Robust Multivariate Tolerance Region: Influence Function and Monte Carlo Study

A. Farall[1] and G. Boente[2]

[1] Fundación Antorchas, Argentina.
[2] CONICET and Departamento de Matemática and Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Ciudad Universitaria, Pabellón 1. C1428EHA, Buenos Aires, Argentina.

Keywords: Tolerance regions, Robust estimation, Donoho–Stahel estimators, Influence function.

Abstract

We define a class of multivariate tolerance regions which turn out to be more resistant than the classical ones to outliers. More precisely, let $\mathbf{t}_n = \mathbf{t}_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $\mathbf{V}_n = \mathbf{V}_n(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ be robust location and scatter estimates. Define the region

$$\mathcal{R}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \left\{\mathbf{y} : (\mathbf{y} - \mathbf{t}_n)' \mathbf{V}_n^{-1} (\mathbf{y} - \mathbf{t}_n) \le K\right\} , \tag{1}$$

where the constant K is choosen such that

$$p_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = P_{n,\boldsymbol{\theta}}\left[P_{\boldsymbol{\theta}}(\mathbf{x} \in \mathcal{R}(\mathbf{x}_1, \ldots, \mathbf{x}_n) | \mathbf{x}_1, \ldots, \mathbf{x}_n) \ge q\right] \ge \delta \quad \forall \quad \boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\Sigma}) ,$$

with $\mathbf{x} \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\mathbf{x}_i \sim N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $1 \le i \le n$, independent, $P_{n,\boldsymbol{\theta}}$ the distribution of $(\mathbf{t}_n, \mathbf{V}_n)$ and $P_{\boldsymbol{\theta}}$ that of $\mathbf{x}$. We will then say that $\mathcal{R}$ is *a robust tolerance region*.

The tolerance factors are evaluated using a simulation study under the central model and the sensitivity to deviations of the normal distribution for moderate samples is studied through a Monte Carlo study.

The coverage probability (Pc), for a given confidence level δ fixed, a fixed theoretical coverage q, and a tolerance factor K, i.e,

$$\begin{aligned} Pc(G, F) &= P_F\left((\mathbf{x} - \mathbf{T}(G))' \mathbf{V}(G)^{-1} (\mathbf{x} - \mathbf{T}(G)) \le K\right) \text{ with } \mathbf{x} \sim F \\ &= \int I_{\mathcal{R}(G)}(\mathbf{x})\, dF(\mathbf{x}) \end{aligned}$$

where

- $\mathcal{R}(G) = \{\mathbf{x} : (\mathbf{x} - \mathbf{T}(G))' \mathbf{V}(G)^{-1} (\mathbf{x} - \mathbf{T}(G)) \le K\}$
- $\mathbf{T}(G)$ is the location functional at the distribution G,
- $\mathbf{V}(G)$ is the scatter functional at G and
- K is the tolerance factor, which is assumed to be fixed.

The influence function of the coverage probability is derived and allows to compare the sensitivity of different proposals to anomalous data.

References

G. Canavos and I. Koutraouvaelis (1984). The robustness of two sided tolerance limits for normal distributions. *J. Qual. Technol.*, 144-149.

L. Fernholz and J. Gillespie (2001). Content-corrected tolerance limits based on bootstrap. *Technometrics*, 43, 147-155.

L.T. Fernholz (2002). Robustness Issues regarding Content Corrected Tolerance Limits. *Metrika*, 53-66.

D. Fraser and I.Guttman (1956). Tolerance regions. *Ann. Math. Statist.*, 162-179.