Robust Independent Component Analysis

M. Hubert[1], G. Brys[2] and P. Rousseeuw[3]

[1] Department of Mathematics, Katholieke Universiteit Leuven, W. de Croylaan 54, B-3001 Leuven
[2] Faculty of Applied Economics, University of Antwerp, Prinsstraat 13, B-2000 Antwerp
[3] Department of Mathematics and Computer Sciences, University of Antwerp, Middelheimlaan 1, B-2020 Antwerpen.

Keywords: PCA, Adjusted boxplot, Blind source separation.

1 Introduction

Independent Component Analysis (ICA) is a statistical method for transforming multidimensional random vectors into components which are statistically as independent from each other as possible. It can be seen as a generalization of Principal Component Analysis, which seeks for uncorrelated factors. In recent years, many algorithms were proposed that perform well in many situations, but lack robustness. We propose a three-step robust ICA procedure. After a robust preprocessing stage, we remove outliers by applying a new outlier rejection rule for skewed multivariate data. Next, we apply the well-known FASTICA method on the clean data set.

2 Independent Component Analysis

The theory of independent component analysis (or blind source separation) is nicely exposed in Hyvärinen et al. (2001). The model states that

$$x_i = As_i \quad i = 1, \ldots, n$$

with x_i the p-dimensional observed data, A the mixing matrix and s_i the unknown independent sources. Typical examples are signals which have been measured by p recorders at n time points, so $p < n$.

The aim of ICA is to find the unmixing matrix $B(= A^{-1})$ such that

$$y_i = Bx_i \approx s_i.$$

It is assumed that the s_i are not jointly normally distributed, since otherwise uncorrelated components are necessarily independent, and then the mixing matrix A can not be determined. As the $y_i = Bx_i = BAs_i$ are linear combinations of the s_i, they tend to be more normal than the s_i according to the CLT. Therefore ICA is concerned with looking for directions with maximal non-gaussianity.

3 Robust ICA

An ICA procedure often starts with preprocessing my means of PCA, as this already yields uncorrelated components. In our robust procedure, we will use the MCD-estimator for this purpose.

Next we try to detect multivariate outliers. Outlier detection rules based on normality assumptions (or more general, elliptical assumptions) are not very accurate here, as the data are assumed not to be normally distributed. Hence, we apply a projection pursuit procedure, with a projection index based on the adjusted boxplot for skewed univariate data (Vandervieren and Hubert, 2004).

Finally, we apply the FASTICA algorithm (Hyvärinen and Oja, 1997) on the clean data. Several simulation results and examples are presented to show the robustness of this approach.

References

A. Hyvärinen, J. Karhunen and E. Oja (2001). *Independent Component Analysis.* John Wiley and Sons, New York.

A. Hyvärinen and E. Oja (1997). A Fast Fixed-Point Algorithm for Independent Component Analysis. *Neural Computation*, 9, 1483-1492.

E. Vandervieren and M. Hubert (2004). An Adjusted Boxplot for Skewed Distributions. In: J. Antoch, editor, *COMPSTAT 2004, Proceedings in Computational Statistics,* pp. 1933-1940, Springer-Verlag, Heidelberg.