Bayesian wavelet estimators in nonparametric regression

Lecture 3

Natalia Bochkina University of Edinburgh

Lecture 3. Wavelet estimators in nonparametric regression

1. Thresholding estimators (universal, SURE, block thresholding; different types of thresholding functions).
2. Empirical Bayes estimators (posterior mean and median, Bayes factor estimator)
3. Different choices of prior distributions
4. Globally optimal wavelet estimators
5. Locally optimal wavelet estimators

Wavelet nonparametric regression

$$y_i = f(i/n) + \epsilon_i, \quad i = 1, \dots, n,$$

assuming $\mathbb{E}\epsilon_i = 0$, $\mathbb{E}\epsilon_i^2 < \infty$ and ϵ_i are independent.

Aim: estimate f.

Decompose $f \in L^2[0,1]$ in orthonormal **wavelet basis:**

$$f(x) = \sum_{k=0}^{2^L-1} \theta_k \phi_{Lk}(x) + \sum_{j=L}^{\infty} \sum_{k=0}^{2^j-1} \theta_{jk} \psi_{jk}(x),$$

where $\phi(x)$ is a **scaling function**, $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$ where ψ is a **wavelet function**, and $\{\theta_k, \theta_{jk}\}$ is a set of wavelet coefficients.

$$\int \phi(x)dx = 1, \quad \int \psi(x)dx = 0.$$

Wavelet transform

Apply orthonormal discrete wavelet transform:

$$\begin{aligned} d_{jk} &= w_{jk} + \varepsilon_{jk}, \\ c_{Lk} &= u_{Lk} + \varepsilon_{L-1,k}. \end{aligned}$$

$j = L, L+1, \ldots, J-1$, $k = 0, 1, \ldots, 2^j - 1$, $n = 2^J$.

If $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ then $\varepsilon_{jk} \sim \mathcal{N}(0, \sigma^2)$ iid.

To estimate f, need to estimate w_{jk} and u_{Lk}.

Relationship to “continuous” wavelet coefficients:

$$\theta_{jk} \approx \tilde{\theta}_{jk} = \frac{w_{jk}}{\sqrt{n}}$$

due to an approximation:

$$\theta_{jk} = \int_0^1 f(x)\psi_{jk}(x)dx \approx \frac{1}{n}\sum_{i=1}^{n} \psi_{jk}(i/n) f(i/n) = \frac{1}{\sqrt{n}}(Wf_n)_{(jk)} = \frac{w_{jk}}{\sqrt{n}}.$$

Minimax paradigm for wavelet estimators

- "global": $\sup_{f\in\mathcal{F}} \mathbb{E}||\hat{f}_n - f||_u^u = \inf_{\tilde{f}_n} \sup_{f\in\mathcal{F}} \mathbb{E}||\tilde{f}_n - f||_u^u,$
- "local": for $t_0 \in (0,1)$,
  $\sup_{f\in\mathcal{F}} \mathbb{E}|\hat{f}_n(t_0) - f(t_0)|^u = \inf_{\tilde{f}_n} \sup_{f\in\mathcal{F}} \mathbb{E}|\tilde{f}_n(t_0) - f(t_0)|^u$

for $u \in (0,\infty)$ and some functional space $\mathcal{F}$. Case $u = \infty$ for global optimality is also of interest.

Choice of functional space $\mathcal{F}$?

Besov space $B^r_{p,q}$

Besov sequence norm with $r > 0$, $1 \le p, q \le \infty$ is defined by

$$||\theta||_{b^r_{p,q}} = ||\theta_k||_p + \left[\sum_{j=L}^{\infty} 2^{qj(r+1/2-1/p)} ||\theta_j||_p^q \right]^{1/q}, \quad \text{if} \quad q < \infty,$$

$$||\theta||_{b^r_{p,q}} = ||\theta_k||_p + \sup_{L \le j < \infty} [2^{j(r+1/2-1/p)} ||\theta_j||_p], \quad \text{if} \quad q = \infty,$$

and the Besov sequence space $b^r_{p,q}(A) = \{\theta : \quad ||\theta||_{b^r_{p,q}} \le A\}$.

