BNP NETWORK

NICOSIA

april 2022

Nonparametric Smoothing

Bayesian Nonparametric Smoothing

f a function

$f \sim \Pi$

$X \mid f \sim P_f$

$\Pi(f \in \cdot \mid X)$

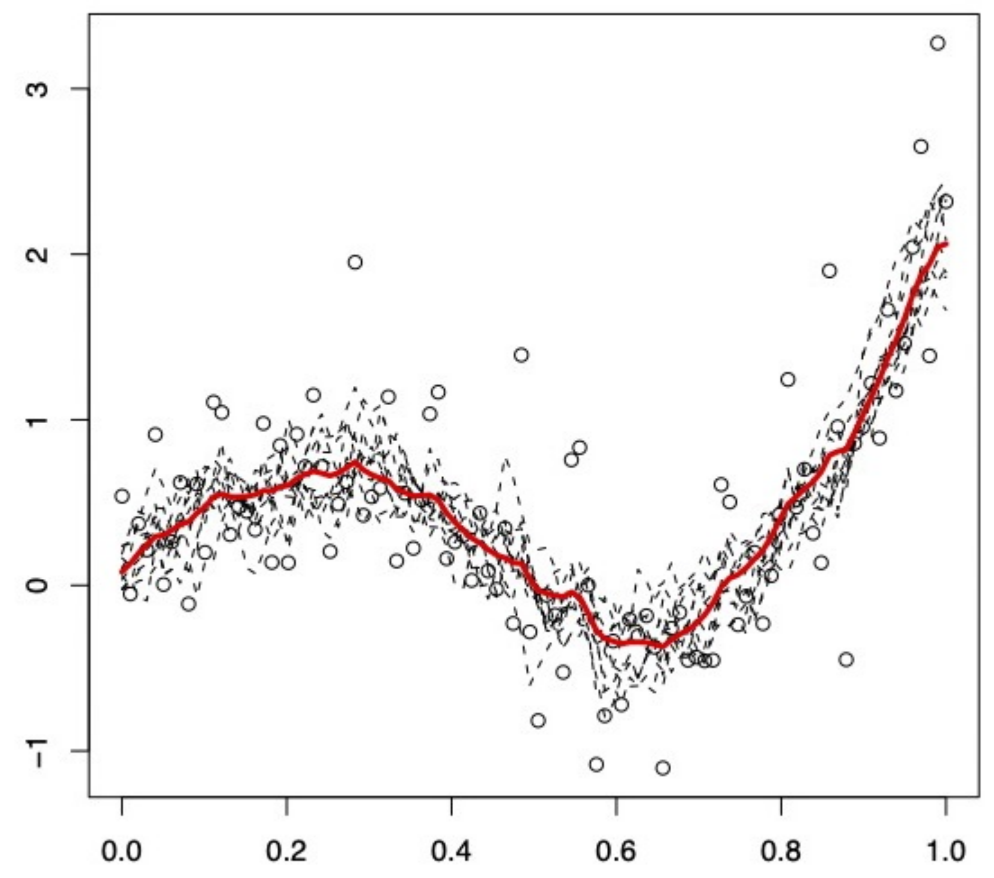
posterior mean/mode/median recovers true function

spread in $\Pi(f \in \cdot \mid X)$ should show confidence in this recovery

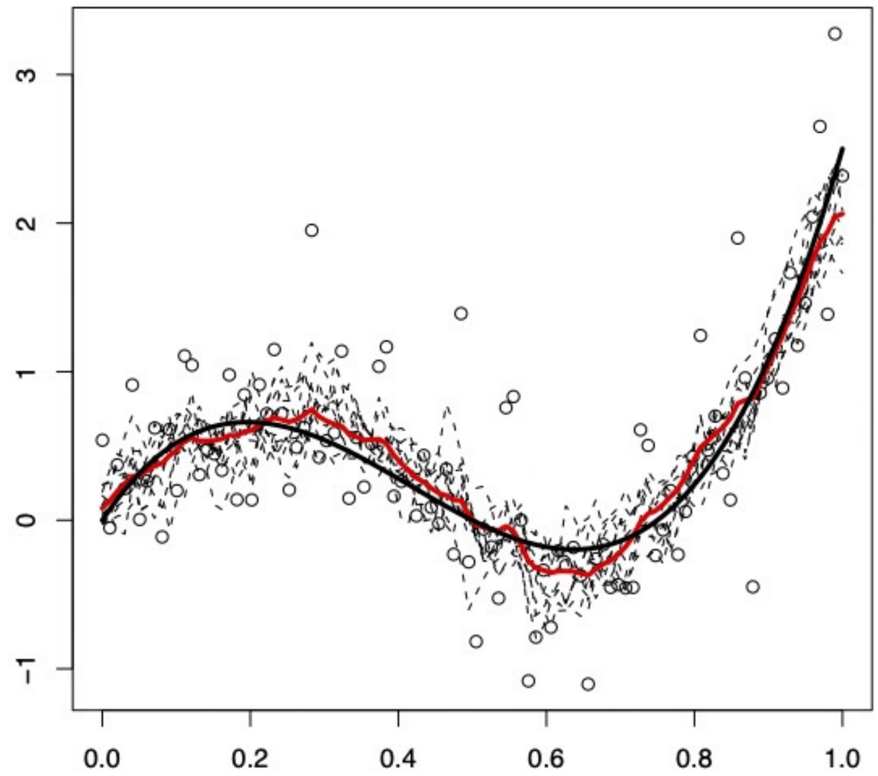
"Smoothing"

$\rightarrow$ bias-variance trade-off (prior "regularises")

$\rightarrow$ no $\sqrt{n}$-rate



3
2
1
0
−1
0.0
0.2
0.4
0.6
0.8
1.0



3
2
1
0
−1
0.0
0.2
0.4
0.6
0.8
1.0

History

Wahba, 1975

J. R. Statist. Soc. B (1983),
45, *No.* 1, *pp.* 133–150

Bayesian "Confidence Intervals" for the Cross-validated Smoothing Spline

By GRACE WAHBA

University of Wisconsin, USA

[Received August 1981. Revised August 1982]

SUMMARY

We consider the model $Y(t_i) = g(t_i) + \epsilon_i$, $i = 1, 2, \ldots, n$, where $g(t)$, $t \in [0, 1]$ is a smooth function and the $\{\epsilon_i\}$ are independent $N(0, \sigma^2)$ errors with σ^2 unknown. The cross-validated smoothing spline can be used to estimate g non-parametrically from observations on $Y(t_i)$, $i = 1, 2, \ldots, n$, and the purpose of this paper is to study confidence intervals for this estimate. Properties of smoothing splines as Bayes estimates are used to derive confidence intervals based on the posterior covariance function of the estimate. A small Monte Carlo study with the cubic smoothing spline is carried out to suggest by example to what extent the resulting 95 per cent confidence intervals can be expected to cover about 95 per cent of the true (but in practice unknown) values of $g(t_i)$, $i = 1, 2, \ldots, n$. The method was also applied to one example of a two-dimensional thin plate smoothing spline. An asymptotic theoretical argument is presented to explain why the method can be expected to work on fixed smooth functions (like those tried), which are "smoother" than the sample functions from the prior distributions on which the confidence interval theory is based.

Keywords: SPLINE SMOOTHING; CROSS-VALIDATION; CONFIDENCE INTERVALS

1. INTRODUCTION

Consider the model

$$Y(t_i) = g(t_i) + \epsilon_i, \quad i = 1, 2, \ldots, n, \quad t_i \in [0, 1], \tag{1.1}$$

where $\epsilon = (\epsilon_1, \ldots, \epsilon_n)' \sim N(0, \sigma^2 I_{n \times n})$, σ^2 is unknown and $g(\cdot)$ is a fixed but unknown function with $m-1$ continuous derivatives and $\int_0^1 (g^{(m)}(t))^2 dt < \infty$. The smoothing spline estimate of g given $Y(t_i) = y_i$, $i = 1, 2, \ldots, n$, which we will call $g_{n,\lambda}$, is the minimizer of

$$n^{-1} \sum_{i=1}^{n} (g(t_i) - y_i)^2 + \lambda \int_0^1 (g^{(m)}(t))^2 dt$$

Works great!

Cox, 1993

The Annals of Statistics
1993, Vol. 21, No. 2, 903–923

AN ANALYSIS OF BAYESIAN INFERENCE FOR NONPARAMETRIC REGRESSION[1]

BY DENNIS D. COX

Rice University

The observation model $y_i = \beta(i/n) + \varepsilon_i$, $1 \le i \le n$, is considered, where the ε's are i.i.d. with mean zero and variance σ^2 and β is an unknown smooth function. A Gaussian prior distribution is specified by assuming β is the solution of a high order stochastic differential equation. The estimation error $\delta = \beta - \hat{\beta}$ is analyzed, where $\hat{\beta}$ is the posterior expectation of β. Asymptotic posterior and sampling distributional approximations are given for $\|\delta\|^2$ when $\|\cdot\|$ is one of a family of norms natural to the problem. It is shown that the frequentist coverage probability of a variety of $(1-\alpha)$ posterior probability regions tends to be larger than $1-\alpha$, but will be infinitely often less than any $\varepsilon > 0$ as $n \to \infty$ with prior probability 1. A related continuous time signal estimation problem is also studied.

