Resistant

In sensitive to localized misbehavior in the data
e.g. median

Robust

In sensitive to departures from assumptions surrounding an underlying probability model

trimmed mean
M-estimates

The Biweight

\[ y^* = \frac{\sum w_i y_i}{\sum w_i} \]

\[ w_i = \begin{cases} 1 - \left( \frac{d_i / c}{\epsilon} \right)^2 & \text{if } \frac{d_i}{c} < \epsilon \\ 0 & \text{otherwise} \end{cases} \]

\[ s = \text{median} \left( |y_i - y^*| \right) \]

\[ a = \frac{1}{2} \text{ IQR} \]

S estimates \( \frac{2}{3} s \)

eg. \( c = 6 \), residuals count up to 4c

Iterate to convergence
The heuristics of robust regression. Consider a weight function like

\[ w(r) = (1 - r^2)^2 \quad 1 \leq |r| < 1 \]
\[ w(r) = 0 \quad \text{otherwise} \]

Consider a preliminary estimate \( \hat{\beta} \) and residuals \( \hat{e}_i = y_i - x_i^T \hat{\beta} \).

Regress \( y_i \) on \( x_i \) with weight \( w(\hat{e}_i) = w_i \).

i.e. \( \hat{\beta} \) to satisfy

\[ \sum_i (y_i - x_i^T \hat{\beta}) w_i x_i = 0 \quad \text{for} \quad (1, x_i^T)^T w_i x_i \neq 0 \]

In the limit

\[ \sum_i (y_i - x_i^T \hat{\beta}) w(\hat{e}_i) x_i = 0 \quad \text{for} \quad (1, x_i^T)^T w_i x_i = 0 \]

i.e.

\[ t(r) = r w(r) \quad \text{or} \quad w(r) = t(r)/r \]

\[ r = (1 - r^2)^2 \quad 1 \leq |r| < 1 \]
\[ r = 0 \quad \text{otherwise} \]

\[ p(t) = \int_0^t (1 - s^2) ds^2 = \int_0^t \frac{1}{2} (1 + s) ds^2 \quad t = 0^+ \]

\[ p(t) = \frac{1}{2} (1 - t^2)^2 \quad 0 \leq t \leq 1 \]
\[ p(t) = \frac{1}{6} \left[ 1 - (1 - t^2)^3 \right] \quad |t| > 1 \]
This approach suggests the approximate variance matrix
\[ \text{var}(-\beta \sim \sigma^2 (X^t WX)^{-1}) \]

e.g., \( \hat{\beta} = \frac{3}{4} \text{IQR} \), median = 31213671, 6745

Asymptotic normality from M-estimate theory

Start eg. with L1 estimate
Robust / Resistant Estimate

Would like an automatic way to handle outliers and long-tailed distributions. Give up full efficiency, as in normal case, but obtain protection against outliers and nonnormality.

\[ y_i = x_i^T \beta + \varepsilon_i \]

M-estimate

\[
\min_{\beta} \sum_{i=1}^{n} \rho \left( \frac{y_i - x_i^T \beta}{\sigma} \right)
\]

or to satisfy

\[
\sum_{i=1}^{n} \left( y_i - x_i^T \hat{\beta} \right) x_i = 0
\]

\[ t = \rho' \]
Examples of $\rho_i$

1. OLS \hspace{1cm} \rho(r) = r^2

2. L_1 \hspace{1cm} \rho(r) = |r|

3. Huber \hspace{1cm} \rho(r) = r^2 \hspace{1cm} 1r1 \leq H
   \hspace{1cm} = H(2|r|-H)

   \[\begin{array}{c}
   H \text{ large: normal} \\
   H \text{ small: L}_1
   \end{array}\]

   $H = 1.345$ gives 95% efficiency at normal

   $\rho$ is convex and mathematically convenient

4. Bisquare \hspace{1cm} \rho(r) = \frac{B^2}{2}\left[1 - \left(1 - \frac{|r|}{B}\right)^2\right] \hspace{1cm} 1r1 \leq B
   \hspace{1cm} = \frac{B^2}{2} \hspace{1cm} \text{otherwise}
Example: Detecting circular arcs in image. Roads.

