

TEST (2015) 24:471–472
DOI 10.1007/s11749-015-0453-3



DISCUSSION

Comments on: Robust estimation of multivariate location and scatter in the presence of cellwise and casewise contamination

Ricardo A. Maronna¹

Published online: 27 June 2015

© Sociedad de Estadística e Investigación Operativa 2015

The authors are to be praised for a very original and elegant contribution to a difficult and important problem.

I would like to point out that one of the fields where the need to consider cellwise outliers arose was computer vision. As far as I know, the efforts in this direction have mainly been concentrated on principal components. Here, the goal is to develop estimators that are resistant to both types of outliers, and also fast enough for high-dimensional data. Before the formalization by [Alqallaf et al. \(2009\)](#), an early proposal is due to [De la Torre and Black \(2001\)](#). [Maronna and Yohai \(2008\)](#) proposed a “perturbed MM estimator”. [Candés et al. \(2011\)](#) propose a quite different approach, which is very fast and is resistant to cellwise outliers, but not to casewise ones.

Estimation of a scatter matrix presents difficulties of a different nature than those in principal components, and the authors have dealt with them very skillfully.

I have just two comments for the authors:

In Step I, instead of declaring the univariate outliers as missing, one could somehow impute them (eventually correcting the imputation after Step II). This would make it possible to apply in Step II a “standard” robust multivariate estimator, with considerable gains in computing speed.

In Step II, a bisquare S estimator is employed. [Rocke \(1996\)](#) showed that the normal efficiency of these estimators tends to the one with increasing dimension, but this gain is paid for by a decrease in robustness. It would be interesting if the procedure by

This comment refers to the invited paper available at: doi:[10.1007/s11749-015-0450-6](https://doi.org/10.1007/s11749-015-0450-6).

✉ Ricardo A. Maronna
rmaronna@retina.ar

¹ Departamento de Matemática Facultad de Ciencias Exactas,
Universidad Nacional de La Plata, Calle 50 y 115,
1900 La Plata, Argentina

Danilov et al. could be adapted to use estimators with a controllable efficiency such as τ -estimators (Lopuhaä 1991) or MM-estimators (Tatsuoka and Tyler 2000), which would ensure a gain in robustness for high dimensions.

References

- Alqallaf F, Van Aelst S, Yohai VJ, Zamar RH (2009) Propagation of outliers in multivariate data. *Ann Stat* 37:311–331
- Candès EJ, Li X, Ma Y, Wright J (2011) Robust principal component analysis. *J ACM* 58(3):11
- De la Torre F, Black MJ (2001) Robust principal components analysis for computer vision. In: Proceedings of international conference on computer vision, 2001. <http://citeseer.ist.psu.edu/torre01robust.html>
- Lopuhaä HP (1991) Multivariate τ -estimators for location and scatter. *Can J Stat* 19:307–321
- Maronna RA, Yohai VJ (2008) Robust lower-rank approximation of data matrices with element-wise contamination. *Technometrics* 50:295–304
- Rocke D (1996) Robustness properties of S-estimators of multivariate location and shape in high dimension. *Ann Stat* 24:1327–1345
- Tatsuoka KS, Tyler DE (2000) On the uniqueness of S-functionals and M-functionals under nonelliptical distributions. *Ann Stat* 28:1219–1243