1. Introduction. In this article we consider Bayesian inference for a class of nonparametric regression models. Suppose we observe

$$Y_{ni} = \beta(t_{ni}) + \varepsilon_i, \qquad 1 \le i \le n, \tag{1.1}$$

where $t_{ni} = i/n$, $\beta: [0, 1] \to \mathbb{R}$ is an unknown smooth function, and $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. random errors with mean 0 and known variance $\sigma^2 < \infty$. The ε_i are modeled as $N(0, \sigma^2)$. A Gaussian prior for β will now be specified. Let $m \ge 2$ and for some constants $a_0, \ldots, a_m$ with $a_m \ne 0$ let

$$L = \sum_{i=0}^{m} a_i D^i$$

Fails miserably!

(penalised least squares with splines
= posterior mode of IBM prior)

White noise model

Observe $X = f + \frac{1}{\sqrt{n}} \dot{W}$ Gaussian white noise

Equivalent

Observe $(X_1, X_2, \cdots)$, $X_i \overset{ind}{\sim} N(f_i, \frac{1}{n})$, $f = \sum_{i=1}^{\infty} f_i e_i$

$(e_1, e_2 \cdots)$ orthonormal . $\|f\|^2 = \int_0^1 |f^2(t)| dt = \sum_{i=1}^{\infty} f_i^2$

$f \in L_2[0,1] \iff (f_1, f_2, \cdots) \in \ell_2$

White noise model

Observe $X_n = f + \frac{1}{\sqrt{n}} \dot{W}$ Gaussian white noise

Equivalent

Observe $(X_{n,1}, X_{n,2}, \dots)$, $X_{n,i} \overset{ind}{\sim} N(f_i, \frac{1}{n})$, $f = \sum_{i=1}^{\infty} f_i e_i$

$(e_1, e_2 \dots)$ orthonormal. $\|f\|^2 = \int_0^1 |f^2(t)| dt = \sum_{i=1}^{\infty} f_i^2$

$f \in L_2[0,1] \iff (f_1, f_2 \dots) \in \ell_2$

Inverse problem $X_n = Kf + \frac{1}{\sqrt{n}} \dot{W}$

Observe $(X_{n,1}, X_{n,2}, \dots)$, $X_{n,i} \overset{ind}{\sim} N(\kappa_i f_i, \frac{1}{n})$, $f = \sum_{i=1}^{\infty} f_i e_i$

eigenvalues of K

Can also project on wavelet basis, spline basis, ...

Smoothness

$$\ell_2 = \left\{ (\beta_1, \beta_2, \dots) : \sum_{i=1}^{\infty} \beta_i^2 < \infty \right\}$$

$$H^{\alpha} = \left\{ (\beta_1, \beta_2, \dots) : \sum_{i=1}^{\infty} i^{2\alpha} \beta_i^2 < \infty \right\}$$

If $e_1, e_2, \dots$ are Fourier basis, then H^{α} are roughly the functions that are α-times differentiable.

Smoothness

$$\ell_2 = \left\{ (f_1, f_2, \ldots) : \sum_{i=1}^{\infty} f_i^2 < \infty \right\}$$

$$H^\alpha = \left\{ (f_1, f_2, \ldots) : \sum_{i=1}^{\infty} i^{2\alpha} f_i^2 < \infty \right\}$$

If $e_1, e_2 \ldots$ are Fourier basis, then H^α are roughly the functions that are α-times differentiable.

LEM $f_i \overset{\text{ind}}{\sim} N\left(0, \frac{1}{i^{2\alpha+1}}\right) \Rightarrow (f_1, f_2 \ldots) \in H^\beta$ a.s. $\forall \beta < \alpha$.

Smoothness

$$\ell_2 = \left\{ (f_1, f_2, \dots) : \sum_{i=1}^{\infty} f_i^2 < \infty \right\}$$

$$H^\alpha = \left\{ (f_1, f_2, \dots) : \sum_{i=1}^{\infty} i^{2\alpha} f_i^2 < \infty \right\}$$

If $e_1, e_2 \dots$ are Fourier basis, then H^α are roughly the functions that are α-times differentiable.

LEM $f_i \overset{\text{ind}}{\sim} N\left(0, \frac{1}{i^{2\alpha+1}}\right) \Rightarrow (f_1, f_2 \dots) \in H^\beta$ a.s. $\forall \beta < \alpha$.

"proof" $E \sum_{i=1}^{\infty} i^{2\beta} f_i^2 = \sum_{i=1}^{\infty} i^{2\beta} \frac{1}{i^{2\alpha+1}} < \infty \iff \alpha > \beta$.

+ a thm on a.s. convergence. $\square$

$\rightarrow$ "prior of roughly smoothness α".

Bernstein–von Mises in ℓ_2

(Freedman, Leahu)

$$f_i \overset{ind}{\sim} N(0, \tau_i^2)$$

$$X_{n,i} \mid f_i \overset{ind}{\sim} N\left(f_i, \frac{1}{n}\right)$$

THM

$$\left\| \mathcal{L}\left(\theta - E(\theta \mid X_n) \mid X_n \right) - \mathcal{L}_{\theta_0}\left(E(\theta \mid X_n) - \theta_0 \right) \right\|_{TV} \xrightarrow{P_{\theta_0}} 0$$

iff

$$\sum_{i=1}^{\infty} \frac{1}{\tau_i^4} < \infty \quad \text{and} \quad \sum_{i=1}^{\infty} \frac{\theta_{0,i}}{\tau_i^2} < \infty$$

Bernstein-von Mises in ℓ_2

(Freedman, Leahu)

$$f_i \overset{ind}{\sim} N(0, \tau_i^2)$$

$$X_{n,i} \mid f_i \overset{ind}{\sim} N(f_i, \tfrac{1}{n})$$

THM

$$\| \mathcal{L}(\theta - E(\theta|X_n) \mid X_n) - \mathcal{L}_{\theta_0}(E(\theta|X_n) - \theta_0) \|_{TV} \xrightarrow{P_{\theta_0}} 0$$

iff

$$\sum_{i=1}^{\infty} \frac{1}{\tau_i^4} < \infty \quad \text{and} \quad \sum_{i=1}^{\infty} \frac{\theta_{0,i}}{\tau_i^2} < \infty$$

So BvM holds only if $\tau_i \to \infty$, fast. USELESS

ℓ_2-Contraction Rate

$$f_i \overset{\text{ind}}{\sim} N\left(0, \frac{1}{i^{1+2\alpha}}\right)$$

$$X_{n,i} \mid f_i \sim N\left(f_i, \frac{1}{n}\right)$$

THM If $f_0 \in H^\beta$, then

$$\Pi\left(f : \|f - f_0\| \gtrsim n^{-\frac{\alpha \wedge \beta}{2\alpha+1}} \mid X_n\right) \xrightarrow{P_{f_0}} 0.$$

ℓ_2-Contraction Rate

$f_i \overset{ind}{\sim} N\left(0, \frac{1}{i^{1+2\alpha}}\right)$

$X_{n,i} \mid f_i \sim N\left(f_i, \frac{1}{n}\right)$

$\sum_{i=1}^{\infty} i^{2\beta} f_{0,i}^2 < \infty$

THM If $f_0 \in H^\beta$, then

$$\Pi\left(f : \|f - f_0\| \gtrsim n^{-\frac{\alpha \wedge \beta}{2\alpha+1}} \mid X_n\right) \xrightarrow{P_{f_0}} 0.$$

ℓ_2-Contraction Rate

$f_i \overset{ind}{\sim} N\left(0, \frac{1}{i^{1+2\alpha}}\right)$

$X_{n,i} \mid f_i \sim N\left(f_i, \frac{1}{n}\right)$

$\sum_{i=1}^{\infty} i^{2\beta} f_{0,i}^2 < \infty$

THM If $f_0 \in H^\beta$, then

$$\Pi\left(f : \|f - f_0\| \gtrsim n^{-\frac{\alpha \wedge \beta}{2\alpha+1}} \mid X_n\right) \xrightarrow{P_{f_0}} 0.$$

- best rate $n^{-\frac{\beta}{1+2\beta}}$ obtained iff $\alpha = \beta$: prior matches truth
- consistency $\forall \alpha$
- $\forall \beta > \alpha$ rate $n^{-\frac{\alpha}{2\alpha+1}}$.

