Robust Statistics 1: Ideas and Tools

GEORGY L. SHEVLYAKOV

Peter the Great St. Petersburg Polytechnic University

Summer School 2018, March 05, Nizhniy Novgorod, RUSSIA

OUTLINE

Generalities: Ontology, History and Ideas

Robust (*Latin*: strong, healthy, vigorous, sturdy, tough)

Robustness (Box 1953) <=> **Stability**

(Tukey 1960): **the Least Squares Method estimates are not stable under small deviations from Gaussianity!**

Generalities: Ontology, History and Ideas

Consider the Cauchy contaminated Gaussian distribution density (**Tukey's gross-error model**)

$$f(x;\theta) = \frac{1-\varepsilon}{\sqrt{2\pi}} \exp\left(-\frac{(x-\theta)^2}{2}\right) + \frac{\varepsilon}{\pi[1+(x-\theta)^2]},$$

where θ is a parameter of location and $0 \leq \varepsilon < 1$ is a parameter of contamination—the probability of outlier occurrence.

The sample mean $\bar{x}$ is the LSM estimate of location for a Gaussian, but for arbitrarily small $\varepsilon > 0$ **it is not even consistent!**

The classical robust estimate is **the sample median** ***med x*****.**

Generalities: Ontology, History and Ideas

1) (Huber 1964, 1981): **Minimax Approach**

Minimax Principle: to search for **the best solution in the least favorable case —** a guaranteed quality result, sometimes too pessimistic.

Huber's minimax approach in robustness is a good example of application of the minimax principle.

Generalities: Ontology, History and Ideas

2) (Hampel 1968, 1986): **The Approach Based on Influence Functions**

Lyapunov: **Stability = Continuity** => **Robustness = Continuity**

Parametric Statistics (1900–1940)

Robust Statistics (1960–2000)

Nonparametric Statistics (1940–1960)

Robust Estimation of Location: *M*-Estimate Tools

Let $X_1, \dots, X_n$ be i.i.d. observations from a symmetric distribution F with a density $f(x - \theta)$, where θ is a parameter of location. Without any loss of generality, we set $\theta = 0$.

M-estimates T_n of location were proposed by (Huber 1964)

$$\sum \psi(X_i - T_n) = 0,$$

where $\psi(x)$ is an estimating (score) function.

Robust Estimation of Location: *M*-Estimate Tools

Consider the following particular cases of *M*-estimates:

Least Squares: $\psi_{LS}(x) = x$, $T_n = \bar{x}$;

Least Absolute Values: $\psi_{LAV}(x) = sign(x)$, $T_n = med\, x$;

Maximum Likelihood: $\psi_{ML}(x) = -f'(x)/f(x)$, $T_n = \widehat{\theta}_{ML}$.

An *M*-estimate is a generalization of the maximum likelihood estimate!

Robust Estimation of Location: *M*-Estimate Tools

Under regularity conditions imposed on estimating functions $\psi \in \Psi$ and distribution densities $f \in \mathcal{F}$, *M*-estimates T_n are consistent and asymptotically normal $N(0, V)$ with the asymptotic variance

$$V(\psi, f) = \frac{\int \psi(x)^2 f(x)\, dx}{\left(\int \psi'(x)\, f(x)\, dx\right)^2}.$$

In the case of maximum likelihood efficient *M*-estimate, we get the minimum value of the Cramer-Rao inequality bound:

$$V(\psi_{ML}, f) = \min_{\psi \in \Psi} V(\psi, f) \quad => \quad V(\psi_{ML}, f) = V(-f'/f, f) = \frac{1}{I(f)},$$

where $I(f)$ is Fisher information for location

$$I(f) = \int \left(\frac{f'(x)}{f(x)}\right)^2 f(x)\, dx.$$

Robust Estimation of Location: Huber's Minimax Approach Tools

The minimax solution means that the asymptotic variance $V(\psi, f)$ has the saddle-point $V(\psi^*, f^*)$

$$V(\psi^*, f) \le V(\psi^*, f^*) \le V(\psi, f^*),$$

where

$$V(\psi^*, f^*) = \inf_{\psi \in \Psi} \sup_{f \in \mathcal{F}} V(\psi, f).$$

The right-hand side inequality in the saddle-point double inequality is just the aforementioned Cramer-Rao inequality, whereas the left-hand side one provides **the property of a guaranteed accuracy of estimation.**

Robust Estimation of Location: Huber's Minimax Approach Tools

This property means that there exists the optimal score function ψ^* such that

$$V(\psi^*, f) \leq V(\psi^*, f^*)$$

for any distribution density f in the class $\mathcal{F}$.
The minimax estimating function ψ^* is defined by the maximum likelihood choice for the least favorable (informative) density f^*

$$\psi^*(x) = \psi_{ML}(x) = -f^{*\prime}(x)/f^*(x),$$

which minimizes Fisher information $I(f)$ over the class $\mathcal{F}$

$$f^* = \arg\min_{f \in \mathcal{F}} I(f)\,.$$

Robust Estimation of Location: Huber's Minimax Approach Tools

Example: Huber's minimax solution for the class of ε-contaminated normal distributions (Huber 1964)

$$\mathcal{F}_H = \{f\colon\ f(x) \geq (1-\varepsilon)\varphi(x), \quad 0 \leq \varepsilon < 1\},$$

where $\varphi(x) = (2\pi)^{-1/2} \exp\left(-x^2/2\right)$ is the standard normal density and ε is a contamination parameter.

