Robust splines and wavelets

Douglas Nychka,
www.image.ucar.edu/~nychka

- A penalized robust smoother
- Pseudo data.
- Empirical pseudo data
- Robust wavelets (and wavelets)

Supported by the National Science Foundation  L’ Ecole D’Ete, Ovoronnaz Sep 2005
The additive model

Given $n$ pairs of observations $(x_i, y_i), \ i = 1, \ldots, n$

$$y_i = g(x_i) + \epsilon_i$$

$\epsilon_i$'s are random errors.

Assume that $g$ is a smooth function or a realization of a Gaussian process and $\epsilon$ are symmetric but potentially heavy tailed.

*Being tied to local averages makes it difficult to think about a robust method.*

*What is the model for the unknown function?*
A univariate robust estimate as a minimization problem

Recall that

\[ \min_a \sum_{i=1}^{n} (y_i - a)^2 \]

is a complicated way to characterize \( \bar{y} \).

Let \( \rho(t) \) be a convex function symmetric about 0.

\( \rho \) is like \( t^2 \) close to 0 but grows like \( |t| \) when \( t \) is large.

\[ \min_a \sum_{i=1}^{n} \rho(y_i - a) \]

will give a robust estimator that down weights the effect of large values of \( Z \).
An example of $\rho$ and its derivative
Transition is at $-2$ and $2$.

A smoother version is $\rho(t) = \log(\cosh(t))$
A reasonable penalized smoother

Just replace the sum of squares with a robust measure of fit.

Recall

\[ \hat{g}(x) = \sum_{l=1}^{n} \hat{\theta}_k \psi_k(x) \]

\[
\min_{\theta} \sum_{i=1}^{n} \rho(y - [W\theta]_i) + \lambda \theta^T B \theta
\]

This is now a nonlinear problem to find \( \theta \)

But in some sense we are done ...
Some details of actually finding $\hat{g}(x)$

For convenience parametrize this as $f = W\theta W^{-1}f = \theta$

$$
\min_{f} \sum_{i=1}^{n} \rho(y_i - f_i) + \lambda f^T W^{-T} B W^{-1} f
$$

Differentiate, set equal to zero and Solve for $f$.

$$\eta = \rho' \text{ and } R = W^{-T} B W^{-1}$$

$$\eta(y_i - f_i) - \lambda[Rf]_i = 0$$
Adding an outlier to the ozone data
Some examples of robust smooths
Analysis of the robust estimator based on pseudo-data

This is theory – not yet practical. Create alternative observations: pseudo-data

\[ \tilde{y}_i = g(x_i) + \frac{\eta(\epsilon_i)}{2} \]

Recall that \( \eta = \rho \) and \( \eta(\epsilon_i) \) will be bounded.

Use a least squares type smoother applied to the pseudo data
Theoretical Result for pseudo data

Consider smoothers that are based on a reproducing kernel (i.e. spatial statistics type estimates )

\( \hat{g} \) robust estimate

\( \tilde{g} \) LS estimate of \( g \) based on pseudo data.

\[ \| f \|_n^2 = \frac{1}{n} \sum_{i=1}^{n} f(x_i)^2 \text{ and } C_n = E \| \tilde{g} - g \|_n^2 \]

\[ \| \hat{g} - \tilde{g} \|_n / \sqrt{C_n} \to 0 \]

as \( n \to \infty \).