DISAPPOINTING

Posterior Spread

$$f_i \overset{ind}{\sim} N\left(0, \frac{1}{i^{1+2\alpha}}\right)$$

$$X_{n,i} \mid f_i \sim N\left(f_i, \frac{1}{n}\right) \qquad i = 1, 2, \ldots$$

actually nonrandom

THM

If $\alpha < \beta$ $\quad E\left(\|f - E(f|X_n)\|^2 \mid X_n\right) \overset{P}{\gg} \|E(f|X_n) - f_0\|^2$, all $f_0 \in H^\beta$

If $\alpha > \beta$ $\quad E\left(\|f - E(f|X_n)\|^2 \mid X_n\right) \underset{P}{\ll} \|E(f|X_n) - f_0\|^2$, some $f_0 \in H^\beta$

Posterior Spread

$f_i \overset{ind}{\sim} N\left(0, \frac{1}{i^{1+2\alpha}}\right)$

$X_{n,i} \mid f_i \sim N\left(f_i, \frac{1}{n}\right) \quad i = 1, 2, \dots$

actually nonrandom

THM

If $\alpha < \beta$ $\quad E\left(\|f - E(f \mid X_n)\|^2 \mid X_n\right) \overset{P}{\gg} \|E(f \mid X_n) - f_0\|^2$, $\quad$ all $f_0 \in H^\beta$

If $\alpha > \beta$ $\quad E\left(\|f - E(f \mid X_n)\|^2 \mid X_n\right) \underset{P}{\ll} \|E(f \mid X_n) - f_0\|^2$, $\quad$ some $f_0 \in H^\beta$

proof $\quad f_i - f_{0,i} \mid X_n \sim N\left(\frac{f_{0,i}}{1 + n i^{-1-2\alpha}}, \frac{i^{-1-2\alpha}}{1 + n i^{-1-2\alpha}}\right)$

$$E\|E(f \mid X_n) - f_0\|^2 = \sum_i \frac{f_{0i}^2}{(1 + n i^{-1-2\alpha})^2} + \sum_i \frac{n i^{-2-4\alpha}}{(1 + n i^{-1-2\alpha})^2}$$

$$E\left(\|f - E(f \mid X_n)\|^2 \mid X_n\right) = \sum_i \frac{i^{-1-2\alpha}}{1 + n i^{-1-2\alpha}}$$

standard deviations are smaller.

$\square$

Posterior Spread

$f_i \overset{\text{ind}}{\sim} N(0, i^{-1-2\alpha})$

$X_{n,i} \mid f_i \sim N(f_i, \frac{1}{n})$

THM

If $\alpha < \beta$ $\quad E\left(\|f - f_0\|^2 \mid X_n\right) \overset{P}{\gg} \|E(f \mid X_n) - f_0\|^2$, $\quad$ all $f_0 \in H^\beta$

If $\alpha > \beta$ $\quad E\left(\|f - f_0\|^2 \mid X_n\right) \underset{P}{\ll} \|E(f \mid X_n) - f_0\|^2$, $\quad$ some $f_0 \in H^\beta$

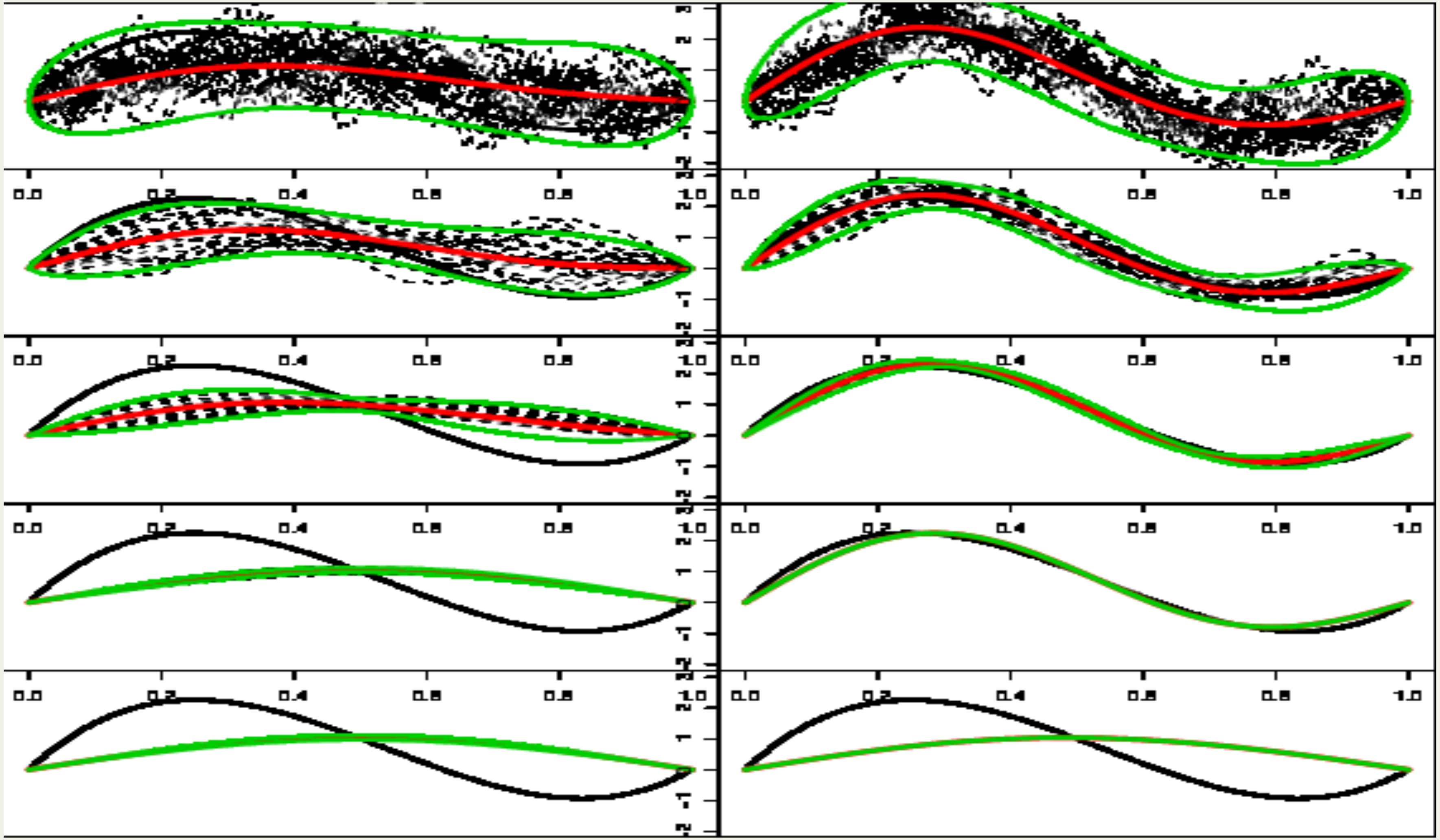
$C_n^\gamma := \{ f : \|f - E(f \mid X_n)\| < R_n^\gamma \}$ for $\Pi(C_n^\gamma \mid X_n) = 1 - \gamma$.

COR

if $\alpha < \beta$ $\quad P_{f_0}(f_0 \in C_n^\gamma) \to 1$, $\quad$ all $f_0 \in H^\beta$

if $\alpha > \beta$ $\quad P_{f_0}(f_0 \in C_n^\gamma) \to 0$, $\quad$ some $f_0 \in H^\beta$

Example: heat equation (n=10 000, n=100 000 000)



True θ_0 (black), posterior mean (red), 20 realizations from the posterior (dashed black), and posterior credible bands (green).
Left: $n = 10^4$; right: $n = 10^8$. Top to bottom: prior of increasing smoothness.

[Knapik, vdV and Van Zanten, 2013.]

with priors of fixed regularity

- get suboptimal recovery if prior regularity $\neq$ true regularity
- get overly optimistic uncertainty quantification if prior regularity $>$ true regularity

Adaptive Priors

As "true regularity" will be unknown,
it is preferable to tune a prior to the data.

We can use a family of priors indexed by one or more hyperparameters*, which are estimated from the data (empirical Bayes) or receive a hyper prior (hierarchical Bayes).

* length scale of Gaussian process, number of basis functions, rate of decrease of variances, bandwidth.

Adaptive Priors

$$f_i \overset{ind}{\sim} N\left(0, \frac{1}{i^{1+2\alpha}}\right)$$

$$X_{n,i} \mid f_i \sim N\left(f_i, \frac{1}{n}\right)$$

$$\log \int p(X_n \mid f)\, d\Pi_\alpha(f)$$

Empirical Bayes $\hat{\alpha}_n = \underset{\alpha}{\operatorname{argmax}} \left(\sum_{i=1}^{\infty} \frac{n^2 X_{ni}^2}{i^{1+2\alpha} + n} - \log\left(1 + \frac{n}{i^{1+2\alpha}}\right)\right)$

Hierarchical Bayes $\frac{1}{\alpha^r} \sim \Gamma(a, b)$

ℓ_2-Contraction Rate with adaptation

$$f_i \overset{ind}{\sim} N\left(0, \frac{1}{i^{1+2\alpha}}\right)$$

$$X_{n,i} \mid f_i \sim N\left(f_i, \frac{1}{n}\right)$$

plug in $\hat{\alpha}_n$ or hierarchical posterior

THM If $f_0 \in H^\beta$

$$\Pi\left(f : \|f - f_0\| \gtrsim n^{-\frac{\beta}{2\beta+1}} \mid X_n\right) \xrightarrow{P_{f_0}} 0.$$

- best rate $n^{-\frac{\beta}{2\beta+1}}$ obtained $\forall \beta$

For hierarchical Bayes there is general theory.
Empirical Bayes is harder: $\hat{\alpha}_n$ does not necessarily settle down.