The least informative density is Gaussian in the center with exponential tails

$$f_H^*(x) = \begin{cases} (1-\varepsilon)\varphi(x) & \text{for} \quad |x| \leq k, \\ (1-\varepsilon)(2\pi)^{-1/2} \exp\left(-k|x| + k^2/2\right) & \text{for} \quad |x| > k. \end{cases}$$

The optimal minimax estimating function is bounded linear

$$\psi_H^*(x) = \max\left[-k, \min(x, k)\right].$$

Robust Estimation of Location: Huber's Minimax Approach Tools

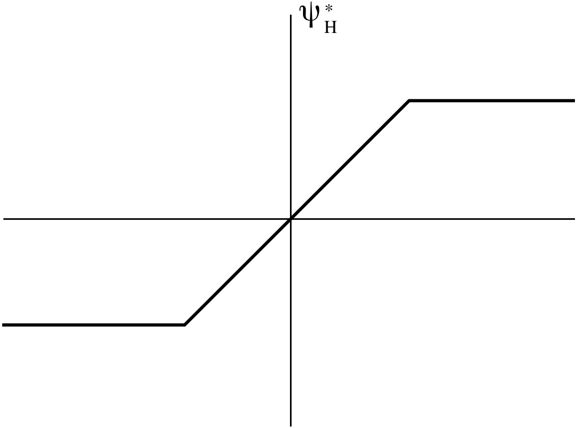

Figure 1: Huber's minimax estimating function

Robust Estimation of Location: Hampel's Influence Function Tools

Let $\{T_n\}$ be a sequence of statistics; $T_n(X)$ denote the statistic from $\{T_n\}$ on the sample $X = (x_1, \ldots, x_n)$, and let $T_{n+1}(x, X)$ denote the same statistic on the sample $(x_1, \ldots, x_n, x)$. Then the function

$$SC_n(x; T_n, X) = (n+1)[T_{n+1}(x, X) - T_n(X)]$$

characterizes the sensitivity of T_n to the addition of one observation at x and is called the **sensitivity curve** for this statistic (Tukey 1977).

In particular,

$$SC_n(x; \overline{x}, X) = x - \frac{1}{n}\sum_{i=1}^{n} x_i = x - \overline{x}$$

for the sample mean $\overline{x}$.

Robust Estimation of Location: Hampel's Influence Function Tools

Let F be a given distribution and $T(F)$ be a functional defined on some set F of distributions, and let the estimate $T_n = T(F_n)$ of $T(F)$ be that functional of the sample distribution function F_n. Then **the influence function** $IF(x; T, F)$ is defined as (Hampel *et al*. 1986)

$$IF(x; T, F) = \lim_{t \to 0} \frac{T((1-t)F + t\Delta_x) - T(F)}{t},$$

where Δ_x is the degenerate distribution at x: $IF(x; T, F)$ **is the Gateaux derivative.**

For the sample mean $\overline{x} = T(F_n) = \int x\, dF_n(x)$, the influence function is

$$IF(x; \overline{x}, F) = x - T(F) = x - \int x\, dF(x).$$

Robust Estimation of Location: Hampel's Influence Function Tools

Under regularity conditions, **the influence function** for the *M*-estimate has the following form (Hampel *et al*. 1986)

$$IF(x; \psi, F) = \frac{\psi(x)}{\int \psi'(x)\, dF(x)}.$$

For *M*-estimates, the relation between the influence function and the estimating function is the simplest.

Main properties of the influence function:

1. Gross-error sensitivity

$$\gamma^*(T, F) = \sup_x |IF(x; T, F)|.$$

Robust Estimation of Location: Hampel's Influence Function Tools

2. Gross-error breakdown point

$$\varepsilon^*(T, F) = \sup\{\varepsilon\colon \sup_{F\colon\, F=(1-\varepsilon)F_0+\varepsilon H} |T(F) - T(F_0)| < \infty\}.$$

This notion defines the largest fraction of gross errors that still keeps the bias bounded (F_0 – an ideal model, H – a contamination): for example, $\varepsilon^*(\bar{x}) = 0$, $\varepsilon^*(med\, x) = 0.5$.

3. Asymptotic variance of M-estimates

$$V(\psi, F) = \int IF(x; \psi, F)^2 dF(x).$$

Robust Estimation of Location: Hampel's Influence Function Tools

Optimal Huberization:
extremal problems of maximization of estimate efficiency under the bounded sensitivity to outliers (Hampel *et al*. 1986)

$$\max_{\psi} eff(\psi, f) \quad \text{under} \quad \gamma(\psi, f) \leq \overline{\gamma}.$$

Example. In the Gaussian case, the optimal solution coincides with the Huber's minimax linear bounded estimating function

$$\psi^*(x) = \psi^*_H(x) = \max\,[-k, \min(x, k)].$$

In general, robust estimates within both Huber's and Hampel's approaches to robustness are close in performance !

Concluding Remarks

Applications in econometrics via robust regression tools:

$$\sum_i \psi^* \left(x_i - \sum_j \phi_{ij} \widehat{\theta}_j \right) = 0.$$

References

Hampel, F.R., Ronchetti, E., Rousseeuw, P.J., Stahel, W.A., 1986. Robust Statistics. The Approach Based on Influence Functions. Wiley, New York.

Huber, P.J., 1964. Robust estimation of a location parameter, Ann. Math. Statist., 35, 1–72.

Huber, P.J., 1981. Robust Statistics. Wiley, New York.

Shevlyakov, G.L., Vilchevski, N.O., 2002. Robustness in Data Analysis: criteria and methods. VSP, Utrecht.

Shevlyakov, G.L., Vilchevski, N.O., 2011. Robustness in Data Analysis. De Gruyter, Boston.

Shevlyakov, G.L., Oja, H., 2016. Robust Correlation: Theory and Applications. Wiley.

THANK YOU !