APIC'95, El Paso, Extended Abstracts, A Supplement to the international journal of Reliable Computing

On the Choice of an Optimization Criterion under Uncertainty in Interval Computations - Nonstochastic Approach

G. L. Shevlyakov and N. O. Vil'chevskiy

To compute tolerance intervals for location and regression parameters, it is necessary to first get the approximate estimates of these characteristic. One of the main statistical approaches to obtaining these estimates consists of using an optimization criterion, e.g., least squares, least modules, Chebyshev, etc. [1]. From the viewpoint of estimating tolerance intervals, this optimization approach has the following additional advantage: not only we can get the estimate itself by finding the values of the parameters Θ in which the criterion $J(\Theta)$ attains its minimum, but we can define the tolerance interval as the set of all the values of Θ for which the value $J(\Theta)$ is sufficiently small (i.e., is $\le J_0$ for an appropriately chosen J_0).

In traditional mathematical statistics, the choice of an optimization criterion is uniquely determined by the choice of an appropriate stochastic model, as in the cases of maximum likelihood, robust or nonparametric approaches [2]. The situation is different, however, in nonstochastic approach to data analysis, when there are no (objective) reasons to use any stochastic model. This situation occurs, e.g., when we are processing unique data samples: this is a typical situation with medical data.

In this paper, we show that it is possible to reduce an apriori uncertainty connected with the choice of an optimization criterion, if we impose natural requirements on M-estimates of location [3, 4]. In some cases, these requirements lead to the unique choice of the optimization method.

In our considerations, we do not assume any stochastic model [4, 5].

I. Univariate Case

An *M-estimate* of the location is defined as

$$\hat{\theta}_n = \arg\min_{\theta} \sum_{i=1}^{n} \rho(x_i - \theta), \qquad (1)$$

where:

- $x_1, ..., x_n$ is a given sample, and
- $\rho(u)$ is a function called a *loss function.*

The author is with Department of Mathematics, St. Petersburg State Technical University, St. Petersburg, 195251, Russia.

The following result is known for M-estimates:

Theorem [3] *If an M-estimate satisfies the following three requirements:*

(i) convexity of $\rho(u)$;

(ii) central symmetry

$$\hat{\theta}_n(-x_1, ..., -x_n) = -\hat{\theta}_n(x_1, ..., x_n);$$

(iii) equivariance to scale transformations $\hat{\theta}_n(\lambda x_1, ..., \lambda x_n) = \lambda \hat{\theta}_n(x_1, ..., x_n),$

then this M-estimate coincides with the M-estimate defined by a loss function $\rho(u) = |u|^p$ for some $p \ge 1$.

Comment. In other words, we get L_p-estimates.

Theorem 1 *The conditions (i), (ii), and equivariance to monotone transformations:*

$$\hat{\theta}_n(f(x_1), ..., f(x_n)) = f(\hat{\theta}_n(x_1, ..., x_n)),$$

where $f(x)$ is an arbitrary strictly monotone function, yield L_1-estimate with $\rho(u) = |u|$.

Theorem 2 *The requirements of differentiability and of minimum norm of the sensibility of estimates to small data deviations*

$$\rho^* = \arg\min_{\rho} \sum_{i=1}^{n} (\frac{\partial \hat{\theta}_n}{\partial x_i})^2$$

yields L_2-estimate with $\rho^*(u) = u^2$.

Theorem 3 *The requirement of recurrency*

$$\hat{\theta}_{n+1} = \hat{\theta}_n + \gamma_n \psi(x_{n+1} - \hat{\theta}_n)$$

yields L_2-estimate with $\rho(u) = u^2,\ \psi(u) = u,\ \gamma_n = n^{-1}$.

II. Multivariate Case

For multivariate data

$$X_1, ..., X_n,\ X_i = (x_{i1}, ..., x_{im}),$$

M-estimate of the mean values is defined as

$$\hat{\Theta}_n = \arg\min_{\Theta} \sum_{i=1}^{n} \rho(x_{i1} - \theta_1, ..., x_{im} - \theta_m),$$

where $\hat{\Theta}_n = \hat{\Theta}_n(X_1, ..., X_n) = (\hat{\theta}_1, ..., \hat{\theta}_m)$ is a vector of estimates, and $\rho(u_1, ..., u_m)$ is a function of m variables called a *loss function*.

