A new way for achieving adaptation

ISBA-BNP Webinar

Sergios Agapiou (joint work with Ismael Castillo)

24th May 2023

Department of Mathematics and Statistics, University of Cyprus

Table of Contents

Motivation

Series Priors

- For (φ_k) an orthonormal basis of $L^2[0,1]$, let

$$f = \sum_{k=1}^{\infty} \sigma_k \zeta_k \varphi_k(\cdot),$$

where $\zeta_k \overset{iid}{\sim} h$ and $\sigma_k = \tau k^{-1/2-S}$, scaling $\tau > 0$, regularity $S > 0$.

- For (ψ_{lk}) an orthonormal wavelet basis of $L^2[0,1]$, let

$$f = \sum_{l=1}^{\infty} \sum_{k \in \mathcal{K}_l} s_l \zeta_{lk} \psi_{lk}(\cdot)$$

where $\zeta_{lk} \overset{iid}{\sim} h$ and $s_l = \tau 2^{-l(1/2+S)}$, scaling $\tau > 0$, regularity $S > 0$.

Literature - Rates of Contraction

- Gaussian: for appropriately tuned τ or S get minimax optimal contraction rates over Sobolev or Hölder regularity classes (not spatially inhomogeneous Besov)
  - A. van der Vaart and H. van Zanten, *Rates of contraction of posterior distributions based on Gaussian process priors*, Annals of Statistics, 2008
  - Many contributions in many settings!
- Laplace (more generally p-exponential): for appropriately tuned τ and/or S get minimax optimal contraction rates over Sobolev, Hölder, Besov classes
  - S. Agapiou, M. Dashti and T. Helin, *Rates of contraction of posterior distributions based on p-exponential priors*, Bernoulli, 2021
  - M. Giordano and K. Ray, *Nonparametric Bayesian inference for reversible multidimensional diffusions*, Annals of Statistics, 2022
  - S. Agapiou and S. Wang, *Laplace priors and spatial inhomogeneity in Bayesian inverse problems*, Bernoulli, 2023+
  - M. Giordano, *Besov priors in density estimation: optimal posterior contraction rates and adaptation*, arXiv:2208.14350
  - S. Agapiou and A. Savva, *Adaptive inference over Besov spaces in the white noise model using p-exponential priors*, arXiv:2209.06045

Literature - Rates of Contraction

- Gaussian: for appropriately tuned τ or S get minimax optimal contraction rates over Sobolev or Hölder regularity classes (not spatially inhomogeneous Besov)
  - A. van der Vaart and H. van Zanten, *Rates of contraction of posterior distributions based on Gaussian process priors*, Annals of Statistics, 2008
  - Many contributions in many settings!

- Laplace (more generally p-exponential): for appropriately tuned τ and/or S get minimax optimal contraction rates over Sobolev, Hölder, Besov classes
  - S. Agapiou, M. Dashti and T. Helin, *Rates of contraction of posterior distributions based on p-exponential priors*, Bernoulli, 2021
  - M. Giordano and K. Ray, *Nonparametric Bayesian inference for reversible multidimensional diffusions*, Annals of Statistics, 2022
  - S. Agapiou and S. Wang, *Laplace priors and spatial inhomogeneity in Bayesian inverse problems*, Bernoulli, 2023+
  - M. Giordano, *Besov priors in density estimation: optimal posterior contraction rates and adaptation*, arXiv:2208.14350
  - S. Agapiou and A. Savva, *Adaptive inference over Besov spaces in the white noise model using p-exponential priors*, arXiv:2209.06045

Heavy-tailed Priors Avoid Tuning?

White Noise Model

Theorem (A., Dashti, Helin '21)

Assume $f_0 \in \mathcal{S}^\beta$. Consider S-regular p-exponential priors $p \in [1,2]$ with $\tau = 1$. Then the posterior contracts at rate

$$\epsilon_n = \begin{cases} n^{-\frac{\beta}{1+2\beta+p(S-\beta)}}, & \text{if } S > \beta \\ n^{-\frac{S}{1+2S}}, & \text{if } S \leq \beta \end{cases}$$

Heavy-tailed Priors Avoid Tuning?

White Noise Model

Theorem (A., Dashti, Helin '21)

Assume $f_0 \in \mathcal{S}^\beta$. Consider S-regular p-exponential priors $p \in [1,2]$ with $\tau = 1$. Then the posterior contracts at rate

$$\epsilon_n = \begin{cases} n^{-\frac{\beta}{1+2\beta+p(S-\beta)}}, & \textit{if } S > \beta \\ n^{-\frac{S}{1+2S}}, & \textit{if } S \leq \beta \end{cases}$$

- If we could take $p \searrow 0$, for a fixed large S we would get the minimax rate without tuning for $\beta \in (0, S)$!

Heavy-tailed Priors Avoid Tuning?