Uncertainty Quantification with Adaptive Priors

Uncertainty Quantification works

Uncertainty Quantification with Adaptive Priors

Uncertainty Quantification works
but only for nice f_0.

Posterior may be tricked into thinking
f_0 is smoother than it is.

ℓ_2 - Credible Balls - positive result (Szabo et al)

$$f_i \overset{ind}{\sim} N\left(0, \frac{1}{i^{1+2\alpha}}\right)$$

$$X_{n,i} \mid f_i \sim N\left(f_i, \frac{1}{n}\right)$$

correction

$$C_n^\gamma := \{ f : \| f - E(f \mid X_n) \| \leq R_n^\gamma \, c \}, \quad \Pi(C_n^\gamma \mid X_n) = 1-\gamma.$$

ℓ_2 - Credible Balls - positive result (Szabo et al)

$$f_i \overset{ind}{\sim} N\left(0, \frac{1}{i^{1+2\alpha}}\right)$$

$$X_{n,i} \mid f_i \sim N\left(f_i, \frac{1}{n}\right)$$

correction

$$C_n^{\gamma} := \{ f : \| f - E(f \mid X_n) \| \le R_n^{\gamma} c \} , \quad \Pi(C_n^{\gamma} \mid X_n) = 1-\gamma .$$

DEF f is polished tail if $\exists$ L, ρ

$$\sum_{i=N}^{\infty} f_i^2 \le L^2 \sum_{i=N}^{\rho N} f_i^2 \qquad \forall \text{ large } N$$

"polished tail" is a weak form of "self-similarity"!

ℓ_2 - Credible Balls - positive result (Szabo et al)

$f_i \overset{\text{ind}}{\sim} N\left(0, \frac{1}{i^{1+2\alpha}}\right)$

$X_{n,i} \mid f_i \sim N\left(f_i, \frac{1}{n}\right)$

$C_n^\gamma := \{ f : \| f - E(f \mid X_n) \| \le R_n^\gamma \, c \}$, $\Pi(C_n^\gamma \mid X_n) = 1-\gamma.$

correction

DEF f is polished tail if $\exists$ L, ρ

$$\sum_{i=N}^{\infty} f_i^2 \le L^2 \sum_{i=N}^{\rho N} f_i^2 \qquad \forall \text{ large } N$$

THM $P_{f_0}(f_0 \in C_n^\gamma) \to 1$ for all polished tail f_0

if $c = c(L, \rho)$ big enough.

"polished tail" is a weak form of "self-similarity".

“Everything” is polished tail..

For the *topologist*:

THEOREM [Giné+Nickl, 2010]
Non self-similar sequences are meagre relative to a natural topology.

For the *minimax expert*:

THEOREM
By intersecting a model with the polished tail sequences the minimax risk decreases by at most

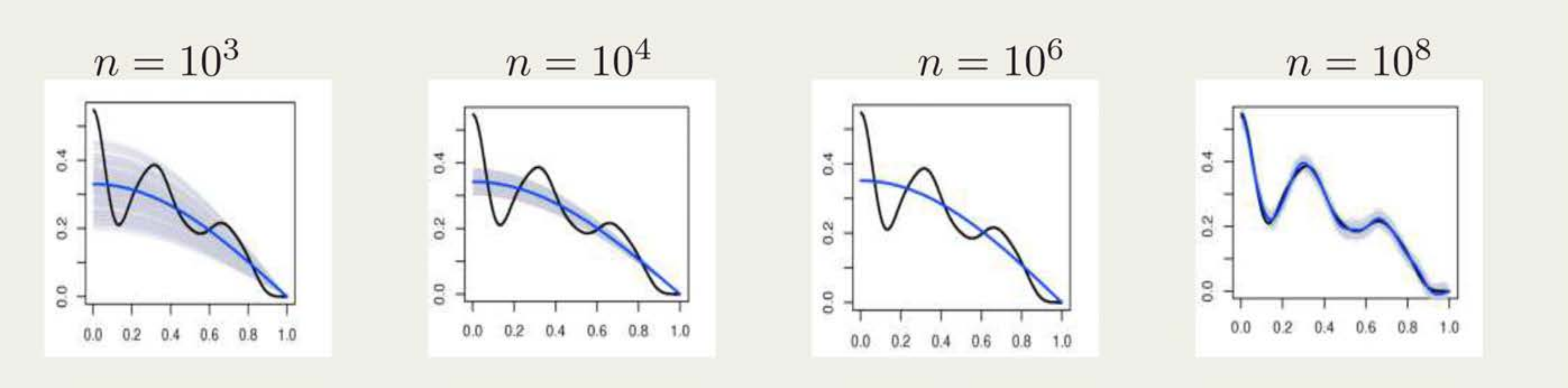
- a constant if the model is a hyperrectangle.
- a logarithmic factor if the model is a Sobolev ball.

For the *Bayesian*:

THEOREM
For every $\alpha > 0$ the prior $\Pi_\alpha = N_\infty(0, \Lambda)$ with $\lambda_i \sim i^{-1-2\alpha}$ satisfies

$$\Pi_\alpha\Big(\cup_{N_0}\{\theta: \theta \in \text{polished tail}(2^{2+2\alpha}, N_0, 2)\}\Big) = 1.$$

Not "Everything" is Polished Tail



Volterra inverse problem, adaptive prior

black : truth

blue : posterior mean

grey : draws from posterior

Not "Everything" is Polished Tail

Counter examples are functions such that $(f_1, f_2, \dots)$ has gaps of zeros of increasing length.

$$f_1, f_2, \dots, f_{n_1}, 0, 0, \dots 0, f_{n_2}, f_{n_2+1}, \dots, f_{n_2}, 0, 0, \dots \dots, 0, f_{n_3}, \dots f_{n_3}, 0, \dots$$

A posterior "knows" that it can only estimate $f_1, \dots, f_{N_n}$, some N_n?
If N_n falls at the end of a gap, then it will oversmooth

