Huber smooth M-estimator

Māra Vēliņa, Jānis Valeinis

University of Latvia

Sigulda, 28.05.2011

Contents

Introduction

Aim: robust estimation of location parameter

- Huber M-estimator (1964) - well known robust location estimator
- Owen (1988) introduced empirical likelihood method, also applicable to M-estimators
- Hampel (2011) proposed a smoothed version of Huber estimator

Work in progress

Two sample problem: empirical likelihood based method for a difference of smoothed Huber estimators
(Valeinis, Velina, Luta: abstract for ICORS 2011 conference)

M-estimator

Let $X_1, X_2, \ldots, X_n \sim \text{iid}$, $X_1 \sim F$. An M-estimator T_n is defined as a solution of

$$\sum_{i=1}^{n} \rho(X_i, t) = \int \rho(x, t) dF_n(x), \qquad (1)$$

for a specific funtion ρ where F_n is the empirical CDF.
If ρ is differentiable in t, then (1) is minimized by the solution of

$$\sum_{i=1}^{n} \psi(X_i, t) = 0,$$

where $\psi(x, t) = \frac{\partial}{\partial t} \rho(x, t)$.

Examples

- Mean. $\psi(x, t) = x - t$ gives $T_n = \bar{X}$.
- ML estimator. $\psi(x, \theta) = -\frac{d}{d\theta} \log f(x, \theta)$ for a class of density functions $f(x, \theta)$, gives T_n is the root of likelihood equation

$$\frac{d}{d\theta} \log \left(\prod_{i=1}^{n} f(X_i, \theta) \right) = 0.$$

- Median. $\psi(x, t) = \psi_0(x - t)$, $\psi_0(z) = k \operatorname{sgn}(z)$, $k > 0$.

Huber estimator for location parameter μ

Huber (1964) combined examples of mean and median.
Let F have a symmetric density

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right),$$

assume $\sigma = 1$. Then M-estimator for the location parameter μ is defined as

$$\sum_{i=1}^{n} \psi\left(\frac{X_i - t}{\sigma}\right) = 0. \qquad (2)$$

Huber M-estimator is defined by the function ψ in (2):

$$\psi_k(x) = \begin{cases} k, & x \geq k \\ x, & -k \leq x \leq k \\ -k, & x \leq -k. \end{cases} \qquad (3)$$

Huber's motivaton:

- Unrestricted ψ-functions have undesired properties (unstable to outliers);
- Cosider the limiting values of k in ψ_k and their respective M-estimators:
  - If $k \to \infty$, then ψ_k is mean;
  - If $k \to 0$, then ψ_k is median.
- k is a tuning constant determining the degree of robustness.
- Huber estimator has minimax assymptotic variance for class of distribution functions

$$(1-\epsilon)\phi(x) + \epsilon h(x),$$

where ϕ is pdf of $N(0,1)$ and h is a symmetric density.

Scaled estimator of location

In reality σ is not known, thus a robust estimate of σ should be used. A common choice is MAD.

MAD

$$\mathrm{S_n} = \mathrm{MAD} = \mathrm{median}(|\mathrm{X_i} - \mathrm{median}(\mathrm{X_i})|).$$

Robust estimator is acquired, even in presence of outliers (up to 50% of the sample).

Smoothed M-estimator (Hampel, 2011)

For a general ψ-function of an M-estimator define

$$\tilde{\psi}(x) = \int \psi(x+u) dQ_n(u), \tag{4}$$

where

- Q_n may be chosen as a the distribution of the initial M-estimator
- Q_n can be approximated by $N(0, V/n)$, where V is assymptotic variance of the M-estimator.
- Need to specify distribution under which the assymptotic variance is computed.
- The smoothing prinicple can be applied to ψ functions already smooth.

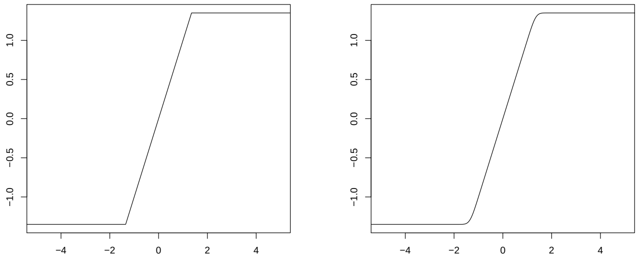
Smoothed Huber estimator

The ψ-function of the smoothed Huber estimator defined by $\psi = \psi_k$ can be written in closed form as

$$\begin{aligned} \tilde{\psi}_k(x) &= k\Phi\left(\frac{x-k}{\sigma_n}\right) - k\left(1 - \Phi\left(\frac{x+k}{\sigma_n}\right)\right) \\ &+x\,\Phi\left(\frac{x+k}{\sigma_n}\right) - \Phi\left(\frac{x-k}{\sigma_n}\right)\Big) \\ &+\sigma_n\left(\phi\left(\frac{x+k}{\sigma_n}\right) - \phi\left(\frac{x-k}{\sigma_n}\right)\right), \end{aligned} \tag{5}$$

where $\sigma_n = \sqrt{V/n}$, and Φ and ϕ denote the cdf and pdf of $N(0,1)$.

Example



(a) (b)

(a) ψ function of Huber M-estimate;
(b) $\tilde{\psi}$ function of smoothed Huber M-estimate. $\mathrm{k}{=}1.35$.

Empirical likelihood method for M-estimators

- Owen (1988) showed that EL method can be applied to certain M-estimators, including Huber estimator.
- Nonparametric Wilk's theorem applies thus EL based confidence intervals for Huber estimate can be obtained.
- Tsao, Zhu (2001) showed that EL based confidence intervals preserves robustness.