If regularity of wavelets s is such that $s > r > 0$, then Besov sequence norm is equivalent to Besov space norm (Donoho and Johnstone, 1998).

Properties of Besov space $B^r_{p,q}$

- Continuous functions: if $r > 1/p$
- Sobolev spaces: $B^r_{2,2} = W^r_2$
- Hölder spaces: $B^r_{\infty,\infty} = \mathbb{H}^r$.
- $B^r_{p,q} \subset L^p$
- Spaces of bounded total variation $\subset B^1_{1,\infty}$
- Spatially inhomogeneous functions: $p \in [1, 2)$
- Besov space embeddings: $B^r_{p,q} \subset B^r_{p,\infty}$

Wavelets and thresholding

Donoho and Johnstone (1994) showed that, for $u = 2$,

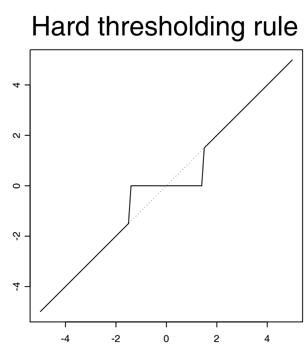
- no linear estimator can achieve such a rate (even up to a log factor) for $p \in [1, 2)$, i.e. for spatially inhomogeneous functions;
- a term-by-term thresholding estimator with threshold $\sigma\sqrt{2 \log n / n}$ (VisuShrink) achieves global minimax rate of convergence, up to a log factor, over Besov spaces $B^r_{p,q}$.

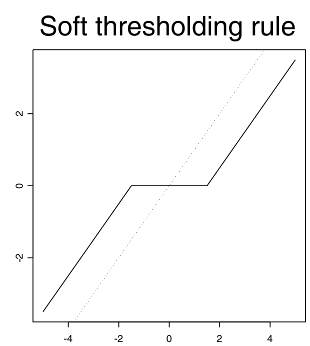
Thresholding estimators

Sparsity $\Rightarrow$ a small number of wavelet coefficients is sufficient to approximate a function well.

Leads to thresholding estimators:

$$\begin{aligned}\text{hard: } \eta_H(x;\lambda) &= xI(|x|>\lambda),\\ \text{soft: } \eta_S(x;\lambda) &= \text{sign}(x)(|x|-\lambda)I(|x|>\lambda).\end{aligned}$$





Some thresholds

- Universal: $\lambda_n = \sigma\sqrt{2\log n/n}$

  For purely white noise $\xi_i \sim N(0,\sigma^2), i = 1,\ldots,n,$

  $$\mathbb{P}(\max_{i=1,\ldots,n} |\xi_i| > \lambda_n) \to 0 \quad \text{as} \quad .n \to \infty.$$

- SURE (Stein's unbiased risk estimator) (Donoho and Johnstone, 1995)

  Stein's phenomenon: $X_i \sim \mathcal{N}(\mu_i, 1), i = 1,\ldots,m$. If $\hat{\mu}(x) = x + g(x)$ and g is weakly differentiable, then

  $\mathbb{E}||\hat{\mu}(X) - \mu||_2^2 = m + \mathbb{E}_\mu[||g(X)||_2^2 + 2\sum_{i=1}^m \frac{\partial g}{\partial x_i}(X_i)$

  If $\hat{\mu}$ is soft thresholding,

  $$SURE(\lambda, x) = m - 2\sum_{i=1}^m I(|x_i| \le \lambda) + \sum_{i=1}^m \min(x_i^2, \lambda^2).$$

  Then, $\lambda^{SURE} = \arg\min_{\lambda>0} SURE(x, \lambda)$.

  Wavelet estimator with soft thresholding and SURE threshold is minimax over Besov spaces with distance $\frac{1}{n}\sum_{i=1}^n |f(x_i) - \hat{f}_n(x_i)|^2$.

Block James-Stein estimator

BlockJS estimator (Cai,1999). Divide each resolution level $j < J$ into nonoverlapping blocks of approximate length $\ell = \log n$.