Fitting a circular arc when outliers severely perturb fit:

\[ C_i = (x - a)^2 + (y - b)^2 = R^2 \]

\[ x^2 + y^2 = (-2a)x + (-2b)y + (a^2 + b^2 - R^2) \]

\[ Z = \Theta_1 x + \Theta_2 y + \Theta_3 \]

CAE: circular arc estimator

DRO: Duda road operator
- nonlinear line detector

Theil-Sen: median of \( (n_i) \) slopes

RM: repeated median

LSM: least median of squared residuals

MM: variant of M-estimate
- estimates scale too

SW: sliding window
Figure 15: Data set 2: (a) Input image; (b) DRO output; (c) OLS fit; (d) Theil-Sen CAE fit (not contained in the displayed window); (e) RM CAE fit; (f) LMS fit; (g) nonlinear MM fit; (h) SW fit using an unbiased OLS estimator.
Robust estimates.

M-estimates. Return to the regression problem

\[ y \sim X \beta + \varepsilon \]

Consider the problem

\[ \min_{\beta} \sum_{j=1}^{n} \rho \left( \frac{y_j - x_j^T \beta}{s} \right) \]

where \( s \) is a known or estimated scale parameter, e.g.,

\[ 1.48 (\text{median} \left| y_j - x_j^T \hat{\beta}_{ols} \right|) \]

1.48 works for normal

Here \( \rho(r) \) is a (robust) loss function

E.g., \( \rho(r) = r^2 \) (OLS)

\( \rho(r) = |r|^p \) (Lp regression)

\( \rho(r) = \begin{cases} r^2 & \text{for } |r| \leq H \\ H^2 & \text{for } |r| > H \end{cases} \) (Huber)
\[ p(r) = \begin{cases} \frac{3}{2} \left[1 - \left(1 - \frac{r}{B}\right)^{3}\right] & \text{if } |r| \leq B \\ \frac{B^2}{2} & \text{if } |r| > B \end{cases} \]

biweight
his;square

Might pick tuning constants \( H, B, \ldots \) so 95% efficient in the normal error case.

Note: 1. The location problem is a particular case with \( \mathbf{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \)

\( L_2 \) regression gives sample mean
\( L_1 \) " " " median

2. \( \mathbb{E} \) other rules for estimating scale

Properties of \( p(r) \):

i) \( p(r) \geq 0 \), \( p(r) \) non-decreasing for \( r \geq 0 \)

ii) \( p(0) = 0 \)

iii) \( p(-r) = p(r) \)

iv) \( p(\cdot) \) continuous for all but a finite number of \( r \)
Differentiate w.r.t \( \beta \) and write \( 4 = \rho' \)

\[ \sum_{j=1}^{n} x_{j} \left[ 4 \left( y_{j} - \hat{x}_{j} \frac{\hat{\beta}}{\delta} \right) \right] = 0 \quad (*) \]

A set of nonlinear equations
- solve iteratively
- good starting values needed

Write \( w(r) = 4(r)/r \)

\( (*) \) becomes

\[ \sum_{j=1}^{n} w \left( \frac{y_{j} - \hat{x}_{j} \frac{\hat{\beta}}{\delta}}{s} \right) x_{j} \left( y_{j} - \hat{x}_{j} \frac{\hat{\beta}}{\delta} \right) = 0 \]

or

\[ \sum_{j=1}^{n} w_{j} x_{j} \left( y_{j} - \hat{x}_{j} \frac{\hat{\beta}}{\delta} \right) = 0 \]

with data dependent weights

\[ w_{j} = w \left( \frac{y_{j} - \hat{x}_{j} \frac{\hat{\beta}}{\delta}}{s} \right) \]

IRLS regresses \( \sqrt{w_{j}} y_{j} \) on \( \sqrt{w_{j}} x_{ij} \)

until "convergence"
\[ w(r) \]
\[ \frac{2}{1/1r/1} \]
\[ 1 \quad 1r/1 \leq H \]
\[ H/1r/1 \quad 1r/1 > H \]

\[ (1 - (r/B)^2)^2 \quad 1r/1 \leq B \]
\[ O \quad 1r/1 > B \]