In the multivariate case, the requirements of equivariance to scale transformations can be formulated in several different ways:

Theorem 4 *If an M-estimate satisfies the following requirements:*

- *equivariance to independent scale transformations of vector components:*
$$\hat{\Theta}_n(\tilde{X}_1, ..., \tilde{X}_n) = (\lambda_1\hat{\theta}_1, ..., \lambda_m\hat{\theta}_m),$$
$$\tilde{X}_i = (\lambda_1 x_{i1}, ..., \lambda_m x_{im});$$
- *convexity of* ρ*;*
- *central symmetry of vector components*
$$\hat{\theta}_j(-x_{1j}, ..., -x_{nj}) = -\hat{\theta}_j(x_{1j}, ..., x_{nj}),$$
$$j = 1, ..., m;$$
- *piece-wise differentiability of* $\psi_j = \partial\rho/\partial u_j$, $j = 1, ..., m$*;*
- *boundary conditions*
$$\rho(u_1, ..., u_{l-1}, 0, u_{l+1}, ..., u_m) = \sum_{j \neq l}^{m} A_j |u_j|^{p_j},$$
$$A_j > 0,\ p_j \geq 1,\ l = 1, ..., m \qquad (2)$$

then

$$\rho(u_1, ..., u_m) = \sum_{j=1}^{m} A_j |u_j|^{p_j},\ A_j > 0,\ p_j \geq 1.$$

Theorem 5 *The equivariance of M-estimate to the scale transformation*

$$\hat{\Theta}_n(\lambda X_1, ..., \lambda X_n) = \lambda\hat{\Theta}_n(X_1, ..., X_n) \qquad (3)$$

implies that the loss function is homogeneous:

$$\rho(\lambda u_1, ..., \lambda u_m) = |\lambda|^p \rho(u_1, ..., u_m).$$

Theorem 6 *The equivariance of M-estimates to arbitrary orthogonal transformations of data*

$$\tilde{X}_i = O_m X_i,\ i = 1, ..., n;$$
$$\hat{\Theta}_n(\tilde{X}_1, ..., \tilde{X}_n) = O_m\hat{\Theta}_n(X_1, ..., X_n), \qquad (4)$$

where O_m *is an orthogonal matrix of dimension* m*, yields the loss functions of the form*

$$\rho(u_1, ..., u_m) = F(\sum_{j=1}^{m} u_j^2);$$

where $F = F(v^2)$ *is an arbitrary convex function.*

Theorem 7 *If an M-estimate satisfies either the condition (4) and one of the conditions (2) or (3), then it is a multivariate* L_p*-optimization criterion*

$$\rho(u_1, ..., u_m) = Ar^p,\ A > 0,\ p \geq 1,$$
$$r = (\sum_{j=1}^{m} u_j^2)^{1/2}. \qquad (5)$$

Another characterization of criterion (5) was obtained in [6]:

Theorem 8 *The equivariance of M-estimates to arbitrary affine transformations yields*

$$\rho(u_1, ..., u_m) = \sum_{j=1}^{m} u_j^2.$$

References

[1] J. R. Rice, **The Approximation of Functions. Linear Theory. Vol. 1,** Addison Wesley, New York, 1964.

[2] J. W. Tukey, "The Future of Data Analysis", *Ann. Math. Statist.*, 1967, Vol. 33, No. 1, pp. 1–67.

[3] P. J. Bickel and E. L. Lehmann, "Descriptive Statistics for Nonparametric Models. 1. Introduction", *Ann. Statist.*, 1975, Vol. 3, pp. 1045–1069.

[4] V. Ya. Kreinovich, "A General Approach to Analysis of Uncertainty in Measurements", *Proceedings of the the 3-rd USSR National Symposium on Theoretical Metrology*, Leningrad, Mendeleev Metrology Institute (VNIIM), 1986, pp. 187–188 (in Russian).

[5] G. L. Shevlyakov and N. O. Vil'chevskiy, "Axiomatic Approach to the Choice of an Optimization Criterion," *Proceedings of the 6-th Seminar on Nonparametric and Robust Methods in Cybernetics*, Tomsk University Press, Tomsk, 1987, Part 1, pp. 93–97 (in Russian).

[6] W. Gehrig and K. Hellwig, "Eine Charakterisierung der gewichteten L_r-Distanz", In: **OR-Spektrum**, Springer Verlag, Berlin - Heidelberg - New York, 1982, pp. 233–237.