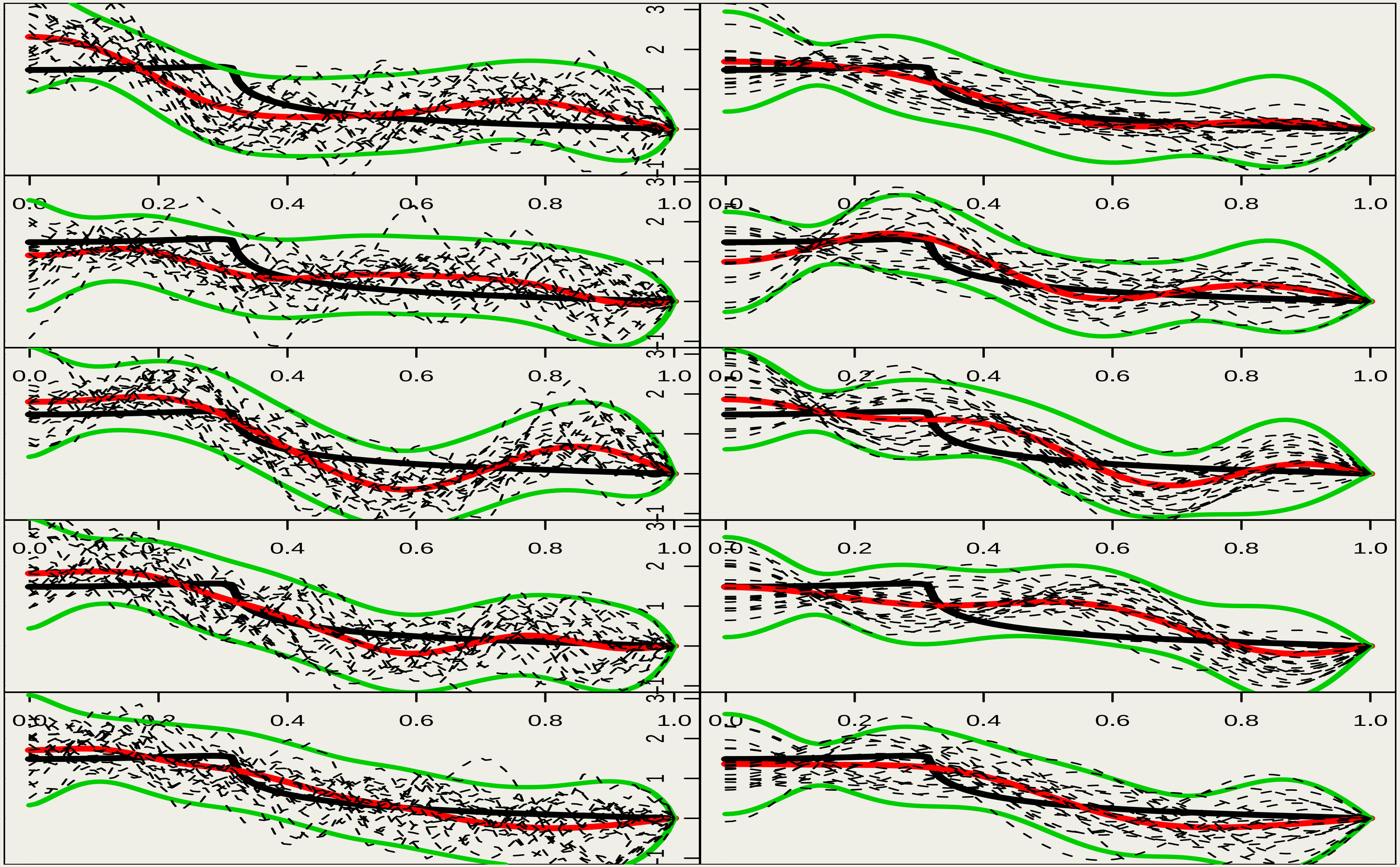
choose big α

EXAMPLE $n_1 \geq 2$, $n_j > n_{j-1}^4$, $j = 2, 3, \dots$

$$f_j = \begin{cases} \frac{1}{n_j} & n_j^{\frac{1}{1+2\beta}} < j < 2n_j^{\frac{1}{1+2\beta}}, \quad j = 1, 2, \dots \\ 0 & \text{otherwise} \end{cases}$$

Then $\sum_{j=1}^{\infty} j^{2\beta} f_j^2 \lesssim 1$ but $P_{f_0}\left(\|f_0 - E(f \mid X_n)\| < \epsilon_n^{\delta} L_n \right) \to 0$
for $L_n \ll n^{\delta}$

Example: reconstruct derivative (n=1000)



True θ_0 (black), posterior mean (red), and 20 realizations from the posterior, repeated 5 times for a rescaled rough prior (left) and a rescaled smooth prior (right).

Other Models

Difficulty in deriving theoretical results is in analysing the empirical Bayes estimator $\hat{\alpha}_n$.

- General setup : Rousseau & Szabó
- Gaussian regression : Sniekers
- Double exponential prior : Hadji & Szabó
- Deep learning : Franssen & Szabó

⋮

Definition of "polished tail" to be adapted to prior + setting.

Dirichlet mixtures for density estimation ??

Weak Bernstein-von Mises

- Castillo & Nickl 2013, 2014
- Ray 2016 (adaptive)
- Nickl 2018
- Nickl et al. 2020

Weak BvM in H^{-s}

(Castillo & Nickl, Ray 2017, Nickl 2018)

$e_1, e_2, \ldots$ orthonormal

$$\ell_2 = \left\{ f = \sum_{j=1}^{\infty} f_j e_j \, , \ \|f\|^2 = \sum_{j=1}^{\infty} f_j^2 < \infty \right\}.$$

$$H^{-s} = \left\{ f = \sum_{j=1}^{\infty} f_j e_j \, , \ \|f\|_{-s}^2 := \sum_{j=1}^{\infty} \left(\frac{1}{j}\right)^{2s} f_j^2 < \infty \right\}$$

Weak BvM in H^{-s}

(Castillo & Nickl, Ray 2017, Nickl 2018)

$e_1, e_2, \dots$ orthonormal

$$\ell_2 = \{ f = \sum_{j=1}^{\infty} f_j e_j \, , \; \|f\|^2 = \sum_{j=1}^{\infty} f_j^2 < \infty \}.$$

$$H^{-s} = \{ f = \sum_{j=1}^{\infty} f_j e_j \, , \; \|f\|_{-s}^2 := \sum_{j=1}^{\infty} \left(\tfrac{1}{j}\right)^{2s} f_j^2 < \infty \}$$

LEM Let $s > 1/2$.
If $Z_j \overset{\text{ind}}{\sim} N(0,1)$, then $\sum_{j=1}^{\infty} Z_j e_j$ is a tight variable in H^{-s}.

proof $\quad E \| \sum_{j=1}^{\infty} Z_j e_j \|_{-s}^2 = E \sum_{j=1}^{\infty} \left(\tfrac{1}{j}\right)^{2s} Z_j^2 = \sum_{j=1}^{\infty} \left(\tfrac{1}{j}\right)^{2s}$

+ a thm on a.s. convergence. $\square$

Weak BvM in H^{-s}

(Castillo & Nickl, Ray 2017, Nickl 2018)

$e_1, e_2, \dots$ orthonormal

$\ell_2 = \{ f = \sum_{j=1}^{\infty} f_j e_j \,,\ \|f\|^2 = \sum_{j=1}^{\infty} f_j^2 < \infty \}$.

$H^{-s} = \{ f = \sum_{j=1}^{\infty} f_j e_j \,,\ \|f\|_{-s}^2 := \sum_{j=1}^{\infty} \left(\frac{1}{j}\right)^{2s} f_j^2 < \infty \}$

LEM Let $s > 1/2$.
If $Z_j \overset{\text{ind}}{\sim} N(0,1)$, then $\sum_{j=1}^{\infty} Z_j e_j$ is a tight variable in H^{-s}.

proof $E \| \sum_{j=1}^{\infty} Z_j e_j \|_{-s}^2 = E \sum_{j=1}^{\infty} \left(\frac{1}{j}\right)^{2s} Z_j^2 = \sum_{j=1}^{\infty} \left(\frac{1}{j}\right)^{2s}$

+ a thm on a.s. convergence. $\square$

$Z = (Z_1, Z_2, \dots)$ is a standard normal variable in H^{-s}.

A posterior can "converge" to it.

Weak Convergence

$(\mathbb{D}, d)$ metric space

$X_n \rightsquigarrow X$ iff $E h(X_n) \to E h(X)$ $\forall\, h : \mathbb{D} \to [0,1]$, continuous

Weak Convergence

$(\mathbb{D}, d)$ metric space

$X_n \rightsquigarrow X$ iff $Eh(X_n) \to Eh(X)$ $\forall\, h: \mathbb{D} \to [0,1]$, continuous

For separable $(\mathbb{D}, d)$ equivalent to convergence of laws in the bounded-Lipschitz metric

$$d_{BL}\big(\mathcal{L}(X_n), \mathcal{L}(X)\big) = \sup_{h \in \mathrm{Lip}} \left| Eh(X_n) - Eh(X) \right|$$

$|h(x) - h(y)| \le d(x,y)$ $\forall\, x, y$

$|h(x)| \le 1$ $\forall\, x$.

The stronger the metric d, the stronger $X_n \rightsquigarrow X$

BvM in H^{-s}

(Ray 2017)

$$f = \sum_{j=1}^{\infty} f_j e_j \,, \qquad f_j \overset{\text{ind}}{\sim} N(0, j^{-1-2\alpha})$$

$$X_n = f + \frac{1}{\sqrt{n}} \dot{W} \,, \qquad \dot{W} = \sum_{j=1}^{\infty} Z_j f_j \equiv Z$$

(standard normal in H^{-s})

Choose α by empirical or hierarchical Bayes.

BvM in H^{-s}

(Ray 2017)

$$f = \sum_{j=1}^{\infty} f_j e_j \,, \qquad f_j \overset{ind}{\sim} N(0, j^{-1-2\alpha})$$

$$X_n = f + \frac{1}{\sqrt{n}} \dot{W} \,, \qquad \dot{W} = \sum_{j=1}^{\infty} Z_j f_j \equiv Z$$ (standard normal in H^{-s})

Choose α by empirical or hierarchical Bayes.

THM If $s > 1/2$ and $\sup_j j^{2\beta+1} f_{0,j}^2 < \infty$, then

$$d_{BL}\left(\Pi(\sqrt{n}(f - X_n) \mid X_n), \mathcal{L}(Z) \right) \xrightarrow{P_{f_0}} 0 \quad \text{in } H^{-s}.$$

Similar results for wavelets.

BvM in H^{-s}

(Ray 2017)

$f = \sum_{j=1}^{\infty} f_j e_j$, $\quad f_j \overset{ind}{\sim} N(0, j^{-1-2\alpha})$

$X_n = f + \frac{1}{\sqrt{n}} \dot{W}$, $\quad \dot{W} = \sum_{j=1}^{\infty} z_j f_j \equiv Z$ (standard normal in H^{-s})

Choose α by empirical or hierarchical Bayes.

THM If $s > 1/2$ and $\sup_j j^{2\beta+1} f_{0,j}^2 < \infty$, then

$$d_{BL}\left(\Pi(\sqrt{n}(f - X_n) \mid X_n), \mathcal{L}(Z)\right) \xrightarrow{P_{f_0}} 0 \quad \text{in } H^{-s}.$$

$C_n^\gamma := \{ f : \|f - X_n\|_{-s} \leq R_n^\gamma \}$, $\quad \Pi(f \in C_n^\gamma \mid X_n) = 1-\gamma$

COR $P_{f_0}(f_0 \in C_n^\gamma) \to 1-\gamma$

Exact coverage!

