Statistical Approaches for Wavelet Shrinkage

Simon Lunagomez
Math 348

March 19, 2007

Outline

- Introduction
- Distributional Invariance
- The Donoho-Johnstone Approach
- Bayesian Methods
- Summary

Introduction

- We observe N noisy samples from f

$$y_i = f(x_i) + \sigma\epsilon_i, \qquad i = 1, \ldots, N.$$

- here $x_i = (i-1)/N$, and $\epsilon_i \sim \mathcal{N}(0,1)$.
- **Goal**: Propose $\widehat{f}$ with small risk $R(\widehat{f}, f) = \frac{1}{N} E\|\widehat{f} - f\|_2^2$.

Introduction

- Find strategy with minimax risk (frequentist way)

$$\bar{R} = \inf_{\widehat{f}} \sup_{f \in \mathcal{F}(M)} E_f \left\| f - \widehat{f}(x) \right\|_2^2$$

- Or propose prior π over space of models and find Bayes rule

$$\min_{\widehat{f}} E^{\pi} \left[\left\| f - \widehat{f}(x) \right\|_2^2 \mid x \right]$$

(this is the Bayesian way).

Distributional Invariance

- Let $y = f + \sigma\epsilon$, with $\epsilon \sim \mathcal{N}(0, I)$
- W is the matrix of the DWT
- Then

$$\begin{aligned} Wy &= Wf + \sigma\widetilde{\epsilon}, \text{ or} \\ d &= \theta + \sigma\widetilde{\epsilon}, \end{aligned}$$

- with $\widetilde{\epsilon} \sim \mathcal{N}(0, I)$.

The Donoho-Johnstone Approach

- Landmark paper by Donoho and Johnstone (1994)
- The model is $y = f + \sigma\epsilon$ (or $d = \theta + \sigma\widetilde{\epsilon}$).
- **Key Assumption**: Only a very few wavelet coefficients needed to describe f.
- **Idea**: Keep only those coefficients above the noise level.
- σ is regarded as known.

The Wavelet-Shrinkage Paradigm

$$\text{Data} \longrightarrow \text{DWT} \longrightarrow \textbf{Shrinkage} \longrightarrow \text{Inv. DWT} \longrightarrow \text{Function Estimate}$$

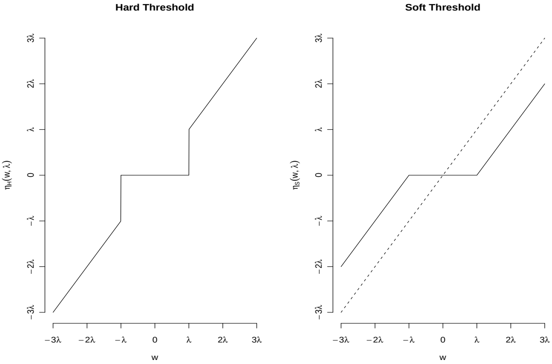
The Donoho-Johnstone Approach

- Define

$$\begin{aligned}\eta_H(w,\lambda) &= w\mathbf{1}\{|w| > \lambda\}\\ \eta_S(w,\lambda) &= sgn(w)(|w| - \lambda)_+\end{aligned}$$

- These are known as **hard** and **soft** thresholding.

Thresholding Policies



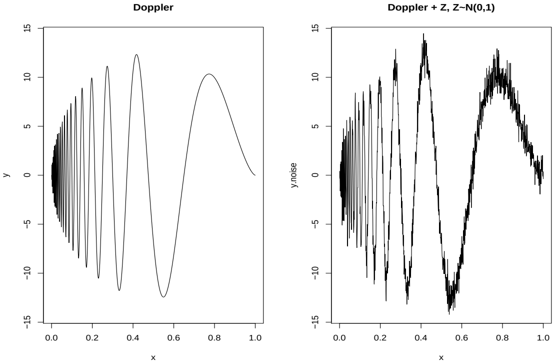
The Donoho-Johnstone Approach

- Donoho *et al* proposed $\widehat{\theta}_i^* = \eta_S(d_i, \sigma\sqrt{2\log n})$
- and proved

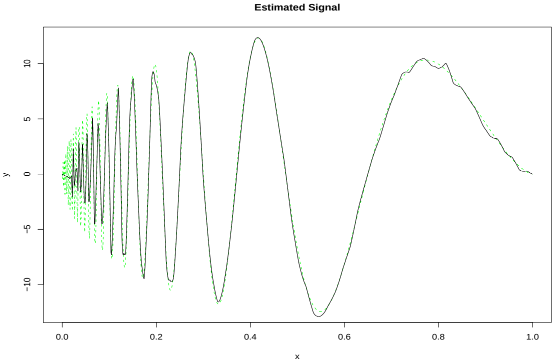
$$E\left\|\widehat{\theta}^* - \theta\right\|_{2,n}^2 \leq (2\log n + 1)\left\{\sigma^2 + \sum_{i=1}^{n} min(\theta_i^2, \sigma^2)\right\}.$$

- $\sum_{i=1}^{n} min(\theta_i^2, \sigma^2)$ is known as the **oracular risk**.