Denote (jb) the b-th block at level j and $S^2_{(jb)} = \sum_{k \in (jb)} y^2_{jk}$.

Let $\lambda^* = 4.50524$ be the root of the equation $\lambda - \log \lambda - 3 = 0$. The BlockJS estimator $\hat{f}_{BJS}$ has the following non-zero wavelet coefficients $\hat{\theta}_{jk}$:

$$\hat{\theta}_{jk} = \left(1 - \frac{\lambda^* \ell \sigma^2 n^{-1}}{S^2_{(jb)}}\right)_+ y_{jk}, \quad k \in (jb), \quad j < J.$$

Block James-Stein estimator: optimality

Then, Cai(1999) showed that $\hat{f}_{BJS}$ achieves minimax global rate for $u = 2$:

$$\limsup_{n\to\infty} n^{2r/(2r+1)} \sup_{f\in B^r_{p,q}(A)} \mathbb{E}||\hat{f}_{BJS} - f||_2^2 < \infty$$

for all $r \in (0, s)$, $p \geq 2$, $q \geq 1$ and $A > 0$ (for $p \in [1, 2)$ the upper bound includes a log factor),

as well as achieves minimax adaptive local rate for $u = 2$ over Hölder classes:

if we fix arbitrary $t_0 \in (0, 1)$, $\delta > 0$ such that $[t_0 - \delta, t_0 + \delta] \subseteq [0, 1]$,

$$\sup_{f\in B^r_{\infty,\infty}[t_0-\delta,t_0+\delta](A)} \mathbb{E}|\hat{f}_{BJS}(t_0) - f(t_0)|^2 \leq C \left(\frac{n}{\log n}\right)^{-\frac{2r}{2r+1}}$$

Bayesian wavelet methods

Can specify threshold and thresholding function from a different point of view, as a point posterior estimator.

Choice of prior for wavelet: must take into account sparsity.

- Prior is a mixture, one component - point mass at 0:

$$w_{jk} \sim (1-\pi_j)\delta_0(\cdot) + \pi_j h_j(\cdot),$$

  where $h_j(\cdot)$ is the prior density function of non-zero wavelet coefficients, and $\pi_j = \mathbb{P}(\theta_{jk} \neq 0)$.

- Shrinkage priors (discontinuity at 0):

$$p(w_{jk}) \propto \exp\{-\tau |w_{jk}|^\beta\},$$

  for appropriate $\beta > 0$, $\tau > 0$.

Scaling coefficients: $\hat{u}_{Lk} = c_{Lk}$ ($p(u_{Lk}) = 1$ - noninformative prior).

Posterior mixture distribution

Posterior distribution function:

$$F_j(w_{jk}|d_{jk}) = \frac{1}{1+\omega_j(d_{jk})} I_{(0,+\infty)}(w_{jk}) + \frac{\omega_j(d_{jk})}{1+\omega_j(d_{jk})} \tilde{H}_j(w_{jk}|d_{jk}), \tag{1}$$

where $\tilde{H}_j(x|y)$ is the posterior distribution of the non-zero component

$$\tilde{H}_j(x|y) = \frac{1}{g_j(y)} \int_{-\infty}^{x} \varphi_j(y-u) h_j(u) du,$$

$g(x)$ is the marginal density:

$$g_j(y) = \int_{-\infty}^{+\infty} \varphi_j(y-u) h_j(u) du,$$

and $\omega_j(y)$ is the posterior odds of the non-zero component:

$$\omega_j(y) = \frac{\pi_j g_j(y)}{(1-\pi_j)\varphi_j(y)}$$

Point estimates

- Posterior mean $\hat{w}_{jk} = E(w_{jk}|d_{jk})$ (Clyde and George, 1998),
- Posterior median $\hat{w}_{jk} = Med(w_{jk}|d_{jk})$ (Abramovich et al, 1998),
- Bayes factor estimate: $\hat{w}_{jk} = d_{jk} I \left\{ \frac{\mathbb{P}(w_{jk} \neq 0|d_{jk})}{\mathbb{P}(w_{jk}=0|d_{jk})} > 1 \right\}$ (Vidakovic 1998). Related to testing hypothesis $H_0 : w_{jk} = 0$ vs $H_1 : w_{jk} \neq 0$.