But unusual (and large) set.

Adaptive confidence set

(Ray 2017)

$$\tilde{C}_n^{\gamma} := \left\{ f : \|f - X_n\|_{-s} < R_n^{\gamma},\ \|f - \hat{f}_n\|_{\hat{\alpha}_n - \varepsilon_n} \le C\sqrt{\log n} \right\}$$

$\frac{1}{\log n}$

big

EB or median(FB)

Adaptive confidence set

(Ray, 2017)

$$\tilde{C}_n^{\gamma} := \left\{ f : \|f - X_n\|_{-s} < R_n^{\gamma},\ \|f - \hat{f}_n\|_{\hat{\alpha}_n - \varepsilon_n} \leq C\sqrt{\log n} \right\}$$

$\frac{1}{\log n}$; big

EB or median(FB)

DEF f_0 is self-similar if $\sup_j j^{2\beta+1} f_j^2 \leq R$ and $\sum_{j=N}^{\rho N} f_j^2 \gtrsim R N^{-2\beta}$

Adaptive confidence set

(Ray, 2017)

$$\tilde{C}_n^{\gamma} := \left\{ f : \|f - X_n\|_{-s} < R_n^{\gamma},\ \|f - \hat{f}_n\|_{\hat{\alpha}_n - \varepsilon_n} \le C\sqrt{\overset{\text{big}}{\log n}} \right\}$$

$\frac{1}{\log n}$

EB or median(FB)

DEF f_0 is self-similar if $\sup_j j^{2\beta+1} f_j^2 \le R$ and $\sum_{j=N}^{\rho N} f_j^2 \gtrsim R N^{-2\beta}$

THM If f_0 self-similar, then

$$P_{f_0}(f_0 \in \hat{C}_n) \to 1-\gamma$$

$$\Pi(f_0 \in \tilde{C}_n \mid X_n) \to 1-\gamma$$

$$\operatorname{diam}_{\ell_2}(\tilde{C}_n) = O_{P_{f_0}}\left(n^{-\frac{\beta}{2\beta+1}} (\log n)^{\cdots}\right)$$

uniformly in f_0, fixed R, M, ρ

Adaptive confidence set

(Ray, 2017)

$$\tilde{C}_n^\gamma := \{ f : \|f - X_n\|_{-s} < R_n^\gamma ,\ \|f - \hat{f}_n\|_{\hat{\alpha}_n - \varepsilon_n} \le C\sqrt{\log n} \}$$

$\frac{1}{\log n}$ big

EB or median (FB)

DEF f_0 is self-similar if $\sup_j j^{2\beta+1} f_j^2 \le R$ and $\sum_{j=N}^{\rho N} f_j^2 \gtrsim R N^{-2\beta}$

THM If f_0 self-similar, then

$$P_{f_0}(f_0 \in \hat{C}_n) \to 1-\gamma$$

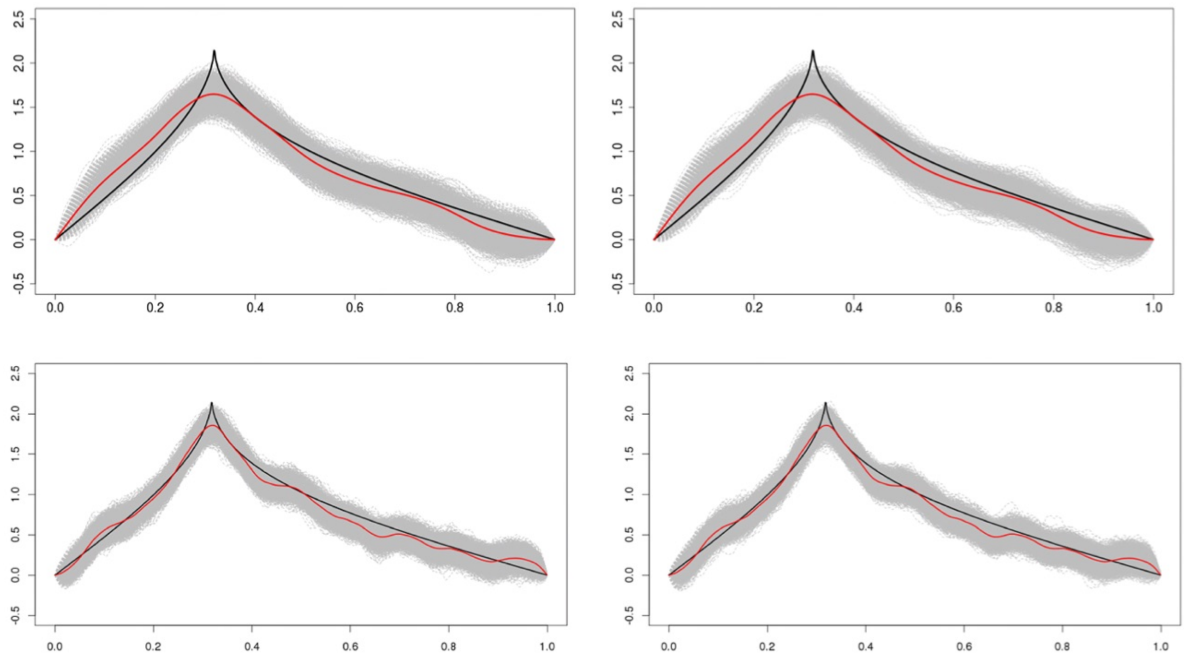
$$\Pi(f_0 \in \tilde{C}_n \mid X_n) \to 1-\gamma$$

$$\mathrm{diam}_{\ell_2}(\hat{C}_n) = O_{P_{f_0}}\left(n^{-\frac{\beta}{2\beta+1}} (\log n)^{\cdot\cdot}\right)$$

uniformly in f_0, fixed R, M, ρ

(Ray, 2017)