White Noise Model

Theorem (A., Dashti, Helin '21)

Assume $f_0 \in \mathcal{S}^\beta$. Consider S-regular p-exponential priors $p \in [1,2]$ with $\tau = 1$. Then the posterior contracts at rate

$$\epsilon_n = \begin{cases} n^{-\frac{\beta}{1+2\beta+p(S-\beta)}}, & \text{if } S > \beta \\ n^{-\frac{S}{1+2S}}, & \text{if } S \le \beta \end{cases}$$

- If we could take $p \searrow 0$, for a fixed large S we would get the minimax rate without tuning for $\beta \in (0, S)$!
- If we could take $S \to \infty$, we would get the minimax rate for $\beta > 0$!

Setup

Heavy-tailed Series Priors

For (φ_k) orthonormal basis, (ψ_{lk}) orthonormal wavelet basis of $L^2[0,1]$

$$f = \sum_{k=1}^{\infty} \sigma_k \zeta_k \varphi_k(\cdot), \qquad f = \sum_{l=1}^{\infty} \sum_{k \in \mathcal{K}_l} s_l \zeta_{lk} \psi_{lk}(\cdot)$$

where

$$\sigma_k = k^{-1/2-S} \quad \text{or} \quad \sigma_k = e^{-(\log k)^2}$$

$$s_l = 2^{-l(1/2+S)} \quad \text{or} \quad s_l = 2^{-l^2}$$

We take i.i.d ζ's from a heavy-tailed pdf h.

Heavy-tailed Series Priors

HT Assumptions

For some constants $c_1, c_2 > 0$ and $\kappa \geq 0$, assume

- h is symmetric, positive, bounded and decreasing on $[0, \infty)$
- $$\log(1/h(x)) \leq c_1(1 + \log^{1+\kappa}(1 + x)), \qquad x \geq 0$$
- $$\overline{H}(x) := \int_x^\infty h(u)du \leq \frac{c_2}{x^2}, \qquad x \geq 1$$

$\kappa = 0$: polynomial tails e.g. Cauchy or Student (for Cauchy $\overline{H}(x) \asymp x^{-1}$)

Heavy-tailed Priors in Applied BNP

- A. Shah, A. Wilson and Z. Ghahramani, *Student-t Processes as Alternatives to Gaussian Processes*, PMLR, 2014
- C. M. Carvalho, N. G. Polson, and J. G. Scott, *The horseshoe estimator for sparse signals*, Biometrika, 2010
- S. van der Pas, B. Szabo and A. van der Vaart, *Uncertainty quantification for the horseshoe*, Bayesian Analysis, 2017
- T. Sullivan, *Well-posed Bayesian inverse problems and heavy-tailed stable quasi-Banach space priors*, Inverse Problems and Imaging, 2017
- M. Markkanen, L. Roininen, J. Huttunen and S. Lasanen, *Cauchy difference priors for edge-preserving Bayesian inversion*, Journal of Inverse and Ill-posed Problems, 2019
- J. Suuronen, N. Chada and L. Roininen, *Cauchy Markov Random Field Priors for Bayesian Inversion*, Statistics and Computing, 2022

Regularity Assumptions for the Truth

We will consider three types of smoothness assumptions:

- Sobolev: $f_0 \in \mathcal{S}(\beta, L)$ for some $\beta, L > 0$ if

$$\sum_{k=1}^{\infty} k^{2\beta} f_{0,k}^2 \leq L^2.$$

- Hölder (Zygmund): $f_0 \in \mathcal{H}(\beta, L)$ for some $\beta, L > 0$ if

$$2^{l(1/2+\beta)} \max_{k \in \mathcal{K}_l} |f_{0,lk}| \leq L.$$

- Besov: $f_0 \in \mathcal{B}(\beta, r, L)$, for some $\beta > 0$, $1 \leq r \leq 2$, $L > 0$ if

$$\sum_{l=1}^{\infty} 2^{rl(\beta+1/2-1/r)} \sum_{k \in \mathcal{K}_l} |f_{lk}|^r \leq L^r.$$

For $r < 2$ spatial inhomogeneity

Contraction in the White Noise Model

White Noise Model

- Equivalently consider the normal sequence model

$$X_k | f_k \sim \mathcal{N}(f_k, 1/n) \qquad \text{(single index)}$$

or

$$X_{lk} | f_{lk} \sim \mathcal{N}(f_{lk}, 1/n) \qquad \text{(double index)}$$

- We denote by $X^{(n)}$ the corresponding observation sequence.

White Noise Model

- Equivalently consider the normal sequence model

$$X_k | f_k \sim \mathcal{N}(f_k, 1/n) \qquad \text{(single index)}$$

or

$$X_{lk} | f_{lk} \sim \mathcal{N}(f_{lk}, 1/n) \qquad \text{(double index)}$$

- We denote by $X^{(n)}$ the corresponding observation sequence.