Under several assumptions, e.g. on \( \lambda, A(\lambda) \)
Empirical pseudo data

Use an estimate in place of \( g \).

\[
\tilde{y}_i = \hat{g}(x_i) + \eta(y_i - \hat{g}(x_i))/2
\]

Empirical pseudo data (EPD) and fixed points

Apply a LS smoother to the EPD to get a robust estimate of \( g \). Our goal is a fixed point where the estimate obtained is the same as that used to construct the EPD

An algorithm:

- Start with initial \( \hat{g}^0 \).
- Repeat:
  1. Form EPD: \( \tilde{y}_i = \hat{g}^J(x_i) + \eta(y_i - \hat{g}^J(x_i))/2 \)
  2. \( \hat{g}^{J+1} \) LS smoother based on \( \tilde{y} \)
Comments

- If the algorithm converges it will give the robust estimator!
- In the smoothing step 2 one can also use Cross validation to choose the smoothing parameter.
- One need not use a spline type smoother any penalized LS smoother can work.
Wavelets

Choose a multiresolution basis: members are similar shape but different sizes of supports at different locations.
A simple shrinkage estimate

Usually the wavelet basis is orthogonal, locations equally spaced.

\[ W^T y = \theta^* \]

These are the empirical coefficients for \( g \). (i.e. \( y = W \theta^* \))

Now shrink and decimate

For some \( C > 0 \)

\[ \hat{\theta}_k = \text{sign}(\theta^*_k)(|\theta^*_k| - C)^+ \]

\((u)^+ = u \text{ if } u > 0 \text{ and } 0 \text{ if } u \leq 0\)
So what about penalized LS?
For orthogonal wavelets. \( W^T W = I \).

\[
\min_{\theta} \sum_{i=1}^{n} (y_i - [W \theta]_i)^2 + \lambda \sum_{i=1}^{n} |\theta_i|
\]
So what about penalized LS?

For orthogonal wavelets. \( W^T W = I \).

\[
\min_{\theta} \sum_{i=1}^{n} (y_i - [W \theta]_i)^2 + \lambda \sum_{i=1}^{n} |\theta_i| 
\]

\[
\min_{\theta} \| (y - W\theta) \|^2 + \lambda \sum_{i=1}^{n} |\theta_i| 
\]

\[
\min_{\theta} \| \theta^* - \theta \|^2 + \lambda \sum_{i=1}^{n} |\theta_i| 
\]
So what about penalized LS?

For orthogonal wavelets. \( W^T W = I. \)

\[
\min_{\theta} \sum_{i=1}^{n} (y_i - [W \theta]_i)^2 + \lambda \sum_{i=1}^{n} |\theta_i|
\]

\[
\min_{\theta} \| (y - W \theta) \|^2 + \lambda \sum_{i=1}^{n} |\theta_i|
\]

\[
\min_{\theta} \| \theta^* - \theta \|^2 + \lambda \sum_{i=1}^{n} |\theta_i|
\]

\[
\min_{\theta} \sum_{i=1}^{n} (\theta^*_i - \theta_i)^2 + \lambda \sum_{i=1}^{n} |\theta_i|
\]
So what about penalized LS?

For orthogonal wavelets. $W^TW = I$.

$$\min_\theta \sum_{i=1}^n (y_i - [W\theta]_i)^2 + \lambda \sum_{i=1}^n |\theta_i|$$

$$\min_\theta \|y - W\theta\|^2 + \lambda \sum_{i=1}^n |\theta_i|$$

$$\min_\theta \|\theta^* - \theta\|^2 + \lambda \sum_{i=1}^n |\theta_i|$$

$$\min_\theta \sum_{i=1}^n (\theta^*_i - \theta_i)^2 + \lambda \sum_{i=1}^n |\theta_i|$$

For each $i$

$$\min_{\theta_i} (\theta^*_i - \theta_i)^2 + \lambda |\theta_i|$$

We know how to do this! And it is the same as shrinkage.
A robust wavelet

Choose a multiresolution basis ....

\[
\min_{\theta} \sum_{i=1}^{n} \rho(y - [W\theta]_i) + \lambda \sum_{i=1}^{n} |\theta_i|
\]

The absolute value penalty really changes the estimator!

EPD also works here.
Summary

We have substituted a robust measure of fit for the sum of squares to get a robust smoother.

There is some theory for the robust smoother based on pseudo data.

Empirical pseudo data provides a computational algorithm and a generalization to more exotic penalized estimators.