Data Fitting under Interval Error: Centroids of Feasible Parameter Set

Sergei Zhilin[1,2]

[1] Altai State University,
[2] CSort LLC,
Barnaul, Russia,
sergei@asu.ru

The classical approach to data fitting is based on stochastic assumptions about the measurement error, and provides estimates that have random nature. In contrast, interval data fitting only assumes the knowledge of deterministic error bounds, and allows obtaining of guaranteed interval estimates for a feasible parameter set (FPS) [1]. Along with interval estimates, in real-life applications, it is often important to choose point estimate in FPS [2].

Generally speaking, all points in FPS are equally possible as we have no information to prefer one or another member of FPS. Nevertheless, in practice some more or less grounded recipes are employed. Most of them recommend to use one or another so called center point or centriod of FPS as a point estimate.

Our work is aimed at the study of properties of estimates constructed as FPS centroids for fitting linear-parameterized dependency $y = x\beta$ ($y \in \mathbb{R},\ x, \beta \in \mathbb{R}^p$) to a set of observations $D_N = \{(x_j, y_j, \varepsilon_j) \mid j = 1, \ldots, N\}$ where inputs x_j are observed without errors while output y_j is measured with the error absolutely bounded by ε_j. The following FPS centroids are considered:

- outer interval estimate center (center of FPS minimum bounding box);
- Chebyshev center (minimax or worst-case estimate);
- L_1-center;
- Oskorbin center (restricted L_1-center);
- center of maximum length diagonal;
- analytic center;
- point minimizing average distance to verticies;
- gravity center;
- Fréchet mean;
- maximum density point for interval outer estimate of FPS.

The study is based on the extensive numerical simulation. For some known p, N and ground truth value β^* we simulate data set D_N where error of y_j has known distribution (uniform, triangle, truncated normal) and compute centroids from the above list. Using centroids as point estimates $\hat{\beta}$ we study the accuracy of estimates (proximity to β^*), their variation and behavior depending on dimension p, observations number N, radius of interval error ε.

The results of simulation study show that in most number of cases analytic center and Oskorbin center give better point estimates than other FPS centroids, i.e. they provide the lowest mean and median distance between $\hat{\beta}$ and β^* as well as the lowest standard deviation and interquartile range.

References

1. Kantorovich, L.V.: On some new approaches to numerical methods and processing observation data. Sib. Math. J. 3(5), 701–709 (1962) (in Russian)
2. Voshchinin, A.P., Bochkov, A.F., and Sotirov, G.R.: Method of Data Analysis for Interval Nonstatistical Error. Zav. Lab. 56(7), 76–81 (1990) (in Russian)