Moment Assumption

$$\int_{-\infty}^{\infty} x^2 h(x) dx < \infty \quad \text{or} \quad \int_{-\infty}^{\infty} |x|^q h(x) dx < \infty, \; q \geq 1$$

Contraction in L^2-loss, White Noise Model

Theorem (A and Castillo '23+)

Let Π be a heavy-tailed series prior defined via an orthonormal basis (φ_k), for h satisfying the *HT Assumptions* and the *Moment Assumption* with $q = 2$.

Let $f_0 \in \mathcal{S}(\beta, L)$ for $L > 0$ and consider one of the next two settings:

- $\sigma_k = k^{-1/2-S}$ for $S \geq \beta$;
- $\sigma_k = e^{-(\log k)^2}$, $\beta > 0$.

Then in either setting, as $n \to \infty$

$$E_{f_0}\left[\int \|f - f_0\|_2^2 d\Pi(f \mid X)\right] \lesssim n^{-\frac{2\beta}{2\beta+1}} (\log n)^d,$$

for some $d > 0$.

Contraction in L^∞-loss, White Noise Model

Theorem (A and Castillo '23+)

Let Π be a heavy-tailed series prior defined via an orthonormal wavelet basis (ψ_{lk}), for h satisfying the *HT Assumptions* and the *Moment Assumption* with $q \geq 1$.

Let $f_0 \in \mathcal{H}(\beta, L)$ for $L > 0$ and consider one of the next two settings:

- $s_l = 2^{-l(1/2+S)}$ for $S > \{(1/q + \frac{\beta}{1+2\beta}) \vee \beta\}$;
- $s_l = 2^{-l^2}$, $\beta > 0$.

Then in either setting, as $n \to \infty$

$$E_{f_0}\left[\int \|f - f_0\|_\infty d\Pi(f \mid X)\right] \lesssim (n/\log n)^{-\frac{\beta}{2\beta+1}}(\log n)^d,$$

for some $d > 0$.

Contraction in White Noise Model

Corollary

By Markov inequality, the two last results imply in their corresponding settings that as $n \to \infty$

$$E_{f_0} \Pi[\{f : \|f - f_0\|_2 > \mathcal{L}_n n^{-\frac{\beta}{2\beta+1}}\} \mid X^{(n)}] \to 0,$$

and

$$E_{f_0} \Pi[\{f : \|f - f_0\|_\infty > \mathcal{L}_n (n/\log n)^{-\frac{\beta}{2\beta+1}}\} \mid X^{(n)}] \to 0,$$

respectively, where $\mathcal{L}_n = (\log n)^d$ for some $d > 0$.

Contraction in White Noise Model - Proof Ideas

- L_2-loss with $\sigma_k = k^{-1/2-S}$ and $|f_{0,k}| \lesssim k^{-1/2-\beta}$, other cases similar
- Need to bound
$$E_{f_0}\left[\int \|f - f_0\|_2^2 d\Pi(f \mid X)\right],$$
work coefficient-wise

Contraction in White Noise Model - Proof Ideas

- L_2-loss with $\sigma_k = k^{-1/2-S}$ and $|f_{0,k}| \lesssim k^{-1/2-\beta}$, other cases similar
- Need to bound

$$E_{f_0}\left[\int \|f - f_0\|_2^2 d\Pi(f \mid X)\right],$$

work coefficient-wise

- Using heavy tails assumption

$$E_{f_0}\int (f_k - f_{0,k})^2 d\Pi(f \mid X) \lesssim n^{-1} \log^{1+\kappa}\left(1 + \frac{L + 1/\sqrt{n}}{\sigma_k}\right)$$

- For $k \le K_n := n^{1/(1+2\beta)}$, since $\sigma_k^{-1} \le \sigma_{K_n}^{-1}$, bound logarithmic in n!

Contraction in White Noise Model - Proof Ideas

- L_2-loss with $\sigma_k = k^{-1/2-S}$ and $|f_{0,k}| \lesssim k^{-1/2-\beta}$, other cases similar
- Need to bound
$$E_{f_0}\left[\int \|f - f_0\|_2^2 d\Pi(f \mid X)\right],$$
work coefficient-wise

- Using heavy tails assumption
$$E_{f_0}\int (f_k - f_{0,k})^2 d\Pi(f \mid X) \lesssim n^{-1} \log^{1+\kappa}\left(1 + \frac{L + 1/\sqrt{n}}{\sigma_k}\right)$$
- For $k \leq K_n := n^{1/(1+2\beta)}$, since $\sigma_k^{-1} \leq \sigma_{K_n}^{-1}$, bound logarithmic in n!
- The total contribution to the error of all $k \leq K_n$, for any $S > 0$, is
$$\lesssim n^{-1}(\log n)^d K_n \lesssim n^{-\frac{2\beta}{1+2\beta}}(\log n)^d$$

