High-efficiency and robust M-estimates of the scale parameter on the Q-estimate basis

Pavel O. Smirnov , Ivan S. Shirokov , Georgiy L. Shevlyakov

PII: S2405-7223(17)30098-1
DOI: 10.1016/j.spjpm.2017.09.012
Reference: SPJPM 156

To appear in: *St. Petersburg Polytechnical University Journal: Physics and Mathematics*

Received date: 18 September 2017
Accepted date: 18 September 2017

Please cite this article as: Pavel O. Smirnov , Ivan S. Shirokov , Georgiy L. Shevlyakov , High-efficiency and robust M-estimates of the scale parameter on the Q-estimate basis, *St. Petersburg Polytechnical University Journal: Physics and Mathematics* (2017), doi: 10.1016/j.spjpm.2017.09.012

High-efficiency and robust *M*-estimates of the scale parameter on the *Q*-estimate basis

Pavel O. Smirnov, Ivan S. Shirokov, Georgiy L. Shevlyakov
29 Politekhnicheskaya St., St. Petersburg, 195251, Russian Federation

The commonly employed highly efficient and robust *Q*-estimate of the scale parameter proposed by Rousseeuw and Croux (1993) has been approximated using computationally fast Huber *M*-estimates. The suggested *M*-estimates were shown to be robust and highly efficient for an arbitary underlying data distribution due to correctly choosing the approximation parameters. The following indicators of the efficiency and robustness of *M*-estimates of scale were computed: their asymptotic variances, influence functions and breakdown points. Special attention was given to the particular cases of the Gaussian and Cauchy distributions. It is noteworthy that for the Cauchy distribution, the suggested robust estimate of scale coincides with the maximal likelihood estimate. Finally, the computation time of these highly efficient and robust estimates of scale is 3-4 times less than for the corresponding *Q*-estimates.

Key words: *M*-estimate; *Q*-estimate, robustness; scale parameter; Gaussian distribution; Cauchy distribution

Introduction

Estimation of the scale parameter is one of the most important problems in statistical analysis [1, 2, 6, 8, 10]. The highly efficient robust Q_n-estimate of the scale parameter (n is the sample size) is currently considered the best [7]. This estimate is defined as the first quartile of the distance between observations:

$$Q_n = c\{|x_i - x_j|\}_{(k)},$$

where c is the constant providing the consistency of the estimate; $k = C_h^2$ ($h = [n/2] + 1$).

The Q_n-estimate is robust, with the highest possible breakdown point $\varepsilon^* = 0.5$ and a high asymptotic efficiency for a normal distribution (82%). Its disadvantages include high asymptotic complexity of the computation algorithm, requiring $O(n\ln(n))$ time and the same amount of memory.