists a sequence of eigenvalues $\lambda_h \downarrow 0$ and an orthonormal system of eigenfunctions $\phi_h$, such that

$$C(s, t) = \sum_{h=1}^{\infty} \lambda_h \phi_h(s) \phi_h(t)$$

Define $\tilde{X}(t) = \sum_{h=1}^{\infty} \lambda_h^{1/2} Z_h \phi_h(t)$, where $Z_h$ i.i.d. $\mathcal{N}(0, 1)$

$$\text{cov}(\tilde{X}_s, \tilde{X}_t) = \sum_{h=1}^{\infty} \lambda_h \phi_h(s) \phi_h(t) = C(s, t)$$

We can start with a series representation by choosing $\lambda_h$ and $\phi_h$. Different choices lead to splines, neural networks, wavelets, etc.
**RKHS of Gaussian processes**

- In np Bayes, want priors to place positive probability around arbitrary neighborhoods of a large class of parameter values (large support property)
- The prior concentration plays a key role in determining the rate of posterior contraction
- The reproducing kernel Hilbert space (RKHS) of a Gaussian process determines the prior support and concentration
- Let $X$ be a zero mean Gaussian process on $[0, 1]$ with covariance kernel $C(s, t) = E(X_sX_t)$
- The RKHS $\mathcal{H}$ is the completion of the linear space

$$f(t) = \sum_{h=1}^{m} a_h C(s_h, t), \quad s_h \in [0, 1], \quad a_h \in \mathbb{R}.$$ 

- Intuitively, a space of functions that are similar to the covariance kernel in terms of smoothness
If \( f_1(t) = C(s_1, t), f_2(t) = C(s_2, t) \), define \((f_1, f_2)_H = C(s_1, s_2)\). Extend linearly and continuously to whole of \( H \).

Finite-dimensional case: let \( X \sim N_k(0, \Sigma) \), \( \Sigma \) pd. Then \( H = \mathbb{R}^k \), \((x, y)_H = x^T \Sigma^{-1} y \) and hence \( ||x||_H^2 = x^T \Sigma^{-1} x \). Same RKHS norm on density contours!!

The support of a mean zero Gaussian process is the closure of the RKHS. For many standard covariance kernels, the support equals \( C[0, 1] \).

The rate of posterior contraction at a function \( f_0 \) depends on

\[
\phi_{f_0}(\epsilon) = \inf_{h \in H: ||h - f_0||_H < \epsilon} ||h||_H^2 - \log \Pr(||X||_\infty < \epsilon)
\]
Back to the rates problem

- Ongoing work (Pati, Bhattacharya & Dunson, 2011) on posterior convergence rates in NL-LVM model
- Only focus on the compactly supported case here
- Analysis of non-compact case more involved as quantile function of a non-compact density not in $C[0, 1]$
- Standard sieve available for GP priors (van der Vaart & van Zanten 2007 onwards) - clever application of Borel’s inequality
- KL condition main hurdle
Assume $f_0$ twice continuously differentiable, optimal minimax rate in that case $n^{-2/5}$

Using a GP prior with squared exponential covariance kernel for $\mu$ & an inverse-gamma prior for $\sigma$, we achieve the minimax rate up to a log-factor

One has

$$\int f_0 \log \left( \frac{f_0}{f_{\mu,\sigma}} \right)^2 \leq h^2(f_0, f_{\mu,\sigma}) \left( 1 + \log \| f_0 / f_{\mu,\sigma} \|_\infty \right)^2$$

With $\epsilon_n = n^{-2/5} (\log n)^\kappa$ and $\sigma_n^4 = \epsilon_n^2$,

$$\{ \sigma \in [\sigma_n, \sigma_n + \sigma_n^b], \| \mu - \mu_0 \|_\infty \lesssim O(\sigma_n^3) \} \subset \left\{ \int f_0 \log \frac{f_0}{f_{\mu,\sigma}} \lesssim \sigma_n^4, \int f_0 \log \left( \frac{f_0}{f_{\mu,\sigma}} \right)^2 \lesssim \sigma_n^4 \right\}.$$
Appendix - list of common inequalities

▶ $\|p - q\|_1^2 \leq 4h^2(p, q) \leq 4\|p - q\|_1$
▶ $\text{KL}(p, q) \geq \|p - q\|_1^2 / 2$
▶ $\text{KL}(p, q) \leq h^2(p, q)\{1 + \log \|p/q\|_\infty\}$
▶ $p = \mathcal{N}(\mu_1, \sigma_1^2), q = \mathcal{N}(\mu_2, \sigma_2^2)$ with $\sigma_2 > \sigma_1 > \sigma_2 / 2$, then $\|p - q\|_1 \leq (2/\pi)^{0.5} |\mu_1 - \mu_2|/\sigma_2 + 3(\sigma_2 - \sigma_1)/\sigma_1$
Key references:

▸ Ghosal’s research page
   \textit{http://www4.stat.ncsu.edu/~sghosal/papers.html}

▸ van der Vaart’s page
   \textit{http://www.few.vu.nl/~aad/research.html}

▸ van Zanten’s page
   \textit{http://www.win.tue.nl/~jzanten/research.html}

▸ Key consistency references: Barron, Schervish & Wasserman, 1999; Ghosal, Ghosh & Ramamurthy, 1999; Tokdar, 2006; Tokdar & Ghosh, 2007


▸ Several others not cited! Our apologies - See references within these articles.