Contraction in White Noise Model - Proof Ideas

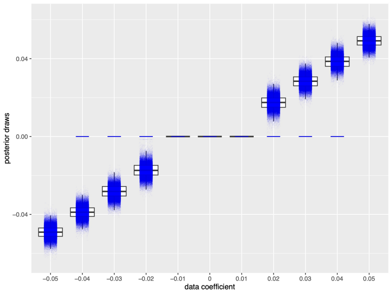
- L_2-loss with $\sigma_k = k^{-1/2-S}$ and $|f_{0,k}| \lesssim k^{-1/2-\beta}$, other cases similar
- Need to bound
$$E_{f_0}\left[\int \|f - f_0\|_2^2 d\Pi(f \mid X)\right],$$
work coefficient-wise

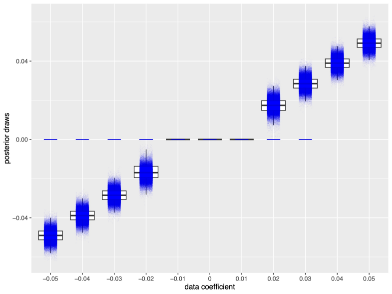
- For $k > K_n$ use $(f_k - f_{0,k})^2 \leq 2f_k^2 + 2f_{0,k}^2$
    - Assumption on f_0 implies 2nd term is small
    - Oversmoothing prior suggests 1st term also small under the posterior
    - Delicate analysis shows that for $S \geq \beta$ the contribution to the squared error is also $n^{-\frac{2\beta}{1+2\beta}}(\log n)^d$

□

Behaviour of Student Prior

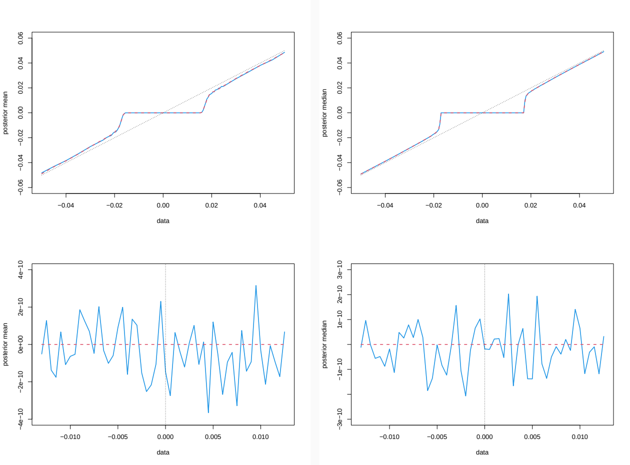
- $X \sim \mathcal{N}(f, 10^{-5})$
- $f \sim \sigma t_3$
- $\sigma = 20^{-5.5}$ (left) and $\sigma = (2e9)^{-5.5}$ (right) $(S = 5)$





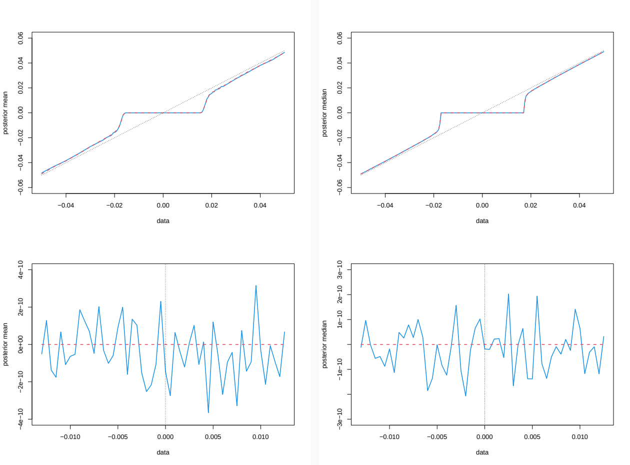
Behaviour of Student Prior

$\sigma = 20^{-5.5}$ (blue) and $\sigma = (2e9)^{-5.5}$ (red dashed)



Behaviour of Student Prior

$\sigma = 20^{-5.5}$ (blue) and $\sigma = (2e9)^{-5.5}$ (red dashed)



Similar to spike and slab prior

I. Johnstone and B. Silverman, Annals of Statistics, 2004, 2005

Contraction in White Noise Model - Besov truth

- Besov spaces $\mathcal{B}^{\beta}_{rr}$ with $1 \leq r < 2$ model spatial inhomogeneity

- Allow for large wavelet coefficients in high frequencies

- Gaussian priors are limited by the linear minimax rate $n^{-\frac{\beta+1/2-1/r}{2+2\beta-2/r}}$

  - S. Agapiou and S. Wang, *Laplace priors and spatial inhomogeneity in Bayesian inverse problems*, Bernoulli, 2023+

- Laplace priors can achieve the minimax rate ($r = 1$) or nearly the minimax rate ($r > 1$), but require tuning both S and τ

  - S. Agapiou, M. Dashti and T. Helin, *Rates of contraction of posterior distributions based on p-exponential priors*, Bernoulli, 2021
  - S. Agapiou and A. Savva, *Adaptive inference over Besov spaces in the white noise model using p-exponential priors*, arXiv:2209.06045

Contraction in White Noise Model - Besov truth

Theorem (A and Castillo '23+)

Let Π be a heavy-tailed series prior defined via an orthonormal wavelet basis (ψ_{lk}), for h satisfying the *HT Assumptions* and the *Moment Assumption* with $q = 2$.