Properties:

- Thresholding estimator: $\exists \lambda > 0$ such that $\hat{w}_{jk} = 0$ for $|d_{jk}| \leq \lambda$, $\hat{w}_{jk}$ - monotonic function of d_{jk}
- Shrinkage estimator: $|\hat{w}_{jk}| \leq |y_{jk}|$, monotonic function of d_{jk}.

Thresholding: Bayes factor and posterior median, $\lambda_{BF} \leq \lambda_{Med}$

Non-zero shrinkage: posterior mean and median

Normal mixture prior distribution

Gaussian likelihood: $d_{jk} \mid w_{jk} \sim \mathcal{N}(w_{jk}, \sigma^2)$,

Gaussian mixture prior: $w_{jk} \sim (1-\pi_j)\delta_0 + \pi_j \mathcal{N}(0, \tau_j^2)$

Gaussian posterior distribution:

$$w_{jk} \mid d_{jk} \sim \frac{1}{1+\omega_j(d_{jk})}\delta_0 + \frac{\omega_j(d_{jk})}{1+\omega_j(d_{jk})}\mathcal{N}\left(d_{jk}\frac{\tau_j}{\sqrt{\tau^2+\sigma_j^2}}, \frac{\tau_j^2\sigma^2}{\tau_j^2+\sigma^2}\right),$$

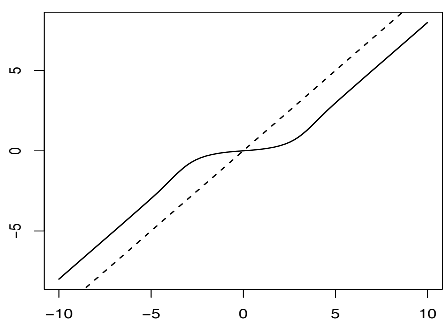
where the posterior odds $\omega_j(d_{jk})$ is

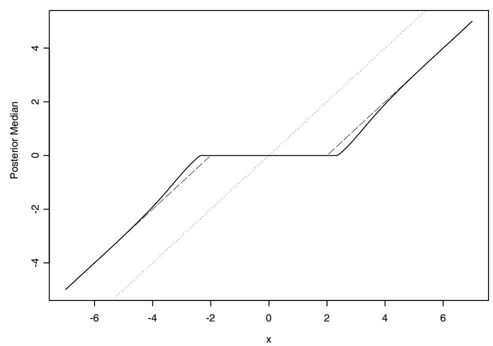
$$\omega_j(y) = \frac{(1-\pi_j)}{\pi_j}\frac{\sigma}{\sqrt{\sigma^2+\tau_j^2}} e^{\frac{\tau_j^2}{\sigma^2(\sigma^2+\tau_j^2)}y^2/2}$$

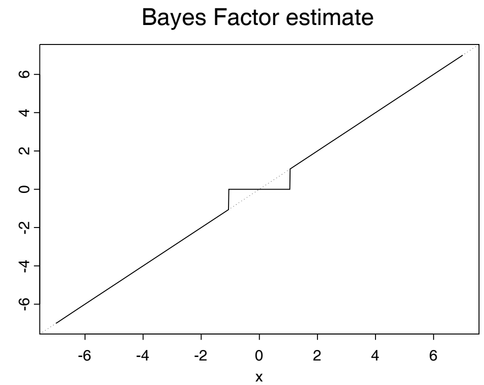
Posterior median for double exponential (Laplace) nonzero prior component and Gaussian noise

Non-zero component prior: $h_j(x) = \frac{a_j}{2} e^{-a_j|x|}$.

