An L1-type estimator of multivariate location and shape

E. Roelant[1], and S. Van Aelst[1]

[1] Ghent University, Department of Applied Mathematics and Computer Science, Krijgslaan 281 S9, B-9000 Ghent, Belgium

Keywords: Multivariate location and shape.

The proposed estimator is a location and shape estimator which generalizes the L1-idea to a multivariate context. Consider a sample $x_1, \ldots, x_n$ of p-variate observations. Then the estimator is defined as the solution $(\hat{\mu}, \hat{V})$ that yields the minimum of the sum of the distances $d_i(\mu, V) = \sqrt{(x_i - \mu)' V^{-1} (x_i - \mu)}$, minimized under the constraint that V has determinant 1. The constraint $\det(V) = 1$ implies that we will get an estimate for the shape of the data cloud. To compute these estimates we use an Iteratively Reweighted Least Squares algorithm.

We can also consider the estimator $(\hat{\mu}, \hat{\Sigma})$ which is obtained as the solution to the problem of minimizing $\det \Sigma$ subject to the constraint

$$\frac{1}{n} \sum_{i=1}^{n} \sqrt{(x_i - \mu)' \Sigma^{-1} (x_i - \mu)} = b_0$$

among all $\mu \in \mathbb{R}^p$, Σ in the class of positive definite symmetric matrices PDS(p) and for some $b_0 > 0$. Both estimators are equivalent in the sense that $\hat{V} = \hat{\Sigma}/((\det \hat{\Sigma})^{1/p})$.

If the observations come from an elliptical distribution with a density of the form

$$\det(\Sigma)^{-1/2} f[(x - \mu)' \Sigma^{-1} (x - \mu)]$$

then the constant b_0 can be chosen equal to $E_{\mu,\Sigma}[\sqrt{(X_1 - \mu)' \Sigma^{-1} (X_1 - \mu)}]$. This choice yields a consistent scatter estimator $\hat{\Sigma}$ for Σ.

Using results of Lopuhaä (1989) and Van Aelst and Willems (2005), we study the properties of the L1-type multivariate location, shape and scatter estimators defined above. We investigate the influence function and asymptotic variances of the estimators at elliptical distributions. We compare the efficiencies of the L1-type estimators with those of the classical estimators for the location and shape at multivariate normal and multivariate t-distributions. We can conclude that we get better efficiencies for multivariate t-distributions.

To get an idea about the robustness of the L1-type estimator, we compare influence functions of location and shape estimators by looking at the γ-functions for the location as defined by Ollila, Oja and Hettmansperger (2002) and the α-functions for the shape/scatter matrix (Salibián-Barrera, Van Aelst, and Willems 2004, Croux and Haesbroeck 2000). These two functions also determine the robustness properties of most multivariate analysis methods such as principal components, multivariate regression, and canonical correlations when based on the corresponding location and shape or scatter estimator. The γ- and α-functions of the L1-type estimator are compared with those of the classical estimator, the MCD estimator and the biweight S-estimator. Finally we give some real data examples to illustrate the behavior of the estimator in practice.

References

C. Croux and G. Haesbroeck (2000). Principal Component Analysis based on Robust Estimators of the Covariance or Correlation Matrix: Influence Functions and Efficiencies. *Biometrika*, 87, 603–618.

H.P. Lopuhaä (1989). On the Relation between S-estimators and M-estimators of Multivariate Location and Covariance. *The Annals of Statistics*, 17, 1662–1683.

E. Ollila, H. Oja and T.P. Hettmansperger (2002). Estimates of Regression Coefficients based on the Sign Covariance Matrix. *Journal of the Royal Statistical Society Series B*, 64, 447–466.

M. Salibián-Barrera, S. Van Aelst and G. Willems (2004). PCA based on Multivariate MM-Estimators with Fast and Robust Bootstrap. Submitted.

S. Van Aelst and G. Willems (2005). Multivariate Regression S-estimators for Robust Estimation and Inference. *Statistica Sinica*, to appear.