Let $f_0 \in \mathcal{B}(\beta, r, L)$ for $1 \leq r \leq 2, L > 0$ and $\beta > 1/r - 1/2$.

Then for $s_l = 2^{-l^2}$, as $n \to \infty$

$$E_{f_0}\left[\int \|f - f_0\|_2^2 d\Pi(f \mid X)\right] \lesssim n^{-\frac{2\beta}{2\beta+1}} (\log n)^d,$$

for some $d > 0$.

General Results - ρ-posteriors

Generic Prior Mass Condition in L^2-loss

Theorem (A and Castillo '23+)

Let Π be a heavy-tailed series prior defined via an orthonormal basis (φ_k), for h satisfying the *HT Assumptions*.

Let $f_0 \in \mathcal{S}(\beta, L)$ for $L > 0$ and consider one of the next two settings:

- $\sigma_k = k^{-1/2-S}$ for $S > 1/2$, $\beta \leq S$;
- $\sigma_k = e^{-(\log k)^2}$, $\beta > 0$.

In either setting there exist $c_1, c_2, d > 0$ such that

$$\Pi[\|f - f_0\|_2 < c_1 \varepsilon_n] \geq e^{-c_2 n \varepsilon_n^2},$$

with

$$\varepsilon_n = (\log n)^d n^{-\frac{\beta}{1+2\beta}}.$$

Generic Prior Mass Condition in L^∞-loss

Theorem (A and Castillo '23+)

Let Π be a heavy-tailed series prior defined via an orthonormal wavelet basis (ψ_{lk}), for h satisfying the *HT Assumptions*.

Let $f_0 \in \mathcal{H}(\beta, L)$ for $L > 0$ and consider one of the next two settings:

- $s_l = 2^{-l(1/2+S)}$ for $S > 1/2$, $\beta \leq S$;
- $s_l = 2^{-l^2}$, $\beta > 0$.

In either setting there exist $c_1, c_2, d > 0$ such that

$$\Pi[\|f - f_0\|_\infty < c_1 \varepsilon_n] \geq e^{-c_2 n \varepsilon_n^2},$$

with

$$\varepsilon_n = (\log n)^d n^{-\frac{\beta}{2\beta+1}}.$$

Generic Prior Mass Condition - Proof Ideas

- Focus on L_2-loss with $\sigma_k = k^{-1/2-S}$ and $f_0 \in \mathcal{S}(\beta, L)$, other cases similar
- For some K to be chosen, split to low and high frequencies

$$\Pi[\|f - f_0\|_2 < \varepsilon] \geq \Pi\left[\|f^{[K]} - f_0^{[K]}\|_2 < \varepsilon/2\right] \Pi\left[\|f^{[K^c]} - f_0^{[K^c]}\|_2 < \varepsilon/2\right]$$

Generic Prior Mass Condition - Proof Ideas

- For some K to be chosen, split to low and high frequencies
$$\Pi[\|f - f_0\|_2 < \varepsilon] \geq \Pi\left[\|f^{[K]} - f_0^{[K]}\|_2 < \varepsilon/2\right] \Pi\left[\|f^{[K^c]} - f_0^{[K^c]}\|_2 < \varepsilon/2\right]$$

- $f_0 \in \mathcal{S}(\beta; L)$ implies even for small k the coefficients $f_{0,k}$ cannot be too large and so prior puts substantial mass around them
- Since h is heavy tailed, even if σ_k decays very quickly (oversmoothing prior), this mass is still substantial!
$$\Pi\left[\|f^{[K]} - f_0^{[K]}\|_2 < \varepsilon/2\right] \geq \varepsilon^K \exp\{-C_1 K \log^{1+\kappa}(C_2/\sigma_K)\}$$
$$\geq \varepsilon^K \exp\{-C_1' K \log^{1+\kappa} K\}$$

Generic Prior Mass Condition - Proof Ideas

- For some K to be chosen, split to low and high frequencies

$$\Pi[\|f - f_0\|_2 < \varepsilon] \geq \Pi\left[\|f^{[K]} - f_0^{[K]}\|_2 < \varepsilon/2\right] \Pi\left[\|f^{[K^c]} - f_0^{[K^c]}\|_2 < \varepsilon/2\right]$$

$$\Pi\left[\|f^{[K^c]} - f_0^{[K^c]}\|_2 < \varepsilon/2\right] \geq \Pi\left[\|f^{[K^c]}\|_2 < \varepsilon/4\right] \mathbb{1}_{\|f_0^{[K^c]}\|_2 < \varepsilon/4}$$