Notes:
1. \( \rho \): nondecreasing \( \rightarrow \) estimate exists and unique
2. \( \rho(r) = \) constant for \( 1r/1 \geq \) constant
   \( \Rightarrow \) possible deletion of observations

Classification:
\[ t(r) = 0 \] for \( 1r/1 \) sufficiently large, e.g., Biweight
  \textit{hard re-descender}

\[ t(r) \rightarrow 0 \] as \( 1r/1 \rightarrow \infty \), e.g., Cauchy
  \textit{soft re-descender} \( w(r) = 1/(1 + (r/c)^2) \)

\textit{non-decreasing}
Biweight estimate of location

\[ \hat{\mu} = \frac{\sum_i w_i y_i}{\sum_i w_i} \]

\[ w_i = \left(1 - \left(\frac{y_i - \hat{\mu}}{c}\right)^2\right)_+ \]

\[ s = \text{med} \{ | y_i - \hat{\mu} | \} \]

\[ c = 6.9 \]

Seems to converge quickly.


Inconsistency results.
Heuristics in location case (Jeffreys)

\[ y_i \sim f \left( \frac{y_i - \mu}{\sigma} \right) \quad \sigma : \text{"known"} \]

log likelihood \[ \sum_i \log f \left( \frac{y_i - \mu}{\sigma} \right) \]

\[ \frac{2}{\sigma} \sum_i \frac{1}{f \left( \frac{y_i - \mu}{\sigma} \right)} f' \left( \frac{y_i - \mu}{\sigma} \right) = 0 \]

so \[ \sum_i w_i (y_i - \hat{\mu}) = 0 \]

\[ w_i = \frac{1}{f \left( \frac{y_i - \hat{\mu}}{\sigma} \right)} f' \left( \frac{y_i - \hat{\mu}}{\sigma} \right) \frac{1}{(y_i - \hat{\mu})} \]

Cauchy: \[ f = \frac{1}{1+y^2}, \quad f' = -\frac{2y}{(1+y^2)^2} \]

\[ w = -\frac{1}{f} f' \frac{1}{y} = \frac{(1+y^2)}{(1+y^2)^2} \frac{2y}{(1+y^2)} \frac{1}{y} = \frac{2}{(1+y^2)} \]

Jeffreys took mixture of normal and uniform
Approximate distribution.

Suppose \( y = x \beta + \epsilon \) distribution \( \epsilon_i \sim \mathcal{N}(0, \sigma^2) \).

\[
\hat{\beta} \sim \mathcal{N}(\beta, \frac{E_p(y^2)}{[E_p(y^4)]\sigma^2}(x'x)^{-1})
\]

Estimate of the covariance matrix

\[
\frac{(n-1)}{(n-p)} \sum_{j=1}^{n} \frac{4(\frac{\hat{\epsilon}_j}{\sigma})^2 (x'x)^{-1}}{\sum_{j=1}^{n} 4(\frac{\hat{\epsilon}_j}{\sigma})^2} 4(r) = rw(r)
\]

Sometimes \( \hat{\sigma}^2 = x'wxx \)

\[\hat{\sigma}^2 = \frac{3}{4} \text{IQR} \]
In practice run OLS and cob/iws in parallel.

If results dissimilar, indication on an invalid assumption (of classical regression)

Look at residuals for possible outliers

eg. index plot

Look at index plot of weights
Stat 131a: Midterm

Y, obtained score

X, guessed score

$Y = 4.4 + .8X$

Stat 131a: Midterm + outlier

Y, obtained score

X, guessed score

solid line OLS, dashed biweighted LS
Summary of Robust/Resistant Methods

1. Can lead to identification of outliers
2. Can suggest where present model does not apply
3. Leads to parameter estimates which are not sensitive to arbitrary changes in any small part of the data
   - breakdown point

Discuss anomalies with subject matter experts
Functions in Splus:

`location.m()`  bisquare, huber

`l1fit()`  L₁

`rrreg()`  converged huber followed by bisquare

`robust()`  bisquare, single X

`ltareg()`  t-norm half 1.5

`rebusd()`  `glm(..., family = robust)`

`gam()`