Mean	Median	Bayes Factor
$E(w_{jk}\|d_{jk})$	$Med(w_{jk}\|d_{jk})$	$\hat{w}_{jk} = d_{jk} I\left\{\omega_{jk} < 1\right\}$







Comparison of mixture priors

Double exponential (Laplace) prior, Gaussian noise: constant shrinkage of size $a\sigma$

Student t distribution, t noise distribution: asymptotically zero shrinkage

If prior distribution tail is heavier than the tail of the noise distribution (Gaussian prior and Laplace error): heavy shrinkage.

(plots from Bochkina & Sapatinas, 2005)

Example: plethysmography data

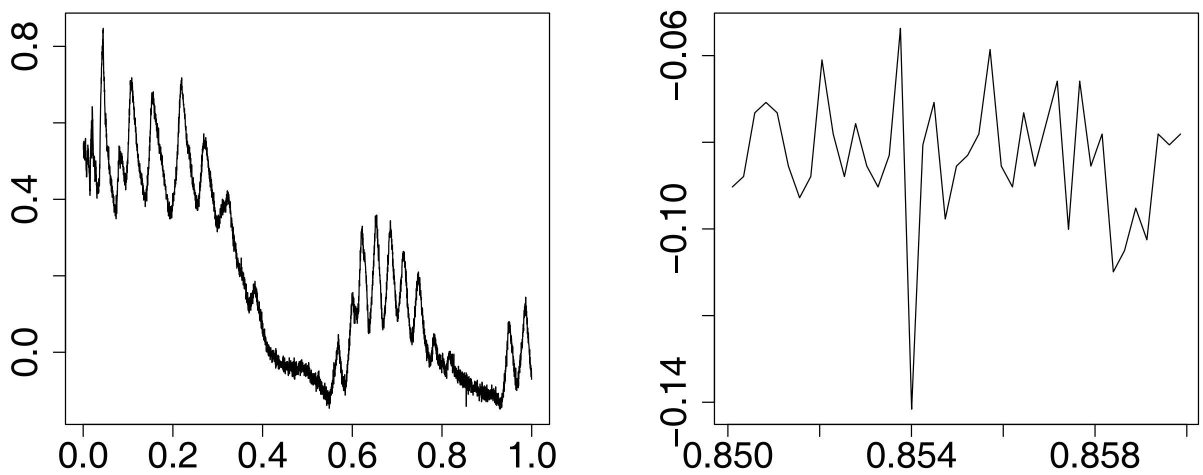


Figure 1: (Left) Section of an inductance plethysmography recording lasting approximately 80 seconds. (Right) Zoom in for the rapid variation near point 0.85.

Bayesian model for wavelet coefficients

Gaussian errors: $d_{jk} \mid w_{jk}, \sigma^2 \sim \mathcal{N}(w_{jk}, \sigma^2)$

Prior: $w_{jk} \mid \nu_j, \pi_j \sim (1-\pi_j)\delta_0 + \pi_j h_j(x)$

with $h_j(x) = \frac{\nu_j}{2} e^{-\nu_j |x|}$ - double exponential, and hyperpriors

$$\begin{aligned} \pi_j &\sim \mathrm{Uniform}[0,1], \\ \nu_j &\sim \mathrm{Exp}(1), \\ \sigma^{-2} &\sim \Gamma(0.001, 0.001). \end{aligned}$$

Fitted in WinBUGS

(Show MCMC output)

