Wavelets

Consider approximating \( y \) by \( f(x|\beta) \) of the form

\[
\sum_j \sum_k \beta_{jk} \psi_{jk}(x)
\]

where

\[
\psi_{jk}(x) = 2^{j/2} \psi(2^{j/2} x - k), \quad j, k = 0, \pm 1, \pm 2, \ldots
\]

is a series of orthonormal functions of compact support

\( \psi \) is called the mother function

E.g. Haar

\[
\psi(x) = -1 \quad 0 < x < \frac{1}{2}
\]

\[
= 1 \quad \frac{1}{2} < x < 1
\]

\[
= 0 \quad \text{otherwise}
\]

The \( \beta_{jk} \) may be determined by OLS
Nonlinear least squares

There are whole books on the topic, e.g. Seber’s

\[ \rho_i(\beta) = (y_i - f(x_i | \beta))^2, \quad f \text{ known} \]

e.g.

\[ \beta_0 \exp{-\beta_1 x} + \beta_2 \exp{-\beta_3 x} \]
\[ \beta_0 \cos(\beta_1 x + \beta_2) \]

nonlinear in \( \beta \)

Differentiating

\[ \sum_i (y_i - f(x_i | \beta))^2 \]
wrt $\beta$ leads to the

"normal" equations

$$\sum_i [y_i - f(x_i | b)] \frac{\partial f(x_i | b)}{\partial b} = 0$$

fitting/solving

locally linear

$$[f(x_i | \beta)] \approx [f(x_i | \beta_0)] + X_0(\beta - \beta_0)$$

iterate (Gauss–Newton)

$$\beta_1 - \beta_0 = (X_0^T X_0)^{-1} X_0^T (y - [f(x_i | \beta_0)])$$

supposing inverse exists

nls()

Example.


Previous researchers had found concentration of DDT to be proportional to trout’s age
\[ f(x|\beta) = \beta_1 + \beta_2 \exp\{\beta_3 x\}, \quad x: \text{age} \]

need good initial values

often need scaling

Fletcher-Powell/Levenberg-Marquart adaptation

Possible difficulties

lack of convergence
ill conditioning
nonidentifiability

residual analysis

Some routines handle linear parameters distinctly

\textit{robust/resistant variant}

\[ \sum_i \rho\left(\frac{y_i - f(x_i|b)}{s}\right) \frac{\partial f(x_i|b)}{\partial b} = 0 \]

with \(s\) a robust scale value
Back to smoothing/nonparametric regression

Parametric form not apparent a priori

Predict $y$ by $f(x)$ with $f$ smooth

Typically wish $f(x)$ for range of $x$’s

Variety of methods:

- polynomial
- fourier
- other functional forms
- spline
- kernel
- local (polynomial) OLS

Difficulties:

- choosing degree of smoothing (try several)
- choosing method (local polynomial preferred to kernel - end effects, theory)
There are forms combining parametric and nonparametric

There are robust/resistant variants

**The sensitivity curve**

How is result, \( T_n(y_1, \ldots, y_n) \) changed when another observation, \( y \), is added?

The sensitivity curve

\[
SC(y) = (n+1)\{T_{n+1}(y_1, \ldots, y_n, y) - T_n(y_1, \ldots, y_n)\}
\]

Describes effect an outlier might have

For the mean, \( \bar{y}_n \), it is

\[
y - \bar{y}_n
\]

which behaves linearly as \( y \) deviates from \( \bar{y}_n \)

For the median

\[
SC(y) = .5*n*(y_{(m)} - y_{(m+1)}) \quad \text{for} \quad y < y_{(m)}
\]

\[
= .5*n*(y - y_{(m+1)}) \quad \text{for} \quad y_{(m)} \leq y \leq y_{(m+2)}
\]

\[
= .5*n*(y_{(m+2)} - y_{(m+1)}) \quad \text{for} \quad y > y_{(m+2)}
\]

Relates to deletion statistics
There is the related concept of the empirical influence curve (Cook and Weisberg).

In the regression case, adding \((x^T, y)\), this is

\[ n(X^T X)^{-1} x (y - x^T b) = n(X^T X)^{-1} xr \]

depends on \(y\) and \(x\)

increases linearly with the residual at \((x^T, y)\)

[More later]

data perturbation

variability of results with data