Universal Thresholding in Action



Universal Thresholding in Action



A Note on Smoothness

- Need to restrict f to $\mathcal{F}(M)$
- For certain $\mathcal{F}(M)$ we can ensure that $\widehat{f}$ will be (with high probability) as smooth as f.

Bayesian Thresholding

- Again $d = \theta + \sigma\tilde{\epsilon}$ (i.e, $d \mid \theta, \sigma^2 \sim \mathcal{N}(\theta, \sigma^2 I)$).
- For thresholding we have to test

$$H_0 : \theta_i = 0 \qquad \text{vs} \qquad H_i : \theta_i \neq 0.$$

- Use prior (Vidakovic, 1998)

$$\theta_i \sim \pi(\theta_i) = \pi_0\delta_0 + \pi_1\xi(\theta_i),$$

- where $\pi_0 + \pi_1 = 1$.

Bayesian Thresholding

- Need $p(d_i \mid \theta_i)$ such that does not depend on σ.
- Define $p(\sigma^2 \mid \mu) = \mu e^{-\mu\sigma^2}$
- then

$$p(d_i \mid \theta_i) = \int_0^\infty p(d_i \mid \theta_i, \sigma^2) p(\sigma^2 \mid \mu) d\sigma^2 = \frac{\sqrt{2\mu}}{2} e^{-\sqrt{2\mu}|d_i - \theta_i|}$$

- $p(d_i \mid \theta_i)$ is called the **marginal likelihood**.
- Compute $p(\theta_i \mid d_i)$ via Bayes Theorem.

Bayesian Thresholding

- The **Bayes Rule** is given by

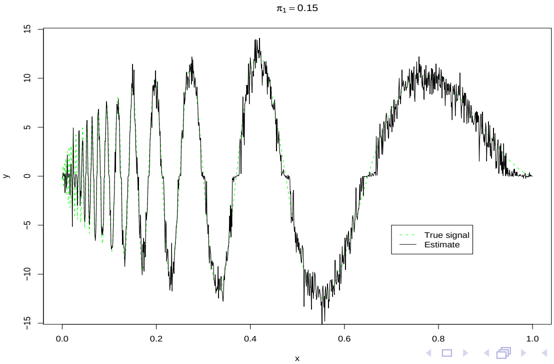
$$\widehat{\theta}_i = d_i \mathbf{1}\left(P(H_0 \mid d_i) < \frac{1}{2}\right)$$

- where

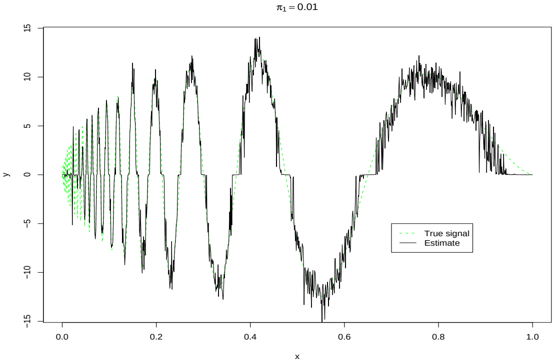
$$P(H_0 \mid d_i) = \left(1 + \frac{\pi_1}{\pi_0} \frac{\int_{\theta_i \neq 0} p(d_i \mid \theta_i)\xi(\theta_i)d\theta_i}{p(d_i \mid \theta_i = 0)}\right)^{-1}.$$

Bayesian Thresholding in Action

- Use Cauchy distribution as ξ



Bayesian Thresholding in Action



Summary and Final Comments

- DWT $\longrightarrow$ **Shrinkage** $\longrightarrow$ Inv. DWT
- **D-J method** is conservative (over smoothing)
- Easy to implement
- Easier to incorporate assumptions for the **Bayesian approach**
- Harder to implement

Summary and Final Comments

- **D-J method** tends to produce estimates nearly as smooth as true signal
- **Bayesian approach** sensitive to prior specification
- Tends to introduce artifacts

References

- Donoho, D.L. and Johnstone, I.M. (1994) Ideal Spatial Adaptation by Wavelet Shrinkage. *Biometrika*, 81, 425-55.
- Vidakovic, B. (1998) Nonlinear Wavelet Shrinkage with Bayes Rules and Bayes Factors. *JASA*, 93, 173-9.
- Vidakovic, B. (1999) *Statistical Modeling by Wavelets*. New York: John Wiley and Sons.