- $f_0 \in \mathcal{S}(\beta; L)$ implies $\|f_0^{[K^c]}\|_2$ is small for large K
- For S large enough and $\varepsilon \asymp K^{-\beta} \log K$

$$\Pi\left[\|f^{[K^c]}\|_2 < \varepsilon/4\right] \geq \exp(-C_3 K)$$

Generic Prior Mass Condition - Proof Ideas

- For some K to be chosen, split to low and high frequencies

$$\Pi[\|f - f_0\|_2 < \varepsilon] \geq \Pi\left[\|f^{[K]} - f_0^{[K]}\|_2 < \varepsilon/2\right] \Pi\left[\|f^{[K^c]} - f_0^{[K^c]}\|_2 < \varepsilon/2\right]$$

- Combining, for large K and $\varepsilon \asymp K^{-\beta} \log K$ it holds

$$\Pi[\|f - f_0\|_2 < \varepsilon] \geq \exp\{-CK \log^{1+\kappa} K\}$$

- Optimize choice $K = K(n)$ so that $\varepsilon \asymp K^{-\beta} \log K$ as small as possible while $K \log^{1+\kappa} K \asymp n\varepsilon^2$

$\square$

Contraction Results for ρ-posteriors

- Both results can be extended to cover Cauchy priors ($\overline{H}(x) \asymp x^{-1}$), provided $S > 1$ for the S-regular cases.

- Combining with Theorem 8.43 of GV17, these prior mass conditions imply contraction results for pseudo-posteriors

$$\Pi^{(\rho)}(\theta \in B|X) = \frac{\int_B p_\theta^\rho(X) d\Pi_n(\theta)}{\int p_\theta^\rho(X) d\Pi_n(\theta)}, \quad \rho \in (0,1).$$

- Example 8.44 of GV17 shows that for i.i.d observations, under the prior mass condition $\Pi(\theta : K(p_{f_0}; p_f) < \varepsilon_n^2) \geq \exp(-n\varepsilon_n^2)$, we have

$$\Pi_n^{(\rho)}(d_H(p_{f_0}, p_f) > M_n \varepsilon_n \,|\, X_1, \ldots, X_n) \to 0,$$

for any $M_n \to \infty$.

S. Ghoshal and A. vd Vaart, *Fundamentals of nonparametric Bayesian inference*, 2017.

Contraction Results for ρ-posteriors

- e.g. density estimation with

$$p_f(x) = \frac{e^{f(x)}}{\int e^{f(x)} dx}$$

  and a prior on f

- Lemma 2.5 of GV17 shows

$$K(p_f; p_g) \lesssim \|f - g\|_\infty^2 e^{\|f-g\|_\infty}(1 + \|f - g\|_\infty)$$

- The prior mass condition in L^∞-loss suffices for showing contraction of ρ-posteriors in Hellinger distance at the nearly minimax rate ε_n over Hölder smoothness

- Partial adaptivity $s_l = 2^{-(1/2+S)l}$ / adaptivity $s_l = 2^{-l^2}$ (up to logs)

Simulations

White Noise Model - Sobolev Truth

- Study linear inverse problem in simulations of
  B. Knapik, B. Szabo, A. van der Vaart and H. Zanten, *Bayes procedures for adaptive inference in inverse problems for the white noise model*, PTRF, 2016
- Equivalent to normal sequence model

$$X_k \sim \mathcal{N}(\lambda_k f_k, 1/n)$$

  defined wrt the eigenbasis of the forward operator

$$\varphi_k(t) = \sqrt{2}\cos(\pi(k - 1/2)t), t \in [0, 1]$$

  where

$$\lambda_k = \pi/(k - 1/2)$$

  are the corresponding eigenvalues
- Coefficients of truth

$$f_{0,k} = k^{-3/2}\sin(k)$$

  "Sobolev regularity $\beta = 1$"

White Noise Model - Sobolev Truth

- Priors
  - Gaussian hierarchical regularity prior:

$$f_k|S \sim \mathcal{N}(0, \sigma_k^2), \quad \sigma_k = k^{-1/2-S}, \quad S \sim Exp(1)$$