Bayes Factor wavelet estimator

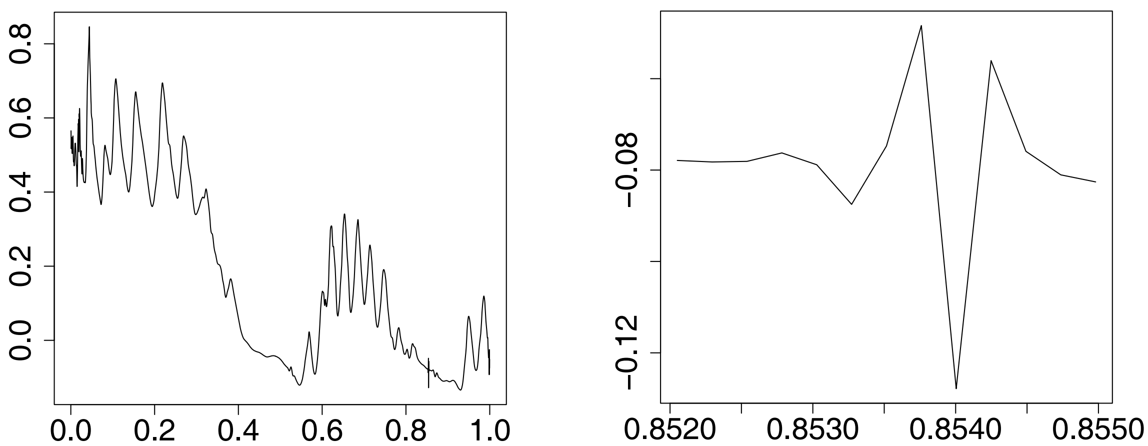


Figure 2: (Left) Smooth estimate obtained using the BF estimator. (Right) Zoom in for the rapid variation near point 0.85 for the panel shown in Figure 1.

How to choose h?

Usually do not have a priori information about h.

- h - normal: closed form posterior distribution
- h: function f has certain properties a priori, e.g. specified regularity - Abramovich, Sapatinas, Silverman (1998), Bochkina (2002); others.
- Posterior estimator of f has "good" properties, e.g. optimal error rates over a class of functions: global - Johnstone & Silverman (2005), Pensky (2006), Pensky & Sapatinas (2007); local - Abramovich, Angelini, de Candiitis (2007), Bochkina & Sapatinas (2006, 2009).
- For a mixture prior $(1-\pi)\delta_0 + \pi h$: study the tail behaviour of wavelet coefficients in practice using Extreme Value Theory (light tail vs heavy tail).
- Posterior predictive checks: Marshall & Spiegelhalter (2003), Lewin, Bochkina, Richardson (2007) - for a mixture model.
- More?

Back to optimality

Ideal aim: Given a loss function, characterision of the set of priors such that Bayesian wavelet estimators are optimal over Besov spaces.

(with some stability to misspecification of error distribution.)

Bayesian wavelet estimators: posterior median and Bayes factor.

Adaptive estimators.

Loss functions:

- local error $d(f, \hat{f}_n) = |f(x_0) - \hat{f}_n(x_0)|^u$, $u \in (0, \infty)$
- global error $d(f, \hat{f}_n) = ||f - \hat{f}_n||_u$, $u \in (0, \infty]$.

Global minimax rate for Besov spaces

For Besov spaces $B^r_{p,q}$ and $u \in (0, \infty)$, global minimax rate is given by

$$\inf_{\tilde{f}_n} \sup_{f \in B^r_{p,q}} \mathbb{E}||\tilde{f}_n - f||^u_u \asymp \Lambda_n(r,p)$$

where

$$\Lambda_n(r,p) = \begin{cases} n^{-\frac{ur}{2r+1}}, & \text{if} \quad u < p(2r+1), \\ \left(\frac{n}{\log n}\right)^{-\frac{u(r-1/p)+1}{2(r-1/p)+1}} \log n, & \text{if} \quad u = p(2r+1), \\ \left(\frac{n}{\log n}\right)^{-\frac{u(r-1/p)+1}{2(r-1/p)+1}}, & \text{if} \quad u > p(2r+1) \end{cases}$$

(Donoho et al. (1995, 1996), Delyon and Juditsky (1996)).

Global optimality Pensky (2006):

- Prior: h - normal, double exponential and t, specific form of hyperparameters:
- Estimator: Posterior mean and median, nonadaptive
- Optimality: for $u \in [1, \infty)$.

Pensky and Sapatinas (2007):

- Prior: h - normal, double exponential and t, specific form of hyperparameters:
- Estimator: Bayes factor estimator, nonadaptive
- Optimality: for $u \in [1, \infty)$.

Johnstone and Silverman (2005):

- Prior: (DE or heavier), π_j is estimated
- Estimator: posterior median or any other thresholded estimator with bounded shrinkage property, adaptive
- Optimality: for $u \in (0, 2]$.