ℓ_2 weak

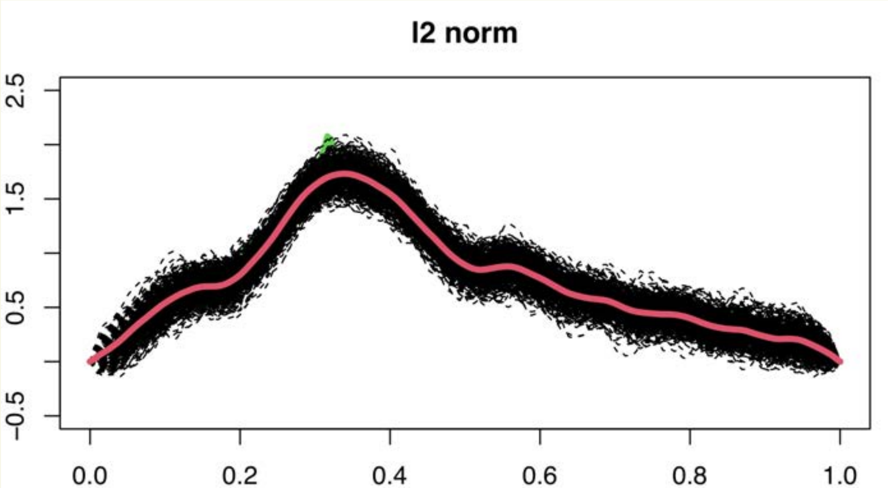


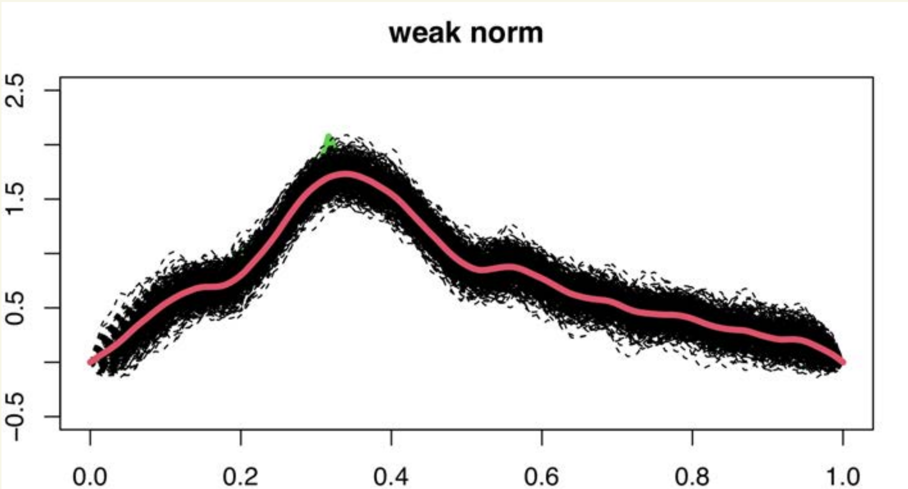
n=500

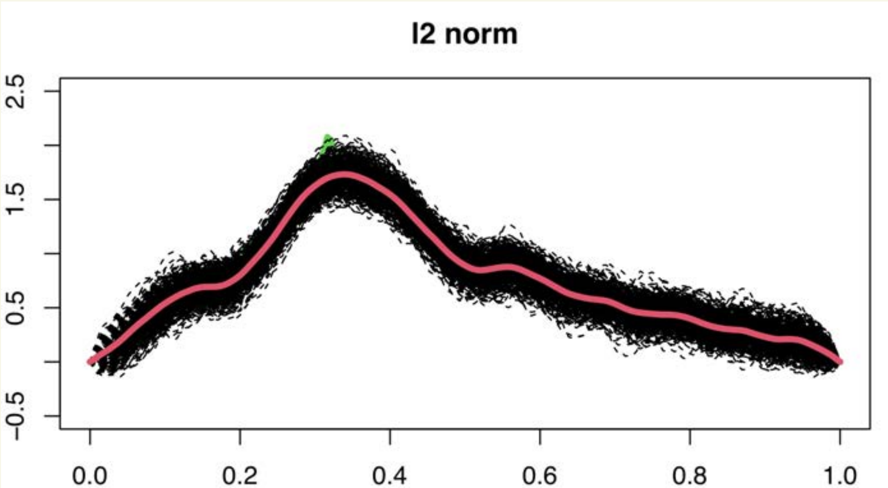
n=2000

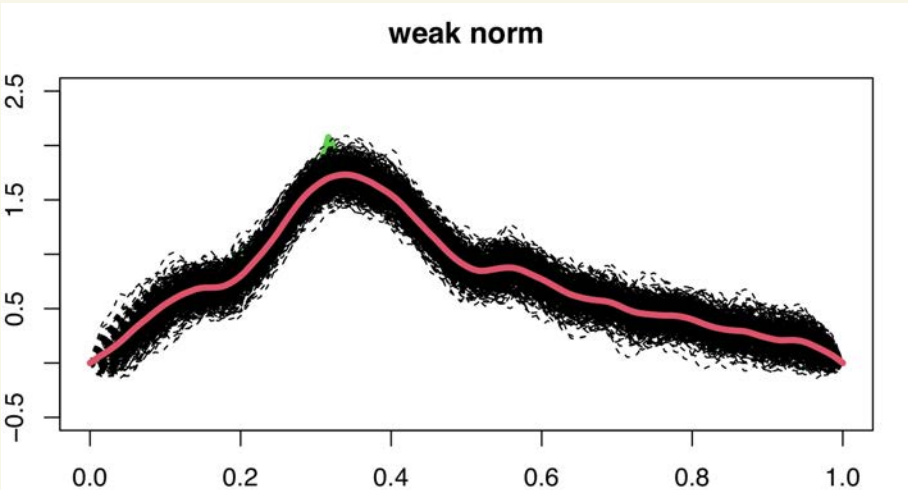
FIG. 1. *Empirical Bayes credible sets for the Fourier sine basis with the true curve (black) and the empirical Bayes posterior mean (red). The left panels contain the ℓ_2 credible ball $C_n^{\ell_2}$ given in* (5.1) *and the right panels contain the set $\tilde{C}_n$ given in* (4.2). *From top to bottom,* $n = 500, 2000$ *and* $\hat{\alpha}_n = 1.29, 1.01$, *with the right-hand side each having credibility* 95%.

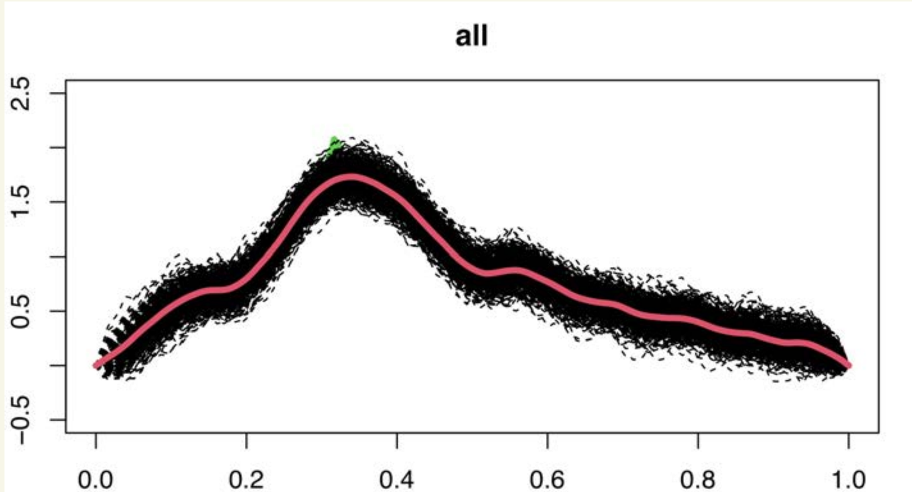
- draw 1000 times from posterior
- draw graphs of 950 functions closest to posterior mean


l2 norm
2.5
1.5
0.5
−0.5
0.0
0.2
0.4
0.6
0.8
1.0

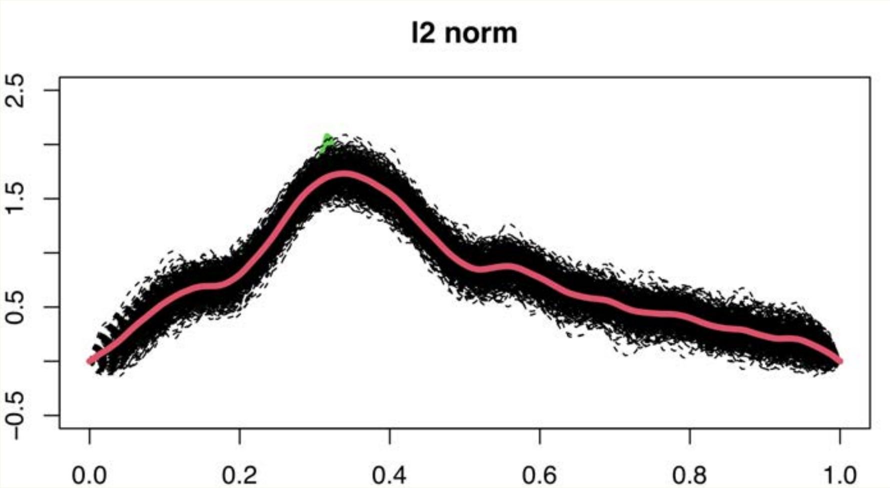

weak norm
2.5
1.5
0.5
−0.5
0.0
0.2
0.4
0.6
0.8
1.0

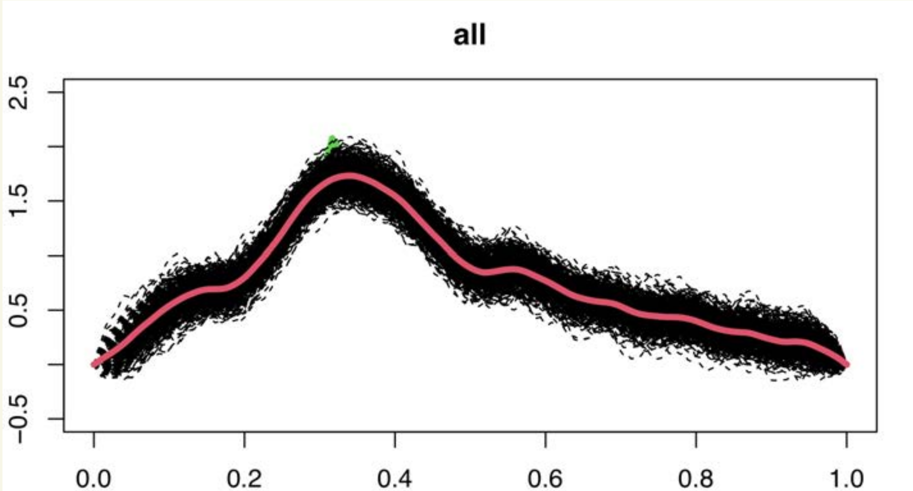

l2 norm
2.5
1.5
0.5
−0.5
0.0
0.2
0.4
0.6
0.8
1.0

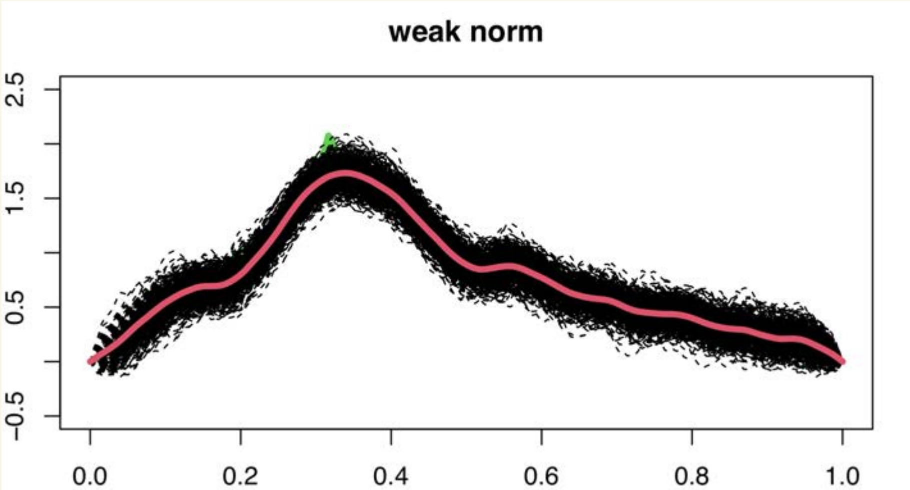

weak norm
2.5
1.5
0.5
−0.5
0.0
0.2
0.4
0.6
0.8
1.0


all
2.5
1.5
0.5
−0.5
0.0
0.2
0.4
0.6
0.8
1.0

(Ray 2017)