EL confidence bands for Huber estimator

Empirical likelihood ratio for parameter t

$$R(t) = \sup\{\prod_{i=1}^{n} \omega_i \sum_{i=1}^{n} \omega_i \psi(X_i, t) = 0, \omega_i \geq 0, \sum_{i=1}^{n} \omega_i = 1\}$$

is maximized by $\prod \omega_i(\lambda)$, where

$$\omega_i(\lambda) = \{n(1 + \lambda Z_i)\}^{-1},$$

and $Z_i = \psi(X_i, t)$ and λ follows from

$$n^{-1} \sum Z_i/(1 + \lambda Z_i) = 0.$$

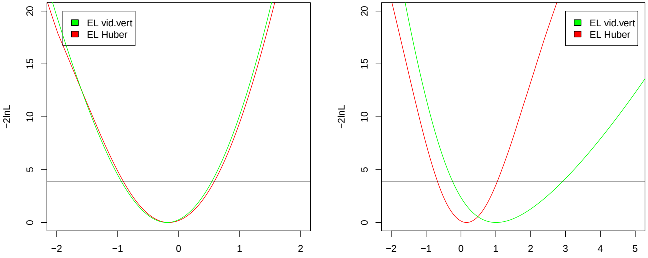


Figure: EL -2*ln,(a) $\mathrm{N}(0,3)$ (b) $0.95 * \mathrm{N}(0,3) + 0.05 * \mathrm{N}(20,3)$

Simulation results for one sample problem

Table: Huber estimation for location parameter and its EL confidence bands, alpha=0.05

$N(0,3)$					$0.95 * N(0,3) + 0.05 * N(20,3)$			
sample		len		estimate		len		estimate
n=50	EL.huber	0.494	EL.huber	-0.055	EL.huber	1.706	EL.huber	0.159
	EL.mean	0.492	EL.mean	-0.064	EL.mean	3.14	EL.mean	1.008
	t-test	0.506	mean	-0.064	t-test	3.117	mean	1.008
	z-test	0.554	huber	-0.076	z-test	0.554	huber	0.159
	Bootstrap	0.497			Bootstrap	3.057		
n=20	EL.huber	0.667	EL.huber	-0.167	EL.huber	2.478	EL.huber	-0.441
	EL.mean	0.667	EL.mean	-0.167	EL.mean	4.894	EL.mean	0.498
	t-test	0.732	mean	-0.167	t-test	4.938	mean	0.498
	z-test	0.877	huber	-0.643	z-test	0.877	huber	-0.441
	Bootstrap	0.699			Bootstrap	4.583		
n=10	EL.huber	1.001	EL.huber	-0.067	EL.huber	4.303	EL.huber	-0.189
	EL.mean	1.001	EL.mean	-0.067	EL.mean	9.68	EL.mean	1.008
	t-test	1.239	mean	-0.067	t-test	11.494	mean	1.799
	z-test	1.24	huber	-0.201	z-test	1.24	huber	-0.189
	Bootstrap	1.039			Bootstrap	9.74		

Two sample EL problem

Consider empirical likelihood-based method for the difference of smoothed Huber estimators.

Given two independent samples X and Y with distribution functions F_1 and F_2, respectively, we have two unbiased estimating functions:

$$\mathrm{E}_{\mathrm{F}_1}\mathrm{w}_1(\mathrm{X},\theta_0,\Delta)=0,\quad \mathrm{E}_{\mathrm{F}_2}\mathrm{w}_2(\mathrm{Y},\theta_0,\Delta)=0,$$

where Δ is the parameter of interest and θ_0 is a nuisance parameter. Specifically, $\Delta=\theta_1-\theta_0$ and

$$\mathrm{w}_1(\mathrm{X},\theta_0,\Delta)=\tilde{\psi}\left(\frac{\mathrm{X}-\theta_0}{\hat{\sigma}_1}\right)\quad \mathrm{w}_2(\mathrm{Y},\theta_0,\Delta)=\tilde{\psi}\left(\frac{\mathrm{Y}-\Delta+\theta_0}{\hat{\sigma}_2}\right),$$

where $\hat{\sigma}_1$ and $\hat{\sigma}_2$ are scale estimators, and $\tilde{\psi}$ corresponds to the smoothed Huber estimator.

Simulation results for two sample problem

Conisder two models:

- $Y_1 \sim (1-\epsilon)\mathrm{Gamma}(\alpha = 5; \sigma = 1) + \epsilon \mathrm{Uniform}[0; 50]$
- $Y_2 \sim \mathrm{Gamma}(\alpha = 1; \sigma = 5)$

Table: Coverage accuray and average confidence interval lengths based on 1000 replicates, $n_1 = n_2 = 50$

	t.int		EL.hub1		EL.hub2		Boot1		Boot2	
	acc	ave	acc	len	acc	len	acc	len	acc	len
$\sigma = 5$	0.62	3.05	0.66	2.99	0.56	2.83	0.36	2.98	0.36	2.98
$\sigma = 6$	0.69	3.56	0.73	3.51	0.65	3.34	0.38	3.46	0.38	3.47
$\sigma = 7$	0.74	4.09	0.77	4.04	0.72	3.85	0.44	3.97	0.45	3.99
$\sigma = 8$	0.78	4.62	0.81	4.56	0.76	4.39	0.48	4.49	0.48	4.50
$\sigma = 9$	0.81	5.19	0.84	5.13	0.80	4.95	0.50	5.00	0.50	5.02

Thank you for your attention!