Optimal pointwise rate for normal errors

Risk: $R_n^u(t_0, B_{p,q}^r, \hat{f}) = \sup_{f \in B_{p,q}^r} \mathbb{E}|\hat{f}(t_0) - f(t_0)|^u$

Pointwise minimax rate:

$$\inf_{\hat{f}} R_n^u(t_0, B_{p,q}^r, \hat{f}) \asymp n^{-\frac{u(r-1/p)}{2(r-1/p)+1}}, \quad n \to \infty.$$

for any $r > 1/p$.

Adaptive pointwise minimax rate:

$$\inf_{\hat{f}} R_n^u(t_0, B_{p,q}^r, \hat{f}) \asymp \left(\frac{n}{\log n}\right)^{-\frac{u(r-1/p)}{2(r-1/p)+1}} \quad n \to \infty,$$

for any $r > 1/p$.

(Cai (2003) - for white noise model, Bochkina & Sapatinas (2009) - for nonparametric model)

Wavelet estimators that achieve optimal rate

Non-adaptive wavelet estimator:

Achieved for a hard thresholding wavelet estimator with threshold:

$$t_{jn} = \begin{cases} \sigma_j n^{-1/2} & \text{if } j \leqslant j_1 \\ \sigma_j n^{-1/2}[2uj \log 2]^{1/2} & \text{if } j > j_1 \end{cases}$$

where $j_1 = \frac{1}{2(r-1/p)+1} \log_2 n$.

Adaptive wavelet estimator

Achieved for a hard thresholding wavelet estimator with threshold

$$t_{jn} = \sigma_j n^{-1/2}[uj]^{1/2}.$$

Approximation

Johnstone & Silverman (2004): constructed boundary coiflets based on orthonormal coiflets in order to give a better approximation error between discrete and continuous wavelet coefficients for Besov spaces.

Orthonormal coiflets (ϕ, ψ) have regularity s, are supported in $[-S+1, S]$, $s < S$, with $L \geq \log_2(6S-6)$, for some integer S.

Approximation error for $f \in B^r_{p,q}(A)$, $1 \leq p, q \leq \infty$ and $r > 1/p$:

$$2^{j(r-1/p+1/2)}||\tilde{\theta}_{jk} - \theta_{jk}||_p \leq AC(\phi, \psi, p, r)2^{-r(J-j)},$$

where J: $n = 2^J$.

Allows to bound the approximation error.

Pointwise optimal Bayesian wavelet estimators

Abramovich et al (2007): $w_{jk} \sim (1-\pi_j)\delta_0 + \pi_j N(0, \tau_{jn}^{-2})$, $\tau_{jn}^2 = C_\tau 2^{j(2(r-1/p)+1)} n$, $\pi_j = \min(1, C_\beta 2^{-\beta j})$.

Non-adaptive model.

The pointwise rate of convergence for posterior mean, median and Bayes factor wavelet estimators is given by

$$R_n^2(t_0, \hat{f}, B_{p,q}^r) = O\left(n^{-\frac{2(r-1/p)}{2(r-1/p)+1}} \log n\right)$$

Settings

Error distribution: $\varepsilon_{jk} \sim \varphi_j(x)$

Prior distribution: $w_{jk} \sim (1-\pi_j)\delta_0(x) + \pi_j \tau_{jn} h(\tau_{jn} x)$, $\tau_{jn} = \nu_j \sqrt{n}$

Use Bayes factor estimator:

$$\hat{w}_{jk} = d_{jk} I \left(\frac{P(w_{jk} \neq 0 \mid d_{jk})}{P(w_{jk} = 0 \mid d_{jk})} > 1 \right)$$

- hard thresholding estimator (Vidakovic, 1998).