THM $\Pi_n(\hat{C}_n \cap C_n^{\ell_2} \mid X_n) - \Pi_n(\hat{C}_n \mid X_n)\, \Pi_n(C_n^{\ell_2} \mid X_n) \xrightarrow{P_{f_0}} 0$

Schrödinger Equation

(Nickl 2018)

$$\begin{cases} \frac{1}{2}\Delta u_f = u_f f & \text{on } \mathcal{O} \subset \mathbb{R}^d \\ u_f = g & \text{on } \partial\mathcal{O} \end{cases}$$

open

$$\log f \sim \sum_{j=1}^{J_n} \sum_{l} f_{j,l} \psi_{j,l} \qquad f_{j,k} \sim \text{Uniform}\left[-2^{-j(\alpha+d/2)}, 2^{-j(\alpha+d/2)}\right]$$

wavelet

Estimate $\chi_\psi(f) = \int \psi(x) f(x)\, dx$, given $\psi \in C_c^q(\mathcal{O})$, $q > 2 + \frac{3d}{2}$

THM if $f_0 \in C^\beta(\mathcal{O})$ for $\beta > (2+\frac{d}{2}) \vee d$ and $2^{J_n} \sim n^{\frac{1}{2\beta+4+d}}$

then BvM holds for $\chi_\psi(f)$.

$$\boxed{\begin{array}{l}\text{THM Infinite-dimensional BvM holds for } \left(\chi_\psi(f) : \|\psi\|_{C_c^q(\mathcal{O})} \le 1\right) \\[2em] \chi_\psi(f),\ \psi \in C_c^q(\mathcal{O}) \text{ identifies } f.\end{array}}$$

Adaptive

CREDIBLE BANDS

- Sniekers & vdV
- Yoo 2017

(Sniekers, 20)

Gaussian regression

$\theta | c \sim \sqrt{c}\, W$, Gaussian process

$y_i | \theta, c \overset{ind}{\sim} N(\theta(x_i), 1)$

Choose c by ℓ_2-EB (Wahba, 1974), lik-EB, hierarch. B.

Gaussian regression

(Szabo, van der Vaart, van Zanten...)

(Sniekers 20)

$\theta \mid c \sim \sqrt{c}\, W$, Gaussian process

$Y_i \mid \theta, c \overset{ind}{\sim} N(\theta(x_i), 1)$

Choose c by ℓ_2-EB (Wahba, 1978), lik-EB, hierarch. B.

$C_n^\gamma(x) = \{ f : |f(x) - E(f(x) \mid \vec{Y}_n)| \le c_n \xi_\gamma \, sd(f(x) \mid \vec{Y}_n) \}$

Gaussian regression

(Sniekers 20)

$\theta | c \sim \sqrt{c}\, W$, Gaussian process

$y_i | \theta, c \overset{\text{ind}}{\sim} N(\theta(x_i), 1)$

Choose c by ℓ_2-EB (Wahba, 1978), lik-EB, hierarch. B.

$$C_n^\gamma(x) = \{ f : |f(x) - E(f(x) | \vec{y}_n)| \le c_n \xi_\gamma \, sd(f(x) | \vec{y}_n) \}$$

THM If f_0 is polished tail, then, for c_n large,

$$P_{f_0}\left(\frac{1}{n} \sum_{i=1}^{n} 1\{ f \in C_n^\gamma(x) \} \ge \gamma \right) \to 1$$

Gaussian regression

(Sniekers 20)

$\theta \mid c \sim \sqrt{c}\, W$, Gaussian process

$y_i \mid \theta, c \overset{\text{ind}}{\sim} N(\theta(x_i), 1)$

Choose c by ℓ_2-EB (Wahba, 1978), lik-EB, hierarch. B.

$$C_n^\gamma(x) = \{ f : |f(x) - E(f(x) \mid \vec{y}_n)| \le c_n \xi_\gamma \, sd(f(x) \mid \vec{y}_n) \}$$

THM If f_0 is polished tail, then, for c_n large,

$$P_{f_0}\left(\frac{1}{n} \sum_{i=1}^n 1\{f \in C_n^\gamma(x_i)\} \ge \gamma \right) \to 1$$

Choose $c_n = \sqrt{\log n}$ and W = Brownian motion

THM. $P_{f_0}\left(f_0 \in \bigcap_x C_n^\gamma(x) \right) \to 1$ for:

- almost every f_0 generated from Gaussian prior,
- every self-similar $f_0 \in C^\beta[0,1]$.

Bayes + Lepski

(Yoo, 2017)

For $J \in \mathbb{N}$: $e_{J,1} \dots e_{J,J}$ "basis", e.g. splines or wavelets

$$f = \sum_{j=1}^{J} f_{J,j}\, e_{J,j} \quad , \quad (f_{J,1} \cdots f_{J,J}) \sim N_J(0, \Omega) \quad , \quad \Omega \asymp I$$

$$\hat{J}_n := \min\Big(j : \| E_j(f \mid y) - E_i(f \mid y) \|_\infty \lesssim l_{1,i} \sqrt{\frac{\log i}{\lambda_i}}, \ \forall i > j \Big)$$

$\in \left(\left(\frac{n}{\log n}\right)^{\gamma}, \frac{n}{(\log n)^{1+}} \right)$

$\| \sum_{j=1}^{J} |e_{J,j}| \|_\infty$

min $\asymp$ max eigenvalues of $\left(\sum_{i=1}^{n} e_{J,k}(X_i)\, e_{J,l}(X_i) \right)_{k,l=1..n}$

$C_J := \{ f : | f(x) - E_J(f(x) \mid Y) | \le R_J \, sd(f(x) \mid Y_n), \ \forall x \}$

R_J such that $\Pi_J(C_J \mid Y_n) = 1 - \gamma$.

Credible set $C_{\hat{J}_n}$.

Bayes + Lepski

(Yoo, 2017)

DATA $y_i = f_0(X_i) + \varepsilon_i$, $i = 1, \dots, n$

$$J^*(f_0) = \min\left(J : \|f_0\|_\alpha J^{-\alpha} \le l_{1,J} \sqrt{\frac{\log J}{\lambda_J}} \right)$$

THM if $\|f_0 - \mathrm{lin}(e_{J,1}, \dots e_{J,J})\|_\infty \lesssim \|f_0\|_\alpha J^{-\alpha}$, $\forall J$, then

- $P_{f_0}\left(\hat{J}_n \gtrsim J_n^*(f_0) \right) \to 0$
- $\Pi_{J_n}\left(f : \|f - f_0\|_\infty \gtrsim l_{1, J_n^*(f_0)} \sqrt{\frac{\log J_n^*(f_0)}{\lambda_{J_n^*(f_0)}}} \,\Big|\, Y_n \right) \to 0.$

Bayes + Lepski

(Yoo, 2017)

DATA $\quad y_i = f_0(X_i) + \varepsilon_i \quad , \; i = 1, \dots, n$

$$j^*(f_0) = \min\left(j : \|f_0\|_\alpha \, j^{-\alpha} \le l_{1,j} \sqrt{\frac{\log j}{\lambda_j}} \right)$$

THM $\quad$ if $\|f_0 - \mathrm{lin}(e_{j,1}, \dots e_{j,j})\|_\infty \lesssim \|f_0\|_\alpha \, j^{-\alpha}$, $\forall j$, then

- $P_{f_0}\left(\hat{j}_n \gtrsim j_n^*(f_0) \right) \to 0$
- $\Pi_{j_n}\left(f : \|f - f_0\|_\infty \gtrsim l_{1, j_n^*(f_0)} \sqrt{\frac{\log j_n^*(f_0)}{\lambda_{j_n^*(f_0)}}} \;\middle|\; y_n \right) \to 0.$

DEF $\quad f_0$ is self-similar if $\forall$ quasi-uniform knot sequence

$\xi_1 = 0 < \xi_1 < \dots < \xi_N = 1$: $\quad \|f_0 - \mathrm{splines}_j(\xi)\|_\infty \gtrsim \|f_0\|_\alpha \max_i |\xi_{i+1} - \xi_i|^\alpha$

THM (splines) $\quad$ If $f_0 \in C^\alpha[0,1]$ and self-similar, then

$$P_{f_0}\left(f_0 \in \mathcal{C}_{j_n} \right) \to 1 - \gamma.$$

Similar results with wavelets.

BNP NETWORK

THANKS

NICOSIA

april 2022