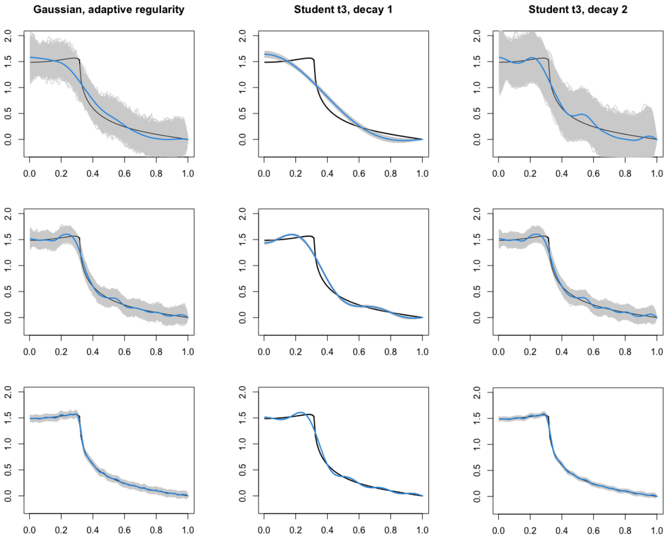
  - Student t_3 oversmoothing prior 1: $\sigma_k = k^{-1/2-S}$, $S = 5$
  - Student t_3 oversmoothing prior 2: $\sigma_k = e^{-(\log k)^{3/2}}$
- Truncate up to $k = 200$
- Gaussian hierarchical: use non-centered Gibbs sampler
- Student: product of univariate problems, use RW Metropolis on each

White Noise Model - Sobolev Truth - Full Posteriors

$n = 10^5$

$n = 10^7$

$n = 10^9$



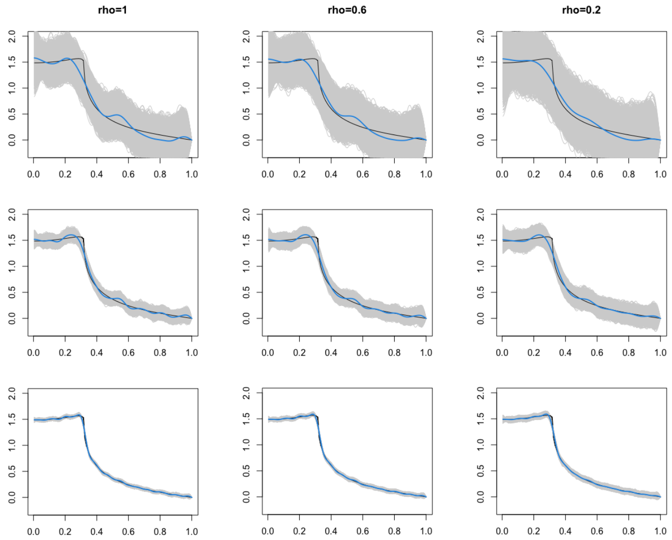
White Noise Model - Sobolev Truth - ρ-posteriors

Student t_3, decay 2

$n = 10^5$

$n = 10^7$

$n = 10^9$



White Noise Model - NMR Signal

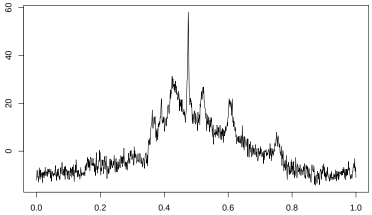
- Denoising (no inversion) NMR signal
- Expand in Symmlet 6 wavelet basis (ψ_{lk}) truncated at $l = 9$

$$X_{lk} \sim \mathcal{N}(f_{lk}, 1/n)$$

- Student t_3 oversmoothing prior on f_{lk} with

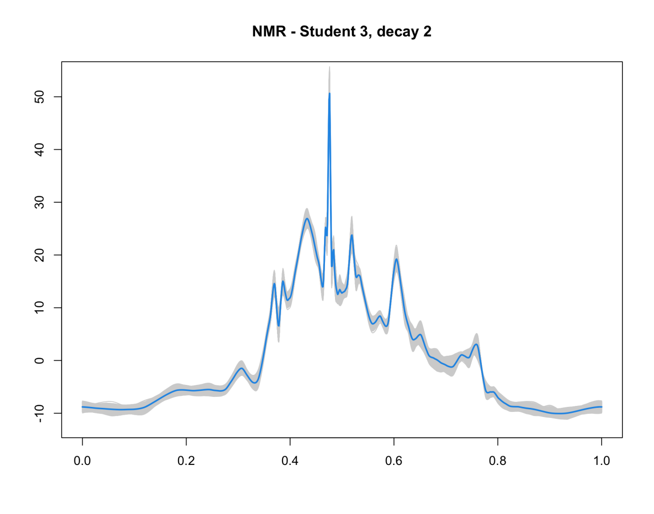
$$s_l = 2^{-l^{3/2}}, \; \forall k \in \mathcal{K}_l$$

- Product of univariate problems, use RW Metropolis to sample each posterior

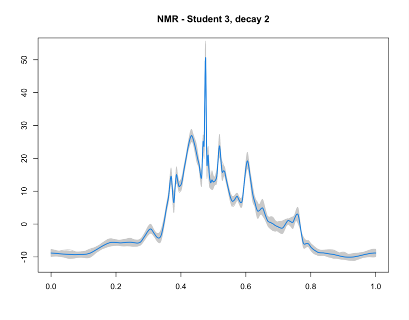


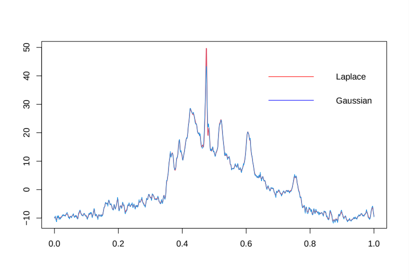
J. Buckheit, S. Chen, D. Donoho, I. Johnstone, and J. Scargle, *Wavelab 850*