Bayes factor wavelet estimator with $\hat{\theta}_k = \hat{u}_{L,k}/\sqrt{n}$, $\hat{\theta}_{jk} = \hat{w}_{jk}/\sqrt{n}$:

$$\hat{f}_{BF}(x) = \sum_{k=0}^{2^L-1} \hat{\theta}_k \phi_{L,k}(x) + \sum_{j=L}^{J-1} \sum_{k=0}^{2^j-1} \hat{\theta}_{jk} \psi_{jk}(x).$$

Pointwise adaptive Bayesian estimator: assumptions

- Loss: pointwise, $1 \leq u < \infty$
- Regression function: $f \in B^r_{pq}(A)$ with $1 \leq p, q \leq \infty$, $A > 0$ and $r > 1/p$.
- Wavelets: boundary coiflets with regularity $s > r$
- Error distributions (in wavelet domain):
  - have symmetric unimodal density φ_j
  - $\mathbb{E}|\varepsilon_{jk}|^u < \infty; \quad \mathbb{V}(\varepsilon_{jk}) \in [\underline{\sigma}^2, \bar{\sigma}^2] \subset (0, \infty)$; $\mathbb{V}(\varepsilon_{L-1,k}) < \infty$.
- Prior:
  - has symmetric unimodal density h
  - $|h'(x)/h(x)| \leqslant C_h$, $C_h > 0$, or h - normal.
  - $|\varphi_j(x)/h(x)| \leq C_{\varphi h}$
  - h is such that $\omega_j(y)$ increases for $y > 0$:

$$\omega_j(y) = C \int \varphi_j(x - y)\tau_{jn} h(\tau_{jn} x) dx / \varphi_j(y)$$

Adaptive Bayesian estimator with normal errors

Theorem 1. *Let $\varphi_j(x)$ be the pdf of $\mathcal{N}(0, \sigma_j^2/2)$. In addition to the assumptions above, assume that the following conditions hold.*

1. $\nu_j/\sqrt{n} \leqslant C$ *for some* $C > 0$;
2. $C_1 n^{b+1/2} 2^{aj} \leqslant \beta_{jn}\sqrt{n}/\nu_j < C_2 n^B$ *for some* $B, C_1, C_2 > 0$ *and* $b + 1/2 - (u/2 - a)_+ \geqslant 0$.

Then, for any $t_0 \in (0, 1)$,

$$R_n^u(\hat{f}_{BF}, B_{pq}^r, t_0) = O\left(\left(\frac{n}{\log n}\right)^{-\frac{u(r-1/p)}{2(r-1/p)+1}}\right) \quad \text{as} \quad n \to \infty.$$

Adaptive Bayesian estimator with power exponential errors

Theorem 2. *Assume that* $\varphi_j(x) = C_\beta \sigma_j^{-1} e^{-(|x|/\sigma_j)^\beta}$, $\beta > 0$, *and that* h *is either heavy-tailed or Gaussian.*

In addition to the assumptions above, assume that the following conditions hold.

1. $\nu_j/\sqrt{n} \to 0$ *as* $n \to \infty$;
2. $C_1 n^{b+1/2} 2^{aj} \leqslant \beta_{jn}\sqrt{n}/\nu_j < C_2 \exp\{B[\log n]^{\beta/2}\}$ *for some* B, C_1, $C_2 > 0$ *and* $b + 1/2 - (u/2 - a)_+ \geqslant 0$.

Then, for any $t_0 \in (0,1)$,

$$R_n^u(\hat{f}_{BF}, B_{pq}^r, t_0) = O\left(\left(\frac{n}{\log n}\right)^{-\frac{u(r-1/p)}{2(r-1/p)+1}}\right) \quad \textit{as} \quad n \to \infty.$$

Conclusions

To achieve optimal pointwise convergence rate, choose prior:

- tail of prior h should not be lighter than the tail of the error distribution;
- for adaptive estimators, variance of the non-zero component of the prior distribution should be greater than the error variance;
- choosing heavy-tailed prior distribution h (plus other conditions) leads to the optimal rate;
- additional constraints on the parameters of the prior distribution π_j and ν_j.

Some robustness to the misspecification of error distribution: conditions for pointwise optimality of nonadaptive Bayes factor estimator overlap.

Further questions

Bayesian wavelet estimators optimal in L^∞ norm?

Considered problem: given a loss function, find optimal adaptive estimator.

Does there exist an estimator that is optimal for both local and global losses?