White Noise Model - NMR Signal



White Noise Model - NMR Signal





Density Estimation - Full Posteriors

- $X^{(n)} = (X_1, \ldots, X_n)$ where $X_j \overset{iid}{\sim} p(x),\ x \in [0,1]$
- $p : [0,1] \to R^+$ unknown density, modelled as

$$p(x) = \frac{e^{f(x)}}{\int e^{f(x)} dx}$$

- True density p_0 defined via f_0, which has coefficients wrt Symmlet 8 wavelet basis

$$f_{0,lk} = 4\cos^3(2^l + k)2^{-(5/2)l}$$

Hölder-Zygmund regularity $\beta = 2$

Density Estimation - Full Posteriors

- Wavelet priors on f
  - Gaussian oversmoothing prior: $s_l = 2^{-(1/2+S)l}$, $S = 5$
  - Cauchy oversmoothing prior 1: $s_l = 2^{-(1/2+S)l}$, $S = 5$
  - Cauchy oversmoothing prior 2: $s_l = 2^{-l^{3/2}}$
- Sampled posterior using Whitened Precondition Crank-Nicolson algorithm, based on orthogonal transformation for Cauchy

  V. Chen, M. Dunlop, O. Papaspiliopoulos and A. Stuart, *Dimension robust MCMC in Bayesian inverse problems*, arXiv:1803.03344

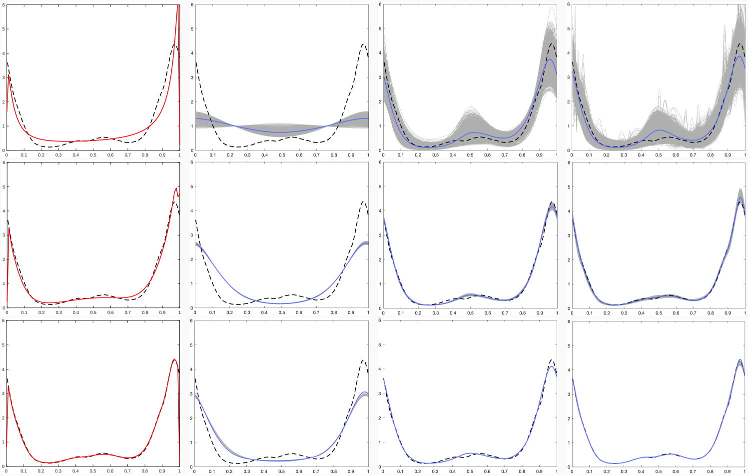
Density Estimation - Full Posteriors

KDE | Gaussian | Cauchy 1 | Cauchy 2

$n = 10^2$

$n = 10^4$

$n = 10^6$



Density Estimation - ρ-posteriors

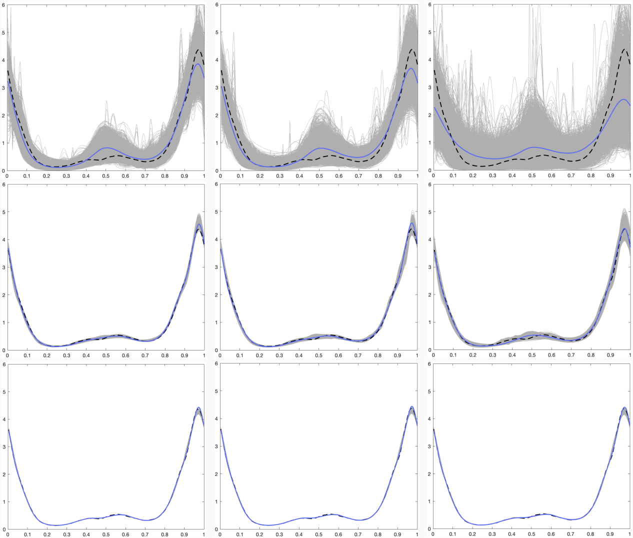
Cauchy prior 2

$\rho = 1$ | $\rho = 0.6$ | $\rho = 0.2$

$n = 10^2$

$n = 10^4$

$n = 10^6$



Conclusion

Outlook

Summary:

- Adaptivity with minimal/no tuning with heavy tailed priors
- Posterior contraction results for WNM
- Generic prior mass condition can give ρ-posterior contraction results for general models
- Results in L^2 and L^∞ losses, for Sobolev, Hölder and Besov truths
- Promising simulations, despite multimodal posteriors

Still to do:

- Uncertainty quantification
- Inverse problems
- Posterior contraction for general models
- Computation

Thank you!